A Honda Civic Comparison

San Francisco
V.
Phoenix

By L C and another CSM Student
Spring 2006

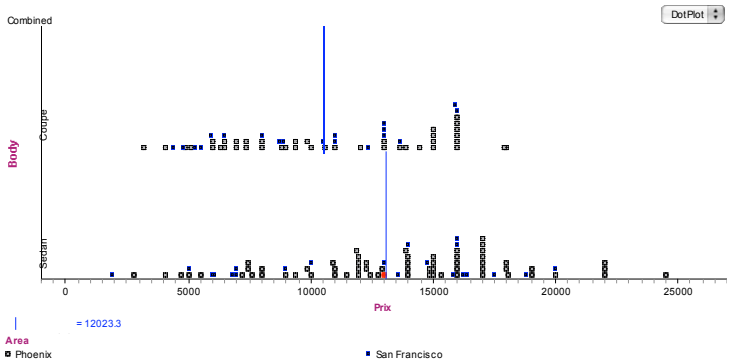
Table of Contents:

In the following report, we compare the sale of used Honda Civics within a fifty-mile radius of both San Francisco, California, and Phoenix Arizona. We will attempt to root out any significant relationship or differences that may occur between the sales of Honda Civics in San Francisco Ca. with those being sold in Phoenix Az., with the caveat that there may be none at all. We will attempt to discover which city is more apt to buy the car, which color and body style is most prevalent in each city, and what are the possible underlying reasons affecting these decisions.

We have taken a sampling of over one-hundred and eighty cars, which range in price, age, mileage, color, body style, and distance from our base zip codes of 94124 for San Francisco, and 85001 for Phoenix. We have downloaded a list of 181 cars from cars.com, and based on that information we have compiled some statistical data about different categories related to the sales of the Honda Civic in each city. After a careful comparison of categories for the Honda Civic; such as the relationship between body style and prix (the French word for price); prix and area (San Francisco, and Phoenix); prix and colour (the French word for Color) of the cars; body style and area; and, mileage and prix, we will try to determine what differences if any there are in the regards to sales of Honda Civics between these two cities.

smcccd 4/4/06 9:34 AM
Comment: Well, no, actually. Colour is the British spelling for color. En français, le mot 'color' en anglais, est couleur.

We start by first comparing a quantitative variable (a variable that "measures a numerical characteristic") and a categorical variable (one that "simply records a category") On the following page we have a Dot Plot displaying a comparison between prix(price) to body style, and the resulting frequency of each.



For the Honda Coupe, the Dot Plot shows a distribution skewed left , with one distinct peak occurring between $14000 and $16000 of which there are 13 cars, representing 21.6% of the coupes being offered for sale. There is another cluster between $6000 and $8000, of which 11 cars are being offered, this represents about 18.3% of the coupes being sold. Together these two price ranges represent about 40% of all coupes being offered. There are also two data points which might appear to be outliers: one of them is a 1995 Honda Civic Coupe priced at $3200 from the Phoenix area, and falls -1.784 standard deviations from the mean. The other subinterval in question occurs between $16000 and $18000, here there two cars; a 2004 and a 2005 Honda Coupe, which fall 2.07 standard deviations from the mean. The mound shape of the coupes distribution confirms the empirical rule that 68% of the coupes fall within one standard deviation, while 96.8% of the coupes fall within two standard deviations, and 100% of the observational units fall within three standard deviations of the mean.

smcccd 4/2/06 7:56 PM
Deleted: distribution shape

smcccd 4/2/06 8:06 PM
Comment: There needs to be some qualification here. The empirical rule is based upon the Normal Model, and that model is symmetrical. Hence, if a distribution based upon data is left skewed, then it for that reason departs from the "Empirical Rule/Normal Model". It may still be in accord with the 68% 95% criteria, but if the distribution is left skewed then to that extent it does not follow the Normal Model. In any case it would have been wise to tell the reader (who is supposed to be ignornant of all of these things) what you were doing.

The final body style is the Honda Sedan, there are 96 sedans being sold. This seems to be the most popular style. Of the 181 observations we recorded, minus the 17 cars that did not report body style, the sedan is by far the most popular body style, which accounts for 58.5% of all the Honda Civics being sold in both San Francisco and Phoenix The shape of the sedans Dot Plot is fairly symmetrical with a slight skew to the left. There is one distinct peak occurring between $14000 and $16000, this peak occurs at the same subinterval as the coupes. The sedans along with the coupes and the hatchbacks show a steady granularity, with each subinterval occurring every $2000. The sedans also appear to have two outliers. The lowest observational unit is priced at $1900 this is -2.319 standard deviations from the mean; while the highest observational unit is priced at $24495 or 2.387 standard deviations from the mean. So although they appear as if they might be outliers they fall with in the parameters of the empirical rule.

