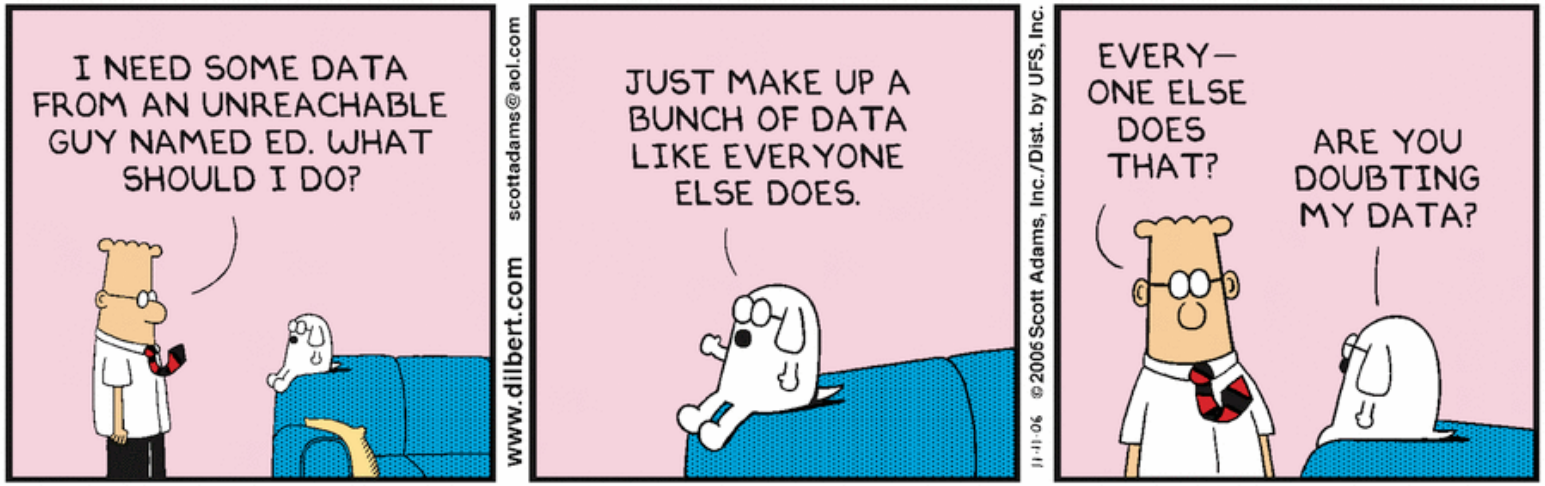


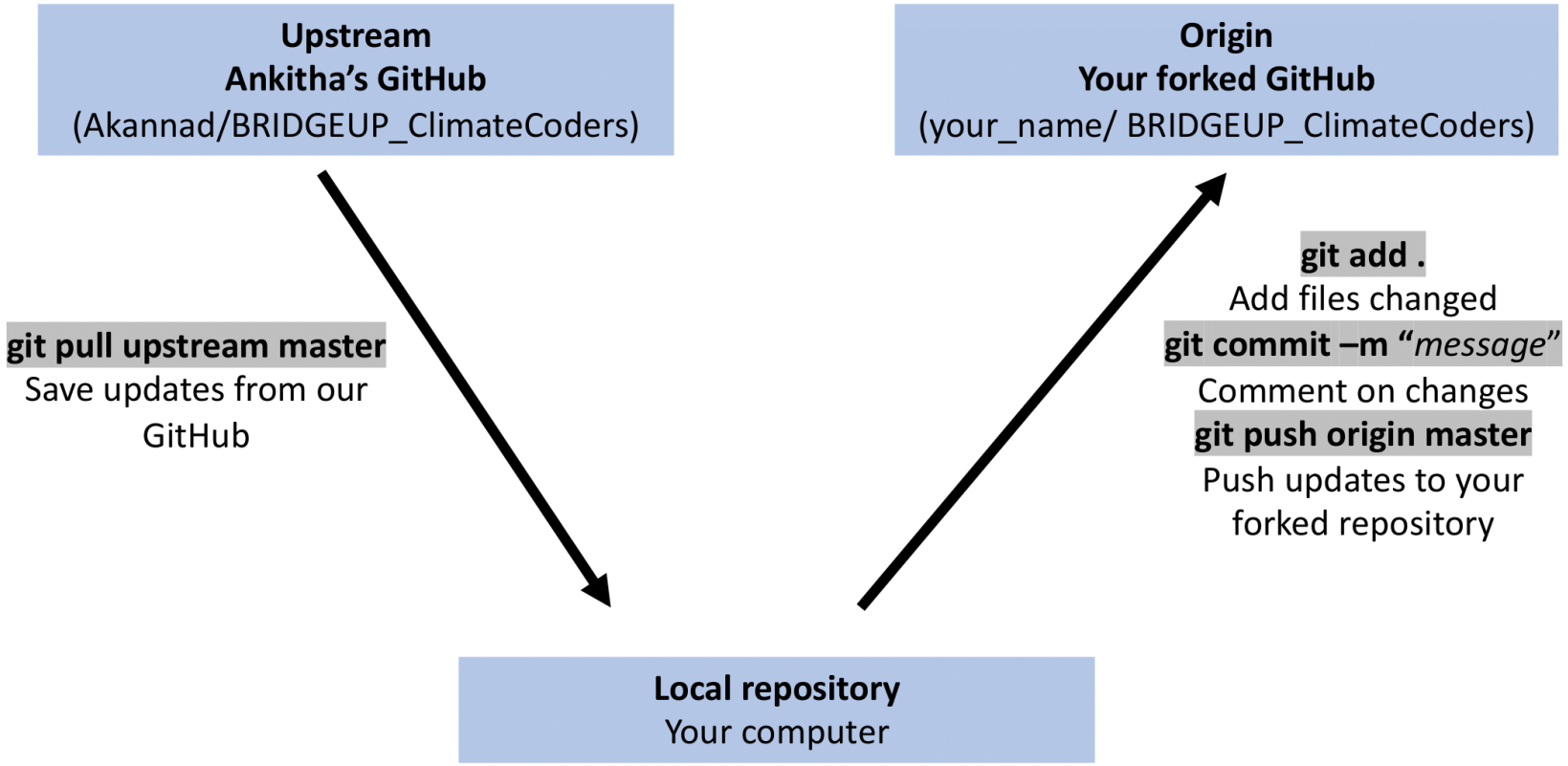
READING IN OUR DATA

UNIT 3: RECONSTRUCTING CORAL CORE DATA

MARCH 17TH 2020

HOUSEKEEPING

- Zoom guidelines:
  - Sharing errors on Slack
    - Short description of error and what you tried out already
    - Error message
    - Section of code/ terminal
  - Break-out rooms
    - Separate chat/ video rooms to work as a group



PLAN FOR TODAY

- ❑ Finish reading in the coral data files (from our Dropbox folder)
- ❑ Introduce linear regression which we will use to convert our Sr/Ca data to SST
- ❑ Update your lab notes
- ❑ Exit survey

BREAK-OUT ROOMS

- In your lab notes,
  - Use pseudo-code, to outline the steps you will need to take to read in data

Hint: A quick summary of some useful functions are in the next slide

- Check if there are any shared errors
  - If you fixed it together, save a copy of the error and code in your lab notes with a short description of how you fixed it
  - Otherwise, post a screenshot of your code (just what is relevant!) and the error message on Slack.

