Implementing Factor Analysis and PCA in Excel and VBA

Vitthal Srinivasan
CO-FOUNDER, LOONYCORN
www.loonycorn.com

Overview

Explain returns of a stock using returns of several other stocks using PCA

Use VBA to create a user-generated function for eigen analysis in Excel

Calculate principal components of the financial data

Relate the principal components to underlying latent factors

Perform a regression using these principal components

PCA in Excel and VBA

Explain Google’s returns
Yahoo finance

Using returns of correlated stocks

Covariance and Correlation
Correlation matrix signals trouble

Multicollinearity problems

Eigen Decomposition
VBA

On covariance matrix

Scree Plot
Number of dimensions

Discard low-value dimensions

Principal Components
From eigen vectors

Uncorrelated components

Interpret and Regress
Beta, bonds, sectors

Now regress Google

Building Is Hard, Using Is Easy

Builder

Building a solver for eigen values and eigen vectors is **hard**

User

Using eigen values and eigen vectors for PCA is **easy**

Demo

Implement Eigen analysis in VBA

Use this to implement PCA in Excel

Apply multiple regression to the principal components found this way

Correlation Matrix

$$[\; X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_k \;]$$

$$\begin{bmatrix} \rho_{x_1} & \rho_{x_1 x_2} & \cdots & \rho_{x_1 x_k} \\ \rho_{x_2 x_1} & \rho_{x_2} & \cdots & \rho_{x_2 x_k} \\ \rho_{x_k x_1} & \rho_{x_k x_2} & \cdots & \rho_{x_k} \end{bmatrix}$$

k rows

k columns

Each element is the **correlation** of two random variables

Correlation Matrix

$$[\ X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_k\ ]$$

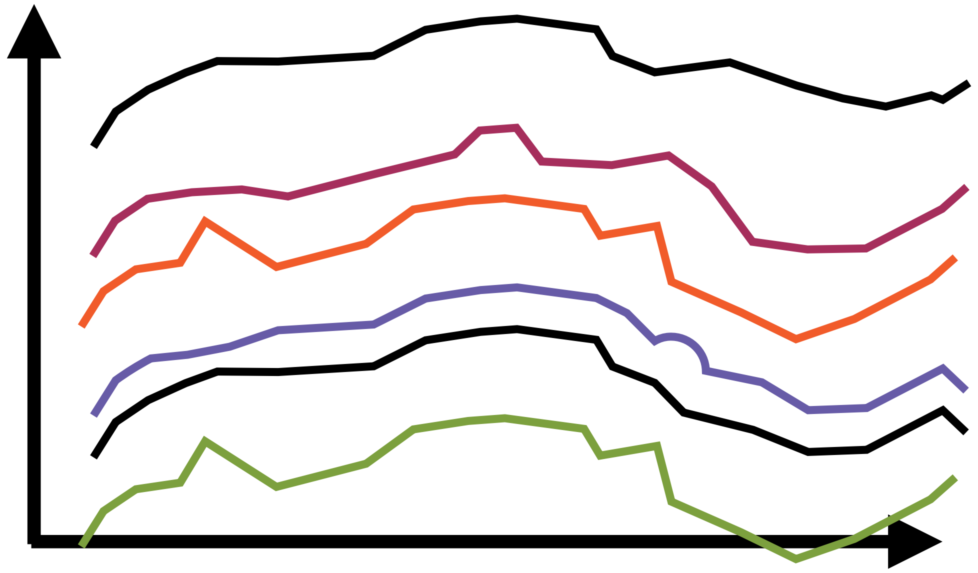
$$\begin{bmatrix} 1 & \rho_{x_1x_2} & \cdots & \rho_{x_1x_k} \\ \rho_{x_2x_1} & 1 & \cdots & \rho_{x_2x_k} \\ \rho_{x_kx_1} & \rho_{x_kx_2} & \cdots & 1 \end{bmatrix}$$

