GRADE 11 ESSENTIAL
UNIT D – STATISTICS

UNIT NOTES

INTRODUCTION

Statistics is one of the many branches of Mathematics. So far you have worked with **Arithmetic** and **Measurement** and **Geometry** and **Algebra** and **Trigonometry**; special branches of mathematics. The branch of **Probability** is very closely related to **Statistics** as well.

One simple definition of statistics is the '*collection, processing, and display of information*'.

If you are adept at Statistics it is possible to credibly and readily fool most people about anything most of the time! "*There are three kinds of lies: lies, damned lies and statistics*"; Mark Twain.

In this course though we will just look at how to display information graphically. You likely learned some collecting and processing of information and surveys and data in Grade 9 Foundation Mathematics.

Objectives. The objectives of this unit are to learn about graphing bar graphs; histograms; line graphs; and circle graphs.

Prior Knowledge. You should be familiar with how to read a table, do percentages and fractions, plot points on a Cartesian grid, and how to use a protractor. A faint familiarity with the EXCEL spreadsheet program is desirable also.

EXAMPLE OF GRAPHING STATISTICS

Here is the result of asking **100** adult women their **shoe size**!

4	6	8	7	7	5	10	7	5	6
8	8	5	7	7	10	6	10	7	11
7	4	9	5	9	8	6	6	5	9
9	9	4	4	9	8	4	10	7	7
5	7	7	11	4	9	9	6	11	5
5	7	6	8	6	9	8	9	10	9
8	6	8	5	6	6	6	10	8	8
6	8	7	7	5	8	10	4	11	8
7	5	8	6	8	6	7	8	5	5
7	8	7	7	6	7	6	6	9	9

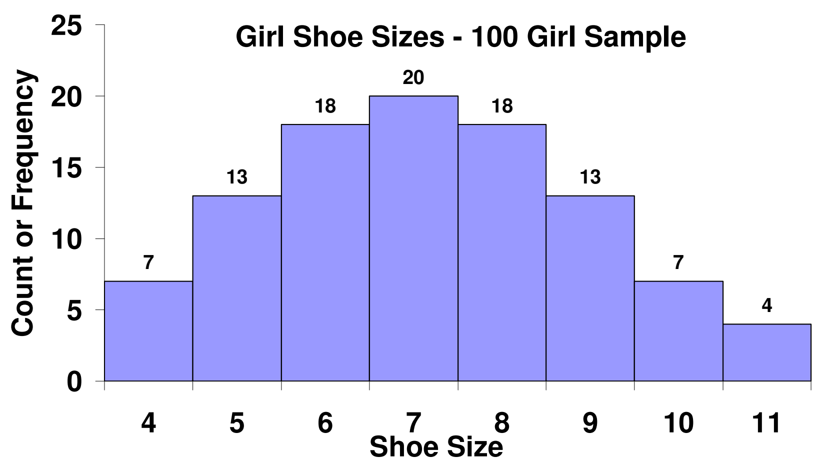
That is a nice mess of numbers that does not readily tell you much!

Often we arrange the numbers into groups and count the number in each group, so below is a **table** in which we **count** the occurrence (or frequency) of each shoe size:

Shoe Size	4	5	6	7	8	9	10	11	
Count	7	13	18	20	18	13	7	4	

You likely are familiar with such frequency tables and tally charts.

But **better yet** is a picture of this same shoe size data presented as a graph.



Now you instantly see the average (*mean and median*) size is about a size **7**, the most common size is a size **7** and that the bigger and smaller sizes are less common (less frequent). You see the **minimum** size, the **maximum** size, and the wide **range** of sizes from size 4 to size 11. You may recall from former studies we had calculations like **mean**, **median**, and **mode** to describe this distribution of data. Regardless of what calculations and processing you do with the numbers, the graph tells you lots of information intuitively and readily.

Graphs are a powerful way to show lots of information (and mis-information too!) in a readily understandable graphical format!

BAR GRAPHS

Bar Graphs are useful for displaying the result of **any** category of thing (not necessarily numeric): chocolate bars; music preferences; favourite colours; favourite hockey teams, etc.

Example: We did a survey of 20 people and asked them their favourite colour. We gave them five choices; **Red**, **Orange**, **Yellow**, **Green**, **Blue**, or **Brown**.

