Ch08 NCHS Case Study

September 5, 2018

1 Embarak Ch08 Case Study --> NCHS Case Study

Prepared by:
Ossama Embarak

```
In [2]: import pandas as pd
        data = pd.read_csv("NCHS.csv")
        data.head(3)

Out[2]:    Year                                   113 Cause Name  \
        0  1999  Accidents (unintentional injuries) (V01-X59,Y8...
        1  1999  Accidents (unintentional injuries) (V01-X59,Y8...
        2  1999  Accidents (unintentional injuries) (V01-X59,Y8...

                     Cause Name    State  Deaths  Age-adjusted Death Rate
        0  Unintentional Injuries  Alabama  2313.0                     52.2
        1  Unintentional Injuries   Alaska   294.0                     55.9
        2  Unintentional Injuries  Arizona  2214.0                     44.8
```

See how many rows and how many columns

```
In [3]: data.shape    # 15028 rows and 6 columns

Out[3]: (15028, 6)
```

Remove all rows with na cases

```
In [4]: data = data.dropna()
        data.shape

Out[4]: (14917, 6)
```

What are the unique causes of death in this data set?

```
In [5]: data.head(2)
```

```
Out[5]:    Year                               113 Cause Name  \
        0  1999  Accidents (unintentional injuries) (V01-X59,Y8...
        1  1999  Accidents (unintentional injuries) (V01-X59,Y8...

                       Cause Name    State  Deaths  Age-adjusted Death Rate
        0  Unintentional Injuries  Alabama  2313.0                     52.2
        1  Unintentional Injuries   Alaska   294.0                     55.9
```

```
In [7]: causes =  data["Cause Name"].unique()
        causes
```

```
Out[7]: array(['Unintentional Injuries', 'All Causes', "Alzheimer's disease",
               'Homicide', 'Stroke', 'Chronic liver disease and cirrhosis',
               'CLRD', 'Diabetes', 'Diseases of Heart',
               'Essential hypertension and hypertensive renal disease',
               'Influenza and pneumonia', 'Cancer', 'Suicide', 'Kidney Disease',
               "Parkinson's disease", 'Pneumonitis due to solids and liquids',
               'Septicemia'], dtype=object)
```

Remove 'All Causes' from the Cause death Name column

```
In [8]: data = data[data["Cause Name"] !="All Causes"]
        causes =  data["Cause Name"].unique()
        causes
```

```
Out[8]: array(['Unintentional Injuries', "Alzheimer's disease", 'Homicide',
               'Stroke', 'Chronic liver disease and cirrhosis', 'CLRD',
               'Diabetes', 'Diseases of Heart',
               'Essential hypertension and hypertensive renal disease',
               'Influenza and pneumonia', 'Cancer', 'Suicide', 'Kidney Disease',
               "Parkinson's disease", 'Pneumonitis due to solids and liquids',
               'Septicemia'], dtype=object)
```

```
In [9]: len(causes)
```

```
Out[9]: 16
```

Find the unique causes of "State",

```
In [10]: data.head(3)
```

```
Out[10]:    Year                               113 Cause Name  \
         0  1999  Accidents (unintentional injuries) (V01-X59,Y8...
         1  1999  Accidents (unintentional injuries) (V01-X59,Y8...
         2  1999  Accidents (unintentional injuries) (V01-X59,Y8...

                        Cause Name    State  Deaths  Age-adjusted Death Rate
         0  Unintentional Injuries  Alabama  2313.0                     52.2
         1  Unintentional Injuries   Alaska   294.0                     55.9
         2  Unintentional Injuries  Arizona  2214.0                     44.8
```

```
In [11]: state = data["State"].unique()
         state

Out[11]: array(['Alabama', 'Alaska', 'Arizona', 'Arkansas', 'California',
                'Colorado', 'Connecticut', 'Delaware', 'District of Columbia',
                'Florida', 'Georgia', 'Hawaii', 'Idaho', 'Illinois', 'Indiana',
                'Iowa', 'Kansas', 'Kentucky', 'Louisiana', 'Maine', 'Maryland',
                'Massachusetts', 'Michigan', 'Minnesota', 'Mississippi',
                'Missouri', 'Montana', 'Nebraska', 'Nevada', 'New Hampshire',
                'New Jersey', 'New Mexico', 'New York', 'North Carolina',
                'North Dakota', 'Ohio', 'Oklahoma', 'Oregon', 'Pennsylvania',
                'Rhode Island', 'South Carolina', 'South Dakota', 'Tennessee',
                'Texas', 'United States', 'Utah', 'Vermont', 'Virginia',
                'Washington', 'West Virginia', 'Wisconsin', 'Wyoming'],
               dtype=object)
```

```
In [12]: data1 = data[data["State"] !="United States"]

         state = data1["State"].unique()
         state

Out[12]: array(['Alabama', 'Alaska', 'Arizona', 'Arkansas', 'California',
                'Colorado', 'Connecticut', 'Delaware', 'District of Columbia',
                'Florida', 'Georgia', 'Hawaii', 'Idaho', 'Illinois', 'Indiana',
                'Iowa', 'Kansas', 'Kentucky', 'Louisiana', 'Maine', 'Maryland',
                'Massachusetts', 'Michigan', 'Minnesota', 'Mississippi',
                'Missouri', 'Montana', 'Nebraska', 'Nevada', 'New Hampshire',
                'New Jersey', 'New Mexico', 'New York', 'North Carolina',
                'North Dakota', 'Ohio', 'Oklahoma', 'Oregon', 'Pennsylvania',
                'Rhode Island', 'South Carolina', 'South Dakota', 'Tennessee',
                'Texas', 'Utah', 'Vermont', 'Virginia', 'Washington',
                'West Virginia', 'Wisconsin', 'Wyoming'], dtype=object)
```

```
In [13]: len(state)

Out[13]: 51
```

1.0.1 What were the total number of deaths in the United States from 1999 to 2015?

```
In [14]: data.head(0)

Out[14]: Empty DataFrame
         Columns: [Year, 113 Cause Name, Cause Name, State, Deaths, Age-adjusted Death Rate]
         Index: []
```

```
In [15]: data["Deaths"].sum()

Out[15]: 69279057.0
```