By comparing the summary tables on the next page, for Age/Prix; Prix/Area; and, Age/Area, there seems to be very little difference between the two cities. For instance the mean prix of the Honda Civics in Phoenix is $12,358.46 while the mean prix for San Francisco is $10,598.81, we feel one reason for the slight disparity is that the cars in Phoenix are slightly newer and thus more expensive, even though the mean mileage for Phoenix is slightly higher; What we did find puzzling was that the standard deviation for miles driven was almost three times greater for Phoenix than it was for San Francisco. We feel this is attributed to the fact that San Francisco which is known as the "world's biggest little city" basically means that people who commute in the Area, travel less distance between destination points than those in Phoenix which is sprawling metropolis where people have to t ravel greater distances between destinations.

smcccd 4/2/06 8:11 PM
Comment: Your summary table is giving you the stats for age for the two places combined. That particular summary table is not very informative. The others are much more useful.

smcccd 4/2/06 8:13 PM
Comment: Actually, you could easily have reverted to using English rather than French. The only reason for the French is that Fathom will not allow you to translate a variable to a variable with the same name.

smcccd 4/2/06 8:07 PM
Comment: Nice analysis.

smcccd 4/2/06 8:07 PM
Deleted: Std

smcccd 4/2/06 8:09 PM
Comment: However, given that you had a radius around SF, you would include all of those people who commute and therefore rack up miles on their cars; this blunts your interpretation somewhat.

Combined

	Prix
Age	-0.8458503181 4.1767956 2.9470725 0.21905417 0

S1 = correlation ( )
S2 = count ( )
S3 = mean ( )
S4 = stdDev ( )
S5 = stdError ( )
S6 = count (missing ( ))

Combined

		Prix
Area	Phoenix	12358.462119
	San Francisco	10598.81844
Column Summary		11883.466163

S1 =
S2 =

Combined

		Age
Area	Phoenix	4.0273438128
	San Francisco	4.537735853
Column Summary		4.1767956181

S1 =
S2 =

Combined

	Area		Row Summary
	Phoenix	San Francisco	
miles	128	53	181
	66561.164	65731.509	66318.227
	91936.537	32795.627	79211.393
	8126.1186	4504.8258	5887.7365
	0	0	0

S1 =
S2 =
S3 =
S4 =
S5 =

We see by looking at the Regression line formula, Prix=-1363Age+17900; r^2= 0.70 for Phoenix, and, -1310Age+16840; r^2= 0.75 for San Francisco that the linear models fitted to the data are almost identical. By a linear model we mean that we are trying to summarize the relationship between the age of the cars being sold and the prices they fetch with a straight line of the form $y=mx+b$, where the m is the slope of the line. The slope for Phoenix is slightly higher then that of San Francisco; this shows that for each year a civic ages in phoenix it loses $1363 in value as compared to san Francisco where the value drops $1310 in value.

smcccd 4/2/06 8:14 PM
Comment: It would have been well to have the standard deviation for both of these summary tables.

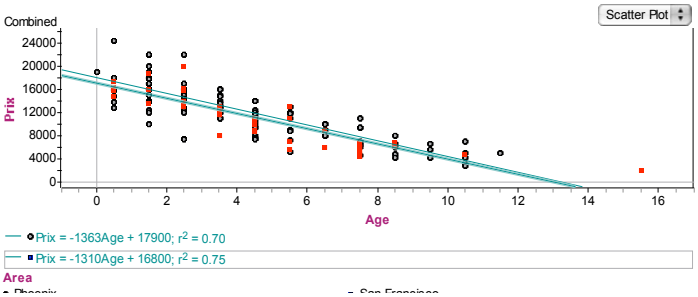
smcccd 4/2/06 8:19 PM
Comment: This paragraph needs some introduction. It is understandable to me, but suppose someone from one of your other classes were to read that. Would that person understand? I doubt it. You need to explain what a regression line is, at least in rudimentary terms. See my insertion.

