Week 12 Data Story Submission

Esther Kho Yining

2024-04-16

Week 8

What is the topic that you have finalized? (Answer in 1 or 2 sentences)

The topic that I have finalized is about the correlation between the popularity of Spotify songs and the elements of the music.

What are the data sources that you have curated so far? (Answer 1 or 2 sentences).

The data source I have obtained is for free, and originates from the following link: https://www.kaggle.com/datasets/sujaykapadnis/spotify-songs

The data set contains information on the track popularity, danceability of the music, energy of the music and liveliness of the music.

```
# Data set
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
Spotify <- read.csv("spotify_songs.csv")

# First 5 rows of the data set
head(Spotify, 5)
```

```
##                 track_id                                              track_name
## 1 6f807x0ima9a1j3VPbc7VN I Don't Care (with Justin Bieber) - Loud Luxury Remix
## 2 0r7CVbZTWZgbTCYdfa2P31                        Memories - Dillon Francis Remix
## 3 1z1Hg7Vb0AhHDiEmnDE79l                        All the Time - Don Diablo Remix
## 4 75FpbthrwQmzHlBJLuGdC7                      Call You Mine - Keanu Silva Remix
## 5 1e8PAfcKUYoKkxPhrHqw4x                Someone You Loved - Future Humans Remix
```

```
##       track_artist track_popularity         track_album_id
## 1       Ed Sheeran               66 2oCs0DGTsRO98Gh5ZSl2Cx
## 2         Maroon 5               67 63rPSO264uRjW1X5E6cWv6
## 3     Zara Larsson               70 1HoSmj2eLcsrR0vE9gThr4
## 4 The Chainsmokers               60 1nqYsOef1yKKuGOVchbsk6
## 5    Lewis Capaldi               69 7m7vv9wlQ4i0LFuJiE2zsQ
##                                       track_album_name
## 1 I Don't Care (with Justin Bieber) [Loud Luxury Remix]
## 2                        Memories (Dillon Francis Remix)
## 3                        All the Time (Don Diablo Remix)
## 4                            Call You Mine - The Remixes
## 5                Someone You Loved (Future Humans Remix)
##   track_album_release_date playlist_name            playlist_id playlist_genre
## 1               2019-06-14     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 2               2019-12-13     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 3               2019-07-05     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 4               2019-07-19     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 5               2019-03-05     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
##   playlist_subgenre danceability energy key loudness mode speechiness
## 1         dance pop        0.748  0.916   6   -2.634    1      0.0583
## 2         dance pop        0.726  0.815  11   -4.969    1      0.0373
## 3         dance pop        0.675  0.931   1   -3.432    0      0.0742
## 4         dance pop        0.718  0.930   7   -3.778    1      0.1020
## 5         dance pop        0.650  0.833   1   -4.672    1      0.0359
##   acousticness instrumentalness liveness valence   tempo duration_ms
## 1       0.1020         0.00e+00   0.0653   0.518 122.036      194754
## 2       0.0724         4.21e-03   0.3570   0.693  99.972      162600
## 3       0.0794         2.33e-05   0.1100   0.613 124.008      176616
## 4       0.0287         9.43e-06   0.2040   0.277 121.956      169093
## 5       0.0803         0.00e+00   0.0833   0.725 123.976      189052
```

Week 9

What is the question that you are going to answer?

Does the popularity of Spotify music and correlation to elements of the music suggest any insightful reflections on how people remember things?

Why is this an important question?

Despite not being able to recall the majority of events in their lives, Dr Timothy Byron and Dr Jadey O'Regan (2022) highlights that many individuals have the ability to remember a song, even if they have not heard it in a long time. This ability to remember a song may be attributed to the catchiness and the tempo of the song which allows the song to remain memorable and popular for a long time once learned. As the popularity of the song is dependent on how memoriable it is to individuals (Byron & O'Regan, 2022), understanding the elements of popular music reflects the essential components that help people to remember things.

Which rows and columns of the dataset will be used to answer this question?

I will use the columns: Track popularity, Playlist Genre, Danceability, Valence, Speechiness, Tempo,

As of submitting this document, I intend to work with the first 50 row from each playlist genre as the data is ordered based on the 6 different playlist genre.

