Vector Geometry

Vectors

Vectors are mathematical idea (although they are rooted in everyday physical experience)

Vectors “live” in a space of some dimension and are often represented as a collection of numbers displayed in a column

$$\mathbf{Y} = \begin{bmatrix} 5 \\ 9 \\ 2 \\ 3 \end{bmatrix}$$

Vector in four-dimensional space

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Vector in n-dimensional space

The number of elements in the vector enumerate the vector's dimension.

Vectors names are denoted in **bold-face**

Dimension

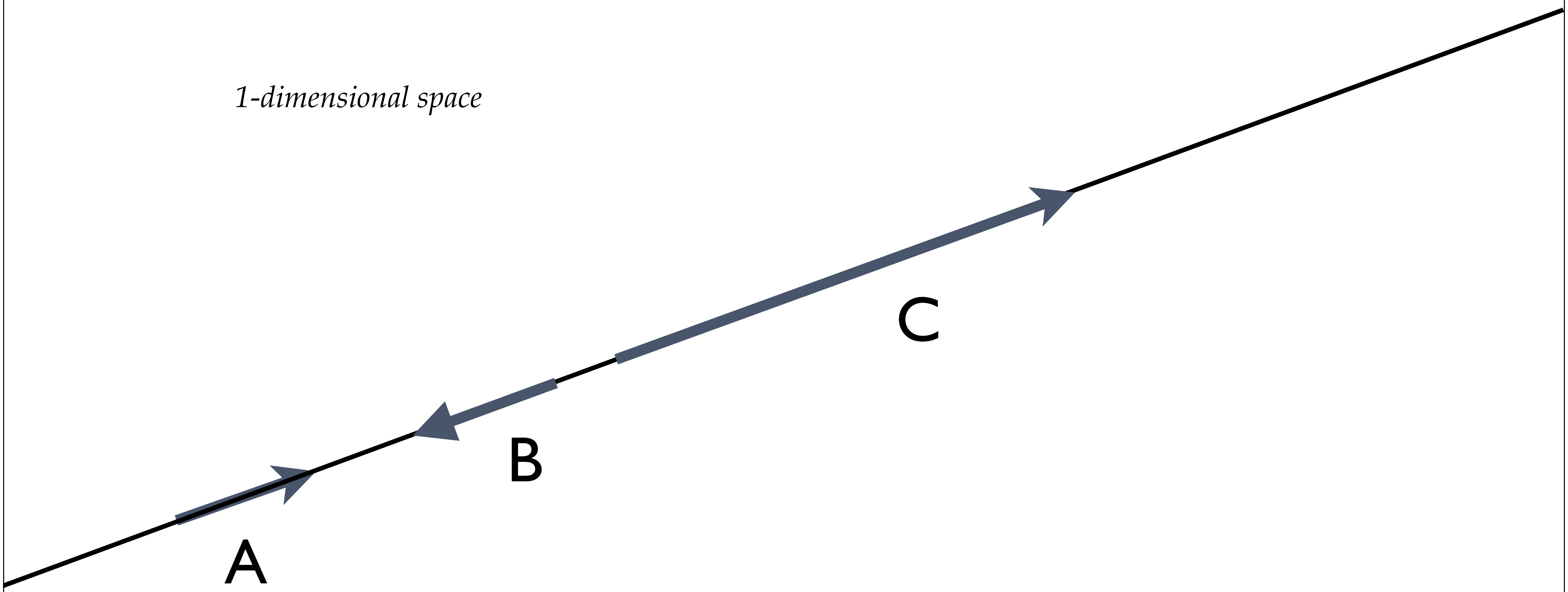
Dimension is a mathematical abstraction

Sometimes it makes intuitive sense and we can represent it

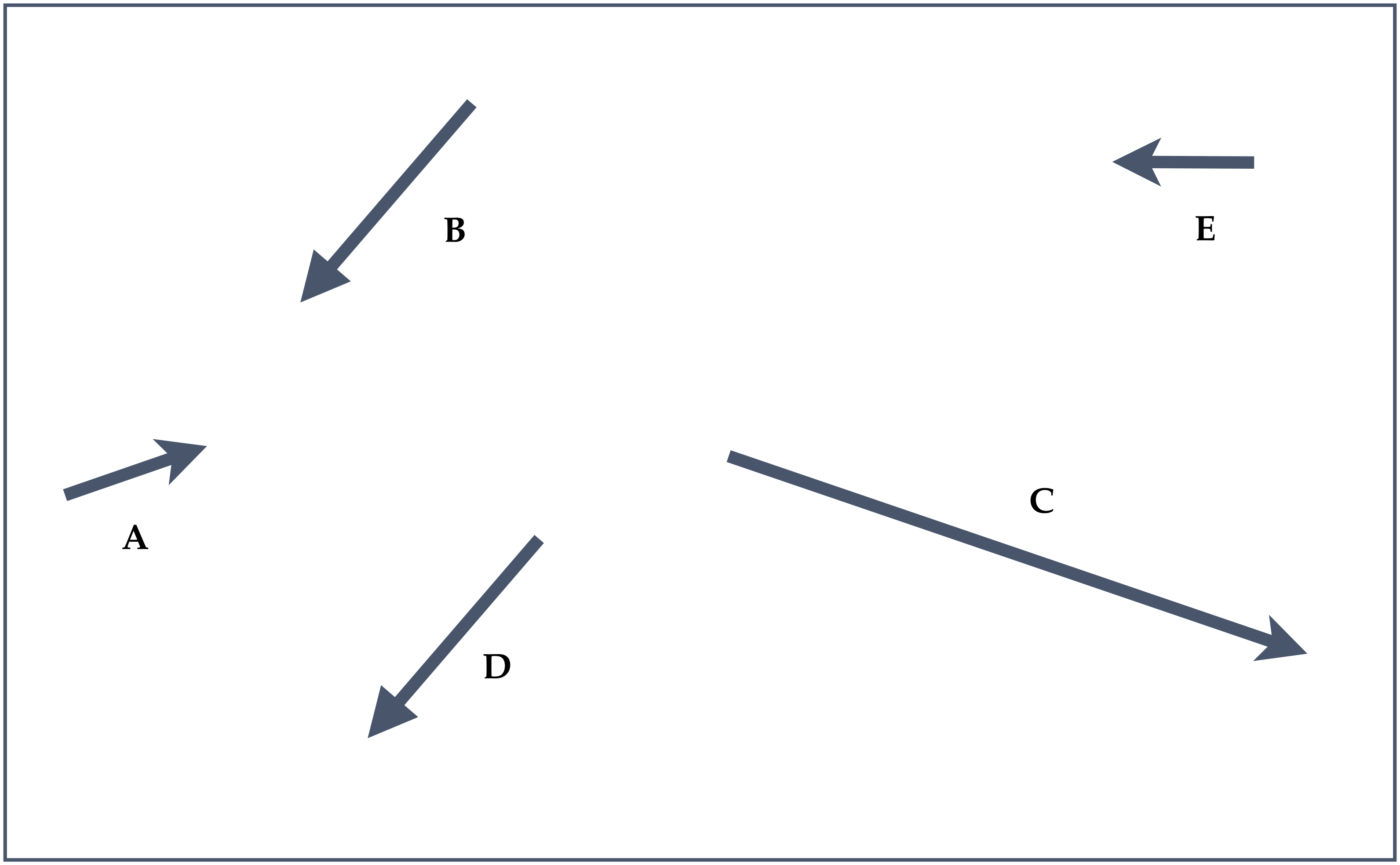
- Drawing an arrow in 2-dimensional space
- A pencil in 3-dimensional space

