Music Insights

Reece Resendez

First, we need tidyverse to handle the data importing and wrangling Since we already have it installed on this system we can go ahead and implement it.

```
suppressMessages(library(tidyverse))
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
## Warning: package 'stringr' was built under R version 3.5.3
```

```
suppressMessages(library(readr))

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

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

Now we want to go ahead and update all the column names for each of the dataframes to make referencing and appearance look better.

```
colnames(survey)[colnames(survey) ==
    "First, we are going to create a pseudonym for you to keep this survey anonymous (more or less). Wh

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) ==
    "Which musical instruments/talents do you play? (Select all that apply)"] <- "instrument_list"

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

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

```
colnames(survey)[colnames(survey) ==
    "YOB"] <- "year_born"

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


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"
```

We want to make the code tidy so we are going to use two packages that come with tidyverse but need to be loaded in to work.

```
suppressMessages(library(dplyr))
suppressMessages(library(tidyr))
```

Now we are going to move into creating seperate tables from these two dataframes we have. We will do this by using dplyr and tibbles. We are also going to manipulate the data for a few variables to standardize them and create proper levels for categorical variables.

```
Person <- tibble(time_submitted = survey$time_submitted,
    pseudonym_generator = survey$pseudonym_generator,
    pseudonym = survey$pseudonym, sex = survey$sex,
    year_born = survey$year_born, academic_level = survey$academic_level,
    academic_major = survey$academic_major,
    favorite_instrument = survey$instrument_list)

Person$academic_level <- as.factor(Person$academic_level)
Person$academic_major <- as.factor(Person$academic_major)

# Standardize spelling of variables for
# major
levels(Person$academic_major)[levels(Person$academic_major) ==
    "Computer information systems"] <- "Computer Information Systems"

colnames(Person)
```

```
## [1] "time_submitted"      "pseudonym_generator" "pseudonym"
## [4] "sex"                 "year_born"           "academic_level"
## [7] "academic_major"      "favorite_instrument"
```

```
#-----------------------------------------------------------------

FavoriteSong <- tibble(pseudonym = survey$pseudonym,
    song_artist = survey$favorite_song_artist,
    favorite_song = survey$favorite_song,
    song_link = survey$favorite_song_link)

colnames(FavoriteSong)
```

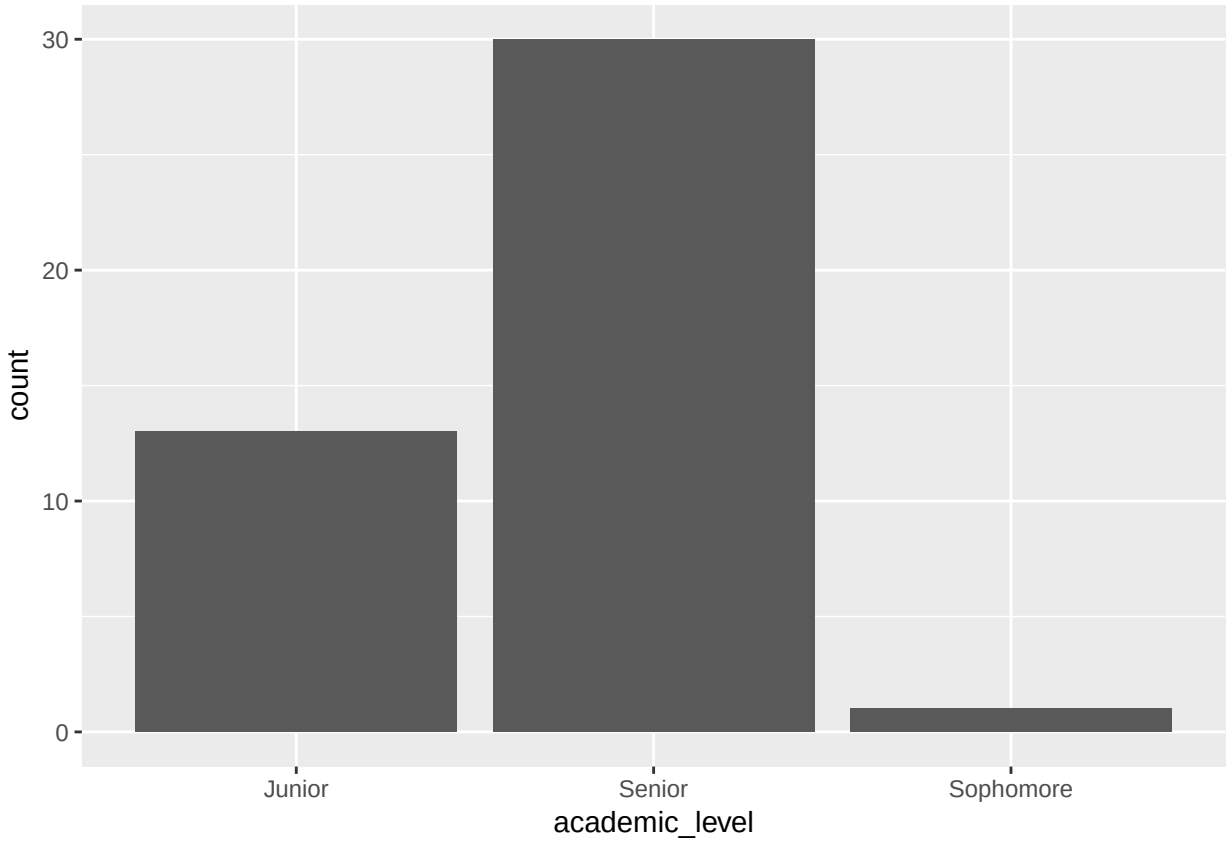
```
## [1] "pseudonym"     "song_artist"   "favorite_song" "song_link"
```

```
#-----------------------------------------------------------------
# This part was tricky. I manipulated the
# original dataframe preferences by using
# the gather function from Dplyr. This
# allowed me to lengthen data that was too
# wide by taking all the columns and
# making those rows under column name
# song_to_rate, then stored the ratings as
# keys called ratings.

preferences <- preferences %>% gather(song_to_rate,
    rating, 3:45)

# using this new manipulated preferences
# dataframe I can easily map the columns
# to my ratings tibble

Ratings <- tibble(pseudonym = preferences$`What was your pseudonym?`,
    song_to_rate = preferences$song_to_rate,
    ratings = preferences$rating)

Ratings$song_to_rate <- as.factor(Ratings$song_to_rate)
```

