Refinement Practice Lab

Ben Goldstone

9/11/2023

Import Taylor Swift Data and Mutate

```
ts <- read_csv("~/CSVs/ts.csv")
```

```
## New names:
## * `` -> ...1
```

```
## Rows: 242 Columns: 18
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (4): name, album, id, uri
## dbl  (13): ...1, track_number, acousticness, danceability, energy, instrumen...
## date  (1): release_date
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(ts)
```

```
## # A tibble: 6 x 18
##    ...1 name     album    release_date track_number id     uri      acousticness
##   <dbl> <chr>    <chr>    <date>              <dbl> <chr>  <chr>           <dbl>
## 1     0 Mine (T~ Speak N~ 2023-07-07              1 7G0gB~ spotify~      0.00444
## 2     1 Sparks ~ Speak N~ 2023-07-07              2 3MytW~ spotify~      0.0251
## 3     2 Back To~ Speak N~ 2023-07-07              3 79uDO~ spotify~      0.00621
## 4     3 Speak N~ Speak N~ 2023-07-07              4 5xXqy~ spotify~      0.248
## 5     4 Dear Jo~ Speak N~ 2023-07-07              5 1zU8j~ spotify~      0.0236
## 6     5 Mean (T~ Speak N~ 2023-07-07              6 30Y4C~ spotify~      0.311
## # ... with 10 more variables: danceability <dbl>, energy <dbl>,
## #   instrumentalness <dbl>, liveness <dbl>, loudness <dbl>, speechiness <dbl>,
## #   tempo <dbl>, valence <dbl>, popularity <dbl>, duration_ms <dbl>
```

```
tally(~album,data=ts)
```

```
## album
##                                                                 evermore
##                                                                       15
##                                                   evermore (deluxe version)
##                                                                       17
##                                                 Fearless (Taylor's Version)
##                                                                       26
##                                                                 folklore
##                                                                       16
```

```
##                                                                folklore (deluxe version)
##                                                                                       17
## folklore: the long pond studio sessions (from the Disney+ special) [deluxe edition]
##                                                                                       34
##                                                                                    Lover
##                                                                                        9
##                                                                                Midnights
##                                                                                       13
##                                                                  Midnights (3am Edition)
##                                                                                       20
##                                                           Midnights (The Til Dawn Edition)
##                                                                                       23
##                                                                   Red (Taylor's Version)
##                                                                                       30
##                                                              Speak Now (Taylor's Version)
##                                                                                       22
```

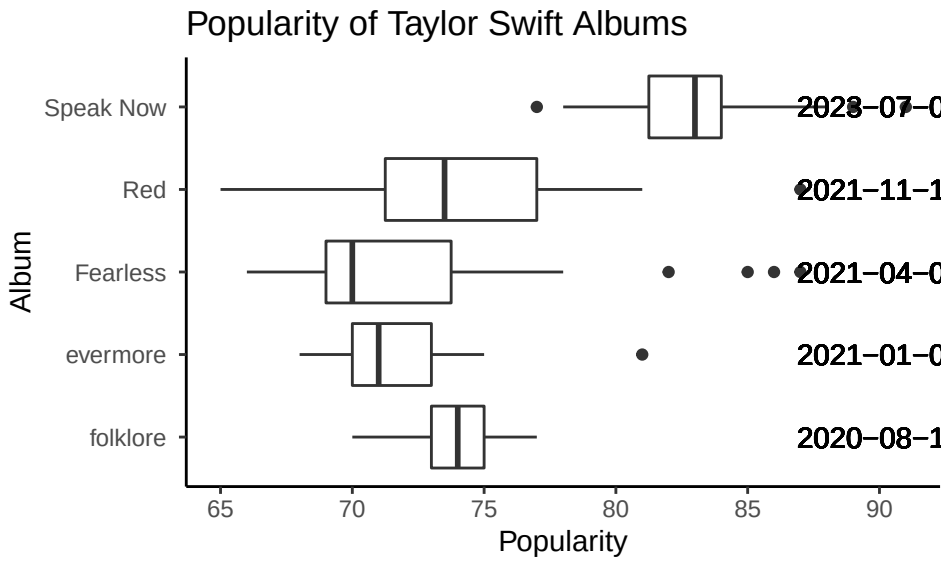
```
ts = ts %>% mutate(Keep = ifelse(str_detect(album,"ersion"),"Yes","No"),
                   ShortAlbum = gsub("[(].*","",album))
```

