LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_n X_n + \varepsilon$$

LET'S PARSE THIS

LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Y CAN BE EXPLAINED BY

LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Y CAN BE EXPLAINED BY

A BASE VALUE

LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Y CAN BE EXPLAINED BY
A BASE VALUE

A LINEAR COMBINATION OF
THE INDEPENDENT VARIABLES

LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Y CAN BE EXPLAINED BY
A BASE VALUE
A LINEAR COMBINATION OF
THE INDEPENDENT VARIABLES

SOME RANDOM VARIATION
THAT IS SUBJECT TO CHANCE

LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Y CAN BE EXPLAINED BY
A BASE VALUE
A LINEAR COMBINATION OF
THE INDEPENDENT VARIABLES
SOME RANDOM VARIATION
THAT IS SUBJECT TO CHANCE

THE OBJECTIVE OF LINEAR REGRESSION IS TO FIND THE PARAMETERS

LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Y CAN BE EXPLAINED BY
A BASE VALUE
A LINEAR COMBINATION OF THE INDEPENDENT VARIABLES
SOME RANDOM VARIATION THAT IS SUBJECT TO CHANCE

THE OBJECTIVE OF LINEAR REGRESSION IS TO FIND THE PARAMETERS

LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

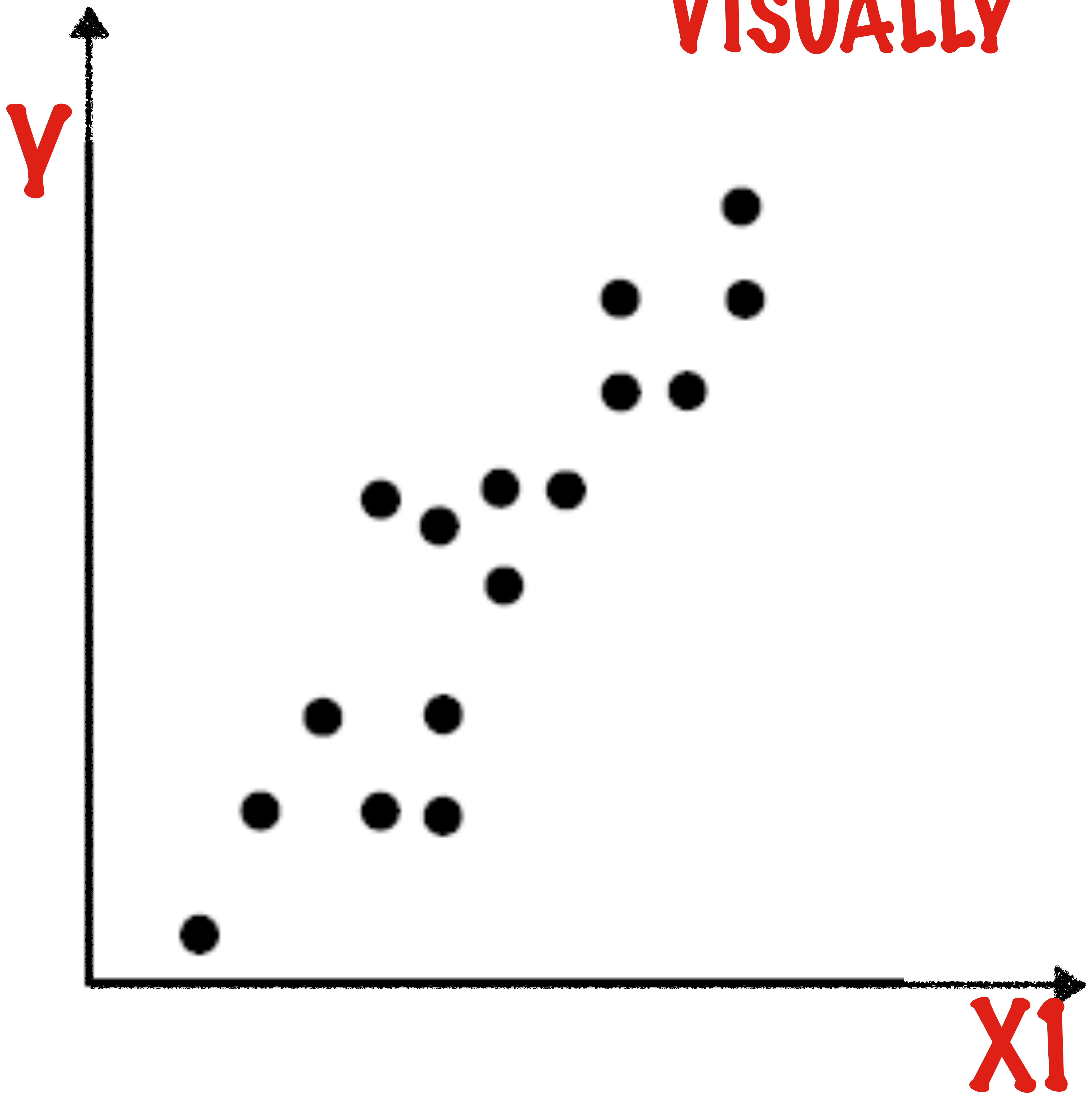
Y CAN BE EXPLAINED BY
A BASE VALUE
A LINEAR COMBINATION OF
THE INDEPENDENT VARIABLES
SOME RANDOM VARIATION
THAT IS SUBJECT TO CHANCE

