MAKING EFFECTIVE GRAPHS IN THE SOCIAL SCIENCES

FUNDAMENTAL PRINCIPLES AND PROCESSES

ERIC JOHANN HIRIS

Making Effective Graphs in the Social Sciences

Making Effective Graphs in the Social Sciences provides the knowledge and skills for creating graphs that are easy to interpret accurately.

This includes: (a) knowledge of the different types of graphs and under what circumstances each graph is appropriate, (b) knowledge of what decisions to make when choosing graph components, such as the type of axis or data symbols, and what evidence supports those decisions, and (c) how to use consistency within and across graphs to make your graphs easier to understand. In addition to developing this knowledge base, practical skills are developed for creating effective graphs in Microsoft Excel, IBM SPSS Statistics software ("SPSS"), and R. For Microsoft Excel and SPSS, this includes illustrated and annotated step-by-step instructions. Electronic resources, including full Excel and SPSS appendices and downloadable datasets hosted on the Routledge product page, support the worked examples in the book. Social science researchers and students in data-based social science courses will benefit from the focus on both knowledge and practical skills.

Instructors will find the book self-contained – allowing students to make more effective graphs with minimal instructor intervention.

Eric Johann Hiris is a researcher in the Department of Psychology at the University of Ostrava, Czech Republic. He studies visual perception and cognitive psychology and has published numerous peer-reviewed journal articles on these topics. He received his BA from Oakland University and his MA and PhD from Vanderbilt University. He was previously an associate professor at St. Mary's College of Maryland, a professor at the University of Northern Iowa, and a professor at the University of Wisconsin – La Crosse.

Hiris' *Making Effective Graphs in the Social Sciences* provides a guide for both students and professionals. Hiris provides guiding principles using examples and humor with the added benefit of clear practical instruction in making graphs across multiple platforms.

Professor Aileen M. Bailey, *St. Mary's College of Maryland, United States*

Eric Hiris, an accomplished cognitive psychologist, has produced an excellent book on *Making Effective Graphs in the Social Sciences*. Full of ideas and resources for creating informative, attractive graphs, the book reifies the maxim that a well-composed picture can be worth a thousand words.

Professor Emeritus Randolph Blake, *Vanderbilt University*

Clear guiding principles and developed examples illustrate how smart graphing choices improve reader experience. Hiris gives instructors, researchers, and students a solid, usable resource that covers the why and how of effective graphs, and thus, directly contributes to science and data literacy.

Dr. Kimberly Epting, *Department of Psychology, Elon University*

This book is an invaluable guide for students and researchers seeking clarity in data visualization. By bridging SPSS, Excel, and R, it provides practical strategies to transform data into meaningful graphics, making complex results easy to understand and communicate across audiences.

Professor Alessandro Quartiroli, *University of Wisconsin - La Crosse*

Making Effective Graphs in the Social Sciences

Fundamental Principles and Processes

Eric Johann Hiris

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Designed cover image: Getty Images

First published 2026
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

ISBN: 978-1-032-74854-2 (hbk)
ISBN: 978-1-032-73803-1 (pbk)
ISBN: 978-1-003-47126-4 (ebk)

DOI: 10.4324/9781003471264

Typeset in Times New Roman
by KnowledgeWorks Global Ltd.

Access the Support Material: www.routledge.com/9781032738031

For my father who taught me the concept of doing something better than right and my mother who taught me kindness.

Contents

Acknowledgment

I am extremely grateful to Dr. Kim Epting and Dr. Dawn Norris for their feedback on early versions of this book, helpful advice, practical suggestions, constructive criticism, and editorial assistance.

I would like to express my appreciation to all the editorial staff at Routledge, and in particular, I extend a special thank-you to Adam Woods and Shraddha Saikia for their patience and their answers to my many questions.

I would also like to acknowledge the University of Wisconsin – La Crosse for providing a sabbatical leave that supported the start of the project that resulted in this book.

I would also like to extend my deepest gratitude to my wife, Dawn Norris, for her encouragement and patience throughout the process.

Finally, I greatly appreciate the reminders provided by Václav, Kafka, and Goffman that sometimes it is necessary to simply relax and pet a cat.

1 Introduction

Why is this book necessary?

Graphs are a powerful way of conveying information and an important aspect of scientific communication (Cleveland, 1984). Graphs may increase persuasiveness (Pandey et al., 2014; Tal & Wansink, 2014**)** and/or understanding (Dragicevic & Jansen, 2018) of scientific and other types of information. However, the social sciences, compared to other sciences, lag in the use of graphs in their published work (Cleveland, 1984). Looking specifically at subdisciplines within psychology, the use of graphs increases as the rated difficulty of the subdiscipline increases (Best et al., 2001; Smith et al., 2002). As Smith et al. (2002) note, this difference in graph use across subdisciplines is not due to differences in quantification: those subdisciplines that use fewer graphs tend to report more tables and inferential statistics. It is not entirely clear why there is a subdisciplinary difference in the use of graphs versus tables and statistics.

This subdisciplinary difference cannot explain the overall lag in graph usage in social sciences behind other sciences discussed by Cleveland (1984). Beniger and Robyn (1978) noted that in the 17th and 18th centuries, the social sciences actively avoided the use of graphs, instead preferring tabular representation of information. However, that began to change for the social sciences in the 19th century and much time has passed since then. It seems unlikely that the current difference in the social science use of graphs can be entirely attributed to events in the 17th and 18th centuries.

I think there is another factor involved: a lack of specific training in making graphs. In my wildly unscientific approach to this question, I have asked many academics in psychology and other social sciences whether they recall ever receiving specific training on how to make a graph. So far, in several decades of asking, no one has ever reported receiving specific training. Certainly, graphs were made as part of a course – but what was lacking was instruction on the principles of what makes a good graph and why particular decisions might be made regarding which graph to use, and so forth. If you do not receive training on how to make graphs, you are not likely to be comfortable making graphs, and therefore will be less likely to make a graph when the opportunity arises. To address this, I think two things are necessary. First, individuals need to be trained on the principles of making good graphs (knowledge). Second, individuals need to know how to actually make that good graph (practical skills). With both the knowledge and the practical skills, I believe social scientists will be more likely to present their data graphically. The goal of this book is to help with both the knowledge and the practical skills.

But some might wonder if this knowledge and practical skills training are really necessary. Is it not possible to just use graphing software to make a good graph? Unfortunately, one cannot expect to use the default settings of graphing software and get good results (e.g., see Su, 2008).

DOI: 10.4324/9781003471264-1

The remedy for this is human knowledge and practical skills of how to make the graph in the way it needs to be for other humans to understand it easily and accurately.

Who is this book for?

This book is written for social science students in a methods, statistics, data visualization, or lab course. However, I think the book will be useful for anyone looking to increase their understanding of graphing and improve their practical skills. If you are not a student, you have the benefit of not needing to worry about being assigned the problems at the end of the chapters – you can still choose to complete them if you like. Some problems even have humorous answers. As I was contemplating the answer to the question of who this book is intended for, I was surprised to realize this book is for me, too. I expect it will be a useful resource for me when I forget the details of how to accomplish a particular graphing task.

What this book will and will not do

This book will:

- Develop your understanding of what makes a graph difficult to interpret accurately.
- Develop your understanding of what makes a graph easier to interpret accurately.
- Provide fundamental knowledge for making choices about which type of graph to use.
- Provide fundamental knowledge for making choices about various graph components.
- Make graph recommendations based on research findings.
- Make graph suggestions based on my preferences and intuition.
- Provide you with practical skills in actually creating graphs in Microsoft® Excel, IBM® SPSS® Statistics software ("SPSS"), or R.

This book will **not**:

- Teach you how to use Microsoft Excel, SPSS, or R. However, you do not need to already know how to use a particular software package because the book gives you step-by-step instructions. You will certainly learn some by doing, but the focus is not on teaching the software.
- Teach you statistics. There is a review (if you know statistics) or an introduction (if you do not know statistics). The goal of this book is to make readers become familiar with the basic statistical ideas that may be relevant to graphing.
- Teach you complex data visualizations. This book focuses purposefully on the fundamentals of graphing in relatively simple graphs. I believe these fundamentals will serve you well if you later learn more complicated visualizations.

How to prepare yourself for this book?

A picture is worth a thousand words.[1] Excellent! But do you want those words to be a disorganized jumble of sentences with no focus, or do you want those words to be clear, concise, and focused? Furthermore, I argue that a graph is worth more words than a picture. A graph has a defined purpose to communicate information. Surely it is worth more than the thousand words of a picture – let's estimate two thousand words. Each page of single-spaced text is approximately 500 words, so a graph can contain the information of about four pages of single-spaced text. It is therefore even more important to ensure the graph is as clear as possible.

Expect that making a good graph will take time and effort, just like writing four good pages of text. There are graphs in this book that took me a day to make – because I was learning something new and trying to determine the exact best way to accomplish what I wanted to do. That same graph is set up for you to make in a minute or even less as you work through the book. Making a graph in a minute should not be your expectation when you graph your own data with no certainty about the best approach. Expect to take some time planning and needing to work through multiple versions of the graph to get the right one. Much like writing, which takes revision, expect a good graph to require experimentation and many iterations (Zacks & Franconeri, 2020). Also, you may not even be the best judge of whether your graph is good, because you already know the intended message of the graph (Zacks & Franconeri, 2020). Consider asking someone you trust to give you feedback on the graph. Specifically, this should be a person you trust to give you the bad news that your graph is not good (if that is what needs to be said). It is even better if that person also fits the profile of the intended audience for the graph.

Approach of the book

It is useful for the reader to know a little about how I made decisions about what to include in the book. When in doubt, I show you how to accomplish a task in a graph. For example, I will show you how to include secondary *x*- and *y*-axes on your graphs and explain when and why you might want to use them. It is not the case that all your graphs should have secondary axes. However, if you do not know how to include secondary axes, there is almost no chance you will include them when they could be beneficial.

As another example, I include instructions on how to use difference-adjusted confidence intervals in Chapters 7, 8, and 9. You may not be aware of what difference-adjusted confidence intervals are as you read this first chapter, but they will be explained and used, and I will provide instructions on how to make them yourself. You certainly do not have to use difference-adjusted confidence intervals, but again, if you received no instruction on them, there would be little chance you would implement them.

The goal is not for you to make your graphs exactly like the graphs in this book. Instead, the goal is for you to have the practical skills to be able to make various types of graphs and the knowledge to be able to decide what works in your particular case.

The main chapters of the book are where knowledge related to graphing is developed. Specifically, Chapters 2 through 4 develop some basic background knowledge, introduce terminology for graphs and tables, and outline the principles of good graphing. Chapter 2 gives a statistical review. Chapter 3 discusses what a reader needs from a graph or table and introduces terminology to discuss the components of graphs and tables. Chapter 4 develops the principles for making graphs that yield an accurate understanding of the information presented.

Chapters 5 through 9 develop relatively simple graphs for specific situations. Chapter 5 focuses on describing the data. Chapter 6 focuses on representing correlations and simple linear regression. Chapter 7 focuses on representing the results of studies that compare different groups. Chapter 8 focuses on representing the results of studies where a single group is compared across different conditions. Chapter 9 focuses on representing the results of studies that have multiple independent variables.

Chapter 10 reviews the principles of graphing, suggests some nongraph figures that could be useful for presenting your work, and offers suggestions for further reading to continue to develop the fundamental skills introduced in the book.

The appendices will help you develop your practical graphing skills. Appendices A, B, and C discuss the basics of using Microsoft Excel, SPSS, and R (respectively) in terms of how you will be asked to use them for this book. The chapter appendices provide instructions for creating specific graphs from the main chapter. Appendices 5a through 9a give Microsoft Excel instructions, Appendices 5b through 9b give SPSS instructions, and Appendices 5c through 9c give R instructions.

If you have a choice in which software package to learn, it might be worthwhile to briefly review several examples from each chapter's appendix to determine which you want to focus on learning. Although this book is not a book that teaches you a particular software package (there are other resources available for that), you will learn something about that software package in the process of using it for this book. You can, of course, always come back and learn another software package later. Consider your long-term goals when making a choice: what will be most useful to you for your career goals?

I hope you find this book enjoyable and useful.

