Data Analysis Project: Status Memo #1  
Oct. 29, 2018

**Overview:**

For this project, I will be analyzing the Virginia Court Data database. My primary interest is in the differences in sentencing between those who plead guilty and those who choose to go to trial, either by judge or jury.

I hope to answer the following questions:

1. Is there a sentencing penalty for those who go to trial (a.k.a. “trial penalty”)? In other words, do people who choose to go to trial instead of taking a plea deal generally end up being sentenced to more time than those who do not?
2. Which crimes have a trial penalty?
3. Is there a stronger trial penalty for people based on race? Gender?
4. How many individual crimes are people usually charged with? Are a large number of those charges dropped when someone takes a plea deal, compared to when they do not?

This is of interest to my readers because a demonstrable trial penalty shows that people who try to use our system of justice as intended are being punished for doing so.

**Process:**

The first thing I had to figure out was why there were so many NULL values in the plea field (90% of all values). I ran a number of queries (attached), but eventually a legal expert explained that in many cases, the prosecutor dismisses the charge before a plea is entered.

There are [XX] tables in the database. After some preliminary work (attached), I determined that the main *CircuitCriminalCase* table is the most useful for this analysis. The *ConcludedBy* field shows whether a case is determined by a guilty plea or a trial.

I next pulled 10 years of data, 2007-2017, out of SQL so I could clean it in Excel. *CodeSection*, the field storing the crime code, will be the most useful in determining which crimes have a penalty. Unfortunately, this field is very “dirty,” in that there are typos and inconsistencies in formatting of the data. After roughly two(?) weeks of cleaning using a combination of hand cleaning, Excel formulas and OpenRefine.org, the field is usable for analysis.

After a preliminary analysis, I discovered a trial penalty on drug use, among other things. However, after making some adjustments, it disappeared. More analysis is required to determine what happened. Also, my preliminary analysis ignores charges that were entered as a group of two. I am still determining how to handle these charges.

Next steps will be to finalize my analysis of the data as grouped by crime type, and later to group by demographic groups such as age, race and gender to determine what other factors influence trial penalty. I will use an ANOVA analysis in R to determine statistical relevancy.

I will also need to count the number of charges per person and figure out a way to determine which were dropped in cases of a plea deal as compared to a trial.

Please see my [process log](https://github.com/shardsofblue/va-court-analysis/blob/master/process-notes.md) on GitHub for complete and ongoing documentation of my progress.

*Memo requirements:*

* *the questions you hope to answer with the database.*
* *why this would be of interest to an audience of readers, viewers or listeners.*
* *what you have discovered in your initial analysis, and what you plan to do to complete the assignment.*
* *also attach a well-commented SQL file with at least five queries you have written as you begin to explore the data.*