

Math 243 Sections 2.1.2-2.2.5 Descriptive Statistics for Numerical Variables

A framework to describe quantitative data:

Describe the Shape, Center and Spread, and Unusual Features

Shape – How is the data distributed? We need to see a picture to determine the shape.

There are four types of graphs for quantitative variables

- Dot Plot
- Stem-and-leaf Plot
- Histogram
- Boxplot

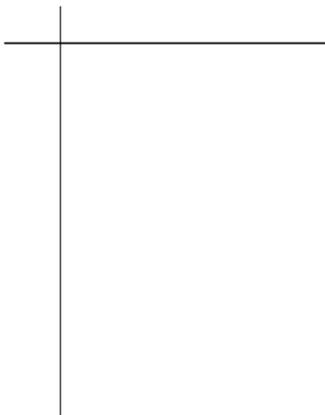
Example 1: Here is a set of 15 exam scores for a fictional MTH 243 Statistics class at PCC.

31, 62, 65, 70, 76, 81, 82, 82, 87, 88, 89, 94, 95, 98, 100

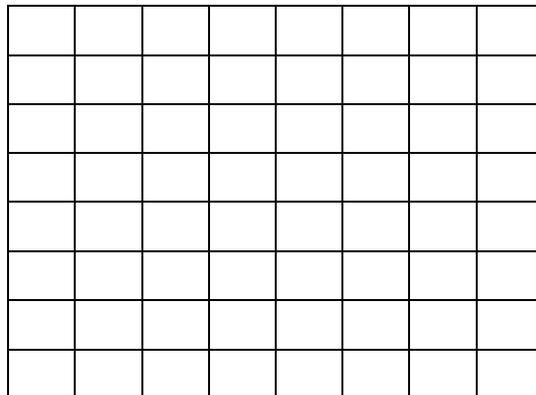
a. Draw a **dot plot** for this data.



b. Draw a **stem-and-leaf plot** using the tens digits as the stem and the ones digits as the leaves.



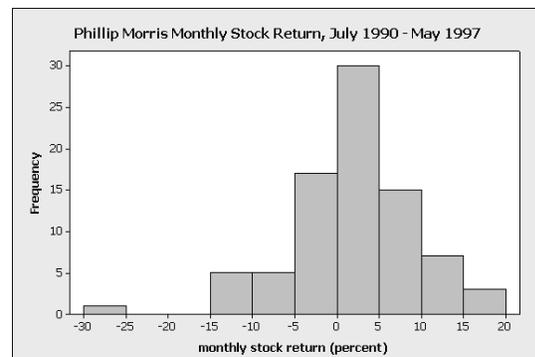
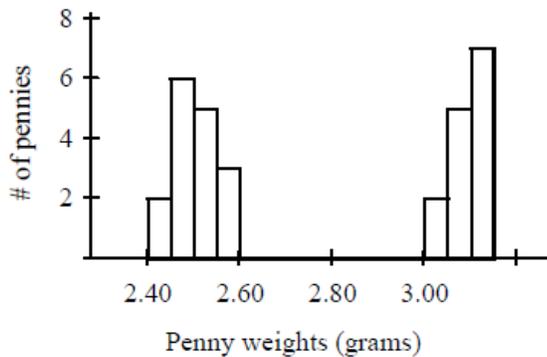
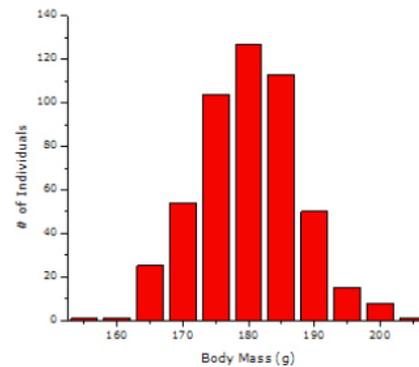
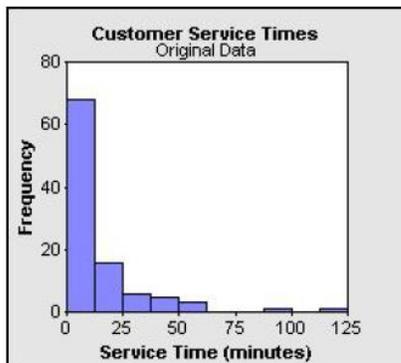
c. Sketch the corresponding **histogram** for this data using a bin width of 10. Scale and label your graph appropriately.