The results were recorded as a 'tally'; a tick for each response in that category; and then a final numerical count of the ticks in each category.

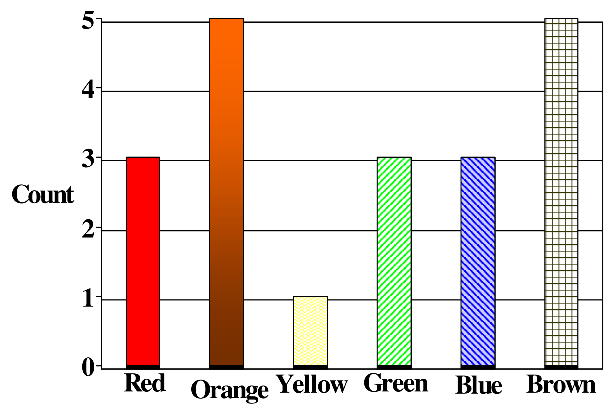
Category	Tally	Count
Red	///	**3**
Orange	卌	**5**
Yellow	/	**1**
Green	///	**3**
Blue	///	**3**
Brown	卌	**5**

Then we can prepare a **bar graph**. Categories along the bottom (horizontal side) axis, and count (or frequency) along the vertical side (axis).

Any graph should be nicely labelled with what each axis represents and what the entire graph represents.

Of course colours and different *textures* or *shading* of the bars make for a nice looking graph.

Count of Favourite Colours of the Class



Notice you can change the *texture* and the *pattern* or '*cross-hatching*' of each category bar to make it classy. You can change the colour too if you have a colour printer or colour pencils. Try not to get too fancy though!

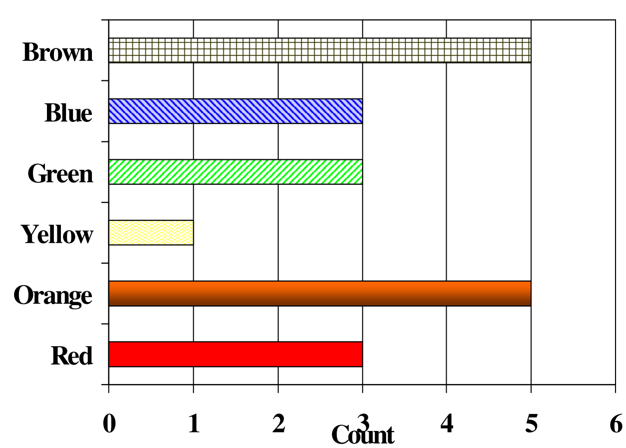
Horizontal Bar Graph

The exact same information could be represented going sideways as a Horizontal Bar Graph.

Don't waste too much making your graph fancy, the proper representation of the data is the goal, not the art work.

This one if you are viewing it in colour is somewhat *chromatically offensive*!

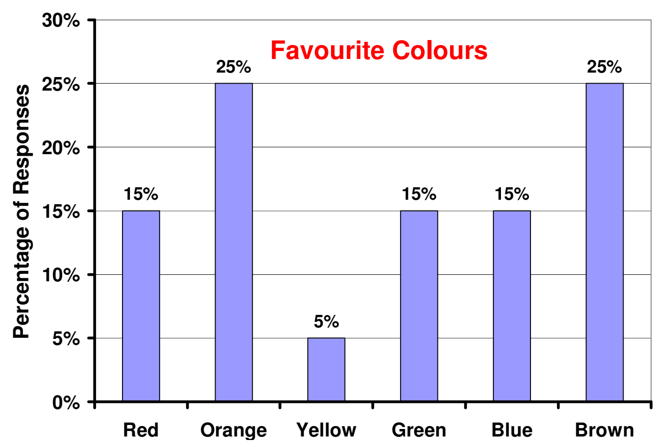
Count of Favourite Colours of the Class



Frequency: Count vs Percentage

Sometimes the frequency of the observations is represented by what **percentage** of the whole data set each category is.

