SMI205 Replication Project (2023)

200382018

22/05/2023

Table of Contents

# [Replication and Extension of ‘Getting the Message?’ Study: Investigating Choice, Self-Selection, and Argument Efficacy in Social Movements]

### Rpubs link: [copy Rpubs url address here]

### GitHub Repository: [add url here if you have created a data or code repository, if not delete this line]

### Study Preregistration form: [copy Rpubs url address here or delete]

## Information about this replication project

* Replication project based on paper [add full citation here and link to its published online version]
* Replication method (select one from below):
  + Used replication package available at [add citation + repository link here]
  + Used materials obtained from authors
  + Own replication following methods section of the paper
  + Other - explain

## Workspace setup

### YAML settings

output:   html\_document:    code\_download: true     toc: true     toc\_depth: 2     toc\_float:      collapsed: false      smooth\_scroll: true

### Global settings of R chunks

# Global options  
opts\_chunk$set(echo=TRUE,  
 cache=TRUE,  
 comment=NA,  
 message=FALSE,  
 warning=FALSE)

### Libraries

# All used libraries  
library("rmarkdown")  
library("knitr")  
library("tidyverse")  
library("car")  
library("Hmisc")  
library("sessioninfo")  
library("ggplot2")  
library("dplyr")

### Versions of used packages

$rmarkdown  
[1] '2.21'  
  
$knitr  
[1] '1.42'

### My enviroment

NULL

## 1. Introduction

*Code and outputs are not included in the wordcount*

**500 words**

Critical and evidence-based reflections on the paper and your new research question(s).

* Clearly state which claim (finding/argument) from the replicated paper you discuss or test robustness of any methodological assumptions.
* What the authors argued, and what could be counter-argued, and why.
* Use at least 5 scientific references to support your critique and your idea for research extension.

Refer to the four types of replications as discussed in Freese and Peterson (2017) and explain what aspects of the original research your replication tests.

Briefly summarise, what results you expected to find and preregistered.

## 2. Data and methods

### 2.1. Data

**250 words**

Transparently **describe data** used in the original paper and in your replication project: full dataset names, years of data collection, sources (who created data), sample sizes, variables used in the model you replicate.

Embed R code to upload your data (recommendation - show code only using results='hide' in the chunk options).

Remember to add all datasets with their doi number to the list of references. See examples in the conditions of use for the ESS data [here](https://www.europeansocialsurvey.org/data/conditions_of_use.html).

Report all steps you have undertaken in **data transformation**:

* Subsetting data by selecting a lower number of variables or excluding any cases.
* Merging data files.
* Steps in recoding and renaming all variables.
* Embed R code below to present the steps in data manipulation. If this would be very long chunk of R code (over 1 page), show just some examples and hide the rest of the code in text (include=FALSE option). The entire code will be still displayed in Appendix 2.

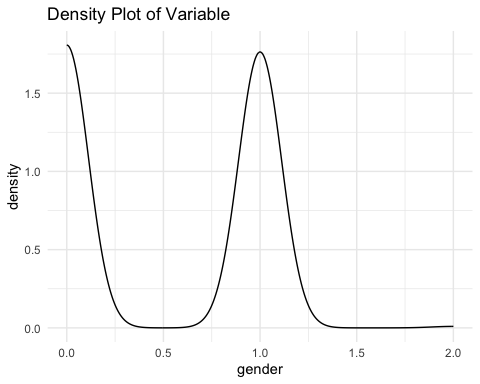
You **have to** display some summary statistics of your key variables of interest. This could be done as a table/s or a graph/s presenting distribution of your dependent variable/s for the entire sample or by a group of interest, boxplots, correlation plots or graphs with means values.

load("/Users/georgiamugglestone/Documents/smi 205/Assessments /200382018 2/dataverse\_files-2/df\_mtg.rda")

summary\_stats <- df\_mtg %>%  
 summarise(  
 mean\_variable = mean(gender),  
 sd\_variable = sd(gender),  
 median\_variable = median(gender),  
 min\_variable = min(gender),  
 max\_variable = max(gender)  
 )  
  
# Display the summary statistics  
print(summary\_stats)

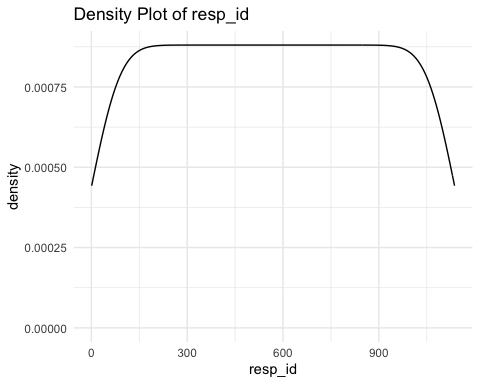
mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.4978012 0.505467 0 0 2

# Create Figure 3 with statistical analysis  
fig3 <- df\_mtg %>%  
 ggplot(aes(x = gender)) +  
 geom\_density() +  
 ggtitle("Density Plot of Variable") +  
 theme\_minimal()  
  
# Display Figure 3  
fig3

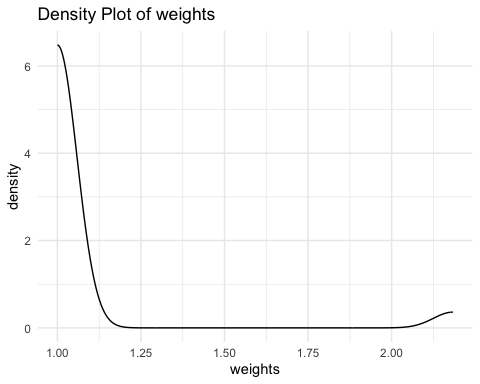


# Perform statistical analysis and create plots for all variables  
  
# Loop over each variable  
for (col in names(df\_mtg)) {  
 # Calculate summary statistics  
 summary\_stats <- df\_mtg %>%  
 summarise(  
 mean\_variable = mean(!!sym(col)),  
 sd\_variable = sd(!!sym(col)),  
 median\_variable = median(!!sym(col)),  
 min\_variable = min(!!sym(col)),  
 max\_variable = max(!!sym(col))  
 )  
   
 # Display the summary statistics  
 cat("Summary Statistics for", col, ":\n")  
 print(summary\_stats)  
   
 # Create plot for the variable  
 fig <- df\_mtg %>%  
 ggplot(aes(x = !!sym(col))) +  
 geom\_density() +  
 ggtitle(paste("Density Plot of", col)) +  
 theme\_minimal()  
   
 # Display the plot  
 print(fig)  
}

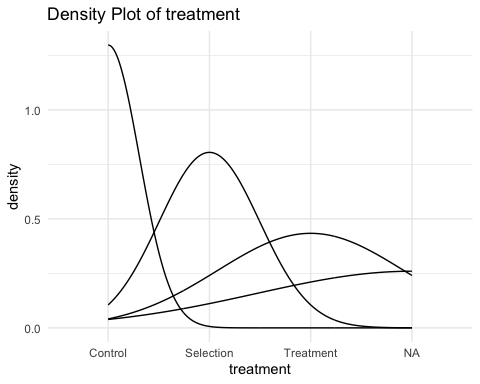
Summary Statistics for resp\_id :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 569 328.3679 569 1 1137



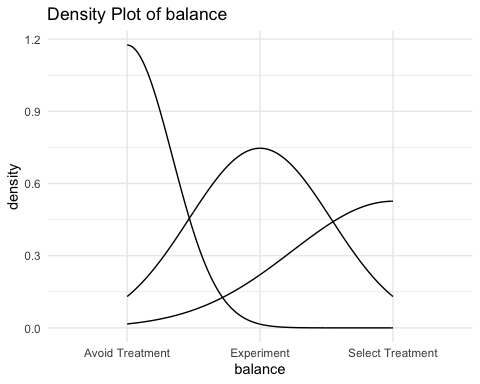
Summary Statistics for weights :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 1.062445 0.26468 1 1 2.183333



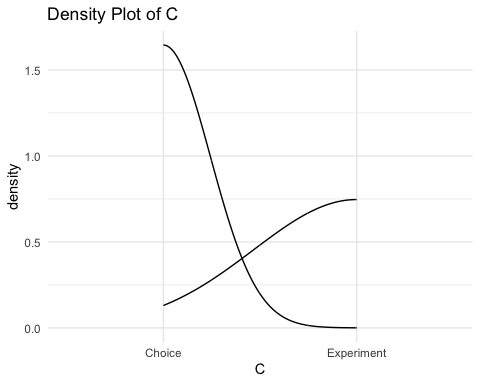
Summary Statistics for treatment :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA <NA> <NA> <NA>



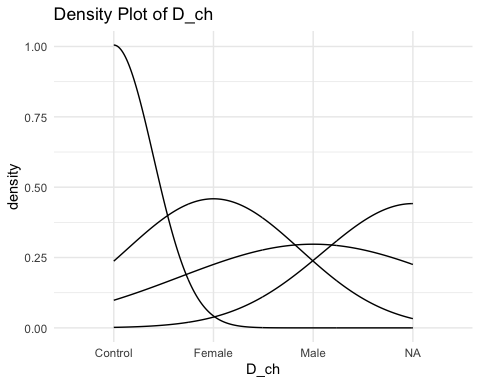
Summary Statistics for balance :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Select Treatment Avoid Treatment Select Treatment



