**Section 2.1 Organizing Qualitative Data**

Chapter 1 discussed how to identify the research objective

and collect data.

We learned that data can be obtained

from either observational studies

or designed experiments.

When data are obtained, they are referred to as raw data.

The purpose of this chapter is to learn

how to organize raw data into a meaningful form

so that we can understand what the data are telling us.

The first step in determining how to organize raw data

is to determine whether the data is qualitative or quantitative.

We know from chapter 1 that there are two types of data--

qualitative and quantitative.

In this section, we will concentrate

on tabular and graphical summaries of qualitative data.

In sections 2.2 and 2.3, we discuss tabular and graphical

summaries of quantitative data.

**When**[**qualitative data**](https://xlitemprod.pearsoncmg.com/assignment/containerassignmentplayer.aspx#xln-lb-lnk_obj2_1_bc0736fe-cbc0-ca31-5c28-041c93316f8b)**are collected, we often first determine the number of occurrences within each category.**

**DEFINITION**

**A frequency distribution lists each category of data and the number of occurrences for each category of data.**

**STATCRUNCH**

**In this example, we'll learn to organize qualitative data**

**into a frequency distribution.**

A physical therapist wants to determine

types of rehabilitation required by her patients.

To do so, she obtains a sample 30 of her patients

and records the body part requiring rehabilitation.

Construct a frequency distribution

for the following data.

Here are the data.

Let's go ahead and open up StatCrunch.

Here we are in StatCrunch.

I've typed the data in the first column labeled Body Part.

And we'll press Stat, Tables, Frequency.

Next, we'll select the column containing the data-- Body

Part.

And we just want the frequency.

So click on Frequency.

And click Compute.

And there's our frequency distribution.

Let's go over the steps for using StatCrunch.

Type the data, one per line, in one column.

Press the Stat button.

And from Tables, select Frequency.

Select the column containing the data.

Select Frequency.

And click Compute.

And you'll have your result.

STATCRUNCH RESULTS

2.1 Organizing Qualitative Data

In any frequency distribution, it is a good idea to add up the frequency column to make sure that it equals the number of observations.

In [Example 1](https://xlitemprod.pearsoncmg.com/assignment/containerassignmentplayer.aspx#xln-lb-lnk_obj2_3_3099b78d-80df-c737-015f-998ccd0207ed), the frequency column totals to 30 as it should because there are 30 body parts (observations).

STATCRUNCH RESULTS

**Frequency table results for DATA:**

Count = 30

| **DATA** | **Frequency** |
| --- | --- |
| Back | 12 |
| Knee | 5 |
| Shoulder | 4 |
| Wrist | 2 |
| Hip | 2 |
| Hand | 2 |
| Neck | 1 |
| Groin | 1 |
| Elbow | 1 |

SHOW SOLUTION

1. A random sample was used to attain each category of data and a recording of the body part requiring rehab.
2. A Sample of 30

Question 1. StatCrunch

**Frequency table results for var1:**

Count = 50

| **var1** | **Frequency** |
| --- | --- |
| O | 24 |
| A | 16 |
| AB | 6 |
| B | 4 |

Ob 1, Part 6.

Often, we want to know the *relative frequency* of the categories rather than the frequency.

**DEFINITION**

**The *relative frequency* is the proportion (or percent) of observations within a category and is found using the formula**

Relative frequency=Frequency /Sum of all frequencies

**A *relative frequency distribution* lists each category of data together with the relative frequency.**

| **TABLE 2** | | |
| --- | --- | --- |
| **Body Part** | **Tally** | **Frequency** |
| Back | |||| |||| || | 12 |
| Wrist | || | 2 |
| Elbow | | | 1 |
| Hip | || | 2 |
| Shoulder | |||| | 4 |
| Knee | |||| | 5 |
| Hand | || | 2 |
| Groin | | | 1 |
| Neck | | | 1 |

Obj1, Pg. 8

5. The Sample Size. Because it is measuring frequency of a piece of data within a full sample size.

OR THAT THE **SUM IS EQUAL TO EXACTLY ONE.**

It is a good idea to add up the relative frequencies to be sure they sum to 1. In fraction form, the sum should be exactly 1. In decimal form, the sum may differ slightly from 1 due to rounding.

**OBJECTIVE 2**

Once raw data are organized in a table, we can create graphs. Just as "a picture is worth a thousand words," pictures of data result in a more powerful message than do tables.

Try the following exercise: Open a newspaper or news website and look at a table and a graph. Study each one. Then put the paper away and close your eyes. What do you see in your mind's eye? Can you recall information more easily from the table or the graph? In general, people are more likely to recall information obtained from a graph than from a table.

