CombiningLifeandWaterResults

September 7, 2020

1 Benefits of Clean Drinking Water for Longer Life Expectancy

In this project, I choose to look at the industry of public health. Specifically, I wanted to analyze the impact that clean drinking water has on life expectancy. I started by looking at the average life expectancy of people in the United States. I then looked at the raw data to see about other countries life expectancies. Nigeria, an African country, has a much lower life expectancy. As I thought about this more, I remembered things I had read the Gates Foundation and their work in helping developing counties. They speculated that some countries did not have the resources to properly dispose of thier waste, causing people to get sick from using the contaminated water to bathe and other chores. So, as the Gates Foundation analyzed the cause of water born diseases, I thought I could look at the effects of clean drinking water on life expectancy. I look at the trend in the life expectancy in Nigeria and saw that the average age for life expectancy has been increasing over the past 15 years as the percentage of people drinking clean water has also increased. I then compared those numbers to the percentage of people in the United States with clean drinking water and the United State's life expectancy. The United States averaged around 99% clean drinking water and a much higher life expectancy. I concluded that having access to clean drinking water could be a factor that increases life expectancy.

```
[1]: import pandas as pd
```

```
[2]: import numpy as np
```

```
[3]: import matplotlib.pyplot as plt
```

To begin this project, I first imported the raw data. The water_N and life_N is the data for the percentage of people who have access to clean water and the average life expectancy for people in Nigeria. I have formatted the data and did some preemtive analysis in the notebooks `LifeExpectancy_Nigeria LifeExpectancy_US PercentWithSafeWater_Nigeria` and `PercentWithSafeWater_US`. These notebooks can be found in the `src` or `notebooks` folder and the raw data can be found in the `data` folder.

Trimming the data required deleting all the years when no data was gathered and all the countries which I was not analyzing. It seemed that almost every country was on the list but I was only interested in the United States and Nigeria. Also, when I recieved the data, there were many years that no data was collected. These years ranged from 1960 to 2000. Since there was no data for the US and Nigeria during those times, I dropped those years.

```
[4]: water_N = pd.read_csv('~/Documents/EECS/EECS_731/HW/EECS731_1/data/water_N.csv')
```

```
[5]: life_N = pd.read_csv('~/Documents/EECS/EECS_731/HW/EECS731_1/data/life_N.csv')
```

I concattenated the data for safe water and life expectancy to make one table.

```
[6]: result_N = pd.concat([water_N,life_N], axis=1, sort=False)
```

```
[7]: result_N = result_N.drop(columns="Unnamed: 0")
```

```
[8]: result_N = result_N.drop(columns="Year")
```

```
[9]: result_N.insert(2,"Year",␣
     ↪[2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017],␣
     ↪True)
```

```
[10]: result_N = result_N.drop(columns="Country Name")
```

```
[11]: result_N.set_index('Year', inplace=True)
```

```
[12]: result_N = result_N.rename(columns={'Percent Safe Water':'Percent Drinking Safe␣
      ↪Water (N)'})
```

```
[13]: result_N = result_N.rename(columns={'Average Age':'Average Life Expectancy␣
      ↪(N)'})
```

