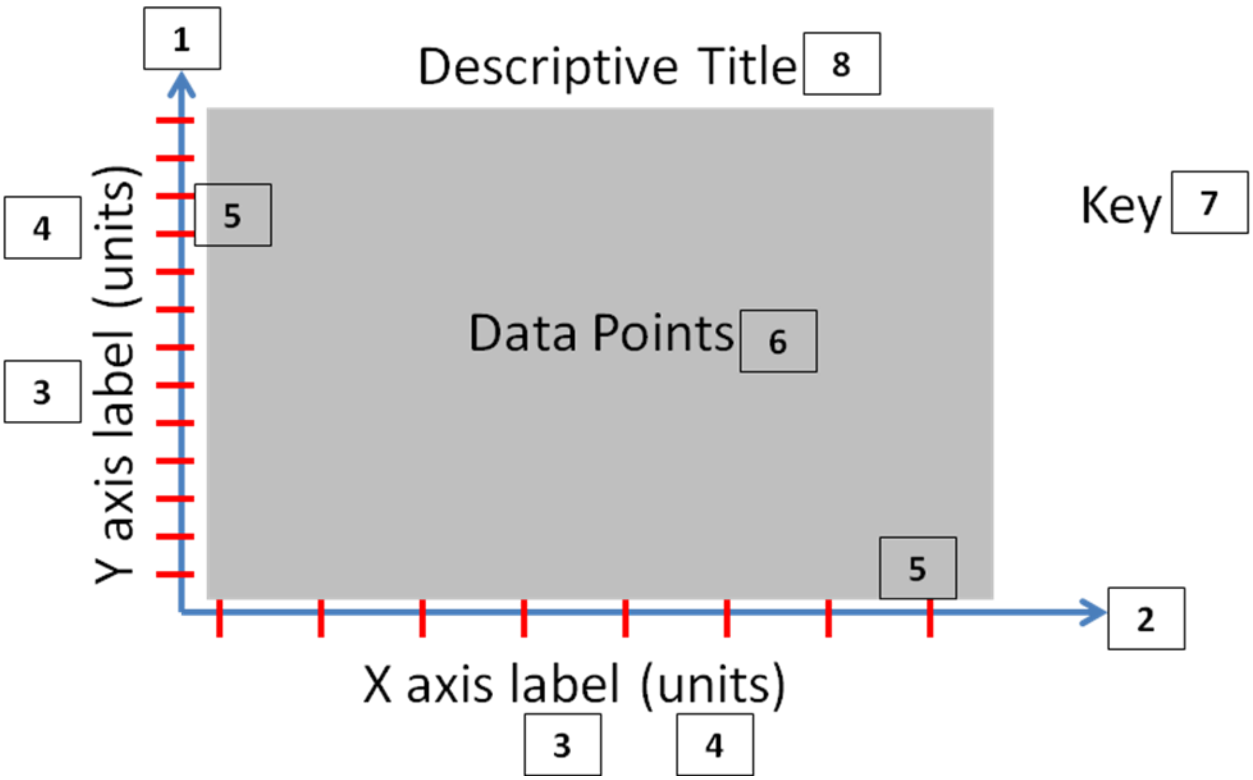
A Guide to Graphs and Tables

Purpose of Graphs

Graphic presentation of data is an important aspect of communicating ideas in science. Whatever the format of presentation, (slide-show, poster, abstract, or manuscript) your audience often makes critical judgments based on the data you present them with. As such, learning to effectively present the data you have collected will help you tremendously in the expression of your own ideas. This handout is designed to reacquaint you with scientific graphing. Before we look at some examples, let us review the basic parts of a graph since they play an important role in communication.

