

2.12: Summary and Some Honesty

Intentional Lies or Accidental Misleading?

It's easy to mislead with graphs. Some of the fancy graphics make it more difficult to see the exact numbers, but there are many ways to be misled with graphs. To avoid being fooled by a fancy chart you should also check out the axes. In a frequency distribution, the y-axis should generally be one or zero. This [article "Lying with Graphs"](http://chmullig.com/2011/04/lying-with-graphs/) (website address: <http://chmullig.com/2011/04/lying-with-graphs/>) by Chris Mulligan in 2011 shows an example of both a confusing type of chart, and mis-leading axes (plus, he didn't think that the original article calculated the actual numbers correctly, either); always be vigilant!

Example of Misleading

We will use data from interviews of computer users to show how graphs in general, but especially bar graphs, can be misleading. Here's the research scenario: When Apple Computer introduced the iMac computer in August 1998, the company wanted to learn whether the iMac was expanding Apple's market share. Was the iMac just attracting previous Macintosh owners? Was it purchased by newcomers to the computer market? Were previous Windows users switching over? To find out, 500 iMac customers were interviewed. Each customer was categorized as a previous Macintosh owner, a previous Windows owner, or a new computer purchaser.

Figure 2.12.1 shows that three-dimensional bars are usually not as effective as two-dimensional bars because it's unclear where the top of the bar is.

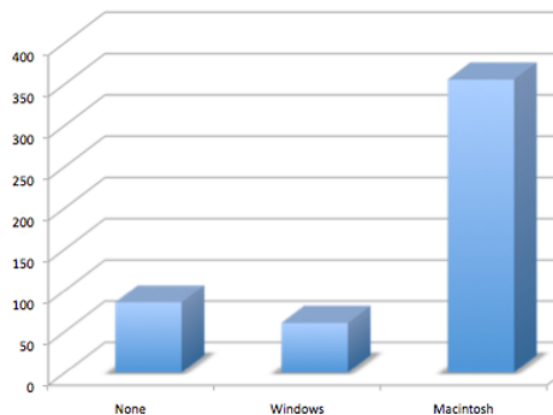


Figure 2.12.1: A three-dimensional bar chart (CC-BY-NC-SA [Foster et al.](#) from [An Introduction to Psychological Statistics](#))

Here is another way that fanciness can lead to trouble. Instead of plain bars, it is tempting to substitute meaningful images. For example, Figure 2.12.2 presents iMac data using pictures of computers. The heights of the pictures accurately represent the number of buyers, yet Figure 2.12.2 is misleading because the viewer's attention will be captured by the area (width) of the computers, not just the height. The areas can exaggerate the size differences between the groups. In terms of percentages, the ratio of previous Macintosh owners to previous Windows owners is about 6 to 1. But the ratio of the two areas in Figure 2.12.2 is about 35 to 1. A biased person wishing to hide the fact that many Windows owners purchased iMacs would be tempted to use Figure \(\PageIndex{2}\)! Edward Tufte coined the term "lie factor" to refer to the ratio of the size of the effect shown in a graph to the size of the effect shown in the data. He suggests that lie factors greater than 1.05 or less than 0.95 produce unacceptable distortion.

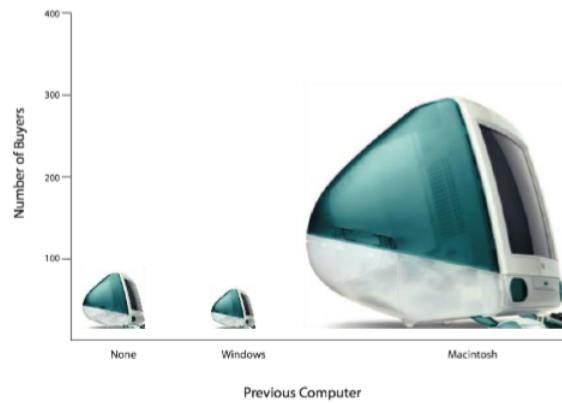


Figure 2.12.2: Example with a lie factor greater than 8. (CC-BY-NC-SA Foster et al. from [An Introduction to Psychological Statistics](#))

Another distortion in bar charts results from setting the bottom of the y-axis to a something other than zero. The bottom of the Y-axis should represent the least number of cases that could have occurred in a category. Normally, this number should be zero. Figure 2.12.3 shows the iMac data with a baseline of 50. Once again, the differences in areas suggests a different story than the true differences in percentages. The number of Windows-switchers seems minuscule compared to its true value of 12%.

