Data_Vis_Prj

2023-12-03

Data Visualization Final Project: Music Taste Analysis 2023

This data set includes a multitude of variables for my most played songs on Spotify for 2023.

```
library(data.table)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v ggplot2 3.4.3     v purrr   1.0.2
## v tibble  3.2.1     v stringr 1.5.0
## v tidyr   1.3.0     v forcats 1.0.0
## v readr   2.1.3
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

```r
# Listening data for the past 7 years

readRenviron('/Users/alarahector/Desktop/Fall 2023/Music Analysis /Listeninghistory.R')

# Top 100 songs imported csv

top2015 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2015.csv', check.names = TRUE)
top2016 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2016.csv', check.names = TRUE)
top2017 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2017.csv', check.names = TRUE)
top2018 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2018.csv', check.names = TRUE)
top2019 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2019.csv', check.names = TRUE)
top2020 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2020.csv', check.names = TRUE)
top2021 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2021.csv', check.names = TRUE)
top2022 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2022.csv', check.names = TRUE)
top2023 <- fread('/Users/alarahector/Desktop/Fall 2023/Music Analysis /2023.csv', check.names = TRUE)

# ------------------------------------------------------------------------------
# Selecting Relevant Columns

top2015 <-subset(top2015, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))
top2016 <-subset(top2016, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))
top2017 <-subset(top2017, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))
top2018 <-subset(top2018, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))
top2019 <-subset(top2019, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))
top2020 <-subset(top2020, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))
top2021 <-subset(top2021, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))
top2022 <-subset(top2022, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))
top2023 <-subset(top2023, select = c(Track.Name, Album.Name, Artist.Name.s., Duration..ms.,
                                     Popularity, Genres, Danceability, Energy, Key, Loudness, Tempo))

# ------------------------------------------------------------------------------
# Summary of Years

summary.unique <- data.frame(year=c(2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023),
                  artists=c(915, 895, 1029, 1148, 1280, 1528, 1278, 760, 1109),
                  albums=c(1411, 1432, 1579, 1783, 2112, 2240, 1904, 1112, 1739),
                  tracks=c(3096, 2886, 4439, 3867, 4418, 3918, 3802, 2941, 4023))

# ------------------------------------------------------------------------------
top2015 <- top2015 %>%
  mutate(Year = 2015)

top2016 <- top2016 %>%
  mutate(Year = 2016)
```

```r
top2017 <- top2017 %>%
  mutate(Year = 2017)

top2018 <- top2018 %>%
  mutate(Year = 2018)

top2019 <- top2019 %>%
  mutate(Year = 2019)

top2020 <- top2020 %>%
  mutate(Year = 2020)

top2021 <- top2021 %>%
  mutate(Year = 2021)

top2022 <- top2022 %>%
  mutate(Year = 2022)

top2023 <- top2023 %>%
  mutate(Year = 2023)

topfull = rbind(top2015, top2016, top2017, top2018, top2019, top2020, top2021, top2022)
```

Diversity of Music Taste over the Years (2015 - 2023)

```r
library(ggplot2)
library(grid)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

