Final Paper Part 2

Team 4: Dhiraj Lahoti, Sravya Lenka, Iskander Lou, Yeri Jeong

Data: https://www.kaggle.com/code/lmorgan95/r-suicide-rates-in-depth-stats-insights/data

1. Your specific research question

- How do demographic (e.g. gender, age, generation) and temporal factors(e.g. year, GDP) in the United States correlate with suicide rates?

2. A brief description of your dataset.
Be sure to mention the unit of analysis in your dataset (e.g., individuals, households, countries), as well as the number of observations in the dataset (the "n" about whom there is information for your key variables)

- This dataset was sourced from the World Health Organisation's Global Suicide Trends & Analysis.
- Unit of analysis: A demographic group (of certain age and gender) in the United states in a certain year. For example, female Americans of age 25-34 in 1994.
- Number of observations: 372.

3. Description of your dependent variable and its descriptive statistics

- Dependent variable: Suicide per 100k population
    - Min:0.260
    - Max:3.973
    - Median:6.890
    - Mean:13.820
    - Sd: 13.22987

4. Description of your independent variable(s) and its descriptive statistics

- Gender
    - 2 levels (male,female)
    - Both levels have 186 observations
- Age
    - 6 levels (5-14, 15-24, 25-34, 35-54, 55-74, 75+)
    - All levels have 62 observations
- Year
    - Levels: 31 (1985~2016)
    - All years have 12 observations
- GDP per capita
    - Median: 39,218

