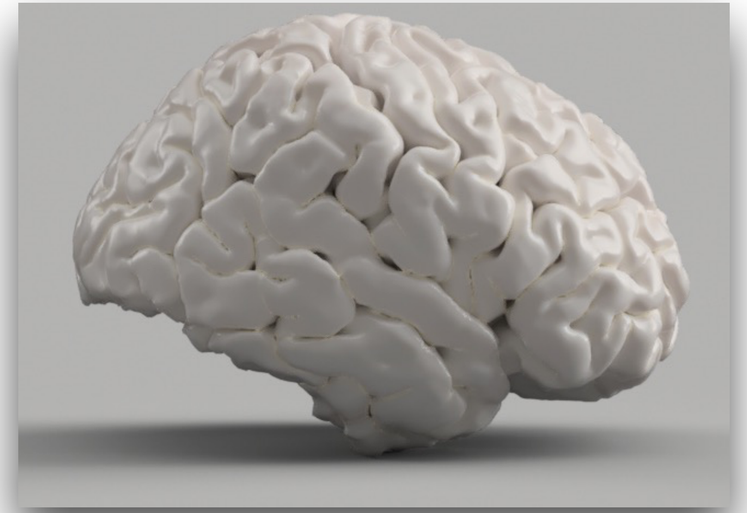
# COGS109: Lecture 11



Big picture revisited of M.D.A., Regression II, Splines, error analysis

July 25, 2023

***Modeling and Data Analysis***

Summer Session 1, 2023

C. Alex Simpkins Jr., Ph.D.

RDPRobotics LLC | Dept. of CogSci, UCSD

# Plan for today

- Announcements
- Review of last time
- Modeling and data analysis - pathways from data to models
- Regression - linear to nonlinear
- Error analysis, goodness of fit
- Splines

Given a scientific question and hypothesis, what is a typical modeling and data analysis path?

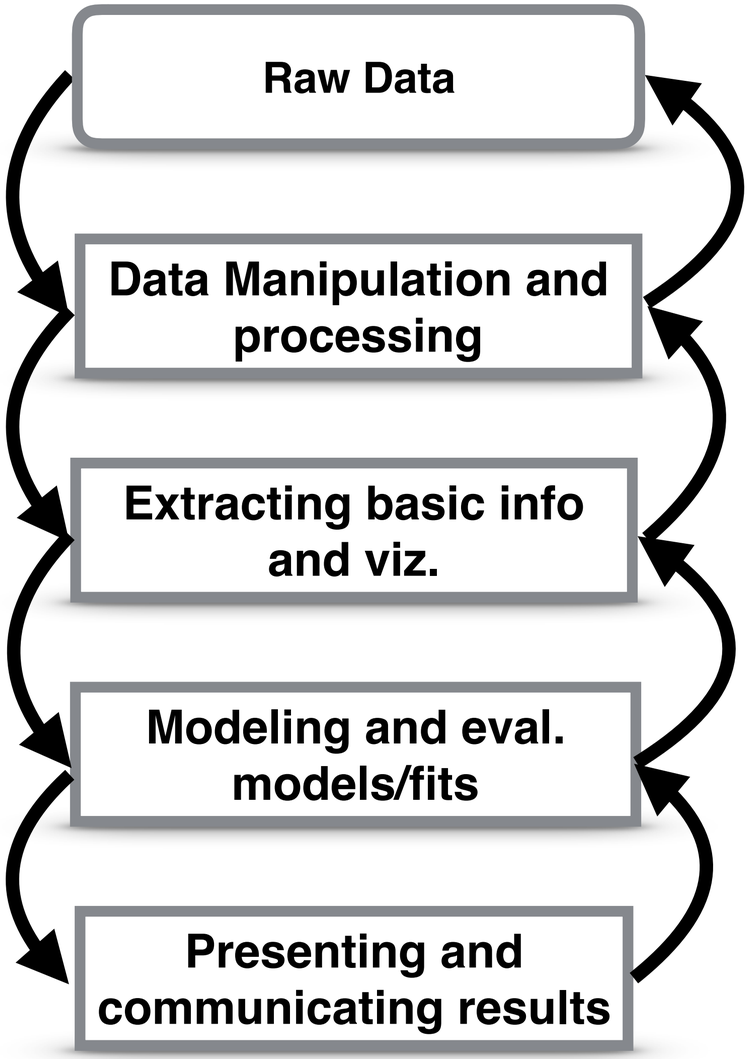
# Aside: Chicken or egg, which is first?

- Sometimes you start with a question
- Sometimes you start with the data, and it helps you generate the questions
- It is not always one way or the other

