



# **Covariance and Correlation: What is it and why it's important**

Charles Marchini, MBM Trainee and PhD Candidate in Bioengineering



## Expected value: taking arithmetic average

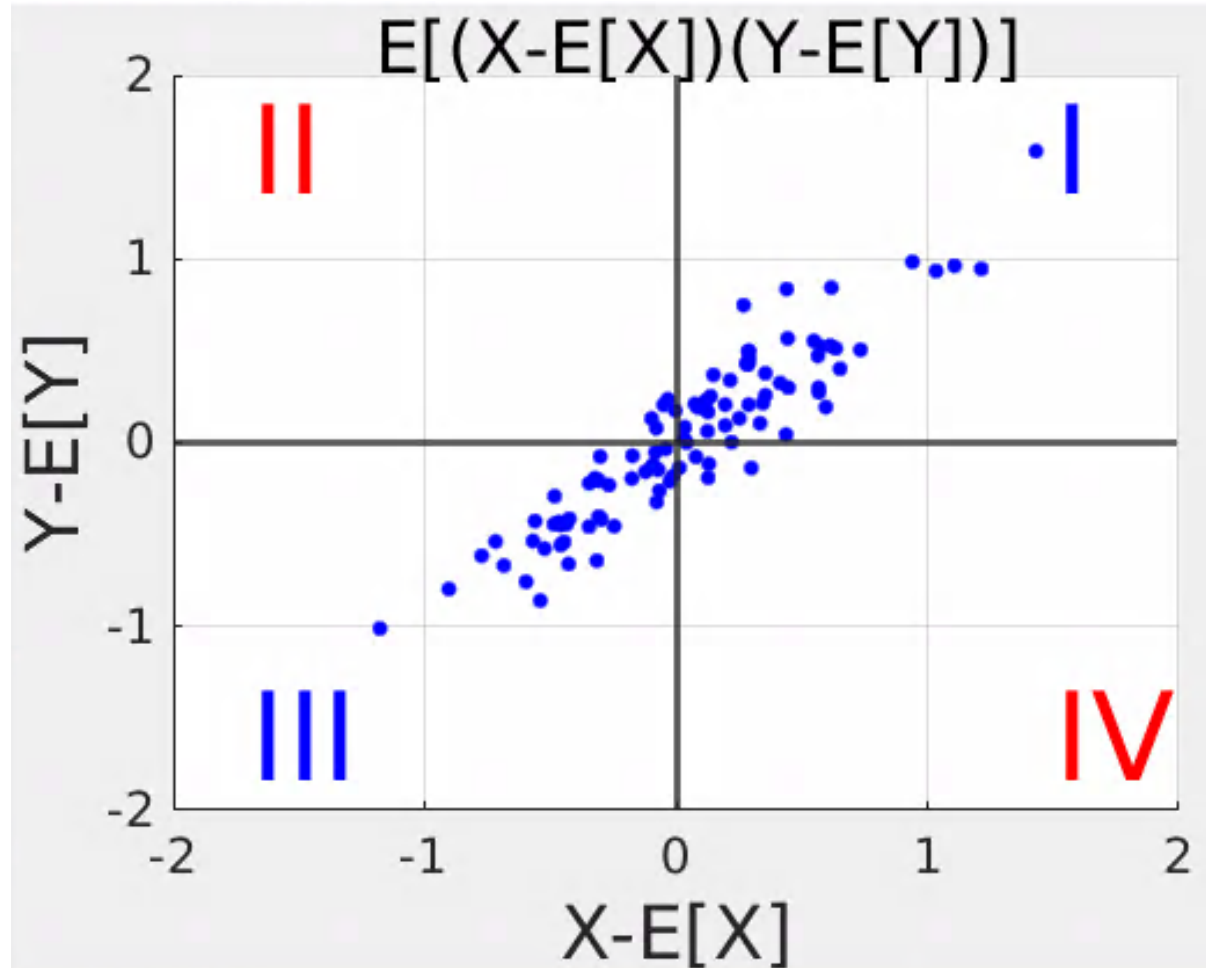
- If have three data points for  $X$ : 5,9,7
- Expected value of  $X = E[X] = \frac{5+9+7}{3} = 7$
- If have  $n$  data points:  $X_1, X_2, \dots, X_n$
- $E[X] = \frac{X_1+X_2+\dots+X_n}{n}$



