

DESCRIPTIVE STATISTICS

Graphical Description of Data

A graphical presentation of data will often reveal information about the sample that is not apparent from a table of data measurements.

Goals

1. To be able to display data using various graphing techniques.
2. To be able to interpret a graphical display.

Important Terms and Definitions

Classes are intervals used to group sample measurements.

Class frequency is the number of observations in a particular class. We denote the frequency for a class by f_i .

Relative class frequency is the proportion of the total measurements that fall in a particular class. That is, the relative frequency for class i is f_i/n .

Type of Graphical Display for Qualitative Data

1. Bar Chart

Types of Graphical Displays for Quantitative Data

1. Relative frequency histogram
2. Stem and leaf display

Some Examples

Example: Below are the Major Course Codes for students in an EX ST 301 class.

109	118	118	118	118	118	118	125	125	125	125	130
130	163	163	220	230	402	402	402	402	402	410	410
410	410	415	420	565	710	710	710	710	710	710	710
710	710	710	710	710	811	865	930	960			

CAFLS	{	109	Agronomy		402	Freshman Engineering	}	CES
		118	Animal Science		410	Ceramic Engineering		
		125	Aquaculture Fisheries Wildlife Bio		415	Chemical Engineering		
		130	Community & Rural Development		420	Civil Engineering		
		163	Horticulture		565	Textile Management		
		811	Biological Sciences					
	}	865	Microbiology		710	Nursing	}	HEHD
		930	Forest Resource Management		960	Parks, Recreation, and Tourism		
AAH	{	220	Design (Landscape Arch -BS)					
		230	Landscape Architecture					

Is this an example of qualitative or quantitative data?
Why?

Example: Below are Math Level II Achievement Test scores for twenty-eight randomly selected students.

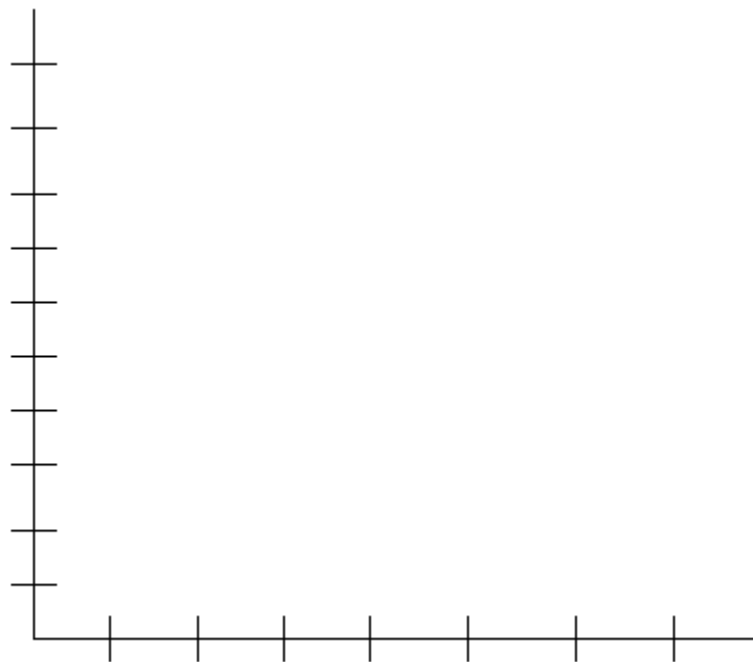
390	400	410	440	450	450	450	450	460	460	460
470	470	470	470	480	480	480	480	480	480	500
500	500	500	500	510	520					

Is this an example of qualitative or quantitative data?
Why?

Constructing a Bar Chart

1. Arrange the categories along the horizontal axis.
2. Construct rectangles over each category, with their height being the frequency for each category.