Popularity Album Boxplot

```
head(ts)
```

```
## # A tibble: 6 x 20
##    ...1 name     album    release_date track_number id     uri      acousticness
##   <dbl> <chr>    <chr>    <date>              <dbl> <chr>  <chr>           <dbl>
## 1     0 Mine (T~ Speak N~ 2023-07-07              1 7G0gB~ spotify~      0.00444
## 2     1 Sparks ~ Speak N~ 2023-07-07              2 3MytW~ spotify~      0.0251
## 3     2 Back To~ Speak N~ 2023-07-07              3 79uDO~ spotify~      0.00621
## 4     3 Speak N~ Speak N~ 2023-07-07              4 5xXqy~ spotify~      0.248
## 5     4 Dear Jo~ Speak N~ 2023-07-07              5 1zU8j~ spotify~      0.0236
## 6     5 Mean (T~ Speak N~ 2023-07-07              6 30Y4C~ spotify~      0.311
## # ... with 12 more variables: danceability <dbl>, energy <dbl>,
## #   instrumentalness <dbl>, liveness <dbl>, loudness <dbl>, speechiness <dbl>,
## #   tempo <dbl>, valence <dbl>, popularity <dbl>, duration_ms <dbl>,
## #   Keep <chr>, ShortAlbum <chr>
```

```
gf_boxplot(fct_reorder(ShortAlbum, release_date)~popularity,data=subset(ts,Keep == "Yes")) %>% gf_text(
```



Import UFO Dataset

```
ufo <- read_csv("~/CSVs/ufo.csv")
```

```
## Rows: 60632 Columns: 16
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (5): Location.City, Location.State, Location.Country, Data.Shape, Data....
## dbl (11): Data.Encounter duration, Location.Coordinates.Latitude, Location.C...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(ufo)
```

```
## # A tibble: 6 x 16
##   Location.City Location.State Location.Country Data.Shape `Data.Encounter dura~
##   <chr>         <chr>          <chr>            <chr>                        <dbl>
## 1 anchor point  AK             US               disk                           300
## 2 anchorage     AK             US               changing                     21600
## 3 anchorage     AK             US               changing                       600
## 4 anchorage     AK             US               cigar                           15
## 5 anchorage     AK             US               circle                         300
## 6 anchorage     AK             US               circle                           4
## # ... with 11 more variables: Data.Description excerpt <chr>,
## #   Location.Coordinates.Latitude <dbl>, Location.Coordinates.Longitude <dbl>,
## #   Dates.Sighted.Year <dbl>, Dates.Sighted.Month <dbl>,
## #   Date.Sighted.Day <dbl>, Dates.Sighted.Hour <dbl>,
## #   Dates.Sighted.Minute <dbl>, Dates.Documented.Year <dbl>,
## #   Dates.Documented.Month <dbl>, Dates.Documented.Day <dbl>
```

```
colnames(ufo)[12] = "SightedHour"
colnames(ufo)
```