USEFUL FUNCTIONS

Read in delimited text files:

pandas.**read_table**(*filepath, sep, header, skiprows, skipfooter*)

- **filepath**: path of file as string
- **sep**: separates the files in a delimited text file. Ex: period, comma, colon, etc.
- **header:** index of row which corresponds to the name of the columns
- **skiprows**: numbers of lines to skip at the start of the file
- **skipfooter**: number of lines at bottom of file to skip

Drop rows or columns:

pandas_dataframe.**drop**(*labels, axis*)

- **labels**: "column name" or if you have multiple ["col_name1", "col_name2", "col_name3"]
- **axis**: axis = 0 for rows, axis = 1 for columns

For example:

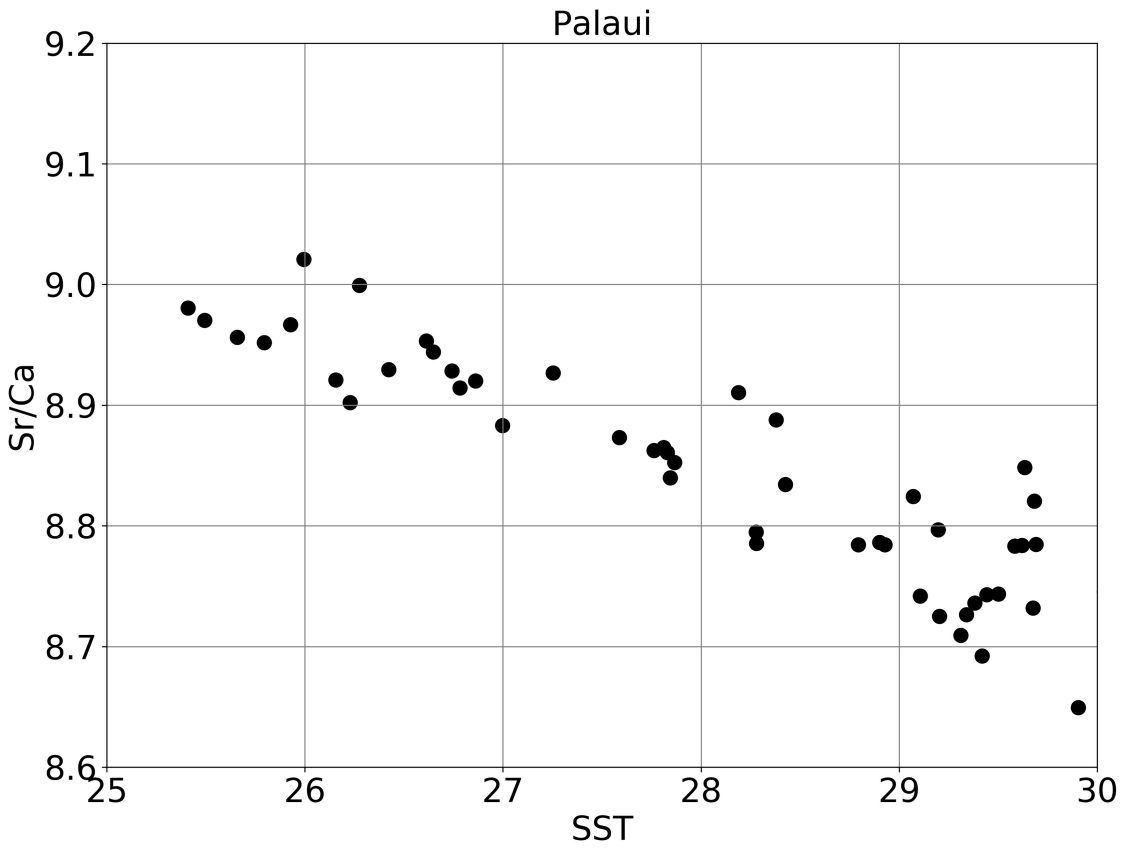
- df.**drop**(labels = ['Date','Topic'], axis = 1)

