

MA Research Commitment

Assignment 3, MA Thesis Data and Methods

Joshua Gary Mausolf and Nora Nickels

Due 11:59 PM CDT on December 14th

Requirements:

The MA Research Data and Methods should be a **5-6 page, double-spaced**, and clearly articulated paper, spending approximately 2-3 pages each, reviewing the (1) data and (2) methods for the assignment. You must also include (3) works cited with your submission. The 5-6 page length requirement does not include the works cited.

Required Sections:

1. Data: ~2-3 pages:

The data section should provide an in-depth overview of your data source including its origins (data sources or methods of obtainment), past uses (if any), scope, and size:

- **Origins** Describe your data source and data origins. For example, if your dataset has been previously compiled or collected, cite the exact source of the data and explain how it was accessed. Alternatively, if you are collecting or generating your data, explain that process. This might be a detailed overview of an experimental or survey protocol, an web-scraping/API collection process, or data simulation models.
- **Past Uses or Approaches** Your project did not originate in an ideological vacuum. *Briefly* articulate whether other researchers have (A) used your exact dataset or (B) produced a similar dataset (via using the same sources Reddit/Twitter; a similar experimental design; or similar simulation). Although this could be a long segment, limit yourself to a paragraph providing some grounding of the *data* being used.
- **Scope** What is your data's scope? In this section, articulate both the specific unit of analysis and time frame for your data. For example, what research population is it applicable to? Does your data represent all American citizens ages 18 and older? Alternatively, are you studying a very select group of individuals: 'Caucasian American men ages 21-29 working in Fortune 100 Investment Banks.' Specify these details about who your data is representative of and also indicate the time-scale represented. For example, do you have annual or monthly data of this population from 1980-2018 or cross-sectional data captured at one in 2018? If you are performing an experiment or survey, what inclusion criteria did you use to determine subjects/respondents?
- **Size** How many observations does your dataset have and at what level are these observations? For example, you might have 2M observations of individual behavior (such as contributions, Tweets, or messages) that occur over a year and represent 50,000 individuals. Often, including a simple table is an easy way to convey this data. If you are still collecting data, then offer your best projections.

2. Methods: ~2-3 pages

Your methods section should articulate what your variables you will use [dependent (outcome / Y) variable; independent (predictive / X / feature) variables; and controls/covariates], what analyses you will conduct, and the justification of those analytic methods.

- **Variables** Clearly articulate what your dependent and primary independent variable(s) are. If you have any important confounder variables that you control for, then you may discuss those as well. You can be more far more general in discussing what other covariates or controls you intend to additionally include in the analysis.
- **Analysis Method** Discuss what specific methods of analysis you will use. In other words, given your data and variables, what type of statistical or computational analysis will you use to demonstrate that a relationship exists between your primary independent variable and some outcome variable? Please be as specific as you can. For example, you might intend to offer both descriptive or bivariate statistics (or figures) of the relationship as well as a more complex statistical or computational approach (or combination thereof) such as ARFIMA time-series models and temporally cross-validated random forests with boosting.

It is okay if you have not settled on the final model parameterization, but you should be able to indicate which types or combination of models or analyses you intend to run and what strategy you might use to refine the analyses.

- **Analysis Justification** In addition to explaining what type of analyses you intend to perform, please provide some justification of your proposed analytic methods. For example, if you want to use ARFIMA models, why did you select ARFIMA and not ARIMA/ARMA or another distributional approach such as negative binomial time-series models? Here, you might both cite past authors who (A) developed the model you intend to use and (B) have used that model for a similar research design as your project. This section can be integrated with the above analytic methods.

3. References (Works Cited)

Include an additional section of works cited. You should follow an accepted and consistent in-text citations and bibliography style such as ASA, APA, or Chicago.

Submission

The assignment should be submitted in PDF format, attached to an email. The email should be sent to both Joshua Mausolf (jmausolf@uchicago.edu) and Nora Nickels (nnickels@uchicago.edu). This assignment is due by 11:59 PM, Central Time on December 14th.

Late submissions will be deducted one letter grade for every day they are late.