

# Regression

## Computer Science Retreat

1. Get the 'JohnsonJohnson' dataset:
  - `from pydataset import data`
  - `dataset = data('JohnsonJohnson')`
2. Plot the data - what is it showing?
3. Make a linear fit to the data and plot it overlaid on the data.
4. Quantify the error, is it a good fit?
5. Now make a polynomial fit to the data. What is the lowest order that gives a reasonable fit?
6. Notice the annual structure to the data, what could be going on?
7. Now make a fit that will include the annual variations as well (tips - what kind of function could you use? - to fit arbitrary functions use `leastsq` from `scipy.optimize` and make sure to give sensible input parameters)