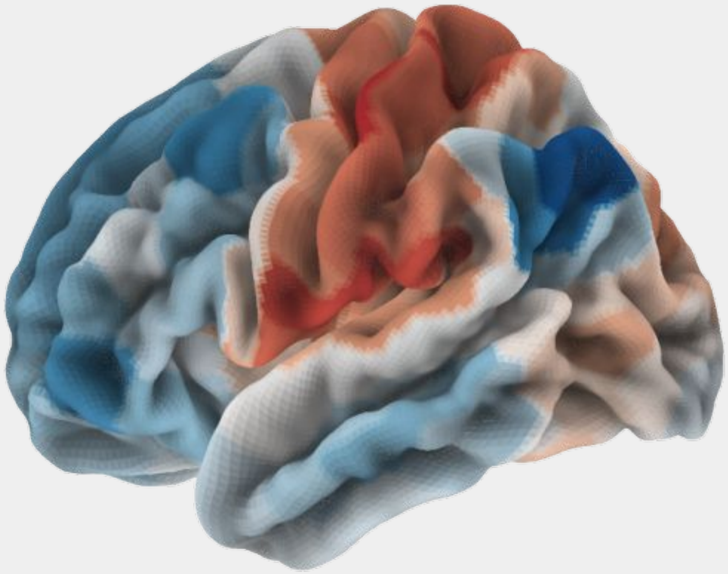
Advanced fMRI data analysis



Karolina Finc

Centre for Modern Interdisciplinary Technologies

Nicolaus Copernicus University in Toruń

COURSE #4: General Linear Model 1 | 24th April 2020 neuromatch

About

Neuromatch Academy 2020

An online school for Computational Neuroscience

Started by the team who created CoSMo summer school, CCN SS, Simons IBRO, and neuromatch conference , we announce a worldwide academy to train neuroscientists to learn computational tools, make connections to real world neuroscience problems, and promote networking with researchers.

Objectives

- Introduce traditional and emerging computational neuroscience tools
- Learn hands-on skills with neuro data
- Understand how these tools relate to biological questions
- Build networking, professional development, and how-to-model skills

https://neuromatch.io/academy/

Study plan

BEFORE

Open science & neuroimaging

fMRI data manipulation in python

fMRI data preprocessing

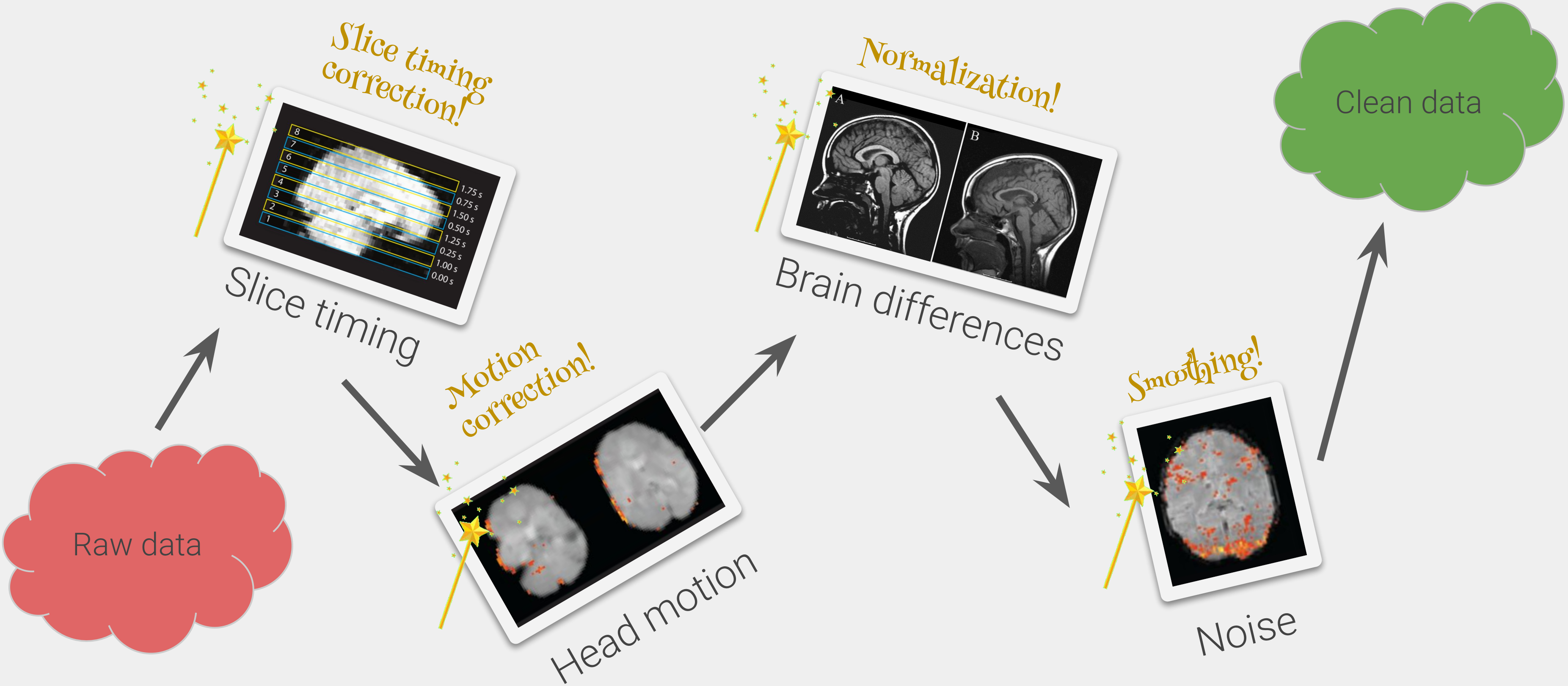
General Linear Model

Functional connectivity

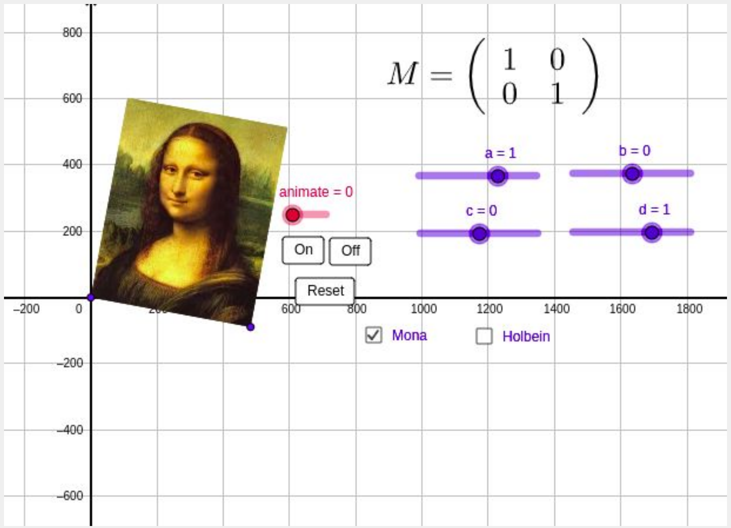
Machine Learning on fMRI data

AFTER

Preprocessing workflow / pipeline



Linear transformations magic!



GeoGebra

https://www.geogebra.org/m/pDU4peV5

Khan Academy

https://www.khanacademy.org/math/linear-algebra/matrix-transformations/linear-transformations/a/visualizing-linear-transformations

https://www.youtube.com/channel/UCYO_jab_esuFRV4b17AJtAw

Software

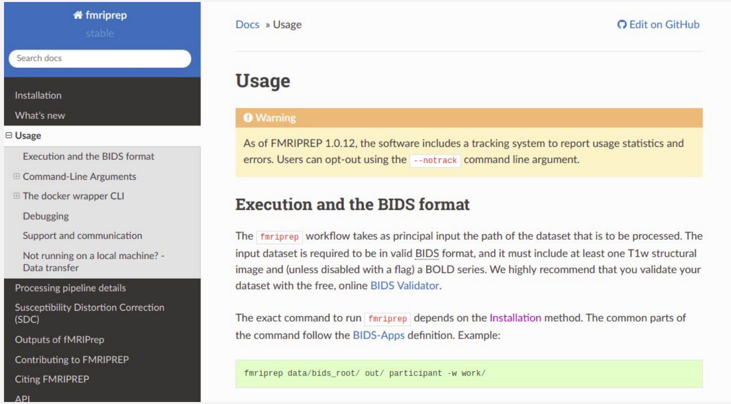
SPM 12

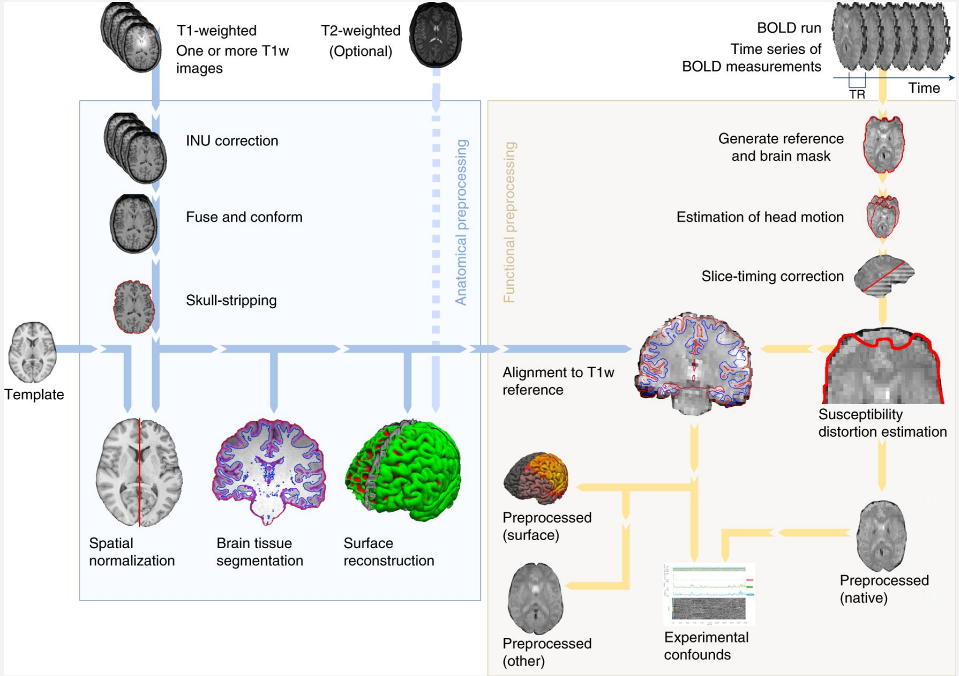
FSL

FreeSurfer

AFNI!

