

CMSE-201 Project_Pharm

April 14, 2020

CMSE 201 Final Project

0.0.1 DOM_PHAM

0.0.2 Section_001

04-17-2020

1 *CMSE 201 Final Project*

1.1 Background and Motivation

As an individual who consumes soft drinks pretty frequently, I wanted to observe the relationship between soft drink consumption and health. Many agree that the consumption of sugar is highly correlated with weight gain and consuming too much of it can have adverse effects on an individual's health. As a result many believe that drinking too much soda can lead to obesity. In this project I aim to answer: 1.) Which countries consume the most soft drinks? 2.) What were their adult obesity rates in the given year? 3.) How are soft drink consumption and obesity related? 4.) What other factors contribute to obesity? 5.) Which factor is the most significant to increasing obesity rates?

1.2 Methodology

```
[37]: # Import tools
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

The cells below load a data set I created that shows the soda consumption for each country as well as the percentage of obese adults relative to the country's population. The data was based on the 10 countries that consumed the most soft drinks per capita in 2017.

```
[6]: # Load data
Drinks = pd.read_csv('Country Drinks.csv')
```

```
[52]: # Display the 5 highest rows
Drinks.tail()
```

```
[52]:
```

	Country	Soda Consumption (L)	Obesity Rate (%)
5	Uruguay	113	26.7
6	Mexico	137	28.1
7	Chile	141	27.8
8	United States	154	33.7
9	Argentina	155	26.3

```
[8]: # Create variables
Country = Drinks['Country']
Soda = Drinks['Soda Consumption (L)']
Obesity = Drinks['Obesity Rate (%)']
```

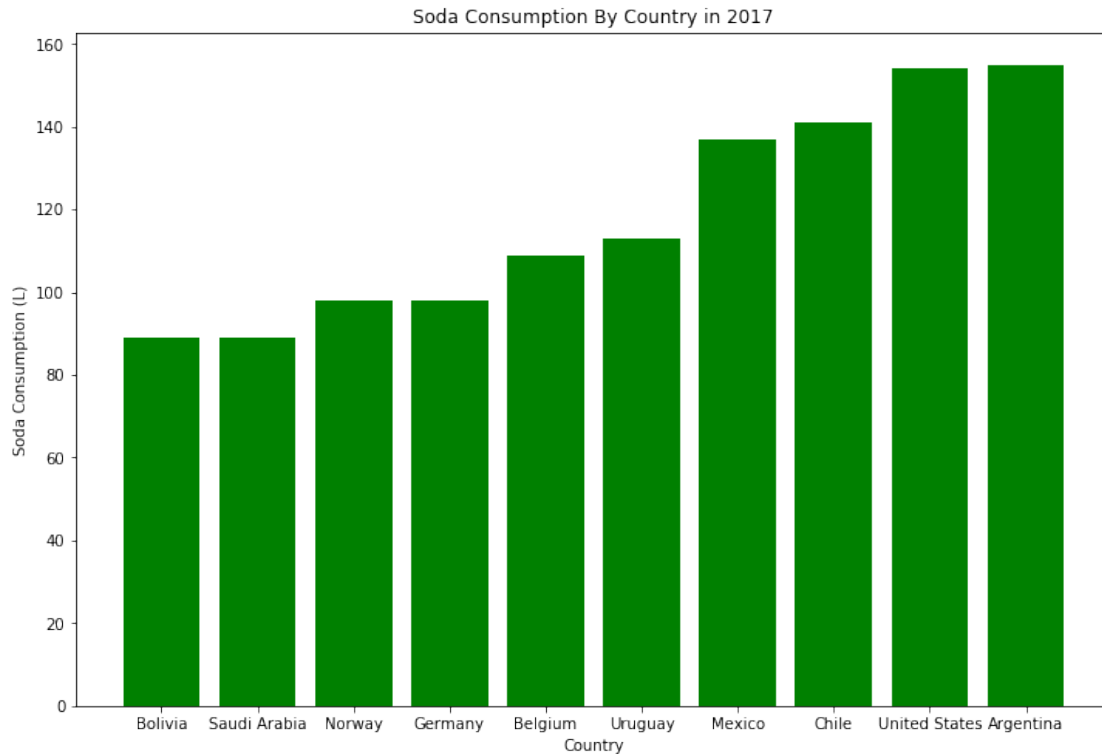
The following cells create and display a bar chart that visualizes each country's soda consumption for the year 2017. The countries are arranged in ascending order with Bolivia being ranked #10 and Argentina ranked #1 in terms of soft drink consumption. The amount of soda consumed is measured in liters for this data set.