Parts of a Graph



1-Y axis- The dependent variable, also known as the response variable is the variable that is measured in the experiment. It is displayed on the vertical axis of the graph in two- dimensional space.
2-X axis- The independent variable, or the variable that is manipulated by the researcher is displayed on the horizontal axis of the graph in two- dimensional space.
3-Axes labels- Labels communicate the variables present on the graph.
4-Units- Units next to axes labels communicate the property of measurement. Units are not needed for categorical independent variables.
5-Tick marks- If both the x and y variables are continuous, then the tick marks create increments from a lower number to a higher number along both axes. However if the x variable is categorical, then the tick marks reference the categories as they are defined by experimental parameters.
6-Data Points- Data points are obtained directly from the experiment and are depicted in two- dimensional space based on the value of the independent variable. Also, data points may be displayed as summaries; for instance: averages accompanied with error bars, percentages, and changes.
7-Key- Helps to identify and label symbolic data in a simplified manner which reduces clutter from direct labeling on the graph.
8-Descriptive Title- Provides a concise statement that summarizes the relationship between the independent and dependent variables in the experimental study.

Data
When assembling graphs for presentation it is often unnecessary to include all accessible data points. Presenting all material collected typically obscures the major point of your graph. To avoid this, before you assemble the outline of your presentation carefully establish the major points you'd like to make. For example, you have measured the resting heart rate in beats per minutes (bpm) of everyone in your lab group. You calculate the mean value to be 71.32 bpm. How should you report it? 71 bpm?, 71.3 bpm? or 71.32 bpm? You need to consider what you are measuring, what your purpose is (test of your hypothesis), and the sensitivity of your instruments. In this case, reporting a fraction of a heartbeat in the average doesn't make sense because you don't have a third of a heart beat! If multiple trials were performed in a given condition, the average of the trials with the standard deviation or standard error should be plotted for each condition.

Examples of Graphs
Below are examples of the most common graphical displays of data for your reference. Each of the computers in the lab has the PowerLab data acquisition and analysis software. Within 'Chart' you should be able to get all the measurements that you are interested in. I would suggest using Microsoft Excel or OriginPro on the laptops to generate figures for data other than example traces of recordings. Appropriate statistics can also be performed within Excel and OriginPro. Please consult with me regarding the appropriate statistical tests to perform on your data and how to display those results on your figures.

Raw data plots

Histograms
An extremely useful type of graph that can be used either for one of your data displays to report with or to get an idea of the data distribution to inform further analysis decisions is the histogram. The measured values are plotted on the x-axis and the number of times that value is observed is plotted on the y-axis. Often the x-axis values are organized into ranges called bins. Below are histograms of nerve conduction velocities in different nerves (Figure 1)

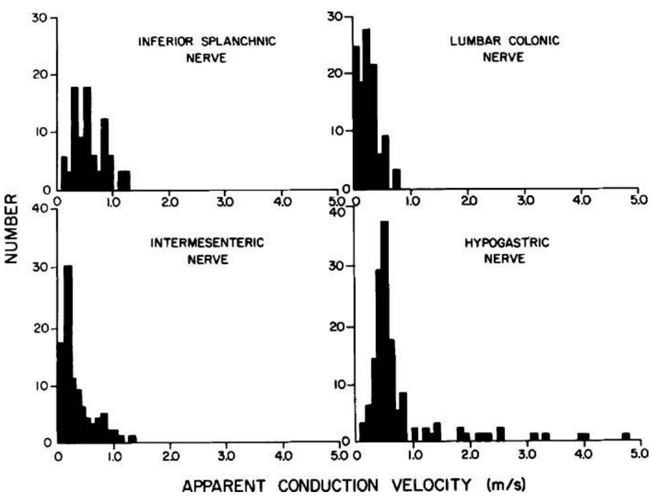


Figure 1 - Example of a histogram

Figure 11. Frequency histogram of conduction velocities of preganglionic fibers in different trunks attached to nerve inferior mesenteric ganglion of guinea pig.
(Source: Data from J. H. Szurszewski and P. J. Crowcroft.)

Scatter plots
Scatter plots are a nice way to illustrate how two variables are related to one another. The individual, raw data are plotted for the two values instead of averages. In the example below, the levels of sodium in the blood serum (liquid portion) are plotted as a function of the change in body weight following an endurance race. Also included are clinically-related classifications for sodium levels and the hydration status of the athletes. You can see that there is a relationship between the two variables.

