**What is the question you hope to answer?**

What is the likelihood of a successful cyber attack on US cars?

**What data are you planning to use to answer that question?**

* Proxy of survey data for successful cyber-attack comparison (see pg. 18)
  + <http://www-935.ibm.com/services/multimedia/SEL03027USEN_Poneman_2014_Cost_of_Data_Breach_Study.pdf>
* Csv file of all data breaches
  + <https://www.privacyrights.org/data-breach>
  + <http://www.privacyrights.org/data-breach/new>
* Other cyber attack website references
  + <http://hackmageddon.com/2012-cyber-attacks-statistics-master-index>
  + <http://www.heritage.org/research/reports/2014/10/cyber-attacks-on-us-companies-in-2014>
  + <http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks>
* Website for scoping/norming total number of businesses
  + <http://www.census.gov/data.html>
* Websites for vehicle data
  + <http://catalog.data.gov/dataset/highway-statistics-data-browser>
  + <http://catalog.data.gov/dataset/deleted-www-fueleconomy-gov>
  + <http://catalog.data.gov/dataset/national-automotive-sampling-system-crashworthiness-data-system-nass-cds-nass-cds-multiyea>
  + <http://catalog.data.gov/dataset/vehicle-snowmobile-and-boat-registrations>
  + <http://catalog.data.gov/dataset/manufacturer-registration-submissions-and-manufacturer-vin-decoding-submissions-mid-ui>
  + <http://catalog.data.gov/dataset/motor-carrier-registrations-census-file>

**What do you know about the data so far?**

My biggest issue doesn’t seem to be the structuring of the data so far, it is more in the shaping of the hypothesis and which sources this data will be pulled from. I know the structured elements of the data columns:Date Made Public, Name, Entity, Type, Street Address, Location, Postal Code, Total Records, and Records Breached

**Why did you choose this topic?**

I work as a cyber-security consultant, I have been interested if and where analytics crosses with cyber security. This is something we want to try and predict to determine what kind of problem the market will face because the 5 year trend looks to have interconnected cars, either car-car or car-object (i.e. city, signs, etc.)