Analysis of Suicide Trends

ggh725

INTRODUCTION

The aim of this project is to examine global trends in suicides in order to determine whether demographic variables are associated with suicide trends. The effectiveness of the developed models is assessed using the Root Mean Squared Error (RMSE).

The Suicides dataset contains global suicide rates from 1985 to 2016, as well as demographic and socio-economic variables. For purposes of this project, the explanatory variables of interest are: sex, age, generation, population, and continent. The Suicides dataset was divided into training and test sets, both of which contain 50% of the observations.

For the first part of this project, the features of the Suicides dataset are explored to determine possible trends in suicide rates. This is followed by the analysis section. Models were developed based on the assumption that the number of suicides ($\hat{Y}$) is a function of demographic variables (X), where (X) includes sex, age, generation, population, and continent.

The project concludes with a summary of its findings, as well as a discussion of its limitation and possible extensions of the analysis.

EXPLORATORY DATA ANALYSIS

In this section, the Suicides dataset is explored and analyzed in order to determine the features of the data and to detect possible trends.

```
library(tidyverse)
```

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

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```

```
## Warning: package 'tidyr' was built under R version 3.6.1
```

```
## Warning: package 'readr' was built under R version 3.6.1
```

```
## Warning: package 'purrr' was built under R version 3.6.1
```

```
## Warning: package 'dplyr' was built under R version 3.6.1
```

```
## Warning: package 'forcats' was built under R version 3.6.1
```

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

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.6.1
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.6.1

##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```
library(countrycode)
```

```
## Warning: package 'countrycode' was built under R version 3.6.1
```

```
library(broom)
```

```
## Warning: package 'broom' was built under R version 3.6.1
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.6.1

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.6.1

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 3.6.1

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':
##
##     combine

## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(ranger)
```

```
## Warning: package 'ranger' was built under R version 3.6.2
```

```
##
## Attaching package: 'ranger'
```

```
## The following object is masked from 'package:randomForest':
##
##     importance
```

At the outset, the data should be downloaded and unzipped. The Suicides dataset can be found in Kaggle Suicide Rates Overview 1985 to 2016.

```
file <- unzip("./suicide-rates-overview-1985-to-2016.zip")
suicides <- read.csv("./master.csv")
```

This is followed by an examination of the dataset.

```
glimpse(suicides)
```

```
## Observations: 27,820
## Variables: 12
## $ ï..country         <fct> Albania, Albania, Albania, Albania, Albania, Alb...
## $ year               <int> 1987, 1987, 1987, 1987, 1987, 1987, 1987, 1987, ...
## $ sex                <fct> male, male, female, male, male, female, female, ...
## $ age                <fct> 15-24 years, 35-54 years, 15-24 years, 75+ years...
## $ suicides_no        <int> 21, 16, 14, 1, 9, 1, 6, 4, 1, 0, 0, 0, 2, 17, 1,...
## $ population         <int> 312900, 308000, 289700, 21800, 274300, 35600, 27...
## $ suicides.100k.pop  <dbl> 6.71, 5.19, 4.83, 4.59, 3.28, 2.81, 2.15, 1.56, ...
## $ country.year       <fct> Albania1987, Albania1987, Albania1987, Albania19...
## $ HDI.for.year       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ gdp_for_year....   <fct> "2,156,624,900", "2,156,624,900", "2,156,624,900...
## $ gdp_per_capita.... <int> 796, 796, 796, 796, 796, 796, 796, 796, 796, 796...
## $ generation         <fct> Generation X, Silent, Generation X, G.I. Generat...
```

```
summary(suicides)
```

```
##        ï..country          year          sex                  age
##  Austria    :  382   Min.   :1985   female:13910   15-24 years:4642
##  Iceland    :  382   1st Qu.:1995   male  :13910   25-34 years:4642
##  Mauritius  :  382   Median :2002                  35-54 years:4642
##  Netherlands:  382   Mean   :2001                  5-14 years :4610
##  Argentina  :  372   3rd Qu.:2008                  55-74 years:4642
##  Belgium    :  372   Max.   :2016                  75+ years  :4642
##  (Other)    :25548
##   suicides_no        population       suicides.100k.pop      country.year
##  Min.   :    0.0   Min.   :     278   Min.   :  0.00    Albania1987:   12
##  1st Qu.:    3.0   1st Qu.:   97498   1st Qu.:  0.92    Albania1988:   12
##  Median :   25.0   Median :  430150   Median :  5.99    Albania1989:   12
##  Mean   :  242.6   Mean   : 1844794   Mean   : 12.82    Albania1992:   12
##  3rd Qu.:  131.0   3rd Qu.: 1486143   3rd Qu.: 16.62    Albania1993:   12
##  Max.   :22338.0   Max.   :43805214   Max.   :224.97    Albania1994:   12
##                                                       (Other)    :27748
##   HDI.for.year            gdp_for_year.... gdp_per_capita....
##  Min.   :0.483   1,002,219,052,968:   12   Min.   :   251
##  1st Qu.:0.713   1,011,797,457,139:   12   1st Qu.:  3447
```

```
##  Median :0.779   1,016,418,229    :   12   Median :  9372
##  Mean   :0.777   1,018,847,043,277:   12   Mean   : 16866
##  3rd Qu.:0.855   1,022,191,296    :   12   3rd Qu.: 24874
##  Max.   :0.944   1,023,196,003,075:   12   Max.   :126352
##  NA's   :19456   (Other)          :27748
##            generation
##  Boomers        :4990
##  G.I. Generation:2744
##  Generation X   :6408
##  Generation Z   :1470
##  Millenials     :5844
##  Silent         :6364
##
```

```
head(suicides)
```

```
##   ï..country year    sex         age suicides_no population suicides.100k.pop
## 1    Albania 1987   male 15-24 years          21     312900              6.71
## 2    Albania 1987   male 35-54 years          16     308000              5.19
## 3    Albania 1987 female 15-24 years          14     289700              4.83
## 4    Albania 1987   male   75+ years           1      21800              4.59
## 5    Albania 1987   male 25-34 years           9     274300              3.28
## 6    Albania 1987 female   75+ years           1      35600              2.81
##   country.year HDI.for.year gdp_for_year.... gdp_per_capita....      generation
## 1  Albania1987           NA    2,156,624,900                796    Generation X
## 2  Albania1987           NA    2,156,624,900                796          Silent
## 3  Albania1987           NA    2,156,624,900                796    Generation X
## 4  Albania1987           NA    2,156,624,900                796 G.I. Generation
## 5  Albania1987           NA    2,156,624,900                796         Boomers
## 6  Albania1987           NA    2,156,624,900                796 G.I. Generation
```

There are 27,820 observations with 12 variables. The dependent variable is number of suicides (suicides_no). The explanatory variables of interest are: country, sex, age, generation, and population.

Trends in Number of Suicides

```
summary(suicides$suicides_no)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0     3.0    25.0   242.6   131.0 22338.0
```

```
sd(suicides$suicides_no)
```

```
## [1] 902.0479
```

```
IQR(suicides$suicides_no)
```

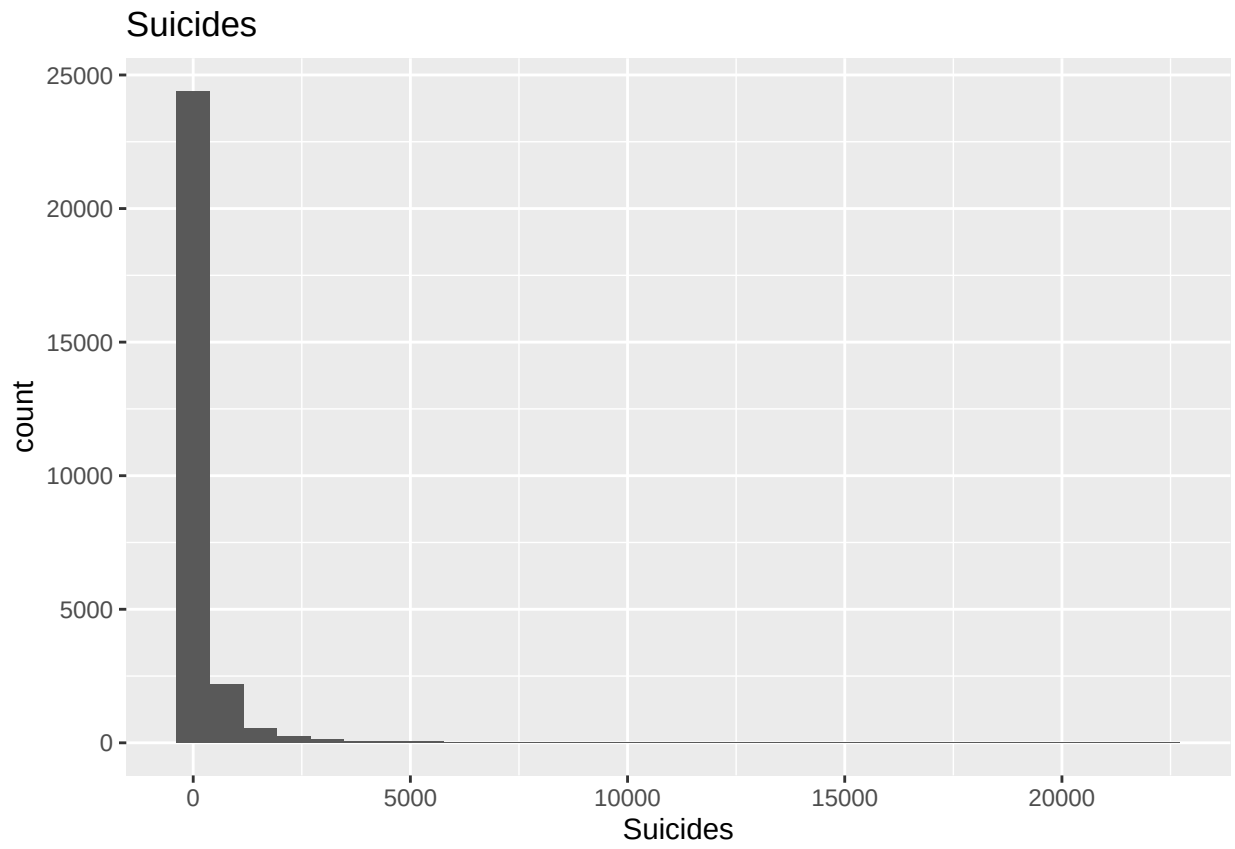
```
## [1] 128
```

The value of the mean is significantly higher than the median. This shows that there are outliers in the data. Additionally, the standard deviation is 902.05. As the variance is larger than the mean, a Quasipoisson analysis might be worth considering.

The histograms below confirm the presence of outliers. Furthermore, the first histogram shows that the suicide numbers are heavily right skewed. Taking the log transformation of the number of suicides elicits a histogram which better illustrates the trends in the data.

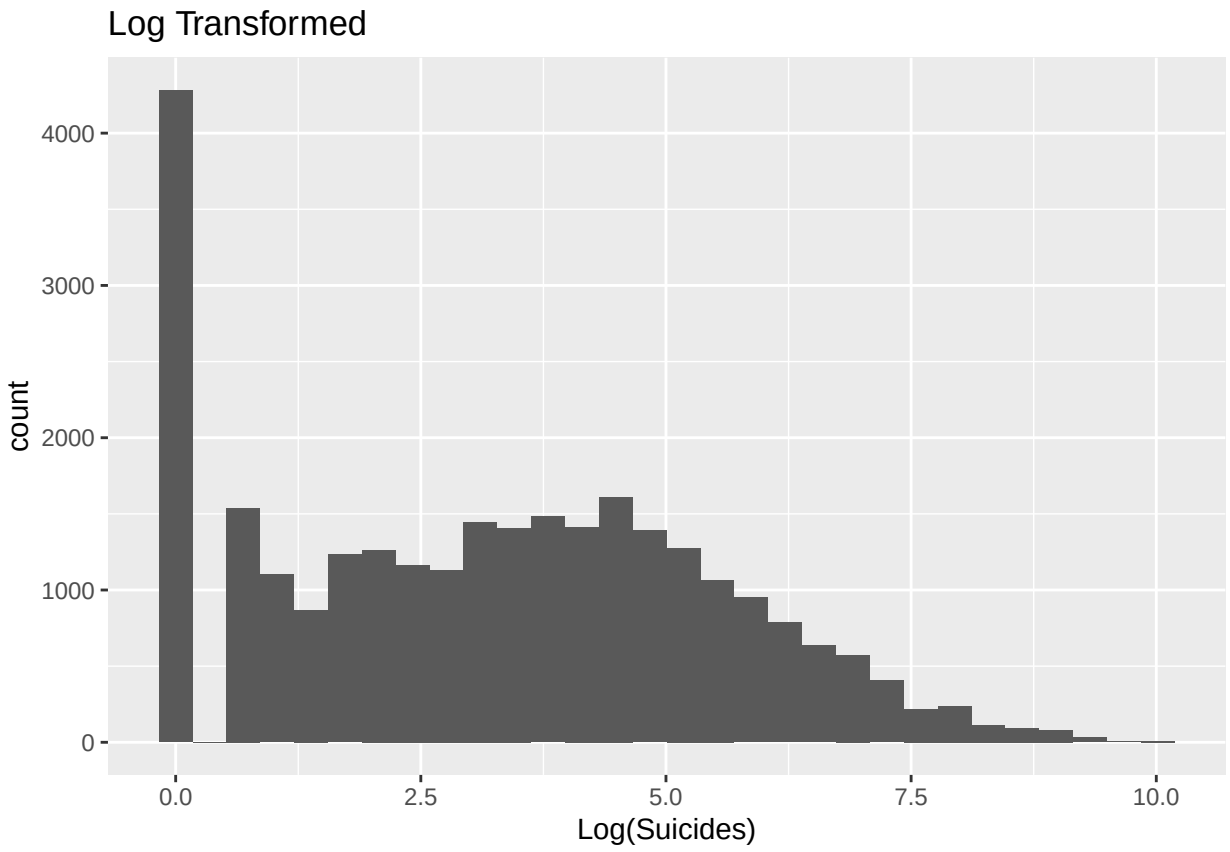
```
suicides %>%
  ggplot(aes(suicides_no)) + geom_histogram() +
    labs(title = "Suicides", x = "Suicides")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
suicides %>%
  ggplot(aes(log(suicides_no + 1))) + geom_histogram() +
    labs(title = "Log Transformed", x = "Log(Suicides)")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

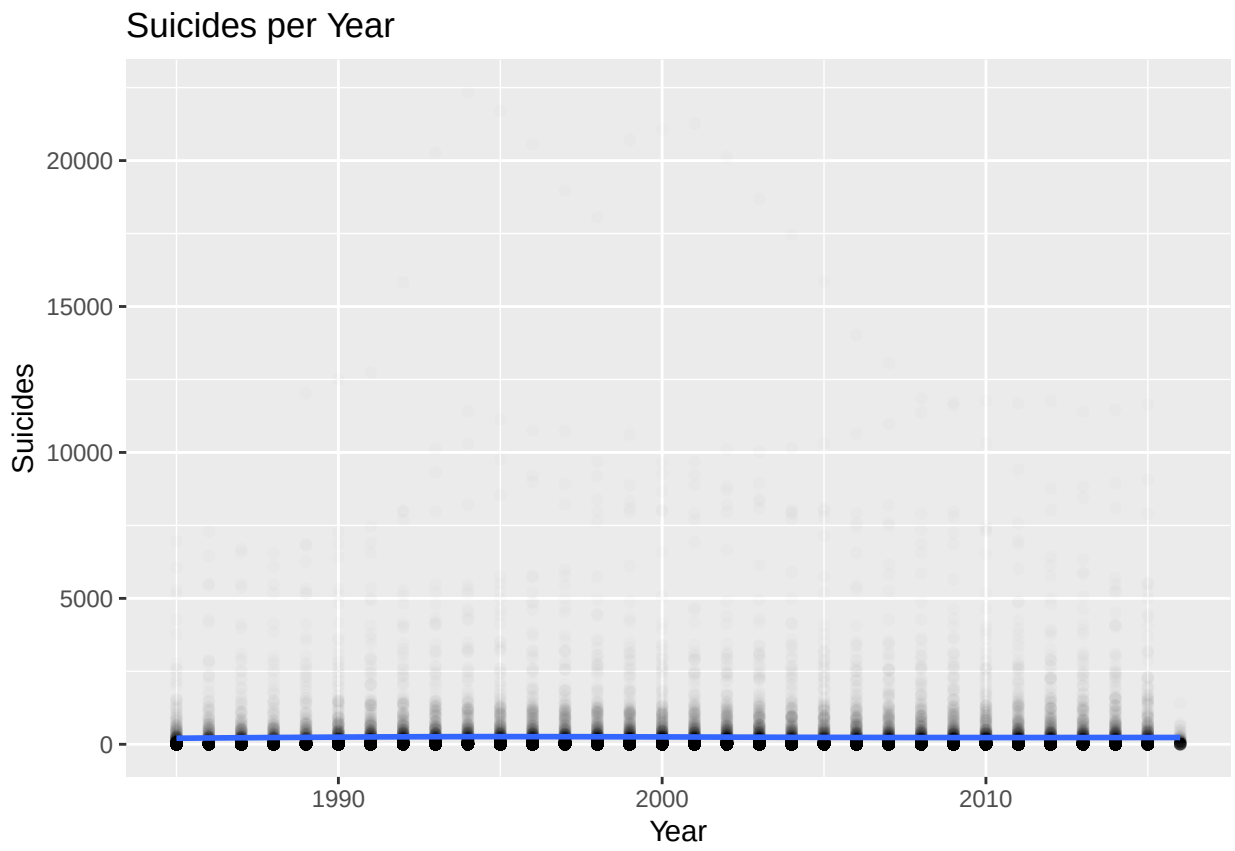


Trends in Suicides per Year

The plot below shows that there is no clear time trend in the number of suicides.

```
suicides$year <- as.Date(as.character(suicides$year), format = "%Y")
suicides$year <- year(suicides$year)

suicides %>%
  ggplot(aes(year, suicides_no)) +
  geom_point(alpha = 0.01) + geom_smooth() +
    labs(title = "Suicides per Year", x = "Year",
         y = "Suicides")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Trends per Country and Continent

```
colnames(suicides)[1] <- "country"
names(suicides)
```

```
##  [1] "country"            "year"               "sex"
##  [4] "age"                "suicides_no"        "population"
##  [7] "suicides.100k.pop"  "country.year"       "HDI.for.year"
## [10] "gdp_for_year...."   "gdp_per_capita...." "generation"
```

```
levels(suicides$country)
```

```
##   [1] "Albania"                    "Antigua and Barbuda"
##   [3] "Argentina"                  "Armenia"
##   [5] "Aruba"                      "Australia"
##   [7] "Austria"                    "Azerbaijan"
##   [9] "Bahamas"                    "Bahrain"
##  [11] "Barbados"                   "Belarus"
##  [13] "Belgium"                    "Belize"
##  [15] "Bosnia and Herzegovina"     "Brazil"
##  [17] "Bulgaria"                   "Cabo Verde"
##  [19] "Canada"                     "Chile"
##  [21] "Colombia"                   "Costa Rica"
##  [23] "Croatia"                    "Cuba"
##  [25] "Cyprus"                     "Czech Republic"
```

```
##  [27] "Denmark"                      "Dominica"
##  [29] "Ecuador"                      "El Salvador"
##  [31] "Estonia"                      "Fiji"
##  [33] "Finland"                      "France"
##  [35] "Georgia"                      "Germany"
##  [37] "Greece"                       "Grenada"
##  [39] "Guatemala"                    "Guyana"
##  [41] "Hungary"                      "Iceland"
##  [43] "Ireland"                      "Israel"
##  [45] "Italy"                        "Jamaica"
##  [47] "Japan"                        "Kazakhstan"
##  [49] "Kiribati"                     "Kuwait"
##  [51] "Kyrgyzstan"                   "Latvia"
##  [53] "Lithuania"                    "Luxembourg"
##  [55] "Macau"                        "Maldives"
##  [57] "Malta"                        "Mauritius"
##  [59] "Mexico"                       "Mongolia"
##  [61] "Montenegro"                   "Netherlands"
##  [63] "New Zealand"                  "Nicaragua"
##  [65] "Norway"                       "Oman"
##  [67] "Panama"                       "Paraguay"
##  [69] "Philippines"                  "Poland"
##  [71] "Portugal"                     "Puerto Rico"
##  [73] "Qatar"                        "Republic of Korea"
##  [75] "Romania"                      "Russian Federation"
##  [77] "Saint Kitts and Nevis"        "Saint Lucia"
##  [79] "Saint Vincent and Grenadines" "San Marino"
##  [81] "Serbia"                       "Seychelles"
##  [83] "Singapore"                    "Slovakia"
##  [85] "Slovenia"                     "South Africa"
##  [87] "Spain"                        "Sri Lanka"
##  [89] "Suriname"                     "Sweden"
##  [91] "Switzerland"                  "Thailand"
##  [93] "Trinidad and Tobago"          "Turkey"
##  [95] "Turkmenistan"                 "Ukraine"
##  [97] "United Arab Emirates"         "United Kingdom"
##  [99] "United States"                "Uruguay"
## [101] "Uzbekistan"
```

The country variable is coded as a factor variable with 101 levels. To make the analysis more tractable, a continent variable will be added. The models will be using the continent variable, rather than the country variable.

```
suicides$continent <- countrycode(sourcevar = suicides[,"country"],
                                  origin = "country.name",
                                  destination = "continent")

str(suicides)
```

```
## 'data.frame':    27820 obs. of  13 variables:
##  $ country           : Factor w/ 101 levels "Albania","Antigua and Barbuda",..: 1 1 1 1 1 1 1 1 1 1
##  $ year              : num  1987 1987 1987 1987 1987 ...
##  $ sex               : Factor w/ 2 levels "female","male": 2 2 1 2 2 1 1 1 2 1 ...
##  $ age               : Factor w/ 6 levels "15-24 years",..: 1 3 1 6 2 6 3 2 5 4 ...
##  $ suicides_no       : int  21 16 14 1 9 1 6 4 1 0 ...
```

```
##  $ population        : int  312900 308000 289700 21800 274300 35600 278800 257200 137500 311000 ...
##  $ suicides.100k.pop : num  6.71 5.19 4.83 4.59 3.28 2.81 2.15 1.56 0.73 0 ...
##  $ country.year      : Factor w/ 2321 levels "Albania1987",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ HDI.for.year      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ gdp_for_year....  : Factor w/ 2321 levels "1,002,219,052,968",..: 727 727 727 727 727 727 727 727
##  $ gdp_per_capita....: int  796 796 796 796 796 796 796 796 796 796 ...
##  $ generation        : Factor w/ 6 levels "Boomers","G.I. Generation",..: 3 6 3 2 1 2 6 1 2 3 ...
##  $ continent         : chr  "Europe" "Europe" "Europe" "Europe" ...
```

```
suicides$continent <- factor(suicides$continent)
levels(suicides$continent)
```

```
## [1] "Africa"   "Americas" "Asia"     "Europe"   "Oceania"
```

```
suicides %>%
    group_by(continent) %>%
    summarize(mean = mean(suicides_no),
              median = median(suicides_no),
              sd = sd(suicides_no),
              n = n()) %>%
    arrange(desc(mean))
```

```
## # A tibble: 5 x 5
##   continent  mean median     sd     n
##   <fct>     <dbl>  <dbl>  <dbl> <int>
## 1 Europe    299.      60 1061.  11418
## 2 Asia      271.      20  838.   5366
## 3 Americas  194.       9  798.   9214
## 4 Oceania    87.3     23  148.    972
## 5 Africa     13.4      4   23.7   850
```

Europe has the highest mean and median suicides, while Africa has the lowest.