VISUALLY REGRESSION
CAN BE SEEN AS
CURVE FITTING

LINEAR REGRESSION

VISUALLY

HERE IS SOME PAST DATA FOR X AND Y

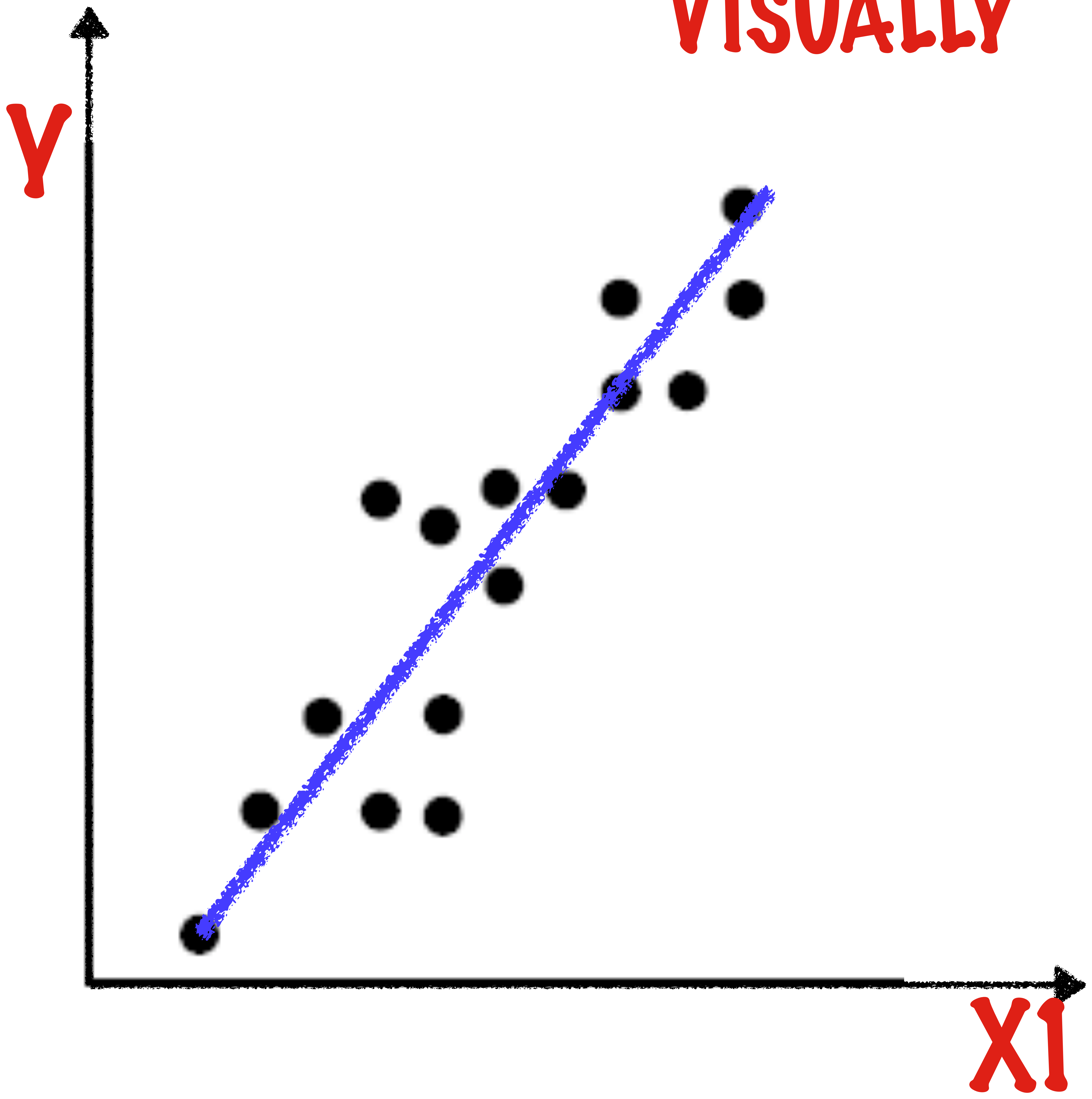


LINEAR REGRESSION

VISUALLY

LINEAR REGRESSION FINDS THE LINE THAT IS THE BEST FIT

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

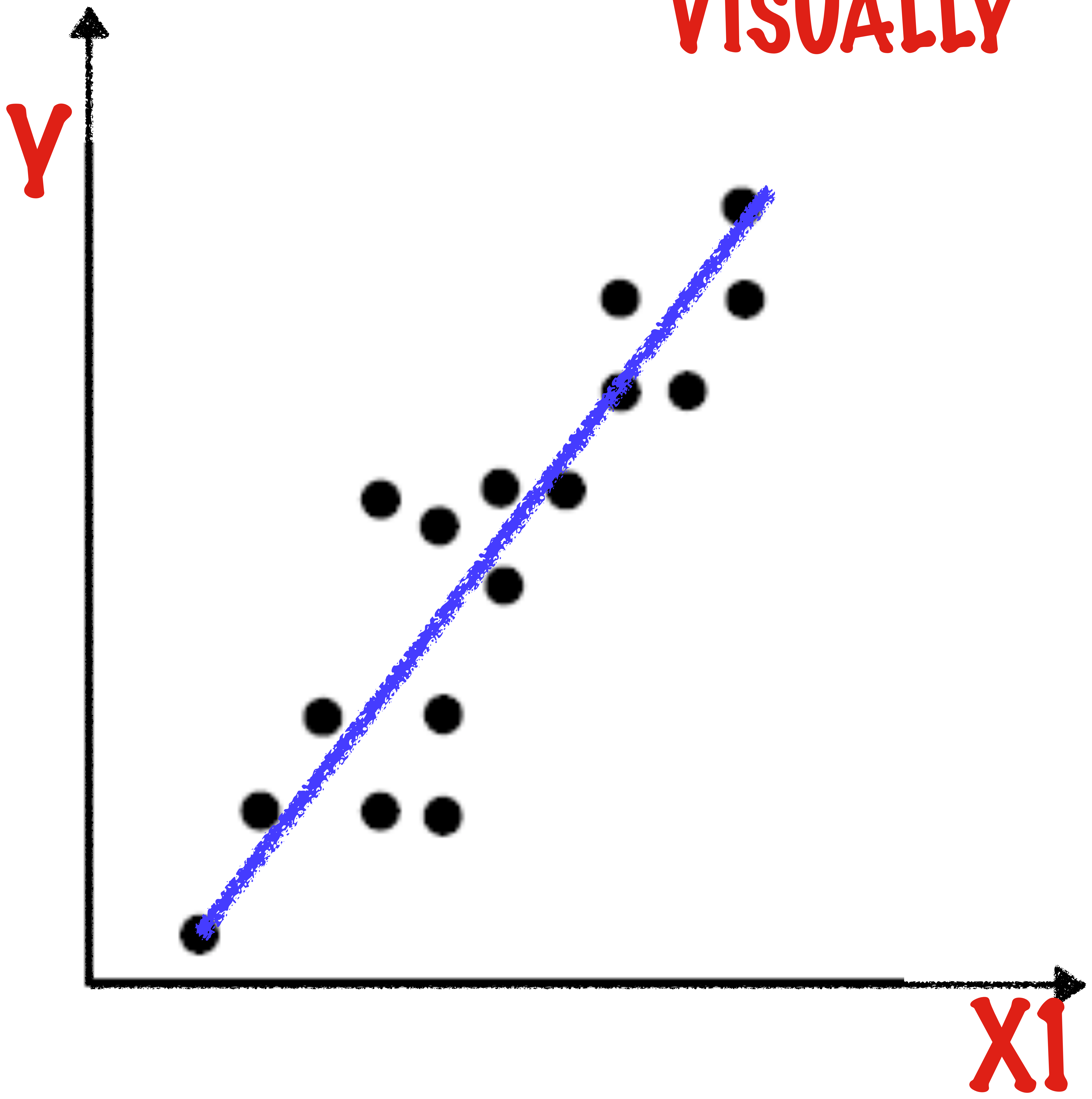


LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

INTERCEPT OF THE LINE

VISUALLY

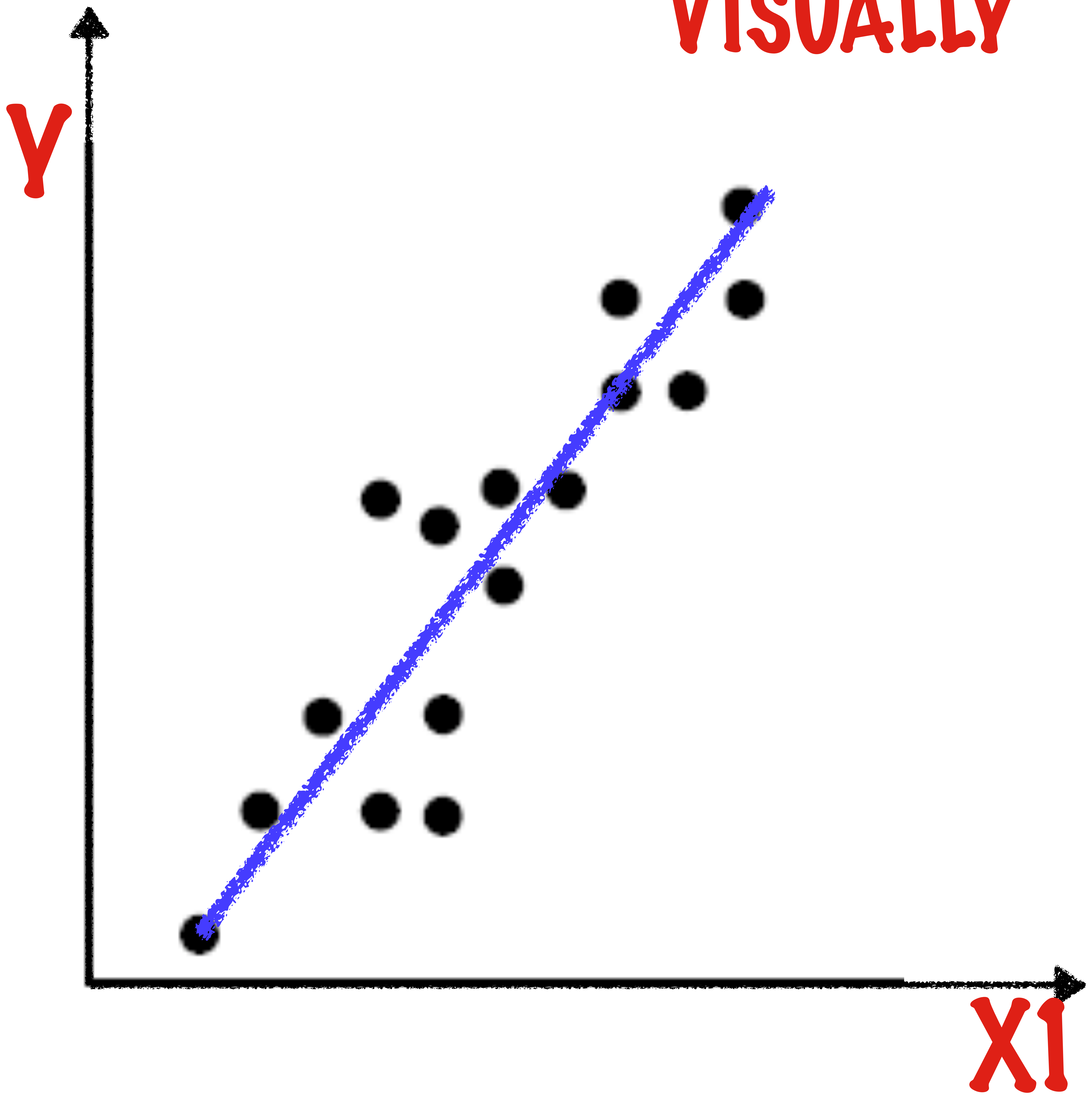


LINEAR REGRESSION

