ST.310, Spring 2024

Homework #8

Due on Wednesday, 4/10/24

***Directions:***

* Follow the ***homework format guidelines*** shown on the syllabus.
* Do ***not*** fill in your source code and answers on this problem set. Use the homework template.
* You may work with one partner if you wish.
* Upload a single Word file (saved as ***doc*** or ***docx***) on Moodle by ***11:59 PM*** on the due date.
* Late homework will ***not*** be accepted without any legitimate excuses.

Load the UsingR package. The package includes all textbook datasets. Show your R command lines and outputs for each question.

**Problem#3.14.** [Page 128]

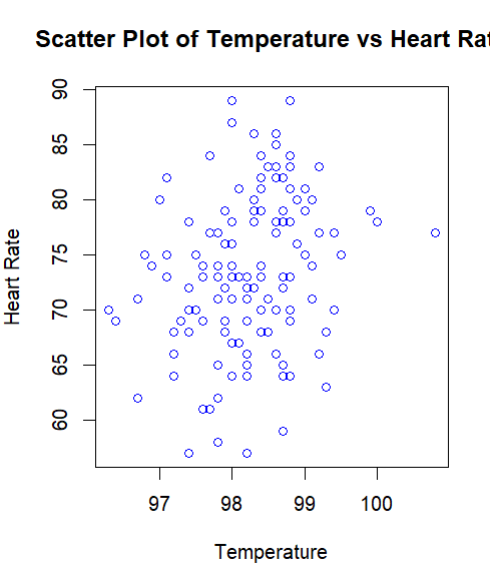
The data set normtemp contains body measurements for 130 healthy, randomly selected individuals. The variable temperature measures normal body temperature, and the variable hr measures resting heart rate. Make a scatter plot of the two variables and find the Pearson correlation coefficient. [10 pts]

data(normtemp)

plot(normtemp$temperature, normtemp$hr, xlab = "Temperature", ylab = "Heart Rate",main = "Scatter Plot of Temperature vs Heart Rate", col = "blue", )

correlation\_coefficient <- cor(normtemp$temperature, normtemp$hr)

print(paste("Pearson Correlation Coefficient:", correlation\_coefficient))



[1] "Pearson Correlation Coefficient: 0.253656402720764"

**Problem#3.18.** [Page 129]

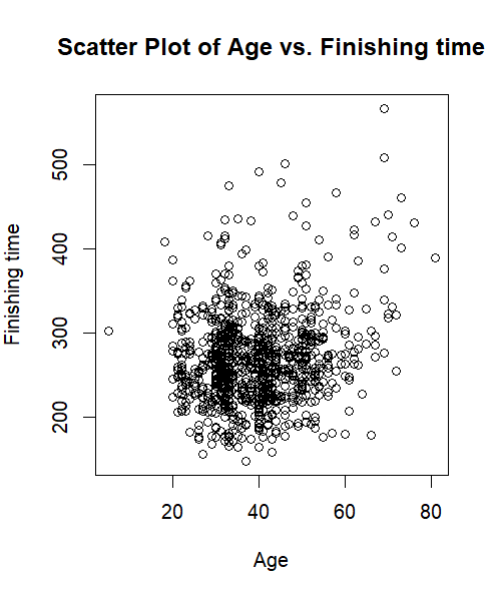
The data set nym.2002 contains information about the 2002 New York City Marathon. What do you expect the correlation between age and finishing time to be? **Find it and see whether you were close. Also, construct a scatterplot of age vs time.** [10 pts]

plot(nym.2002$age, nym.2002$time, xlab = "Age", ylab = "Finishing time", main = "Scatter Plot of Age vs. Finishing time")

correlationCoefficient <- cor(nym.2002$age, nym.2002$time)

print(paste("Pearson Correlation Coefficient:", correlationCoefficient))

I believe that people who are younger are most likely to finish earlier than the people who are older. This is somewhat true those who are between the age 20-60 finished before people who are 60+



[1] "Pearson Correlation Coefficient: 0.189867151009491"

**Problem#3.24.** [Page 130]

The data frame x77 contains data from each of the fifty United States. First coerce the state.x77 variable into a data frame with

x77 <- data.frame(state.x77)