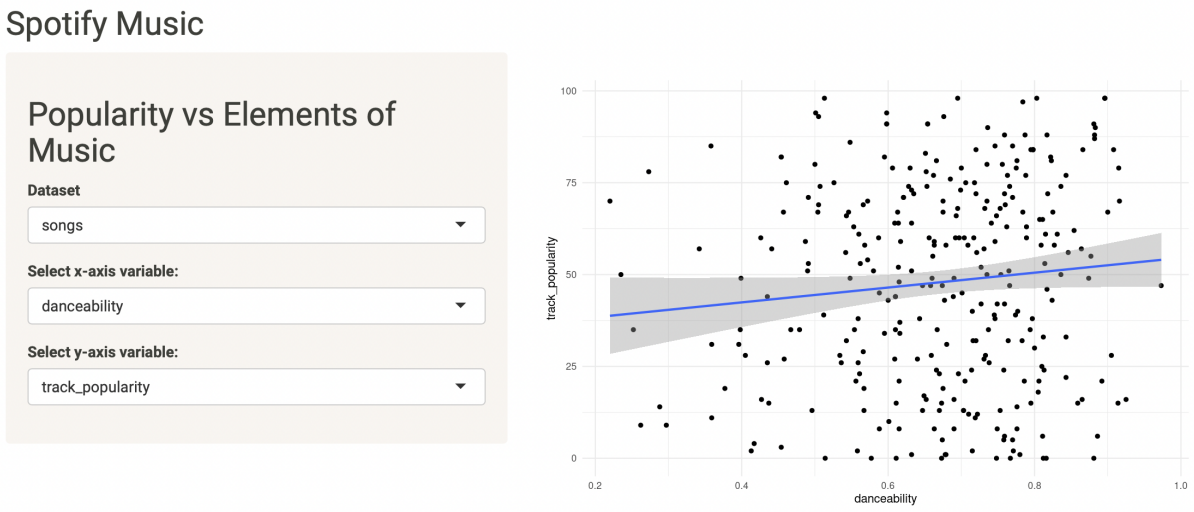
Week 10

List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?)

By looking at the elements of songs such as danceability and valence, I want to find out if there is a significant element that affects the popularity of the music. Through this comparison, I am to find out if there is any patterns in how people remember things.

How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

I will repurpose the Shiny app used in Week 9's Challenge to allow other users to plot the graphs based on the variables that they are interested in my data sets. This enables them to see the different patterns shown in the ggplot that I will create in the Shiny app



What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

Topics	Weeks
ggplot	2,7
basics of Shiny app	8
Quarto website set-up	8
conditional outputting different variables from a curated list on Shiny	

Week 11

Update: Changes made in the dataset

In order to clearly depict the correlations between the music's popularity and its elements, I have selected another dataset which originates from the following link: https://www.kaggle.com/datasets/yukawithdata/spotify-top-tracks-2023

