Explore the relationship between victim and police officer demographics in the likelihood to seek or offer medical assistance in use of force incidents.

Relational Analytics: What is the average value of X?   How many distinct values of Y are in the data set?   What groups of individuals are most affected by Z?  Propose at least four questions.

1. What percentage of use of force incidents result in injury for citizens? Police officers? Broken down by race, age, and other demographics? Broken down by neighborhood (or any other surrogates for SES)? Broken down by type of use of force (physical, taser, firearm, etc.)?
2. Difference in percentage of different types of uses of force, taser vs. physical vs. firearms, broken down by race and likelihood to result in medical attention.

Visualization:  What are some relationships in the data that you'd like to explore?  A UI of the data is [here (Links to an external site.)](https://data.cpdp.co/data/Lovo7d/citizens-police-data-project) to familiarize you with its contents. Tableau, the first visualization tool with which we will work, offers these [chart types (Links to an external site.)](https://towardsdatascience.com/the-ultimate-cheat-sheet-on-tableau-charts-642bca94dde5).   Propose at least two visualizations.

1. Stacked bar chart showing use of force incidents by race and percentage that result in calling emergency medical attention
2. Symbol/bubble map – assess geographic distribution for percentage of use of force incidents that result in medical care – compare to proximity of medical centers or trauma centers

Interactive visualization: Sometimes it is easier to explore the data when we can examine more than two dimensions at a time.   Interactive visualization examples are [here (Links to an external site.)](https://github.com/d3/d3/wiki/Gallery).   Propose at least 2 interactive visualizations.

1. Expanding Treemap, circle packing graph showing nested racial groups, use of force encounters, type of use of force, injuries sustained

Machine learning:Identify a couple of things in keeping with your theme that you would like to predict from the CPDB.   Describe the fields you think will be useful for making your predictions.  Propose two things you are interested in modeling or predicting.

1. Predict police encounters (based on citizen and police demographics, alleged crime, neighborhood) where use of force incidents are less likely to result in injury.
2. Predict police encounters more likely to result in lethal uses of force vs. less-than-lethal uses of force.
3. Predict police encounters (based on citizen and police demographics, alleged crime, neighborhood) where use of force incidents are less likely to result in appropriate medical attention.

Natural Language Processing:  This dataset has a well-known disconnect in that many of the complaint reports filed by civilians describe an incident of one type, but the report is classified as something different because the reports are generated by the officer taking their statement.  For example, if someone files a complaint saying that the cops planted drugs on them, a complaint report may list it as an illegal search [[link] (Links to an external site.)](https://cpdp.co/complaint/1044999/).   For our last checkpoint, we will be building models over the freeform text (e.g. complaint report narratives) in the CPDB.   Many of these are available in CPDB - the database you will load for Checkpoint 1 - in the data\_allegation table in the column summary.   You may find additional narratives that are still in process [here (Links to an external site.)](https://github.com/invinst/cpdp_ocr/blob/master/6_reports/output/narratives.csv).   New ones are coming in regularly owing to the recent [Green Ruling (Links to an external site.)](https://www.chicagoreporter.com/charles-green-says-his-fight-for-decades-of-chicago-police-misconduct-records-is-about-clearing-his-name/).

Propose one model NLP model for this.  An example of this is identifying narratives pertaining to your theme and/or adjacent themes using topic modeling.  We recommend checking out the Invisible Institute's current work on this for advice on starting points [[link].  (Links to an external site.)](https://github.com/invinst/documentAnalysis) If we are effective with this assignment, it will open up a new window on the complaint reports

1. NLP to identify encounters that result in emergency medical care and if possible mode and outcome of that care – EMS (ambulance), hospital admission, emergency room; time/date of emergency medical services.

Throughout this course, you will learn the skills you need to answer these questions and you will incrementally work on them with checkpoints throughout the quarter.