Lab 3: Barplots for Film Dialogue Analysis

Elijah Russell

9/25/25

Load data and libraries

I've included some code below to load, restructure, and clean the data we'll use. This chunk also includes the chunk option `echo=FALSE` which hides it in your final submission. *Don't edit the code in this chunk.*

0. Update the YAML header

Make appropriate updates to the author and date fields at the very top of your .Rmd file.

1. Print out the `lotr` dataset

The `lotr` dataset tracks **the number of words spoken by characters** in the three films that make up the Lord of the Rings (LOTR) trilogy: 1. The Fellowship of the Ring, 2. The Two Towers, 3. The Return of the King.

```
lotr
```

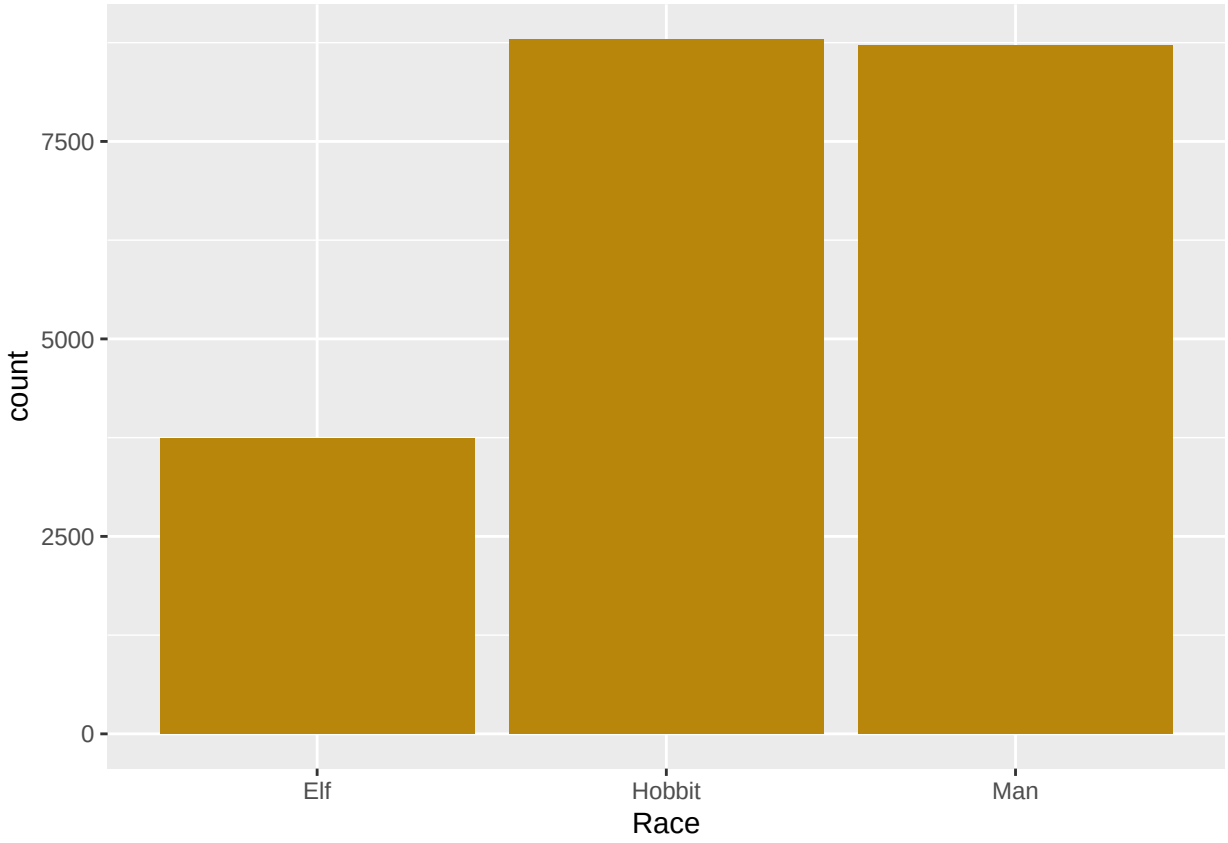
```
## # A tibble: 18 x 4
##    Film                      Race   Gender Words
##    <fct>                     <chr>  <chr>  <dbl>
##  1 The Fellowship Of The Ring Elf    Female  1229
##  2 The Fellowship Of The Ring Elf    Male     971
##  3 The Fellowship Of The Ring Hobbit Female    14
##  4 The Fellowship Of The Ring Hobbit Male    3644
##  5 The Fellowship Of The Ring Man    Female     0
##  6 The Fellowship Of The Ring Man    Male    1995
##  7 The Two Towers             Elf    Female   331
##  8 The Two Towers             Elf    Male     513
##  9 The Two Towers             Hobbit Female     0
## 10 The Two Towers             Hobbit Male    2463
## 11 The Two Towers             Man    Female   401
## 12 The Two Towers             Man    Male    3589
## 13 The Return Of The King     Elf    Female   183
## 14 The Return Of The King     Elf    Male     510
## 15 The Return Of The King     Hobbit Female     2
## 16 The Return Of The King     Hobbit Male    2673
## 17 The Return Of The King     Man    Female   268
## 18 The Return Of The King     Man    Male    2459
```

