

Case study rubric

Due: TBD

Submission format: Submit the link to your repository on Canvas

Why am I doing this? This study serves as an opportunity for you to practice your skills in text data analysis. Relevant across industry sectors and academia, this case study provides you the skills necessary to tackle big data projects when working with text.

What am I going to do? This study will guide you through the process of giving numerical value to text data. This step will be followed with relevant techniques to extract knowledge from the newly transformed data. You will have the opportunity to showcase your skills in data science through a sociological lens. At the end of the project, you will produce a deliverable outlining countries that are grouped by constitutional similarity. This will help the UVA Department of Sociology structure their new course series on the alignment of countries' policies with their constitutions.

How will I know I will have succeeded? You will meet expectations for the case study when you follow the criteria outlined below:

Category	Details
Formatting	<p>Submit each component listed in the rest of this rubric as advised below.</p> <p>Written Portion</p> <ul style="list-style-type: none">- Submit the written portion as a PDF file. <p>Data & Code</p> <ul style="list-style-type: none">- Submit code created for all portions in a GitHub repository.- Include any additional data that was used in GitHub if necessary.- The GitHub should be titled "CS-[insert first & last name]" <p>References</p> <ul style="list-style-type: none">- References should be included on a separate page at the end of the Written Portion PDF file.- IEEE citation style

Written Portion	<p>Discuss your interpretation of the study and thought process by answering the questions below.</p> <ul style="list-style-type: none"> - In a small paragraph, summarize the problem presented in the study as well as its importance. - In a small paragraph, discuss your plan to meet the demands of the deliverable. - Include a simple graphic that outlines your analysis plan that you will follow for this study. - Finally, discuss the results of your study in complete sentences. <p>Reflection:</p> <ul style="list-style-type: none"> - In a small paragraph, discuss the challenges you met while completing this study and how you overcame them. - Also discuss what you could have done differently and how you plan to improve in the future.
Code	<p>Your code should...</p> <ul style="list-style-type: none"> - Use dataset preambles_data.csv - Located in "DATA" folder on GitHub - An exploratory data analysis with respect to frequency of certain words across constitutions and number of constitutions by continent - Feature extraction of the constitutions into dense vectors - Utilization of the cosine-similarity matrix to build a hierarchical clustering algorithm - Outcome of distinct clusters with word clouds of each - Be sure to include comments so a viewer can understand.
References	<p>At the end of the written portion of this assignment please include a list of references in IEEE citation style that were NOT in the already given reference materials.</p>

