

## Homework Answers:

### 25. Super Bowl wins.

Super Bowl Winning Margins	
4	5
4	
3	556
3	2
2	5779
2	123
1	567777899
1	000122344
0	567779
0	133333344444

Key : 4|5 = 45  
points

### 27. Wins again.

- The median winning margin in the first 48 Super Bowl games is 12.5 points.
- The first quartile of the winning margin in the first 43 Super Bowl games is 4.5 points. The third quartile is 21 points.

### 29. Describing Super Bowl wins.

The distribution of winning margins in the Super Bowl is skewed high. The lowest winning margin was 1 point and the highest winning margin was 45 point, which was an outlier. The second highest winning margin was only 36 points. The median winning margin was 12.5 points, with the middle 50% of winning margins between 4.5 and 21 points.

pg. 68

## Day 11 - Review

## Tell -- Draw a Picture

- When telling about quantitative variables, start by making a histogram, dotplot, or stem-and-leaf display and discuss the shape of the distribution.

## Tell -- Shape, Center, and Spread

- Next, always report the *shape* of its distribution, along with a *center* and a *spread*.
- \* { ■ If the shape is *skewed*, report the *median* and *IQR*.
- If the shape is *symmetric*, report the *mean* and *standard deviation* and possibly the median and IQR as well.

## Tell -- What About Unusual Features?

- If there are multiple modes, try to understand why. If you identify a reason for the separate modes, it may be good to split the data into two groups.
- If there are any clear outliers and you are reporting the mean and standard deviation, report them with the outliers present and with the outliers removed. The differences may be quite revealing.



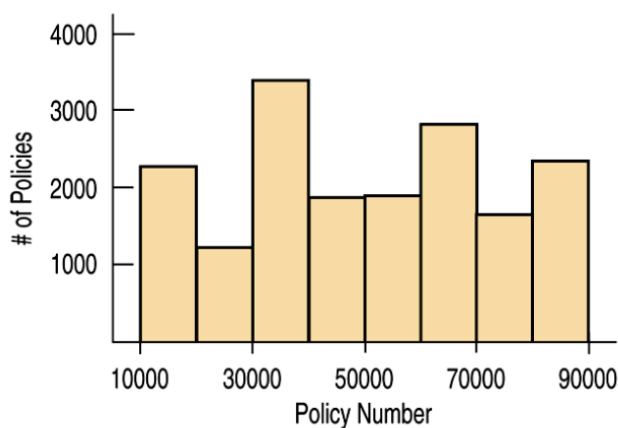
Note: The median and IQR are not likely to be affected by the outliers.

## Thinking About Variation

- Since Statistics is about variation, spread is an important fundamental concept of Statistics.
- Measures of spread help us talk about what we *don't* know.
- When the data values are tightly clustered around the center of the distribution, the IQR and standard deviation will be small.
- When the data values are scattered far from the center, the IQR and standard deviation will be large.

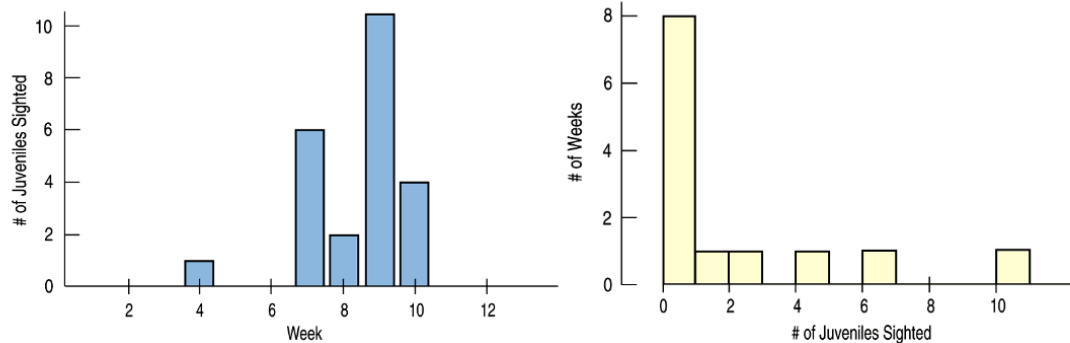
## What Can Go Wrong?

- Don't make a histogram of a categorical variable—bar charts or pie charts should be used for categorical data.
- Don't look for shape, center, and spread of a bar chart.



