

ETC1000 Business and Economic Statistics
Homework Week 1
(to be completed before your Computer Lab class in Week 2)

This homework is designed to help you:

- Explore the use of Business Analytics being used in practice
- Understand and work with data in an Excel spreadsheet
- Use Excel to draw a random sample
- Understand the importance of selecting a representative sample
- Describe numerical data
- Scale and adjust numerical data to enable meaningful comparisons
- Construct and work with pivot tables to describe categorical data
- Construct and interpret bar charts

A. Business Analytics in Practice

First, take some time to browse some of these websites to see examples of Business Analytics being used in practice:

<http://www.allanalytics.com/>

<http://www.scoop.it/t/big-data-technology-applications-and-analytics>

A1. Working with Data in a Spreadsheet

Open the *Timor-Leste* worksheet in the “Homework 1.xlsx” file in Excel. You should see a spreadsheet as follows:

	A	B	C	D	E
1	HH ID	HH Size	Electricity	Mobile Phone	Time to Road
2	1	7	No	No	0
3	2	10	No	No	0
4	3	2	No	No	0
5	4	3	Yes	No	0
6	5	7	Yes	No	0
7	6	12	Yes	Yes	0
8	7	3	Yes	Yes	0

This data is census data from a particular village in Timor-Leste – that is, the data contains some characteristics of every household that lives in the village.

The dataset contains 375 rows (1 row of headings and 374 rows of data) and 5 columns. Each row represents a household, and each column provides data on characteristics of the household. The column headings are short descriptions; more detail on what the data in the column captures can be

found by hovering your mouse over the little red triangle in the corner of the heading (called a “comment”). Look at the description in each of these comments – there are 2 things to note here:

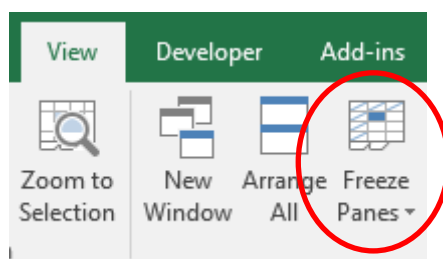
1. The data in each column was obtained by asking the household (typically 1 respondent answering on behalf of the household) a question – that is, this data was collected by surveying every household in the village.
2. The units of the data – number of, minutes, etc. These units will be important when it comes to understanding, analysing and making predictions with this data – particularly when reporting or interpreting the data.

Take a look at Household 1 in row 2. This household has 7 members, does not have an electricity connection, does not have a mobile phone, and lives on a vehicle-passable road. Household 374 in row 375 has 1 member, has electricity but no mobile phone, and lives 5 minutes’ walk from the nearest road.

Now let’s introduce some useful Excel tools to make looking at such a large dataset easier.

Freeze Panes

Freeze Panes allows you to keep headings visible while you scroll through the data. To freeze the headings in the first row of your spreadsheet, go to the *View* tab and when you click on *Freeze Panes*, select *Freeze Top Row* from the drop-down list.



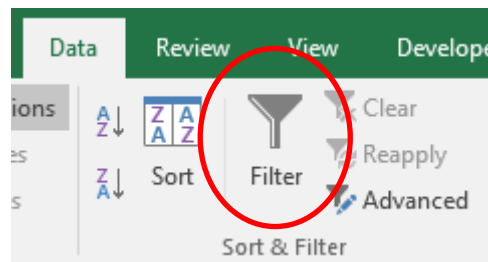
Now when you scroll down to row 375, you will be able to see the headings as well as the data.

Filter

Filter allows you to quickly look at subsets of the data. To enable filtering, select all 5 columns at once:

	A	B	C	D	E
1	HH ID	HH Size	Electricity	Mobile Phone	Time to Road
2	1	7	No	No	0
3	2	10	No	No	0
4	3	2	No	No	0
5	4	3	Yes	No	0
6	5	7	Yes	No	0
7	6	12	Yes	Yes	0
8	7	3	Yes	Yes	0

and from the *Data* tab click on the *Filter* button:



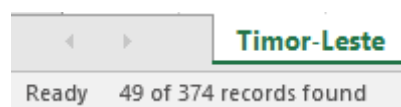
Your dataset should now look like this:

	A	B	C	D	E
1	HH ID	HH Size	Electricity	Mobile Phone	Time to Road
2	1	7	No	No	0
3	2	10	No	No	0
4	3	2	No	No	0
5	4	3	Yes	No	0
6	5	7	Yes	No	0

First click on the drop-down arrow for *Electricity*. Un-tick the *No* box and press *OK*. You will now only see the households that have access to electricity:

	A	B	C	D	E
1	HH ID	HH Size	Electricity	Mobile Phone	Time to Road
5	4	3	Yes	No	0
6	5	7	Yes	No	0
7	6	12	Yes	Yes	0
8	7	3	Yes	Yes	0
9	8	7	Yes	Yes	0
10	9	11	Yes	Yes	0

In the *Status Bar* at the bottom left-hand side of your window you will see:

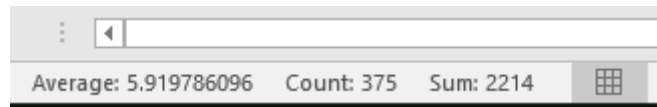


This tells us that 49 of the 374 households are visible – that is only 49 out of the 374 households in this village have an electricity connection, or 13.1%.

Click on the drop-down arrow for *Electricity* again and re-tick the *No* box to view all 374 household records (we call these our 374 ‘observations’).

Next click on the *HH Size* drop-down arrow. Notice the range of values that this household characteristic takes: 1-14 household members! Use the filter function to see how many households have 10 or more members (you should get 28 households, or 7.5% – note the use of units here).

Return to the full set of observations and select the entire *HH Size* column. The right-hand side of the *Status Bar* will show you the average number for the column.



This tells us the average household size in this village is 5.9 people.

One bonus of the Filter tools is that it enables you to perform a relatively 'safe' sort of the data. However, you must still ensure that you had selected all data (i.e. all columns) when enabling the filter.

Check that you can obtain the following:

- In this village, there is 1 household that lives 120 minutes' walk from the nearest road.
- Only 16 of the 374 households own a mobile phone.
- Of those 16 households who own a mobile phone, 7 do not have an electricity connection to charge it with - they must have nice friends and neighbours!
- The average time to walk to the nearest vehicle-passable road is 13.5 minutes.
- For those with electricity, the average time to walk is 4.1 minutes, while for those without electricity, the average is 14.9 minutes.