Note

1 See https://www2.cs.uregina.ca/~hepting/projects/pictures-worth for an interesting history of this saying.

References

Beniger, J. R. & Robyn, D. L. (1978). Quantitative graphs in statistics: A brief history. *The American Statistician*, *32*, 1–11. https://doi.org/10.1080/00031305.1978.10479235

Best, L. A., Smith, L. D., & Stubbs, D. A. (2001). Graph use in psychology and other sciences. *Behavioural Processes*, *54*, 155–165. https://doi.org/10.1016/s0376-6357(01)00156-5

Cleveland, W. S. (1984). Graphs in scientific publications. *The American Statistician*, *38*, 261–269. https://doi.org/10.2307/2683400

Dragicevic, P., & Jansen, Y. (2018). Blinded with science or informed by charts? A replication study. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 781–790. https://doi.org/10.1109/TVCG.2017.2744298

Pandey, A. V., Manivannan, A., Nov, O., Satterthwaite, M., & Bertini, E. (2014). The persuasive power of data visualization. *IEEE Transactions on Visualization and Computer Graphics*, *20*, 2211–2220. https://doi.org/10.1109/TVCG.2014.2346419

Smith, L. D., Best, L. A., Stubbs, D. A., Archibald, A. B., & Roberson-Nay, R. (2002). Constructing knowledge. The role of graphs and tables in hard and soft psychology. *The American Psychologist*, *57*, 749–761. https://doi.org/10.1037/0003-066X.57.10.749

Su, Y. S. (2008). It's easy to produce chartjunk using Microsoft® Excel 2007 but hard to make good graphs. *Computational Statistics and Data Analysis*, 52, 4594–4601. https://doi.org/10.1016/j.csda.2008.03.007

Tal, A., & Wansink, B. (2014). Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*, *25*, 117–125. https://doi.org/10.1177/0963662514549688

Zacks, J. M., & Franconeri, S. L. (2020). Designing graphs for decision-makers. *Policy Insights from the Behavioral and Brain Sciences*, *7*, 52–63. https://doi.org/10.1177/2372732219893712

2 Statistical review

This chapter reviews statistical concepts that will be useful in making good graphics, including helping you make informed choices about what type of graphic to create. You may be familiar with many of the concepts, but note that in this book, you might be using the concepts in a new way. The goal here is not to teach the calculation of the statistics, but to review each statistic conceptually. You should not be concerned if you are not familiar with all of the statistical concepts or methods discussed here. Knowledge of the concepts or methods is helpful, but not critical, to understanding later chapters.

Descriptive statistics

Descriptive statistics are used to describe various aspects of a data set, including central tendency, variability, and distribution shape.

Central tendency

Measures of central tendency are used to describe the center (or "average") score of a data set. Commonly used measures of central tendency include the mean, median, and mode.

The **mean** is calculated by taking the sum of all the scores divided by the number of scores. This is sometimes referred to as the arithmetic average. In most cases, the exact value of the mean will not match any of the actual scores in your data set. However, the mean is the value that balances the weight of all the scores lower than the mean versus all those higher than it. The symbol μ will be used throughout this book for the mean.

The **median** is the middle score of an ordered list of all scores in a data set. The median is the score that divides the distribution such that half of the scores are less than and half the scores are greater than the median. The actual value (or weight) of the scores does not matter in determining the median (unlike the mean). Another way of defining the median is that it is the 50th percentile.

The **mode** is the most frequent score. If your data are given to several decimal places, it is likely that there will not be a clear mode (that is, when each score happens once in your data set, all scores are the mode).

Which measure of central tendency works best depends, in part, on how your data are distributed. The choice of measure of central tendency will be discussed in the section on distributions later in this chapter.[1]

DOI: 10.4324/9781003471264-2

Variability

Measures of variability are also important in describing a data set. The following story illustrates that point: A statistician and a bow hunter are strolling through the woods. When they come to a clearing, the statistician and bow hunter see a pot of gold coins at the exact same moment. After some discussion about splitting the gold coins evenly, they decide instead to have a contest to see which one of them gets all the gold. The bow hunter is surprised when the statistician suggests that they have an arrow shooting contest where they each shoot four arrows at a target. The bow hunter shoots four arrows first, and the bow hunter's results are shown on the left side of Figure 2.1. The statistician shoots four arrows next, and the statistician's results are shown on the right side of Figure 2.1. The statistician begins celebrating after the last arrow hits the target, and the bow hunter asks, "What are you doing?" The statistician excitedly replies, "On average I got a perfect bullseye!" Obviously, there is more to describing a data set than describing its central tendency. A measure of the variability of the data is also important for more completely understanding the data.

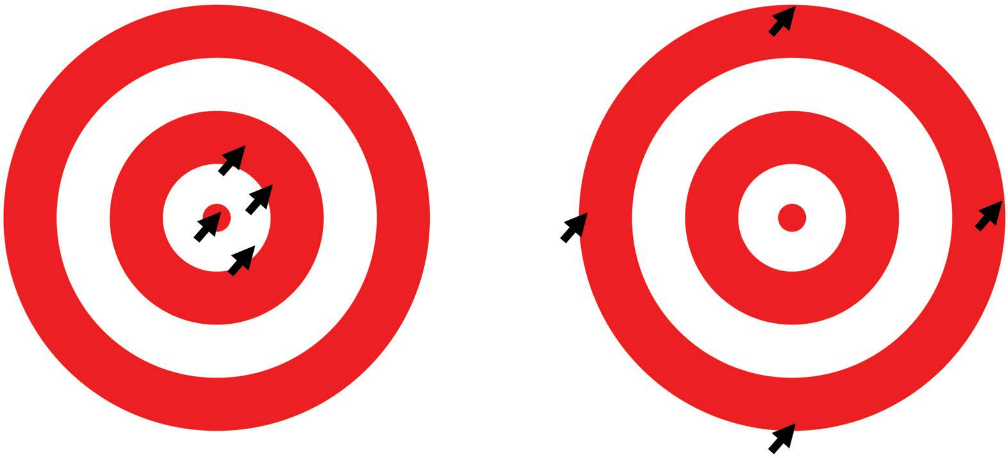


Figure 2.1 The results of the arrow shooting competition. The left side shows the bow hunter's results and the right side shows the statistician's results.

A measure of variability that is based on the mean is standard deviation. The **standard deviation** gives the usual (or average) distance of the scores in your distribution from the mean of the distribution. Standard deviation is always a positive number because it gives the usual distance (but not the direction) of a score from the mean. The symbol, σ, will be used throughout this book to indicate standard deviation.

A measure of variability that is based on the median is the interquartile range. The **interquartile range** is the difference between the value of the 75th percentile and the 25th percentile (recall that the median is the value of the 50th percentile). It can be useful here to remind yourself that the "regular" range of the data is the difference between the value of the largest score (100th percentile) and the smallest score (0th percentile). In other words, the "regular" range gives the range of all the data. The interquartile range, therefore, might be better described as the range of the middle 50% of the data.

Distributions

Although measures of central tendency can describe where a distribution is on a number line and measures of variability can describe how spread out a distribution is on a number line, they cannot completely describe the shape of a distribution unless the data are distributed normally.

A **normal distribution** is a distribution where most scores are in the middle of the distribution with fewer scores in the tails (extremes) of the distribution.[2] As our early definitions would suggest, the mean of a normal distribution describes where the distribution is on a number line, and the standard deviation indicates how spread out the distribution is. Figure 2.2 gives three examples of normal distributions.

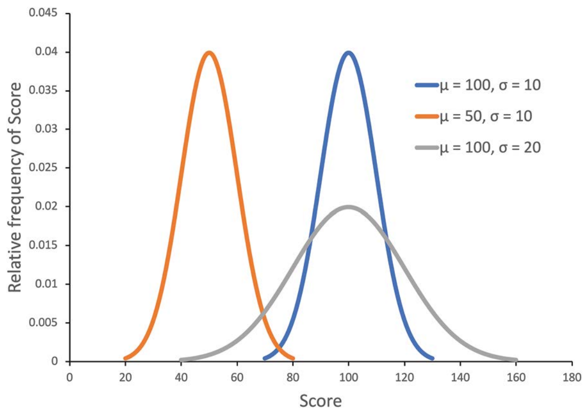


Figure 2.2 Three normal distributions.

The normal distributions given by the orange and blue lines in Figure 2.3 differ only by a difference in the mean (they have the same shape and spread, but differ in location on a number line). The normal distributions given by the blue and gray lines in Figure 2.3 differ only by the standard deviation (they have the same location (that is, central tendency) but have different shapes and spreads). Finally, the normal distributions given by the orange and gray lines in Figure 2.2 differ in both the mean and standard deviation (that is, they differ in location and shape/spread).

The mean and standard deviation describe a normal distribution. You have probably viewed graphics where the mean and standard deviation are shown. However, one difficulty is that almost no data strictly follow a normal distribution. This is unfortunate given that the normal distribution has several useful features: The mean, median, and mode all have the same value and the interquartile range is approximately $\pm 0.67\sigma$. It is important to note that all of the above is only true for normally distributed data and often data are not normally distributed.

Distributions of data can deviate from a normal distribution in many ways. We will consider extreme values/outliers, skewness, and kurtosis.

Extreme values/outliers

In the context of a distribution, an **extreme value** or **outlier** is a score that differs from the rest of the scores. There is no single accepted definition of what an outlier is. In many ways, it depends on the situation, the data, and what you know about the variable you are measuring. However, it is well known that extreme values have particular influences on descriptive statistics. For example, an outlier with an extremely large value will bias the mean to be larger (because the mean involves adding up all the scores and dividing by the number of scores, the value of each score in the data set influences the calculated mean). As illustrated by the following small data set of five scores: 4, 6, 7, 8, 10 – has a mean of 7. However, if we change one score in the data set to be an outlier, the mean drastically changes the data set: 4, 6, 7, 8, 100 – has a mean of 25. The value of the median in each of the previous data sets is 7 – the value of the median is not influenced by a single outlier. The median is therefore a more desirable measure of central tendency than the mean when the data set contains outliers.

Likewise, the standard deviation will become a less desirable measure of variability when outliers are present. In the second data set, the mean is 25 and does not represent the "center" of the data well. We should not expect the usual distance of the scores from a mean that does not represent the scores well to be particularly helpful.

In general, outliers require careful thinking about why they occurred (Could it be a data entry error? Is the outlier just random chance? Is the outlier telling us something important about the variable? Is the outlier telling us something important about the world?). The answers to these questions might indicate what approach to take. If it is a data entry error, by all means, fix the data. Otherwise, it may be best to use the median and interquartile range to describe the data, including graphically (see Chapter 5).

Skewness

Skewness is not about having an outlier (or even several outliers). **Skewness** is about the shape of the data clearly being nonnormal. **Positive skewness** means that there are more scores than expected that are large (the positive tail is longer and contains more scores than expected compared to a normal distribution). **Negative skewness** means that there are more scores than expected that are small. Figure 2.3 shows skewed and normal distributions.

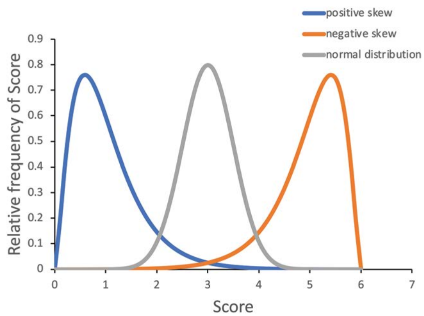


Figure 2.3 Two skewed distributions and a normal distribution.

Negative skewness will make the value of the mean smaller (and also the median to a lesser extent). Likewise, positive skewness will inflate the value of the mean (and also the median to a lesser extent). Figure 2.4 illustrates the influence of positively skewed distributions on measures of central tendency.

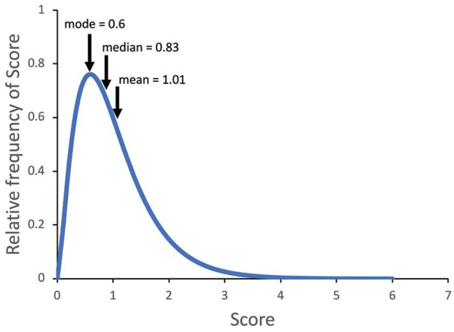


Figure 2.4 Central tendency and skew.

Because the median (and the interquartile range) are less influenced by skewness, they are better descriptors when skewness is present in the data.

Kurtosis

Kurtosis is about the spread of a distribution relative to a normal distribution. **Leptokurtic** (positive kurtosis) distributions have more values in the extreme (tails) of the distribution than a normal distribution. This will result in a higher calculated standard deviation (and interquartile range) than a normal distribution. **Platykurtic** (negative kurtosis) distributions have fewer values in the extremes (tails) of the distribution than a normal distribution. This will result in a lower calculated standard deviation (and interquartile range) than a normal distribution. Because kurtotic distributions are still symmetrical, the mean and median are not affected.

Inferential statistics

Inferential statistics are used to make inferences using a data set. Specifically, **inferential statistics** use a sample to make statements about the population the sample represents. For example, inferential statistics include determining whether a correlation between two variables exists, whether two sample means are different, and other inferences.

Useful terms

There are many terms that will be helpful to review to make this chapter, as well as future chapters, easier to understand.

A **population** is all members of the group of interest, for example, all people in the Czech Republic who have children. A **sample** is a subset of the population, for example, all the people from the population that participated in your study. Ideally, the sample is a representative subset of the population, meaning that the sample characteristics do not differ from the population characteristics.