## What Can Go Wrong? (cont.)

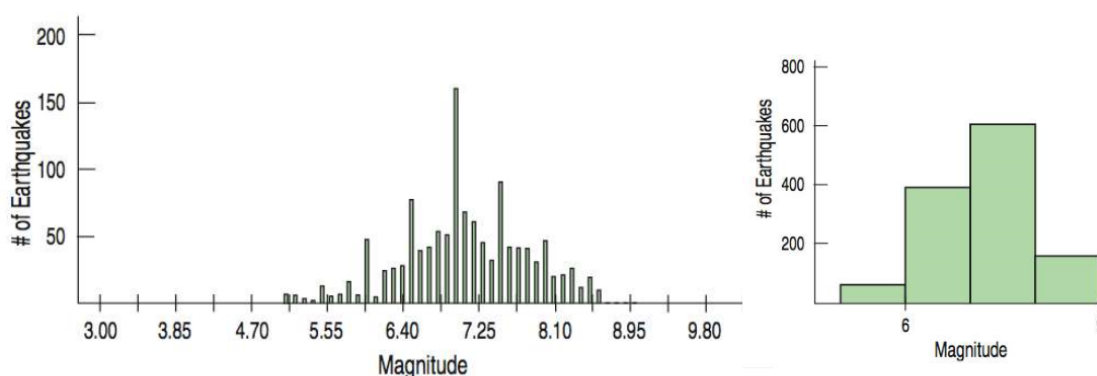
- Don't use bars in every display—save them for histograms and bar charts.
- Below is a badly drawn plot and the proper histogram for the number of juvenile bald eagles sighted in a collection of weeks:





## What Can Go Wrong? (cont.)


- Choose a bin width appropriate to the data.
  - Changing the bin width changes the appearance of the histogram:



## What Can Go Wrong? (cont.)

- Don't forget to do a reality check – don't let the calculator do the thinking for you.
- Don't forget to sort the values before finding the median or percentiles.
- Don't worry about small differences when using different methods.
- Don't compute numerical summaries of a categorical variable.
- Don't report too many decimal places.
- Don't round in the middle of a calculation.
- Watch out for multiple modes
- Beware of outliers
- Make a picture ... make a picture . . . make a picture !!!

## What have we learned?

- We've learned how to make a picture for quantitative data to help us see the story the data have to *Tell*.
  - We can display the distribution of quantitative data with a *histogram, stem-and-leaf display, dotplot, or boxplot*.
  - We've learned how to summarize distributions of quantitative variables numerically.
- 
- Measures of center for a distribution include the median and mean.
  - Measures of spread include the range, IQR, and standard deviation.
  - Use the median and IQR when the distribution is skewed. Use the mean and standard deviation if the distribution is symmetric.

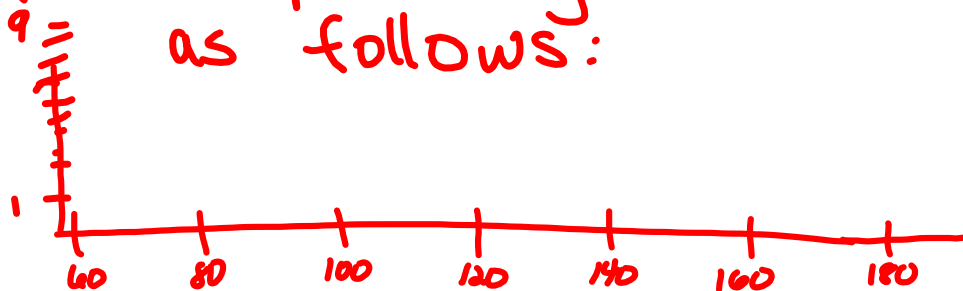
## What have we learned? (cont.)

- We've learned to *Think* about the type of variable we are summarizing.
  - All methods of this chapter assume the data are quantitative.
  - The **Quantitative Data Condition** serves as a check that the data are, in fact, quantitative.

Classwork <sup>Today</sup> ~~Tomorrow~~  
Chapter 3 Review B

## Classwork - Chapter 3 Review B (in packet)

#3 No calculations

#4 Set up Histogram X-axis  
as follows:

**Statistics: Chapter 3 Review B – KEY**

- Describe when it is more accurate to use the median and IQR to measure the center and spread of a set of data opposed to using the mean and standard deviation.

If a distribution is skewed or has outliers, use the median and IQR instead of mean and standard deviation.

- Here is a histogram of the horsepower of cars ([www.shodor.org/interactivate](http://www.shodor.org/interactivate)). Which measures of center and spread are appropriate to use for this data and why?

Since the distribution is skewed to the right, use the median and IQR instead of mean and standard deviation. The median and IQR are more resistant to the effects of skewness than the mean and standard deviation.

- If a new car with 300 horsepower was added to the histogram above, how would each of the following summary statistics change: mean, median, range, IQR, and standard deviation.

**Mean, Range, SD. will ↑**  
**Median, IQR will stay approx. the same**

- Here are the number of floors in the 20 tallest buildings of the world in 2014. Make a histogram of this data, find the five number summary, and write a complete shape, center, spread description.

163	116	101	98	88
121	108	101	89	88
120	104	101	88	80
118	103	100	88	66

Min = 66  
 Q1 = 88  
 Median = 101  
 Q3 = 112  
 Max = 163.

The distribution of the number of floors in the world's 20 tallest buildings is roughly unimodal and symmetric. The median number of floors is 88 and the IQR is 24 floors. The distribution contains an outlier (Burj Khalifa in Dubai, UAE) that has 163 floors. This is 42 floors more than the next highest number of floors.