A common device for graphically representing qualitative data is a *bar graph*

**DEFINITION**

A **bar graph** is constructed by labeling each category of data on either the horizontal or vertical axis and the frequency or relative frequency of the category on the other axis. Rectangles of equal width are drawn for each category.

1. ***A bar graph* is constructed by labeling each category of data on either the horizontal or vertical axis and the frequency or relative frequency of the category on the other axis. The height of each rectangle represents the category's *frequency or relative frequency.***

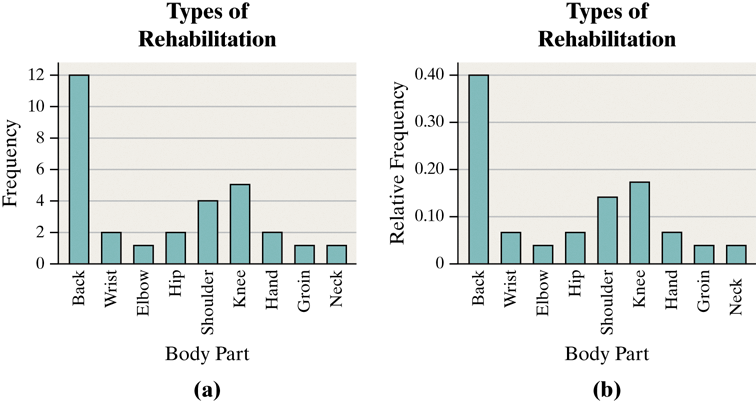
Both bar graphs are labeled “Types of Rehabilitation”. The bar graph with the x-axis labeled “Body Part (a)” and the y-axis labeled “Frequency” is summarized below:

|  |  |
| --- | --- |
| Body part | Frequency |
| Back | 12 |
| Wrist | 2 |
| Elbow | 1 |
| Hip | 2 |
| Shoulder | 4 |
| Knee | 5 |
| Hand | 2 |
| Groin | 1 |
| Neck | 1 |

Bar graph with the x-axis labeled “Body Part (b)” and the y-axis labeled “Relative Frequency” is summarized below:

|  |  |
| --- | --- |
| Body part | Relative frequency |
| Back | 0.40 |
| Wrist | 0.07 |
| Elbow | 0.04 |
| Hip | 0.07 |
| Shoulder | 0.14 |
| Knee | 0.18 |
| Hand | 0.07 |
| Groin | 0.04 |
| Neck | 0.04 |

All data are approximate.



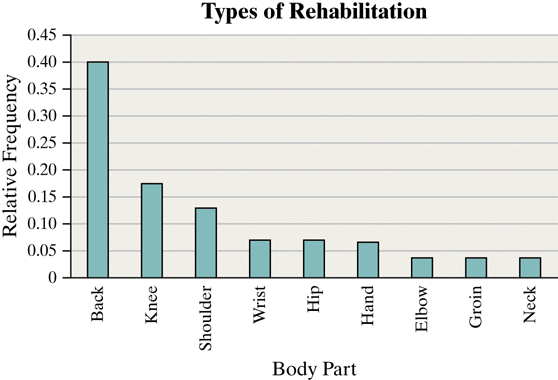
7. Pareto Chart?

In bar graphs, the order of the categories does not usually matter. However, bar graphs that have categories arranged in decreasing order of frequency help prioritize information for decision-making purposes.

**DEFINITION**

A **Pareto chart** is a bar graph whose bars are drawn in decreasing order of frequency or relative frequency

DOWN BELOW



8. Explain why it is best to use relative frequencies when comparing data sets.

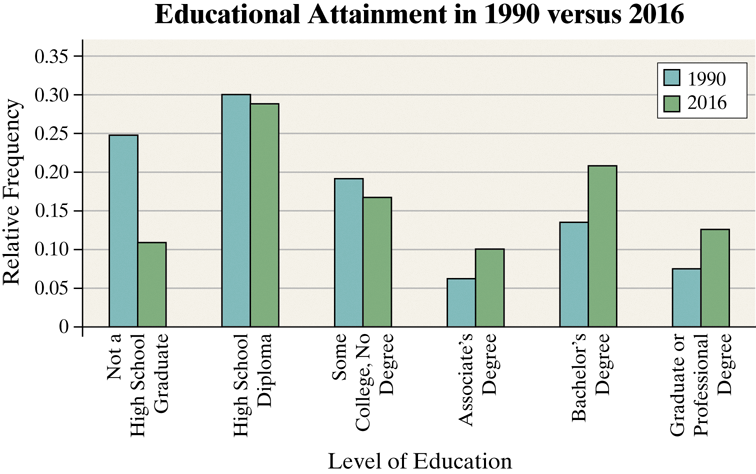
When comparing data sets, it is best to use relative frequencies because different sample or population sizes make comparisons using frequencies difficult or misleading.

