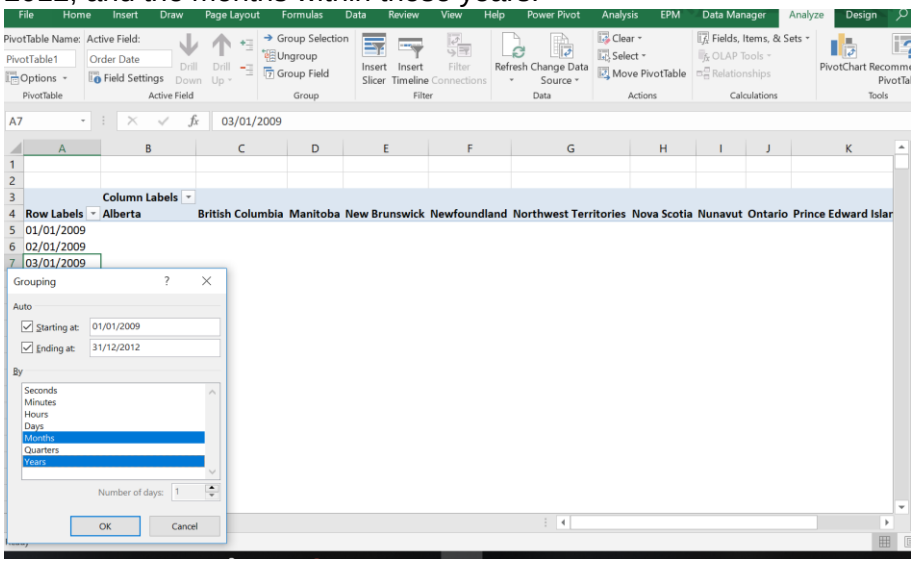
Lab 04 – Microsoft Excel

Introduction to Charts

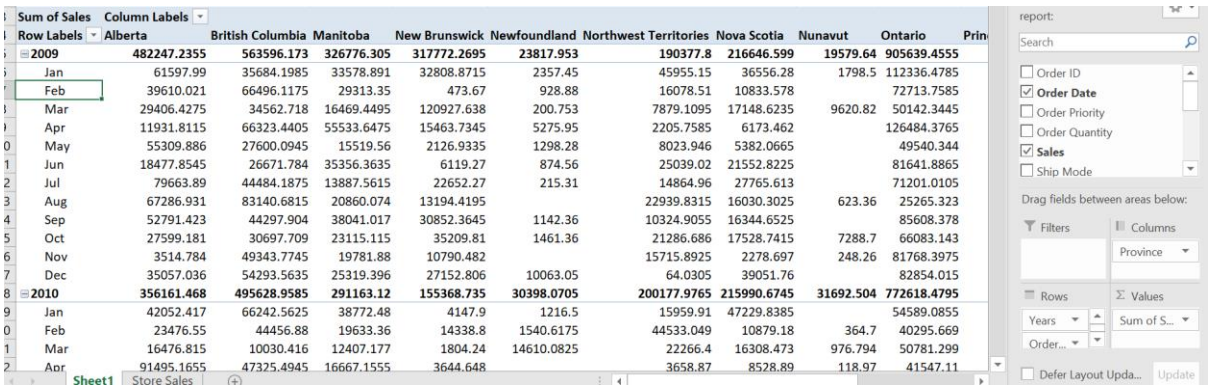
Here we will introduce the popular charts in Excel. Charts, also known as graphs, are a very useful way of visualizing data in Excel.

We will try popular charts such as line charts, column charts and pie charts using the **Store-Sales.xls data that you can download.**

- We will use a pivot table again for summarising the data.
- Summarize sales across different provinces and different time periods.
- Insert pivot table.
- Prepare the pivot table as follows:
- Since my order date is a particular date, you see a whole lot of dates in my vertical axis.
- Collapse them by putting my cursor anywhere on the vertical axis, right click, Group, and then group them by month and years. So years, and within those years, particular months. OK, the structure gets prepared with the four years, 2009 through 2012, and the months within those years.



