Homework 4

Reyna Orellana

5/08/2021

R Markdown for Homework 4

Homework 4 Instructions:Your goal is to build on the previous assignments, and create a data visualizati
  1. What the visualization demonstrates.
2. Why you chose the visualization approach you selected, and what alternative approaches you considered but decided not to pursue.
3. What you wished, if anything, you could have executed but found limited capability to do.

```
#Add gss data
library(haven)
GSS2018 <- read_dta("GSS2018.dta")
#dataset and creating new variables in the next chunk
```

```
## # A tibble: 1,559 x 1
##         race
##    <dbl+lbl>
##  1 1 [white]
##  2 1 [white]
##  3 1 [white]
##  4 1 [white]
##  5 1 [white]
##  6 1 [white]
##  7 1 [white]
##  8 3 [other]
##  9 1 [white]
## 10 2 [black]
## # ... with 1,549 more rows
```

```
##
##    1    2    3
## 1120  252  187
```

```
## # A tibble: 6 x 1
##        race
##   <dbl+lbl>
## 1 1 [white]
## 2 1 [white]
## 3 1 [white]
## 4 1 [white]
## 5 1 [white]
## 6 1 [white]
```

```
##     race_3cat
## race Black Other White
##    1     0     0  1120
```

```
##    2   252     0     0
##    3     0   187     0
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   35.00   50.00   50.12   64.00   89.00       7
```

```
##
## 18-29 30-39 40-49   50+
##   235   275   255   787
```

```
##         age
## age_4cat 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
##    18-29 13 16  6 16 18 14 23 24 20 28 28 29  0  0  0  0  0  0  0  0  0  0  0
##    30-39  0  0  0  0  0  0  0  0  0  0  0  0 33 20 29 27 40 25 21 25 22 33  0
##    40-49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 28
##    50+    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##         age
## age_4cat 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63
##    18-29  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    30-39  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    40-49 27 27 26 24 23 30 23 24 23  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    50+    0  0  0  0  0  0  0  0  0 24 27 22 38 24 32 33 30 28 27 28 28 16 27
##         age
## age_4cat 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86
##    18-29  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    30-39  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    40-49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    50+   25 27 28 17 22 20 33 27 15 18 18 17 13 22  7  8 11  8 10  6  5  8 10
##         age
## age_4cat 87 88 89
##    18-29  0  0  0
##    30-39  0  0  0
##    40-49  0  0  0
##    50+    2  5 21
```

```
## Warning in mean.default(race_3cat, na.rm = TRUE): argument is not numeric or
## logical: returning NA
```

```
## Warning in mean.default(age_4cat, na.rm = TRUE): argument is not numeric or
## logical: returning NA
```

```
## # A tibble: 1 x 38
##   abany abdefect abfelegl abhelp1 abhelp2 abhelp3 abhelp4 abhlth abinspay
##   <dbl>    <dbl>    <dbl>   <dbl>   <dbl>   <dbl>   <dbl>  <dbl>    <dbl>
## 1  1.50     1.23     2.07    1.28    1.70    1.43    1.12   1.09     1.50
## # ... with 29 more variables: abmedgov1 <dbl>, abmedgov2 <dbl>, abmelegl <dbl>,
## #   abmoral <dbl>, abnomore <dbl>, abpoor <dbl>, abpoorw <dbl>, abrape <dbl>,
## #   absingle <dbl>, abstate1 <dbl>, age <dbl>, ballot <dbl>, bible <dbl>,
## #   born <dbl>, class <dbl>, cohort <dbl>, coldeg1 <dbl>, degree <dbl>,
## #   depress <dbl>, occ10 <dbl>, partyid <dbl>, race <dbl>, sex <dbl>,
## #   hispanic <dbl>, maeduc <dbl>, paeduc <dbl>, income <dbl>, race_3cat <dbl>,
## #   age_4cat <dbl>
```

```
##
##   1   2
## 381 377
```

```
##
```

```
##         1         2
## 0.5026385 0.4973615
```

```
##
##         1         2
## 0.7743658 0.2256342
```

```
##
##   1   2   3
## 329 126 387
```

```
##
##         1         2         3
## 0.3907363 0.1496437 0.4596200
```

```
##
##   1   2
## 718 706
```

```
##
##   1   2   3
## 272  96 302
```

```
##
##         1         2         3
## 0.4059701 0.1432836 0.4507463
```

```
##
##   1   2   3
## 601 222 689
```

```
##
##   1   2   3
## 425 415 681
```

```
##
##   1   2   3   4
## 149 654 695  53
```

```
##
##   lower class  middle class   upper class working class
##           149           695            53           654
```

```
##
## Female   Male
##    868    691
```

```
## # A tibble: 1,559 x 1
##                coldeg1
##              <dbl+lbl>
##  1     1 [associate's]
##  2     2 [bachelor's]
##  3 NA(i) [IAP]
##  4     2 [bachelor's]
##  5 NA(i) [IAP]
##  6 NA(i) [IAP]
##  7     4 [mba]
##  8 NA(i) [IAP]
##  9 NA(i) [IAP]
## 10 NA(i) [IAP]
```

```
## # ... with 1,549 more rows
```

```
##
##   1   2   3   4   5   6   7   8
##  82 215  95   9   4  15   5   7
```

```
##
## Not Available     Strong No    Strong Yes       Weak No      Weak Yes
##           375           150           420           227           387
```

```
##
##  No Yes
## 150 420
```