# Examples of Correlations

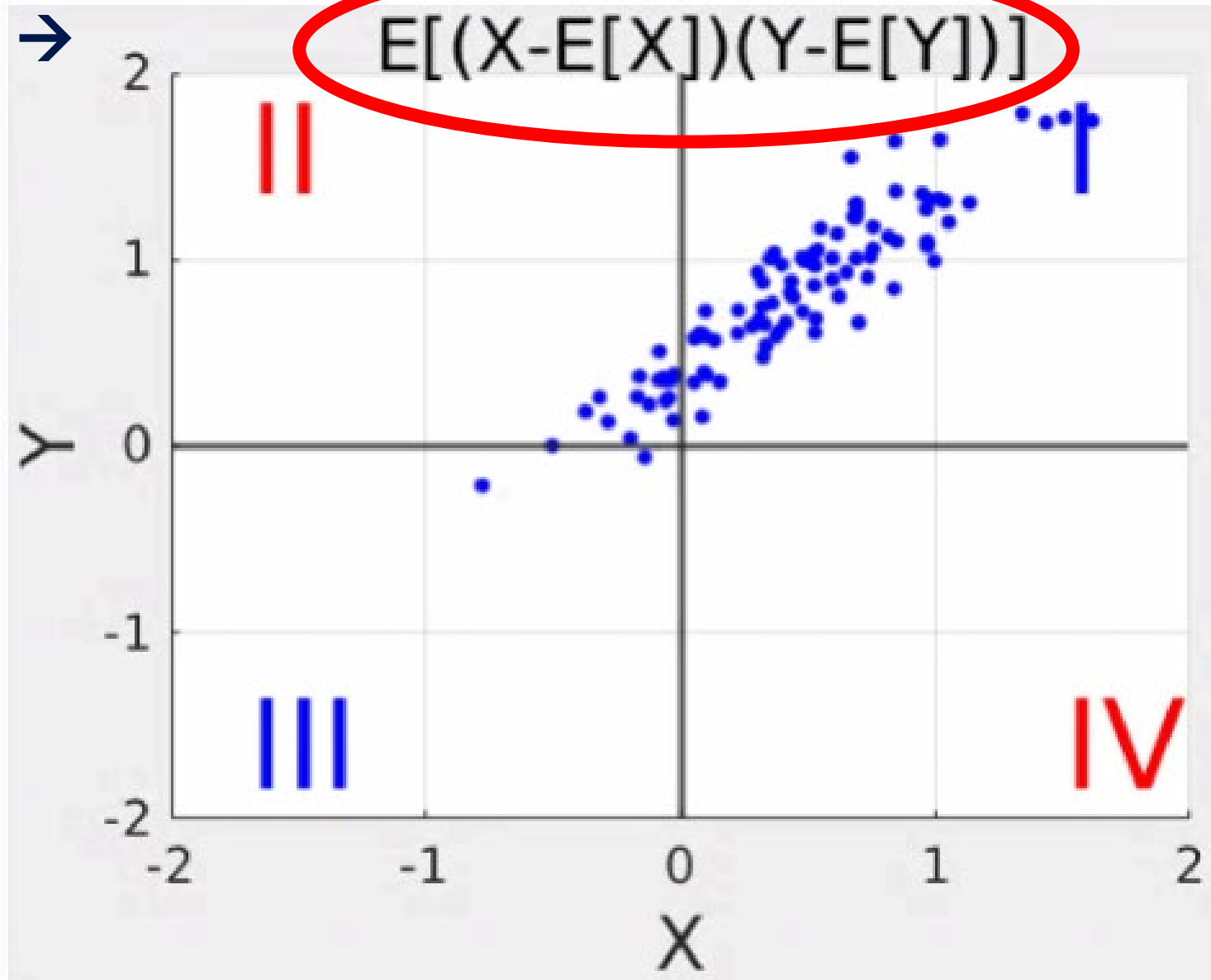
Plot X and Y for data

- Example 1:
  - X: Height
  - Y: Weight
- Example 2:
  - X: Cognition
  - Y: Physical Fitness