Figure 2. Example of a scatter plot. (From: Rosner and Kriven, 2007)

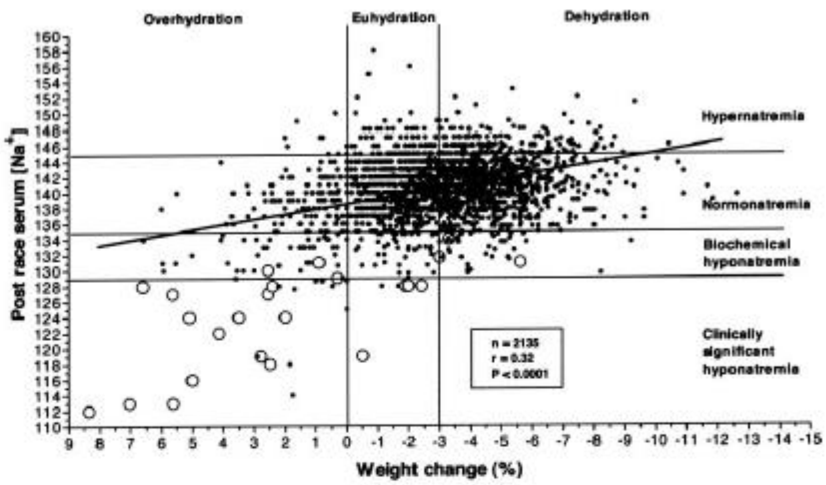


Figure 2. Relationship between serum sodium after racing and the weight change (in %) during exercise in 2135 athletes who competed in endurance events. ●, asymptomatic athletes; ○, athletes with symptoms compatible with EAH encephalopathy (EAHE). The majority of athletes who develop clinically significant hyponatremia have positive weight changes. Reprinted from reference (34), with permission. Copyright 2005 National Academy of Sciences.

Categorical data sets

Bar graphs

Bar graphs are useful in comparing (or contrasting) a series of values collected under different circumstances that cannot be conveniently arranged in a quantitative manner. They are also useful in comparing a series of values collected at the same time. For example, the data shown in Figure 3 below are from an experiment evaluating the time it took for subjects to exert themselves to exhaustion (y-axis, **dependent variable**) depending on prior activity or recovery activities (x-axis categories, **independent variables**). Note that the x and y axes are clearly labeled, each group is easily distinguishable, and the standard error is shown as a line on each bar. Significant differences are indicated with asterisks. Also note the figure legend under the figure describing in text what is plotted with some more information to aid in its interpretation.

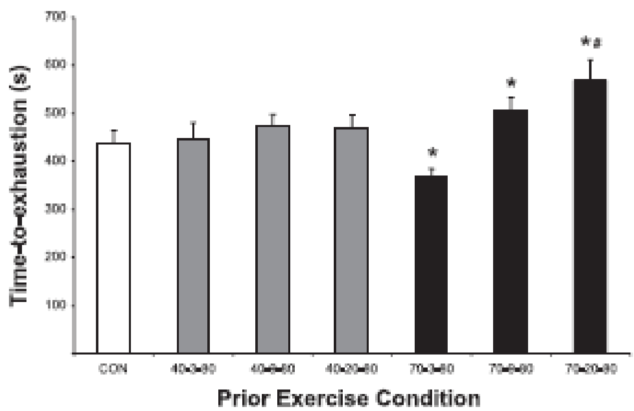


Fig. 6. Mean ± SE times to exhaustion during severe-intensity exercise in control and after the various prior exercise and recovery permutations. Relative to control, the time to exhaustion was increased by 2% in the 40-3-80 condition [nonsignificant difference (NSD)], by 8% in the 40-9-80 condition (NSD), by 7% in the 40-20-80 condition (NSD), by 15% in the 70-9-80 ($P < 0.05$), and by 30% in the 70-20-80 condition ($P < 0.05$). The time to exhaustion in the 70-20-80 condition was significantly greater than in the 70-9-80 condition ($P < 0.05$). Relative to control, the time-to-exhaustion was 16% lower in the 70-3-80 condition ($P < 0.05$)

Figure 3 - Example of a bar graph (From Bailey et al., 2009)

Dot plots

Dot plots are a great way to show all of the data in a data set so the viewer can see the distribution easily. Similar to the histogram, the data points falls into discrete bins. The independent variable is categorical and the dependent variable is continuous. The mean is often denoted with a bar. Figure 4 is an example of a dot plot.

