Guide to using Geogebra for MTH243

This guide covers how to use Geogebra to make calculations and produce graphs that will come up during MTH243, as taught at Portland Community College (Cascade) in 2019. All references to "textbook" and the section numbers are referring to *Advanced High School Statistics* by Diez, Barr, Çetinkaya-Rundel, & Dorazio.

Note that there are some calculations and graphs that you cannot make with Geogebra (using basic methods anyway), which you will have to do "by hand" or with some other technology, for example: building a two-way table from categorical data, making a pie chart, calculating the expected value of a random variable, etc. Note also that even for those procedures which you can carry out using Geogebra, you still need to check that the conditions are met for using your chosen procedure.

Step 1: Getting Geogebra

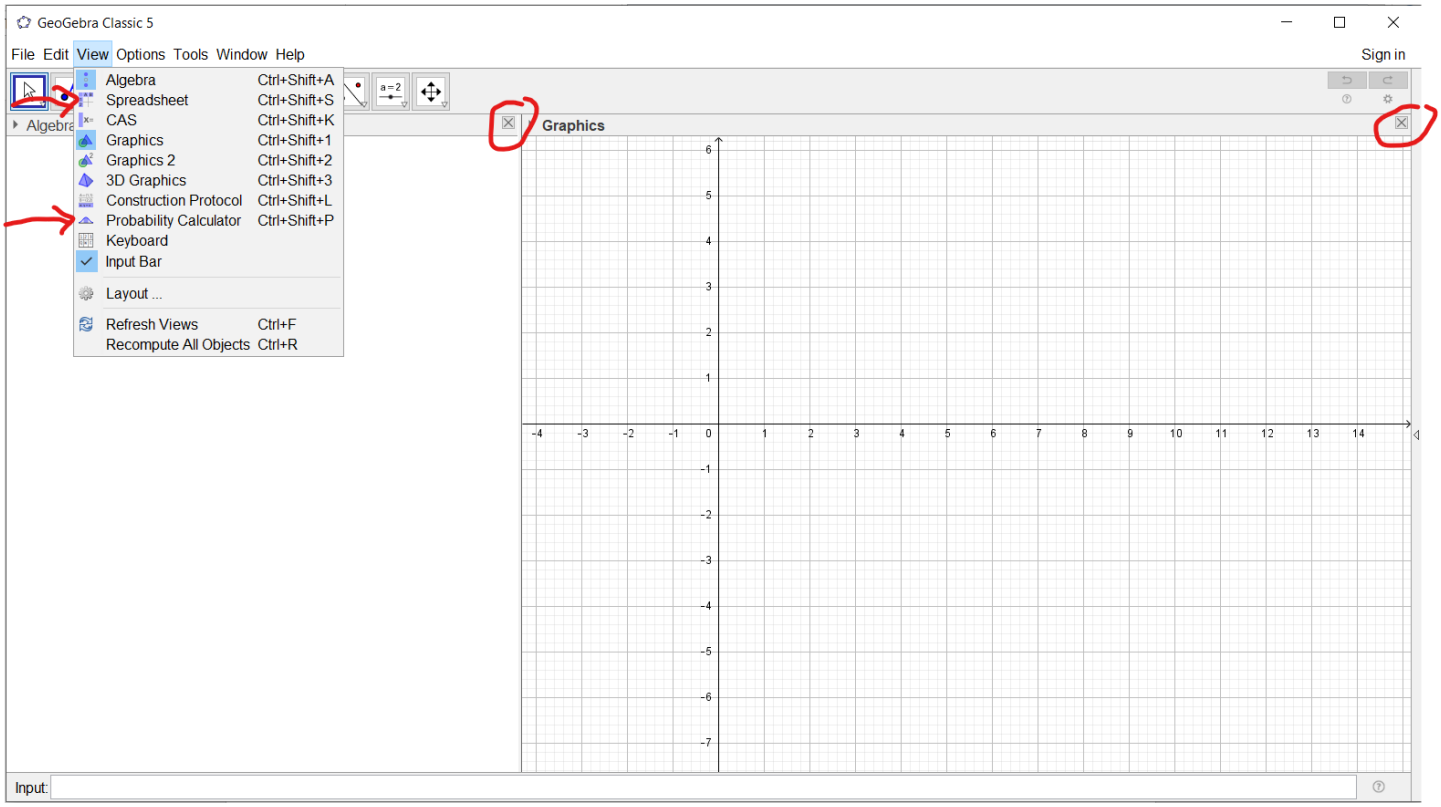
Geogebra is available on PCC computers. If there is no desktop shortcut, check the start menu and find it alphabetically in the list of programs.

Geogebra is free for download onto your own computer. Visit geogebra.org/download, and download "GeoGebra Classic 5". Choose a location to download the installation file which you will remember (e.g. the desktop or the downloads folder). Once it is fully downloaded, double-click on the file and follow the on-screen instructions to install the program.

Note that although there is a phone app available, this unfortunately does not include the statistics tools that we'll need for this course.

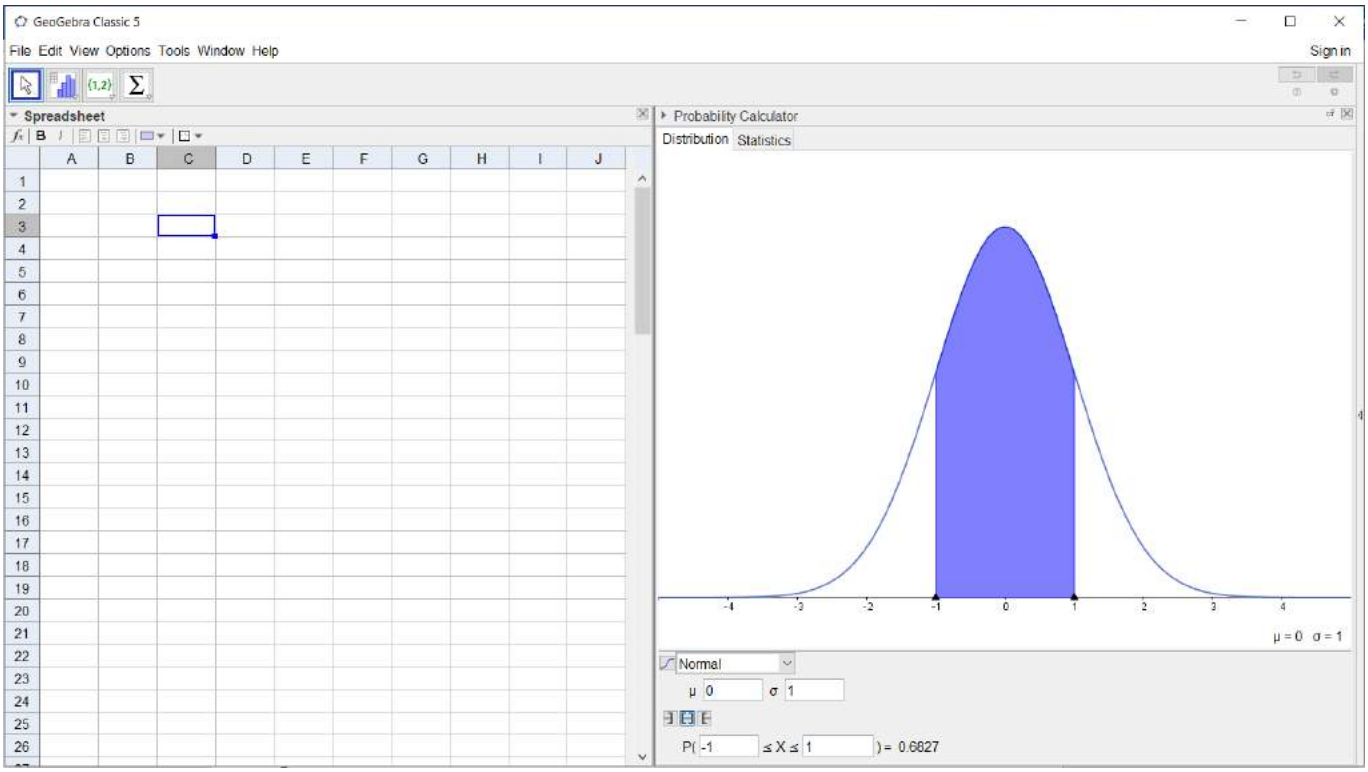
General orientation

When you open Geogebra, it will (usually) have the "Algebra" and "Graphics" windows open. We will not need either of these. So you can always close those first. What we will need is the "Spreadsheet" and/or "Probability Calculator" windows – depending on which statistical procedure we are carrying out – and these are both available through the "View" menu.



The Spreadsheet window is very similar to Excel. To be completely honest though, if you want to do anything complex, it is much easier to use Excel. For the relatively simple operations outlined in this guide though, Geogebra will do just fine.

The boxes in the spreadsheet are referred to as "cells", and are identified by their column letter and row number. For example, the box outlined in blue below is cell C3.



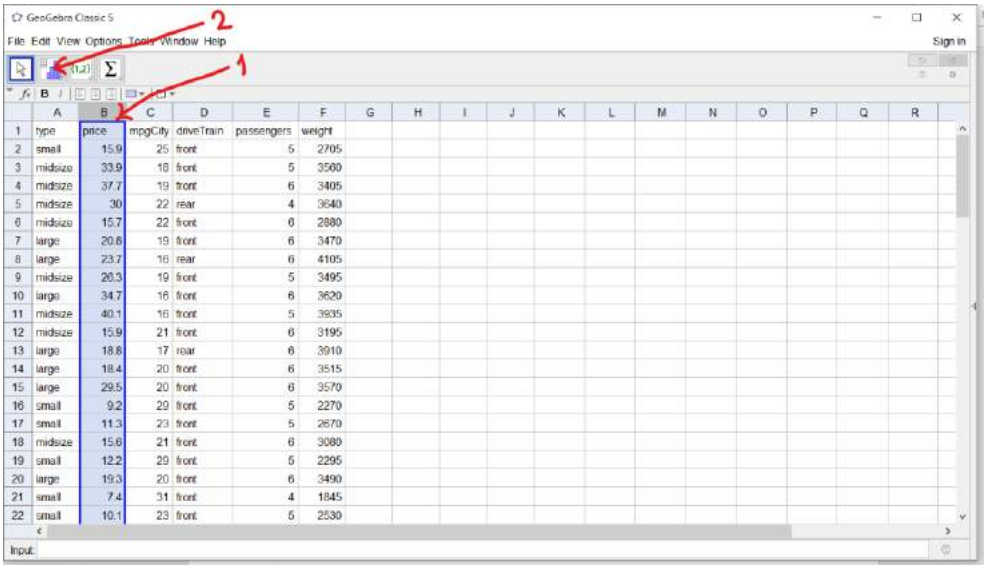
It is easy to enter data into the spreadsheet. Simply click on a cell, type a number, and hit enter. You can navigate to other cells either by clicking with the mouse, or using the arrow keys. However, care must be taken when attempting to copy/paste data into Geogebra from some other source, such as a web page, txt or doc file, etc. It will not always be pasted as intended. For example, if columns of numbers are separated by spaces, Geogebra will paste them into the same cells for each row, which is not what we want. However if the columns are separated by tabs, then all will go as planned. Likewise if the data is copied from an Excel file (an actual xls or xlsx file that is, not a csv file), it can be pasted into Geogebra without any trouble.

Everything we'll use Geogebra for in this course can be done by pointing and clicking, and some typing of data values – as opposed to writing computer code like in many other statistical analysis programs. So besides what is specifically covered in this guide, you can also just explore what's in the menus and what happens when pressing each icon button.

For a couple useful examples, consider the "Options" menu at the top (between "View" and "Tools"). The "Rounding" can be changed there; for example if you paste or type some data with 5 decimal places and are wondering why it's not displaying properly, this is why. Just go to "Options" → "Rounding" and change it to 5 or more decimals. You can also change the font size in the "Options" menu, as the default is quite small. You can even change the program to use a different language. (But, as always in the modern world, beware because the translation might be horrible!)

