Homework Assignment 2

Helinda He

October 21, 2022

Question 1:

a)

This histogram is unimodal with the symmetric shape which means that the data only has one peak. The range of the data is from 50-70. The histogram show that there are potential outlier. This histogram is matching with (2) in box plots.

b)

This histogram is uniform because those data are very close to each other. The range of the data is from 0 to 100. This histogram is matching with (3) in the box plots.

c)

This histogram is unimodal with the right skew which means that the data only has one peak and its mean is greater than the median. The range of the data is from 0 to 6. This histogram is matching with (1) in box plots.

Question 2:

a)

The distribution of the graph will be right skewed because there will be some potential outlier. The median will best represent of the data because there are some potential huge outlier that will influence the mean. Since the median is a good representation of the data, the IQR will be more suitable for the data.

b)

The distribution of the graph will be symmetric because there are not so many of he expensive house that can influence the mean. The mean will best represent the observation of the data since the distribution is symmetric. With the symmetric structure, the standard division will be suitable under this situation.

c)

The distribution of the graph will be right skewed because there are a few people who drink excessively when most of people do not drink. The people who drink a lot will influence the mean so the median will better represent the typical observation. The IQR will best represent the variability of observation.

d)

The distribution of the graph will be right skewed because there are only a few people that earn much more money than other people. Since there are potential outliers, the median will be a good representation of the typical observation and the IQR will best represent the variability of observation.

Question 3:

a)

The median will best represent as a typical income of the coffee shop because there are two huge potential outliers which will highly influence the mean and that two huge amount of income makes the histogram shift from symmetric to right skewed. People can say that the median is suitable for a data set that is skewed and the mean is more suitable for symmetric graphs. The median do not easily be influence by the outliers.

b)

The IQR will best represent the amount of variability in the incomes because the standard division is based on the mean and the mean is influenced by the two new people in the coffee shop. People can say that the IQR does not influenced by the two extreme value and it is more stable compare to standard division.

Question 4:

a)

372 / 910 * 100% = 41%

b)

278 / 910 * 100% = 31%

c)

57 / 910 * 100% = 6%

d)

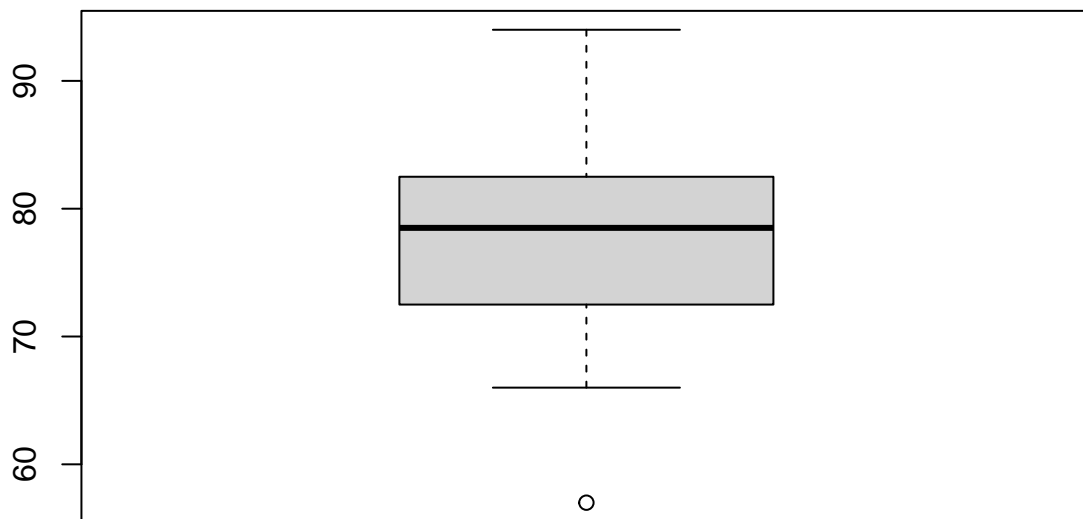conservatives: 57 / 371 * 100% = 15% moderates: 120 / 363 * 100% = 33% liberal: 101 / 175 * 100% = 58%

e)

The political ideology and views on immigration are dependent because the political ideology can have influence on people's thoughts about immigration.

