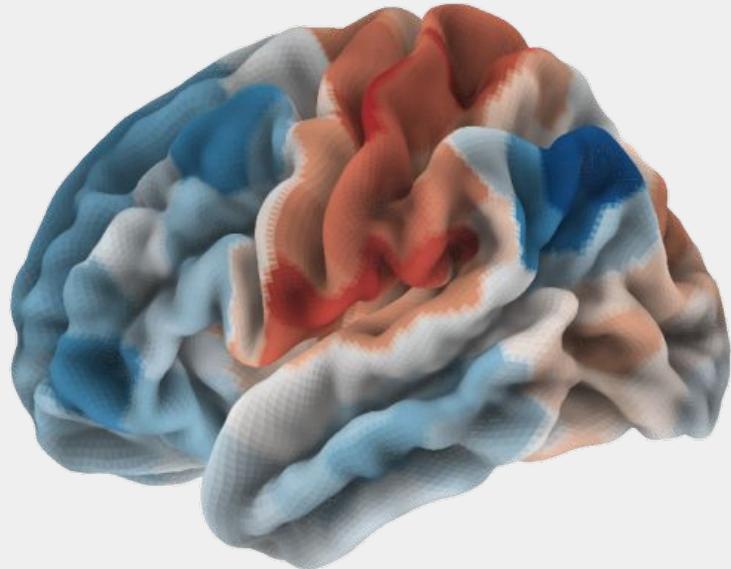


# Fundamentals of fMRI data analysis

Karolina Finc

Centre for Modern Interdisciplinary Technologies  
Nicolaus Copernicus University in Toruń



COURSE #4: **General Linear Model 1** | 30<sup>th</sup> November 2020

# Study plan

Open science & neuroimaging



**BEFORE**

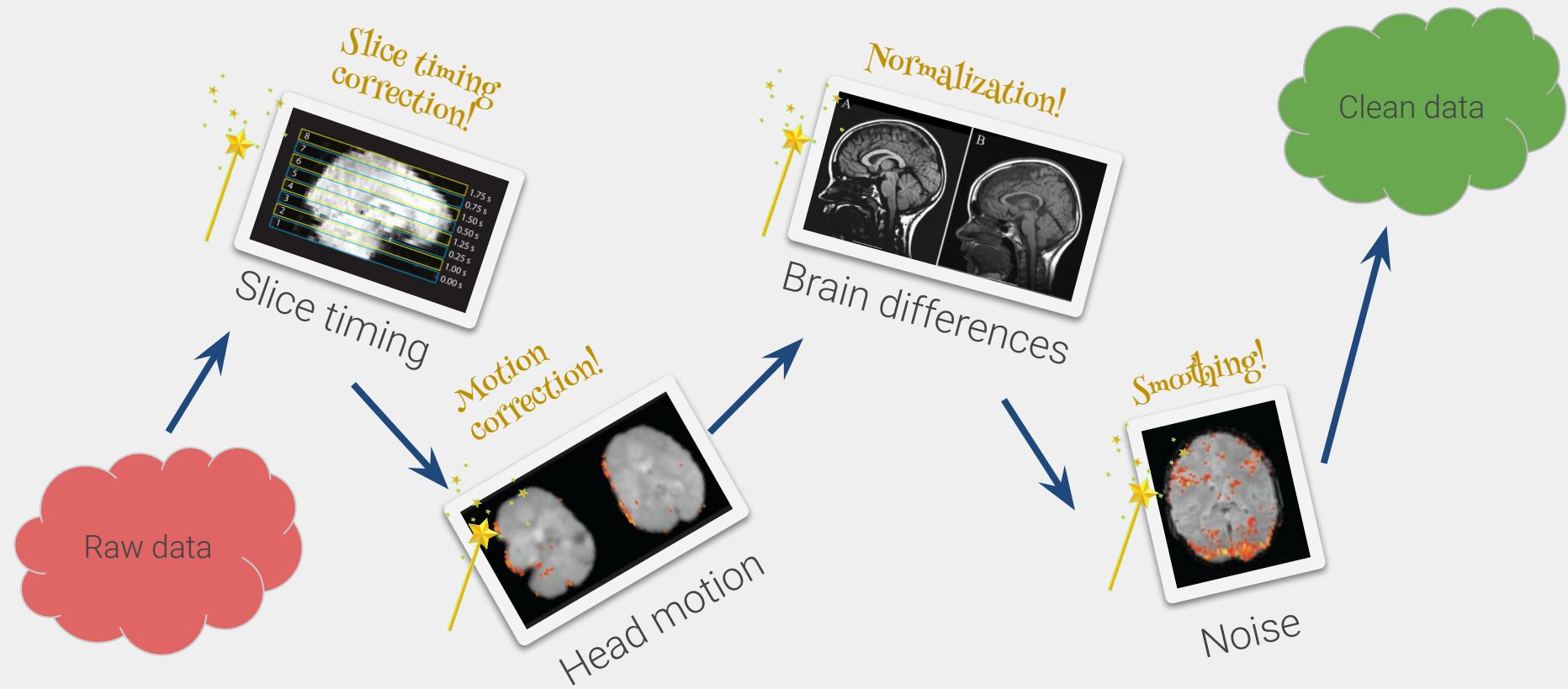
fMRI data manipulation  
in python



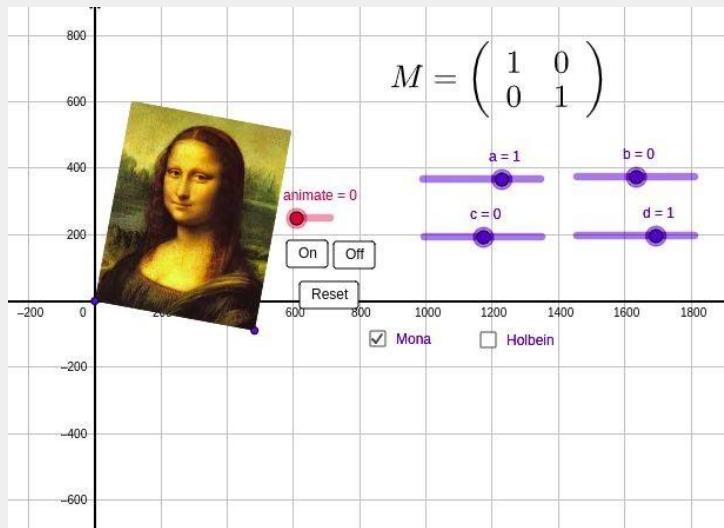
fMRI data  
preprocessing



# Preprocessing workflow / pipeline



# Linear transformations magic!

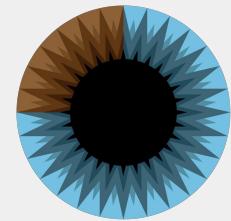


# GeoGebra

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

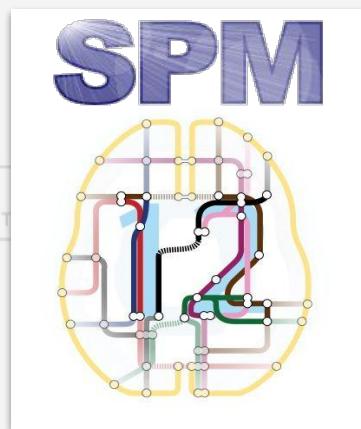


[https://www.khanacademy.org/math/linear-algebra/matrix-transformation\\_s/linear-transformations/a/visualizing-linear-transformations](https://www.khanacademy.org/math/linear-algebra/matrix-transformation_s/linear-transformations/a/visualizing-linear-transformations)



[https://www.youtube.com/channel/UCYO\\_jabesuFRV4b17AJtAw](https://www.youtube.com/channel/UCYO_jabesuFRV4b17AJtAw)

# Software



?

?

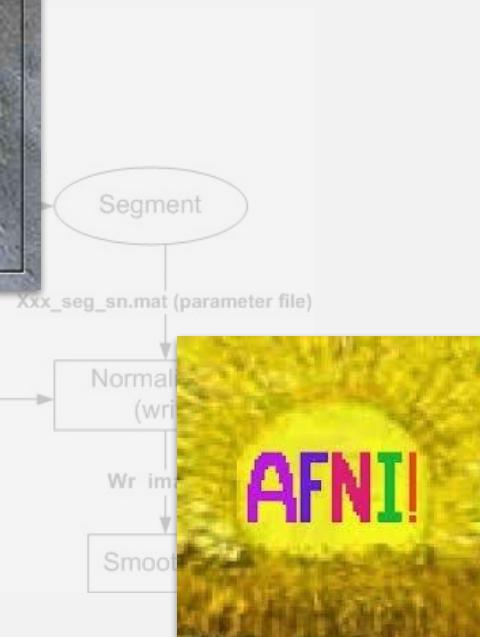
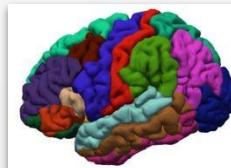
?

?

?