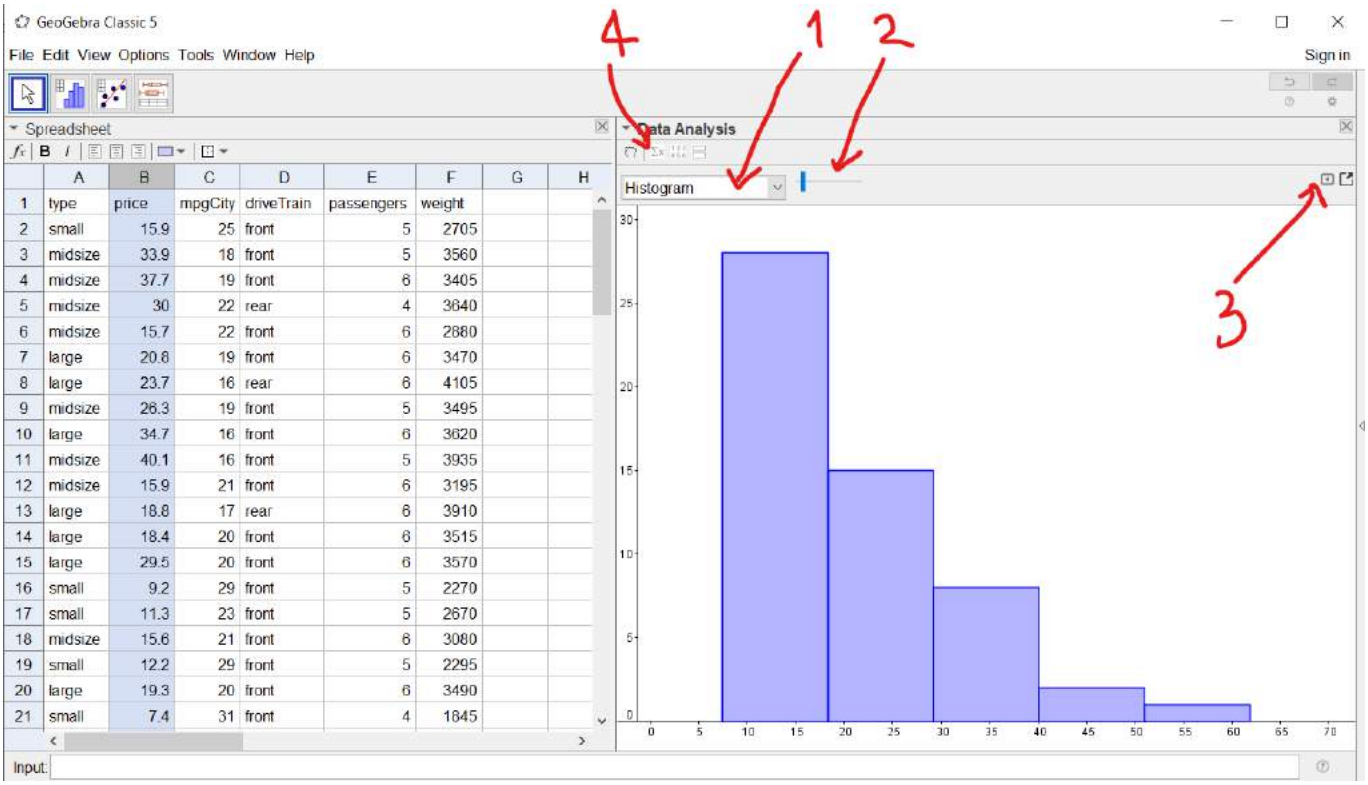
Section 2.1

The "cars" data set from the textbook will be used for this example. The file is a text file with columns separated by tabs, so simply copying and pasting gets the data set into Geogebra for us. We won't need the probability calculator, so that window can be closed. (In fact, we won't need that until chapter 3.3.) Let's make a *stem & leaf plot*, *dot plot*, and *histogram* using the "price" column of data, located in column B, since that is numerical data. Select the entire column by clicking on the "B". Although cell B1 has the word "price", Geogebra will automatically ignore this and compute using only the actual numerical entries in column B (marked "1" below). Next, click on the second icon, "One Variable Analysis" (marked "2" below), and then the "Analyze" button.



	A	B	C	D	E	F
1	type	price	mpgCity	driveTrain	passengers	weight
2	small	15.9	25	front	5	2705
3	midsize	33.9	18	front	5	3560
4	midsize	37.7	19	front	6	3405
5	midsize	30	22	rear	4	3640
6	midsize	15.7	22	front	6	2880
7	large	20.8	19	front	6	3470
8	large	23.7	16	rear	6	4105
9	midsize	26.3	19	front	5	3495
10	large	34.7	16	front	6	3620
11	midsize	40.1	16	front	5	3935
12	midsize	15.9	21	front	6	3195
13	large	18.8	17	rear	6	3910
14	large	18.4	20	front	6	3515
15	large	29.5	20	front	6	3570
16	small	9.2	29	front	5	2270
17	small	11.3	23	front	5	2670
18	midsize	15.6	21	front	6	3080
19	small	12.2	29	front	5	2295
20	large	19.3	20	front	6	3490
21	small	7.4	31	front	4	1845
22	small	10.1	23	front	5	2530

This will bring up a new "Data Analysis" window on the right. You may need to resize this by clicking and dragging the left edge of the window, so that the contents are reasonably visible. At first it contains a *histogram*. There are some noteworthy items here: (1) a dropdown menu for making different types of graphs of the same data, (2) a slider for changing the bin size of the histogram on the fly, (3) an "Options" area for changing the histogram in a more detailed way, and (4) a "Show Statistics" button.



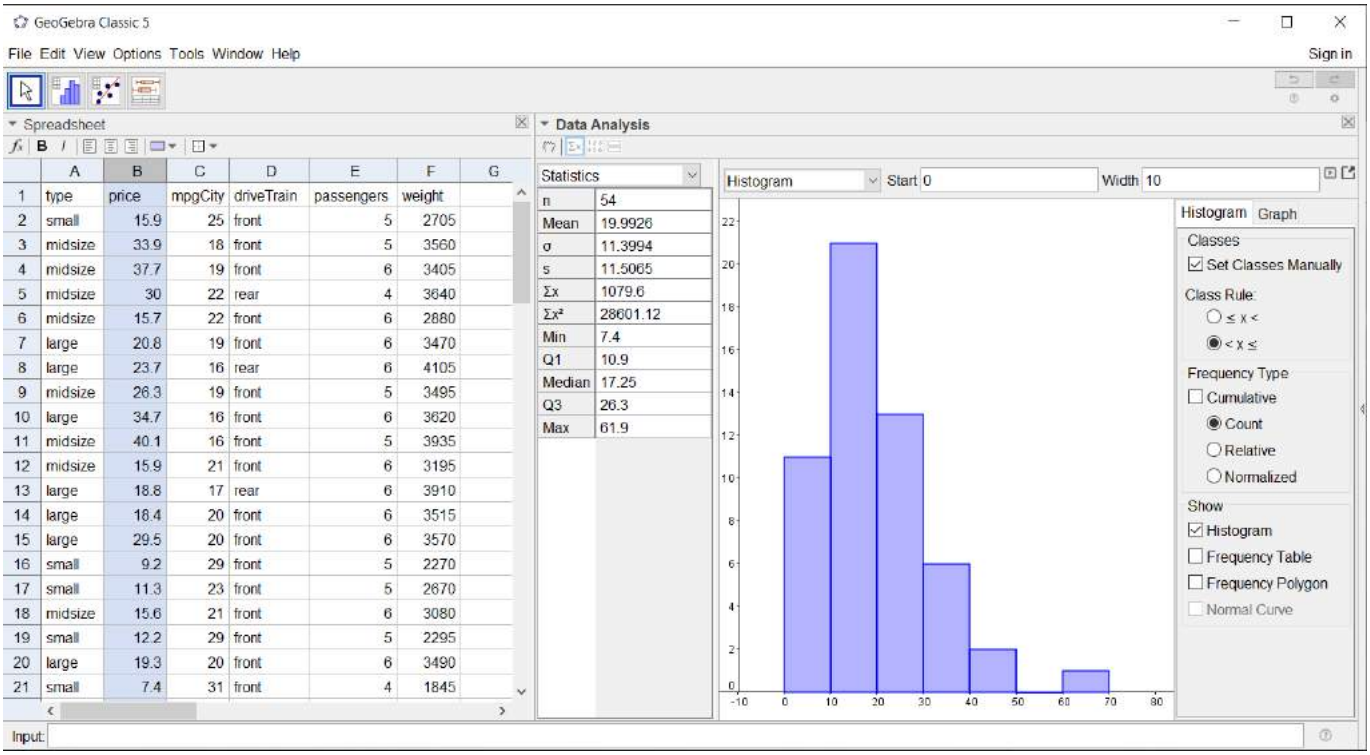
	A	B	C	D	E	F	G	H
1	type	price	mpgCity	driveTrain	passengers	weight		
2	small	15.9	25	front	5	2705		
3	midsize	33.9	18	front	5	3560		
4	midsize	37.7	19	front	6	3405		
5	midsize	30	22	rear	4	3640		
6	midsize	15.7	22	front	6	2880		
7	large	20.8	19	front	6	3470		
8	large	23.7	16	rear	6	4105		
9	midsize	26.3	19	front	5	3495		
10	large	34.7	16	front	6	3620		
11	midsize	40.1	16	front	5	3935		
12	midsize	15.9	21	front	6	3195		
13	large	18.8	17	rear	6	3910		
14	large	18.4	20	front	6	3515		
15	large	29.5	20	front	6	3570		
16	small	9.2	29	front	5	2270		
17	small	11.3	23	front	5	2670		
18	midsize	15.6	21	front	6	3080		
19	small	12.2	29	front	5	2295		
20	large	19.3	20	front	6	3490		
21	small	7.4	31	front	4	1845		

(1) Let's leave this alone for now, but we'll return there to make a dot plot and a stem & leaf plot.

(2) Click and drag on the slider to change the number of bins used to make the histogram. Notice that a number appears to the right of the slider when you are holding down the click button: this is the number of bins that the data is divided up between the min and the max. So for example, if you move the slider to 5, there are 5 bins. The min=7.4 and the max=61.9, so the width of each bin is (61.9-7.4)/5=10.9. But how did I know what the min and max were? Well I clicked on the "Σx" icon (marked "4" above).

You should always try moving the slider around when first investigating a data set. Using different bin widths may give different shapes: the modality and skew can change.

(3) Instead of the slider, we can set the Options. Always select the conventional practice of "<x≤" (which for some reason is not the Geogebra default), unless you have a specific reason not to. Check the "Set Classes Manually" box, put the left endpoint of the first bin in the "Start" blank and the bin width (or right endpoint of the first bin" in the "Width" blank. Under "Frequency Type", leaving "Count" selected displays frequency in the y-axis, whereas selecting "Relative" would show proportions (not percentages!) in the y-axis instead.

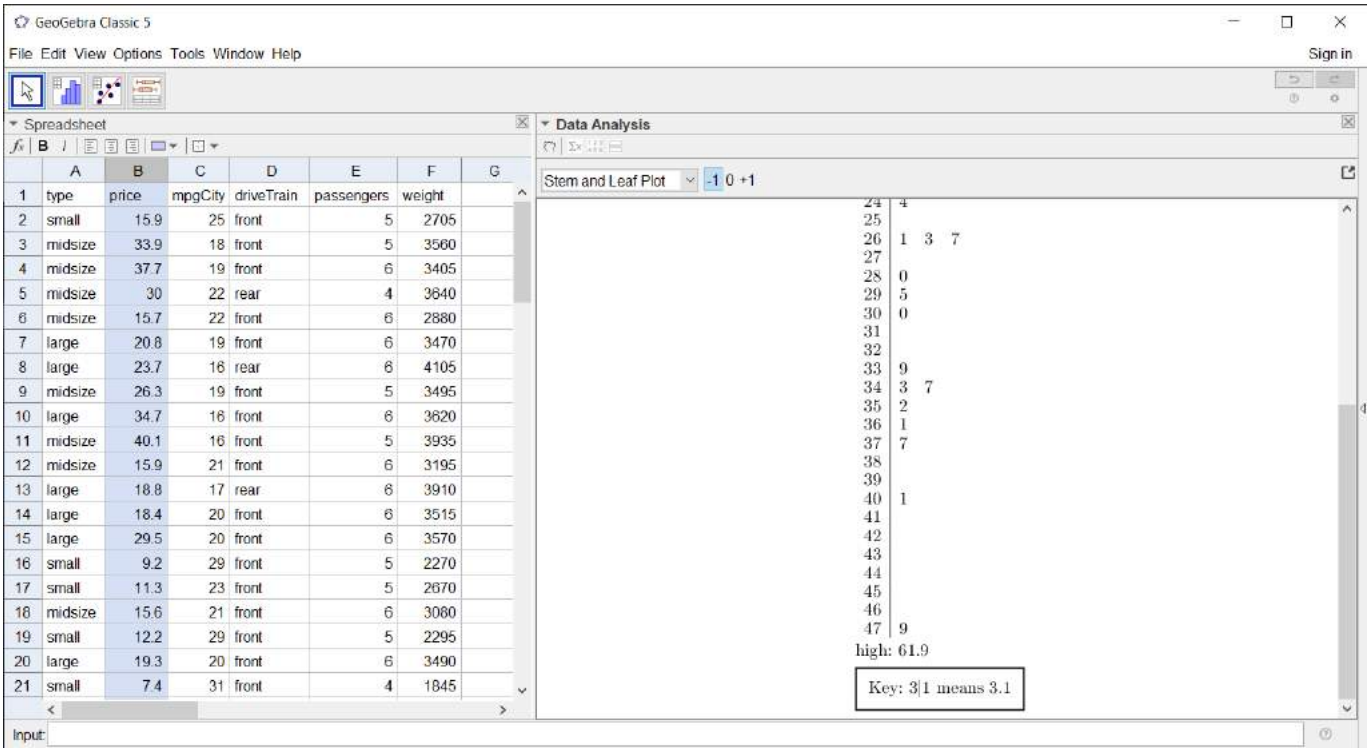


	A	B	C	D	E	F	G
1	type	price	mpgCity	driveTrain	passengers	weight	
2	small	15.9	25	front	5	2705	
3	midsize	33.9	18	front	5	3560	
4	midsize	37.7	19	front	6	3405	
5	midsize	30	22	rear	4	3640	
6	midsize	15.7	22	front	6	2880	
7	large	20.8	19	front	6	3470	
8	large	23.7	16	rear	6	4105	
9	midsize	26.3	19	front	5	3495	
10	large	34.7	16	front	6	3620	
11	midsize	40.1	16	front	5	3935	
12	midsize	15.9	21	front	6	3195	
13	large	18.8	17	rear	6	3910	
14	large	18.4	20	front	6	3515	
15	large	29.5	20	front	6	3570	
16	small	9.2	29	front	5	2270	
17	small	11.3	23	front	5	2670	
18	midsize	15.6	21	front	6	3080	
19	small	12.2	29	front	5	2295	
20	large	19.3	20	front	6	3490	
21	small	7.4	31	front	4	1845	

Click on the "Graph" tab (in "Options", right of the "Histogram" tab) to make more changes to the display, such as displaying grid lines or zooming in or out.