2. Words spoken by race

Make a bar plot of the total number of words spoken by each race across all three movies (i.e., summing up across all three movies). *No stacked barplots allowed.* Look carefully at the data before beginning.

Which of the two functions for making bar plots, `geom_bar` or `geom_col`, is most appropriate here? Why?
"geom_col" because "geom_col" is clearer when you have an x and a y, "geom_bar" is better for when you just have x and count of x.

```
ggplot(data = lotr,aes(x=Race, weight=Words))+
  geom_bar(fill = "darkgoldenrod")
```

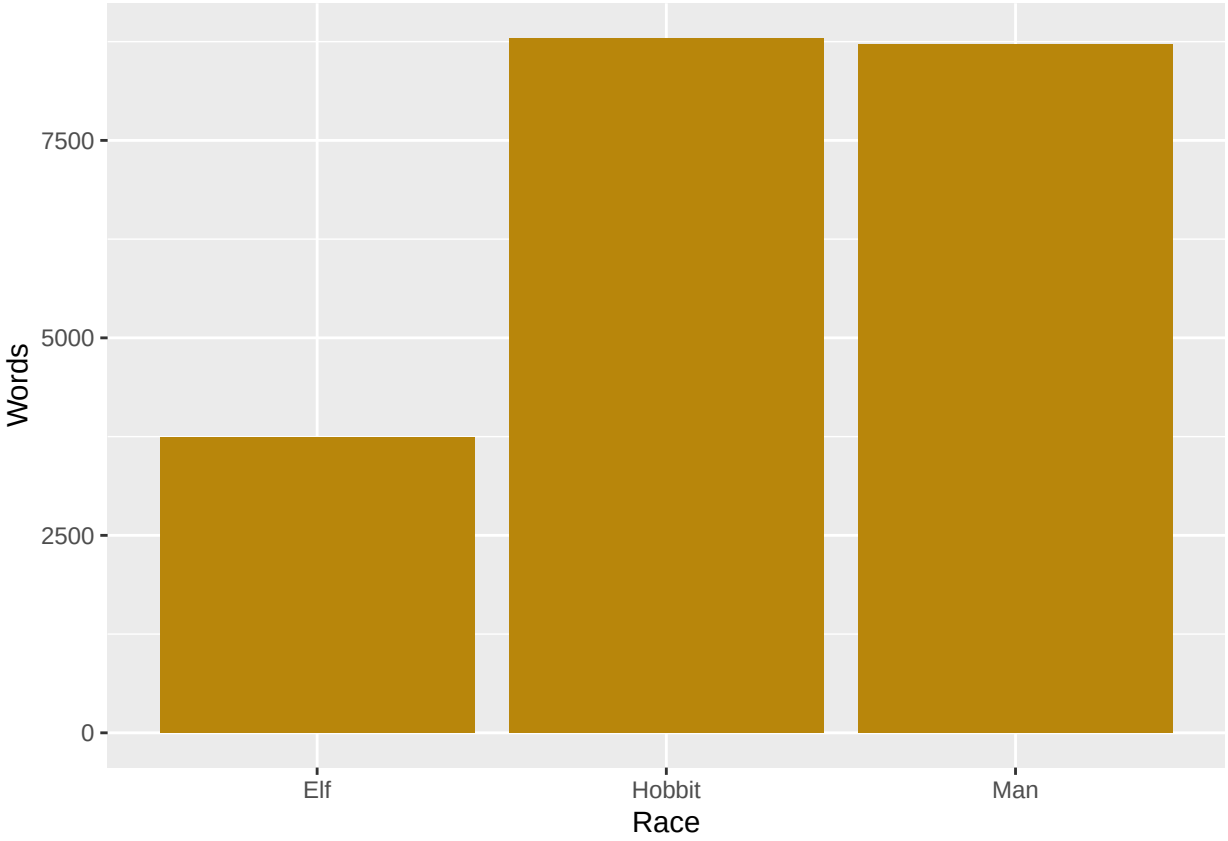


3. Words spoken by race again

Remake your plot using the other `geom_` function for making bar plots.

