Logistic Regression

Fundamentals

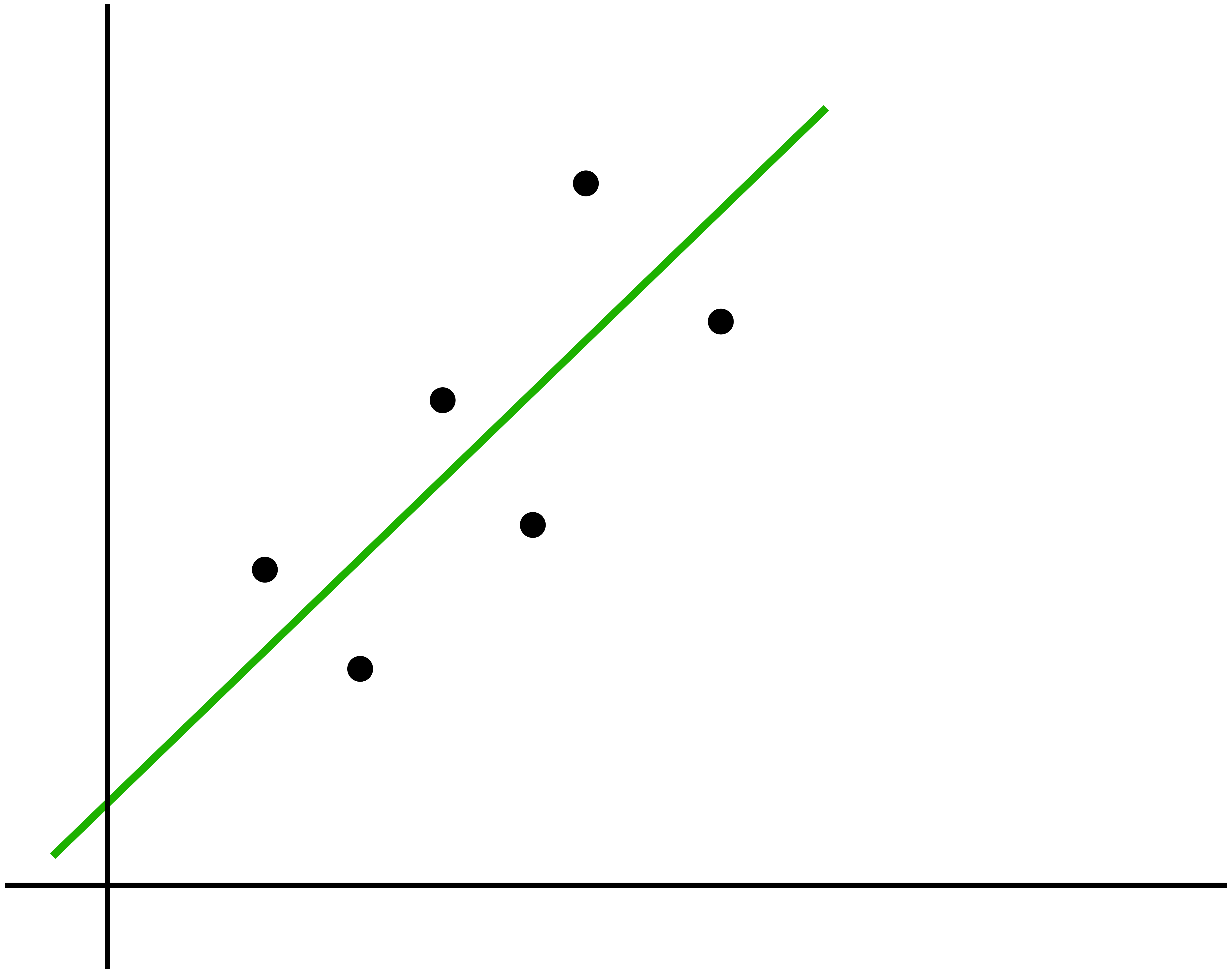
Rahul Singh
rsingh@arrsingh.com

x is the independent variable

$$y = f(x)$$

y is the dependent variable

Linear Regression



The dependent variable y is **continuous**.
It can hold any value.

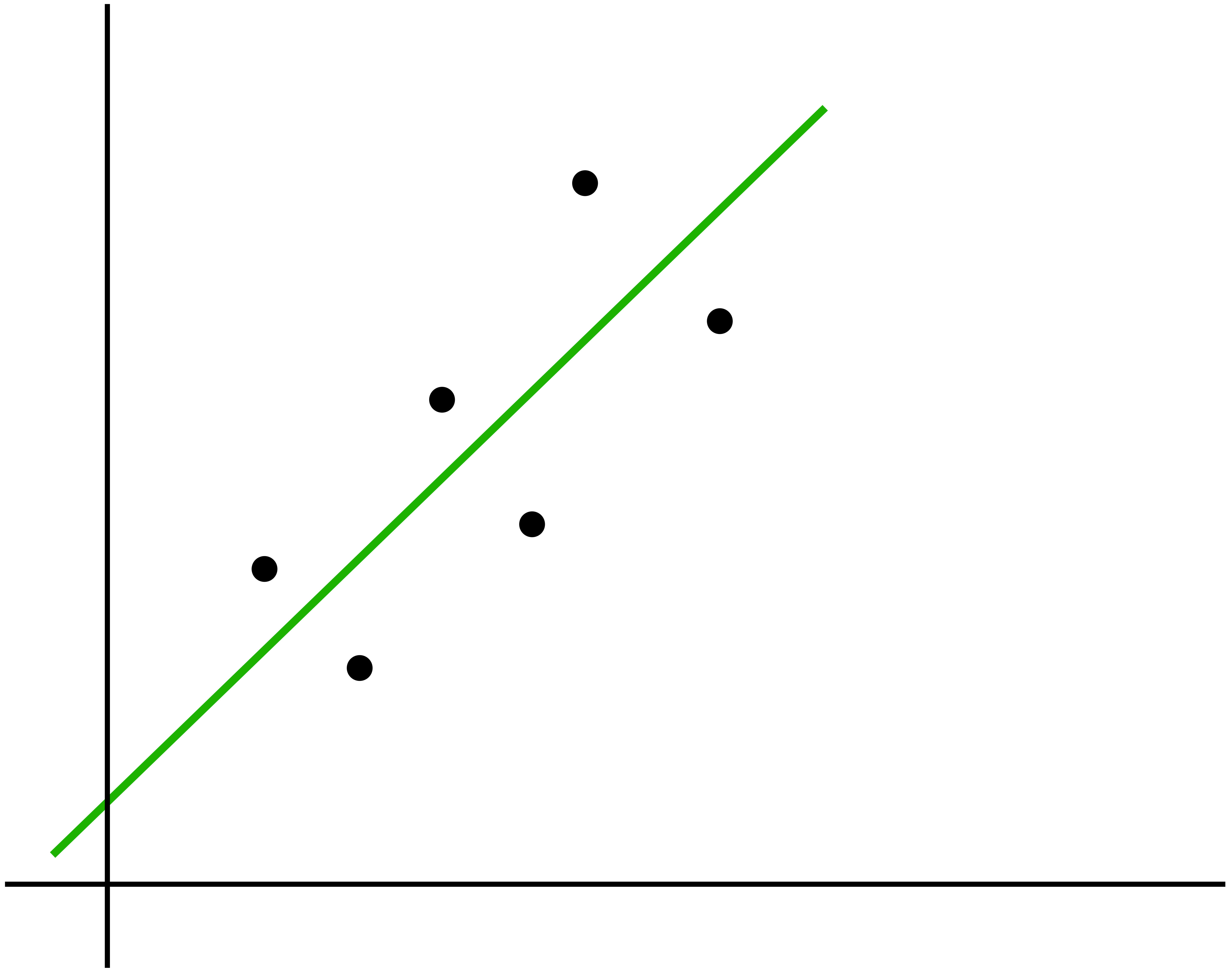
x is the independent variable

$$y = f(x)$$

y is the dependent variable

Linear Regression

Linear Models can be used to **predict** the values of y (dependent variable) for any value of x (independent variables)



The dependent variable y is **continuous**.

It can hold any value.

What if we wanted to predict a binary value

How do we build a model that returns a binary value (true / false)

x is the independent variable

$$y = f(x)$$

y (dependent variable) that returns 0 (false) or 1 (true)

What if we wanted to predict a binary value

How do we build a model that returns a binary value (true / false)

x is the independent variable

$$y = f(x)$$

y (dependent variable) that returns 0 (false) or 1 (true)

Real world applications…

Is this email Spam or not?

Is this transaction Fraud or not?

Will I have heart disease or not?

Lets take a simple example…

A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

Question: Can we **predict** whether a person that drives at 69 mph (avg) will be involved in an accident?

A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

Lets begin by plotting this data.

A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

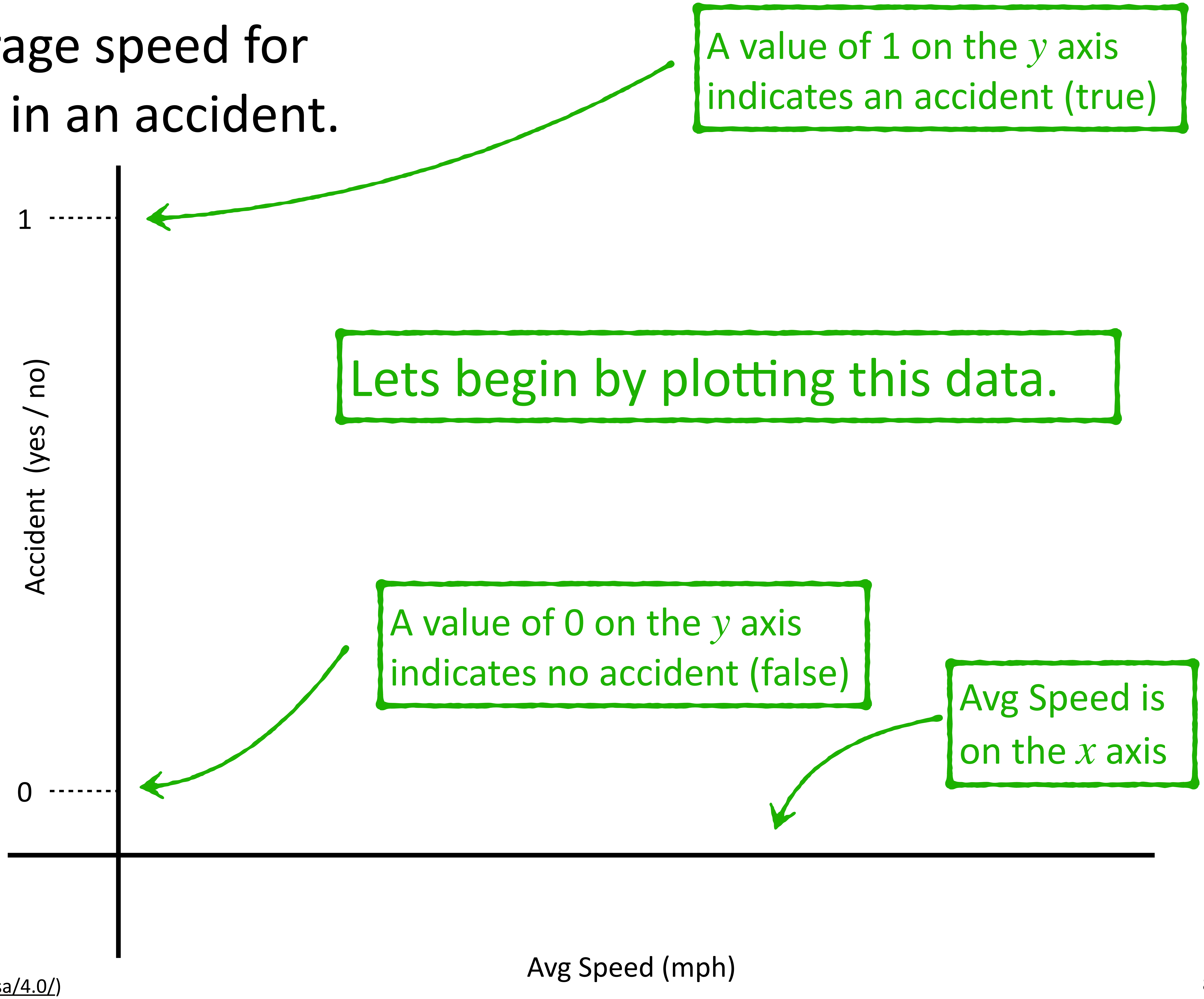
Lets begin by plotting this data.

Question: Can we **predict** whether a person that drives at 69 mph (avg) will be involved in an accident?

A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

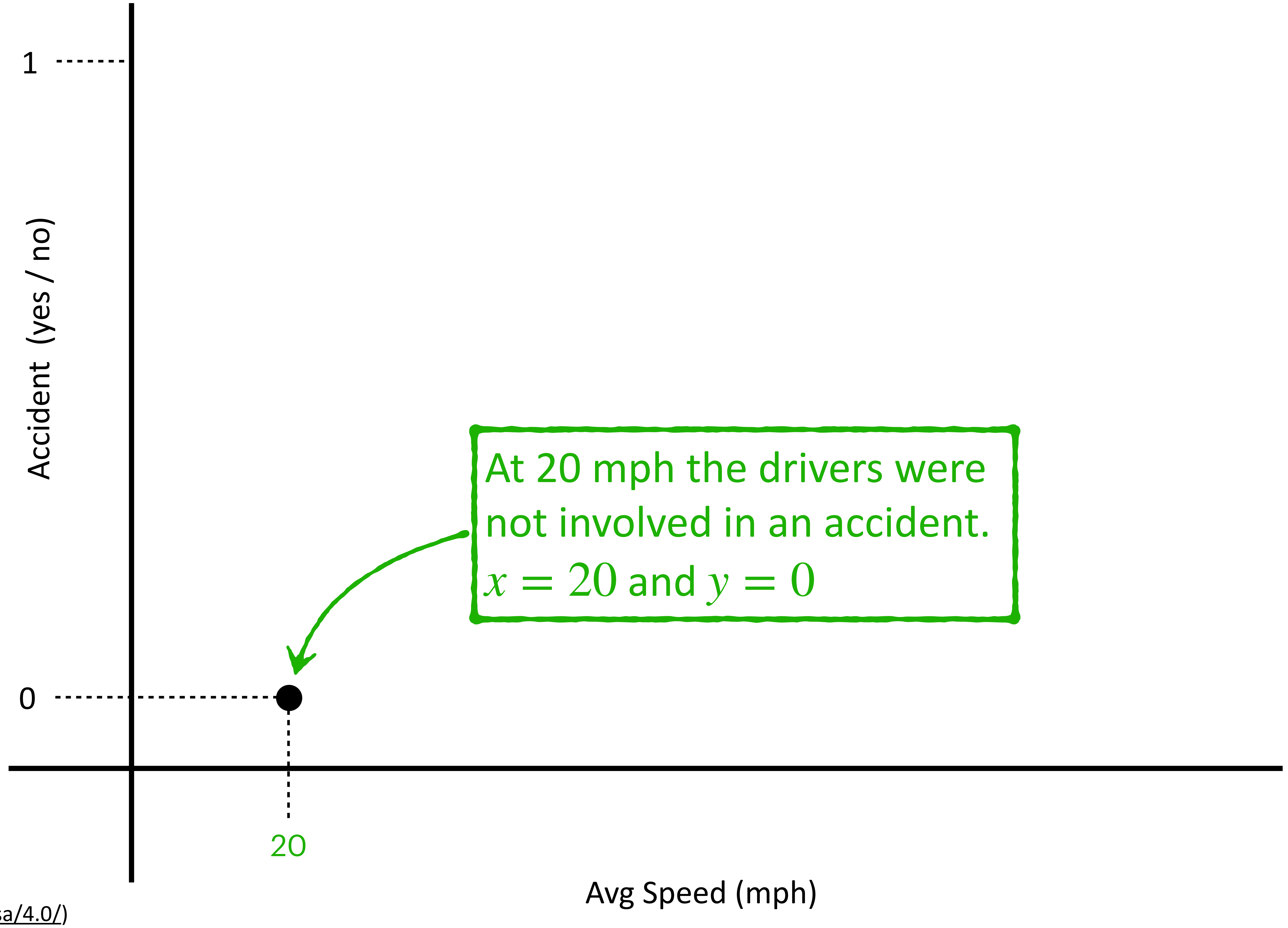
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example...

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

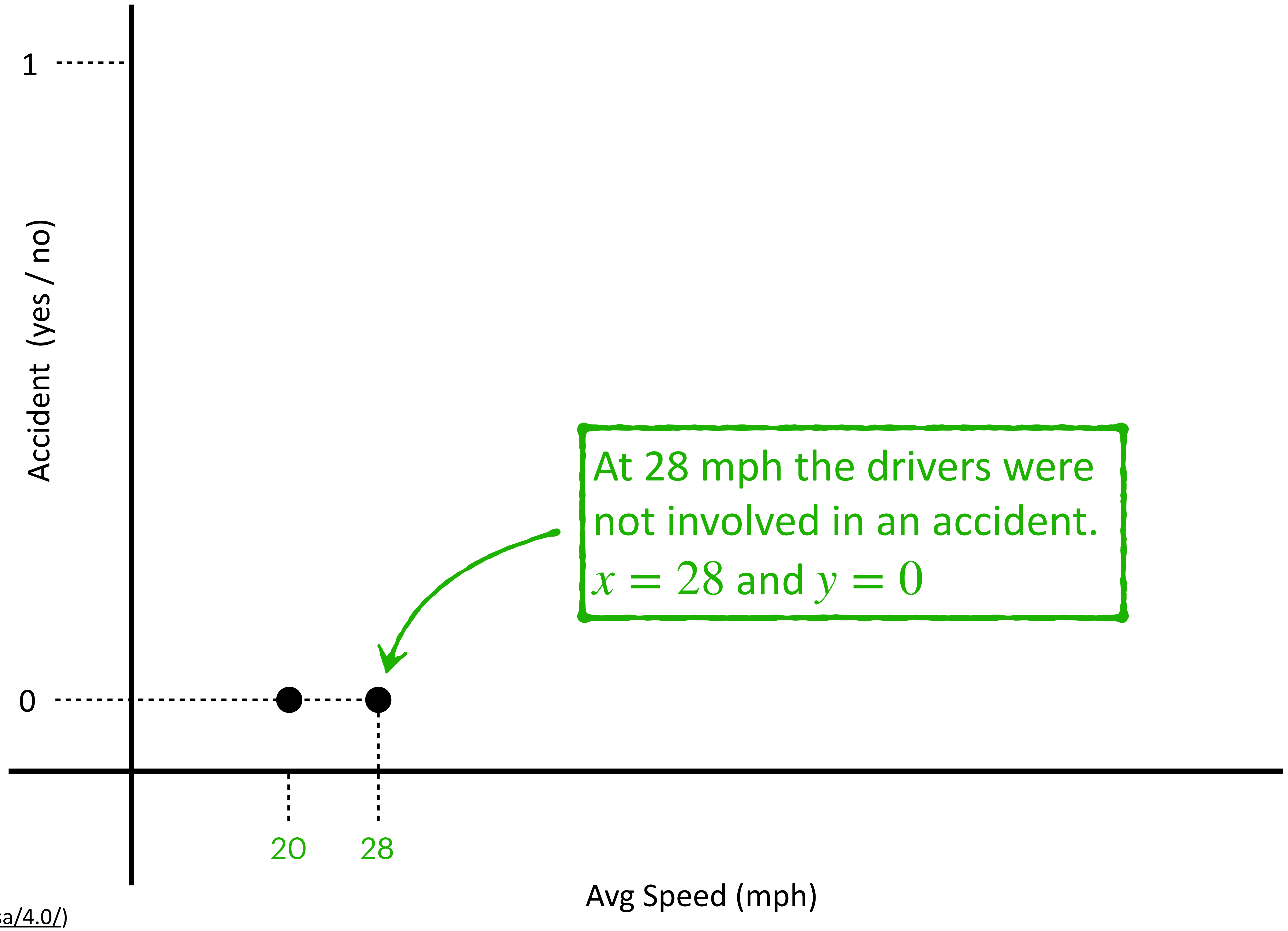
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

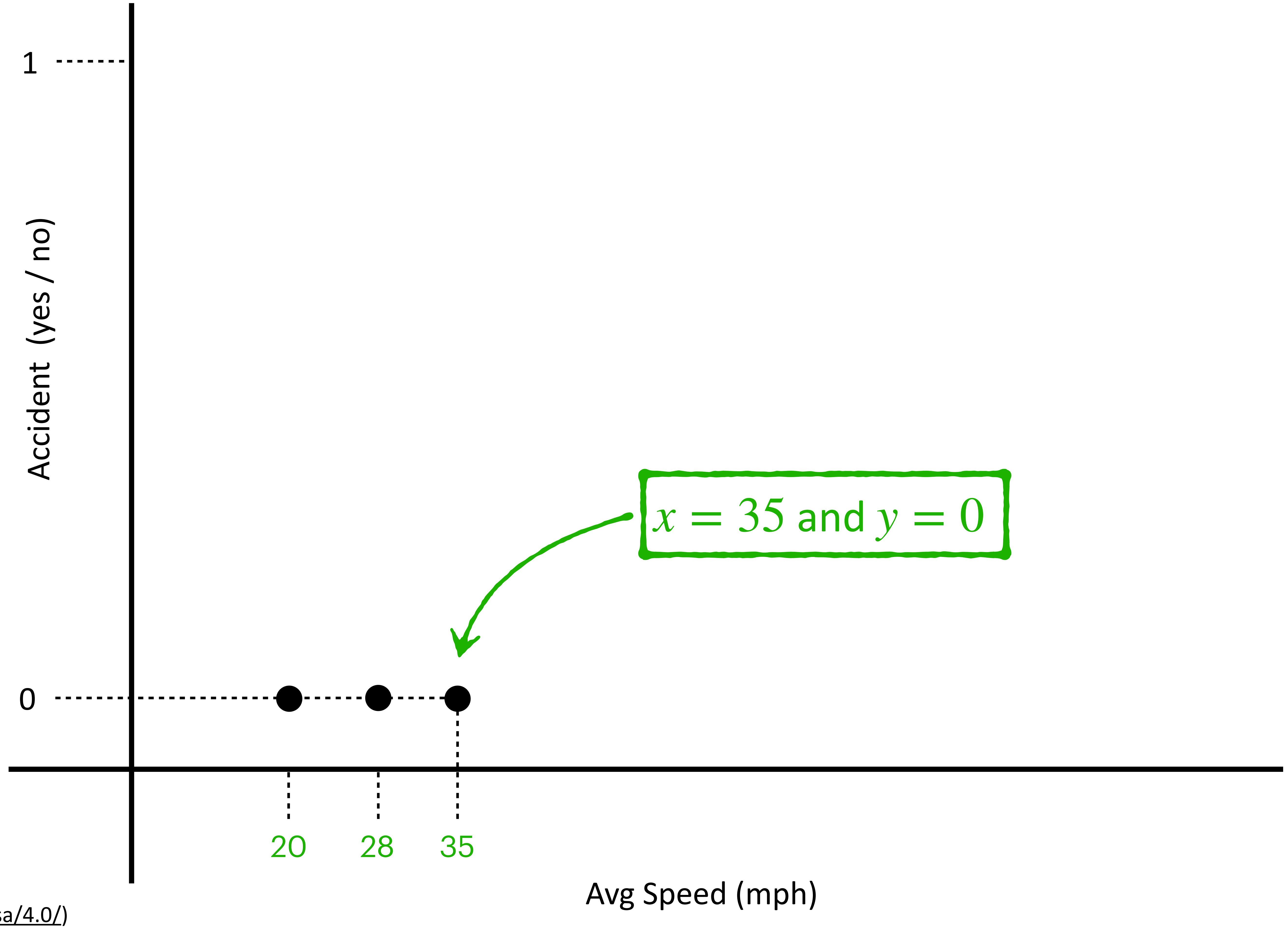
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example...