Click on the "Options" button again to close it when you have the histogram looking how you want.

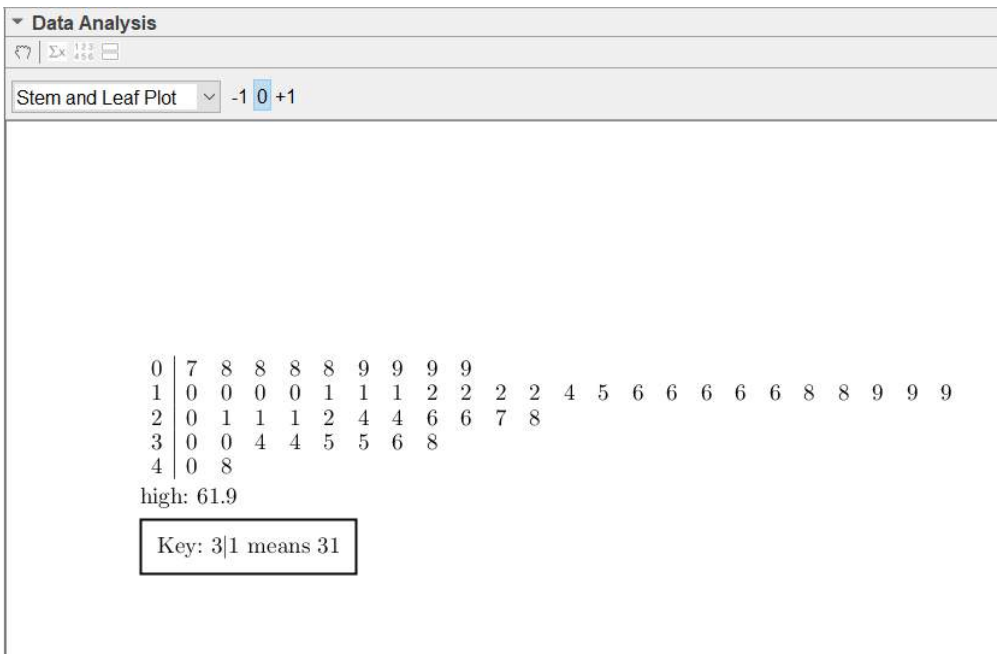
Now let's make a stem & leaf plot. Just go to drop-down menu and select "Stem and Leaf Plot" instead of "Histogram". The "-1 0 +1" which appears to the right is clickable, allowing you to choose the place of the stem; in this case "0" has tens as stems and ones as leaves, "-1" has ones as stems and tenths as leaves, and "+1" has hundreds as stems and tens as leaves. Below the "-1" selection is shown:



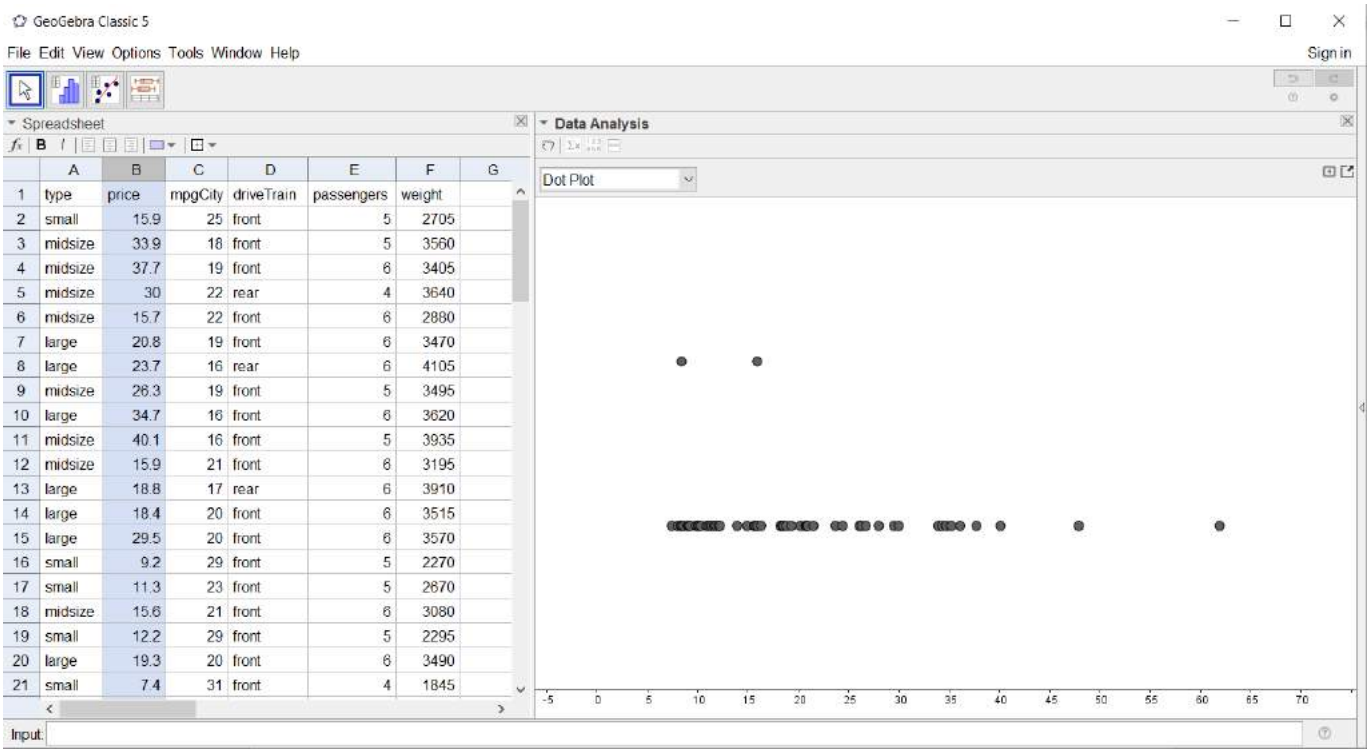
	A	B	C	D	E	F	G
1	type	price	mpgCity	driveTrain	passengers	weight	
2	small	15.9	25	front	5	2705	
3	midsize	33.9	18	front	5	3560	
4	midsize	37.7	19	front	6	3405	
5	midsize	30	22	rear	4	3640	
6	midsize	15.7	22	front	6	2880	
7	large	20.8	19	front	6	3470	
8	large	23.7	16	rear	6	4105	
9	midsize	26.3	19	front	5	3495	
10	large	34.7	16	front	6	3620	
11	midsize	40.1	16	front	5	3935	
12	midsize	15.9	21	front	6	3195	
13	large	18.8	17	rear	6	3910	
14	large	18.4	20	front	6	3515	
15	large	29.5	20	front	6	3570	
16	small	9.2	29	front	5	2270	
17	small	11.3	23	front	5	2670	
18	midsize	15.6	21	front	6	3080	
19	small	12.2	29	front	5	2295	
20	large	19.3	20	front	6	3490	
21	small	7.4	31	front	4	1845	

Geogebra has a couple of shortcomings here, that we humans need to use common sense to interpret and fix. Firstly, it does provide a key, but in this case it isn't quite right. Although "3|1" does represent "3.1" for the data as entered. But actually, these are car prices in thousands of dollars, so a better key would be "3|1 means $3100". Secondly, we can see that although it says "high: 61.9", the stem & leaf plot actually stops at 47.9, so apparently it doesn't always have room to display the full data set.

Picking the "0" option, notice that the values are rounded. For example, "2|0" could be 20, but it could also be anything from 1.5 to 2.4.



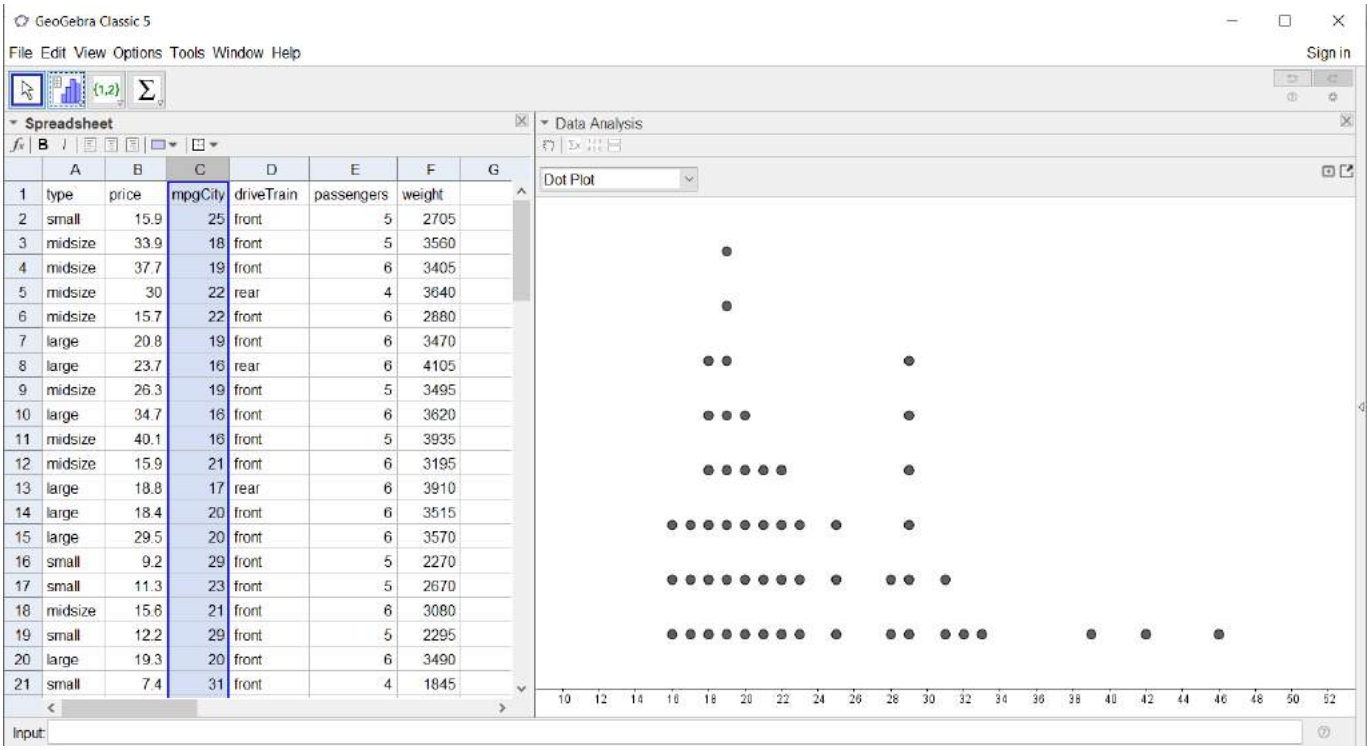
Now let's make a dot plot. Go back to that drop-down menu and select "Dot Plot".



	A	B	C	D	E	F	G
1	type	price	mpgCity	driveTrain	passengers	weight	
2	small	15.9	25	front	5	2705	
3	midsize	33.9	18	front	5	3560	
4	midsize	37.7	19	front	6	3405	
5	midsize	30	22	rear	4	3640	
6	midsize	15.7	22	front	6	2880	
7	large	20.8	19	front	6	3470	
8	large	23.7	16	rear	6	4105	
9	midsize	26.3	19	front	5	3495	
10	large	34.7	16	front	6	3620	
11	midsize	40.1	16	front	5	3935	
12	midsize	15.9	21	front	6	3195	
13	large	18.8	17	rear	6	3910	
14	large	18.4	20	front	6	3515	
15	large	29.5	20	front	6	3570	
16	small	9.2	29	front	5	2270	
17	small	11.3	23	front	5	2670	
18	midsize	15.6	21	front	6	3080	
19	small	12.2	29	front	5	2295	
20	large	19.3	20	front	6	3490	
21	small	7.4	31	front	4	1845	

It's not very interesting, is it? However it is functioning as it should. This is just an example of how a histogram is a much better graph for this particular data set than a dot plot.

