Nobel Prize Gender Breakdown Lab

Tiffany Xiao

10/28/2022

The Gender Divide in Nobel Prizes

Introduction

Nobel Prizes are among the most renowned accolades for scholars in physics, chemistry, medicine, literature, economics, and peace. Each year, a prize ceremony takes place, where individuals are awarded medallions, diplomas, and monetary gifts to recognize significant contributions to their field.

Most, if not all, global societies were built on a patriarchal structure. Even now, the term "masculinity" is associated with a sense of power, competence, and courage while "femininity" is associated with fragility, beauty, and ignorance. Understanding the widespread impact of sexist history, policy, and stereotypes, I was immediately drawn to the topic of gender divide (in this context: binary) when I saw the data set detailing Nobel Prizes.

My Questions

My overarching question: Is there a gender divide within Nobel Prize winners?

Anticipating that there would be a gender divide, I was also curious: 1. How does the gender divide translate into different fields of study? Is it more prominent in one than the other? 2. Is there an increase in the number of women who win Nobel Prizes over time with the evolution of social norms and women's increased accessibility to education?

The Data Set

```
nobel.data <- read.csv("nobel_final.csv")
head(nobel.data)
```

```
##          firstname    surname born_country_code died_country_code gender year
## 1 Wilhelm Conrad    Röntgen                DE                DE   male 1901
## 2     Hendrik A.    Lorentz                NL                NL   male 1902
## 3         Pieter     Zeeman                NL                NL   male 1902
## 4           Henri  Becquerel                FR                FR   male 1903
## 5          Pierre      Curie                FR                FR   male 1903
## 6           Marie      Curie                PL                FR female 1903
##    category share
## 1  physics     1
## 2  physics     2
## 3  physics     2
## 4  physics     2
## 5  physics     4
## 6  physics     4
##                                                                         name_o
```

```
## 1                                                                                              Munich
## 2                                                                                              Leiden
## 3                                                                                              Amsterdam
## 4                                                                                              École Po
## 5 École municipale de physique et de chimie industrielles (Municipal School of Industrial Physics and
## 6
##   city_of_university country_of_university born_month age age_get_prize
## 1             Munich               Germany        Mar  78            56
## 2             Leiden       the Netherlands        Jul  75            49
## 3          Amsterdam       the Netherlands        May  78            37
## 4              Paris                France        Dec  56            51
## 5              Paris                France        May  47            44
## 6                                                Nov  67            36
```

The variables I use in my analysis include: 1. gender of award recipient (female or male) 2. year Prize was won (year) 3. category of award (physics, chemistry, medicine, economics, literature, peace)

Prior to using the data set to research relationships, I reworked variables into factors so that I would be able to make barplots, which are most helpful in helping to draw conclusions on my questions.

```
#changing data into factors
nobel.data$gender <- as.factor(nobel.data$gender)
nobel.data$category <- as.factor(nobel.data$category)
head(nobel.data$category)
```

```
## [1] physics physics physics physics physics physics
## Levels: chemistry economics literature medicine peace physics
```

```
nobel.data$country_of_university <- as.factor(nobel.data$country_of_university)
head(nobel.data$country_of_university)
```

```
## [1] Germany          the Netherlands the Netherlands France
## [5] France
## 30 Levels:  Argentina Australia Austria Belgium Canada ... USSR (now Russia)
```

Question 1: Is there a gender divide within Nobel Prize winners?

To create a stacked proportion graph with two entries in the *prop.table*, I first created a new variable recognizing "winners" (everybody is a winner).

```
#everyone is a winner!
nobel.data$winner <- nobel.data$age_get_prize[nobel.data$age_get_prize > 0]
head(nobel.data)
```

```
##          firstname    surname born_country_code died_country_code gender year
## 1 Wilhelm Conrad    Röntgen                DE                DE   male 1901
## 2     Hendrik A.    Lorentz                NL                NL   male 1902
## 3         Pieter     Zeeman                NL                NL   male 1902
## 4          Henri  Becquerel                FR                FR   male 1903
## 5         Pierre      Curie                FR                FR   male 1903
## 6          Marie      Curie                PL                FR female 1903
##   category share
## 1  physics     1
## 2  physics     2
## 3  physics     2
## 4  physics     2
## 5  physics     4
## 6  physics     4
```

```
##                                                                                     name_o
## 1                                                                             Munich
## 2                                                                             Leiden
## 3                                                                          Amsterdam
## 4                                                                          École Po
## 5 École municipale de physique et de chimie industrielles (Municipal School of Industrial Physics and
## 6
##   city_of_university country_of_university born_month age age_get_prize winner
## 1             Munich              Germany        Mar  78            56     56
## 2             Leiden      the Netherlands        Jul  75            49     49
## 3          Amsterdam      the Netherlands        May  78            37     37
## 4              Paris               France        Dec  56            51     51
## 5              Paris               France        May  47            44     44
## 6                                               Nov  67            36     36
```

```r
breaks.for.winner <- c(0,1000000)
nobel.data$winner <- cut(nobel.data$winner, breaks = breaks.for.winner, right = TRUE)
levels(nobel.data$winner) <- ("winner!")
head(nobel.data)
```

```
##          firstname    surname born_country_code died_country_code gender year
## 1 Wilhelm Conrad    Röntgen                DE                DE   male 1901
## 2    Hendrik A.    Lorentz                NL                NL   male 1902
## 3        Pieter     Zeeman                NL                NL   male 1902
## 4         Henri  Becquerel                FR                FR   male 1903
## 5        Pierre      Curie                FR                FR   male 1903
## 6         Marie      Curie                PL                FR female 1903
##    category share
## 1  physics     1
## 2  physics     2
## 3  physics     2
## 4  physics     2
## 5  physics     4
## 6  physics     4
##                                                                                     name_o
## 1                                                                             Munich
## 2                                                                             Leiden
## 3                                                                          Amsterdam
## 4                                                                          École Po
## 5 École municipale de physique et de chimie industrielles (Municipal School of Industrial Physics and
## 6
##   city_of_university country_of_university born_month age age_get_prize  winner
## 1             Munich              Germany        Mar  78            56 winner!
## 2             Leiden      the Netherlands        Jul  75            49 winner!
## 3          Amsterdam      the Netherlands        May  78            37 winner!
## 4              Paris               France        Dec  56            51 winner!
## 5              Paris               France        May  47            44 winner!
## 6                                               Nov  67            36 winner!
```

