

# An Analysis of the Effects of The 2007-2009 Financial Crisis on Suicide Rates in the United States

Paul A. Hodgetts

25/03/2020

## Abstract

Using a differences in differences approach, the effects of the 2007-2009 Financial Crisis on suicide rates in people of investing age was examined. An initial analysis revealed no significant effects, with gender and age accounting for variation in suicide rates. A differences in differences in differences model revealed a minorly significant correlation between suicides for males of investing age and the Financial Crisis. Limitations and ethical concerns are discussed regarding model and dataset designs.

## Introduction

An apocryphal story of the stock market crash that occurred in 1929 and contributed to the Great Depression is that investment bankers began to throw themselves from rooftops and out windows. The idea being that the crash caused their futures to be so bleak, they would rather commit suicide than face what may come. In 2007, the United States housing market bubble burst, leading to the markets crashing in 2009 (Amadeo, 2019). This caused a global economic crisis that has come to be known as the Great Recession, with some economists calling it the worst economic crisis since the Great Depression (Pendery, 2009). Using data on global suicide rates from between 1985 to 2016, available from Kaggle.com, this paper examines the effects of an economic crisis on suicide rates in the United States. Using a differences in differences approach, this paper examines the effects of the Great Recession on suicide rates between people of investing age and non-investing age. The assumption being that people who are reliant upon investments for cash flow and financial support would feel the effects of the 2007-2009 Financial Crisis more so than people who have no personal cash flow, are dependent upon others for finances, or are not of an age where investments are financially significant. Furthermore, as completed suicide rates for men have historically been higher than those of women (Moscicki, 1994) a differences in differences in differences analysis was made of the effects of the 2007-2009 Financial Crisis on suicide rates in males of investing age. A linear regression revealed the differences in differences model showed no significant difference in suicide rates between people of investing age and non-investing age before and during the 2007-2009 Financial Crisis. However, the differences in differences in differences model revealed a significant correlation between the Financial Crisis and suicide rates in males of investing age. However, the validity of the model is questioned, and the effect found may be better explained by age and gender differences.

## Research Question

The question this paper seeks to answer is, does an economic crisis increase the suicide rate for individuals who rely on investments for financial support when compared to individuals who are less reliant or do not rely on investments for financial support?

## Dataset

The dataset used in this paper is the 'Suicide Rates Overview 1985 to 2016' available from Kaggle.com and uploaded by the user 'Rusty'. This dataset contains data on country, year, sex (gender), age, total number of suicides for a country, population, suicides per 100,000 people, Human Development Index rating (HDI), Gross Domestic Product (GDP) for the year, GDP per capita, and generation name. Age groups were designated by the dataset uploader as 5-14 years, 15-24 years, 25-34 years, 35-54 years, 55-74 years, and 75+ years.

## Data Cleaning and Structure

For the purposes of this analysis, only country, year, sex (relabelled as gender), age, suicides per 100,000 people, and GDP per capita were used. Sex was relabelled to gender as the label sex is presumptive and exclusionary. Both country and GDP per capita were renamed to clean the original column names. For simplicity and speed of analysis, country data was filtered to only retain data referring to the United States. The range of years for analysis was chosen to be 2004 to 2009. This was done for two reasons. First, the Financial Crisis occurred within the United States between 2007 and 2009, with GDP per capita reaching its lowest in 2009 and climbing back to pre-recession levels in 2010 (see Figure 1). Second, the dot-com bubble bursting in the early 2000's may have influenced investor behaviour. As the burst of the dot-com bubble could possibly confound data from the 2007-2009 Financial Crisis, a period of relative economic stability was necessary to compare to the economic turmoil of the Financial Crisis. As population of a country can change year to year, a normalized rate of suicide was required. As such, suicides per 100,000 people was used over total number of suicides. An analysis on the created United States dataset using the skim function from the skimr package revealed no missing values, nor any duplicate values (see Table 1). Additionally, as can be seen in Table 1, suicides per 100,000 people (*suicides.100k.pop*) has a mean value of  $M = 12.43$  with a standard deviation of  $SD = 11.30$ . Furthermore, the data summary in Table 1 shows that suicides per 100,000 people has a range of 0.27 to 37.92. A range that can be explained by the differences in suicides between the age groups 5-14 years and 75+ years. With the former consistently experiencing close to 0 suicides across all years, and the latter experiencing close to 40 suicides over the years (see Figure 2).

Data summary	
Name	suicides_us
Number of rows	72
Number of columns	11
Column type frequency:	
factor	3
numeric	3
Group variables	None

### Variable type: factor

skim_variable	n_missing
country	0
gender	0
age	0

### Variable type: numeric

skim_variable	n_missing	mean	sd	p0	p25	p50	p75	p100
year	0	2006.50	1.72	2004.00	2005.00	2006.50	2008.00	2009.00
suicides.100k.pop	0	12.43	11.30	0.27	3.59	6.89	21.58	37.92
gdp_per_capita	0	49372.00	2537.32	44867.00	47423.00	50114.50	51585.00	52128.00

Table 1. Summary Statistics of United States Suicides, 2004 - 2009.