Let's look at a dot plot of "mpgCity" instead of "price". Just select column C and click on the "One Variable Analysis" button:

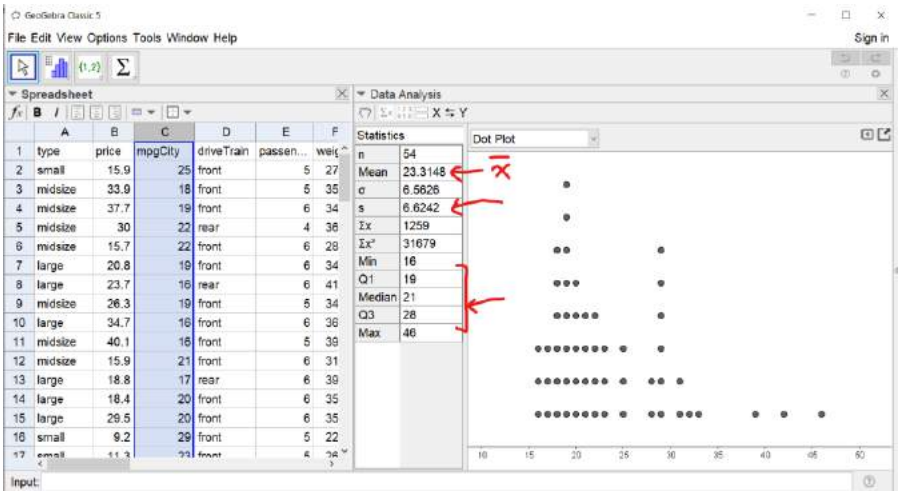


	A	B	C	D	E	F	G
1	type	price	mpgCity	driveTrain	passengers	weight	
2	small	15.9	25	front	5	2705	
3	midsize	33.9	18	front	5	3560	
4	midsize	37.7	19	front	6	3405	
5	midsize	30	22	rear	4	3640	
6	midsize	15.7	22	front	6	2880	
7	large	20.8	19	front	6	3470	
8	large	23.7	16	rear	6	4105	
9	midsize	26.3	19	front	5	3495	
10	large	34.7	16	front	6	3620	
11	midsize	40.1	16	front	5	3935	
12	midsize	15.9	21	front	6	3195	
13	large	18.8	17	rear	6	3910	
14	large	18.4	20	front	6	3515	
15	large	29.5	20	front	6	3570	
16	small	9.2	29	front	5	2270	
17	small	11.3	23	front	5	2670	
18	midsize	15.6	21	front	6	3080	
19	small	12.2	29	front	5	2295	
20	large	19.3	20	front	6	3490	
21	small	7.4	31	front	4	1845	

Section 2.2

We'll continue with the "mpgCity" variable of the "cars" data set from the textbook. Let's find the *mean*, *median*, *range*, *standard deviation*, *first quartile*, *third quartile*, *interquartile range*, and make a *boxplot*.

Just click on that "Σx" button and it shows us almost all of these statistics.



Note that "σ" is the population standard deviation, and should only be used if your data is from a census. In most cases, you only have a sample standard deviation and therefore are interested in "s".

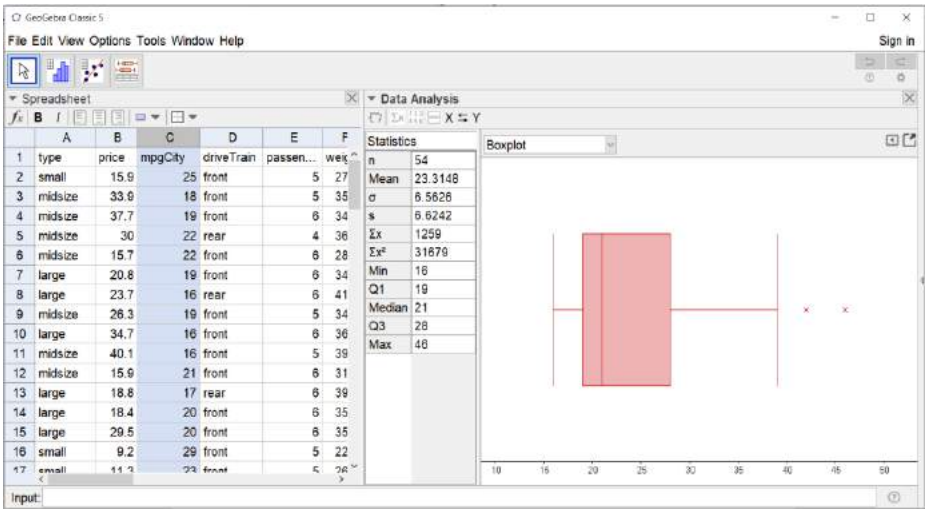
Note that the values with decimals are almost always actually rounded. The computer does not display anything indicating what is approximate and what is exact. But consider *s*: it added up a bunch of squared differences from the mean, then divided by *n-1*, then took the square root. There's no way 6.6242 is the exact answer. So on paper you should write "$s \approx 6.6242$".

There are some statistics on our list that are missing in Geogebra's output. We will need to calculate those with a calculator (or the Windows "calculator" program, accessible from the start menu). However we don't need to start from scratch and do it all by hand. The output does include everything in the right-hand side of the formulas for the two missing stats:

$range = max - min$

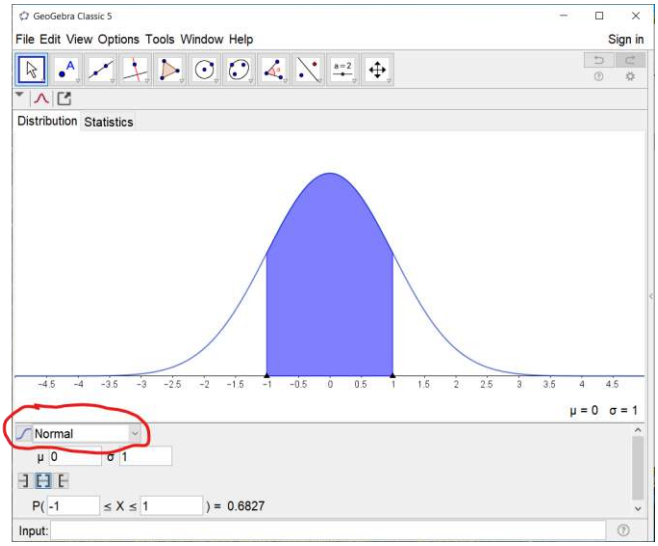
$IQR = Q_3 - Q_1$

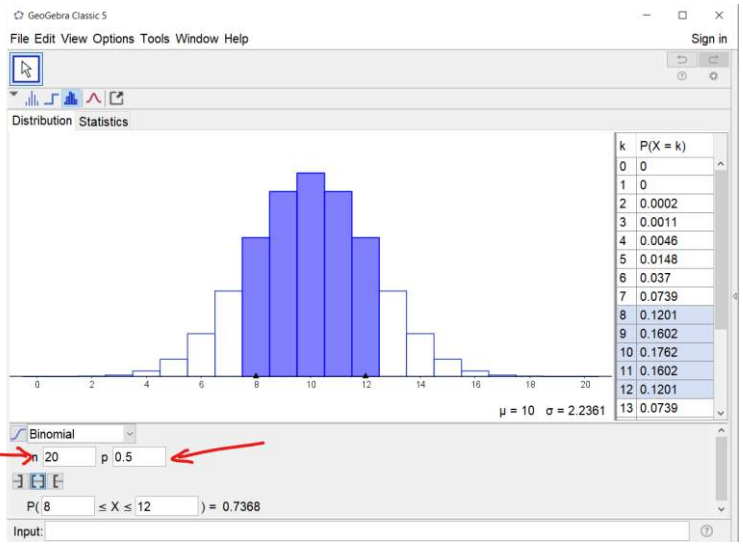
Now let's make a boxplot. Note that Geogebra won't output what the upper and lower fences are, or the specific values of any outliers. However it will display the boxplot correctly, including by showing any outliers using the 1.5 IQR rule. Simply select "Boxplot" in the dropdown menu (instead of "Histogram", "Dot Plot", etc).



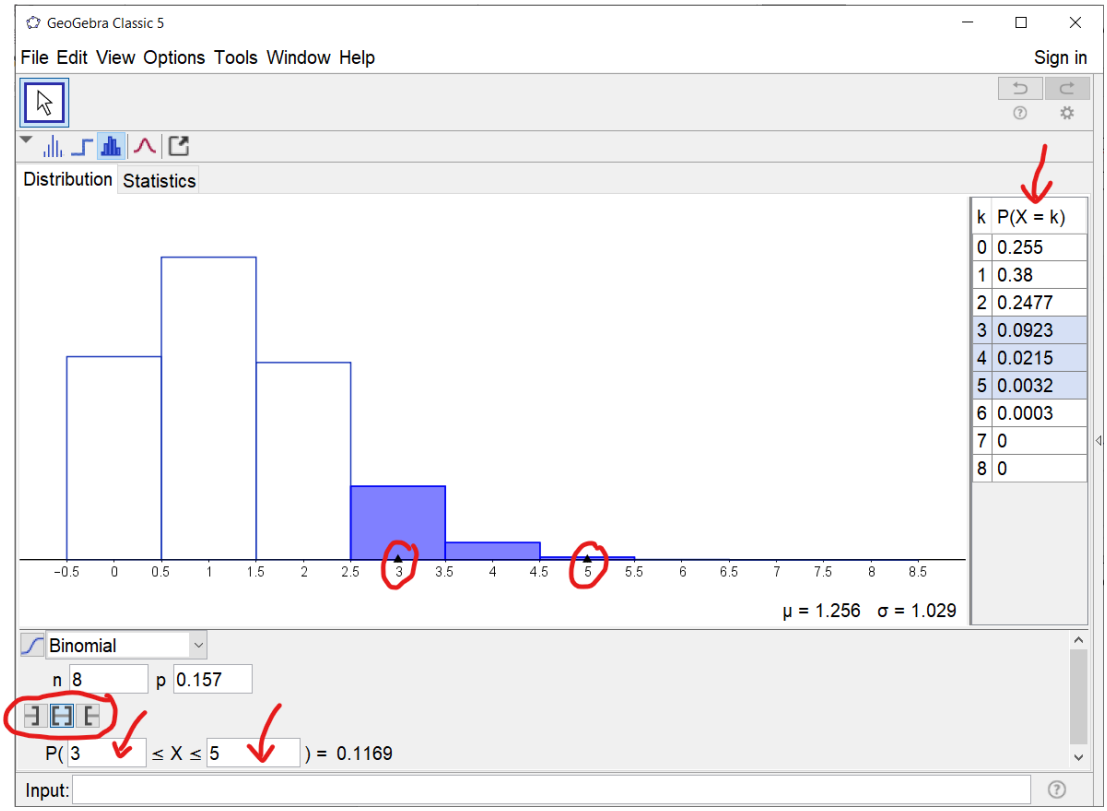
Section 3.3

Open the "Probability Calculator" (in the "View" menu). Go ahead and close any other Geogebra windows, unless you need them for other (non-binomial) problems you're working on. The default setting is to display the normal distribution, so go to the drop-down menu and select "Binomial" instead. Notice the fields for inputting *n* and *p*.





For example, let's say $n = 8$ and $p = .157$: just highlight the corresponding blanks, type the numbers, and press enter. (Also, if *p* is a fraction with a repeating decimal like 1/3, 42/987, etc: you can just type it in like so using the slash button, and the computer will compute the rounded decimal for you.) Notice that the graph, the table on the right, and the probability statement on the bottom all change.



First of all, keep in mind that this graph isn't 100% in the appropriate format. It looks like a histogram for sample data, but really this is a graph for the theoretical situation where some *X* follows Bin(*n*,*p*) exactly. To be the most mathematically correct, the graph should just have dots where at the centres of the tops of each bar. However, to be fair, Geogebra's version is actually easier to see and use.

Secondly, keep in mind also that all of the output probabilities are rounded (unless there is some lucky coincidence). If you want the super exact value as a fraction, you need to use the formula by hand, or use a different program such as wolframalpha.com.

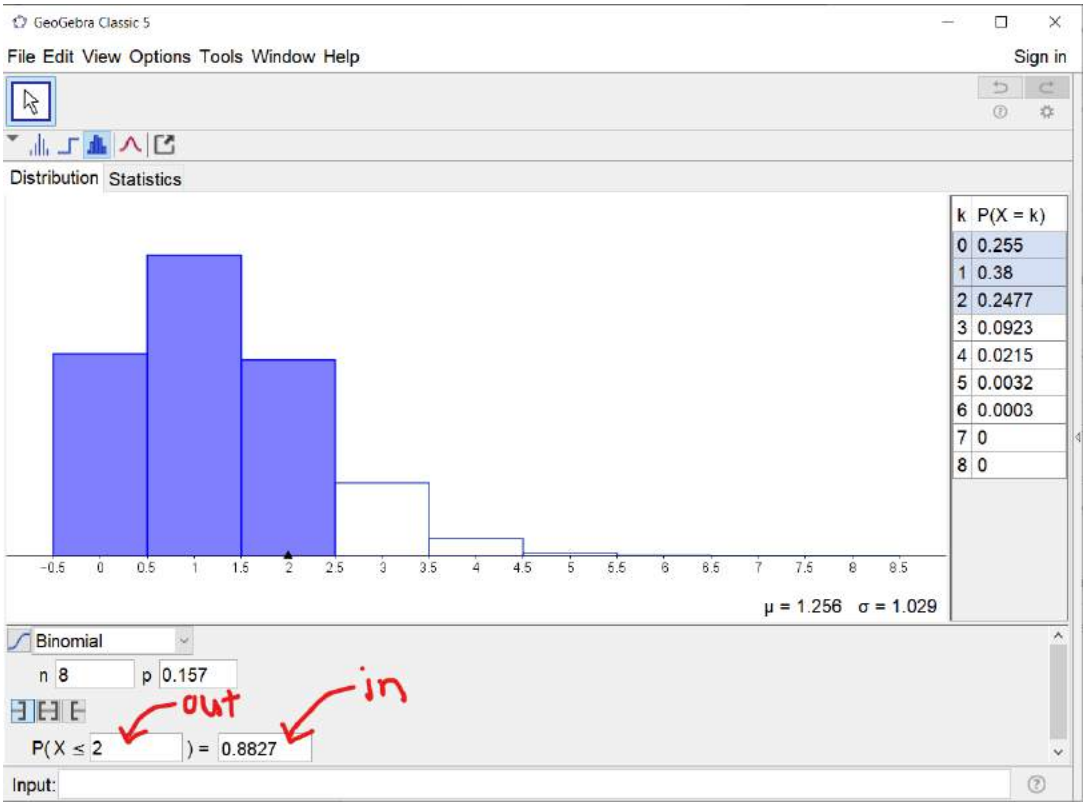
Anyway, notice the three buttons below the *n* and *p* fields. Currently the middle one is selected, allowing us to find the probability of *X* being between two constants, inclusive of the endpoints.