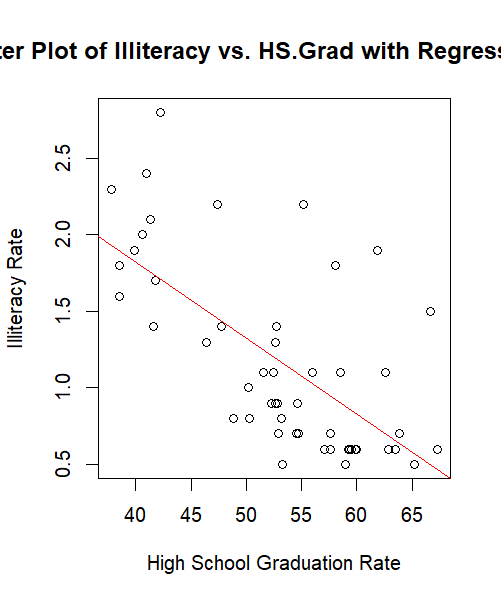
For the following, make a scatter plot with regression line:

1. The model of illiteracy rate (Illiteracy) modeled by high school graduation rate (HS.Grad).

x77 <- data.frame(state.x77)

plot(x77$HS.Grad, x77$Illiteracy, xlab = "High School Graduation Rate", ylab = "Illiteracy Rate", main = "Scatter Plot of Illiteracy vs. HS.Grad with Regression Line")

abline(lm(Illiteracy ~ HS.Grad, data = x77), col = "red")

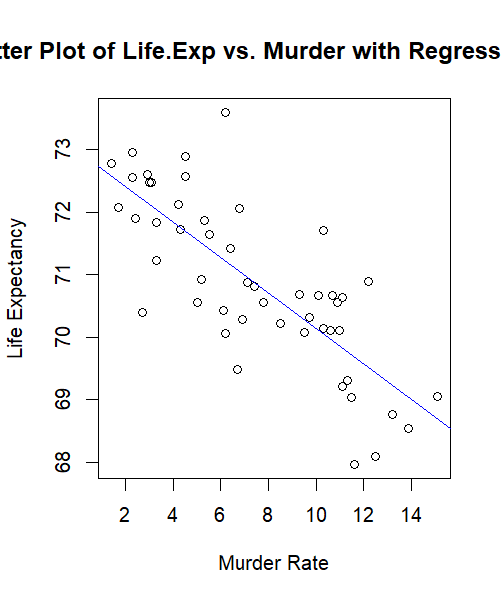


1. The model of life expectancy (Life.Exp) modeled by the murder rate (Murder).

x77 <- data.frame(state.x77)

plot(x77$Murder, x77$Life.Exp, xlab = "Murder Rate", ylab = "Life Expectancy", main = "Scatter Plot of Life.Exp vs. Murder with Regression Line")

abline(lm(Life.Exp ~ Murder, data = x77), col = "blue")

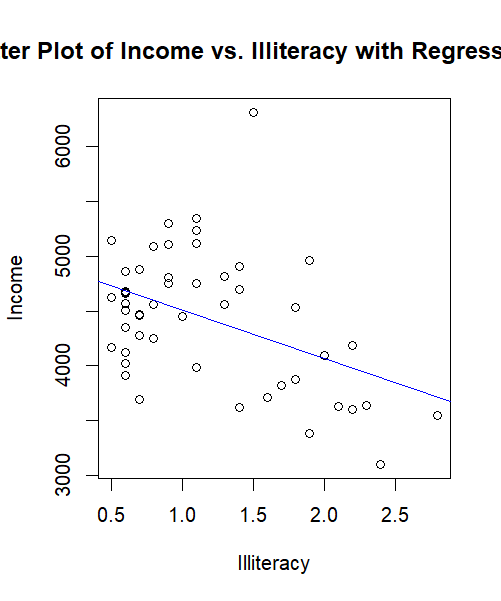


1. The model of income (Income) modeled by the illiteracy rate (Illiteracy).

x77 <- data.frame(state.x77)

plot(x77$Illiteracy, x77$Income, xlab = "Illiteracy", ylab = "Income", main = "Scatter Plot of Income vs. Illiteracy with Regression Line")

abline(lm(Income ~ Illiteracy, data = x77), col = "blue")



**Write a sentence or two describing any relationship. In particular, do you find it as expected or is it surprising? [30 pts]**

Most of the plots are like to each others in looks. The plots do not line up on the Stright line, instead they are scattered around the lines. The relationships between of best fit and the data itself.

