



# Python for Data Analysis

Hunter Owens ([howens@uchicago.edu](mailto:howens@uchicago.edu))



# Why Python?

1. This is a good time to check that
  - You have the dataset (<http://bit.ly/19Tm8AQ>, unzipped: <http://bit.ly/1huLQzy>)
  - You have the required libraries
    - pandas, requests, matplotlib
    - run `python -> import [[library name]]` to test.
    - Raise your hand and one of the TA's can help if you are having troubles
2. Python integrates well with all toolsets
  - ie, Hadoop, Web Apps, etc
  - Not a DSL

# Problem 1

- go to [data.cityofchicago.org](https://data.cityofchicago.org) and download a dataset
- import the dataset into pandas, and create either a series or dataframe out of it.
- Run the Describe function on the dataset
- Compute something interesting on the dataset.

# Problem 2

- Download a 2nd, related dataset from [data.cityofchicago.gov](https://data.cityofchicago.gov)
  - or, use just the divvy sets
- Join the two datasets on a common value
- use the describe method learn about the combined dataframe
- Compute something interesting

# Problem 3

- Create a graph of station average over time for 1 particular station, rather than the entire systems. Print out the station name on the axis of the graph.