We can change the endpoints in three ways. (1) We can type them into the blanks in the probability statement at the bottom. (2) We can click and drag the black triangles in the graph. (3) We can click and drag to highlight rows in the table at the top right.

What if we just want the probability that *X* equals a single value? Just read that off of the table.

What if we want the probability that *X* is less than or equal to something? Choose the left button. And what if we want the probability that *X* is greater than or equal to something? Choose the right button. You don't have to memorise that choice: notice how the shading of the graph changes.

For either of the latter two situations, i.e. a one-sided inequality, Geogebra can also help us solve the inverse problem. That is, we can find the percentile. For example, let's find the 80th percentile when $n = 8$ and $p = .157$. Select the left button and type ".8" into the blank for the right-hand side of the probability statement at the bottom (labelled "in" below) and hit enter.



After hitting enter, it actually changes to .8827 instead of .8 – this is not an error! The closest we can get is that $P(X \leq 2) \approx .8827$, i.e. 2 is the 88.27th percentile. If *n* were higher, or if we were using some continuous distribution, we could get closer to 80.

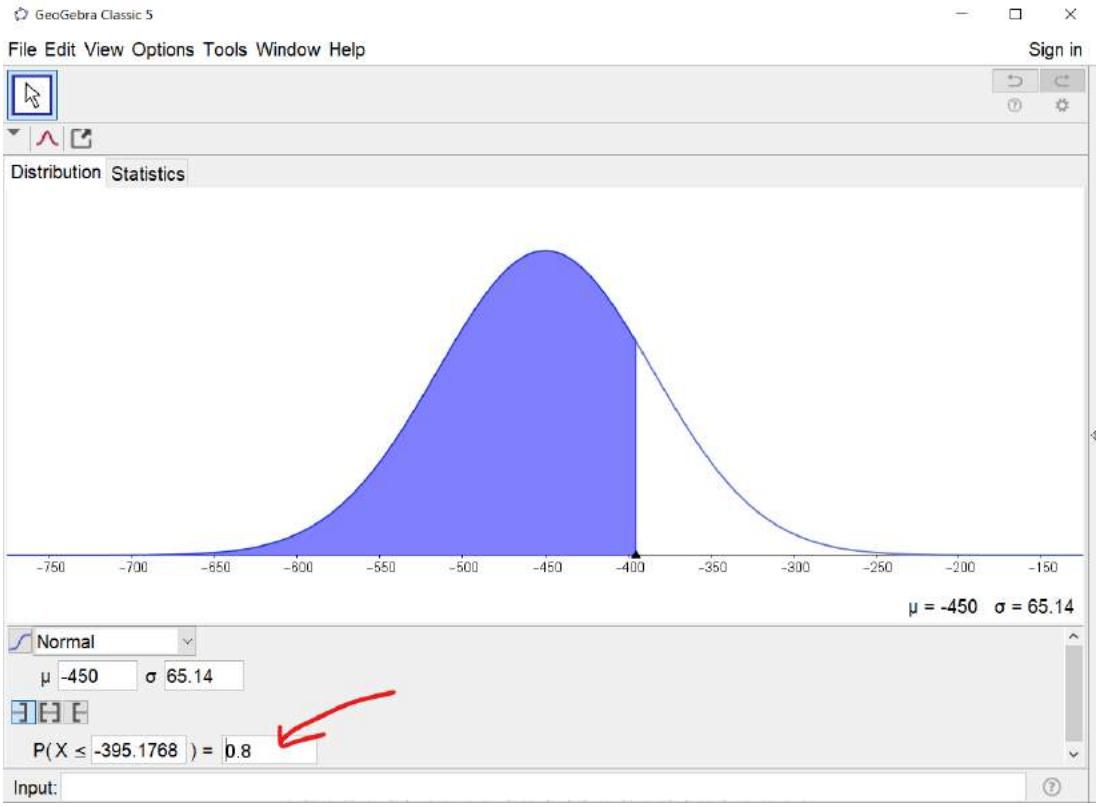
Section 4.1

For the normal distribution, as noted earlier this will pop up by default when using the "Probability Calculator". Or if you've been using binomial, go to the drop-down and select "Normal" instead.

By default, the *standard normal* distribution is shown, i.e. $\mu = 0$ and $\sigma = 1$.

Let's use some other normal distribution, e.g. $\mu = -450$ and $\sigma = 65.14$. Like with the binomial n and p, just type the values into their respective blanks. In fact, everything works the same as for the binomial. The only difference is that there is no table, which makes sense because this is a continuous distribution and we can't very well have a table with infinitely many rows in it.

Also as with the binomial, the output values are actually rounded despite the "=" on the screen. For example, let's find the 80th percentile.



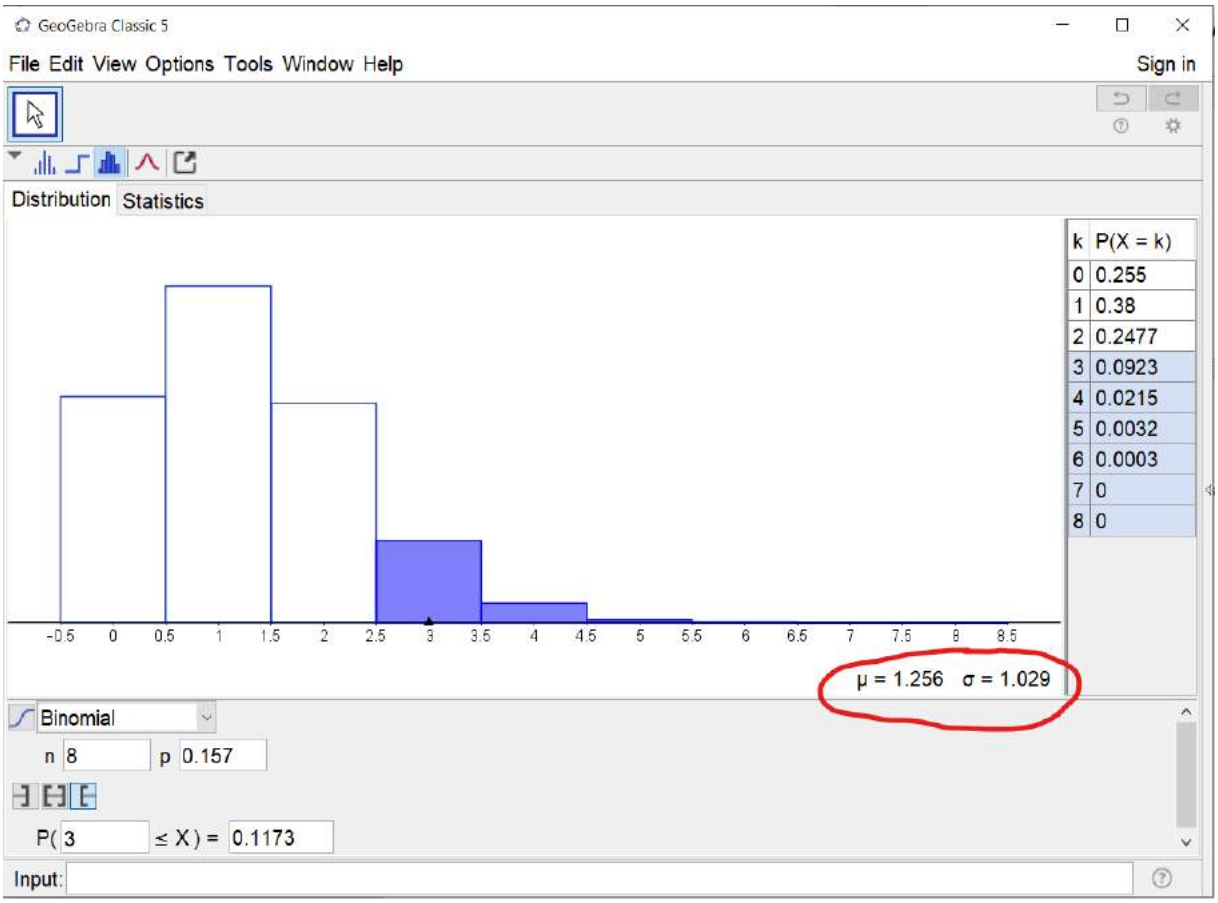
I typed in .8, so that is exact. However, the cutoff value that Geogebra computed when I hit enter is not exact. (Recall that the normal cdf does not even have a closed-form formula! And the pdf formula was pretty complicated too!) So in reality I should say that the 80th percentile (for *this* normal distribution) is approximately -395.1768.

Section 4.2

For a sampling distribution of $\bar{X}$, this is just a normal distribution with particular values for μ and σ. You must check if the conditions are met first! Geogebra will not do that part for you! Use the formula and a calculator to find the correct values for these parameters, and then proceed with finding any probabilities as above.

Section 4.4

Revisiting the binomial distribution, notice the μ and σ on the graph. These aren't normal parameters; they are the expected value and standard deviation of the binomial distribution:



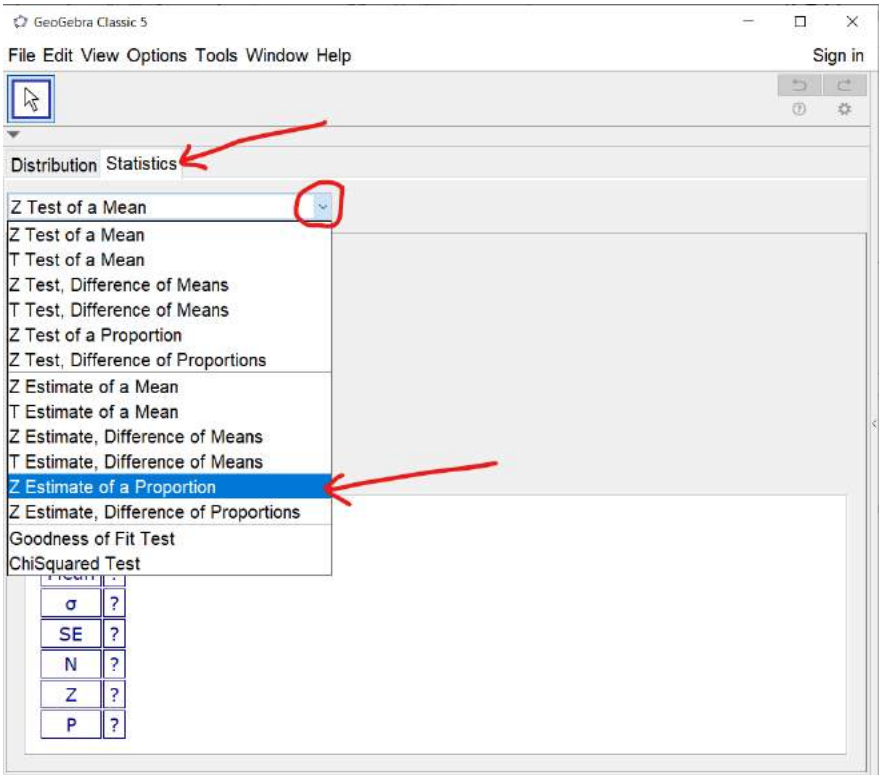
These Geogebra values are rounded and not exact. Use your calculator and the formulas to find exact answers, and/or answers rounded to more decimal places.

