# Spotify Streaming Data Visualization

Kyle Albert

This project is anaylzes my extended listening history from Spotify. This data includes all songs I have streamed on Spotify, starting back in 2013. Additionally, after requesting access to the Web API, I was able to pull information about each song which provided further interesting analysis. Some of the additional variables I was able to access are Danceability, a measure of how suitable a track is for dancing, between 0 and 1 where 0 is least danceable and 1 is most danceable, Energy, a measure of intensity and activity, something like death metal would score close to 1 while a Bach prelude would score closer to 0, and Valence, the musical positiveness of a track, 0 is more negative (sad, depressed, angry, etc.) and 1 is more positive (happy, cheerful, euphoric, etc.).

This analysis focuses mainly on data visualizations and examining trends in my data, seeking interesting patterns. This data set provides many unique ways to examine it, there is no clear endpoint with which to stop this Exploratory Data Analysis. I plan on continuing this data exploration in the future, where I will want to try and take on more of a behavior prediction approach, but for the scope of this project I mainly focus on trying to understand and get an appreciation for the music I have listened to in the past.

One last note before examining figures is that this data was requested in September, so there is no streaming data for October, November, or December for 2022.

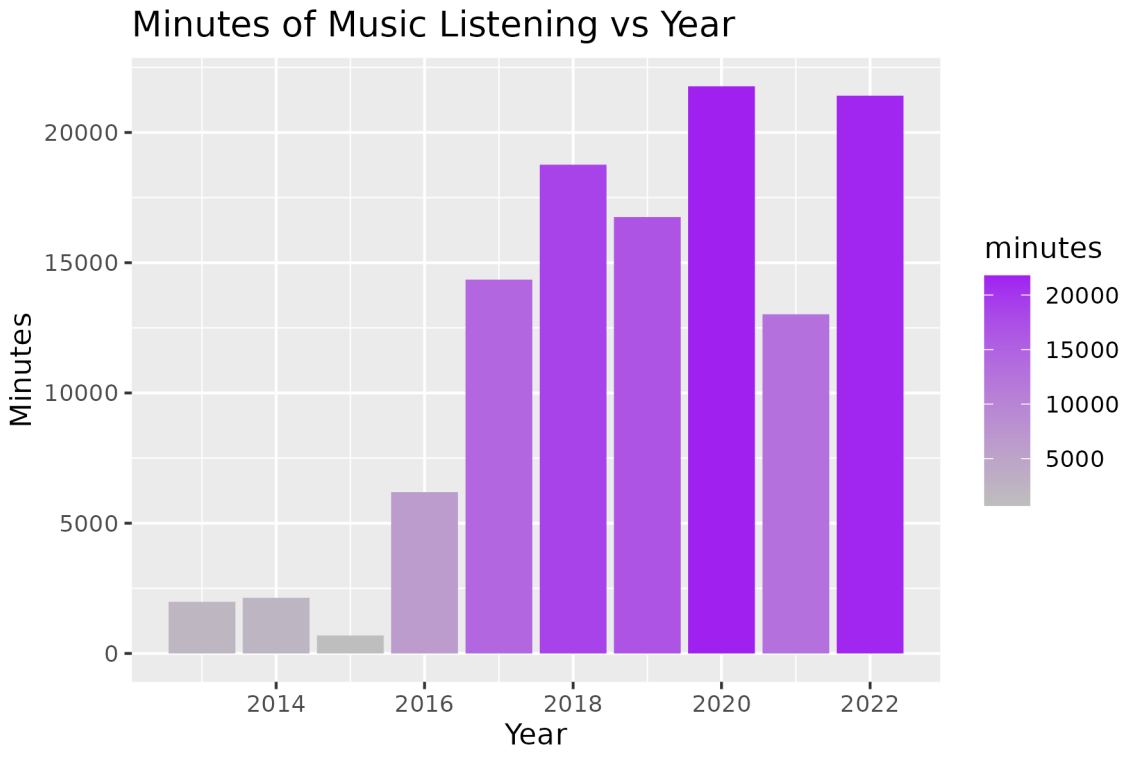


Figure 1: Minutes per year

(Figures 1 and 2) I started using Spotify as my main, and essentially exclusive, music streaming service