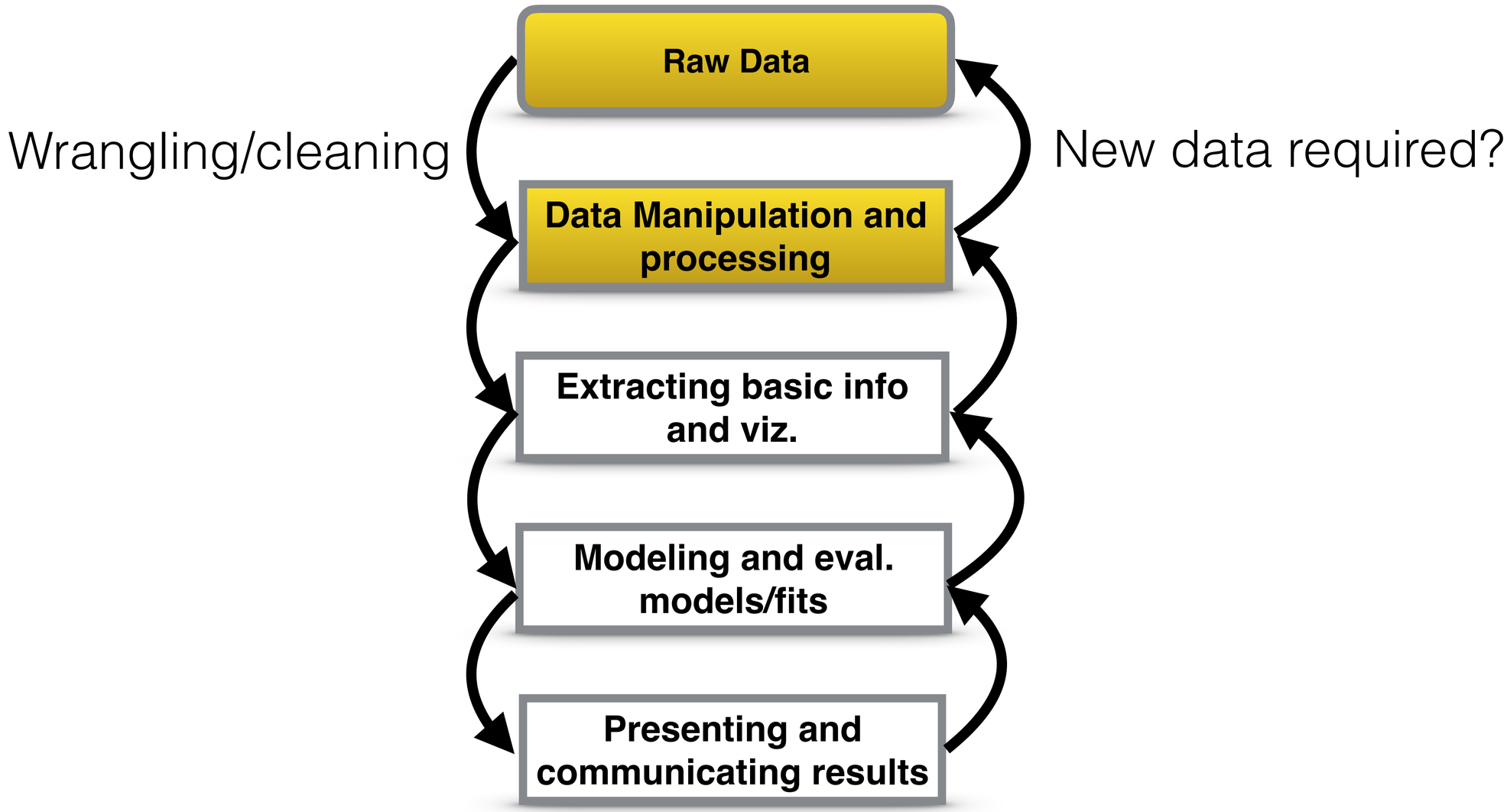
# 4 main parts to modeling and data analysis

| ***Steps*** | ***M.A.D. Topics*** |
|---|---|
| Step 1 | Data manipulation and processing |
| Step 2 | Extracting basic information from data and visualizing that info |
| Step 3 | Modeling the data and evaluating models, data fits |
| Step 4 | Presenting and communicating results |