1.0.1 Percentage of People Drinking Safe Water and the Average Life Expectancy per Year for Nigeria

```
[14]: result_N
```

```
[14]:       Percent Drinking Safe Water (N)  Average Life Expectancy (N)
      Year
      2001                      15.747336                       46.510
      2002                      16.064639                       46.835
      2003                      16.376387                       47.242
      2004                      16.682519                       47.720
      2005                      16.982870                       48.252
      2006                      17.277628                       48.812
      2007                      17.566862                       49.373
      2008                      17.850645                       49.913
      2009                      18.128638                       50.422
      2010                      18.401244                       50.896
      2011                      18.667351                       51.346
      2012                      18.927109                       51.786
      2013                      19.180478                       52.228
      2014                      19.427614                       52.672
      2015                      19.668675                       53.112
      2016                      19.903542                       53.541
```
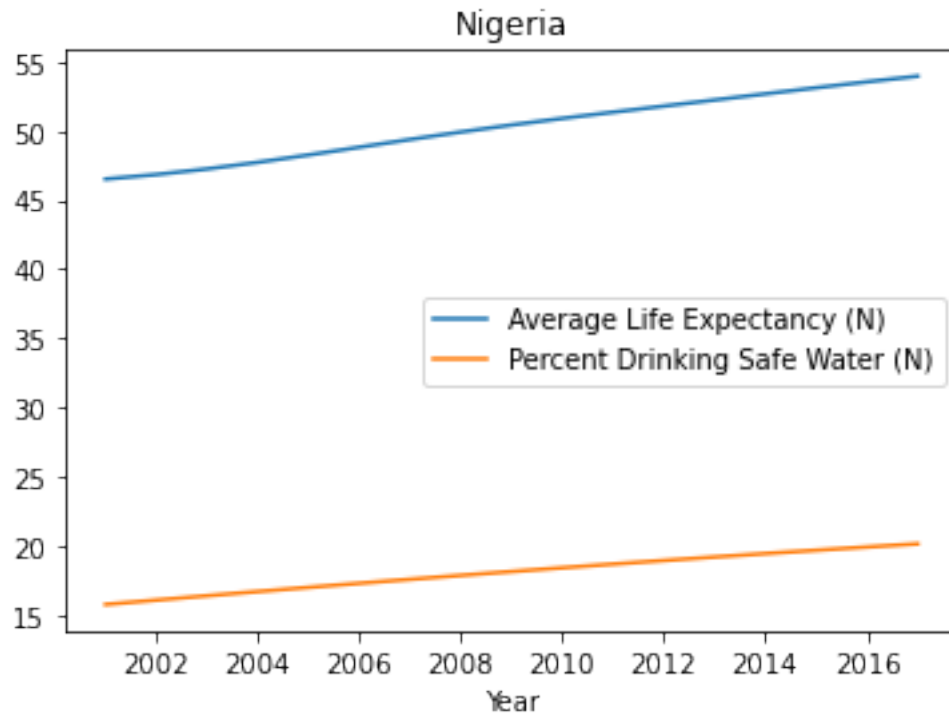
2017	20.132560	53.950

I then plotted the data to determine that as the percentage of people drinking clean water increases, the average life expectancy increases.

```python
[15]: result_N.plot(y=['Average Life Expectancy (N)', 'Percent Drinking Safe Water⏎
      ↪(N)'], title='Nigeria')
```

```
[15]: <matplotlib.axes._subplots.AxesSubplot at 0x11ea516d0>
```


I visualized that the average life expectany increases as the amount of people drinking safe water increases in Nigeria. This confirmed my inital thoughts. After looking at the trend in data for Nigeria, I decided to look at the trends for the United States.

```python
[16]: water_US = pd.read_csv('~/Documents/EECS/EECS_731/HW/EECS731_1/data/water_US.
      ↪csv')
```

```python
[17]: life_US = pd.read_csv('~/Documents/EECS/EECS_731/HW/EECS731_1/data/life_US.csv')
```

```python
[18]: life_US = life_US.drop(columns="Unnamed: 0")
```

```python
[19]: water_US = water_US.drop(columns="Unnamed: 0")
```

```python
[20]: water_US.set_index('Year', inplace=True)
```

```
[21]: life_US.set_index('Year', inplace=True)
```

I combined the data for life expectancy and percentage of people drinking clean water.

```
[22]: result_US = pd.concat([water_US,life_US], axis=1, sort=False)
```

```
[23]: result_US
```