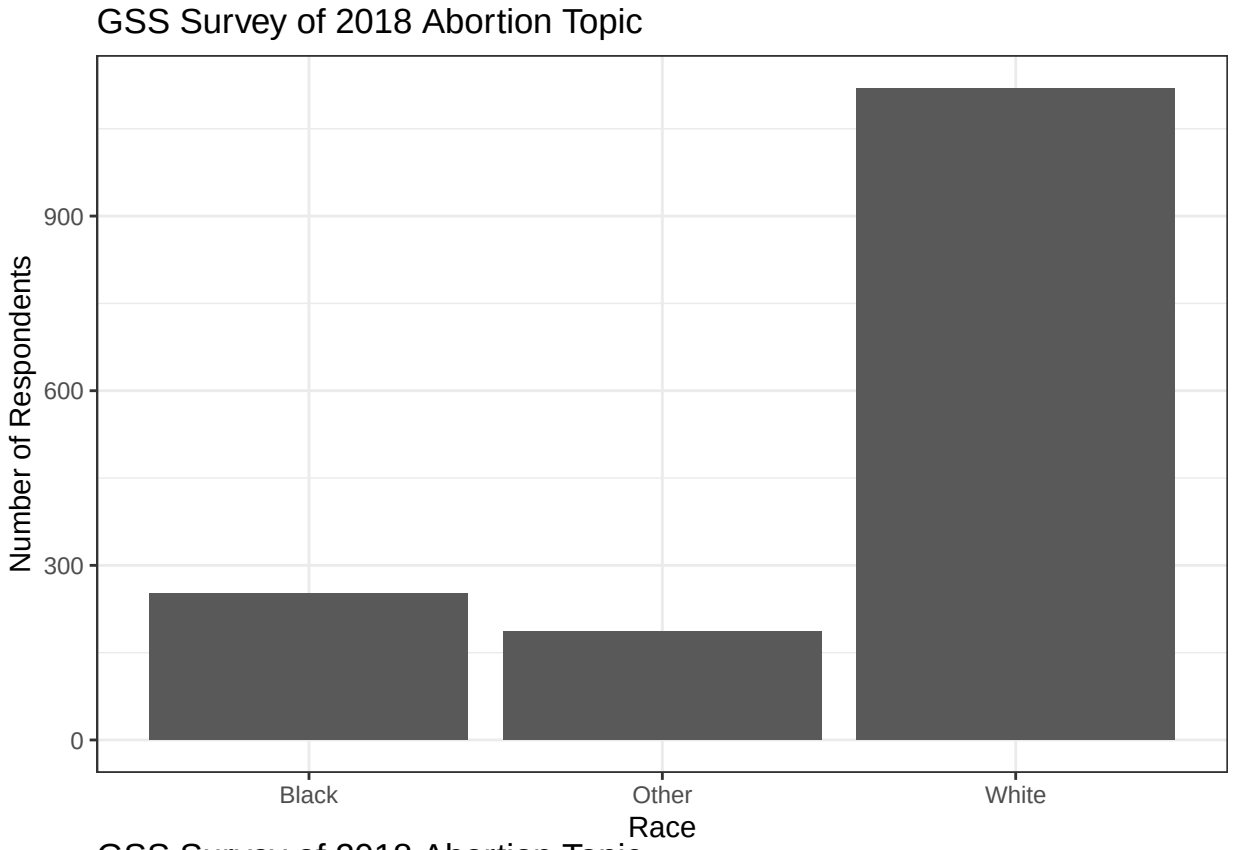
Including Visualizations

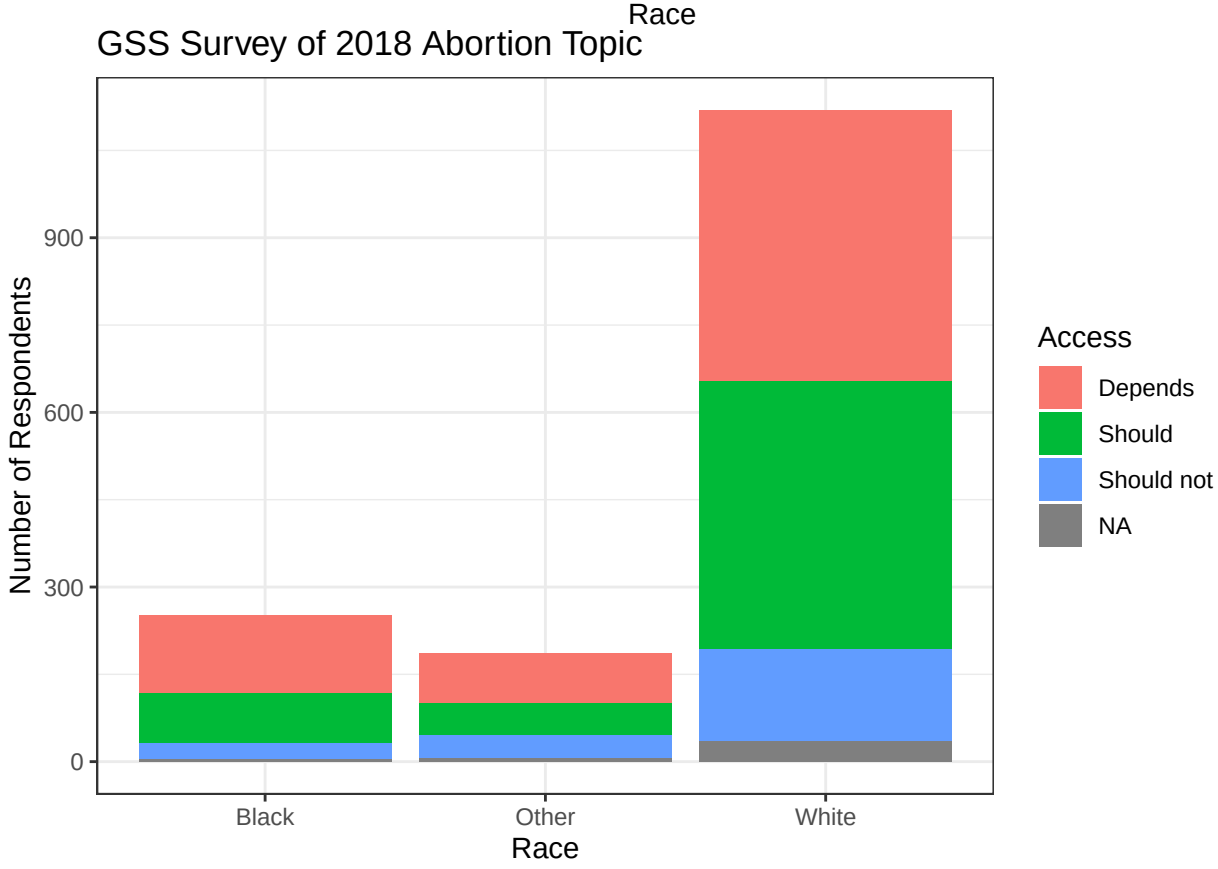
The first graph is a simple bar graph. I used the variable race. Although it seems fine, I would have liked to see the number per category above or inside the bars themselves. Not sure how to add that feature. The next graph is also a bar graph however this one is a stacked bar graph. This presents how each racial category varied by the variable about whether they think abortion should be legal. In this graph I could not remove the NAs. I tried filtering but it did not function so I left the NAs.