k rows

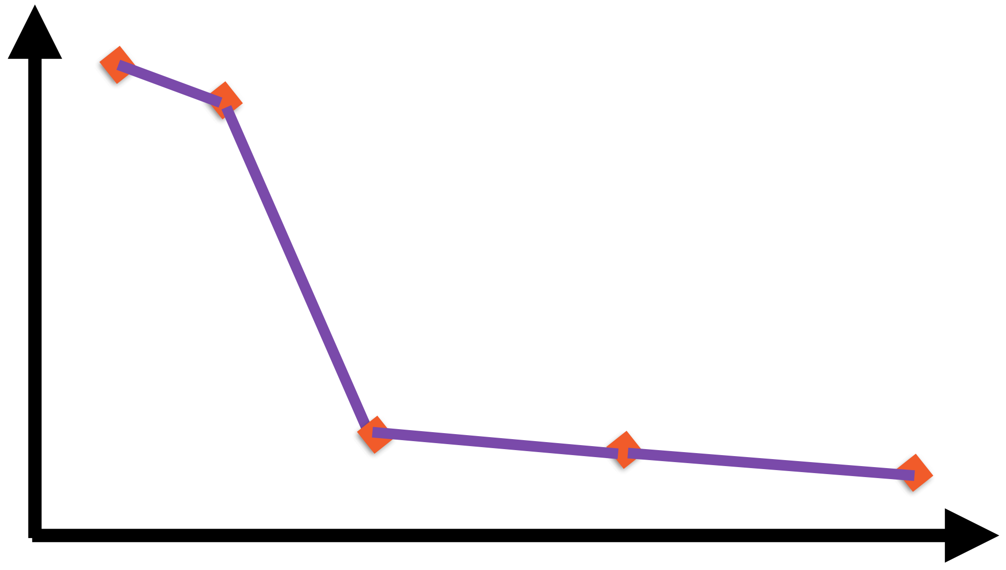
k columns

Diagonal elements are always 1

PCA’s Forte

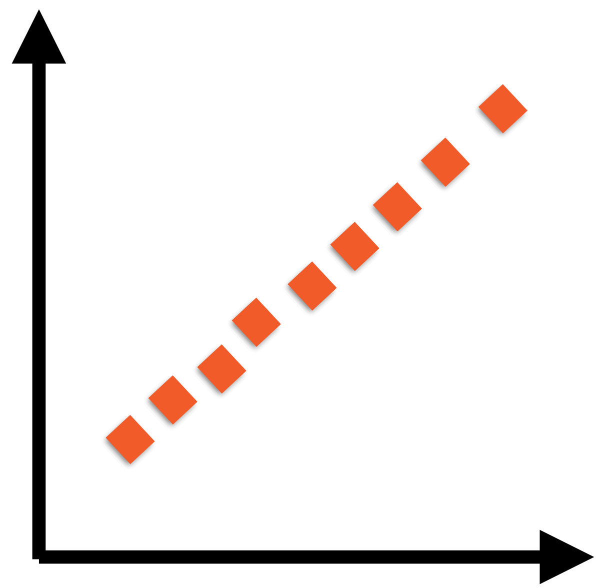


Many, Highly Correlated Xi



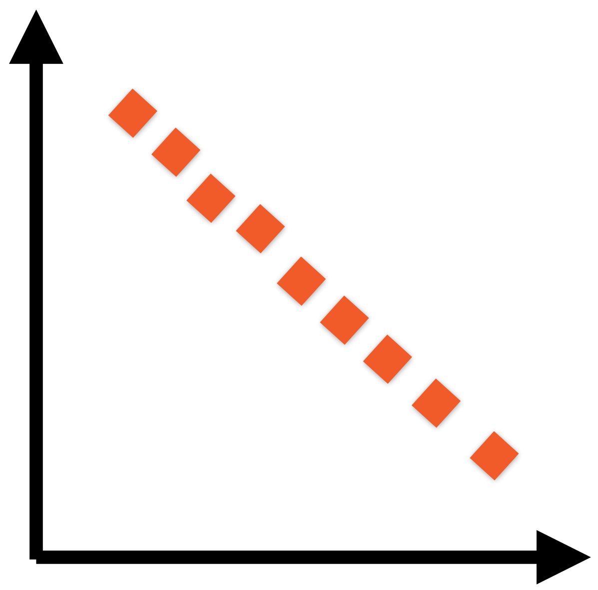
Unequal Eigenvalues

PCA for Highly Correlated Data



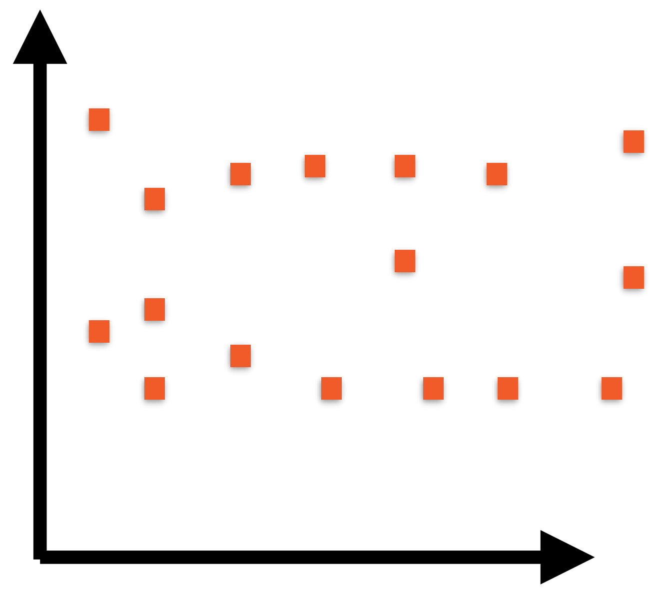
Correlation = +1

As X increases, Y increases linearly



Correlation = -1

As X increases, Y decreases linearly



Correlation = 0

Changes in X independent* of changes in Y

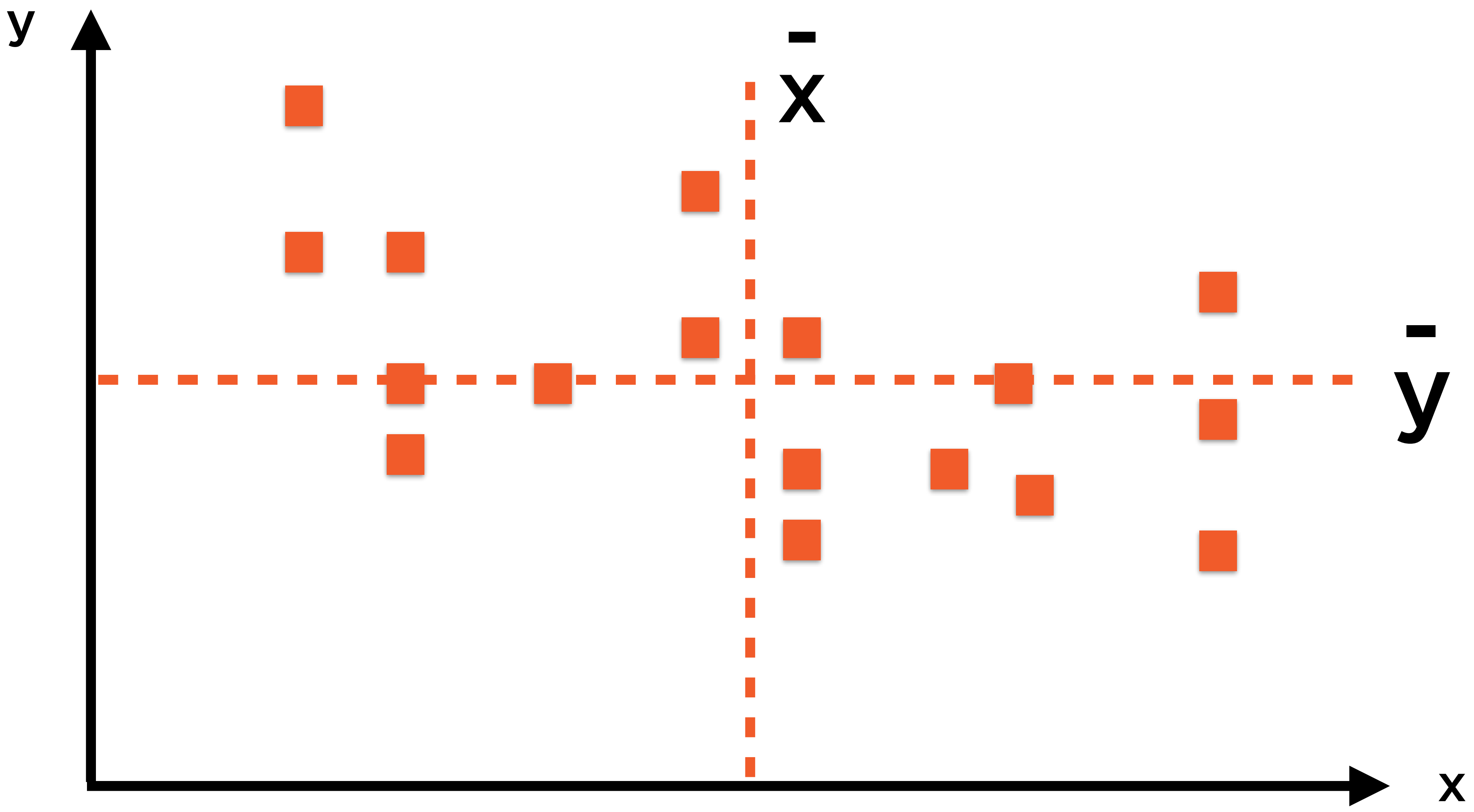
PCA for Highly Correlated Data

$$[\ X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_k\ ]$$

$$\begin{bmatrix} 1 & \rho_{x_1x_2} & \cdots & \rho_{x_1x_k} \\ \rho_{x_2x_1} & 1 & \cdots & \rho_{x_2x_k} \\ \rho_{x_kx_1} & \rho_{x_kx_2} & \cdots & 1 \end{bmatrix}$$

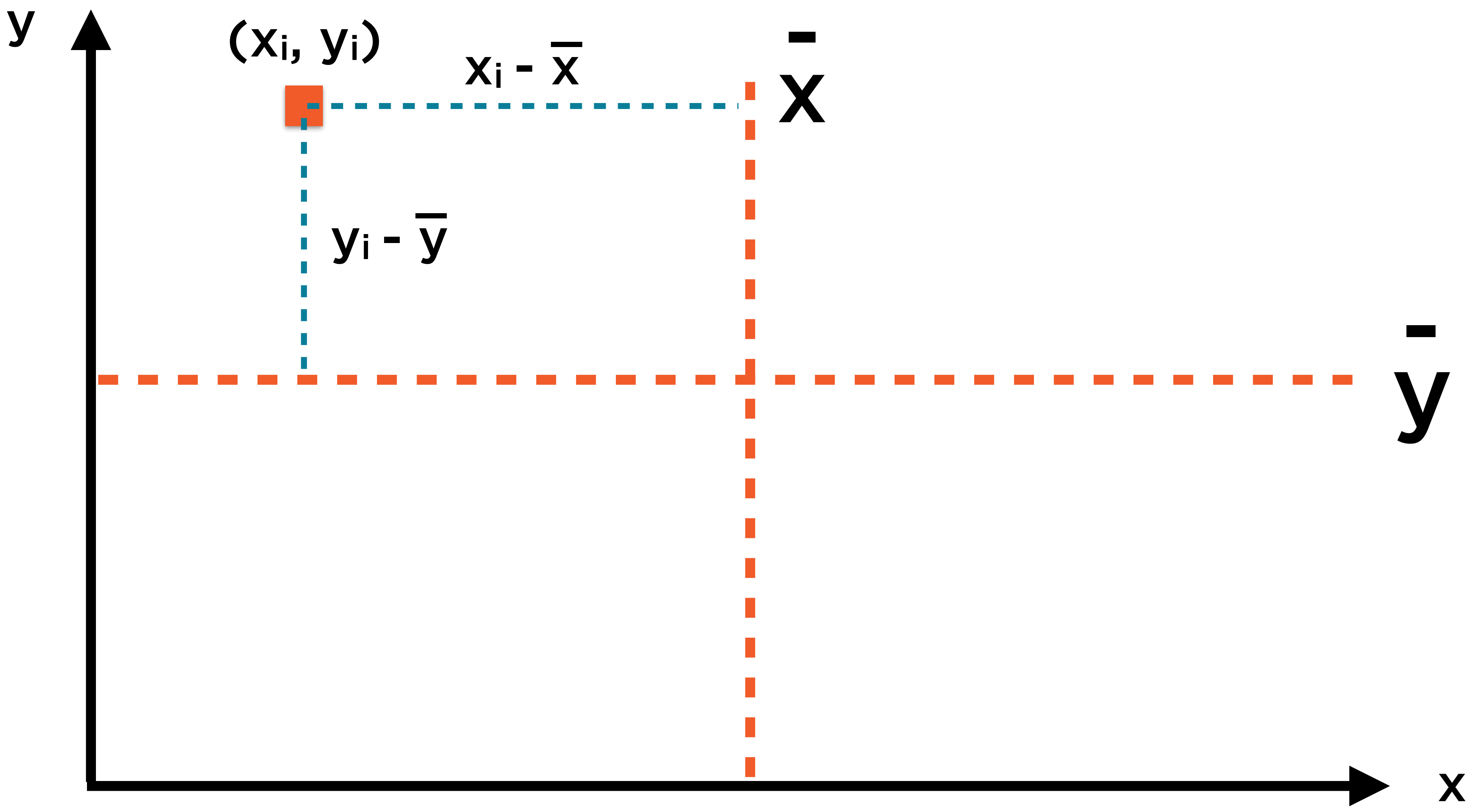
Rule-of-thumb: If average absolute values of off-diagonal entries is less than 0.3, PCA not a great idea

