Record labels, competition, and diversity in Spotify Charts:

Evidence from ten Latin American countries

Student name: Miguel Tomas-Miranda

Student number: 620721

Supervisor: Dr. Christian W. Handke

Cultural Economics and Entrepreneurship

Erasmus School of History, Culture, and Communication

Erasmus Universiteit Rotterdam

Master thesis

December 18th, 2023

Word count: 21,264

Abstract

Over the past few years, the Latin American music industry has experienced significant growth in the generation of royalties from digital music. Nonetheless, it remains underexplored academically, especially regarding the influence of major record labels such as Universal, Sony, and Warner on market concentration, product standardization, and diversity. By integrating three databases, two from Spotify and one we constructed, covering ten Latin American countries, this study aimed to address two key issues: the evolution of concentration and inequality in Spotify's music consumption, and the contribution of different record labels to diversity in popularity charts. Through descriptive and inferential statistics, including fixed-effects panel data regressions, we obtained exciting findings that shed light on the dynamics of the region. Firstly, we found evidence of overall decreasing concentration levels over time on a label-level, although this result was overshadowed by the upper portion of the distribution retaining a massive proportion of streams. Secondly, we highlighted the varied contribution to diversity by record labels, controlling for other macroeconomic and technological variables, with Universal Music Group being the only entity with a positive correlation. As a plausible explanation, we hypothesise on the relevance of contract types between artists and record labels —production versus distribution—, which our study didn't capture. We conclude with a discussion on the case of Brazil, the spurious long-tail phenomena, and offer further insights on the adoption of monetisation caps on the side of the streaming platforms, and future avenues of research in this thriving field, mainly focusing on more sophisticated modelling that incorporate possible non-linearities when relating success to the songs' musicological characteristics.

Keywords: Latin America, superstar effects, long-tail, music industry, concentration, inequality, diversity, Spotify, record labels

Resumen

Durante los últimos años, la industria musical latinoamericana ha experimentado un crecimiento significativo en cuanto a la generación de regalías en el ámbito digital. Sin embargo, aún se mantiene poco explorada académicamente, especialmente en lo que respecta a la influencia que tienen los principales sellos discográficos como Universal, Sony y Warner en la concentración del mercado, la estandarización de productos y la diversidad. Mediante la integración de tres bases de datos —dos de Spotify y una que construimos— y abarcando diez países latinoamericanos, este estudio tuvo como objetivo abordar dos temas clave: la evolución de la concentración de mercado y la desigualdad en el consumo de música de Spotify y la contribución de las *major labels* a la diversidad en sus ránquines. A través de estadísticas descriptivas e inferenciales, incluidas regresiones de datos de panel con efectos fijos, obtuvimos hallazgos interesantes que dan luces sobre las dinámicas de la región latinoamericana. En primer lugar, encontramos evidencia de concentración decreciente a lo largo del tiempo a nivel de los sellos discográficos, aunque este resultado se vio eclipsado debido a que la parte superior de la distribución retuvo una proporción masiva de reproducciones. En segundo lugar, destacamos la variada contribución a la diversidad por parte de los sellos discográficos, controlando por otras variables macroeconómicas y tecnológicas, siendo Universal Music Group la única entidad con una correlación positiva. Como explicación plausible, planteamos una hipótesis relacionada con la relevancia de los tipos de contratos entre artistas y sellos discográficos —producción versus distribución— que nuestro estudio no capturó. Concluimos con una discusión sobre el caso de Brasil, los espurios fenómenos de cauda larga, y ofrecemos algunas reflexiones sobre la adopción de umbrales de monetización por parte de las plataformas de *streaming* y futuros caminos de investigación en esta área en desarrollo, centrándonos principalmente en modelos estadísticos más sofisticados que incorporen posibles no linealidades al relacionar el éxito con las características musicológicas de los fonogramas.

Palabras clave: Latinoamérica, superestrellato, cauda larga, industria musical, concentración, desigualdad, diversidad, Spotify, sellos discográficos

This work is dedicated to my family, whose support and encouragement have been fundamental in helping me navigate through the many efforts that writing this dissertation entailed.

And to my supervisor, Dr. Christian W. Handke, for his extraordinary patience and guidance, even when the journey extended far beyond expected timelines. His support during this challenging yet illuminating process has been invaluable.

Table of Contents

1. Introduction

Concentration and inequality are topics of interest for cultural economists while also representing a concern for creative practitioners. Historically, there have been notoriously high concentration levels within the recorded music industry, showcasing the predominance of major record labels in producing, distributing, and disseminating records. In recent years, however, digitisation and other disintermediation processes have emerged as a way to potentially mitigate inequality and provide independent artists with a more advantageous position. This goes with the notion that indie labels have overall demonstrated better compatibility with the tastes and preferences of the public through the years.

Within the academic field, most of the available studies have been conducted in developed countries. In this regard, and despite its continuous growth for thirteen consecutive years, there have been relatively few efforts to characterise the Latin American music industry's structure and dynamics. Little is known about the majors' role in this context and whether the [potential] occurrence of superstar and long-tail effects is comparable to Western countries. Despite the relative obscurity of this topic, the recent availability of high-quality, country-differentiated datasets resulting from massive music consumption within streaming services provides a unique and exciting opportunity to study such relatively unexplored areas of the music industry.

This study aims to better understand the industrial organisation economics and demand/ supply characteristics of the global music industries by analysing a component of the Latin American recording markets: streaming services. Given the distinctive features of Latin American countries in terms of cultural production and consumption, the notion that music communicates highly-context-dependant messages, and assuming a variegated impact of the major labels within these countries in terms of market concentration and diversity, we aim to uncover some of the differential points of the recording industry in Latin America through answering the following main research questions: *How have label concentration and ownership of streams evolved within the Latin American digital music charts over the 2016-2023 period?*

Regarding the type of music supplied based on its musicological characteristics, to what extent do specific record label types and major conglomerates contribute to the diversity of consumed products?

We employed several metrics to assess concentration and inequality, utilising a high-quality, custom-built label database to categorise record companies. For characterising the music, we harnessed the audio features provided by the Spotify API, which elucidate both the musical and sonic attributes of the tracks hosted on the platform. In terms of diversity and considering the characteristics of our main database, we focused on two of the three parameters proposed by Stirling (1998) — balance and disparity — with a particular interest in song-level indicators[1]. Regarding the longitudinal evolution of chart diversity, we identified only a slight upward trend concerning consumed song diversity throughout the study's duration. However, intriguing correlations emerged between label affiliations and our preferred diversity indicator, the Rao-Stirling Index (RSI). We performed two fixed effects regressions on aggregate weekly data to address the relationship between consumed diversity and label dominance across all countries, controlling for macroeconomic and platform-specific confounding variables. Our findings support the assertion that long-tail effects become noticeable when online music services attain a particular level of penetration, however we hypothesise that the relative weakening of the "head" is explained by revenue dilution for the entire supply landscape, which we could not test because of data constraints.

[1] The variety property was fixed at 200 entities (rank positions) for the entire panel dataset.

2. Ownership of streams and concentration – Literature review

In this section, we will consider the fundamental theoretical notions that will be used throughout this work and offer a critical review of the existing empirical studies related to our topic of interest.

2.1. The current international music industry

Among the creative industries, music has received relatively significant attention, despite its modest contribution to the economy (Towse, 2019). The music industry comprises a series of economic activities generally divided into three main areas: the recording industry, live concerts, and publishing. The recorded music industry has traditionally exploited products derived from master recordings. During most of its history, the primary format for its commercialisation was a physical product. However, due to digitisation, consumption has shifted towards non-physical products, most substantially through consumers' adoption of streaming services, with companies such as Spotify, Apple Music and Tencent Music leading the market (RouteNote, 2022). The recording sector lost a considerable proportion of its revenues at the turn of the century, but streaming platforms have taken a central role in its recent recovery (Krueger, 2019).

We often consider major record labels the most influential entities within the recorded music industry. Nowadays, there are three conglomerates: Universal Music Group (UMG), Sony Music Entertainment (SME) and Warner Music Group (WMG). Due to their high individual market share, these three companies are the basis of the prevailing oligopoly. Their power includes resources that smaller companies cannot readily access, especially in terms of distribution reach and the possibility of increasing the visibility of their catalogue artists. Division of labour into highly specialised departments has provided them with an expanded reach and nowadays it is not unusual for major labels to capitalise on revenue sources beyond the exploitation of sound recordings, particularly through 360 deals.

Towse (2019) mentions that music is "an industry whose production and international trade are concentrated in developed countries" (p. 479). Accordingly, much of the existing academic literature and statistical reports by private entities are condensed around these economies. Prominent scholarly work has often concentrated on Western nations, notably in countries such as the United States. While data sources such as Pollstar and Nielsen SoundScan provide some quantitative insights, the availability of comprehensive information of such nature is limited, which Krueger (2019) attributes largely to the reluctance of private companies to disclose financial details. This issue is compounded by the inherent challenges of forecasting success in the music industry (Cameron, 2016), which aligns with the 'nobody knows' principle, establishing uncertainty as a fundamental aspect of the creative industries (Caves, 2000).

2.2. The Latin American music industry

There has been a more limited chance to study the music industry's structure and dynamics in contrasting realities such as the Latin American countries. Information sources are usually more fragmented compared to the Western world, especially regarding live events. Despite this, the proliferation of digital means of consumption, including downloads and streaming services, generates massive quantitative data that can be analysed to better comprehend the supply and demand dynamics of digital phonographic markets.

When discussing the Latin American recording industry, we have two possible focuses: the Western Latin sector (predominantly corresponding to the United States) and the Latin American market. Nonetheless, given our ambitions for this study, we will solely present a succinct overview of the local Latin American music industry. We believe that delving into this area represents a more significant contribution that may open a myriad of possibilities for academic research within these emerging markets.

2.2.1. Local Latin American markets: A history with numerous challenges

The local Latin markets have not been exempted from diverse challenges, some of them hindering the investment of foreign labels. For instance, historically, there have been all sorts of operational difficulties in South America due to the ups and downs of its economies, political instability (Bernstein & Weissman, 2007) and problems with royalty collections (Fisher, 1980). Also, in the early 2000s, the Latin recording markets –as well as the rest of the world– were experiencing a dramatic decline in revenues, while the IFPI frowned upon their prevailing piracy practices and ineffective copyright enforcement policies (IFPI, 2000; IFPI, 2001; IFPI, 2002). The overwhelming piracy rates were even blamed for Mexico's dropping out from the top ten global markets in 2003 (IFPI, 2003).

Figure 1. Constructed with data from IFPI (2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023) and Smirke (2014). To the best of our knowledge, IFPI does not use inflation-adjusted numbers.

Less than two decades ago, Latin America only represented a minor role in the global recording industry (Bernstein & Weissman, 2007). However, a more encouraging future was yet to come: In recent years, the International Federation of the Phonographic Industry (IFPI) has recognised its growing importance, which has shifted Latin America to the position of a robust market, especially regarding digital products. Since 2010, and "[a]fter almost 17 years of continuous decline" (IFPI, 2018), the Latin American region has experienced sustained growth, particularly in digital revenues (IFPI, 2014). According to a recent report issued by this organisation using nominal values, in 2022, "[r]evenues in Latin America rose by 25.9%" (IFPI, 2023, p. 14). During this year "[t]he Latin American region saw its thirteenth consecutive year of growth with its market now worth US$1.3 billion" (p. 41). To further illustrate this point, we provide growth rates of the Latin American recording industry during the 2012-2022 period (Figure 1). Unfortunately, no absolute values were found in the consulted reports, therefore only percentual changes are displayed. Despite this, a clear upward pattern is visible, with values being higher compared to Western countries' growth in digital music[2].

2.3. Record labels

The incorporation of record labels to the cultural economics academic agenda has profound implications: it has been proposed that research on record labels is capable of helping unveil the tensions between art and commerce in the music industry (Mall, 2018). Further, record labels play a pivotal role in shaping the structure of the music industry. A record label is a company that oversees one or more of the aspects related to the creation/exploitation of a musical product. In the traditional model, record labels "usually sign[ed] the artist to a recording contract promising to pay the artist a royalty for recordings sold in return for the artist's promise to record exclusively for that particular label" (Hull et al, 2011, p. 20). However, labels can also avoid the risk of production investment, and focus only on the distribution of the record. Moreover, in the past few years, major record labels have purchased digital aggregators such as The Orchard by SME (Ingham, 2015), AWAL by SME (SME, 2022b) and, more recently, the Spanish company Altafonte by SME, which had significant presence in Latin America (Casado, 2023).

2.3.1. Shaping the industry through mergers and acquisitions

The history of record labels contributes to the understanding of the recording industry. In the dawn of phonography during the late 19th century, the industry's power quickly became concentrated in just a few players, in line with a world where transnational corporations were becoming increasingly important (Negus, 1992). During most of its history, the recording industry has been a "tight oligopoly" (Hull et al., 2011, p. 111). However, there were times of lesser and greater concentration. For example, in the United States, an increase from 11 to more than 200 record labels was observed between 1949 and 1954, which was explained by the multiple independent record labels

---

[2] It is reasonable to assume that, for digital revenues (particularly streaming), the growth rates of countries such as the United States would be lower, as many of their users have been using digital services for a longer time.

that entered to compete in the market during that time. Nonetheless, these less powerful entities were thought of having a better perception of consumers' tastes and preferences and were generally more flexible with their cultural output –in terms of genre adoption– and organisational affairs (Negus, 1996). The bigger companies, on the other hand, followed strictly hierarchical procedures and relevant decisions were often made solely by humdrum workers, while also focusing on a handful of mainstream genres. By the end of the 1970s, concentration had gone up again, and mergers would continue past the turn of the century (Tschmuck, 2021; Baskerville & Baskerville, 2019).

Consequently, over the past few decades, the number of leading record labels has dwindled. Today, just three companies—UMG, SME, and WMG—command the majority of the global market share. Yet, mergers and acquisitions persist within the music industry, with updates about consolidations regularly appearing in specialised music industry publications. Such developments have gained recent significance in South American music markets, particularly in Brazil. Considering this background, we point out that the increased granularity of our dataset, compared to the historical accounts we presented, offers us a more nuanced approach to evaluating these cycles of fluctuating concentration levels, particularly within the charts.

2.3.2.Majors versus indies

We can distinguish two types of record companies. On the one hand, major labels manage their own distribution systems, while usually exhibiting a complex organisational structure with numerous departments[3] and relatively large budgets for production and promotion. The current recording industry continues to be dominated by the major labels, companies "owned by large transnational corporations which have interests encompassing enterprises and firms providing domestic products and services" (Negus, 1992, p. 2). Even in the digital era, core strategies followed by the majors still include mergers, acquisitions, and alliances (Rogers, 2017).

On the other hand, independent "indie" labels are smaller in scope and generally need the power of majors' distribution chains so their artists can reach comparable levels of exposure. Another

[3] Common divisions of labour within a major label subsidiary typically include departments such as Artist & Repertoire (A&R), Legal Affairs, Distribution, Marketing, and others.

group of independent record labels often use different intermediaries, such as digital aggregators, to upload the music to digital service providers (DSP) like Spotify and Apple Music. On Table 1, we provide a table with criteria to distinguish between major record labels and indie labels, mainly based on Tschmuck (2021) and Wikström (2013).

Table 1. Characteristics of record label types		
	Major labels	**Indie labels**
Parent company	Vivendi (UMG), Sony (SME), Access Industries (WMG)	All others. Also includes self-represented recording artists
Control over distribution	Own distribution channels, both physical and digital	Celebrate deals with major distributors and/or digital aggregators
Geographic scope	Transnational	Generally local
Market share	Higher (generally over 15% of the total market)	Lower. The combined forces of indie labels rarely surpass the most powerful major conglomerates
Preferred music genres	Mainstream genres (e.g., pop, rock, and urban styles in the US market)	Underground and niche musical styles (e.g., jazz and classical music)
Budget	Higher, with significant funds providing for substantial advances and widespread promotion	Generally modest, advances can be on the lower side or absent
Organisational structure	Complex and with multiple departments	Generally simpler. Multiple roles can be assumed by a single department/person.

For operationalisation purposes, we chose market share as the sole distinguishing feature between majors and indies, establishing a cut-off value of 15%. The specific value is based on the worldwide market structure of the recorded music industry during the 9-year period from 2013-2022 (Figure 2, based on Music & Copyright, 2023). The start of the period was chosen taking into consideration the partitioning and incorporation of EMI into the three remaining majors, which concluded in 2012.

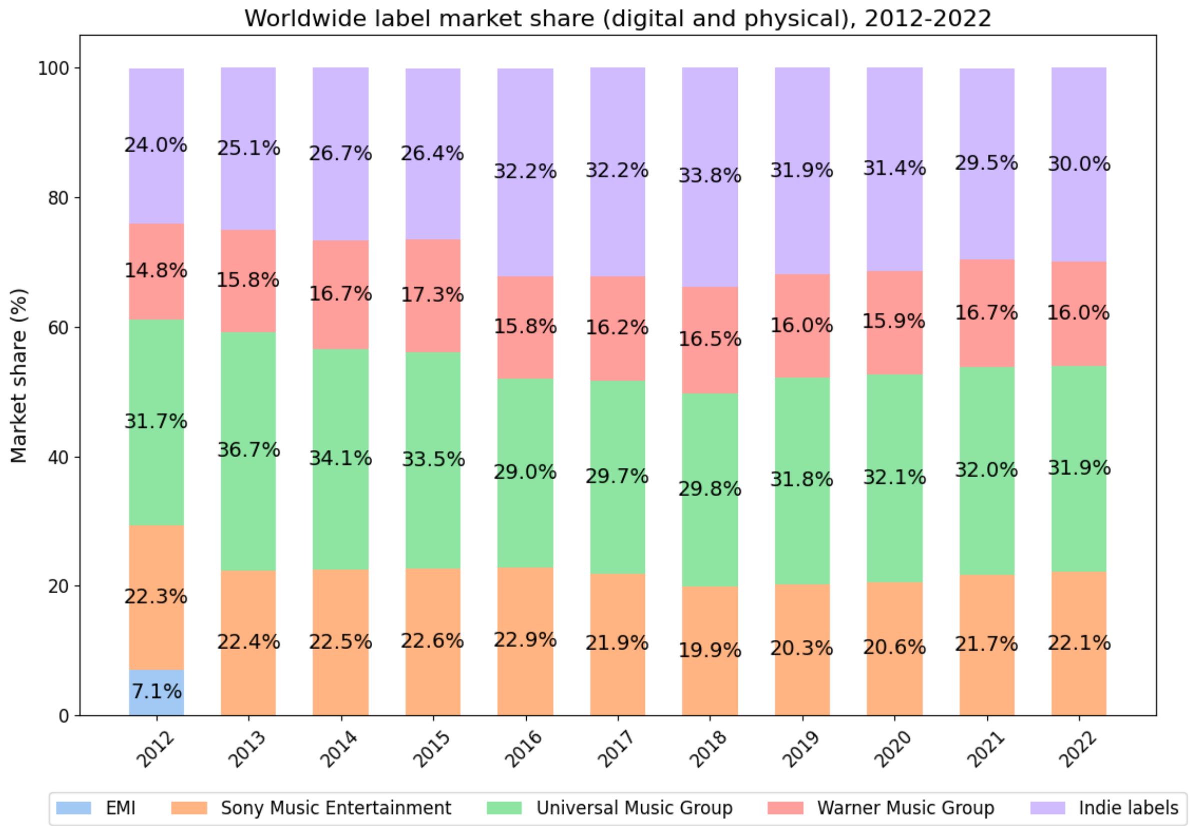


Figure 2. Label market share based on digital and physical revenue on a global scale. Note the disappearance of EMI as a separate conglomerate in 2012, as its catalogue was divided between the remaining companies.

2.3.3. Major labels in Latin America

Latin music and the Latin American market did not become an important part of the majors' interest until later. Brazil was a premature case, as SME, at that time through its predecessor RCA Victor, started operating in the country in 1929. In the case of UMG, we can trace back its origins to 1945, as the Sociedade Interamericana de Representações (Sinter), through reestablishments and subsequent acquisitions, would end up being part of Universal Music Brazil, which was conformed in 1997. WMG existed since the 1960s through its predecessors and subsidiary labels, but the current conformation only began in 2004 (Howard-Spink, 2012).

For the rest of the territories in Latin America, Sony apparently became aware of the potential of the Latin scene prematurely, as Sony Discos long-time former president Frank Welzer explained in a 1999 interview:

> We were years ahead of the game. We were the only company that realized that this big business was out there waiting to be developed. We had the confidence to turn the company into something while the other majors were still in the licensee era or beginning to think of their U.S. Latin companies as distributors for product from elsewhere in the world. We felt we could do both: have a burgeoning market for product outside of the United States, but, in addition to that, really get into this U.S. Latin business. (Lannert, 1999, p. 58).

Following the recent rise of Latin artists, majors are more inclined toward acquiring shares in successful independent labels, such as SME's interest in Bad Bunny's Puerto Rican label, Rimas Entertainment (Christman, 2023).

2.4. Superstardom, success, and long tails

It is a common notion that attaining [and maintaining] success in the music industry is a difficult endeavour (Tschmuck, 2012) and that most competing artists will largely remain obscure during their careers, compared to the mainstream paradigm. More generally speaking, the concept of a handful of players/products commanding entire markets is widely recognised. This is exemplified in principles such as the Pareto rule (80% of sales come from 20% of products), "winner-takes-all" markets (Frank & Cook, 1996), and the phenomenon of "superstardom" –the superstar being a disproportionately successful firm–. It is also documented that the steeply skewed income distributions of records typically resemble a power law (Krueger, 2019; Haampland, 2017). In some studies, efforts have been devoted to mathematically model revenue distributions, usually finding good fits with expressions that resemble inverse-square laws, such as Lotka's Law (Cox and Felton, 1995).

The existing research on these phenomena, however, mostly derives from two theories: (1) Sherwin Rosen's (1981) "enough" talent differences/economies of scale exploitation, and (2) Adler's (1985, 2006) luck/search costs theory. We provide a brief overview of both approaches.

2.4.1. The "original" superstardom hypothesis: Rosen's "enough" talent differences

According to American labour economist Sherwin Rosen (1981), superstar effects refer to the "concentration of output among a few individuals, marked skewness in the associated distributions of income and very large rewards at the top" (p. 845). Within the arts and culture field, Rosen mentioned classical soloists being scarce and earning more than substantial incomes. Additionally, he proposed a relationship between talent and skewness, further expanded by the notion of convexity. For Rosen, convexity referred to the magnification of minor differences in talent, which translated into disproportionate earnings differences. The author considered that imperfect substitution explained such differences because "lesser talent often is a poor substitute for greater talent" (p. 846). Talent, however, has been consistently difficult to measure, and most often the "gold standards" are the opinions of specialised critics. Despite that initial notion, Rosen recognised that preferences alone were "incapable of explaining [...] the market concentration of output" (p. 847) and proposed technology as the causal factor: those in control of means of production and distribution that allow for a more widespread reach would be in a significantly more advantageous position. In this regard, the history of the recorded music industry has been a synonym of technological development: since its beginnings, emphasis has been placed on efficiency and lowering production costs, an early example being the endeavours of talking machine producers to design more durable record matrices, enabling the production of a greater number of copies to be generated from a single master before deteriorating. The shift to digitisation brought about even more substantial improvements, as the marginal cost for each additional copy of a digital product, such as downloads, essentially became zero. Given its recent nature, this is compatible with the notion that superstar effects "are not new but have become increasingly important in the platform economy of the twenty-first century" (Abbing, 2022, p. 288).

The original definition for superstardom appears to us somehow questionable, as no clear cut-off points were established for claims such as "relatively small numbers of people earn enormous amounts of money and the activity in which they engage" (Rosen, 1981, p. 845). However, the issues derived from Rosen's superstardom have been an important topic of interest in cultural economics for quite some time with several empirical studies attempting to validate such constructs within the creative sector. However, the lack of clarity in the definition, especially in relation to the "talent" component of the hypothesis, limits its applicability in the context of this research.

2.4.2. Adler's luck and search costs theory

Adler's expanded perspective on superstardom, as articulated in his works from 1985 and 2006, goes beyond the simplistic narrative that winner-take-all markets are solely a product of inherent talent. He attributes the rise of superstars to the intricacies of consumption behaviours. Specifically, he posits that once a listener has committed to an artist, the associated costs—both in time and cognitive effort—to gravitate towards a potentially more talented alternative can be relatively high. In our case, this framework introduces the idea of "search costs" in the domain of music consumption.

Moreover, Adler infuses an element of serendipity into the superstardom equation. He posits that luck, combined with the momentum generated by snowball effects, plays a pivotal role. From an efficiency standpoint, having a limited number of superstars at a particular moment is more practical as it simplifies the consumers' choice. This dynamic might show an amplified relevance in the era of streaming platforms. Given how these platforms structure user interfaces—with pronounced emphasis on visibility for a select few artists (i.e., appearing in editorial playlists)—it is plausible that listeners might pivot towards these easily accessible choices. This not only minimises their search efforts (i.e., navigating through the vast number of available options within the catalogue) but also associates with a desire to align with broadly accepted and popular selections, avoiding the potential social friction of an unconventional choice.

2.4.3. Empirical evidence on superstar effects

Descriptive statistical reports on superstars-dominated music markets and their evolution over time include the work by Connolly and Krueger (2005), where the authors reported that, according to an analysis of the Pollstar database, income from ticket sales went from 26% for the top 1% in 1982, to 56 % for the top 1% of artists in 2003. Krueger (2019) further shows that, by 2016, the top 1% artists took 60% of concert revenue. The author also noted power law distributions within streaming platforms, based on data obtained from BuzzAngle: in 2016, the most-streamed artist accumulated 6.1 billion plays, but the hundredth artist was listened 0.5 billion times, less than 10% of the what the most successful obtained.

More evidence on varying concentration levels at the label level within the recorded music industry is available. Earlier reports following Peterson and Berger's (1975) work focus on the fluctuating nature of concentration (i.e., label diversity) in the phonographic industry, but solely based on the number of participants at any given time, using measures such as the 4-firm and 8-firm concentration ratios. The authors concluded that vertical integration, understood as being in control of "the total production flow from raw materials to wholesale sales" (p. 161), explained oligopolisation. In a similar fashion, Belinfante and Johnson (1982), utilising several ratios and the Herfindahl-Hirschmann Index (HHI), found increasing concentration in the US market for albums and singles at several points during the 1954 to 1981 period. It is also possible to see that, gradually, studies became more sophisticated by incorporating sales data or sales equivalents plus different concentration indicators, offering a greater window of opportunity compared to using rather rough measures such as rank positions.

Further, some evidence utilising and/or supporting Rosen's theory is available. In a study in the cinematographic industry, Hofmann and Opitz (2018) found that the income of actors and actresses in what they called the "talent segment" (i.e., having received nominations and/or awards) varied depending on the "talent levers" in a directly proportional way, although with noticeable skewness. Interestingly, evidence supporting Adler's theory was also found within the same study, but in a separate study group, as the income for the "publicity cohort" (i.e., actors and actresses not having received a distinction) depended more on their popularity ranking with an even more skewed distribution compared to the talent group.

2.4.4. Success factors in the music industry

Following Adler's theory on superstardom which focuses more on extrinsic variables than intrinsic talent, we observe that there is a body of academic literature that aims to comprehend the external factors related to success. A couple of years after publishing the study that brought vocal upper harmonics as the quality parameter to test for evidence of superstar effects, Hamlen (1994) described several variables positively or negatively correlated with success in the singles and albums market. Among the findings, the release date was important as "singers beginning later in the period had less time to accumulate hit records" (p. 401). Also, being female correlated with a higher chance of getting a hit single or album. Movie appearances were also positively correlated with success. On the contrary, being a black singer was significantly associated to a lesser chance to obtain a hit record in the albums market. Additionally, writing their own material wasn't significantly correlated with success.

Fox and Kochanowski (2004) pointed out that "those artists producing more singles during their career are likely to have more gold or platinum awards", positively associating success with productivity, and that "artists with longer careers are likely to have accumulated more awards" (p. 519), positively associating success with career longevity. Cho et al (2019) observed that artists participating in singing contests, specifically American Idol, were perceived as being less risky by the record labels, which increased their chances of being offered a recording contract. Also, past contest participants showed more success rates, in terms of digital song sales, especially within the Korean music industry.

In Latin America, Santos et al. (2019) analysed the determinants of success regarding Brazilian "music content producers" (including both the recorded and live music industry). In contrast to other studies, the authors differentiated between the "head" and "tail" portions of the distribution. They found that, overall, factors such as gender and race ("factors related to discrimination") weren't determinant in the hits market, suggesting that other variables related to talent and productive characteristics were more important. However, gender and race did influence success in the niche markets. Also, the flexibility of not being subject to formal contractual agreements (i.e., informality) was only advantageous for the actors of the hits market, while more stable work conditions were more important for the niche markets.

Two additional works shed light on supplementary factors that may correlate with success, specifically in the streaming domain. McKenzie et al. (2020) conducted a study to determine whether artist collaborations on streaming platforms influenced success in terms of stream counts. They found that songs featuring a guest artist generally outperformed those lacking the collaborative element, attributing the differentiated outcome mainly to promotion. Finally, Kaimann et al. (2020) found a positive correlation between chart survival and affiliation with a major label. They also reported that tracks that obtained popularity more quickly faded away faster, and that repeated simultaneous releases from an artist seem to cannibalise chart survival times.

2.4.5. Long-tail effects and long-tail players

The long-tail hypothesis, as popularized by Anderson (2004, 2008), posits that, with the rise of digital marketplaces and reduced entry barriers, niche products previously overshadowed would gain more attention. Conversely, the head of the distribution, traditionally dominated by the most popular items, would diminish in significance over time. Such a shift promised consumers a broader product variety (Waldfogel, 2018). As a result, one would anticipate a gradual "fattening" of the tail, benefitting lesser-known entities no longer restricted by physical shelf space constraints. Translating this to the music industry suggests that even niche artists, outside mainstream circles, could experience a surge in "purchases" or streams. However, evidence supporting this claim has been conflicting, and the supposedly positive effects of the long tail hypothesis on niche artists have been questioned by empirical studies. For example, Elberse (2008), using data from Nielsen VideoScan found that, "[f]rom 2000 to 2005 the number of titles in the top 10% of weekly sales [of home videos] dropped by more than 50%" (p. 92). In an article written for lay-public, Page and Bud (2008, cited in Day, 2011), reported that on iTunes Music Store, 80% of the sales came from 0.4% of the songs, and over one year of observations, 85% of all albums hosted in the platform sold zero units. It would seem that Anderson's claims didn't hold up during the first wave of digitisation in the music industry.

On the other hand, long-tail players are those commercially less-favoured artists who remain at the end of the distribution and experience a degree of separation from mainstream participants[4]. It is worth mentioning that our study does not effectively consider the most obscure long-tail players, as our database only includes the top 200 artists, but draws from the notion that widespread

[4] Empirical evidence supports the claim that even broader phenomena, such as music piracy, has impacted both groups differently. See Savelkoul (2020).

technological development and increasing access to “Do-It-Yourself” (DIY) digital aggregators could make the “head” of the distribution less concentrated, even in light of the conflicting evidence, as shown in this section. Therefore, it would be more accurate to say that from now on we will be focusing on the relative long tail (Im et al., 2019). Moreover, for the first part of this research, we will not be referring strictly to technology-driven effects, but mostly concentration changes in time following a descriptive approach. Thus, “relative long-tail players” is a more useful construct than the “effects” component of the long-tail theory.

3. Data – Ownership of streams and concentration

We begin this section by describing an ideal source for calculating concentration within the Latin American recorded music industry. We needed a database containing the various available consumption sources that generate digital revenues (i.e., non-interactive streaming, interactive streaming, and digital downloads), separated by country.

Unfortunately, the easily available sources of information on streams/plays/downloads that we came across were heterogeneous (i.e., coming from different streaming platforms in an ununified way). Obtaining an integrated dataset requires a third-party entity to compile and process the data to generate a coherent report. In this regard, companies such as Alpha Data –previously known as BuzzAngle (Dredge, 2020)–, Music & Copyright, and BMAT specialise in data provision and analytics for music industry professionals, specifically regarding chart information. However, according to their representatives, such cohesive datasets do not currently exist in Latin American countries, or there are contractual issues that prevent these companies from providing more flexible access to the information they possess.

Considering these initial limitations, we obtained the primary dataset for the present study, from Spotify Charts (http://www.spotifycharts.com). Spotify Charts is an official free compilation of the most successful songs and artists on the streaming platform. There are six types of datasets available: “Weekly Top Songs”, “Weekly Top Artist”, “Weekly Top Albums”, “Daily Top Songs”, “Daily Top Artists”, and “Daily Viral Songs”. The data comes as a list of the top 200 tracks/artists for all datasets, except for “Daily Viral Songs,” which only displays the top 100 tracks. The information is available on a global scale, but Spotify Charts also includes data specific to 72 countries, and even to some cities within those countries. It displays the summarised information on the web platform, which can be expanded by clicking on each track/artist that appears on the list. It also includes a downloadable .csv file for each week. The data is updated regularly, and the information is available from the last week of 2016 to the present day.

We chose streaming as our proxy for income/attention. According to the International Federation of the Phonographic Industry (IFPI, 2021), streaming showed a 41.3% Compound Annual Growth Rate from 2010 to 2020 (p. 62). In the case of Latin America, the revenue share of streaming was 84.1% as of 2020, compared to its global counterpart at 62.1% (p. 88). Also, although there are some other free databases available from different companies (e.g., Apple Music's "Top 100"), we chose Spotify Charts because the company holds a significant market share in Latin America, with 46% of streaming users captured, according to Global Web Index (GWI, n.d.).

When deciding between "top artists" and "top songs", both available on Spotify Charts, we opted for "top songs", first, because the relationships between artists and labels can vary considerably over time (e.g., an artist can start their appearance in the charts as an independent, then be signed to indie label "A", then to major label subsidiary "B", and so on). The second reason lies in our subsequent level of analysis, which includes several indices that measure musical/sonic aspects of the phonograms. For this second layer, the particularities of each track are deemed essential, and they would be lost if only the artists were considered. Finally, we believe there can be high intra-artist variability regarding success, with not every song released by a single artist/band resulting in chart inclusion. In addition, we chose "Weekly Top Songs" instead of "Albums", as singles have recently become quite a popular form of music consumption.

The start of our sample period will correspond to the first data collection entry available on Spotify Charts, which is the last week of December 2016, ending the 29th day of that month. The end of our sample period will be the week concluding on January 19th, 2023, with a total of 317 weeks. To the best of our knowledge, there are no significant variations in Spotify's data collection method. Still, it is also worth mentioning that the company does not disclose how they discriminate between "chartable" and "non-chartable" tracks (Spotify, n.d.-b).

Overall, compared to years before the proliferation of online music services, excellent data is available. Digital technologies allow for a better characterization of the demand of music (Collins & O'Grady, 2016). For our specific purposes, they allow accessing more sophisticated data on charts composition and behaviour, for example, by providing streams as an indicative of attention/popularity, instead of solely chart positions, doing so on a daily or weekly basis, which in turn allows for a better identification of trends over time.

3.1. Country selection

We have included the top ten Latin American nations regarding streaming revenue. We used the data offered in the IFPI report from 2021. We completed missing information for Costa Rica and Guatemala with data from the IFPI Latina site (https://ifpilatina.org/), as IFPI (2021) grouped some countries as part of "Central America" and "The Caribbean" without offering further detail. After this procedure, we ended up with the following countries in our list (descending streaming revenue order): Brazil, Mexico, Chile, Argentina, Colombia, Peru, Costa Rica, Ecuador, Guatemala, and Uruguay (Table 2).

Table 2. Streaming revenues for selected Latin American countries, according to IFPI (2021) and associated regional reports

Nº	Country	Revenue (million US dollars)
1	Brazil	256.6
2	Mexico	193.9
3	Chile	48.4
4	Argentina	40.9
5	Colombia	33.6
6	Peru	19.7
7	Costa Rica	14.7
8	Ecuador	10.5
9	Guatemala	8.2
10	Uruguay	6.3

3.2. Custom label database construction

In a recent study, Aguiar and Waldfogel (2021) noticed the difficulty of classifying labels in the Spotify database as majors or independents. To deal with this issue, they compared entity names with a list of twenty-two well-known major labels' subsidiaries to separate them from the independents (Table 3). However, they acknowledged that their method, despite ensuring that major labels ended up being correctly identified, could potentially end up misclassifying "some of the non-obvious majors […] as independent labels” (p. 663). Moreover, they did not distinguish whether label subsidiaries came from UMG, SME or WMG.

Table 3. Aguiar and Waldfogel’s (2021) major label selection

Asylum	Parlophone	Universal
Atlantic	Republic	Astralwerks
Capitol	Big Machine	WM
Epic	Sony	Trinidad & Tobago
Interscope	Polydor	RCA
Warner	Big Beat	Columbia
Motown	Def Jam	
Virgin	MCA	

In contrast, our work followed a more thoroughly approach, although with similarities at various point of the process. First, we used the “source” column of the Spotify database, which contained information on the record labels for each song included in the charts. We discovered that 1,837 entities in the “source” column had a non-coincident spelling, which we assumed as generally representing different entities, although misspells and slight variations on name writing were occasionally found. As expected, there were no categorisations as “indie” or “major”. Furthermore, in most cases, there was no information on the role of the record company, whether they were acting solely as distributors, publishers, phonographic producers, or any combination of the three. Consequently, our next step was to find a comprehensive, up-to-date database with information on record labels, specifically on the distinction between majors and indies, the production companies and the extent of distribution deals between indies and majors. As we couldn’t find one, our second-best option was to construct the major/indie categorisation database manually. For this purpose, we used MusicBrainz as our main source of information. MusicBrainz is a free relational database containing information on artists, recordings, songs, labels, etc. It also offers a search engine that can discriminate whether the input terms are referring to an artist, song, release, label, etcetera.

We followed a specific process to categorise each of the 1,837 entities. Firstly, within MusicBrainz, if the name contained a particle that was evidently related to one of the “big three” – UMG, SME, and WMG–, we considered it as being a [subsidiary of a] major label. For example: “Universal Music Argentina S.A.”, “Universal Music Mexico”, and “Universal Music New Zealand

Limited" all have a particle that refers explicitly to UMG. We also considered acronyms and abbreviations when they did not lead to significant ambiguity (i.e., "UMG" for "Universal Music Group"). Second, we focused on the "Overview" and "Relationships" tabs in MusicBrainz. The former, when available, contained information on the labels' history, country of origin, preferred genres, etcetera, usually obtained from Wikipedia. Due to the inconsistency of the "Overview" tab, we often decided based on the "Relationships" tab. This section contained different types of information ("Founders", "Signed [artists]", "Parent label", "Subsidiaries", "Distributors" etcetera). The amount of information varied considerably between labels, thus, we focused only on the "parent label" information, which was more consistent. We considered a label as "major" if "Parent label" led to one of the "big three" and to one of the "indie" categories if no such connection could be made from the entire "Relationships" section.

The case of Som Livre is worth mentioning. This record company, which used to be independent, represented up to 22.8% of the Brazilian market. However, SME acquired Som Livre from the Globo media conglomerate in March 2021 (Music & Copyright, 2021b), and completed the transaction on March 4th, 2022 (SME, 2022a). Therefore, starting on week 272, we changed Som Livre's condition in our dataset from indie to a major label subsidiary.

The careful procedure that we followed was relevant for more accurate label classification in our dataset. For comparison purposes, we calculated the number of misclassified labels that could come up from applying Aguiar and Waldfogel's (2021) method to our dataset. By doing this, 81 of the total 1,837 labels ended up being erroneously classified as indies. Moreover, when calculating the number of entries affected by such inaccuracy, we ended up with a total of 96,666 entries where the label was misclassified, which accounted for approximately 15% of the entire dataset. Therefore, by following our approach, a better-quality dataset was ultimately constructed.

Finally, it is important to highlight that we are aware of the organisational fragmentation that characterises the recorded music industry. There are three influential conglomerates, but there are also "hundreds (if not thousands) of individual labels that appear to function autonomously" (Hull, 2004,

p. 18). Also, despite “labels”, “distributors”, and “publishers” being fundamentally different roles, we could hardly infer that sort of information from the “source” column on the Spotify database, the only information being an ambiguous “powered by” particle added to selected entities appearing in a handful of cases. Because of that, regarding their functioning, we limited ourselves to consider all entities as “labels” (in charge of the production, marketing and/or distribution of the records, indistinctly).

4. Ownership of streams – Variables

This section describes the variables used in the first part of our research. For organisation purposes, we have grouped the variables into four categories: chart songs data, number of streams, share of streams, and concentration and inequality. We follow with a detailed account, but a summary can be found in the Table of variables at the end of the section.

4.1. Number of streams

Weekly song plays are displayed on the "streams" column in Spotify Charts-generated .csv files. A single stream is counted with 30 seconds or more of listening time (Spotify, n.d.-c). Chart positions are established solely on this basis, not considering other factors (e.g., expert opinion). As we have suggested before, stream counts reflect consumption more accurately, but they are not immune to distortion/external influence. For example, playlist inclusion and "passive listening" can temporarily increase the stream count based on the algorithmic recommendation systems, while "active listening" more accurately reflects the intention of consumers to listen to specific songs and artists (Aguiar & Waldfogel, 2021).

Streams are a good indicator of royalty payment on Spotify (and, therefore, a proxy for digital music revenue, considering the decline of downloads and the predominance of the Swedish platform). It is worth clarifying that according to their own statements, Spotify does not pay based on a per-stream rate. Nonetheless, streams are mostly homogeneous in their monetary counterpart as they currently follow a *pro-rata* model by which all listeners' stream numbers are "pooled" and then used to calculate how much each artist receives (Meyn et al., 2023). Specifically, Spotify distributes their net revenue by "tallying the total number of streams in a given month and determining what proportion of those streams were people listening to music owned or controlled by a particular rightsholder" (Spotify, n.d.-d). Of course, there are factors than can cause heterogeneity, such as cross-country price discrimination regarding premium subscriptions, which could ultimately cause variations regarding the "pool of money" that Spotify ends up distributing (Waldfogel, 2020).

Our initial variable of interest is the raw stream count, as provided by Spotify Charts. We've aggregated these stream counts on a weekly basis. Over time, these cumulative stream numbers illustrate both the growth of the platform in terms of user count and country-specific revenue growth trends for major labels, their conglomerates, and independent labels. On the other hand, the "share of streams" variables align more closely with how label market shares are typically reported by music data compiling/processing organizations like Music & Copyright, Luminate, and BMAT. This offers a more direct method for assessing the relative significance of each label type and its evolution over time.

4.2. Concentration and inequality

4.2.1. Concentration ratios

Concentration ratios (CR) are popular metrics used to gauge market share composition/structure based on the dominant influence of the top "N" firms (e.g., CR3, CR4, CR5, CR10, etc.) (Rhoades, 1995). Such metrics have been employed in the music industry, as seen in research by Peterson and Berger (1996), Alexander (2002) and Rothenbuhler and Dimmick (1982). Though CRs are straightforward, widely recognized, and easy to interpret, they do possess a notable limitation: they neglect the market share of other, potentially numerous, firms— which can have important implications especially when a market features a pronounced long tail. Further, focusing on a large number of firms (like CR50) might downplay the influence of the powerful entities when only a few dominate the market (OECD, 2021). Despite these drawbacks, we have utilized CR3 as a simple and forthright measure of concentration, calculating it weekly across our entire dataset. In every instance, CR3 equalled the cumulative stream share of the top three major record labels within the Spotify Charts. Hence, throughout this study, we will refer to it simply as the majors' share of streams ('share_st_majors').

4.2.2. The Gini coefficient

Lorenz curves (Figure 3) and Gini coefficients have been utilised mostly for measuring income inequality but have found further use in determining firm market share inequality (Rhoades, 1995). The Lorenz curve "shows the percentage of total income earned by cumulative percentage of the population", and the Gini coefficient is "equivalent to the size of the area between the Lorenz curve and the 45º line of equality, divided by the total area under [it]" (De Maio, 2007, p. 850). Its values vary from 0 (perfect equality) to 1 (perfect inequality), or from 0 to 100. We computed the Gini coefficient using the formula: Gini = Area A / (Area A + Area B). Here, Area "A" represents the space between the Lorenz curve and the line of perfect equality, while Area "B" is the area under the Lorenz curve. To find Area "A", we subtracted Area "B" from the total area beneath the equality line (which is 50). For these calculations, we took into account the total weekly streams for each label. Furthermore, we aggregated the streams of the major subsidiaries under their respective parent labels.

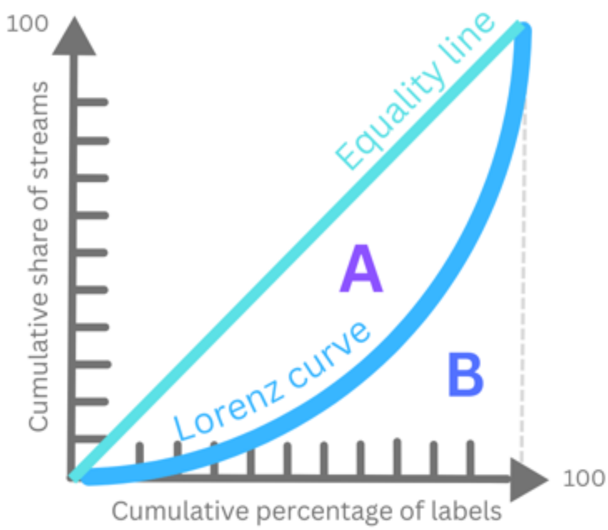


Figure 3. The Lorenz curve.

While the Gini coefficient provides an intuitive understanding of inequality, it is particularly sensitive to variations in the median of the population distribution. This could unintentionally divert our attention away from songs at both extremes of the charts. Notably, the top songs typically garner most of the attention, a trait that is clearly reflected in our Spotify dataset. Given that it comprises solely chart information, it naturally only includes "the best of the best".

4.2.3. The Herfindahl-Hirschmann Index

In contrast, the HHI is a slightly more sophisticated measure of concentration. It is calculated by summing the squares of the relative shares of every firm within an industry. The HHI "emphasises the importance of larger firms by assigning them a greater weight than smaller ones" (OECD, 2021, p.

12), which in this case would generally correspond to giving more importance to major record labels, due to their well-known presence on the charts. It is also well worth highlighting that not having the total composition of the market, but only their top values, will limit our conclusions to "chart territory" (i.e., the most popular songs) and not the total Spotify market. We calculated the HHI as the sum of squares of the percentage of streams corresponding to each label for any given week. Also, we grouped major subsidiaries according to their parent labels.

Table 4. Selected variables for ownership of streams and concentration						
Chart songs data						
Name (abbreviation)	Variable type	Definition	Unit	Data source	Notes	Descriptives
Country (country)	Nominal categorical	The country where the stream count was generated. All countries are from Latin America (South America, Central America, and North America). Not to be confused with the country of origin of the sound recording.	None	Spotify Charts	0=Argentina 1=Brazil 2=Chile 3=Colombia 4=Costa Rica 5=Ecuador 6=Guatemala 7=Mexico 8=Peru 9=Uruguay	
Week number (week_number)	Ordinal continuous	Week of the year where the stream count was generated (not accumulative).	None	Spotify Charts	From 29-12-2016 to 19-01-2023. See Appendix for details on the exact dates corresponding to each week	Min=1 Max=317
Number of streams						
Name (abbreviation)	Variable type	Definition	Unit	Data source	Notes	Descriptives
RAW data						
Individual stream count (streams)	Discrete numerical	Refers to the number of times a track has been played or streamed by the Spotify users during a period of a week.	None	Spotify Charts		Mean=348,294.87 Median=115,269 SD=607,317.026 Min=5,127 Max=19,816,644
Aggregate – All						
Sum of streams – All labels (sum_st)	Continuous numerical	The weekly sum of streams of all 200 chart positions.	None	Spotify Charts		Mean=69,649,195.64 Median=30,294,486.5 SD=87,067,125.57 Min=2,501,795 Max=451,313,134

Aggregate - Majors						
Sum of streams – Major record labels (sum_st_majors)	Continuous numerical	The weekly sum of streams of the three major record label conglomerates (UMG, SME and WMG) in Spotify Charts.	None	Spotify Charts MusicBrainz		Mean=37,223,795.59 Median=18,219,763.50 SD=44,794,021.26 Min=2,022,968 Max=244,249,188
Sum of streams – UMG (sum_st_universal)	Continuous numerical	The weekly sum of streams of songs produced and/or distributed by UMG which were included in the charts.	None	Spotify Charts MusicBrainz		Mean=13,503,475.43 Median=6,234,202.50 SD=17,057,266.89 Min=528,448 Max=85,703,392
Sum of streams - SME (sum_st_sony)	Continuous numerical	The weekly sum of streams of songs produced and/or distributed by SME which were included in the charts.	None	Spotify Charts MusicBrainz		Mean=15,523,888.78 Median=7,577,746.50 SD=21,134,157.02 Min=732,525 Max=170,368,951
Sum of streams - WMG (sum_st_warner)	Continuous numerical	The weekly sum of streams of songs produced and/or distributed by WMG which were included in the charts.	None	Spotify Charts MusicBrainz		Mean=8,196,431.38 Median=3,750,566.00 SD=9,825,838.49 Min=319,729 Max=54,796,027
Aggregate - Indies						
Sum of streams – Indie labels (sum_st_indies)	Continuous numerical	The weekly sum of streams of songs produced and/or distributed by them which were included in the charts.	None	Spotify Charts MusicBrainz		Mean=32,425,400.05 Median=12,694,342.50 SD=45,292,003.21 Min=477,875 Max=304,285,849
Sum of streams – Artists as labels (sum_st_artist_label)	Continuous numerical	The weekly sum of streams of songs where the artist assumed the role of a label (financing, production and/or distribution of the record), which were included in the charts	None	Spotify Charts MusicBrainz		Mean=3,703,472.24 Median=618,611.50 SD=7,423,134.40 Min=0 Max=50,360,988
Share of streams						

Name (abbreviation)	Variable type	Definition	Unit	Data source	Notes	
Majors						
Share of streams – Major record labels (share_st_majors)	Continuous numerical	Refers to the sum of the relative share of streams of the three major record label conglomerates (UMG, SME and WMG) in Spotify Charts.	None	Spotify Charts MusicBrainz	Calculated by taking the total number of streams that all three major record label conglomerates accumulated during a specific week, then dividing it by the total amount of streams during that week, and multiplying it by 100.	Mean=57.63% Median=57.13% SD=11.34% Minimum=26.39% Maximum=83.67%
Share of streams – UMG (share_st_universal)	Continuous numerical	Refers to the relative share of streams of UMG and its subordinate labels, considering the amount of streams for songs produced and/or distributed by UMG which were included in the charts.	None	Spotify Charts MusicBrainz	Calculated by taking the total number of streams that UMG and its subordinate labels accumulated during a specific week, then dividing it by the total amount of streams during that week and multiplying it by 100.	Mean=20.68% Median=21.27% SD=6.27% Min=5.43% Max=40.96%
Share of streams - SME (share_st_sony)	Continuous numerical	Refers to the relative share of streams of SME and its subordinate labels, considering the amount of streams for songs produced and/or distributed by SME which were included in the charts.	None	Spotify Charts MusicBrainz	Calculated by taking the total number of streams that SME and its subordinate labels accumulated during a specific week, then dividing it by the total amount of streams during that week, and multiplying it by 100.	Mean=23.65% Median=23.71% SD=5.62% Min=7.87% Max=45.95%
Share of streams - WMG (share_st_warner)	Continuous numerical	Refers to the relative share of streams of WMG and its subordinate labels, considering the amount of streams for songs produced and/or distributed by WMG which were included in the charts.	None	Spotify Charts MusicBrainz	Calculated by taking the total number of streams that WMG and its subordinate labels accumulated during a specific week, then dividing it by the total amount of streams during that week, and multiplying it by 100.	Mean=13.29% Median=13.10% SD=4.72% Min=2.75% Max=29.98%
Indies						

Share of streams – Indie labels (share_st_indies)	Continuous numerical	Refers to the relative share of streams of independent record labels, considering the amount of streams for songs produced and/or distributed by them which were included in the charts.	None	Spotify Charts MusicBrainz	Calculated by taking the total number of streams that indie labels accumulated during a specific week, then dividing it by the total amount of streams during that week, and multiplying it by 100.	Mean=42.37% Median=42.87% SD=11.34% Min=16.33% Max=73.61%
Share of streams – Artists as labels (share_st_artist_label)	Continuous numerical	Refers to the percentage of streams where the artist assumed the role of a label (financing, production and/or distribution of the record), considering the total amount of streams for songs produced and/or distributed which were included in the charts.	None	Spotify Charts MusicBrainz	Calculated by taking the total number of streams accumulated during a specific week where artists assumed the role of a record label, then dividing it by the total amount of streams during that week, and multiplying it by 100.	Mean=3.39% Median=2.55% SD=3.23% Min=0.00% Max=24.39%
Concentration and inequality						
Name (abbreviation)	Variable type	Definition	Unit	Data source	Notes	
Three-Firm Concentration Ratio (cr3)	Continuous numerical	The sum of the relative share of streams of the top three firms (a, b and c) on a weekly basis: $$CR3 = \frac{S_a + S_b + S_c}{T}$$	Index	Spotify Charts MusicBrainz	For this study, the top three firms corresponded to the major conglomerates in all cases, therefore this measure is numerically identical to 'share_st_majors'	Mean=57.63% Median=57.13% SD=11.34% Minimum=26.39% Maximum=83.67%
Gini coefficient (gini_index)	Continuous numerical	A measure for quantifying inequality, in this case within Spotify Charts. $$G = \frac{A}{A + B}$$ Where A is the area between the Lorenz curve and the line of perfect equality and B is the area under the Lorenz curve.	Index	Spotify Charts MusicBrainz	The values go from 0 (perfect equality) to 100 (perfect inequality).	Mean=74.68 Median=74.97 SD=3.48 Range=23.91 Min=61.53 Max=85.44
Herfindahl-Hirschmann Index (hhi)	Continuous numerical	The sum of squares of the relative market shares for each label, weekly, considering the total number of streams that each	Index	Spotify Charts MusicBrainz	The relative share of streams for each label was calculated as the sum of streams for all the songs that any particular	Mean=1,542.82 Median=1,485.53 SD=397.85 Range=2,984.04

		label accumulated for the specific week analysed. For any specific week, considering "n" labels (major conglomerates included), the HHI is expressed as: $HHI = S_1^2 + S_2^2 + S_3^2 \ldots + S_n^2$ Where S is the proportion of each label or conglomerate.			indie label or major conglomerate accumulated, divided by the total streams during that week for known labels, and multiplied by 100.	Min=713.63 Max=3,697.68

5. Ownership of streams – Descriptive statistics

In this section, we present descriptive statistics for the key variables used in our analysis regarding ownership of streams. The total number of observations for the ten countries was amounted to 633,911[5]. Of these, Argentina, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Peru, and Uruguay each had 63,393 observations. Due to certain missing entries, Brazil recorded 63,385 observations and Mexico accounted for 63,382 observations over the entire 317-week period. Be that as it may, the total number of missing entries accounted for only about 0.014% of the total dataset, didn't follow any systematisation that we are aware of, and therefore are deemed negligible.

5.1. Charts songs data

Even though our dataset includes variables like "weeks on chart" and "peak rank", we believe that the stream count is the most accurate representation of income/attention. Consequently, we've centred our attention on stream-related variables.

5.1.1. Streams

A combination of positive skewness and leptokurtosis (a kurtosis value over 3) was found in the distribution of streams, individually considering the songs that entered the charts (i.e., raw data). Positive skewness may correspond to a few extremely popular songs driving the stream count, while leptokurtosis indicates a flatter tail and a sharper peak, corresponding to a greater chance of finding extreme numbers and more stream counts closer to the mean. To illustrate, in a study by Elberse and Oberholzer-Gee (2007) on long-tail phenomena in video sales, the year-to-year increase in skewness and kurtosis on DVD sales was consistent with a distribution that became "more dispersed, more asymmetrical, and develop[ed] a sharper peak and a longer tail over time" (p. 58). Together, a leptokurtic and right-skewed distribution could preliminarily be suggestive of high concentration levels, when looking at the data transversally. However, the literature suggests that distributions around a mean appear to have "little economic meaning […] and probably little relevance as a tool for

[5] There were 89 entries missing for the total Spotify Charts dataset (see Appendix).

describing the size distribution of firms in markets […] because those measures do not necessarily capture the differences in market shares" (Rhoades, 1995, p. 662). Furthermore, it is inadequate to equal concentration to a simplification of dispersion (Adelman, 1959). Therefore, because this part of our study focuses on firms (labels) and their relative share of streams (analogous to size distribution), we will give more weight to more specific descriptors of concentration and inequality that aggregate the data, rather than on the raw data itself, although some other graphical accounts may be of guidance, such as what we present in Figure 4:

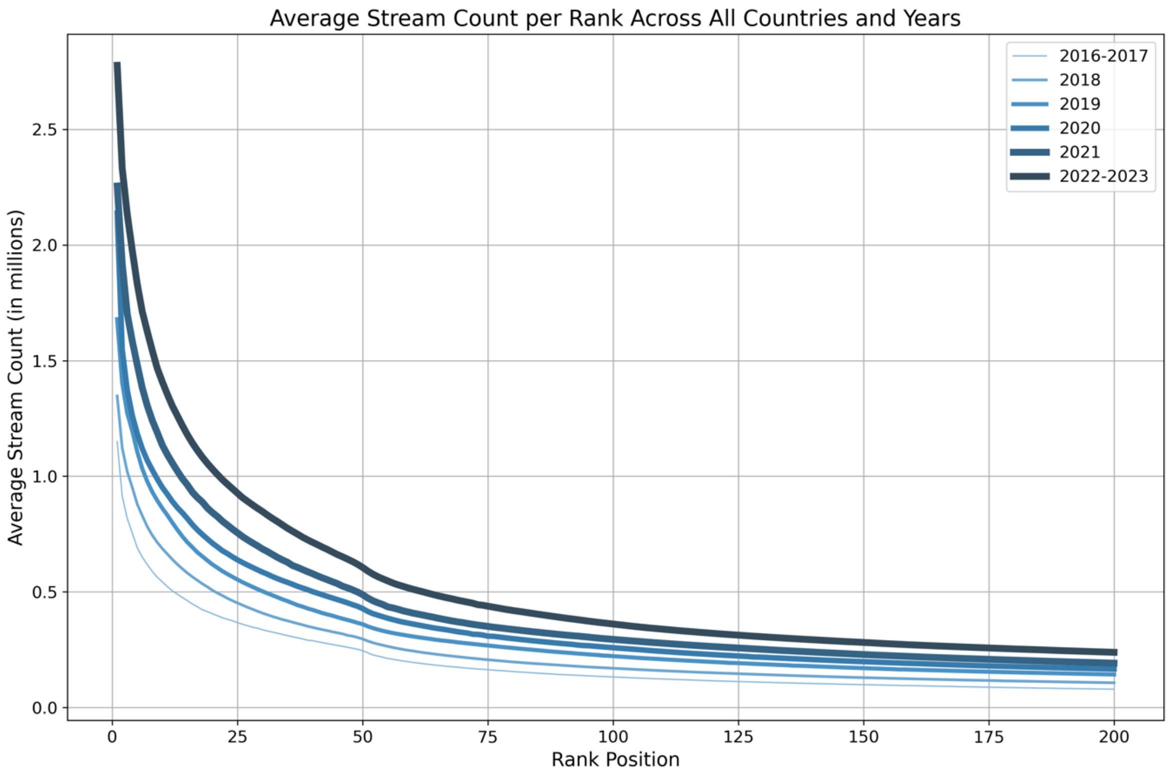


Figure 4. Yearly trends in stream growth across the entire distribution. Both the head and tail of the distribution have seen increased attention over the years (indicating market expansion in terms of consumption, as reflected by stream counts). However, the head portion seems to keep its dominance undisputed.

This shows the year-to-year growth in total stream numbers, but also suggests that, on average across all countries and over time, a high number of streams have accumulated at the top of the rank distribution, regardless of the affiliation with a specific record label type, compared to the tail of the distribution. A more in-depth longitudinal examination using concentration indicators will help ascertain whether this preliminary assessment translates into varying concentration levels over time.

5.1.2.Sum of streams

The aggregate stream indicators were designed to assess both the overall and country-specific streaming market growth trends throughout the duration of our study. It is also a proxy to how the number of Spotify users has changed over time, albeit not discriminating between free and premium accounts. In this sense, the first prevalent trend we observed was a consistent rise in total weekly streams: in the 2016-2023 period, the overall absolute difference in average sum of streams was 99,125,165.6, corresponding to a 310.96% total growth occurring between week 1 and week 317. The growth patterns, however, varied substantially between countries, as shown in Table 5. In this regard, Guatemala showcased the highest percentual variation –nearly six times that of Costa Rica–, which suggests that the penetration of online music services in Latin America varies significantly across countries. We suspect that long-tail effects, as per the original definition, could be more noticeable in such countries.

Table 5. Streaming count growth on Spotify by country, 2016-2023

Country	Absolute difference in average sum of streams	Relative growth
Argentina	112,157,369	+240.37%
Brazil	290,906,568	+275.62%
Chile	81,144,890	+261.33%
Colombia	52,611,343	+415.62%
Costa Rica	7,670,799	+141.76%
Ecuador	20,254,111	+410.46%
Guatemala	20,563,240	+821.94%
Mexico	356,457,046	+375.79%
Peru	41,647,184	+360.3%
Uruguay	7,839,106	+218.03%

On the other hand, Brazil was the leader in terms of average weekly streams, with Mexico closely trailing. By comparison, Argentina, ranking third in terms of average weekly stream aggregates, amassed only about 40% of Mexico's stream count and approximately a third of Brazil's total. This brings about the well-known size supremacy of Mexico and Brazil in the Latin American recording industry, which accounted for the preferred attention that the IFPI gave to these two countries in their reports since the early 2000s. It is also consistent with more general regional states: in 2021, 66.5% of the total recording revenues in Latin America corresponded to Brazil and Mexico (IFPI, 2022).

Two related facts should be considered when commenting on the evolution of aggregate plays. First, the number of monthly streaming hours on Spotify has progressively increased: in the 2015-2021 period, it went from 1.7 billion up to 9.8 billion (Dredge, 2021b). Second, Spotify's consumer numbers have significantly increased over time, particularly regarding premium users (Spotify, 2023). Specifically in Latin America, the number of users nearly doubled in the 2018-2021 period, from 42 to 83.8 million monthly active users and 17.4 to 34.4 million premium subscribers (Dredge, 2021a). Our results are therefore also compatible with these platform tendencies.

We have also considered some anecdotal accounts. For instance, Brazil's sum of streams peaked at week 54[6], two weeks after the release of "Vai malandra", a single by Brazilian singer Anitta, which debuted at number 1 on the charts, entered Spotify's Top 20 (UOL, 2017), and was promoted by the platform itself (Adnews, 2017; Torres, 2017). This also was on pair with an increase in major labels' sum of streams, most evidently in WMG, as WM Brazil was the label behind the release. Anitta's song debuted number 1 on this week. A second peak was found on week 106, with no new release driving the stream count, but a handful of Brazilian *funk*[7] and *sertanejo* songs. Similarly, more peaks occurred on weeks 158, 210, 246, 262, 271 and 314. It is worth noting that most of these peaks occurred in the proximity of New Year's Eve.

Other peaks occurred synchronically between countries. On week 167, less than one week after the release of "YHLQMDLG", the second solo studio album of Puerto Rican artist Bad Bunny, we observed notably higher stream counts in Mexico, Chile, Colombia, Ecuador, and Guatemala. When looking at the positions, we noticed that many of the songs entering the charts during that week came from "YHLQMDLG", which was also "the first all-Spanish-language album to reach No. 2 on

---

[6] A complete list of week numbers and their corresponding dates can be found in Appendix L.
[7] A Brazilian electronic music style from Rio de Janeiro.

the Billboard 200 chart" (Flores, 2020, para. 1). On week 206, a few days after the release of "El Último Tour Del Mundo", Bad Bunny's third solo studio album, a new peak occurred, most notably in Mexico, Chile, Peru and Colombia, with several album's songs being featured during that week in the top chart positions. Finally, a third peak occurred in week 281, closely after the release of "Un Verano Sin Ti", Bad Bunny's fourth solo studio album. It is worth mentioning that none of these tendencies were present in Brazil and were least evident in Uruguay. We expect an increase in our concentration measures coinciding with the stream "leaps" we described, but a subsequent in-depth inferential analysis would be desirable, not being covered in this work.

5.1.3. Share of streams according to label type

In our study, major labels were the most relevant entities in the charts regarding their respective share of streams, consistent with the accumulation of power that they have showcased during their history. Therefore, at the label level, this is anticipatory for considerable degrees of concentration within Spotify Charts, even before adhering to any hypotheses on superstardom. Nonetheless, we observed a general tendency for the majors' share of streams to decline over time during our period of interest (see Appendix). This aligns with a global diminishing tendency observed in digital music services over recent years: within Spotify, the majors' market share went from 87% in 2017, to 77% in 2021 (MBW, 2022). In our study, a notable exception to this trend was Brazil, with more frequent ups-and-downs and a leap in the average major labels' share of streams on week 272, following the acquisition of Som Livre by SME. While the academic literature and private statistical reports highlight the dominance of major labels in market shares for digital music and overall music sales, the majors' streaming share in our selected countries fell below global figures, even when limited to the most popular songs (i.e., chart territory): the mean value for majors' share of streams was 57.63%. Also, it is worth highlighting the overall dominance of SME regarding digital revenues: SME had, on average, 23.65% of total streams, while UMG and WMG had 20.68% and 13.29%, respectively. This is not described in Western countries or at a global scale (Music & Copyright, 2021a). Although we lack more granular data to elucidate the apparent dominance of SME over UMG and WMG in Latin America, historical accounts shed light on the regional dynamics between these companies. Firstly, SME's interest in Latin music roots back to the early 1980s, with the establishment of CBS Discos, the Latin division of international company CBS, in the United States (Moreno, 1979). CBS and their Latin catalogue became officially under the control of SME with the acquisition of the company in 1988. Secondly, UMG only got notoriety in the region after acquiring Polygram. By that time, according to Latin region insiders, "the combined market shares of Polygram and Universal –pegged in the 20% range– could make the combination nearly as large (italics added) as perennial market leader Sony, whose market share percentage is believed to be in the lower 20s" (Lannert, 1998, p. 78).

Corroborating our findings, Billboard's 2023 Year-End charts reported Sony Music Latin as the top Latin label, with Universal Music Latin Entertainment in third place, and Warner Latina ranking fifth.

5.2. Concentration and inequality

Throughout the duration of our study, the GI values consistently registered on the higher end, frequently surpassing 50. The mean GI value across all the countries analysed stood at 74.68, a figure that inherently signifies a market marked by pronounced inequality. This disproportion becomes even more notable when juxtaposed with findings from a comparable creative domain. Specifically, a study by Fernández-Blanco et al. (2014), which examined movie industry sales using the same indicator, reported GI values consistently under 50.

Over our study's course, some countries, notably Argentina, Chile, and Mexico, exhibited a noticeable downward trend in GI values. Argentina presented significant GI fluctuations, with a range of 19.06 across the span of 317 weeks and a minimum GI value of 61.53, the lowest in our dataset. In contrast, Brazil showcased a more stable landscape, registering a GI range of merely 8.13 throughout the study, coupled with marginally higher mean and median values in comparison to its counterparts. Variability in the GI, as seen in countries like Argentina, may be interpreted as chart turbulence, suggesting frequent chart entries of new tracks with divergent stream counts week-on-week.

In relation to the HHI, the values primarily indicated moderately concentrated markets (with values <2500), although the shape of the line chart was pretty similar to the GI's. It has been postulated, both hypothetically and through empirical evidence, that markets can manifest blatant inequalities in terms of individual firm market share while concurrently displaying only moderate concentration (Rhoades, 1995). Thus, considering the progression of the rank-level distribution of streams over time, the GI emerges as a more sensitive descriptor of the competitive landscape within the charts.

In Rosen's foundational hypothesis, control over technology endowed the more resource-rich firms with disproportionate market influence. When applied to the recorded music industry and viewed in the context of the industry's increasing reliance on data, it is the major labels that are better prepared to leverage data most extensively. With their vast resources, these entities can, for instance, acquire prominent playlist-generating companies and deeply analyse the consequent consumption data (Hagen, 2022). Nonetheless, our study revealed an interesting trajectory: while major labels unarguably reigned supreme in the charts, their grip started showing signs of loosening over the examined 317-week period. However, when assessing the average streams per rank longitudinally, independent of label affiliation, we observed that the percentages remained relatively constant over time, and that the head portion of the distribution continued amassing a disproportionate amount of streams. This trend resonates with prevailing academic narratives, such as the insights presented by Elberse (2008), questioning the accuracy of the long tail hypothesis during the digital era.

6. Musicological characteristics and diversity – Literature review

In this section, we provide a brief theoretical overview of the musicological characteristics of music, while also discussing on the use of new measures in the academic field.

6.1. The musicological characteristics of recorded music

Musical products convey an array of symbolic messages (Wikström, 2009). These messages are anticipated to be deeply influenced by the characteristics of the societies in which they are crafted and/or marketed, regardless of whether they originated or were adopted therein. Cultural products, including music, vary according to their "traits, moods, styles", up to an "infinite variety" (Caves, 2000, p. 6).

We generally consider the social function of music as an area of interest separated from its aesthetics values. Social sciences such as sociology and cultural economics have tried to comprehend, although with different methods, how popular music impacts individuals and communities, while also trying to explain success in operational terms. However, academics such as Simon Frith have previously recognised that, despite success being somehow relatable to sales strategies, promotion resources and audience development, the reasons behind fan fascination towards music itself are not as easily explained (Frith, 1987, cited in Adell, 1998). Musicology has traditionally tried to overcome such gaps by "examining the 'substance' of the music" (p. 42), although with tools that rarely cover the details of the popular genres, as they were designed to be applied to academic music.

Nonetheless, we are aware that musical styles can be broken down into smaller components. When describing the technical features of musical products, we have preferred using musical and/or sonic descriptors. To a certain extent, the differences between the various existing styles of music (e.g., the commonplace conga pattern "tumbao" of salsa music), and even between works within those styles (e.g., the different claves, 2-3 and 3-2, that can be used in salsa), can be expressed by means of a traditional music notation system (e.g., writing down the percussive patterns in a score) and other categorical parameters: features such as time signature, tempo, overall harmonic structure, and melodic scope have been widely used for the analysis of musical pieces. In this regard, scores can acquire a degree of sophistication to even accurately describe dynamics and articulations that closely represent the actual playing of instruments. However, in virtually all cases scores are insufficient to represent timbre and other peculiarities of sound such as loudness. Furthermore, the modes of production of popular music today do not routinely follow a writing-performance-recording process, as it has been the case in academic music. Popular music "is neither conceived nor designed to be stored or distributed as notation, a large number of important parameters of musical expression being either difficult or impossible to encode in traditional notation" (Tagg, 1982, p. 41). Although this author recognises that "a holistic approach to the analysis of popular music is the only viable one if one wishes to reach a full understanding of all factors interacting with the conception, transmission and reception of the object of study" (p. 44), the large volumes of output that we have nowadays, especially those distributed over digital services, creates the need of a mode of analysis that can handle large quantities of data.

For several decades, the recording studio has been the space where arrangements are constructed and even where the songs are created, a tendency that started shortly after World War 2 and that was heavily linked to technological development (Burgess, 2014). Consequently, the recognition of any aesthetic feature of a musical product and further linking to its popularity/attention/success in the marketplace would necessarily require some sort of analysis of the sound recordings. In this regard, there are sonic properties that can be measured with reasonable consistency between different tracks, such as loudness, but others are much more "subjective", "emotional" or "mood-related", such as the degree of "movement" of a piece or how "happy" or "sad" a song can be. Moreover, consumers are not necessarily aware or mindful of the aesthetic qualities of music when deciding which artists/bands to follow. As an example, in 1991, William Hamlen Jr. analysed the relationship between record sales and "vocal quality", measured as the upper harmonics contained in the word "love" sung by a roster of artists that were included in the study. One could argue that a single word may not represent the complete listening experience of a song, or that there are styles that are less dependent on melodic vocal quality, such as rap music and other forms of urban music, but the author nonetheless operationalised a purely aesthetic feature into numerical terms. The results, however, showed that listeners recognised but did not care about vocal quality.

6.2. Objectivising musical features

There have been several attempts to operationalise the musical and/or sound characteristics of recordings. The initiatives are grouped under the umbrella term Music Information Retrieval (MIR), which include several applications such as indexing, cataloguing, recommendation, copyright protection, categorisation –even considering emotional patterns–, et cetera (Murthy & Koolagudi, 2018). The various characteristics that can be extracted and/or computed from an audio signal are classified as low, middle, and high-level features[8]. Low-level features are obtained from very brief audio segments, which make them non-representative for entire songs or significant portions. They correspond to timbral and temporal characteristics such as the root mean square (RMS) energy, a peak loudness parameter.

The low-level features are subsequently integrated into more complex mid-level characteristics, which mainly correspond to the "intrinsic" properties of music. They are then organised in three broad categories: pitch (i.e., fundamental frequency), rhythm (i.e., regularities within the pattern of accents), and harmony (i.e., several notes played simultaneously).

The highest degree of sophistication in this regard is achieved with the high-level features, whose content description encapsulates "the knowledge that an experienced or professional listener would have about [a] piece of music" (Zheng et al., 2017, p. 671). Their nature relates to the perceptual and affective domains (e.g., how energetic and/or joyful a song can feel), and their practical use often involves classification (e.g., by musical genre) or delineating the structure of a composition (i.e., musical form) (Hsu & Huang, 2015). Most indices within the Spotify Audio Features fit into this category, with numerous studies attempting to correlate these variables with

[8] The literature presents diverse views on the classification of features as low, mid, or high-level. For instance, some studies, like Casey et al. (2008), categorise pitch and harmony as high-level features. However, our research primarily draws upon the categorisations set forth by Murthy and Koolagudi (2018).

commercial success (Sciandra & Spera, 2022). However, despite the promising potential of these sophisticated metrics, tasks such as music structure analysis encounter notable challenges related to subjectivity and ambiguity (Nieto et al., 2020). Furthermore, the accuracy of these metrics seems to be moderate at best. For example, during the MIREX[9] 2007 Exchange, the highest accuracy achieved for mood recognition by the evaluated algorithms was 61.5% (Casey et al., 2008).