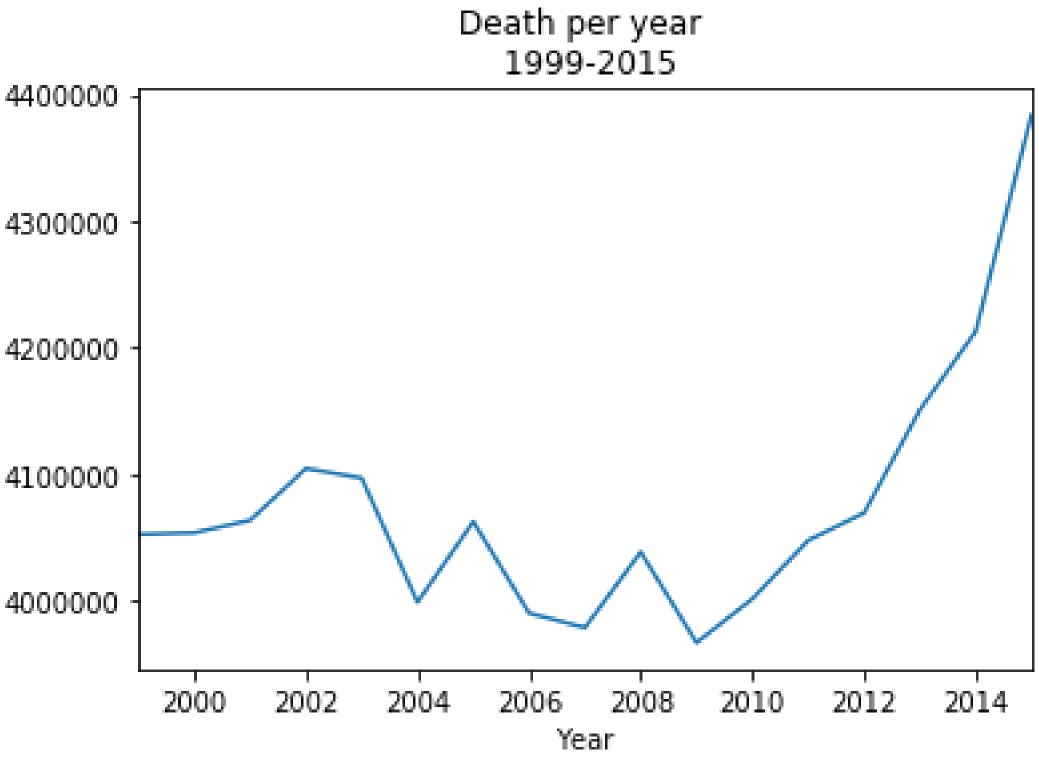
1.0.2 What is the trend of number of deaths per year?

```
In [16]: dyear= data.groupby(["Year"]).sum()
         dyear
```

```
Out[16]:          Deaths  Age-adjusted Death Rate
         Year
         1999  4052876.0                  38550.3
         2000  4054097.0                  38136.3
         2001  4063971.0                  37645.3
         2002  4104796.0                  37503.0
         2003  4097245.0                  36904.3
         2004  3999321.0                  35359.7
         2005  4062908.0                  35368.7
         2006  3990647.0                  34113.0
         2007  3979212.0                  33405.3
         2008  4038942.0                  33270.1
         2009  3967369.0                  32052.5
         2010  4001895.0                  31929.8
         2011  4048145.0                  31522.9
         2012  4069794.0                  30965.9
         2013  4151064.0                  30930.9
         2014  4213058.0                  30862.1
         2015  4383717.0                  31496.7
```

```
In [18]: dyear["Deaths"].plot(title="Death per year \n 1999-2015")
```

```
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6012d30208>
```



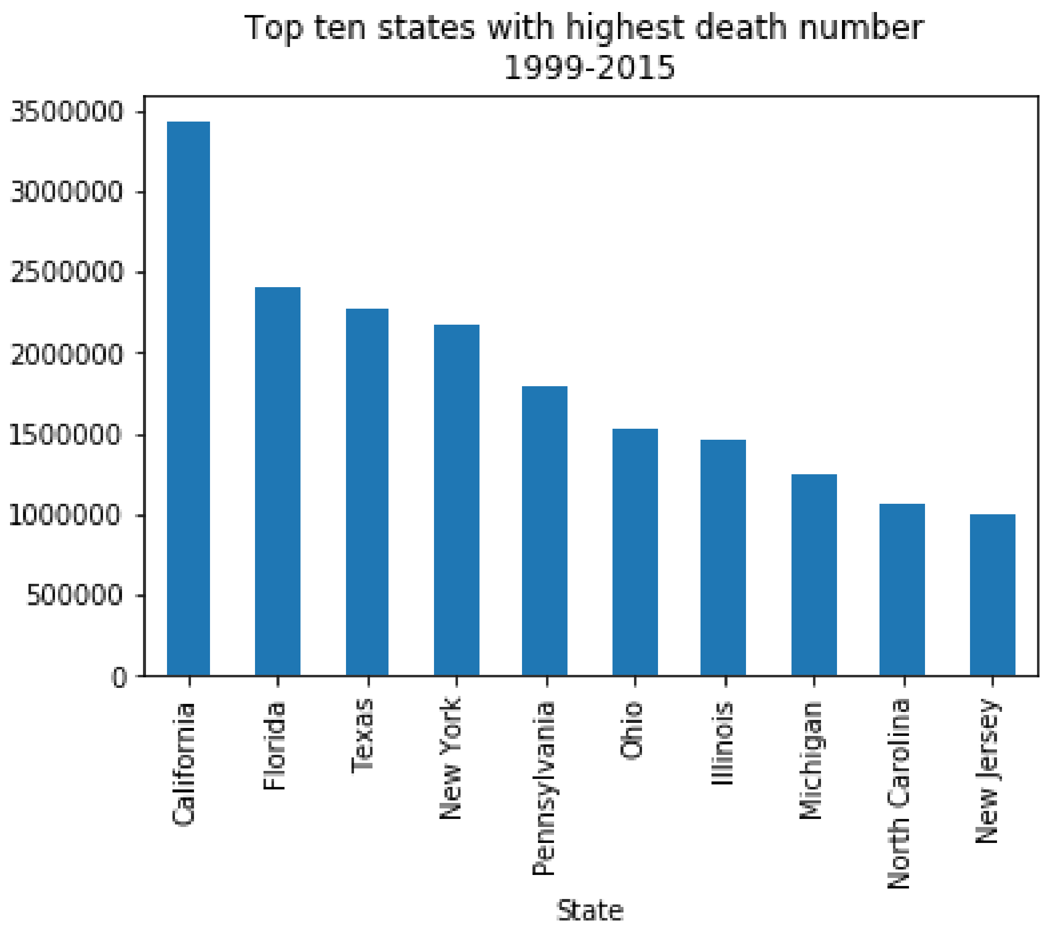
Which 10 states had the highest number of deaths in all years?

```
In [19]: data1 = data[data["State"] !="United States"]
         dataset2 = data1.groupby("State").sum()
         dataset2.sort_values("Deaths", ascending=False , inplace = True)
         dataset2.head(10)
```

```
Out[19]:                   Year     Deaths  Age-adjusted Death Rate
         State
         California      545904  3422459.0                  10101.2
         Florida         545904  2397507.0                  10156.8
         Texas           545904  2270961.0                  11339.7
         New York        545904  2170019.0                  10226.5
         Pennsylvania    545904  1785982.0                  11334.1
         Ohio            545904  1529552.0                  11931.3
         Illinois        545904  1460489.0                  11170.8
         Michigan        545904  1248155.0                  11645.7
         North Carolina  545904  1063835.0                  11737.3
         New Jersey      545904  1003709.0                  10446.7
```

```
In [20]: dataset2["Deaths"].head(10).plot.bar(title="Top ten states with highest death number \n
```

```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6012d30e48>
```



