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.

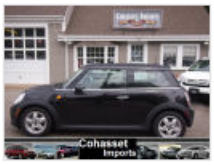
24 Photos/Video

2012 MINI Cooper Base **$22,500**

Highclass Gray Metallic, 2 door, FWD, Convertible, 6-Speed Manual, 1.6L I4 16V MPFI DOHC, Stock# MI265375. 9,844 mi.

Autobahn USA ~ 47 mi. away
888-233-5057 Email Dealer

Save/Compare ☑ Free CARFAX Report



15 Photos/Video

2011 MINI Cooper Base **$22,500**

Midnight Black Metallic, 2 door, FWD, Hatchback, Automatic, 1.6L I4 16V MPFI DOHC, Stock# 093365. 13,370 mi.

Cohasset Imports ~ 87 mi. away
888-586-6530 Email Dealer

Save/Compare ☑ Free CARFAX Report

20 Photos/Video

2012 MINI Cooper Base **$22,165**

Chili Pepper Red, 2 door, FWD, Hatchback, 6-Speed, 1.6L I4 16V MPFI DOHC, Stock# G2806. 16,737 mi.

MINI of Warwick ~ 70 mi. away
877-861-3067 Email Dealer

Save/Compare ☑ Free CARFAX Report $ Click for Specials

In this activity, in groups of 2, you are going to investigate the influence of a variety of factors on the asking price for used cars:

- Model year
- Model
- Mileage
- Geographical location

1. For your assigned location, use the web site to find prices and the other variables for used Toyota Prius's (you can specify the city under "Advanced Search"). I want you to collect data for at least 40 cars, using data from `cars.com`. Record these in a spreadsheet (e.g., Microsoft Excel) with six variables: `car` (which will always be `Toyota Prius`), `model`, `price`, `year`, `mileage`, and `location` (your assigned city).

   The model will always be `Toyota Prius`. The model name will typically be of the form `none`, `c`, `plug-in`, `v Two`, `v Three`, `v Four`, `v Five` or similar.

   I would like you to restrict your scraping to used vehicles with model years between 2011 and 2016.

   Recording this information in the spreadsheet will help in combining data for different types of Toyota Prius's.

It's important to ensure that all of your variables (in particular `price`) are provided as numbers (no dollar sign, commas, etc) and that you use the exact names given above. Here's an example for the Mini's from above (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
```

2. Save your spreadsheet as a `.csv` file, then upload it to RStudio.
3. Generate two interesting plots of your data that represent an interesting relationship between two variables. There are some examples to get you started in the R Markdown template.
4. Using the data, build a simple regression model for the patterns seen in used-car prices. Some basic claims that you might make are in this form:
   (a) Looking just at price versus mileage, the price of car model XXX falls by 12 cents per mile driven.
   (b) Looking just at price versus age, the price of car model XXX falls by 1000 dollars per year of age driven.
5. Note whether there are outliers in your data and indicate whether these are having a strong influence on the coefficients you find.

(Hint: you may want to set the option: `options(scipen=5)` to avoid scientific notation for some of your coefficients.)

Your deliverable for this project will be your dataset (as a .csv file: please double check format), plus the formatted Markdown file from your analyses, which includes:

- a summary of what you collected (an overview of what you collected: price, model, mileage and year from n=XX used Toyota Corollas for sale within YY miles of ZZ (your city),
- a summary of the distribution of two of the variables that you collected (for categorical variables counts and proportions; for continuous the standard summary of shape, center, and spread),
- two interesting plots, along with your interpretation, plus
- results from a single model, along with interpretation of the coefficients from that model.

These two files should be emailed to `nhorton@amherst.edu` before class on Thursday, February 23rd and cc'ed to your project member using subject line "STAT135: cars". Please be sure that your csv file has the *exact* format given above (same names, same order).