6.3. The diversity of musical products

In economics, diversity has been a central subject of interest. It has been recognised that higher levels of institutional and technological diversity favour innovation (Stirling, 2007), and, in a broader sense, "diversity is prominent in crucial efforts to promote religious, cultural, racial, and gender equality […] and pluralism" (p. 708). Specifically, cultural diversity "creates a climate in which different cultures can engage in a mutually beneficial dialogue" (Parekh, 2000, p. 168).

Stirling (1998) distinguishes three general properties of diversity: variety, balance, and disparity (Figure 2). Variety refers to "the number of categories into which the quantity in question can be partitioned" (e.g., the number of different record labels that appear on the charts, number of musical styles or the number of songs that enter the charts during a specified period) (p. 39). Balance is "the pattern in the apportionment of that quantity across the relevant categories"; in this case, a more even distribution refers to a more diverse system (e.g., a less concentrated market or less skewedly distributed in terms of intellectual property ownership, chart appearances, etc.) (p. 39).

Finally, disparity refers to "the nature and degree to which the categories themselves are different from each other" (e.g., how different the labels are in terms of organisational hierarchy, overall production methods, etc., or how distant from each other are songs based on their musicological properties) (p. 40). Stirling also recognises that, individually, these properties fail to

[9] Music Information Retrieval Evaluation Exchange: an annual event aimed at evaluating the performance of MIR algorithms.

fully distinguish markets in terms of diversity. To solve it, he proposed a quantitative heuristic that incorporates the three elements, which eventually came to be known as the Rao-Stirling Index.

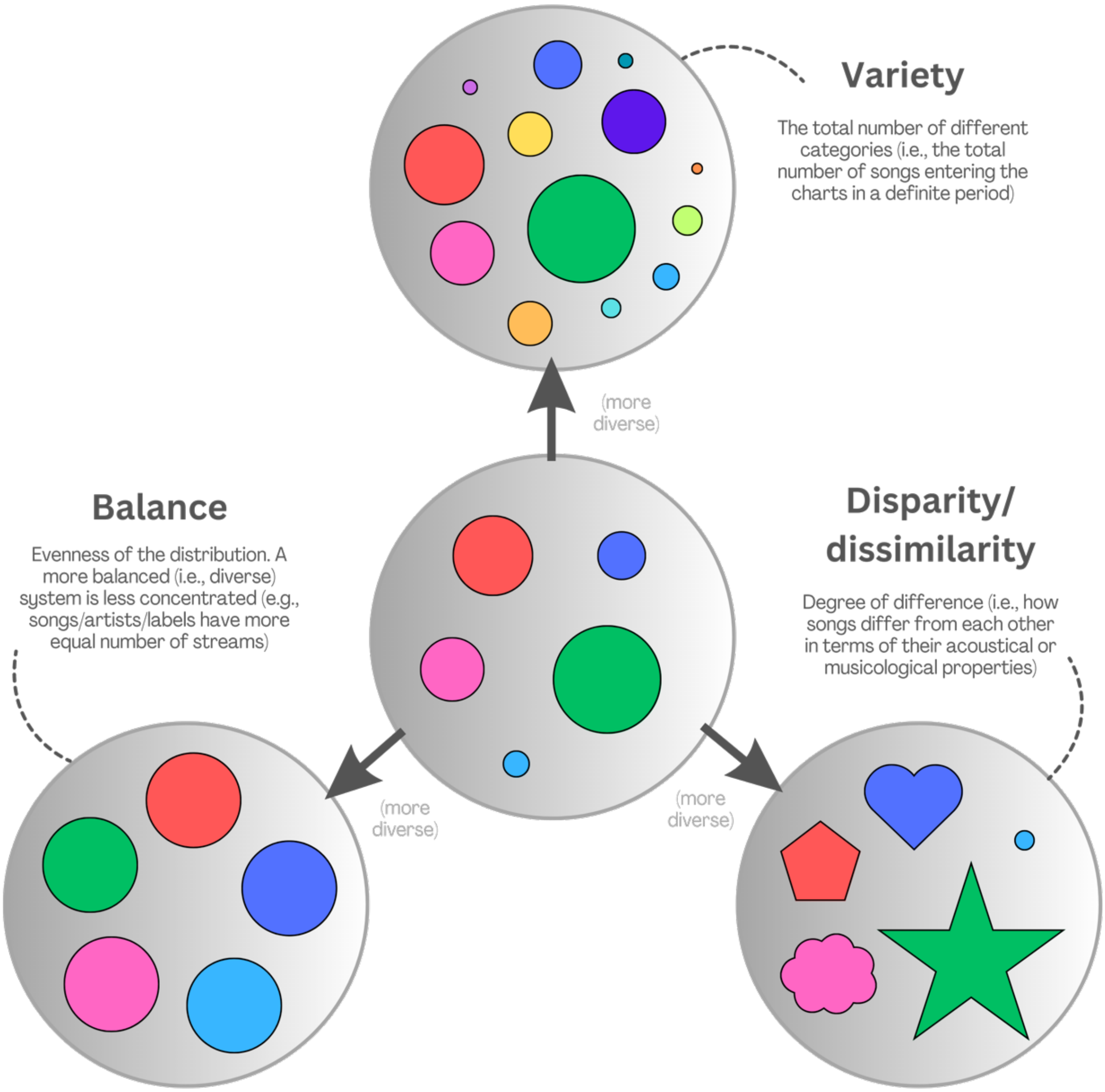


Figure 5. Stirling's properties of diversity with examples from the music industry. Based on Stirling (2007) and Leydesdorff et al. (2019).

While there's ample empirical evidence about diversity in the music industry, many earlier studies, particularly from the sociology realm, are based on limited datasets. Moreover, some research approaches the topic with an oversimplified lens. Most strikingly, there's a notable absence of studies using "objective" measures of disparity.

Using data from Billboard's weekly top lists from 1948-1973, Peterson and Berger (1975) argued that diversity (i.e., Stirling's variety), measured as the number of labels and firms operating in the market at a given year, was inversely proportional to market concentration (i.e., balance), countering the Schumpeterian claim that "the rate of innovation in an industry is a function of market structure" (p. 159), which favoured oligopolistic states. A similar study was conducted by Rothenbuhler and Dimmick (1982), this time during the 1974-1980 period, with comparable findings and conclusions.

Challenging such precedents, Lopes (1992) found that the re-oligopolisation of the phonographic industry wasn't accompanied by a decrease in diversity (i.e., variety): on the contrary, the number of Top 10 hit singles increased dramatically in the 1987-1982 period, despite the decreasing number of labels. Burnett (1990, 1992) found similar results. Christianen (1995), on the other hand, reported that indie labels contribute more to song variety and balance in terms of share of supply (i.e., more titles released) and genres (i.e., indies supply more evenly across different musical styles), respectively. The author also found that, in terms of innovation, debut albums were more frequently released by indie labels, and local artists were more frequently signed by the independent labels.

More recent evidence by Handke (2006) highlighted an increase in label variety in the German market, with more record companies entering the competition, even amidst the profound recession the recording industry faced globally towards the end of the twentieth century. The author highlighted an inverse association between the number of record labels and industry revenues, according to the IFPI. Furthermore, Handke (2010) noted an uptick in title variety during this downturn. Specifically, in 2006 –a recession year–, the number of new titles saw an increase of 54.7% compared to 1998, the final year of the industry's "boom" period. The availability of comprehensive and complete data in the German music market allowed the author to more closely evaluate the full multidimensional label space.

Gallego (2016) studied the situation of the radiophonic industry in Spain, with diversity defined by the number of weeks that the songs survived on radio airplay (i.e., more airplay time corresponding to less diversity). Using data from Promusicae (an institution representing the Spaniard phonographic industry), as provided by BMAT on an annual basis, the author found worsening standardization, as per the Adornoian concept, during the 1978-2012 period. Specifically, there was a notable increase in the average number of weeks songs stayed within the Top 40 positions, rising from a mean of 3 weeks in 1978 to 13.5 weeks in 2012. Interestingly, this trend of prolonged airplay was restricted to individual songs and was not mirrored at the album level: The average chart survival time for albums only slightly increased from 18.1 to 19.3 weeks over the same period.

Academic studies have also sought to understand the connection between variables such as socio-demographic characteristics, cultural influences, and music consumption, which in turn reflect the diversity of music people consume. Early investigations in this field were constrained by small sample sizes and relied on self-reported data (Liu et al., 2018), potentially limiting their scope and not reflecting “real world” consumption, but data quality has improved over time, and research has progressively begun to adopt a multi- or cross-country approach.

Using data from a 1987 survey on cultural preferences among the Dutch population, Van Eijck (2001) reported that musical “omnivorousness” –the variety of music genres an individual appreciates– tends to increase with higher levels of education and improved occupational status. On the other hand, Ranaivoson (2010) attempted to study diversity in the music industry in 69 countries, using a heterogeneous/self-constructed database with data collected from national unions of phonographic producers, and complemented with “more general data” from United Nations the World Bank (p. 6). The author found that cultural diversity didn’t always correlate with the better conditions present in the “most favoured countries” (p. 9), pointing out that, paradoxically, “a higher index of human development leads to less supplied diversity” (p. 225). However, more recently, Woolhouse and Bansal (2013), analysing data on over 180 million mobile phone downloads from Nokia (i.e., more closely reflecting real world consumption), discovered a positive correlation between countries’

Human Development Index (HDI) and the variability of music downloaded. Conversely, they noted that greater download diversity inversely correlated with higher unemployment rates.

7. Musicological characteristics and song diversity – Data

In this section, we provide a description for the gathering of data regarding the musicological characteristics of the Spotify chart tracks, which will serve as the basis of our most sophisticated diversity indicator.

7.1. The Spotify Web API

According to their site, the Spotify Web API "enables the creation of applications that can interact with Spotify's streaming service, such as retrieving content metadata, getting recommendations, creating and managing playlists, or controlling playback" (Spotify, n.d.-e, para. 1). To use the platform, a Spotify account –free or premium– is the first requirement, followed by the creation of an app and an access token request. Although the API is free to use, there are rate limits for the requests made to the platform.

The API is colloquially used to examine the characteristics of individual users' preferences –for example by analysing playlists–, as it can be seen in popular developer sites such as GitHub and Kaggle. It offers both straightforward, composite parameters such a custom popularity index, as well as "raw" numerical data, such as playlist followers. The API has also been previously used as a source of data for research, for example by Pyun et al. (2020), Chun et al. (2021), and Sciandra and Spera (2022).

7.2. Spotify Audio Features database

Spotify Audio Features, the operationalised expression of the musicological characteristics of tracks, are available at the Spotify API, where the included user manual details a total of thirteen different musicological characteristics that are included (Spotify, n.d.-a). Each set of audio features are directly linked to every song unique URI code, which we obtained from the 'uri' column in our Spotify Charts database.

Even though it is possible to connect the Spotify API to external software such as R or use Python to gather the data, we opted to accomplish such task with the help of Stevesie's web scraping tools (https://stevesie.com), a third-party site with a more friendly user interface that can batch-scrap data for multiple songs in a short amount of time. To do so, we filtered out all repeated songs using Microsoft Excel and ended up with a total of 11,824 unique titles that were featured across the 317-week span of our study. We imported the corresponding URI codes directly to Stevesie's platform and downloaded a .csv file containing the thirteen audio features. Despite the process being successful for most of the included songs, it was not possible to retrieve the audio features for three songs (see Appendix).

7.2.1.Data preparation

The data for seven of the thirteen provided variables —specifically, 'danceability', 'energy', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', and 'valence'— originally ranged from 0 to 1. The interpretation differs slightly among these variables (refer to Table), but in general, the attribute's intensity increases as the value approaches 1 (for instance, a track becomes increasingly 'danceable' and 'festive' as its value nears 1). For clarity and consistency, we rescaled these values to a 0-100 range, aligning with how we presented other variables in our research, like the GI.

For the 'key' variable, values were originally numbers ranging from 0 to 11, each corresponding to one of the twelve musical keys (C, C sharp/D flat, D, D sharp/E flat, E, F, F sharp/G flat, G, G sharp/A flat, A, A sharp/B flat, and B). We transformed these into dummy variables to facilitate subsequent modelling. We also adjusted the 'duration' unit from milliseconds to seconds to aid in clarity. As for the 'mode' variable, following its categorical nature, we didn't need further modifications since it assumed only two values: '0' for a minor key and '1' for a major key.

8. Musicological characteristics and song diversity – Variables

In this section, we delve into the musicological attributes of the tracks in our dataset using the variables provided by the Spotify API. At present, Spotify’s audio features analysis is driven by The Echo Nest analyser. The foundational principles of this audio analysis are rooted in the music cognition segments of Tristan Jehan’s (2005) doctoral thesis. Jehan’s research focused on the “extraction and use of acoustic metadata” or the “objective” content (p. 32), adopting a psychoacoustic approach to the listening model. Originally an independent entity, The Echo Nest was later integrated into Spotify's portfolio in 2014 (Cookson, 2014).

While Jehan aims to provide an objective assessment of musicological attributes of audio tracks, Spotify's Audio Features inherently present stark differences in their measurement goals. Therefore, instead of categorising them as “objective” and “subjective” based on rudimentary music theory, we classify all thirteen features considering their automated extraction and/or calculation. We bifurcate them into (1) low and mid-level features, and (2) high-level characteristics.

Further, we introduce the concept of Stirling’s disparity and build several diversity indices. However, it is essential to contextualize the nature of our dataset: it primarily reflects a subset of consumed diversity. Our emphasis is on the tracks that have soared in popularity on Spotify, rather than the exhaustive list of titles the platform offers. This focus narrows our lens to the hits, offering insights into prevailing tastes but potentially side-lining the vast array of lesser-known tracks that contribute to Spotify’s rich tapestry of offerings.

8.1. The low and mid-level features

The tonality of a song (‘key’) corresponds to one of the twelve musical keys used in Western music, following the standard Pitch Class notation (numbers ranging from 0 to 11), each one of them separated by a semitone. In the case of the Spotify API, altered keys are nominated as sharp keys, and not flat keys. It is also specified that, when no key is detected, a value of -1 is assigned. It is also worth noting that we couldn’t find any indication in the Spotify API documentation as to how the algorithm deals with tonality changes within a song, potentially misrepresenting more complex arrangements where more than one key is used throughout the musical piece.

The duration of a track is a relatively straightforward measure and is originally displayed in milliseconds. However, to facilitate the reading, we converted such values to seconds. The mode of a song is "the modality (major or minor) of a track, the type of scale from which its melodic content is derived" (Spotify, n.d.-a, key section). The tempo of a track is "the speed or pace of a given piece and derives directly from the average beat duration" (Spotify, n.d.-a, tempo section), it is expressed in beats per minute (BPM). In this regard, we should mention that the tempo of a song is dependent of the time signature, or "a notational convention to specify how many beats are in each bar (or measure)" (Spotify, n.d.-a, time_signature section). For example, a song can be written in 4/4, with a tempo of 80. However, the same song can be in 2/2, and in this case the tempo would double up to 160. The Spotify API only considers quarter note divisions for the time signature, so in essence it cannot capture these subtleties.

Loudness is measured in decibels (dB) in the Spotify API. Here, it is defined as "the quality of a sound that is the primary psychological correlate of physical strength (amplitude)" (Spotify, n.d.-a, loudness section). There is no data regarding which specific type of unit is being used by the name "decibel", but it is reasonable to assume that it is not a peak measurement.

8.2. The high-level features

The other characteristics found within the Spotify API are indices that assign numerical values from 0 to 1 to a group of variables that are related to the emotional, mood-related, and other "subjective" perceptual content of the tracks, most likely constructed using the previously described low and mid-level features. Acousticness is described as "a confidence measure from 0.0 to 1.0 of whether the track is acoustic" (Spotify, n.d.-a, acousticness section). Danceability "describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity" (Spotify, n.d.-a, danceability section). Energy "represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. […] Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy” (Spotify, n.d.-a, energy section). Instrumentalness “predicts whether a track contains no vocals” (Spotify, n.d.-a, instrumentalness section). Liveness “detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live” (Spotify, n.d.-a, liveness section). Speechiness “detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g., talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. […] Values below 0.33 most likely represent music and other non-speech-like tracks” (Spotify, n.d.-a, speechiness section). Finally, valence describes “the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g., happy, cheerful, euphoric, while tracks with low valence sound more negative (e.g., sad, depressed, angry)” (Spotify, n.d.-a, valence section).

8.3. Diversity indices

Numerous methods are available for measuring diversity, with a primary focus on balance—often termed 'evenness'—and disparity or dissimilarity. We offer a summary of several conventional two-dimensional measures, yet our emphasis is on a three-dimensional integrative index, drawing inspiration from Stirling (2007). This index captures the aesthetic differences of music reasonably well, while maintaining rigorous quantitative integrity.

8.3.1. The Simpson Diversity Index (SDI)

The SDI originates from the Simpson Dominance Index, with values from 0 to 1 in its primordial formulation that are inversely proportional to the degree of diversity of a system. We transformed such values to a “true” Diversity Index (SDI), where diversity becomes directly proportional to the SDI value. The SDI is described as “the probability of any two individuals drawn at random from an infinitely large community belonging to the same species” (Magurran, 2004, p. 114) that is obtained using the proportions of the individuals. The mathematical expression is:

$$\mathrm{SDI} = 1 - \sum p_i^2$$

Where p_i is the proportion of individuals in the ith species. By subtracting 1 minus the Dominance Index, we obtained the SDI. Furthermore, it is worth mentioning that, numerically, the Dominance Index is the same as the HHI. For this application we have not considered labels but only the songs as separate entities (i.e., song balance and not label balance).

8.3.2. The Shannon-Wiener Diversity Index (SWI)

Another frequently used diversity index based on species richness and evenness, the SWI is "based on the rational that the diversity, or information, in a natural system can be measured in a similar way to the information contained in a code or a message" (Magurran, 2004, p. 106). The SWI increases when there are more species in the sample, and with a more balanced sample. It is calculated as minus the sum of the natural logarithm[10] of the relative abundance of each class within a system. The relative abundance in our case is measured by the number of streams of each song relative to the total number of streams for any specific week. The formula of the SWI is:

$$\mathrm{SWI} = -\sum_{i=1}^{S} p_i \ln(p_i)$$

Where 'S' is the total number of classes (in our case, S=200 as we have 200 rank positions each week) and p_i is the proportion of the ith class.

8.3.3. Mean coefficient of variation (MCV)

Our simplest measure for determining diversity in terms of [unweighted] disparity is the coefficient of variation (CV), expressed in the following formula:

$$MCV = (\frac{\sigma_1}{\mu} + \frac{\sigma_2}{\mu} \dots + \frac{\sigma_n}{\mu}) \div n$$

[10] Although the SWI can be calculated using other logarithm types, it is the natural logarithm that gets employed more frequently.

In this expression, σ is the standard deviation of a group of datapoints, μ is its mean, and n is the total number of features. Because in our case we had nine variables, then n =9. We excluded dummy categorical variables ('key', 'mode', and 'time_signature') and the 'instrumentalness' variable, as its values were very close to zero, even after rescaling by a factor of 100. Therefore, the audio features included for the calculations were 'danceability', 'energy', 'loudness', 'speechiness', 'acousticness', 'liveness', 'valence', 'tempo', and 'duration_s'. We calculated the mean of all their CVs for each week and each country. It is important to highlight however, that by calculating the mean CV across all the selected audio features we didn't assign a specific weight to any of them, therefore assuming their impact to song disparity is the same.

8.3.4. The Rao-Stirling Diversity Index (RSI)

Stirling (2007) proposed a series of criteria that a general diversity index should have, which included properties of monotonicity (i.e., the increase in diversity based on a single property when the others were constant), allowance of aggregation of the three properties in divergent contexts, et cetera. He proposed a "quantitative heuristic" consisting of "the sum of pairwise disparities, weighted in proportion to contributions of individual system elements" (p. 712)[11].

$$D = \sum_{ij(i \neq j)} d_{ij} \cdot p_i \cdot p_j$$

In this formula, "p_i and p_j are proportional representations of elements i and j in the system (balance/evenness) and d_{ij} is the degree of difference (disparity/dissimilarity) attributed to elements i and j" (p. 712). Researchers have called this formula the Rao-Stirling Diversity Index. In our case, because the number of different categories remains constant (i.e., there are always 200 positions each week on Spotify Charts), there is no contribution of the "variety" parameter. Therefore, we are only considering balance (the relative proportion of streams for each song, each week, on the total number of weekly streams) and disparity (the degree of difference between

[11] Although Stirling (2007) further proposed a more complex index (Δ) which incorporated two additional exponents (alpha and beta) that aimed to address "all the possible relative weightings on balance and variety/disparity", for simplicity, we have assumed, as many others, a value of 1 for both exponents.

songs). In our research, the RSI quantifies the relative uniqueness of each track compared to its counterparts appearing on the charts for any given week. This disparity measure, however, is based solely on the expression of extreme values within a seven-day period and might not fully encompass the intricate aesthetic attributes of individual tracks in a broader context. Nonetheless, given the nature of our dataset, we believe that the RSI represents the most suitable measure for articulating the extent of "musical" differentiation between songs.

As to measuring song uniqueness, the attributes we considered were 'danceability', 'energy', 'loudness', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo', 'key', and 'mode'. For simplicity, because most of the songs were in 4/4, we created a dummy variable for time signature ('time_sig_cat'), where '1' represents tracks in 4/4 according to The Echo Nest algorithm, while '0' accounts for songs that are written in every other time signature (e.g., 5/4, 3/4, et cetera), or not recognised at all. Again, we didn't consider any weighting scheme for the musicological characteristics, assuming that all of them contribute to the relative song uniqueness in the same proportion.

Past research has employed the Euclidean distance to determine disparity between paired elements within a system. In our study, given the presence of three categorical variables ('key', 'mode', and 'time signature') among our selected attributes, we instead relied on using the Gower distance. This method is better suited for mixed data types, both continuous and categorical (Akay & Yüksel, 2017). Our calculations were executed in Python 3, using packages such as 'pandas', 'sklearn.impute', 'sklearn.preprocessing', 'numpy', and 'gower'. Given the sporadic missing values in our dataset, which comprised less than 0.1% of the total, we engaged in an imputation process: substituting missing continuous data with the mean and missing categorical data with the mode. This step was crucial since the Python function for Gower distance isn't compatible with missing values. Once we structured the Gower matrix, we weighted it according to the relative proportions of each element. This process culminated in a singular D value (RSI) for each week across every country in our panel. For ease of interpretation, we amplified the indices by a factor of 100.

Table 6. Selected variables for audio features and song diversity						
Audio features						
Name (abbreviation)	Variable type	Definition	Unit	Data source	Notes	Descriptives
Key (from c to b)	Binary categorical (dummy)	Indicates whether a song is in any of the twelve musical keys, either major or minor (specified by "mode"). This uses the standard Pitch Class Notation.	None	Spotify API	c_key=C c_sharp_key=C# or Db d_key=D d_sharp_key=D# or Eb e_key=E f_key=F f_sharp_key=F# or Gb g_key=G g_sharp_key=G# or Ab a_key=A a_sharp_key=A# or Bb b_key=B For any of the dummy variables: 0=No 1=Yes	Mode=1
Acousticness (acousticness)	Continuous numerical	Measure of how acoustic (i.e., using instruments such as acoustic guitars, piano, etc.) a track is.	Index	Spotify API	From 0 (less acoustic) to 100 (more acoustic).	Mean=24.77 Median=18.4 SD=21.44 Min=0 Max=99.2
Danceability (danceability)	Continuous numerical	Measure of how suitable a track is for dancing. Uses a combination of other features (not specified by Spotify).	Index	Spotify API	From 0 (less danceable) to 100 (more danceable).	Mean=71.8 Median=73.8 SD=11.41 Min=7.83 Max=98.5
Duration (duration_s)	Continuous numerical	The duration of a track.	Seconds	Spotify API		Mean=213.93 Median=207.61 SD=48.68 Min=33.87 Max=3,653.96

Energy (energy)	Continuous numerical	A measure of intensity and activity. Other features such as dynamic range, loudness and timbre are contributory.	Index	Spotify API	From 0 (less energetic) to 100 (more energetic).	Mean=68.7 Median=70.9 SD=14.62 Min=2.17 Max=99.9
Instrumentalness (instrumentalness)	Continuous numerical	The probability of a track containing vocals versus being purely instrumental.	Index	Spotify API	From 0 (“less instrumental” or “more vocal”) to 100 (more likely to be purely instrumental). A value over 50 is intended to represent an instrumental track.	Mean=0.4 Median=0 SD=3.57 Min=0 Max=99
Liveness (liveness)	Continuous numerical	The presence of an audience in the track, or the probability that the track was performed live instead of purely recorded in a studio.	Index	Spotify API	From 0 (less likely a live performance) to 100 (more likely a live performance).	Mean=17.78 Median=12 SD=15.09 Min=1.34 Max=99
Loudness (loudness)	Continuous numerical	A feature that subjectively correlates to the strength or intensity of sound.	Decibels (dB)	Spotify API	A negative number. Values that are closer to 0 represent a louder track.	Mean= -5.28 Median= -4.93 SD=2.06 Min= -23.02 Max= 1.91
Mode (mode)	Binary categorical	Modality of the track. In the case of Spotify, it only intends to distinguish between major and minor keys.	None	Spotify API	0=Minor key. 1=Major key.	Mode=1
Speechiness (speechiness)	Continuous numerical	Detects the presence of spoken words and exhibits the probability of a track being made completely out of spoken words (i.e., a talk show).	Index	Spotify API	From 0 (“less spoken”) to 100 (“more spoken”). A value between 33 and 66 could correlate with music and spoken words (i.e., rap music), and a value below 33 most likely represents music.	Mean=10.75 Median=7.25 SD=8.88 Min=2.32 Max=88.4
Tempo (tempo)	Continuous numerical	Determines how fast or slow a track is, and it is related to the subdivision length for any song’s musical measure.	Beats per minute (BPM)	Spotify API		Mean=123.28 Median=113.06 SD=33.31 Min=48.75

						Max=214.03
Time signature (time_signature)	Binary categorical (dummy)	Specifies how many beats appear in each musical measure.	None	Spotify API	1=A song written in 4/4, according to the Spotify algorithm 0=A song written in any other time signature	Mode=4
Valence (valence)	Continuous numerical	Represents the musical "positiveness" of a track (i.e., how "happy" or "sad" a track is).	None	Spotify API	From 0 to 100. A lower value represents a "sad", "depressed", or "angry" song, whereas a high value represents a "happy", "cheerful", or "euphoric" song.	Mean=61.97 Median=64.8 SD=20.93 Min=3.2 Max=98.9
Song diversity						
Name (abbreviation)	**Variable type**	**Definition**	**Unit**	**Data source**	**Notes**	
Simpson Diversity Index (simpson)	Continuous numerical	The probability of any two individuals drawn at random belonging to the same species. $$\text{SDI} = 1 - \sum p_i^2$$ Where p_i is the proportion of individuals in the ith species.	None	Spotify Charts	This is a measure of evenness (balance), as the number of classes was fixed in our dataset.	Mean=99.14 Median=99.14 SD=0.11 Min=98.66 Max=99.38
Shannon-Wiener Diversity Index (shannon)	Continuous numerical	Minus the sum of the natural logarithm of the relative abundance of each class within a system. $$\text{SWI} = -\sum_{i=1}^{S} p_i \ln(p_i)$$ Where 'S' is the total number of classes and p_i is the proportion of the ith class.	None	Spotify Charts	This is a measure of evenness (balance), as the number of classes was fixed in our dataset.	Mean=5.04 Median=5.04 SD=0.06 Min=4.77 Max=5.2
Mean coefficient of variation	Continuous numerical	The mean ratio of the standard deviation (σ) to the mean (μ) of every included audio feature (total features: n).	None	Spotify API	A simple, unweighted measure of disparity.	Mean=35.95 Median=35.88 SD=1.39

		$$MCV = (\frac{\sigma_1}{\mu} + \frac{\sigma_2}{\mu} ... + \frac{\sigma_n}{\mu}) \div n$$				Min=32.54 Max=47.1
Rao-Stirling Diversity Index (rao_stirling)	Continuous numerical	Balance/disparity-weighted variety. $$D = \sum_{ij(i \neq j)} d_{ij} \cdot p_i \cdot p_j$$ d_{ij}=disparity for elements i and j p_i, p_j=proportion of elements i and j	None	Spotify Charts and Spotify API	A compound measure of diversity based on variety, balance, and disparity. Again, variety was constant due to the fixed 200 song positions in the charts.	Mean=22.16 Median=22.05 SD=1.08 Min=18.62 Max=25.5

9. Musicological characteristics and song diversity – Descriptive statistics

In this section, we outline descriptive statistics of the primary variables used in our analysis, focusing on the musicological attributes of tracks featured in the Spotify Charts from our selected ten Latin American countries. Additionally, we provide descriptive statistics for our diversity metrics. To facilitate a clearer understanding, we have included line charts highlighting the most salient findings. For a visual overview of the nuances of each of the ten countries, please refer to the Appendix.

9.1. Musicological characteristics

Table 7. Descriptive statistics for Spotify Audio Features across all countries

	Mean	Median	Mode	SD	Range	Minimum	Maximum
Key	5.32	6.00	1.00	3.67	11.00	0.00	11.00
Danceability	71.89	73.80	74.40	11.41	90.67	7.83	98.50
Energy	68.70	70.90	77.30	14.62	97.73	2.17	99.90
Loudness	-5.28	-4.93	-6.33	2.06	24.93	-23.02	1.91
Mode	0.58	1.00	1.00	0.49	1.00	0.00	1.00
Speechiness	10.75	7.25	4.32	8.88	86.08	2.32	88.40
Acousticness	24.77	18.40	17.60	21.44	99.20	0.00	99.20
Instrumentalness	0.40	0.00	0.00	3.57	99.00	0.00	99.00
Liveness	17.78	12.00	10.10	15.09	97.66	1.34	99.00
Valence	61.97	64.80	68.00	20.93	95.70	3.20	98.90
Tempo	123.28	113.06	104.82	33.31	165.27	48.75	214.03
Duration (s)	213.93	207.61	205.72	48.68	3620.09	33.87	3653.96
Time Signature	3.96	4.00	4.00	0.26	4.00	1.00	5.00

Danceability scores consistently ranked high across all countries, boasting an average of 71.89 and a modest standard deviation of 11.41. This narrow deviation highlights the prevalence of danceable tracks, indicating limited outliers. Nonetheless, there were some tunes with very low danceability values, the most extreme of them scoring 7.83 on this measure. Curiously, this specific outlier corresponded to "Weightless Part 1", an ambient instrumental track by Marconi Union, dubiously advertised as "the most relaxing song in the world" (Shepherd et al., 2022).

The line charts revealed a distinct upward trend for danceability across the duration of the study (see Figure 6). This trend was particularly pronounced in the initial half of the 317-week span. It reached a plateau around week 150, followed by a phase of oscillation, then culminating in a final increase towards the study's conclusion. Despite these fluctuations, the values consistently stayed towards the higher end of the spectrum.

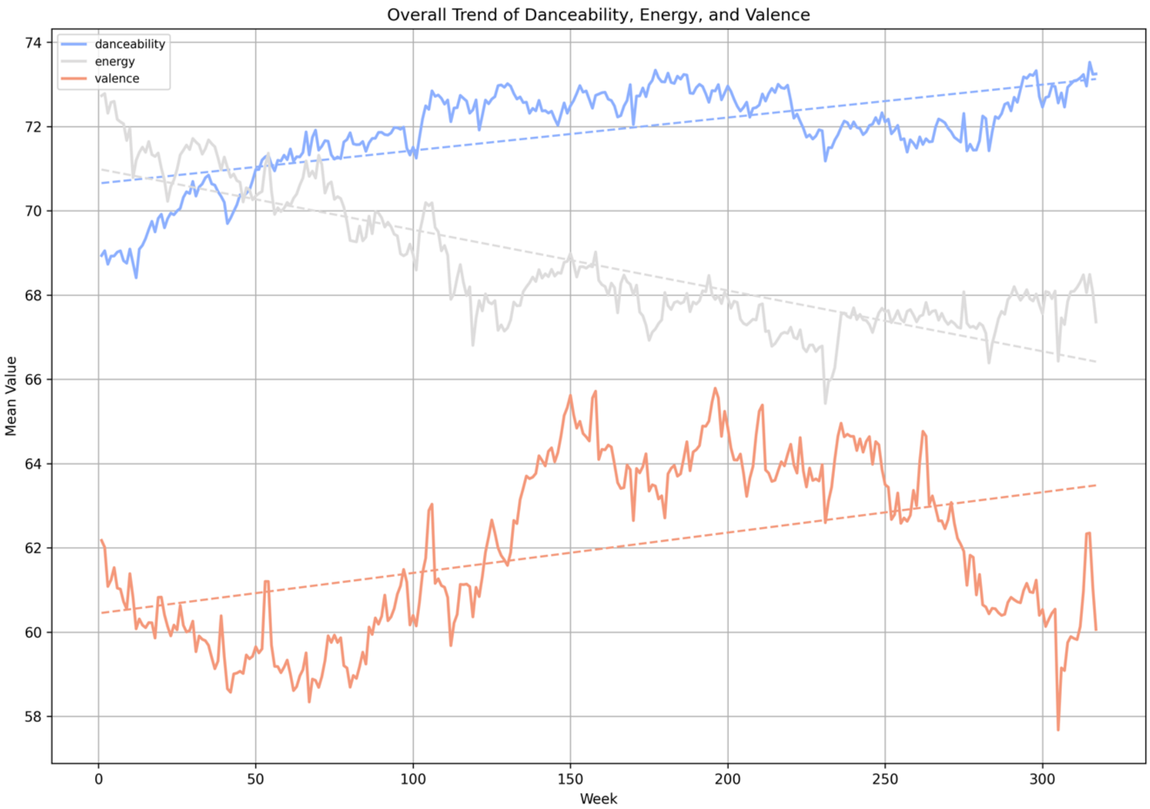


Figure 6. Line charts illustrating trends in three high-level features. Observe the divergent trajectory of energy in contrast to the trajectories of valence and danceability.

Energy levels were notably high across all countries, averaging at 68.7 with a relatively modest SD of 14.62, indicating some variability but not to an extreme extent. Nonetheless, a general downward trend in energy levels was observed throughout the study, which seemed counterintuitive against the backdrop of rising danceability values. In contrast, valence displayed an upward trend, with similar patterns in all countries and still consistently high values over 50, though with more variability (SD of 20.93). This trend points to a prevailing taste for music that is perceived as more optimistic and uplifting, despite the fluctuating pattern that such feature exhibited over time.

While valence and danceability generally increased together, suggesting a correlation between the move-inducing quality and joyfulness in songs, a closer examination of the data revealed occasional divergences in their trajectories, particularly at the outset and conclusion of the study. This mismatch, especially between the overall song energy and the other two features, suggests there might be underlying issues with the precision of these high-level features. We recognise, however, that the assumption that more danceable songs are inherently "happier" may be an oversimplification.

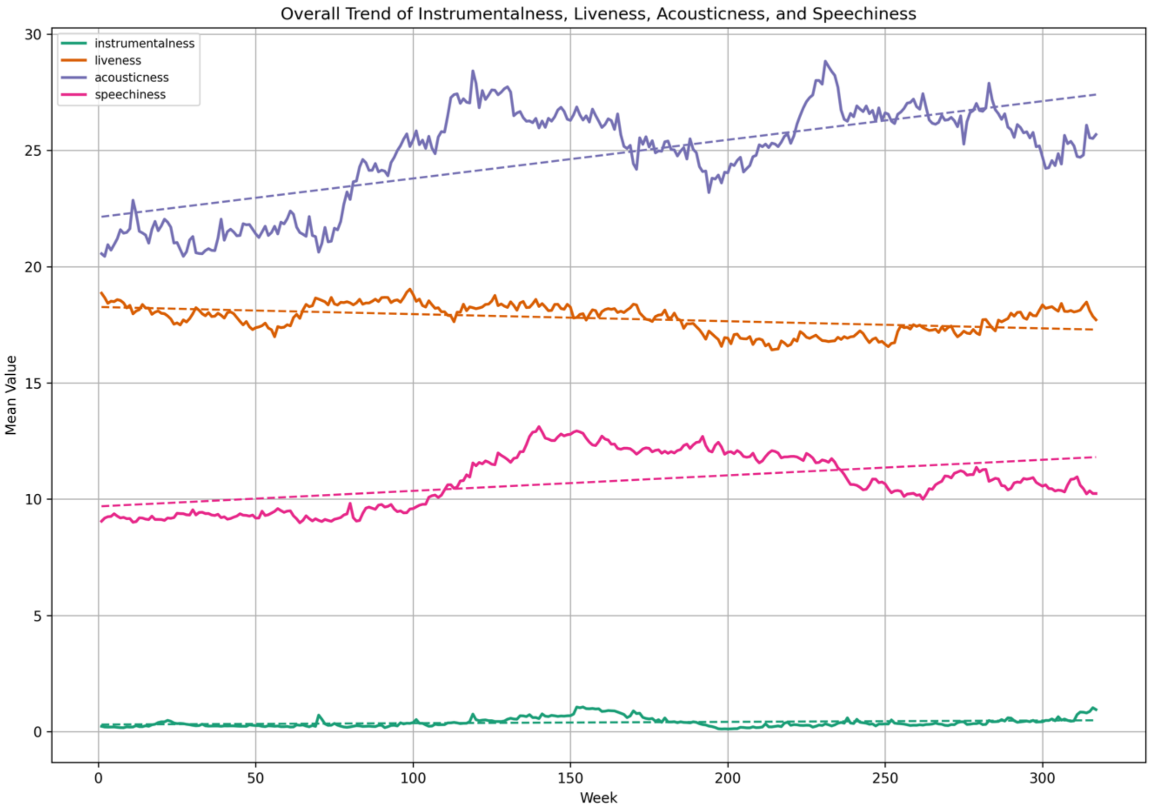


Figure 7. Line charts illustrating trends in the remaining high-level features.

The values for the remaining four high-level features consistently fell on the lower end of the spectrum, yet certain trends were discernible (see Figure 7). For instance, acousticness and speechiness followed an upward trend, potentially indicating a rising preference among listeners for genres traditionally rich in acoustic instrumentation – a hallmark of popular Latin American styles like salsa, cumbia, and Latin pop. Simultaneously, there appears to be an increased interest towards genres where spoken word prevails over sung vocals, such as trap, reggaeton, and Brazilian funk *carioca*, popular urban music styles in Latin America. On the other hand, instrumentalness values were generally very low in all countries, and we did not find any remarkable trends worth mentioning. Regarding the liveness variable, most countries exhibited comparable behaviours and maintained similar values during the entire period of the study. This time, a diminishing trend was observed throughout the 317-week period, which could suggest an increasing preference for the particularities of studio-recorded music over live performances such as concerts, festivals, and recitals.

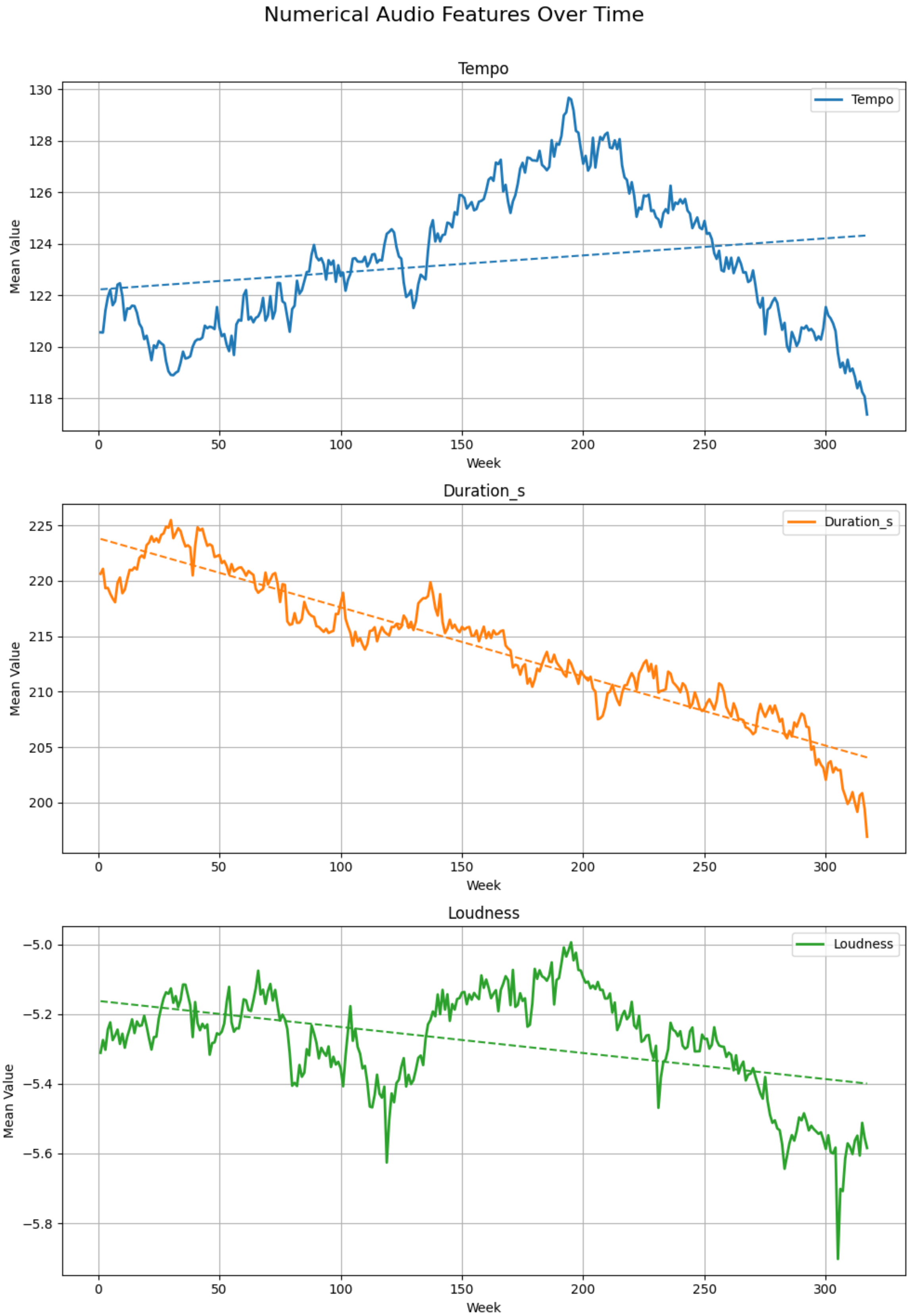


Figure 8. Line charts illustrating trends in three of the low and mid-level features, particularly the numerical variables. Note that song duration exhibited a clear diminishing trend over time, while tempo has an inverted "V" shape.

Figure 9. Line charts illustrating tendencies in the categorical audio features over time.

Latin genres, particularly those of Afro-Caribbean, Afro-Cuban, and Afro-Antillean descent, are renowned for their festive nature. While we didn't have direct access to data specifying song genres for this study, we can draw insights from a 2021 consumer preference survey conducted across twelve Latin American countries. Apart from Brazil, this survey indicated a significant preference for "Latin music" in most of these nations, although the precise styles remained unspecified (Statista, 2021). Given this backdrop, it's not surprising to find high danceability, energy and valence values in our results. However, it is crucial to note that streaming platforms, from which this data likely originates, may not capture the full spectrum of musical consumption in these regions.

Regarding the more "objective" variables, comprising low and mid-level features (see Figure 8), song tempo was consistently high, with a mean of 123.28 beats per minute (BPM) across all countries, although it is unclear to what extent the algorithms can capture the half-time or double-time feeling that the songs may exhibit. On the other hand, the weekly mean loudness values across countries were consistently similar and remained relatively high, at -5.28, significantly above Spotify's recommendation of -14 LUFS for mastering. This might be indicative of a prevailing trend among mixing and mastering engineers to produce tracks with diminished dynamics and a more pronounced, aggressive sound – a practice rooted in the "loudness wars" that began around the 1990s. Finally, it appears that over time, songs became shorter.

In terms of the categorical low-level audio features (refer to Figure 9), there was a marked preference for songs in the key of C# or Db, with the least favoured being F, though both keys exhibited an increase in popularity over time. The preference for C# or Db is particularly intriguing, given that these keys are not commonly preferred in musical practice due to their complexity—C# includes seven sharps, and Db has five flats, both of which can be more challenging for musicians. When considering song modality, there was a clear inclination towards major keys, consistently exceeding 55%, with the exception of the latter part of the study period. Conversely, minor-key songs made up less than 40% of chart presence for the majority of the timeframe, except for a rise noted after around week 230, suggesting a declining trend for major-key songs and a corresponding rise for minor-key ones. Additionally, a vast majority of songs were composed in 4/4 time, aligning with the characteristic rhythmic patterns of popular music genres, including various locally prevalent Latin American styles.

9.1.1. Aesthetic characteristics according to label type

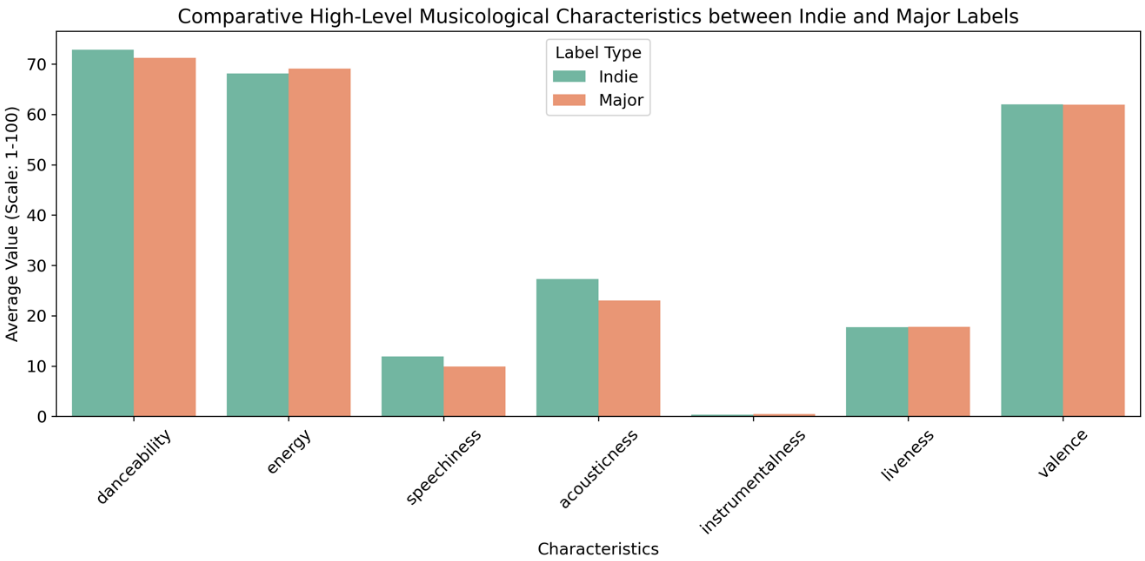


Figure 10. Bar graphs illustrating differences between major and indie labels regarding their high-level features.

Although average values were broadly similar across indie and major label chart tracks, there were some differences (Figure 10). Indie label hits, on average, were slightly more danceable (72.86 vs. 71.22 mean danceability) but paradoxically less energetic (68.11 vs. 69.11 mean energy). They also featured more spoken content (11.94 vs. 9.91 mean speechiness value), had a higher acoustic presence (27.29 vs. 23.01 mean acousticness), and were less instrumental (instrumentalness average of 0.33 vs. 0.44). The demand for live content in the tracks was nearly identical (liveness of 17.73 for indies vs. 17.81 for majors), and both had similar levels of "joyfulness", as indicated by the mean valence (62.02 for indies vs. 61.93 for majors).

Additionally, examining specific low and mid-level features (Figure 11) revealed that major label tracks, on average, had a similar loudness level compared to indie tracks in the charts, with mean values of -5.24 dB LUFS for majors and -5.34 dB for indies. Indie songs on the charts also tended to be slightly faster, averaging a tempo of 123.91 BPM compared to 122.83 for major label songs. Most notably, songs from indie labels were, on average, longer in duration, clocking in at 219.69 seconds compared to the major label average of 209.89 seconds.

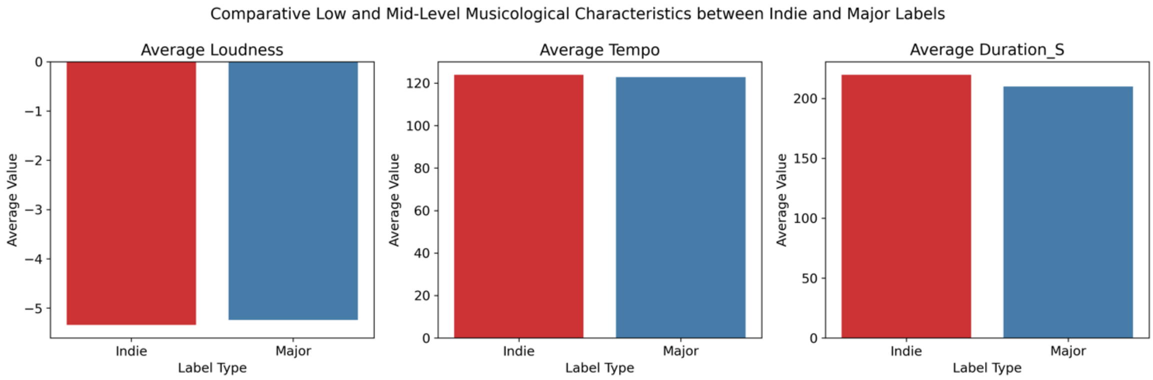


Figure 11. Bar graphs showing differences between major and indie labels regarding selected low and mid-level features.

In our deeper analysis of major industry players, distinct variations emerged among the conglomerates. Notably, SME tracks stood out for their louder average volume, registering a mean loudness of -4.85 dB LUFS, in contrast to UMG's -5.57 dB and WMG's -5.35 dB. Furthermore, SME's songs were, on average, faster, with a tempo of 124.90 BPM, compared to 123.28 BPM for UMG and 118.80 BPM for WMG. For more detailed information, refer to Appendix F, Table F1.

9.2. Diversity indices

We plotted the different values over time using line charts (Figure 12), very much like our previous sections. For the MCV charts, a single significant outlier in Peru that lied in the extreme upper end had to be filtered out. We established a threshold of 42 before creating the plots in Python to avoid distortion on the rest of the countries' charts.

Our results showed that the evenness measures indicated only a very faint upward trend in diversity throughout the study. Specifically, the SDI varied only slightly, with a narrow a range of 0.72 across all countries, with an average value of 99.16 and a notably low standard deviation of 0.11. The SWI yielded similar results, suggesting that there were mostly no changes in product balance: individual successful songs continued to receive roughly the same proportion of streams over the length of the study. In contrast, when assessing the variation of music and sound characteristics over time as an indicator of diversity, we observed more pronounced variability. The MCV registered higher standard deviations (an overall SD of 1.39), began with fluctuating patterns, and demonstrated an upward trend from the mid-point of the study period. Meanwhile, the RSI exhibited a clearer overall upward trend during the study, implying that, when weighted against the number of streams, the most consumed music became aesthetically more diverse as the study period progressed.

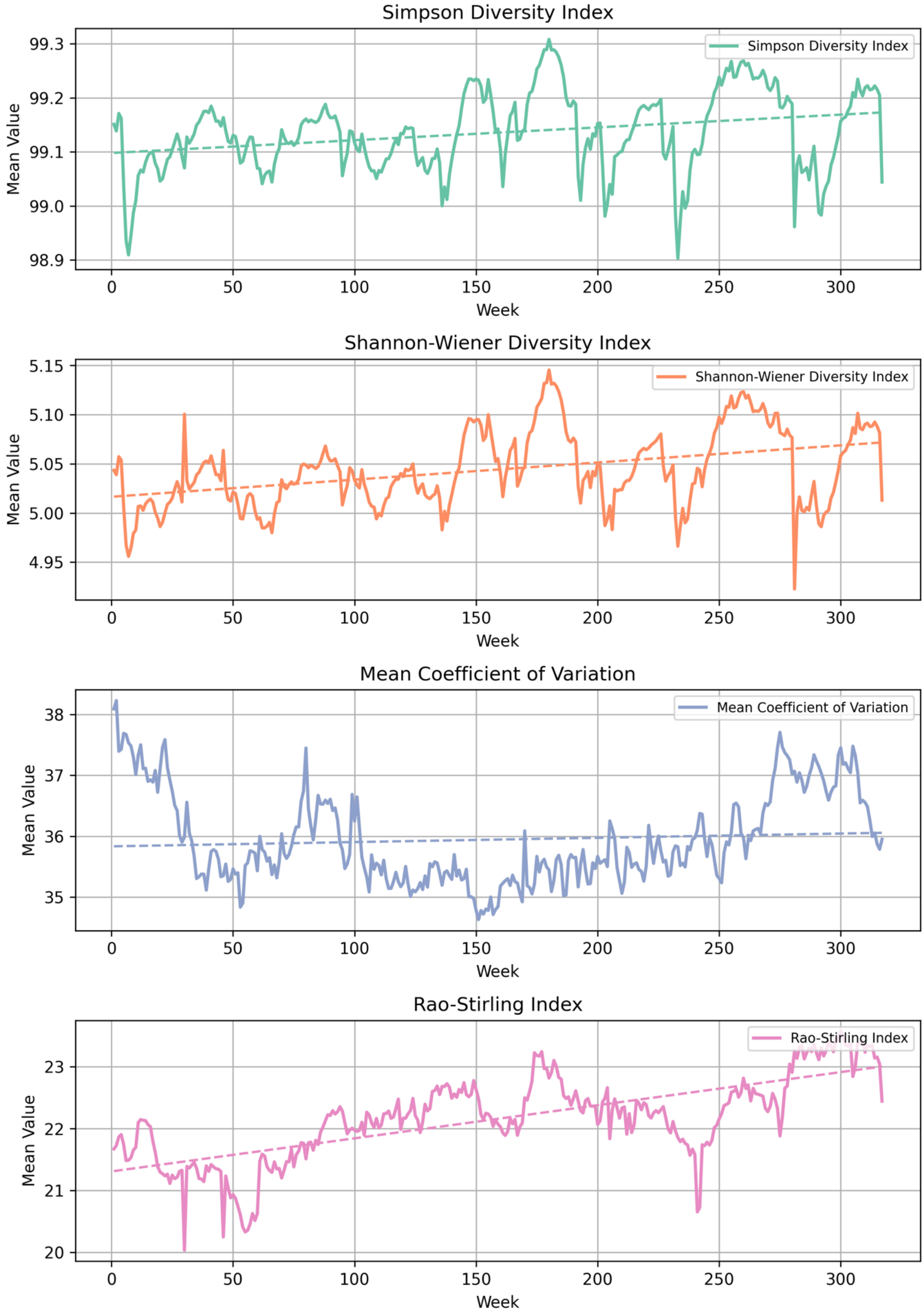


Figure 12. The evolution of diversity measures over time.

10.Labels, concentration, and diversity – Inferential statistics

Utilising inferential statistics, we aim to shed light on two intertwined phenomena. On the one hand, label balance, as denoted by the weekly percentages of label affiliation over the 200 positions, was our main independent variable. On the other hand, an integrative metric for diversity will serve as the dependent variable and confronted against concentration in our attempt to explain our second research objective: uncover to what extent concentration of ownership of streams is correlated with song diversity.

10.1. Diversity panel data regression

Because we will be working with weekly indices, our 3-way panel dataset (i.e., country, rank position, and week) was effectively converted into a 2-way panel, retaining 'country' as the entity dimension, and 'week_number' as the time dimension. The 'rank' column deemed no longer meaningful as each index was calculated using the data from the weekly top 200 positions in every case.

10.1.1. Data preparation and model

In choosing the optimal model for our analysis, we postulated that each country, while sharing regional similarities, would possess a unique set of unobservable characteristics –spanning cultural, economic, and technological dimensions–. These intrinsic factors could notably influence the relationship between stream concentration/ownership and diversity. Considering our study's timeframe of approximately six years, we hypothesise that such unobservable attributes would remain largely constant throughout the period. Consequently, we opted for a fixed effects regression model (FE) and performed two staged regressions. The first only considers the aggregate market share of major record labels as the independent variable ('share_st_majors'), while the second regression incorporated the three main conglomerates' shares ('st_share_universal', 'st_share_sony', and 'st_share_warner'). We established 'st_share_indies' as our reference category and excluded it from our model to avoid perfect collinearity.

In both cases, we chose 'rao_stirling' as our dependent variable, as it was the more sophisticated measure we calculated for song diversity, and the only one that included two of the three Stirling's diversity dimensions (variety being fixed by the nature of the dataset): it relied on audio features for the disparity/dissimilarity component and assigned a relative weight as per the stream count, while also being more independently calculated by summarising variables that were not accounted for in the independent variables' calculations (i.e., share of streams mainly uses the stream count, but the RSI relies on a separate database as provided by the Spotify API).

Regarding control variables, we focused on demand-side sociodemographic, macroeconomic, and technological factors that may influence consumption diversity. We have previously mentioned that there is a body of academic literature supporting the relationship between such "macro" factors with individual-level musical preferences. Specifically for this work, we chose gross national income (GNI) converted to [2017] international dollars based on purchasing power parity (PPP), as expressed by the indicator "GNI per capita, PPP (constant 2017 international $)" ('gni_percapita_ppp_2017'). Regarding technological factors, we chose internet penetration, measured with the indicator labelled "Individuals using the Internet (% of population)" ('internet_access'), following the rationale that a more widespread availability of internet access would account for greater access to information, nurturing more diverse tastes, and a comparable argument for the portability and convenience of music consumption through mobile devices, especially smartphones.

We performed linear interpolation using Python 'scipy.interpolate' package to transform the yearly data into weekly data. Further, to control for specific confounding factors related to platform growth, we used the weekly sum of streams for each of the 317 weeks in each country, as a proxy of the differences between different nations' market sizes and how they evolved over time. As the source for the macro-level data, we relied on the World Bank Open Data, which offers yearly numbers on a per-country basis.

Unfortunately, we could not control for supply-side variables, specifically regarding changes in the total number of songs in Spotify, as this information isn't disclosed beyond sporadic and inconsistent announcements by company representatives. Further, any interpolation attempt would be riskier in this case, as per the recent surge of AI-powered massive content creation and uploading to DSPs, which would likely follow a non-linear behaviour.

The models, therefore, acquired the following final configuration:

Model 1:

$$\text{rao\_stirling} = \alpha + \beta_1 \text{share\_st\_majors}_{it} + \beta_2 \text{gni\_percapita\_ppp\_2017}_{it} + \beta_3 \text{internet\_access}_{it} + \beta_4 \text{sum\_st}_{it} + u_i + \epsilon_{it}$$

Model 2:

$$\text{rao\_stirling} = \alpha + \beta_1 \text{share\_st\_universal}_{it} + \beta_2 \text{share\_st\_sony}_{it} + \beta_3 \text{share\_st\_warner}_{it} + \beta_4 \text{gni\_percapita\_ppp\_2017}_{it} + \beta_5 \text{internet\_access}_{it} + \beta_6 \text{sum\_st}_{it} + u_i + \epsilon_{it}$$

10.1.2. Diagnostic tests for model assumptions and corrections

To assess multicollinearity, we conducted a Variable Inflation Factor (VIF) test. All independent variables returned VIF values under 5, indicating no significant multicollinearity concerns. To inspect the presence of first-order autocorrelation, the Wooldridge test—a method tailored for panel data (as supported by Drukker, 2003)—was conducted. The test yielded p-values exceeding 0.05, suggesting the absence of autocorrelation. However, the Wooldridge test for heteroskedasticity (Wooldridge, 2001) produced p-values approaching zero, strongly providing evidence of this problem. Consequently, we employed robust standard errors, incorporating them into our Python code.

10.1.3. Results

When examining the model fit and significance, we see that the overall R-squared for our first model is 0.3424. This means that our model explains approximately 34.24% of the variation in the dependent variable, RSI. Although this isn't a particularly high value, due to the complexity of the phenomenon we are studying, we still consider the results as insightful. The FE model seems adequate considering the value of F-statistic (robust) at 409.31, and the associated P-value of <0.0001. In contrast, our second model, which disaggregates the major labels' share of streams by conglomerate, performed better in explaining the variance, at 38.56%.

When looking at the individual predictors, we found remarkable results. In the case of ownership of streams, it seems that, overall, the majors' presence in the charts was associated with decreased diversity: a 1% increase in majors' share of streams corresponded to an RSI drop of around

Table 8. Results for aggregate majors regression (model 1)

	Coefficient	Std. Error	95% CI	p
Intercept	18.778	(0.5348)	[17.730, 19.827]	<.001
Share of Majors	-0.0246	(0.0018)	[-0.0281, -0.0211]	<.001
GNI Per Capita PPP 2017	0.0001	(0.00002448)	[0.00008899, 0.0002]	<.001
Internet Access	0.0355	(0.0030)	[0.0296, 0.0415]	<.001
Sum of Streams	2.466e-09	(5.554e-10)	[1.377e-09, 3.555e-09]	<.001
R-squared		0.3424		
No. Observations		3158		
F-Statistic (robust)		409.31		

Note: CI = Confidence Interval. The dependent variable is the Rao-Stirling Index value. The model includes entity fixed effects.

Table 9. Results for major conglomerates regression (model 2)

	Coefficient	Std. Error	95% CI	p
Intercept	18.657	(0.5496)	[17.580, 19.735]	< .001
Share of Universal	0.0132	(0.0034)	[0.0066, 0.0199]	< .001
Share of Sony	-0.0442	(0.0028)	[-0.0497, -0.0387]	< .001
Share of Warner	-0.0209	(0.0033)	[-0.0273, -0.0144]	< .001
GNI Per Capita PPP 2017	0.00008724	(0.00002518)	[0.00003788, 0.0001]	< .001
Internet Access	0.0417	(0.0030)	[0.0358, 0.0476]	< .001
Sum of Streams	4.697e-09	(5.381e-10)	[3.642e-09, 5.752e-09]	< .001
R-squared		0.3856		
No. Observations		3158		
F-Statistic (robust)		328.71		

Note: CI = Confidence Interval. The dependent variable is the Rao-Stirling Index value. The model includes entity fixed effects. P-values are for two-tailed tests.

0.0246. More specifically, both SME and WMG have a negative correlation with the RSI. In the case of SME, for a 1% increase in stream ownership, RSI was expected to decrease by 0.0442 units, with a significance level of $p<0.05$. Similarly, a 1% increase in WMG ownership of streams correlated with RSI decreasing by 0.0209 units. However, most surprisingly, a 1% increase in UMG's weekly share of streams correlated with increased diversity –although with a small effect size– as it led to an RSI uptick of 0.0132 units.

A correlation between independent labels and increased diversity was expected. Handke (2010) stated that, within the music industry, it is fringe suppliers who "generate a disproportionally large share of radical innovations concerning musical content" (p. 171). However, for the case of UMG's contribution to diversity, two plausible explanations arise. The first and most simple explanation lies at the number of sublabels that are part of UMG: in our dataset, UMG had the most significant number of unique sublabels at 162 (8.82%) compared to the other majors (see Table 9). While we have focused on label conglomerates from the beginning of our work, in practice each sublabel functions as a relatively autonomous unit, which can also be reflected in the aesthetic characteristics of their output.

Table 10. Majors' absolute and relative number of subordinate labels		
Label	**Count**	**Percentage of total (%)**
Universal Music Group	162	8.82
Sony Music Entertainment	95	5.17
Warner Music Group	67	3.65
Independent labels	1157	62.98
Total	1837	100

The second possible explanation lies in the proliferation of exclusive distribution agreements between major labels over the past few years, which effectively guarantee that the creative control of the musical productions remain in the hands of the independent record companies and artists. If these types of contracts between art and commerce are more common in UMG than SME or WMG, this would explain why one major conglomerate can contribute to diversity, while the others cannot.

Regarding our control variables, we noted a marked improvement in both the fit and significance of our model upon their inclusion. Prior to this, the R-squared for model 2 accounted for merely 22.68% of the variance. Moving forward, our results show that, while better economic conditions —as expressed by the 'gni_percapita_ppp_2017' indicator— were associated with increased diversity, the effect size was minuscule when compared to the other variables, although retaining statistical significance. Further, we found that increasing internet access correlated with a more diverse consumption: a 1% increase in our measure was associated with un uptick of 0.0355 and 0.0417 in RSI for model 1 and model 2, respectively.

In sum, these results support the notion that there are quite intricate dynamics between market dominance, as represented here by ownership of streams, and diversity, as expressed by the RSI. Furthermore, we advise caution when attempting to draw more general applications from these results, as our model seem to have left a significant proportion of the variance unexplained. For a better predictive power, other variables should be taken into consideration in further studies, while also diving deeper into the types of relationships between the majors and their catalogue artists: the mere distinction of the type of contract (i.e., production versus exclusive distribution) could shed some light on how majors, being incentivised by relatively less financial risk adoption, avoid standardisation focusing on the distribution of fringe artists.

11.Further discussion

Overall, the longitudinal analysis of the countries in our study revealed consistent trends in both stream ownership and musicological characteristics, with subtle and occasional differences. However, while the core aim of this research is to illuminate the wider landscape of digital music consumption in Latin America, this section will concentrate on Brazil. This country presents a unique case, often deviating from the regional trends in several respects.

11.1. The case of Brazil

Music as a cultural expression in Brazil has been always important, as “the songs, genres and dances that are the product and raw material of centuries of cultural hybridisation are essential to how Brazilians perceive themselves as a people in relation to the rest of the world” (Howard-Spink, 2012, p. 77). It is also well known that Brazilians consume a high proportion of their own music: in 2010, local artists represented 65% of the total market.

11.1.1. Record labels and Spotify consumption trends in Brazil

Historically, Brazil emerged as a significant player in the recording industry quite early. For instance, Casa Edison, Brazil's first record label (*gravadora* in Portuguese), was the third most important phonograph producer worldwide in 1903. Founded by Fred Figner, the label initially offered a catalogue of imported music along with some recordings made in Rio de Janeiro, which were then sent to Germany for conversion into actual phonographs (De Almeida, 2002; Darbilly & Vieira, 2012). A milestone was reached in 1912 when Casa Edison, after signing a contract with the International Talking Machine company, facilitated the establishment of a phonograph factory in Rio de Janeiro, known as Fábrica Odeon. This led to the release of the first entirely Brazilian phonograph. Other companies soon entered the market, with Fábrica Phonographica União launching in 1919 and Fábrica Popular in 1920.

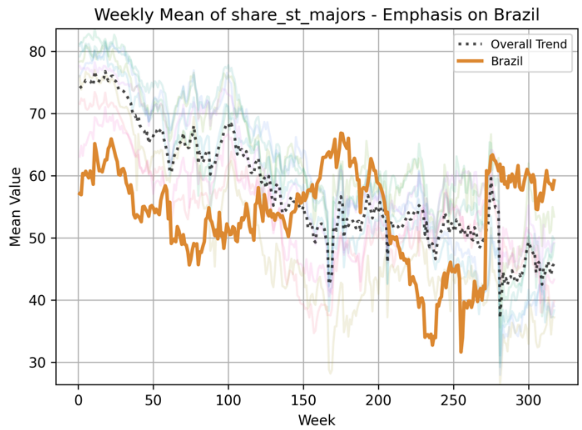
Further descriptions of the Brazilian market are available through more recent reports. According to the Observatorio Latinoamericano de Música Independiente (OLMI):

> Brazil is the largest market in Latin America in terms of recorded music revenue, with sales in 2020 of USD 306 million. Regarding the independent industry, it is the Latin American country with the oldest labels, given that approximately 56% of these companies were created before 2010 (in Chile only 6% and in Argentina only 40% [were created before that year]) […] The Brazilian music industry represents the largest music market in Latin America. Despite this, its integration with the rest of the region is rather limited, since its most popular genres nowadays (sertanejo, pisadinha, funk carioca) have only managed to position themselves in their own country. Among the reasons that could explain it are the language and the lack of a cultural link with these musical expressions outside of Brazil. (OLMI, n.d.)

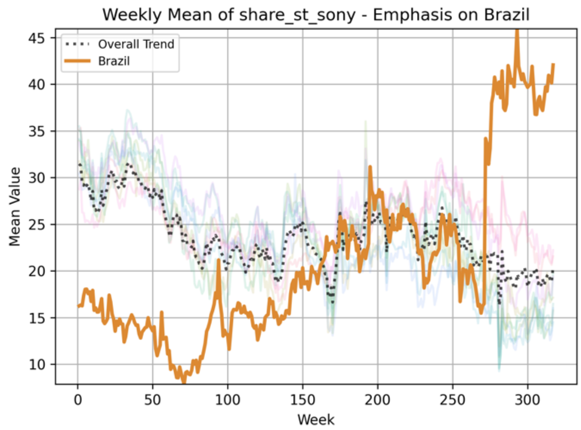
It is not surprising that one plausible explanation of the prevailing cultural separation of Brazil from the rest of Latin America, or at least from the other South American countries, is its language. In this regard, Brazil is the only country in the American continent where Portuguese is an official language (World Data, n.d.), whereas in most of the other Latin American countries, Spanish is the official language.

Another indicator of Brazil's relevance in the global music industries is how often it has been included in music industry publications before the most recent Latin American expansion, and especially in reports issued by the IFPI. The differentiated attention given to Brazil by institutions such as the IFPI was also related to the privileged place the country occupied in the fight against piracy, especially during the early 2000s, as "the cultural industry in Brazil formed as part of the country's general process of economic development" (Bishop, 2005, p. 465). The only other country enjoying this differentiated preference was Mexico, as the two were disproportionately larger markets compared to other Latin American countries.

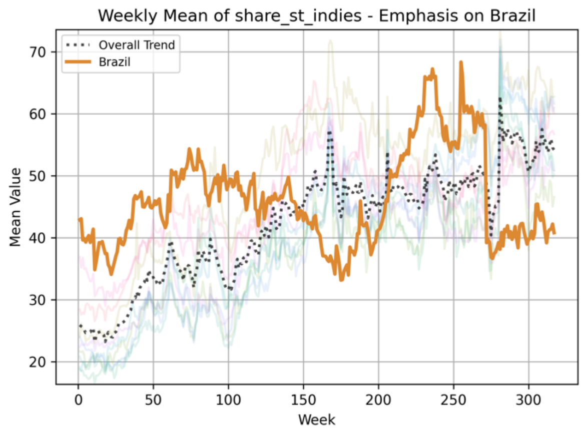
The importance of the independent sector in Brazil has been recognised before and described as representing the "true dynamism" of its music industry (Howard-Spink, 2012, p. 85). In line with this notion, although only considering data on Spotify's streaming consumption behaviours, independent artists and labels seem to have been relatively resilient during recent years. Compared to the other countries, Brazil started strong regarding indie labels' chart dominance. However, after the acquisition of Som Livre by SME, the other countries caught up and even surpassed Brazil in terms of indie presence. The dynamics of record labels in Brazilian territory, particularly regarding the majors, are likely different from the other Latin American countries (Darbilly & Vieira, 2012).



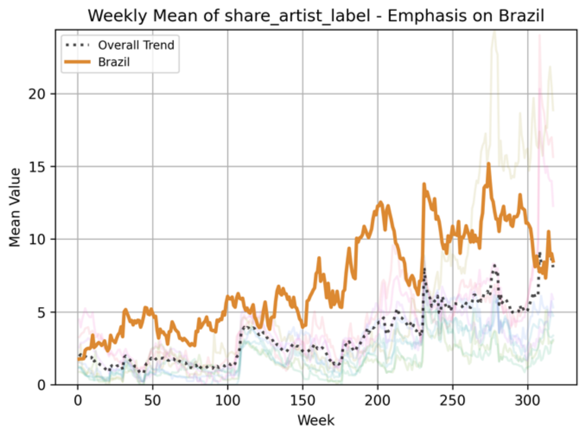
(a)



(b)



(c)



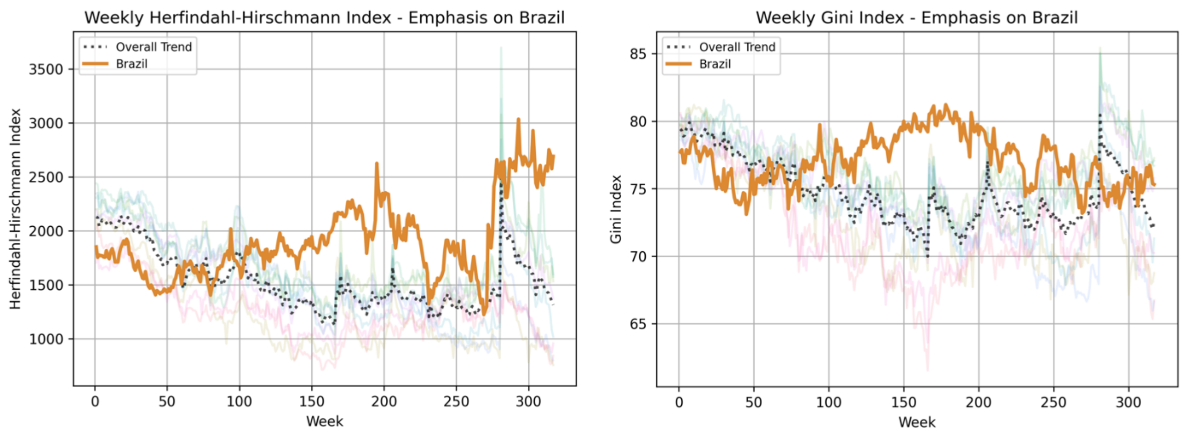
(d)

Figure 13. Line charts illustrating the contrasting patterns of selected record labels in Spotify Charts.