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

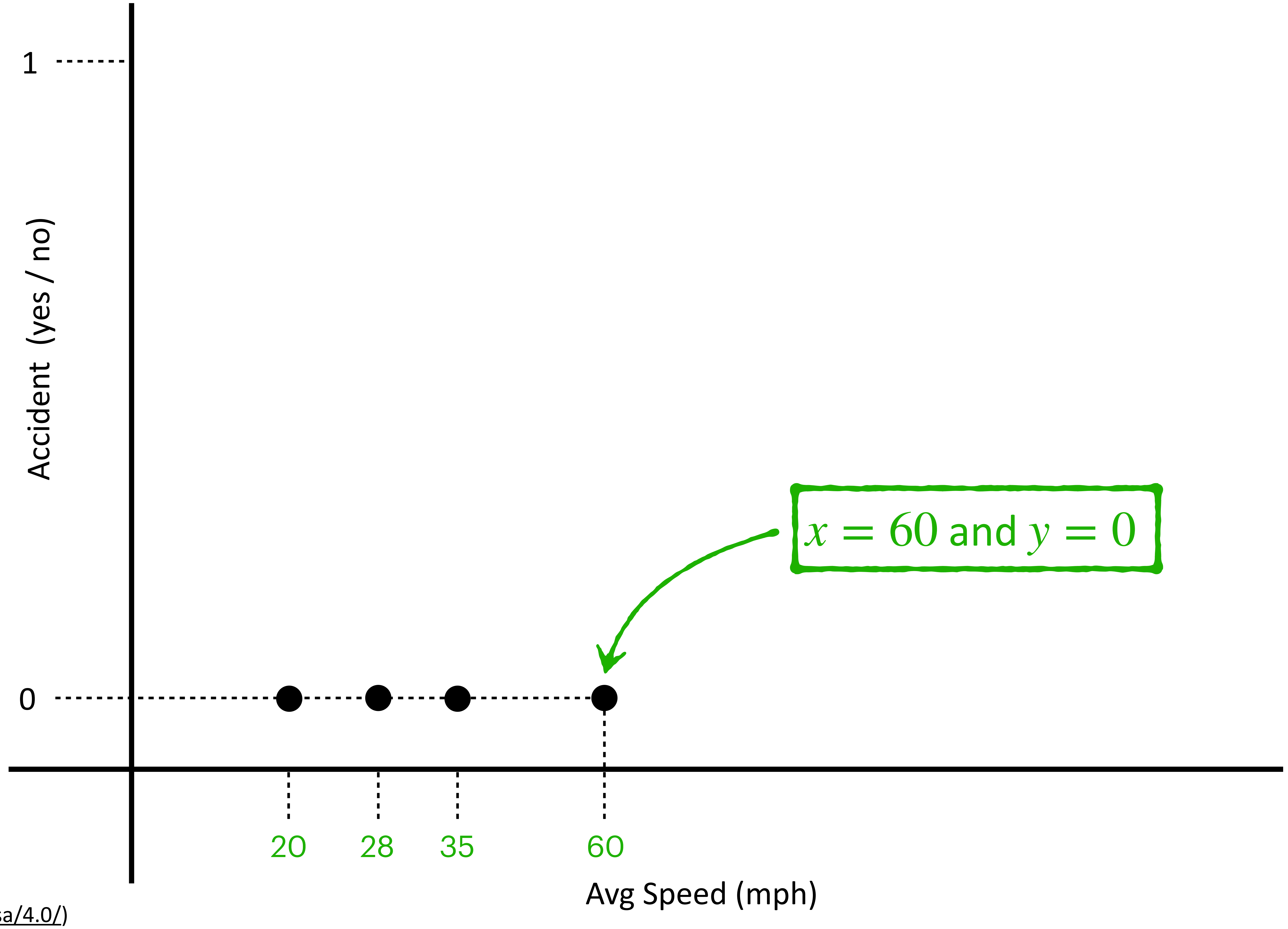
Avg Speed (mph)	Accident
20	No
28	No
35	**No**
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

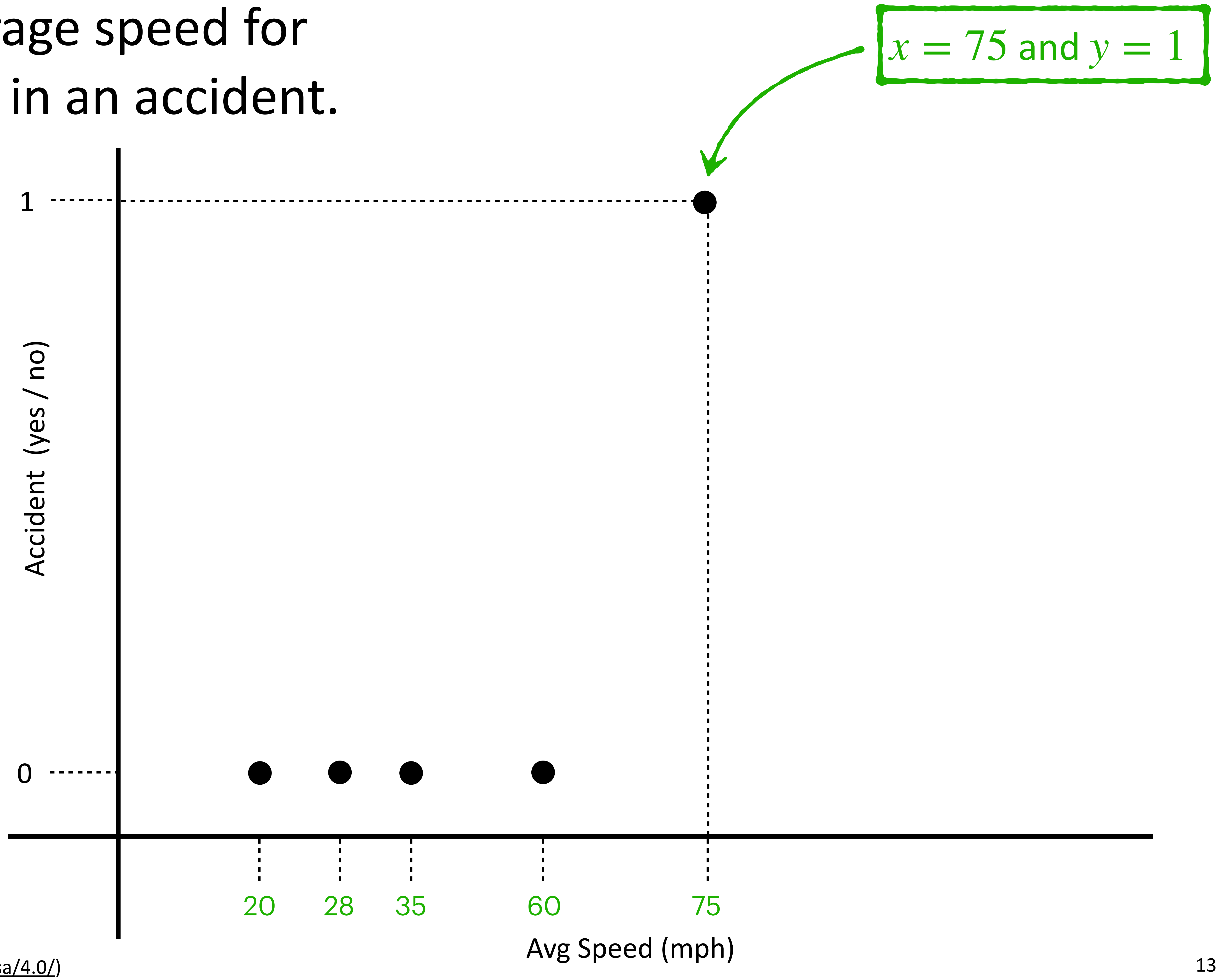


A simple example…

Drivers that speed, tend to get into more accidents.

An insurance company has data for average speed for drivers and whether they were involved in an accident.

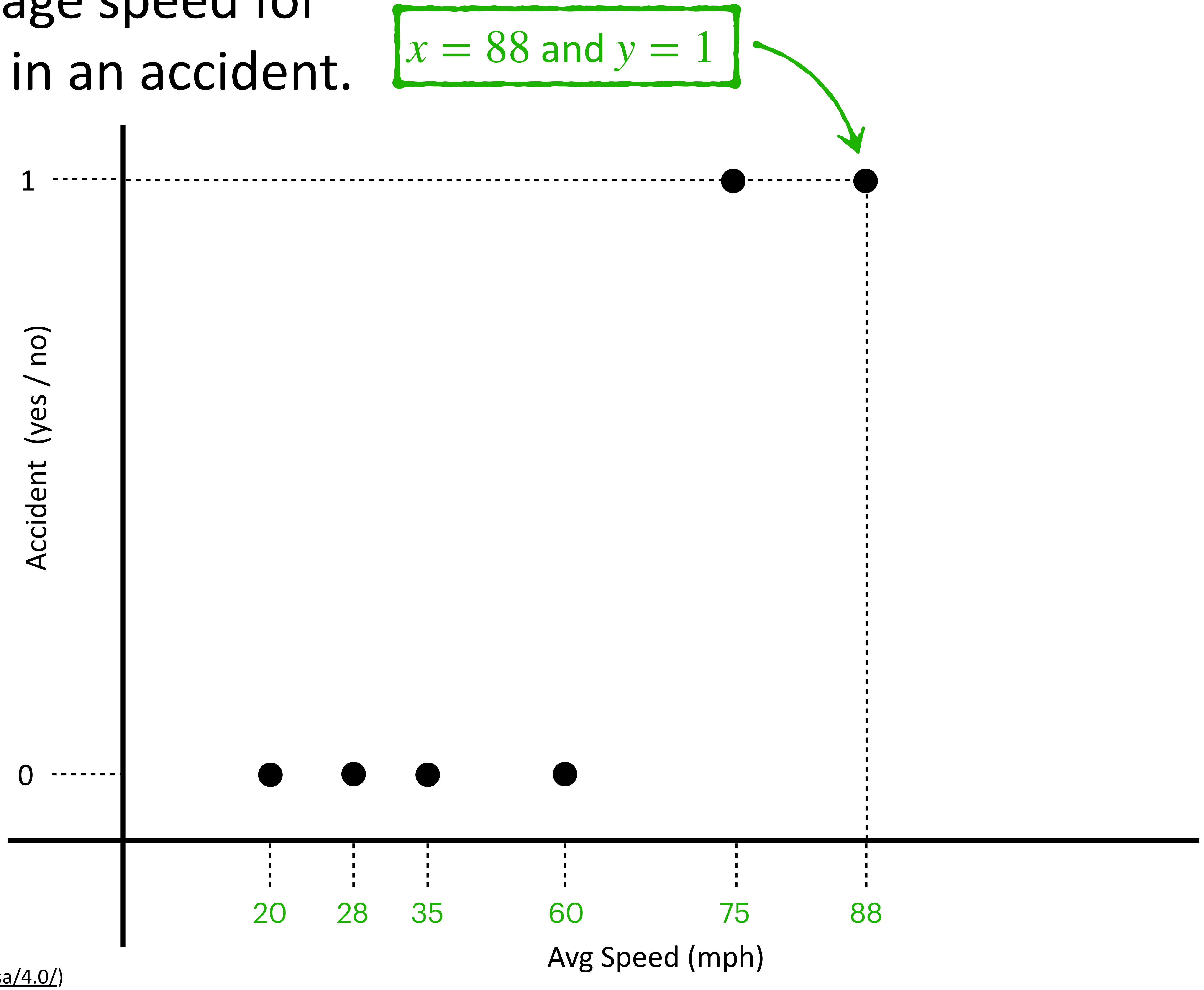
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	**Yes**
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

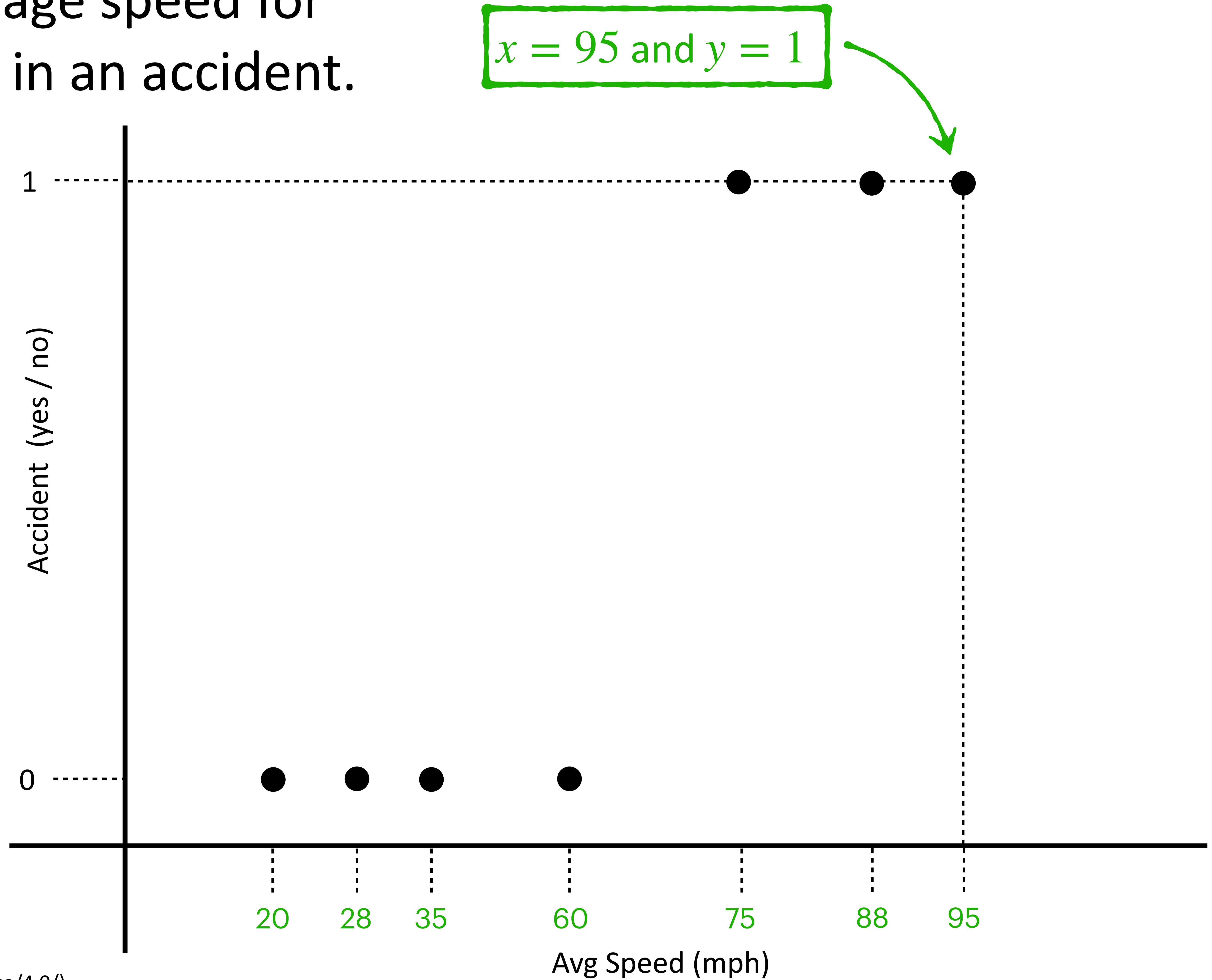
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

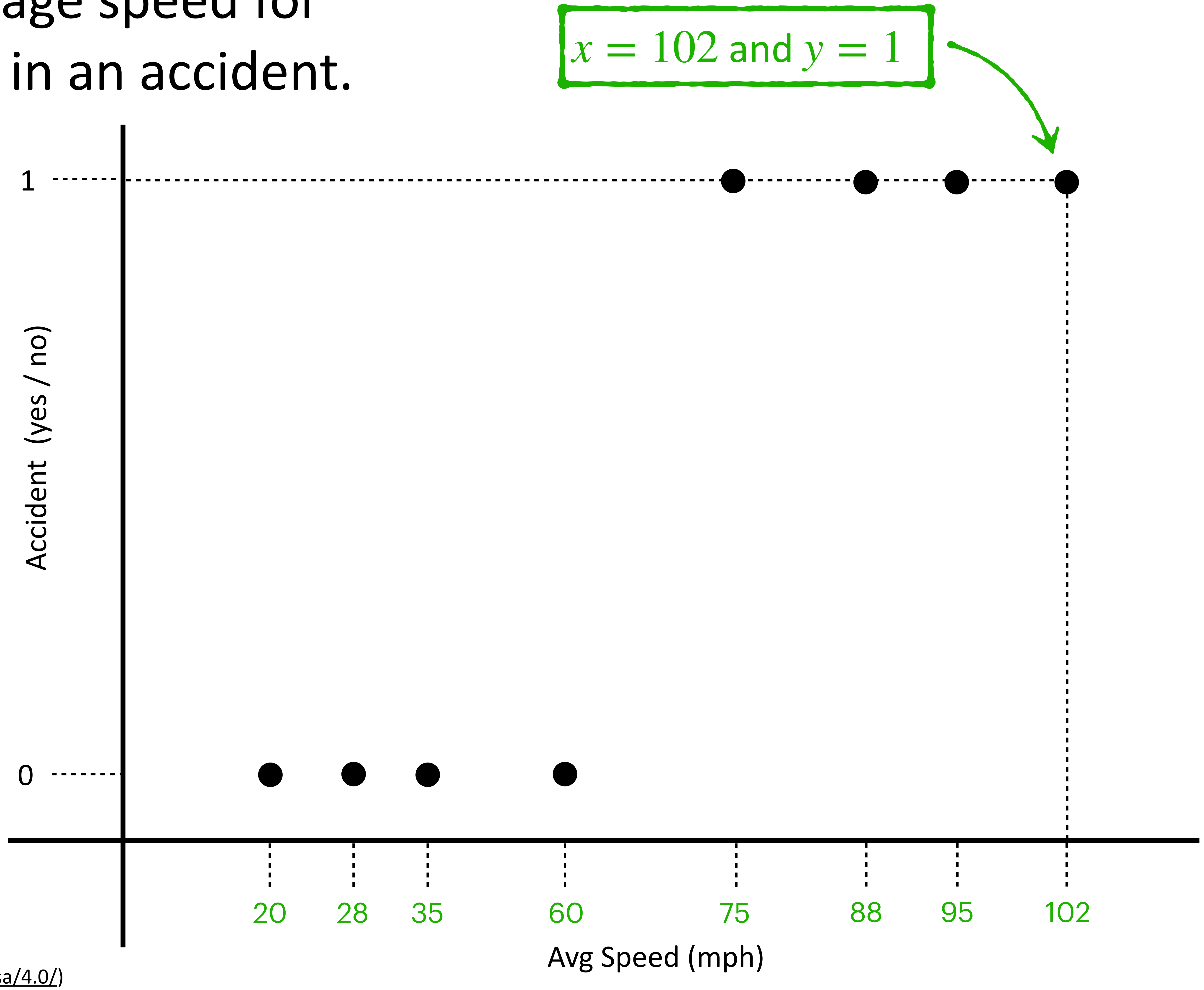
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

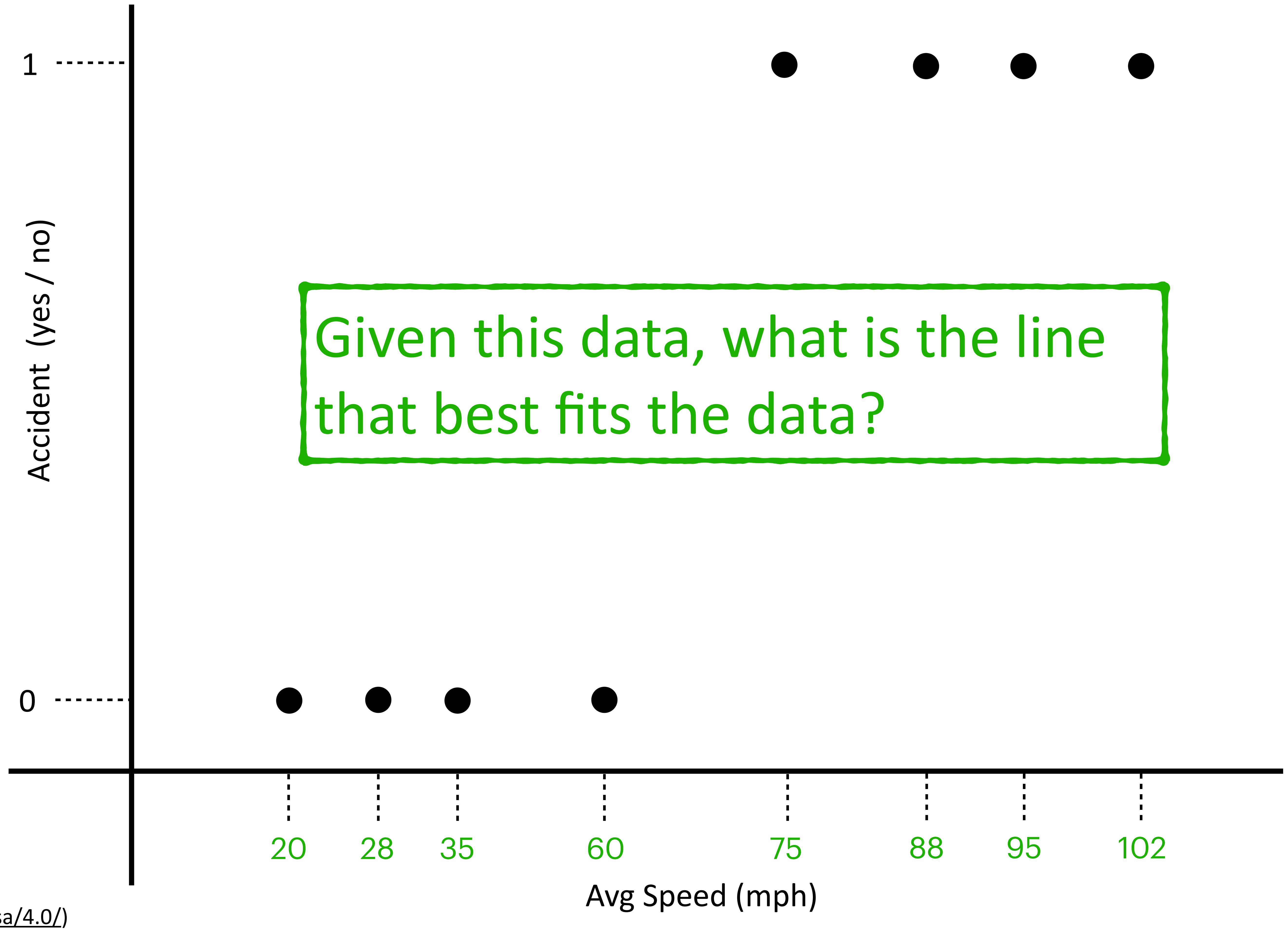
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

