

## 2.2: Using and Understanding Graphs

### Learning Objectives

- Assess graphs for information and quality

### Introduction to Graphs

**Florence Nightingale (1820 – 1910)**, although primarily known as a nurse, analyzed data in the service of her patients. She was the first woman to be a fellow of the **Royal Statistical Society**. She was one of the first nurses to use graphical representations to illustrate the causes of mortality. She created the graph below to share the true cause of British soldier mortality during the Crimean War clearly.

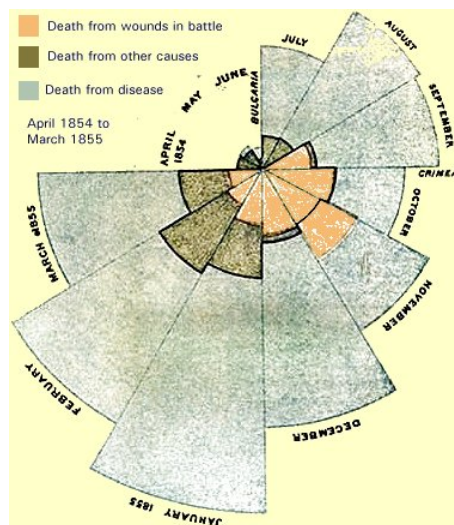


Figure 2.2.1: Polar-area diagram for British soldier mortality from April 1854 through March 1855

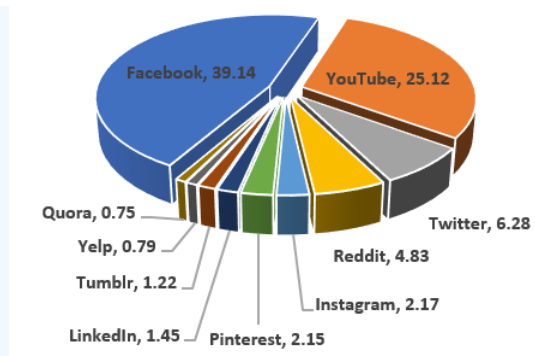
From the graph, we know that causes of mortality are broken into three classifications: wounds in battle, disease, and other causes. We can conclude relatively quickly that being wounded in battle was not the leading cause of death; disease caused most of the deaths. When Nightingale arrived, the conditions of the military hospitals were awful, and more people died from disease than their wounds. Once her sanitation practices were implemented, the mortality rate dropped from 42.7% to 2.2%. The question we must address is: how do we trust that the graph relays the truth?

We were not provided any information about the construction of the graph; please be careful of graphs without enough explanation. We guessed that the number of mortalities was related to the slice size for any given month. This graph has two components to size: the radius and the area. By which measure are we making a comparison? Most of us naturally make mental comparisons based on the area of each slice. If the graph were constructed using the radius to indicate the number of mortalities, we could still tell which months had more deaths, but we would misjudge the magnitude of these differences. Nightingale constructed the graph so that the areas represented the amount of mortalities; this information is necessary for quality interpretation and understanding.

Graphs and charts are excellent ways to share information quickly with a larger audience. In this section, we will look at several different types of graphs. However, our goal is to identify certain types of graphs to become informed consumers of information and critical thinkers actively engaged in the world around us.

### ? Text Exercise 2.2.1

The **pie graph below** shows the percentage of visits to social media sites in 2017. There are various issues with this particular pie graph; identify some of the problems.



### Answer

Pie graphs effectively display the relative frequencies of a small number of categories. However, pie graphs with a large number of categories are not recommended. This example uses too many categories.

Pie graphs are most commonly used to compare parts to the whole. This example has two components working against this comparison. The pie slices are not touching and the visual comparison of areas is skewed because of the three-dimensional formatting. When adding the third dimension to a graph, it must add information and not hide important information.

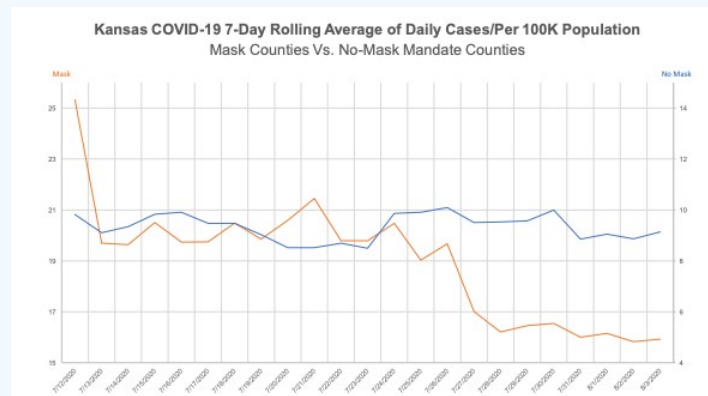
Furthermore, the numbers here do not add to 100. This means that the number given does not represent the relative size of the slice of the pie. For example, Facebook appears to take up nearly half of the pie, but the number given is 39.14. This makes it seem as if the websites listed here account for all the websites that see a relatively non-negligible amount of visits. There should be another slice, perhaps labeled "other," which takes up 16.1% of the pie.