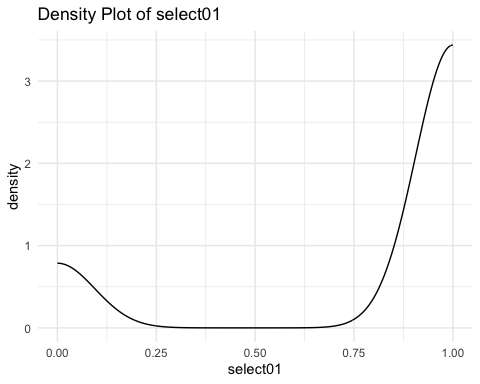
Summary Statistics for C :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Choice Choice Experiment



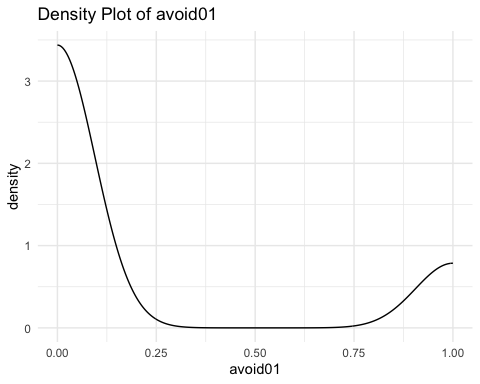
Summary Statistics for D\_ch :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA <NA> <NA> <NA>



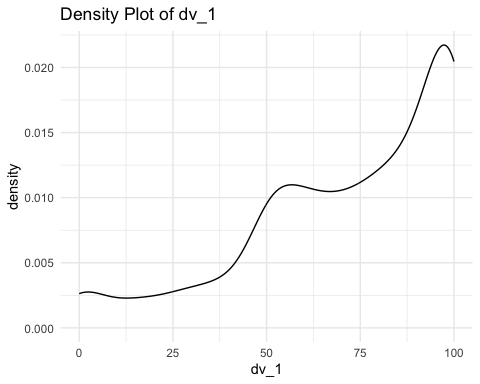
Summary Statistics for select01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



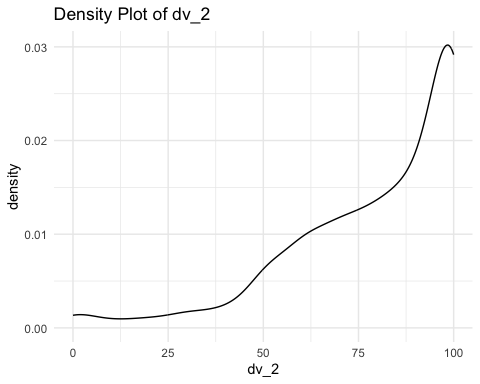
Summary Statistics for avoid01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



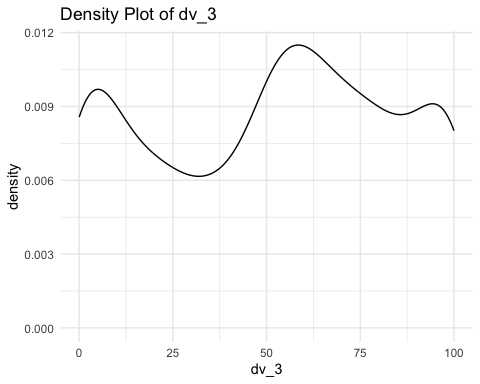
Summary Statistics for dv\_1 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



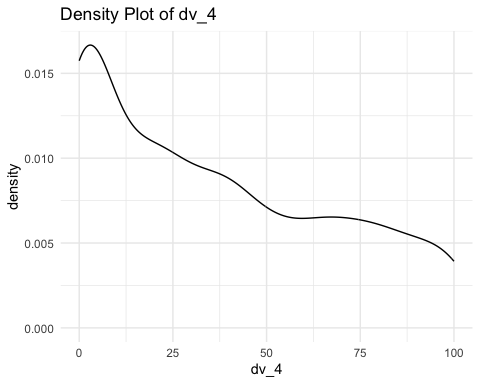
Summary Statistics for dv\_2 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 79.19877 22.69061 85 0 100



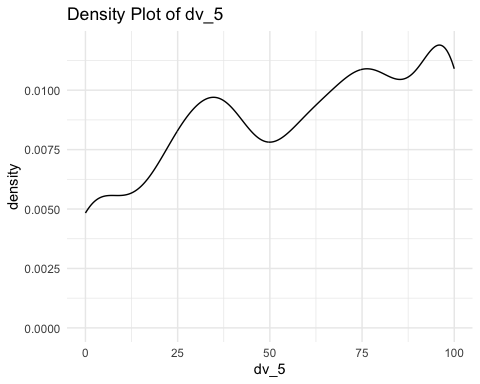
Summary Statistics for dv\_3 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



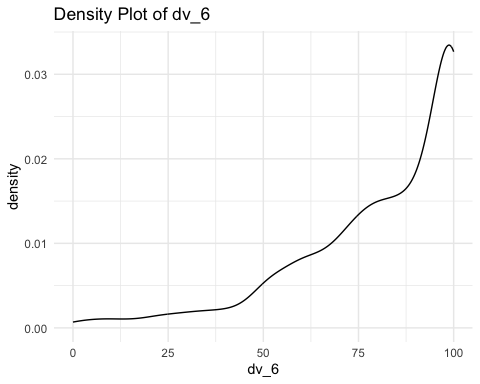
Summary Statistics for dv\_4 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 36.20405 31.62426 30 0 100



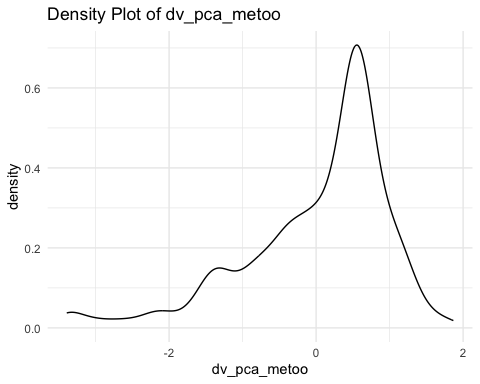
Summary Statistics for dv\_5 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



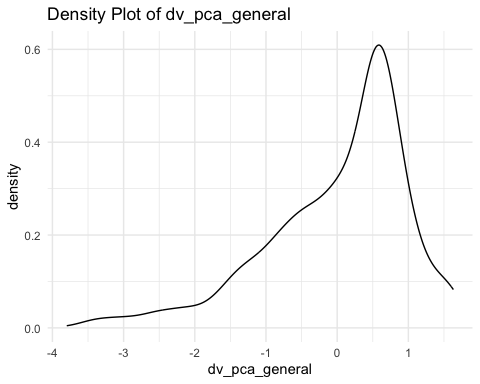
Summary Statistics for dv\_6 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



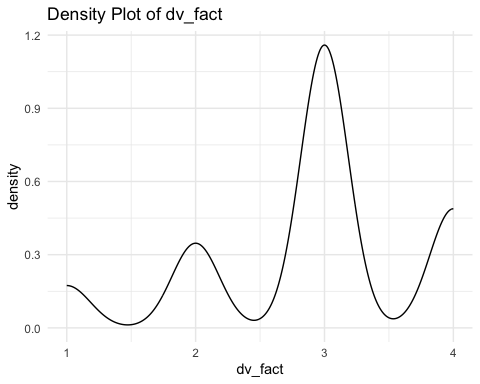
Summary Statistics for dv\_pca\_metoo :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



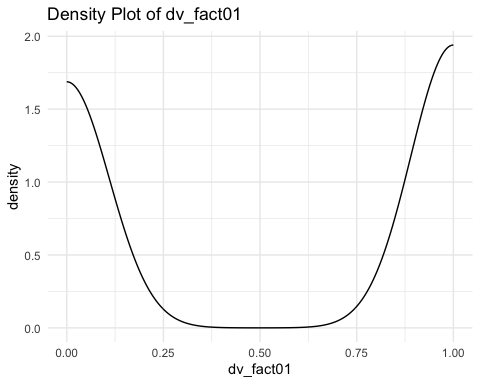
Summary Statistics for dv\_pca\_general :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



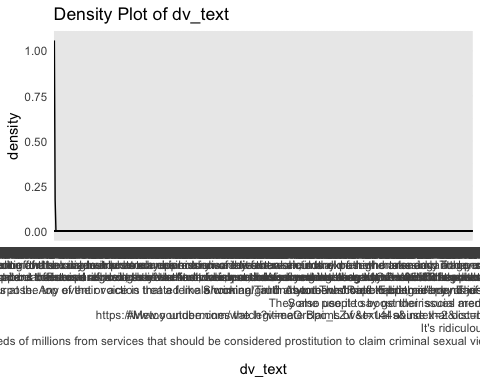
Summary Statistics for dv\_fact :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 2.905013 0.8348386 3 1 4



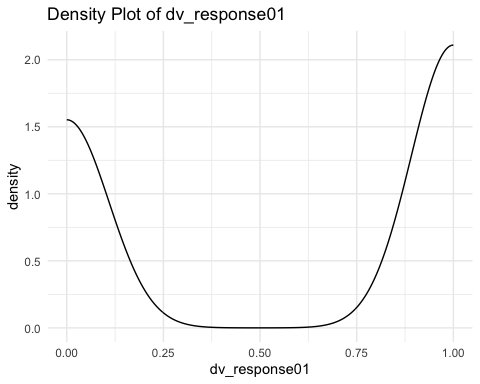
Summary Statistics for dv\_fact01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.5347405 0.4990111 1 0 1



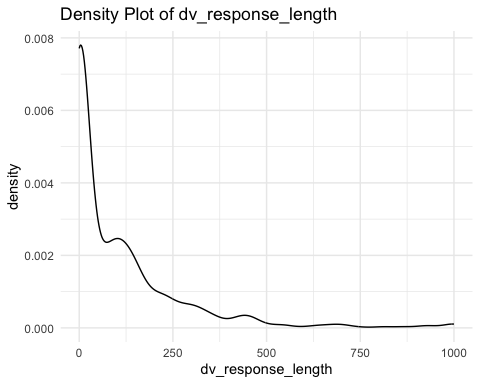
Summary Statistics for dv\_text :  
 mean\_variable sd\_variable  
1 NA NA  
 median\_variable  
1 I think it's positive that people are made aware of the prevalence of sexual harassment. I don't necessarily think it will stop the perpetrators because many are sociopaths who don't care. But it will help entities such as law enforcement and human resources understand what is happening.  
 min\_variable max\_variable  
1 yes.