Does one race's speech dominate the films?

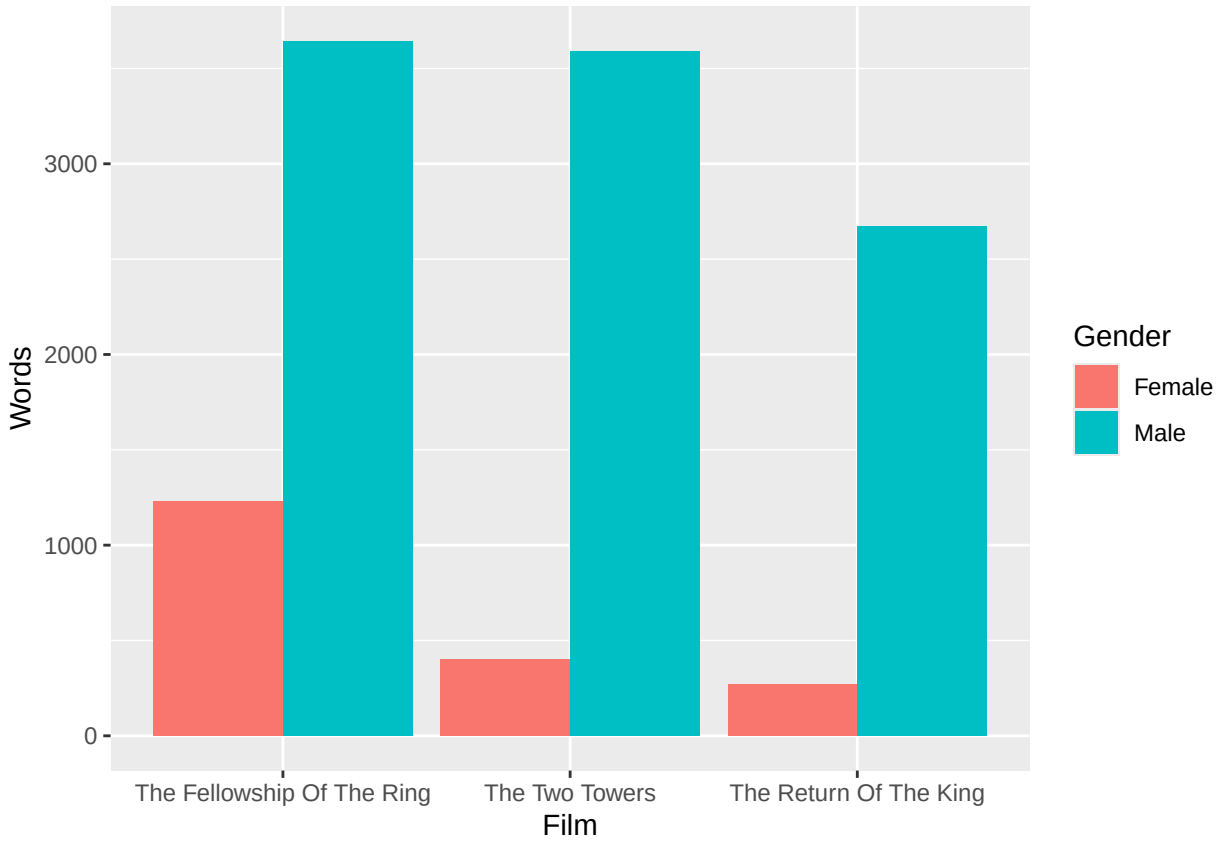
```
ggplot(data = lotr,aes(x=Race, y=Words))+
  geom_col(fill = "darkgoldenrod")
```