smcccd 4/2/06 8:16 PM
Deleted: ,

smcccd 4/2/06 8:17 PM
Deleted: they

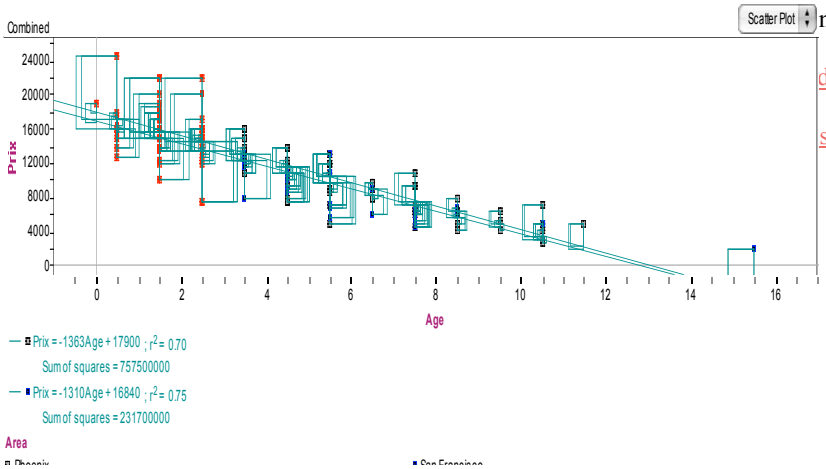
smcccd 4/2/06 8:18 PM
Formatted: Font:Italic

smcccd 4/2/06 8:19 PM
Deleted: it



smcccd 4/2/06 8:20 PM
Deleted: line

There is a very negligible difference, which could be attributed to the extremely hot weather of Phoenix, but more than likely it can be explained by a sampling variation, and if we were to collect a new set of data, we could very well see a difference in the other