The third graph is a population pyramid. I used age and gender but not much difference exists. The female looks slightly larger and it makes sense because the there were more females that took this particular survey.

The last two graphs are histograms. The first is age with should abortion be legal. The last I removed NAs but it is same.

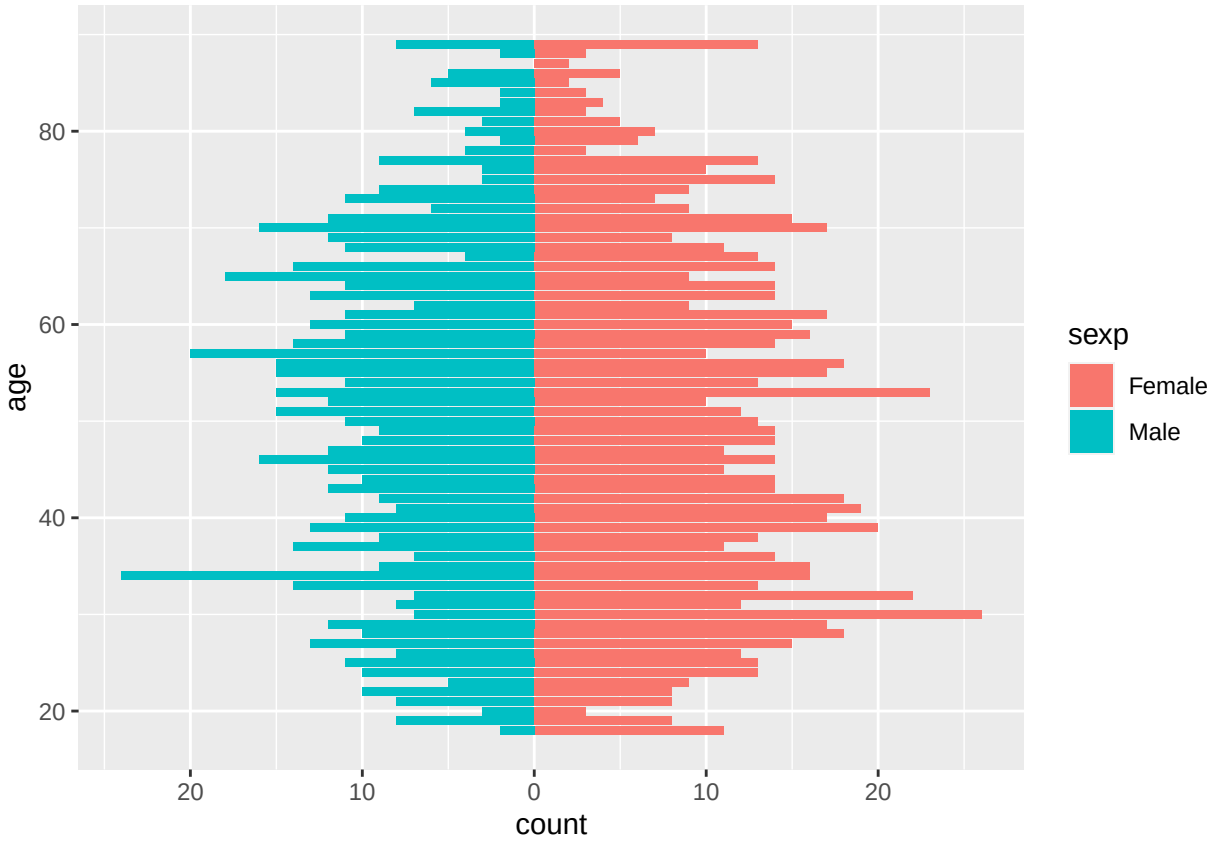
I used only a couple of variables at the time I thought were interesting but I created more variables that I want to add. For now I only added 5 graphs that I wanted to learn how to do. I want to be more creative especially around the legend and colors. I will be working on