Category	Tally	Count	Percent
Red	///	**3**	**15%**
Orange	~~////~~	**5**	**25%**
Yellow	/	**1**	**5%**
Green	///	**3**	**15%**
Blue	///	**3**	**15%**
Brown	~~////~~	**5**	**25%**
	TOTAL	**20**	**100%**

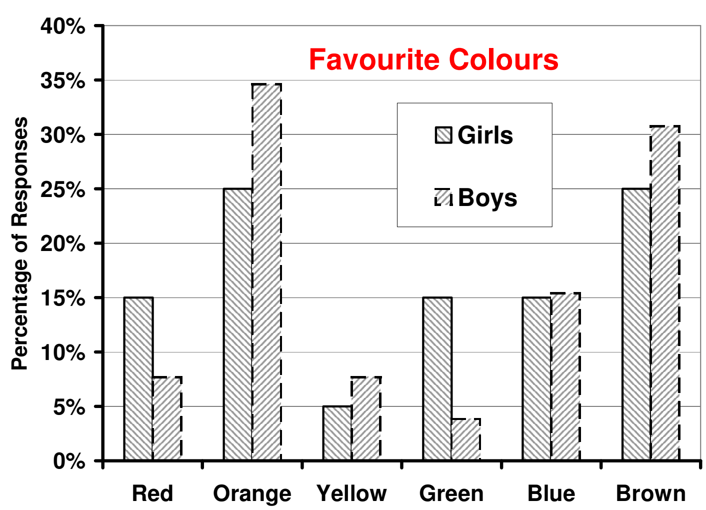


MULTI-BAR GRAPHS

Sometimes your data is two separate but related sets of data that you want to compare such as **girls** favourite colour **vs** **boys** favourite colour. The data and the graph might look like this:

	Girls		Boys	
Red	3	15%	2	8%
Orange	5	25%	9	35%
Yellow	1	5%	2	8%
Green	3	15%	1	4%
Blue	3	15%	4	15%
Brown	5	25%	8	31%
TOTALS:	20		26	

Notice a **legend** has been included in the graph to show which data bar is girls and which data bar is boys.

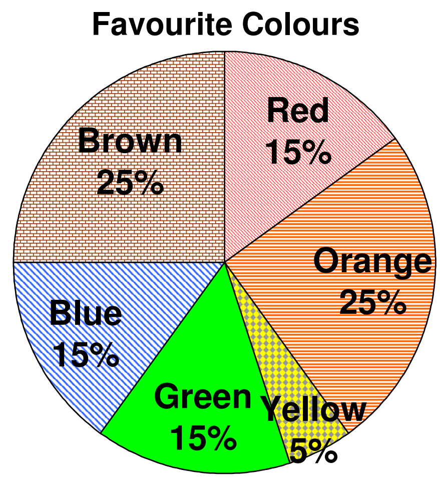


Notice in the multi-bar graph also that it is the **percentages** of each category that are graphed, not the count. The reason is that since the size of the boys data set (26 samples surveyed) and the size of the girls data (20 samples surveyed) is not the same, it could be misleading if you do not graph them as a percentage and it would be difficult to compare boys responses to girls responses, so we just adjust the counts so they are out of **100**.

CIRCLE GRAPHS

Circle graphs represent the frequency of each category but as sectors within a circle. Also known as pie charts.

Notice that circle graphs normally represent the percentage of each category.



To calculate a circle graph (or pie chart) it is necessary to change your counts into percentages and then your percentages into a proportion of a circle in degrees.

It is not possible to have simple circle graphs that represent multiple data sets like when we showed boy's versus girl's data.

To get the RED percentage you calculate the number of red as a percentage of the total. 3 out of 20 is the same as 15 out of 100; so 15%.

So: $\frac{3}{20}=\frac{x}{100}$; $\therefore \frac{3*100}{20}=x=15$

Compare with the circle graph above.

	Girls	%	Degrees
Red	3	15%	54
Orange	5	25%	90
Yellow	1	5%	18
Green	3	15%	54
Blue	3	15%	54
Brown	5	25%	90
TOTALS:	20		360

To get the RED angle (in degrees) you calculate the equivalent fraction that 3 parts out of 20 is to **360°**.

$\frac{3}{20}=\frac{x}{360°}$; $\therefore \frac{3*360}{20}=x$; $\therefore$ x = 54°. So the RED sector would be a 54° slice.

You Try: You do a circle graph for the following data.