1.1 6. What were the top causes of deaths in the United States during this period?

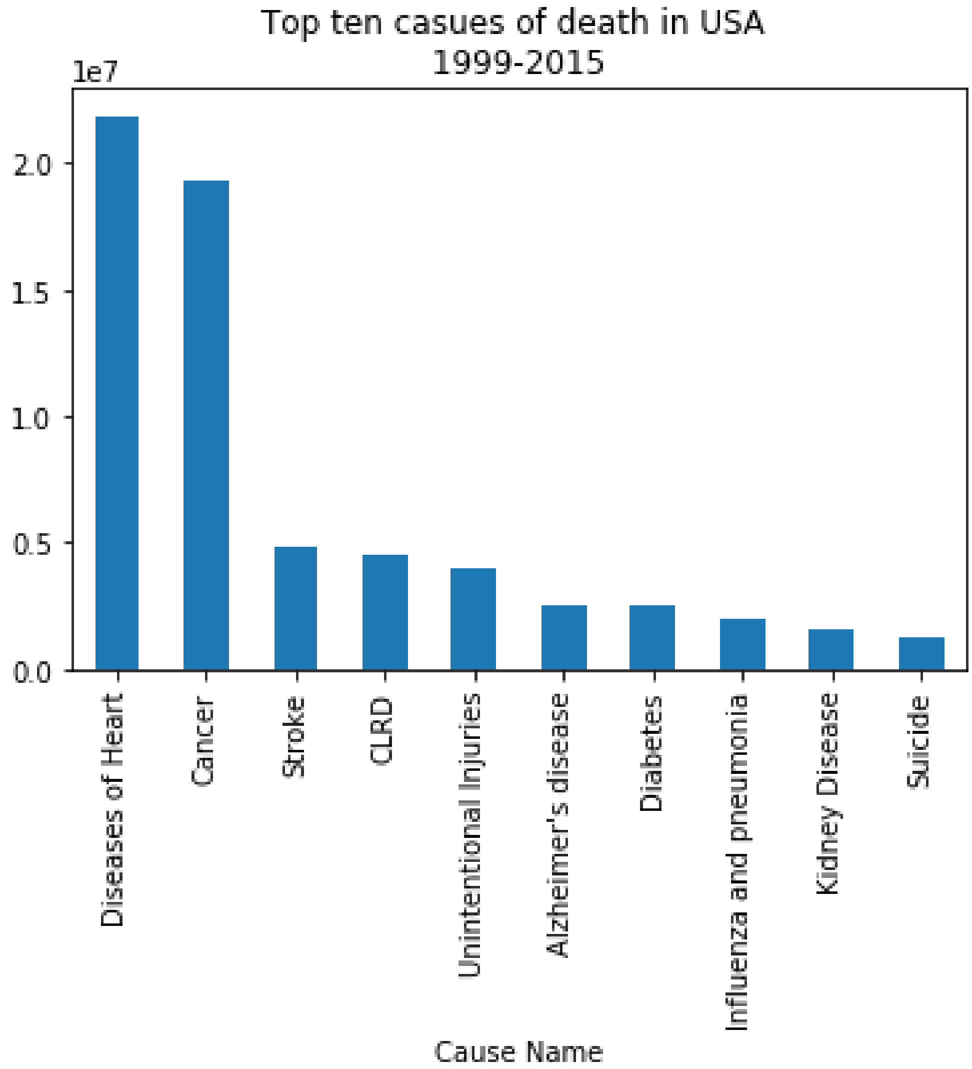
```
In [21]: dataset1 = data[data["Cause Name"] !="All Causes"]
         dataset2 = dataset1.groupby("Cause Name").sum()
         dataset2.sort_values("Deaths", ascending=False , inplace = True)
         dataset2.head(10)
```

```
Out[21]:                             Year      Deaths  Age-adjusted Death Rate
         Cause Name
         Diseases of Heart        1774188  21879846.0                 178315.3
         Cancer                   1774188  19292996.0                 160163.8
         Stroke                   1774188   4875996.0                  41458.8
         CLRD                     1774188   4560260.0                  39545.5
         Unintentional Injuries   1774188   4033020.0                  37368.6
         Alzheimer's disease      1774188   2514618.0                  21435.6
         Diabetes                 1774188   2472642.0                  20851.9
         Influenza and pneumonia  1774188   1974864.0                  16498.5
```

```
Kidney Disease           1774188   1515868.0                    12555.4
Suicide                  1774188   1209756.0                    11580.1
```

```
In [22]: dataset2["Deaths"].head(10).plot.bar(title="Top ten casues of death in USA \n 1999-2015
```

```
Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x7f60129c1cf8>
```