In our study, we observed that in Brazil, independent labels initially had a greater share of streams compared to major labels, primarily due to SME's notably low market share (Figure 13). Even after SME expanded its portfolio by incorporating Som Livre, its average market share remained the lowest among the ten countries studied, at 21.11%. However, this was accompanied by a comparatively higher SD of 8.89%, understandable after the sudden increase in market share after the acquisition of the Brazilian independent company. We also noted a comparatively more significant number of artists ascending to the charts who appeared not to be affiliated with any label, whether major or indie. This was indicated by entries in the 'source' column that matched the artist's name, suggesting self-made releases.

11.1.2. Concentration, inequality, and diversity in Brazil's Spotify Charts

The evolution of concentration and inequality in Brazil at the label level was also unique, as denoted by the contrary motion that it exhibited from the beginning of the study, compared to all the other countries (Figure 14). According to the GI values that we obtained, overall, the Brazilian streaming landscape is more unequal than the other countries, especially during the middle portion of the study.



(a) (b)

Figure 14. Concentration and inequality indicators at the label level in Brazil. Again, observe the contrary motion of the evolution of market concentration compared to the other countries.

Because so many of the preferred artists and genres are local, it is understandable that indies are a better fit to produce records in such distinctive musical styles. Despite the relatively long history of major record labels in Brazilian territory, their catalogues seemed to be an imperfect match to satisfy the local tastes and preferences in Spotify. The exception, of course, was SME, which gradually acquired more relevance in the charts and significantly boosted its presence after the acquisition of Som Livre. This may also reflect in the striking differences between the preferences for certain songs' musicological characteristics across time (Figure 13). While we didn't dive into categorical descriptors of tracks (i.e., song genre), evidence of Brazilians' preferences for national music is available. Through a 2021 online survey with 1455 participants in the hands of the research department at Globo media conglomerate, it was determined that *sertanejo* was the most popular genre among Brazilians, with 43% of positive responses. Other Brazilian styles also ranked high on consumers' preferences: *música popular brasileira* (MPB) accounted for 28% of respondents' affinity, *samba/pagode* for 24%, and *forró/piseiro/arrocha* for 21% (Sintonia com a Sociedade, 2021). Another study (Opinion Box, 2022) showed similar results, with Brazilian styles taking good proportions of audience preferences: sertanejo (48%), MPB (44%), pagode (32%), samba (25%) and piseiro (19%). All of this may account for the profound song differences between Brazil and the rest of our selected Latin American countries.

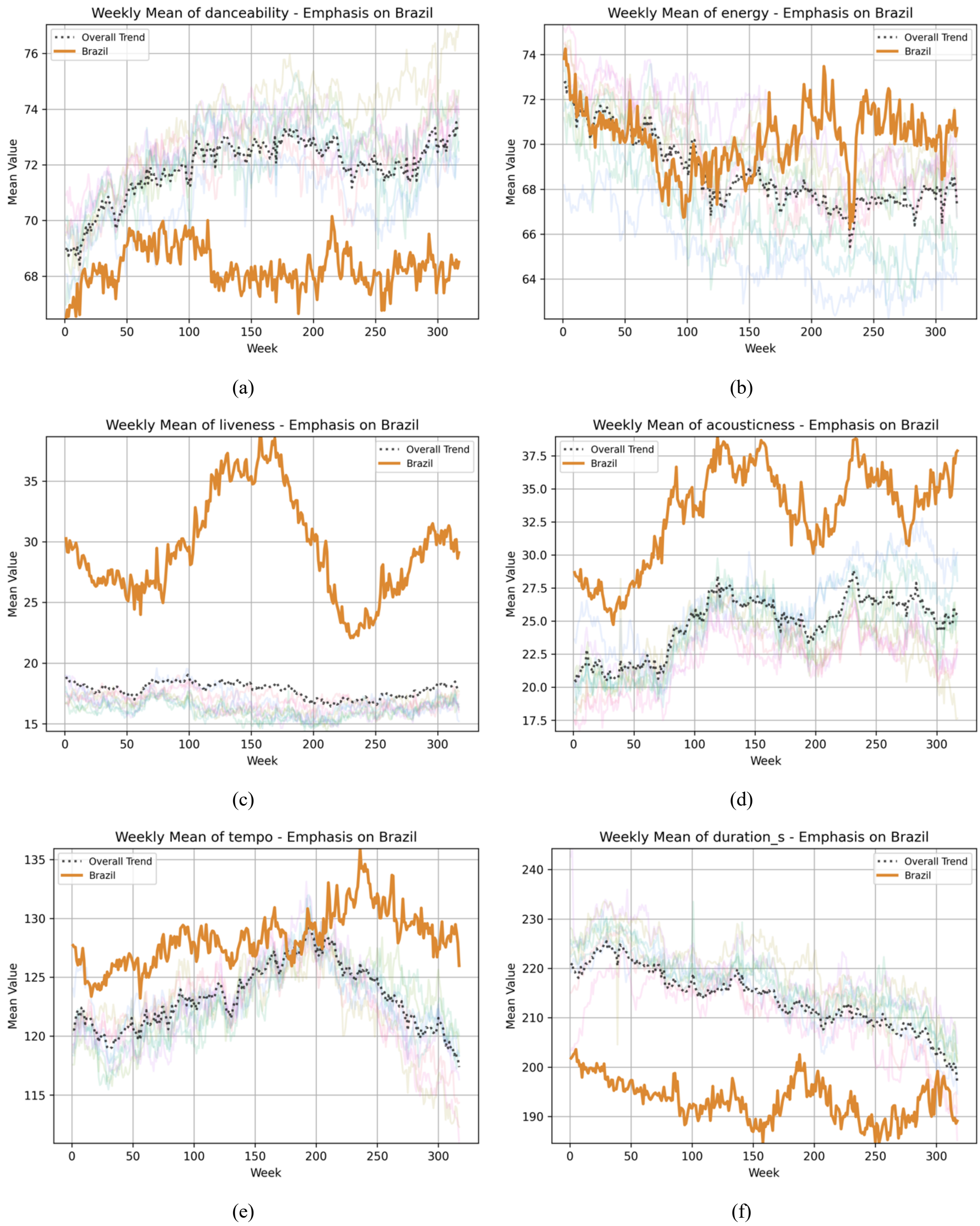


Figure 15. Selected audio features for tracks appearing in Spotify Charts Brazil.

In our examination of individual audio features in Brazil, a distinct aesthetic preference emerged (Figure 15). On average, Brazilian listeners showed a lower preference for danceable tracks, with a mean danceability value of 68.3%, compared to the overall average of 71.89%. However, this finding appears to contradict the musical genre preferences reported in surveys, as most popular Brazilian local genres seem to be very festive, euphoric, and dynamic. This discrepancy might indicate limitations in The Echo Nest's algorithm to accurately identify 'true' danceability in popular Brazilian genres. Additionally, from the midpoint of our study period, Brazilian tracks exhibited an increase in energy. Furthermore, there was a notable preference for live and acoustic music among Brazilian consumers, more so than in other countries included in our study.

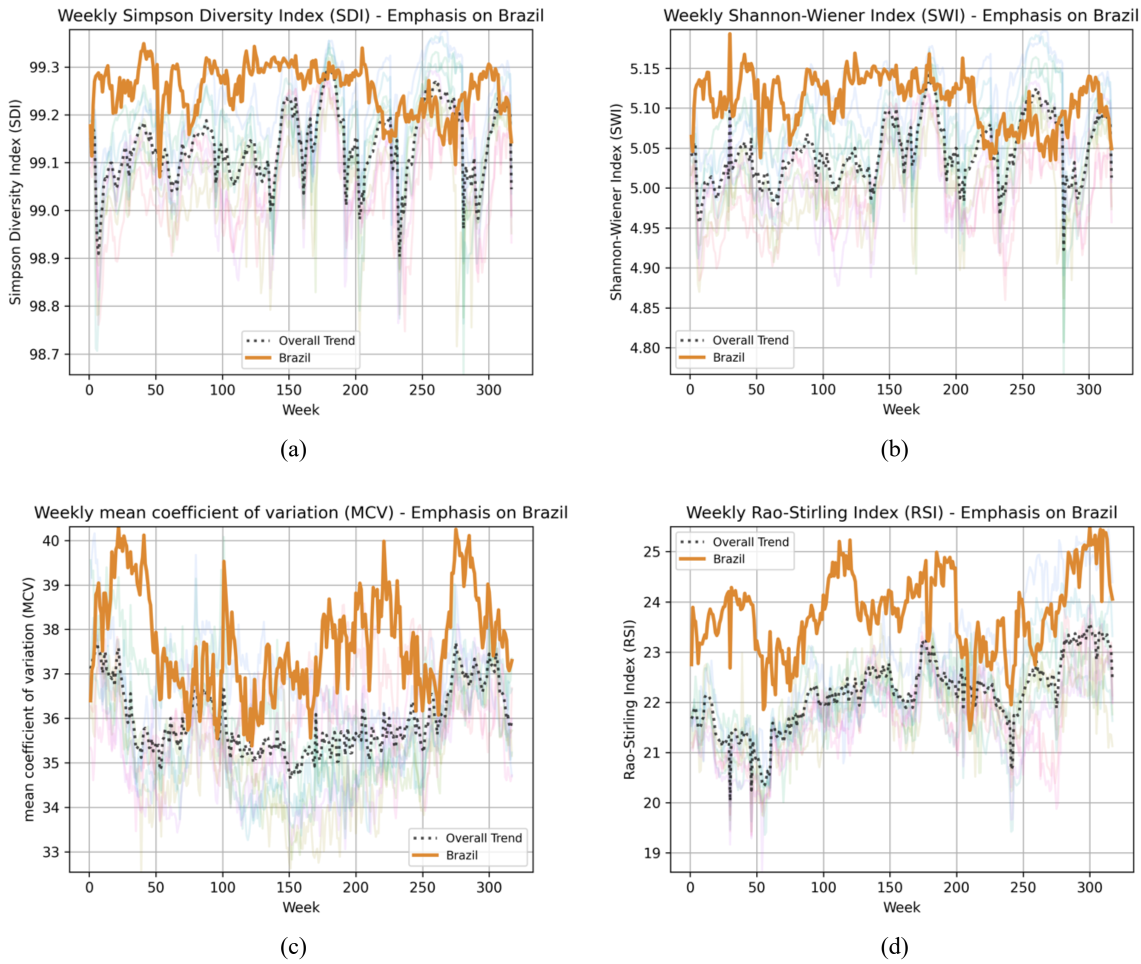


Figure 16. Song diversity measures with emphasis on Brazil. Notice the overall higher values in all measures.

Finally, Brazil also accounted for the maximum song diversity across all measures, especially when considering the evenness-plus-disparity indicators such as the RSI (Figure 16). This is perhaps the strongest indicator, coming from our study, that Brazil demand-side characteristics for digital music are also unlike any other Latin American country.

11.2. Spurious long-tail effects?

In our study, evidence seems to support the preliminary notion that relative long-tail suppliers and independent artists are acquiring a more advantageous position following the growth of the streaming platforms, as denoted in Chapter 5. Label-level concentration indicators showed a general declining tendency over time, with majors losing presence in the charts while indie tracks gained more visibility/success through the years. An initial evaluation of these findings could make them appear promising for independent artists, in line with the notion that the dissemination of new technologies and wide availability of digital distribution services allow for more amateur artists to find niche markets and fight against obscurity. However, when focusing on the song level, the additional streams that the least popular tracks obtained over time, independent of their label affiliation, remained inferior compared to what the "head" portion of the distribution retained at the end of the study. Therefore, despite the popularisation of digital platforms, fringe offerings would seem to be still at a disadvantage compared to the most popular tracks.

A second issue should be considered. Recent reports on streaming services have shown an alarming rise regarding track uploads to the platforms. In 2019, Spotify's CEO Daniel Ek claimed that around 40,000 tracks were uploaded daily to the platform (Ingham, 2019). In 2021, estimations raised to 60,000 tracks per day (Ingham, 2021). In May 2023, music data company Luminate issued a report claiming that, on average, 120,000 tracks are now being uploaded to streaming services daily, with artificial intelligence accounting for a significant proportion of them (Ruza, 2023a). Spotify and other streaming platforms follow a payment model where "everybody gets something" and "everybody weights (roughly) the same". Until recently, there was no threshold for royalty distribution: long-tail artists, having minuscule proportions of monthly listeners, received a proportional amount of money for the few reproductions they get, following the *pro rata* model. Because so many tracks are uploaded, and considering the payment approach, the monetary rewards are expected to get noticeably diluted. In this study, because we didn't have access to data on the extreme low end of the distribution, we remain ignorant of how the “true” tail portion has been affected by the recent proliferation of “automated” music, but it is reasonable to speculate that it has become even flatter.

On a related account, there are two ways we can think of major labels to respond to the apparent threat that such dilution poses. Firstly, following the logic of optimising search costs, new features within the streaming platforms design such as the recently added Spotify’s vertical discovery (Dredge, 2023) could allow for more visibility to be attained by a small roster of artists, particularly coming from major labels. Secondly, partnership programs could be established, following the YouTube model, where, to qualify for monetisation, a few requirements could be made mandatory, such as having a minimum number of all-time streams for an artist/band or a minimum number of subscribers. This would translate into major record label artists immediately qualifying for royalty payments. Interestingly, as this thesis was approaching completion, Spotify announced a shift towards a new royalty model that introduces a threshold: starting early 2024, only songs with a minimum of 1,000 streams will be eligible for monetisation (Ingham, 2023; Mulligan, 2023).

Other related actions such as the takedown and/or labelling of AI-generated tracks are already in motion, by companies such as Spotify and Deezer, the latter following an initiative by UMG (Ruza, 2023b). However, the time and resources needed for such distilling actions may make the whole task inefficient.

11.3. Towards a profile of the non-Western music industries

Our approach has been atomised and thorough in our attempts to better comprehend the nature and dynamics of the Latin American music industry. Several remarkable insights have been drawn from our results and even from the historical accounts we have included in our literature review, which support the notion that Latin America is a region with a significant degree of differentiation compared to the global and Western music industries. As we stated in the introductory portion of this work, the interest in these nations is justified following the development of the digital music platforms and increasing affiliation of users that have established streaming as one of their preferred means of music consumption.

Yet, Latin America isn't unique in this upward trajectory. Recent IFPI reports indicate that regions outside traditional Western markets have also blossoming musical economies. For example, in 2022, the Middle Eastern African region was spotlighted for its exceptional growth in streaming numbers, and by 2023, Sub-Saharan Africa had assumed this leading position (IFPI 2022, 2023). We posit that nuanced, accurate characterizations lead to more informed decision-making, especially when considering creative industries as significant contributors to the economy of a country. Nevertheless, it is crucial that these assessments incorporate other revenue streams of the recording industry, while also integrating data from both the live and publishing sectors.

12. Conclusions, recommendations, and limitations of the study

In our research, which was focused on the Latin American music industries, we pursued two primary areas of inquiry. First, in terms of ownership concentration, we sought to discern if the rise of online music services had manifested either superstar or long-tail effects. Our second line of inquiry centred on the type of music provided, particularly concerning its musicological characteristics. Here, our focus was on the evolution of the tracks' aesthetic profiles, while also examining the trends regarding song-level diversity. Further, we aimed to investigate on the extent to which labels have enriched or diminished the diversity of offerings, controlling for other related factors. To comprehensively address these queries, we utilised a high-quality label database we developed, alongside data extracted from Spotify Charts. Our study spanned from 2016 to 2023, focusing on ten countries.

12.1. Main findings and conclusions

This research has enabled us to appreciate some distinctive characteristics of the Latin American market. Among them, the dominance of major record labels stood out. Descriptive statistics were particularly useful in this regard. Contrary to what most reports mention, SME had a more significant dominance over ownership of streams compared to UMG. We attempted to come up with some historical accounts that could potentially explain SME's advantage in Latin America.

Our findings highlighted a decline in concentration levels on ownership of streams by major labels. This [pseudo] long tail effect may be related to the increasing widespread adoption of streaming platforms and musical content flooding by independent artists and generative artificial intelligence. Although not having access to information of fringe suppliers, we hypothesise that the subtle gains of the end portion of the distribution will continue to pale against the strengthening of the "head", once again supporting the "winner-take-all" character of the music industry. Moreover, some anecdotal accounts (i.e., the case of Bad Bunny and Anitta) suggest that individual superstar artists are still relevant and that their influence can be powerful in the charts, especially considering the top positions, effectively driving concentration to noticeable, although transitory peaks. Further inquiries are required regarding specific releases and the magnitude and duration of their impact over competition.

Continuing with our descriptive accounts, this time regarding musicological characteristics and diversity, we witnessed several regional trends for certain audio features. An overall increased interest for more danceable tracks over time, in conjunction with a gradual decrease of song duration, reveals a changing, dynamic digital music market, following the well-known phenomenon of fads and fashions within the cultural industries. While user preferences could account for decreasing duration of the tracks, suppliers may also have stronger incentives to "cut content" as the current model greatly favours number of plays, instead of "sales" in the traditional sense. On the other hand, diversity measures remained relatively stable over time and only a slight upward trend was evident when using the RSI. The trends were comparable across countries, with the exception of Brazil's higher overall scores in song diversity.

In broadening our analytical scope, we employed inferential statistics to better comprehend the relationship between major labels' presence in the charts and song diversity. We conducted two fixed-effects (FE) panel data regression models. They revealed the detrimental effects of SME and WMG on weekly diversity indices, whereas UMG appeared to contribute to a more diverse landscape. When examining our control variables, we noted the positive correlation between internet access and diversity. Internet access, in this regard, could be seen as a proxy of digitalisation. A more widespread access to digital services such as DSPs could effectively contribute to consumed diversity by eliminating the well-known constraints of shelf space.

12.2. Limitations

Paradoxically, despite access to high-quality data, our study primarily shed light on the "head" of the distribution. While our efforts surpassed merely examining "the tip of the iceberg" (Christianen, 1995, p. 56) as earlier studies did, the vast universe of music on platforms like Spotify still pose research challenges. As of now, Spotify boasts over a hundred million tracks, a number that's continually surging. Regrettably, our data didn't encompass detailed stream counts for newer songs that, despite not charting, might still have amassed significant listens. Additionally, we lacked insights into the extent of major record label ownership in the distribution's tail-end and the cumulative number of weekly streams across all Spotify releases, both chart-toppers and less popular tracks.

A common notion within the music industry is that only 10% of the time record labels can recoup their production and marketing costs whenever they finance and/or support a new artist (Krueger, 2019). To the best of our knowledge, we lack an organised database containing numeric data on the artists that conform the remaining —roughly— 90%. Furthermore, we hardly know whether the musicological characteristics of the unsuccessful major label products are similar to those of the superstar songs, which could allow for further inquiries on the explanatory variables for success in the music industry.

Even though we constructed a label database, on many occasions MusicBrainz, our main source for the categorisation, wasn't consistent: the amount of information for some labels was disproportionately larger than for others, even within the major realm, and we had to use other sources, like the labels' social media or Wikipedia pages, to classify them. Also, we were unable to find a separate source for measuring the majors' market shares in Latin America, despite having tried contacting the local offices in some of the countries. An expert source pointed out that market shares are very sensitive for the major labels and that they never cooperate with trade groups that want to publish those numbers. Therefore, third-party estimations are usually the best alternative, but were unavailable for our countries of interest.

While employing numeric indices from the Spotify API to profile a vast array of songs offers convenience, such an approach is not without its limitations. A prime concern is the ambiguity around Spotify's treatment of songs with fluctuating key or tempo. Even though a majority of popular genres might maintain stability in these aspects, chart songs with intricate musical arrangements might not receive an accurate profile compared to aesthetically less sophisticated tracks. Furthermore, the regions we examined —most notably Brazil— frequently exhibit a preference for local genres, often marked by intricate rhythmic patterns, diverse instruments, and nuanced performances which usually differ from Western pop. It is plausible to presume that the algorithms developed by the Echo Nest might not be ideally tailored to these unique musical styles. Evidence supporting this perspective can be found in studies that critically assess the accuracy of these algorithms. Notably, in a research effort evaluating the “danceability” metric against public opinions on the Brazilian rhythm *forró*, the authors highlighted the inadequacy of the Echo Nest audio feature in gauging a song’s aptitude for dancing (Vasconcelos et al., 2018).

12.3. Future avenues of research

Tschmuck (2012) states that innovation is highly dependent on context, and that, within the music industry, “we can observe a succession of waves that were caused by innovative thrusts” (p. 214). Nowadays, vertiginous changes are currently occurring regarding digital technologies that are starting to impact the production and commercialisation of cultural products. New digital assets emerge as potential sources of income for labels, artists, and musicians: NFTs, other blockchain-derived products, Metaverse-derived experiences, etcetera. At the same time, artificial intelligence is allowing creators to rapidly increase output, which is already posing a threat to the business models of the DSPs. While we don’t focus on “the end” of these innovative processes, we remain curious as to how our findings could change soon, following the ongoing trends.

The ambition of our study has been to explore the dynamics of the Latin American music industry as deployed in streaming services, but it barely paints the whole picture of music consumption in territories where other consumption means are more relevant, such as radio airplay. Further research could expand on our findings using more sophisticated statistical analysis, particularly incorporating machine learning methods that can deal with non-linearities and voluminous sources of data. The future of this subfield of study appears promising, as the music industry seems to be slowly moving towards an era where numerical data is used more consistently even by non-major players, mainly due to increased accessibility and a higher supply of data compilers and providers.

13.References

Abbing, H. (2022). *The Economies of Serious and Popular Art: How They Diverged and Reunited*. Springer Nature. https://doi.org/10.1007/978-3-031-18648-6

Adell, J. (1998). *La música en la era digital: La cultura de masas como simulacro*. Editorial Milenio.

Adelman, M. A. (1959). Differential rates and changes in concentration. *The Review of Economics and Statistics*, *41*(1), 68. https://doi.org/10.2307/1925461

Adler, M. (1985). Stardom and Talent. *The American Economic Review*, *75*(1), 208–2012. https://www.jstor.org/stable/1812714

Adler, M. (2006). Stardom and Talent. In V. A. Ginsburgh & D. Throsby (Eds.), *A Handbook of the Economics of Art and Culture* (pp. 895–906). Elsevier. https://doi.org/10.1016/S1574-0676(06)01025-8

Adnews. (2017, December 18). Spotify faz campanha de lançamento da nova música de Anitta. *Adnews*. Retrieved April 1, 2023, from https://adnews.com.br/spotify-faz-campanha-de-lancamento-da-nova-musica-de-anitta/

Aguiar, L., & Waldfogel, J. (2021). Platforms, Power, and Promotion: Evidence from Spotify Playlists. *Journal of Industrial Economics*, *69*(3), 653–691. https://doi.org/10.1111/joie.12263

Akay, Ö., & Yüksel, G. (2017). Clustering the mixed panel dataset using Gower's distance and k-prototypes algorithms. *Communications in Statistics - Simulation and Computation*, *47*(10), 3031–3041. https://doi.org/10.1080/03610918.2017.1367806

Anderson, C. (2004). The Long Tail. *Wired*. https://www.wired.com/2004/10/tail/

Anderson, C. (2008). *The Long Tail: Why the Future of Business is Selling Less of More* (2nd ed.). Hyperion.

Baskerville, D., & Baskerville, T. (2019). *Music Business Handbook and Career Guide* (12th ed.). SAGE Publications, Incorporated.

Belinfante, A., & Johnson, R. L. (1982). Competition, pricing and concentration in the U.S. recorded music industry. *Journal of Cultural Economics*, *6*(2), 11–24. https://doi.org/10.1007/bf02511597

Bernstein, A., Sekine, N., & Weissman, D. (2007). *The Global Music Industry: Three Perspectives*. Routledge.

Billboard. (2023). Year-End Charts.: Top Latin Labels. *Billboard*. Retrieved November 26, 2023, from https://www.billboard.com/charts/year-end/top-latin-labels/

Bishop, J. (2005). Building International Empires of Sound: Concentrations of Power and Property in the “Global” Music Market. *Popular Music and Society*, *28*(4), 443–471. https://doi.org/10.1080/03007760500158957

Burgess, R. J. (2014). *The History of Music Production*. Oxford University Press.

Burnett, R. (1990). *Concentration and Diversity in the International Phonogram Industry* [PhD dissertation]. University of Gothenburg.

Burnett, R. (1992). The Implications of Ownership Changes on Concentration and Diversity in the Phonogram Industry. *Communication Research*, *19*(6), 749–769. https://doi.org/10.1177/009365092019006005

Cameron, S. (2016). Past, present and future: music economics at the crossroads. *Journal of Cultural Economics*, *40*(1), 1–12. https://doi.org/10.1007/s10824-015-9263-4

Casado, R. (2023, November 24). Sony revoluciona la música en España al adquirir Altafonte, distribuidora de Julieta Venegas y Mónica Naranjo. *Expansión*. Retrieved November 26, 2023, from https://www.expansion.com/empresas/distribucion/2023/11/24/655fae53e5fdeace058b461c.html

Casey, M. A., Veltkamp, R. C., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-Based Music Information Retrieval: current directions and future challenges. *Proceedings of the IEEE*, *96*(4), 668–696. https://doi.org/10.1109/jproc.2008.916370

Caves, R. E. (2000). *Creative Industries: Contracts between Art and Commerce*. Harvard University Press.

Cho, D., Lee, S. H., Yoo, Y., & Chu, H. Y. (2019). Television singing competitions create stars? Empirical evidence from the digital music chart in South Korea. *Journal of Cultural Economics*, *43*(1), 1–20. https://doi.org/10.1007/s10824-018-9327-3

Christianen, M. (1995). Cycles in symbol production? A new model to explain concentration, diversity and innovation in the music industry. *Popular Music*, *14*(1), 55–93. https://doi.org/10.1017/s0261143000007637

Christman, E. (2023, April 21). Sony Music in Advanced Talks to Acquire Stake in Bad Bunny Label and Management Firm Rimas. *Billboard*. Retrieved November 26, 2023, from https://www.billboard.com/pro/bad-bunny-label-management-rimas-sell-sony/

Chun, Z., Li, H., Yi, W. J., & Liu, W. (2021). The development trend of musicians' influence and music genres of big data. *E3S Web of Conferences*, *253*, 02085. https://doi.org/10.1051/e3sconf/202125302085

Collins, S., & O'Grady, P. (2016). Off the Charts: The Implications of Incorporating Streaming Data into the Charts. In R. Nowak & A. Whelan (Eds.), *Networked Music Cultures: Contemporary Approaches, Emerging Issues* (pp. 151–169). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-58290-4

Connolly, M., & Krueger, A. (2005). Rockonomics: The Economics of Popular Music. *National Bureau of Economic Research Working Paper Series*. https://doi.org/10.3386/w11282

Cookson, R. (2014, March 7). Spotify's purchase of The Echo Nest prompts talk of potential IPO. *Financial Times*, 18.

Cox, R. a. K., Felton, J. M., & Chung, K. H. (1995). The concentration of commercial success in popular music: An analysis of the distribution of gold records. *Journal of Cultural Economics*, *19*(4), 333–340. https://doi.org/10.1007/bf01073995

Darbilly, L. V. C., & Vieira, M. M. F. (2012). Technological evolution and the music industry in brazil: current situation and future prospects in a changing field. *Creative Industries Journal*. https://doi.org/10.1386/cij.5.1-2.69_1

Day, B. R. (2011). In Defense of Copyright: Creativity, Record Labels, and the Future of Music. *Seton Hall Journal of Sports and Entertainment Law*, *21*(1), 61–103. https://doi.org/10.2139/ssrn.1609689

De Almeida, M. (2002, July 11). *Como o homem registrou o som*. Forró Em Vinil. Retrieved November 26, 2023, from https://www.forroemvinil.com/textos/texto-linha-do-tempo-da-evolucao-musical/

De Maio, F. (2007). Income inequality measures. *Journal of Epidemiology and Community Health*, *61*(10), 849–852. https://doi.org/10.1136/jech.2006.052969

Dredge, S. (2020, April 24). *Music data firm BuzzAngle relaunches as Alpha Data*. Music Ally. Retrieved November 25, 2021, from https://musically.com/2020/04/24/music-data-firm-buzzangle-relaunches-as-alpha-data/

Dredge, S. (2021a, November 10). *Latin tracks are streamed over 20bn times a month on Spotify*. Music Ally. Retrieved April 5, 2023, from https://musically.com/2021/11/10/latin-tracks-are-streamed-over-20bn-times-a-month-on-spotify/

Dredge, S. (2021b, December 14). *Spotify notched up 107.9bn hours of music listening in 2021*. Music Ally. Retrieved April 6, 2023, from https://musically.com/2021/12/14/spotify-107-9bn-hours-music-listening-2021/

Dredge, S. (2023). Spotify’s new vertical discovery feed is more than just a TikTok clone. *Music Ally*. https://musically.com/2023/03/07/spotifys-new-vertical-discovery-feed-is-more-than-just-a-tiktok-clone/

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal*, *3*(2), 168–177. https://doi.org/10.1177/1536867x0300300206

Elberse, A. (2008). *Should You Invest in the Long Tail?* Harvard Business Review. https://hbr.org/2008/07/should-you-invest-in-the-long-tail

Elberse, A., & Oberholzer-Gee, F. (2007). Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales. *MSI Reports: Working Paper Series*, *4*, 49–72. https://www.hbs.edu/faculty/Pages/item.aspx?num=39668

Fernández-Blanco, V., Ginsburgh, V., Prieto-Rodriguez, J., & Weyers, S. (2014). As Good as it Gets?: Blockbusters and the Inequality of Box Office Results Since 1950. In J. Kaufman & D. K. Simonton (Eds.), *The Social Science of Cinema* (pp. 269–285). Oxford. https://doi.org/10.1093/acprof:oso/9780199797813.003.0012

Fisher, M. (1980, April 30). Gas Topper Cites Royalty Woes: Acosta: Collections Must Improve in Latin America. *Billboard*, 80.

Flores, G. (2020, March 9). Behind Bad Bunny's History-Making 'YHLQMDLG': 'We Didn't Expect This Big of an Acceptance.' *Billboard*. Retrieved April 1, 2023, from https://www.billboard.com/pro/bad-bunny-yhlqmdlg-streaming-manager-noah-assad/

Fox, M. A., & Kochanowski, P. (2004). Models of Superstardom: An Application of the Lotka and Yule Distributions. *Popular Music and Society*, *27*(4), 507–522. https://doi.org/10.1080/0300776042000264694

Frank, R. T., & Cook, P. J. (1996). *The Winner-Take-All Society: Why the Few at the Top Get So Much More Than the Rest of Us*. Penguin Books.

Gallego, J. I. (2016). Una mirada a la diversidad en las industrias radiofónica y musical en España. *Cuadernos De Información Y Comunicación*, *21*, 139–155. https://doi.org/10.5209/ciyc.52879

Haampland, O. (2017). Power Laws and market shares: Cumulative advantage and the Billboard Hot 100. *Journal of New Music Research*, *46*(4), 356–380. https://doi.org/10.1080/09298215.2017.1358285

Hagen, A. N. (2022). Datafication, literacy, and democratization in the music industry. *Popular Music and Society*, *45*(2), 184–201. https://doi.org/10.1080/03007766.2021.1989558

Hamlen, W. A. (1991). Superstardom in Popular Music: Empirical Evidence. *The Review of Economics and Statistics*, *73*(4), 729–733. https://doi.org/10.2307/2109415

Hamlen, W. A. (1994). Variety and Superstardom in Popular Music. *Economic Inquiry*, *32*(3), 395–406. https://doi.org/10.1111/j.1465-7295.1994.tb01338.x

Handke, C. (2006). Plain destruction or creative destruction? Copyright erosion and the evolution of the record industry. *Review of Economic Research on Copyright Issues*, *3*(2), 29–51. https://ssrn.com/abstract=1144318

Handke, C. (2010). *The Creative Destruction of Copyright: Innovation in the Record Industry and Digital Copying* [Doctoral thesis]. Erasmus University Rotterdam.

Hofmann, K. H., & Opitz, C. (2018). Talent and publicity as determinants of superstar incomes: empirical evidence from the motion picture industry. *Applied Economics*, *51*(13), 1383–1395. https://doi.org/10.1080/00036846.2018.1527452

Howard-Spink, S. (2012). Brazil. In L. Marshall (Ed.), *The International Recording Industries* (pp. 77–94). Taylor & Francis Group.

Hsu, J., & Huang, C. (2015). Designing a graph-based framework to support a multi-modal approach for music information retrieval. *Multimedia Tools and Applications*, *74*(15), 5401–5427. https://doi.org/10.1007/s11042-014-1860-2

Hull, G. P. (2004). *The recording industry* (2nd ed.). Rotledge.

Hull, G. P., Hutchison, T., & Strasser, R. (2011). *The Music Business and Recording Industry*. Routledge.

Im, H., Song, H., & Jung, J. (2019). The effect of streaming services on the concentration of digital music consumption. *Information Technology & People*, *33*(1), 160–179. https://doi.org/10.1108/itp-12-2017-0420

Ingham, T. (2015, March 18). Sony fully acquires The Orchard in $200m deal. *Music Business Worldwide*. https://www.musicbusinessworldwide.com/sony-fully-acquires-the-orchard-in-200m-deal/

Ingham, T. (2019, April 29). Nearly 40,000 tracks are now being added to Spotify every single day. *Music Business Worldwide*. https://www.musicbusinessworldwide.com/nearly-40000-tracks-are-now-being-added-to-spotify-every-single-day/

Ingham, T. (2021, March 1). Over 60,000 tracks are now uploaded to Spotify every day. That’s nearly one per second. *Music Business Worldwide*. https://www.musicbusinessworldwide.com/over-60000-tracks-are-now-uploaded-to-spotify-daily-thats-nearly-one-per-second/

Ingham, T. (2023, October 25). Spotify is changing its royalty model to crush streaming fraud and introduce a minimum payment threshold. Its. *Music Business Worldwide*. https://www.musicbusinessworldwide.com/spotify-is-changing-its-royalty-model-to-crush-streaming-fraud/

International Federation of the Phonographic Industry [IFPI]. (2000). *IFPI Music Piracy Report 2000*.

International Federation of the Phonographic Industry [IFPI]. (2001). *IFPI Music Piracy Report 2001*.

International Federation of the Phonographic Industry [IFPI]. (2002). *IFPI Music Piracy Report 2002*.

International Federation of the Phonographic Industry [IFPI]. (2003). *The recording industry world sales 2003*.

International Federation of the Phonographic Industry [IFPI]. (2004). *The recording industry world sales 2004*.

International Federation of the Phonographic Industry [IFPI]. (2013). *IFPI Digital Music Report 2013: Engine of a digital world*.

International Federation of the Phonographic Industry [IFPI]. (2014). *IFPI Digital Music Report 2014: Lighting up new markets*.

International Federation of the Phonographic Industry [IFPI]. (2015). *IFPI Digital Music Report 2015: Charting the path to sustainable growth*.

International Federation of the Phonographic Industry [IFPI]. (2016). *Global Music Report: Music consumption exploding worldwide*.

International Federation of the Phonographic Industry [IFPI]. (2017). *Global Music Report 2017: Annual state of the industry*.

International Federation of the Phonographic Industry [IFPI]. (2018). *Global Music Report 2018: Annual state of the industry*.

International Federation of the Phonographic Industry [IFPI]. (2019). *Global Music Report 2019: State of the industry*.

International Federation of the Phonographic Industry [IFPI]. (2020). *Global Music Report: The industry in 2019*.

International Federation of the Phonographic Industry [IFPI]. (2021). *Global Music Report 2021: State of the industry*.

International Federation of the Phonographic Industry [IFPI]. (2022). *Global Music Report 2022: State of the industry*.

International Federation of the Phonographic Industry [IFPI]. (2023). *Global Music Report 2023: State of the industry*.

Jehan, T. (2005). *Creating Music by Listening* [PhD dissertation]. Massachusetts Institute of Technology.

Kaimann, D., Tanneberg, I., & Cox, J. (2020). "I will survive": Online streaming and the chart survival of music tracks. *Managerial and Decision Economics*, *42*(1), 3–20. https://doi.org/10.1002/mde.3226

Krueger, A. B. (2019). *Rockonomics: A Backstage Tour of What the Music Industry Can Teach Us about Economics and Life*. Currency.

Lannert, J. (1998, December 19). P'Gram Execs Take Lead In Latin America: Reflects Firm's Dominance In Region; Manolo Díaz Still In Charge. *Billboard*, 78.

Lannert, J. (1999, November 20). Frank Welzer: The Billboard Interview. The Head Of Sony Music Latin America Saw Sony Discos Through Its Most Formative Years. *Billboard*, 58–64.

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics*, *13*(1), 255–269. https://doi.org/10.1016/j.joi.2018.12.006

Lopes, P. (1992). Innovation and Diversity in the Popular Music Industry, 1969 to 1990. *American Sociological Review*, *57*(1), 56–71. https://doi.org/10.2307/2096144

Magurran, A. E. (2004). *Measuring Biological Diversity*. Blackwell Science.

Mall, A. (2018). Concentration, diversity, and consequences: Privileging independent over major record labels. *Popular Music*, *37*(3), 444–465. https://doi.org/10.1017/s0261143018000375

McKenzie, J., Crosby, P., & Lenten, L. J. A. (2020). It takes two, baby! Feature artist collaborations and streaming demand for music. *Journal of Cultural Economics*, *45*(3), 385–408. https://doi.org/10.1007/s10824-020-09396-y

Meyn, J., Kandziora, M., Albers, S., & Clement, M. (2023). Consequences of platforms' remuneration models for digital content: initial evidence and a research agenda for streaming services. *Journal of the Academy of Marketing Science*, *51*(1), 114–131. https://doi.org/10.1007/s11747-022-00875-6

Moreno, T. (1979, December 1). CBS Records Sets Plans For Latin Product Distribution. *Billboard*, 14.

Mulligan, M. (2023, October 24). *Two-tier licensing is about to become a reality*. https://www.midiaresearch.com/blog/two-tier-licensing-is-about-to-become-a-reality

Murthy, Y. V. S., & Koolagudi, S. G. (2018). Content-Based Music Information Retrieval (CB-MIR) and Its Applications toward the Music Industry. *ACM Computing Surveys*, *51*(3), 1–46. https://doi.org/10.1145/3177849

Music & Copyright. (2021a, April 21). *UMG and SME put the market share squeeze on WMG and the independent sector*. Music & Copyright's Blog. Retrieved December 9, 2021, from https://musicandcopyright.wordpress.com/2021/04/21/umg-and-sme-put-the-market-share-squeeze-on-wmg-and-the-independent-sector/

Music & Copyright. (2021b). Music & Copyright Newsletter 29 June 2021. In *Music & Copyright* (No. 668). Omdia. https://musicandcopyright.wordpress.com/

Music & Copyright. (2023, July 6). *Digital and physical revenue market share of the largest record companies worldwide from 2012 to 2022*. Statista. https://statista.upc.elogim.com/statistics/422926/record-companies-market-share-worldwide-physical-digital-revenues/

Music Business Worldwide [MBW]. (2022, June 6). The major labels face an uphill battle for streaming market share. Here's how they might fight back. *Music Business Worldwide*. Retrieved April 1, 2023, from https://musicbusinessworldwide.com/podcast/the-major-labels-face-an-uphill-battle-for-streaming-market-share-heres-how-they-might-fight-back/

Negus, K. (1992). *Producing Pop: Culture and Conflict in the Popular Music Industry*. Hodder Arnold.

Negus, K. (1996). *Popular Music in Theory: An introduction*. Wesleyan University Press.

Nieto, O., Mysore, G. J., Wang, C., Smith, J. B. L., Schlüter, J., Grill, T., & McFee, B. (2020). Audio-Based Music Structure Analysis: Current trends, open challenges, and applications. *Transactions of the International Society for Music Information Retrieval*, *3*(1), 246–263. https://doi.org/10.5334/tismir.54

Observatorio Latinoamericano de Música Independiente [OLMI]. (n.d.). *Brasil*. OLMI. Retrieved June 7, 2023, from https://olmi.la/brasil/

Opinion Box. (2022). *Consumo de música no Brasil: Dados sobre o consumo e hábitos dos fãs de música no Brasil*. Tumpats. https://tumpats.com.br/consumo-de-musica-no-brasil/

Organisation for Economic Co-operation and Development [OECD]. (2021). *Methodologies to Measure Market Competition, OECD Competition Committee Issues Paper*. https://www.oecd.org/daf/competition/methodologies-to-measure-market-competition.htm

Parekh, B. (2000). *Rethinking Multiculturalism: Cultural Diversity and Political Theory*. Macmillan Press.

Peterson, R. E., & Berger, D. H. (1975). Cycles in Symbol Production: The Case of Popular Music. *American Sociological Review*, *40*(2), 158. https://doi.org/10.2307/2094343

Peterson, R. E., & Berger, D. H. (1996). Measuring Industry Concentration, Diversity, and Innovation in Popular Music. *American Sociological Review*, *61*(1), 175. https://doi.org/10.2307/2096413

Pyun, M., Kim, D., Lim, C., Lee, E., Kwon, J., & Lee, S. Y. (2020). Examining the relationship between songs and psychological characteristics. In *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-030-60128-7_8

Ranaivoson, H. R. (2010). The determinants of the diversity of cultural expressions. An international quantitative analysis of diversity of production in the recording industry. *Observatorio (OBS*)*, *4*(4), 215–249. https://doi.org/10.15847/obsobs442010434

Rhoades, S. A. (1995). Market share inequality, the HHI, and other measures of the firm-composition of a market. *Review of Industrial Organization*, *10*(6), 657–674. https://doi.org/10.1007/bf01024300

Rogers, J. (2017). Deconstructing the Music Industry Ecosystem. In S. Sparviero, C. Peil, & G. Balbi (Eds.), *Media Convergence and Deconvergence* (pp. 217–239). Springer Publishing. https://doi.org/10.1007/978-3-319-51289-1_11

Rosen, S. (1981). The Economics of Superstars. *The American Economic Review*, *71*(5), 845–858. https://www.jstor.org/stable/1803469

Rothenbuhler, E. W., & Dimmick, J. W. (1982). Popular Music: Concentration and Diversity in the Industry, 1974–1980. *Journal of Communication*, *32*(1), 143–149. https://doi.org/10.1111/j.1460-2466.1982.tb00485.x

RouteNote. (2022, November 1). *Share of music streaming subscribers worldwide in the 2nd quarter of 2022, by company*. Statista. Retrieved April 13, 2023, from https://statista.upc.elogim.com/statistics/653926/music-streaming-service-subscriber-share/

Ruza, J. H. (2023a, May 26). *Aumenta a 120.000 el número de canciones que se suben a streaming diariamente - Industria Musical*. Industria Musical. https://industriamusical.com/aumenta-a-120-000-el-numero-de-canciones-que-se-suben-a-streaming-diariamente/

Ruza, J. H. (2023b, June 9). *Deezer identificará y etiquetará música generada por inteligencia artificial en su servicio - Industria Musical*. Industria Musical. https://industriamusical.com/deezer-identificara-y-etiquetara-musica-generada-por-inteligencia-artificial-en-su-servicio/

Santos, V., Monsueto, S. E., & Pereira Da Silva, B. C. (2019). Análise dos determinantes salariais dos profissionais brasileiros produtores de conteúdo musical: uma visão pela abordagem da cauda longa. *Nova Economia*, *29*(1), 103–134. https://doi.org/10.1590/0103-6351/3039

Savelkoul, R. (2020). Superstars vs the long tail: How does music piracy affect digital song sales for different segments of the industry? *Information Economics and Policy*, *50*, 1–17. https://doi.org/10.1016/j.infoecopol.2019.100847

Sciandra, M., & Spera, I. C. (2022). A model-based approach to Spotify data analysis: a Beta GLMM. *Journal of Applied Statistics*, *49*(1), 214–229. https://doi.org/10.1080/02664763.2020.1803810

Shepherd, D., Hautus, M. J., Giang, E., & Landon, J. (2022). "The most relaxing song in the world"? A comparative study. *Psychology of Music*, *51*(1), 3–15. https://doi.org/10.1177/03057356221081169

Sintonia com a Sociedade. (2021, October 21). O consumo de música no Brasil. *Globo*. https://gente.globo.com/o-consumo-de-musica-no-brasil/

Smirke, R. (2014, March 18). IFPI Music Report 2014: Global Recorded Music Revenues Fall 4%, Streaming and Subs Hit $1 Billion. *Billboard*. Retrieved April 18, 2023, from https://www.billboard.com/music/music-news/ifpi-music-report-2014-global-recorded-music-revenues-fall-4-5937645/

Sony Music Entertainment [SME]. (2022a, March 4). *Sony Music Entertainment Completes Its Acquisition of Som Livre*. Sony Music Entertainment. Retrieved April 1, 2023, from https://www.sonymusic.com/sonymusic/som-livre-acquisition-closes/

Sony Music Entertainment [SME]. (2022b, March 16). *Sony Music Entertainment acquisition of AWAL and neighbouring rights cleared in CMA Final report - Sony Music*. Sony Music Entertainment. Retrieved November 26, 2023, from https://www.sonymusic.com/sonymusic/cma-clears-awal-neighbouring-rights-acquisition/

Spotify. (n.d.-a). *Get Track's Audio Features: Get audio feature information for a single track identified by its unique Spotify ID*. Spotify for Developers. https://developer.spotify.com/documentation/web-api/reference/get-audio-features

Spotify. (n.d.-b). *How we calculate charts*. Spotify for Artists. Retrieved November 25, 2021, from https://artists.spotify.com/help/article/how-we-calculate-charts

Spotify. (n.d.-c). *How we count streams*. Spotify for Artists. Retrieved April 6, 2023, from https://artists.spotify.com/help/article/how-we-count-streams

Spotify. (n.d.-d). *Royalties*. Spotify for Artists. Retrieved April 6, 2023, from https://artists.spotify.com/help/article/royalties

Spotify. (n.d.-e). *Web API*. Spotify for Developers. Retrieved September 9, 2023, from https://developer.spotify.com/documentation/web-api

Spotify. (2023, January 31). *Spotify Reports Fourth Quarter 2022 Earnings*. Retrieved April 5, 2023, from https://newsroom.spotify.com/2023-01-31/spotify-reports-fourth-quarter-2022-earnings/

Statista. (2021). Music in Latin America. In *Statista*.

Stirling, A. (1998). On the Economics and Analysis of Diversity. *SPRU Electronic Working Papers Series*, *28*.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, *4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213

Tagg, P. (1982). Analysing popular music: theory, method and practice. *Popular Music*, *2*, 37–67. https://doi.org/10.1017/s0261143000001227

Torres, L. (2017, December 21). Spotify põe ônibus personalizado na rua para promover "Vai Malandra" da Anitta. *POPline*. Retrieved April 1, 2023, from https://portalpopline.com.br/anitta-poe-onibus-personalizado-na-rua-para-promover-vai-malandra/

Tschmuck, P. (2012). *Creativity and Innovation in the Music Industry* (2nd ed.). Springer. https://doi.org/10.1007/978-3-642-28430-4

Tschmuck, P. (2021). *The Economics of Music* (2nd ed.). Agenda Publishing.

UOL. (2017, December 20). Nova música de Anitta faz história e entra no Top 20 do Spotify. *UOL Entretenimento*. Retrieved April 1, 2023, from https://musica.uol.com.br/noticias/redacao/2017/12/20/nova-musica-de-anitta-faz-historia-e-entra-no-top-20-do-spotify.htm

Van Eijck, K. (2001). Social differentiation in musical taste patterns. *Social Forces*, *79*(3), 1163–1185. https://doi.org/10.1353/sof.2001.0017

Vasconcelos, L. L. P. M., Cunha, C. C., Ferreira, G. M. D., Rodrigues, L., & Souza, J. T. (2018). Avaliação da eficácia do Danceability do Echo Nest aplicado a músicas de forró. *Revista Principia*, *1*(38), 138–147. https://doi.org/10.18265/1517-03062015v1n38p138-147

Waldfogel, J. (2018). *Digital Renaissance: What Data and Economics Tell Us about the Future of Popular Culture*. Princeton University Press.

Waldfogel, J. (2020). The Welfare Effects of Spotify's Cross-Country Price Discrimination. *Review of Industrial Organization*, *56*(4), 593–613. https://doi.org/10.1007/s11151-020-09748-0

Wikstrom, P. (2009). *The Music Industry: Music in the Cloud*. Wiley.

Wooldridge, J. M. (2001). *Econometric analysis of cross section and panel data*. The MIT Press.

Woolhouse, M., & Bansal, J. (2013). Work, rest and (press) play: Music consumption as an indicator of human economic development. *Journal of Interdisciplinary Music Studies*, *7*(1–2), 45–71. https://doi.org/10.4407/jims.2015.05.003

World Data. (n.d.). *Portuguese speaking countries*. Retrieved June 14, 2023, from https://www.worlddata.info/languages/portuguese.php

Zheng, E., Moh, M., & Moh, T. (2017). Music Genre Classification: A N-Gram Based Musicological Approach. *IEEE 7th International Advance Computing Conference*. https://doi.org/10.1109/IACC.2017.132

14. Appendices

14.1. Appendix A – Raw and aggregate data on streams: tables and graphs

Table A1. Streams – Descriptive statistics

N	Valid	633911
	Missing	89
Mean		348294.87
Median		115269.00
Std. Deviation		607317.026
Skewness		4.730
Std. Error of Skewness		.003
Kurtosis		39.901
Std. Error of Kurtosis		.006
Minimum		5127 [a]
Maximum		19816644[b]

[a] The minimum stream number was found in Guatemala, on week 1. The song was “Tu No Vive Asi (feat. Mambo Kingz & DJ Luian)” by Arcangel, Bad Bunny, Mambo Kingz and DJ Luian.
[b] This maximum stream number occurred in Mexico, on week 317. The song was Bizarrap and Shakira’s “Shakira: Bzrp Music Sessions, Vol. 53”

Table A2. Streams - Descriptive statistics - By country

			Argentina	Brazil	Chile	Colombia	Costa Rica	Ecuador	Guatemala	Mexico	Peru	Uruguay
streams	N	Valid	63,393	63,385	63,393	63,393	63,393	63,393	63,393	63,382	63,393	63,393
		Missing	7	15	7	7	7	7	7	18	7	7
	Mean		415,240.09	1,163,563.58	370,247.41	145,822.01	45,370.42	62,412.43	50,032.96	1,031,480.70	170,152.92	28,847.63
	Median		268,247.00	852,351.00	243,305.00	96,561.00	31,688.00	43,449.00	36,243.00	789,102.50	112,676.00	18,709.00
	Std. Deviation		426,214.76	952,955.96	376,782.71	156,950.72	39,955.83	65,837.17	53,844.36	906,080.17	167,540.26	28,575.18
	Skewness		3.49	2.79	3.59	4.17	3.84	4.36	4.21	4.09	3.40	3.48
	Std. Error of Skewness		.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
	Kurtosis		19.66	11.37	22.11	28.73	24.22	31.80	30.74	30.17	18.34	20.97
	Std. Error of Kurtosis		.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	Minimum		78,897	199,060	55,783	23,265	11,376	9,133	5,127	201,084	23,380	5,675
	Maximum		7,669,613	13,937,355	6,976,383	2,644,828	665,302	1,180,668	932,591	19,816,644	2,837,236	596,879

Table A3. Descriptive statistics – Sum of streams per week

	sum_st	sum_st_majors	sum_st_universal	sum_st_sony	sum_st_warner	sum_st_indies	sum_st_artist_label
Mean	69,649,195.64	37,223,795.59	13,503,475.43	15,523,888.78	8,196,431.38	32,425,400.05	3,703,472.24
Median	30,294,486.50	18,219,763.50	6,234,202.50	7,577,746.50	3,750,566.00	12,694,342.50	618,611.50
Std. Deviation	87,067,125.57	44,794,021.26	17,057,266.89	21,134,157.02	9,825,838.49	45,292,003.21	7,423,134.40
Skewness	1.730	1.578	1.551	2.835	1.595	2.283	3.035
Std. Error of Skewness	.003	.003	.003	.003	.003	.003	.003
Kurtosis	2.335	1.638	1.065	10.837	1.955	5.719	9.603
Std. Error of Kurtosis	.006	.006	.006	.006	.006	.006	.006
Minimum	2,501,795	2,022,968	528,448	732,525	319,729	477,875	0
Maximum	451,313,134	244,249,188	85,703,392	170,368,951	54,796,027	304,285,849	50,360,988

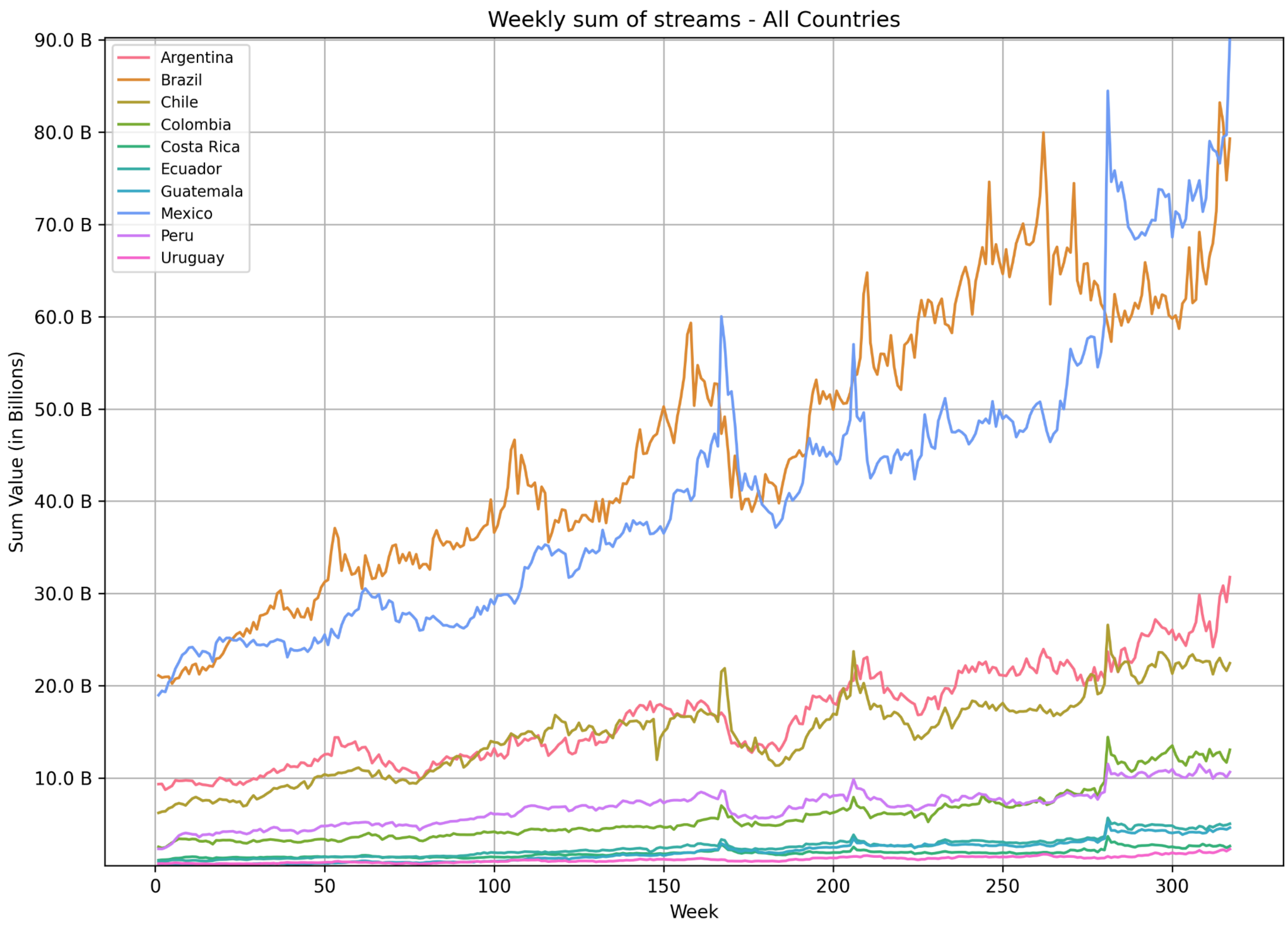
Table A4. Descriptive statistics – Sum of streams per week – By country

		Argentina	Brazil	Chile	Colombia	Costa Rica	Ecuador	Guatemala	Mexico	Peru	Uruguay
sum_st	Mean	83,038,848.75	232,657,657.06	74,041,306.67	29,161,182.89	9,073,082.09	12,481,107.85	10,005,486.33	206,237,569.66	34,026,825.81	5,768,889.27
	Median	80,640,426.00	224,579,811.00	76,462,573.00	24,107,312.00	9,042,957.00	11,761,669.00	9,372,212.00	198,262,898.00	34,659,223.00	5,505,830.00

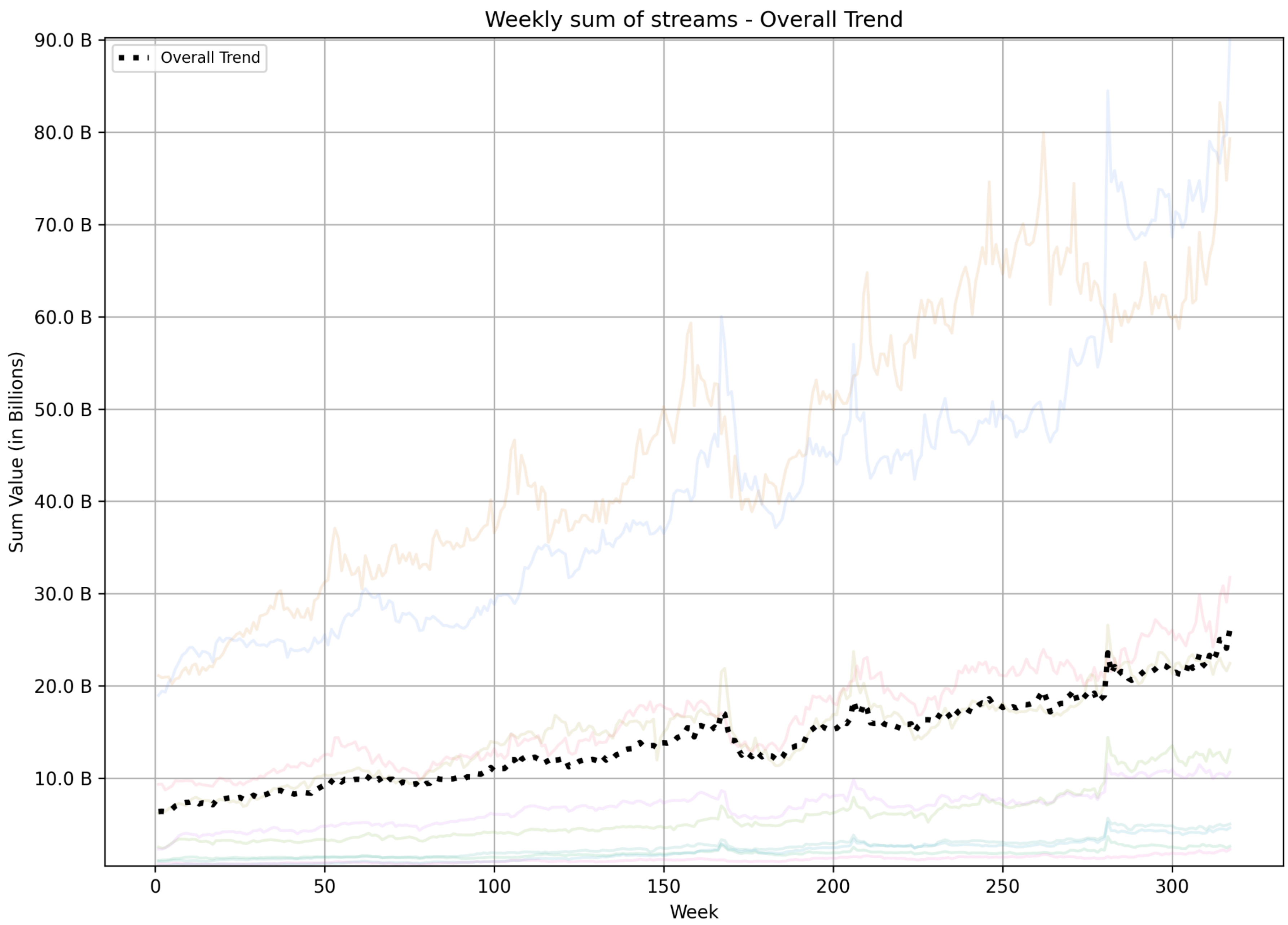
	Std. Deviation	25,987,729.87	74,203,447.17	22,548,412.09	13,910,364.28	2,104,838.27	5,447,283.15	5,746,171.39	76,463,992.87	9,791,959.75	1,751,719.97
	Skewness	.468	.119	.042	1.278	.977	.788	.587	.880	.236	.675
	Std. Error of Skewness	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
	Kurtosis	-.669	-1.011	-.718	.778	.961	.005	-.630	.158	-.389	-.080
	Std. Error of Kurtosis	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
	Minimum	43,605,101	101,132,988	31,050,512	12,023,153	5,410,986	4,816,977	2,501,795	94,856,088	11,499,984	3,207,660
	Maximum	158,817,160	416,031,973	132,938,016	72,156,124	18,352,314	28,345,064	26,100,627	451,313,134	57,567,014	11,434,461
sum_st_majors	Mean	42,684,360.92	122,974,223.38	34,490,450.52	17,435,032.55	5,386,548.21	7,162,672.98	5,024,094.61	114,000,307.40	20,015,286.62	3,064,978.75
	Median	38,453,571.00	121,098,709.00	34,994,520.00	15,030,690.00	5,429,756.00	7,174,458.00	4,876,769.00	114,916,098.00	20,287,652.00	2,803,422.00
	Std. Deviation	10,140,244.87	38,708,842.28	5,499,247.75	6,156,120.12	551,962.56	1,863,338.47	1,939,977.98	16,771,588.75	3,559,301.25	704,859.89
	Skewness	.561	.525	-.086	1.446	-.144	.475	.200	.354	-.159	.579
	Std. Error of Skewness	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
	Kurtosis	-1.101	-.350	-.177	1.490	1.029	-.104	-1.237	.241	-.074	-.876
	Std. Error of Kurtosis	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
	Minimum	27,564,447	59,464,863	21,293,163	9,434,119	3,821,709	3,848,433	2,023,920	72,061,685	8,868,357	2,022,968
	Maximum	64,735,576	244,249,188	53,396,992	41,879,841	7,622,381	12,529,043	8,954,852	176,082,790	30,227,188	4,949,380
sum_st_universal	Mean	10,536,468.31	45,581,384.06	11,278,002.07	7,513,833.81	2,153,537.75	2,769,636.65	2,034,348.87	45,559,912.94	6,775,869.31	831,760.57
	Median	10,613,355.00	45,864,648.00	10,792,971.00	5,995,693.00	2,107,838.00	2,598,724.00	1,858,368.00	44,682,347.00	6,367,594.00	806,758.00
	Std. Deviation	2,153,181.65	8,724,174.81	2,675,636.64	4,315,146.68	446,691.46	1,142,450.97	993,468.28	10,437,499.09	1,967,167.73	152,647.82
	Skewness	.408	-.127	.548	1.999	.802	1.765	.622	.448	1.285	.547
	Std. Error of Skewness	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
	Kurtosis	-.213	-.006	.189	3.844	.592	3.585	-.587	-.413	1.750	-.351
	Std. Error of Kurtosis	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
	Minimum	6,380,006	19,385,723	5,612,861	3,286,263	1,375,559	1,338,961	752,008	29,229,551	3,241,129	528,448
	Maximum	18,272,854	72,707,823	23,010,242	27,668,068	3,767,618	7,503,136	4,567,548	85,703,392	14,368,321	1,330,378

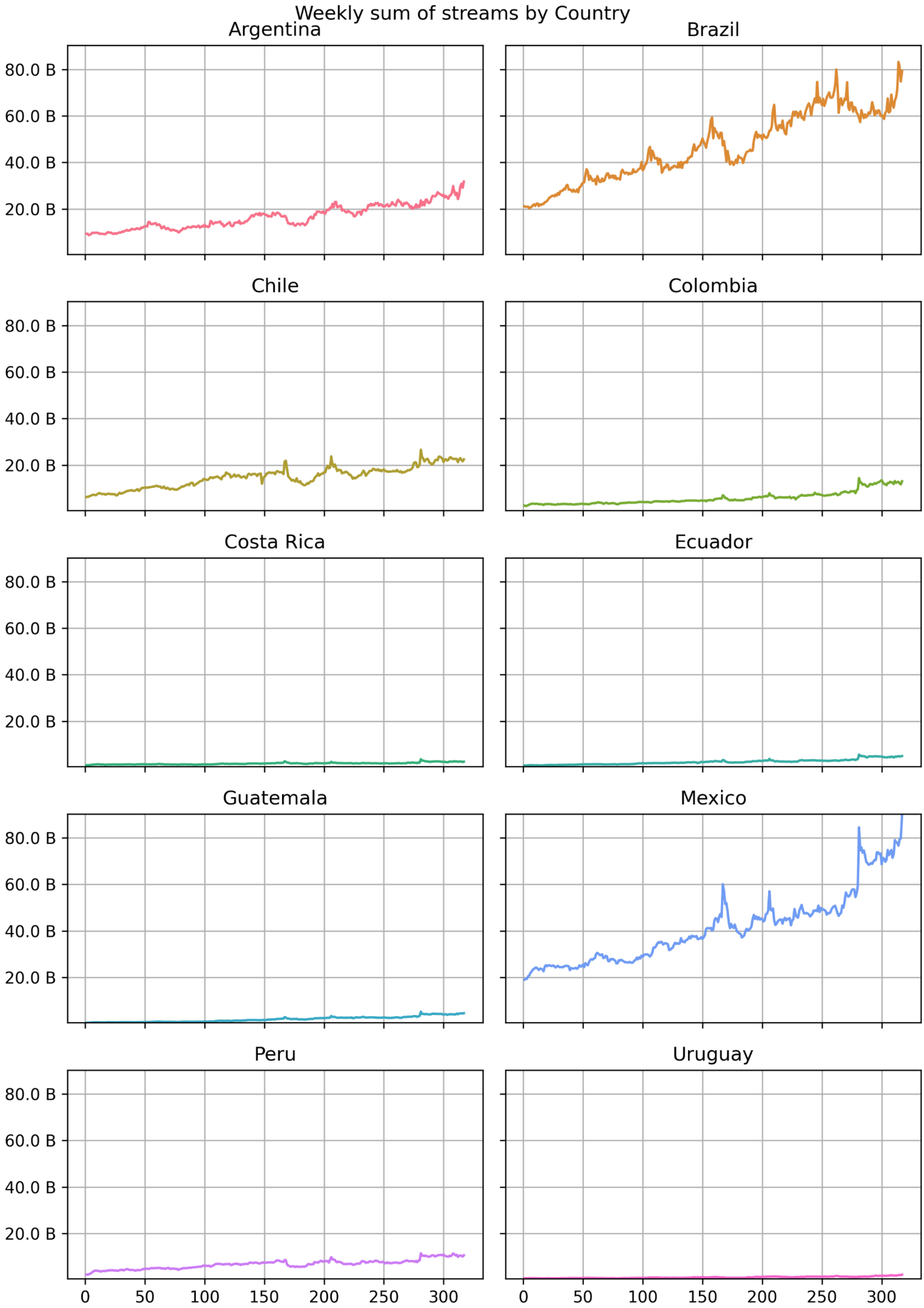
sum_st_sony	Mean	20,594,965.95	53,676,057.44	15,782,352.14	6,356,192.47	2,055,267.07	2,826,188.41	1,966,715.34	41,907,456.79	8,609,062.69	1,464,629.49
	Median	17,599,549.00	46,804,656.00	15,618,222.00	5,936,443.00	2,069,908.00	2,806,346.00	2,061,195.00	40,764,022.00	8,506,289.00	1,292,933.00
	Std. Deviation	6,714,187.78	36,246,556.41	3,667,418.65	1,663,855.76	395,682.83	716,419.96	761,359.37	7,330,847.19	1,669,804.14	480,689.83
	Skewness	.511	1.027	.211	.368	-.170	.241	.080	.685	.039	.447
	Std. Error of Skewness	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
	Kurtosis	-1.186	.267	-.757	-1.080	-.384	-.826	-1.433	.727	-.169	-1.206
	Std. Error of Kurtosis	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
	Minimum	10,688,629	13,642,593	9,137,871	3,360,733	1,179,196	1,460,660	816,005	27,723,315	3,765,886	732,525
	Maximum	34,991,331	170,368,951	25,909,262	10,973,140	3,102,135	4,578,807	3,574,382	72,223,874	13,012,009	2,540,476
sum_st_warner	Mean	11,552,926.67	23,716,781.88	7,430,096.32	3,565,006.26	1,177,743.39	1,566,847.92	1,023,030.40	26,532,937.66	4,630,354.62	768,588.69
	Median	10,253,208.00	22,773,279.00	6,961,918.00	3,300,653.00	1,184,119.00	1,474,385.00	949,602.00	26,658,751.00	4,286,166.00	689,318.00
	Std. Deviation	4,886,254.18	8,113,796.15	2,321,624.29	941,213.31	329,659.56	435,384.81	369,299.34	6,663,305.28	1,262,995.32	290,486.95
	Skewness	.956	1.368	.385	1.613	1.386	1.093	1.078	.688	.558	.741
	Std. Error of Skewness	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
	Kurtosis	-.112	2.141	-.889	3.263	7.365	1.404	1.436	.845	-.295	-.395
	Std. Error of Kurtosis	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
	Minimum	5.227.214	10.277.211	3.397.174	2.050.815	522.356	758.572	417.459	13.805.886	1.834.358	319.729
	Maximum	25.650.661	54.796.027	14.264.810	7.528.828	3.494.479	3.291.575	2.465.730	54.112.189	8.714.494	1.586.517
sum_st_indies	Mean	40,354,487.83	109,683,433.68	39,550,856.15	11,726,150.34	3,686,533.87	5,318,434.87	4,981,391.73	92,237,262.26	14,011,539.19	2,703,910.52
	Median	41,926,485.00	97,509,778.00	41,002,759.00	9,948,635.00	3,627,045.00	5,018,737.00	4,656,570.00	85,534,749.00	14,565,383.00	2,746,922.00
	Std. Deviation	17,747,883.88	46,036,973.23	19,075,209.06	8,123,435.86	1,825,505.54	3,739,907.51	3,899,979.15	62,812,451.91	6,751,289.13	1,135,607.39
	Skewness	.436	.773	.167	1.352	1.212	1.101	.874	1.073	.490	.675
	Std. Error of Skewness	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
	Kurtosis	-.171	-.080	-.527	1.461	2.199	.847	.009	.446	-.285	.422
	Std. Error of Kurtosis	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019

	Minimum	12,180,049	38,696,545	7,438,339	2,589,034	1,015,336	968,544	477,875	22,260,374	2,582,477	959,751
	Maximum	96,801,310	244,552,535	97,852,249	46,472,105	12,870,499	19,534,829	18,476,105	304,285,849	33,918,676	6,504,929
sum_st_artist_label	Mean	4,138,678.53	18,519,751.15	4,922,066.93	557,616.52	148,038.48	322,297.02	266,996.29	6,782,705.60	1,106,643.16	269,928.68
	Median	2,538,576.00	13,366,889.00	2,027,742.00	379,180.00	112,087.00	244,239.00	143,413.00	5,182,748.00	1,037,570.00	184,343.00
	Std. Deviation	4,736,643.42	12,602,725.68	6,380,443.02	514,475.07	119,672.47	270,656.07	283,408.34	5,747,091.49	720,245.73	295,516.21
	Skewness	3.285	.452	1.623	1.676	1.775	.955	1.032	.711	.407	2.983
	Std. Error of Skewness	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
	Kurtosis	13.766	-1.187	1.353	3.764	5.120	.078	-.037	-.465	-.786	11.792
	Std. Error of Kurtosis	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
	Minimum	492,514	1,857,053	66,743	0	0	15,246	0	273,575	41,762	0
	Maximum	35,815,200	50,360,988	25,135,139	3,261,818	846,495	1,116,723	1,108,050	23,964,355	3,138,469	2,184,461