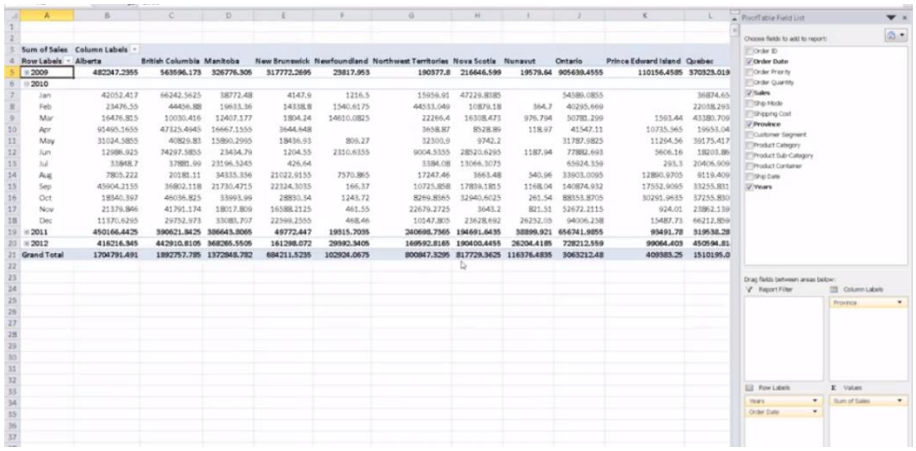
- Next bring sales field to the values box.
- The data summary is complete, where I've organized the sales data across different territories, across years, and within those years, the particular months:



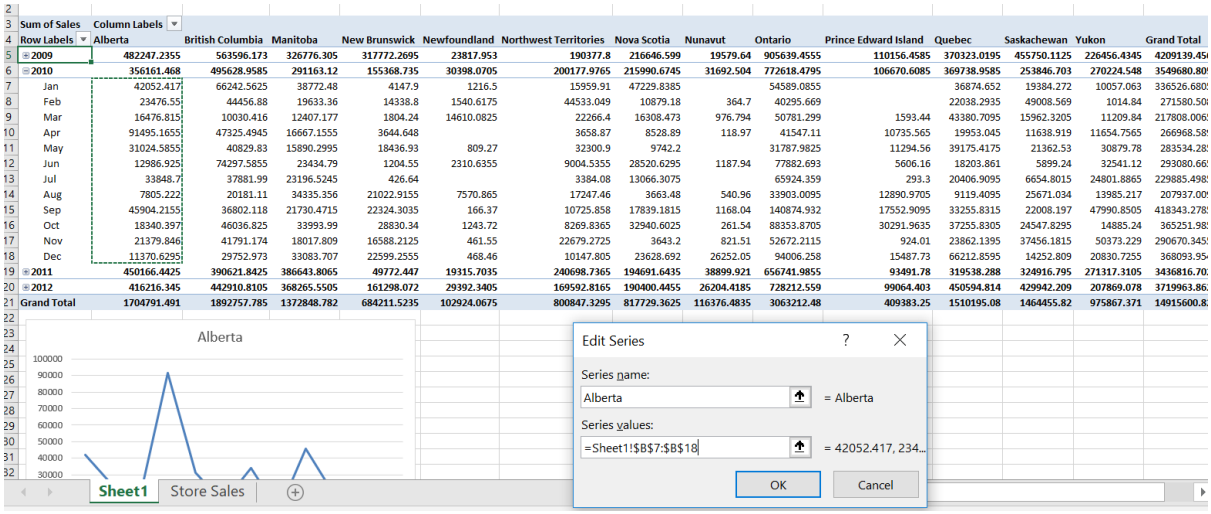
Sum of Sales	Column Labels									
Row Labels	Alberta	British Columbia	Manitoba	New Brunswick	Newfoundland	Northwest Territories	Nova Scotia	Nunavut	Ontario	Prin
2009	482247.2355	563596.173	326776.305	317772.2695	23817.953	190377.8	216646.599	19579.64	905639.4555	
Jan	61597.99	35684.1985	33578.891	32808.8715	2357.45	45955.15	36556.28	1798.5	112336.4785	
Feb	39610.021	66496.1175	29313.35	473.67	928.88	16078.51	10833.578		72713.7585	
Mar	29406.4275	34562.718	16469.4495	120927.638	200.753	7879.1095	17148.6235	9620.82	50142.3445	
Apr	11931.8115	66323.4405	55533.6475	15463.7345	5275.95	2205.7585	6173.462		126484.3765	
May	55309.886	27600.0945	15519.56	2126.9335	1298.28	8023.946	5382.0665		49540.344	
Jun	18477.8545	26671.784	35356.3635	6119.27	874.56	25039.02	21552.8225		81641.8865	
Jul	79663.89	44484.1875	13887.5615	22652.27	215.31	14864.96	27765.613		71201.0105	
Aug	67286.931	83140.6815	20860.074	13194.4195		22939.8315	16030.3025	623.36	25265.323	
Sep	52791.423	44297.904	38041.017	30852.3645	1142.36	10324.9055	16344.6525		85608.378	
Oct	27599.181	30697.709	23115.115	35209.81	1461.36	21286.686	17528.7415	7288.7	66083.143	
Nov	3514.784	49343.7745	19781.88	10790.482		15715.8925	2278.697	248.26	81768.3975	
Dec	35057.036	54293.5635	25319.396	27152.806	10063.05	64.0305	39051.76		82854.015	
2010	356161.468	495628.9585	291163.12	155368.735	30398.0705	200177.9765	215990.6745	31692.504	772618.4795	
Jan	42052.417	66242.5625	38772.48	4147.9	1216.5	15959.91	47229.8385		54589.0855	
Feb	23476.55	44456.88	19633.36	14338.8	1540.6175	44533.049	10879.18	364.7	40295.669	
Mar	16476.815	10030.416	12407.177	1804.24	14610.0825	22266.4	16308.473	976.794	50781.299	
Apr	91495.1655	47325.4945	16667.1555	3644.648		3658.87	8528.89	118.97	41547.11	