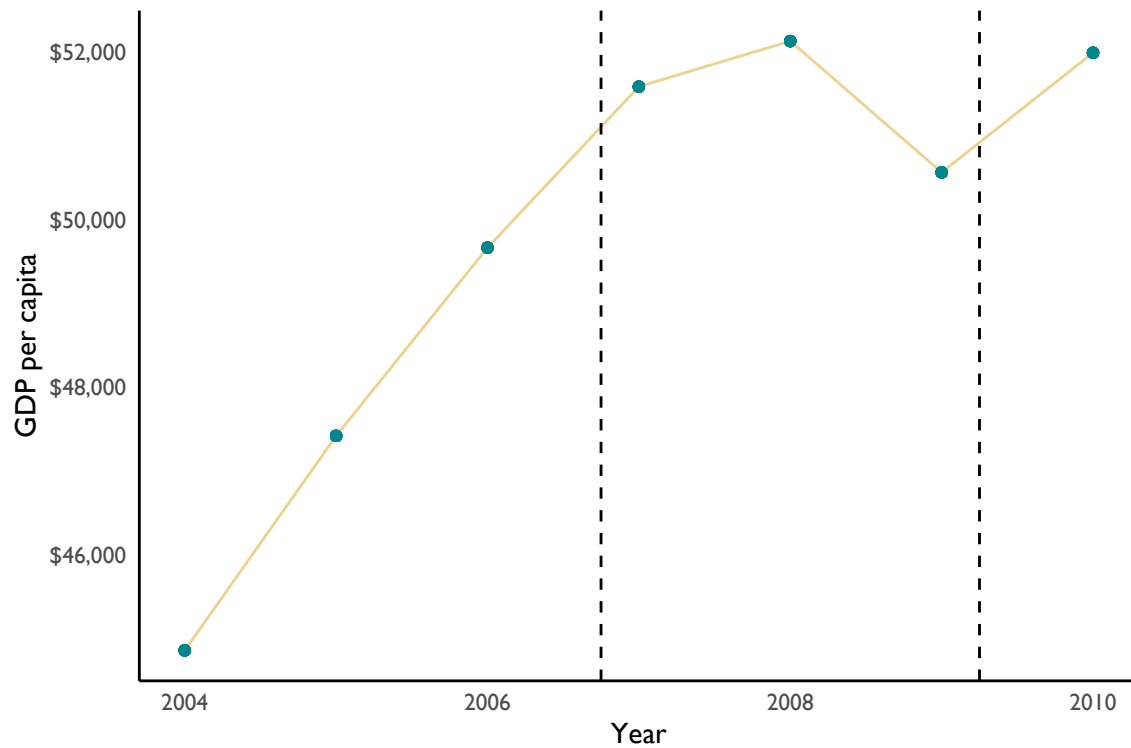


Figure 1. United States GDP per Capita, 2004 - 2010.