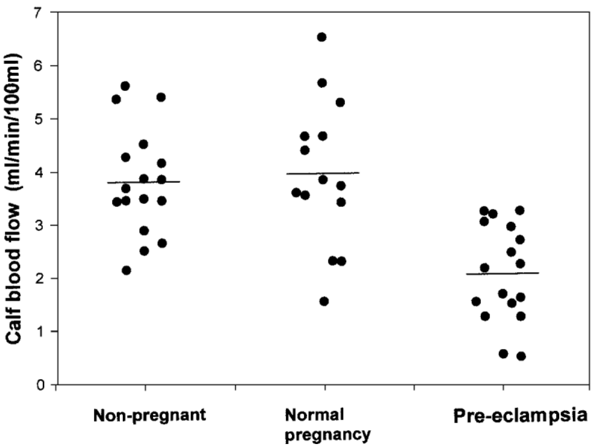


Figure 4 - Example of a dot plot (From N. ANIM-NYAME et al., 2000)

Figure 1 Dot plot comparing calf nutritional blood flow in non-pregnant controls, normally pregnant controls and women with pre-eclampsia

Box plots

Box plots are a statistical plot that allows you to provide very detailed descriptive statistics about a data set in a concise visual. The median of the data set (line in the middle of the box) is determined and then the data set is divided into four, equal parts (quartiles) along the range of the data values. The number of data points that fall into each of those quartiles is represented by the size of the box (for the two middle quartiles around the median) or whisker for the outer quartiles. The mean is often denoted with an asterisk or other symbol.

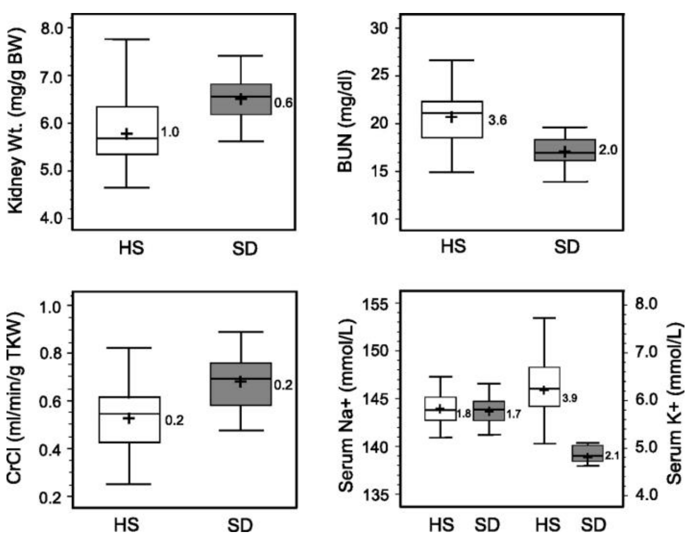


Figure 5. Example of a box plot (From: http://ajprenal.physiology.org/cgi/content-nw/full/298/6/F1484/F2)

Fig. 2. Box plot comparing physical traits and renal parameters of HS and Sprague-Dawley (SD) rats. Open plots are HS data, and shaded plots are SD data. The box itself contains the middle 50% of the data. The upper quartile (UQ) indicates the 75th percentile of the data set, and the lower quartile (LQ) represents the 25th percentile. The interquartile range (IQR = UQ – LQ) represents a measure of variability and is displayed on *right* of each box plot. The line within the box indicates the median value of the data; + represents the mean value of the population. BW, body weight; BUN, blood urea nitrogen; CrCl, creatinine clearance.

