

# Practice Problems: Week 1

1. Consider the diamonds data set, available on the class website.
  - (a) Is the data set structured or unstructured? Why?
  - (b) How many tables does the data set have?
  - (c) What are the cases?
  - (d) What are the variables?
  - (e) What kind of variable is each variable? Be specific!
  - (f) What kind of graph(s) would you use to visualize each kind of variable?
  - (g) If you wanted to use StatCrunch to visualize the distribution of the cut variable with a pie chart, would you choose “with data,” or “with summary?” How would the data need to be structured to compel the other choice?
  - (h) Are there any labels in the diamonds data set?
2. In the diamonds data set, you create a new variable called “ideal.” You assign the value 1 to this new variable for every diamond in the data set with the ideal cut and you assign 0 to this new variable for every other diamond.
  - (a) How could you more simply describe the sum of the values of the ideal variable in terms of statistics that are important for categorical variables?
  - (b) How could you more simply describe the average (mean) of the values of the ideal variable in terms of statistics that are important for categorical variables?
3. What does the distribution of a variable tell us? Your answer should be the loose definition given in class.
4. What graphs help us visualize the distribution of each kind of variable we studied (different answers for different kinds of variables)? When would you choose one type of graph over the other?
5. Describe a bar plot with Pareto ordering. Why would such an ordering be desirable?
6. Name two things you can do with StatCrunch (or other statistical software) to better visualize a distribution of a quantitative variable with a histogram?
7. Put the following data into stemplot. Trim and split stems, and include all data points. You may use StatCrunch, but you do not have to.

236, 375, 383, 412, 413, 426, 440, 458, 459, 465, 491, 507