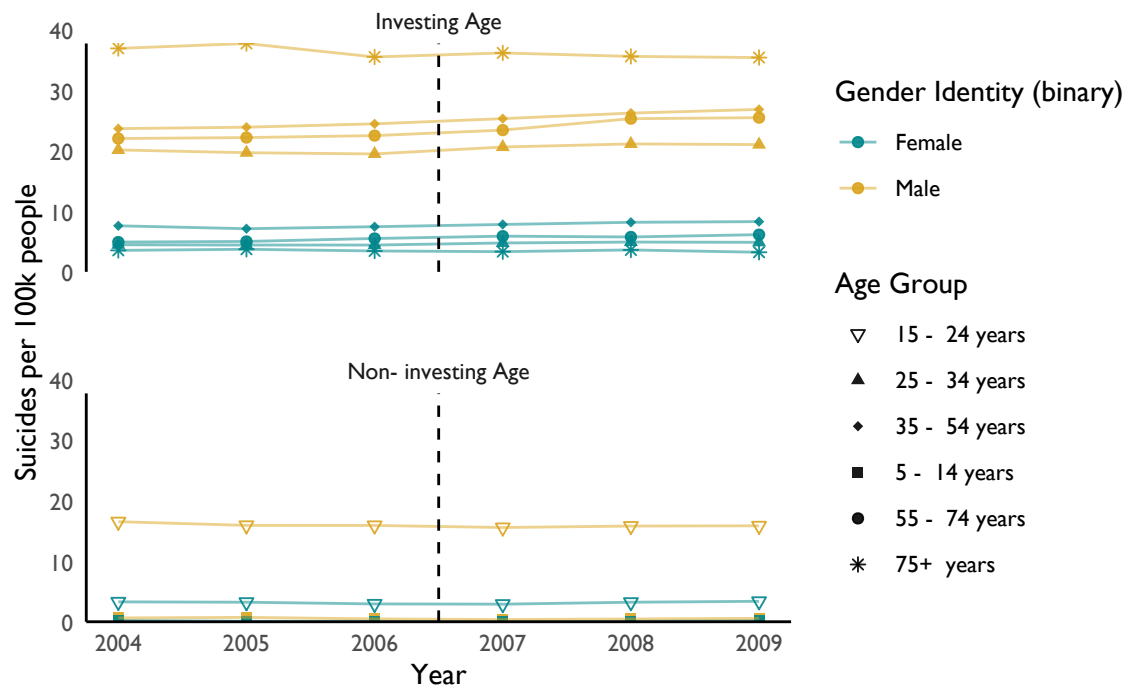


Figure 2. United States Suicides per 100k People, 2004 - 2009, by Investing Age.

## Limitations of the Data

The biggest limitation to the dataset is the designation of the age groups. The rigidity of these groups does not lend itself to nuanced analysis. There is no designation of which data points correspond to a specific age within an age group, so rearrangement of age groups cannot be done. This means the designation of the age groups as they are for the investing age and non-investing age groups is possibly confounded by spill-over. Not all individuals who were included in the non-investing age group would be completely devoid of the influence of investment. As such, the separation of the groups is not exact. A further limitation of the dataset is the binary designation of the gender variable. This is limited in the sense that it is not representative of the population and confines those people to a label to which they may not subscribe. In general, this limitation is indicative that the dataset does not lend itself to a more specific analysis of the individual.

## Model

### Dummy Variable Creation

Four binary dummy variables were created for this analysis. An investing age variable, a recession year variable, a post recession variable, and a post recession males variable. Investing age was defined by the age groups 25-34 years, 35-54 years, 55-74 years, and 75+ years. Non-investing age was defined by the age groups 5-14 years and 15-24 years. The assumption being that the investing age group was made up of individuals of an age who generally had full employment, had built up a stock portfolio, and were dependent upon that portfolio for financial support. Whereas, the non-investing age group was composed of people of an age who were in school, too young to be fully employed, and were not dependent upon stocks for financial support. This variable was created logically filtering for the chosen age groups and assigning investing age groups as 1 and non-investing age groups as 0. Differences in suicides per 100,000 people between investing age groups can be seen in *Figure 2*. The separation between pre- and post-recession was specified as 2007, as this is the first year of the Financial Crisis. As such, pre-recession years were designated as 2004, 2005, 2006, while post-recession years were 2007, 2008, and 2009. The year 2009 was chosen as the cut-off as this marks the lowest point in GDP per capita and the financial crisis was considered to have ended in the United States in 2010. This provided three lead-ins and two lags. This variable was created by logically filtering for and designating years greater than and equal to 2007 with a 1 and years less than 2007 with a 0. The post-recession variable was created using by multiplying investing age by recession year. In this way, whenever both investing age and recession year were equal to 1, the result for post-recession would be a 1. Whenever either investing age or recession year were 0, post-recession would equal 0. Gender was factorized by designating all male values with a 1 and all female values with a 0. A new post-recession gender variable was created by multiplying the values in the factorized gender variable by investing age and recession year. Differences in suicides per 100,000 people between the genders for pre- and post-recession years can be seen in *Figure 3*.

### Formulas and Regression Models

Two models were made for this analysis. The first model is a differences in differences model regressing suicides per 100,000 people onto post-recession and controlling for investing age, year, and gender. The formula for this model can be seen below, with  $i$  being suicides per 100,000 people of an investing age group pre- or post-recession,  $a$  being an investing age constant,  $t$  being a time constant, and  $g$  being a gender constant. The main focus of this equation is  $\beta_1(A \times R)$ , with this being the slope of investing age ( $A$ ) by post-recession ( $R$ ).

$$Y_{iatg} = \beta_0 + \beta_1(A * R) + \lambda_a + \tau_t + \gamma_g + \epsilon$$

The second model is a differences in differences in differences model regressing suicides per 100,000 people onto post-recession males of investing age and controlling for investing age, year, and gender. The formula for this model can be seen below. It is similar to the previous formula except for the inclusion of the variable  $j$  which represents the category of interest (i.e. males of investing age post-recession). The focus of this formula is  $\beta_1(A \times R \times G)$  with  $G$  representing the gender factor variable.

$$Y_{ijatg} = \beta_0 + \beta_1(A * R * G) + \lambda_a + \tau_t + \gamma_g + \epsilon$$