A **variable** is something that can take on different values. A variable can either be measured or manipulated. An example of a **measured variable** would be asking someone how far away an object is (the experimenter "measures" the participant's response). A **manipulated variable** could be how many eyes your participant is allowed to use to make the judgment (the experimenter tells the participant to use either one or two eyes, the experimenter "manipulates" the value of the variable by assigning a participant to use either one or two eyes). An experiment has at least one manipulated variable, which is called an **independent variable**. An experiment will also have at least one **measured variable**, which is called a dependent variable. In other studies, such as correlational studies, only measured variables are present and usually those measured variables are simply called "variables."

It is important to note that whether a variable is measured or manipulated is about what is done with the variable, not what the variable is. For example, an experimenter could measure the eye color of participants in a study by using a color meter; here, eye color would be a measured variable (because the experimenter is just recording [measuring] the participant's eye color). In a different study, an experimenter could take images and manipulate a person's eye color using a photo-editing tool; here, eye color would be a manipulated variable (because the experimenter is assigning values of eye color to the image).

There are many types of experiments, but one critical factor is whether the experiment had its independent variable levels between-subjects or within-subjects. This distinction in experimental design influences what type of inferential statistics should be performed. When an experiment has its independent variable levels **between-subjects**, that means different participants experienced each level of the independent variable. When an experiment has its independent variable levels **within-subjects**, that means the same participant experienced each level of the independent variable. Note that you may have learned this distinction using other names. Between-subjects variables are sometimes called independent groups, independent measures, or independent samples. Within-subjects variables are sometimes called within groups, repeated measures, or within samples.

Imagine you wanted to test how well people could perceive how far away an object was. You wanted to know whether having one eye versus two eyes open made any difference in your participants' distance judgments. You randomly assigned half of your participants to perform the judgments with one eye open (the other covered with an eye patch) and the other half of your participants perform the judgments with two eyes open. In this study, distance was your dependent (measured) variable, the number of eyes open was your independent variable, and your independent (manipulated) variable had two levels: One eye or two eyes. Finally, because different participants were in the one eye condition versus the two eyes condition, your independent variable is between-subjects (the name comes from the fact that to compare your two independent variable levels, your comparison takes place between [different] subjects).

Note that nearly the exact same experiment could be performed, except now all participants complete the task once with one eye uncovered and once with two eyes open. The dependent and independent variables are the same, except now the independent variable is within-subjects (the name comes from the fact that to compare your two independent variable levels, your comparison takes place within [the same] subjects).

Statistics can be parametric or nonparametric. **Parametric statistics** have more assumptions, but are more powerful, that is, more likely to find a difference. Depending on the parametric statistic, the assumptions can include having normal distributions, equality of variance, and more. **Nonparametric statistics** make fewer assumptions, can often be used when the assumptions of the parametric test are not met, and are often based on ranks. We will focus primarily on parametric statistics in this chapter.

Sampling distributions

Many inferential statistics are based on sampling distributions. A "regular" distribution is a distribution of scores (for example, a distribution of values for a measured variable). A **sampling distribution** is a theoretical distribution of a sample statistic. For example, the sampling distribution of the mean is a theoretical distribution of sample means from all possible samples of a given size from a population. Reread the previous sentence with the goal of answering the question: Is this ever actually done by a researcher? Hopefully, your answer was "no" and this makes it clear what is meant by a sampling distribution being a theoretical distribution of a sample statistic – it is not actually done in the experiment, but the theoretical distribution (worked out by statisticians) is used to make inferences about the study that was performed.

The sampling distribution of the mean has three important properties. 1) The mean of the sampling distribution of the mean is equal to the population mean. 2) The standard deviation of the sampling distribution of the mean is equal to the standard deviation (σ) divided by the square root of the sample size (n): $\sigma/\sqrt{n}$. In order to distinguish between the standard deviation of scores and the standard deviation of the sampling distribution of the mean, the latter is called the **standard error of the mean** (σ_m), or standard error for short. The standard error is sometimes plotted in graphics, and in some cases used to form confidence intervals that are sometimes plotted in graphics. 3) The sampling distribution of the mean becomes more normal (even if the underlying distribution of scores is not normal) as the sample size increases.

Distributions of scores and sampling distributions can be difficult to understand and keep clear. Figure 2.5 attempts to help with distinguishing these distributions; pay particular attention to how the curves are labeled in each panel. In particular, note that the x-axis gives scores in the left panel, and sample means in the right panel. Also note that the left panel shows the standard deviation of the scores, while the right panel shows the standard error of the sample means. Finally, note that the right panel is actually the sampling distribution of the mean if you took all possible samples of size 25 from a normal population with a $\mu = 100$ and $\sigma = 10$. In other words, if we took all possible samples of size 10 from the distribution of scores in the left panel, we would obtain the sampling distribution of means shown in the right panel.

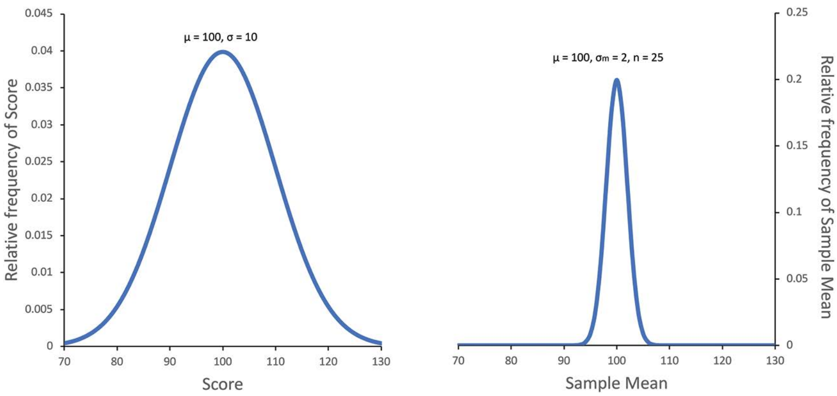


Figure 2.5 A distribution of scores (left) and a sampling distribution of the mean (right).

Specific inferential statistics

One of the potential purposes for a graphic is to represent the results of an inferential statistical test. Below, common inferential statistical tests are reviewed based on the type of study it is. Correlational designs use measured variables, and experimental designs have at least one manipulated variable. Note that the corresponding graphics will be discussed in later chapters.

Correlation

There are many types of correlations, but in all cases, the researcher is using inferential statistics to ask whether the sample data indicate that the variables are correlated. Correlation means that as one variable changes, so does the other variable. A positive correlation means that the variables change in the same way (e.g., as one increases in value, and so does the other). A negative correlation means that the variables change in the opposite way (e.g., as one increases in value, the other decreases in value). In some situations, the interpretation of the sign of the correlation (positive vs. negative) depends critically on how the data was coded.

Pearson's correlation (r) is used to determine if there is a linear relationship between two quantitative variables (e.g., height and weight). A positive correlation would indicate that as height increased, so did weight (which is not surprising at all).

Spearman's correlation (r_s) is used to determine if there is a monotonic relationship between two quantitative variables. A monotonic relationship can be nonlinear, but monotonic means that it only has "one tone" – meaning, for example, that although the relationship might be nonlinear it is always positive.

Point-biserial correlation is used when one variable has categories and the other variable is quantitative (e.g., a correlation between a personality factor: Extravert or introvert and height). Note that often a point biserial correlation might be noted in the text of a paper, but the statistical symbol given is r, just like Pearson's correlation (this is in part because the calculations for a point-biserial correlation are the same as for a Pearson's correlation). Whether a point-biserial correlation is positive or negative is determined in part by how the categories were coded (for example, was extravert encoded as a 1 and introvert a 0, or was it the other way around?).

A **Chi-square test of independence** (χ^2) is used to determine if two variables are related. This test of correlation is primarily used when the data are counts of events in various categories for each variable. For example, an experimenter may want to know whether a personality factor (categories: extraversion, introversion) is correlated with gender (categories: male, female, transgender, nonbinary, other, prefer not to answer). A Chi-square statistic is always positive, so the interpretation of what type of correlation is present ("positive" or "negative") requires carefully examining the observed and expected counts.

Phi-coefficient (φ) is an alternative to a chi-square test of independence when each of the two variables has two and only two categories (e.g., personality factor and pet owner/not a pet owner). Whether a phi-coefficient is positive or negative is determined in part by how both categories were coded (for example, was extravert encoded as a 1 and introvert a 0, or was it the other way around? Also, was pet owner coded as 1 or 0?).

Regression

There are many types of regression. Two will be discussed here, but much of what we apply to regression graphics in later chapters can be applied to most types of regression. In general, the researcher is using inferential statistics to determine if a given variable is a reliable (statistically significant) predictor of another variable.

Simple linear regression is the predictive version of Pearson's correlation. In Pearson's correlation, the researcher determines whether two variables are related. In simple linear regression, the researcher uses one variable (the predictor) to make a prediction of what value the other variable will be (the criterion). Simple linear regression involves one predictor variable and one criterion variable. One of the end results is a regression equation that takes the general form of $\acute{Y} = bX + a$, where $\acute{Y}$ is the predicted value of Y (the criterion), b is the slope of the line, X is the value of the predictor, and a is the y-intercept. For example, the researcher may want to predict your weight ($\acute{Y}$) based on your height (X).

Multiple linear regression is similar to simple linear regression except that multiple predictors are used. Therefore, in multiple linear regression, the researcher uses two or more variables (the predictors) to make a prediction of the value of the criterion. For multiple linear regression involving two predictors, the equation takes the general form of $\acute{Y} = b_1X_1 + b_2X_2 + a$, where the new subscripts for b and X indicate which predictor variable is being used in the equation. For example, the researcher may want to predict your weight ($\acute{Y}$) based on your height (X_1) and your family's average income while you were ages 0–12 (X_2).

Nonlinear regression (simple or multiple) can also be used as well. Here, the equation used to predict the criterion variable is not based on fitting a straight line to the data.

Experiments

Experiments involve at least one manipulated variable (the independent variable) and at least one measured variable (the dependent variable). In general, the descriptions in this section will begin with the simplest experimental questions/designs and then move toward the more complicated designs. In these designs, inferential statistics are used to determine if there is a difference from a known value or a difference between conditions (levels of the independent variable) in the experiment.

A **one-sample t-test** is used when a researcher wants to determine if a sample mean differs from an expected mean value. The source of the expected mean value can come from previous research, population values, or theory. The test uses the sampling distribution of the mean to determine if it appears that the sample comes from a population with the expected mean value. The statistical test is based on the t-distribution which is more leptokurtic (tail heavy) than a normal distribution. The degree of leptokurtosis decreases as the sample size increases.

An **independent-samples t-test** is used when a researcher wants to determine if the means of two samples are statistically different. An independent-samples t-test would be used with an experiment with one between-subjects independent variable with two levels and a dependent variable. Unfortunately, this t-test goes by many names, including independent t-test, independent measures t-test, independent groups t-test, two-groups t-test, two-samples t-test, between-subjects t-test, unpaired-samples t-test, and student's t-test. In an independent-samples t-test, the statistical test is typically based on whether the difference between the means of the groups is zero or not. This statistical test is also based on the t-distribution and the sample sizes of the groups.

A **paired-samples t-test** is used when a researcher wants to determine if the mean difference score between two conditions is zero or not. A paired-samples t-test would be used with an experiment with one within-subjects independent variable with two levels and a dependent variable. This t-test is also known by several names, including dependent samples t-test, repeated measures t-test, dependent groups t-test, and within-groups t-test. This statistical test is similar to a one-sample t-test given that all calculations are based on a single set of difference scores (one difference score for each participant). A paired-samples t-test is based on the t-distribution and the sample size of the group.

A **one-way independent measures analysis of variance** (one-way independent measures ANOVA) is used when a researcher wants to determine if there are any differences between three or more groups on the dependent variable. A one-way independent measures ANOVA would be used with an experiment with one between-subjects independent variable with three or more levels and a dependent variable. The resulting statistic is F (rather than t). Technically, ANOVA (any of the types) can also be used in cases where the independent variable has only two levels. However, traditionally, t-tests are used in those cases.[3] A one-way independent measures ANOVA is also known by several other names, including one-way between-subjects ANOVA, one-way independent groups ANOVA, one-way independent-samples ANOVA, and others. The statistical test is based on the F-distribution (a positively skewed distribution) where the exact shape depends on the number of independent variable levels and the sample sizes of those groups. It is important to remember that the one-way independent measures ANOVA only tells you whether some of your means differ, not which ones. A follow-up test, such as Tukey's honestly significant difference test is necessary to determine which specific means statistically differ from one another (for a history of Tukey's contributions to multiple comparison testing, see Benjamini & Braun, 2002).

One-way repeated measures ANOVA is used when a researcher wants to determine if there are any differences between three conditions on the dependent variable. A one-way repeated measures ANOVA would be used with an experiment with one within-subjects independent variable with three or more levels and a dependent variable. A one-way repeated measures ANOVA is also known by several other names, including one-way within-subjects ANOVA, one-way dependent groups ANOVA, one-way dependent samples ANOVA, and others. The statistical test is also based on the F-distribution that depends on the number of independent variable levels and the sample size. A one-way repeated measures ANOVA also only tells you whether some of your means differ, not which ones. A follow-up test is necessary to determine which specific means statistically differ from one another.