n other

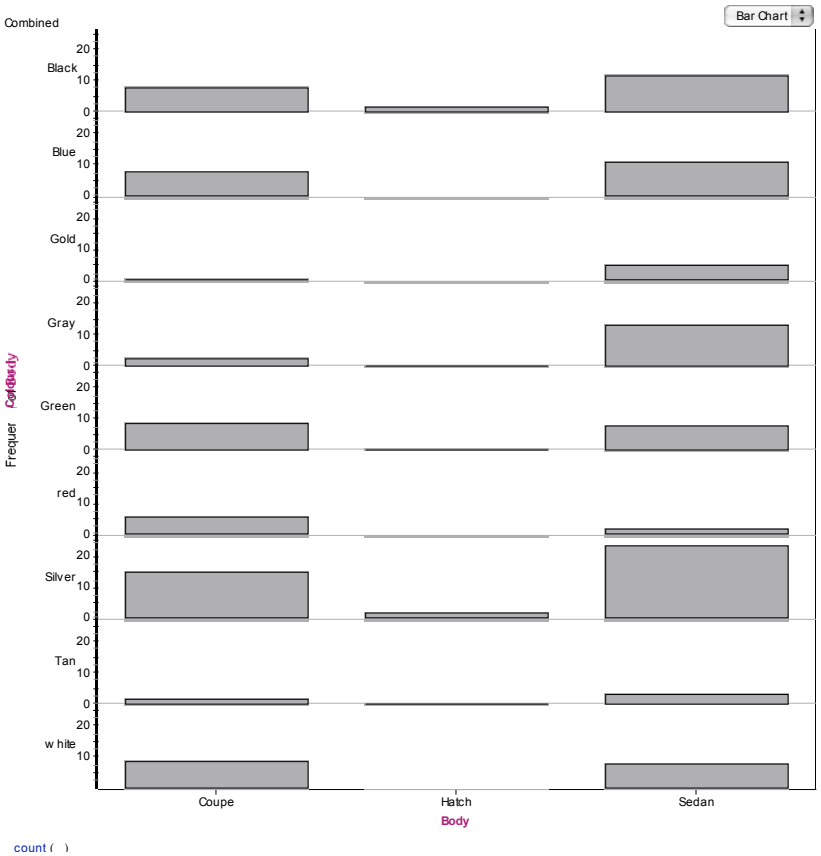
del, whereas

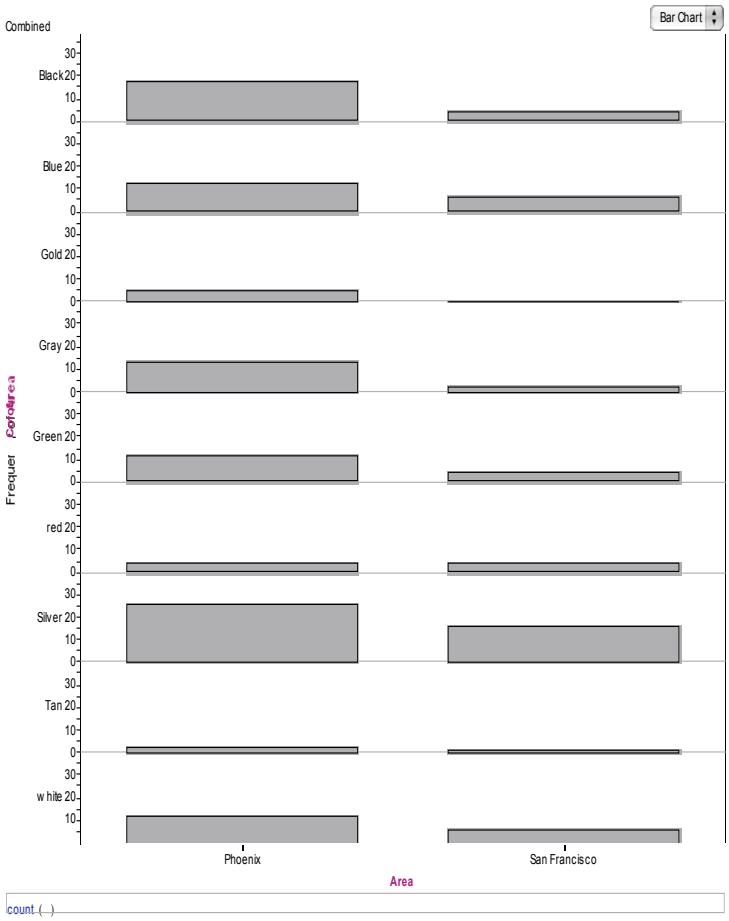
ssion model.

In the next few pages we examine the relationship between the various colors for the Honda Civic and the various body styles, area, and price to see if there is any pattern of differences or similarities between the two cities, and if so, what might explain these variances.

smcccd 4/2/06 8:22 PM
Comment: Your fellow student readers who have not had stats (even some who have!!) will wonder what all the squares are. You did not explain.

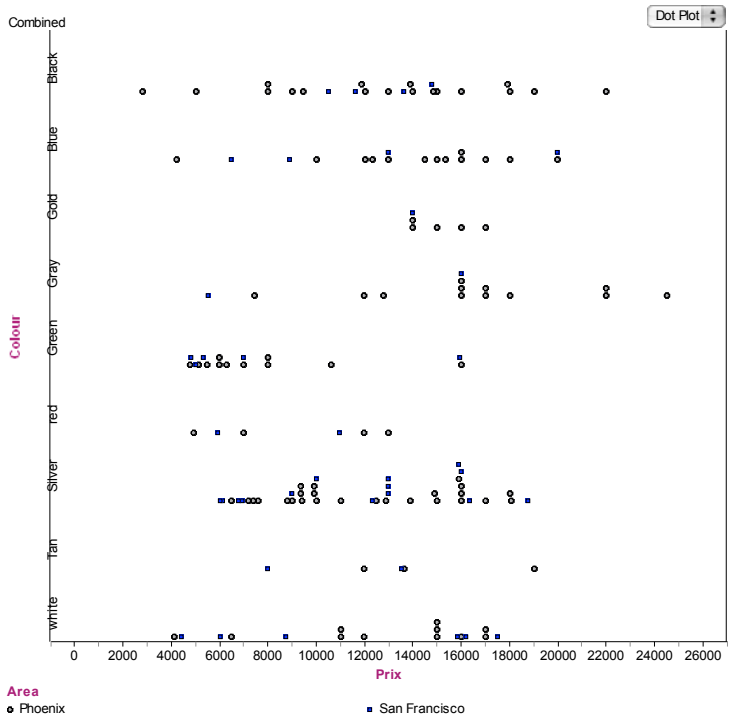
smcccd 4/2/06 8:24 PM
Comment: Nice lead in. I think that I messed up your pagination with my comments, though.


Combined
Bar Chart
Black
Blue
Gold
Gray
Green
red
Silver
Tan
w hite
Coupe
Hatch
Sedan
Body
count ( )



smcccd 4/4/06 9:37 AM

Comment: This graphical analysis should have been supplemented by more numerical comparisons.



smcccd 4/4/06 9:43 AM
Comment: Again, you rely too heavily on the graphical analysis. The obvious thing to do here is to compare the mean or median prices for the various colors. I suspect that the means will not differ much, although perhaps the green cars are on average cheaper. Are the green ars also older. Did they leave of making green cars after a certain time?