As for the slopes of theses plots: income vs illiteracy, illiteracy vs H.S Grad have a relationship while life exp vs murder has a strong relationship.

**Problem#3.31.** [Page 149]

The data set coins contains the number of coins in a change bin and the years they were minted. Do the following:

1. How much money is in the change bin? [10 pts]

data(coins)

totalValue <- sum(coins$value)

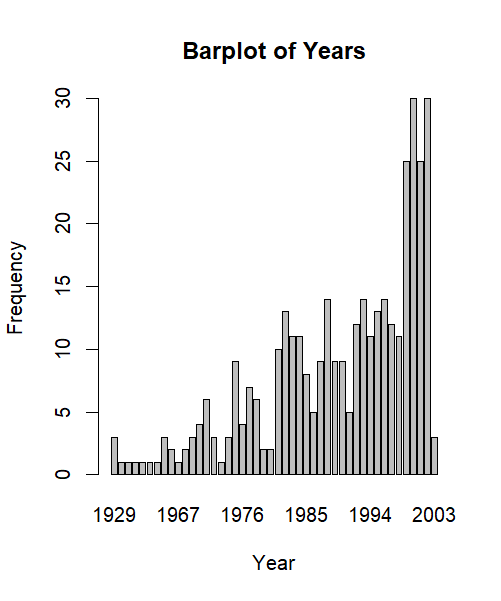
cat("Total amountof coins in the change bin:", totalValue, "cents\n")

Total amountof coins in the change bin: 25.93 cents

1. Make a barplot of the years. **Is there a trend?** [10 pts]

barplot(table(coins$year), xlab = "Year", ylab = "Frequency", main = "Barplot of Years")

There is a trend, there a more coins in the coin bin minted from recent years.



1. *Skipped*.
2. Make a contingency table of the year and the value. Does it look like the two variables are associated? Why?[10 pts]

contingency<- table(coins$year, coins$value)

print(contingency)

0.01 0.05 0.1 0.25

1929 2 1 0 0

1936 0 0 1 0

1939 0 0 1 0

1955 0 0 0 1

1959 1 0 0 0

1960 1 0 0 0

1964 1 0 0 0

1965 2 1 0 0

1966 1 1 0 0

1967 0 1 0 0

1968 1 0 0 1

1969 2 0 1 0

1970 1 1 0 2

1971 3 0 2 1

1972 2 0 0 1

1973 1 0 0 0

1974 1 1 1 0

1975 4 1 0 4

1976 2 1 0 1

1977 4 2 0 1

1978 5 1 0 0

1979 1 1 0 0

1980 1 1 0 0

1981 7 1 2 0

1982 9 1 2 1

1983 7 3 1 0

1984 7 3 0 1

1985 2 4 1 1

1986 3 0 1 1

1987 7 1 1 0

1988 7 3 2 2

1989 4 0 3 2

1990 3 3 1 2

1991 2 1 1 1

1992 6 0 0 6

1993 8 3 1 2

1994 7 1 2 1

1995 6 2 3 2

1996 7 3 0 4

1997 6 2 1 3

1998 5 1 0 5

1999 16 2 6 1

2000 15 4 3 8

2001 13 5 0 7

2002 19 3 5 3

2003 1 0 0 2

>

**Problem#4.7.** [Page 166]

Use the subset function to return a data frame made from the Cars93 data frame consisting only of non-USA cars in origin, with 4 cylinders and a maximum price of $15,000 or less. [10 pts]

subsetDf <- subset(Cars93, origin != "USA" & cylinders == 4 & Price <= 15000)

print(subsetDf)