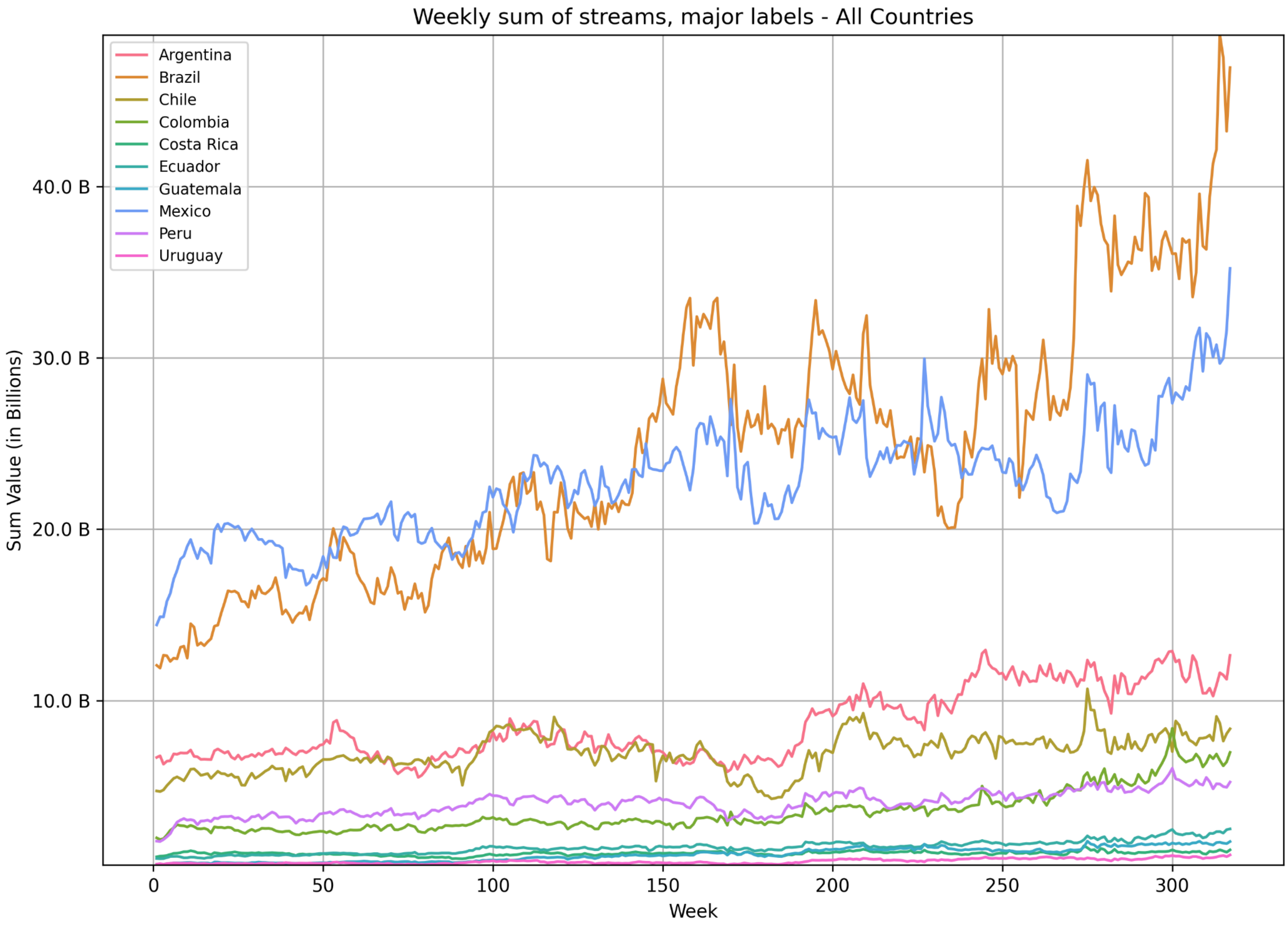


Weekly sum of streams - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
Sum Value (in Billions)
Week
90.0 B
80.0 B
70.0 B
60.0 B
50.0 B
40.0 B
30.0 B
20.0 B
10.0 B
0
50
100
150
200
250
300

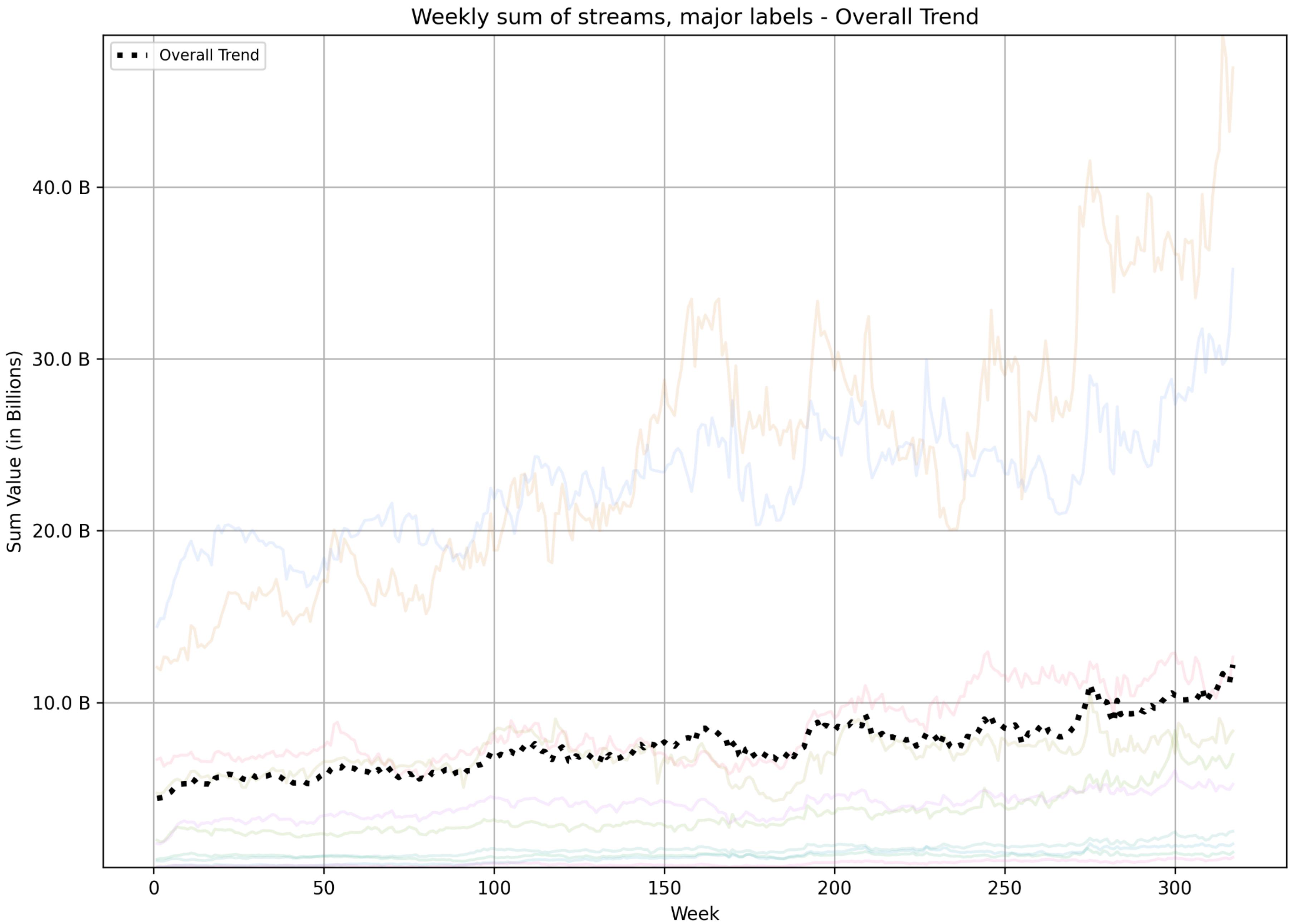

Weekly sum of streams - Overall Trend
Overall Trend
90.0 B
80.0 B
70.0 B
60.0 B
50.0 B
40.0 B
30.0 B
20.0 B
10.0 B
Sum Value (in Billions)
0
50
100
150
200
250
300
Week

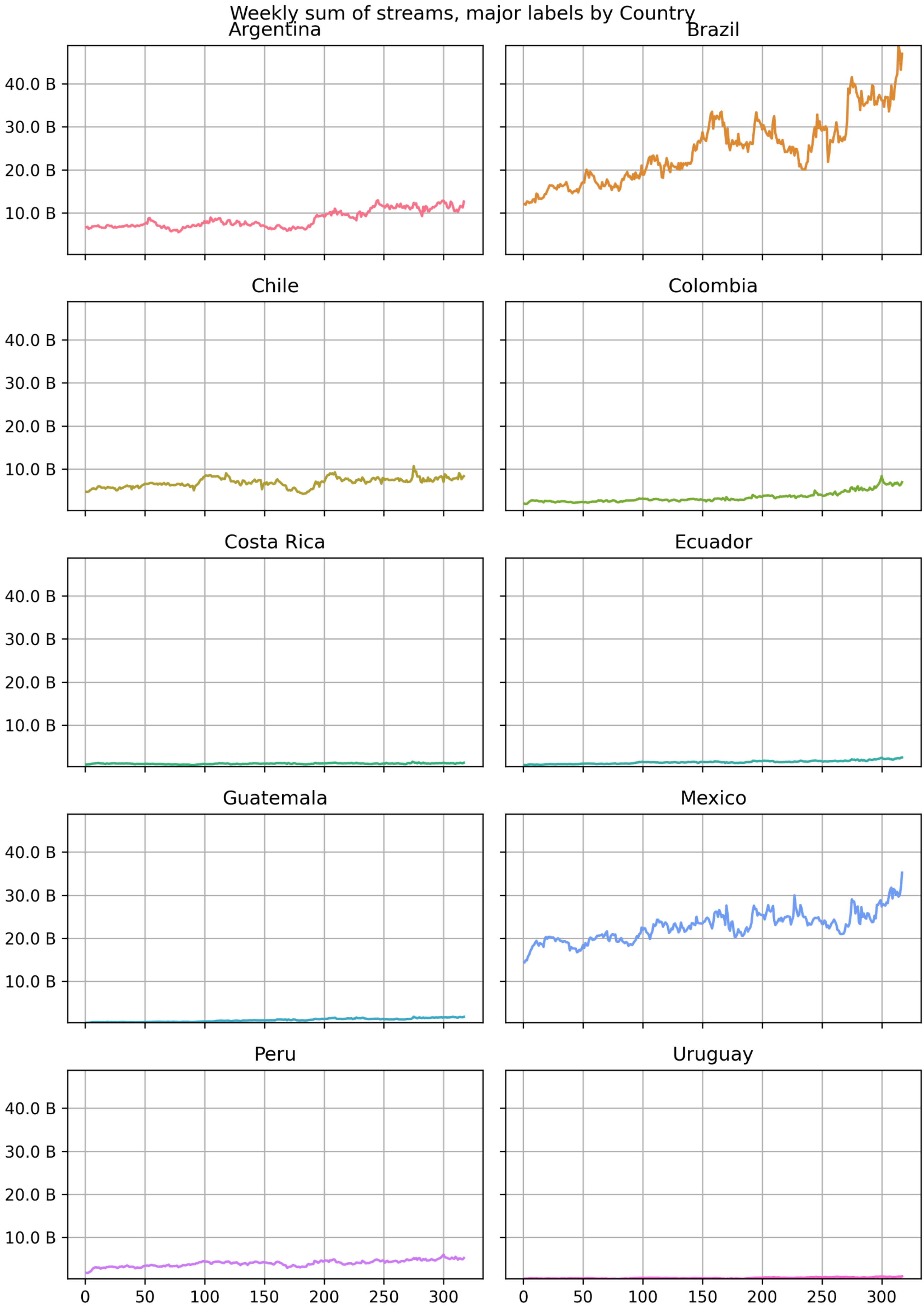


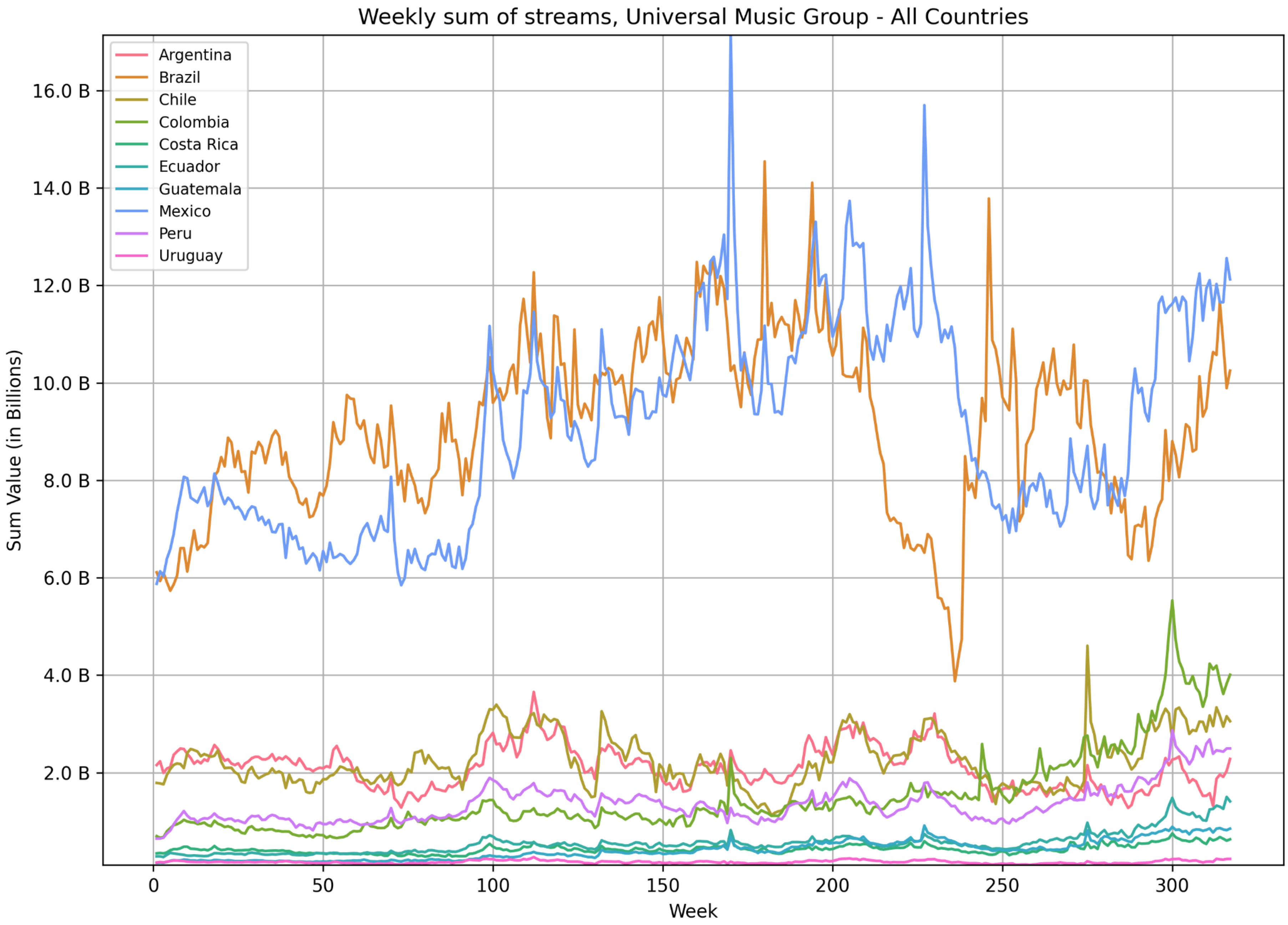
Weekly sum of streams by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
80.0 B
60.0 B
40.0 B
20.0 B
0
50
100
150
200
250
300



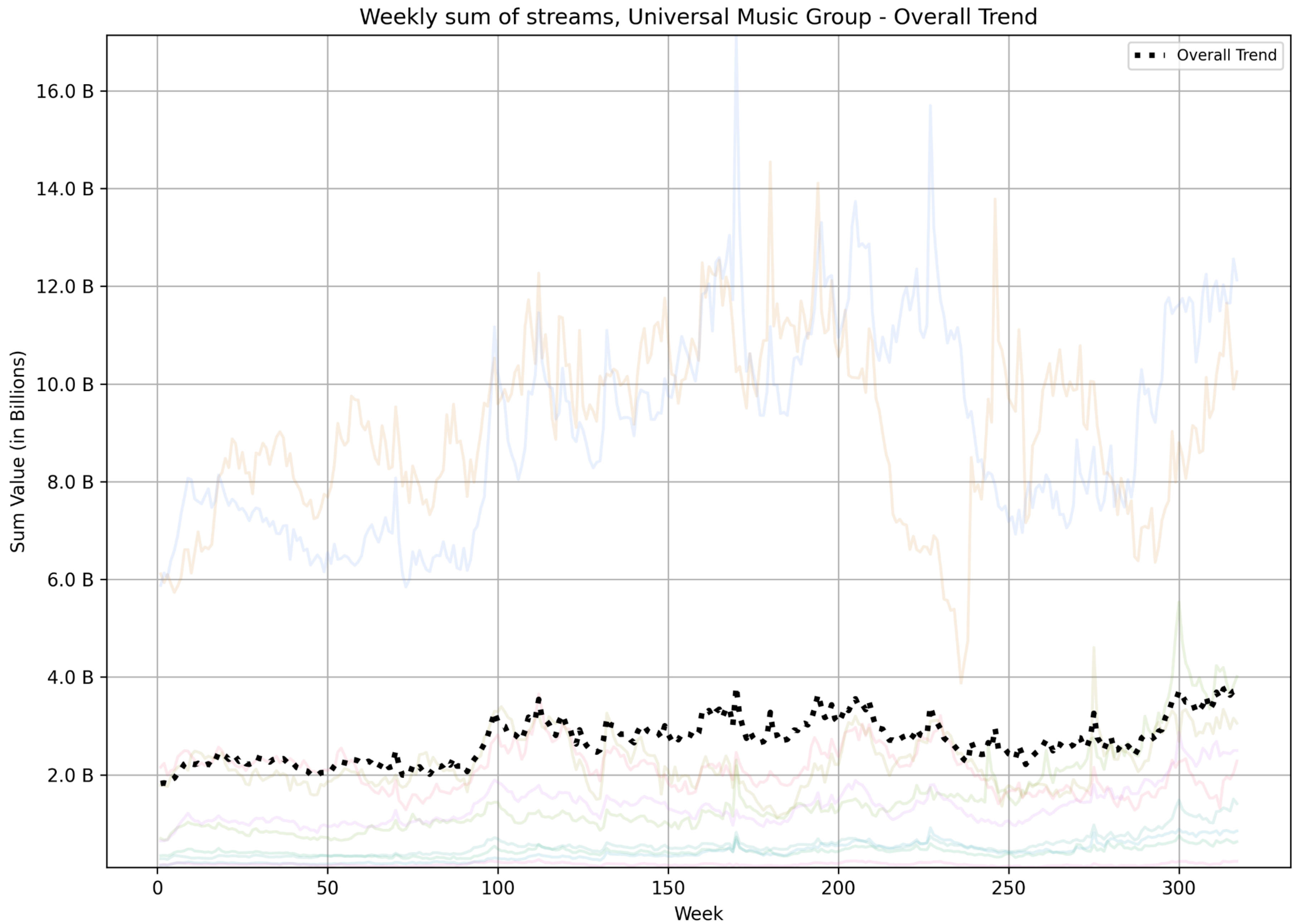
Weekly sum of streams, major labels - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
40.0 B
30.0 B
20.0 B
10.0 B
Sum Value (in Billions)
0
50
100
150
200
250
300
Week


Weekly sum of streams, major labels - Overall Trend
Overall Trend
Sum Value (in Billions)
40.0 B
30.0 B
20.0 B
10.0 B
0
50
100
150
200
250
300
Week

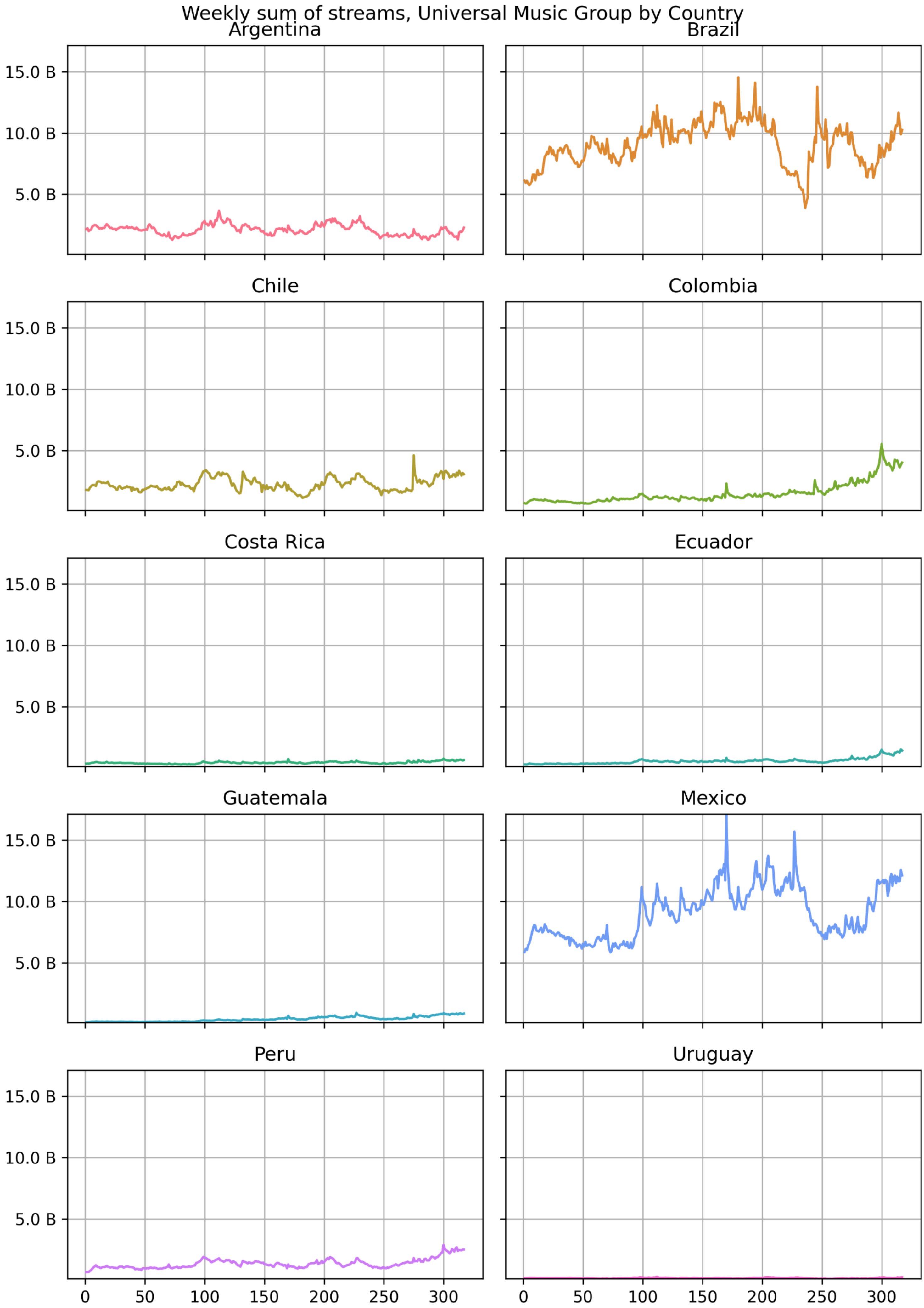

Weekly sum of streams, major labels by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
40.0 B
30.0 B
20.0 B
10.0 B
0
50
100
150
200
250
300

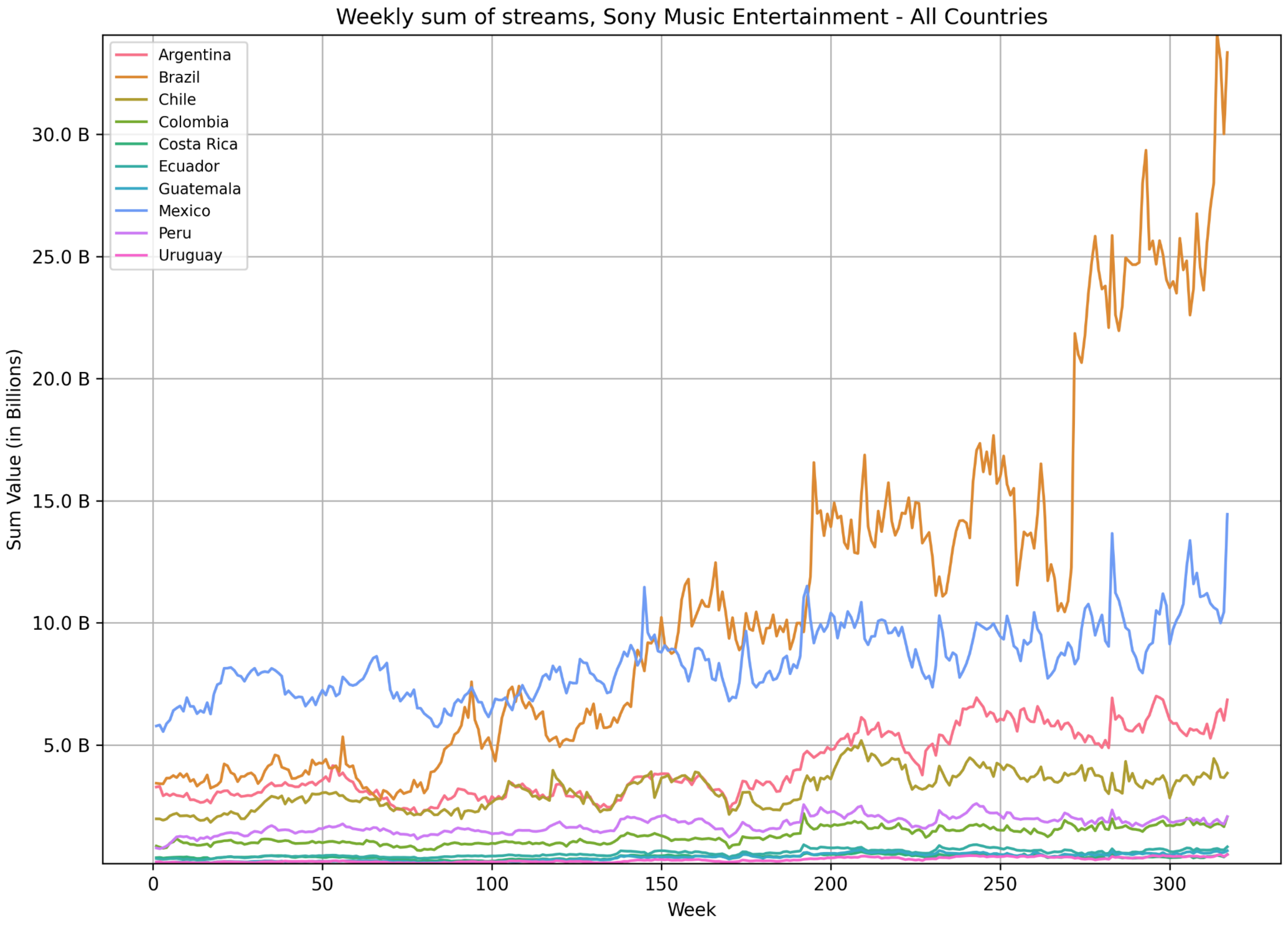


Weekly sum of streams, Universal Music Group - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
16.0 B
14.0 B
12.0 B
10.0 B
8.0 B
6.0 B
4.0 B
2.0 B
Sum Value (in Billions)
0
50
100
150
200
250
300
Week

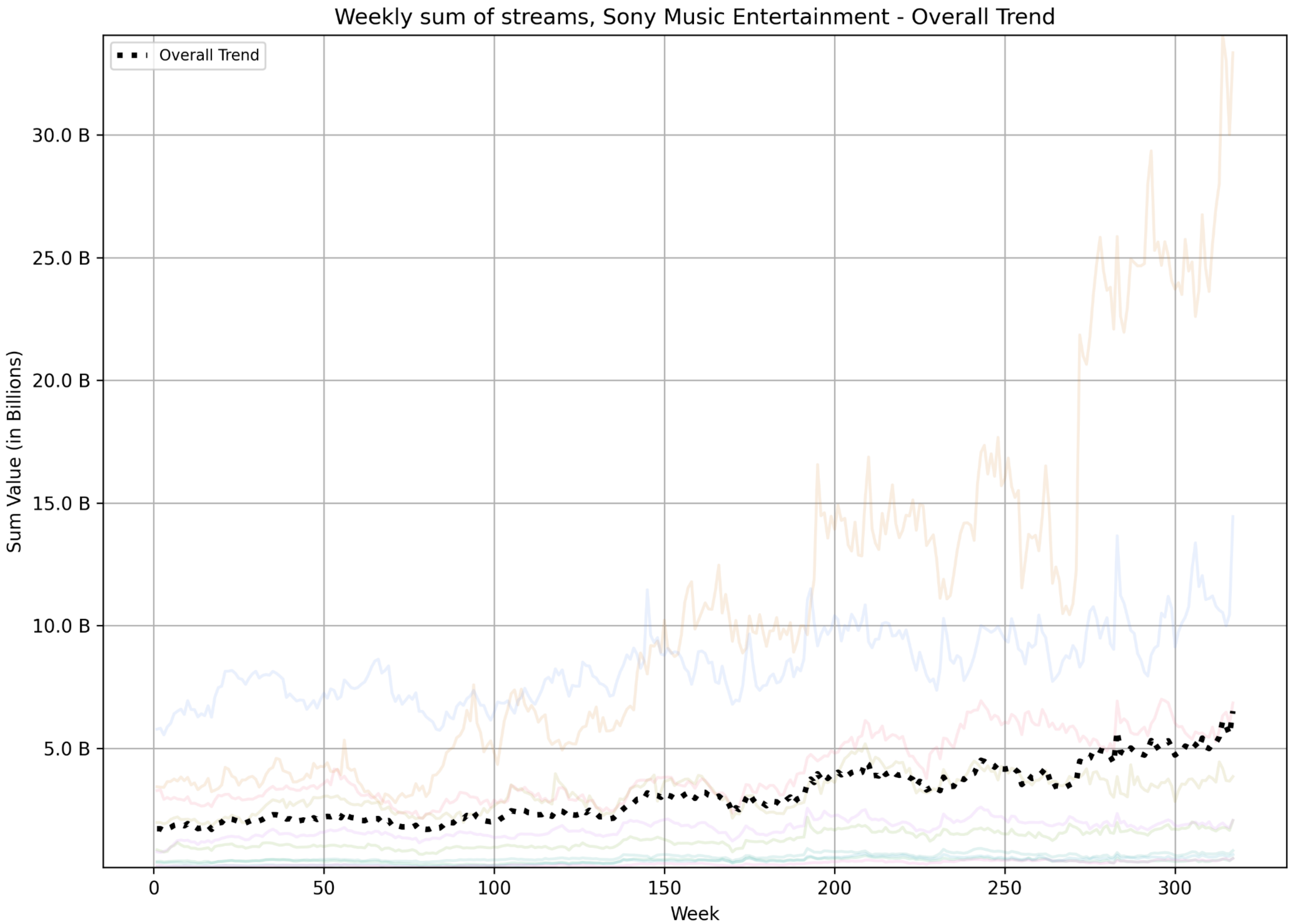


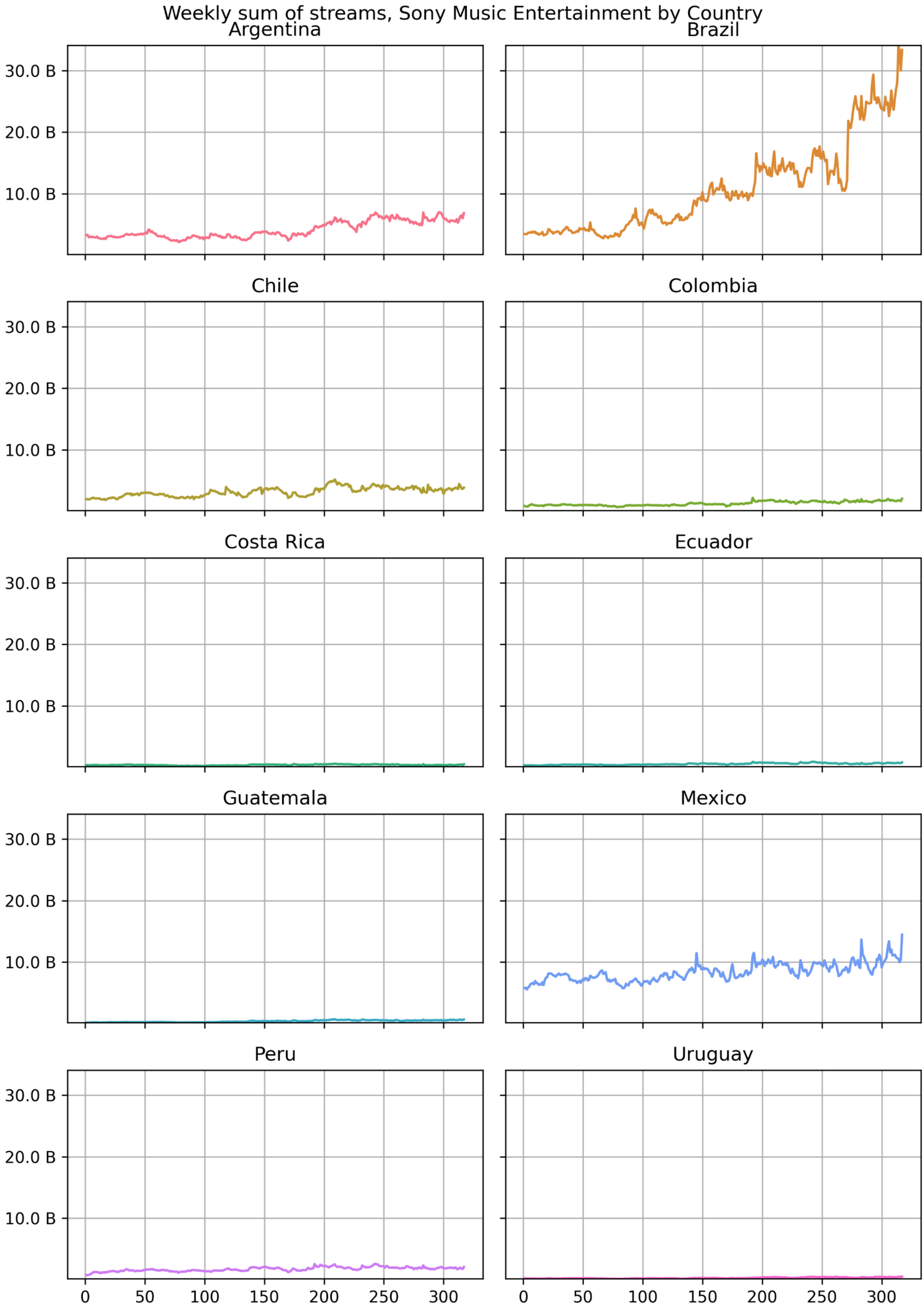
Weekly sum of streams, Universal Music Group - Overall Trend
Overall Trend
Sum Value (in Billions)
16.0 B
14.0 B
12.0 B
10.0 B
8.0 B
6.0 B
4.0 B
2.0 B
0
50
100
150
200
250
300
Week

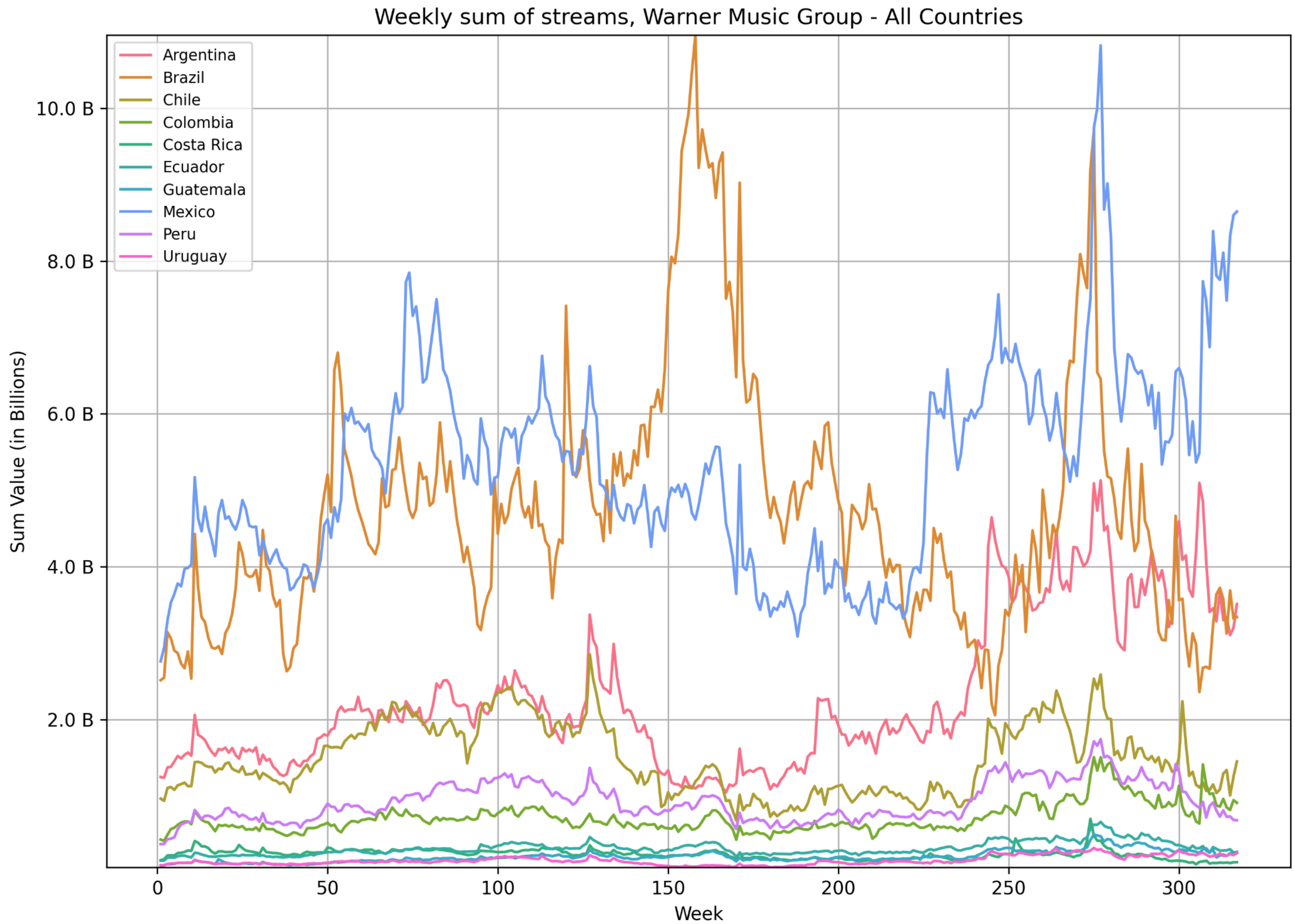

Weekly sum of streams, Universal Music Group by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
15.0 B
10.0 B
5.0 B
0
50
100
150
200
250
300

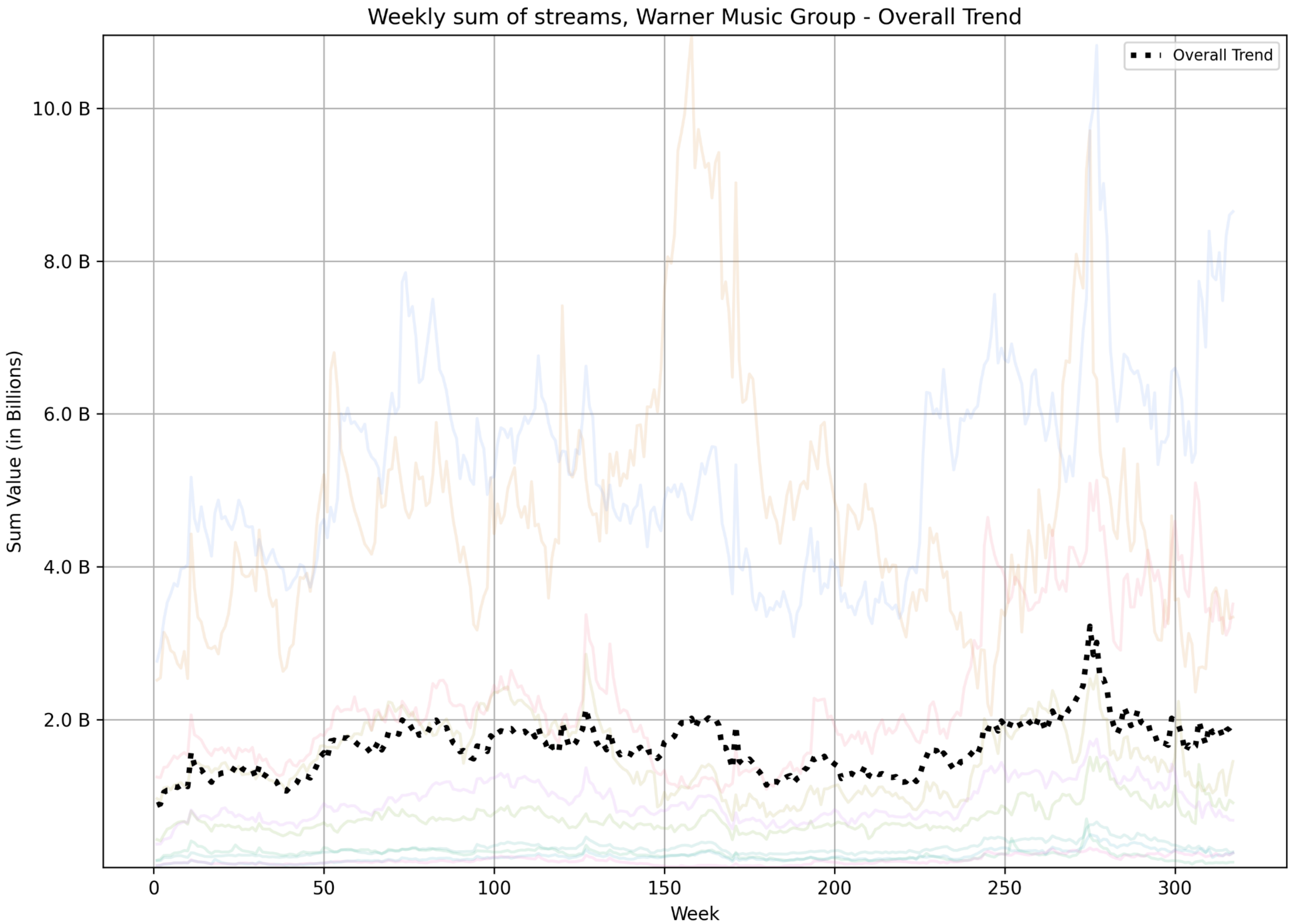


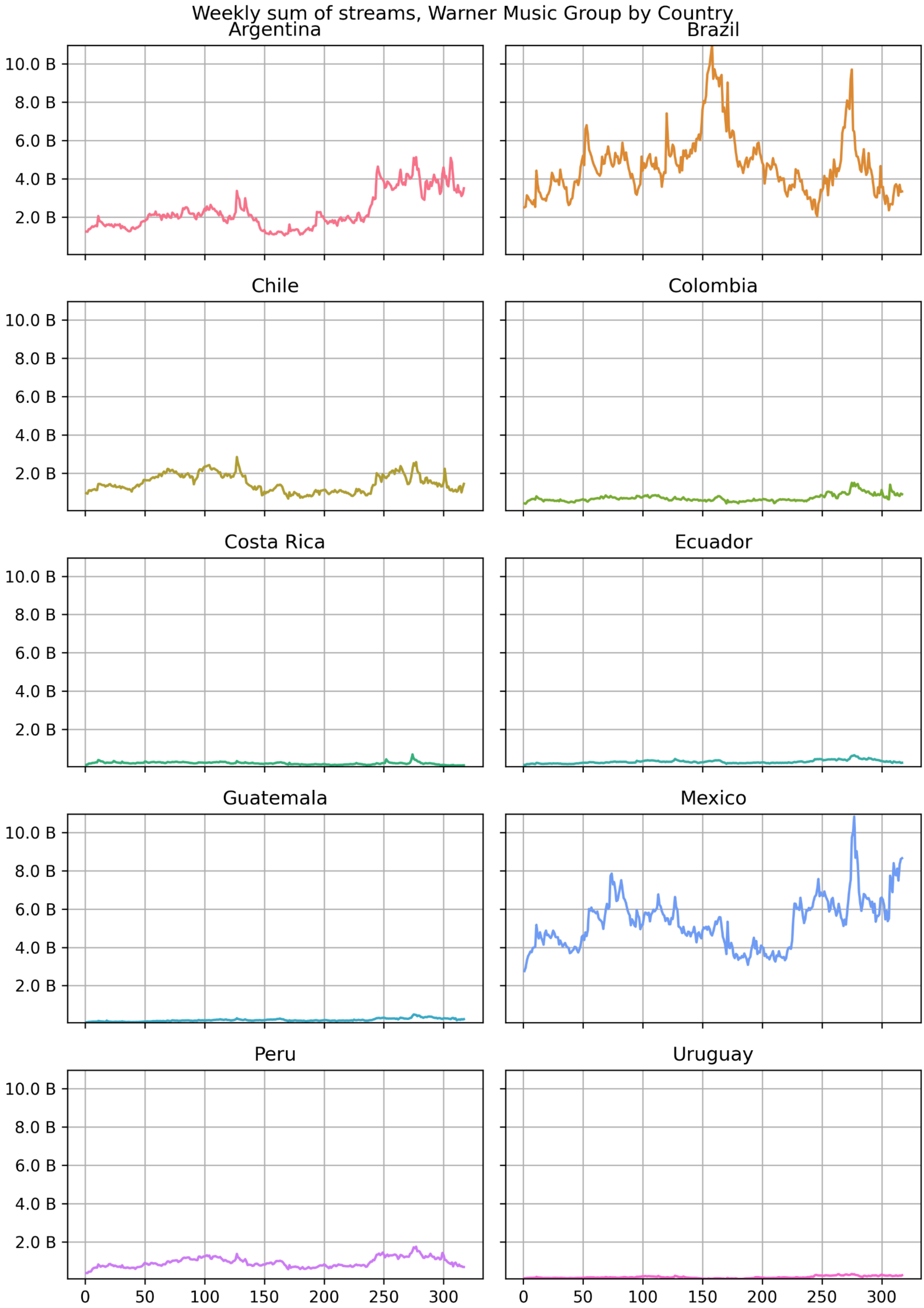
Weekly sum of streams, Sony Music Entertainment - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
Sum Value (in Billions)
30.0 B
25.0 B
20.0 B
15.0 B
10.0 B
5.0 B
0
50
100
150
200
250
300
Week

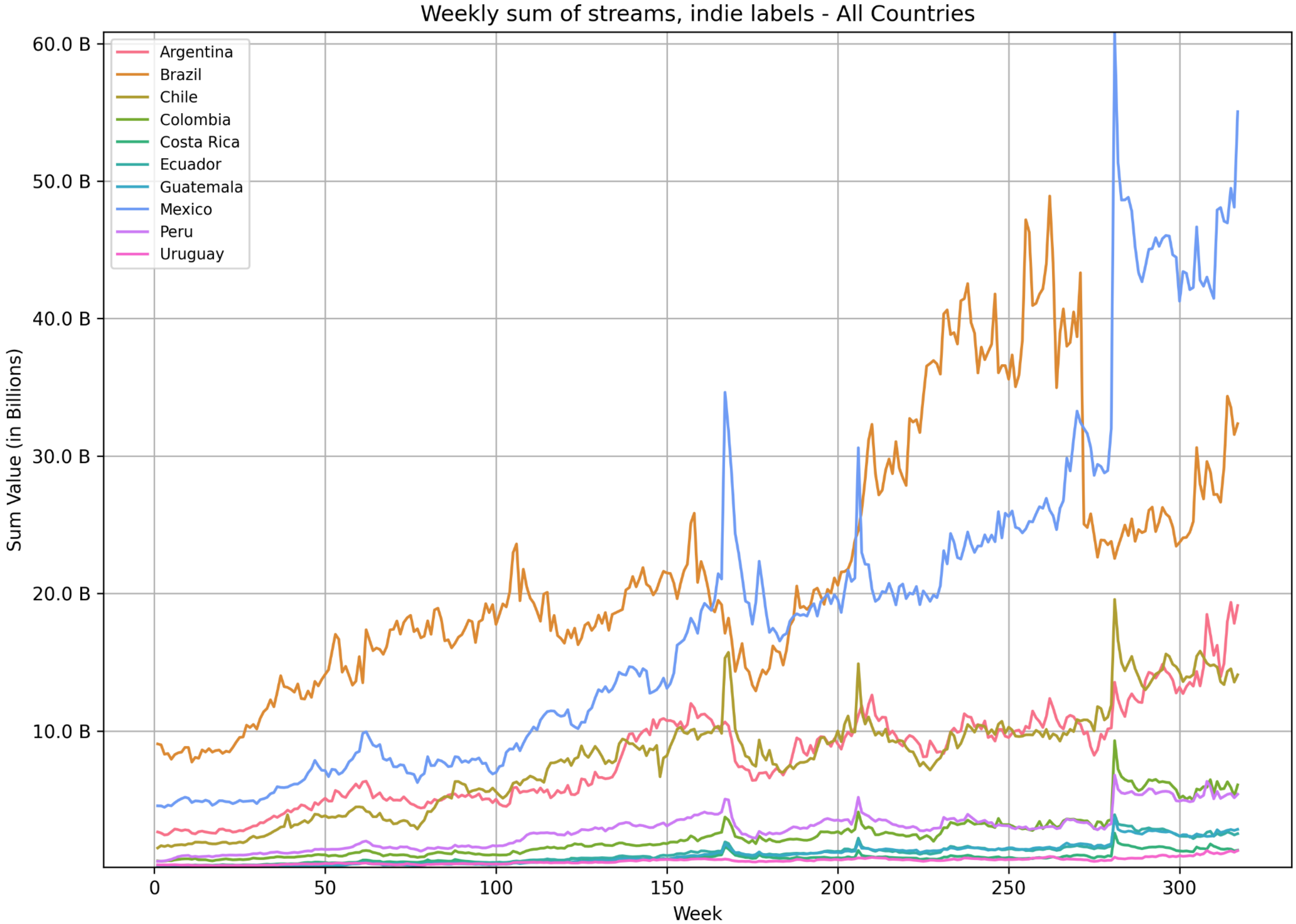

Weekly sum of streams, Sony Music Entertainment - Overall Trend
Overall Trend
Sum Value (in Billions)
30.0 B
25.0 B
20.0 B
15.0 B
10.0 B
5.0 B
0
50
100
150
200
250
300
Week


Weekly sum of streams, Sony Music Entertainment by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
30.0 B
20.0 B
10.0 B
0
50
100
150
200
250
300

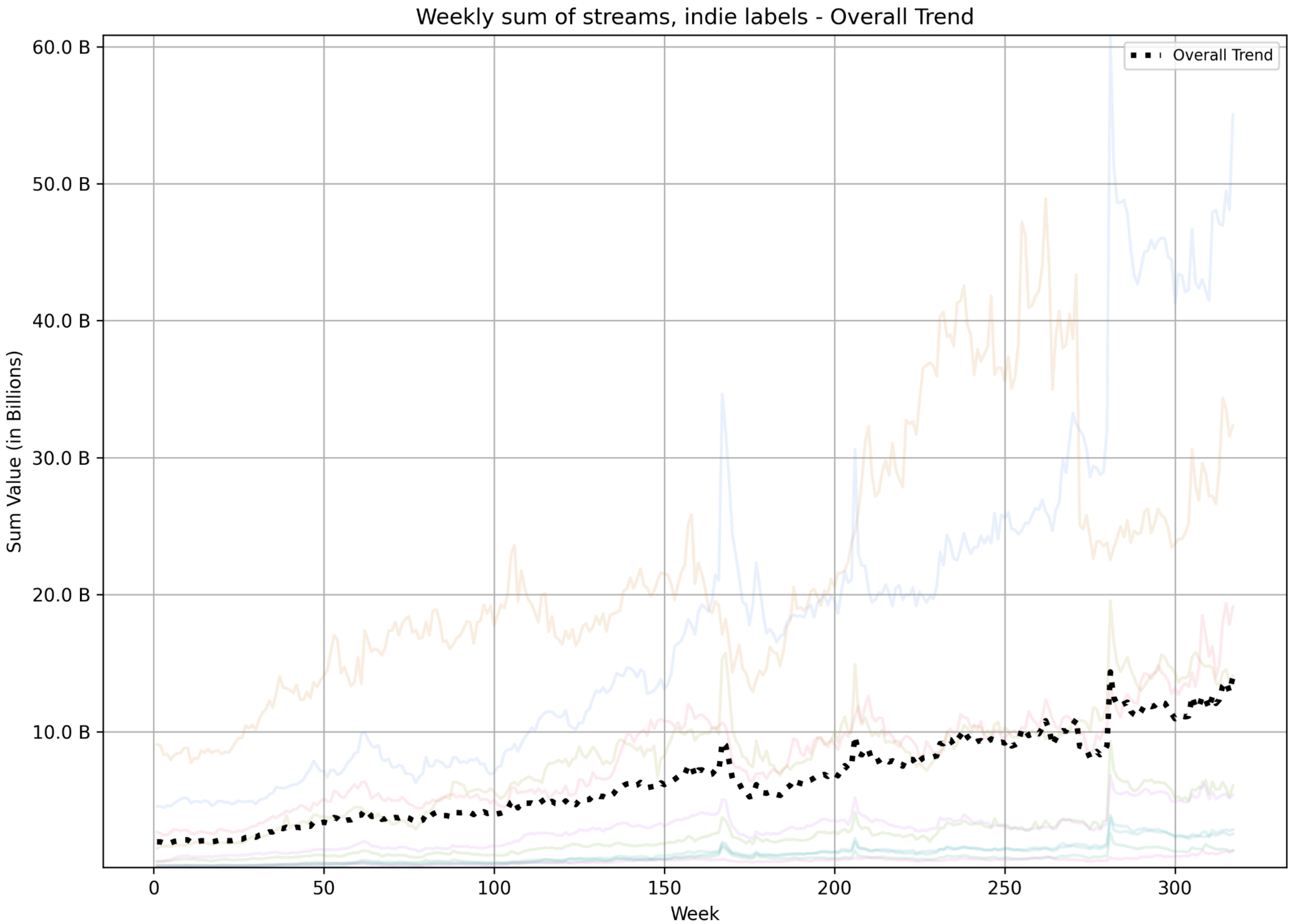

Weekly sum of streams, Warner Music Group - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
Sum Value (in Billions)
10.0 B
8.0 B
6.0 B
4.0 B
2.0 B
0
50
100
150
200
250
300
Week

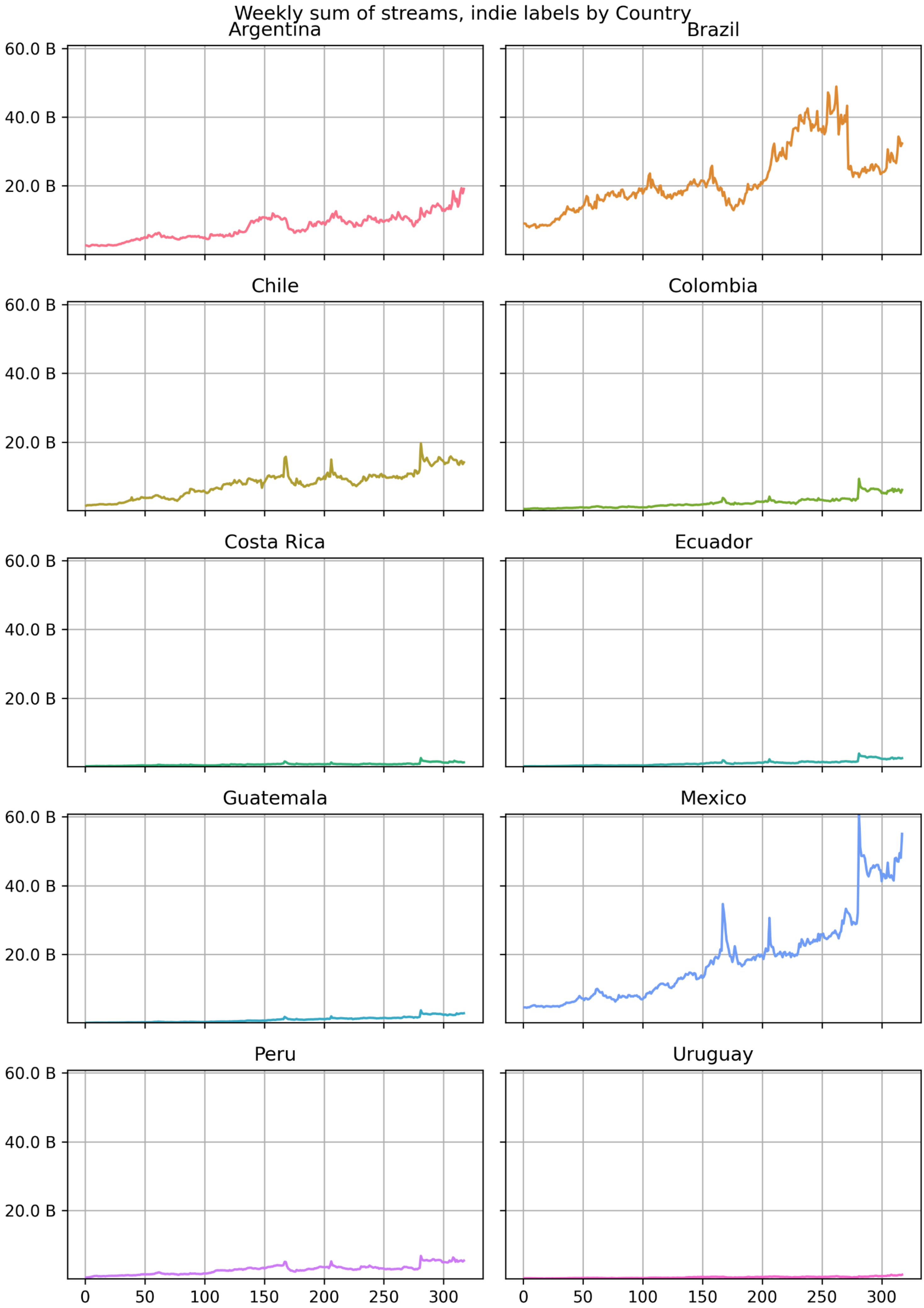

Weekly sum of streams, Warner Music Group - Overall Trend
Overall Trend
Sum Value (in Billions)
10.0 B
8.0 B
6.0 B
4.0 B
2.0 B
0
50
100
150
200
250
300
Week


Weekly sum of streams, Warner Music Group by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
10.0 B
8.0 B
6.0 B
4.0 B
2.0 B
0
50
100
150
200
250
300

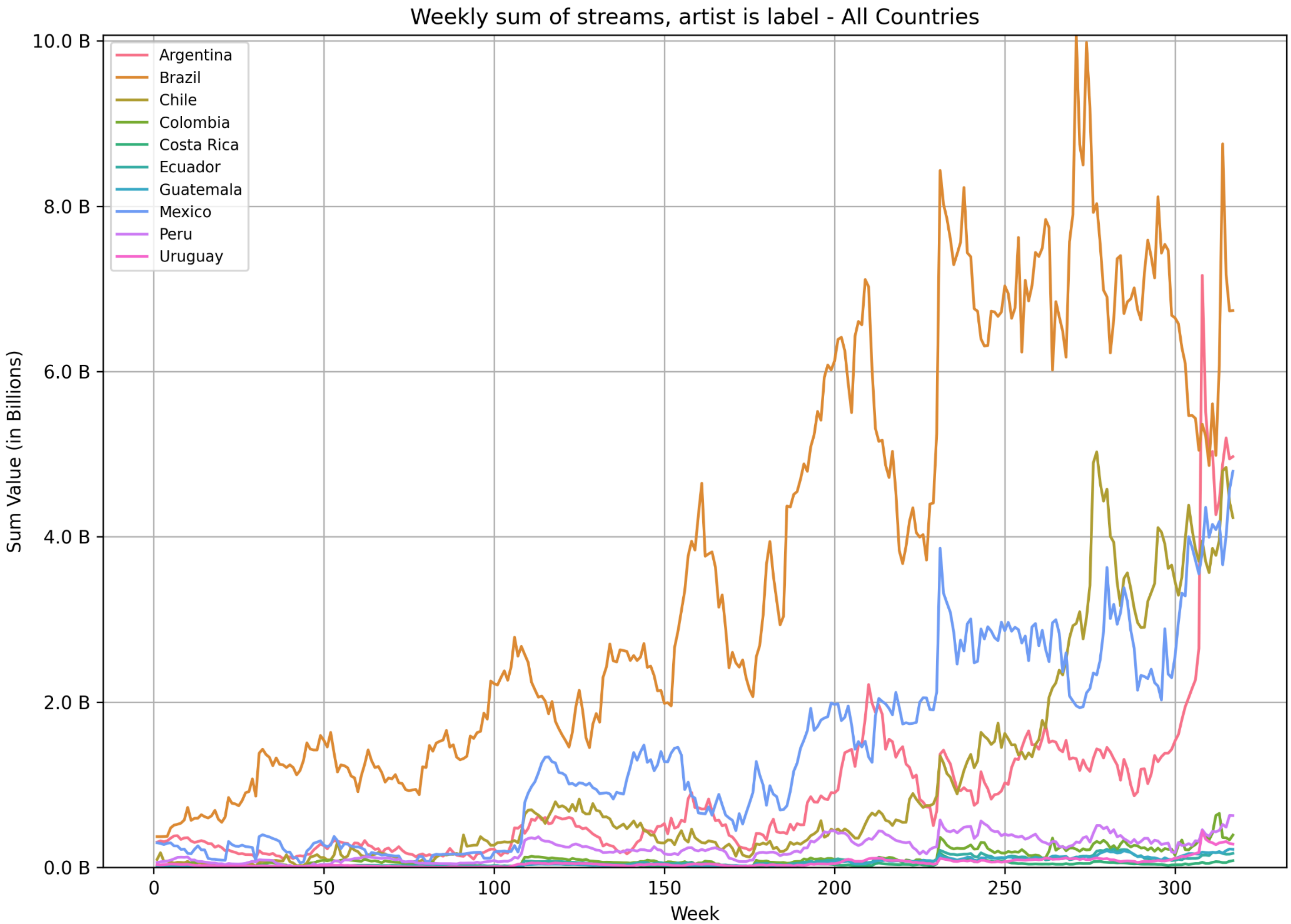


Weekly sum of streams, indie labels - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
60.0 B
50.0 B
40.0 B
30.0 B
20.0 B
10.0 B
Sum Value (in Billions)
0
50
100
150
200
250
300
Week

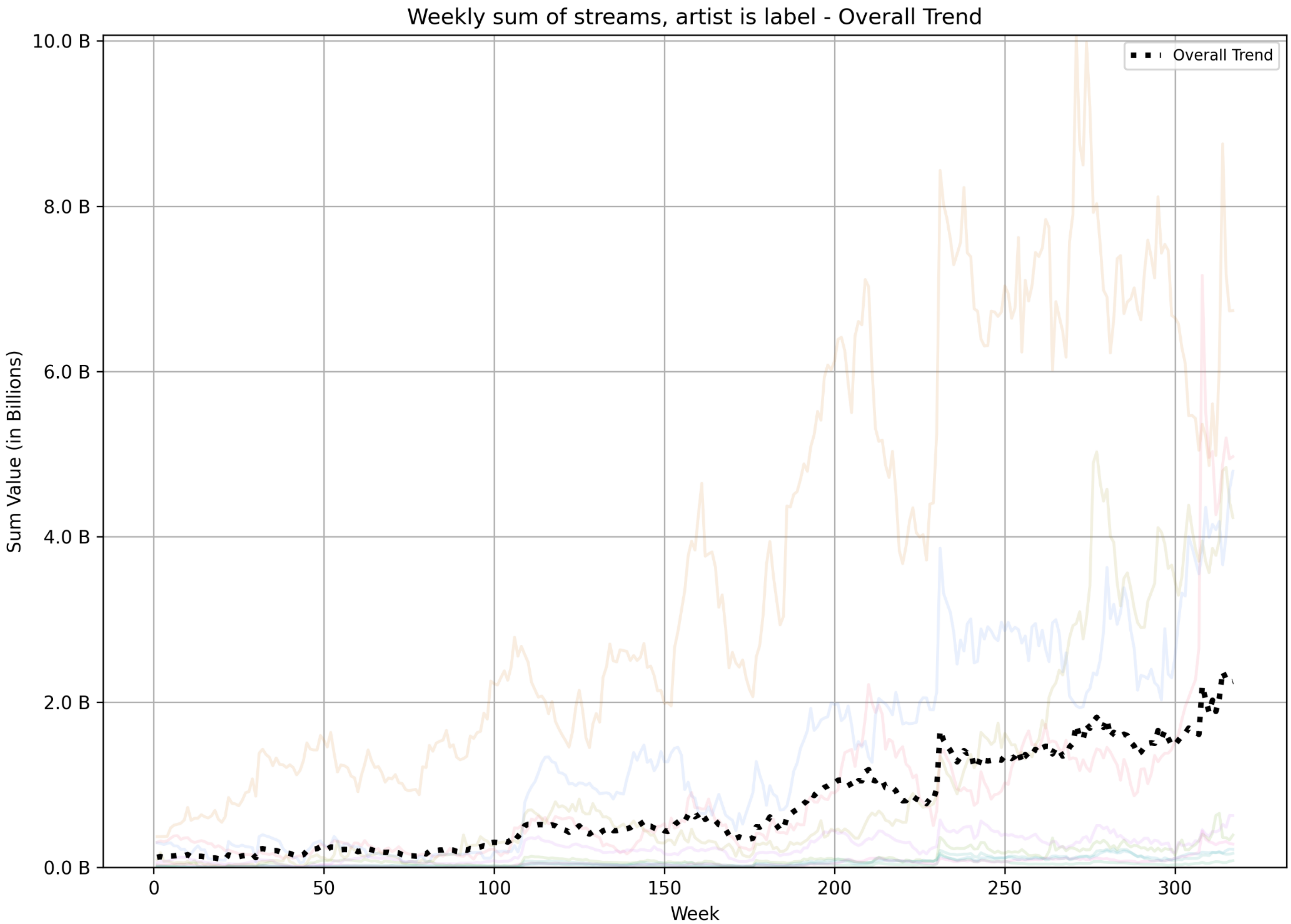

Weekly sum of streams, indie labels - Overall Trend
Overall Trend
Sum Value (in Billions)
60.0 B
50.0 B
40.0 B
30.0 B
20.0 B
10.0 B
0
50
100
150
200
250
300
Week

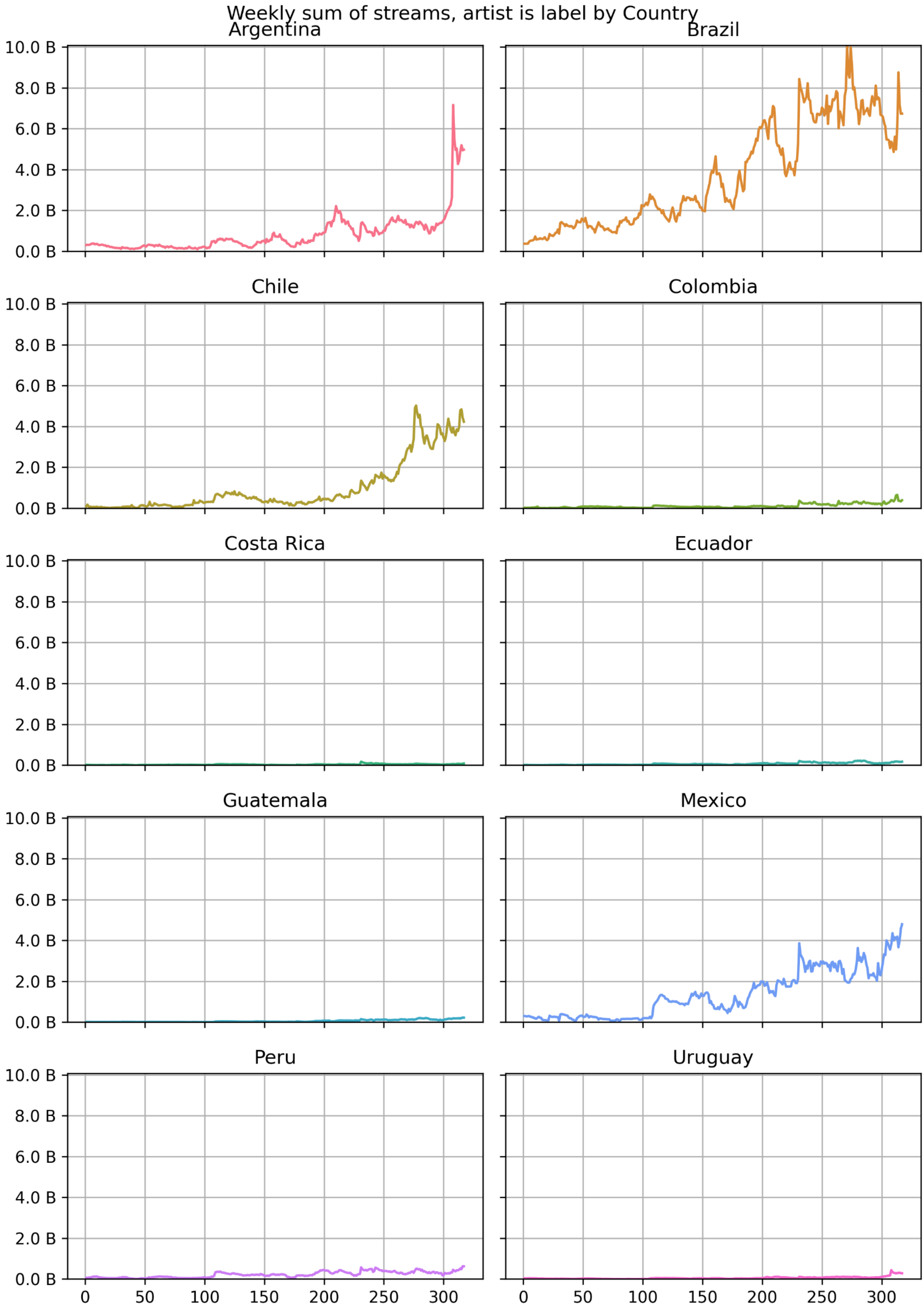


Weekly sum of streams, indie labels by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
60.0 B
40.0 B
20.0 B
0
50
100
150
200
250
300



Weekly sum of streams, artist is label - All Countries
Sum Value (in Billions)
10.0 B
8.0 B
6.0 B
4.0 B
2.0 B
0.0 B
0
50
100
150
200
250
300
Week
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay


Weekly sum of streams, artist is label - Overall Trend
Overall Trend
Sum Value (in Billions)
0.0 B
2.0 B
4.0 B
6.0 B
8.0 B
10.0 B
Week
0
50
100
150
200
250
300



Weekly sum of streams, artist is label by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
10.0 B
8.0 B
6.0 B
4.0 B
2.0 B
0.0 B
0
50
100
150
200
250
300

14.2. Appendix B – Share of streams: tables and graphs

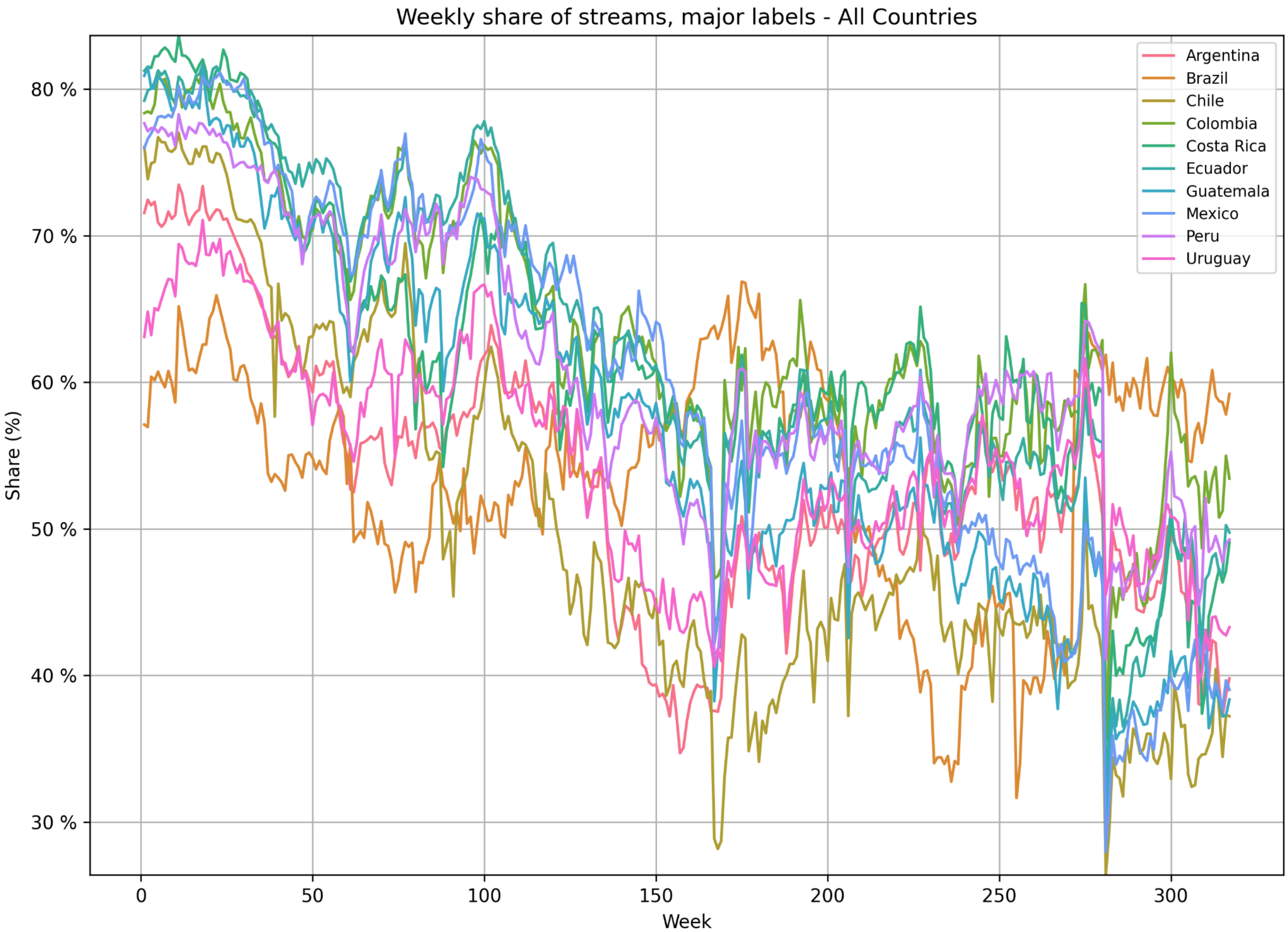
Table B1. Descriptive statistics – Share of streams

	share_st_majors	share_st_universal	share_st_sony	share_st_warner	share_st_indies	share_artist_label
Mean	57.63%	20.68%	23.65%	13.29%	42.37%	3.39%
Median	57.13%	21.27%	23.71%	13.10%	42.87%	2.55%
Std. Deviation	11.34%	6.27%	5.62%	4.72%	11.34%	3.23%
Minimum	26.39%	5.43%	7.87%	2.75%	16.33%	0.00%
Maximum	83.67%	40.96%	45.95%	29.98%	73.61%	24.39%

