SPPUA 5262

Second Mid-Term Project

***Measurement:* Constructing a Novel Metric from Raw Data**

This unit of the semester has focused on using a theoretical concept to guide the construction of one or more novel metrics from our raw data sets, describing a particular aggregate unit of analysis (e.g., parcel, street, neighborhood). This will culminate in our second midterm assignment, which will include two parts: the individual portion will be a short paper summarizing the measure and its construction; the group portion will be an updated version of the database and accompanying documentation, including the record-level data set *as well as* any newly constructed datasets at the aggregate level.

Project: 15% of final grade
Due Date: November 7th in class

Individual Portion:

The paper should be organized in a series of short sections:

- An *Overview* of the measure and why it is interesting. About 1-2 paragraphs.
- A textual description of *Measure Construction*, justifying any specific decisions that were made (for example, categorization of case types). This will include:
    - A summary of new variables at the record level (i.e., the original database) that were constructed first as part of the overall calculation of your aggregate measure(s).
    - A summary of the new aggregate measure(s) that you've calculated and how you have done so.
- A short description of the new variable's distribution and/or values. What's the mean? Where is it highest? Anything else fun or interesting? Etc. This should include at least one tabular visualization (e.g., histogram) and a map, if appropriate.
- An appendix with an annotated R syntax articulating all steps required to create the measure(s) (should be your portion of the R syntax copy-and-pasted; see below).

Rubric (Total 10 pts.)

Measurement-Concept: 3 pts.
Measurement-Execution: 2.5 pts.
Visualization: 2.5 pts.
Details (Grammar, etc.): 2 pts.

Group Portion:

As with the first midterm, the group will work together to combine all modifications made by group members into an updated data set and documentation. It will include four parts that will be uploaded to the Dataverse:

- The record-level file with all new and modified variables. **This is an update of the record-level file submitted for the first midterm.**
- Any files at the aggregate level (e.g., census tracts) and the variables describing them.
- An updated Data Dictionary that describes the new variables and, if necessary, modifies the description of any already existing variables, *as well as* variables at the aggregate level of the database. **This is an update of the Data Dictionary submitted for the first midterm.**
- Annotated R syntax clearly articulating steps for all data cleaning and variable creation. This should be efficient and complete such that the code could be run all at once on raw data to recreate the updated data set. **This is an update of the R Syntax submitted for the first midterm.**

Rubric (Total 5 pts.)

Data: 2 pts.
Syntax: 2 pts.
Updated Documentation: 1 pt.