Snake Eyes

Seeing your data with Python

To Do

- Noel Hidalgo, BetaNYC
- Final project
- Class meetup
- Note on previous homework
- Last module's homework
- Moving to Python
- This module's homework

Noel Hidalgo

- Executive Director, betaNYC http://betanyc.us/

Final Project

- Posted on Blackboard page
- Create a public visualization
- Use data relevant to a current policy, business, or justice issue
    - Find data
    - Get sign-off on project
    - Clean/transform data
    - Create visualization
    - Write about its importance
    - Get it up on our site

Final Project

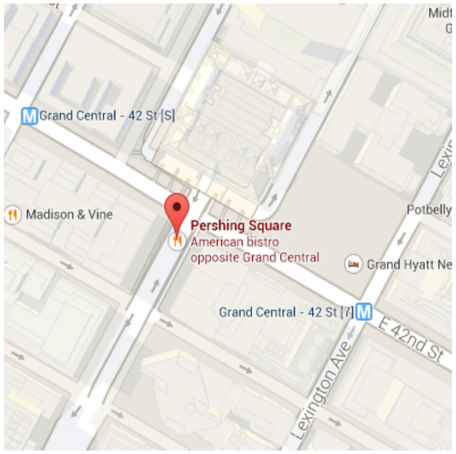
- Consider this a portfolio piece
- Will stay up either as long as I can keep it or as long as you want
- Min 1 month public
- Proposal due 4/10
- Final project due 5/15

Suggested Data Sources

- UN http://data.un.org/
- World Bank http://datacatalog.worldbank.org/
- NYC open data https://nycopendata.socrata.com/
- NYS open data https://data.ny.gov/
- DataKind data (discussing next week)
- Anywhere else (just run it by me)

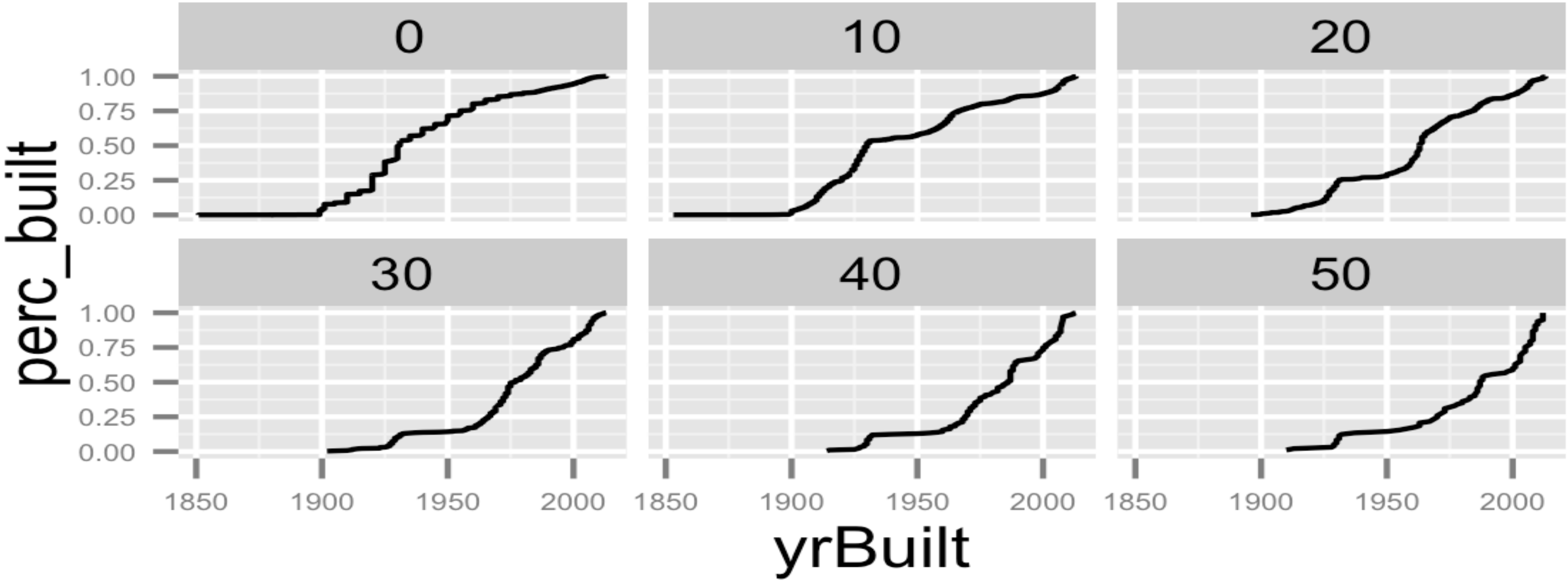
Class Meetup

- Thanks to those who responded
- We are meeting `Monday, March 24 at 6PM`
- Coffee/tea/other refreshments at Pershing Square http://www.pershingsquare.com/
- Right by Grand Central Terminal
- Come if you can
- I'm around for coffee otherwise