Sometimes it does not and we cannot represent it

- 4-dimensional or 100-dimensional space isn't as intuitive


1-dimensional space
A
B
C

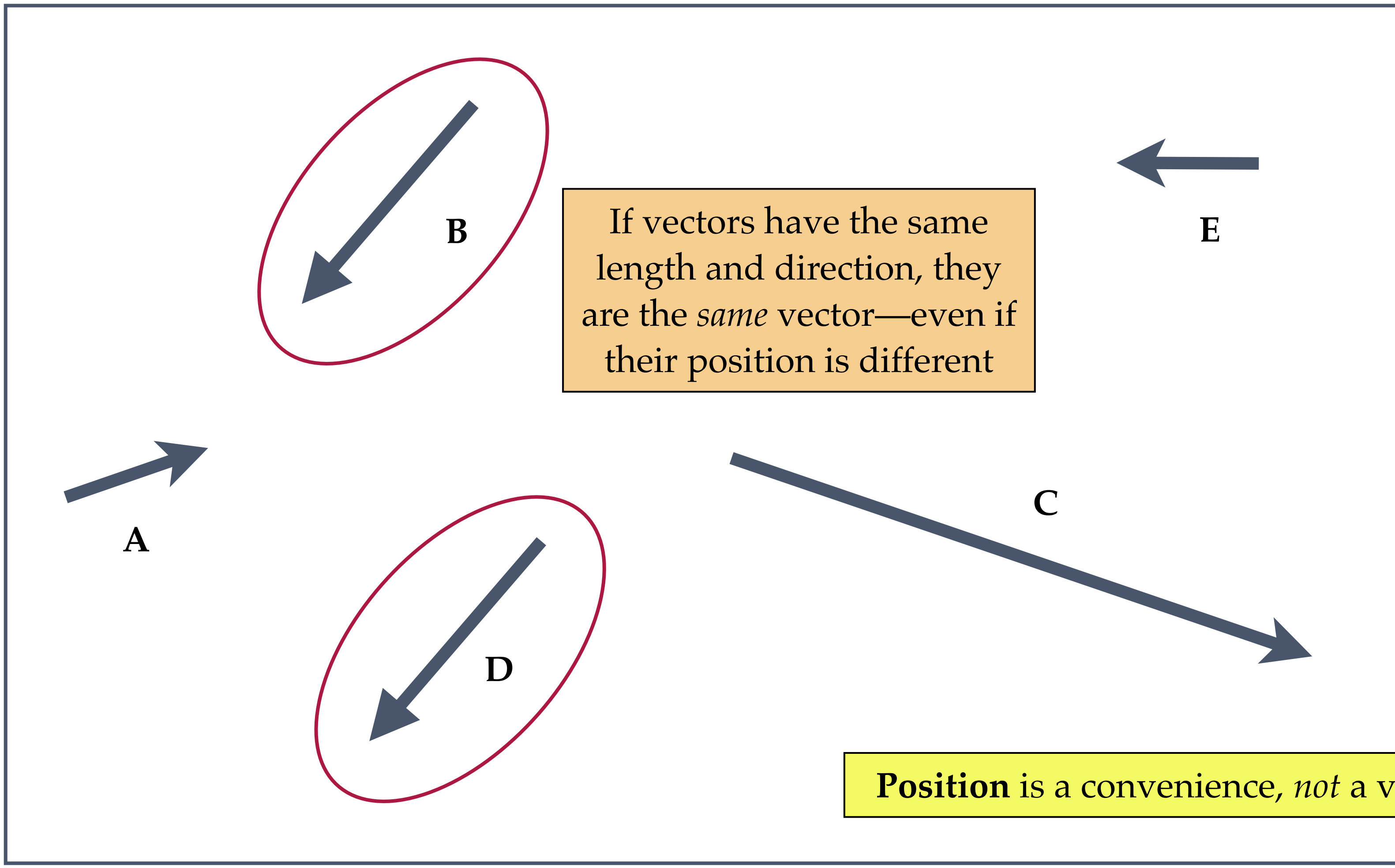
2-Dimensional space



Drawing vectors outside of two dimensions requires tricks of perspective.

Other Properties of Vectors

Aside from dimension, the only other two properties of a vector are its **length** and a **direction**

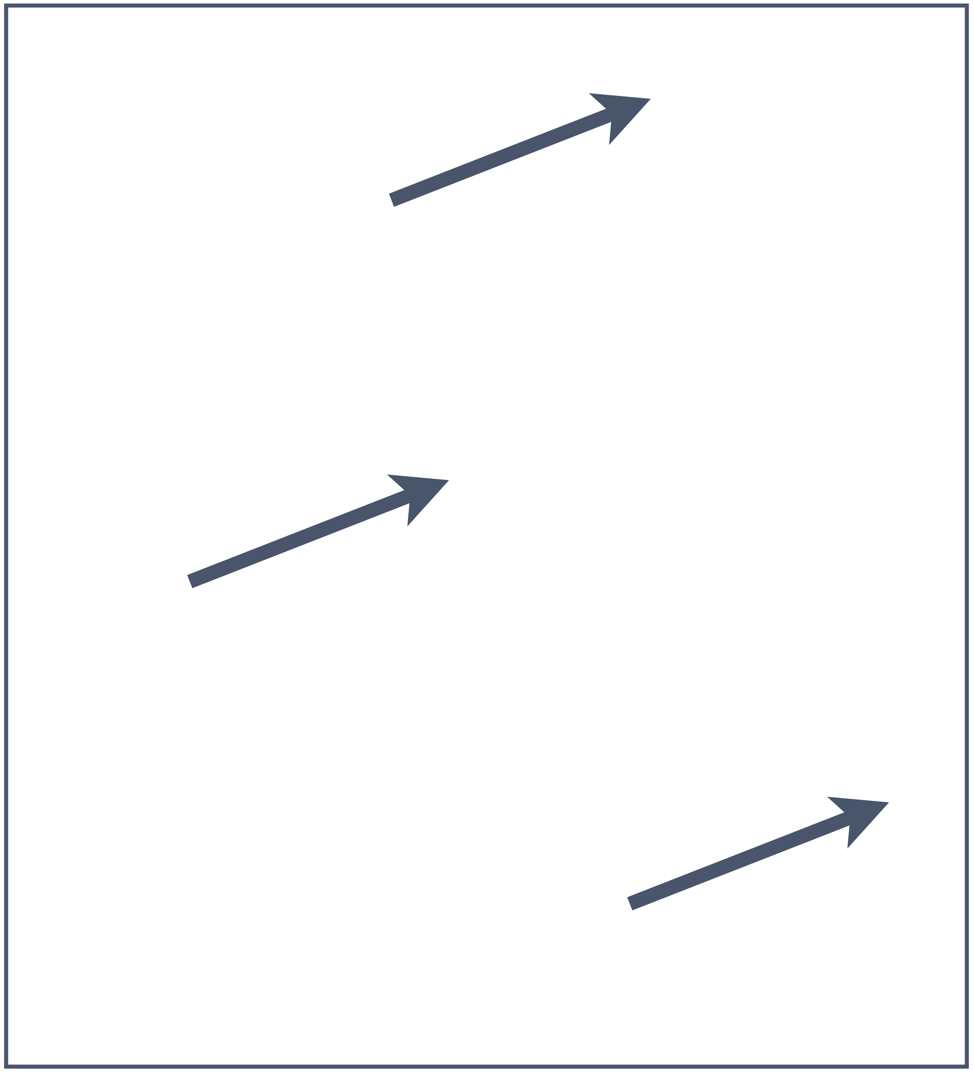


Position is a convenience, *not* a vector property.

To help with this concept, think about vectors is in terms of movement...

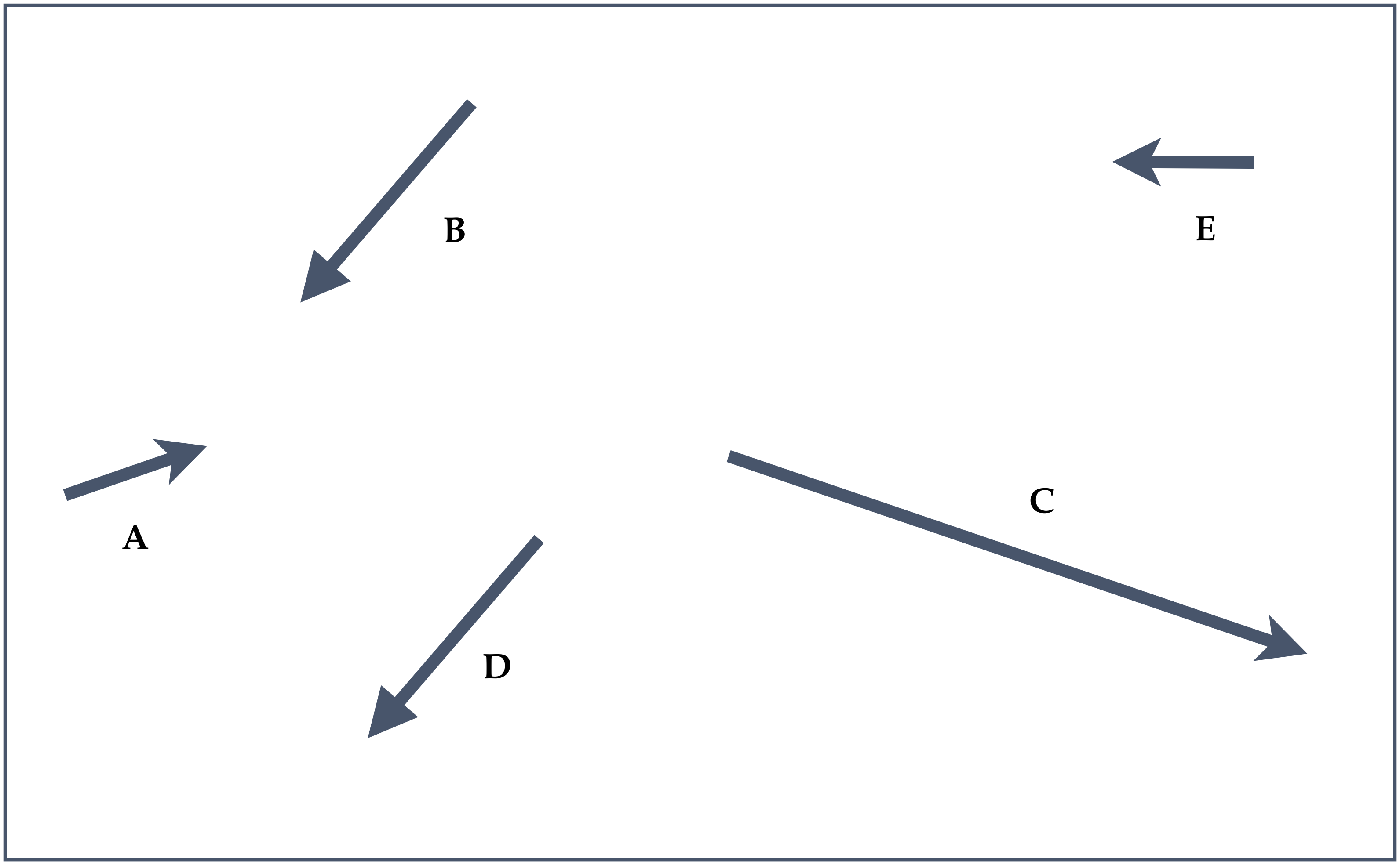
Vector is instruction to **take *x* steps** (length) in a **particular direction** (direction).
For example, move 1 meter to the NE.

