Schedules

* Mark has one take home final exam and one other final exam, not sure when but he should be free to work on this for the next week.
* Chrystal has her prelims Friday afternoon and Saturday morning. From Saturday afternoon and on she can work full-time on the project, but before that she can do minimal work.
* Andrew has a draft paper due Monday afternoon, but should be able to work on both the project and the paper at the same time.
* Next meeting: **Sunday 5/10, 3pm**

Report

* Use Google Docs

Presentation

* Andrew and Mark have experience with video editing
* Probably can use screen cast + voiceover

Questions

* Can be modified as we work

EDA/Data Cleaning/Visualization

* Limit data to the 50 states and D.C.
* For each dataset, describe the data: what the columns represent, the data types (strings, ints, nulls, datetimes)
* Make some visuals to see how the data is distributed in general
  + Andrew: heat map of rate or final number of cases with respect to geographical position
  + Andrew: distribution of demographics with respect to counties: average income, population density, population/land area, rural/urban (column 13 of counties: rural/urban continuum score), Dem/Rep governor, insurance coverage
  + Mark: distribution of intervention methods
  + Chrystal: time series of case growth in each region
  + Everyone: output variables such as final case numbers, mortality rate, incident rate
* Make cleaner, nicer dataframes from initial datasets

Method and Experiments

* Primarily using linear regression
* Feature engineering
* Cross validation
* Regularization (Ridge/LASSO regression)