A2. Sampling

Remember this dataset represents a census – a list of all the households in this particular village. Suppose we want to find out more about the characteristics of households in this village, but do not have the budget to re-interview everyone – we want to take a sample that is representative of the whole population. How do we take a representative sample (of say, 50 households) from this list (population) of households? There are a few ways we could approach this, some a lot better than others.

(1) Take the first 50 households?

In this case, taking ('sampling') the first 50 households in the list would produce a particularly **BIASED** sample. Why? Take a look at the characteristics of the first 50 households in the spreadsheet. Household ID is a unique identifier – just a number – allocated to the household. But it turns out (as per the description) that this was essentially the order in which households were interviewed – imagine the census interviewer arriving in the village, interviewing house by house. Looking at the Time to Road column, the interviewer clearly started interviewing along the main road. So if we were to take the first 50 households, we would essentially be taking the 50 households close to the main road. And we know from our analysis above that those on the main road are quite different to those who do live away from it (much more likely to have electricity, for example).

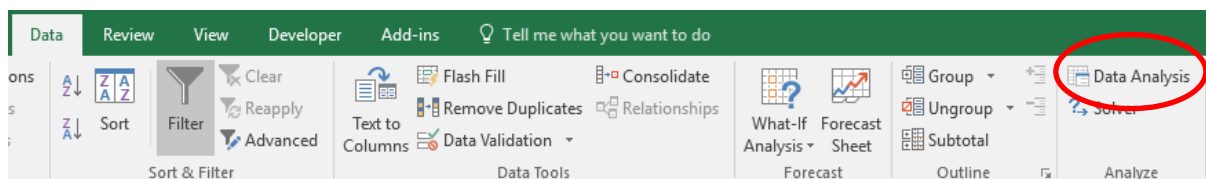
(2) Take every j th house?

This is an approach that is often used in practice, as it can be administered easily. We have 374 households in this village, so selecting every 7-8th house ($374/50=7.5$) will give us 50. If there is some kind of systematic ordering of the data, this method will not produce a representative sample.

(3) Use a random number generator

Excel can be used to draw a random sample. There are a few ways of doing this in Excel; we will use the random number generator in the Data Analysis ToolPak.

First, ensure you have the Data Analysis ToolPak add-in installed. If it is installed, it will appear as follows on the *Data* tab:



If the Data Analysis ToolPak is not yet installed, follow these steps to install it:

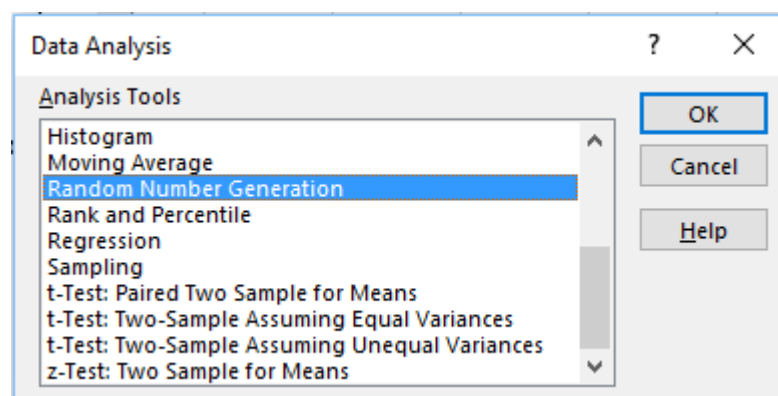
- On the *File* tab, select *Options* and *Add-Ins*.
- Press *GO* to manage Excel Add-ins, and select *Analysis ToolPak* from the list.

It should now appear under the *Data* tab.

The process for drawing a random sample of 50 households from our population of 374 is as follows: we assign each household a number randomly. What is important is that each household (row in our spreadsheet) has equal chance of being selected. Once each household has been assigned a number, we rank (sort) households based on their number, smallest to largest. The top 50 households in the list become our sampled households.

Now let's create 374 random numbers in Excel, one for each household in the population, and list them in Column F.

First, give Column F in your spreadsheet the heading "Random Number". From the *Data* tab, select *Data Analysis*, and *Random Number Generation*.



Complete the dialog box as follows:

Random Number Generation

Number of Variables: 1

Number of Random Numbers: 374

Distribution: Uniform

Parameters

Between 0 and 1

Random Seed: 1234

Output options

☒ Output Range: F2

☐ New Worksheet Ply:

☐ New Workbook

Buttons: OK, Cancel, Help

Here we are asking Excel to give us 374 draws (numbers) from a standard uniform distribution. The (0, 1) uniform distribution is a rectangular distribution in that all values between 0 and 1 are equally likely to occur (Google it!). Specifying the *Random Seed* ensures that you obtain the same numbers as this example.

Here is the output:

	A	B	C	D	E	F
1	HH ID	HH Size	Electricity	Mobile Phone	Time to Road	Random Number
2	1	7	No	No	0	0.124149297
3	2	10	No	No	0	0.006500443
4	3	2	No	No	0	0.389446699
5	4	3	Yes	No	0	0.267281106
6	5	7	Yes	No	0	0.703634754
7	6	12	Yes	Yes	0	0.235511338
8	7	3	Yes	Yes	0	0.466139714
9	8	7	Yes	Yes	0	0.74794763
10	9	11	Yes	Yes	0	0.123783074

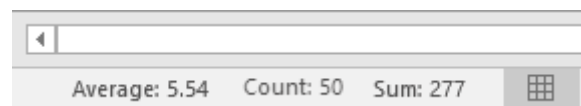
Expand the filter range to include Column F by selecting the column and going to *Data, Filter*. Looking at the values in the filter, we can see that Household 2 has been assigned the 4th smallest number. Use the filter on Column F to sort the data from smallest to largest. Your data should now look like this:

	A	B	C	D	E	F
1	HH ID	HH Size	Electricity	Mobile Phone	Time to Road	Random Number
2	32	7	No	No	3	0.000122074
3	332	4	No	No	35	0.004974517
4	133	8	No	No	2	0.005371258
5	2	10	No	No	0	0.006500443
6	234	8	No	No	10	0.009308145
7	51	1	No	No	5	0.010773034
47	160	8	No	No	3	0.111850337
48	227	4	No	No	10	0.112430189
49	34	9	No	No	3	0.115024262
50	14	4	Yes	No	1	0.121555223
51	65	9	No	No	2	0.122379223
52	9	11	Yes	Yes	0	0.123783074
53	1	7	No	No	0	0.124149297
54	180	4	No	No	5	0.127079073
55	334	6	No	No	35	0.127262185
56	142	5	No	No	2	0.12735374

(rows 8-46 have been hidden to save space)

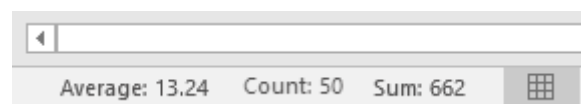
The first 50 households in this list are your sample – rows 2-51. Let's look at the characteristics of these 50 households compared to the population values we obtained earlier. If our sample is indeed representative of the population, then we should see similar characteristics in the sample to what we saw in the population.

HH Size: select cells B2:B51 so that the *Status Bar* shows the average value for HH Size for your sample of households:



The average number of household members for our sample is 5.5 people. This is reasonably close to the population figure we obtained earlier of 5.9 people – because it is a sample, we don't expect to see the exact population value.

Time to Road: select cells E2:E51. The *Status Bar* shows the average time to walk to the main road for your sample of households:



The average time of 13.2 minutes is very close to the population figure we obtained earlier of 13.5 minutes. This suggests that, geographically speaking, our sample would seem to be representative of the population from which we drew.

B. Data Types

In order to use data to adequately address a question of interest, we need to ensure the data is appropriate for answering the question, and that any summary tools being used are appropriate for the type of data.

Numerical data is data that takes a numerical value. It can be discrete (counted: e.g. number of people in a household, number of children in school, etc.) or continuous (measured: e.g. your height, the value of the Consumer Price Index (CPI), the unemployment rate, etc.).

Categorical data do not take on numerical values, but can be classified into distinct categories (e.g. country of birth, gender, day of the week, etc.).

B1. Describing Numerical Data

In this exercise, you will scale and adjust numerical data from the Australian Bureau of Statistics to make meaningful descriptions and comparisons using the data. The data you will use is Australian average weekly earnings.

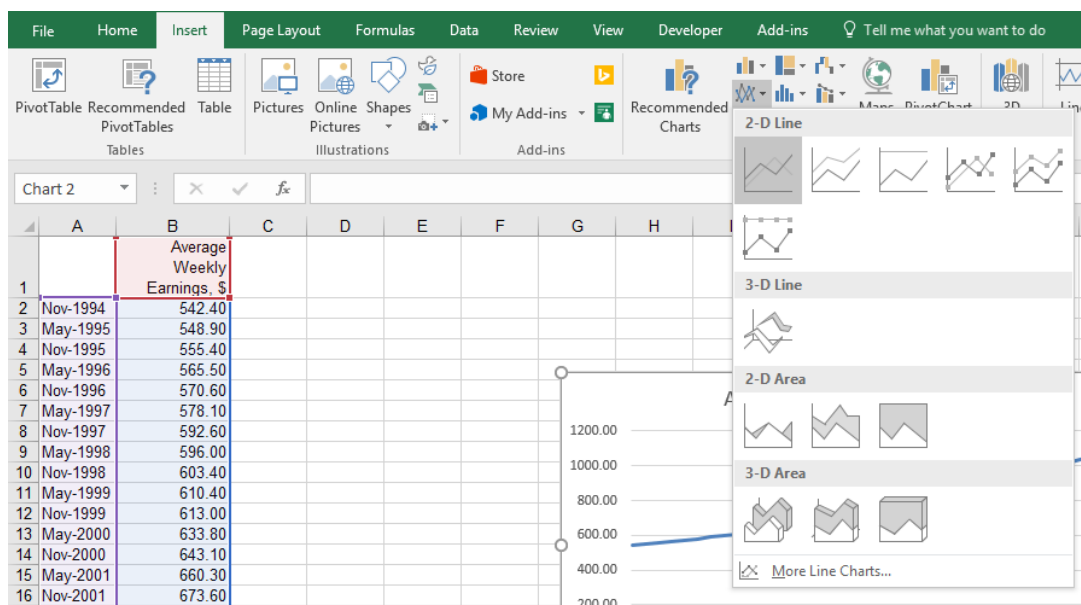
Nominal Wages

Open the *Wages* worksheet and look at the data. Notice the timespan, frequency and units of the data:

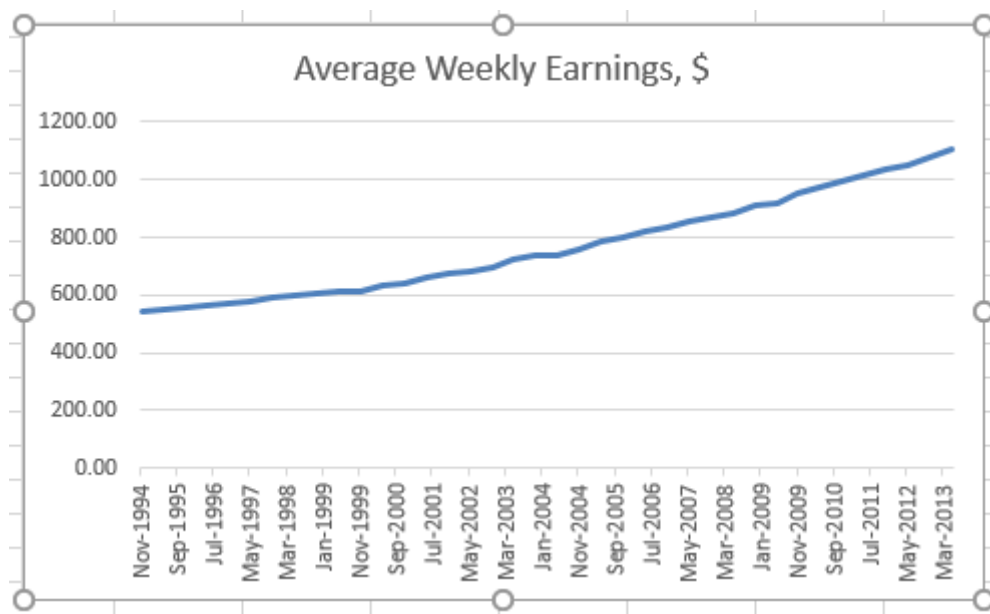
- The dataset contains bi-annual data spanning November 1994 to May 2013.
- Values are average weekly earnings in Australian dollars.