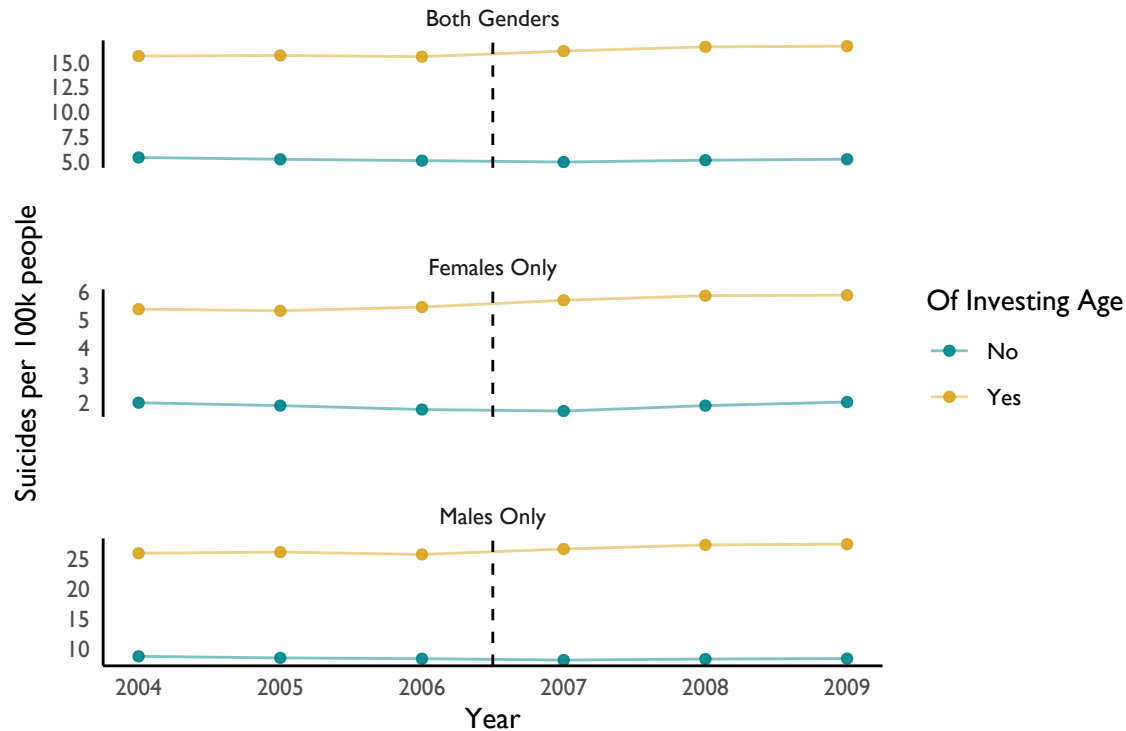


Figure 3. United States Suicides per 100k People, 2004 - 2009, by Gender.

### Testing Assumptions

**Differences in Differences** The main assumption of a differences in differences approach is the parallel trends assumption. This being, in the case of no intervention the trends experienced before intervention would continue as is. Looking at *Figure 3*, the pre-recession years give little to question that those trends would not have continued given no financial crisis. Whereas looking at *Figure 2*, the trend of suicides for 75+ over the years is variable enough to question how that would continue without intervention; however, all other age groups appear to show consistency over the years that support the assumption of this trend.

**OLS Regression** A VIF test was performed to test for multicollinearity of independent variable for both models. All independent variables were revealed to have values below 5, indicating that there was little collinearity between the independent variables. To test the assumption of heteroscedacity, a Breusch-Pagan test was run on both the differences in differences and the differences in differences in differences models. Both models were found to have *p-values* of  $p < .01$  indicating issues of heteroscedacity. Although an examination of the normality of the data showed the data followed a normal distribution. However, an examination of linearity revealed a pattern in the data. Suggesting issues of non-linearity of the model. Given that not all assumptions were met, the validity of the models was brought to question.

### Results

Regression results in *Table 2* show that although there is a positive correlation between suicides and the Great Recession ( $\beta_1 = 0.79$ ), this correlation is not significant. That any variation in suicide rates can be accounted for by age ( $\beta_2 = 10.47, p < .001$ ) and gender ( $\beta_3 = 16.12, p < .001$ ). It can also be seen in *Table 2* that the differences in differences in differences model revealed a minorly significant correlation between males of investing age post-recession and suicide rates ( $\beta_1 = 6.08, p < .01$ ). Although this is can again be accounted for by the differences in suicide rates by age ( $\beta_2 = 9.34, p < .001$ ) and gender ( $\beta_3 = 14.09, p < .001$ ) moreso than the effects of the Great Recession. Indeed, looking at *Figure 3* it can be seen that males in the 75+ years age group have a higher rate of suicide than the other age groups, and males in general have a higher rate of suicide. Thereby, skewing the results of the DDD model. It is important to note that although a correlation was revealed that does not mean the financial crisis was the cause of the increase in suicides.

	Suicides After Recession	Suicides DDD
Post Recession	0.79 (2.53)	
Post Recession Male		6.08 * (2.50)
Investing Age	10.47 *** (1.98)	9.34 *** (1.59)
Year	0.01 (0.60)	-0.38 (0.46)
Male Suicides	16.12 *** (1.44)	14.09 *** (1.61)
N	72	72
R2	0.73	0.75
logLik	-229.73	-226.75
AIC	471.45	465.49

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ .

Table 2. Regression Table of Recession DD and DDD Models.

### Limitation of Model Approach

The main limitation of this approach is that the Financial Crisis was a global economic crisis. It is difficult to say that any one person or age was not affected in some way. A second limitation of this approach is its use of age. Firstly, it assumes dependency on investments. It may entirely be the case that people within the investing age group are not as dependent upon investments as people in the non-investing age group. Not everyone has the capital to build a stock portfolio. Secondly, the 5-14 years and 75+ years age groups both being on extreme ends of the range could possibly skew the results. That 75+ years age group experiences a rate of suicides beyond any other group may skew the results towards the effect of age on suicide. Especially when compared to a group such as 5-14 years. An option would be to remove the 75+ years age group, although this removes a group that could possibly be highly dependent upon investments for financial support. Whereas removing the 5-14 years age group would remove one of two age groups in the comparison group. The ultimate solution would be to restructure the age groups.

### Ethics

Regarding ethics, this analysis deals with incredibly sensitive data. These are quite literally peoples lives that are being examined. It is one reason that the label sex was relabelled as gender and the variable suicides.100k.pop was referred to as people instead of population. It is important to recognize the people represented by these numbers. However, while it is important to understand what contributes to a person's decision to take their own life, this paper does nothing to suggest intervention or treatment for individuals considering suicide. The paper does reveal concerning trends regarding suicide rates in the 75+ age group and in the male gender designation; however, given that the data examined is over a decade old, these trends may not be representative of current trends. Although the trends are concerning, little can be drawn from them in terms of current implementation of suicide intervention initiatives. Furthermore, even though there is some indication of a correlation between economic crisis and male suicides, it cannot be said that economic crisis is the cause of this increase in suicides. In actuality greater concern should be given to understanding the greater correlation between age and suicide as well as gender and suicide. Secondly, while this paper examines the possible consequences of an economic crisis on suicide rates, it does not consider other factors that may lead to an individual's decision to take their life. This is partially due to the design of the dataset not lending itself to a nuanced examination of the reasons why a person committed suicide, nor the socio-economic resources available to that person. It provides a general view of the situation with little to no indication of individual factors. There is little possibility to introduce intervention measures that pertain to the individual. And while lumping individuals into age groups may provide an easier means of statistical analysis, the individual is almost wholly removed. Making their death a little more tragic. Lastly, while the purpose of the stated dataset is suicide

prevention, no outreach resources are provided by the dataset uploader in this regard nor does the dataset lend itself to the analysis of preventative measures. At best, the dataset allows for the analysis of global suicide rates by economic conditions for a country. Which is a very limited scope of the conditions that contribute to an individual's decision to commit suicide. It would be pertinent to at least provide a list of community resources by country on the dataset description page. A differences in differences approach was selected for this analysis as it provided the best means of comparison of the effects of an economic crisis on suicide rates. The nature of the dataset did not lend itself to randomized control nor A/B testing. Furthermore, matching would exclude variables of interest such as the extremes of the 75+ years age group.

## **Conclusion**

Suicide is a tragedy that is often seen as the last option for those who chose it, and leaves those closest wondering if there was something more they could have done. Understanding conditions that influence an individual's decision to commit suicide is important to preventing future suicides. This paper used data on global suicide rates from 1985 to 2015 available from Kaggle.com to better understand the effects of an economic crisis on suicide rates. Using a differences in differences approach, suicide rates in the United States from before and during the 2007-2009 Financial Crisis were examined between people of investing age and non-investing age. A secondary analysis using a differences in differences in differences approach was also performed, examining the effects of the Great Recession on suicide rates for males of investing age. For the purposes of this analysis, and due to limitations of the design of the dataset, investing age was defined as age groups which would consist of individuals who felt the financial burden of the Financial Crisis more so than those people of non-investing ages. A linear regression model of people of investing age post-recession showed there was no significant correlation between the Financial Crisis and suicide rates, with any variation in suicide rates being accounted for by age or gender. A linear regression model of males of investing age post-recession shows a minorly significant correlation between the effects of the 2007-2009 Financial Crisis and suicide rates. However, this effect was again better accounted for by age or gender alone. Validity of the models is brought to question regarding dataset design limitations and not all assumptions being met. Ethical concerns are that there is no greater understanding of the conditions that lead to an individual committing suicide, and an understanding of the individual is lost in the design of the dataset.