Factorial ANOVA is used when a researcher is testing the effects of two or more independent variables on a dependent variable. In one-way ANOVAs, "one-way" refers to having one independent variable in the experiment. Therefore, in factorial ANOVAs, the number of independent variables can be specified in a similar way: "two-way" for two independent variables, "three-way" for three independent variables, etc. Each independent variable can have two or more levels and be either between- or within-subjects. These parameters give many possible specific designs. For example, a two-way independent measures ANOVA is a design with two between-subjects independent variables. A three-way repeated measures ANOVA is a design with three within-subjects independent variables. A two-way mixed factors ANOVA is a design with one between-subjects and one within-subjects independent variable. The statistical approach in each case is different, but in general follows the basic principles outlined for the simpler one-way ANOVAs. Note that sometimes factorial ANOVAs are named with a different convention that specifies both the number of independent variables and the number of levels for each independent variable. For example, a $2 \times 4 \times 3$ independent measures ANOVA specifies that there are three independent variables (because there are three numbers) and it also specifies that those three independent variables have 2, 4, and 3 levels, respectively. Finally, the independent measures portion of the name tells us that all the independent variables are between-subjects.

If you have a study that is not an experiment (i.e., you only have measured variables, no independent [manipulated] variables), Table 2.1 organizes the inferential statistics discussed in this chapter. If you have an experiment (i.e., independent variable(s) and a dependent variable), Table 2.2 organizes the most likely inferential statistics. As you may be aware, there are also nonparametric tests that correspond to the parametric tests given in Table 2.2. Table 2.3 gives the nonparametric equivalents to the parametric statistical tests given in Table 2.2. None of the tables prepares you for the following question, though: What is a pirate's favorite statistic? The answer is in this note.[4]

Table 2.1 Correlation and Regression

Variable 1	*Variable 2*	*Goal*	*Statistic(s)*
Quantitative	Quantitative	Establish linear relationship	Pearson's (r)
Quantitative	Quantitative	Predict one variable from another in a linear relationship	Simple linear regression[a]
Quantitative	Quantitative	Establish monotonic relationship	Spearman's (r_s)
Two categories	Quantitative	Establish relationship	Point-biserial correlation
Categories	Categories	Establish relationship	Chi-square test of independence (χ^2)
Two categories	Two categories	Establish relationship	Chi-square test of independence (χ^2) or phi coefficient (φ)

[a] Multiple linear regression if more than one predictor variable is used.

Table 2.2 Parametric Statistics for Experiments

Independent Variables	*Number of Independent Variable Levels*	*Type of Independent Variable Levels*	*Statistic*
None	None	Comparison to population value	One-sample t-test
One	Two	Between-subjects	Independent-samples t-test
One	Two	Within-subjects	Paired-samples t-test
One	Three or more	Between-subjects	One-way independent measures ANOVA
One	Three or more	Within-subjects	One-way repeated measures ANOVA
Two or more	Two or more	Between-subjects	X[a]-way independent measures ANOVA
Two or more	Two or more	Within-subjects	X[a]-way repeated measures ANOVA
Two or more	Two or more	Some between-subjects and some within-subjects	X[a]-way mixed factors ANOVA

[a] The value of X should correspond to the number of independent variables.

Table 2.3 Parametric Statistics and Corresponding Nonparametric Statistics for Experiments

Parametric Statistic	*Corresponding Nonparametric Statistic*
One-sample t-test	Single-sample Wilcoxon signed-rank test
Independent-samples t-test	Mann-Whitney U test
Paired-samples t-test	Wilcoxon signed-rank test
One-way independent measures ANOVA	Kruskal-Wallis one-way ANOVA
One-way repeated measures ANOVA	Friedman's ANOVA on ranks
X[a]-way independent measures ANOVA	None available
X[a]-way repeated measures ANOVA	None available
X[a]-way mixed factors ANOVA	None available

[a] The value of X should correspond to the number of independent variables.

Scales of measurement

Scales of measurement can inform decisions about what graphics to use. **Scales of measurement** describe how numbers are being used. Stevens (1946) first developed the four classic scales of measurement we will consider: nominal, ordinal, interval, and ratio. Nominal scales are the simplest use of numbers and each subsequent scale – ordinal, interval, ratio – adds another aspect of how numbers are being used to the previous scale.

Nominal scales of measurement use numbers as labels for categories. For example, if a variable in a study was whether someone self-identified as an extravert or an introvert, we could assign a 1 for extravert and a 2 for introvert. Note that there is no meaning to the choice of numbers; we could have just as easily chosen 32 and 41 for extravert and introvert. In other words, 1 and 2 just serve as labels; there is nothing "first" or better about 1 compared to 2 in terms of how we are using the numbers. Likewise, a 2 is not twice as anything as a 1. Another way to think about nominal scales is that you have categories that you assign numbers to, and any number could potentially go with any category. In Stevens's (1946) terms, a nominal scale allows one to determine whether there is equality of category (are two people both extraverts?).

Ordinal scales of measurement use numbers not only to categorize (the observation assigned 1), but to indicate rank (order). For example, one could rank restaurants from best to worst, with a 1 given to the best restaurant, a 2 to the second best, etc. The numbers are used to tell you about preference, but the numbers do not tell you how much the restaurant ranked 1 differs from the restaurant ranked 2. It is possible that it was a very difficult decision on which to rank first (first and second are very similar in quality), or it could have been that it was easy to rank which restaurant was first versus second (first and second were very different in quality). The numbers as used here do not capture anything about the amount of difference between rank 1 and 2, just their order. In Stevens's (1946) terms, an ordinal scale allows one to determine whether something is greater or less than something else (which restaurant is better?).

Interval scales of measurement use numbers to not only categorize and order (the temperature of x°C, which is smaller or larger than y°C), but also indicate the amount of difference between categories. For example, a temperature of 10°C is colder than a temperature of 15°C. However, the numbers also indicate something about the amount of difference as well. Specifically, on an interval scale, the same difference in numbers (10 vs. 15 = 5°C difference) means the same thing anywhere on the scale: 17°C versus 22°C is also a 5°C difference, and that difference of 5°C means the same as any other difference of 5°C. In Stevens's (1946) terms, an interval scale allows one to determine whether differences are equal (is the difference between 2 and 4°C the same as the difference between 15 and 17°C?).

Ratio scales of measurement not only categorize, order, and indicate magnitude (a specific height, which is taller than other heights, and where 2 centimeters means the same thing anywhere on the scale) – but also have a true zero point. Height has a true zero point (0 cm is no height) and this allows statements that make comparisons of magnitudes in terms of ratios: 2 meters is twice as tall as 1 meter. Notice that the same cannot be said with temperature in Celsius: 10°C is not twice as much temperature as 5°C. It is not temperature per se as a thing that is not a ratio scale, though, it is just how we are using the numbers – the Kelvin scale for temperature does have a true zero point. In Stevens's (1946) terms, a ratio scale allows one to determine whether ratios are equal (is the ratio of 2 to 4 meters the same as the ratio of 6 to 12 meters?).

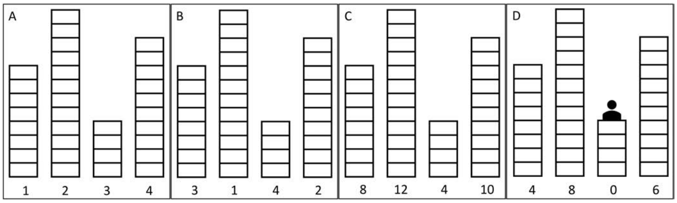


Figure 2.6 Scales of measurement for buildings

The panels of Figure 2.6 allow you to test yourself on your understanding of nominal, interval, and ratio scales of measurement. In Figure 2.6, let us assume that each stack of rectangles represents a building with each rectangle in the stack representing a floor of the building. Write down your answers to what scale of measurement is being used in each of the panels before reading further. In panel A of Figure 2.6, the buildings are labeled 1 through 4 from left to right, without regard to the number of floors each building has. This is using numbers in a nominal fashion (labeling the buildings: Building 1, building 2, etc.). In panel B of Figure 2.6, the same four numbers are used, but they are used differently. A 1 was assigned to the building with the most floors, a 2 for the building with the second most floors, etc. This is using numbers as an ordinal scale of measurement. Note that comparing panel A and panel B is perhaps the most important part of Figure 2.6 because it makes clear that it is not which numbers are used that determines what the scale of measurement is (the same four numbers are used in panels A and B), but how the numbers are used (as labels for buildings or as ranks of building heights). In panel C of Figure 2.6, the numbers used correspond to the number of floors the building has. This is using numbers on a ratio scale of measurement – there is a true zero point (no floors) and it makes sense to say one building has twice as many floors as another building. In panel D of Figure 2.6, the buildings seem to be numbered based on how many floors there are equal to or above the person on the shortest building. This is an interval scale of measurement – a difference of two floors means the same across the buildings (compare all the nonzero-labeled buildings). However, there is no true zero point as in panel C, so therefore it would not make sense to say that the building labeled with an 8 is twice as tall as the building labeled with a 4 (because it is not true!).

Scales of measurement, or more particularly, how scales of measurement have been used with regard to inferential statistics has been controversial in some fields (e.g., psychology). In Stevens's (1946) paper establishing nominal, ordinal, interval, and ratio scales, Stevens claimed that only certain statistics are permissible depending on the scale. For example, nominal scales only allow counts and modes to be calculated, ordinal scales only allow the calculation of counts, modes, medians, and percentiles, while interval and ratio scales allow the calculation of counts, modes, medians, percentiles, means, and standard deviation. Stevens (1946) points out that many of the scales of measurement psychologists use are probably best considered ordinal scales.

Stevens's (1946) proscription about inappropriate statistics was controversial and has generated much discussion in the literature. Borgatta and Bohrnstedt (1980) have argued that although the specific measures used by psychologists may appear to be ordinal in some circumstances, one should consider the nature of the underlying variable too when determining what is statistically appropriate. Gaito (1980) notes that associating scales of measurement with particular tests is not found in the mathematics and statistics literature, but is common in the psychology literature. Velleman and Wilkinson (1993) argue that scales of measurement are not a quality of the data itself, but of the questions the researcher intends to ask about the data. Scales, therefore, should not determine what statistical approach should be taken.

Others have argued that different theories of measurement are the cause of the conflicting advice concerning scales of measurement and statistical procedures (see, for example, Hand, 1996; Michell, 1986). In general, be careful about being too proscriptive or prescriptive regarding which statistics should or should not be performed based on scales of measurement. As Lord (1953) cleverly and humorously pointed out – the numbers are not aware of where they came from.[5]

A quick demonstration of the proscription being okay to break can be made with the personality factor variable (extravert or introvert, coded as 1 and 2), which we said was a nominal scale. Imagine we had three extraverts and seven introverts in a small sample (that is, the data are: 1, 1, 1, 2, 2, 2, 2, 2, 2, 2). Stevens (1946) would suggest that I should not calculate a mean on nominal data. Let us see what happens if I do: The sum of all the scores is 17, there are 10 scores, so the mean is 1.7. Nothing bad happened from this calculation, and the value of 1.7 is useful. Obviously, no one can be a "1.7" on my nominal scale, but I can tell you that 70% of my sample was introverts, given how the data were coded as 1 and 2. Care does need to be taken when interpreting numbers, but that applies to any scale of measurement and any calculations.

Scales of measurement certainly are important to consider, though, because they do influence what operations make sense to do (Townsend & Ashby, 1984). In fact, not knowing what scale of measurement you are working with can ruin a gift. The following is a true story: I wear a size 12 shoe (US size); it is important to note that US shoe sizes are on an interval scale. In particular, a US shoe size 6 is not half of a shoe size 12 (see Figure 2.7). However, "hypothetically" your mother may have a pattern for size 6 slippers and is not aware of the difference between interval and ratio scales. Therefore, she was also not aware that doubling a size 6 pattern would not result in size 12 slippers. The slippers she created were quite comical, given that they were about 18 inches long. Although not intended, everyone present when I received the slippers did get to enjoy the gift of laughter.

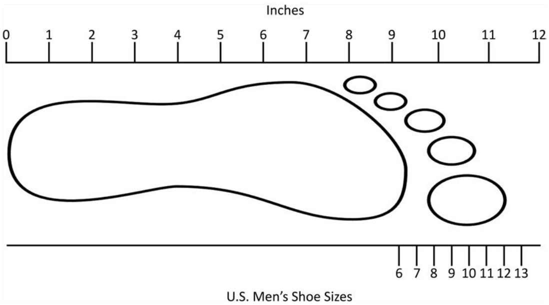


Figure 2.7 US shoe size scale. Note that the top scale gives inches, not shoe size. The bottom scale gives men's adult foot sizes on the far right of the scale.

