

CHAPTER TWELVE

TABLES, CHARTS, AND GRAPHS

Tables, charts, and graphs are frequently used in statistics to visually communicate data. Such illustrations are also a frequent first step in evaluating raw data for trends, data entry errors, and outlying values which might impact on the statistical interpretation of the data. A properly drawn chart or graph can answer a number of questions in a minimal amount of space as well as suggest questions not previously considered.

The goal in creating tables, charts or graphs is to present data in a clear and accurate format which is easily interpreted. The multitude of computer graphics programs currently available has greatly simplified the creation of such illustrations. Nevertheless, a poorly designed table or figure will only serve to confuse and even mislead the reader. The following are some guidelines for presenting data visually.

DESIGNING TABLES

Tables are useful for summarizing the raw data upon which a study's conclusions are based. They effectively use a minimum of space to communicate a large amount of information. There are, however, a few important points to keep in mind when creating a table:

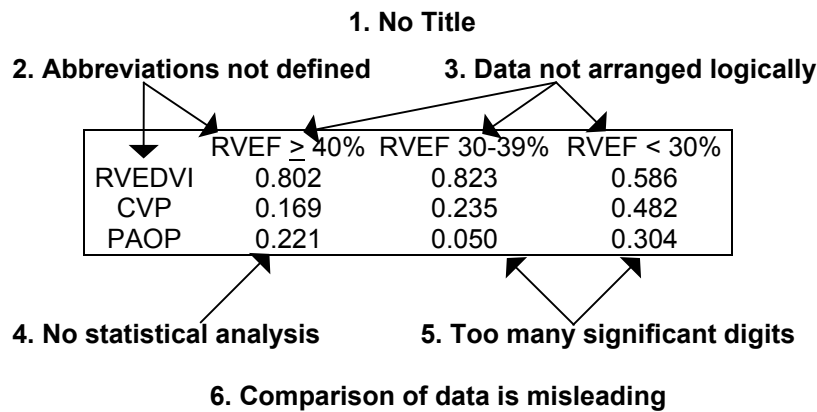
1. **Arrange table rows and columns logically.** This will make the table more easily interpreted and aesthetically pleasing. We each have a preconceived impression of how a table should be arranged. We expect the data to be presented in a logical order. If this is not the case, we must first reorient ourselves to the table before we can begin to interpret the data. Rows and columns should therefore be ordered by size, time, or variable of interest as the reader would normally expect them to be. The order of presentation can also be used to emphasize important data.
2. **Simplify data entries as much as possible.** In most situations, presenting data with more than two significant digits only serves to make the table appear "cluttered". In fact, it is rare that we are capable of measuring a physiologic parameter to more than two decimal places. As a general rule, a number should never have more significant digits than that of the raw data which were used to calculate it or more digits than we are physically able to measure. For example, since we can only measure arterial carbon dioxide tension (PaCO_2) in round numbers (i.e., $\text{PaCO}_2 = 35$ torr), we should not report mean PaCO_2 values of 38.42 torr; rather, we need to round off the mean to the number of significant digits we are capable of measuring and therefore present the mean PaCO_2 as 38 torr.
3. **Important data should be emphasized.** A table is primarily used to convey important information which is used to support the study's conclusions. Emphasizing the table with lines and/or bold face type styles will draw the reader's attention to this data. Important statistics such as mean, median, standard deviation, confidence intervals, and p values can be emphasized in this manner, or set apart in a summary row or column. This is especially important in scientific abstract preparation where space is limited. Statistical significance should be illustrated through the use of symbols which are described in the table legend.
4. **Summarize the data to assist in comparison.** Include row and column means or medians, where applicable, to summarize the information contained in the table. This allows the reader to quickly interpret the table and find important data.

To illustrate these points, consider the appearance of the following table:

	RVEF ≥ 40%	RVEF 30-39%	RVEF < 30%
RVEDVI	0.802	0.823	0.586
CVP	0.169	0.235	0.482
PAOP	0.221	0.050	0.304

Without any accompanying text, this table is uninterpretable as well as misleading. From the table, there is no way of knowing that these are the correlation coefficients (r values) from a study comparing cardiac index with right ventricular end-diastolic volume index (RVEDVI), pulmonary artery occlusion pressure (PAOP) and central venous pressure (CVP) at various levels of right ventricular ejection fraction (RVEF).

A table should be self-explanatory and self-supporting. It should be easily understood apart from its accompanying text. It is unclear in the above example exactly what is being presented and why. Abbreviations are not defined, data are not arranged logically, and there is no statistical analysis given. Furthermore, the table appears to be comparing RVEF with RVEDVI, CVP, and PAOP and is therefore misleading.



