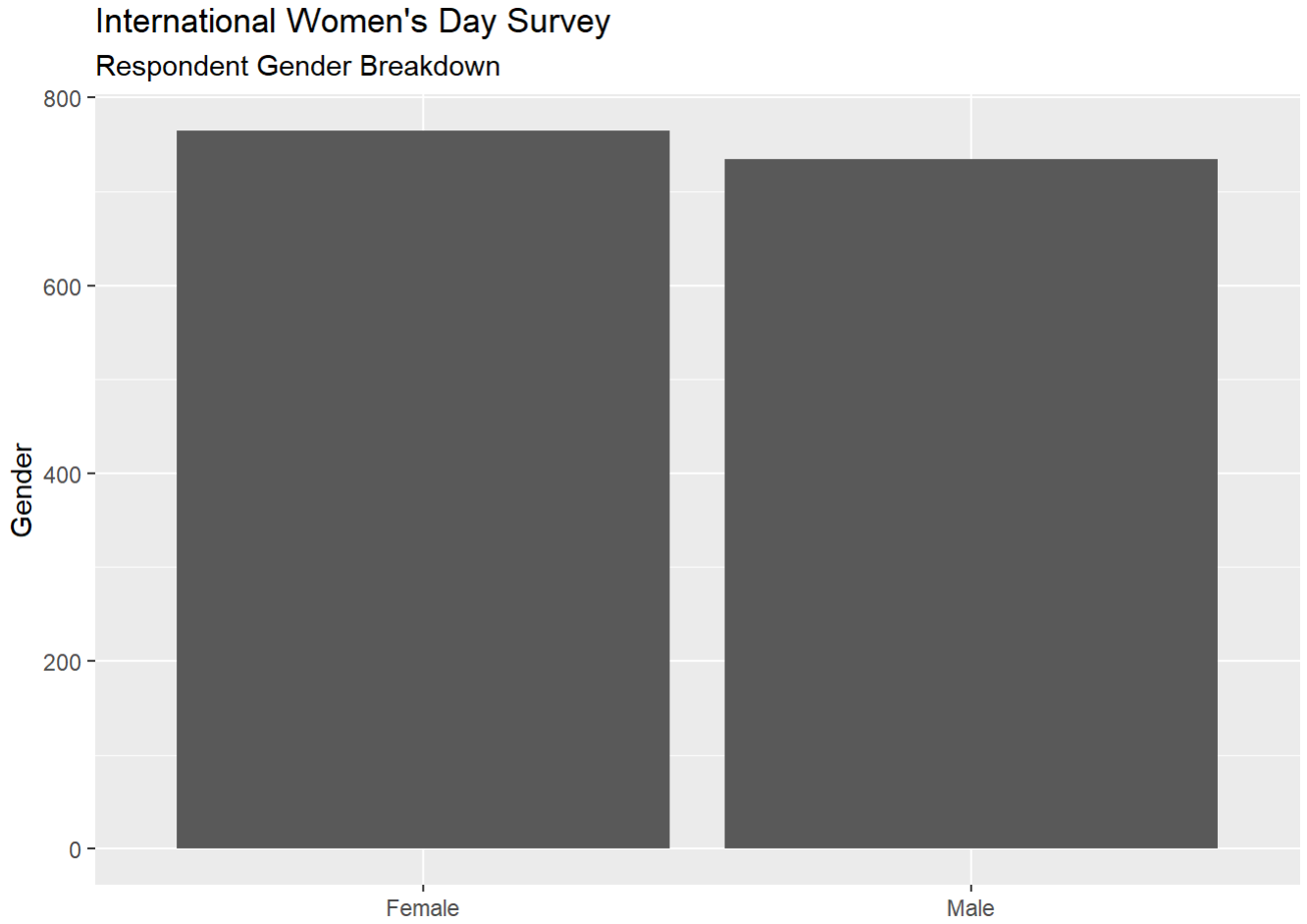
Homework Week 3

Renee Zale

3/24/2021

R Markdown

The following data comes from a survey conducted by C Space, a market research firm. In honor of International Women's Day, this survey evaluates how both male and female respondents evaluate their employers and overall satisfaction in the workplace. There are 1500 total respondents who completed this survey.



The mean age in this dataset is 41.282, and the median age is 40. The age and gender variables did not require any recoding or wrangling.

