1. 1. > BABIES = read.table("http://www.stat.berkeley.edu/users/statlabs/data/babies.data",header = T)  
      > dim(BABIES)

[1] 1236 7

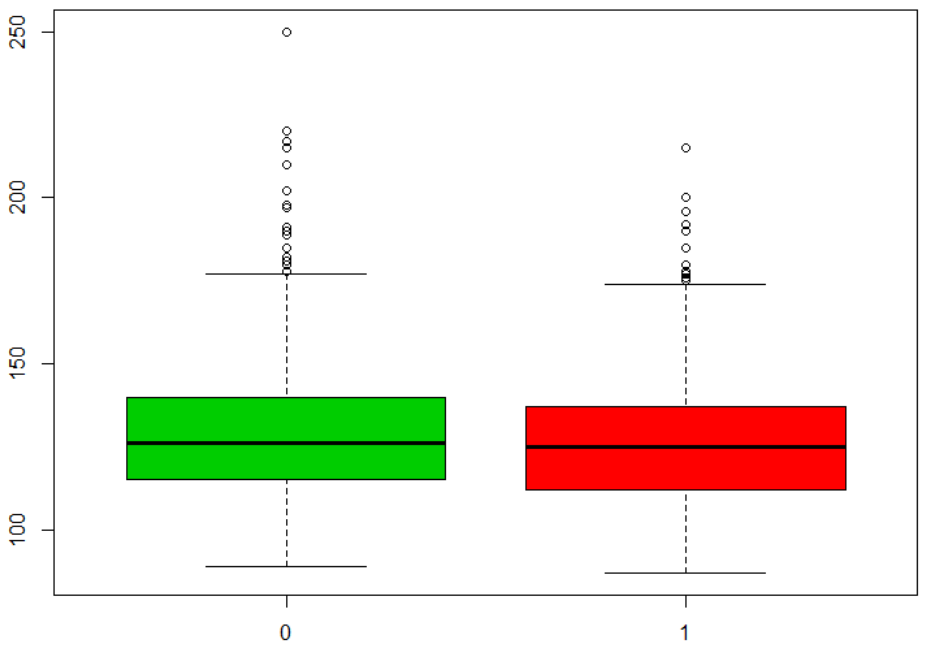
> CLEAN = BABIES[-(which(gestation==999|gestation==999|parity==9|height==99|weight==999|smoke==9)),]

> dim(CLEAN)

[1] 1175 7



> boxplot(CLEAN$weight~factor(CLEAN$smoke),col = c(3,2))





> summary(CLEAN[CLEAN$smoke==1,'weight'])

Min. 1st Qu. Median Mean 3rd Qu. Max.

87.0 112.0 125.0 126.9 137.0 215.0

> summary(CLEAN[CLEAN$smoke==0,'weight'])

Min. 1st Qu. Median Mean 3rd Qu. Max.

89.0 115.0 126.0 129.5 140.0 250.0



> data = c(140, 125, 150, 102,143, 170,120,94,53,115)

> t.test(data,alternative = 'greater',mu=100)

One Sample t-test

data: data

t = 2.0261, df = 9, p-value = 0.0367

alternative hypothesis: true mean is greater than 100

95 percent confidence interval:

102.0193 Inf

sample estimates:

mean of x

121.2  
H0: test book cost is not greater than $100 per class  
Ha: test book cost is greater than $100 per class  
p-value < 0.05, reject null hypothesis



> t.test(data,alternative = 'greater',mu=100,conf.level = 0.95)$conf.int

[1] 102.0193 Inf

attr(,"conf.level")

[1] 0.95

> t.test(data,alternative = 'greater',mu=100,conf.level = 0.9)$conf.int

[1] 106.7287 Inf

attr(,"conf.level")

[1] 0.9



> method1 = scan()

1: 45.9 47.6 54.9 38.7 35.7 39.2 45.9 43.2 45.4 54.8

11:

Read 10 items

> method2 = scan()

1: 48.2 64.2 56.8 47.2 43.7 45.7 53.0 52.0 45.1 57.5

11:

Read 10 items

> t.test(method1,method2)

Welch Two Sample t-test

data: method1 and method2

t = -2.1399, df = 17.977, p-value = 0.04633

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-12.307351 -0.112649

sample estimates:

mean of x mean of y

45.13 51.34  
H0: the mean of two methods are not different

Ha: the mean of two methods are different  
p-value < 0.05, reject null hypothesis, so the mean of two methods are different

Then, let’s test which mean is greater.

> t.test(method1,method2,alternative = 'greater')

Welch Two Sample t-test

data: method1 and method2

t = -2.1399, df = 17.977, p-value = 0.9768

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

-11.24253 Inf

sample estimates:

mean of x mean of y

45.13 51.34  
H0: the mean of method 1 is not greater than the mean of method 2  
Ha: the mean of method 1 is greater than the mean of method 2  
p-value > 0.05, do not have enough evidence to reject the null hypothesis, so the mean of method 1 is not greater than the mean of method 2, since the two means are different, so the mean of method 1 is greater than the mean of method 2.

