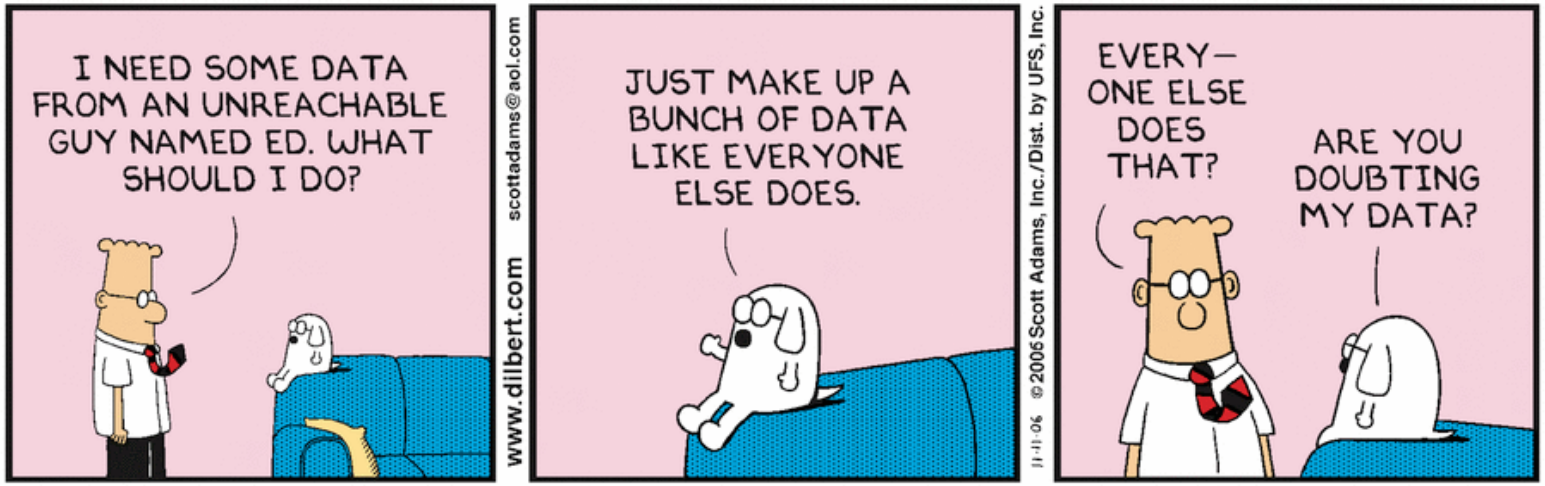


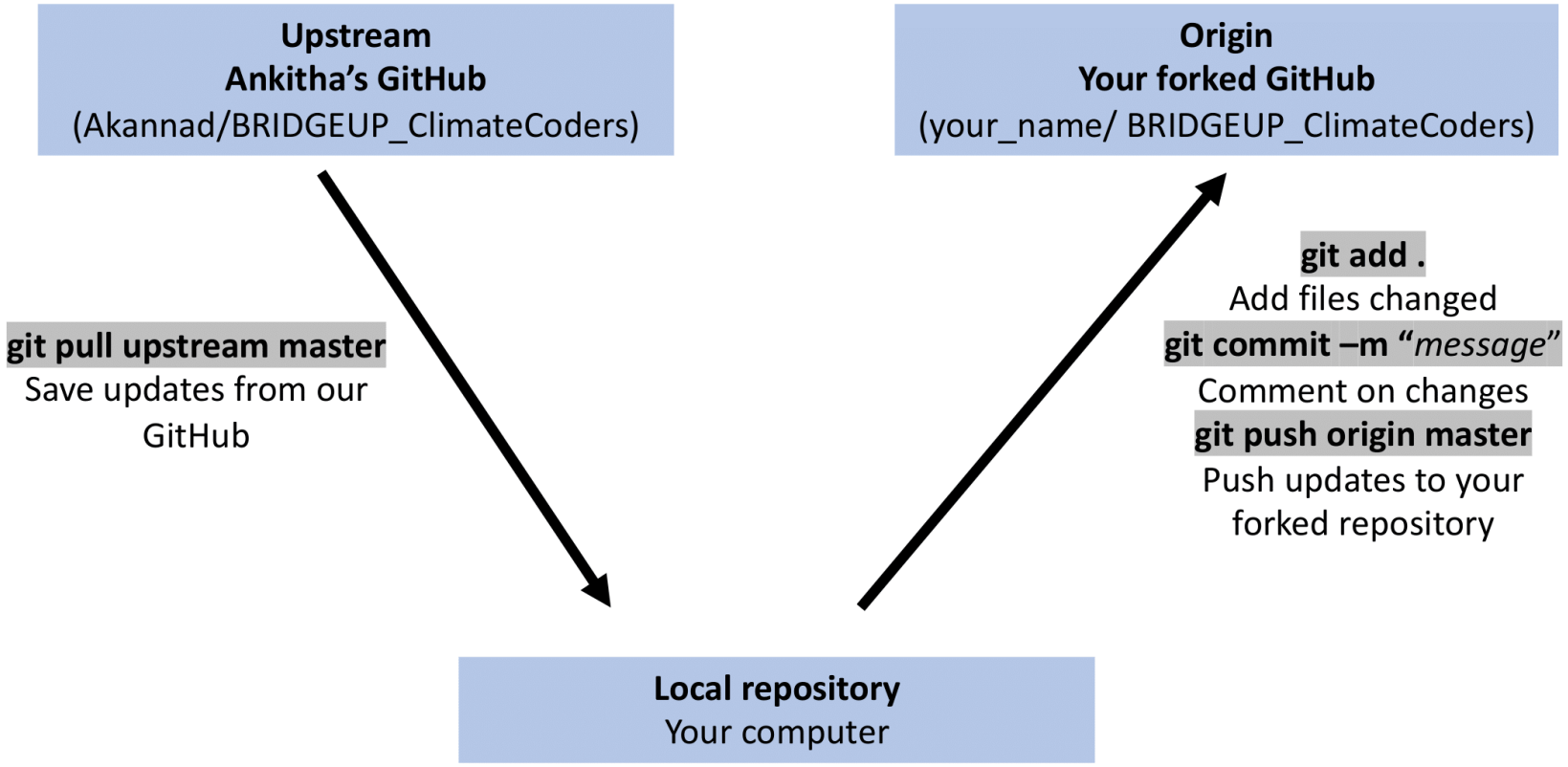
CONVERTING TIME FORMAT

UNIT 3: RECONSTRUCTING CORAL CORE DATA

MARCH 19TH 2020

HOUSEKEEPING

- Headphones
- Zoom guidelines



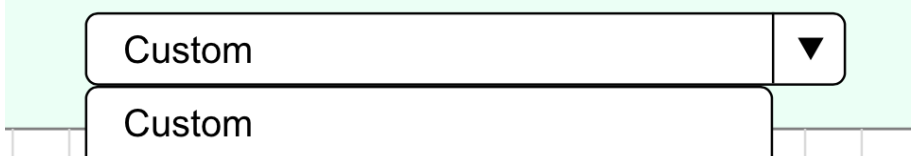
PLAN FOR TODAY

- ❑ Continue discussion on linear regression
- ❑ Finish reading in the coral data files (from our Dropbox folder)
- ❑ Convert time column to a readable format
- ❑ Update lab notes
- ❑ Exit survey