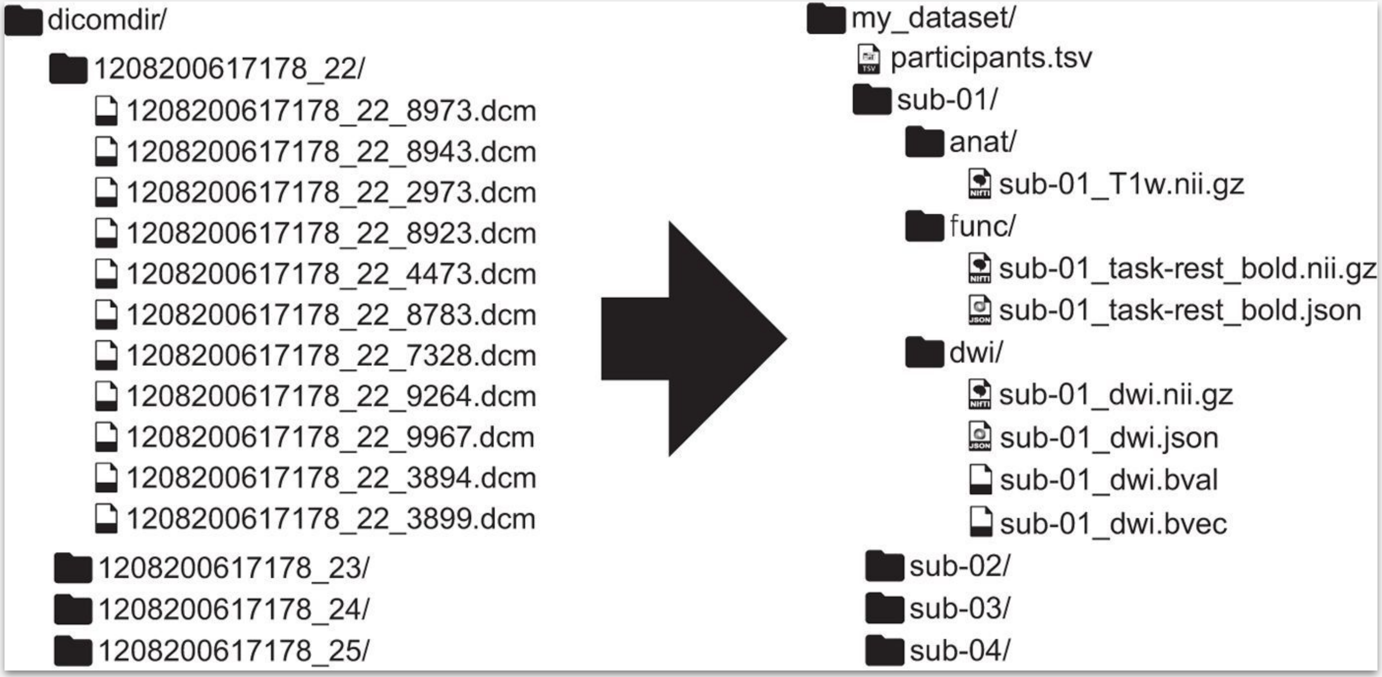
fMRIPrep!





https://fmriprep.readthedocs.io/en/stable/

Brain Imaging Data Structure (BIDS)



Study plan

BEFORE

Open science & neuroimaging

fMRI data manipulation in python

fMRI data preprocessing

General Linear Model

Functional connectivity

Machine Learning on fMRI data

AFTER

Study plan

BEFORE

Open science &
neuroimaging

fMRI data manipulation
in python

fMRI data
preprocessing

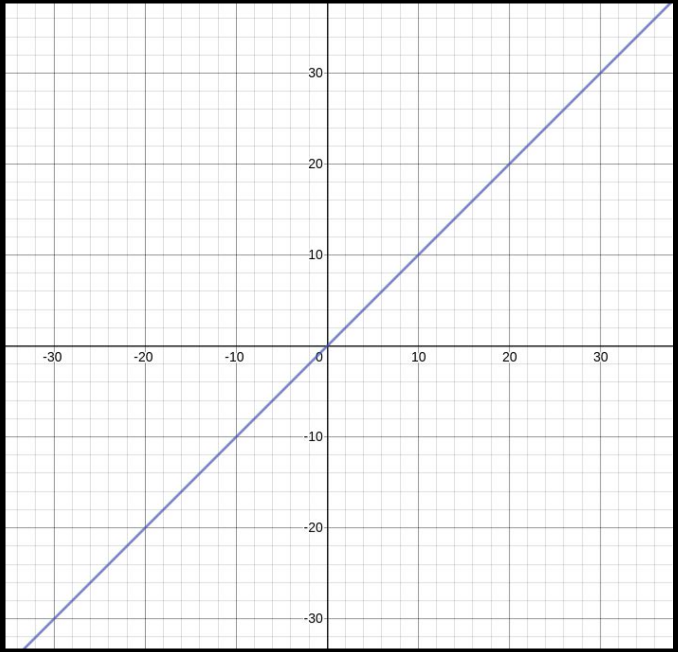
General
Linear Model

Functional
connectivity

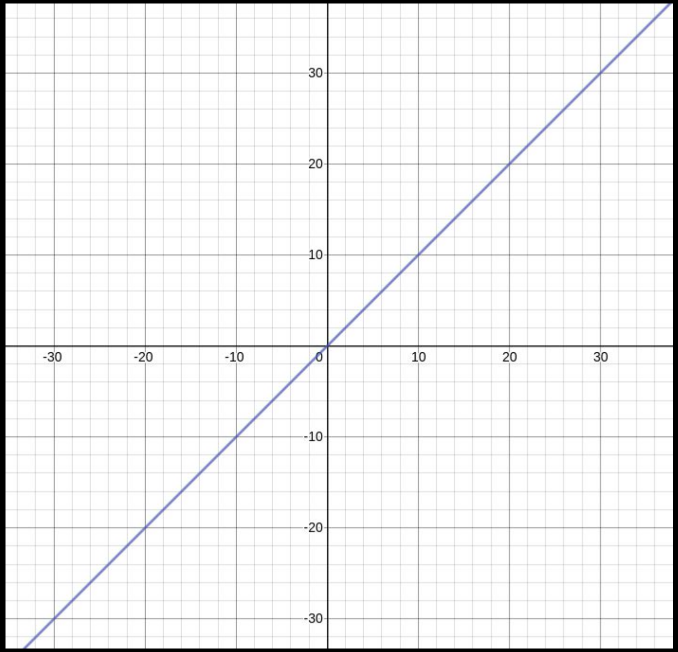
Machine Learning
on fMRI data

AFTER

Guess the function formula!

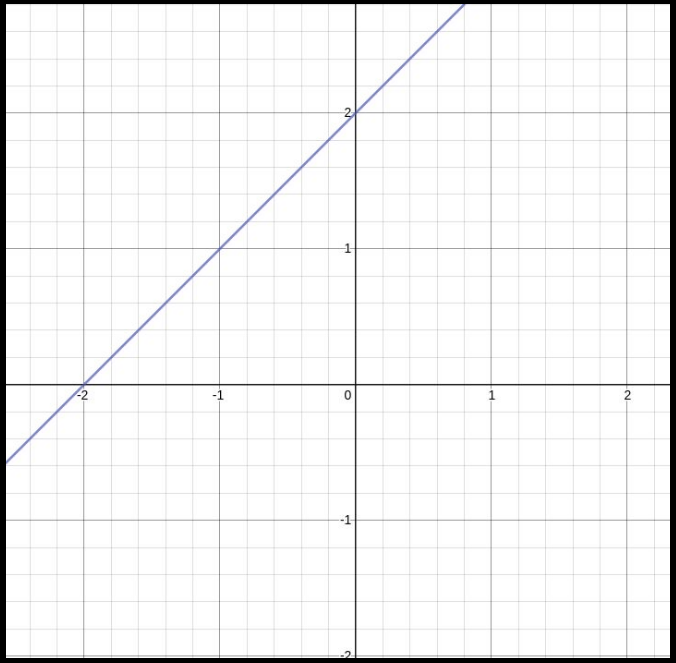


Guess the function formula!

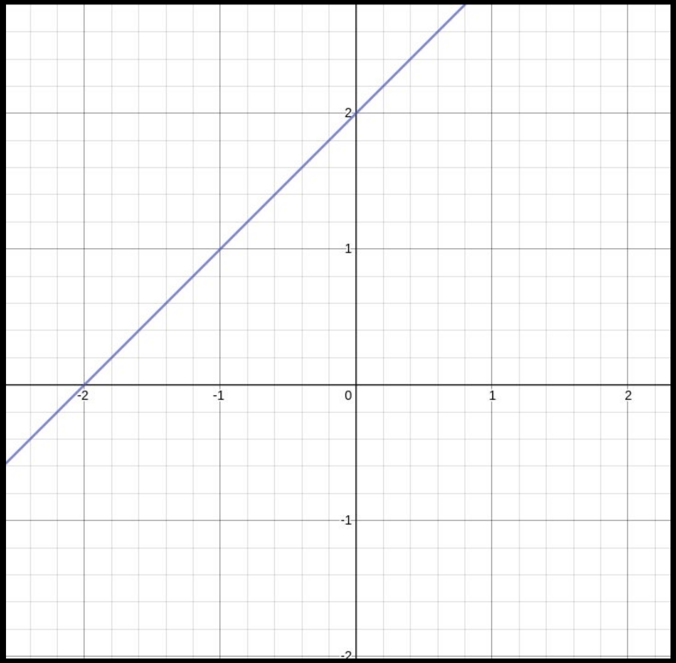


$y = x$

Guess the function formula!

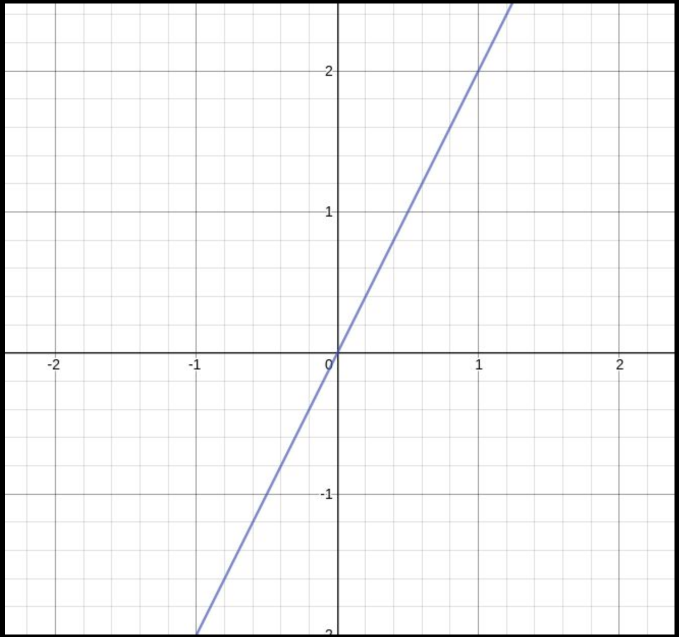


Guess the function formula!

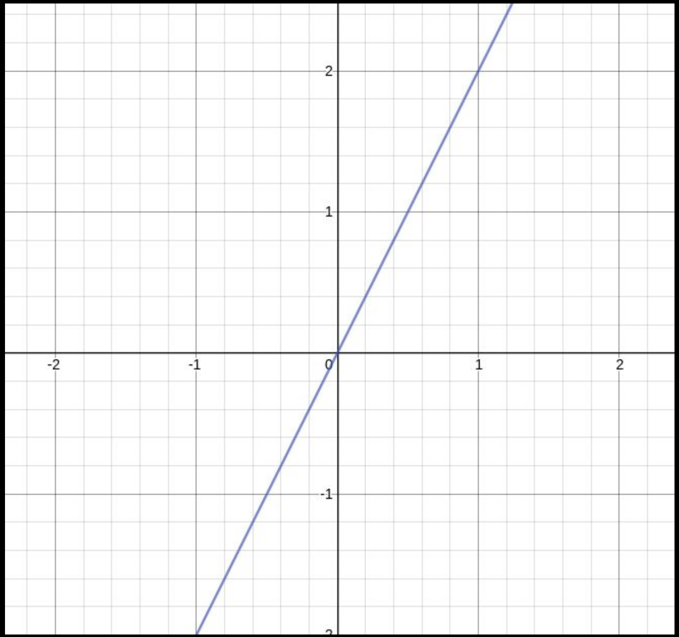


$y = x + 2$

Guess the function formula!

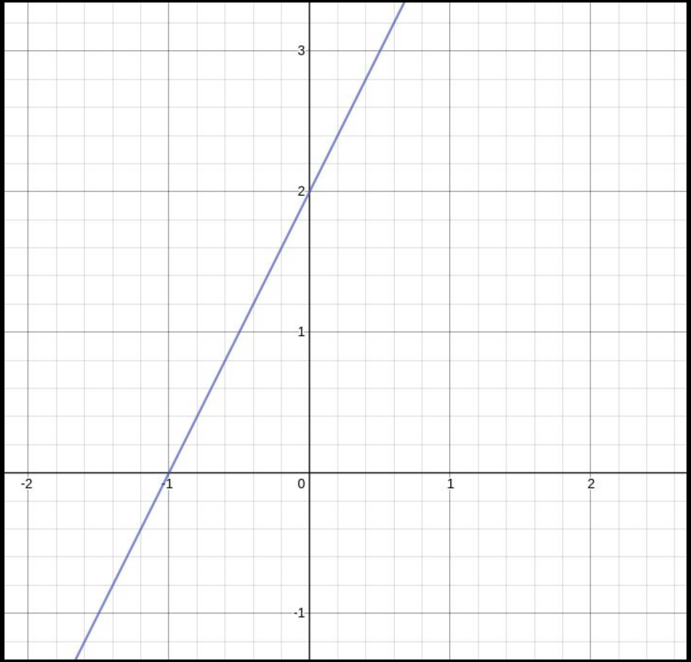


Guess the function formula!