```
[9]: # Create bar chart
plt.figure(figsize = (12,8))
x = ['Bolivia', 'Saudi Arabia', 'Norway', 'Germany', 'Belgium', 'Uruguay',
     → 'Mexico', 'Chile', 'United States', 'Argentina']
y = [89, 89, 98, 98, 109, 113, 137, 141, 154, 155]
plt.bar(x, y, color = 'g')

# Title
plt.title("Soda Consumption By Country in 2017")

# X and Y axes
plt.xlabel('Country')
plt.ylabel('Soda Consumption (L)')

#Plot
plt.show()
```



The barplot below shows the obesity rates in each country based on how much soda was consumed in 2017.

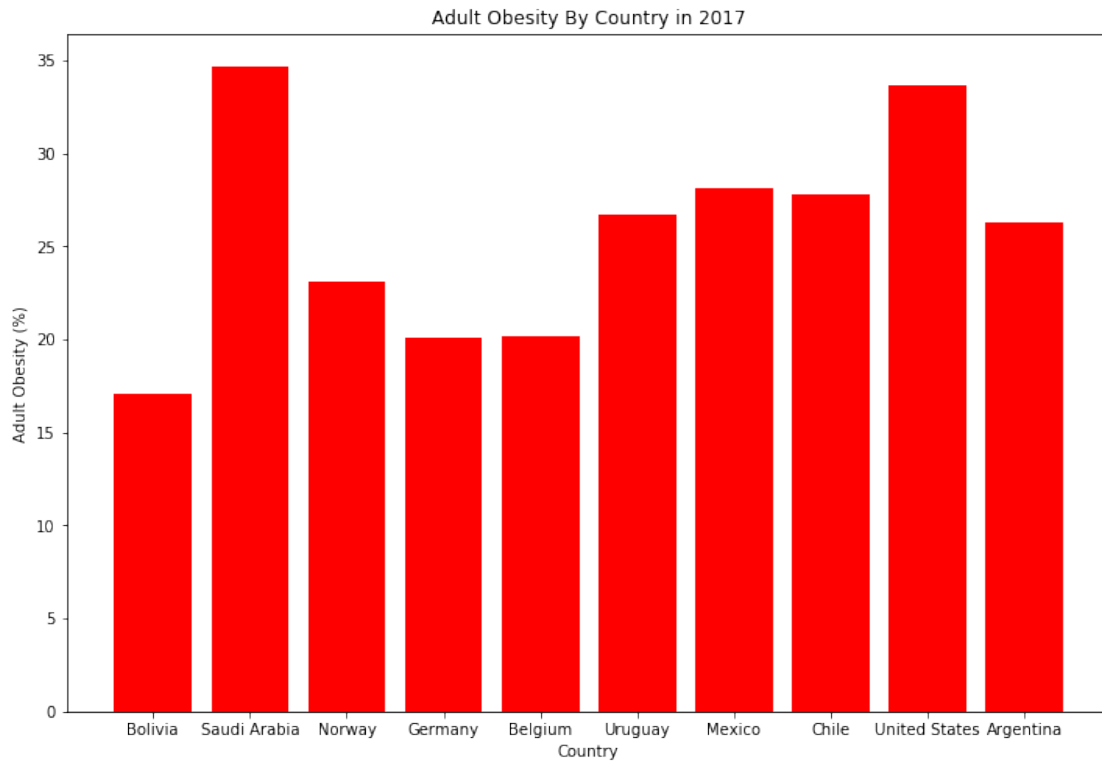
```
[10]: # Figure size
plt.figure(figsize = (12,8))

