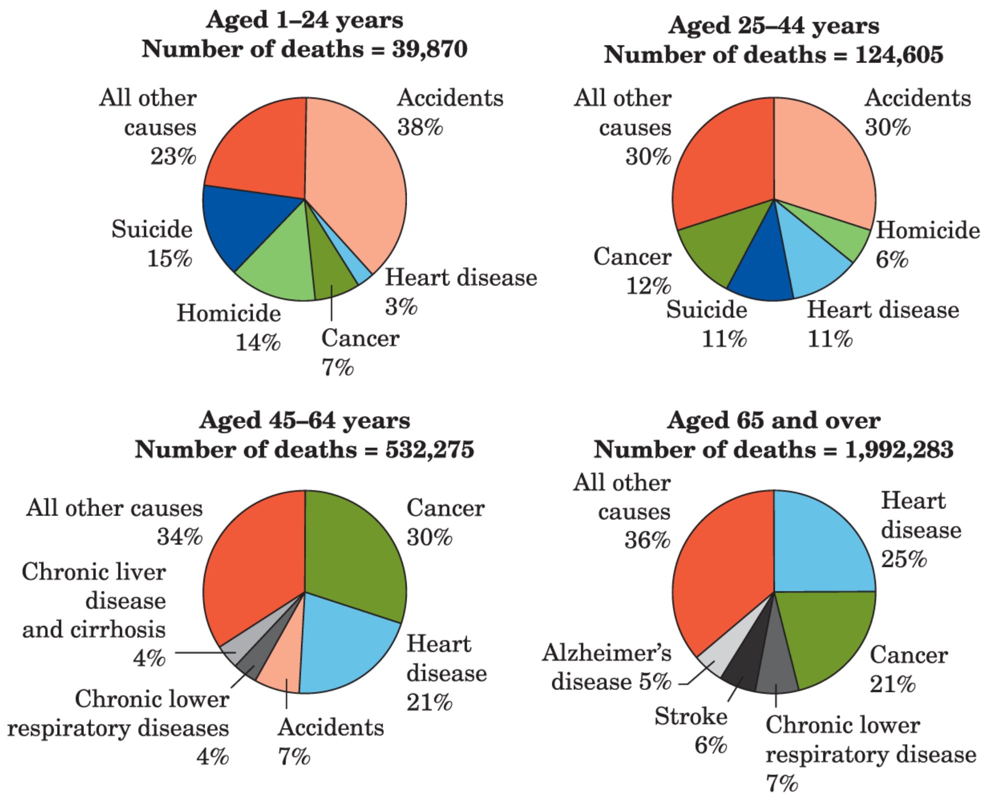
Chapter 10

Graphs, Good and Bad

Lecture Slides

Case Study: Graphs, Good and Bad 1

In its brief, "Deaths: Final Data for 2015," the National Center for Health Statistics (NCHS) produced the following figure:



Case Study:
Graphs, Good and Bad 2

A clear graphical display can make it easy for policy and decision makers to decide how to allocate resources for prevention or intervention.

A quick glance at the pie chart for ages 1–24 shows that a large portion, 15% of deaths in that age range, are attributed to suicide.

A clear display of this staggering number makes a strong argument for allocation of resources.

Case Study: Graphs, Good and Bad 3

Tables and graphs help us see what the data say.

Not all tables and graphs do so accurately or clearly.

In this chapter you will learn some basic methods for displaying data and how to assess the quality of the graphics you see in the media.

By the end of the chapter you will be able to determine whether the side-by-side pie charts make a good or a bad graphic.

Data Tables 1

Has the number of private elementary and secondary schools grown over time? What about minority enrollments in these schools? How many college degrees were given in each of the past several years, and how were these degrees divided among fields of study and by the age, race, and sex of the students?

You can find all this and more in the education section of the *Statistical Abstract*. The tables summarize data.

We don’t want to see information on every college degree individually, only the counts in categories of interest to us.

Example: What makes a clear table? 1

What makes a clear table?

The following table presents education data for people aged 25 years and over.

Table 10.1 Education of people aged 25 years and over in 2017

Level of education	Number of persons (thousands)	Percent
Less than high school	22,540	10.4
High school graduate	62,512	28.8
Some college, no degree	35,455	16.4
Associate's degree	22,310	10.3
Bachelor's degree	46,262	21.3
Advanced degree	27,841	12.8
Total	216,921	100.0

Data from Census Bureau, *Educational Attainment in the United States: 2017*.

Example: What makes a clear table? 2

It is clearly labeled so that we can see the subject of the data at once.

The source of the data appears at the foot of the table.

Table 10.1 Education of people aged 25 years and over in 2017

Level of education	**Number of persons (thousands)**	**Percent**
Less than high school	22,540	10.4
High school graduate	62,512	28.8
Some college, no degree	35,455	16.4
Associate's degree	22,310	10.3
Bachelor's degree	46,262	21.3
Advanced degree	27,841	12.8
Total	216,921	100.0

Data from Census Bureau, *Educational Attainment in the United States: 2017*.