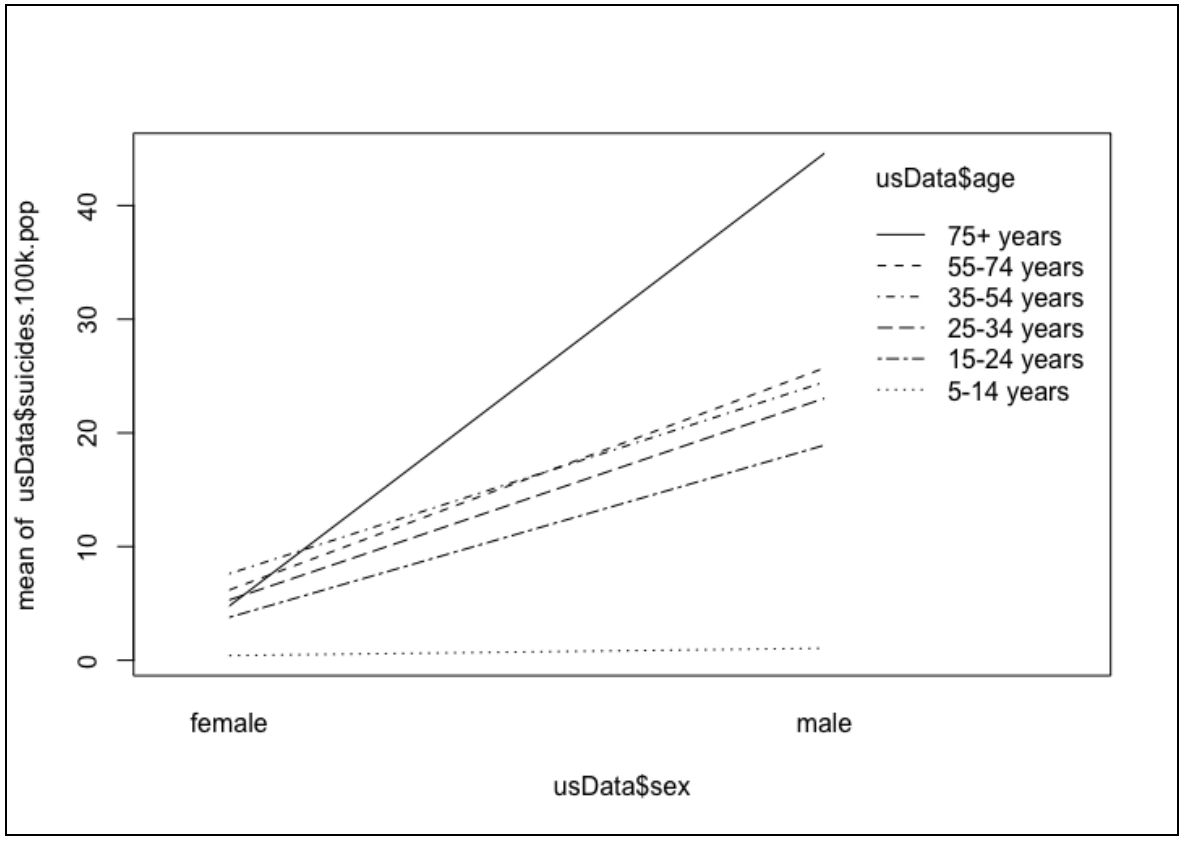
- Mean: 39,270
- SD: 12,334.12

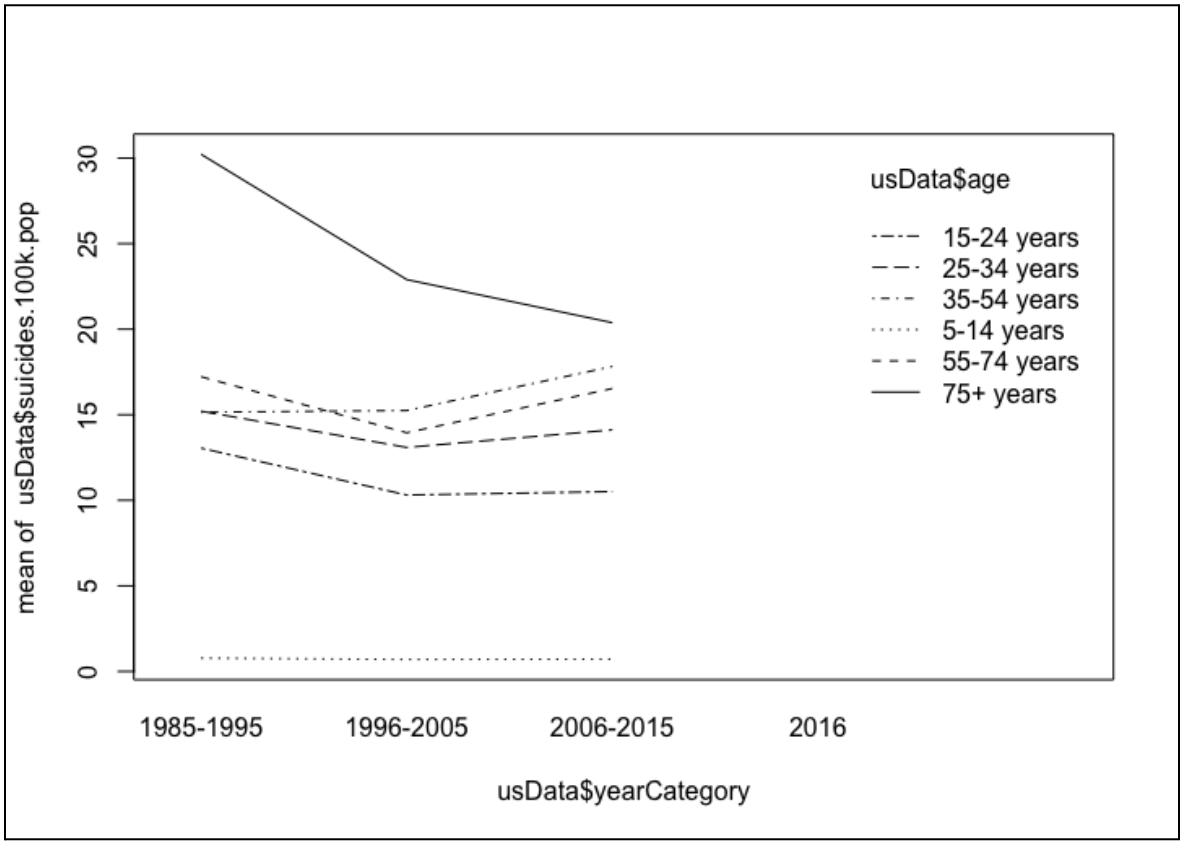
5. Preliminary/rough answer to your research question