We asked **32** girls how many children they had; here is the raw data response for each response

0	**3**	**3**	**1**	**4**	**0**	**2**	**2**
0	**1**	**5**	**0**	**0**	**1**	**4**	**3**
2	**5**	**2**	**1**	**1**	**1**	**5**	**0**
4	**3**	**2**	**1**	**0**	**1**	**1**	**5**

Tally and Table

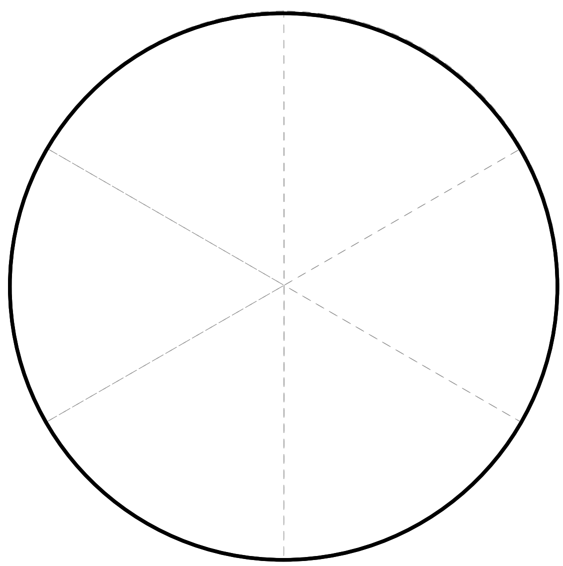
Do a tally of each observation and complete the table.

Category	Tally	Count	Percent	Degrees
ZERO				
ONE				
TWO				
THREE				
FOUR				
FIVE				
	TOTAL			

Now do the circle graph. Give it a **proper title**, a **legend**, **values for each category**, ***distinct pattern*** (eg: cross hatching) for each category.

(The pie is already dotted in with 60° sectors so you can readily more readily graph your data)

Title: ______________



GRAPHING PURE NUMERIC (QUANTITATIVE) OBSERVATIONS

Often the data you are collecting is not a colour or a brand of chocolate bar, it is a numeric value that you can quantify (A quantitative value) like a shoe size or math mark or an annual income or a month (when numbered in order). Plotting data of that nature where you are looking for trends and relationships in data is the purpose of Line Graphs and Histograms.

LINE GRAPHS

Line graphs require that you plot the value of your numeric observation on a vertical axis, it is not simply a count or frequency on the vertical axis.

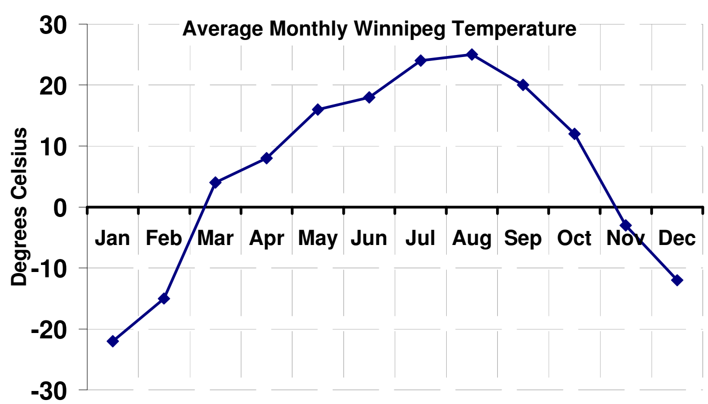
Let us look at the average monthly temperature in Winnipeg in 2012! Notice there is only one temperature for any particular month.

Month	Mid Jan	Mid Feb	Mid Mar	Mid Apr	Mid May	Mid Jun	Mid Jul	Mid Aug	Mid Sep	Mid Oct	Mid Nov	Mid Dec
Temp [°C]	¯22	¯15	4	8	16	18	24	25	20	12	**¯3**	¯12

Notice that the months form a quantifiable continuum of time and the temperature is a numeric value of degrees Celsius.

A line graph of the data looks like this.

The advantage of the line graph is that you can see trends from the **slope** of the line segments.



SLOPE

Slope is calculated by how much something rises or falls every time you take a step of one unit to the right. (You may recall calling this a tangent in Grade 10 trigonometry! Exactly the same idea)

For example the temperature changes by 7 degrees upwards in the one month from Jan to Feb, so we say there is a slope of plus 7, stepping forward one month from Feb has an increase of 7°C, so the temperature goes up 7°C per month at that time of year.