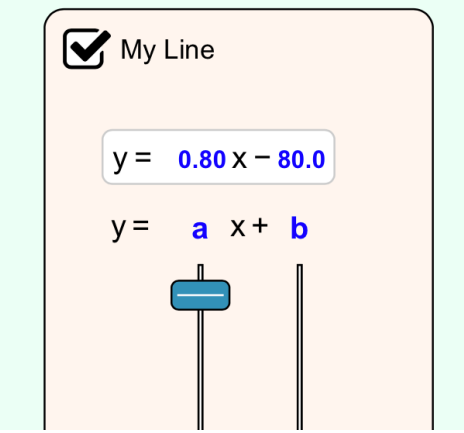
LINEAR REGRESSION ACTIVITY

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



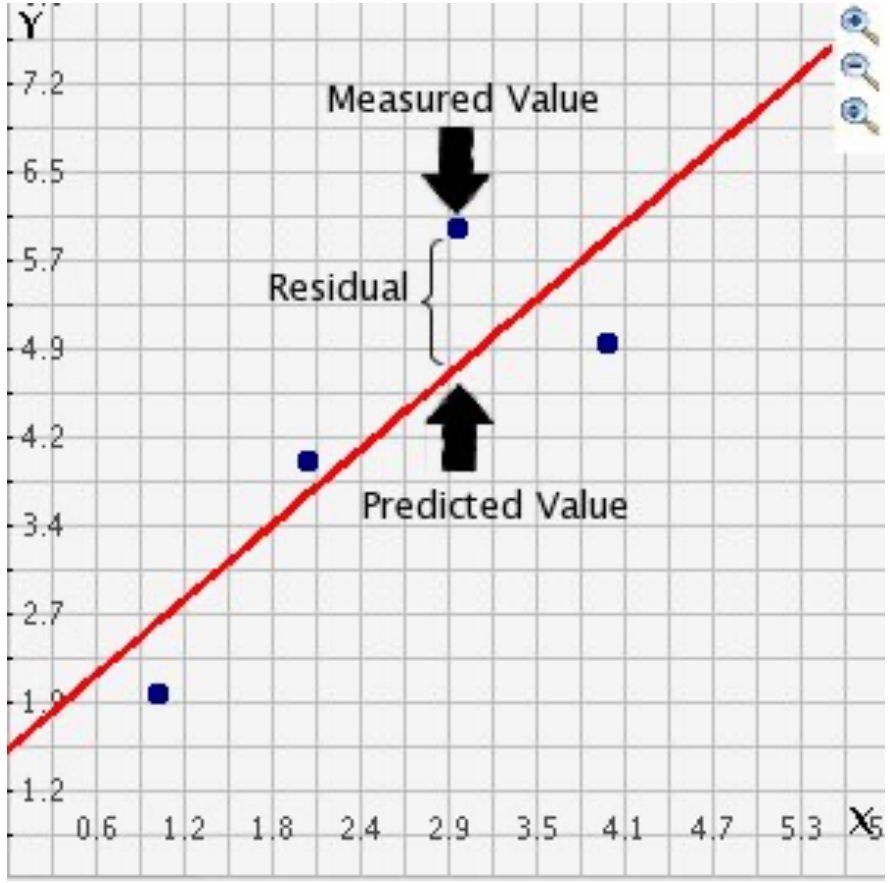
How do we decide which line is a better fit?

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



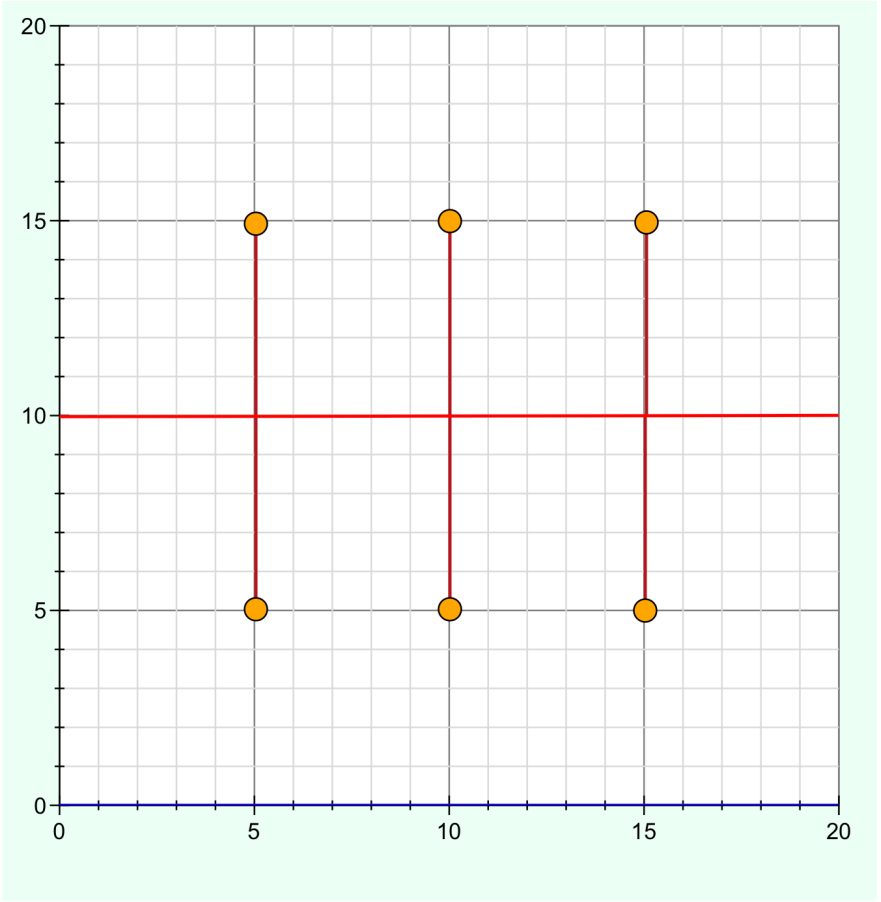
TO GET AN OVERALL MEASURE OF BEST FIT

We can sum up the residuals

What might be potential issue with this method? What is the sum of the residuals for the graph shown?

