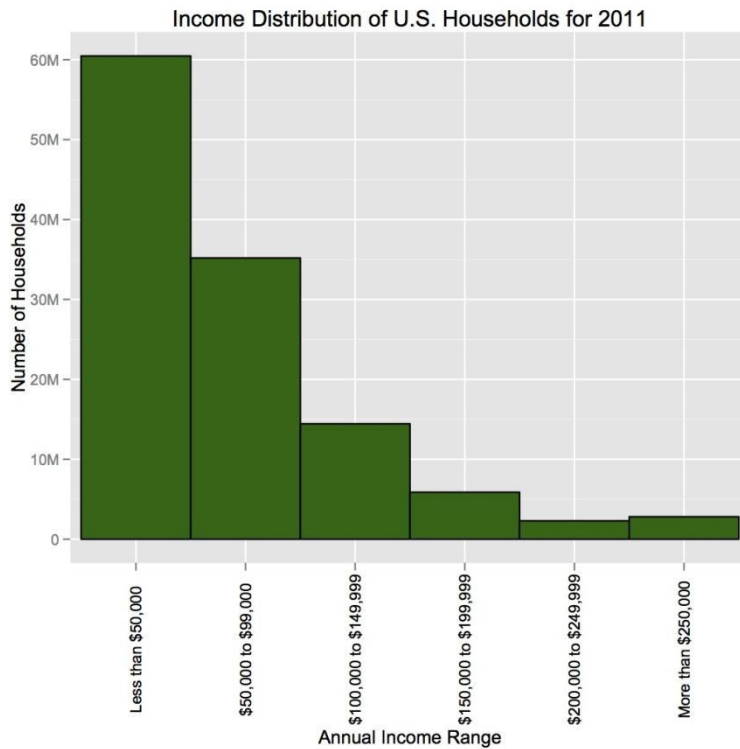


**STAT 113 Week 6 (Feb 11)**  
**Work Sheet 4: Chapter 11 & 12**

➤ **Graphs for Quantitative Variables**

- The following data was obtained from the census website. Data on the household income for approximately 121 million households were collected. The mean income for this data set is \$69,677.



- Is the histogram symmetric, left skewed or right skewed?

**Right skewed**

- Is the histogram unimodal, bimodal, or multimodal?

**Unimodal**

- Approximately how many households made more than \$50,000? What does this tell you about the median income?

**Approximately 60 million households made more than \$50,000, median is about \$50,000.**

- Is the mean income for 2011 higher or lower than the median income?

**Mean is higher than median**

- Which would be more appropriate description of center and spread for this data set: The mean and standard deviation or the 5-number summary? Why?

**5-number summary would be more appropriate because the data is skewed**

➤ **Describing Data with Numbers**

- Below are survival times (in days) of 13 guinea pigs that were injected with a bacterial infection in a medical study:

91 83 84 79 91 93 95 97 97 111 101 105 98

- Find the 5-number summary for this data set.

**79 83 84 91 91 93 95 97 97 98 101 105 111**

Min    Q1    Q2    Q3    Max  
 \_\_\_79\_\_\_    \_\_\_87.5\_\_\_    \_\_\_95\_\_\_    \_\_\_99.5\_\_\_    \_\_\_111\_\_\_

(c) Draw a stemplot of the data and describe the shape of the distribution.

7 | 9  
 8 | 3 4  
 9 | 1 1 3 5 7 7 8  
 10 | 1 5  
 11 | 1

It is unimodal and symmetric.

(d) Are there any outliers in the data set above? Use the 1.5 IQR rule to check.

$$\text{Upper limit} = Q3 + 1.5(\text{IQR}) = 99.5 + 1.5 * (99.5 - 87.5) = 117.5$$

$$\text{Lower limit} = Q1 - 1.5(\text{IQR}) = 87.5 - 1.5 * (99.5 - 87.5) = 69.5$$

No outliers.

(e) Which would be more appropriate description of center and spread for this data set: The mean and standard deviation or the 5-number summary? Why?

The mean and standard deviation since the data is symmetric without outliers.

3. We have a class of 30 students and the data below shows the height (in cm) distribution of those people. The data has already been sorted from lowest to highest.

132	151	151	152	156	156	157	160	161	162
163	163	165	167	167	169	171	172	175	175
177	177	178	183	186	189	189	189	197	206

(a) Find the 5-number summary for this data set.