# Modeling and Data Analysis pathway



# Modeling and Data Analysis pathway

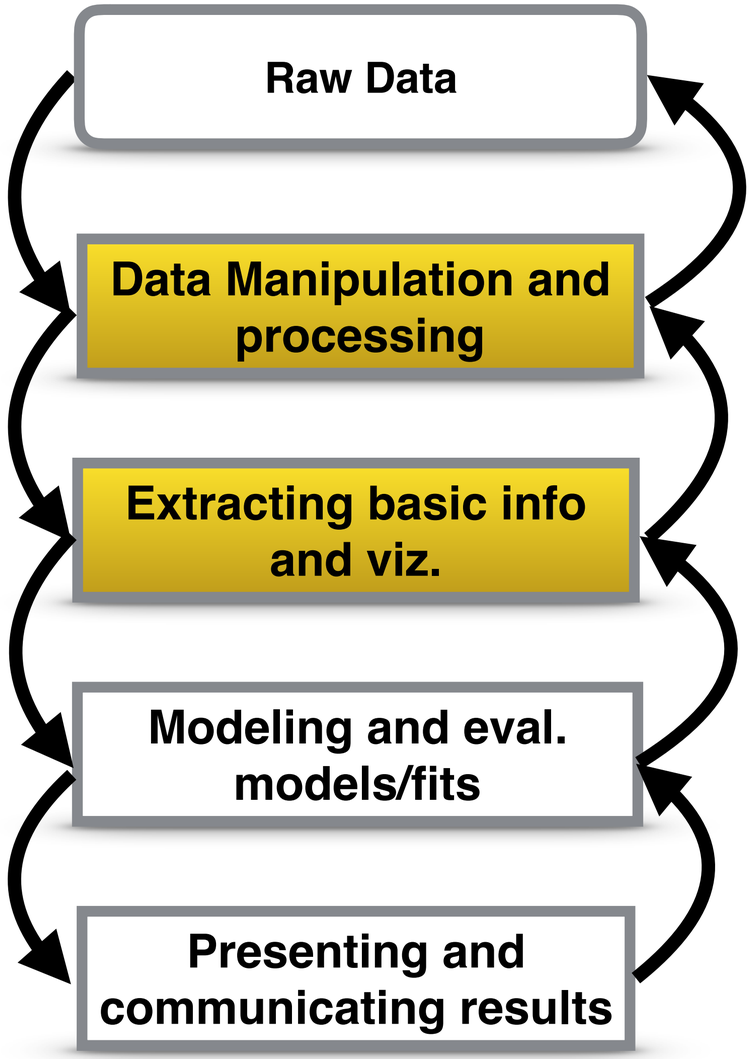


# Data manipulation and processing

- We must import the data into our computational package (python, matlab, C code, R, SPSS, etc)
- We have discussed the challenges and strategies of wrangling and cleaning
- Raw data needs to be transformed into a computationally useful structure (rectangular)
- Raw data need to be cleaned and filtered - removing NaNs, missing data, inconsistencies
- At this stage we sometimes discover issues that require new data, or have insights that change the M.D.A. path

# Modeling and Data Analysis pathway

Filtering, basic statistics (descriptive), Basic plots and tables



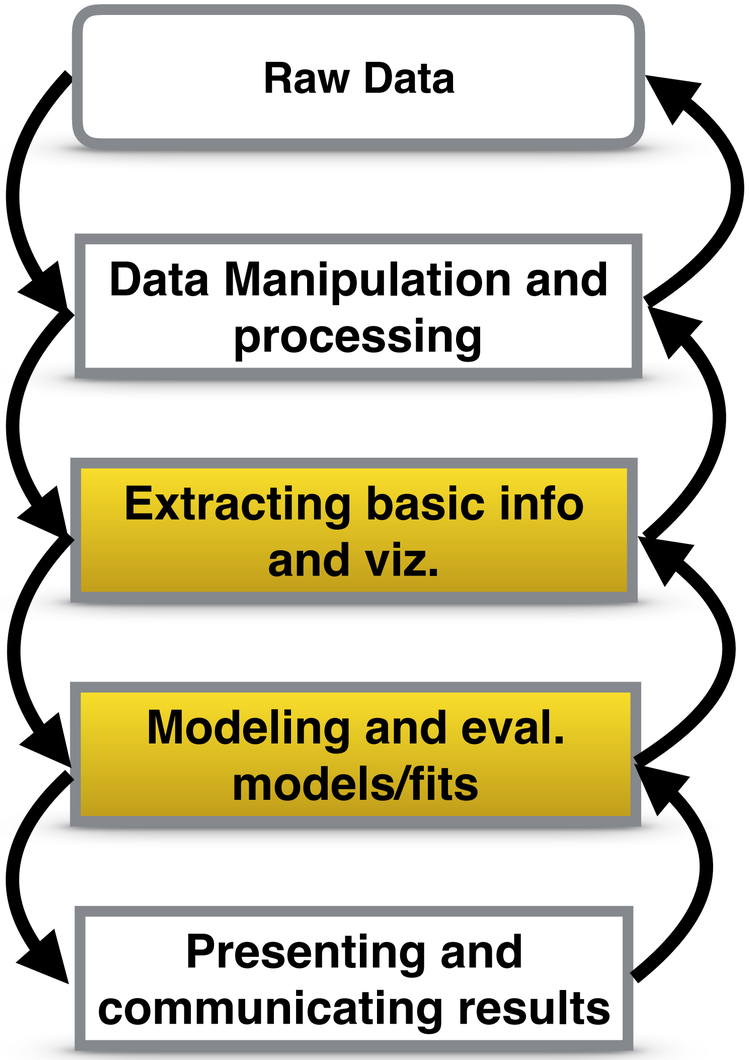

New data required?
Restructure data?
Manipulate differently?

# Extracting basic info and visualizing: Descriptive analysis

- Understand the data set's characteristics
- Describe what they are
- Explain that description to others so they can understand the data
- Filtering can be here as well
- Classical statistics without inferential conclusions (central tendency and variability), shape of distribution
- Basic visualizations (charts, plots, tables) to visualize the data

# Modeling and Data Analysis pathway

From EDA to modeling: Regression, curve fits, parametrization, ML, error analysis