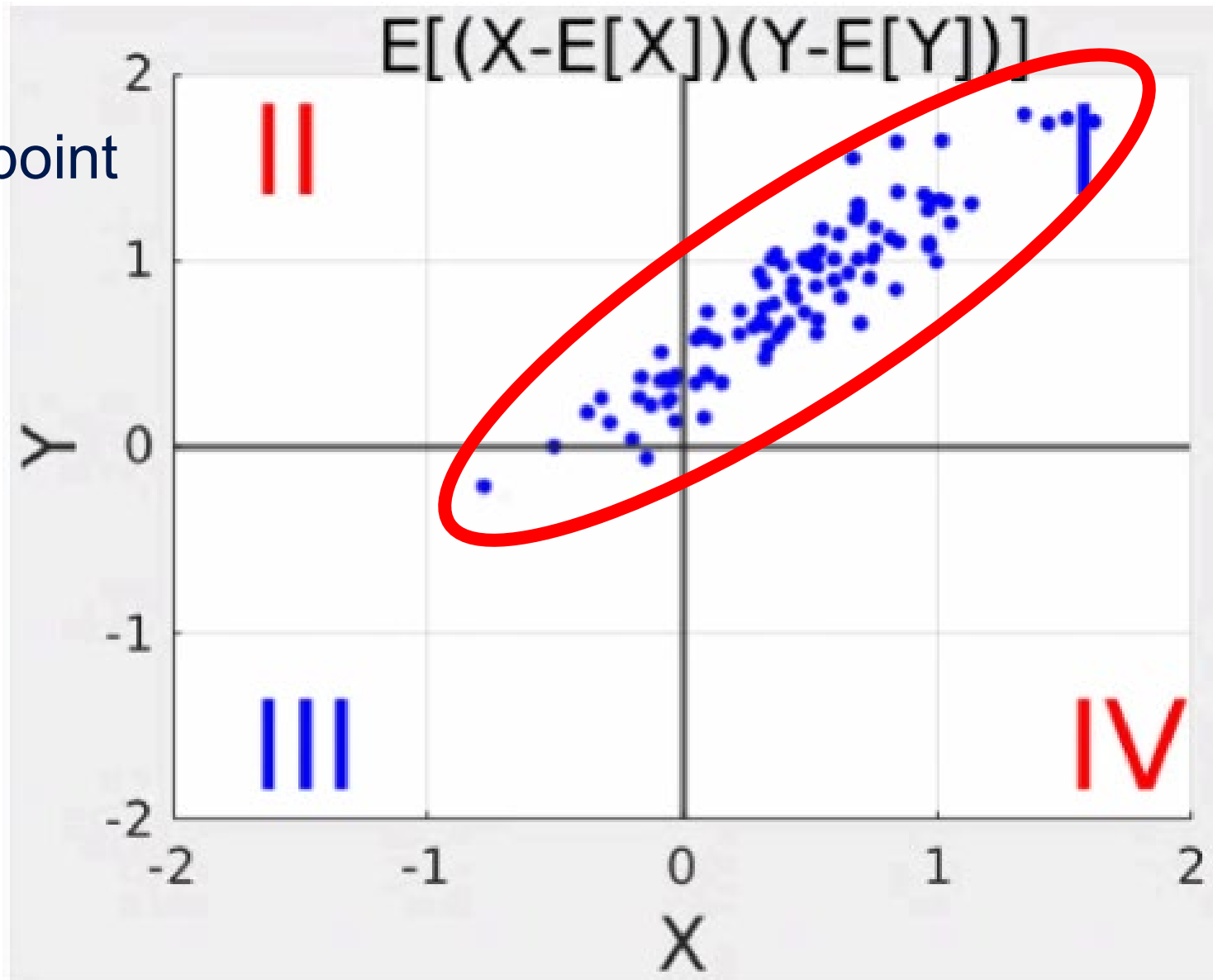


Covariance →



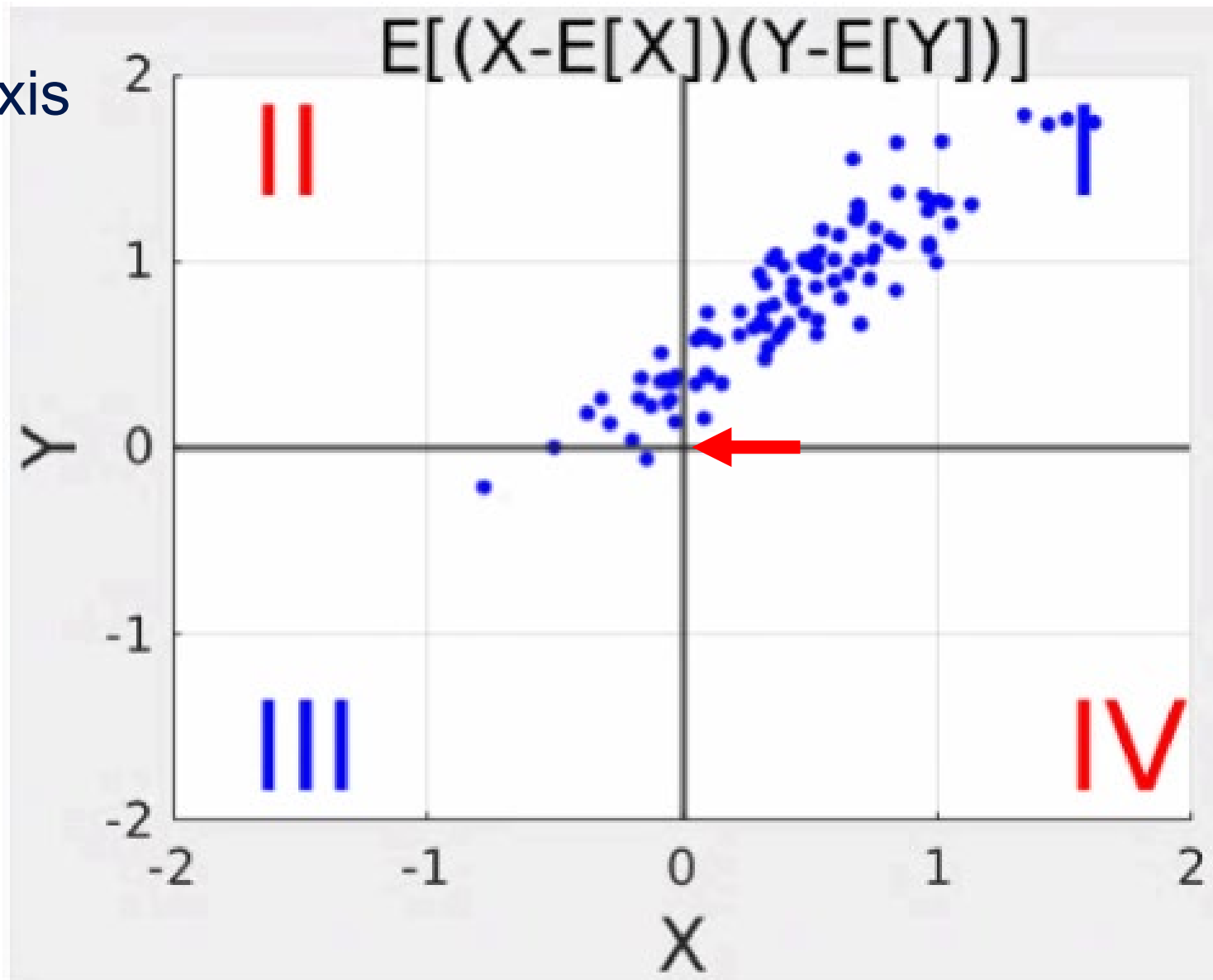


Plot each data point



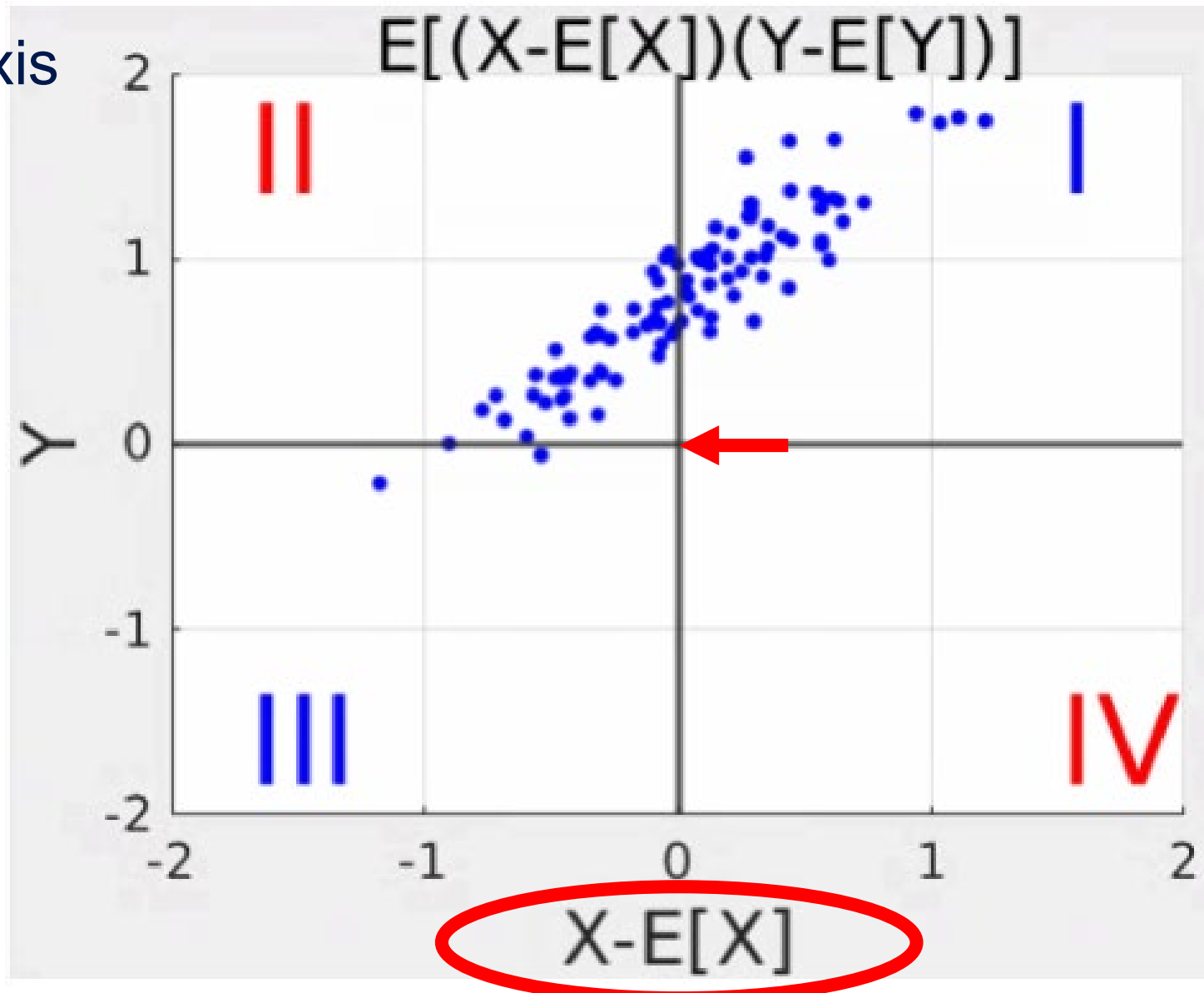


Shift along X axis



