> # Read the CSV into a data frame object

> (toy\_data <- read.csv("toy\_dataset.csv"))

Number City Gender Age Income Illness

1 1 Dallas Male 41 40367 No

2 2 Dallas Male 54 45084 No

3 3 Dallas Male 42 52483 No

4 4 Dallas Male 40 40941 No

5 5 Dallas Male 46 50289 No

6 6 Dallas Female 36 50786 No

7 7 Dallas Female 32 33155 No

8 8 Dallas Male 39 30914 No

9 9 Dallas Male 51 68667 No

10 10 Dallas Female 30 50082 No

11 11 Dallas Female 48 41524 Yes

12 12 Dallas Male 47 54777 No

13 13 Dallas Male 46 62749 No

14 14 Dallas Female 42 50894 No

15 15 Dallas Female 61 38429 No

16 16 Dallas Male 43 34074 No

17 17 Dallas Male 27 50398 No

18 18 Dallas Male 38 46373 Yes

19 19 Dallas Male 47 51137 No

20 20 Dallas Female 35 23688 No

21 21 Dallas Male 57 17378 No

22 22 Dallas Male 33 45919 No

23 23 Dallas Female 33 23001 No

24 24 Dallas Female 27 34292 Yes

25 25 Dallas Male 58 55190 No

26 26 Dallas Male 64 26169 No

27 27 Dallas Male 58 57322 No

28 28 Dallas Male 44 61704 No

29 29 Dallas Male 34 53619 No

30 30 Dallas Male 45 47421 Yes

31 31 Dallas Female 44 40353 No

32 32 Dallas Male 39 28125 No

33 33 Dallas Female 55 42630 No

34 34 Dallas Male 27 56645 No

35 35 Dallas Female 63 41946 No

36 36 Dallas Male 41 50312 No

37 37 Dallas Male 64 47872 No

38 38 Dallas Female 41 29538 No

39 39 Dallas Female 61 39881 Yes

40 40 Dallas Female 59 48518 No

41 41 Dallas Female 26 16168 Yes

42 42 Dallas Male 41 68522 No

43 43 Dallas Female 47 50750 No

44 44 Dallas Female 58 49614 No

45 45 Dallas Male 33 56169 No

46 46 Dallas Female 30 40661 No

47 47 Dallas Male 51 53730 No

48 48 Dallas Female 45 34613 No

49 49 Dallas Female 38 35249 No

50 50 Dallas Female 56 52218 No

51 51 Dallas Female 55 47702 No

52 52 Dallas Male 42 62512 No

53 53 Dallas Female 58 48655 No

54 54 Dallas Male 57 55133 No

55 55 Dallas Male 32 39309 No

56 56 Dallas Female 55 41036 No

57 57 Dallas Male 62 60594 No

58 58 Dallas Female 43 46373 No

59 59 Dallas Male 35 53303 No

60 60 Dallas Female 31 42281 No

61 61 Dallas Female 46 48268 No

62 62 Dallas Male 49 36223 No

63 63 Dallas Female 54 36552 No

64 64 Dallas Female 63 36222 No

65 65 Dallas Female 58 53333 No

66 66 Dallas Female 53 40244 No

67 67 Dallas Male 61 53713 No

68 68 Dallas Male 44 42823 No

69 69 Dallas Male 51 44030 Yes

70 70 Dallas Female 44 25604 No

71 71 Dallas Female 38 31953 Yes

72 72 Dallas Female 57 46137 No

73 73 Dallas Male 44 50089 No

74 74 Dallas Male 27 53213 No

75 75 Dallas Female 41 27897 No

76 76 Dallas Male 43 26997 No

77 77 Dallas Male 46 29374 No

78 78 Dallas Male 55 44003 No

79 79 Dallas Male 53 58625 No

80 80 Dallas Female 43 52935 No

81 81 Dallas Male 46 60637 No

82 82 Dallas Male 46 27092 Yes

83 83 Dallas Male 43 49741 No

84 84 Dallas Female 51 31416 No

85 85 Dallas Female 40 46882 No

86 86 Dallas Female 34 33246 No

87 87 Dallas Female 58 32567 No

88 88 Dallas Male 29 49798 No

89 89 Dallas Male 28 38119 No

90 90 Dallas Female 30 27446 No

91 91 Dallas Male 28 54937 No

92 92 Dallas Female 25 34653 No

93 93 Dallas Male 37 63495 No

94 94 Dallas Female 62 38385 No

95 95 Dallas Male 45 40789 No

96 96 Dallas Male 48 62287 No

97 97 Dallas Female 30 47982 No

98 98 Dallas Male 44 47860 Yes

99 99 Dallas Male 62 49231 No

100 100 Dallas Female 48 42252 Yes

101 101 Dallas Male 46 68275 No

102 102 Dallas Male 48 48433 No

103 103 Dallas Male 61 42379 No

104 104 Dallas Male 63 44965 No

105 105 Dallas Female 56 40122 No

106 106 Dallas Male 63 48512 No

107 107 Dallas Female 48 47466 Yes

108 108 Dallas Male 47 30382 No

109 109 Dallas Female 38 46733 No

110 110 Dallas Female 64 40960 No