Describing Shape

- Symmetric, skewed left or skewed right
- Modes: no mode, unimodal, bimodal, multimodal

Example 2: In the examples below we'll explore the shape of various histograms. Is it symmetric? What is the shape of the histogram? Where are the modes (if it has any)?



Center and Spread – These should always be reported together.

Center is the position or location of the data (The average or typical value).

Spread is how much variation is in the data or how spread out the data is.

There are two different sets of measures for center and spread:

For skewed distributions: Center: Median Spread: IQR (Interquartile Range)	For symmetrical distributions: Center: Mean, \bar{x} Spread: Standard Deviation, s
-----------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------

Median and IQR

Find the median and quartiles of the data. This divides the data into four equal sections.

Example 1 continued: Using the previous data set (repeated here), find the median and quartiles by hand. The data is already in order. If there is an odd number of values, do not include the median when finding the quartiles. If there is an even number of values, average the middle two.

31, 62, 65, 70, 76, 81, 82, 82, 87, 88, 89, 94, 95, 98, 100

The median (Q_2) is the middle value or ____th percentile. ____% of the data are below that value.

The first quartile (Q_1) is the ____th percentile. ____% of the data are below that value.

The third quartile (Q_3) is the ____th percentile. ____% of the data are below that value.

5-Number Summary:

Min
Q_1
Median (Q_2)
Q_3
Max

Interquartile Range or IQR (Spread):

The width of the middle 50% of the data

$$IQR = Q_3 - Q_1$$

The IQR is a better measure of spread than the range (max - min) because it is less sensitive to outliers.

How to draw a Boxplot. Some books call this a modified boxplot because outliers are shown.

1. Draw the Box. Determine the scale and draw vertical lines at the Median, Q_1 and Q_3 . Connect these to form the box. Label your horizontal axis and scale (there is no vertical scale).

2. Calculate the Fences using $1.5 \cdot IQR$.

We use 1.5 times the Interquartile range on each side of the box to determine the fences. Any data values outside of the fences are considered outliers. The whiskers are drawn to the nearest data values inside each fence.

$$\text{Upper Fence: } Q_3 + 1.5 \cdot IQR$$

$$\text{Lower Fence: } Q_1 - 1.5 \cdot IQR$$

Draw in the fences lightly or below the graph. They are not technically part of the graph.

3. Draw the whiskers to the nearest data value inside each fence and make a short vertical bar. Label each value outside the fences (outliers) with a dot.

Activity 1. Use **GeoGebra** to calculate the 1-variable statistics.

View > Spreadsheet > Type in your data (or copy and paste)

Select your data > Click on the histogram for **One Variable Analysis** > Click **Analyze**

Use **GeoGebra** to draw a histogram and boxplot.

Use the dropdown menu to select **Histogram**. Change the **options** using the icon in the upper right corner

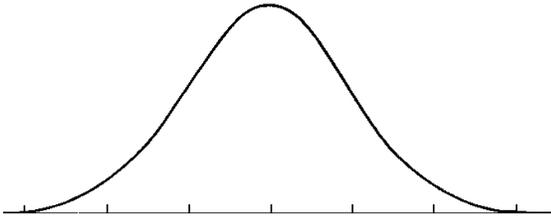
Select **Show 2nd Plot** and select **Boxplot**. Use the same scale for both plots so they match.

GeoGebra Results. Select Copy to Clipboard to insert a graph into a document.



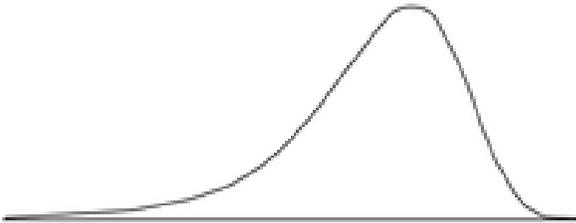
How do the mean, median and mode relate to the shape of the distribution? (2.6)

Symmetric Distributions

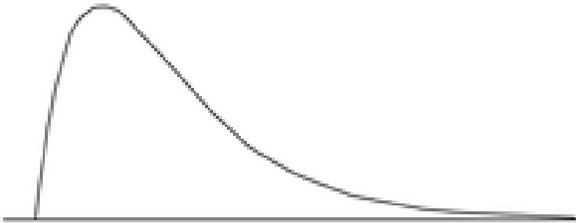


Skewed Distributions

Skewed left



Skewed right

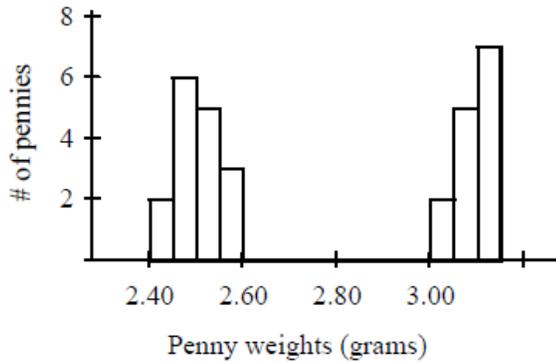


Unusual Features:

