1.

data(iris)

set.seed(123)

split\_index <- createDataPartition(iris$Species, p = 0.8, list = FALSE)

train\_data <- iris[split\_index, ]

test\_data <- iris[-split\_index, ]

library(caret)

model <- train(Species ~ Petal.

2.

a = c(90, 50, 70, 80, 70, 60, 20, 30, 80, 90, 20)

> print(mean(a))

[1] 60

> print(median(a))

[1] 70

> getmode <- function(v) {

+ uniqv <- unique(v)

+ uniqv[which.max(tabulate(match(v, uniqv)))]

+ }

> print(getmode(a))

[1] 90

> print(max(a))

[1] 90

> x = sort(a)

> print(x[10])

[1] 90

> print(x[9])

[1] 80

3.

x = airquality

> mymedian <- function(lst) {

+ n <- length(lst)

+ s <- sort(lst)

+ ifelse(n%%2==1,s[(n+1)/2],mean(s[n/2+0:1]))

+ }

> head(airquality)

Ozone Solar.R Wind Temp Month Day

1 41 190 7.4 67 5 1

2 36 118 8.0 72 5 2

3 12 149 12.6 74 5 3

4 18 313 11.5 62 5 4

5 NA NA 14.3 56 5 5

6 28 NA 14.9 66 5 6

> print(mymedian(x$Temp))

[1] 79

> print(x[1:5,])

Ozone Solar.R Wind Temp Month Day

1 41 190 7.4 67 5 1

2 36 118 8.0 72 5 2

3 12 149 12.6 74 5 3

4 18 313 11.5 62 5 4

5 NA NA 14.3 56 5 5

> print(x[,1:2:2])

> print(x[order(x[,1]),])