These type of directions make sense regardless of your position….you have moved the same amount in the same direction regardless of where you started.

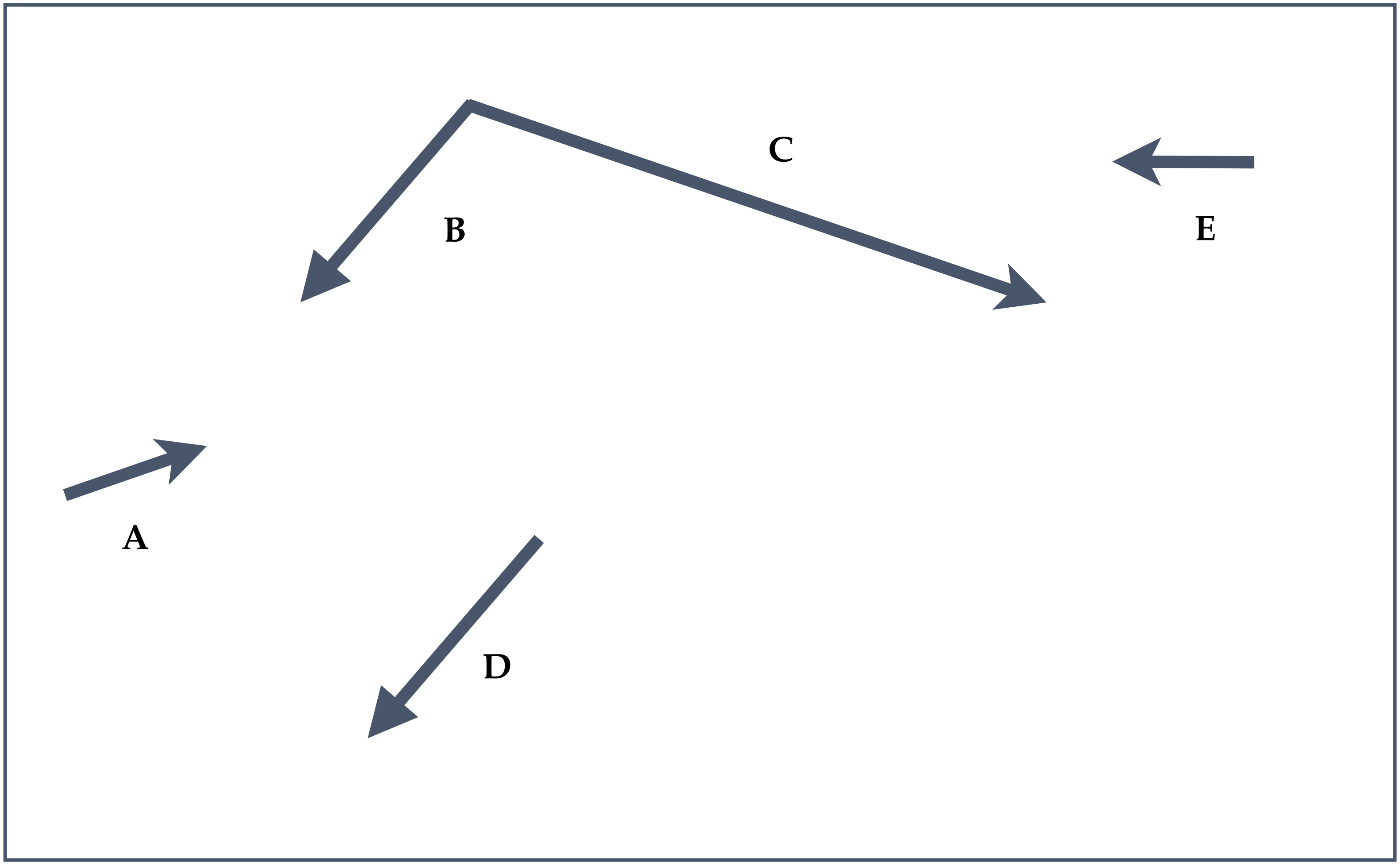


Vectors can be positioned wherever it is convenient to put them.

For example, suppose you are asked to find the angle between vector **B** and vector **C**...



...Vector **B** and vector **C** can be moved tail-to-tail and then the angle between them can be measured with a protractor.

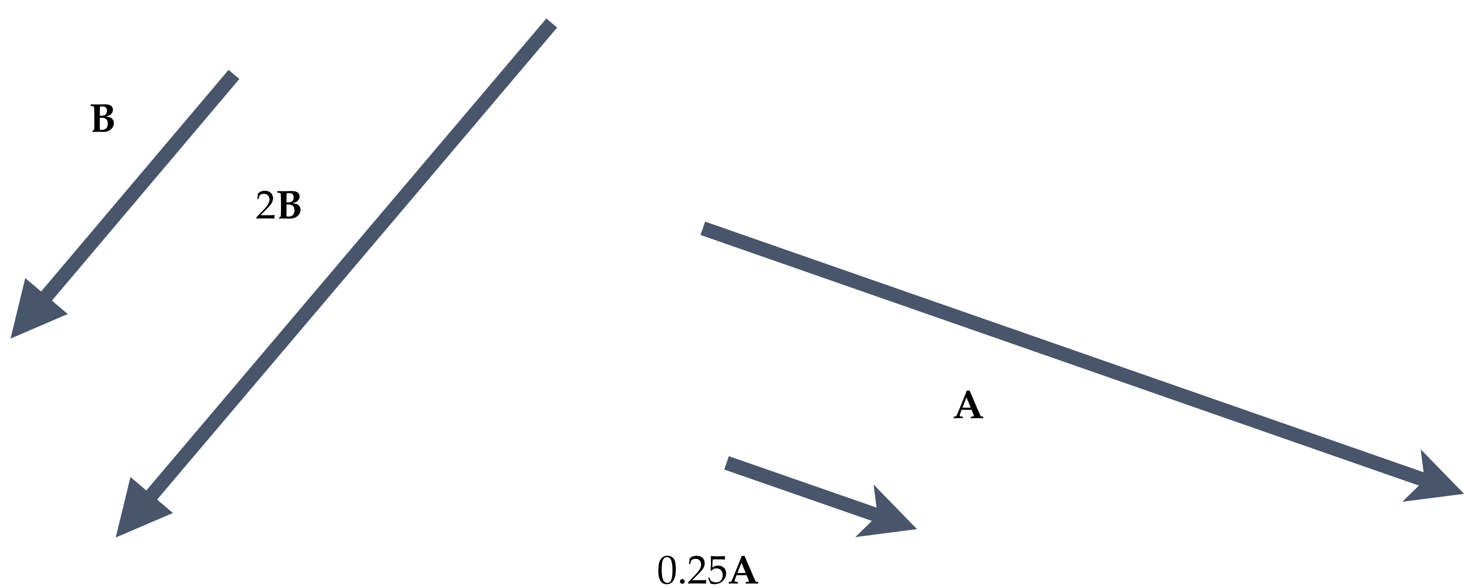


Scaling Vectors

Scaling a vector by a **positive factor** *changes the length* (makes it shorter or longer), but not the direction

