

Attitudes Towards Globalization Based on Gender

Aiden Dickson

04/26/2021

***Now with additional GGPlot Infographs appended towards end for HW4 - building on HW3**

Introduction: This analysis of attitudes regarding globalization, based on gender, is taken from a sample of 1008 survey participants, 18 years of age or older, living in the United States, (304 respondents were interviewed on a land-line telephone, and 704 were interviewed on a mobile phone, including 469 who had no landline telephone). The survey was conducted under the direction of SSRS. It is a study of overall attitudes towards globalization with a larger scope of demographic features such as religion, race and political persuasion.

Importing and Setting Up Libraries:

```
if (!require('tidyverse')) install.packages('tidyverse')

## Loading required package: tidyverse

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.0      v dplyr   1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

if (!require('haven')) install.packages('haven')

## Loading required package: haven

if (!require('dplyr')) install.packages('dplyr')
if (!require('knitr')) install.packages('knitr')

## Loading required package: knitr
```

```
if (!require('tinytex')) install.packages('tinytex')
```

```
## Loading required package: tinytex
```

```
if (!require('ggplot2')) install.packages('ggplot2')
```

```
library(tidyverse)
library(haven)
library(dplyr)
library(knitr)
library(tinytex)
library(ggplot2)
```

```
knitr::opts_chunk$set(echo = TRUE)
```

Importing Dataset from Google Drive and Viewing Dimensions:

```
us_data<-read_sav("/Volumes/GoogleDrive/My Drive/DataSets/United_States_April_2020_Topline_WEB_FINAL.sav")
dim(us_data)
```

```
## [1] 1008 25
```

Creating Table of Analytic Interest with Two Columns: Sex (Sex of respondent): '1' for 'Male', '2' for 'Female', and Q4 (Response to 'Overall has globalization in the past few years been good for the United States?'): '1' = 'Good', '2' = 'Bad', '3' = 'Both Good and Bad', and '9' = 'DK/Refused to Q4'. (Additionally, viewing new dimensions at top, still including 1008 rows, but only printing out top 10 here):

```
tibble_Q4Ans_eaGender <-select(us_data, sex, Q4)
dim(tibble_Q4Ans_eaGender)
```

```
## [1] 1008 2
```

```
print(tibble_Q4Ans_eaGender, n = 10)
```

```
## # A tibble: 1,008 x 2
##       sex      Q4
##   <dbl> <lbl> <dbl> <lbl>
## 1 2 [Female] 2 [Bad]
## 2 1 [Male] 1 [Good]
## 3 1 [Male] 1 [Good]
## 4 2 [Female] 1 [Good]
## 5 1 [Male] 2 [Bad]
```

```
## 6 2 [Female] 2 [Bad]
## 7 2 [Female] 1 [Good]
## 8 2 [Female] 1 [Good]
## 9 1 [Male] 2 [Bad]
## 10 1 [Male] 1 [Good]
## # ... with 998 more rows
```

Mutate for new 'Gender' Column in larger dataset to Show Labels 'Male' or 'Female' instead of numerical code:

```
us_data<-us_data %>%
  mutate(Gender = case_when(
    sex == 1 ~ 'Male',
    sex == 2 ~ 'Female'))

select (us_data, sex, Gender)
```

```
## # A tibble: 1,008 x 2
##       sex Gender
##   <dbl> <lbl> <chr>
## 1 2 [Female] Female
## 2 1 [Male] Male
## 3 1 [Male] Male
## 4 2 [Female] Female
## 5 1 [Male] Male
## 6 2 [Female] Female
## 7 2 [Female] Female
## 8 2 [Female] Female
## 9 1 [Male] Male
## 10 1 [Male] Male
## # ... with 998 more rows
```

Mutate for new 'Answers' Column in larger dataset to Show Labels 'Bad','Good', 'Both Good and Bad' and 'DK/Refused' in place of numerical code:

```
us_data <- us_data %>%
  mutate(Answers = case_when(
    Q4 == 1 ~ 'Good',
    Q4 == 2 ~ 'Bad',
    Q4 == 3 ~ 'Both good and bad',

  ))

select (us_data, Q4, Answers)
```

```
## # A tibble: 1,008 x 2
##       Q4 Answers
##   <dbl> <lbl> <chr>
## 1 2 [Bad] Bad
## 2 1 [Good] Good
```

```
## 3 1 [Good] Good
## 4 1 [Good] Good
## 5 2 [Bad] Bad
## 6 2 [Bad] Bad
## 7 1 [Good] Good
## 8 1 [Good] Good
## 9 2 [Bad] Bad
## 10 1 [Good] Good
## # ... with 998 more rows
```