1. 1. > library(MASS)

> head(quine,5)

Eth Sex Age Lrn Days

1 A M F0 SL 2

2 A M F0 SL 11

3 A M F0 SL 14

4 A M F0 AL 5

5 A M F0 AL 5



> table(data.frame(quine$Eth,quine$Sex))

quine.Sex

quine.Eth F M

A 38 31

N 42 35

> prop.test(x=c(31,38),n=c(80,66),alternative='two.sided')

2-sample test for equality of proportions with continuity correction

data: c(31, 38) out of c(80, 66)

X-squared = 4.4147, df = 1, p-value = 0.03563

alternative hypothesis: two.sided

95 percent confidence interval:

-0.36212634 -0.01438881

sample estimates:

prop 1 prop 2

0.3875000 0.5757576

H0: the proportion of aboriginal female not different from that of male  
Ha: the proportion of aboriginal female different from that of male

P-value < 0.05, reject null hypothesis, the proportion of aboriginal female different from that of male



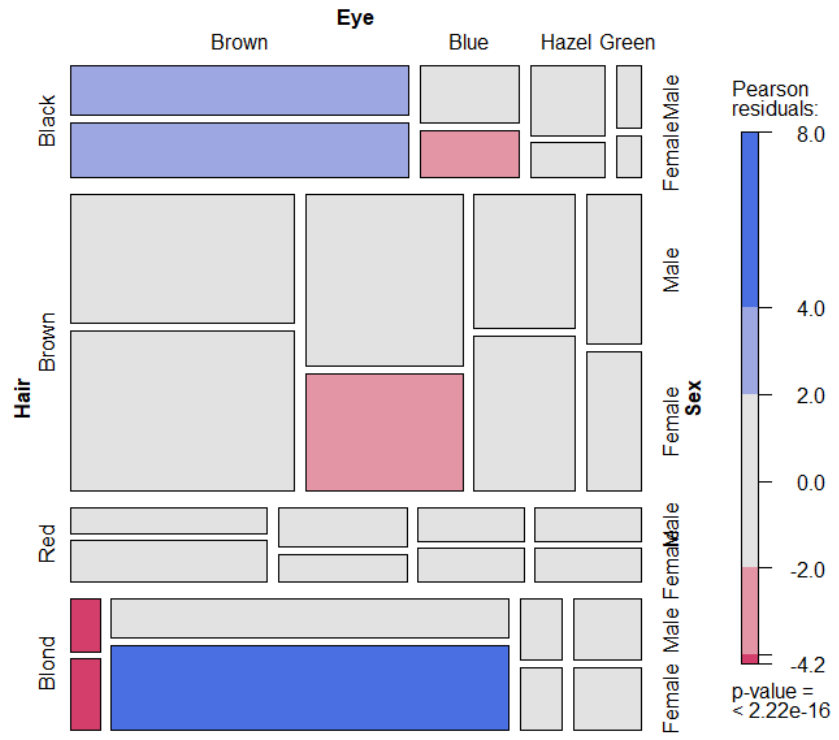
> sum(HairEyeColor)

[1] 592



> library(vcd)

Loading required package: grid

> mosaic(HairEyeColor, shade=TRUE, legend=TRUE)  




> chisq.test(HairEyeColor[,,1])

Pearson's Chi-squared test

data: HairEyeColor[, , 1]

X-squared = 41.28, df = 9, p-value = 4.447e-06

H0: hair color independent of eye color for men  
Ha: hair color dependent of eye color for men

p-value < 0.05, reject null hypothesis, hair color dependent of eye color for men



> chisq.test(HairEyeColor[,,2])

Pearson's Chi-squared test

data: HairEyeColor[, , 2]

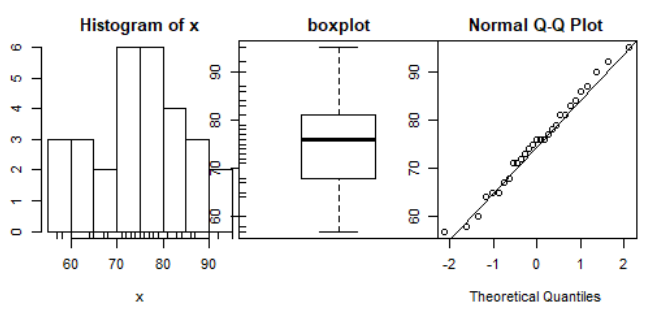
X-squared = 106.66, df = 9, p-value < 2.2e-16

H0: hair color independent of eye color for women

Ha: hair color dependent of eye color for women

p-value < 0.05, reject null hypothesis, hair color dependent of eye color for women



library(‘UsingR’)  
> simple.eda(TestScores[,])  




> shapiro.test(TestScores[,])

Shapiro-Wilk normality test

data: TestScores[, ]

W = 0.9852, p-value = 0.9461  
H0: test score distribution is normal

Ha: test score distribution is not normal

p-value > 0.05, do not have enough evidence to reject null hypothesis, the test score distribution is normal



library(UsingR)  
> shapiro.test(OBP)

Shapiro-Wilk normality test

data: OBP

W = 0.97092, p-value = 1.206e-07

H0: on-base percentage distribution is normal

Ha: on-base percentage distribution is not normal

p-value < 0.05, reject the null hypothesis, the on-base percentage distribution is not normal