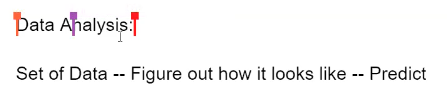
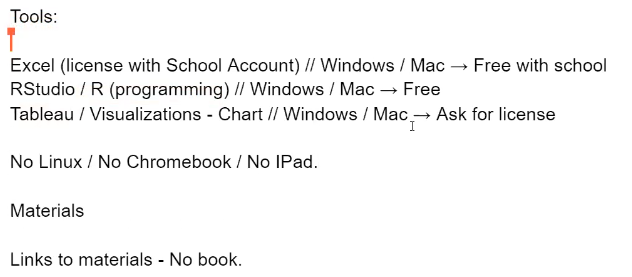
[Video Notes](https://zoom.us/rec/play/-6HbgtNDj1XEY5_kGLHoXEXJnUr0A5kCPf96zigojEZCswfCpmG1K8jFx9jpqjykjUUtsRgHzP4g_NFe.aLSKPglQCwLUHAmk)

[00:00:15]

* Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.
* Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.

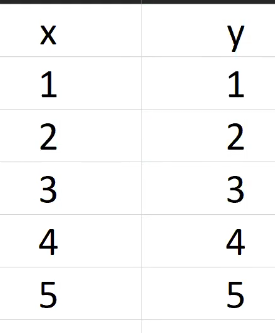


[00:01:07]

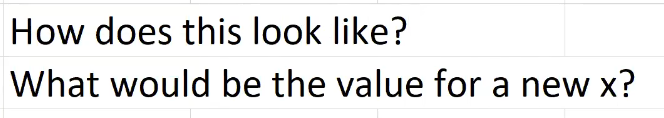


[00:03:32]

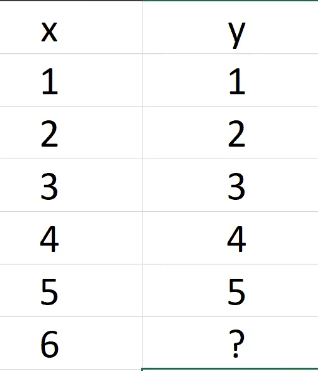
* Here we have two variables x and y.



Two questions:



* What will be the value of y if x=6?



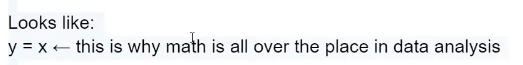
[00:05:55]

**Question**: How does this look like?

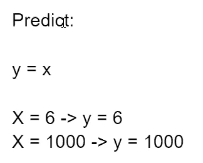
**Answer**: looks like a line



The equation will be:

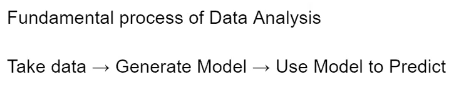


[00:07:07]

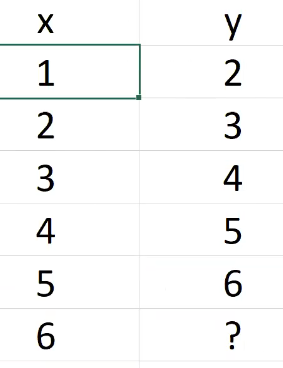


[00:07:56]

Data Analysis is a process of collecting, transforming, cleaning, and modeling data with the goal of discovering the required information.



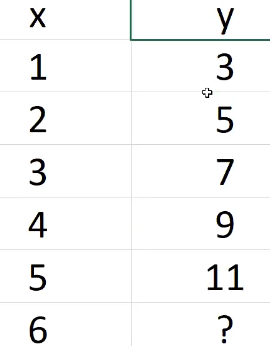
[00:08:15]



**Question**: how does this look?

**Answer**: y=x+1

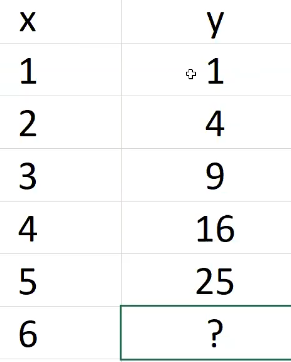
[00:09:30]