Where a line segment rises or drops at a sharp angle is where there is a big rise or drop in temperature. So you can readily see a rapid warming up in February to March, and a nice almost constant temperature in July and August. The slope gives us a good idea of the rate of change, how much something changes for a certain number of steps forward.

Interpolation. Another advantage of a Line Graph is that you can readily interpolate measurements. If you know that the Mid Apr temp is 8°C and the Mid May temp is 16°C, you can calculate that since **1 May** is midway between Mid Apr and Mid May that the temperature on **1 May** should be half way between 8°C and 16°C or 12°C.

$$\frac{8^\circ \ increase}{1 \ month} = \frac{x \ increase}{\frac{1}{2} \ month}; \quad \therefore \ x = 4^\circ \text{ increase in half a month, so } 8 + 4 = 12!$$

You Interpolate:

What is the expected temperature end of October? ________________ [4.5°]

What is the expected temperature beginning of February? _____________ [-5.5°]

EXTRAPOLATION

Extra means 'outside'. When you ***extrapolate*** you predict a value for outside of the range of the measured data. For example; here is a line graph for the world record high jump heights:

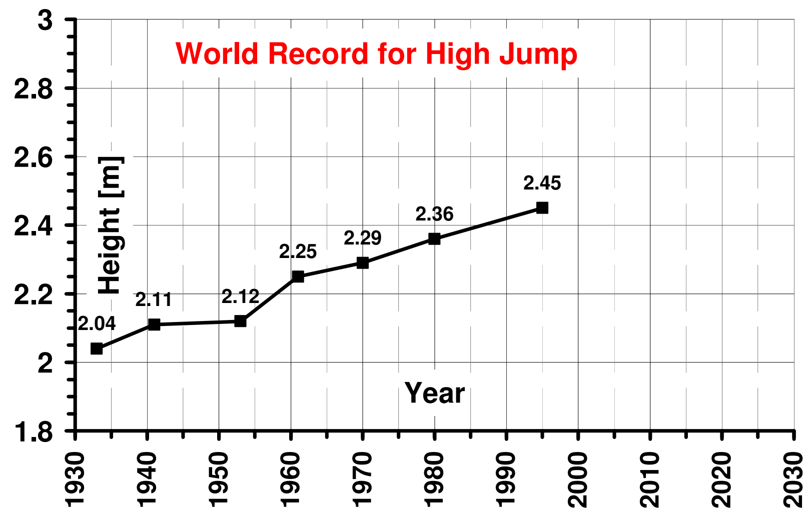
What is the slope of the data from 1980 to 1995?

____

If we follow that trend what should have been the world record in 2010? _________

In 2025? __________

in 2100? _________



What was the world record if we follow the trend backwards in 1803? __________

You likely suspect that extrapolating data is a very dangerous thing to do! Yet you will find that some people do it often.

Be very cautious and suspicious of extrapolated data!

HISTOGRAMS

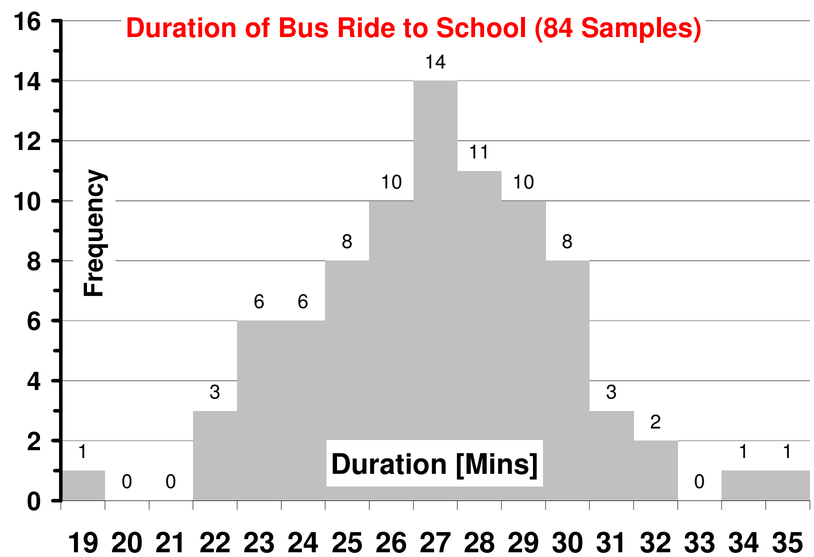
Histograms look similar to Bar Graphs except that histograms are only used when the categories are a numerical or measurable amount. The first graph we saw, the Shoe Size, was actually a **histogram** since we were counting the occurrence of quantifiable (numerical) categories; ie: shoe size. Histograms are very closely related to line graphs; since they both graph numbers vs numbers, except histograms graph counts of numbers vs numbers.