```
suicides %>%
  group_by(country) %>%
  summarize(total = sum(suicides_no)) %>%
  arrange(desc(total)) %>%
  top_n(10)
```

```
## Selecting by total
```

```
## # A tibble: 10 x 2
##    country              total
##    <fct>                <int>
##  1 Russian Federation 1209742
##  2 United States      1034013
##  3 Japan               806902
##  4 France              329127
##  5 Ukraine             319950
##  6 Germany             291262
##  7 Republic of Korea   261730
##  8 Brazil              226613
##  9 Poland              139098
## 10 United Kingdom      136805
```

The country with the highest total number of suicides is Russia, followed by the United States and Japan.

Trends per Sex

```
levels(suicides$sex)
```

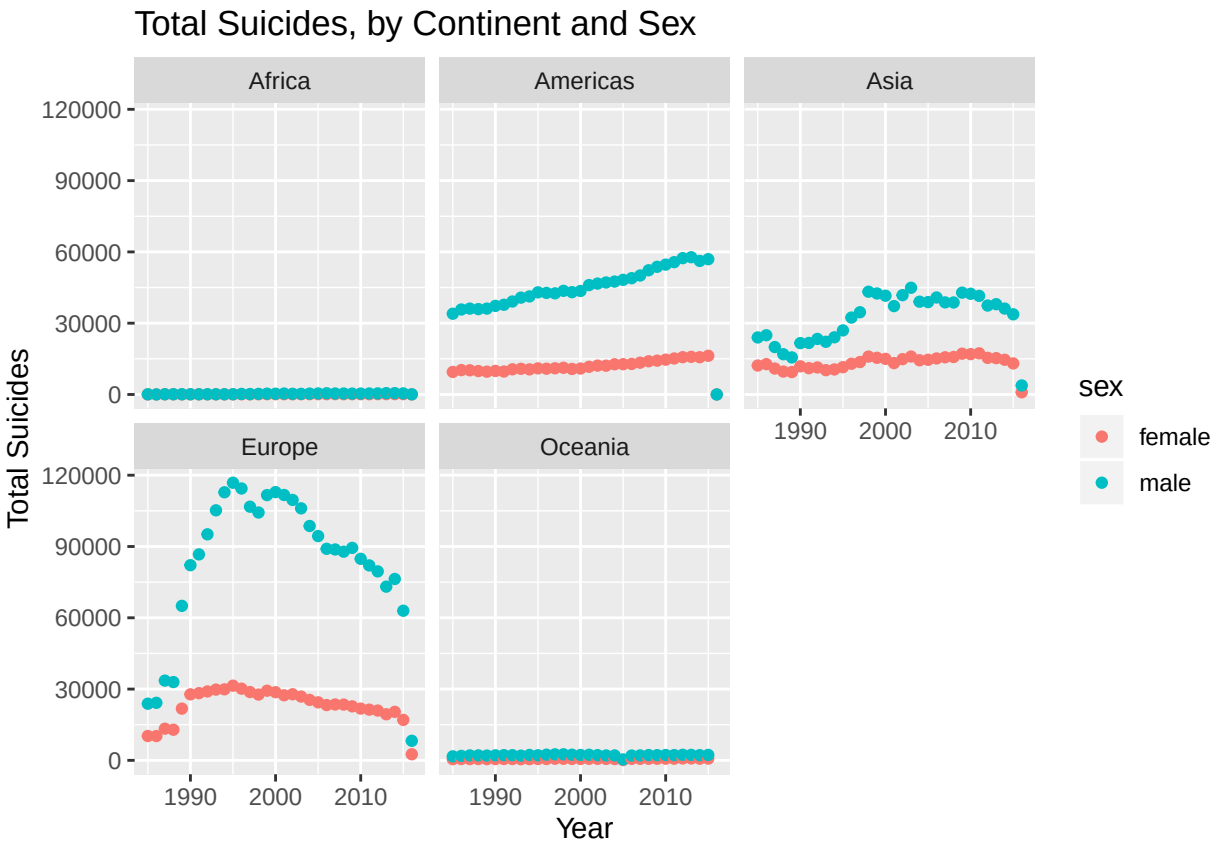
```
## [1] "female" "male"
```

```
suicides %>%
    group_by(sex) %>%
    summarize(total = sum(suicides_no))
```

```
## # A tibble: 2 x 2
##   sex      total
##   <fct>    <int>
## 1 female 1559510
## 2 male   5188910
```

Males have a higher number of suicides than females. This trend can be seen in the Americas, Asia and Europe.

```
suicides %>%
    group_by(year, continent, sex) %>%
    summarize(total = sum(suicides_no)) %>%
    ggplot(aes(year, total, color = sex)) +
    geom_point() + facet_wrap(~continent) +
    labs(title = "Total Suicides, by Continent and Sex",
         x = "Year", y = "Total Suicides")
```



Trends per Age

```
levels(suicides$age)
```

```
## [1] "15-24 years" "25-34 years" "35-54 years" "5-14 years"  "55-74 years"
## [6] "75+ years"
```

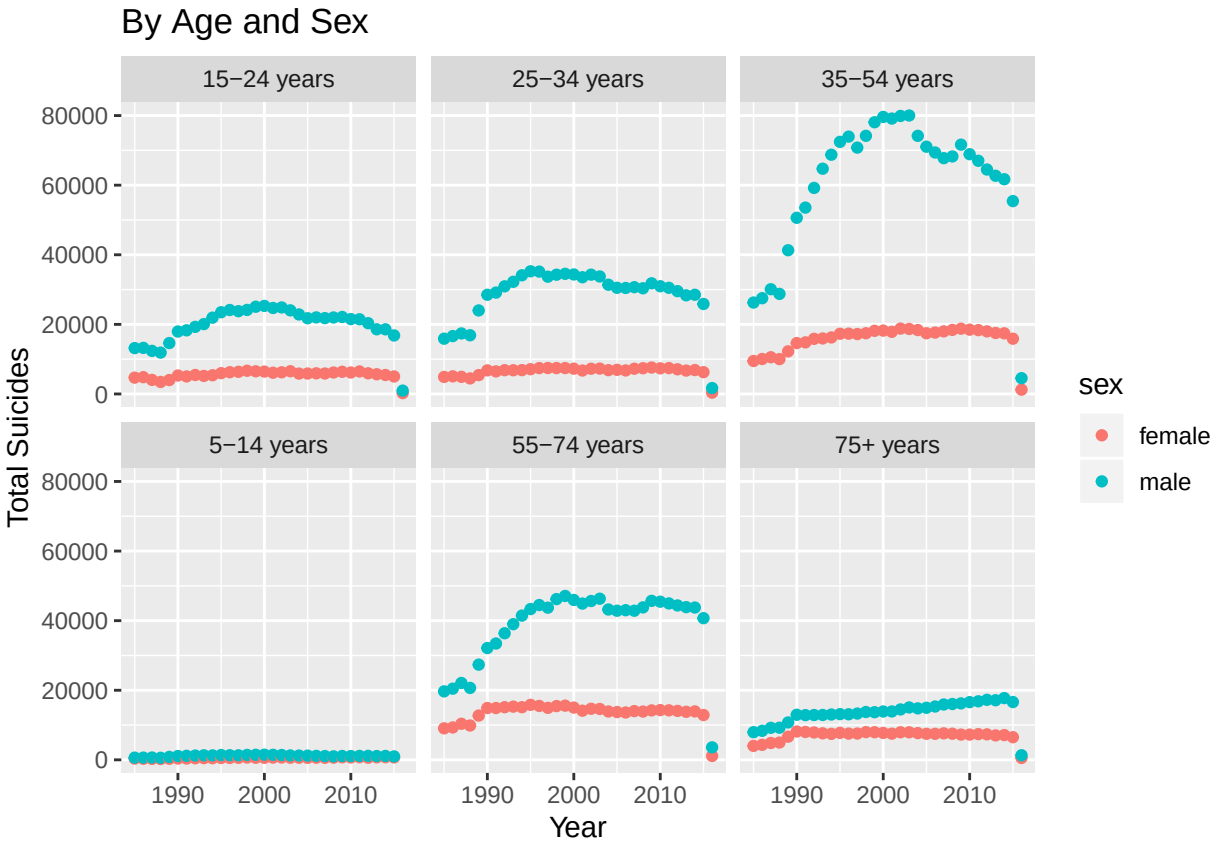
Age is encoded as a factor variable with 6 levels: 5-14 years, 15-24 years, 25-34 years, 35-54 years, 55-74 years, and 75+ years.

```
suicides %>%
    group_by(age) %>%
    summarize(total = sum(suicides_no)) %>%
    arrange(desc(total))
```

```
## # A tibble: 6 x 2
##   age           total
##   <fct>         <int>
## 1 35-54 years 2452141
## 2 55-74 years 1658443
## 3 25-34 years 1123912
## 4 15-24 years  808542
## 5 75+ years    653118
## 6 5-14 years    52264
```

The highest number of total suicides are in the age groups which correspond to adults, i.e., 25-74 years. The trend that suicide rates for males are higher can also be seen in the adult age groups.

```
suicides %>%
    group_by(year, age, sex) %>%
    summarize(total = sum(suicides_no)) %>%
    ggplot(aes(year, total, color = sex)) +
    geom_point() + facet_wrap(~ age) +
    labs(title = "By Age and Sex", x = "Year",
         y = "Total Suicides")
```

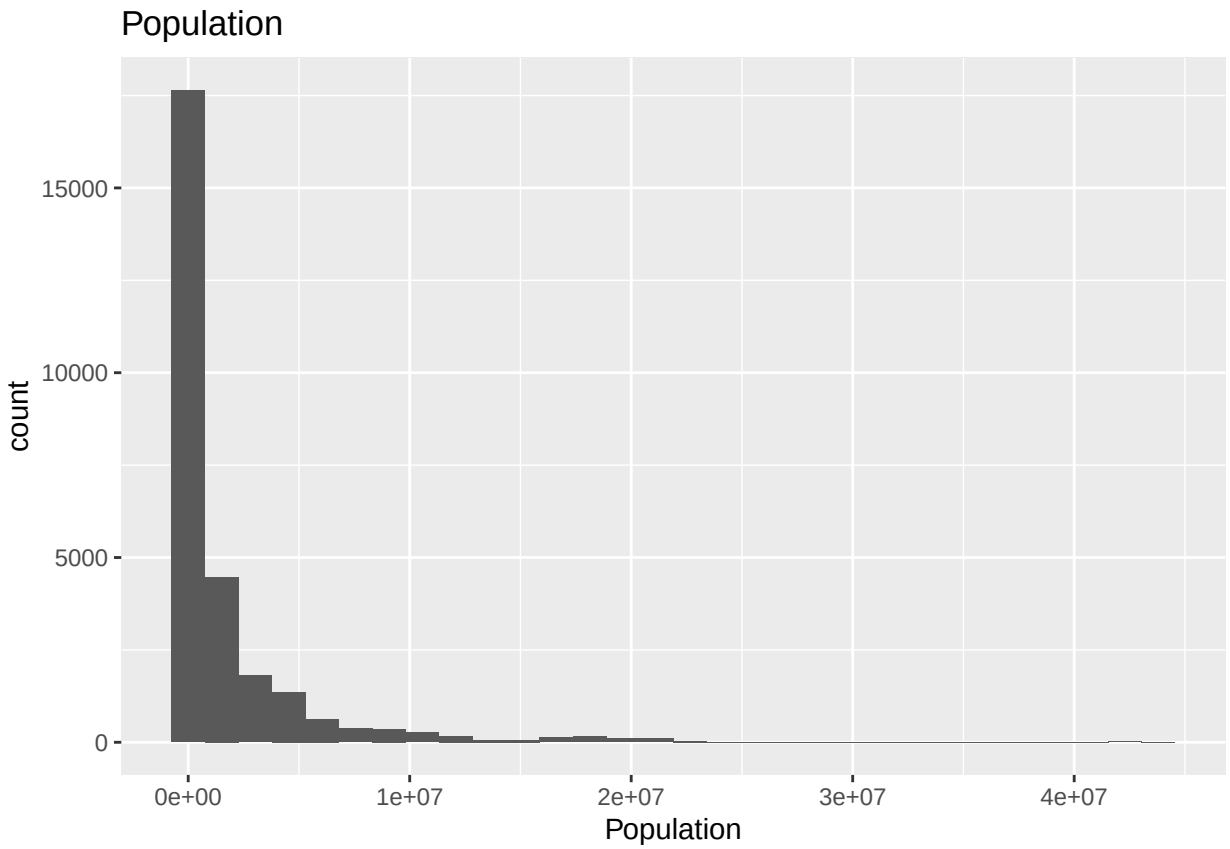


Trends by Population

The population is also heavily right skewed. The log transformation of this variable can better show its distribution.

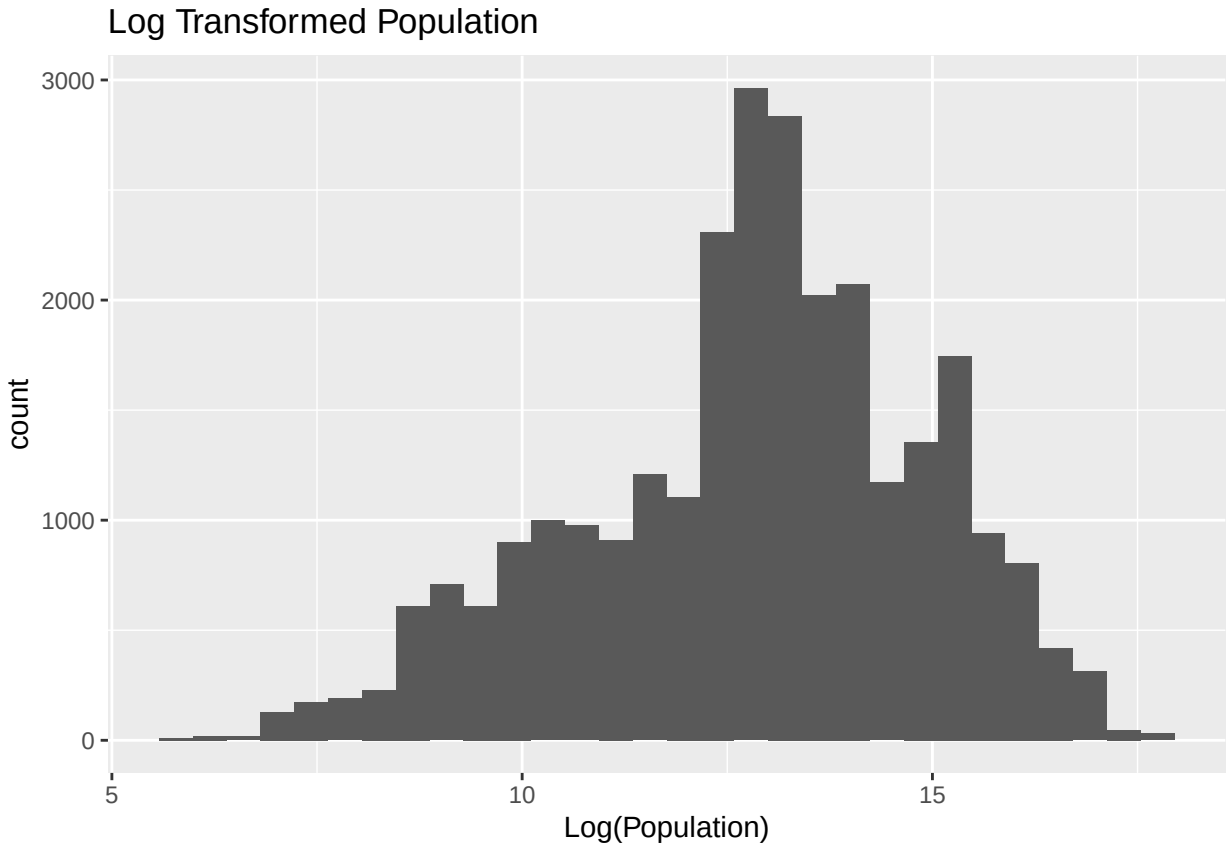
```
suicides %>%
  ggplot(aes(population)) + geom_histogram() +
  labs(title = "Population", x = "Population")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



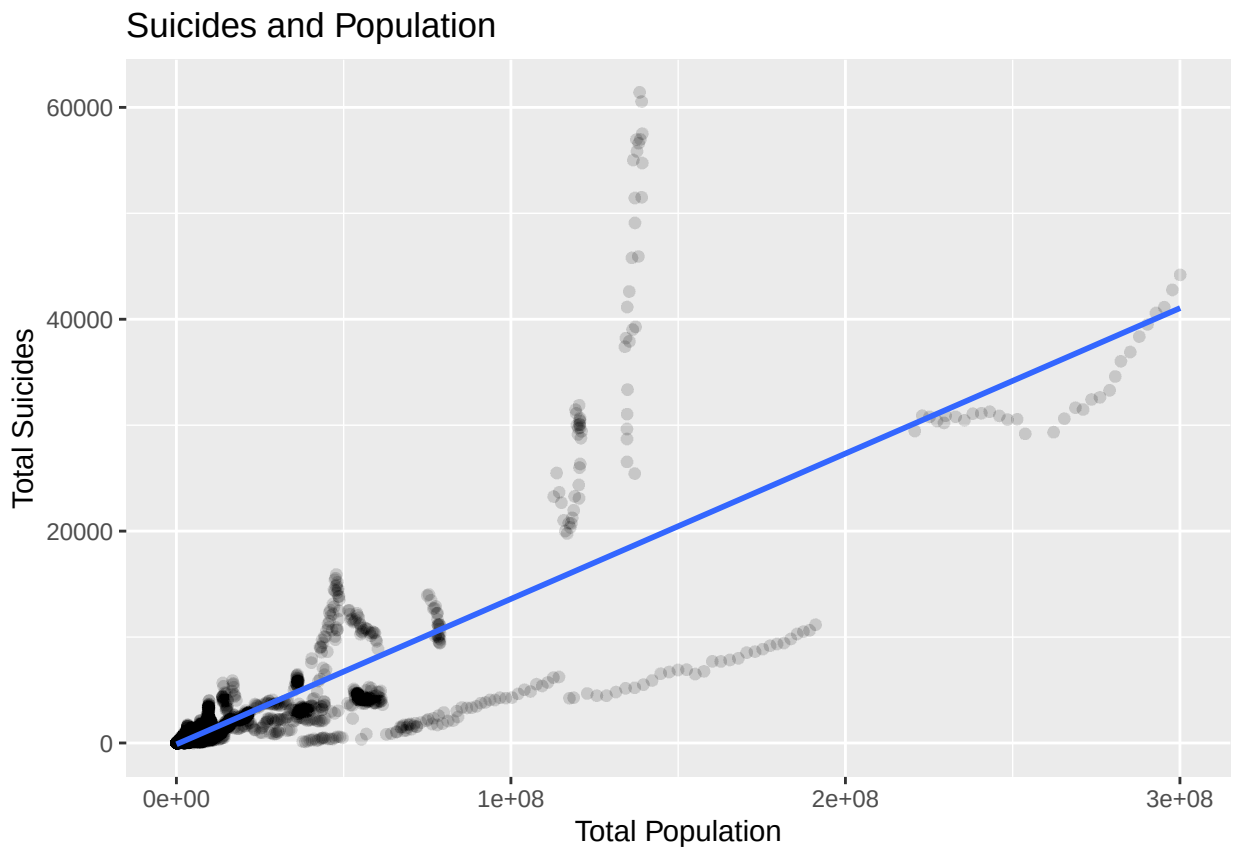

```
suicides %>%
  ggplot(aes(log(population))) + geom_histogram() +
  labs(title = "Log Transformed Population", x = "Log(Population)")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



A cursory look at the data suggests that there is a positive association between population and suicide rates.

```
suicides %>%
    group_by(year, country) %>%
    mutate(total_pop = sum(population), total_suicides = sum(suicides_no)) %>%
    ggplot(aes(total_pop, total_suicides)) +
    geom_point(alpha = 0.01, position = "jitter") +
    geom_smooth(method = "lm", se = FALSE) +
    labs(title = "Suicides and Population",
       x = "Total Population",
       y = "Total Suicides")
```



Trends by Generation

```
levels(suicides$generation)
```

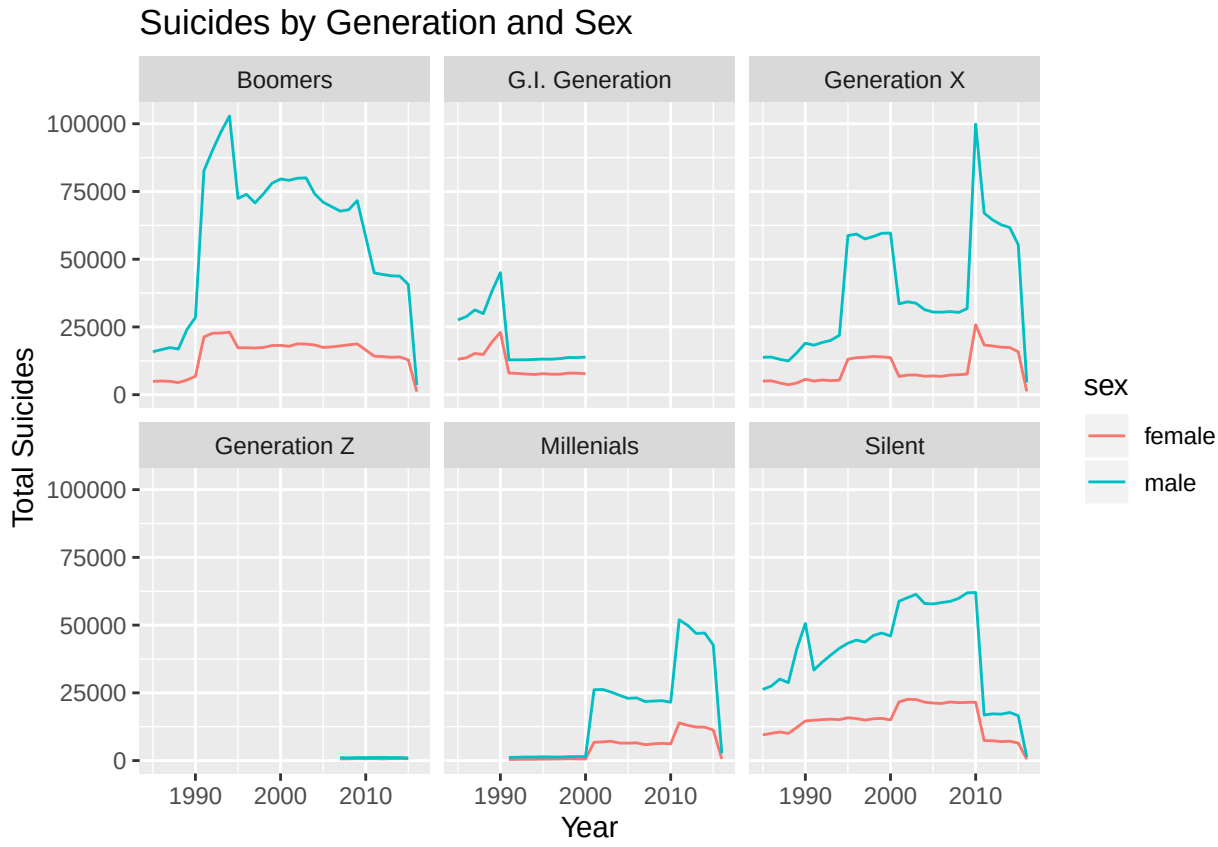
```
## [1] "Boomers"         "G.I. Generation" "Generation X"    "Generation Z"
## [5] "Millenials"      "Silent"
```

Generation is encoded as a factor variable with 6 levels: Boomers, G.I. Generation, Generation X, Generation Z, Millenials, and Silent.

