Multivariate: Matrix algebra 2

Table of contents

1 Goals **2**
1.1 Goals . . . 2

2 Maximizing R^2 instead **2**
2.1 Solutions via maximizing . . . 2

3 Geometric representation of vectors **6**
3.1 Geometric representation of vectors . . . 6
3.2 Basis of a space . . . 8
3.3 Independence and orthogonality . . . 10
3.4 Standardization . . . 13
3.5 Geometric representation of correlations . . . 21
3.6 Summary . . . 25

4 Determinant and rank **26**
4.1 Determinant . . . 26
4.2 Rank . . . 33
4.3 Summary . . . 34

5 Eigenvectors and eigenvalues **34**
5.1 Eigenvectors and eigenvalues . . . 34

6 Example and Conclusion **39**
6.1 Example . . . 39
6.2 Summary of this week . . . 43
6.3 Next few weeks . . . 43

1 Goals

1.1 Goals

1.1.1 Goals of this section

- Dimension reduction
  - We have **many** measures of a thing
    - Multi-item scale
    - Multiple sources / reporters
  - How can we reduce the number of variables while keeping as much information as possible?

1.1.2 Goals of this lecture

- Matrices have a lot of parts – how can we understand them?
  - Matrices and vectors as geometric objects
  - Numerical summaries of matrices
- Much of this is in service of assessing:
  - How much independent information we have in a matrix
  - How we can divide up the variance in a matrix
  - Leads into PCA and FA – dimension reduction

2 Maximizing R^2 instead

2.1 Solutions via maximizing

2.1.1 Maximizing the multiple correlation

- Ordinary least squares (OLS) estimation **minimizes** the sum of squared residuals to get regression coefficients
  - This also **maximizes** the multiple correlation
- Other multivariate techniques use **maximizing** functions to find solutions
  - For example, **maximum likelihood estimation**

2.1.2 Maximizing the multiple correlation

- So why don't we **maximize** the multiple correlation instead?
  - Cut out the "middle man" that is minimizing the sum of squared residuals
  - Two **related** reasons
    - *Lack of unique solutions*
    - *Fewer knowns than unknowns*

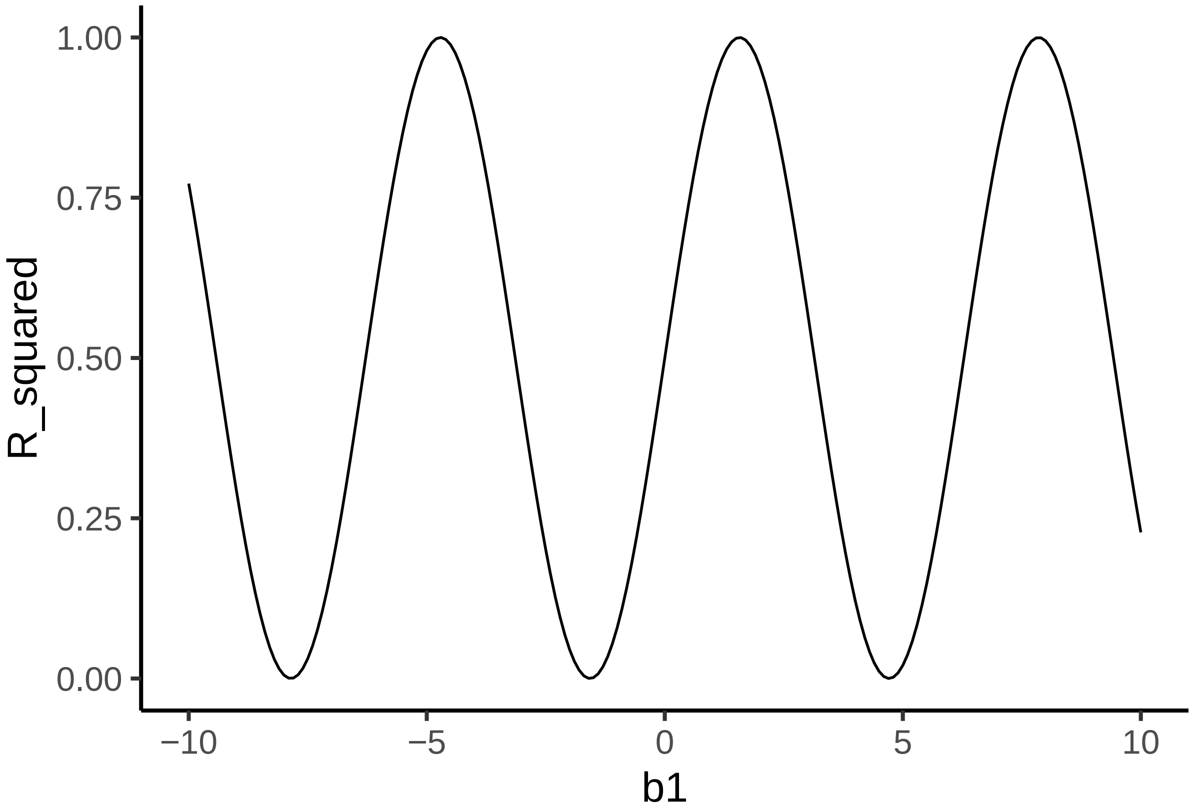
2.1.3 More unknowns than equations

- The lack of uniqueness is due to **more unknowns than equations**
- Two unknowns, two equations = solvable for all unknowns:
  - $y = 2x + 5$
  - $x = 4y$
- Two unknowns, one equation = NOT solvable for all unknowns:
  - $y = 2x + 5$
- If you take an SEM course, this idea is called **identification**

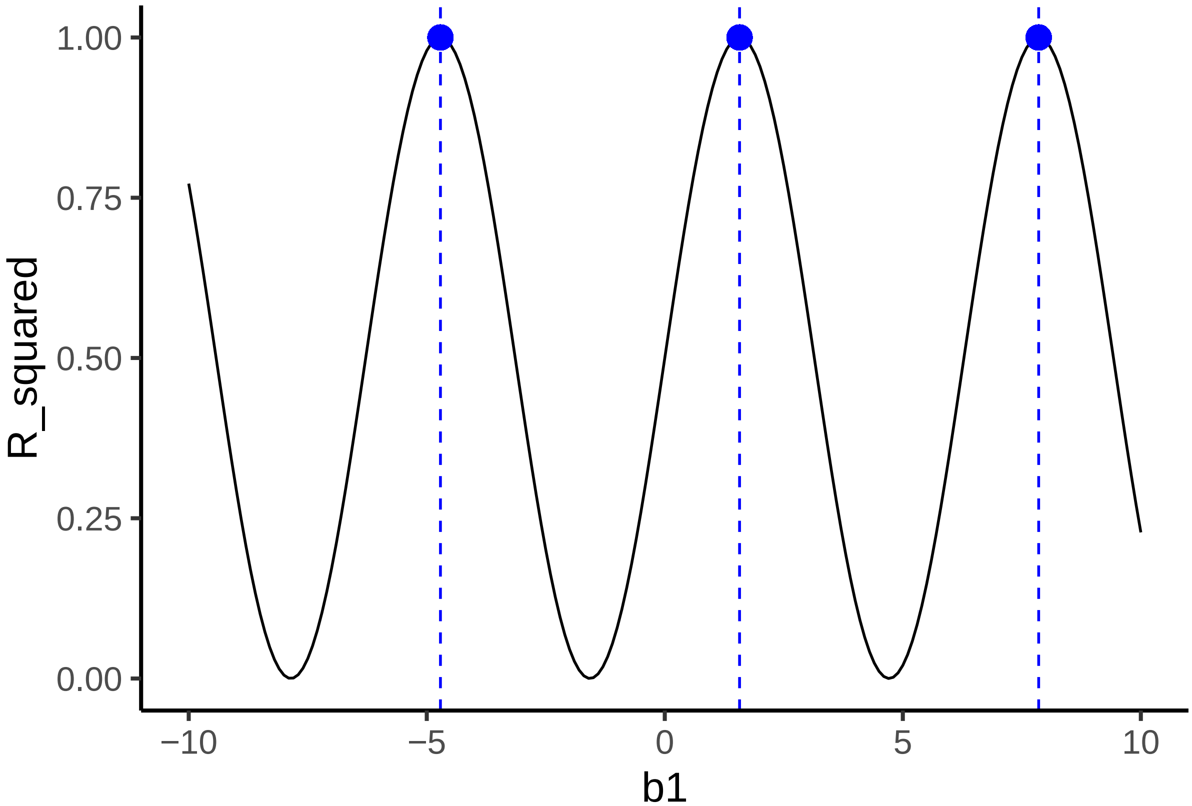
2.1.4 Unique solutions

- The OLS regression weights are unique
  - They are the **ONLY** regression weights that *minimize* the sum of squared residuals
  - They **also** produce the maximum possible multiple correlation
- *But it doesn't work the opposite direction*
  - **Any multiple of the regression coefficients** will also *maximize* the multiple correlation
  - **Only** the least squares weights will **BOTH** minimize the sum of the squared deviations **AND** maximize the multiple correlation

2.1.5 No unique solution



2.1.6 No unique solution



2.1.7 Can we make it happen?

- **Could** we get a unique solution by maximizing $R^2_{multiple}$?
  - Regression coefficients will change if the **scale** (i.e., variance) of *either* Y or $\hat{Y}$ changes
  - BUT if we fix or **constrain** the variance of both Y and $\hat{Y}$, we get a unique solution
- Approach: we want to **simultaneously**
  - **Maximize** the multiple correlation
  - **Constrain** the variance of Y and $\hat{Y}$

2.1.8 Constraints on the solution

- **Constrain** means that we set some part of the model to a value **instead of estimating it**

- Regression
  - Fix the variance of both Y and $\hat{Y}$ to 1
  - Include a **scaling constant** (covariance between Y and $\hat{Y}$)
- We will use this *general approach* for other methods, such as factor analysis

3 Geometric representation of vectors

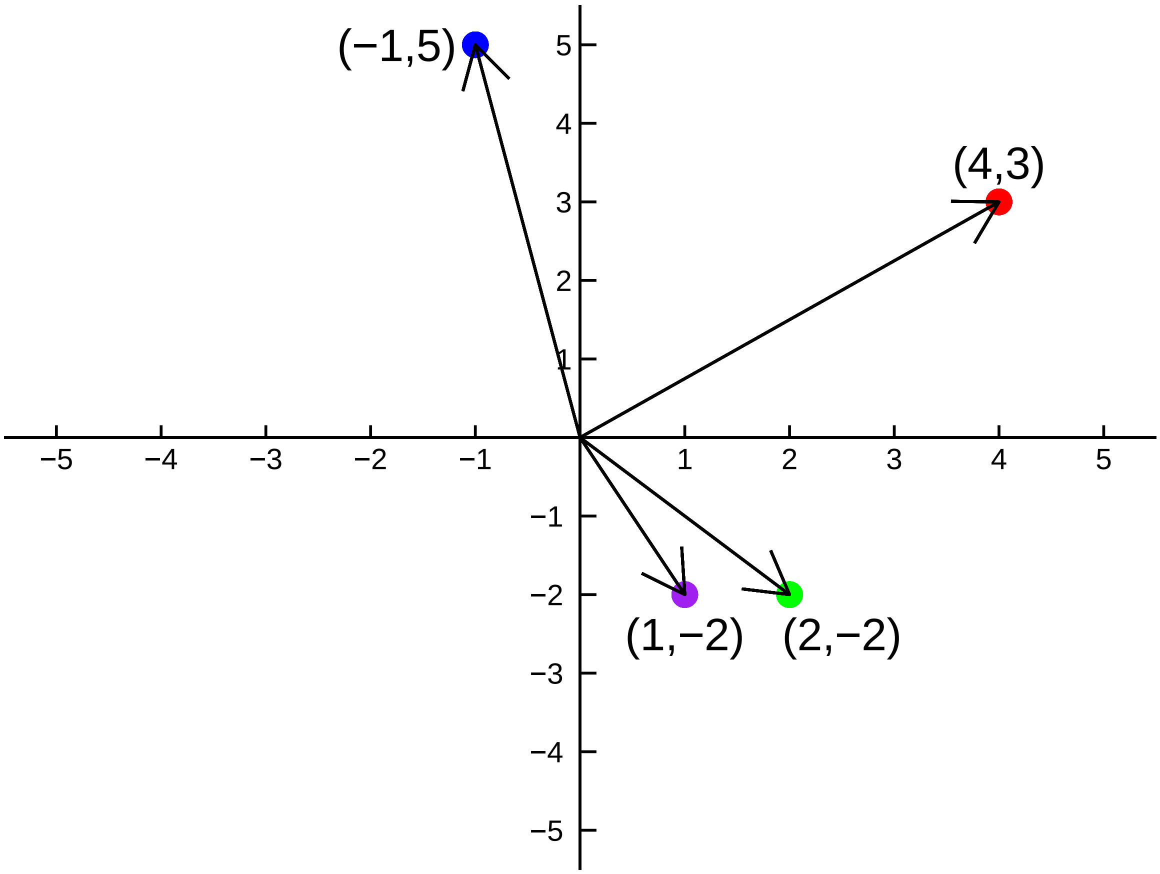
3.1 Geometric representation of vectors

3.1.1 Vectors as geometric objects

- **Vectors** have an **algebraic** interpretation
  - Add, subtract, multiply
- **Vectors** also have a **geometric** interpretation
  - We can think about vectors as **objects in space**
  - This can help us think about the **structure** of data
- PCA, factor analysis: large number of variables reduced to a smaller number of dimensions
- Regression diagnostics: distance between points in space