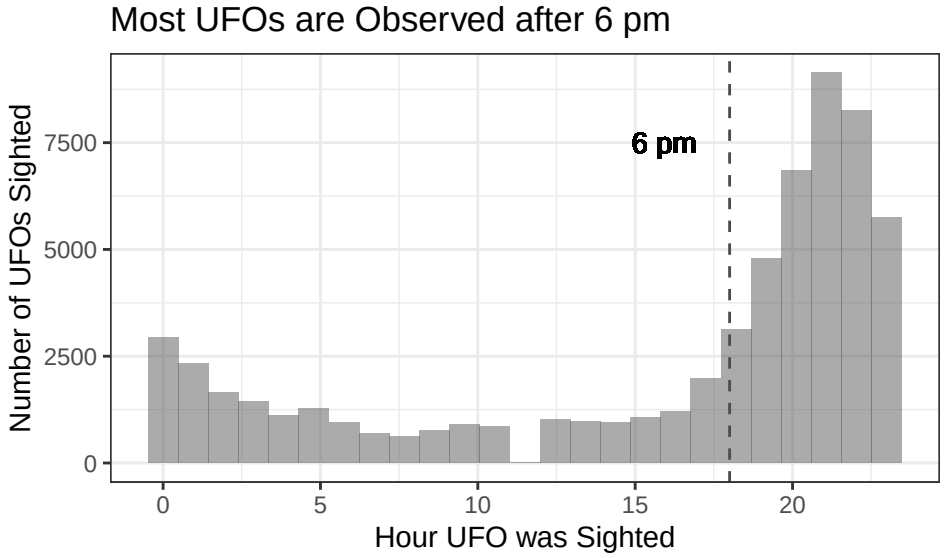
```
##  [1] "Location.City"                 "Location.State"
##  [3] "Location.Country"              "Data.Shape"
##  [5] "Data.Encounter duration"       "Data.Description excerpt"
##  [7] "Location.Coordinates.Latitude" "Location.Coordinates.Longitude"
##  [9] "Dates.Sighted.Year"            "Dates.Sighted.Month"
```

```
## [11] "Date.Sighted.Day"                "SightedHour"
## [13] "Dates.Sighted.Minute"            "Dates.Documented.Year"
## [15] "Dates.Documented.Month"          "Dates.Documented.Day"
```

UFOs Observed Histogram

```
gf_histogram(~SightedHour, data=ufo) %>%
  gf_vline(xintercept = 18, linetype = "dashed", color="gray30") %>%
  gf_text(y=7500,x=18,hjust=1.5,label="6 pm") %>%
  gf_theme(theme_bw()) %>%
  gf_labs(x="Hour UFO was Sighted",y="Number of UFOs Sighted",title="Most UFOs are Observed after 6 pm")
```

```
## Warning: geom_vline(): Ignoring `mapping` because `xintercept` was provided.
```



Import StudentSurvey

```
require(Lock5Data)
data("StudentSurvey")
head(StudentSurvey)
```

```
##        Year Sex Smoke   Award HigherSAT Exercise TV Height Weight Siblings
## 1    Senior   M    No Olympic      Math       10  1     71    180        4
## 2 Sophomore   F   Yes Academy      Math        4  7     66    120        2
## 3 FirstYear   M    No   Nobel      Math       14  5     72    208        2
## 4    Junior   M    No   Nobel      Math        3  1     63    110        1
## 5 Sophomore   F    No   Nobel    Verbal        3  3     65    150        1
## 6 Sophomore   F    No   Nobel    Verbal        5  4     65    114        2
##   BirthOrder VerbalSAT MathSAT  SAT  GPA Pulse Piercings
## 1          4       540     670 1210 3.13    54         0
## 2          2       520     630 1150 2.50    66         3
## 3          1       550     560 1110 2.55   130         0
## 4          1       490     630 1120 3.10    78         0
## 5          1       720     450 1170 2.70    40         6
## 6          2       600     550 1150 3.20    80         4
```

Student Awards

```
StudentSurvey$Year = factor(StudentSurvey$Year, levels=c("FirstYear","Sophomore","Junior","Senior"))
gf_bar(~Award,data=na.omit(StudentSurvey), fill=~Award,show.legend = FALSE) %>%
  gf_facet_wrap(~Year, nrow=1) %>%
  gf_theme(theme_bw()) %>% gf_labs(title="Awards Received by Class Year",y="Number of Awards Received")
```