Tibble for Labeled Answers per Labeled Gender. *Only printing top ten here:

```
Answers_Per_Gendr_Tibble<- select (us_data, Gender, Answers)
print(Answers_Per_Gendr_Tibble, n=10)
```

```
## # A tibble: 1,008 x 2
##   Gender Answers
##   <chr> <chr>
## 1 Female Bad
## 2 Male Good
## 3 Male Good
## 4 Female Good
## 5 Male Bad
## 6 Female Bad
## 7 Female Good
## 8 Female Good
## 9 Male Bad
## 10 Male Good
## # ... with 998 more rows
```

One Proportional Table for Each Variable, each converted to percent. Then a Crosstab-type table for both, converted to proportional and finally percent:

```
Perc_Gendr_Tibble<- Answers_Per_Gendr_Tibble %>%
  select(Gender) %>%
  table() %>%
  prop.table()*100
```

```
Perc_Ans_Tibble<- Answers_Per_Gendr_Tibble %>%
  select(Answers) %>%
  table() %>%
  prop.table()*100
```

```
print(Perc_Gendr_Tibble)
```

```
## .
##   Female    Male
## 48.61111 51.38889
```

```
print(Perc_Ans_Tibble)
```

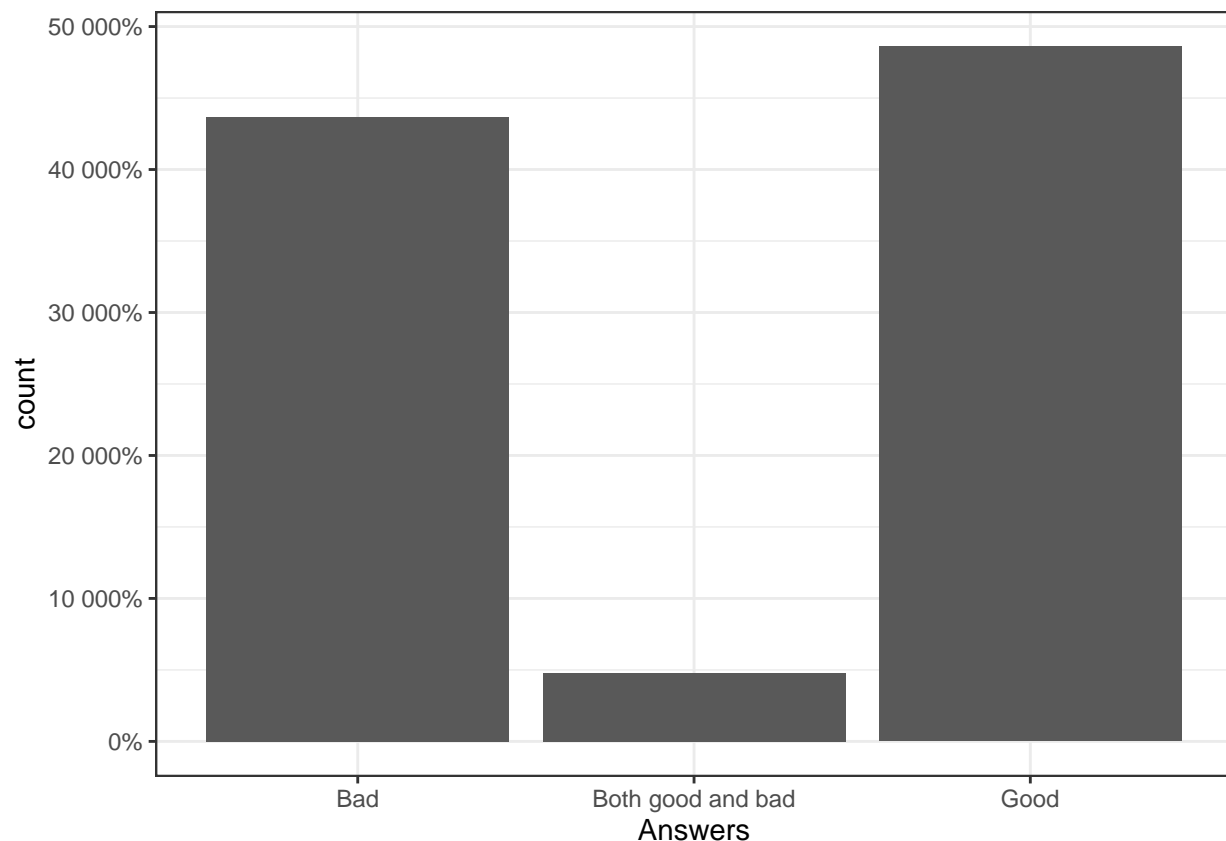
```
## .
##           Bad Both good and bad           Good
##      45.005149      4.943357      50.051493
```

```
perc_tibble <-prop.table(xtabs(~ Gender + Answers, us_data))*100
print(perc_tibble)
```

```
##           Answers
## Gender           Bad Both good and bad           Good
## Female 22.760041      2.471679 22.760041
## Male   22.245108      2.471679 27.291452
```