3.1.2 Vectors as lines in space

- Represent any vector as a **line** from the **origin** (0,0) to a **point**

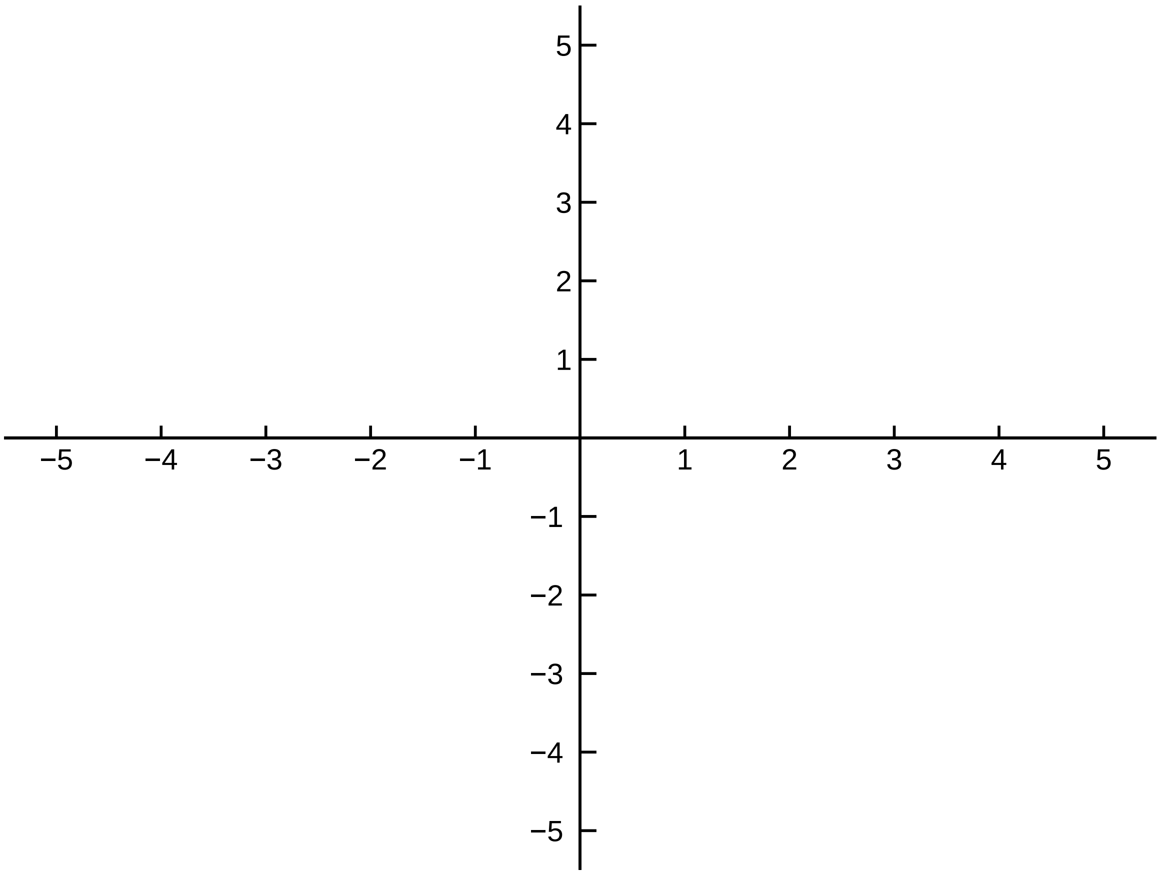


3.1.3 Vectors have a direction and length

- Every vector has a **direction** and a **length**
  - The *direction* of a vector is *where it is*
  - The *length* of a vector will become more relevant when we talk about **standardization**

3.2 Basis of a space

3.2.1 Define a space: 2D



3.2.2 Three dimensional space

- X, Y, and Z axes represent a 3 dimensional space
- The three axes can be written as **vectors**:
  - X-axis: $(1, 0, 0) = \underline{e}_1'$
  - Y-axis: $(0, 1, 0) = \underline{e}_2'$
  - Z-axis: $(0, 0, 1) = \underline{e}_3'$
  - These are **standard axes** of **unit length**

3.2.3 Reference axes

- Reference axes are the **basis** of a space
  - All vectors in a space can be **created from** the reference axes
  - Specifically, a **composite** or **linear combination** of them
- References axes need to be **linearly independent**
  - More on linear dependence / independence in a few minutes
  - We are used to thinking of **orthogonal** (i.e., right angle) axes
    - Orthogonal axes **are** linearly independent, but non-orthogonal axes can be linearly independent too

3.2.4 Basis of a space example

- Test that measures these three uncorrelated abilities
  - Test score is a **composite** or **linear combination**
  - 1 part X, 2 parts Y, 2 parts Z
- This composite is represented by a **vector** $\underline{a}' = (1, 2, 2)$
  - In the standard references axes, the test can be represented as
    - $\underline{a}' = 1 \times \underline{e}_1' + 2 \times \underline{e}_2' + 2 \times \underline{e}_3'$

3.2.5 Basis summary

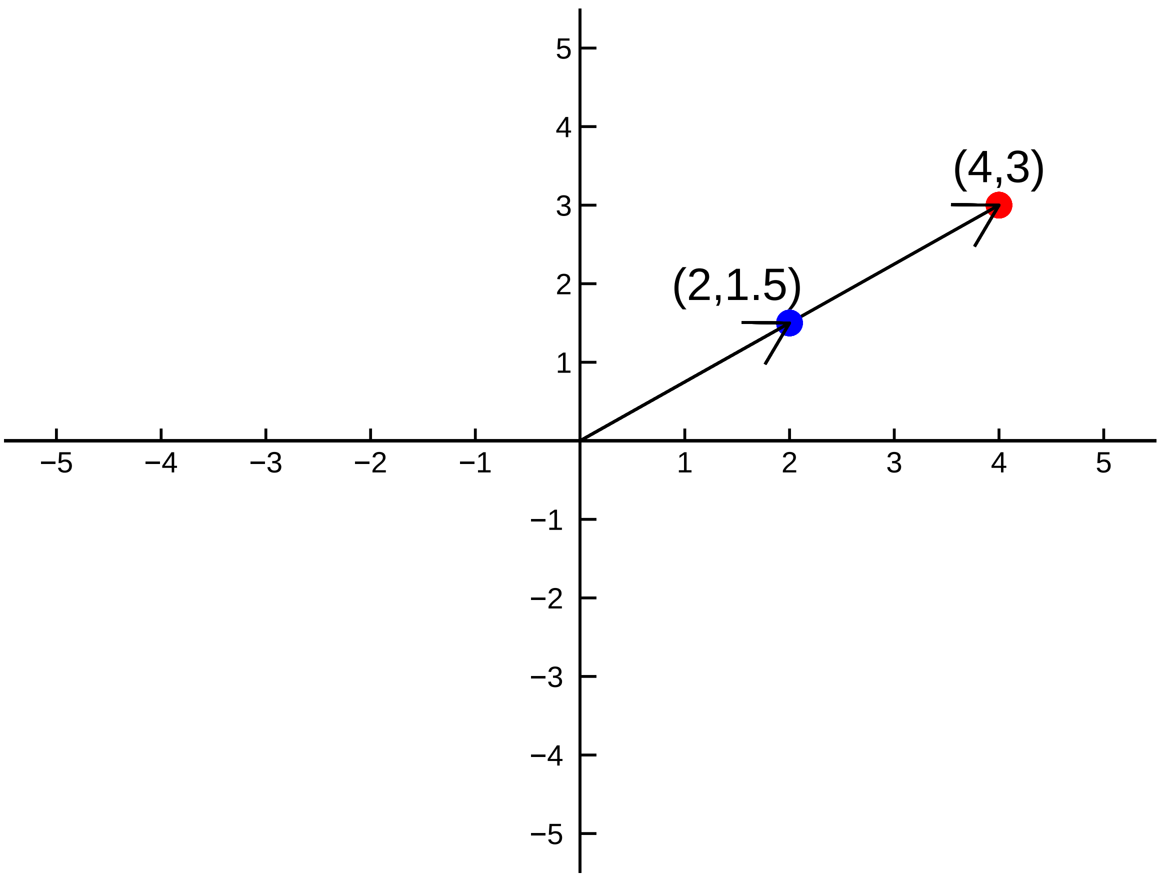
- Data (i.e., vectors and matrices) can be represented **geometrically**
- We have to **define** our geometric space
  - Often, we use reference axes (X, Y, Z), which are orthogonal
  - But we don't have to
- Anything in the space is a function of the axes

3.3 Independence and orthogonality

3.3.1 Linear dependence and linear independence

- **Linear independence**: no vectors are *multiples* or *sums* of another
- **Linear dependence** (also called "collinearity"): they **are**
  - $\begin{bmatrix} 1 & 1 & 2 \\ 3 & 4 & 7 \\ 1 & 3 & 4 \end{bmatrix}$: Third column is the sum of the first two columns
  - $\begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix}$: Second column is exactly 2 times the first column

3.3.2 Collinear vectors: (4,3) and (2, 1.5)



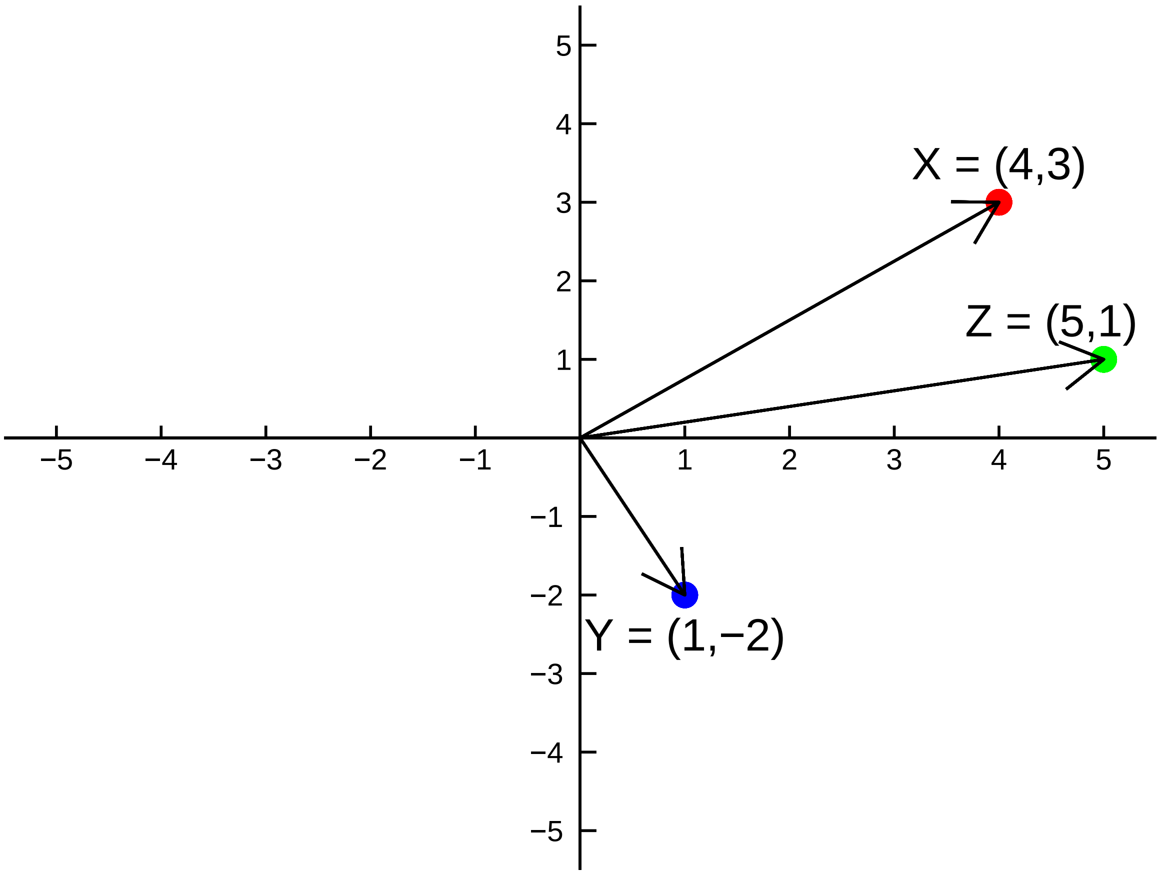
3.3.3 Orthogonality

- References axes (basis) need to be **linearly independent**
  - No vector is a multiple or sum of others
- The standard reference axes are also **orthogonal**
  - **Orthogonal** = *perpendicular* or *right angle*
    - X, Y axes in 2 dimensions
    - X, Y, Z axes in 3 dimensions
    - Extends to 4+ dimensions

3.3.4 Oblique dimensions

- **Oblique axes** are **not orthogonal**
  - Not at right angles
- Oblique axes can be used as reference axes
  - They just been to be **linearly independent**

3.3.5 Oblique axes



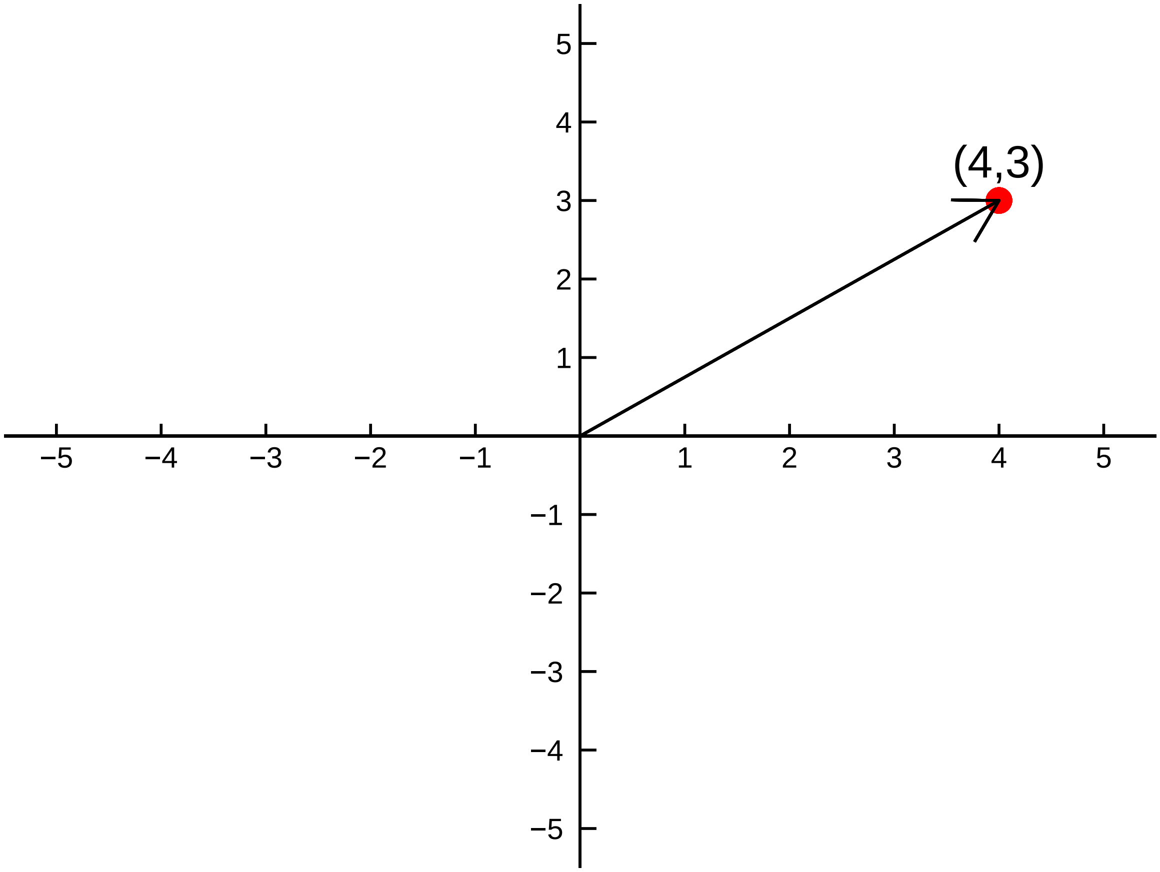
- X and Y are **not orthogonal**, but they are **linearly independent**
- Z is **not linearly independent** of X and Y: 4 + 1 = 5 and 3 - 2 = 1
- Using all 3 axes would result in **linear dependence**