The main differences of a histogram from a bar graph are that the bars are always vertical and they must touch to make a nice continuous numeric measure of some quantity that we measure in a survey or a large statistical sample. The shape of a histogram relates very much to a *probability curve* and is **not** discrete non-numeric categories such as favourite colours or favourite chocolate bars. Favourite colours and chocolate bar brands have no numeric order and cannot be represented as a histogram unless maybe you are measuring the colour 'wavelength' or the sugar calories of the chocolate bar.

This histogram shows the frequency of the duration of bus trips to school based on 84 different trips (samples).

It gives you a sense of the probability, the likelihood, it will take you more than 30 minutes to get to school for example!

You will learn to such type of statistics in Grade 12.



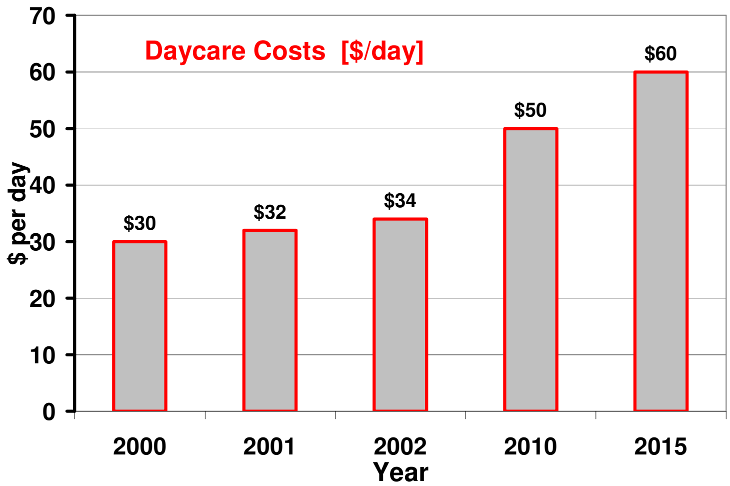
The histogram also gives you a very clear graph of what the most common trip duration is: ___________; the average (mean) duration is: ______________; the **range** of the possible trip durations: _____________; the least duration (minimum): ____________; the longest duration (maximum): ________________; and you even get a sense of the durations that will occur half of the occasions you take the bus (inter-quartile range): __________. **Half of the time** the bus takes **between** 25 and 29 minutes! You can even see the '*outliers*'; those rogue trip durations that stick out as being rather unusual because they only happen once (*or maybe you recorded them wrong*!)

But you will study these types of numerical statistics in Grade 12!

MIS-REPRESENTING DATA WITH A GRAPH

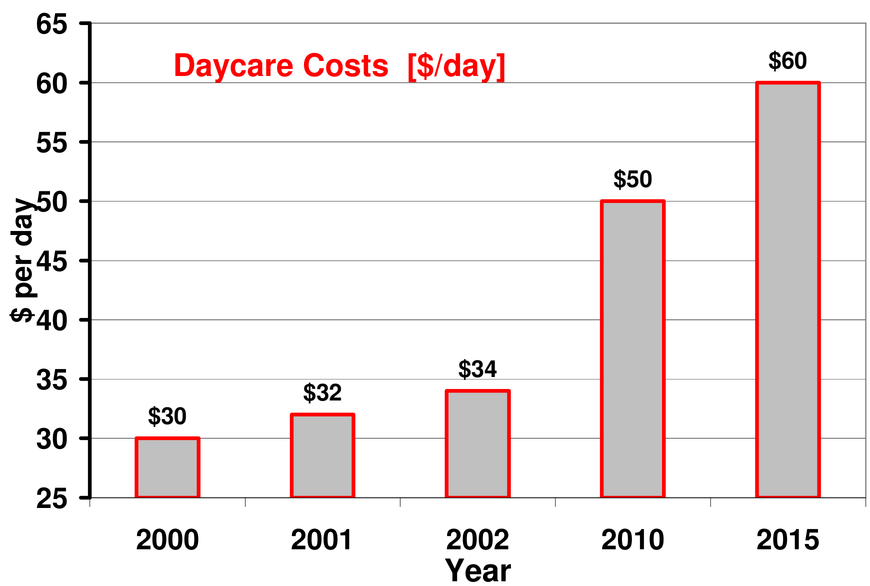
Graphs readily show lots of information, they can also be twisted to mis-represent information (whether accidentally or on purpose)