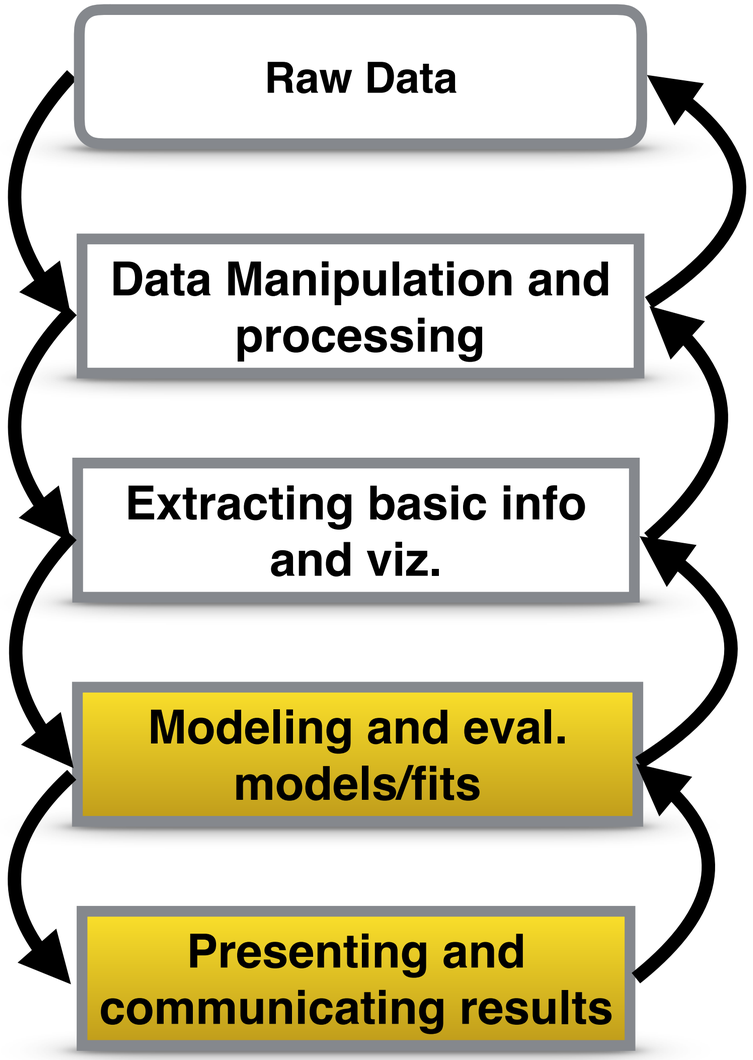



Change perspective? Another aspect to explore? Poor fit? Another candidate?

# Modeling and evaluating model fits

- EDA gives insight into which model set to select - you don't know during EDA what final model you will be choosing
  - Shines the light on the road
- Regression, linear and nonlinear, multivariate regression, curve fits, error analysis (interpolation, approximation)
- Black box, grey box, and white box modeling (approximation, extrapolation)
  - Parameterizing model structures and fitting them (linear/nonlinear, ANN, dynamic/static, kinematic, feedback/feedforward, decision tree
  - ML is one way to fit them
  - Error analysis and assessing goodness of fit (error criteria - RMSE, simple error, etc)

# Modeling and Data Analysis pathway



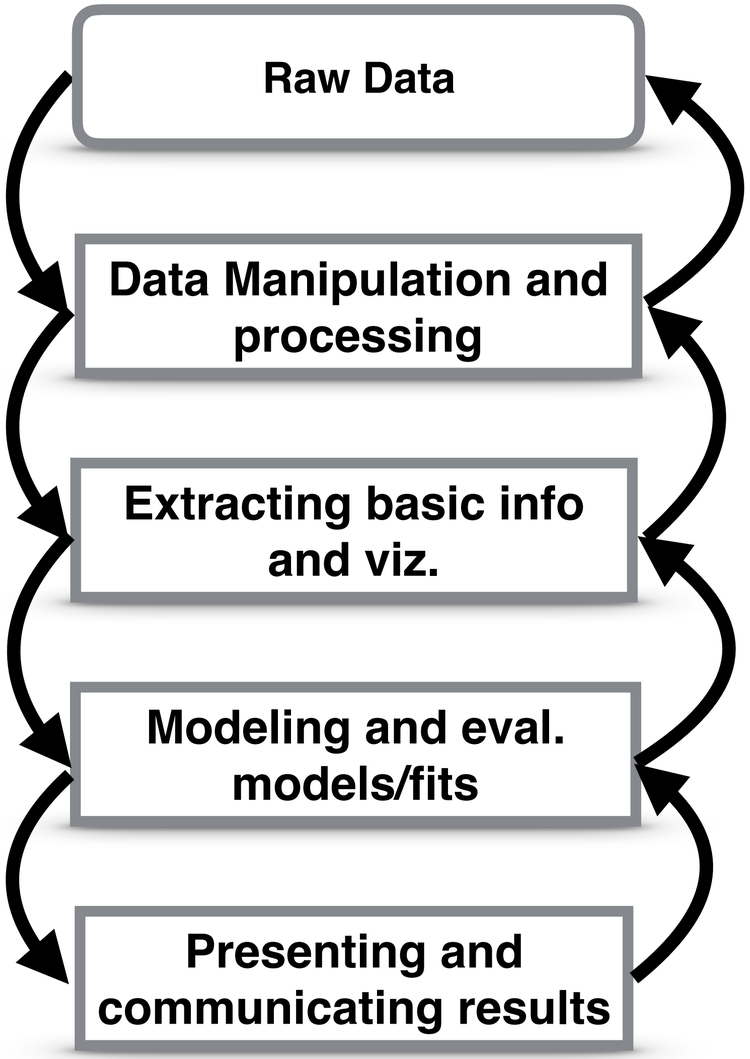

Use model to make inferences and predictions. Come to conclusion about question. Hypothesis reject null? Share with the world : Papers, software, products, talks, courses, etc

Communicating may bring insights. Try another model? Refine? Didn't turn out, is there a way to learn more?