Manufacturer Model Type

1 Acura Integra Small

5 BMW 535i Midsize

40 Geo Storm Sporty

41 Honda Prelude Sporty

42 Honda Civic Small

43 Honda Accord Compact

44 Hyundai Excel Small

45 Hyundai Elantra Small

46 Hyundai Scoupe Sporty

47 Hyundai Sonata Midsize

53 Mazda 323 Small

54 Mazda Protege Small

55 Mazda 626 Compact

58 Mercedes-Benz 190E Compact

62 Mitsubishi Mirage Small

64 Nissan Sentra Small

65 Nissan Altima Compact

78 Saab 900 Compact

81 Subaru Loyale Small

82 Subaru Legacy Compact

84 Toyota Tercel Small

85 Toyota Celica Sporty

86 Toyota Camry Midsize

87 Toyota Previa Van

88 Volkswagen Fox Small

90 Volkswagen Passat Compact

92 Volvo 240 Compact

Min.Price Price Max.Price

1 12.9 15.9 18.8

5 23.7 30.0 36.2

40 11.5 12.5 13.5

41 17.0 19.8 22.7

42 8.4 12.1 15.8

43 13.8 17.5 21.2

44 6.8 8.0 9.2

45 9.0 10.0 11.0

46 9.1 10.0 11.0

47 12.4 13.9 15.3

53 7.4 8.3 9.1

54 10.9 11.6 12.3

55 14.3 16.5 18.7

58 29.0 31.9 34.9

62 7.7 10.3 12.9

64 8.7 11.8 14.9

65 13.0 15.7 18.3

78 20.3 28.7 37.1

81 10.5 10.9 11.3

82 16.3 19.5 22.7

84 7.8 9.8 11.8

85 14.2 18.4 22.6

86 15.2 18.2 21.2

87 18.9 22.7 26.6

88 8.7 9.1 9.5

90 17.6 20.0 22.4

92 21.8 22.7 23.5

MPG.city MPG.highway

1 25 31

5 22 30

40 30 36

41 24 31

42 42 46

43 24 31

44 29 33

45 22 29

46 26 34

47 20 27

53 29 37

54 28 36

55 26 34

58 20 29

62 29 33

64 29 33

65 24 30

78 20 26

81 25 30

82 23 30

84 32 37

85 25 32

86 22 29

87 18 22

88 25 33

90 21 30

92 21 28

AirBags DriveTrain

1 None Front

5 Driver only Rear

40 Driver only Front

41 Driver & Passenger Front

42 Driver only Front

43 Driver & Passenger Front

44 None Front

45 None Front

46 None Front

47 None Front

53 None Front

54 None Front

55 Driver only Front

58 Driver only Rear

62 None Front

64 Driver only Front

65 Driver only Front

78 Driver only Front

81 None 4WD

82 Driver only 4WD

84 Driver only Front

85 Driver only Front

86 Driver only Front

87 Driver only 4WD

88 None Front

90 None Front

92 Driver only Rear

Cylinders EngineSize Horsepower

1 4 1.8 140

5 4 3.5 208

40 4 1.6 90

41 4 2.3 160

42 4 1.5 102

43 4 2.2 140

44 4 1.5 81

45 4 1.8 124

46 4 1.5 92

47 4 2.0 128

53 4 1.6 82

54 4 1.8 103

55 4 2.5 164

58 4 2.3 130

62 4 1.5 92

64 4 1.6 110

65 4 2.4 150

78 4 2.1 140

81 4 1.8 90

82 4 2.2 130

84 4 1.5 82

85 4 2.2 135

86 4 2.2 130

87 4 2.4 138

88 4 1.8 81

90 4 2.0 134

92 4 2.3 114

RPM Rev.per.mile Man.trans.avail

1 6300 2890 Yes

5 5700 2545 Yes

40 5400 3250 Yes

41 5800 2855 Yes

42 5900 2650 Yes

43 5600 2610 Yes

44 5500 2710 Yes

45 6000 2745 Yes

46 5550 2540 Yes

47 6000 2335 Yes

53 5000 2370 Yes

54 5500 2220 Yes

55 5600 2505 Yes

58 5100 2425 Yes

62 6000 2505 Yes

64 6000 2435 Yes

65 5600 2130 Yes

78 6000 2910 Yes

81 5200 3375 Yes

82 5600 2330 Yes

84 5200 3505 Yes

85 5400 2405 Yes

86 5400 2340 Yes

87 5000 2515 Yes

88 5500 2550 Yes

90 5800 2685 Yes

92 5400 2215 Yes

Fuel.tank.capacity Passengers

1 13.2 5

5 21.1 4

40 12.4 4

41 15.9 4

