MA 200 *Descriptive Statistics Project Handout 2: Car Report Introduction*

1. **What is the idea of the car reports?**

At this point in the course you should be able to intelligently interpret data. The car report gives you the opportunity to:

- collect some data,
- analyze these data using Fathom (or Excel – but Fathom is better)
- interpret the results of your analysis
- write up your report in an informative and attractive fashion.

2. **What are the data? And where do we get the data?**

The data are the advertisements for used cars posted on the Internet in places such as www.cars.com, although you could get the data from newspapers or other sources.

Here is what the data look like when they are entered into Fathom

Ford Explorer 050916 Boston.jsp

	Year	Vehicle	Price	Mileage	Seller	Body	Color	Distance
1	2005	FordExplÉ	$28,995	--	Metro Honda 8É	SUV	Gray	44 mi.
2	2004	FordExplÉ	$27,300	34643	York Ford 866-É	SUV	White	7 mi.
3	2004	FordExplÉ	$25,381	17552	York Ford 866-É	SUV	White	7 mi.
4	2004	FordExplÉ	$24,995	14794	Watertown Ford É	SUV	Fire Red É	8 mi.
5	2005	FordExplÉ	$23,980	15178	Framingham Ford É	SUV	Blue	21 mi.
6	2004	FordExplÉ	$23,625	25789	Herb Chambers HoÉ	SUV	ObsidianÉ	40 mi.

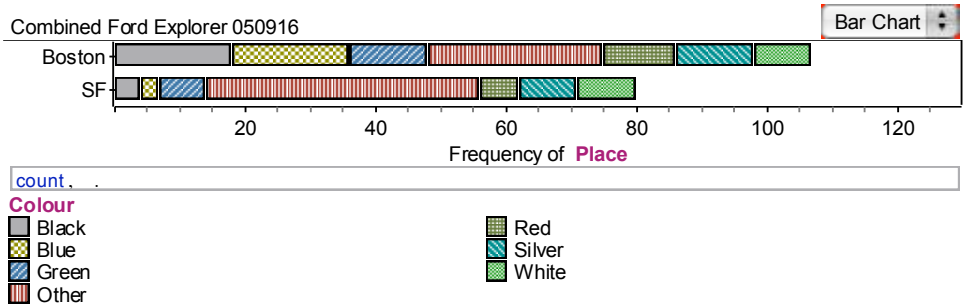
3. **What kinds of analyses will we do?**

You will be comparing the *markets* for used cars either between two places, or between two different cars. Here are some examples:

- Here we have chosen one car, Ford Explorer (well, actually, SUV) and looked at all of Ford Explorers for sale in the SF bay Area compared with all of the Ford Explorers for sale in the Boston area. (This is easy to do using www.cars.com.) The plot compares the ages of the cars in the two places:

Does it appear that on average, the Ford Explorers in the bay Area are a bit older than the ones being sold in Boston? The boxplots seem to say that. What about the spread? Are the IQRs different? What about the standard deviations?

- Are the cars for sale in Boston about the same in the kinds of colors, or are they different? Here is a ribbon plot, and a percentage table.



count , .

Colour
Black
Blue
Green
Other
Red
Silver
White

- In this case, since there are so many colors, the numbers are more informative.

Combined Ford Explorer 050916

		Colour							Row Summary
		Black	Blue	Green	Other	Red	Silver	White	
Place	Boston	18 0.1682243	18 0.1682243	12 0.11214953	27 0.25233645	11 0.10280374	12 0.11214953	9 0.08411215	107 1
	SF	4 0.05	3 0.0375	7 0.0875	42 0.525	6 0.075	9 0.1125	9 0.1125	80 1
Column Summary		22 0.11764706	21 0.11229947	19 0.10160428	69 0.36898396	17 0.090909091	21 0.11229947	18 0.096256684	187 1

S1 = count, .
S2 = rowProportion

- Here is another analysis. This time, we are comparing Toyota 4Runners and Ford Explorers but in just one place---the SF Bay Area.

