

# Descriptive Statistics Supplement: Boxplots and Skew

Ben Babcock  
University of Minnesota

# Percentiles and Quartiles

Percentile: A score under which a certain percentage of scores fall. H, 117

Quartiles: scores at which 25%, 50%, or 75% of the rest of the scores are below.

1<sup>st</sup> quartile: 25% below

2<sup>nd</sup> quartile: 50% below

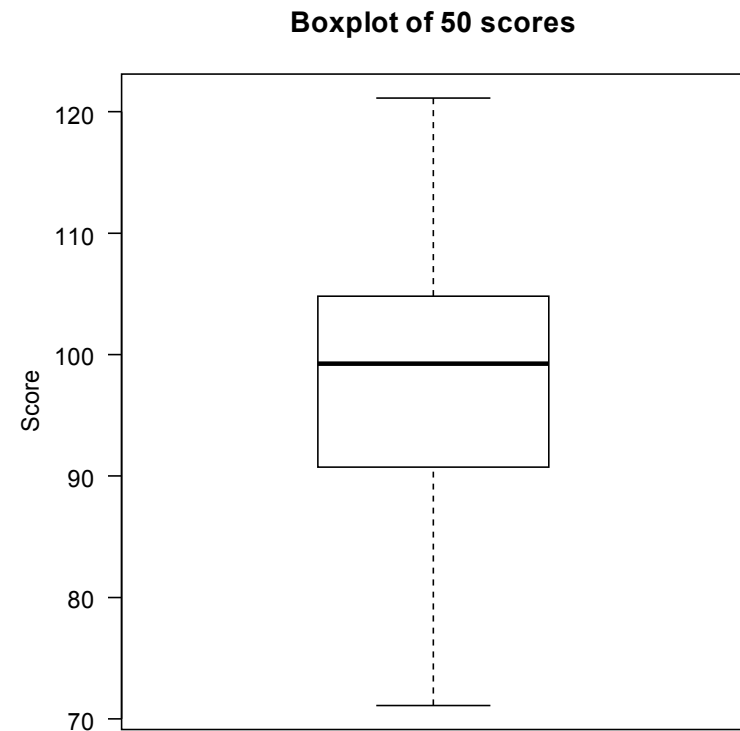
3<sup>rd</sup> quartile: 75% below

# Boxplots

Boxplots (a.k.a. box and whisker plots) are plots which represent several descriptive measures graphically.

The center line is the median. The lower line of the box is the 1<sup>st</sup> quartile. The upper line of the box is the 3<sup>rd</sup> quartile. The box contains 50% of the data points.

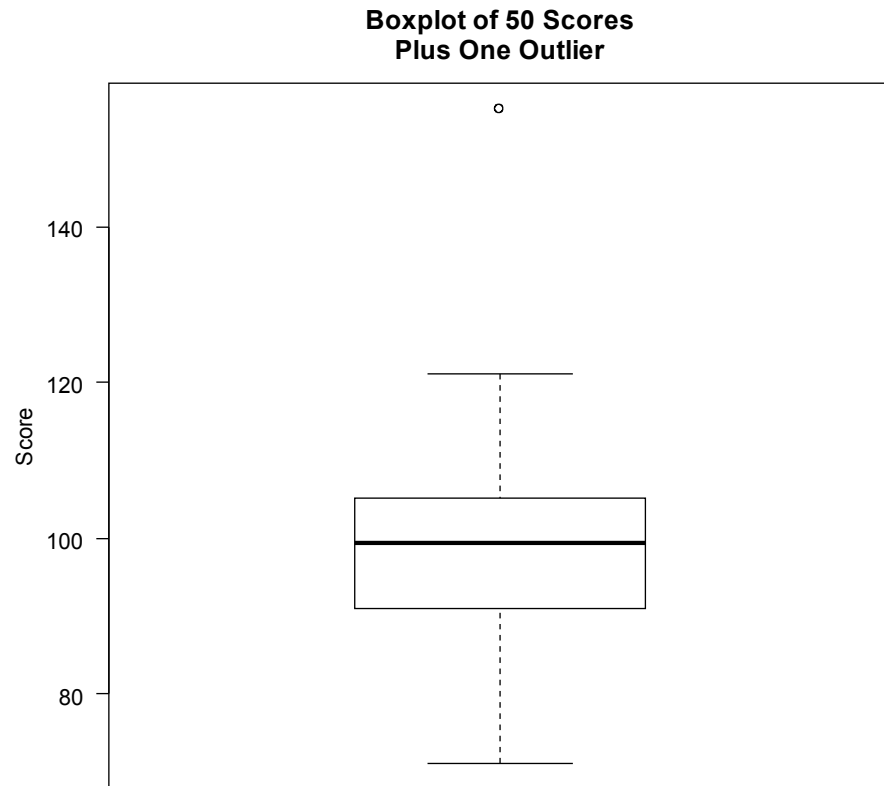
The whiskers represent roughly the upper and lower 25% of scores (the remaining 50% of the data). Howell, p 87-91.



# Boxplots and Outliers

One nifty feature about boxplots is that they help identify outliers.

**Outlier:** A point that is well above or below the rest of the sample.



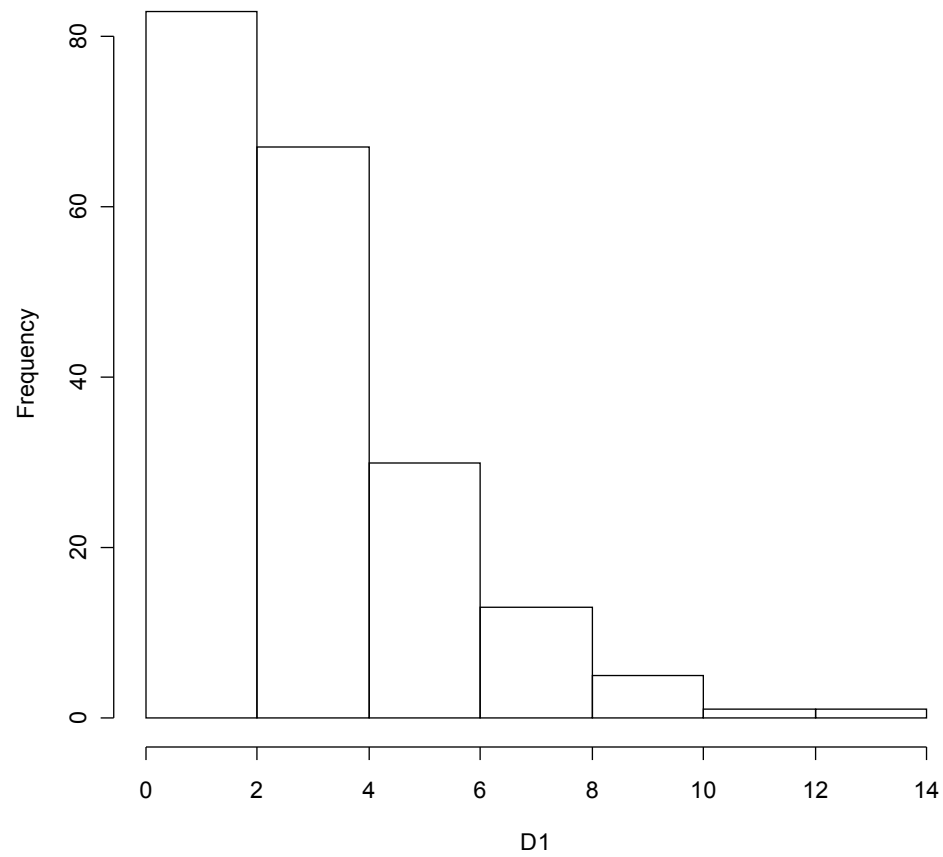
# Skew

When talking about distributions, skew can be an important topic. Skew is the extent to which a distribution is asymmetrical (one of the tails of the distribution is longer).

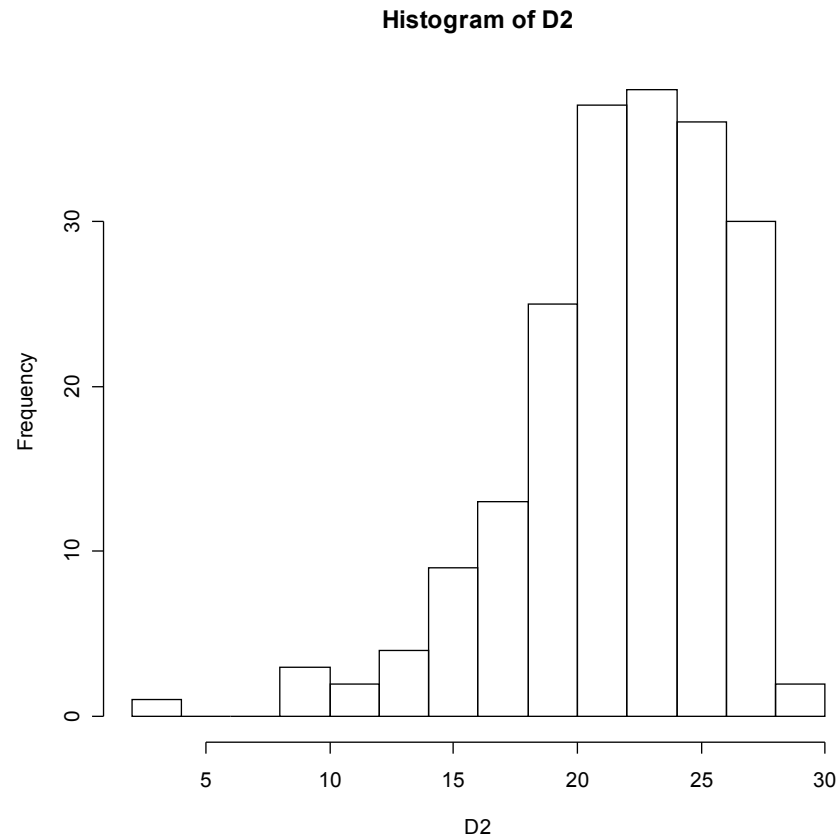
Skew is itself a descriptive statistic. However, the calculation of skew is not extremely important. In this class, we will use skew as a word to generally describe distributions.

The direction of skew, positive or negative, always goes with the direction of the long tail.

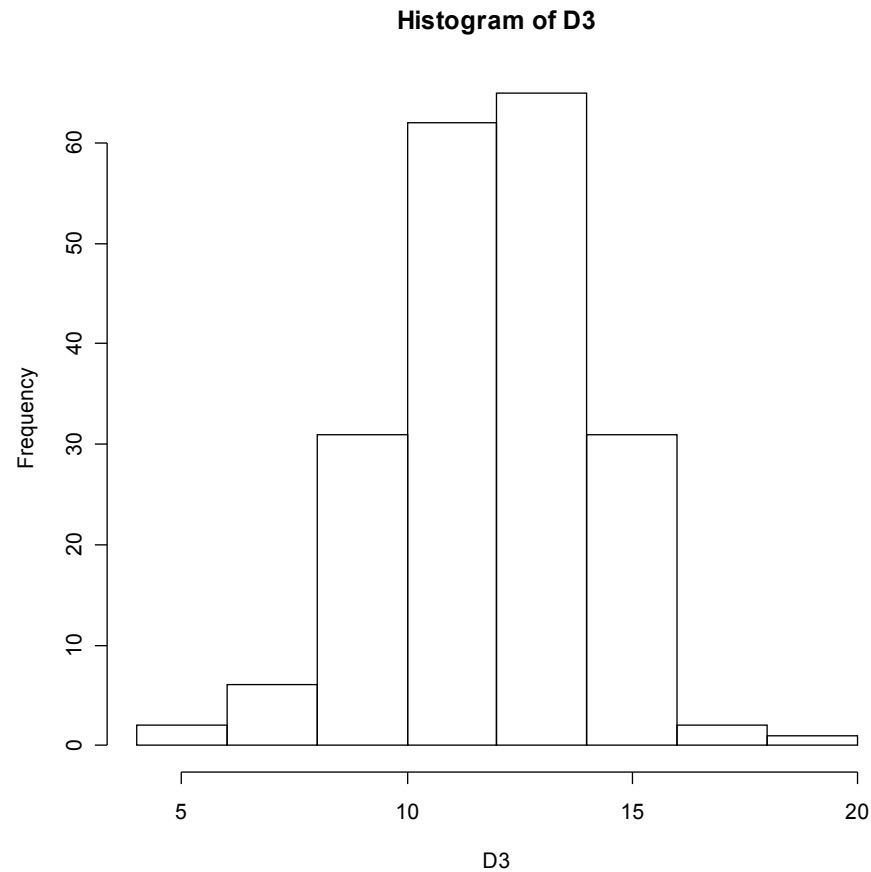
Histogram of D1



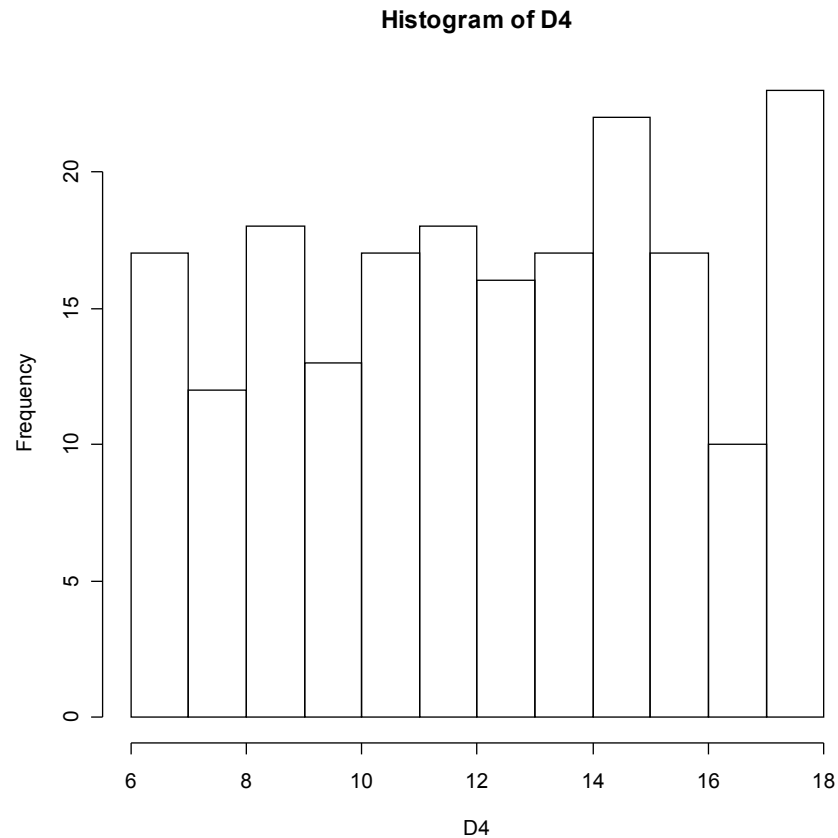
This distribution has a great deal of positive skew.



This distribution is somewhat negatively skewed.

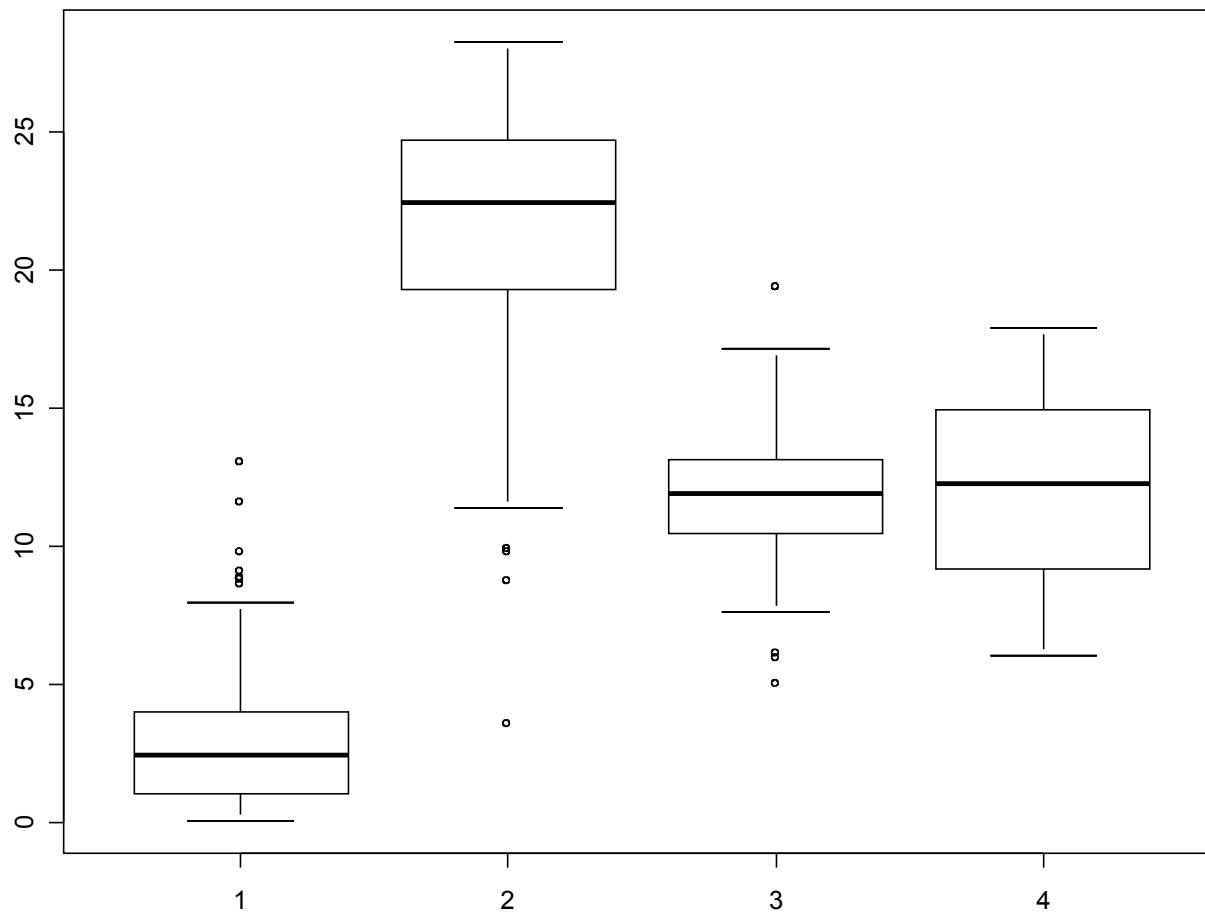


This distribution does not appear to have any skew. In fact, this distribution looks to be appropriately normal.



This distribution also does not appear to have much skew. These data could be from a uniform (straight across) distribution.

Here are the boxplots for all of those distributions:



# Think About It...Hard!

How would skew, particularly major skew, affect the mean?

The median?

The mode?

Which of the Three Ms would be affected the most?

The least?