2 Presenting Data in Tables and Charts

2.1 Tables and Charts for Categorical Data

**Bar charts, Pareto charts,** and **pie charts** graph the relative frequency of categorical data.

relative frequency= $\frac{\text{frequency in the category}}{\text{total \# of observations}}$

Example: Poll of 400 students at a university, we have 138 Democrats, 146 Republicans, and 116 Independents.
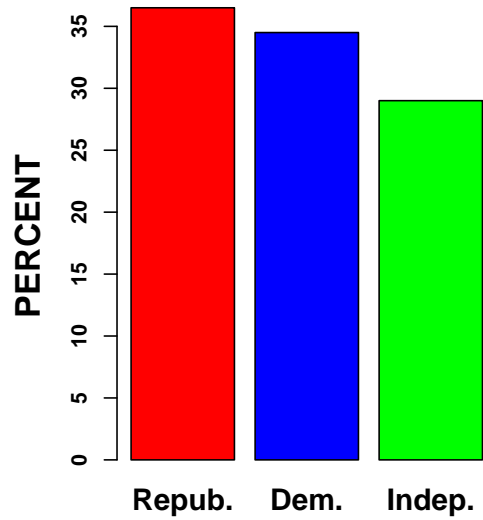
Construct the **bar chart, Pareto chart,** and **pie chart.**

"A **summary table** indicates frequency, amount, or percentage of items in a set of categories so that you can see differences between categories."
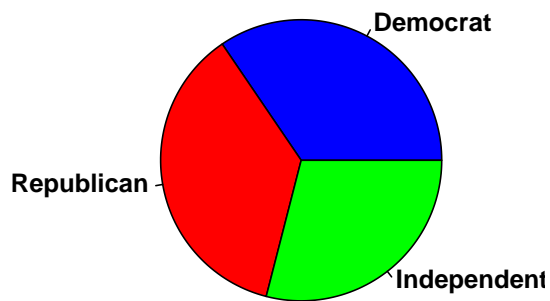
BAR CHART
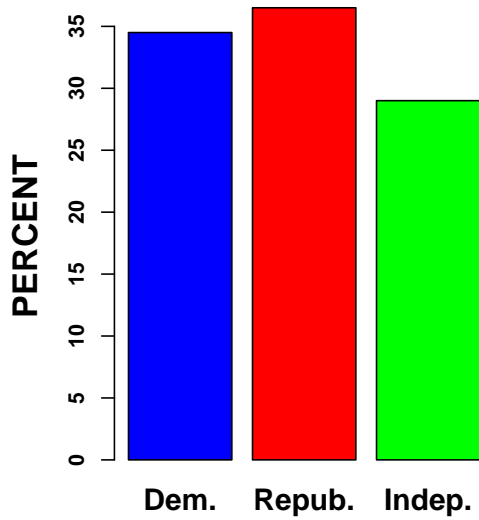
PARETO CHART

PIE CHART

What is the **mode** in the above chart?

☐

2.2 Organizing Numerical Data

Graphs for Quantitative Variables

Example:

Data: 43, 38, 25, 41, 13, 24, 30, 10, 17, 5, 46, 33, 5, 29, 58, 62, 95, 26, 21, 6, 46

Ordered array: 5, 5, 6, 10, 10, 13, 17, 21, 24, 25, 26, 29, 30, 33, 34, 38, 41, 43, 46, 46, 57, 58, 62, 95

Stem-and-Leaf Display

Each stem represents tens, and each leaf represents ones, in this example.

Stems	Leaves
0	5 5 6
1	0 0 3 7
2	1 4 5 6 9
3	0 3 4 8
4	1 3 6 6
5	7 8
6	2
7	
8	
9	5

□

2.3 Tables and Charts for Numerical Data

Histogram

A **histogram** is a graph that uses bars to portray the frequencies or the relative
frequencies of the possible outcomes for a numerical variable.

Discrete case with a small number of possible outcomes:

Example: Construct a relative frequency histogram for data on household
sizes.

# of people	frequency
1	34
2	51
3	42
4	30
5	20
6	13
7	4
8	6

Discrete case with a large number of possible outcomes, OR *continuous* case:

Note: Select the intervals of the histogram to be of equal width, for simplicity.