For example,

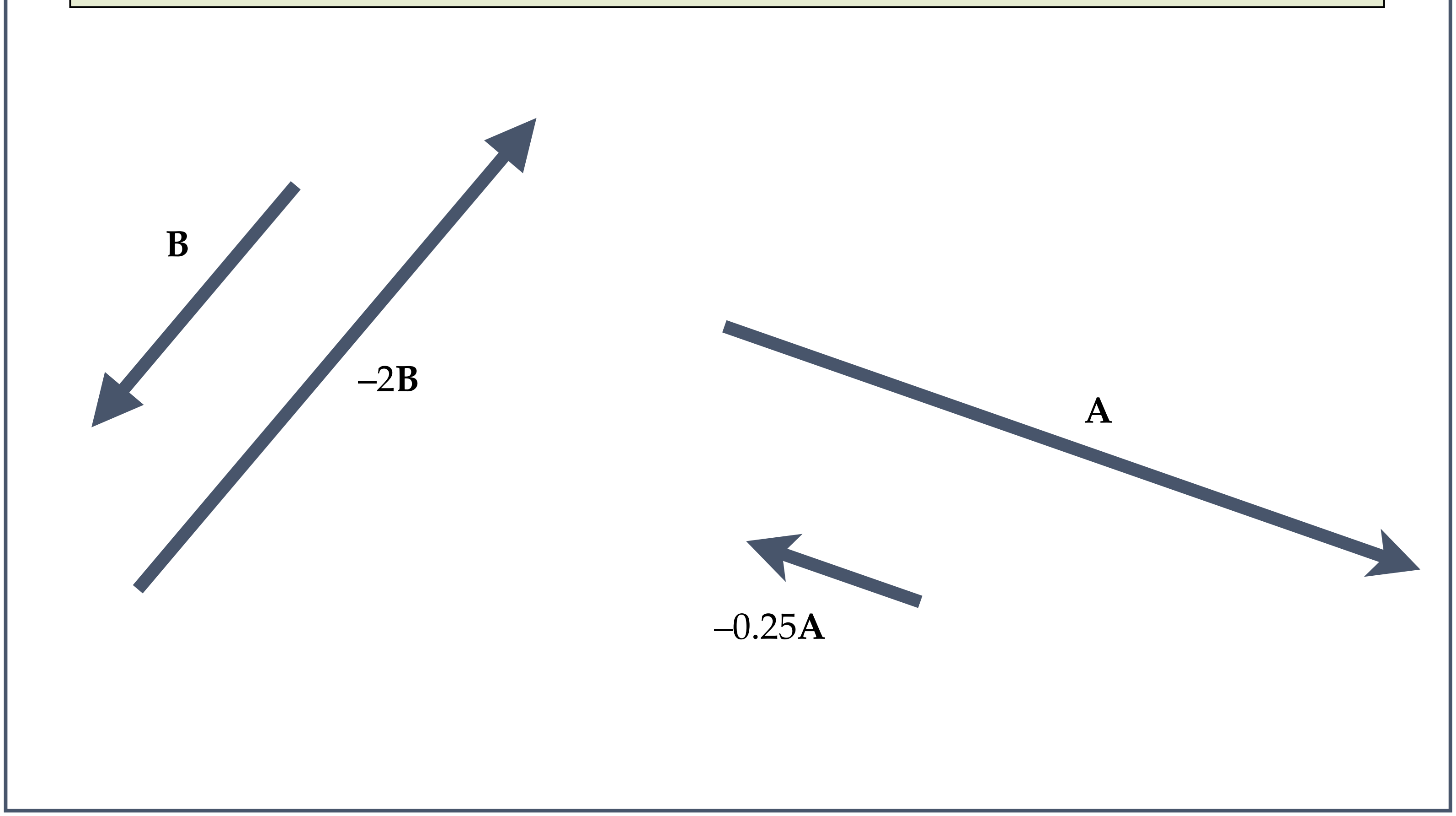
- "*Move 1 meter SW*" scaled by a factor of 2 means "*move 2 meters SW*"
- "*Move 1 meter SE*" scaled by a factor of 0.25 means "*move 1/4 meter SE*"



Scaling vectors by a **negative factor** changes the length (makes it shorter or longer), and the direction (step in the opposite direction)

For example,

- "*Move 1 meter SW*" scaled by a factor of –2 means "*move 2 meters NE*"
- "*Move 1 meter SE*" scaled by a factor of –0.25 means "*move 1/4 meter NW*"



Mathematics of Scaling Vectors

Scaling a vector boils down to multiplication

$$3 \begin{bmatrix} 5 \\ 9 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 15 \\ 18 \\ 6 \\ 9 \end{bmatrix}$$

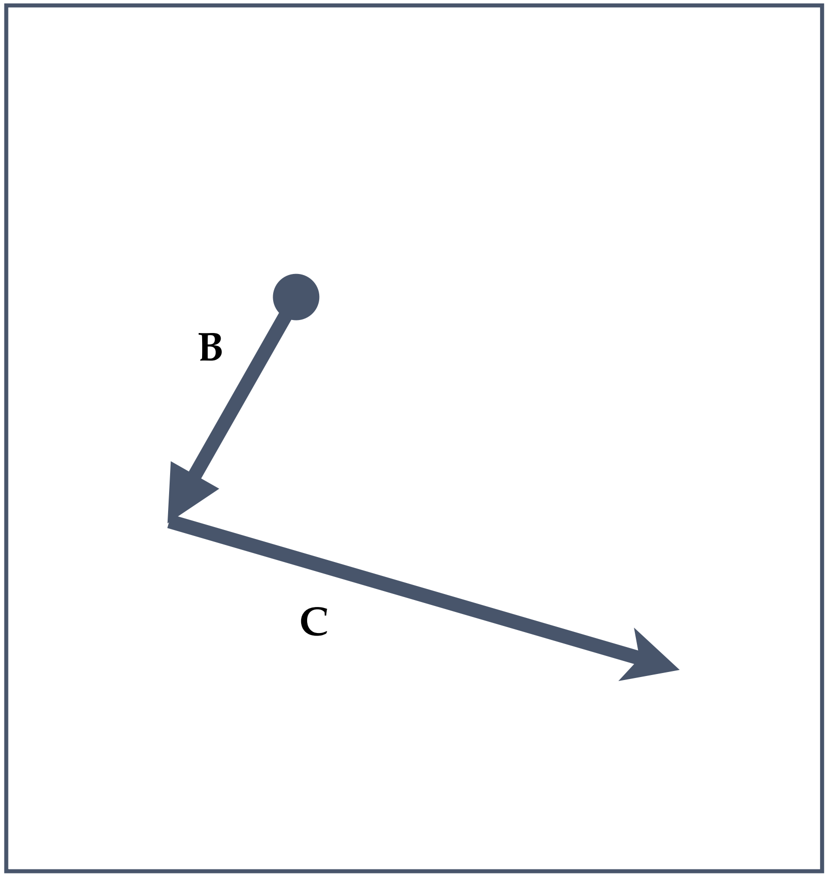
Scaling a vector by 3

The factor the vector is being scaled by, in this example, 3, is referred to as a **scalar**.

In general

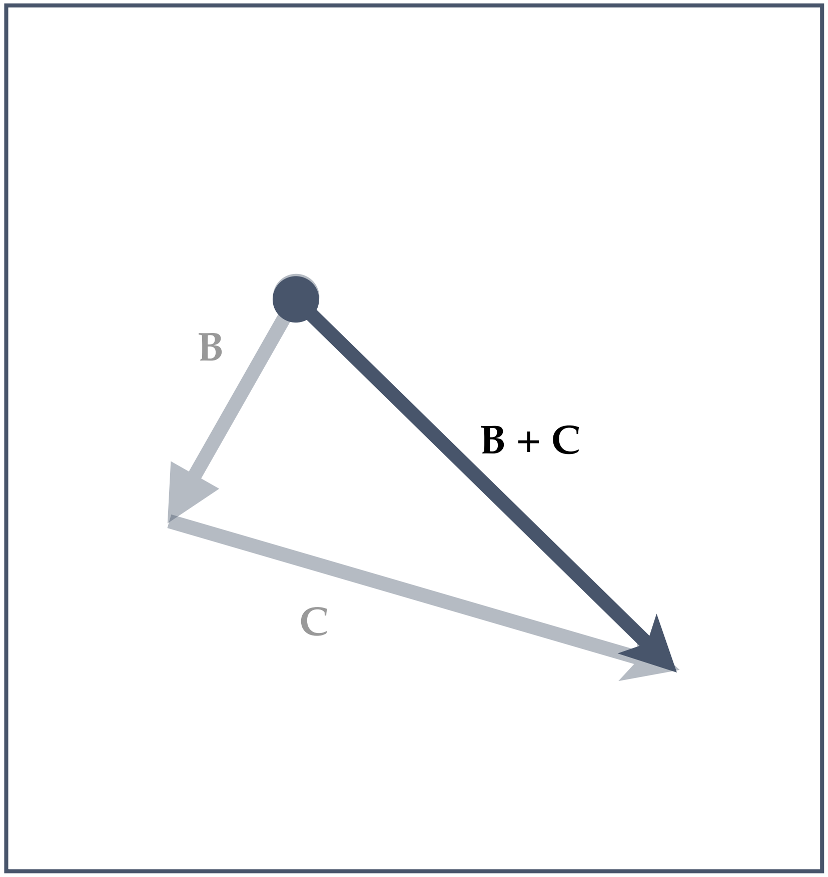
$$k\mathbf{X} = \begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{bmatrix}$$

Adding Vectors



- Pick a place to start
- Step according to **length and direction of first vector**
- From the point you ended, step according to length and direction of second vector

Adding Vectors



- The overall **distance and direction** ***from the place you started*** is the result of adding the two vectors

Mathematics of Adding Vectors

Mathematically, adding vectors boils down to adding their corresponding elements

$$\begin{bmatrix} 1 \\ 0 \\ -3 \\ 4 \end{bmatrix} + \begin{bmatrix} 5 \\ 9 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 9 \\ -1 \\ 7 \end{bmatrix}$$