# Presenting and communicating results

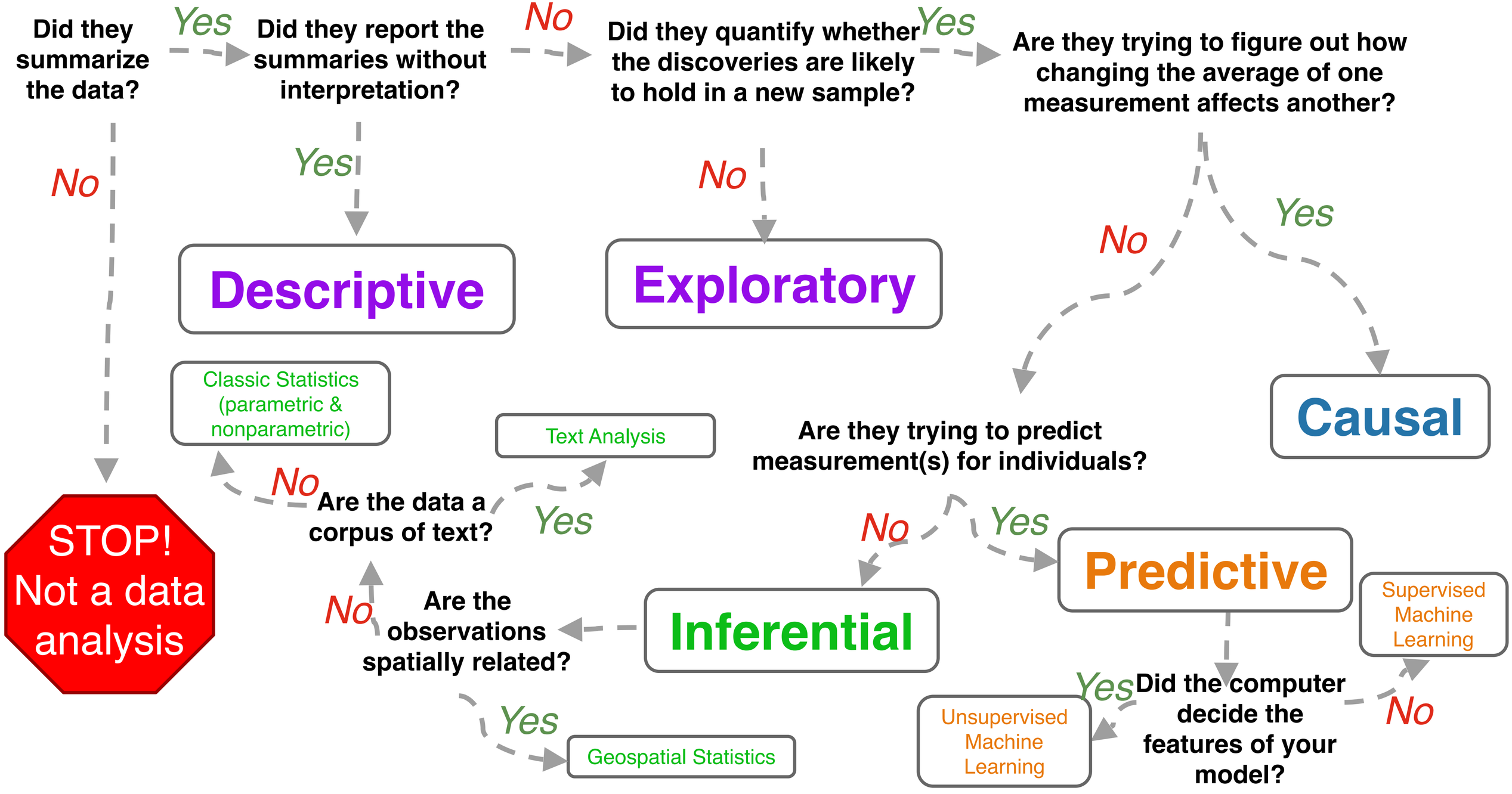
- You summarize what you did
- You interpret the results of your study -
  - *Do you reject or fail to reject the null?*
  - *What is the significance?*
  - *What are the implications?*
  - *What are the limitations of the data, modeling and analysis?*
  - *Where would you go from here (future work)?*
- Share with the world as papers, software, products, talks, seminars, workshops, courses, etc
- Do you go back up the chain and redo, get new data, perform more modeling, etc? Or do you have an idea for another study?

# Modeling and Data Analysis pathway



# Pathways into data analysis

- There is no one path
- The set of possibilities is complex
- One potentially useful flow diagram follows to orient you - what analysis should I be doing?
- Blends the stages of M.D.A. but useful for deciding which analysis to do


Did they summarize the data?
Yes
Did they report the summaries without interpretation?
No
Did they quantify whether the discoveries are likely to hold in a new sample?
Yes
Are they trying to figure out how changing the average of one measurement affects another?
No
Yes
Descriptive
No
Exploratory
Causal
Classic Statistics (parametric & nonparametric)
No
Are the data a corpus of text?
Yes
Text Analysis
Are they trying to predict measurement(s) for individuals?
No
Yes
STOP! Not a data analysis
No
Are the observations spatially related?
Yes
Geospatial Statistics
Inferential
Predictive
Supervised Machine Learning
Yes
Unsupervised Machine Learning
Did the computer decide the features of your model?
No

# Summary: Analytical Approaches

1. Descriptive (and Exploratory) Data Analysis are the first step(s) of analysis
2. Inference establishes relationships
    a. Classic Statistics
    b. Geospatial Analysis
    c. Text Analysis
3. Machine Learning and modeling is for prediction
    a. Supervised
    b. Unsupervised
4. Experiments best way to establish the likelihood of causality
    a. Remember you ***cannot*** establish causality with computational methods only correlations along with statistical beliefs
    b. More you are establishing if they are NOT related or ‘NOT NOT’ related