**Question**: how does this look?

**Answer**: y=2x+1

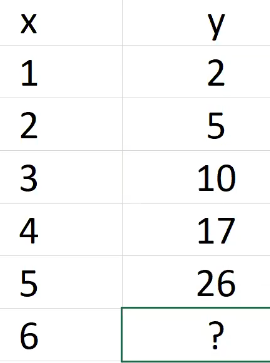
[00:10:33]



**Question**: how does this look?

**Answer**: y=x^2

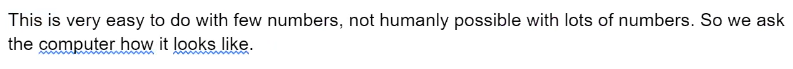
[00:11:11]



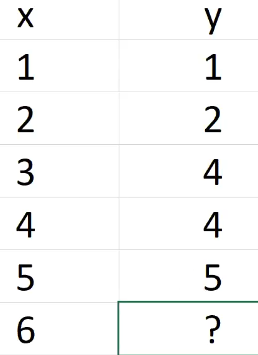
**Question**: how does this look?

**Answer**: y=x^2 + 1

[00:12:05]



[00:12:55]



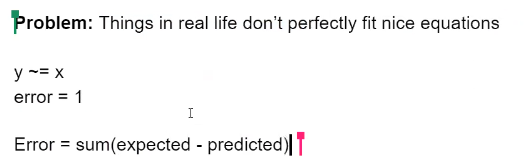
**Question**: how does this look?

**Answer**:

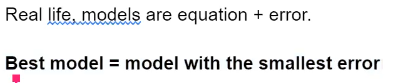
int y = x;

if(y==3) y++;

[00:14:15]



[00:16:10]



[00:16:56]

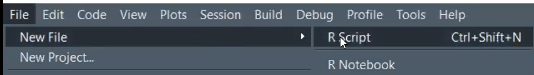
* Download R from <https://cloud.r-project.org/>

[00:21:58]

* Download RStudio from <https://rstudio.com/products/rstudio/download/#download>

[00:25:27]

* Open RStudio
* Go to New File -> R Script



[00:26:28]

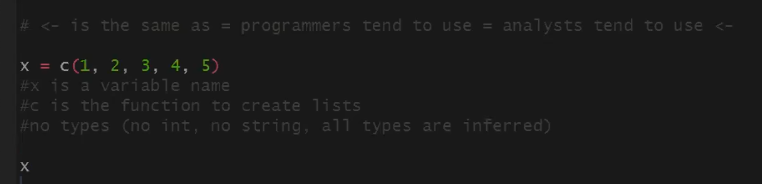
* In RStudio, use CTRL+Enter to run the code.



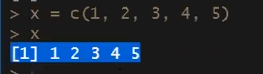


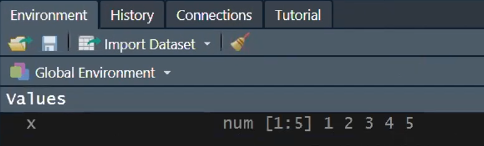
[00:28:32]

* Here we created variable x using R language.



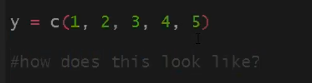
* We can print the value of variable x.





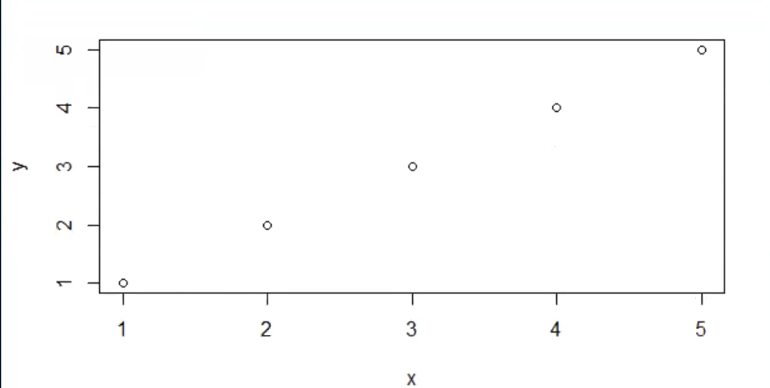
[00:30:42]

* Now we will create y variable.