The positive and negative residuals are going to cancel each other out (even though it is a good fit)

How might we fix this?



SQUARED RESIDUALS

We square them!

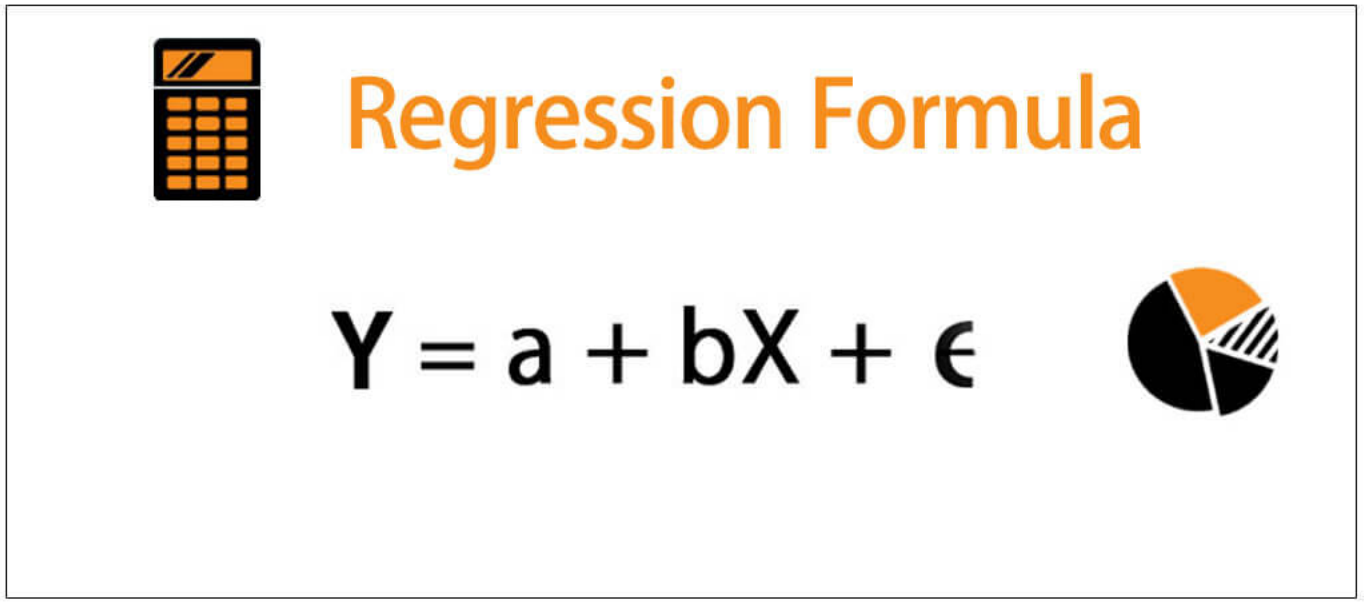
Why is this a good idea?

Try it out:

- Check "Squared Residuals". Further fine tune your fit by minimizing the sum of the squared residuals
- Compare to the "Best-Fit Line"

LINEAR REGRESSION

- Linear regression uses calculus to find the minimum sum and give you the equation for your best-fit line
- What's
  - **a =**
  - **b =**
  - **ϵ = error**



BREAK-OUT ROOMS

- Finish your script to read in data files
  - **Make sure you comment your code!**
  - Drop unnecessary columns
  - Those who have finished, share you screen and explain your steps
  - Plot your data – any interesting trends? Talk about them in your lab notes

- Vietnam team: I wanted you to practice reading in .txt files but you will be working with unpublished data in excel format which extends back to 1600s!
  - File name: **HonTre_Vietnam_SrCa.xlsx**
  - Sheet name: SrCa
  - Make sure you can read this in

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)

CODING CHALLENGE

Dates are stored in a digital time format

For example

1880.5 = 1880 + ½ of a year = 06/1880 or June 1880

Goal: Convert your date column into two separate columns of year and month

- Work through the "200319_convert_time_format" Jupyter notebook

UPDATE LAB NOTES

EXIT SURVEY