```
suicides %>%
    group_by(generation) %>%
    summarize(total = sum(suicides_no)) %>%
    arrange(desc(total))
```

```
## # A tibble: 6 x 2
##   generation        total
##   <fct>             <int>
## 1 Boomers         2284498
## 2 Silent          1781744
## 3 Generation X    1532804
## 4 Millenials       623459
## 5 G.I. Generation  510009
## 6 Generation Z      15906
```

```
suicides %>%
    group_by(year, generation, sex) %>%
    summarize(total = sum(suicides_no)) %>%
    ggplot(aes(year, total, color = sex)) +
    geom_line() + facet_wrap(~generation) +
    labs(title = "Suicides by Generation and Sex",
         x = "Year", y = "Total Suicides")
```



The Boomer generation has the highest number of total suicides, followed by the Silent generation and Generation X. Even when viewed according to generations, males still have a higher number of suicides than females.

Trends by GDP

Before the trends can be analyzed, the variables must be cleaned.

```
colnames(suicides)[10] <- "gdp_per_year"
colnames(suicides)[11] <- "gdp_per_capita"
names(suicides)
```

```
##  [1] "country"           "year"              "sex"
##  [4] "age"               "suicides_no"       "population"
##  [7] "suicides.100k.pop" "country.year"      "HDI.for.year"
## [10] "gdp_per_year"      "gdp_per_capita"    "generation"
## [13] "continent"
```

```r
levels(suicides$gdp_per_year)
```

```
##    [1] "1,002,219,052,968"  "1,011,797,457,139"  "1,016,418,229"
##    [4] "1,018,847,043,277"  "1,022,191,296"      "1,023,196,003,075"
##    [7] "1,025,211,803,414"  "1,033,561,654"      "1,052,121,055"
##   [10] "1,052,584,601,611"  "1,052,696,282,279"  "1,056,188,593"
##   [13] "1,057,801,282,051"  "1,057,845,500"      "1,059,498,884"
##   [16] "1,061,388,722,256"  "1,061,445,225,791"  "1,065,826,670"
##   [19] "1,069,555,500,372"  "1,078,402,128"      "1,080,642,033"
##   [22] "1,093,169,389,205"  "1,094,499,338,703"  "1,095,590,833,694"
##   [25] "1,101,897,918"      "1,107,640,297,890"  "1,109,989,038,339"
##   [28] "1,114,222,550"      "1,117,835,286"      "1,122,679,154,632"
##   [31] "1,134,193,630"      "1,140,489,745,944"  "1,141,759,996,315"
##   [34] "1,142,797,178,131"  "1,144,260,547,873"  "1,157,005,444"
##   [37] "1,157,276,458,152"  "1,159,869,246"      "1,160,900,000"
##   [40] "1,162,317,852,349"  "1,163,362,438"      "1,169,357,979,865"
##   [43] "1,169,622,672,463"  "1,170,787,352,906"  "1,177,326,294,441"
##   [46] "1,179,659,529,660"  "1,180,489,563,964"  "1,192,925,407"
##   [49] "1,197,789,902,774"  "1,201,090,018,604"  "1,201,313,201"
##   [52] "1,202,463,682,634"  "1,211,411,704"      "1,215,640,315"
##   [55] "1,217,467,600"      "1,222,643,696,992"  "1,222,807,284,485"
##   [58] "1,224,253,000"      "1,226,829,563"      "1,228,071,038"
##   [61] "1,239,050,932,242"  "1,242,109,397,534"  "1,248,563,179,203"
##   [64] "1,249,062,025"      "1,262,973,407"      "1,264,551,499,185"
##   [67] "1,266,309,245,009"  "1,266,510,634,293"  "1,269,179,616,914"
##   [70] "1,269,276,828,276"  "1,271,196,078"      "1,272,577,456"
##   [73] "1,274,443,078,609"  "1,280,133,333"      "1,282,215,407"
##   [76] "1,284,685,051"      "1,290,573,400"      "1,299,705,247,686"
##   [79] "1,305,604,981,272"  "1,308,929,351,236"  "1,311,401,333"
##   [82] "1,314,385,330,073"  "1,315,158,670"      "1,315,415,197,461"
##   [85] "1,315,806,985,860"  "1,320,670,391"      "1,322,815,612,694"
##   [88] "1,328,091,524"      "1,335,218,557,677"  "1,336,018,949,806"
##   [91] "1,336,957,250"      "1,342,997,306"      "1,349,034,029,453"
##   [94] "1,361,854,206,549"  "1,362,248,940,483"  "1,364,863,037"
##   [97] "1,368,400,705,491"  "1,368,431,037"      "1,368,625,150"
##  [100] "1,369,693,171"      "1,371,153,004,986"  "1,375,604,279"
##  [103] "1,376,465,324,385"  "1,376,910,811,041"  "1,381,968,259"
##  [106] "1,382,764,027,114"  "1,393,982,750,473"  "1,394,280,784,778"
##  [109] "1,397,084,345,950"  "1,397,113,450"      "1,398,892,744,821"
##  [112] "1,401,465,923,172"  "1,408,781,591,264"  "1,411,333,926,201"
##  [115] "1,417,575,550"      "1,431,616,749,640"  "1,435,079,200"
##  [118] "1,436,803,333"      "1,437,684,815"      "1,452,884,917,959"
##  [121] "1,458,449,453"      "1,464,955,475,994"  "1,464,977,190,206"
##  [124] "1,468,317,350"      "1,469,046,115"      "1,479,341,637,011"
##  [127] "1,484,092,538"      "1,487,005,600"      "1,488,067,258,325"
##  [130] "1,489,928,889"      "1,492,647,560,196"  "1,494,286,655,374"
##  [133] "1,499,099,749,931"  "1,503,108,739,159"  "1,524,916,112,079"
##  [136] "1,525,112,242"      "1,543,411,012,580"  "1,549,131,208,997"
##  [139] "1,551,921,037"      "1,552,483,628,029"  "1,556,321,742"
##  [142] "1,569,649,661,400"  "1,573,670,250"      "1,573,696,522,007"
##  [145] "1,596,968,913"      "1,601,094,756,210"  "1,605,640,633"
##  [148] "1,605,675,086,550"  "1,612,573,850"      "1,613,464,422,811"
##  [151] "1,621,510,004,318"  "1,635,015,380,108"  "1,638,511,096,390"
##  [154] "1,639,492,424"      "1,645,963,750"      "1,647,951,278,560"
```

```
##  [157] "1,660,102,346"      "1,660,844,408,500"  "1,661,018,519"
##  [160] "1,665,623,685,488"  "1,667,019,780,934"  "1,681,006,993"
##  [163] "1,695,824,571,927"  "1,703,951,050"      "1,707,710,053"
##  [166] "1,713,771,321"      "1,722,798,883"      "1,740,334,782"
##  [169] "1,751,247,763"      "1,764,967,948,917"  "1,767,864,036"
##  [172] "1,768,408,273,381"  "1,778,567,600"      "1,788,647,906,048"
##  [175] "1,793,754,805"      "1,798,314,750,435"  "1,802,214,373,741"
##  [178] "1,822,815,594"      "1,824,288,757,448"  "1,827,570,586"
##  [181] "1,832,868,490,534"  "1,840,480,812,641"  "1,842,628,005,830"
##  [184] "1,842,691,481"      "1,845,482,181"      "1,852,661,982,341"
##  [187] "1,861,873,895,109"  "1,864,824,081"      "1,868,383,461"
##  [190] "1,873,452,514"      "1,888,754,655"      "1,893,726,437"
##  [193] "1,911,563,665"      "1,919,012,781"      "1,920,262,570"
##  [196] "1,922,598,121"      "1,941,094,972"      "1,942,633,797,516"
##  [199] "1,949,953,934,034"  "1,950,648,769,575"  "1,950,900,000"
##  [202] "1,958,711,603"      "1,985,673,798"      "10,003,670,690"
##  [205] "10,070,450,000"     "10,095,760,000"     "10,109,612,142"
##  [208] "10,120,274,493"     "10,128,112,401"     "10,142,111,334"
##  [211] "10,145,114,180"     "10,150,978,155"     "10,167,250,000"
##  [214] "10,172,869,680"     "10,183,317,625"     "10,221,705,900"
##  [217] "10,277,598,152"     "10,284,779,000,000" "10,350,515,464"
##  [220] "10,391,504,709"     "10,432,619,390"     "10,440,842,165"
##  [223] "10,526,000,000"     "10,532,001,130"     "10,546,135,160"
##  [226] "10,553,337,673"     "10,574,026,838"     "10,614,455,232"
##  [229] "10,618,340,000"     "10,619,320,049"     "10,621,824,000,000"
##  [232] "10,627,600,000"     "10,646,157,920"     "10,677,286,100"
##  [235] "10,701,011,897"     "10,720,500,000"     "10,766,809,099"
##  [238] "10,829,710,145"     "10,936,669,900"     "10,943,548,387"
##  [241] "10,972,878,636"     "10,977,514,000,000" "10,982,972,256"
##  [244] "100,163,995,151"    "100,273,097,170"    "100,324,627,215"
##  [247] "100,343,219,507"    "100,351,670,000"    "100,948,236,941"
##  [250] "101,550,654,721"    "101,564,800,000"    "101,726,331,000"
##  [253] "101,900,260,856"    "102,445,800,000"    "102,450,000,000"
##  [256] "102,633,789,558"    "103,143,500,000"    "103,321,570,859"
##  [259] "103,904,537,815"    "104,066,609,518"    "104,849,886,826"
##  [262] "105,143,232,380"    "105,899,930,507"    "106,659,508,271"
##  [265] "107,602,689,041"    "107,753,069,307"    "108,809,058,859"
##  [268] "109,103,056,148"    "109,133,512,304"    "109,957,321,960"
##  [271] "11,074,822,074"     "11,121,465,767"     "11,183,458,131"
##  [274] "11,195,612,105"     "11,204,416,000"     "11,234,045,376"
##  [277] "11,240,360,898"     "11,284,197,000"     "11,305,459,802"
##  [280] "11,316,441,660"     "11,399,942,453"     "11,401,351,420"
##  [283] "11,510,670,000,000" "11,513,472,694"     "11,518,393,367"
##  [286] "11,539,211,480"     "11,575,486,400"     "11,609,512,940"
##  [289] "11,618,286,553"     "11,638,536,834"     "11,662,040,714"
##  [292] "11,668,685,524"     "11,692,287,066"     "11,748,433,157"
##  [295] "11,784,927,700"     "11,880,438,824"     "11,905,525,197"
##  [298] "11,926,953,259"     "110,803,391,517"    "110,934,442,763"
##  [301] "111,106,191,358"    "111,452,869,378"    "111,608,845,081"
##  [304] "112,625,431,378"    "113,035,361,317"    "113,675,706,127"
##  [307] "114,188,557,567"    "114,641,097,818"    "114,718,721,397"
##  [310] "114,724,862,034"    "115,270,054,945"    "115,295,199,392"
##  [313] "115,308,661,143"    "115,419,050,942"    "115,537,126,326"
##  [316] "115,748,110,113"    "115,979,230,547"    "116,224,673,043"
```

```
##  [319] "116,601,802,107"    "117,046,198,971"    "117,074,863,822"
##  [322] "117,189,920,212"    "117,227,769,792"    "118,133,634,072"
##  [325] "118,358,489,958"    "119,064,708,328"    "119,162,172,468"
##  [328] "119,791,683,308"    "12,030,023,548"     "12,038,829,246"
##  [331] "12,044,212,904"     "12,045,631,093"     "12,059,201,242"
##  [334] "12,129,642,296"     "12,130,252,200"     "12,138,486,532"
##  [337] "12,232,463,656"     "12,252,498,921"     "12,267,175,481"
##  [340] "12,274,928,000,000" "12,282,533,600"     "12,304,115,000"
##  [343] "12,354,820,144"     "12,502,013,400"     "12,552,071,367"
##  [346] "12,611,087,031"     "12,664,165,103"     "12,664,190,300"
##  [349] "12,736,856,828"     "12,803,445,934"     "12,881,352,688"
##  [352] "12,899,156,991"     "12,941,297,376"     "12,944,114,736"
##  [355] "12,983,235,568"     "12,994,310,400"     "120,579,072,751"
##  [358] "120,661,220,336"    "121,069,378,150"    "121,338,622,025"
##  [361] "121,545,880,984"    "121,600,818,310"    "121,872,464,483"
##  [364] "122,210,716,310"    "122,629,812,841"    "122,879,042,002"
##  [367] "122,964,812,046"    "123,533,036,668"    "123,981,736,420"
##  [370] "124,168,442,860"    "125,122,306,346"    "125,539,893,127"
##  [373] "125,816,640,421"    "126,206,817,196"    "126,392,308,498"
##  [376] "126,668,932,160"    "126,833,123,354"    "126,864,966,909"
##  [379] "127,131,461,120"    "127,417,688,056"    "127,465,545,493"
##  [382] "127,586,973,492"    "127,856,647,108"    "127,866,490,222"
##  [385] "127,945,379,258"    "129,250,111,857"    "129,440,993,789"
##  [388] "13,051,886,552"     "13,063,422,619"     "13,093,726,000,000"
##  [391] "13,099,013,836"     "13,150,166,755"     "13,153,016,531"
##  [394] "13,191,645,686"     "13,229,247,948"     "13,243,892,200"
##  [397] "13,280,275,123"     "13,310,567,803"     "13,350,468,917"
##  [400] "13,360,607,918"     "13,495,075,366"     "13,606,494,599"
##  [403] "13,617,405,420"     "13,677,622,222"     "13,686,329,890"
##  [406] "13,693,981,200"     "13,722,824,251"     "13,724,810,900"
##  [409] "13,760,374,488"     "13,794,910,634"     "13,855,888,000,000"
##  [412] "13,890,828,708"     "13,897,738,375"     "13,945,431,882"
##  [415] "13,948,892,216"     "13,972,676,841"     "13,993,546,732"
##  [418] "130,133,845,771"    "130,593,960,612"    "130,706,147,871"
##  [421] "130,838,040,068"    "130,922,638,689"    "132,099,404,608"
##  [424] "132,339,311,284"    "133,105,805,928"    "133,279,679,483"
##  [427] "133,339,397,080"    "133,441,612,247"    "133,503,411,376"
##  [430] "133,936,359,591"    "134,199,346,405"    "134,228,697,534"
##  [433] "134,300,851,255"    "135,215,704,419"    "135,225,868,315"
##  [436] "135,445,033,199"    "136,013,155,905"    "136,191,353,468"
##  [439] "136,280,689,891"    "136,361,854,808"    "136,631,966,609"
##  [442] "136,878,366,230"    "137,316,087,308"    "137,774,361,015"
##  [445] "139,552,983,249"    "139,850,794,387"    "14,006,088,297"
##  [448] "14,075,616,789"     "14,128,408,566"     "14,195,623,425"
##  [451] "14,278,357,284"     "14,292,008,745"     "14,307,509,839"
##  [454] "14,321,878,795"     "14,373,269,156"     "14,418,739,000,000"
##  [457] "14,434,619,982"     "14,439,910,353"     "14,477,635,000,000"
##  [460] "14,594,249,023"     "14,630,992,956"     "14,655,404,433"
##  [463] "14,698,001,400"     "14,718,582,000,000" "14,732,689,535"
##  [466] "14,744,603,774"     "14,756,846,154"     "14,932,024,169"
##  [469] "14,949,514,585"     "14,964,372,000,000" "14,988,971,211"
##  [472] "140,118,140,455"    "140,263,387,027"    "140,782,064,609"
##  [475] "141,352,368,715"    "141,517,648,888"    "142,009,922,306"
##  [478] "142,137,319,588"    "142,462,925,574"    "142,540,728,958"
```

```
##  [481] "142,719,009,901"    "143,157,600,025"    "144,428,172,835"
##  [484] "145,641,704"        "145,861,612,826"    "146,133,338,196"
##  [487] "146,566,264,837"    "146,583,831,538"    "146,683,499,006"
##  [490] "147,395,833,333"    "147,607,982,695"    "147,797,218,201"
##  [493] "148,047,348,241"    "149,359,918,060"    "15,002,106,518"
##  [496] "15,013,381,700"     "15,091,913,884"     "15,222,012,660"
##  [499] "15,239,278,100"     "15,314,143,988"     "15,431,288,006"
##  [502] "15,517,926,000,000" "15,548,321,544"     "15,656,327,860"
##  [505] "15,674,835,615"     "15,749,753,805"     "15,794,972,847"
##  [508] "15,846,474,596"     "15,913,363,335"     "15,929,902,138"
##  [511] "15,968,726,862"     "15,982,282,462"     "15,999,886,400"
##  [514] "150,180,268,649"    "150,194,077,688"    "152,027,402,450"
##  [517] "152,280,653,544"    "152,451,923,077"    "152,586,154,514"
##  [520] "153,830,947,017"    "153,966,916,311"    "154,027,536,232"
##  [523] "154,165,219,812"    "154,788,024,806"    "155,463,807,113"
##  [526] "156,162,311,732"    "157,998,423,132"    "159,942,880,457"
##  [529] "16,065,740,778"     "16,116,843,146"     "16,140,047,012"
##  [532] "16,155,255,000,000" "16,276,456,428"     "16,280,991,736"
##  [535] "16,330,814,180"     "16,374,393,900"     "16,452,201,101"
##  [538] "16,486,900,187"     "16,504,795,453"     "16,509,305,828"
##  [541] "16,536,535,647"     "16,691,490,114"     "16,691,517,000,000"
##  [544] "16,870,817,135"     "16,922,504,045"     "16,942,247,374"
##  [547] "16,963,625,016"     "16,988,535,268"     "160,846,667"
##  [550] "161,354,369,893"    "162,286,003,693"    "162,590,146,096"
##  [553] "162,631,412,509"    "163,159,671,670"    "163,167,853,538"
##  [556] "163,517,783,497"    "164,158,800,460"    "164,285,112,867"
##  [559] "164,641,483,516"    "164,657,930,453"    "164,791,416,350"
##  [562] "164,964,195,260"    "165,100,094,595"    "166,463,386,663"
##  [565] "166,658,327,827"    "167,728,444"        "167,775,274,725"
##  [568] "168,334,601,260"    "168,461,998,741"    "168,887,539"
##  [571] "169,278,552,851"    "169,717,677,901"    "17,011,750,900"
##  [574] "17,037,037,037"     "17,048,647,568"     "17,078,465,982"
##  [577] "17,149,094,590"     "17,155,421,053"     "17,195,867,540"
##  [580] "17,304,033,021"     "17,362,857,684"     "17,422,375,000"
##  [583] "17,427,609,000,000" "17,474,647,792"     "17,538,461,538"
##  [586] "17,601,616,000"     "17,658,408,759"     "17,702,885,394"
##  [589] "17,790,026,222"     "17,827,791,321"     "17,903,681,693"
##  [592] "17,986,886,200"     "171,071,106,095"    "171,315,639,983"
##  [595] "171,664,638,717"    "171,885,598,583"    "172,389,498,445"
##  [598] "172,611,845,342"    "172,895,476,153"    "173,537,588,008"
##  [601] "173,605,968,179"    "173,794,177,725"    "174,003,247,439"
##  [604] "174,070,025,009"    "174,161,142,454"    "174,195,136,253"
##  [607] "175,256,916,996"    "175,580,630"        "175,781,379,051"
##  [610] "175,933,642,292"    "176,192,886,551"    "176,992,000,955"
##  [613] "177,911,101,680"    "177,965,224,621"    "178,635,160,297"
##  [616] "178,706,717,753"    "179,638,496,279"    "179,981,288,567"
##  [619] "179,992,405,832"    "18,000,000,000"     "18,094,238,119"
##  [622] "18,120,714,000,000" "18,141,666,300"     "18,284,194,680"
##  [625] "18,291,990,619"     "18,318,412,251"     "18,325,791,415"
##  [628] "18,327,764,882"     "18,369,070,085"     "18,409,364,147"
##  [631] "18,447,922,400"     "18,504,130,753"     "18,505,053,191"
##  [634] "18,529,767,934"     "18,558,343,508"     "18,569,292,305"
##  [637] "18,644,723,861"     "18,702,802,395"     "18,703,146,375"
##  [640] "18,802,576,988"     "18,881,765,437"     "18,891,048,819"
```