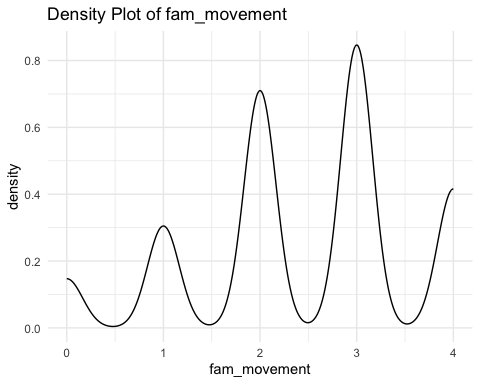
Summary Statistics for dv\_response01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.5760774 0.4943958 1 0 1



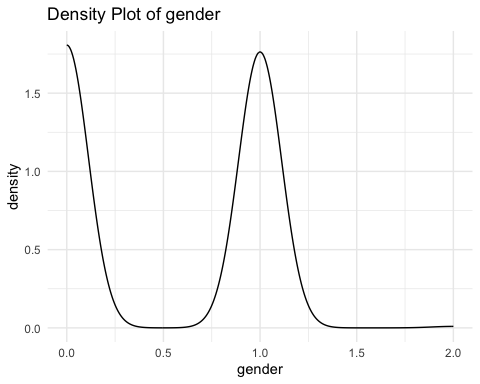
Summary Statistics for dv\_response\_length :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 109.2023 166.992 42 0 1000



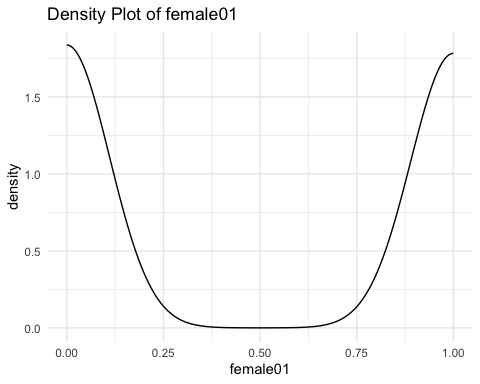
Summary Statistics for fam\_movement :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 2.445031 1.0985 3 0 4



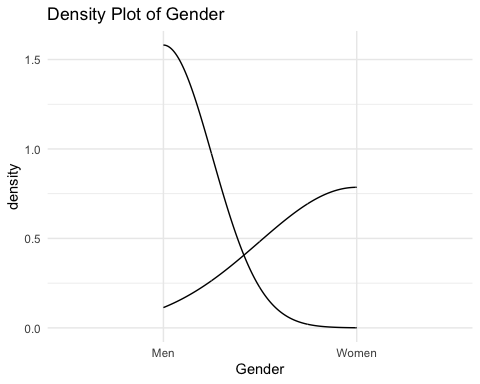
Summary Statistics for gender :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.4978012 0.505467 0 0 2



Summary Statistics for female01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.4925242 0.5001641 0 0 1



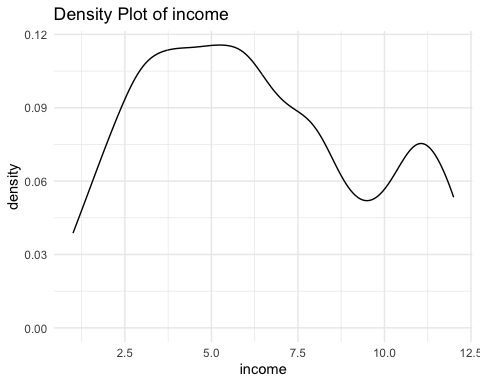
Summary Statistics for Gender :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Men Men Women



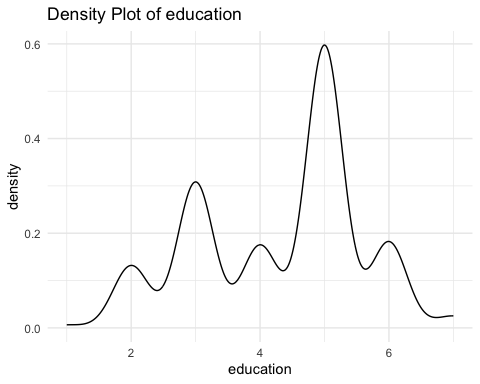
Summary Statistics for age :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 38.11873 11.69104 35 18 82



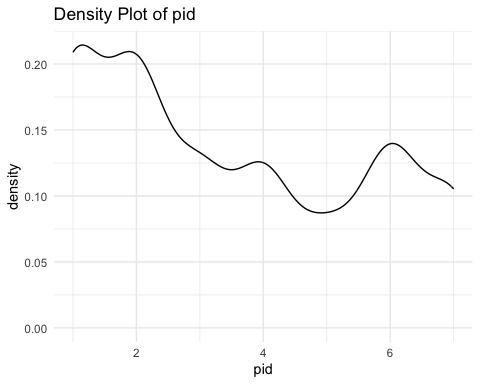
Summary Statistics for income :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



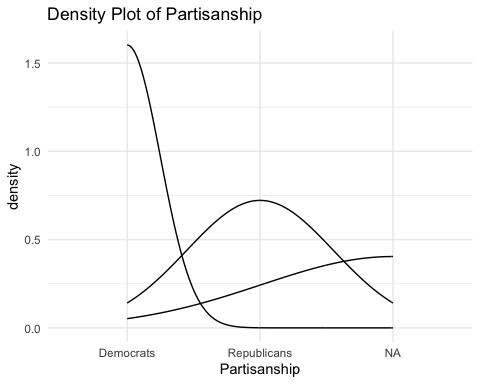
Summary Statistics for education :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 4.313105 1.271483 5 1 7



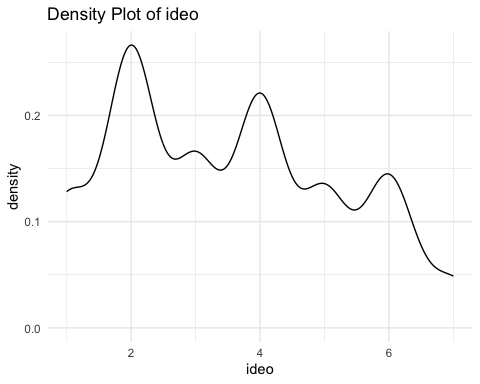
Summary Statistics for pid :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA NA NA



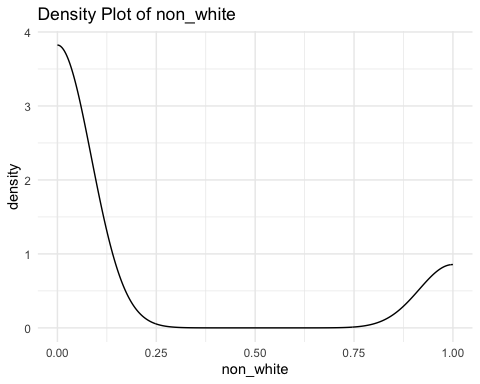
Summary Statistics for Partisanship :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA <NA> <NA> <NA>



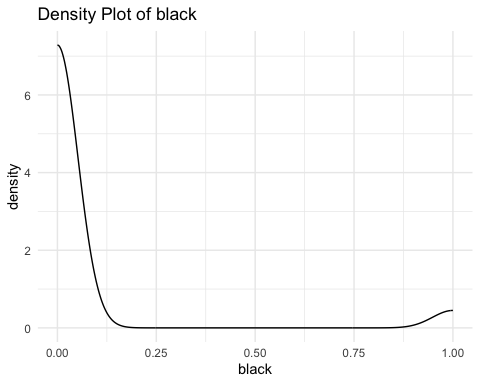
Summary Statistics for ideo :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 3.532982 1.723137 3 1 7



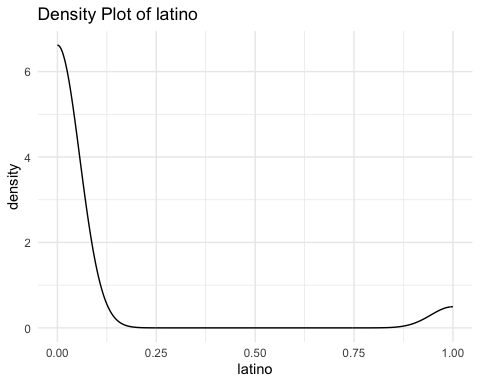
Summary Statistics for non\_white :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.1829376 0.3867854 0 0 1