**Example 4 *Comparing Two Data Sets***

The frequency data in Table 4 represent the educational attainment (level of education) in 1990 and 2016 of adults 25 years and older who are U.S. residents. The data are in thousands. So 39,344 represents 39,344,000.

**Table 4**

| **Educational Attainment** | **1990** | **2016** |
| --- | --- | --- |
| Not a high school graduate | 39,344 | 23,453 |
| High school diploma | 47,643 | 62,002 |
| Some college, no degree | 29,780 | 36,003 |
| Associate’s degree | 9792 | 21,657 |
| Bachelor’s degree | 20,833 | 44,778 |
| Graduate or professional degree | 11,478 | 27,122 |
| Totals | 158,870 | 215,015 |



The frequency data in Table 4 represent the educational attainment (level of education) in 1990 and 2016 of adults 25 years and older who are U.S. residents. The data are in thousands. So 39,344 represents 39,344,000

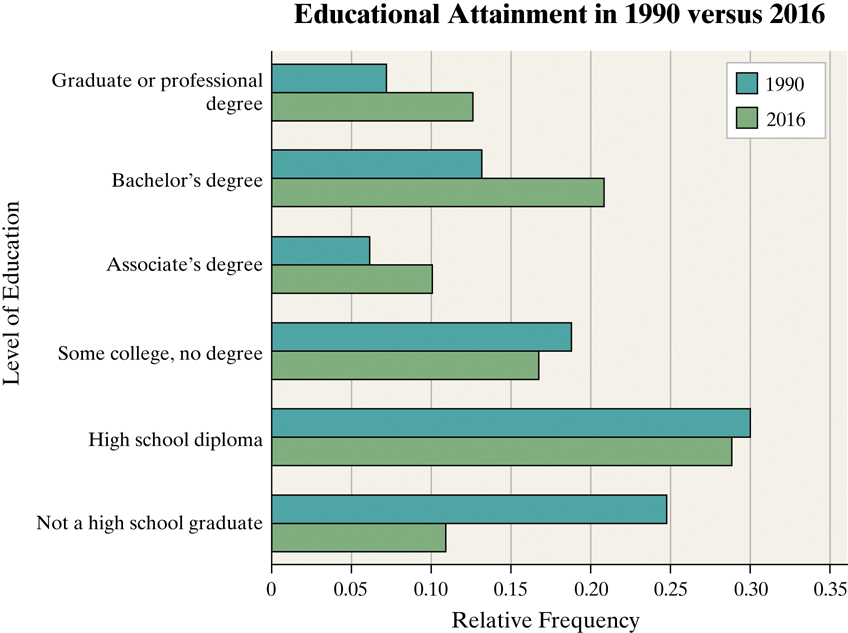
|  |  |  |
| --- | --- | --- |
| **Educational Attainment** | 1990 | 2016 |
| Not a high school graduate | 39,344 | 23,453 |
| High school diploma | 47,643 | 62,002 |
| Some college, no degree | 29,780 | 36,003 |
| Associate's degree | 9,792 | 21,657 |
| Bachelor's degree | 20,833 | 44,778 |
| Graduate or professional degree | 11,478 | 27,122 |
| **Totals** | 158,870 | 215,015 |

Solution

The side-by-side bar graph illustrates that the proportion of Americans 25 years and older who had some college but no degree was higher in 1990. This information is not clear from the frequency table (Table 4) because the total population sizes are different. The increase in the number of Americans who did not complete a degree is due partly to the increases in the size of the population. In addition, the number of individuals with a Bachelor's degree more than doubled (20,833 to 44,778). However, from the side-by-side bar graph, we see that the proportion of Americans 25 years and older who had a Bachelor's degree did not double. It is also clear that adult Americans have more education in 2016 than in 1990 with a much higher percentage of the population having at least a bachelor's degree (20.33% in 1990 versus 33.44% in 2016).

#### 9. **Question Horizontal Bars**

Bar graphs may also be drawn with horizontal bars. Horizontal bars are preferable when category names are lengthy. For example, Figure 4 uses horizontal bars to display the same data as in [Figure 3](https://xlitemprod.pearsoncmg.com/assignment/containerassignmentplayer.aspx#xln-lb-lnk_obj3_8_0e0bae58-d7c3-047a-d700-cba51c15430d).



1. What is a pie chart?

*Pie charts* are typically used to present the relative frequency of qualitative data. In most cases, the data are nominal, but ordinal data can also be displayed in a pie chart.