4. Gender and film

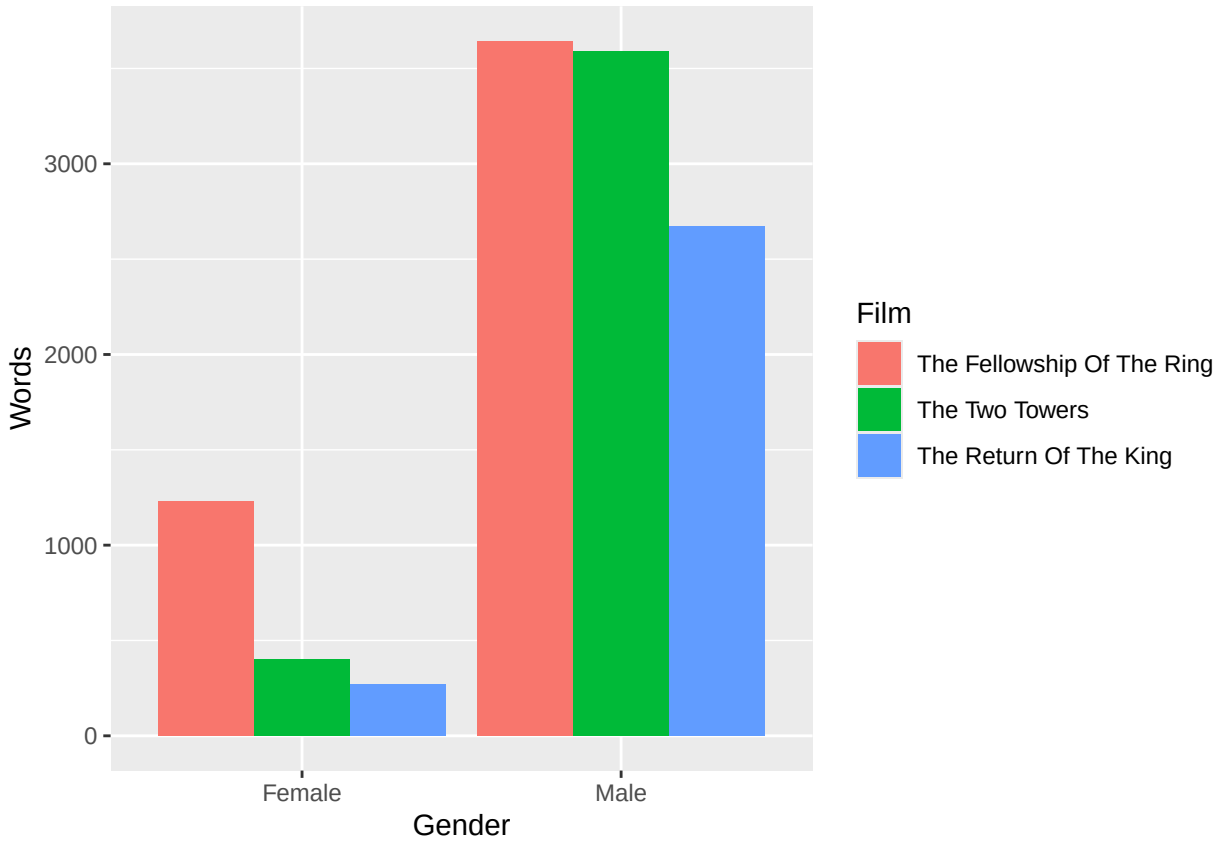
Does a certain gender dominate a movie? (lolz of course it does, but still, make a bar plot to look at this). *No stacked barplots allowed.* In fact, make four versions of this plot: - filling by Gender

```
ggplot(data = lotr,aes(x=Film, y=Words, fill=Gender))+
  geom_col(position = "dodge")
```