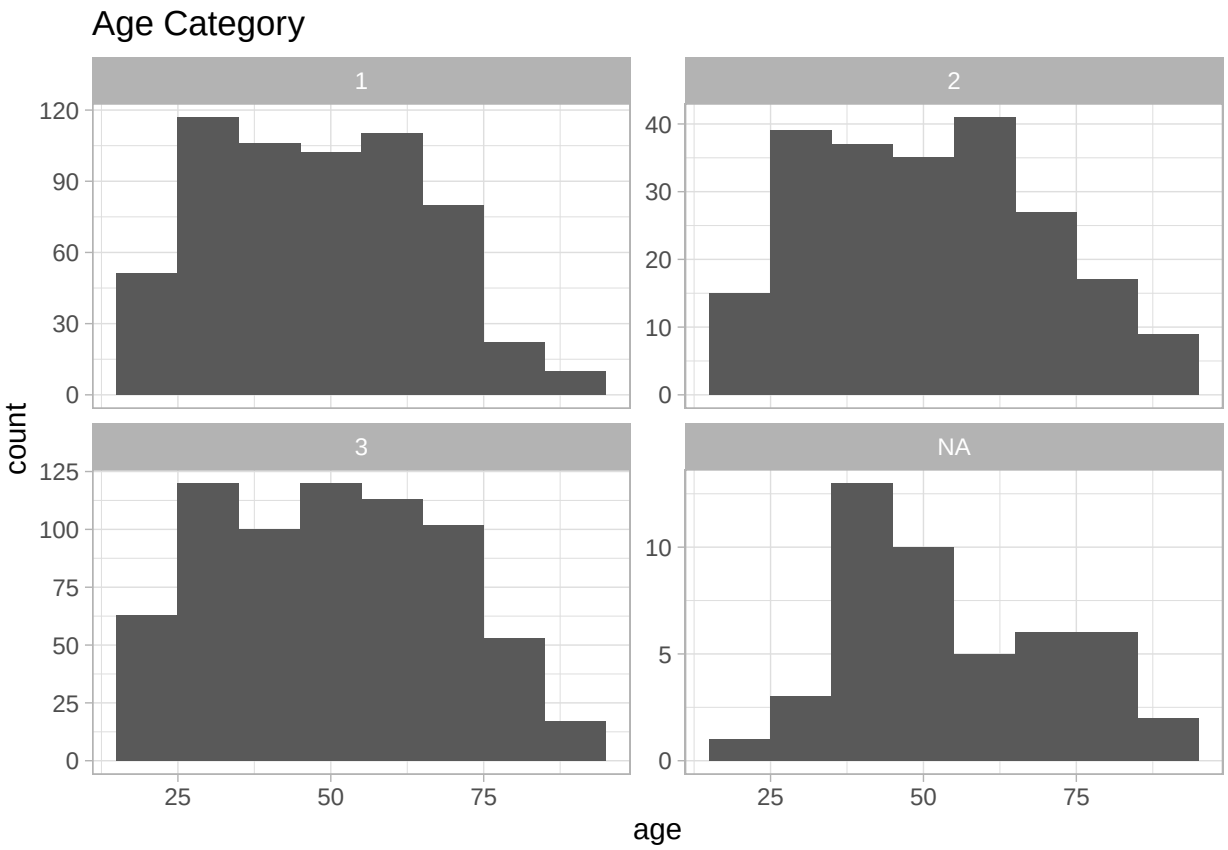



```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
```