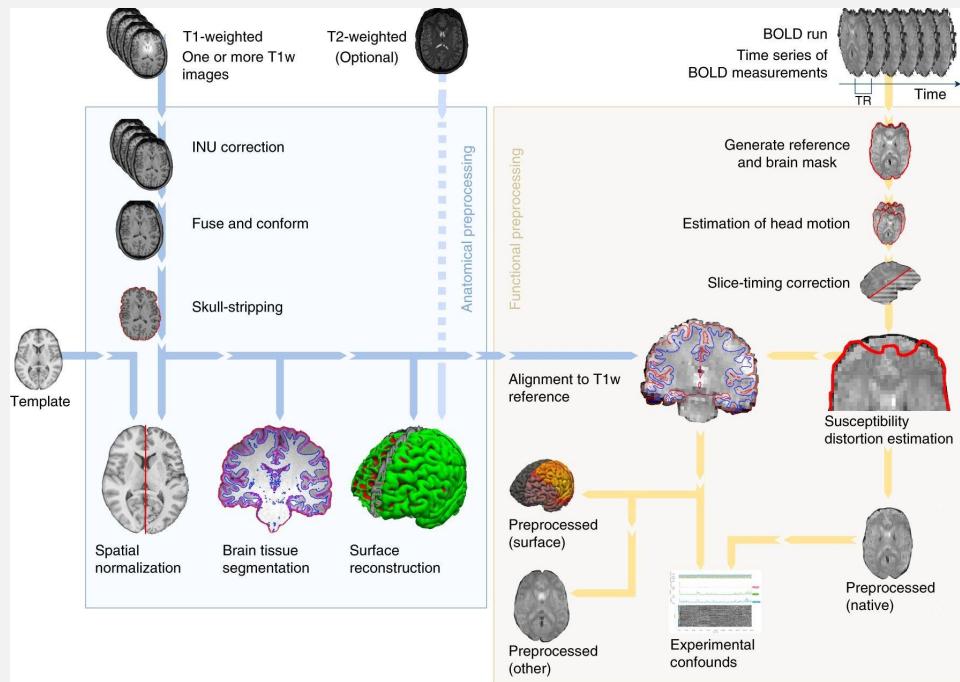
?

FreeSurfer



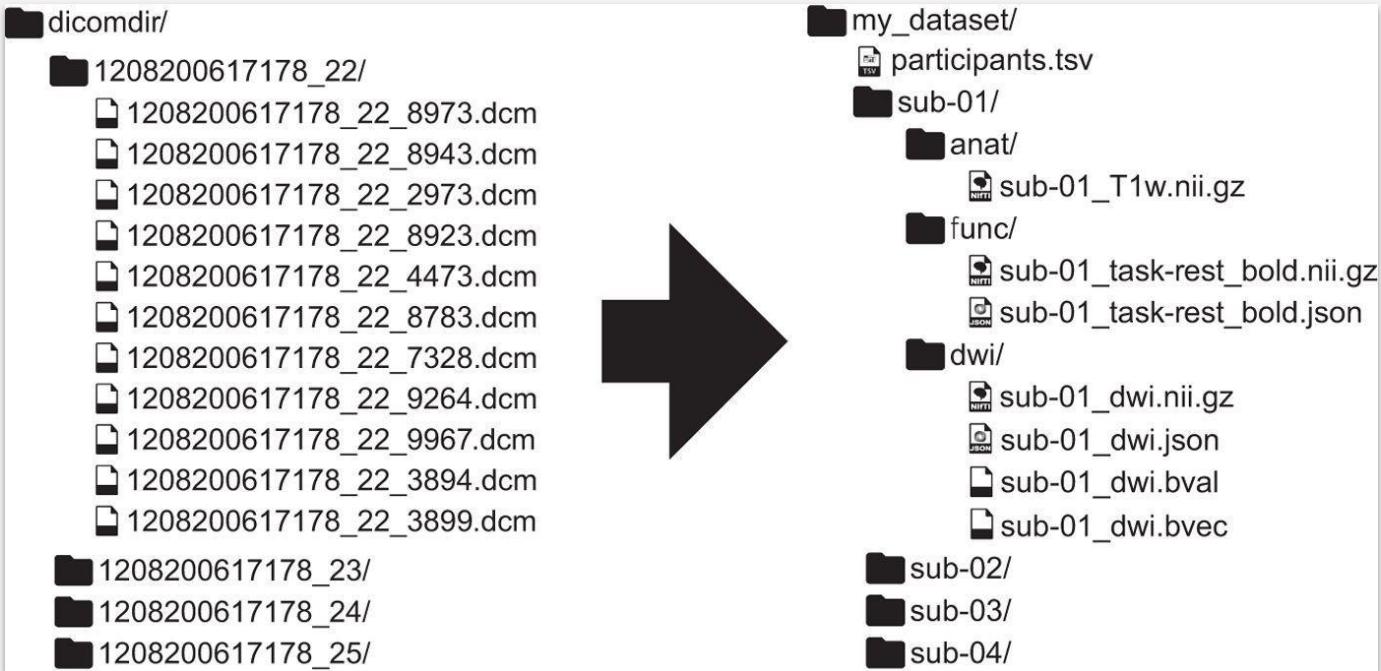
# fMRIPrep!

The screenshot shows the fMRIprep stable documentation page. The left sidebar includes links for Installation, What's new, Usage (Execution and the BIDS format, Command-Line Arguments, Docker CLI, Debugging, Support and communication), Processing pipeline details (Susceptibility Distortion Correction (SDC), Outputs of fMRIPrep, Contributing to fMRIPREP, Citing fMRIPREP, API), and a search bar. The main content area has a "Usage" heading with a warning about usage statistics and a command-line example: `fmriprep data/bids_root/ out/ participant -w work/`.