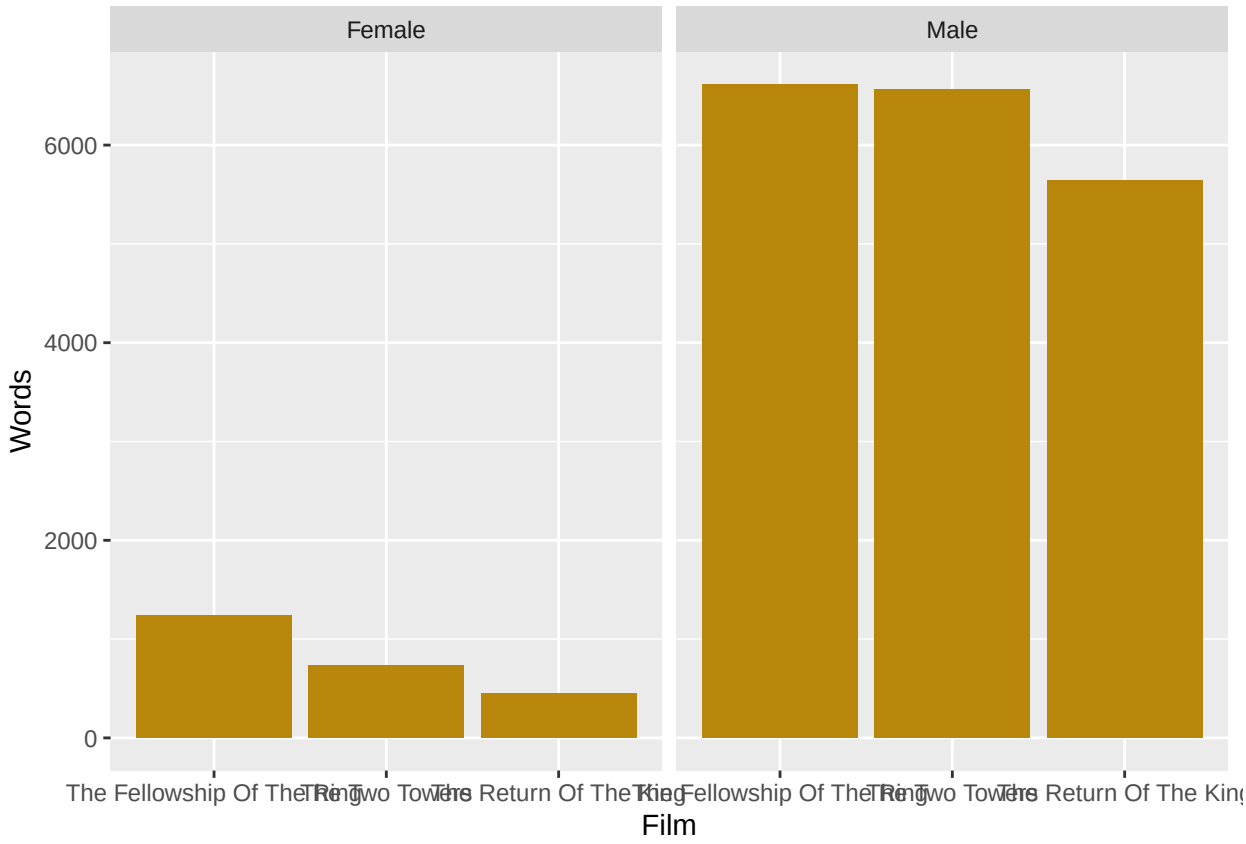
- filling by Film

```
ggplot(data = lotr,aes(x=Gender, y=Words, fill=Film))+
  geom_col(position = "dodge")
```