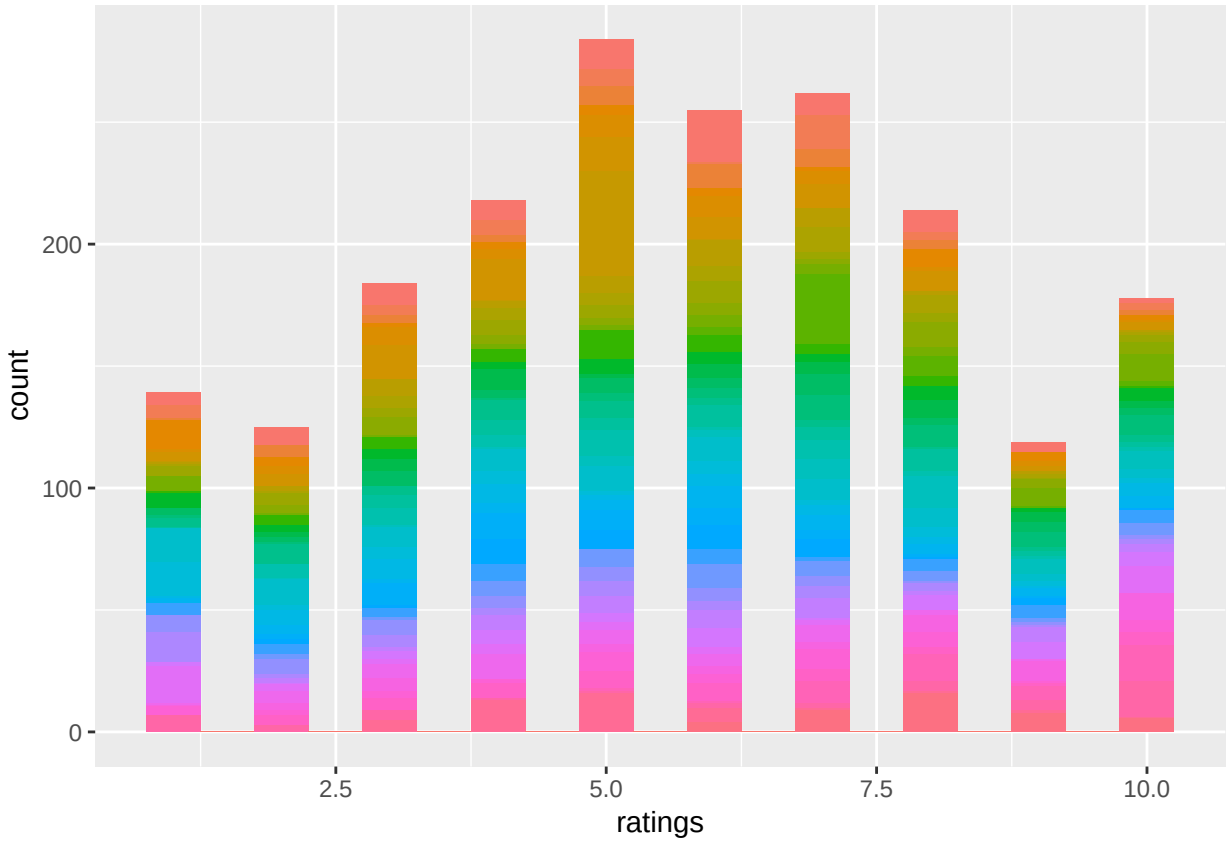
Time to manipulate the timestamps column in Person. We will be using a couple functions to format this column properly.