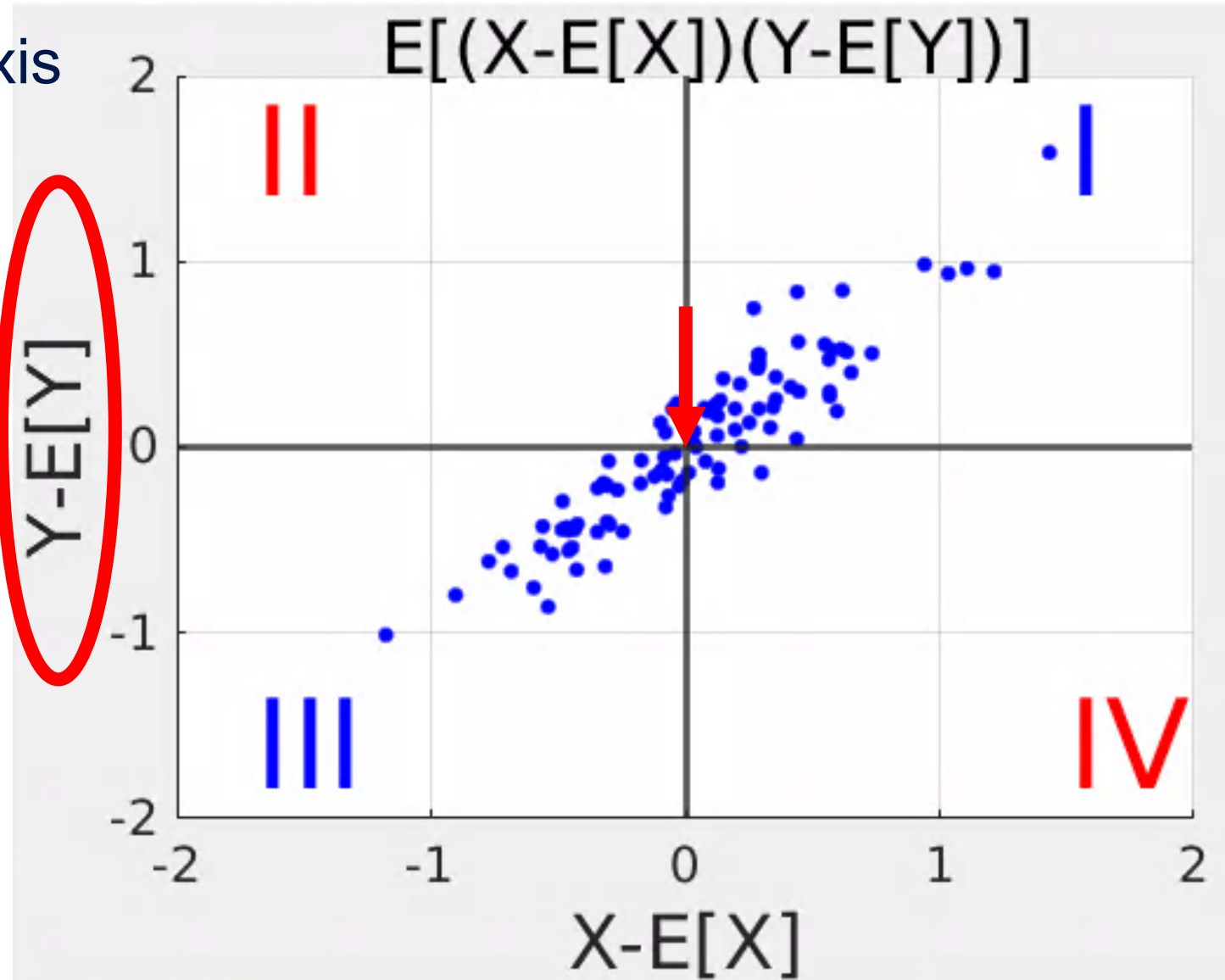


Shift along X axis





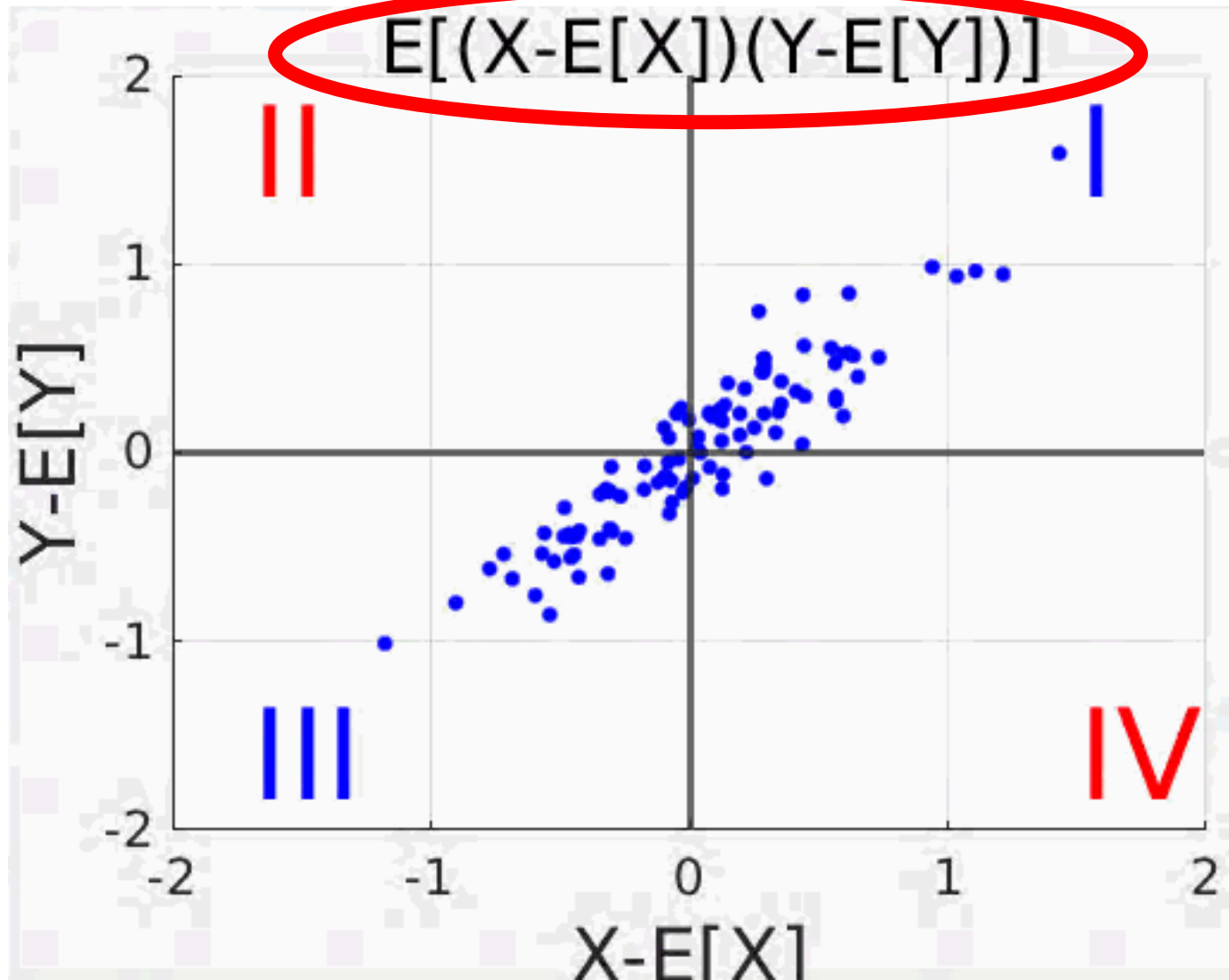
Shift along Y axis







$$\frac{(X_1 - E[X])(Y_1 - E[Y]) + (X_2 - E[X])(Y_2 - E[Y]) + \dots + (X_n - E[X])(Y_n - E[Y])}{n}$$