Graph the data in a line graph to see the sorts of values that the series takes:

- Select all the data, including headings, in columns A and B of the worksheet.
- On the *Insert* tab, click the *Line* icon and select the first *2D Line Graph* option.



The following chart will appear:



Play with the appearance of the graph to improve the presentation:

- Remove the decimal points from the vertical axis by right-clicking on the vertical axis and selecting *Format Axis*. Scroll to the bottom and expand the *Number* item, select the *Number* category and change the number of decimal places to 0:

The screenshot shows the "Format Axis" task pane in Microsoft Excel. The "Axis Options" tab is selected. Under the "Number" category, the "Category" is set to "Number", "Decimal places" is set to 0, and "Use 1000 Separator (,)" is checked. The "Negative numbers" section shows four options: "-1,234", "1,234", "(1,234)", and "(1,234)". The "Format Code" field contains "#,##0" and the "Add" button is visible. The "Linked to source" checkbox is unchecked.

- Change the horizontal axis so that it shows the date corresponding to the datapoints by right-clicking on the horizontal axis and selecting *Format Axis*. Under *Axis Options*, change the *Major* unit to Fixed 12 months:

Format Axis

Axis Options Text Options

Axis Options

Axis Type

☒ Automatically select based on data

☐ Text axis

☐ Date axis

Bounds

Minimum 01-11-94 Auto

Maximum 01-05-13 Auto

Units

Major 12 Months Reset

Minor 1 Months Auto

Base Months Auto

Vertical axis crosses

☒ Between dates

☐ At date 01-11-94

Your graph should now look like this:



The graph makes it easy to see the overall trajectory of average weekly earnings over time. In particular, we see that earnings have been rising over the 19 years.

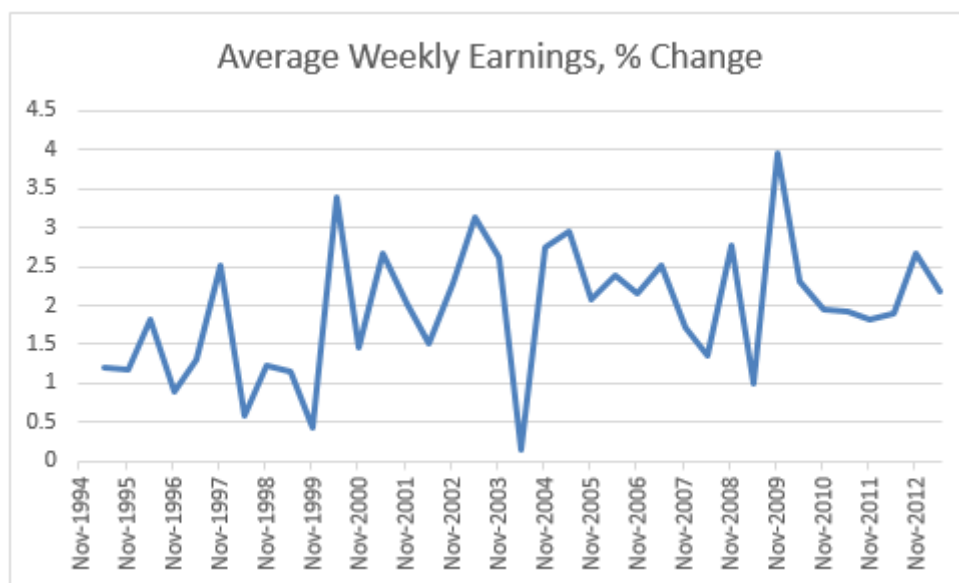
Converting to Percentages

Numerical figures are often most meaningfully described in percentage terms or rates, as this allows us to make relative comparisons across population sizes, company net worth, duration of time, etc.

Let's see whether the apparent increase in average weekly earnings is increasing at an increasing rate. For data that is observed over time (called a 'time series') such as this, a percentage change is calculated relative to the previous value of the series. That is, the percentage change in average weekly earnings between November 1994 and May 1995 is calculated as the difference in the value of the series, divided by (relative to) the value of the series in November 1994 (the earlier value). The resultant value of 1.2% therefore tells us that average weekly earnings in May 1995 increased by 1.2% over the 6 months.

C3			$=((B3-B2)/B2)*100$
	A	B	C
		Average Weekly Earnings, \$	% Change
1			
2	Nov-1994	542.40	
3	May-1995	548.90	1.2
4	Nov-1995	555.40	1.2
5	May-1996	565.50	1.8
6	Nov-1996	570.60	0.9
7	May-1997	578.10	1.3
8	Nov-1997	592.60	2.5

Graphing the percentage change values over time, we see that generally wages have been rising at an increasing rate (around the 1% mark in the late 1990s, but more like 2.5% by 2012). Average weekly earnings have never decreased in these 19 years, but the rate of change between 6-monthly recordings is quite variable.



Adjusting for Prices Using the CPI

When comparing numerical data across time, it is important to take into account that other factors may have also changed with time – a country's GDP may increase over time, but so does the population (workforce) that produces it. In the case of wages, even though earnings may have risen, so would have prices: some of the increase in earnings we see is due to inflation, and some is actual increase in earnings.