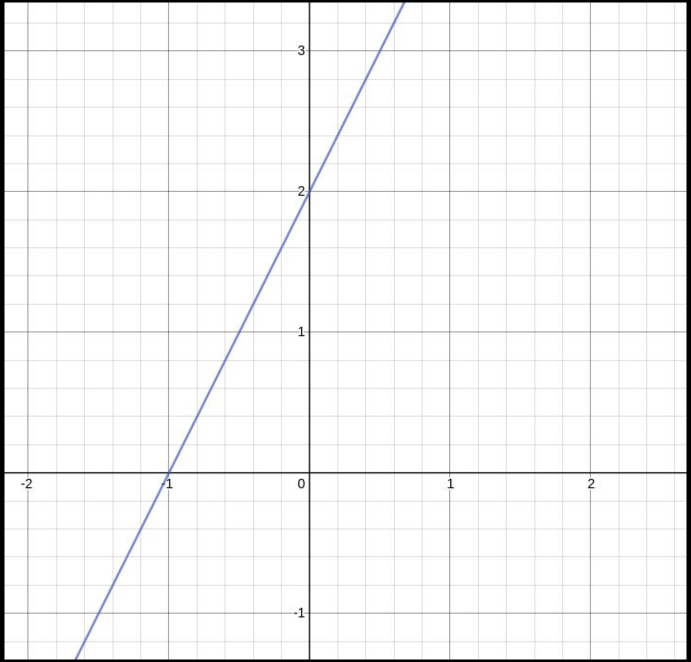


$y = 2x$

Guess the function formula!



Guess the function formula!



$y = 2x + 2$

Slope-intercept form of linear function

Slope-intercept form of linear function

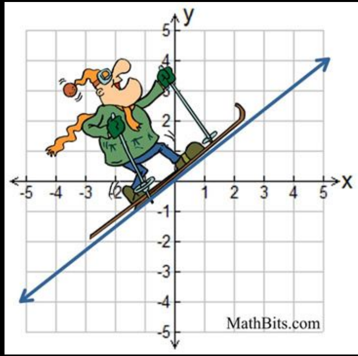
$$y = mx + b$$

Slope-intercept form of linear function

slope

$$y = mx + b$$

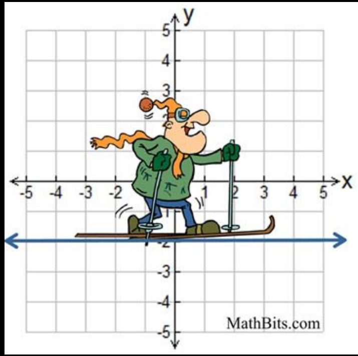
Slope-intercept form of linear function



slope

$$y = mx + b$$

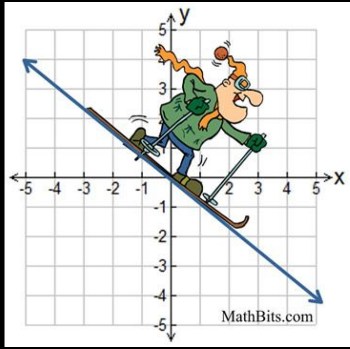
Slope-intercept form of linear function



slope

$$y = mx + b$$

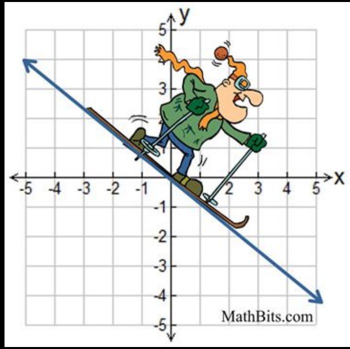
Slope-intercept form of linear function



slope

$$y = mx + b$$

Slope-intercept form of linear function

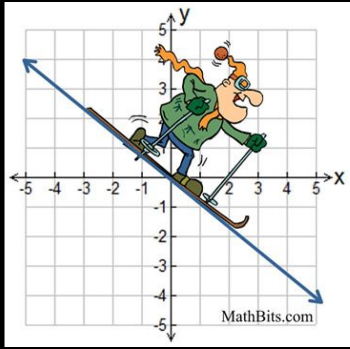


slope

intercept

$$y = mx + b$$

Slope-intercept form of linear function



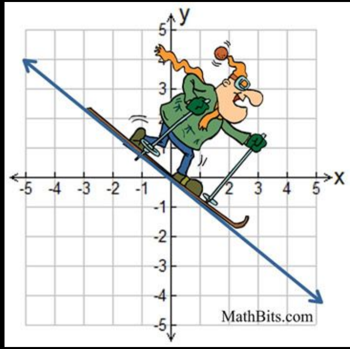
slope

intercept

$$y = mx + b$$

coefficients

Slope-intercept form of linear function



slope

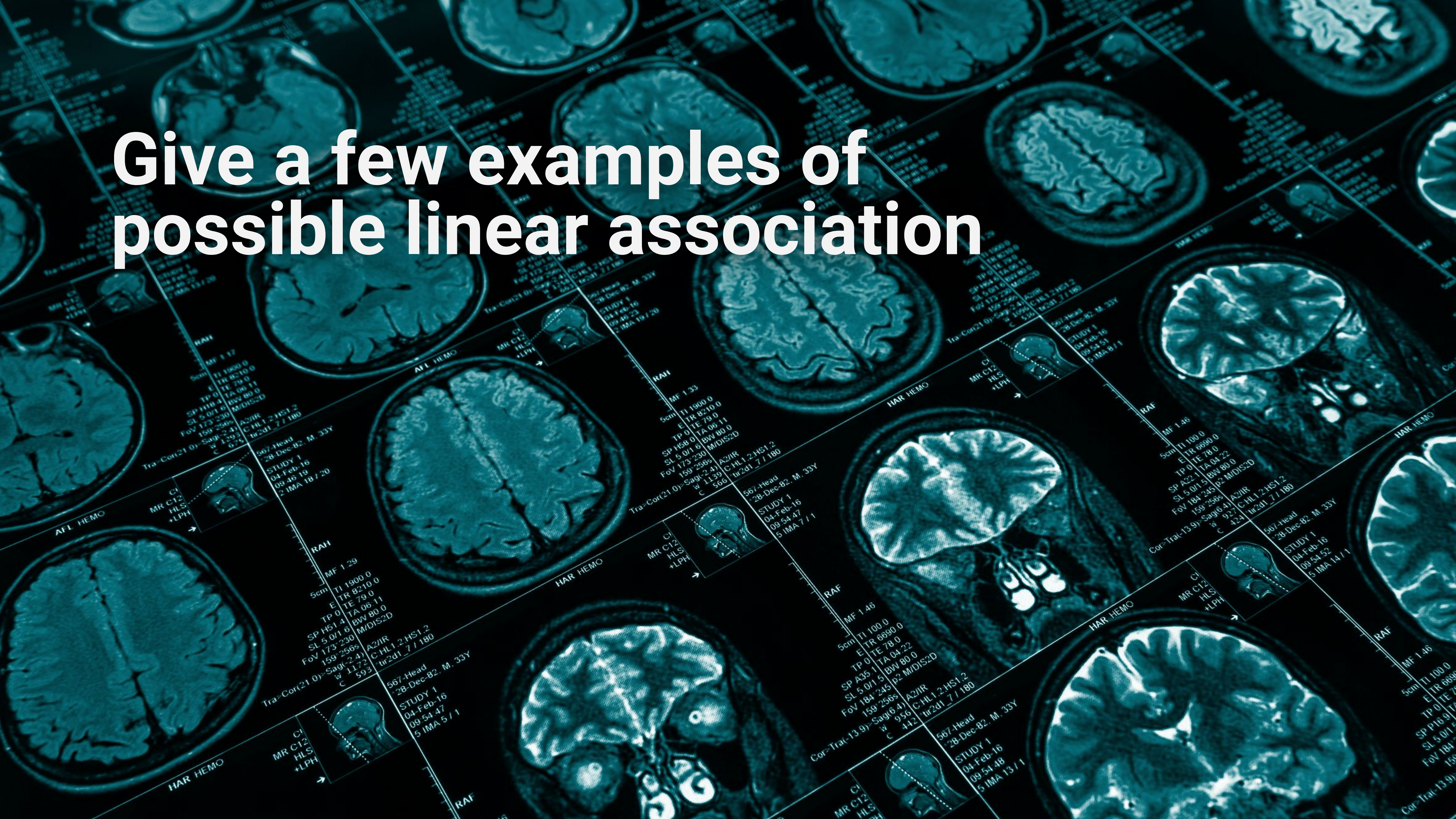
intercept

$$y = mx + b$$

coefficients

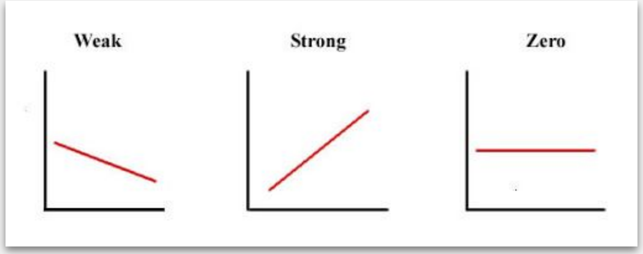
https://www.desmos.com/calculator

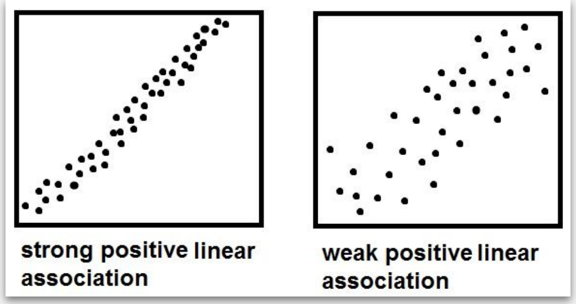
Give a few examples of possible linear association

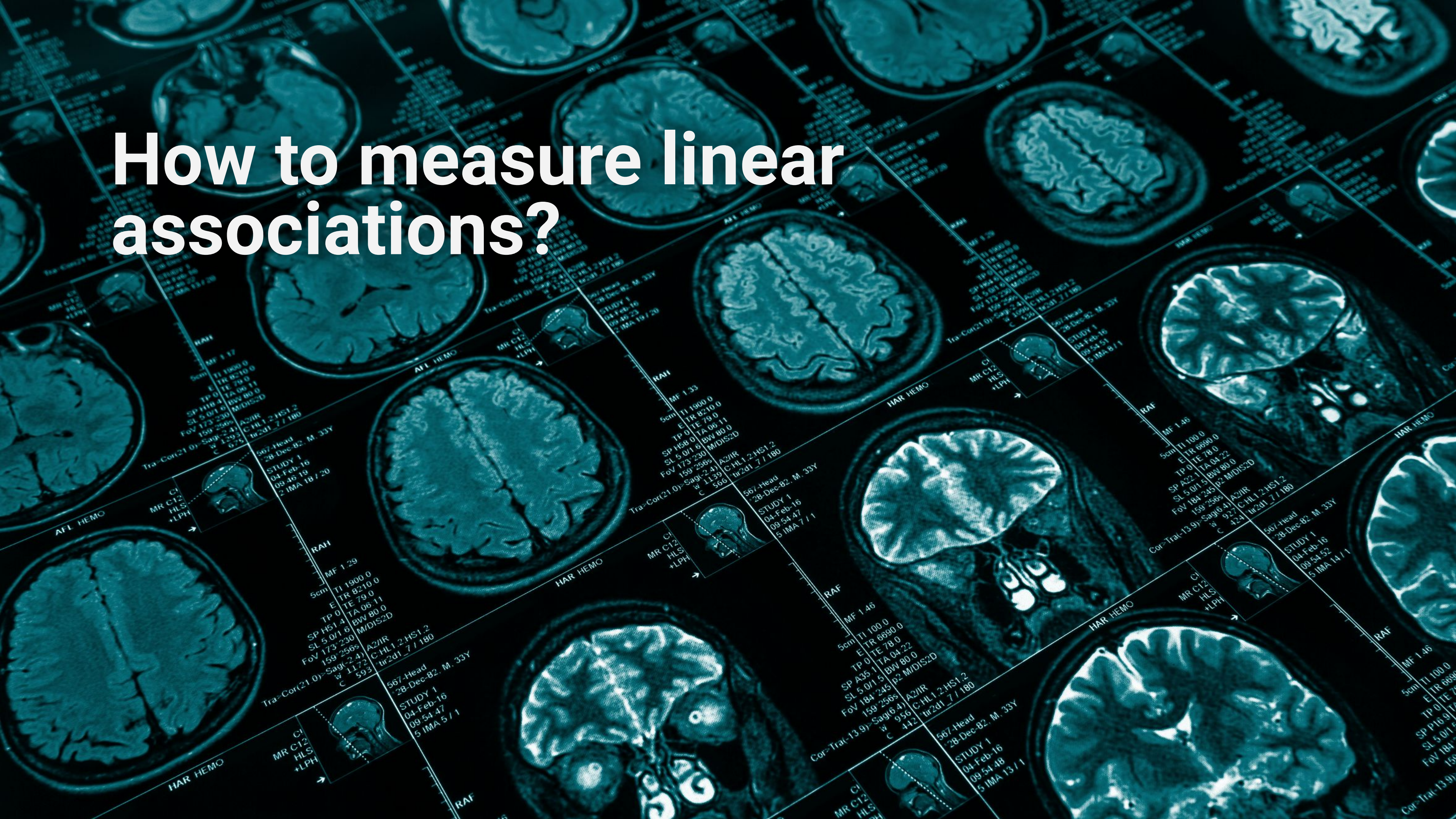


Examples of linear associations

- Positive linear association between height and foot size
- Negative linear association between cortical thickness and age
- Positive linear association between network modularity and cognitive plasticity
- etc.

