

Music Insights

Tanya Bonilla

First, we need tidyverse to handle data importing and wrangling

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

```
nrow(survey) #44 Rows
```

```
## [1] 44
```

```
ncol(survey) #11 Columns
```

```
## [1] 11
```

```
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"] <- "pseudonym"
colnames(survey)[colnames(survey)=="What is your pseudonym?"] <- "pseudonym"
colnames(survey)[colnames(survey)=="Sex"] <- "sex"
colnames(survey)[colnames(survey)=="Major"] <- "academic_major"
colnames(survey)[colnames(survey)=="Academic Year"] <- "academic_level"
colnames(survey)[colnames(survey)=="Major"] <- "academic_major"
colnames(survey)[colnames(survey)=="Year you were born (YYYY)"] <- "year_born"
colnames(survey)[colnames(survey)=="Which musical instruments/talents do you play? (Select all that apply)"] <- "instrument_list"
colnames(survey)[colnames(survey)=="Artist"] <- "favorite_song_artist"
colnames(survey)[colnames(survey)=="Song"] <- "favorite_song"
colnames(survey)[colnames(survey)=="Link to song (on Youtube or Vimeo)"] <- "favorite_song_link"
colnames(survey)
```

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

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

```
person <- tibble(time_submitted = survey$time_submitted,
                 generator_name = survey$generator_name,
                 pseudonym = survey$pseudonym, sex = survey$sex,
                 major = survey$academic_major,
                 academic_level = survey$academic_level,
                 year_born = survey$year_born)
person$time_submitted <- as.POSIXlt(parse_datetime(person$time_submitted, format="%m/%d/%y %H:%M"))

favoritesong <- tibble(pseudonym = survey$pseudonym,
                      song = survey$favorite_song,
                      artist = survey$favorite_song_artist,
                      link = survey$favorite_song_link)
```

```

preferences <- gather(preferences, "artist_song", "rating", 3:45)
colnames(preferences)[colnames(preferences)=="What was your pseudonym?"] <- "pseudonym"
ratings <- tibble(pseudonym = preferences$pseudonym,
                  artist_song = preferences$artist_song,
                  rating = preferences$rating
)

person$academic_level <- as.factor(person$academic_level)
person$major <- as.factor(person$major)
levels(person$major)[levels(person$major) == "Computer information systems"]<-"Computer Information Systems"

```