In your thought journal for "Reading: Exploratory Data Analysis" please define:

\* Exploratory Data Analysis

approach to analyzing data sets to summarize their main characteristics, often with visual methods

\* Histogram

approximate representation of the distribution of numerical data

\* Kernel Density

non-parametric way to estimate the probability density function of a random variable

Although a histogram can be very useful for examining the distribution of avariable, the graph can differ dramatically depending on the number of binsused. This problem can be overcome (partially) using nonparametric densityestimation. Density estimation is an attempt to estimate the probability den-sity function of a variable based on the sample, but less formally it can bethought of as a way of averaging and smoothing the histogram. Kernel density estimation is essentially a sophisticated form of locally weightedaveraging of the distribution. It uses a weight function (i.e., kernel) that en-sures the enclosed area of the curve equals one.

\* Winsorized/Trimmed Mean

One strategy for dealingwith this problem is to give less weight to values in the tails of the distribution,and pay more attention to the values near the center. A specific strategy forimplementing this idea is to winsorize4the distribution. winsorizing the distribution changes the highest x% of the scoresto the next smallest score, and changes the x% smallest scores to the nextlargest score. Another strategy for reducing the effects of the tails of adistribution is simply to remove them. This is the strategy employed by trim-ming. To find a trimmed mean, the x% largest and smallest scores are deletedand the mean is computed using the remaining scores.

\* Boxplots

method for graphically depicting groups of numerical data through their quartiles

\* Mean/Error bars and how they are different from Confidence Intervals

a graph of the sample means is provided as a vi-sual depiction of the differences between multiple groups. The use of error bars helps us facilitate a more accurateinterpretation of what is actually happening in the population.