Final Project Code by AJ Acacio

Code

Loading and Cleaning Data

Preparing Packages:

Hide

```
library(tidyverse)
library(haven)
library(infer)
library(descr)
library(stargazer)
```

```
Please cite as:

 Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
 R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

Loading the Data from GSS2021:

Hide

```
gss2021 <- read_dta("gss2021.dta")
View(gss2021)
```

Selecting the columns I need for my analysis:

Hide

```
#Selecting my IV (age) and DV (natspac) and dropping the NAs
spaceage <- gss2021 %>%
  select(age, natspac) %>%
  drop_na()
spaceage
```

age <S3: haven_labelled>	natspac <S3: haven_labelled>
65	2
33	2
20	1
37	1
71	3
21	1
75	2
65	1
31	1
55	2

1-10 of 1,801 rows Previous **1** 2 3 4 5 6 … 100 Next

Binning the age to convert it into categorical variables:

Hide

```
#Here, I will bin the ages into categories, using generally agreed upon cutoffs for the generations, using this s
ource: https://www.beresfordresearch.com/age-range-by-generation/

generations <- spaceage %>%
  mutate(natspac = case_when(natspac == 1 ~ "Too Little",
                             natspac == 2 ~ "About Right",
                             natspac == 3 ~ "Too Much" )) %>%
  mutate(generation = ifelse(age > 0 & age <= 24, "Gen Z",
                      ifelse(age > 24 & age <= 40, "Millennial",
                      ifelse(age > 40 & age <= 56, "Gen X",
                      ifelse(age > 56 & age <= 75, "Boomers",
                      "Post-War")))))
```

Creating the contingency table, conducting the chi-squared test, and constructing the crosstab visualization:

Hide

```
#Creating a Contingency Table
contingency <- table(generations$natspac, generations$generation)
contingency
```

```
             Boomers Gen X Gen Z Millennial Post-War
  About Right     353   231    51        219       95
  Too Little      131   128    26        137       34
  Too Much        171   100    12         89       24
```

Hide

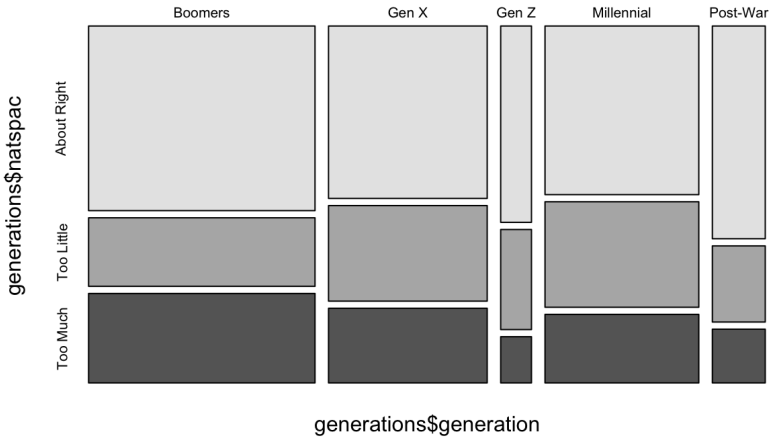
```
#Conducting the Chi-Squared Test
chisq.test(generations$generation, generations$natspac)
```

```
    Pearson's Chi-squared test

data:  generations$generation and generations$natspac
X-squared = 31.047, df = 8, p-value = 0.0001378
```

Hide

```
#Creating a crosstab visualization
crosstab(generations$natspac, generations$generation)
```

```
   Cell Contents
|-------------------------|
|                   Count |
|-------------------------|

=============================================================================
                        generations$generation
generations$natspac     Boomers   Gen X   Gen Z   Millennial   Post-War   Total
-----------------------------------------------------------------------------
About Right                 353     231      51          219          95     949
-----------------------------------------------------------------------------
Too Little                  131     128      26          137          34     456
-----------------------------------------------------------------------------
Too Much                    171     100      12           89          24     396
-----------------------------------------------------------------------------
Total                       655     459      89          445         153    1801
=============================================================================
```



Constructing the Univariate and Multivariate Regression:

Hide

```
#Creating the univariate regression
modelA <- (lm(natspac ~ age, data = spaceage))
summary(modelA)
```

```

Call:
lm(formula = natspac ~ age, data = spaceage)

Residuals:
     Min       1Q   Median       3Q      Max
-1.08878 -0.85970  0.01746  0.12370  1.14694

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.7933048  0.0512632  34.982  < 2e-16 ***
age         0.0033199  0.0009317   3.563 0.000376 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.685 on 1799 degrees of freedom
Multiple R-squared:  0.007009,  Adjusted R-squared:  0.006457
F-statistic:  12.7 on 1 and 1799 DF,  p-value: 0.0003756
```

Hide