* To answer the “how does this look like” question, we will use the R plot function.





* According to the chart, this looks like a line

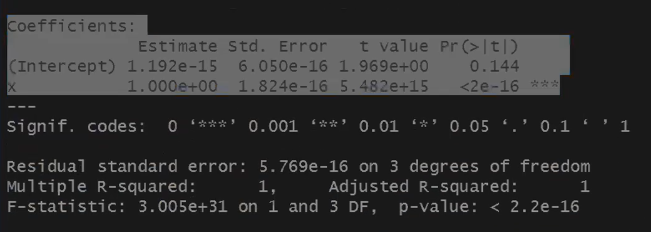
[00:31:33]

* We can use the R lm() function to get the equation of the line.
* In R, the lm(), or “linear model,” function can be used to create a simple regression model.

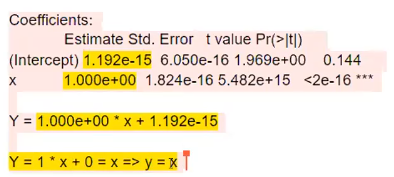


* Here lm is the function
* The ~ symbol basically means “who predicts who”. So y goes on the left and x on the right.
* Now we will summarize the model by using the R summary() function.



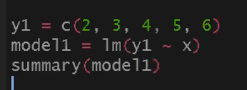


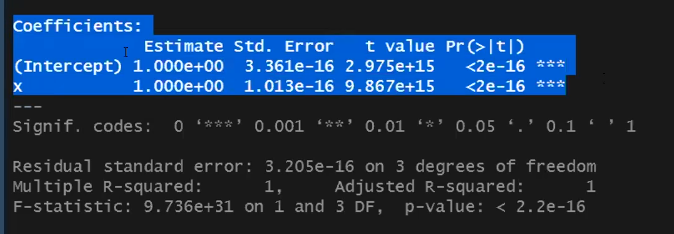
[00:33:03]

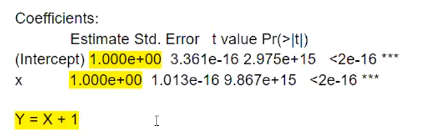


[00:34:01]

* Now we will calculate the equation of the model for different values of y.





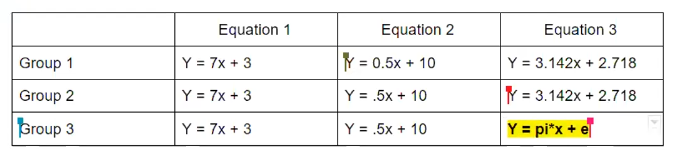


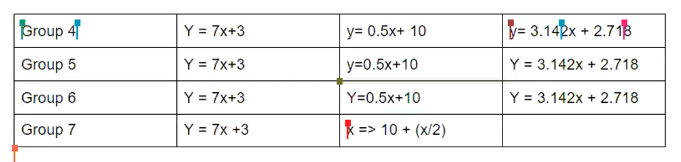
[00:35:33]

**Activity**:

* Use R and find the equations for given different values.

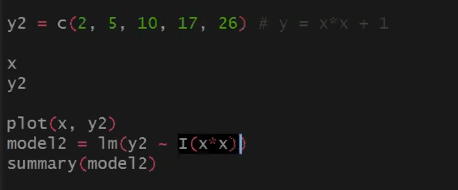
[00:49:38]