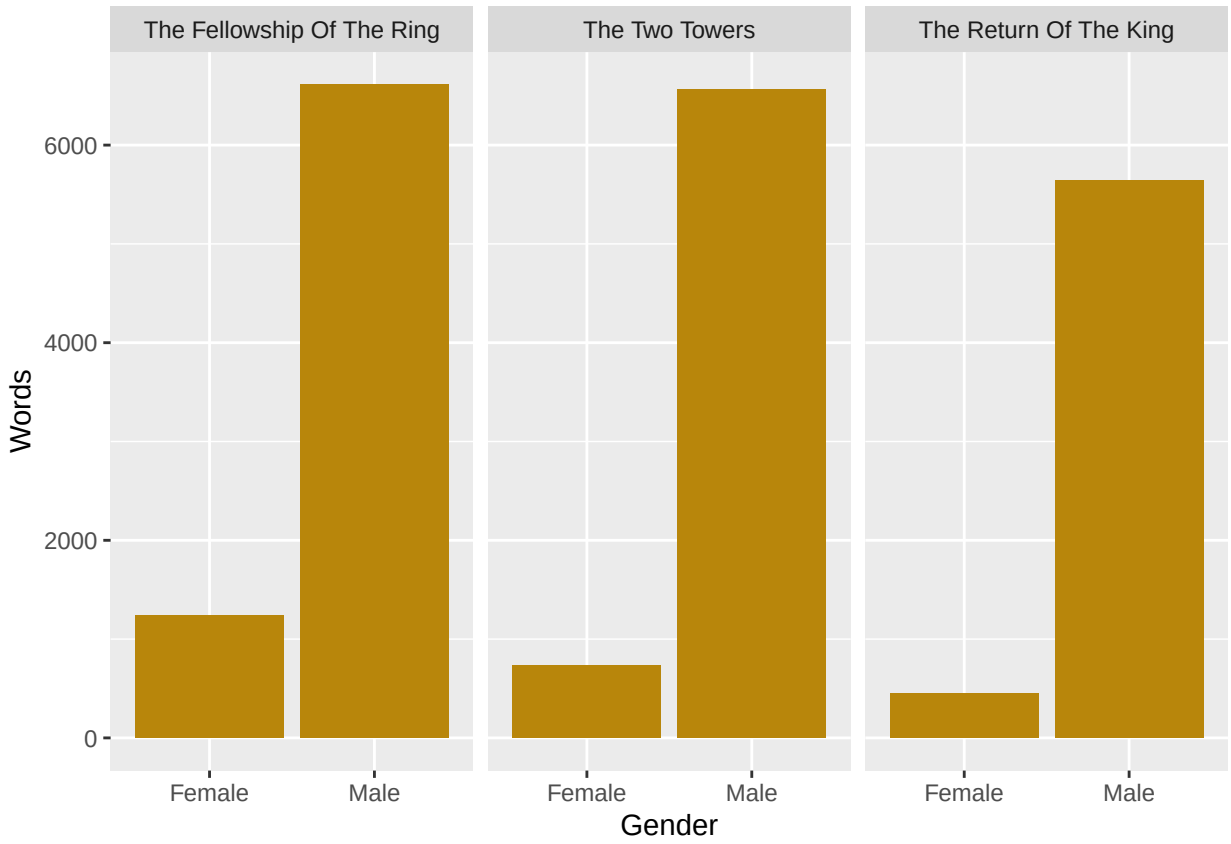
- faceting by Gender

```
ggplot(data = lotr,aes(x=Film, y=Words))+
  geom_col(fill = "darkgoldenrod") +
  facet_wrap(~ Gender)
```



- faceting by Film

```
ggplot(data = lotr,aes(x=Gender, y=Words))+
  geom_col(fill = "darkgoldenrod") +
  facet_wrap(~ Film)
```