# Create bar chart
x = ['Bolivia', 'Saudi Arabia', 'Norway', 'Germany', 'Belgium', 'Uruguay', 'Mexico', 'Chile', 'United States', 'Argentina']
y = [17.1, 34.7, 23.1, 20.1, 20.2, 26.7, 28.1, 27.8, 33.7, 26.3]
plt.bar(x, y, color = 'r')

# Title
plt.title("Adult Obesity By Country in 2017")

# X and Y axes
plt.xlabel('Country')
plt.ylabel('Adult Obesity (%)')

#Plot
plt.show()
```



[]:

1.3 Results

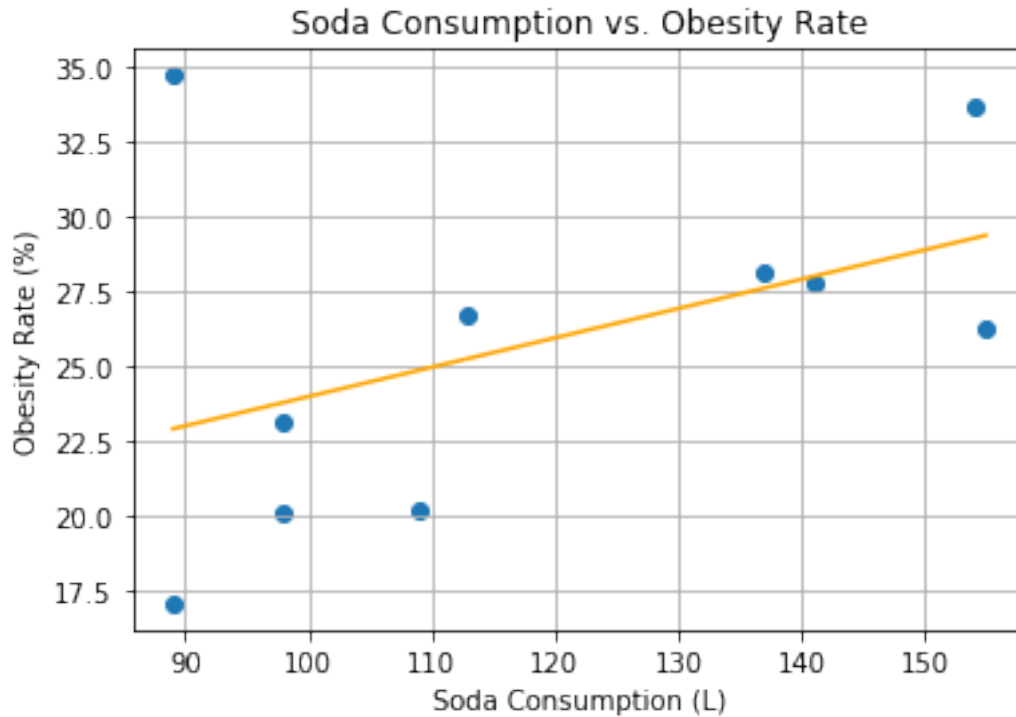
```
[35]: # Create scatter plot
plt.scatter(Soda, Obesity)