Beyond these practical gift-making concerns, in later chapters, we will see that considering the scales of measurement of our variables may influence our choices about how to represent a variable in a graphic.

What does this have to do with graphics?

In this chapter, we have reviewed descriptive and inferential statistics, and also considered scales of measurement. One of the goals of a graph is to help the reader understand the properties of your data through representations of descriptive statistics and/or the distribution of your data. Another common goal of a graph is to help the reader understand what your inferential statistical analysis shows. In later chapters, we will develop such graphs. However, to develop good graphs, we need to understand what the reader needs from a graph (Chapter 3) and what principles we can develop to make graphs that help the reader's understanding (Chapter 4).

Notes

1 A good way to test your knowledge of measures of central tendency is to consider the following set of riddles, which are best considered in the order given (the answers appear at the end of this footnote): How does a statistician like her apple pie most of the time? How does a statistician like her apple pie half of the time? How does a statistician like her apple pie on average?
Answers: A la mode, a la median, a la mean (note that at no point did I say these were good riddles).

2 Technically a normal distribution is one that follows the function below, where μ is the mean of the distribution and σ is the standard deviation of the distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

3 Fun fact: If you perform an independent-samples t-test and a one-way independent measures ANOVA on the same data (collected from a study with two independent groups), you will find that $F = t^2$. Note that my definition of fun may differ from the reader's definition of fun.

4 Said in a pirate's voice: Pearson's Arrrrrrrrrr!

5 Lord's paper also gives definitive evidence that academic humor has existed since at least 1953.

References

Benjamini, Y. & Braun, H. (2002). John W. Tukey's contributions to multiple comparisons. *The Annals of Statistics*, *30*, 1576–1594. https://doi.org/10.1002/j.2333-8504.2002.tb01891.x

Borgatta, E. F. & Bohrnstedt, G. W. (1980). Level of measurement-once over again. *Sociological Methods and Research*, *9*, 147–160. https://doi.org/10.1177/004912418000900202

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin*, *87*, 564–567. https://doi.org/10.1037/0033-2909.87.3.564

Hand, D. (1996). Statistics and the theory of measurement. *Journal of the Royal Statistical Society, A*, *159*, 486–487. https://doi.org/10.2307/2983326

Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist*, *8*, 750–751. https://doi.org/10.1037/h0063675

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, *100*, 398–407. https://doi.org/10.1037/0033-2909.100.3.398

Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680. https://doi.org/10.1126/science.103.2684.677

Townsend, J. T. & Ashby, F. G. (1984). Measurement scales and statistics: the misconception misconceived. *Psychological Bulletin*, *96*, 394–401. https://doi.org/10.1037/0033-2909.96.2.394

Velleman, P.F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading for classifying statistical methodology. *The American Statistician*, *47*, 65–72. https://doi.org/10.2307/2684788

Key concepts

Between-subjects
Chi-square Test of Independence
Descriptive Statistics
Extreme Value
Factorial ANOVA
Independent Samples t-test
Independent Variable
Inferential Statistics
Interquartile Range
Interval
Kurtosis
Leptokurtic
Manipulated Variable
Mean
Measured Variable
Median
Mode

Multiple Linear Regression
Negative Skewness
Nominal
Nonlinear Regression
Non-parametric Statistics
Normal Distribution
One sample t-test
One-way Independent Measures Analysis of Variance
One-way Repeated Measures ANOVA
Ordinal
Outlier
Sampling Distribution
Paired samples t-test
Parametric Statistics
Pearson's Correlation
Phi-coefficient
Platykurtic
Point-biserial Correlation
Population
Positive Skewness
Ratio
Sample
Scales of Measurement
Simple Linear Regression
Skewness
Spearman's Correlation
Standard Deviation
Standard Error of the Mean
Variable
Within subjects

Practice problems

1 What are the measures of central tendency? What do they tell you about the data? Which involves dividing scores into equal halves (but ignoring the value of each score)?
2 What are the measures of variability? What do they tell you about the data? Which measure of variability gives you the usual distance scores are from the mean?
3 What are the properties of a normal distribution? (shape? spread? measures of central tendency?)
4 Describe three general ways that a distribution can be nonnormal.
5 In terms of inferential statistics, what are a population and a sample?
6 What are measured and manipulated variables? How do they relate to independent and dependent variables?
7 What does it mean for an independent variable to be between-subjects versus within-subjects?
8 What is the difference between parametric versus nonparametric statistics?
9 What is a sampling distribution?
10 What is the difference between a standard deviation and the standard error of the mean?
11 Describe the five types of correlation.
12 What is the difference between simple linear regression, multiple regression, and nonlinear regression?
13 What are the three types of t-tests and under what circumstances are each used?
14 What are the three types of ANOVAs and under what circumstances are each used?
15 What is the nonparametric equivalent of an independent-samples t-test? What is the non-parametric equivalent of a one-way repeated measures ANOVA?
16 What are the differences between nominal, ordinal, interval, and ratio scales of measurement?

3 What does the reader need from a graphic?

In this chapter, we explore what the reader needs from a graphic. Understanding the needs of the reader will help you to decide what graphic to make and what decisions to make when creating the graphic.

Recall that our definition of a graphic is any illustration of data. In the context of this book, this usually means either a table or a graph. In order to discuss tables and graphs, we first need to define the parts of a graph and a table.

The parts of a table

A table contains data in cells organized in rows and columns. Table 3.1 is a table that, rather than showing data, illustrates the parts of a table. Below the **table title** ("Table 3.1") is the **table caption** that describes the contents of the table (in this case, "Illustration of the Main Portions of a Table"). Horizontal lines are used to indicate the beginning and the end of the table as well as to separate column labels from data cells.

A **stub heading** labels what is given in the left-most column of the table. An optional **column spanner** labels multiple columns and reduces repetition in labeling columns if they were not present. Column spanners are often followed by a horizontal line that does not include the stub heading column. **Column headings** (and any associated column spanners) label the data that appear in each column. Likewise, **row headers** label the data that appears in each

Table 3.1 Illustration of the Main Portions of a Table

Stub Heading	*Column Spanner*		*Column Spanner*	
	Column 1 Heading	*Column 2 Heading*	*Column 1 Heading*	*Column 2 Heading*
Table Spanner				
Row 1 Header	Data Cell	Data Cell	Data Cell	Data Cell
Row 2 Header	Data Cell	Data Cell	Data Cell	Data Cell
Row Average or Total	Data Cell	Data Cell	Data Cell	Data Cell
Table Spanner[a]				
Row 1 Header	Data Cell	Data Cell	Data Cell	Data Cell
Row 2 Header	Data Cell	Data Cell	Data Cell	Data Cell
Row Average or Total	Data Cell	Data Cell	Data Cell	Data Cell

Note: Information that refers to the entire table.

[a] A specific note about a portion of the table.

DOI: 10.4324/9781003471264-3

row. A **data cell** is the location where data is given, with the source of the data indicated by the column and row heading. Rows often include an average or total row (of the data above the average/total row).

A **table spanner** covers the width of the table and labels a portion of the table. A **table note** provides information about the entire table, while a **specific note** (linked by a symbol to a portion of the table) gives information that applies to only a portion of the table.

Although Table 3.1 lays out this information visually, it is helpful to see that table as it might actually be used. Table 3.2 shows a version of Table 3.1 "in use" for a study recording flight distances of paper airplanes in cm for two airplane types (Basic Dart and the Raven, see foldnfly.com), underhand versus overhand throwing, and expert and novice throwers on two different days.

Table 3.2 Distances in Centimeters Paper Airplanes Flew Based on Thrower Experience Level, Throwing Method, Plane Type, and Day of Throw

Experience Level	*Underhand*		*Overhand*	
	Basic Dart	*The Raven*	*Basic Dart*	*The Raven*
Day 1 Distances				
Expert	434	417	505	478
Novice	418	399	441	432
Average	426	408	473	455
Day 2 Distances[a]				
Expert	454	432	530	502
Novice	448	416	514	494
Average	451	424	522	498

Note: Throwers were the same on Days 1 and 2.

[a] Day 2 throws were completed three days after Day 1 throws.

Table 3.2 allows the reader to see how each of the components of the table might look in use. The title gives information about all the variables, including the units of measure (centimeters). The column spanners allow the quick visual separation of underhand versus overhand throwing without having to use the names twice as often in the column headings in the next row. Column headings should follow the same order under each column spanner. The table spanners quickly and easily separate data from Day 1 versus Day 2. The average rows are probably unnecessary in this table, given that each is the average of just the two rows immediately above, but they are included here for illustration purposes. Average or total rows are often visually indicated by horizontal lines above and below them. The table note indicates information about the entire table, while the specific note gives information about Day 2 throws specifically.

Note that many tables are much simpler than the ones given in Tables 3.1 and 3.2. Table spanners, column spanners, and multiple horizontal lines are often unnecessary in relatively simple tables. Let the data guide your decisions and expect that you may need to try several approaches to making a table before determining what arrangement best presents the data. Chapter 8 in *Show Me the Numbers* by Stephen Few (2012) is an excellent guide to table design.

The parts of a graph

A graph looks very different than a table, even if it is representing the same data. Figure 3.1 is a labeled version of the data from Day 1 of Table 3.2 and the labels are modeled after the approach of Cleveland (1994). Figure 3.1 shows the basic components of a graph. The **data rectangle** is the minimum rectangle that includes all the plotted data. The ends of the labeled horizontal scale and vertical scale in the graph would give the **scale-line rectangle** (not labeled for clarity). In general, the data rectangle is normally slightly smaller than the scale-line rectangle (in Figure 3.1, it is significantly smaller to give room for the legend and panel label). Each scale has a **scale line** and associated **scale labels** that indicate what is plotted on the graph. **Tick marks** anchor particular values given as **tick mark labels** to the scale line. A **data symbol** indicates the actual data in the graph. The **legend**, when necessary, gives information on differences in the data symbols plotted. In Figure 3.1, the experience level of the thrower is shown by the data labels within the legend. Although Figure 3.1 consists of a single panel, I have provided a **panel label,** given that the reader knows that these data are only a portion of the data from Table 3.2. The title and caption serve a similar purpose to the table title and caption. A **reference line** is sometimes included on a graph. For example, if there was something special about 400 cm (for example, the average distance flown from a past study), a horizontal reference line at 400 cm could be included in the graph (and the reference line would be described either in the caption or in the main text associated with the graph).

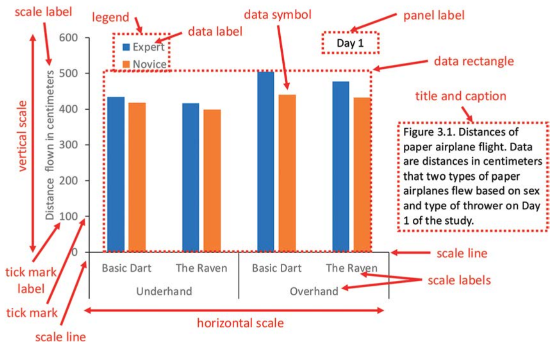


Figure 3.1 The parts of a graph.

Figure 3.2 is a two-panel graph giving all the data from Table 3.2, and removes the labels of the graph components. Although this is not a great graph (as we will see later in the book), the graph does allow one to quickly see the pattern of the data. Specifically, the consistent expert/novice difference in distances. With a little more effort, it can also be noted that the size of the expert/novice difference decreases between Day 1 and Day 2 for overhand throws. Even after knowing this difference exists for overhand throws between Days 1 and 2, it is almost impossible to notice this difference by examining the values in Table 3.2. This underscores the big advantage of graphs over tables: One can easily assess the overall patterns in the data in graphs. What advantage then do tables have over graphs? It can be a challenge to determine the exact values of any of the bars in Figure 3.2, but exact values are easy to obtain in Table 3.2. Graphs are (or at least can be) great for patterns. Tables are great for exact values.

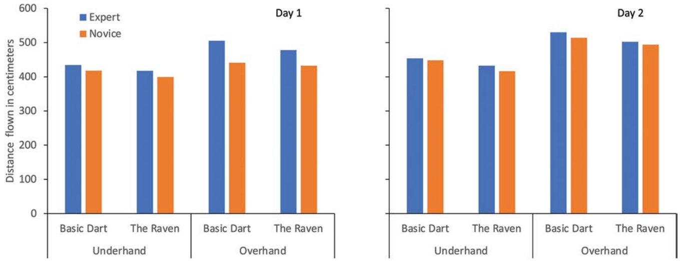


Figure 3.2 A graph of the data in Table 3.2.

What does a reader need (from a graphic)?

In the process of labeling the components of a table and a graph, we have already determined that tables are excellent for learning about exact values, while graphs are excellent for learning about the overall pattern of the data (see, for example, Jarvenpaa & Dickson, 1988). However, there are more fundamental tasks that a reader needs to be able to perform successfully in order to extract the information from a table or perceive the pattern in a graph.

