Mat 354

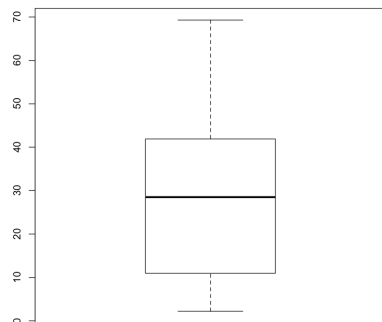### Homework 1: Chapter 1) 56, 64, 66, 70, 88

Kenny Roffo

Due September 2, 2015

56) Analyze the given tree data.
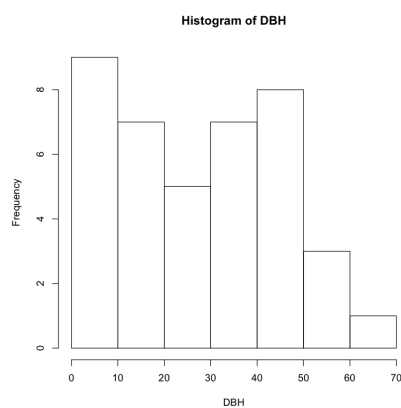
a. Find the five number summary:

```
> fivenum(DBH)
[1]  2.20 10.95 28.50 41.90 69.30
```

b. Make a box plot:

c. Make a histogram:

d. Write a short summary of the major features of the distribution:

The distribution seems to have many values on the at the low end and the numbers decrease towards the higher end (I guess this is called left skewed?). I prefer the histogram since I am used to them, and am very inexperienced with boxplots.

64) Analyze the CO2 data:

a. Find the five number summary:

```
> fivenum(co2)
[1]  0.00  0.75  3.20  7.80 19.90
```

This summary suggests that the distribution is right skewed because the distance between the third quartile and the mean is much greater than the distance between the first quartile and the mean.

b. What are the outliers. Make a histogram. Do you agree with the suspected outliers?:

The IQR is 7.05, and Q_3 is 7.80 so any values above 14.85 are suspected outliers. These are Australia, Canada, and the United States.

Histogram of co2
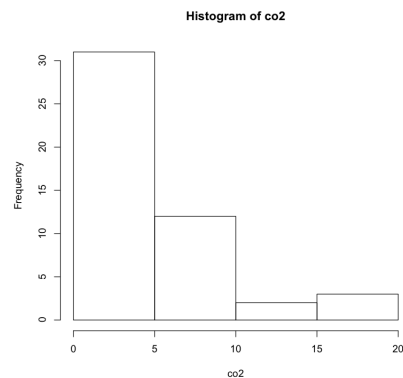
I agree with the suspected outliers, since according to my knowledge of statistics that is what is correct.

66) Analyze oil data.

a. Find the mean and median. Explain how their relationship shapes the distribution:

```
> mean(oil)
[1] 48.24859
> median(oil)
[1] 37.8
```

The median being lower than the mean implies that most of the values are below the mean. This means the distribution is skewed towards the left.

b. Give the five number summary and explain how it reflects the shape of the distri-

bution:
```
> fivenum(oil)
[1]    2.000  21.505  37.800  60.100 204.900
```
The mean is closer to the first quartile than the third, which reflects a leftward skew. Also the very high maximum displays that there is a small number of extremely high values.

70) Calculate the mean and standard deviation for the following metabollic rates: 1792 1666 1362 1614 1460 1867 1439

The mean is calculated by simply summing the values and dividing by the number of how many there are (here 7). The mean here is 1600. The deviations are found by subtracting the mean from each value:

$$1792 - 1600 = 192 \rightarrow (1792 - 1600)^2 = 36864$$
$$1666 - 1600 = 66 \rightarrow (1666 - 1600)^2 = 4356$$
$$1362 - 1600 = -238 \rightarrow (1362 - 1600)^2 = 56644$$
$$1614 - 1600 = 14 \rightarrow (1614 - 1600)^2 = 196$$
$$1460 - 1600 = -140 \rightarrow (1460 - 1600)^2 = 19600$$
$$1867 - 1600 = 267 \rightarrow (1867 - 1600)^2 = 71289$$
$$1439 - 1600 = -161 \rightarrow (1439 - 1600)^2 = 25921$$

Note that the sum of the deviations (the unsquared value) is 0. Since we are trying to get some sort of a feel for an average ërrorf̈rom the mean, we square the values to prevent this. Now we get the variance by simply dividing the sum of the squared deviations by $n-1$, so here by 6. This gives a variance of

$$\sigma^2 = 5968.61$$

Now to get the standard deviation we take the square root to give

$$\sigma = 77.25$$

.

88) Show that the average distance from the mean of a value x, $(x_i - \bar{x})$, is always 0: Consider a number of values x. The mean of these values, $\bar{x}$, is the sum of them all divided by the number of them, n. The sum of their deviations is $x_i - \bar{x}$ where x_i is the ith value. Adding these looks like this:

$$(x_0 - \frac{\Sigma x_i}{n}) + (x_1 - \frac{\Sigma x_i}{n}) + ... + (x_n - \frac{\Sigma x_i}{n}) = \Sigma x_i - n\left(\frac{\Sigma x_i}{n}\right)$$
$$= \Sigma x_i - \Sigma x_i$$
$$= 0$$

3