# Line of best fit
m, b = np.polyfit(Soda, Obesity, 1)
plt.plot(Soda, m*Soda + b, color = 'orange')

# Title
plt.title('Soda Consumption vs. Obesity Rate')

# X and Y axes
plt.xlabel('Soda Consumption (L)')
plt.ylabel('Obesity Rate (%)')

# Plot
plt.grid()
plt.show()
```



```
[13]: x_values = Soda
      y_values = Obesity

      correlation_matrix = np.corrcoef(x_values, y_values)
      correlation = correlation_matrix[0,1]
      r_squared = correlation**2

      print('The correlational coefficient between soda consumption and the
            ↳percentage of overweight adults is', r_squared, 'this means that about', "%.
            ↳2f" % (r_squared*100),
            '% of the variation in the percentage of overweight adults can be
            ↳explained by the amount of soda consumed by a country.')
```

The correlational coefficient between soda consumption and the percentage of overweight adults is 0.19665742805935835 this means that about 19.67 % of the variation in the percentage of overweight adults can be explained by the amount of soda consumed by a country.

```
[40]: # Load data
      Health = pd.read_csv('data.csv')
```

```
[41]: # Show first 10 rows
      Health.head(10)
```

```
[41]:
```

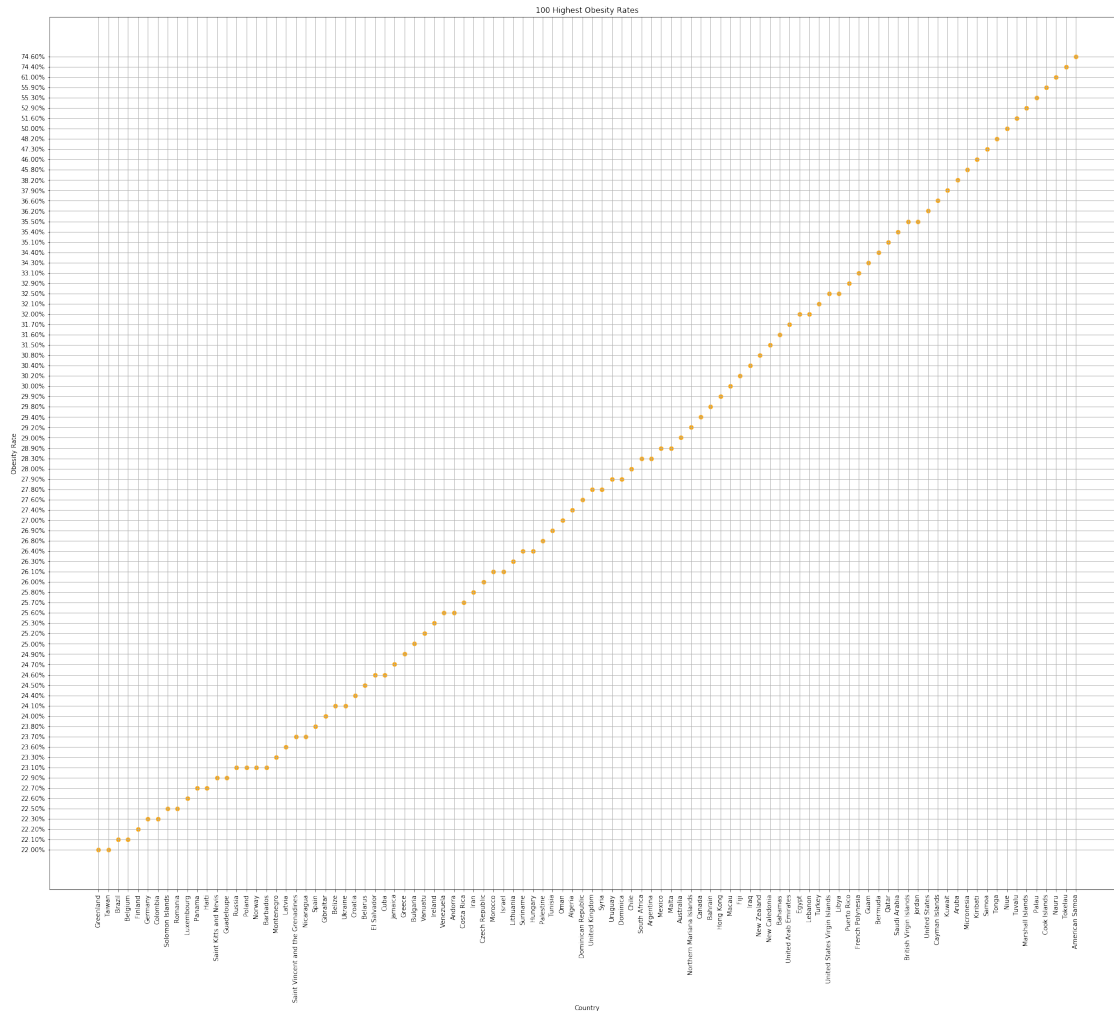
	Name	Obesity Rate
0	Greenland	22.00%
1	Taiwan	22.00%
2	Brazil	22.10%
3	Belgium	22.10%
4	Finland	22.20%
5	Germany	22.30%
6	Colombia	22.30%
7	Solomon Islands	22.50%
8	Romania	22.50%
9	Luxembourg	22.60%

The cell below shows the 100 countries with the highest obesity rates in 2020. As you can see the US, one of the top 10 soft drink consuming countries in 2017, ranked 16th. This could imply several things. According to an article by the New York Times the consumption of soda has been in a decline since 2014. This would also suggest that soda consumption is not a major cause of obesity. Based on the graph, Pacific Islander nations lead the world in obesity rates. This could mean that cultural differences such as diet might play a bigger role. Although their populations are much smaller, they also suffer from increasing demands for cheap, fast food and convenient goods.