```
#Creating the multivariate regression that will control for income, political affiliation, education

confounds <- gss2021 %>%
  select(age, natspac, educ, incom16, polviews) %>%
  drop_na()

confounds
```

age <S3: haven_labelled>	natspac <S3: haven_labelled>	educ <S3: haven_labelled>	incom16 <S3: haven_labelled>	polviews <S3: haven_labelled>
65	2	12	2	4
33	2	17	3	3

age <S3: haven_labelled>	natspac <S3: haven_labelled>	educ <S3: haven_labelled>	incom16 <S3: haven_labelled>	polviews <S3: haven_labelled>
20	1	12	3	3
37	1	11	2	5
71	3	16	5	2
21	1	14	4	2
75	2	18	3	3
65	1	12	1	2
31	1	16	3	2
55	2	16	4	6

1-10 of 1,734 rows Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
modelB <- lm(natspac ~ age + educ + incom16 + polviews, data = confounds)
summary(modelB)
```

```
Call:
lm(formula = natspac ~ age + educ + incom16 + polviews, data = confounds)

Residuals:
     Min       1Q   Median       3Q      Max
-1.40448 -0.70251  0.01846  0.22000  1.31578

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.3594431  0.1172541  20.122  < 2e-16 ***
age          0.0032246  0.0009542   3.379 0.000743 ***
educ        -0.0253768  0.0060582  -4.189 2.95e-05 ***
incom16     -0.0689473  0.0175228  -3.935 8.66e-05 ***
polviews     0.0005619  0.0108716   0.052 0.958785
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6788 on 1729 degrees of freedom
Multiple R-squared:  0.03026,   Adjusted R-squared:  0.02802
F-statistic: 13.49 on 4 and 1729 DF,  p-value: 7.868e-11
```

Hide

```
#Age (in years), education (in years attained), incom16 (likert scale from well below average to well above avera
ge), political views (scale of 1 (most liberal) to 7 (most conservative))


#I use the stargazer package to make nicer visualization of my regression tables.
stargazer(modelA, modelB, type = "html", title = "OLS Regression Analysis",
          dep.var.caption = "DV: Sentiment on Space Exploration (1-3)",
         out = "bothregressions.txt")
```

```
<table style="text-align:center"><caption><strong>OLS Regression Analysis</strong></caption>
<tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"></td><td col
span="2">DV: Sentiment on Space Exploration (1-3)</td></tr>
<tr><td></td><td colspan="2" style="border-bottom: 1px solid black"></td></tr>
<tr><td style="text-align:left"></td><td colspan="2">natspac</td></tr>
<tr><td style="text-align:left"></td><td>(1)</td><td>(2)</td></tr>
<tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">age</td><td>
0.003<sup>***</sup></td><td>0.003<sup>***</sup></td></tr>
<tr><td style="text-align:left"></td><td>(0.001)</td><td>(0.001)</td></tr>
<tr><td style="text-align:left"></td><td></td><td></td></tr>
<tr><td style="text-align:left">educ</td><td></td><td>-0.025<sup>***</sup></td></tr>
<tr><td style="text-align:left"></td><td></td><td>(0.006)</td></tr>
<tr><td style="text-align:left"></td><td></td><td></td></tr>
<tr><td style="text-align:left">incom16</td><td></td><td>-0.069<sup>***</sup></td></tr>
<tr><td style="text-align:left"></td><td></td><td>(0.018)</td></tr>
<tr><td style="text-align:left"></td><td></td><td></td></tr>
<tr><td style="text-align:left">polviews</td><td></td><td>0.001</td></tr>
<tr><td style="text-align:left"></td><td></td><td>(0.011)</td></tr>
<tr><td style="text-align:left"></td><td></td><td></td></tr>
<tr><td style="text-align:left">Constant</td><td>1.793<sup>***</sup></td><td>2.359<sup>***</sup></td></tr>
<tr><td style="text-align:left"></td><td>(0.051)</td><td>(0.117)</td></tr>
<tr><td style="text-align:left"></td><td></td><td></td></tr>
<tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">Observations
</td><td>1,801</td><td>1,734</td></tr>
<tr><td style="text-align:left">R<sup>2</sup></td><td>0.007</td><td>0.030</td></tr>
<tr><td style="text-align:left">Adjusted R<sup>2</sup></td><td>0.006</td><td>0.028</td></tr>
<tr><td style="text-align:left">Residual Std. Error</td><td>0.685 (df = 1799)</td><td>0.679 (df = 1729)</td></tr>
<tr><td style="text-align:left">F Statistic</td><td>12.698<sup>***</sup> (df = 1; 1799)</td><td>13.490<sup>***</s
up> (df = 4; 1729)</td></tr>
<tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"><em>Note:</e
m></td><td colspan="2" style="text-align:right"><sup>*</sup>p<0.1; <sup>**</sup>p<0.05; <sup>***</sup>p<0.01</td>
</tr>
</table>
```