[Discovered I needed to change the 'Mutate' function above to not include 'NA' explicitly assigned, to simply not account for it in 'Mutate', so it will be omitted]. In the bar graph below is the distribution of answers without discerning gender.

```
us_data %>%
  filter(! is.na(Answers)) %>%
  ggplot(aes(Answers)) + geom_bar() +
  theme_bw() + scale_y_continuous(labels = scales:: percent_format(accuracy=1))
```



```
labs(title = "Attitudes Regarding if Globalization in US in the Past Few Years Has Been 'Good'", y =

## $y
## [1] "Magnitude"
##
## $x
## [1] "Answers"
##
## $title
## [1] "Attitudes Regarding if Globalization in US in the Past Few Years Has Been 'Good'"
##
## attr(,"class")
## [1] "labels"

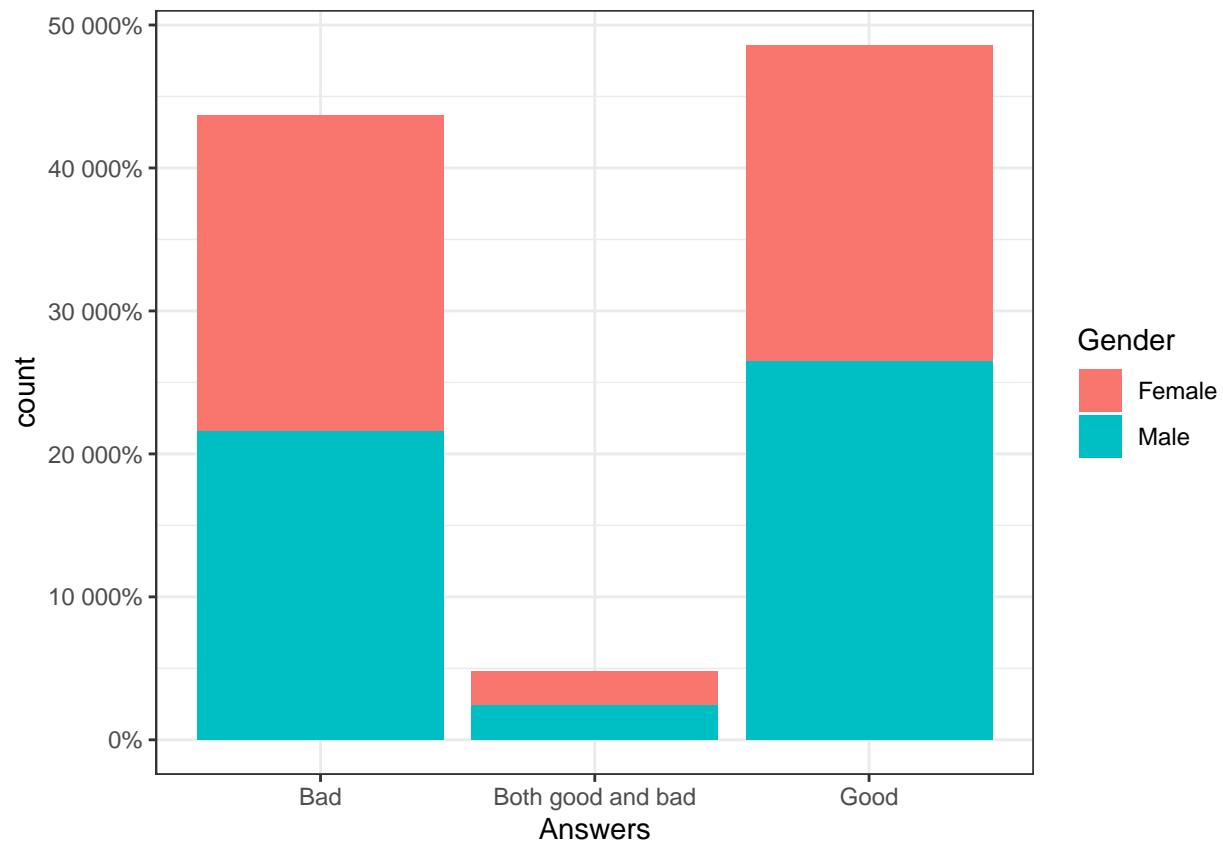
# Is not displaying specified y-axis label
```

