Segment: Descriptive Statistics

Topic: Visual Representation of Data

Table of Contents

Introduction

A picture is worth a thousand words, indeed. A large data set might be meaningless unless it is organised into tables, charts, or graphs. This topic introduces some of the most common types of tables, charts, and graphs that are useful in organising and presenting data.

In this topic, we will investigate a few methods of displaying data, some of which are descriptive only.

Let us first look at the overview of this topic.

The United States government publishes the pie charts for filing tax returns. It is an effective way of communicating to the taxpayers what proportion of taxes is collected from different sources and how the tax revenue is spent.

This communication is more effective than presenting the numbers involved because there will be very many numbers and each number would be so large with more than ten digits that it would be difficult to comprehend them.

A comprehension of the facts and the realisation of the effort the government puts into communicating those facts offer some comfort, however little, to millions of taxpayers.

In many presentations that you are going to make, you might find yourself in a similar situation: you have to communicate large sets of numerical data to an important audience. You will then have to use charts, graphs, and tables in your presentation.

In this topic, we will study different types of charts and graphs and their suitability for the type of data communicated. You will also learn how to use spreadsheet templates to create many types of charts.

Learning Objectives

At the end of this topic, you will be able to:

- distinguish between frequency and pivot tables
- evaluate the importance of using different types of charts
- assess the significance of using different types of graphs.

1. Types of Tables

The two types of tables normally used to present quantitative data are frequency and pivot.

Frequency Tables

An effective method of presenting data that is essentially quantitative is to present it as a frequency table, where the data is grouped into various categories or classes.

The frequency table lists the number of observations for some variables that fall into various categories. For example, suppose that data was collected on the ages of 25 employees of a particular company. The data is quantitative and could be presented as a frequency table, where the data is grouped into categories based on age groups as shown in the table below.

Table 1: Frequency Table

Age	Frequency
Less than 35	1
36 - 40	3
41 - 45	2
46 - 50	7
51 - 55	5
56 - 60	3
61 - 65	1
66 - 70	3

There is no set rule for determining the correct number of categories or classes, although common sense should prevail. As a guide, somewhere between 5 to 15 categories might be suggested.

Pivot Tables

One of the most powerful tools in Excel for analysing data is a pivot table. Pivot tables enable us in a variety of ways in order to investigate relationships in data. Tables generated using this technique are often called contingency tables or cross tabs. Excel, however, provides more flexibility with pivot tables than is usually reflected with simple contingency tables. Contingency tables list only counts, whereas pivot tables can list counts, averages, sums, and other measures.

As an example, let us say we have been performing some market research to investigate the characteristics of Internet users. Suppose we were interested in investigating the relationship between Age and Gender (coded as 1 for female and 0 for male). One of the questions of interest may have been the proportion of Internet users that were men under thirty. Pivot tables allow us to group these variables as shown in the table below, where the answer to our question can be identified with ease.

Table 2: Pivot Table

Count of Gender	Gender		
Age	0 (Male)	1 (Female)	Grand Total
16-29	19.20%	12.60%	31.80%
30-43	22.20%	15.10%	37.30%
44-57	13.30%	7.20%	20.50%
58-71	5.10%	3.80%	8.90%
72-85	1.00%	0.50%	1.50%
Grand total	**60.80%**	**39.20%**	**100.00%**

2. Types of Charts

Additionally, charts may also be used. In this section, we will look at three different types of charts:

1. Histogram
2. Pie chart
3. Bar chart

Pie and bar charts can be used for both qualitative and quantitative data, while histograms are for quantitative data.

Read below to find out more about these charts.

Types of Charts

1. Histograms

When data is grouped into categories or classes, we could plot a frequency distribution of the data. Such a frequency plot is called a histogram.

Visually, a histogram is a chart made up of bars of different heights. The height of each bar represents the frequency of values in the class represented by the bar. Adjacent bars share

sides. For example, if you refer back to the frequency table given earlier in this topic, we could represent the data on the ages for the employees as shown in the figure below.

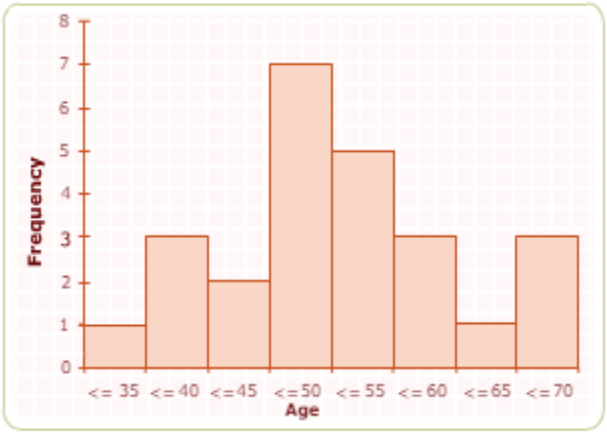


Fig. 1: Histogram

2. Pie Charts

Pie charts are used to show the proportion of different parts that make up the whole, for example, we can show the proportion of different ethnic groups in a community or the proportion of different types of expenses within a budget.