Bar Chart for College of the Major



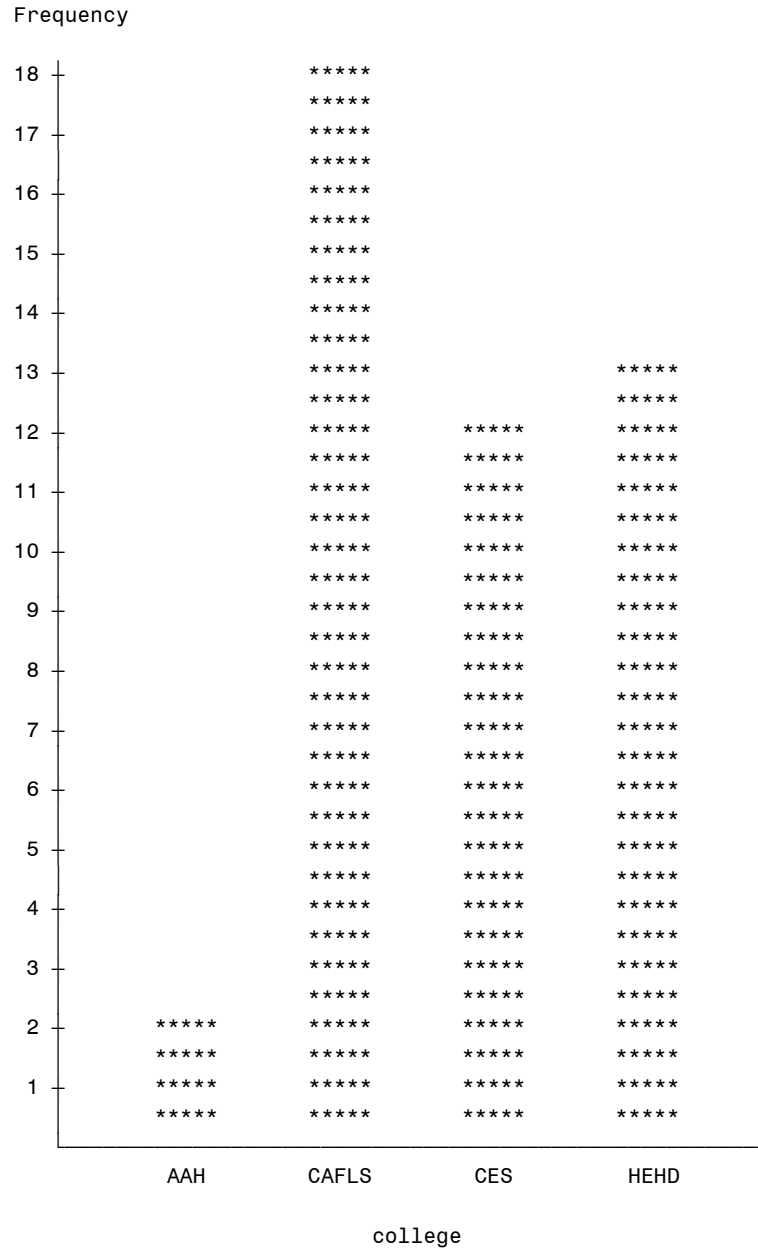
Below is the SAS code for producing a bar chart for college of the major. The bar chart produced from this code can be found on the next page.

```

data barchart;
input code $ @@;
if code='109' or code='118' or code='125' or code='130' or code='163'
or code='811' or code='865' or code='930'
then college='CAFLS';
if code='220' or code='230'
then college='AAH';
if code='402' or code='410' or code='415' or code='420' or code='565'
then college='CES';
if code='710' or code='960'
then college='HEHD';
cards;
109 118 118 118 118 118 118 125 125 125 125 130 130 163 163 220 230 402
402 402 402 402 410 410 410 410 415 420 565 710 710 710 710 710 710 710
710 710 710 710 710 811 865 930 960
;
proc sort;
by college;
proc chart;
vbar college/type=freq;
title Bar Chart for College of the Major;
run;
quit;