Covariance as Variance in Two Dimensions



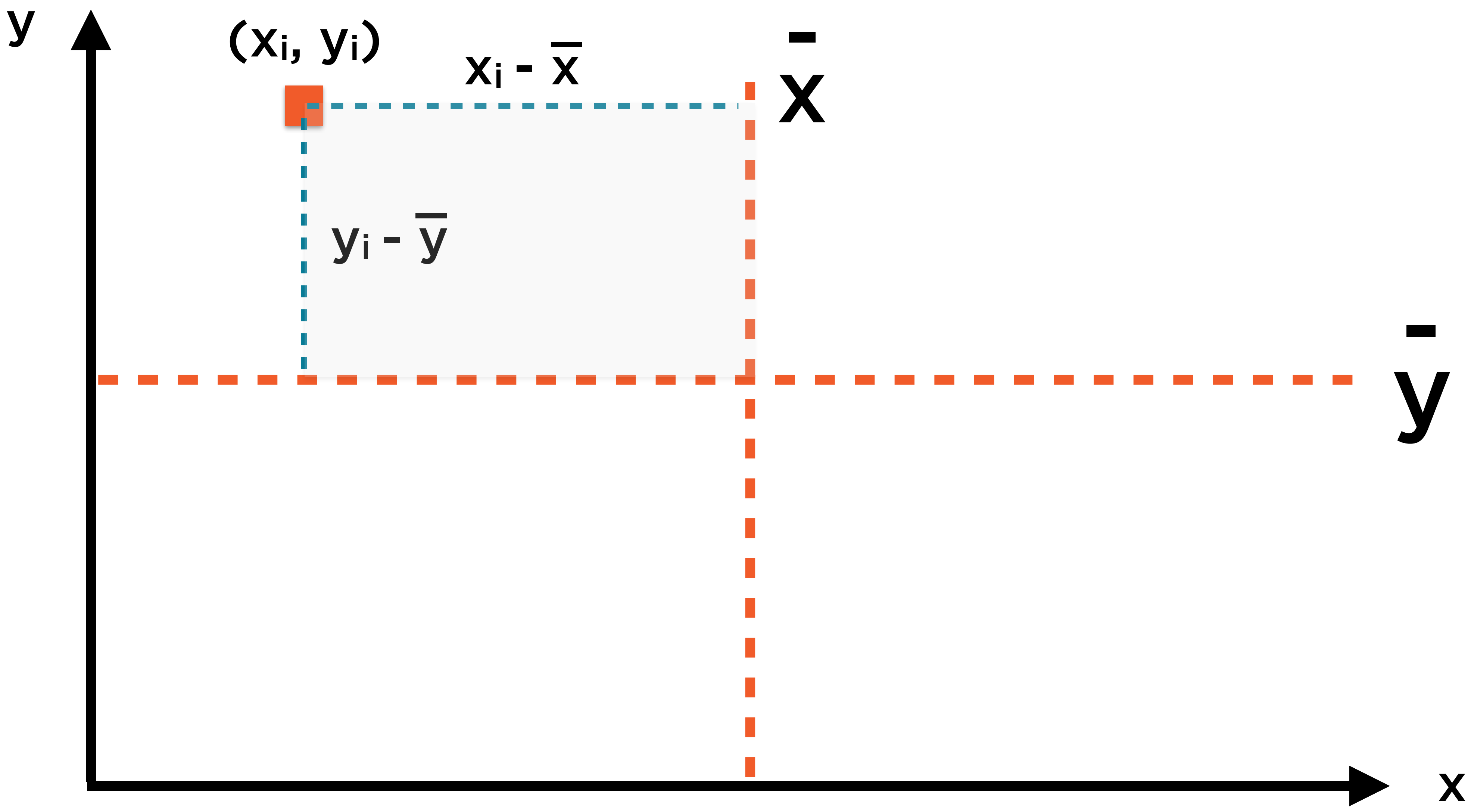
$$\text{Covariance}(x,y) = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Covariance as Variance in Two Dimensions



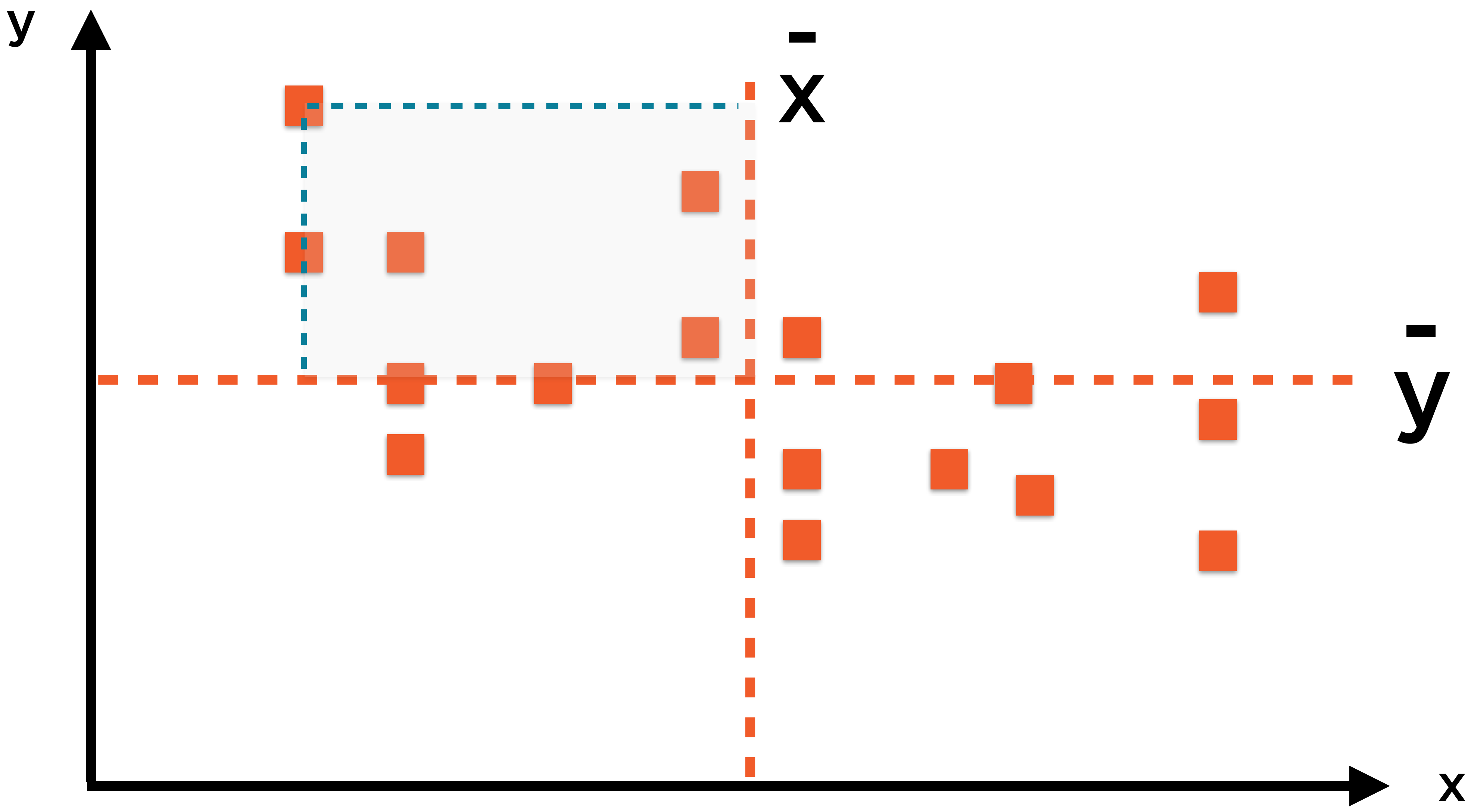
$$\text{Covariance }(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Covariance as Variance in Two Dimensions



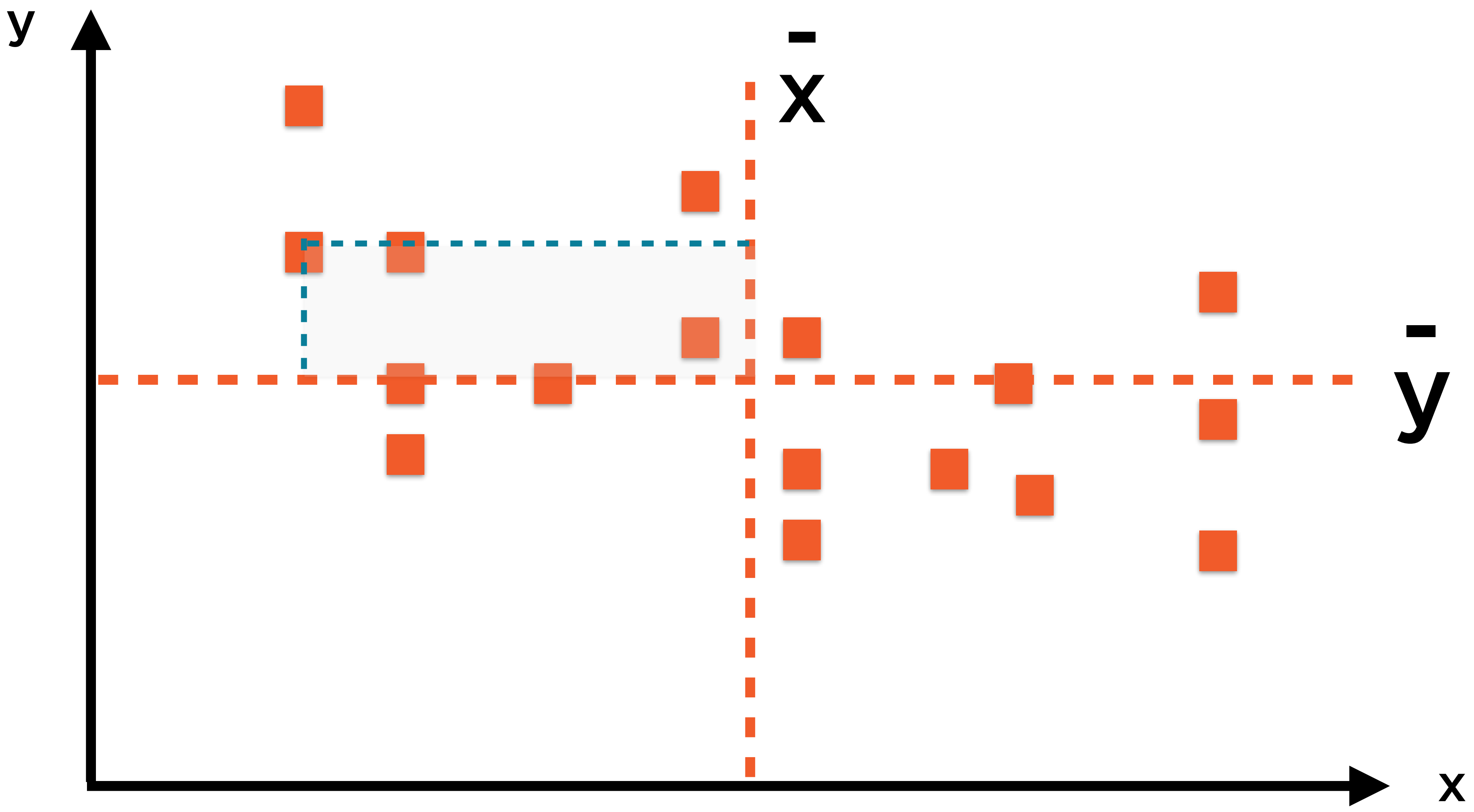
$$\text{Covariance }(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Covariance as Variance in Two Dimensions



