How to Compare Distributions

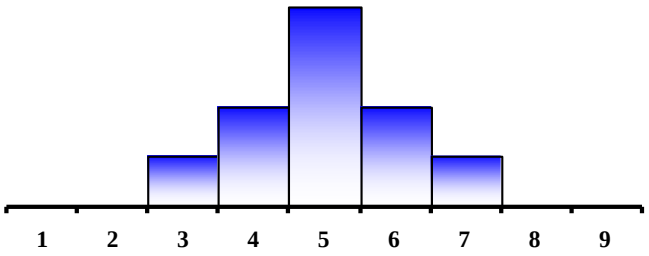
Name: _____ Class Period:_____

When you compare two or more data sets, focus on four features:

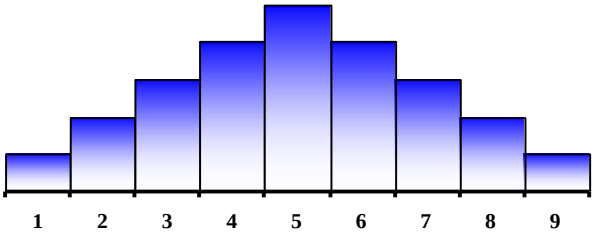
- **Center.** Graphically, the center of a distribution is the point where about half of the observations are on either side.
- **Spread**. The spread of a distribution refers to the variability of the data. If the observations cover a wide range, the spread is larger. If the observations are clustered around a single value, the spread is smaller.
- **Shape**. The shape of a distribution is described by symmetry, skewness, number of peaks, etc.
- **Unusual features**. Unusual features refer to gaps (areas of the distribution where there are no observations) and outliers.

SPREAD

The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set.



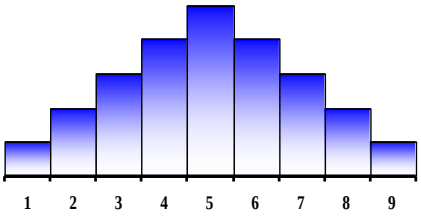
Less Spread



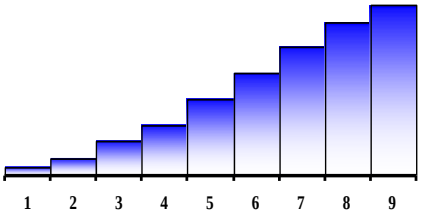
More Spread

SHAPE

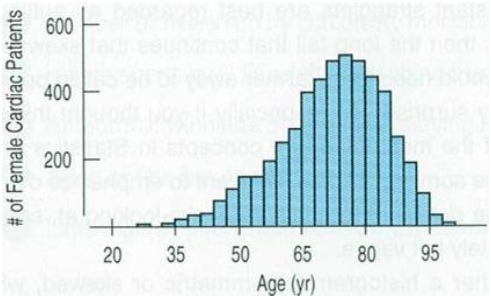
The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity

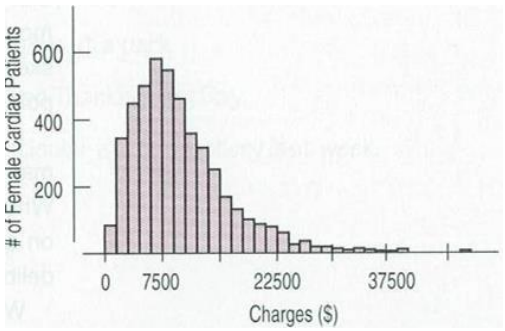


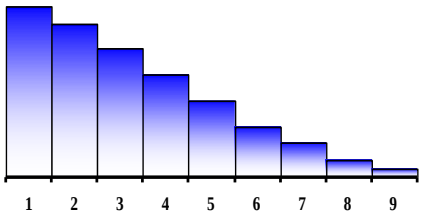
Symmetric, Unimodal, Bell-shaped



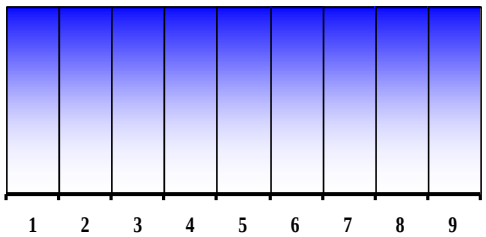
Skewed Left





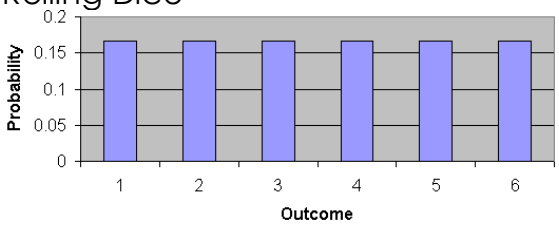


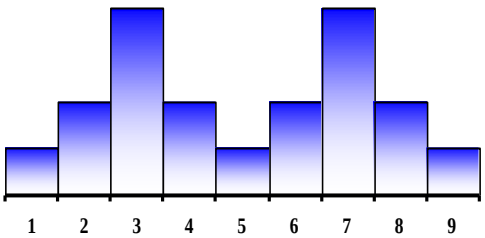
Skewed Right



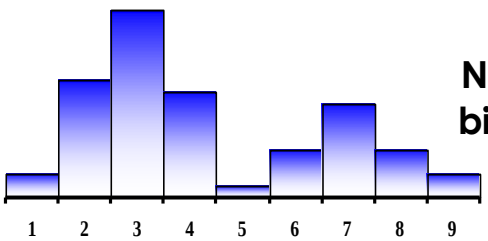
Uniform

Rolling Dice

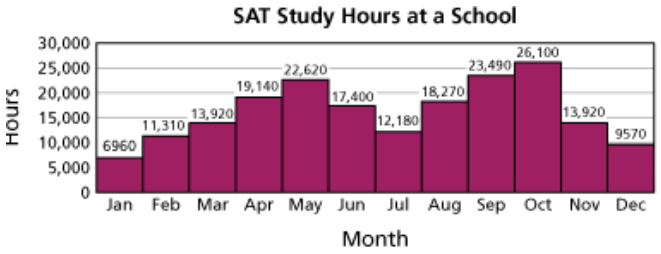




Symmetric, Bimodal

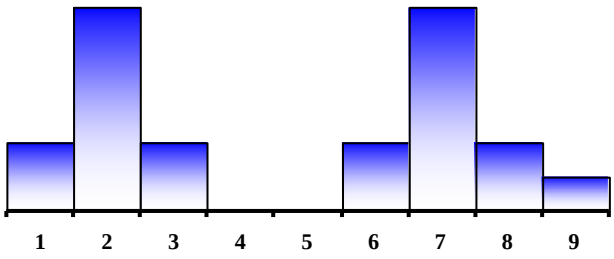


Non-Symmetric, bimodal



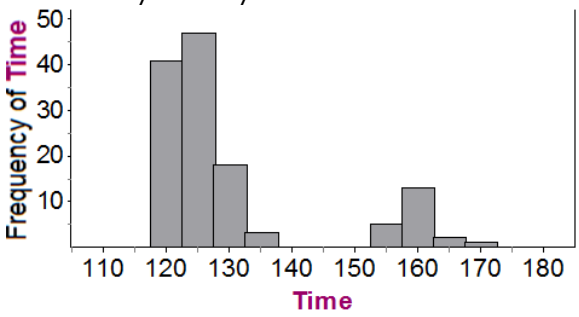
UNUSUAL FEATURES

Sometimes, statisticians refer to unusual features in a set of data. The two most common unusual features are gaps and outliers.

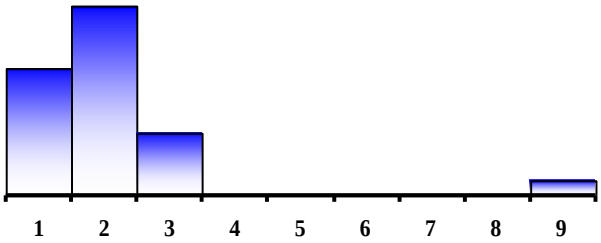


Gap

Kentucky Derby Times

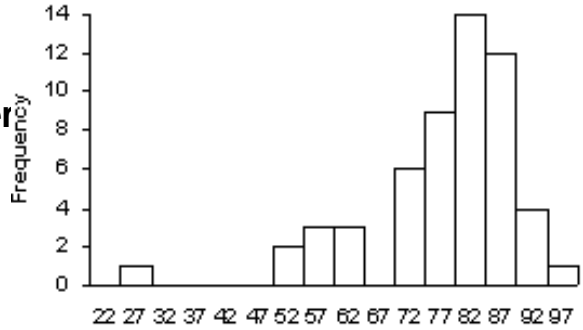


What could have caused this shift in times?



Outlier

Exam Scores



<u>Practice Problems</u>:

What shape would the following situations have?

1) A really hard test
2) A really easy test
3) Results of rolling a 6 sided die 1000 times
4) Heights for each gender at LHS?
5) Combined heights of male and female students at LHS?