STATS 101
Introductory Statistics

Lecture 1

VISUAL SUMMARIZATION OF DATA

GRAPHICAL SUMMARIZATION OF DATA

There are two ways to summarize data – by using a numerical measure (e.g., sample mean and sd) or visually using graphs.

Numerical measures obviously (e.g., sample mean and sd) can only be calculated for numerical (called QUANTITATIVE) variables.

Not all variables are numerical or quantitative – variables can be QUALITATIVE or CATEGORICAL as well.

Examples: (1) Type of credit card a customer carries (AX, Master card, Visa, Discover, Diners) – no ordering exists, this is NOMINAL data.

(2) Grades in a class (A > B > C > D > F) – this is ORDINAL data.

COMMON TOOLS FOR VISUAL SUMMARIZATION OF -

QUANTITATIVE DATA:

- Box plot
- Histogram

QUALITATIVE DATA:

- Bar Chart
- Pie Chart

Age of the presidents at inauguration

President	Age	President	Age	President	Ag
Washington	57	Buchanan	65	Harding	55
J. Adams	61	Lincoln	52	Coolidge	51
Jefferson	57	A. Johnson	56	Hoover	54
Madison	57	Grant	46	F. D. Roosevelt	51
Monroe	58	Hayes	54	Truman	60
J. Q. Adams	57	Garfield	49	Eisenhover	61

President	Age	President	Age	President	Age
Jackson	61	Arthur	51	Kennedy	43
Van Buren	54	Cleveland	47	L. Johnson	55
W. H. Harrison	68	B. Harrison	55	Nixon	56
Tyler	51	Cleveland	55	Ford	61
Polk	49	McKinley	54	Carter	52
Taylor	64	T. Roosevelt	42	Reagan	69
Filmore	50	Taft	51	Bush	64
Pierce	48	Wilson	56	Clinton	46

Plotting a histogram for quantitative (continuous) data

1. Divide sample range into a few equal sub-intervals (called CLASS INTERVALS or BINS) – say 6 intervals.
2. Count # of sample values in each BIN – this number is called the FREQUENCY of the bin.
3. The histogram is a plot of FREQUENCY over the CLASS INTERVALS.

HISTOGRAM PLOT IN EXCEL

1) Type the in Table of slides 5 and 6 – Name of the President in Column A, AGE in Column B. Ages then will be in the cells B2 – B43.

2) Go to a blank cell, say (E5), type in: =min(B2:B43)

3) Go to cell (E6), type in: = max(B2:43)

4) In cell (E7), calculate RANGE by typing = B6 – B5

5) In cell (E8), calculate class width h by typing = B7/6 (for 6 class intervals)

6) Next create BIN values as follows:

(a) In cell D10 type = E5 (1^{st} value = a_1 = min)

(b) In cell D11 type = D10 + E8 ($a_2 = a_1 + h$)

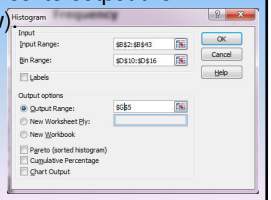
(c) Copy D11 in cells D12 – D16.

The BIN values are now in the cells D11 – D16.

Now we are ready to compute the FREQUENCIES IN EXCEL: click on -

DATA/DATA ANALYSIS/HISTOGRAM – this will open following window.

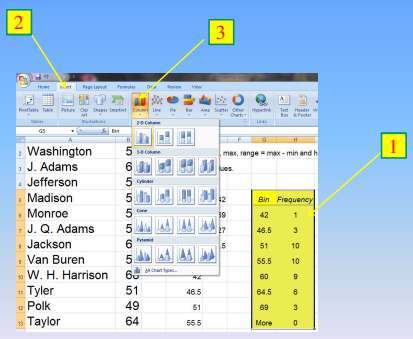
Select data in INPUT RANGE, Bin values in Bin range, and select a cell for excel to output the frequencies (as shown below)

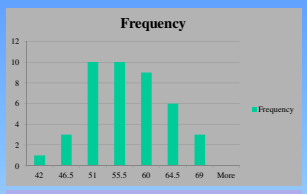


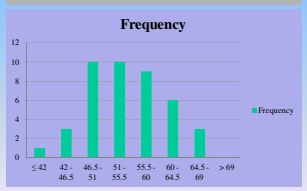
Histogram can now be plotted in EXCEL as follows:

Highlight the frequency table (including header) produced by excel.

Click on INSERT







The bottom histogram is obtained by manually typing bin values as '≤ 42', '42 – 46.5', … and plotting another histogram.
See excel file HISTOGRAM.XLSX

PIE CHART –Example

Marital Status	Count	%
Single	41.8 mil	22.6
Married	113.3 mil	61.1
Widowed	13.9 mil	7.5
Divorced	16.3 mil	8.8

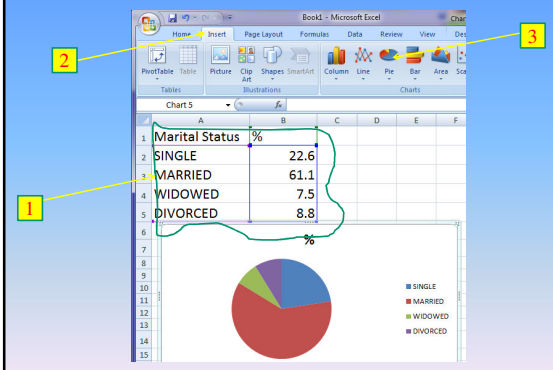
PIE CHART IN EXCEL

(1) Type data in EXCEL as shown in the table on the right.

(2) Highlight the data including the HEADER Cells.

(3) INSERT/PIE will create the first PIE CHART shown on the next slide.

Marital Status	%
SINGLE	22.6
MARRIED	61.1
WIDOWED	7.5
DIVORCED	8.8



To show category names and % in the pie chart:

(a) Right click on the pie chart itself., and click on 'Add Data Labels'. This will show the % values on the pie chart.
(b) Right click on the pie chart again, then click on ' Format Data Labels'. The following window will open. Click on the boxes shown in the window to get the following pie chart (see EXCEL file PIE CHART.xlsx).