Next we will use some charting techniques to visualise this data:

- Place your cursor in an empty cell
- Insert ribbon- charts - line chart, the first option is the most common.
- You will get a blank box
- Draw your line chart from here
- We are interested in the 2010 data so collapse the other years:

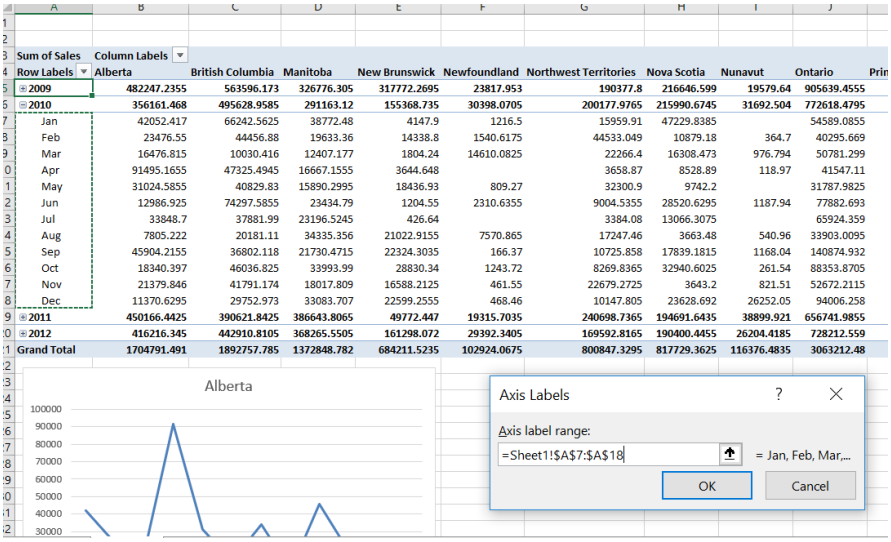


- Drag the chart box up.
- Right-click over box, Select Data > Add.
- Plot the monthly sales in the year 2010 for the province of Alberta.
- Give series a name, Alberta
- In the series values area, select the sales data for Alberta, the monthly sales data for that year, January through December. Select ok twice.



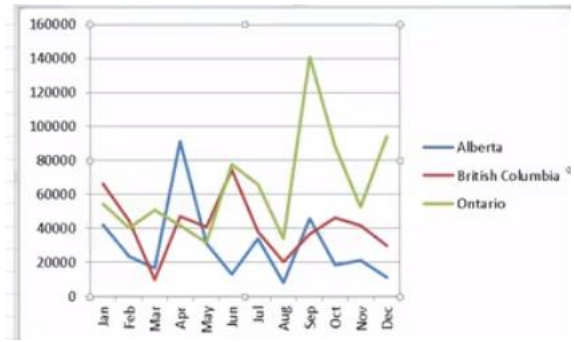
Line plot is drawn whereby I have sales on the vertical axis, the dollar sales volume, the horizontal axis has months labelled 1 through 12.

- You can change this numbering of months, rather than having it 1 to 12, January through December might be preferred.
- Right-click once again, select Data, and on the right hand side where it says Horizontal Category Access Labels, click on Edit, and Access Label Range, select January through December.