Continuous data sets

Line Graphs

Line graphs are most often used to express how one measured variable varies continuously with another variable. In general one variable such saturation of hemoglobin with O_2 will depend on another variable such as the amount of O_2 available (P_{O2} , the Figure 6). The hemoglobin saturation in this example is the **dependent variable** and P_{O2} is the **independent variable**. Convention favors plotting the independent variable along the x-axis or abscissa, and the dependent variable along the y-axis, or ordinate.

To distinguish between the different groups to be plotted, it is recommended that obviously different symbols be used for each group or that they be clearly labeled if space permits you to do so clearly. Below are two examples of line graphs. The figure of left shows the hemoglobin saturation vs. P_{O2} for in different pH (top) and in different birds (bottom). The figure on the right is a plot of several dependent variables measured during a dive by an emperor penguin as a function of time. Note how the different groups and measure are plotted in a way that makes them easily distinguishable from one another with different colors and/or symbols).

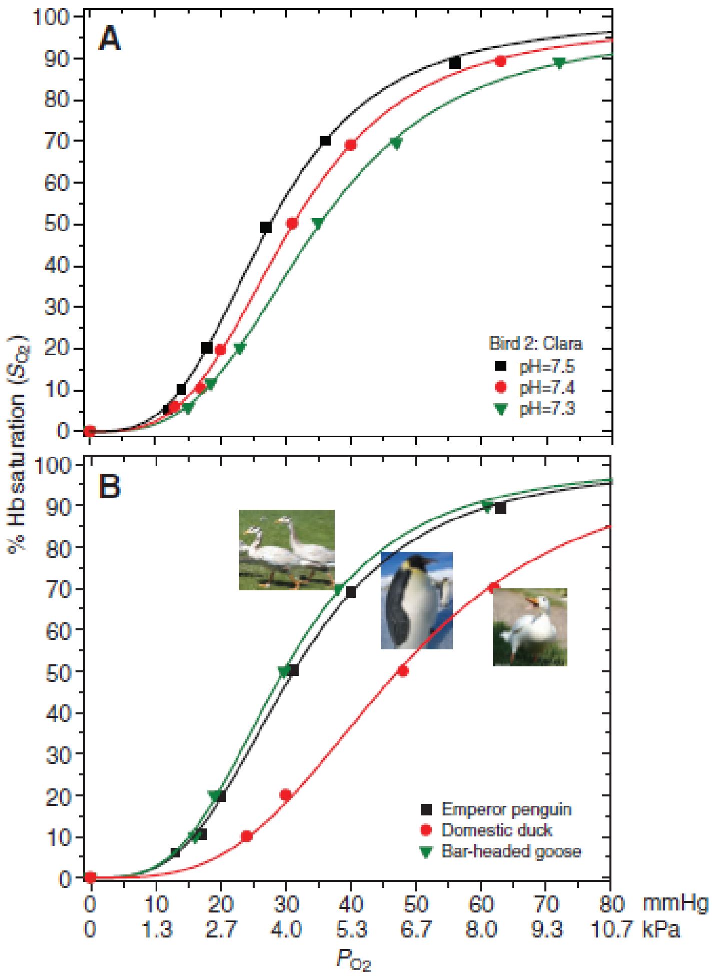


Fig. 1. An example of oxygen–hemoglobin (O_2–Hb) dissociation curves from (A) one penguin at pH 7.5, 7.4 and 7.3, and (B) the emperor penguin, the bar-headed goose (*Anser indicus*) (Black and Tenney, 1980) and the domestic duck (*Anas platyrhynchos*, forma domestica) (Hudson and Jones, 1986) at pH 7.4. Note that as for the bar-headed goose, the O_2–Hb dissociation curve of the emperor penguin is significantly left-shifted as compared with the domestic duck (and most birds). The bar-headed goose photo is courtesy of Graham Scott; the domestic duck photo is by Maren Winter (licensed under the terms of the GNU Free Documentation License, Version 1.2 or any later version); the penguin photo is by J.M.

