230917 DATA 607 Siebecker Assignment 2

Molly Siebecker

2023-09-17

Overview

I distributed a survey through Google Forms, asking participants for their gender identity and their movie ratings. I wanted to see specifically if women were more likely to rate Barbie higher than men, and whether there were any other apparent gender-based trends for other recent movies. After downloading the Google Sheets data as a csv, I created a table for the data in PostgreSQL. Below, I connect R to PostgreSQL and read the table into R as a data frame.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(dbplyr)
```

```
##
## Attaching package: 'dbplyr'

## The following objects are masked from 'package:dplyr':
##
##     ident, sql
```

```
library(odbc)
library(RODBC)
library(DBI)
library(knitr)
```

Checking to See if R Can Connect to PostgreSQL

```
con <- dbCanConnect(RPostgres::Postgres(),
                    dbname = "my_database",
                    port = 5432,
                    user = "postgres",
                    password = [redacted])
con
```

```
## [1] TRUE
```

Connecting R To PostgreSQL

```
con <- dbConnect(RPostgres::Postgres(),
                    dbname = "my_database",
                    port = 5432,
                    user = "postgres",
                    password = [redacted])
con
```

```
## <PqConnection> my_database@/tmp:5432
```

Reading the SQL Table into an R Data Frame

```
movie_survey <- dbReadTable(con, "movie_survey")
```

Renaming Columns and Replacing Values I renamed the columns to make them clearer to read and indicate that each value was a rating. I also replaced the single letters indicating gender identity with the entire word or phrase.

```
movie_survey <- movie_survey %>%
rename("Gender_Identity" = "Gender.Identity", "The_Banshees_of_Inisherin_Rating" = "The.Banshees.of.Inis
movie_survey$Gender_Identity[movie_survey$Gender_Identity=="M"] <- "Man"
movie_survey$Gender_Identity[movie_survey$Gender_Identity=="W"] <- "Woman"
movie_survey$Gender_Identity[movie_survey$Gender_Identity=="P"] <- "Prefer Not to Say"
```

**Creating New Data Frame without Timestamp Column for Table" I focused on the columns that would be of interest to someone viewing a table of the data.

```
all_movie_survey <- select (movie_survey, c(Gender_Identity,The_Banshees_of_Inisherin_Rating,Barbie_Rati
```

Creating a Table of the Data

```
kable(all_movie_survey, col.names=c("Gender Identity","The Banshees of Inisherin Rating","Barbie Rating
```

Table 1: Gender Identity and Popular Movie Ratings

Gender Identity	The Banshees of Inisherin Rating	Barbie Rating	Everything Everywhere All At Once Rating	Glass Onion A Knives Out Mystery Rating	Oppenheimer Rating	Top Gun Maverick Rating
Man	5	4	5	3	4	3
Prefer Not to Say	NA	4	NA	5	NA	NA
Woman	NA	NA	NA	2	NA	NA
Woman	NA	3	4	NA	NA	NA
Woman	NA	5	3	3	NA	NA
Man	NA	4	5	3	4	NA
Woman	NA	4	5	NA	NA	NA
Woman	NA	4	NA	5	NA	NA
Man	NA	NA	NA	NA	4	4
Man	5	4	3	2	5	4
Woman	NA	NA	4	NA	NA	NA
Man	NA	NA	NA	4	NA	NA
Woman	4	4	NA	3	5	3
Woman	NA	NA	NA	4	NA	NA
Woman	5	4	NA	5	5	3
Man	4	5	5	4	NA	NA
Woman	3	5	5	3	NA	NA

Creating Subsets of Women and Men Since only one respondent selected "Prefer Not to Say," I did not create a subset of that response.

```
men_movie_survey <- select (filter(all_movie_survey, Gender_Identity=="Man"), c(Gender_Identity,The_Bans
women_movie_survey <- select (filter(all_movie_survey, Gender_Identity=="Woman"), c(Gender_Identity,The
```

Calculating Summary Statistics of All Respondents

```
summary(all_movie_survey)
```

```
##  Gender_Identity    The_Banshees_of_Inisherin_Rating Barbie_Rating
##  Length:17          Min.   :3.000                    Min.   :3.000
##  Class :character   1st Qu.:4.000                    1st Qu.:4.000
##  Mode  :character   Median :4.500                    Median :4.000
##                     Mean   :4.333                    Mean   :4.167
##                     3rd Qu.:5.000                    3rd Qu.:4.250
##                     Max.   :5.000                    Max.   :5.000
##                     NA's   :11                       NA's   :5
##  Everything_Everywhere_All_At_Once_Rating
##  Min.   :3.000
##  1st Qu.:4.000
##  Median :5.000
##  Mean   :4.333
##  3rd Qu.:5.000
##  Max.   :5.000
##  NA's   :8
##  Glass_Onion_A_Knives_Out_Mystery_Rating Oppenheimer_Rating
```