1. What the Visualization demonstrates:

We can observe that there is an even split between the answers of ‘Good’ or ‘Bad’ to the question ‘Has globalization in the past few years been good for the United States?’ within female participants (Females: ‘Good’=22.03% and ‘Bad’=22.03%), unlike the the clearly uneven split among male participants (Males: ‘Good’= 26.42% and ‘Bad’= 21.54%)

It can also be noted, more female participants answered ‘Bad’ in response to the question of globalization value for the US in past few years than male participants. 22.03% female participants answered ‘Bad’ vs 21.53% of male participants responding ‘Bad’. Additionally, more male participants responded to the question, of globalization in past few years as being good for the US (Males: ‘Good’=26.42, Females: ‘Good’=22.03). Any difference of value in these 4 comparisons are so slight that they are most likely not statistically significant, but before making this conclusion with confidence, this significance will be measured in a later analysis.

```
us_data %>%
  filter(! is.na(Answers)) %>%
  ggplot(aes(Answers, fill = Gender)) + geom_bar() +
  theme_bw() + scale_y_continuous(labels = scales:: percent_format(accuracy=1))
```



```
labs(title = "Attitudes Regarding if Globalization in US in the Past Few Years Has Been 'Good'", y =
```

```
## $y
## [1] "Percent"
##
## $x
## [1] "Answers"
##
## $title
## [1] "Attitudes Regarding if Globalization in US in the Past Few Years Has Been 'Good'"
##
## attr("class")
## [1] "labels"
```

```
# Is not displaying specified y-axis label
```

2. I chose the Bar Chart / Histogram, due to the way it lends well to a comparison of numeric values [a percentage of summation of survey responses depending on category, based on closed-ended answers] between discrete nominal variables [Bad, Both Good and Bad, Good]. The choice of color coding of male and female responses also lended well, basically overlapping 2 histograms with the potential for more categories depending on the demographic variable. At first I thought of doing the grouping type histogram, one grouping for males and a second grouping for females which would result in only two histograms side by side, so I decided this would not lend itself as well to comparison as the bi-chromatic overlap-type histogram.

3. What I wished if anything I could have executed but found limited capacity to do: I attempted to do a ‘gather’ operation so graph displays percentages along the y axis, using perc_tibble but it didn’t work due to the r not able to treat ‘xtabs object’ like a tibble. Figured out how to do this within the ggplot code and added ‘+ labs(title = "Attitudes Regarding if Globalization in US in the Past Few Years above’. Precision of sig figs 3 zeroes past decimal is a bit much but have yet to find how to pare this.

```
#Gathered_Perc_Tibble <- gather(perc_tibble, key = "Answer", value = "Percent", -c("Gender"))  
#print(Gathered_Perc_Tibble)
```

Next in Homework 5: Will Attempt to clean and tidy data by doing “Gather” on Questions from us_data, so can group by answers in the resultant infograph and then differentiate the answers to each question based on Race within the grouping.