3.3.6 Basis and dimension

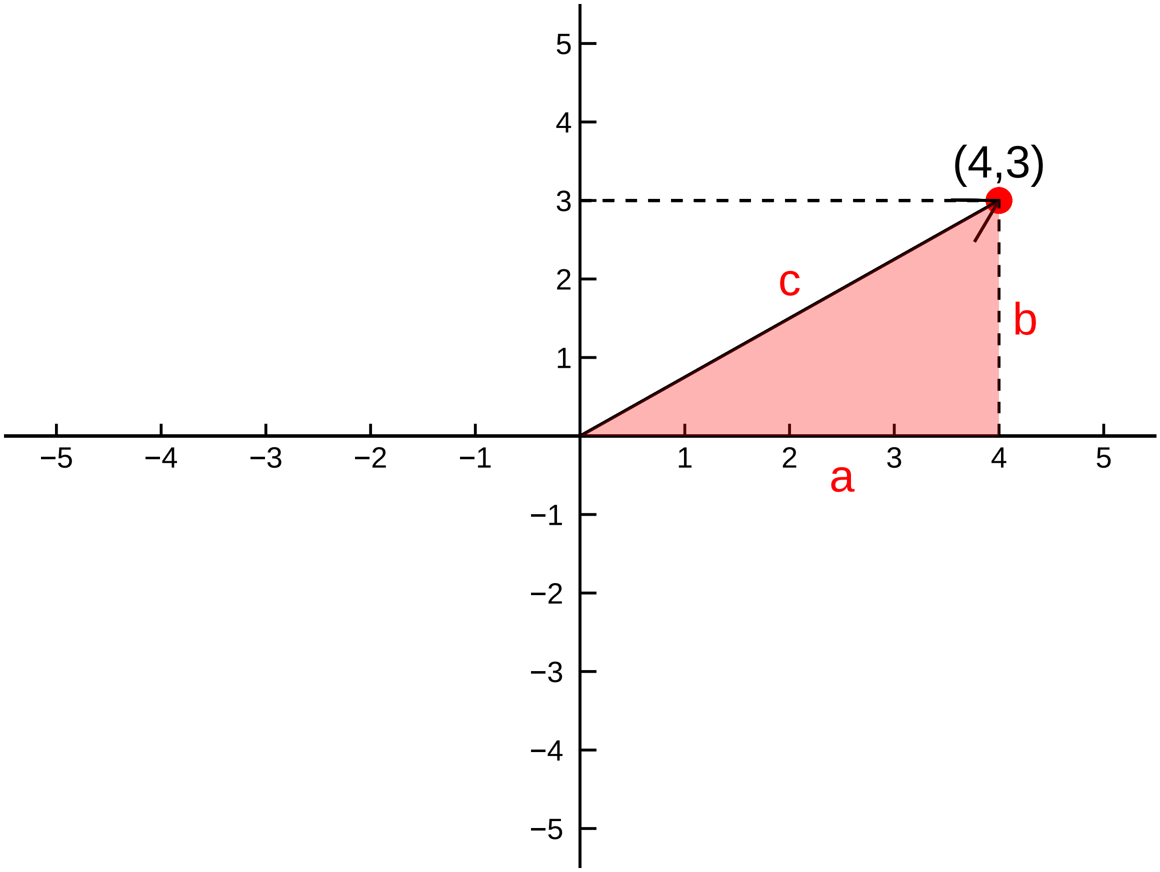
- 2 dimensions in a flat plane, so the basis can be any 2 linearly independent vectors
- 3 dimensions in a space, so the basis can be any 3 linearly independent vectors
- Same for more dimensions, only it's harder to imagine it
- Use this in PCA and factor analysis to reduce *many measures* to *fewer linearly independent vectors* (factors)

3.4 Standardization

3.4.1 Length of a vector



3.4.2 Length of a vector: Algebra



- Pythagorean theorem
  - $a^2 + b^2 = c^2$
  - $c = \sqrt{a^2 + b^2}$
- $length = \sqrt{a^2 + b^2}$
  - $length = \sqrt{4^2 + 3^2} = 5$

3.4.3 Length of a vector: Matrix

- $\underline{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$
- The length of $\underline{z}$ is:
  - $length(\underline{z}) = ||\underline{z}|| = \sqrt{z_1^2 + z_2^2} = (\underline{z}'\underline{z})^{1/2}$

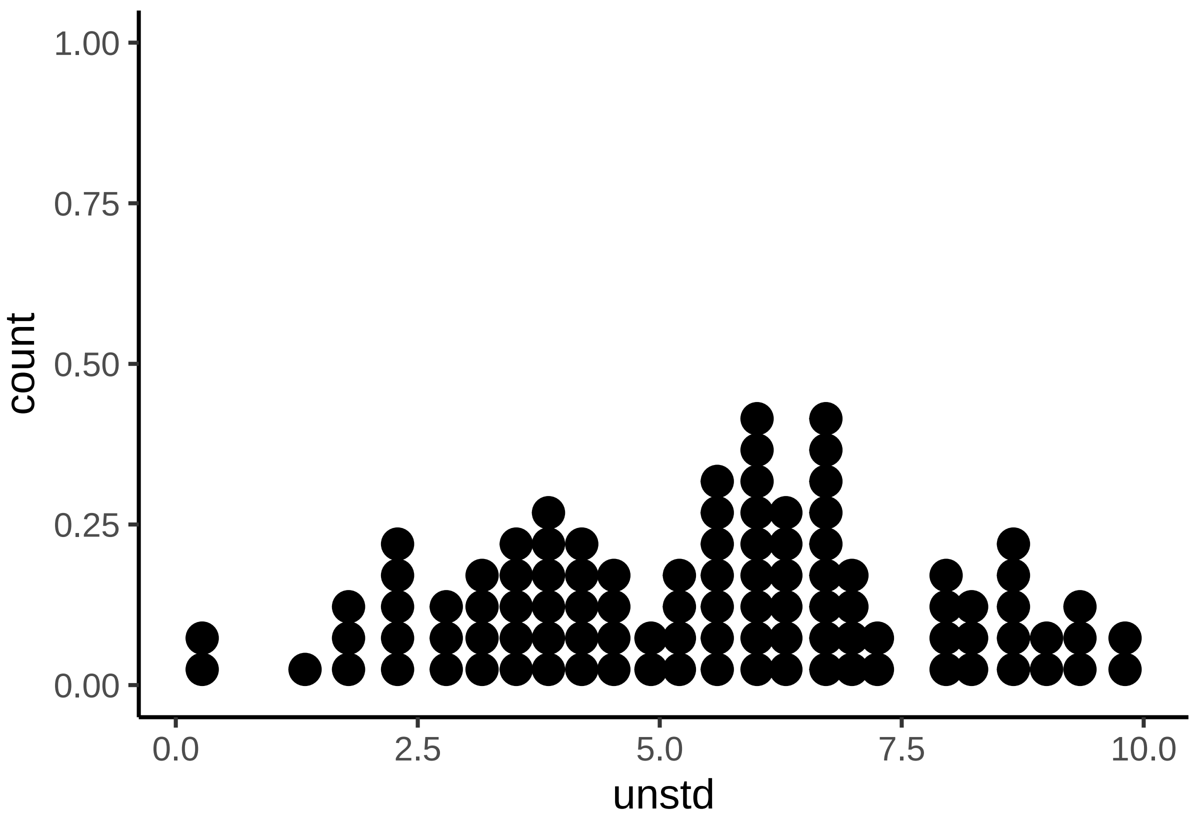
  - Last expression generalizes to more than 2 dimensions

3.4.4 Standardization

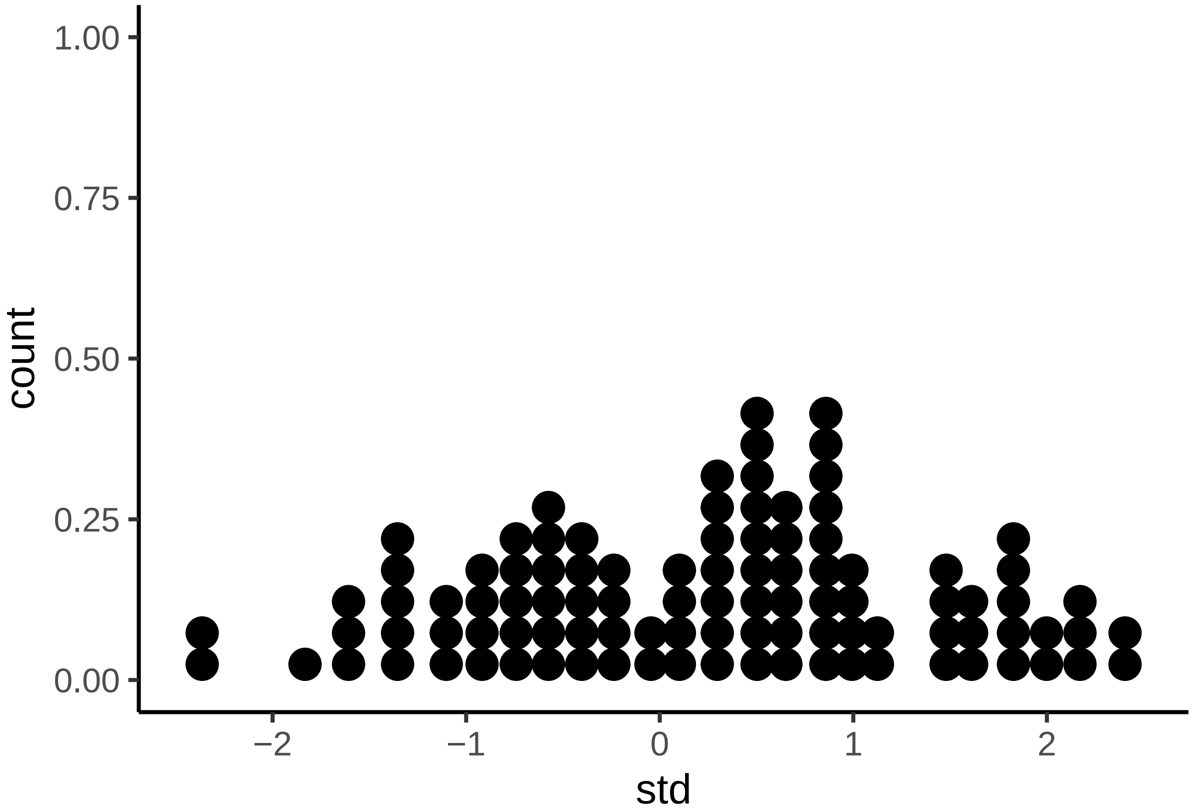
- Standardizing variables:
  - Subtract mean and **divide by standard deviation**
  - For standardized variable: Mean = 0 and variance = 1 (and SD = 1)
  - **Changes the variance to 1**
- But standardization **doesn't change**:
  - Overall *shape* of the distribution
  - *Relations* with other variables

3.4.5 Standardization

Mean = 5, SD = 2 (Variance = 4)