Consider the revised table after the above problems have been addressed:

Correlation with Cardiac Index at Various Levels of RVEF

	RVEF < 30%	RVEF 30-39%	RVEF ≥ 40%
n	46	48	55
RVEDVI	0.59***	0.82***	0.80***
PAOP	0.48**	0.24	0.22
CVP	0.30*	0.05	0.17

*p < 0.05 **p < 0.01 ***p < 0.001
 (Linear correlation compared to cardiac index)

n - number of observations
 RVEF - right ventricular ejection fraction
 RVEDVI - right ventricular end-diastolic volume index
 PAOP - pulmonary artery occlusion pressure
 CVP - central venous pressure

This table is easily interpreted apart from its accompanying text. The title clearly defines the information that is being presented. Abbreviations are explained in the accompanying legend. Data are logically arranged by increasing RVEF. The number of observations in each group is reported. The statistical significance for each group is presented and the statistical methods used described. The important values are set apart through the use of bold-face type. The data are also simplified to the appropriate number of significant digits. Careful design of tables in this manner makes the conclusions of an abstract or manuscript more easily recognized and applicable to clinical use.

SUGGESTIONS FOR CHART AND GRAPH DESIGN

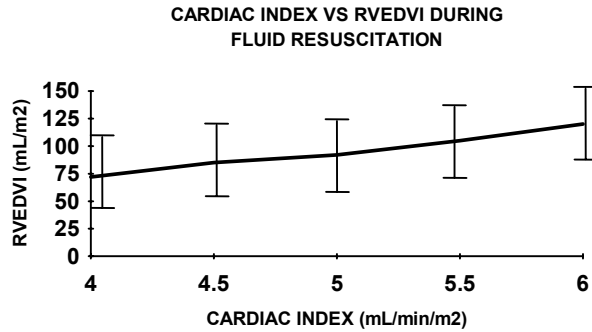
1. **Charts and graphs should be simple.** In the past, charts and graphs were simple by necessity. They were often drawn by hand, requiring hours of work. With the widespread use of computers, creating elaborate and colorful charts is now quick and easy. Complicated, colorful, and intricate charts and graphs are not necessarily better, however, and may only serve to confuse or mislead the reader. Charts should therefore be kept as simple as possible.
2. **Maximize the data in each chart.** One of the real advantages of using visual representations of data is that a maximum of information can be conveyed in a minimal amount of space. It is possible, however, to include too much data and create a graph that overwhelms the reader by trying to answer several questions at once. Charts and graphs should be designed to be easily understood at first glance. At the same time, using multiple graphs to present data that could just as easily be presented in a single graph is an inefficient use of space.
3. **Label rows and columns concisely and accurately.** Numbers by themselves are meaningless. Accurate labeling of the data is essential if the reader is to interpret the data correctly.
4. **Use legends to explain the data being illustrated.** Legends should be used to define which study groups are being illustrated and in what manner.
5. **Use an appropriate scale.** The type of data variable and the statistical methods used to analyze the data will frequently determine the way in which data is displayed. Data which is binary or dichotomous will most likely require a **nominal scale** (i.e., disease vs no disease). When the data is ordered in categories, such as in tumor staging, an **ordinal scale** consisting of several categories is appropriate. The majority of data encountered in medical research is continuous and requires a **numerical scale**.

Manipulating the scale can make any data look better, even if there is no significant difference to demonstrate. At the same time, using an inappropriate scale may minimize the appearance of a significant difference or may hide clinically useful variations in the data. Thus, the scale utilized can profoundly affect the way in which the data is interpreted.

One common mistake in creating charts and graphs is changing the scale in mid-axis. This may distort the graphical representation of the data and can be misleading to the reader. Another common error is beginning the scale at a point other than zero. This is known as **suppression of zero error** and can also result in distortion of the data.

6. **Arrange data logically.** As with tables, the reader will expect to see data presented in a logical order. If the data has a natural order, rearranging that order may result in a confusing graph. Although it might seem natural and helpful, displaying data alphabetically or numerically from smallest to largest may thus upset the natural order of the data and make it less clear.
7. **Use the same baseline for multiple graphs of the same type.** If multiple graphs are being used to display data, use of a common baseline will allow the graphs to be compared much more easily and will make them less confusing. It can also allow multiple graphs to be condensed into one graph thus saving space.
8. **Use significant digits appropriately.** As described above for tables, only use as many significant digits as are appropriate.
9. **Avoid using “double Y-axis” graphs.** These graphs are commonly used to display two or more groups of data which are measured on different scales. Unfortunately, they can easily be misunderstood. It may not be immediately obvious, for example, which Y axis belongs to which data set, especially if multiple data sets are being presented. In such a situation, it may be better to display the data in two separate graphs, each with a different scale on the Y axis.
10. **Avoid logarithmic scales** (or other transformed scales) unless the data analysis required transformation (i.e., the data did not assume a normal distribution). Logarithmic scales can similarly distort the data and make it confusing.

11. **Error bars should be used to convey variability.** Most charts and graphs provide information about the mean of the data. We are also interested, however, in the variability associated with the data mean. Error bars, as illustrated in the graph below, provide a way of displaying this variability in the data.