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

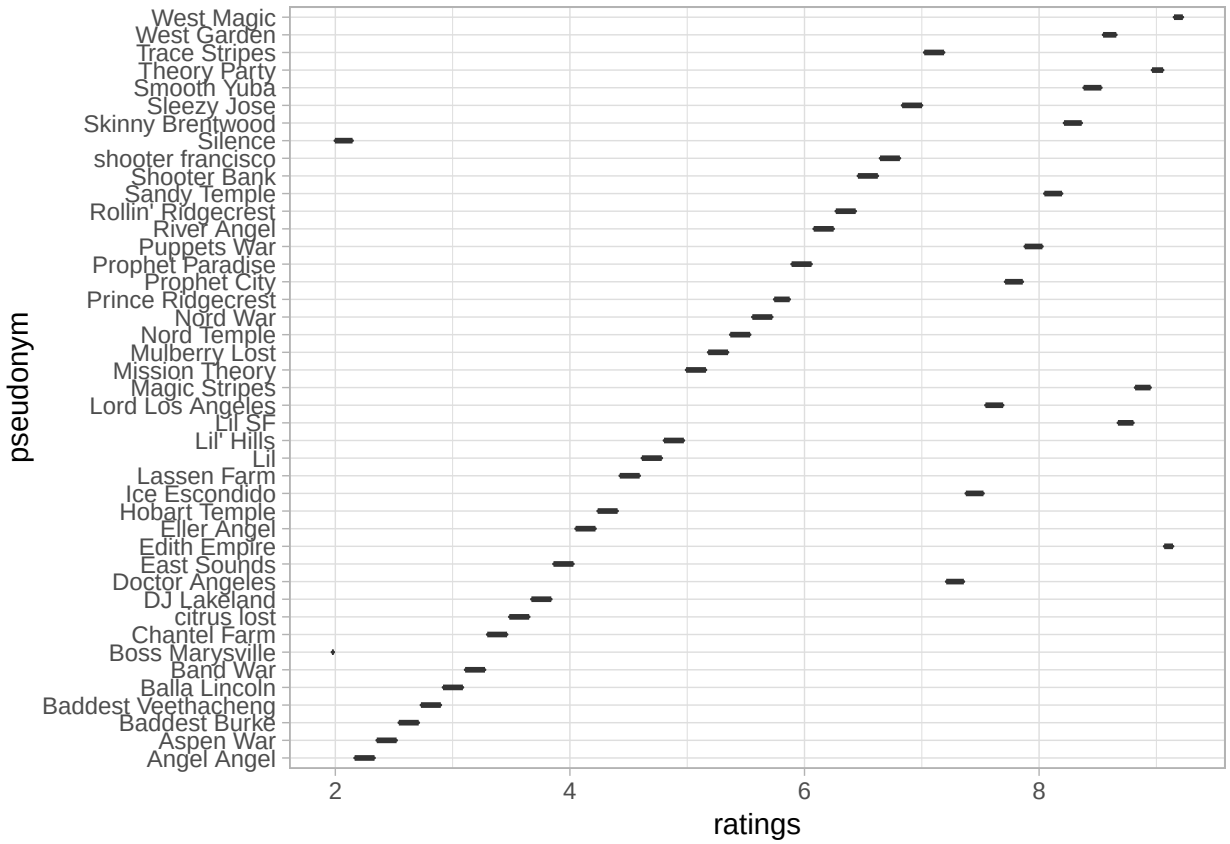
```
p <- ggplot(data=Person) + geom_bar(aes(academic_level))
p
```




```
p1 <- ggplot(data=Ratings,aes(ratings,fill=pseudonym)) +
      geom_histogram(binwidth = .5,show.legend = FALSE)

p1
```




```
p2 <- ggplot(data=Ratings,aes(ratings,pseudonym)) + geom_boxplot() + theme_light()
p2
```