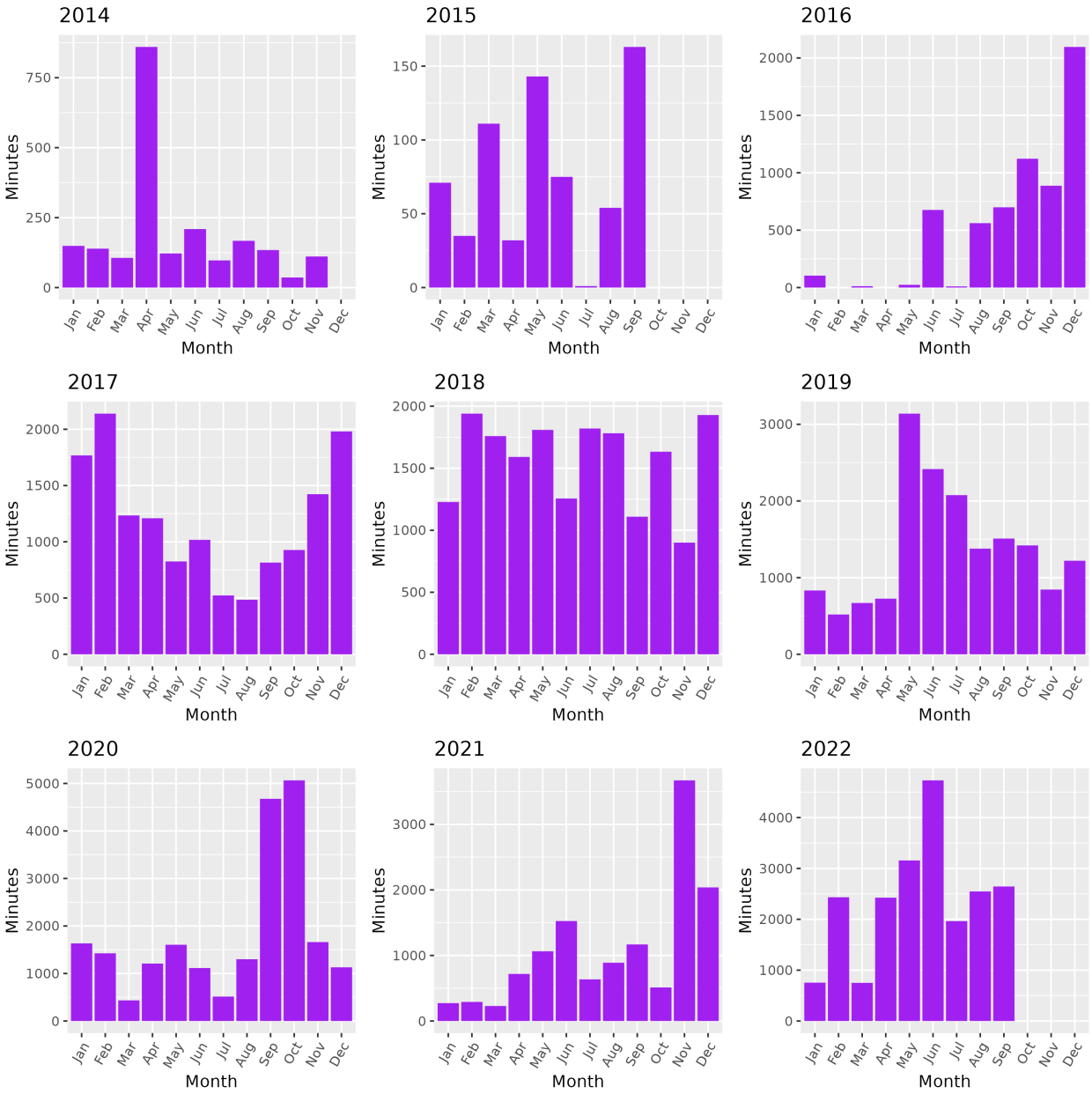


Figure 2: Minutes Per Month by Year

when I started undergrad in the Fall of 2016. Even though 2022 is missing 3 months of listening time, it is already on par with 2020 which is the year in which I have listened to the most music. There does not appear to be any clear trends with when I listen to music based on the month, and most of these peaks and valleys correlate more with life change rather than the month itself. For example, 2017 and 2019 almost have opposite distributions, in 2017 I listened to the most music in the winter months and less in the summer, but this trend is reversed for 2019. In 2020, I have a spike in music listening for whenever I started a full time job in September.

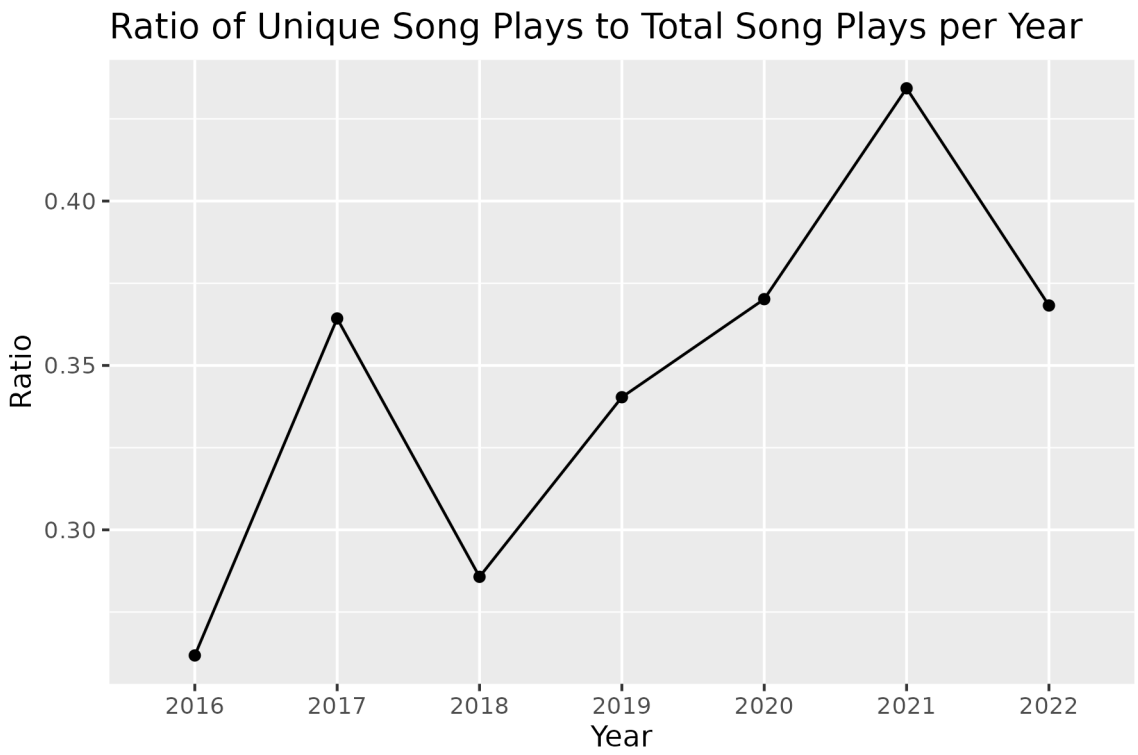


Figure 3: Ratio of unique song streams to total song streams per year

(Figure 3: Ratio of unique song streams to total song streams per year) The ratio of unique song streams to total song streams per year figure is a proxy I measure to assess how much new music I am listening to. A year like 2018 for instance, with a low ratio, implies that I listened to the same set of songs, but a year like 2021 shows that I listened to a lot of new music, not just songs that I already am familiar with. Part of what I love about Spotify is the music discovery algorithm and their "Discover Weekly" playlists, which helps me to discover new music that I enjoy easily. Discovering new songs and artists is one of my favorite things about listening to music and is a major factor in why I enjoy music listening as a hobby. From 2018 on wards, I increased the ratio of new music I listened to.

(Figure 4: Minutes per hour of day) This figure shows how many of minutes of music I have listened to at each hour of the day. The shape of this is about what I would expect, I listen to the most music in the afternoon around 1pm and 2pm.

(Figure 5 Minutes per day of week) This figure is interesting because it shows that I listen to the least music overall consecutively on Monday and Sunday, and then follow that up with the day I listen to the most music, Tuesday. I do not think there are any meaningful connections to be drawn from this, but interesting nonetheless.