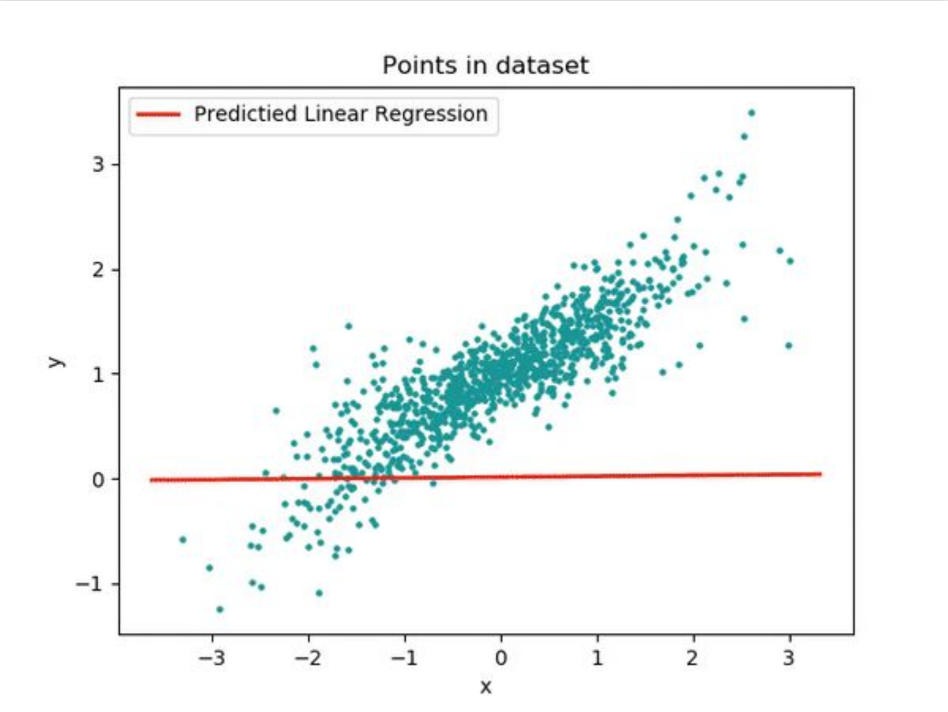




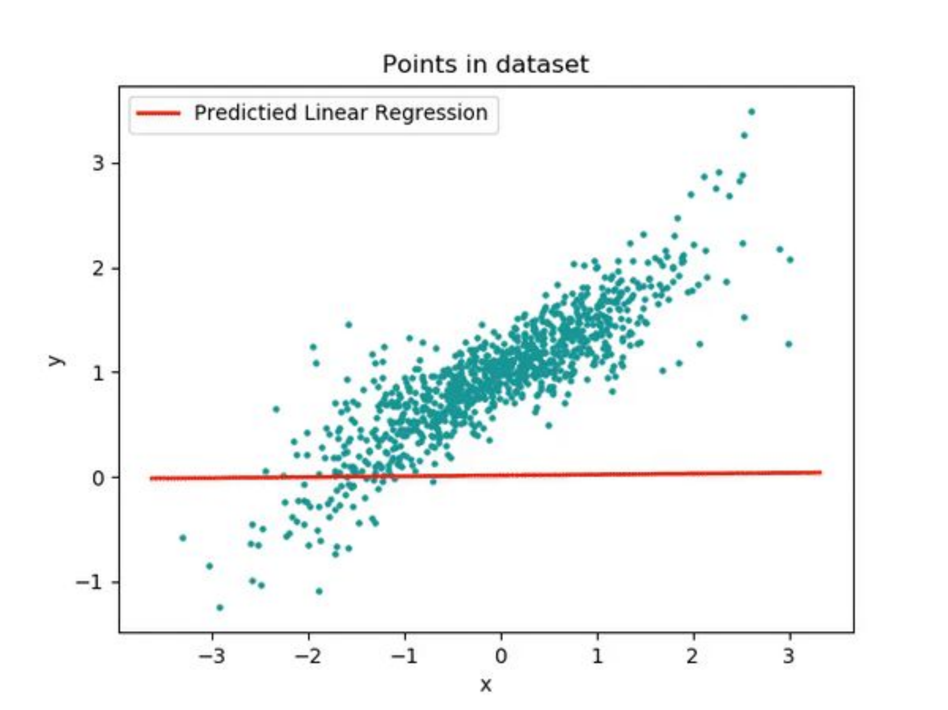


How to measure linear associations?

Linear regression

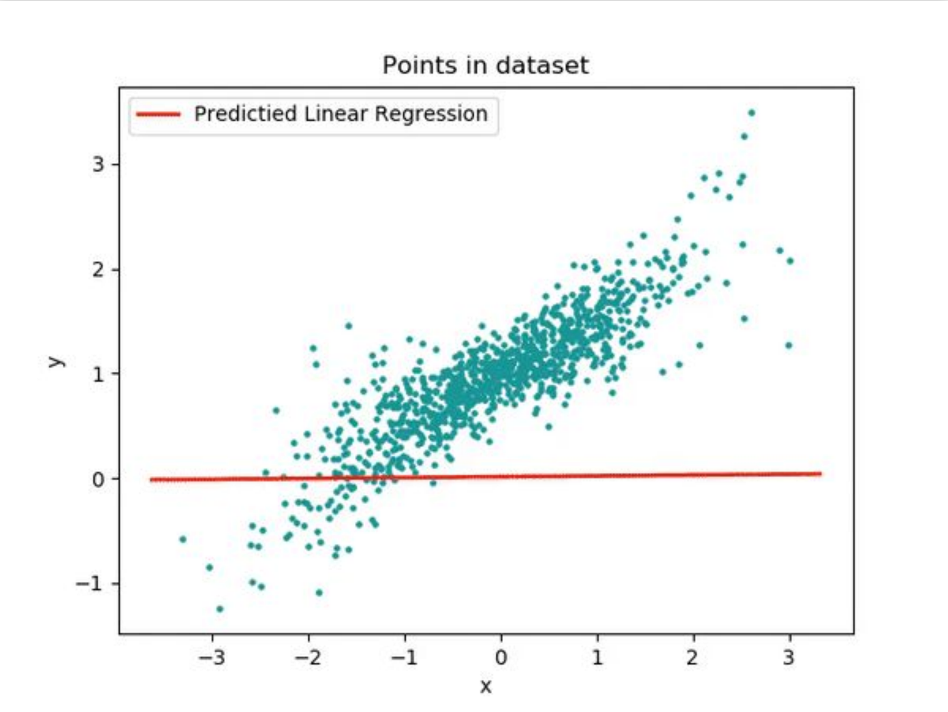


Linear regression



Regression line provides a **model** of the data

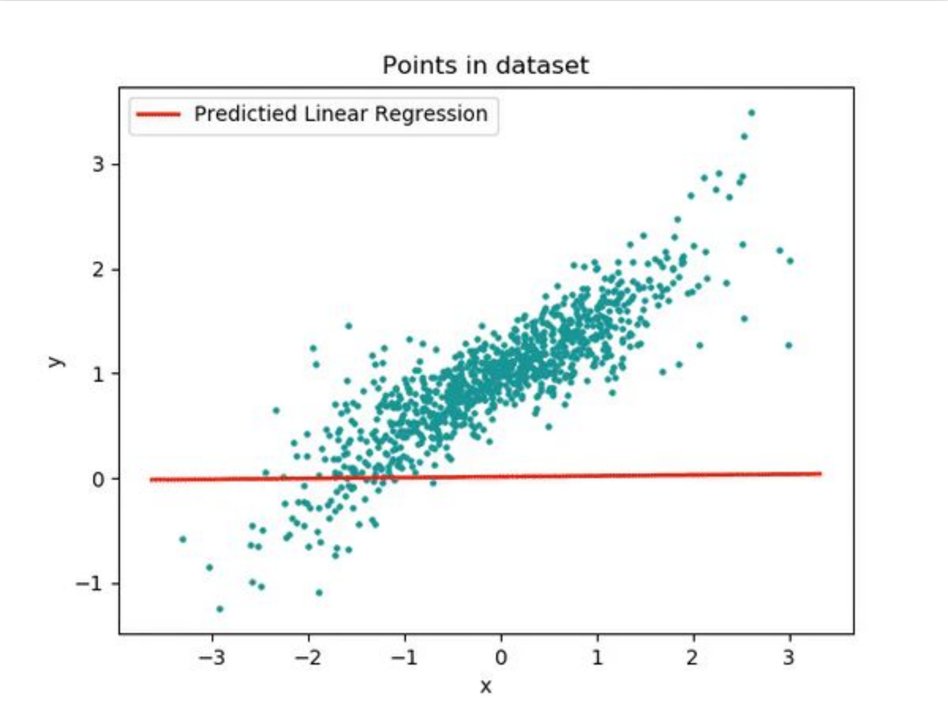
Linear regression



Regression line provides a **model** of the data

Regression problem: predict real-valued output

Linear regression

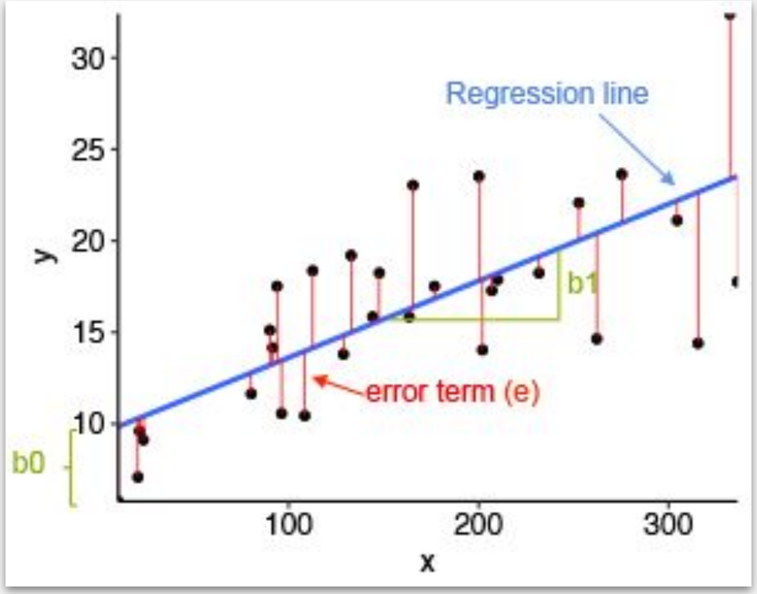


Regression line provides a **model** of the data

Regression problem: predict real-valued output

Regression is an example of **supervised learning** (answers are given)