Data Tables 2

The table from the previous example gives counts and rates.

Rates (percentages or proportions) are often clearer than counts—it is more helpful to hear that 10.4% of this age group did not finish high school than to hear that there are 22,540,000 such people.

The last two columns of the table present the distribution of the variable “level of education” in two alternate forms. The **distribution** of a variable tells us what values it takes and how often it takes these values.

Example: Roundoff errors 1

Roundoff errors: recall the table from the previous example.

The total number of people should be 22,540 + 62,512 + 35,455 + 22,310 + 46,262 + 27,841 = 216,920 (thousands)

The table gives the total as 216,921. What happened?

Example: Roundoff errors 2

Each entry is rounded to the nearest thousand. The rounded entries don't quite add to the total, which is rounded separately. Such roundoff errors will be with us from now on as we do more arithmetic.

It is not uncommon to see roundoff errors in tables.

For example, when table entries are percentages or proportions, the total may sum to a value slightly different from 100% or 1, often to 99.9 or 100.1%.

Types of Variables

When we think about graphs, it is helpful to distinguish between variables that place individuals into categories (such as gender, occupation, or education level) and those whose values have a meaningful numerical scale (such as height in centimeters or SAT scores).

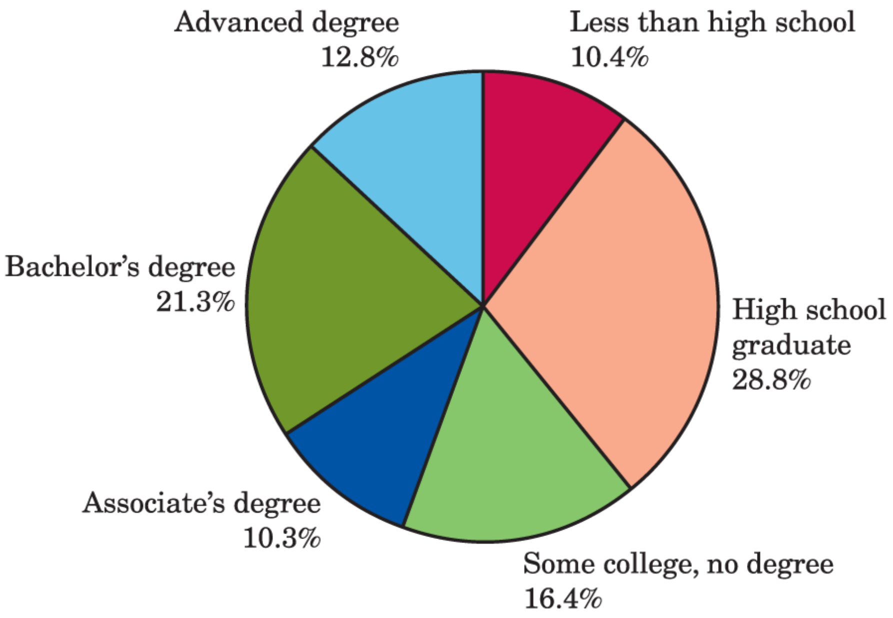
A **categorical variable** places an individual into one of several groups or categories.

A **quantitative variable** takes numerical values for which arithmetic operations such as adding and averaging make sense.

Pie Charts and Bar Graphs 1

"Level of education" is a categorical variable. There are six possible values of the variable from our previous example.

To picture this distribution in a graph, we might use a pie chart.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Pie Charts and Bar Graphs 2

Pie charts show how a whole is divided into parts.

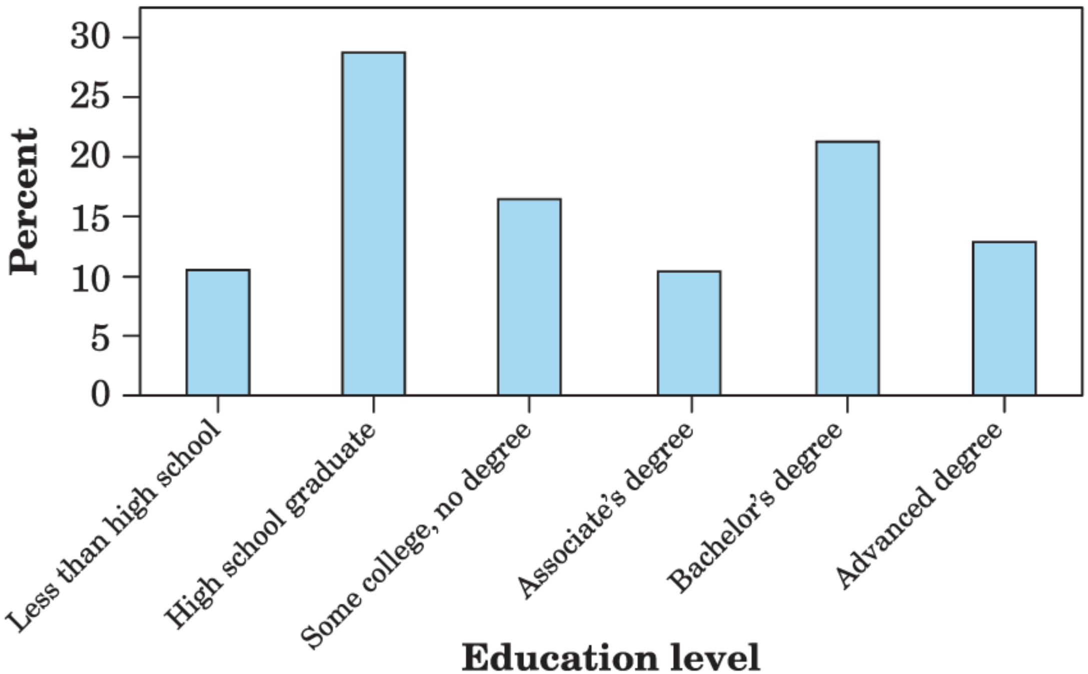
To make a pie chart, draw a circle. The circle represents the whole.

