Data Bootcamp: Code Practice #3
Revised: January 18, 2016

Answer each of the questions below. Answers should include any code you wrote to find your answer. They can be handwritten, but they must be readable and look professional.

Run this code to produce the dataframe `weo`:

```
import pandas as pd
data = {'BRA': [13.37, 13.30, 14.34, 15.07, 15.461625, 15.98, 16.10],
        'JPN': [33.43, 31.83, 33.71, 34.29, 35.60, 36.79, 37.39],
        'USA': [48.30, 46.91, 48.31, 49.72, 51.41, 52.94, 54.60],
        'Year': [2008, 2009, 2010, 2011, 2012, 2013, 2014]}
weo  = pd.DataFrame(data)
```

The numbers are GDP per person in thousands of US dollars, 2008 to 2014, variable PPPPC in the World Economic Outlook database. The code is at the bottom of this program; you can cut and paste it from there to save time and effort.

1. *Review.* Something about: objects and methods. Help. Dictionaries. Types.

2. In the code at the top, explain these components:
   (a) The `import` statement.
   (b) The data dictionary.
   (c) The `pd.` in the last line.

3. Describe the structure of `weo`. What are its dimensions? Its column labels? Its row labels?

4. What "type" is `weo['BRA']`? What does this mean in non-technical language?

5. Create a new variable equal to the ratio of Brazilian to US GDP per capita. What is its value in 2014?

6. What method would you use to "export" the dataframe as an Excel spreadsheet? Verify that it works by opening the spreadsheet you create.

7. Describe the result of the statement `t = weo.tail(3)`. What kind of object is `t`? What does it look like?

8. How would you create a new dataframe that consists of the first 4 rows of `weo`?

9. Set the "index" to the `Year` variable. Does this change the dimensions of the dataframe?

10. Change the variable names from country codes to country names: Brazil, Japan, and the United States.

11. What method would you use to compute the mean for each country? What are the means?

12. *Challenging.* How would you compute the means for each date across countries?

13. Read the first 10 rows of the 538 college majors data. What are the dimensions of the resulting dataframe?

14. Create a dataframe of the first ten rows of the 538 college majors data that contains only the variables numbered [2, 15, 16, 17]. (Remind yourself that numbering starts at zero.) What are the names of these variables? *Bonus points:* What do the last two variables represent?

15. *Challenging.* The data come sorted by median income, the variable `Median`. How does the ranking change if we sort by `P75th`? What comes in second place? What do we learn about this major?

16. Read from your computer...