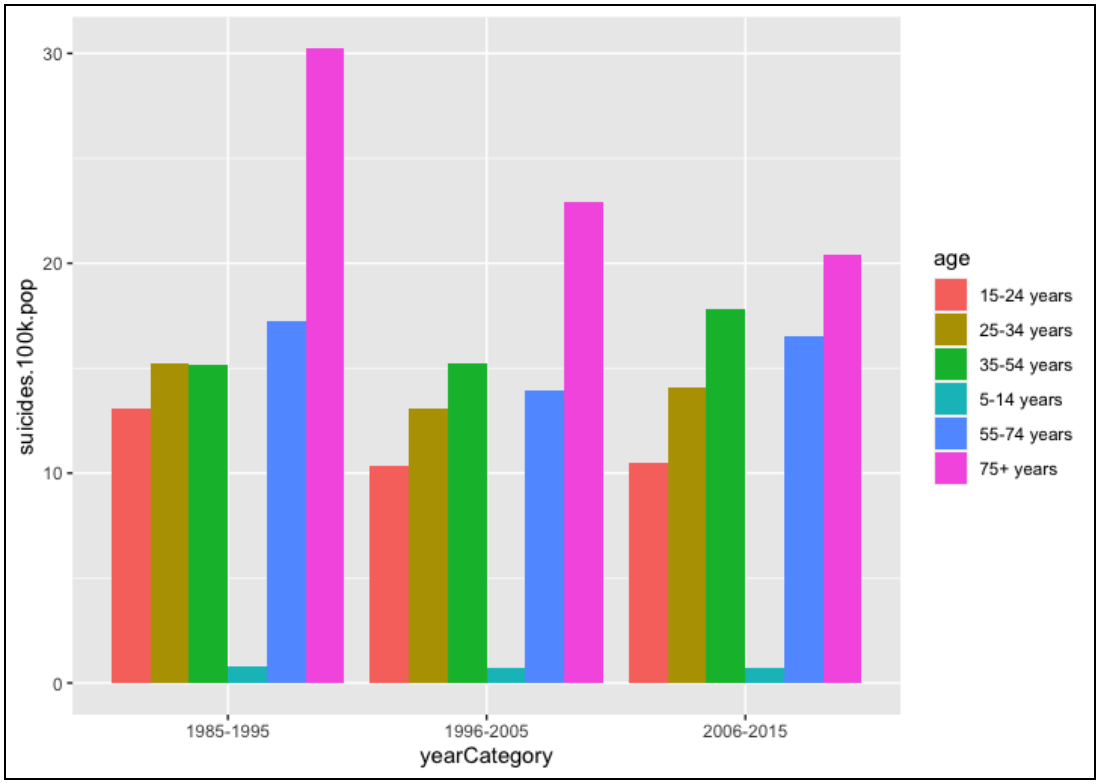
We found a significant relationship between most factors that we tested (e.g. sex, year, generation, and age) and suicide rates, however there was no relationship between GDP per capita and suicide rates. Between female and males, males had higher suicide rates compared to females. Among the age groups, the 75+ age subset had the highest suicide rates compared to other age groups.

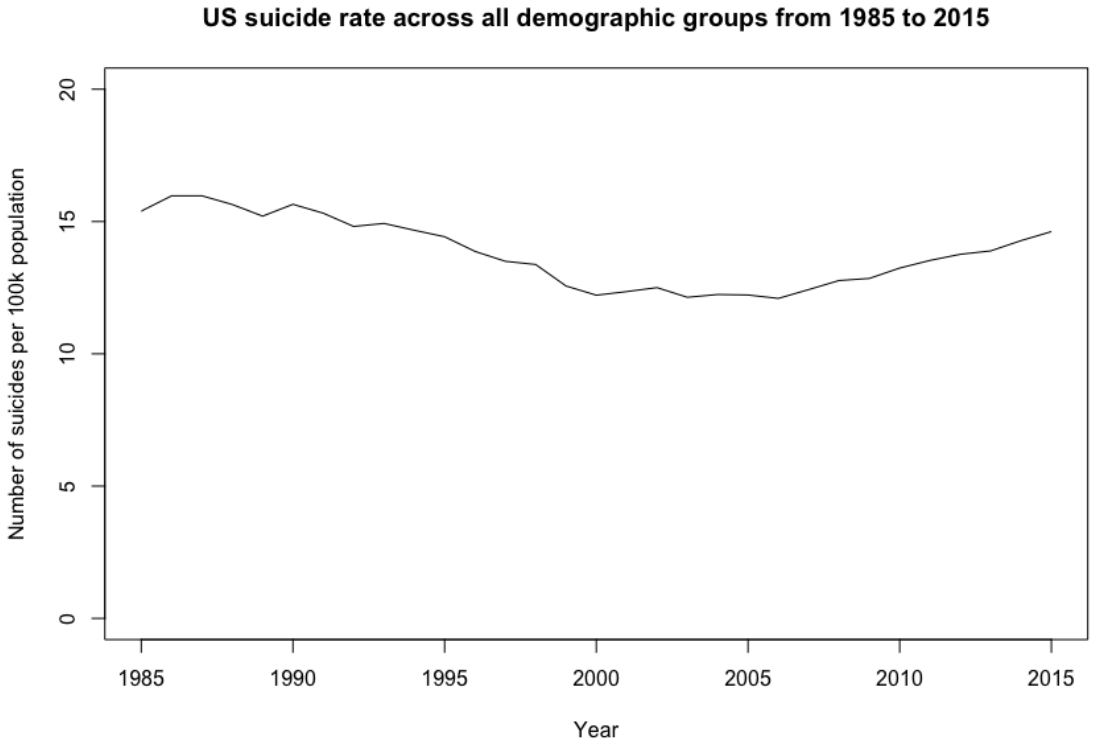
```
> summary(aov(usData$suicides.100k.pop ~ usData$sex + usData$year + usData$generation +usData$age + usData$gdp_per_ca
pita....))
                         Df Sum Sq Mean Sq F value Pr(>F)
usData$sex                1  31100   31100 799.917 <2e-16 ***
usData$year               1    240     240   6.163 0.0135 *
usData$generation         5  15862    3172  81.598 <2e-16 ***
usData$age                5   3815     763  19.626 <2e-16 ***
usData$gdp_per_capita....   1      0       0   0.007 0.9314
Residuals               358  13919      39
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