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

# Brain Imaging Data Structure (BIDS)



# Study plan

Open science & neuroimaging



**BEFORE**

fMRI data manipulation  
in python



fMRI data  
preprocessing



Functional  
connectivity



General  
Linear Model



**AFTER**



Machine Learning  
on fMRI data

# Study plan

Open science & neuroimaging



**BEFORE**

fMRI data manipulation  
in python



fMRI data  
preprocessing



Functional  
connectivity



**AFTER**



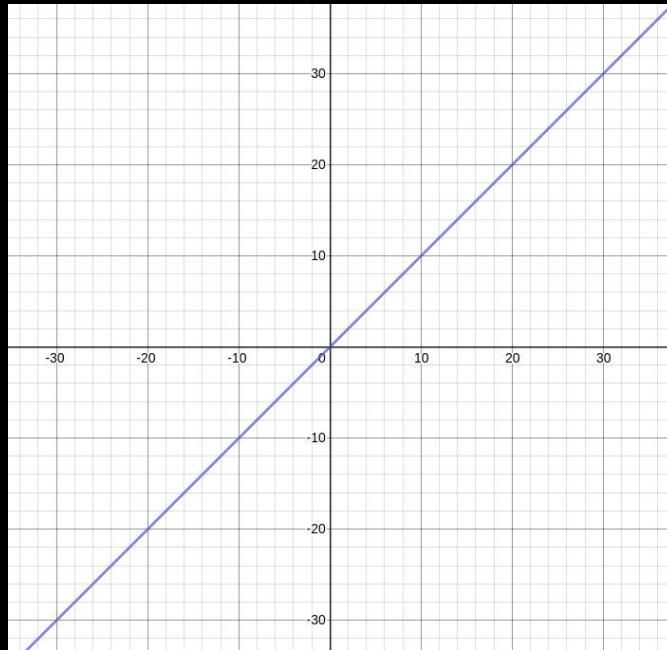
General  
Linear Model



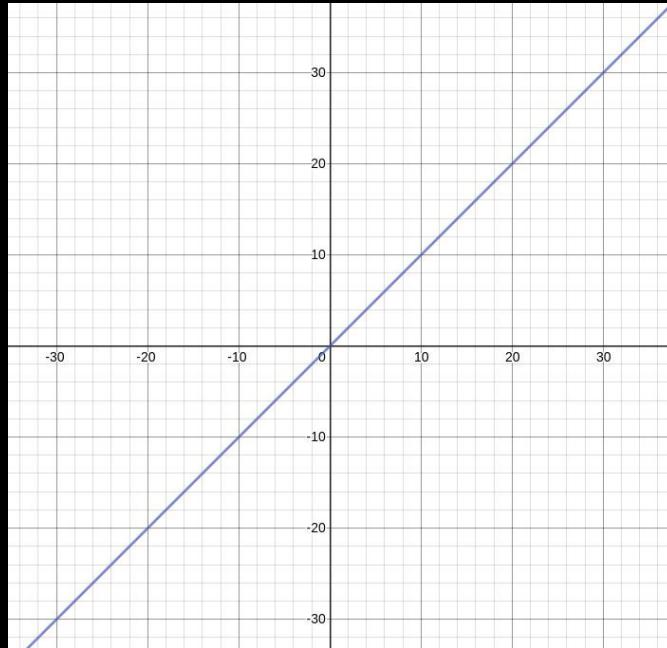
Machine Learning  
on fMRI data



# Guess the function formula!

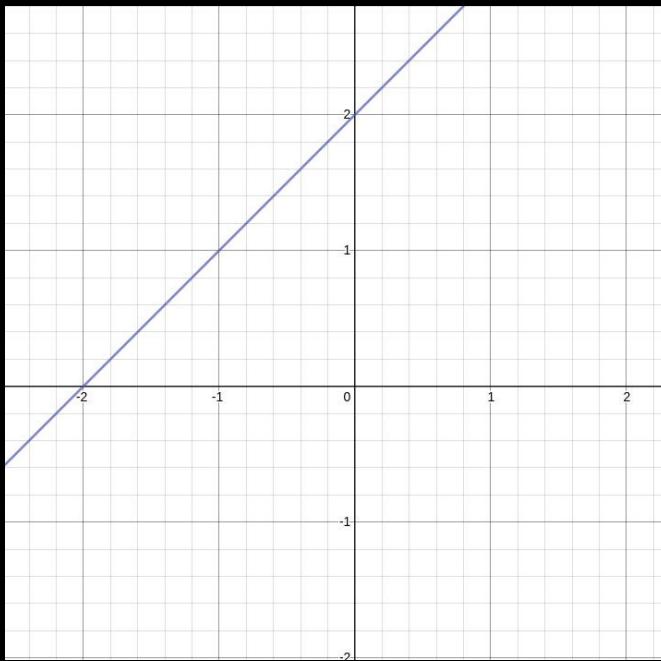


# Guess the function formula!

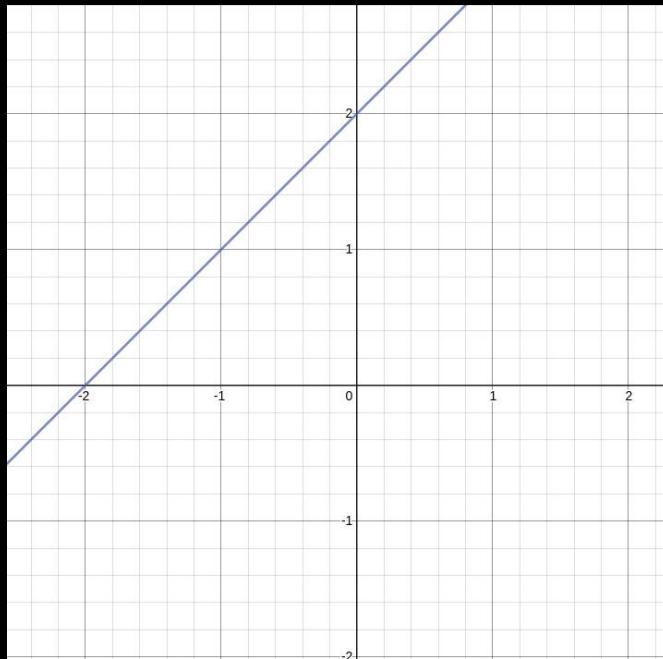


$$y = x$$

# Guess the function formula!

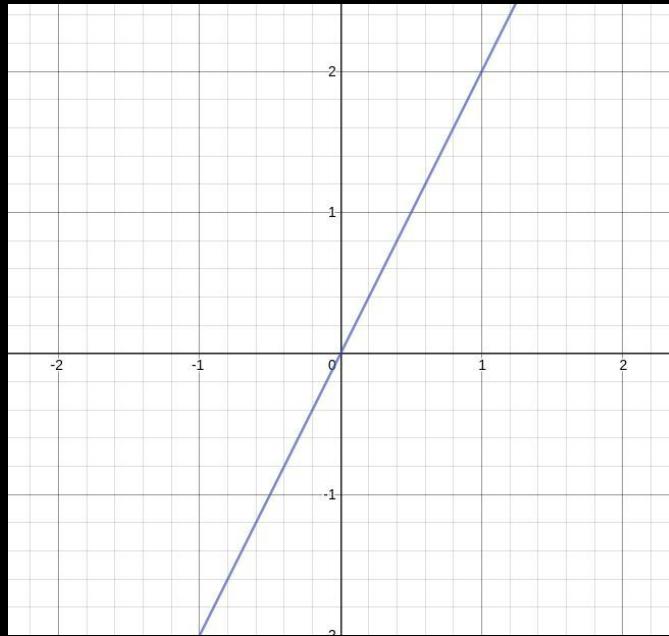


# Guess the function formula!

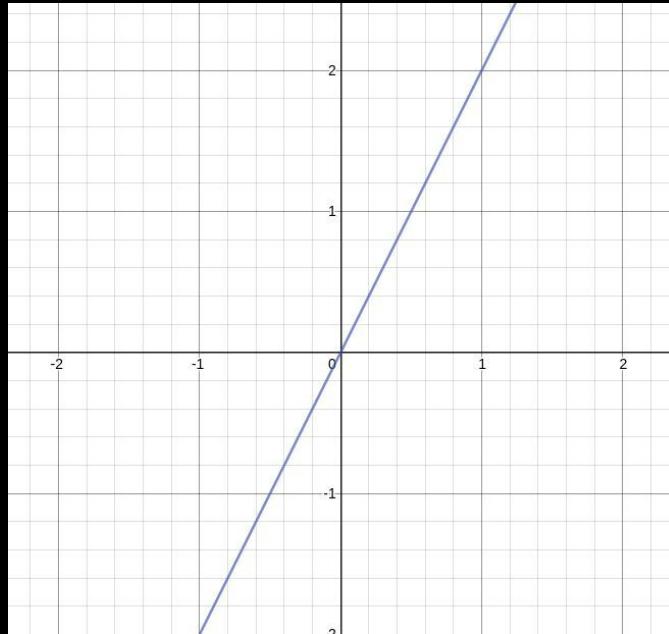


$$y = x + 2$$

# Guess the function formula!

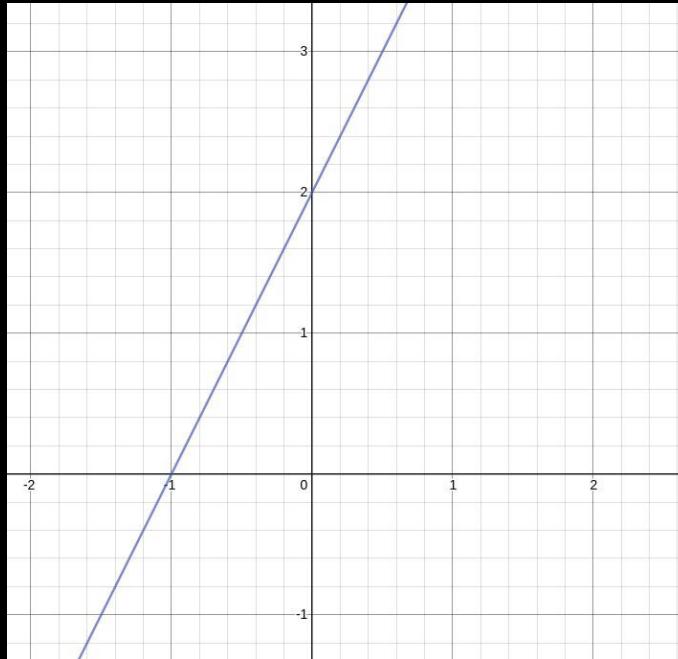


# Guess the function formula!

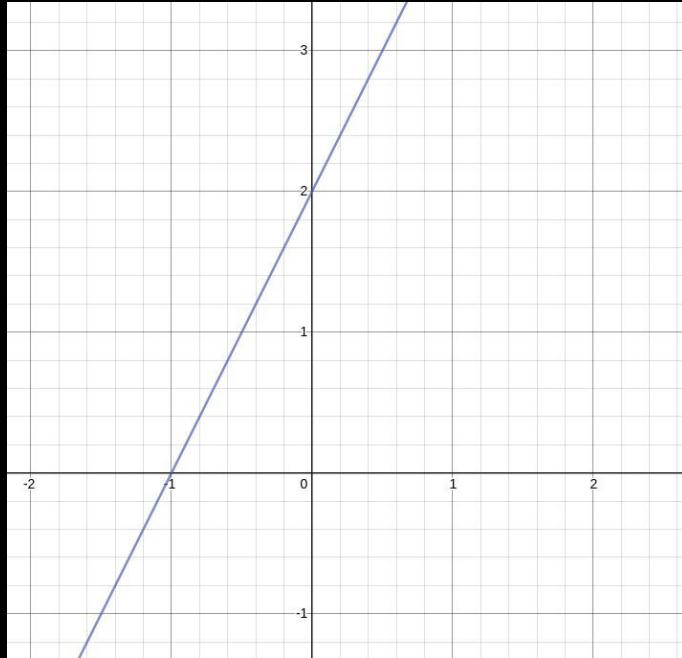


$$y = 2x$$

# 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

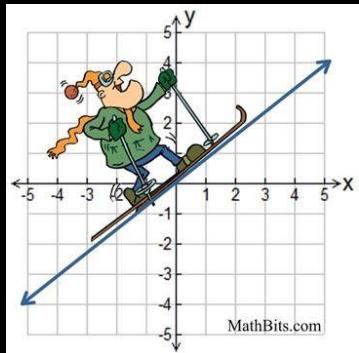
slope



$$y = \textcolor{green}{m}x + \textcolor{blue}{b}$$

# Slope-intercept form of linear function

slope



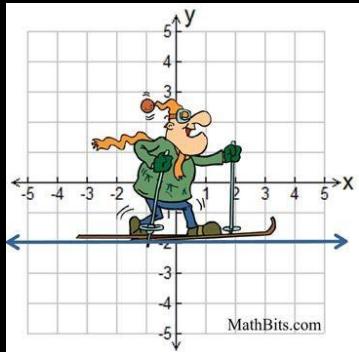
$$y = mx + b$$



# Slope-intercept form of linear function

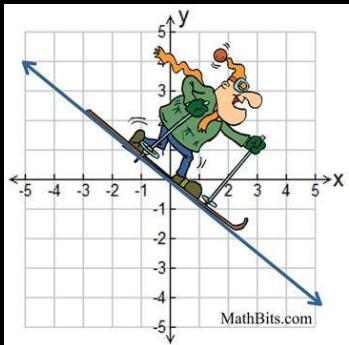
slope

$$y = mx + b$$



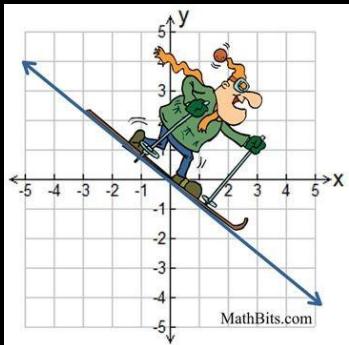
# Slope-intercept form of linear function