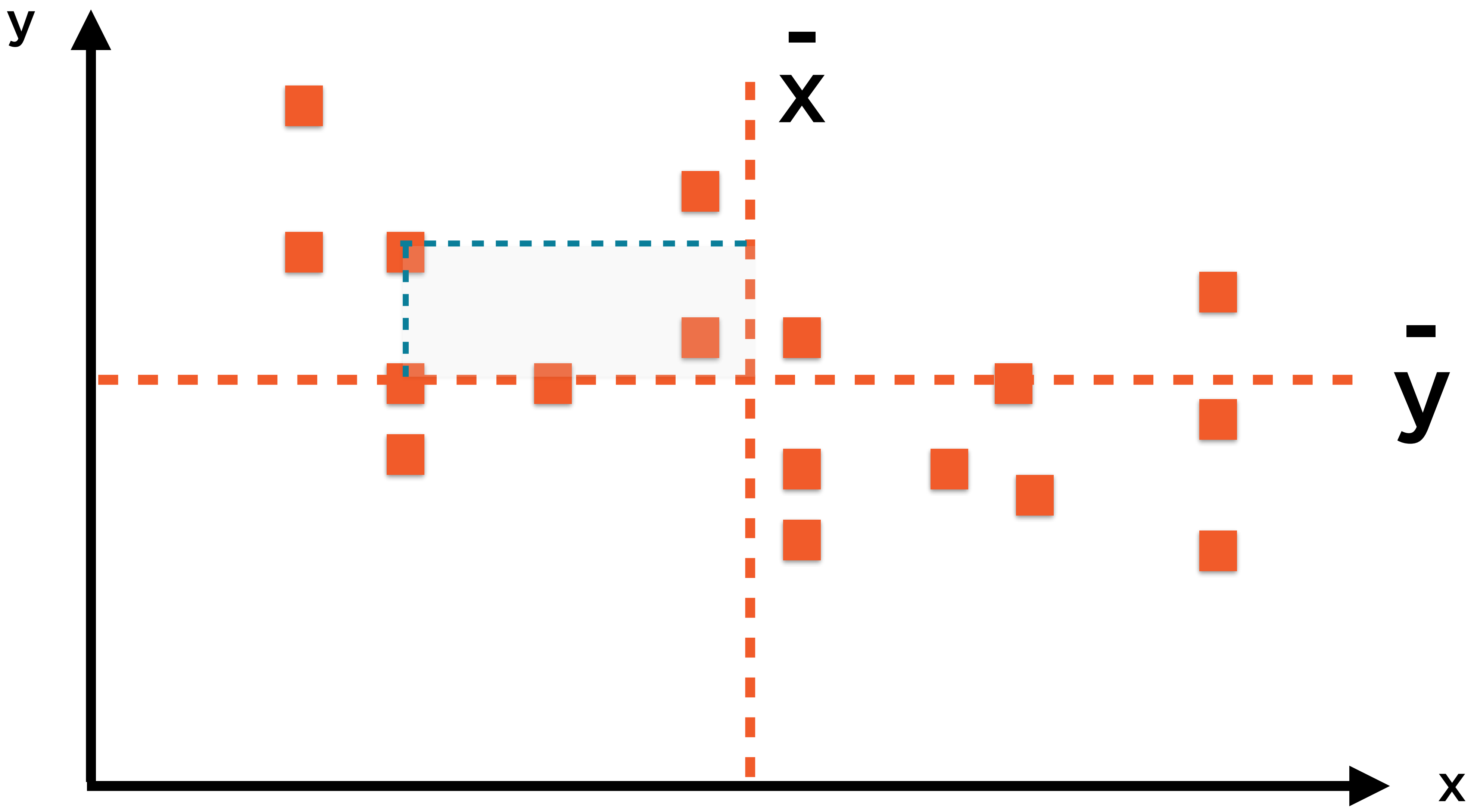
$$\text{Covariance}(x,y) = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Covariance as Variance in Two Dimensions



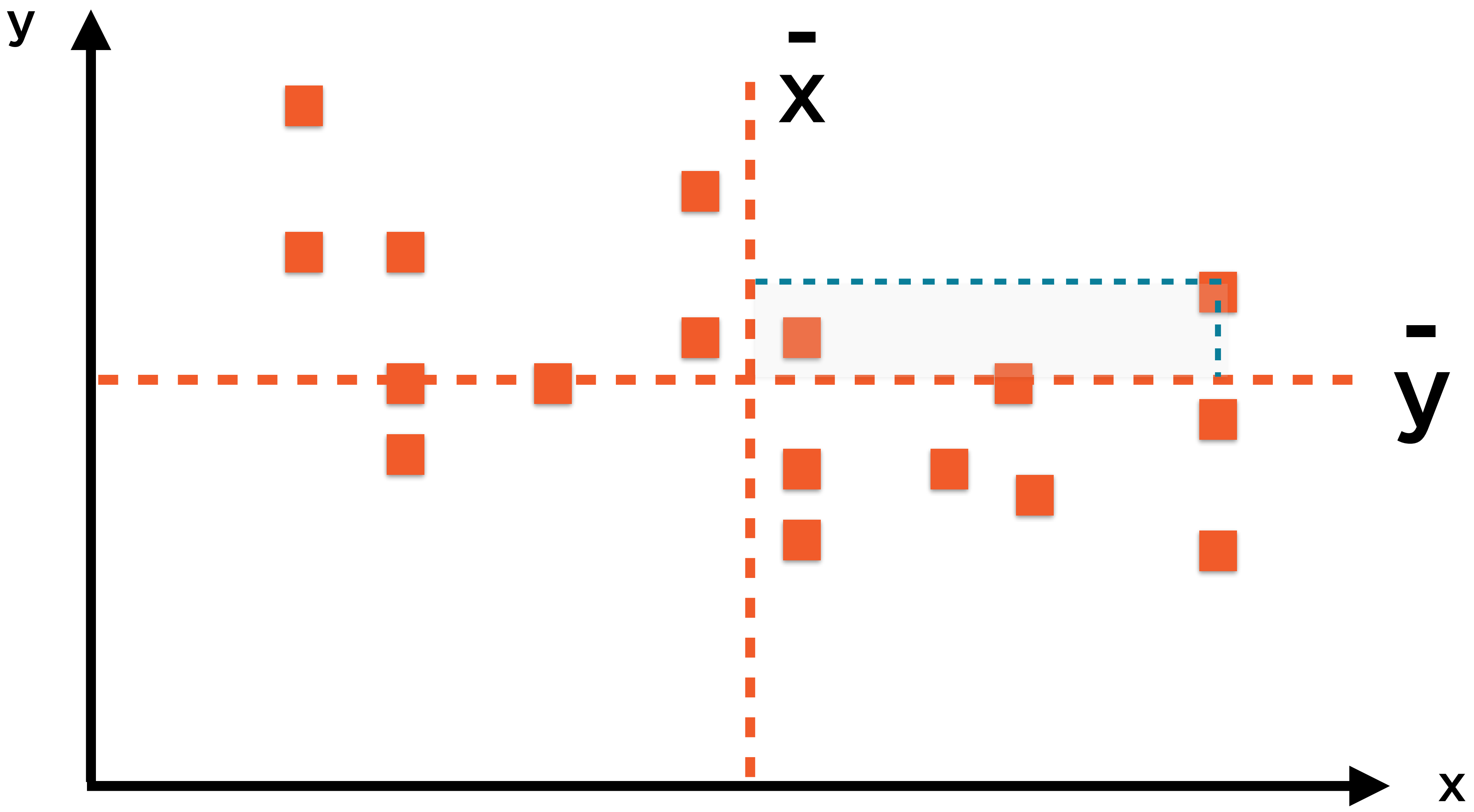
$$\text{Covariance}(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Covariance as Variance in Two Dimensions