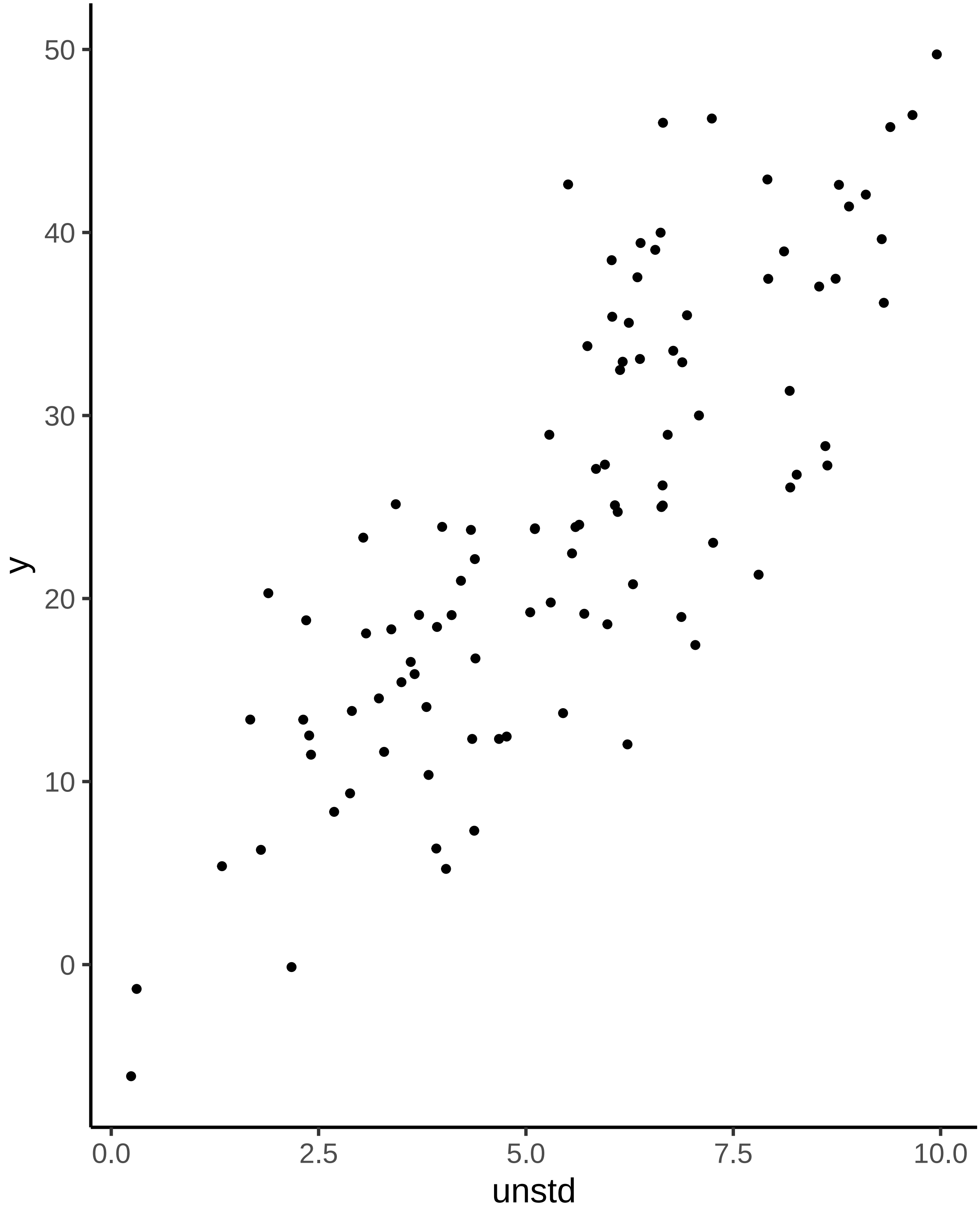


Mean = 0, SD = 1 (Variance = 1)

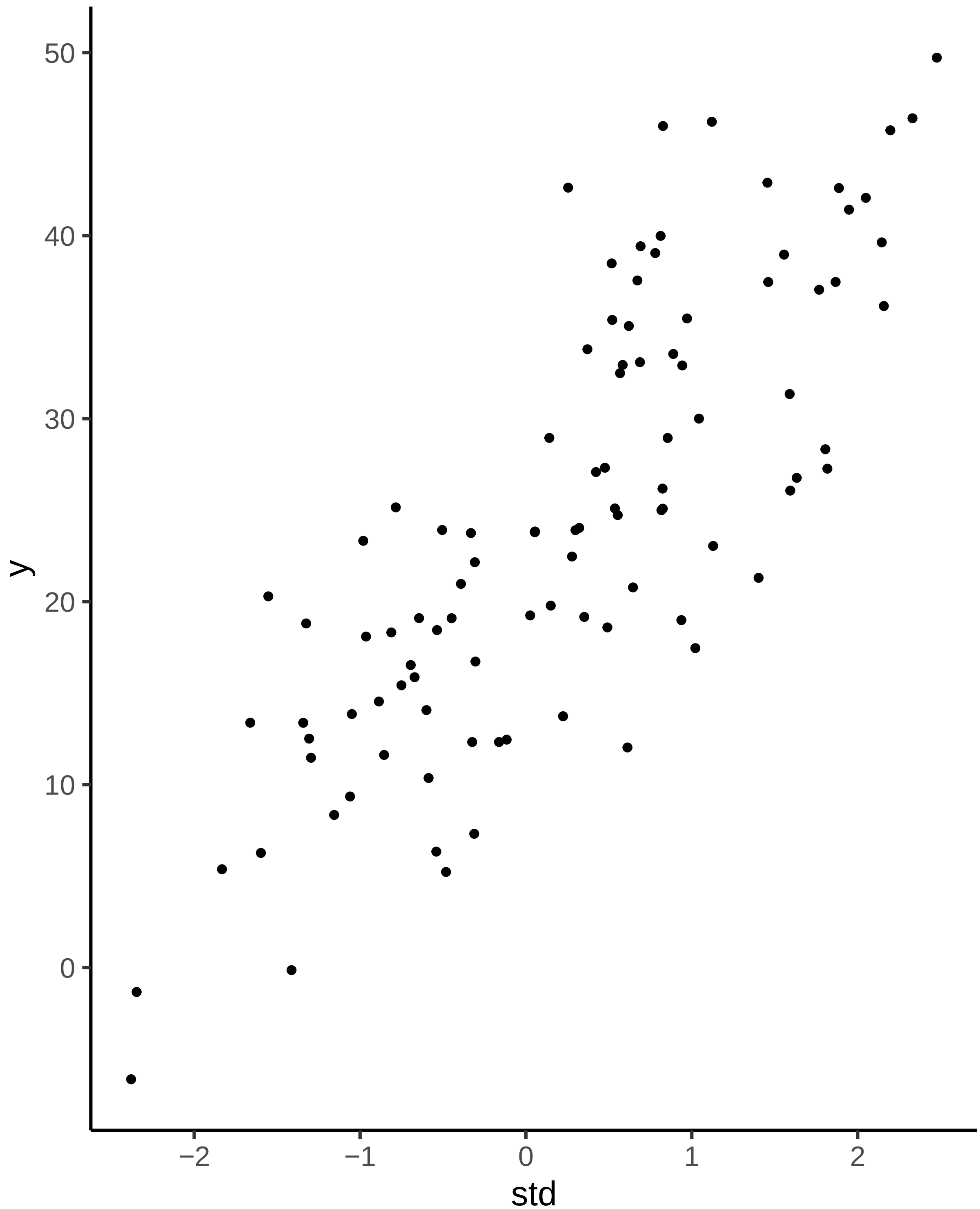


3.4.6 Standardization

Mean = 5, SD = 2 (Variance = 4)



Mean = 0, SD = 1 (Variance = 1)


50
40
30
20
10
0
y
−2
−1
0
1
2
std

3.4.7 Standardization

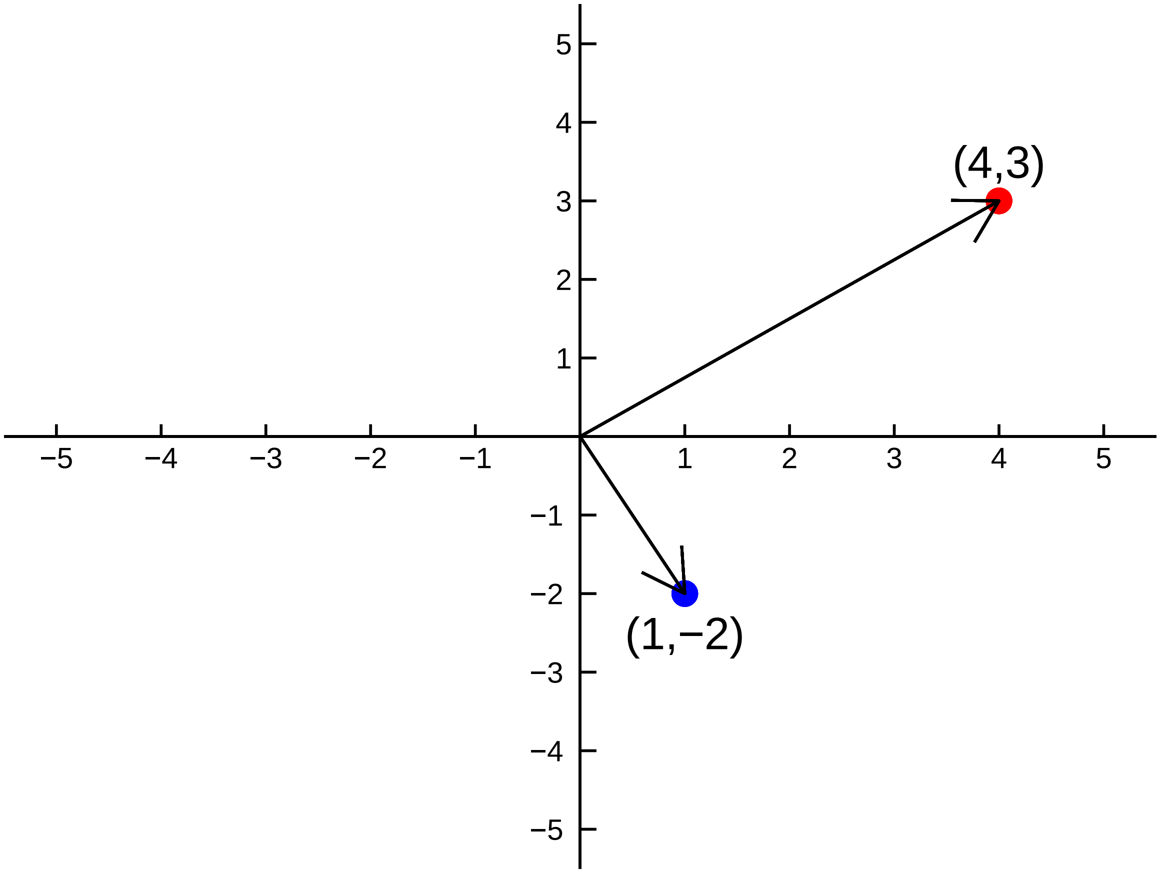
- Standardizing vectors is about the **length of the vector**
  - **Change the length of a vector to 1**
- Vector: $\underline{z} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$
- Length of vector: $||\underline{z}|| = \sqrt{4^2 + 3^2} = 5$

3.4.8 Standardization

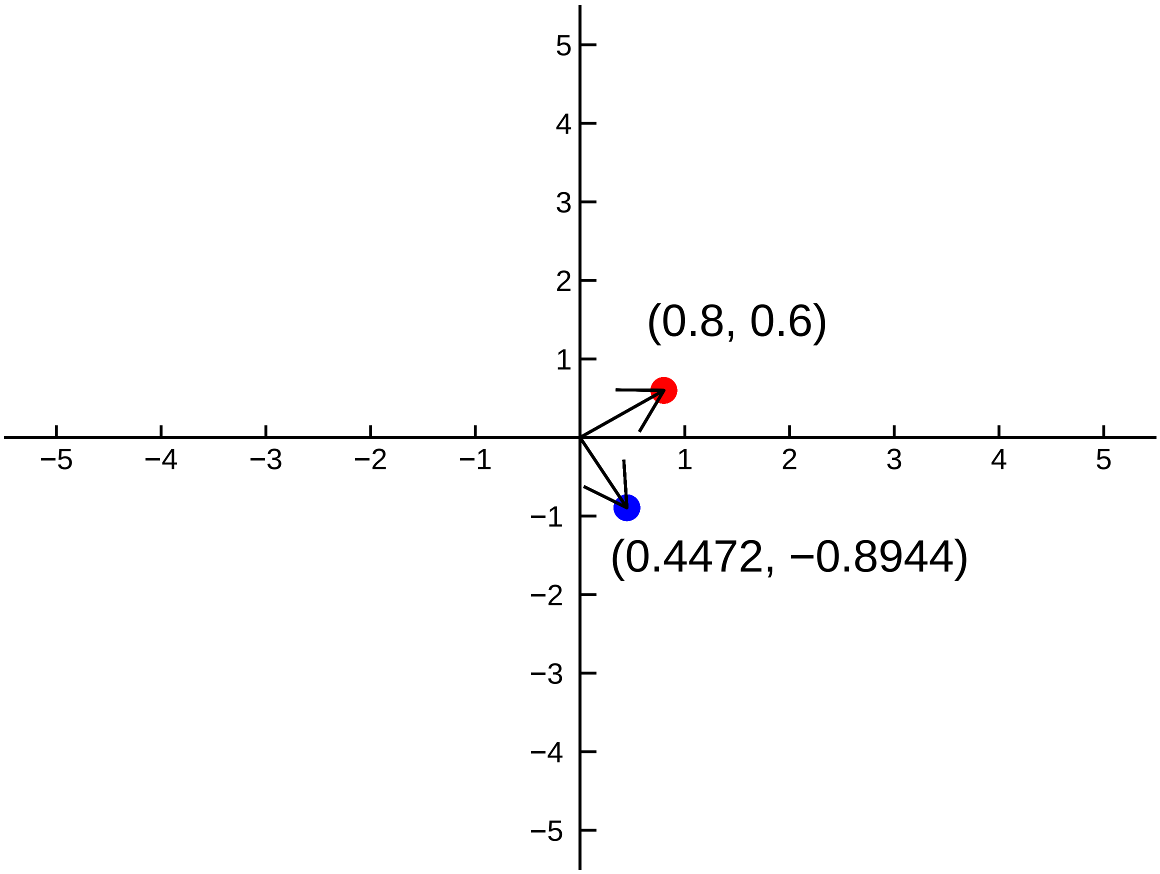
- To standardize, **divide each element by the length** of the vector:
  - Vector: $std\ \underline{z} = \begin{bmatrix} \frac{4}{5} \\ \frac{3}{5} \end{bmatrix}$
  - Length of vector: $||std\ \underline{z}|| = \sqrt{\left(\frac{4}{5}\right)^2 + \left(\frac{3}{5}\right)^2} = 1$
- The **length** of the vector $\begin{bmatrix} \frac{4}{5} \\ \frac{3}{5} \end{bmatrix}$ is 1
  - But its **direction** is **unchanged**