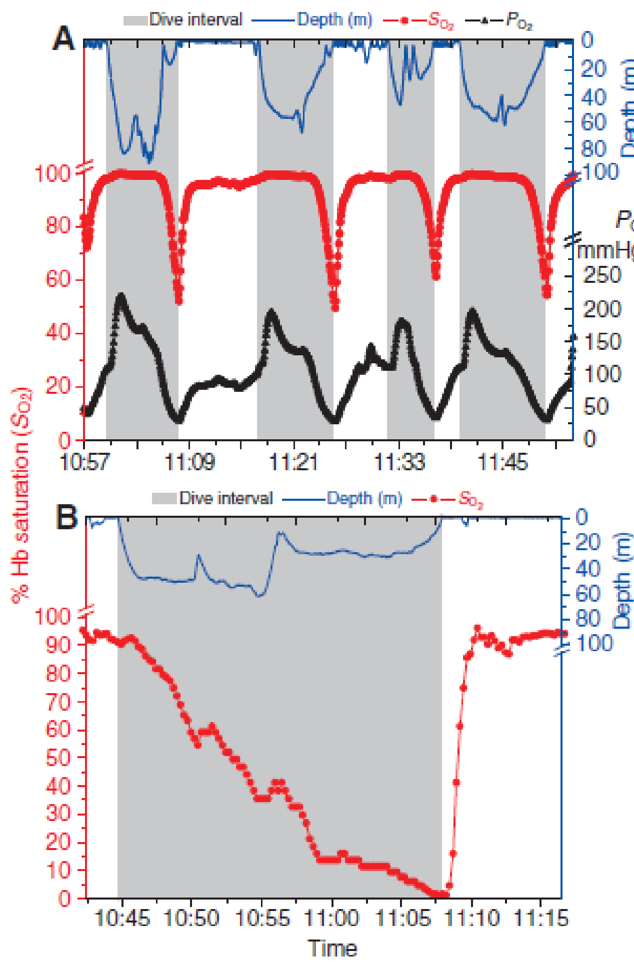


Fig. 2. The S_{O2} profile during (A) 1 h of diving of bird 1 (2008) (arterial S_{O2}) with $P_{a,O2}$ superimposed. Note that $P_{a,O2}/S_{a,O2}$ are at low levels at the start of the measurements because a dive had occurred just before the series recorded, and (B) the current record dive (23.1 min) of an emperor penguin [bird 19 (Ponganis et al., 2007), venous S_{O2}]. Note the arterialization of the venous blood O_2 store in this dive, as $S_{v,O2}$ before the dive is as high as 95% and the initial $S_{v,O2}$ of the dive is 91%. $S_{v,O2}$ decreased to 1% by the end of this long dive. S_{O2} was determined at pH 7.5 throughout the entire dive to maintain consistency and to provide a conservative estimate of continuous S_{O2}.

Figure 6. Examples of line graphs. (From: Meir and Ponganis, 2009).

Other data displays (not an exhaustive list)

<u>Pie Charts</u>

Pie charts are effective at quickly communicating relative frequencies or percentages. The relative area of the pie is divided into pieces, which represent the proportions of the whole. When constructing a pie chart, it is important to construct the pie in two-dimensions, as adding the third-dimension distorts the area and causes miscommunication. It is also important to label the wedges with names, percentages, and to differentiate each wedge with a distinct color. Unlike other types of graphs, the pie chart does not use an L-shaped framework; instead, it is primarily based on the polar coordinate system, which limits its utility in biology. Thus, the pie chart cannot communicate distribution in data, cannot illustrate uncertainty via error bars or standard deviation, and cannot be used for displaying nested data.

