Music Insights

Asheela Magwili

```
library("tidyverse")
```

```
## ── Attaching packages ─────────────────────────────────────────────────
──── tidyverse 1.2.1 ──
```

```
## ✔ ggplot2 3.2.1     ✔ purrr   0.3.2
## ✔ tibble  2.1.3     ✔ dplyr   0.8.3
## ✔ tidyr   1.0.0     ✔ stringr 1.4.0
## ✔ readr   1.3.1     ✔ forcats 0.4.0
```

```
## ── Conflicts ────────────────────────────────────────────────────────── t
idyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library("dplyr")
library("tidyr")
library("readr")

survey <- read_csv("https://raw.githubusercontent.com/introdsci/MusicSurvey/master/music
-survey.csv")
```

```
## Parsed with column specification:
## cols(
##   Timestamp = col_character(),
##   `First, we are going to create a pseudonym for you to keep this survey anonymous (m
ore or less). Which pseudonym generator would you prefer?` = col_character(),
##   `What is your pseudonym?` = col_character(),
##   Sex = col_character(),
##   Major = col_character(),
##   `Academic Year` = col_character(),
##   `Year you were born (YYYY)` = col_double(),
##   `Which musical instruments/talents do you play? (Select all that apply)` = col_char
acter(),
##   Artist = col_character(),
##   Song = col_character(),
##   `Link to song (on Youtube or Vimeo)` = col_character()
## )
```

```
preferences <- read_csv("https://raw.githubusercontent.com/introdsci/MusicSurvey/master/
preferences-survey.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Timestamp = col_character(),
##   `What was your pseudonym?` = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

Cleaning Variable Names

```
colnames(survey)[colnames(survey) == "Timestamp"] <- "time_submitted"

colnames(survey)[colnames(survey) == "First, we are going to create a pseudonym for you
 to keep this survey anonymous (more or less). Which pseudonym generator would you prefe
r?"] <- "pseudonym_generator"

colnames(survey)[colnames(survey) == "What is your pseudonym?"] <- "pseudonym"

colnames(survey)[colnames(survey) == "Which musical instruments/talents do you play? (Se
lect all that apply)"] <- "instrument_list"

colnames(survey)[colnames(survey) == "Major"] <- "academic_major"

colnames(survey)[colnames(survey) == "Year you were born (YYYY)"] <- "year_born"

colnames(survey)[colnames(survey) == "Artist"] <- "favorite_song_artist"

colnames(survey)[colnames(survey) == "Link to song (on Youtube or Vimeo)"] <- "favorite_
song_link"

colnames(survey)[colnames(survey) == "Sex"] <- "sex"

colnames(survey)[colnames(survey) == "Academic Year"] <- "academic_level"

colnames(survey)[colnames(survey) == "Song"] <- "favorite_song"

colnames(survey)
```

```
##  [1] "time_submitted"       "pseudonym_generator"  "pseudonym"
##  [4] "sex"                  "academic_major"       "academic_level"
##  [7] "year_born"            "instrument_list"      "favorite_song_artist"
## [10] "favorite_song"        "favorite_song_link"
```

Creating Tables

```
Person <- tibble(time_submitted = survey$time_submitted, pseudonym = survey$pseudonym, s
ex = survey$sex, academic_major = survey$academic_major, academic_level = survey$academi
c_level, year_born = survey$year_born)

FavoriteSong <- tibble(pseudonym = survey$pseudonym, artist = survey$favorite_song_artis
t, song = survey$favorite_song, link = survey$favorite_song_link)
```

Cleaning Data

```
Person$time_submitted <- as.POSIXlt(parse_datetime(Person$time_submitted, format="%m/%
d/%y %H:%M"))
```

Cleaning academic major's levels by making the letter cases more consistent:

```
Person$academic_major <- as.factor(Person$academic_major)
levels(Person$academic_major)[levels(Person$academic_major) == "Computer information sys
tems"] <- "Computer Information Systems"

levels(Person$academic_major)
```

```
## [1] "Computer Engineering"         "Computer Information Systems"
## [3] "Computer Science"             "Math"
```

Create a 'Rating' Table

Create a Ratings table using the R function gather()

```
preferences$Timestamp <- NULL
Rating <- gather(preferences,artist_song,rating,"40 crew\tNot Enough" :"Wheezer\tBuddy H
olly")
Rating <- tibble(pseudonym = Rating$`What was your pseudonym?`, artist_song = Rating$art
ist_song, rating = Rating$rating)
```