Sum of Sales	Column Labels									
Row Labels	Alberta	British Columbia	Manitoba	New Brunswick	Newfoundland	Northwest Territories	Nova Scotia	Nunavut	Ontario	Prin
2009	482247.2355	563596.173	326776.305	317772.2695	23817.953	190377.8	216646.599	19579.64	905639.4555	
2010	356161.468	495628.9585	291163.12	155368.735	30398.0705	200177.9765	215990.6745	31692.504	772618.4795	
Jan	42052.417	66242.5625	38772.48	4147.9	1216.5	15959.91	47229.8385		54589.0855	
Feb	23476.55	44456.88	19633.36	14338.8	1540.6175	44533.049	10879.18	364.7	40295.669	
Mar	16476.815	10030.416	12407.177	1804.24	14610.0825	22266.4	16308.473	976.794	50781.299	
Apr	91495.1655	47325.4945	16667.1555	3644.648		3658.87	8528.89	118.97	41547.11	
May	31024.5855	40829.83	15890.2995	18436.93	809.27	32300.9	9742.2		31787.9825	
Jun	12986.925	74297.5855	23434.79	1204.55	2310.6355	9004.5355	28520.6295	1187.94	77882.693	
Jul	33848.7	37881.99	23196.5245	426.64		3384.08	13066.3075		65924.359	
Aug	7805.222	20181.11	34335.356	21022.9155	7570.865	17247.46	3663.48	540.96	33903.0095	
Sep	45904.2155	36802.118	21730.4715	22324.3035	166.37	10725.858	17839.1815	1168.04	140874.932	
Oct	18340.397	46036.825	33993.99	28830.34	1243.72	8269.8365	32940.6025	261.54	88353.8705	
Nov	21379.846	41791.174	18017.809	16588.2125	461.55	22679.2725	3643.2	821.51	52672.2115	
Dec	11370.6295	29752.973	33083.707	22599.2555	468.46	10147.805	23628.692	26252.05	94006.258	
2011	450166.4425	390621.8425	386643.8065	49772.447	19315.7035	240698.7365	194691.6435	38899.921	656741.9855	
2012	416216.345	442910.8105	368265.5505	161298.072	29392.3405	169592.8165	190400.4455	26204.4185	728212.559	
Grand Total	1704791.491	1892757.785	1372848.782	684211.5235	102924.0675	800847.3295	817729.3625	116376.4835	3063212.48	

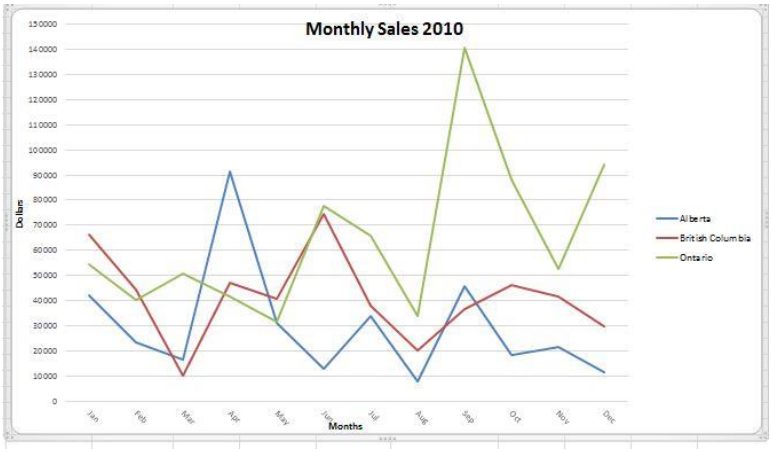
- Which month had the highest sales? _____
- Next, add the sales in the province British Columbia as well as the province Ontario.
- Right-click on my line chart, select Data. Add the province of British Columbia, give the series a name, select the sales data for British Columbia, January through December. Add, series name Ontario, select the sales data for Ontario, January through December.



- Which province has highest sales? _____
- After noticing the trend in sales… insert the following:

  - Chart title – Monthly Sales 2010
  - Y axis title ‘Dollars’
  - Change the Y axis range to 10,000. Click Y axis and choose Format Axis – change major unit to 10,000
  - Give a title to X axis, ‘Months’.
  - You could change alignment of X axis where months are at a 45 degree angle. Select X axis, right click, format…

- Legend – right click, format, you can change its position.
- You can change colours of line by selecting formatting, line style.

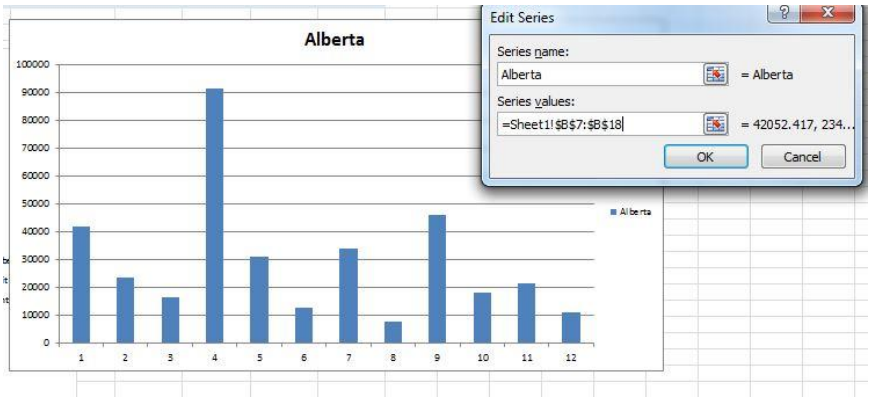


The next section focuses on two other popular charts: Column and Pie charts.