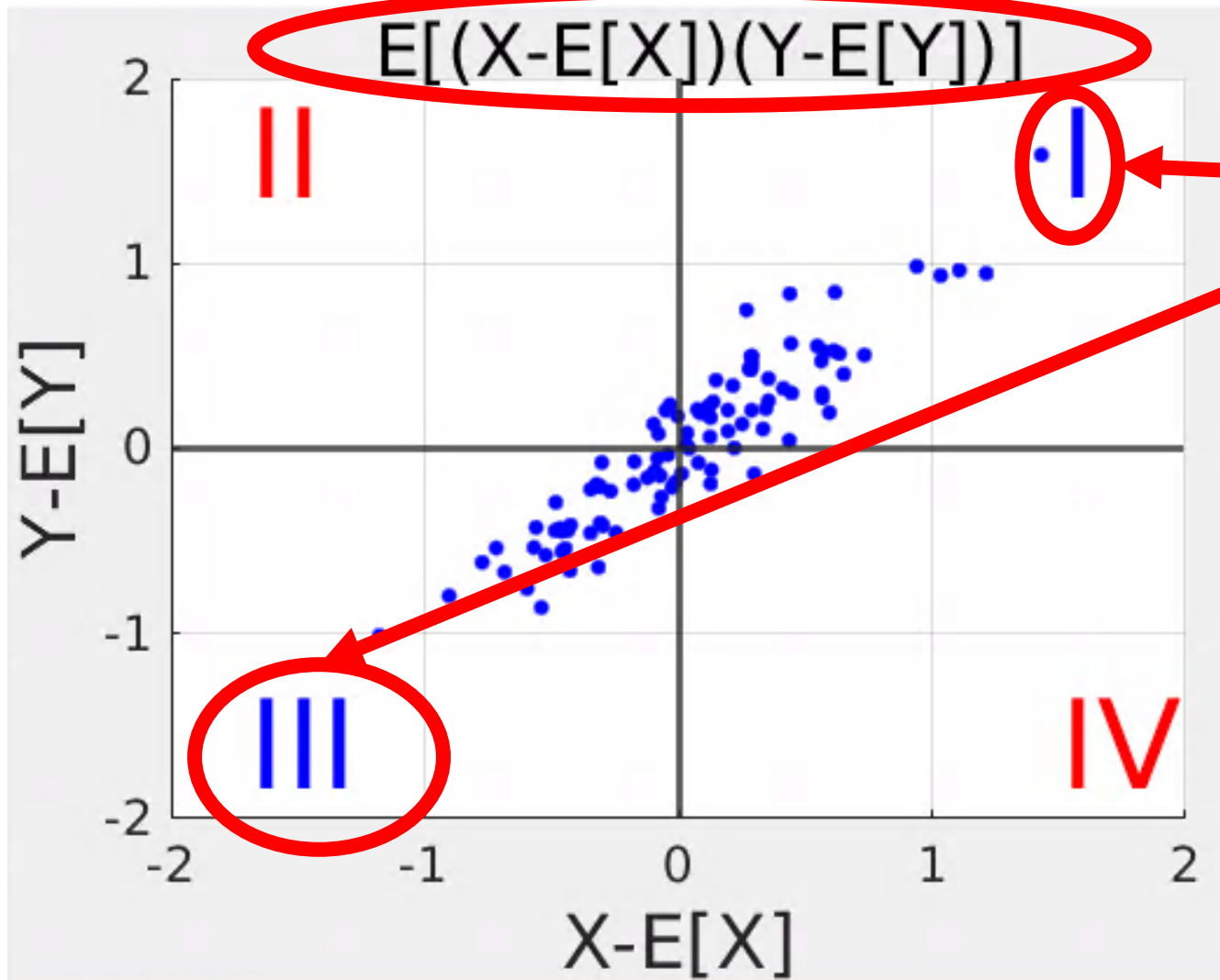


Add up areas according to the expression



→  $(X_1 - E[X])(Y_1 - E[Y]) + (X_2 - E[X])(Y_2 - E[Y]) + \dots + (X_n - E[X])(Y_n - E[Y])$

