

12.9. This means 50% of the US households had an income of less than \$68,145 in 2017. Fifty percent (50%) of the US households had an income more than \$68,145. Note that this is a statement about the population of all US households in 2017 (not a sample of households).

12.14. (a) We constructed this stemplot in Exercise 11.13:

```
> mileage = c(24,22,27,13,21,22,14,22,29,31,18,88,88,42,31,47,37,22,18,30,24,25,
  30,84,23,31,114,15,50,95,22)
> stem(mileage,scale=2)

 1 | 34588
 2 | 122222344579
 3 | 001117
 4 | 27
 5 | 0
 6 |
 7 |
 8 | 488
 9 | 5
10 |
11 | 4
```

(b) Here is the five-number summary for these data:

```
> quantile(mileage,type=2)
 0%  25%  50%  75% 100%
13   22   27   42  114
```

There are 31 cars listed in Table 11.2. This means there are 15 cars in the lower half of the data set and there are 7 cars in the lower quartile. These 7 are listed below:

Model	mpg
Bentley Mulsanne	13
Cadillac CTS	14
Rolls Royce Wraith	15
Dodge Challenger	18
Jaguar XF	18
Buick LaCrosse	21
Audi A8	22

Note: There are five cars in the data which have 22 mpg (Audi A8, Buick Regal, Cadillac CTS AWD, Infiniti Q40, and Volvo S80). Any of these could go in the table above (but only one).

(c) The stemplot shows the shape of the distribution, and it shows there are multiple outliers (the electric cars). This information is not readily available from the five-number summary.

12.17. Here is a stemplot of the hot dog calories data:

```
> calories = c(173,191,182,190,172,147,146,139,175,136,179,153,107,195,135,140,138)
> stem(calories,scale=2)
```

```

10 | 7
11 |
12 |
13 | 5689
14 | 067
15 | 3
16 |
17 | 2359
18 | 2
19 | 015
```

Here is the five-number summary for these data:

```
> quantile(calories,type=2)
 0%  25%  50%  75% 100%
107  139  153  179  195
```

The stemplot shows the shape of the distribution. There appears to be different clusters of hot dogs: the low-calorie hot dog at 107, those ranging from 135-153 calories, and those hot dogs whose calories exceed 172. This information cannot be seen in the five-number summary.

12.21. I first calculated the five-number summary and the mean and standard deviation of the data:

```
> immigrants = c(3.8,11.9,2.2,127.6,26.2,40.5,10.1,7.5,1.6,25.1,37.0,18.9,1.8,2.3,
  54.4,139.4,21.1,16.9,27.8,4.1,5.0,9.8,0.8,29.5,0.8,6.7)
>
> quantile(immigrants,type=2) # Five-number summary
 0%  25%  50%  75% 100%
 0.8   3.8  11.0  27.8 139.4
>
> mean(immigrants) # Mean
[1] 24.3
> sd(immigrants) # Standard deviation
[1] 35.1
```

I can already tell the distribution will be skewed to the right (high side) because the mean is so much larger than the median. On the next page, I prepared both a histogram and a boxplot of the immigrant data. States whose observations are larger than

$$Q_3 + 1.5(IQR) = 27.8 + 1.5(24.0) = 63.8$$

would be considered an outlier using our 1.5(IQR) rule of thumb. There are two outliers: Florida (127.6) and New York (139.4). There are no outliers on the low side.

R code:

```
bins = seq(0,140,10) # specify class intervals
# Histogram
hist(immigrants,breaks=bins,xlab="Number of legal immigrants (in thousands)",
     ylab="Count",main="",col="lightblue")
# Boxplot
boxplot(immigrants,range=1.5,xlab="",ylab="Number of legal immigrants (in thousands)",
       col="lightblue")
```

