Mini Project #1

Kalide Endale

10/5/2021

Contents

```
library(readr)
library(tibble)
library(ggplot2)
library(tidyr)
```

Importing dataset into R

Source: *https://www.kaggle.com/bidyutchanda/does-time-prefer-men-or-women-on-their-covers*

```
dir<-"/Users/Kalide/Documents/Northeastern/Introduction to Data Management and Processing"
dir00<- "Homeworks and Exercsies/TIMEGenderData.csv"
path0 <- file.path(dir,dir00)
TIMEGenderData <- read_csv(path0)
```

```
## Rows: 91 Columns: 6
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Female %, Male %
## dbl (4): Year, Female, Male, Total
```

```
##
## 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.
```

I pulled this dataset from Kaggle. My dataset is a record of all the appearances that were made by both male(variable 3) and female(variable 2) personalities for the cover of Time Magazine from 1923(when Time magazine was founded) to 2013(variable 1). The Time is a popular American news magazine publishing company that is famous for its magazines which were published weekly and as of 2020 published bi-weekly. According to a 2012 statistic, Time had a circulation of 3.3 million, making it the second-most circulated magazine behind "People" magazine. I choose this dataset to see if Time magazine has any gender-specific bias when choosing personalities for their cover pages. Additionally, I wanted to see if a bias exists, are these biases being exacerbated or is Time magazine making an effort to uphold equality in their selection process? As stewards of a very popular magazine, the role Time magazine has in shaping our perception and culture is immense so it's profoundly important for this institution to practice equality in its selection process. Additionally, these cover pages are also snapshots of American news makers so this dataset could lend us a window into analyzing American culture moments. *Showcase dataset*

```
## # A tibble: 91 x 6
##     Year Female  Male Total `Female %` `Male %`
##    <dbl>  <dbl> <dbl> <dbl> <chr>      <chr>
##  1  1923      1    34    35 2.86%      97.14%
##  2  1924      4    48    52 7.69%      92.31%
```

```
##  3  1925      1    51    52 1.92%      98.08%
##  4  1926      7    46    52 13.46%     88.46%
##  5  1927      4    49    52 7.69%      94.23%
##  6  1928      5    47    52 9.62%      90.38%
##  7  1929      7    45    52 13.46%     86.54%
##  8  1930      6    46    52 11.54%     88.46%
##  9  1931      6    45    50 12.00%     90.00%
## 10  1932      3    48    51 5.88%      94.12%
## # ... with 81 more rows
```

Quick review of the dataset

```
summary(TIMEGenderData)
```

```
##       Year          Female            Male           Total
##  Min.   :1923   Min.   : 1.000   Min.   :20.00   Min.   :26.00
##  1st Qu.:1946   1st Qu.: 4.000   1st Qu.:33.00   1st Qu.:41.00
##  Median :1968   Median : 5.000   Median :41.00   Median :47.00
##  Mean   :1968   Mean   : 5.835   Mean   :39.55   Mean   :45.36
##  3rd Qu.:1990   3rd Qu.: 7.000   3rd Qu.:46.00   3rd Qu.:51.00
##  Max.   :2013   Max.   :18.000   Max.   :51.00   Max.   :52.00
##    Female %           Male %
##  Length:91          Length:91
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

Change column names 'female %' and 'male %'

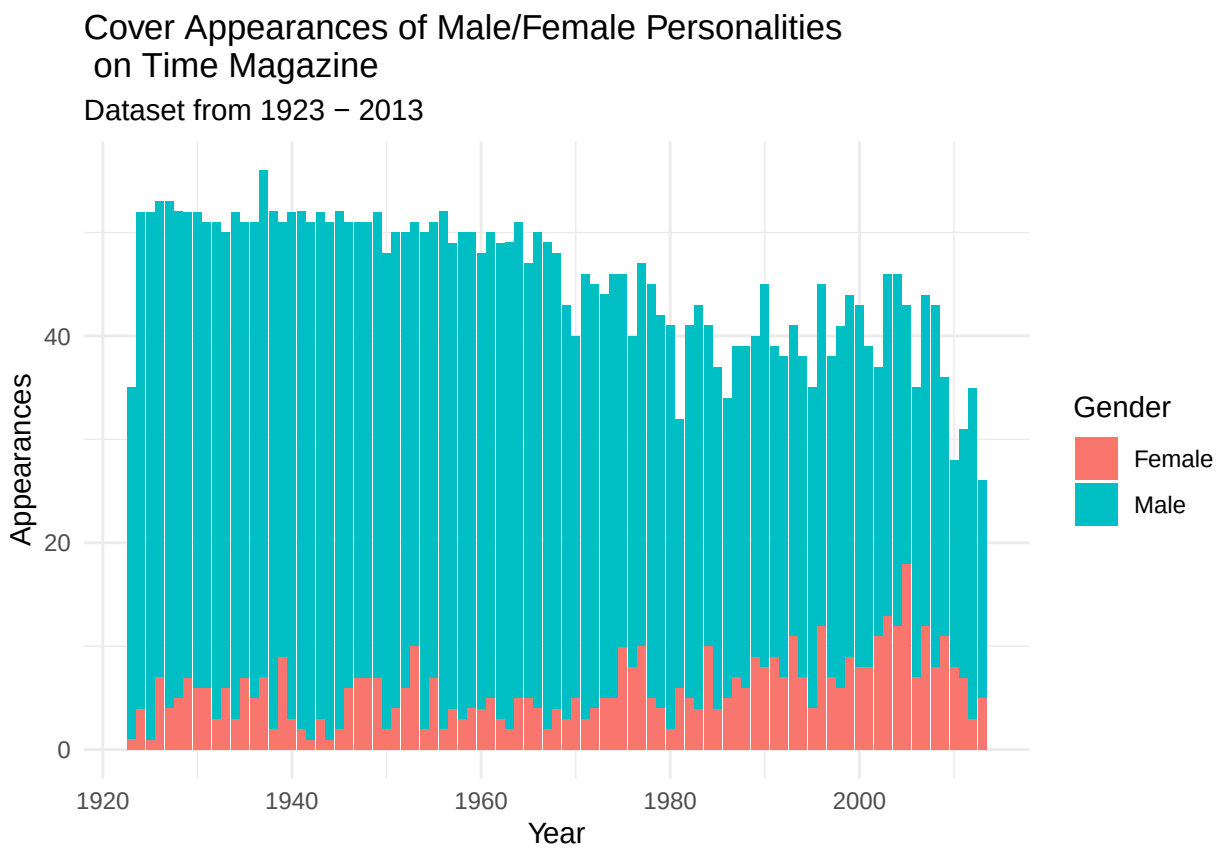
```
colnames(TIMEGenderData)[5] <- "Female_pct"
colnames(TIMEGenderData)[6] <- "Male_pct"
```

Tidy the data for simple analysis: put male and female under one column

