AP Statistics Chapter 4 Notes and Practice
Quantitative Data

**Quantitative Data** – data that can be measured

**Distribution** provides the possible values of the variable and the frequency or relative frequency of each value.

Take out all the pennies that you have in your pocket or purse. Record the ages of the pennies (2007 – date on penny) for your pennies.

We are going to record the age of 50 pennies from the class. It would be nice to have some organization for this information. There are many ways that we can organize this.

Start with a list:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

Before making a graph of quantitative data, we need to state the condition (or assumption) that allows us to use this graph. Just like categorical data.

- I have measured data for the _____
- State the units_____.

A **dotplot** is a simple display that places dots along an axis. Label your axis from 0 until the oldest penny. Place a dot for each penny's age.

**Stem-and Leaf plots** are similar to dotplots but they retain the original data values (this is the cool part about stem-and-leaf plots). We usually repesent the last digit and the stem represents the other digits. In this case, the leaf will be the ones digit of age and the stem will be the tens digit. Note: You must provide a key for the stem-and-leaf plot!!!

| | |
|---|---|
| 0 | |
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |

Key 1|6 = 16 year old penny

| | |
|---|---|
| 0 | |
| 0 | |
| 1 | |
| 1 | |
| 2 | |
| 2 | |
| 3 | |

Note: If the rows are too long, we can split the information over two rows for each stem.

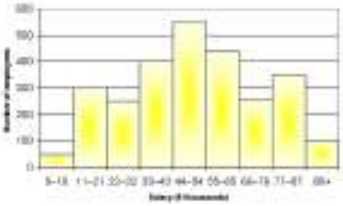
**Histograms** organize the data into "bins". A histogram looks like a bar graph but the bars are touching since the data is continuous. Also, we can decide on the width of the bars. The height of a histogram is often the "count" for each bin or the percentage of items in that bin. Let's make a histogram for the age of our pennies.

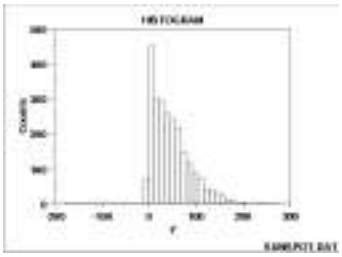
Notice that a histogram looks like a stem-and-leaf plot that has been outlined with rectangles.

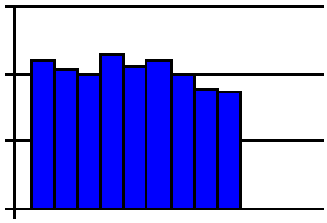
The following vocabulary words can be used to describe graphical displays.

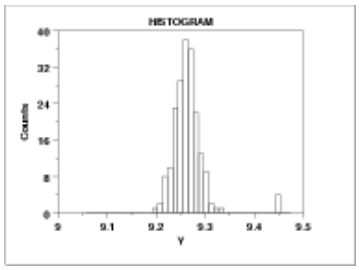
| | | | |
|---|---|---|---|
| **Uniform**<br>Each bin has approximately the same height | **Gaps**<br>Spaces between data points | **Uni-modal**<br>One bin has a highest value | **Bi-modal**<br>Two bins are ties for the highest vale |
| **Multi-modal**<br>There are more than two ties for the highest bin | **Mound-shapes**<br>The histogram looks like a hill with the highest peak near the middle | **Outliers**<br>Extreme values that don't appear to belong with the rest of the data | **Symmetric**<br>The two halves look like approximately like mirror images. |
| **Long Tails**<br>The edges drop off slowly. | **Short tails**<br>The edges drop off quickly | **Skewed Left**<br>The longer tail reaches to the left | **Skewed Right**<br>The longer tail reaches to the right. |

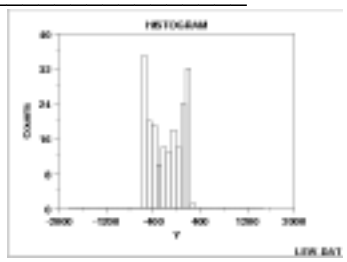
Using as many vocabulary words from the preceding page to describe these histograms.