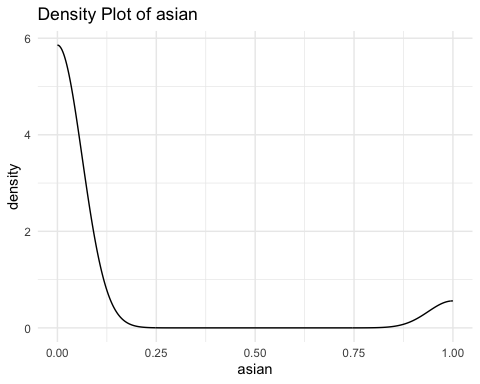
Summary Statistics for black :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.05804749 0.2339361 0 0 1



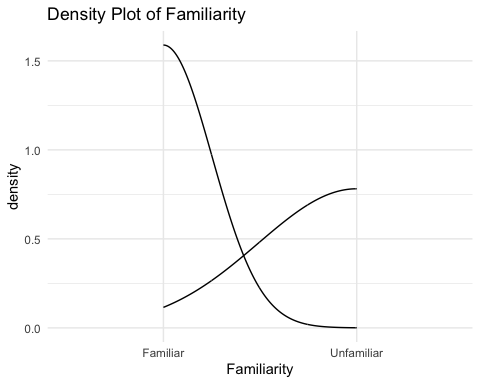
Summary Statistics for latino :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.06948109 0.2543824 0 0 1



Summary Statistics for asian :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.08707124 0.2820635 0 0 1

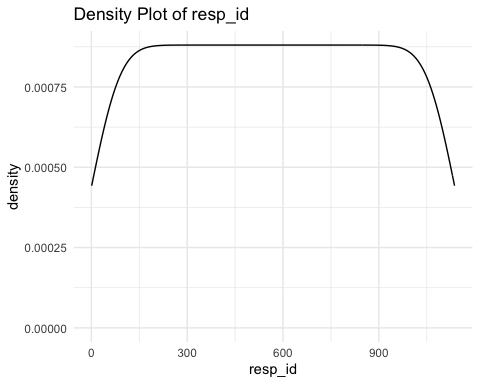


Summary Statistics for Familiarity :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Familiar Familiar Unfamiliar

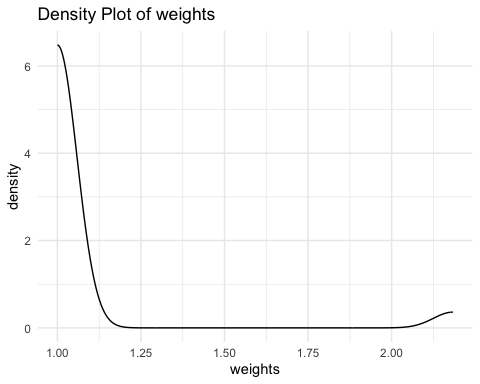


# Perform statistical analysis and create plots for all variables  
  
# Loop over each variable  
for (col in names(df\_mtg)) {  
 # Calculate summary statistics  
 summary\_stats <- df\_mtg %>%  
 summarise(  
 mean\_variable = mean(!!sym(col), na.rm = TRUE),  
 sd\_variable = sd(!!sym(col), na.rm = TRUE),  
 median\_variable = median(!!sym(col), na.rm = TRUE),  
 min\_variable = min(!!sym(col), na.rm = TRUE),  
 max\_variable = max(!!sym(col), na.rm = TRUE)  
 )  
   
 # Display the summary statistics  
 cat("Summary Statistics for", col, ":\n")  
 print(summary\_stats)  
   
 # Check if there is sufficient variability in the data  
 if (!is.na(summary\_stats$sd\_variable) && summary\_stats$sd\_variable > 0.001) {  
 # Perform t-test  
 ttest\_result <- t.test(df\_mtg[[col]], mu = 0, na.rm = TRUE)  
   
 # Display the t-test results  
 cat("T-Test Results for", col, ":\n")  
 print(ttest\_result)  
 } else {  
 cat("Insufficient variability in the data or missing values. Skipping t-test for", col, "\n")  
 }  
   
 # Create plot for the variable  
 fig <- df\_mtg %>%  
 ggplot(aes(x = !!sym(col))) +  
 geom\_density() +  
 ggtitle(paste("Density Plot of", col)) +  
 theme\_minimal()  
   
 # Display the plot  
 print(fig)  
}

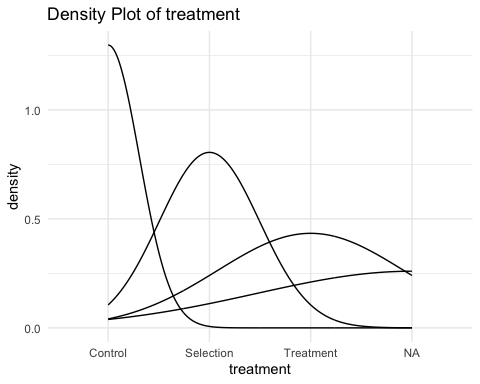
Summary Statistics for resp\_id :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 569 328.3679 569 1 1137  
T-Test Results for resp\_id :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 58.429, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 549.893 588.107  
sample estimates:  
mean of x   
 569



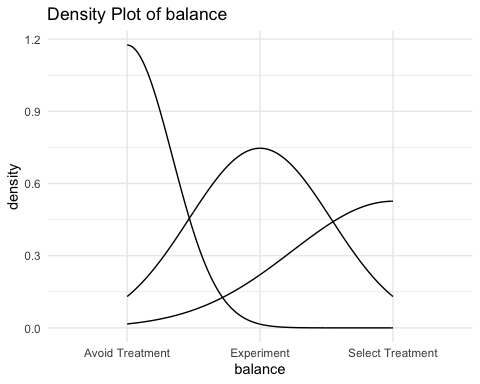
Summary Statistics for weights :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 1.062445 0.26468 1 1 2.183333  
T-Test Results for weights :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 135.35, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 1.047044 1.077846  
sample estimates:  
mean of x   
 1.062445



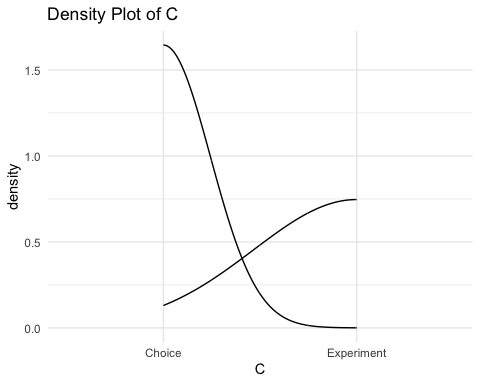
Summary Statistics for treatment :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA Control Treatment  
Insufficient variability in the data or missing values. Skipping t-test for treatment



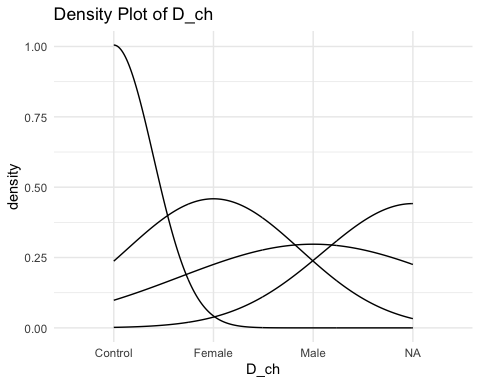
Summary Statistics for balance :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Select Treatment Avoid Treatment Select Treatment  
Insufficient variability in the data or missing values. Skipping t-test for balance



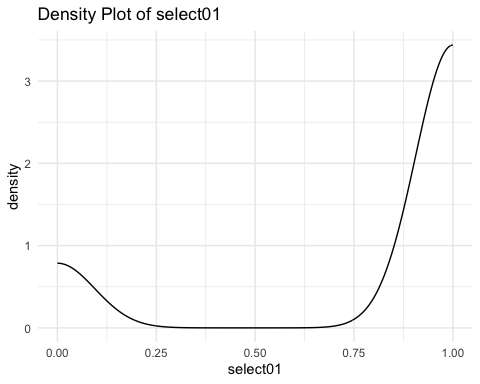
Summary Statistics for C :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Choice Choice Experiment  
Insufficient variability in the data or missing values. Skipping t-test for C



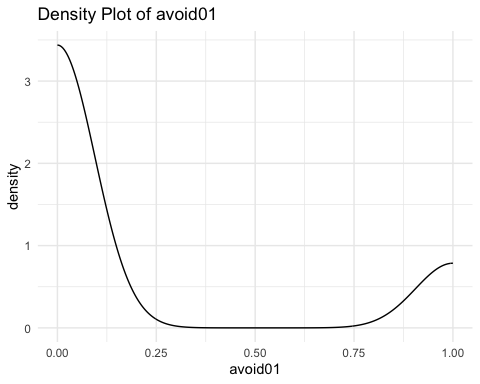
Summary Statistics for D\_ch :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Female Control Male  
Insufficient variability in the data or missing values. Skipping t-test for D\_ch



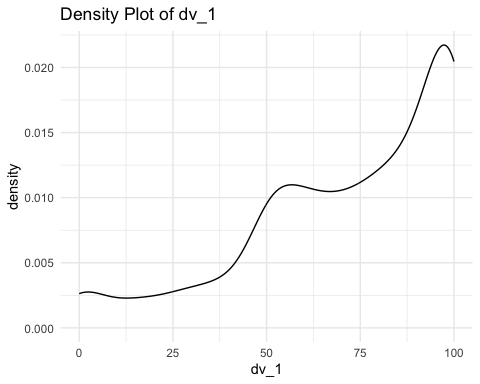
Summary Statistics for select01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.8139205 0.3894475 1 0 1  
T-Test Results for select01 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 55.452, df = 703, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.7851028 0.8427381  
sample estimates:  
mean of x   
0.8139205



