Analyzing two variables

Data

Let’s use the schools data again. But this time we’re interested in how two variables move *together*.

Types of questions:

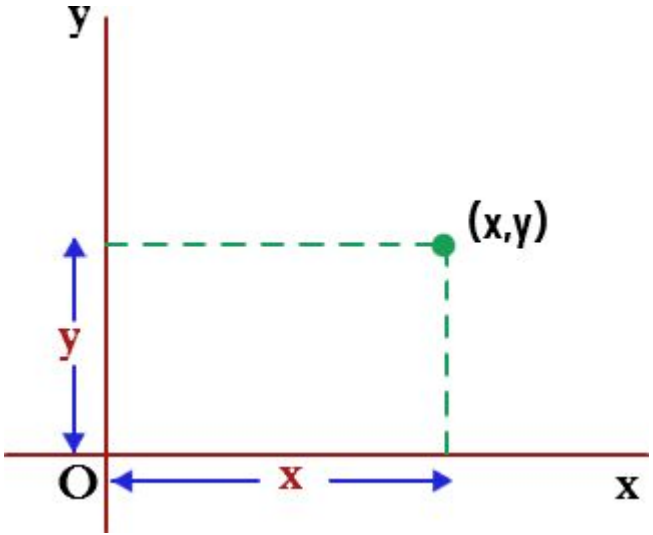
As the rate of poor students at a school increases, what happens to test scores?

Is there a relationship between student teacher-ratio and math test scores?

Cartesian coordinate review

In general, the x-axis is the thing we think is a predictor.

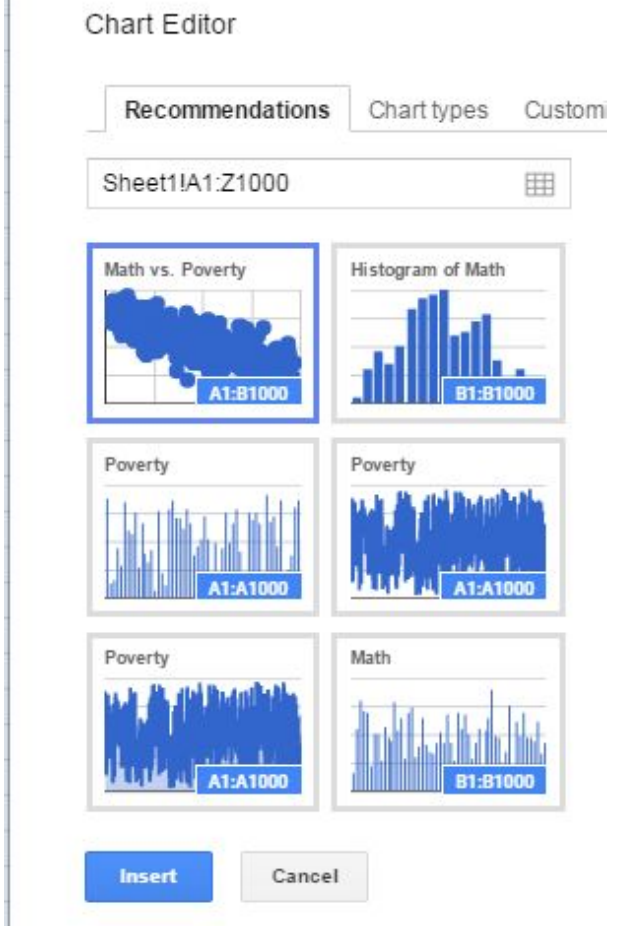
The y-axis is the thing we think might be related to the x-axis.



Are poverty and test scores related?

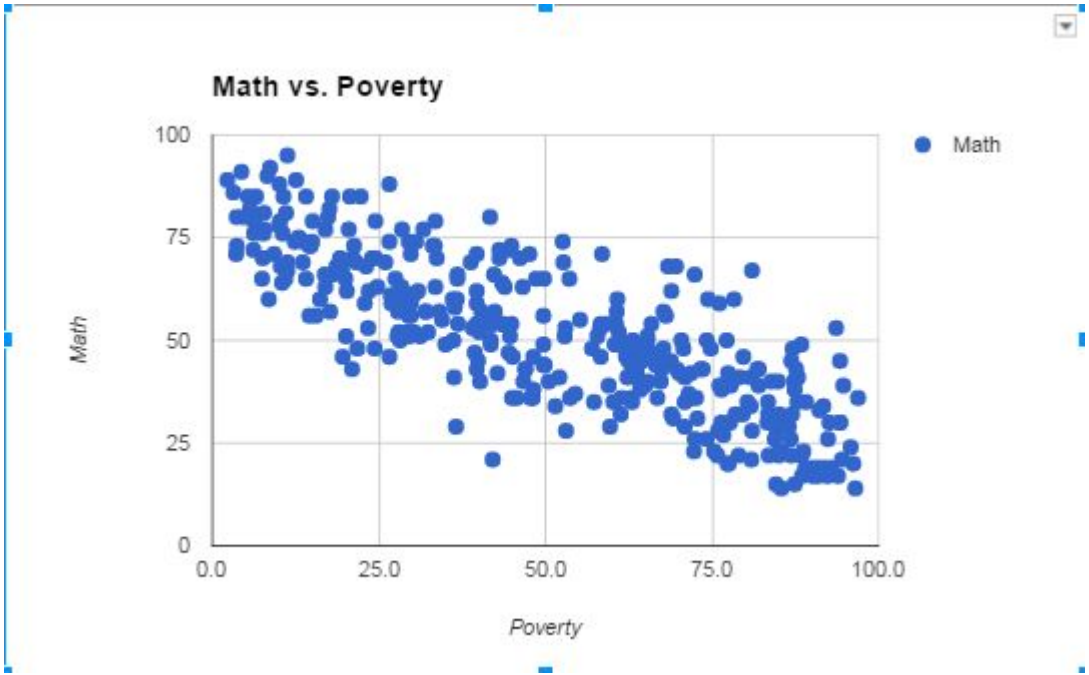
To assess this, we want to plot poverty rates and school average test scores on our x and y axis.

