

Midterm Report for DATA 11800 - Spring 2023

(due April 26)

The midterm report will assess data visualization, analysis, and interpretation skills. Using the provided data, students are expected to perform an analysis that produces informative and understandable graphs and tables. This midterm report is not a group project; you may discuss the assignment with other students, but you should code and compose your report independently. Your score will be based on the accompanying rubric uploaded to Canvas. The data that is to be used for this report, along with any relevant codebooks and documentation, will also be uploaded to Canvas.

About the Data

The PRIO Battledeaths Dataset is a dataset on battle deaths (number of soldiers and civilians killed in combat) in state-based armed conflicts for the period 1946–2008. The dataset is compatible with the UCDP/PRIO Armed Conflict Dataset for the period 1946–2008. This dataset was originally published in 2005 and is cited as follows:

Lacina, Bethany & Nils Petter Gleditsch (2005) 'Monitoring trends in global combat: A new dataset of battle deaths', *European Journal of Population* 21(2–3): 145–166.

For academic exchanges on the PRIO data, see the following articles and rebuttals:

Lacina & Gleditsch (2005) 'Monitoring trends in global combat: A new dataset of battle deaths'.

Obermeier et al. (2008) 'Fifty years of violent war deaths from Vietnam to Bosnia'.

Spagat et al. (2009) 'Estimating war deaths: An arena of contestation'.

Gohdes & Price (2012) 'First things first: Assessing data quality before model quality'.

Lacina & Gleditsch (2012) 'The waning of war is real: A response to Gohdes and Price'.

Guidelines

Perform analyses on this dataset (or a subset of this dataset) that will yield the following:

- 5 graphs/plots:
 - must have at least 3 different kinds of visualizations. You may choose from the following discussed in class: histogram, line plot, bar graph, box and whisker plot, pie chart, or scatter plot
 - at least 1 figure must depict multiple variables by either using subplots or plotting multiple variables on the same axes. (See *Data Visualization Part 2* lecture for examples)
- 2 tables
 - tables must provide summative information, must be less than 20 rows displayed
- A new column of data created by you based on data already in the dataset

For each graph and table made, a markdown cell describing the following is required:

- A statement about what the graph/table shows
- A description of the analysis done to produce the graph/table

The markdown cell should be placed immediately below the graph/table output.

Any analysis or processing done to the data must be shown within your report. You are encouraged to use resources online (e.g., StackOverflow, library documentation) or in-person (e.g., TAs, instructor office hours) to brainstorm ways that you may wish to analyze and process the data.

Submission

Your report should be submitted as a Python Noteboook (.ipynb file extension) with your last name and first name in the file name like so: LastnameFirstname_Midterm.ipynb.

To be eligible to receive full credit, all code outputs should be shown and all graphs should be properly annotated and constructed. Upload your Python Notebook to Gradescope by April 26th at 5:00 pm CST.