Note on Previous Homework

- There was a building collapse yesterday
- info at http://news.yahoo.com/nyc-building-collapse-140122865.html
- 5 story buildings built in 1910



8/20

Last Module's Homework

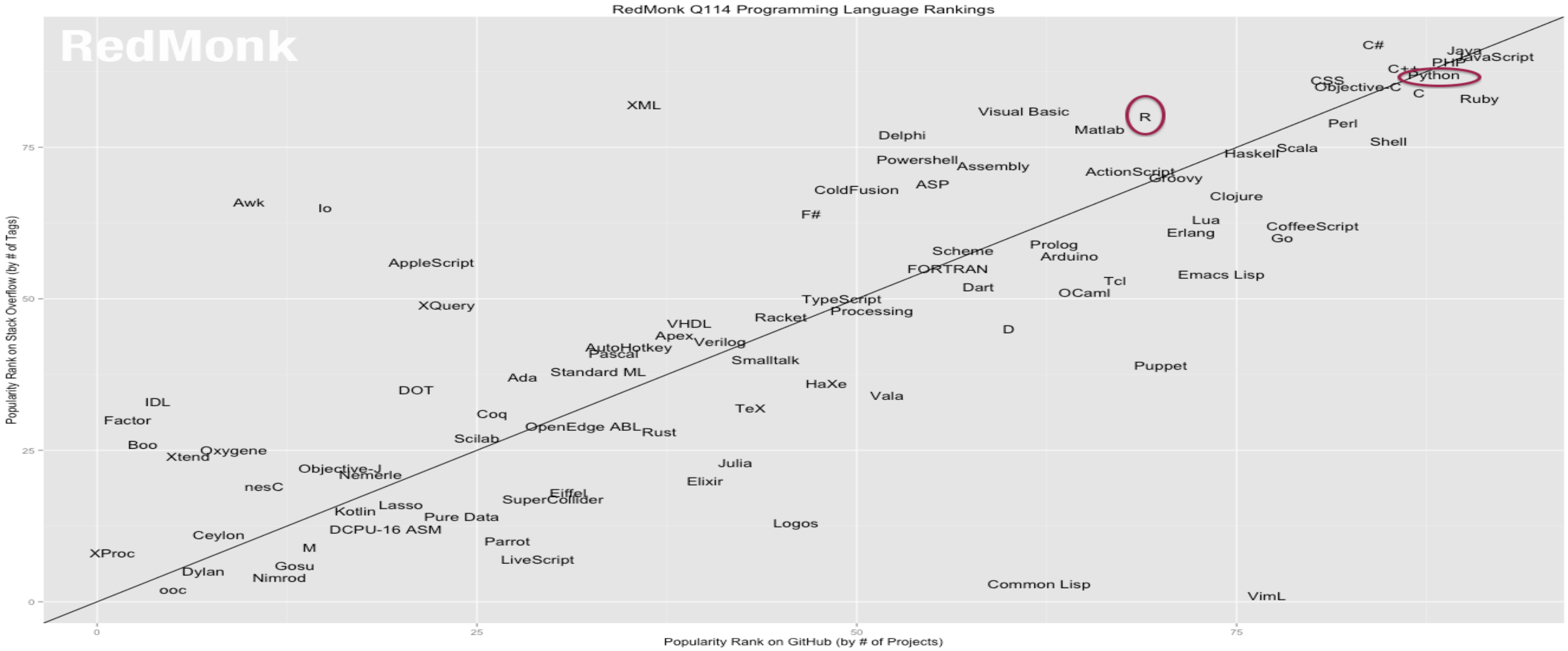
- Will go over in the app

Q1: As a researcher, you frequently compare mortality rates from particular causes across different States. You need a visualization that will let you see (for 2010 only) the crude mortality rate, across all States, from one cause (for example, Neoplasms, which are effectively cancers). Create a visualization that allows you to rank States by crude mortality for each cause of death.

Q2: Often you are asked whether particular States are improving their mortality rates (per cause) faster than, or slower than, the national average. Create a visualization that lets your clients see this for themselves for one cause of death at the time. Keep in mind that the national average should be weighted by the national population.

Moving to Python

- Switching from R to Python for this module
- Python is a great general purpose language, very popular



http://redmonk.com/sogrady/2014/01/22/language-rankings-1-14/

Visualization/Data Exploration is not Python's strength

- Not primarily a Read-Eval-Print Loop (REPL) environment
- Primarly viz tool is `matplotlib`: much lower level than `ggplot2`
- Much poorer set of baseline tools to analyze data

So Why Python?