# Back to regression

# Approximating a curve

- Recall that in regression we want to approximate data with a curve
- The curve can be linear (affine), or nonlinear (polynomial or other shapes)
- Let's go back over regression quickly from the linear perspective then expand to parametrically linear fits of nonlinear functions
- We looked at the linear algebra approach, let's back up to the partial differential equation method (less abbreviated, but might be more clear for those less familiar with linear algebra)

# Linear regression revisited

- Let us begin by considering a set of pairs of x and y data points
- The concept of least squares is to fit a linear or nonlinear curve which fits that data the best according to some criterion.
- One such common criterion is the minimization of sum of the squared differences between the actual data and the predicted data

# Error criterion

- Thus the error is given by

$$E = \sum_{i=1}^{M} [y_i - y_{LS}(x_i)]^2$$

- Where i = 1,2,...M is the number of data points, and $y_{LS}$ is the approximating curve's predicted y at the point $x_i$.

- There are several reasons why we square the error and then find the minimum of that function.

  - One of the most notable is to create a function with an attractive bowl shape and a single global minimum with respect to each line fitting parameter

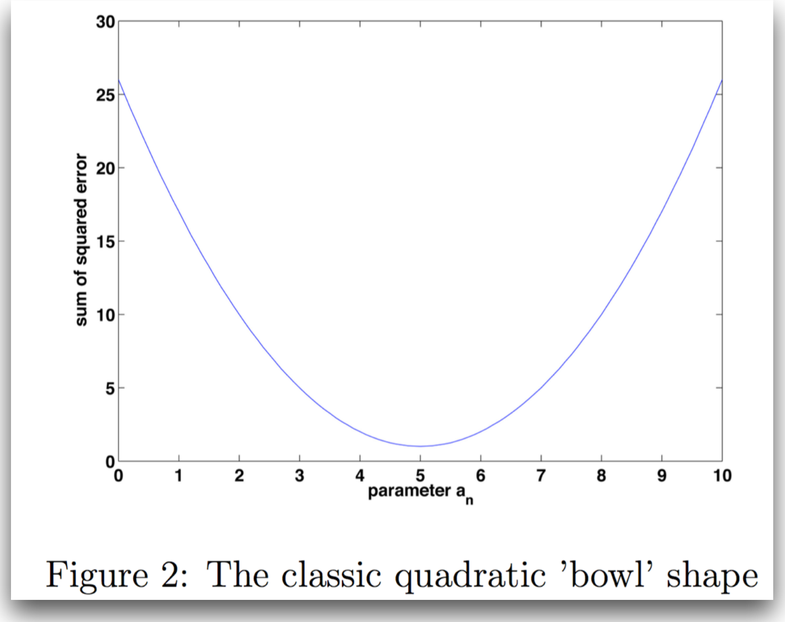



Figure 2: The classic quadratic 'bowl' shape

# Error criterion

- This serves to make positive and negative errors all positive
- i.e. if y is bigger than yLS at point i, the un-squared error is positive, and if yLS is bigger than y at point i, the un-squared error would be negative- but if we square the error, negative numbers become positive, so we measure the magnitude of the error
- A larger negative error is still a bigger error!
- Thus by looking at the sum of the squared error we know when we are reducing error (making our fit better) or increasing error (making our fit worse) as we vary our parameters
- We can define a measure of how to find the ’best fit.’

# Minimum of a function

- The minimum of a function occurs when the derivative (slope) of that function is zero and the second derivative is positive
    - Think of it as the bottom of a valley versus the top of a hill - we want a valley shape
- We have more than one parameter we can vary, so we have to consider the partial derivative with respect to each parameter.

If we want to fit a line, we have only two parameters, $a_0$ and $a_1$ to fit the line defined by

$$y_{LS} = a_0 + a_1 x \tag{2}$$

Thus our error equation becomes, after substituting $y_{LS}$ in,

$$E = \sum_{i=1}^{M} [y_i - (a_0 + a_1 x_i)]^2 \tag{3}$$

To find the parameters $a_0$ and $a_1$ which minimize the error E, we take the partial derivative with respect to each parameter, $a_0$ and $a_1$ and set each resulting equation equal to zero.

$$\frac{\partial E}{\partial a_0} = 0 \tag{4}$$

$$\frac{\partial E}{\partial a_1} = 0 \tag{5}$$

$$\frac{\partial E}{\partial a_0} = \frac{\partial}{\partial a_0} \left\{ \sum_{i=1}^{M} [y_i - (a_0 + a_1 x_i)]^2 \right\} \tag{6}$$

$$\frac{\partial E}{\partial a_1} = \frac{\partial}{\partial a_1}\left\{ \sum_{i=1}^{M} \left[y_i - (a_0 + a_1 x_i)\right]^2 \right\} \tag{7}$$

Which gives us

$$\frac{\partial E}{\partial a_0} = 2\sum_{i=1}^{M} \left[y_i - (a_0 + a_1 x_i)\right](-1) \tag{8}$$

$$\frac{\partial E}{\partial a_1} = 2\sum_{i=1}^{M} \left[y_i - (a_0 + a_1 x_i)\right](-x_i) \tag{9}$$