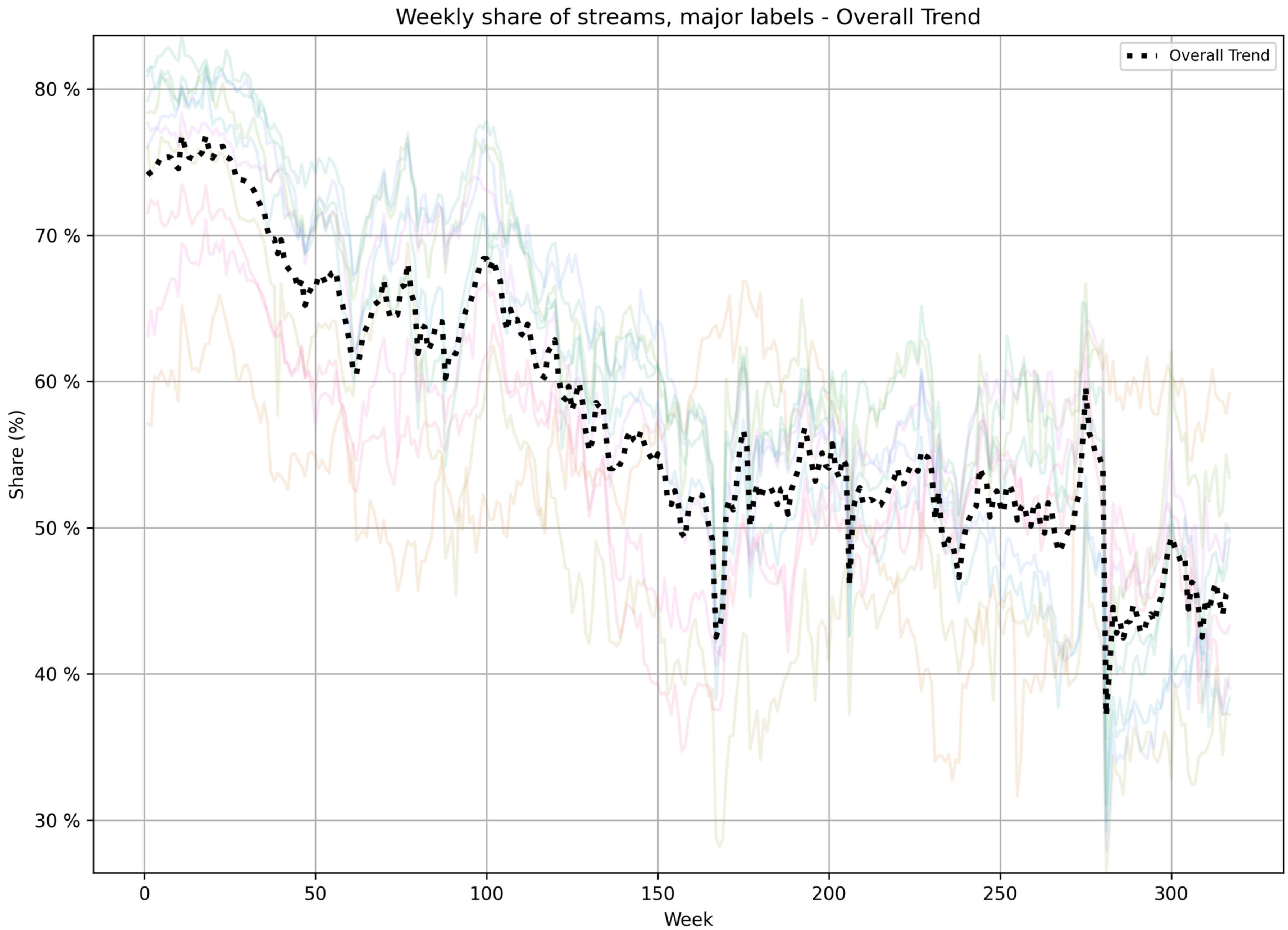
Table B2. Descriptive statistics – Share of streams – By country

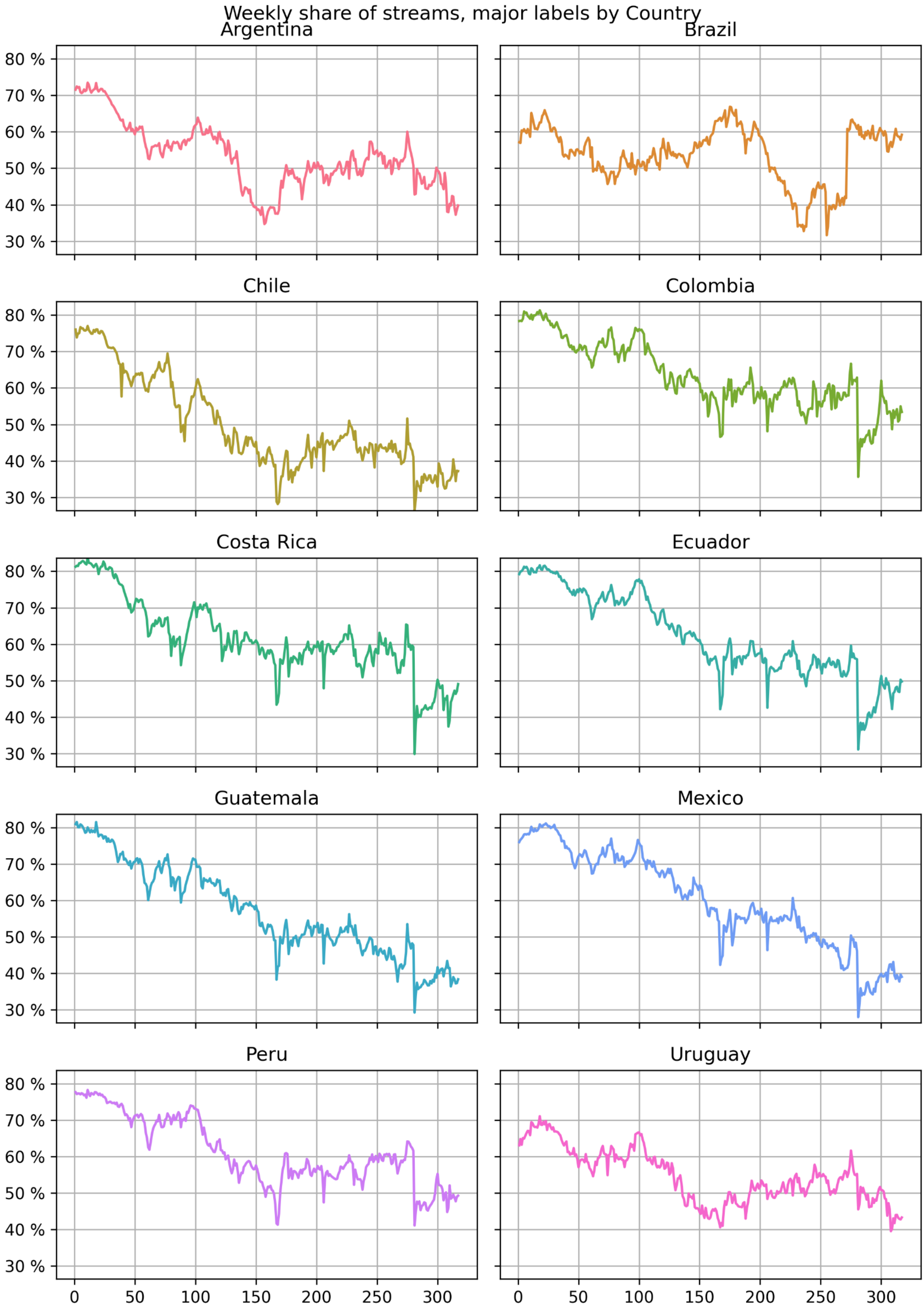
		Argentina	Brazil	Chile	Colombia	Costa Rica	Ecuador	Guatemala	Mexico	Peru	Uruguay
share_st_majors	Mean	53.29%	53.76%	49.90%	63.22%	61.44%	62.08%	56.92%	59.89%	61.15%	54.65%
	Median	52.47%	54.65%	45.23%	61.36%	60.01%	59.78%	53.38%	58.02%	59.02%	54.01%
	Std. Deviation	8.92%	7.63%	12.62%	9.30%	10.43%	11.68%	12.57%	13.22%	9.23%	7.34%
	Minimum	34.69%	31.64%	26.39%	35.60%	29.87%	31.08%	29.21%	27.96%	41.08%	39.52%
	Maximum	73.47%	66.86%	77.01%	81.32%	83.67%	81.64%	81.49%	81.11%	78.27%	71.06%
share_st_universal	Mean	14.12%	21.53%	16.49%	25.50%	24.17%	23.20%	21.96%	23.79%	20.45%	15.61%
	Median	13.70%	23.60%	14.83%	25.26%	23.25%	23.07%	21.62%	25.21%	20.16%	15.08%
	Std. Deviation	5.40%	6.81%	5.73%	3.99%	4.09%	4.52%	4.19%	5.83%	4.00%	4.86%
	Minimum	5.43%	6.31%	7.83%	16.38%	13.65%	12.75%	13.08%	8.86%	12.38%	7.00%
	Maximum	26.65%	36.09%	32.37%	40.96%	36.97%	36.99%	31.79%	38.05%	30.53%	25.54%
share_st_sony	Mean	25.02%	21.11%	22.27%	23.88%	23.32%	24.82%	22.43%	21.97%	26.38%	25.34%
	Median	24.98%	18.65%	21.88%	23.94%	24.29%	25.07%	21.97%	21.49%	26.66%	25.32%
	Std. Deviation	3.75%	8.89%	4.42%	5.46%	4.92%	5.79%	5.70%	5.12%	4.74%	2.90%
	Minimum	16.73%	7.87%	11.91%	10.79%	9.50%	9.63%	9.26%	10.96%	15.99%	17.99%
	Maximum	35.33%	45.95%	32.70%	36.02%	35.04%	37.23%	34.75%	33.30%	36.33%	32.03%
share_st_warner	Mean	14.14%	11.12%	11.13%	13.84%	13.95%	14.05%	12.53%	14.14%	14.32%	13.70%
	Median	14.84%	11.59%	10.51%	13.15%	13.10%	14.05%	11.93%	13.31%	14.48%	14.43%
	Std. Deviation	4.15%	4.11%	4.91%	4.53%	5.36%	4.37%	4.91%	4.89%	4.11%	4.15%
	Minimum	6.04%	2.75%	4.28%	5.20%	3.90%	5.26%	4.37%	5.91%	6.24%	6.12%
	Maximum	24.72%	20.09%	22.49%	23.67%	29.98%	23.86%	24.08%	28.30%	22.94%	23.93%
share_st_indies	Mean	46.71%	46.24%	50.10%	36.78%	38.56%	37.92%	43.08%	40.11%	38.85%	45.35%
	Median	47.53%	45.35%	54.77%	38.64%	39.99%	40.22%	46.62%	41.98%	40.98%	45.99%
	Std. Deviation	8.92%	7.63%	12.62%	9.30%	10.43%	11.68%	12.57%	13.22%	9.23%	7.34%
	Minimum	26.53%	33.14%	22.99%	18.68%	16.33%	18.36%	18.51%	18.89%	21.73%	28.94%

	Maximum	65.31%	68.36%	73.61%	64.40%	70.13%	68.92%	70.79%	72.04%	58.92%	60.48%
share_artist_label	Mean	4.23%	7.09%	5.31%	1.69%	1.56%	2.24%	2.02%	2.81%	3.01%	3.99%
	Median	3.49%	6.26%	2.59%	1.58%	1.27%	1.91%	1.68%	2.99%	2.87%	3.36%
	Std. Deviation	3.29%	3.29%	5.81%	0.99%	1.16%	1.34%	1.48%	1.84%	1.57%	2.96%
	Minimum	0.79%	1.76%	0.17%	0.00%	0.00%	0.23%	0.00%	0.23%	0.20%	0.00%
	Maximum	24.01%	15.19%	24.39%	5.49%	8.88%	6.68%	5.90%	7.93%	7.36%	20.31%

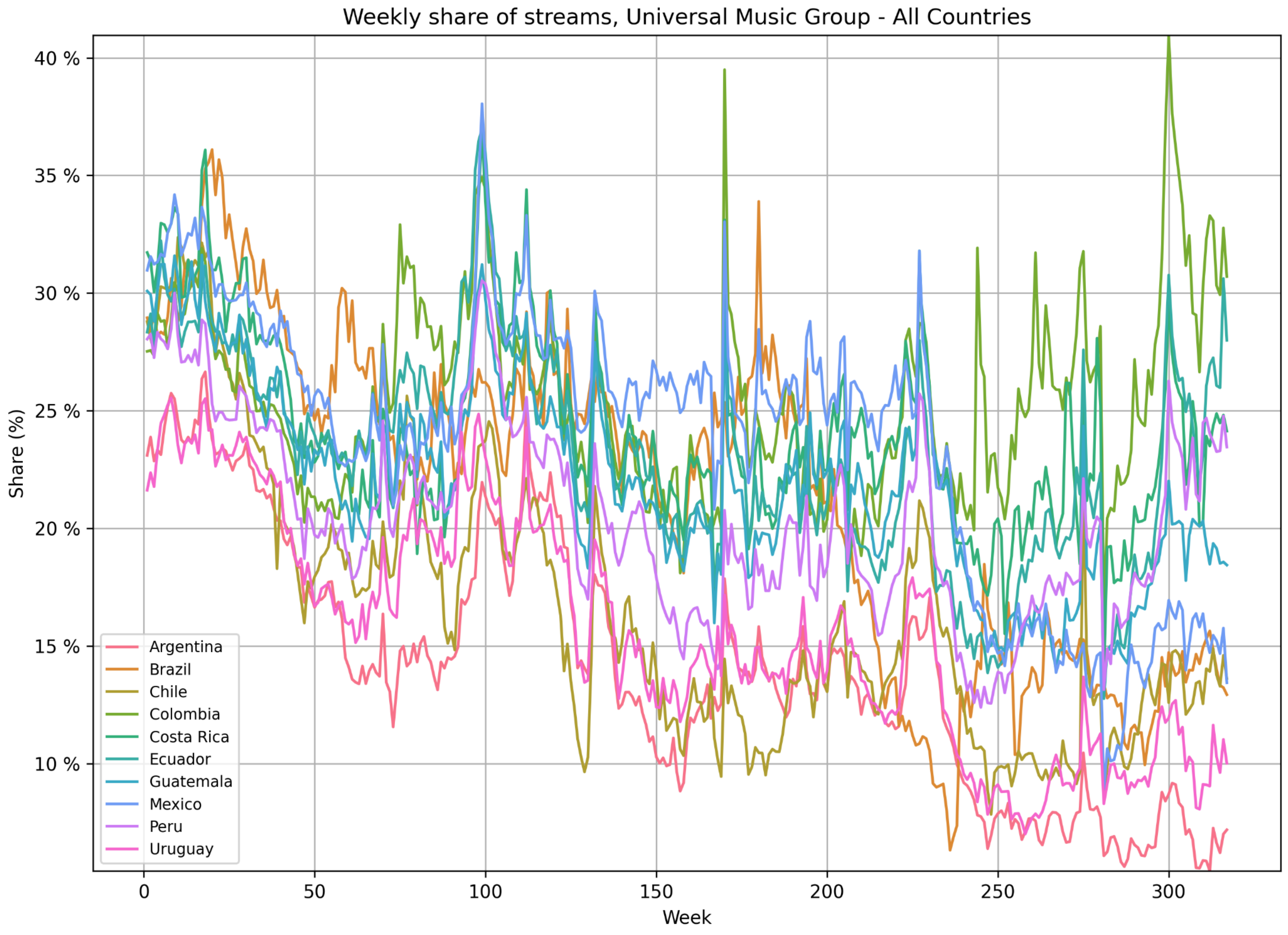


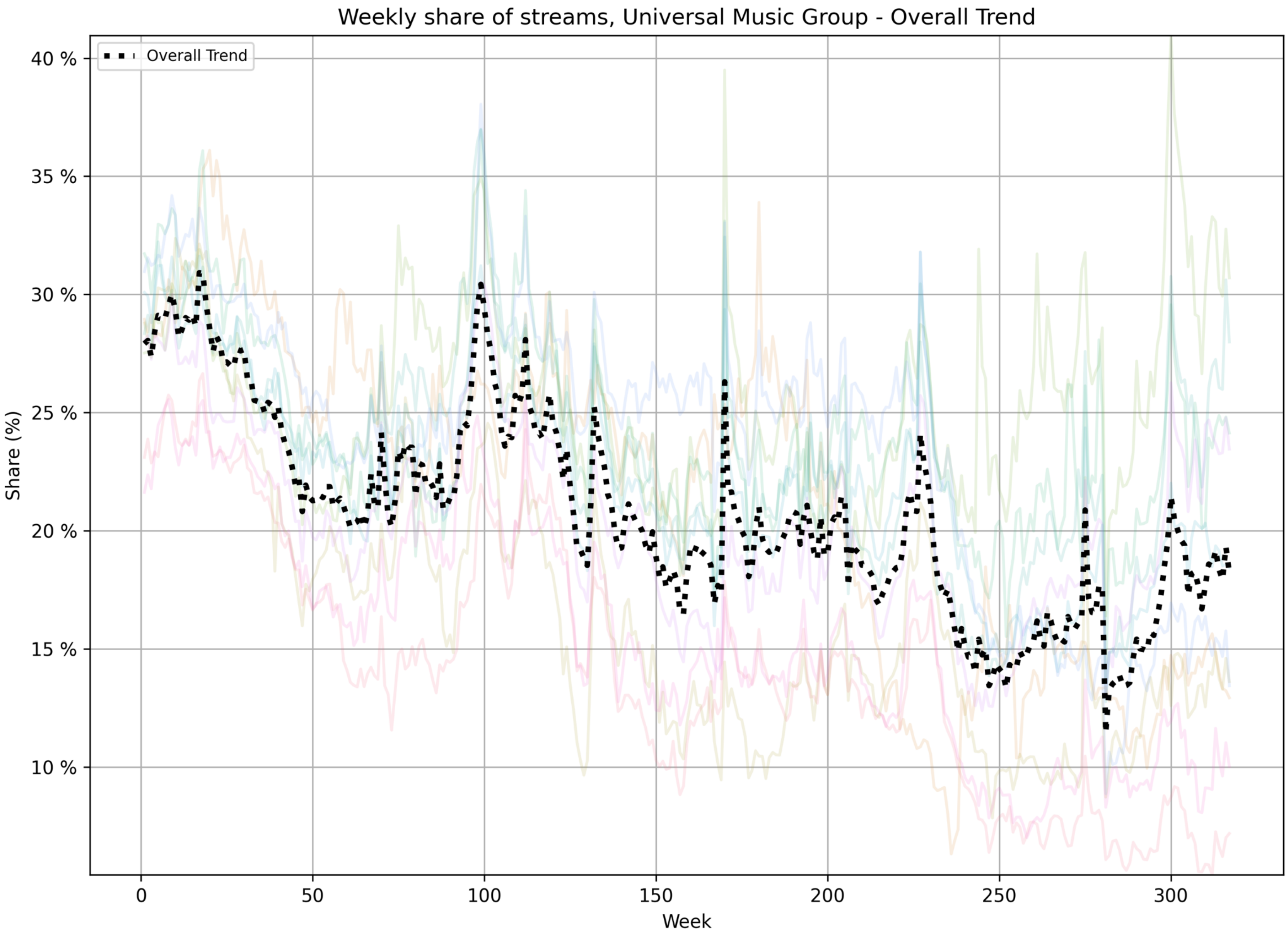
Weekly share of streams, major labels - All Countries
Share (%)
Week
80 %
70 %
60 %
50 %
40 %
30 %
0
50
100
150
200
250
300
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay


Weekly share of streams, major labels - Overall Trend
Overall Trend
Share (%)
Week
80 %
70 %
60 %
50 %
40 %
30 %
0
50
100
150
200
250
300



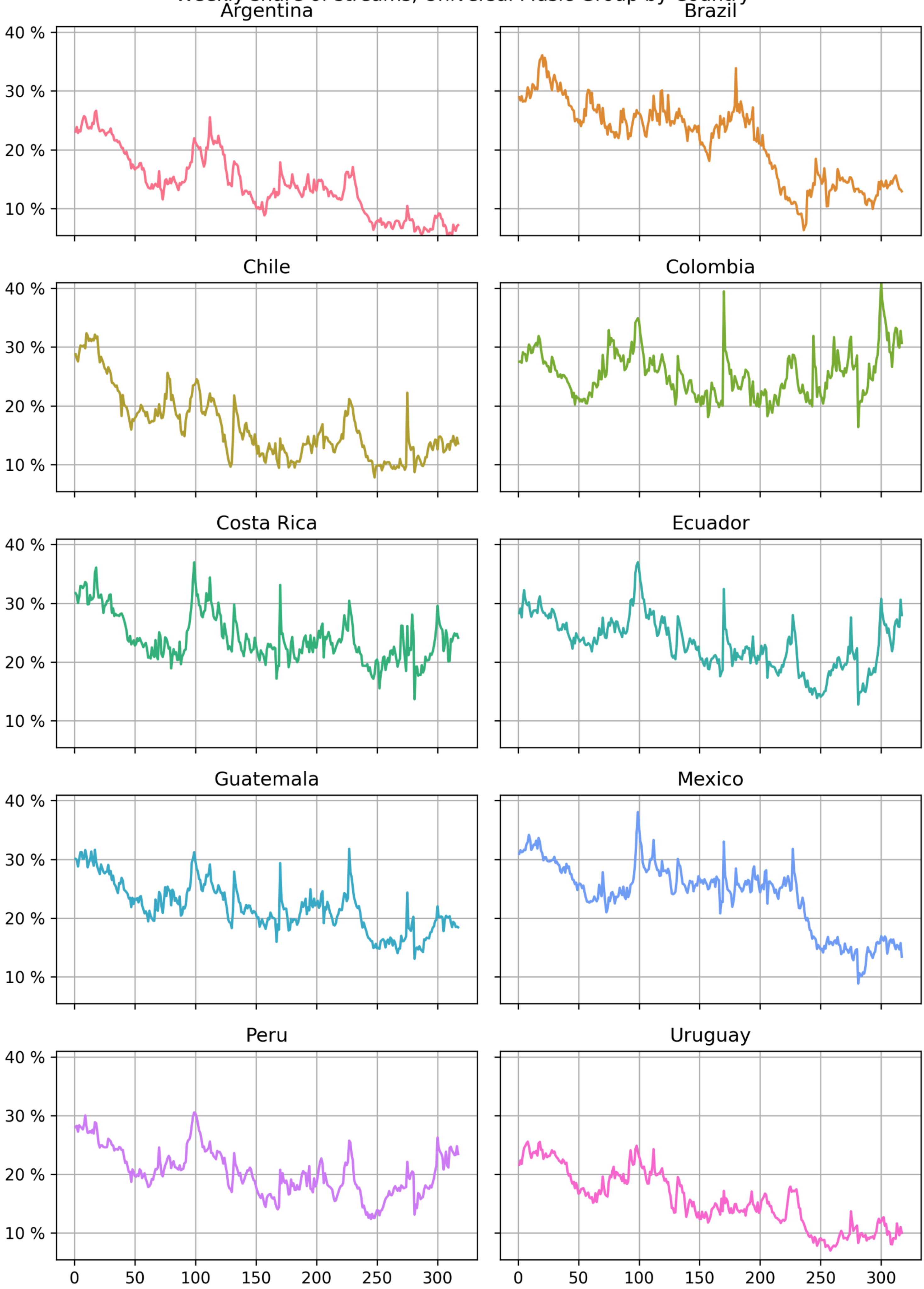
Weekly share of streams, major labels by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
80 %
70 %
60 %
50 %
40 %
30 %
0
50
100
150
200
250
300

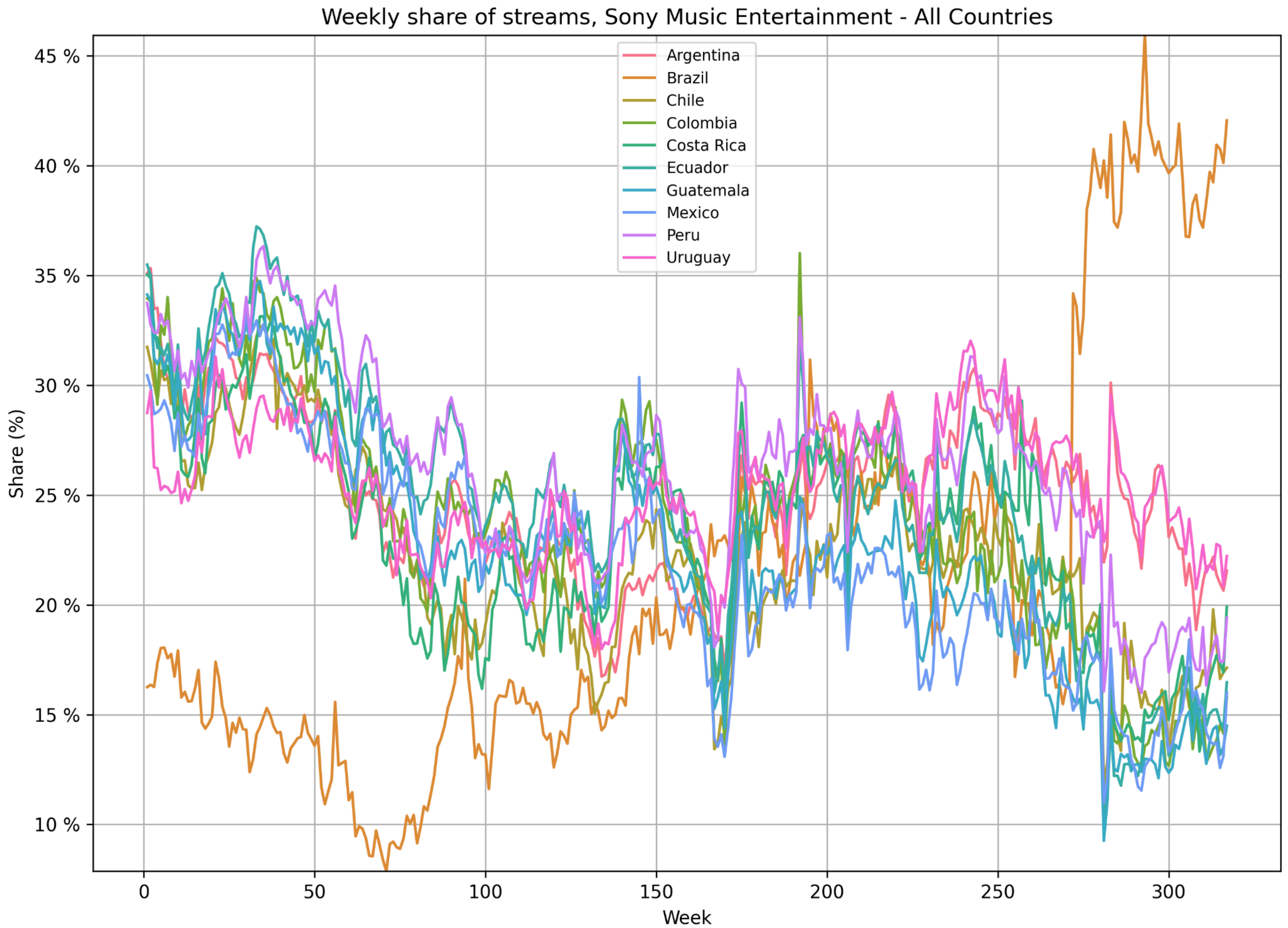

Weekly share of streams, Universal Music Group - All Countries
Share (%)
40 %
35 %
30 %
25 %
20 %
15 %
10 %
0
50
100
150
200
250
300
Week
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay



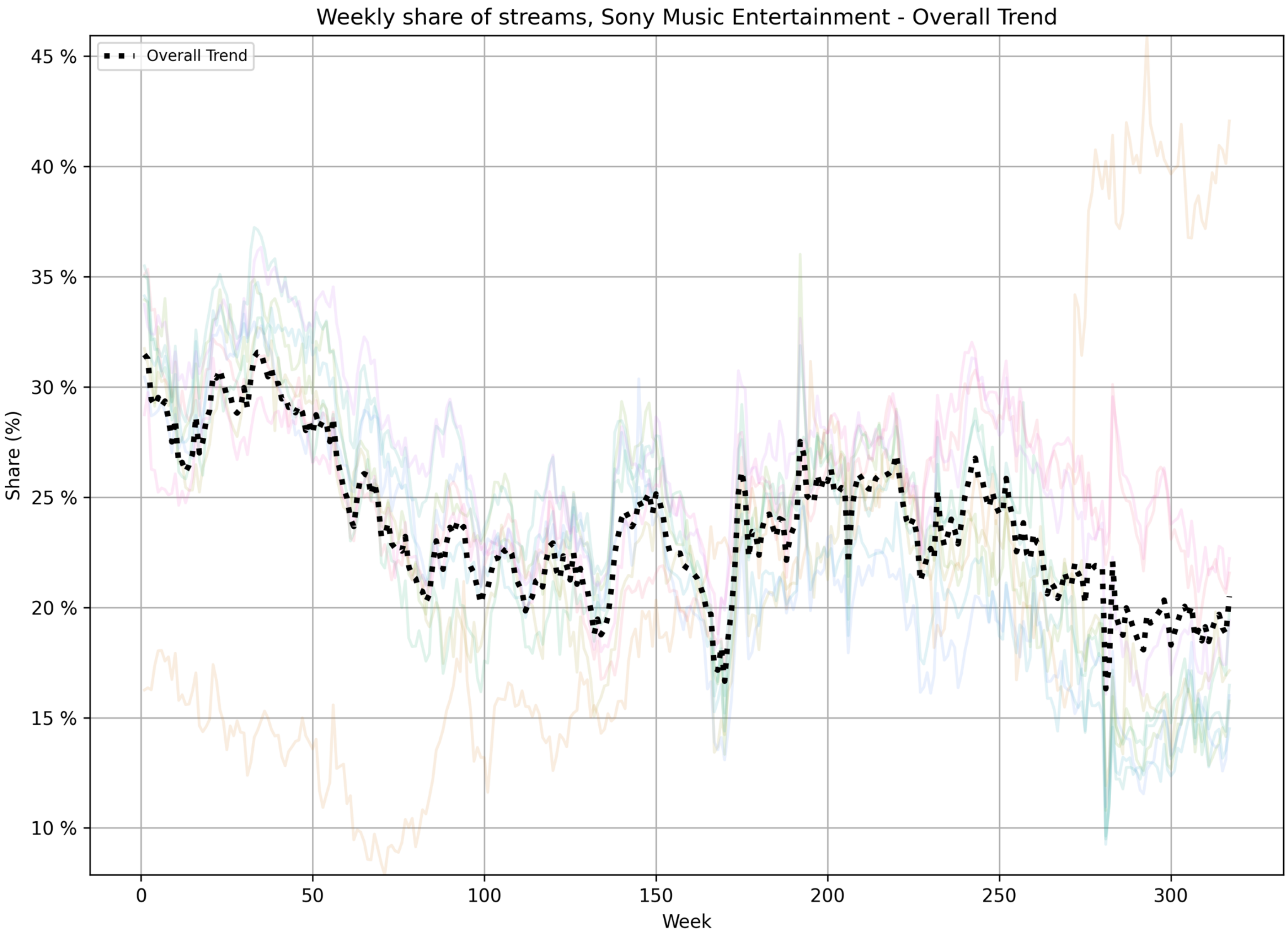
Weekly share of streams, Universal Music Group - Overall Trend
Overall Trend
40 %
35 %
30 %
25 %
20 %
15 %
10 %
Share (%)
0
50
100
150
200
250
300
Week

Weekly share of streams, Universal Music Group by Country

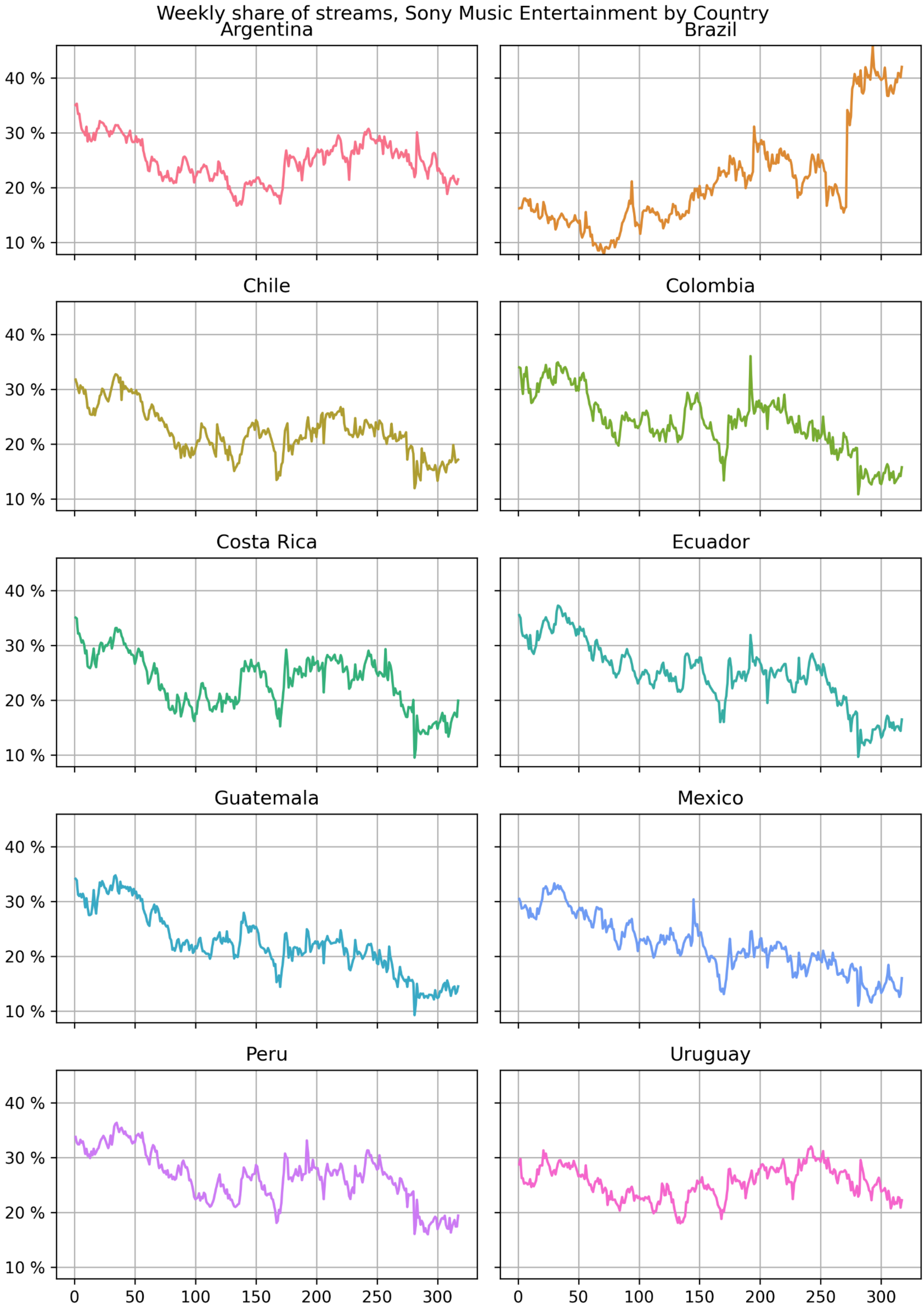




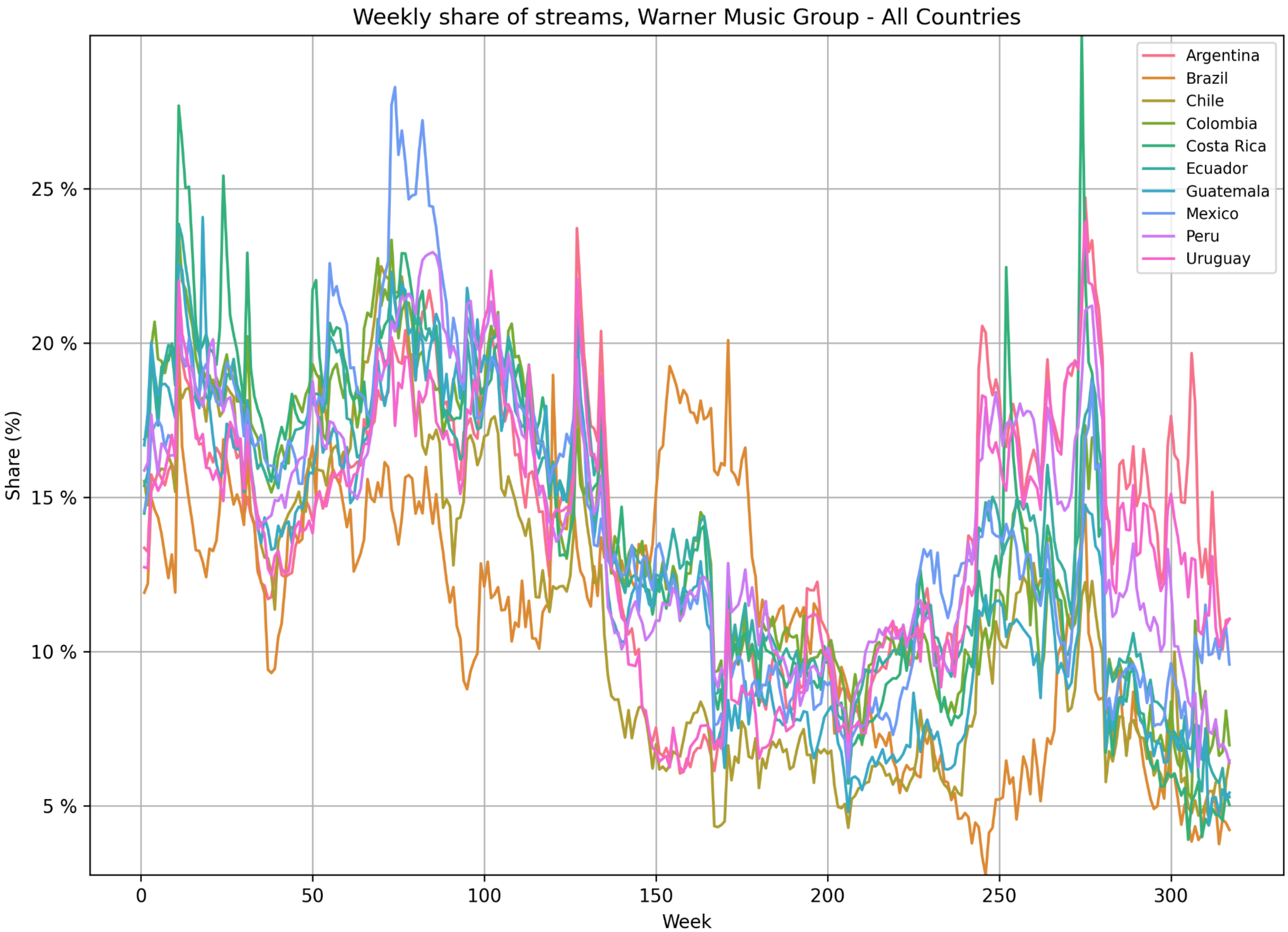
Weekly share of streams, Sony Music Entertainment - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
45 %
40 %
35 %
30 %
25 %
20 %
15 %
10 %
Share (%)
0
50
100
150
200
250
300
Week

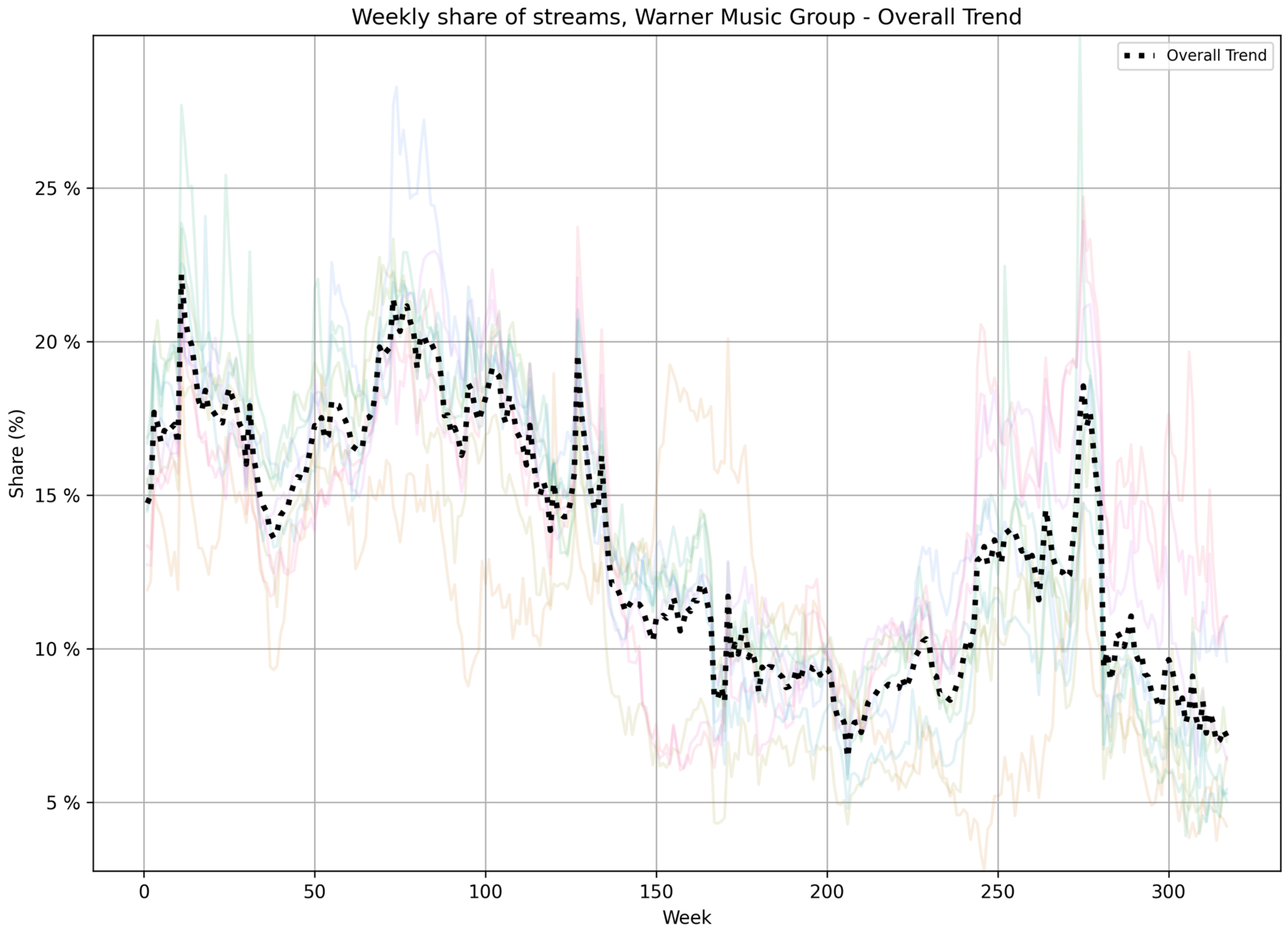


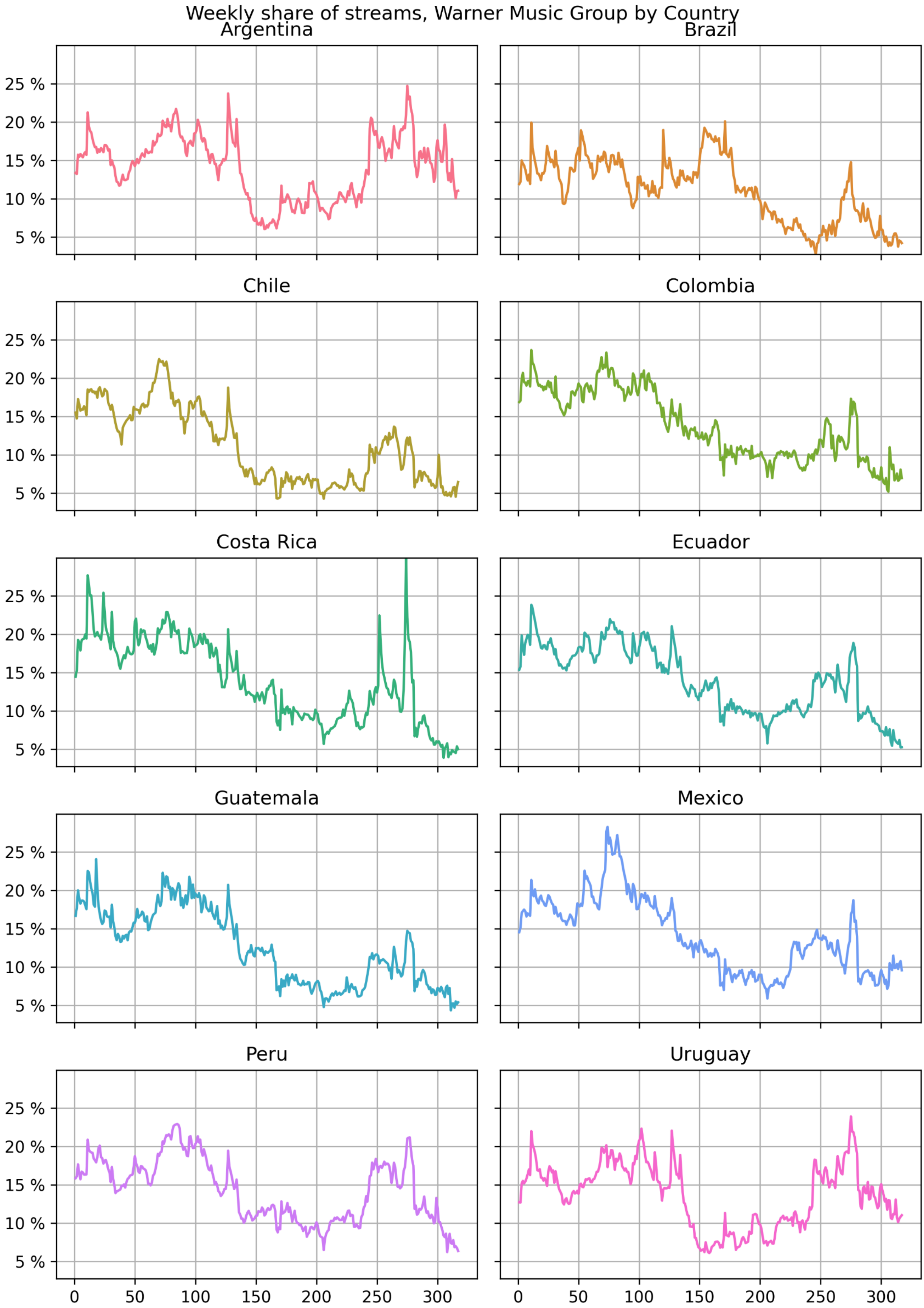
Weekly share of streams, Sony Music Entertainment - Overall Trend
Overall Trend
45 %
40 %
35 %
30 %
25 %
20 %
15 %
10 %
Share (%)
0
50
100
150
200
250
300
Week



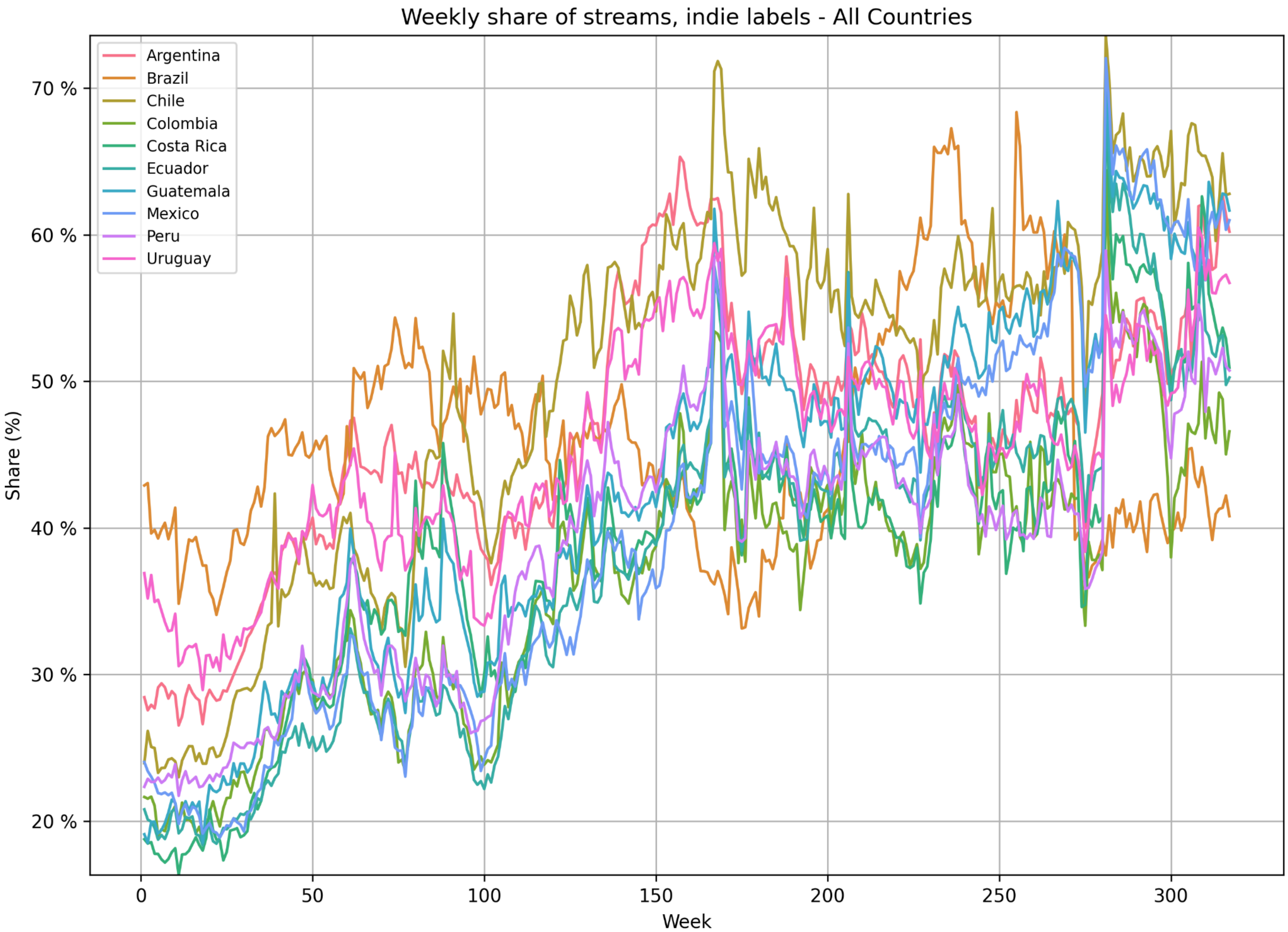
Weekly share of streams, Sony Music Entertainment by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
10 %
20 %
30 %
40 %
0
50
100
150
200
250
300

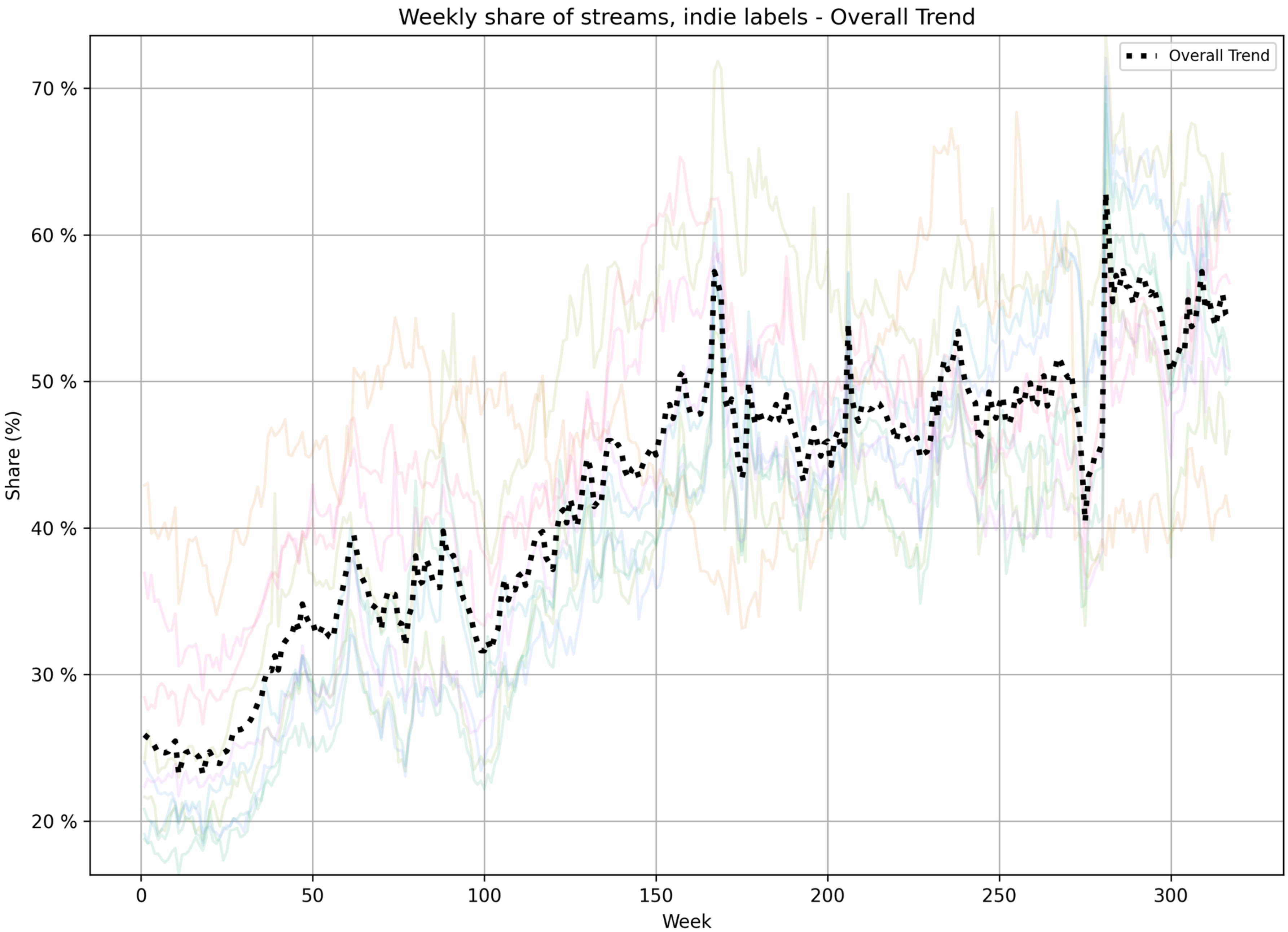

Weekly share of streams, Warner Music Group - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
Share (%)
25 %
20 %
15 %
10 %
5 %
0
50
100
150
200
250
300
Week


Weekly share of streams, Warner Music Group - Overall Trend
Overall Trend
Share (%)
25 %
20 %
15 %
10 %
5 %
Week
0
50
100
150
200
250
300

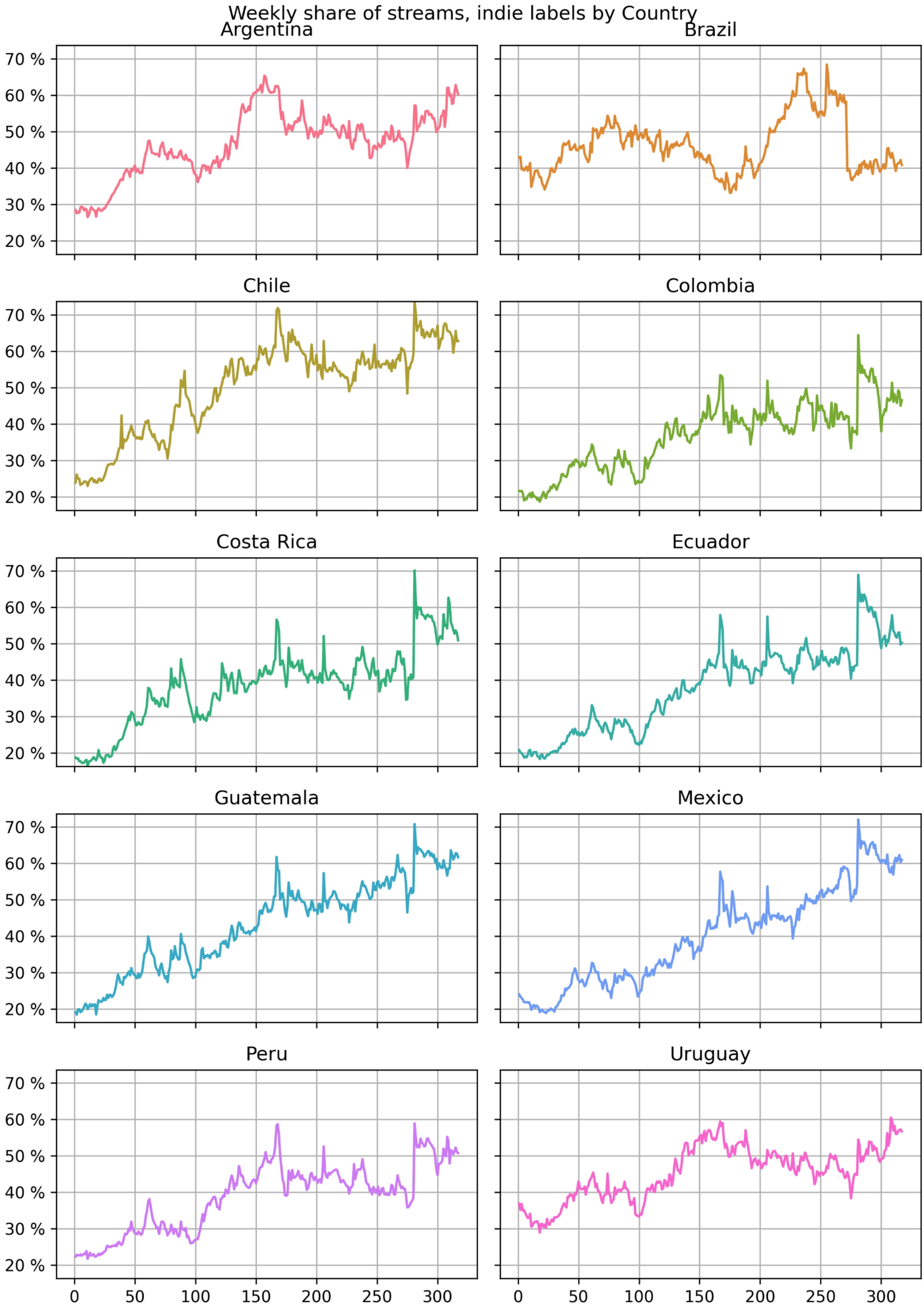


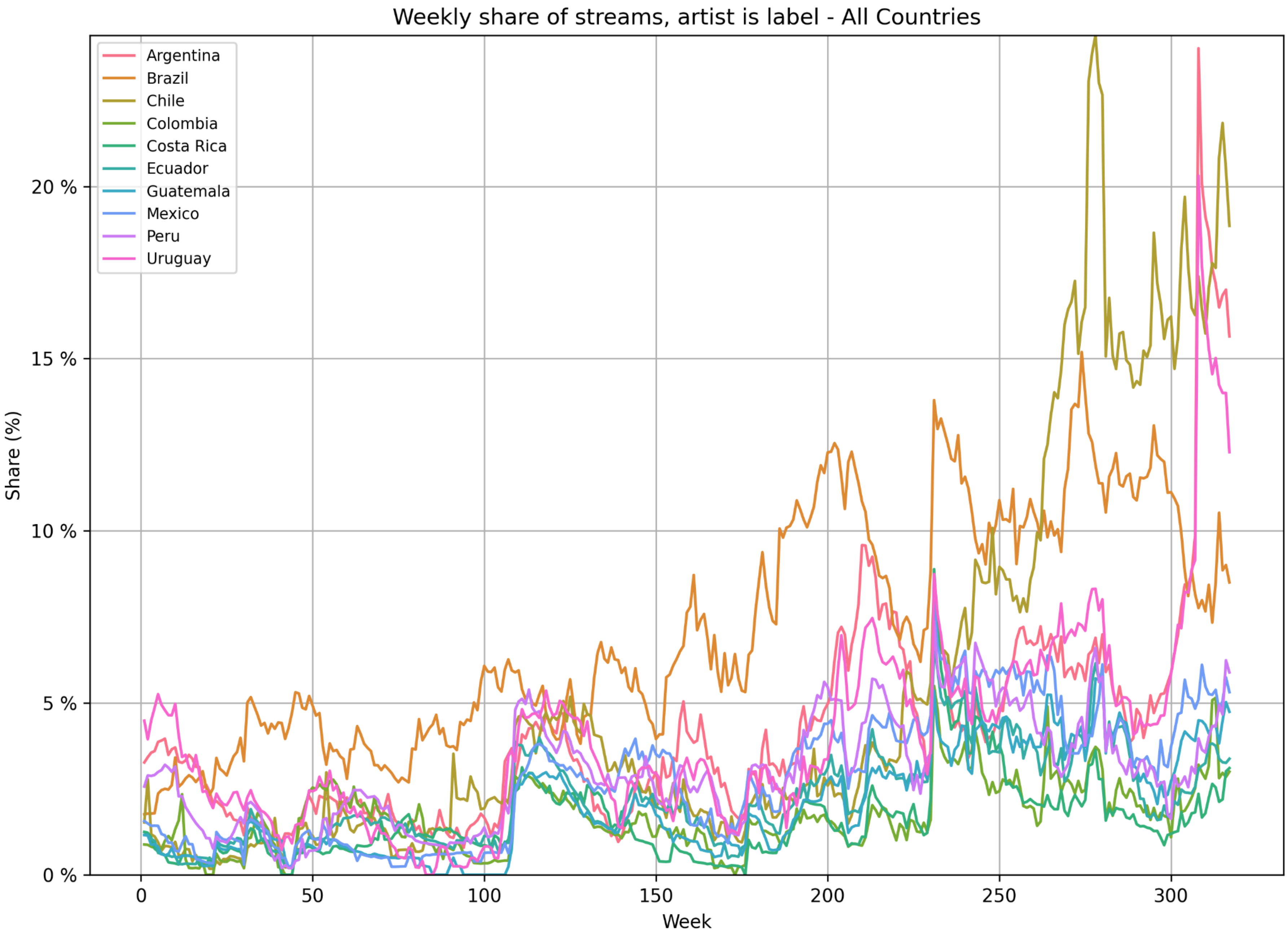
Weekly share of streams, Warner Music Group by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
25 %
20 %
15 %
10 %
5 %
0
50
100
150
200
250
300


Weekly share of streams, indie labels - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
70 %
60 %
50 %
40 %
30 %
20 %
Share (%)
0
50
100
150
200
250
300
Week

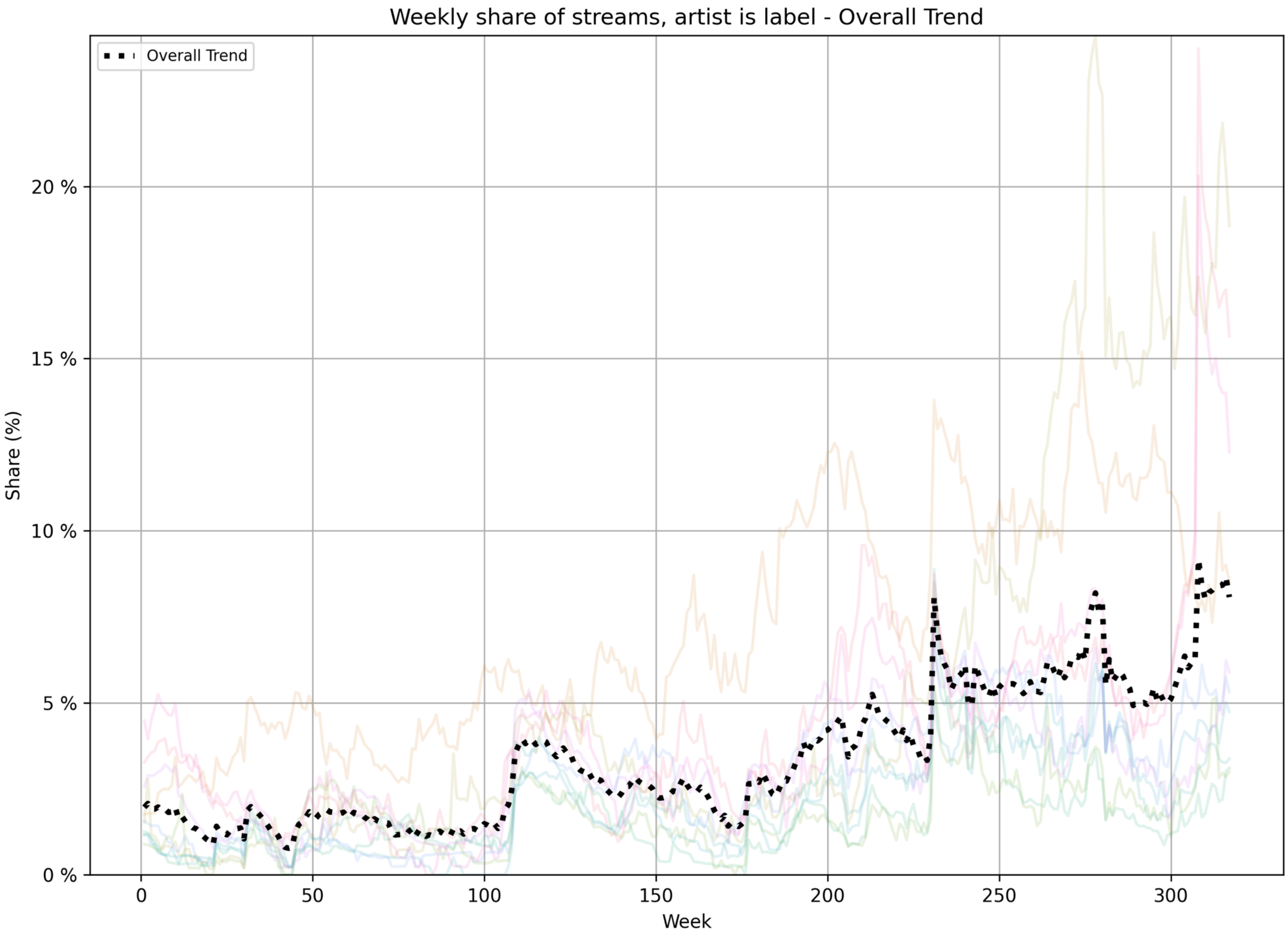


Weekly share of streams, indie labels - Overall Trend
Overall Trend
Share (%)
70 %
60 %
50 %
40 %
30 %
20 %
Week
0
50
100
150
200
250
300

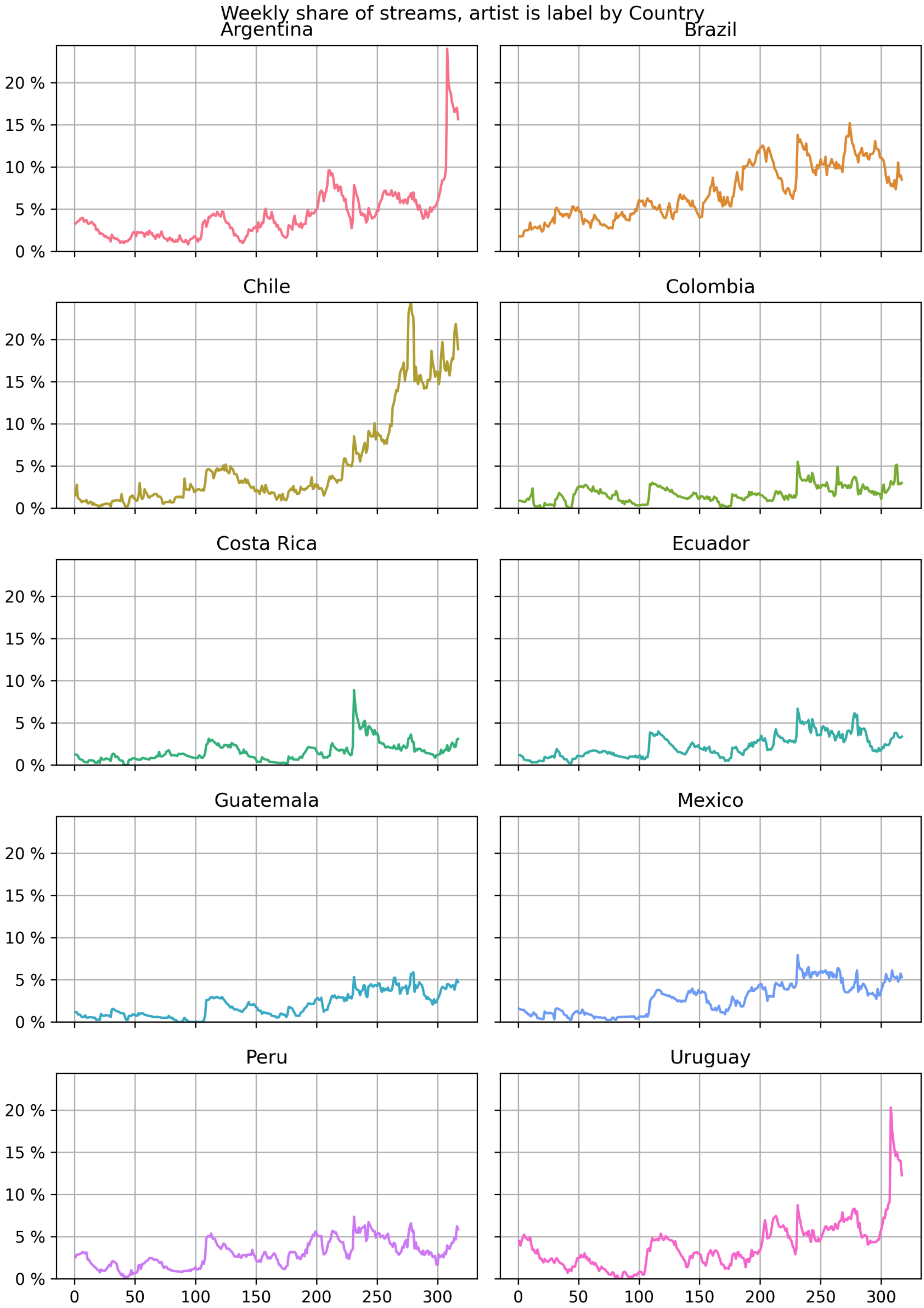

Weekly share of streams, indie labels by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
70 %
60 %
50 %
40 %
30 %
20 %
0
50
100
150
200
250
300



Weekly share of streams, artist is label - All Countries
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
Share (%)
Week
0 %
5 %
10 %
15 %
20 %
0
50
100
150
200
250
300



Weekly share of streams, artist is label - Overall Trend
Overall Trend
Share (%)
20 %
15 %
10 %
5 %
0 %
0
50
100
150
200
250
300
Week


Weekly share of streams, artist is label by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
0 %
5 %
10 %
15 %
20 %
0
50
100
150
200
250
300

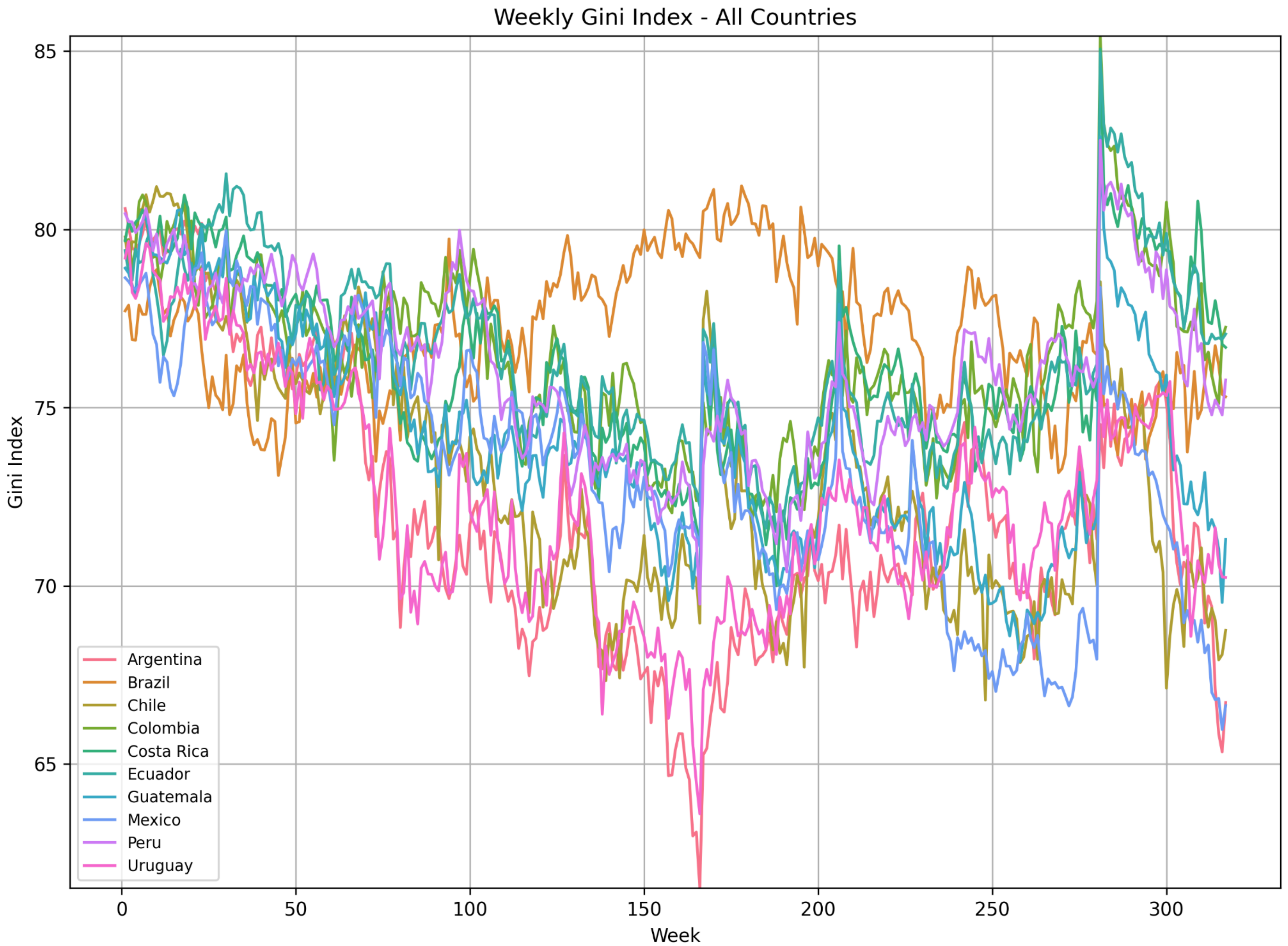
14.3. Appendix C – Gini Index: tables and graphs