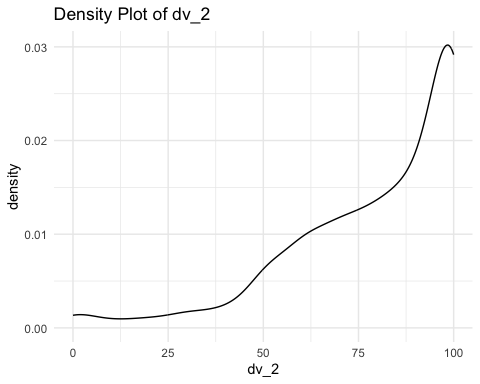
Summary Statistics for avoid01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.1860795 0.3894475 0 0 1  
T-Test Results for avoid01 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 12.678, df = 703, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.1572619 0.2148972  
sample estimates:  
mean of x   
0.1860795



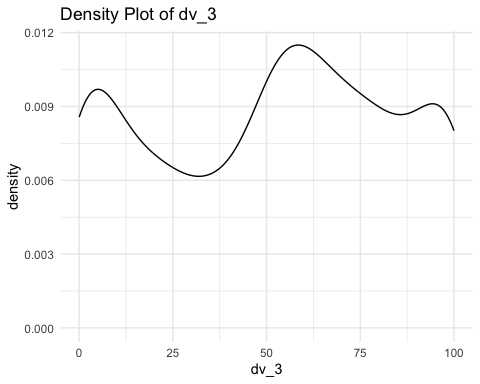
Summary Statistics for dv\_1 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 71.85878 27.19514 80 0 100  
T-Test Results for dv\_1 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 88.941, df = 1132, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 70.27356 73.44400  
sample estimates:  
mean of x   
 71.85878



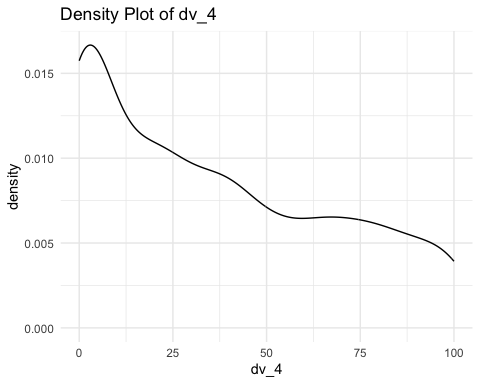
Summary Statistics for dv\_2 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 79.19877 22.69061 85 0 100  
T-Test Results for dv\_2 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 117.69, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 77.87846 80.51908  
sample estimates:  
mean of x   
 79.19877



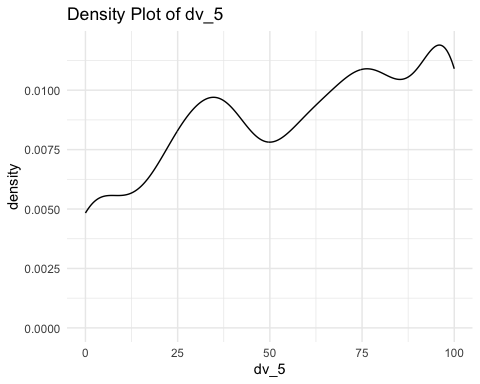
Summary Statistics for dv\_3 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 51.42254 32.44926 55 0 100  
T-Test Results for dv\_3 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 53.412, df = 1135, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 49.53356 53.31151  
sample estimates:  
mean of x   
 51.42254



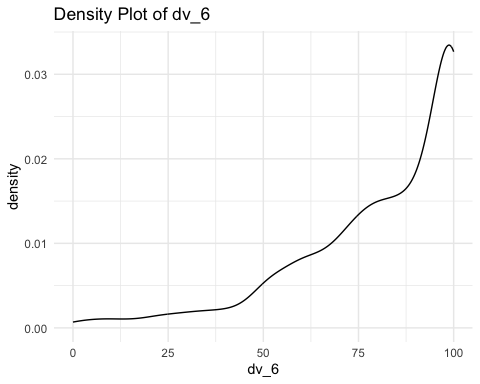
Summary Statistics for dv\_4 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 36.20405 31.62426 30 0 100  
T-Test Results for dv\_4 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 38.603, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 34.36390 38.04419  
sample estimates:  
mean of x   
 36.20405



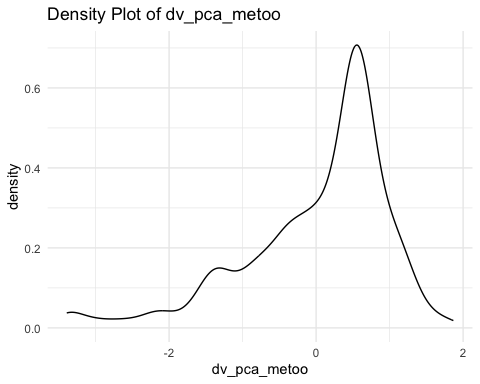
Summary Statistics for dv\_5 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 57.53128 31.12611 60 0 100  
T-Test Results for dv\_5 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 62.27, df = 1134, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 55.71852 59.34403  
sample estimates:  
mean of x   
 57.53128



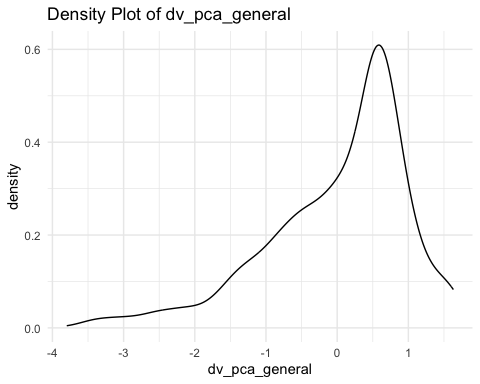
Summary Statistics for dv\_6 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 80.91373 21.72535 88 0 100  
T-Test Results for dv\_6 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 125.53, df = 1135, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 79.64903 82.17844  
sample estimates:  
mean of x   
 80.91373



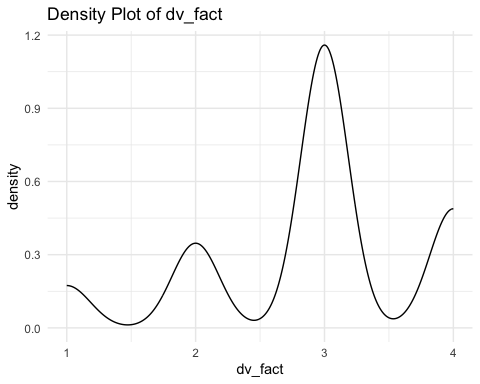
Summary Statistics for dv\_pca\_metoo :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.001410983 1.000532 0.3330572 -3.392406 1.867306  
T-Test Results for dv\_pca\_metoo :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 0.047469, df = 1132, p-value = 0.9621  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 -0.05691054 0.05973251  
sample estimates:  
 mean of x   
0.001410983



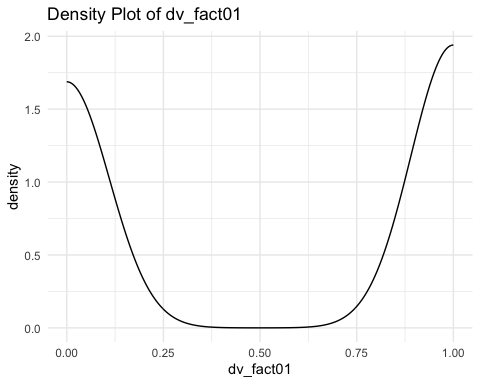
Summary Statistics for dv\_pca\_general :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.001338544 1.000029 0.2896232 -3.798618 1.632108  
T-Test Results for dv\_pca\_general :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 0.045074, df = 1133, p-value = 0.9641  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 -0.05692787 0.05960496  
sample estimates:  
 mean of x   
0.001338544



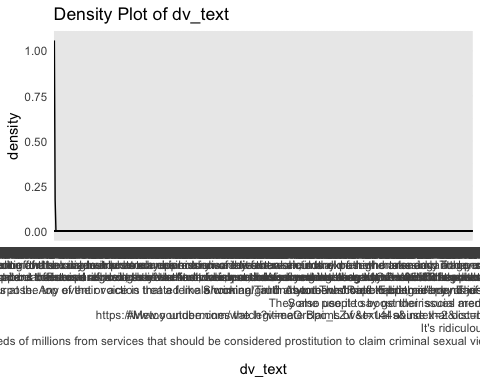
Summary Statistics for dv\_fact :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 2.905013 0.8348386 3 1 4  
T-Test Results for dv\_fact :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 117.33, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 2.856436 2.953590  
sample estimates:  
mean of x   
 2.905013



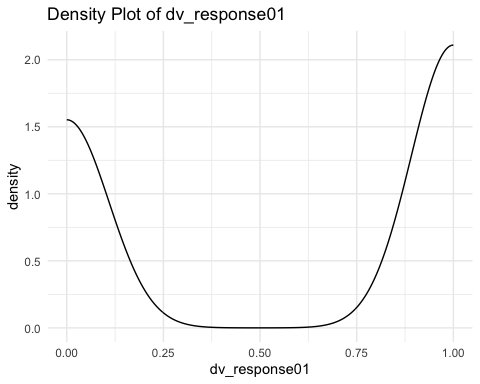
Summary Statistics for dv\_fact01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.5347405 0.4990111 1 0 1  
T-Test Results for dv\_fact01 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 36.134, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.5057043 0.5637768  
sample estimates:  
mean of x   
0.5347405