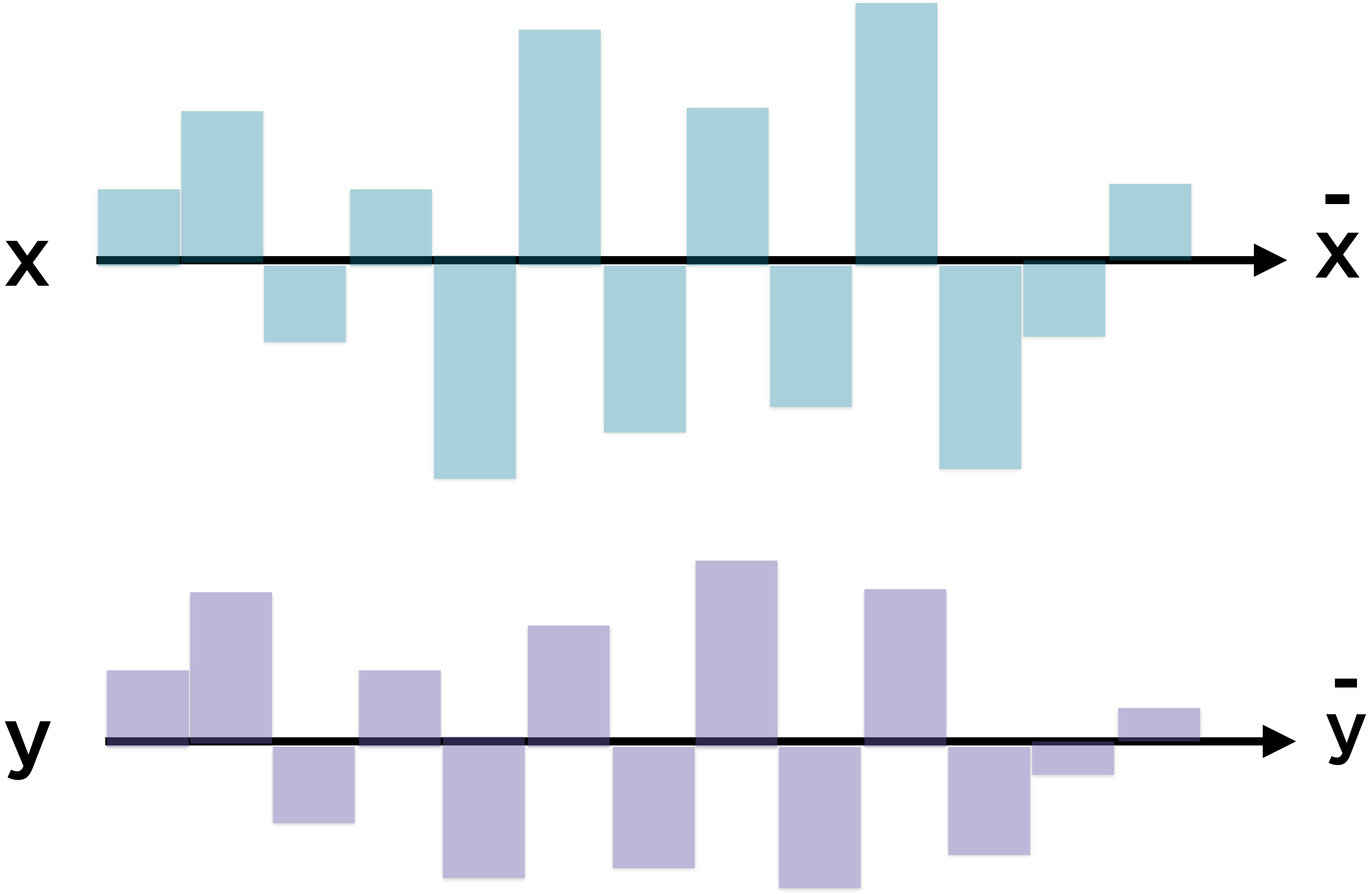
$$\text{Covariance}(x,y) = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Covariance as Variance in Two Dimensions

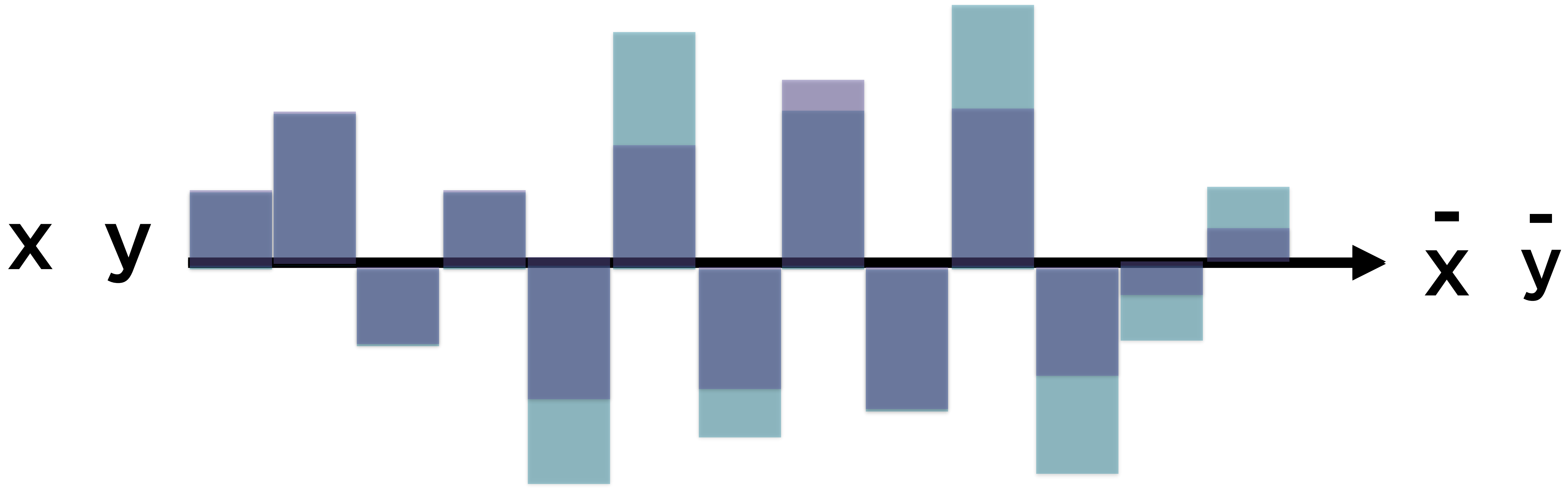


$$\text{Covariance}(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Intuition: Positive Covariance