- Select the poverty and math score columns and paste them into a new tab.
- Choose Math vs. Poverty from the options



Are poverty and test scores related?

Your result should look like this:



Interpreting a scatterplot

If y goes up as X goes up, that is a positive correlation.

If y goes down as X goes up, that is a negative correlation.

There are two uses for a scatterplot:

1. What is the relationship between the variables overall?
2. Are there ‘outliers’, or observations that don’t fit the pattern. This can be the source of good reporting leads.

correlation

Calculate the correlation between two columns in Google Sheets with

```
=CORREL(A2:A,B2:B)
```

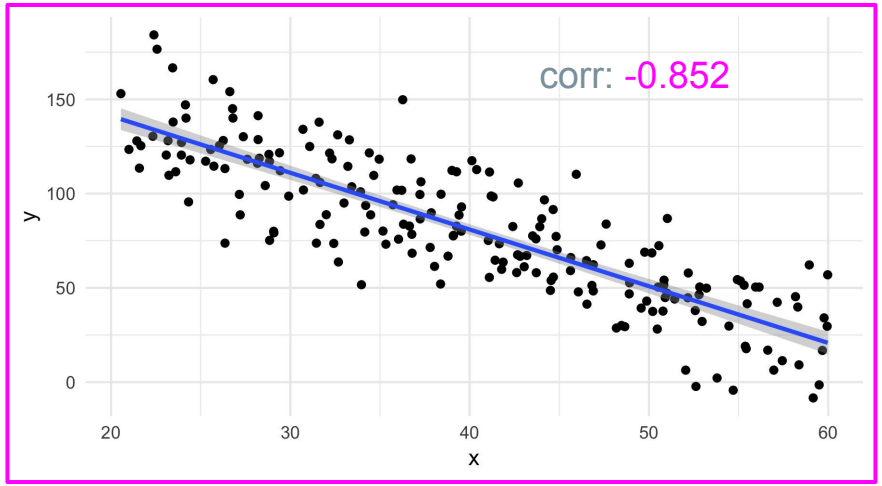
This will return a single value, a correlation coefficient. The value measures how close the two variables are to having a perfectly linear relationship with each other.

It will always be between -1 and 1.

- -1 : perfectly negatively correlated
- 0 : no correlation
- 1 : perfectly positively correlated

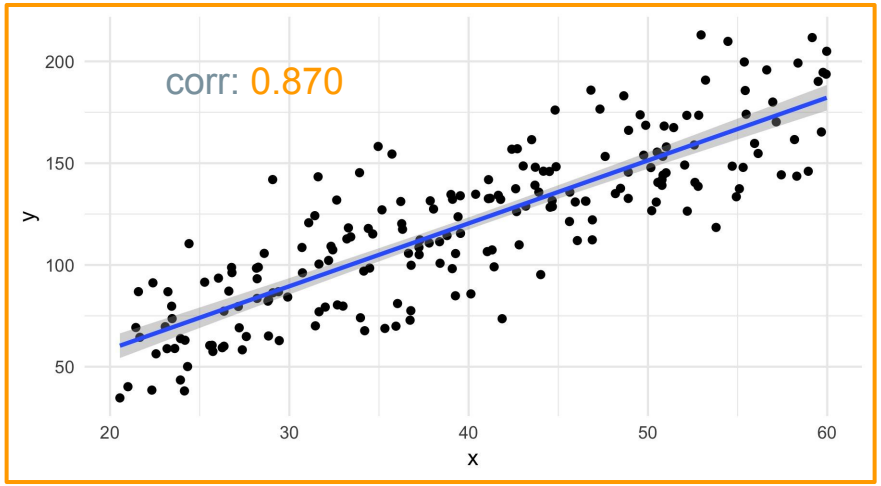
Negative correlation:

as one variable goes up, the other goes down



Positive correlation:

as one variable goes up, the other also goes up



Slope and correlation are different concepts