Analyze guns deaths in the US

```
In [3]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        sns.set(style='white', color_codes=True)
        %matplotlib inline
```

```
In [2]: dataset = pd.read_csv('Death data.csv', index_col=0)
        print(dataset.shape)
        dataset.index.name = 'Index'
        dataset.columns = map(str.capitalize, dataset.columns)
        dataset.head(5)
```

```
(100798, 10)
```

```
Out[2]:        Year  Month   Intent  Police Sex   Age                    Race  \
        Index
        1      2012      1  Suicide       0   M  34.0  Asian/Pacific Islander
        2      2012      1  Suicide       0   F  21.0                   White
        3      2012      1  Suicide       0   M  60.0                   White
        4      2012      2  Suicide       0   M  64.0                   White
        5      2012      2  Suicide       0   M  31.0                   White

               Hispanic            Place     Education
        Index
        1           100             Home           BA+
        2           100           Street  Some college
        3           100  Other specified           BA+
        4           100             Home           BA+
        5           100  Other specified        HS/GED
```

```
In [5]: # Organizing the data by the year, then by month:
        dataset_Gun = dataset
        dataset_Gun.sort_values(['Year', 'Month'], inplace=True)
```

Annual U.S. suicide gun deaths 2012-2014, by gender

```
In [6]: dataset_Gun.Sex.value_counts(normalize=False)
```

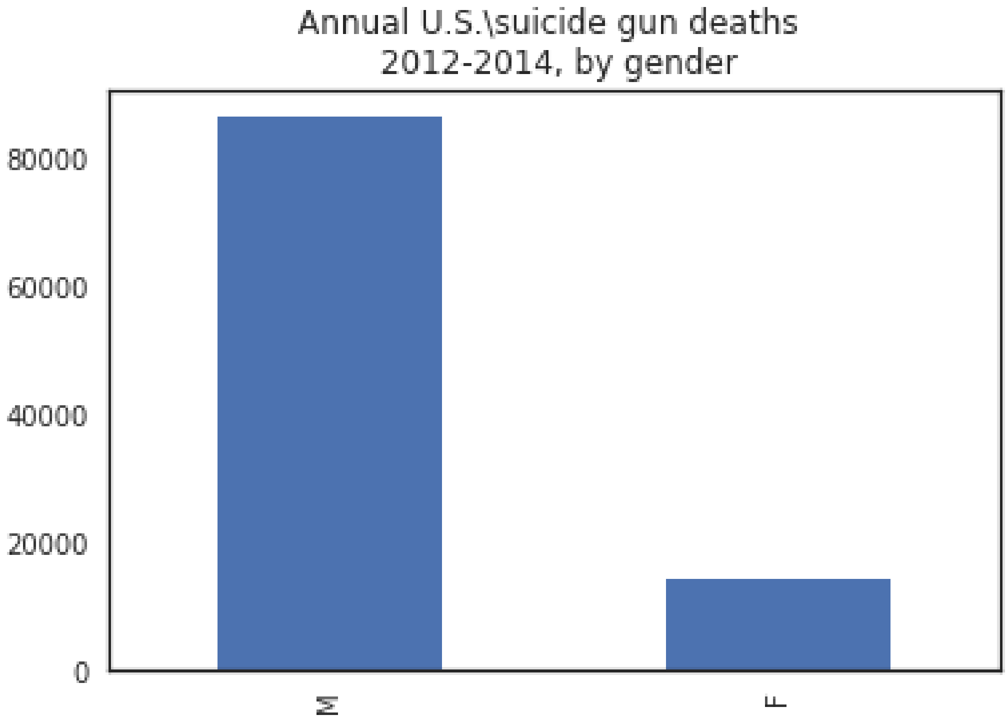
```
Out[6]: M    86349
        F    14449
        Name: Sex, dtype: int64
```

```
In [8]: dataset_byGender = dataset_Gun.groupby('Sex').count()
        dataset_byGender
```

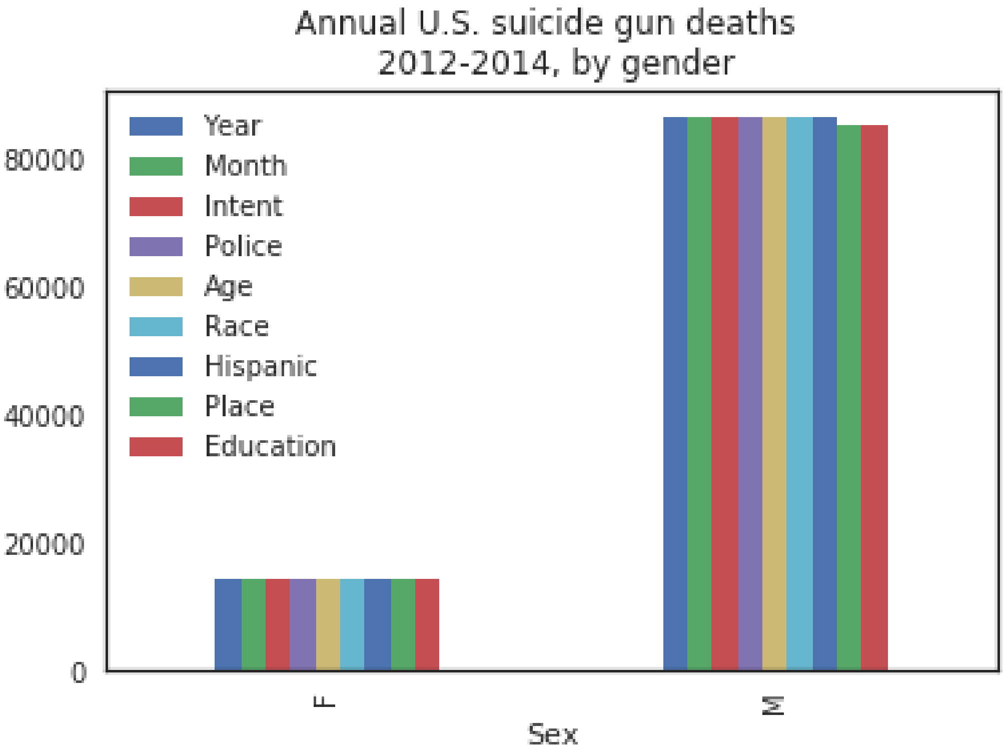
```
Out[8]:       Year  Month  Intent  Police    Age   Race  Hispanic  Place  Education
        Sex
        F    14449  14449   14449   14449  14446  14449     14449  14386      14243
        M    86349  86349   86348   86349  86334  86349     86349  85028      85133
```

```
In [29]: dataset_Gun.Sex.value_counts(normalize=False).plot.bar(title='Annual U.S.\\
         suicide gun deaths \n 2012-2014, by gender')
```

```
Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6010b7d278>
```