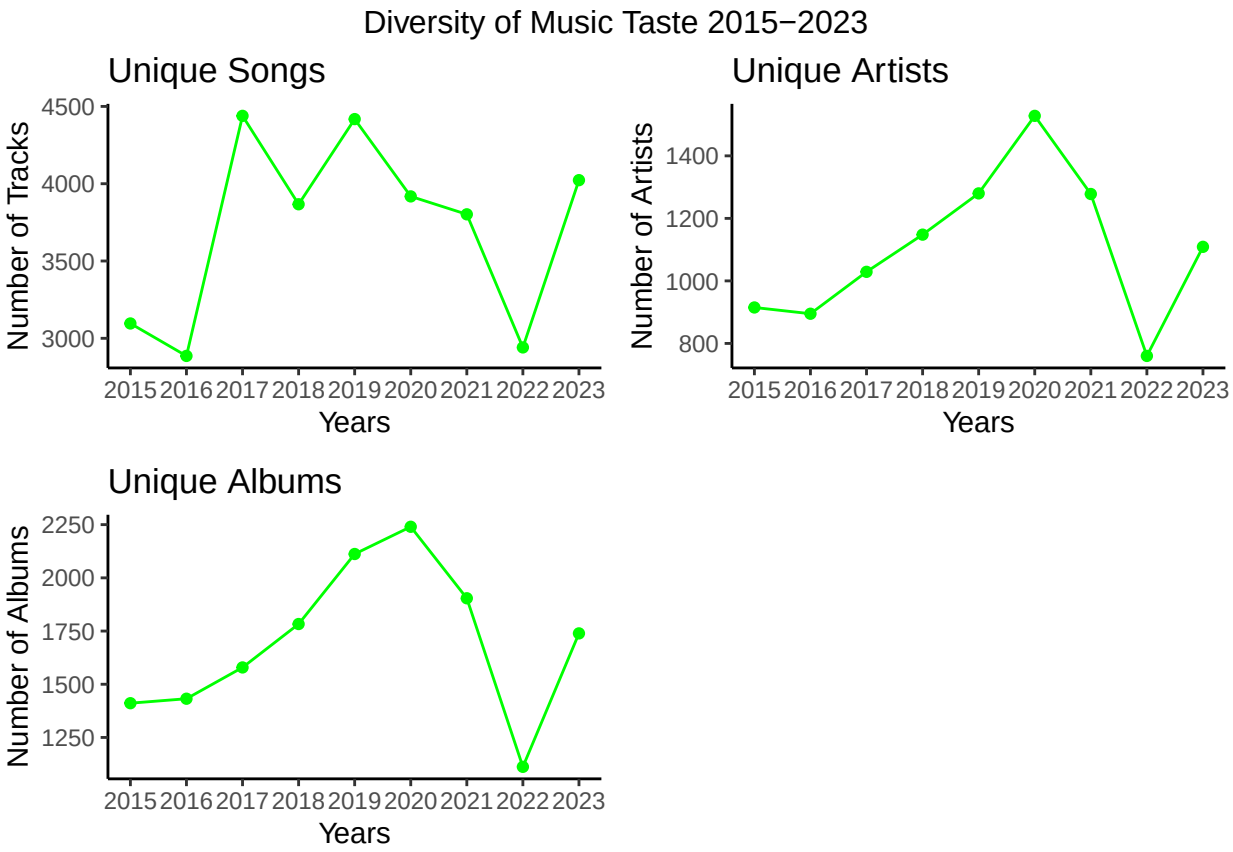
```r
x <- ggplot(summary.unique, aes(x = year,
                                y = tracks)) +
  geom_point(color = "green") +
  geom_line(aes(x = year, y = tracks), color = "green") +
  theme_classic() + labs(y = "Number of Tracks", title = "Unique Songs") +
  scale_x_continuous(name = "Years",
                     breaks = c(2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023))

y <- ggplot(summary.unique, aes(x = year,
                                y = artists)) +
  geom_point(color = "green") +
  geom_line(aes(x = year, y = artists), color = "green") +
  theme_classic() + labs(y = "Number of Artists", title = "Unique Artists") +
  scale_x_continuous(name = "Years",
```

```
                      breaks = c(2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023))

z <- ggplot(summary.unique, aes(x = year,
                            y = albums)) +
  geom_point(color = "green") +
  geom_line(aes(x = year, y = albums), color = "green") +
  theme_classic() + labs(y = "Number of Albums", title = "Unique Albums") +
  scale_x_continuous(name = "Years",
                      breaks = c(2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023))

grid.arrange(x, y, z, ncol = 2, top= "Diversity of Music Taste 2015-2023")
```



