**Proposed Project:**

**Jeff Brown**

**Topic:** Regional variations in car types for sale. Why am I interested in this – my son just turned 16 so we were looking at cars on craiglist. Started looking at the HTML behind the scenes and noticed could extract this information. Made we wonder if types of cars varied by regions (are subarus more popular in Minnesota than Florida?).

**Information to Extract and Source**

**Locations:** State Capitals, taken from <https://www.nationsonline.org/oneworld/capitals-usa.htm> pulled as html table using pandas and converted to dataframe.

**Population of Location:** Also provided in table from: <https://www.nationsonline.org/oneworld/capitals-usa.htm>. Time permits would prefer to get from Census (see below in additional time).

**Car Types for Sale:** Using Craigslist will find the total number of cars for sale in a given city. This is available using a search query for all car types:

<https://minneapolis.craigslist.org/d/cars-trucks/search/cta>. Site gives total in HTML as

<span class="totalcount">3000</span>

Then search for specific car manufacturers rather than all cars: Ford, Dodge/Chrysler, Chevrolet/Chevy, Honda, Toyota, Subaru, Mercedes, BMW. Example for Subaru:

<https://minneapolis.craigslist.org/search/cta?query=Subaru>.

<span class="totalcount">522</span>

Query structure is easy to implement – combine basic query with city drawn from a list.

**Database to create:**

PostgresSQL with two tables: 1) City information (ID, City, State, Population); 2) Car Breakdown (ID, City, #Ford, #Dodge, #Chev, #Honda, #Toyota, #Subaru, #Mercedes, #BMW.

**Additional (Time permits)**

Query Census with API call to get population and average income for each city. Data would be added to the City Information Table.