```
##  [643] "18,938,717,359"     "180,190,994,861"    "180,302,412,231"
##  [646] "180,429,286,796"    "180,617,018,380"    "181,977,476,217"
##  [649] "183,035,154,107"    "183,144,164,357"    "184,367,381,749"
##  [652] "184,388,432,149"    "184,472,557,415"    "185,006,961,302"
##  [655] "186,829,940,546"    "186,833,516,484"    "187,632,400,366"
##  [658] "187,805,922,349"    "189,041,159,344"    "189,227,050,760"
##  [661] "189,318,499,954"    "189,719,984,268"    "19,138,296,376"
##  [664] "19,175,196,446"     "19,271,523,179"     "19,288,827,159"
##  [667] "19,297,663,097"     "19,395,491,993"     "19,490,936,349"
##  [670] "19,579,457,966"     "19,645,272,636"     "19,652,492,637"
##  [673] "19,676,167,240"     "19,731,912,494"     "19,952,156,475"
##  [676] "190,379,720,927"    "190,521,263,343"    "190,785,204,764"
##  [679] "191,549,024,911"    "192,225,881,688"    "192,408,387,762"
##  [682] "192,626,507,972"    "193,870,350,137"    "195,078,126,346"
##  [685] "195,305,084,919"    "195,418,347,153"    "195,541,761,243"
##  [688] "195,905,767,669"    "196,768,065,557"    "196,799,778,883"
##  [691] "196,964,195,387"    "197,337,879,195"    "198,298,498,021"
##  [694] "198,680,637,255"    "198,727,747,253"    "198,737,095,012"
##  [697] "199,493,490,983"    "199,590,775,189"    "2,007,109,829"
##  [700] "2,015,731,029"      "2,019,474,244"      "2,021,301,676"
##  [703] "2,023,318,252"      "2,028,295,455"      "2,038,395,102,041"
##  [706] "2,046,189,032"      "2,051,661,732,060"  "2,061,323,854"
##  [709] "2,063,662,665,172"  "2,068,526,522"      "2,068,555,542,411"
##  [712] "2,072,823,157,060"  "2,073,234,418"      "2,078,953,333,674"
##  [715] "2,079,136,081,310"  "2,100,223,150"      "2,115,742,488,205"
##  [718] "2,118,467,913"      "2,118,574,772"      "2,123,130,870,382"
##  [721] "2,125,058,244,243"  "2,126,000,000"      "2,130,491,320,659"
##  [724] "2,143,484,488"      "2,151,295,182"      "2,151,732,868,243"
##  [727] "2,156,624,900"      "2,185,160,183,384"  "2,191,096,860"
##  [730] "2,196,126,103,718"  "2,199,957,383,337"  "2,203,053,380,783"
##  [733] "2,205,966,011,812"  "2,208,871,646,203"  "2,210,256,976,945"
##  [736] "2,211,535,312"      "2,218,689,375,141"  "2,228,279,330"
##  [739] "2,243,225,519,618"  "2,256,838,858"      "2,257,174,481"
##  [742] "2,261,969,870"      "2,271,646,188"      "2,273,225,042"
##  [745] "2,276,292,404,601"  "2,297,128,039,058"  "2,315,935,753"
##  [748] "2,316,562,500"      "2,318,593,651,988"  "2,320,699,900"
##  [751] "2,331,005,587"      "2,331,358,820"      "2,335,124,988"
##  [754] "2,359,903,108"      "2,376,335,048"      "2,379,281,768"
##  [757] "2,382,825,985,356"  "2,390,729,163,615"  "2,398,555,474,185"
##  [760] "2,421,474,860"      "2,424,499,009"      "2,441,173,394,730"
##  [763] "2,450,084,970"      "2,450,686,660"      "2,455,993,625,159"
##  [766] "2,460,248,026"      "2,465,188,674,415"  "2,467,703,911"
##  [769] "2,472,806,919,902"  "2,482,228,440"      "2,498,932,961"
##  [772] "2,503,665,193,657"  "2,505,733,634,312"  "2,513,870,587"
##  [775] "2,520,701,818,182"  "2,532,808,573,157"  "2,547,163,582"
##  [778] "2,561,118,608"      "2,570,833,333"      "2,576,024,116"
##  [781] "2,584,463,687"      "2,588,176,055"      "2,591,620,035,485"
##  [784] "2,605,688,065"      "2,616,201,578,192"  "2,619,700,404,733"
##  [787] "2,623,726,257"      "2,626,380,435"      "2,642,609,548,930"
##  [790] "2,653,480,001"      "2,657,213,249,384"  "2,662,085,168,499"
##  [793] "2,674,000,000"      "2,683,825,225,093"  "2,690,222,283,968"
##  [796] "2,692,612,695,492"  "2,693,731,866"      "2,707,123,772"
##  [799] "2,709,178,327"      "2,713,999,900"      "2,721,904,403"
##  [802] "2,727,507,213"      "2,739,818,680,930"  "2,750,041,434"
```

```
##  [805] "2,754,566,176"      "2,774,351,760"      "2,791,960,894"
##  [808] "2,800,024,314"      "2,807,061,009"      "2,811,077,725,704"
##  [811] "2,819,245,095,605"  "2,834,168,889"      "2,852,165,760,630"
##  [814] "2,856,890,681"      "2,861,408,170,265"  "2,861,410,272,354"
##  [817] "2,861,562,266"      "2,870,624,636"      "2,885,570,309,161"
##  [820] "2,890,564,338,235"  "2,904,662,605"      "2,918,382,891,460"
##  [823] "2,936,612,022"      "2,939,845,038"      "2,966,234,106"
##  [826] "2,990,006,534"      "2,998,570,147"      "20,030,528,043"
##  [829] "20,079,363,626"     "20,153,871,681"     "20,209,122,027"
##  [832] "20,214,385,965"     "20,249,294,118"     "20,283,783,700"
##  [835] "20,289,200,000"     "20,304,098,101"     "20,342,201,356"
##  [838] "20,374,307,047"     "20,403,713,461"     "20,515,543,039"
##  [841] "20,632,090,909"     "20,749,140,606"     "20,776,669,467"
##  [844] "20,875,387,068"     "20,897,630,201"     "20,898,788,417"
##  [847] "20,982,685,981"     "201,924,270,316"    "203,535,242,742"
##  [850] "204,068,257,818"    "204,436,015,421"    "206,179,982,164"
##  [853] "206,224,725,275"    "206,430,841,502"    "207,376,427,021"
##  [856] "207,416,494,642"    "207,419,370,241"    "207,477,857,919"
##  [859] "207,818,330,724"    "207,850,624"        "207,998,568,866"
##  [862] "208,740,444"        "209,402,444,996"    "209,880,963"
##  [865] "209,950,792,713"    "21,035,357,833"     "21,188,704,081"
##  [868] "21,250,839,258"     "21,263,514,833"     "21,272,418,792"
##  [871] "21,273,055,398"     "21,295,486,383"     "21,295,984,200"
##  [874] "21,319,946,740"     "21,386,153,000"     "21,442,619,681"
##  [877] "21,447,021,570"     "21,480,023,017"     "21,588,170,498"
##  [880] "21,642,304,046"     "21,650,000,000"     "21,730,000,000"
##  [883] "21,774,273,832"     "21,776,609,772"     "21,917,706,491"
##  [886] "21,977,401,900"     "21,988,444,444"     "211,637,816,539"
##  [889] "211,650,764,830"    "212,790,348,405"    "213,377,771,504"
##  [892] "213,605,065,703"    "214,875,344,910"    "215,840,354,627"
##  [895] "216,368,178,659"    "216,552,502,823"    "217,185,787,343"
##  [898] "217,513,049,292"    "218,095,997,085"    "218,259,904,402"
##  [901] "218,537,551,220"    "22,009,300,000"     "22,085,858,243"
##  [904] "22,125,435,372"     "22,135,245,413"     "22,152,694,162"
##  [907] "22,157,948,396"     "22,165,932,063"     "22,235,929,043"
##  [910] "22,237,061,730"     "22,255,407,685"     "22,365,734,482"
##  [913] "22,367,254,865"     "22,387,561,845"     "22,555,941,176"
##  [916] "22,566,956,982"     "22,583,157,895"     "22,585,841,200"
##  [919] "22,600,431,878"     "22,649,930,576"     "22,689,994,990"
##  [922] "22,708,673,337"     "22,823,255,802"     "22,938,218,085"
##  [925] "221,415,572,820"    "221,758,486,880"    "221,951,354,762"
##  [928] "222,116,541,865"    "222,942,790,435"    "222,977,046,516"
##  [931] "224,143,083,707"    "225,571,853,194"    "225,924,679,921"
##  [934] "226,073,492,966"    "226,079,963,712"    "227,850,000"
##  [937] "227,948,349,666"    "228,752,436,372"    "228,788,617,202"
##  [940] "228,937,347,866"    "229,434,519"        "229,629,822,122"
##  [943] "23,043,864,510"     "23,170,239,901"     "23,289,671,102"
##  [946] "23,359,294,149"     "23,386,945,597"     "23,409,027,476"
##  [949] "23,410,572,634"     "23,563,576,758"     "23,616,328,816"
##  [952] "23,678,012,697"     "23,757,368,290"     "23,822,087,053"
##  [955] "23,941,391,391"     "23,965,275,996"     "23,969,823,010"
##  [958] "23,983,945,191"     "232,085,535,065"    "232,464,833,066"
##  [961] "232,557,260,817"    "233,609,522,091"    "233,821,670,544"
##  [964] "235,699,941,793"    "235,718,586,901"    "236,038,384,442"
```

```
##  [967] "236,311,336,482"    "236,357,519"        "236,421,782,178"
##  [970] "236,634,552,078"    "236,741,715,015"    "237,029,579,261"
##  [973] "237,250,948,791"    "237,841,968,680"    "237,904,919,845"
##  [976] "238,303,443,425"    "239,018,536,582"    "239,389,340,720"
##  [979] "239,862,011,450"    "239,917,320,967"    "24,025,800,000"
##  [982] "24,077,470,572"     "24,084,572,491"     "24,194,039,256"
##  [985] "24,432,884,442"     "24,468,324,000"     "24,595,319,574"
##  [988] "24,636,593,223"     "24,848,483,838"     "24,861,483,281"
##  [991] "24,881,135,586"     "24,906,939,560"     "24,950,895,141"
##  [994] "240,169,336,162"    "240,365,259"        "240,521,260,988"
##  [997] "240,923,926"        "241,038,283,063"    "241,918,791,123"
## [1000] "243,526,047,717"    "243,745,748,819"    "243,982,437,871"
## [1003] "243,999,477,738"    "244,895,101,712"    "245,670,666,639"
## [1006] "246,194,938,751"    "247,783,001,865"    "247,799,815,768"
## [1009] "249,267,040"        "25,017,300,000"     "25,041,372,222"
## [1012] "25,099,681,461"     "25,121,666,667"     "25,137,153,149"
## [1015] "25,155,888,600"     "25,226,393,197"     "25,337,226,971"
## [1018] "25,366,200,000"     "25,385,928,198"     "25,432,144,406"
## [1021] "25,433,011,405"     "25,562,251,656"     "25,710,877,660"
## [1024] "25,713,271,277"     "25,733,043,137"     "25,736,331,247"
## [1027] "25,941,929,462"     "25,942,622,951"     "25,957,970,922"
## [1030] "250,638,463,467"    "251,373,036,671"    "251,499,027,508"
## [1033] "252,251,992,029"    "253,547,358,747"    "254,813,599,459"
## [1036] "255,039,560,740"    "255,102,252,843"    "255,384,615,385"
## [1039] "256,706,466,091"    "257,296,579,579"    "257,440,000,000"
## [1042] "257,671,413,751"    "257,916,133,424"    "258,031,750,000"
## [1045] "258,099,014,042"    "258,154,283,909"    "258,567,751,143"
## [1048] "258,860,436,665"    "259,708,496,267"    "259,802,012,617"
## [1051] "26,040,229,793"     "26,125,575,942"     "26,169,854,045"
## [1054] "26,224,622,451"     "26,252,007,830"     "26,362,894,737"
## [1057] "26,385,800,000"     "26,743,874,206"     "26,878,499,206"
## [1060] "26,932,728,899"     "26,972,863,394"     "260,202,429,150"
## [1063] "260,584,090,571"    "260,601,911,536"    "261,628,548,084"
## [1066] "261,695,778,781"    "262,007,590,450"    "262,709,948,091"
## [1069] "262,942,650,544"    "263,926,220,333"    "264,051,981,551"
## [1072] "264,357,494,659"    "264,467,308,109"    "264,477,727,279"
## [1075] "266,800,462,899"    "267,122,320,057"    "267,327,630"
## [1078] "268,137,224,730"    "268,696,750,000"    "269,917,519,847"
## [1081] "269,980,111,643"    "27,116,635,600"     "27,118,476,174"
## [1084] "27,191,689,008"     "27,211,377,225"     "27,427,161,523"
## [1087] "27,493,591,484"     "27,660,149,541"     "27,821,913,815"
## [1090] "27,839,460,964"     "27,870,257,894"     "27,981,896,948"
## [1093] "270,362,531,377"    "270,847,937,645"    "270,953,116,950"
## [1096] "271,811,088,781"    "272,055,499,941"    "272,149,750,000"
## [1099] "272,609,288,690"    "273,317,737,047"    "273,674,236,773"
## [1102] "275,020,023,436"    "275,966,926,379"    "276,550,000"
## [1105] "278,098,778"        "278,384,332,694"    "278,631,271,391"
## [1108] "279,349,355,714"    "28,101,000,000"     "28,119,996,053"
## [1111] "28,161,200,000"     "28,162,053,027"     "28,223,552,825"
## [1114] "28,267,410,543"     "28,364,615,242"     "28,448,326,757"
## [1117] "28,548,945,000"     "28,748,960,471"     "28,776,595,745"
## [1120] "28,965,906,502"     "28,998,684,211"     "280,312,318,915"
## [1123] "281,358,175,896"    "281,710,095,725"    "282,884,912,894"
## [1126] "283,523,000,000"    "283,742,493,042"    "284,203,750,000"
```

```
## [1129] "286,519,135,327"    "287,018,184,638"    "287,099,991,517"
## [1132] "287,530,508,431"    "287,797,822,093"    "288,103,936,773"
## [1135] "289,567,323,481"    "289,689,704"        "289,787,338,325"
## [1138] "289,912,492,345"    "29,114,875,622"     "29,233,333,333"
## [1141] "29,440,287,600"     "29,557,325,056"     "29,636,617,750"
## [1144] "29,675,502,270"     "29,697,448,108"     "29,828,899,206"
## [1147] "29,885,685,143"     "290,440,148"        "290,673,681,684"
## [1150] "291,383,081,232"    "291,519,591,533"    "292,489,185,194"
## [1153] "292,859,000,000"    "294,084,112,393"    "295,045,152,021"
## [1156] "297,216,730,669"    "298,948,250,000"    "299,033,511,000"
## [1159] "299,093,839,690"    "299,317,645,143"    "299,361,576,558"
## [1162] "3,002,446,368,084" "3,006,725,015"      "3,006,988,217"
## [1165] "3,010,982,414"      "3,021,910,217"      "3,022,827,781,881"
## [1168] "3,052,467,522"      "3,054,914,166,263" "3,057,453,483"
## [1171] "3,071,683,013,179" "3,074,359,743,898" "3,092,000,000"
## [1174] "3,094,915,506"      "3,102,741,451"      "3,116,650,000"
## [1177] "3,121,500,000"      "3,132,817,652,848" "3,169,600,000"
## [1180] "3,176,749,593"      "3,179,225,949"      "3,189,539,641"
## [1183] "3,200,539,816"      "3,208,098,919"      "3,219,487,747"
## [1186] "3,224,267,548"      "3,259,000,000"      "3,263,368,410"
## [1189] "3,274,850,000"      "3,286,987,552"      "3,313,739,674"
## [1192] "3,314,898,292"      "3,375,611,100,742" "3,392,021,011"
## [1195] "3,405,050,612"      "3,414,760,915"      "3,417,094,562,649"
## [1198] "3,418,005,001,389" "3,429,000,000"      "3,439,716,562"
## [1201] "3,439,931,907"      "3,439,953,462,907" "3,510,540,809"
## [1204] "3,514,350,000"      "3,523,185,920"      "3,532,969,035"
## [1207] "3,534,803,922"      "3,543,983,909,148" "3,558,137,040"
## [1210] "3,570,271,558"      "3,584,420,077,101" "3,609,000,000"
## [1213] "3,613,500,117"      "3,632,043,908"      "3,651,861,702"
## [1216] "3,680,710,375"      "3,690,328,964"      "3,702,393,617"
## [1219] "3,705,372,039"      "3,752,365,607,148" "3,752,513,503,278"
## [1222] "3,757,698,281,118" "3,774,530,616"      "3,778,316,380"
## [1225] "3,802,566,171"      "3,828,310,735"      "3,875,409,836"
## [1228] "3,890,606,893,347" "3,897,300,000"      "3,908,809,463,464"
## [1231] "3,923,637,971"      "3,930,518,366"      "3,962,710,163"
## [1234] "3,972,012,571"      "3,973,027,397"      "30,074,440,483"
## [1237] "30,121,879,435"     "30,216,060,233"     "30,231,249,362"
## [1240] "30,254,677,297"     "30,354,434,553"     "30,366,213,119"
## [1243] "30,415,095,887"     "30,423,573,842"     "30,429,803,651"
## [1246] "30,562,361,123"     "30,565,400,000"     "30,603,919,000"
## [1249] "30,604,668,357"     "30,607,991,862"     "30,612,932,876"
## [1252] "30,703,017,450"     "30,749,308,511"     "30,833,699,703"
## [1255] "30,881,166,852"     "30,901,399,261"     "300,757,889"
## [1258] "300,904,221,505"    "301,298,464,861"    "301,416,810,214"
## [1261] "304,097,759,674"    "304,454,327,499"    "306,125,173,853"
## [1264] "306,602,673,980"    "308,416,982,140"    "308,722,079,938"
## [1267] "309,812,185"        "31,261,527,363"     "31,335,013,752"
## [1270] "31,493,987,642"     "31,580,639,045"     "31,660,911,277"
## [1273] "31,682,400,000"     "31,734,065,934"     "310,160,444"
## [1276] "310,838,014,621"    "311,528,948,848"    "311,539,499,645"
## [1279] "314,267,667,675"    "314,454,015,372"    "315,474,615,739"
## [1282] "315,974,418,605"    "316,008,481"        "317,741,039,198"
## [1285] "317,882,187,037"    "318,497,936,901"    "319,002,821,670"
## [1288] "319,423,370,134"    "319,998,987"        "32,287,031,000"
```

```
## [1291] "32,432,858,000"     "32,539,547,872"     "321,241,396,034"
## [1294] "321,995,350,347"    "322,201,314,424"    "323,320,449,906"
## [1297] "324,878,105,053"    "325,111,815"        "325,358,292,128"
## [1300] "325,734,233,313"    "327,148,899,962"    "33,050,343,783"
## [1303] "33,113,887,818"     "33,387,712,766"     "33,590,500,000"
## [1306] "33,689,223,673"     "33,961,142,322"     "330,000,252,153"
## [1309] "330,091,688,350"    "330,397,381,998"    "331,108,912,605"
## [1312] "331,489,704"        "332,976,484,578"    "335,415,156,702"
## [1315] "335,998,557,270"    "336,708,419"        "337,174,852"
## [1318] "34,083,532,195"     "34,113,107,086"     "34,130,122,491"
## [1321] "34,470,227,454"     "34,590,052,812"     "34,630,430,000"
## [1324] "34,658,113,497"     "34,685,281,848"     "34,686,224,300"
## [1327] "34,748,508,332"     "34,890,772,742"     "34,894,411,352"
## [1330] "34,942,489,684"     "341,105,009,515"    "342,172,519"
## [1333] "342,617,007,104"    "343,584,385,594"    "344,003,209,696"
## [1336] "344,748,646,558"    "345,110,438,692"    "345,424,664,369"
## [1339] "347,770,000"        "349,037,818,106"    "35,083,608,131"
## [1342] "35,164,210,526"     "35,596,016,664"     "35,799,628,571"
## [1345] "35,838,588,170"     "35,901,200,000"     "350,051,111,253"
## [1348] "350,904,575,292"    "352,914,820,747"    "352,993,633,221"
## [1351] "353,361,056,080"    "354,460,802,549"    "358,330,385,840"
## [1354] "36,152,027,893"     "36,183,003,978"     "36,346,974,008"
## [1357] "36,373,307,085"     "36,591,661,000"     "36,922,456,000"
## [1360] "360,073,909,244"    "361,558,037,110"    "363,150,000"
## [1363] "363,609,163,462"    "364,756,499,451"    "366,541,091"
## [1366] "366,829,390,479"    "366,911,444"        "367,216,364,716"
## [1369] "368,281,379"        "369,659,700,376"    "37,120,517,694"
## [1372] "37,160,332,465"     "37,182,938,696"     "37,268,635,287"
## [1375] "37,347,394,603"     "37,438,527,800"     "37,440,673,478"
## [1378] "37,662,075,750"     "37,711,864,407"     "37,733,994,976"
## [1381] "37,818,134,253"     "37,834,793,730"     "370,818,747,397"
## [1384] "370,885,026,074"    "373,619,852"        "374,241,351,752"
## [1387] "374,291,430,318"    "375,138,723,325"    "375,298,134,440"
## [1390] "377,437,927,312"    "378,195,716,714"    "378,215,090,694"
## [1393] "379,779,390"        "38,009,344,577"     "38,137,545,245"
## [1396] "38,203,000,000"     "38,724,945,368"     "38,749,715,722"
## [1399] "38,900,692,712"     "38,995,454,545"     "380,191,881,860"
## [1402] "381,705,425,302"    "382,065,930,308"    "386,302,839,274"
## [1405] "386,622,457,580"    "386,663,139,403"    "387,365,750,529"
## [1408] "388,300,000"        "388,395,162,301"    "388,691,445,387"
## [1411] "389,042,298,377"    "39,136,893,345"     "39,197,543,860"
## [1414] "39,212,550,050"     "39,285,385,088"     "39,332,770,929"
## [1417] "39,460,357,731"     "39,540,080,200"     "39,587,732,029"
## [1420] "39,690,630,000"     "39,738,180,077"     "390,719,148"
## [1423] "390,799,991,147"    "391,719,993,757"    "391,892,746,545"
## [1426] "394,163,688,621"    "394,486,709,920"    "395,077,301,248"
## [1429] "395,531,066,563"    "396,270,000"        "396,332,702,639"
## [1432] "396,582,263"        "397,558,094,270"    "398,637,741"
## [1435] "398,899,138,574"    "4,032,509,760,873"  "4,040,345,933"
## [1438] "4,060,758,804"      "4,071,219,198"      "4,115,116,279,070"
## [1441] "4,127,313,818"      "4,159,330,370"      "4,169,664,285"
## [1444] "4,187,367,602"      "4,246,600,000"      "4,291,172,816"
## [1447] "4,303,544,259,843"  "4,306,192,436"      "4,323,058,824"
## [1450] "4,331,870,648"      "4,346,734,000,000"  "4,368,398,048"
```