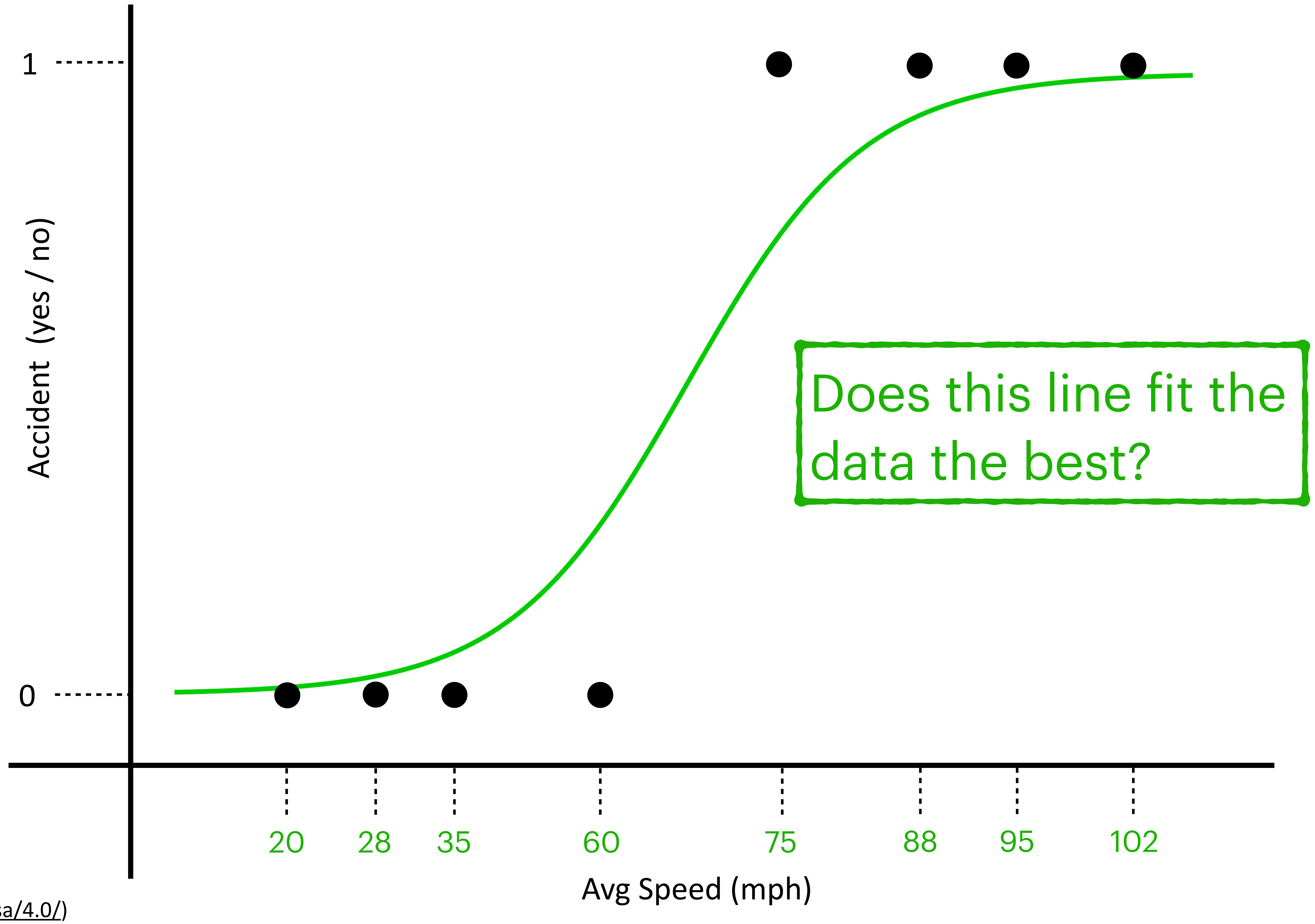


Logistic Regression

A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

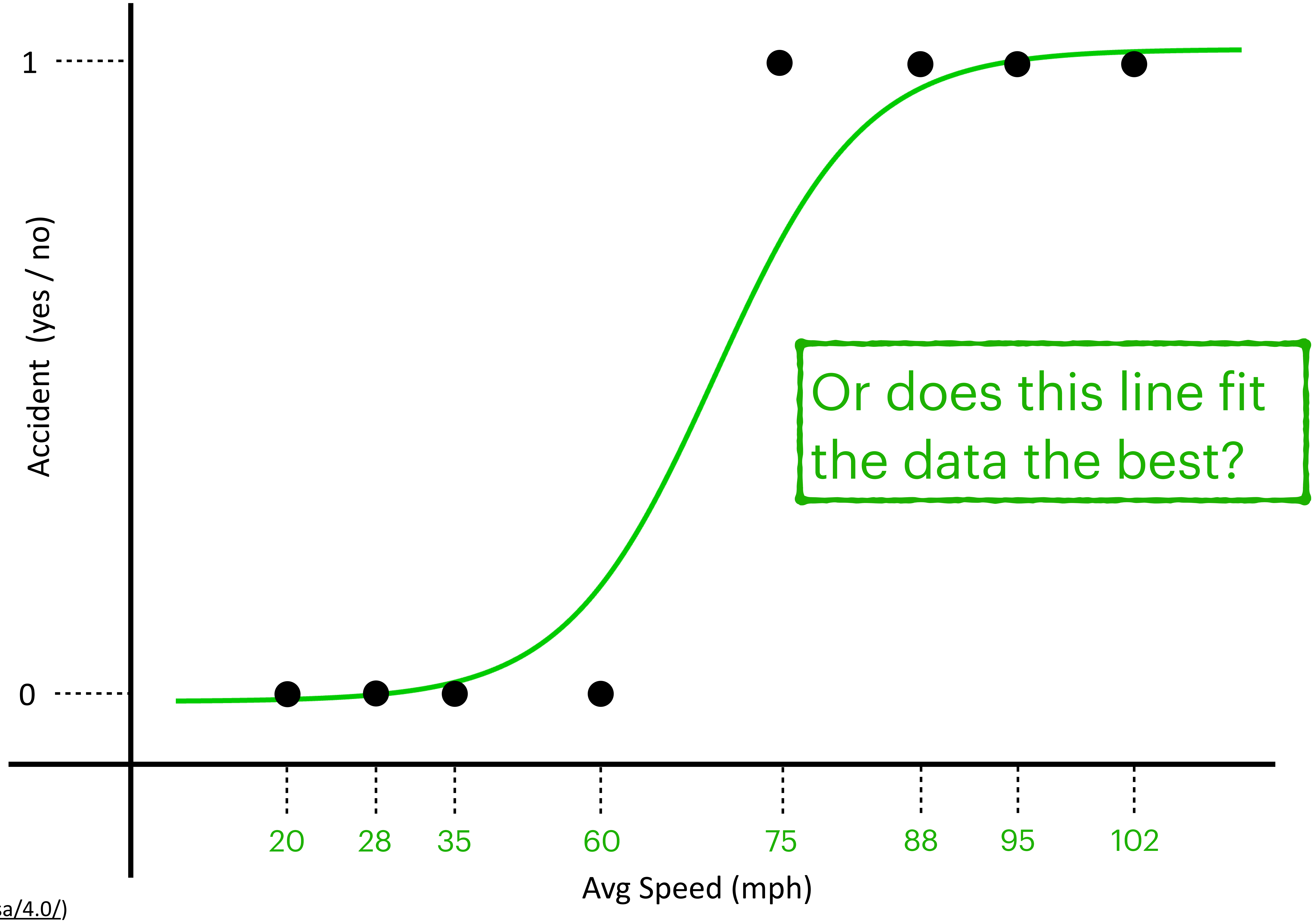
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

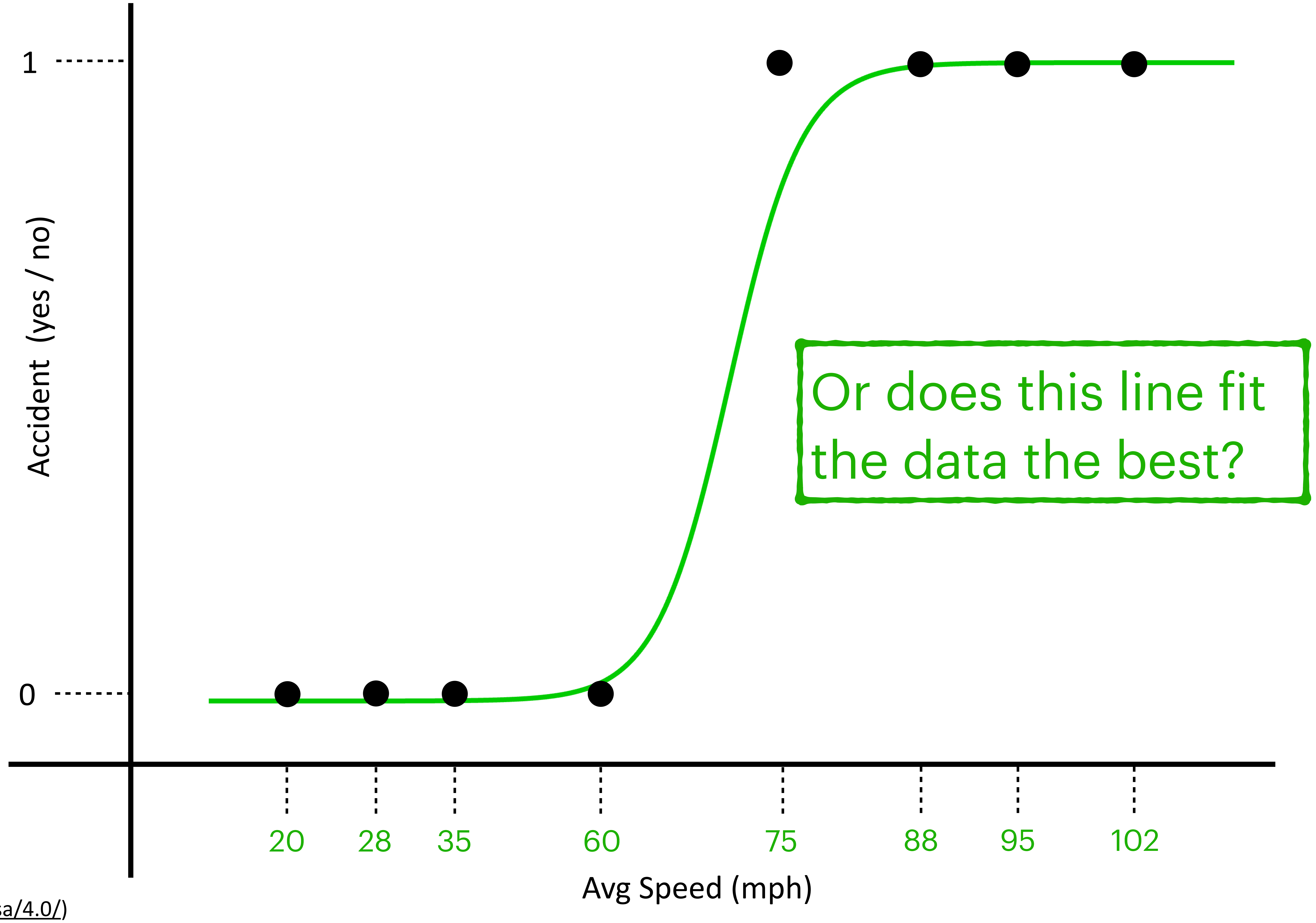
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

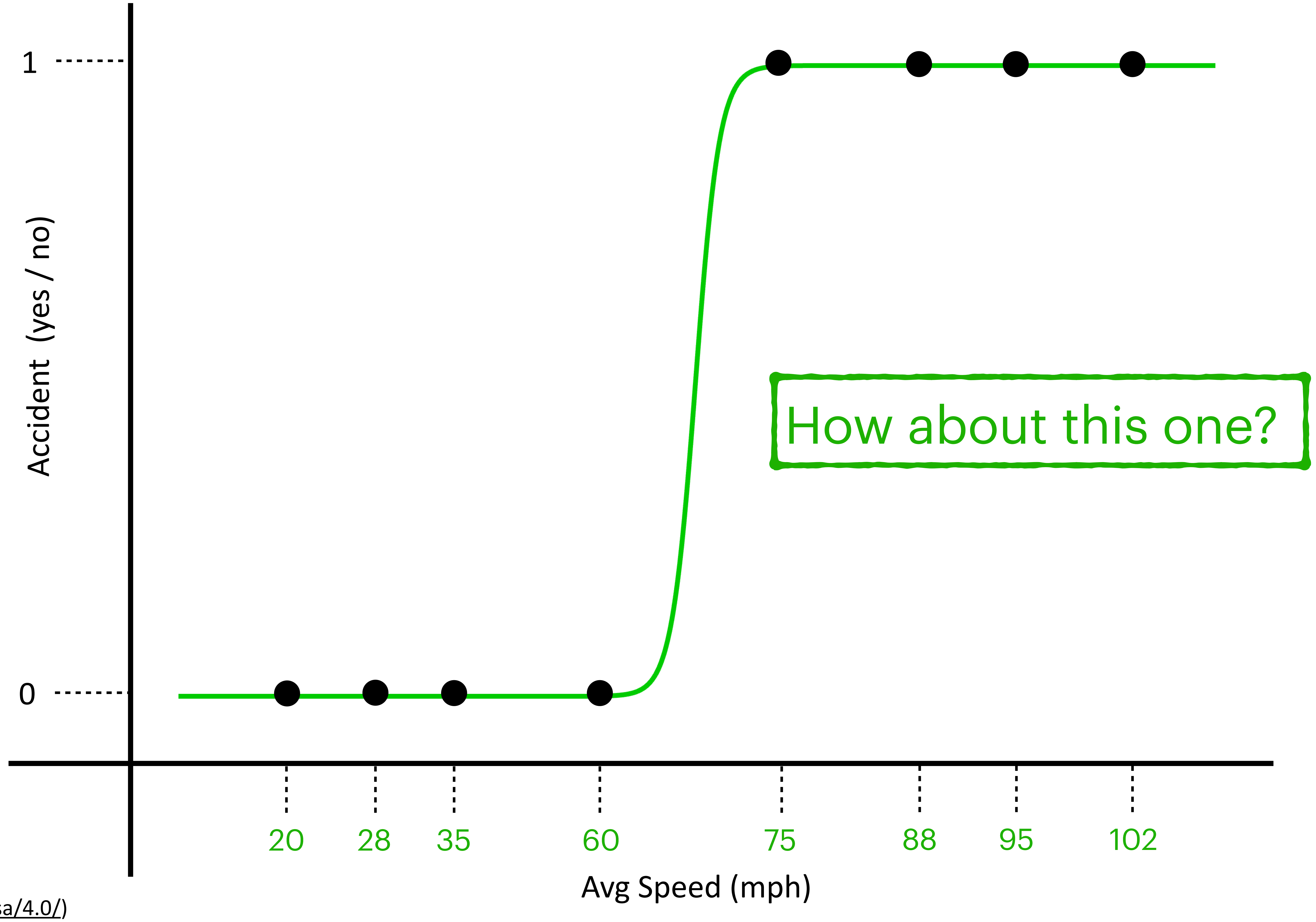
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

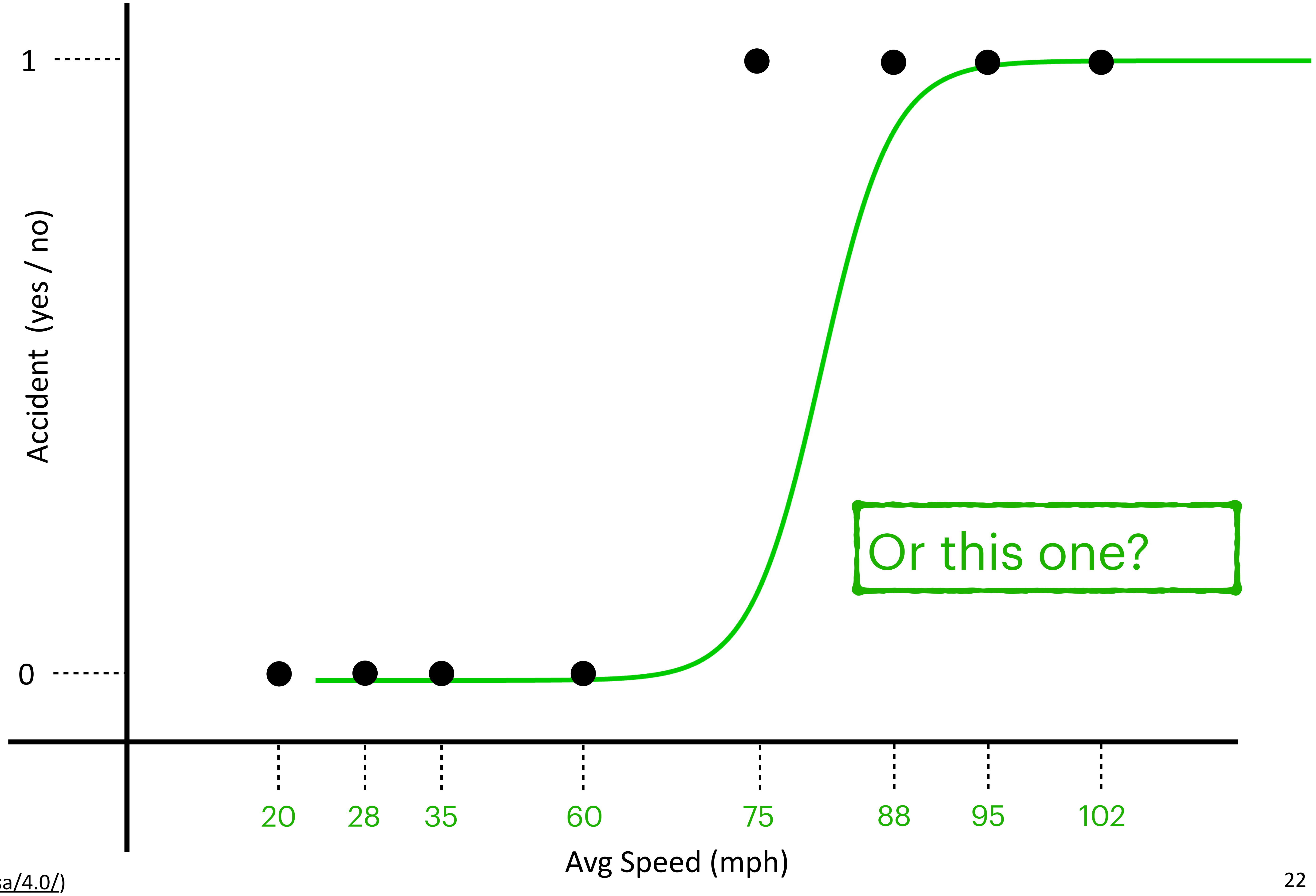
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

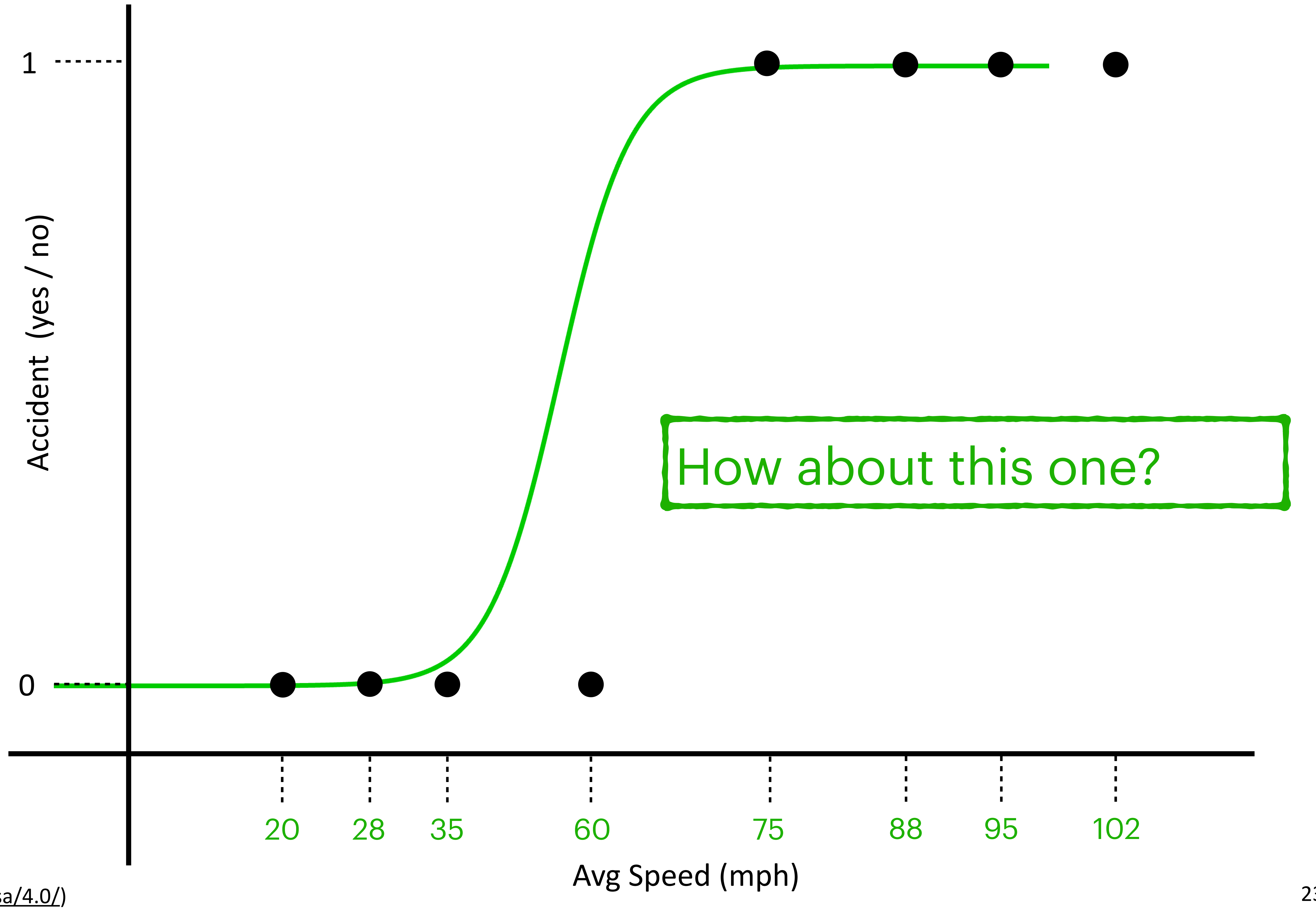
Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



