4 Using Graphics for Exploratory Data Analysis

"Use a picture. It's worth a thousand words."
Arthur Brisbane (1864-1936) Newspaper Editor"

There are a lot of good reasons to use a graphical-based approach to data mining. The first is that you are able to process much more information graphically rather than when reading data from tables. Prior to any complex statistical analysis of data sets, an exploratory glance of the data is required to understand the distribution characteristics of the data, in an effort to apply the best available method of statistical testing. This chapter further expands on the development of more complex graphics and statistical analysis within R.

There are two types of variables: discrete and continuous. This chapter focuses on the exploratory graphics available for both types of variables and recounts methods of descriptive statistics and distribution diagnostics. It also presents an introduction to working with time variables in R because time classes in R can be a daunting unless you have a base understanding of how R handles the two time classes.

The majority of the chapter is concerned with the exploration of data sets, both graphically and statistically. A basic review of descriptive statistics and an introduction to the statistical functions within R are presented. Following this, confidence intervals and hypothesis testing are presented through an in depth look into the Z and t distributions for single continuous variables. Multivariate analysis is also presented in the form of the ANOVA test and appropriate post-hoc tests. The chapter concludes with activities introducing and expanding upon multivariate graphics using previously presented functions and introduces several packages that generate refined graphics for two continuous variables.

ACTIVITY LIST

Activity Type	Number and Title	Assessment Type
In Class	Activity 24: Interactive Review of Basic Statistics Using R	
Out-of-class	Activity 25: Basic Charts for Single Discrete Variable	Short quiz
In Class	Activity 26: Exploring Single Discrete Variable Plots	Participation
Out-of-class	Activity 27: Exploratory and Diagnostic Plots for the Distribution of a Single Discrete Variable	Short quiz
Out-of-class	Activity 28: Probability Distributions	Short quiz
In Class	Activity 29: Kernel Density Estimates and Histograms	Short Response
In Class	Activity 30: Diagnosing a Distribution	Short Response

Activity Type	Number and Title	Assessment Type
Out-of-class	Activity 31: Depicting the Distribution Involving Discrete Variables	Short quiz
Out-of-class	Activity 32: Depicting the Distribution of Two Continuous Variables	Short quiz
In Class	Activity 33: Introduction to Final Project Topics	Short Response
Out-of-class	Activity 34: One and Two Sample Tests	Short quiz
I/O Class	Activity 35: Exploring Confidence Intervals & Simple Hypothesis Testing	Short Response
Out-of-class	Activity 36: An Application of Hypothesis Testing	Discovery

Student Notes