```
## [1453] "4,373,665,146"      "4,394,977,752,878"  "4,404,970,059"
## [1456] "4,414,732,843,544"  "4,421,943,910"      "4,422,276,622"
## [1459] "4,434,050,000"      "4,435,078,648"      "4,445,658,071,222"
## [1462] "4,446,396,218"      "4,454,143,876,947"  "4,461,650,000"
## [1465] "4,461,978,499"      "4,496,852,073"      "4,515,264,514,431"
## [1468] "4,523,750,000"      "4,530,377,224,970"  "4,536,538,211"
## [1471] "4,545,674,528"      "4,562,078,822,335"  "4,581,222,442"
## [1474] "4,582,555,125"      "4,590,155,000,000"  "4,592,224,067"
## [1477] "4,599,970,618"      "4,607,300,000"      "4,612,500,000"
## [1480] "4,656,350,000"      "4,660,900,000"      "4,669,488,516"
## [1483] "4,689,832,690"      "4,690,062,255"      "4,732,017,873"
## [1486] "4,738,559,685"      "4,746,109,767"      "4,755,410,630,912"
## [1489] "4,767,303,154"      "4,794,357,795"      "4,794,444,444"
## [1492] "4,796,628,461"      "4,797,777,778"      "4,815,148,854,362"
## [1495] "4,817,542,204"      "4,833,712,542,207"  "4,850,413,536,038"
## [1498] "4,870,217,000,000"  "4,887,519,660,745"  "4,907,039,384,470"
## [1501] "4,947,205,860"      "4,980,000,000"      "4,991,350,458"
## [1504] "40,119,073,326"     "40,274,204,595"     "40,284,481,652"
## [1507] "40,289,556,656"     "40,338,594,862"     "40,429,734,400"
## [1510] "40,477,403,220"     "40,614,350,197"     "40,716,836,998"
## [1513] "40,742,313,861"     "40,946,424,370"     "400,172,297,861"
## [1516] "400,302,731,411"    "400,599,250,000"    "401,082,621,083"
## [1519] "401,399,422,443"    "404,600,000"        "404,926,534,140"
## [1522] "407,339,361,696"    "408,689,353,999"    "409,425,234,155"
## [1525] "409,813,197,842"    "41,239,551,378"     "41,338,595,381"
## [1528] "41,507,085,000"     "41,508,609,233"     "41,574,530,816"
## [1531] "41,636,005,955"     "41,883,241,472"     "41,976,002,704"
## [1534] "410,975,595,310"    "411,755,164,833"    "412,199,006,099"
## [1537] "412,807,259,996"    "413,050,000"        "413,630,538,018"
## [1540] "413,799,990"        "415,034,227,218"    "415,518,112"
## [1543] "416,878,162,441"    "42,105,263,158"     "42,262,697,840"
## [1546] "42,392,896,031"     "42,414,308,117"     "42,616,653,300"
## [1549] "42,643,836,075"     "42,647,331,000"     "42,756,020,707"
## [1552] "42,773,029,835"     "42,848,195,256"     "420,032,121,656"
## [1555] "420,333,333,333"    "423,627,422,092"    "425,595,310,000"
## [1558] "426,573,601,790"    "429,063,549,984"    "429,130,952,709"
## [1561] "429,657,033,108"    "43,072,415,017"     "43,096,746,122"
## [1564] "43,151,647,003"     "43,160,392,124"     "43,476,873,413"
## [1567] "43,524,210,526"     "430,040,370"        "430,068,712,972"
## [1570] "430,294,287,388"    "430,347,770,732"    "430,921,192,375"
## [1573] "431,120,310,089"    "431,316,742,081"    "432,476,116,419"
## [1576] "434,568,007,513"    "435,083,713,851"    "437,798,577,640"
## [1579] "438,794,778"        "439,796,160,379"    "44,170,562,822"
## [1582] "44,210,806,366"     "44,411,476,557"     "44,530,494,505"
## [1585] "44,558,077,827"     "441,885,415,806"    "441,975,282,335"
## [1588] "442,273,433"        "444,720,750"        "445,704,575,163"
## [1591] "446,528,959,649"    "448,300,000"        "45,277,399,814"
## [1594] "45,340,835,000"     "45,416,358,502"     "45,474,442,836"
## [1597] "45,493,075,684"     "45,519,034,244"     "45,519,650,911"
## [1600] "45,599,994,000"     "45,915,191,189"     "45,964,327,559"
## [1603] "450,725,816,043"    "455,039,563,133"    "455,602,962,225"
## [1606] "459,469,074"        "46,174,557,556"     "46,352,802,766"
## [1609] "46,417,340,375"     "46,418,916,500"     "46,466,728,667"
## [1612] "46,473,128,286"     "46,659,796,773"     "46,712,018,141"
```

```
## [1615] "46,731,767,494"     "46,802,044,000"     "46,919,625"
## [1618] "460,290,556,901"    "461,883,444"        "461,951,782,000"
## [1621] "462,554,432,624"    "465,368,906,456"    "466,294,700,058"
## [1624] "47,290,180,589"     "47,310,623,887"     "47,515,189"
## [1627] "47,554,674,591"     "47,654,841,113"     "47,737,955"
## [1630] "47,820,850,975"     "47,850,551,149"     "47,875,838,926"
## [1633] "471,821,105,940"    "477,355,617,456"    "479,321,460,551"
## [1636] "479,913,034,322"    "48,013,606,745"     "48,114,688,201"
## [1639] "48,116,256,926"     "48,187,039,000"     "48,187,667,853"
## [1642] "48,213,868,178"     "48,388,296,489"     "48,516,371,721"
## [1645] "48,770,466,839"     "481,706,333"        "481,806,296"
## [1648] "483,548,031,197"    "483,962,431"        "484,552,792,442"
## [1651] "485,248,229,337"    "487,816,328,342"    "488,377,689,565"
## [1654] "489,285,164,271"    "49,170,434,391"     "49,209,523,810"
## [1657] "49,259,526,053"     "49,279,585,355"     "49,297,773,130"
## [1660] "49,364,681,256"     "49,425,513,611"     "49,745,088,112"
## [1663] "49,847,128,533"     "49,881,433,991"     "49,904,928,335"
## [1666] "49,921,464,400"     "497,884,216,569"    "497,918,109,302"
## [1669] "498,831,558,926"    "499,281,148"        "499,338,534,779"
## [1672] "5,037,908,465,114" "5,066,240,419"      "5,068,000,000"
## [1675] "5,125,363,001"      "5,139,957,785"      "5,145,757,576"
## [1678] "5,155,717,056,271" "5,231,382,674,594" "5,240,606,061"
## [1681] "5,252,342,400"      "5,252,629,000,000" "5,272,617,196"
## [1684] "5,307,905,882"      "5,329,214,163"      "5,438,537,482"
## [1687] "5,439,552,941"      "5,449,116,304,981" "5,456,583,589"
## [1690] "5,477,895,475"      "5,588,533,007"      "5,609,831,328"
## [1693] "5,617,109,245"      "5,657,693,000,000" "5,685,774,809"
## [1696] "5,695,201,563"      "5,700,098,114,744" "5,707,720,391"
## [1699] "5,726,897,998"      "5,737,751,332"      "5,746,945,913"
## [1702] "5,759,537,726"      "5,788,368,511"      "5,813,399,300"
## [1705] "5,841,132,962"      "5,880,112,788"      "5,902,783,400"
## [1708] "5,918,469,800"      "5,970,044,666"      "5,977,560,877"
## [1711] "5,978,460,972"      "5,979,589,000,000" "50,078,767,903"
## [1714] "50,132,953,288"     "50,150,399,792"     "50,244,793,832"
## [1717] "50,388,454,861"     "50,453,577,898"     "50,535,438,696"
## [1720] "50,577,769,838"     "50,610,058,210"     "50,888,134,410"
## [1723] "500,360,816,828"    "500,413,483,109"    "500,736,065,605"
## [1726] "506,500,173,960"    "507,354,351,182"    "507,962,487,700"
## [1729] "51,007,777,000"     "51,264,390,116"     "51,290,792,018"
## [1732] "51,338,524,831"     "51,370,543,206"     "51,821,573,338"
## [1735] "51,884,481,410"     "510,229,136,227"    "513,753,818"
## [1738] "513,965,650,650"    "514,459,972,806"    "516,814,274,022"
## [1741] "517,962,962,963"    "518,625,897,173"    "52,156,414,979"
## [1744] "52,480,253,169"     "52,549,555,149"     "52,622,842,840"
## [1747] "52,742,800,000"     "52,976,344,929"     "520,444,185"
## [1750] "520,925,468,953"    "521,975,111"        "523,502,127,660"
## [1753] "523,649,481,762"    "524,234,322,597"    "526,319,673,732"
## [1756] "526,502,129,378"    "527,008,453,887"    "527,813,238,126"
## [1759] "528,832,185,770"    "529,121,577,320"    "53,212,476,812"
## [1762] "53,274,304,222"     "53,821,315,066"     "53,851,058,955"
## [1765] "53,872,425,917"     "53,903,022,339"     "530,163,281,575"
## [1768] "530,770,977,634"    "533,052,076,314"    "533,815,789,474"
## [1771] "535,101,248,776"    "535,172,778"        "54,086,400,000"
## [1774] "54,315,722,500"     "54,368,083,953"     "54,409,138,498"
```

```
## [1777] "54,832,578"          "540,336,926"        "540,874,934"
## [1780] "541,506,500,414"     "542,520,000"        "543,880,647,757"
## [1783] "545,179,584,720"     "545,982,375,701"    "55,154,160,816"
## [1786] "55,589,849,128"      "55,758,754,072"     "55,802,538,219"
## [1789] "55,849,686,539"      "55,985,506,499"     "550,728,667"
## [1792] "552,025,140,252"     "553,138,414,367"    "554,363,487,120"
## [1795] "556,130,926,913"     "557,503,074,772"    "558,111,997,497"
## [1798] "558,319,920,832"     "559,372,276,082"    "559,858,250"
## [1801] "56,227,696,195"      "56,338,028"         "56,352,797,354"
## [1804] "56,379,593,720"      "56,565,475,275"     "56,677,961,787"
## [1807] "56,731,990,232"      "561,633,125,840"    "563,109,663,291"
## [1810] "565,055,743,243"     "566,894,749"        "57,008,425,296"
## [1813] "57,123,671,734"      "57,166,037,102"     "57,222,490,769"
## [1816] "57,236,013,086"      "57,240,535,138"     "57,418,414,504"
## [1819] "57,531,233,351"      "57,629,518,806"     "57,690,453,461"
## [1822] "57,784,495,265"      "57,841,000,000"     "571,863,431,151"
## [1825] "573,817,719,109"     "574,870,405"        "575,598,537,070"
## [1828] "577,170,761,956"     "577,280,741"        "578,139,279,438"
## [1831] "578,742,001,488"     "58,085,856,019"     "58,603,900,000"
## [1834] "58,642,392,718"      "58,722,323,918"     "58,762,260,626"
## [1837] "58,986,997,917"      "580,863,700"        "583,782,977,866"
## [1840] "588,692,045,455"     "589,429,593"        "59,440,108,585"
## [1843] "59,775,697,061"      "59,829,574,391"     "59,867,801,205"
## [1846] "591,016,690,743"     "591,018,407"        "592,387,689,253"
## [1849] "593,929,550,908"     "594,749,285,413"    "595,402,616,547"
## [1852] "598,099,073,901"     "599,118,593"        "599,388,579,986"
## [1855] "6,016,168,896"       "6,043,694,330"      "6,062,780,269"
## [1858] "6,063,759,371"       "6,126,456,175"      "6,157,459,594,824"
## [1861] "6,174,043,000,000"   "6,183,776,596"      "6,197,766,119"
## [1864] "6,203,213,121,334"   "6,235,795,104"      "6,245,069,734"
## [1867] "6,265,844,252"       "6,283,803,256"      "6,294,803,497"
## [1870] "6,325,151,760"       "6,332,360,000"      "6,349,202,394"
## [1873] "6,357,615,894"       "6,372,905,073"      "6,384,451,606"
## [1876] "6,385,695,187"       "6,394,851,387"      "6,405,210,564"
## [1879] "6,410,852,596"       "6,525,676,264"      "6,539,299,000,000"
## [1882] "6,540,247,190"       "6,541,517,100"      "6,588,103,836"
## [1885] "6,605,139,933"       "6,621,010,372"      "6,678,178,340"
## [1888] "6,680,269,200"       "6,757,119,558"      "6,797,834,200"
## [1891] "6,807,365,898"       "6,808,982,521"      "6,827,665,300"
## [1894] "6,833,220,000"       "6,838,351,088"      "6,866,402,028"
## [1897] "6,878,718,000,000"   "6,921,264,132"      "6,976,080,331"
## [1900] "6,984,367,763"       "60,004,630,234"     "60,093,222,709"
## [1903] "60,600,056,659"      "60,644,572,348"     "60,763,483,146"
## [1906] "60,806,300,000"      "60,882,142,857"     "600,005,564"
## [1909] "600,232,874,043"     "601,452,653,181"    "602,860,000,000"
## [1912] "604,031,623,433"     "609,020,054,512"    "61,649,492,817"
## [1915] "61,701,800,000"      "61,739,352,212"     "61,757,788,945"
## [1918] "61,762,635,000"      "61,792,161,168"     "610,328,183,643"
## [1921] "610,930,037"         "611,904,253,806"    "612,939,685,081"
## [1924] "617,041,986,858"     "62,080,000,000"     "62,375,044,443"
## [1927] "62,519,686,000"      "62,697,540,107"     "62,703,143,057"
## [1930] "620,140,400"         "621,626,786"        "622,262,057"
## [1933] "624,337,145"         "625,975,838,926"    "628,546,387,972"
## [1936] "629,202,392,004"     "63,067,077,179"     "63,101,272"
```

```
## [1939] "63,767,597,194"     "63,918,039,320"     "631,813,279,407"
## [1942] "633,194,118,900"     "638,273,986,102"    "64,087,694,038"
## [1945] "64,328,000,000"      "64,867,483,193"     "64,883,060,726"
## [1948] "640,449,501"         "640,998,292,395"    "641,383,800"
## [1951] "644,639,902,581"     "65,334,841"         "65,648,559,903"
## [1954] "65,652,751,132"      "65,977,749,037"     "650,532,654,582"
## [1957] "651,833,333"         "652,825,364,726"    "654,314,350"
## [1960] "655,420,645,477"     "66,074,513,018"     "66,327,344,189"
## [1963] "66,490,372,105"      "66,515,377"         "66,627,729,311"
## [1966] "66,985,765,439"      "662,196,185"        "668,043,614,123"
## [1969] "669,316,654,017"     "67,254,174"         "67,523,642,262"
## [1972] "67,530,220,219"      "67,537,480"         "67,716,887,203"
## [1975] "674,922,481"         "676,082,654,641"    "676,129,407"
## [1978] "676,949,594"         "678,533,764,457"    "679,289,166,858"
## [1981] "68,990,000,000"      "680,520,724,062"    "681,225,963"
## [1984] "688,504,173,431"     "688,992,450"        "69,032,258"
## [1987] "69,208,400,000"      "69,222,626,263"     "69,386,774,475"
## [1990] "69,555,367,000"      "69,736,811,435"     "692,933,741"
## [1993] "693,970,588"         "694,754,988"        "695,370,296"
## [1996] "695,428,852"         "696,281,472"        "697,518,248"
## [1999] "698,700,667"         "699,579,638,638"    "7,018,100,153"
## [2002] "7,028,803,366"       "7,084,399,840"      "7,168,999,428"
## [2005] "7,174,985,107"       "7,231,963,516"      "7,276,013,032"
## [2008] "7,308,755,000,000"   "7,314,865,176"      "7,322,069,511"
## [2011] "7,330,965,239"       "7,335,027,592"      "7,367,494,080"
## [2014] "7,375,918,367"       "7,403,457,319"      "7,468,096,567"
## [2017] "7,533,187,605"       "7,548,912,105"      "7,650,125,217"
## [2020] "7,662,595,076"       "7,664,060,000,000"  "7,683,870,000"
## [2023] "7,745,406,201"       "7,753,501,868"      "7,841,602,824"
## [2026] "7,870,782,261"       "7,870,982,171"      "7,880,509,171"
## [2029] "7,937,758,980"       "70,140,835,299"     "70,481,451,449"
## [2032] "70,596,729,394"      "70,979,923,960"     "705,145,868,624"
## [2035] "705,406,001"         "705,704,816"        "707,906,744,575"
## [2038] "709,182,559,935"     "709,452,584"        "71,623,500,000"
## [2041] "717,530,683"         "72,207,028,767"     "721,207,148"
## [2044] "722,460,886"         "725,185,185"        "726,649,102,998"
## [2047] "727,860,593"         "729,336,319,677"    "73,141,000,000"
## [2050] "73,447,063,319"      "73,777,792,327"     "73,942,235,330"
## [2053] "732,732,350"         "736,379,777,893"    "737,554,872"
## [2056] "74,703,517,903"      "74,827,400,000"     "741,929,343"
## [2059] "742,293,448,253"     "743,808,076"        "745,162,608,269"
## [2062] "745,521,862,833"     "748,346,606"        "749,138,010"
## [2065] "75,173,794,497"      "75,527,984,234"     "75,575,902,932"
## [2068] "75,643,459,840"      "75,880,632,560"     "75,954,641,355"
## [2071] "756,666,667"         "756,706,300,590"    "757,950,678,647"
## [2074] "757,999,453,314"     "758,683,593"        "76,216,441,462"
## [2077] "76,261,278,405"      "76,262,072,468"     "76,636,898,036"
## [2080] "763,465,550"         "764,017,107,992"    "764,880,644,711"
## [2083] "766,198,926"         "769,305,386,183"    "77,148,000,000"
## [2086] "77,464,561,150"      "77,860,932,152"     "771,015,889"
## [2089] "771,278,111"         "771,470,783,218"    "771,901,768,698"
## [2092] "772,106,378,935"     "775,934,813"        "777,227,541,581"
## [2095] "778,648,667"         "78,039,572,222"     "78,693,253,276"
## [2098] "78,721,607,509"      "78,813,839,984"     "782,240,601,985"
```

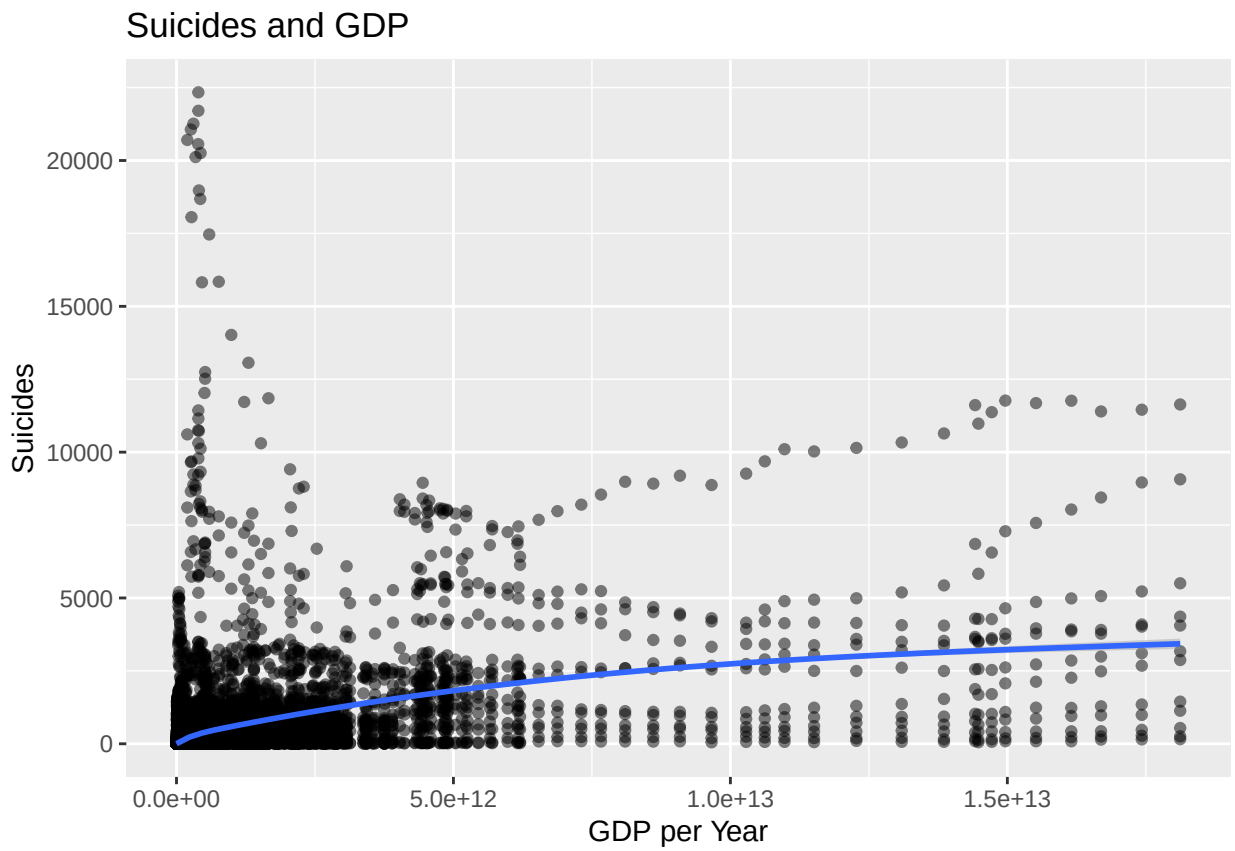
```
## [2101] "784,159,229"       "785,918,770"       "79,169,043,642"
## [2104] "79,276,664,000"    "79,712,087,912"    "799,882,148"
## [2107] "8,033,877,360"     "8,076,470,000"     "8,100,201,000,000"
## [2110] "8,104,355,717"     "8,136,345,144"     "8,137,911,978"
## [2113] "8,150,138,757"     "8,154,338,233"     "8,158,548,717"
## [2116] "8,195,993,231"     "8,213,515,459"     "8,317,830,000"
## [2119] "8,350,252,966"     "8,385,109,020"     "8,392,549,702"
## [2122] "8,410,724,361"     "8,438,951,476"     "8,453,704,212"
## [2125] "8,528,202,278"     "8,528,593,084"     "8,608,515,000,000"
## [2128] "8,614,215,559"     "8,647,936,748"     "8,680,472,169"
## [2131] "8,734,653,809"     "8,741,059,603"     "8,758,622,329"
## [2134] "8,792,365,811"     "8,824,873,259"     "8,858,006,036"
## [2137] "8,870,090,000"     "8,881,160,000"     "8,917,286,036"
## [2140] "8,921,947,100"     "8,927,140,438"     "8,976,207,713"
## [2143] "8,977,149,553"     "8,985,352,832"     "8,992,642,349"
## [2146] "80,322,313,000"    "80,656,100,000"    "80,797,945,205"
## [2149] "800,740,259"       "803,055,418,883"   "81,003,884,545"
## [2152] "81,026,300,311"    "81,076,723,017"    "81,357,605,642"
## [2155] "81,456,918,679"    "81,577,430,181"    "81,703,500,846"
## [2158] "81,910,771,994"    "814,615,333"       "82,344,260,571"
## [2161] "82,709,161,099"    "82,826,146,132"    "82,995,145,792"
## [2164] "823,837,141"       "824,880,550"       "825,977,889"
## [2167] "828,946,812,397"   "83,908,206,648"    "83,914,521,300"
## [2170] "830,158,778"       "832,072,450"       "832,523,681,194"
## [2173] "836,389,937,229"   "839,319,927"       "839,419,655,078"
## [2176] "84,952,360,922"    "842,620,111"       "847,397,850"
## [2179] "85,324,771,841"    "85,343,063,966"    "85,500,935,935"
## [2182] "85,707,636,233"    "850,426,432,992"   "851,962,785,585"
## [2185] "855,643,111"       "857,932,759,100"   "859,796,872,794"
## [2188] "86,142,018,069"    "86,186,158,685"    "86,283,126,844"
## [2191] "86,304,245,825"    "86,307,135,997"    "86,730,038,793"
## [2194] "860,630,923"       "863,723,411,633"   "866,680,000,367"
## [2197] "87,132,800,000"    "87,276,164,400"    "87,890,009,877"
## [2200] "87,924,544,000"    "870,179,739"       "871,860,600"
## [2203] "873,250,000"       "873,982,246,102"   "877,476,221,382"
## [2206] "879,635,084,125"   "88,250,885,550"    "88,416,668,900"
## [2209] "88,945,625,174"    "883,199,625,325"   "885,444,186"
## [2212] "888,667,913,419"   "89,242,382,961"    "89,255,751,015"
## [2215] "89,286,208,629"    "89,501,012,916"    "89,524,131,600"
## [2218] "891,000,000"       "891,630,175,813"   "892,164,394"
## [2221] "892,380,986,368"   "893,107,211"       "893,757,287,202"
## [2224] "897,031,250"       "898,137,194,716"   "9,008,273,721"
## [2227] "9,024,567,484"     "9,055,290,000"     "9,062,131,308"
## [2230] "9,062,906,915"     "9,089,168,000,000" "9,128,843,109"
## [2233] "9,178,016,493"     "9,206,301,700"     "9,207,689,916"
## [2236] "9,209,559,296"     "9,260,284,938"     "9,298,839,655"
## [2239] "9,399,447,609"     "9,406,097,735"     "9,507,645,260"
## [2242] "9,537,297,507"     "9,546,441,564"     "9,586,327,800"
## [2245] "9,632,155,053"     "9,660,624,000,000" "9,687,951,055"
## [2248] "9,694,169,757"     "9,697,416,974"     "9,750,161,053"
## [2251] "9,774,316,692"     "9,788,391,733"     "9,833,870,709"
## [2254] "9,836,200,000"     "9,870,494,000"     "9,965,225,497"
## [2257] "9,981,960,000"     "9,990,370,016"     "90,082,034,316"
## [2260] "900,045,362,045"   "901,934,953,365"   "906,853,273,138"
```