Wedges within the circle represent the parts, with the angle spanned by each wedge in proportion to the size of that part.

Pie charts force us to see that the parts do make a whole. However, it is much easier for our eyes to compare the heights of the bars on a bar graph than it is to compare the size of angles on a pie chart.

Pie Charts and Bar Graphs 3

The figure below is a bar graph of the same “education level” data.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Pie Charts and Bar Graphs 4

A bar graph displays a single number representing each category.

The height of each bar will represent that number.

On a bar graph, each bar has the same width. This is always the case with a bar graph. Also, there is a space between the bars.

The bars on a bar graph can be vertical or horizontal.

Pie Charts and Bar Graphs 5

Bar graphs are better for making comparisons of the sizes of categories.

In addition, if there is a natural ordering of the variable, such as how much education a person has, this order can be displayed along the horizontal axis of the bar graph but cannot be displayed in an obvious way in a pie chart.

Pie Charts and Bar Graphs 6

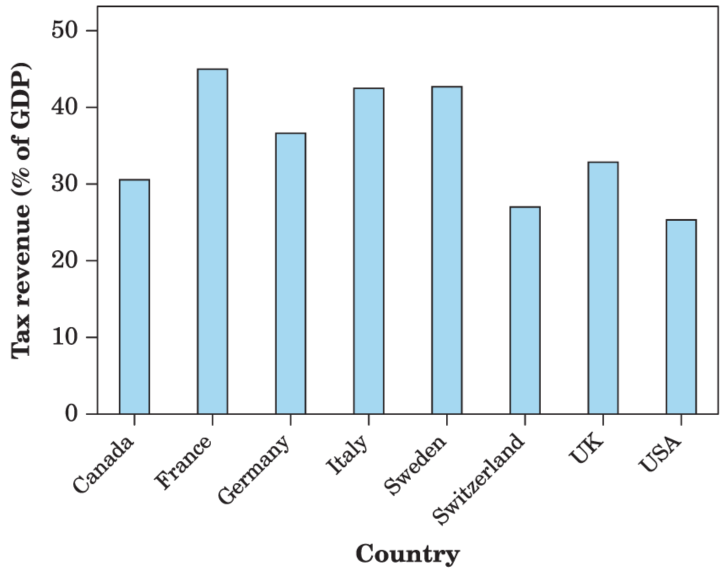
Pie charts and bar graphs can both show the distribution (either counts or percentages) of a categorical variable such as level of education.

A pie chart usually displays the percentage for each category (rather than the count) and only works if you have all the categories (the percentages add to 100).

A bar graph can display an entire distribution or only a few categories. A bar graph can also compare the size of categories that are not parts of one whole. If you have one number to represent each category, you can use a bar graph.

Example: High taxes? 1

Level of taxation in eight democratic nations by gross domestic product that is taken in taxes.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Notice that a pie chart is not possible for these data since we are displaying eight separate quantities, not the parts of a whole.

Example: High taxes? 2

In its 2014 report, the Tax Foundation analyzed data on tax revenue for 2011 from the Organization for Economic Cooperation and Development (OECD).

The U.S. tax revenue is comprised of individual income tax (37.1%), social insurance tax (22.8%), consumption tax (18.3%), property tax (12.4%), and corporate income tax (9.4%).

A bar graph is appropriate to display these data since we have one value to explain the size of each tax category. Notice a pie graph is appropriate as well since these are all parts of a whole.

