WHAT IT IS

Graphs are pictorial representations of the relationships between two (or more) variables and are an important part of descriptive statistics. Different types of graphs can be used for illustration purposes depending on the type of variable (nominal, ordinal, or interval) and the issues of interest. This module discusses some of the more common types of graphs.

WHEN TO USE IT

Graphs can be used any time one wants to visually summarize the relationships between variables, especially if the data set is large or unmanageable. They are routinely used in reports to underscore a particular statement about a data set and to enhance readability. Graphs can appeal to visual memory in ways that mere tallies, tables, or frequency distributions cannot. However, if not used carefully, graphs can misrepresent relationships between variables or encourage inaccurate conclusions. (See the module on Deception with Graphs for more information.)

TYPES OF GRAPHS

Bar Graphs

Bar graphs are commonly used to show the number or proportion of nominal or ordinal data which possess a particular attribute. They depict the frequency of each category of data points as a bar rising vertically from the horizontal axis. Bar graphs most often represent the number of observations in a given category, such as the number of people in a sample falling into a given income or ethnic group. They can be used to show the proportion of such data points, but the pie chart is more commonly used for this purpose. Bar graphs are especially good for showing how nominal data change over time.

Advantages - Bar graphs can:
•   show each nominal or ordinal category in a frequency distribution
•   display relative numbers or proportions of multiple categories
•   summarize a large data set in visual form
•   clarify trends better than do tables or arrays
•   estimate key values at a glance
•   permit a visual check of the accuracy and reasonableness of calculations
•   be easily understood due to widespread use in business and the media

Disadvantages - Bar graphs can:
•   require additional written or verbal explanation
•   be easily manipulated to yield false impressions
•   be inadequate to describe the attribute, behavior, or condition of interest
•   fail to reveal key assumptions, norms, causes, effects, or patterns

An example of a bar graph follows on the next page.

Team Sales
1993

Annual Sales $ (Thousands)
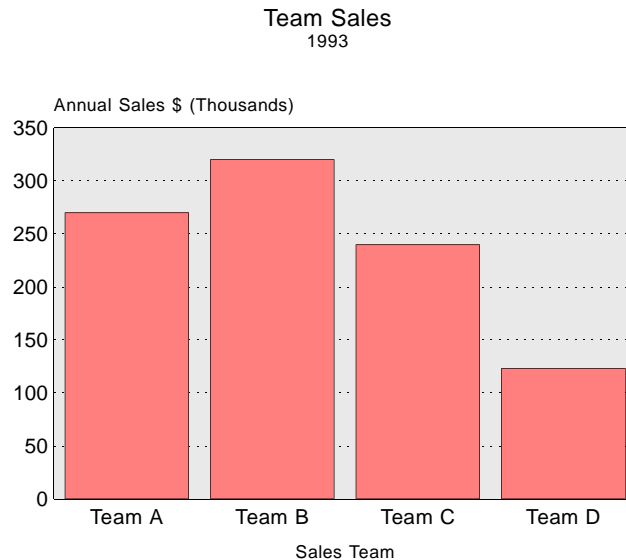
Frequency Polygons

Frequency polygons are the preferred way to graph the frequency distribution of ungrouped (raw) interval data. They represent the frequency of each class of data points as a line connecting the midpoints of the bars of a histogram. The normal (bell) curve is the most common type of frequency polygon. Frequency polygons can describe the behavior of the same interval variable under different circumstances, as in before-after situations, if superimposed on each other.

**Advantages** - Frequency polygons can:
•   begin to show central tendency, dispersion, and clustering/modality
•   estimate key values, especially the mean, and show skew and kurtosis
•   summarize a large data set in visual form
•   clarify trends better than do tables, arrays, and most other graphs
•   show the behavior and distribution of a variable
•   become more smooth as classes or data points are added
•   be easily understood due to widespread use in business and the media

Disadvantages - Frequency polygons can:
•   fail to delineate each interval in a frequency distribution
•   lose details on relative numbers and proportions vis-a-vis the histogram
•   require additional written or verbal explanation
•   be inadequate to describe the attribute, behavior, or condition of interest
•   fail to reveal key assumptions, norms, causes, effects, or patterns
•   be easily manipulated to yield false impressions
•   fail as a visual check of the accuracy or reasonableness of calculations

Examples of both a frequency polygon and histogram follow on the next page.