Cleveland (1994) discusses three visual operations necessary for pattern perception in processing a graph: Detection, assembly, and estimation. **Detection** is the visual recognition of an object that encodes information in a graphic. This includes perceiving the detected object as different from other objects. In Figure 3.2, this is perceiving the difference between the blue and the orange bars. **Assembly** is the grouping of detected objects. In Figure 3.2, this would be grouping all the blue bars together as being from one condition. **Estimation** is the visual assessment of the relative magnitudes of two or more values. In Figure 3.2, this would be perceiving that in each pair of bars, the blue bars are longer than the orange bars. Note that this particular estimation is the easy-to-see pattern that we noted as an advantage of graphs over tables. Estimation was further developed by Cleveland (1994) as a process that can involve **discrimination** (determining whether $a = b$ or $a \neq b$), **ranking** (determining whether $a > b$, $a < b$, or $a = b$), and/or **ratioing** (determining a/b). The easy-to-see pattern in Figure 3.2 is supported by the blue and orange bars being easy to perceive (detection), easy to group (assembly), and easy to estimate which bars are longer in each pair (estimation, specifically: Ranking blue > orange in each pair).

Figure 3.2 was constructed well to allow the reader to perceive the interesting patterns in the data. However, it is instructive to imagine presenting the same data when choices are made that make the reader's task more difficult. Figure 3.3 shows an example of poorer choices. The choice of two similar shades of blue makes it difficult to detect and assemble the bars in the graph. The choice of the y-axis scale (0 to 3000 cm) makes the data rectangle too small vertically, which makes accurate estimation difficult. Specifically, the graph reader might read Figure 3.3 as indicating that for all pairs of bars, the bars are approximately equal. If that is the initial impression, it would be impossible for the graph reader to note the decrease in the difference between expert and novice throwers for overhand throws that happens between Days 1 and 2. As a graph maker, your task is to make sure your choices allow the graph reader to successfully recognize the patterns in the data (detection, assembly, and estimation).

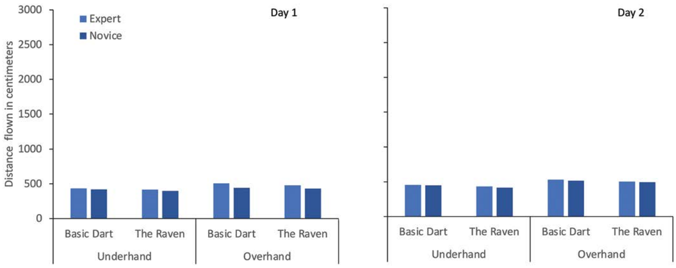


Figure 3.3 Poor choices make a graph less useful.

An additional task for graphic perception is table look-up (Cleveland, 1994). **Table look-up** refers to determining (or estimating) the value of each data symbol in a graph. In a graph, this involves **scanning** (looking up/down and left/right from a data symbol to scale-line tick marks), **interpolating** (estimating the distance between tick marks for the data symbols), and **matching** (determining how the data symbol is related to data labels and any other labels). For example, in Figure 3.2, the graph reader may want to estimate the distance flown from the leftmost blue bar. The graph reader would scan down to the horizontal axis to see that the data corresponds to the Basic Dart airplane type thrown underhand. The graph reader would scan left to the vertical axis to see that the top of the bar falls between the 400 and 500 tick marks. The graph reader would then interpolate and perhaps estimate that the bar is about 1/3 of the way to the 500 tick mark and estimate the value of the top of the bar as 433 cm. Finally, the graph reader will need to match the blue to the blue in the legend to determine that this is an expert throwing, and read the panel label to determine that the data are for Day 1. Figure 3.4 shows these table look-up tasks graphically. Referring back to Table 3.2 shows that we did a great job in terms of "table look-up" (Table 3.2 shows that on Day 2, expert underhand Basic Dart throws were 434 cm). Note, however, a different graph reader might perceive the top of the bar as only ¼ of the way to the 500 tick mark and estimate the top of the bar as being 425 cm – a worse estimate. Earlier, we made the observation that graphs are not particularly good at giving exact values; now we can say that another way: Graphs are not particularly good at providing accurate table look-up. Note that table look-up becomes more difficult in Figure 3.3, where the data rectangle is unusually small.

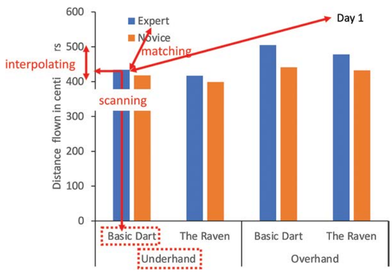


Figure 3.4 Table look-up in a graph.

Table look-up can also be applied to tables. In the case of tables, scanning involves looking up/down and left/right from a data cell to the stub heading and column heading). Interpolating is normally not involved in a table because the data cell gives exact values, and matching may be involved if column spanners and table spanners are present. In Table 3.2, the upper leftmost data cell is 434 cm (no interpolation necessary) and scanning indicates this value is for an expert throwing the Basic Dart. Matching the data cells to the column spanner and table spanner shows that these expert Basic Dart data are for underhand throws on Day 1. Figure 3.5 illustrates the table look-up tasks for Table 3.2.

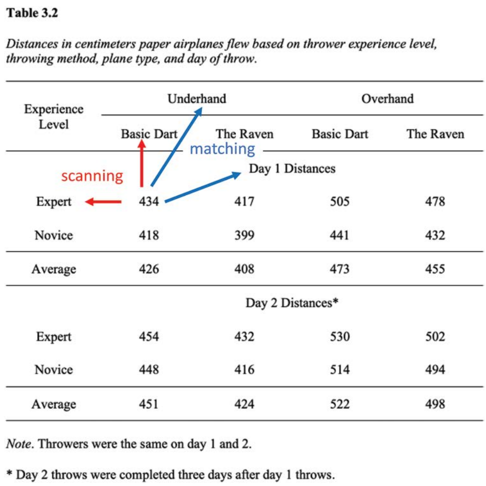


Table 3.2

Distances in centimeters paper airplanes flew based on thrower experience level, throwing method, plane type, and day of throw.

Experience Level	Underhand		Overhand	
	Basic Dart	The Raven	Basic Dart	The Raven
	Day 1 Distances			
Expert	434	417	505	478
Novice	418	399	441	432
Average	426	408	473	455
	Day 2 Distances*			
Expert	454	432	530	502
Novice	448	416	514	494
Average	451	424	522	498

Note. Throwers were the same on day 1 and 2.

* Day 2 throws were completed three days after day 1 throws.

Figure 3.5 Table look-up in a table.

To illustrate why tables are not particularly useful for perceiving patterns in the data, it is informative to attempt to apply Cleveland's three visual operations necessary for pattern perception (detection, assembly, and estimation) to a table. Figure 3.6 shows the detection and assembly tasks involved in a portion of Table 3.2. Figure 3.6 shows that the table reader needs to detect the differences in stub headings (expert vs. novice), column headings (Basic Dart vs. the Raven), column spanners (Underhand vs. Overhand), and table spanners (Day 1 vs. Day 2). Assembly is shown by the left dashed box, where all underhand throws have been grouped for possible comparison to the right dashed box of all overhand throws. There are also other assembly tasks not illustrated (e.g., the expert rows versus the novice rows, the Basic Dart column**s** vs. the Raven column**s**). Furthermore, Figure 3.7 shows **a few** (but not all) of the estimation tasks that could potentially be involved. In Figure 3.7, values between rows show the difference between expert and novice throw distances, while numbers between columns give the difference between the Basic Dart and the Raven distances for the same type of throw. Missing are the estimates between underhand versus overhand throwing (keeping all else constant), as well as Day 1 versus Day 2 distances. Referring back to Table 3.2 (where none of these calculated values are shown for you), it is apparent how difficult it would be for a table reader to make all the calculations that could be made, remember them, and make the relevant comparisons for estimation (discrimination, ranking, and ratioing). Psychologists would describe this as the **cognitive load** (the amount of working memory resources needed for a task) being too high (Sweller, 1988).

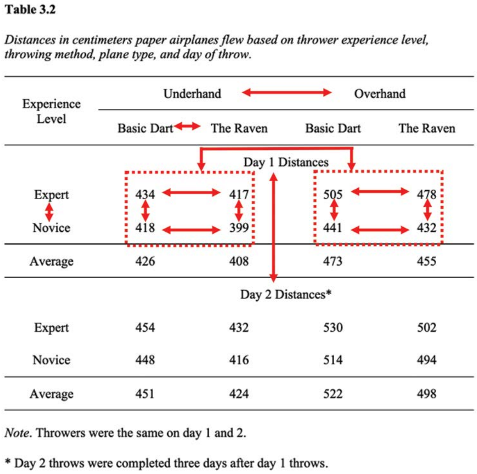


Table 3.2

Distances in centimeters paper airplanes flew based on thrower experience level, throwing method, plane type, and day of throw.

Experience Level	Underhand		Overhand	
	Basic Dart	The Raven	Basic Dart	The Raven
	Day 1 Distances			
Expert	434	417	505	478
Novice	418	399	441	432
Average	426	408	473	455
	Day 2 Distances*			
Expert	454	432	530	502
Novice	448	416	514	494
Average	451	424	522	498

Note. Throwers were the same on day 1 and 2.

* Day 2 throws were completed three days after day 1 throws.

Figure 3.6 Table detection and assembly.

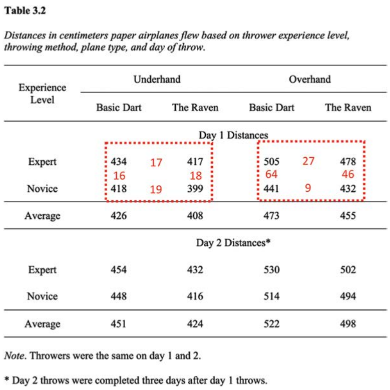


Table 3.2

Distances in centimeters paper airplanes flew based on thrower experience level, throwing method, plane type, and day of throw.

Experience Level	Underhand		Overhand	
	Basic Dart	The Raven	Basic Dart	The Raven
	Day 1 Distances			
Expert	434	417	505	478
Novice	418	399	441	432
Average	426	408	473	455
	Day 2 Distances*			
Expert	454	432	530	502
Novice	448	416	514	494
Average	451	424	522	498

Note. Throwers were the same on day 1 and 2.

* Day 2 throws were completed three days after day 1 throws.

Figure 3.7 Table estimation.

Summary

We have seen that tables are excellent at giving exact values, but are difficult to use for perceiving patterns. You also now understand the specific reasons why that is true. Exact values are given in tables, so no interpolation is necessary – only scanning and matching are required in your table look-up task (see Figure 3.5). Perceiving patterns within a table is difficult because even though the calculations might be simple, there are too many calculations and comparisons to possibly make (estimation task) to easily determine what the patterns are (see Figures 3.6 and 3.7). For graphs, table look-up tasks are going to be more prone to error (see Figure 3.4). However, a graph can make overall patterns in the data visually apparent without any calculations (see Figure 3.2) as long as poor decisions are avoided by the graph maker (see Figure 3.3).

Kelly, Jasperse, and Westbrooke (2005) suggest that data be represented either in the text (1–5 numbers), as indented text tables (3–8 numbers), or as full tables (5–100 numbers). Large amounts of data in general should be presented in a graph and only in rare circumstances presented both in a table and a graph (Kelly et al., 2005).

In addition to the information you are learning about being a graphic maker, what you have learned can also make you a better graphic reader. Although you may have performed many of the tasks of pattern perception and table look-up before, having explicit knowledge of the functions and how the graph maker likely intends you accomplish those tasks may make it easier to accurately and completely read a graph or table. Knowing the difference between graphs and tables (and what each is better at conveying to the reader) may make it clearer to you as a reader why some graphs/tables are difficult to use (because perhaps the wrong graphic has been given to you for your task).

In the next chapter, principles are developed that will help you design graphs that ensure that the graph reader has all the necessary information to accurately interpret the graph, minimize the demands on working memory, and guide the graph reader to the appropriate aspect of the graph.

References

Cleveland, W. S. (1994). *The elements of graphing data*. Hobart Press.

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Analytics Press.

