

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; ordinal), smoke (categorical), amtWeekends (numerical; discrete), amtWeekdays (numerical; discrete)

```
smoking <- read_csv("~/GitHub/DATA606Spring2017/Data/Data from openintro.org/  
Ch 1 Exercise Data/smoking.csv")
```

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. Causal relationships can only be inferred from randomized experiments and it doesn't appear they controlled for the 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. In addition, the members voluntary participants which may introduce bias.
- b) No, the statement is not justified. The study suggests an association between sleep disorders and behavioral issues (bullying).

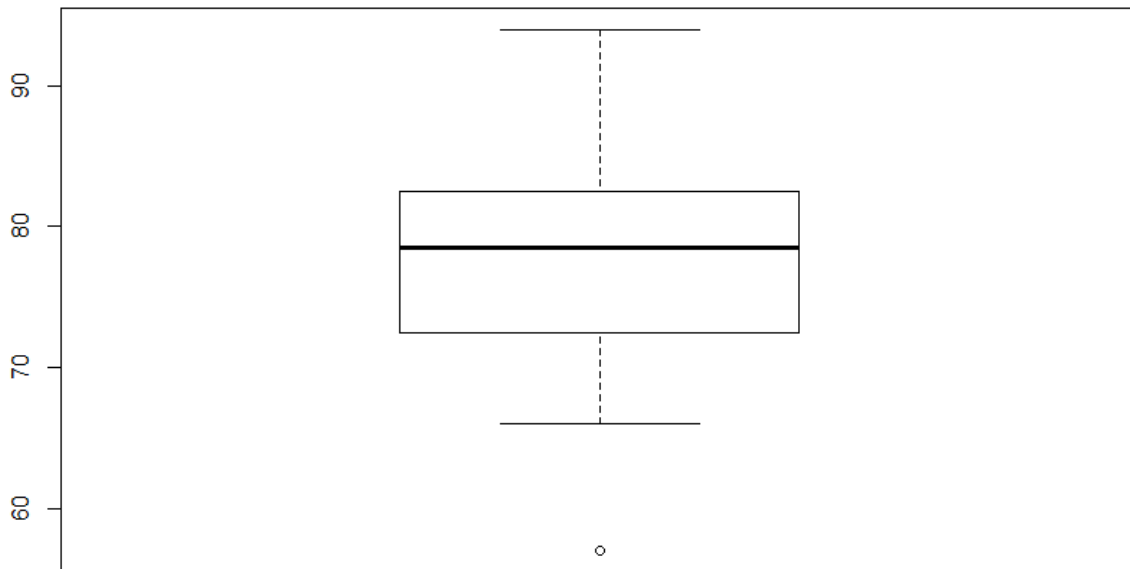
1.36

- 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 a causal relationship. The sample is also random, so the results can be generalized to the population.
- f) Yes, the experiment does not ensure that the participants are blind to the study. There is no "placebo" for not working out so that group may be aware they are in the control group. Also, it is a human study which can introduce bias to the results. Additionally, it doesn't account for whether the person regularly exercises. If they are placed in the control group, it may have a larger effect on the mental health exam.

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, unimodal - 2
- b) uniform - 3
- c) right skewed, unimodal - 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) The boxplots suggest that median number of days of survival is higher for the treatment group than for the control. They also suggest there is larger variability observed by the longer IQR. Overall, besides the few outliers in the control group, 75% of the participants in the treatment group, survived longer than the participants in the control group.

	Control	Treatment	Total
Alive	4	24	28
Dead	30	45	75
<b>Total</b>	<b>34</b>	<b>69</b>	<b>103</b>

`table(heartTr$transplant,heartTr$survived)`

```

      alive dead
control    4   30
treatment 24   45

```

c) Proportion of patients in treatment group that died =  $p_t = \frac{45}{69} \approx 0.6522$

Proportion of patients in control group that died =  $p_c = \frac{30}{34} \approx 0.8824$

d) i)  $H_0$ : The experimental heart transplant program has no effect on lifespan.  
 $H_A$ : The experimental heart transplant program has an effect on lifespan.

ii) 28; 75; 69; 34; 0; equal to or more extreme than the difference we observed.

iii)  $p_t - p_c = 0.6522 - 0.8824 = -0.2302$

The simulations show that a difference of at least -0.2302 due to chance alone would only happen about 2% of the time (2 out of the 100 simulations). This is a low probability which indicates a rare event. We can conclude that the evidence is sufficiently strong to reject  $H_0$ . Therefore, there the experimental heart transplant program does have an effect on lifespan.