OUR DATA

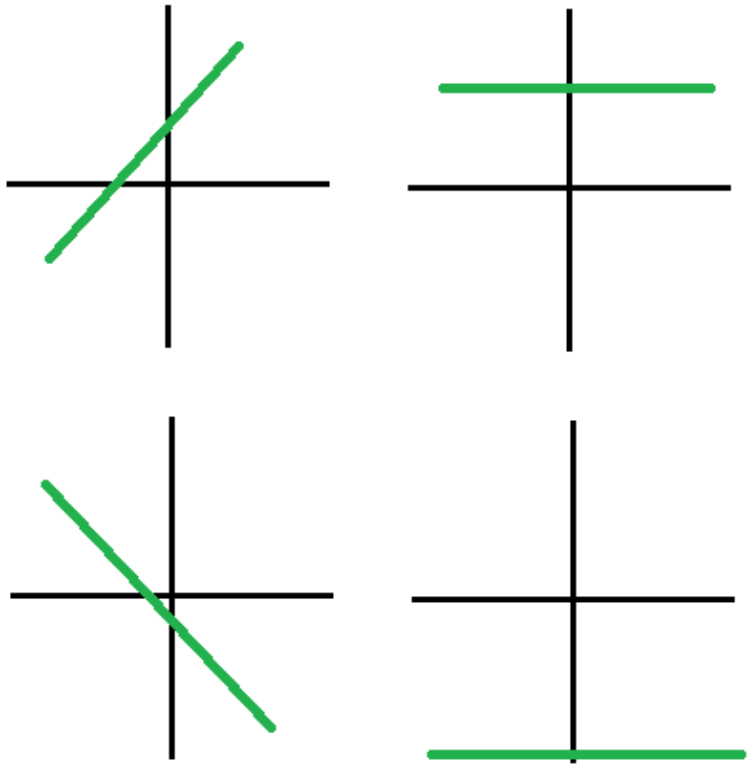
Sr/Ca → coral data

SST → SODA SST

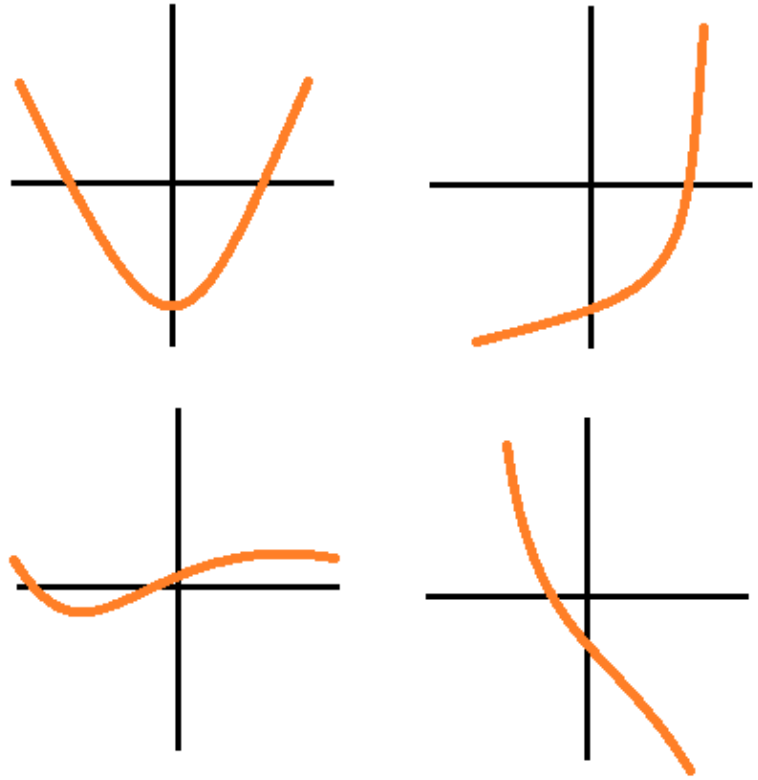
What is the relationship between Sr/Ca and SST?