Min    Q1    Q2    Q3    Max  
 \_\_\_132\_\_\_    \_\_\_160\_\_\_    \_\_\_168\_\_\_    \_\_\_178\_\_\_    \_\_\_206\_\_\_

(b) Find the mode for this data set.

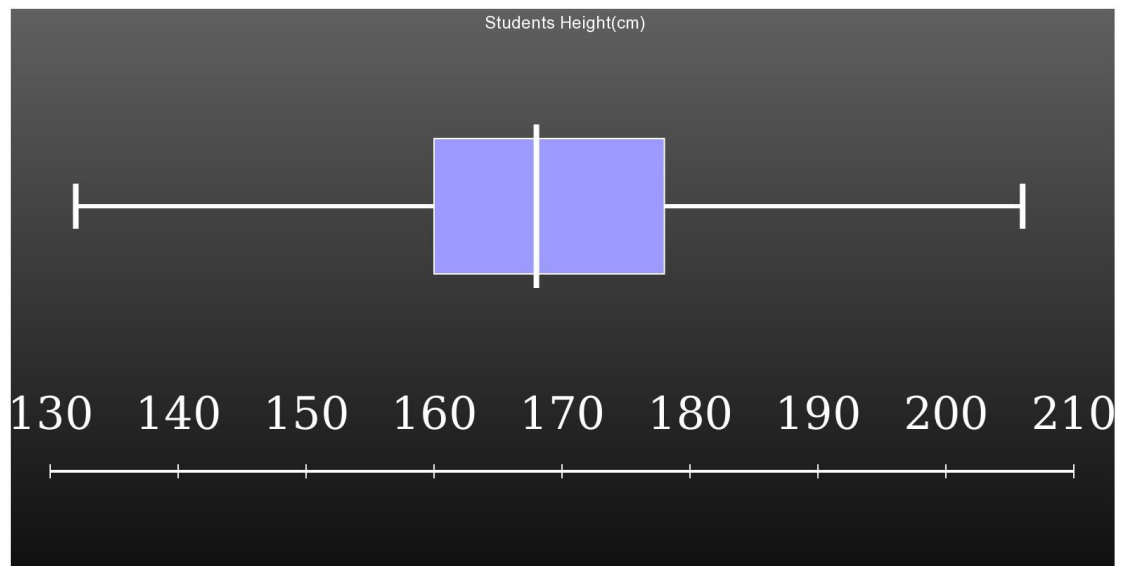
Mode=189

(c) Are there any outliers in the data set above? Use the 1.5 IQR rule to check.

$$\text{Upper limit} = Q3 + 1.5(\text{IQR}) = 178 + 1.5 * (178 - 160) = 205 \Rightarrow \text{upper outlier: } 206$$

$$\text{Lower limit} = Q1 - 1.5(\text{IQR}) = 160 - 1.5 * (178 - 160) = 133 \Rightarrow \text{lower outlier: } 132$$

(d) Draw a boxplot of the data.



4. For the following set of 20 numbers:

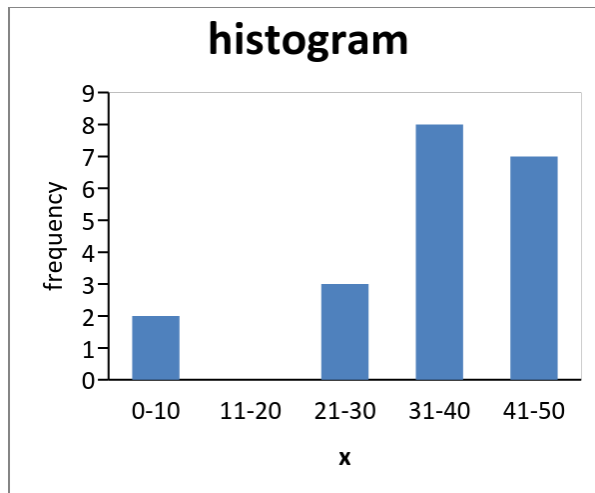
1, 3, 20, 23, 25, 30, 30, 31, 32, 33, 33, 34, 34, 40, 40, 42, 43, 43, 44, 44

(a) Draw a stemplot of the data and describe the shape of the distribution.

```
0| 1 3
1|
2| 0 3 5
3| 0 0 1 2 3 3 4 4
4| 0 0 2 3 3 4 4
```

The shape is left skewed.

(b) Create a histogram by hand.



- (c) Between stemplot and histogram, which plot would better display the data?  
For a small number of observations like this data set, a stemplot would be preferred, it is quicker to make and presents more detailed information.