## References

- Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. *R News* 2(3), 7-10. URL <https://CRAN.R-project.org/doc/Rnews/>
- Amadeo, K (2019, November, 20). *2009 Financial Crisis Explanation with Timeline*. The Balance. <https://www.thebalance.com/2009-financial-crisis-bailouts-3305539>
- David Hugh-Jones (2020). *huxtable: Easily Create and Style Tables for LaTeX, HTML and Other Formats*. R package version 4.7.1. <https://CRAN.R-project.org/package=huxtable>
- David Robinson and Alex Hayes (2020). *broom: Convert Statistical Analysis Objects into Tidy Tibbles*. R package version 0.5.5. <https://CRAN.R-project.org/package=broom>
- Elin Waring, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis (2020). *skimr: Compact and Flexible Summaries of Data*. R package version 2.1. <https://CRAN.R-project.org/package=skimr>
- H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- Hadley Wickham and Dana Seidel (2019). *scales: Scale Functions for Visualization*. R package version 1.1.0. <https://CRAN.R-project.org/package=scales>
- Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.5. <https://CRAN.R-project.org/package=dplyr>
- John Fox and Sanford Weisberg (2019). *An {R} Companion to Applied Regression, Third Edition*. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Kirill Müller (2017). *here: A Simpler Way to Find Your Files*. R package version 0.1. <https://CRAN.R-project.org/package=here>
- Max Kuhn (2020). *caret: Classification and Regression Training*. R package version 6.0-86. <https://CRAN.R-project.org/package=caret>
- Moscicki, E. K. (1994). Gender differences in completed and attempted suicides. *Annals of Epidemiology*, (4)2, 152-158.
- Pendery, D. (2009, February 13). *Three Top Economists Agree 2009 Worst Financial Crisis Since Great Depression; Risks Increase if Right Steps are Not Taken*. Businesswire. <https://www.businesswire.com/news/home/20090213005161/en/Top-Economists-Agree-2009-Worst-Financial-Crisis>
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sam Firke (2020). *janitor: Simple Tools for Examining and Cleaning Dirty Data*. R package version 1.2.1. <https://CRAN.R-project.org/package=janitor>
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Winston Chang, (2014). *extrafont: Tools for using fonts*. R package version 0.17. <https://CRAN.R-project.org/package=extrafont>



## Appendices

### Appendix A

```
knitr::opts_chunk$set(echo = TRUE, dev = "cairo_pdf")

library(here) # Load in library 'here' to make loading CSV file(s) easier.
library(tidyverse)
library(broom)
library(ggplot2)
library(janitor)
library(skimr)
library(huxtable)
library(scales)
library(dplyr)
library(extrafont)
library(car)
library(lmtest)
library(caret)

# Load in CSV file.
suicides_full <- read.csv(here("data", "suicide-rates-overview-1985-to-2016",
                              "master.csv"))

# Set up working dataset.
suicides_working <- suicides_full
# Apply full dataset to new, working dataset so original remains untouched.

# Build main dataset.
suicides_us <- suicides_working %>%
  select(1:4, 7, 11) %>% # Select only columns that will be used.
  rename(country = i..country) %>%
  # Odd symbol and column name could cause issues later.
  # Rename i..country to country to remove odd symbol.
  rename(gdp_per_capita = gdp_per_capita....) %>% # Rename odd column name.
  rename(gender = sex) %>% # Rename sex column to gender.
  filter(country == "United States") %>%
  # Filter so only United States remains and apply to new dataset.
  filter(year >= 2004 & year <= 2009) %>%
  # Filter the years so only those occurring at the end of the dotcom bubble
  # to a few years after the Great Recession are included.
  arrange(year, age) %>%
  # Arrange the dataset so that it is ordered first by year then by age group.
  mutate(investing_age = as.integer(age == '25-34 years' | age == '35-54 years' |
                                     age == '55-74 years' | age == '75+ years')) %>%
  # Create dummy binary variable for working age vs. non-working age.
  mutate(recession_year = as.integer(year >= 2007)) %>%
  # Create dummy binary variable for whether a year is pre recession or in/post
  # Great Recession.
  mutate(post_recession = as.integer(recession_year * investing_age)) %>%
  # Create dummy binary variable for working age and affected by the Great Recession.
  mutate(gender_factor = case_when(gender == 'male' ~ 1,
                                    gender == 'female' ~ 0)) %>%
```

```

# Create numeric binary variable for gender.
mutate(post_male = as.integer(recession_year * investing_age * gender_factor))
# Create dummy binary variable for males of working age affected by the Great Recession.

suicides_us_distinct = suicides_us %>% distinct()

# Build dataset for GDP.
suicides_us_gdp <- suicides_working %>%
  select(1:4, 7, 11) %>% # Select only columns that will be used.
  rename(country = i..country) %>%
  # Odd symbol and column name could cause issues later.
  #Rename i..country to country to remove odd symbol.
  rename(gdp_per_capita = gdp_per_capita....) %>% # Rename odd column name.
  rename(gender = sex) %>% # Rename sex column to gender.
  filter(country == "United States") %>% # Filter so only United States remains
  #and apply to new dataset.
  filter(year >= 2004 & year <= 2010)
  # Filter the years so only those occurring at the end of the dotcom bubble
  #to a few years after the Great Recession are included.

# Build datasets that average suicide rate by investing age.
suicides_noninvesting <- suicides_us %>%
  select(1:3, 5:11) %>% # Select only the necessary columns.
  filter(investing_age == "0") %>% # Filter data so only non-investing age data remain.
  group_by(year, investing_age) %>% # Group variables by year and investing age.
  summarise(avg_suicides.100k.pop = mean(suicides.100k.pop))
  # Summarise means of suicides.

suicides_investing<- suicides_us %>%
  select(1:3, 5:11) %>% # Select only the necessary columns.
  filter(investing_age == "1") %>% # Filter data so only investing age data remain.
  group_by(year, investing_age) %>% # Group variables by year and investing age.
  summarise(avg_suicides.100k.pop = mean(suicides.100k.pop))
  # Summarise means of suicides.

suicides_investing_combined <- rbind(suicides_noninvesting, suicides_investing) %>%
  # bind the two created datasets by row.
  mutate(investing_age = as.factor(investing_age)) %>%
  # Convert investing_age to factor variable type.
  mutate(gender = "combined")

# Build datasets that average suicide rate by investing age.
suicides_noninvesting_male <- suicides_us %>%
  select(1:3, 5:11) %>% # Select only the necessary columns.
  filter(investing_age == "0") %>% # Filter data so only non-investing age data remain.
  filter(gender == "male") %>% # Filter data so only male data remain.
  group_by(year, investing_age) %>% # Group variables by year and investing age.
  summarise(avg_suicides.100k.pop = mean(suicides.100k.pop))
  # Summarise means of suicides.

suicides_noninvesting_female <- suicides_us %>%
  select(1:3, 5:11) %>% # Select only the necessary columns.