Jarvenpaa, S. L. & Dickson, G. W. (1988). Graphics and managerial decision making: Research-based guidelines. *Communications of the ACM*, *31*(6), 764–774. https://doi.org/10.1145/62959.62971

Kelly, D., Jasperse, J., & Westbrooke, I. (2005). Designing science graphs for data analysis and presentation. *Department of Conservation Technical Series*, *32*, 1–68.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Key concepts

Assembly
Cognitive Load
Column Headings
Column Spanner
Data Cell
Data Rectangle
Data Symbol
Detection
Discrimination
Estimation
Interpolating
Legend
Matching
Panel Label
Reference Line
Ranking
Ratioing
Row Headers
Scale Labels
Scale Line
Scale-line Rectangle
Scanning
Stub Heading
Specific Note
Table Caption
Table Look-up
Table Note
Table Spanner
Table Title
Tick Mark Labels
Tick Marks

Practice problems

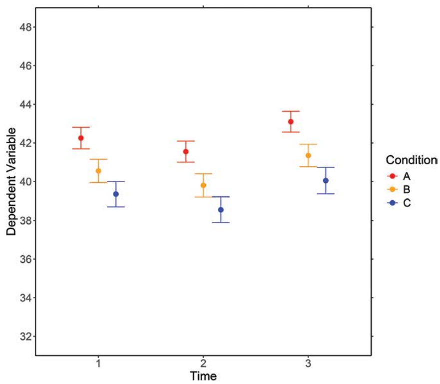
1 What is the difference between a table spanner and a column spanner? What are their purposes?
2 What is the difference between the scale-line rectangle and the data rectangle? Ideally, how should they be related?

3 What are the three visual operations necessary for pattern perception in processing a graph, according to Cleveland (1994)?
4 What are each of the following in processing/perceiving a graph: Discrimination, ranking, and ratioing?
5 What problem does a small data rectangle cause?
6 What is table look-up?
7 What are scanning, interpolating, and matching in table look-up?
8 What is cognitive load and should it be minimized or maximized in a graph or table?
9 If you want to show exact values, which is better, a table or a graph? Why?
10 If you want to show the overall pattern of results, which is better, a table or a graph? Why?
11 Label the parts of the table. Use the following terms: Column spanner, column heading, row average/total, row header, stub heading.

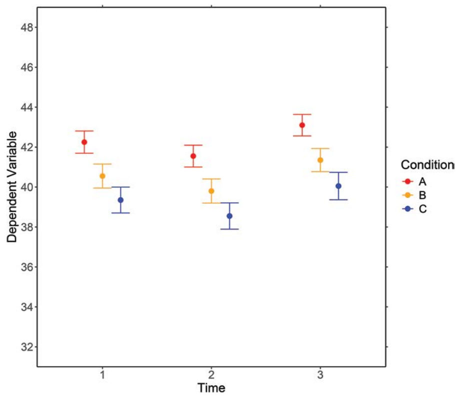
Salaries in Thousands of People Based on Whether a Degree Was Earned, Family of Origin Socioeconomic Status (SES), University Type, and Career Stage

University Type	*No Degree*		*Degree*	
	Low SES	*Middle SES*	*Low SES*	*Middle SES*
Early Career				
Public	434	417	505	478
Private	418	399	441	432
Average	426	408	473	455
Late Career				
Public	454	432	530	502
Private	448	416	514	494
Average	451	424	522	498

12 Label the parts of the figure. Use the following terms: Data label, data rectangle, data symbol, horizontal scale, legend, scale label, scale line, tick mark, tick mark label, vertical scale.



13 Use the figure to illustrate the concepts of scanning, interpolating, and matching.



4 Guiding principles for graphs

In this chapter, the principles for graph making used in the subsequent chapters are developed. Just as the rules of grammar enable the comprehension of language, guidelines for graphing help make your graphs easier to interpret. Cleveland (1994) suggests that a large amount of quantitative information can be presented in a small region. Although this may be true, the data will not be clear, nor will the purpose of the graph be apparent, unless the graph is created thoughtfully.

Imagine for a moment that I wrote the following: "Writing really well should result in a reasonably constructed sentence. A sentence should not have a bunch of extra words that do not advance the meaning of the sentence, and likewise a paragraph should not have a bunch, or even a few, extra sentences. This is because it is like a drawing that should not have any extra lines that are not helpful. Writing well is also like a machine that does not have extra parts that do not make the machine work any better or worse. Writing well does not mean all the sentences should be short though, sometimes they can be long. However, the writer should also try to include all the detail that is needed and not just say a little bit about the topic. It is important that the words help the reader understand." Although the meaning might be apparent, it is difficult to read. Strunk (2006) said it much better in his classic on writing style: "Vigorous writing is concise. A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts. This requires not that the writer make all his sentences short, or that he avoid all detail and treat his subjects only in outline, but that he make every word tell."[1]

Your graph should be like the Strunk quote, not like my (purposeful) butchering of the quote. To translate Strunk to graph making: Your graph should be concise and have no unnecessary parts.[2] It should provide the reader with all the important details, ensuring that every part of your graph tells part of the story.

You may be hoping that this guide will ensure that you will make the perfect graph on your first attempt. However, there is no best graph for a given data set. What makes a good graph depends on what you want to illustrate to the reader (Hegarty, 2011). Furthermore, creating a graphic is an iterative process (Cleveland, 1994; Kelly et al., 2005) – you might try a particular approach to create a graph, but may find that a different approach is better once you see the result of the initial attempt. You should expect to modify your graph to make it better, even if you follow all the guidelines in this chapter. Also, guidelines are not rules. The reader should use these guidelines in this chapter as a default approach, but deviations are encouraged if they serve the purpose of a good comprehension of the data you are representing. In future chapters, deviations from the guidelines will be illustrated and justified.

DOI: 10.4324/9781003471264-4

The guidelines

There are two main principles underlying the guidelines: Accuracy of comprehension and ease of processing. Accuracy of comprehension can refer to accuracy in illustration, description, and comparison: Are the data represented and discussed correctly, and does the graph design support appropriate and accurate comparisons? A more difficult aspect of accuracy is taking into account the perception of the graph reader. It is not a useful graph if, despite being plotted technically accurately, it is easily misperceived by the intended audience.

An example of a graph that is easily misinterpreted is the National Hurricane Center's 5-day track forecast cone (see Figure 4.1). If you track hurricane weather, this forecast cone is familiar to you and plots the probable track of the center of a hurricane (https://www.nhc.noaa.gov/aboutcone.shtml). However, this cone is often misperceived as indicating the size and areas the storm will impact (Cox et al., 2013; Ruginski et al., 2016). The size of the hurricane is not indicated in any way by the cone. However, the *current* size of the hurricane and likely area of wind damage are indicated by the extent of tropical storm-force and hurricane-force winds shown in orange and brown on the map in Figure 4.1. Likely coastal effects of the hurricane are indicated by the warnings and watches on the map (red and blue in Figure 4.1). Everything on the graph is plotted and described accurately; however, misinterpretation of this graph is common and can have deadly consequences.[3]

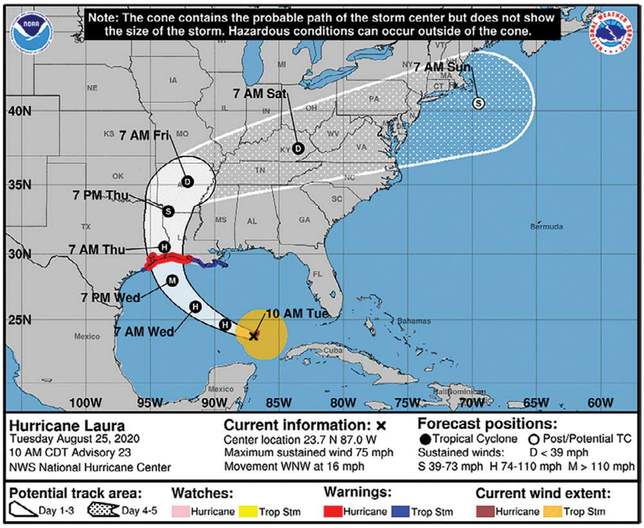


Figure 4.1 National Hurricane Center's easy-to-misperceive 5-day track forecast cone used to communicate risk to the public. Credit: NOAA/NWS

Ease of processing refers to taking steps to make the graph as easy to comprehend as possible. Psychologists would call ease of processing "cognitive load." **Cognitive load** is the amount of mental effort (working memory) required to complete a task (Sweller, 1988). Graphs should be designed to minimize the cognitive load for the task that the graph maker intends for the graph reader. For example, a graph with high cognitive load might require a graph reader to remember the meaning of several colors, several symbols (that are labeled elsewhere in the graph), and the meaning of several abbreviations. The graph reader may need to refer to keys/legends on the graph or even the main text of a paper. This increases the likelihood that important information will be forgotten and therefore the graph either be forgotten completely or misinterpreted. Figure 4.2 illustrates graphs that differ in cognitive load.

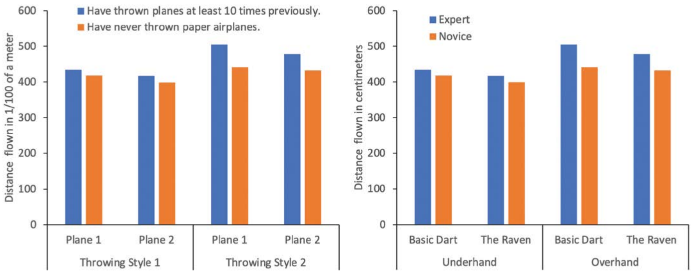


Figure 4.2 Graphs with different cognitive loads. The left panel shows a graph with a high cognitive load. Here, the y-axis is labeled as 1/100 of a meter (in other words, centimeters). The x-axis is labeled with Plane 1 and 2 and Throwing Style 1 and 2. There is no indication on the graph what 1 and 2 mean in either case, so the reader will need to scan the main text that referenced the figure to determine how planes and throwing styles were defined, remember those definitions, and return to reading the graph. The key and data labels are too long to remember easily and interfere with reading the graph. The right panel shows a graph that reduces the cognitive load of the graph in the left panel.

Accuracy of comprehension and ease of processing are the main goals for the guidelines. For ease of later reference, these guidelines are organized by either a design goal (general principles, guiding attention, consistency, etc.) or an aspect of the graph the guideline pertains to (scales, labels, etc.).

General principles

Know your audience

Kelly, Jasperse, and Westbrooke (2005) suggest that the graph maker assume the audience is intelligent. However, Hegarty (2011) and Kosslyn (2006) suggest ensuring the reader has the ability to interpret the graph accurately by taking the time and space to explain the graphic conventions used as necessary. Note that this requires some knowledge of your audience, including what the audience likely knows and likely does not know. For example, Figure 4.1 would need no explanation for a room of meteorologists and the meaning would be interpreted correctly. For science and technology audiences, Cleveland (1994) states that one can assume that the viewer will process and understand tick marks and tick mark labels. Therefore, in most situations, tick marks and labels do not need an explanation for a science and technology audience. Also, different disciplines have different conventions for how data are displayed. A familiar design for your audience will help your audience understand the graph (Franconeri et al., 2021, 2022). In some cases, you may decide to present a graph design that is unfamiliar to your audience based on the recommendations made in this book. I encourage you to do so, as long as you offer your audience the proper support in interpreting and understanding that graph.

Know which graph features support accurate decoding

To represent your data, you must encode its value in your graph and that value then must be decoded by your graph reader. Cleveland and McGill (1984a, 1985, 1986) provide a hierarchy of features (ways of representing the data), from most to least accurately decoded and perceived: 1) Position on a common scale, 2) position on identical nonaligned scales, 3) length, 4) angle or slope, 5) area, 6) volume, density, or color saturation, and 7) color hue. Figure 4.3 illustrates these features. Use the most accurate feature that your data allows. Most commonly in this book, we will be discussing the first three: Position on a common scale, position on identical nonaligned scales, and/or length.

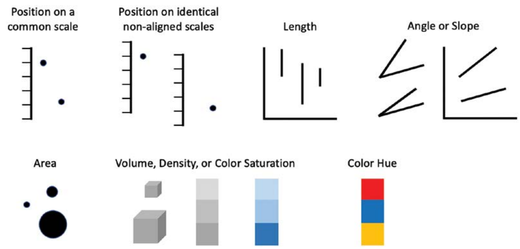


Figure 4.3 Hierarchy of features for representing data. Position on a common scale is the easiest to decode and color hue is the most difficult to decode.

Ensure that all information is present

A graph should have all the components necessary to understand it. This includes fully labeled axes, necessary keys, and data labels, figure titles/captions, clearly discriminable symbols, and tick mark labels (Cleveland, 1994; Franconeri et al., 2021, 2022; Hegarty, 2011; Kelly et al., 2005; Kosslyn, 1989). The top panel of Figure 4.4 shows the importance of including the necessary information.

Present just the relevant information

Hegarty (2011) and Kosslyn (2006) suggest that you present no more or less than the information needed for the point that the graph is meant to illustrate. Figure 4.4 shows an example of graphs that present too little (top panel), too much (middle panel), and the right amount of information (bottom panel). It is important to note that this is not a recommendation to alter the data set; altering the data set by definition results in an inaccurate representation of the data.[4] Finally, design the graph to support perceptual accuracy across a wide range of tasks (Franconeri et al., 2021, 2022). The reader of your graph may visually investigate questions beyond your intended point, and your graph should support accurate perception in these additional tasks.