```
[43]: # Set figure size
plt.figure(figsize = (30, 25))

# Create a plot
plt.scatter(Health['Name'], Health['Obesity Rate'], color = 'orange')
plt.title('100 Highest Obesity Rates')
plt.xlabel('Country')
plt.ylabel('Obesity Rate')

# Adjust x-axis labels
degrees = 90
plt.xticks(rotation=degrees)
plt.grid()
plt.show()
```



With such a small correlation between soda consumption and obesity I decided to determine if other factors had a more significant impact on a country's increasing obesity. In the cell below I decide to compare access to fast food and obesity. In this case I based fast food access on how many McDonalds franchises we're in a country since the chain is arguably the largest fast food chain in the world.

```
[44]: # Load data
McD = pd.read_csv('McDonalds.csv')
```

The cell below shows the first 10 countries out of 25 with the most McDonalds restaurants in the world. The data is organized by increasing obesity rates.

```
[45]: # Show first 10 rows
McD
```

```
[45]:
```

	Top25	Number of McDonalds	Obesity Rate (%)
0	India	400	3.9
1	Japan	2975	4.3
2	South Korea	447	4.7
3	China	2700	6.2
4	Philippines	640	6.4
5	Thailand	240	0.1
6	Malaysia	282	10.0
7	Italy	598	19.9
8	Netherlands	249	20.4
9	France	1450	21.6
10	Taiwan	413	22.1
11	Brazil	1000	22.1
12	Germany	1480	22.3
13	Russia	645	23.1
14	Poland	445	23.1
15	Spain	509	23.8
16	United Kingdom	1274	27.8
17	South Africa	275	28.3
18	Mexico	402	28.9
19	Australia	920	29.0
20	Canada	1419	29.4
21	Hong Kong	237	29.9
22	Turkey	253	32.1
23	Saudi Arabia	304	35.4
24	US	14146	36.2

The following cell produces a scatterplot comparing the number of McDonalds resturants to the obesity rate in each country. Notice how there is not a strong visible relationship between the two. This would suggest that access to fast food and soda consumption are not good methods of measuring a country's obesity.

