Turning Data into Information Summary

		Categorical/Ordinal	*Quantitative*
Single Variable	Question(s)	How many/category?	Location, Spread, Shape.
	Summary	Counts, percentages	5# summary, x-bar, S, IQR
	Graphics	Pie chart, Bar chart	Box, histogram, stem & leaf
Two Variables	Question(s)	How many in sub-groups	Do they vary together?
	Summary	Counts, percents (row,col)	5#, x-bar, S, IQR
	Graphics	Grouped bar chart	Scatterplot
Combined relationships	Question(s)	What is the distribution within each category?	
	Summary	Single variable summaries for each category.	
	Graphics	Box plot of quantitative by categorical variable.	

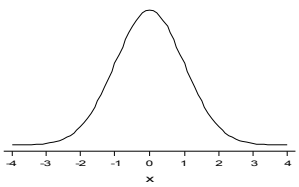
Definitions

Raw data: Data not yet processed or analyzed.
Population: The entire collection of individuals of interest.
Sample: A subset of a population.
Categorical Variable: Values are labels for categories within which individual units fall.
Ordinal Variable: A categorical variable which has a logical order to the categories.
Quantitative Variable: A measurement variable, sometimes numerical or continuous.
Response Variable: Usually the primary variable we're interested in (often on vertical (Y) or dependent axis).
Explanatory Variable: A variable which helps to partially explain the variation of the response (often on the horizontal (X) or independent axis).
Symetric: Quantitative data for which the probability or a histogram is symetric about the center.
Right/Left Skew: Quantiative data wich tends to trail off to either the right or left.
Bell Shaped: Quantitative data which tends toward a central value.
Outlier: A data value 'outside of the norm'.

	Median	
1st Quartile		3rd Quartile
Min		Max

5# Summary

Empirical Rule



For any bell-shaped curve, approximately

68% of the values fall within **1 standard deviation** of the mean in either direction

95% of the values fall within **2 standard deviations** of the mean in either direction

99.7% of the values fall within **3 standard deviations** of the mean in either direction