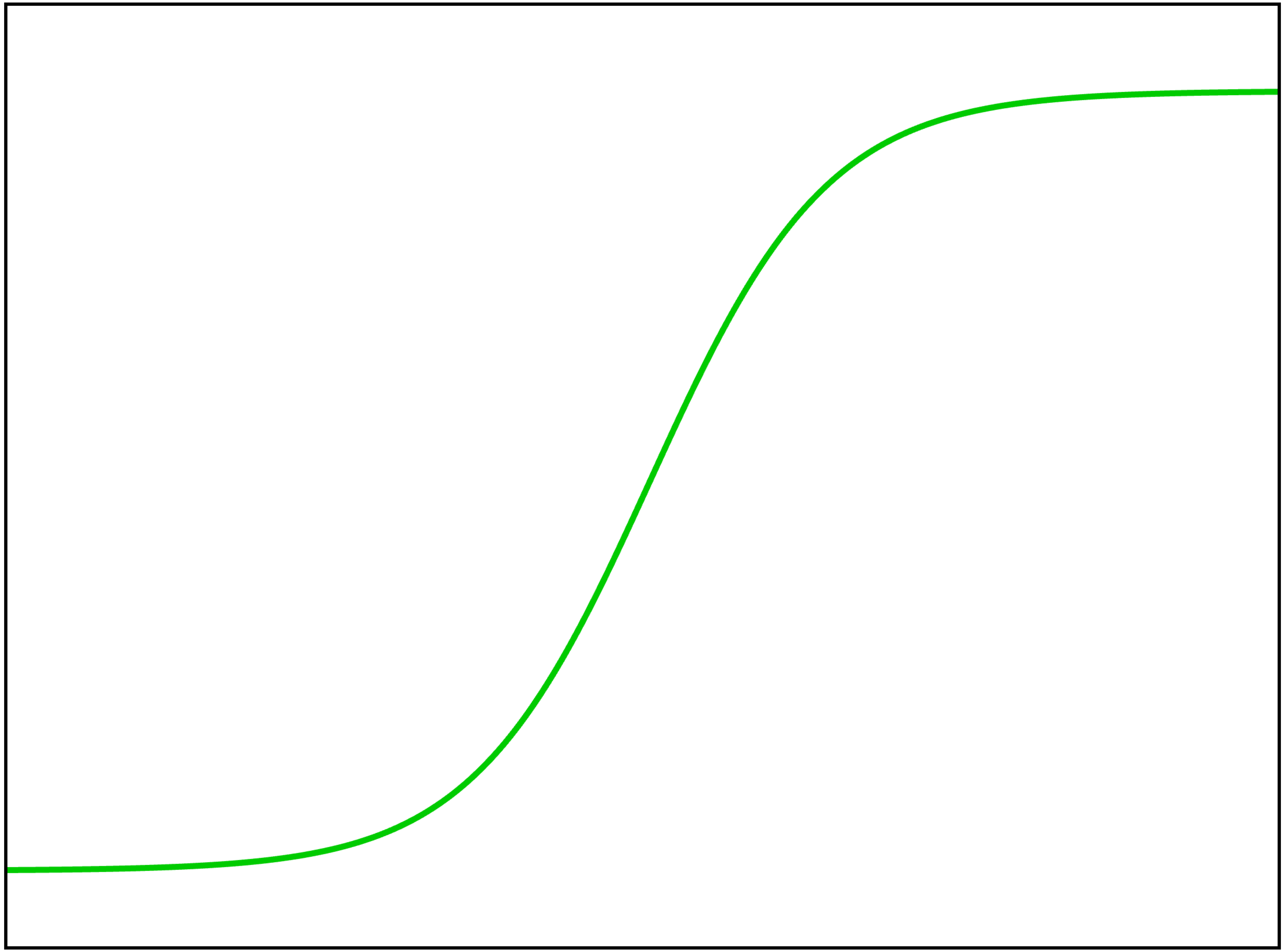
A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes



Lets dive deep into this shape…

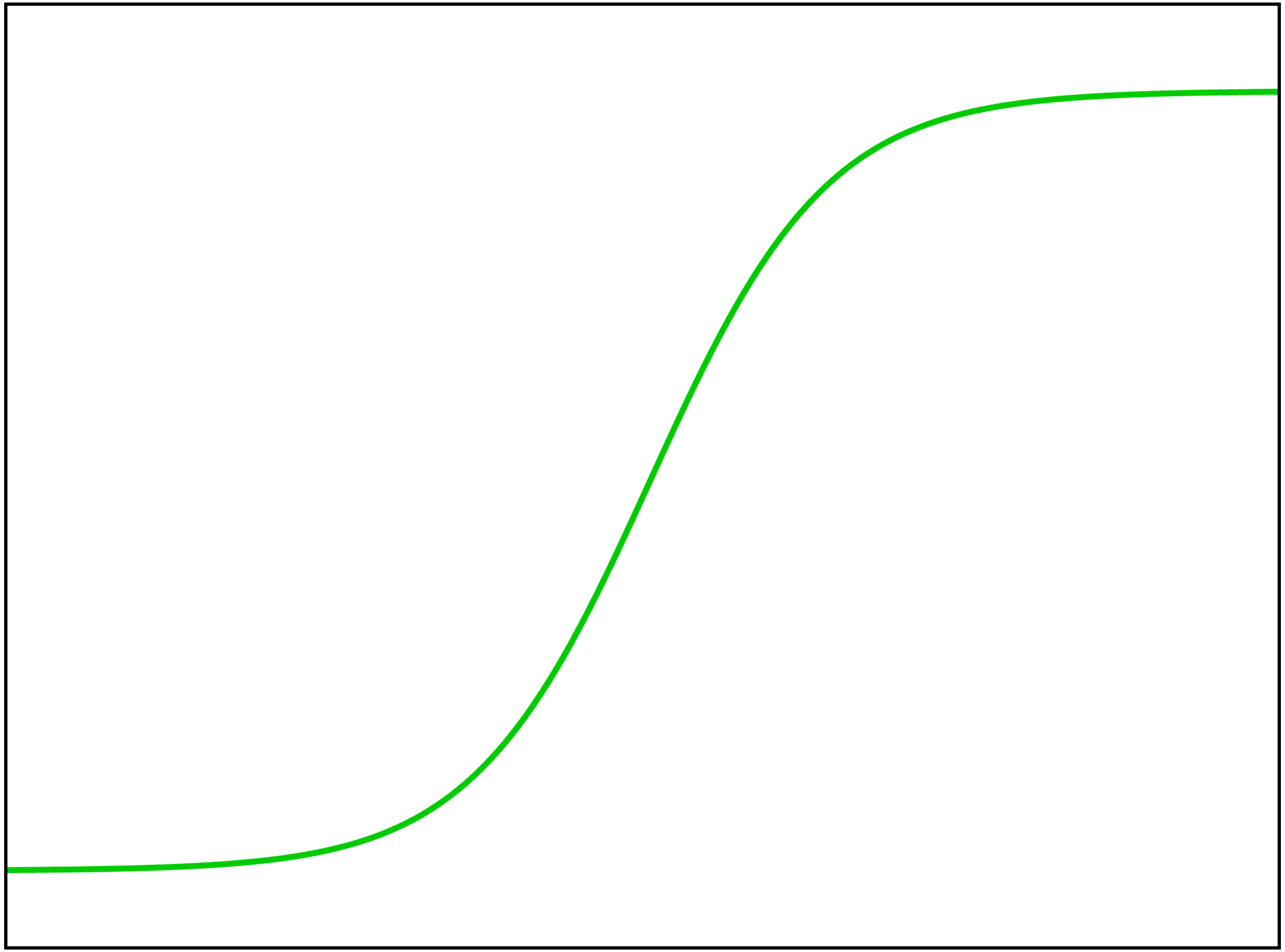


Sigmoid Curve (S - shaped)

There are many different types of sigmoid functions. We'll use the Logistic Function

Logistic Function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

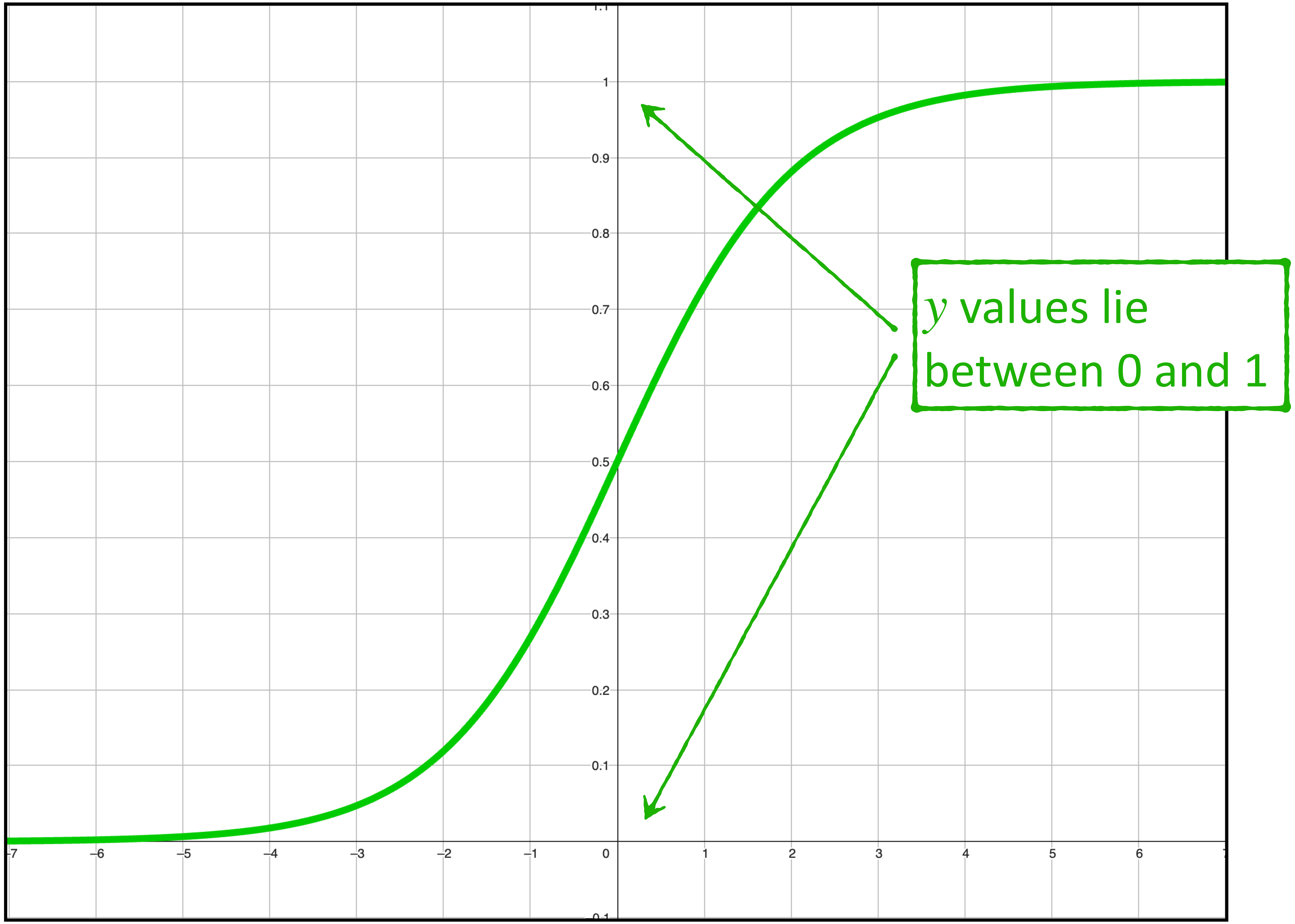


Sigmoid Curve (S - shaped)

There are many different types of sigmoid functions. We'll use the Logistic Function

Logistic Function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

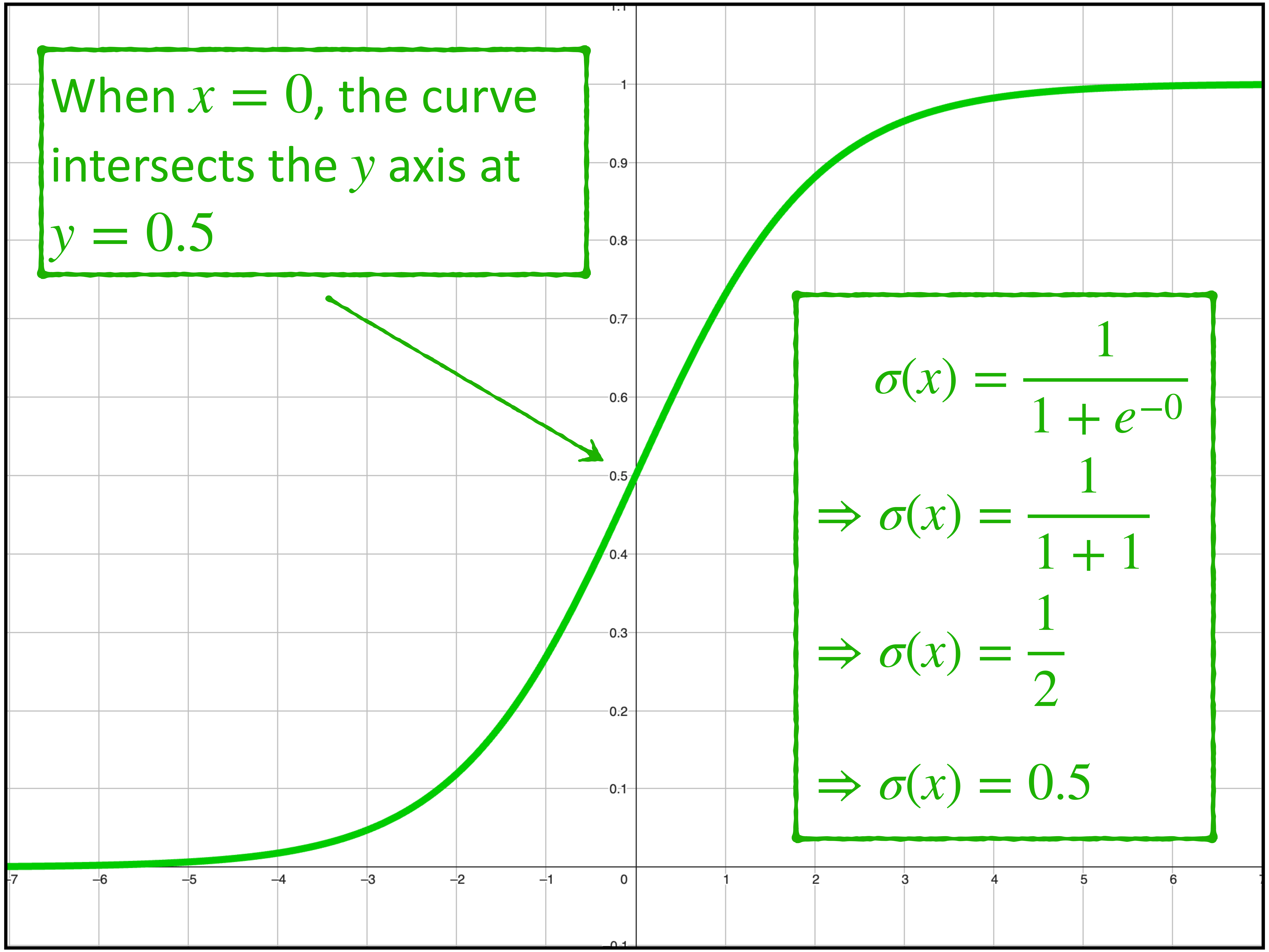


Sigmoid Curve (S - shaped)

There are many different types of sigmoid functions. We'll use the Logistic Function

Logistic Function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



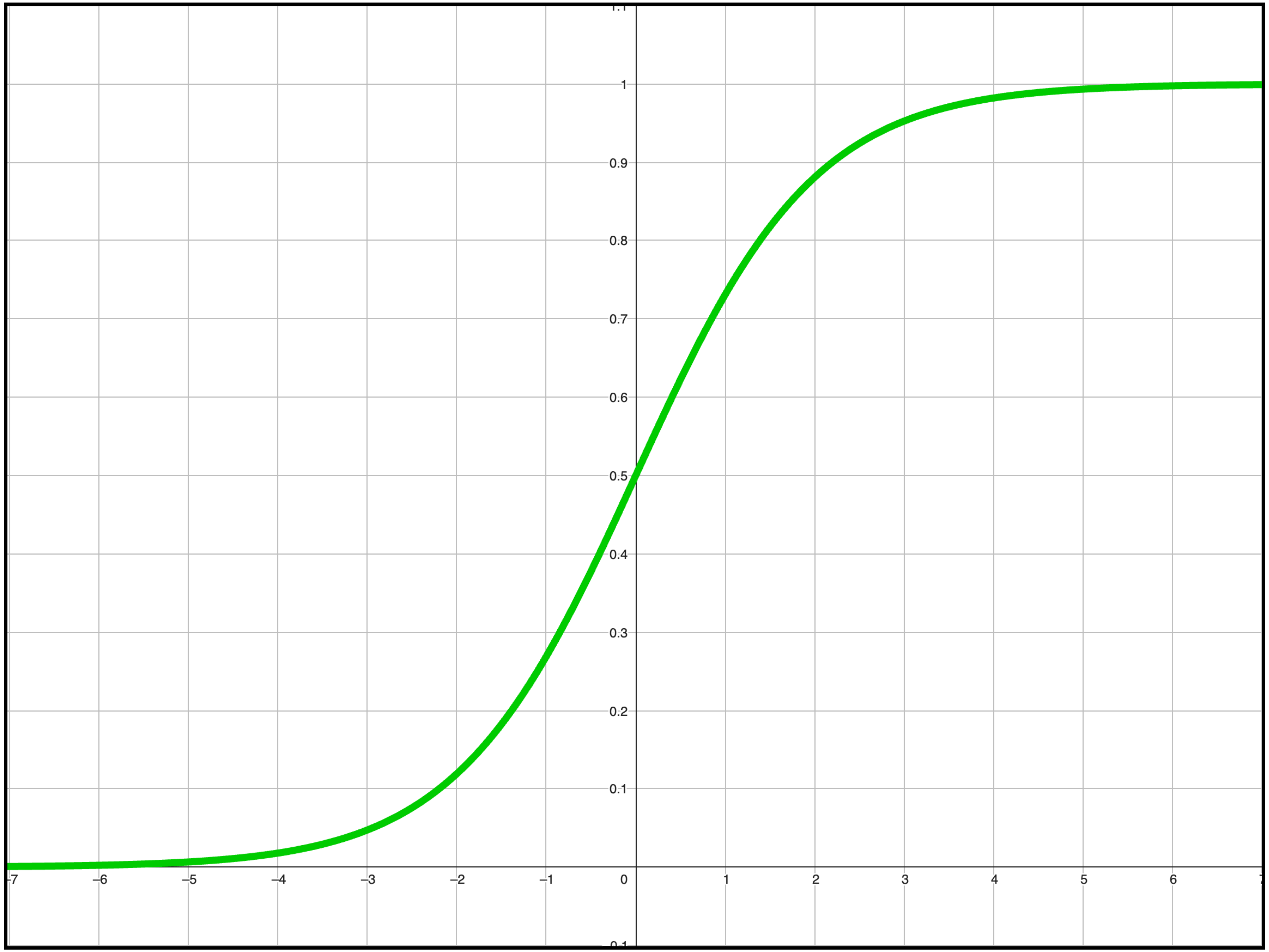
Sigmoid Curve (S - shaped)

There are many different types of sigmoid functions. We'll use the Logistic Function

Logistic Function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

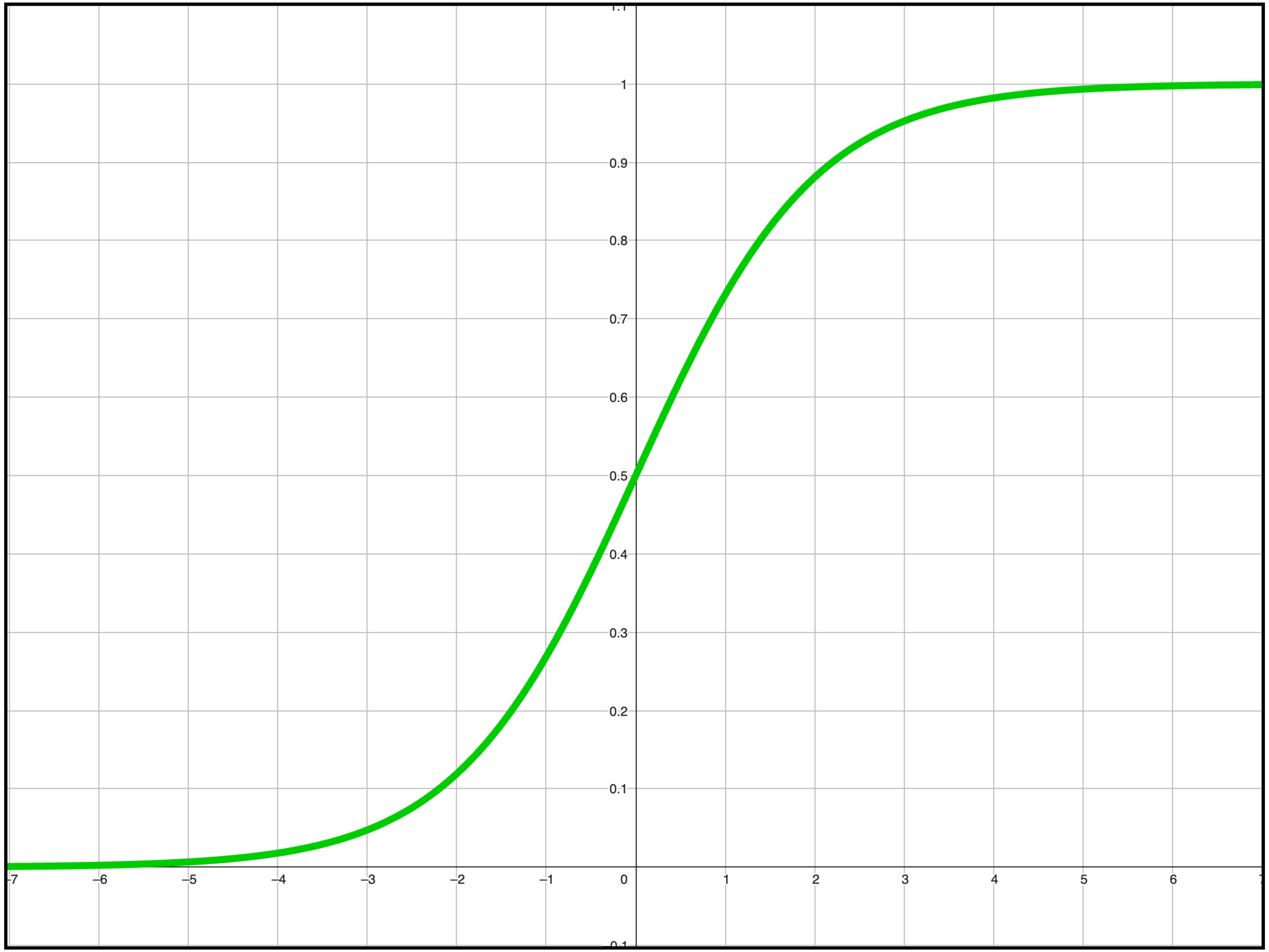
The Logistic Function, converts inputs into a range between 0 and 1.



