**Project Proposal Outline**
Fall 2016
CS3943/9223
Prof. Rumi Chunara
Due: 10/20/16
Total: 25 points (25% of project grade)

The project will be evaluated based on achievement of the stated goals. So, don't get over excited and list every single possible thing you can explore. Be specific and detailed. Pick one concrete question to answer and use descriptive statistics to describe the issue and motivate your choice of analyses.

At the proposal stage, you should already have selected and explored your data. Thus, the problem, data and approach can be fully defined and should be in the proposal. We will evaluate your proposed approach for clarity and thoughtfulness of approach are key parts of the grade, and make suggestions where applicable. Each point below should be addressed clearly.

The length and number of figures/tables should be sufficient to explain your project. This should be approximately 600-700 words and up to 2-3 figures/tables.

1. Describing the problem.
    - What is the problem?
    - What type of problem(s) is this (text processing, classification, regression, etc.)

2. Background.
    - Provide some initial motivation for the project.
    - How are you learning the background? A*re there specific publications or documents that discuss related information (what are those), or is there a domain expert you will engage (what is their relevant experience)?*

3. Describing the data.
    - What kinds of data will you use?
    - Where do you obtain it?
    - What features will be included?*(describe the features fully including it's temporal and spatial dimensions, types and scales (e.g. numerical or text, ordinal or nominal, etc.)*

4. Outcome.
    - What outcome will be computed? *It is not an appropriate problem if there are existing solutions available or proposed.*

- What type of variables will be in the output?
- Why is this outcome useful?
- What confounders exist?
- How will you evaluate the outcome?

5. Describing the Methods.
    - What kind of model will you build?
    - Why is this approach appropriate? *(What approach will you take for solving the problem, including how data will be cleaned, what specific algorithm(s) and any parameters used, and how you will evaluate your approach)*
    - What assumptions are safe to make? *This could be in terms of features considered, variable types, what will be done with missing data, etc.*