2 Graphical Summaries of Data

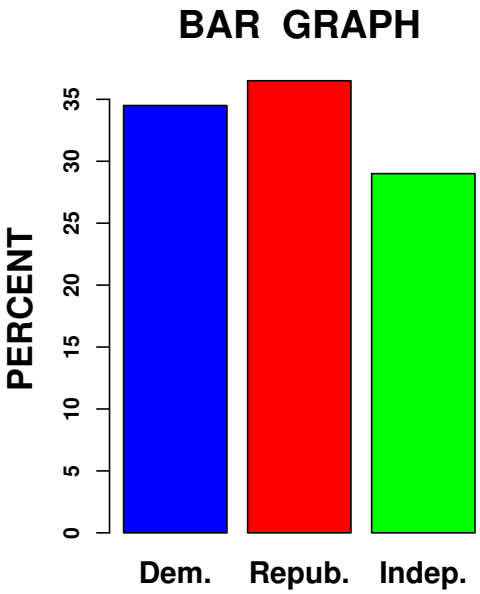
2.1 Graphical Summaries of Qualitative Data

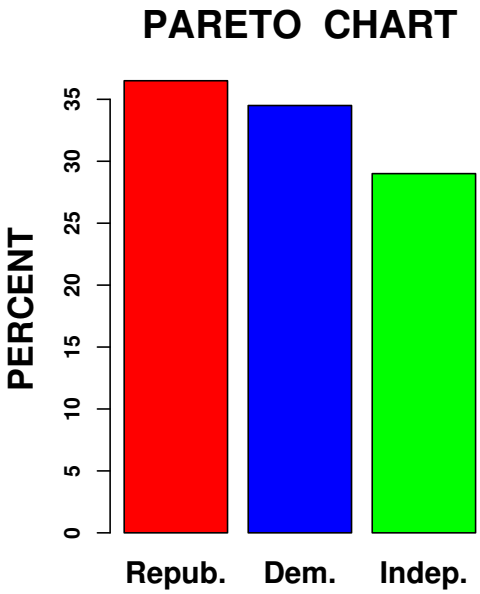
Bar graphs, Pareto charts, and **pie charts** graph the relative frequency of qualitative data.

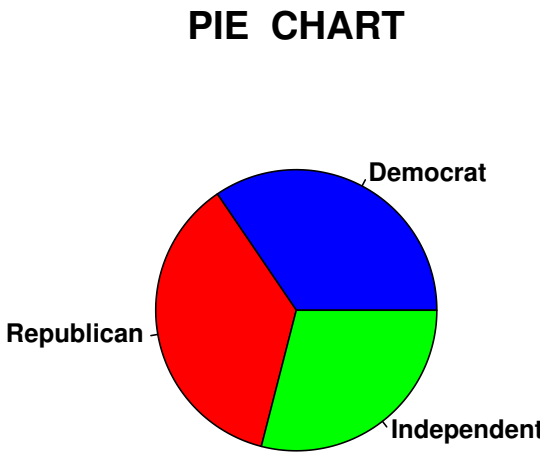
$$\textbf{relative frequency} = \frac{\text{frequency in the category}}{\text{total \# of observations}}$$

Example: Poll of 400 students at a university, we have 138 Democrats, 146 Republicans, and 116 Independents.

Construct the **bar graph, Pareto chart,** and **pie chart.**







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

□

2.2 Frequency Distributions and Their Graphs

Consider **quantitative** variables.

Histograms

A **histogram** is a graph that uses rectangles to portray the frequencies or the relative frequencies of the possible outcomes for a quantitative 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

Sample histograms and population histograms

Compare and contrast sample histograms with population histograms.

The shape of a distribution

A histogram might be described as

1. unimodal

2. bimodal

3. multimodal

A **sample or population histogram** might be described as

1. symmetric
2. skewed to the right
3. skewed to the left

2.3 More Graphs for Quantitative Data

Stem-and-Leaf Plots

Example:

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

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

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

□

Stem-and-leaf plots can be used to observe the shape (and location and spread) of the data or detect *outliers*.

An **outlier** is an observation that falls well above or well below the overall bulk of the data.

Example: In a random sample of heights of 100 women, a 6-foot-8-inch-tall woman is recorded.

Dotplots

Example: Construct the *dotplot* for the following data on personal income (in thousands of dollars): 35, 49, 70, 21, 49, 80, 57, 160.

Time-Series Plots

A *time* plot is the plot of a variable against time.