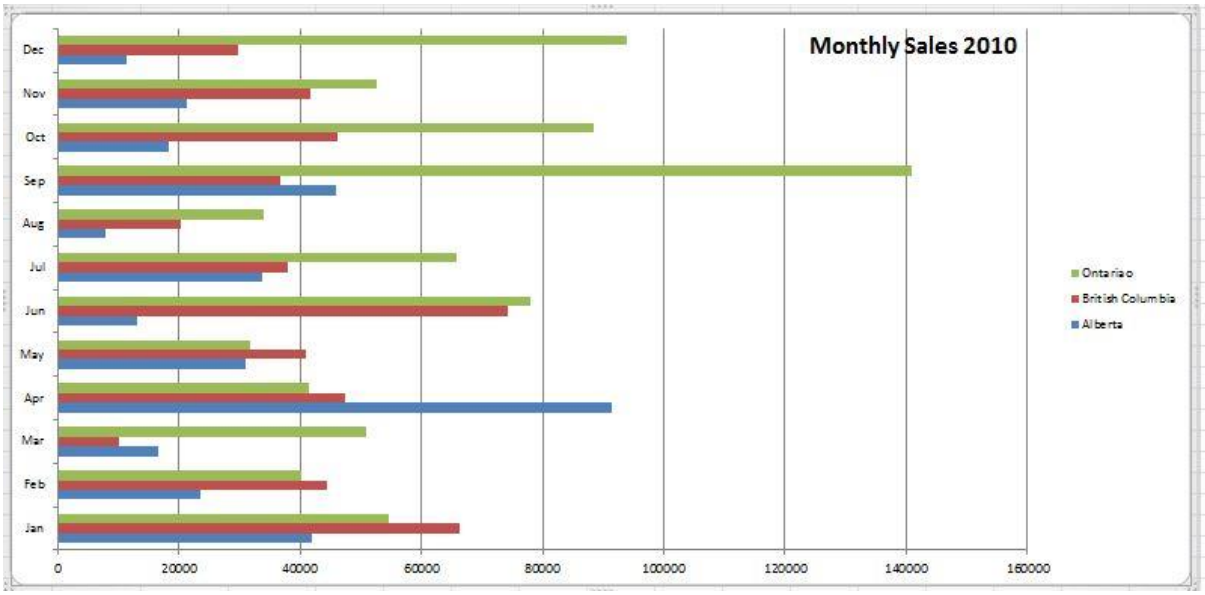
Insert Column Chart

Using the same data, go to an empty cell and choose insert, column chart. The first one is the most common.

- Same as before, you have a blank box. Right click and choose ‘select data’.
- Name the series ‘Alberta’
- Select the sales values from January to December.



- Continue with selecting British Columbia and Ontario.
- Change the horizontal axis to months labels instead of the numbers.
- Give the chart a title
- Change the orientation of bar chart – chart type, try horizontal bar:

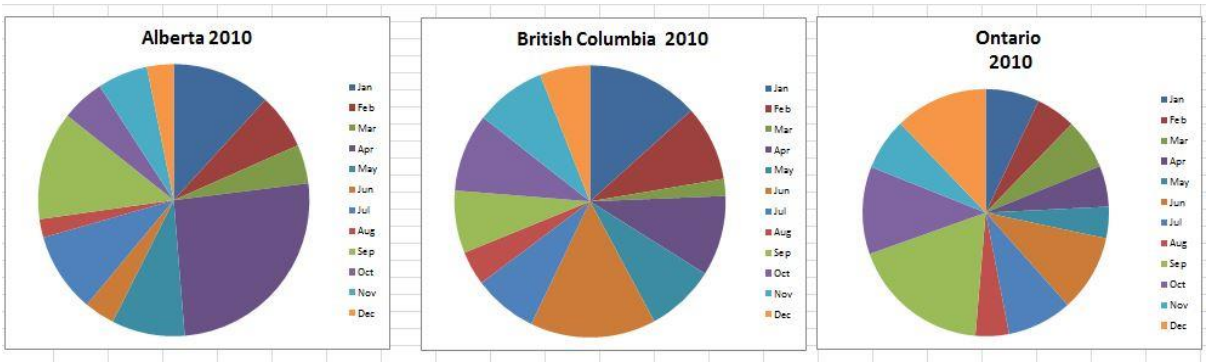


- This can be changed to a line chart or other charts

Pie Chart

A pie chart is a circle divided into sections, or pie slices representing a proportion of the whole pie.