3.4.9 Unstandardized and standardized axes

Length = 5 and 2.236



Length = 1 and 1

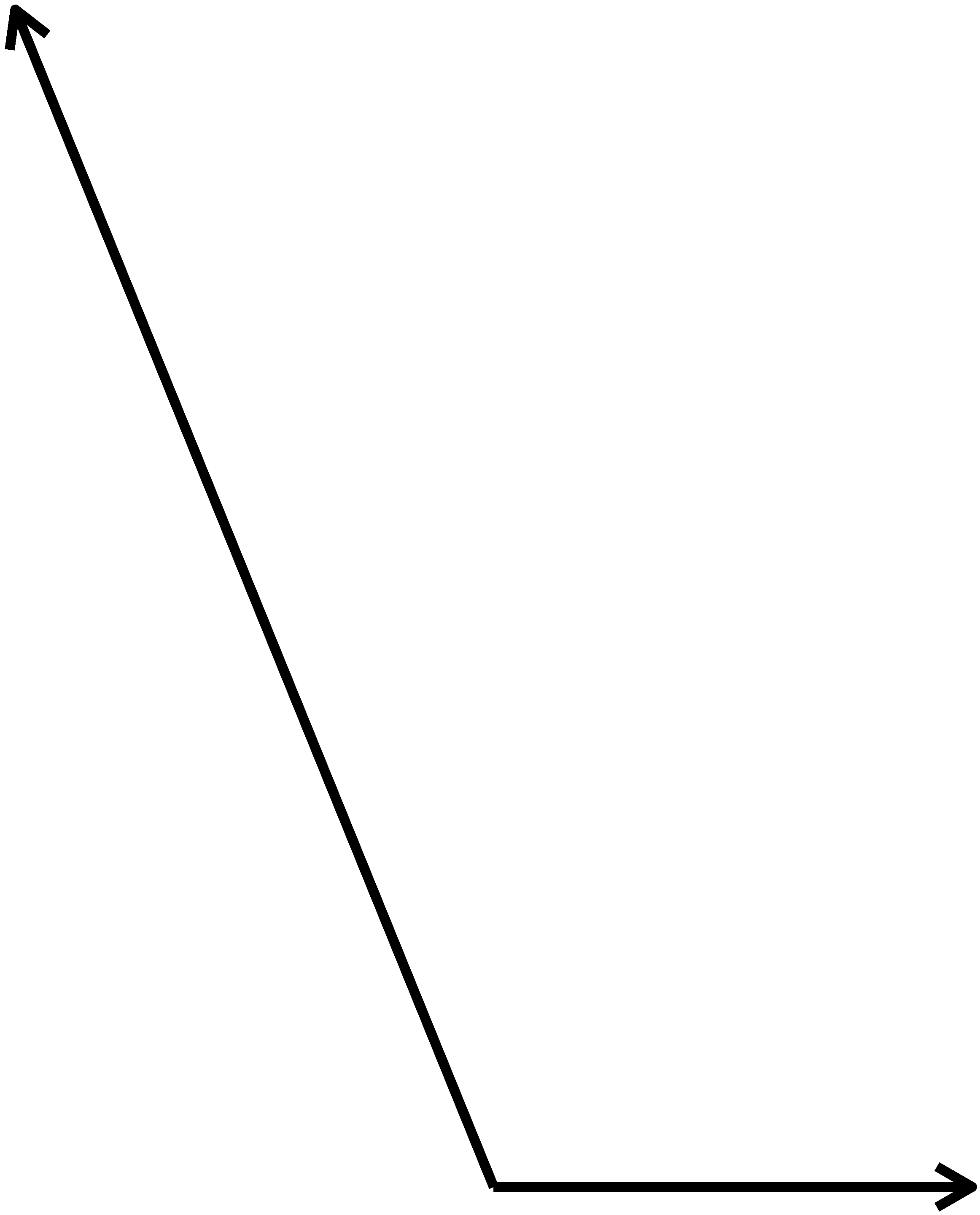


3.5 Geometric representation of correlations

3.5.1 Geometric representation of correlations

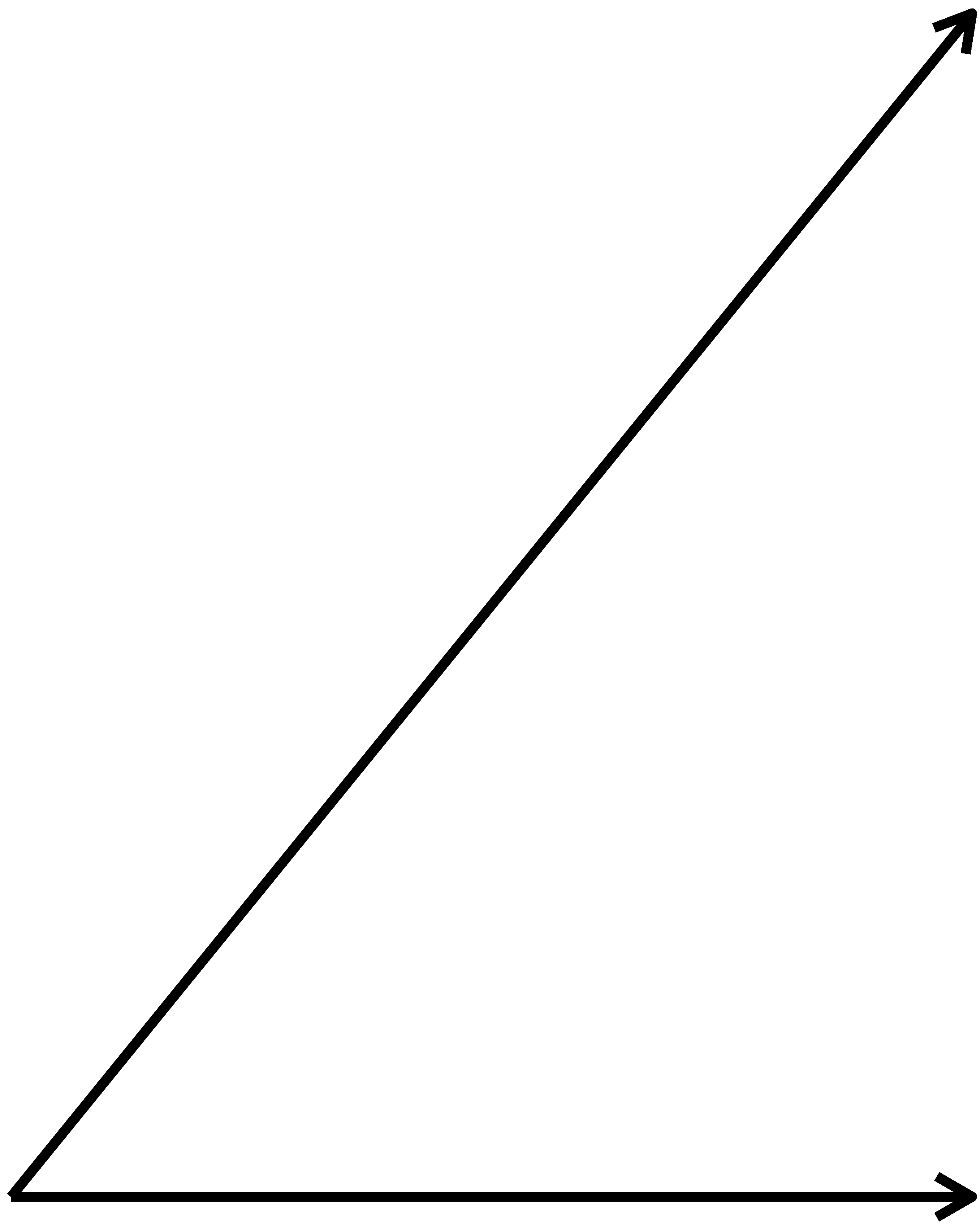
The **angle** between vectors reflects their **correlation**

Angle > 90: $r \to$ -1

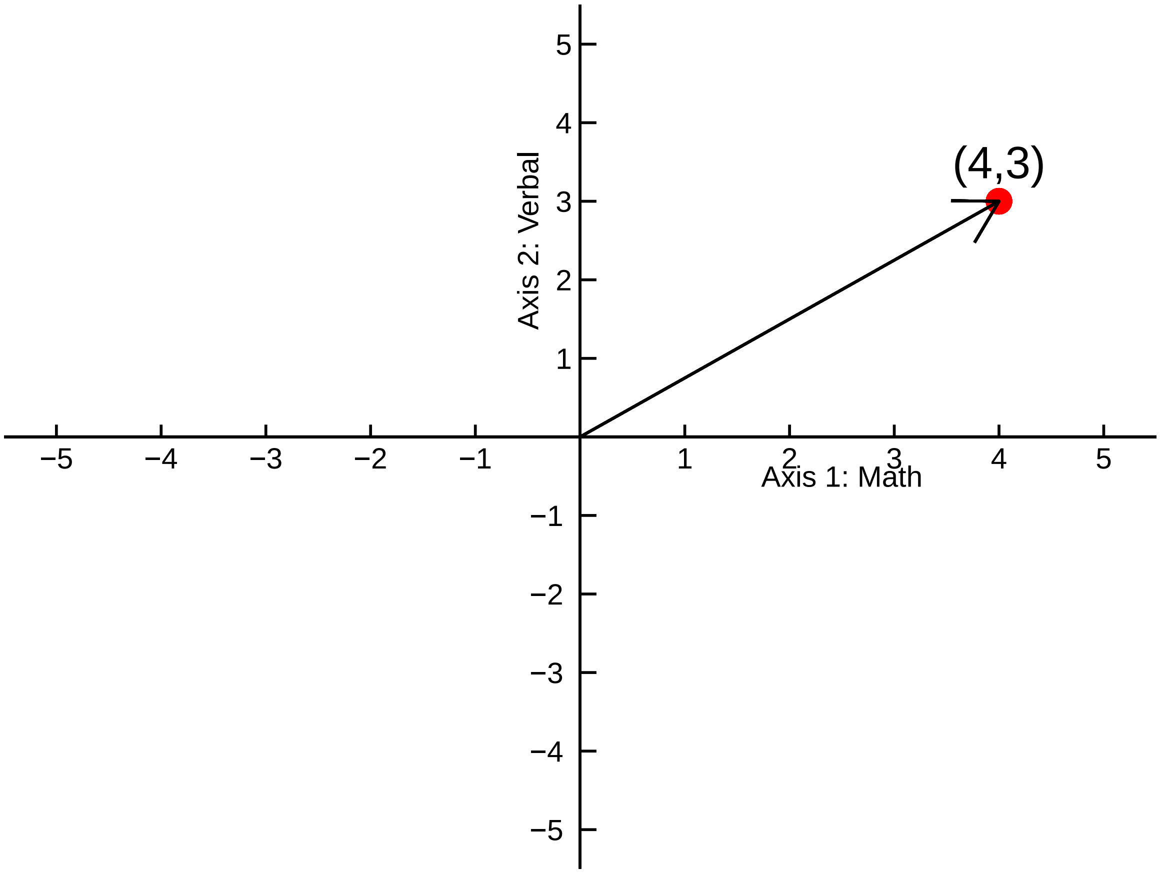


Angle = 90: $r = 0$

Angle < 90: $r \to +1$



3.5.2 Correlation with axis



3.6 Summary

3.6.1 Summary: Vectors and geometry

- **Vectors** are geometric objects with **direction** and **length**
  - **Length** of a vector is its **scale**
    * Standardize by dividing by the length (i.e., length = 1)
- **Reference axes** (like X-Y axes) form the **basis** of a space
  - Things in the space are **linear combinations** of reference axes
- Reference axes need to be **linearly independent** (NOT collinear) and **may be orthogonal** (but don't need to be)
- **Angles** between vectors and/or axes represent their **correlation**

4 Determinant and rank

4.1 Determinant

4.1.1 Multivariate = multiple variables

- Multivariate means many variances and covariances
    - Hard to look at a large matrix and get information out of it
- Determinant of a matrix does **2 things**:
    1. Screens for **linear dependency**
    2. **Summarizes** all variance in the matrix with **one number** ("generalized variance")
- Can get the determinant of any **square matrix**
    - More on that in a minute