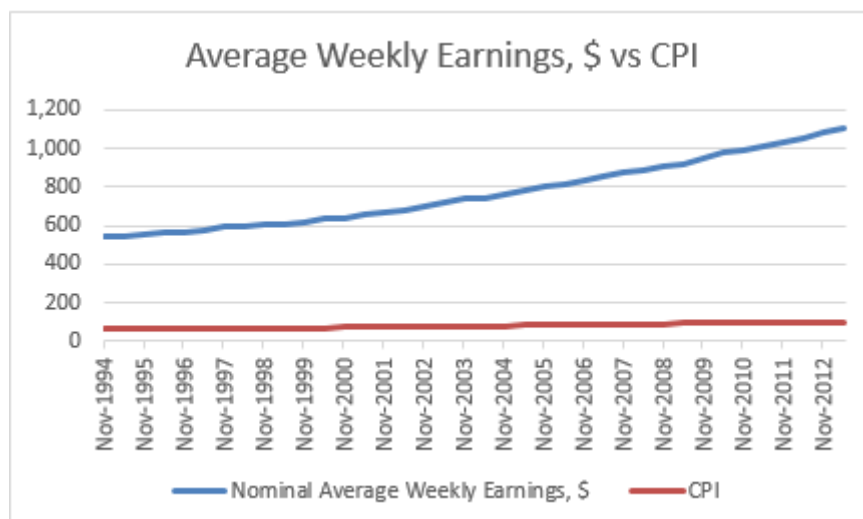
The CPI is an index that tells us how much prices have changed across time. When we adjust (nominal) wages by the CPI, we are essentially taking out the part of the increase that is due to inflation.

First let's take a look at the sorts of values that the CPI takes, in the worksheet named *CPI*:

	A	B	C
		Nominal Average Weekly Earnings, \$	CPI
1			
2	Nov-1994	542.40	62.8
3	May-1995	548.90	64.7
35	May-2011	1015.20	99.2
36	Nov-2011	1033.70	99.8
37	May-2012	1053.20	100.4
38	Nov-2012	1081.30	102.0
39	May-2013	1105.00	102.8

First note what the base period is – that is when CPI=100. Since we only have bi-annual values, the closest we have is May 2012 where the CPI value was 100.4 – the value of 100 probably occurred in the first quarter of 2012. Looking back to the 1990s, the series was in the 60s range during the 1990s – this tells us that, relative to those in early 2012, prices were much lower in the 1990s – something that cost \$1 in 2012 cost only \$0.63 in 1994.

Graphing the CPI series alongside our earnings data, we see that prices (the CPI) have risen alongside average weekly earnings.



How would average weekly earnings look if we were to adjust for the CPI? Based on the graph, the increase in earnings over time would be dampened, but not totally removed (the rate of increase in average weekly earnings seems larger than the rate of increase in the CPI).

Let's do the calculation to see.

We calculate *real* average weekly earnings by scaling *nominal* average weekly earnings by the CPI using the following formula:

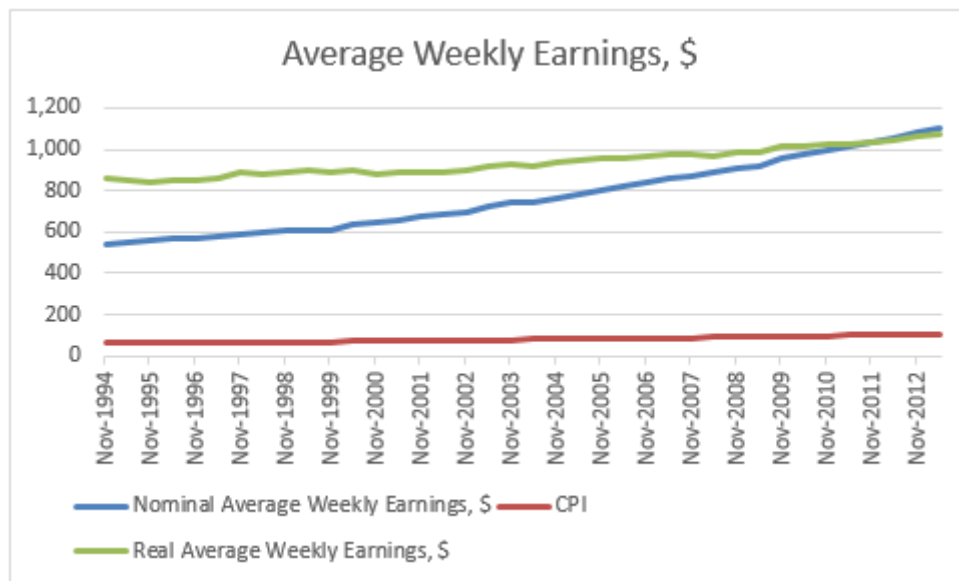
$$\text{Real Average Weekly Earnings} = \frac{\text{Nominal Average Weekly Earnings}}{(\text{CPI}/100)}$$

In Excel, we use:

D2 : ✕ ✓ <i>f_x</i> =B2/(C2/100)				
	A	B	C	D
1		Nominal Average Weekly Earnings, \$	CPI	Real Average Weekly Earnings, \$
2	Nov-1994	542.40	62.8	863.69
3	May-1995	548.90	64.7	848.38
35	May-2011	1015.20	99.2	1023.39
36	Nov-2011	1033.70	99.8	1035.77
37	May-2012	1053.20	100.4	1049.00
38	Nov-2012	1081.30	102.0	1060.10
39	May-2013	1105.00	102.8	1074.90

What do these values for real average weekly earnings mean? First, look at the value for May 2012. The CPI value in May 2012 is very close to the base value of 100, so when we adjust for prices, the nominal and real values are virtually the same. Back in November 1994, you could buy a lot more with \$1, so earning \$542.40 per week is worth more than it looks: you could buy an equivalent of \$863.60 in 2012 prices. By converting all the nominal earnings to their equivalent value in 2012, we can directly compare the real buying power of earnings over the 19 years.

Graph all 3 series (columns B, C and D) on the same graph:



Here we see that, indeed, the upward trend in average weekly earnings is dampened once we take into account inflation. While real average weekly earnings have increased over time, in terms of purchasing power (how much the wage can buy), they have not increased by as much as the nominal values suggest.

B2. Describing Categorical Data

We use different tools to describe categorical data than the ones used above.

Pivot Tables

A pivot table is mainly used to analyse categorical data, either univariate or bivariate. It is a tabulation showing the number of times each category appears.

Previously, we used Excel's filter tool to count the number of times (the frequency) a particular category, or groups of categories, occurred.

However, when we are interested in making comparisons across multiple characteristics of interest, or summarising the data for inclusion in a report, the filtering approach can become cumbersome and unhelpful. Instead, we can produce a summary table – a pivot table – using Excel.

