

STA 320 Final Project

Dr. Kari Lock Morgan and Dr. Fan Li

Due date: Friday 4/25, 5pm

Please submit a hard copy under Prof Li or Prof Morgan's office door, as well as an electronic copy on Sakai by the deadline. Submit your R Script and dataset on Sakai, as well as your paper. No late projects accepted.

Grade: 30% of final course grade

The final project comprises 30% of your final grade in the course. Your grade for the project will take into account at least the following factors (among other considerations):

- quality and accuracy of statistical methods used
- demonstrated level of understanding
- accuracy
- depth of thought and analysis
- effort
- difficulty of topic and approach
- quality of writing and organization
- explanation of reasoning and methods
- interpretation of results and suggested implications
- acknowledgement of limitations

Topic: Any topic related to causal inference

Your final project in this class can be any topic related to causal inference. Some suggestions for topics are given below, but topics are not limited to these suggestions and you should feel free to discuss other project ideas with Prof Morgan or Prof Li. Possible topics:

- Pose a causal question and answer it using available observational data
- Analyze data from a randomized experiment
- Improve upon an existing causal analysis
- Pose a causal question and answer it by designing, conducting, and analyzing a randomized experiment (talk to us first if humans will be involved)
- Further explore any topic discussed in class (e.g. model-based imputation)
- Research a topic in causal inference not covered in class
- Anything else you want to propose!

Finding Data

You may use any dataset of interest to you that can be made relevant to this course. Duke Libraries Data and GIS Lab can be a great resource for finding data. If you are unsure of where to look for data, take a look at the guide they put together for STA 101 to help students at Duke find data for stat projects:

<http://guides.library.duke.edu/stat101>

In particular, look at the tabs for Duke Data and GIS Collections page which contains many datasets searchable by topic, and tabs for two large datasets with many variables: the General Social Survey and the American National Election Survey.

Length: Maximum of 10 pages

The project should not exceed 10 pages, double-spaced. This includes figures and references. R code and the data should be uploaded separately and do not count towards these 10 pages.

R Script

Anything you do for your project should be included in your R Script that you upload to Sakai. We should be able to simply run your R Script to reproduce your results. For example, if you cycle through the trimming code 3 times to eliminate units, this code should be pasted in three times. Just before submitting your project and code, clear your history and rerun your R Script to make sure you get the same results. Your R Script should be well commented and organized. If we cannot easily reproduce your results from your script you will be penalized.

Policy: Project is individual

This project is large portion of your grade, and to be fair to all students, we want to assess what you can do as an individual. You may consult with the Duke Data and GIS Lab regarding finding data and getting the data into R, and may talk with Professor Morgan or Professor Li regarding your project, but may not discuss any aspect of the project with anyone else. Any communication of any kind with anyone other than Duke Data and GIS staff regarding getting data, or with Professor Morgan or Li, is not permitted. This includes R code - all coding must be done entirely on your own. Moreover, plagiarism will not be permitted, and all writing must be entirely in your own words, or else quoted appropriately. If you suspect that a classmate is potentially violating this policy, please contact Professor Morgan or Professor Li immediately. You will be asked to sign an honor statement upon submission stating that you have abided by this policy.