4.1.2 Geometric interpretation of a determinant

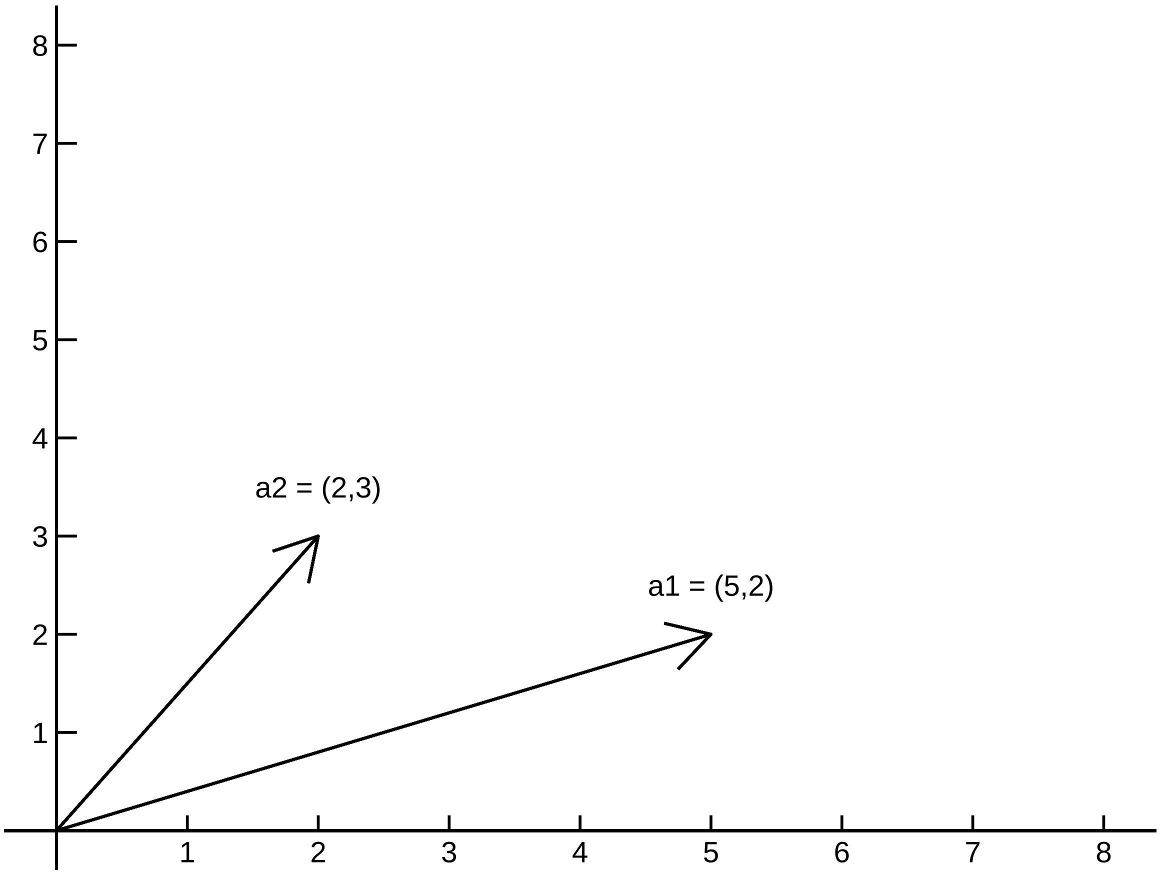
- For the simplest case of 2 dimensions

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$$

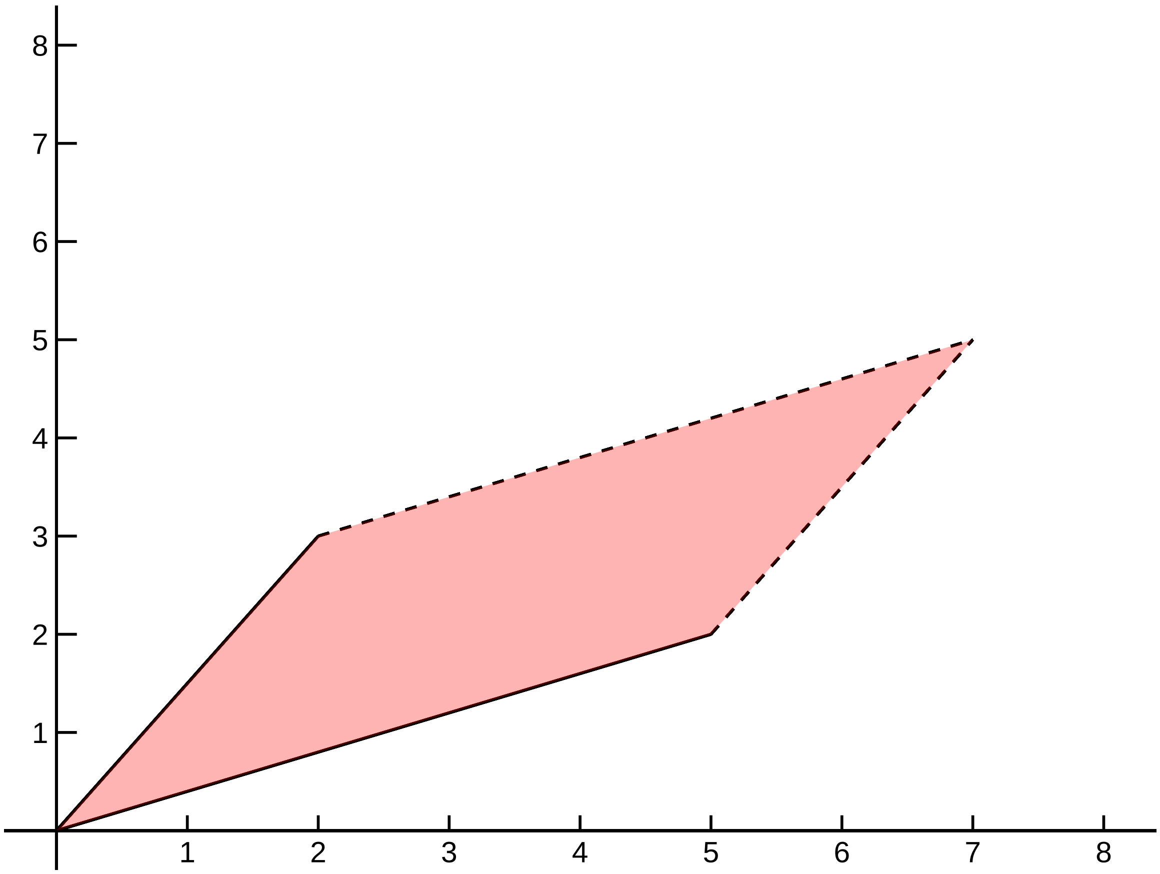
Vector 1: $\underline{a}_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

Vector 2: $\underline{a}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

4.1.3 Geometric interpretation of a determinant



4.1.4 Geometric interpretation of a determinant

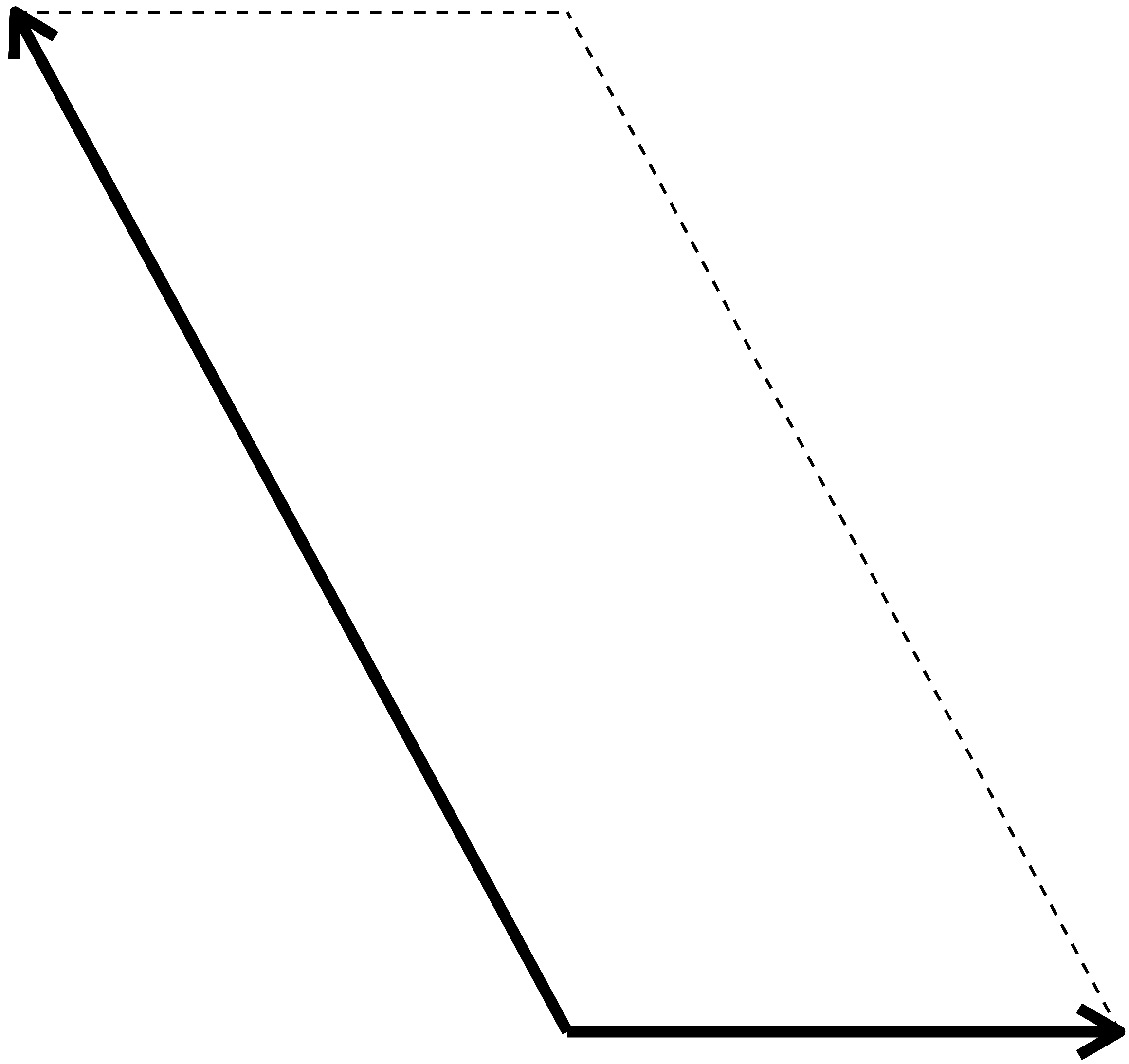


4.1.5 Determinant

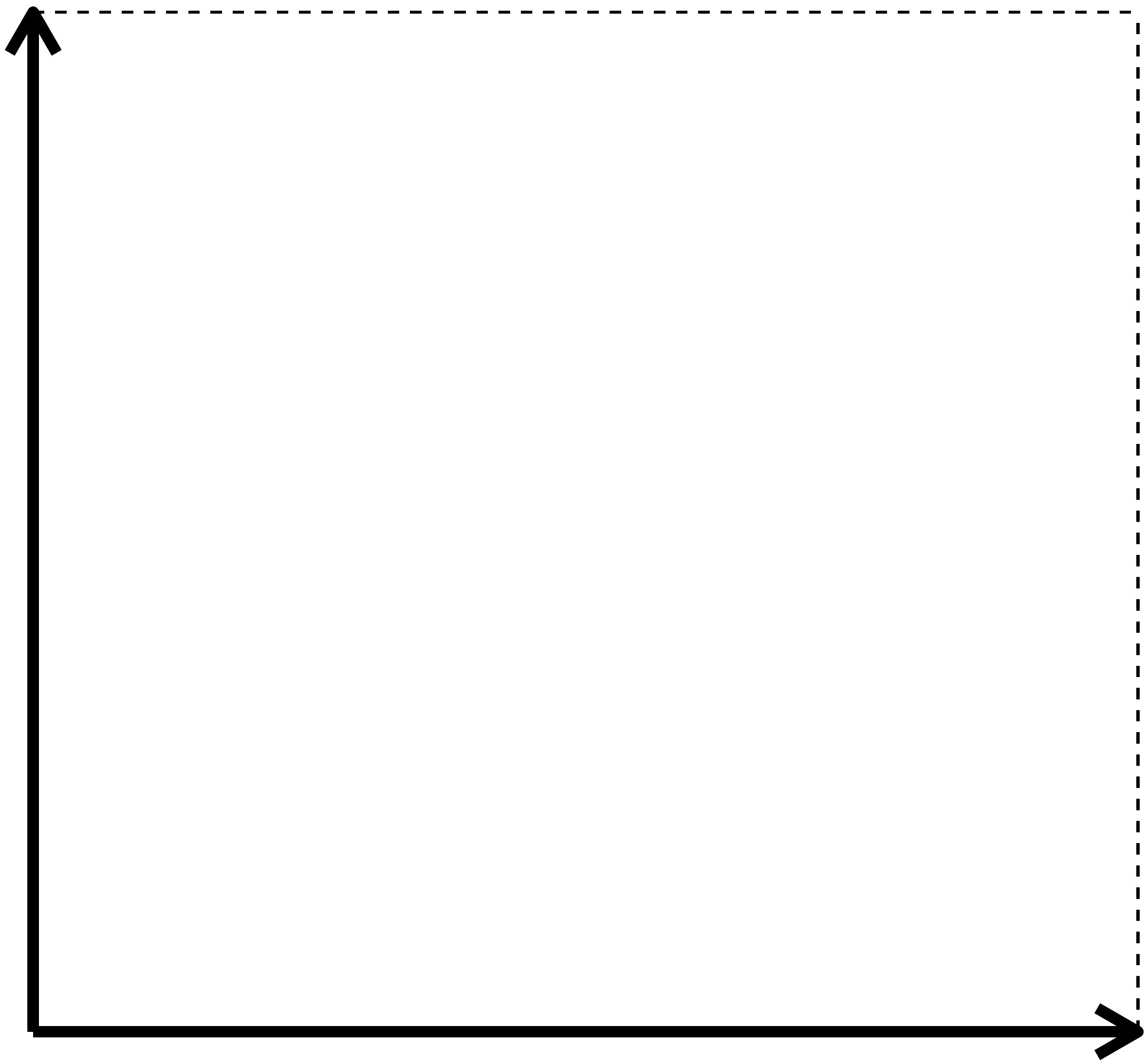
- Determinant is the **area of the parallelogram** created by the vectors
  - **Larger area** for parallelogram as vectors approach **90 degrees**
  - **Smaller area** for parallelogram as vector approach **0 degrees** or **180 degrees**
- Determinant = 0 when vectors are linearly dependent
  - Determinant is *close to 0* when vectors are highly correlated