Graph A: Proportion of Nobel Prizes Earned by Gender

```r
barplot(prop.table(table(nobel.data$gender, nobel.data$winner), 2),
        main = "Proportion of Nobel Prizes Earned by Gender",
        ylab = "proportion",
```

```
    xlim = c(0, 2),
    col = c("lightcoral", "steelblue"),
    legend.text = TRUE)
```

Proportion of Nobel Prizes Earned by Gender

winner!
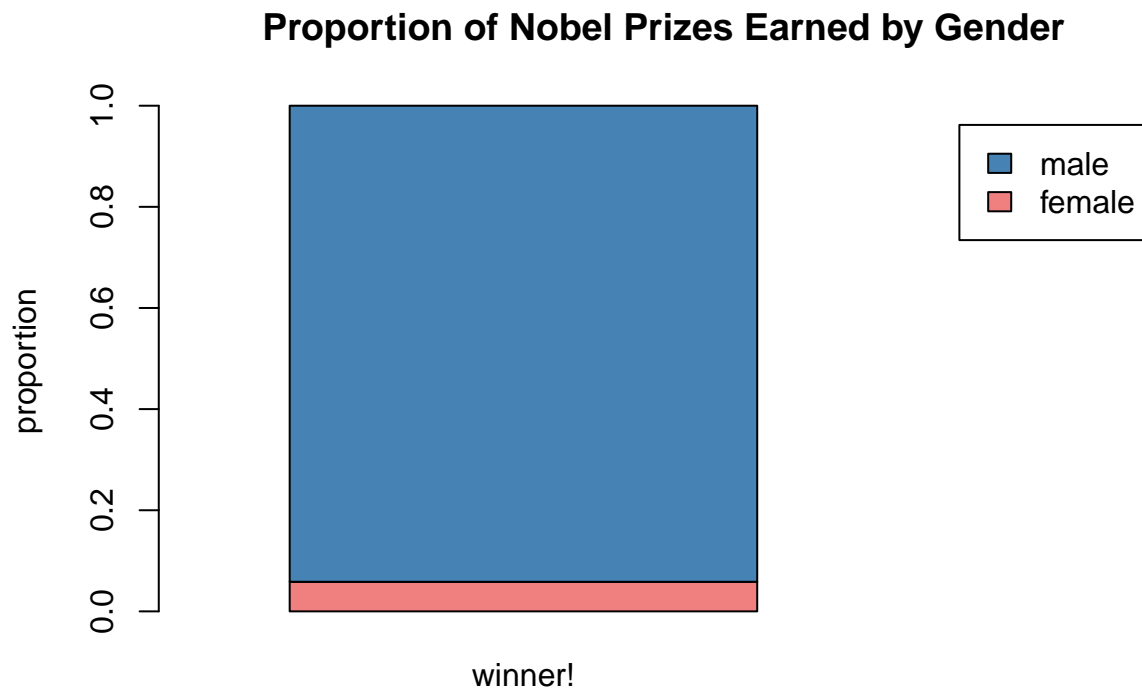
I chose to use a stacked proportion graph for my initial evaluation as it provided the most clear overall breakdown of gender in relation to Nobel Prizes awarded. From the graph, it is evident that there IS a gender divide — an immense one at that. Nearly 95% of Nobel Prizes were won by men, leaving only around 5% to be won by women.

Disappointingly, perhaps, these results did not shock me, although the magnitude of difference did take me by surprise. Most fields where Nobel Prizes are awarded (particularly in STEM) are male-dominated, and most Nobel Prize winners I've previously heard about have been male. Even then, the gender divide shown exceeded my expectations.