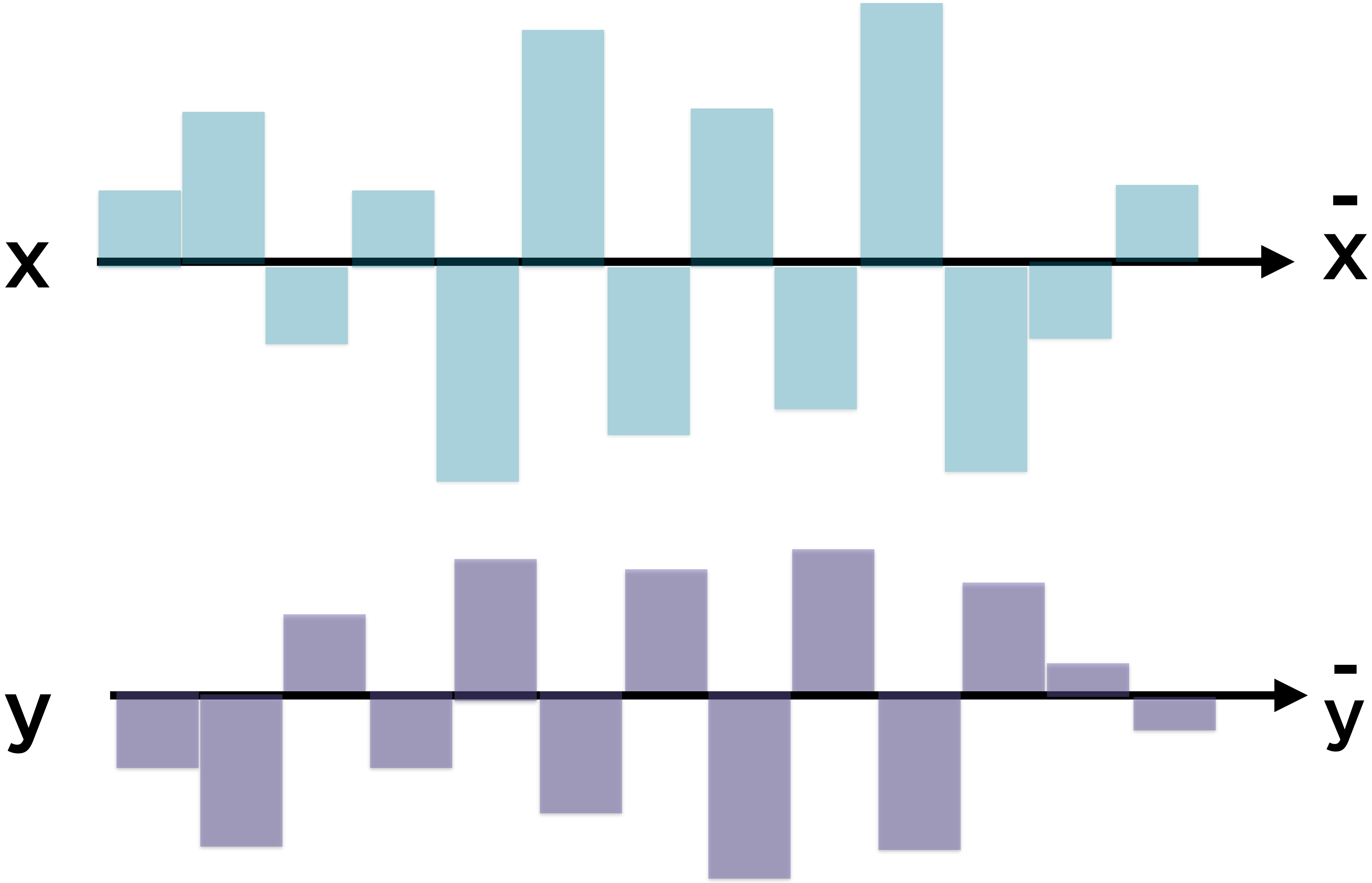


Intuition: Positive Covariance

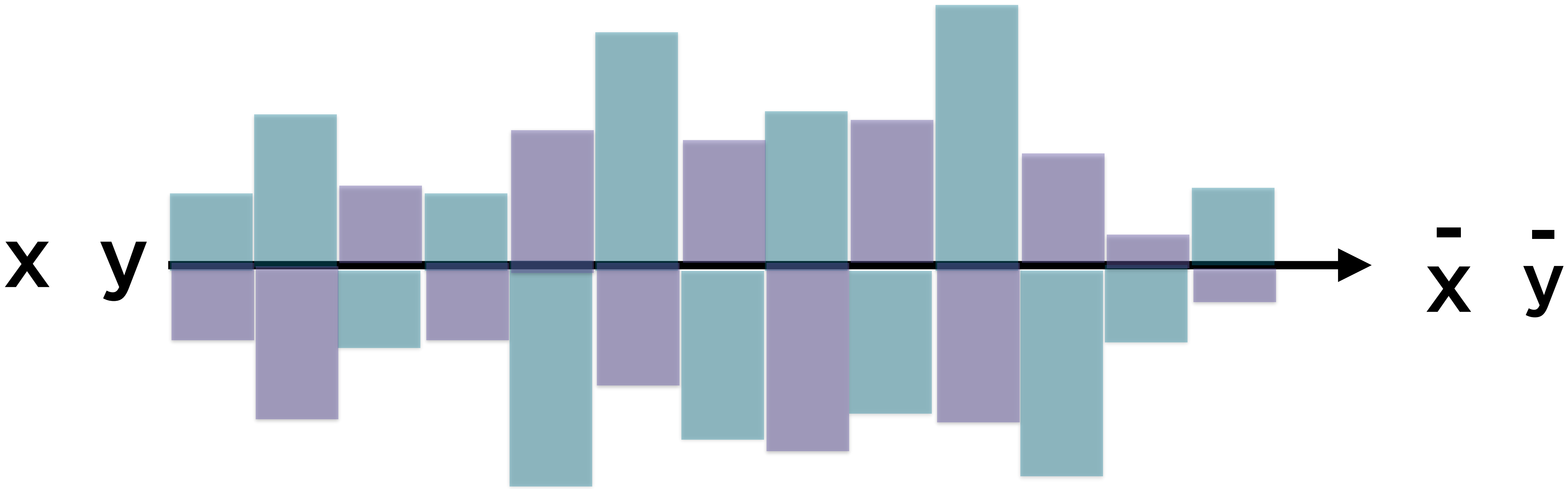


The deviations around the means of the two series are in-sync

Intuition: Negative Covariance



Intuition: Negative Covariance



The deviations around the means of the two series are out-of-sync

Principal Components Analysis

$[\ X_1\ X_2\ X_3 \dots X_k\ ]$

Eigenvalue Decomposition

Principal Components:

$[\ F_1\ F_2\ F_3 \dots F_k\ ]$ n rows

k columns

Eigenvectors:

$[\ v_1\ v_2\ v_3 \dots v_k\ ]$ k rows

k columns

Eigenvalues:

$[\ e_1\ e_2\ e_3 \dots e_k\ ]$ 1 row

k columns

Interpreting Eigenvalues

$$[\ F_1 \quad F_2 \quad F_3 \quad \dots \quad F_k\ ]$$

$$var(F_1) > var(F_2) > var(F_3) \quad > var(F_k)$$

These vectors F_i are arranged in order of decreasing variance

The greater the variance of a principal component, the more important it is

Interpreting Eigenvalues

$$[\ F_1 \quad F_2 \quad F_3 \quad \ldots \quad F_k\ ]$$

$$\mathrm{var}(F_1) > \mathrm{var}(F_2) > \mathrm{var}(F_3) \quad > \mathrm{var}(F_k)$$