Table C1. Gini Index, all countries

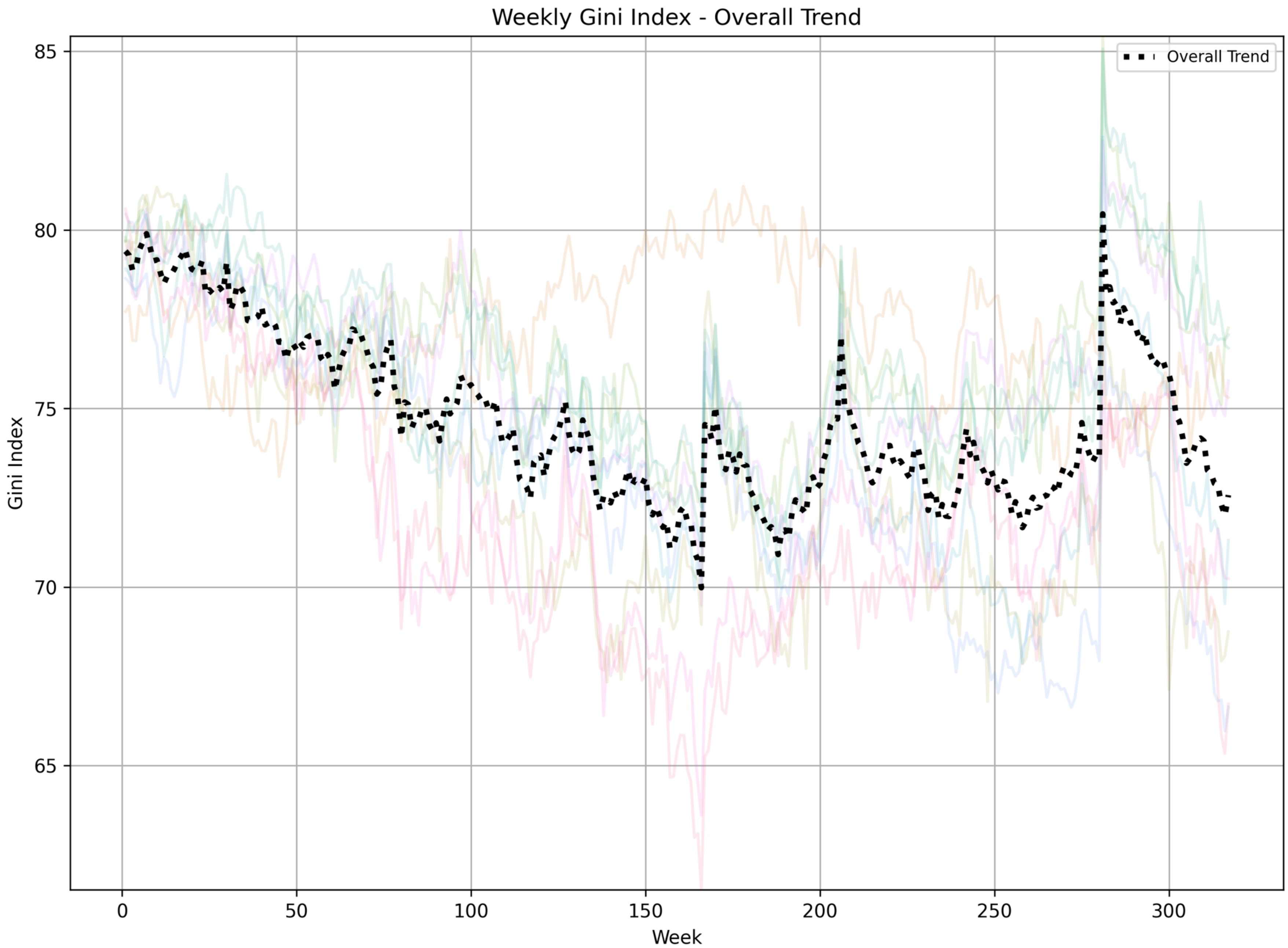
Mean	74.68
Median	74.97
Std. Deviation	3.48
Range	23.91
Minimum	61.53
Maximum	85.44

Table C2. Gini Index, by country

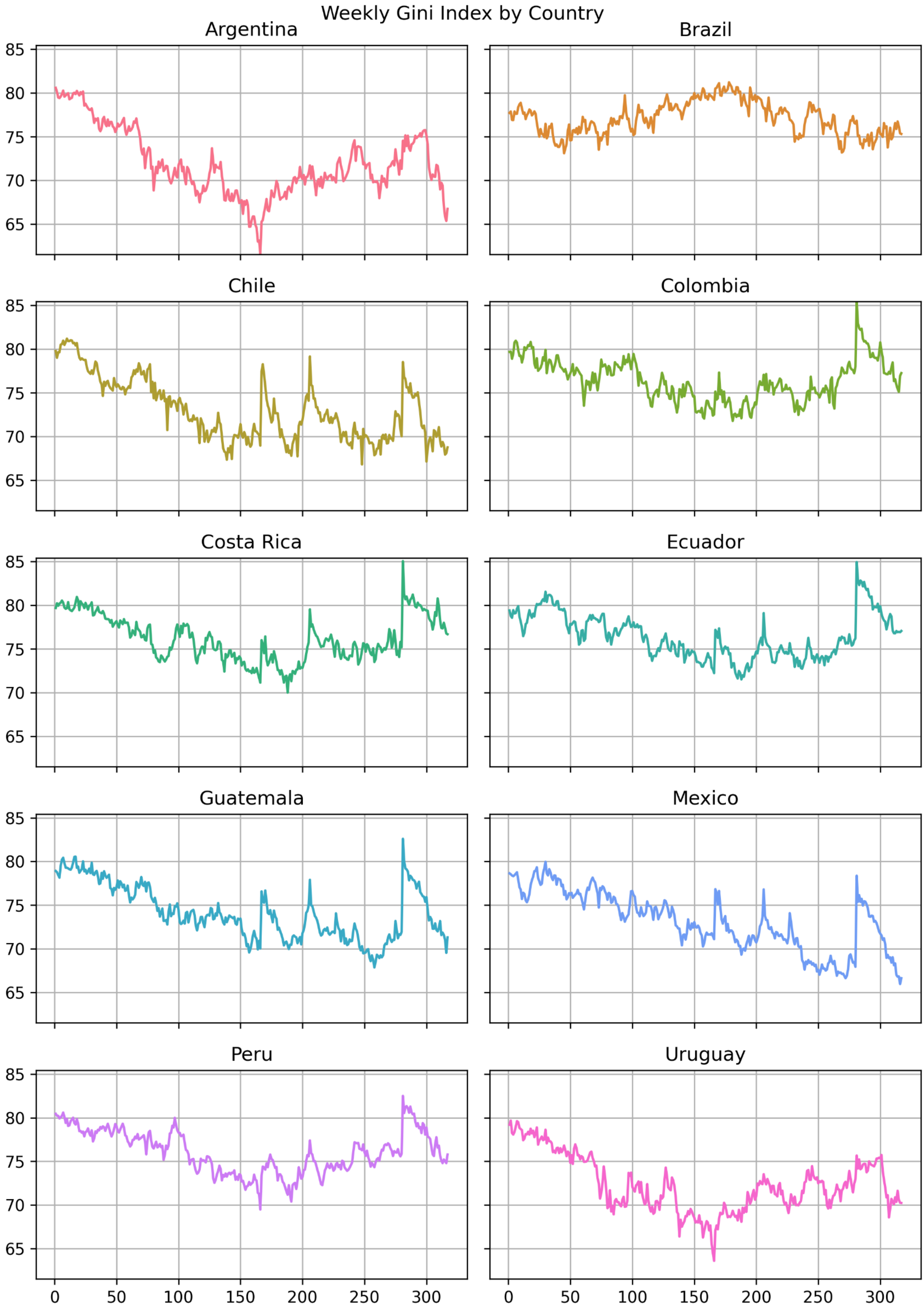
	Argentina	Brazil	Chile	Colombia	Costa Rica	Ecuador	Guatemala	Mexico	Peru	Uruguay
Mean	71.95	77.09	72.99	76.43	76.20	76.45	74.15	73.17	76.09	72.34
Median	71.31	77.00	72.19	76.29	75.90	76.15	73.65	73.32	75.88	71.83
Std. Deviation	3.83	1.88	3.54	2.36	2.57	2.55	3.08	3.37	2.46	3.18
Range	19.06	8.13	14.41	13.65	15.06	13.44	14.73	13.99	13.02	16.06
Minimum	61.53	73.10	66.80	71.78	70.00	71.52	67.87	65.97	69.49	63.61
Maximum	80.59	81.22	81.20	85.44	85.06	84.96	82.60	79.97	82.50	79.67



Weekly Gini Index - All Countries
85
80
75
70
65
Gini Index
0
50
100
150
200
250
300
Week
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay



Weekly Gini Index - Overall Trend
Overall Trend
Gini Index
Week
85
80
75
70
65
0
50
100
150
200
250
300



Weekly Gini Index by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
85
80
75
70
65
0
50
100
150
200
250
300

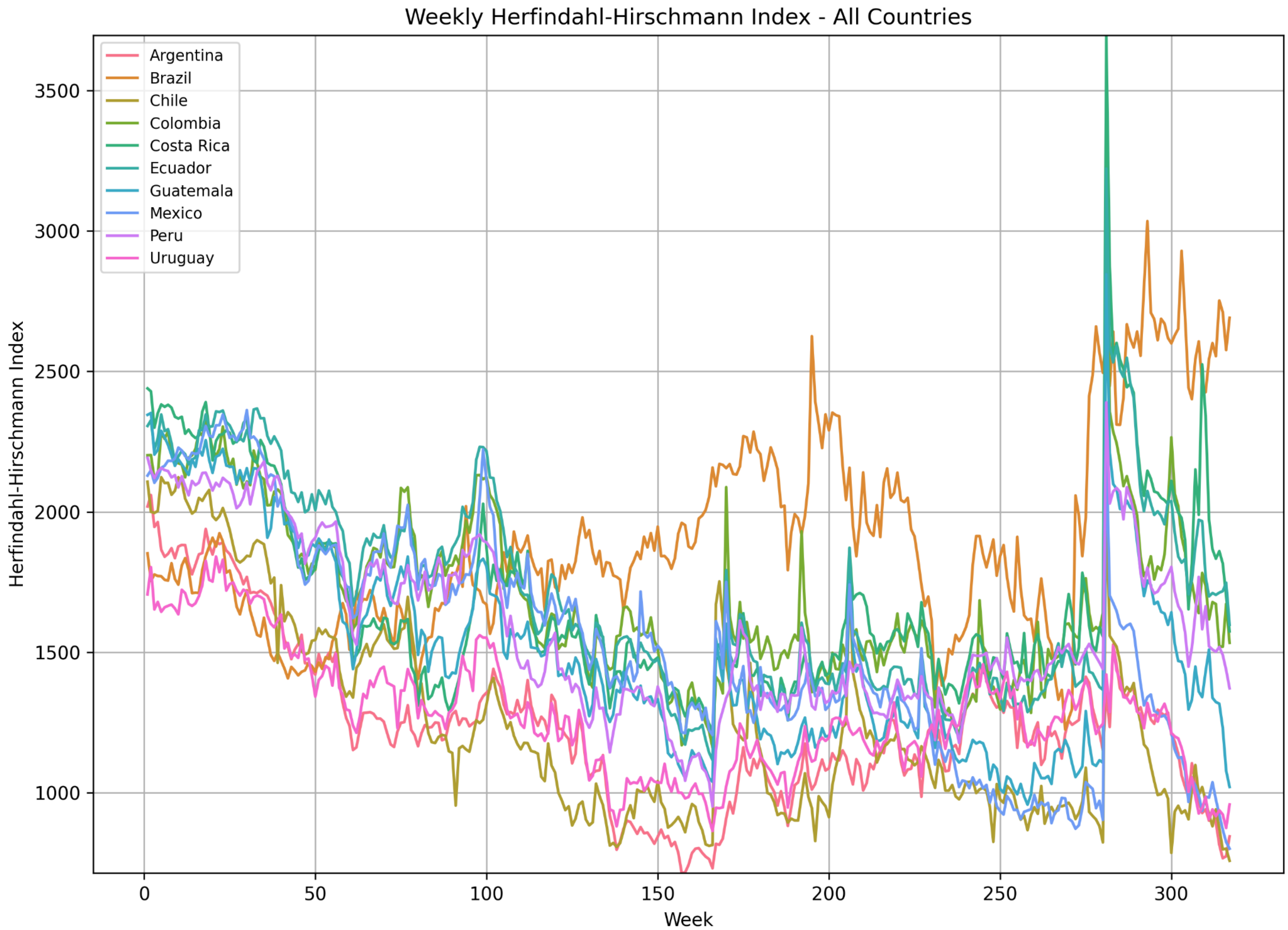
14.4. Appendix D – Herfindahl-Hirschman Index: tables and graphs

Table D1. Herfindahl-Hirschmann Index , all countries

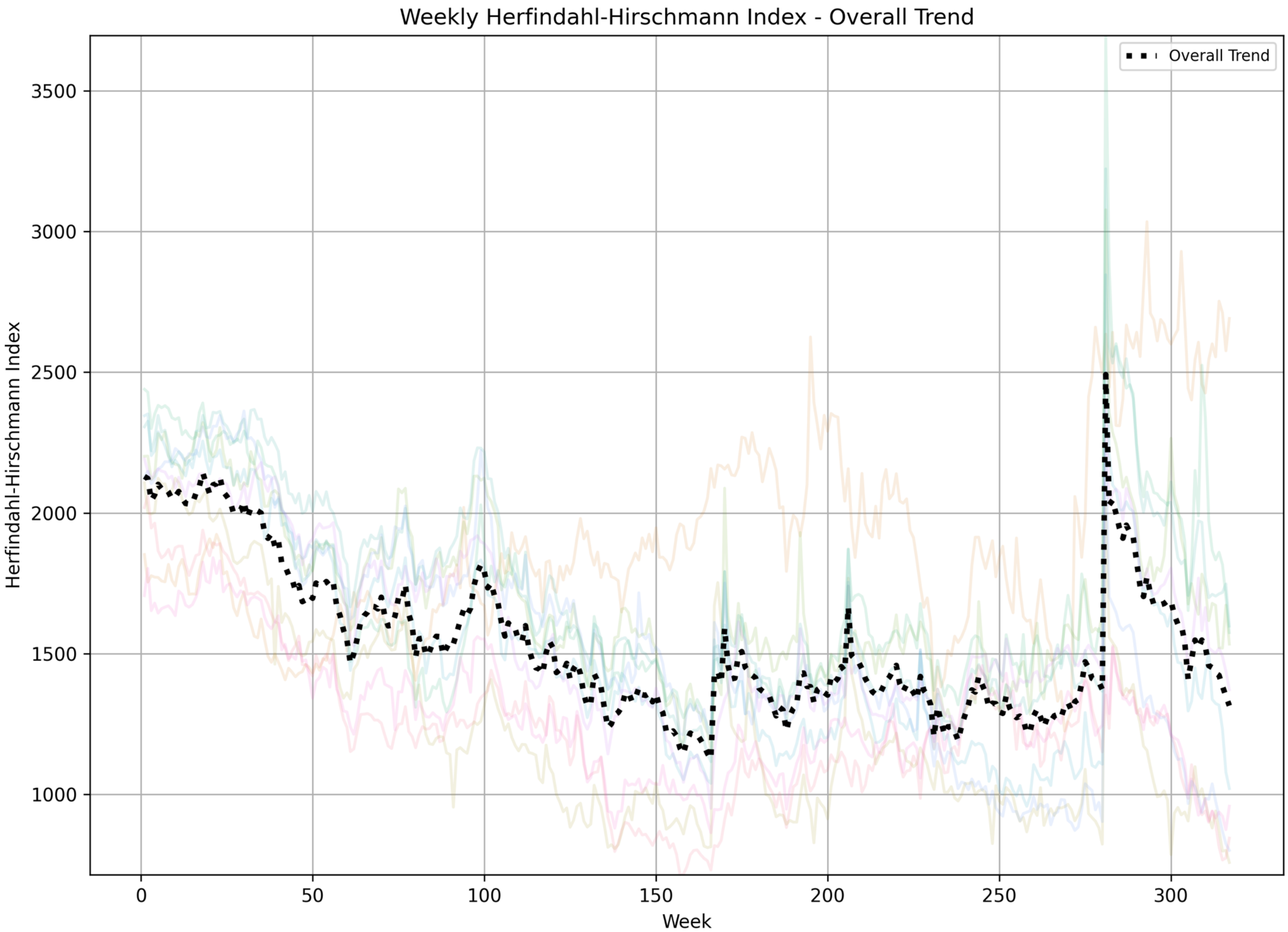
Mean	1,542.82
Median	1,485.53
Std. Deviation	397.85
Range	2,984.04
Minimum	713.63
Maximum	3,697.68

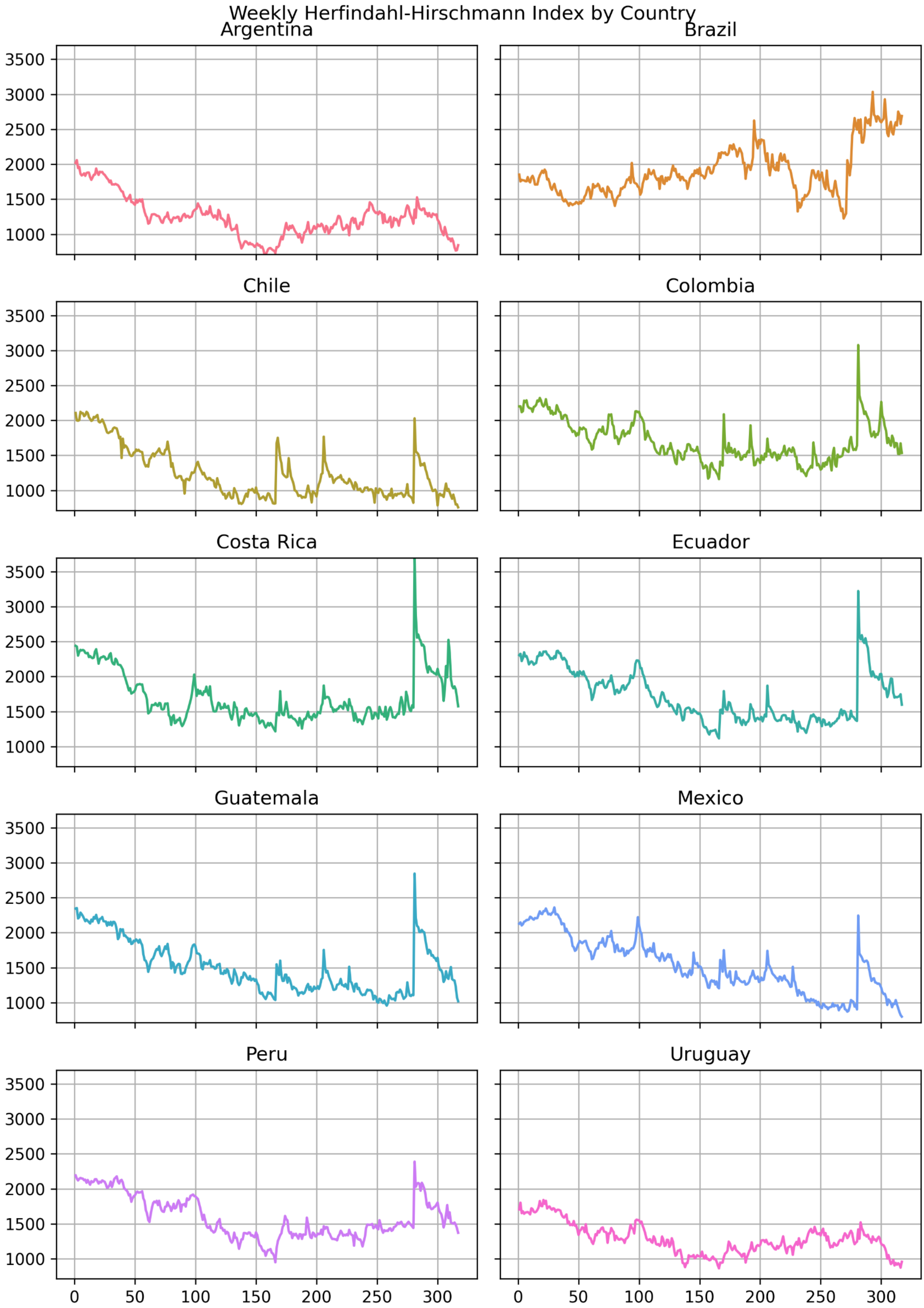
Table D2. Herfindahl-Hirschmann Index, by country

	Argentina	Brazil	Chile	Colombia	Costa Rica	Ecuador	Guatemala	Mexico	Peru	Uruguay
Mean	1,242.52	1,902.48	1,238.54	1,705.84	1,703.15	1,724.95	1,505.25	1,526.04	1,594.28	1,285.10
Median	1,225.65	1,828.21	1,141.37	1,637.10	1,567.93	1,635.40	1,422.09	1,475.64	1,490.73	1,266.39
Std. Deviation	284.24	353.87	355.73	294.87	355.77	367.59	365.34	400.89	302.77	217.87
Range	1,346.09	1,810.81	1,365.69	1,917.18	2,481.20	2,107.21	1,889.29	1,561.83	1,439.52	977.73
Minimum	713.63	1,224.38	757.27	1,160.17	1,216.47	1,115.50	956.94	800.20	950.05	862.92
Maximum	2,059.72	3,035.19	2,122.96	3,077.35	3,697.68	3,222.71	2,846.23	2,362.03	2,389.57	1,840.65



Weekly Herfindahl-Hirschman Index - All Countries
Herfindahl-Hirschman Index
Week
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
3500
3000
2500
2000
1500
1000
0
50
100
150
200
250
300


Weekly Herfindahl-Hirschmann Index - Overall Trend
Overall Trend
Herfindahl-Hirschmann Index
3500
3000
2500
2000
1500
1000
0
50
100
150
200
250
300
Week



Weekly Herfindahl-Hirschmann Index by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
3500
3000
2500
2000
1500
1000
0
50
100
150
200
250
300

14.5. Appendix E – Spotify Audio Features: tables and graphs

Table E1. Descriptive statistics – Spotify Audio Features – All countries

		key	danceability	energy	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_s	time_signature
N	Valid	633,874	633,874	633,874	633,874	633,874	633,874	633,874	633,874	633,874	633,874	633,874	633,874	633,874
	Missing	126	126	126	126	126	126	126	126	126	126	126	126	126
Mean		5.32	71.89	68.70	-5.28	.58	10.75	24.77	.40	17.78	61.97	123.28	213.93	3.96
Median		6.00	73.80	70.90	-4.93	1.00	7.25	18.40	.00	12.00	64.80	113.06	207.61	4.00
Mode		1	74.40	77.30	-6.33	1	4.32	17.60	.00	10.10	68.00	104.82	205.72	4
Std. Deviation		3.67	11.41	14.62	2.06	.49	8.88	21.44	3.57	15.09	20.93	33.31	48.68	.26
Range		11	90.67	97.73	24.93	1	86.08	99.20	99.00	97.66	95.70	165.27	3,620.09	4
Minimum		0	7.83	2.17	-23.02	0	2.32	.00	.00	1.34	3.20	48.75	33.87	1
Maximum		11	98.50	99.90	1.91	1	88.40	99.20	99.00	99.00	98.90	214.03	3,653.96	5

Table E2. Descriptive statistics – Spotify Audio Features – All countries

			country_name									
			Argentina	Brazil	Chile	Colombia	Costa Rica	Ecuador	Guatemala	Mexico	Peru	Uruguay
key	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		5.30	5.41	5.35	5.26	5.28	5.27	5.34	5.35	5.31	5.35
	Median		5.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
	Mode		1	1	1	1	1	1	1	0	1	1
	Std. Deviation		3.71	3.64	3.68	3.65	3.67	3.67	3.66	3.65	3.68	3.70
	Range		11	11	11	11	11	11	11	11	11	11
	Minimum		0	0	0	0	0	0	0	0	0	0
	Maximum		11	11	11	11	11	11	11	11	11	11
danceability	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		71.87	68.30	73.61	72.48	71.63	72.81	72.10	70.79	72.79	72.55
	Median		73.90	68.20	75.30	74.30	73.90	74.40	73.90	72.80	74.40	74.40

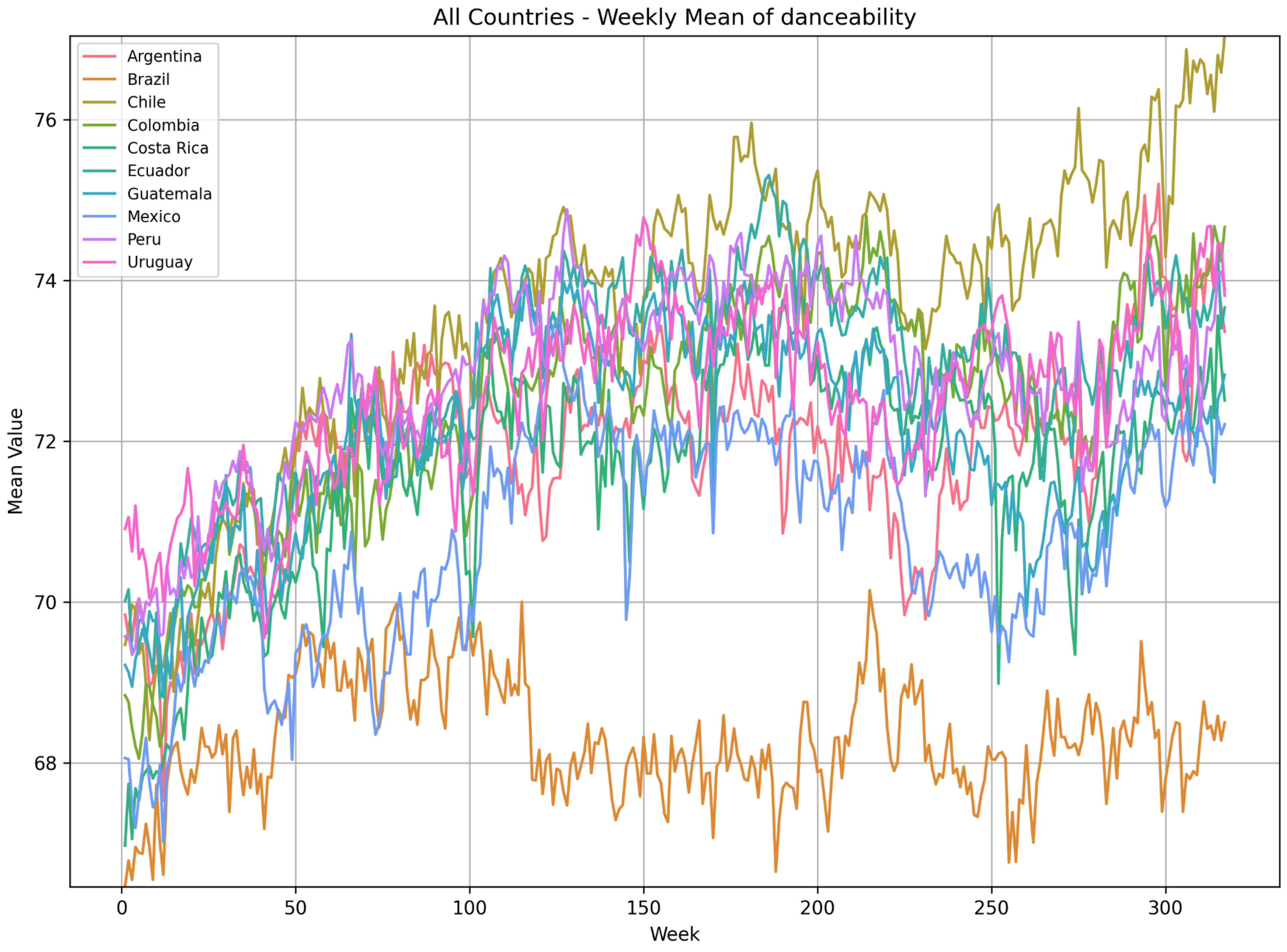
	Mode		79.50	67.60	74.40	71.40	74.40	74.40	74.40	74.40	76.00	79.50
	Std. Deviation		11.61	12.80	10.86	10.92	12.03	10.78	10.79	11.74	10.53	11.00
	Range		76.20	80.10	75.60	89.34	88.57	74.90	88.87	88.87	75.50	72.80
	Minimum		20.90	18.40	20.90	8.66	7.83	21.50	7.83	7.83	20.90	23.80
	Maximum		97.10	98.50	96.50	98.00	96.40	96.40	96.70	96.70	96.40	96.60
energy	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		68.78	70.23	70.09	69.28	66.68	69.36	66.94	65.06	70.88	69.70
	Median		71.00	72.50	71.50	71.20	68.80	71.50	69.40	67.20	72.70	71.50
	Mode		80.00	74.30	71.20	77.30	77.30	77.30	77.30	64.60	77.30	74.50
	Std. Deviation		14.14	16.15	13.26	14.04	14.81	14.59	15.04	15.61	13.80	13.43
	Range		88.50	97.73	88.50	90.48	96.24	94.77	88.50	94.77	88.30	88.50
	Minimum		10.40	2.17	10.40	8.42	3.16	4.13	10.40	4.13	11.10	10.40
	Maximum		98.90	99.90	98.90	98.90	99.40	98.90	98.90	98.90	99.40	98.90
loudness	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		-5.36	-5.23	-5.18	-5.20	-5.37	-5.24	-5.31	-5.65	-5.12	-5.15
	Median		-4.96	-4.96	-4.83	-4.89	-5.04	-4.83	-4.96	-5.27	-4.78	-4.82
	Mode		-6.66	-5.46	-3.76	-7.12	-6.33	-7.12	-7.12	-3.68	-7.12	-6.66
	Std. Deviation		2.19	2.27	1.92	1.91	2.03	2.05	2.04	2.22	1.98	1.93
	Range		19.63	24.59	18.51	22.14	23.20	19.37	20.79	20.79	16.75	18.47
	Minimum		-18.99	-22.69	-17.87	-22.21	-23.02	-19.45	-20.61	-20.61	-17.24	-17.83
	Maximum		.64	1.91	.64	-.07	.18	-.07	.18	.18	-.48	.64
mode	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		.56	.58	.55	.59	.59	.58	.60	.61	.59	.55
	Median		1	1	1	1	1	1	1	1	1	1
	Mode		1	1	1	1	1	1	1	1	1	1
	Std. Deviation		.50	.49	.50	.49	.49	.49	.49	.49	.49	.50
	Range		1	1	1	1	1	1	1	1	1	1
	Minimum		0	0	0	0	0	0	0	0	0	0
	Maximum		1	1	1	1	1	1	1	1	1	1
speechiness	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377

		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		10.73	10.79	11.60	11.01	10.98	10.97	10.39	9.37	10.63	11.03
	Median		7.29	6.76	8.25	7.53	7.35	7.46	7.01	6.09	7.35	7.60
	Mode		2.72	12.80	10.00	10.00	4.32	3.83	4.32	3.52	6.75	11.10
	Std. Deviation		8.98	10.07	8.80	8.67	9.13	8.71	8.54	8.29	8.46	8.82
	Range		69.38	85.58	69.38	86.08	69.48	86.08	86.08	69.38	86.08	69.38
	Minimum		2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32
	Maximum		71.70	87.90	71.70	88.40	71.80	88.40	88.40	71.70	88.40	71.70
acousticness	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		22.64	33.26	23.16	23.44	24.63	23.80	25.33	26.65	22.97	21.82
	Median		17.20	30.50	17.60	17.40	17.60	17.40	18.80	20.10	17.20	16.70
	Mode		21.00	42.70	12.20	17.60	17.60	17.60	17.60	17.60	17.60	18.60
	Std. Deviation		19.82	23.62	19.82	21.04	22.01	21.71	21.67	22.48	20.63	18.91
	Range		98.39	98.60	98.50	98.50	98.50	99.20	99.18	99.18	98.50	98.38
	Minimum		.01	.00	.00	.00	.00	.00	.02	.02	.00	.02
	Maximum		98.40	98.60	98.50	98.50	98.50	99.20	99.20	99.20	98.50	98.40
instrumentalness	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		.39	.49	.45	.38	.43	.43	.36	.51	.29	.25
	Median		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	Mode		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	Std. Deviation		3.13	4.60	3.80	3.33	3.57	3.81	3.45	4.02	2.95	2.62
	Range		91.50	89.30	91.50	89.30	99.00	82.80	90.10	92.10	91.00	82.80
	Minimum		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	Maximum		91.50	89.30	91.50	89.30	99.00	82.80	90.10	92.10	91.00	82.80
liveness	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		17.17	29.49	16.61	16.17	16.01	16.11	16.31	16.89	16.30	16.71
	Median		12.20	16.00	11.70	11.70	11.60	11.60	11.80	11.80	11.80	11.90
	Mode		10.80	10.10	10.30	10.10	10.10	10.10	10.10	10.60	10.10	10.30
	Std. Deviation		13.37	27.47	12.52	11.51	11.48	11.53	11.73	13.60	11.99	12.99
	Range		95.99	97.66	95.65	95.53	95.85	95.53	94.55	96.10	93.43	95.59

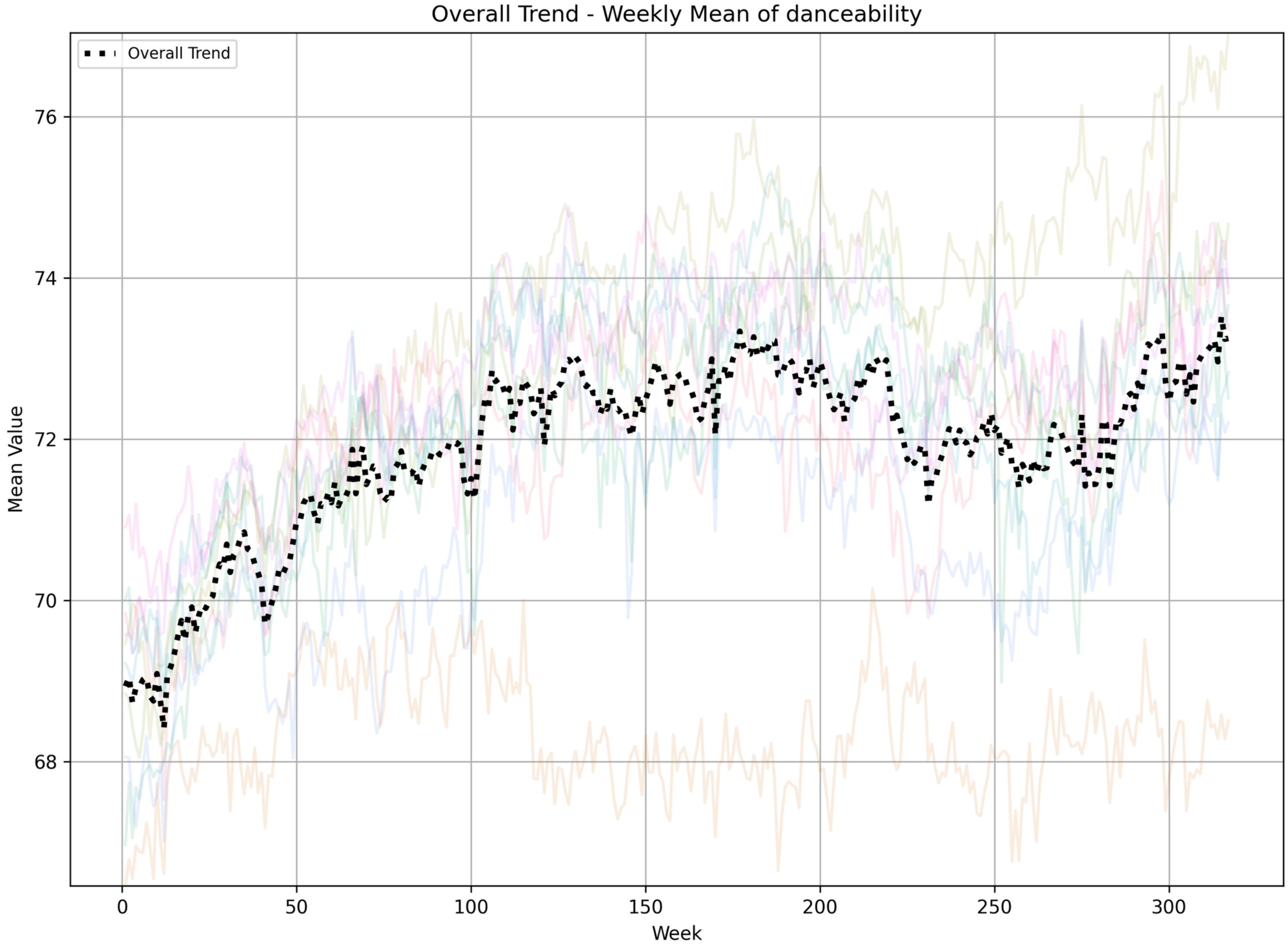
	Minimum		1.81	1.34	2.15	2.07	2.15	2.07	2.15	1.90	2.07	1.81
	Maximum		97.80	99.00	97.80	97.60	98.00	97.60	96.70	98.00	95.50	97.40
valence	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		61.82	62.70	61.41	62.72	59.12	61.91	61.22	62.10	62.99	63.72
	Median		64.60	65.20	63.50	65.60	61.80	65.10	64.00	64.60	66.20	66.40
	Mode		70.60	96.30	76.10	68.00	44.60	68.00	68.00	90.90	68.00	70.60
	Std. Deviation		20.54	21.00	20.48	20.86	21.48	20.98	21.24	21.59	20.66	20.10
	Range		94.70	94.30	94.80	95.70	94.10	94.20	94.40	94.40	94.40	94.70
	Minimum		3.20	3.20	3.20	3.20	3.20	3.20	3.20	3.20	3.20	3.20
	Maximum		97.90	97.50	98.00	98.90	97.30	97.40	97.60	97.60	97.60	97.90
tempo	N	Valid	63.375	63.385	63.393	63.391	63.393	63.393	63.393	63.382	63.392	63.377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		122.74	128.34	122.13	122.95	122.89	122.43	123.07	123.02	122.84	122.35
	Median		112.86	129.03	105.03	106.20	112.95	105.11	112.14	116.23	106.03	108.29
	Mode		118.00	122.00	104.82	104.82	104.82	102.79[a]	123.95	123.95	102.79	109.33
	Std. Deviation		33.30	27.52	34.51	34.87	33.32	34.38	33.74	31.90	34.22	34.29
	Range		163.69	161.42	151.54	151.54	149.67	151.51	151.54	157.37	151.51	163.69
	Minimum		48.75	50.70	62.48	62.48	62.45	62.52	62.48	54.75	62.48	48.75
	Maximum		212.44	212.12	214.03	214.03	212.12	214.03	214.03	212.12	213.99	212.44
duration_s	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		213.58	193.34	220.70	216.76	217.70	215.83	217.31	213.83	220.32	209.90
	Median		206.50	181.86	212.44	211.63	211.63	210.61	211.63	209.76	213.03	203.67
	Mode		260.37	204.35	205.72	205.72	205.72	205.72[a]	205.72	191.97	205.72	262.00
	Std. Deviation		52.99	62.51	48.82	43.51	45.76	41.35	42.98	40.40	49.17	49.42
	Range		938.16	707.09	564.51	579.16	987.39	571.57	891.15	571.57	3558.49	890.63
	Minimum		61.52	37.64	48.52	33.87	37.64	41.46	48.52	41.46	95.47	37.64
	Maximum		999.68	744.73	613.03	613.03	1025.03	613.03	939.67	613.03	3653.96	928.27
time_signature	N	Valid	63,375	63,385	63,393	63,391	63,393	63,393	63,393	63,382	63,392	63,377
		Missing	25	15	7	9	7	7	7	18	8	23
	Mean		3.97	4.00	3.98	3.98	3.97	3.98	3.94	3.88	3.98	3.97
	Median		4	4	4	4	4	4	4	4	4	4

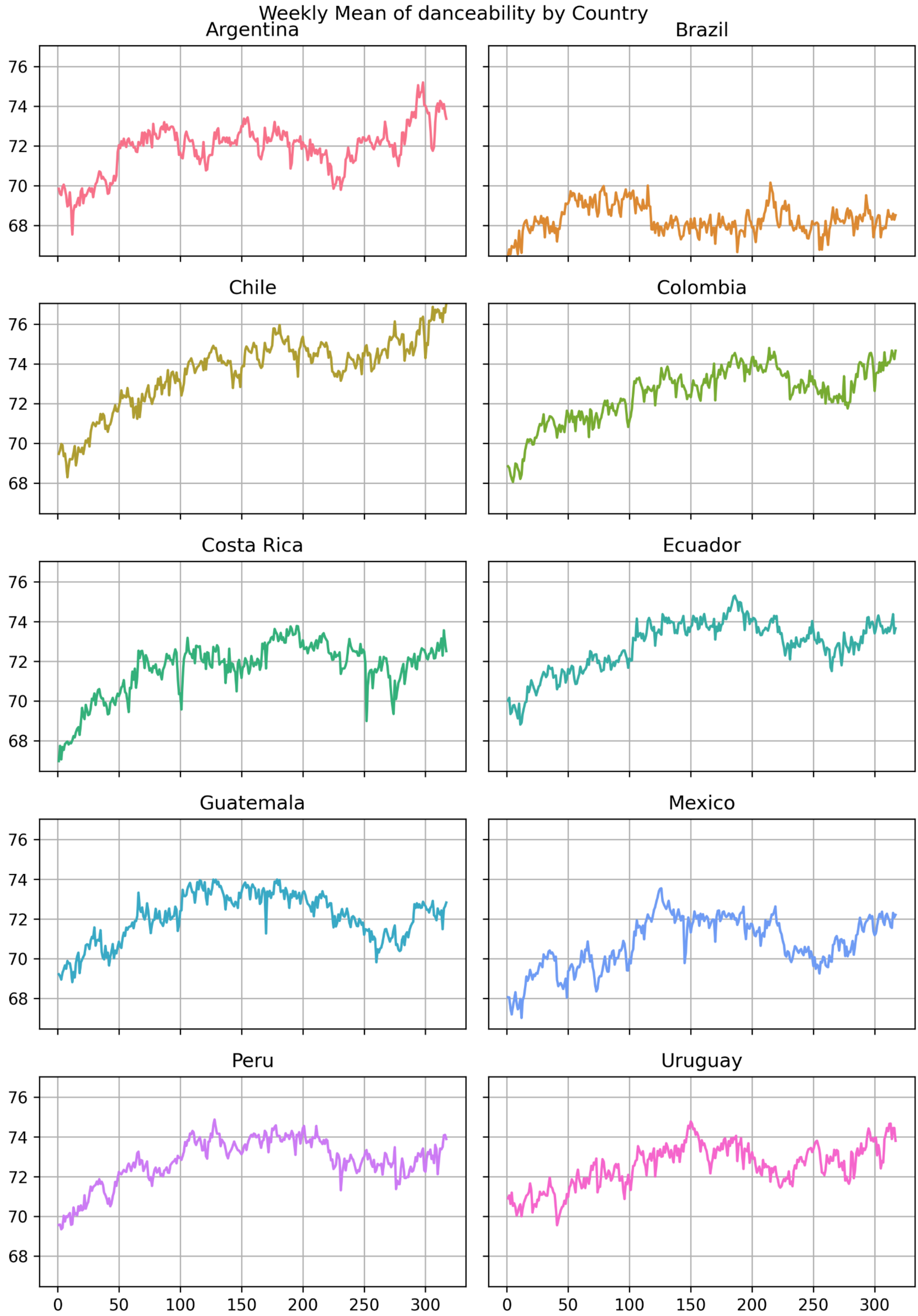
	Mode	4	4	4	4	4	4	4	4	4	4
	Std. Deviation	.21	.25	.21	.22	.23	.22	.31	.40	.22	.23
	Range	4	4	4	4	4	4	4	4	4	4
	Minimum	1	1	1	1	1	1	1	1	1	1
	Maximum	5	5	5	5	5	5	5	5	5	5

a. Multiple modes exist. The smallest value is shown

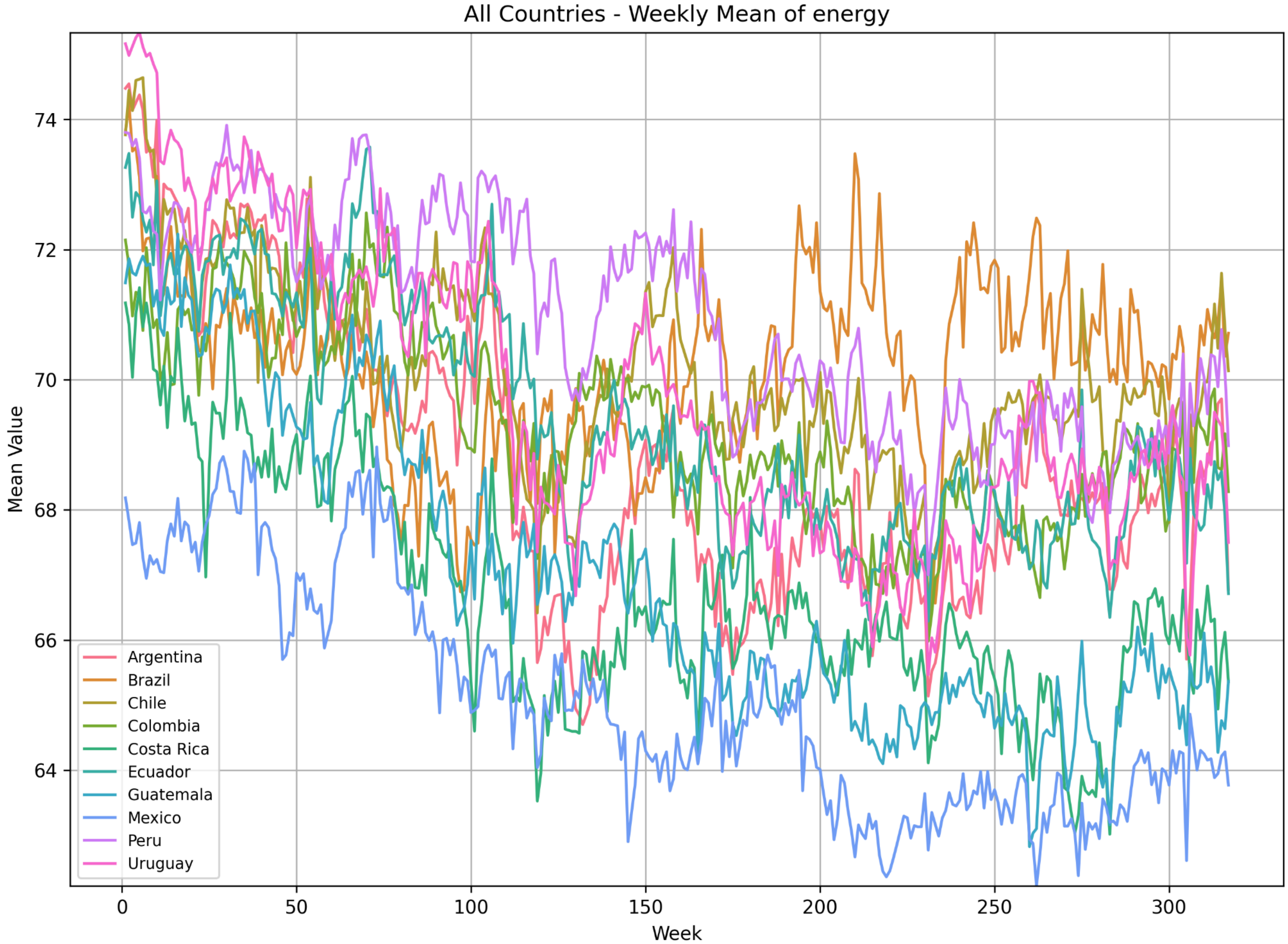


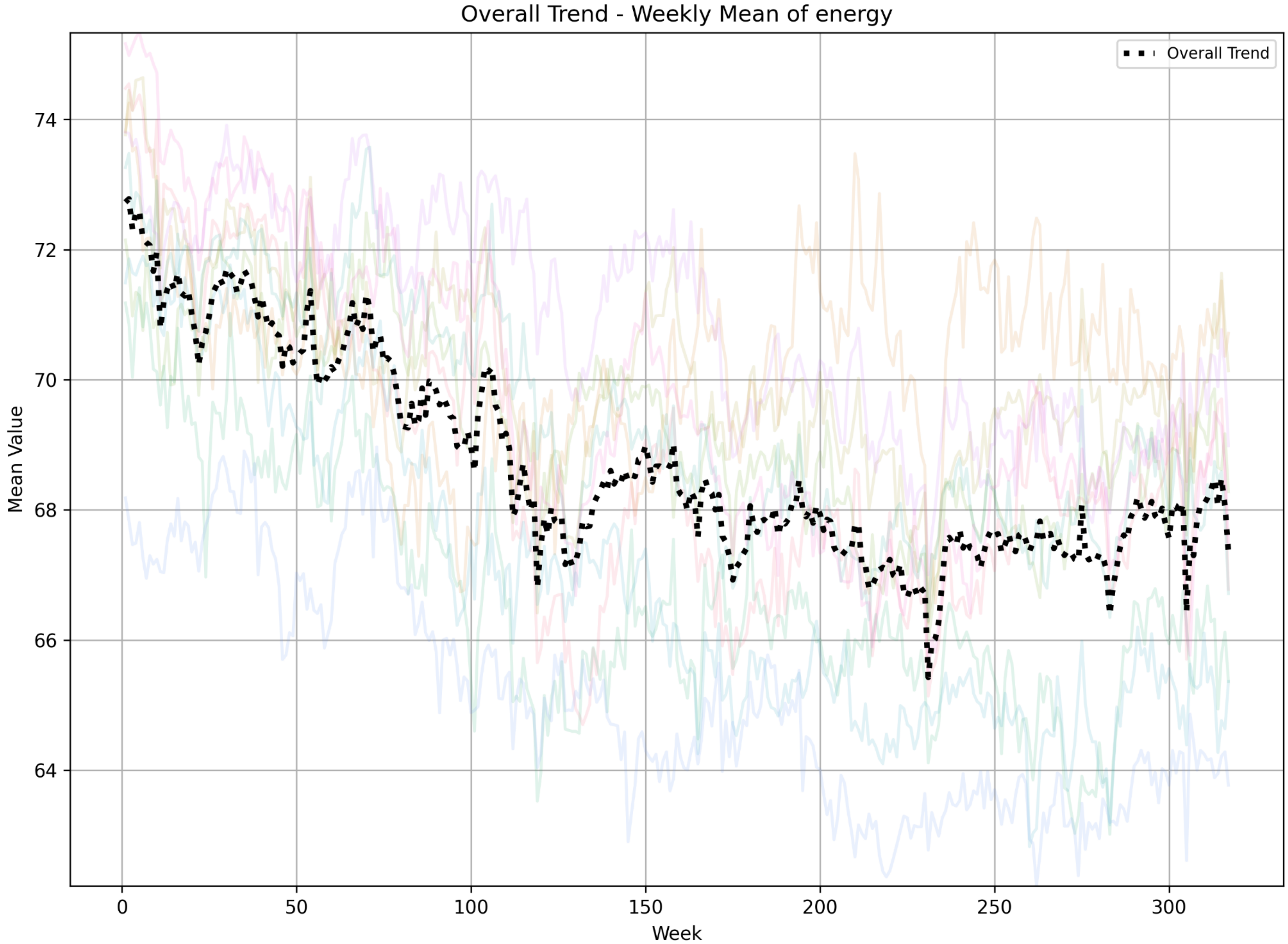
All Countries - Weekly Mean of danceability
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
Mean Value
Week
76
74
72
70
68
0
50
100
150
200
250
300


Overall Trend - Weekly Mean of danceability
Overall Trend
Mean Value
Week
76
74
72
70
68
0
50
100
150
200
250
300

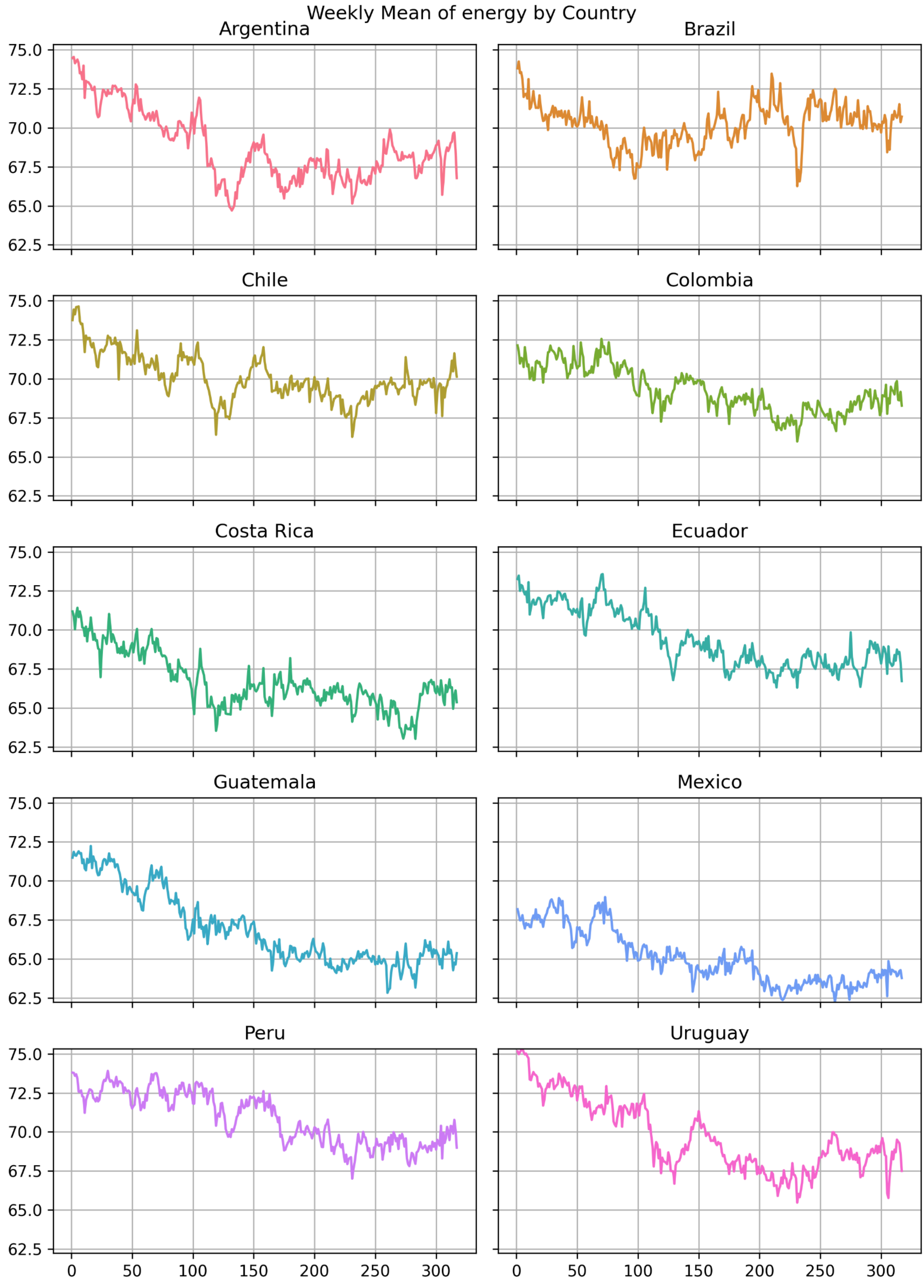


Weekly Mean of danceability by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
68
70
72
74
76
0
50
100
150
200
250
300

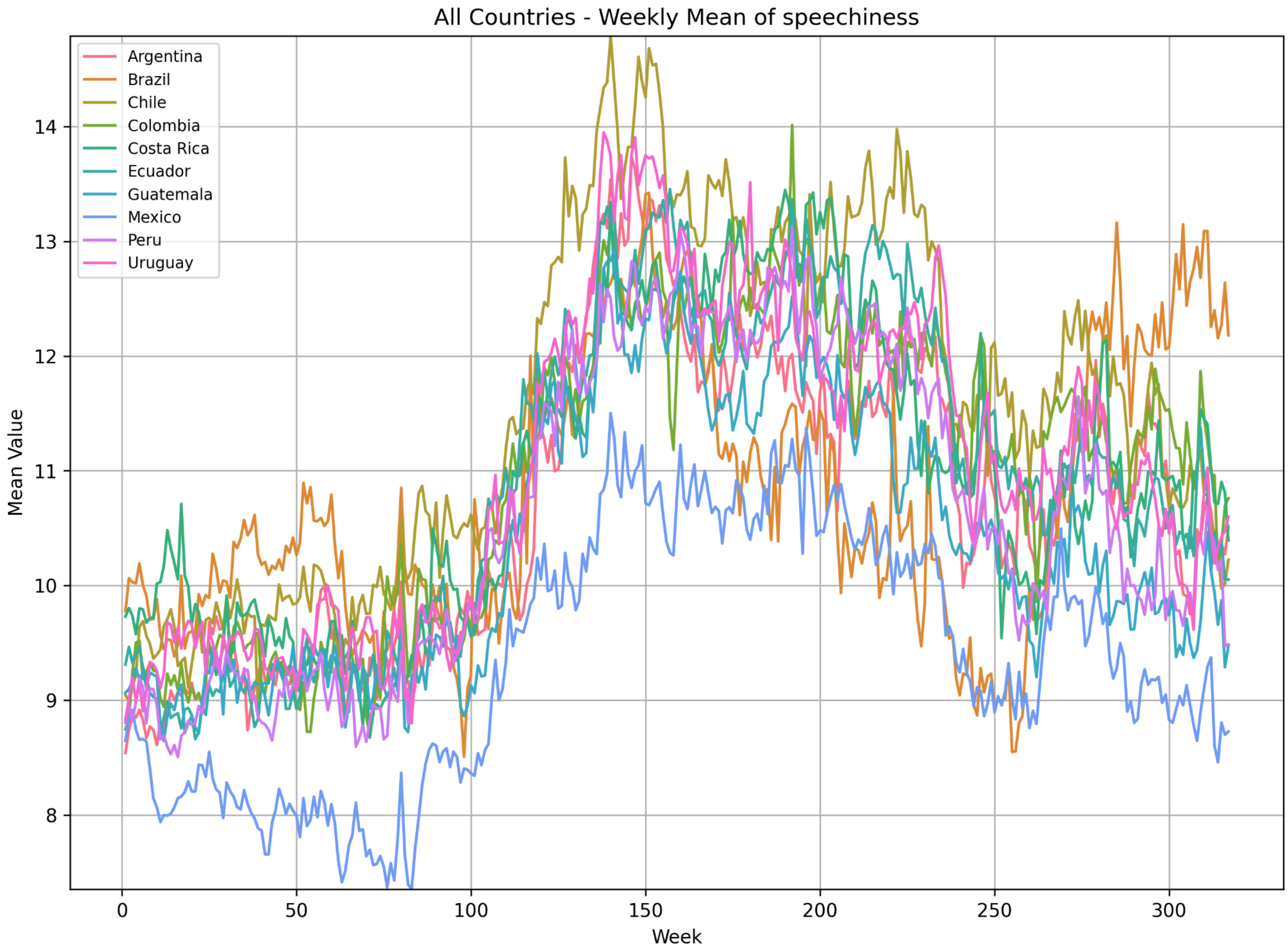

All Countries - Weekly Mean of energy
Mean Value
Week
74
72
70
68
66
64
0
50
100
150
200
250
300
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay

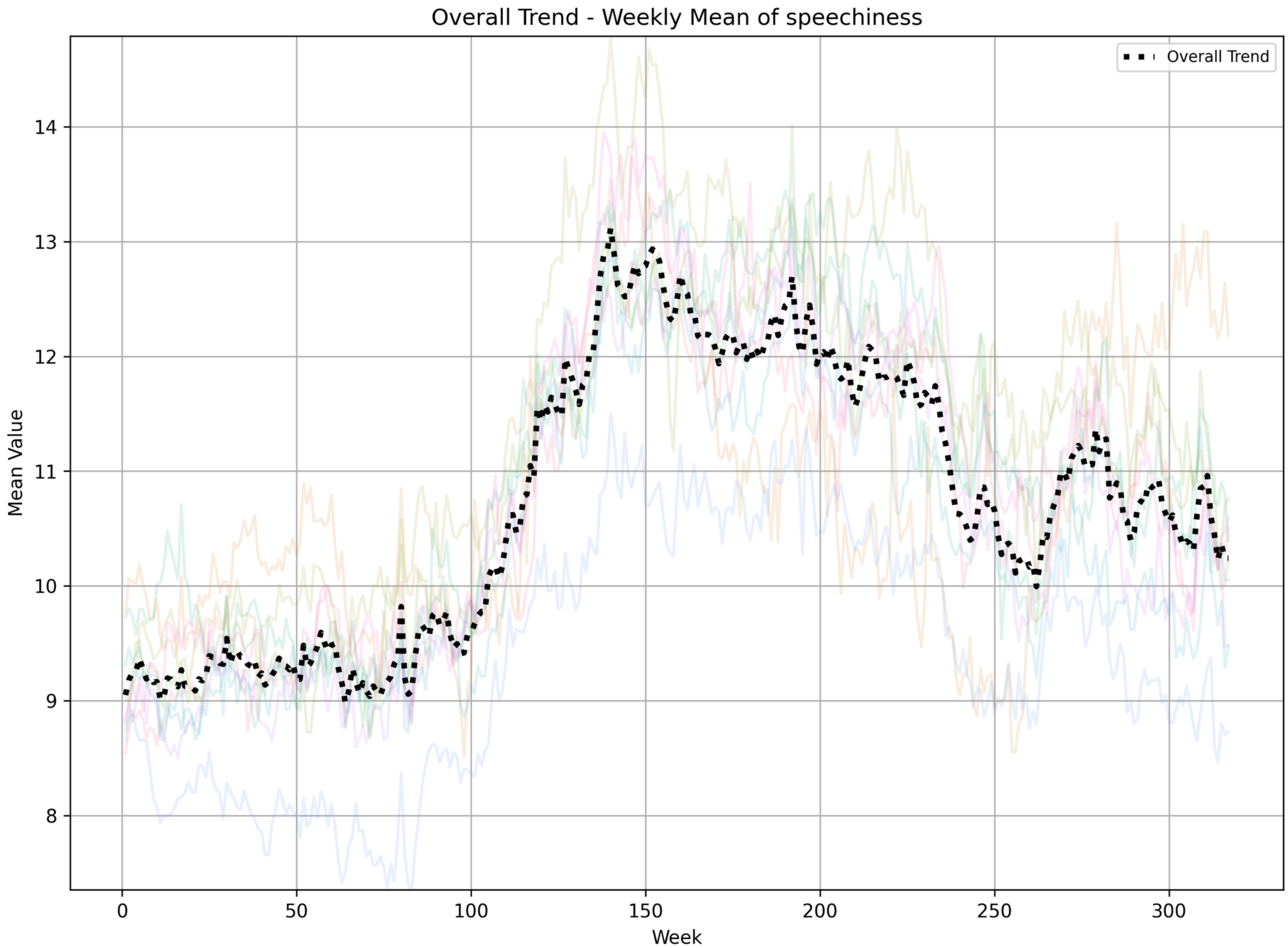


Overall Trend - Weekly Mean of energy
Overall Trend
Mean Value
Week
74
72
70
68
66
64
0
50
100
150
200
250
300

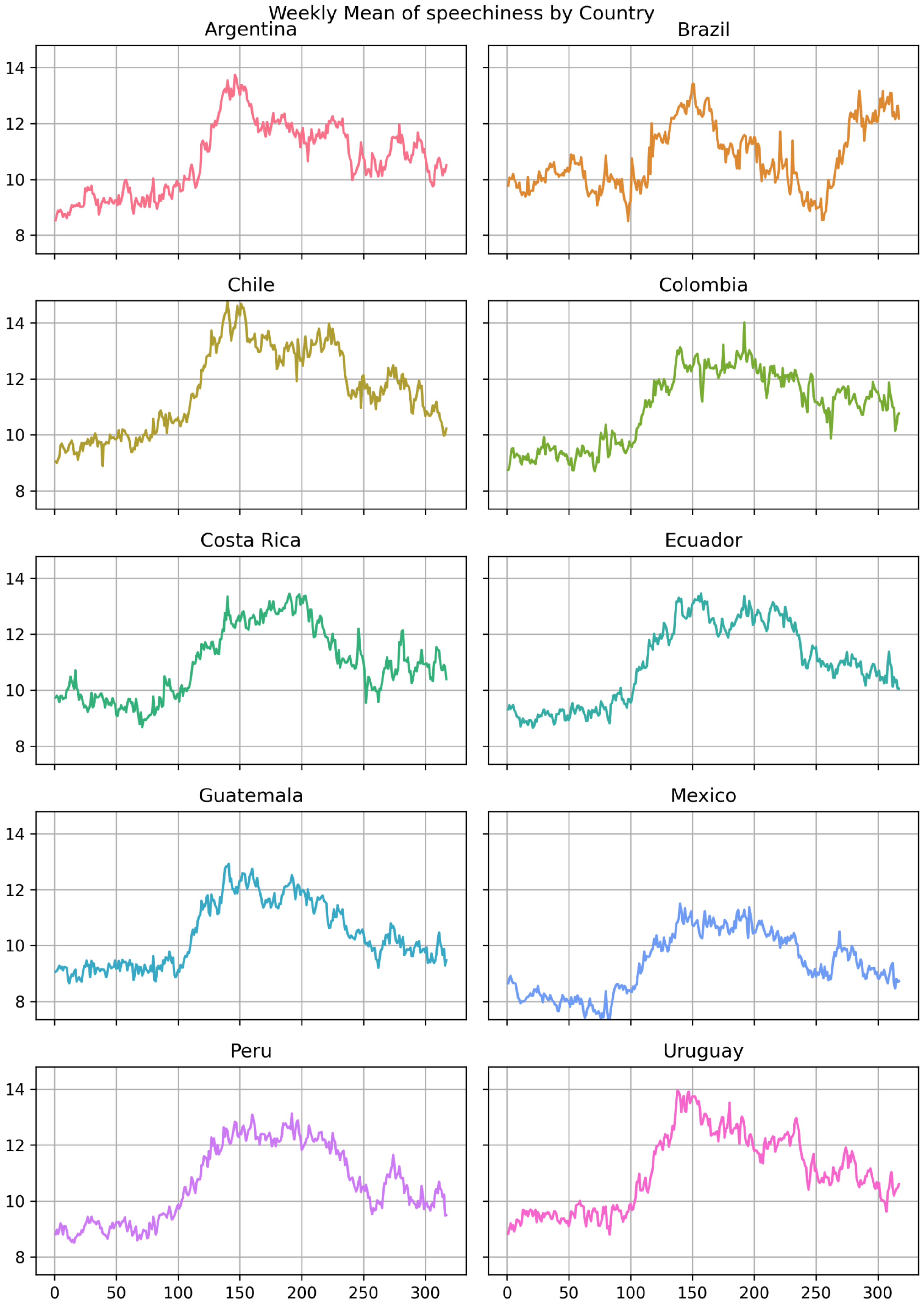


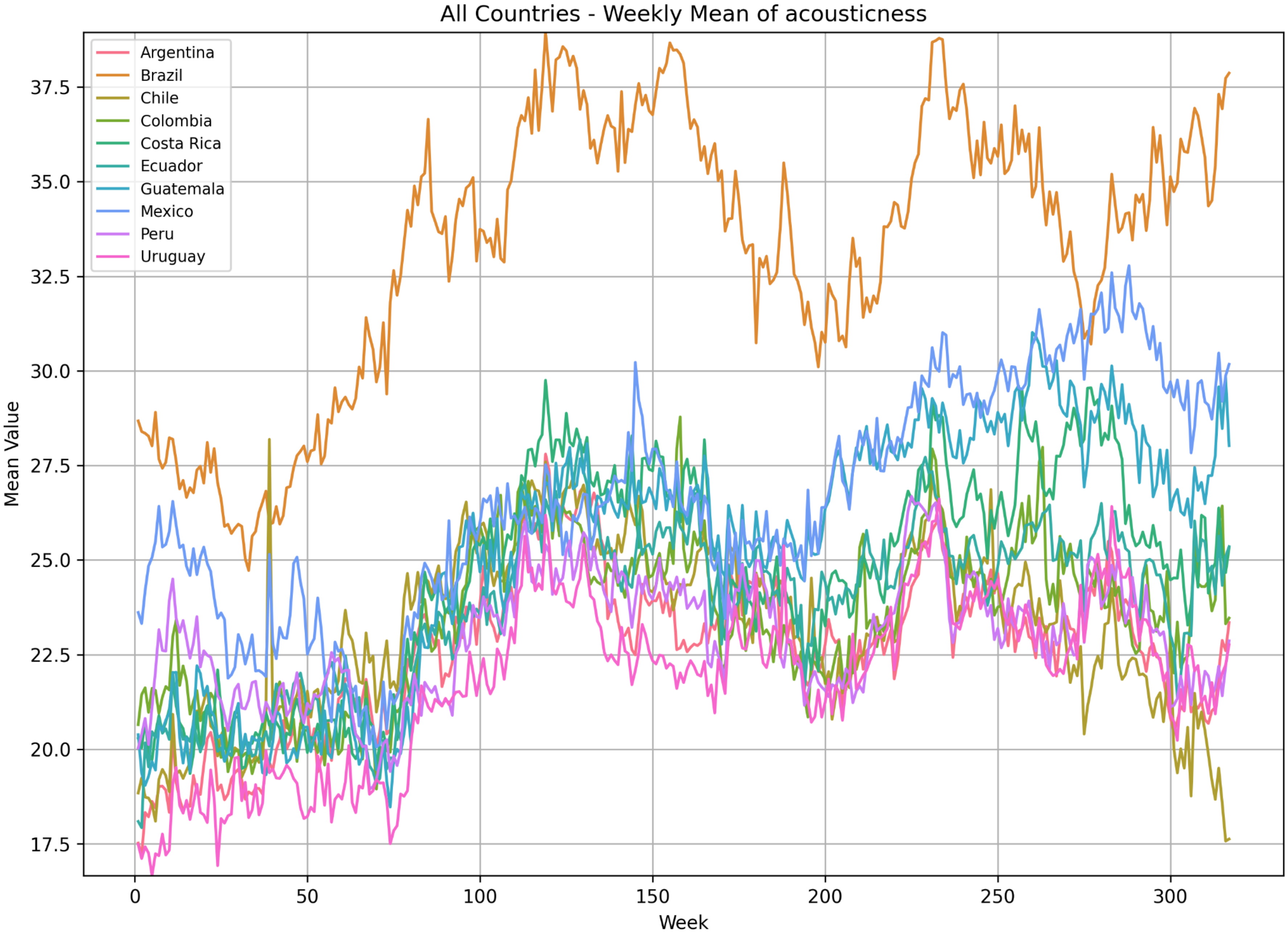
Weekly Mean of energy by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
75.0
72.5
70.0
67.5
65.0
62.5
0
50
100
150
200
250
300


All Countries - Weekly Mean of speechiness
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
Mean Value
Week
14
13
12
11
10
9
8
0
50
100
150
200
250
300