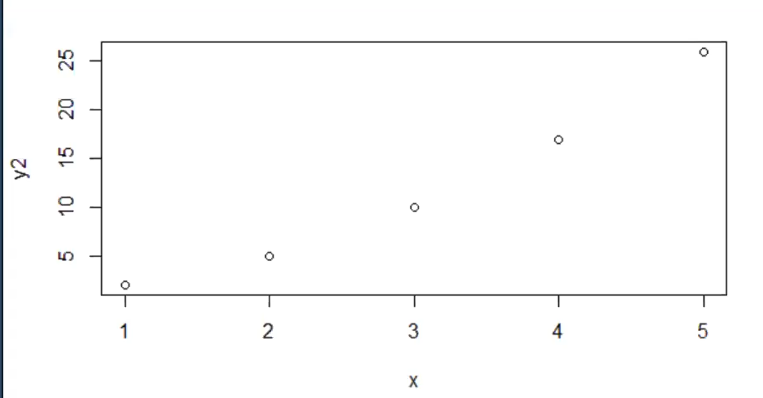


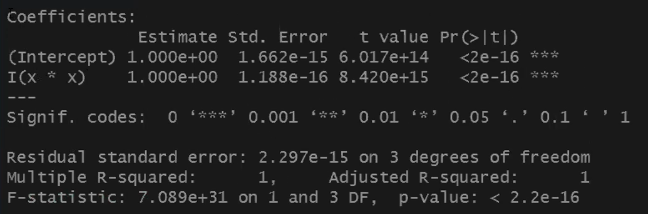


[00:50:36]

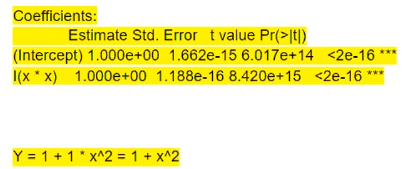
* Here we plotted a chart for the equation y = x\*x + 1





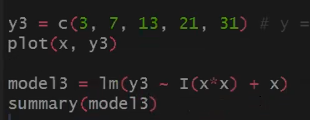


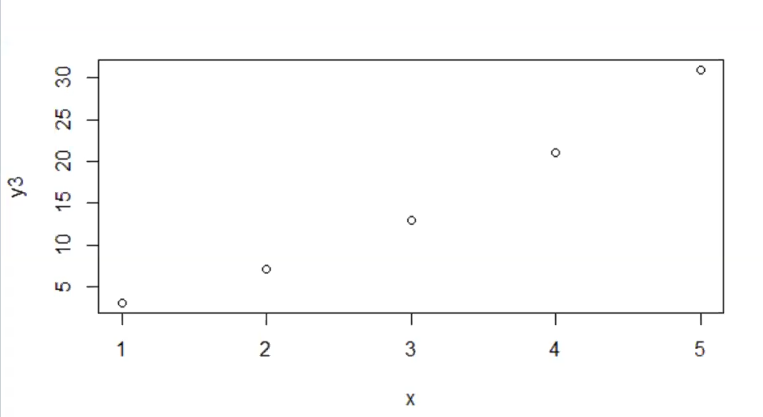
[00:56:25]

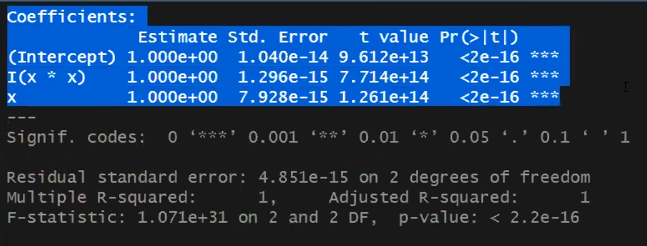


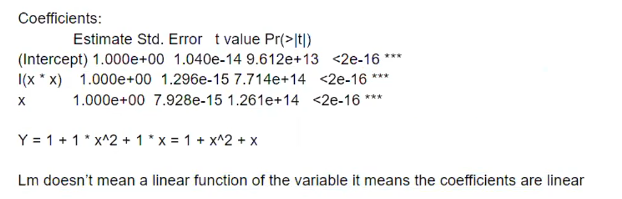
[00:56:55]

* Calculating model for the equation y = x\*x + x



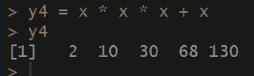






[00:59:23]

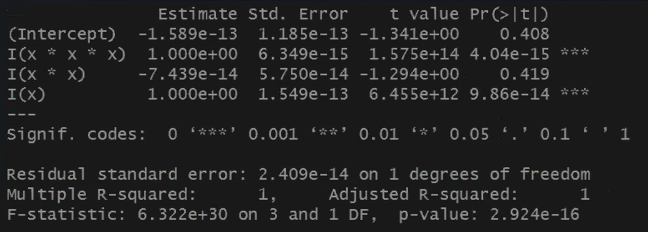




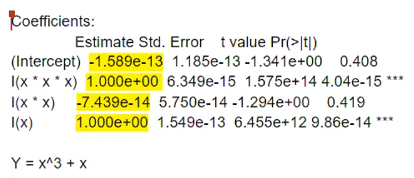
[01:00:19]

* Now we will try to create a model for y4.