```

```

filter(investing_age == "0") %>% # Filter data so only non-investing age data remain.
filter(gender == "female") %>% # Filter data so only female data remain.
group_by(year, investing_age) %>% # Group variables by year and investing age.
summarise(avg_suicides.100k.pop = mean(suicides.100k.pop))
# Summarise means of suicides.

suicides_investing_male<- suicides_us %>%
select(1:3, 5:11) %>% # Select only the necessary columns.
filter(investing_age == "1") %>% # Filter data so only investing age data remain.
filter(gender == "male") %>% # Filter data so only male data remain.
group_by(year, investing_age) %>% # Group variables by year and investing age.
summarise(avg_suicides.100k.pop = mean(suicides.100k.pop))
# Summarise means of suicides.

suicides_investing_female<- suicides_us %>%
select(1:3, 5:11) %>% # Select only the necessary columns.
filter(investing_age == "1") %>% # Filter data so only investing age data remain.
filter(gender == "female") %>% # Filter data so only female data remain.
group_by(year, investing_age) %>% # Group variables by year and investing age.
summarise(avg_suicides.100k.pop = mean(suicides.100k.pop))
# Summarise means of suicides.

suicides_investing_combined_male <- rbind(suicides_noninvesting_male,
                                           suicides_investing_male) %>%

# bind the two male datasets by row.
mutate(investing_age = as.factor(investing_age)) %>%
# Convert investing_age to factor variable type.
mutate(gender = "male")

suicides_investing_combined_female <- rbind(suicides_noninvesting_female,
                                           suicides_investing_female) %>%

# bind the two female datasets by row.
mutate(investing_age = as.factor(investing_age)) %>%
# Convert investing_age to factor variable type.
mutate(gender = "female")

suicides_avg_all <- rbind(suicides_investing_combined,
                          suicides_investing_combined_male,
                          suicides_investing_combined_female) %>%
mutate(type = case_when(gender == "combined" ~ "Both Genders",
                        gender == "male" ~ "Males Only",
                        gender == "female" ~ "Females Only")) %>%
mutate(recession_year = as.integer(year >= 2007)) %>%
# Create dummy binary variable for whether a year is pre recession or in/post
#Great Recession.
mutate(investing_age = case_when(investing_age == 0 ~ 0,
                                investing_age == 1 ~ 1)) %>%
# Convert investing_age variable to integer to allow for later calculations.
mutate(after_recession = as.integer(recession_year * investing_age)) %>%
# Create dummy binary variable for working age and affected by the Great Recession.
mutate(gender_factor = case_when(gender == 'male' ~ 1,
                                gender == 'female' ~ 0)) %>%
# Create numeric binary variable for gender.

