

# NCSU Datathon

---

Vidusha Rao and Michael McKibben

March 20, 2021

North Carolina State University

# Approach

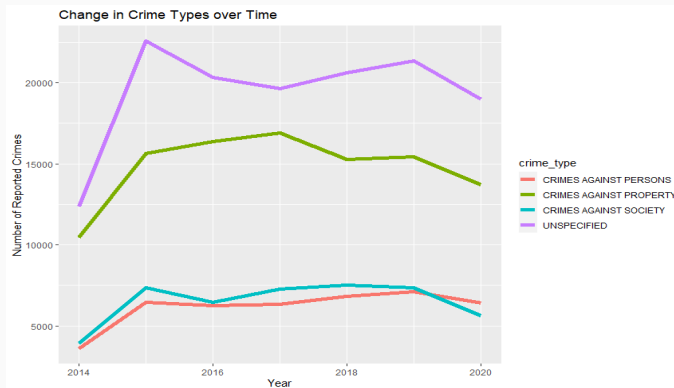
While the police data is incredibly rich, visualizations make data much easier to understand.

In particular, we want to provide decision makers with an easier way to answer the 5 W's and H question.

Specifically, we'll look at answering When, Where, and What.

# Looking at the data

We have an extremely rich dataset. As the prompt is open-ended, we'll start by exploring the dataset.

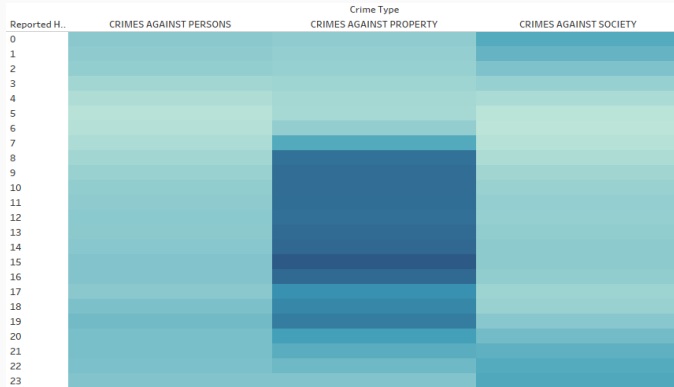


**Figure 1:** Crime types over year

- The number of crimes is relatively steady and slightly declines from 2015 onward. The largest amount of crimes is "Unspecified."
- Unspecified may include crimes which are hard to classify, such as a crime described as "Humane/Chemical Deployment."
- Some of this is also due to inconsistencies in the data, i.e. property damage that is not classified as "Crimes Against Property."

# When?

A useful descriptor would look at the 5 W's and H- Who, what, when, where, why, and How? This is in the police data, but it's not easy to see in that format. This more readily communicates when:



**Figure 2:** Crime density by hour reported

# When?

Of particular interest:

1. Crimes against property seem to be particularly dense around 3pm, right after school gets out.
2. There is a secondary spike around 7pm as well.
3. Crimes against society are generally in the late night/early morning hours.

The data is slightly limited in that the time of the incident being recorded may not match the time of the crime. However, if there is a density of reported crimes, it is more easy to target interventions.

# What? Where?

Cluster analysis is often used to analyze latent clusters. For example, in a business, cluster analysis is often used to identify who is actually buying or utilizing a product.



**Figure 3:** Gower Distance Clustering

# What? Where?

Description of our clusters