- Create a pie chart for the monthly sales in 2010 for Alberta
- Give the series the name Alberta
- Again, make sure the months names are used instead of numbers
- You will notice the individual pie slices represent the proportion of sales in that particular month.
- What month has max sales? _____
- What month has min sales? _____
- If you wish to plot similar pie charts for the other 2 provinces British Columbia and Ontario, you need to do separate pie charts for each.

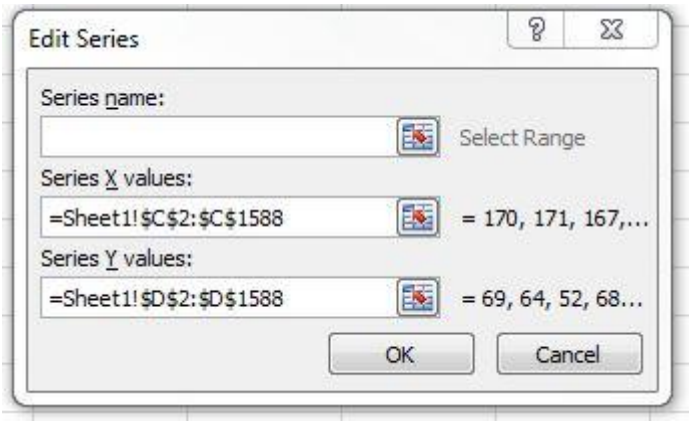


Scatter Plot

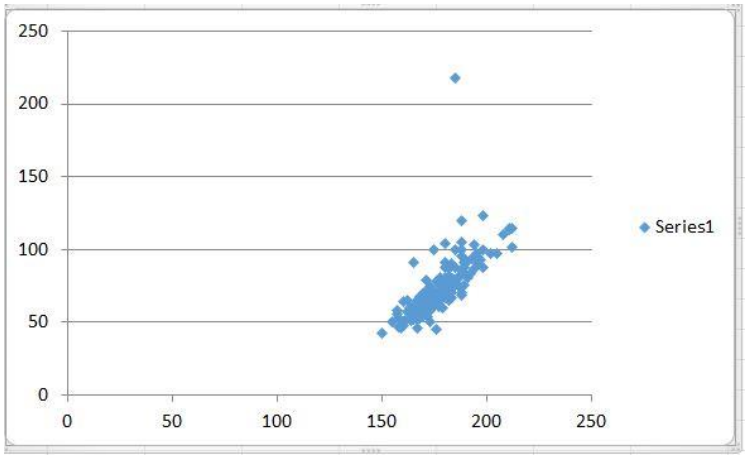
We will use Scatter Plot charts to investigate the relationship between two variables. Also known as the XY plot. We will learn how to plot and interpret a Scatter Chart in Excel.

In addition to showing the relationship of two variables, it shows if the relationship is positive or negative or no relationship. The two axis in a scatter chart represent the two variables. And, the coordinates of every point on the chart, are the values of the two variables.

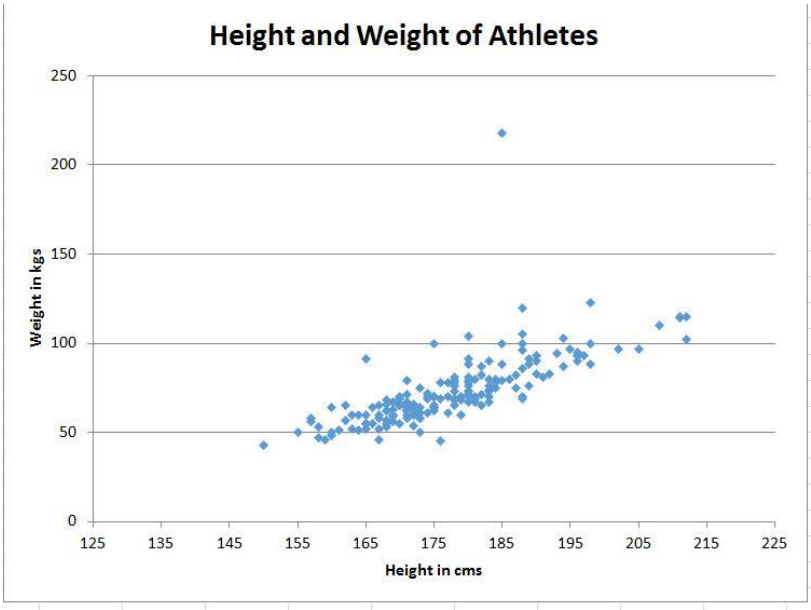
- Download the file **Height and weight**.xls
- This file contains the names of certain Olympic athletes from a recent Olympic Games. Specifically what is contained is the name of the athlete, the second name, the first name, the height of the athlete in centimeters, the weight of the athlete in kilograms, the country, and the gender of the athlete.
- We are interested in studying the relationship between height and weight, to do this we will plot the two variables on a scatter plot
- Go to an empty space
- Insert- scatter chart or scatter plot, choose the first.
- Right click the blank area, select data, add
- Specify the horizontal axis and what goes into vertical axis
- Height for horizontal and weight for vertical.



