Problem Set 2b - Pooja Amin

Worked with the whole class

Case & Deaton: Mortality and morbidity in the 21st century

Use the Compressed or Detailed Mortality tool on Wonder. According to this tool, how many people died in the United States in 2017? What was the (crude, not age-adjusted) death rate?

According to CDC Wonder, there were 2,813,503 deaths in 2017. The crude rate was 863.78 per 100,000.

Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2017 on CDC WONDER Online Database, released December, 2018. Data are from the Multiple Cause of Death Files, 1999-2017, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at http://wonder.cdc.gov/ucd-icd10.html (http://wonder.cdc.gov/ucd-icd10.html) on Feb 10, 2019 3:23:34 PM

Download the "public use mortality microdata" found on the NBER's website for the year 2017. Note that the website helpfully provides a PDF documentation file to help you understand the variables in the dataset. Replicate in a few lines of code the finding in part 1 regarding number of deaths. There is a category/type of death that is included in the Public Use mortality files, but excluded in the official mortality rates published by the CDC and on Wonder. What is it?

The NBER includes an extra category of foreign resident deaths that occured in the States, while the CDC does not.

Source: http://www.nber.org/data/vital-statistics-mortality-data-multiple-cause-of-death.html (http://www.nber.org/data/vital-statistics-mortality-data-multiple-cause-of-death.html)

Suppose you were a researcher who wanted to calculate the U.S. crude mortality rate using the number of deaths you just got from public use files. What other number do you need to calculate the rate? Where can you find it? Show that you can calculate the exact same number as the number that which was generated by the CDC Wonder tool.

In order to calculate the crude mortality rate, I need the total population in 2017. The US population, according to the World Bank, was 325,719,178 in 2017.

Source: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=US (https://data.worldbank.org/indicator/SP.POP.TOTL?locations=US)

In [1]:

```
import pandas as pd
import urllib.parse
```

In [2]:

```
df= pd.read_csv('http://www.nber.org/mortality/2017/mort2017.csv.zip', usecols
=['restatus'])
number_died = df[df.restatus <= 3].count()
```

In [3]:

```
number_died
```

Out[3]:

```
restatus    2813503
dtype: int64
```

In [4]:

```
USpop=325719178
#Source: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=US
```

In [5]:

```
calculated_cruderate=(number_died/USpop)*100000
calculated_cruderate
```

Out[5]:

```
restatus    863.781807
dtype: float64
```

The calculated crude rate using these numbers is 863.78 per 100,000.

Use the Compressed Mortality tool to create a similar (but not identical) graph. You should create a graph that depicts the same years, for the same age group, except instead of ending at 2013, include years from 1990 to 2017. Instead of including USW, USH, and the international comparison lines in they way they do (see the Figure footnote and supporting text), you will generate a graph containing three lines – US-White, US-Black or African American, and US-Other Race.

In [6]:

```
df_race1990=pd.read_csv('Compressed Mortality, 1979-1998 (4).txt', delimiter='
\t',usecols=['Year','Race','Deaths','Population','Crude Rate'])
df_race1990.drop(df_race1990.tail(25).index,inplace=True)
df_race1990.replace('Other Race','Other',inplace=True)
```

In [7]:

```
df_race1999=pd.read_csv('Compressed Mortality, 1999-2016 (4).txt', delimiter='
\t')
df_race1999.drop(df_race1999.tail(39).index,inplace=True)
```

In [8]:

```
df_race1999.replace('American Indian or Alaska Native','Other',inplace=True)
df_race1999.replace('Asian or Pacific Islander','Other',inplace=True)
mortby_race=df_race1990.append(df_race1999)
mortby_race=mortby_race.groupby(['Race','Year'],as_index=False).sum()
mortby_race
```

```
/anaconda3/lib/python3.7/site-packages/pandas/core/frame.py:6211:
FutureWarning: Sorting because non-concatenation axis is not align
ed. A future version
of pandas will change to not sort by default.

To accept the future behavior, pass 'sort=False'.

To retain the current behavior and silence the warning, pass 'sort
=True'.

  sort=sort)
```

Out[8]:

	Race	Year	Crude Rate	Deaths	Population	Year Code
0	Black or African American	1990.0	921.4	23871.0	2590791.0	0.0
1	Black or African American	1991.0	910.9	24364.0	2674576.0	0.0
2	Black or African American	1992.0	886.4	25005.0	2820889.0	0.0
3	Black or African American	1993.0	909.2	26918.0	2960565.0	0.0
4	Black or African American	1994.0	905.4	28271.0	3122545.0	0.0
5	Black or African American	1995.0	898.6	29619.0	3296228.0	0.0
6	Black or African American	1996.0	855.3	29771.0	3480929.0	0.0
7	Black or African American	1997.0	807.8	29614.0	3665910.0	0.0
8	Black or African American	1998.0	786.3	30265.0	3848854.0	0.0
9	Black or African American	1999.0	782.4	31759.0	4059306.0	1999.0
10	Black or African American	2000.0	786.9	33289.0	4230425.0	2000.0
11	Black or African American	2001.0	767.4	34518.0	4498003.0	2001.0
12	Black or African American	2002.0	761.4	35623.0	4678394.0	2002.0
13	Black or African American	2003.0	763.5	37006.0	4847081.0	2003.0
14	Black or African American	2004.0	733.2	36710.0	5006690.0	2004.0
15	Black or African American	2005.0	730.4	37749.0	5168047.0	2005.0
16	Black or African American	2006.0	704.6	37419.0	5310576.0	2006.0
17	Black or African American	2007.0	675.3	36818.0	5451695.0	2007.0
18	Black or African American	2008.0	644.3	35960.0	5581542.0	2008.0
19	Black or African American	2009.0	622.5	35447.0	5694521.0	2009.0
20	Black or African American	2010.0	591.5	34093.0	5763931.0	2010.0
21	Black or African American	2011.0	579.7	33531.0	5783915.0	2011.0
22	Black or African American	2012.0	571.8	32955.0	5763270.0	2012.0
23	Black or African American	2013.0	559.8	32111.0	5736357.0	2013.0
24	Black or African American	2014.0	557.2	31815.0	5709721.0	2014.0
25	Black or African American	2015.0	555.5	31638.0	5695622.0	2015.0
26	Black or African American	2016.0	564.7	32106.0	5685146.0	2016.0
27	Other	1990.0	288.7	2585.0	895366.0	0.0
28	Other	1991.0	273.9	2659.0	970871.0	0.0
29	Other	1992.0	274.1	2885.0	1052693.0	0.0
...	...	...	...	...	...	...
51	Other	2014.0	595.5	6605.0	3122581.0	4028.0
52	Other	2015.0	612.6	6852.0	3218283.0	4030.0
53	Other	2016.0	604.6	6943.0	3273843.0	4032.0
54	White	1990.0	427.2	92152.0	21570493.0	0.0
55	White	1991.0	422.3	93653.0	22174705.0	0.0

	Race	Year	Crude Rate	Deaths	Population	Year Code
56	White	1992.0	410.3	97140.0	23674136.0	0.0
57	White	1993.0	411.4	101870.0	24762562.0	0.0
58	White	1994.0	411.8	106227.0	25798029.0	0.0
59	White	1995.0	408.8	109784.0	26854334.0	0.0
60	White	1996.0	397.6	110879.0	27888249.0	0.0
61	White	1997.0	384.4	111404.0	28979261.0	0.0
62	White	1998.0	377.1	112156.0	29740654.0	0.0
63	White	1999.0	380.4	117011.0	30763465.0	1999.0
64	White	2000.0	388.3	122674.0	31590952.0	2000.0
65	White	2001.0	391.3	128699.0	32889387.0	2001.0
66	White	2002.0	396.8	131816.0	33218143.0	2002.0
67	White	2003.0	398.2	134512.0	33780855.0	2003.0
68	White	2004.0	395.2	135719.0	34342429.0	2004.0
69	White	2005.0	401.2	140224.0	34953243.0	2005.0
70	White	2006.0	399.9	141969.0	35504802.0	2006.0
71	White	2007.0	395.9	142233.0	35922086.0	2007.0
72	White	2008.0	399.8	144776.0	36209377.0	2008.0
73	White	2009.0	401.4	146108.0	36401990.0	2009.0
74	White	2010.0	392.9	143049.0	36406838.0	2010.0
75	White	2011.0	398.0	143426.0	36039826.0	2011.0
76	White	2012.0	394.2	140139.0	35554537.0	2012.0
77	White	2013.0	397.2	139086.0	35016167.0	2013.0
78	White	2014.0	397.1	137497.0	34626549.0	2014.0
79	White	2015.0	396.8	136004.0	34274256.0	2015.0
80	White	2016.0	397.5	134467.0	33827690.0	2016.0

81 rows × 6 columns

In [9]:

```
whitemort=mortby_race[mortby_race['Race']=='White']
blackmort=mortby_race[mortby_race['Race']=='Black or African American']
othermort=mortby_race[mortby_race['Race']=='Other']
othermort['Crude Rate']=(othermort['Deaths']/othermort['Population'])*100000
```

```
/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4: Se
ttingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pan
das-docs/stable/indexing.html#indexing-view-versus-copy
  after removing the cwd from sys.path.
```

In [10]:

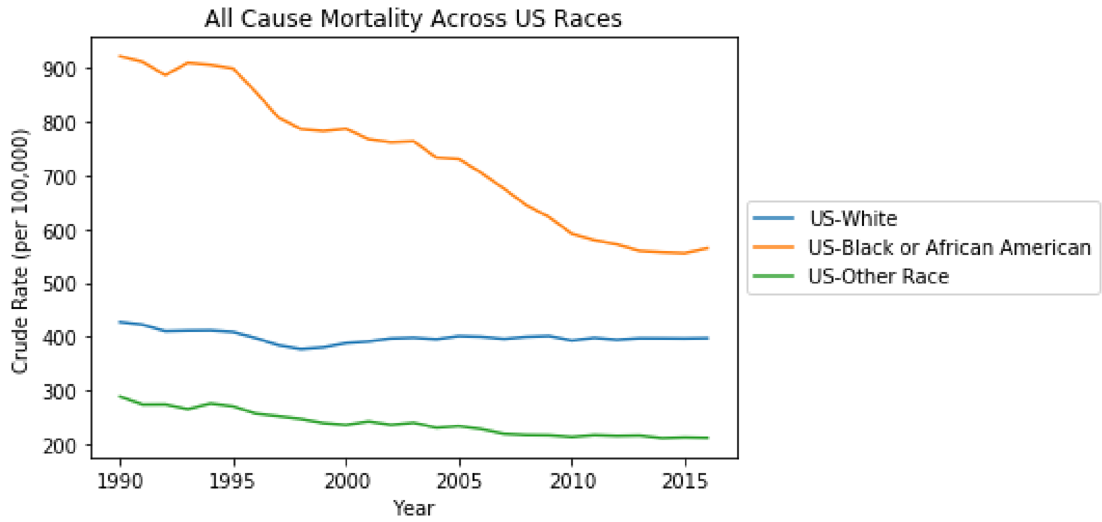
```
import matplotlib.pyplot as plt
import numpy as np
```

In [14]:

```
whiteplt=whitemort.plot('Year','Crude Rate')
blackplt=blackmort.plot('Year','Crude Rate',ax=whiteplt)
desiredplt=othermort.plot('Year','Crude Rate',ax=whiteplt)
plt.title('All Cause Mortality Across US Races')
plt.ylabel('Crude Rate (per 100,000)')
desiredplt.legend(['US-White','US-Black or African American','US-Other Race'],
loc='center left', bbox_to_anchor=(1, 0.5))
```

Out[14]:

```
<matplotlib.legend.Legend at 0x118946f28>
```



Why isn't it possible using the Compressed Mortality tool to exactly replicate Case & Deaton Fig. 1?

We do not have the data for the other countries in order to perfectly replicate the figure.

Examine a Public Use Microdata file and codebook between 1990 and 1998. Describe in detail, without actually doing, the procedure by which you could reconstruct their figure exactly for USW and USH.

Download the txt file. Go to the pdf file that describes the variables/codes used in the Microfile data. On page 33 of the 1991 pdf, there's infomation regarding Hispanic Origin and Race Re-code. Find the tape locations for those (80-82) and include the fields 2 and 1 in order to recode the races to take into account hispanic versus non-hispanic data.

Discuss your figure, and what you learn from it, and compare and contrast with Case and Deaton Figure 1. You may find it helpful to generate a second version of your graph with only two lines – US-White and US-Other, alongside your graph with all three lines. Focus in on the key points/arguments Case and Deaton are making with their version of this graph, and comment on their choice of comparison US racial/ethnic group. How do you think their paper and discussion would have shifted had they included the US-Black or African American line? Do you think that its inclusion undermines or bolsters their point?

Black mortality rates are higher than that of any other race, but are decreasing. Races other than White or Black are decreasing. White mortality rates are higher than that of other races, but the white mortality rates from CDC include White Hispanic mortality. Because the hispanic mortality rate is decreasing, it pulls the white mortality rate down from increasing as much in this graph. If Case & Deaton had included the US - Black or African American line, the increase in US - White would not be as clearly seen as the scale for the Black or African American rate would diminish the effects on crude rate over time. Omitting the black mortlity rate, allows for the scale of the crude death rate better show the effects, and emphasize the increase in white mortality rate over time by zooming in on the graph.

Case & Deaton (2017) Figure 7 depicts "Deaths of Despair for White Non-Hispanics with Less Than a Bachelor's Degree, by Birth Cohort." Use the Detailed Mortality tool to generate a graph like the bottom left, for suicides. What demographic option is the tool missing that you would need to fully replicate the Case & Deaton figure?

CDC Wonder is missing education data. In order to fully replicate the Case & Deaton figure, we need data for individuals with less than a bachelor's degree.