Example: Government tax revenue breakdown 1

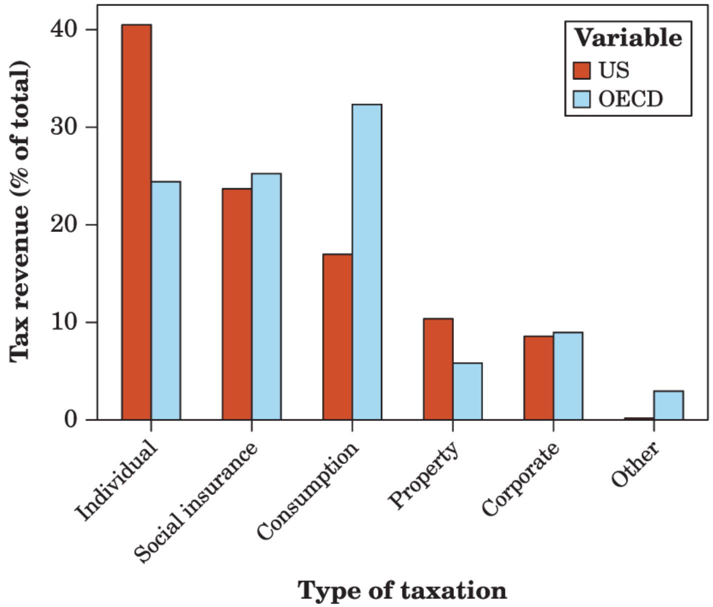
What if we wanted to compare the distribution of tax revenue for the United States to the average for all the other OECD countries?

Pie charts are not good for comparisons.

We can create a bar graph for the distribution of government tax revenue for the United States (USA) with a second set of bars representing the average for all other OECD nations adjacent to the bars for the United States (USA).

This will be called a side-by-side bar graph.

Example: Government tax revenue breakdown 2



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

A side-by-side bar graph is useful for making comparisons. It is now clear that the United States relies more heavily on individual income tax, while the other countries rely more heavily on consumption taxes.

Beware the Pictogram

Bar graphs compare several quantities by means of the differing heights of bars that represent the quantities.

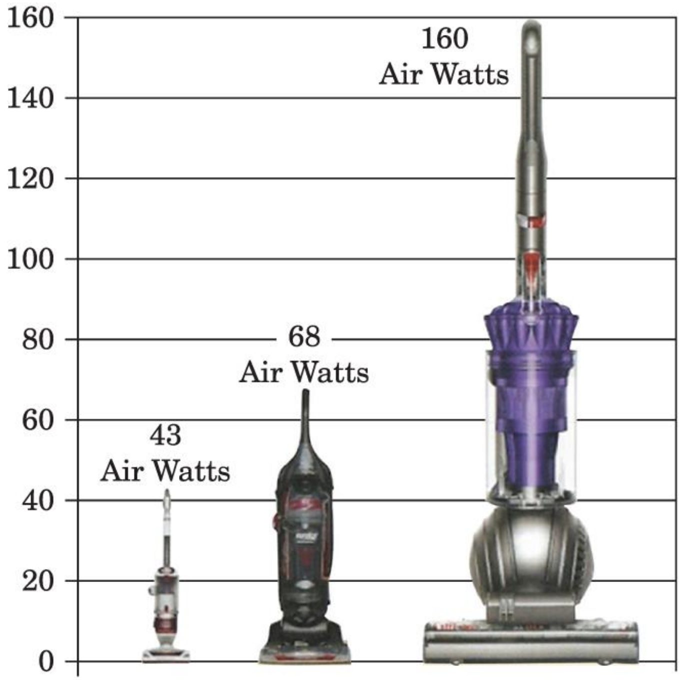
When all bars have the same width, the area (width × height) varies in proportion to the height and our eyes receive the right impression. When you draw a bar graph, make the bars equally wide.

Artistically speaking, bar graphs are a bit dull. It is tempting to replace the bars with pictures for greater eye appeal.

Avoid replacing the bars with a picture!

Example: A misleading graph 1

Dyson’s ad claiming their vacuum has more than twice the suction of any other vacuum.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Example: A misleading graph 2

To magnify a picture, the artist must increase both height and width to avoid distortion of the image.

By increasing both the height and width of the Dyson vacuum, it appears to be 4 x 4 = 16 times larger.

Remember a bar graph should have bars of equal width; only the heights of the bars should vary.

Replacing the bars on a bar graph with pictures is tempting, but it is difficult to keep the ”bar” width equal without distorting the picture.

Change over Time: Line Graphs 1

Many quantitative variables are measured at intervals over time.

To display how a quantitative variable changes over time, make a line graph.

A line graph of a quantitative variable plots each observation against the time at which it was measured.

Always put time on the horizontal scale of your plot and the variable you are measuring on the vertical scale. Connect the data points by lines to display the change over time.