VISUALLY

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

SLOPE OF THE LINE

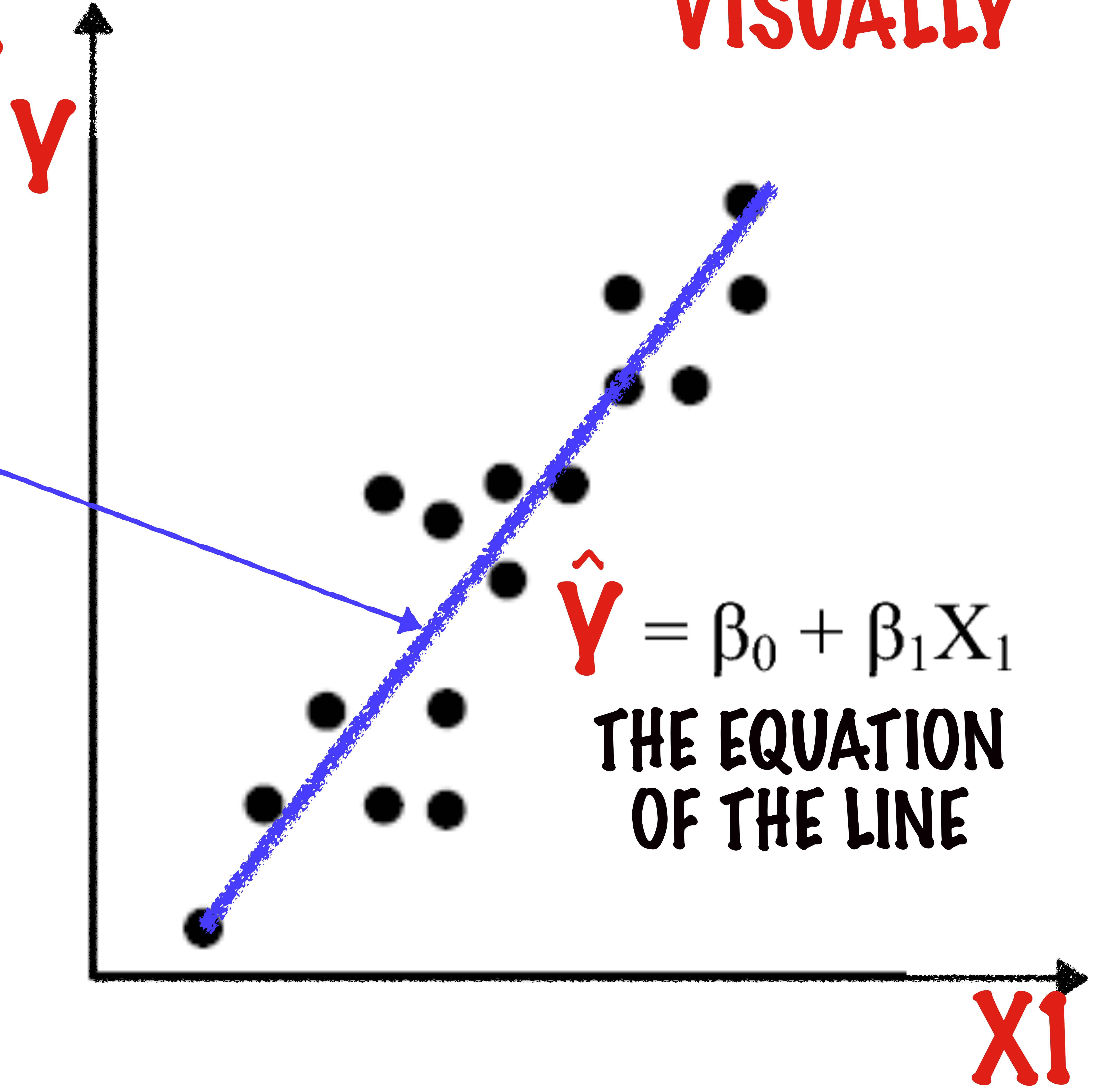


LINEAR REGRESSION VISUALLY

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

THE PREDICTED VALUE OF Y USING THE LINE

$$Y = \hat{Y} + \varepsilon$$



LINEAR REGRESSION

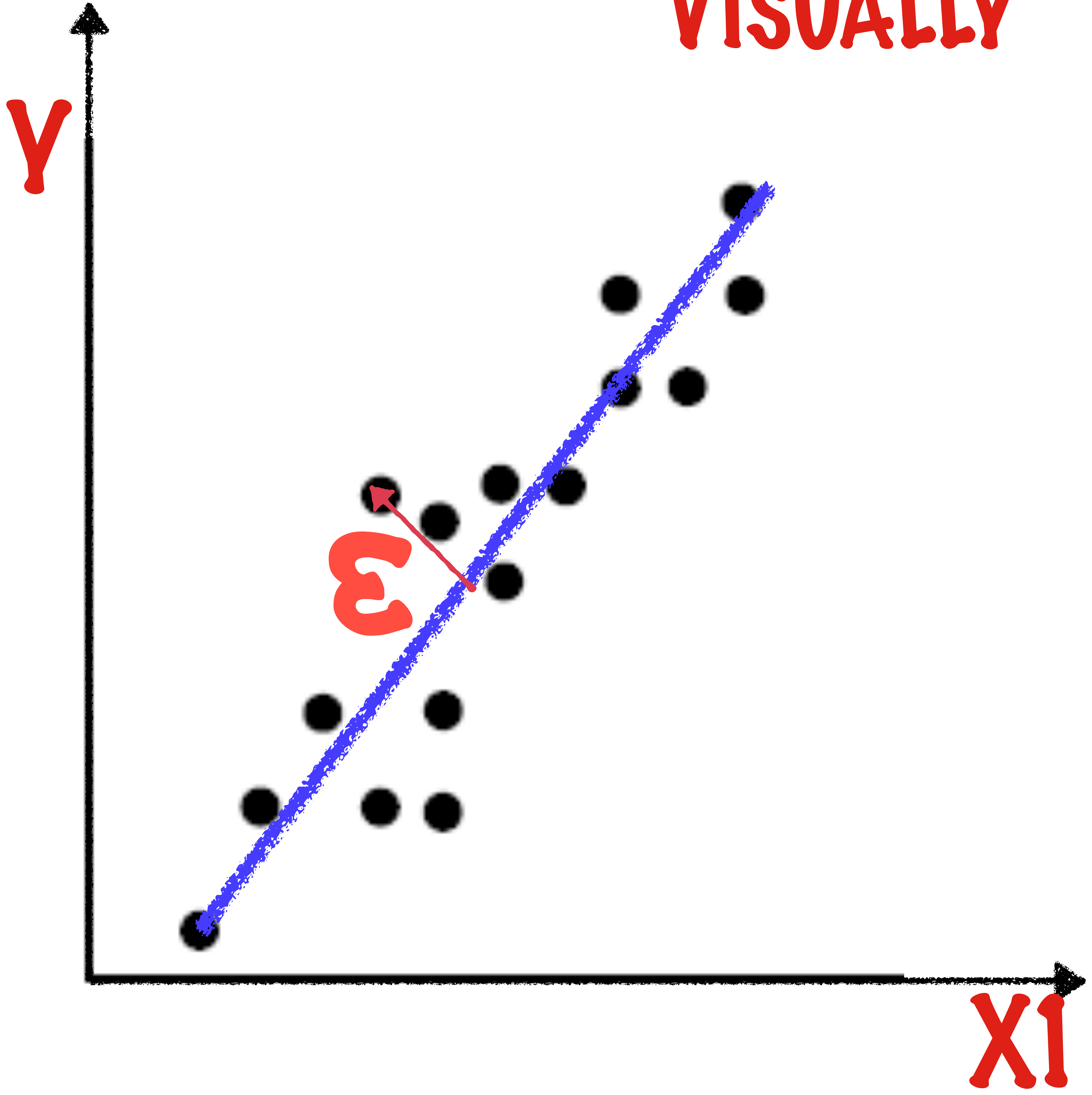
VISUALLY

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

$$Y = \hat{Y} + \varepsilon$$

ERROR

(DISTANCE BETWEEN THE ACTUAL POINT AND THE LINE)



