series, financial data, or geographic maps. A useful resource for exploring such packages is the Task Views http://cran.r-project.org/web/views/.

Python's matplotlib can display additional types of graphs, including bar charts, two dimensional scatter plots, three dimensional surface plots. See http://matplotlib.org for details. Python's pandas module has some graphics functionality that is useful for graphing dataframes. See http://pandas.pydata.org for details. Python also has additional graphics module for specialized graphics, such as interactive graphics.

10.17 Exercises

1. Using the `mpg` data, describe the relationship between highway mpg and car manufacturer. Describe which companies produce the most and least fuel efficient cars, and display a graph supporting your conclusion.

2. Using the `mpg` data, explore the three-way relationship between highway mpg, city mpg, and model class. What are your observations? Display a graph supporting these observations.

3. What are the pros and cons of using a histogram vs a box plot? Which one will you prefer for what purpose?

4. Generate two sets of N random points using the function `runif` and display a corresponding scatter plot. If you save the file to disk, what is the resulting file size for the following file formats: ps, pdf, jpeg, png? How do these values scale with increasing N?

5. The `diamonds` dataset within ggplot2 contains 10 columns (price, carat, cut, color, etc.) for 53940 different diamonds. Type `help(diamonds)` for more information. Plot histograms for color, carat, and price, and comment on their shapes. Investigate the three-way relationship between price, carat, and cut. What are your conclusions? Provide graphs that support your conclusions. If you encounter computational difficulties, consider using a smaller dataframe whose rows are sampled from the original `diamonds` dataframe. Use the function `sample` to create a subset of indices that may be used to create the smaller dataframe.