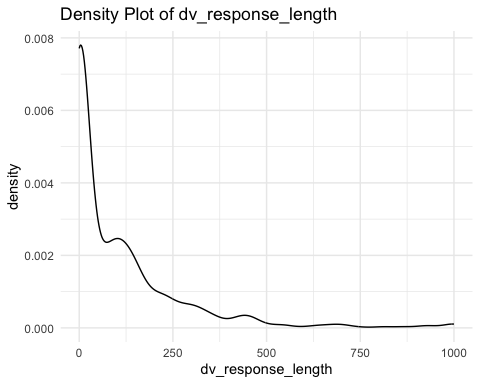
Summary Statistics for dv\_text :  
 mean\_variable sd\_variable  
1 NA NA  
 median\_variable  
1 I think it's positive that people are made aware of the prevalence of sexual harassment. I don't necessarily think it will stop the perpetrators because many are sociopaths who don't care. But it will help entities such as law enforcement and human resources understand what is happening.  
 min\_variable max\_variable  
1 yes.  
Insufficient variability in the data or missing values. Skipping t-test for dv\_text



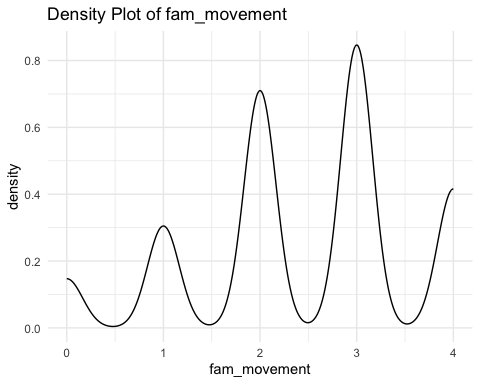
Summary Statistics for dv\_response01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.5760774 0.4943958 1 0 1  
T-Test Results for dv\_response01 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 39.29, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.5473097 0.6048451  
sample estimates:  
mean of x   
0.5760774



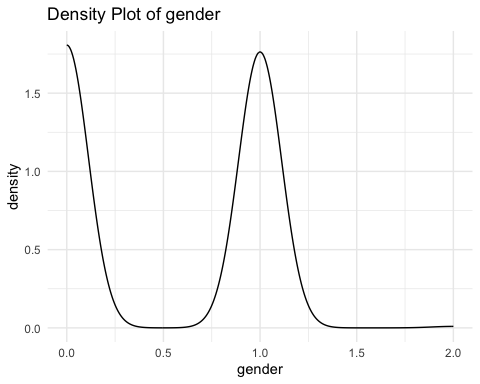
Summary Statistics for dv\_response\_length :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 109.2023 166.992 42 0 1000  
T-Test Results for dv\_response\_length :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 22.05, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 99.48541 118.91916  
sample estimates:  
mean of x   
 109.2023



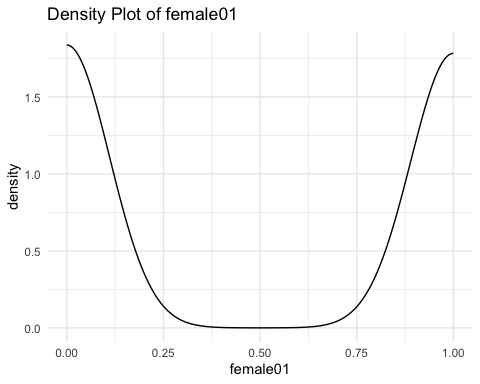
Summary Statistics for fam\_movement :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 2.445031 1.0985 3 0 4  
T-Test Results for fam\_movement :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 75.052, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 2.381112 2.508950  
sample estimates:  
mean of x   
 2.445031



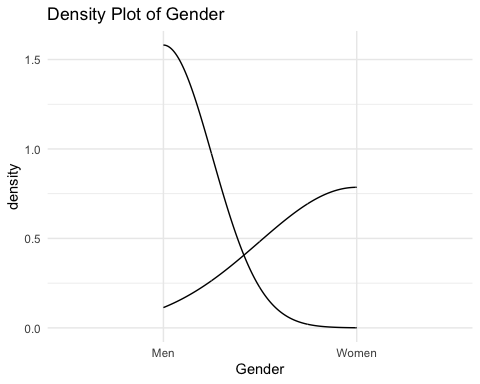
Summary Statistics for gender :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.4978012 0.505467 0 0 2  
T-Test Results for gender :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 33.208, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.4683893 0.5272132  
sample estimates:  
mean of x   
0.4978012



Summary Statistics for female01 :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.4925242 0.5001641 0 0 1  
T-Test Results for female01 :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 33.204, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.4634208 0.5216276  
sample estimates:  
mean of x   
0.4925242



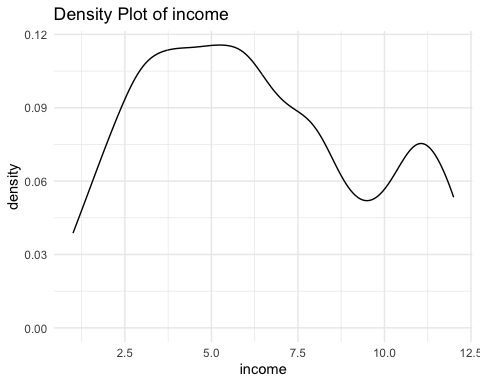
Summary Statistics for Gender :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Men Men Women  
Insufficient variability in the data or missing values. Skipping t-test for Gender



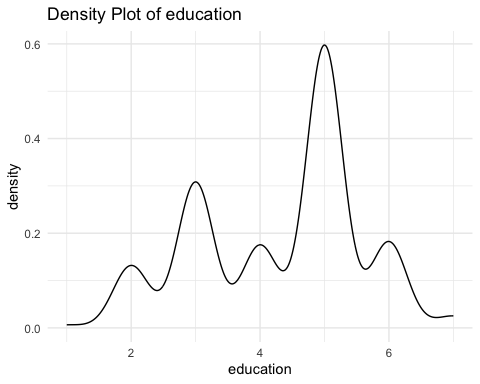
Summary Statistics for age :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 38.11873 11.69104 35 18 82  
T-Test Results for age :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 109.94, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 37.43846 38.79901  
sample estimates:  
mean of x   
 38.11873



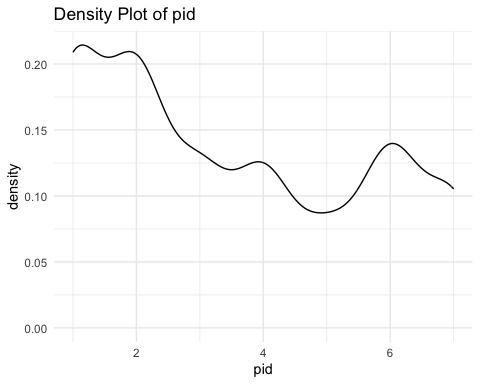
Summary Statistics for income :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 6.226415 3.127239 6 1 12  
T-Test Results for income :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 66.424, df = 1112, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 6.042493 6.410337  
sample estimates:  
mean of x   
 6.226415



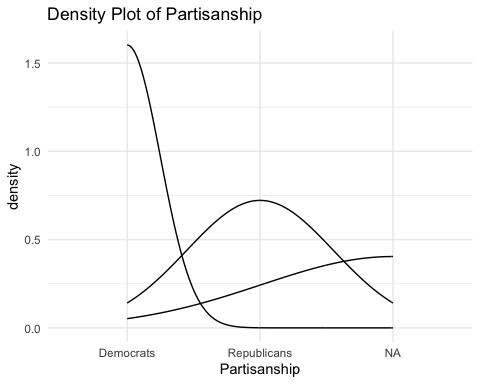
Summary Statistics for education :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 4.313105 1.271483 5 1 7  
T-Test Results for education :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 114.38, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 4.239120 4.387089  
sample estimates:  
mean of x   
 4.313105



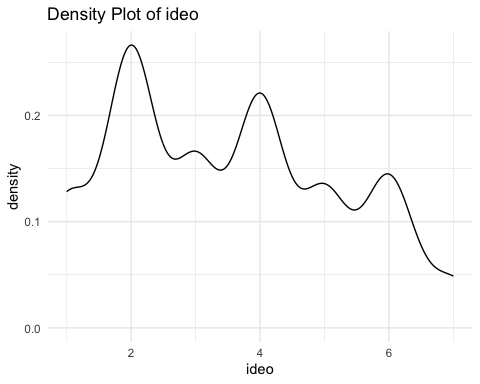
Summary Statistics for pid :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 3.488889 2.071049 3 1 7  
T-Test Results for pid :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 56.503, df = 1124, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 3.367737 3.610041  
sample estimates:  
mean of x   
 3.488889



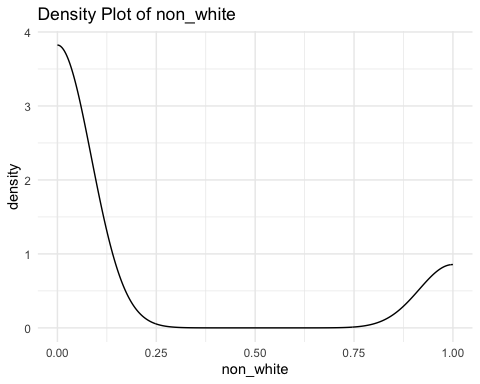
Summary Statistics for Partisanship :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA NA Democrats Republicans  
Insufficient variability in the data or missing values. Skipping t-test for Partisanship