```
[46]: # Create scatter plot
McDonalds = McD['Number of McDonalds']
Obesity_2 = McD['Obesity Rate (%)']
plt.scatter(McDonalds,Obesity_2, color = 'r')

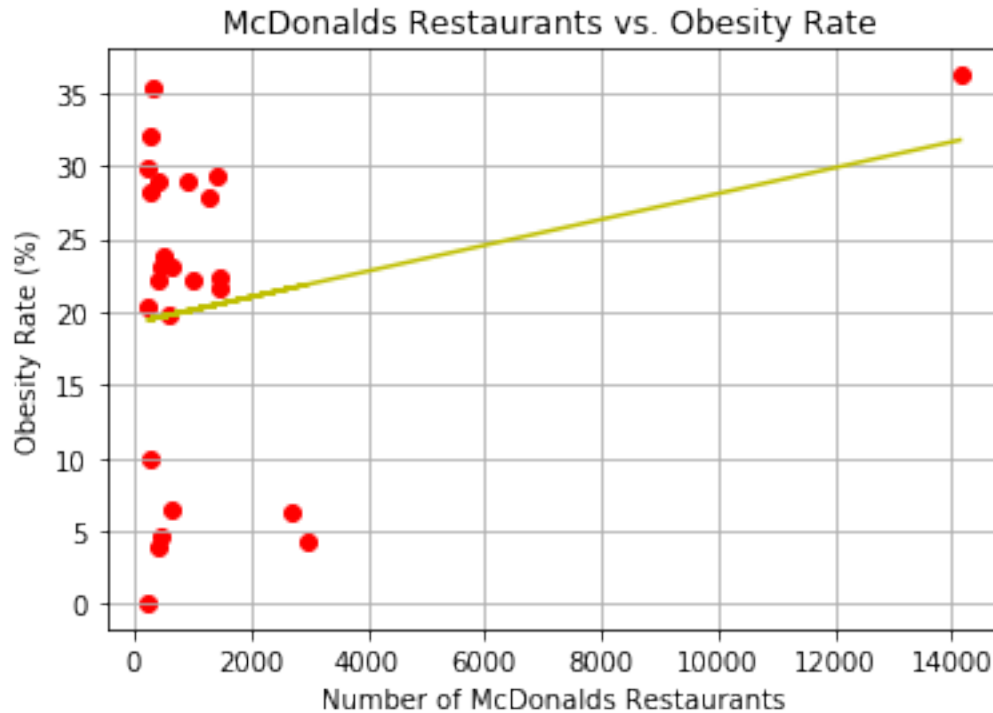
# Line of best fit
m, b = np.polyfit(McDonalds,Obesity_2,1)
plt.plot(McDonalds, m*McDonalds + b, color = 'y')

# Title
plt.title('McDonalds Restaurants vs. Obesity Rate ')

# X and Y axes
plt.xlabel('Number of McDonalds Restaurants')
plt.ylabel('Obesity Rate (%)')
```



```
# Plot
plt.grid()
plt.show()
```



We can quantify the correlational value between the number of McDonalds restaurants a country has and its obesity rate by running the cell below.

```
[47]: x_value = McDonalds
      y_value = Obesity_2

      correlation_matrix = np.corrcoef(x_value, y_value)
      correlation = correlation_matrix[0,1]
      r_squared = correlation**2

      print('The correlational coefficient between the number of McDonalds
      ↪restuarants in a country and the percentage of overweight adults is',
      ↪r_squared, 'this means that about', "%.2f" % (r_squared*100),
      ↪'% of the variation in the percentage of overweight adults can be 
      ↪explained by the numbers of McDonalds restuarants in a country.')
```

The correlational coefficient between the number of McDonalds restuarants in a country and the percentage of overweight adults is 0.051701811244402936 this means that about 5.17 % of the variation in the percentage of overweight

adults can be explained by the numbers of McDonalds restuarants in a country.

The final factor I chose to analyze was the percent of a country's population that is living in poverty. The cells below load a csv file that displays the 10 countries with the highest percent of people living in poverty and the obesity rates for each.

