Riley Payung

CDS 301

03/02/2020

Assignment 5

Problem 1

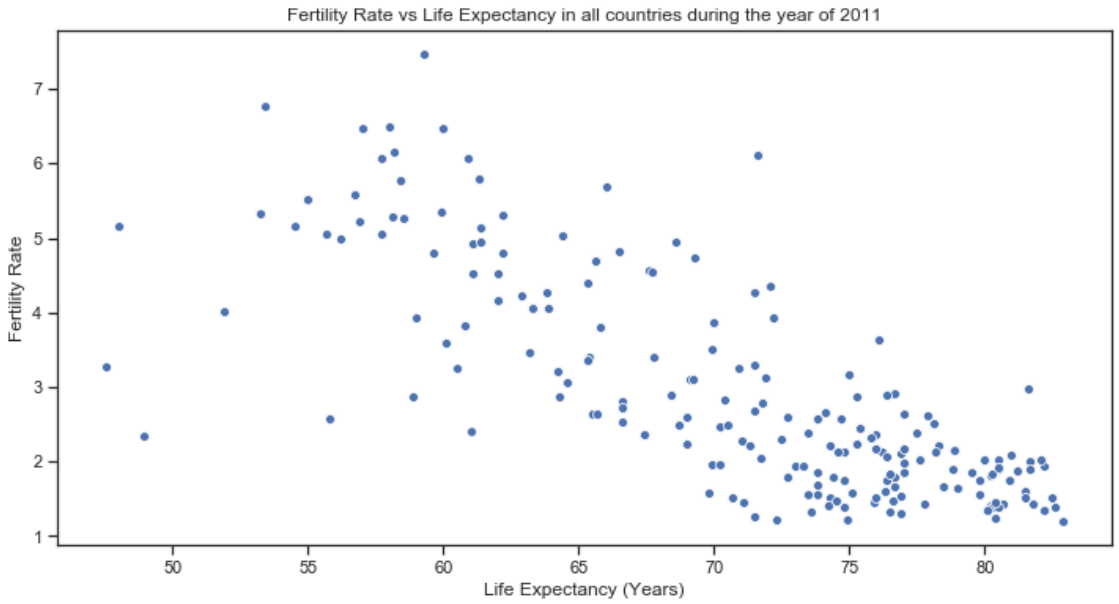
I wrote a function for selecting based on the year that visualizes the life-expectancy and fertility rate of ALL countries by the year:

```
def scatterYear(year):
    df2 = pd.read_csv("./data/gapminder-1.csv")
    df2.dropna()
    df2 = df2.loc[df2['year'] == year];
    fig = plt.figure(figsize=(12,6))
    sns.scatterplot(df2['life-expectancy'],df2['fertility-rate'])
    plt.title("Fertility Rate vs Life Expectancy in all countries during the year of " + str(year))
    plt.xlabel("Life Expectancy (Years)")
    plt.ylabel("Fertility Rate")
    plt.show()

scatterYear(2010)
```

Interestingly, the year 2010 had the best rate, if we were to apply a linear regression, we would likely find that 2010 had the best fit.

```
[71]: scatterYear(2011)
```



Straight after the year of the best rate, it went back to normal.