Question 5:

```
data1 <- c(57, 66, 69, 71, 72, 73, 74, 77, 78, 78, 79, 79, 81, 81, 82, 83, 83, 88, 89, 94)
```

```
boxplot(data1)
```


Question 6:

a)

In the histogram, it shows the number of the peak and it can also show the number of people with times. People can see the distribution of data. In the box plot, there is the median, Q1 and Q3. The box plot can also show the outliers and whiskers.

b)

The bimodal means that there are 2 peaks in the histogram. From the histogram of the given data, there are 2 peaks in the data because the data include the time of both women and men. Men usually runs faster

than women.

c)

From the separate box plot we can see that on average that men takes less time to finish the race. The time of men has a smaller range while the time of women has a larger range.

d)

People can see from the time series plot that there is a gap between women and men. Also, from the plot people can tell that the time that both men and women use are decreasing.

Question 7:

a)

median - there is a heart rates which is much higher than the other 4

b)

mean - the journalist wants to show that the basketball player is overpaid by average. We need to use mean to fit her need.

c)

median - there might be some people that lives in a very expensive house which can be potential outliers of the data set

d)

median - because there might be some patients who can live much longer than other patients. Under that situation, those patients data will influence the mean.

Question 8:

```
data2 <- c(25, 27, 31, 26, 33, 38, 28, 48, 24, 32, 38, 71, 62, 40)
```

a)
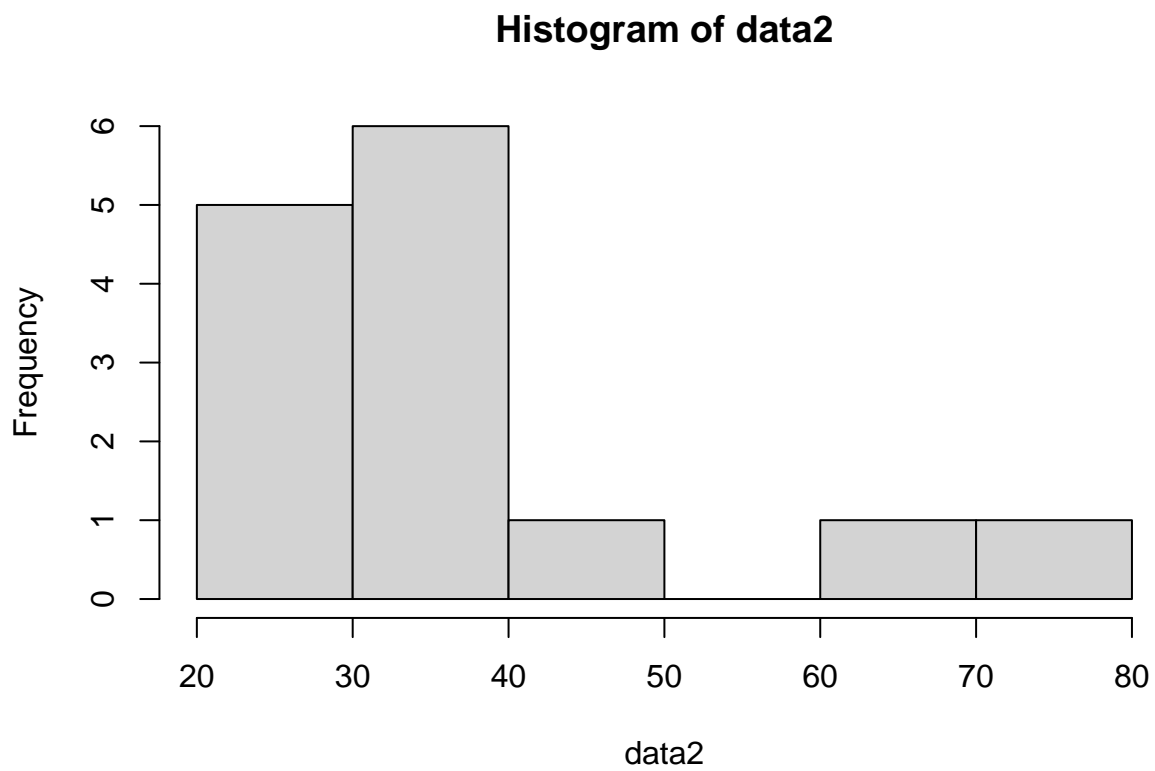
The mean is 37.35714. The median is 32.5.

```
mean(data2)
```

```
## [1] 37.35714
```

```
median(data2)
```

```
## [1] 32.5
```

```
hist(data2)
```

Histogram of data2

b)

The IQR is 12.25. The variance is 200.2473. The standard division is 14.15087.

```
IQR(data2)
```

```
## [1] 12.25
```

```
var(data2)
```

```
## [1] 200.2473
```

```
sd(data2)
```

```
## [1] 14.15087
```