In general

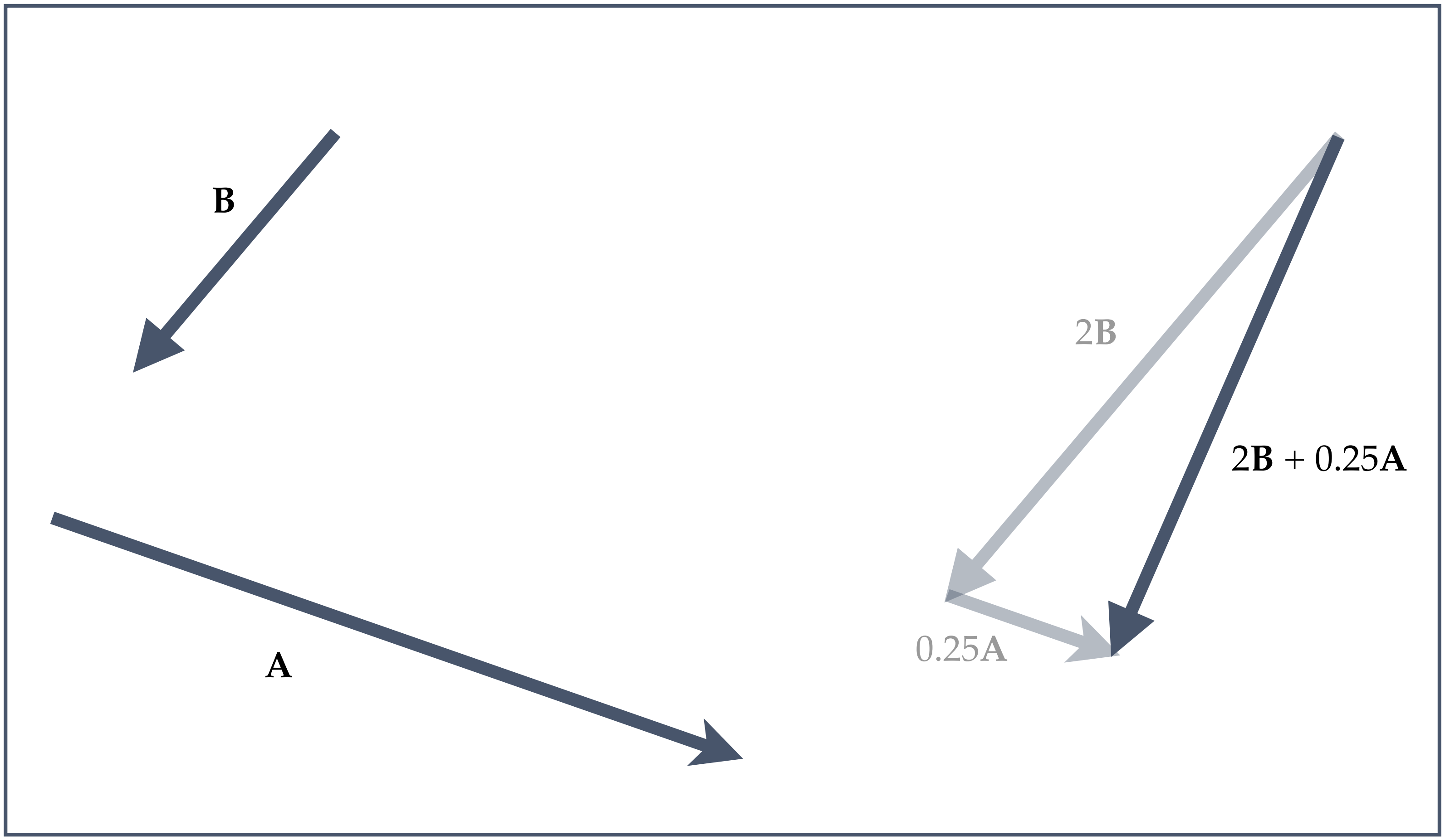
$$\mathbf{X} + \mathbf{Y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

Adding vector together requires that the vectors being added have the **exact same dimensions**

Linear Combinations

Scaling two vectors and adding the resulting vectors

2B + 0.25**A**



Mathematics of Linear Combinations

Scaling both vectors and then add the resulting vectors

$$3\begin{bmatrix}1\\0\\1\\1\end{bmatrix} - 6\begin{bmatrix}2\\4\\-1\\8\end{bmatrix} = \begin{bmatrix}3\\0\\3\\3\end{bmatrix} + \begin{bmatrix}-12\\-24\\6\\-48\end{bmatrix} = \begin{bmatrix}-9\\-24\\9\\-45\end{bmatrix}$$

If the original vectors have the exact same dimensions, then the scaled vectors will have the same dimensions as the originals and will be able to be added together.

In general

$$A\mathbf{X} + B\mathbf{Y} = \begin{bmatrix} A(x_1) + B(y_1) \\ A(x_2) + B(y_2) \\ \vdots \\ A(x_n) + B(y_n) \end{bmatrix}$$

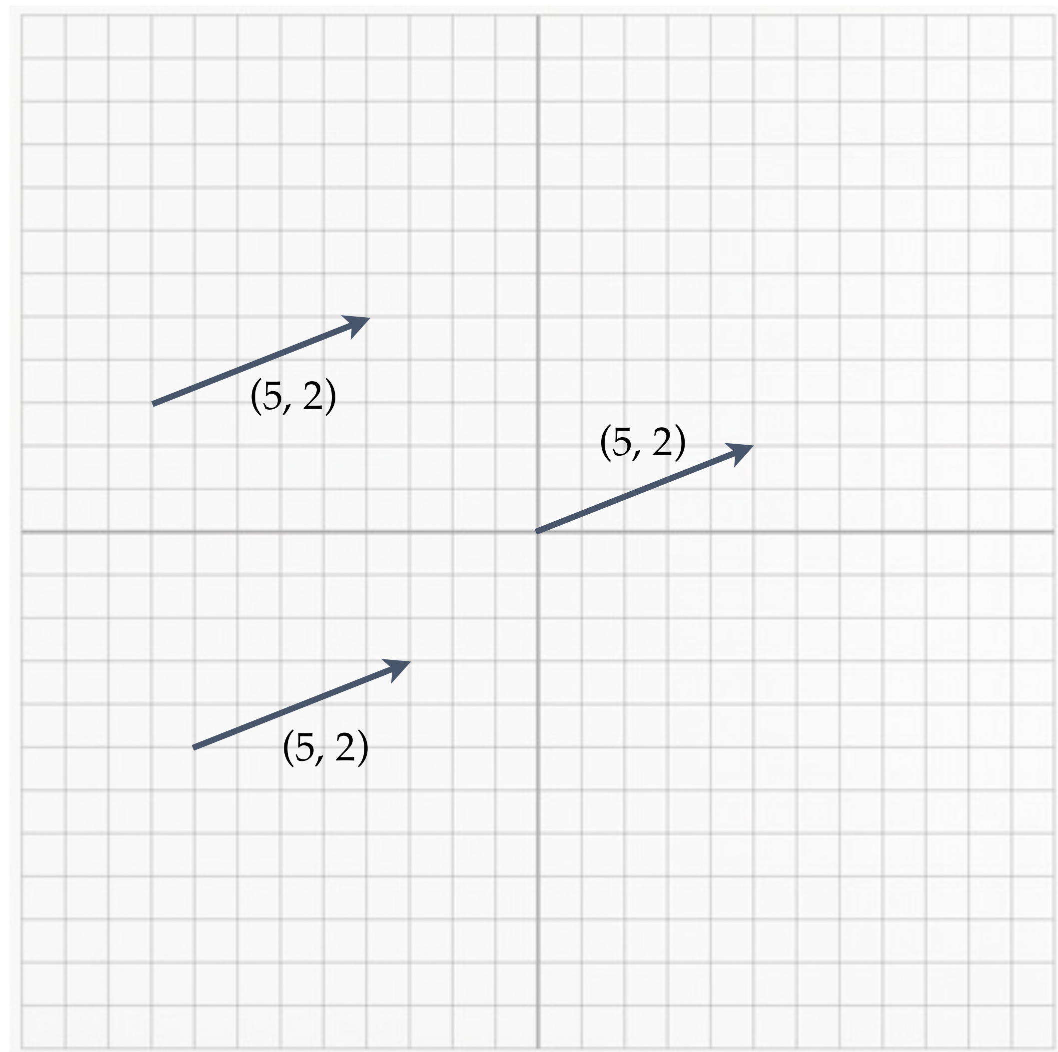
Linear combinations are at the heart of regression.