Logistic Function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

The shape of the curve changes for different values of x

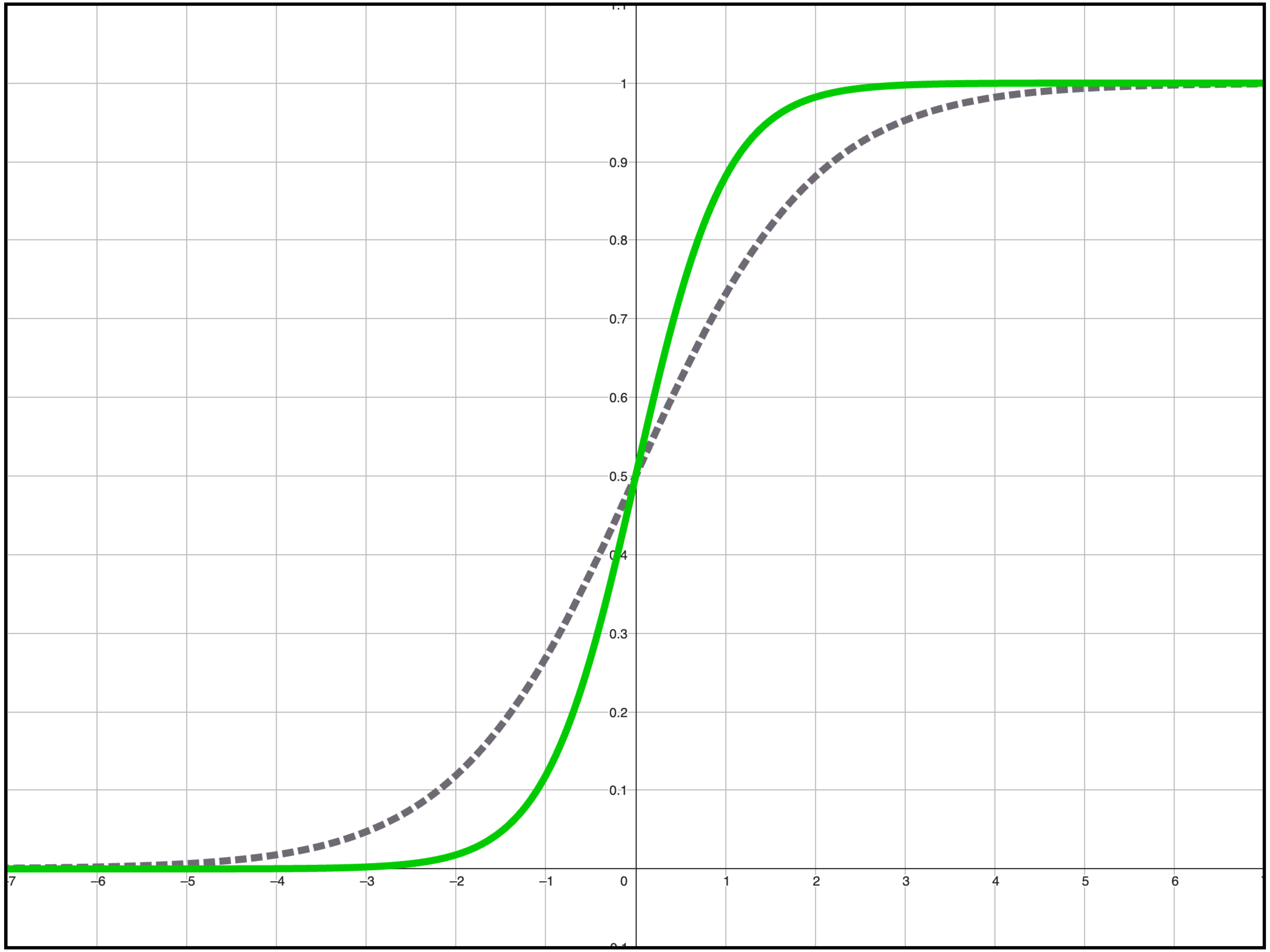


Logistic Function

$$\sigma(2x) = \frac{1}{1 + e^{-2x}}$$

Multiplying x by 2 makes the curve steeper

It still intersects the y axis at $y = 0.5$

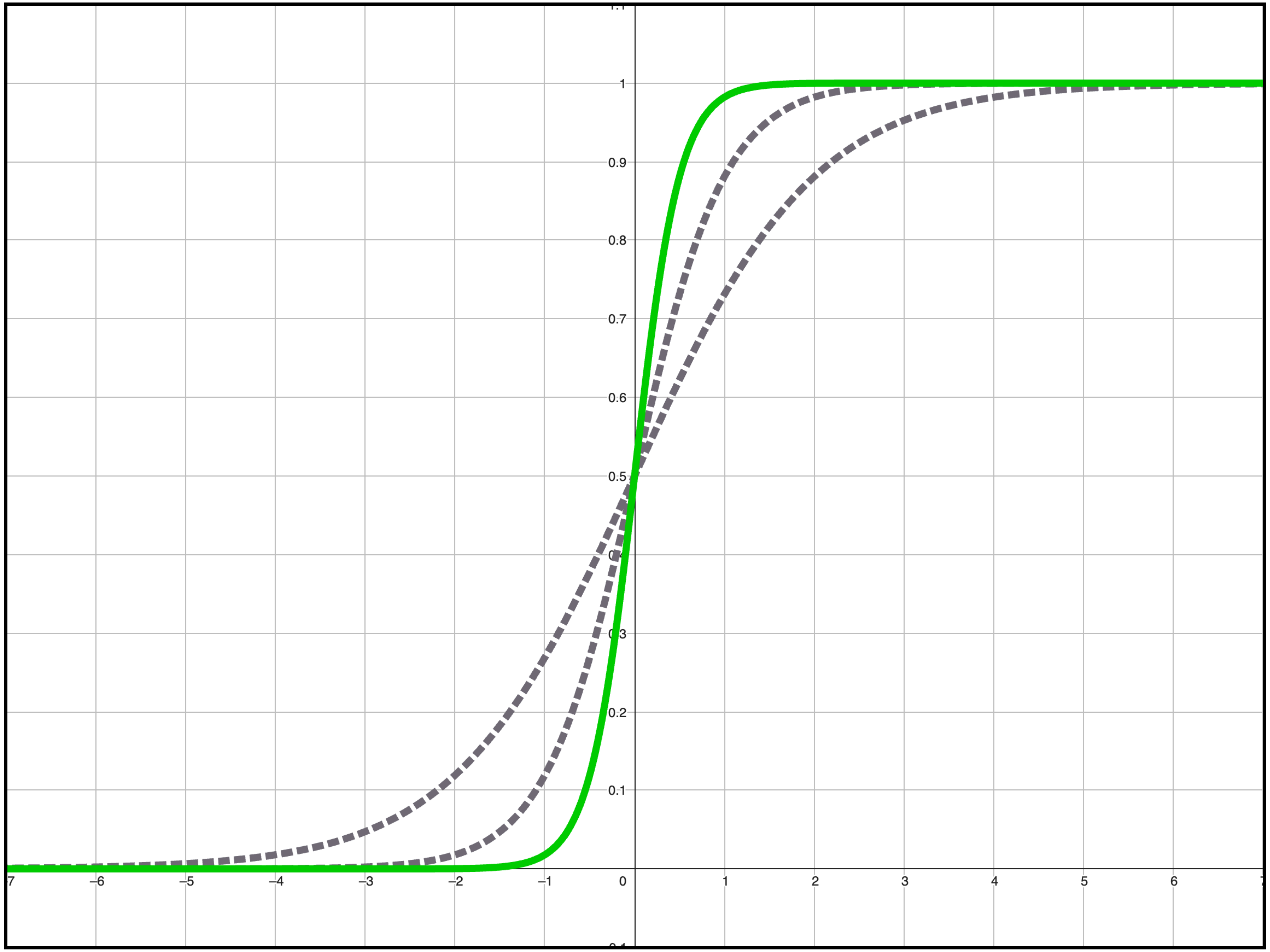


Logistic Function

$$\sigma(4x) = \frac{1}{1 + e^{-4x}}$$

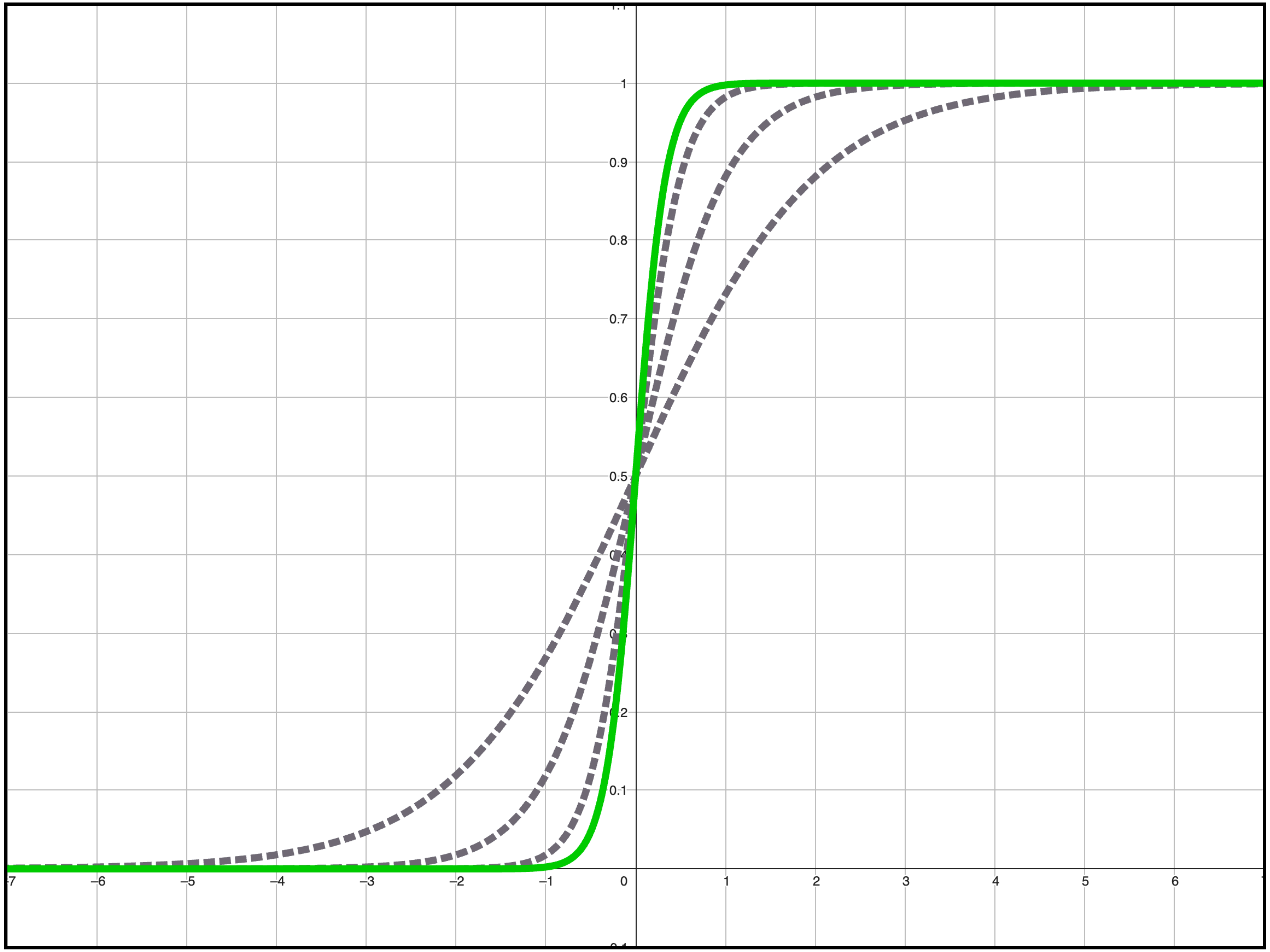
Multiplying x by 4 makes the curve even steeper

It still intersects the y axis at $y = 0.5$



Logistic Function

$$\sigma(6x) = \frac{1}{1 + e^{-6x}}$$



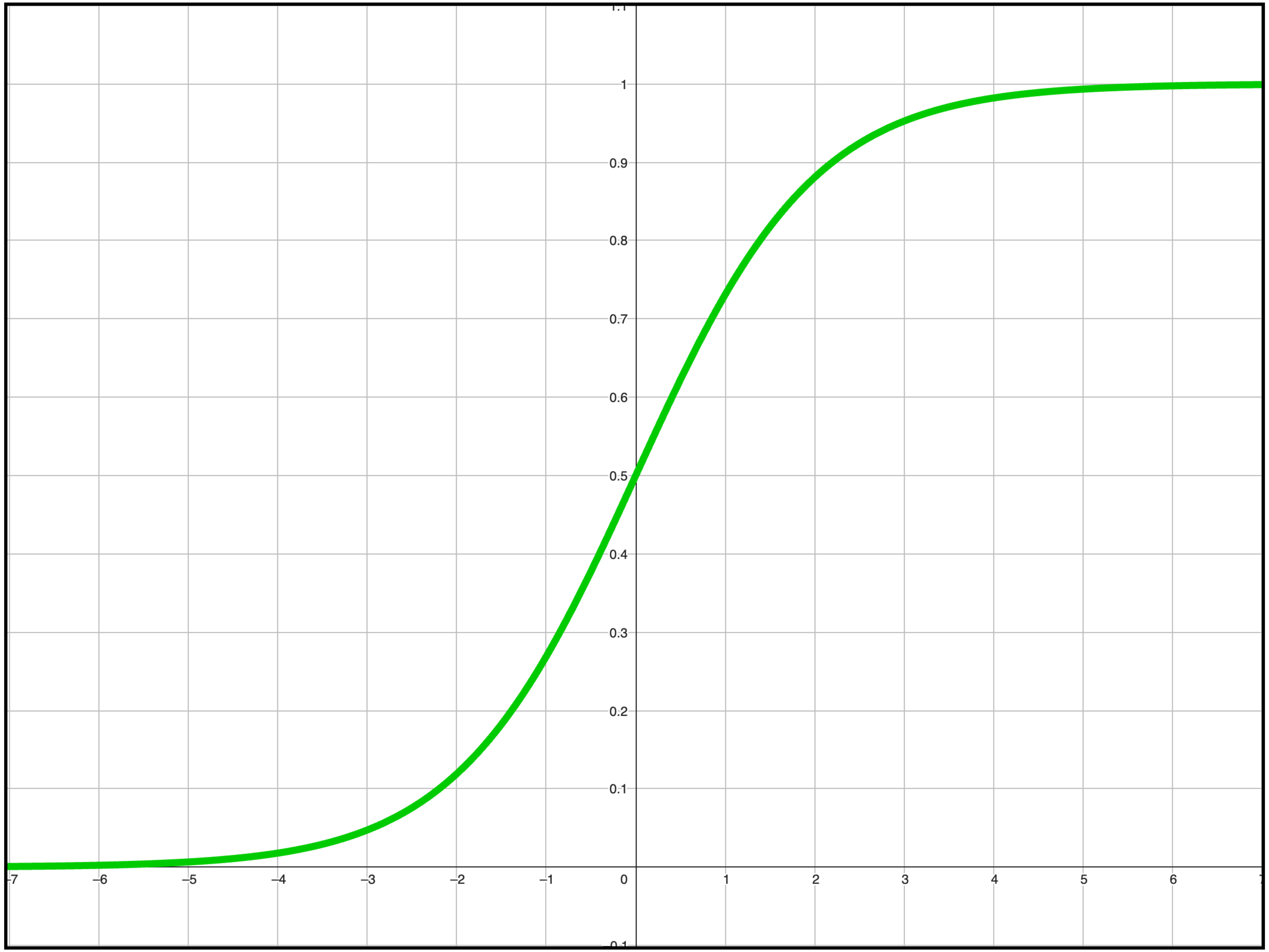
Multiplying x by 6 makes the curve even steeper

It still intersects the y axis at $y = 0.5$

Logistic Function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

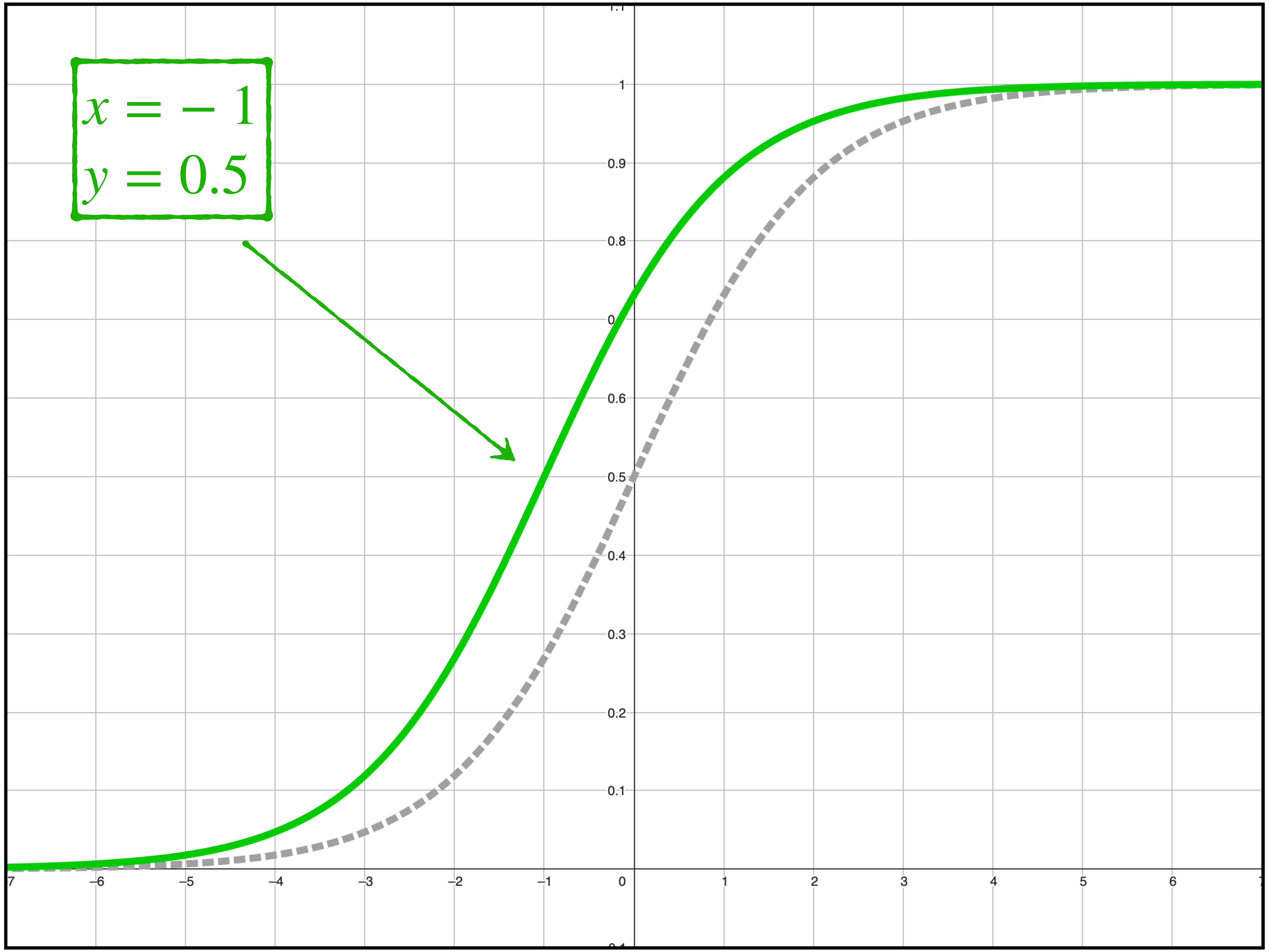
Adding a term shifts the curve to the left



Logistic Function

$$\sigma(x+1) = \frac{1}{1+e^{-(x+1)}}$$

Adding a term shifts the curve to the left



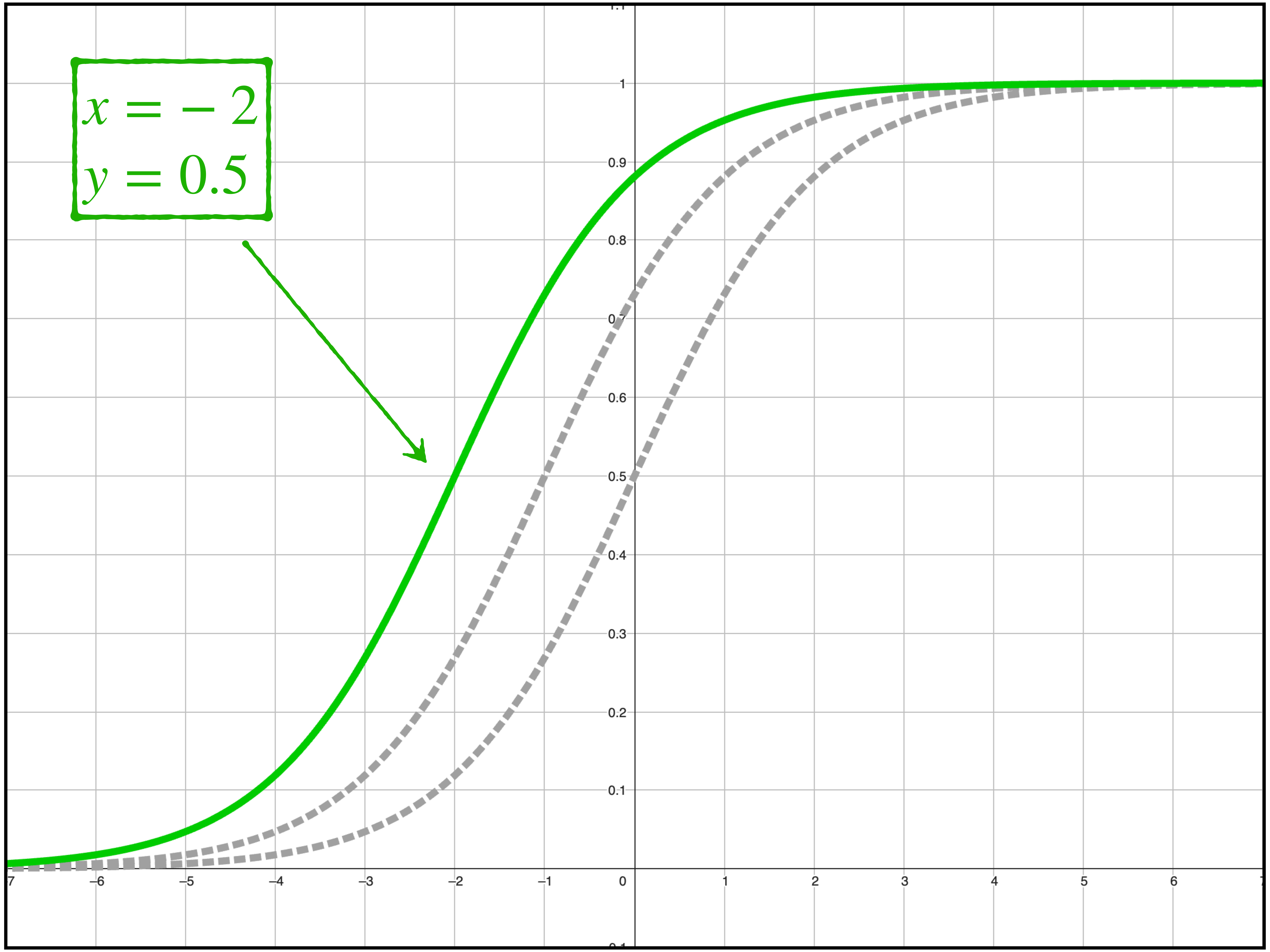
The curve intersects the y axis at $y = 0.5$ when $x = -1$