Question 2: How does the gender divide translate into different fields of study? Is it more prominent in one than the other?

Bouncing off of my first observation - the immense gap in the number of Prizes won by women and men - I was excited to explore how the gender divide manifested in various areas of study. My prediction was that more women would have won prizes in subjects like peace or literature than in economics or physics.

Graph B: Proportion of Nobel Prizes Earned by Gender by Category

```
#gender different in categories
barplot(prop.table(table(nobel.data$gender, nobel.data$category), 2),
    main = "Proportion of Nobel Prizes Earned by Gender by Category",
    xlab = "category",
    ylab = "proportion",
    xlim = c(0, 9),
    legend.text = TRUE,
    cex.names = 0.6,
    #ylim = c(0, 130),
```

```
        col = c("lightcoral", "steelblue"))
```

Proportion of Nobel Prizes Earned by Gender by Category
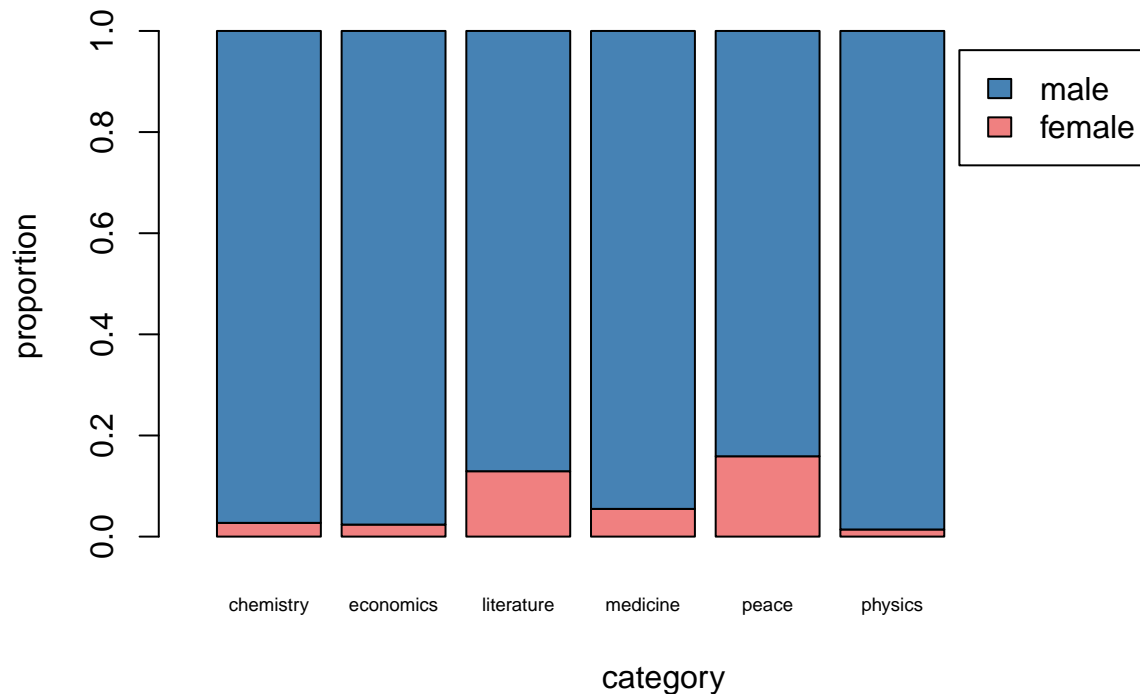
Graph B confirmed my hypothesis. Proportionally, women won the most awards in peace and literature and the least in economics and physics. My hypothesis, however, was based on personal perspective rather than objectivity. For example, economics has a infamous reputation because of its strong connections to Wall Street, wealth, and government, which are often characterized by hyper-masculinity. Physics, also, is well-known to be outstandingly difficult and tedious. Difficult and tedious tasks - those requiring backbone and discipline - are less commonly attributed to women. In comparison, peace and literature are subject areas that tend to be more creative and amicable, and, thus, feminine.

Question 3: Is there an increase in the number of women who win Nobel Prizes over time with the evolution of social norms and women's increased accessibility to education?

My first instinct was to graph the gender breakdown in award winners by year.

Graph C: Number of Nobel Prizes Earned by Gender by Year

```
#gender breakdown by year
barplot((table(nobel.data$gender, nobel.data$year)),
        main = "Number of Nobel Prizes Earned by Gender by Year",
        xlab = "year",
        ylab = "frequency",
        xlim = c(0, 182),
        las = 2,
        legend.text = TRUE,
        col = c("lightcoral", "steelblue"))
```

