Sample Final Part 1 - Solution

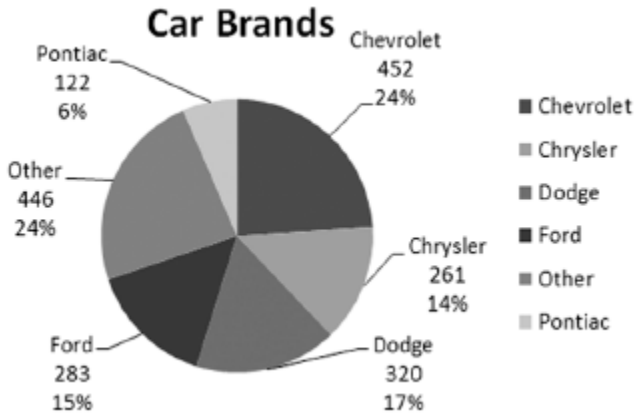
1)

(a)

	Frequency	Relative Freq
BUICK	16	0.00849
CHEVROLET	452	0.23992
CHRYSLER	261	0.13854
DODGE	320	0.16985
FORD	283	0.15021

(b) Chevrolet.
(c) The pie chart has too many slices.
(d)

Brand	Count
Chevrolet	452
Chrysler	261
Dodge	320
Ford	283
Other	446
Pontiac	122



2)

(a) Median=$220 million, mean = $2.2 billion, SD = $9.6 billion
(b) Heavily right skewed; outliers conceal most of the data.
(c) Three most extreme outliers are ATT, Verizon, and Microsoft.
(d) Dominated by very large telecoms and Microsoft.

3)

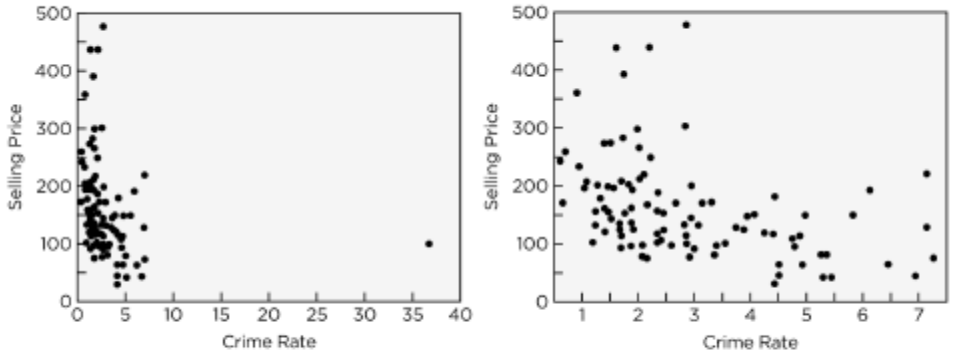
(a) Type of Day By Grade of Gasoline

Count	Premium	Plus	Regular	Total
Weekday	126	103	448	677
Weekend	63	29	115	207
Total	189	132	563	884

(b) Among weekday purchases, 126/677 = 19% are for premium, 103/677 = 15% are for plus, and the remaining 66% are for regular.
(c) 126/189=67% of premium purchases are on weekdays, and 33% on weekends.
(d) No. These conditional distributions are not directly comparable. One refers to the type of gas given that the purchase happens on the weekday, whereas the second refers to the timing of the purchase given that premium gas was bought. To illustrate association, one can, for example, compare the conditional distribution of purchases of regular gas to the answer in part c. For regular purchases 448/563 = 80% occur on weekdays, compared to 67% of premium purchases.
(e) Weekends. Though more premium gas is sold during the week, there's a greater concentration of premium sales on weekend days.

4)

(a) The plot on the left below plots the price on the crime rate. The outlier is Center City, Philadelphia, and it is unusual in terms of the crime rate, but not the selling price.



(b) The correlation using all of the data is $r = -0.25$.
(c) The refocused scatterplot on the right shows a great deal of variation around a weak, negative trend that appears to bend. The price seems to drop off faster on the left (few crimes) than the right (more crimes), either that or there are new outliers (such as the cluster of expensive districts at the upper left).

(d) The correlation without Center City, Philadelphia, is much stronger than previously found, -0.43.
(e) No, we cannot for several reasons. First, this is aggregated data. We do not see the prices for individual homes, only for communities. Second, correlation measures association, not causation.