(Figure 6 Minutes of Music Listening Per Month) I have listened to the most music in June and September and the least music in March. Making this chart is difficult because there are multiple approaches, if all years were weighted evenly, I am sure the distribution would look different. Ancedotally, my experience with

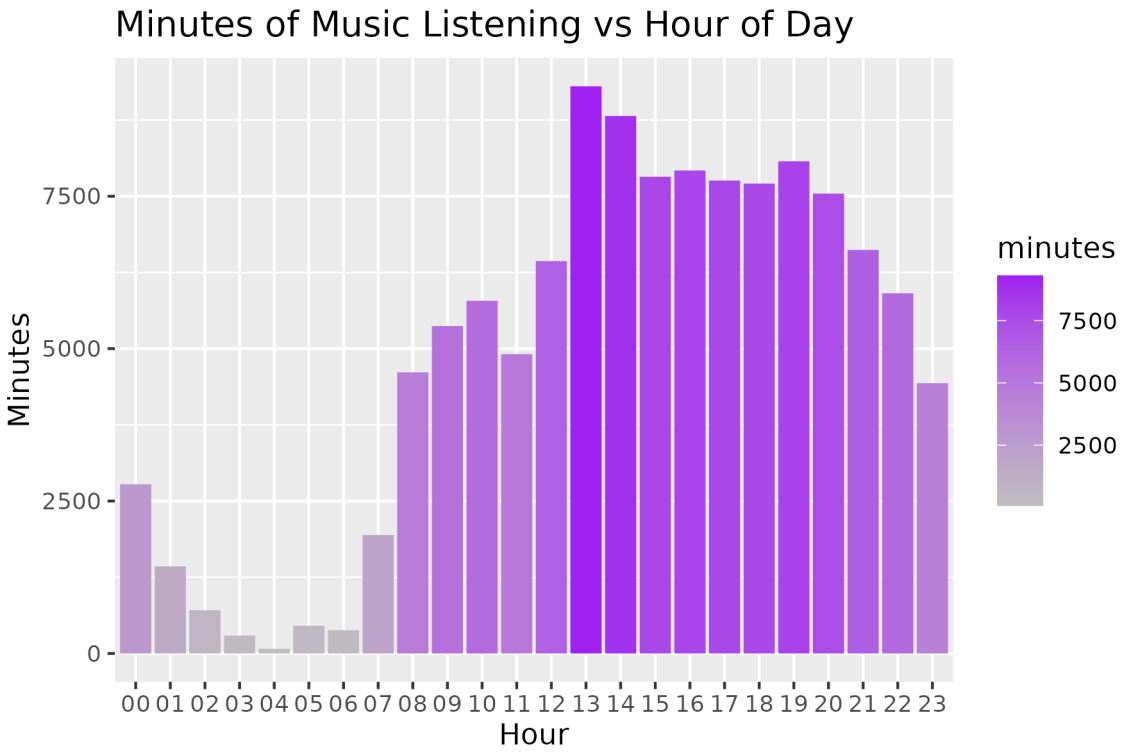


Figure 4: Minutes per hour of day

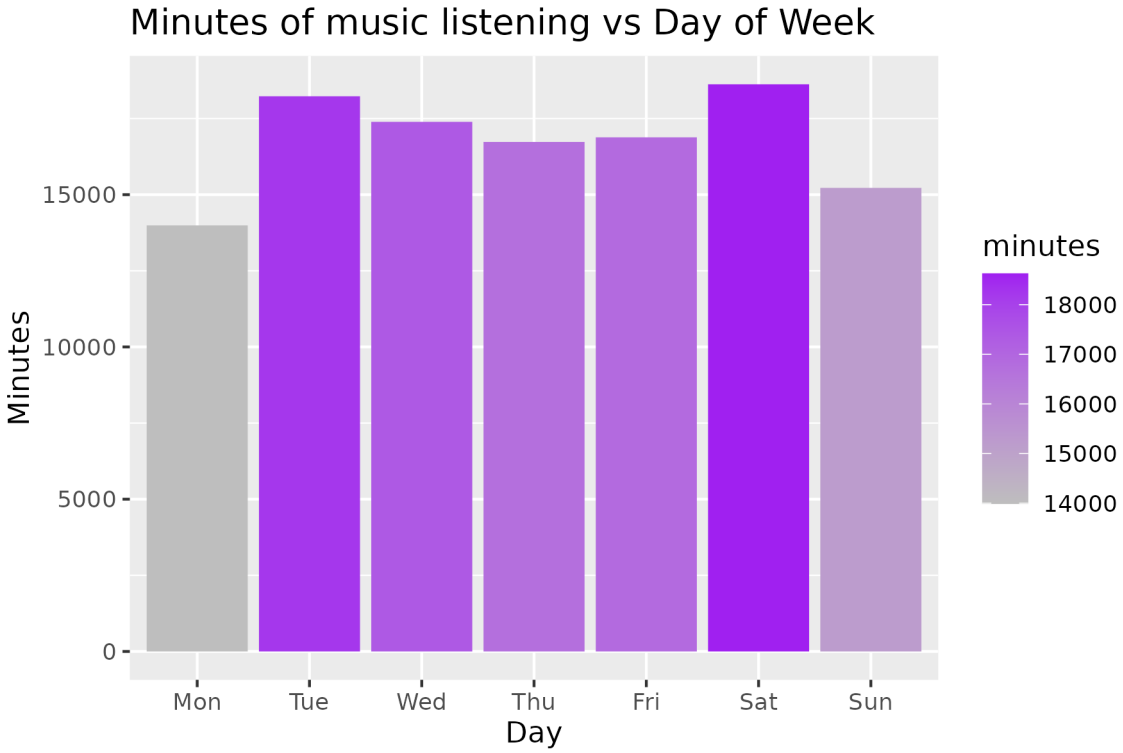


Figure 5: Minutes per day of week

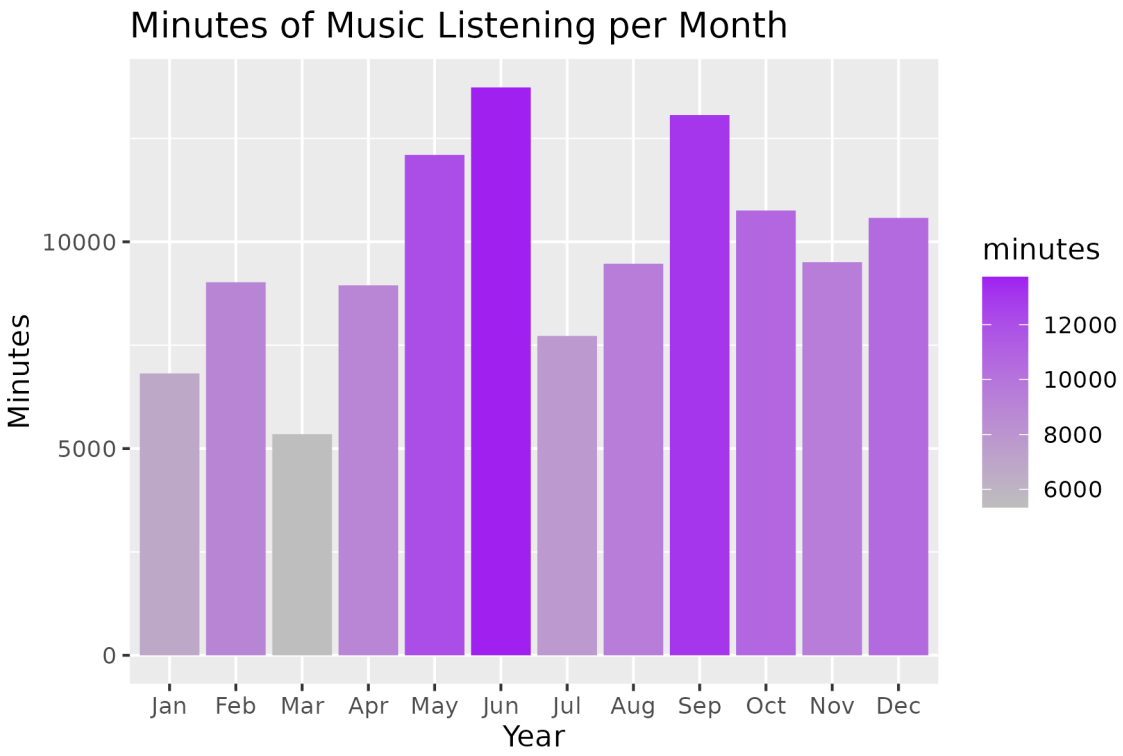


Figure 6: Minutes per month, all Years

music listening tends to agree with this observation, I feel I listen to a lot of music in September's, usually corresponding to the start of an academic year, and June since it is the end of an academic year. This figure is interesting but again, I do not think this is a super insightful into my listening habits.

(Figure 7 Number of Songs Played vs Release Year) This distribution is about what I expected, I spend plenty of time listening to indie and pop rock artists, which lends itself to modern releases, with the peak at 2016. I listen to more music from the 70s/80s such as ABBA, Hall & Oates, Billy Joel, etc. so I expected there to be a peak here, albeit much smaller than the peak focused about 2016. Additionally, I have thought that I do not prefer much music from the 90s and this chart appears to reflect that.

(Figure 8 Average Valence per Year) Valence is Spotify's variable for a measure of happiness, 1 being happy and 0 being sad. This is the chart that I have been most interested in seeing, I have hypothesized that this is the trend I would see, a decrease in valence overtime after 2018. This average valence was calculated by multiplying the valence of each song by the number of minutes played, summing the total valence for each song and dividing it by the number of minutes played. From a personal perspective, it would be interesting to somehow back trace these trends and see if there is a connection between this and the onset of depression, but nonetheless to see this trend when I have hypothesized it is very validating.

(Figure 9 Average Valence per Month) It is interesting that July has a much lower Valence than any of the other months, by a decent margin, but I cannot hypothesize an explanation for why this is the case.

(Figure 10 Average Valence per Day) The average valence per day barely differs, but Thursday is the lowest valence and Saturday is the highest. Possible explanation could be playing more happy and upbeat music on Saturday since there is no class/work, and Thursday can sometimes feel like the most difficult day of the week to get through, playing music to which relfects this.

(Figure 11 Average Valence per Hour) I have listened to an incredibly small amount of music at 5 am, I hypothesize that there is a peak here because if I am up at 5am, it is either from the night before in a social gathering, or it would be because I had 6am track practices and would listen to music to help get myself