Distribution of Personal Income

# of respondents
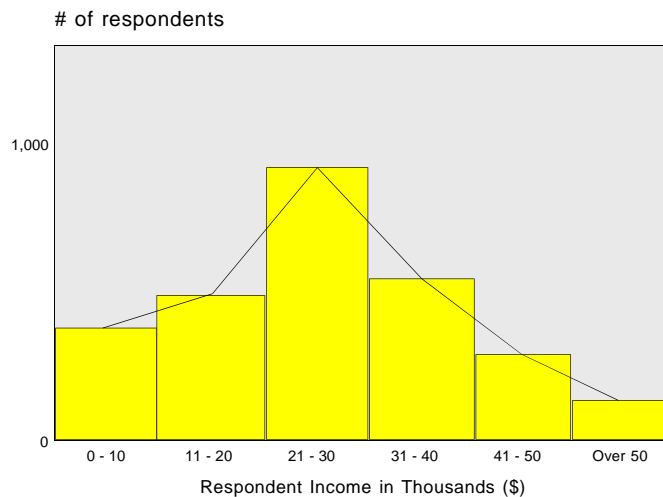
Histograms

Histograms are the preferred method for graphing grouped interval data. They depict the number or proportion of data points falling into a given class. For example, a histogram would be appropriate for depicting the number of people in a sample aged 18-35, 36-60, and over 65. While both bar graphs and histograms use bars rising vertically from the horizontal axis, histograms depict continuous classes of data rather than the discrete categories found in bar charts. Thus, there should be no space between the bars of a histogram.

Advantages - Histograms can:
- begin to show the central tendency and dispersion of a data set
- closely resemble the bell curve if sufficient data and classes are used
- show each interval in the frequency distribution
- summarize a large data set in visual form
- clarify trends better than do tables or arrays
- estimate key values at a glance
- permit a visual check of the accuracy and reasonableness of calculations
- be easily understood due to widespread use in business and the media
- use bars whose areas reflect the proportion of data points in each class

Disadvantages - Histograms can:
- require additional written or verbal explanation
- be easily manipulated to yield false impressions
- be inadequate to describe the attribute, behavior, or condition of interest
- fail to reveal key assumptions, norms, causes, effects, or patterns

An example of a histogram is found with the frequency polygon above.

Line Graphs

Line graphs use a single line to connect plotted points of interval and, at times, nominal data. Since they are most commonly used to visually represent trends over time, they are sometimes referred to as time-series charts.

Advantages - Line graphs can:
- clarify patterns and trends over time better than most other graphs
- be visually simpler than bar graphs or histograms
- summarize a large data set in visual form
- become more smooth as data points and categories are added
- be easily understood due to widespread use in business and the media
- require minimal additional written or verbal explanation

Disadvantages - Line graphs can:
- be inadequate to describe the attribute, behavior, or condition of interest
- fail to reveal key assumptions, norms, or causes in the data
- be easily manipulated to yield false impressions
- reveal little about key descriptive statistics, skew, or kurtosis
- fail to provide a check of the accuracy or reasonableness of calculations

An example of a line graph follows.

Bank Failures
By Year

Thousands

1989	1990	1991	1992	1993
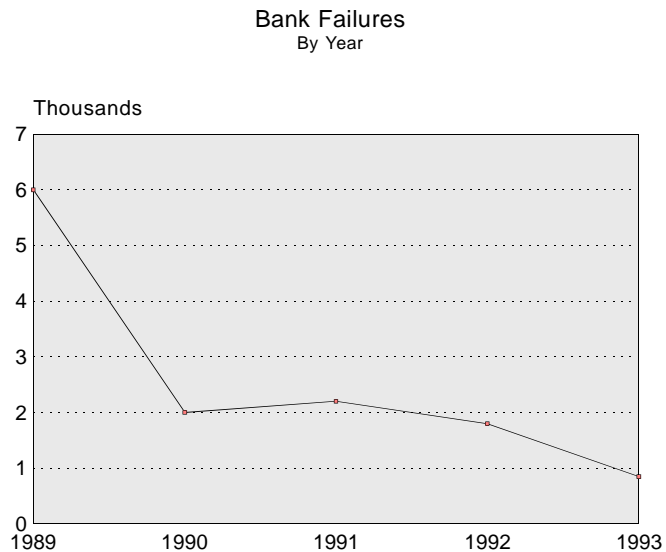
Ogives

Ogives are a type of frequency polygon used to depict the cumulative number or cumulative proportion of grouped interval data. They represent the cumulative frequency of each class of data points as a line connecting the upper limit of each class. Ogives let us make such statements as "X number (or X%) of the people sampled make below $Y." They are also quite useful for discussing rates of change between classes in a data set by measuring how quickly the effect of a variable accumulates across groups of data points.

