

Tidyverse Assignment - Masculinity Data

Amber Ferger

11/28/2019

The Data: What Do Men Think It Means to be a Man?

For this assignment, I decided to use the dataset that corresponds to the *What Do Men Think It Means To Be A Man?* article on FiveThirtyEight.com. The article can be found here: <https://fivethirtyeight.com/features/what-do-men-think-it-means-to-be-a-man/>

It contains the results of a survey of 1,615 adult men conducted by SurveyMonkey in partnership with FiveThirtyEight and WNYC Studios from May 10-22, 2018.

```
library(tidyverse)
dat <- as_tibble(read.csv('https://raw.githubusercontent.com/amberferger/DATA607_Masculinity/master/raw/
```

We have quite a bit of questions in this survey, so we will focus on just a few. For the purpose of this vignette, let's see what role demographics play in the answer to the question **How important is it to you that others see you as masculine?** We'll use the `select` command (from the tidyverse dependency *dplyr*) to return only the columns we are interested in looking at (race and orientation). We'll also use the `filter()` command to subset our data to only individuals that provided a response to these question.

```
dat <- dat %>%
  select(race2, orientation, q0002) %>%
  filter(q0002 != 'No answer' & race2 != 'No answer' & orientation != 'No answer')
```

Data Aggregation

Our final data set has 1 response variable (the answer to the question) and 2 explanatory variables (our demographic data). We'll use the `group_by` function with the `count()` function to summarize our data. We will then transform our values by creating a percent for each of the of the groupings.

```
raceCount <- dat %>%
  group_by(race2, q0002) %>%
  count()

raceCount <- raceCount %>%
  group_by(race2) %>%
  mutate(RACE_PCT = n/sum(n))

raceCount
```

```
## # A tibble: 8 x 4
## # Groups:   race2 [2]
##   race2      q0002          n RACE_PCT
##   <fct>    <fct>    <int>   <dbl>
## 1 Non-white Not at all important    46   0.178
## 2 Non-white Not too important      68   0.264
## 3 Non-white Somewhat important    99   0.384
```

```
## 4 Non-white Very important      45    0.174
## 5 White      Not at all important 193    0.144
## 6 White      Not too important   471    0.353
## 7 White      Somewhat important  523    0.391
## 8 White      Very important      149    0.112
```

We'll do the same thing for the **orientation** variable.

```
orientationCount <- dat %>%
  group_by(orientation, q0002) %>%
  count()

orientationCount <- orientationCount %>%
  group_by(orientation) %>%
  mutate(ORIENTATION_PCT = n/sum(n))

orientationCount
```

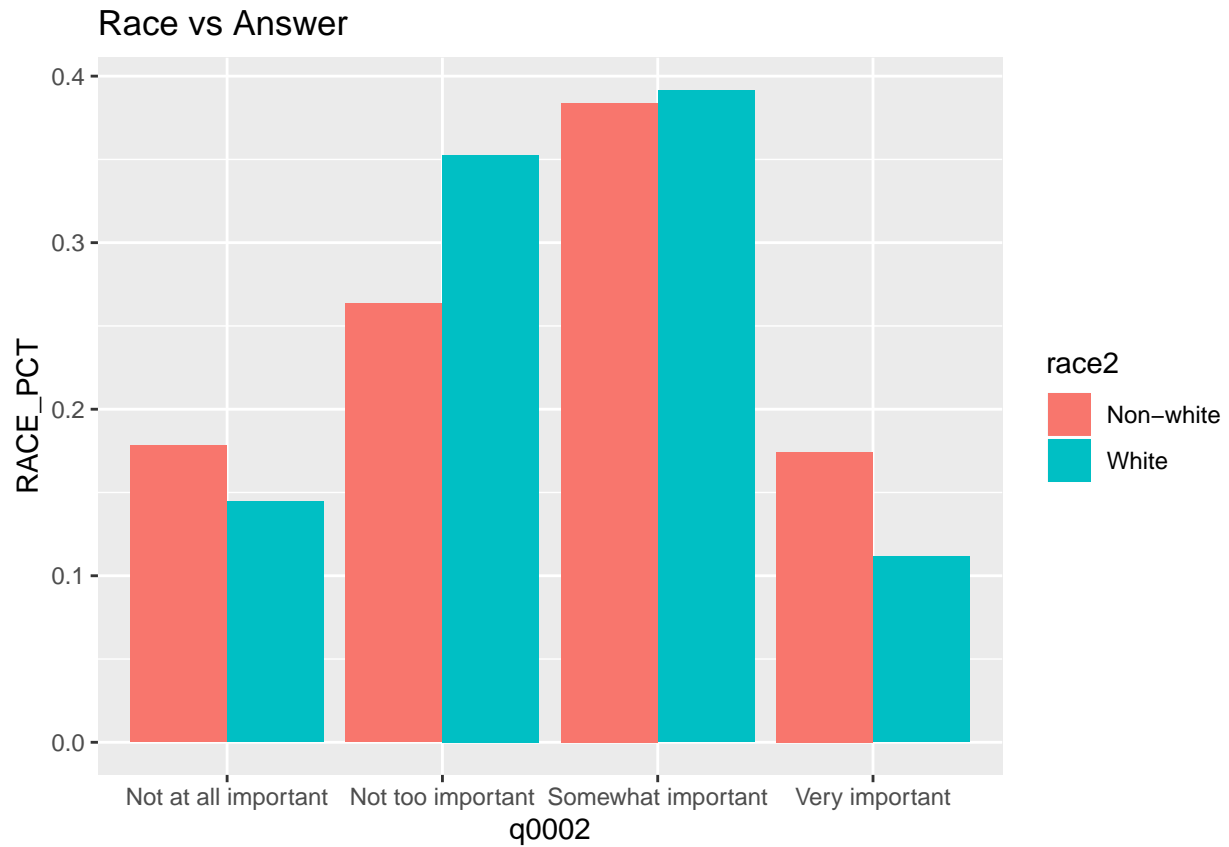
```
## # A tibble: 12 x 4
## # Groups:   orientation [3]
##   orientation q0002          n ORIENTATION_PCT
##   <fct>       <fct>      <int>         <dbl>
## 1 Gay/Bisexual Not at all important    33         0.206
## 2 Gay/Bisexual Not too important     58         0.362
## 3 Gay/Bisexual Somewhat important    54         0.338
## 4 Gay/Bisexual Very important     15         0.0938
## 5 Other        Not at all important    10         0.323
## 6 Other        Not too important      8         0.258
## 7 Other        Somewhat important      5         0.161
## 8 Other        Very important         8         0.258
## 9 Straight     Not at all important   196         0.140
## 10 Straight     Not too important    473         0.337
## 11 Straight     Somewhat important   563         0.401
## 12 Straight     Very important     171         0.122
```

Visualization

Now let's visualize our data! We'll use the **ggplot** library to take a look:

```
library(ggplot2)

ggplot(raceCount, aes(fill=race2, y=RACE_PCT, x=q0002)) +
  geom_bar(position="dodge", stat="identity") +
  ggtitle("Race vs Answer")
```



```
ggplot(orientationCount, aes(fill=orientation, y=ORIENTATION_PCT, x=q0002)) +  
  geom_bar(position="dodge", stat="identity") +  
  ggtitle("Orientation vs Answer")
```

