Chapter 2: Descriptive Statistics: Tabular and Graphical Methods

Example 1

```
C2_1 <- c(rep('Latin', 3), rep('Europe', 4), rep('US', 8),
      rep('India', 10), rep('China', 15))

head(C2_1, 8)

C2_2 <- table(C2_1)
```

Example 2

```
n <- sum(C2_2)

C2_3 <- C2_2 / n
```

Example 3

```
barplot(C2_2, col = c('purple', 'red', 'blue', 'yellow', 'green'),
    xlab = 'Country/Region', ylab = 'Count', ylim = c(0, 20),
    main = 'Bar Graph of Student Origin Data')
```

Example 4

```
dotchart(sort(C2_3), xlab = 'Relative Frequencies of Nation/Region',
      main = 'A Dot Chart of Student Origin Data',
      xlim = c(0, 0.40), pch = 20, col = 'blue')
```

Example 5

```
head(tv_hours, 2)

C2_4 <- cut(tv_hours$hours, c(0, 5, 10, 15, 20, 25))
C2_5 <- table(C2_4)
```

Example 6

```
C2_6 <- C2_5 / sum(C2_5)
```

Example 7

```
hist(tv_hours$hours, breaks = 5, xlab = "Hours of TV Viewing",
     ylim = c(0, 50), main = "Histogram of Hours of TV
     Viewing", col = "blue")
```

Example 8

```
names(students)

C2_7 <- cut(students$rating, c(0, 2, 4, 6, 8))

C2_8 <- table(students$year, C2_7)

Total <- rowSums(C2_8)
```

```
C2_8 <- cbind(C2_8, Total)

Total <- colSums(C2_8)

C2_9 <- rbind(C2_8, Total)

C2_10 <- rbind(C2_9[1, ], C2_9[4, ], C2_9[2, ], C2_9[3, ],
     C2_9[5, ])

rownames(C2_10) <- c('Freshmen', 'Sophomores', 'Juniors',
     'Seniors', 'Column Total')

colnames(C2_10) <- c('Poor', 'Below.Avg', 'Above.Avg',
     'Excellent', 'Row Total')
```

Example 9

```
names(poverty)

plot(poverty$Poverty, poverty$Obesity, xlab = 'Percent in
     Poverty', ylab = 'Percent Designated Obese', pch = 19, col =
     'red')
```

Example 10

```
plot(poverty$Education, poverty$Obesity, xlab = 'Percent with
     College Degree', ylab = 'Percent Designated Obese', pch =
     19, col = 'red')
```

End-of-Chapter 2 Exercises

Exercise 1

```
E2_1 <- c(rep('A', 272), rep('B', 212), rep('C', 297),
     rep('D', 38), rep('E', 181), rep('F', 95))

fd <- table(E2_1)
```

Exercise 2

```
options(digits = 2)

rf <- fd / sum(fd)

barplot(fd, col = c('red','blue', 'green', 'violet', 'orange',
     'cyan'), ylim = c(0, 300), main = 'Number of Households
     Preferring Brand', xlab = 'Brands', ylab = 'Frequencies')
```

Exercise 3

```
barplot(rf, col = c('brown', 'purple', 'brown', 'purple',
     'brown', 'purple'), ylim = c(0, 0.35), main = 'Percentages',
     xlab = 'Brand', ylab = 'Relative Frequencies')
```

Exercise 4

```
library("MASS")

names(Cars93)

fd <- table(Cars93$Type)
```

rf <- fd / sum(fd)

table(Cars93$Origin, Cars93$Type)

Exercise 5

plot(poverty$Wind, poverty$Poverty, xlab = 'Wind', ylab = 'Percent Below Poverty Line', pch = 1, col = 'blue')

R Functions

abline() Provides a reference line to a scatter plot.

barplot() Produces a bar graph of the count or frequency of observations falling   into two or more mutually-exclusive, collectively exhaustive categories.

. cbind() Binds together a column of data values with a separate object.

. colnames() Changes the names of the columns in a table.

. colSums() Provides sum of values down all columns of a table for each row category.

. cut() Divides the range of quantitative values into interval categories and then assigns those values to the interval in which they fall.

. dotchart() Provides a simple graphical representation of the frequency, or relative frequency, of qualitative data.

. head() Displays the first six lines of a data object.

. hist() Provides a histogram of the count or frequency of quantitative data values;   bar cans be either verically or horizontally arranged.

. length() Can be used to count the number of data values in a data object.

. lm() Runs linear model methods (e.g., linear regression, analysis of variance, etc.)

. mosaicplot() Provides a mosaic plot of the relationship between two (or more) categorical variables.

. ncol() Can be used to count the number of columns in a data object.

. nrow() Can be used to count the number of rows in a data object.

. plot() Creates a scatter diagram of points based on the position of each observation on two quantitative variables.

. prod() Provides the product of a sequence of numbers. That is, prod(n:N) reports the product of all numbers running from the smallest value n to the largest value N.

. rbind() Binds together a row of data values with a separate object.

. rep() Replicates or repeats values defined in ().

. rnorm() Generates a set of data values that are normally distributed.

. rownames() Changes the names of the rows of a table.

. rowSums() Provides sum of values across all rows of a table for each column category.

. sort() Sorts data into ascending (or descending, if preferred) order.

. sum() Provides the sum of data values defined in ().

. summary() Reports the Five-Number Summary (the minimum and maximum values plus the 1st, 2nd, and 3rd quartiles) as well as the mean.

. table() Creates tabular representations of categorical data.

Appendix

```
a <- c(1, 2, 3)
  b <- c(4, 5, 6)
c <- c(7, 8, 9)
d <- c(a, b, c)
e <- a + b + c
f <- a - b
g <- a * b
h <- a / b
i <- c("Red", "White", "Blue")
j <- rep(c(1, 2, 3), 3)
```

```
k <- c(5 : 15)

l <- seq(from = 5, to = 6, by = 0.1)

m <- rnorm(10000, 100, 15)

hist(m, breaks = 50, ylim = c(0, 600), col = 'red')


n<-runif(20000,75,125)

hist(n, breaks = 50, xlim = c(70, 130), ylim = c(0, 500),
    col = 'blue',)

names(longley)

longley[ , 3 : 7]

longley[longley$Year > 1959, 7]

mean(longley[longley$Year > 1959, 7])

longley[longley$Year < 1950, 7]

mean(longley[longley$Year < 1950, 7])


C2_11 <- longley[longley$Armed.Forces>300, c('Year','GNP')]
```