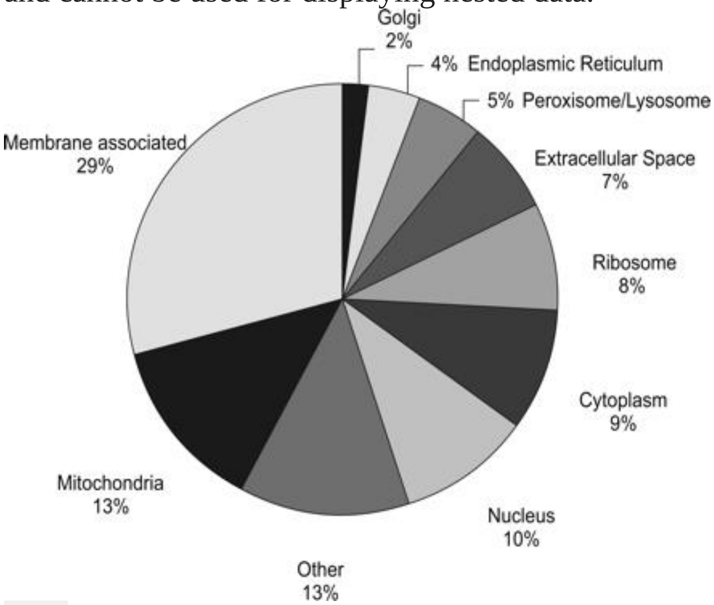


Fig. 2. Pie chart illustrating the global functional distribution of genes in rat LV myocardium whose expression was significantly altered by diabetes. This graph is based on GO Cellular Component classification (**1**). The total number of genes with a GO Cellular Component Classification term and at least one other associated GO term were used in this analysis, yielding a total of n = 394 out of an overall total of n = 1,614 significantly modified genes. Genes were sorted according to their GO classification and the total in each category then divided by the total number of genes analyzed (n = 394) to generate respective percentages, as shown. LV, left ventricle; GO, gene ontology.

Figure 7. Example of a pie chart (From Jones et al., 2007)

<u>Tables</u>

Tables are used to describe large amounts of data that would ordinarily be too burdensome to graph. The format of a table allows many variables which may have distinctly different units of measure to be assembled for quick reading. However tables can be challenging to sift through and the take home message is not immediately clear. The table below is an example of a table summarizing heart rate, blood pressure, and exercise duration and intensity in men and women. Note the descriptive title at the top and the caption below providing some more information to assist in evaluating the data in the table.

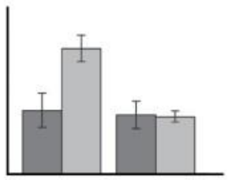
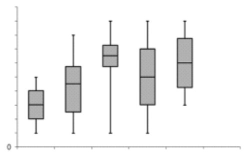
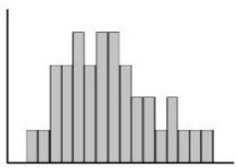
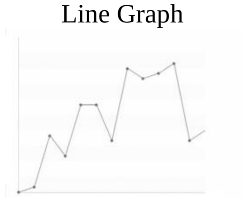
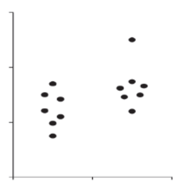
Table 4. Treadmill electrocardiographic exercise stress test characteristics of the men and women of the study sample.

Variables	Men		Women	
	N	Mean ± SD	N	Mean ± SD
Baseline heart rate (bpm)	200	77.9 ± 0.9	241	82.9 ± 0.8*
Peak exercise heart rate (bpm)	200	171.5 ± 1	241	170.1 ± 0.9
Heart rate recovery 1st min (bpm)	155	143.5 ± 1.1	207	139.2 ± 1.3*
Heart rate recovery 2nd min (bpm)	180	123.2 ± 1.1	216	117.2 ± 1.2*
Baseline systolic blood pressure (mmHg)	200	124.5 ± 0.7	241	119.5 ± 0.7*
Peak exercise systolic blood pressure (mmHg)	200	181.0 ± 1.6	241	166.8 ± 1.7*
Baseline diastolic blood pressure (mmHg)	200	78.7 ± 0.6	241	77.1 ± 0.6
Peak exercise diastolic blood pressure (mmHg)	200	84.2 ± 0.9	241	81.5 ± 0.8*
Rate-pressure product (bpm x mmHg)	200	31023.6 ± 321.7	241	28378.7 ± 313.1*
Exercise duration (min)	200	8.4 ± 0.1	241	6.9 ± 0.1*
Exercise capacity (MET)	188	12.7 ± 0.2	228	10.4 ± 0.2*

MET = metabolic equivalent. *P < 0.05 compared to men (Student *t*-test).

