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.

10/14 Discussion via slack. Determine if current dataset usable.

10/15 Best school district by state does it correlate with google search.

Graduation Rates Alexander, Teacher Evaluation Anthony, Assessment Jack

10/17 Goto Meeting Discussion. Decided to go with Graduations rates and Teacher assessments. 2015 to 2016….Per\_pupil expenditure vs Overall\_Rating..teacher..Substitute Suppressed with some other analysis.

10/18 Slack, uploaded code and csv. Additional analysis and graphs needed.

10/20 Slack discussion on finalzing PPT

10/21 finishing touches

CORE Funcntion link : <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/cor.test.html>

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#](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.

Slide 1 Anthony Introduction

our team consisted of myself, Alexander Niculescu and Jack Russo. For our team project we decided to focus on Educational data focusing on teachers Ratings in the Albany district

We’ll start off by handing the presentation to Alexander

Slide 2 Alexander

Slide 4 – 6 Jack

Slide 7-9 Anthony

After some searches for data we settled on using data from data.nsyed.gov/. We were pressed for time and decided we needed to narrow scope to Albany school district for 2015. Getting the data included

* downloading data locally and saving to data frame
* it was a little more comfortable analyzing data in SQL so we create a my SQL db and imported the data.
* There were some issues with table keys on the 3 tables. So we joined the data with 2 keys and used a combination of the 2 keys to get rows for teachers.
* Once that was completed we wrote the data to a csv file for further analysis in R.

Will now hand off the presentation back to Jack.

Slide 10-13 Jack

Slide 14 Alexander