



Visual Displays- One Variable

Before you Watch

Visual displays if done well can more effectively or succinctly communicate key information than text.

Visual displays in Statistics are referred to as graphs, and the type of graph, or visual display, we use is determined by ... yes ... the **variable type** upon which we have data to display.

Whether reading or drawing a graph for a single variable, there are 3 labels that are very important.

- The **title** – provides details of what the graph is about eg “Distribution of the ages of women in the Hunter region diagnosed with coeliac disease in June 2016”
- The **vertical axis** (y) – this is often a count (frequency) eg “Number of women”
- The **horizontal axis** (x) – the **variable of interest** eg “Age”

Graphs of single variables often demonstrate the relative frequency of the varying values that a variable can assume. This is often referred to as the **distribution** of the variable and is a vital piece of information; it is also required when determining most appropriate methods of analysis to use.

The Video Content

This video provides some key visual displays, and descriptive measures, for reporting upon a single variable. In video 1 we described the concept of a Random Variable, identifying what you’re interested in measuring.

For example ‘reaction time’...
...and determining the variable’s type.

For example, reaction time is continuous. Once data are recorded for a variable, we need to communicate the key information the data provides. For a variable that is continuous we can consider a histogram of the data, which shows the relative frequency, or distribution, of the recorded data once they are grouped into short

intervals- in this example 2-second intervals from 18 to 32 seconds.

...And we describe three key things: shape, center and spread ---- the 3 Ss. The shape of the distribution in this case is symmetric about the centre, and bell-shaped, like a Normal curve. The center of the data may be described using the mean or median.

The mean is found by adding the 21 values and dividing by 21. The median is the middle value, once they are sorted from lowest to highest, in this case the value that has 10 values below it and 10 values above it.

And the spread (or variation) of the data is often described by the standard deviation (which essentially measures how close the values are to the mean on average). A small standard deviation indicates the data are all relatively close to the center whilst a larger standard deviation indicates the data are not all so close, and are spread further away from the center.

Recall the four variable types, which were described in the video on random variables. We've just considered a continuous variable, reaction time, and how to display and describe the data in summary. For each variable type, certain visual displays and descriptions are appropriate.

Consider the nominal variable 'Primary complaint' which records the primary complaint of customers staying at a hotel. A frequency table lists the variable's outcomes, and how often each was identified, thus categorising the individual responses, and simplifying the information.

Better yet we re-order these categories based on frequency (or percentage) to more clearly see underlying messages, Note that the category 'Other' is a compilation of all responses where each had a smaller frequency than the smallest individually listed response 'Telephone', so the category 'Other' appears last, and we know in this case it would have at least 9 categories within it.

A Pareto Chart further assists in prioritising action. The blue bars indicate the relative frequency for each category. The red line represents the cumulative percentage as we include more and more of the complaint types from left to right.

So what are the key messages? The reservation process is a key concern, reported by over half the respondents. Further, over 70% report one of two primary concerns, reservation process and cleaning- these seem priority areas to consider for improvement. The Pareto Principle states that the larger majority of outcomes are



due to a smaller minority of reasons and is exemplified here. In this case 70% of responses are due to about 13% (or two) of the 15 or more categories...considering that 'Other' includes at least 9 categories.

Now let's consider an ordinal variable, 'Risk of developing disease'. If the possible responses are 'Low', 'Low-Medium', 'Medium', 'Medium-High', or 'High' then we can see the categories have a natural order, Low comes before Low-Medium which comes before Medium, and so on on the scale of Low through to High.

We may then assess the levels of risk among a group of patients, and consider their distribution of risk, via a Bar chart, where each bar represents the frequency of each category- notice that the ordering of the categories is maintained- we don't reorder them based on frequency. Since we have an ordinal variable- we must retain the order!

We can describe the shape, centre and spread of such a variable.

We quickly see the shape is 'skewed to the right' - with the peak nearer the left, and then tailing out with decreasing frequency for categories to the right, the concentration (or where it is centered) is nearer the lower end of the scale. The spread is from low all the way through to high, we don't have values only at one end.

For a discrete variable, like number of electrical circuit failures per month ... or number of workplace injuries per month... we also construct a Bar chart, when we have relatively few possible responses. Or we may treat it more like a continuous variable when we have many possible responses and hence produce a histogram.

So, that's a brief look at the visual displays and descriptive measures used for each of the variable types. The next video looks at describing pairs of variables.

Now What?

Now that you have learnt about visual displays for a single variable you are ready to consider visual displays for two variables and consider hypothesis tests (see the videos [Visual Displays - Two Variables](#) and [Hypothesis Tests](#)).



But, when am I going to use this?

Across all fields of study we measure characteristics, or random variables. Accordingly, we need to present the data visually to understand patterns within the abundance of data and effectively communicate the key information. This may be used in organisational reports, presentations to key stakeholders, education or any other forum. Say less, show more!

Other Links

- Differences between bar graphs and histograms - more information about the bar graph and histogram showing the differences if you were to create each graph using the same data.

<https://www.youtube.com/watch?v=iYluquvGvAw>

- Bar charts, pie charts, histograms, stemplots, timeplots – detailed explanation of the creation of histograms with some information on other types of graphs

<https://www.youtube.com/watch?v=uHRqkGXX55I>