[23]:
Year	Country Name	Percent Safe Water	Country Name	Average Age
2001	NaN	NaN	United States	76.836585
2002	NaN	NaN	United States	76.936585
2003	NaN	NaN	United States	77.036585
2004	NaN	NaN	United States	77.487805
2005	United States	98.942332	United States	77.487805
2006	United States	98.947103	United States	77.687805
2007	United States	98.951852	United States	77.987805
2008	United States	98.956569	United States	78.039024
2009	United States	98.961258	United States	78.390244
2010	United States	98.965887	United States	78.541463
2011	United States	98.970687	United States	78.641463
2012	United States	98.975572	United States	78.741463
2013	United States	98.980598	United States	78.741463
2014	United States	98.985731	United States	78.841463
2015	United States	98.990978	United States	78.690244
2016	United States	98.996310	United States	78.539024
2017	United States	99.001778	United States	78.539024

From this table, I observed that the United States has almost a 100% of the population drinking clean water.

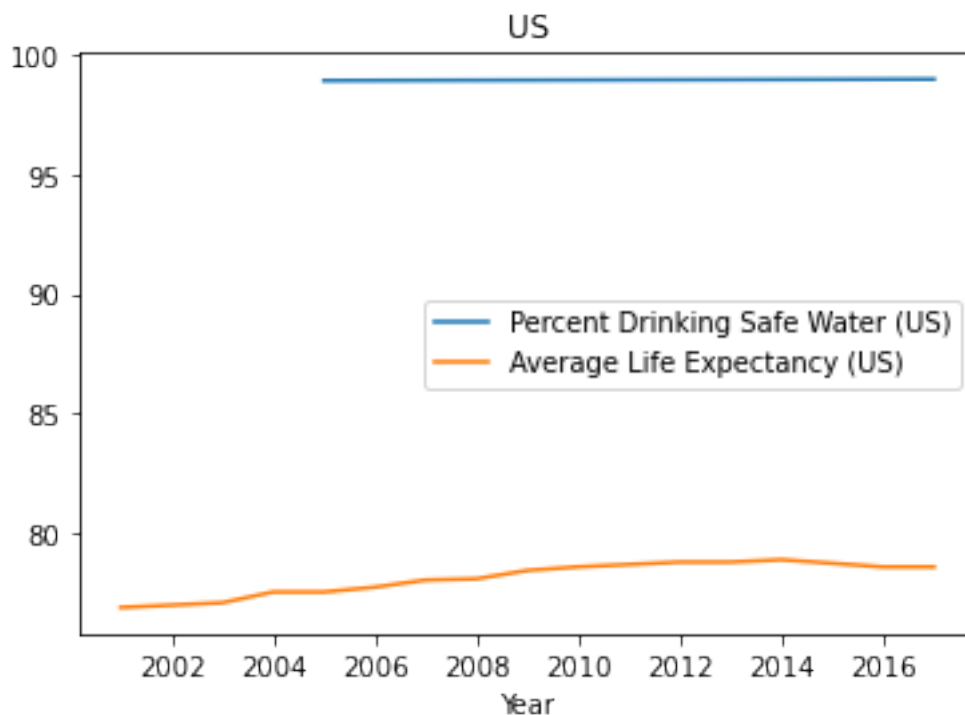```
[24]: result_US = result_US.drop(columns="Country Name")
```

```
[25]: result_US = result_US.rename(columns={'Percent Safe Water':'Percent Drinking␣
      ↪Safe Water (US)'})
```

```
[26]: result_US = result_US.rename(columns={'Average Age':'Average Life Expectancy␣
      ↪(US)'})
```

```
[27]: result_US.plot(y= ['Percent Drinking Safe Water (US)', 'Average Life Expectancy␣
      ↪(US)'], title='US')
```

```
[27]: <matplotlib.axes._subplots.AxesSubplot at 0x11eb54850>
```

For the United States, I observed no evident trend in life expectancy and clean drinking water. So, even though this graph does not show much, I think it is still useful to see that the US has almost 100% of people drinking safe water. Therefore, since the percent of people drinking safe water remains steady, it is difficult to infer what correlation is produced between the average life expectacy and the percentage of people drinking safe water.

