

Name: \_\_\_\_\_

Math 227 / Fall 2019 / Prof. David Soto

Module 3. Sections 3.6, 3.7, 3.8

**Question 1.**

Use the instruction **sort(CommuteAtlanta\$Time)** to sort and display the values of the Time variable from the **CommuteAtlanta** data frame. What is the smallest value? What is the largest value?

**Question 2.**

Use the instruction **favstats(~Time, data = CommuteAtlanta)** to display the Five-number summary.

a) What is the value of the median for this data? What is the meaning of median?

b) What is the value of the first quartile for this data? What is the meaning of it?

c) What is the value of the third quartile for this data? What is the meaning of it?

d) What is the value of the range for this data?

e) What is the value of the interquartile range, IQR, for this data?

### Question 3.

Any data point bigger than the  $Q3 + 1.5 * IQR$  is considered a large outlier. What would be this breaking point for the Time variable? Do we have any large outliers in this data?

### Question 4.

Any data point bigger than the  $Q1 - 1.5 * IQR$  is considered a small outlier. What would be this breaking point for the Time variable? Do we have any large outliers in this data?

### Question 5.

Use the instruction `gf_boxplot(Time~1, data = CommuteAtlanta)` to draw a boxplot of the Time variable in the CommuteAtlanta data frame.

Which instruction would you use to draw a boxplot of the Age variable in the CommuteAtlanta data frame? Do you see any outlier on this graph?

### Question 6.

Use the instruction `gf_bar(~Sex, data = CommuteAtlanta)` to create a bar graph (kind of a histogram for a categorical variable) of the Sex variable. Would you say that you have significantly more drivers of one sex than the other?

### Question 7.

Use the instruction `tally(~Sex, data = CommuteAtlanta, format = "proportion")` to obtain the proportion of female and male commuters. Which percent of commuters are female? Male?