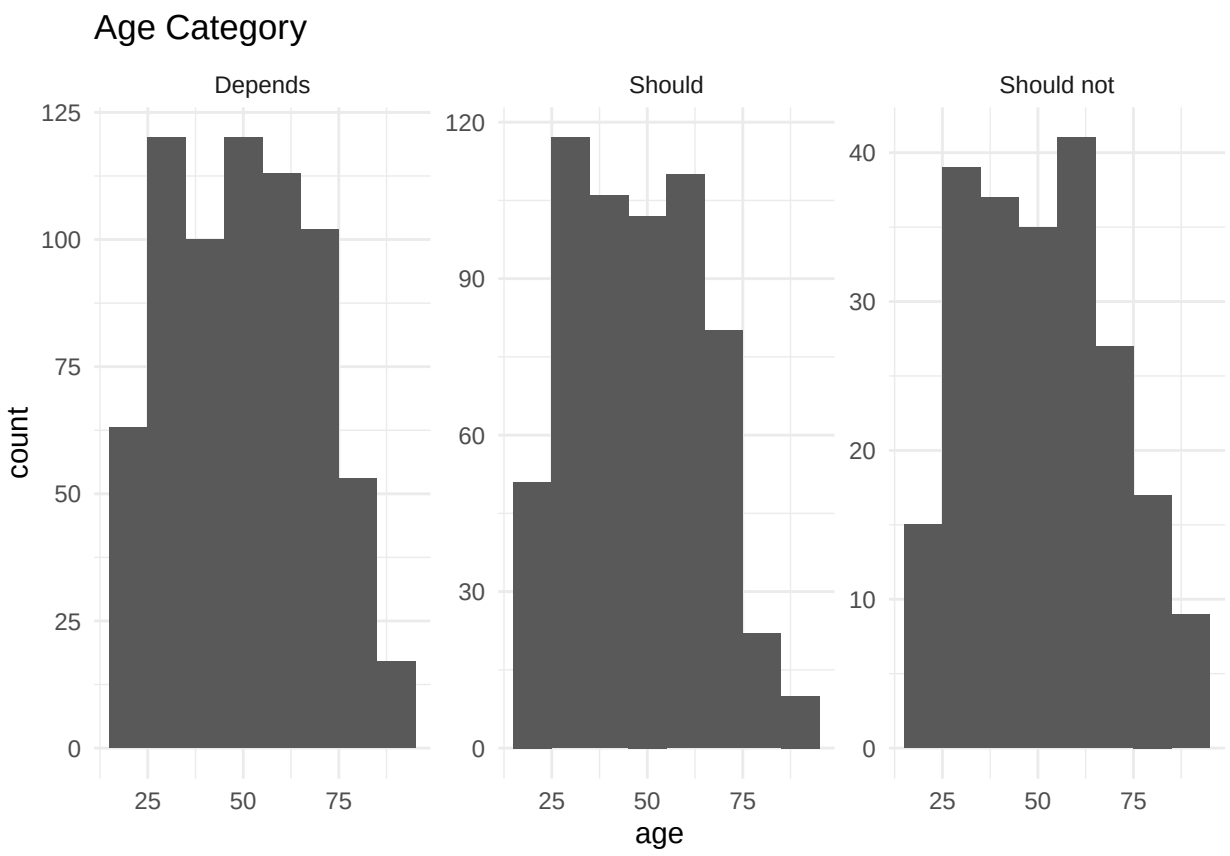
```
## Warning: Removed 4 rows containing non-finite values (stat_count).
## Warning: Removed 3 rows containing non-finite values (stat_count).
```




```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
## Warning: Removed 7 rows containing non-finite values (stat_bin).
```




```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
```


Age Category
Depends
Should
Should not
count
age
125
100
75
50
25
0
120
90
60
30
40
30
20
10
25
50
75