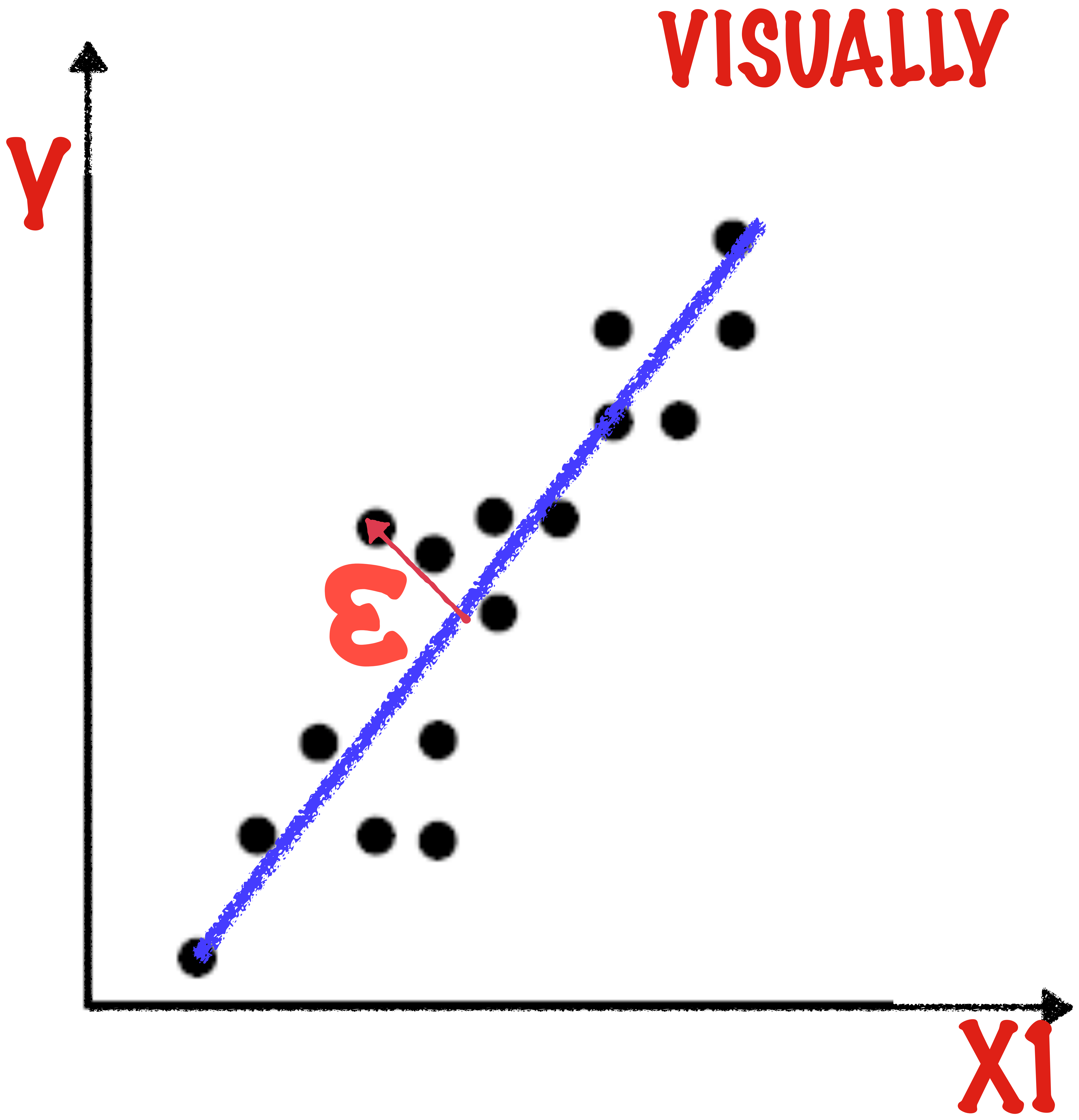
LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

$$Y = \hat{Y} + \varepsilon$$

ERROR

THESE ARE ALSO CALLED RESIDUALS



LINEAR REGRESSION

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

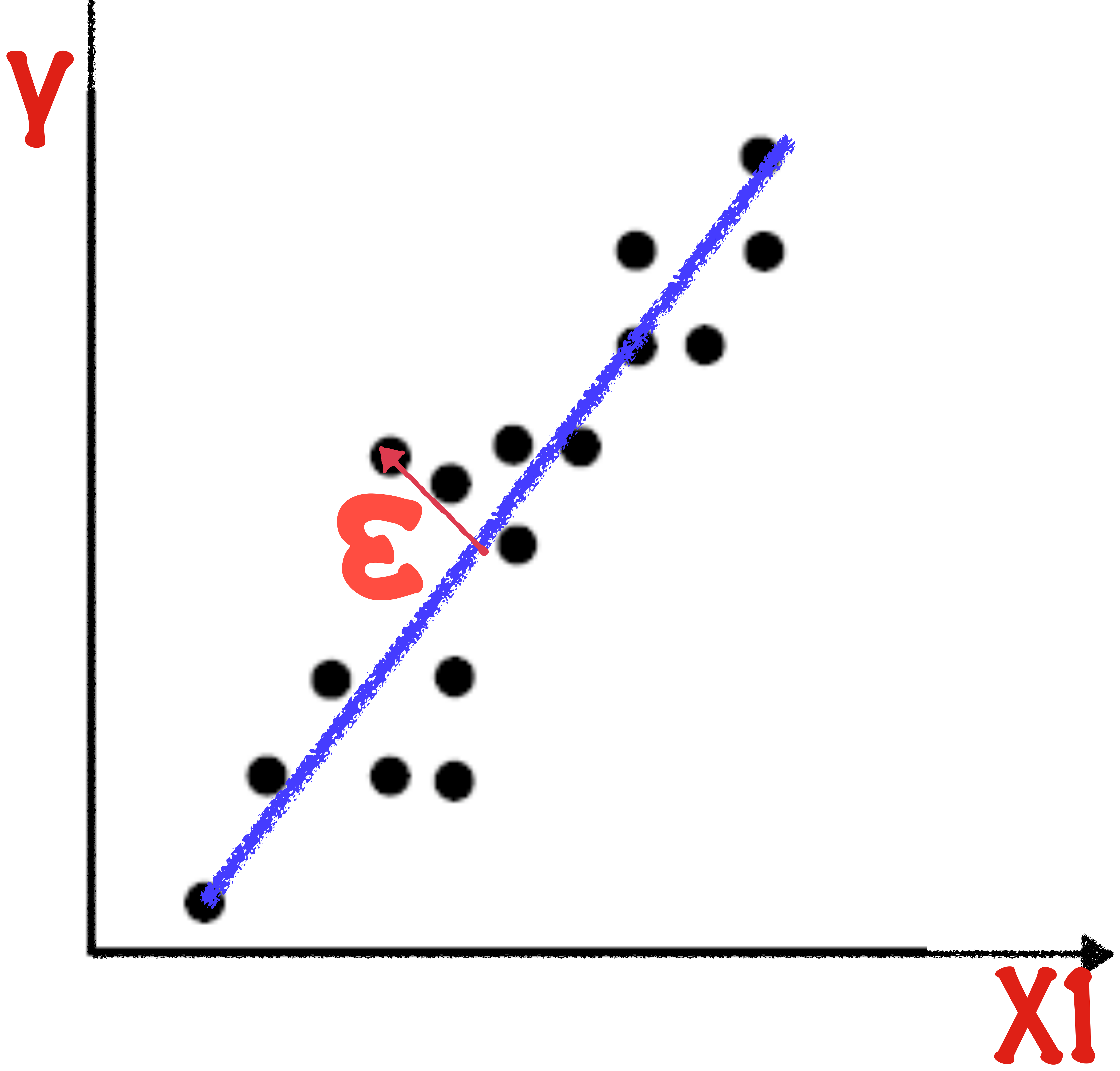
$$Y = \hat{Y} + \varepsilon$$

(RESIDUALS)