The figure below shows a pie chart of the geographical locations of the world's largest telecommunications companies.

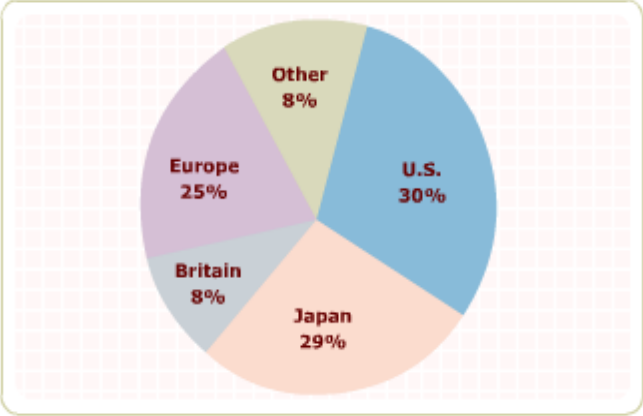


Fig. 2: Pie Chart

3. Bar Chart

Bar charts use horizontal or vertical rectangles to display categorical data when there is no emphasis on the percentage of a total represented by each category. The scale of measurement is nominal or ordinal. A bar chart is a good way to show how different categories stack up against one another. The figure below shows how a bar chart can be used to display and interpret operating expenses and revenues of the top US airline carriers.

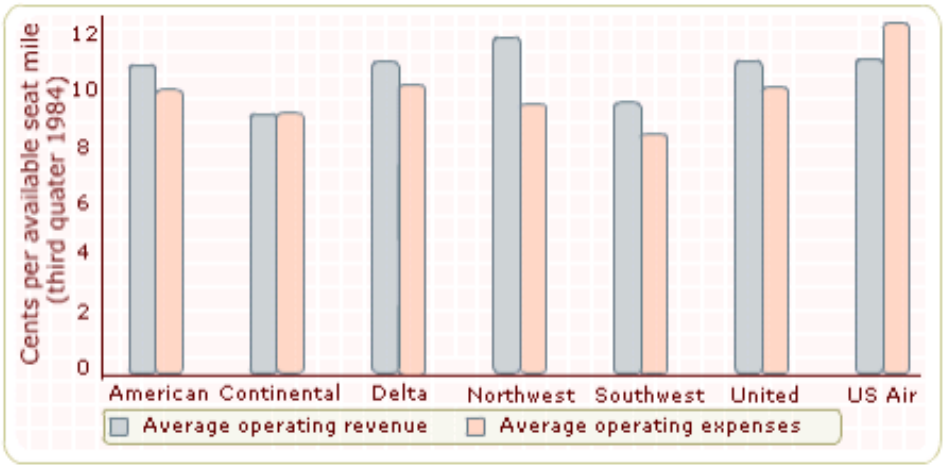


Fig. 3: Bar Chart

The following are key differences between histograms and bar charts:

- In a (vertical) bar chart, the horizontal axis does not represent a continuous scale, as each bar represents a discrete value or a qualitative measurement.
- In a histogram, the horizontal axis is in a numerical scale (interval or ratio) and the measurement is continuous.

For this reason, a bar chart is drawn with gaps between successive bars, and a histogram is drawn without any gaps.

3. Types of Graphs

Let us now look at two different types of graphs, namely:

1. Scatter plot
2. Line graph

Scatter Plot

In some instances, we are interested in investigating the relationship between two variables. One way is to prepare a plot where each pair of data points is represented as a point on the graph. The resulting graph is called a scatter plot as shown in the following figure.