IS THIS LINEAR OR NON-LINEAR?



Linear Functions



Nonlinear Functions

OUR DATA SEEMS PRETTY LINEAR!

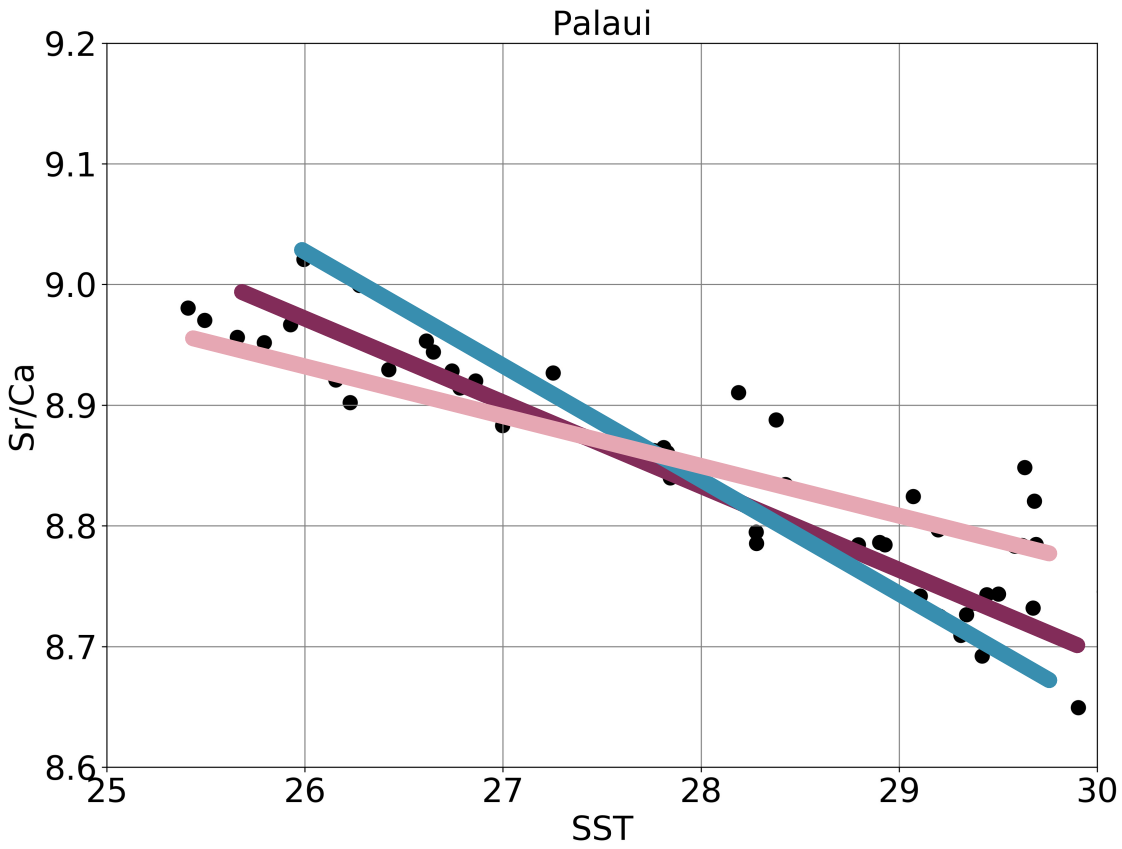
Linear equation:

$y = m x + C$

m : slope of the line

C : y-intercept

But which line describes the data best?