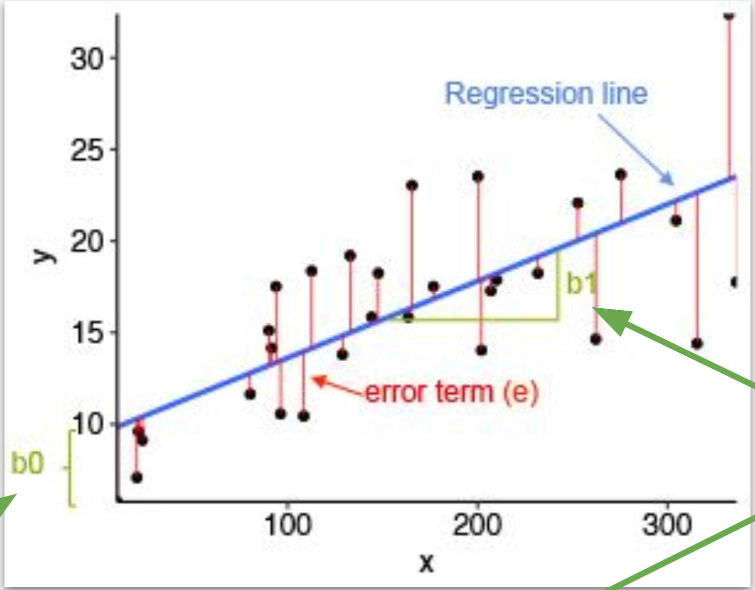
Fitting regression line



Linear Regression: Single Variable

$$\underbrace{\boxed{\hat{y}}}_{\text{Predicted output}} = \underbrace{\beta_0 + \beta_1}_{\text{Coefficients}} \underbrace{\boxed{x}}_{\text{Input}} + \underbrace{\boxed{\epsilon}}_{\text{Error}}$$

Fitting regression line



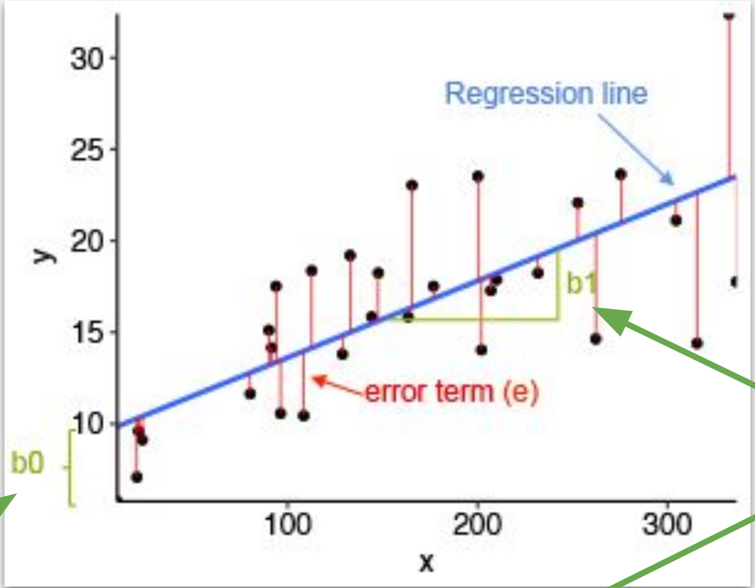
Linear Regression: Single Variable

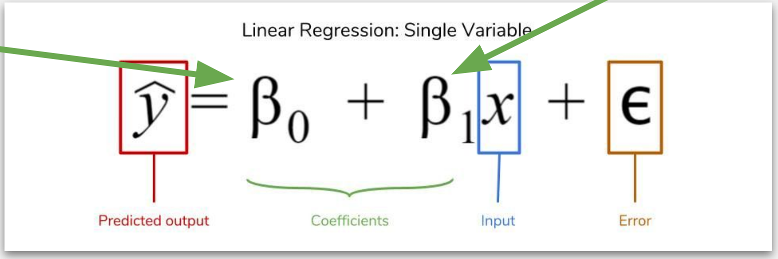
$$\underbrace{\boxed{\hat{y}}}_{\text{Predicted output}} = \underbrace{\beta_0 + \beta_1}_{\text{Coefficients}} \underbrace{\boxed{x}}_{\text{Input}} + \underbrace{\boxed{\epsilon}}_{\text{Error}}$$

intercept

slope

Fitting regression line



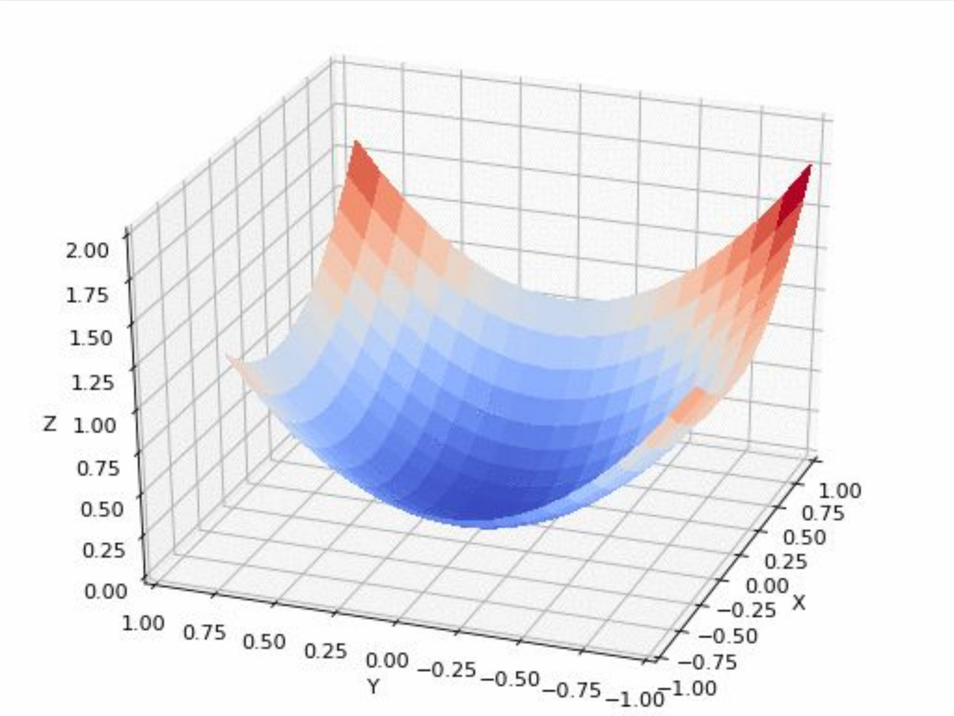


intercept

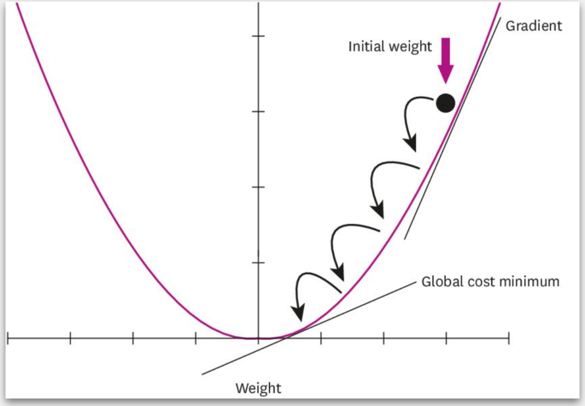
slope

Find such β_0 and β_1 that minimize *cost function*: **sum of squared errors** function

Gradient descent

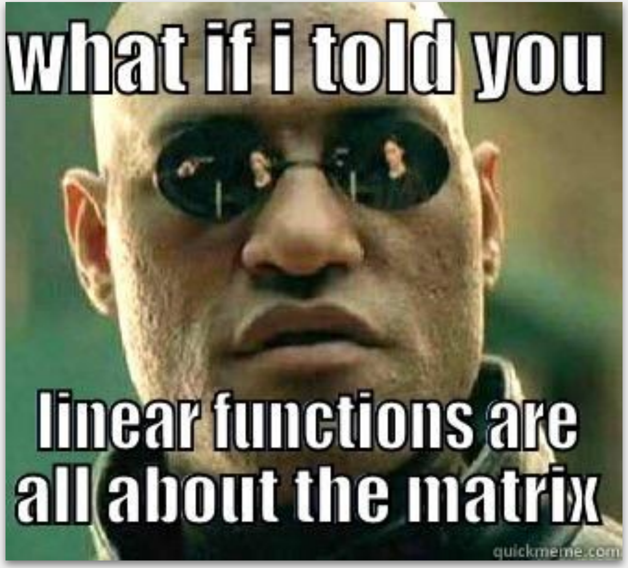


- Algorithm for minimizing cost function
- Is used not only in linear regression

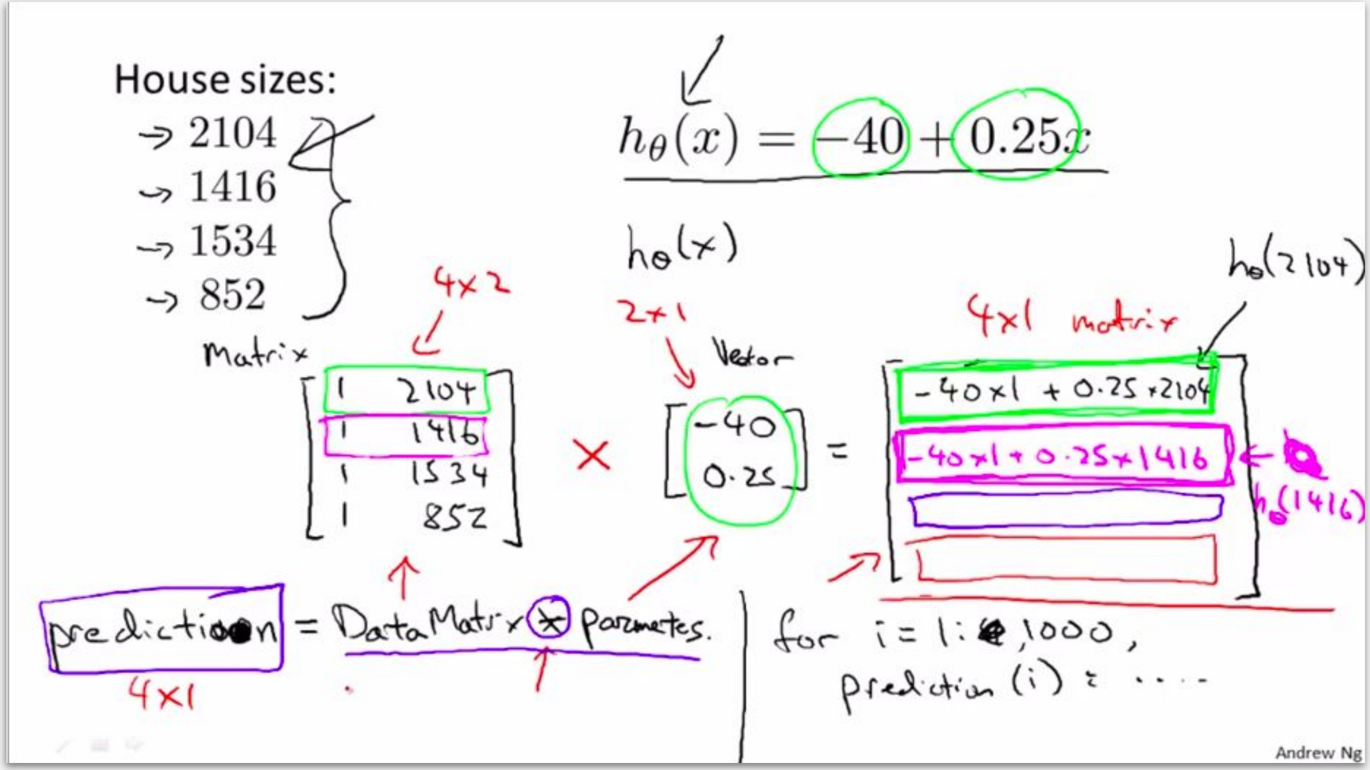


Matrix vector multiplication

$$\begin{bmatrix} A & B \\ C & D \\ E & F \end{bmatrix} \times \begin{bmatrix} G \\ H \end{bmatrix} = \begin{bmatrix} A \times G + B \times H \\ C \times G + D \times H \\ E \times G + F \times H \end{bmatrix}$$