```

Bar Chart for College of the Major



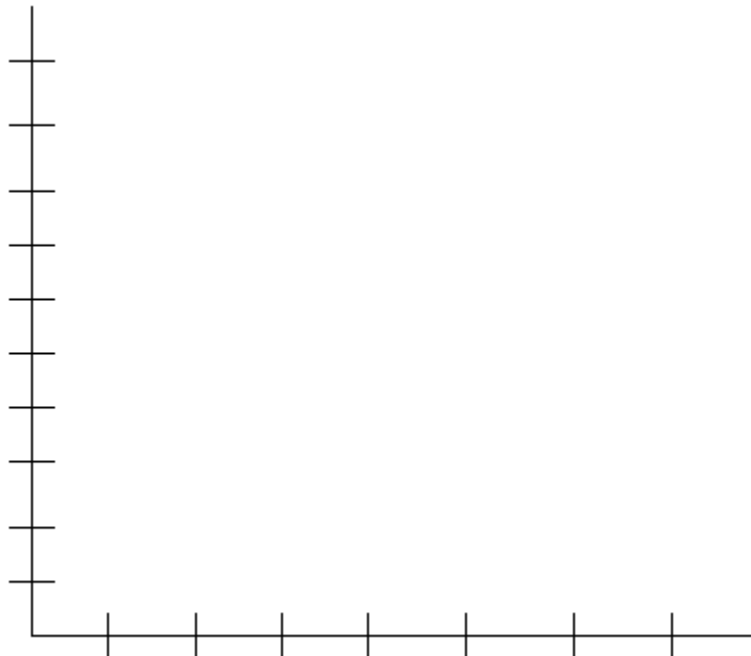
Constructing a Relative Frequency Histogram

1. Determine the number of classes. Usually 5 to 20 classes. The chart below gives an idea of how many classes should be formed.

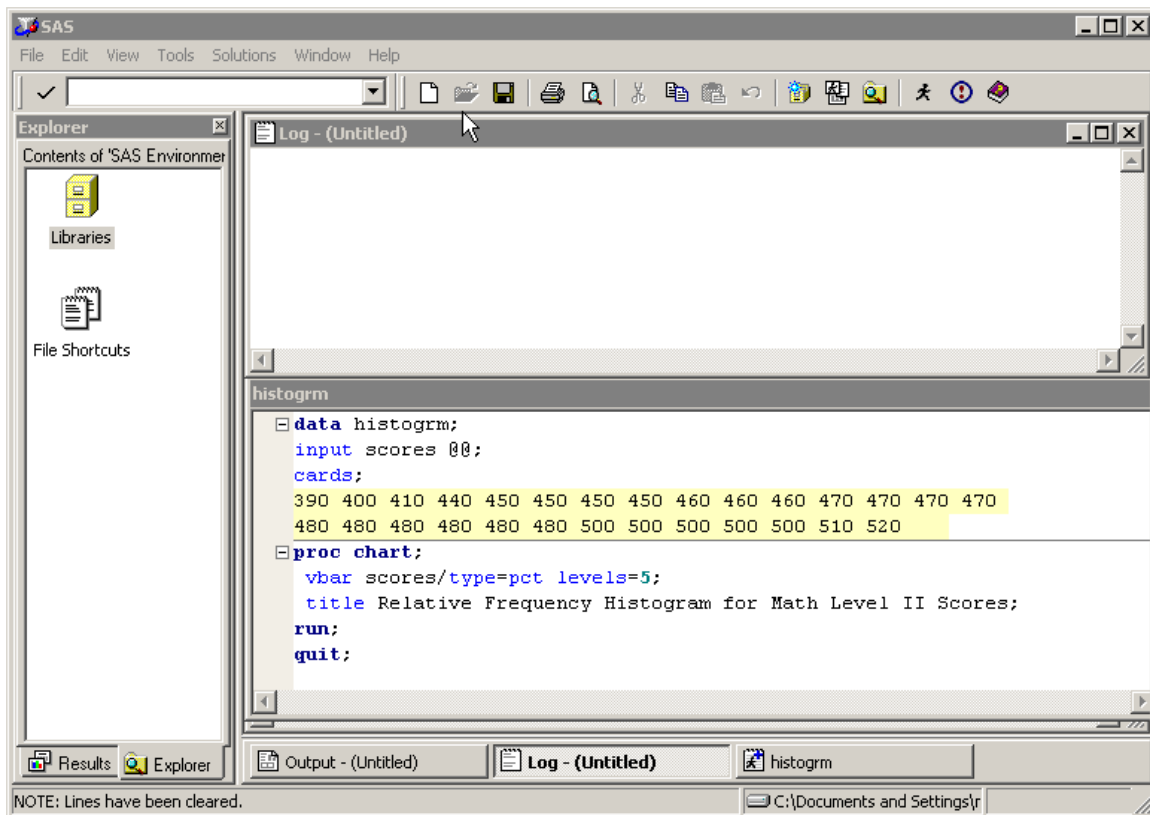
n	Classes
9 - 16	4
17 - 32	5
33 - 65	6
66 - 132	7

