

## PROBLEMS

- 3.1 Shipments of Household Appliances: Line Graphs.** The file *ApplianceShipments.csv* contains the series of quarterly shipments (in millions of dollars) of US household appliances between 1985 and 1989.
- Create a well-formatted time plot of the data using R.
  - Does there appear to be a quarterly pattern? For a closer view of the patterns, zoom in to the range of 3500–5000 on the  $y$ -axis.
  - Using R, create one chart with four separate lines, one line for each of Q1, Q2, Q3, and Q4. In R, this can be achieved by generating a data.frame for each quarter Q1, Q2, Q3, Q4, and then plotting them as separate series on the line graph. Zoom in to the range of 3500–5000 on the  $y$ -axis. Does there appear to be a difference between quarters?
  - Using R, create a line graph of the series at a yearly aggregated level (i.e., the total shipments in each year).
- 3.2 Sales of Riding Mowers: Scatter Plots.** A company that manufactures riding mowers wants to identify the best sales prospects for an intensive sales campaign. In particular, the manufacturer is interested in classifying households as prospective owners or nonowners on the basis of Income (in \$1000s) and Lot Size (in 1000 ft<sup>2</sup>). The marketing expert looked at a random sample of 24 households, given in the file *RidingMowers.csv*.
- Using R, create a scatter plot of Lot Size vs. Income, color-coded by the outcome variable owner/nonowner. Make sure to obtain a well-formatted plot (create legible labels and a legend, etc.).
- 3.3 Laptop Sales at a London Computer Chain: Bar Charts and Boxplots.** The file *LaptopSalesJanuary2008.csv* contains data for all sales of laptops at a computer chain in London in January 2008. This is a subset of the full dataset that includes data for the entire year.
- Create a bar chart, showing the average retail price by store. Which store has the highest average? Which has the lowest?
  - To better compare retail prices across stores, create side-by-side boxplots of retail price by store. Now compare the prices in the two stores from (a). Does there seem to be a difference between their price distributions?
- 3.4 Laptop Sales at a London Computer Chain: Interactive Visualization.** The next exercises are designed for using an interactive visualization tool. The file *LaptopSales.txt* is a comma-separated file with nearly 300,000 rows. ENBIS (the European Network for Business and Industrial Statistics) provided these data as part of a contest organized in the fall of 2009.
- Scenario:** Imagine that you are a new analyst for a company called Acell (a company selling laptops). You have been provided with data about products and sales. You need to help the company with their business goal of planning a product strategy and pricing policies that will maximize Acell's projected revenues in 2009. Using an interactive visualization tool, answer the following questions.
- Price Questions:**
    - At what price are the laptops actually selling?
    - Does price change with time? (*Hint:* Make sure that the date column is recognized as such. The software should then enable different temporal aggregation

choices, e.g., plotting the data by weekly or monthly aggregates, or even by day of week.)

- iii. Are prices consistent across retail outlets?
- iv. How does price change with configuration?

**b. Location Questions:**

- i. Where are the stores and customers located?
- ii. Which stores are selling the most?
- iii. How far would customers travel to buy a laptop?
  - *Hint 1:* You should be able to aggregate the data, for example, plot the sum or average of the prices.
  - *Hint 2:* Use the coordinated highlighting between multiple visualizations in the same page, for example, select a store in one view to see the matching customers in another visualization.
  - *Hint 3:* Explore the use of filters to see differences. Make sure to filter in the zoomed out view. For example, try to use a “store location” slider as an alternative way to dynamically compare store locations. This might be more useful to spot outlier patterns if there were 50 store locations to compare.
- iv. Try an alternative way of looking at how far customers traveled. Do this by creating a new data column that computes the distance between customer and store.

**c. Revenue Questions:**

- i. How do the sales volume in each store relate to Acell’s revenues?
- ii. How does this relationship depend on the configuration?

**d. Configuration Questions:**

- i. What are the details of each configuration? How does this relate to price?
- ii. Do all stores sell all configurations?