Logistic Function

$$\sigma(x+2) = \frac{1}{1+e^{-(x+2)}}$$

Adding a term shifts the curve to the left

The curve intersects the y axis at $y = 0.5$ when $x = -2$

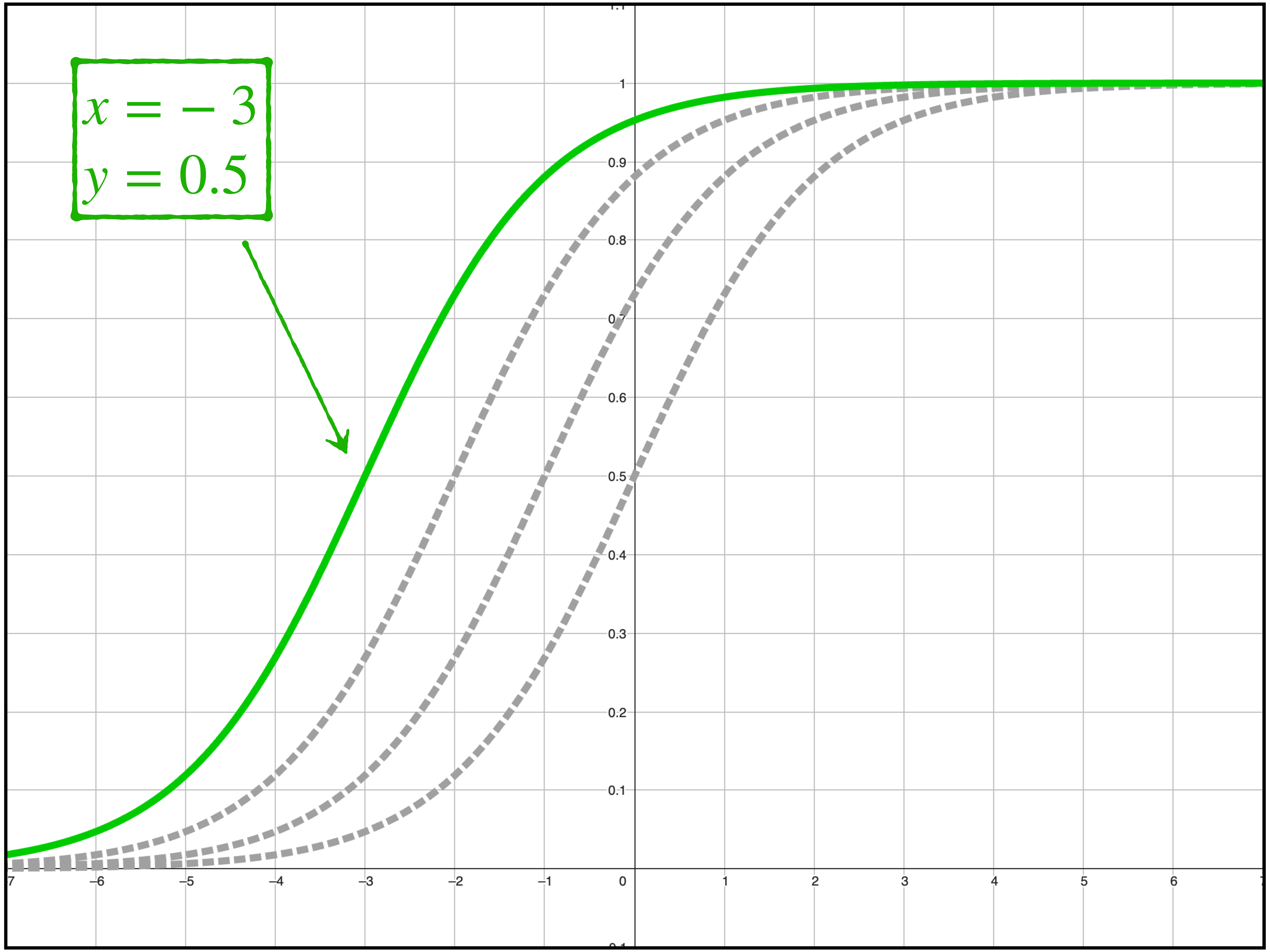


Logistic Function

$$\sigma(x+3) = \frac{1}{1+e^{-(x+3)}}$$

Adding a term shifts the curve to the left

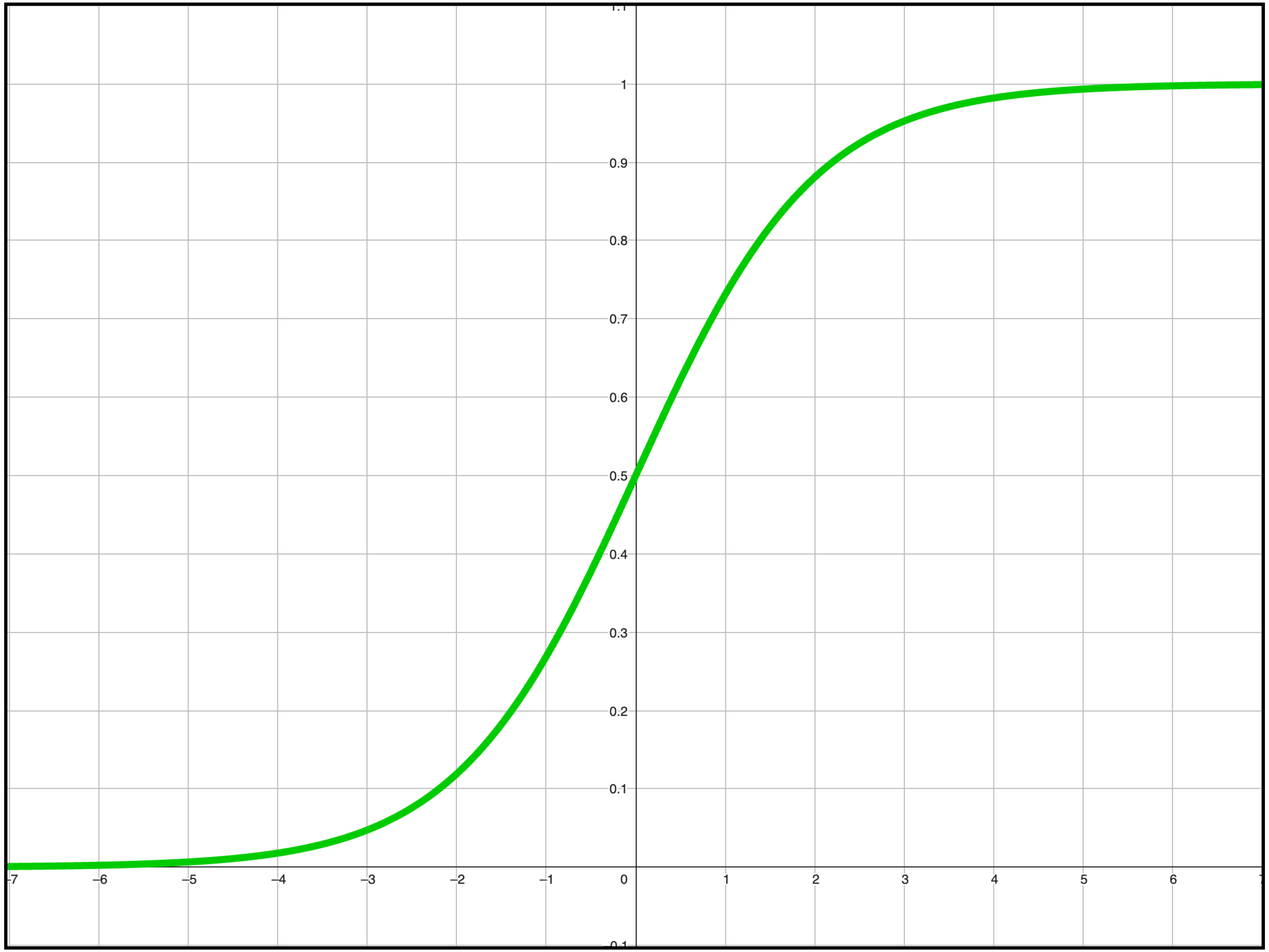
The curve intersects the y axis at $y = 0.5$ when $x = -3$



Logistic Function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Subtracting a term shifts the curve to the right

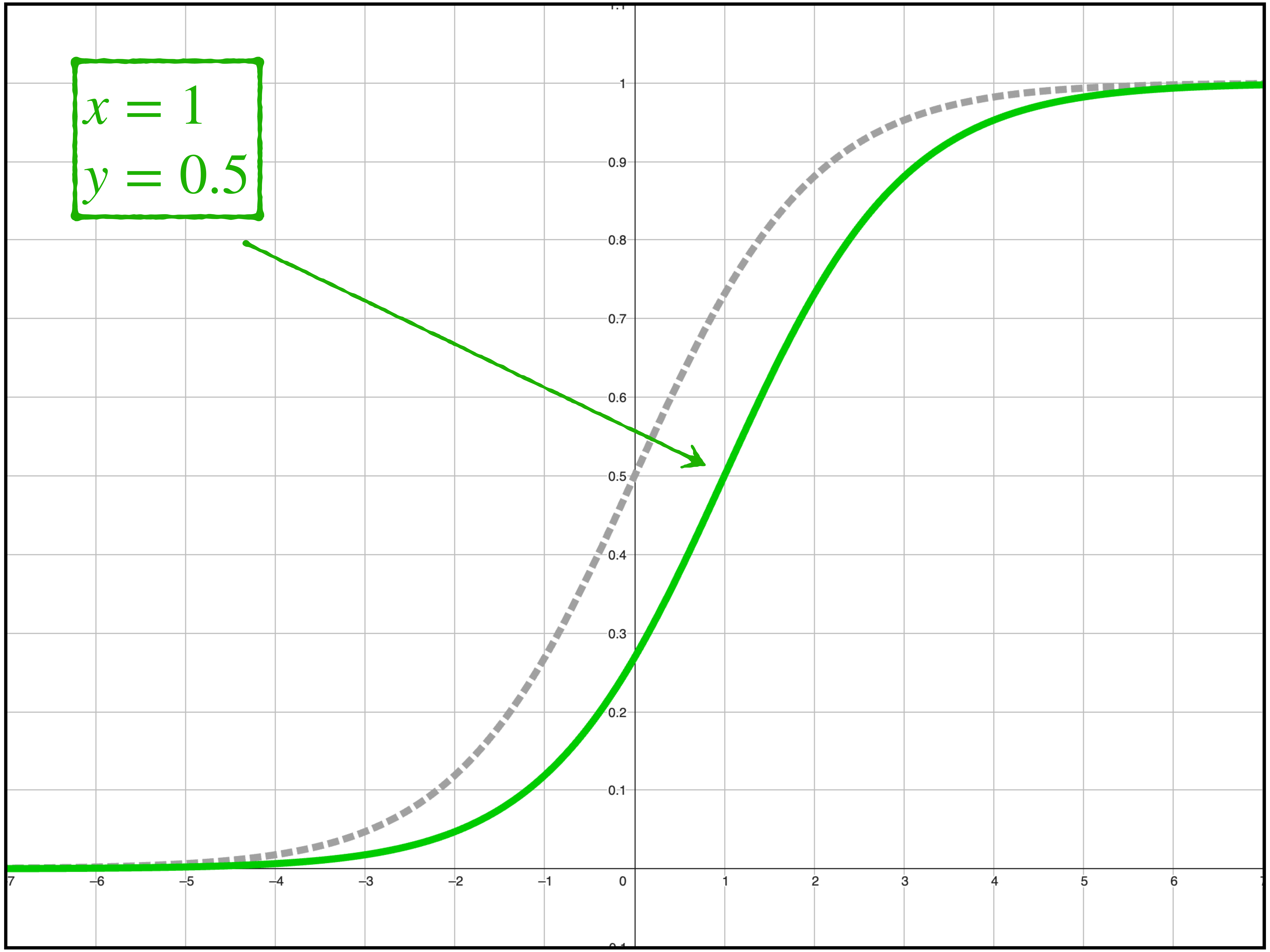


Logistic Function

$$\sigma(x-1) = \frac{1}{1+e^{-(x-1)}}$$

Subtracting a term shifts the curve to the right

The curve intersects the y axis at $y = 0.5$ when $x = 1$

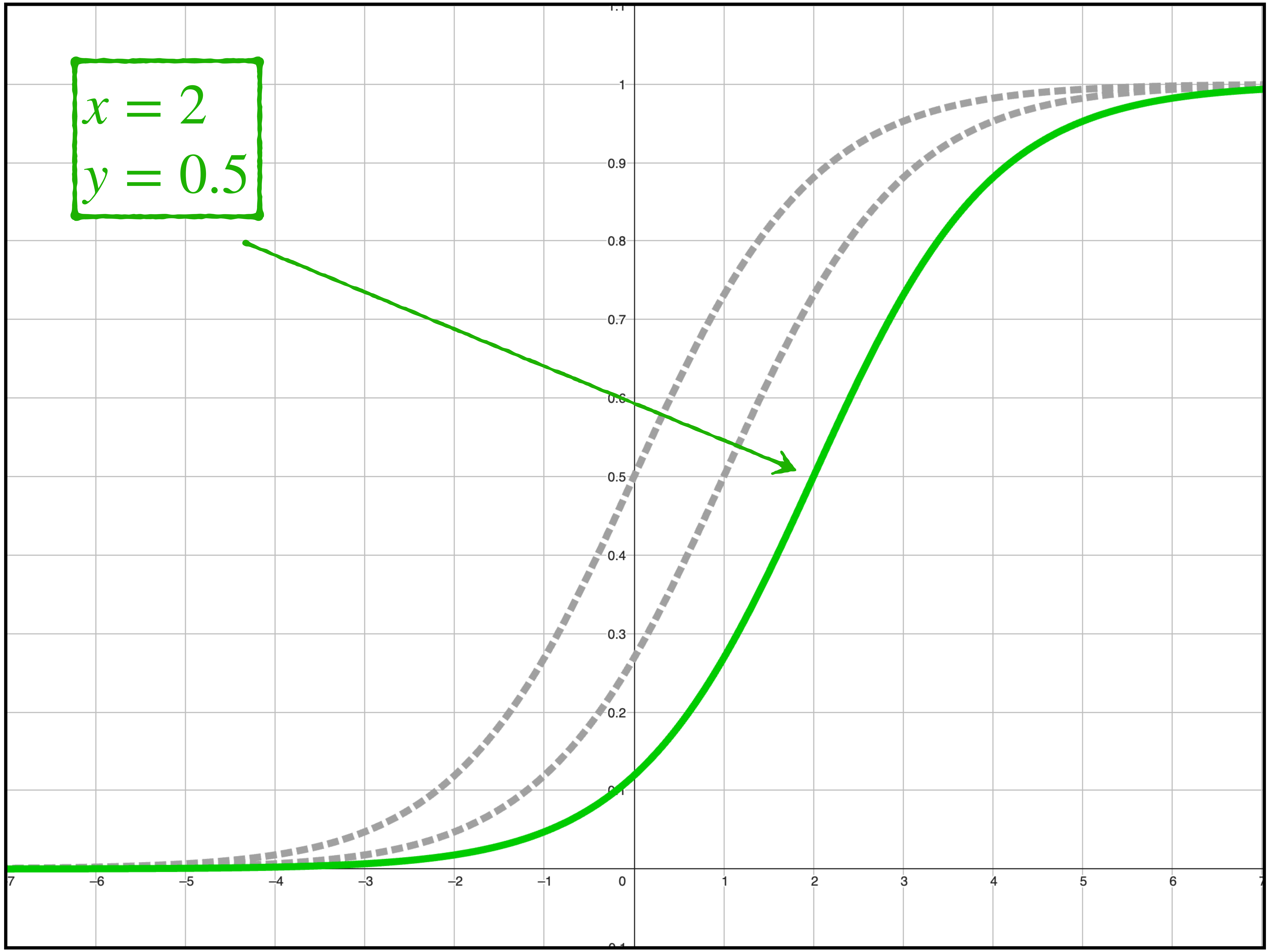


Logistic Function

$$\sigma(x-2) = \frac{1}{1+e^{-(x-2)}}$$

Subtracting a term shifts the curve to the right

The curve intersects the y axis at $y = 0.5$ when $x = 2$

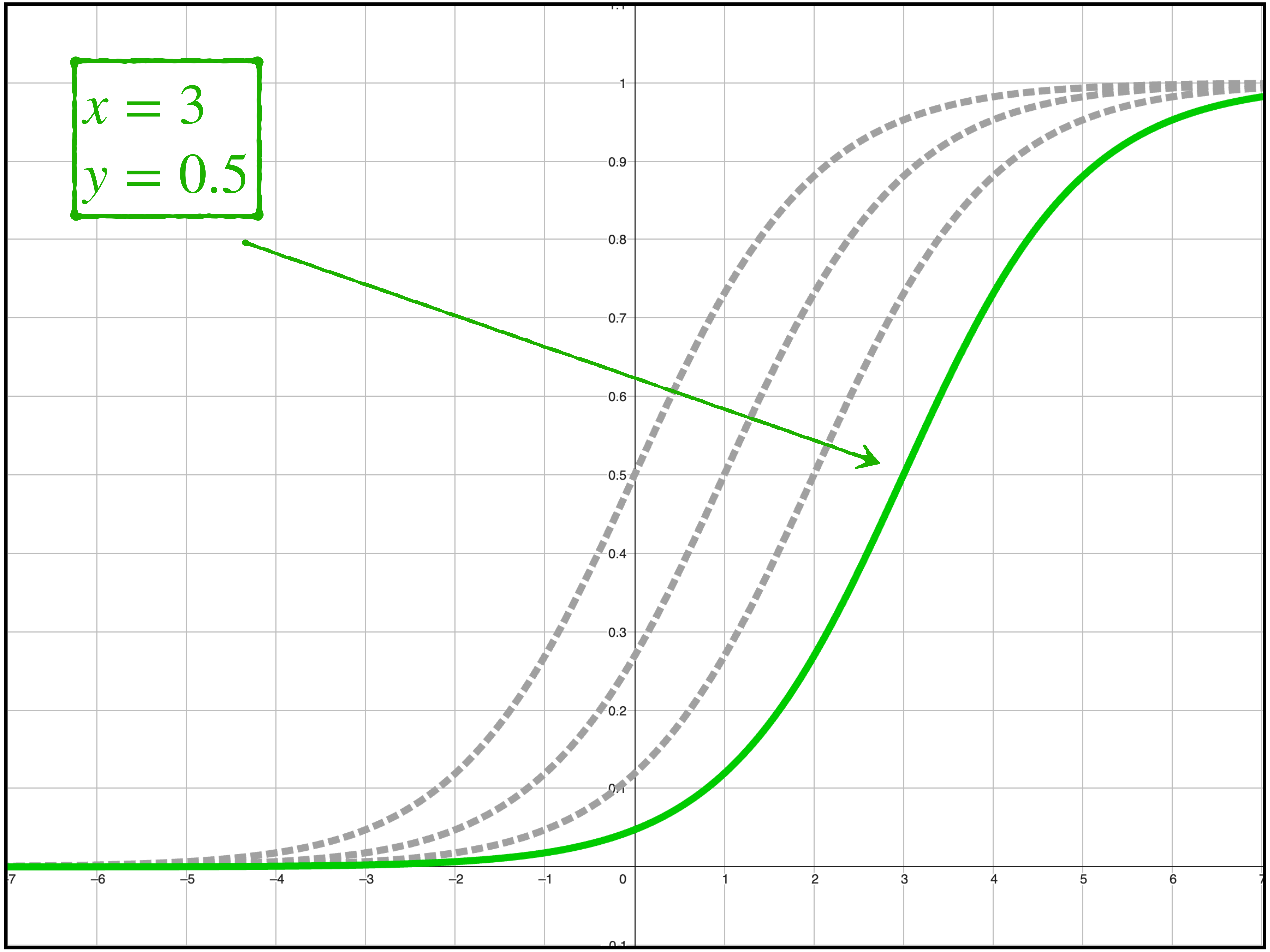


Logistic Function

$$\sigma(x-3) = \frac{1}{1+e^{-(x-3)}}$$

Subtracting a term shifts the curve to the right

The curve intersects the y axis at $y = 0.5$ when $x = 3$



Logistic Function

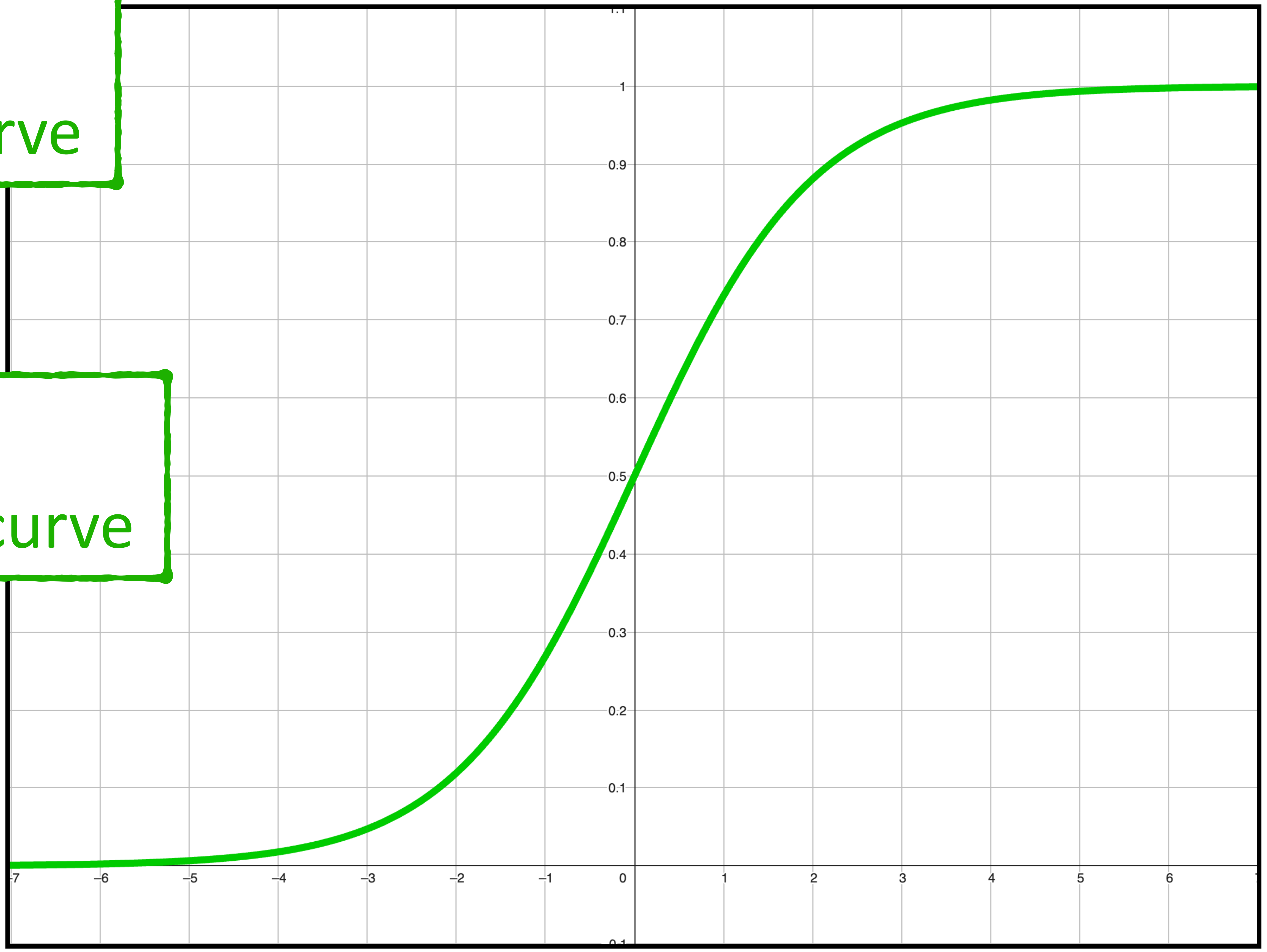
The shape and position of the sigmoid curve is dependent on two parameters:

$$y = \beta + \omega x$$

$$\sigma(y) = \frac{1}{1 + e^{-(\beta + \omega x)}}$$

β determines the position of the curve

ω determines the steepness of the curve



Logistic Function

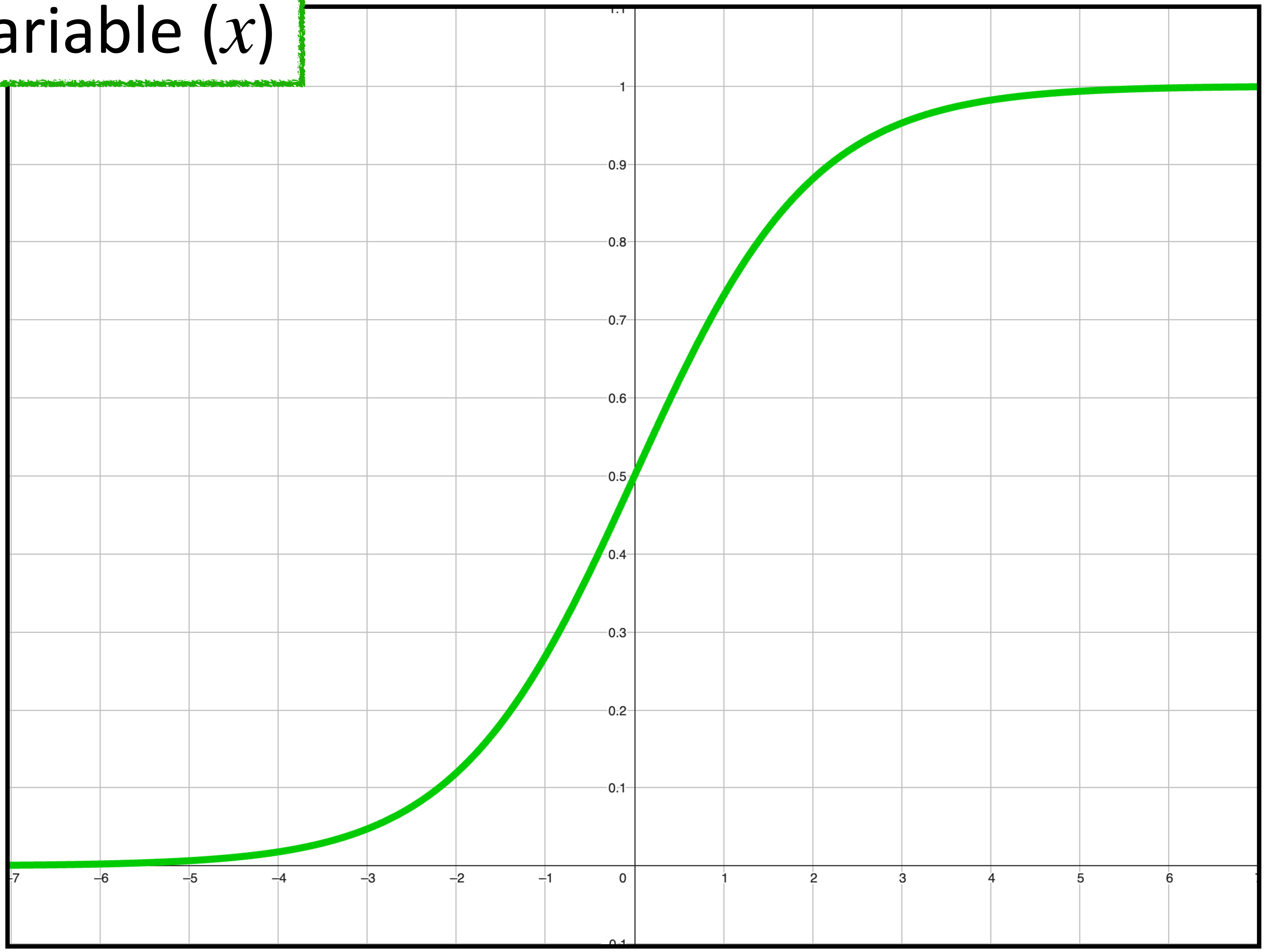
The shape and position of the sigmoid curve is dependent on two parameters:

$$y = \beta + \omega x$$

1 independent variable (x)

$$\sigma(y) = \frac{1}{1 + e^{-(\beta + \omega x)}}$$

1 dependent variable (y)