2. Determine the class width. To approximate the class width, divide the difference between the largest and smallest measurements by the number of classes. Round this number to a convenient figure. All classes should be the same width.
3. Locate the class boundaries. The lowest class must include the smallest observation, and the highest class must include the largest observation. Boundaries should be chosen so that measurements do not fall on the boundary.
4. Construct rectangles over each class interval, with their height being the relative class frequencies.

Relative Frequency Histogram for Math Level II Scores



Below is the SAS code for producing a histogram using the Math Level II scores. The histogram produced from this code can be found on the next page.



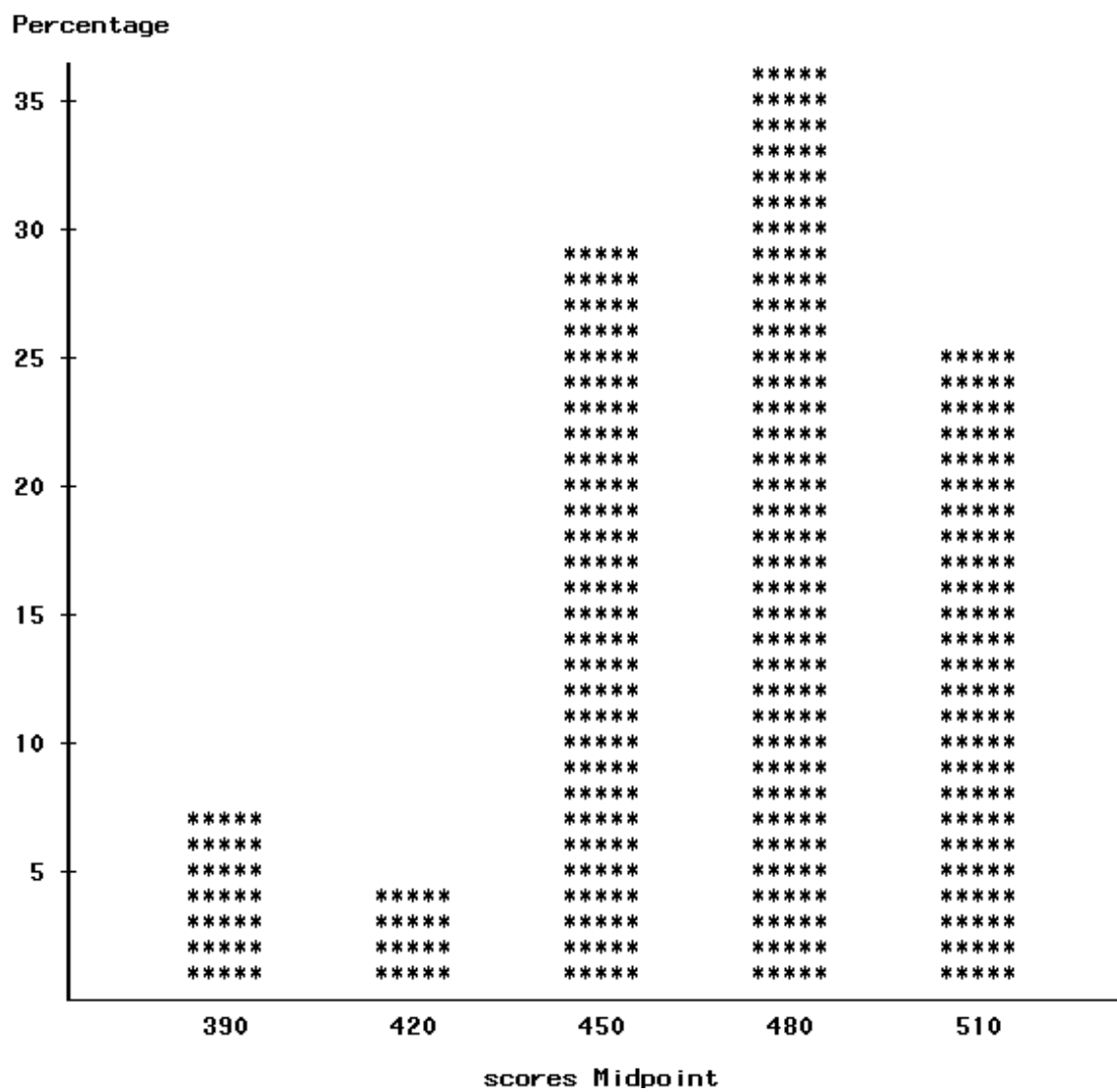
The screenshot shows the SAS software interface. The main window is titled "Log - (Untitled)". On the left, there is an "Explorer" pane showing the "Contents of 'SAS Environment'" with "Libraries" and "File Shortcuts". The main editor area contains the following SAS code:

```
data histogrml;
input scores @@;
cards;
390 400 410 440 450 450 450 450 460 460 460 470 470 470 470
480 480 480 480 480 480 500 500 500 500 500 510 520
proc chart;
vbar scores/type=pct levels=5;
title Relative Frequency Histogram for Math Level II Scores;
run;
quit;
```

The code is written in a monospaced font. The data is entered using the "cards" statement. The "proc chart" statement uses "vbar" to create a vertical bar chart (histogram) with "scores" on the x-axis, "type=pct" for relative frequency, and "levels=5" for the number of bins. The title is "Relative Frequency Histogram for Math Level II Scores".

At the bottom of the window, there is a status bar that says "NOTE: Lines have been cleared." and a file path "C:\Documents and Settings\r".