Section 4.5

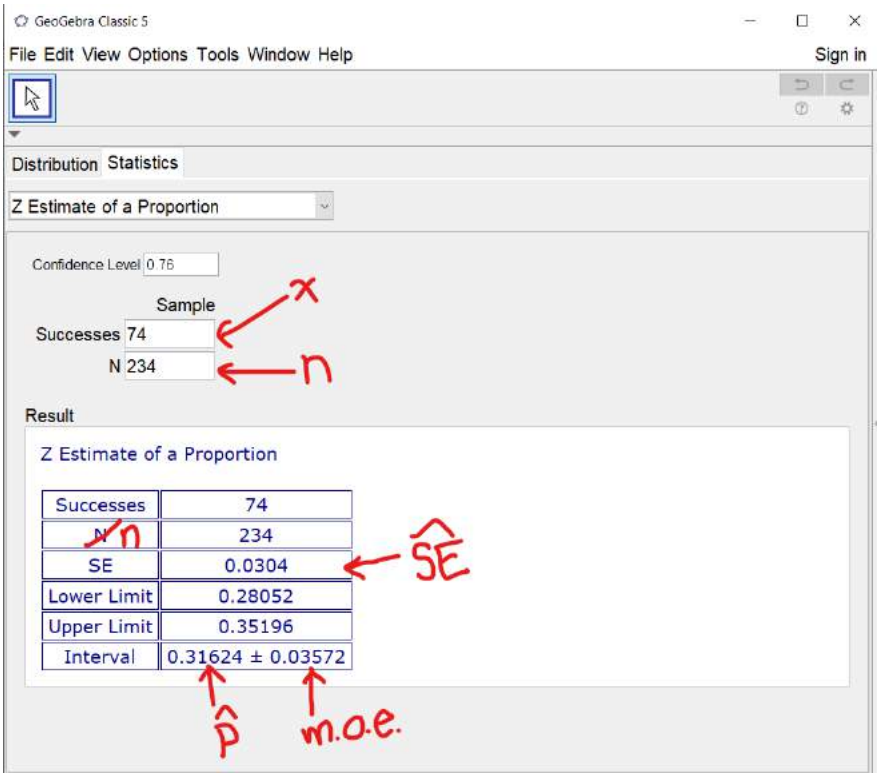
For a sampling distribution of $\hat{p}$, this is just a normal distribution with particular values for μ and σ. You must check if the conditions are met first! Geogebra will not do that part for you! Use the formula and a calculator to find the correct values for these parameters, and then proceed with finding any probabilities as usual.

Section 5.1, 5.2, 6.1: Confidence intervals for a proportion

In the "Probability Calculator", click on the "Statistics" tab, which is to the right of the default "Distribution" tab. Then select "Z Estimate of a Proportion". As usual, check the conditions yourself before blindly diving in to use the computer program: it will not display any error message if the conditions aren't met.



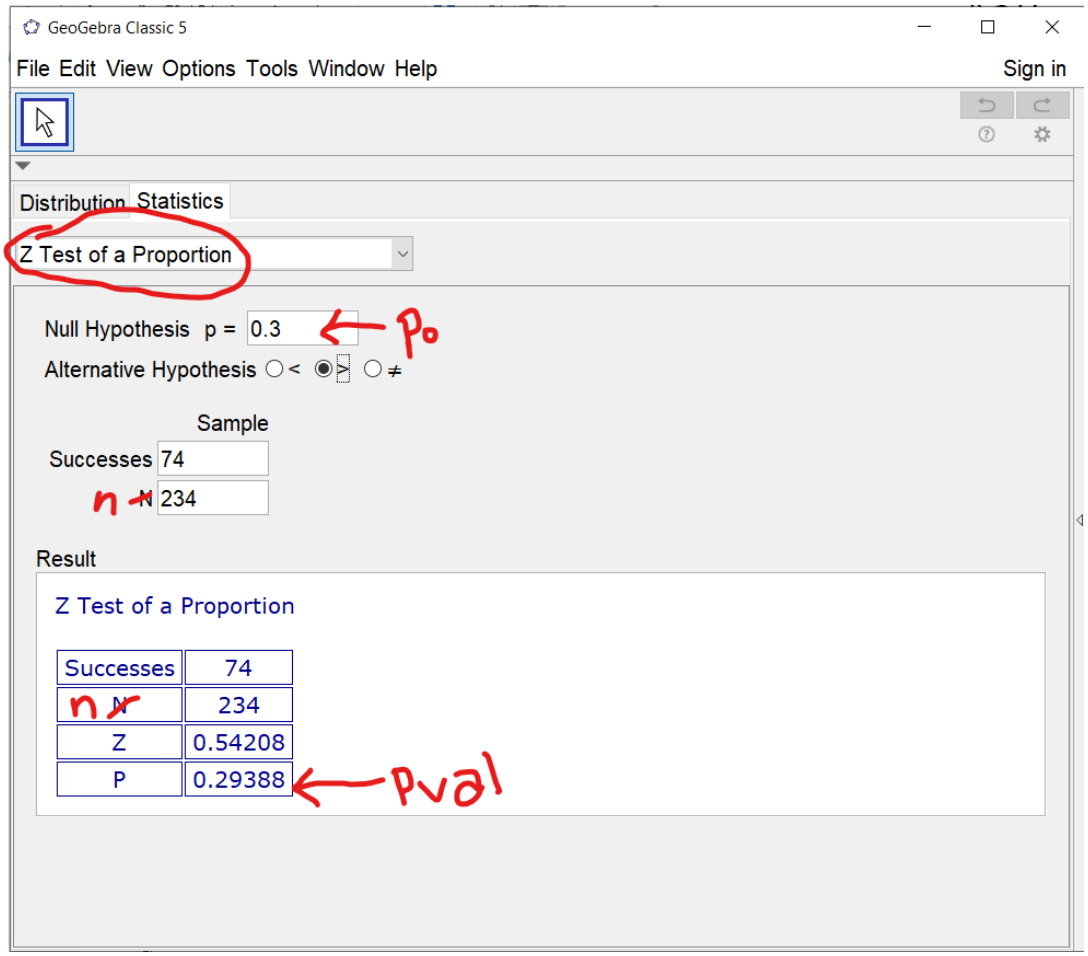
Enter the confidence level – as a number between 0 and 1, *not* a percent! –, the number of successes *x*, and the sample size *n*, and Geogebra does all the work for you. Note that there is a notation mistake here: Geogebra's "N" refers to the sample size *n*, and not the population size *N*. The confidence interval is (Lower Limit, Upper Limit).



If you also need to know z^*, there a couple ways to do this. (1) Use the normal distribution to find the value that has the appropriate amount of area to its right or left. (2) Use the fact that $moe = z^* \widehat{SE}$ and solve for z^*, since you know moe and $\widehat{SE}$ from the Geogebra output. Note however that method (1) will give you a more precise answer, since method (2) uses rounded values and this rounding error can accumulate when you do the calculation.

Section 5.3, 5.4, 6.1: Hypothesis tests about a proportion

In the "Probability Calculator" view, "Statistics" tab, select "Z Test of a Proportion" from the drop-down menu. Again, check the conditions yourself before proceeding. If the conditions are met, plug in the numbers, choose the alternate hypothesis direction (< or > or ≠) and Geogebra will calculate the *Pval* for you. The example below tests H_0: $p = p_0$ vs H_A: $p > p_0$, where $p_0 = 0.3$, $x = 74$, and $n = 234$.

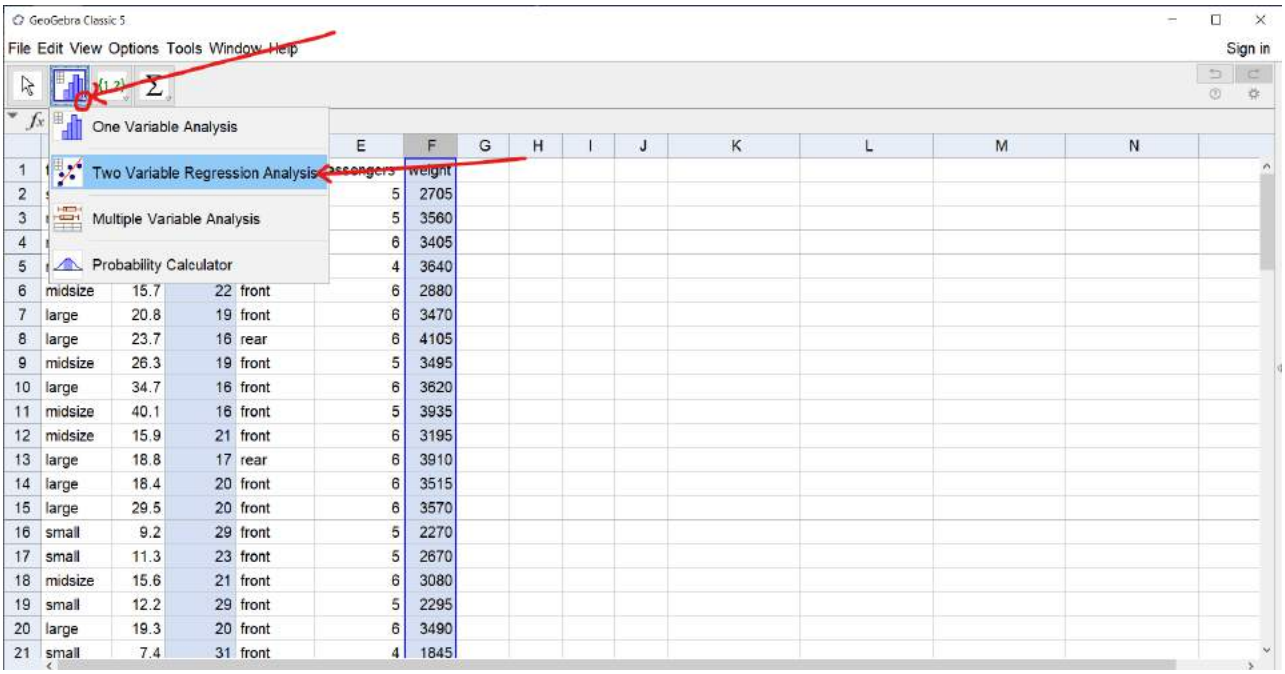


Note that Geograbra erroneously displays "N" instead of *n* for the sample size, and "P" for the *Pval* (or "p-value"). Of course there is no way to calculate the true population proportion *p* (besides going out and taking a census of the full population, anyway).

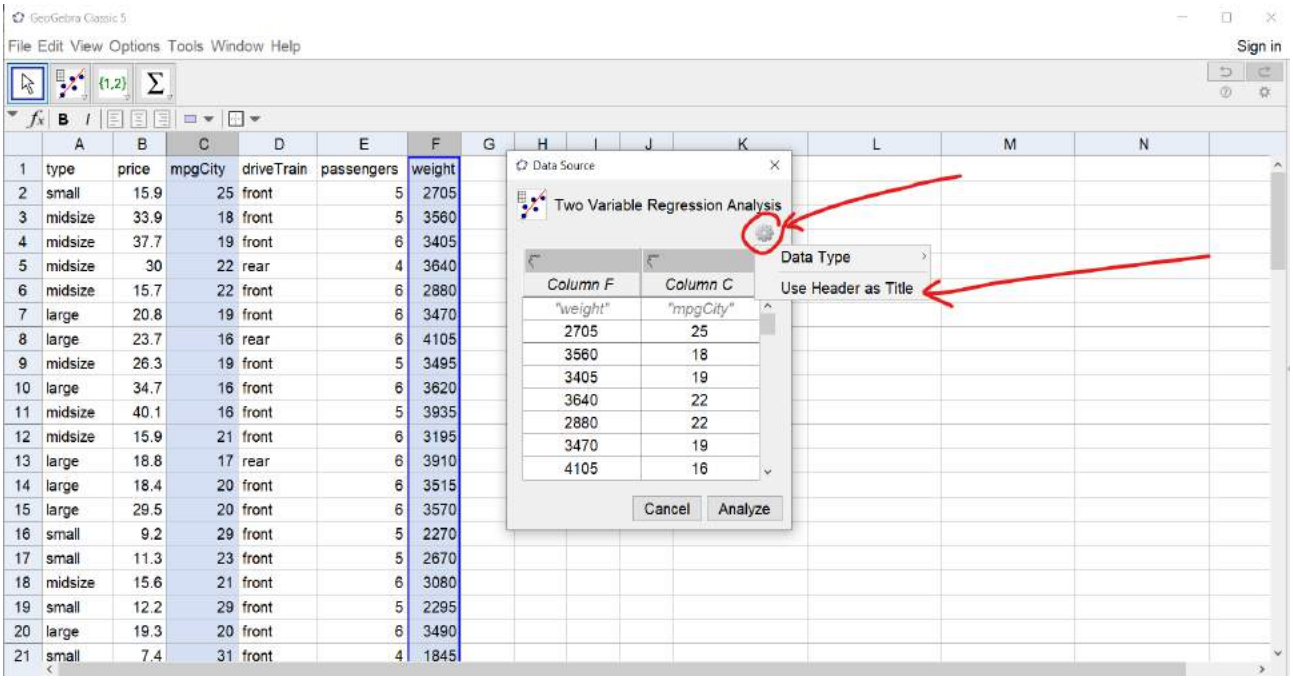
Section 8.1-8.2

For this example we'll use the "cars" data set again (as in section 2.1). Perhaps there is a linear relationship between city mpg and weight of the cars. After copying and pasting the data into the "Spreadsheet" view, select the columns for these two variables. Two columns that aren't next to each other, columns C and F in this case, can be simultaneously selected by holding down the Ctrl button and then clicking on the column headers.