Ozone Solar.R Wind Temp Month Day

21 1 8 9.7 59 5 21

23 4 25 9.7 61 5 23

18 6 78 18.4 57 5 18

11 7 NA 6.9 74 5 11

76 7 48 14.3 80 7 15

147 7 49 10.3 69 9 24

9 8 19 20.1 61 5 9

94 9 24 13.8 81 8 2

114 9 36 14.3 72 8 22

137 9 24 10.9 71 9 14

73 10 264 14.3 73 7 12

13 11 290 9.2 66 5 13

20 11 44 9.7 62 5 20

22 11 320 16.6 73 5 22

3 12 149 12.6 74 5 3

50 12 120 11.5 73 6 19

51 13 137 10.3 76 6 20

138 13 112 11.5 71 9 15

141 13 27 10.3 76 9 18

144 13 238 12.6 64 9 21

14 14 274 10.9 68 5 14

16 14 334 11.5 64 5 16

148 14 20 16.6 63 9 25

151 14 191 14.3 75 9 28

12 16 256 9.7 69 5 12

82 16 7 6.9 74 7 21

95 16 77 7.4 82 8 3

143 16 201 8.0 82 9 20

4 18 313 11.5 62 5 4

15 18 65 13.2 58 5 15

140 18 224 13.8 67 9 17

152 18 131 8.0 76 9 29

8 19 99 13.8 59 5 8

49 20 37 9.2 65 6 18

87 20 81 8.6 82 7 26

130 20 252 10.9 80 9 7

153 20 223 11.5 68 9 30

47 21 191 14.9 77 6 16

113 21 259 15.5 77 8 21

132 21 230 10.9 75 9 9

135 21 259 15.5 76 9 12

108 22 71 10.3 77 8 16

7 23 299 8.6 65 5 7

28 23 13 12.0 67 5 28

44 23 148 8.0 82 6 13

110 23 115 7.4 76 8 18

131 23 220 10.3 78 9 8

145 23 14 9.2 71 9 22

133 24 259 9.7 73 9 10

142 24 238 10.3 68 9 19

74 27 175 14.9 81 7 13

6 28 NA 14.9 66 5 6

105 28 273 11.5 82 8 13

136 28 238 6.3 77 9 13

38 29 127 9.7 82 6 7

19 30 322 11.5 68 5 19

149 30 193 6.9 70 9 26

111 31 244 10.9 78 8 19

24 32 92 12.0 61 5 24

64 32 236 9.2 81 7 3

129 32 92 15.5 84 9 6

17 34 307 12.0 66 5 17

78 35 274 10.3 82 7 17

97 35 NA 7.4 85 8 5

2 36 118 8.0 72 5 2

146 36 139 10.3 81 9 23

31 37 279 7.4 76 5 31

48 37 284 20.7 72 6 17

41 39 323 11.5 87 6 10

93 39 83 6.9 81 8 1

67 40 314 10.9 83 7 6

1 41 190 7.4 67 5 1

104 44 192 11.5 86 8 12

112 44 190 10.3 78 8 20

134 44 236 14.9 81 9 11

29 45 252 14.9 81 5 29

116 45 212 9.7 79 8 24

139 46 237 6.9 78 9 16

128 47 95 7.4 87 9 5

77 48 260 6.9 81 7 16

63 49 248 9.2 85 7 2

90 50 275 7.4 86 7 29

88 52 82 12.0 86 7 27

92 59 254 9.2 81 7 31

109 59 51 6.3 79 8 17

79 61 285 6.3 84 7 18

81 63 220 11.5 85 7 20

66 64 175 4.6 83 7 5

91 64 253 7.4 83 7 30

106 65 157 9.7 80 8 14

98 66 NA 4.6 87 8 6

40 71 291 13.8 90 6 9

118 73 215 8.0 86 8 26

126 73 183 2.8 93 9 3

120 76 203 9.7 97 8 28

68 77 276 5.1 88 7 7

96 78 NA 6.9 86 8 4

125 78 197 5.1 92 9 2

80 79 187 5.1 87 7 19

85 80 294 8.6 86 7 24

89 82 213 7.4 88 7 28

122 84 237 6.3 96 8 30

71 85 175 7.4 89 7 10

123 85 188 6.3 94 8 31

100 89 229 10.3 90 8 8

127 91 189 4.6 93 9 4

124 96 167 6.9 91 9 1

69 97 267 6.3 92 7 8

70 97 272 5.7 92 7 9

86 108 223 8.0 85 7 25

101 110 207 8.0 90 8 9

30 115 223 5.7 79 5 30

121 118 225 2.3 94 8 29

99 122 255 4.0 89 8 7

62 135 269 4.1 84 7 1

117 168 238 3.4 81 8 25

5 NA NA 14.3 56 5 5

10 NA 194 8.6 69 5 10

25 NA 66 16.6 57 5 25

26 NA 266 14.9 58 5 26

27 NA NA 8.0 57 5 27

32 NA 286 8.6 78 6 1

33 NA 287 9.7 74 6 2

34 NA 242 16.1 67 6 3

35 NA 186 9.2 84 6 4

36 NA 220 8.6 85 6 5

37 NA 264 14.3 79 6 6

39 NA 273 6.9 87 6 8

42 NA 259 10.9 93 6 11

43 NA 250 9.2 92 6 12

45 NA 332 13.8 80 6 14

46 NA 322 11.5 79 6 15

52 NA 150 6.3 77 6 21

53 NA 59 1.7 76 6 22

54 NA 91 4.6 76 6 23

55 NA 250 6.3 76 6 24

56 NA 135 8.0 75 6 25

57 NA 127 8.0 78 6 26

58 NA 47 10.3 73 6 27

59 NA 98 11.5 80 6 28

60 NA 31 14.9 77 6 29

61 NA 138 8.0 83 6 30

65 NA 101 10.9 84 7 4

72 NA 139 8.6 82 7 11

75 NA 291 14.9 91 7 14

83 NA 258 9.7 81 7 22

84 NA 295 11.5 82 7 23

102 NA 222 8.6 92 8 10

103 NA 137 11.5 86 8 11

107 NA 64 11.5 79 8 15

115 NA 255 12.6 75 8 23

119 NA 153 5.7 88 8 27

150 NA 145 13.2 77 9 27

> print(min(x$Temp))

[1] 56

> print(max(x$Wind))

[1] 20.7

4.

> summary(airquality)

Ozone Solar.R Wind Temp Month

Min. : 1.00 Min. : 7.0 Min. : 1.700 Min. :56.00 Min. :5.000

1st Qu.: 18.00 1st Qu.:115.8 1st Qu.: 7.400 1st Qu.:72.00 1st Qu.:6.000

Median : 31.50 Median :205.0 Median : 9.700 Median :79.00 Median :7.000

Mean : 42.13 Mean :185.9 Mean : 9.958 Mean :77.88 Mean :6.993

3rd Qu.: 63.25 3rd Qu.:258.8 3rd Qu.:11.500 3rd Qu.:85.00 3rd Qu.:8.000

Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000

NA's :37 NA's :7

Day

Min. : 1.0

1st Qu.: 8.0

Median :16.0

Mean :15.8

3rd Qu.:23.0

Max. :31.0

(ii)Melt airquality data set and display as a long – format data?

> View(ChickWeight)

> names(airquality) <- tolower(names(airquality))

> head(airquality)

ozone solar.r wind temp month day

1 41 190 7.4 67 5 1

2 36 118 8.0 72 5 2

3 12 149 12.6 74 5 3

4 18 313 11.5 62 5 4

5 NA NA 14.3 56 5 5

6 28 NA 14.9 66 5 6

> aql <- melt(airquality)

Using as id variables

> head(aql)

variable value

1 ozone 41

2 ozone 36

3 ozone 12

4 ozone 18

5 ozone NA

6 ozone 28

(iii)Melt airquality data and specify month and day to be “ID variables”?

> aql <- melt(airquality, id.vars = c("month", "day"))

> head(aql)

month day variable value

1 5 1 ozone 41

2 5 2 ozone 36

3 5 3 ozone 12

4 5 4 ozone 18

5 5 5 ozone NA

6 5 6 ozone 28

(iv)Cast the molten airquality data set with respect to month and date features

aql <- melt(airquality, id.vars = c("month", "day"))

> aqw <- dcast(aql, month + day ~ variable)

> ## month day ozone solar.r wind temp

> ## 1 5 1 41 190 7.4 67

> ## 2 5 2 36 118 8.0 72

> ## 3 5 3 12 149 12.6 74

> ## 4 5 4 18 313 11.5 62

> ## 5 5 5 NA NA 14.3 56

> ## 6 5 6 28 NA 14.9 66

5.

library(dplyr)

data("airquality")

summary(airquality)

airquality\_clean <- airquality %>%

drop\_na()

prop\_na <- sum(is.na(airquality)) / nrow(airquality)

if (prop\_na > 0.1) {

airquality[is.na(airquality)] <- sapply(airquality, function(x) mean(x, na.rm = TRUE))

}

model <- lm(Ozone ~ Solar.R, data = airquality\_clean)

library(ggplot2)

ggplot(airquality\_clean, aes(x = Solar.R, y = Ozone)) +

geom\_point() +

geom\_smooth(method = "lm", formula = y ~ x, se = FALSE)

6.

data("ChickWeight")

grouped\_data <- group\_by(ChickWeight, diet)

ordered\_data <- arrange(grouped\_data, weight)

last\_6\_records <- slice\_tail(ordered\_data, 6)