PURPOSE

The main purpose of the Project Proposal is to give you the opportunity to brainstorm a list of questions for your final project and then write a brief proposal on how you plan to answer 1 -2 main questions of interest for your final project.

DELIVERABLE *(Due March 23rd 11:59 pm) [5 % of grade]*

A **one page** google doc file shared with your classmates, me and the TA that contains the following information:

1) Team Name

2) **Your MAIN question (Maximum 2)**

3) **Your SUB questions (Minimum 5)**

4) **Proposed Data sets** (*inputs*) that you might use to answer your above questions (Minimum Three)

- Mention the dataset (with weblink wherever applicable) and a one line description of the variables you plan to extract.

5) **Proposed *outputs* (Minimum 5)**

- such as tables/plots/maps/models from the identified data sets to answer your questions. Mention the output and a one liner about what function fo the output.

SUGGESTED APPROACH

The question generating exercise we conducting in class is a good starter. You can use questions from the list you generated or come up with new ones. Since your group members will have access to google docs, each member can generate a few Main and few sub questions individually and put on google docs. Then through google docs and other communication tools such as Piazza (I will create individual piazza groups for each project team) and Zoom, you can vote to choose 2-3 main questions.

Once those 2-3 main questions are selected, you can generate a list of sub questions together (these will generally be a LOT of sub questions) for your TOP choice. Then, you will do some detective work to see if there are any public datasets or other available datasets (including those in an api etc) that you can use to answer some sub-questions). While you are searching for datasets, also think about what kind of output you potentially envision to answer your question. In case you don't find much data for your TOP main question, you can then move on to the 2nd or 3rd main question and repeat the process.

I suggest that in your google doc, you create an **appendix,** which basically stores all the other main questions, subquestions, datasets etc that you brainstormed but did not put in your final proposal. This way, in case I feel that your main proposal is unrealistic, you will still have your other brainstorming ideas for your project.

Remember that it is totally OK and expected that your final project may end up being different than what you proposed. That is completely OK. But you will realize that if you spend some good quality time thinking about your project proposal, it will immensely speed up the final

project and give clarity of thought. This proposal does not have to be perfect. I just want to make sure you start thinking about your projects early enough and not leave them for late. If you submit in time, you will get extensive feedback on your proposal to help you make a really good final project.

The proposal is **worth 5 %** of your grade. You will get full points as long as you follow the instructions, and demonstrate that you have sufficiently thought about your project. Otherwise I will deduct points based on either any missing portions of the proposal or if I feel that you have not given enough thought to your final projects. I have also attached a 1 page example proposal. Hopefully, it will give you a good template for your own proposals.

********** **EXAMPLE PROJECT PROPOSAL** **********

Team name: Foodies FoR All

Main Question

How does the food and the transportation in *North Carolina* affect childhood obesity rates?

Sub Questions

1. Where are grocery stores located? Where are convenience stores location
2. Which places, groups of people travel furthest to get food?
3. What is the dominant mode of transportation in each census Tract in north Carolina.
4. What is the relationship between number of grocery stores and percentage of population having access to public transport?
5. As distance to a grocery store decreases, does childhood obesity increase
6. Are certain communities that are disproportionately exposed to obesogenic environments

Proposed Datasets

1) Census Api through R - To get information on mode of transportation, and demographics
2) USDA food access database [https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/] - To get GIS information on location of food deserts and food access index
3) Nielsen convenience/grocery stores database [https://catalog.data.gov/dataset/nielsen-tdlinx] - Contains addresses of all super markets, grocery stores and convenience stores in the US.

Proposed Outputs

1) Choropleth map of census level travel time to nearest grocery store
2) Box plot comparing differences in travel time to grocery stores and convenience stores among black and white majority areas
3) Scatter plot by race to measure relationship between distance to grocery store and rates of child hood obesity in each census tract in North Carolina
4) Map showing hotspots or clusters of census tracts which have both high childhood obesity and low access to healthy foods. This will show places that can be targeted for a health intervention.
5) Map showing the geographical differences in ratio of people with lower social economic status to those with high SES, living within one mile of a grocery store. This will help

provide information on whether grocery stores are mainly situated in richer neighborhoods vs poorer neighborhoods.