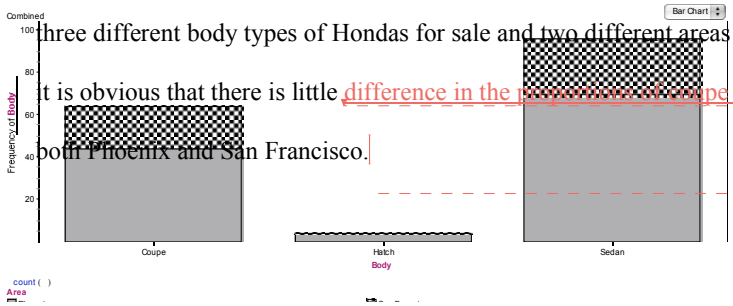
After reviewing the two histograms, and the scatter plot, regarding the influence of color on prix, area, and body style there have been very few differences between both areas. When focusing on color related to body style both cities had mostly the same demand for colors on all three styles of Honda Civics. The only difference was the Coupe had less of a demand for the colors gray and gold. Taking a next look on the histogram showing the influence of color on area showed a stronger relationship between the two variables. In Phoenix all of the colors were in high demand except for tan and red. The

smcccd 4/4/06 9:43 AM
Comment: You need percentages here to drive home your point.

other colors on the graph were darker colors. It was surprising to see such a high demand for dark colors in such a warm climate when darker colors absorb heat. In San Francisco all colors were pretty equal with just silver, blue and black in a slightly higher demand which shows no significant relationship between color and area. Next, we took a look at the relationship between color and prix (price) between both areas and all three body styles of the Honda Civic. Like the other histograms there was little to no relationship between the colors and prices in either area. There was only one color, gold, that had no prices under $15,000 but only above. Green and red also was shown to be sold only on lower valued cars unlike the colors blue, black and silver which sold strongly in both cheaper and more expensive cars in both cities.

smcccd 4/4/06 9:44 AM
Comment: Calculate!! You have an entire computer program to do it.

Referring to the histogram chart before showing the relationship between the three different body types of Hondas for sale and two different areas where they are sold, it is obvious that there is little difference in the proportions of coupe and sedans between both Phoenix and San Francisco.



smcccd 4/4/06 9:46 AM
Deleted: relationship

smcccd 4/4/06 9:47 AM
Comment: This sentence is not clear. Replace "relationship" with "difference" but again, calculate! I think that it would be better to have the ribbon graph based upon the place, since that is your explanatory variable.

After reviewing the three different body types being sold, the coupe, hatch and sedan, Phoenix had a larger proportion of Sedan's being sold than San Francisco. In fact

97% of the cars being sold in Phoenix were Honda Sedan's, where as only 38% of the cars being sold in San Francisco were Sedans. Another interesting observation from this data was that Phoenix had a small proportion of Honda Hatch cars for sale while San Francisco had none at all. The only explanation I could come up with is that the Hatch model is a bit outdated. If you go onto Honda's website you won't even see a picture of a Hatch. San Francisco being a famous city with a lot of marketing influence is also an expensive city.

smcccd 4/4/06 9:53 AM
Comment: Are you sure? The bar charts do not look like this. You could easily have got a summary table to show the numbers. I thought that

smcccd 4/4/06 10:01 AM
Comment: Hatchbacks are very popular in Europe but not in North America. In Europe the Jetta is known as a car for Americans, the Golf being far more popular there. Part of the reason for this is that compactness is more important in Europe.

smcccd 4/4/06 9:59 AM
Comment: And so?

To sum up; we found no real significant differences between the various variables that we used to compare between the two cities. This surprised us, because initially we thought we would see more differences in the data because of the drastic differences in climate between the two areas. However, one strong connection we did find was because these are two busy cities where commuting is a way of life for most, the Honda Civic and its various body styles, seem to be very popular for its low gas mileage and its legacy for being known as a "commuter" car.

After completing the research it was obvious to us that if we had done a comparison between San Francisco and a more rural town, such as Pocatello Idaho, we believe that we would have found significant differences in our data and conclusions. The reason for this hypothesis is mainly because in cities smaller cars are more in demand for their economic value, whereas rural areas that are more agrarian based, tend to demand larger utility vehicles and trucks to meet their farming and ranching needs.

smcccd 4/4/06 10:06 AM
Comment: The market in Pocotello Idaho wwill probably be too small to analyze. For 100 miles around Boise, I got $n = 42$.