- It looks like the Toyotas are more expensive, on average. But, is that necessarily so? Suppose the Fords for sale are on average older than the Toyotas? Then, overall, they would be cheaper, but maybe because as a whole, the SUVs being sold are older. Look at the distribution of ages as well.

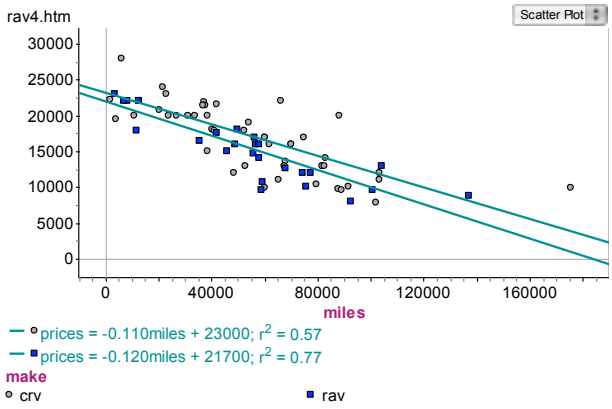
- Here are the numerical summaries for the variable age. What do they tell you?

Combined Explor and 4Run SÉ

		Age
SUV	Ford	5.825 3.7069043 80
	Toyota	5.5 3.7183447 93
Column Summary		5.650289 3.7058192 173

S1 = mean, .
S2 = s, .
S3 = count, .

- Below are regression lines for the price vs. mileage of the Honda CR-V and Toyota RAV4 in the same geographic region. What do the regression lines tell you?



.

4. **But what analyses do you want?**

- The actual choice of analyses is up to you. Part of the idea of this project is that it lets your team take some initiative in producing a quality piece of work.

- Having said that, we will be looking for <u>at least one comparison of quantitative variables</u> (like the example above of the age distributions between Boston and SF), <u>at least one between two categorical variables</u>, like the analysis of colors, and <u>at least one regression analysis</u> (like price v mileage for the RAV4 and Toyota CRV).

- You should first of all do some one variable analyses to look for errors in the data, and to get a feel for your data

- Mostly, we will be looking at what you say about the graphical and especially the numerical summaries that Fathom produces. Does what you write make good statistical sense? Is it clear? Is it true to what the data say or do not say?

5. **How many cars do we need to look at?**

   - Answer: all of them! That is, if you choose to compare Honda Civics between SF bay and Seattle, then you must get all the data from the two places. Remember, you have software: you do not have to do all the calculations yourself, and your software actually enjoys large amounts of data.

   - You could make a bad choice of cars that would give you too few. Stay away from analyzing Ferraris and Plymouths. But these are details: consult with your instructors. As a rule of thumb, if your search yields less than 25 cars or an age -spread of less than 10 years, try either a different vehicle or market.

6. **How many pages does the report have to be?**

   - The number of pages depends on how thorough you are, how wordy you are, what you do with graphics and Fathom output. Even the font size may control the number of pages. You probably will not be able to do everything in less than two pages, and thirty is probably overdoing it.

7. **Do we have to do this as a team? And could our team do something else, since we hate cars?**

   - Answer: yes! And yes!

   ♥ We know that it makes life harder for some people, and that you fear doing the work of freeloaders.

   - One of the reasons for having the work done in teams is that the marking load would be too heavy for your instructors otherwise.

   - But there are other reasons: we think that it is a good experience for people to work in teams. Employers say that it is characteristic of the work place, so students should also have this experience. We know that it is harder, but it is closer to real life, unless your future is that of a hermit poet, or hermit geek.

   - We have some policies that can be instituted to curb freeloading.

   - Yes, your team may do a different kind of project but that project should also involve the collection and analysis of data and the report should meet all the criteria outlined above.

8. **Will we be able to get more detailed help?**

   - Yes, this is just the introduction. The next one will be about data collection.

9. **When is it due?**

   - There are several due dates, and a space here to write them: they are listed on the First Handout.