```
TIMEGconsolidated <- TIMEGenderData[1:3]
TIMEGconsolidated.long <- pivot_longer(TIMEGconsolidated,
                                       cols = 2:3, names_to = "Gender",
                                       values_to = "Appearances")
```

The bar chart shows that males have dominated Time magazine cover pages for a long period of time. This is likely due to many factors, but it points to a sexism culture that persists in the US. We can see that during the 1920s and 1940s, Men were presented on nearly all of the cover pages. In the later years, we can see spikes of female personalities gracing the cover pages of Time magazine.

```
ggplot(TIMEGconsolidated.long, aes(x = Year, y = Appearances,
                                   group = Gender, fill = Gender)) +
  geom_col(position = position_stack(reverse = TRUE)) +
  labs(x = "Year", y= "Appearances",
       title = "Cover Appearances of Male/Female Personalities \n on Time Magazine",
       subtitle = "Dataset from 1923 - 2013") +
  theme(plot.title = element_text(hjust = 0), face = "bold") +
  theme_minimal()
```



tidy dataset, by removing the percentage sign from our two columns

```
TIMEGenderData2 <- TIMEGenderData
TIMEGenderData2$Female_pct <- as.numeric(gsub("\\%", "", TIMEGenderData$Female_pct))
TIMEGenderData2$Male_pct <-  as.numeric(gsub("\\%", "", TIMEGenderData$Male_pct))
```

Additional Analysis: Although males have dominated the cover pages of Time magazine for a long period of time, it does seem like this trend is decreasing. This is shown by the negative linear regression slope for males and the positive linear regression slope for females. This is a positive step and shows that Time magazine is making progress in portraying both genders equally. Although, much more work is still needed.

```
ggplot(TIMEGenderData2) +
  geom_point(data = TIMEGenderData2, aes(x = Year, y = Female_pct, color = 'Female')) +
  geom_point(data = TIMEGenderData2, aes(x = Year, y= Male_pct, color = 'Male'),
             shape = 18) +

  geom_smooth(data = TIMEGenderData2, aes(x=Year, y = Female_pct),
              fromula = y~x, method = "lm") +
  geom_smooth(data = TIMEGenderData2, aes(x=Year, y = Male_pct),
              formula = y~x, method = "lm") +

   labs(x = "Year", y= "Percentage of Personalities on Cover",
       title = "Trend of Males on Time Magazine Cover Diminishing",
       subtitle = "Dataset from 1923 - 2013",
       color = "Gender") +
  theme_minimal()
```

```
## Warning: Ignoring unknown parameters: fromula
```

```
## `geom_smooth()` using formula 'y ~ x'
```