Number of Nobel Prizes Earned by Gender by Year
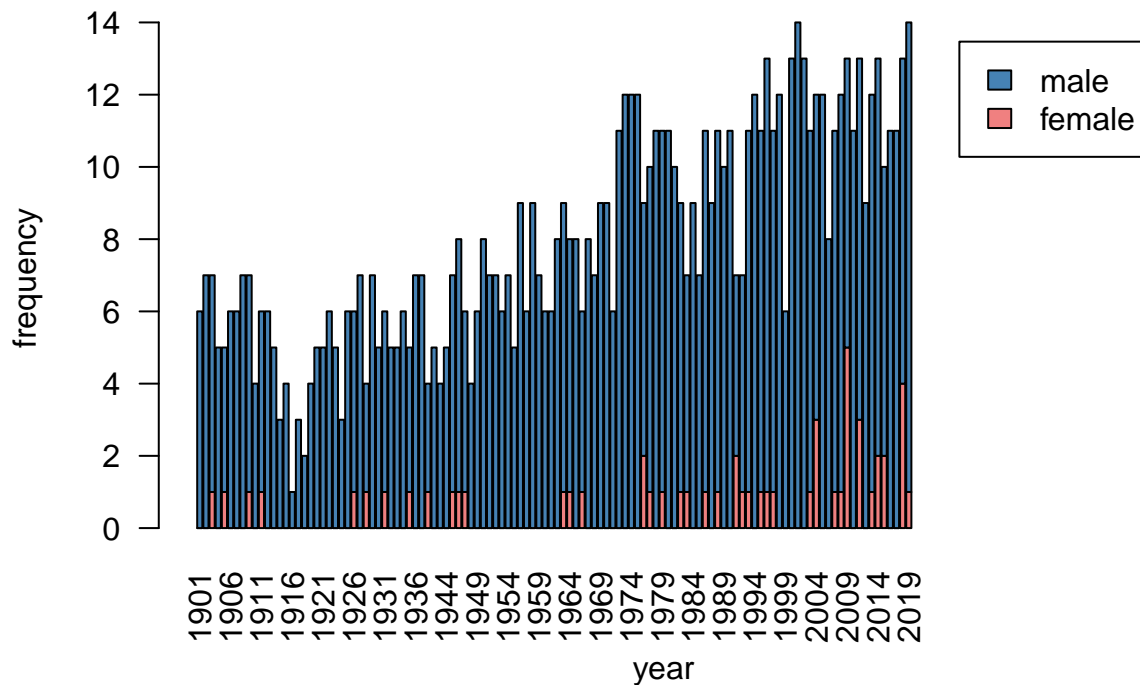
However, I quickly noticed that Graph C appeared immensely crammed, making it hard to observe. To remedy this, I created a new variable that grouped years into decades.

```
#years to decades
breaks.for.nobel.years <- c(0, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010, 2000000
nobel.data$decades <- as.factor(nobel.data$year)
nobel.data$decades <- cut(nobel.data$year, breaks = breaks.for.nobel.years, right = FALSE)
levels(nobel.data$decades) <- c("1910s", "1920s", "1930s", "1940s", "1950s", "1960s", "1970s", "1980s",
head(nobel.data$decades)
```

```
## [1] 1910s 1910s 1910s 1910s 1910s 1910s
## 12 Levels: 1910s 1920s 1930s 1940s 1950s 1960s 1970s 1980s 1990s ... after 2010
```

From there, I created a new graph of the gender breakdown in award winners by decade.

Graph D: Number of Nobel Prizes Earned by Gender by Decade

```
#gender breakdown by decade
barplot(table(nobel.data$gender, nobel.data$decades),
        main = "Number of Nobel Prizes Earned by Gender by Decade",
        xlab = "year",
        ylab = "frequency",
        xlim = c(0, 19),
        ylim = c(0, 130),
        legend.text = TRUE,
        las = 2,
      col = c("lightcoral", "steelblue"))
```

Number of Nobel Prizes Earned by Gender by Decade
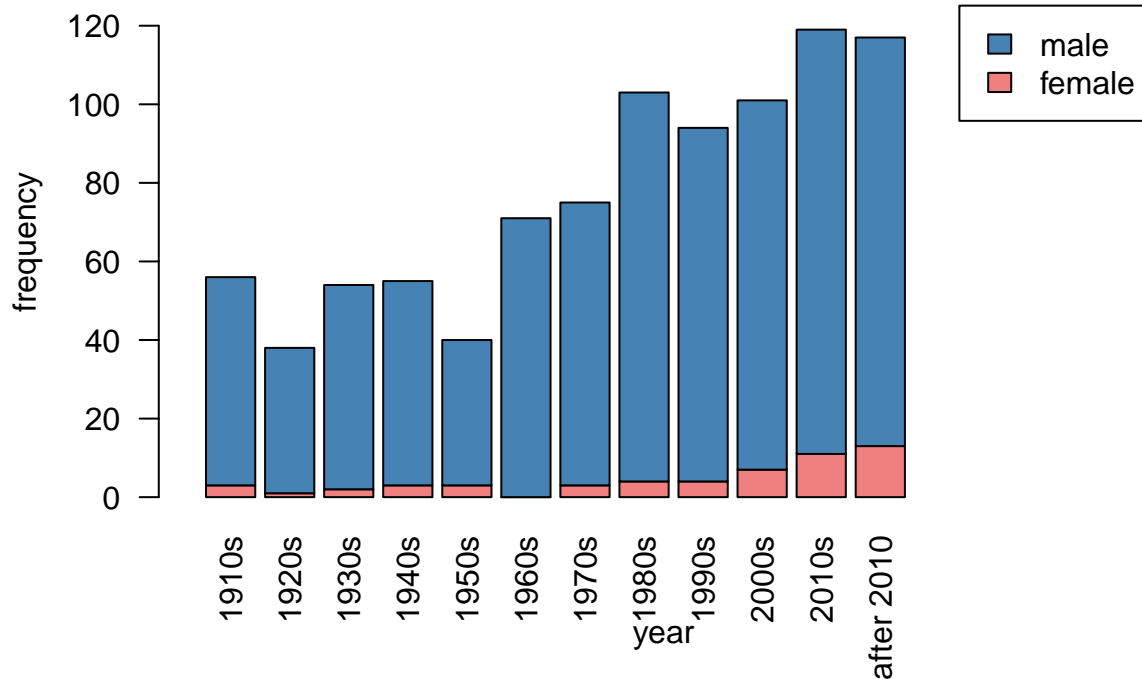
From Graph D, I noticed a upwards trend in the number of Nobel Prizes awarded to women over time and decided to create a graph isolating that relationship, so I could observe more clearly.

