Introduction:
Throughout the semester, we have learned how to use a variety of analytical tools to evaluate quantitative and statistical analyses prepared by others. This exercise provides you an opportunity to conduct an analysis on a small dataset yourself, as well as to interpret your findings in a short writing assignment.

Assignment:
We have prepared a dataset of information from 100 households with 18 variables. (The variables included the dataset, along with their associated questions and codes, are attached to these instructions.) One half of the observations were drawn (using a simple random sample) from a small community (population ≤ 15,000), and one half were drawn (also with an SRS) from a large city (population > 500,000). Both communities are located in the same state, although you do not need to know which state this is in order to complete the assignment.

Using the tools we have covered throughout the course, conduct a comparative analysis of the two subsamples (“small town” and “big city”). Your objective is to draft a short paper comparing and contrasting the two groups. You will not be able to cover all the similarities and differences between the groups; you must, in the course of your analysis, decide which are the most interesting issues that arise and that deserve to be included in your paper.

The first step in such an analysis is to get to know your data. Using the various techniques and tools that we’ve covered, you will want to generate descriptive statistics for the two groups. Next, think about relationships among the variables included in the dataset. Are there any interesting links between different household characteristics? Once you have a good understanding of the two groups, compare and contrast their characteristics with one another. What are the most interesting or surprising findings in your analysis?

While you are encouraged to include a few well-chosen and well-constructed tables and graphs in your paper, it is not enough simply to provide numerical summaries of your output. You must translate those findings into a well-written narrative that tells the reader what you’ve learned about these communities. Choose your language carefully. Remember to use terminology correctly, including words such as “significant” and “substantial,” and when you discuss concepts such as association and causality.

Grading:
Grading of the paper will be based on the following criteria (please pay attention to these!):
• Application of statistical tools (e.g., Were the correct analytical techniques used given the question you were asking and the type of data you were working with?)
Interpretation of findings (e.g., Do your interpretations make sense? Are they consistent with your data? Is your narrative supported by your data, or have you made generalizations that are beyond the scope of your analysis?)

Choice of findings presented (e.g., Have you included interesting findings from your analysis, or have you just included all the output you could generate? Note that a paper that makes a manageable number of points well and thoroughly will receive a better grade than one that includes a laundry list of findings but little effort to explore them or to think about the linkages among them.)

Quality of writing (e.g., Did you use correct grammar, spelling, and punctuation? Is your writing, clear, tight, and to the point, or is it rambling and difficult to follow?) We strongly suggest that you ask a colleague to proofread and/or edit your paper before you hand it in—you will be penalized for poor writing.

Quality of graphical presentation (e.g., Are graphs well-labeled and easy to understand? Are they too simple or too complex (see below)? Do they support or detract from the analysis put forward in your text?)

Format:

All papers must be composed using a computer and word processing software. They must also conform to the following guidelines:

- Maximum of 8 double-spaced pages, which includes all text, tables, and figures (but not the title page). (If you submit more than 8 pages, the excess pages will not be graded.) Please see the attached sheet for an example of paragraph formatting.

- Minimum of 4 double-spaced pages of text.

- 8½ X 11” white paper (recycled is fine)

- 1-inch margins on all sides

- Printing on one side only

- Title page included

- 12-point font, Times or Times New Roman for your text (you may use other fonts for your graphs if you like)

- Numbered pages

Rules on group work:

This assignment is an opportunity for us to assess your individual ability to make sense of a dataset and to write about it clearly. You may discuss the data with one another, and you may offer and receive help on getting Excel (or any other data analysis software) to do what you want it to do. However, all decisions regarding which findings to include in the paper and how to interpret the findings of your analysis must be your own. You must conduct your own analysis, create your own tables, figures, or charts, and write your own paper. If you have any questions on this, ask one of the course staff.
Important notes and hints:

It is essential that you have facility with Excel (or a statistical software program) before taking on this assignment. If your Excel skills are weak, find a colleague to help you with them, or ask your TA if you need advice on conducting a specific piece of analysis. Although our dataset is considered “small” by most social science standards, it is definitely too large for you to be trying to compute statistics by hand. The better you are with Excel or another software program, the quicker you will be able to get through the generation of tables and figures and move on to the interpretation and analysis of your findings. (We will be offering a tutorial in SPSS—a statistical software package that could save you a lot of time on this assignment. Watch for announcements about the scheduling of this session.)

When you first begin the assignment, you will probably generate a lot of tables and graphs. These are important for you to do the work, but most of them should probably not be included in your final paper. Resist the temptation to load up your paper with evidence of all the work you have done. We recognize how much “behind the scenes” effort will be required for you to produce a high-quality analysis; just include those tables and figures that support the main points you are making in your text. Similarly, be careful about including tables or figures that are very simple (i.e., you could have summarized their content in a sentence or two within your paper) or overly complicated (i.e., they are more confusing than helpful to the reader).

You might find it useful to split the dataset into two Excel files so you can more easily deal with one community at a time. Remember, however, that you can save a lot of time by copying and pasting formulas that you write for one of the files into the second spreadsheet (rather than starting from scratch). Again, if you are not sure how to do this, check with a colleague or a TA.

There are only 18 variables in the dataset, and most likely you will have questions about these communities that you cannot answer because a relevant variable was not included. Recognize that this happens in most quantitative analysis; we rarely have all the information we would like to complete an analysis. You can certainly “flag” these lingering questions in your paper if you like, but we recommend that you focus your energies on what you can say based on the data that you do have. Similarly, do not get distracted in a discussion of the sampling or elicitation procedures used to generate these data. If you feel strongly about issues of validity with respect to particular indicators, you may include a few concise comments on this topic. Otherwise, just focus on interpreting the data you have rather than second-guessing their origin.

For each variable, there are some individuals who gave “Don’t know” answers or whose code is NA, meaning “not answered.” This is very common in social science research and as analysts we must decide how to deal with these data. Use any (reasonable) approach you like, but be sure to (1) state how such individuals were treated and (2) take the proper steps when doing computations to make sure the codes for these individuals aren’t treated as if they had numerical “meaning.”

Finally, if you have any questions or problems with this assignment, please get in touch with one of the course staff. We want this to be a useful learning experience for you, and if there are ways we can help we’d like to do so.