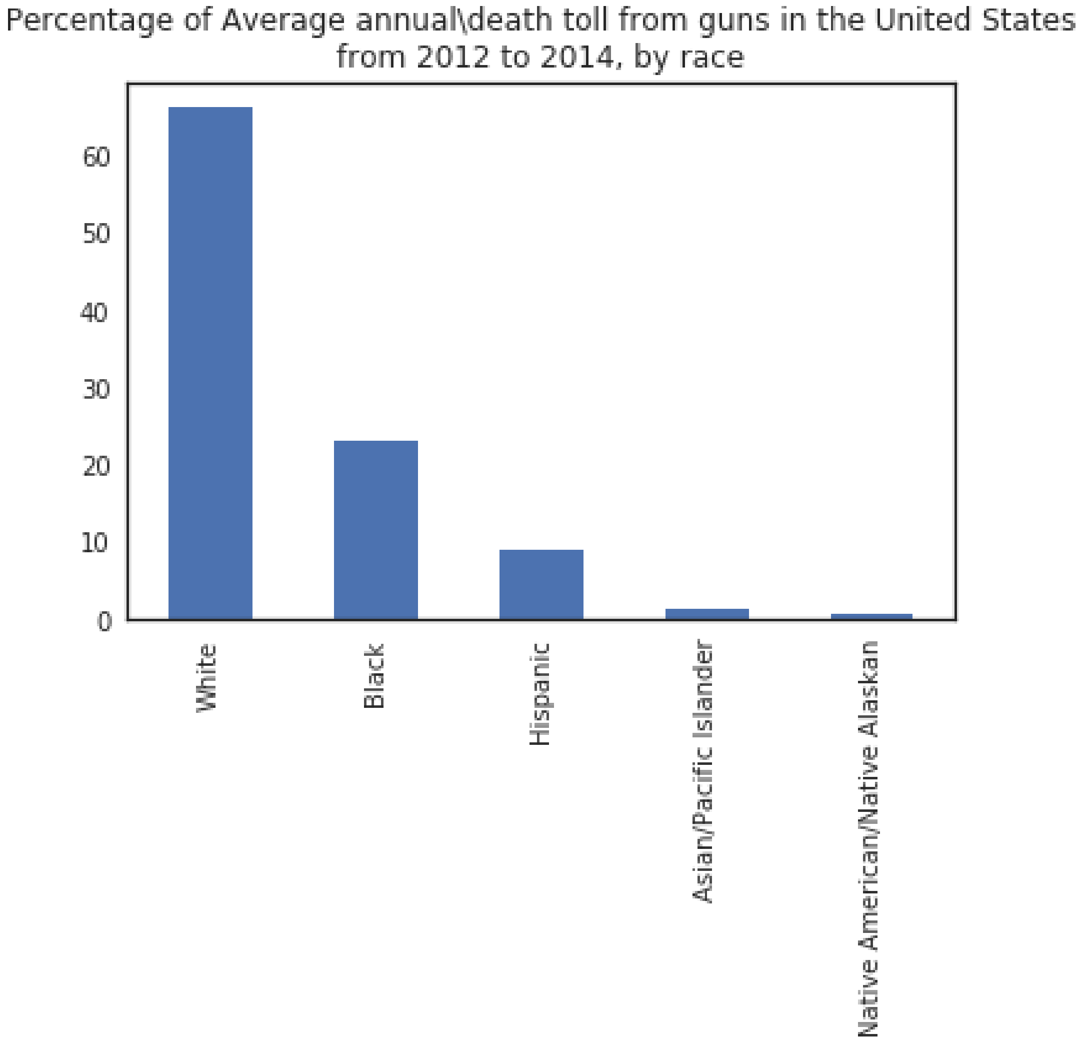


```
In [30]: dataset_byGender = dataset_Gun.groupby(['Sex']).count()
         dataset_byGender

Out[30]:       Year  Month  Intent  Police    Age   Race  Hispanic  Place  Education
         Sex
         F    14449  14449   14449   14449  14446  14449     14449  14386      14243
         M    86349  86349   86348   86349  86334  86349     86349  85028      85133

In [31]: dataset_byGender.plot.bar(title='Annual U.S. suicide gun deaths \n 2012-2014, by gender

Out[31]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6013d76710>
```