To do so, I first created subsets of the different genders.

```
#subsetting for women
nobel.women <- nobel.data[nobel.data$gender == "female",]

#subsetting for men
nobel.men <- nobel.data[nobel.data$gender == "male",]
```

Graph E: Number of Nobel Prizes Earned by Women Over Time

```
barplot(table(nobel.women$decades),
        main = "Number of Nobel Prizes Earned by Women Over Time",
        xlab = "year",
        ylab = "frequency",
        xlim = c(0, 14),
        ylim = c(0, 15),
        cex.names = 0.8,
        las = 2,
        col = c("lightcoral"))
```

Number of Nobel Prizes Earned by Women Over Time
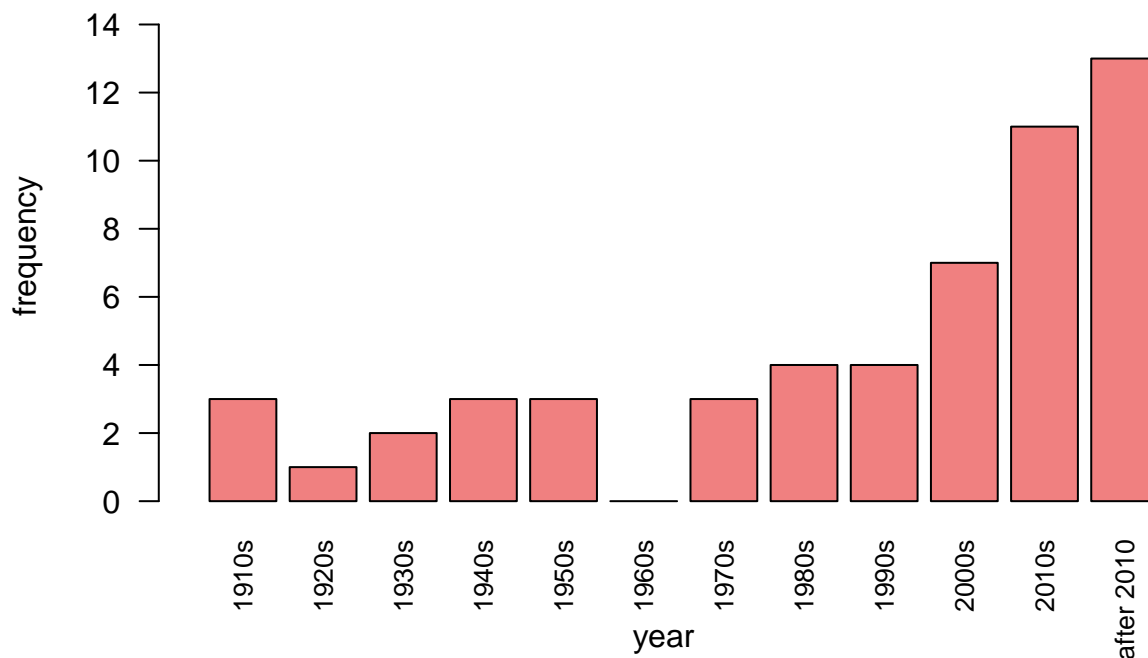
Graph E illustrated a clear and substantial growth in the number of women who were awarded Nobel Prizes over the course of time, aligning with social progressiveness. However, as I reexamined the graph, I noticed that the total frequency of Prizes being awarded each year also increased substantially later on. To compare, I created a graph of the number of men who were awarded Nobel Prizes over the course of time.

