MusicInsights

Jasmin Mousavi

9/17/2019

##Packages and Installations

First, we need tidyverse to handle data importing and wrangling…

```r
#install.packages("tidyverse")
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   0.8.3      ✔ stringr 1.4.0
## ✔ readr   1.3.1      ✔ forcats 0.4.0
```

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

##Load Data

```r
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()
## )
```

```r
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.
```

Clean up column names for survey

```
#colnames(survey)

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 prefer?"]
<- "pseudonym_generator"
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)=="Year you were born (YYYY)"] <- "year_born"
colnames(survey)[colnames(survey)=="Which musical instruments/talents do you play? (Sele
ct 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_so
ng_link"

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"
```

Clean up preferences column names

```
colnames(preferences)
```

```
##  [1] "Timestamp"
##  [2] "What was your pseudonym?"
##  [3] "40 crew\tNot Enough"
##  [4] "Arctic Monkeys Arabella"
##  [5] "Avatar\tThe Eagle Has Landed"
##  [6] "Ben Folds\tStill"
##  [7] "blink-182\tShe's Out Of Her Mind"
##  [8] "brian jonestown massacre / sarabeth tucek\tSeer"
##  [9] "BROCKHAMPTON\tNO HALO"
## [10] "Brother Ali\tCan't Take That Away"
## [11] "Daboii\tBeat in a Chokehold"
## [12] "Daniel Caesar\tOpen Up"
## [13] "De La Soul\tPatty Dooke"
## [14] "Dead Kennedys\tKill The Poor"
## [15] "Dominic Fike\t3 Nights"
## [16] "Gorillaz\tAny but here is Humility"
## [17] "Dream Theater\tA Change of Seasons"
## [18] "Ed Sheeran\tDive"
## [19] "General Levy\tIncredible (Remix)"
## [20] "Hiroyuki Sawano\tS-ave"
## [21] "Jai Paul\tDo You Love Her Now"
## [22] "Johnny Cash\tWe'll Meet Again"
## [23] "Kane Brown\tWhat's Mine is Yours"
## [24] "Kim Petras\tI dont want it at all"
## [25] "Led Zeppelin\tHey Hey What Can I Do"
## [26] "Marshmello\tOne Thing Right"
## [27] "Matisyahu\tOne Day"
## [28] "MF DOOM\tPotholderz"
## [29] "mozzy\tbladadah"
## [30] "Oliver Tree\tHurt"
## [31] "Radiohead\tweird fishes arpeggi"
## [32] "Rainbow Kitten Surprise\tCocaine Jesus"
## [33] "Rick Astley\tNever Going to Give You Up"
## [34] "Sia\tElastic Heart"
## [35] "Sixtoo\tJackals and Vipers in Envy of Man [Album]"
## [36] "Snakehips\tEither Way"
## [37] "Sufjan Stevens\tShould Have Known Better"
## [38] "Tame Impala\tPatience"
## [39] "The Black Keys\tUnknown Brother"
## [40] "The Byrds\tMr. Spaceman"
## [41] "Tool\tFear Inoculum"
## [42] "Traffic\tDear Mr. Fantasy"
## [43] "Travis Scott\tAstrothunder"
## [44] "Virtual Self\tAngel Voices"
## [45] "Wheezer\tBuddy Holly"
```

```
colnames(preferences)[colnames(preferences)=="Timestamp"] <- "time_submitted"
colnames(preferences)[colnames(preferences)=="What was your pseudonym?"] <- "pseudonym"
```

##Create tables

```r
library("dplyr")

Person <- tibble(pseudonym_generator=survey$pseudonym_generator, pseudonym=survey$pseudo
nym, sex=survey$sex, academic_major=survey$academic_major, academic_level=survey$academi
c_level, year_born=survey$year_born, time_submitted=survey$time_submitted)

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

#Ratings
Ratings <- preferences[2:45]
Ratings <- Ratings %>% gather(colnames(Ratings[2:44]), key="artist_song", value="rating
s")
Ratings
```

```
## # A tibble: 1,978 x 3
##    pseudonym          artist_song          ratings
##    <chr>              <chr>                  <dbl>
##  1 Baddest Burke      "40 crew\tNot Enough"      3
##  2 Shooter Bank       "40 crew\tNot Enough"      4
##  3 Prophet Paradise   "40 crew\tNot Enough"      4
##  4 Smooth Yuba        "40 crew\tNot Enough"      8
##  5 Silence            "40 crew\tNot Enough"      3
##  6 West Garden        "40 crew\tNot Enough"      4
##  7 East Sounds        "40 crew\tNot Enough"      4
##  8 Balla Lincoln      "40 crew\tNot Enough"      1
##  9 Theory Party       "40 crew\tNot Enough"      8
## 10 Rollin' Ridgecrest "40 crew\tNot Enough"      8
## # … with 1,968 more rows
```

```r
Person$time_submitted<- as.POSIXlt(parse_datetime(Person$time_submitted, format="%m/%d/%
y %H:%M"))
Person$time_submitted[1]$min
```

```
## [1] 22
```

```r
#Levels -> categorical variable
Person$academic_level <- as.factor(Person$academic_level)
Person$academic_major <- as.factor(Person$academic_major)
Ratings$artist_song <- as.factor(Ratings$artist_song)
Ratings$pseudonym <- as.factor(Ratings$pseudonym)
levels(Person$academic_major)[levels(Person$academic_major) == "Computer information sys
tems"] <- "Computer Information Systems"
```