The worksheet, *Travellers*, contains raw data collected from 3999 travellers as they arrive at Melbourne airport. The sheet contains the country (region) they came from and the main purpose of their visit (work, study or tourism), so there are two categorical variables to be examined: one is Region and the other is Purpose.

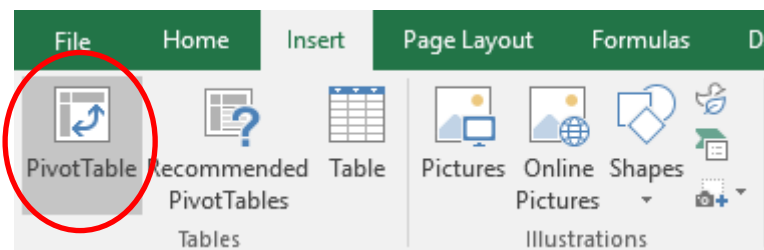
Univariate Pivot Tables

Suppose we are interested in the characteristic of Purpose at the moment. In this case a univariate pivot table will suffice. Follow these 4 steps to create your univariate pivot table for the categories of Purpose:

1. Open the spreadsheet and highlight the whole data range, A1: B4000. Make sure each column is labelled with the name of the categorical variable, and that no blank cells are included.

	A	B	C	D	E
1	region	purpose			
2	Middle East	work			
3	Middle East	work			
4	Americas	study			
5	Europe	study			
6	Europe	study			
7	Middle East	work			
8	Europe	work			
9	Europe	work			
10	Middle East	study			

2. Click the *Insert* tab, select the *PivotTable* button from the *Tables* group.

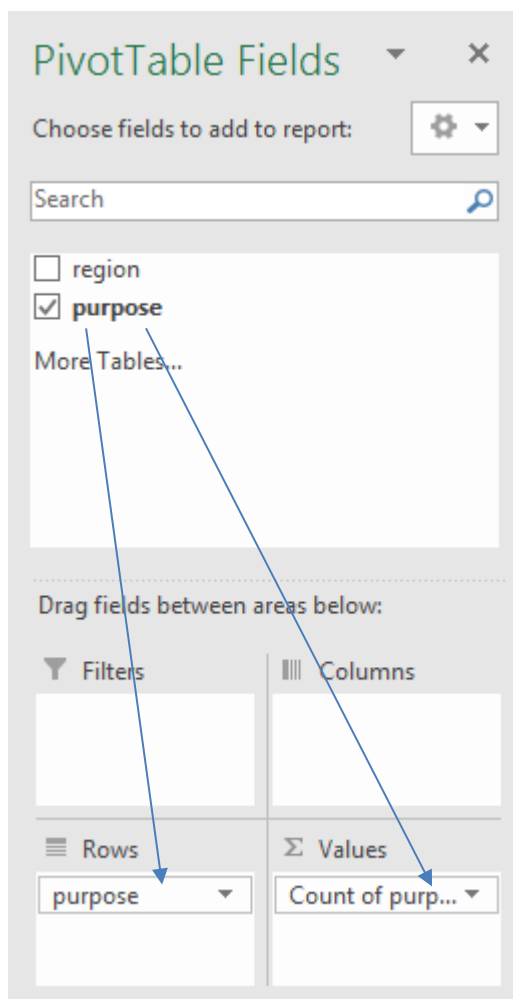


3. Edit the *Create PivotTable* dialog box:

The highlighted data should automatically appear under *Table/Range*, which is the input data to be used in the table (check that it contains *Travellers!\$A\$1:\$B\$4000*). Select where you want Excel to report the pivot table – in this example a *New Worksheet* – and click *OK*.

The screenshot shows the 'Create PivotTable' dialog box. The 'Choose the data that you want to analyze' section has the 'Select a table or range' radio button selected. The 'Table/Range' field contains the text 'Travellers!\$A\$1:\$B\$4000'. The 'Choose where you want the PivotTable report to be placed' section has the 'New Worksheet' radio button selected. The 'Location' field is empty. The 'Choose whether you want to analyze multiple tables' section has the 'Add this data to the Data Model' checkbox unchecked. The 'OK' and 'Cancel' buttons are at the bottom right.

4. A new worksheet opens with a blank pivot table, and you will see that the two categorical variables country and purpose have been carried over to the *PivotTable Fields* dialog box.
 - Drag the *purpose* label to the *Rows* quadrant.
 - Ignore the *Columns* quadrant for the moment since we are creating a univariate table. The left side of the Excel spreadsheet should show the three categories for the Purpose variable: study, tourism and work.
 - Tell Excel that you want to see the count for each category, which can be done by dragging *purpose* into the *Values* quadrant. Excel will work out the count of each category and update the results in the table.



	A	B
1		
2		
3	Row Labels	Count of purpose
4	study	1559
5	tourism	1247
6	work	1193
7	Grand Total	3999
8		
9		

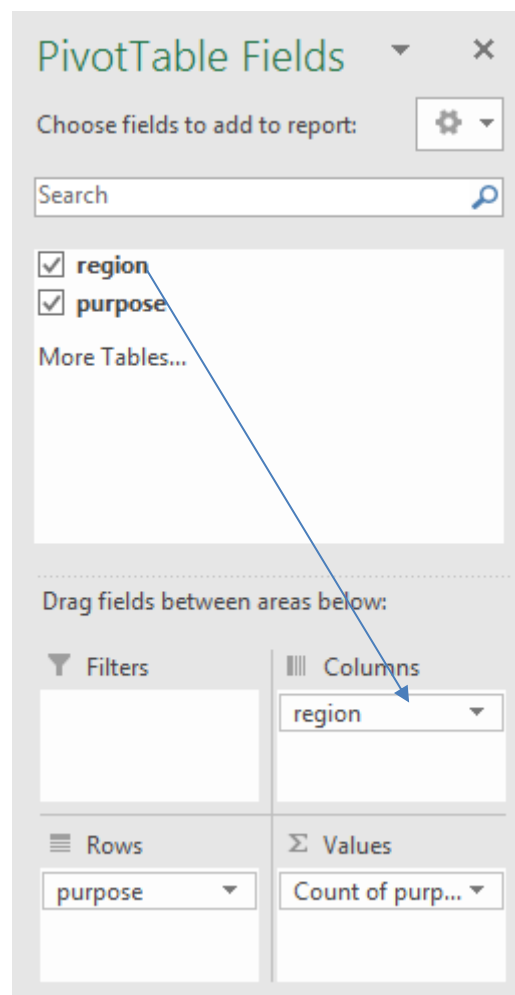
Having created the pivot table, the next step is to analyse it and this is the skill you are required to develop for this unit. Based on your pivot table, what could you say about the travellers' purpose for travel to Melbourne? It appears that travellers tend to visit Melbourne mainly for the purpose of study, and they are less likely to visit Melbourne for work.

