After completing this material, you should be able to:

- describe the distribution of a quantitative variable by discussing its shape, center, spread, and unusual characteristics.

- calculate (using StatCrunch) measures of center and measures of spread.

- apply the Empirical Rule or Chebyshev's Rule to a distribution when discussing the standard deviation.

- compare distributions using boxplots.

Recall: What is a quantitative variable?

To summarize a quantitative variable, we need a new graphical display – a bar graph cannot be used. We will first look at **histograms** for graphically summarizing quantitative data. What exactly is a histogram?

Example: *Money* magazine undertook a study in 2009 to estimate the average cost for a visit to a hospital emergency room.   A random sample of 175 emergency room visits in a certain urban area was taken, and the out-of-pocket costs associated with that visit were recorded. A histogram for the collected data is given below:

Histogram of Out-of-Pocket ER Costs for 175 Visits
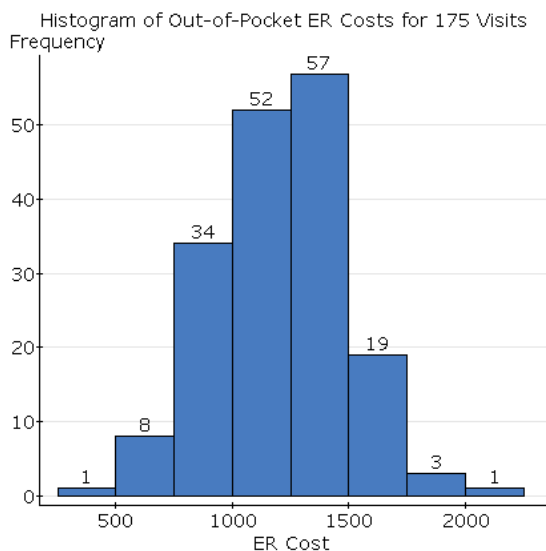
When summarizing or describing a distribution, the following four characteristics must be discussed:

1.

2.

3.

4.

> **When asked to describe a distribution, make sure you address these four characteristics *in context* and *in complete sentences.***

Let's consider each of these four characteristics individually.

Shape of the distribution

When describing a distribution, the first think we should consider is what shape the distribution has. We will consider **five** common shapes (shown below):

| Shape	Histogram	Description

Measures of Center

Once we know the shape of a distribution, it is common to summarize it by finding a "typical" value of the distribution – these values are generally referred to as measures of center. There are two common measures of center which are used:

Measure of Center	Notation	Description

The calculation of these measures, while not difficult, can be tedious. Instead of calculating these summary statistics by hand, we will rely on the use of StatCrunch.

How does the shape of the distribution affect measures of center?
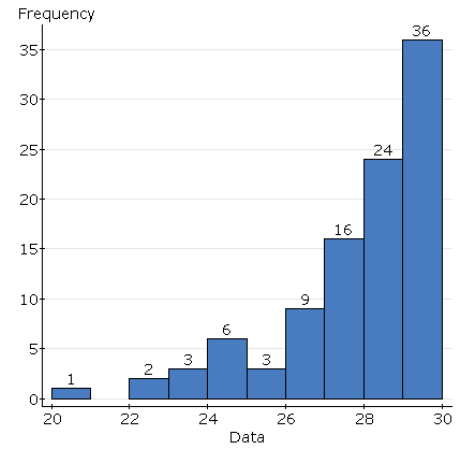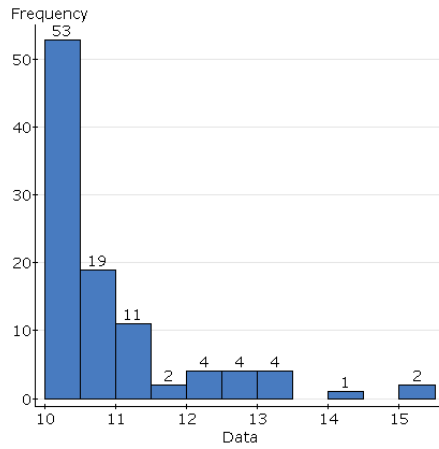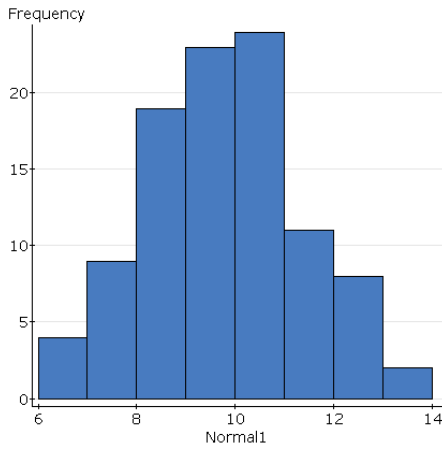
Measures of Spread

Unfortunately, a measure of center doesn't adequately describe a distribution. We also must have some idea how the values in the distribution vary. This requires a measure of spread. There are three common measures of spread which are used:

Measure of Spread	Notation	Description

The calculation of these measures can be quite difficult – the formula for standard deviation is quite tedious. Instead of calculating these summary statistics by hand, we will rely on the use of StatCrunch.