Solving linear equations with linear algebra



Matrix matrix multiplication

House sizes:

$$\begin{cases} 2104 \\ 1416 \\ 1534 \\ 852 \end{cases}$$

Have 3 competing hypotheses:

1. $h_\theta(x) = -40 + 0.25x$
2. $h_\theta(x) = 200 + 0.1x$
3. $h_\theta(x) = -150 + 0.4x$

Matrix $\qquad\qquad$ Matrix

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \times \begin{bmatrix} -40 & 200 & -150 \\ 0.25 & 0.1 & 0.4 \end{bmatrix} = \begin{bmatrix} 486 & 410 & 692 \\ 314 & 342 & 416 \\ 344 & 353 & 464 \\ 173 & 285 & 191 \end{bmatrix}$$

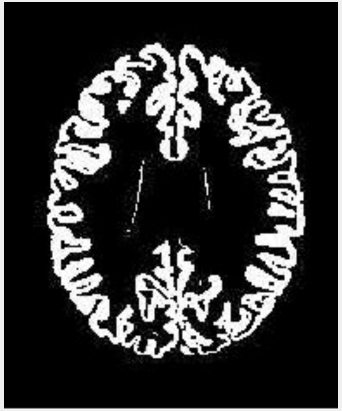
Prediction of first h_θ

Predictions of 2^{nd} h_θ

Andrew Ng

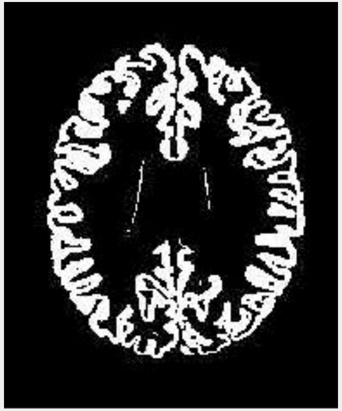
Challenge 1

Gray matter

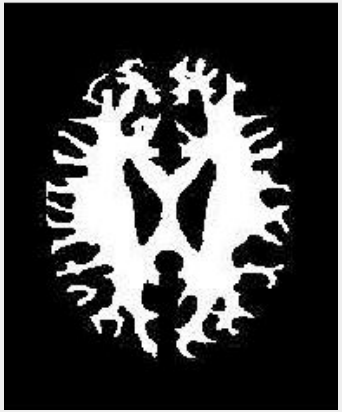


Challenge 1

Gray matter

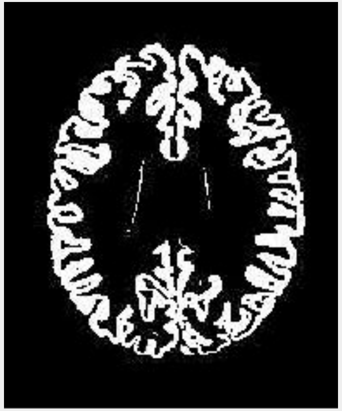


White matter

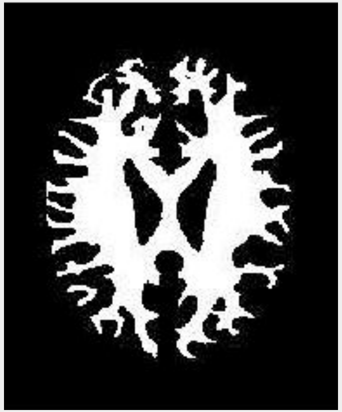


Challenge 1

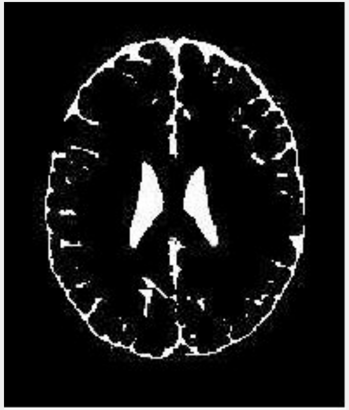
Gray matter



White matter

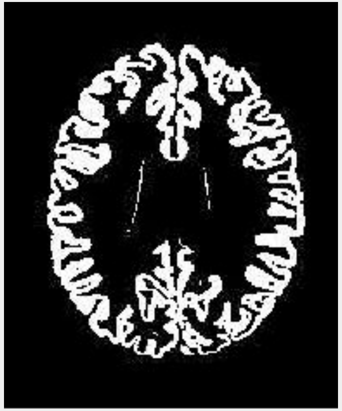


CSF

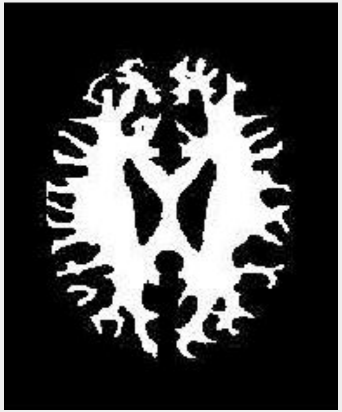


Challenge 1

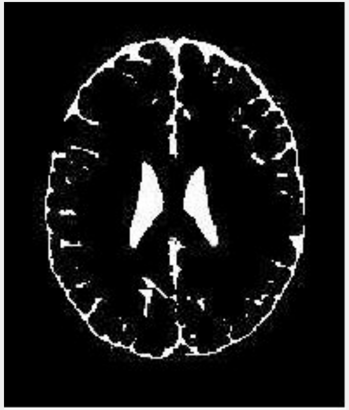
Gray matter



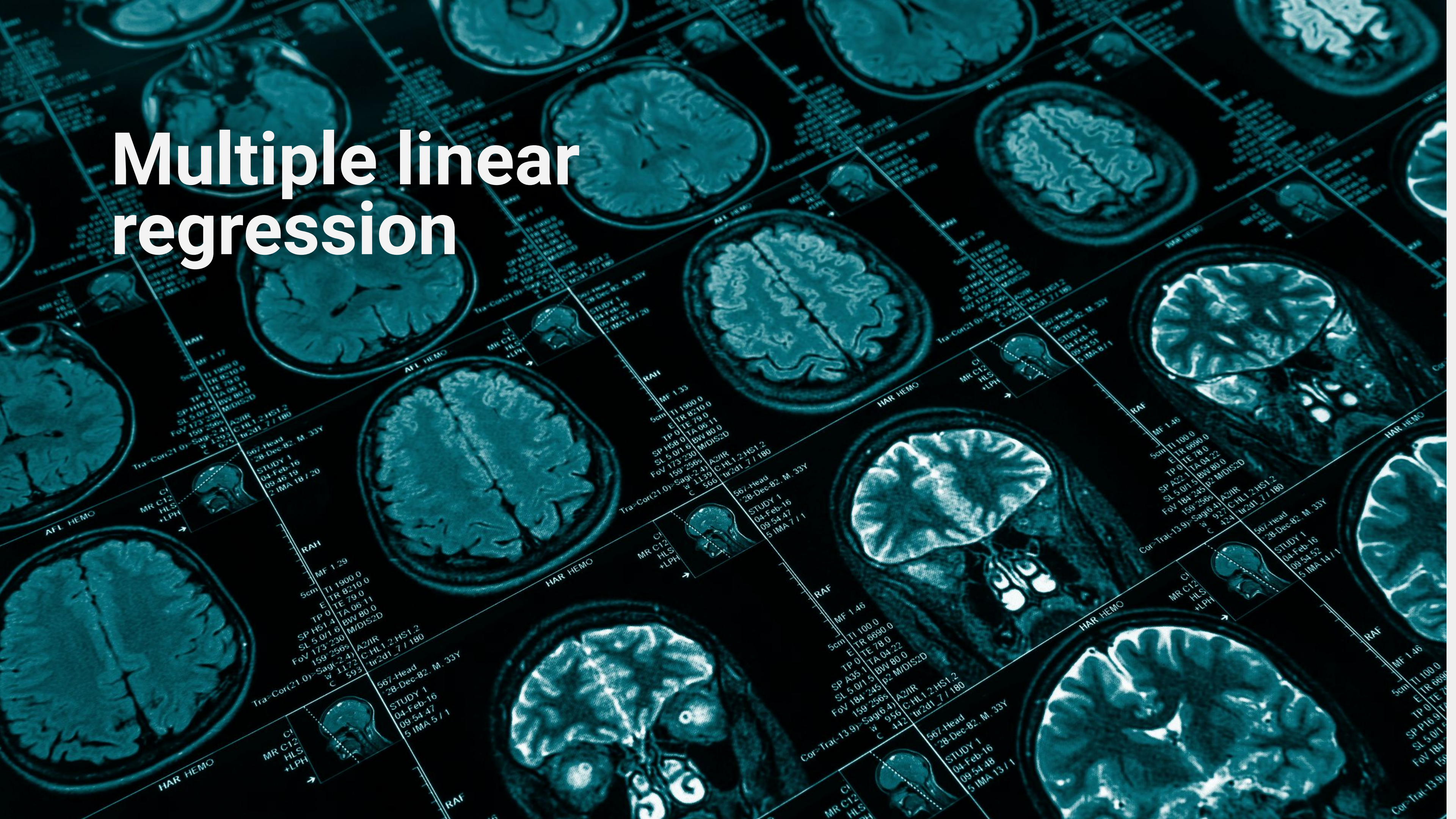
White matter



CSF



Use `LinearRegression` from SciPy to fit regression line between BOLD signal from white matter and BOLD signal from cerebrospinal fluid (CSF).



Multiple linear regression

Linear combination

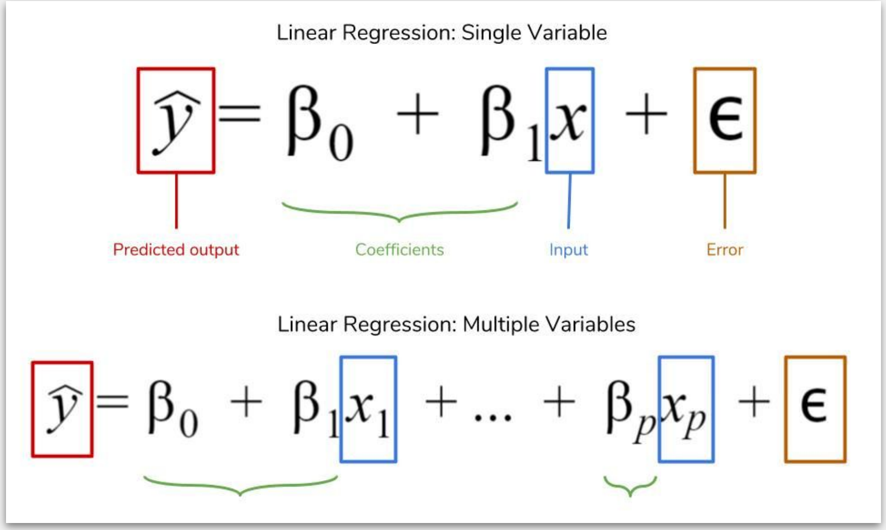
= a

+ b

+ ε

What **combination** of Lily & James gives a better **prediction** of harry?

Multiple linear regression



Each parameter β_i is interpreted as the effect of x_i controlling for all other variables in the model.