**Pie graphs** are most helpful in communicating relative sizes or proportions of a few categories. They are not useful if there are numerous categories or if one wants to communicate absolute, not relative, measures. Note that relative measures are more likely to be misleading when the sample size is small.

A **time series graph** consists of the measurement of the same variable of the same subject taken over regular time intervals for a given period. Time series data frequently occurs in contexts where variables change: stock market prices, national economic figures, population tracking, etc. During the COVID-19 pandemic, time series graphs were frequently used to communicate transmission and mortality rates. Such graphs can serve as natural and valuable visual aids, but they can be misleading if not constructed appropriately.

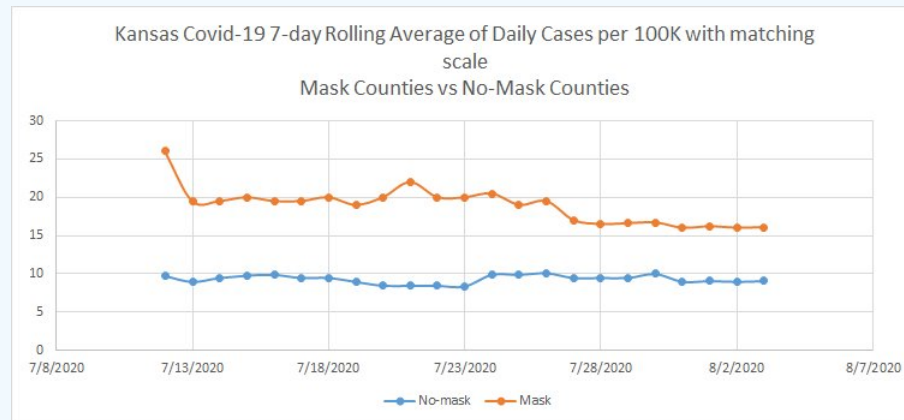
### ? Text Exercise 2.2.2

In August 2020, Dr. Norman, the Secretary of the Kansas Department of Health and Environment, used the [following graph](#) to show the number of COVID-19 cases in counties with mask mandates vs counties without mask mandates. Study the graph and see what conclusions can be made. Are there any issues?



### Answer

First, notice that the scales on each side of the graph represent the same information, but the numbers differ. The line representing 25 cases for masked counties (scale on the left) corresponds to 14 cases for unmasked counties (scale on the right). This 11 point difference is consistent throughout the scale. Without careful attention, a consumer might conclude that the 7-day rolling averages for masked counties dropped below those of the unmasked counties. This, however, is not the case. Consider the following graph, where the same data is plotted using the same axes.



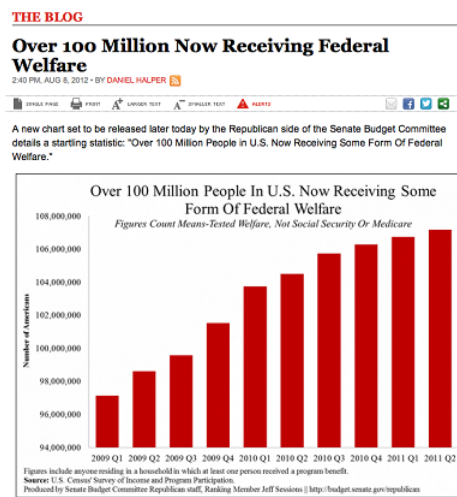
We now see clearly that the masked county 7-day rolling averages stay above those of the unmasked counties. There can be a variety of factors at play in this difference. While the 7-day rolling averages should be scaled to account for differences in population, other related factors could also be at play. Smaller and rural counties were less prone to mask mandates. Larger and more densely populated counties were more likely to have mask mandates. These differences alone could explain the differences in the 7-day rolling averages.

The changes in the 7-day rolling averages could be a better measure for the success of the mandates, but notice how the inclusion of 0 on the vertical scale helps us better gauge the change over time. The variation in the 7-day rolling averages seems less dramatic in the second graph than in the original. The no-mask counties seem to hover around 9 or 10 while the mask counties seem to hover between 16 and 19. The first data point makes the change seem quite stark, but a change in a 7-day rolling average indicates that the first measurement was significantly higher than the subsequent days. Would one day of wearing masks have such an immediate effect? Possibly. Such a question warrants further analysis, preferably with each day's raw data.

