1.8

a) Each row represents a single case, or single UK resident.

b) 1691 participants were included

c) sex (categorical), age (numerical; discrete), marital (categorical), grossIncome (categorical), smoke (categorical), amtWeekends (categorical), amtWeekdays (categorical)

\*For the last 2 variables, amtWeekends & amtWeekdays, I’m assuming they selected from options in the survey on how many cigs/day. If the fields only included the numbers (ie 12, 6, 8, etc) I would consider these discrete numerical fields.

1.10

a) population: all children between the ages of 5 and 15; sample: 160 children between the ages of 5 and 15

b) if the 160 children in the sample are representative of all children between the ages of 5 and 15, then the results can be generalized to the population. However, causal relationships can only be inferred from randomized experiments and I am not sure this experiment was randomized because there are notable differences in characteristics between the two groups.

1.28

a) no, the data were obtained through an observational study so we cannot conclude causal relationships.

b) No, the statement is not justified. The study suggests an association between sleep disorders and behavioral issues (bullying)

1.36

a) experimental

b) treatment: group instructed to exercise twice a week; control: group told not to exercise

c) yes, the age groups.

d) no

e) It is a randomized experiment so you can establish casual relationship and since the sample is random, the results can be generalized to the population.

f) yes, cant just tell people to and not to exercise

1.48

>exams <- c(57,66,69,71,72,73,74,77,78,78,79,79,81,81,82,83,83,88,89,94)

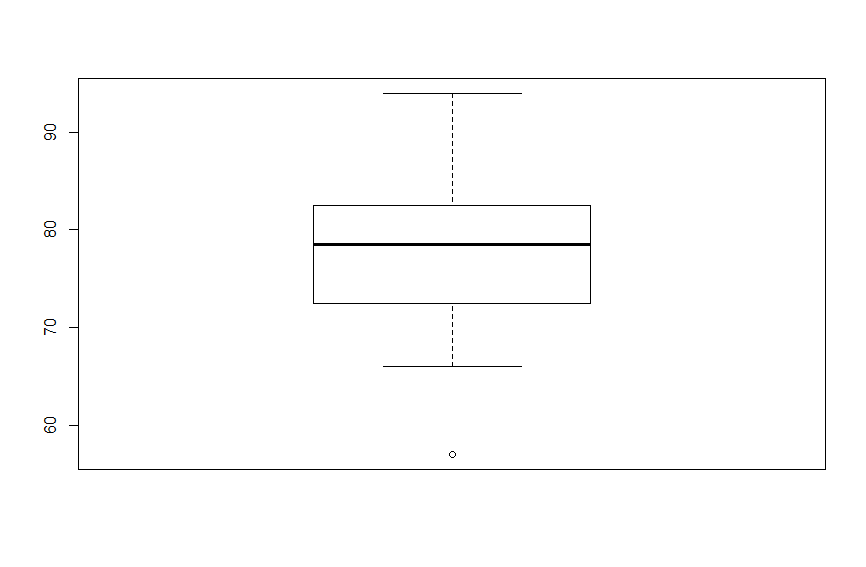
> summary((exams))

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

57.00 72.75 78.50 77.70 82.25 94.00

> boxplot(exams)

(see next page)



1.50

a) symmetric - 2

b) uniform - 3

c) right skewed - 1

1.56

a) The distribution is likely right skewed. The center would be best described by the median and variability would be best described by the IQR.

b) The distribution is likely symmetric. The center would be best described by the mean and variability would be best described by the standard deviation.

c) The distribution is likely right skewed. The center would be best described by the median and variability would be best described by the IQR.

d) The distribution is likely symmetric. The center would be best described by the mean and variability would be best described by the standard deviation.

1.70

a) no, we can see that survival and transplant are associated (or dependent) because the columns are divided in different vertical locations

b)

c) Proportion of patients in treatment group that died = 0.6522

Proportion of patients in control group that died = 0.8824

d) i)

ii) 28; 75; ; ; 0;

iii)