This survey asks respondents to evaluate a series of statements on a five point Likert scale from "Strongly Agree" to "Strongly Disagree." Unfortunately, rather than create a column for each statement with a row for each respondent, this dataset has split the 5 point Likert scale for each statement into 5 columns. For example, the Likert scale responses are spread across five columns for the the statement "At work, I feel respected" (Statement 1). In the table below, each respondent has their answer stored in one of six potential columns.

```
select(womensday, starts_with("Q2_S1")) %>%
  print(n = 10, width = Inf)
```

```
## # A tibble: 1,500 x 12
##    Q2_S1_R1       Q2_S1_R2 Q2_S1_R3                  Q2_S1_R4 Q2_S1_R5
##    <chr>          <chr>    <chr>                     <chr>    <chr>
##  1 <NA>           <NA>     Neither Agree nor Disagree <NA>     <NA>
##  2 <NA>           Agree    <NA>                      <NA>     <NA>
##  3 <NA>           Agree    <NA>                      <NA>     <NA>
##  4 <NA>           <NA>     <NA>                      Disagree <NA>
##  5 <NA>           <NA>     <NA>                      Disagree <NA>
##  6 <NA>           Agree    <NA>                      <NA>     <NA>
##  7 Strongly Agree <NA>     <NA>                      <NA>     <NA>
##  8 <NA>           <NA>     <NA>                      <NA>     Strongly Disagree
##  9 Strongly Agree <NA>     <NA>                      <NA>     <NA>
## 10 <NA>           <NA>     <NA>                      Disagree <NA>
##    Q2_S1_R6_Skip Q2_S10_R1      Q2_S10_R2 Q2_S10_R3                  Q2_S10_R4
##    <chr>         <chr>          <chr>     <chr>                      <chr>
##  1 <NA>          <NA>           Agree     <NA>                       <NA>
##  2 <NA>          <NA>           <NA>      Neither Agree nor Disagree <NA>
##  3 <NA>          <NA>           <NA>      Neither Agree nor Disagree <NA>
##  4 <NA>          <NA>           <NA>      <NA>                       Disagree
##  5 <NA>          <NA>           Agree     <NA>                       <NA>
##  6 <NA>          <NA>           <NA>      Neither Agree nor Disagree <NA>
##  7 <NA>          Strongly Agree <NA>      <NA>                       <NA>
##  8 <NA>          <NA>           <NA>      Neither Agree nor Disagree <NA>
##  9 <NA>          Strongly Agree <NA>      <NA>                       <NA>
## 10 <NA>          <NA>           <NA>      <NA>                       Disagree
##    Q2_S10_R5 Q2_S10_R6_Skip
##    <chr>     <chr>
##  1 <NA>      <NA>
##  2 <NA>      <NA>
##  3 <NA>      <NA>
##  4 <NA>      <NA>
##  5 <NA>      <NA>
##  6 <NA>      <NA>
##  7 <NA>      <NA>
##  8 <NA>      <NA>
##  9 <NA>      <NA>
## 10 <NA>      <NA>
## # ... with 1,490 more rows
```

I will attempt to merge these columns together so that we can analyze “I feel respected at work” cut by gender.

```
unite(womensday, "Statement2", 21:26, remove = TRUE, na.rm = FALSE) %>%
  select("Statement2")
```

```
## # A tibble: 1,500 x 1
##    Statement2
##    <chr>
##  1 NA_NA_Neither Agree nor Disagree_NA_NA_NA
##  2 NA_Agree_NA_NA_NA_NA
##  3 NA_Agree_NA_NA_NA_NA
##  4 NA_NA_NA_Disagree_NA_NA
##  5 NA_NA_NA_Disagree_NA_NA
##  6 NA_Agree_NA_NA_NA_NA
##  7 Strongly Agree_NA_NA_NA_NA_NA
##  8 NA_NA_NA_NA_Strongly Disagree_NA
##  9 Strongly Agree_NA_NA_NA_NA_NA
## 10 NA_NA_NA_Disagree_NA_NA
## # ... with 1,490 more rows
```

This has combined the text of each of the six columns, but leaves us with many “NA” strings in each cell. Instead I will try recoding the variables as numbers from 1-5, and then summing the columns into one new column.

```
womensday<-mutate(womensday, NumQ2_S1_R1 = recode(Q2_S1_R1, `Strongly Agree` = "5"))
table(select(womensday, NumQ2_S1_R1))
```

```
##
##   5
## 359
```