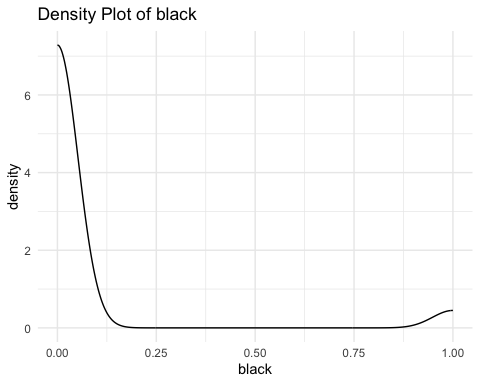
Summary Statistics for ideo :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 3.532982 1.723137 3 1 7  
T-Test Results for ideo :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 69.136, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 3.432716 3.633247  
sample estimates:  
mean of x   
 3.532982



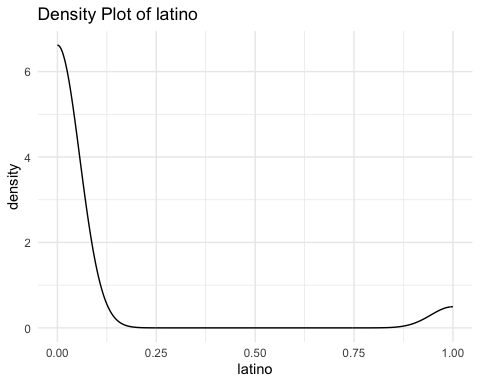
Summary Statistics for non\_white :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.1829376 0.3867854 0 0 1  
T-Test Results for non\_white :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 15.948, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.1604314 0.2054437  
sample estimates:  
mean of x   
0.1829376



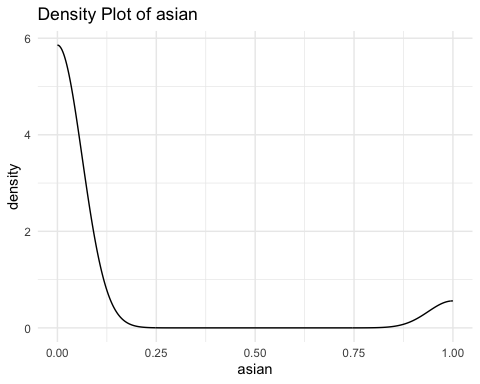
Summary Statistics for black :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.05804749 0.2339361 0 0 1  
T-Test Results for black :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 8.3669, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.04443530 0.07165969  
sample estimates:  
 mean of x   
0.05804749



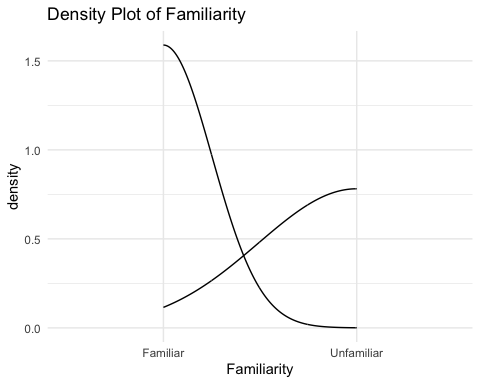
Summary Statistics for latino :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.06948109 0.2543824 0 0 1  
T-Test Results for latino :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 9.21, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.05467918 0.08428300  
sample estimates:  
 mean of x   
0.06948109



Summary Statistics for asian :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 0.08707124 0.2820635 0 0 1  
T-Test Results for asian :  
  
 One Sample t-test  
  
data: df\_mtg[[col]]  
t = 10.409, df = 1136, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 0.07065863 0.10348385  
sample estimates:  
 mean of x   
0.08707124



Summary Statistics for Familiarity :  
 mean\_variable sd\_variable median\_variable min\_variable max\_variable  
1 NA NA Familiar Familiar Unfamiliar  
Insufficient variability in the data or missing values. Skipping t-test for Familiarity



function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Rest of the function code...  
}

function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Rest of the function code...  
}

my\_function <- function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Function code...  
}

fig <- my\_function(df\_mtg, bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), comparison = c("Select Treatment", "Avoid Treatment"))

result <- my\_function(df\_mtg)

### 2.2. Methods

**250 words**

Describe what methods were used in the original paper and what methods you used and why.

Support any changes you made in the analytical strategy with literature and/or displaying any relevant data diagnostics.

If suitable, embed statistical tests comparing model fit of both methods.

function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Rest of the function code...  
}

function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Rest of the function code...  
}

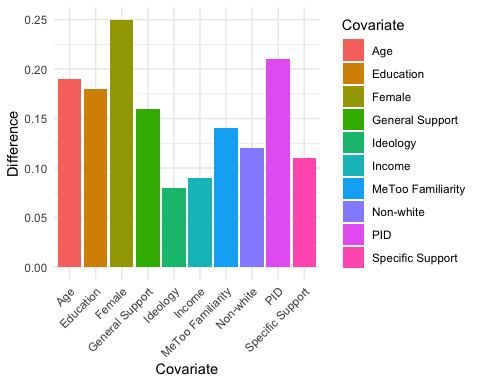
my\_function <- function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Function code...  
}

fig <- my\_function(df\_mtg, bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), comparison = c("Select Treatment", "Avoid Treatment"))

result <- my\_function(df\_mtg)

fig3 <- function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"),  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Descriptives Differences in Selecting Treatment Overall  
 bal\_gen <- data.frame(  
 rbind(  
 balance\_fn(the\_data = d, "gender", comparison),  
 balance\_fn(the\_data = d, "non\_white", comparison),  
 balance\_fn(the\_data = d, "education", comparison),  
 balance\_fn(the\_data = d, "income", comparison),  
 balance\_fn(the\_data = d, "pid", comparison),  
 balance\_fn(the\_data = d, "ideo", comparison),  
 balance\_fn(the\_data = d, "fam\_movement", comparison),  
 balance\_fn(the\_data = d, "age", comparison)  
 ))  
 bal\_gen$Covariate <- factor(bal\_labs[1:nrow(bal\_gen)], levels = rev(bal\_labs))  
 rownames(bal\_gen) <- bal\_gen$Covariate  
   
 # Create data frame for plotting  
 fig\_df <- data.frame(  
 bal\_gen,  
 Group = "Overall",  
 Type = "Overall"  
 )  
   
 # Set labels as factors for plotting  
 fig\_df$Type <- factor(fig\_df$Type, levels = unique(fig\_df$Type))  
 fig\_df$Group <- factor(fig\_df$Group, levels = unique(fig\_df$Group))  
   
 # Create Figure  
 fig <- fig\_df %>%   
 ggplot(aes(Covariate, Difference, col = Group, shape = Group)) +  
 geom\_hline(yintercept = 0, linetype = "dashed", alpha = 0.5) +  
 geom\_point(aes(shape = Group), position = position\_dodge(width = 0.5), size = 2) +  
 geom\_linerange(aes(ymin = ll, ymax = ul), size = 0.3, position = position\_dodge(width = 0.5)) +  
 geom\_linerange(aes(ymin = ll90, ymax = ul90), size = 0.6, position = position\_dodge(width = 0.5)) +  
 facet\_wrap(~Type, ncol = 3, drop = TRUE) +  
 theme(  
 axis.text.x = element\_text(angle = 0, hjust = 1),  
 panel.grid.minor = element\_blank(),  
 legend.position = "bottom"  
 ) +  
 ylab("Difference\n(Treatment Selectors - Treatment Avoiders)") +  
 coord\_flip() +  
 scale\_color\_grey(start = 0, end = 0.75) +  
 scale\_shape\_manual(values = c(16, 17, 15, 23, 4))  
   
 # Display Figure  
 print(fig)  
}

library(ggplot2)  
  
# Create a bar plot using ggplot2 with clearer x-axis labels  
graph\_data <- data.frame(  
 Covariate = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"),  
 Difference = c(0.25, 0.12, 0.18, 0.09, 0.21, 0.08, 0.14, 0.11, 0.16, 0.19)  
)  
  
graph <- ggplot(graph\_data, aes(x = Covariate, y = Difference, fill = Covariate)) +  
 geom\_bar(stat = "identity") +  
 labs(x = "Covariate", y = "Difference") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1, vjust = 1))  
  
graph



## 3. Results

**500 words**

Display and describe your research results.

Show R code and professionally looking outputs, e.g. tables prepared in tab\_model or stargazer functions, or graphs as coefplots or predicted values.

Provide in-depth discussion and interpretation of your results.

## 4. Conclusions

**250 words**

Discuss you findings in the light of literature you introduced in the introduction.

Reflect why your results are similar/different from the original study.

## References

**min. 5 scientific references, excluding the replicated paper**

Freese, J., & Peterson, D. (2017). Replication in social science. *Annual Review of Sociology*, 43, 147-165, [doi: 10.1146](https://www.annualreviews.org/doi/abs/10.1146/annurev-soc-060116-053450).