At the beginning of this exercise, we just had raw information on 3999 travellers. It was difficult to extract any reliable information from the raw data without summarising it first. With the help of the pivot table, we are able to summarise the overall pattern in the data and see a bigger picture.

We can do deeper analysis of this data using a pivot table. For example, an interesting question to ask might be: do travellers from all regions tend to visit Melbourne for study? Alternatively, do Asian travellers and European travellers visit Melbourne for the same reason? To answer such questions, we need to examine the two characteristics of Purpose and Region together, and this is the focus of the next section.

Bivariate Pivot Tables

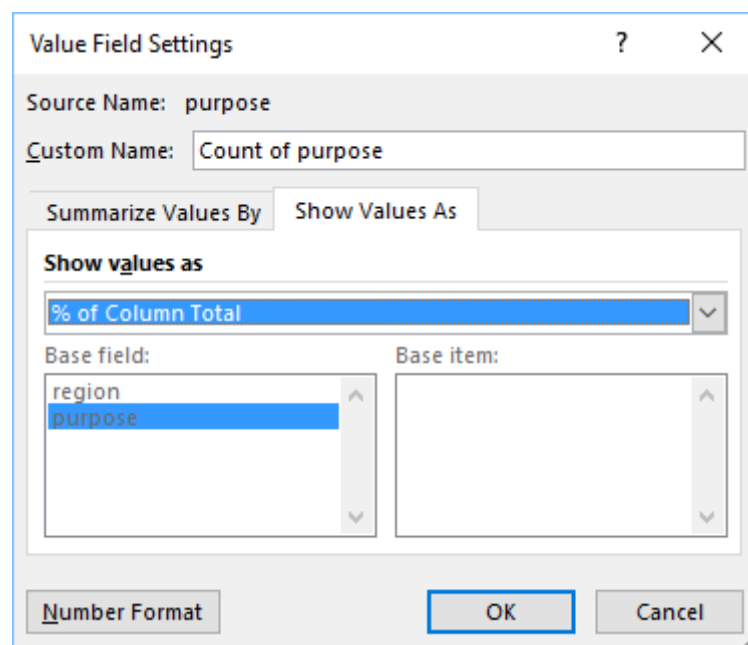
The way to tell Excel to break down each category of Purpose according to Region is simply to drag Region from the *PivotTable Fields* to the *Columns* quadrant. Excel will again work out the count for each entry and update the results.



	A	B	C	D	E	F
1						
2						
3	Count of purpose	Column Labels				
4	Row Labels	Americas	Asia	Europe	Middle East	Grand Total
5	study	181	831	225	322	1559
6	tourism	182	390	597	78	1247
7	work	32	373	411	377	1193
8	Grand Total	395	1594	1233	777	3999

Since each category of Purpose has been broken down by Region, we are provided with more information compared with that in the univariate table, which allows us to draw a more detailed picture of the data. For example, one can compare Asian and European travellers in terms of their purposes by examining the Asia and Europe columns. Some caution however, is required before directly comparing the values in the two columns. The column totals of Asia and Europe are different, that is, there are more travellers from Asia than from Europe, so the comparison across columns could be misleading without taking into account the different column totals. A possible solution is to standardise the data by converting count into proportion.

The conversion can be done manually, but Excel easily facilitates this by changing the setting in the *Values* quadrant. Click the label of *Count of purpose* in the *Values* quadrant, then select *Value Field Settings...* from the popup menu. You will see the following *Value Field Settings* dialog box, where you can change *Show values as* from *No Calculation* to *% of column* and click *OK*.



The pivot table presented in percentage of column is shown below – check that you have the same results. Note that the formatting and labels have been edited for presentation. What can we take away from the pivot table in this form? We can now correctly compare purpose of travel across regions. It is clear that in terms of traveling purpose, more than half of the Asian travellers visit Melbourne for study, while almost half of European travellers are tourists. To sum up, travellers from different regions visit Melbourne for different purposes, so the two characteristics certainly do not appear to be independent.

Purpose	Region				Grand Total
	Americas	Asia	Europe	Middle East	
study	45.82%	52.13%	18.25%	41.44%	38.98%
tourism	46.08%	24.47%	48.42%	10.04%	31.18%
work	8.10%	23.40%	33.33%	48.52%	29.83%
Grand Total	100.00%	100.00%	100.00%	100.00%	100.00%

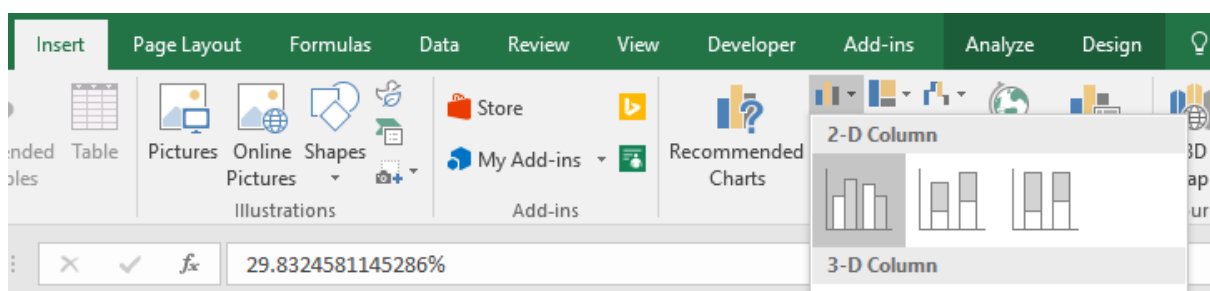
One can draw other conclusions by analysing other values in the table. It might also be a good exercise to present the table in percentage of row (% of row). Consider the difference between percentage of column and percentage of row, and in what situations one might be more appropriate than the other.