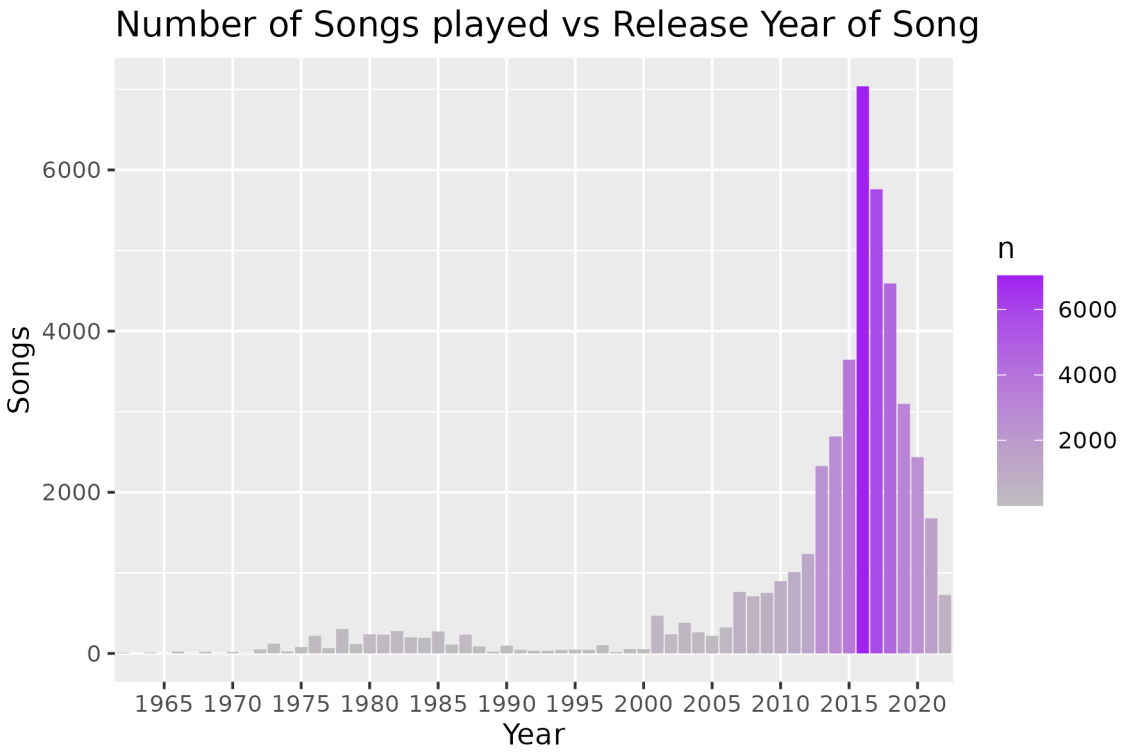


Figure 7: Songs Streamed per Year vs Unique Songs Streams per Year

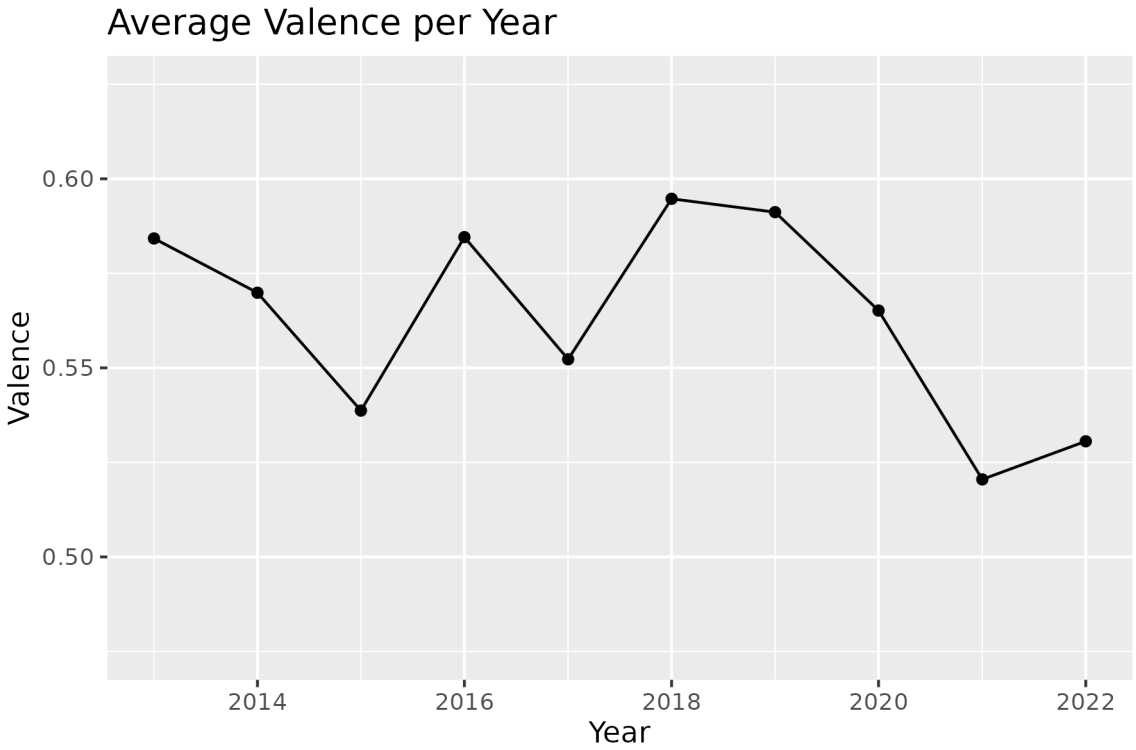


Figure 8: Valence by year

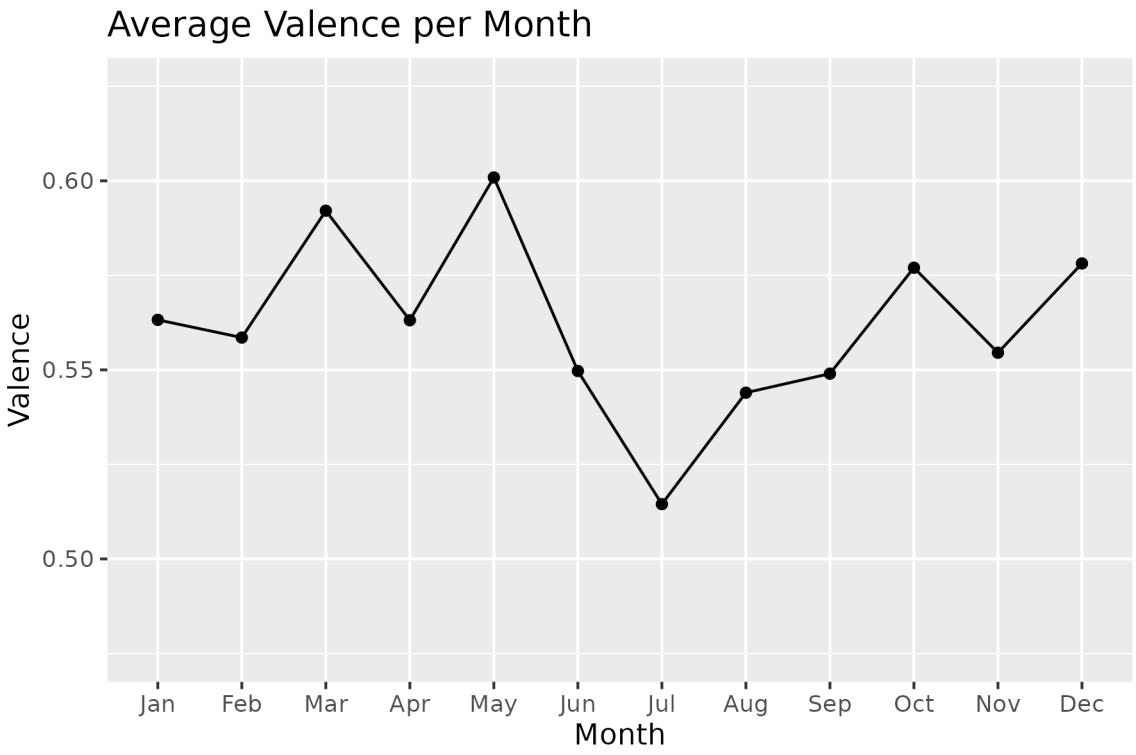


Figure 9: Valence by Month

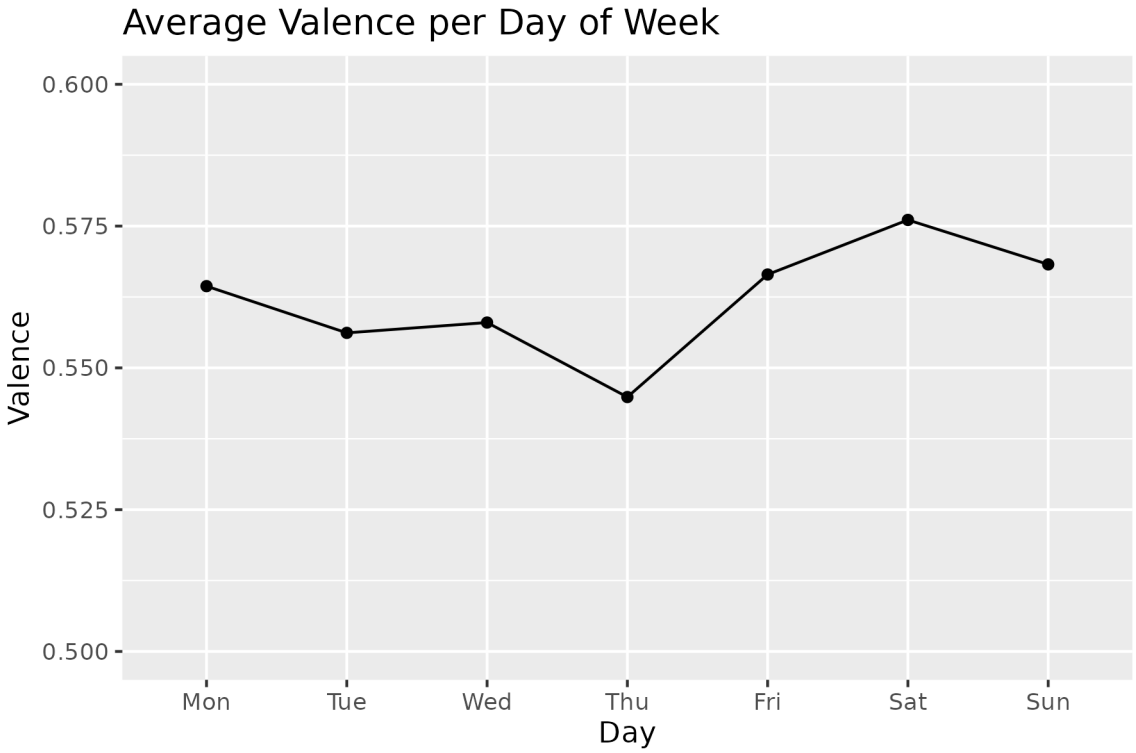


Figure 10: Valence by Day

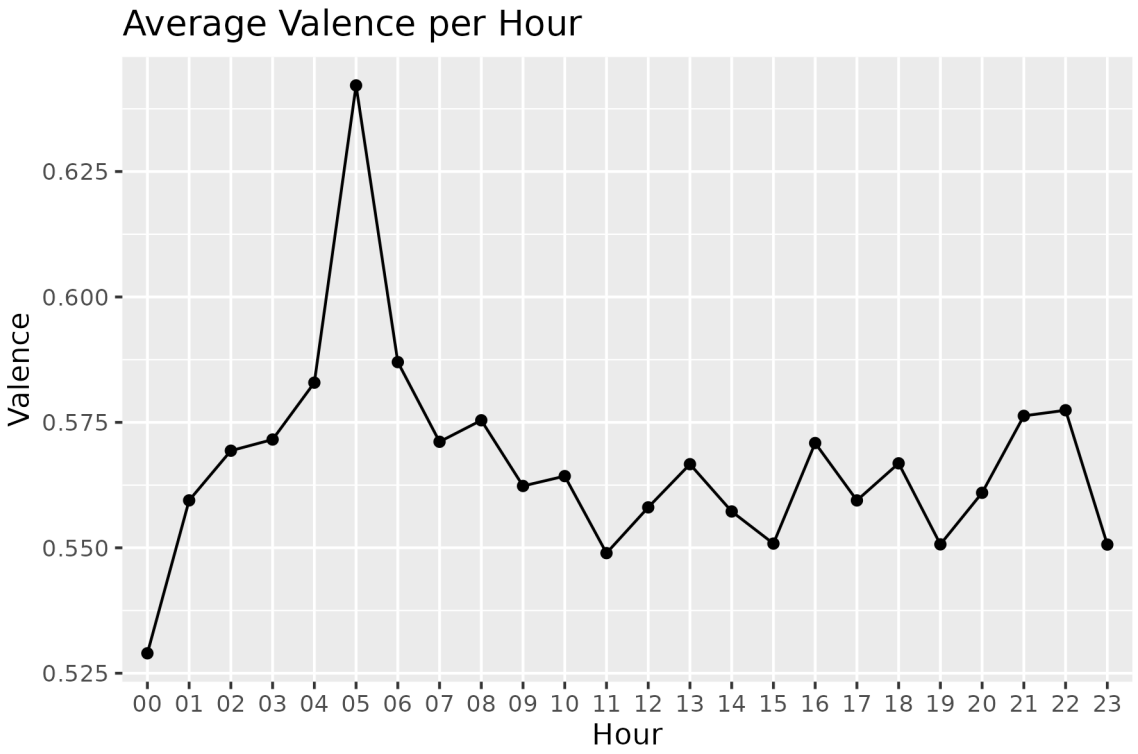


Figure 11: Valence by Hour

energized prior to it.

(Figure 12 Average Energy per Hour) I hypothesize the peak at 5am is for similar reasons mentioned above. I could listen to higher energy music in the morning (8am or 9am) to give myself energy for the day, and throughout the day the peak tends to be much lower.

(Figure 13 Average Energy per Day of Week) The y-axis scale here is very small to try and highlight the variation in energy, but there is no interesting trends here.

(Figure 14 Average Danceability per Hour) I hypothesize this peak at 5am is for the same reason mentioned in the hourly valence plot. The danceability of music does not vary much throughout the day, but is higher at 6am and 7am. Similar to energy, I hypothesize this is to help wake me up in the morning.

(Figure 15 Average Danceability per Day of Week) Examine the y-axis to see that there is barely any variation in Danceability vs Day of week.

The remaining figures are *.html files, accessible by your local browser, within the "/work/html/" directory.

For each of the following tables, songs, albums, and artists are ranked based on minutes of listening, rather than number of plays. Philosophically, I value the amount of time spent listening to a track more than number of listens, as I also find this makes comparisons between tracks more honest, the number of plays for a longer song should naturally be less than that of an "equally played" shorter song. (An argument for the converse could be made, if I listen to a song I usually listen to the full thing which would inherently benefit longer tracks.)

(top_songs.html) This is a table of my 15 most played songs, I was surprised to see "Secrets" by "The Weeknd" in my top 10 as I have not listened to that song much recently, but it used to be a daily repeat. I am incredibly surprised to see "Africa" by "Toto" and "True" by "Spandau Ballet" on this list, whenever I listen to older music I would have thought I trended towards different artists. I would not have expected either of these to be in my top 30 most listened to songs.