**DEFINITION**

A **pie chart** is a circle divided into sectors. Each sector represents a category of data. The area of each sector is proportional to the frequency of the category.

Problem

The frequency data presented in Table 6 represent the educational attainment of U.S. residents 25 years and older in 2016. The data are in thousands so 23,453 represents 23,453,000. Construct a pie chart of the data.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **TABLE 6** | | | **Educational Attainment** | **2016** | | Not a high school graduate | 23,453 | | High school diploma | 62,002 | | Some college, no degree | 36,003 | | Associate's degree | 21,657 | | Bachelor's degree | 44,778 | | Graduate or professional degree | 27,122 | | **Total** | 215,015 | | Data from U.S. Census Bureau | | |

**CONSTRUCT PIE CHART**

In this example, we'll learn to construct a pie

chart using StatCrunch.

The following data represent the educational attainment

of residents of the United States 25 years

and older in 2016 based on data from the U.S. Census Bureau.

The data are in thousands.

Construct a pie chart of the data.

Here are the data.

Let's go to StatCrunch.

I've already typed the data into StatCrunch.

**1. The educational attainment categories**

**are in the first column,**

**2.the frequencies for 2016**

**in the second column.**

**3.To make the pie chart, I click Graph, Pie Chart With Summary.**

We use With Summary whenever we have

a total frequency for each category

rather than the raw data, like high school

graduate, bachelor's degree, bachelor's degree,

some college, et cetera.

**4.The categories are in the first column.**

**5. I label that educational attainment.**

**6. The counts are in 2016.**

And I want to make sure that I display the count

and the percent of total.

**7. Order by-- I'm going to stick with the worksheet order**

**8. and press Compute.**

And there is our pie chart.

**Now we'll go over the StatCrunch steps, type**

**the categories in one column,**

**and the counts in another.**

**Press the Graph button,**

**and from Pie Chart, select With Summary.**

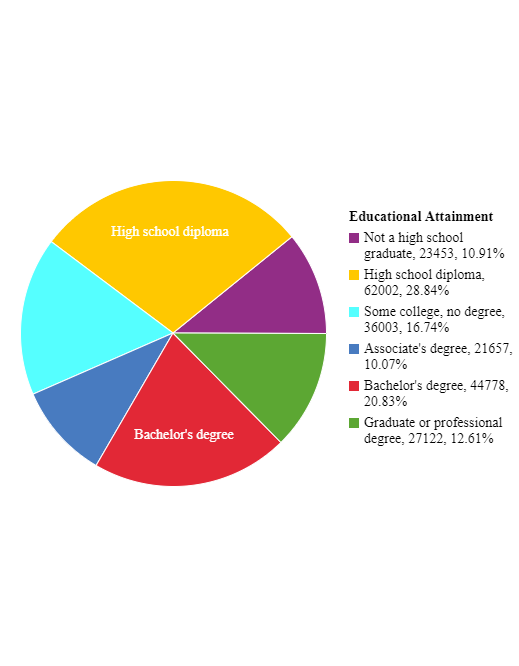
**Select the columns containing the categories and counts,**

**and click Next.**

**Display the percent of total for each category, click Compute,**

**and you'll have your pie chart.**

1. Which graph, a pie chart or a bar graph, is better at comparing one category to another category?
2. Which graph, a pie chart or a bar graph, is better at comparing one category to the whole?

**11**

11.Which graph, a pie chart or a bar graph, is better at comparing one category to another category?

Bar

12.Which graph, a pie chart or a bar graph, is better at comparing one category to the whole?

**Pie**

In this video, we'll talk about when

should a bar graph or pie chart be used.

Pie charts are useful for showing

the division of all possible values

of a qualitative variable into its parts.

However, because angles are often

hard to judge in pie charts, they

are not as useful in comparing two specific values

of the qualitative variable.

Instead, the emphasis is on comparing the part

to the whole.

Bar graphs are useful when we want

to compare the different parts, not necessarily

the parts to the whole.

For example, to get the big picture regarding

educational attainment, a pie chart is a good visual summary.

Here we can tell what proportion of the total

is represented by adults who had a bachelor's degree.

However, it's difficult to compare the adults who

got a bachelor's degree to the adults who got a high school

diploma only.

To compare bachelor's degrees to high school diplomas,

a bar graph is a good visual summary.

Here we can compare the height of each bar.

And we can determine how much taller the high school diploma

bar is than the bachelor's degree bar.

Since bars are easier to draw and compare,

some practitioners forego pie charts

in favor of Pareto charts when comparing parts to the whole.