Focus on the data

Your data should be the most important and noticeable feature of your graph; Do not distract from the data with irrelevant information or images (Kelly et al., 2005). Avoid chartjunk – for example, background images that serve no purpose or cartoons to represent the data rather than bars or other data symbols (Cleveland, 1994). Maximize the data-ink ratio – as much as reasonable, make sure the ink on your graph is about the data (data symbols), or ink necessary to interpret the data (scale lines, tick marks, etc.) (Hegarty, 2011; Tufte, 2001). Unnecessarily three-dimensional bars are chartjunk and do not maximize the data-ink ratio (see Figure 4.5). Decluttering a graph refers to removing irrelevant features from your graph that appear by

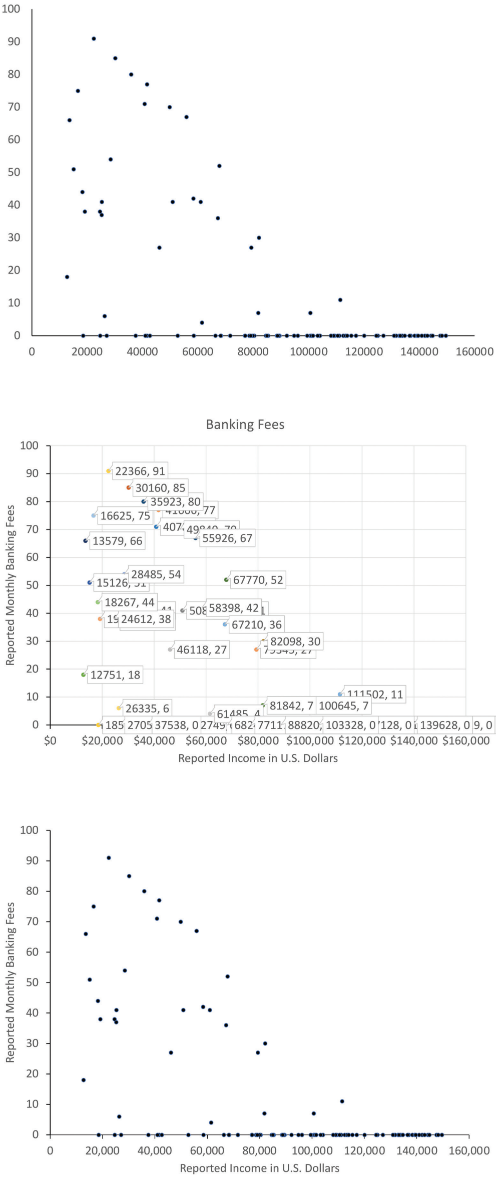


Figure 4.4 Three graphs of the same data. In the top panel, there is not enough information to know what the graph is about. The axes have numbers, but are otherwise unlabeled. No tick marks are given, so it is difficult to know where to connect the numbers to the axes. In the middle panel, the axes are labeled and tick marks are given. On the x-axis, the numbers do not need a "$" given the axis is labeled "Reported Income in U. S. Dollars." The graph title "Banking Fees" is unnecessary (it is redundant with the y-axis label). The light gray gridlines are unnecessary. Perhaps most egregiously, each data point is labeled with a callout giving its x and y values (if exact values are important, perhaps a table would be better – see Chapter 2). The data callouts make it impossible to see most of the data and most of the x-axis. Finally, each data point is a different color for no apparent reason. In the bottom panel, the excesses of the middle panel are removed and now the data can be seen and interpreted more easily. Many people do not pay monthly banking fees, but those who do pay such fees tend to have lower reported income.

default in many graphing programs. Decluttering can include removing background images or lines, removing color that is not used to encode data, removing boxes around the frame of the graph, and other non-data ink. Figure 4.4 (middle to bottom panel) shows the effect of removing background lines and color that is not used to encode data. Decluttering improves the perceived quality of the graph (Ajani et al., 2022).

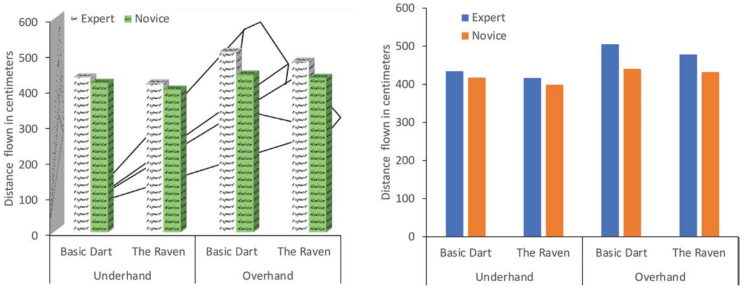


Figure 4.5 The effect of chartjunk on a graph. In the left panel, the three-dimensional bars make determining the distance flown more difficult – the bars' tops are ambiguous and their interpretation depends on how far back in depth you perceive the bars to be. Having the words "expert" and "novice" repeated within the bars is somewhat distracting. The paper airplane image in the background does not inform the graph reader in any way, and could be confusing if the graph reader recognizes that it is an image of the "Basic Dart" paper airplane (but "The Raven" does not appear anywhere in the background). Under duress, I would admit that the graph in the left panel is interesting to look at. However, the interest actually detracts from attending to the actual data. In the right panel, we have removed the distracting chartjunk. Under duress, I would admit that the graph in the right panel is less interesting to look at – all I can see are the data – but only seeing the data is an advantage if the goal is to have me understand the data. Visually interesting is not the same thing as visually or cognitively useful.

Design for accessibility

In addition to using clearly discriminable data symbols, strive to account for potential difficulties individuals might have reading your graph. For example, make the text large enough to be easily readable by all, taking into account the presentation method (e.g., a printed manuscript vs. a projected oral presentation in a large room). About 5% of the population is red-green color deficient (Neitz & Neitz, 2011). Do not use red versus green as an indicator of a difference in category. Figure 4.6 simulates what a graph using red and green bars could look like to those people with red-green color deficiency.[5] If you are using color in your graph, you can simulate how those with different types of color deficiency would see your image by using the smartphone application available here: https://asada.website/cvsimulator/e/index.html. I used this smartphone application to guide my creation of Figure 4.6.

Scales and ticks

Know the standard shape for a graph

The scale-line rectangle of a graph is usually a little wider than tall, with the data rectangle slightly smaller than the scale-line rectangle (Cleveland, 1994; Kelly et al., 2005). Figure 4.7 shows examples of the importance of this guideline. There are situations where the standard-shaped graph is not ideal; in later chapters, those exceptions will be discussed.

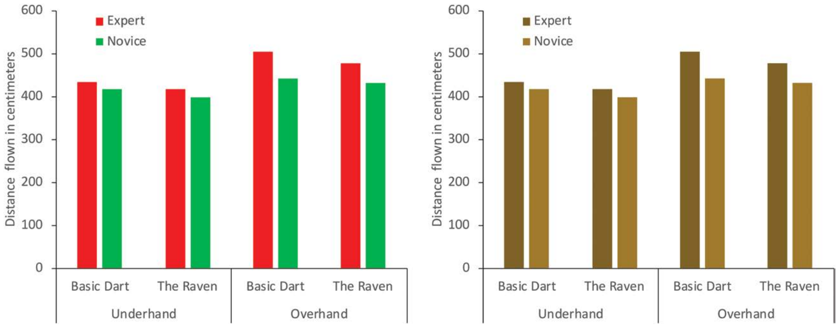


Figure 4.6 Designing accessible graphs. The graph in the left panel is perceived as having vividly distinctive colored bars (red and green) for about 95% of people. Those with red-green color deficiency (about 5% of people) perceive the graph in the left panel similar to the graph shown in the right panel. There are two types of red-green color deficiency; which type an individual has determines whether the left or right bar in each pair appears as the darker shade of brown. Note that in the earlier blue-orange version of this graph, the image would be perceived as blue and brown by someone with red-green color deficiency, which still allows the categories to be easily distinguished.

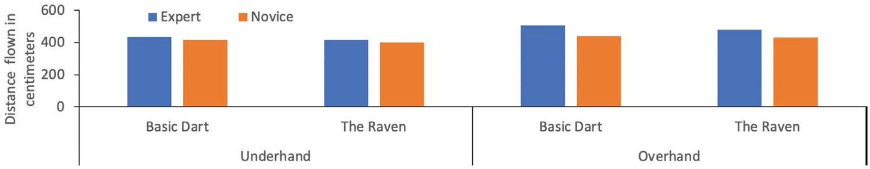


Figure 4.7 An unusually wide and short graph. There is nothing drastically wrong per se with this graph; it is just not the usual shape. Note that the data are the same as the data in the graph in the right panel of Figure 4.5. You might note that the shape in Figure 4.7 does make the difference between expert and novice appear to be smaller. This shape is allowable if you have a good reason; we will discuss some possible good reasons in later chapters.

Ensure statistical differences look perceptually different

Statistical differences, for example, a statistically significant result of an independent measures t-test, should appear to be different in your graph (Hegarty, 2011; Kosslyn, 2006). This may require careful decisions about the scales of the axes, see Figure 4.8.

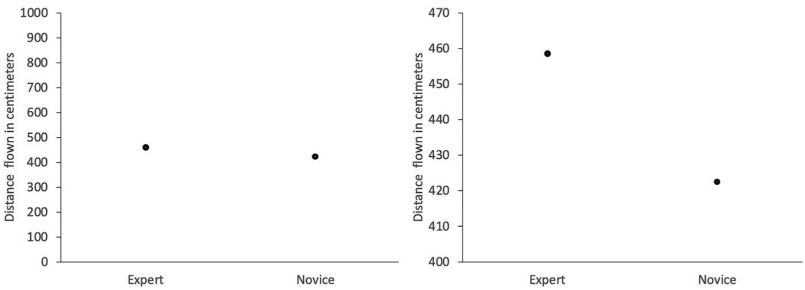


Figure 4.8 Choices made about the range of the scale can influence whether a difference is apparent in a graph. Even though the data are the same, the graph in the left panel seems to suggest that there is no difference between expert and novice throwing distances, while the graph in the right panel suggests that there is a difference. Note that the left panel also violates the previous guideline that the data rectangle is just slightly smaller than the scale-line rectangle.

Use (four) scale lines effectively

Cleveland (1994) suggests using four scale lines encompassing the data rectangle on the left, right, above, and below. Having four scale lines allows for more accurate estimation of the value of data symbols. When necessary, a reference line can be added to the graph when there is a particularly important scale value that serves as a comparison for the data symbols. For example, imagine an experiment where an observer has to say which of four lights is brightest. Three of the four lights are 100 lumens, and a fourth is some variable amount of lumens. Figure 4.9 plots the hypothetical results. Cleveland (1994) also suggests that it can be useful to pair scale lines to show different scales, for example, a log y-axis scale on the right side and a y-axis scale in the original measurement values on the left side of the graph, allowing the reader two options for how to read the data, see Figure 4.10.

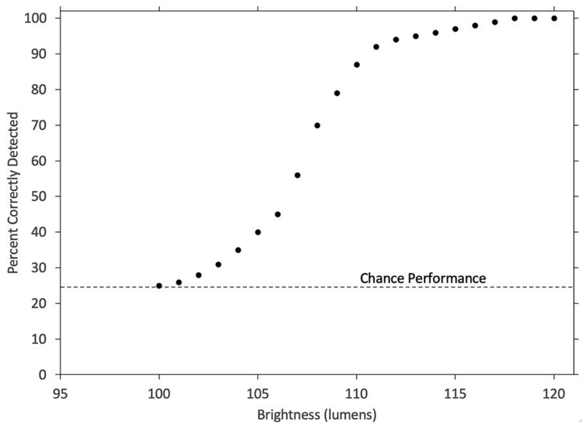


Figure 4.9 Graph illustrating the use of four scale lines and a reference line. It is not necessary to label both x-axes or both y-axes. The top and right-hand scales are present to provide the tick marks that help the reader judge the approximate value of data symbols, particularly those data symbols nearer those scales. Given there are four choices in the hypothetical experiment, chance performance is 25% correct. A chance performance line is added to help make it clear to the graph reader why there is no data below 25% correct. Note that the chance performance line is dashed to distinguish it from scale lines and is labeled directly on the graph. An astute reader might note that the data rectangle of this graph could be expanded by adjusting the scale ranges.

Use a sufficient but limited number of tick marks

Approximately four to six tick marks are sufficient for an axis to support data symbol table look-up (Kelly et al., 2005). Three (nonzero) tick marks are required to determine whether a scale is linear or logarithmic (Cleveland, 1994). Figure 4.10 illustrates the use of slightly more than six tick marks in a graph. In the right panel of Figure 4.10, the left axis shows why at least three tick marks are required for determining if the scale is linear or logarithmic. For example, if just ticks for 128 and 256 were present, it would be ambiguous whether it was linear or logarithmic. If the next tick higher on the scale is 384, then the scale is linear; if the next tick is 512, then the scale is logarithmic.