```
## [2263] "91,030,959,455"     "91,642,093,873"     "91,941,192,896"
## [2266] "910,122,732,124"    "911,481,481"        "919,103,255"
## [2269] "919,577,148"        "92,507,279,383"     "925,598,068,021"
## [2272] "926,448,240,318"    "926,884,816,754"    "929,607,500"
## [2275] "93,413,992,956"     "93,639,316,000"     "932,551,850"
## [2278] "934,173,305,686"    "934,185,915,467"    "936,228,211,513"
## [2281] "94,230,055,659"     "94,337,050,693"     "94,684,584,163"
## [2284] "945,000,000"        "95,019,103,603"     "95,129,659,000"
## [2287] "95,833,932,715"     "950,579,413,279"    "951,207,366"
## [2290] "96,045,645,026"     "96,385,638,000"     "96,403,758,865"
## [2293] "967,199,594"        "969,936,525"        "97,001,377,569"
## [2296] "97,160,109,278"     "97,724,004,252"     "97,798,351,648"
## [2299] "97,891,090,929"     "97,933,391,976"     "975,387,131,716"
## [2302] "979,850,000"        "98,203,546,156"     "98,234,695,722"
## [2305] "98,381,268,000"     "98,443,739,941"     "98,478,349,315"
## [2308] "98,585,185"         "98,691,849,563"     "984,297,589"
## [2311] "989,930,542,279"    "99,036,165,210"     "99,210,392,858"
## [2314] "99,290,381,000"     "99,627,140,274"     "99,697,566,668"
## [2317] "99,698,453,261"     "99,853,528,653"     "99,886,577,331"
## [2320] "990,374,050"        "997,007,926"
```

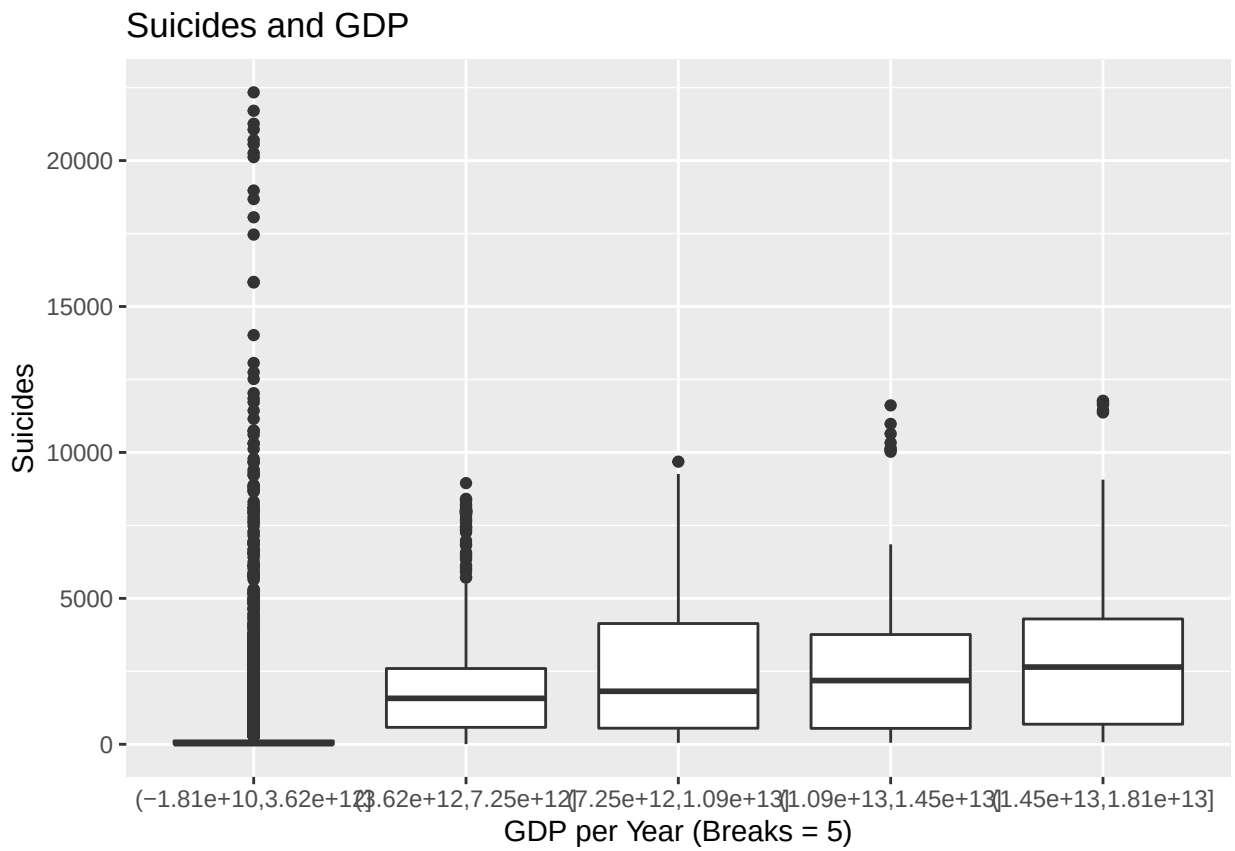
```
suicides$gdp_per_year <- gsub(",", "", suicides$gdp_per_year)
suicides$gdp_per_year <- as.numeric(suicides$gdp_per_year)
```

```
suicides %>%
  ggplot(aes(gdp_per_year, suicides_no)) +
  geom_point(alpha = 0.5, position = "jitter") +
  geom_smooth() +
  labs(title = "Suicides and GDP",
       x = "GDP per Year",
       y = "Suicides")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
suicides %>%
  ggplot(aes(cut(gdp_per_year, breaks = 5), suicides_no)) +
  geom_boxplot() +
  labs(title = "Suicides and GDP",
       x = "GDP per Year (Breaks = 5)",
       y = "Suicides")
```



There is no clear trend between suicides and GDP per year. The scatterplot suggests a slightly positive association. However, the boxplot hints that this might be due mainly to outliers.

MODELS AND RESULTS

In this section, different models are trained with the end goal of determining whether suicide rates are associated with demographic variables.

First, the variables of interest are selected from the main dataset.

```
suicides <- suicides %>%
  select(suicides_no, sex, age, population, generation, continent)

colnames(suicides)[1] <- "suicides"
names(suicides)
```

```
## [1] "suicides"   "sex"        "age"        "population" "generation"
## [6] "continent"
```

Next, the training and validation sets are created.

```
set.seed(725, sample.kind = "Rounding")
```

```
## Warning in set.seed(725, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```
y <- suicides$suicides
test_index <- createDataPartition(y, times = 1, p = 0.5, list = FALSE)
```

```
test_set <- suicides %>% slice(test_index)
train_set <- suicides %>% slice(-test_index)
```

The RMSE function, which will be used to assess the performance of the models, is defined next. The RMSE refers to the residual mean squared error of the predicted suicide rates against the actual rates. The aim is to minimize the RMSE.

```
RMSE <- function(actual_suicides, predicted_suicides){
  sqrt(mean((actual_suicides - predicted_suicides)^2))
}

actual_suicides <- test_set$suicides
```

Using the Mean

The simplest model is to predict the same number of suicides for all continents, with random variation explained by an error term:

$$\hat{Y} = \mu + \varepsilon$$

The estimate that would minimize the RMSE is the average of all suicides.

```
mu <- mean(train_set$suicides)
mu
```

```
## [1] 237.4253
```

Using mu to predict the suicide rates, the RMSE is 913.738.

```
RMSE_mean <- RMSE(actual_suicides, mu)

options(pillar.sigfig = 6)
rmse_results <- tibble(Method = "Mean",
                       RMSE = RMSE_mean)
rmse_results
```

```
## # A tibble: 1 x 2
##   Method    RMSE
##   <chr>    <dbl>
## 1 Mean   913.738
```

The RMSE is quite high. It is higher than the standard deviation of the suicide rates. This ineffectiveness may be due to the presence of significant outliers in the data.

Linear Regression

The next option is to develop a linear model which predicts the number of suicides ($\hat{Y}$) based on the demographic variables (X), while at the same time minimizing the error.

As there are 5 explanatory variables, the linear model is selected using the nested models approach. Under this approach, models which add a new variable are compared to each other using the analysis of variance (ANOVA) method. The ANOVA results can show whether the added term is necessary for the model. The variance inflation factor (VIF) is also examined to determine whether the additional term significantly increases the variance of the model.

This nested model approach showed that it is not necessary to add generation as an explanatory variable.

```
fit1 <- lm(suicides ~ sex, train_set)
fit2 <- update(fit1, suicides ~ sex + age, train_set)
vif(fit2) ## to check the variance inflation ##
```

```
##         GVIF Df GVIF^(1/(2*Df))
## sex 1.000178  1        1.000089
## age 1.000178  5        1.000018
```

```
anova(fit1, fit2) ## model 2's terms are necessary over model 1
```

```
## Analysis of Variance Table
##
## Model 1: suicides ~ sex
## Model 2: suicides ~ sex + age
##   Res.Df        RSS Df Sum of Sq      F    Pr(>F)
## 1  13907 1.0801e+10
## 2  13902 1.0442e+10  5 358470902 95.449 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit2) ## Adj. R-squared = 0.05223
```

```
##
## Call:
## lm(formula = suicides ~ sex + age, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -626.8  -260.4  -127.5    19.9 21079.2
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)       37.08      19.43   1.909  0.05631 .
## sexmale          255.25      14.70  17.365  < 2e-16 ***
## age25-34 years    68.52      25.45   2.692  0.00711 **
## age35-54 years   334.44      25.31  13.212  < 2e-16 ***
## age5-14 years   -154.84      25.48  -6.076 1.27e-09 ***
## age55-74 years   206.11      25.40   8.113 5.34e-16 ***
## age75+ years     -31.89      25.42  -1.255  0.20959
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 866.7 on 13902 degrees of freedom
## Multiple R-squared:  0.05264,    Adjusted R-squared:  0.05223
## F-statistic: 128.7 on 6 and 13902 DF,  p-value: < 2.2e-16
```

```
fit3 <- update(fit1, suicides ~
                 sex + age + generation, train_set)
vif(fit3) ## variance increased, for age and generation
```

```
##                 GVIF Df GVIF^(1/(2*Df))
## sex         1.000294  1        1.000147
## age        22.215400  5        1.363532
## generation 22.213557  5        1.363521
```

```r
anova(fit2, fit3) ## model 2 over model 3, generation not necessary ##
```

```
## Analysis of Variance Table
##
## Model 1: suicides ~ sex + age
## Model 2: suicides ~ sex + age + generation
##   Res.Df        RSS Df Sum of Sq     F Pr(>F)
## 1  13902 1.0442e+10
## 2  13897 1.0436e+10  5   5737584 1.528 0.1773
```

```r
summary(fit3) ## Adj. R-squared = 0.05241 ##
```

```
##
## Call:
## lm(formula = suicides ~ sex + age + generation, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -640.2  -260.8  -124.9    19.4 21067.8
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  50.765     37.689   1.347   0.1780
## sexmale                     254.833     14.699  17.337  < 2e-16 ***
## age25-34 years               58.460     27.952   2.091   0.0365 *
## age35-54 years              332.567     36.734   9.053  < 2e-16 ***
## age5-14 years              -147.633     28.934  -5.102 3.40e-07 ***
## age55-74 years              265.604     48.703   5.453 5.02e-08 ***
## age75+ years                 42.862     53.430   0.802   0.4225
## generationG.I. Generation   -87.668     44.523  -1.969   0.0490 *
## generationGeneration X        2.021     30.547   0.066   0.9472
## generationGeneration Z      -21.574     52.338  -0.412   0.6802
## generationMillenials        -26.075     37.726  -0.691   0.4895
## generationSilent            -88.680     35.320  -2.511   0.0121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 866.6 on 13897 degrees of freedom
## Multiple R-squared:  0.05316,    Adjusted R-squared:  0.05241
## F-statistic: 70.93 on 11 and 13897 DF,  p-value: < 2.2e-16
```

```r
fit4 <- update(fit1, suicides ~
                  sex + age +
                  population, train_set)
vif(fit4) ## less variance
```

```
##                GVIF Df GVIF^(1/(2*Df))
## sex        1.000391  1        1.000196
## age        1.034409  5        1.003389
## population 1.034483  1        1.017096
```

```r
anova(fit2, fit4) ## model 4 over model 2 ##
```

```
## Analysis of Variance Table
##
## Model 1: suicides ~ sex + age
```

```
## Model 2: suicides ~ sex + age + population
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1  13902 1.0442e+10
## 2  13901 6.3686e+09  1 4073551731 8891.6 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit4) ## Adj. R-squared = 0.4219 ##
```

```
##
## Call:
## lm(formula = suicides ~ sex + age + population, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3112.8  -172.2    -8.4   103.5 18713.8
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.345e+02  1.544e+01 -15.187  < 2e-16 ***
## sexmale         2.711e+02  1.148e+01  23.609  < 2e-16 ***
## age25-34 years  7.904e+01  1.988e+01   3.976 7.04e-05 ***
## age35-54 years  1.688e+02  1.985e+01   8.503  < 2e-16 ***
## age5-14 years  -1.516e+02  1.990e+01  -7.615 2.80e-14 ***
## age55-74 years  1.988e+02  1.984e+01  10.018  < 2e-16 ***
## age75+ years    1.477e+02  1.994e+01   7.407 1.36e-13 ***
## population      1.448e-04  1.535e-06  94.295  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 676.9 on 13901 degrees of freedom
## Multiple R-squared:  0.4222, Adjusted R-squared:  0.4219
## F-statistic:  1451 on 7 and 13901 DF,  p-value: < 2.2e-16
```

```
fit5 <- update(fit1, suicides ~
                sex + age +
                population + continent, train_set)
vif(fit5)
```

```
##                GVIF Df GVIF^(1/(2*Df))
## sex        1.000501  1        1.000250
## age        1.036088  5        1.003552
## population 1.043113  1        1.021329
## continent  1.009527  4        1.001186
```

```
anova(fit4, fit5) ## model 5 over model 4 ##
```

```
## Analysis of Variance Table
##
## Model 1: suicides ~ sex + age + population
## Model 2: suicides ~ sex + age + population + continent
##   Res.Df        RSS Df Sum of Sq      F    Pr(>F)
## 1  13901 6368559423
## 2  13897 6279289798  4  89269624 49.392 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit5) ## Adj. R-squared = 0.4299 ##
```

```
##
## Call:
## lm(formula = suicides ~ sex + age + population + continent, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3040.9  -163.7   -19.6   130.2 18618.6
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -3.530e+02  3.479e+01 -10.147  < 2e-16 ***
## sexmale            2.707e+02  1.140e+01  23.736  < 2e-16 ***
## age25-34 years     8.243e+01  1.974e+01   4.175 3.00e-05 ***
## age35-54 years     1.690e+02  1.971e+01   8.573  < 2e-16 ***
## age5-14 years     -1.504e+02  1.977e+01  -7.610 2.91e-14 ***
## age55-74 years     1.973e+02  1.970e+01  10.013  < 2e-16 ***
## age75+ years       1.507e+02  1.981e+01   7.608 2.97e-14 ***
## population         1.458e-04  1.531e-06  95.193  < 2e-16 ***
## continentAmericas  1.588e+01  3.331e+01   0.477   0.6336
## continentAsia      1.372e+02  3.440e+01   3.988 6.69e-05 ***
## continentEurope    1.951e+02  3.300e+01   5.911 3.47e-09 ***
## continentOceania   1.310e+02  4.382e+01   2.989   0.0028 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 672.2 on 13897 degrees of freedom
## Multiple R-squared:  0.4303, Adjusted R-squared:  0.4299
## F-statistic: 954.3 on 11 and 13897 DF,  p-value: < 2.2e-16
```

Using this linear model generated the RMSE of 691.42, which is better than the RMSE from simply using the mean.

```
lm_fit1 <- lm(suicides ~ sex + age +
               population + continent, train_set)

y_hat1 <- predict(lm_fit1, test_set)

RMSE_lm <- RMSE(actual_suicides, y_hat1)

rmse_results <- bind_rows(rmse_results,
                          tibble(Method = "Linear Model",
                       RMSE = RMSE_lm))
rmse_results
```

```
## # A tibble: 2 x 2
##   Method          RMSE
##   <chr>          <dbl>
## 1 Mean         913.738
## 2 Linear Model 691.420
```

The figures above suggest that certain variables (i.e., population) might have a different effect on the number of suicides based on the values of another variable (i.e., continent, sex). As such, the possibility of interactions should also be accounted for.

The nested models approach is likewise used in selecting the linear regression model with interaction terms.

```
fit6 <- update(fit1, suicides ~ sex + age +
                 population + continent + population:sex,
               train_set)
vif(fit6)
```

```
##                     GVIF Df GVIF^(1/(2*Df))
## sex             1.225158  1        1.106868
## age             1.036587  5        1.003600
## population      2.011642  1        1.418324
## continent       1.009613  4        1.001197
## sex:population 2.186076  1        1.478538
```

```
anova(fit5, fit6) ## with interactions
```

```
## Analysis of Variance Table
##
## Model 1: suicides ~ sex + age + population + continent
## Model 2: suicides ~ sex + age + population + continent + sex:population
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1  13897 6279289798
## 2  13896 4859616789  1 1419673009 4059.5 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit6) ## adj. r2 = 0.5587
```

```
##
## Call:
## lm(formula = suicides ~ sex + age + population + continent +
##     sex:population, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4322.7  -120.8    -8.8   104.9 17123.5
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -2.016e+02  3.070e+01  -6.569 5.26e-11 ***
## sexmale            -3.222e+01  1.110e+01  -2.902 0.003711 **
## age25-34 years      8.402e+01  1.737e+01   4.837 1.33e-06 ***
## age35-54 years      1.685e+02  1.734e+01   9.718  < 2e-16 ***
## age5-14 years      -1.523e+02  1.739e+01  -8.757  < 2e-16 ***
## age55-74 years      2.015e+02  1.733e+01  11.623  < 2e-16 ***
## age75+ years        1.697e+02  1.743e+01   9.739  < 2e-16 ***
## population          6.305e-05  1.871e-06  33.708  < 2e-16 ***
## continentAmericas   1.187e+01  2.930e+01   0.405 0.685531
## continentAsia       1.338e+02  3.026e+01   4.422 9.86e-06 ***
## continentEurope     1.974e+02  2.903e+01   6.799 1.10e-11 ***
## continentOceania    1.288e+02  3.855e+01   3.341 0.000836 ***
## sexmale:population  1.681e-04  2.639e-06  63.714  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 591.4 on 13896 degrees of freedom
```

```
## Multiple R-squared:  0.5591, Adjusted R-squared:  0.5587
## F-statistic:  1469 on 12 and 13896 DF,  p-value: < 2.2e-16
```

```
fit7 <- update(fit1, suicides ~ sex + age +
                 population + continent + population:sex +
               population:age, train_set)
vif(fit7) ## increased variance with interaction
```

```
##                     GVIF Df GVIF^(1/(2*Df))
## sex             1.226332  1        1.107399
## age             3.116415  5        1.120380
## population      8.094232  1        2.845036
## continent       1.017783  4        1.002206
## sex:population  2.195574  1        1.481747
## age:population 17.673750  5        1.332701
```

```
anova(fit6, fit7)
```

```
## Analysis of Variance Table
##
## Model 1: suicides ~ sex + age + population + continent + sex:population
## Model 2: suicides ~ sex + age + population + continent + sex:population +
##     age:population
##   Res.Df        RSS Df Sum of Sq      F    Pr(>F)
## 1  13896 4859616789
## 2  13891 3943719561  5 915897229 645.21 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit7) ## adj. r2 = 0.6418
```

```
##
## Call:
## lm(formula = suicides ~ sex + age + population + continent +
##     sex:population + age:population, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3926.1   -68.2   -14.4    80.2 16495.7
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -9.055e+01  2.820e+01  -3.211 0.001326 **
## sexmale                    -3.475e+01  1.001e+01  -3.473 0.000517 ***
## age25-34 years             -8.697e+00  1.770e+01  -0.491 0.623176
## age35-54 years             -3.366e+01  1.754e+01  -1.918 0.055087 .
## age5-14 years               1.188e+01  1.765e+01   0.673 0.500877
## age55-74 years             -1.973e+01  1.755e+01  -1.125 0.260754
## age75+ years                3.386e+01  1.733e+01   1.954 0.050716 .
## population                  9.578e-06  3.381e-06   2.833 0.004617 **
## continentAmericas           1.845e+01  2.641e+01   0.699 0.484636
## continentAsia               1.200e+02  2.727e+01   4.402 1.08e-05 ***
## continentEurope             1.594e+02  2.618e+01   6.088 1.17e-09 ***
## continentOceania            1.003e+02  3.474e+01   2.886 0.003907 **
## sexmale:population          1.733e-04  2.383e-06  72.742  < 2e-16 ***
## age25-34 years:population   5.057e-05  4.638e-06  10.904  < 2e-16 ***
```

```
## age35-54 years:population  8.773e-05  3.698e-06  23.722  < 2e-16 ***
## age5-14 years:population  -9.203e-05  4.493e-06 -20.483  < 2e-16 ***
## age55-74 years:population  1.198e-04  4.337e-06  27.634  < 2e-16 ***
## age75+ years:population    1.257e-04  8.825e-06  14.240  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 532.8 on 13891 degrees of freedom
## Multiple R-squared:  0.6422, Adjusted R-squared:  0.6418
## F-statistic:  1467 on 17 and 13891 DF,  p-value: < 2.2e-16
```

```
fit8 <- update(fit1, suicides ~ sex + age +
                population + continent + population:sex +
                population:age + population:generation, train_set)
vif(fit8)
```

```
##                            GVIF Df GVIF^(1/(2*Df))
## sex                    1.226515  1        1.107482
## age                    3.132044  5        1.120941
## population            18.969014  1        4.355343
## continent              1.018870  4        1.002340
## sex:population         2.198017  1        1.482571
## age:population       108.286476  5        1.597561
## population:generation  21.497319  5        1.359059
```

```
anova(fit7, fit8)
```

```
## Analysis of Variance Table
##
## Model 1: suicides ~ sex + age + population + continent + sex:population +
##     age:population
## Model 2: suicides ~ sex + age + population + continent + sex:population +
##     age:population + population:generation
##   Res.Df        RSS Df Sum of Sq      F    Pr(>F)
## 1  13891 3943719561
## 2  13886 3932559048  5  11160512 7.8816 2.013e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit8) ## adj. r2 = 0.6427
```