```
[48]: # Load data
pov = pd.read_csv('Poverty.csv')
```

```
[49]: # show data
pov
```

```
[49]:
```

	Country	Poverty Rate (%)	Obesity Rate (%)
0	South Africa	26.6	28.3
1	Costa Rica	20.9	25.7
2	Israel	17.9	26.1
3	United States	17.8	36.2
4	Korea	17.4	4.7
5	Turkey	17.2	32.1
6	Lithuania	16.9	26.3
7	Latvia	16.8	23.6
8	Mexico	16.6	28.9
9	Chile	16.5	28.0

Once again the following cell will create a scatter plot. However, this plot will visualize the relationship between a country's poverty rate and the obesity rate for each nation.

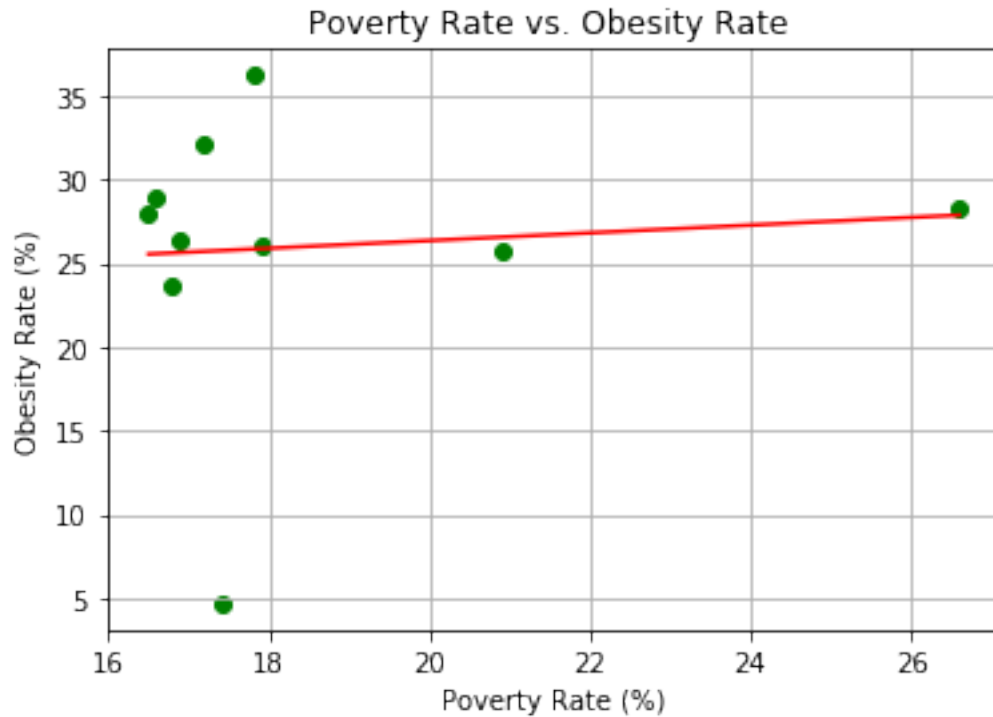
```
[50]: # Create scatter plot
Poverty = pov['Poverty Rate (%)']
Obesity_3 = pov['Obesity Rate (%)']
plt.scatter(Poverty,Obesity_3, color = 'g')

# Line of best fit
m, b = np.polyfit(Poverty,Obesity_3,1)
plt.plot(Poverty, m*Poverty + b, color = 'r')

# Title
plt.title('Poverty Rate vs. Obesity Rate ')

# X and Y axes
plt.xlabel('Poverty Rate (%)')
plt.ylabel('Obesity Rate (%)')

# Plot
plt.grid()
plt.show()
```



We will quantify the relationship between these two variables again following the same format as before.