slope



$$y = mx + b$$

# Slope-intercept form of linear function

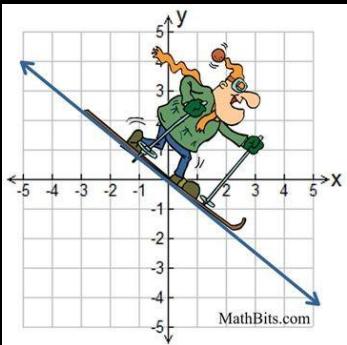


slope

intercept

$$y = \textcolor{green}{m}x + \textcolor{blue}{b}$$

# Slope-intercept form of linear function



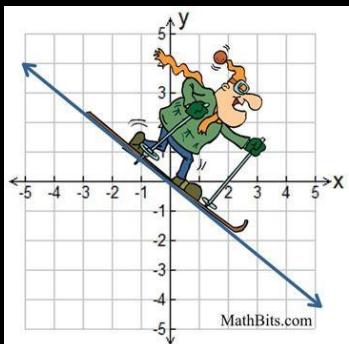
slope

intercept

$$y = \textcolor{green}{m}x + \textcolor{blue}{b}$$

coefficients

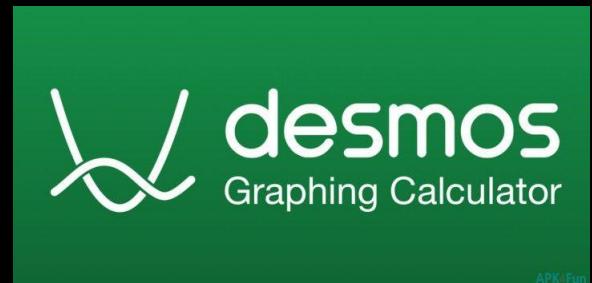
# Slope-intercept form of linear function



$$y = \text{slope} \cdot x + \text{intercept}$$

↓                    ↓                    ↓

coefficients



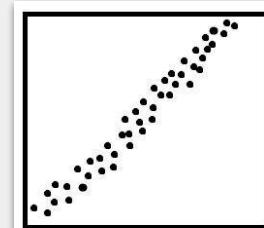
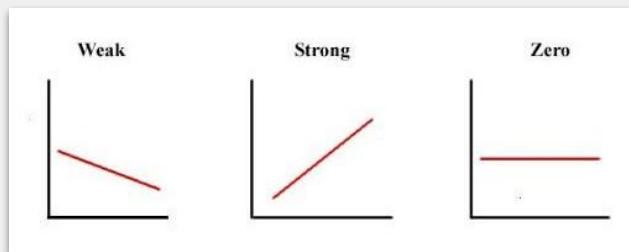
<https://www.desmos.com/calculator>

# Give a few examples of possible linear association

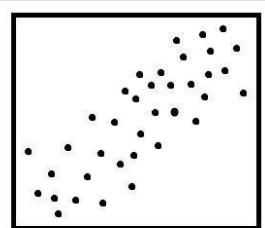


# Examples of linear associations

- 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.