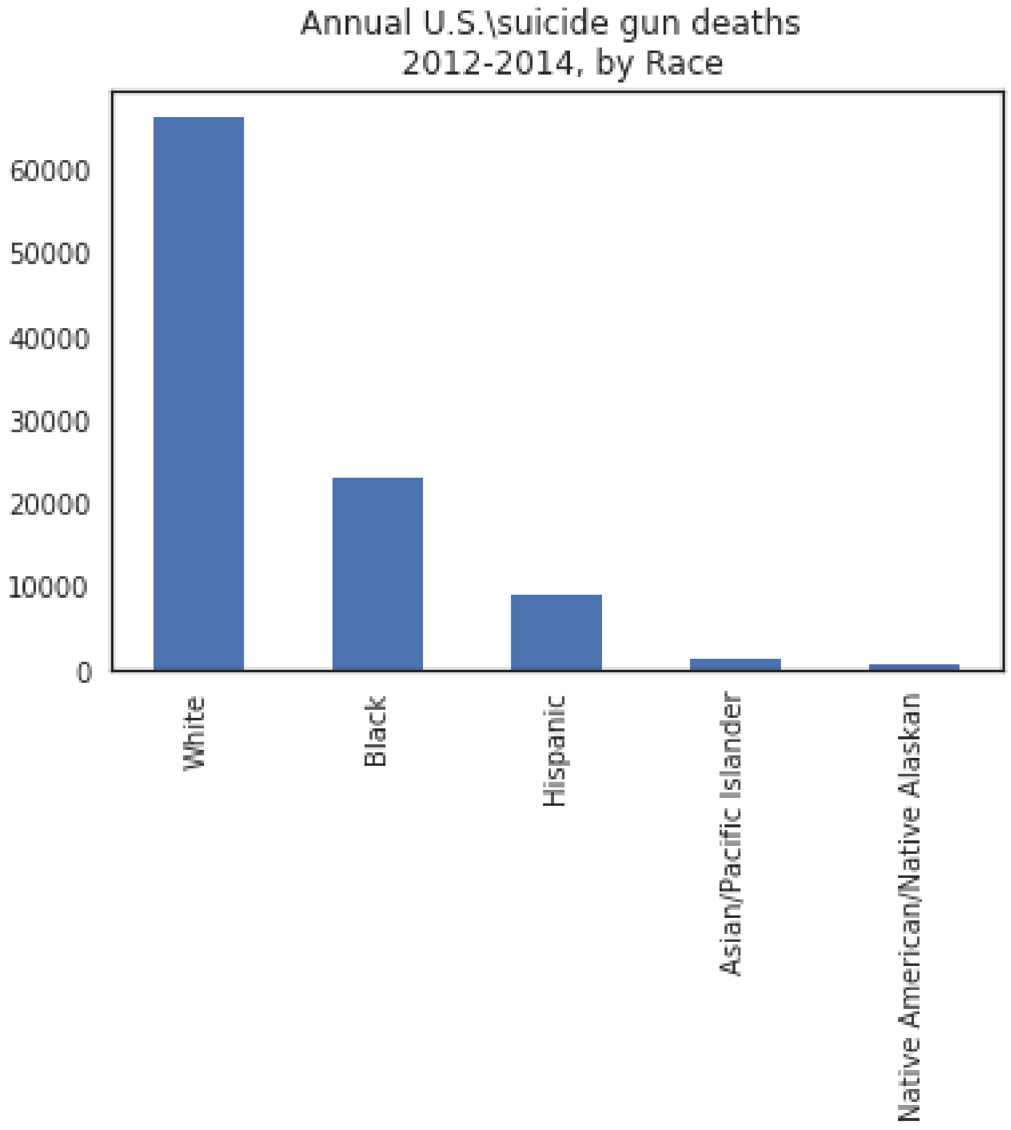


1.1.1 Average annual death toll from guns in the United States from 2012 to 2014, by race

```
In [12]: dataset_byRace = dataset
         (dataset_byRace.Race.value_counts(ascending=False) *100/100000)

Out[12]: White                              66.237
         Black                              23.296
         Hispanic                            9.022
         Asian/Pacific Islander              1.326
         Native American/Native Alaskan      0.917
         Name: Race, dtype: float64

In [13]: (dataset_byRace.Race.value_counts(ascending=False) *100/100000).plot.bar(title=' Percen
         death toll from guns in the United States \nfrom 2012 to 2014, by race')

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7f1c197f4ac8>
```



```
In [34]: dataset_byRace.Race.value_counts(normalize=False)
         dataset_byRace.Race.value_counts(normalize=False).plot.bar(title='Annual U.S.\\
         suicide gun deaths \n 2012-2014, by Race')

Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6010b1e278>
```



3. Rate of gun deaths in the U.S. per 100,000 population 2012-2014, by race.

```
In [35]: dataset_byRace = dataset
         print (dataset_byRace.shape)
         dataset_byRace.head(2)
```

```
(100798, 10)
```

```
Out[35]:        Year  Month   Intent  Police Sex   Age                    Race  \
         Index
         1      2012      1  Suicide       0   M  34.0  Asian/Pacific Islander
```