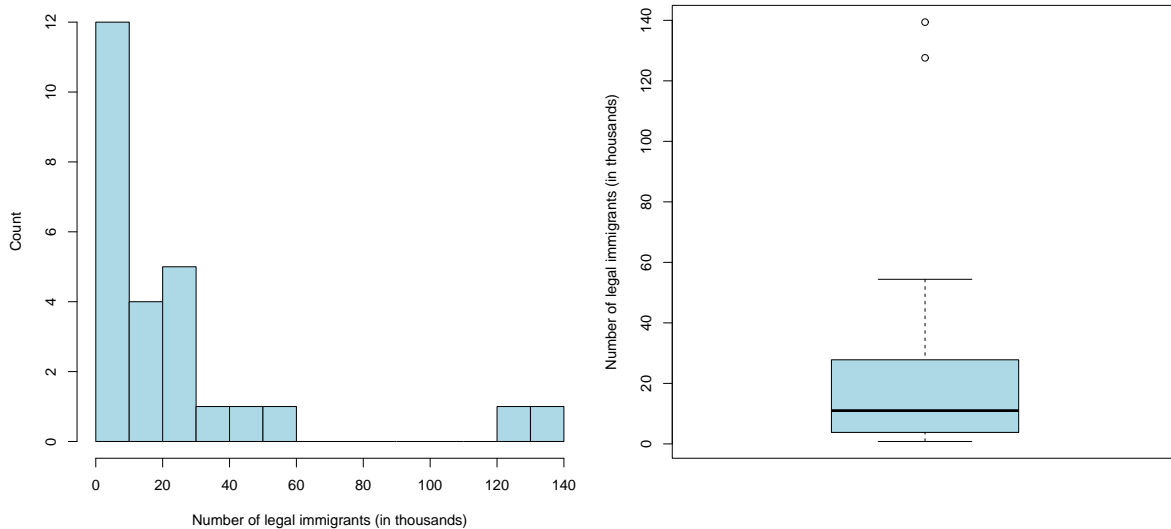


Figure 1: Immigrant data. Left: Histogram. Right: Boxplot. I used the 1.5(IQR) rule to identify outliers in the boxplot.

Interpretation: The distribution of the number of legal immigrants among these states is strongly skewed to the right. There are two outliers. Therefore,

- we should not use the mean and standard deviation as measures of “center” and “spread,” respectively. These will be heavily impacted by the outliers.
- we should use instead the median as a measure of center ($M = 11.0$) and the IQR as a measure of variability. In particular, 50% of the states have immigrant counts between 3.8 and 27.8. These measures are in thousands.

12.30. In this problem, the authors want you to calculate the mean \bar{x} and the standard deviation (s) by hand to get practice. We will do this and then use R to check our work.

Here are the levels of phosphate (measured in mg/dl) for the patient:

5.6 5.2 4.6 4.9 5.7 6.4

The sum of the observations is

$$\sum x = 5.6 + 5.2 + 4.6 + 4.9 + 5.7 + 6.4 = 32.4.$$

Therefore,

$$\bar{x} = \frac{\sum x}{6} = \frac{32.4}{6} = 5.4.$$

The mean phosphate reading is 5.4 mg/dl. Now, let's calculate the standard deviation:

Observation	Squared distance from the mean
5.6	$(5.6 - 5.4)^2 = (0.2)^2 = 0.04$
5.2	$(5.2 - 5.4)^2 = (-0.2)^2 = 0.04$
4.6	$(4.6 - 5.4)^2 = (-0.8)^2 = 0.64$
4.9	$(4.9 - 5.4)^2 = (-0.5)^2 = 0.25$
5.7	$(5.7 - 5.4)^2 = (0.3)^2 = 0.09$
6.4	$(6.4 - 5.4)^2 = (1.0)^2 = 1.00$
$\sum (x - \bar{x})^2 = 2.06$	

Therefore,

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{2.06}{5}} \approx \sqrt{0.412} \approx 0.64.$$

The standard deviation is 0.64 mg/dl. This is the six observations' average distance from the mean.

Implementation in R:

```
# Enter data
phosphate = c(5.6, 5.2, 4.6, 4.9, 5.7, 6.4)
mean(phosphate)
[1] 5.4
> sd(phosphate)
[1] 0.64
```

12.36. (a) If all teachers get a \$3,000 raise, then the mean and median will also increase by \$3,000. For example, suppose we have five teachers with salaries

42000 45000 48000 51000 54000

The median salary is \$48,000. The mean salary is \$48,000. If each salary is increased by \$3,000, then the salaries become

45000 48000 51000 54000 57000

The median salary is now \$51,000. The mean salary is now \$51,000. Each salary has increased by \$3,000.

(b) Calculate the IQR for the two data sets above. For the first data set,

$$\text{IQR} = Q_3 - Q_1 = 51000 - 45000 = 6000.$$

For the second data set,

$$\text{IQR} = Q_3 - Q_1 = 54000 - 48000 = 6000.$$

The IQR has not changed! If all the salaries increase by the same amount, then the quartiles will change, but the distance between them will not change.

(c) The same is true with the standard deviation. If all the salaries increase by the same amount, then the average distance from the mean will not change. Let's check this with the two data sets above (using R):

```
> salaries = c(42000,45000,48000,51000,54000)
> sd(salaries)
[1] 4743.416
> salaries = c(45000,48000,51000,54000,57000)
> sd(salaries)
[1] 4743.416
```

The standard deviation remains the same. Remember the standard deviation is a measure of variability (as the IQR is). If all the salaries increase by the same amount, the center will change, but the variability (or “spread”) will not.