In [12]:

```
suicide=pd.read_csv('Underlying Cause of Death, 1999-2017 (1).txt', delimiter=
'\t',usecols=['Single-Year Ages','Year','Deaths','Population','Crude Rate'],sk
ipfooter=19)
suicide['Single-Year Ages']=suicide['Single-Year Ages'].str.replace('years',''
).astype(int)
suicide['Birth Cohort']=suicide['Year']-suicide['Single-Year Ages']
suicide=suicide.groupby(['Birth Cohort','Year'],as_index=False).sum()
BC1935 = suicide[suicide['Birth Cohort']==1935]
BC1940= suicide[suicide['Birth Cohort']==1940]
BC1945= suicide[suicide['Birth Cohort']==1945]
BC1950= suicide[suicide['Birth Cohort']==1950]
BC1955= suicide[suicide['Birth Cohort']==1955]
BC1960= suicide[suicide['Birth Cohort']==1960]
BC1965= suicide[suicide['Birth Cohort']==1965]
BC1970= suicide[suicide['Birth Cohort']==1970]
BC1975= suicide[suicide['Birth Cohort']==1975]
BC1980= suicide[suicide['Birth Cohort']==1980]
```

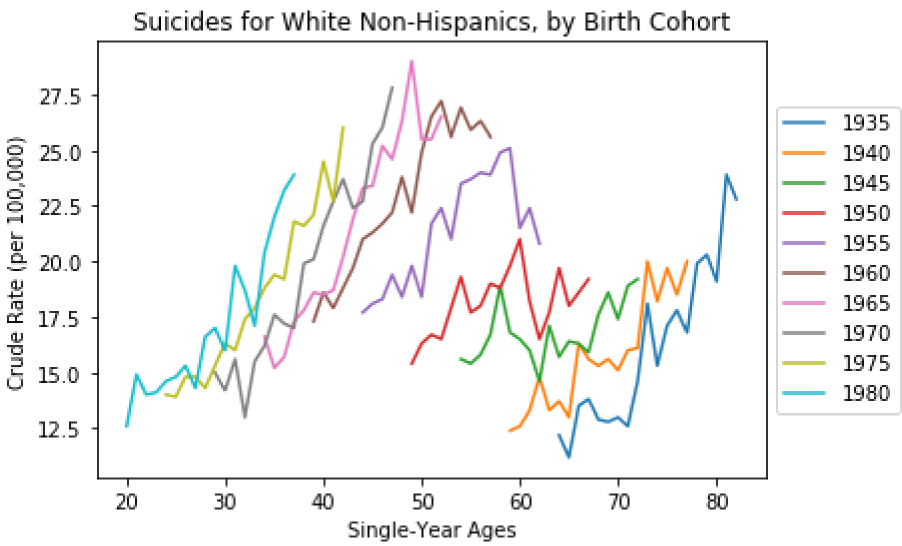
```
/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:1: Pa
rserWarning: Falling back to the 'python' engine because the 'c' e
ngine does not support skipfooter; you can avoid this warning by s
pecifying engine='python'.
  """Entry point for launching an IPython kernel.
```

In [15]:

```
plot1935=BC1935.plot('Single-Year Ages','Crude Rate')
plot1940=BC1940.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1945=BC1945.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1950=BC1950.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1955=BC1955.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1960=BC1960.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1965=BC1965.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1970=BC1970.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1975=BC1975.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1980=BC1980.plot('Single-Year Ages','Crude Rate',ax=plot1935)
plot1980.legend(['1935','1940','1945','1950','1955','1960','1965','1970','197
5','1980'],loc='center left', bbox_to_anchor=(1, 0.5))
plt.ylabel('Crude Rate (per 100,000)')
plt.title('Suicides for White Non-Hispanics, by Birth Cohort')
```

Out[15]:

```
Text(0.5, 1.0, 'Suicides for White Non-Hispanics, by Birth Cohor
t')
```



Name several things you learn from the graph. Be specific. Consider what we have discussed in class about the Case & Deaton thesis with regards to educational attainment, and speculate about what this graph might look like if we could split it into one graph for people with less than a Bachelor’s, and a second graph for people with a Bachelor’s degree or more.

For the Case and Deaton thesis, the rate at which mortality rises with age is higher in every successive birth cohort for individuals with less than a bachelor's degree. The rise in mortality by birth cohort is not simply a level shift but also a steepening of the age-mortality profiles, at least until the youngest cohorts. However, in this created graph without education taken into account, the rates at which mortality rises with age rises, then decreases, then rises a bit to approximately match the slope of the earlier birth cohorts. The level shift is not as consistent as it was with the data for individuals with less than a bachelor's degree. This graph has more of a parabolic shape regarding level shifts, and the between birth cohorts 1960 and 1980, is appears that each successive birth cohort is doing slightly better than the previous birth cohort. If the graph was split, the graph for people with less than a bachelor's would have steeper slopes and increasing level shifts with each birth cohort. The graph with people with a bachelor's degree or more would show each successive birth cohort doing better than the previous one.

In [ ]:

```

```