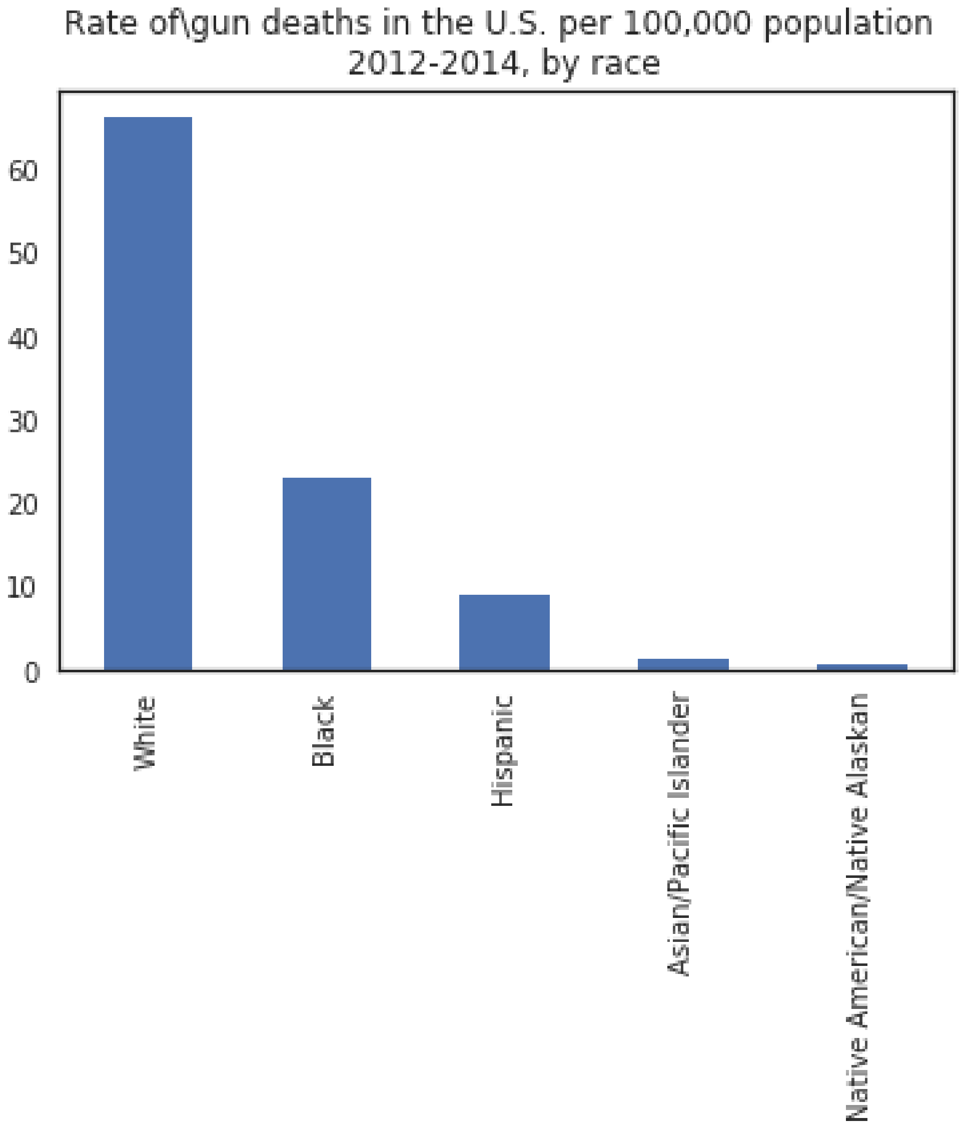
```
2      2012     1  Suicide      0   F  21.0                  White

       Hispanic   Place    Education
Index
1           100    Home          BA+
2           100  Street Some college
```

```
In [36]: dataset_byRace = dataset
         (dataset_byRace.Race.value_counts(ascending=False) *100/100000)
```

```
Out[36]: White                             66.237
         Black                             23.296
         Hispanic                           9.022
         Asian/Pacific Islander             1.326
         Native American/Native Alaskan     0.917
         Name: Race, dtype: float64
```

```
In [37]: (dataset_byRace.Race.value_counts(ascending=False) *100/100000).plot.bar(title='Rate of
         gun deaths in the U.S. per 100,000 population \n2012-2014, by race')
```

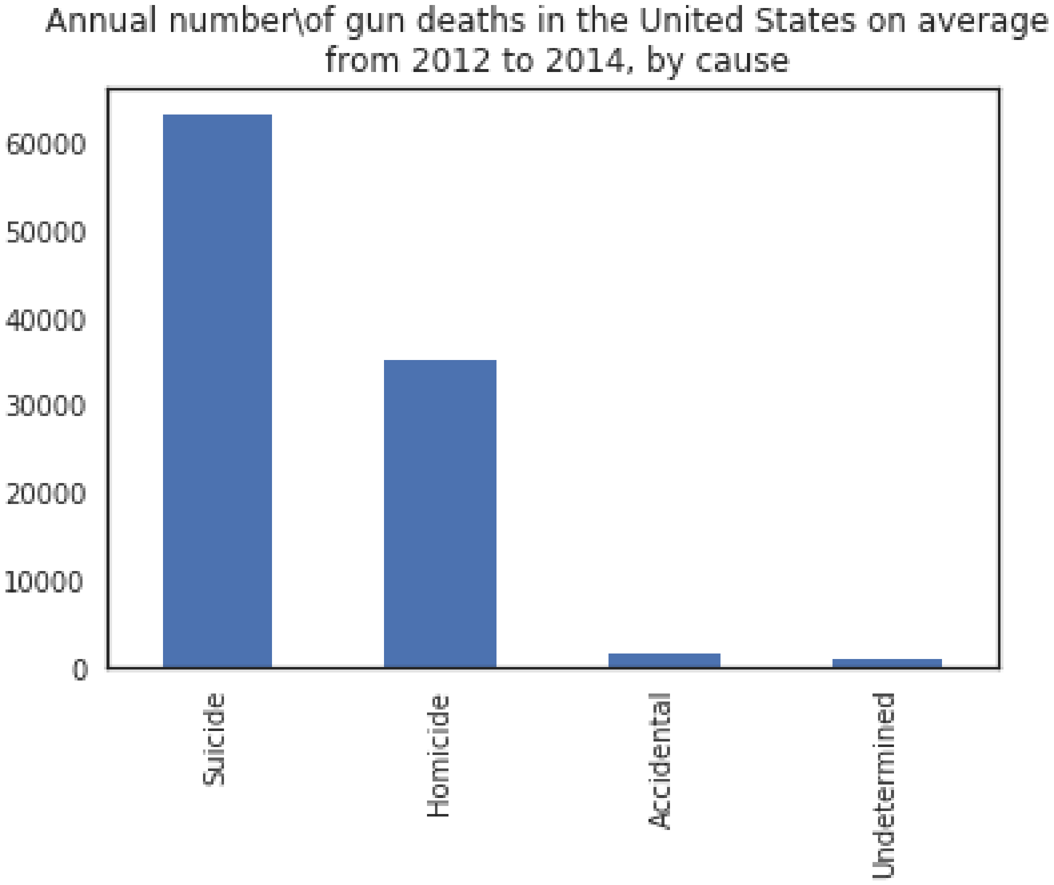
```
Out[37]: <matplotlib.axes._subplots.AxesSubplot at 0x7f60107baeb8>
```