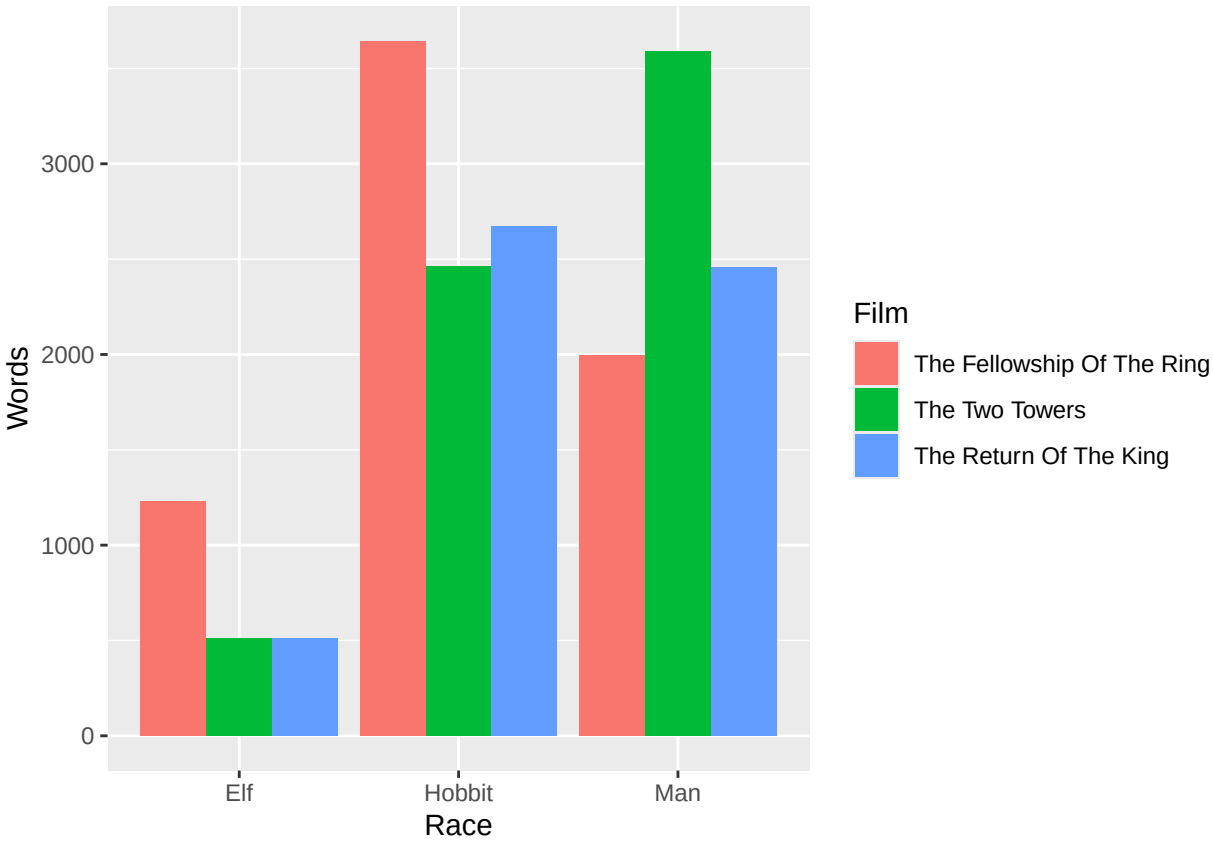


Which bar plot do you prefer? Briefly explain why. I prefer the plot filled by film, because you see all the data on one graph and can easily compare dialogue amounts by film and by gender.

5. Race and film

Make a bar plot of the number of words spoken by each race in each of the three movies. Repeat your experiments from above (changing fill or facet by either variable) and include only your favorite plot.

```
ggplot(data = lotr,aes(x=Race, y=Words, fill=Film))+
  geom_col(position = "dodge")
```



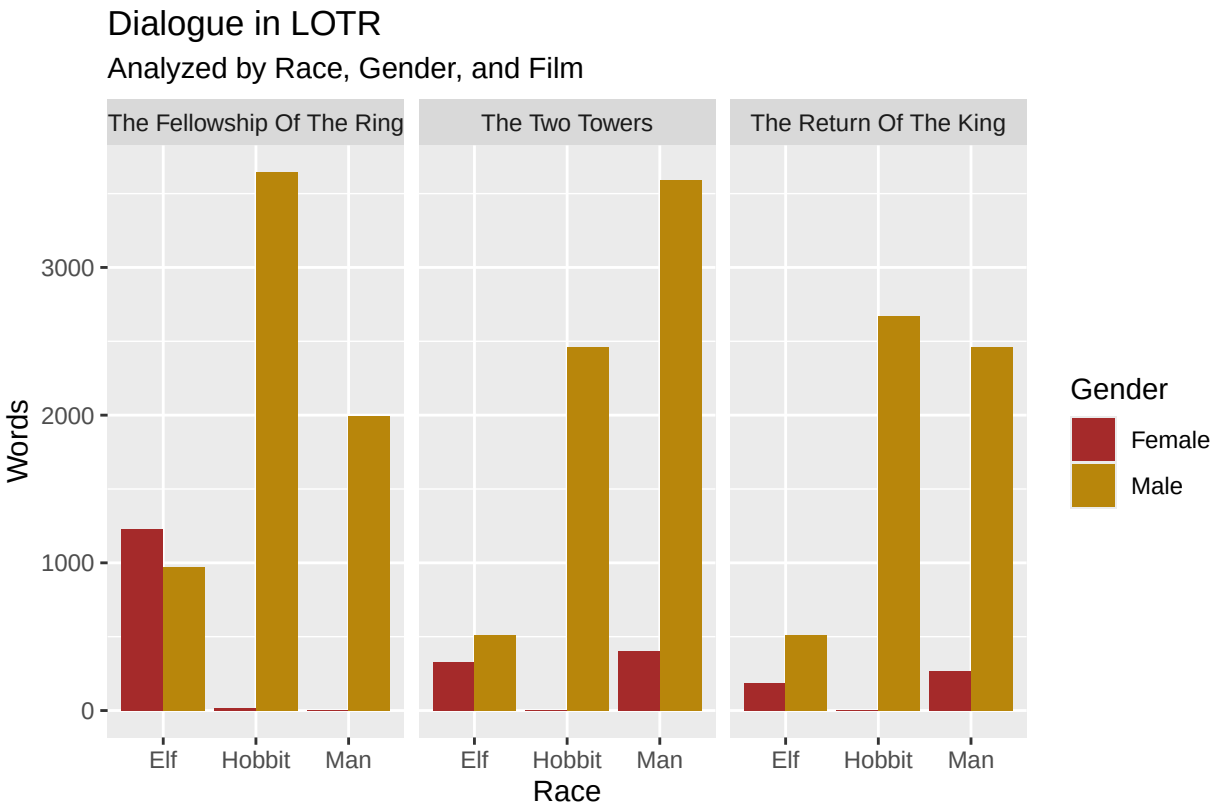
Does the dominant race differ across the three movies? Yes, the first movie is primarily hobbit dialogue, although elves play a larger role than in the latter two. The second movie is primarily human dialogue, and the third is an almost even split between hobbits and men.