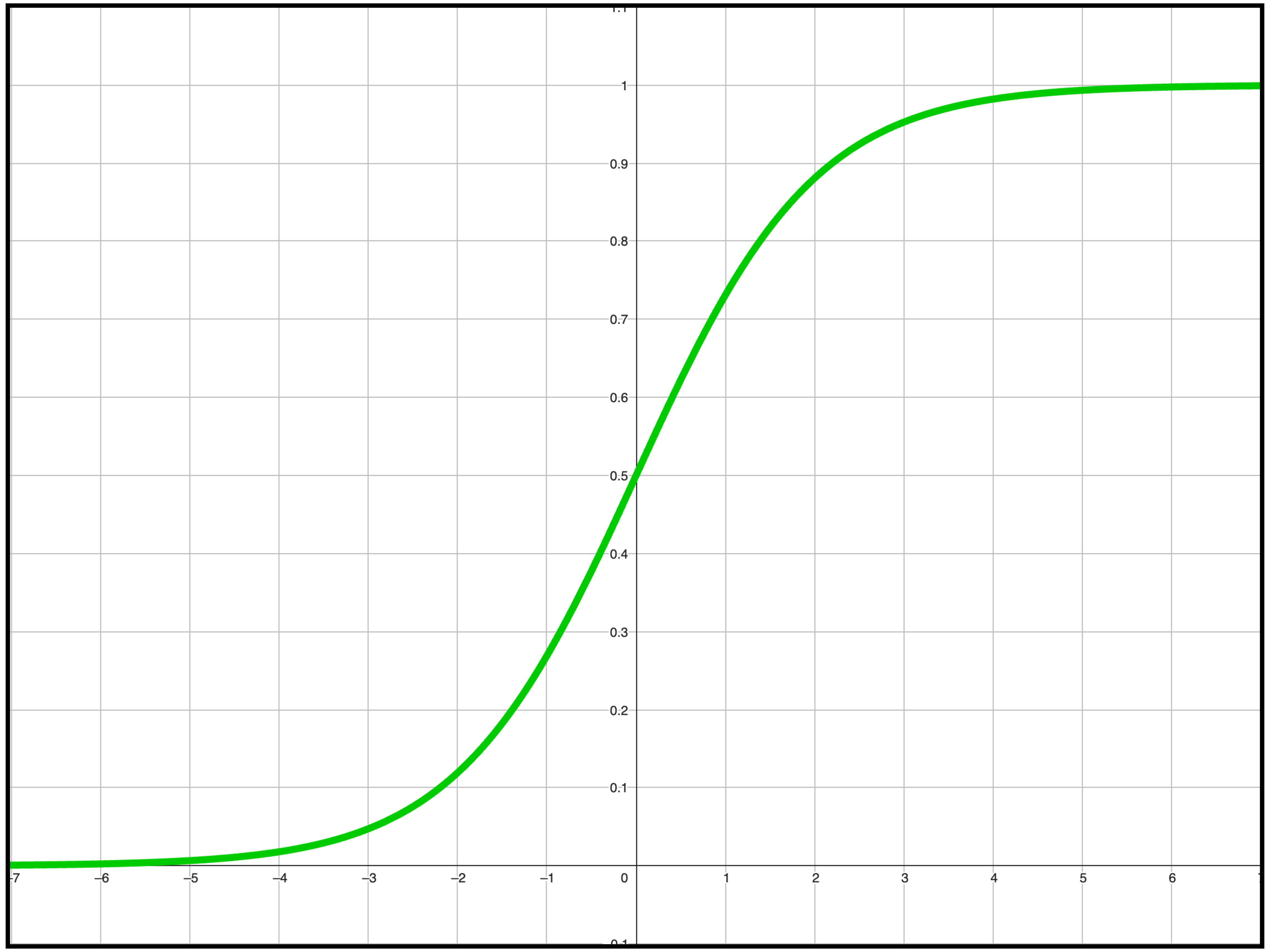
Logistic Function

The shape and position of the sigmoid curve is dependent on two parameters:

$$y = \beta + \omega x$$

$$\sigma(y) = \frac{1}{1 + e^{-(\beta + \omega x)}}$$

Let's generalize this to k independent variables



The Logistic function converts the linear combination of input features (x values) into probabilities (y values)

Logistic Function

The model is generalized to k independent variables and $k + 1$ parameters

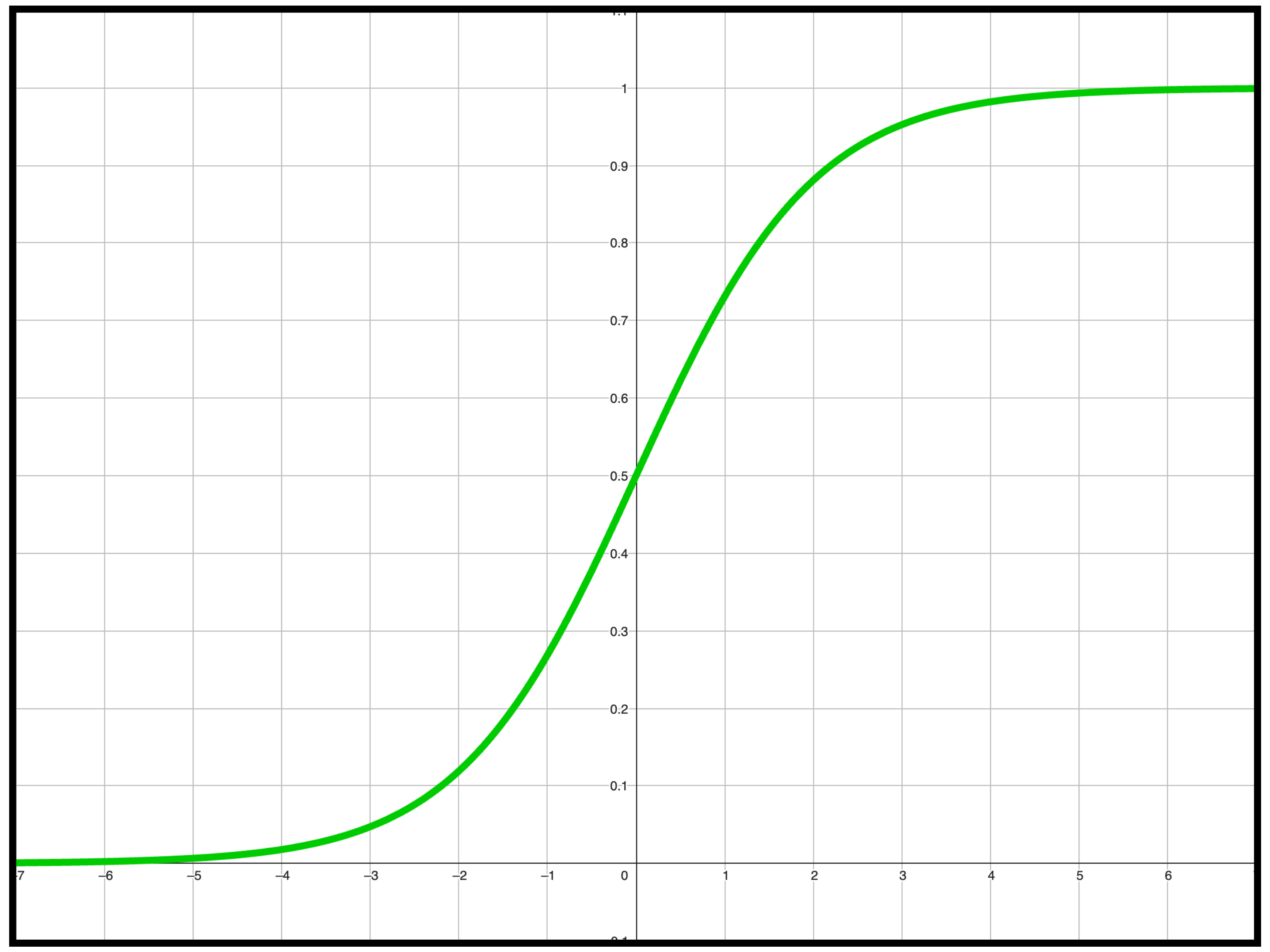
$$y = \beta + \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \ldots + \omega_k x_k$$

1 dependent variable y

k independent variables $x_1 \ldots x_k$

$$\sigma(y) = \frac{1}{1 + e^{-(\beta + \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \ldots + \omega_k x_k)}}$$

This model has
$k + 1$ parameters
$\beta, \omega_1, \omega_2 \ldots \omega_k$



Logistic Function

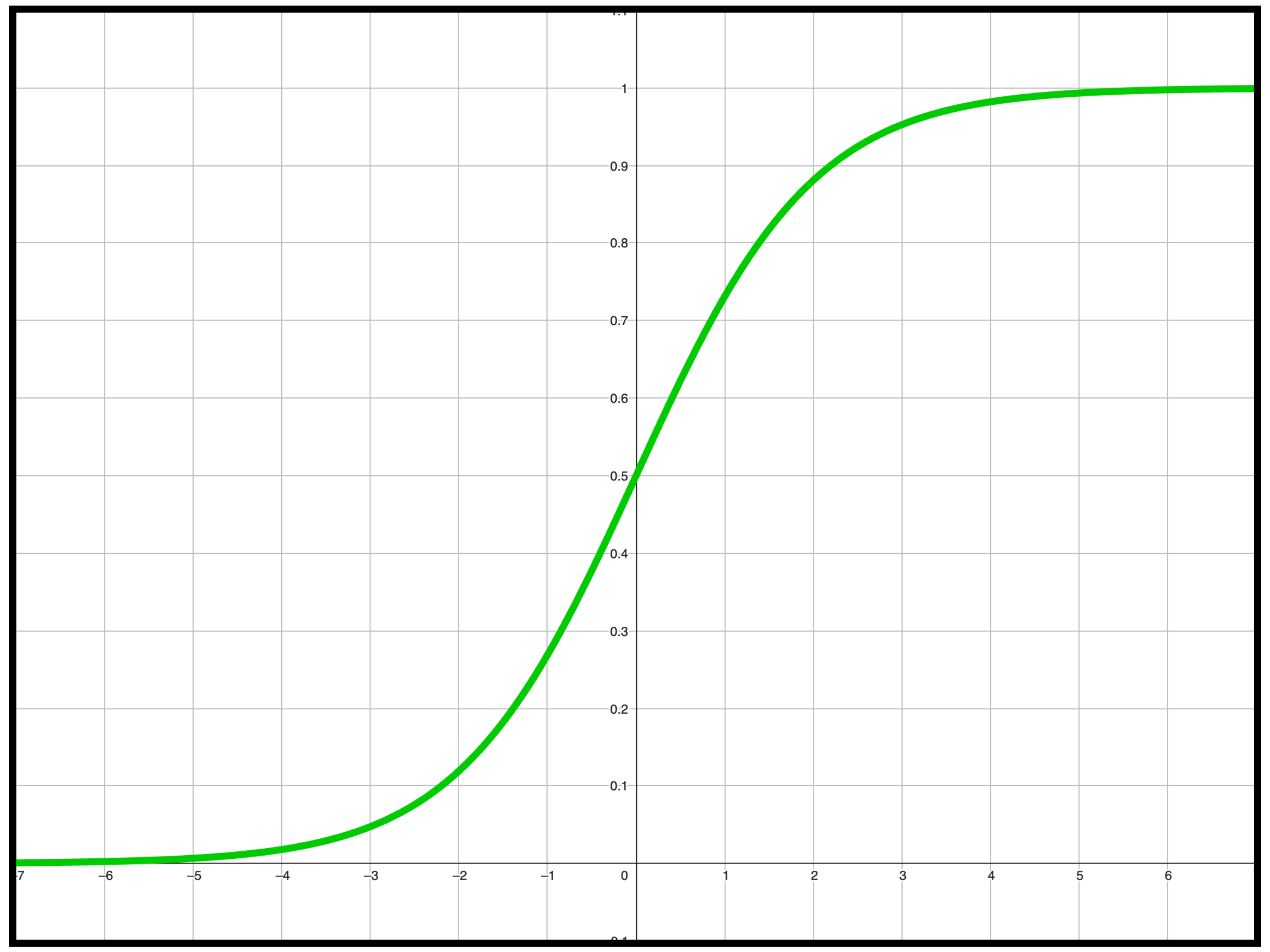
The model is generalized to k independent variables and $k+1$ parameters

$$y = \beta + \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \ldots + \omega_k x_k$$

General Matrix form:

W is a $k \times 1$ vector of weights $\omega_1, \omega_2, \omega_3 \ldots \omega_k$
X is a $k \times n$ matrix of observations $x_1, x_2, x_3 \ldots x_k$
β is a scalar

$$\hat{Y} = \sigma(W^T X + \beta)$$



Logistic Function

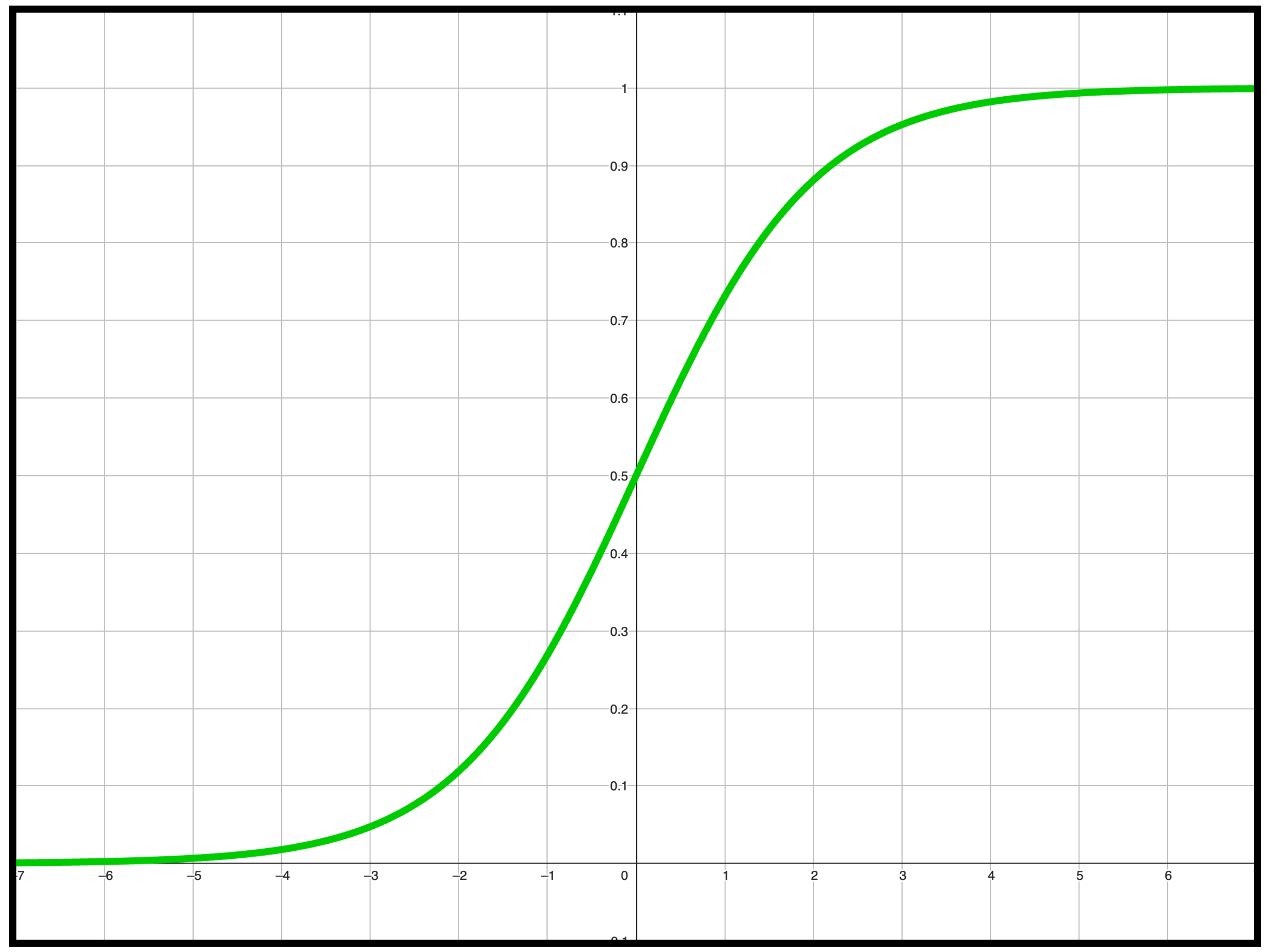
The model is generalized to k independent variables and $k+1$ parameters

$$y = \beta + \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \ldots + \omega_k x_k$$

General Matrix form:

$$\hat{Y} = \sigma(W^T X + \beta)$$

$\hat{Y}$ is the vector of predicted values from the model.
$\hat{Y}$ is a vector of probabilities each between 0 and 1



Logistic Function

We can convert the probabilities ($\hat{Y}$) into binary values (true / false) using a threshold.

$$y = \beta + \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \ldots + \omega_k x_k$$

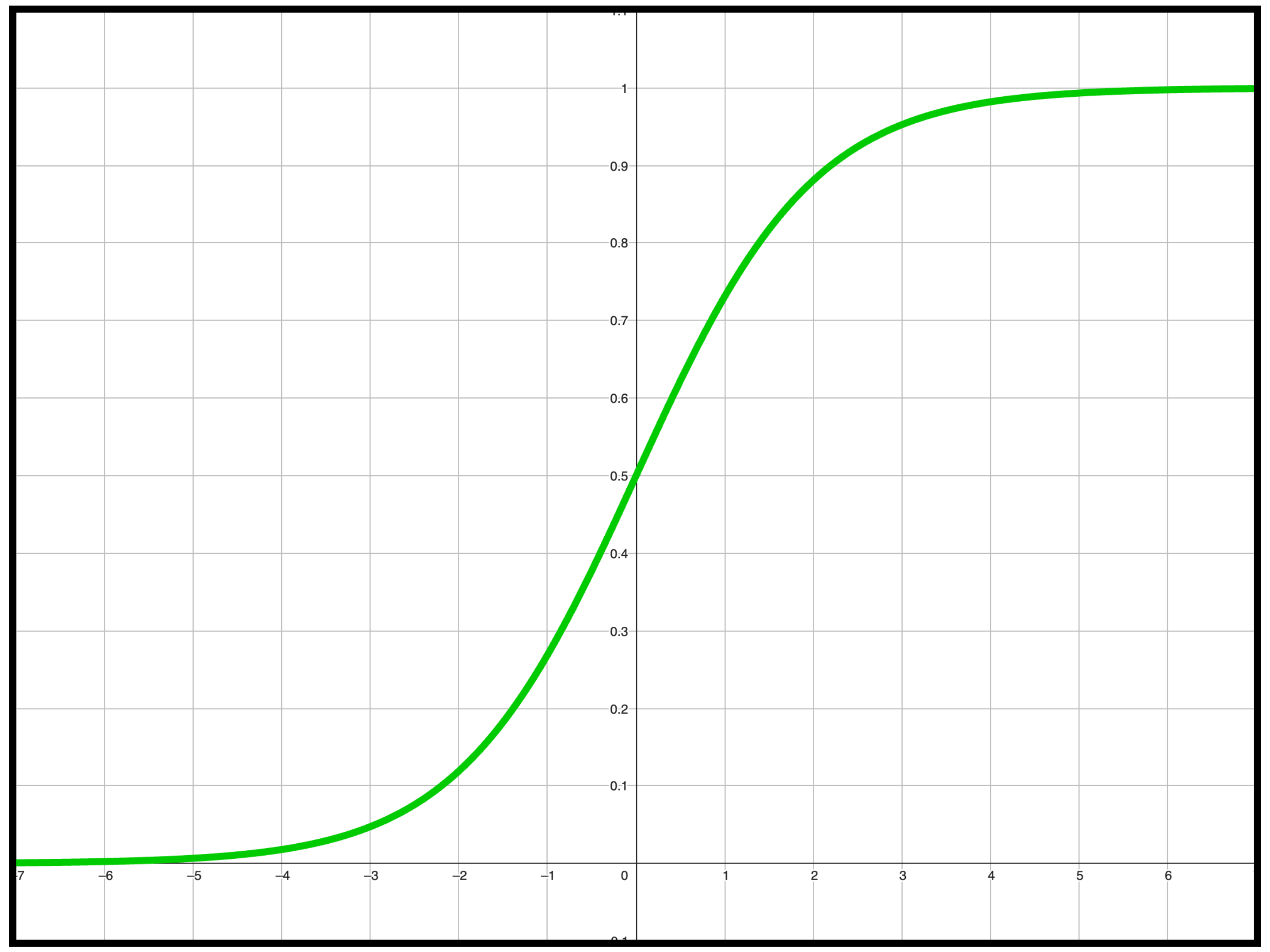
General Matrix form:

$$\hat{Y} = \sigma(W^T X + \beta)$$

$\hat{Y}$ is the vector of predicted values from the model.
$\hat{Y}$ is a vector of probabilities each between 0 and 1