That worked, so I will do this same process for the rest of the Statement 1 answer options.

```
womensday<-mutate(womensday, NumQ2_S1_R1 = recode(Q2_S1_R1, `Strongly Agree` = "6"))
womensday<-mutate(womensday, NumQ2_S1_R2 = recode(Q2_S1_R2, `Agree` = "5"))
womensday<-mutate(womensday, NumQ2_S1_R3 = recode(Q2_S1_R3, `Neither Agree nor Disagree` = "4"))
womensday<-mutate(womensday, NumQ2_S1_R4 = recode(Q2_S1_R4, `Disagree` = "3"))
womensday<-mutate(womensday, NumQ2_S1_R5 = recode(Q2_S1_R5, `Strongly Disagree` = "2"))
womensday<-mutate(womensday, NumQ2_S1_R6_Skip = recode(Q2_S1_R6_Skip, `Skipped` = "1"))
select(womensday, "NumQ2_S1_R1", "NumQ2_S1_R2", "NumQ2_S1_R3", "NumQ2_S1_R4", "NumQ2_S1_R5", "Nu
mQ2_S1_R6_Skip")
```

```
## # A tibble: 1,500 x 6
##    NumQ2_S1_R1 NumQ2_S1_R2 NumQ2_S1_R3 NumQ2_S1_R4 NumQ2_S1_R5 NumQ2_S1_R6_Skip
##    <chr>       <chr>       <chr>       <chr>       <chr>       <chr>
##  1 <NA>        <NA>        4           <NA>        <NA>        <NA>
##  2 <NA>        5           <NA>        <NA>        <NA>        <NA>
##  3 <NA>        5           <NA>        <NA>        <NA>        <NA>
##  4 <NA>        <NA>        <NA>        3           <NA>        <NA>
##  5 <NA>        <NA>        <NA>        3           <NA>        <NA>
##  6 <NA>        5           <NA>        <NA>        <NA>        <NA>
##  7 6           <NA>        <NA>        <NA>        <NA>        <NA>
##  8 <NA>        <NA>        <NA>        <NA>        2           <NA>
##  9 6           <NA>        <NA>        <NA>        <NA>        <NA>
## 10 <NA>        <NA>        <NA>        3           <NA>        <NA>
## # ... with 1,490 more rows
```

Now I need to convert the characters into numeric format and replace the NAs with 0s. (This took me a few attempts before finding the right solution!).

```
womensday$NumQ2_S1_R1 <- womensday$NumQ2_S1_R1 %>%
  as.numeric() %>%
  replace_na(0)
womensday$NumQ2_S1_R2 <- womensday$NumQ2_S1_R2 %>%
  as.numeric() %>%
  replace_na(0)
womensday$NumQ2_S1_R3 <- womensday$NumQ2_S1_R3 %>%
  as.numeric() %>%
  replace_na(0)
womensday$NumQ2_S1_R4 <- womensday$NumQ2_S1_R4 %>%
  as.numeric() %>%
  replace_na(0)
womensday$NumQ2_S1_R5 <- womensday$NumQ2_S1_R5 %>%
  as.numeric() %>%
  replace_na(0)
womensday$NumQ2_S1_R6_Skip <- womensday$NumQ2_S1_R6_Skip %>%
  as.numeric() %>%
  replace_na(0)
select(womensday, "NumQ2_S1_R1", "NumQ2_S1_R2", "NumQ2_S1_R3", "NumQ2_S1_R4", "NumQ2_S1_R5", "Nu
mQ2_S1_R6_Skip")
```

```
## # A tibble: 1,500 x 6
##    NumQ2_S1_R1 NumQ2_S1_R2 NumQ2_S1_R3 NumQ2_S1_R4 NumQ2_S1_R5 NumQ2_S1_R6_Skip
##          <dbl>       <dbl>       <dbl>       <dbl>       <dbl>             <dbl>
##  1           0           0           4           0           0                 0
##  2           0           5           0           0           0                 0
##  3           0           5           0           0           0                 0
##  4           0           0           0           3           0                 0
##  5           0           0           0           3           0                 0
##  6           0           5           0           0           0                 0
##  7           6           0           0           0           0                 0
##  8           0           0           0           0           2                 0
##  9           6           0           0           0           0                 0
## 10           0           0           0           3           0                 0
## # ... with 1,490 more rows
```

Now I will sum up the 6 columns into a new column to consolidate for analysis.

```
womensday$Q2_S1_Likert <- rowSums(womensday[c("NumQ2_S1_R1", "NumQ2_S1_R2", "NumQ2_S1_R3", "NumQ
2_S1_R4", "NumQ2_S1_R5", "NumQ2_S1_R6_Skip")])
select(womensday, "Q2_S1_Likert")
```

```
## # A tibble: 1,500 x 1
##    Q2_S1_Likert
##           <dbl>
##  1            4
##  2            5
##  3            5
##  4            3
##  5            3
##  6            5
##  7            6
##  8            2
##  9            6
## 10            3
## # ... with 1,490 more rows
```

After a lot of manipulation I was able to consolidate the 6 columns into one Likert Scale column for statement 1. **However, this was a very time consuming process and there are 49 total statements in the dataset that this would need to be done for to do a complete analysis. I would love any tips for how to make this a quicker process, or potentially complete the process for all 49 statements at once.**

Now I'd like to compare how men and women respond to the question of whether they feel respected at work (statement 1).

```
womensday %>%
  select(Q2_S1_Likert) %>%
  table() %>%
  prop.table()*100
```

```
## .
##         1         2         3         4         5         6
##  1.066667  5.666667  8.000000 19.733333 41.600000 23.933333
```

Looking at the overall results, nearly 24% strongly agree they feel respected at work, and 41% agree.

```
ggplot(womensday, aes(Gender, Q2_S1_Likert)) + geom_violin()
```

This plot doesn't make it easy to see the differences between the two groups. My next steps for analysis would be to convert this variable into a categorical variable and then experiment with different types of visualizations.