Change over Time: Line Graphs 2

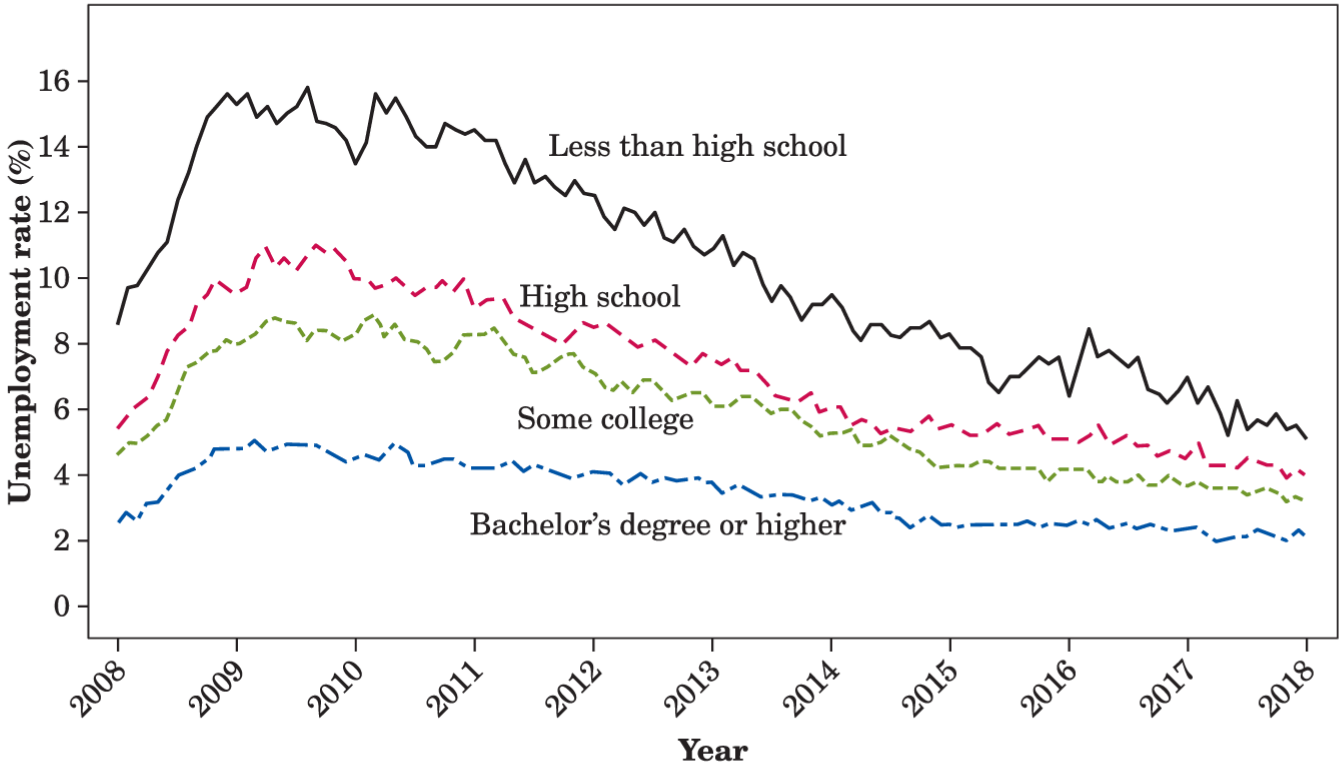
When constructing a line graph, make sure to use equally spaced time intervals on the horizontal axis to avoid distortion.

Line graphs can also be used to show how a quantitative variable changes over time, broken down according to some other categorical variable.

When displaying several categories, use a separate line for each category.

Example: Unemployment by education level

How has unemployment changed over time?



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Change over Time: Line Graphs 3

How should we describe patterns on a line graph?

First, look for an **overall pattern**. For example, a trend is a long-term upward or downward movement over time.

From the previous example, unemployment was at its highest for all education groups in 2009 and 2010, due to the Great Recession. Since then, the overall trend is showing a decrease in the unemployment rate for all education levels (each line is generally decreasing).

Change over Time: Line Graphs 4

Next, look for **striking deviations** from the overall pattern.

From the previous example, there is a noticeable increase from 2008 to the beginning of 2009. This was a side effect of the recession economy. Unemployment hovered around these record highs through 2010, when the rates finally started their decline to the current levels. There is a striking deviation (a drastic dip) in the unemployment rate for those with less than a high school degree around mid-2010.

Change over Time: Line Graphs 5

A pattern that repeats itself at known regular intervals of time is called **seasonal variation**.

