HW1 Chapter 1

anjal hussan

9/11/2017

```
smoking <- read.csv('https://raw.githubusercontent.com/jbryer/DATA606Fall2017/master/Dat
a/Data%20from%20openintro.org/Ch%201%20Exercise%20Data/smoking.csv')
```

1.8:

a.

```
names(smoking)
```

```
##  [1] "gender"              "age"              "maritalStatus"
##  [4] "highestQualification" "nationality"      "ethnicity"
##  [7] "grossIncome"         "region"           "smoke"
## [10] "amtWeekends"         "amtWeekdays"       "type"
```

Each row represents a study on an individual looking into their gender, age, matiral status, highestQualification, nationality, ethnicity, gross income, region, smoking habits, etc.

b.

```
nrow(smoking)
```

```
## [1] 1691
```

Total Number of participants were included in the survery is: 1,691

c.

Gender: Categorical - not ordinal Age: Numerical - Continious Marital Status: Categorical - not ordinal Highest Qualification: Categorical - ordinal Nationality: Categorical - not ordinal Ethnicity: Categorical - not ordinal GrossIncome - Numerical - numerical - continious Region: Categorical - not ordinal Smoke - Categorical - not ordinal amtWeekends: Numerical Discrete amtWeekedays: Numerical Discrete

---

1.10:

```
countyComplete <- read.csv('https://raw.githubusercontent.com/jbryer/DATA606Fall2017/mas
ter/Data/Data%20from%20openintro.org/Ch%201%20Exercise%20Data/countyComplete.csv')
```

a.

Population: children between the ages of 5 and 15 Sample size: 160 children

b.

Any results from this study cannot be generalized for the population since the sample was not randomly selected. The children were not selected randomly to give the instruction not to cheat. No casual relationships can be established with these findings.

1.28

a.

We can't conclude that smoking causes dementia later in life by using this study. The subjects were voluntary and not chosen randomly. Also there could be other factors involved other than smoking, family history, health, etc.

b.

This statement is not justified because this article doesn't imply that sleeping disorders directly causes bullying. There may be a correlation between the two but Certainly there are more factors involved than just sleep disorder.

1.36

a.

Experiment ####b. Treament: Exercise Control:No Exercise ####c. Yes, age is the blocking variable. ####d. No blinding was not use since one group knows that they need to exercise. ####e. Selection and assignment were done randomly, hence, the results of this study can be used to establish a causal relationship between exercise and mental health, and the conclusions can be generalized to the population at large. ####f. I would not have any reservations about this proposal since the study was conducted using the random populations.

1.48

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   57.00   72.75   78.50   77.70   82.25   94.00
```

```
boxplot(Listofstats)
```
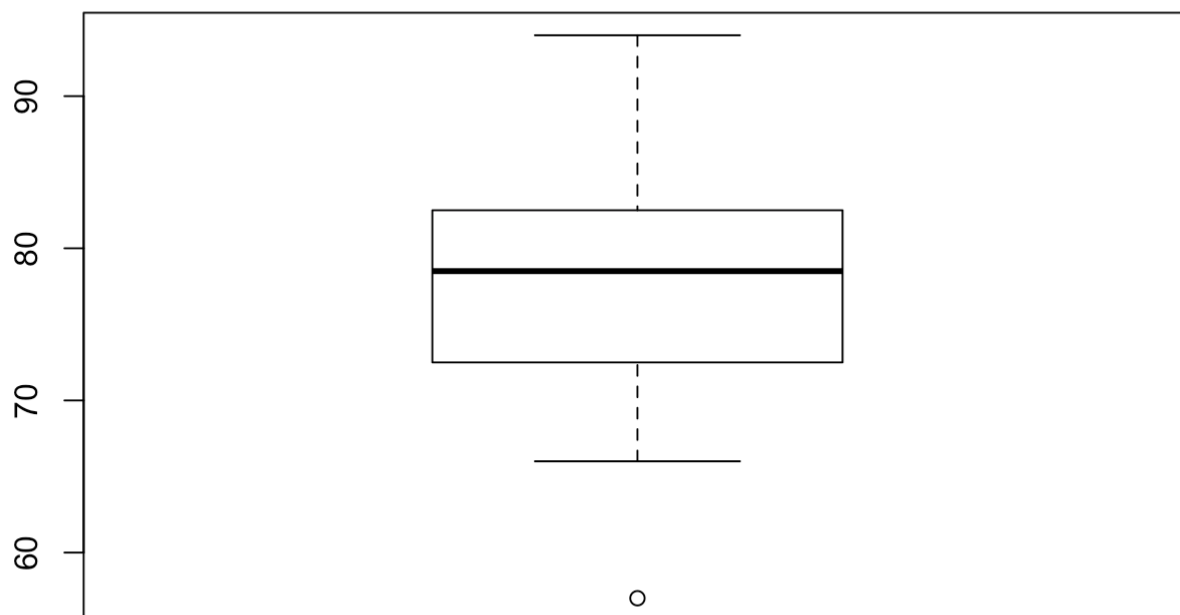
1.50

a:

Symmetrical = box plot 2 ####b: Multimodal = box plot 3 ####c: Right Skewed = box plot 1

1.56

a.

```
1000000 + 1.5*(1000000-650000)
```

```
## [1] 1525000
```

This will be skewed to the right since a significant number of houses are greater than $6M. ####b. This will be a symmetric distribution with most house prices evenly distributed. Mean and standard deviation sholdb be used. ####c This will be skewed to the right with since there is no excesive drinking. Median and IQR can be used. ####d. This should be symmetric distribution will salaries evenly distributed with some outliers. Median and IQR can be used to deal with outliers.

End of file