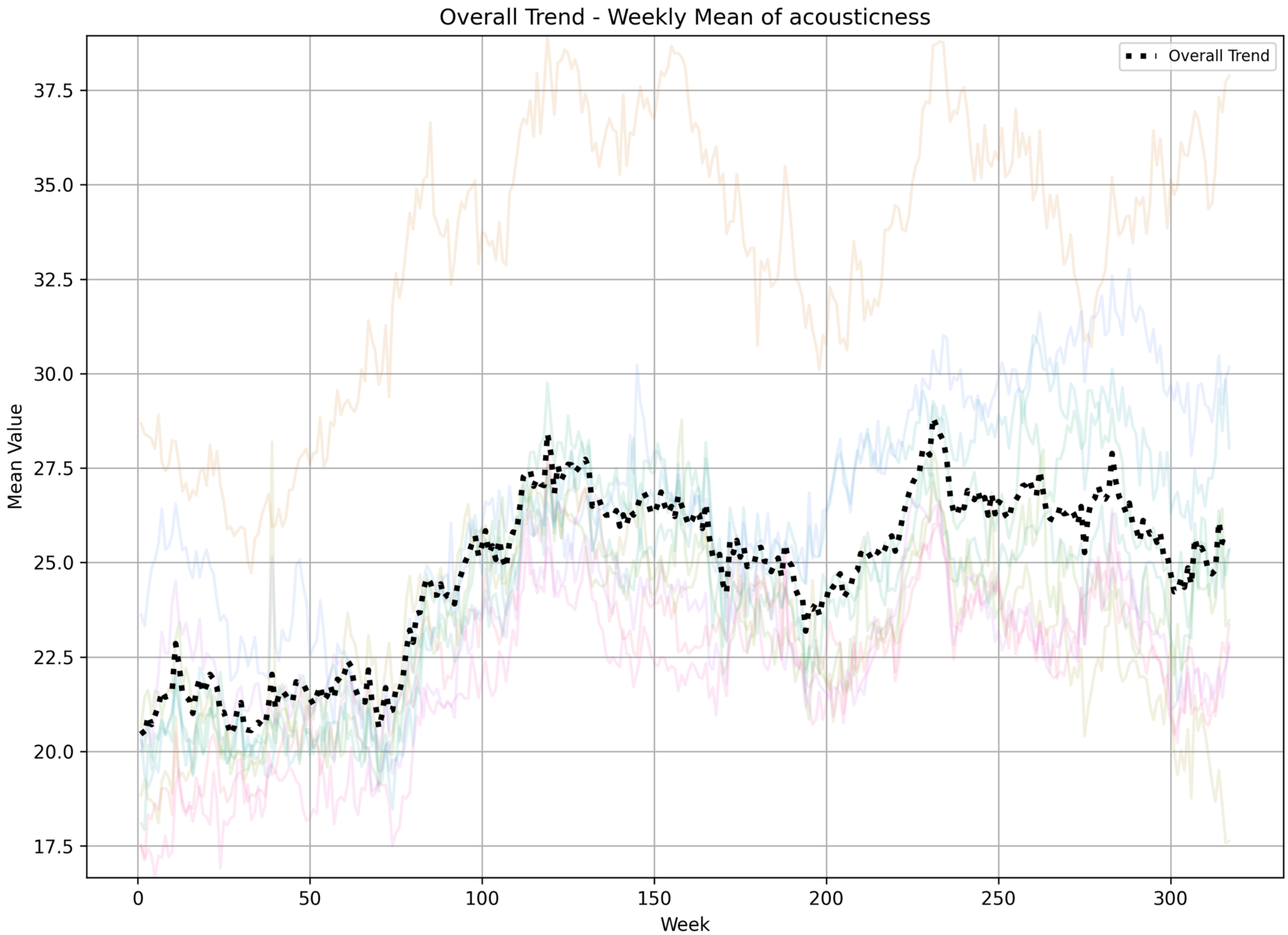


Overall Trend - Weekly Mean of speechiness
Overall Trend
Mean Value
Week
14
13
12
11
10
9
8
0
50
100
150
200
250
300

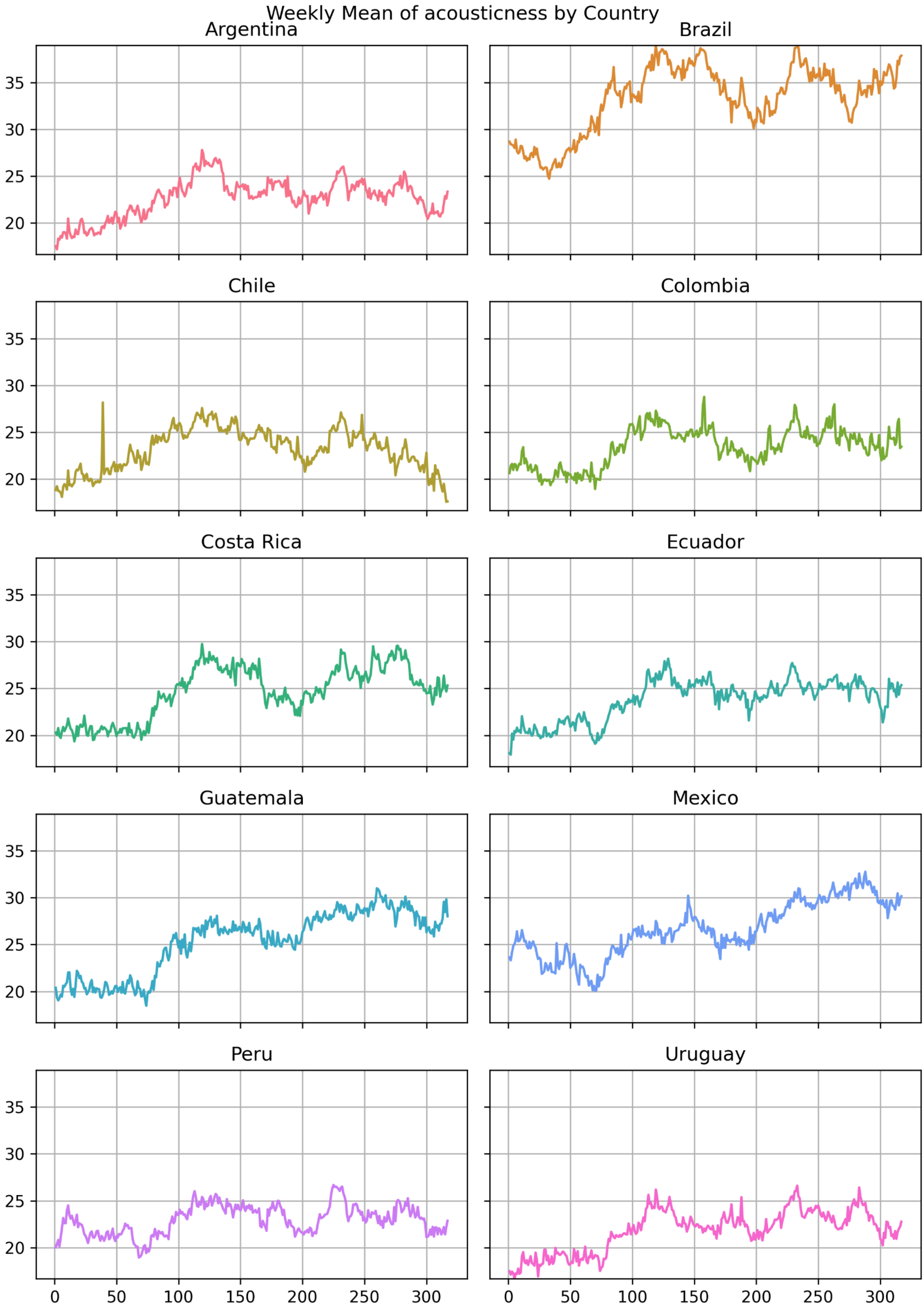

Weekly Mean of speechiness by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
8
10
12
14
0
50
100
150
200
250
300

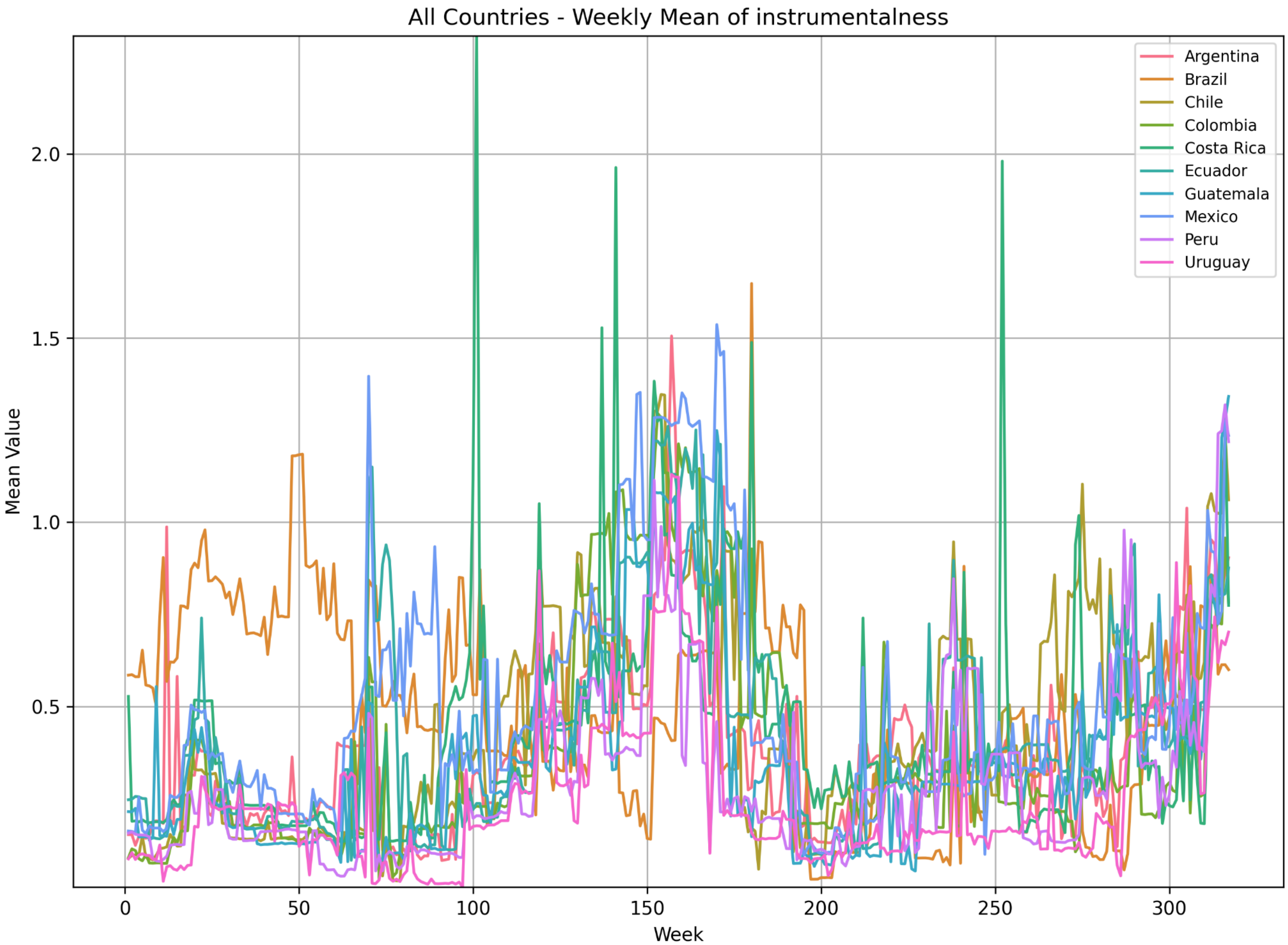


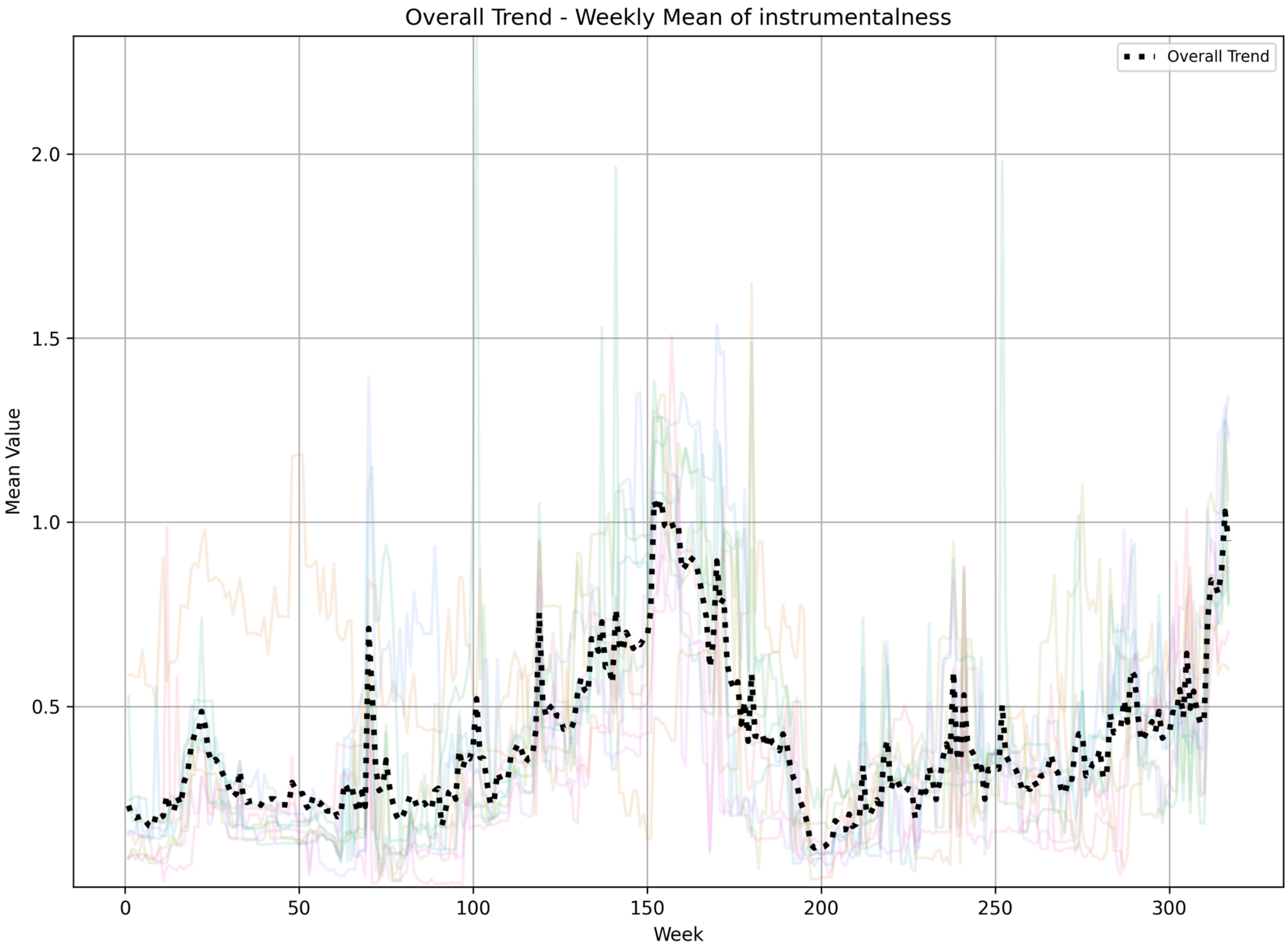
All Countries - Weekly Mean of acousticness
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
Mean Value
0
50
100
150
200
250
300
Week



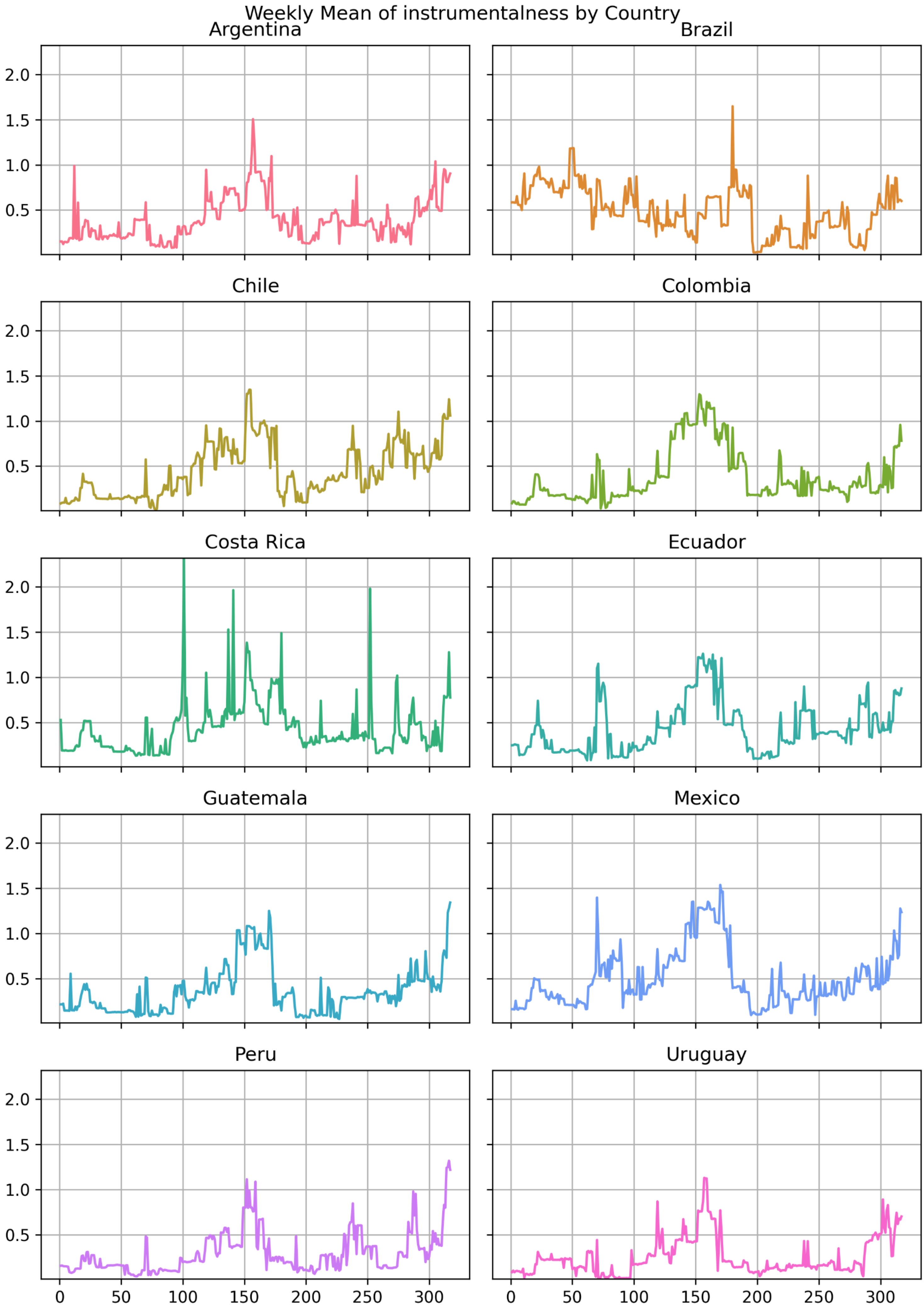
Overall Trend - Weekly Mean of acousticness
Overall Trend
Mean Value
Week
17.5
20.0
22.5
25.0
27.5
30.0
32.5
35.0
37.5
0
50
100
150
200
250
300


Weekly Mean of acousticness by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
20
25
30
35
0
50
100
150
200
250
300

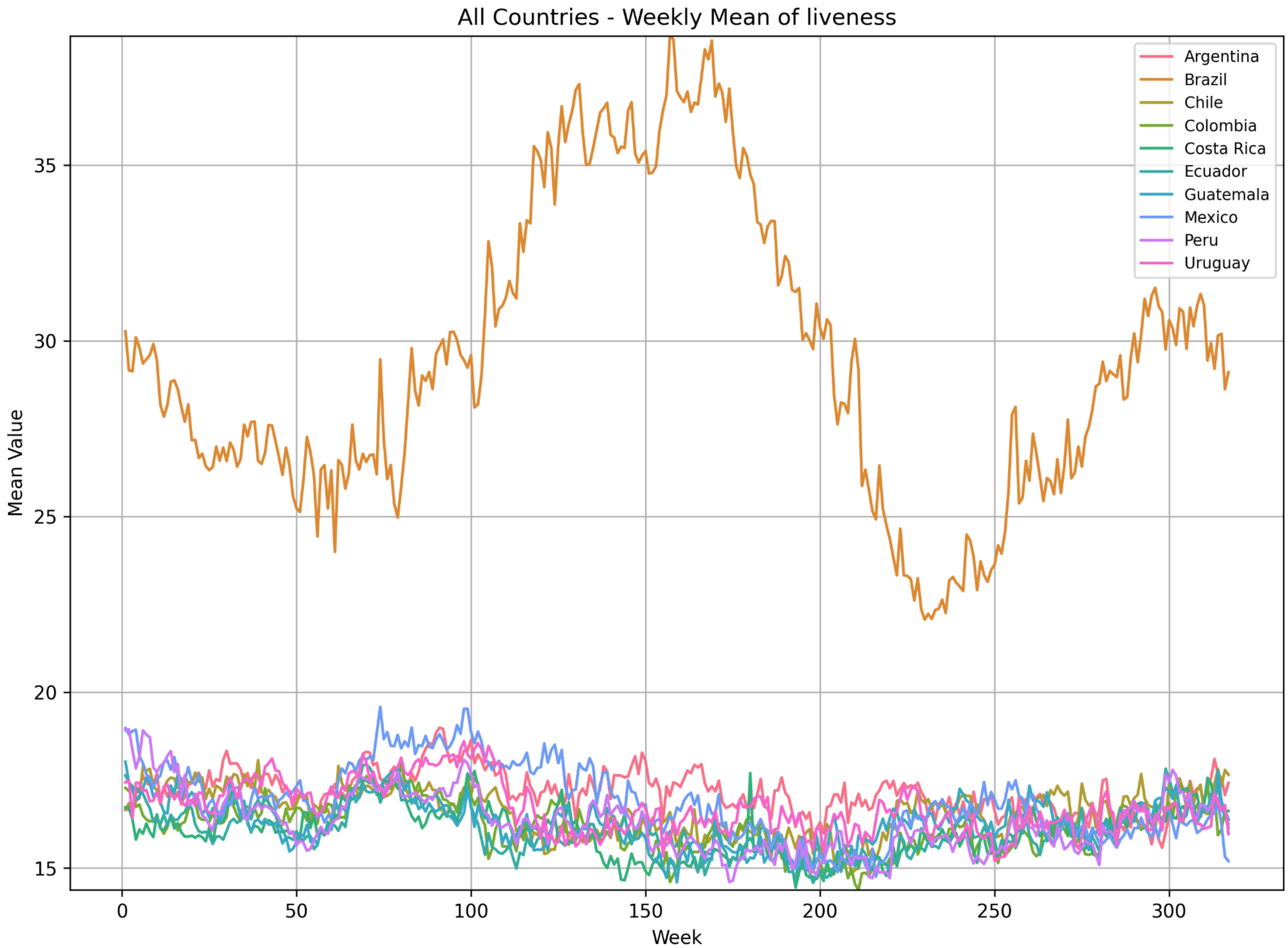

All Countries - Weekly Mean of instrumentalness
Mean Value
Week
0
50
100
150
200
250
300
0.5
1.0
1.5
2.0
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay

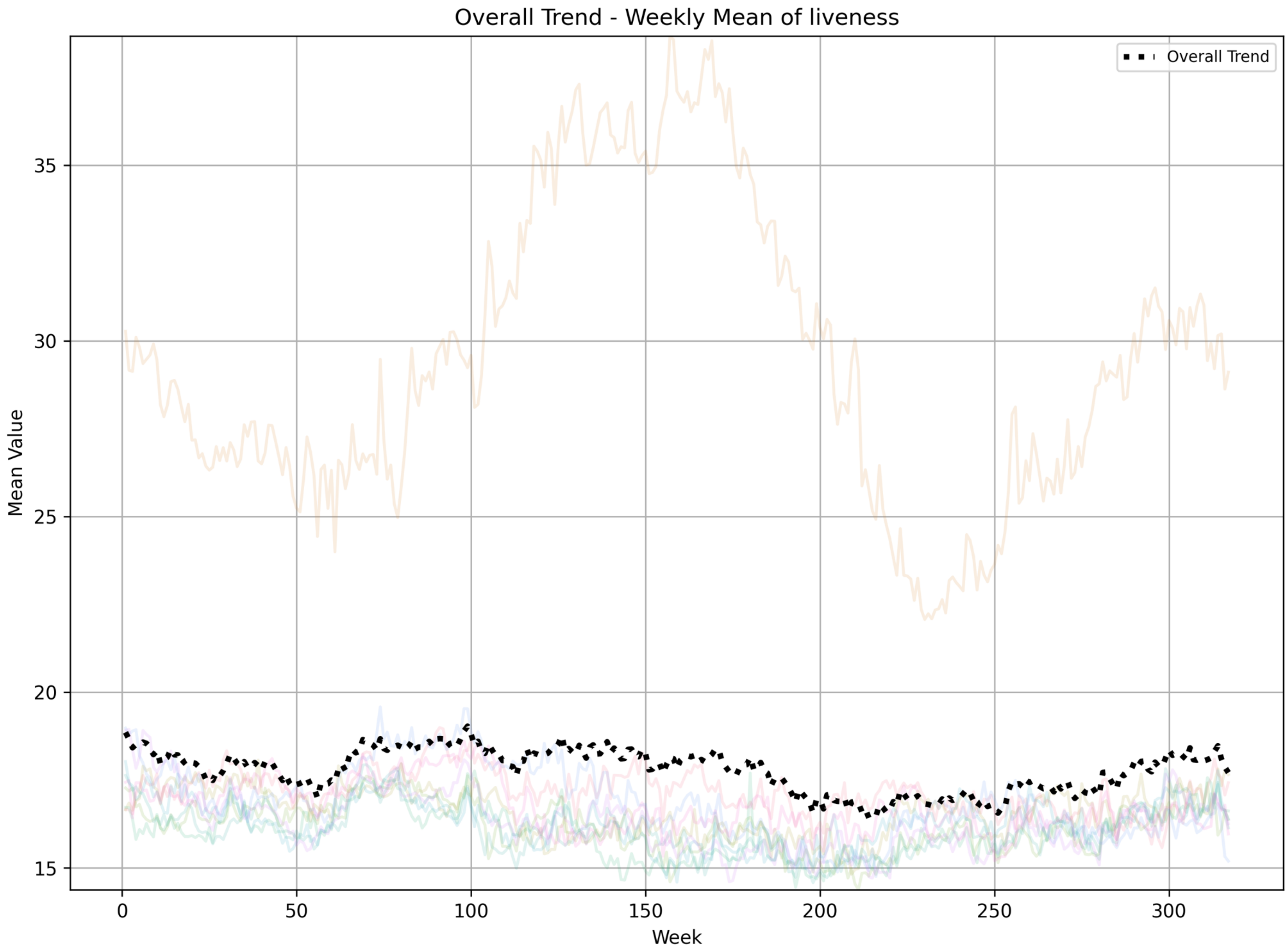


Overall Trend - Weekly Mean of instrumentalness
Overall Trend
Mean Value
Week
0
50
100
150
200
250
300
0.5
1.0
1.5
2.0

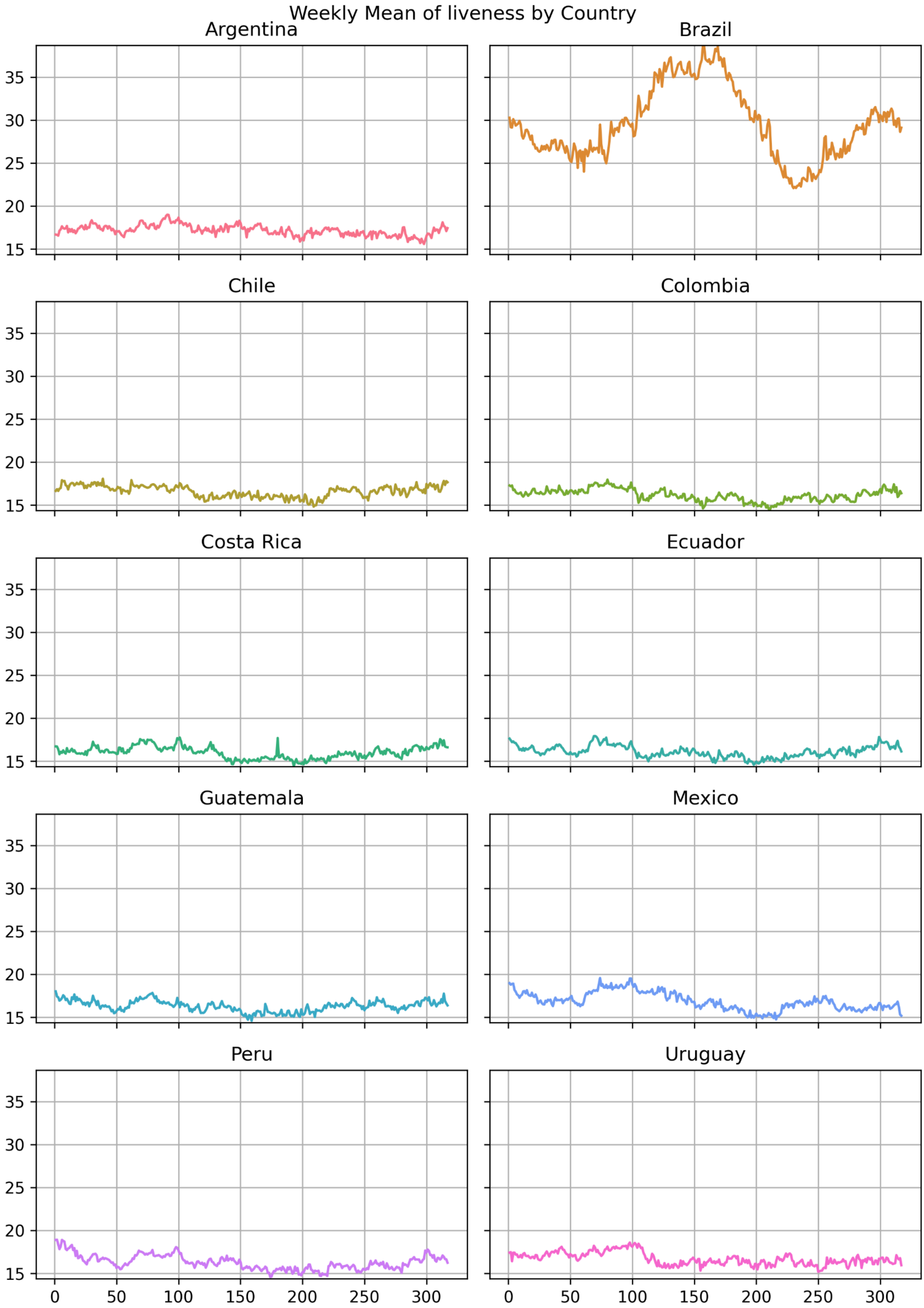


Weekly Mean of instrumentalness by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
2.0
1.5
1.0
0.5
0
50
100
150
200
250
300

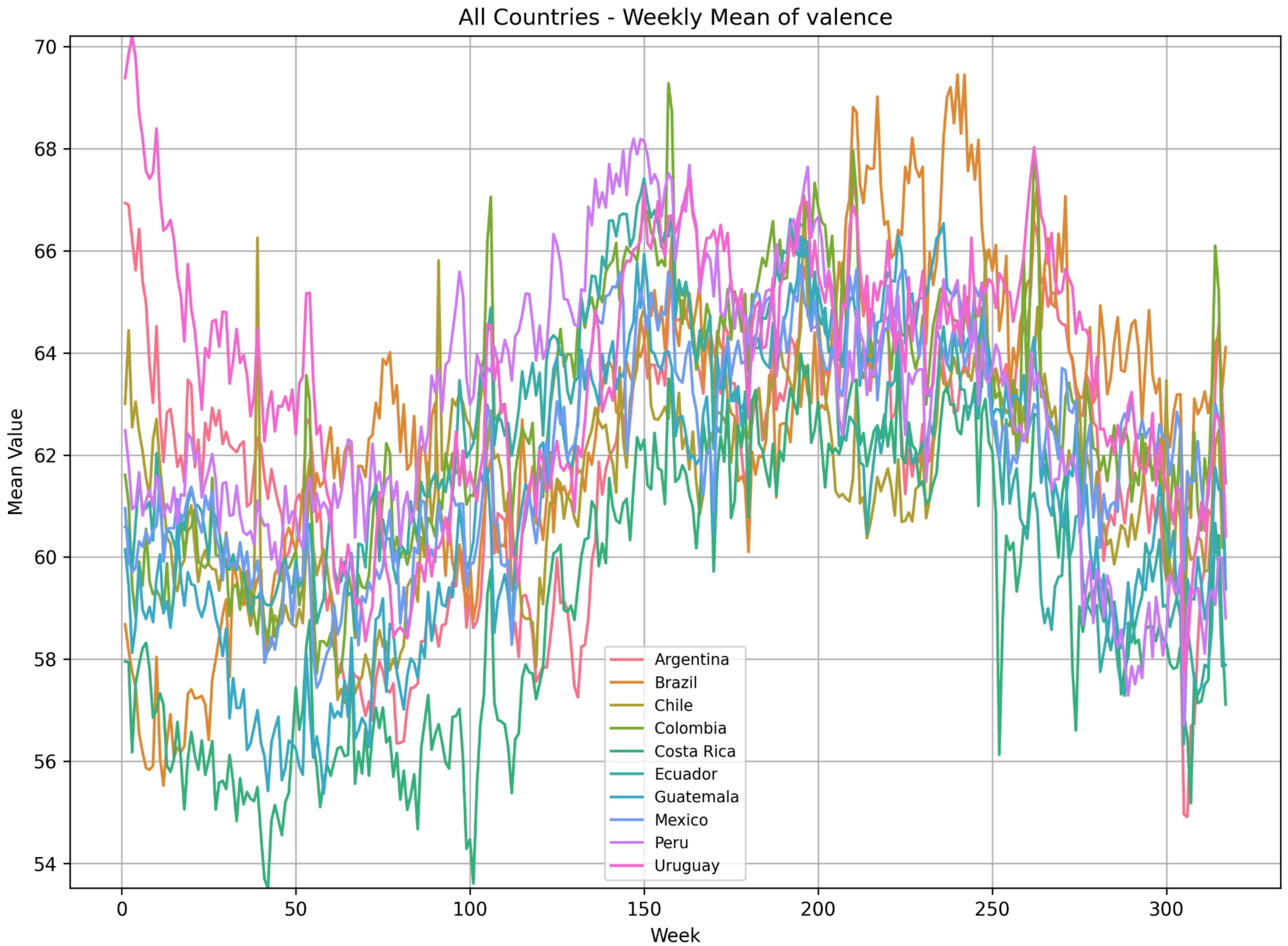

All Countries - Weekly Mean of liveness
Mean Value
Week
35
30
25
20
15
0
50
100
150
200
250
300
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay



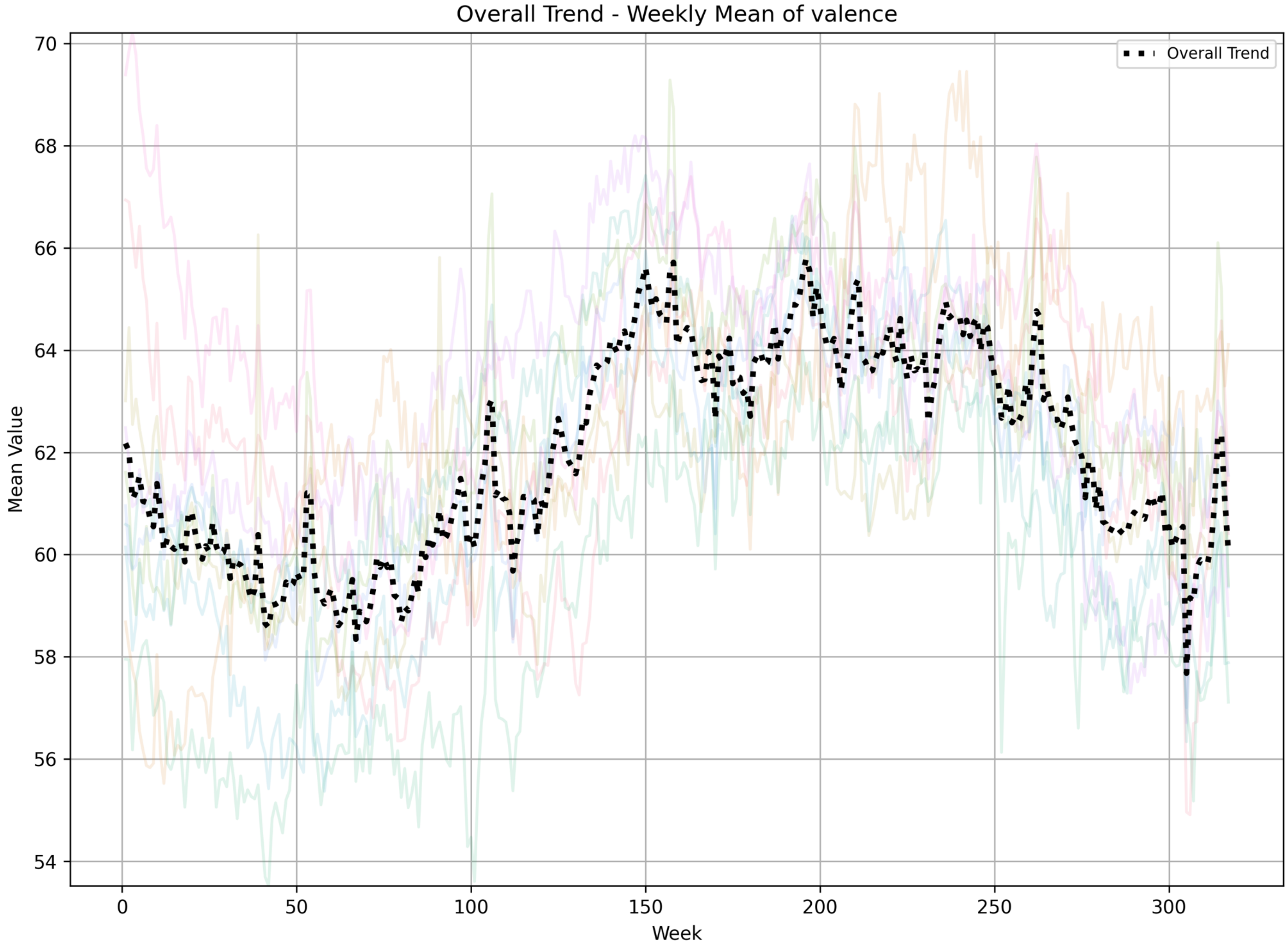
Overall Trend - Weekly Mean of liveness
Overall Trend
Mean Value
Week
15
20
25
30
35
0
50
100
150
200
250
300



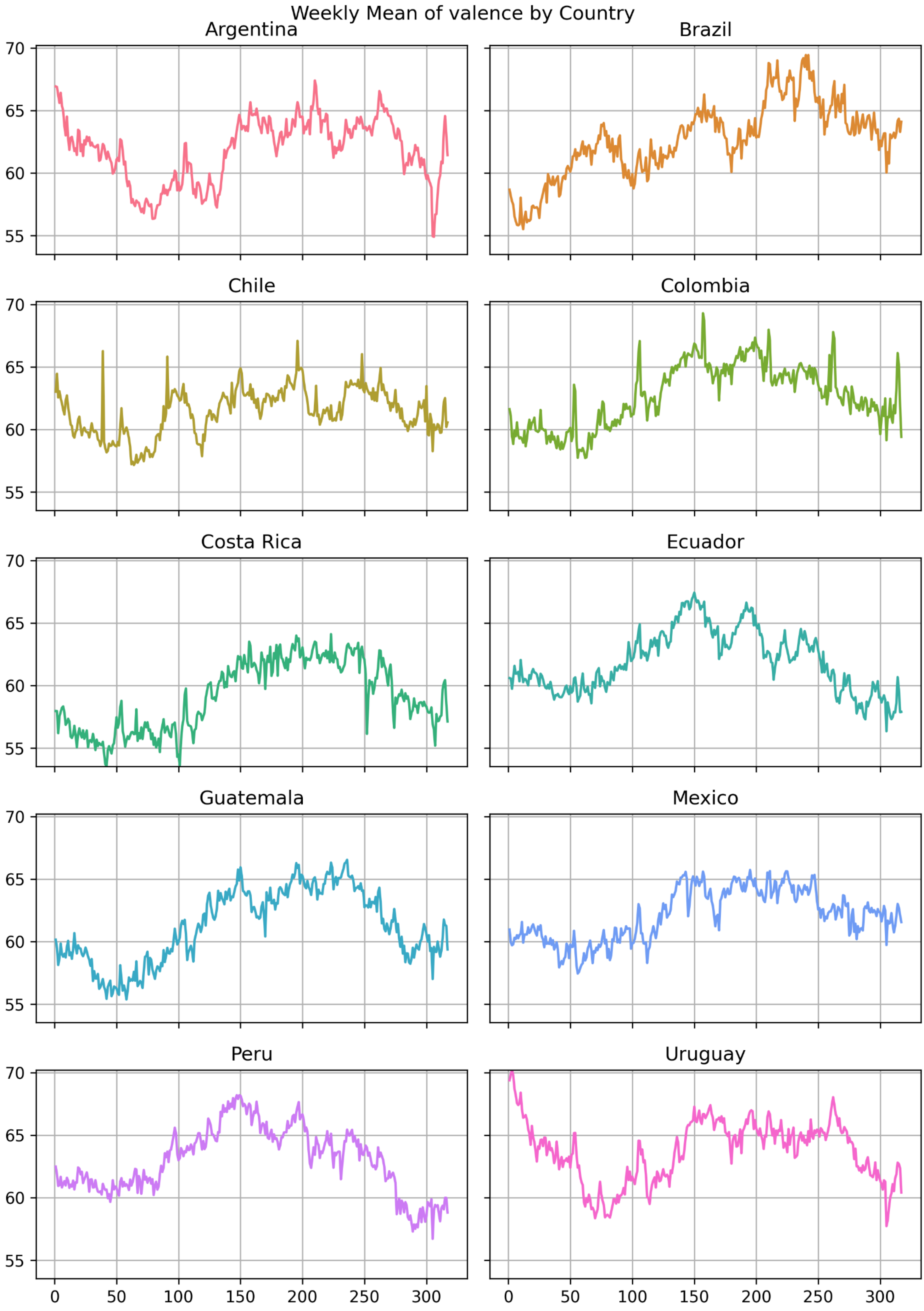
Weekly Mean of liveness by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
15
20
25
30
35
0
50
100
150
200
250
300



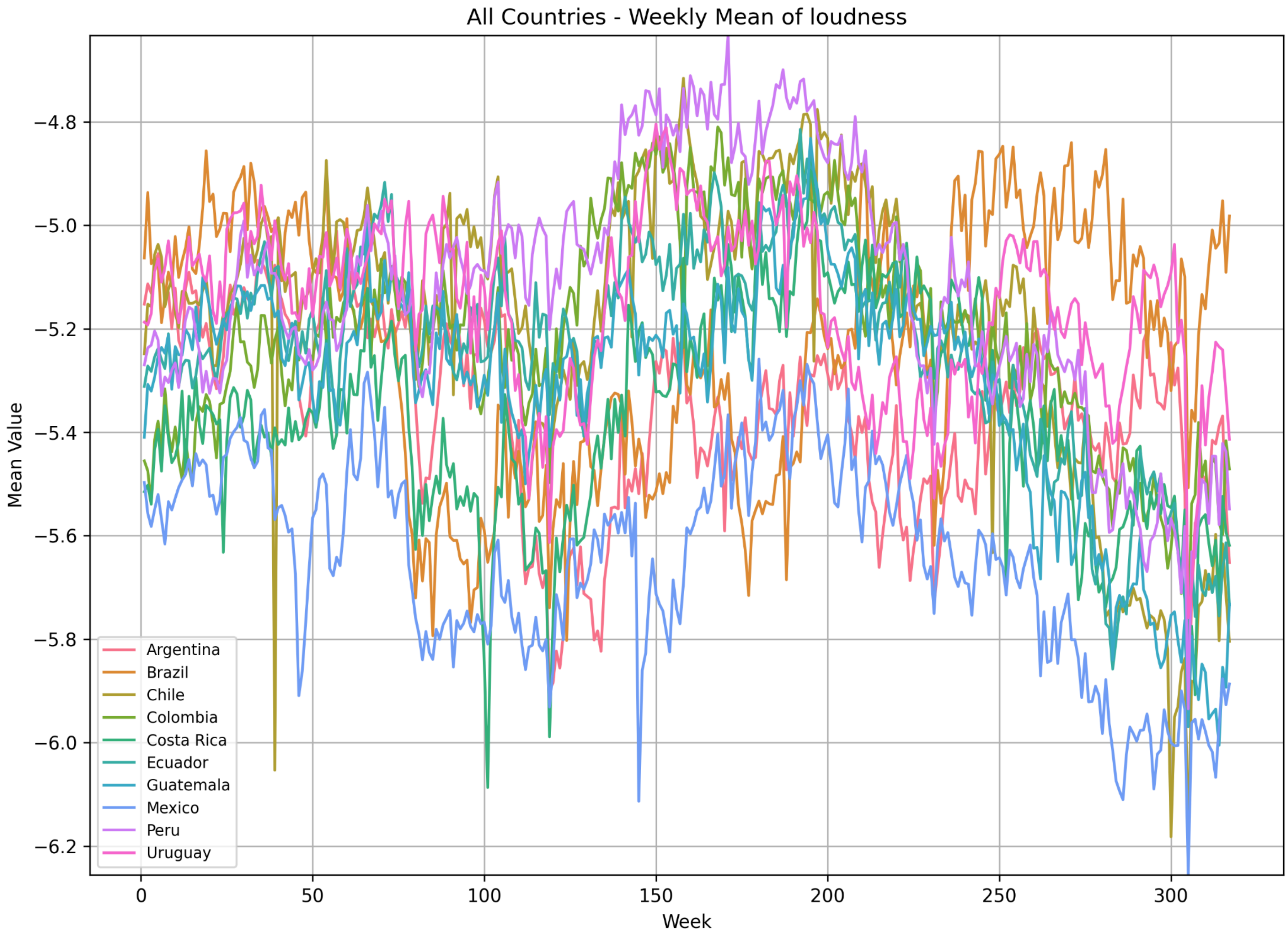
All Countries - Weekly Mean of valence
Mean Value
Week
70
68
66
64
62
60
58
56
54
0
50
100
150
200
250
300
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay

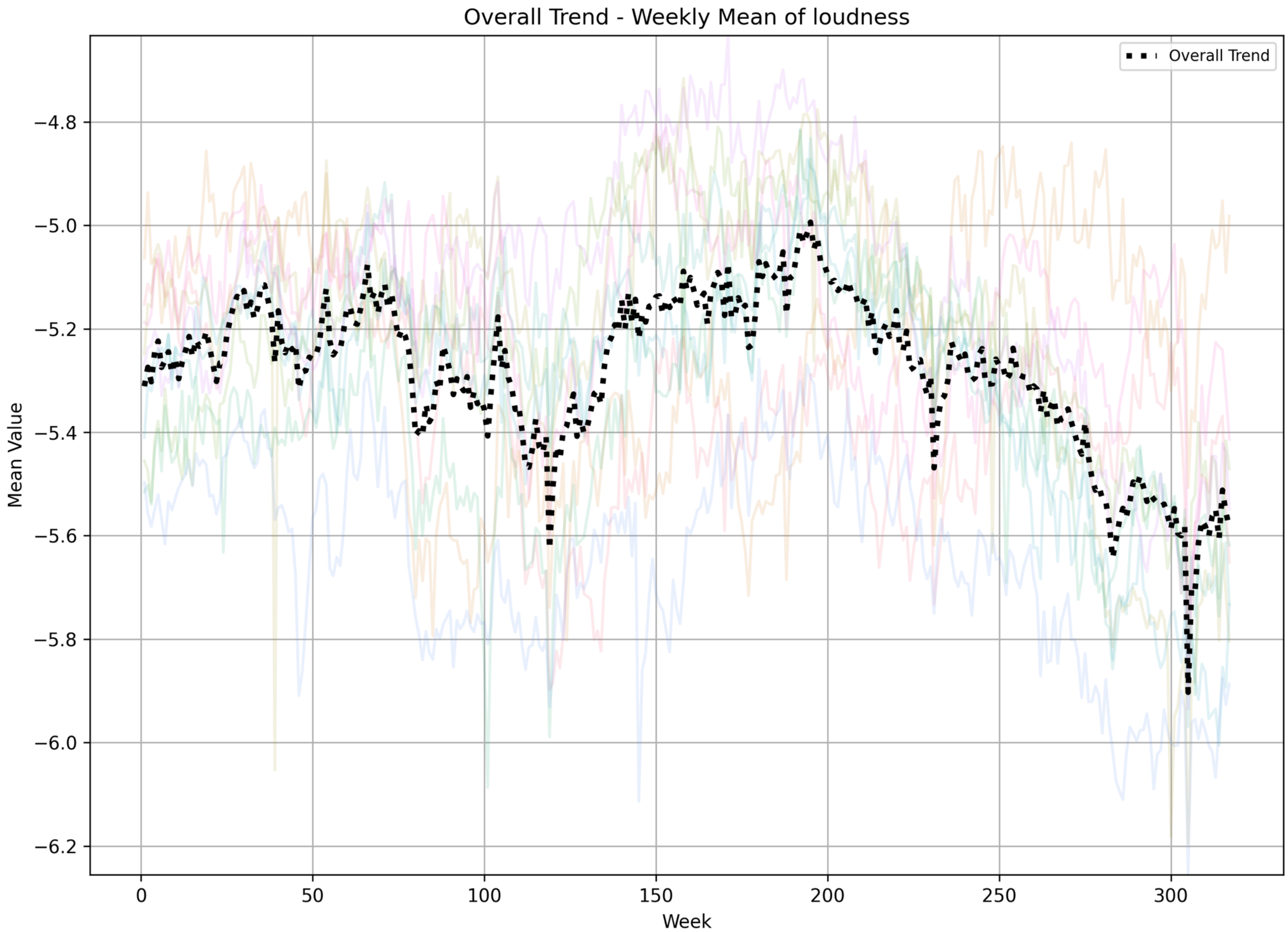


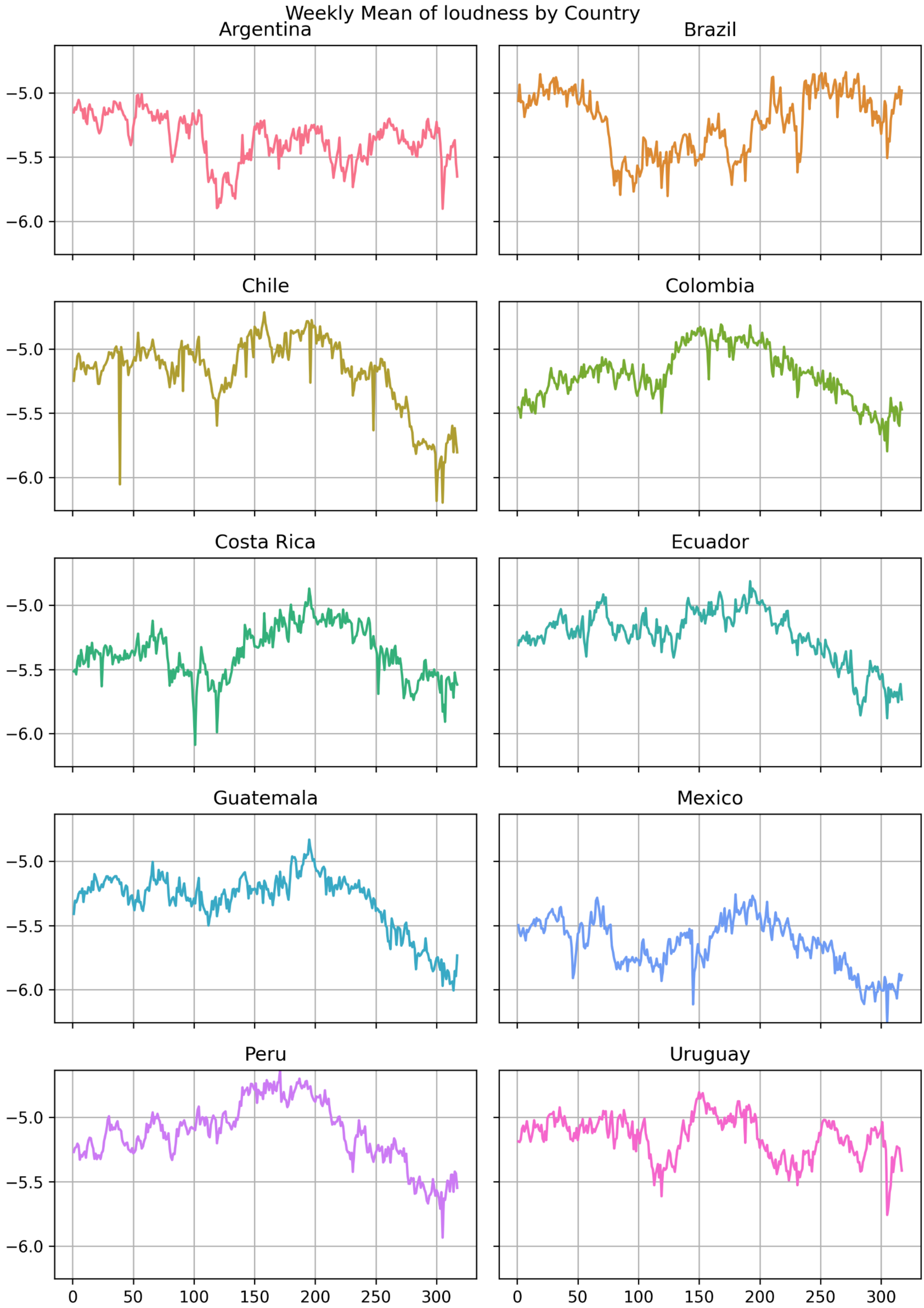
Overall Trend - Weekly Mean of valence
Overall Trend
Mean Value
Week
70
68
66
64
62
60
58
56
54
0
50
100
150
200
250
300

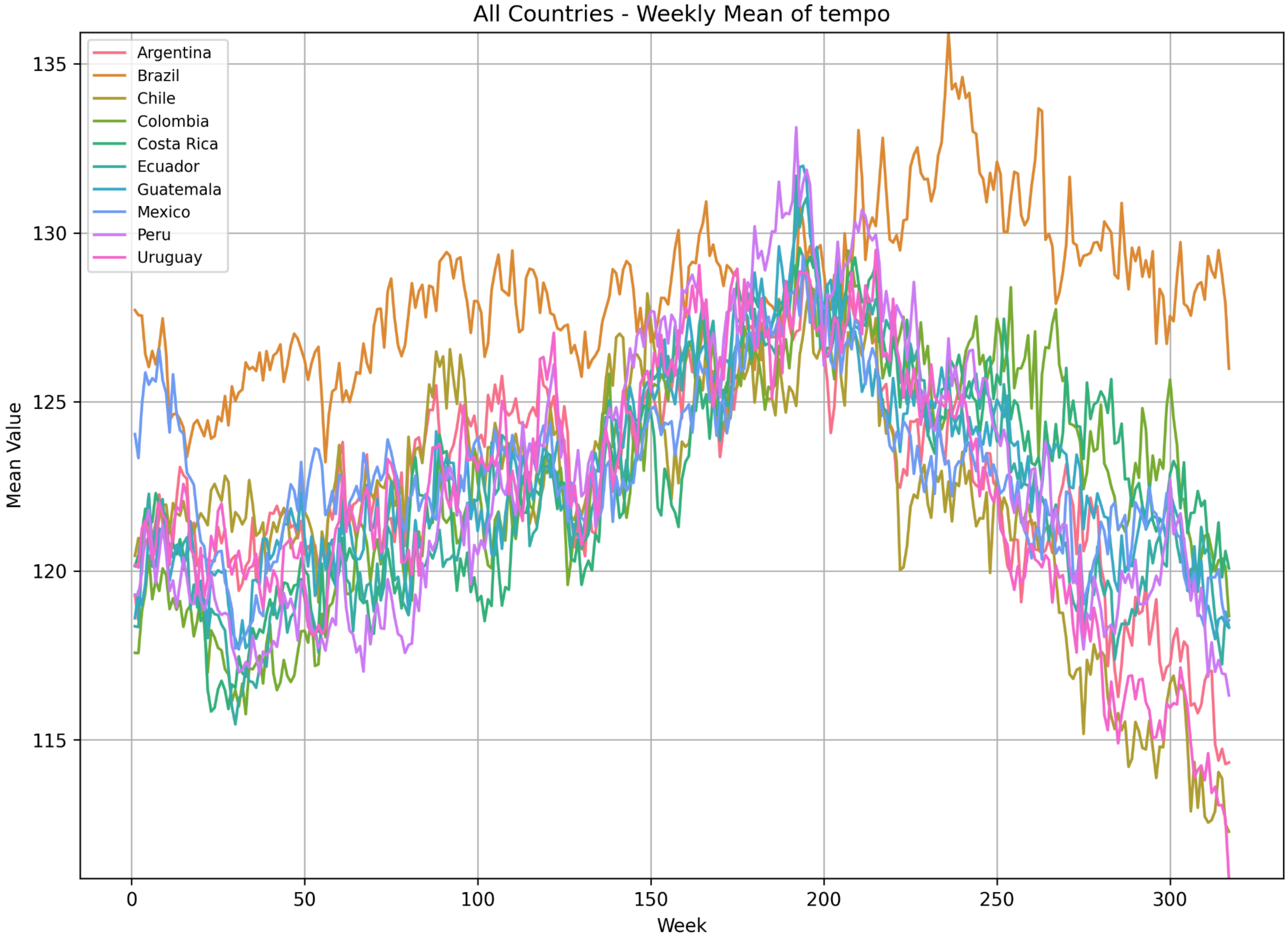


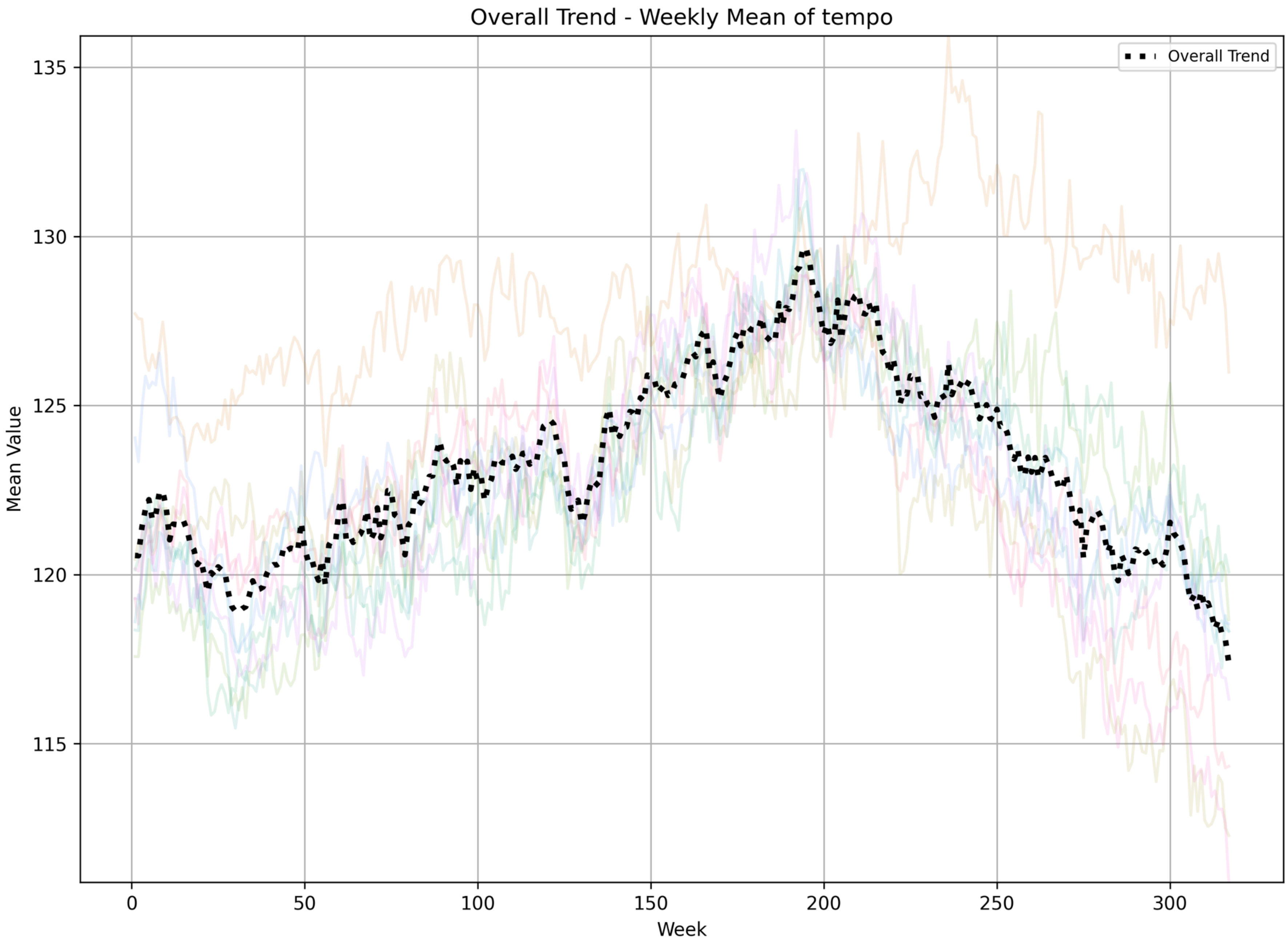
Weekly Mean of valence by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
70
65
60
55
0
50
100
150
200
250
300


All Countries - Weekly Mean of loudness
Mean Value
Week
−4.8
−5.0
−5.2
−5.4
−5.6
−5.8
−6.0
−6.2
0
50
100
150
200
250
300
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay

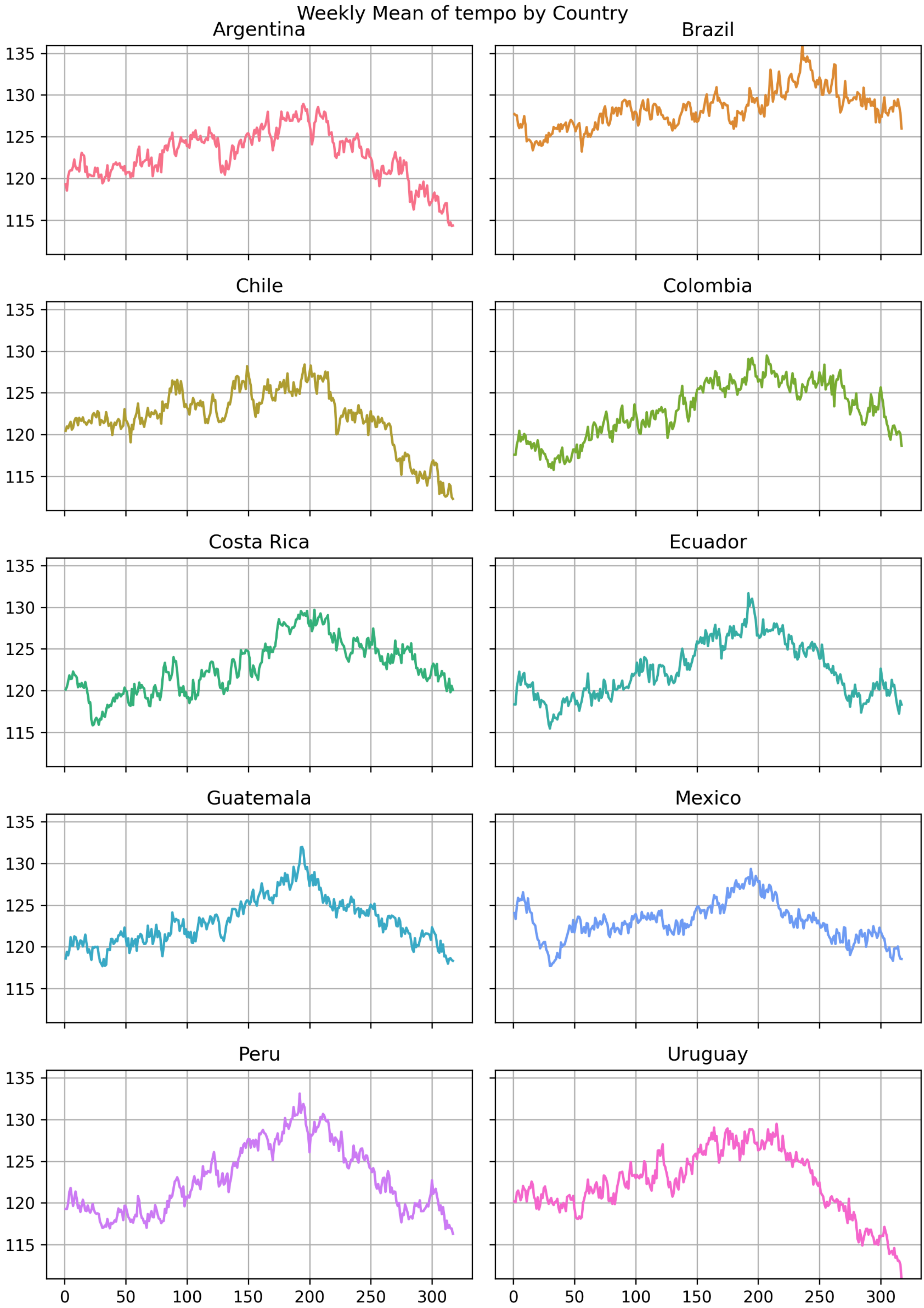

Overall Trend - Weekly Mean of loudness
Overall Trend
Mean Value
Week
−4.8
−5.0
−5.2
−5.4
−5.6
−5.8
−6.0
−6.2
0
50
100
150
200
250
300

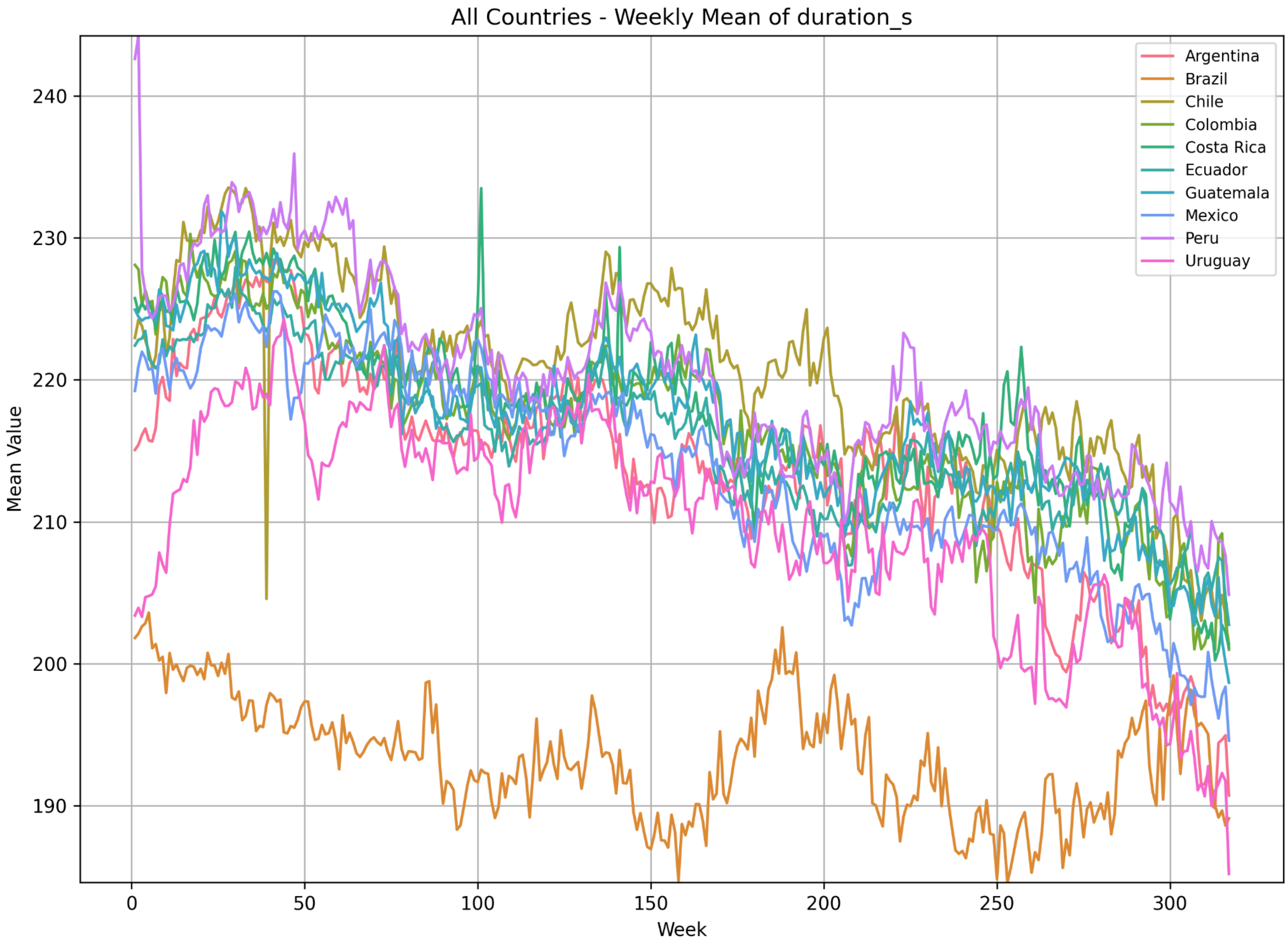

Weekly Mean of loudness by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
−5.0
−5.5
−6.0
0
50
100
150
200
250
300

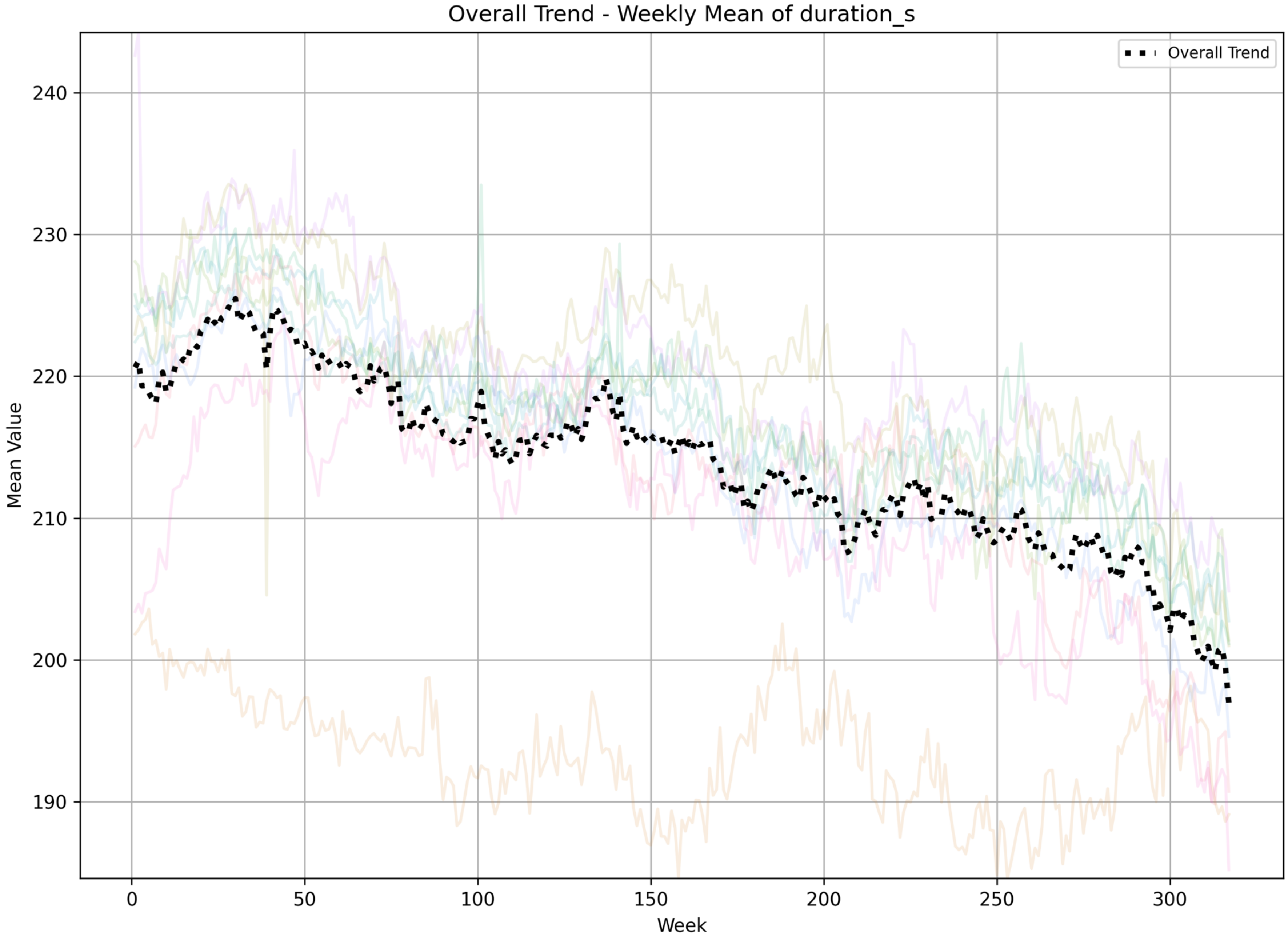

All Countries - Weekly Mean of tempo
Mean Value
Week
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
135
130
125
120
115
0
50
100
150
200
250
300



