

Graphs 2. One variable with groups

Please work through “Graphs 1” before this. That includes introductory material which will not be repeated here.

Example 2.9 in Chapter 2, under section 2.10 Organizing a Statistical Problem, asks you to compare data on flower length from three groups.

It is common to organize such data with all the lengths in one column and the group names in another column. Thus, we need to see how to tell the software to look at only the data in one of these groups at a time.

For graphing, in CrunchIt, we do that with a “Filter.”

Here is the example (right) and the data in CrunchIt (left).

The left screenshot shows the CrunchIt software interface. The data table is as follows:

Row	Col 1	Col 2	Col 3	Col 4	Col
1	Variety	Length			
2	bhai	47.12			
3	bhai	46.75			
4	bhai	46.81			
5	bhai	46.67			
6	bhai	47.43			
7	bhai	46.44			
8	bhai	46.84			
9	bhai	45.07			
10	bhai	48.34			
11	bhai	48.15			
12	bhai	50.26			
13	bhai	50.12			
14	bhai	46.54			
15	bhai	46.84			
16	bhai	43.36			
17	red	41.9			
18	red	42.81			
19	red	41.53			
20	red	43.09			
21	red	41.67			
22	red	41.69			

The right screenshot shows a textbook page titled "2.10 Organizing a statistical problem". It includes a table of mean lengths and standard deviations for three varieties: bhai, red, and yellow.

Variety	Mean length	Standard deviation
bhai	47.60	1.213
red	39.71	1.799
yellow	36.18	0.975

Below the table is a stemplot showing the distribution of flower lengths for each variety. The stemplot is as follows:

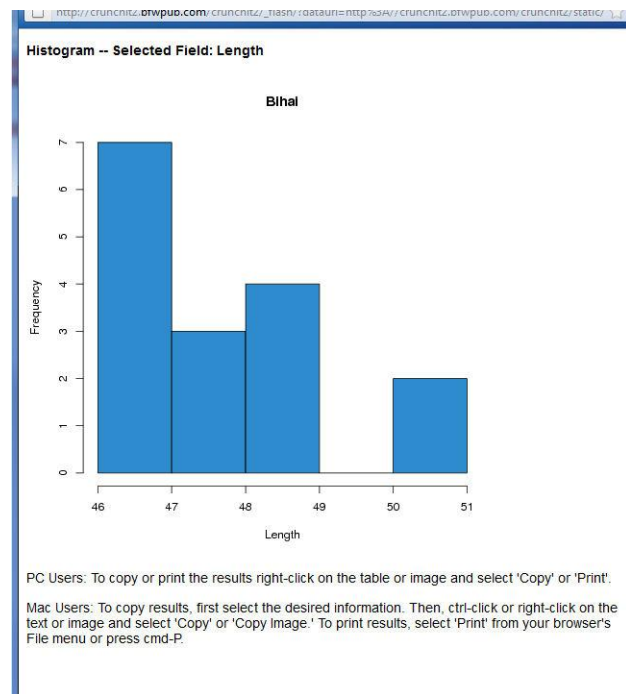
bhai	red	yellow
34	34	34 6 6
35	35	35 2 6 7
36	36	36 0 0 1 5 7 8 8
37	37	37 0 1
38	38	38 0 0 1 1 2 2 6 8
39	39	39 2 6 8
40	40	40 6 7

We will make a frequency histogram of the data for the Bhai group of lengths. In the dialog box for a histogram, we choose “filter” and choose “Variety = bhai” in order to look at only that part of the data.

The screenshot shows the CrunchIt! software interface. On the left is a navigation menu with options like Data, Statistics, Graphics, Bar Plot, Pie Chart, Histogram, Stem and Leaf, Box Plot, Dot Plot, Scatter Plot, Pairs Plot, QQ Plot, Parallel Coordinates, Stars Plot, Line Plot, Preferences, and Help. The main window displays a data table with columns for Row, Variety, Length, and others. A 'Histogram' dialog box is open, allowing the user to choose a variable (Length) and an option (Frequency). The dialog also includes fields for Plot Labels (Title: 'bihai length', X Label: 'Length', Y Label: 'Frequency'), Optional Parameters (Number of Bins, Bin Width, Start Bins At), and a Filter data section.

Row	Col 1	Col 2	Col 3	Col 4	Col 5	Col 6
#	Variety	Length				
1	bihai	47.12				
2	bihai	46.75				
3	bihai	46.81				
4	bihai	47.12				
5	bihai	46.67				
6	bihai	47.43				
7	bihai					
8	bihai					
9	bihai					
10	bihai					
11	bihai					
12	bihai					
13	bihai					
14	bihai					
15	bihai					
16	bihai					
17	red					
18	red					
19	red					
20	red					
21	red					
22	red					
23	red					
24	red					
25	red					
26	red					
27	red					
28	red					
29	red					
30	red					
31	red					
34	red	38.2				

That produces this histogram with frequencies on the vertical axis.



To compare the flower length of these three different groups, compare graphs and numerical summaries. Read the example in the textbook for that discussion.