Which can be simplified by the following steps (note we are working with pairs of equations)

$$\sum_{i=1}^{M} y_i = \sum_{i=1}^{M} a_0 + \sum_{i=1}^{M} a_1 x_i \tag{10}$$

$$\sum_{i=1}^{M} y_i x_i = \sum_{i=1}^{M} a_0 x_i + \sum_{i=1}^{M} a_1 x_i^2 \tag{11}$$

and

and

$$\sum_{i=1}^{M} y_i = a_0 M + a_1 \sum_{i=1}^{M} x_i \tag{12}$$

$$\sum_{i=1}^{M} y_i x_i = a_0 \sum_{i=1}^{M} x_i + a_1 \sum_{i=1}^{M} x_i^2 \tag{13}$$

Which we can solve by realizing that we have a familiar matrix algebra problem:

$$a_0 M + a_1 \sum_{i=1}^{M} x_i = \sum_{i=1}^{M} y_i \tag{14}$$

$$a_0 \sum_{i=1}^{M} x_i + a_1 \sum_{i=1}^{M} x_i^2 = \sum_{i=1}^{M} y_i x_i \tag{15}$$

$$\begin{bmatrix} M & \sum_{i=1}^{M} x_i \\ \sum_{i=1}^{M} x_i & \sum_{i=1}^{M} x_i^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{M} y_i \\ \sum_{i=1}^{M} y_i x_i \end{Bmatrix} \tag{16}$$

Which, if we set $A = \begin{bmatrix} M & \sum_{i=1}^{M} x_i \\ \sum_{i=1}^{M} x_i & \sum_{i=1}^{M} x_i^2 \end{bmatrix}$, $x = \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix}$, and $b = \begin{Bmatrix} \sum_{i=1}^{M} y_i \\ \sum_{i=1}^{M} y_i x_i \end{Bmatrix}$ is equivalent to

$$Ax = b \tag{17}$$

# Solving for the parameters

- Note that the A, x, and b are not the same variables as the x's and a's in the above equations.
- We can solve this problem by finding the inverse of A ($A^{-1}$), and multiplying both sides by $A^{-1}$, as we did last time
- This gives us the parameters of $a_0$ and $a_1$ $$x = A^{-1}b$$

# Computing the inverse

- Gaussian elimination (or another algorithm)
- Cramer's rule (by hand)

# Recall from last time this approach…

- We have 2 (could be 1000) points, and want to fit a line to them
- (1,2) , (3,4)
- How would you solve this problem?
- We want y=mx+b (we need **m** and **b)**
  - **Substitute each point in**

$$2 = m(1) + b$$
$$4 = m(3) + b$$

# Example continued

- And solve for **b** first, then **m**

$$b = 2 - m$$

$$4 = 3m + 2 - m$$

$$4 = 2m + 2$$

$$m = 1$$

$$b = 2 - m$$

$$b = 1$$

# Example continued

- We have two equations and two unknowns (m, b)
- This can be written compactly as

$$\begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix} \begin{Bmatrix} m \\ b \end{Bmatrix} = \begin{Bmatrix} 2 \\ 4 \end{Bmatrix}$$

- Which is of the basic form

$$Ax = b$$

- We want to find

$$x = A^{-1}b$$

# Python implementation

- https://numpy.org/doc/stable/reference/generated/numpy.linalg.lstsq.html
- In discussion we are going to take this further, along with more tools that are even easier and are powerful
- We want you to understand the steps and what is happening

# OK but what if we have more than 2 points?

- Let's check in python using the functions we discussed last time
- To the notebook!
  - http://localhost:8888/notebooks/Documents/teaching/cogs109/lectures/week4/Untitled.ipynb?kernel_name=python3

# Nonlinear regression

# But what about nonlinear systems?

- Not all processes are linear (in fact most real processes are not linear).
- It is therefore more appropriate in such cases to have nonlinear data fitting methods.

# Polynomial fits

One basic nonlinear function is a polynomial. Consider that we have a set of data $\{x_i, y_i\}$ with $i = 0, 1, ...M$ data points, and we would like to fit a polynomial of order n. The general equation for a polynomial is given by

$$P_n(x) = a_0 x^0 + a_1 x^1 + a_2 x^2 + ... + a_n x^n \tag{22}$$

or more compactly written as

$$P_n(x) = \sum_{k=0}^{n} a_k x^k \tag{23}$$

# Same path

- Let us find once again the least squares approximation of our data using this polynomial fit rather than a linear fit.

- It should be noted that a linear fit is merely a polynomial of degree one (recall that the degree or order of a polynomial is the number of the highest power to which a variable in the polynomial equation is raised) .

- To compute the least square polynomial fit we will use the same approach as the linear case.

- It is important to choose n < M (do not try to fit a polynomial with a higher order than the number of points of data).