```
[28]: result_all = pd.concat([result_US,result_N], axis=1, sort=False)
```

Next, I combined the data for Nigeria and the United States.

1.1 Average Life Expectancy and the Percentage of People Drinking Clean Water in the United States and Nigeria

```
[29]: result_all
```

```
[29]:       Percent Drinking Safe Water (US)  Average Life Expectancy (US)  \
      Year
      2001                            NaN                      76.836585
      2002                            NaN                      76.936585
      2003                            NaN                      77.036585
      2004                            NaN                      77.487805
      2005                      98.942332                      77.487805
      2006                      98.947103                      77.687805
      2007                      98.951852                      77.987805
```

2008	98.956569	78.039024
2009	98.961258	78.390244
2010	98.965887	78.541463
2011	98.970687	78.641463
2012	98.975572	78.741463
2013	98.980598	78.741463
2014	98.985731	78.841463
2015	98.990978	78.690244
2016	98.996310	78.539024
2017	99.001778	78.539024

	Percent Drinking Safe Water (N)	Average Life Expectancy (N)
Year		
2001	15.747336	46.510
2002	16.064639	46.835
2003	16.376387	47.242
2004	16.682519	47.720
2005	16.982870	48.252
2006	17.277628	48.812
2007	17.566862	49.373
2008	17.850645	49.913
2009	18.128638	50.422
2010	18.401244	50.896
2011	18.667351	51.346
2012	18.927109	51.786
2013	19.180478	52.228
2014	19.427614	52.672
2015	19.668675	53.112
2016	19.903542	53.541
2017	20.132560	53.950

I plotted the results for Nigeria and United States to see how the data compared between the counties.

1.2 Average Life Expectancy and the Percentage of People Drinking Clean Water in the United States and Nigeria Graphed

```
[30]: result_all.plot(y= ['Percent Drinking Safe Water (N)', 'Average Life Expectancy␣
      ↪(N)','Percent Drinking Safe Water (US)', 'Average Life Expectancy (US)'],␣
      ↪title='United States vs Nigeria')
```

```
[30]: <matplotlib.axes._subplots.AxesSubplot at 0x11ecd3e80>
```

United States vs Nigeria
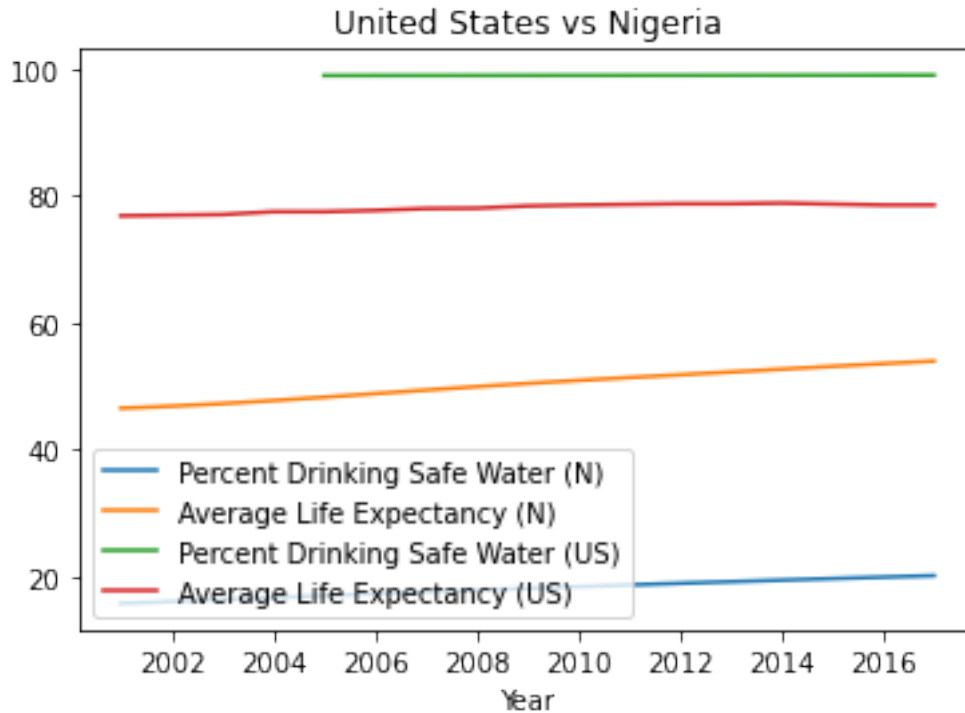
The y-axis is unlabeled. The values for percent drinking water range from 0% - 100% of the population. The values for average life expectancy are in years. Since the value of years ranges in between 0 and 100, it makes sense to graph all the data points together.

With that, I was able to see the percentage of people drinking safe water is for the people in Nigeria is much lower when compared to the United States. This is showcased in the difference between the blue and green lines. I then wanted to focus on the data from a specific year between the United States and Nigeria to better visualize the disparity. I choose the year 2017 as it was the most recent year I had data.