111 111 Dallas Male 46 39085 No

112 112 Dallas Female 59 49691 No

113 113 Dallas Male 39 53649 No

114 114 Dallas Male 38 37010 No

115 115 Dallas Male 58 46788 Yes

116 116 Dallas Male 59 61700 No

117 117 Dallas Male 48 58460 Yes

118 118 Dallas Female 31 19155 No

119 119 Dallas Female 54 56868 No

120 120 Dallas Male 46 46954 No

121 121 Dallas Male 43 64376 Yes

122 122 Dallas Male 38 35911 No

123 123 Dallas Female 28 48836 No

124 124 Dallas Male 62 62174 No

125 125 Dallas Female 62 46385 No

126 126 Dallas Female 32 52215 No

127 127 Dallas Male 45 36702 No

128 128 Dallas Female 31 36979 No

129 129 Dallas Male 37 58357 No

130 130 Dallas Male 44 59426 No

131 131 Dallas Female 59 48994 No

132 132 Dallas Male 37 61390 No

133 133 Dallas Female 35 49629 No

134 134 Dallas Female 65 33676 No

135 135 Dallas Female 40 45797 No

136 136 Dallas Male 59 45130 Yes

137 137 Dallas Male 56 43573 No

138 138 Dallas Female 56 57028 No

139 139 Dallas Male 53 68679 No

140 140 Dallas Male 45 51435 No

141 141 Dallas Male 60 48559 No

142 142 Dallas Female 53 38692 Yes

143 143 Dallas Female 37 46382 No

144 144 Dallas Female 40 48550 No

145 145 Dallas Female 56 59945 No

146 146 Dallas Female 28 51127 No

147 147 Dallas Male 42 26301 No

148 148 Dallas Male 60 50057 No

149 149 Dallas Female 58 31472 No

150 150 Dallas Female 48 42926 No

151 151 Dallas Male 37 60150 No

152 152 Dallas Male 51 38416 No

153 153 Dallas Female 54 43810 No

154 154 Dallas Male 63 71551 Yes

155 155 Dallas Female 36 25564 Yes

156 156 Dallas Female 32 28551 No

157 157 Dallas Male 48 45591 No

158 158 Dallas Female 51 47559 No

159 159 Dallas Male 59 53835 No

160 160 Dallas Male 29 55545 No

161 161 Dallas Female 27 57903 Yes

162 162 Dallas Male 49 63544 No

163 163 Dallas Female 54 39273 No

164 164 Dallas Female 63 38841 No

165 165 Dallas Female 47 48520 No

166 166 Dallas Male 55 54849 No

[ reached 'max' / getOption("max.print") -- omitted 149834 rows ]

> # Attach the data frame to the environment

> attach(toy\_data)

> # Count the number of rows with City = New York City

> (nrow(toy\_data[toy\_data$City == "New York City",]))

[1] 50307

> #or this

> (nrow(subset(toy\_data,City == "New York City")))

[1] 50307

> # Display rows that have top 5 income values

> (head(toy\_data[order(-toy\_data$Income),], 5))

Number City Gender Age Income Illness

109351 109351 Mountain View Male 58 177157 No

105282 105282 Mountain View Male 41 176746 No

109061 109061 Mountain View Male 61 173991 No

110878 110878 Mountain View Male 52 173826 No

112193 112193 Mountain View Male 58 172825 No

> # Find out the number of rows with income greater than average income

> (avg\_income <- mean(toy\_data$Income))

[1] 91252.8

> (nrow(toy\_data[toy\_data$Income > avg\_income,]))

[1] 83631

> #or this

> (nrow(subset(toy\_data,Income>avg\_income)))

[1] 83631

> # Find the highest salary for the female population

> (max(toy\_data[Gender == "Female",]$Income))

[1] 168440

> #or this

> (max(Income[Gender == "Female"]))

[1] 168440

> sample\_rows <- sample(nrow(toy\_data), 10)

> (toy\_data[sample\_rows,])

Number City Gender Age Income Illness

148397 148397 Austin Male 31 89773 No

147429 147429 Austin Male 50 82498 No

64337 64337 New York City Female 28 70567 No

76557 76557 Los Angeles Female 50 74851 Yes

52845 52845 New York City Male 61 116935 No

23757 23757 New York City Male 47 104273 No

33129 33129 New York City Male 32 110785 No

94117 94117 Los Angeles Male 41 100183 No

67179 67179 New York City Male 36 82791 No

148284 148284 Austin Male 63 91102 No

> # Detach the data frame object from the environment

> detach(toy\_data)