Andrew Ivanov

April 12ᵗʰ 2020

CIS 4170

Professor O' Connell

For this assignment I found a dataset on Kaggle which consisted of health statistics gathered by the World Health Organization. I wanted to visualize how significant the differences are between developed countries such as the United States and developing countries.

https://www.kaggle.com/kumarajarshi/life-expectancy-who

The first variables I decided to investigate was the life expectancy in developed countries versus developing countries using a box plot. The life expectancy in developed countries is much more concentrated, the median age is around 82. The highest recorded average age in a developed country is around 87 and the lowest is around 73. In developing countries, the range of life expectancy is much more scattered. The average age of 72 is around the lowest recorded age for developed countries. An interesting observation is that there is a developing country with an average life expectancy of 85 which is higher than the average life expectancy for developed countries.
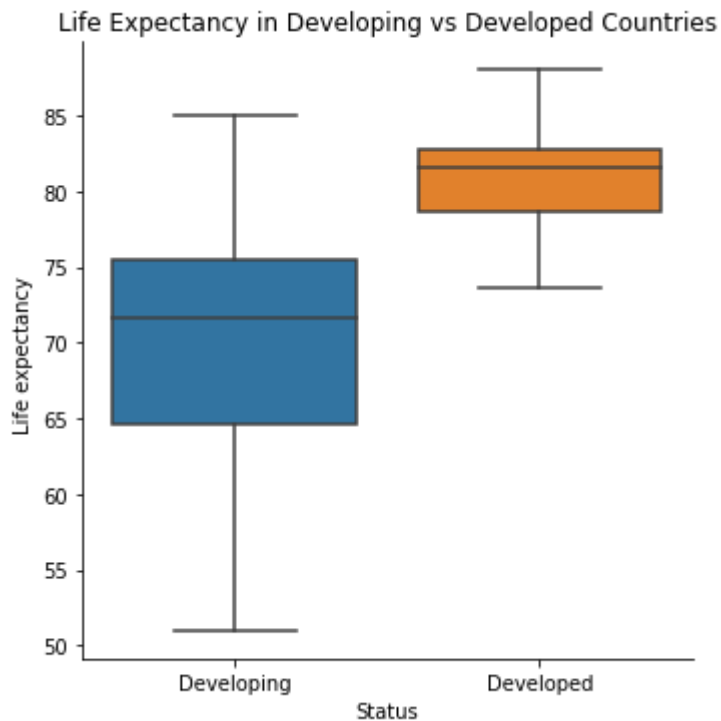
Next, I wanted to look at how another factor influences life expectancy. I decided to look at the level of schooling versus life expectancy. I thought that a higher level of schooling would mean that there are more people in a country with a significant knowledge of medicine and other fields which are beneficial to health.

