



IIT Madras
ONLINE DEGREE

Statistics for Data Science -1

Describing Categorical Data- Single Variable

Usha Mohan

Indian Institute of Technology Madras

Charts of categorical data

- ▶ The two most common displays of a categorical variable are a bar chart and a pie chart.
- ▶ Both describe a categorical variable by displaying its frequency table.

A **pie chart** is a circle divided into pieces proportional to the relative frequencies of the qualitative data.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻ 3 / 14

Definition

A **pie chart** is a circle divided into pieces proportional to the relative frequencies of the qualitative data.

- The steps to construct a pie-chart¹
 - Step 1 Obtain a relative-frequency distribution of the data.
 - Step 2 Divide a circle into pieces proportional to the relative frequencies.
 - Step 3 Label the slices with the distinct values and their relative frequencies.

¹Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited. 2014.

Example

Use a protractor and the fact that there are 360° in a circle. Thus, for example, the first slice of the circle is obtained by marking off $0.4 \times 360 = 144^\circ$.

Example

Use a protractor and the fact that there are 360° in a circle. Thus, for example, the first slice of the circle is obtained by marking off $0.4 \times 360 = 144^\circ$.

1. A,A,B,C,A,D,A,B,D,C

Category	Tally mark	Freq	Rel freq	Degrees
A		4	0.4	144
B		2	0.2	72
C		2	0.2	72
D		2	0.2	72
Total		10	1	360°

Example

1. A,A,B,C,A,D,A,B,D,C, A,B,C,D,A

Category	Tally mark	Freq	Relative freq	Degrees
A		6	0.4	144
B		3	0.2	72
C		3	0.2	72
D		3	0.2	72
Total		15	1	360°

1. A pie chart is used to show the proportions of a categorical variable.
2. A pie chart is a good way to show that one category makes up more than half of the total.

Bar chart

Definition

A bar chart displays the distinct values of the qualitative data on a horizontal axis and the relative frequencies (or frequencies or percents) of those values on a vertical axis. The frequency/relative frequency of each distinct value is represented by a vertical bar whose height is equal to the frequency/relative frequency of that value. The bars should be positioned so that they do not touch each other.

Steps to construct a bar chart

To Construct a Bar Chart²

- Step 1 Obtain a frequency/relative-frequency distribution of the data.
- Step 2 Draw a horizontal axis on which to place the bars and a vertical axis on which to display the frequencies/relative frequencies.
- Step 3 For each distinct value, construct a vertical bar whose height equals the frequency/relative frequency of that value.
- Step 4 Label the bars with the distinct values, the horizontal axis with the name of the variable, and the vertical axis with “Frequency” / “Relative frequency.”

²Weiss, Neil A. Introductory Statistics: Pearson New International Edition. Pearson Education Limited, 2014.

Bar chart in a google sheet

Step 1 Select/Highlight the cells having data you want to visualize.

Step 2 Click the Insert Chart option in Google Sheets toolbar.

Step 3 Change the visualization type in Chart editor.

Step 4 Select in Chart Editor Chart type to Bar chart.

Pareto charts

Definition

*When the categories in a bar chart are sorted by frequency, the bar chart is sometimes called a **Pareto chart**. Pareto charts are popular in quality control to identify problems in a business process.*

- ▶ If the categorical variable is ordinal, then the bar chart must preserve the ordering.

Example- Pareto chart

A,B,B,C,A,D,B,B,A,C, B,B,C,D,A

Category	Tally mark	Freq	Relative freq
A		4	0.26
B		6	0.40
C		3	0.20
D		2	0.14
Total		15	1

Example- ordinal data

The T-shirt sizes (Small-S, Medium-M, Large-L) of twenty students is listed below:

L,M,M,S,L,S,S,M,L,M,M,S,S,L,M,S,M,S,L,M

Size	Tally mark	Freq	Relative freq
Small		7	0.35
Medium		8	0.40
Large		5	0.25
Total		20	1

Sectional summary

1. A bar chart is used to show the frequencies/relative frequencies of a categorical variable.
2. If ordinal, the order of categories is preserved.
3. The bars can be oriented either horizontally or vertically.
4. A Pareto chart is a bar chart where the categories are sorted by frequency.