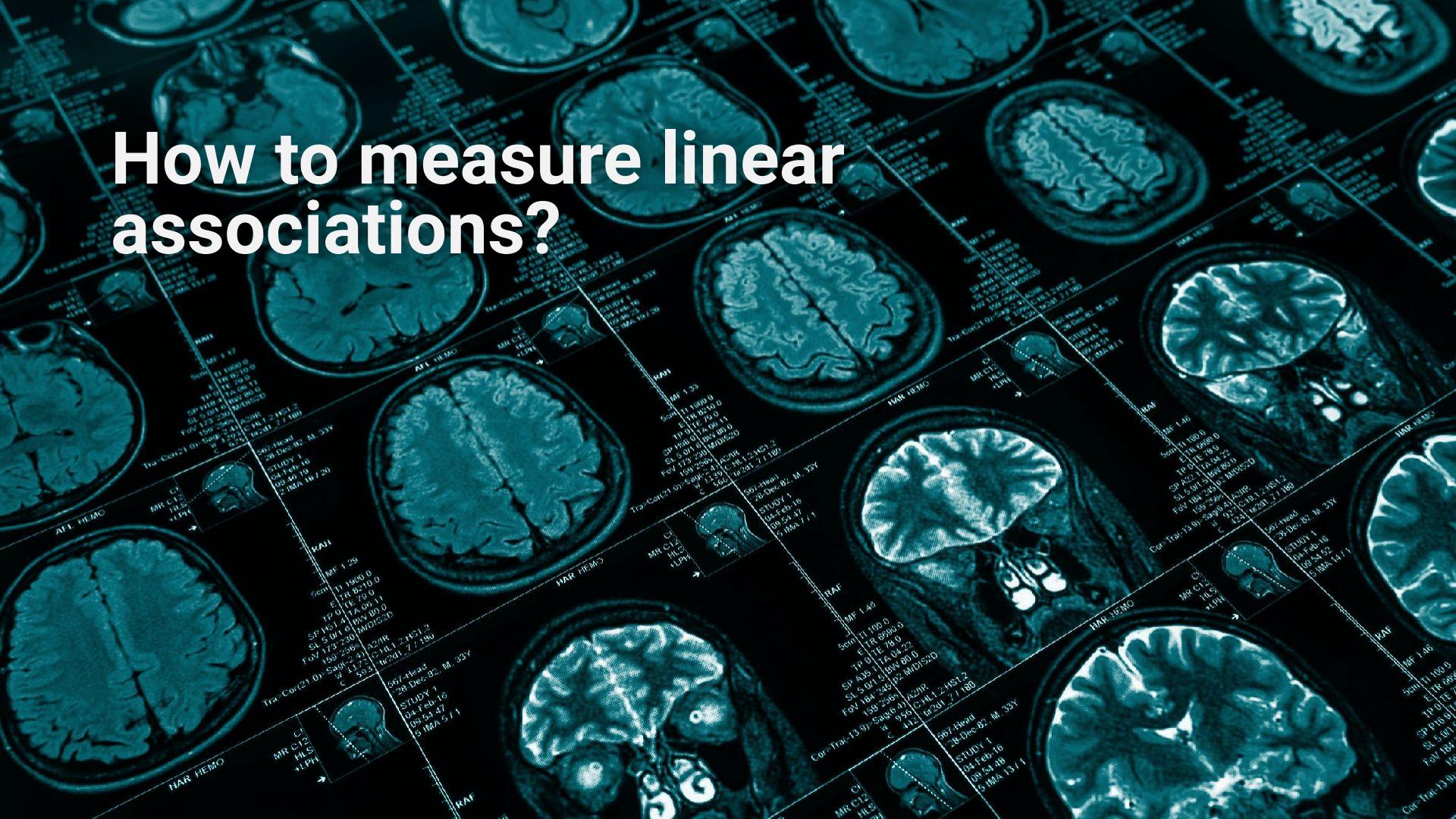


strong positive linear association



weak positive linear association

# 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

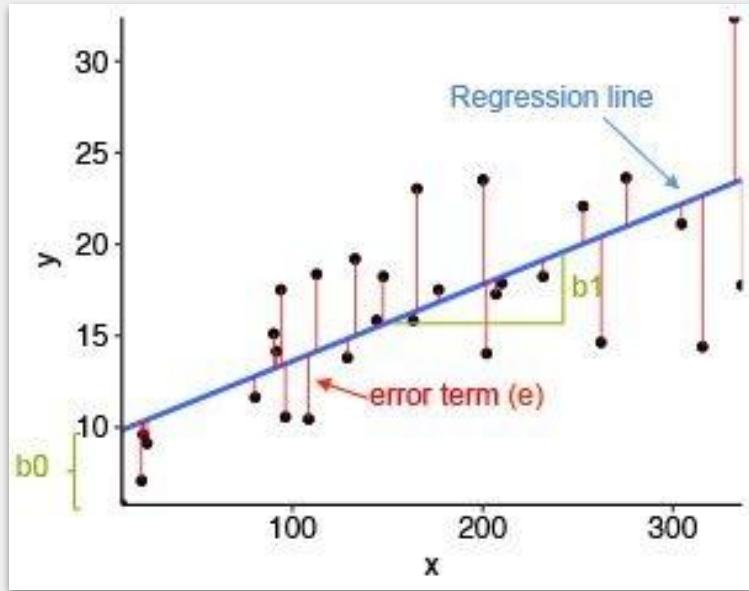


Regression line provides a **model** of the data

**Regression problem:** predict real-valued output

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

# Fitting regression line

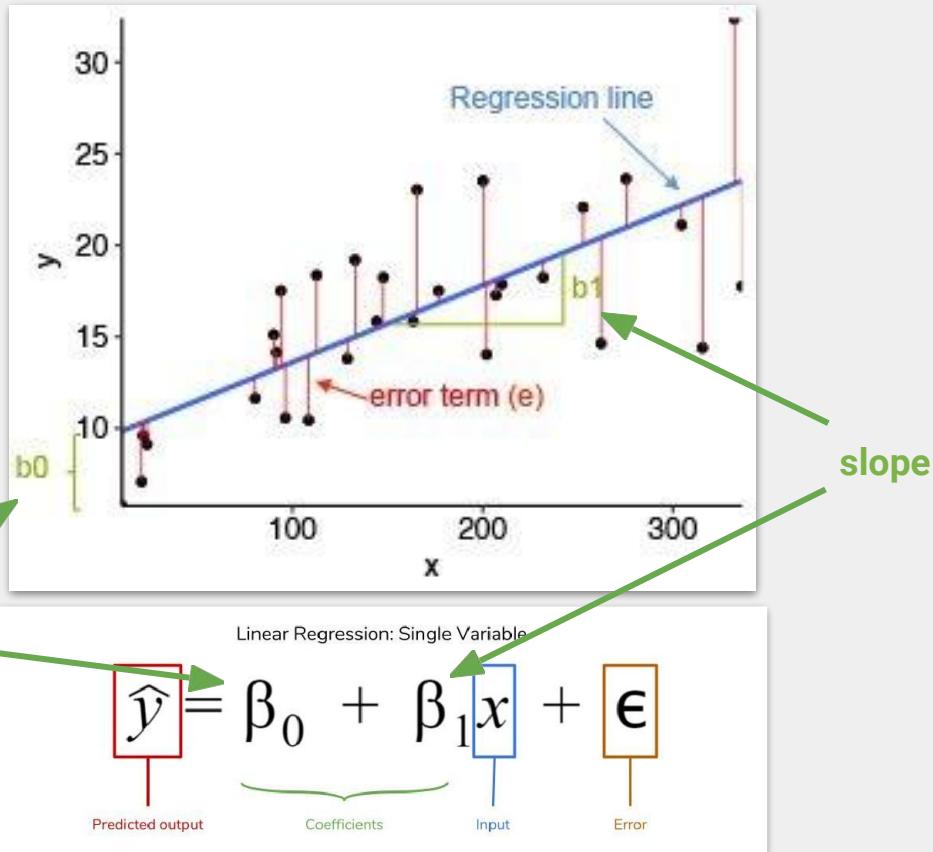


Linear Regression: Single Variable

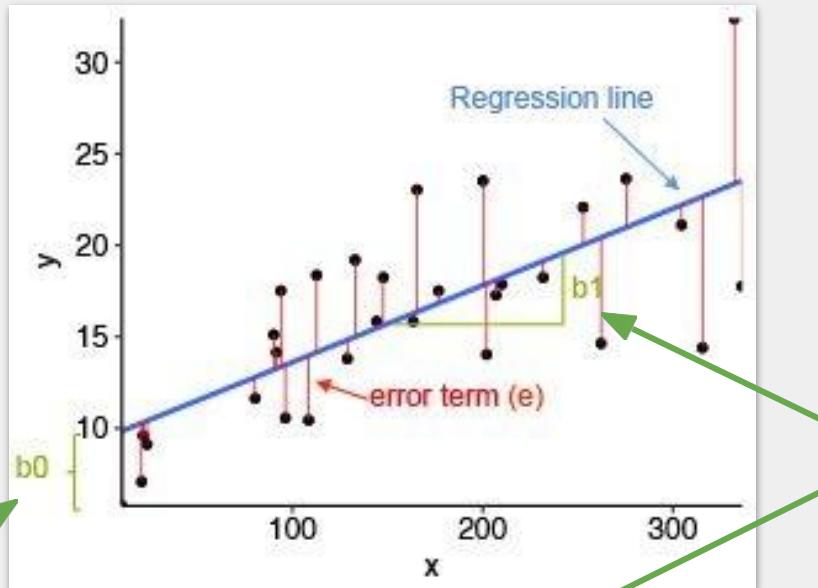
$$\hat{y} = \beta_0 + \beta_1 x + \epsilon$$

Predicted output      Coefficients      Input      Error

# Fitting regression line



# Fitting regression line



Find such  $\beta_0$  and  $\beta_1$  that minimize cost function: **sum of squared errors** function

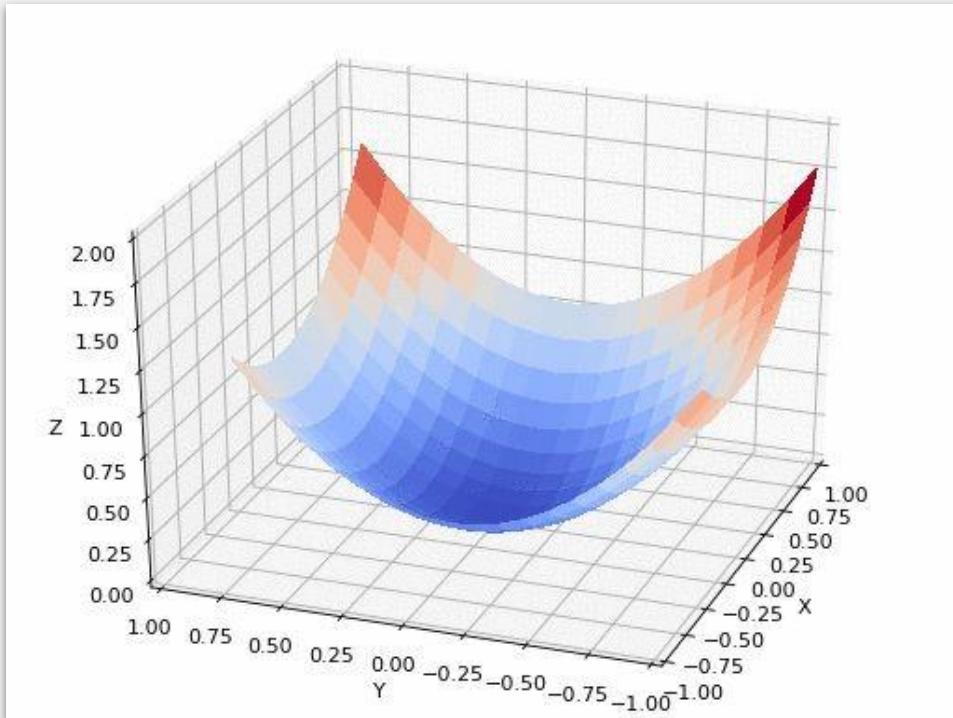
intercept

Linear Regression: Single Variable

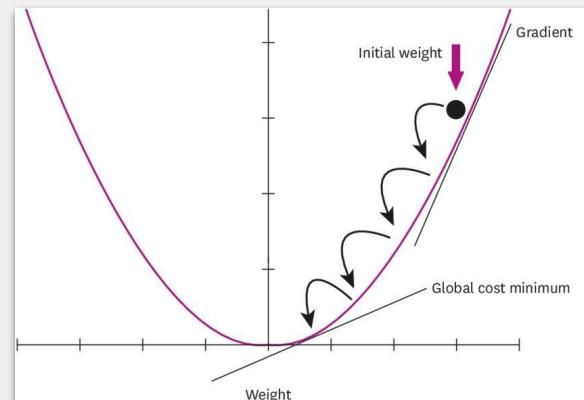
$$\hat{y} = \beta_0 + \beta_1 x + \epsilon$$

Predicted output      Coefficients      Input      Error

# Gradient descent

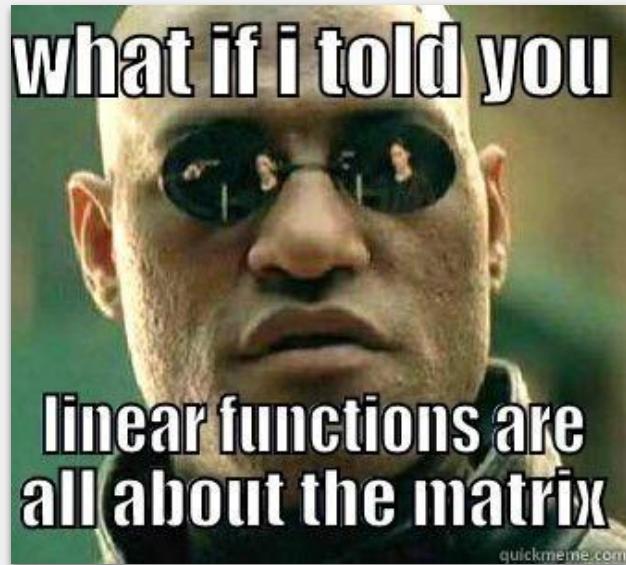


- Algorithm for minimizing cost function
- Is used not only in linear regression



# 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

House sizes:

$$\rightarrow 2104$$

$$\rightarrow 1416$$

$$\rightarrow 1534$$

$$\rightarrow 852$$

Matrix  $x$

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix}$$

$$h_{\theta}(x) = -40 + 0.25x$$

$$h_{\theta}(x)$$

2+1  
Vector

$$\begin{bmatrix} -40 \\ 0.25 \end{bmatrix}$$

$$h_{\theta}(2104)$$

4x1 matrix

$$\begin{bmatrix} -40 \times 1 + 0.25 \times 2104 \end{bmatrix}$$

$$\begin{bmatrix} -40 \times 1 + 0.25 \times 1416 \end{bmatrix}$$

$$\begin{bmatrix} h_{\theta}(1416) \end{bmatrix}$$

Prediction

4x1

$$= \text{DataMatrix} \times \text{Parameters}$$

for  $i = 1:1000$ ,  
 $\text{prediction}(i) = \dots$

# Matrix matrix multiplication

House sizes:

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

Have 3 competing hypotheses:

$$\begin{aligned} 1. \quad h_{\theta}(x) &= -40 + 0.25x \\ 2. \quad h_{\theta}(x) &= 200 + 0.1x \\ 3. \quad h_{\theta}(x) &= -150 + 0.4x \end{aligned}$$

Matrix

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix}$$

Matrix

$$\begin{bmatrix} -40 \\ 0.25 \end{bmatrix} \quad \begin{bmatrix} 200 \\ 0.1 \end{bmatrix} \quad \begin{bmatrix} -150 \\ 0.4 \end{bmatrix}$$

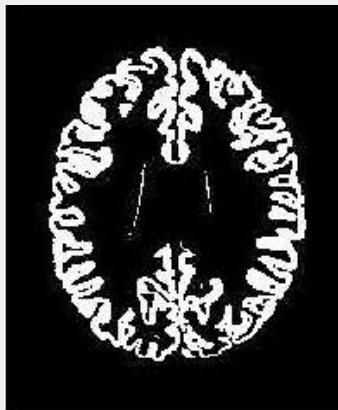
$$\begin{bmatrix} 486 \\ 314 \\ 344 \\ 173 \end{bmatrix} \quad \begin{bmatrix} 410 \\ 342 \\ 353 \\ 285 \end{bmatrix} \quad \begin{bmatrix} 692 \\ 416 \\ 464 \\ 191 \end{bmatrix}$$

Prediction  
of 1<sup>st</sup>  
 $h_{\theta}$

Predictions  
of 2<sup>nd</sup>  
 $h_{\theta}$

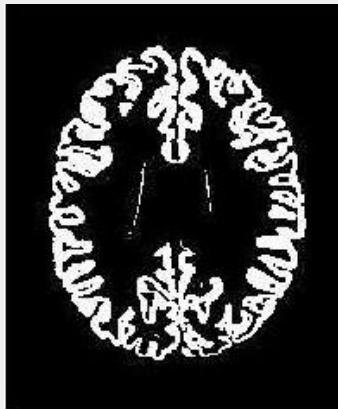
# Challenge 1

Gray matter

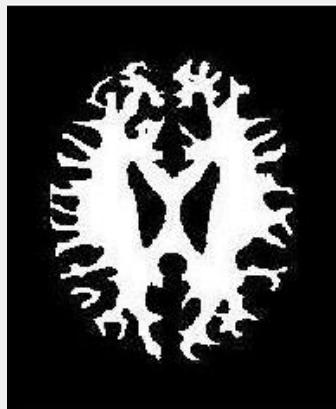


# Challenge 1

Gray matter

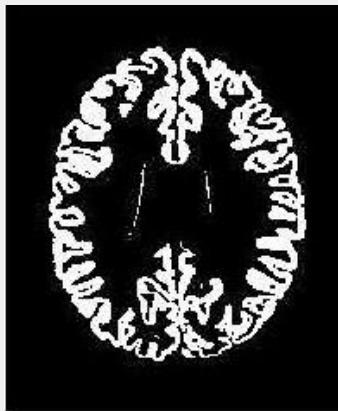


White matter

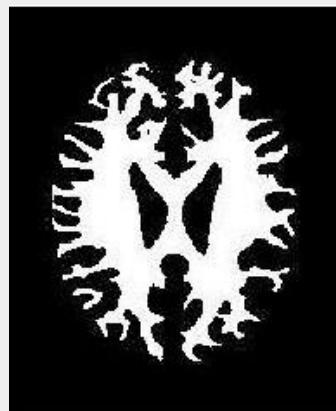


# Challenge 1

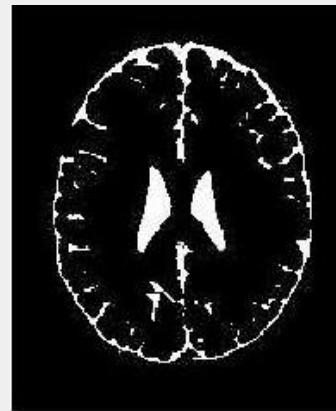
Gray matter



White matter



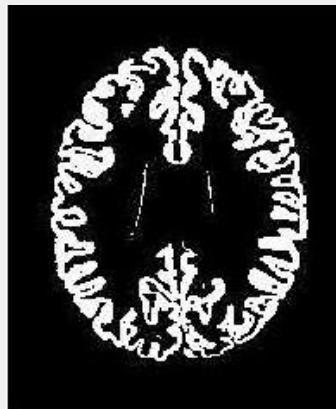
CSF



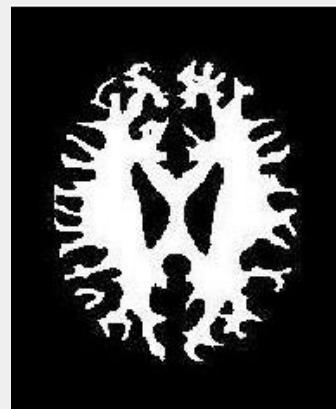
# Challenge 1



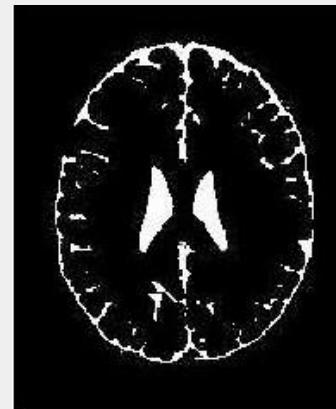
Gray matter



White matter



CSF

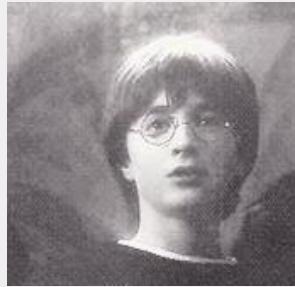


Use [LinearRegression](#) from SciPy to fit regression line between BOLD signal from white matter and BOLD signal from cerebrospinal fluid (CSF).

# Multiple linear regression



# Linear combination



= a



+ b



+ ε

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

# Multiple linear regression

Linear Regression: Single Variable

$$\hat{y} = \beta_0 + \beta_1 x + \epsilon$$

Predicted output      Coefficients      Input      Error

Linear Regression: Multiple Variables

$$\hat{y} = \beta_0 + \underbrace{\beta_1 x_1}_{\text{Coefficients}} + \dots + \underbrace{\beta_p x_p}_{\text{Coefficients}} + \epsilon$$

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

# Matrix notation

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

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{np} & \cdots & X_{np} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Observed Data

Design matrix

Model parameters

Residuals

# Linear algebra magic

Examples:  $m = 4$ .

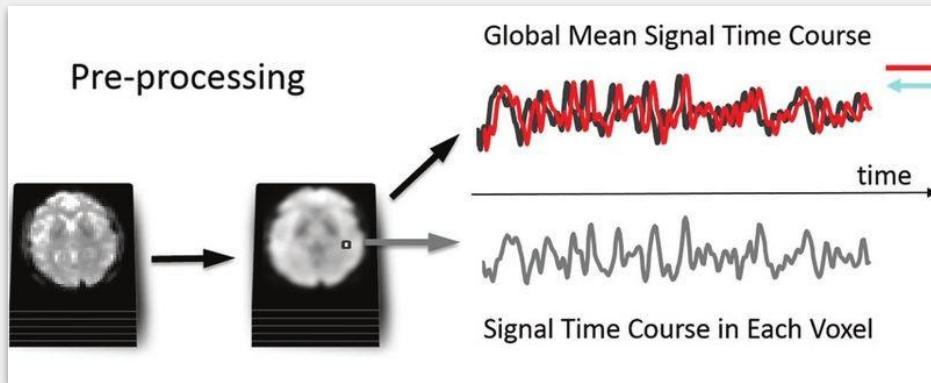
| $x_0$ | Size (feet <sup>2</sup> ) | Number of bedrooms | Number of floors | Age of home (years) | Price (\$1000) |
|-------|---------------------------|--------------------|------------------|---------------------|----------------|
| 1     | 2104                      | 5                  | 1                | 45                  | 460            |
| 1     | 1416                      | 3                  | 2                | 40                  | 232            |
| 1     | 1534                      | 3                  | 2                | 30                  | 315            |
| 1     | 852                       | 2                  | 1                | 36                  | 178            |