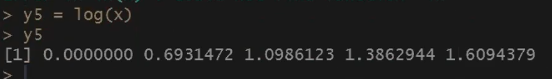


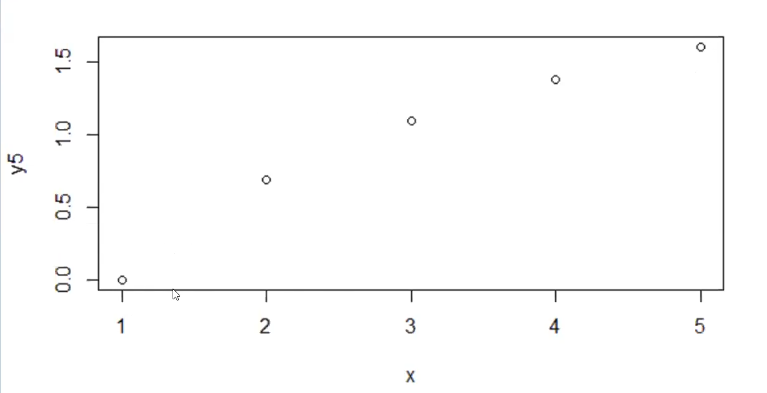
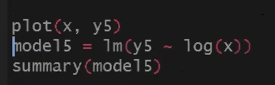
[01:01:16]

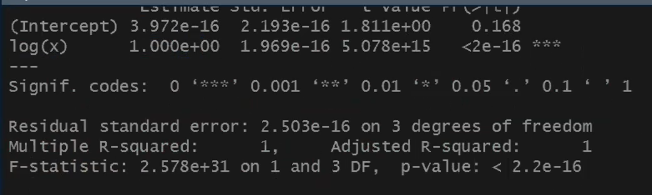


[01:02:27]





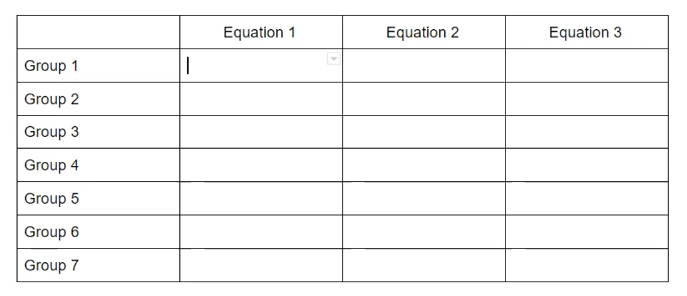




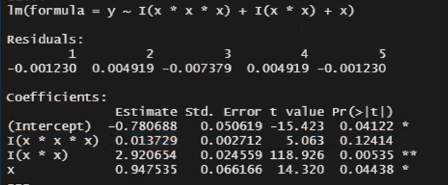
[01:22:46]

**Activity**:

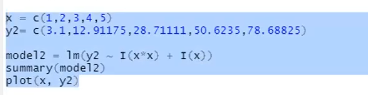
* Try reverse engineer the given formula for the given values.

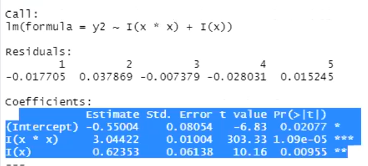


[01:43:00]

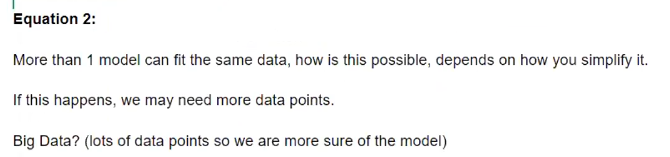


[01:45:40]