Level of Schooling in Developing vs Developed Countries
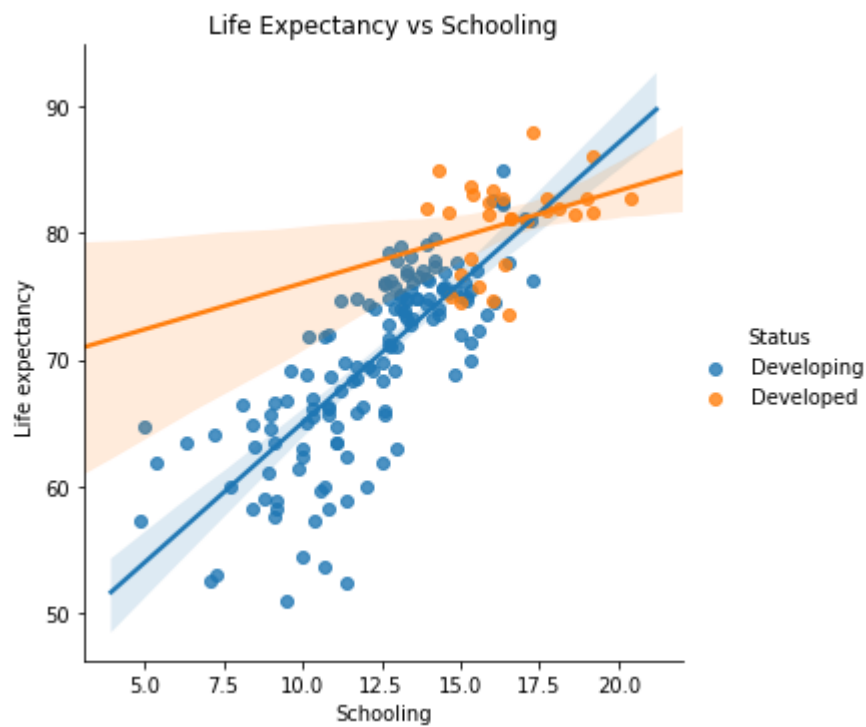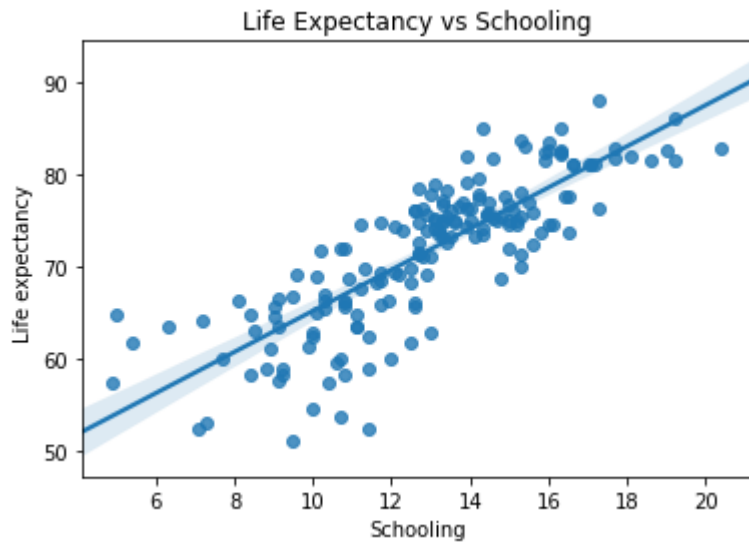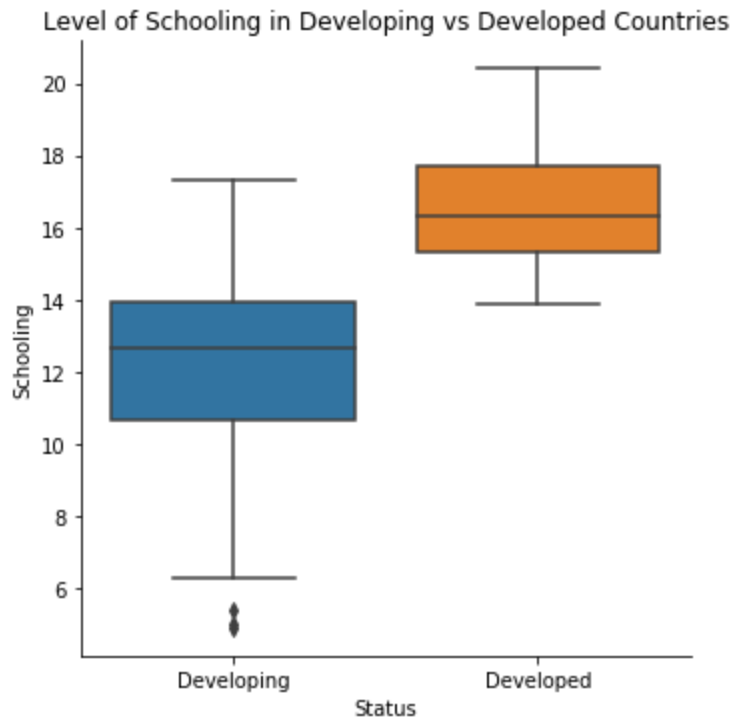
The box plot for the level of schooling resembles the box plot for life expectancy. There appears to be a strong correlation between schooling and life expectancy for all countries. When we separate the graph by developing and developed countries, we see that the correlation is actually steeper for developing countries.