Eigenvalue 1

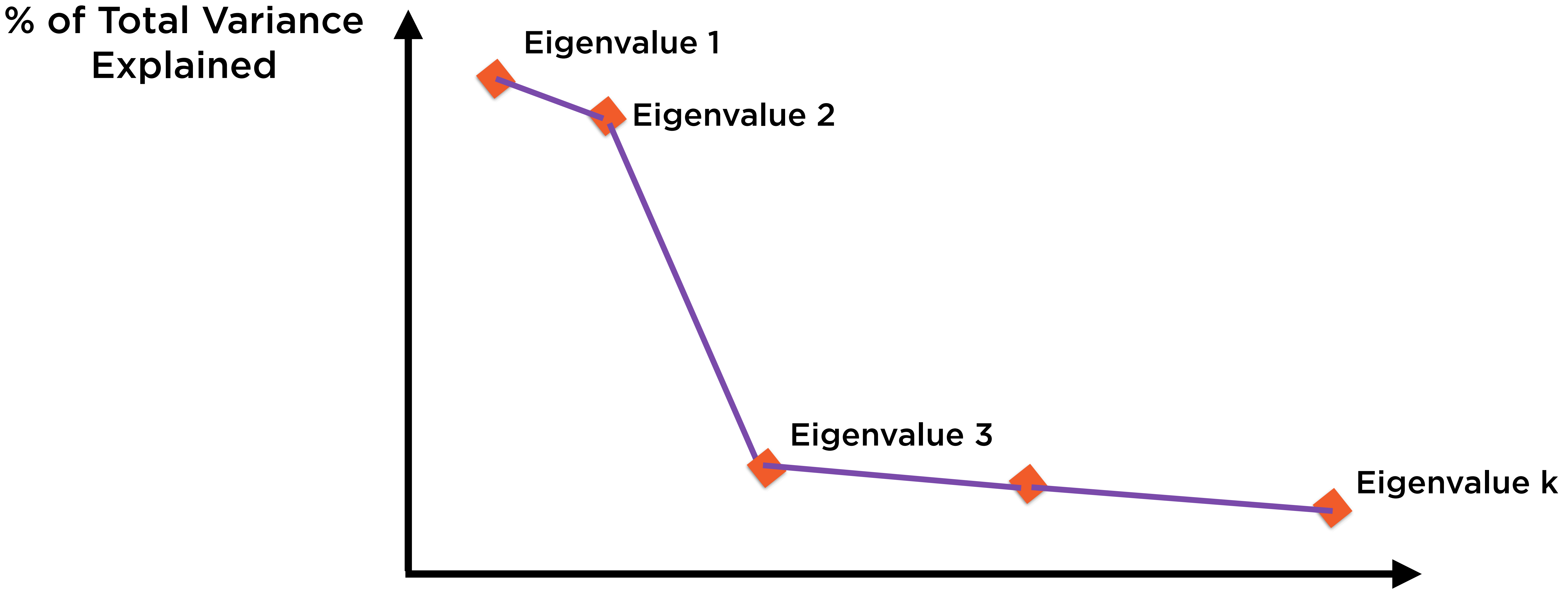
Eigenvalue 2

Eigenvalue 3

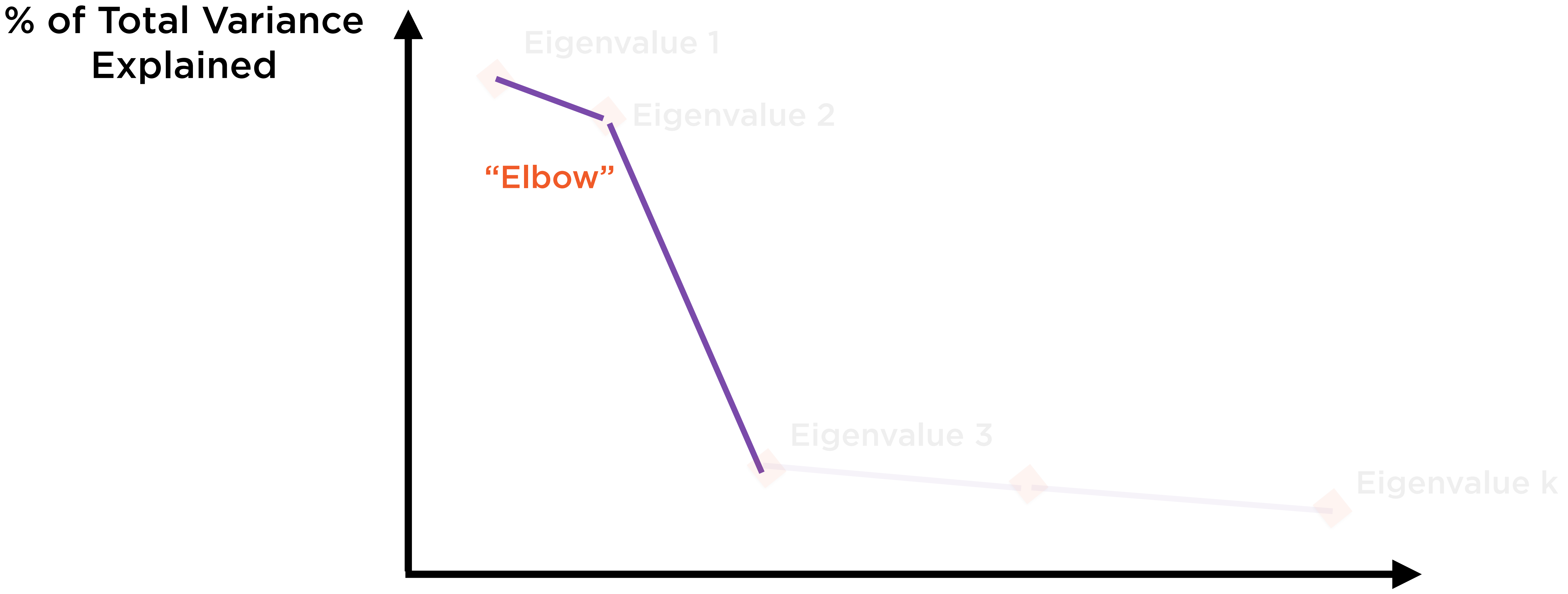
Eigenvalue k

The greater the eigenvalue of a principal component, the more important it is

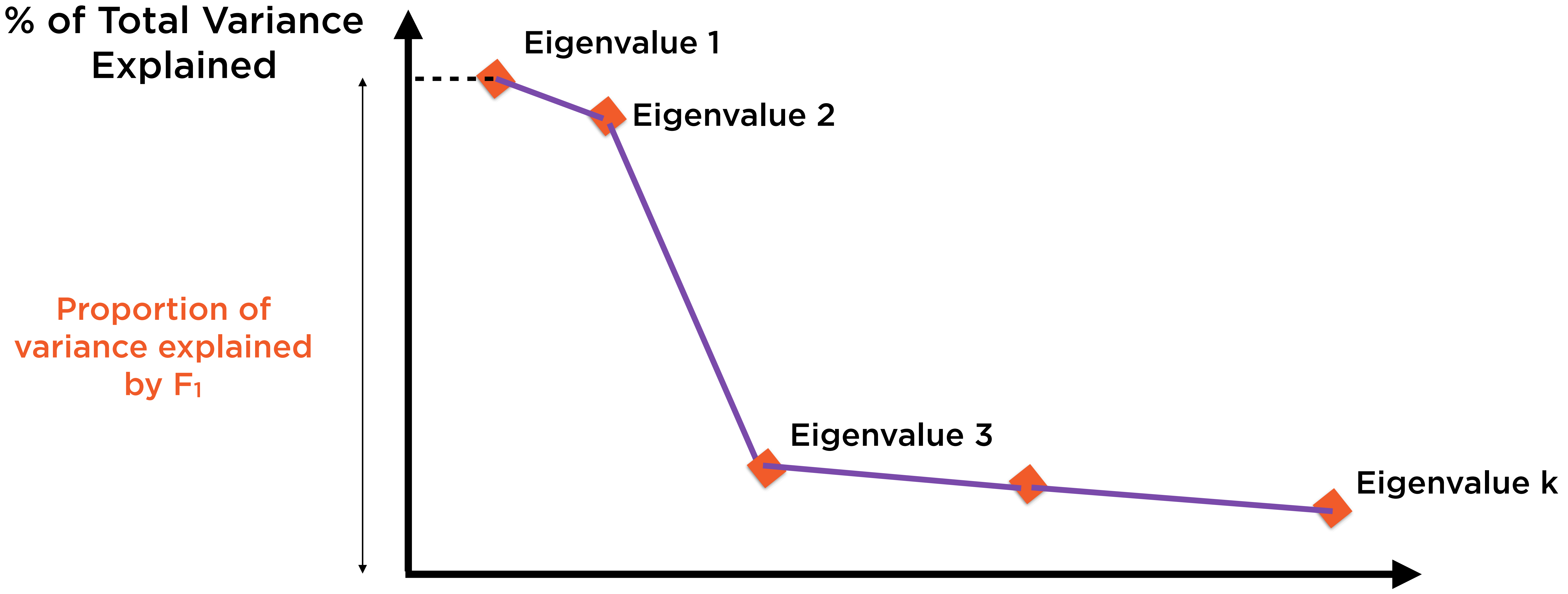
Scree Plots


% of Total Variance
Explained
Eigenvalue 1
Eigenvalue 2
Eigenvalue 3
Eigenvalue k

Scree Plots


% of Total Variance Explained
Eigenvalue 1
Eigenvalue 2
"Elbow"
Eigenvalue 3
Eigenvalue k

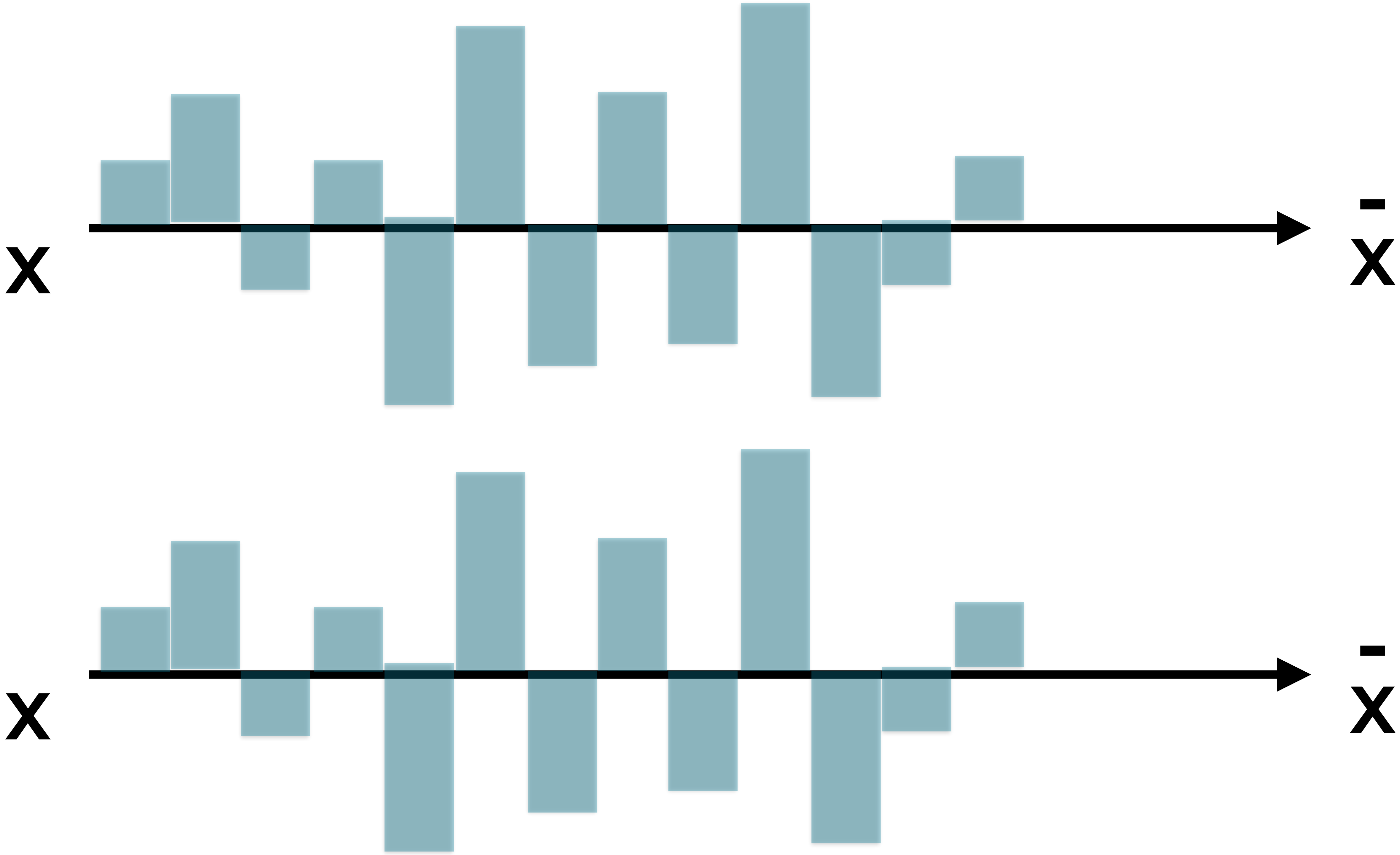
Scree Plots



% of Total Variance
Explained
Proportion of
variance explained
by F1
Eigenvalue 1
Eigenvalue 2
Eigenvalue 3
Eigenvalue k

Use the Scree plot to determine how many principal components to discard

Intuition: Covariance and Variance



Matrix Multiplication

$$F = X v$$

$$= \begin{array}{cc} \begin{bmatrix} x_{11} & & x_{1k} \\ x_{21} & & x_{2k} \\ x_{31} & \dots & x_{3k} \\ \dots & & \dots \\ x_{n1} & & x_{nk} \end{bmatrix} & \text{n rows} \\ \text{k columns} & \end{array} \begin{array}{cc} \begin{bmatrix} v_1 & v_2 & \dots & v_k \end{bmatrix} & \text{k rows} \\ \text{k columns} & \end{array}$$