6. Race and gender and film

You need to show `Race`, `Gender`, and `Film` at the same time, but you only have two possible aesthetics (`x` and `fill`), so you'll also need to facet by the third. Play around with different combinations (e.g. try `x = Race`, then `x = Film`) until you find one that tells the clearest story.

Make your plot look as nice as possible by adding a title, subtitle, caption and any other changes you find helpful.

```
ggplot(data = lotr,aes(x=Race, y=Words, fill=Gender))+
  geom_col(position = "dodge") +
  facet_wrap(~ Film) +
  scale_fill_manual( values = c("Male" = "darkgoldenrod", "Female" = "brown")) +
  labs(title= "Dialogue in LOTR", subtitle="Analyzed by Race, Gender, and Film", caption= "Words spoken
```



Words spoken by Lord of the Rings characters across the main trilogy

List three takeaways (e.g., patterns, trends, comparisons) about race and gender in the LOTR films 1. The dialogue is male dominated 2. The dialogue is hobbit and human dominated 3. The Elvish dialogue is split almost evenly between male and female

7. More films

Outside of the Lord of the Rings trilogy, you may have seem some coverage of film dialogue analyses (and visualizations) by gender (and other attributes) in popular media. Some analyses are much more detailed than the ones we've looked at here. For example, the Bechdel test tracks whether (or more generally, how often) two women in a film or text have a conversation about something other than a man.

More similar to our look at counts of spoken words, The Pudding analyzed over 2,000 screenplays in 2016 and created some interesting data visualizations. Scroll through the article and notice how the visualizations at the top of the page change as you go. Slide your cursor along the blue-red bar that describes the in the "Only High-Grossing Films: Ranked in the Top 2,500 by US Box Office"

Find one blue-ish and one red-ish film that you've watched and report on the percentage of words spoken by female characters, according to their analysis. A blue-ish film is The Jungle Book, and a red-ish film is Inside Out.

8. Recreate The Pudding's bar plot

Look at the The Pudding's plots labelled "Screenplay Dialogue, Broken-down by Gender". We can recreate this in ggplot for the LOTR films.

The code below creates *a new dataset* that sums up the words spoken by each gender in each film (across races). Use this to create a vertical (e.g., rotated 90 degrees) version of the stacked bar plots as in The Pudding. Then, try adding `coord_flip()` and `geom_hline( yintercept = .5, lty=2)` to your plot to finalize your reproduction.

```
lotr_across_race <- lotr %>%
  group_by( Film, Gender) %>%
  summarize( Words = sum(Words) ) %>%
  mutate( Words_prop = Words / sum(Words) )
```

```
## `summarise()` has grouped output by 'Film'. You can override using the
## `.groups` argument.
```

```
lotr_across_race
```

```
## # A tibble: 6 x 4
## # Groups:   Film [3]
##   Film                      Gender Words Words_prop
##   <fct>                     <chr>  <dbl>      <dbl>
## 1 The Fellowship Of The Ring Female  1243     0.158
## 2 The Fellowship Of The Ring Male    6610     0.842
## 3 The Two Towers             Female   732     0.100
## 4 The Two Towers             Male    6565     0.900
## 5 The Return Of The King     Female   453     0.0743
## 6 The Return Of The King     Male    5642     0.926
```

```
ggplot(data = lotr_across_race,aes(x=Film, y=Words_prop, fill=Gender))+
  geom_col() +
  coord_flip() +
  geom_hline( yintercept = .5, lty=2) +
  labs(y="Words")
```