ERROR

THE LEFT OVER PARTS OF Y THAT ARE NOT EXPLAINED BY X1

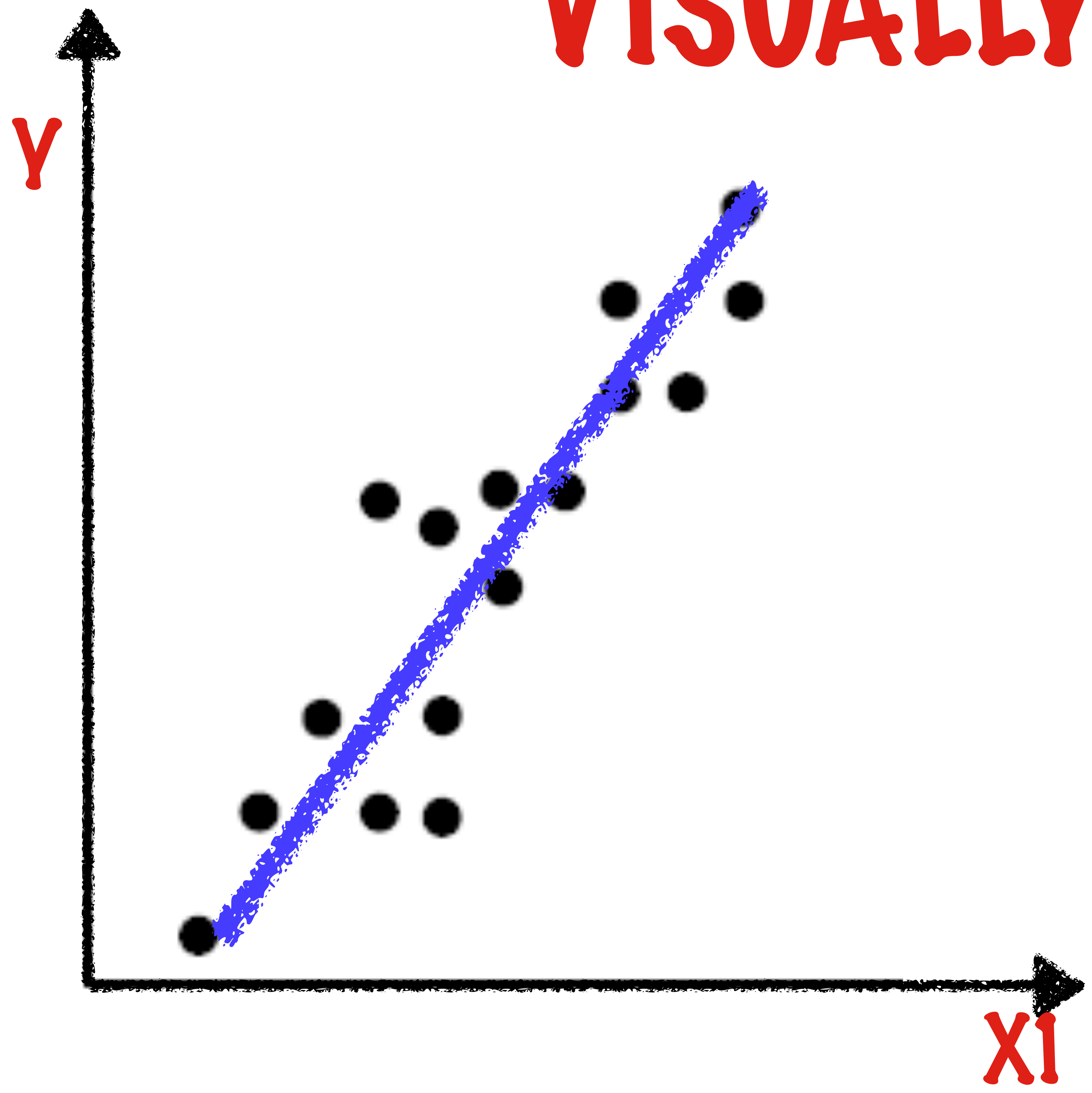
VISUALLY



LINEAR REGRESSION

VISUALLY

HOW DO WE FIND THE "BEST FIT" LINE?



$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

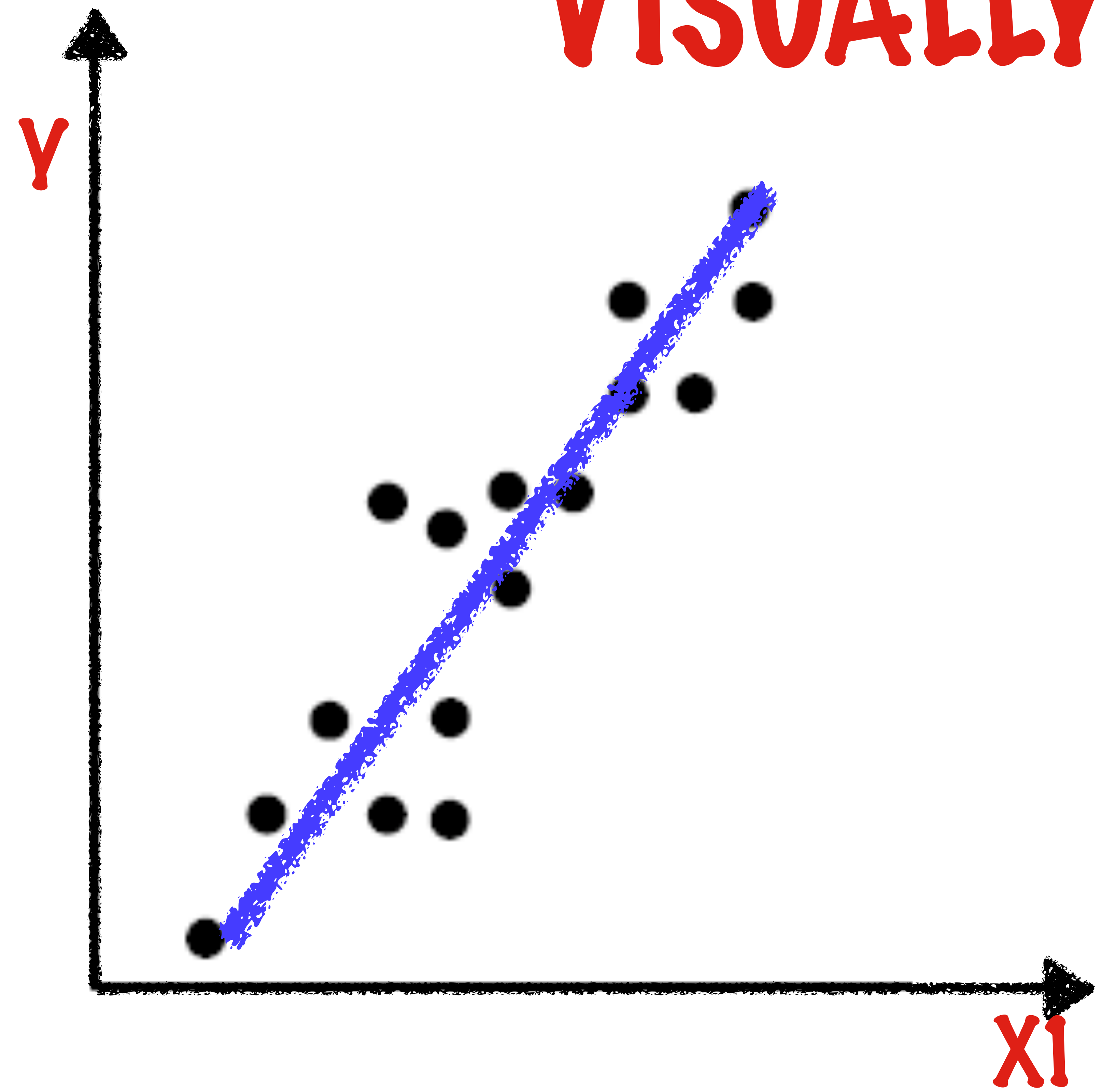
$$Y = \hat{Y} + \varepsilon$$

LINEAR REGRESSION

VISUALLY

HOW DO WE FIND THE "BEST FIT" LINE?