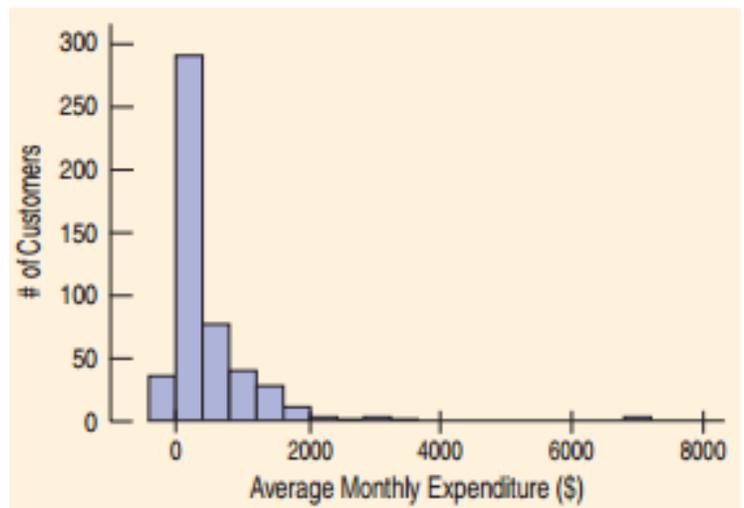
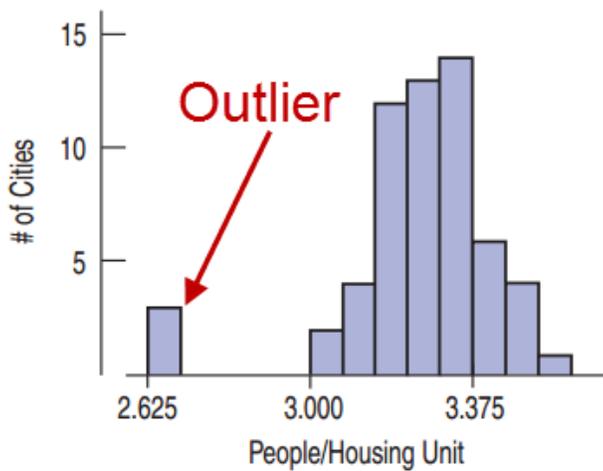
Mention anything unusual about the data or state that there aren't any unusual features

- Multiple modes (look for subpopulations)
- Gaps and Outliers

Multiple Modes:



Gaps and Outliers:

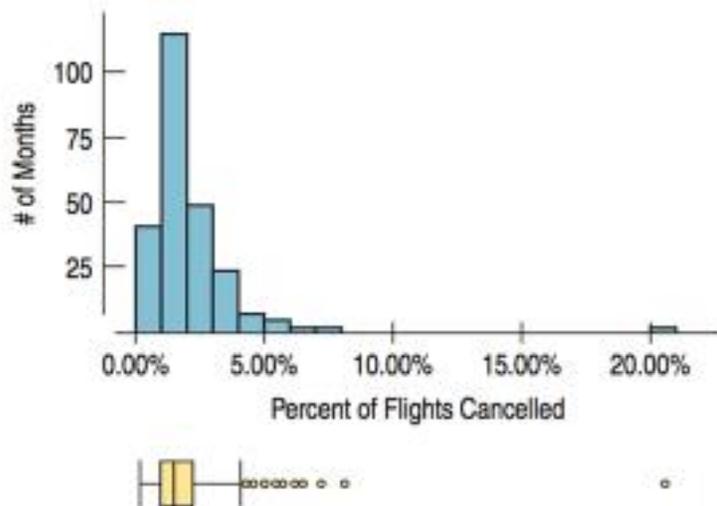


Should you toss out outliers?

Example 1 Continued: Continuing with our test score data set, write a paragraph describing the distribution. Be sure to talk about the shape, center and spread, and any unusual features (or say that there are none).

