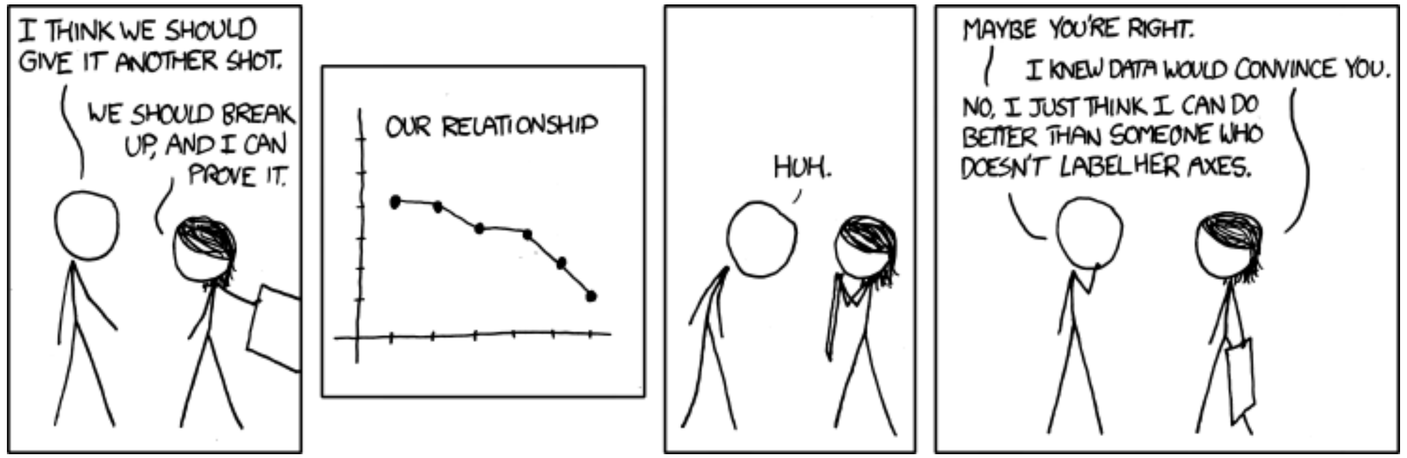
Introduction to Data Science
CS 5963 / Math 3900

Alexander Lex
alex@sci.utah.edu

Braxton Osting
osting@math.utah.edu

THE UNIVERSITY OF UTAH



[xkcd]

Project

It’s time to start thinking about your project.

What you need:

- A team
- An idea
- A dataset (that you actually can get!)
- http://datasciencecourse.net/2016/resources/

Project Phases

1. Announce your team and title (Friday, Oct 21)
2. Submit your project proposal (Friday, Oct 28)
3. Get feedback from staff (week 11)
4. Get/give peer feedback, (Lecture on Wed, Nov 9)
5. Submit project milestone
6. Get staff feedback (week 14)
7. Submit final project (Friday, Dec 2)
8. Present project (Mo/We week 16)

Project Requirements

Scope as agreed upon with TAs

Should contain:

- Data acquisition (scraping, API)
- Data cleanup
- Exploratory Visualization
- Two different analysis methods (classification, regression, clustering, NLP)
  - Evaluate alternative approaches for each one (e.g., compare two or more classification methods)

You can skip one of these, but you have to make up in other areas

- E.g., if you work with clean & existing dataset, the analysis has to be more sophisticated

Be ambitious! Define your goals and categorize them:

- must have, nice to have, etc.

Proposal Sections

Basic Info.

Background and Motivation

Project Objectives

Provide the primary questions you are trying to answer in your project.

Data

Data Processing

Exploratory Analysis

Analysis Methodology

Project Schedule

Submit as PDF or Jupyter notebook to Canvas

Milestone

Acuqired, cleaned data

EDA

Sketches of your analysis methods

Submit zip file with Jupyter Notebook, data, other resources.

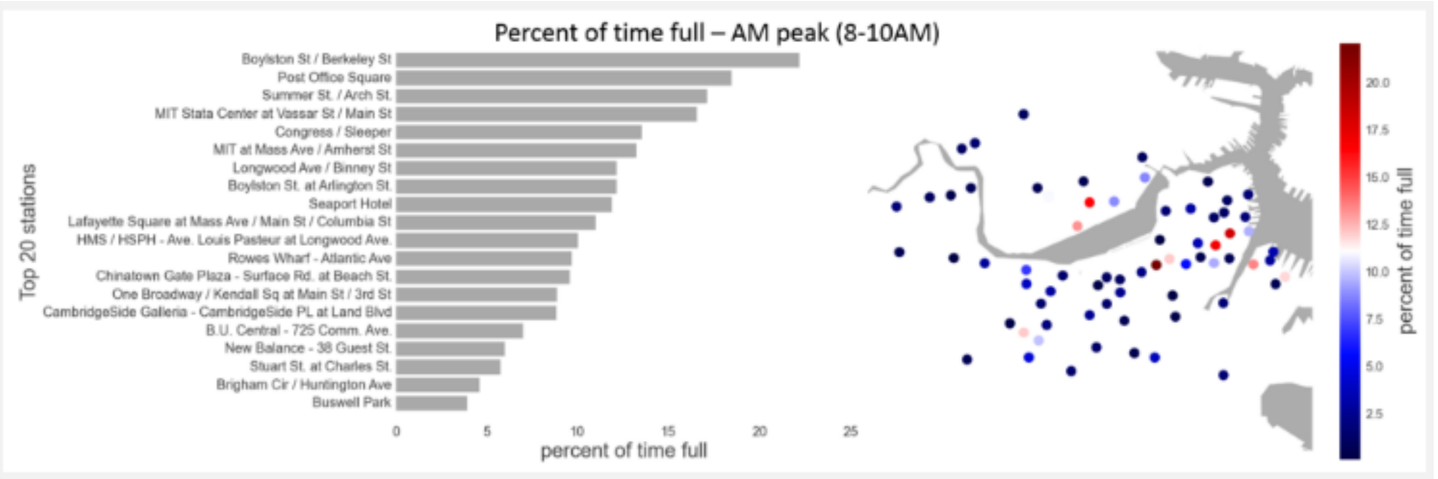
Final Submission

Whole story in a notebook

Include interpretation!

Three minute video that narrates project

Example Project: Hubway Analysis



http://cs109hubway.github.io/classp/

https://github.com/CS109Hubway/classp/blob/master/Hubway_Station_Analysis.ipynb