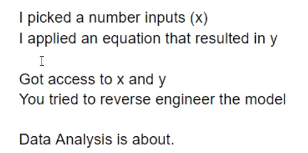




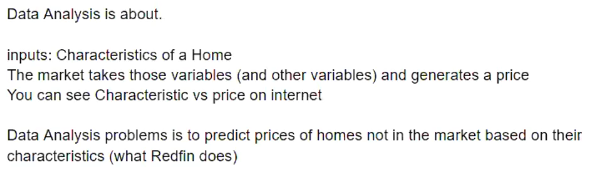
[01:46:26]



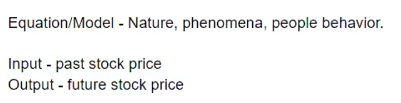
[01:48:00]



[01:48:37]

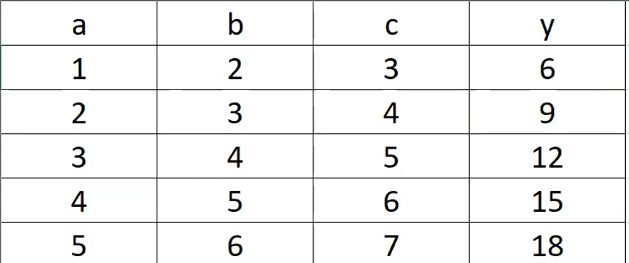


[01:49:56]



[01:51:28]

* Here we will have 3 variables.



The equation of the above model can be:

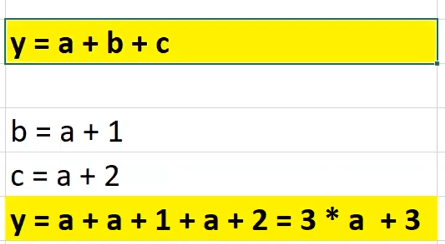
**y = a + b + c**

b = a + 1

c = a + 2

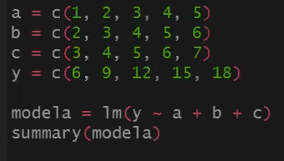
y = a + a + 1 + a + 2

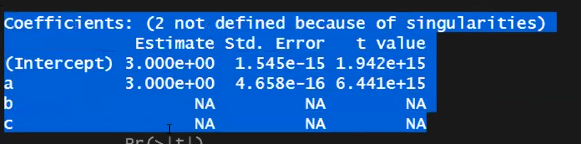
**y = 3 \* a + 3**

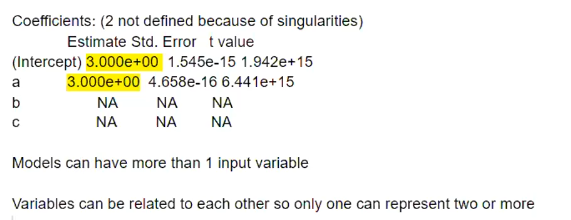
****

[01:53:50]

* Creating the model in R using those points and then summarizing it.



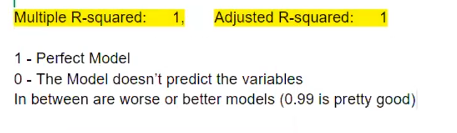




[01:56:15]

**Question**: how do we know that the model is good?

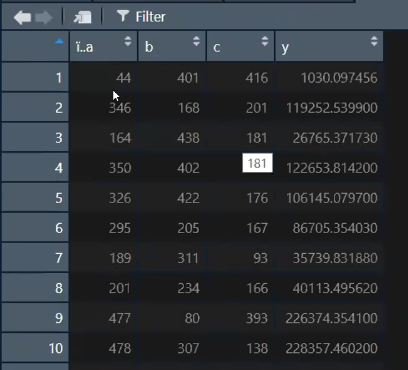
* For that, we will look at the R-squared values.
* R-squared is a statistical measure of how close the data are to the fitted regression line.
* If the value of R-squared is 1, then it is a perfect model.
* If the value of R-squared is 0, then the model cannot predict the variables.



[01:59:48]

* In the R language, the read.csv() function is used to read csv files.





[02:01:30]

**Activity**:

* Find the equation for y as a function of a,b and c.

*Here to access any column, the $ symbol is used.*