We have encountered several **bar graphs** in the previous section as we studied relative frequency and frequency distributions. While pie graphs help make comparisons from a part to the whole, bar charts help make comparisons between parts and across different distributions.

### ? Text Exercise 2.2.3

The 2009 – 2011 data from the U.S. Census Survey of Income and Program Participants was released in USA Today in 2012, and the bar graph below is used. At first glance, the information on the increased number of people on welfare over two years is staggering. What else do you notice about this graph?



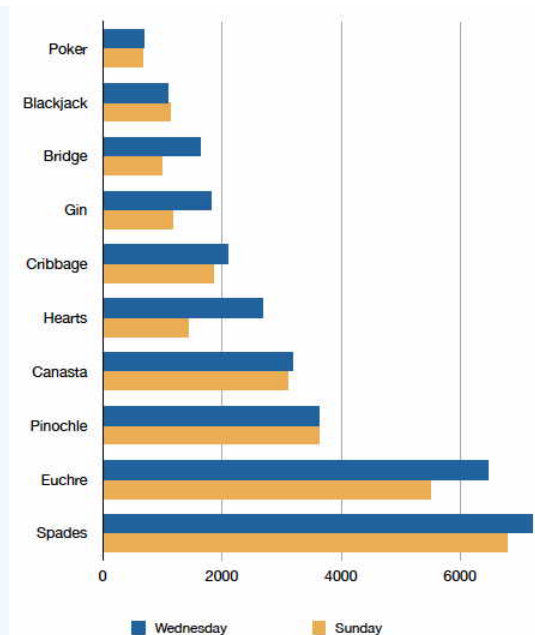
## Answer

Although having 108million people on welfare is not good, we must recognize in this particular graph that the vertical axis begins at 94 million, making the increase look more extreme than it is. Be sure to check the vertical axis to see where it begins. We naturally make mental length comparisons and ratios when looking at bar charts, but recall that ratios are only meaningful if we have a meaningful zero value. When bar charts do not start at a meaningful zero value, like in the graph above, the mental ratio comparisons have no meaning.

We remark here that bar charts displaying ratio data should begin at 0. Unless the reason for beginning at some nonzero value is highlighted and explained, such a chart is more likely to be misleading than informative. We should notice if we are ever shown a bar chart that does not start at 0, as it typically means the differences between the heights are being exaggerated. Even starting at 0 with non-ratio data is misleading, as multiplicative relationships are inherently meaningless. For this reason, we recommend avoiding bar charts for non-ratio data.

## ? Text Exercise 2.2.4

Consider the graph below that shows the number of people playing card games on the Yahoo website on a Sunday and Wednesday in the Spring of 2001. Here, we have two distributions of game frequencies based on the day. Bar graphs are helpful when comparing distributions taking on the same classes. What can we conclude from such a graph?



### Answer

The number of people playing Pinochle was the same on these two days. In contrast, about twice as many people were playing hearts on Wednesday as on Sunday. Blackjack was the only game with more players on Sunday than on Wednesday. Facts like these emerge clearly from a well-designed bar graph. The bars are oriented horizontally rather than vertically. The horizontal format is recommended when you have many categories because there is more room for the category labels.

We can also conclude that there were more players overall on Wednesday than on Sunday. This is because Blackjack was the only game with more players on Sunday than Wednesday, and the margin wasn't large enough to offset the rest. Making conclusions about all the categories collectively is not always easy because bar charts are most conducive to comparing parts to parts.

### ? Text Exercise 2.2.5

Graphs are pictorial representations of numbers. The following graph, originally from [Erickson Times](#), shows the number of medals per country in the Summer Olympics. Take a minute to observe the graph and see why the pictures in this particular graph may be misleading.



### Answer

We naturally expect the representation of the numbers to be proportional to the actual numbers. When looking at Germany's medal count, we see two medals on the graph equal to about 500 medals; we reason that each medal pictured would equal about 250 medals. France only has 523 medals with three units on the graph, and Russia has 999 medals with five units on the graph. The numerical order is correct, but there is no consistent, intuitive correspondence between the units on the graph and the actual number of medals. When using pictures, a graph may look fancy, innovative, or visually appealing but can render the graph misleading and ineffective. The purpose of graphical representations is to disseminate information clearly and quickly. Be cautious.

2.2: Using and Understanding Graphs is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by The Math Department at Fort Hays State University.

- 2.1: Graphing Qualitative Variables by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.