

Introduction to Data Science
Homework 2: Due Wednesday September 5 at 2:00pm

Exercises:

1. Read the Structured Data handout and the first few sections through *Visualising distributions of R for Data Science*.
2. Find a data set or data sets that exemplify each of the terms: rectangular or tabular data, continuous, discrete, categorical, binary, ordinal.
3. True or False: The `diamonds` data set in the `ggplot2` package is a data frame. One way to answer this question is by typing the command `str(diamonds)` in R.
4. True or False: The `diamonds` data set in the `ggplot2` package is “tidy.”
5. How many variables are there in the `diamonds` data set?
6. Classify each of the variables in the `diamonds` data set. That is, state if the variable is continuous, discrete, categorical, etc.
7. Describe the difference between a bar plot and a histogram. Under what circumstances would you use each?