Figure 8. Example of a table (From Chalela et al., 2009)

Guide to Data Displays*

Data Display Type	Usage	Advantages	Disadvantages
Bar Graph 	To compare categorical data, percentages, or summary statistics from multiple groups. [42] Each bar represents a category; shape can be changed by moving the categories around.[25]	Useful for understanding distributions from large datasets. [42]. Stacked bars or shading of bars can be used to distinguish the different levels within the data.[42]	Obscures the distribution of data[42, 49], number of data points, and their values.[17, 18]
Box and Whisker Plot 	To show distribution of data from one or multiple groups[18, 42].	Shows and compares distributions of large datasets.[18, 42]	Should not be used for small datasets.[42] Does not show individual data(except for outliers).[42]
Histogram 	To show a distribution of data with the independent variable as continuous.[25] Uses numerical data instead of categorical data [25].	Shows the shape of the distribution of data with a continuous variable.[25]	Must choose the bin size wisely to avoid influencing the shape being too compressed or too dispersed.[25]
Line Graph 	To show how a single variable or multiple variables changes over time or to show how a variable deviates from a set baseline[20] X axis portrays categories while the Y axis portrays quantitative values.[20]	Shows direct relationships and may be used to predict relationships between continuous variables. [34]	Not appropriate for representing ranked, part-to-whole, or correlational data. [20]
Dot Plot 	To show distribution of small data sets from multiple groups .[17, 49] The independent variable is categorical and the dependent variable is continuous.	Shows all data from multiple categories and the distribution within each category.[17, 49]	Not appropriate for representing a large data set because the plot will become cluttered and it will be difficult to see the individual points.[17]
Scatterplot 	To show individual data points from bivariate data.[42]	Shows the relationship between variables. [18, 42] Shows trends in the data and any noticeable outliers.	It may be difficult to extract individual data points if they fall on the same or nearby coordinates.[18, 42]
Tables 	To display simple relationships between numerical values and categorical groups, so that individual values can be easily extracted from the rows and columns[20]. Often used for small data sets. [19, 47]	Since values in a table are encoded as text, it is easy to extract individual values. [19, 20] Numbers in a table can be displayed with decimal precision.[20, 47] A table can also communicate multiple sets of data with different units. [20]	Tables may make it difficult to interpret the take home message if not organized properly.[19]

Name	Thread pitch (mm)	Minor diameter tolerance	Nominal diameter (mm)	Head shape	Price for 50 screws	Available at factory outlet?	Number in stock	Flat or Phillips head?
M4	0.7	4g	4	Pan	$10.08	Yes	276	Flat
M5	0.8	4g	5	Round	$13.89	Yes	183	Both
M6	1	5g	6	Button	$10.42	Yes	1043	Flat
M8	1.25	5g	8	Pan	$11.98	No	298	Phillips
M10	1.5	6g	10	Round	$16.74	Yes	488	Phillips
M12	1.75	7g	12	Pan	$18.26	No	998	Flat
M14	2	7g	14	Round	$21.19	No	235	Phillips
M16	2	8g	16	Button	$23.57	Yes	292	Both
M18	2.1	8g	18	Button	$25.87	No	664	Both
M20	2.4	8g	20	Pan	$29.09	Yes	486	Both
M24	2.55	9g	24	Round	$33.01	Yes	982	Phillips
M28	2.7	10g	28	Button	$35.66	No	1067	Phillips
M36	3.2	12g	36	Pan	$41.32	No	434	Both
M50	4.5	15g	50	Pan	$44.72	No	740	Flat

* From: Angra A and Gardner SM (2016). Development of a Framework for Graph Choice and Construction. Advances in Physiology Education 40: 123–128. doi:10.1152/advan.00152.2015