Relative Frequency Histogram for Math Level II Scores

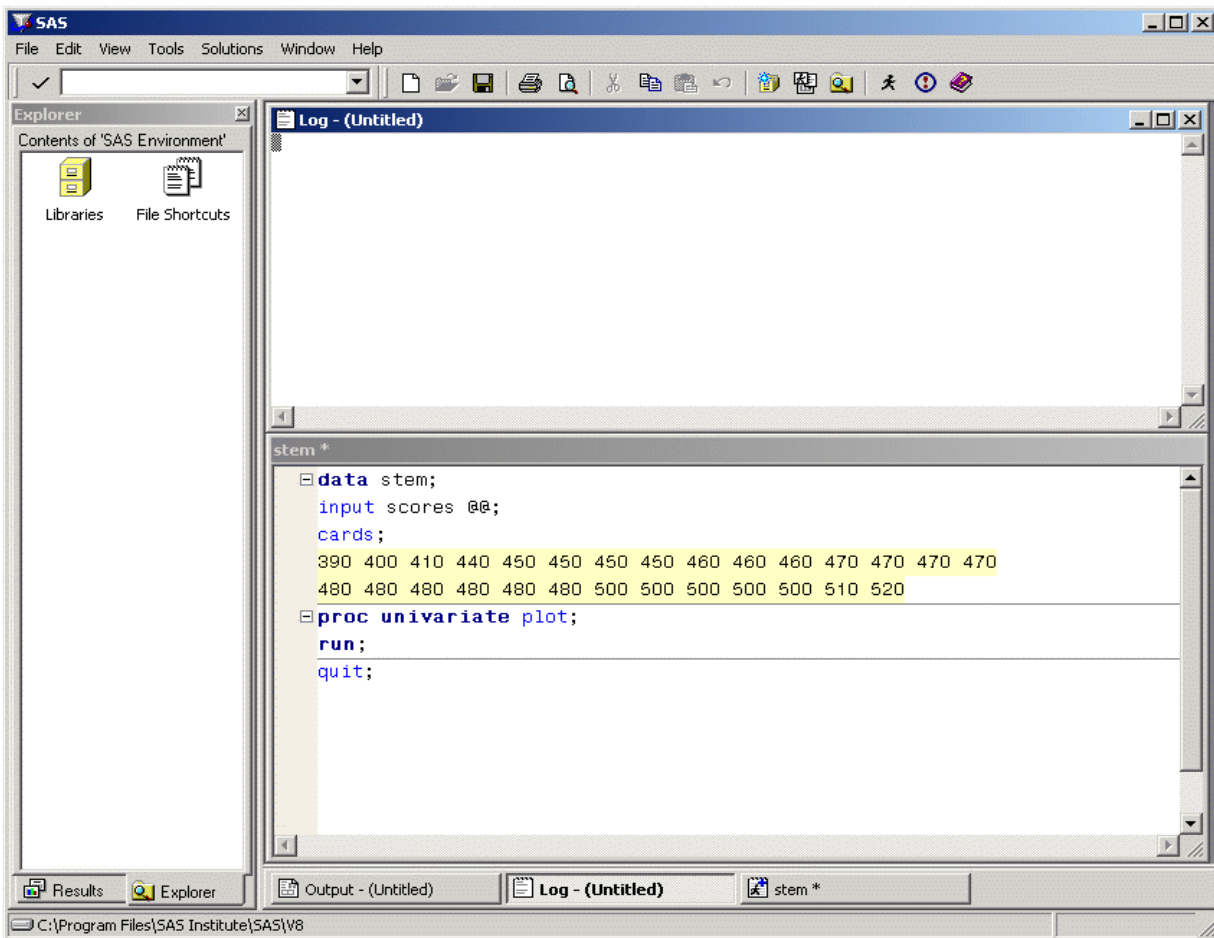


A stem and leaf display is an alternative method for describing a set of data. It looks similar to the relative frequency histogram, but it retains the actual data values.

Constructing a Stem and Leaf Display

1. List the stem values, in order, in a vertical column.
2. Draw a vertical line to the right of the stem values.
3. For each observation, record the leaf portion of the observation in the row corresponding to the appropriate stem.
4. Reorder the leaves from lowest to highest within each stem row.
5. If the number of leaves appearing in each stem is too large, divide the stems into two groups, the first corresponding to leaves with 0 through 4, and the second corresponding to leaves beginning with digits 5 through 9. (This subdivision can be increased to five groups if necessary.)
6. Provide a key to your stem and leaf coding, so the reader can reconstruct the actual measurements.

The following SAS program generated a stem and leaf plot for the Math Level II scores.



Stem	Leaf	#
52	0	1
51	0	1
50	00000	5
49		
48	000000	6
47	0000	4
46	000	3
45	0000	4
44	0	1
43		
42		
41	0	1
40	0	1
39	0	1

-----+-----+-----+-----+
 Multiply Stem.Leaf by 10**+1