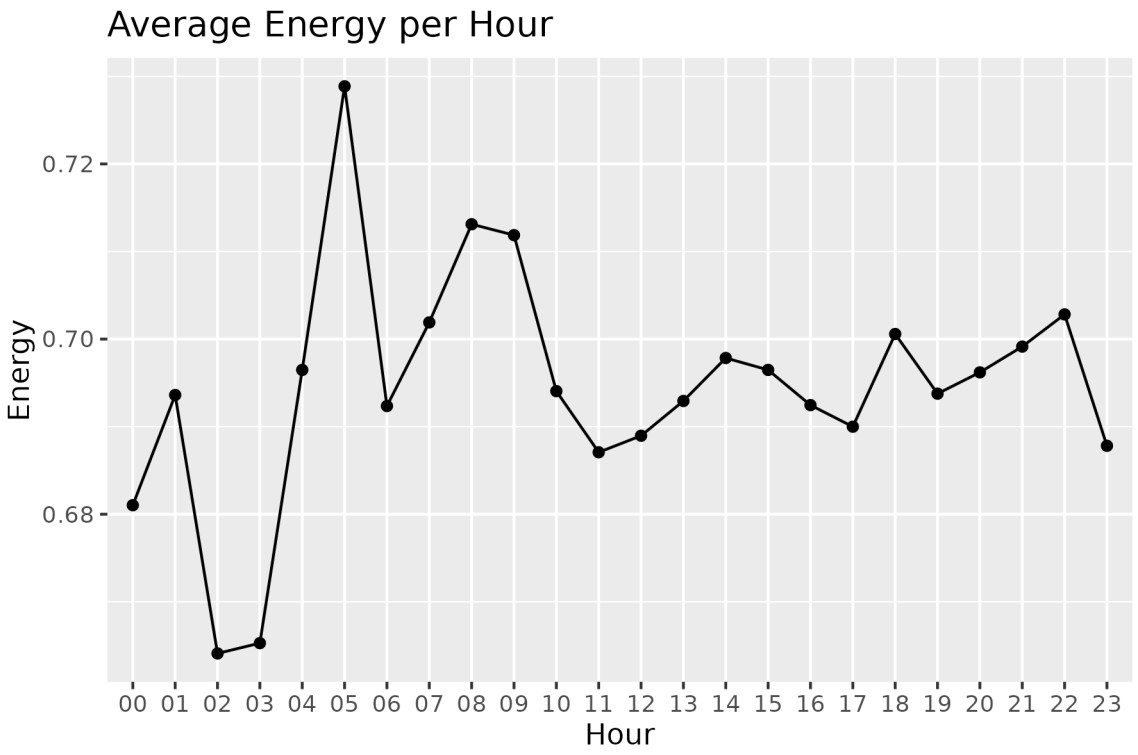


Figure 12: Energy by Hour

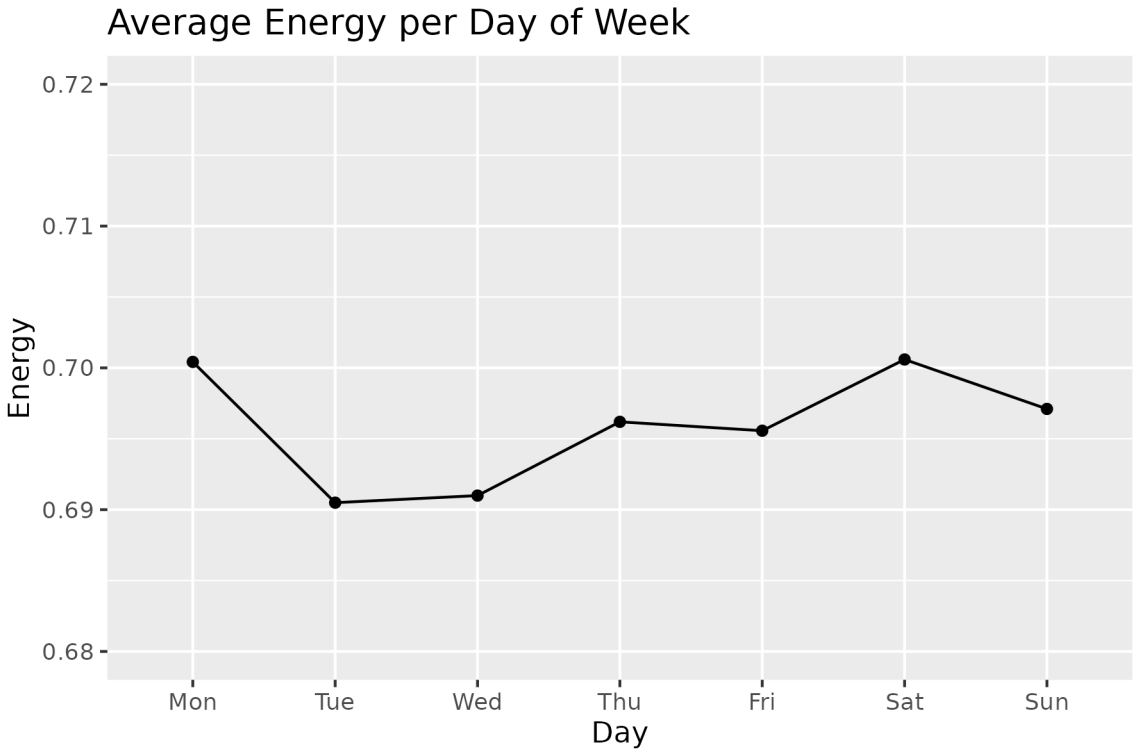


Figure 13: Energy by Day

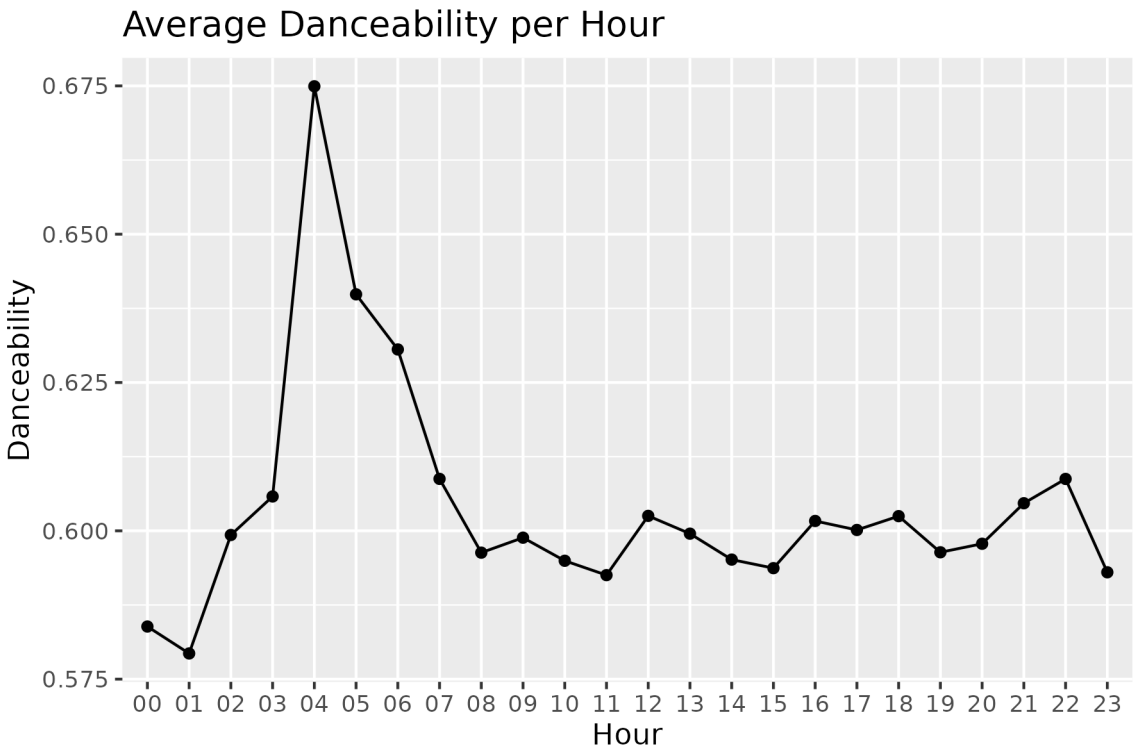


Figure 14: Danceability by Hour

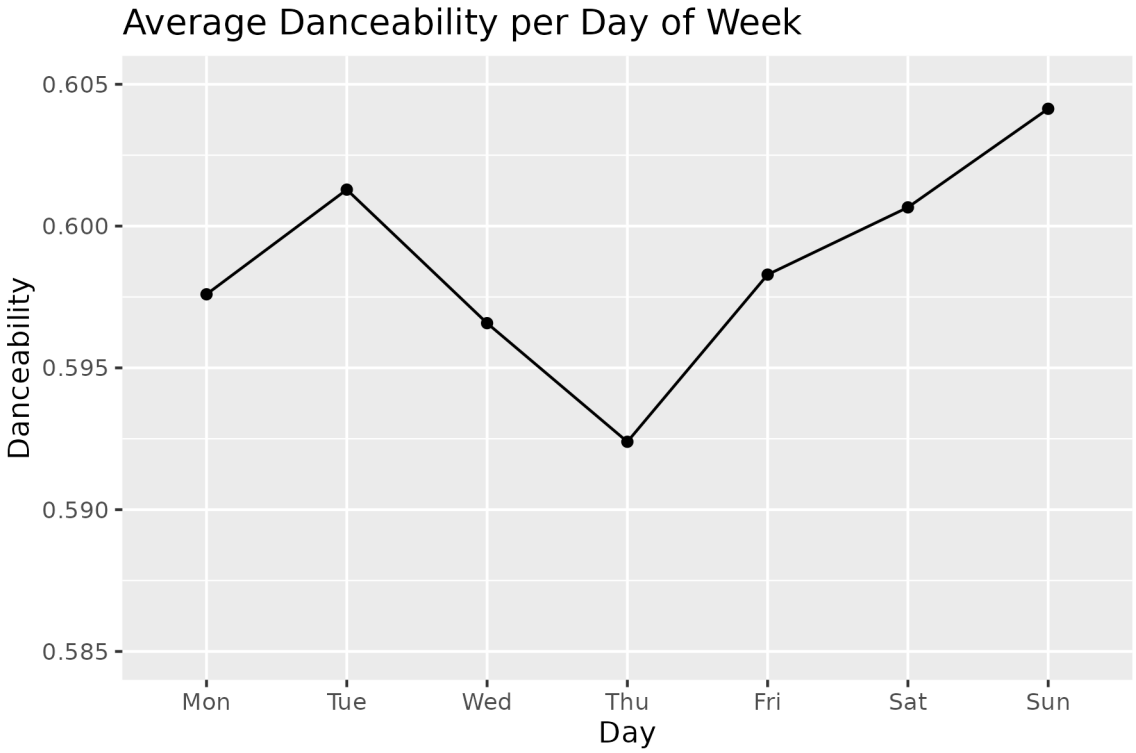


Figure 15: Danceability by Day

(top_albums.html) This is a table of my 15 most played albums. Philosophically, I would like to see which are my top albums as a measure of which Albums do I put on and listen to as a collective unit, rather than a summation of the time spent listening to singles, but Spotify does not track music listening this way, and I am not even sure how I would want this tracked in my ideal situation. For instance, my top Album, "ABBA Gold" by "ABBA", is a compilation album, and most of my time spent listening comes from "ABBA" songs being on a playlist which I would shuffle through, rather than exclusively listening to the "ABBA Gold" album. Despite my qualms with how this list is constructed, it feels like an honest list which reflects my Philisophy of how I would rank my albums. I am surprised to see "Gameshow" by "Two Door Cinema Club" as high up on this list as it is, and by a decent margin over 3rd.

(top_artists.html) This is a table of my 15 most listened to artists. There are no surprises on this list and I am happy to see some of my 80s influences popping up.

(top_podcasts.html) This is a table of my 15 most listened to podcasts. I spend plenty of time listening to podcasts during commute or while working. I knew "The Adam Friedland Show" would be high on this list, but given that I have only started listening to it in 2022, and 2022 only include listening data from January-September, I am surprised to see it has over 15 days worth of playtime.

(top_summary_year.html) This table shows my most listened to songs, albums, and artists from each year. I discovered and fell in love with "Beach House" in 2022, and this table reflects that given they have my top song, album, and artist. I am surprised to see some of the songs and artists which were the top for the year, like "Daydream/Wetdream/Nightmare" by "Saint Motel" in 2019 as my most played song, or "Hoodie Allen" as my most played artist in 2017.