## Appendix

### Appendix 1. My enviroment (full information)

# Detailed information about my environment  
sessionInfo()

R version 4.3.0 (2023-04-21)  
Platform: aarch64-apple-darwin20 (64-bit)  
Running under: macOS Ventura 13.2.1  
  
Matrix products: default  
BLAS: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib   
LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib; LAPACK version 3.11.0  
  
locale:  
[1] en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8  
  
time zone: Europe/London  
tzcode source: internal  
  
attached base packages:  
[1] stats graphics grDevices utils datasets methods base   
  
other attached packages:  
 [1] sessioninfo\_1.2.2 Hmisc\_5.1-0 car\_3.1-2 carData\_3.0-5   
 [5] lubridate\_1.9.2 forcats\_1.0.0 stringr\_1.5.0 dplyr\_1.1.2   
 [9] purrr\_1.0.1 readr\_2.1.4 tidyr\_1.3.0 tibble\_3.2.1   
[13] ggplot2\_3.4.2 tidyverse\_2.0.0 knitr\_1.42 rmarkdown\_2.21   
  
loaded via a namespace (and not attached):  
 [1] utf8\_1.2.3 generics\_0.1.3 stringi\_1.7.12 hms\_1.1.3   
 [5] digest\_0.6.31 magrittr\_2.0.3 evaluate\_0.21 grid\_4.3.0   
 [9] timechange\_0.2.0 fastmap\_1.1.1 backports\_1.4.1 nnet\_7.3-19   
[13] Formula\_1.2-5 gridExtra\_2.3 fansi\_1.0.4 scales\_1.2.1   
[17] codetools\_0.2-19 abind\_1.4-5 cli\_3.6.1 rlang\_1.1.1   
[21] munsell\_0.5.0 base64enc\_0.1-3 withr\_2.5.0 yaml\_2.3.7   
[25] tools\_4.3.0 tzdb\_0.4.0 checkmate\_2.2.0 htmlTable\_2.4.1   
[29] colorspace\_2.1-0 vctrs\_0.6.2 R6\_2.5.1 rpart\_4.1.19   
[33] lifecycle\_1.0.3 htmlwidgets\_1.6.2 foreign\_0.8-84 cluster\_2.1.4   
[37] pkgconfig\_2.0.3 pillar\_1.9.0 gtable\_0.3.3 data.table\_1.14.8  
[41] glue\_1.6.2 xfun\_0.39 tidyselect\_1.2.0 rstudioapi\_0.14   
[45] htmltools\_0.5.5 compiler\_4.3.0

### Appendix 2. Entire R code used in the project

# Opening key libraries first  
library(rmarkdown)  
library(knitr)  
# Global options  
opts\_chunk$set(echo=TRUE,  
 cache=TRUE,  
 comment=NA,  
 message=FALSE,  
 warning=FALSE)  
# All used libraries  
library("rmarkdown")  
library("knitr")  
library("tidyverse")  
library("car")  
library("Hmisc")  
library("sessioninfo")  
library("ggplot2")  
library("dplyr")  
# Versions of used packages  
packages <- c("rmarkdown", "knitr")  
names(packages) <- packages  
lapply(packages, packageVersion)  
# What is my R version?  
version[['R version 4.3.0 (2023-04-21)']]  
load("/Users/georgiamugglestone/Documents/smi 205/Assessments /200382018 2/dataverse\_files-2/df\_mtg.rda")  
summary\_stats <- df\_mtg %>%  
 summarise(  
 mean\_variable = mean(gender),  
 sd\_variable = sd(gender),  
 median\_variable = median(gender),  
 min\_variable = min(gender),  
 max\_variable = max(gender)  
 )  
  
# Display the summary statistics  
print(summary\_stats)  
# Create Figure 3 with statistical analysis  
fig3 <- df\_mtg %>%  
 ggplot(aes(x = gender)) +  
 geom\_density() +  
 ggtitle("Density Plot of Variable") +  
 theme\_minimal()  
  
# Display Figure 3  
fig3  
# Perform statistical analysis and create plots for all variables  
  
# Loop over each variable  
for (col in names(df\_mtg)) {  
 # Calculate summary statistics  
 summary\_stats <- df\_mtg %>%  
 summarise(  
 mean\_variable = mean(!!sym(col)),  
 sd\_variable = sd(!!sym(col)),  
 median\_variable = median(!!sym(col)),  
 min\_variable = min(!!sym(col)),  
 max\_variable = max(!!sym(col))  
 )  
   
 # Display the summary statistics  
 cat("Summary Statistics for", col, ":\n")  
 print(summary\_stats)  
   
 # Create plot for the variable  
 fig <- df\_mtg %>%  
 ggplot(aes(x = !!sym(col))) +  
 geom\_density() +  
 ggtitle(paste("Density Plot of", col)) +  
 theme\_minimal()  
   
 # Display the plot  
 print(fig)  
}  
# Perform statistical analysis and create plots for all variables  
  
# Loop over each variable  
for (col in names(df\_mtg)) {  
 # Calculate summary statistics  
 summary\_stats <- df\_mtg %>%  
 summarise(  
 mean\_variable = mean(!!sym(col), na.rm = TRUE),  
 sd\_variable = sd(!!sym(col), na.rm = TRUE),  
 median\_variable = median(!!sym(col), na.rm = TRUE),  
 min\_variable = min(!!sym(col), na.rm = TRUE),  
 max\_variable = max(!!sym(col), na.rm = TRUE)  
 )  
   
 # Display the summary statistics  
 cat("Summary Statistics for", col, ":\n")  
 print(summary\_stats)  
   
 # Check if there is sufficient variability in the data  
 if (!is.na(summary\_stats$sd\_variable) && summary\_stats$sd\_variable > 0.001) {  
 # Perform t-test  
 ttest\_result <- t.test(df\_mtg[[col]], mu = 0, na.rm = TRUE)  
   
 # Display the t-test results  
 cat("T-Test Results for", col, ":\n")  
 print(ttest\_result)  
 } else {  
 cat("Insufficient variability in the data or missing values. Skipping t-test for", col, "\n")  
 }  
   
 # Create plot for the variable  
 fig <- df\_mtg %>%  
 ggplot(aes(x = !!sym(col))) +  
 geom\_density() +  
 ggtitle(paste("Density Plot of", col)) +  
 theme\_minimal()  
   
 # Display the plot  
 print(fig)  
}  
function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Rest of the function code...  
}  
my\_function <- function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Function code...  
}  
fig <- my\_function(df\_mtg, bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), comparison = c("Select Treatment", "Avoid Treatment"))  
result <- my\_function(df\_mtg)  
function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Rest of the function code...  
}  
my\_function <- function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), # Added "Age" to the covariate labels  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Function code...  
}  
fig <- my\_function(df\_mtg, bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"), comparison = c("Select Treatment", "Avoid Treatment"))  
result <- my\_function(df\_mtg)  
fig3 <- function(d,  
 bal\_labs = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"),  
 comparison = c("Select Treatment", "Avoid Treatment")  
) {  
 # Descriptives Differences in Selecting Treatment Overall  
 bal\_gen <- data.frame(  
 rbind(  
 balance\_fn(the\_data = d, "gender", comparison),  
 balance\_fn(the\_data = d, "non\_white", comparison),  
 balance\_fn(the\_data = d, "education", comparison),  
 balance\_fn(the\_data = d, "income", comparison),  
 balance\_fn(the\_data = d, "pid", comparison),  
 balance\_fn(the\_data = d, "ideo", comparison),  
 balance\_fn(the\_data = d, "fam\_movement", comparison),  
 balance\_fn(the\_data = d, "age", comparison)  
 ))  
 bal\_gen$Covariate <- factor(bal\_labs[1:nrow(bal\_gen)], levels = rev(bal\_labs))  
 rownames(bal\_gen) <- bal\_gen$Covariate  
   
 # Create data frame for plotting  
 fig\_df <- data.frame(  
 bal\_gen,  
 Group = "Overall",  
 Type = "Overall"  
 )  
   
 # Set labels as factors for plotting  
 fig\_df$Type <- factor(fig\_df$Type, levels = unique(fig\_df$Type))  
 fig\_df$Group <- factor(fig\_df$Group, levels = unique(fig\_df$Group))  
   
 # Create Figure  
 fig <- fig\_df %>%   
 ggplot(aes(Covariate, Difference, col = Group, shape = Group)) +  
 geom\_hline(yintercept = 0, linetype = "dashed", alpha = 0.5) +  
 geom\_point(aes(shape = Group), position = position\_dodge(width = 0.5), size = 2) +  
 geom\_linerange(aes(ymin = ll, ymax = ul), size = 0.3, position = position\_dodge(width = 0.5)) +  
 geom\_linerange(aes(ymin = ll90, ymax = ul90), size = 0.6, position = position\_dodge(width = 0.5)) +  
 facet\_wrap(~Type, ncol = 3, drop = TRUE) +  
 theme(  
 axis.text.x = element\_text(angle = 0, hjust = 1),  
 panel.grid.minor = element\_blank(),  
 legend.position = "bottom"  
 ) +  
 ylab("Difference\n(Treatment Selectors - Treatment Avoiders)") +  
 coord\_flip() +  
 scale\_color\_grey(start = 0, end = 0.75) +  
 scale\_shape\_manual(values = c(16, 17, 15, 23, 4))  
   
 # Display Figure  
 print(fig)  
}  
library(ggplot2)  
  
# Create a bar plot using ggplot2 with clearer x-axis labels  
graph\_data <- data.frame(  
 Covariate = c("Female", "Non-white", "Education", "Income", "PID", "Ideology", "MeToo Familiarity", "Specific Support", "General Support", "Age"),  
 Difference = c(0.25, 0.12, 0.18, 0.09, 0.21, 0.08, 0.14, 0.11, 0.16, 0.19)  
)  
  
graph <- ggplot(graph\_data, aes(x = Covariate, y = Difference, fill = Covariate)) +  
 geom\_bar(stat = "identity") +  
 labs(x = "Covariate", y = "Difference") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1, vjust = 1))  
  
graph  
# Detailed information about my environment  
sessionInfo()