Lec 05: In-class Exercise: Boxplots

SDS 192: Introduction to Data Science

Shiya Cao

Step 1: Load packages

```
#Load packages
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(spotifyr)
```

Step 2: Create an Spotify Developer Account

Copy client id and secret from your previous in-class exercise into the chunk below to replace the current client id and secret, and then run the code chunk.

```
id <- 'c283abfcc1c94026a6075358ab4ecfbe'
secret <- '3ee9ec4cd1cf433a8d3be549212f2a67'
Sys.setenv(SPOTIFY_CLIENT_ID = id)
Sys.setenv(SPOTIFY_CLIENT_SECRET = secret)
access_token <- get_spotify_access_token()
```

Step 3: Get Song Features for Your Favorite Artist

Replace the text Janelle Monae below with the name of your favorite artist and then run the code chunk.

```
artist <- get_artist_audio_features(artist = "Janelle Monae") |>
  select(-c(album_images, artists, available_markets))
```

Step 4: Create Grouped Boxplots Visualizing the Distribution of Values in a Song Feature of Your Choice, Grouped by Album Name

Refer to the `ggplot()` cheatsheet to find the geom function you need to create this plot.

Be sure to add labels to your plot.

```
# Fill your code here.
```

Step 5: Interpret Results

Turn to a neighbor and discuss the plots. * Check for outliers. * Compare medians. * Compare the ranges. * Compare the IQRs. * Compare the symmetry.

> Call us over if you are having trouble interpreting!