Advantages - Ogives can:
- visually approximate raw data
- summarize a large data set in visual form
- delineate each interval in the frequency distribution
- clarify rates of change between classes better than other graphs
- provide a visual check of the accuracy or reasonableness of calculations
- become more smooth as data points or classes are added
- be easily understood due to widespread use in business and the media
- show the number or proportion of the data points above/below a particular value

Disadvantages - Ogives can:
- be somewhat complicated to prepare
- fail to reflect all data points in a data set
- reveal little about central tendency, dispersion, skew, or kurtosis
- often require additional written or verbal explanation
- be inadequate to describe the attribute, behavior, or condition of interest
- fail to reveal key assumptions, norms, causes or effects
- be somewhat easily manipulated to yield false impressions

An example of an ogive follows.

Personal Income
Cumulative Distribution

# of respondents

Income at or Below
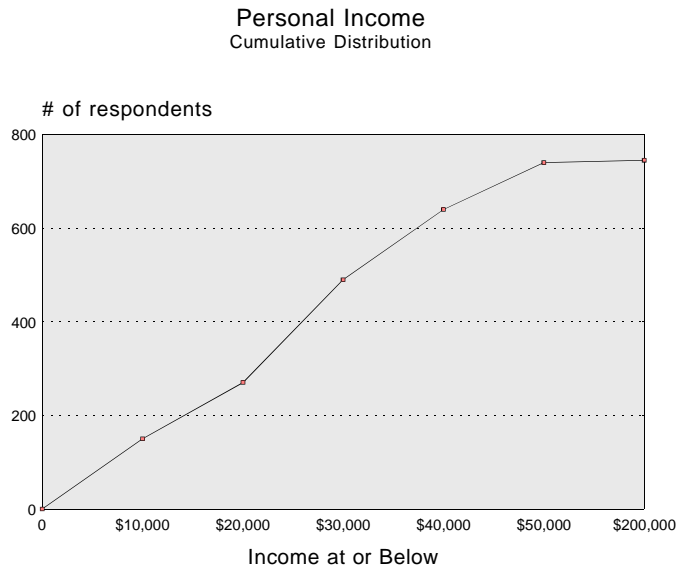
Pie Charts

Pie charts are circles subdivided into a number of "slices." The area of each represents the relative proportion data points falling into a given category. Pie charts are the preferred method for graphing both nominal data and percentages.
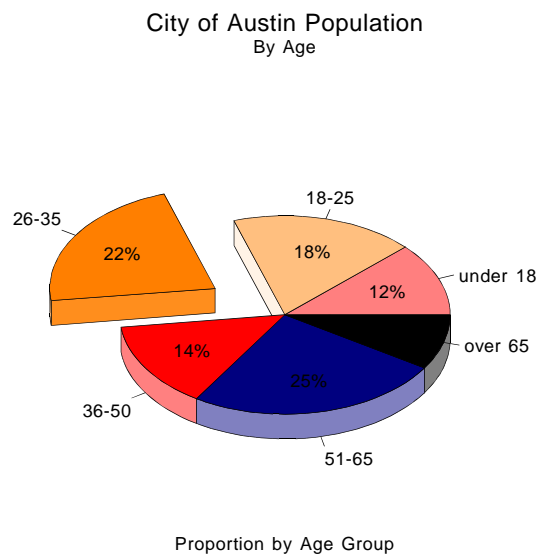
Advantages - Pie charts can:
- display relative proportions of multiple classes of data
- show areas proportional to the number of data points in each category
- summarize a large data set in visual form
- be visually simpler than other types of graphs
- permit a visual check of the reasonableness or accuracy of calculations
- require minimal additional verbal or written explanation
- be easily understood due to widespread use in business and the media

Disadvantages - Pie charts can:
- reveal little about central tendency, dispersion, skew, or kurtosis
- fail to reveal key assumptions, norms, causes, effects, or patterns
- fail to describe the attribute, behavior, or condition of interest
- be easily manipulated to yield false impressions

An example of a pie chart follows:

City of Austin Population
By Age

18-25

26-35   18%

22%   12%   under 18

14%   over 65

25%

36-50

51-65

Proportion by Age Group

Stacked Bar Graphs

Stacked bar graphs (component bar graphs) add information to standard bar graphs by representing multiple categories of data in a single bar. Each bar is divided into sections whose heights indicate the proportion of observations falling into a given category. Since they deal with categories, these graphs are generally used for nominal data.

Stacked bar graphs are useful for comparing the contributions of different groups to the total value of a variable. They can also compare two identical nominal data sets under different conditions, especially at different times. Stacked bar

graphs are often a good alternative to using multiple pie charts since they depict how much the size of a given category changes under varying circumstances. Such graphs also eliminate the need to construct a separate bar for each category, as would occur when using multiple bar graphs.