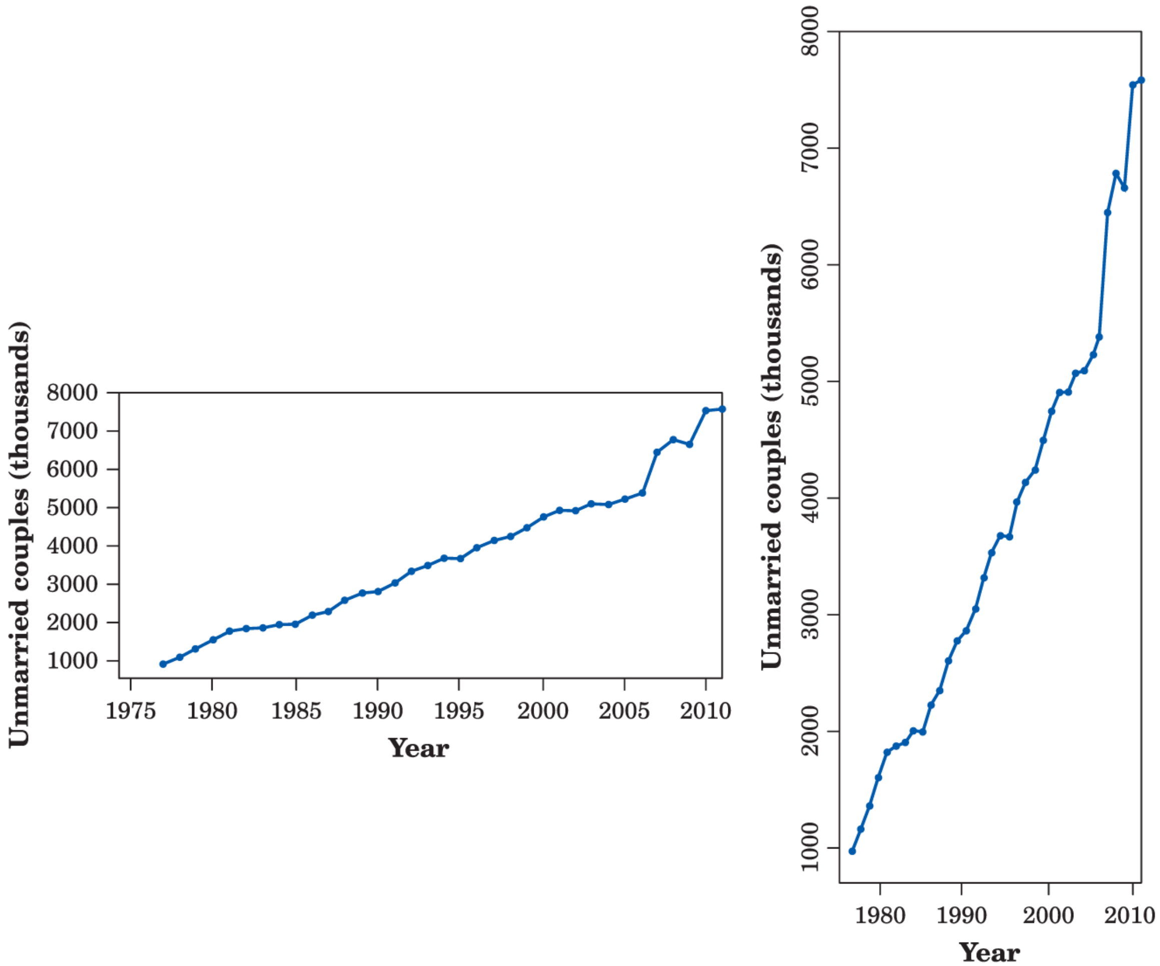
Calculating the unemployment rate depends on the size of the workforce and the number of those in the workforce who are working. The unemployment rate rises every year in January as holiday sales jobs end and outdoor work slows in the North due to winter weather. This is regular and predictable. As such, the Bureau of Labor Statistics makes seasonal adjustments to the monthly unemployment rate before reporting it to the general public.

Watch Those Scales! 1

Because graphs speak so strongly, they can mislead the unwary.

The careful reader of a line graph looks closely at the scales marked off on the axes.

Example: Living together



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Watch Those Scales! 2

Another important issue concerning scales is the following.

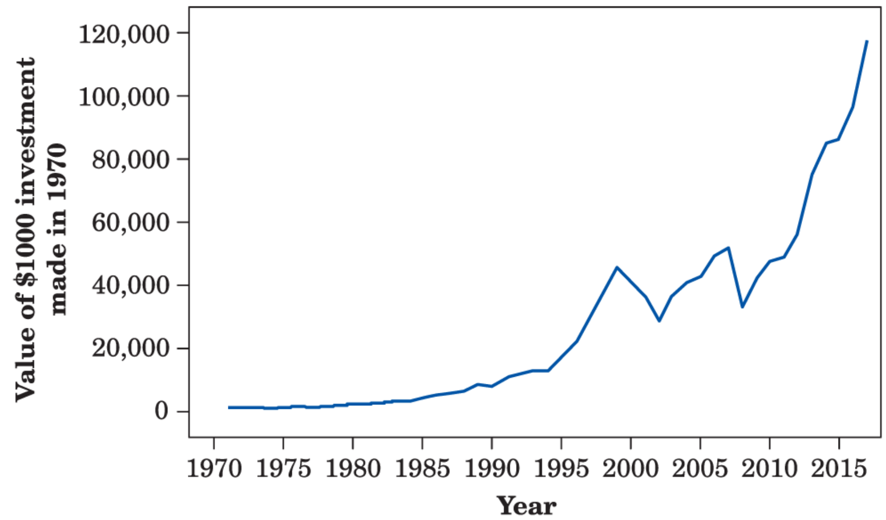
When examining the change in the price or value of an item over time, plotting the actual increase can be misleading.