Practice: The US Bureau of Transportation Statistics reports data on airline flights. Let's look at data giving the percentage of flights cancelled each month between 1994 and 2013. Use a few sentences to describe how often flights are cancelled. Be sure to talk about the shape, center and spread, and any unusual features (or say that there are none).

Count	238
Max	20.24
Q3	2.39
Median	1.68
Q1	1.16
Min	0.38
IQR	1.23



Summary of Test Scores for a Math 243 Class

Summary Statistics

Mean: 80 points

Standard Deviation: 17.76 points

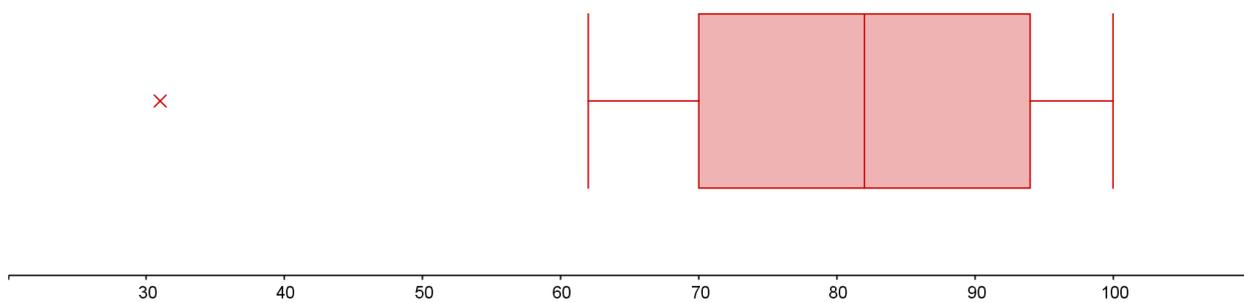
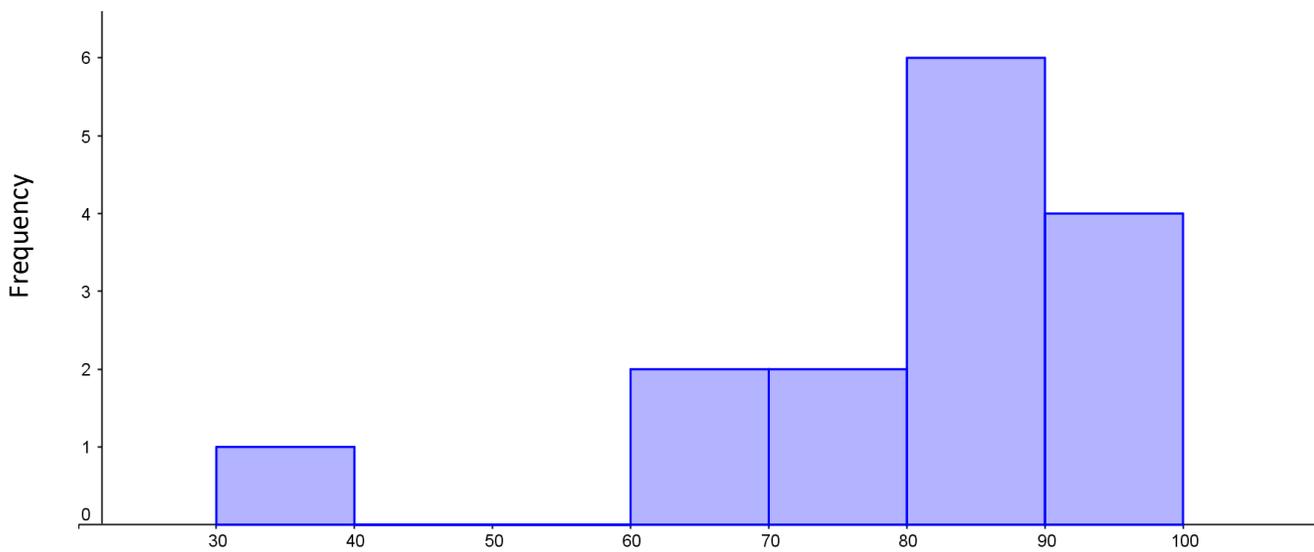
Min: 31 points

Q1: 70 points

Median: 82 points

Q3: 94 points

Max: 100 points



Test Scores for Math 243 Students (Points)

The distribution of test scores is unimodal and skewed to the left. The mode is between 80 and 90 points. The median is a score of 82 points and the width of the middle 50% is 24 points (interquartile range). There is one outlier score of 31 points to look into. The student may have gotten behind or had an extremely bad day.