FIND THE LINE THAT **MINIMIZES** THE DISTANCES BETWEEN THE POINTS AND THE LINE



$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

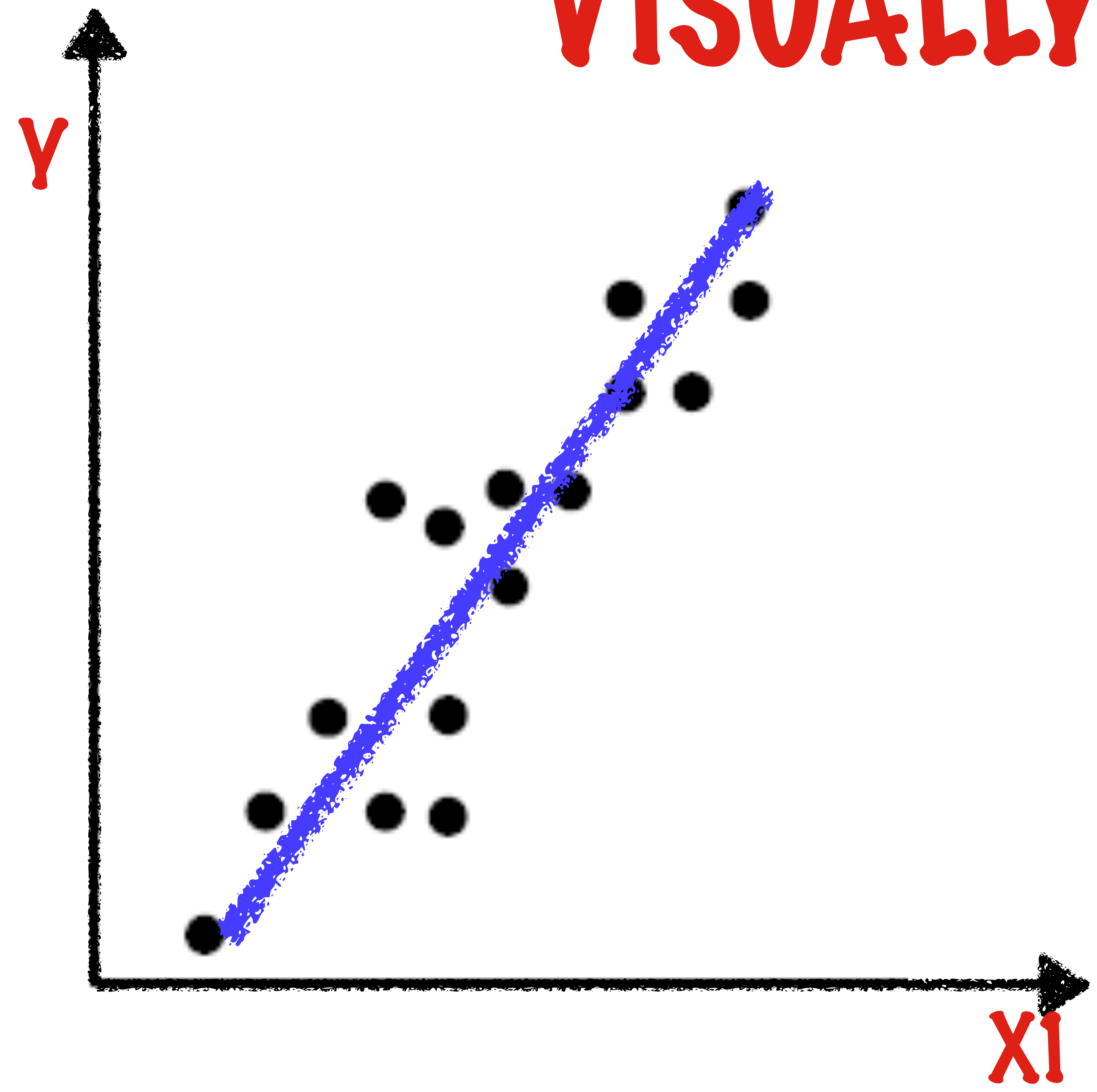
$$Y = \hat{Y} + \varepsilon$$

LINEAR REGRESSION

HOW DO WE FIND THE "BEST FIT" LINE?

FIND THE LINE THAT MINIMIZES THE **DISTANCES BETWEEN THE POINTS AND THE LINE**

VISUALLY



$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

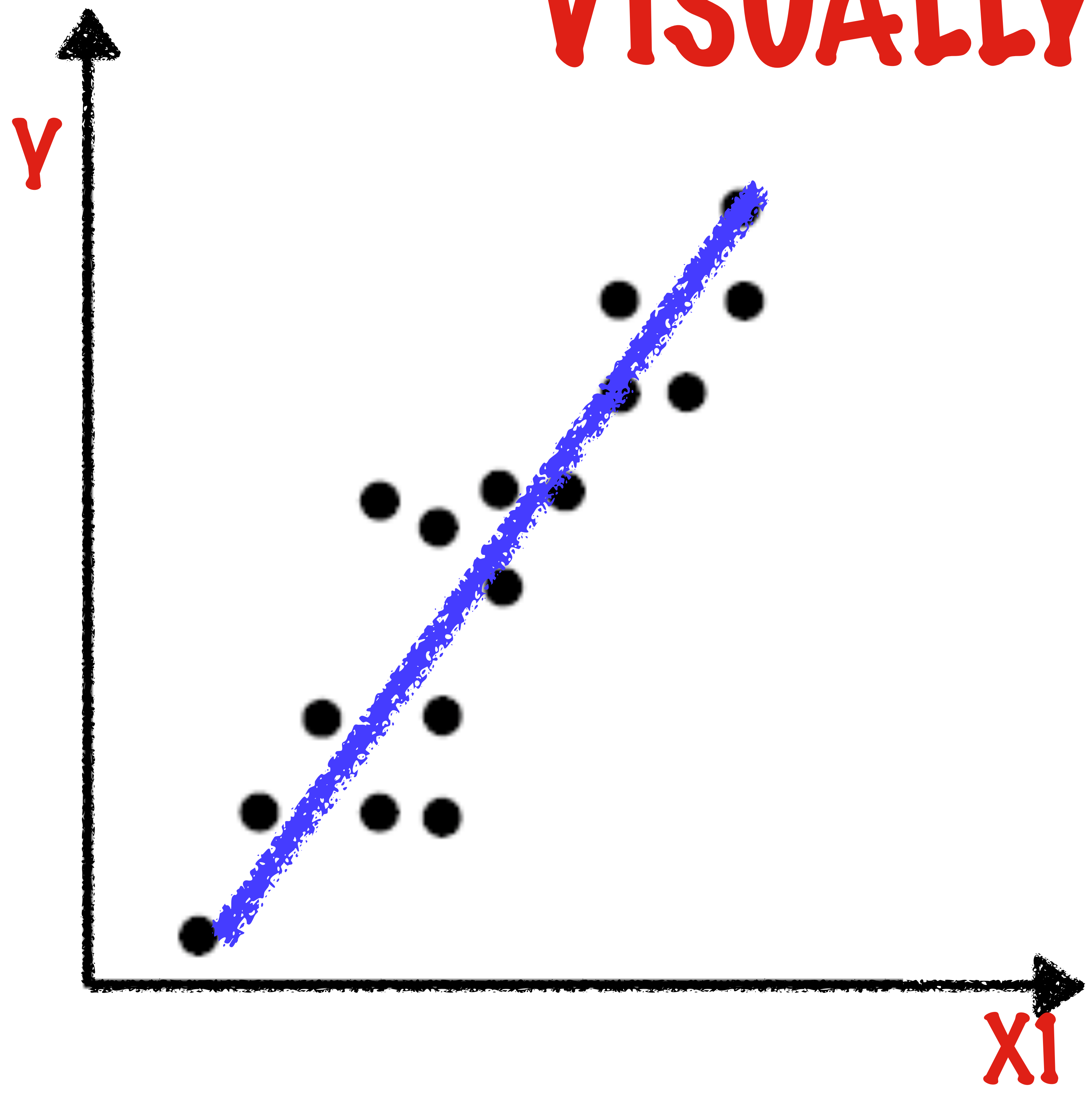
$$Y = \hat{Y} + \varepsilon$$

LINEAR REGRESSION

HOW DO WE FIND THE "BEST FIT" LINE?