- Select ok twice.
- The height is on the horizontal axis. And the weight of the athlete in kilograms is on the vertical axis:

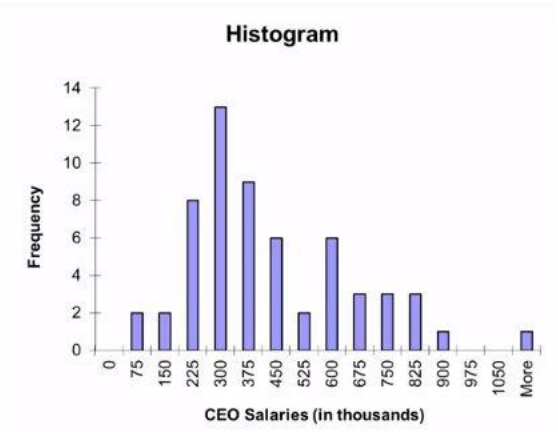


- To make the presentation a bit more visually appealing, minor modifications of this chart are required.
- Add Chart title
- Add the horizontal axis labels, Layout >Axis Titles >Primary Horizontal Axis Title >Below Axis and call it height in centimeters.
- Vertical axis labels - Go to Layout, Axis Titles, a rotated title. This is the weight in kilograms.
- As you can see the points are clustered together. Change the horizontal axis where it begins at 125cm. Right click, format axis, min set to 125cm
- So what does this trend indicate?
- As expected, grading the height of a athlete we can see that the weight also increases. This visual representation of this scatter plot tells me that there's a positive relationship between the height of an athlete and his or her weight. How strong is it? That's for another section.



Histogram

- Histogram are less intuitive than column charts and other graphs
- It is not as easy to create using Excel as some of the other charts.
- Histogram are similar to statistical distributions which will come again in another section, knowing about them now will help with later work.
- Consider the following histogram, a list of CEO salaries of small businesses.
- The basic idea of this histogram is to pictorially show how many salaries are low, high, and in between.

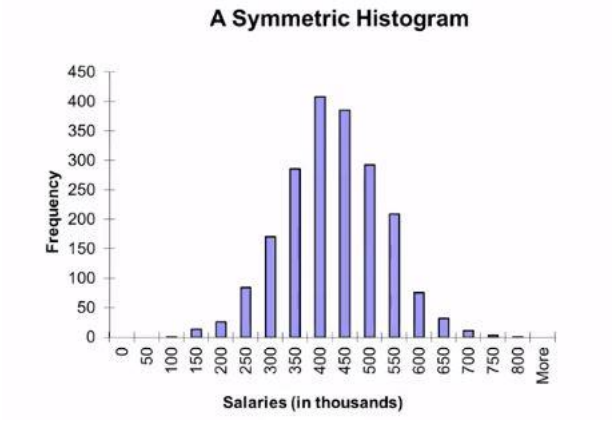


- The horizontal axis gives the salaries and the vertical axis tells how many salaries were approximately to those on the horizontal axis.
- For instance, it appears that there were two salaries around $75,000, two around $150,000, eight around 225,000 and so on.
- What does it mean when we say there were two salaries around $75,000?
- Excel here shows boundaries of ranges. The first range or band goes from 0 to $75,000. That is, there were two CEOs with salaries between 0 and $75,000.
- Similarly there were two CEOs between $75,000 and $150,000. Two CEOs having salaries between $75,000 and $150,000. This means there is two CEOs having salaries greater than $75,000 and less than, equal to $150,000.
- Similarly, eight CEOs with salaries greater than $150,000, but less than or equal to $225,000.

This example of a histogram shows relative frequencies of observations that fall into set bins, also known as classes or ranges. It is quite typical of income histograms. Some CEOs have very high salaries, but no CEO has an extremely low salary.