Graph F: Number of Nobel Prizes Earned by Men Over Time

```
#men and time
barplot(table(nobel.men$decades),
        main = "Number of Nobel Prizes Earned by Men Over Time",
        xlab = "year",
        ylab = "frequency",
        xlim = c(0, 14),
        ylim = c(0, 120),
        cex.names = 0.8,
        las = 2,
      col = c("steelblue"))
```

Number of Nobel Prizes Earned by Men Over Time
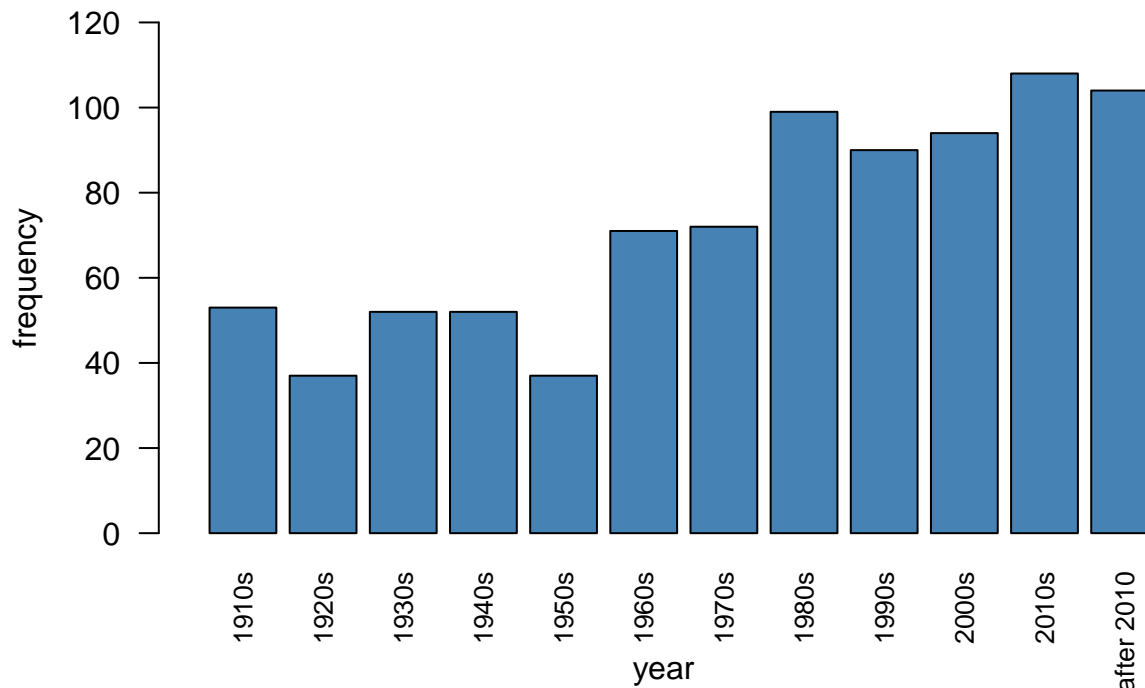
Sure enough, the trend for the number of Prizes being awarded to men also progressed posiively. Both Graphs E and F shared a similar shape: decreasing slightly, increasing up, dropping down, and gradually rising. This supported the idea that although women were, numerically, receiving more awards from the 2000s and onwards, the proportion of men and women receiving Prizes was not drastically different. Graph E is therefore easily misleading.

Wanting to further explore the progression of Prizes won by women over time, I decided to graph the breakdown of gender in different areas of study over time.

