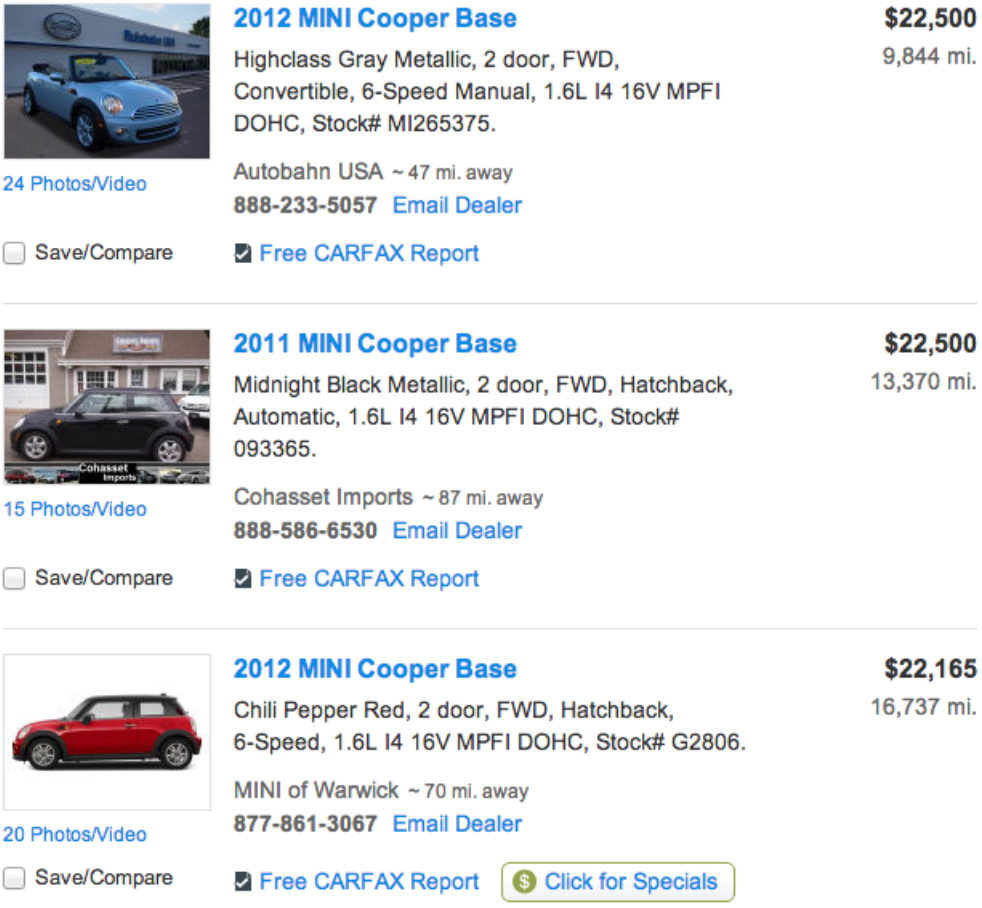
Cars Car prices vary. They vary according to the model of car, the optional features in the car, the geographical location, and the respective bargaining abilities of the buyer and the seller.

In this activity, you will work in groups of two to investigate the influence of a variety of factors on the asking price for at least 25 USED (not new!) cars using data collected from `cars.com`. Record these in a spreadsheet (e.g., Google Sheets or Microsoft Excel) with six variables:

- `car`: will always be `Toyota Prius`, you need to specify the city under "Advanced Search"
- `year`: Model year (restrict your scraping to used vehicles with model years between 2011 and 2017)
- `price`: Price of the car provided as a number of dollars (no dollar sign, commas, etc.)
- `model`: Model will always be Prius, the model name will typically be of the form `none`, `c`, `plug-in`, `v Two`, `v Three`, `v Four`, `v Five` or similar.
- `mileage`: Mileage (in miles)
- `location`: Assigned geographical location (name: you'll enter the zip code for your city when you search cars.com)

I have provided a template in `cars.csv`. Recording this information in the spreadsheet will help in combining data for different types of Toyota Prius's. You will also need to identify a Zip (postal) code for your specified city.

Here's an example for the Mini Coopers (note that you'll be scraping data on Toyota Prius's):



```
ds

##           car model price year mileage location
## 1 Mini Cooper  Base 22500 2012    9844  Amherst
## 2 Mini Cooper  Base 22500 2011   13370  Amherst
## 3 Mini Cooper  Base 22165 2012   16737  Amherst
```

1. Save your spreadsheet as a `.csv` file (I have provided a template as `cars.csv`), then upload it to the RStudio server using the `Upload` button on the `Files` tab. Render the file and review the results: do the summary statistics make sense? Are there any missing values? Please provide a terse summary of the distribution of two of your variables.

2. Generate a plot of your data that represent an interesting relationship between three variables. There are some examples with two variables to get you started in the Quarto template `cars.qmd`.

3. Using the data that you scraped, build a simple multiple regression model for the patterns seen in used-car prices. Some basic bivariate claims that you might build from could be described as:

   (a) Looking just at price versus mileage, the price of car model XXX falls by 12 cents per kilometer driven.

   (b) Looking just at price versus age, the price of car model XXX falls by 1000 dollars per year of age driven.

4. Note whether there are outliers in your data and indicate whether these are having a strong influence on the coefficients you find. (If you note errors in your coding of your data please fix them in the csv file and re-upload the data.)

You will submit three deliverables for this project (1) your dataset (as a .csv file: please double check format), (2)formatted Quarto file from your analyses (pdf) and (3) the source (Quarto) file, which includes:

1. a brief summary of what you collected (an overview of what you collected: price, model, mileage and year from n=XX used Toyota Prius's for sale within YY miles of ZZ (your city and zip code),

2. a summary of the distribution of two of the variables that you collected (for categorical variables counts and proportions; for continuous the usual summary of shape, center, and spread),

3. an interesting plot (that uses at least three of the variables in the dataset), along with your interpretation, plus

4. results from a single multiple regression model (with at least two predictors), along with interpretation of the coefficients from that model.

Please be sure to upload these files to Moodle (see "cars submission") as a group submission (with both names specified) by the end of the day on Thursday, February 27th.