This histogram is an example of a skewed distribution, skewed to the right distribution. It is not symmetric. A symmetric distribution has roughly the same number of observations on either side of the mean.

Here is an example of a symmetric distribution:



If I were to slice this distribution somewhere in the centre. I would get approximately two mirror images on both sides of the centre. This is not real data. It's made up to give you an illustration of a symmetric histogram.

Construct a Histogram using data on salaries.

- Download the file called CEO-Salary_small-firms.xls
- This file contains the salaries in units of $1,000
- Click on Data ribbon and then Data Analysis.
- If you do not see the icon Data Analysis when you click on Data, you need to make sure that the icon is added to your spreadsheet. Go to:
    - File > Options > Add Ins. At the bottom you go to Manage Excel Add Ins
    - Make sure the Analysis ToolPak and the Analysis ToolPak VBA both are checked
    - You may have to close your worksheet and open it again
- Before you click Data Analysis, you need to set up bins.

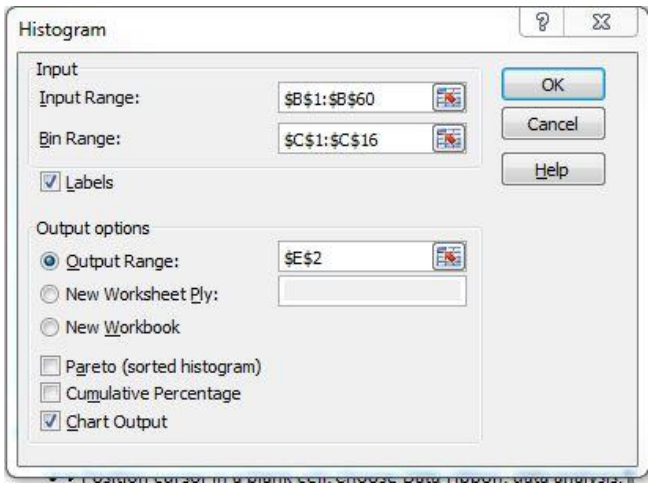
Set up Bins:

- The bins should be so set up so 75that all the data gets represented under one bin or another.
- Begin by putting in a 0 in C2.
- The Bin interval will be $75,000 so the next number is 75 in cell C3.
- The first bin, goes form 0 to 75, and the same bin interval spread of $75,000 is used for each
- Copy and paste it all the way down, this is done by select the 0 and 75, drag it until my bins are increasing to 535, 675. Drag and stop at 1050. There's no reason why we stop here.

	A	B	C
1	S.No	Salary '000$	Bin
2	1	621	0
3	2	145	75
4	3	736	150
5	4	808	225
6	5	149	300
7	6	659	375
8	7	208	450
9	8	573	525
10	9	482	600
11	10	750	675
12	11	300	750
13	12	339	825
14	13	543	900
15	14	296	975
16	15	350	1050
17	16	58	
18	17	350	
19	18	572	
20	19	155	
21	20	370	
22	21	298	
23	22	204	
24	23	1103	
25	24	362	
26	25	536	
27	26	388	
28	27	206	
29	28	21	
30	29	802	
31	30	862	
32	31	242	

Prepare the Histogram:

- Position cursor in a blank cell. Choose Data ribbon, data analysis.
- Select Histogram
- Specify input range **Salary '000$** (the data which I wish to plot in the histogram)
- Tick labels, this will allow the column name to appear.
- Next specify bin range.
- Out put options, arrange the output to an empty cell
- Check the chart output so the Histogram will be plotted
- Click ok.
- Widen the Histogram chart.
- Edit the horizontal title to CEO Salaries (in thousands)



- A few observations: the table that's produced next to the Histogram is also known as a Frequency Table. It tells me the distribution of the number of observations across those specs. E.g. there are 2 CEO's with salaries in excess of 0 dollars and less than or equal to $70,000. Similarly, there are 9 CEOs with a salary greater than $300,000 but less than or equal to $375,000. And, there is a single CEO with salary in excess of $1.05 million.
- One final point about setting up your bins. How big a bin interval should I choose? In this example the bin interval was $75,000. But as a general practice, there is no one particular rule in choosing your bin interval. However, you should choose bins such that you get a good visual representation of the data. Too many bins lead to bad Histograms. Too few bins also lead to bad Histograms. So, try out a different set of bins and go along with the bin interval or the set of bins that gives you the best visual representation

Bin	Frequency
0	0
75	2
150	2
225	8
300	13
375	9
450	6
525	2
600	6
675	3
750	3
825	3
900	1
975	0
1050	0
More	1