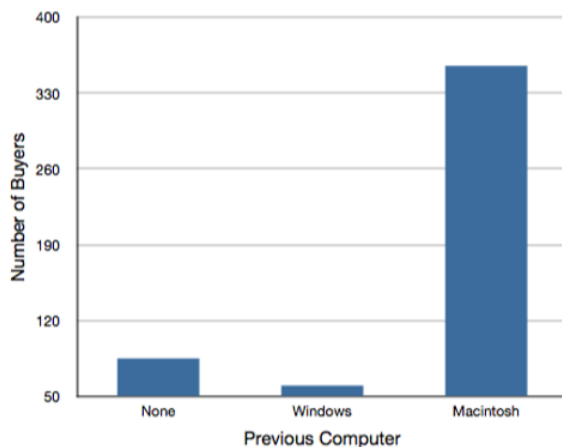


Figure 2.12.3: Example with a baseline of 50. (CC-BY-NC-SA Foster et al. from [An Introduction to Psychological Statistics](#))

Note

Why don't you use this iMac data to play around and create your own bar chart by hand or with computer software, without making it misleading?

Examples of Misleading, or Lies?

✓ Example 2.12.1

Go through the examples from this 2014 BuzzFeed report by Katie Notopoulos (<https://www.buzzfeednews.com/article/katienotopoulos/graphs-that-lied-to-us>), and decide one thing that is wrong with each chart.

1. The time an upside down y-axis made "Stand Your Ground" seem much more reasonable.
2. The time 7 million was 5x more than 6 million.
3. The Governor race where one guy's 37% was WAY more than just 37%
4. This bar graph that shows the devastating drop in this pitcher's speed after one year.
5. The time when Scotland really gave 110%.

6. This poll which gives about the same infuriating response as when you ask someone what they want for dinner.
7. The pollster who really doesn't want people to think A levels are getting harder.
8. This graph that measure..... units of "innovation"?
9. The number skipping y-axis on this totally legit graph.
10. This graph showing the astonishing growth rate to .06 hawks?
11. Whatever happened in this school paper.
12. The graph where last year, last week, and today are equally far apart.
13. This chart showing the giant gulf between 35% and 39.6%.

Solution

1. Check out the y-axis. Instead of zero on the bottom, with the higher numbers being higher, the y-axis is “upside down” from what is typical.
2. Check out the y-axis. Not only is it not labeled, but if the y-axis had started with zero, the 1 million difference would have looked like much.
3. The two 37%'s probably look difference chart doesn't show decimal points, but if the y-axis had started with zero, the two 37%'s would be indistinguishable.
4. It's the missing y-axis problem again, and not starting the y-axis at zero.
5. There's nothing technically wrong with these numbers, but you should always have someone else check your math.
6. There's nothing technically wrong with these numbers, but it is a reminder that designing survey questions is harder than it seems!
7. This is an appropriate way to use a pie chart (showing percentages that add up to close to 100%), but the pie slices need to represent the percentage, so the blue should only be 49% of the pie. It looks like 60% or about two-thirds...
8. It's important to label your y-axis with units that make sense!
9. The y-axis should have equal intervals, meaning that the difference between the first number and the second number (2 points in this example, from 1 to 3) is the same as the difference between the highest number and the second highest number (which is $1803 - 516 = 1287$ in this example).
10. There is nothing technically wrong with this chart, but even going from zero to 1.0 would have seemed more accurate when dealing with such small numbers on the y-axis.
11. The colors in the pie charts should take up the portion of the pie that the numbers show. Also, pie charts show 100% of the sample, but $26\% + 26\% + 26\% + 26\% = 104\%$, so someone needed a friend to check their math here, as well.
12. Here is an example of a line graph showing changes through time, except that the x-axis doesn't have equal intervals. It should probably have been a bar chart showing the three time categories, rather than trying to show a change through time.
13. Another example of a missing y-axis that is misleading because the missing y-axis doesn't start at zero.

You can also read this great 2014 blog by Ravi Parikh of Heap showing some of the same examples of graphs that seem to be trying to intentionally mislead readers, and an explanation about what is inappropriate or missing: <https://heap.io/blog/data-stories/how-to-lie-with-data-visualization>

Be careful to avoid misleading your audience, or being misled!

Graphs Summary

You've learned about many types of graphs, and why they should be used. You also should have practiced interpreting what they can tell you.

✓ Example 2.12.1

Play around with this [interactive website](https://www.zoology.ubc.ca/~whitlock/Kingfisher/SamplingNormal.htm) (website address: <https://www.zoology.ubc.ca/~whitlock/Kingfisher/SamplingNormal.htm>), then answer what kind of graphs they are.

Solution

Although both have lines to show a Normal Distribution, the actual graphs of the data are histograms because they are bars that touch, and the bars show the frequency of different quantitative variables.

Now that you've learned so much about charts and graphs, go through these [charts from the New York Times](https://www.nytimes.com/column/whats-going-on-in-this-graph) (website address: <https://www.nytimes.com/column/whats-going-on-in-this-graph>) and determine if the x-axis is showing qualitative or quantitative variables and if you can describe the shape of any frequency distributions.

Contributors and Attributions

- [Foster et al.](#) (University of Missouri-St. Louis, Rice University, & University of Houston, Downtown Campus)
- [Dr. MO](#) (Taft College)

This page titled [2.12: Summary and Some Honesty](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Foster et al.](#) (University of Missouri's Affordable and Open Access Educational Resources Initiative) via [source content](#) that was edited to the style and standards of the LibreTexts platform.