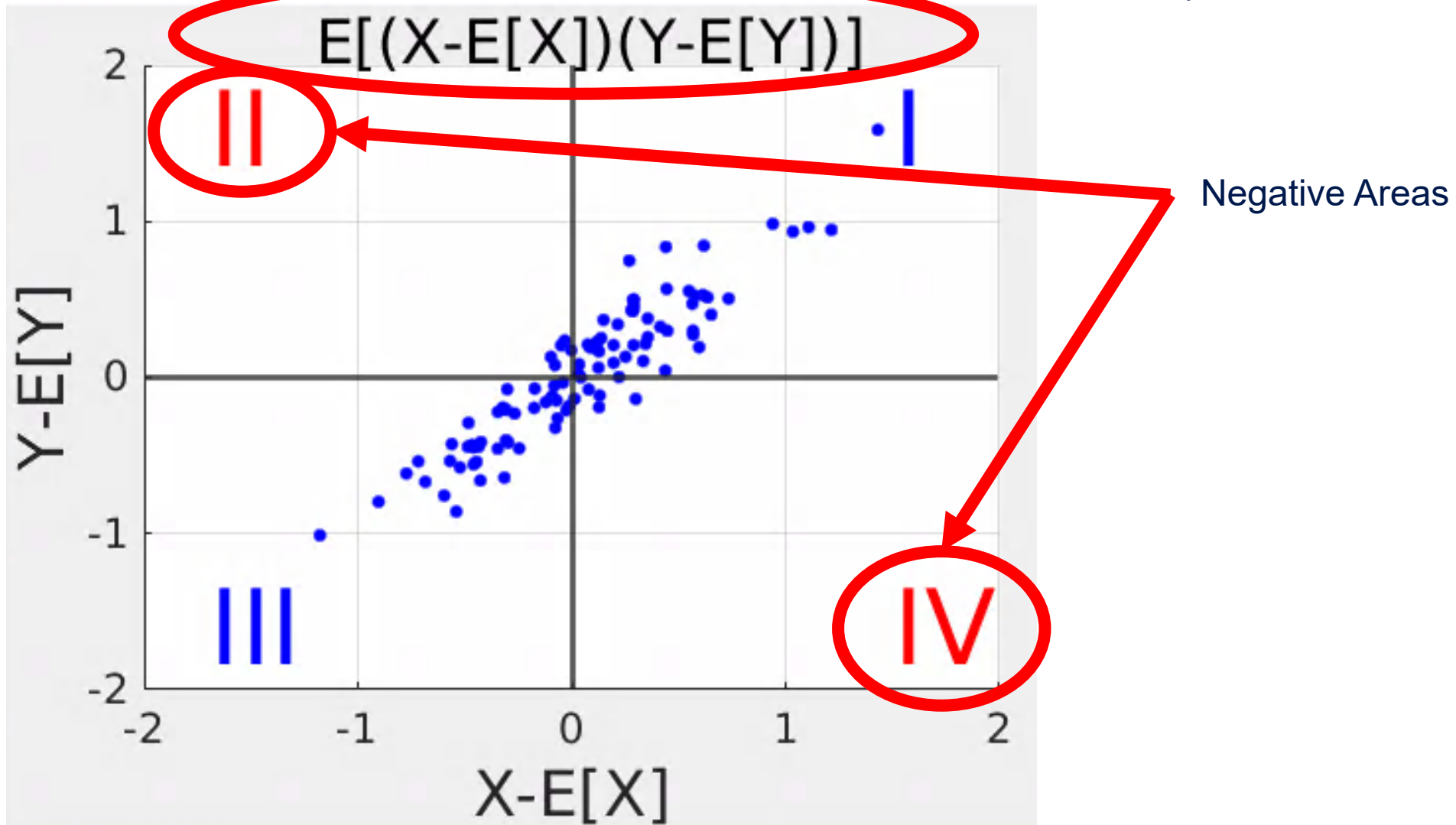
$n$



Positive Areas



$$\rightarrow \frac{(X_1 - E[X])(Y_1 - E[Y]) + (X_2 - E[X])(Y_2 - E[Y]) + \dots + (X_n - E[X])(Y_n - E[Y])}{n}$$



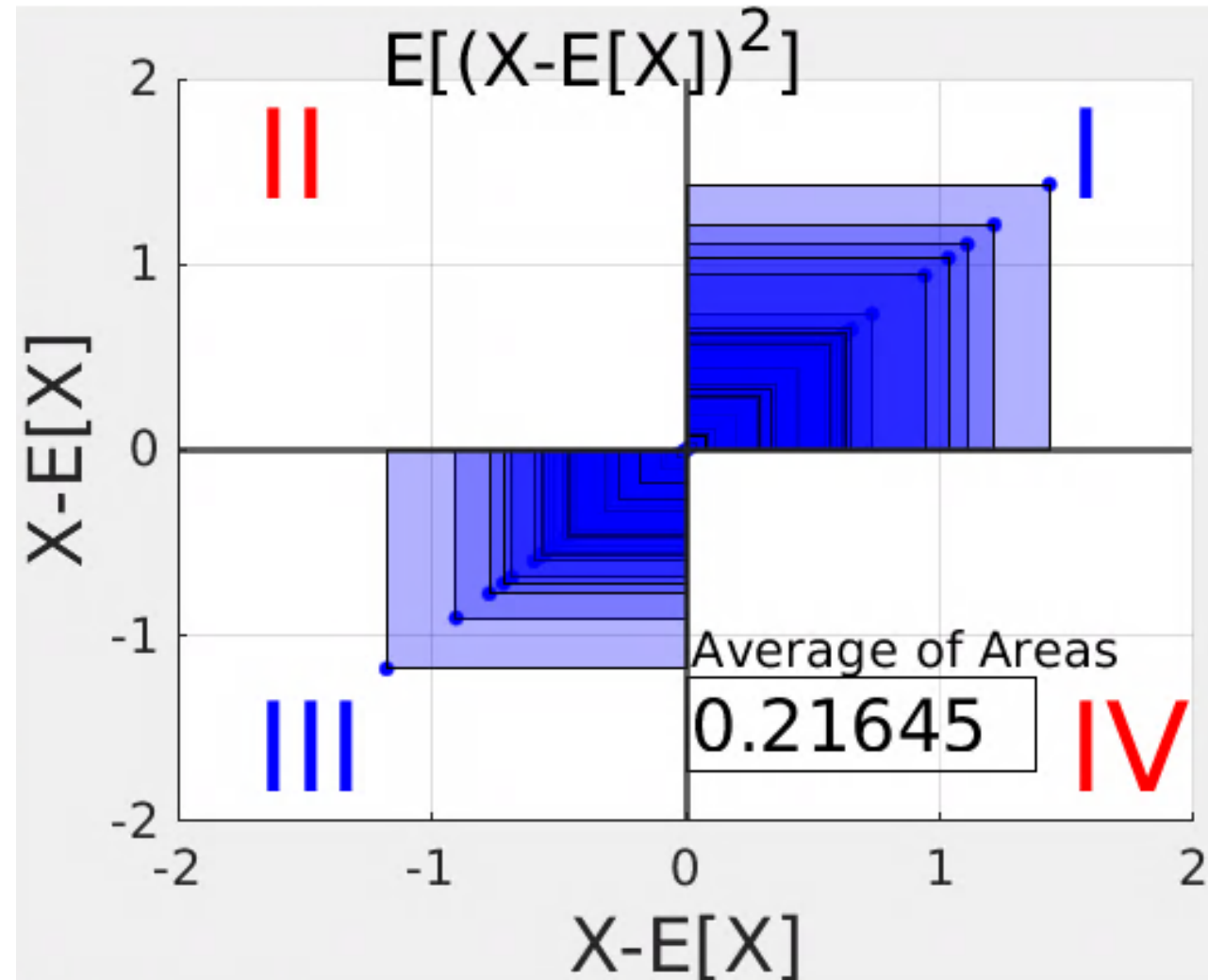


## Correlation = $r$

- Covariance: how  $X$  and  $Y$  co-vary
- Variance,  $\sigma_X^2 = E[(X - E[X])^2]$ , how  $X$  and  $Y$  vary with themselves