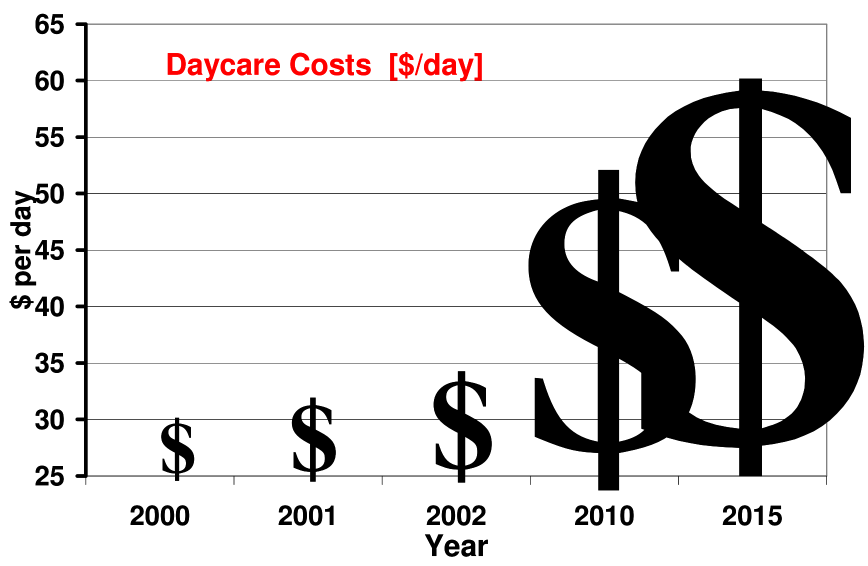
What is wrong with these graphs?



This data may be correct, but how is the graph mis-leading? Are day care costs rising more and more rapidly the last few years?

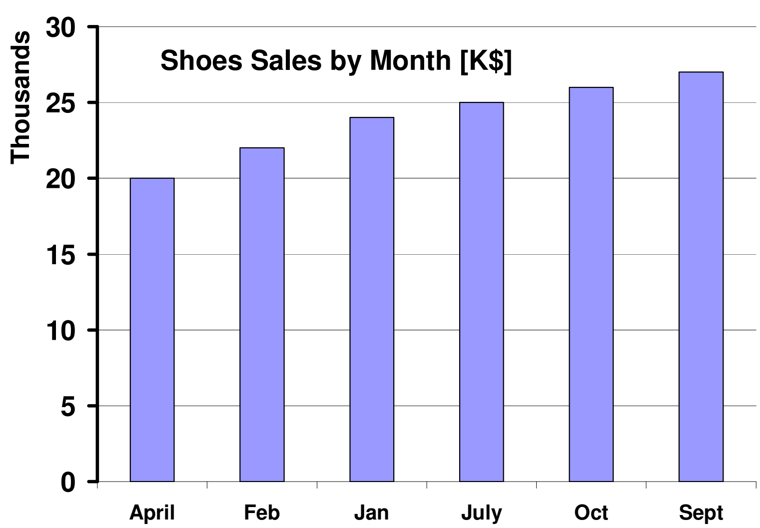
What has been done in this graph to make the information a bit more mis-leading?





This is a popular way to baffle brains too!

What additional thing is mis-leading about this graph now?



What is mis-represented here in this graph of monthly shoe sales performance?

CONCLUSION

We have studied here some very basic graphs. There exist some way more demanding graphs you will find in technical journals! Go find them. Have fun!