Plotting Vectors on Cartesian Coordinates

Although traditionally used to plot position, Cartesian coordinates can, just as easily, be used to plot **direction** and **length**

$$\mathbf{Y} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

In three dimensions, three coordinates are required to indicate length and direction



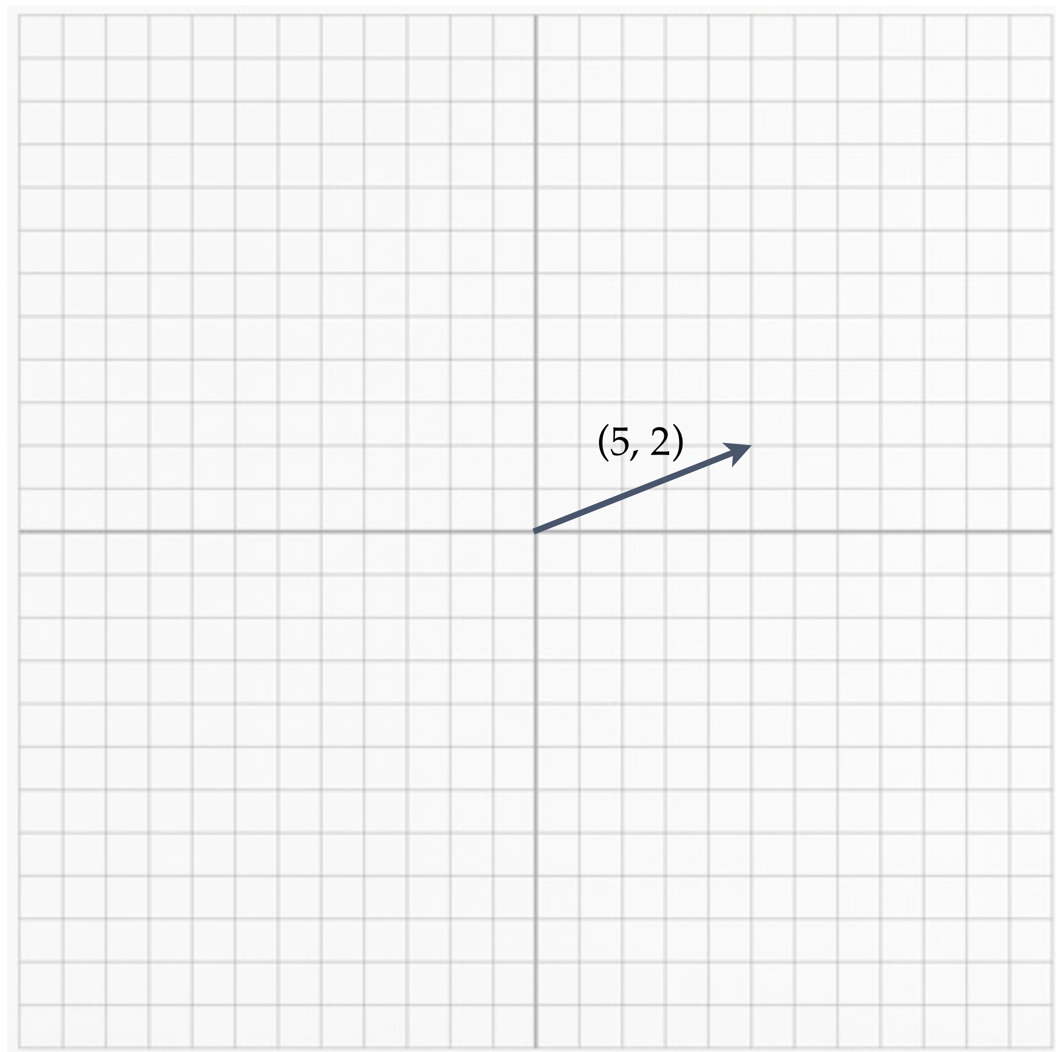
Length of a Vector

Using the Pythagorean Theorem, the length of the vector shown is

$$\sqrt{5^2 + 2^2} = \sqrt{29} = 5.39$$

In general, the length of a *n*-dimensional vector is

$$||\mathbf{X}|| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$



Dot Product

The dot product is a mathematical operation for vectors that multiplies the corresponding elements of two vectors together (element-wise multiplication) and then sums the results

$$\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} \bullet \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix} = 12 - 3 + 0 = 9$$

In general

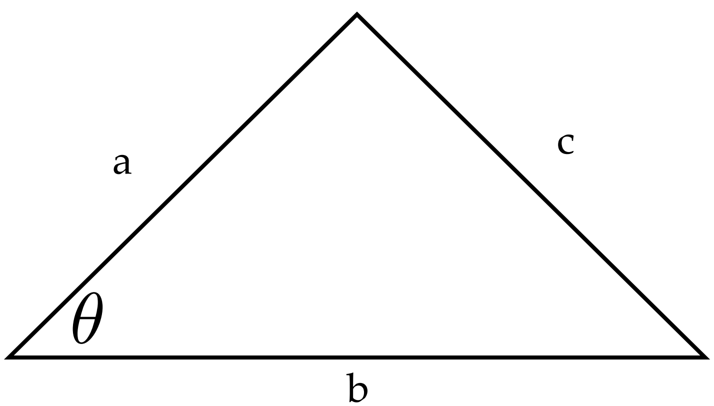
$$\mathbf{X} \bullet \mathbf{Y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \bullet \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1(y_1) + x_2(y_2) + \vdots + x_n(y_n)$$

The **length** of a vector can then be expressed as the *square root of a vector dotted with itself*

$$||\mathbf{A}|| = \sqrt{\mathbf{A} \bullet \mathbf{A}}$$

Law of Cosines

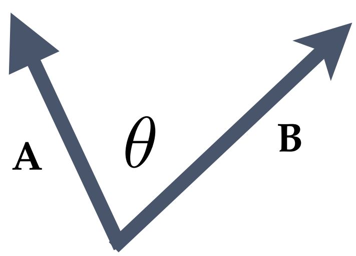
The *Law of Cosines* expresses the mathematical relationship between the side lengths of a triangle and one of its interior angles.



$$c^2 = a^2 + b^2 - 2ab\cos(\theta)$$

Re-expressing this equation, we can solve for theta

The Law of Cosines can also be used to express the mathematical relationship between any two vectors and the angle between them.



It turns out that, mathematically,

$$\mathbf{A} \bullet \mathbf{B} = ||\mathbf{A}||\ ||\mathbf{B}|| \cos\theta$$

Re-expressing this, $$\cos(\theta) = \frac{\mathbf{A} \bullet \mathbf{B}}{||\mathbf{A}||\ ||\mathbf{B}||}$$

$$\theta = \cos^{-1}\left(\frac{\mathbf{A} \bullet \mathbf{B}}{||\mathbf{A}||\ ||\mathbf{B}||}\right)$$

See the proof at http://en.wikipedia.org/wiki/Law_of_cosines

Vectors and Vector Mathematics with R

Vectors can be created in R using the `c()` function.
Each element is separated by a comma.

$$\mathbf{A} = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix}$$

```
# create vectors A and B
> A = c(3, 1, 0)
> B = c(4, -3, 2)

# scaling a vector
> 4 * A
[1] 12  4  0

# vector addition/subtraction
> A + B
[1]  7 -2  2

# linear combination
> 3 * A + 2 * B
[1] 17 -3  4
```

```r
# finding the dimension of a vector
> length(A)

[1] 3
```

```r
# element-wise multiplication/division
> A * B

[1] 12 -3  0

# dot product between A and B
> sum(A * B)

[1] 9

# length of vector A
> sqrt(sum(A * A))

[1] 3.162278
```

$$\cos\theta = \frac{\mathbf{A} \bullet \mathbf{B}}{||\mathbf{A}||\,||\mathbf{B}||}$$

```
# dot product between A and B
> sum(A * B)
[1] 9

# compute cos(theta)
> sum(A * B) / (sqrt(sum(A * A)) * sqrt(sum(B * B)))
[1] 0.5284982
```

$$\cos\theta = 0.53$$

To determine theta we use the arccosine. We will also convert from radians to degrees.

```
# find angle and convert to degrees
> acos(0.5284982) * 180 / pi

[1] 58.09596
```