$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix}$   $m \times (n+1)$

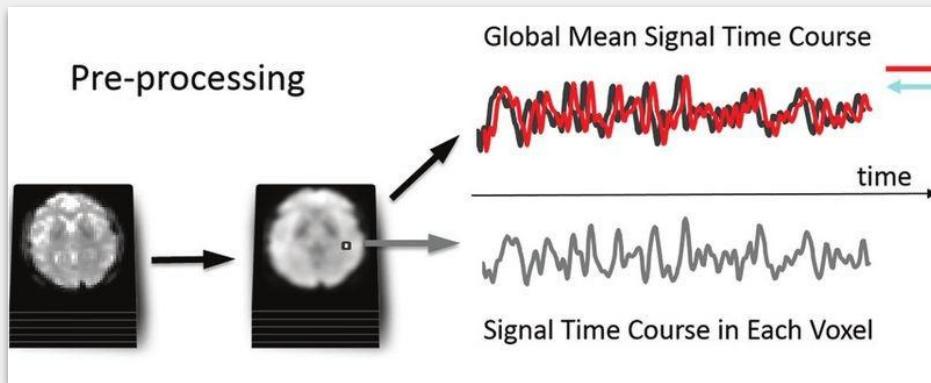
$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$   $m$ -dimensional vector

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

# Challenge 2



# Challenge 2



www.nature.com/scientificreports/

**SCIENTIFIC REPORTS**  
nature research

**OPEN**

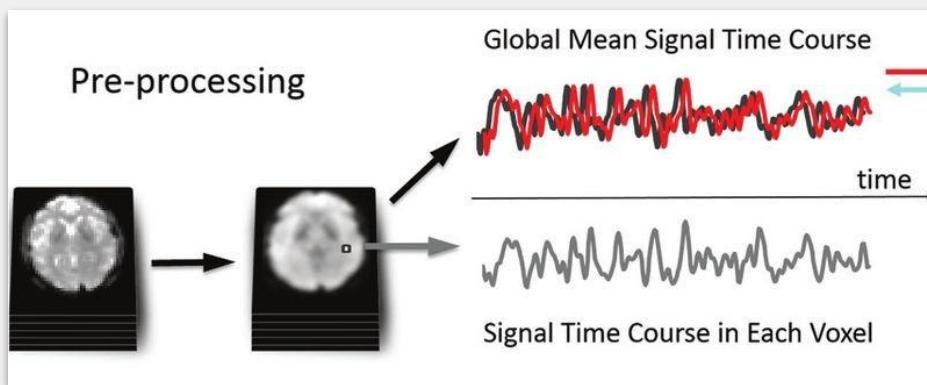
**Topography and behavioral relevance of the global signal in the human brain**

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

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

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**  
nature research

**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<sup>1</sup>, Taylor Bolt<sup>2</sup>, Danilo Bzdok<sup>3,4,5</sup>, Jason S. Nomi<sup>6</sup>, B. T. Thomas Yeo<sup>1</sup>, R. Nathan Spreng<sup>7,8</sup> & Lucina Q. Uddin<sup>6,9</sup>

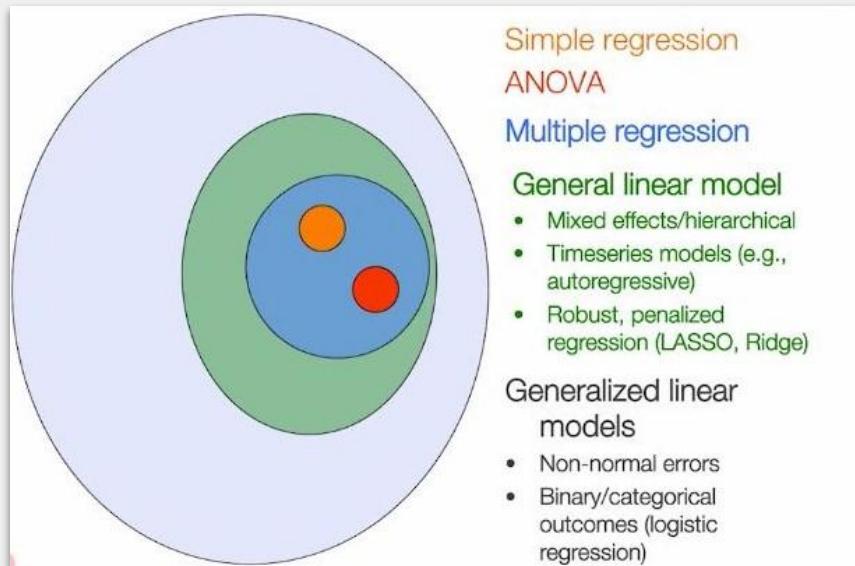
The global signal in resting-state functional MRI data is considered to be dominated by physiological noise and artifacts, yet a growing literature suggests that it also carries information about widespread neural activity. The biological relevance of the global signal remains poorly understood. Applying 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?

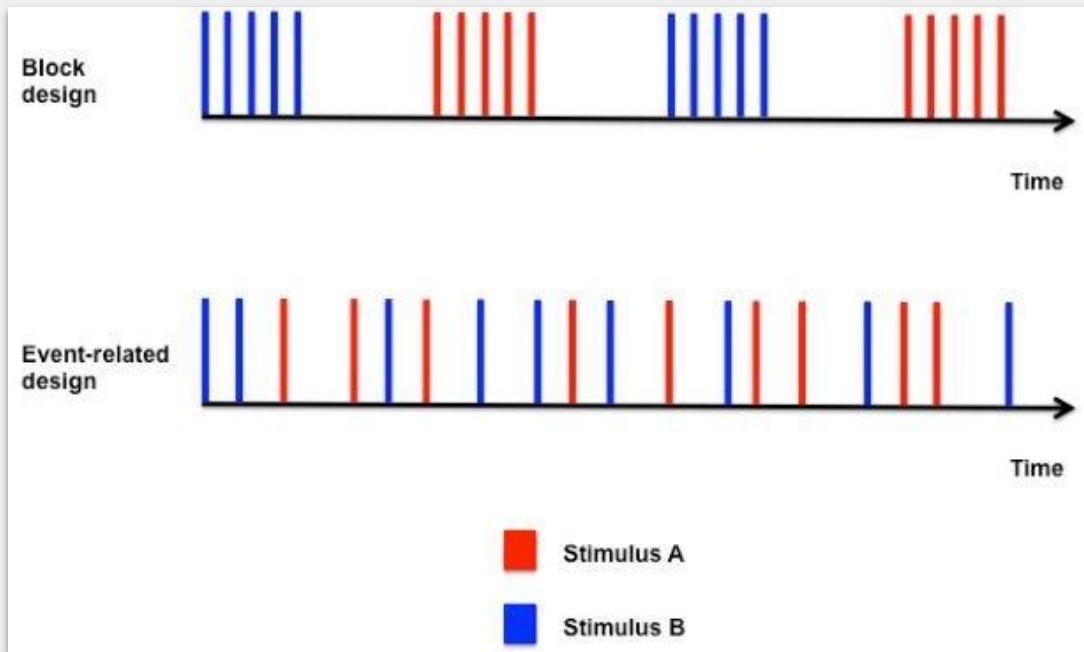


# Generalized Linear Model

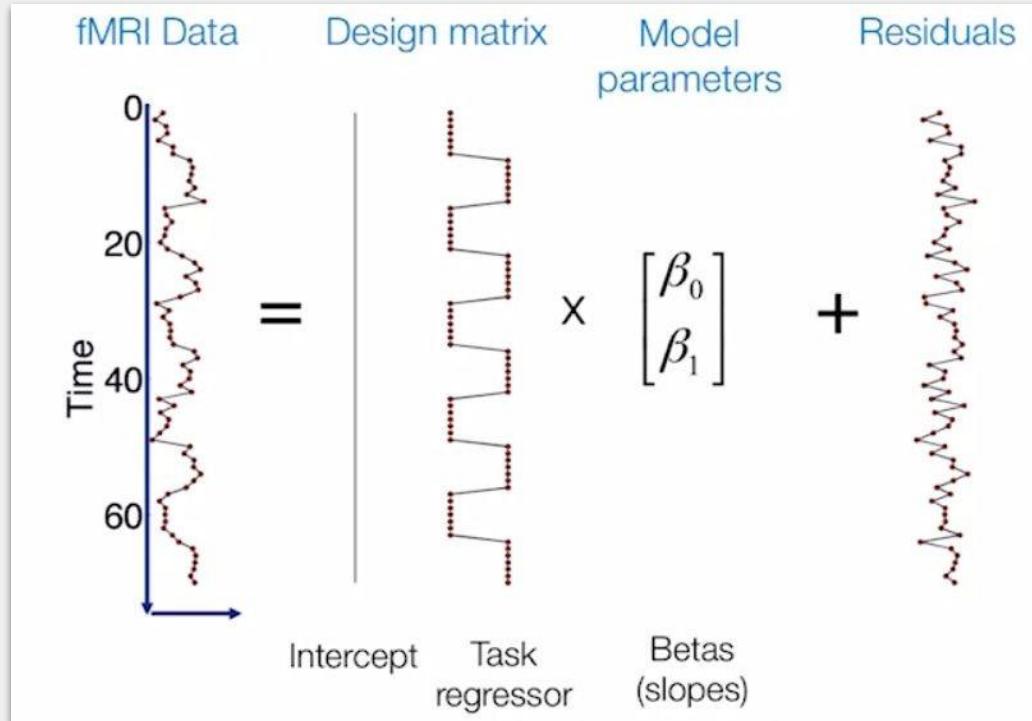


The general linear model (GLM) approach treats the fMRI data as a linear combination of model functions, predictors, plus noise, or error.