4. Annual number of gun deaths in the United States on average from 2012 to 2014, by cause

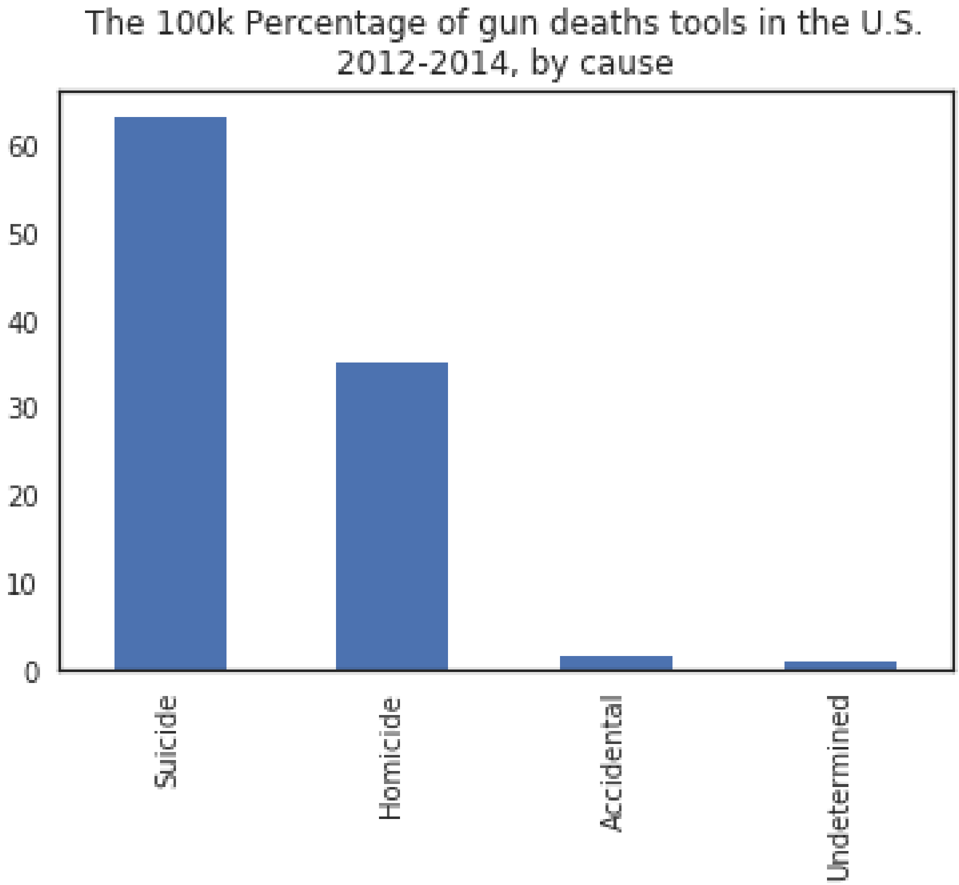
```
In [18]: dataset_byRace.Intent.value_counts(sort =True , ascending=False)

Out[18]: Suicide         63175
         Homicide        35176
         Accidental       1639
         Undetermined      807
         Name: Intent, dtype: int64

In [17]: dataset_byRace.Intent.value_counts(sort=True).plot.bar(title='Annual number\\
         of gun deaths in the United States on average \n from 2012 to 2014, by cause')
```

```
Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x7f1c19aba860>
```



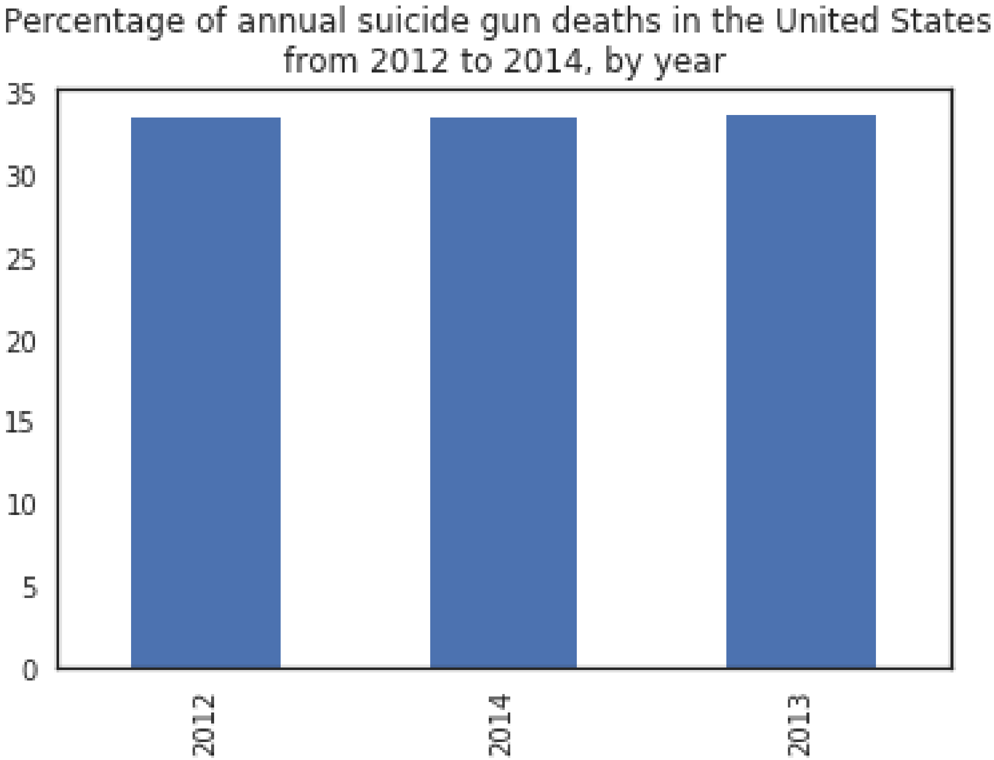
5. Average annual death toll from guns in the United States from 2012 to 2014, by cause

```
In [40]: dataset_byRace.Intent.value_counts(ascending=False) *100/100000

Out[40]: Suicide         63.175
         Homicide        35.176
         Accidental       1.639
         Undetermined     0.807
         Name: Intent, dtype: float64

In [21]: (dataset_byRace.Intent.value_counts(ascending=False) *100/100000).plot.bar(title='The 1

Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x7f1c19738160>
```



6. Percentage of annual suicide gun deaths in the United States from 2012 to 2014, by year

```
In [42]: dataset_byRace.Year.value_counts(ascending=True) *100/100000

Out[42]: 2012    33.563
         2014    33.599
         2013    33.636
         Name: Year, dtype: float64

In [22]: (dataset_byRace.Year.value_counts(ascending=True) *100/100000).plot.bar(title='Percenta

Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x7f1c18dde828>
```


Percentage of annual suicide gun deaths in the United States from 2012 to 2014, by year
35
30
25
20
15
10
5
0
2012
2014
2013