```
##  Min.   :2.000                          Min.   :4.0
##  1st Qu.:3.000                          1st Qu.:4.0
##  Median :3.000                          Median :4.5
##  Mean   :3.538                          Mean   :4.5
##  3rd Qu.:4.000                          3rd Qu.:5.0
##  Max.   :5.000                          Max.   :5.0
##  NA's   :4                              NA's   :11
##  Top_Gun_Maverick_Rating
##  Min.   :3.0
##  1st Qu.:3.0
##  Median :3.0
##  Mean   :3.4
##  3rd Qu.:4.0
##  Max.   :4.0
##  NA's   :12
```

Calculating Summary Statistics of Male Respondents

```
summary(men_movie_survey)
```

```
##  Gender_Identity    The_Banshees_of_Inisherin_Rating Barbie_Rating
##  Length:6           Min.   :4.000                    Min.   :4.00
##  Class :character   1st Qu.:4.500                    1st Qu.:4.00
##  Mode  :character   Median :5.000                    Median :4.00
##                     Mean   :4.667                    Mean   :4.25
##                     3rd Qu.:5.000                    3rd Qu.:4.25
##                     Max.   :5.000                    Max.   :5.00
##                     NA's   :3                        NA's   :2
##  Everything_Everywhere_All_At_Once_Rating
##  Min.   :3.0
##  1st Qu.:4.5
##  Median :5.0
##  Mean   :4.5
##  3rd Qu.:5.0
##  Max.   :5.0
##  NA's   :2
##  Glass_Onion_A_Knives_Out_Mystery_Rating Oppenheimer_Rating
##  Min.   :2.0                             Min.   :4.00
##  1st Qu.:3.0                             1st Qu.:4.00
##  Median :3.0                             Median :4.00
##  Mean   :3.2                             Mean   :4.25
##  3rd Qu.:4.0                             3rd Qu.:4.25
##  Max.   :4.0                             Max.   :5.00
##  NA's   :1                               NA's   :2
##  Top_Gun_Maverick_Rating
##  Min.   :3.000
##  1st Qu.:3.500
##  Median :4.000
##  Mean   :3.667
##  3rd Qu.:4.000
##  Max.   :4.000
##  NA's   :3
```

Calculating Summary Statistics of Female Respondents

```r
summary(women_movie_survey)
```

```
##  Gender_Identity    The_Banshees_of_Inisherin_Rating Barbie_Rating
##  Length:10          Min.   :3.0                      Min.   :3.000
##  Class :character   1st Qu.:3.5                      1st Qu.:4.000
##  Mode  :character   Median :4.0                      Median :4.000
##                     Mean   :4.0                      Mean   :4.143
##                     3rd Qu.:4.5                      3rd Qu.:4.500
##                     Max.   :5.0                      Max.   :5.000
##                     NA's   :7                        NA's   :3
##  Everything_Everywhere_All_At_Once_Rating
##  Min.   :3.0
##  1st Qu.:4.0
##  Median :4.0
##  Mean   :4.2
##  3rd Qu.:5.0
##  Max.   :5.0
##  NA's   :5
##  Glass_Onion_A_Knives_Out_Mystery_Rating Oppenheimer_Rating
##  Min.   :2.000                           Min.   :5
##  1st Qu.:3.000                           1st Qu.:5
##  Median :3.000                           Median :5
##  Mean   :3.571                           Mean   :5
##  3rd Qu.:4.500                           3rd Qu.:5
##  Max.   :5.000                           Max.   :5
##  NA's   :3                               NA's   :8
##  Top_Gun_Maverick_Rating
##  Min.   :3
##  1st Qu.:3
##  Median :3
##  Mean   :3
##  3rd Qu.:3
##  Max.   :3
##  NA's   :8
```

Findings and Conclusions Women actually rated Barbie slightly lower on average than men did (4.143 compared to 4.25.) Similarly, the women and men surveyed saw Barbie at similar rates (around 30-33% of each group had not seen it.) A noticeable trend was that the subset of women were more likely to have seen Barbie than the other films (the other only movie that had 7 female viewers was Glass Onion, which was released on streaming.) Men saw the six films at relatively similar rates. Out of the 6 men surveyed, each movie was seen by between 3-5 of them. However, the sample size is too small (and not random enough) to perform any meaningful analysis, and distribution at a larger scale might show different results.

In my original survey, I had required a response for each movie, allowing participants to select "Have Not Seen" for movies they had not seen. However, this became a problem when I was creating a table in SQL and had different data types in the same column, which forced me to go back and manually delete those strings from the spreadsheet before downloading a new csv. However, this still made it possible to treat each null value as a positive indication that they had not seen the movie, and not as a non-response.