# Task designs



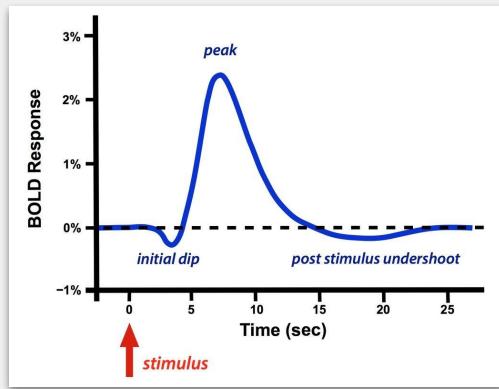
# First level GLM



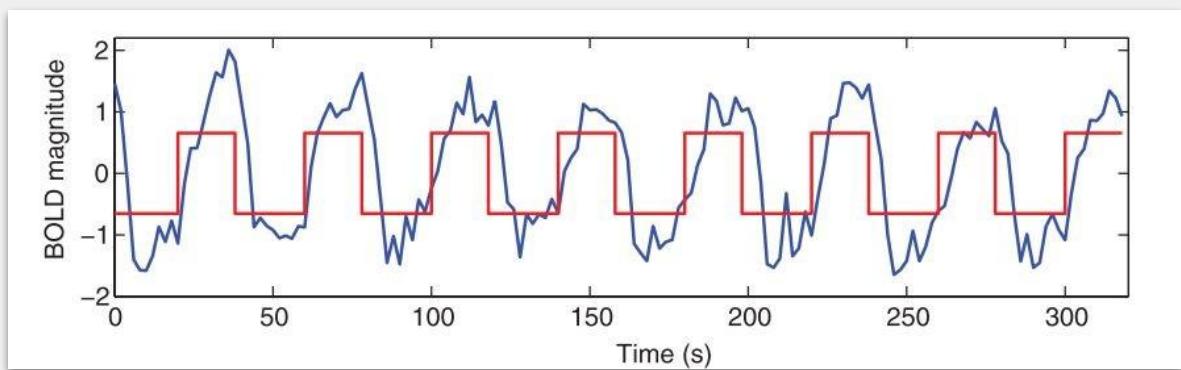
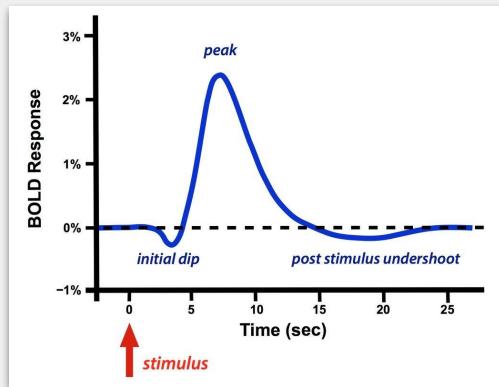
# What we have missed?



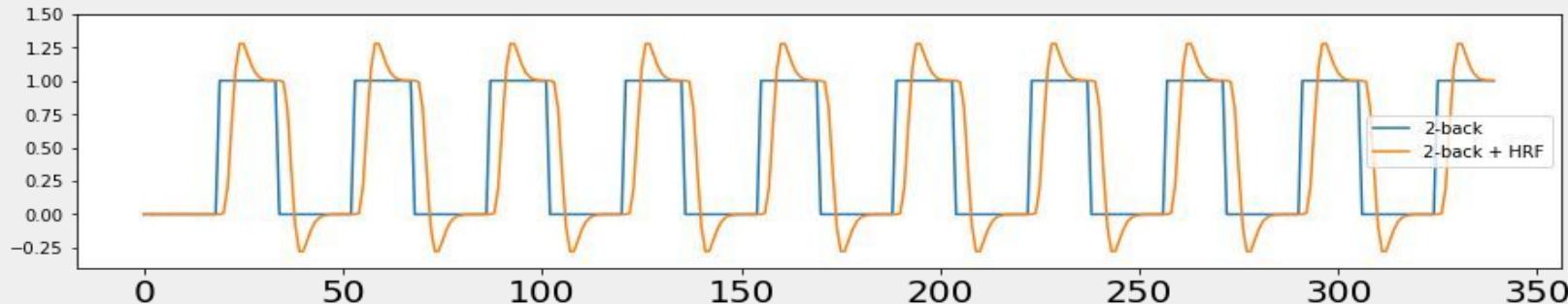
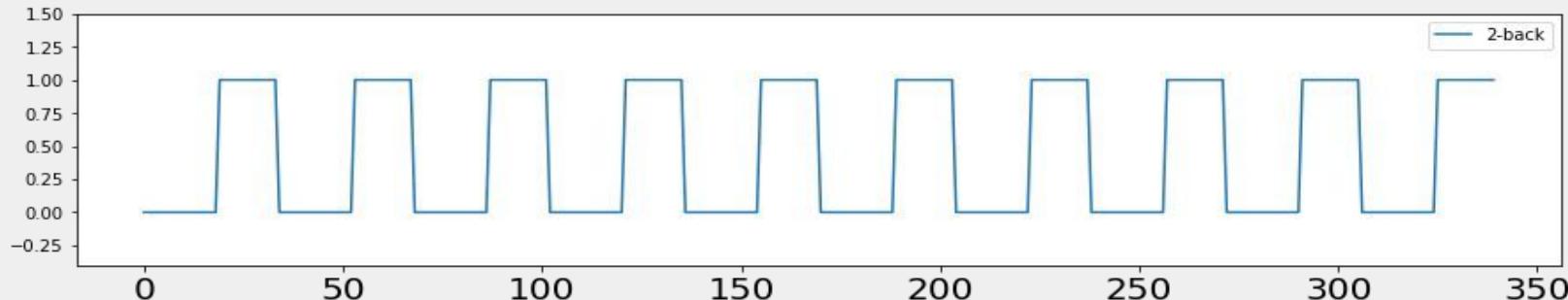
# Haemodynamic delay!



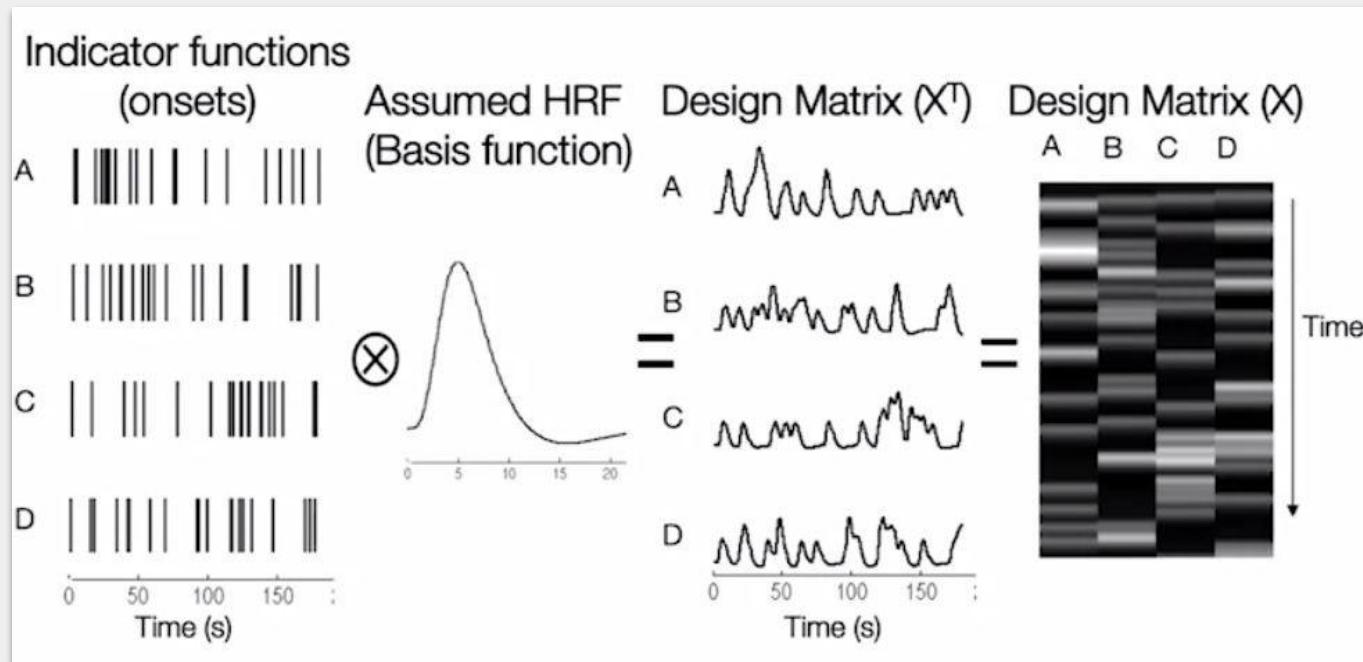
# Haemodynamic delay!



# Convolution



# Building design matrix from events data



# Homework

## 1. GitHub Classroom

Linear and multiple linear regression

Deadline: 06-12-2020

Classroom for **GitHub**

Your course assignments on GitHub



Next



## General Linear Model 2