FIND THE LINE THAT MINIMIZES THE **ERRORS**

VISUALLY



$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

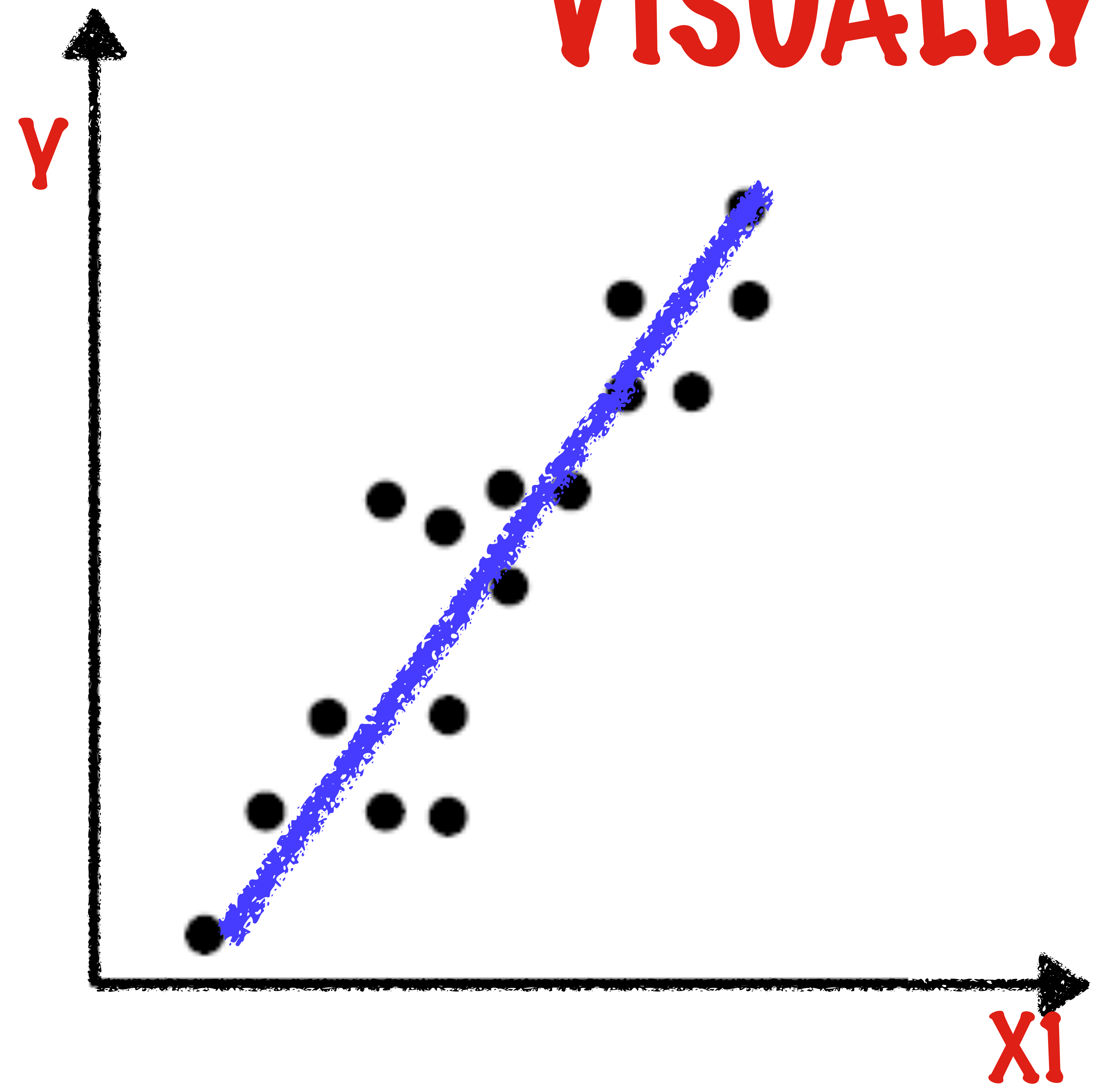
$$Y = \hat{Y} + \varepsilon$$

LINEAR REGRESSION

HOW DO WE FIND THE
"BEST FIT" LINE?

FIND THE LINE THAT **MINIMIZES** THE **ERRORS**

VISUALLY



$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

$$Y = \hat{Y} + \varepsilon$$

LINEAR REGRESSION

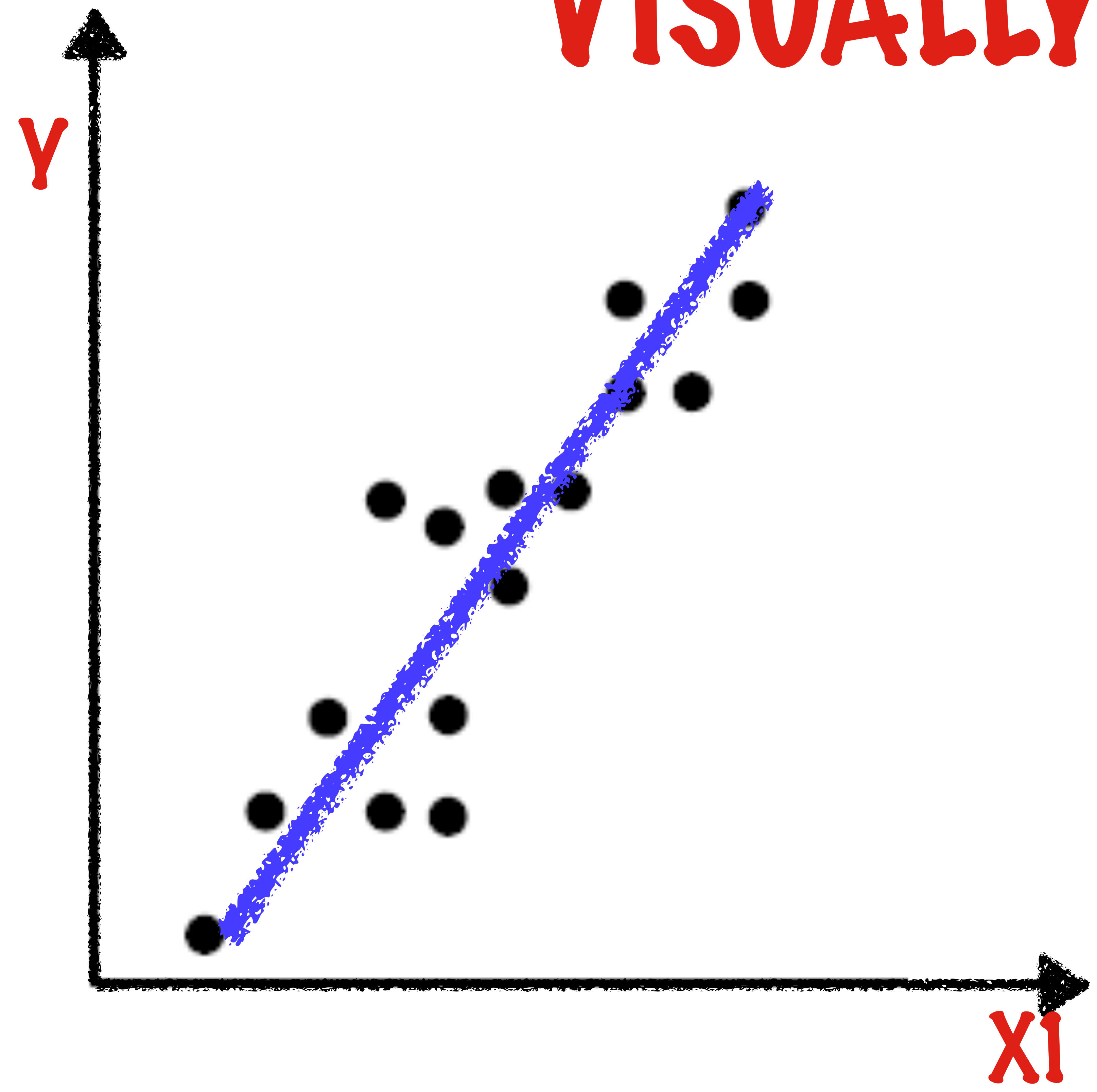
VISUALLY

HOW DO WE FIND THE "BEST FIT" LINE?