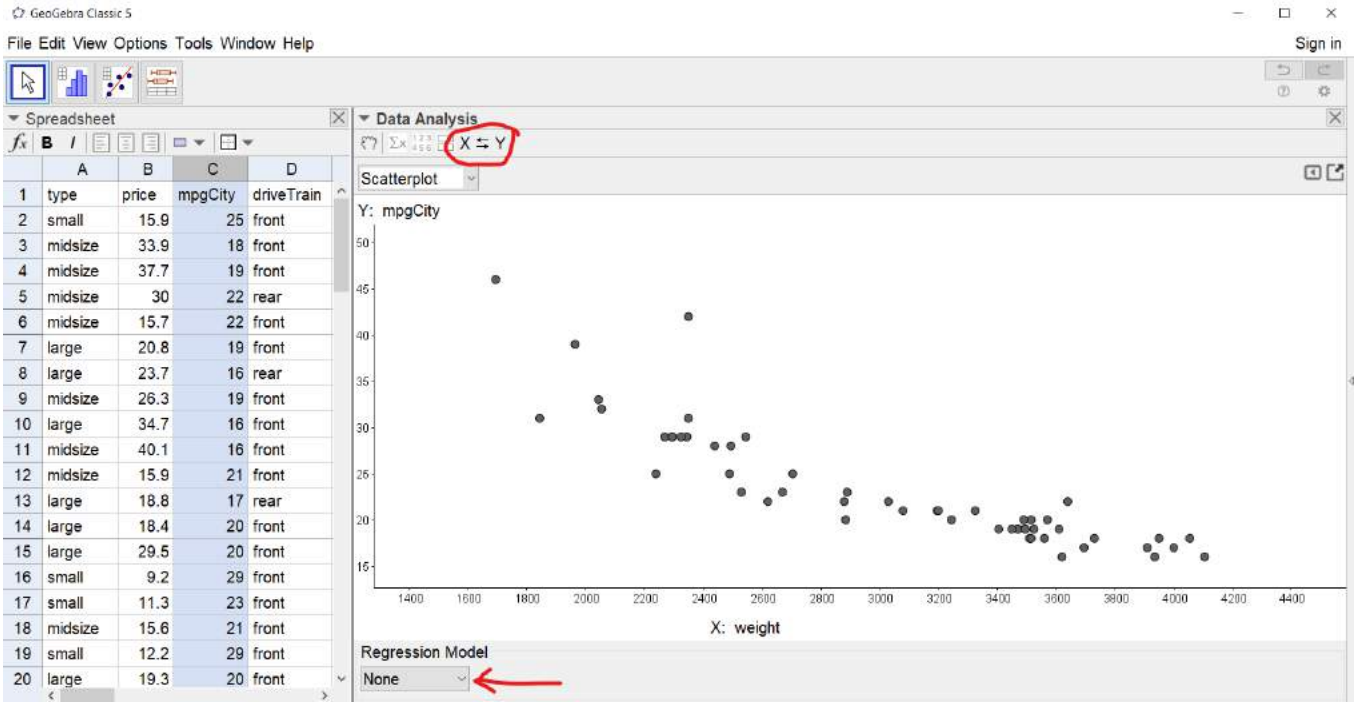
Then, click on the small downward triangle in the bottom right corner of the icon with the histogram on it. A menu will appear; select "Two Variable Regression Analysis".



Before clicking on "Analyze", click on the gear icon and then "Use Header as Title". This tells Geogebra to use the column titles as axis labels on the graph.



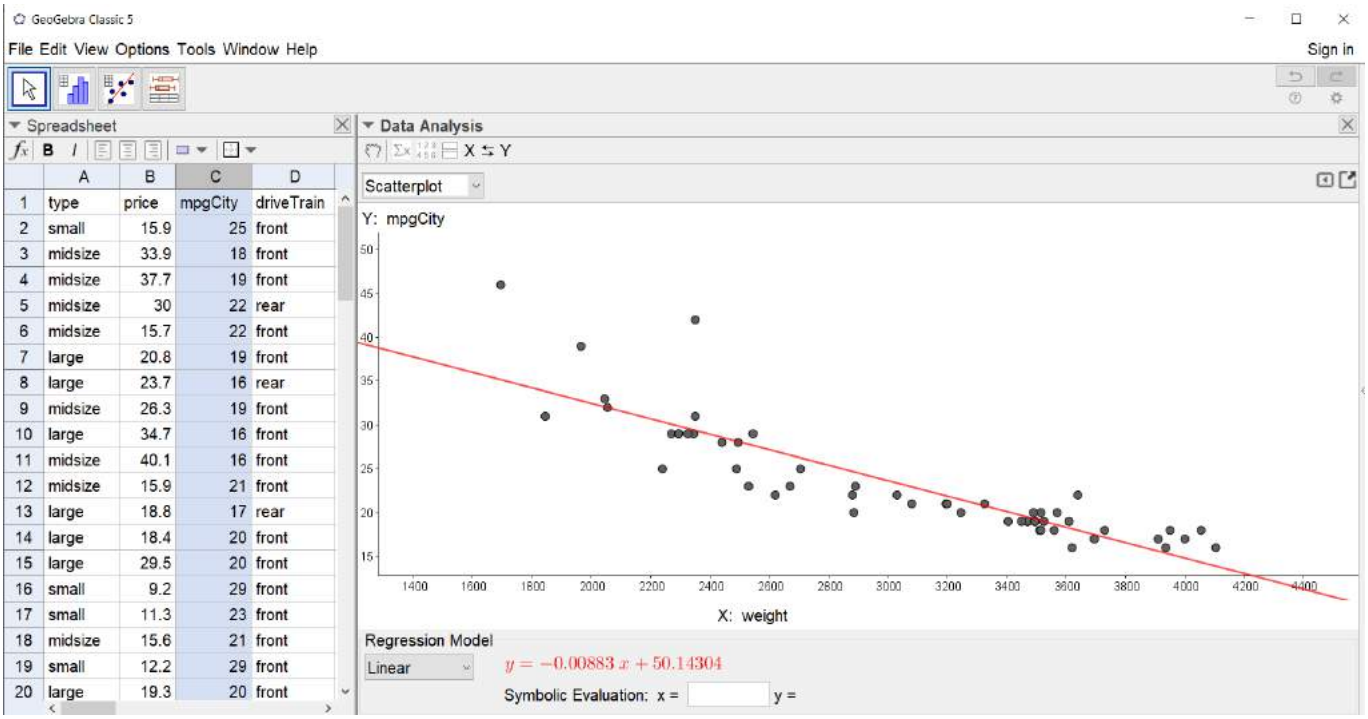
This produces a scatterplot. It is up to you to decide if this looks linear enough to proceed.



	A	B	C	D
1	type	price	mpgCity	driveTrain
2	small	15.9	25	front
3	midsize	33.9	18	front
4	midsize	37.7	19	front
5	midsize	30	22	rear
6	midsize	15.7	22	front
7	large	20.8	19	front
8	large	23.7	16	rear
9	midsize	26.3	19	front
10	large	34.7	16	front
11	midsize	40.1	16	front
12	midsize	15.9	21	front
13	large	18.8	17	rear
14	large	18.4	20	front
15	large	29.5	20	front
16	small	9.2	29	front
17	small	11.3	23	front
18	midsize	15.6	21	front
19	small	12.2	29	front
20	large	19.3	20	front

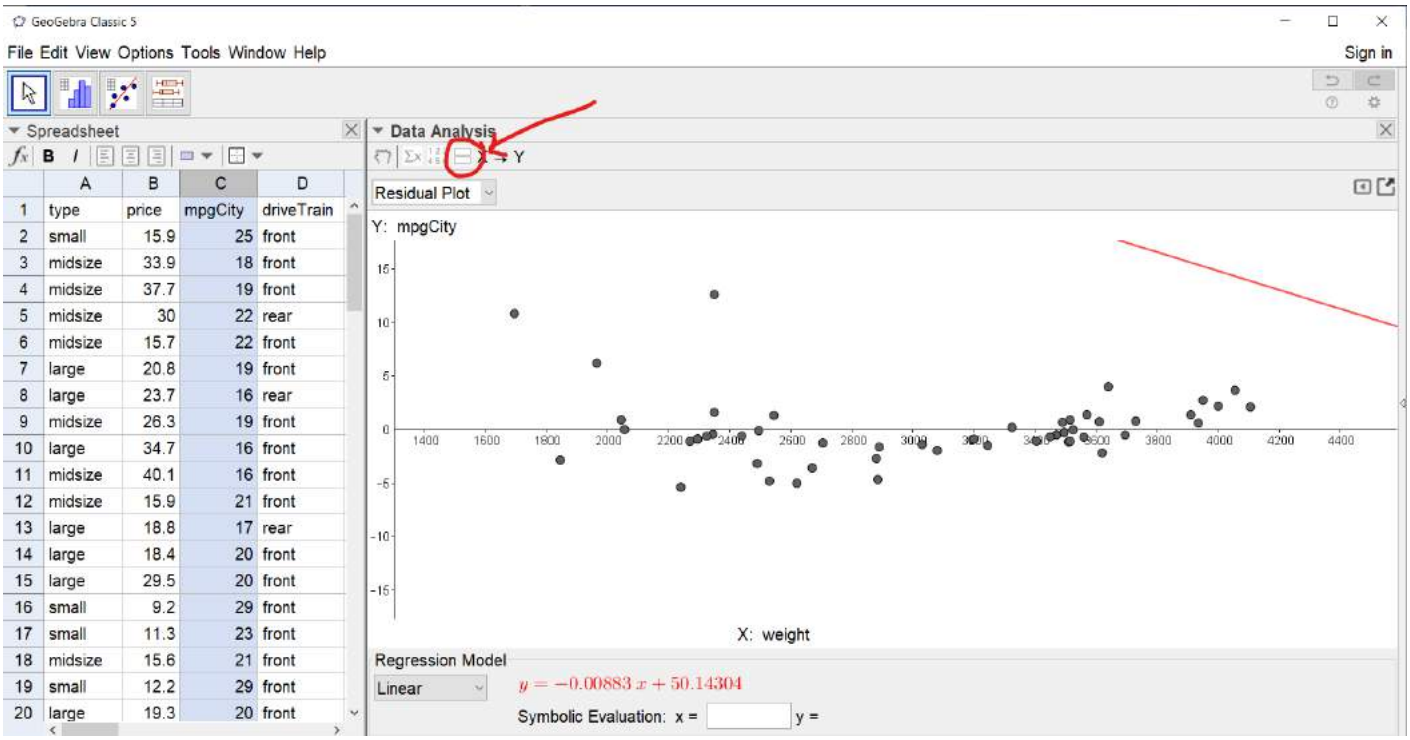
Note that you can easily switch which column of data is *x* and which is *y*. Although even a "successful" linear regression only shows correlation and not causation, it is conventional to choose the *x*-variable to be the one which would be the cause and the *y*-variable to be the one which would be the effect, if there were to exist a cause-and-effect relationship. In this case, "weight causes mpg" is more reasonable than "mpg causes weight" so we'll leave the *x* and *y* as is.

In my opinion, this does look linear enough to proceed. So the next step is to choose "Linear" from the "Regression Model" drop-down menu at the bottom.



	A	B	C	D
1	type	price	mpgCity	driveTrain
2	small	15.9	25	front
3	midsize	33.9	18	front
4	midsize	37.7	19	front
5	midsize	30	22	rear
6	midsize	15.7	22	front
7	large	20.8	19	front
8	large	23.7	16	rear
9	midsize	26.3	19	front
10	large	34.7	16	front
11	midsize	40.1	16	front
12	midsize	15.9	21	front
13	large	18.8	17	rear
14	large	18.4	20	front
15	large	29.5	20	front
16	small	9.2	29	front
17	small	11.3	23	front
18	midsize	15.6	21	front
19	small	12.2	29	front
20	large	19.3	20	front

It looks like a good fit, but as usual we need to check all the conditions. So the residual plot must also be examined. In the drop-down menu at the top, choose "Residual Plot" instead of "Scatterplot".