Bar Charts

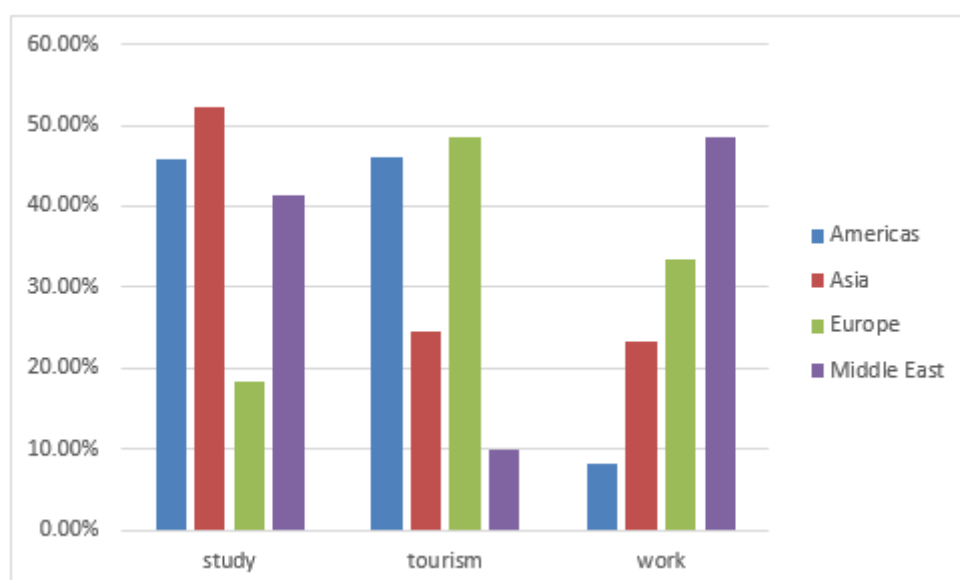
A bar chart is usually used for categorical data, showing the number or percentage of times each category appears. As a graphical representation of a pivot table, it is easy to interpret. This section illustrates how to draw a bar chart for the pivot table created by Excel, and how to label the chart and axis.

Creating a Bar Chart from a Pivot Table

1. Click any cell inside the pivot table to select the table.
2. With the table selected, click the *Insert* tab, select the *Column* icon from the *Charts* group and select the first option under *2-D Column*.

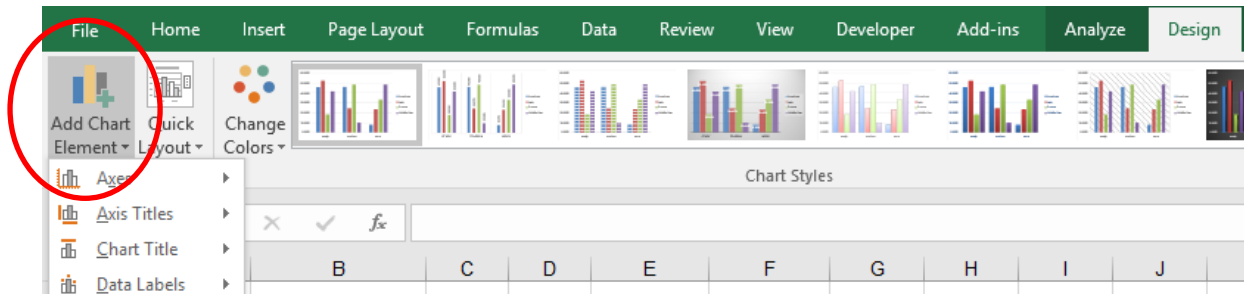


3. Excel will create a bar chart by default. The chart carries the same information as the pivot table, with the bar height proportional to the value in the table. Excel does not label the vertical axis clearly. Since the values in the table are presented in percentage of column, it is necessary to further edit the plot so that this sort of information is clear.

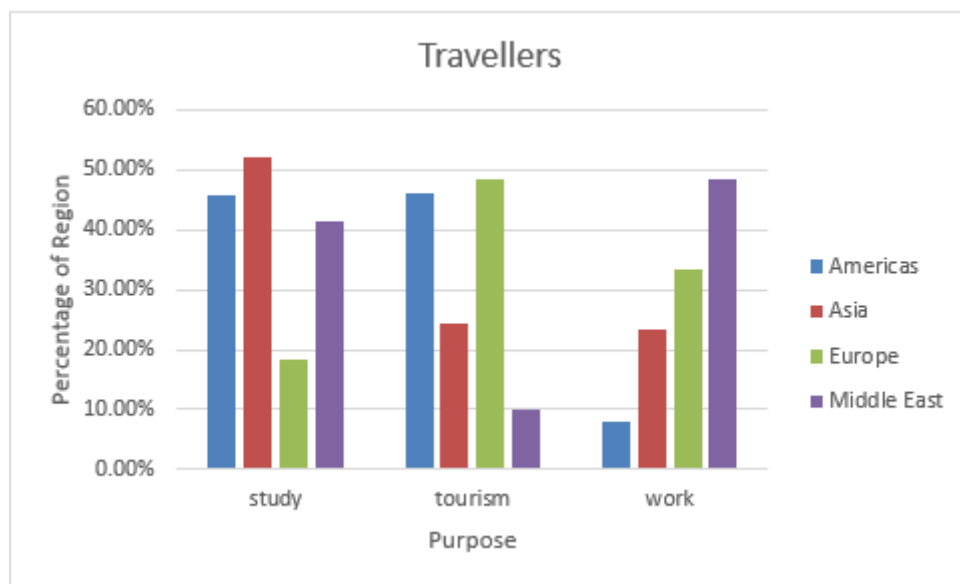


Bar Chart Layout

1. Click anywhere on the chart to select it.
2. With chart selected, click the *Design* tab, click the *Add Chart Element* button.



3. Use the *Axis Titles* and *Chart Title* options to add these to your chart.
 - Click *Chart Title* and type a meaningful title.
 - Click *Axis Title* – for the vertical axis type an appropriate title to indicate the unit on the axis.As previously mentioned, it's important to clearly state the units of your variables when presenting your results.



Now your chart is clearly and appropriately labelled. Do not disregard presentation in your graph — there is no point doing wonderful analysis if the results are not clear and meaningful to the reader. You can change many other aspects of the chart by right-clicking on various areas of the chart. Have a play with it yourself!