Computer lab 2: Sprint 0
Reading in an Excel spreadsheet and plotting data
16-Dec-2014

Create a folder and place inside the spreadsheet 'World firearms murders and ownership.xlsx'

Step 1)
The following root code loads the data into a *Dataframe* object (called “df”). Make sure your Python script is located in the same folder as the spreadsheet

```
import pandas as pd
df = pd.read_csv("worldFirearmsMurdersAndOwnership.csv")
```

...to look at what you just loaded in, simply type `df` and the command prompt

Step 2)
Next, let's make some initial scatter-plots of the data.
In this example, we specify two columns by name, copy them to an x and y variable, and plot them. We'll also overlay the country data for each data point

```
xColumn = 'Average firearms per 100 people'
yColumn = 'Homicide by firearm rate per 100,000 pop'
x = df[xColumn]
y = df[yColumn]
import matplotlib.pyplot as plt
fig1 = plt.figure()
ax1 = plt.subplot()
plt.scatter(x,y,)
ax1.set_ylabel(yColumn)
ax1.set_xlabel(xColumn)
plt.grid(True)
import numpy as np
for ii in range(0,len(df)):
     cond1 = np.isnan(y.iloc[ii])
     cond2 = np.isnan(x.iloc[ii])
     if ~cond1 and ~cond2:
           plt.text(x.iloc[ii],y.iloc[ii],
           df['Country/Territory'].iloc[ii],
             fontsize=8)
plt.show()
```

Step 2a)

Go ahead and explore other variables, two at a time, that MIGHT be of interest. save relevant plots into your working directory. As a specific example, here are two derived variables of possible interest:

```
lethality = df['Number of homicides by firearm'] / df['Average total
all civilian firearms']
population = 100 * df['Average total all civilian firearms'] /
df['Average firearms per 100 people']
x = population + 1
y = lethality + 1
xColumn = 'population'
yColumn = 'Number of homicides by firearm'
fig2 = plt.figure()
ax2 = plt.subplot()
plt.scatter(x,y,)
ax2.set_ylabel(yColumn)
ax2.set_xlabel(xColumn)
plt.grid(True)
for ii in range(0,len(df)):
   cond1 = np.isnan(y.iloc[ii])
   cond2 = np.isnan(x.iloc[ii])
   if ~cond1 and ~cond2:
      plt.text(x.iloc[ii],y.iloc[ii],
      df['Country/Territory'].iloc[ii],fontsize=8)
plt.show()
```

Step 3)

Next: we're going to establish two *canonical* variables of interest for use in the rest of this computer exercise. For clarity, we'll also specify a logarithmic scaling of axes (to be covered in Chapter 4 of the text). Here's the modified plotting code::

```
population = 100 * df['Average total all civilian firearms'] /
df['Average firearms per 100 people']
x = population
xColumn = 'population'
yColumn = 'Number of homicides by firearm'
y = df[yColumn] + 1
fig3 = plt.figure()
ax3 = plt.subplot()
plt.scatter(x,y,)
ax3.set_ylabel(yColumn)
ax3.set_xlabel(xColumn)
ax3.set_yscale('log')
ax3.set_xscale('log')
plt.grid(True)
for ii in range(0,len(df)):
   cond1 = np.isnan(y.iloc[ii])
   cond2 = np.isnan(x.iloc[ii])
   if ~cond1 and ~cond2:
      plt.text(x.iloc[ii],y.iloc[ii],
      df['Country/Territory'].iloc[ii],
      fontsize=8)
plt.show()
```

Save a copy of your source code and figures by e-mailing to *jschuler@yeshivaschools.com*