```
##
## Call:
## lm(formula = suicides ~ sex + age + population + continent +
##     sex:population + age:population + population:generation,
##     data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3706.3   -67.5   -13.7    80.6 16432.9
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       -8.843e+01  2.817e+01  -3.139 0.001699 **
## sexmale                           -3.403e+01  9.995e+00  -3.405 0.000665 ***
## age25-34 years                    -8.536e+00  1.768e+01  -0.483 0.629226
```

```
## age35-54 years                        -3.509e+01  1.753e+01  -2.002 0.045266 *
## age5-14 years                          1.172e+01  1.763e+01   0.665 0.506132
## age55-74 years                        -2.211e+01  1.755e+01  -1.260 0.207630
## age75+ years                           3.252e+01  1.731e+01   1.879 0.060281 .
## population                             2.651e-05  5.169e-06   5.130 2.94e-07 ***
## continentAmericas                      1.573e+01  2.638e+01   0.596 0.551127
## continentAsia                          1.182e+02  2.724e+01   4.337 1.45e-05 ***
## continentEurope                        1.566e+02  2.615e+01   5.989 2.16e-09 ***
## continentOceania                       9.803e+01  3.470e+01   2.825 0.004735 **
## sexmale:population                     1.730e-04  2.381e-06  72.653  < 2e-16 ***
## age25-34 years:population              4.546e-05  4.915e-06   9.250  < 2e-16 ***
## age35-54 years:population              7.445e-05  4.842e-06  15.376  < 2e-16 ***
## age5-14 years:population              -9.127e-05  4.990e-06 -18.290  < 2e-16 ***
## age55-74 years:population              9.690e-05  6.248e-06  15.508  < 2e-16 ***
## age75+ years:population                9.487e-05  1.045e-05   9.075  < 2e-16 ***
## population:generationG.I. Generation   3.421e-05  8.233e-06   4.155 3.27e-05 ***
## population:generationGeneration X     -1.474e-05  3.405e-06  -4.328 1.51e-05 ***
## population:generationGeneration Z     -1.629e-05  7.789e-06  -2.091 0.036530 *
## population:generationMillenials       -1.873e-05  5.154e-06  -3.633 0.000281 ***
## population:generationSilent            4.980e-06  4.090e-06   1.218 0.223365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 532.2 on 13886 degrees of freedom
## Multiple R-squared:  0.6432, Adjusted R-squared:  0.6427
## F-statistic:  1138 on 22 and 13886 DF,  p-value: < 2.2e-16
```

```
fit9 <- update(fit1, suicides ~ sex + age +
                population + continent + population:sex +
                population:age + population:generation +
                population:continent, train_set)
vif(fit9) ## increased variance
```

```
##                              GVIF Df GVIF^(1/(2*Df))
## sex                      1.226573  1        1.107508
## age                      3.180552  5        1.122665
## population             152.601126  1       12.353183
## continent                3.904729  4        1.185629
## sex:population           2.198875  1        1.482861
## age:population         109.366630  5        1.599147
## population:generation   21.754237  5        1.360675
## population:continent   419.839471  4        2.127578
```

```
anova(fit8, fit9)
```

```
## Analysis of Variance Table
##
## Model 1: suicides ~ sex + age + population + continent + sex:population +
##     age:population + population:generation
## Model 2: suicides ~ sex + age + population + continent + sex:population +
##     age:population + population:generation + population:continent
##   Res.Df        RSS Df Sum of Sq      F    Pr(>F)
## 1  13886 3932559048
## 2  13882 3410423709  4 522135339 531.33 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit9) ## adj. r2 = 0.69
```

```
##
## Call:
## lm(formula = suicides ~ sex + age + population + continent +
##     sex:population + age:population + population:generation +
##     population:continent, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2944.8   -29.6    24.9    70.7 14816.2
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          1.148e+01  2.927e+01   0.392 0.695002
## sexmale                             -3.233e+01  9.310e+00  -3.473 0.000517 ***
## age25-34 years                      -7.692e+00  1.647e+01  -0.467 0.640435
## age35-54 years                      -6.333e+01  1.635e+01  -3.873 0.000108 ***
## age5-14 years                        1.291e+01  1.642e+01   0.786 0.431918
## age55-74 years                      -1.357e+01  1.636e+01  -0.830 0.406749
## age75+ years                         6.216e+01  1.615e+01   3.849 0.000119 ***
## population                          -6.396e-05  1.366e-05  -4.684 2.84e-06 ***
## continentAmericas                   -1.928e+01  2.791e+01  -0.691 0.489797
## continentAsia                       -4.537e+01  2.904e+01  -1.562 0.118286
## continentEurope                     -9.728e+01  2.788e+01  -3.489 0.000486 ***
## continentOceania                     9.299e-01  4.006e+01   0.023 0.981481
## sexmale:population                   1.749e-04  2.218e-06  78.854  < 2e-16 ***
## age25-34 years:population            3.811e-05  4.582e-06   8.317  < 2e-16 ***
## age35-54 years:population            5.986e-05  4.522e-06  13.237  < 2e-16 ***
## age5-14 years:population            -9.138e-05  4.649e-06 -19.658  < 2e-16 ***
## age55-74 years:population            7.516e-05  5.840e-06  12.871  < 2e-16 ***
## age75+ years:population              8.497e-05  9.741e-06   8.723  < 2e-16 ***
## population:generationG.I. Generation 2.323e-05  7.674e-06   3.027 0.002474 **
## population:generationGeneration X   -1.883e-05  3.174e-06  -5.933 3.05e-09 ***
## population:generationGeneration Z   -1.452e-05  7.260e-06  -2.001 0.045464 *
## population:generationMillenials     -2.186e-05  4.807e-06  -4.548 5.45e-06 ***
## population:generationSilent         -6.280e-06  3.824e-06  -1.642 0.100592
## population:continentAmericas         6.988e-05  1.282e-05   5.452 5.07e-08 ***
## population:continentAsia             1.360e-04  1.312e-05  10.368  < 2e-16 ***
## population:continentEurope           1.965e-04  1.299e-05  15.126  < 2e-16 ***
## population:continentOceania          9.722e-05  3.134e-05   3.102 0.001928 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 495.7 on 13882 degrees of freedom
## Multiple R-squared:  0.6906, Adjusted R-squared:   0.69
## F-statistic:  1192 on 26 and 13882 DF,  p-value: < 2.2e-16
```

The linear model which accounts for interactions generated an RMSE of 512.80, which is better than the linear model without interaction terms.

```r
lm_fit2 <- lm(suicides ~ sex + age +
                population + continent +
                population:sex + population:age +
                population:generation +
```

```
                population:continent,
              train_set)

y_hat2 <- predict(lm_fit2, test_set)

RMSE_lm_int <- RMSE(actual_suicides, y_hat2)

rmse_results <- bind_rows(rmse_results,
                          tibble(Method = "Linear Model, Interactions",
                                 RMSE = RMSE_lm_int))
rmse_results
```

```
## # A tibble: 3 x 2
##   Method                         RMSE
##   <chr>                         <dbl>
## 1 Mean                         913.738
## 2 Linear Model                 691.420
## 3 Linear Model, Interactions 512.796
```

Poisson Regression

Since the Y in this case is count data with many zeros, a Poisson regression is a viable option. However, we earlier saw that the variance of the number of suicides is greater than its mean. As such, we should use Quasipoisson regression.

The quasipoisson model is developed in an analogous manner to how the linear model was developed. In this case, however, the pseudo R-squared values are compared in order to find which quasipoisson model better predicts the outcome.

```
glm_fit <- glm(suicides ~ sex, train_set,
               family = "quasipoisson")
summary(glm_fit)
```

```
##
## Call:
## glm(formula = suicides ~ sex, family = "quasipoisson", data = train_set)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -26.92  -21.05  -13.86   -4.98  367.40
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.69949    0.05721   82.14   <2e-16 ***
## sexmale      1.19326    0.06516   18.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 2475.964)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance: 10874403  on 13907  degrees of freedom
## AIC: NA
```

```
##
## Number of Fisher Scoring iterations: 7
```

```
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##    pseudoR2
##       <dbl>
## 1 0.0832665
```

```
  ## pseudoR2 = 0.0832665
```

```
glm_fit <- glm(suicides ~ sex + age, train_set,
               family = "quasipoisson")
summary(glm_fit)
```

```
##
## Call:
## glm(formula = suicides ~ sex + age, family = "quasipoisson",
##     data = train_set)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -39.19  -18.19   -9.85   -1.59  321.25
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.33492    0.07628  56.827  < 2e-16 ***
## sexmale         1.20482    0.05271  22.856  < 2e-16 ***
## age25-34 years  0.34338    0.08450   4.064 4.85e-05 ***
## age35-54 years  1.10377    0.07433  14.849  < 2e-16 ***
## age5-14 years  -2.69280    0.25879 -10.405  < 2e-16 ***
## age55-74 years  0.80964    0.07771  10.419  < 2e-16 ***
## age75+ years   -0.20852    0.09635  -2.164   0.0305 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1620.425)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance:  9145589  on 13902  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##   pseudoR2
##      <dbl>
## 1 0.229009
```

```
  ## pseudoR2 = 0.229009
```

```r
glm_fit <- glm(suicides ~ sex + age +
                 generation, train_set,
               family = "quasipoisson")
summary(glm_fit)
```

```
##
## Call:
## glm(formula = suicides ~ sex + age + generation, family = "quasipoisson",
##     data = train_set)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -39.82  -18.15   -9.92   -1.55  319.81
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 4.42896    0.11401  38.847  < 2e-16 ***
## sexmale                     1.20340    0.05261  22.873  < 2e-16 ***
## age25-34 years              0.27454    0.09240   2.971  0.00297 **
## age35-54 years              1.03311    0.10347   9.985  < 2e-16 ***
## age5-14 years              -2.62333    0.30668  -8.554  < 2e-16 ***
## age55-74 years              0.87345    0.12740   6.856 7.37e-12 ***
## age75+ years               -0.11090    0.15142  -0.732  0.46396
## generationG.I. Generation -0.18224    0.11861  -1.536  0.12446
## generationGeneration X      0.01012    0.07368   0.137  0.89079
## generationGeneration Z     -0.21100    0.55529  -0.380  0.70396
## generationMillenials       -0.18528    0.11709  -1.582  0.11359
## generationSilent           -0.19767    0.07953  -2.485  0.01295 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1614.082)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance:  9128759  on 13897  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```r
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##   pseudoR2
##      <dbl>
## 1 0.230428
```

```r
  ## pseudoR2 = 0.230428
```

```r
glm_fit <- glm(suicides ~ sex + age +
                 generation + population,
               train_set, family = "quasipoisson")
summary(glm_fit)
```

```
##
```

```
## Call:
## glm(formula = suicides ~ sex + age + generation + population,
##     family = "quasipoisson", data = train_set)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -191.185   -13.056    -6.882    -0.498   230.225
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 4.073e+00  6.345e-02  64.187  < 2e-16 ***
## sexmale                     1.249e+00  2.924e-02  42.719  < 2e-16 ***
## age25-34 years              2.993e-01  5.142e-02   5.821 5.99e-09 ***
## age35-54 years              2.792e-01  5.972e-02   4.675 2.97e-06 ***
## age5-14 years              -2.643e+00  1.704e-01 -15.508  < 2e-16 ***
## age55-74 years              4.595e-01  7.098e-02   6.473 9.94e-11 ***
## age75+ years               -2.546e-01  8.389e-02  -3.035 0.002409 **
## generationG.I. Generation   2.272e-01  6.651e-02   3.417 0.000636 ***
## generationGeneration X     -5.633e-02  4.035e-02  -1.396 0.162697
## generationGeneration Z     -2.117e-01  3.084e-01  -0.686 0.492494
## generationMillenials       -2.078e-01  6.465e-02  -3.214 0.001311 **
## generationSilent            1.874e-01  4.504e-02   4.160 3.20e-05 ***
## population                  1.188e-07  1.031e-09 115.145  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 497.7751)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance:  4717179  on 13896  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##   pseudoR2
##      <dbl>
## 1 0.602333
```

```
  ## pseudoR2 = 0.602333
```

```
glm_fit <- glm(suicides ~ sex + age +
                 generation + population +
                 continent, train_set, family = "quasipoisson")
summary(glm_fit)
```

```
##
## Call:
## glm(formula = suicides ~ sex + age + generation + population +
##     continent, family = "quasipoisson", data = train_set)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
```

```
## -199.064   -11.172    -5.062     0.783   150.790
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.698e+00  2.421e-01   7.016 2.40e-12 ***
## sexmale                     1.279e+00  2.434e-02  52.526  < 2e-16 ***
## age25-34 years              2.782e-01  4.259e-02   6.531 6.74e-11 ***
## age35-54 years             -1.448e-02  5.060e-02  -0.286 0.774696
## age5-14 years              -2.606e+00  1.417e-01 -18.387  < 2e-16 ***
## age55-74 years              1.889e-01  6.025e-02   3.134 0.001725 **
## age75+ years               -4.289e-01  7.071e-02  -6.065 1.35e-09 ***
## generationG.I. Generation   3.676e-01  5.659e-02   6.496 8.51e-11 ***
## generationGeneration X     -1.259e-01  3.348e-02  -3.759 0.000171 ***
## generationGeneration Z     -3.212e-01  2.564e-01  -1.253 0.210363
## generationMillenials       -2.919e-01  5.351e-02  -5.455 4.97e-08 ***
## generationSilent            2.617e-01  3.901e-02   6.707 2.06e-11 ***
## population                  1.539e-07  1.264e-09 121.760  < 2e-16 ***
## continentAmericas           1.174e+00  2.388e-01   4.916 8.94e-07 ***
## continentAsia               2.644e+00  2.374e-01  11.137  < 2e-16 ***
## continentEurope             2.829e+00  2.367e-01  11.950  < 2e-16 ***
## continentOceania            1.928e+00  2.517e-01   7.660 1.98e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 344.1494)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance:  3565530  on 13892  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##   pseudoR2
##      <dbl>
## 1 0.699419
```

```
  ## pseudoR2 = 0.602333
```

The quasipoisson model with the continent variable does not improve upon the pseudo R-squared of the model with just the sex, age, generation and population variables. As such, the latter model is the better one and it generates an RMSE of 1110.51.

```
glm_fit <- glm(suicides ~ sex + age +
                 generation + population +
                 continent, train_set,
               family = "quasipoisson")
y_hat3 <- predict(glm_fit, test_set, type = "response")

RMSE_glm <- RMSE(actual_suicides, y_hat3)

rmse_results <- bind_rows(rmse_results,
                          tibble(Method = "Quasipoisson Model",
```

```
                                RMSE = RMSE_glm))
rmse_results
```

```
## # A tibble: 4 x 2
##   Method                          RMSE
##   <chr>                          <dbl>
## 1 Mean                           913.738
## 2 Linear Model                   691.420
## 3 Linear Model, Interactions     512.796
## 4 Quasipoisson Model            1032.60
```

As with the Linear Model, the Quasipoisson Model might be significantly improved by accounting for interaction terms. As such, interaction terms are added to the model and their respective pseudo R-squared values are compared to identify the best model.

```
glm_fit <- glm(suicides ~ sex + age +
                 generation + population +
                 continent + population:age,
               train_set, family = "quasipoisson")
summary(glm_fit)
```

```
##
## Call:
## glm(formula = suicides ~ sex + age + generation + population +
##     continent + population:age, family = "quasipoisson", data = train_set)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -103.115    -8.454    -3.194     1.714   136.226
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  1.165e+00  2.045e-01   5.696 1.25e-08 ***
## sexmale                      1.367e+00  2.102e-02  65.007  < 2e-16 ***
## age25-34 years               2.755e-01  4.941e-02   5.576 2.51e-08 ***
## age35-54 years               9.899e-01  4.953e-02  19.988  < 2e-16 ***
## age5-14 years               -2.588e+00  1.600e-01 -16.171  < 2e-16 ***
## age55-74 years               3.948e-01  6.006e-02   6.574 5.08e-11 ***
## age75+ years                -5.656e-01  6.888e-02  -8.212 2.37e-16 ***
## generationG.I. Generation    6.873e-01  4.728e-02  14.536  < 2e-16 ***
## generationGeneration X      -6.789e-02  2.815e-02  -2.412  0.01588 *
## generationGeneration Z      -2.386e-01  2.152e-01  -1.109  0.26756
## generationMillenials        -2.733e-01  4.507e-02  -6.065 1.35e-09 ***
## generationSilent             3.326e-01  3.180e-02  10.460  < 2e-16 ***
## population                   2.619e-07  3.413e-09  76.718  < 2e-16 ***
## continentAmericas            6.334e-01  2.009e-01   3.153  0.00162 **
## continentAsia                2.431e+00  1.991e-01  12.206  < 2e-16 ***
## continentEurope              2.707e+00  1.986e-01  13.635  < 2e-16 ***
## continentOceania             1.968e+00  2.111e-01   9.324  < 2e-16 ***
## age25-34 years:population    6.460e-10  4.065e-09   0.159  0.87374
## age35-54 years:population   -1.234e-07  3.312e-09 -37.265  < 2e-16 ***
## age5-14 years:population    -3.225e-10  1.259e-08  -0.026  0.97956
## age55-74 years:population   -6.131e-08  3.578e-09 -17.134  < 2e-16 ***
## age75+ years:population      2.970e-07  8.625e-09  34.435  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 241.9741)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance:  2474451  on 13887  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##   pseudoR2
##      <dbl>
## 1 0.791399
```

```
  ## pseudoR2 = 0.791399
```

```
glm_fit <- glm(suicides ~ sex + age +
                 generation + population +
                 continent + population:age +
                 population:continent, train_set, family = "quasipoisson")
summary(glm_fit)
```

```
##
## Call:
## glm(formula = suicides ~ sex + age + generation + population +
##     continent + population:age + population:continent, family = "quasipoisson",
##     data = train_set)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -68.657   -8.189   -3.453    1.141   89.002
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   8.222e-01  2.667e-01   3.082 0.002059 **
## sexmale                       1.399e+00  1.766e-02  79.207  < 2e-16 ***
## age25-34 years                2.487e-01  4.228e-02   5.883 4.13e-09 ***
## age35-54 years                6.419e-01  4.344e-02  14.778  < 2e-16 ***
## age5-14 years                -2.605e+00  1.362e-01 -19.124  < 2e-16 ***
## age55-74 years                3.002e-01  5.153e-02   5.827 5.77e-09 ***
## age75+ years                 -4.324e-01  5.835e-02  -7.410 1.34e-13 ***
## generationG.I. Generation     5.895e-01  3.998e-02  14.744  < 2e-16 ***
## generationGeneration X       -1.276e-01  2.353e-02  -5.421 6.01e-08 ***
## generationGeneration Z       -2.336e-01  1.806e-01  -1.293 0.195927
## generationMillenials         -3.020e-01  3.759e-02  -8.033 1.02e-15 ***
## generationSilent              2.645e-01  2.725e-02   9.708  < 2e-16 ***
## population                    4.686e-07  7.234e-08   6.478 9.64e-11 ***
## continentAmericas             2.005e+00  2.646e-01   7.578 3.73e-14 ***
## continentAsia                 2.531e+00  2.647e-01   9.562  < 2e-16 ***
## continentEurope               2.869e+00  2.638e-01  10.877  < 2e-16 ***
## continentOceania              1.552e+00  2.869e-01   5.412 6.35e-08 ***
## age25-34 years:population     1.305e-09  3.526e-09   0.370 0.711397
```

```
## age35-54 years:population    -1.001e-07  2.896e-09 -34.582  < 2e-16 ***
## age5-14 years:population     -2.216e-09  1.086e-08  -0.204 0.838310
## age55-74 years:population    -5.181e-08  3.149e-09 -16.452  < 2e-16 ***
## age75+ years:population       2.545e-07  7.180e-09  35.442  < 2e-16 ***
## population:continentAmericas -2.602e-07  7.229e-08  -3.599 0.000321 ***
## population:continentAsia     -1.677e-07  7.232e-08  -2.319 0.020389 *
## population:continentEurope   -1.684e-07  7.229e-08  -2.329 0.019870 *
## population:continentOceania   4.982e-07  9.294e-08   5.360 8.44e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 170.5721)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance:  1931383  on 13883  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##   pseudoR2
##      <dbl>
## 1 0.837181
```

```
  ## pseudoR2 = 0.837181
```

```
glm_fit <- glm(suicides ~ sex + age +
                 generation + population +
                 continent + population:age +
                 population:continent + population:sex,
               train_set, family = "quasipoisson")
summary(glm_fit)
```

```
##
## Call:
## glm(formula = suicides ~ sex + age + generation + population +
##     continent + population:age + population:continent + population:sex,
##     family = "quasipoisson", data = train_set)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -70.961   -8.440   -3.595    0.926   89.686
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   9.320e-01  2.658e-01   3.506 0.000456 ***
## sexmale                       1.266e+00  2.498e-02  50.662  < 2e-16 ***
## age25-34 years                2.474e-01  4.206e-02   5.883 4.13e-09 ***
## age35-54 years                6.351e-01  4.324e-02  14.690  < 2e-16 ***
## age5-14 years                -2.605e+00  1.355e-01 -19.226  < 2e-16 ***
## age55-74 years                2.922e-01  5.130e-02   5.696 1.25e-08 ***
## age75+ years                 -4.386e-01  5.809e-02  -7.550 4.62e-14 ***
## generationG.I. Generation     5.928e-01  3.980e-02  14.893  < 2e-16 ***
```

```
## generationGeneration X       -1.277e-01  2.342e-02  -5.452 5.07e-08 ***
## generationGeneration Z       -2.325e-01  1.797e-01  -1.294 0.195831
## generationMillenials         -3.020e-01  3.741e-02  -8.073 7.41e-16 ***
## generationSilent              2.693e-01  2.713e-02   9.924  < 2e-16 ***
## population                    4.582e-07  7.200e-08   6.364 2.03e-10 ***
## continentAmericas             2.005e+00  2.633e-01   7.613 2.84e-14 ***
## continentAsia                 2.529e+00  2.634e-01   9.603  < 2e-16 ***
## continentEurope               2.865e+00  2.625e-01  10.916  < 2e-16 ***
## continentOceania              1.551e+00  2.855e-01   5.433 5.63e-08 ***
## age25-34 years:population     1.311e-09  3.507e-09   0.374 0.708564
## age35-54 years:population    -9.960e-08  2.880e-09 -34.576  < 2e-16 ***
## age5-14 years:population     -2.298e-09  1.080e-08  -0.213 0.831476
## age55-74 years:population    -5.106e-08  3.134e-09 -16.296  < 2e-16 ***
## age75+ years:population       2.536e-07  7.163e-09  35.395  < 2e-16 ***
## population:continentAmericas -2.602e-07  7.194e-08  -3.616 0.000300 ***
## population:continentAsia     -1.675e-07  7.197e-08  -2.328 0.019947 *
## population:continentEurope   -1.676e-07  7.194e-08  -2.329 0.019852 *
## population:continentOceania   4.996e-07  9.250e-08   5.401 6.74e-08 ***
## sexmale:population            1.245e-08  1.724e-09   7.225 5.27e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 168.8917)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance:  1922293  on 13882  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##   pseudoR2
##      <dbl>
## 1 0.837947
```