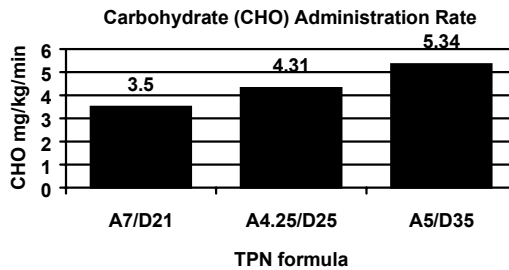
COMMON USES OF VARIOUS CHARTS AND GRAPHS

Various examples of tables, charts, and graphs have been illustrated in previous chapters as listed below:

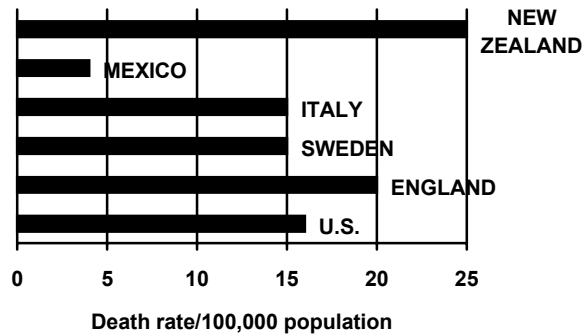
- frequency histogram Chapter 2
- data tables Chapters 3, 4, 5, 6, 9, 11
- contingency tables Chapters 4, 5
- ROC curves Chapter 4
- X-Y scatter graphs Chapter 8

The following are examples of other types of charts and graphs which are commonly seen in the medical literature along with their uses.

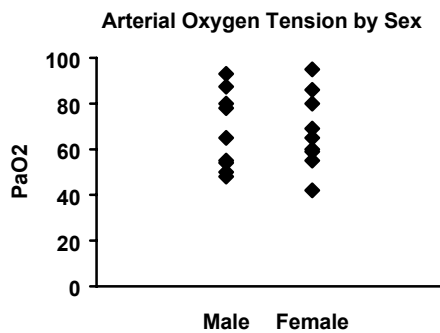
- Bar charts
- used to display nominal or ordinal data



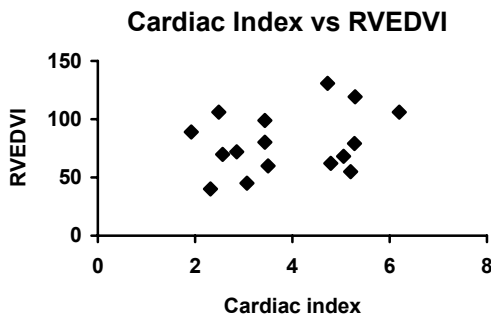
- Histograms
 - used to present information about ordinal data
 - the measurement of interest is graphed on the X axis and the number or percentage of observations on the Y axis



- Dot plots
 - used to display binary or nominal data



- Scattergrams or Scatterplots
 - also known as bivariate plots
 - the risk factor is placed on the X axis and the characteristic or outcome to be explained is placed on the Y axis
 - different subgroups can be represented by different symbols maximizing information exchange without an increase in space

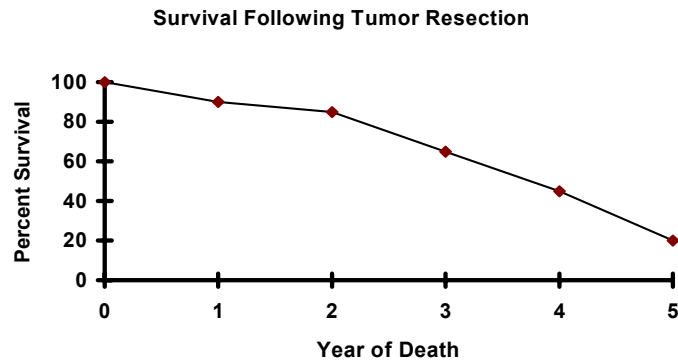


• **Survival Curves**

- used to illustrate survival rates at various periods of time.
- Kaplan-Meier survival curves plot the proportion of patients still surviving each time a patient dies.



- Actuarial survival curves are curvilinear as they plot the proportion of patients still surviving at specific time intervals since the study began, averaging the occurrence of death during the previous time period.



SUGGESTED READING

1. Wainer H. Understanding graphs and tables. Educ Research 1992;21:14-23.
2. Wainer H. How to display data badly. Am Stat 1984;38:137-147.
3. Altman DG. Statistics and ethics in medical research: VI - Presentation of results. Brit Med J 1980;281:1542-1544.
4. O'Brien PC, Shampo MA. Statistics for clinicians. 2. Graphic displays - histograms, frequency polygons, and cumulative distribution polygons. Mayo Clin Proc 1981;56:126-128.
5. O'Brien PC, Shampo MA. Statistics for clinicians. 3. Graphic displays - Scatter diagrams. Mayo Clin Proc 1981;56:196-197.