FIND THE LINE THAT MINIMIZES THE ERRORS

THIS IS DONE USING THE

ORDINARY LEAST SQUARES METHOD



$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

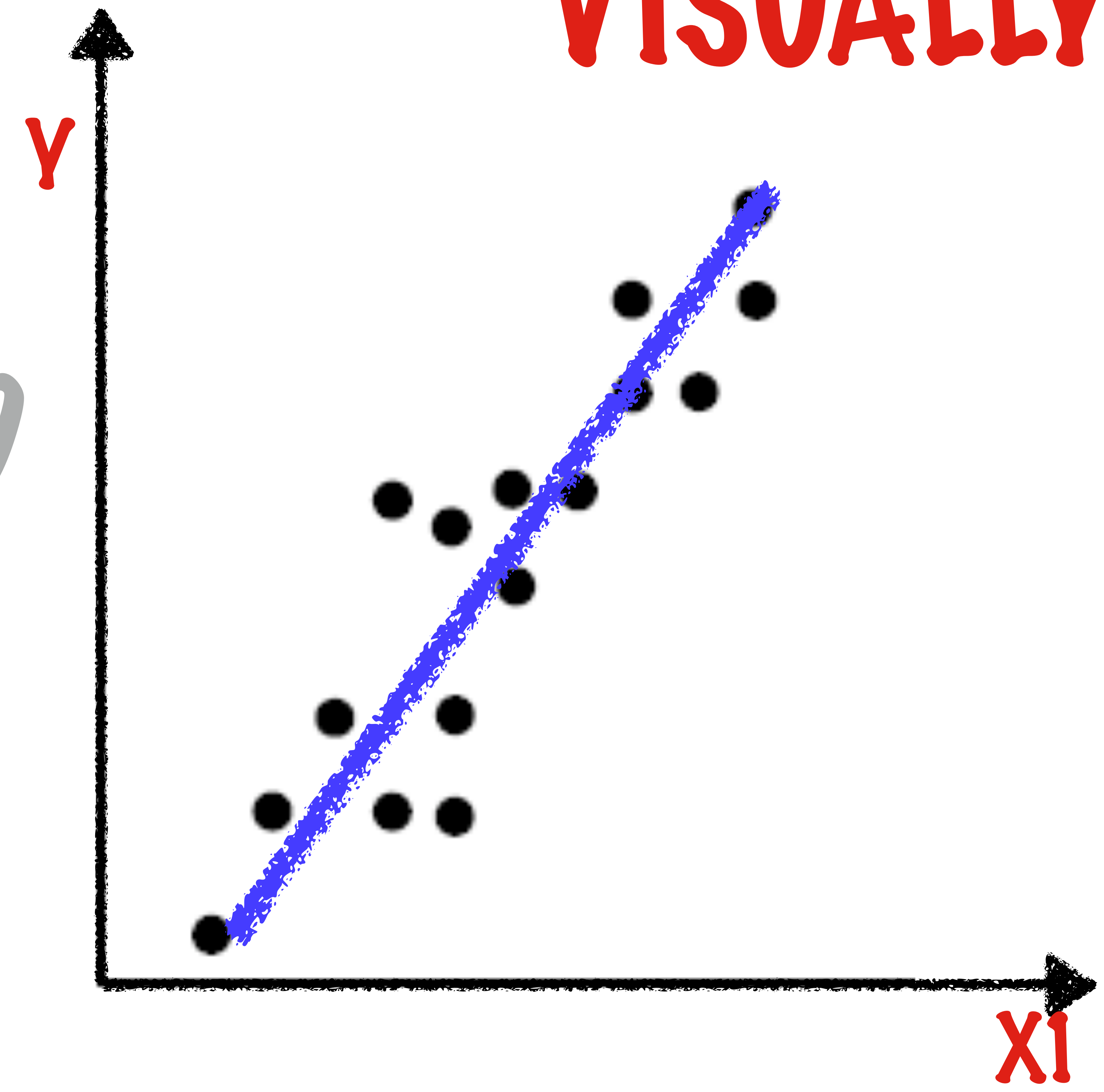
$$Y = \hat{Y} + \varepsilon$$

LINEAR REGRESSION

VISUALLY

THIS IS DONE USING THE

ORDINARY LEAST SQUARES METHOD



OLS MINIMIZES THE SUM OF SQUARES OF THE ERRORS

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

$$Y = \hat{Y} + \varepsilon$$

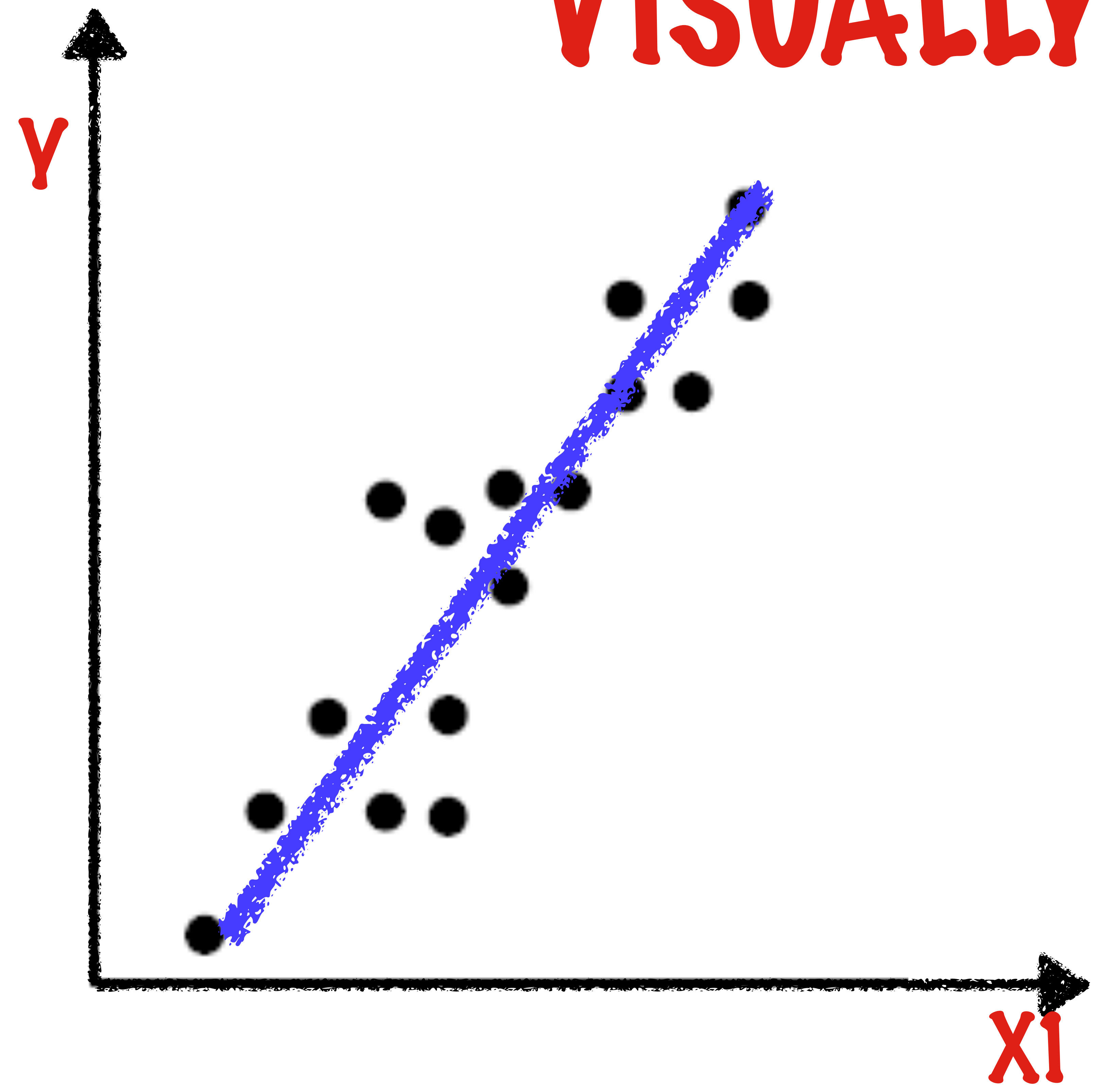
LINEAR REGRESSION

VISUALLY

ORDINARY LEAST SQUARES METHOD

R-SQUARED

IS A MEASURE OF HOW WELL THE LINE FITS THE PAST DATA



$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

$$Y = \hat{Y} + \varepsilon$$

LINEAR REGRESSION

ORDINARY LEAST SQUARES METHOD

R-SQUARED

TELLS US HOW MUCH OF THE VARIATION IN Y IS EXPLAINED BY THE INDEPENDENT VARIABLES

LINEAR REGRESSION

ORDINARY LEAST SQUARES METHOD

R-SQUARED

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

SUM OF SQUARES OF TOTAL DEVIATION FROM MEAN

VARIATION IN Y

LINEAR REGRESSION

ORDINARY LEAST SQUARES METHOD

R-SQUARED

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

SUM OF SQUARES OF REGRESSION

VARIATION IN Y EXPLAINED BY Ŷ