To make this possible, I created subsets of the initial data set into "science" and "humanities" categories.

```
#subsetting for different categories
nobel.chemistry <- nobel.data[nobel.data$category == "chemistry",]
nobel.economics <- nobel.data[nobel.data$category == "economics",]
nobel.literature <- nobel.data[nobel.data$category == "literature",]
nobel.medicine <- nobel.data[nobel.data$category == "medicine",]
nobel.peace <- nobel.data[nobel.data$category == "peace",]
nobel.physics <- nobel.data[nobel.data$category == "physics",]

#subsetting for different areas of study
nobel.science <- nobel.data[(nobel.data$category == "chemistry") | (nobel.data$category == "physics") |
nobel.humanities <- nobel.data[(nobel.data$category == "economics") | (nobel.data$category == "literatu
```
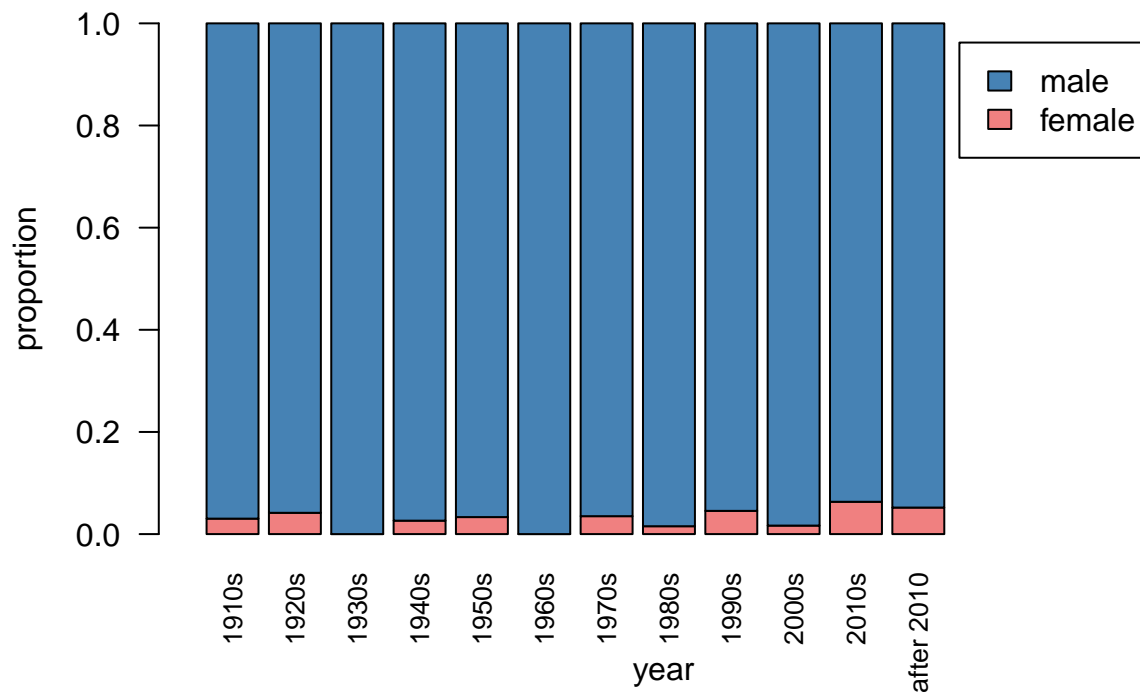
Graph G: Proportion of Nobel Prizes Earned by Gender in Science by Decade

```
#gender differences in science by time (proportions)
barplot(prop.table(table(nobel.science$gender, nobel.science$decades), 2),
        main = "Proportion of Nobel Prizes Earned by Gender in Science by Decade",
        xlab = "year",
        ylab = "proportion",
```

```
      xlim = c(0, 18),
      ylim = c(0, 1),
      legend.text = TRUE,
      cex.names = 0.8,
      las = 2,
    col = c("lightcoral", "steelblue"))
```

Proportion of Nobel Prizes Earned by Gender in Science by Decade

Graph H: Proportion of Nobel Prizes Earned by Gender in Humanities by Decade

```
#gender differences in humanities by time frequency (proportions)
barplot(prop.table(table(nobel.humanities$gender, nobel.humanities$decades), 2),
      main = "Proportion of Nobel Prizes Earned by Gender in Humanities by Decade",
      xlab = "year",
      ylab = "proportion",
      xlim = c(0, 18),
      ylim = c(0, 1),
      legend.text = TRUE,
      cex.names = 0.8,
      las = 2,
    col = c("lightcoral", "steelblue"))
```

**Proportion of Nobel Prizes Earned by Gender in Humanities by Decade**
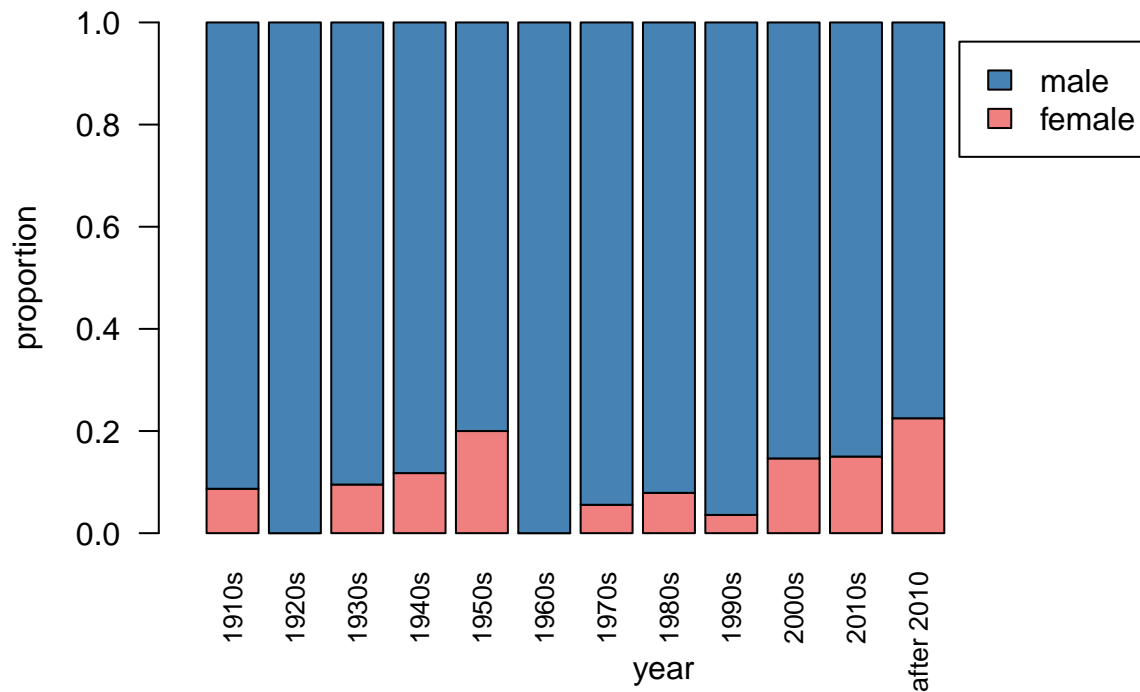
Taking into account all of my graphs, these were the most initially confusing to me. I expected to see a clear spike in Prizes awarded to women after 2010. Instead, there was a reoccuring pattern in both the graphs for humanities and science. Over the course of time, there were slight fluctuations in Prizes awarded to women - minimal for science and more clear for humanities - but there was an absence of an easily identifiable upwards trend. In fact, the number of humanities Prizes won by women in the 1950s and after 2010 seemed nearly identical.