For this new dataset, I will use all the rows and the following columns: Popularity, Beats.Per.Minute, Danceability, Valence, Speechiness

```
# Data set
library(tidyverse)
songs <- read.csv("top50.csv")

# First 5 rows of the data set
head(Spotify, 5)
```

```
##                 track_id                                                track_name
## 1 6f807x0ima9a1j3VPbc7VN I Don't Care (with Justin Bieber) - Loud Luxury Remix
## 2 0r7CVbZTWZgbTCYdfa2P31                      Memories - Dillon Francis Remix
## 3 1z1Hg7Vb0AhHDiEmnDE79l                      All the Time - Don Diablo Remix
## 4 75FpbthrwQmzHlBJLuGdC7                    Call You Mine - Keanu Silva Remix
## 5 1e8PAfcKUYoKkxPhrHqw4x                 Someone You Loved - Future Humans Remix
##       track_artist track_popularity         track_album_id
## 1       Ed Sheeran               66 2oCs0DGTsRO98Gh5ZSl2Cx
## 2         Maroon 5               67 63rPSO264uRjW1X5E6cWv6
## 3     Zara Larsson               70 1HoSmj2eLcsrR0vE9gThr4
## 4 The Chainsmokers               60 1nqYsOef1yKKuGOVchbsk6
## 5    Lewis Capaldi               69 7m7vv9wlQ4i0LFuJiE2zsQ
##                                    track_album_name
## 1 I Don't Care (with Justin Bieber) [Loud Luxury Remix]
## 2                       Memories (Dillon Francis Remix)
## 3                       All the Time (Don Diablo Remix)
## 4                           Call You Mine - The Remixes
## 5                Someone You Loved (Future Humans Remix)
##   track_album_release_date playlist_name            playlist_id playlist_genre
## 1               2019-06-14     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 2               2019-12-13     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 3               2019-07-05     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 4               2019-07-19     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 5               2019-03-05     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
##   playlist_subgenre danceability energy key loudness mode speechiness
## 1         dance pop        0.748  0.916   6   -2.634    1      0.0583
## 2         dance pop        0.726  0.815  11   -4.969    1      0.0373
## 3         dance pop        0.675  0.931   1   -3.432    0      0.0742
## 4         dance pop        0.718  0.930   7   -3.778    1      0.1020
## 5         dance pop        0.650  0.833   1   -4.672    1      0.0359
##   acousticness instrumentalness liveness valence   tempo duration_ms
## 1       0.1020         0.00e+00   0.0653   0.518 122.036      194754
## 2       0.0724         4.21e-03   0.3570   0.693  99.972      162600
## 3       0.0794         2.33e-05   0.1100   0.613 124.008      176616
## 4       0.0287         9.43e-06   0.2040   0.277 121.956      169093
## 5       0.0803         0.00e+00   0.0833   0.725 123.976      189052
```

Week 12

Reflect on your experience doing the project using R

My experience doing the project using R was a challenging yet fulfilling one. Throughout the journey, there were moments when I encountered frustrations, particularly when my code was not functioning as expected, and I struggled to implement the desired interactive elements through the Shiny app. However, these obstacles were valuable learning experiences for me. As I seek help from my friends and the TAs, I gain a better understanding of how the code functions for the Shiny app and become more aware of the potential pitfalls when adapting the code from the given Shiny examples. Despite the challenges, I was eventually able to overcome them and felt a sense of accomplishment when I successfully produced an interactive website. This journey instilled a belief in my ability to develop my own interactive websites in the future, fueling a desire to further explore and expand my skills in R.

What have you learned in NM2207, and how can that be helpful in your future career? Try to identify use-cases within your specialization and future industry that you might benefit by using R.

In NM2207, I have acquired valuable skills in producing and manipulating datasets using functions like filter, arrange and mutate. These functions allow me to extract the specific data I need for analysis. In addition, I have also learnt how to create visualisations through ggplot and Shiny. These tools will allow me to effectively examine and present my datasets in the future.

The knowledge and skills I have gained in NM2207 will undoubtedly benefit me as a Sociology student. In Sociology, statistics are often analysed through the use of R studio in order to understand patterns in social phenomena. According to Clark (2022), Sociology often utilised the igraph function to create a network graph. While this function was not taught in NM2207, the skills learnt in NM2207 still provides me with the basic knowledge required to create such graphs. Hence, I can apply these basic knowledge to produce the necessary graphs and visualisations to analyse the statistics for my sociological research in the future with my understanding of R from NM2207.

With a keen interest in pursuing a career in the interactive media industry, the knowledge and skills acquired from NM2207 have equipped me with a valuable skill set that aligns with roles such as web developer and app developer. With the acquired knowledge of Shiny from NM2207, I am able to develop interactive website applications and dashboards, which are highly sought-after features in the industry. Such interactive elements include user-input functionalities, data-driven visualisations and personalised user experience. Thus, the knowledge gained through NM2207 have provided me with a good foundation for pursuing a career in the interactive media industry.

In conclusion, I have learnt valuable skills and knowledge in NM2207 which provides me with the basics and good foundation for my future studies and careers

References

Byron, T., & O'Regan, J. (2022). *Hooks in popular music*. NSW, Australia. Springer Nature.

Clark, M. (2022). R & SOCIAL SCIENCE: GETTING STARTED WITH APPLIED USE OF R IN THE SOCIAL SCIENCES. Retrieved from https://m-clark.github.io/docs/RSocialScience.pdf