

Student Name
Prof. Vicky Steeves
Data Librarianship and Management
October 30, 2018

My working method for this project will be the performance of a mixture of qualitative and quantitative analysis of my datasets. Qualitatively, I will use a Python script to scrape death row information for the last words of condemned prisoners in Texas. I will perform a content analysis of the text by running a statistical analysis of word frequency as well as sentiment analysis using the TM package in R to see what trends if any exist in this data. For my quantitative analysis, I will investigate my datasets for information on prison facilities and the individuals incarcerated in them. I will perform simple statistical analyses and attend to information regarding the education levels and length of sentences of incarcerated individuals. I will work to place Texas in context with data drawn from other top-ten incarceration states in order to understand how it fits into larger trends. I will take the information derived from these analysis modes and will produce visualizations of my findings in Tableau and CARTO.

These processes will, of necessity, be iterative and will require that I stay close to the data captured in the sources I am using. These constraints lend themselves to employing grounded theory. Grounded theory advocates a form of comparative data gathering and analysis that is focused on a “systemic asking of generative and concept-relating questions, theoretical sampling, [and] systematic coding procedures” (Strauss and Corbin, 1997, p. 274). Such analyses are well-suited to questions of power and the consequences of its use (Strauss and Corbin, 1997) and are applicable to discussions of social phenomena like mass incarceration. Grounded theory is primarily associated with qualitative analysis, though its main proponents acknowledge that it can accommodate both qualitative and quantitative (Strauss and Corbin, 1997). My use of a mixed method, and my preference for content analysis as well as data visualization, has led me to use a modified version of grounded theory. Specifically, I will work along the lines espoused by Yu, Jannash-Pennel, and DiGangi (2011) who note that the use of content analysis and data visualization are compatible with the qualitative focus of grounded theory. These authors note, for example, that “...both grounded theory and text mining utilize an iterative process. In the former, initial categories extracted from the data must be constantly compared against new data (p. 732), and thus the researcher is open to the possibility that previous categories might be collapsed and revised, and new categories might be added. By the same token, a text mining

algorithm is designed to learn from the data by revising the categories” (p. 732). As my use of these methods is ultimately in service of “telling the stories” (Strauss and Corbin 2011, p 281) of those affected by mass incarceration, grounded theory—modified or not—seems an appropriate approach.

Works Cited:

Anselm Strauss, Juliet Corbin, eds. *Grounded Theory in Practice* (Thousand Oaks, CA: Sage Publications, 1997).

Chong Ho Yu, Angel Jannasch-Pennel, Samuel DiGangi, [Compatibility between Text Mining and Qualitative Research in the Perspectives of Grounded Theory, Content Analysis, and Reliability](#), *The Qualitative Report*, 16, 3 (2011): 730-744.

Student Name

LIS 628 – Data Librarianship and Management

Prof. Vicky Steeves

14 October 2018

Scope

The state of Texas is the seventh most incarcerated state in the nation (Kansas is the most). Unlike other high incarceration states, Texas is a bastion of capital punishment and has a higher rate of execution than its peers. With each execution, from 1982 to present, the state has recorded the condemned person's last words, if any. As a means of shedding light on the human aspect of capital punishment, I will scrape death row information from the Texas Department of Criminal Justice website, including the last words of executed individuals. I plan to do some text analysis on these last words to gain a deeper insight into word frequency and sentiment. Although it is too early to know what I will find, I am interested in the ways that these final statements may help to humanize the numbers presented in the data. Capital punishment in Texas is only part of the story, however. State-sanctioned executions, like those in Texas, take place within a system of mass incarceration. In order to understand this context, my project investigates datasets regarding prison facilities and incarcerated individuals in the United States. The datasets that I will be using are provided by the U.S. Department of Justice in conjunction with the Office of Justice Programs and the Bureau of Justice Statistics and are made publicly available by the Inter-University Consortium for Political and Social Research (ICPSR). The specific datasets I am using include the National Prison Statistics, 1978-2016 (ICPSR 37003) and the Annual Survey of Jails, 2015 (ICSPR 6760). From these datasets, I will be examining information on the state of Texas and analyzing the demographics, lengths of sentences, and education levels of incarcerated individuals and comparing them with similar data from other top-ten incarceration states.

Data Management Plan

The datasets available in the ICPSR portal make use of complex (and confusing) acronyms for column headers and often use numbers to signify string values (i.e. -9 to mean "blank", 0 to

mean “Reported”, 1 to mean “Estimated” etc.). I will have to clean this data in OpenRefine to make it usable. The “tidy” version of ICPSR 37003 and ICSPR 6760 will be refined further to isolate a subset representing information about Texas. Additionally, scraping the Texas Department of Criminal Justice website will result in a file containing information on individuals executed by the state and their last words. All scraped data and derivative spreadsheets will be saved as comma-separated (.csv) file formats and versions will be carefully managed using consistent file-naming convention leading with ISO 8601 formatted dates. The Python script used to scrape the site will be saved as a .py file. All cleaned datasets, Texas subset, scraped data, and code will be made available on Github and the Open Science Framework. A Full description of this project will be provided on this projects OSF’s wiki and all respective files will be made available on this platform. All files will have a corresponding README and a codebook stating all changes made from the original, acronym-based column headers (numbering of column headers will remain unchanged for cross-referencing purpose). In keeping with data management best practices, three copies of the data will be saved in difference locations—an external drive, a cloud-based platform, and the corresponding OSF page for this project.

Methodology

The United States currently imprisons more people per capita than any other country in the world. The nation holds approximately 2.3 million individuals in various state prisons, federal prisons, juvenile correctional facilities, and local jails as well as in military prisons, ICE facilities, state psychiatric hospitals, and prisons in the U.S. territories. Given the rate of imprisonment in the United States, and the reverberating effects of being labeled an “ex-offender” after release (e.g. felony disenfranchisement), this project aims to underscore the need to address mass incarceration by placing it within the discourse of social justice and criminal justice reform. My research will focus on Texas in particular, but will place that analysis in a larger context by locating it within the Gulf Coast, which is the most incarcerated geographical region in the most incarcerated nation in the world.

This project will use a mixed methods approach and will utilize publically available data as well as information gathered from the Texas Department of Criminal Justice website, including the last words of executed individuals. The quantitative component of this project includes

analyzing demographic information, sentence length, and education levels found in two ICPSR datasets (ICPSR 37003 and ICSPR 6760). Visualizations will be made using this quantitative data, likely using Tableau and CARTO. In keeping with the work of Yu, Jannasch-Pennell, and DiGangi, this project approaches the text analysis component of this project as qualitative research, epistemologically similar to content analysis and aligned with grounded theory.¹ In using a mixed method approach, and limiting my scope to Texas, I aim to shed light on mass incarceration as issue that needs immediate attention both in Texas and nationally. It is my hope that this research will be reused and built upon by others.

¹ Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between Text Mining and Qualitative Research in the Perspectives of Grounded Theory, Content Analysis, and Reliability. *The Qualitative Report*, 16(3), 730-744. Retrieved from <https://nsuworks.nova.edu/tqr/vol16/iss3/6>