Here we want to use join to get an observation for the following information. Pseudonym, Favorite Song, Rating

```
songs <- str_c(FavoriteSong$song_artist," ", FavoriteSong$favorite_song)
songs
```

```
##  [1] "40 crew Not Enough"
##  [2] "Arctic Monkeys Arabella"
##  [3] "Avatar The Eagle Has Landed"
##  [4] "Ben Folds Still"
##  [5] "blink-182 She's Out Of Her Mind"
##  [6] "brian jonestown massacre / sarabeth tucek Seer"
##  [7] "BROCKHAMPTON NO HALO"
##  [8] "Brother Ali Can't Take That Away"
##  [9] "Daboii Beat in a Chokehold"
## [10] "Daniel Caesar Open Up"
## [11] "De La Soul Patty Dooke"
## [12] "Dead Kennedys Kill The Poor"
## [13] "Dominic Fike 3 Nights"
## [14] "Don't have a favorite but here is Gorillaz Any but here is Humility"
## [15] "Dream Theater A Change of Seasons"
## [16] "Ed Sheeran Dive"
## [17] "General Levy Incredible (Remix)"
## [18] "Hiroyuki Sawano S-ave"
## [19] "Jai Paul Do You Love Her Now"
## [20] "Johnny Cash We'll Meet Again"
```

```
## [21] "Kane Brown What's Mine is Yours"
## [22] "Kim Petras I dont want it at all"
## [23] "Led Zeppelin Hey Hey What Can I Do"
## [24] "Marshmello One Thing Right"
## [25] "Matisyahu One Day"
## [26] "MF DOOM Potholderz"
## [27] "mozzy bladadah"
## [28] "Oliver Tree Hurt"
## [29] "Radiohead weird fishes arpeggi"
## [30] "Rainbow Kitten Surprise Cocaine Jesus"
## [31] "Rick Astley Never Going to Give You Up"
## [32] "Sia Elastic Heart"
## [33] "Sixtoo Jackals and Vipers in Envy of Man [Album]"
## [34] "Snakehips Either Way"
## [35] "Sufjan Stevens Should Have Known Better"
## [36] "Tame Impala Patience"
## [37] "The Black Keys Unknown Brother"
## [38] "The Byrds Mr. Spaceman"
## [39] "Tool Fear Inoculum"
## [40] "Traffic Dear Mr. Fantasy"
## [41] "Travis Scott Astrothunder"
## [42] "Virtual Self Angel Voices"
## [43] "Wheezer Buddy Holly"
## [44] "Martyrd<U+0326>d Harmagedon"
```

```
test <- semi_join(FavoriteSong, Ratings, by=  )
```

```
## Joining, by = "pseudonym"
```

```
test
```

```
## # A tibble: 43 x 4
##    pseudonym    song_artist            favorite_song   song_link
##    <chr>        <chr>                  <chr>           <chr>
##  1 Lord Los An~ 40 crew                Not Enough      https://www.youtube~
##  2 Puppets War  Avatar                 The Eagle Has ~ https://www.youtube~
##  3 West Magic   Ben Folds              Still           https://www.youtube~
##  4 Eller Angel  blink-182              She's Out Of H~ https://www.youtube~
##  5 Trace Strip~ brian jonestown massa~ Seer            https://youtu.be/C-~
##  6 Lil SF       BROCKHAMPTON           NO HALO         https://www.youtube~
##  7 Baddest Bur~ Brother Ali            Can't Take Tha~ https://youtu.be/mn~
##  8 Prince Ridg~ Daboii                 Beat in a Chok~ https://www.youtube~
##  9 Ice Escondi~ Daniel Caesar          Open Up         https://www.youtube~
## 10 Lassen Farm  De La Soul             Patty Dooke     https://www.youtube~
## # ... with 33 more rows
```