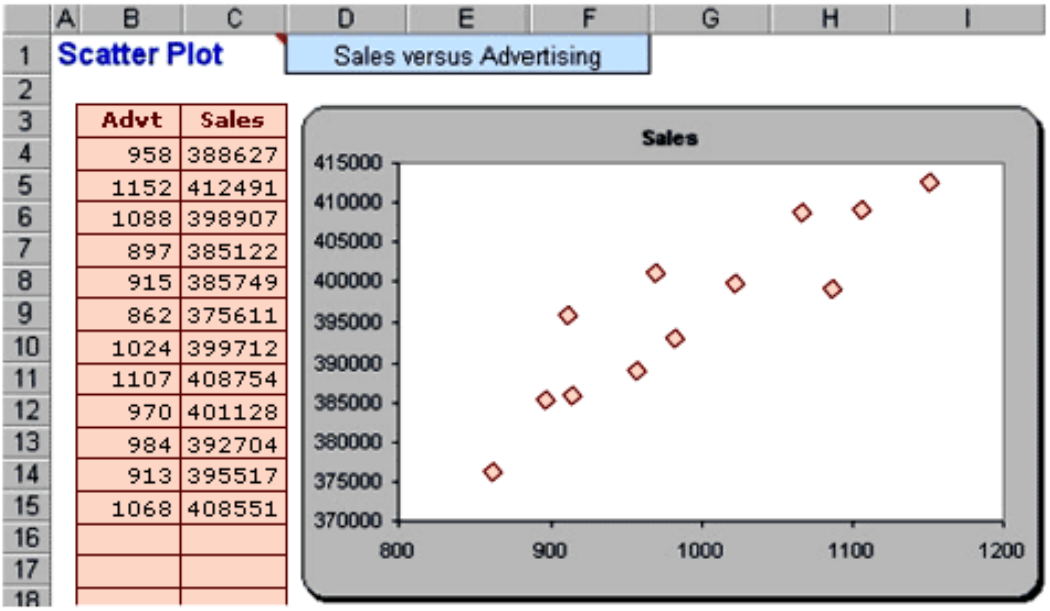


Fig. 4: Scatter Plot

Scatter plots can be generated using either the Excel® Chart Wizard or using the StatTools add-in. To use the Excel® Chart Wizard, first place your cursor anywhere in the data that you wish to plot and then choose Insert/Chart. The chart wizard dialog box will then open up and you can select the type of chart that you wish to insert. Choose the XY Scatter plot graph type and then click on Next. The chart wizard will show a preview of the scatter plot. If this is correct you can click on OK and the scatter plot will be inserted into the worksheet.

Alternatively, StatTools can be used to generate a scatter plot. To do this, first place your cursor anywhere in the data that you wish to plot ad choose StatTools/Chart/Scatter plot(s). Then follow the instructions provided and the scatter plot will be inserted into the worksheet.

A scatter plot is used to identify underlying relationships among variables. For example, if we have the data on annual sales of a product and the annual advertising budgets for that product during the same period, we can plot them on a scatter plot to identify any underlying patterns in the data. One would expect that whenever the advertising budget was high, the sales would also be high. This relationship can be explored using the scatter plot.

The plot consists of a scatter of points, each point representing an observation. For instance, if the advertising budget in one year was x and the sales in the same year was y, then a point is marked on the plot at coordinates (x, y).

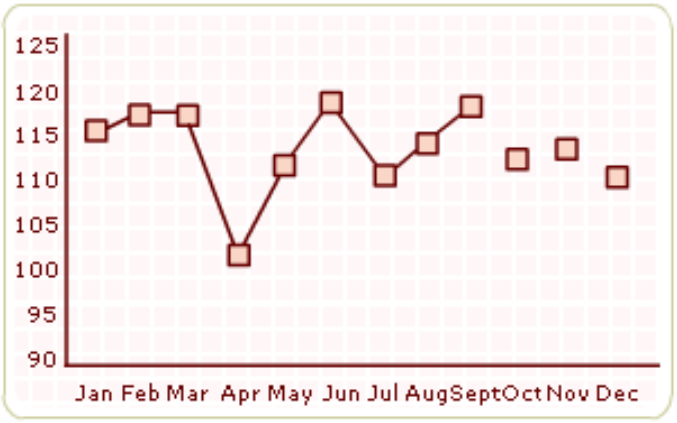
The figure above shows a scatter plot of sales versus advertising budget observed over a period of 12 years measured in thousands of dollars. It shows that whenever the advertising budget was high, sales were also high. This type of relationship is known as a positive correlation. We will learn more details about correlation in the topic, “Numerical representation of data”. A negative correlation between two variables means that when one increases, the other tends to decrease.

Line Graphs

Line graphs are an effective way to represent the relationship between two variables particularly when time is involved.

A time plot is an example of a line graph. Time plots display the changes in a variable that occurred over time. In these graphs, the time variable is plotted along the horizontal axis of the graph.

A line graph of monthly sales figures is shown in the figure below.



Jan	Feb	Mar	Apr	May	Jun
115	116	116	101	112	119

Jul	Aug	Sept	Oct	Nov	Dec
110	115	118	114	115	110

Fig. 5: Line Graph

4. Summary

Here is a quick recap of what we have learnt so far:

- The two types of tables normally used to present quantitative data are frequency and pivot.
- Graphical methods can be used to present statistical information. Some of these methods are
  - Histograms
  - Bar charts
  - Pie charts
- Histograms are very commonly used and enable us to plot a frequency distribution of grouped data.
- A bar chart is a good way to show how different categories stack up against one another. The difference between a histogram and a bar chart is that in a bar chart, the horizontal axis is not in a continuous scale.
- Pie charts are used to show the proportion of different parts that make up the whole.
- A scatter plot is used to identify underlying relationships among variables.
- Line graphs such as time plots represent the relationship between two variables.
- Excel provides a powerful analysis tool called pivot tables, which enable us to summarise and present data in a variety of forms.

5. Glossary

Time plot	A graph of the value of a variable is placed on the vertical axis versus time on the horizontal axis.