$$r = \frac{E[(X - E[X])(Y - E[Y])]}{\sigma_X \sigma_Y}$$

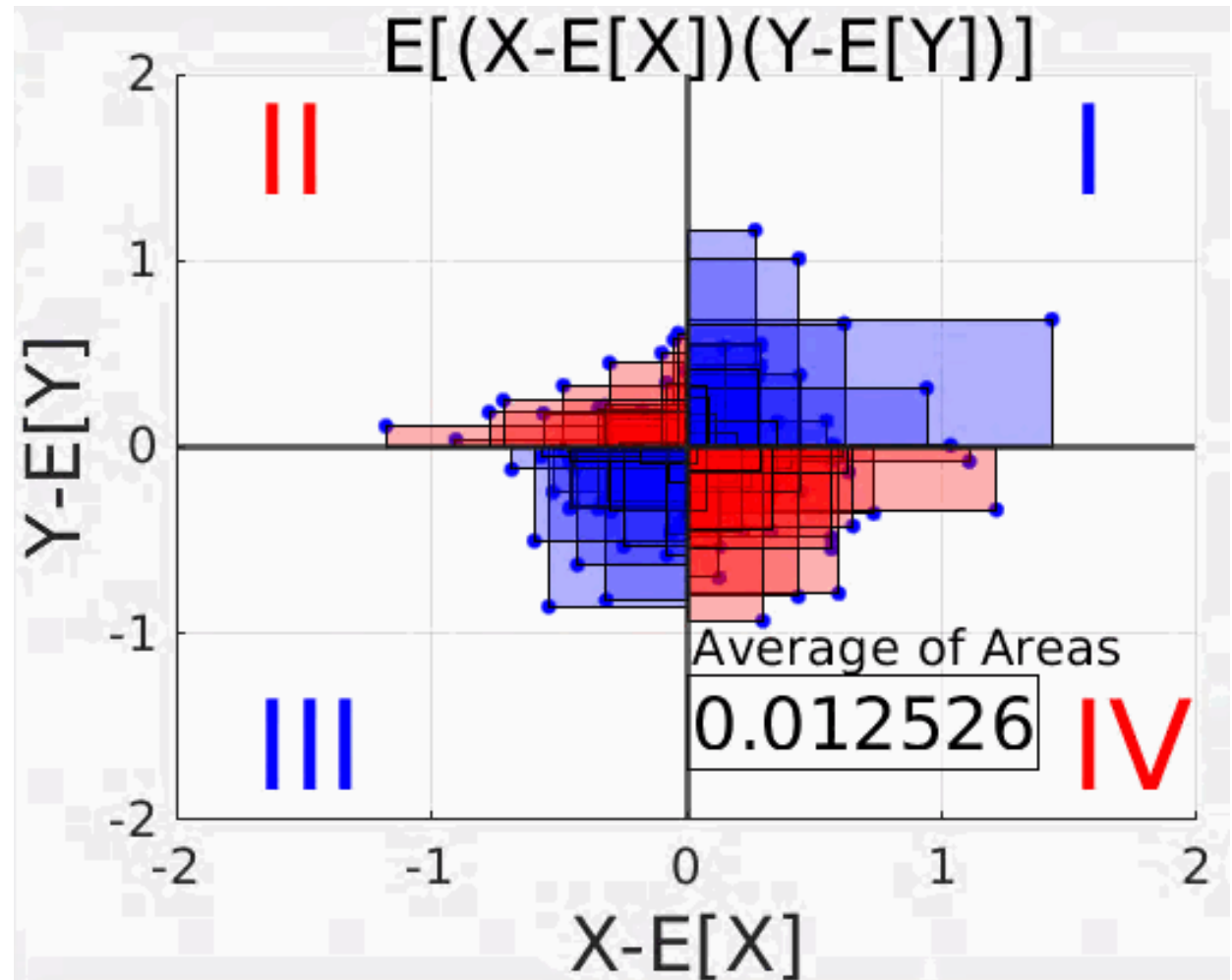
$$-1 \leq r \leq 1$$





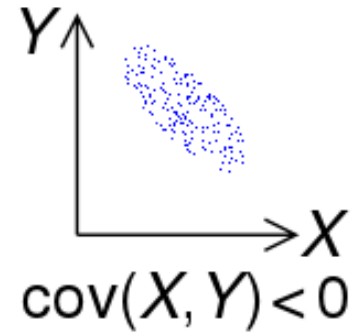
# When there is no correlation

- Example:
  - X: roll of die 1
  - Y: roll of die 2

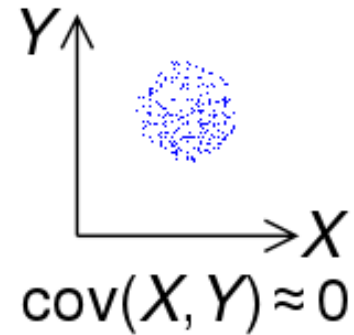




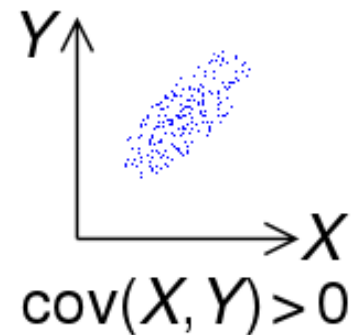
## Negative Covariance



## Zero Covariance



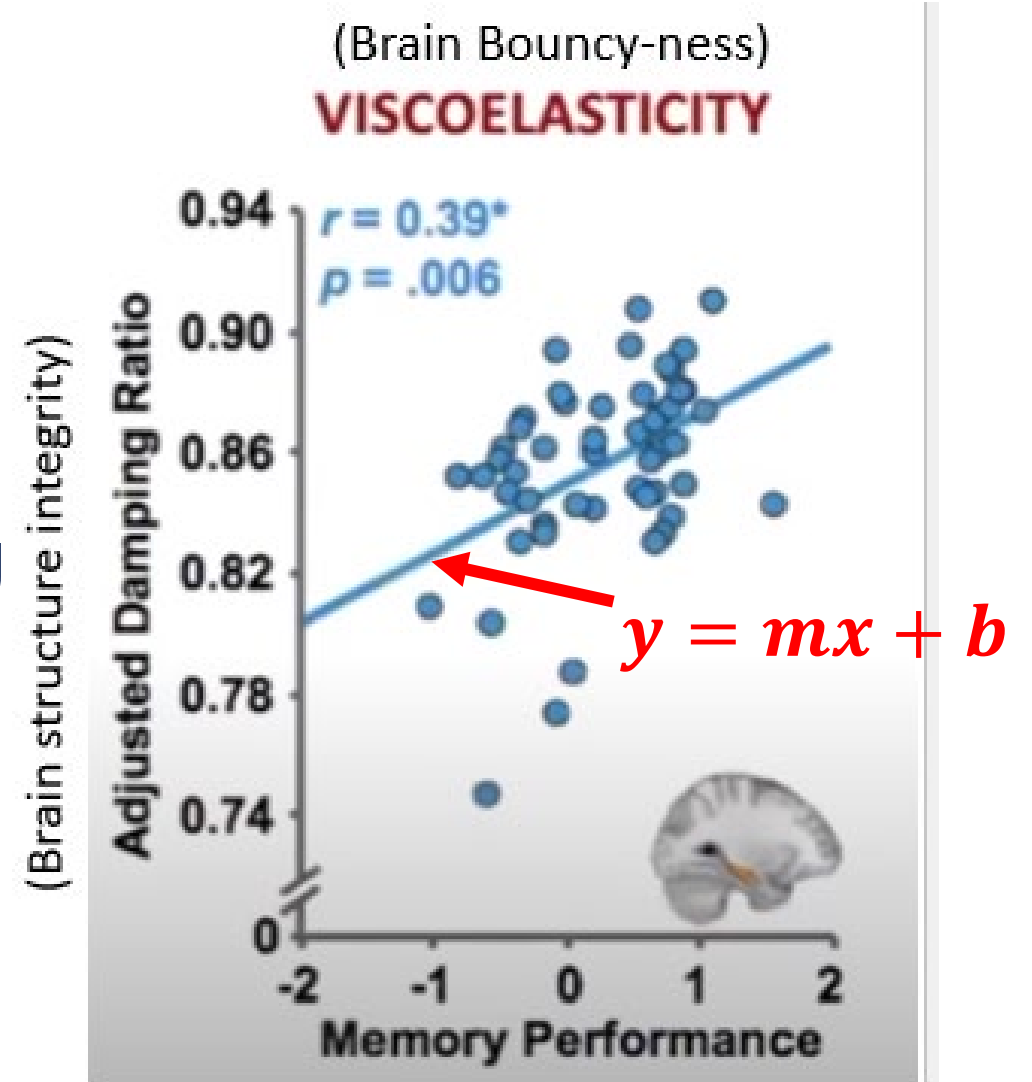
## Positive Covariance





# Other things

- Correlation is not causation
  - What is causing what?
  - Both have common cause?
- p-values = probability of falsely detecting correlation