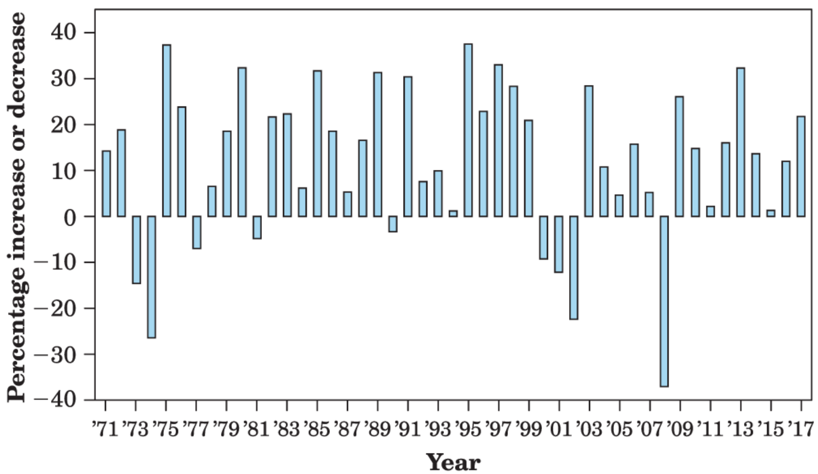
$P_{t+1} - P_t$

It is often better to plot the percentage increase from the previous period.

$$\frac{P_{t+1} - P_t}{P_t}$$

Example: Getting rich in hindsight





Watch Those Scales! 3

The line graph gives the impression that increases between 1970 and 1995 were negligible but that increases between 1995 and 1999 and after 2008 were dramatic.

While it is true that the actual value of our investment increased much more between 1995 and 1999 and after 2008 than it did between 1970 and 1995, it would be incorrect to conclude that investments in general increased much more dramatically between 1995 and 1999 and after 2008 than in any of the years between 1970 and 1995.

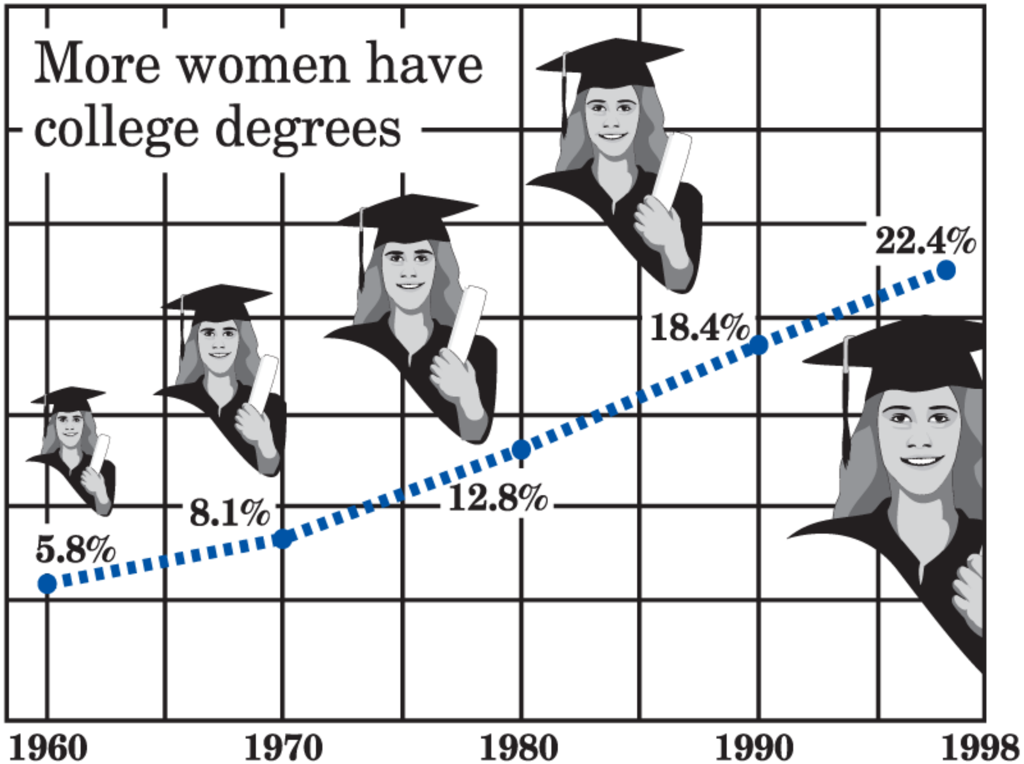
Watch Those Scales! 4

The second graph tells a different, and more accurate, story.

For example, the percentage increase in 1975 (approximately 37%) rivaled that in any of the years between 1995 and 1999. However, in 1975 the actual value of our investment was relatively small ($1170) and a 37% increase in such a small amount is nearly imperceptible on the scale used in the line graph.

By 1995 the actual value of our investment was about $14,000, and a 37% increase appears much more striking.