```
[51]: x_val = Poverty
y_val = Obesity_3

correlation_matrix = np.corrcoef(x_val, y_val)
correlation = correlation_matrix[0,1]
r_squared = correlation**2

print('The correlational coefficient between the poverty rate in a country and_
→its obesity rate is', r_squared, 'which means that about', "%.2f" %_
→(r_squared*100),
      '% of the variation in the percentage of overweight adults can be _
→explained by the percent of the population living below the poverty line.')
```

The correlational coefficient between the poverty rate in a country and its obesity rate is 0.007553138174443815 which means that about 0.76 % of the variation in the percentage of overweight adults can be explained by the percent of the population living below the poverty line.

1.4 Discussion and Conclusion

The results of my project showed me that soda consumption, fast food access, and poverty are not individually strong indicators of increasing obesity. Although soda consumption account for nearly

20% of obesity rate increases in 2017, it could not account for countries that consume little soda but still have high obesity rates. I concluded that even though these factors may have a strong impact on obesity as a whole they cannot be used individually to determine if obesity rates are increasing. That is, an increase in one of these factors alone will not have a significant effect on a country's obesity rate.

When working on my project I found it hard to find data sources that fit each factor. I couldn't find how many liters of soda were consumed in every single country, for example, as not all countries will report that information. If I were to do this project again I would probably find other factors that all countries can report. Factors like how many grams of fat are consumed on a daily basis or how much sugar was consumed in a given year might have been more available for all countries.

1. I found that Argentina, the United States, and Chile consumed the most soda in 2017.
2. These 3 countries had obesity rates of 23.6%, 33.7%, and 27.8% respectively in 2017.
3. When correlating the data I determined that drinking soda accounted for 19.7% of the variation in the obesity rate for these countries in 2017. This means that soda consumption was responsible for 20% of increasing obesity rates in the countries that consumed the most soda.
4. From doing independent research I discovered that increasing demand for fast food and increasing poverty rates can increase obesity in a country, especially countries that have small populations.
5. In terms of my project, soda consumption accounted for 19.7% of the increases in obesity in 2017 while the number of McDonalds restaurants was only accountable for 5.17% and a country's poverty rate only accounted for 0.76%. Based on numbers alone I would say soda consumption had the most significant impact on increasing obesity rates.

1.4.1 References

*(List the source(s) for any data and/or literature cited in your project. Ideally, this should be formatted using a formal citation format (MLA or APA or other, your choice!). Multiple free online citation generators are available such as <http://www.easybib.com/style>. **Important:** if you use **any** code that you find on the internet for your project you **must** cite it or you risk losing most/all of the points for you project.)*

"Code Faster with Line-of-Code Completions, Cloudless Processing." Kite, kite.com/python/answers/how-to-plot-a-linear-regression-line-on-a-scatter-plot-in-python.

"COUNTRY COMPARISON :: OBESITY - ADULT PREVALENCE RATE." Central Intelligence Agency, Central Intelligence Agency, www.cia.gov/library/publications/the-world-factbook/rankorder/2228rank.html.

"List of Countries with McDonald's Restaurants." Wikipedia, Wikimedia Foundation, 5 Apr. 2020, en.wikipedia.org/wiki/List_of_countries_with_McDonald%27s_restaurants.

Most Obese Countries 2020, worldpopulationreview.com/countries/most-obese-countries/.

Sheth, Khushboo. "Countries With The Highest Levels Of Soft Drink Consumption." WorldAtlas, WorldAtlas, 25 Apr. 2016, www.worldatlas.com/articles/countries-with-the-highest-levels-of-soft-drink-consumption.html.

Singh, Maanvi. "Is Samoa's Obesity Epidemic A Harbinger For Other Developing Nations?" NPR, NPR, 7 Apr. 2016, www.npr.org/sections/thesalt/2016/04/07/473371279/is-samoas-obesity-epidemic-a-harbinger-for-other-developing-nations.

Written by renewbariatric Published Sep 23. "World Rankings: Obesity Rates by Country (July 2017)." Renew Bariatrics, 27 July 2019, renewbariatrics.com/obesity-rank-by-countries/.

[]: