

GRAPHIC STATISTICS

STEM AND LEAF PLOT

Given a data set, say: 13.3, 13.4, 14.0, 14.1, 14.3, 14.3, 14.7, 15.3, 15.5, 15.9, 16.1, 18.3, 18.6

1. Sort data set from minimum to maximum.
2. All digits except the rightmost digit is known as the STEM. The rightmost digit is the LEAF. The STEM can be any length, but the LEAF is only one digit.
3. Starting with the minimum value, list the lowest stem and draw a vertical line. List the STEM and LEAFS in order .

EXAMPLE:

13 | 3 4

14 | 0 1 3 3 7

15 | 3 5 9

16 | 1

17 |

18 | 3 6

Notice that we include 17 as a STEM although there are no data points. There are no LEAFS in the 17 row.

HISTOGRAM

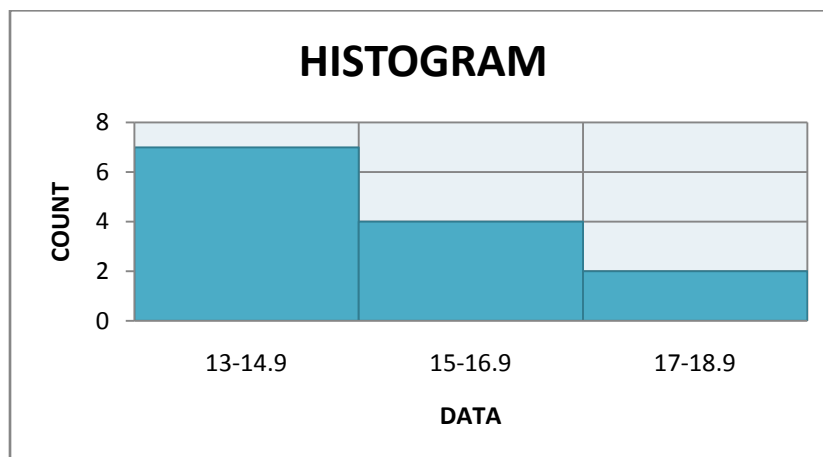
1. Divide the data into classes of equal width. Our data ranges from 13.3 to 18.6, so let's make our range from 13 to 19.

2. Now we can choose any number of classes we desire. Let us choose three classes for our example, and then divide our range into those classes:

Since $\frac{19-13}{3} = 3$, we'll add 3 to 13 to get 15, and make our first class just under 15- i.e., 13-14.9. We'll create our other two classes the same way:

CLASS	FREQUENCY (count in each class)	PERCENT ($\frac{freq}{n}$)
13-14.9	7	53.8
15-16.9	4	30.8
17-18.9	2	15.4

3. Now plot each class against the frequency (or percent, to obtain a *relative frequencies* histogram):



BASIC BOX PLOTS

Boxplots are a good way to visualize the spread and central tendency of a data set, but they can also be used to compare different data sets. Boxplots are based on the *five-number summary*, which consists of the:

Minimum Q_1 Median Q_3 Maximum

of a data set.

For our data set, the five number summary is:

13.3 14.05 14.7 16.0 18.6

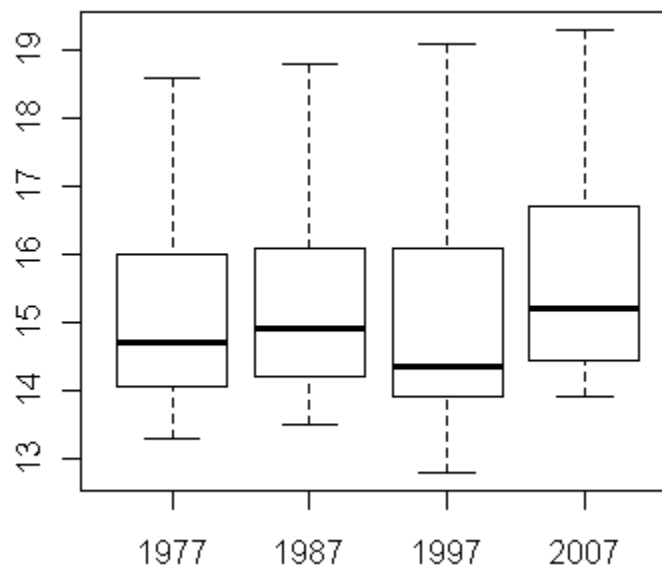
Let's say that our data is (made-up) data for gas mileage for a 4WD vehicle in 1977. Let's compare the data for a 4WD vehicle in 1987, 1997 and 2007:

1987: 13.5 14.2 14.9 16.1 18.8

1997: 12.8 13.9 14.35 16.1 19.1

2007: 13.9 14.45 15.2 16.7 19.3

The box plots for each year are illustrated below:



The bottom line indicates the MINIMUM.

The bottom of the box is Q_1 .

The line in the middle of the box is the MEDIAN.

The top line of the box is Q_3 .

The top line indicates the MAXIMUM.

If the box is relatively centered, and the MEDIAN is in or near the center of the box, this is a good indication that our data is normally distributed.