PUI 2015 HW 3. Due W 9/30 at 5:00PM. Automatic scripts will fork your repositories at deadline time, no excuses or delays accepted.

Reading:

Data Jujitsu (DJ Patil)
http://www.oreilly.com/data/free/data-jujitsu.csp
Sections:
- Introduction
- No data vomit

Assignment 1 : Distributions

Following the ipython notebook
https://github.com/fedhere/PUI2015_fbianco/blob/master/HW3/assignment1_distributions_instructions.ipynb

1. GENERATE 100 samples of different sizes N (N>10 & N<2000) from each of 6 different distributions (600 samples in total), all with the same *population* mean.
Include a Normal, a Poisson, a Binomial, a Chi-Squared distribution, and 2 more of your choice.

2. For each sample plot the sample mean against the sample size N (if you want you can do it with the sample standard deviation as well).
Describe the behavior you see in the plots - do they look as you expected? why?

3. PLOT the distributions of all sample means (together for all distributions).
Mandatory - plot is as a histogram,
Optional - plot it in any other way you think is convincing

Optional: FIT a gaussian to the distribution of means
e.g. how to fit function to data in numpy:
http://glowingpython.blogspot.com/2012/07/distribution-fitting-with-scipy.html
http://stackoverflow.com/questions/7805552/fitting-a-histogram-with-python
Delivery: create a new Github directory called HW3 inside of your PUI2015_<your name> repo. upload an ipython notebook, with the rendered plots.
Include a README.md which describes what you are doing, and, if appropriate, how to run the notebook (input variables? global variables that need to be setup?).
75% of the grade will be based on the rendered version of the plot, 25% will be awarded if the TA can download and run the notebook (if you include any package that was not in the standard Anaconda distribution state that in your README.md, so that the TAs can install them).
Assignment 2 : **Hypothesis testing**

Follow and understand the notebook
https://github.com/fedhere/PUI2015_fbianco/blob/master/HW3/effectivenes%20of%20NYC%20Post-Prison%20Employment%20Programs.ipynb
and repeat the z-test and chisq test for other data included in the paper we are working with.
by fill in missing cells in (your own copy of) the notebook

Delivery: upload the notebook on your PUI2015_<your name>/HW3 github repo (see Assignment 1)

Assignment 3: From idea to scientific result
NOTE: ONLY THIS PART deadline Sunday 10/4 11:59:59PM to facilitate working in groups

All assignments can be done in groups, but i *strongly recommand* you do this one in group (5 ppl max) - brainstorming ideas works best if there are more then 1 brains!

Work on CitiBikes data https://www.citibikenyc.com/system-data
Come up with an idea of something that can be tested with a proportion, or mean problem, on the citibike data.
I prepared an example here:
https://github.com/fedhere/PUI2015_fbianco/blob/master/citibikes/citibikes_1950s.ipynb

1. Describe your idea
2. Turn the idea into a *testable hypothesis* and state your *Null and alternative hypotheses*
3. choose a confidence level
4. mangle your data as needed
5. choose a statistical test. Use e.g. z-score, or chi sq, but also other tests are available. Follow Statistics in a Nutshell to choose the appropriate test depending on your problem and on your data.
6. assess whether you can reject the Null Hypothesis

Delivery: create a new Github directory called citibikes inside of your PUI2015_<your name> repo. I want you to do that because we will probably build on this exercise and work more tiwht the citibike data incrementally.
Deliver the exercise described above as a rendered notebook (75% grade on rendered version 25% on executable as for Assignments 1 and 2)
Include a README.md in your repo that describes the project and states your specific contribution, instructions for the TA to run the notebook etc....

NOTE: the ipython notebooks can be identical if you work in the same group but the README.md must be individual, and it must state your specific contribution. be honest, and learn as much as ou can from the homework!