1. State your question. What is it that you are curious about? What are you looking for in the data?

   The Chicago police department has a long history of questionable policing tactics; there is currently a federal lawsuit filed by the ACLU that claims differential responses in the form of response time based on geographic location. Unfortunately, the response times are not easily obtainable from the department. There are however other data available from the department to address response questions.

   Are there patterns in Chicago crime (arrests, reports of domestic violence) that indicate differential response to crime reports? Are crime/arrests/DV more likely to occur on particular day of week, time of year, time of day, or location? Are rates as we predict with a predictive model?

2. Demonstrate that you have looked at your data. What are your columns? Are they numerical/catagorical? How many Nan's are there? Have you made a couple of plots?

   The Chicago crime dataset is available on a public portal and contains all crimes except for murders. Data are uptodate within 7 days. The columns of interest include date/time of occurrence, beat, ward, community area, primary type of offense, offense description, location type, whether there was an arrest, whether the offense was DV, and latitude/longitude of occurrence.

3. State your MVP. MVP is your Minimum Viable Product. What's the minimum that you hope to accomplish? Then, feel free to expand on MVP+, and MVP++.

   There are three objectives I'd like to accomplish.

   A. The data are well suited for visualizations, so I want to focus on graphing/mapping to describe the temporal and geographical patterns.

   B. In unsupervised learning, I'll be performing data reduction techniques, such as LDA or NMF on the dataset to identify a reduced set of features to work with in my third objective.

   C. In supervised learning, I'll build a predictive model to predict arrest rates for various types of crime, and assess whether rates are as predicted.