4.1.6 Determinant

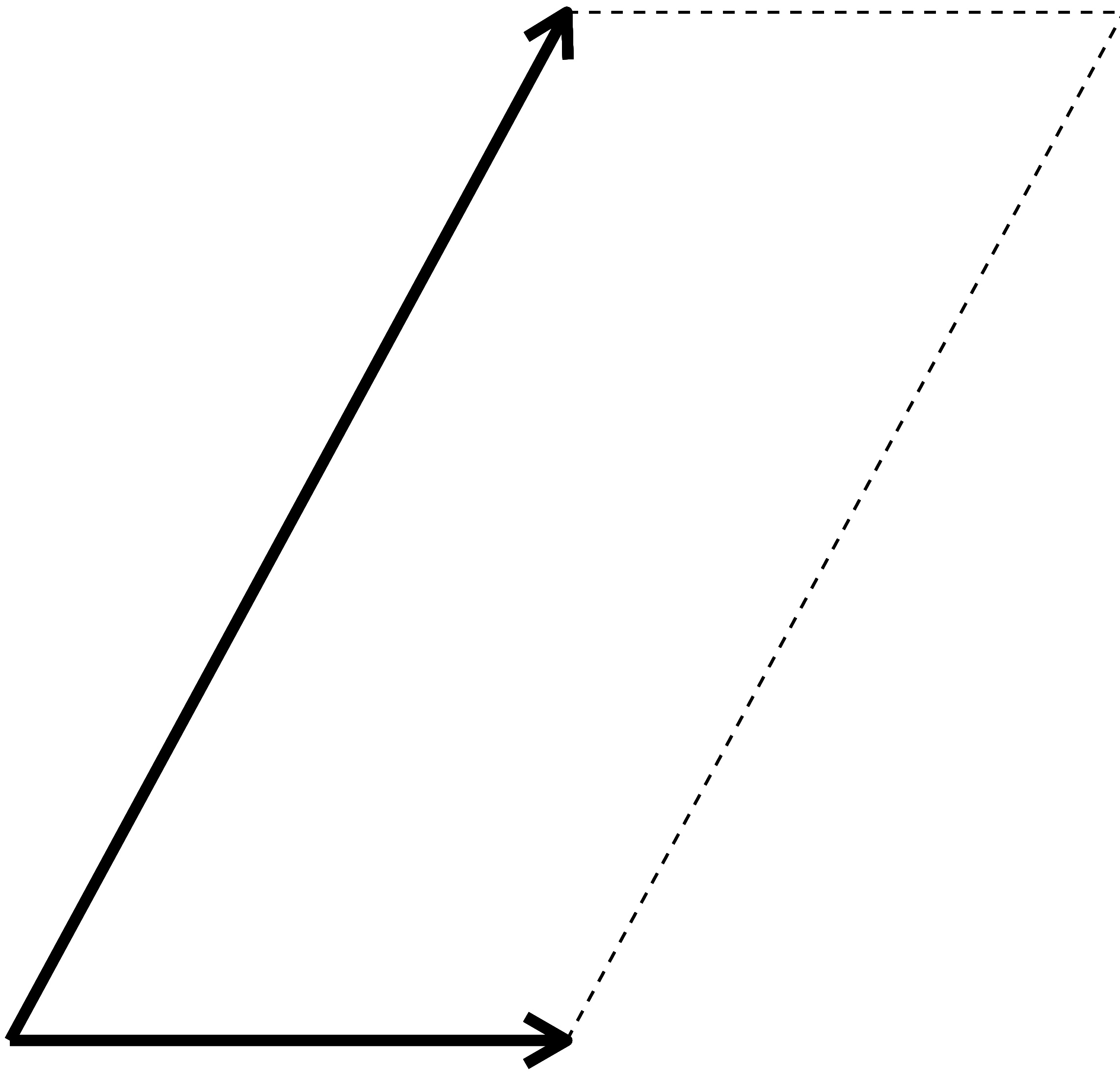
r approaches -1



$r = 0$



r appraoches $+1$



- Easy to **see** for only 2 vectors, but as number of vectors increases, need to use the determinant

4.1.7 Determinant

- Determinant can be calculated for any **square matrix**
  - But the data matrix is $n \times p$ (typically not square)

$$\mathbf{Q} = \underset{(n,p)}{\mathbf{X}} \; \underset{(p,n)}{\mathbf{X}'}$$

- If determinant($\mathbf{Q}$) = 0, then linear dependence in $\mathbf{X}$
  - Can also use any other square matrix based on $\mathbf{X}$, like $\mathbf{P}_{XX}$, $\mathbf{S}_{XX}$, $\mathbf{R}_{XX}$

4.1.8 Calculating the determinant

- The determinant uses **all elements in the matrix** to provide a summary of the relationships in the matrix
- For a 2×2 matrix, the determinant is straightforward:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$
$$det(\mathbf{A}) = |\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$$

4.1.9 Determinant and correlations

With a 2×2 matrix, it's easy to see how the determinant relates to correlations between variables

Highly correlated variables

$\mathbf{R}_{XX} = \begin{bmatrix} 1 & 0.99 \\ 0.99 & 1 \end{bmatrix}$

$|\mathbf{R}_{XX}| =$

$1 \times 1 - 0.99 \times 0.99 =$

$1 - 0.9801 = 0.0199$

Moderately correlated variables

$\mathbf{R}_{XX} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

$|\mathbf{R}_{XX}| =$

$1 \times 1 - 0.5 \times 0.5 =$

$1 - 0.25 = 0.75$

4.1.10 Determinant in linear regression

- **What does this have to do with regression?**
  - Inverse of covariation matrix: $\mathbf{P}_{XX}^{-1} = \frac{1}{|\mathbf{P}_{XX}|}\mathcal{A}_{\mathbf{P}_{XX}}$
  - where $\mathcal{A}_{\mathbf{P}_{XX}}$ is the "adjoint matrix" of $\mathbf{P}_{XX}$
- If there is **linear dependence** in $\mathbf{X}$:
  - Determinant of $\mathbf{P}_{XX} = 0$
  - Divide by 0 to get inverse of $\mathbf{P}_{XX}$
  - Can't get inverse of $\mathbf{P}_{XX}$
  - **Can't solve for regression coefficients**

4.1.11 Error messages

- If there is **linear dependency** (or just highly correlated variables) in your regression, **you will get an error message**
- The message varies depending on the program and the procedure
  - **Linear dependence** present in data matrix
  - **Determinant** approaching 0
  - Predictor matrix **cannot be inverted**
  - Data matrix is **rank deficient**
  - Predictor matrix is **not of full rank**
  - Predictor matrix is **singular**
  - Predictor matrix is **ill conditioned**
  - The matrix is **not positive definite**

4.2 Rank

4.2.1 Rank of a matrix

- **Rank** of a matrix is *related* to the *determinant*
  - Number of **independent pieces of information** in a matrix
- **Maximum rank** of a matrix = lesser of # of rows and # of columns
  - Maximum rank = "full rank" = "nonsingular"
- **Linear dependence** means there is **less information** in the matrix than there appears
  - The matrix is "rank deficient" or "singular"

4.3 Summary

4.3.1 Summary: Determinants and rank

- **Determinant** tells you if any variables are **linearly dependent**
- Determinant summarizes the matrix with one number
- If any vectors in the matrix are highly correlated (i.e., linearly dependent), the determinant is close to 0
- If determinant = 0, cannot solve for regression coefficients
- Matrix with determinant = 0 is **rank deficient**

5 Eigenvectors and eigenvalues

5.1 Eigenvectors and eigenvalues

5.1.1 Eigenvectors and eigenvalues

	Eigenvector	Eigenvalue
they are	vector	scalar
they are	reference axes	amount of variance that is associated with that reference axis
also called	characteristic vector	characteristic root
also called	latent vector	latent root

5.1.2 Motivation for eigenvectors and eigenvalues

- We want to **maximize** functions while also building in **constraints**
- Expand the normal equations from least squares estimation
  - Homogenous equations: $[\mathbf{A} - \lambda\mathbf{I}]\underline{v} = 0$
    - $\mathbf{A}$ is the covariation, covariance, or correlation matrix
    - λ is the **eigenvalues** of $\mathbf{A}$
    - $\mathbf{I}$ is the identity matrix
    - $\underline{v}$ is the **eigenvectors** of $\mathbf{A}$
- Homogenous equations solution: eigenvectors / eigenvalues

5.1.3 What are eigenvectors and eigenvalues?

- Partition the variance in a matrix into **linearly independent portions**
  - **Eigenvectors** create a **basis** for the matrix
    * Each eigenvector is **axis** that is *orthogonal* with all others
    * We will also look at axes that are not mutually orthogonal later
  - **Eigenvalues** show how much **variance** is on each axis/eigenvector
    * Eigenvectors with higher corresponding eigenvalues contain **more of the variance** in the matrix

5.1.4 Eigenfaces

- Face Recognition using Eigenfaces and Distance Classifiers: A Tutorial
  - Every face image is made of *linearly independent* components
    * Gross oversimplification: shape, nose, eyes, mouth
  - Can describe every face as a linear combination of the components and some weights
    * Big eyes: high weight for eyes
    * Small mouth: low weight for mouth

5.1.5 Properties of eigenvalues

- From a $p \times p$ matrix (e.g., covariance or correlation)
  - If matrix is full rank: p eigenvalues
  - Otherwise, $< p$ eigenvalues
- Covariation, covariance, and correlation matrices will only have **positive** or **zero** eigenvalues ("positive definite")
  - Some will be zero if the matrix is **not full rank** (see above)
- Number of non-zero eigenvalues = rank of the matrix
- First eigenvalue is the largest, second is next largest, etc.

5.1.6 Properties of eigenvalues

- Eigenvalues change with the values in the matrix
  - e.g., eigenvalues from covariance matrix are different from eigenvalues from correlation matrix

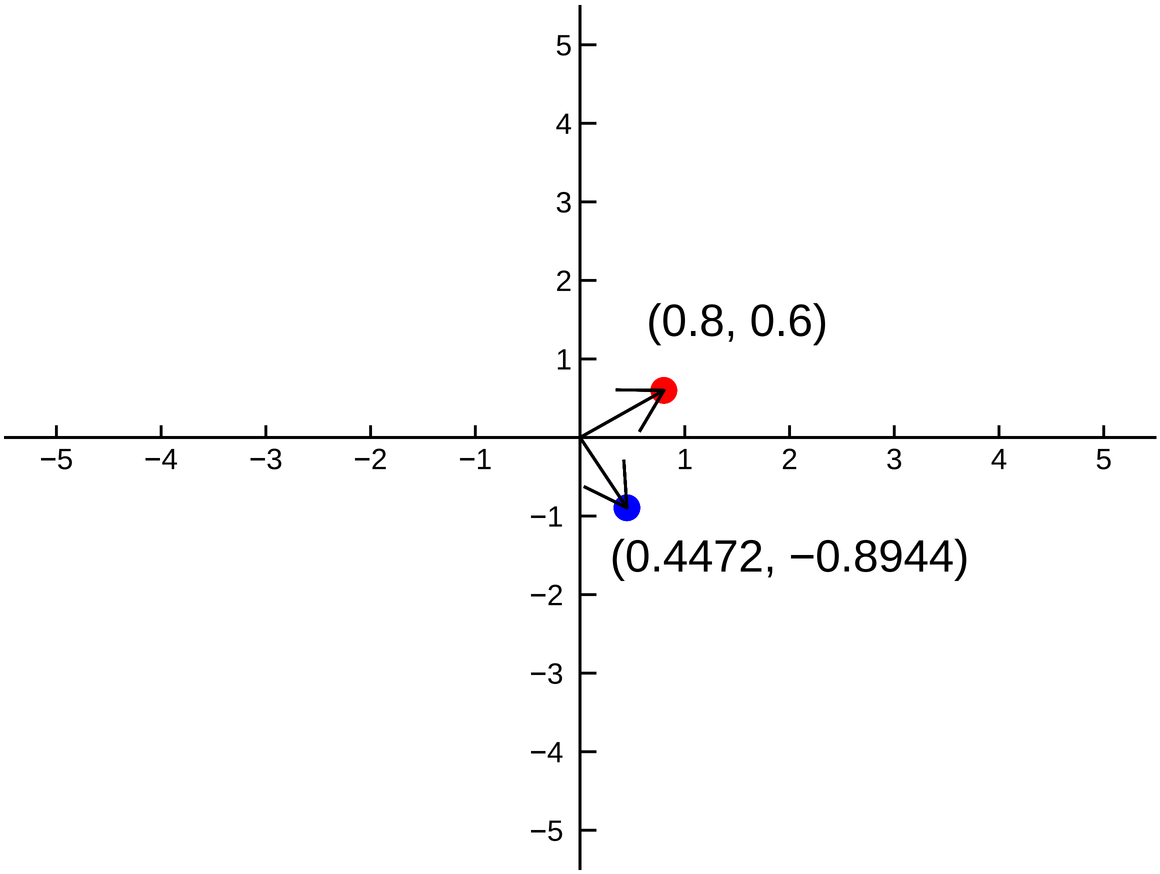
- **Sum** of the p eigenvalues = **sum of the diagonal elements**
  - For covariance matrix, sum of eigenvalues = sum of variances
  - For correlation matrix, sum of eigenvalues = number of variables
- **Product** of the p eigenvalues = **determinant of the matrix**
  - If any eigenvalue = 0, determinant is 0 too

5.1.7 Properties of eigenvectors

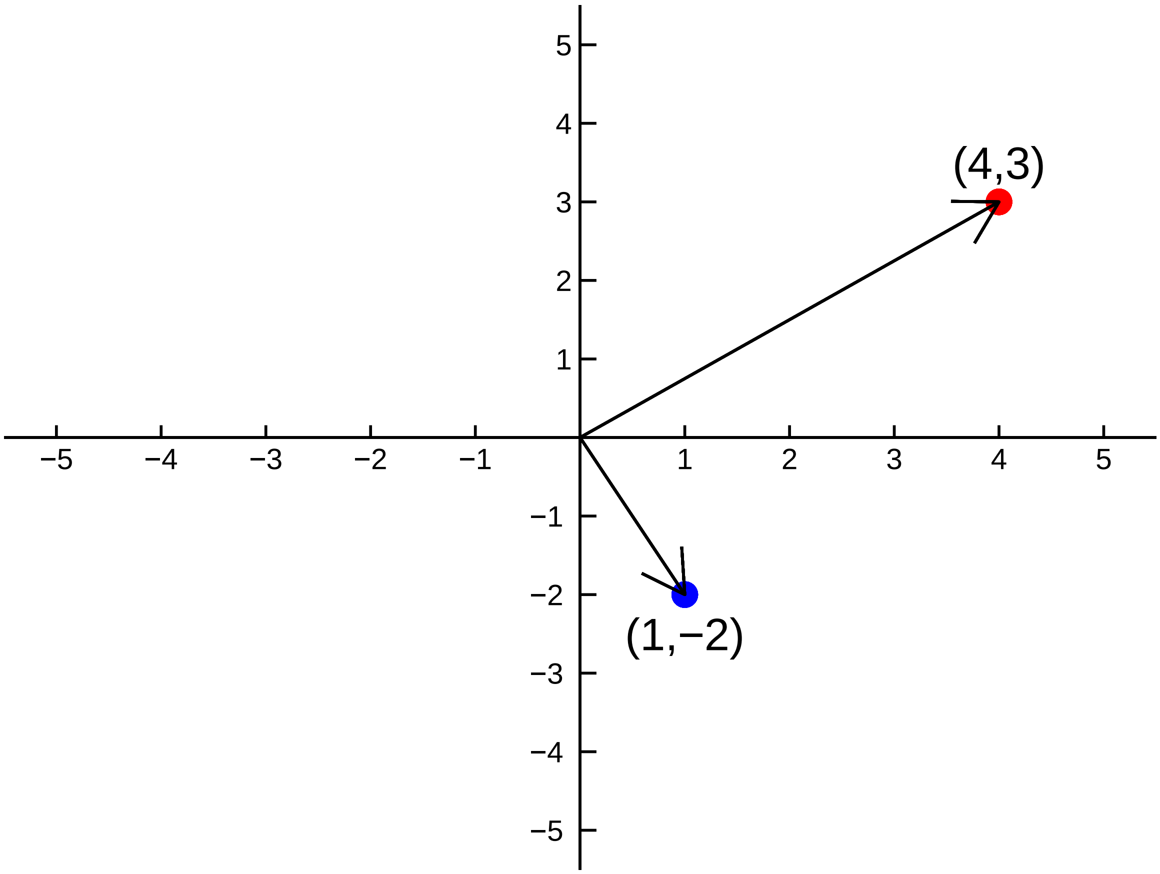
- One eigenvector for each eigenvalue
- Eigenvectors are mutually orthogonal
  - i.e., eigenvectors form an orthogonal basis for the matrix they were derived from
- Eigenvectors must be standardized ("normed")
  - Either to 1 (unity) or their root (the eigenvalue of that eigenvector)
  - SPSS and R give eigenvectors "normed to unity"