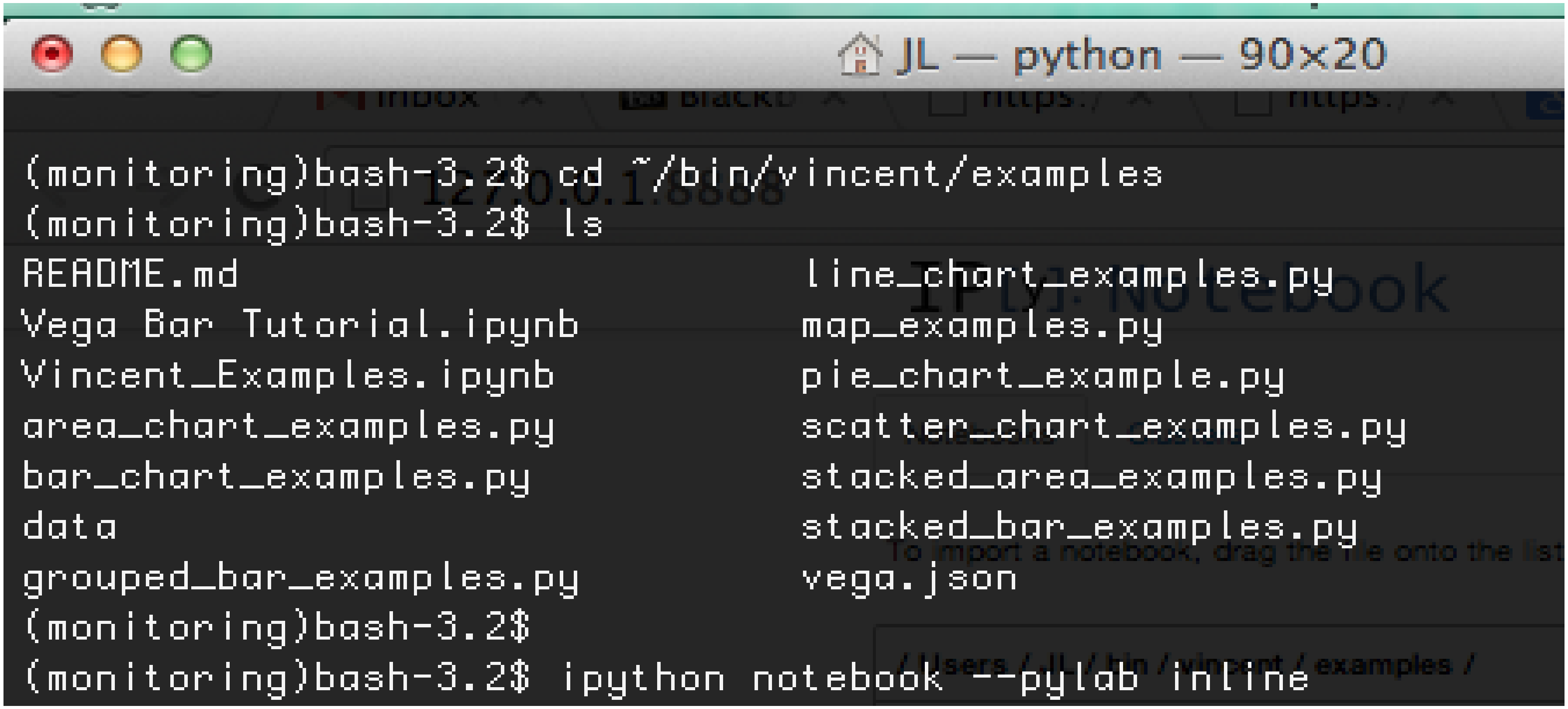
- Namespacing for multiple libraries (remember Hmsic vs plyr `summarize`)
- More libraries for working with other languages/web
- Considered to be better when dealing with big data sets (debatable)
- I use Python primarily, but still use R for some problems
- Makes you more hire-able in some industries

Ways We Can Improve Python

- Use `iPython notebook` to create REPL environment http://ipython.org/install.html
- Sister project `Anaconda` sort of acts as CRAN http://continuum.io/downloads
- Both from Continuum Analytics: US Government funded via DARPA
- NEITHER is required for this module's work: iPython strongly encouraged

