Notes from Apr 26 Meeting

Schedule Overview

Timeline
4/29 – Week 5
5/6 – Week 6 – Check In [Data Ready to Use and Analysis Roadmap]
5/13 - Week 7
5/20 – Week 8 – Check in
5/27 – Week 9
6/3 - Week 10 – Presentation
6/10 – Week 11 Project Software Due

Things to do:
- Get Data
- Determine best place to download data source
  - Data source: Explore these two options and get
  - From Kaggle using BigQuery to write to csv file
  - Stack Exchange Data Dump - torrent/7z
- Determine Workflow:
  - Download Data
  - Get the data and save it on Adam's server
  - Then explore if we want to store it on the server or store on cloud
- Subsetting Data
- Storing Data
- Distributing Data
- Finalize Questions
- Descriptive Statistics
- Create a Few Mockups
- Run the full analysis
- First Draft
  - Write up
  - Data Viz
- Editing
- Final Draft

Question:
Evolution of user from one who asks questions to one who answers question
- Do questioners use similar language as their answerers?
- As one becomes a matured answerer, is there language – stern, mean, short? [address a typical complain of stackoverflow]
    - Differentiate between users that start from questioner to answerer versus purely there to answer
    - Explore the heterogenous effect of gender, occupation, region/country
    - If the answerer is mean, do people who he/she has influence to answer, also answer meanly

Stack overflow users get meaner over time
Duplicate question, negative vote, read documentation, google it

Tasks to Do:
- Everyone: ZenHub Tutorial
- Dhruval - ZenHub populate the tasks
- Nikki - Look at tags and how to best use it.
- Before Week 6 Check In Meeting:
    - Adam & Nikki - Data collection and Storage
    - Dhruval & Li - Analysis Road Map