Afterwards I wanted to see if GDP had any impact on life expectancy.

Life Expectancy vs GDP

Life Expectancy vs GDP

Life Expectancy in Developing vs Developed Countries

GDP vs Life Expectancy
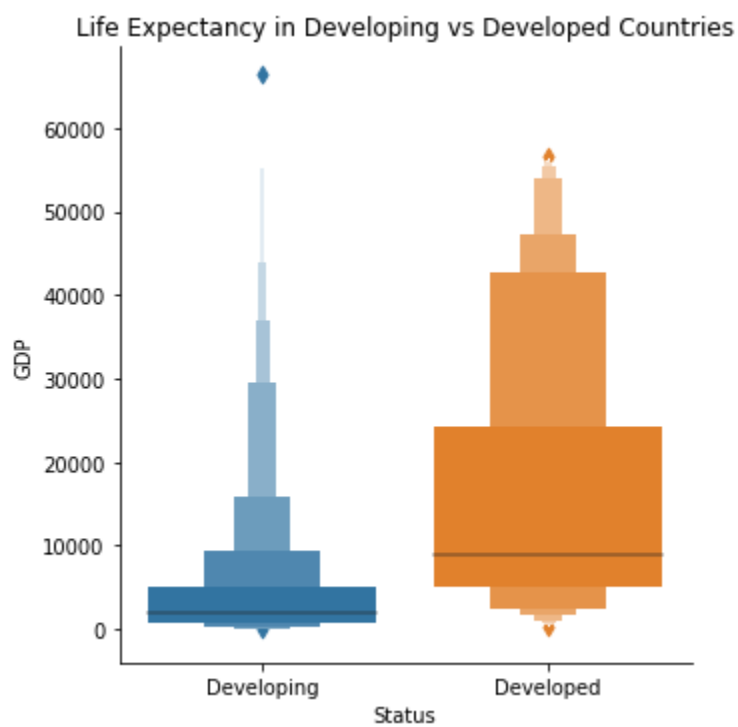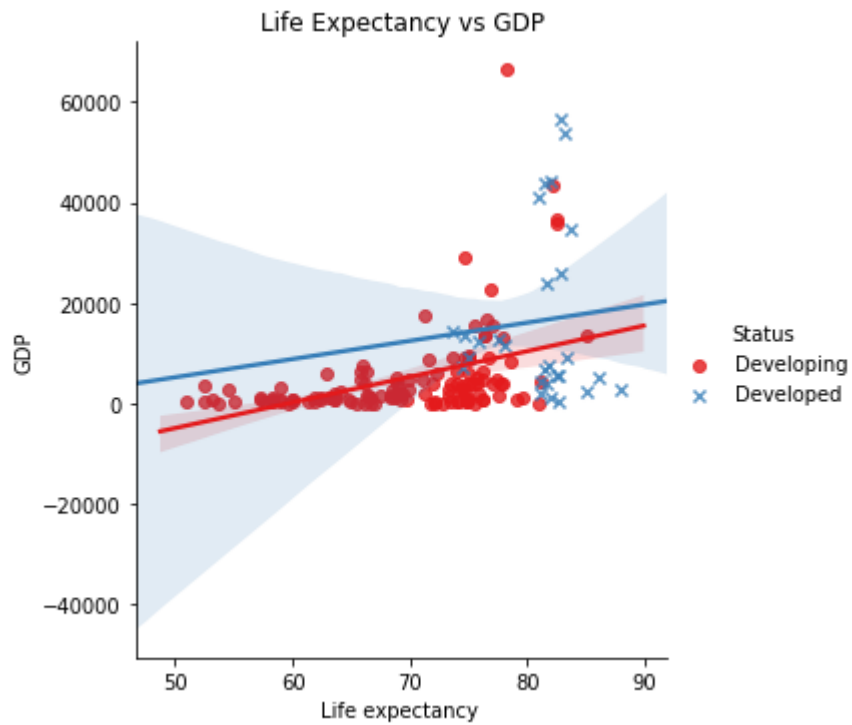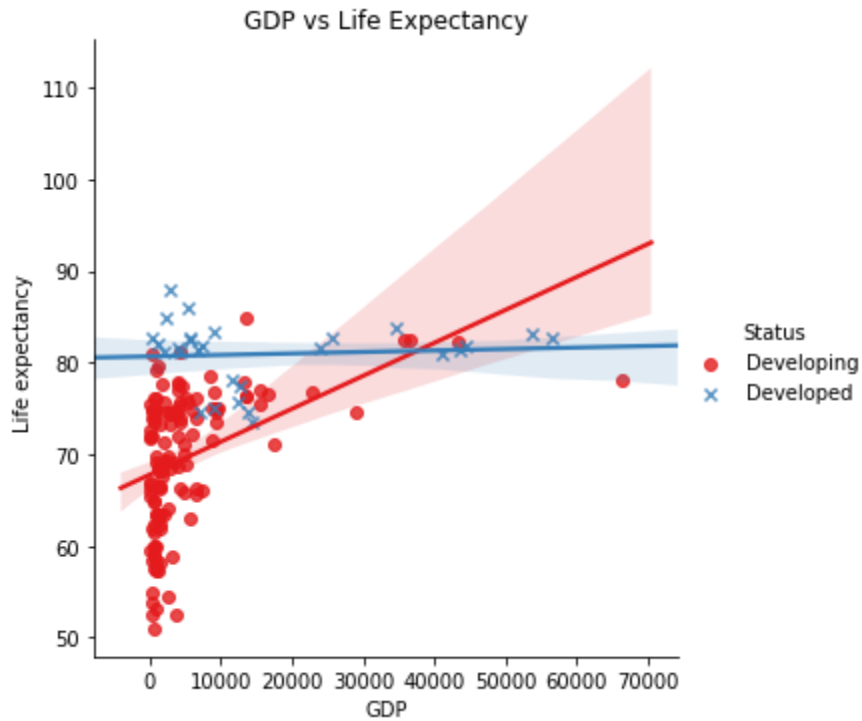
Many developing countries have very low GDPs close to 0. There are some developed countries with GDPs close to 0 as well but the median is 10,000 for developed countries. Interestingly, the country with highest GDP is a developing country. I filtered the dataset to find this country, and apparently China is considered a developing country by the WHO.

```
Code:
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
data = pd.read_csv(r'C:\Users\Andrew\Desktop\Python\LifeExpectancyData.csv')


df =data[data.Year == 2015]
sns.catplot(x="Status", y="Life expectancy ", kind="box", data= df)
plt.title('Life Expectancy in Developing vs Developed Countries')


sns.regplot(x="Schooling", y="Life expectancy ", data=df);
plt.title('Life Expectancy vs Schooling')


sns.lmplot(x="Schooling", y="Life expectancy ", hue = "Status", data=df);
plt.title('Life Expectancy vs Schooling')


sns.catplot(x="Status", y= "Schooling", kind="box", data= df)
plt.title('Level of Schooling in Developing vs Developed Countries')


sns.regplot(x="Life expectancy ", y="GDP", data=df);
plt.title('Life Expectancy vs GDP')


sns.lmplot(x="Life expectancy ", y="GDP", hue = "Status", data=df, markers=["o", "x"], palette="Set1");
plt.title('Life Expectancy vs GDP')


sns.catplot(x="Status", y="GDP", kind="boxen", data= df)
plt.title('Life Expectancy in Developing vs Developed Countries')
```

```python
sns.lmplot(x="GDP", y="Life expectancy ", hue = "Status", data=df, markers=["o", "x"], palette="Set1");

plt.title('GDP vs Life Expectancy')
```