Keys to Using iPython

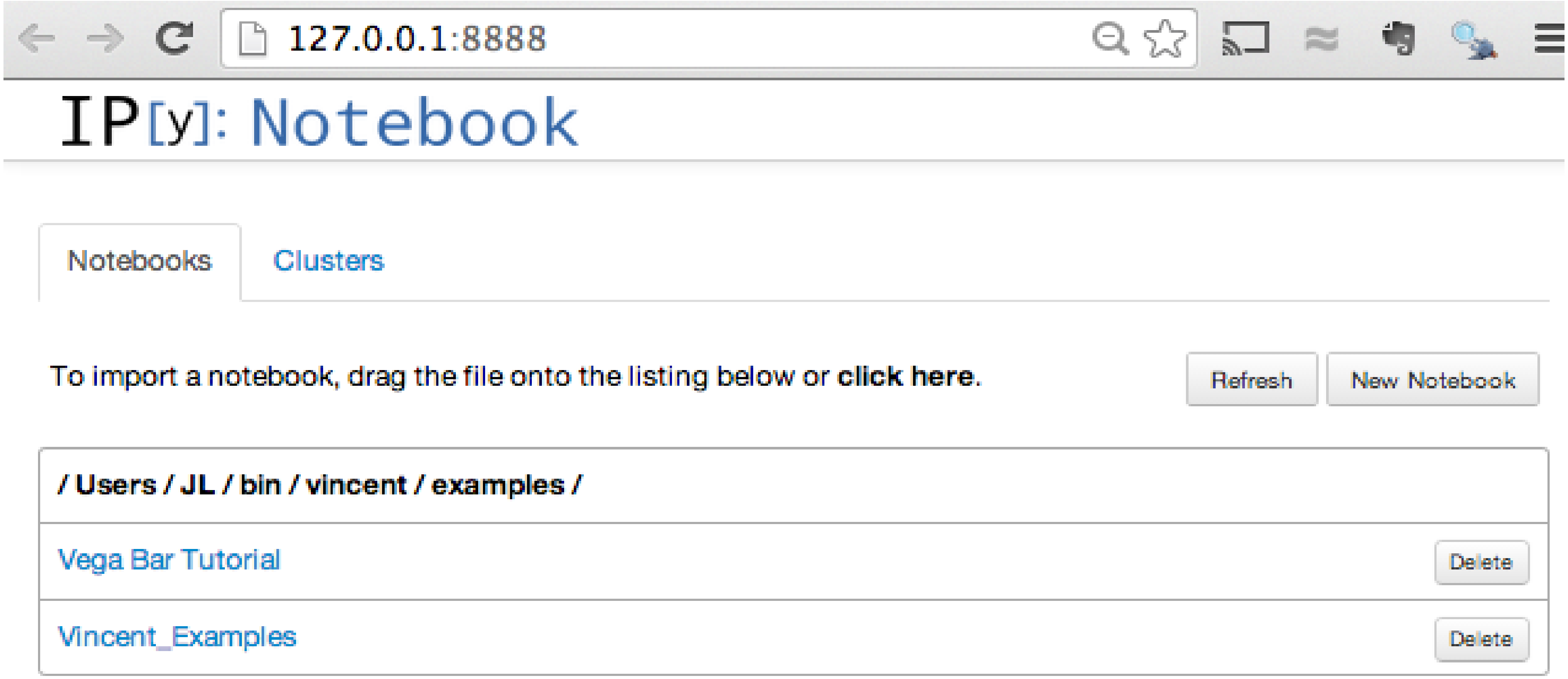
- Navigate to correct directory and run `ipython notebook --pylab inline`



```
(monitoring)bash-3.2$ cd ~/bin/vincent/examples
(monitoring)bash-3.2$ ls
README.md                         line_chart_examples.py
Vega Bar Tutorial.ipynb           map_examples.py
Vincent_Examples.ipynb            pie_chart_example.py
area_chart_examples.py            scatter_chart_examples.py
bar_chart_examples.py             stacked_area_examples.py
data                              stacked_bar_examples.py
grouped_bar_examples.py           vega.json
(monitoring)bash-3.2$
(monitoring)bash-3.2$ ipython notebook --pylab inline
```

Keys to Using iPython

- This will open a notebook viewer



To import a notebook, drag the file onto the listing below or **click here**.

/ Users / JL / bin / vincent / examples /

Vega Bar Tutorial

Vincent_Examples

Keys to Using iPython

- Once you have typed in code, run it line-by-line, like R
- Vizualization libraries will appear in-line!

Libraries to Assist With Graphing

- `Vincent` & `Vega` https://github.com/wrobstory/vincent https://github.com/trifacta/vega
- `ggplot2` ported to Python https://github.com/yhat/ggplot

This module's homework

- Hudson River Water Pollution
- Data from Riverkeeper http://www.riverkeeper.org/
- Extra Credit: find some context