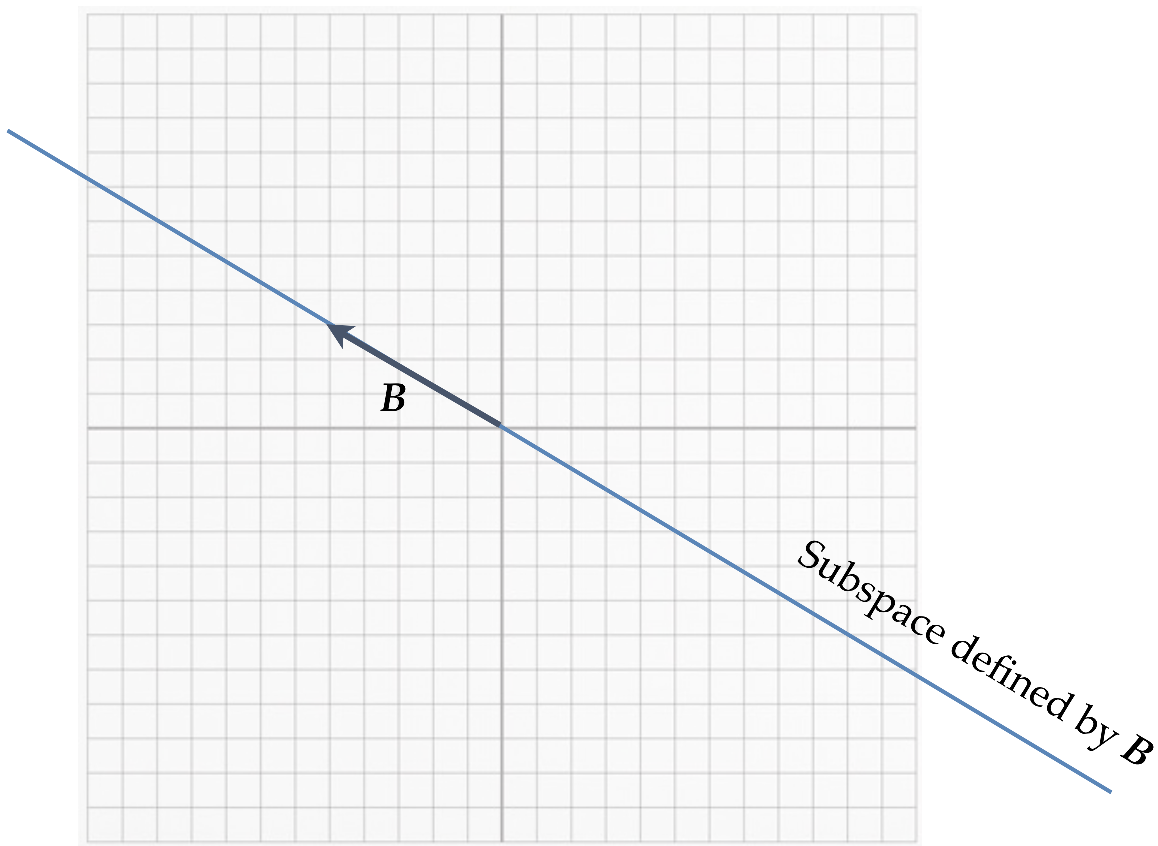
$$\theta = 58.1$$

The angles between **A** and **B** is 58.1°

Other Concepts Related to Vectors that are Important for Statistical Modeling

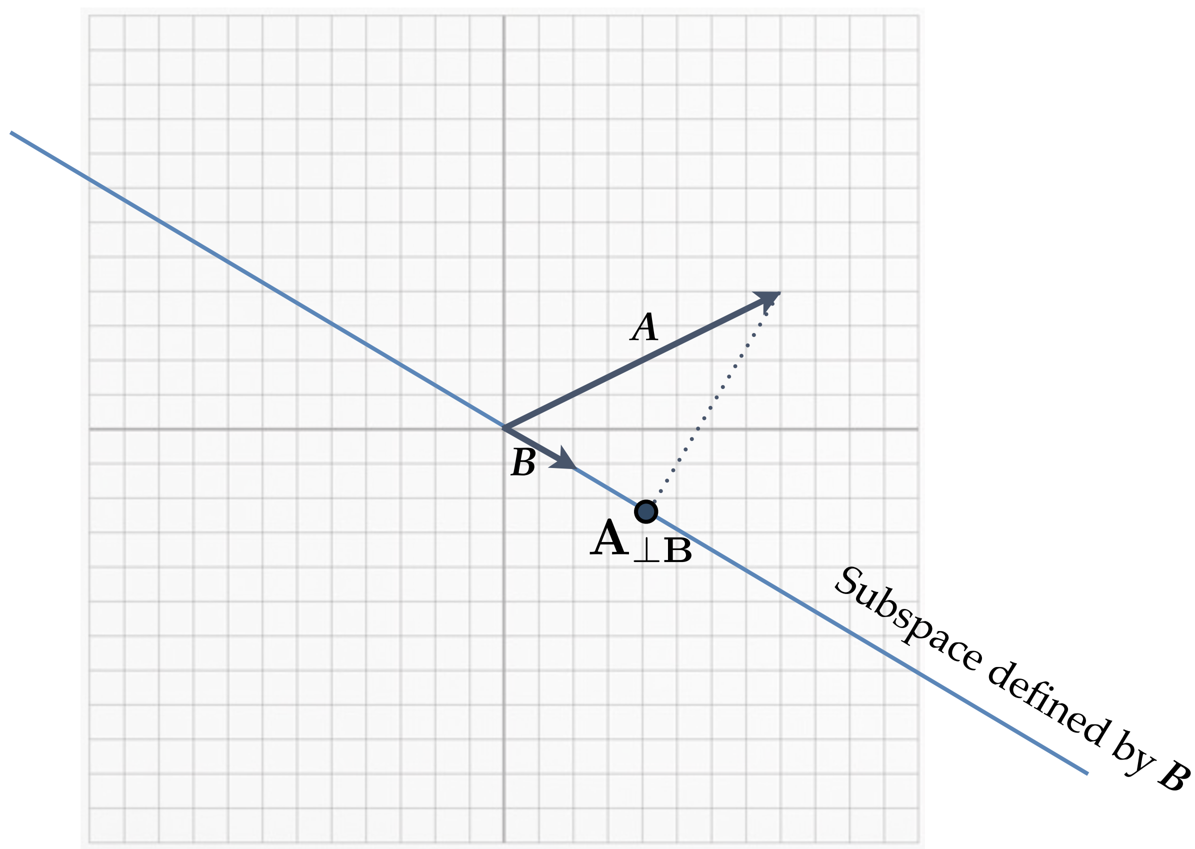
Subspace

A subspace is a part of the entire dimensional space defined by all of the points that can be reached by scaling a particular vector.



Projection

Projecting vector **A** onto a subspace defined by vector **B** finds the point in the subspace of **B** as close as possible to **A**



Unit Vector

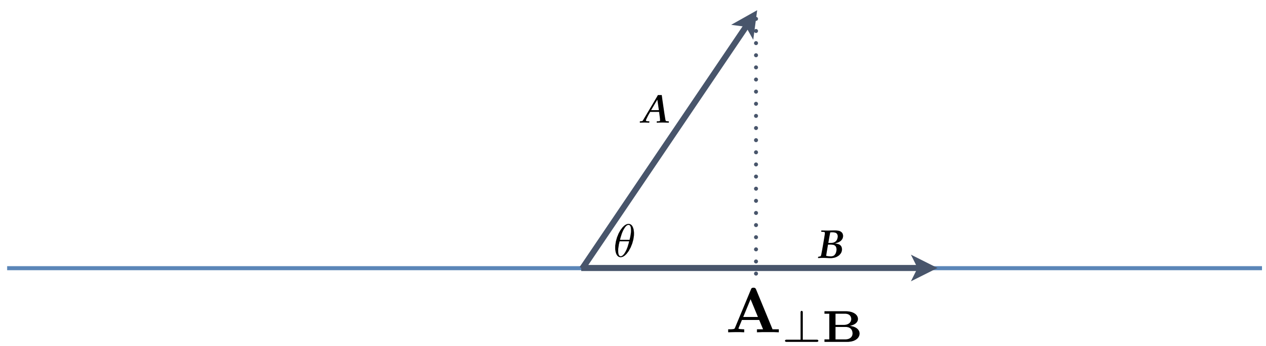
A unit vector is a vector that has length = 1.

Any vector can be scaled to have length 1.

$$\mathbf{1} = \frac{\mathbf{B}}{||\mathbf{B}||}$$

The resulting unit vector will be in the same direction as the original vector.

Length of a Projection



Definition of cosine,

$$\cos(\theta) = \frac{\text{adjacent side}}{\text{hypotenuse}}$$

from earlier,

$$\mathbf{A} \bullet \mathbf{B} = ||\mathbf{A}||\,||\mathbf{B}|| \cos(\theta)$$

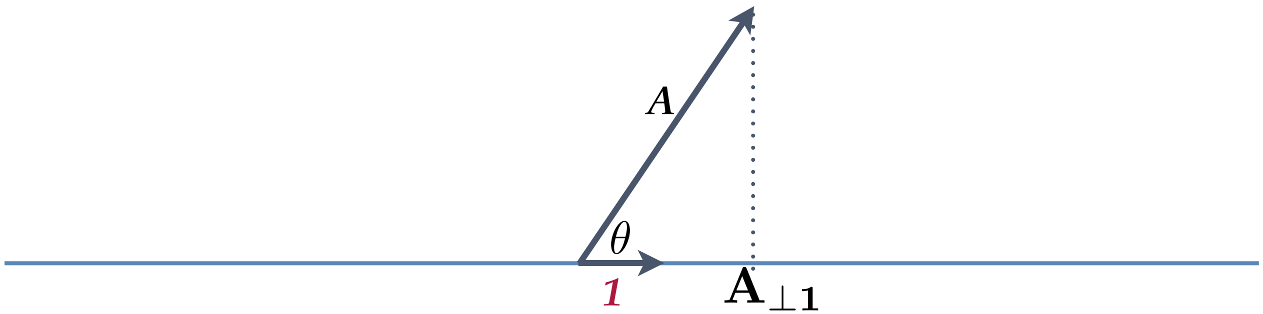
$$\text{adjacent side} = \cos(\theta) \times \text{hypotenuse}$$

$$||\mathbf{A}_{\perp \mathbf{B}}|| = ||\mathbf{A}|| \cos(\theta)$$

$$\frac{\mathbf{A} \bullet \mathbf{B}}{||\mathbf{B}||} = ||\mathbf{A}|| \cos(\theta)$$

$$||\mathbf{A}_{\perp \mathbf{B}}|| = \frac{\mathbf{A} \bullet \mathbf{B}}{||\mathbf{B}||}$$

Length of a Projection: Unit Vector



$$||\mathbf{A}_{\perp \mathbf{1}}|| = \frac{\mathbf{A} \bullet \mathbf{1}}{||\mathbf{1}||}$$

$$||\mathbf{A}_{\perp \mathbf{1}}|| = \mathbf{A} \bullet \mathbf{1}$$

The length of the projection of **A** onto the unit vector is equal to the dot product between **A** and the unit vector

Statistical Models and Vectors

Data

Consider using this "toy" dataset to model the variation in wages

```
> toy = read.csv(file = "toyData.csv")
> toy

  wage educ sex  status age   sector
 12.00   12   M Married  32    manuf
  8.00   12   F Married  33  service
 16.26   12   M  Single  32  service
 13.65   16   M Married  33     prof
  8.50   17   M  Single  26 clerical
```

Both the outcome (wage) and some of the predictors are quantitative (educ, age), and others are categorical (sex, status, sector)

In building the model you could include main-effects and/or interactions between predictors

Model Vectors: Quantitative Variables

Model vectors are the translation of the model terms into vectors. They are also referred to as **indicator variables**.