Matrix notation

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\text{Observed Data}} = \underbrace{\begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{np} & \cdots & X_{np} \end{bmatrix}}_{\text{Design matrix}} \times \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\text{Model parameters}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\text{Residuals}}$$

Linear algebra magic

Examples: $\underline{m = 4.}$

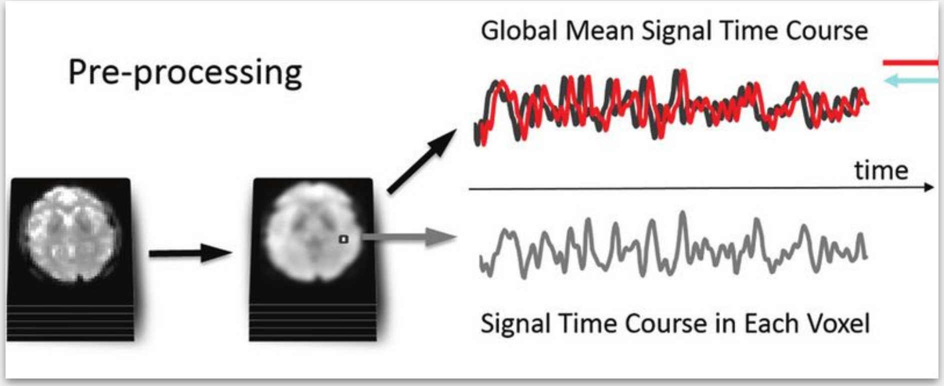
x_0	Size (feet²) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4	Price ($1000) y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

$$X = \underbrace{\begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix}}_{m \times (n+1)} \qquad y = \underbrace{\begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}}_{m\text{-dimensional vector}}$$

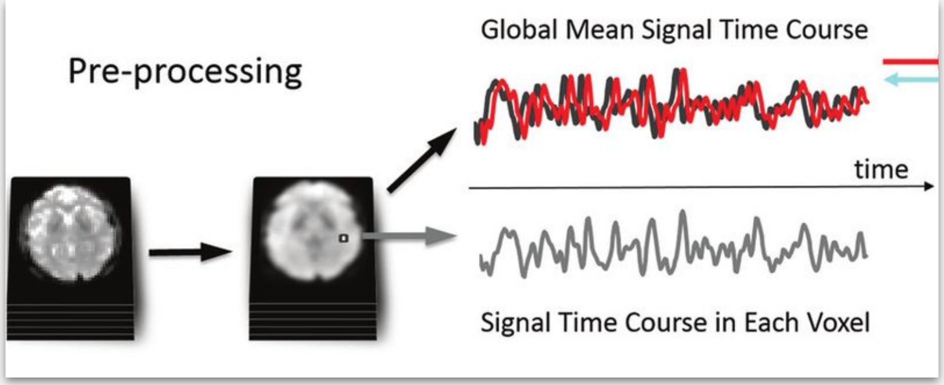
$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

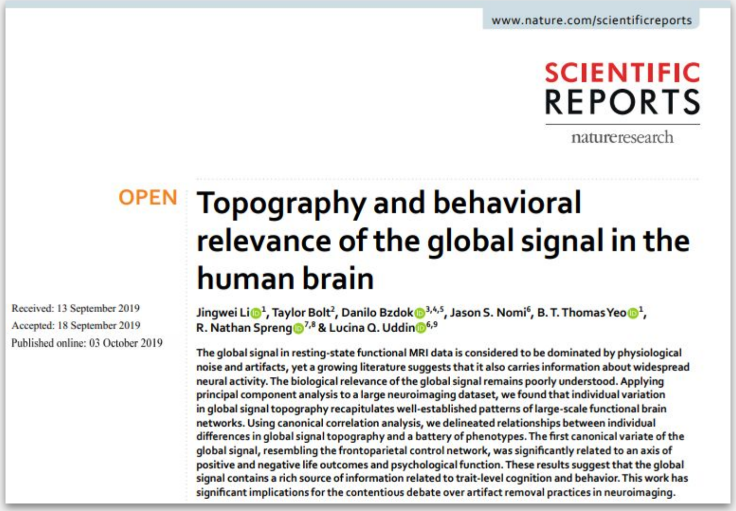
Andrew Ng

Challenge 2


Global Mean Signal Time Course
Pre-processing
time
Signal Time Course in Each Voxel

Challenge 2





www.nature.com/scientificreports

SCIENTIFIC REPORTS
natureresearch

OPEN

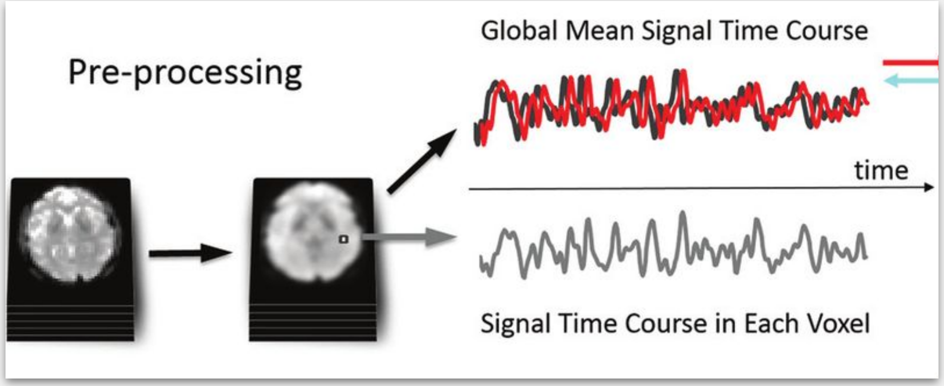
Topography and behavioral relevance of the global signal in the human brain

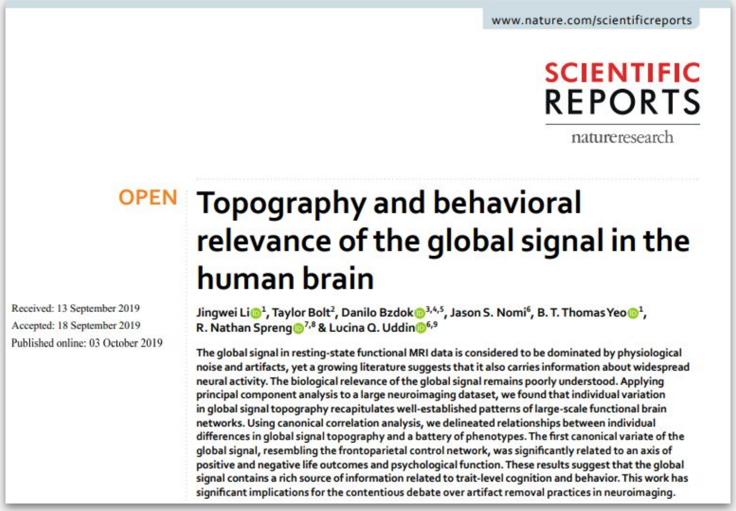
Received: 13 September 2019
Accepted: 18 September 2019
Published online: 03 October 2019

Jingwei Li[1], Taylor Bolt[2], Danilo Bzdok[3,4,5], Jason S. Nomi[6], B. T. Thomas Yeo[1], R. Nathan Spreng[7,8] & Lucina Q. Uddin[6,9]

The global signal in resting-state functional MRI data is considered to be dominated by physiological noise and artifacts, yet a growing literature suggests that it also carries information about widespread neural activity. The biological relevance of the global signal remains poorly understood. Applying principal component analysis to a large neuroimaging dataset, we found that individual variation in global signal topography recapitulates well-established patterns of large-scale functional brain networks. Using canonical correlation analysis, we delineated relationships between individual differences in global signal topography and a battery of phenotypes. The first canonical variate of the global signal, resembling the frontoparietal control network, was significantly related to an axis of positive and negative life outcomes and psychological function. These results suggest that the global signal contains a rich source of information related to trait-level cognition and behavior. This work has significant implications for the contentious debate over artifact removal practices in neuroimaging.

Challenge 2




www.nature.com/scientificreports

SCIENTIFIC REPORTS

natureresearch

OPEN

Topography and behavioral relevance of the global signal in the human brain

Received: 13 September 2019
Accepted: 18 September 2019
Published online: 03 October 2019

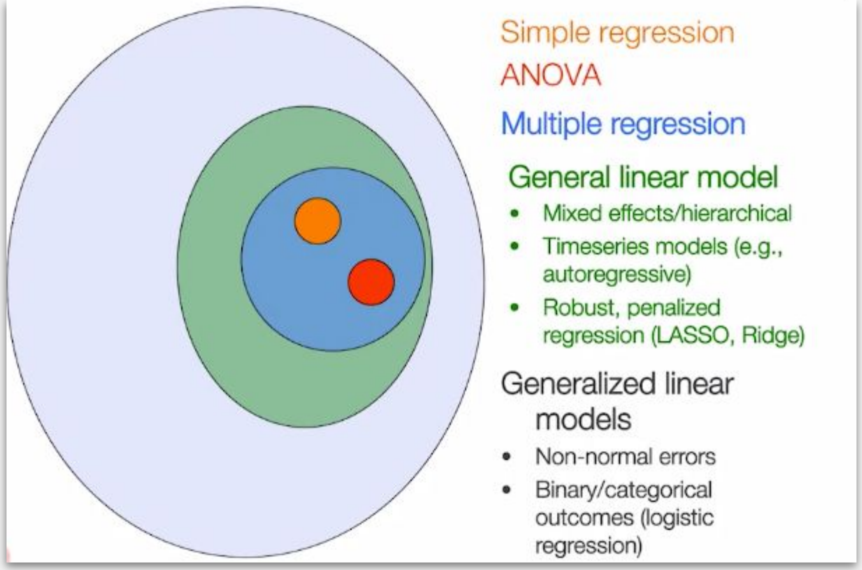
Jingwei Li[1], Taylor Bolt[2], Danilo Bzdok[3,4,5], Jason S. Nomi[6], B. T. Thomas Yeo[1], R. Nathan Spreng[7,8] & Lucina Q. Uddin[6,9]

The global signal in resting-state functional MRI data is considered to be dominated by physiological noise and artifacts, yet a growing literature suggests that it also carries information about widespread neural activity. The biological relevance of the global signal remains poorly understood. Applying principal component analysis to a large neuroimaging dataset, we found that individual variation in global signal topography recapitulates well-established patterns of large-scale functional brain networks. Using canonical correlation analysis, we delineated relationships between individual differences in global signal topography and a battery of phenotypes. The first canonical variate of the global signal, resembling the frontoparietal control network, was significantly related to an axis of positive and negative life outcomes and psychological function. These results suggest that the global signal contains a rich source of information related to trait-level cognition and behavior. This work has significant implications for the contentious debate over artifact removal practices in neuroimaging.

Use `LinearRegression` from SciPy to predict global signal from 3 rotations and 3 translation parameters. Which motion parameter has highest beta value?

What about brain activity analysis?

