Homework 6

1. At this point, we've seen a slew of different data structures in Python and R. It's worth taking a moment to sit down and lay them out in a manner that hopefully reveals their similarities and differences.

   Construct a visual represention of the properties and relationships between the data structures in both languages. This could take the form of a mind map or a flowchart, or whatever other representation that you think best conveys the picture. You may want to start with a draft to try out your layout before spending time on the final version. You can use any technology you like: pencil and paper, powerpoint, R (there is a useful package called `diagrammer`), etc., but include a picture of the final version in your Rmd/pdf document.

   Include the following data structures:

   Python: list, tuple, dictionary, numpy array, pandas series, pandas dataframe.

   R: atomic vectors, matrix/array, list, data frame, tibble.

2. In lecture we demonstrated how you can build a pandas dataframe from a dictionary, but looked at fairly well-behaved examples. Consider the following dataframe and describe what rules pandas appears to rely upon to turn an unruly dictionary into a dataframe. Some suggestions for things to try:

   - Query the types of the columns of a dataframe using the `.dtypes` attribute.
   - Like an array, you can check shape with `.shape`.
   - Try pulling the resulting dictionary apart with subsetting and checking `type()`.

```
import pandas as pd
df = pd.DataFrame({'A': [[1, 2, 3], 'foo', [], [3, 4]], 'B': 1})
print(df)
```

```
##            A  B
## 0  [1, 2, 3]  1
## 1        foo  1
## 2         []  1
## 3     [3, 4]  1
```

3. In addition to building dataframes from dictionaries, you can also build them from arrays. Convert the following arrays from homework 5 into dataframes and use them to provide examples as you answer the following questions.

```
import numpy as np
A = np.array([1])
B = np.array([1, 2, 3])
C = np.array([[4],
              [5],
              [6]])
D = np.array([[4, 1],
              [5, 0],
              [6, 1]])
```

- What happens when you convert an array into a dataframe? How can you add row and column names? (try looking through `dir()` as well as consulting the textbook/internet)
- What happens when you add an array to a dataframe that share the same shape?
- What happens when you add two dataframes that do not share the same shape? Consider both cases: different number of rows and different number of columns.

4. The last problem of the previous homework had you standardize random normally distribution data using z-scores. Repeat that exercise, but work with `X` as a dataframe with variables/columns called X1, X2, X3, and X4.