```
[31]: index = ['US', 'Nigeria',]
```
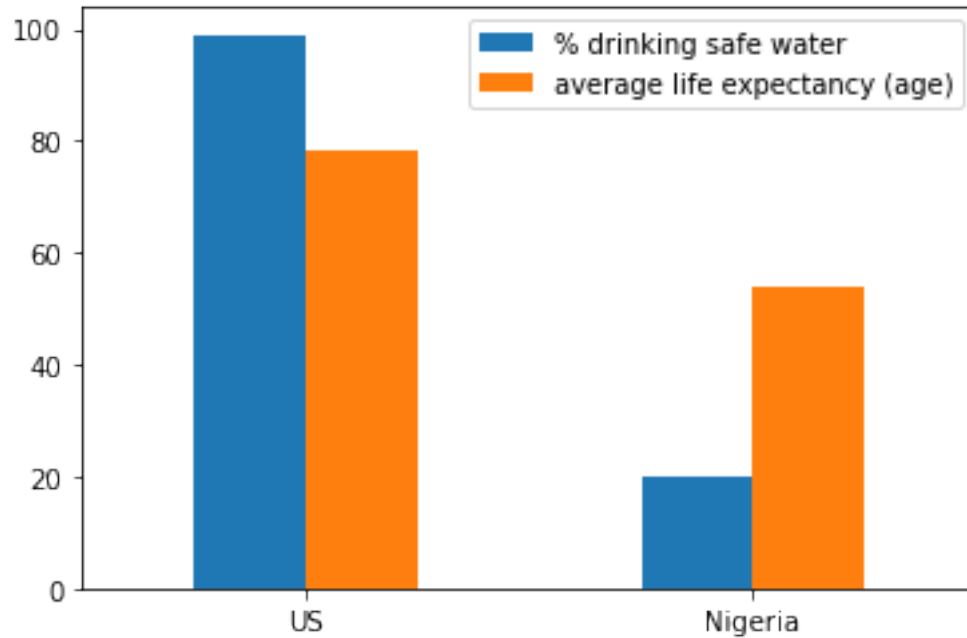
```
[32]: water = [99.001778, 20.132560]
```

```
[33]: age = [78.539024, 53.950]
```

```
[34]: df = pd.DataFrame({'% drinking safe water':water, 'average life expectancy␣
      ↪(age)' : age}, index = index)
```

1.3 Average Life Expectancy vs Percentage of People Drinking Clean Water in the United States and Nigeria in 2017

```
[35]: ax = df.plot.bar(rot=0)
```


With the blue bar, it is evident that almost 5 times the amount of people have access to clean drinking water in the United States when compared to Nigeria. It is also evident that the United States has a higher average life expectacy, by around 20 years. Therefore, I came to conclude that the aviability of safe drinking water may increase the average life expectancy.