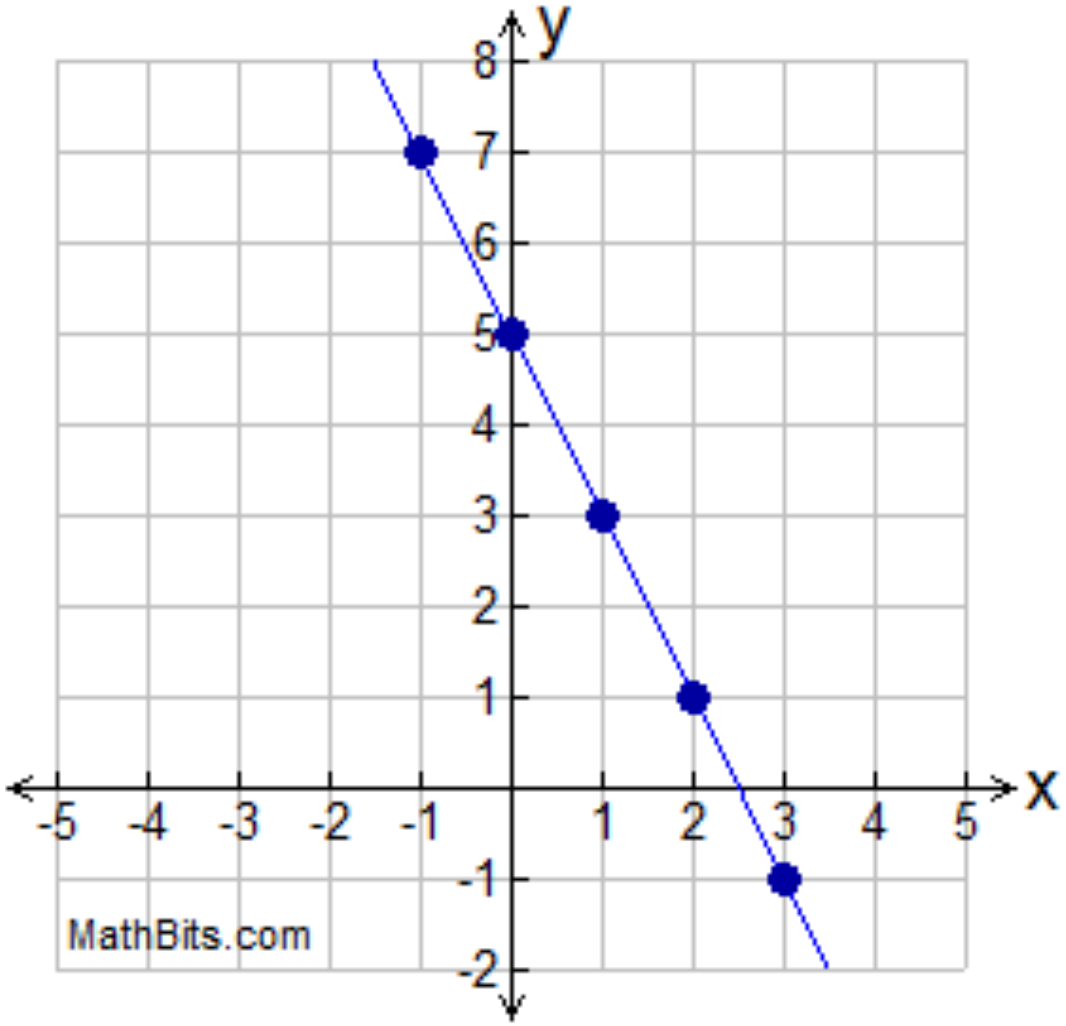


QUICK REFRESHER

$m = \frac{1 - 5}{2 - 0} = -2$

C = 5

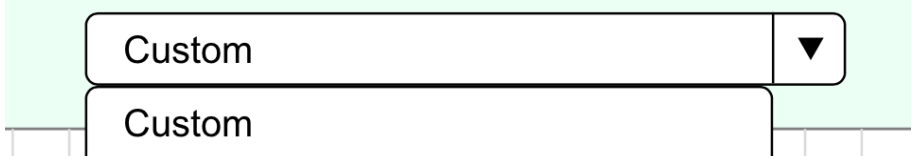
y = -2 x + 5



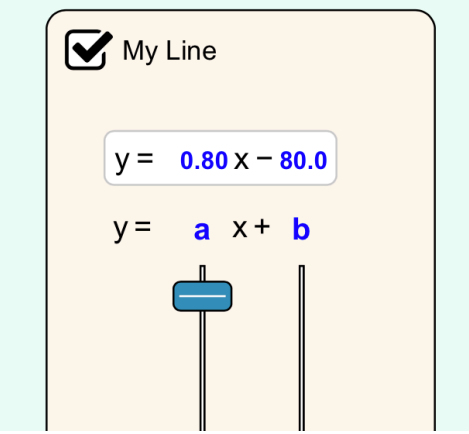
LINEAR REGRESSION ACTIVITY

Go to https://phet.colorado.edu/sims/html/least-squares-regression/latest/least-squares-regression_en.html

- Select a dataset from the dropdown menu →



- Select "My Line" and try to fit your data by changing a and b