Example: Construct a relative frequency histogram for income (in terms of
hourly wage).

income level	# of people
$0 to $9.99	35
$10 to $19.99	50
$20 to $29.99	70
$30 to $39.99	115
$40 to $49.99	100
$50 to $59.99	130

2.4 Cross Tabulations

Set up a **contingency table**, for comparing two categorical variables.

Within a contingency table, we can determine **conditional proportions**; i.e.,
the proportion of the time that a variable takes on a particular value, conditional
on some value of the other variable.

Example: Consider the following contingency table regarding the gender of an
unborn baby (The data are hypothetical but are somewhat consistent with
untrasounds from the 1990's.):

|	Ultrasound Predicted Female	Ultrasound Predicted Male
Actual gender is female	432	48
Actual gender is male	130	390

Based on the above contingency table:

(a) What proportion of babies are *female*, given that the ultrasound predicted
that the baby would be *female*?

(b) What proportion of babies are *male*, given that the ultrasound predicted that
the baby would be *male*?

(c) Is the ultrasound equally reliable for predicting gender for boys and for girls?

2.5 Scatter Plots and Time-Series Plots

Scatter Plot

A **scatter plot** graphically illustrates the relationship between two numerical
variables.

Example: Heights of 1078 fathers and sons, England, around year 1900.
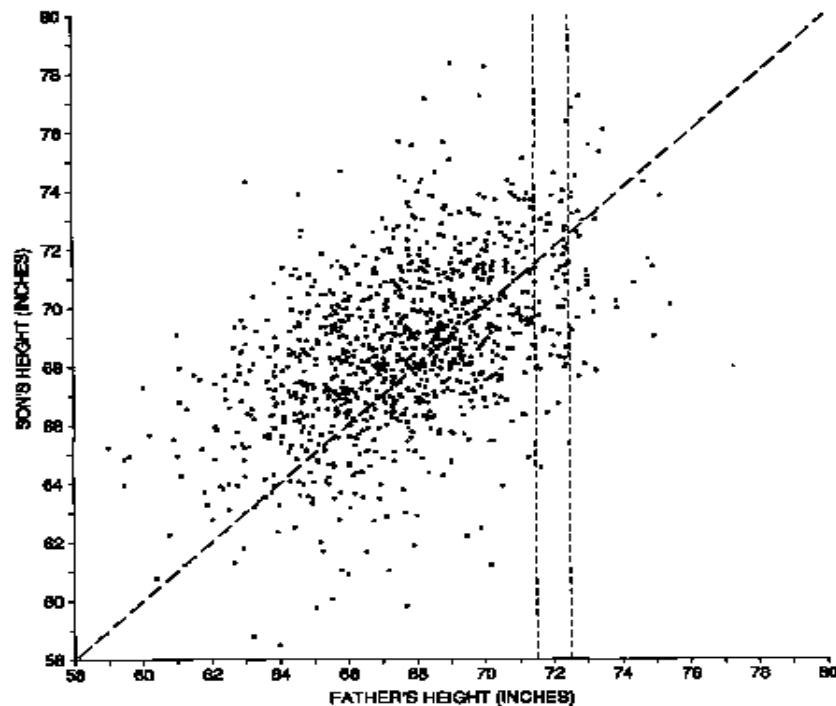
Example: In the Presidential Election of 2000, George W. Bush earned 537 votes more than Al Gore in Florida, granting the Presidency to Bush. However, the Palm Beach County, Florida, the "butterfly ballot" possibly caused some individuals to mistakenly vote for Pat Buchanan rather than Gore.
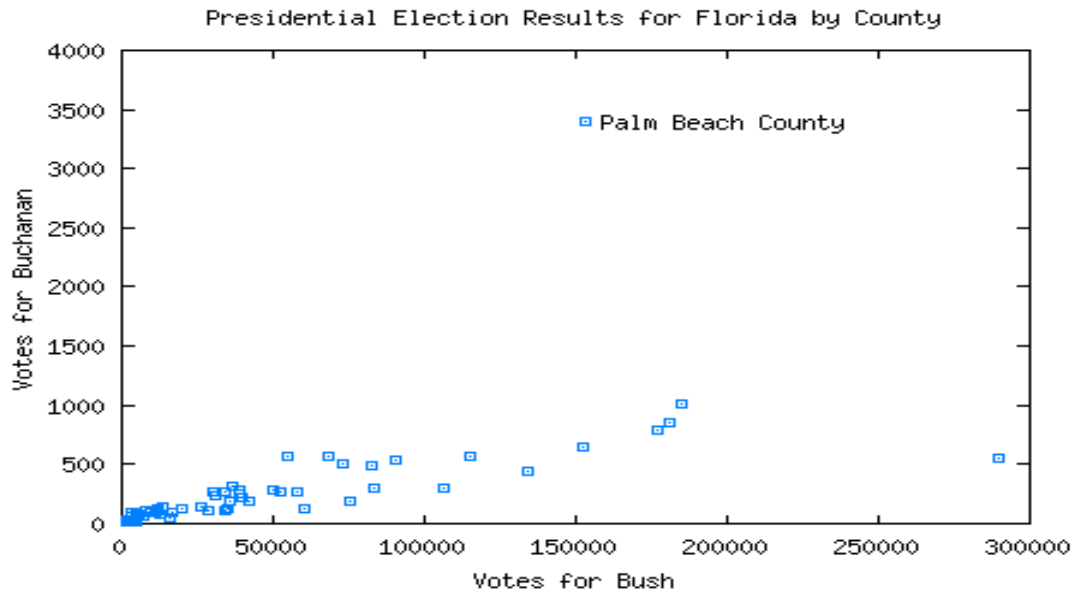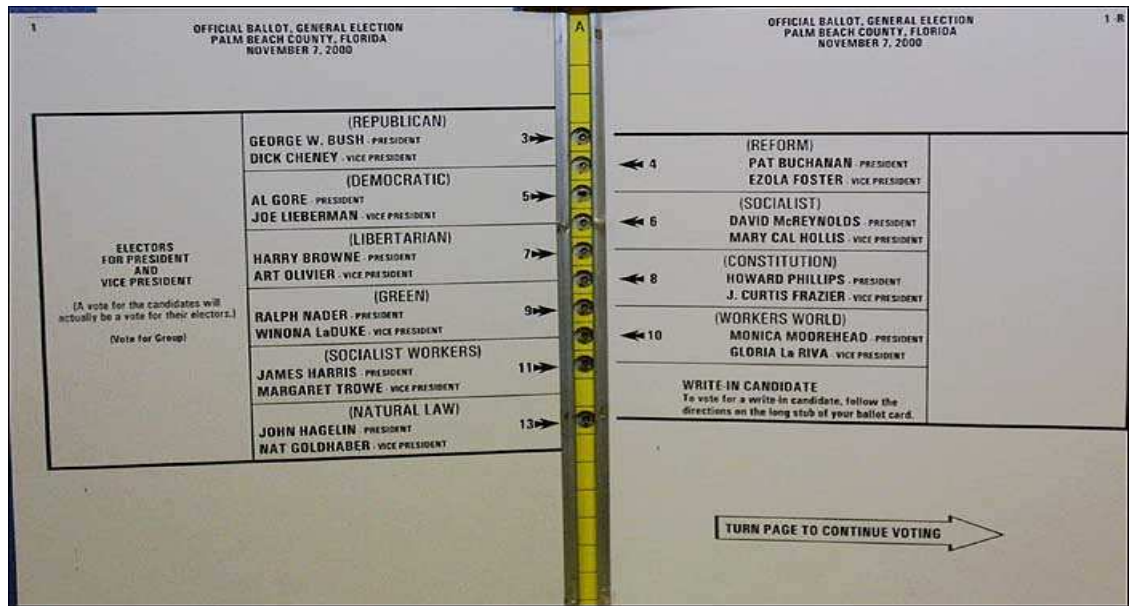
Time-series plot

A *times-series* plot is the plot of a variable against time.

Read pp. 69–79, Appendix E2: Using Microsoft Excel for Tables and Charts.