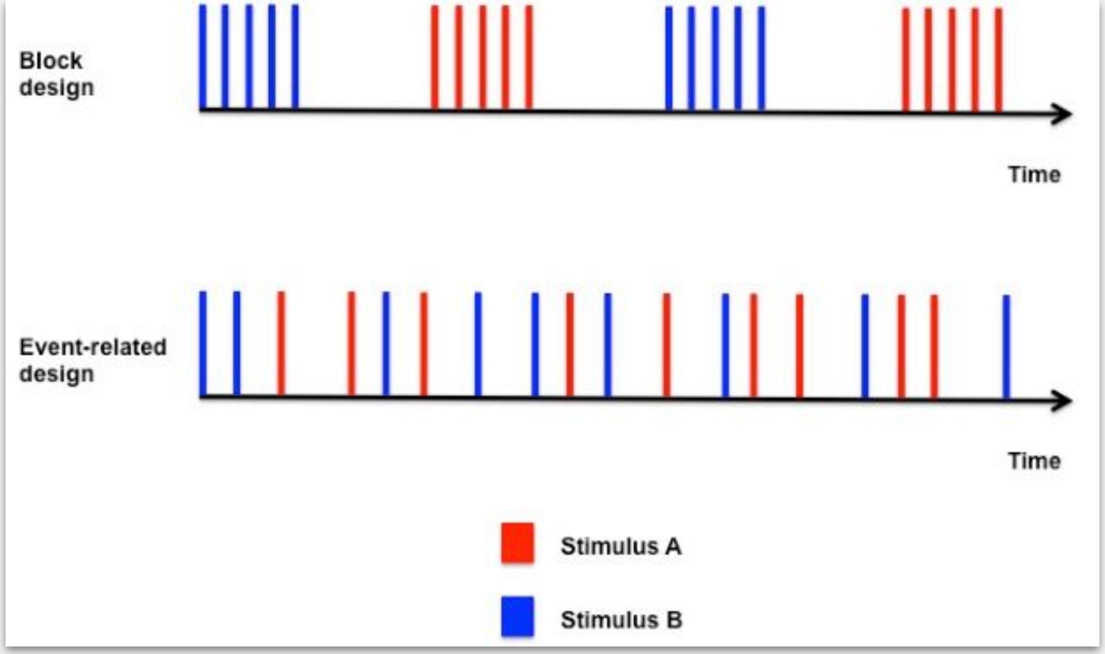
Generalized Linear Model



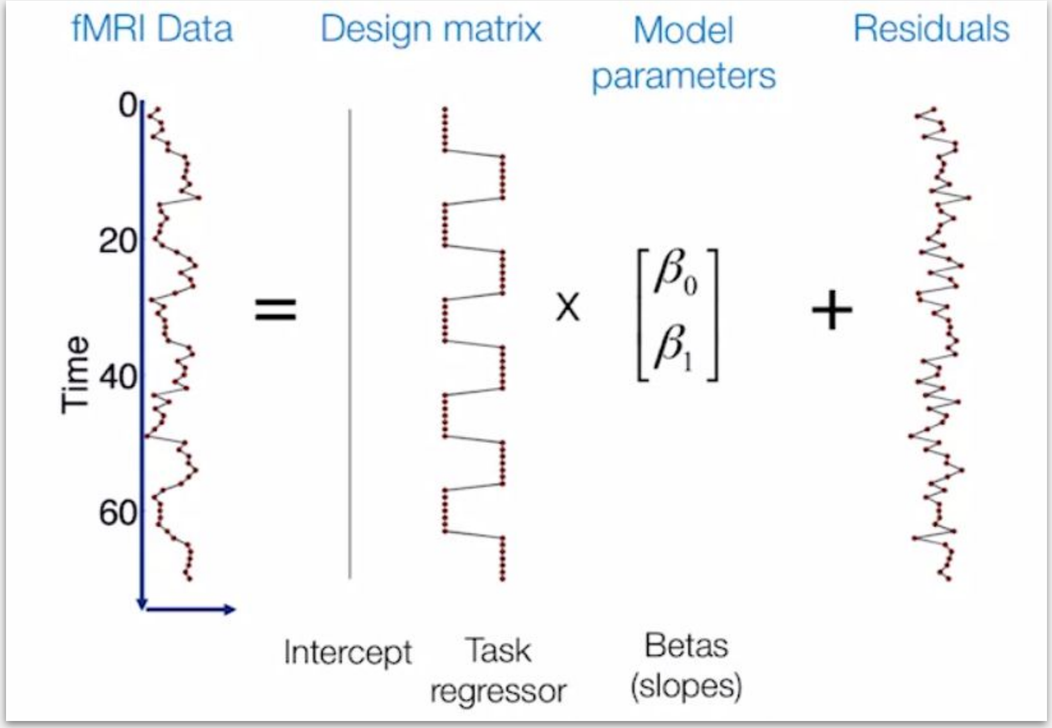
Coursera, *Principles of fMRI*

The general linear model (GLM) approach treats the fMRI data as a linear combination of model functions, predictors, plus noise, or error.

Task designs

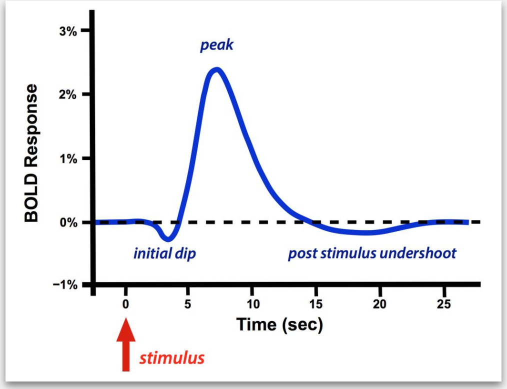


First level GLM

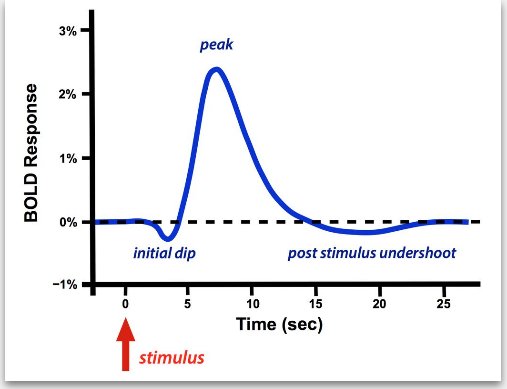


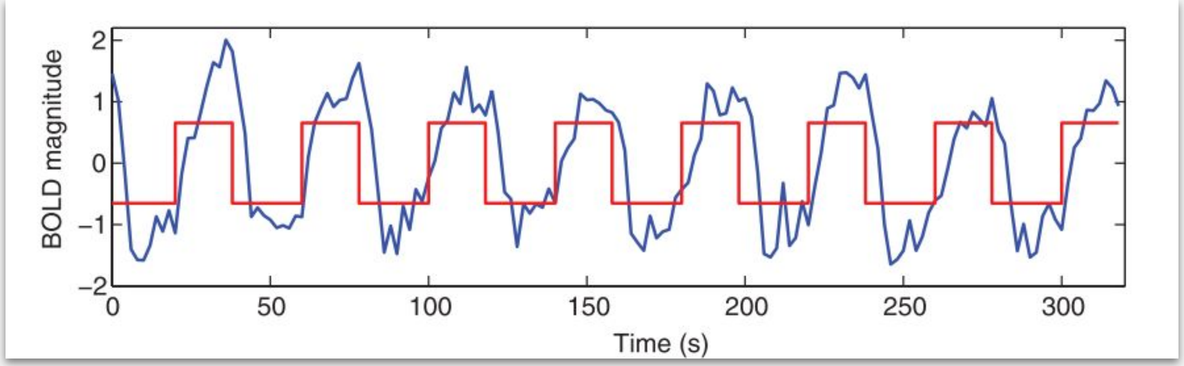
What we have missed?

Haemodynamic delay!

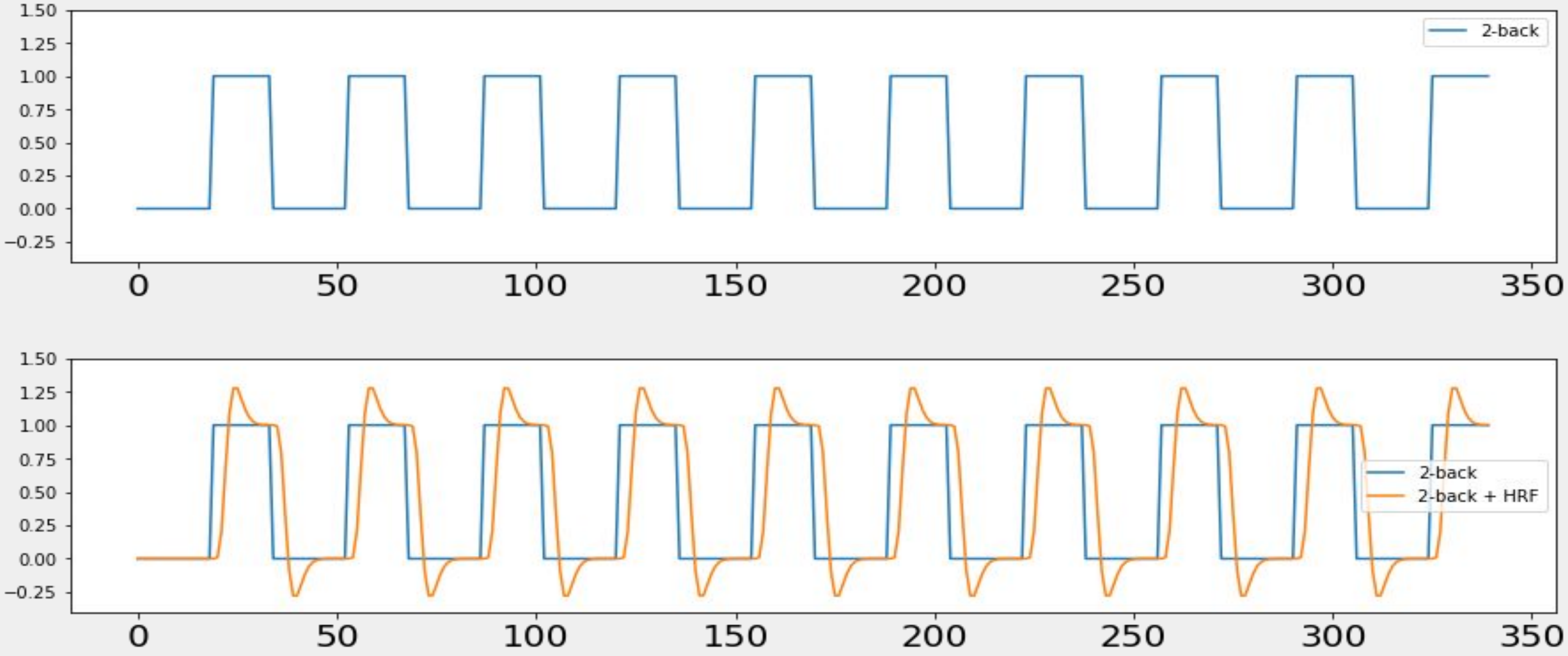


Haemodynamic delay!

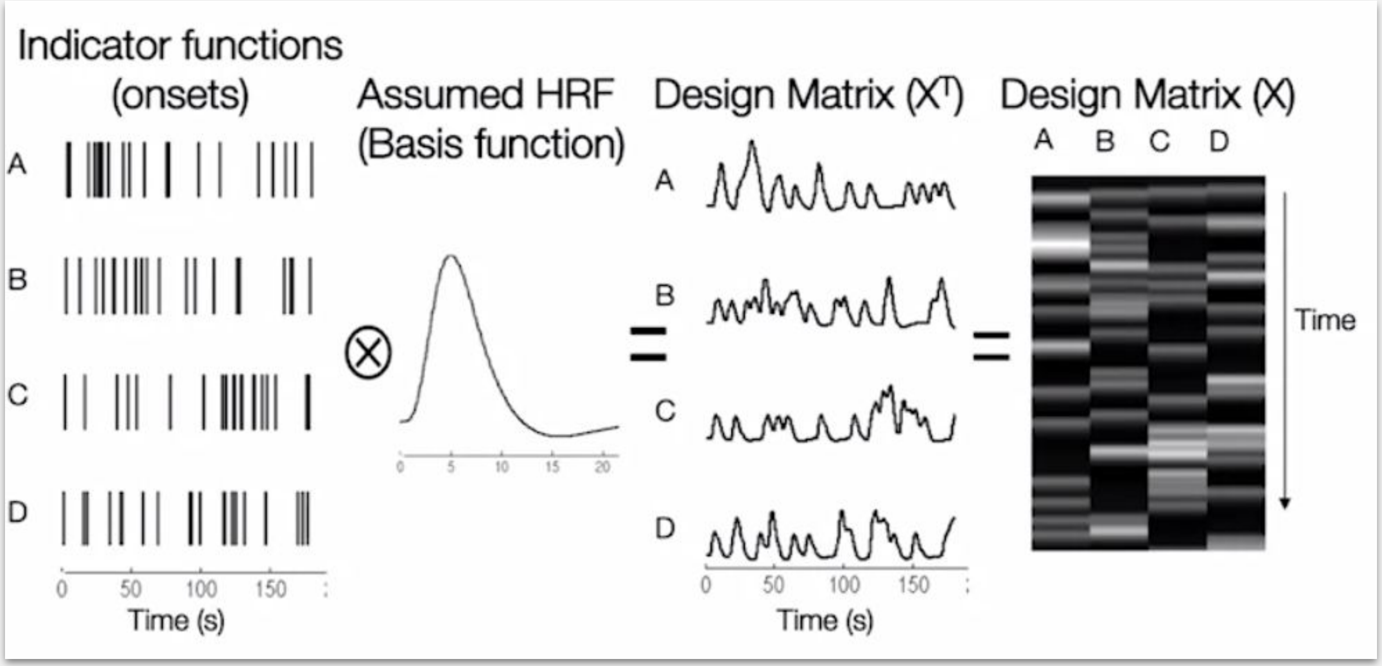




Convolution



Building design matrix from events data



Homework

1. **GitHub Classroom**

    Linear and multiple linear regression
    Deadline: 08-05-2020

2. **Data Camp Classroom**

    `https://www.datacamp.com/enterprise/advanced-fmri-data-analysis/assignments`

    Introduction to Linear Modeling in Python
    Deadline: 08-05-2020

Next

General Linear Model 2