Each **quantitative variables** has a *single* model vector or indicator. For the three quantitative variables in our example, the model vectors are:

wage

$$\begin{bmatrix} 12.00 \\ 8.00 \\ 16.26 \\ 13.65 \\ 8.50 \end{bmatrix}$$

educ

$$\begin{bmatrix} 12 \\ 12 \\ 12 \\ 16 \\ 17 \end{bmatrix}$$

age

$$\begin{bmatrix} 32 \\ 33 \\ 32 \\ 33 \\ 26 \end{bmatrix}$$

Model Vectors: Categorical Variables

Categorical variables (factors) have *multiple* model vectors (indicators). There is **one model vector per level** of the factor.

The sex main-effect would be composed of two model vectors.

$$\begin{array}{cc} \texttt{sexF} & \texttt{sexM} \\ \begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix} & \begin{bmatrix}1\\0\\1\\1\\1\end{bmatrix} \end{array}$$

The status main-effect would also be composed of two model vectors.

$$\begin{array}{cc} \texttt{statusMarried} & \texttt{statusSingle} \\ \begin{bmatrix}1\\1\\0\\1\\0\end{bmatrix} & \begin{bmatrix}0\\0\\1\\0\\1\end{bmatrix} \end{array}$$

The sector main-effect would be composed of five model vectors.

$$\begin{array}{ccccc} \texttt{sectorclerical} & \texttt{sectorconstr} & \texttt{sectormanuf} & \texttt{sectorprof} & \texttt{sectorservice} \\ \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix} & \begin{bmatrix}0\\1\\1\\0\\0\end{bmatrix} \end{array}$$

Model Vectors: Interaction Terms

The model vectors for any interaction terms included are the pairwise products between all of the model vectors for the predictors included in the interaction.

$$
\begin{array}{ccccc}
\texttt{educ} & & \texttt{age} & & \texttt{educ:age} \\
\begin{bmatrix} 12 \\ 12 \\ 12 \\ 16 \\ 17 \end{bmatrix} & \times & \begin{bmatrix} 32 \\ 33 \\ 32 \\ 33 \\ 26 \end{bmatrix} & = & \boxed{\begin{bmatrix} 384 \\ 396 \\ 384 \\ 528 \\ 422 \end{bmatrix}}
\end{array}
$$

There is only one model vectors for the interaction between two quantitative predictors.

$$\begin{array}{ccccc} \texttt{age} & & \texttt{statusMarried} & & \texttt{age:statusMarried} \\ \begin{bmatrix}32\\33\\32\\33\\26\end{bmatrix} & \times & \begin{bmatrix}1\\1\\0\\1\\0\end{bmatrix} & = & \boxed{\begin{bmatrix}32\\33\\0\\33\\0\end{bmatrix}} \end{array}$$

and

$$\begin{array}{ccccc} \texttt{age} & & \texttt{statusSingle} & & \texttt{age:statusSingle} \\ \begin{bmatrix}32\\33\\32\\33\\26\end{bmatrix} & \times & \begin{bmatrix}0\\0\\1\\0\\1\end{bmatrix} & = & \boxed{\begin{bmatrix}0\\0\\32\\0\\26\end{bmatrix}} \end{array}$$

There would be two model vectors for the interaction between age and marital status.

There would be four model vectors for the interaction between sex and marital status.

$$\begin{array}{ccccc} \texttt{sexM} & & \texttt{statusSingle} & & \texttt{sexM:statusSingle} \\ \begin{bmatrix}1\\0\\1\\1\\1\end{bmatrix} & \times & \begin{bmatrix}0\\0\\1\\0\\1\end{bmatrix} & = & \boxed{\begin{bmatrix}0\\0\\1\\0\\1\end{bmatrix}} \end{array}$$

$$\begin{array}{ccccc} \texttt{sexM} & & \texttt{statusMarried} & & \texttt{sexM:statusMarried} \\ \begin{bmatrix}1\\0\\1\\1\\1\end{bmatrix} & \times & \begin{bmatrix}1\\1\\0\\1\\0\end{bmatrix} & = & \boxed{\begin{bmatrix}1\\0\\0\\1\\0\end{bmatrix}} \end{array}$$

$$\begin{array}{ccccc} \texttt{sexF} & & \texttt{statusSingle} & & \texttt{sexF:statusSingle} \\ \begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix} & \times & \begin{bmatrix}0\\0\\1\\0\\1\end{bmatrix} & = & \boxed{\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}} \end{array}$$

$$\begin{array}{ccccc} \texttt{sexF} & & \texttt{statusMarried} & & \texttt{sexF:statusMarried} \\ \begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix} & \times & \begin{bmatrix}1\\1\\0\\1\\0\end{bmatrix} & = & \boxed{\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}} \end{array}$$

The resulting model vectors for the interaction terms between categorical predictors are indicators of the different combinations of the initial predictors.

Model Vectors: Intercept

The intercept is a special model vector of all ones.

$$\begin{array}{c} \texttt{Intercept} \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{array}$$

Model Matrix

When constructing a model, you supply a list of model terms, such as

```
lm.1 = lm(wage ~ 1 + age + sex, data = toy)
```

The software (e.g., R) then translates each of the predictors to their model vectors

$$
\begin{array}{cccc}
\texttt{Intercept} & \texttt{age} & \texttt{sexM} & \texttt{sexF} \\
\end{array}
$$

$$
\begin{bmatrix}
1 & 32 & 1 & 0 \\
1 & 33 & 0 & 1 \\
1 & 32 & 1 & 0 \\
1 & 33 & 1 & 0 \\
1 & 26 & 1 & 0
\end{bmatrix}
$$

A set of vectors having the same dimension is called a *matrix*. The matrix composed of all the model vectors is called the **model matrix** or **design matrix**.

When an intercept is included in the model, not all of the indicators for categorical predictors are used!

Fitted Values

```
> summary(lm.1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.5402    13.6620  -1.430    0.289
age           0.8346     0.4083   2.044    0.178
sexM          6.4802     2.6928   2.407    0.138
```

To obtain fitted values, the model vectors are scaled by the coefficients (scalars) and added. The fitted values are a **linear combination** of the model vectors.

$$\begin{array}{cccccccc} & \texttt{Intercept} & & \texttt{age} & & \texttt{sexM} & & \texttt{fitted} \\ -19.54 & \begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix} & +0.83 & \begin{bmatrix}32\\33\\32\\33\\26\end{bmatrix} & +6.48 & \begin{bmatrix}1\\0\\1\\1\\1\end{bmatrix} & = & \begin{bmatrix}13.65\\8.00\\13.65\\14.48\\8.64\end{bmatrix}\end{array}$$

Model Vectors and Redundancy

Vectors in the model matrix cannot be redundant. Vectors are redundant when one vector can be written as a *linear combination of the others*.

One common source of redundancy is when *all* indicator vectors for categorical variables are included in the model along with an intercept.

Consider the two sex indicator variables

$$
\begin{array}{ccccc}
 & \texttt{sexF} & & \texttt{sexM} & & \texttt{Intercept} \\
1 & \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} & +\,1 & \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} & = & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}
\end{array}
$$

The intercept vector is just a linear combination of the two sex indicator vectors.

To fit a model that includes both indicators for sex, the intercept would need to be dropped from the model.

To drop the intercept, we include `-1` in the formula of the `lm()` function

```
> lm(wage ~ age + sex - 1, data = toy)

Coefficients:
     Estimate Std. Error t value Pr(>|t|)
age    0.8346     0.4083   2.044    0.178
sexF -19.5402    13.6620  -1.430    0.289
sexM -13.0600    12.6054  -1.036    0.409
```

This results in the coefficients for the categorical predictors having different interpretations.

```
> model.matrix(lm.1)

  (Intercept) age sexM
1           1  32    1
2           1  33    0
3           1  32    1
4           1  33    1
5           1  26    1
```

The **model matrix** or **design matrix** can be obtained by inputting the name of a fitted model to the `model.matrix()` function