Can you think of some possible real-life situations that would produce a histogram with these characteristics?

| | | | |
|---|---|---|---|
| **Uniform** | **Gaps** | **Uni-modal** | **Bi-modal** |
| **Multi-modal** | **Mound-shapes** | **Outliers** | **Symmetric** |
| **Long Tails** | **Short tails** | **Skewed Left** | **Skewed Right** |

A Timeplot displays data that change over time. Time is on the x-axis and the data is on the y-axis. Estimate a graph of gas rates over the past 2 years.

To describe a graph, we use three categories: Shape, Center, Spread.

- Center is tries to summarize all the data with one "typical value
- Spread tells how tightly clustered the data is
- Shape describes the symmetry, modes, and skewness

We will do more of this next chapter.

**Comparing distributions Using Stem-and-Leaf Plots**

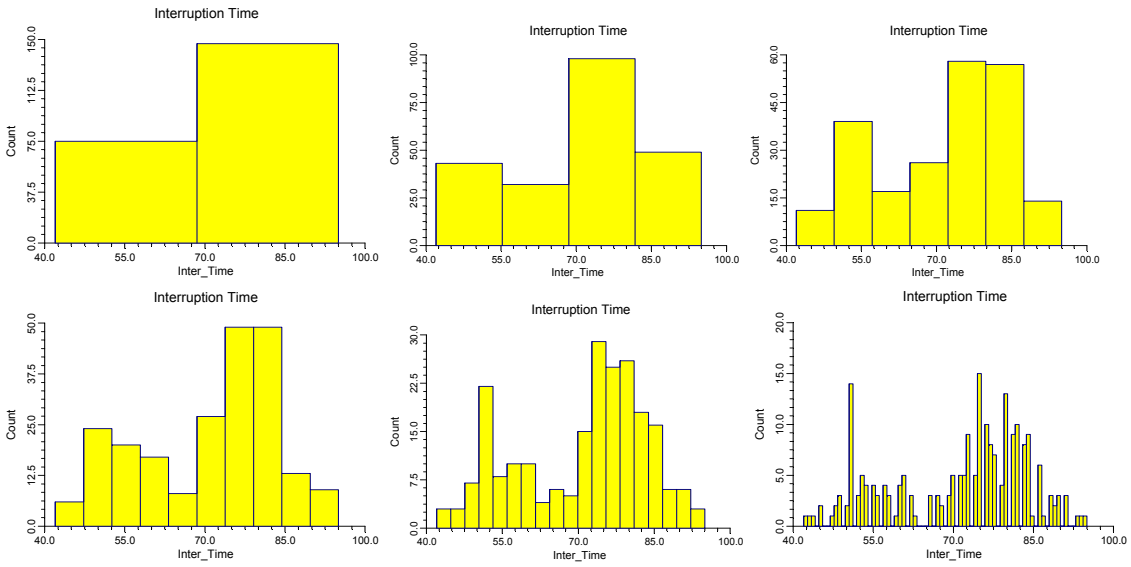
Key: 3|0 means 30 points scored in a game

| Scores | | |
|---|---|---|
| Leaf | Stem | Leaf |
| Tigers | | Sharks |
| 9 7 3 0 | 3 | 2 2 |
| 8 2 | 4 | 3 5 5 |
| 9 7 3 1 | 5 | 4 6 8 8 9 |

1. Which team had the highest score in a game?
2. Which team had the lowest score in a game?
3. Make a decision on which team is the best scoring team. Why?

## Histograms of Interruption Time of the Old Faithful Geyser Dataset

Because data is grouped into 'bins' of equal width, a histogram can give a false impression of the shape of the distribution of a dataset. A classic dataset to illustrate this is the 'Old Faithful' dataset, which contains data on both the duration and the interruption time (i.e. time betweens eruptions) of this famous geyser in Yellowstone National Park in Wyoming, USA.




Which histogram probably gives the most information about the time between eruptions? Why?

Extra Practice: Read pages 54 and 55. Make sure you can make a histogram on the calculator. If you run out of time, try to finish these on your own and check your answers.