42 11.9 4

43 17.0 4

44 11.9 5

45 13.7 5

46 11.9 4

47 17.2 5

53 13.2 4

54 14.5 5

55 15.5 5

58 14.5 5

62 13.2 5

64 13.2 5

65 15.9 5

78 18.0 5

81 15.9 5

82 15.9 5

84 11.9 5

85 15.9 4

86 18.5 5

87 19.8 7

88 12.4 4

90 18.5 5

92 15.8 5

Length Wheelbase Width

1 177 102 68

5 186 109 69

40 164 97 67

41 175 100 70

42 173 103 67

43 185 107 67

44 168 94 63

45 172 98 66

46 166 94 64

47 184 104 69

53 164 97 66

54 172 98 66

55 184 103 69

58 175 105 67

62 172 98 67

64 170 96 66

65 181 103 67

78 184 99 67

81 175 97 65

82 179 102 67

84 162 94 65

85 174 99 69

86 188 103 70

87 187 113 71

88 163 93 63

90 180 103 67

92 190 104 67

Turn.circle Rear.seat.room

1 37 26.5

5 39 27.0

40 37 24.5

41 39 23.5

42 36 28.0

43 41 28.0

44 35 26.0

45 36 28.0

46 34 23.5

47 41 31.0

53 34 27.0

54 36 26.5

55 40 29.5

58 34 26.0

62 36 26.0

64 33 26.0

65 40 28.5

78 37 26.5

81 35 27.5

82 37 27.0

84 36 24.0

85 39 23.0

86 38 28.5

87 41 35.0

88 34 26.0

90 35 31.5

92 37 29.5

Luggage.room Weight Origin

1 11 2705 non-USA

5 13 3640 non-USA

40 11 2475 non-USA

41 8 2865 non-USA

42 12 2350 non-USA

43 14 3040 non-USA

44 11 2345 non-USA

45 12 2620 non-USA

46 9 2285 non-USA

47 14 2885 non-USA

53 16 2325 non-USA

54 13 2440 non-USA

55 14 2970 non-USA

58 12 2920 non-USA

62 11 2295 non-USA

64 12 2545 non-USA

65 14 3050 non-USA

78 14 2775 non-USA

81 15 2490 non-USA

82 14 3085 non-USA

84 11 2055 non-USA

85 13 2950 non-USA

86 15 3030 non-USA

87 NA 3785 non-USA

88 10 2240 non-USA

90 14 2985 non-USA

92 14 2985 non-USA

Make

1 Acura Integra

5 BMW 535i

40 Geo Storm

41 Honda Prelude

42 Honda Civic

43 Honda Accord

44 Hyundai Excel

45 Hyundai Elantra

46 Hyundai Scoupe

47 Hyundai Sonata

53 Mazda 323

54 Mazda Protege

55 Mazda 626

58 Mercedes-Benz 190E

62 Mitsubishi Mirage

64 Nissan Sentra

65 Nissan Altima

78 Saab 900

81 Subaru Loyale

82 Subaru Legacy

84 Toyota Tercel

85 Toyota Celica

86 Toyota Camry

87 Toyota Previa

88 Volkswagen Fox

90 Volkswagen Passat

92 Volvo 240

**Problem#4.13.** [Page 179]

The batting data set has baseball statistics from 2002. Compute the team batting average (total number of hits (see the column H) divided by total number of at bats (see the column AB)) for each team? [10 pts]

**Note.** Consider the tapply function. Do NOT use the sapply function.

batting\_average <- tapply(batting$H, batting$teamID, function(x) sum(x) / sum(batting$AB[batting$teamID == unique(batting$teamID)]))

print(batting\_average)

ANA ARI ATL   
0.1905004 0.1745830 0.1719088   
 BAL BOS CHA   
0.1619763 0.1952120 0.1860436   
 CHN CIN CLE   
0.1524258 0.1559913 0.1600662   
 COL DET FLO   
0.1749650 0.1563734 0.1682160   
 HOU KCA LAN   
0.1683433 0.1650325 0.1687253   
 MIL MIN MON   
0.1581561 0.1906278 0.1655418   
 NYA NYN OAK   
0.1889724 0.1647778 0.1792945   
 PHI PIT SDN   
0.1747103 0.1562460 0.1629950   
 SEA SFN SLN   
0.1866802 0.1777665 0.1791672   
 TBA TEX TOR   
0.1726729 0.1832421 0.1804406