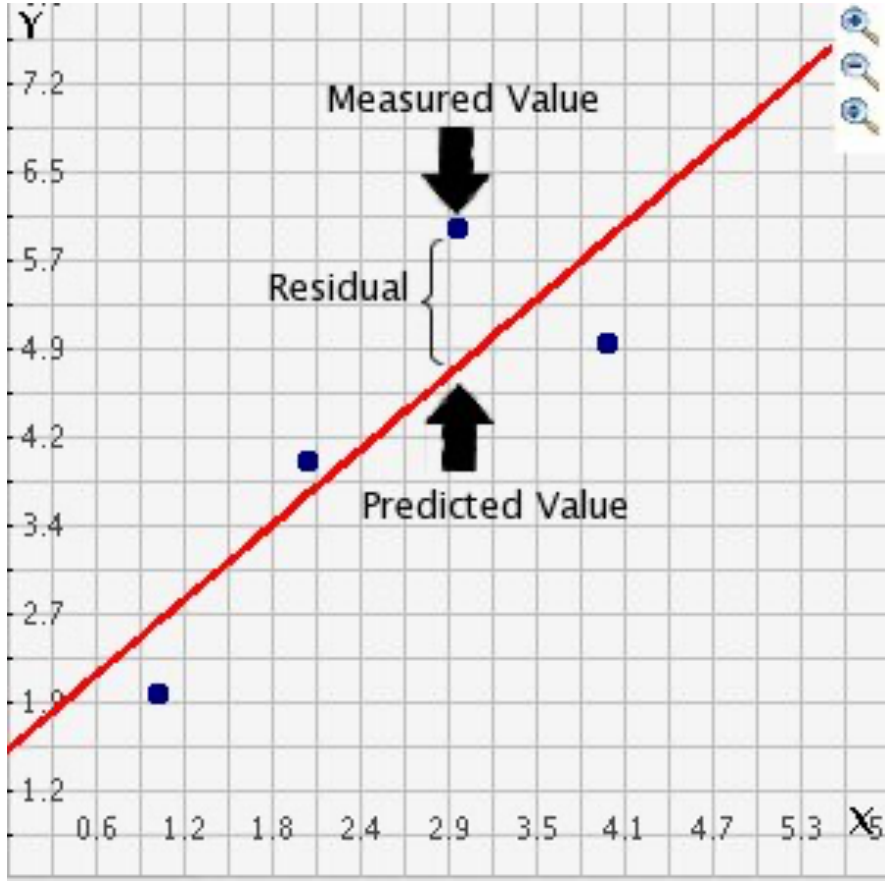
How do we decide which line is a better fit?

ONE WAY IS TO MINIMIZE RESIDUALS!

- **Residual = distance of point from its predicted value on the line**

Check your understanding:

https://www.khanacademy.org/math/statistics-probability/describing-relationships-quantitative-data/regression-library/e/calculating-interpreting-residuals



UPDATE LAB NOTES

EXIT SURVEY