Unusual Observations

Unusual observations are often referred to as *outliers*. When determining if unusual observations are present in the data, look for observations which do not follow the overall patter of the data. These will generally be observations which are in the tail of the distribution – either very large or very small.

Example: *Money* magazine undertook a study in 2009 to estimate the average cost for a visit to a hospital emergency room. A random sample of 175 emergency room visits in a certain urban area was taken, and the out-of-pocket costs associated with that visit were recorded. A histogram for the sample is given below, as well as summary statistics:
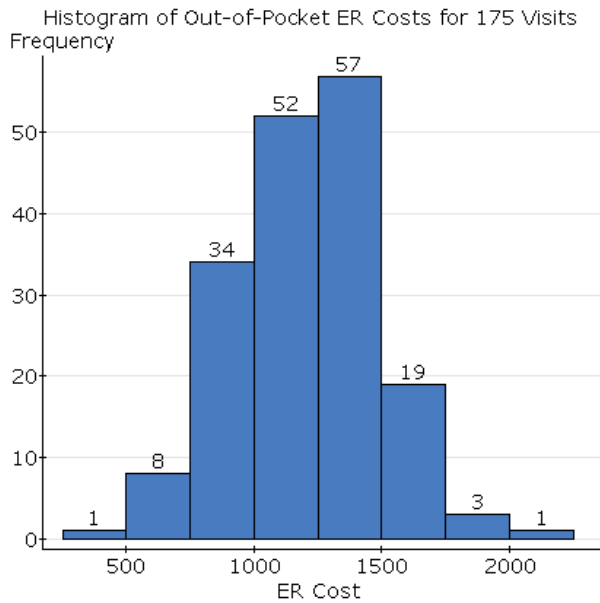
Describe the distribution of out-of-pocket costs.

Summary statistics:

Column	n	Mean	Variance	Std. dev.	Std. err.	Median	Range	Min	Max	Q1	Q3
ER Cost	175	1203.9886	78595.161	280.34828	21.192338	1214	1686	471	2157	1000	1414

How can the standard deviation be interpreted?

Empirical Rule

Chebyshev's Rule

Example: Pelican Stores, a division of National Clothing, is a chain of women's apparel stores operating throughout the country. The chain recently sampled 100 in-store credit transactions in order to gain insight about the spending habits of their customers. A histogram of these transactions is given below, as well as various summary statistics:

Histogram of 100 Credit Card Transactions
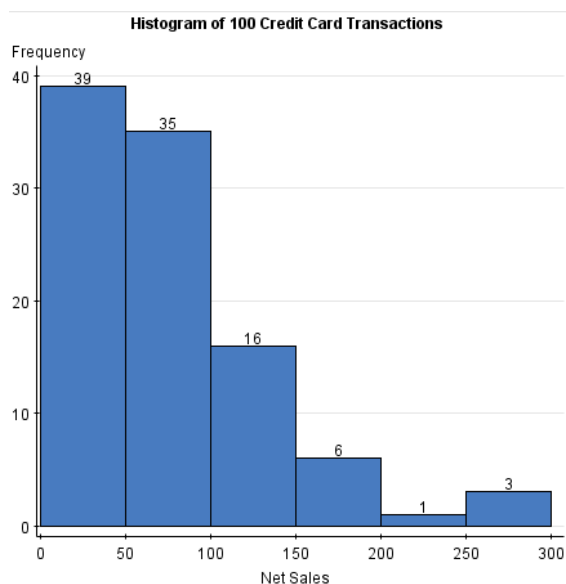
Summary statistics:

Column	n	Mean	Variance	Std. Dev.
Net Sales	100	77.6005	3098.5854	55.66494

Median	Range	Min	Max	Q1	Q3
59.705	274.36	13.23	287.59	39.6	101.4

The distribution of _____ has a shape that is _____. The average purchase in this sample of _____ transactions is _____, and 50% of the transactions have an amount of _____or less. The minimum purchase was _____, while the maximum was _____. The middle _____% of purchases in this sample were between $39.60 and _____. According to _____ Rule, we expect _____% of purchases at this store to fall between _____ and _____ (two standard deviations from the mean).

Comparing Distributions

Textbook pages: 88 – 95

Generally, we can answer much more interesting questions when we compare two or more distributions. It can be cumbersome to compare several different histograms, so a different graphical display called a **boxplot** is often used. A boxplot is based on the 5-number summary which consists of the following five staitsitcs:

These statistics are then used to construct a boxplot. A generic boxplot is shown below:

Example: A large number of fast-food restaurants with drive-through windows offer drivers and their passengers the advantages of quick service. To measure the quality of service, an organization called QSR planned a study in which the amount of time taken by a sample of drive-through customers at each of five restaurants (Popeye's, Wendy's, McDonald's, Hardee's, and Jack in the Box) was recorde

- Which fast food chains had the most similar median drive through times?

- Which fast food chain has a distribution of drive thru times that is likely right skewed?

- Which fast food chain had the smallest innerquartile range of drive through times?

- Which fast food chain had the most variability according to their ranges?

- Which fast food chain was responsible for the overall fastest service in the sample? Which was responsible for the overall slowest service?