Team emails:

[Jack.Russo88@spsmail.cuny.edu](mailto:Jack.Russo88@spsmail.cuny.edu)

[Anthony.pagan@spsmail.cuny.edu](mailto:Anthony.pagan@spsmail.cuny.edu)

[Alexander.niculescu@spsmail.cuny.edu](mailto:Alexander.niculescu@spsmail.cuny.edu)

Timelines:

10/10 Team creation, Goto meeting account created

10/11 First team Goto meeting. Posted Notes in Slack. Selected Data.

10/12 Second team goto Meeting. Reviewed data . Each will take time to decide how to tidy/transform data.

Project 3

* You will need to determine what tool(s) you’ll use as a group to effectively collaborate, share code and any project documentation (such as motivation, approach, findings).
  + GotoMeeting
  + GitHub or post docs on Slack
* You will have to determine what data to collect, where the data can be found, and how to load it.
  + <https://www.kaggle.com/theworldbank/education-statistics>
  + <https://databank.worldbank.org/data/source/education-statistics-%5e-all-indicators#>
    - Academic achievements and expenditures by country/completion
    - How does it change literacy
    - How does population effect progression

The data that you decide to collect should reside in a relational database, in a set of normalized tables.

* Plyr to remove
* You should perform any needed tidying, transformations, and exploratory data analysis in R.
  + Is comp
* Your deliverable should include all code, results, and documentation of your motivation, approach, and findings.
* As a group, you should appoint (at least) three people to lead parts of the presentation.
* While you are strongly encouraged (and will hopefully find it fun) to try out statistics and data models, your grade will not be affected by the statistical analysis and modeling performed (since this is a semester one course on Data Acquisition and Management).
* Every student must be prepared to explain how the data was collected, loaded, transformed, tidied, and analyzed for outliers, etc. in our Meetup. This is the only way I’ll have to determine that everyone actively participated in the process, so you need to hold yourself responsible for understanding what your class-size team did! If you are unable to attend the meet up, then you need to either present to me one-on-one before the meetup presentation, or post a 3 to 5 minute video (e.g. on YouTube) explaining the process. Individual students will not be responsible for explaining any forays into statistical analysis, modeling, data mining, regression, decision trees, etc.