An example of a stacked bar graph follows.

Truck Owners
By U.S. Region
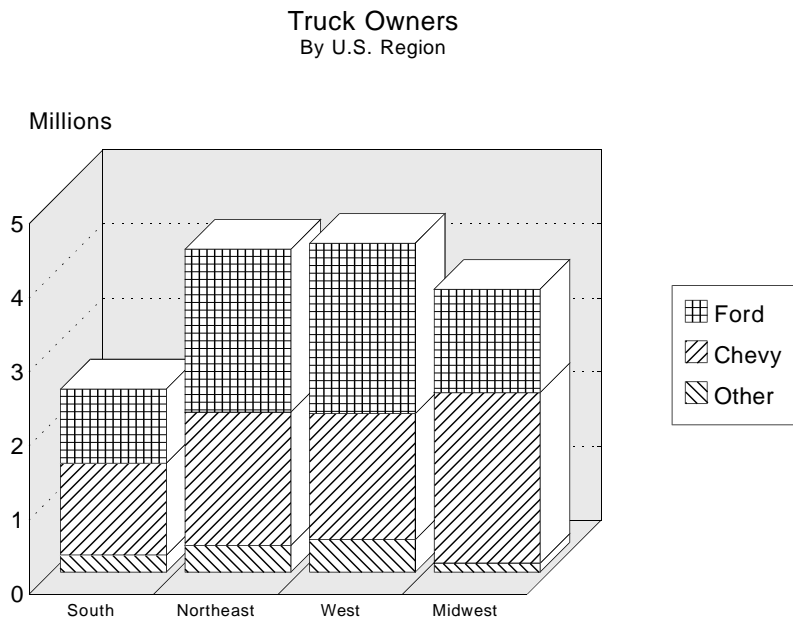
Advantages - Stacked bar graphs can:
- combine richer information than bar graphs or multiple pie charts
- show how each category contributes to the overall value of a variable
- show each nominal or ordinal category in frequency distribution
- display relative proportions of multiple categories
- summarize a large data set in visual form
- clarify trends better than do tables or arrays
- estimate key values at a glance
- permit a visual check of the accuracy and reasonableness of calculations
- be easily understood due to widespread use in business and the media

Disadvantages - Stacked bar graphs can:
- be visually intimidating
- require additional verbal or written explanation
- be somewhat complicated to prepare
- be easily manipulated to yield false impressions
- be inadequate to describe the attribute, behavior, or condition of interest
- fail to reveal key assumptions, norms, causes, effects, or patterns
- reveal little about central tendency, dispersion, skew, or kurtosis

SUMMARY TABLE

The following table outlines when it is most appropriate to use a particular kind of graph. The first column lists the graph type. The second column lists whether a particular graph type is best suited for numbers (#'s) or proportions (%'s). The third column indicates whether numbers are best in raw (R) or grouped (G) form. The last three columns list the type(s) of variables for which the graph is most appropriate (i.e. nominal, ordinal, or interval). A double check mark in a given box indicates the best use of a particular type of graph. A single check mark in a box indicates a suitable but not optimum use.

Graph Type	# / %	R / G	Nominal	Ordinal	Interval
Bar Graph	#	R	✓✓	✓	
Freq. Polygon	#	R			✓✓
Histogram	# / %	G		✓	✓✓
Line Graph	# / %	R	✓		✓✓
Ogive	# / %	G			✓✓
Pie Chart	%	–	✓✓	✓	✓
Stacked Bar	%	–	✓✓	✓	

IN-HOUSE ASSISTANCE

Graph preparation can be complicated. The State Auditor's Office has one full-time professional dedicated to preparing graphics for reports and presentations. For this reason, we do not get into a detailed technical discussion in this module. While creating basic graphs is generally not difficult, producing high-quality, report-ready graphics can be quite time consuming for persons not familiar with the software. Presently, the best in-house software for experimentation is Harvard Graphics. Harvard also has the best in-house support. The spreadsheet software Quattro Pro for Windows also has an extensive graphics capability.

DEALING WITH POST SCRIPT FILES

In the past, it has been common to manually cut and paste graphs and charts into reports. In a computer environment where all software is written to operate with Microsoft Windows, cutting and pasting graphs and charts between software programs is fairly routine. When this is not the case, the alternative to manually inserting graphs is to import Encapsulated Post Script Files (EPS) into word processing software. Graphics created in most software can be exported as EPS files. This is true of both Harvard Graphics and Quattro Pro for Windows.