  - practical certainty of a correlation?

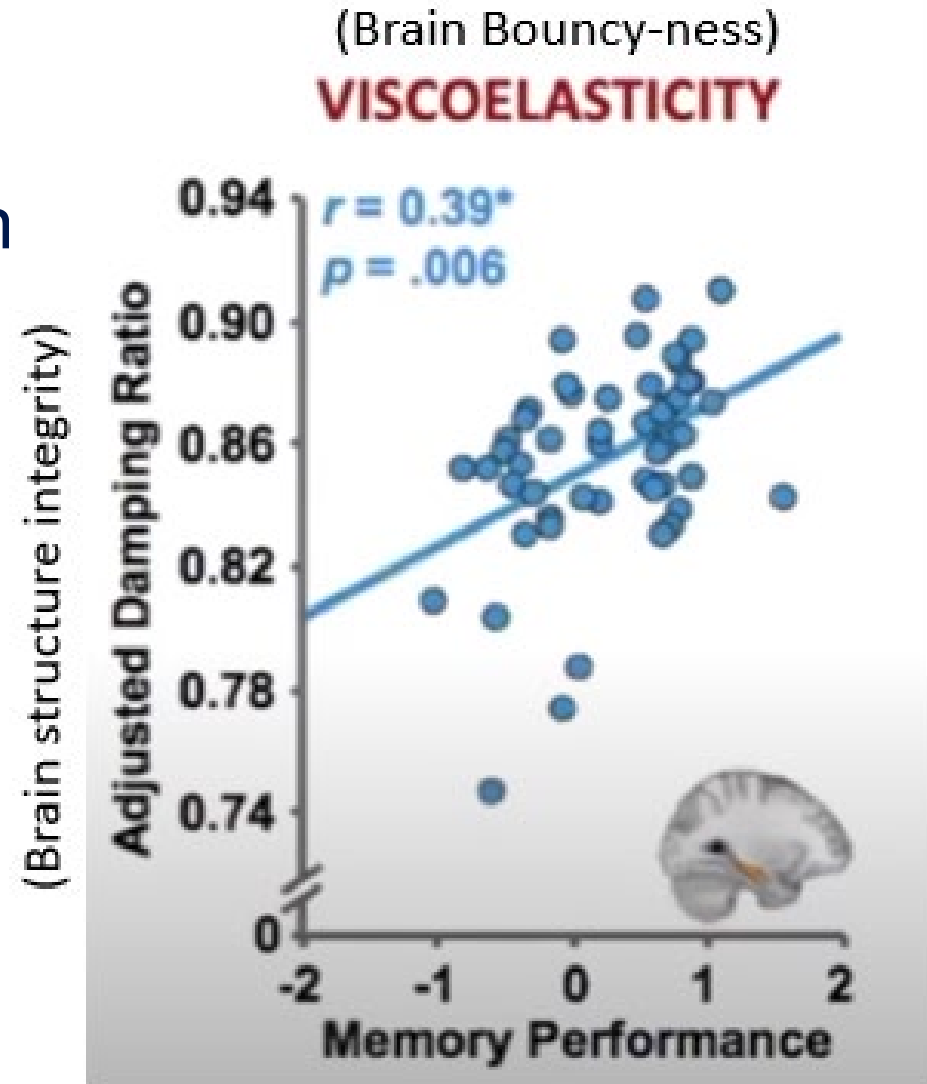






## Importance

- What two variables to find correlation
- Real-world applications
  - Understanding the brain
  - Machine learning (artificial intelligence)
  - Tumor Diagnosis
  - Many more







## End – Thank you!

This lesson is supported by Understanding the Brain: Training the Next Generation of Researchers in Engineering and Deciphering of Miniature Brain Machinery” – a Research Traineeship program supported by the National Science Foundation under grant 1735252.