Example: The rise in college education 1



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Example: The rise in college education 2

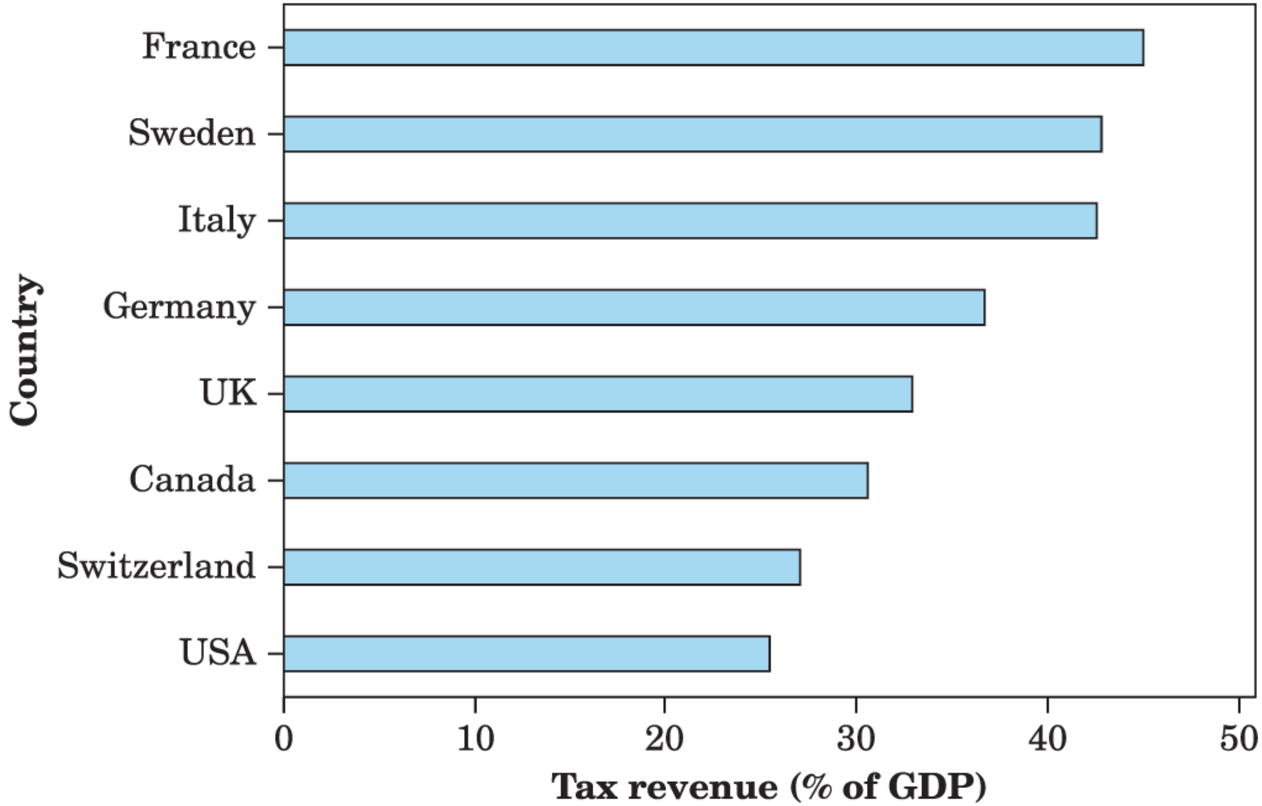
The graph shows the rise in the percentage of women aged 25 years and over who have at least a bachelor’s degree.

The graph isn’t simple. The artist couldn’t resist a nice background sketch and also cluttered the graph with grid lines. Grid lines on a graph serve no purpose.

Using pictures of the female graduate at each time point distorts the values in the same way as a pictogram.

Example: High taxes, reconsidered 1

The tax data displayed again….



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Example: High taxes, reconsidered 2

The first bar graph we saw displayed this tax data with the countries in alphabetical order.

The second bar graph rearranges the categories in order of the tax burdens. This simple change improves the graph by making it clearer where each country stands in the group of eight countries.

The second bar graph also demonstrates the ability to display a bar graph horizontally.

Statistics in Summary 1

To see what data say, start with graphs.

The choice of graph depends on the type of data. Do you have a **categorical variable**, such as level of education or occupation, which puts individuals into categories? Or do you have a **quantitative variable** measured in meaningful numerical units?

Check data presented in a table for **roundoff errors**.

The **distribution** of a variable tells us what values it takes and how often it takes those values.

Statistics in Summary 2

To display the **distribution** of a categorical variable, use a pie chart or a bar graph. **Pie charts** always show the parts of some whole, but bar graphs can compare any set of numbers measured in the same units. **Bar graphs** are better for comparisons. Bar graphs can be displayed vertically or horizontally.

Statistics in Summary 3

To show how a quantitative variable changes over time, use a **line graph** that plots values of the variable (vertical scale) against time (horizontal scale). If you have values of the variable for different categories, use a separate line for each category. Look for **trends** and **seasonal variation** in a line graph, and ask whether the data have been **seasonally adjusted**.

Statistics in Summary 4

Graphs can mislead the eye. Avoid **pictograms** that replace the bars of a bar graph by pictures whose height and width both change. Look at the scales of a line graph to see if they have been stretched or squeezed to create a particular impression. Avoid clutter that makes the data hard to see.