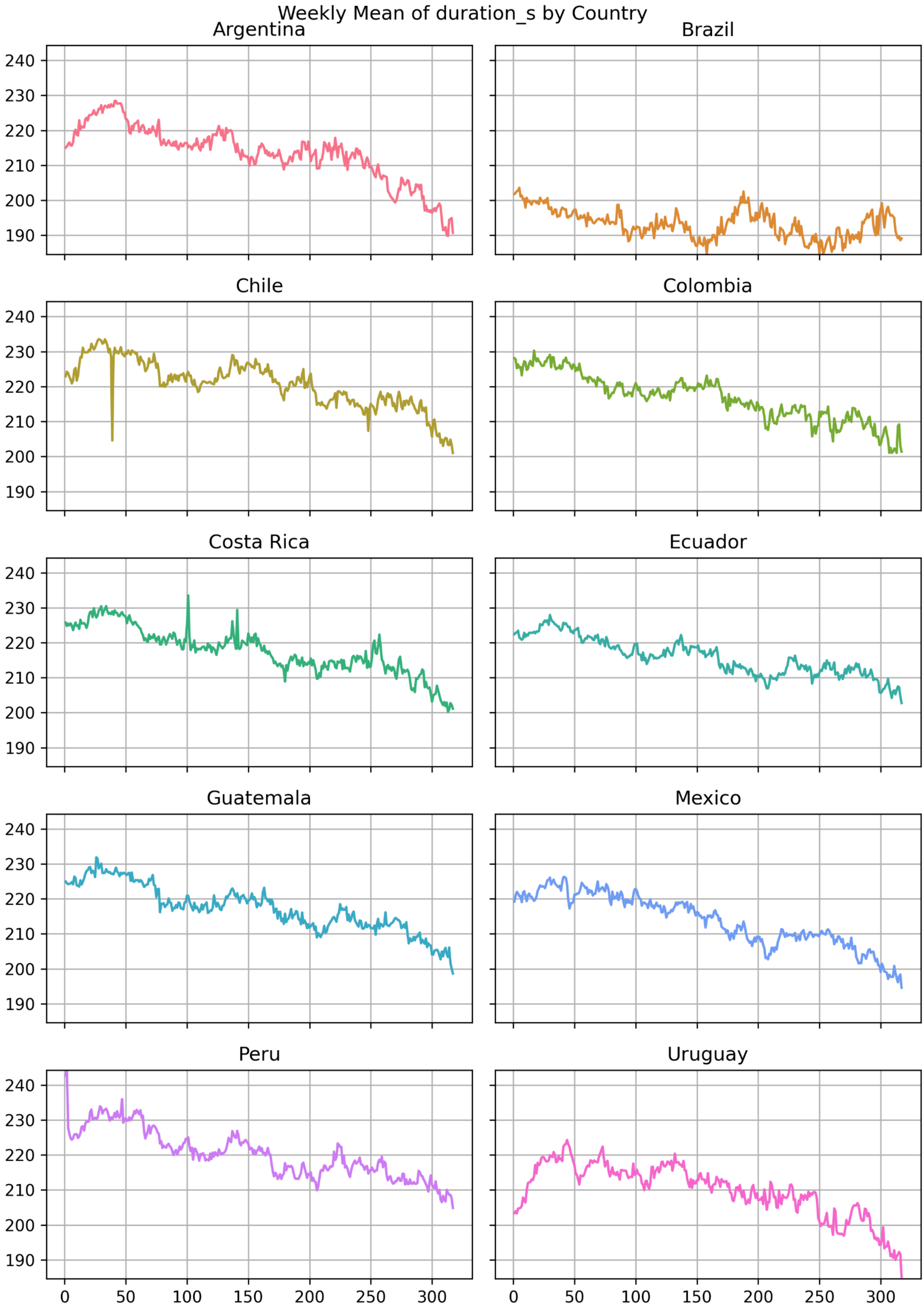
Overall Trend - Weekly Mean of tempo
Overall Trend
Mean Value
Week
135
130
125
120
115
0
50
100
150
200
250
300


Weekly Mean of tempo by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
135
130
125
120
115
0
50
100
150
200
250
300


All Countries - Weekly Mean of duration_s
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
Mean Value
Week
240
230
220
210
200
190
0
50
100
150
200
250
300



Overall Trend - Weekly Mean of duration_s
Overall Trend
240
230
220
210
200
190
Mean Value
0
50
100
150
200
250
300
Week


Weekly Mean of duration_s by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
240
230
220
210
200
190
0
50
100
150
200
250
300

14.6. Appendix F – Musicological characteristics by label: tables

Table F1. Descriptive statistics – Musicological characteristics by label affiliation

		parent_label			
		Indie Label	Sony Music Entertainment	Universal Music Group	Warner Music Group
key	Mean	5.18	5.73	5.11	5.41
	Median	5.00	6.00	5.00	6.00
	Mode	1	11	0	1
	Std. Deviation	3.71	3.62	3.77	3.43
	Range	11	11	11	11
	Minimum	0	0	0	0
	Maximum	11	11	11	11
danceability	Mean	72.86	71.73	69.38	73.25
	Median	74.30	74.20	70.90	75.90
	Mode	79.50	79.10	71.40	74.40
	Std. Deviation	11.20	11.02	11.84	11.34
	Range	90.67	74.60	87.64	87.91
	Minimum	7.83	23.40	8.66	8.59
	Maximum	98.50	98.00	96.30	96.50
energy	Mean	68.11	70.87	67.73	68.46
	Median	69.30	73.10	70.60	71.00
	Mode	67.20	77.30	64.60	80.40
	Std. Deviation	14.19	13.82	16.27	14.02
	Range	97.23	85.60	95.73	96.74
	Minimum	2.17	13.30	3.07	3.16
	Maximum	99.40	98.90	98.80	99.90
loudness	Mean	-5.34	-4.85	-5.57	-5.35
	Median	-5.01	-4.52	-5.16	-5.08
	Mode	-4.77	-4.21	-7.12	-6.33
	Std. Deviation	2.13	1.80	2.24	1.84
	Range	24.59	16.97	23.64	20.17
	Minimum	-22.69	-16.80	-23.02	-20.19
	Maximum	1.91	.18	.61	-.02
mode	Mean	.56	.55	.62	.60
	Median	1.00	1.00	1.00	1.00
	Mode	1	1	1	1
	Std. Deviation	.50	.50	.48	.49
	Range	1	1	1	1
	Minimum	0	0	0	0
	Maximum	1	1	1	1
speechiness	Mean	11.94	10.35	9.99	9.09
	Median	8.12	7.11	6.38	6.75
	Mode	4.32	4.83	11.90	6.77
	Std. Deviation	9.69	8.32	8.42	7.38
	Range	85.96	46.83	69.43	50.68
	Minimum	2.44	2.37	2.37	2.32
	Maximum	88.40	49.20	71.80	53.00
acousticness	Mean	27.29	23.35	23.10	22.32
	Median	21.50	18.20	15.60	16.80
	Mode	14.50	39.00	17.60	2.31
	Std. Deviation	21.60	20.00	22.29	21.11
	Range	99.20	94.70	98.60	97.00
	Minimum	.00	.00	.00	.00
	Maximum	99.20	94.70	98.60	97.00

instrumentalness	Mean	.33	.36	.67	.24
	Median	.00	.00	.00	.00
	Mode	.00	.00	.00	.00
	Std. Deviation	3.52	3.11	4.51	2.61
	Range	91.80	83.80	92.10	99.00
	Minimum	.00	.00	.00	.00
	Maximum	91.80	83.80	92.10	99.00
liveness	Mean	17.73	18.98	18.26	15.25
	Median	11.90	13.20	12.20	10.60
	Mode	10.10	10.30	10.80	4.94
	Std. Deviation	15.33	15.58	15.54	12.30
	Range	97.19	96.73	96.60	95.16
	Minimum	1.81	2.17	1.90	1.34
	Maximum	99.00	98.90	98.50	96.50
valence	Mean	62.02	64.15	60.67	60.36
	Median	65.60	66.70	62.90	62.40
	Mode	58.00	59.20	68.00	42.60
	Std. Deviation	21.21	20.44	20.77	20.84
	Range	95.30	94.30	92.43	93.60
	Minimum	3.60	3.20	5.17	3.80
	Maximum	98.90	97.50	97.60	97.40
tempo	Mean	123.91	124.90	123.28	118.80
	Median	115.00	114.68	115.93	105.14
	Mode	90.01	94.00	176.09	104.82
	Std. Deviation	33.96	34.50	32.95	29.44
	Range	161.11	149.02	163.32	155.41
	Minimum	48.75	63.42	50.70	54.75
	Maximum	209.86	212.44	214.03	210.16
duration_s	Mean	219.69	208.86	206.71	216.50
	Median	210.35	206.08	203.00	212.46
	Mode	309.12	174.00	204.35	205.72
	Std. Deviation	60.83	37.34	36.58	37.90
	Range	3603.39	850.48	503.64	616.13
	Minimum	50.57	89.19	33.87	53.03
	Maximum	3653.96	939.67	537.51	669.16
time_signature	Mean	3.96	3.98	3.96	3.94
	Median	4.00	4.00	4.00	4.00
	Mode	4	4	4	4
	Std. Deviation	.28	.21	.25	.26
	Range	4	4	4	4
	Minimum	1	1	1	1
	Maximum	5	5	5	5

14.7. Appendix G – Diversity indices: tables and graphs

Table G1. Diversity indices - all countries

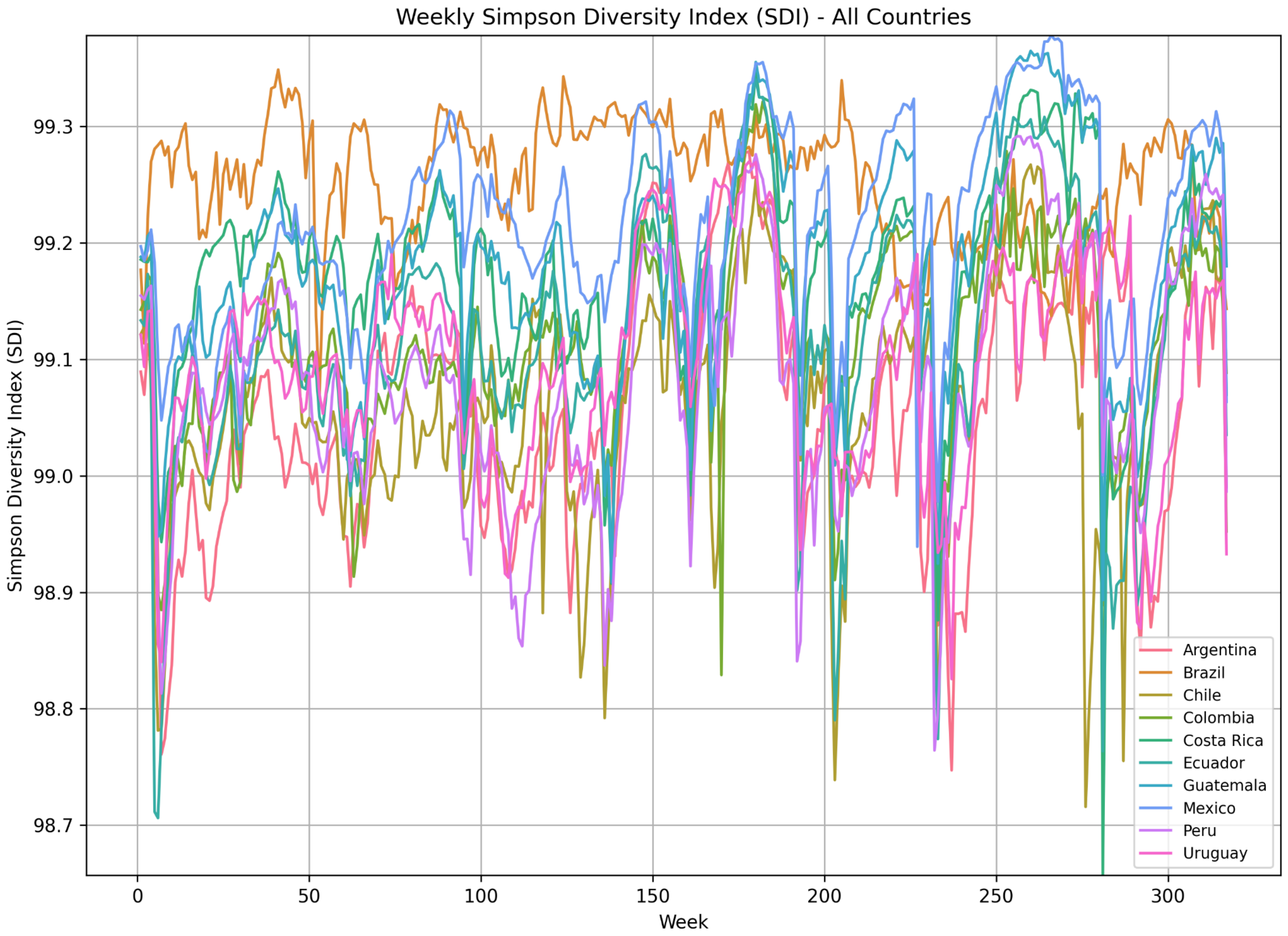
	simpson_diversity	shannon_index	mean_coefvar	rao_stirling
Mean	99.14	5.04	35.95	22.16
Median	99.14	5.04	35.88	22.05
Std. Deviation	.11	.06	1.39	1.08
Range	.72	.43	14.55	6.88
Minimum	98.66	4.77	32.54	18.62
Maximum	99.38	5.20	47.10	25.50

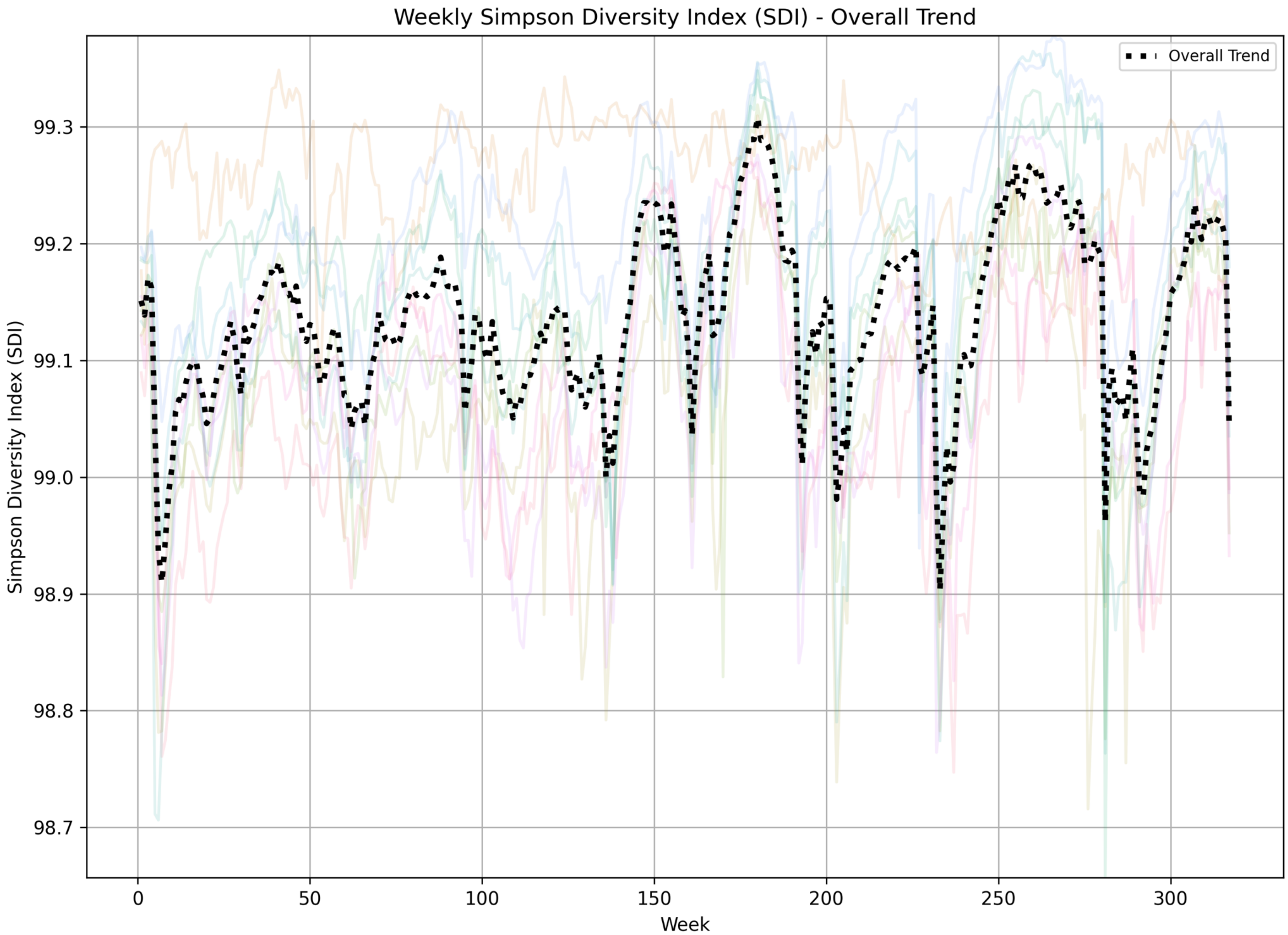
[a] A single significant outlier was found in Peru and removed in further uses of the measure.

Table G2. Diversity indices - by country

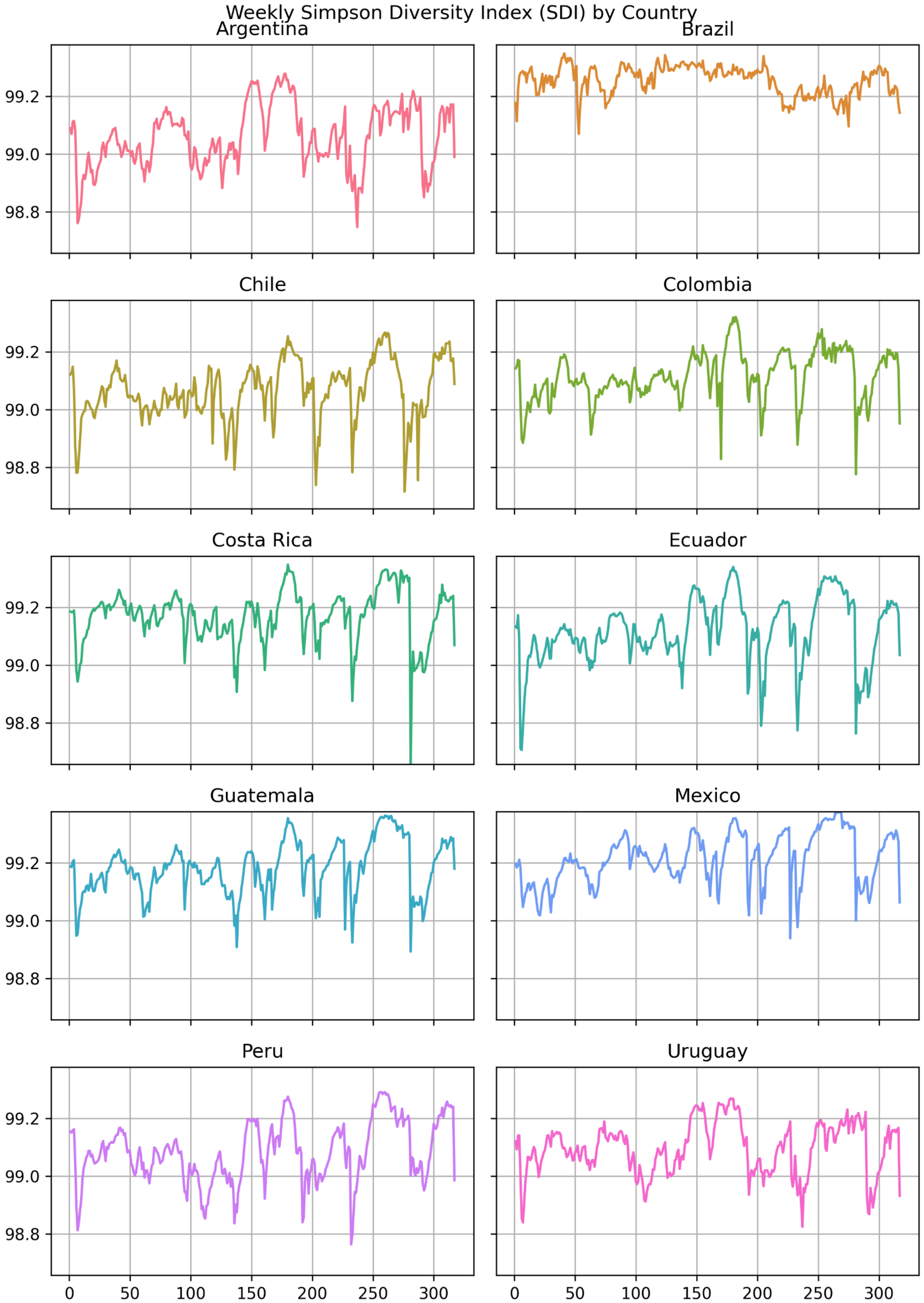
		country_name									
		Argentina	Brazil	Chile	Colombia	Costa Rica	Ecuador	Guatemala	Mexico	Peru	Uruguay
simpson_diversity	Mean	99.06	99.25	99.06	99.12	99.17	99.12	99.18	99.21	99.08	99.10
	Median	99.05	99.27	99.07	99.11	99.18	99.11	99.18	99.22	99.08	99.10
	Std. Deviation	.11	.05	.10	.09	.09	.11	.09	.09	.10	.09
	Range	.53	.28	.55	.55	.69	.63	.47	.44	.53	.44
	Minimum	98.75	99.07	98.72	98.78	98.66	98.71	98.89	98.94	98.76	98.83
	Maximum	99.28	99.35	99.27	99.32	99.35	99.34	99.36	99.38	99.29	99.27
shannon_index	Mean	5.00	5.11	5.00	5.03	5.07	5.04	5.08	5.10	5.01	5.02
	Median	4.99	5.12	5.00	5.02	5.07	5.03	5.07	5.10	5.00	5.01
	Std. Deviation	.06	.03	.05	.05	.05	.06	.05	.05	.06	.05
	Range	.27	.16	.26	.35	.40	.36	.30	.25	.26	.23
	Minimum	4.86	5.03	4.86	4.81	4.77	4.81	4.89	4.95	4.88	4.89
	Maximum	5.13	5.19	5.12	5.15	5.17	5.16	5.19	5.20	5.13	5.12

mean_coefvar	Mean	36.37	37.71	35.02	35.36	36.44	35.33	35.40	36.81	35.27	35.75
	Median	36.29	37.60	34.94	35.21	36.30	35.03	35.13	36.77	34.98	35.75
	Std. Deviation	.95	1.14	1.22	1.08	1.16	1.12	1.12	.95	1.55	.75
	Range	4.90	4.93	5.50	5.45	5.94	4.84	5.00	5.91	14.04	4.24
	Minimum	33.68	35.38	32.54	32.82	34.14	33.30	33.26	34.25	33.05	33.46
	Maximum	38.58	40.31	38.04	38.27	40.08	38.14	38.26	40.16	47.10	37.70
rao_stirling	Mean	21.75	23.77	21.86	21.92	22.26	21.62	22.25	22.90	21.63	21.63
	Median	21.73	23.78	21.90	22.02	22.38	21.73	22.14	22.79	21.73	21.52
	Std. Deviation	.80	.80	.72	.77	.75	.70	.94	1.24	.81	.86
	Range	4.49	4.06	3.84	4.11	3.80	3.83	4.45	5.39	4.60	4.63
	Minimum	19.36	21.44	19.73	19.42	20.01	19.36	19.92	20.06	18.62	19.33
	Maximum	23.84	25.50	23.56	23.53	23.81	23.19	24.37	25.46	23.21	23.95

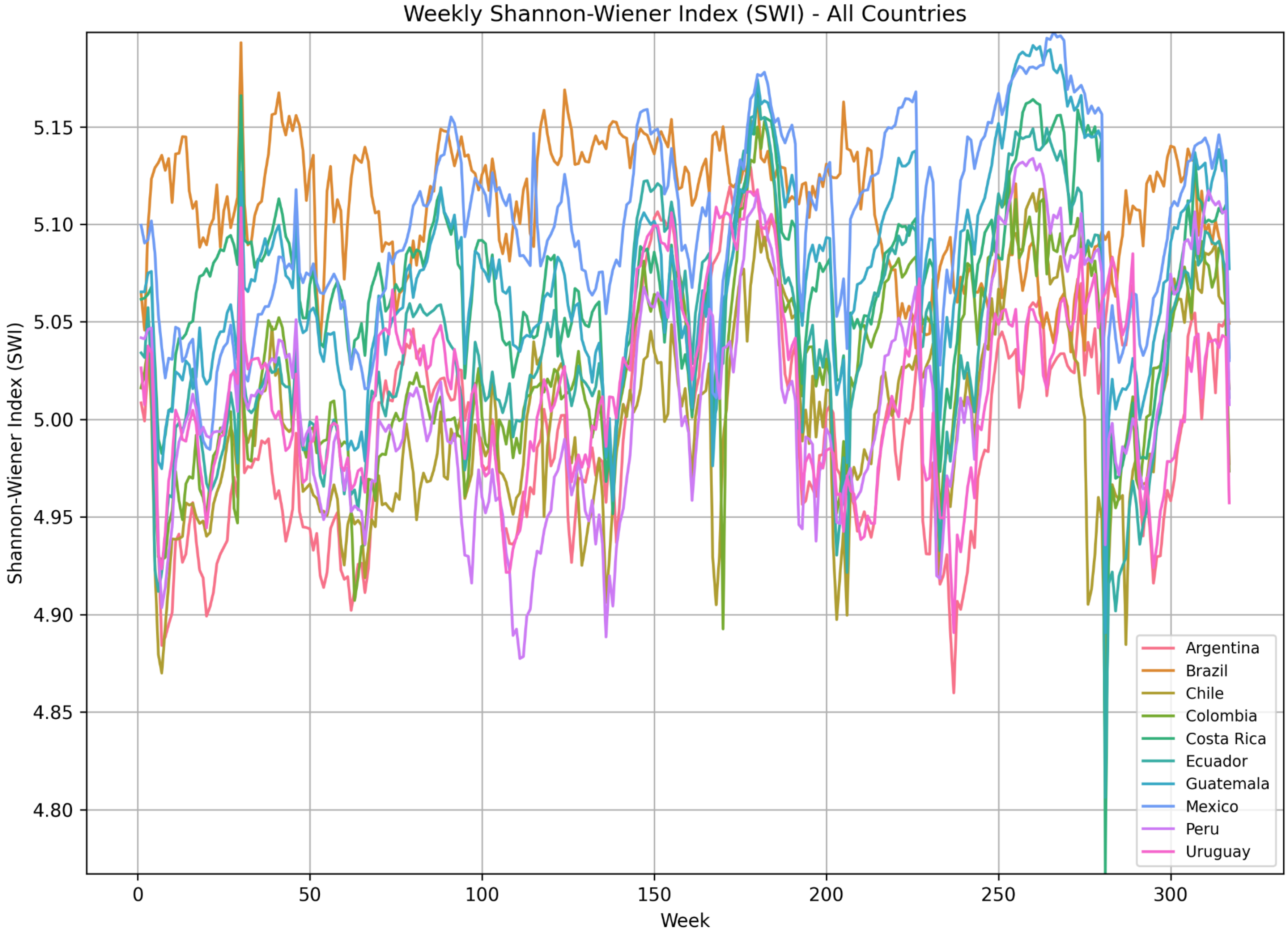

Weekly Simpson Diversity Index (SDI) - All Countries
Simpson Diversity Index (SDI)
99.3
99.2
99.1
99.0
98.9
98.8
98.7
0
50
100
150
200
250
300
Week
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay



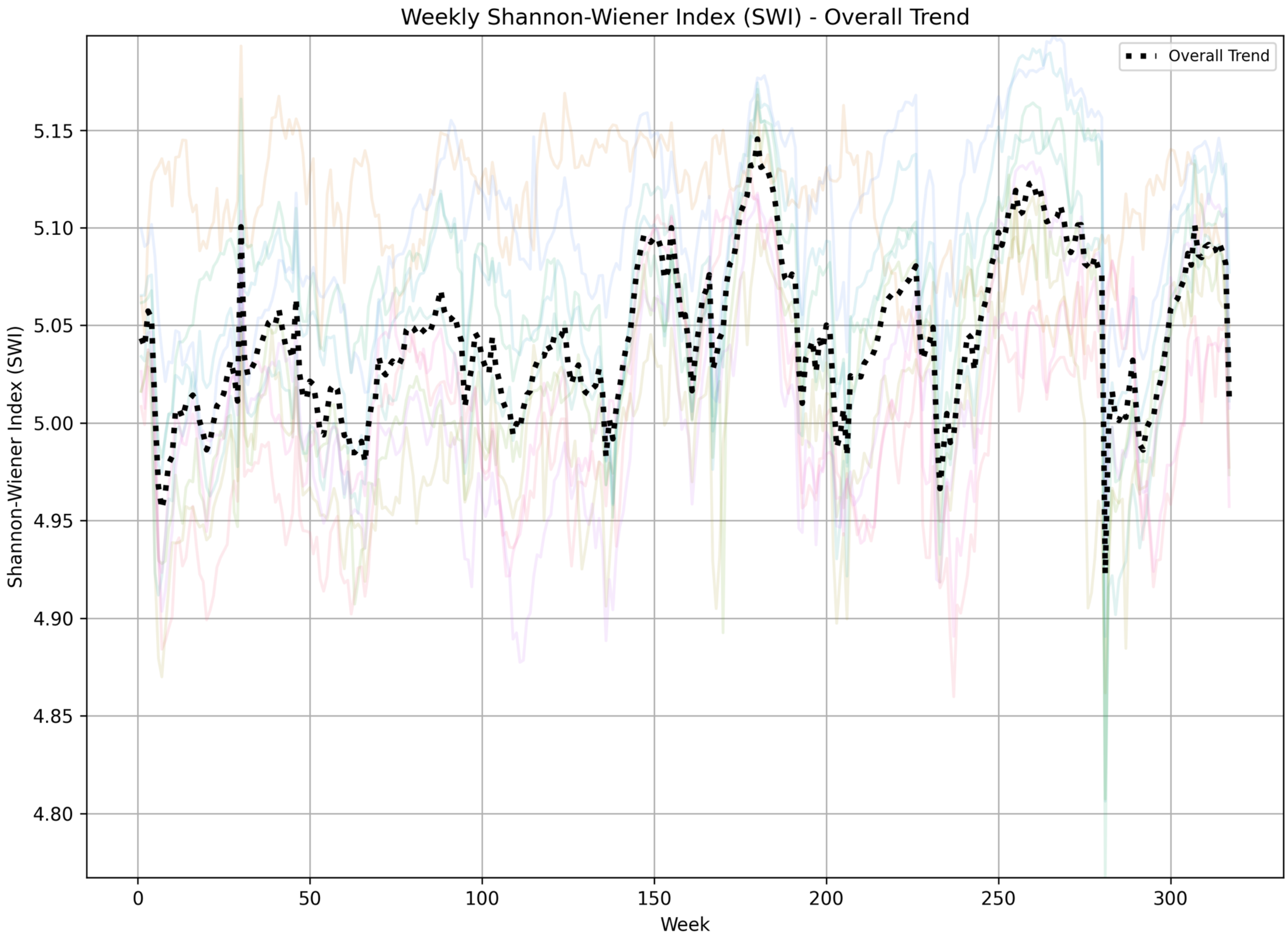
Weekly Simpson Diversity Index (SDI) - Overall Trend
Overall Trend
Simpson Diversity Index (SDI)
99.3
99.2
99.1
99.0
98.9
98.8
98.7
0
50
100
150
200
250
300
Week

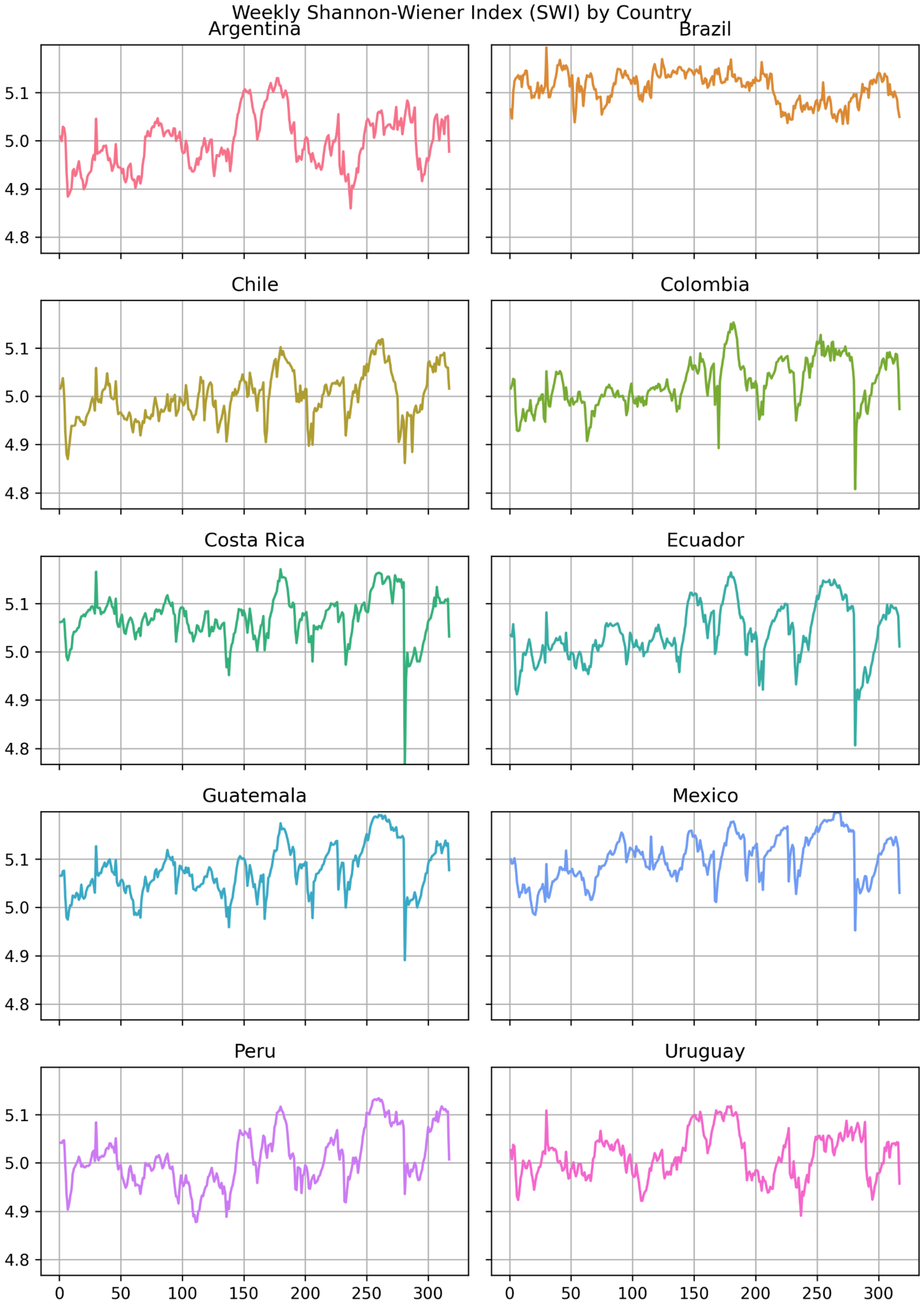


Weekly Simpson Diversity Index (SDI) by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
98.8
99.0
99.2
0
50
100
150
200
250
300

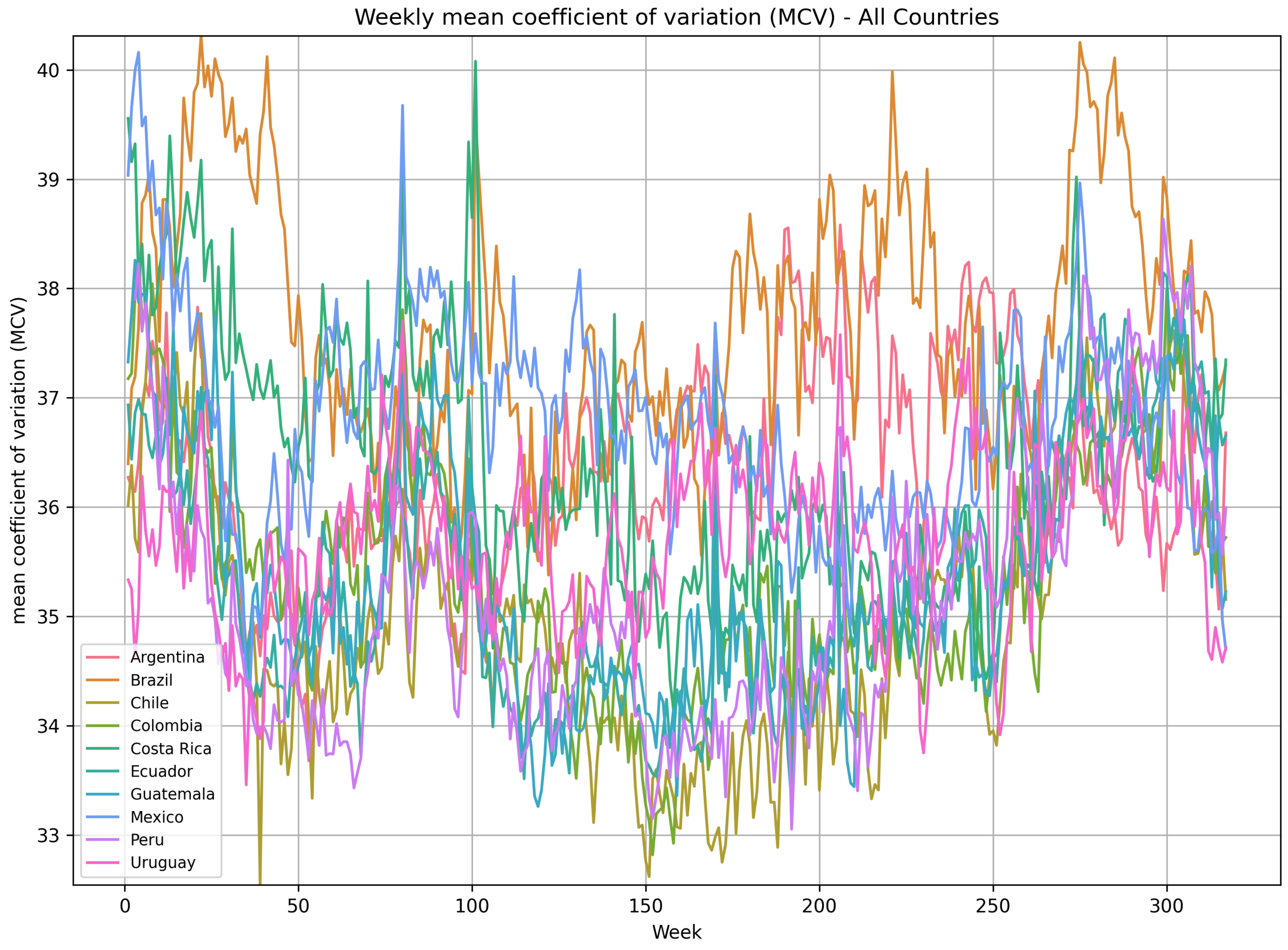


Weekly Shannon-Wiener Index (SWI) - All Countries
Shannon-Wiener Index (SWI)
5.15
5.10
5.05
5.00
4.95
4.90
4.85
4.80
0
50
100
150
200
250
300
Week
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay

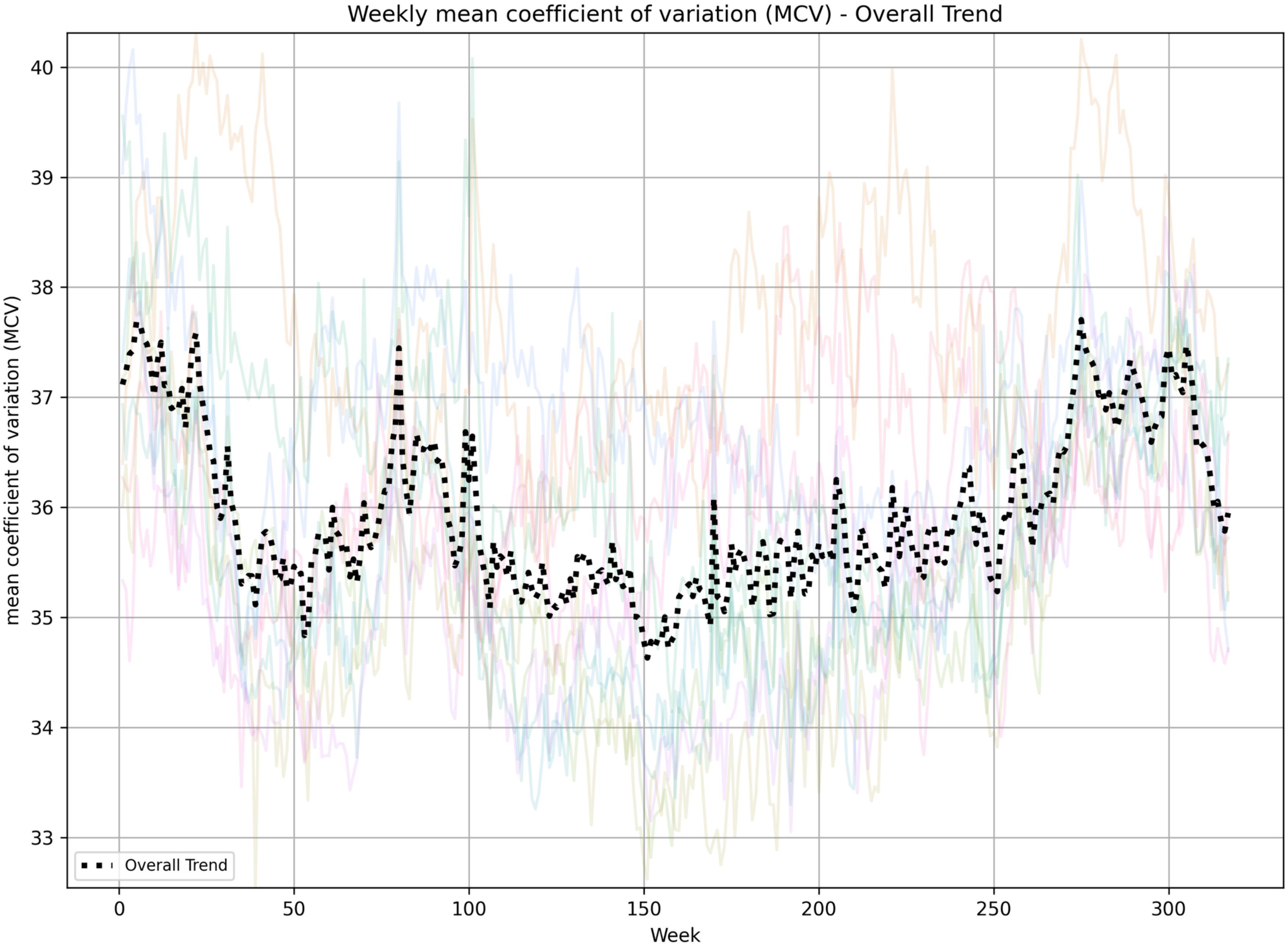

Weekly Shannon-Wiener Index (SWI) - Overall Trend
Overall Trend
Shannon-Wiener Index (SWI)
5.15
5.10
5.05
5.00
4.95
4.90
4.85
4.80
Week
0
50
100
150
200
250
300

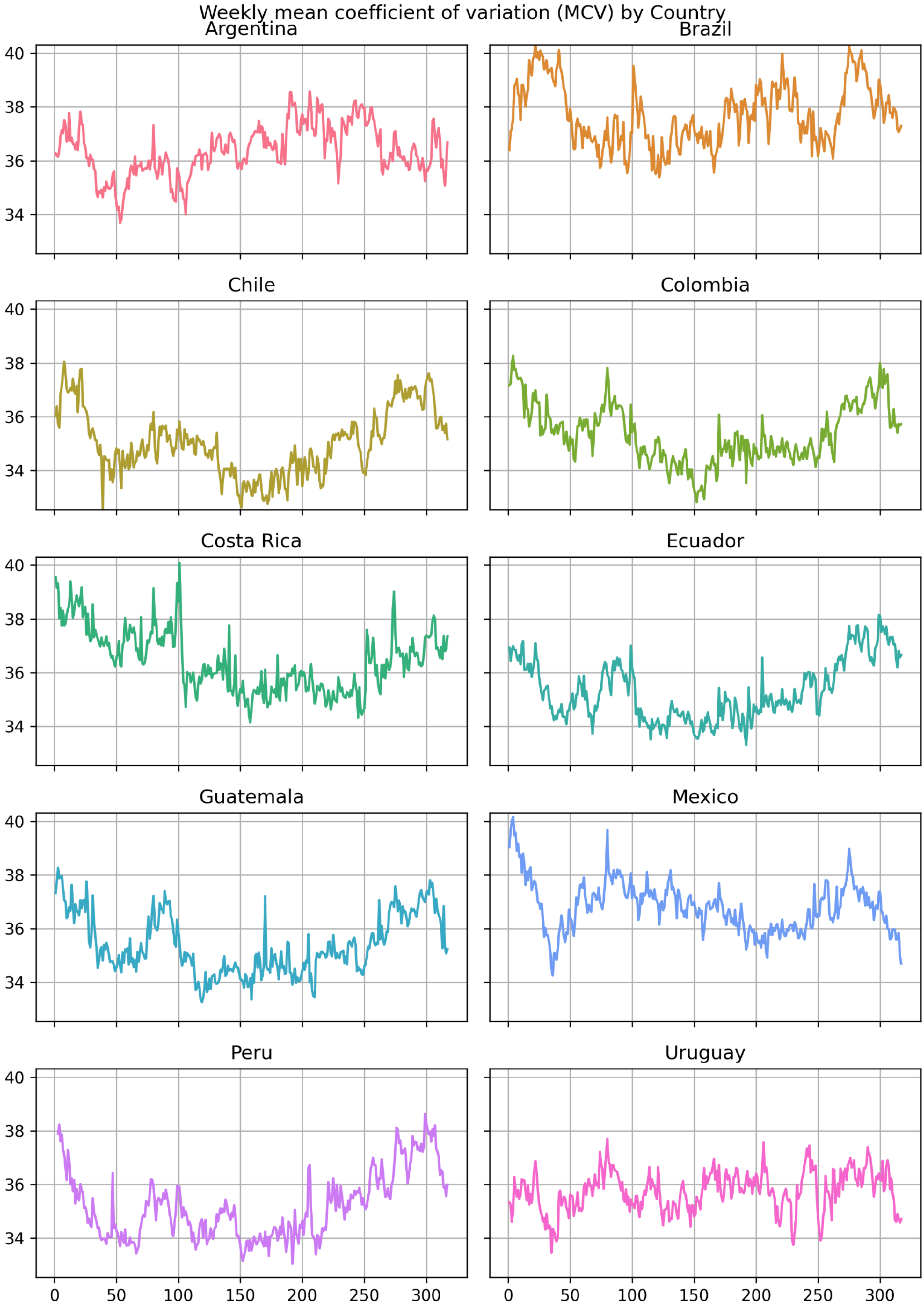


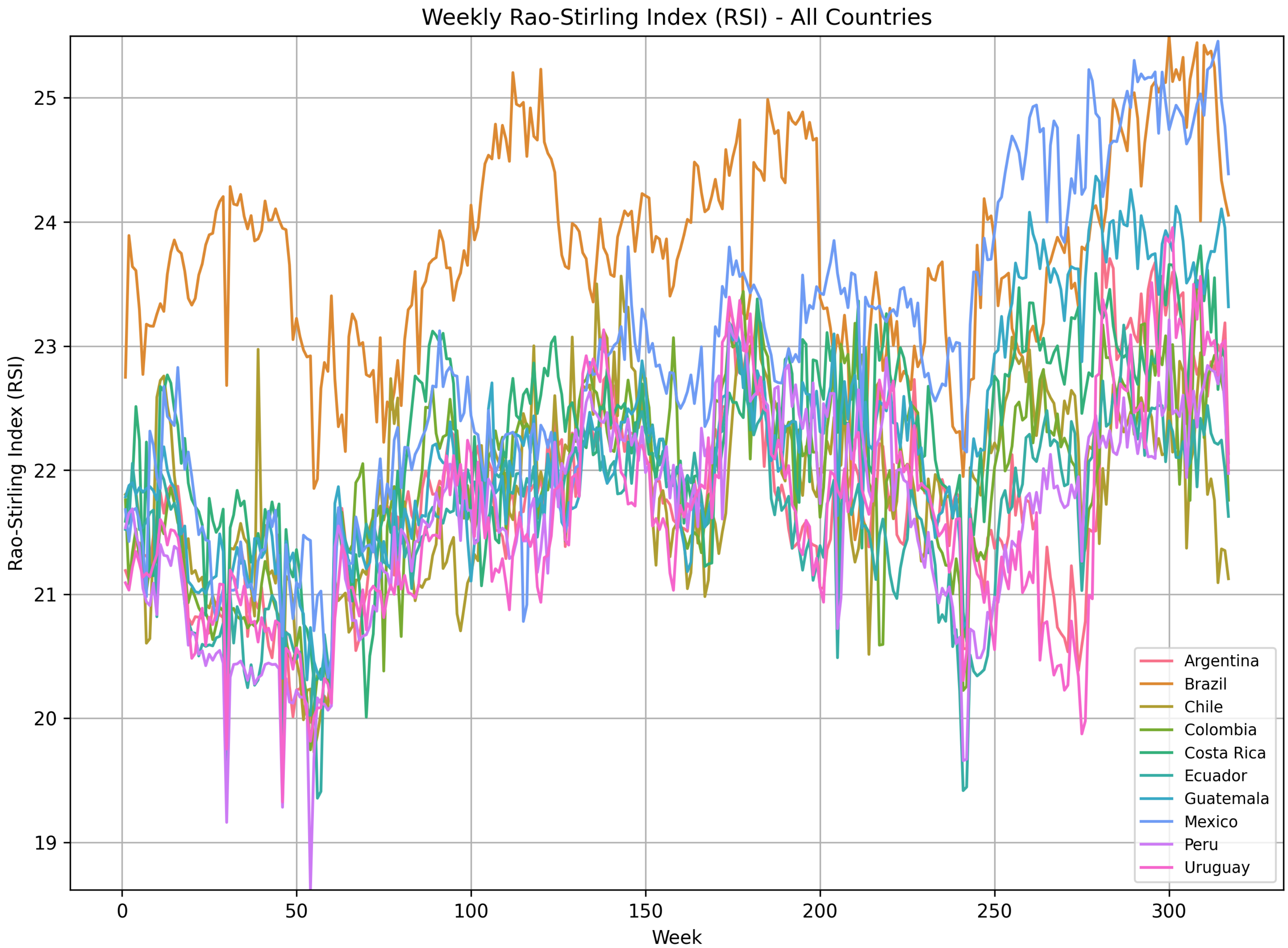
Weekly Shannon-Wiener Index (SWI) by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
5.1
5.0
4.9
4.8
0
50
100
150
200
250
300



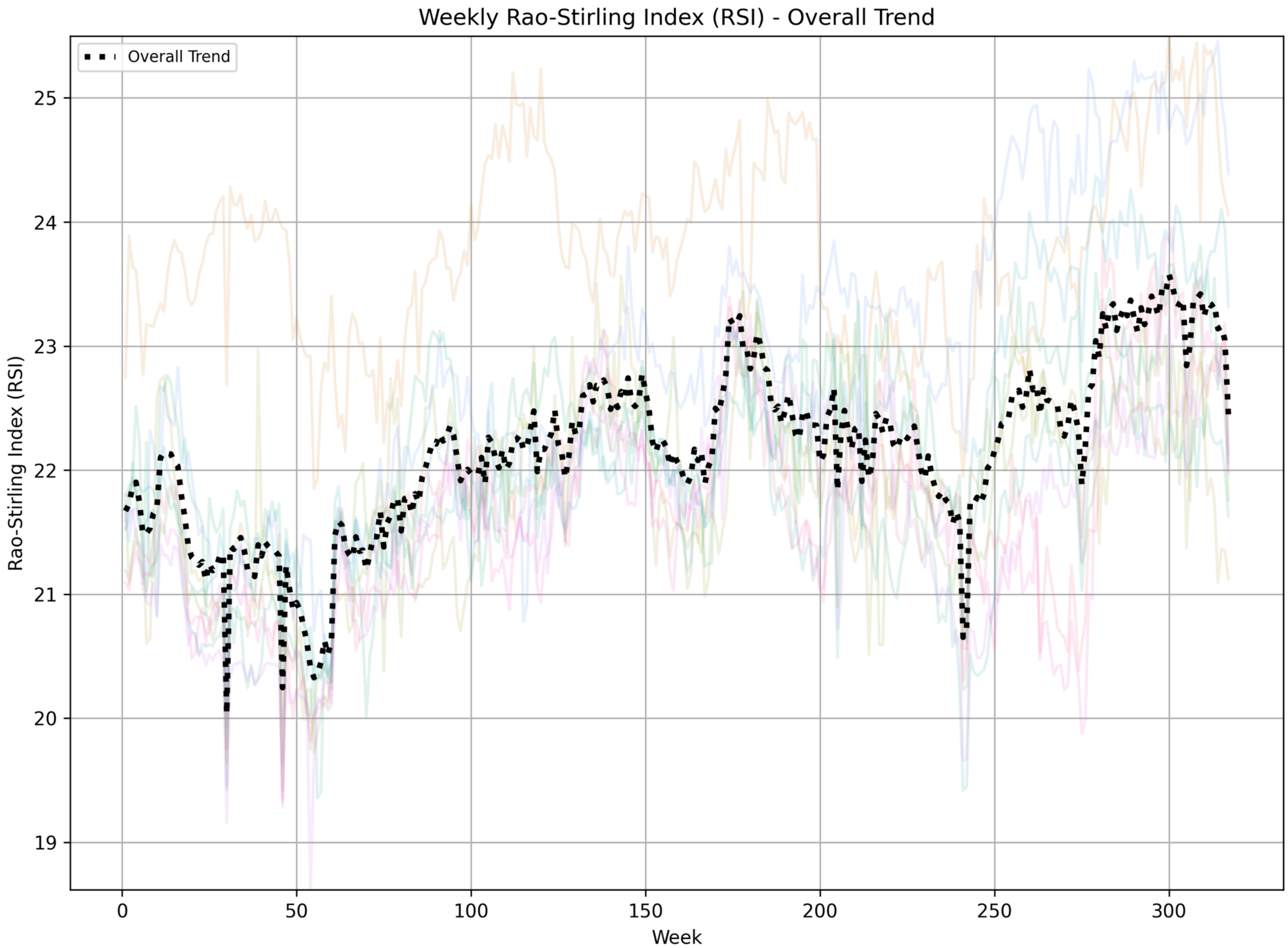
Weekly mean coefficient of variation (MCV) - All Countries
mean coefficient of variation (MCV)
Week
40
39
38
37
36
35
34
33
0
50
100
150
200
250
300
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay


Weekly mean coefficient of variation (MCV) - Overall Trend
mean coefficient of variation (MCV)
Week
Overall Trend
40
39
38
37
36
35
34
33
0
50
100
150
200
250
300

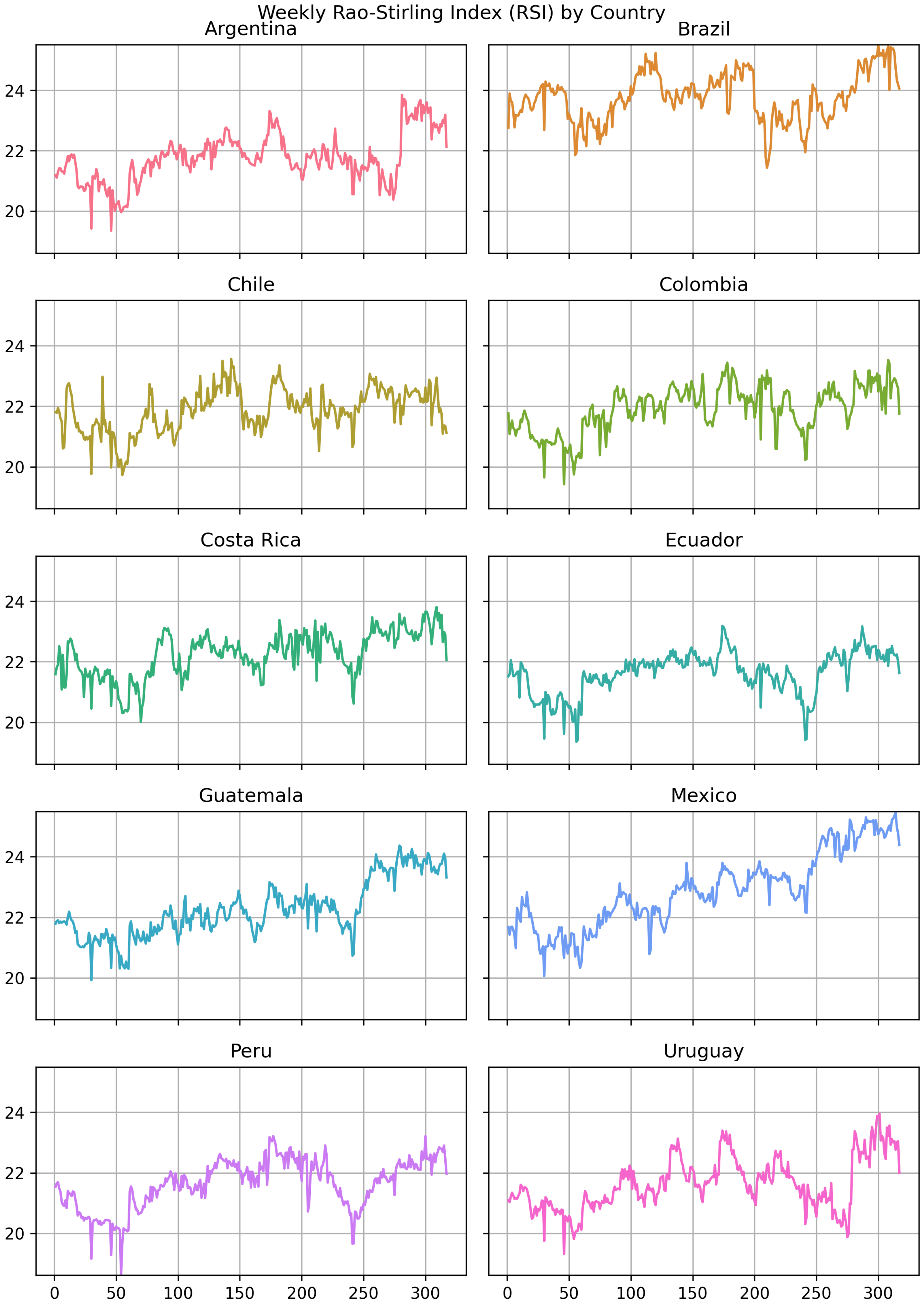

Weekly mean coefficient of variation (MCV) by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
40
38
36
34
0
50
100
150
200
250
300



Weekly Rao-Stirling Index (RSI) - All Countries
Rao-Stirling Index (RSI)
Week
19
20
21
22
23
24
25
0
50
100
150
200
250
300
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay



Weekly Rao-Stirling Index (RSI) - Overall Trend
Overall Trend
Rao-Stirling Index (RSI)
Week
19
20
21
22
23
24
25
0
50
100
150
200
250
300


Weekly Rao-Stirling Index (RSI) by Country
Argentina
Brazil
Chile
Colombia
Costa Rica
Ecuador
Guatemala
Mexico
Peru
Uruguay
20
22
24
0
50
100
150
200
250
300

14.8. Appendix H – Diagnostic tests for model assumptions

Variable Inflation Factor

Model 1:

Variable	VIF
const	119.2398
share_st_majors	1.611951
gni_percapita_ppp_2017	2.679215
internet_access	3.541949
sum_st	1.11104

Model 2:

Variable	VIF
const	123.4514
share_st_universal	1.52693
share_st_sony	1.140444
share_st_warner	1.421754
gni_percapita_ppp_2017	2.758256
internet_access	3.684697
sum_st	1.124907

Wooldridge Test for Autocorrelation in Panel Data

Model 1:

Test statistic: 1.3547142072150193
p-value: 0.17550862961022462

Model 2:

Test statistic: 1.3407448066833108
p-value: 0.18000332156291654

Wooldridge Test for Heteroskedasticity in Panel Data

Model 1:

Test statistic: 189.17284151418377
p-value: 6.131855913237286e-36

Model 2:

Test statistic: 344.6642904197995
p-value: 1.0277730708483801e-65

14.9. Appendix I – Detailed regression output

Model 1

```
Mean Absolute Error (MAE) for model 1: 1.2333027343879508
Root Mean Squared Error (RMSE) for model 2: 1.4874814185456193

Regression results for model 1:
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:           rao_stirling   R-squared:                        0.3424
Estimator:                   PanelOLS   R-squared (Between):             -3.0823
No. Observations:                3158   R-squared (Within):               0.3424
Date:                Thu, Nov 16 2023   R-squared (Overall):             -0.9269
Time:                        22:16:50   Log-likelihood                   -3300.6
Cov. Estimator:                Robust
                                        F-statistic:                      409.31
Entities:                          10   P-value                           0.0000
Avg Obs:                       315.80   Distribution:                  F(4,3144)
Min Obs:                       313.00
Max Obs:                       317.00   F-statistic (robust):             393.92
                                        P-value                           0.0000
Time periods:                     317   Distribution:                  F(4,3144)
Avg Obs:                       9.9621
Min Obs:                       7.0000
Max Obs:                      10.0000

                                  Parameter Estimates
=========================================================================================
                        Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
-----------------------------------------------------------------------------------------
const                      18.778     0.5348     35.112     0.0000      17.730      19.827
share_st_majors           -0.0246     0.0018    -13.778     0.0000     -0.0281     -0.0211
gni_percapita_ppp_2017     0.0001  2.448e-05     5.5950     0.0000   8.899e-05      0.0002
internet_access            0.0355     0.0030     11.766     0.0000      0.0296      0.0415
sum_st                  2.466e-09  5.554e-10     4.4394     0.0000   1.377e-09   3.555e-09
=========================================================================================

F-test for Poolability: 100.26
P-value: 0.0000
Distribution: F(9,3144)

Included effects: Entity
```

Model 2

```
Mean Absolute Error (MAE) for Major Conglomerates Regression: 1.0014874918479337
Root Mean Squared Error (RMSE) for Major Conglomerates Regression: 1.2381195462781485

Regression Results for model 2:
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:           rao_stirling   R-squared:                        0.3856
Estimator:                   PanelOLS   R-squared (Between):             -1.5637
No. Observations:                3158   R-squared (Within):               0.3856
Date:                Thu, Nov 16 2023   R-squared (Overall):             -0.3350
Time:                        22:16:51   Log-likelihood                   -3193.3
Cov. Estimator:                Robust
                                        F-statistic:                      328.71
Entities:                          10   P-value                           0.0000
Avg Obs:                       315.80   Distribution:                  F(6,3142)
Min Obs:                       313.00
Max Obs:                       317.00   F-statistic (robust):             357.50
                                        P-value                           0.0000
Time periods:                     317   Distribution:                  F(6,3142)
Avg Obs:                       9.9621
Min Obs:                       7.0000
Max Obs:                      10.0000

                                  Parameter Estimates
=========================================================================================
```

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	18.657	0.5496	33.949	0.0000	17.580	19.735
share_st_universal	0.0132	0.0034	3.9016	0.0001	0.0066	0.0199
share_st_sony	-0.0442	0.0028	-15.759	0.0000	-0.0497	-0.0387
share_st_warner	-0.0209	0.0033	-6.3523	0.0000	-0.0273	-0.0144
gni_percapita_ppp_2017	8.724e-05	2.518e-05	3.4652	0.0005	3.788e-05	0.0001
internet_access	0.0417	0.0030	13.882	0.0000	0.0358	0.0476
sum_st	4.697e-09	5.381e-10	8.7288	0.0000	3.642e-09	5.752e-09

```
F-test for Poolability: 61.352
P-value: 0.0000
Distribution: F(9,3142)

Included effects: Entity
```

14.10. Appendix J – Missing entries

Table K1. Missing entries – Spotify Charts data		
Country	**Week number**	**Rank position(s)**
Argentina	30	29, 62, 76, 96
	46	2
	241	24
	242	16
Brazil	30	10, 39, 52, 89
	55	178
	56	47
	178	128
	179	93
	180	94
	209	75
	210	72
	211	55
	212	46
	213	36
	309	73
Chile	30	37, 76, 108, 127
	46	2
	241	11
	242	11
Colombia	30	23, 40, 52, 69
	46	3
	241	21
	242	15
Costa Rica	30	24, 33, 45, 69
	46	1
	241	23
	242	18
Ecuador	30	29, 56, 57, 81
	46	6
	241	18
	242	15
Guatemala	30	24, 40, 51, 59
	46	2
	241	21
	242	15
Mexico	1	158
	2	169
	3	185
	4	190
	5	189
	6	197
	7	199
	30	18, 45, 55, 81
	46	2
	115	98, 123, 157
	116	136
	241	10

	242	8
Peru	30	45, 75, 88, 95
	46	4
	241	14
	242	11
Uruguay	30	29, 42, 78, 102
	46	2
	241	22
	242	14

Table K2. Missing songs – Spotify Audio Features

Country	Week	Rank	Spotify URI	Artist names	Track name
Argentina	263	43	5saUoeceT2zau6ScmU77e6	Big Apple, Damas Gratis, Kaleb Di Masi, Homer El Mero Mero, Omar Varela	“Que a Pasao – Remix”
	264	38			
	265	41			
	266	45			
	267	35			
	268	38			
	269	37			
	270	44			
	271	44			
	272	49			
	273	71			
	274	86			
	275	106			
	276	117			
	277	129			
	278	135			
	279	170			
	280	181			
Colombia	217	77	6zQhJcyuZGX7ADNMZF1VHL	Feid	“14 De Febrero”
	218	121			
Peru	54	186	2DEYFawpGha5Zn54Fx6dX5	DJ Krlos Berrospi	“Año Nuevo 2018”
Uruguay	263	63	5saUoeceT2zau6ScmU77e6	Big Apple, Damas Gratis, Kaleb Di Masi, Homer El Mero Mero, Omar Varela	“Que a Pasao – Remix”
	264	55			
	265	53			
	266	58			
	267	59			
	268	63			
	269	68			
	270	74			
	271	81			
	272	88			
	273	103			
	274	112			
	275	138			

	276	148			
	277	153			
	278	164			

14.11. Appendix K – Weeks and starting dates

1	29-Dec-2016	40	28-Sep-2017	79	28-Jun-2018	118	28-Mar-2019
2	5-Jan-2017	41	5-Oct-2017	80	5-Jul-2018	119	4-Apr-2019
3	12-Jan-2017	42	12-Oct-2017	81	12-Jul-2018	120	11-Apr-2019
4	19-Jan-2017	43	19-Oct-2017	82	19-Jul-2018	121	18-Apr-2019
5	26-Jan-2017	44	26-Oct-2017	83	26-Jul-2018	122	25-Apr-2019
6	2-Feb-2017	45	2-Nov-2017	84	2-Aug-2018	123	2-May-2019
7	9-Feb-2017	46	9-Nov-2017	85	9-Aug-2018	124	9-May-2019
8	16-Feb-2017	47	16-Nov-2017	86	16-Aug-2018	125	16-May-2019
9	23-Feb-2017	48	23-Nov-2017	87	23-Aug-2018	126	23-May-2019
10	2-Mar-2017	49	30-Nov-2017	88	30-Aug-2018	127	30-May-2019
11	9-Mar-2017	50	7-Dec-2017	89	6-Sep-2018	128	6-Jun-2019
12	16-Mar-2017	51	14-Dec-2017	90	13-Sep-2018	129	13-Jun-2019
13	23-Mar-2017	52	21-Dec-2017	91	20-Sep-2018	130	20-Jun-2019
14	30-Mar-2017	53	28-Dec-2017	92	27-Sep-2018	131	27-Jun-2019
15	6-Apr-2017	54	4-Jan-2018	93	4-Oct-2018	132	4-Jul-2019
16	13-Apr-2017	55	11-Jan-2018	94	11-Oct-2018	133	11-Jul-2019
17	20-Apr-2017	56	18-Jan-2018	95	18-Oct-2018	134	18-Jul-2019
18	27-Apr-2017	57	25-Jan-2018	96	25-Oct-2018	135	25-Jul-2019
19	4-May-2017	58	1-Feb-2018	97	1-Nov-2018	136	1-Aug-2019
20	11-May-2017	59	8-Feb-2018	98	8-Nov-2018	137	8-Aug-2019
21	18-May-2017	60	15-Feb-2018	99	15-Nov-2018	138	15-Aug-2019
22	25-May-2017	61	22-Feb-2018	100	22-Nov-2018	139	22-Aug-2019
23	1-Jun-2017	62	1-Mar-2018	101	29-Nov-2018	140	29-Aug-2019
24	8-Jun-2017	63	8-Mar-2018	102	6-Dec-2018	141	5-Sep-2019
25	15-Jun-2017	64	15-Mar-2018	103	13-Dec-2018	142	12-Sep-2019
26	22-Jun-2017	65	22-Mar-2018	104	20-Dec-2018	143	19-Sep-2019
27	29-Jun-2017	66	29-Mar-2018	105	27-Dec-2018	144	26-Sep-2019
28	6-Jul-2017	67	5-Apr-2018	106	3-Jan-2019	145	3-Oct-2019
29	13-Jul-2017	68	12-Apr-2018	107	10-Jan-2019	146	10-Oct-2019
30	20-Jul-2017	69	19-Apr-2018	108	17-Jan-2019	147	17-Oct-2019
31	27-Jul-2017	70	26-Apr-2018	109	24-Jan-2019	148	24-Oct-2019
32	3-Aug-2017	71	3-May-2018	110	31-Jan-2019	149	31-Oct-2019
33	10-Aug-2017	72	10-May-2018	111	7-Feb-2019	150	7-Nov-2019
34	17-Aug-2017	73	17-May-2018	112	14-Feb-2019	151	14-Nov-2019
35	24-Aug-2017	74	24-May-2018	113	21-Feb-2019	152	21-Nov-2019
36	31-Aug-2017	75	31-May-2018	114	28-Feb-2019	153	28-Nov-2019
37	7-Sep-2017	76	7-Jun-2018	115	7-Mar-2019	154	5-Dec-2019
38	14-Sep-2017	77	14-Jun-2018	116	14-Mar-2019	155	12-Dec-2019
39	21-Sep-2017	78	21-Jun-2018	117	21-Mar-2019	156	19-Dec-2019

157	26-Dec-2019
158	2-Jan-2020
159	9-Jan-2020
160	16-Jan-2020
161	23-Jan-2020
162	30-Jan-2020
163	6-Feb-2020
164	13-Feb-2020
165	20-Feb-2020
166	27-Feb-2020
167	5-Mar-2020
168	12-Mar-2020
169	19-Mar-2020
170	26-Mar-2020
171	2-Apr-2020
172	9-Apr-2020
173	16-Apr-2020
174	23-Apr-2020
175	30-Apr-2020
176	7-May-2020
177	14-May-2020
178	21-May-2020
179	28-May-2020
180	4-Jun-2020
181	11-Jun-2020
182	18-Jun-2020
183	25-Jun-2020
184	2-Jul-2020
185	9-Jul-2020
186	16-Jul-2020
187	23-Jul-2020
188	30-Jul-2020
189	6-Aug-2020
190	13-Aug-2020
191	20-Aug-2020
192	27-Aug-2020
193	3-Sep-2020
194	10-Sep-2020
195	17-Sep-2020
196	24-Sep-2020
197	1-Oct-2020
198	8-Oct-2020
199	15-Oct-2020
200	22-Oct-2020
201	29-Oct-2020
202	5-Nov-2020
203	12-Nov-2020
204	19-Nov-2020
205	26-Nov-2020
206	3-Dec-2020
207	10-Dec-2020
208	17-Dec-2020
209	24-Dec-2020
210	31-Dec-2020
211	7-Jan-2021
212	14-Jan-2021
213	21-Jan-2021
214	28-Jan-2021
215	4-Feb-2021
216	11-Feb-2021
217	18-Feb-2021
218	25-Feb-2021
219	4-Mar-2021
220	11-Mar-2021
221	18-Mar-2021
222	25-Mar-2021
223	1-Apr-2021
224	8-Apr-2021
225	15-Apr-2021
226	22-Apr-2021
227	29-Apr-2021
228	6-May-2021
229	13-May-2021
230	20-May-2021
231	27-May-2021
232	3-Jun-2021
233	10-Jun-2021
234	17-Jun-2021
235	24-Jun-2021
236	1-Jul-2021
237	8-Jul-2021
238	15-Jul-2021
239	22-Jul-2021
240	29-Jul-2021
241	5-Aug-2021
242	12-Aug-2021
243	19-Aug-2021
244	26-Aug-2021
245	2-Sep-2021
246	9-Sep-2021
247	16-Sep-2021
248	23-Sep-2021
249	30-Sep-2021
250	7-Oct-2021
251	14-Oct-2021
252	21-Oct-2021
253	28-Oct-2021
254	4-Nov-2021
255	11-Nov-2021
256	18-Nov-2021
257	25-Nov-2021
258	2-Dec-2021
259	9-Dec-2021
260	16-Dec-2021
261	23-Dec-2021
262	30-Dec-2021
263	6-Jan-2022
264	13-Jan-2022
265	20-Jan-2022
266	27-Jan-2022
267	3-Feb-2022
268	10-Feb-2022
269	17-Feb-2022
270	24-Feb-2022
271	3-Mar-2022
272	10-Mar-2022
273	17-Mar-2022
274	24-Mar-2022
275	31-Mar-2022
276	7-Apr-2022
277	14-Apr-2022
278	21-Apr-2022
279	28-Apr-2022
280	5-May-2022
281	12-May-2022
282	19-May-2022
283	26-May-2022
284	2-Jun-2022
285	9-Jun-2022
286	16-Jun-2022
287	23-Jun-2022
288	30-Jun-2022
289	7-Jul-2022
290	14-Jul-2022
291	21-Jul-2022
292	28-Jul-2022
293	4-Aug-2022
294	11-Aug-2022
295	18-Aug-2022
296	25-Aug-2022
297	1-Sep-2022
298	8-Sep-2022
299	15-Sep-2022
300	22-Sep-2022
301	29-Sep-2022
302	6-Oct-2022
303	13-Oct-2022
304	20-Oct-2022
305	27-Oct-2022
306	3-Nov-2022
307	10-Nov-2022
308	17-Nov-2022
309	24-Nov-2022
310	1-Dec-2022
311	8-Dec-2022
312	15-Dec-2022
313	22-Dec-2022
314	29-Dec-2022
315	5-Jan-2023
316	12-Jan-2023
317	19-Jan-2023