```
> summary(aov(usData$suicides.100k.pop ~ usData$sex * usData$year *  usData$generation* usData$age * usData$gdp_per_capita....))
                                                                      Df Sum Sq Mean Sq   F value   Pr(>F)
usData$sex                                                             1  31100   31100 93476.582  < 2e-16 ***
usData$year                                                            1    240     240   720.227  < 2e-16 ***
usData$generation                                                      5  15862    3172  9535.412  < 2e-16 ***
usData$age                                                             5   3815     763  2293.418  < 2e-16 ***
usData$gdp_per_capita....                                              1      0       0     0.868 0.352463
usData$sex:usData$year                                                 1    227     227   683.551  < 2e-16 ***
usData$sex:usData$generation                                           5   9471    1894  5693.588  < 2e-16 ***
usData$year:usData$generation                                          5    138      28    83.071  < 2e-16 ***
usData$sex:usData$age                                                  5   3090     618  1857.214  < 2e-16 ***
usData$year:usData$age                                                 5    370      74   222.490  < 2e-16 ***
usData$generation:usData$age                                           5     68      14    41.089  < 2e-16 ***
usData$sex:usData$gdp_per_capita....                                   1      4       4    11.204 0.000945 ***
usData$year:usData$gdp_per_capita....                                  1     47      47   141.007  < 2e-16 ***
usData$generation:usData$gdp_per_capita....                            5     18       4    10.616 3.06e-09 ***
usData$age:usData$gdp_per_capita....                                   5     12       2     7.455 1.58e-06 ***
usData$sex:usData$year:usData$generation                               5     48      10    29.141  < 2e-16 ***
usData$sex:usData$year:usData$age                                      5    183      37   109.922  < 2e-16 ***
usData$sex:usData$generation:usData$age                                5     37       7    22.249  < 2e-16 ***
usData$year:usData$generation:usData$age                               5      4       1     2.594 0.026204 *
usData$sex:usData$year:usData$gdp_per_capita....                       1      9       9    27.138 4.03e-07 ***
usData$sex:usData$generation:usData$gdp_per_capita....                 5      9       2     5.325 0.000115 ***
usData$year:usData$generation:usData$gdp_per_capita....                5     38       8    22.990  < 2e-16 ***
usData$sex:usData$age:usData$gdp_per_capita....                        5      7       1     4.028 0.001570 **
usData$year:usData$age:usData$gdp_per_capita....                       5     17       3    10.194 6.97e-09 ***
usData$generation:usData$age:usData$gdp_per_capita....                 5      1       0     0.657 0.656219
usData$sex:usData$year:usData$generation:usData$age                    5      5       1     2.792 0.017935 *
usData$sex:usData$year:usData$generation:usData$gdp_per_capita....     5     19       4    11.635 4.28e-10 ***
usData$sex:usData$year:usData$age:usData$gdp_per_capita....            5     11       2     6.581 9.14e-06 ***
usData$sex:usData$generation:usData$age:usData$gdp_per_capita....      5      1       0     0.555 0.734814
usData$year:usData$generation:usData$age:usData$gdp_per_capita....     5      2       0     1.199 0.310131
usData$sex:usData$year:usData$generation:usData$age:usData$gdp_per_capita....   5      0       0     0.257 0.936258
Residuals                                                            244     81       0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```


mean of usData$suicides.100k.pop
0
10
20
30
40
usData$age
75+ years
55-74 years
35-54 years
25-34 years
15-24 years
5-14 years
female
male
usData$sex


mean of usData$suicides.100k.pop
0
5
10
15
20
25
30
usData$age
15-24 years
25-34 years
35-54 years
5-14 years
55-74 years
75+ years
1985-1995
1996-2005
2006-2015
2016
usData$yearCategory


suicides.100k.pop
30
20
10
0
1985-1995
1996-2005
2006-2015
yearCategory
age
15-24 years
25-34 years
35-54 years
5-14 years
55-74 years
75+ years


US suicide rate across all demographic groups from 1985 to 2015
Number of suicides per 100k population
0
5
10
15
20
1985
1990
1995
2000
2005
2010
2015
Year