# We begin with

$$E = \sum_{i=1}^{M} [y_i - P(x_i)]^2 \tag{24}$$

$$E = \sum_{i=1}^{M} y_i^2 - 2\sum_{i=1}^{M} y_i P(x_i) + \sum_{i=1}^{M} P^2(x_i) \tag{25}$$

$$E = \sum_{i=1}^{M} y_i^2 - 2\sum_{i=1}^{M} y_i \left[\sum_{k=0}^{n} a_k x_i^k\right] + \sum_{i=1}^{M} \left[\sum_{k=0}^{n} a_k x_i^k\right]^2 \tag{26}$$

Which we can simplify to

$$E = \sum_{i=1}^{M} y_i^2 - 2\sum_{k=0}^{n} a_k \sum_{i=1}^{M} y_i x_i^k + \sum_{h=0}^{n}\sum_{k=0}^{n} a_h a_k \left[\sum_{i=1}^{M} x_i^{h+k}\right] \tag{27}$$

Now similar to before, to find the minimum of the error, E, we must find where the partial derivative with respect to each variable is zero:

$$\frac{\partial E}{\partial a_k} = 0 \tag{28}$$

for each $k = 0, 1, 2, ...n$

$$0 = \frac{\partial E}{\partial a_k} = -2\sum_{i=1}^{M} y_i x_i^k + 2\sum_{h=0}^{n} a_h \sum_{i=1}^{M} x_i^{h+k} \tag{29}$$

We now have n+1 equations in n+1 unknown parameters $a_k$. These are referred to as the normal equations:

$$\sum_{h=0}^{n} a_h \sum_{i=1}^{M} x_i^{h+k} = \sum_{i=1}^{M} y_i x_i^k \tag{30}$$

$k = 0, 1, 2, ...n$

Written in this form we may not initially recognize this equation as being simple to solve. Let us consider this equation in a less short shorthand notation, then take a step back and look for a pattern.

$$a_0 \sum_{i=1}^{M} x_i^0 + a_1 \sum_{i=1}^{M} x_i^1 + a_2 \sum_{i=1}^{M} x_i^2 + ... + a_n \sum_{i=1}^{M} x_i^n = \sum_{i=1}^{M} y_i x_i^0 \tag{31}$$

$$a_0 \sum_{i=1}^{M} x_i^1 + a_1 \sum_{i=1}^{M} x_i^2 + a_2 \sum_{i=1}^{M} x_i^3 + ... + a_n \sum_{i=1}^{M} x_i^{n+1} = \sum_{i=1}^{M} y_i x_i^1 \tag{32}$$

.

.

$$a_0 \sum_{i=1}^{M} x_i^n + a_1 \sum_{i=1}^{M} x_i^{n+1} + a_2 \sum_{i=1}^{M} x_i^{n+2} + ... + a_n \sum_{i=1}^{M} x_i^{2n} = \sum_{i=1}^{M} y_i x_i^n \tag{33}$$

If we consider this in a matrix algebra perspective, we see that we can put this in the $Ax = b$ canonical form:

let us define

$$x = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \tag{34}$$

and

$$b = \begin{pmatrix} \sum_{i=1}^{M} y_i \\ \sum_{i=1}^{M} y_i x_i \\ \sum_{i=1}^{M} y_i x_i^2 \\ \vdots \\ \sum_{i=1}^{M} y_i x_i^n \end{pmatrix} \tag{35}$$

and finally

$$A = \begin{bmatrix} \sum_{i=1}^{M} x_i^0 & \sum_{i=1}^{M} x_i^1 & \sum_{i=1}^{M} x_i^2 & \cdots & \sum_{i=1}^{M} x_i^n \\ \sum_{i=1}^{M} x_i^1 & \sum_{i=1}^{M} x_i^2 & \sum_{i=1}^{M} x_i^3 & \cdots & \sum_{i=1}^{M} x_i^{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{M} x_i^n & \sum_{i=1}^{M} x_i^{n+1} & \sum_{i=1}^{M} x_i^{n+2} & \cdots & \sum_{i=1}^{M} x_i^{2n} \end{bmatrix} \tag{36}$$

Then we can write this:

$$\begin{bmatrix} \sum_{i=1}^{M} x_i^0 & \sum_{i=1}^{M} x_i^1 & \sum_{i=1}^{M} x_i^2 & \cdots & \sum_{i=1}^{M} x_i^n \\ \sum_{i=1}^{M} x_i^1 & \sum_{i=1}^{M} x_i^2 & \sum_{i=1}^{M} x_i^3 & \cdots & \sum_{i=1}^{M} x_i^{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{M} x_i^n & \sum_{i=1}^{M} x_i^{n+1} & \sum_{i=1}^{M} x_i^{n+2} & \cdots & \sum_{i=1}^{M} x_i^{2n} \end{bmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{M} y_i \\ \sum_{i=1}^{M} y_i x_i \\ \sum_{i=1}^{M} y_i x_i^2 \\ \vdots \\ \sum_{i=1}^{M} y_i x_i^n \end{pmatrix} \tag{37}$$

Or more simply...

$$Ax = b \tag{38}$$

x is found by

$$x = A^{-1}b \tag{39}$$

Which can be solved, as before by Gaussian elimination, or another (more efficient) algorithm.

# Solving these equations

- The exact same approach is used in python to solve this!
- Back to our notebook:
  - http://localhost:8888/notebooks/Documents/teaching/cogs109/lectures/week4/Untitled.ipynb?kernel_name=python3

# Why not just use a built-in function only?

- I want you to understand how to do this regardless of the toolset available
- This allows complete flexibility of terms
  - Can also be done with an exponential function and other terms - how would you do this?
- You'll also learn the single line type commands in section with Sagarika today (which are also powerful!)
- Portability of technique (same in matlab other than the actual least squares operation - one line difference)
  - Same setup

# Next let’s consider how good the fit is

- Time for a break