Matrix Multiplication

$$F = X\,v$$

$$= \underbrace{\begin{bmatrix} x_{11} & & x_{1k} \\ x_{21} & & x_{2k} \\ x_{31} & \ldots & x_{3k} \\ \ldots & & \ldots \\ x_{n1} & & x_{nk} \end{bmatrix}}_{\text{k columns}} \text{n rows} \quad \underbrace{\begin{bmatrix} a_1 & b_1 & k_1 \\ a_2 & b_2 & k_2 \\ a_3 & b_3 & k_3 \\ \ldots & \ldots & \ldots \\ a_k & b_k & k_k \end{bmatrix}}_{\text{k columns}} \text{k rows}$$

$v_1 \quad v_2 \;\ldots\; v_k$

Matrix Multiplication

$$\underbrace{\begin{bmatrix} F_{11} & & F_{1k} \\ F_{21} & & F_{2k} \\ F_{31} & \dots & F_{3k} \\ \dots & & \dots \\ F_{n1} & & F_{nk} \end{bmatrix}}_{\text{k columns}} \updownarrow^{\text{n rows}} = \underbrace{\begin{bmatrix} x_{11} & & x_{1k} \\ x_{21} & & x_{2k} \\ x_{31} & \dots & x_{3k} \\ \dots & & \dots \\ x_{n1} & & x_{nk} \end{bmatrix}}_{\text{k columns}} \updownarrow^{\text{n rows}} \underbrace{\begin{bmatrix} a_1 & b_1 & k_1 \\ a_2 & b_2 & k_2 \\ a_3 & b_3 & k_3 \\ \dots & \dots & \dots \\ a_k & b_k & k_k \end{bmatrix}}_{\text{k columns}} \updownarrow^{\text{k rows}}$$

$$V_1 \quad V_2 \; \dots \; V_k$$

Matrix Multiplication

$$\begin{bmatrix} F_{11} & & F_{1k} \\ F_{21} & & F_{2k} \\ F_{31} & \cdots & F_{3k} \\ \cdots & & \cdots \\ F_{n1} & & F_{nk} \end{bmatrix} = \begin{bmatrix} x_{11} & & x_{1k} \\ x_{21} & & x_{2k} \\ x_{31} & \cdots & x_{3k} \\ \cdots & & \cdots \\ x_{n1} & & x_{nk} \end{bmatrix} \begin{bmatrix} a_1 & b_1 & k_1 \\ a_2 & b_2 & k_2 \\ a_3 & b_3 & k_3 \\ \cdots & \cdots & \cdots \\ a_k & b_k & k_k \end{bmatrix}$$

Matrix Multiplication

$$\begin{bmatrix} F_{11} & & F_{1k} \\ F_{21} & & F_{2k} \\ F_{31} & \dots & F_{3k} \\ \dots & & \dots \\ F_{n1} & & F_{nk} \end{bmatrix} = \begin{bmatrix} x_{11} & & x_{1k} \\ \boxed{x_{21} \quad \quad x_{2k}} \\ x_{31} & \dots & x_{3k} \\ \dots & & \dots \\ x_{n1} & & x_{nk} \end{bmatrix} \begin{bmatrix} \boxed{\begin{matrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_k \end{matrix}} & \begin{matrix} b_1 \\ b_2 \\ b_3 \\ \dots \\ b_k \end{matrix} & \begin{matrix} k_1 \\ k_2 \\ k_3 \\ \dots \\ k_k \end{matrix} \end{bmatrix}$$

Matrix Multiplication

$$\begin{bmatrix} F_{11} & & F_{1k} \\ F_{21} & & F_{2k} \\ F_{31} & \dots & F_{3k} \\ \dots & & \dots \\ F_{n1} & & F_{nk} \end{bmatrix} = \begin{bmatrix} x_{11} & & x_{1k} \\ x_{21} & & x_{2k} \\ x_{31} & \dots & x_{3k} \\ \dots & & \dots \\ x_{n1} & & x_{nk} \end{bmatrix} \begin{bmatrix} a_1 & b_1 & k_1 \\ a_2 & b_2 & k_2 \\ a_3 & b_3 & k_3 \\ \dots & \dots & \dots \\ a_k & b_k & k_k \end{bmatrix}$$

Matrix Multiplication

$$\underbrace{\begin{bmatrix} F_{11} \\ F_{21} \\ F_{31} \\ \cdots \\ F_{n1} \end{bmatrix}}_{\text{1 column}}^{\text{n rows}} = \underbrace{\begin{bmatrix} x_{11} & & x_{1k} \\ x_{21} & & x_{2k} \\ x_{31} & \cdots & x_{3k} \\ \cdots & & \cdots \\ x_{n1} & & x_{nk} \end{bmatrix}}_{\text{k columns}}^{\text{n rows}} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \cdots \\ a_k \end{bmatrix}}_{\text{1 column}}^{\text{k rows}}$$

Each principal component is the matrix product of the original data and the corresponding eigenvector

PCA should always be applied on the covariance matrix of standardised vectors

Standardising Data

$$\begin{bmatrix} x_{11} & & x_{1k} \\ x_{21} & & x_{2k} \\ x_{31} & \cdots & x_{3k} \\ \cdots & & \cdots \\ x_{n1} & & x_{nk} \end{bmatrix}$$

$avg(X_1) \quad \ldots \quad avg(X_k)$

$stdev(X_1) \quad \ldots \quad stdev(X_k)$

Standardising Data

$$
\begin{bmatrix}
\dfrac{x_{11} - avg(X_1)}{stdev(X_1)} & \cdots & \dfrac{x_{1k} - avg(X_k)}{stdev(X_k)} \\
\cdots & & \cdots \\
\dfrac{x_{n1} - avg(X_1)}{stdev(X_1)} & & \dfrac{x_{nk} - avg(X_k)}{stdev(X_k)}
\end{bmatrix}
$$

Each column of the standardised data has mean 0 and variance 1

PCA for Latent Factor Identification

$$[\ F_1 \quad F_2 \quad F_3 \quad \ldots \quad F_k\ ]$$

$$\downarrow$$

$$\downarrow$$

$$[\ L_1 \quad L_2 \quad L_3 \quad \ldots \quad L_k\ ]$$

Exploratory Factor Analysis: Experts trace back principal components to observable factors

3 Latent Factors in Stock Returns

Market Movements

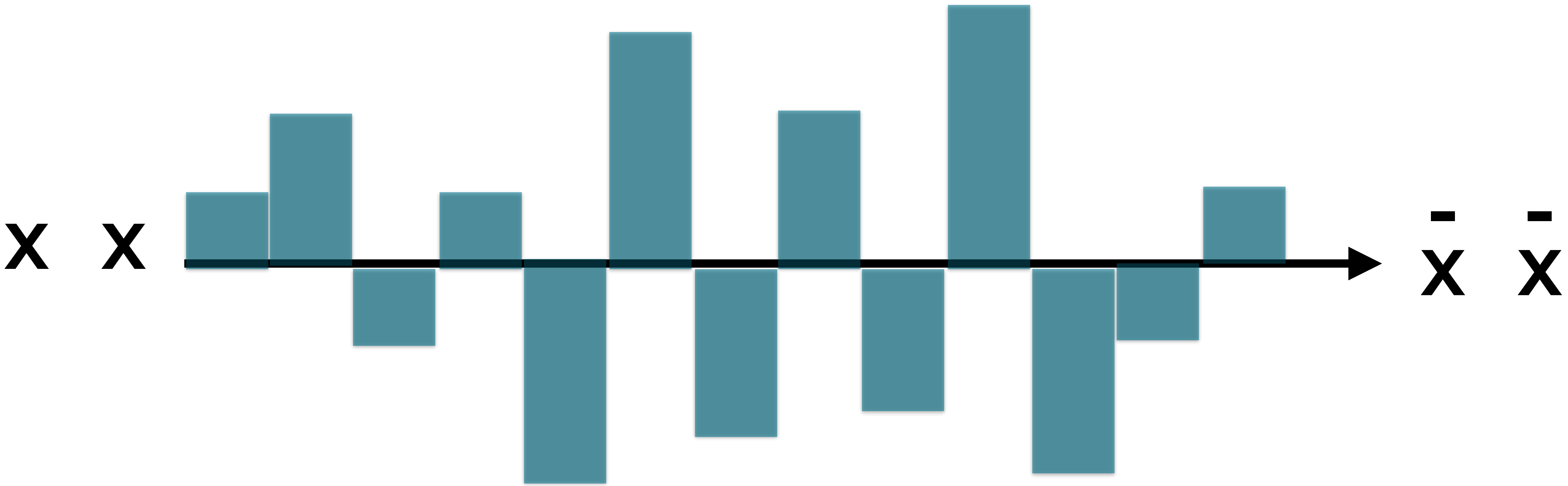
Interest Rates

Industry Sectors

Matrix Multiplication

F_i	$=$	X	v_i
n rows, 1 column		n rows, k columns	k rows, 1 column

Intuition: Positive Covariance



Variance is the covariance of a series with itself

Summary

VBA can be used to add powerful user-generated functions to Excel

Eigen analysis of covariance matrices is easy to implement via VBA

Such analysis of equity returns reveals three important principal components

These closely correlate with underlying economic factors

Regression using these principal components is free of multicollinearity issues