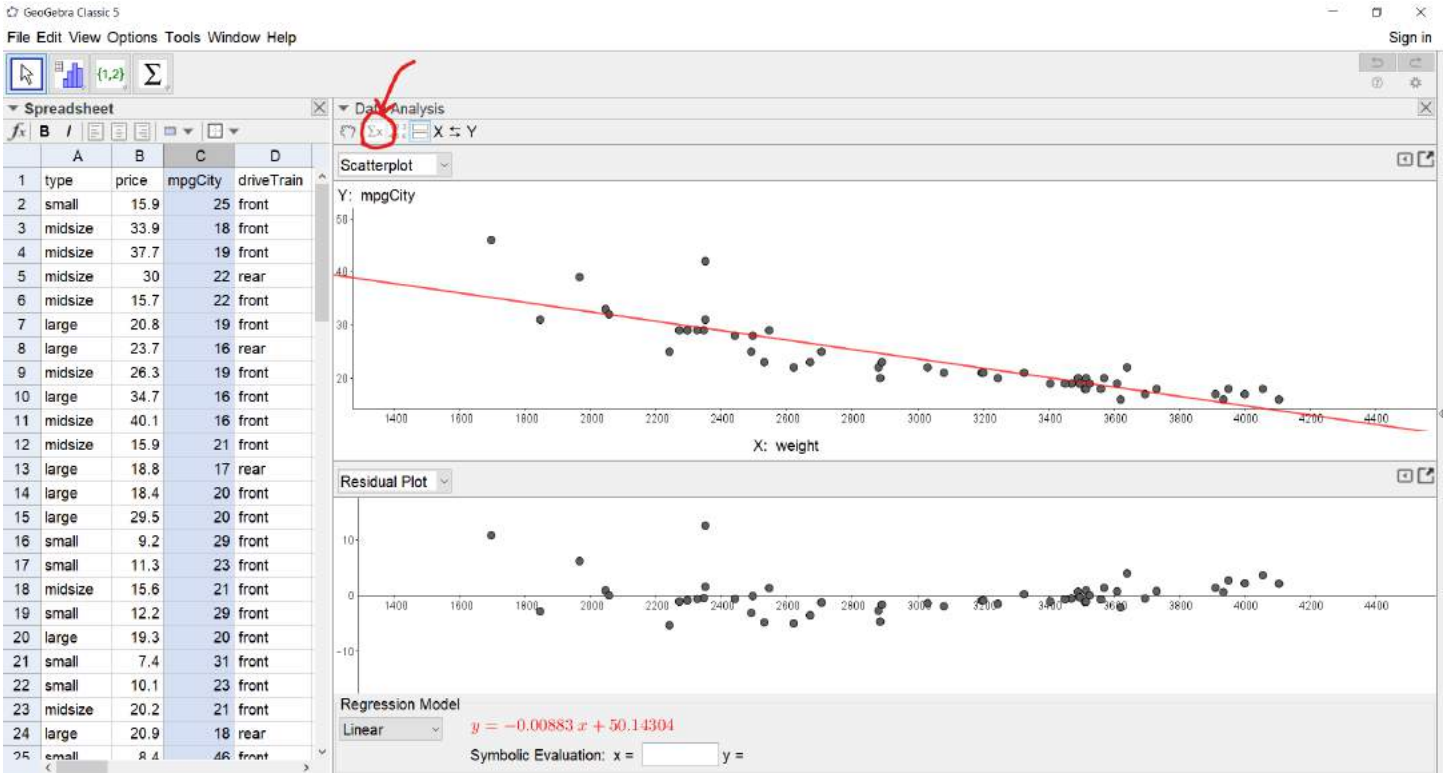


	A	B	C	D
1	type	price	mpgCity	driveTrain
2	small	15.9	25	front
3	midsize	33.9	18	front
4	midsize	37.7	19	front
5	midsize	30	22	rear
6	midsize	15.7	22	front
7	large	20.8	19	front
8	large	23.7	16	rear
9	midsize	26.3	19	front
10	large	34.7	16	front
11	midsize	40.1	16	front
12	midsize	15.9	21	front
13	large	18.8	17	rear
14	large	18.4	20	front
15	large	29.5	20	front
16	small	9.2	29	front
17	small	11.3	23	front
18	midsize	15.6	21	front
19	small	12.2	29	front
20	large	19.3	20	front

The residual plot looks like it meets the conditions, so we can consider this to be a valid linear model.

Note however that Geogebra includes the red best-fit line from the scatterplot on the residual plot. It should not do this. Residual plots should only have dots on them, and no lines (except for the axes of course).

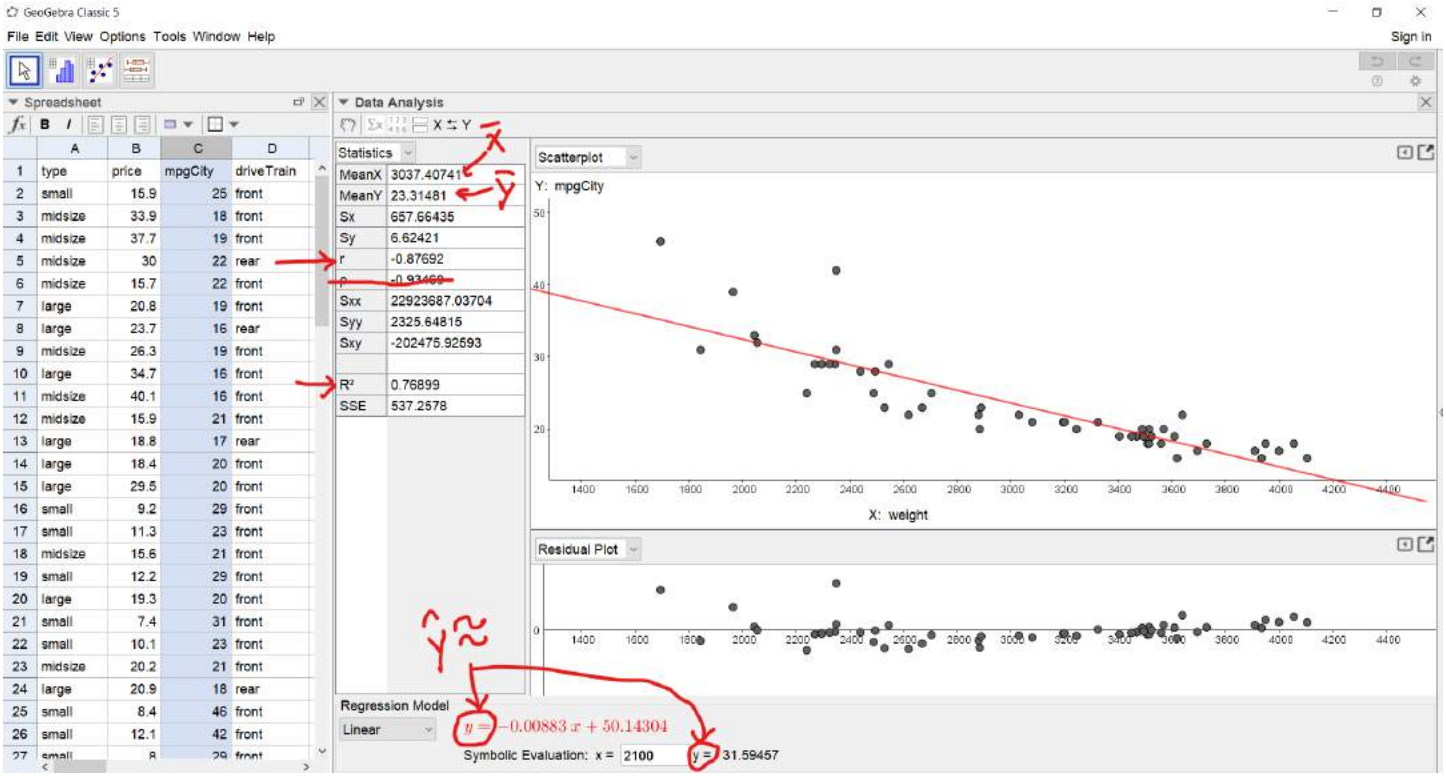
After examining the residual plot on its own, you may find it useful to view both the residual plot and the scatterplot next to each other at the same time. (You should decide if the residual plot meets the conditions by viewing this residual plot at full size though.) To view them side-by-side, click on the icon with the two rectangles.



	A	B	C	D
1	type	price	mpgCity	driveTrain
2	small	15.9	25	front
3	midsize	33.9	18	front
4	midsize	37.7	19	front
5	midsize	30	22	rear
6	midsize	15.7	22	front
7	large	20.8	19	front
8	large	23.7	16	rear
9	midsize	26.3	19	front
10	large	34.7	16	front
11	midsize	40.1	16	front
12	midsize	15.9	21	front
13	large	18.8	17	rear
14	large	18.4	20	front
15	large	29.5	20	front
16	small	9.2	29	front
17	small	11.3	23	front
18	midsize	15.6	21	front
19	small	12.2	29	front
20	large	19.3	20	front
21	small	7.4	31	front
22	small	10.1	23	front
23	midsize	20.2	21	front
24	large	20.9	18	rear

Initially, Geogebra might just display two copies of the residual plot. To produce the view above, I had to choose "Scatterplot from the top drop-down menu. I also had to click and drag the horizontal border between the two plots downward, so that the top plot wasn't so squished.

Click on the "Σx" icon to display summary statistics for the linear model. You can click and drag the vertical border between the summary statistics and the plots, to improve the view of both graphs and stats.



Note that Geogebra has some notation errors. As usual, it uses the exactly equals sign "=" even when presenting rounded values, i.e. when it should use the approximately equals sign "≈". Additionally, it displays the regression equation with a plain "y" instead of a "$\hat{y}$". This is a very important distinction of course, since the former represents values from the actual data set, whereas the latter represents predicted values.

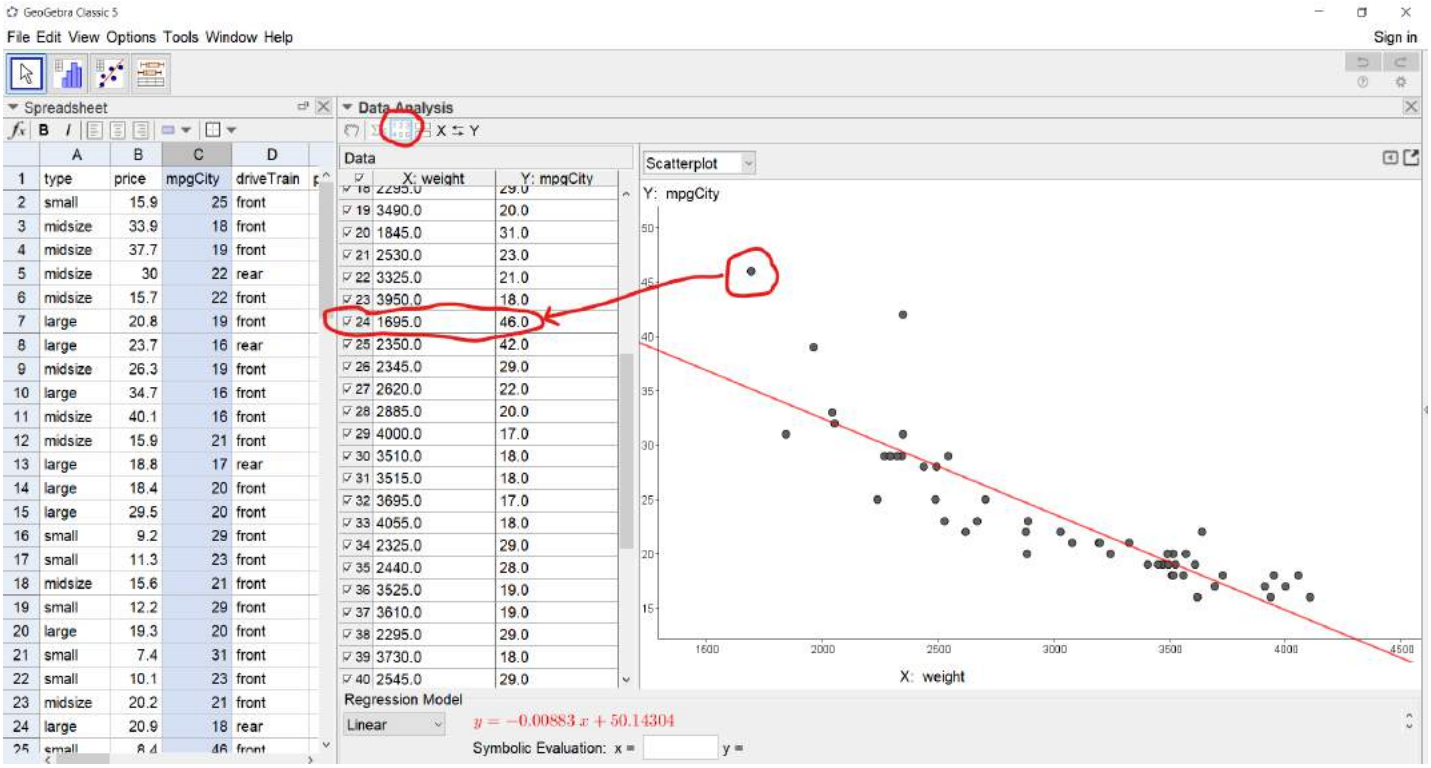
(You should also ignore the "ρ". This is not the population correlation, as you might expect since it's a Greek "r" like how "σ" is a Greek "s". It is actually something called "Spearman's rank correlation coefficient", which we aren't covering in this course.)

Anyway, we now have all the information we'd need about a linear regression on this data set. If you want to predict response variable values for given explanatory variable values, just enter the x-value into the blank at the bottom and hit enter. As noted above, what pops out is not really a y-value, it's a $\hat{y}$-value.

If you want to view just one graph again, click on the two-rectangles icon to get rid of the second graph, and the "Σx" icon to get rid of the stats.

If we have a really really good reason to do so, we can remove outlier(s) to improve the linear model. Of course we can't just do this because we want a better r^2 value, there must be some justifiable practical reason also. For example, if there is an obvious typo in the data or if one of the cases in the sample isn't actually from the population under investigation.

Suppose that the top left data point in our graph is a hybrid car, and we are only interested in studying fully gasoline-powered cars. We can remove that data point in a couple clicks using Geogebra.



Click on the "123456" icon to bring up the data, but this time it's in a table with check boxes, in the "Data Analysis" window instead of the original "Spreadsheet" window. Unfortunately we can't do something so easy as click on the dot in the graph; but since it's an outlier way at the top left, we know that it has the biggest y-value and the smallest x-value; in fact it's the only one with $y > 45$. Find this point and uncheck the box next to it.

… and nothing happens!

We need to remind Geogebra to do something. In the "Regression Model" drop-down menu, select "None". Then, re-select "Linear". Now the regression model is displayed without the outlier. Of course, it's a new model now, so everything has changed a bit. You'd need to examine the residuals again, take note of the new equation, correlation and other stats, etc.