1. Reading On-Campus data file into SAS and cleaning the data.

Create formats, validation, check for errors, as needed, only need 2010 for each violation

2. Reading FBI records data into SAS and cleaning and verifying data

-using arrays as necessary to create rates per population

-I’ve already done some cleaning in the weekly discussions on this data

-I had created a new variable for total violent crime

3. Merging the Datasets together and only including city/state in both data sets

Validation of Merged data

(Possibly merge with the income data set as well)

Possibly Conditional processing or do loops (possibly create tier variable based on crime classification, low/ medium/ high)

Possibly sum the crime variables by State

(Look at total crimes committed on-campus/ city for all cities in the state with universities

then adjust for population)

4. Some sort of very brief data analysis with proc means, freq, and sorts

Maybe just do proc sorts of cities with most on campus crime per population and also sort by most violent crime in surrounding city

Classify total violent crime or total property crime from FBI data as Medium/High/ Low in level and same for university crime and then do a proc freq. (How many cities with High City, High University crime… High City, Medium University crime… etc)

Any other ideas?

5. Merging all code and make sure it runs

6. Writing the report (Introduction, and methods section in report, also results section) Note the length specifications of the report (4 pages max so must be very concise)

(Copied from instructions)

c. Introduction section:

Statement of the business, science, or research interest in preparing the data you’ve chosen. Background information on the data source(s), including a description and links of where it originated.

d. Methods section: Description of the original data files. Description of the guidelines used to validate the data. Description of the issues needed to be cleaned and how it will be done (though not needing to explain the programming specifically). Description of additional data preparation that you performed (e.g. merging, joining, subsetting). Description of variables to be analyzed including attributes such as name and type. You do not have to list all the variables in the original sourced files, but do mention the ones you bring to SAS for the creation of the SAS data set.

e. Results section: Charts and tables pertaining to validation and cleaning. Write-up of the results. Point out notable information from the charts and tables.