```
  ## pseudoR2 = 0.837947
```

```
glm_fit <- glm(suicides ~ sex + age +
                 generation + population +
                 continent + population:age +
                 population:continent + population:sex +
                 population:generation, train_set, family = "quasipoisson")
summary(glm_fit)
```

```
##
## Call:
## glm(formula = suicides ~ sex + age + generation + population +
##     continent + population:age + population:continent + population:sex +
##     population:generation, family = "quasipoisson", data = train_set)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -84.250   -8.304   -3.606    0.898   84.814
```

```
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             8.836e-01  2.647e-01   3.338 0.000848 ***
## sexmale                                 1.270e+00  2.489e-02  51.031  < 2e-16 ***
## age25-34 years                          2.881e-01  4.465e-02   6.452 1.14e-10 ***
## age35-54 years                          7.394e-01  5.110e-02  14.469  < 2e-16 ***
## age5-14 years                          -2.620e+00  1.472e-01 -17.794  < 2e-16 ***
## age55-74 years                          6.261e-01  6.238e-02  10.035  < 2e-16 ***
## age75+ years                            8.850e-02  7.159e-02   1.236 0.216425
## generationG.I. Generation               2.449e-03  5.529e-02   0.044 0.964667
## generationGeneration X                 -1.284e-01  3.554e-02  -3.612 0.000304 ***
## generationGeneration Z                 -1.819e-01  2.570e-01  -0.708 0.478928
## generationMillenials                   -2.075e-01  5.557e-02  -3.734 0.000189 ***
## generationSilent                       -1.043e-01  3.974e-02  -2.624 0.008691 **
## population                              4.573e-07  7.119e-08   6.424 1.37e-10 ***
## continentAmericas                       1.983e+00  2.603e-01   7.617 2.77e-14 ***
## continentAsia                           2.530e+00  2.603e-01   9.720  < 2e-16 ***
## continentEurope                         2.865e+00  2.594e-01  11.043  < 2e-16 ***
## continentOceania                        1.601e+00  2.819e-01   5.678 1.39e-08 ***
## age25-34 years:population              -3.767e-09  3.963e-09  -0.951 0.341830
## age35-54 years:population              -1.086e-07  4.002e-09 -27.134  < 2e-16 ***
## age5-14 years:population               -2.207e-10  1.296e-08  -0.017 0.986413
## age55-74 years:population              -7.768e-08  4.565e-09 -17.018  < 2e-16 ***
## age75+ years:population                 2.028e-07  8.148e-09  24.893  < 2e-16 ***
## population:continentAmericas           -2.523e-07  7.109e-08  -3.549 0.000389 ***
## population:continentAsia               -1.643e-07  7.112e-08  -2.311 0.020869 *
## population:continentEurope             -1.621e-07  7.109e-08  -2.280 0.022638 *
## population:continentOceania             4.755e-07  9.113e-08   5.218 1.83e-07 ***
## sexmale:population                      1.287e-08  1.732e-09   7.431 1.14e-13 ***
## generationG.I. Generation:population    7.165e-08  5.150e-09  13.914  < 2e-16 ***
## generationGeneration X:population       1.682e-10  2.035e-09   0.083 0.934126
## generationGeneration Z:population      -6.549e-09  2.232e-08  -0.293 0.769246
## generationMillenials:population        -1.118e-08  4.462e-09  -2.506 0.012234 *
## generationSilent:population             3.074e-08  2.529e-09  12.155  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 165.074)
##
##     Null deviance: 11862120  on 13908  degrees of freedom
## Residual deviance:  1880745  on 13877  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
glance(glm_fit) %>%
  summarize(pseudoR2 = 1 - (deviance/null.deviance)) ## equivalent to R2 ##
```

```
## # A tibble: 1 x 1
##   pseudoR2
##      <dbl>
## 1 0.841450
```

```
  ## pseudoR2 = 0.841450
```

The final model, with 4 interaction terms, generated the highest pseudo R-squared. This quasipoisson model generated the RMSE of 407, which is a significant improvement from both the first Quasipoisson Model and the Linear Model with Interactions.

```
glm_fit <- glm(suicides ~ sex + age +
                 generation + population +
                 continent + population:age +
                 population:continent + population:sex +
                 population:generation, train_set,
               family = "quasipoisson")
y_hat4 <- predict(glm_fit, test_set, type = "response")

RMSE_glm_int <- RMSE(actual_suicides, y_hat4)

rmse_results <- bind_rows(rmse_results,
                          tibble(Method = "Quasipoisson Model, Interactions",
                                 RMSE = RMSE_glm_int))
rmse_results
```

```
## # A tibble: 5 x 2
##   Method                                RMSE
##   <chr>                                <dbl>
## 1 Mean                                913.738
## 2 Linear Model                        691.420
## 3 Linear Model, Interactions          512.796
## 4 Quasipoisson Model                 1032.60
## 5 Quasipoisson Model, Interactions    406.995
```

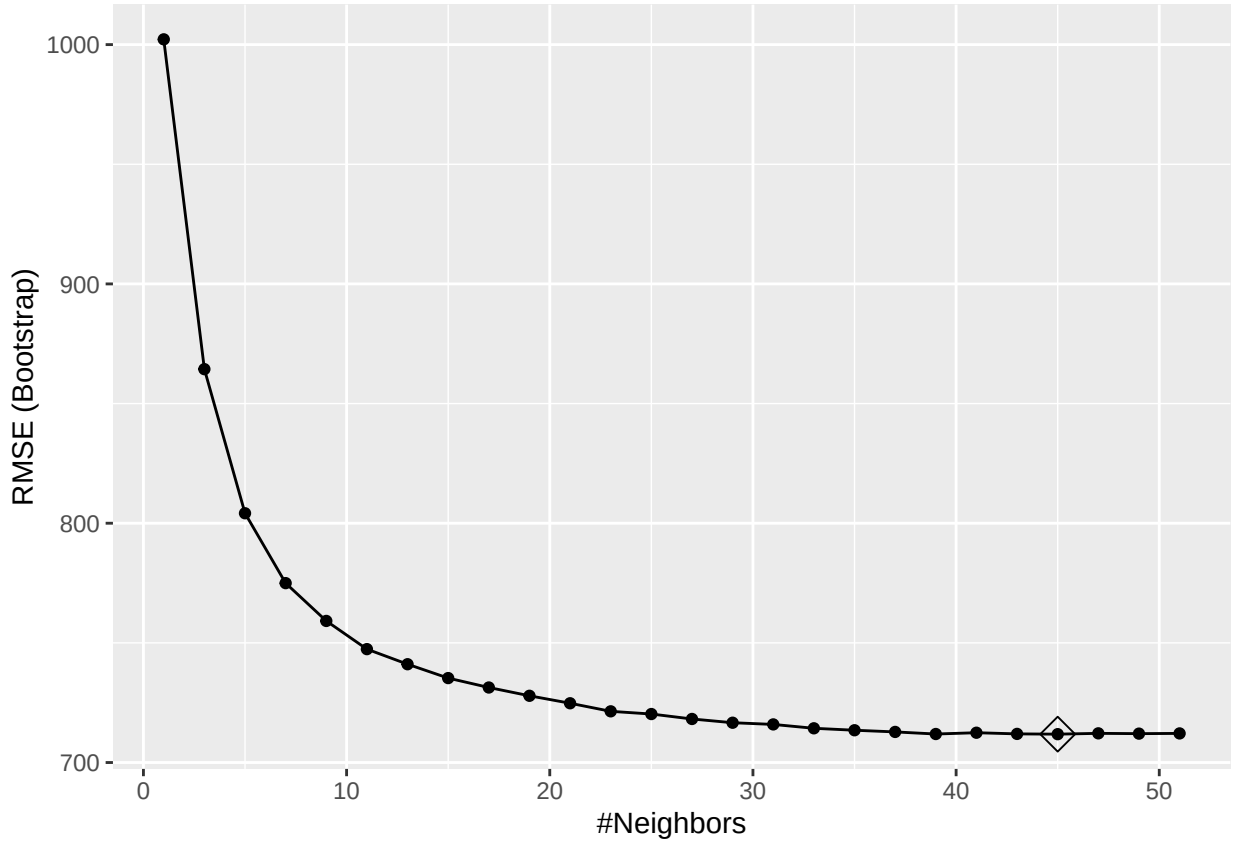
K-Nearest Neighbors

The next model is developed using K-Nearest Neighbors. Under this model, the suicide rate estimates are based on the average of its k closest points. This approach is suitable here, considering that K-Nearest Neighbors is a non-parametric approach which models complex non-linear situations well.

```
set.seed(123, sample.kind = "Rounding")
```

```
## Warning in set.seed(123, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```
train_knn <- train(suicides ~ sex + age + generation +
                     population + continent,
                   method = "knn",
                   data = train_set,
                   tuneGrid = data.frame(k = seq(1, 51, 2)))

ggplot(train_knn, highlight = TRUE)
```




```
train_knn$bestTune ## k = 45
```

```
##     k
## 23 45
```

The KNN model generated an RMSE of 718.76, which is not better than either the Linear Model with Interactions or the Quasipoisson Model with Interactions.

```
y_hat5 <- predict(train_knn, test_set)

RMSE_knn <- RMSE(actual_suicides, y_hat5)

rmse_results <- bind_rows(rmse_results,
                          tibble(Method = "K-Nearest Neighbors",
                                 RMSE = RMSE_knn))
rmse_results
```

```
## # A tibble: 6 x 2
##   Method                              RMSE
##   <chr>                              <dbl>
## 1 Mean                               913.738
## 2 Linear Model                       691.420
## 3 Linear Model, Interactions         512.796
## 4 Quasipoisson Model                1032.60
## 5 Quasipoisson Model, Interactions   406.995
## 6 K-Nearest Neighbors                718.762
```

Random Forest

The final method uses the random forest approach. This approach is a meta-estimator which takes the average of multiple decision trees, thereby improving accuracy without overfitting.

The first random forest model is developed using the ranger package. This package works well with larger datasets and is faster than the randomForest package.

```
set.seed(123, sample.kind = "Rounding")
```

```
## Warning in set.seed(123, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```
fit_rf1 <- ranger(suicides ~ sex + age + generation +
                    population + continent,
                  data = train_set,
                    num.trees = 500,
                  respect.unordered.factors = "order",
                  seed = 1234)

print(fit_rf1)
```

```
## Ranger result
##
## Call:
##  ranger(suicides ~ sex + age + generation + population + continent,      data = train_set, num.trees
##
## Type:                             Regression
## Number of trees:                  500
## Sample size:                      13909
## Number of independent variables:  5
## Mtry:                             2
## Target node size:                 5
## Variable importance mode:         none
## Splitrule:                        variance
## OOB prediction error (MSE):       79700.62
## R squared (OOB):                  0.899433
```

```
y_hat6 <- predict(fit_rf1, test_set)$predictions

RMSE_rf_ranger <- RMSE(actual_suicides, y_hat6)

rmse_results <- bind_rows(rmse_results,
                          tibble(Method = "Random Forest - Ranger",
                                 RMSE = RMSE_rf_ranger))
rmse_results
```

```
## # A tibble: 7 x 2
##   Method                              RMSE
##   <chr>                              <dbl>
## 1 Mean                               913.738
## 2 Linear Model                       691.420
## 3 Linear Model, Interactions         512.796
## 4 Quasipoisson Model                1032.60
## 5 Quasipoisson Model, Interactions   406.995
## 6 K-Nearest Neighbors                718.762
```

```
## 7 Random Forest - Ranger             276.148
```

The first Random Forest model generated an RMSE of 276.15, which is better than Quasipoisson Model with Interactions and the Linear Model with Interactions. Under this model, the population variable is the most important as it has the highest variable importance value.

```
fit_rf1$variable.importance
```

```
## NULL
```

The Random Forest model can be tuned in order to generate better results. For example, the mtry value can be optimized. Mtry refers to the number of variables which are selected at each split.

The randomForest package is used to develop this second tuned Random Forest mode. First, we examine the un-tuned Random Forest model.

```
set.seed(1234, sample.kind = "Rounding")
```

```
## Warning in set.seed(1234, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```
fit_rf2 <- randomForest(suicides ~ sex + age + generation +
                          population + continent,
                        data = train_set,
                        ntree = 500)
which.min(fit_rf2$mse)
```

```
## [1] 23
```

```
print(fit_rf2)
```

```
##
## Call:
##  randomForest(formula = suicides ~ sex + age + generation + population +      continent, data = trai
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 1
##
##           Mean of squared residuals: 283131.4
##                     % Var explained: 64.27
```

```
y_hat7 <- predict(fit_rf2, test_set)

RMSE_rf_1 <- RMSE(actual_suicides, y_hat7)

rmse_results <- bind_rows(rmse_results,
                          tibble(Method = "Random Forest - Not Tuned",
                                 RMSE = RMSE_rf_1))
rmse_results
```

```
## # A tibble: 8 x 2
##   Method                              RMSE
##   <chr>                              <dbl>
## 1 Mean                               913.738
## 2 Linear Model                       691.420
## 3 Linear Model, Interactions         512.796
## 4 Quasipoisson Model                1032.60
## 5 Quasipoisson Model, Interactions   406.995
## 6 K-Nearest Neighbors                718.762
```

```
## 7 Random Forest - Ranger            276.148
## 8 Random Forest - Not Tuned         543.567
```

This un-tuned Random Forest model generated an RMSE of 543.57, which is higher than that generated by the un-tuned Random Forest - Ranger model.

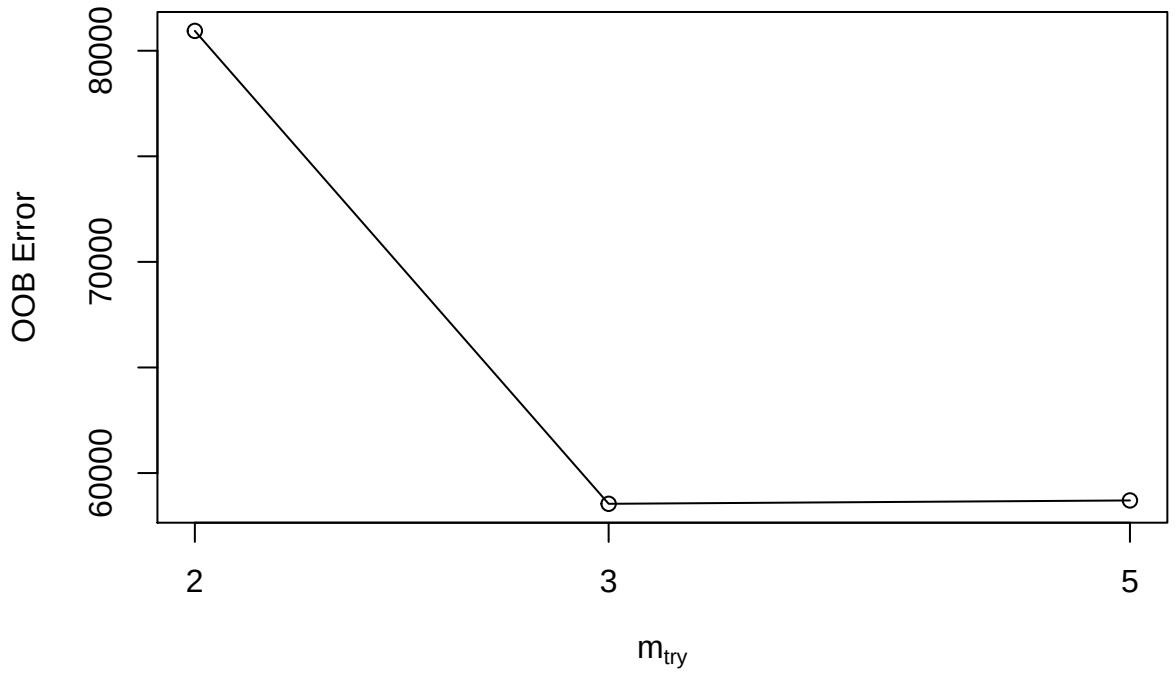
Next, the Random Forest model is tuned in order to determine the optimal mtry value. We start with an mtry of 3 (mtryStart), which is increased by a factor of 2 (stepFactor) until the out-of-bag (OOB) error stops improving by 1%.

```
features <- setdiff(names(train_set), "suicides")

set.seed(1234, sample.kind = "Rounding")
```

```
## Warning in set.seed(1234, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```
mtry <- tuneRF(x = train_set[features],
               y = train_set$suicides,
               ntreeTry = 500,
               mtryStart = 3,
               stepFactor = 2,
               improve = 0.01,
               trace = FALSE)
```

```
## -0.3825264 0.01
## -0.002713164 0.01
```



```
best_m <- mtry[mtry[,2] == min(mtry[,2]), 1]
print(mtry)
```

```
##   mtry OOBError
## 2    2 80939.13
## 3    3 58544.36
## 5    5 58703.20
```

```
print(best_m)
```

```
## [1] 3
```

Here, optimal mtry = 3 as it has the lowest OOB error. Next, we develop the model using the best mtry value.

```
set.seed(123, sample.kind = "Rounding")
```

```
## Warning in set.seed(123, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```
fit_rf3 <- randomForest(suicides ~ sex + age + generation +
                          population + continent,
                        data = train_set,
                        mtry = best_m,
                        importance = TRUE,
                        ntree = 500)
print(fit_rf3)
```

```
##
## Call:
##  randomForest(formula = suicides ~ sex + age + generation + population +      continent, data = trai
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 57907.79
##                     % Var explained: 92.69
```

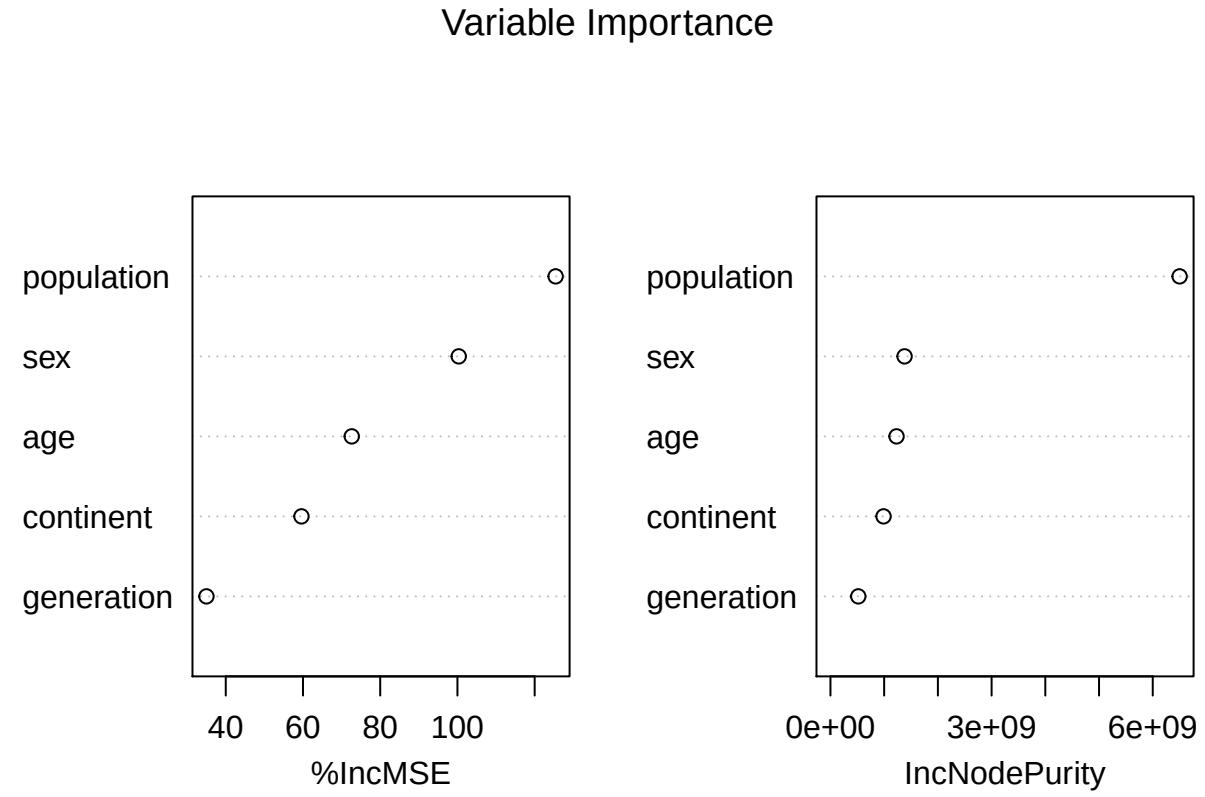
```
y_hat8 <- predict(fit_rf3, test_set)

RMSE_rf_2 <- RMSE(actual_suicides, y_hat8)

rmse_results <- bind_rows(rmse_results,
                          tibble(Method = "Random Forest - Tuned",
                                 RMSE = RMSE_rf_2))
rmse_results
```

```
## # A tibble: 9 x 2
##   Method                              RMSE
##   <chr>                              <dbl>
## 1 Mean                               913.738
## 2 Linear Model                       691.420
## 3 Linear Model, Interactions         512.796
## 4 Quasipoisson Model                1032.60
## 5 Quasipoisson Model, Interactions   406.995
## 6 K-Nearest Neighbors                718.762
## 7 Random Forest - Ranger             276.148
```

```
## 8 Random Forest - Not Tuned         543.567
## 9 Random Forest - Tuned             219.898
```

The tuned Random Forest model generated the lowest RMSE of 219.90. As with the first Random Forest model, the population variable is the most important in the Tuned Model.

```
varImpPlot(fit_rf3, main ="Variable Importance")
```




```
varImp(fit_rf3)
```

```
##              Overall
## sex        100.34178
## age         72.63087
## generation  35.00050
## population 125.43243
## continent   59.60981
```

SUMMARY AND CONCLUSIONS

This project aimed to examine global suicide trends using machine learning models in order to determine whether demographic variables were associated with suicide rates. The effectiveness of the models were assessed using the RMSE.

K-Nearest Neighbors is among the least effective of the approaches, as the KNN model generated a high RMSE.

We found that linear models did not adequately capture the relationship between suicide rates, on the one hand, and demographic variables on the other. This suggests that the relationship between suicide rates and

these variables are non-linear in nature. Nevertheless, the adjusted R-squared values of the linear models improved when (1) the population variable and (2) population's interaction with the other variables were included.

The Quasipoisson Model with Interactions fared marginally better. However, there seems to still be a significant lack of fit considering the residual deviance of 1,880,745 with 13,877 degrees of freedom.

The Random Forest:Ranger and Random Forest:Tuned Models performed best, generating an RMSE of 276.15 and 219.90 respectively. These models show that the population variable is the most important, followed by sex.

This analysis was limited to an exploration of the association between suicide rates and the available demographic variables. The models developed here suggest that suicide rates are associated with population and sex. Specifically, males are more prone to suicides.

The relationship between population and suicides, however, is more complex, due to the interaction effect of other variables. Thus, this project can be expanded upon by looking into the interaction of population with other, possibly, confounding variables.