```

```

mutate(after_male = as.integer(recession_year * investing_age * gender_factor)) %>%
mutate(investing_age = as.factor(investing_age))

skim(suicides_us) %>%
  select(skim_type, skim_variable, n_missing,
         numeric.mean, numeric.sd, numeric.p0,
         numeric.p25, numeric.p50, numeric.p75,
         numeric.p100) %>%
  filter(skim_variable != "investing_age") %>%
  filter(skim_variable != "recession_year") %>%
  filter(skim_variable != "post_recession") %>%
  filter(skim_variable != "gender_factor") %>%
  filter

suicides_us_gdp %>%
  ggplot(aes(x = year, y = gdp_per_capita)) +
  geom_line(alpha = 0.5, colour = "goldenrod") +
  geom_point(colour = "turquoise4") +
  labs(x = "Year",
       y = "GDP per capita") +
  theme_minimal() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"),
        panel.spacing.y = unit(2, "lines"),
        axis.title = element_text(margin = margin(t = 0, r = 10, b = 0, l = 0),
                                   family = "Gill Sans MT"),
        axis.text = element_text(family = "Gill Sans MT")) +
  geom_vline(xintercept = 2006.75, linetype = "dashed") +
  geom_vline(xintercept = 2009.25, linetype = "dashed") +
  scale_y_continuous(labels = dollar_format(prefix = "$"))

suicides_us %>%
  mutate(type = if_else(investing_age == 1, "Investing Age", "Non- investing Age")) %>%
  ggplot(aes(x = year, y = suicides.100k.pop, colour = gender, shape = age)) +
  geom_point(alpha = 0.9) +
  geom_line(alpha = 0.5) +
  labs(x = "Year",
       y = "Suicides per 100k people",
       colour = "Gender Identity (binary)",
       shape = "Age Group",
       title = "") +
  theme_minimal() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"),
        axis.title = element_text(margin = margin(t = 0, r = 10, b = 0, l = 0),
                                   family = "Gill Sans MT"),
        axis.text = element_text(family = "Gill Sans MT"),
        strip.text = element_text(family = "Gill Sans MT"),
        legend.title = element_text(family = "Gill Sans MT"),
        legend.text = element_text(family = "Gill Sans MT"),
        panel.spacing.y = unit(2, "lines")) +

```

```

scale_colour_manual(labels = c("Female", "Male"), values = c("turquoise4",
                                                             "goldenrod")) +
scale_shape_manual(labels = c("15 - 24 years", "25 - 34 years",
                              "35 - 54 years", "5 - 14 years",
                              "55 - 74 years", "75+ years"),
                  values = c(6, 17, 18, 15, 19, 8)) +
geom_vline(xintercept = 2006.5, linetype = "dashed") +
scale_y_discrete(limits = c(0, 10, 20, 30, 40)) +
facet_wrap(vars(type),
           nrow = 2)

suicides_avg_all %>%
  ggplot(aes(x = year, y = avg_suicides.100k.pop, colour = investing_age)) +
  geom_point(alpha = 0.9) +
  geom_line(alpha = 0.5) +
  labs(x = "Year",
       y = "Suicides per 100k people",
       colour = "Of Investing Age") +
  theme_minimal() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"),
        axis.title = element_text(margin = margin(t = 0, r = 10, b = 0, l = 0),
                                   family = "Gill Sans MT"),
        axis.text = element_text(family = "Gill Sans MT"),
        strip.text = element_text(family = "Gill Sans MT"),
        legend.title = element_text(family = "Gill Sans MT"),
        legend.text = element_text(family = "Gill Sans MT"),
        panel.spacing.y = unit(2, "lines")) +
  scale_colour_manual(labels = c("No", "Yes"), values = c("turquoise4", "goldenrod")) +
  geom_vline(xintercept = 2006.5, linetype = "dashed") +
  facet_wrap(vars(type),
           nrow = 3,
           scales = "free_y")

suicides_dd <- lm(suicides.100k.pop ~ post_recession + investing_age + year + gender,
                 data = suicides_us)
# Full dataset model using all age groups controlling for gender, investing age, and year.

suicides_ddd <- lm(suicides.100k.pop ~ post_male + investing_age + year + gender,
                  data = suicides_us)
# Full dataset DDD model using all age groups and controlling for gender,
# investing age, and year.

ht <- huxreg("Suicides After Recession" = suicides_dd,
            "Suicides DDD" = suicides_ddd,
            coefs = c("Post Recession" = "post_recession",
                      "Post Recession Male" = "post_male",
                      "Investing Age" = "investing_age",
                      "Year" = "year",
                      "Male Suicides" = "gendermale"),
            number_format = 2) %>%
  set_top_border(1, 1:3, 3) %>%

```

```

set_bottom_border(3, 1:3, 0.5) %>%
set_bottom_border(5, 1:3, 0.5) %>%
set_bottom_border(7, 1:3, 0.5) %>%
set_bottom_border(9, 1:3, 0.5) %>%
set_bottom_border(11, 1:3, 1.5) %>%
set_bottom_border(15, 1:3, 3) %>%
set_width(.55) %>%
set_align(1:10, "Suicides DDD", "right") %>%
set_position("center")
ht

vif(suicides_dd)
vif(suicides_ddd)
bptest(suicides_dd)
bptest(suicides_ddd)
plot(suicides_dd)
plot(suicides_ddd)

```