

# Day 21: Increasing Specificity in Projects

Raymond Yee

April 8, 2014 (<http://is.gd/wwod1421>)

# Announcements

- ▶ AJ is no longer the Tutor of the class. I'm in the process of hiring a new one. We'll miss AJ in this class.
- ▶ I read and commented on all the project proposals. See Project Proposal.
- ▶ Note: I expected each of you to submit a project proposal in bCourses. I won't assign you a grade until you've done so. Also, I strongly prefer that you paste your project proposal into the HTML text entry provided.
- ▶ I recommend tomorrow's talk – Victoria Stodden is a (the?) leading figure in the area of reproducible research. Toward Reproducible Computational Science: Reliability, Re-Use, and Readability | School of Information
- ▶ People interested in Amazon Web Services (AWS), consider going to AWS Workshop by Amazon - AWS Berkeley (Berkeley, CA) - Meetup/

# Clustering of Challenges

Thanks to Prabha Matta for clustering the post-its we compiled last week:

These are the problem categories I compiled based on the students' project feedback exercise held on last Thursday:

1. collaboration of ipython notebooks
2. statistics - intro to stats packages, normalizing data, comparison, correlations and regression analysis
3. creating interactive visualization/user interactivity
4. large datasets - possible use of AWS
5. data retrieval - wiki scraping, learn to use API - Flickr API, yelp API, google API,
6. Others: data cleaning, database design

# Some Geo work

I'll update the notes here based on what I show in the Day 21 classnotes.

A focus problem: generating topojson for census entities.

# Goal Today

- ▶ Each project should create a github repository in <https://github.com/working-with-open-data-2014> and add a link to each repo at <https://github.com/working-with-open-data-2014/project-organization/wiki/Projects>.
- ▶ In each repository, deposit a “project starter” IPython notebook in which you show to access at least one data set relevant to your project. Write at least a paragraph on how this data set is a good starting point for your project.

# Assignments / Homework

Don't forget Project Progress Report 1 – due tonight.