Characterizing Music Taste By Year

Descriptive Variables: Popularity, Danceability, Energy, Loudness, Temp

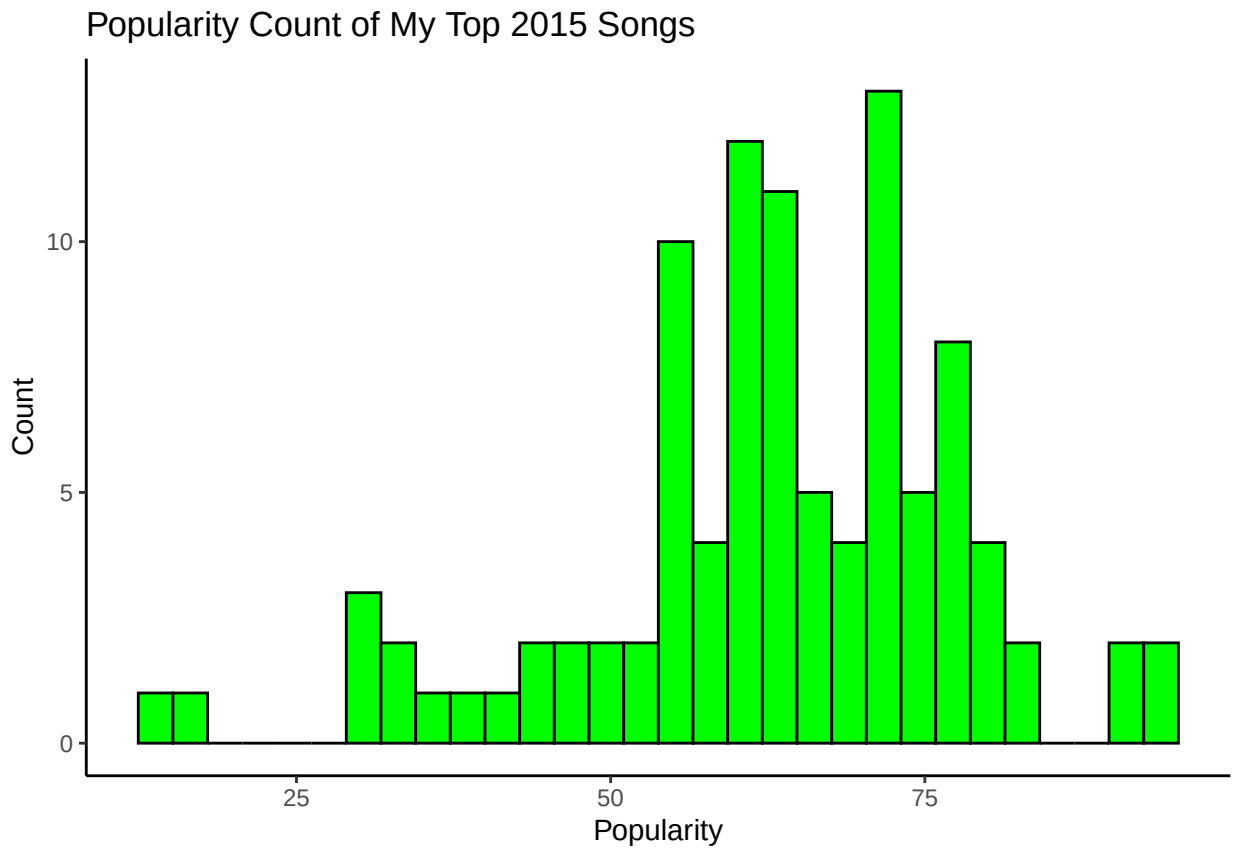
```
library(ggplot2)
library(grid)
library(gridExtra)
#library(corrplot)

# Correlation Plot of the Different Variables
#cormat = cor(topfull[,-c(1, 2, 3, 6)])
#corrplot(cormat)
```

```
# Scatter Plot by Year Danceability and Popularity


# Histogram
ggplot(top2015) +
    geom_histogram(aes(x = Popularity,
                       y = after_stat(count)),
                       fill = "green",
                       color = "black") + theme_classic () + labs(x = "Popularity", y = "Count",
                                                                  title="Popularity Count of My Top 2015
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# Box and Scatter Plots by Year
library(ggforce)
a <- ggplot(topfull, aes(x = factor(Year), y = Energy)) + geom_sina(color = "green", size = 1.5)+ theme_

b <- ggplot(topfull, aes(x = factor(Year), y = Popularity)) + geom_sina(color = "red", size = 1.5)+ ther

c <- ggplot(topfull, aes(x = factor(Year), y = Danceability)) + geom_sina(color = "blue", size = 1.5)+ t

d <- ggplot(topfull, aes(x = factor(Year), y = Tempo)) + geom_sina(color = "magenta", size = 1.5)+ theme

e <- ggplot(topfull, aes(x = factor(Year), y = Loudness)) + geom_sina(color = "orange", size = 1.5)+ the
```

Distribution

```
library(ggplot2)
```