Understanding these findings, I created my final graph.

Graph I: Proportion of Nobel Prizes Earned by Gender by Decade

```
barplot(prop.table(table(nobel.data$gender, nobel.data$decades), 2),
        main = "Proportion of Nobel Prizes Earned by Gender by Decade",
        xlab = "year",
        ylab = "proportion",
        xlim = c(0, 18),
        #ylim = c(0, 130),
        legend.text = TRUE,
        las = 2,
      col = c("lightcoral", "steelblue"))
```

Proportion of Nobel Prizes Earned by Gender by Decade
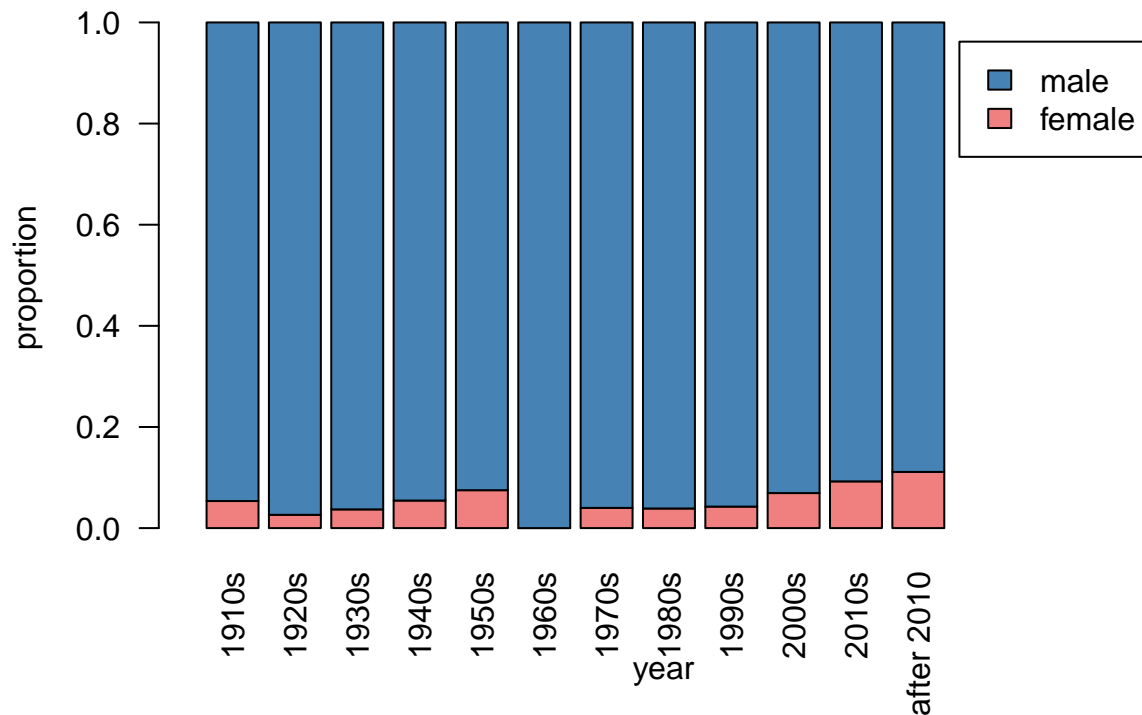
Once again, I noticed the patterns I recognized in Graphs H and I (specifically, I). There is undoubtedly an increased number of women winning Prizes after 2010; however, the progression of the gender proportions of award winners over time does not suggest a strong relationship between time and increased number of female winners. Similar to Graph I, the proportion of Prizes earned by women in the 1950s is comparable to that in the 2010s or later.

Conclusion

There is a massive gender divide with Nobel Prize recipients, with far too less female representation. The gap is most notable in categories like economics, but it is noticeably present in every area. When observed in comparison to time, the number of Prizes awarded to women increases by the early 2000s. But, the number of Prizes awarded to men also steadily increases because the number of total awards presented gradually grew through the late 1900s and into the 2000s. Thus, there is no majorly distinguishable relationship between time and the proportion of Prizes awarded to women.