5.1.8 Normed to unity (1) vs normed to root

Normed to unity



Normed to root



5.1.9 Summary: Eigenvalues and eigenvectors

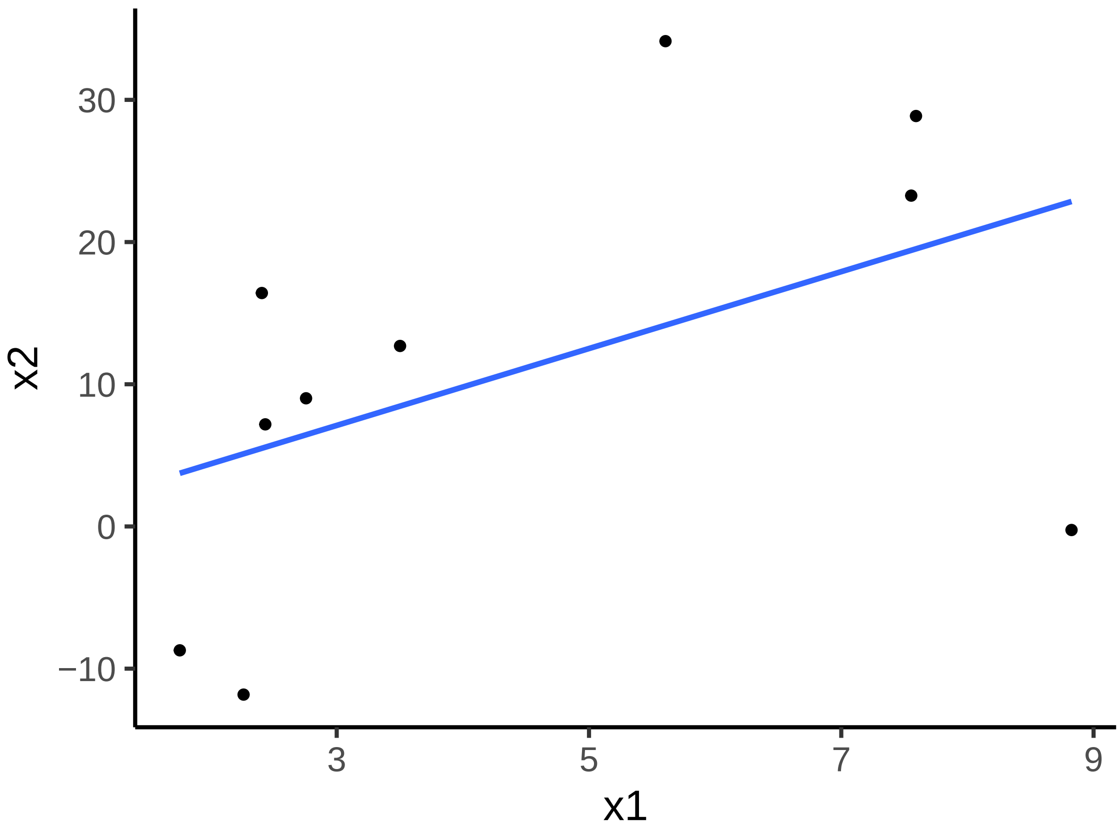
- Eigenvalues and eigenvectors divide up the variance in a matrix
  - Eigenvectors create a set of linearly independent axes
  - Eigenvalues tell us how much variance is on each axis
- If variables in the matrix are more correlated:
  - Determinant gets closer to 0
  - First eigenvalue is even larger relative to the others
  - Relatively more variance is explained by the first eigenvalue

6 Example and Conclusion

6.1 Example

6.1.1 Data matrix A

x1	x2
7.555	23.265
2.406	16.416
1.756	-8.710
2.262	-11.823
3.502	12.694
7.592	28.863
8.825	-0.250
2.757	9.012
2.434	7.180
5.606	34.126



6.1.2 Covariance and correlation matrices of A

Covariance matrix of **A**

	x1	x2
x1	7.116	19.240
x2	19.240	232.329

Correlation matrix of **A**

	x1	x2
x1	1.000	0.473
x2	0.473	1.000

6.1.3 Determinant and rank of A

- Determinant of cov(**A**)
  - $|cov(\mathbf{A})| = 1283.0759$
- Determinant of cor(**A**)
  - $|cor(\mathbf{A})| = 0.776$
- Determinants of **uncorrelated** variables with same variances: 1653.2532 and 1, respectively
- Rank of $\mathbf{A} = 2$
  - Number of pieces of independent info in the matrix
  - Less of # rows (10) and # columns (2)
  - Two variables that are only moderately correlated = rank 2

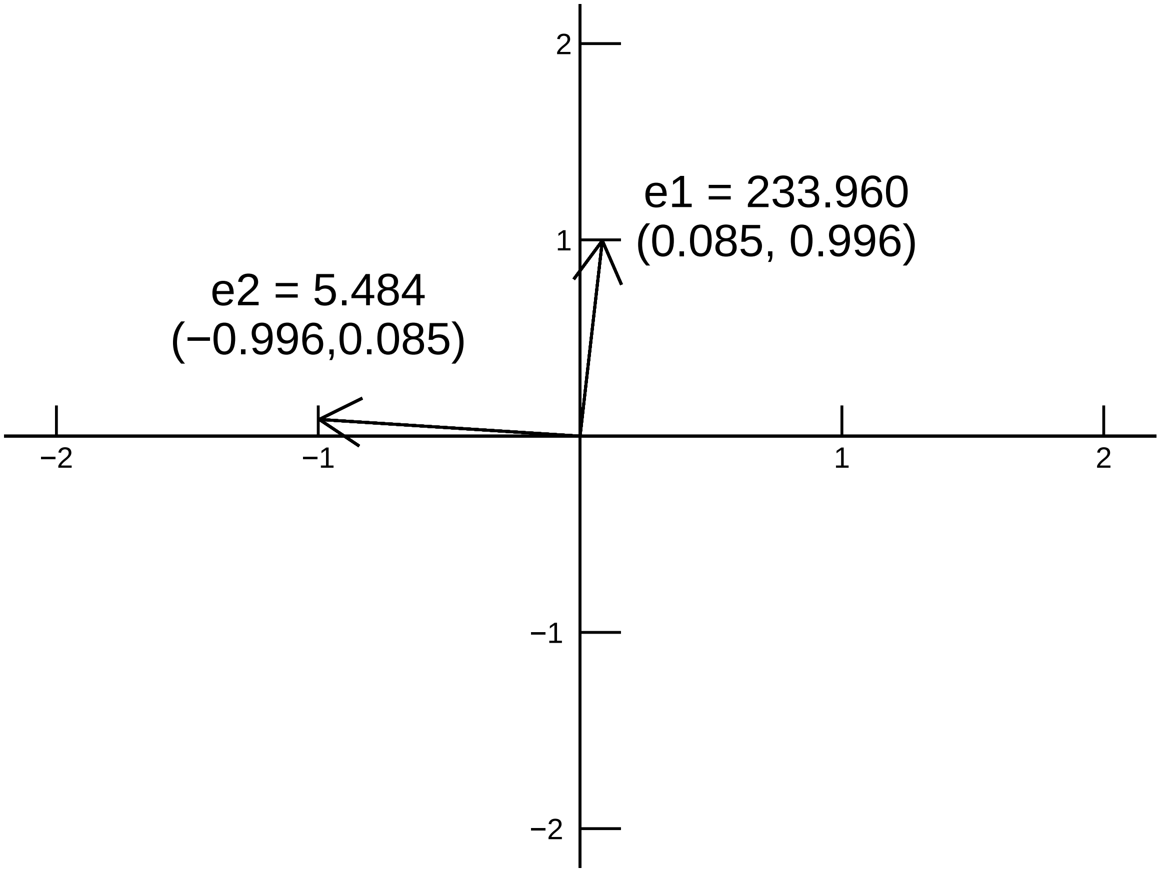
6.1.4 Eigenvector & eigenvalues: Covariance matrix

- Eigenvalues

1	233.960
2	5.484

- Eigenvectors

v1	v2
0.085	-0.996
0.996	0.085



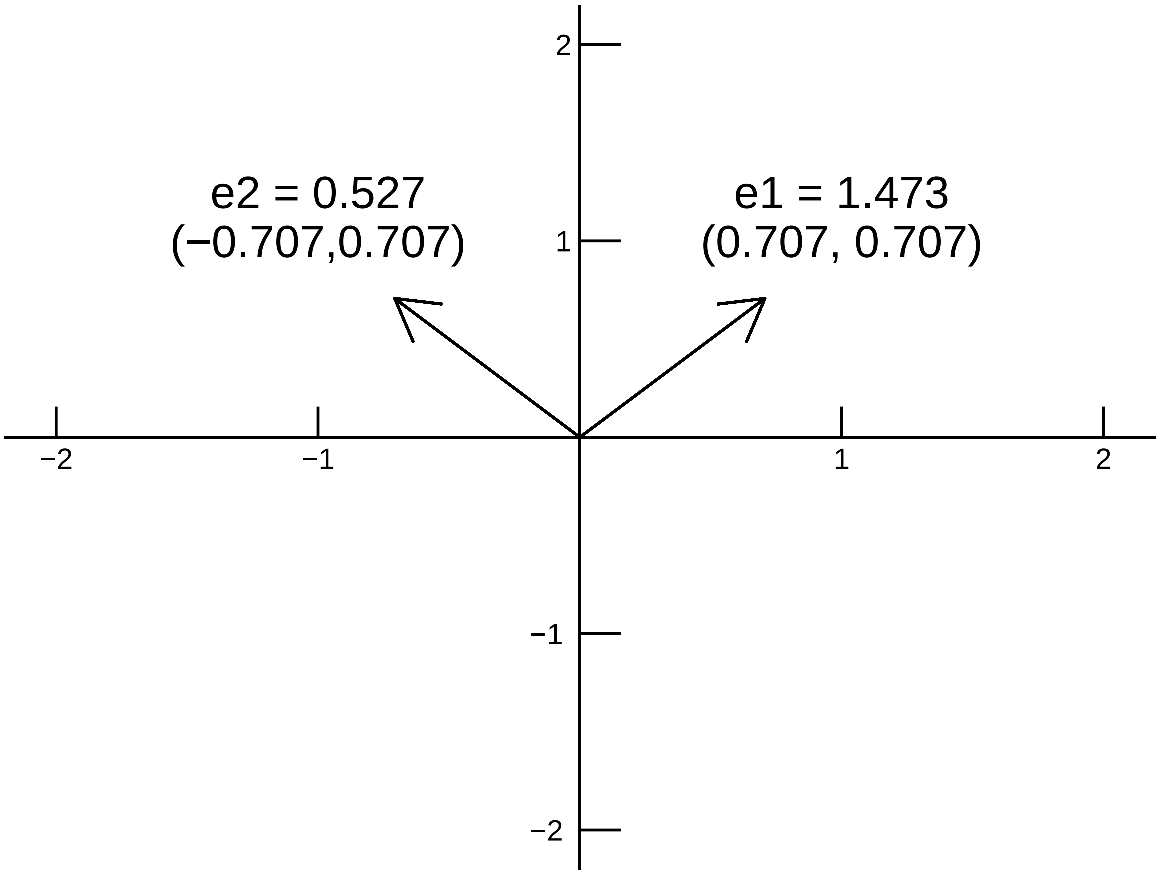
6.1.5 Eigenvector & eigenvalues: Correlation matrix

- Eigenvalues

1	1.473
2	0.527

- Eigenvectors

v1	v2
0.707	-0.707
0.707	0.707



6.1.6 Properties of eigenvalues

- Full rank matrix, so rank = number of variables = $p = 2$
- Covariance, correlation: only **positive** eigenvalues
- **Sum** of the p eigenvalues = **sum of the diagonal elements**
  - Covariance: $233.960 + 5.484 = 7.116 + 232.329 = 239.444$
  - Correlation: $1.473 + 0.527 = 1 + 1 = 2$
- **Product** of the p eigenvalues = **determinant of the matrix**
  - Covariance: $233.960 * 5.484 = 1283.04$ (w/in rounding)
  - Correlation: $1.473 * 0.527 = 0.776$ (w/in rounding)

6.2 Summary of this week

6.2.1 Summary of this week

- Matrices have a lot of parts – how can we understand them?
  - Vectors and matrices are objects with length and direction
  - Determinant and rank tell how vectors in a matrix are related
- How many pieces of independent information?
  - Eigenvectors tell us where independent info is
  - Eigenvalues tell us how much independent info there is

6.3 Next few weeks

6.3.1 Next few weeks

- Eigenvalues and eigenvectors are central to principal components analysis (PCA) and factor analysis (FA)
- PCA and FA seek to **reduce the dimension** of a set of variables by finding a **smaller set of axes** that can represent all the variables
  - For example, 10 variables → 2 axes, with each variable represented as a **composite of the 2 axes and specific weights**