A Threshold converts a given probability to a binary value

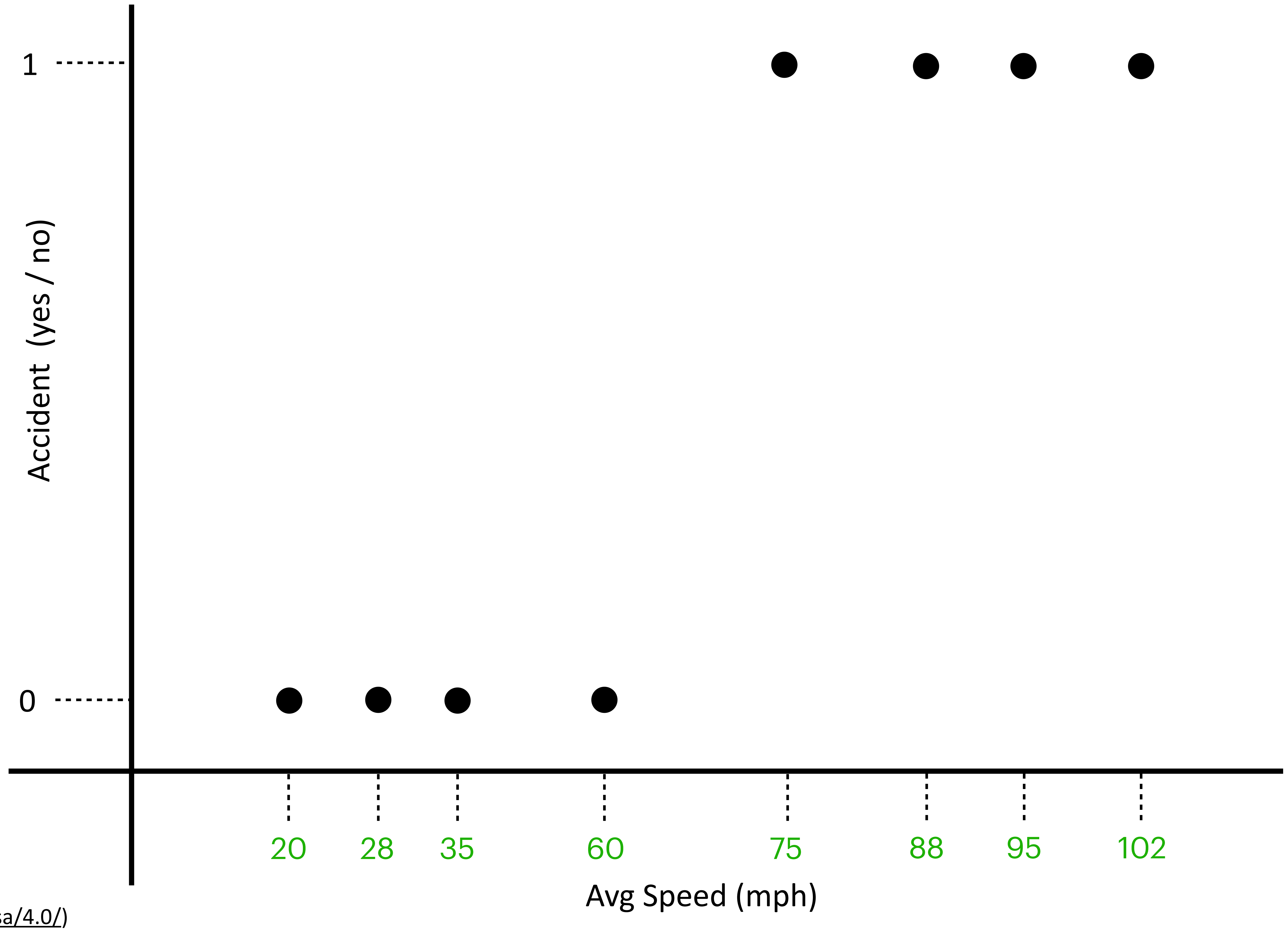
if $\hat{y}_i \geq 0.5$ then 1
if $\hat{y}_i < 0.5$ then 0



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

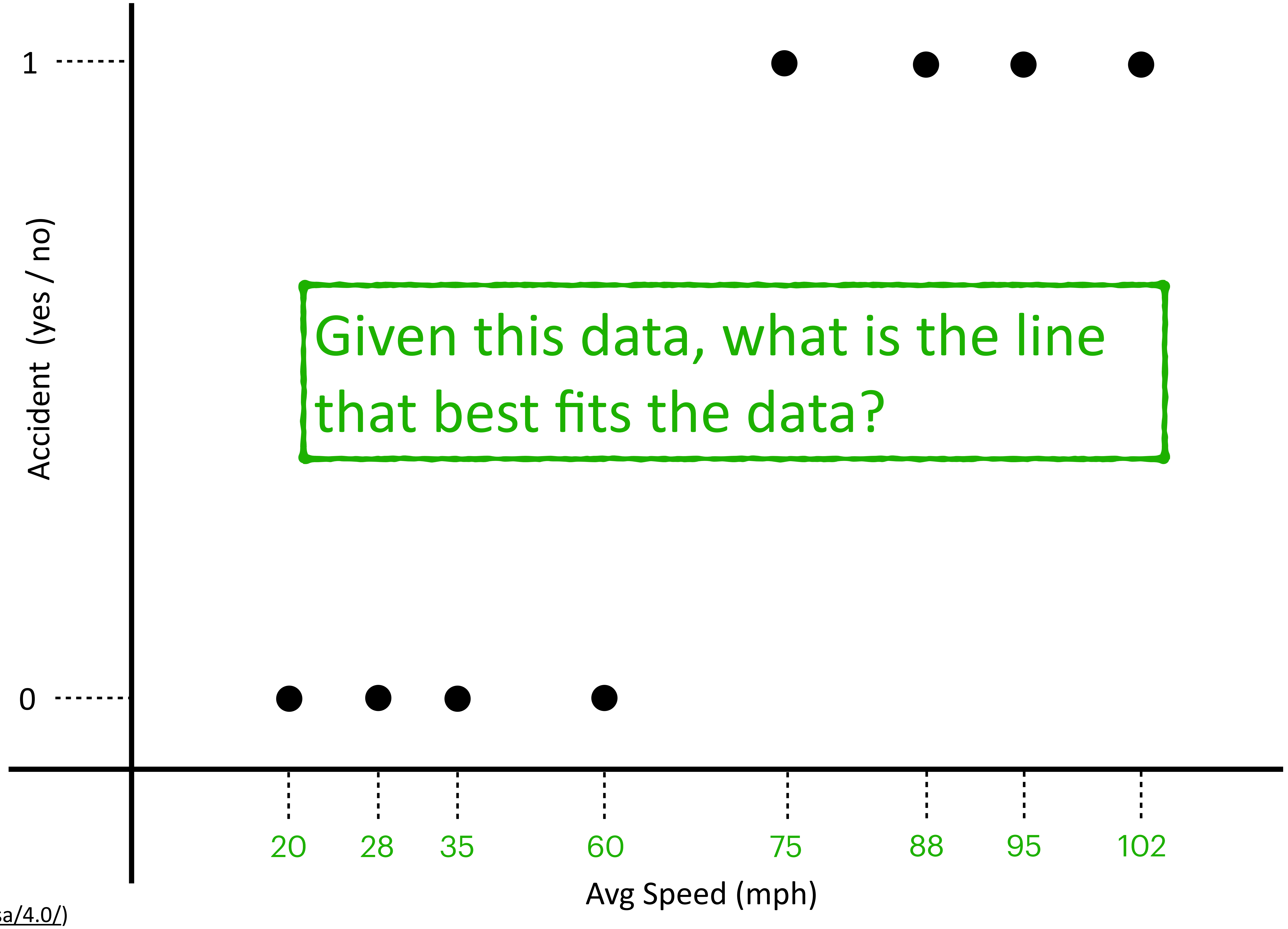
Lets apply this to our original problem



A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

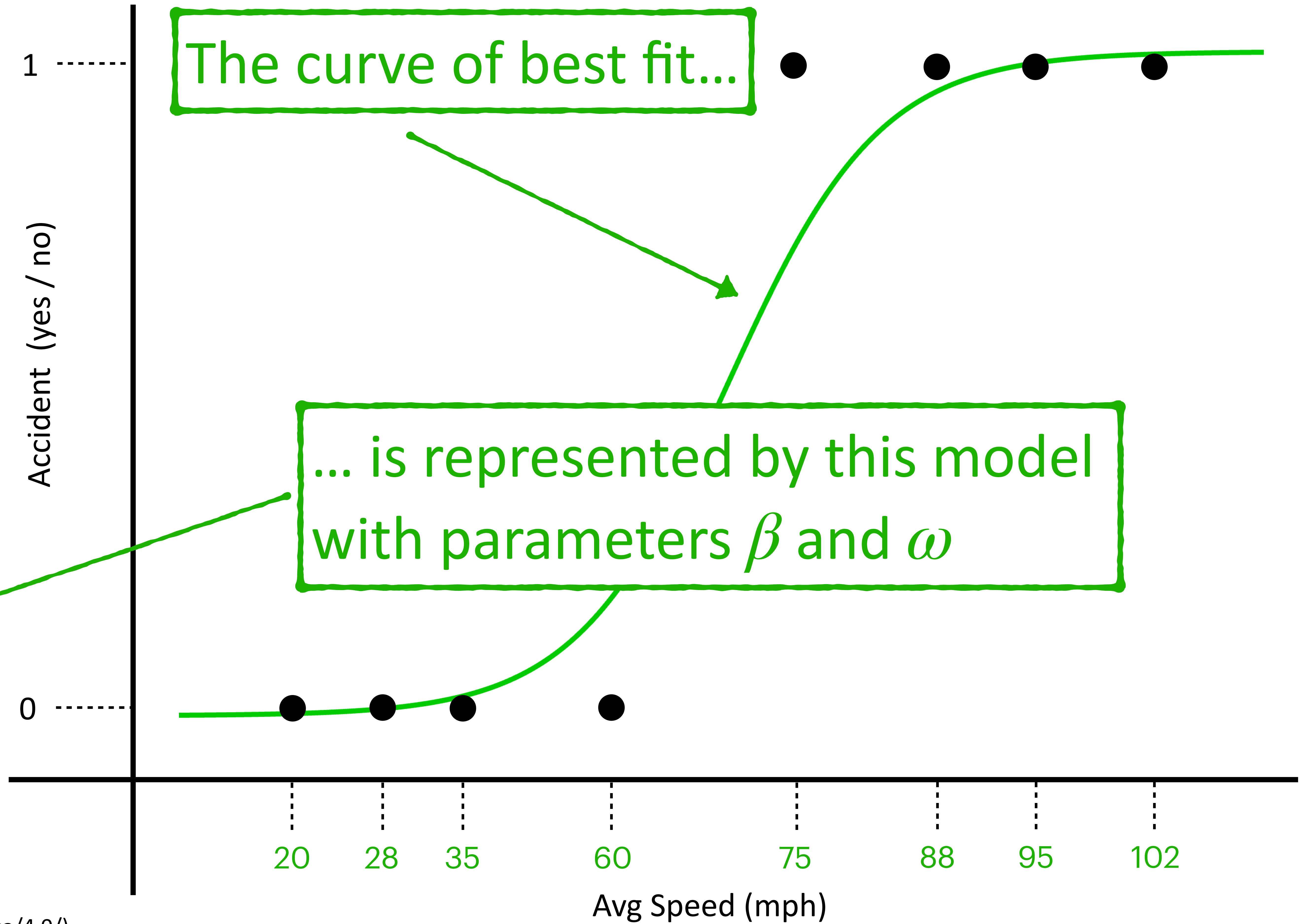


A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

$$\hat{y} = \frac{1}{1 + e^{-(\beta + \omega x)}}$$



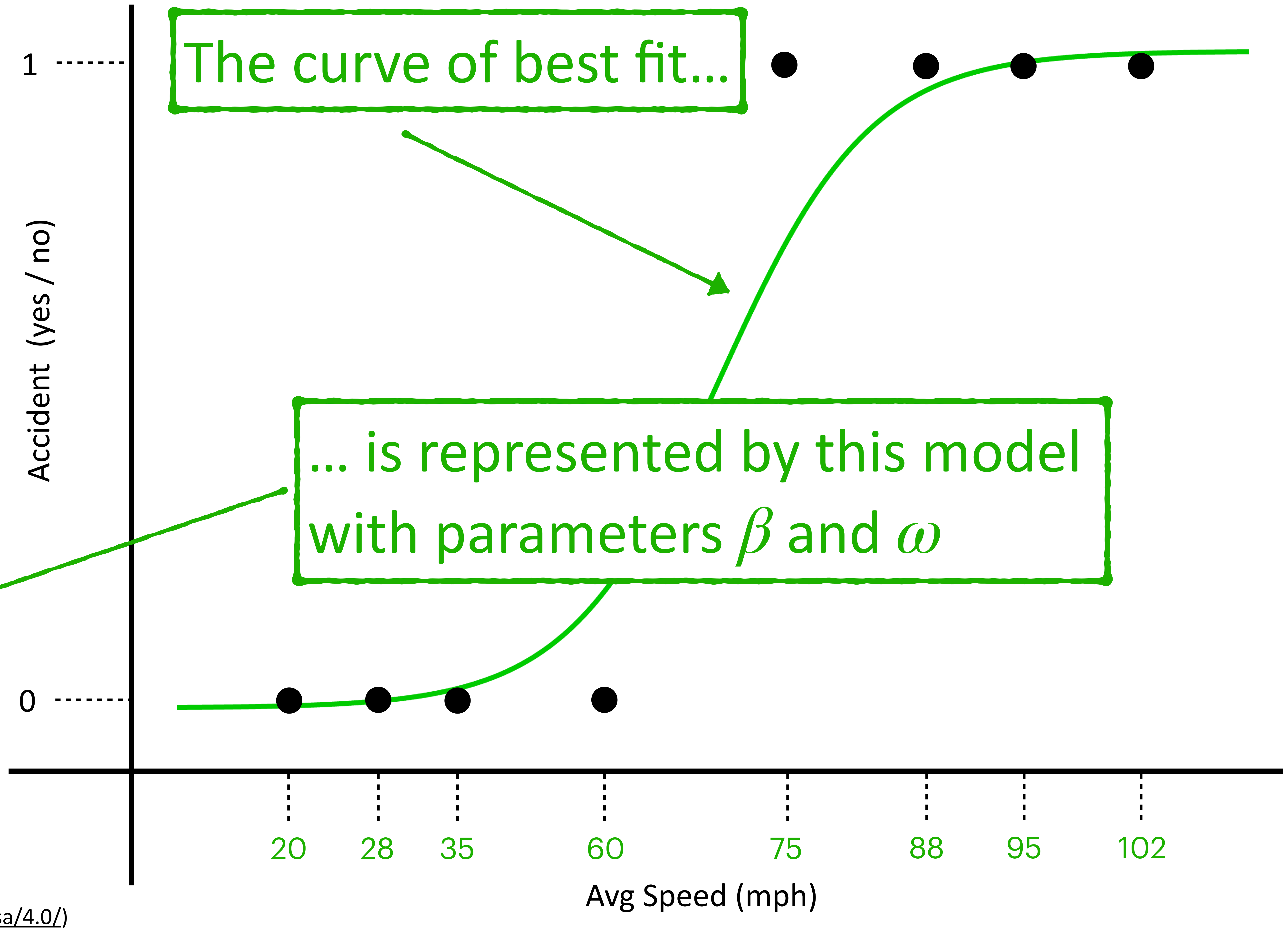
A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

$$\hat{y} = \frac{1}{1 + e^{-(\beta + \omega x)}}$$

x represents the Avg Speed (mph)

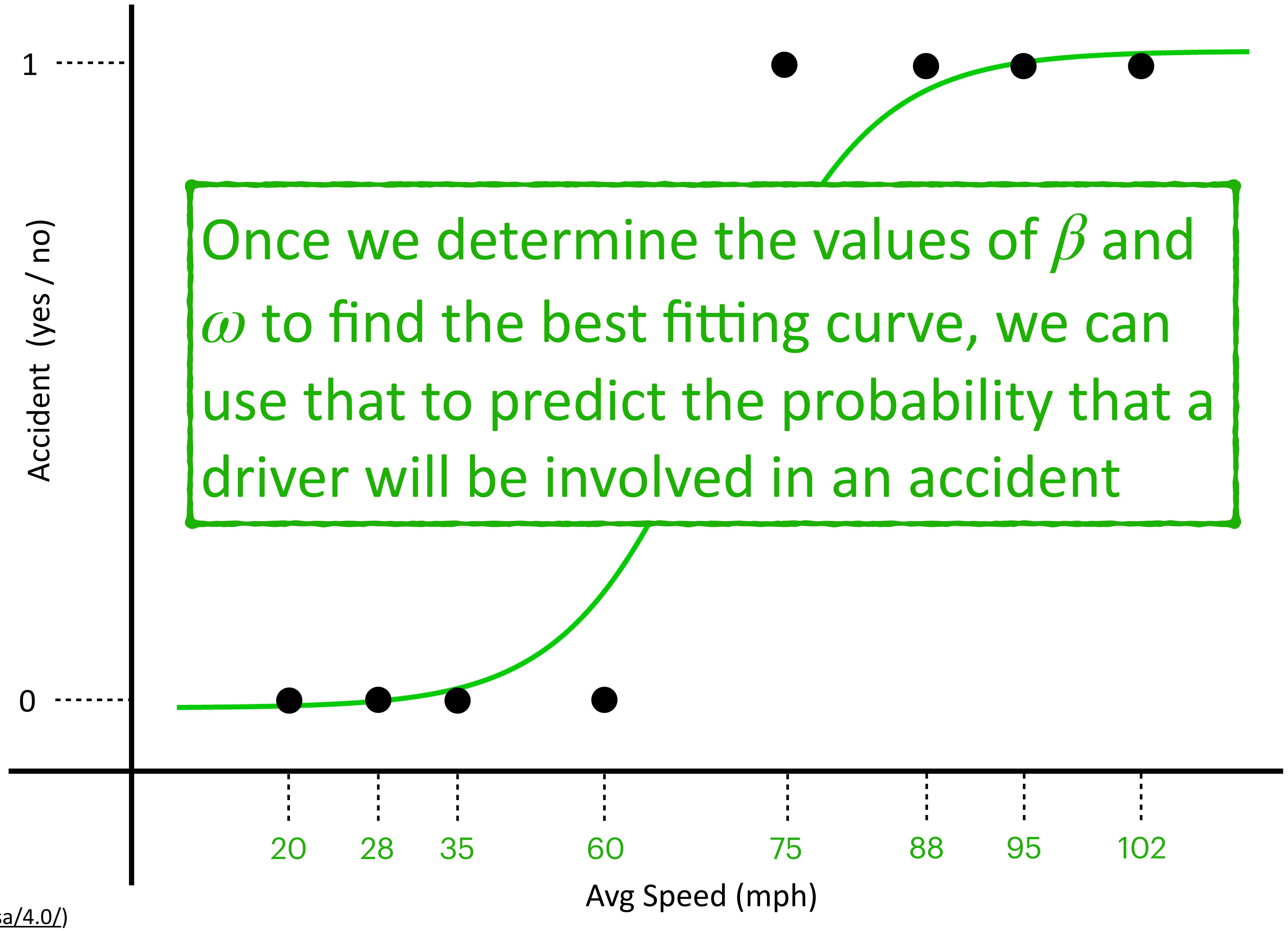


A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

$$\hat{y} = \frac{1}{1 + e^{-(\beta + \omega x)}}$$



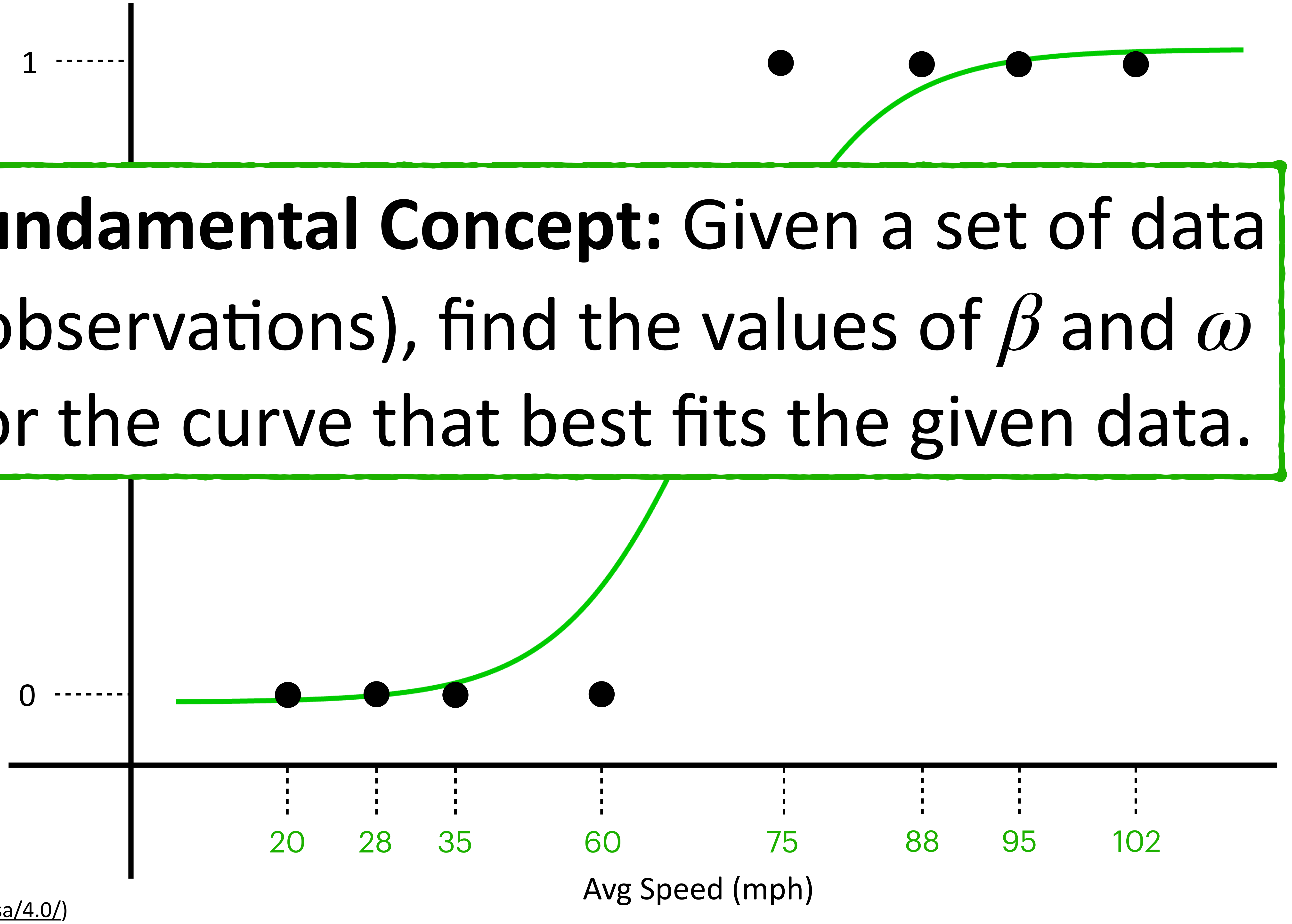
Once we determine the values of β and ω to find the best fitting curve, we can use that to predict the probability that a driver will be involved in an accident

A simple example…

Drivers that speed, tend to get into more accidents.
An insurance company has data for average speed for drivers and whether they were involved in an accident.

Avg Speed (mph)	Accident
20	No
28	No
35	No
60	No
75	Yes
88	Yes
95	Yes
102	Yes

$$\hat{y} = \frac{1}{1 + e^{-(\beta + \omega x)}}$$



Fundamental Concept: Given a set of data (observations), find the values of β and ω for the curve that best fits the given data.

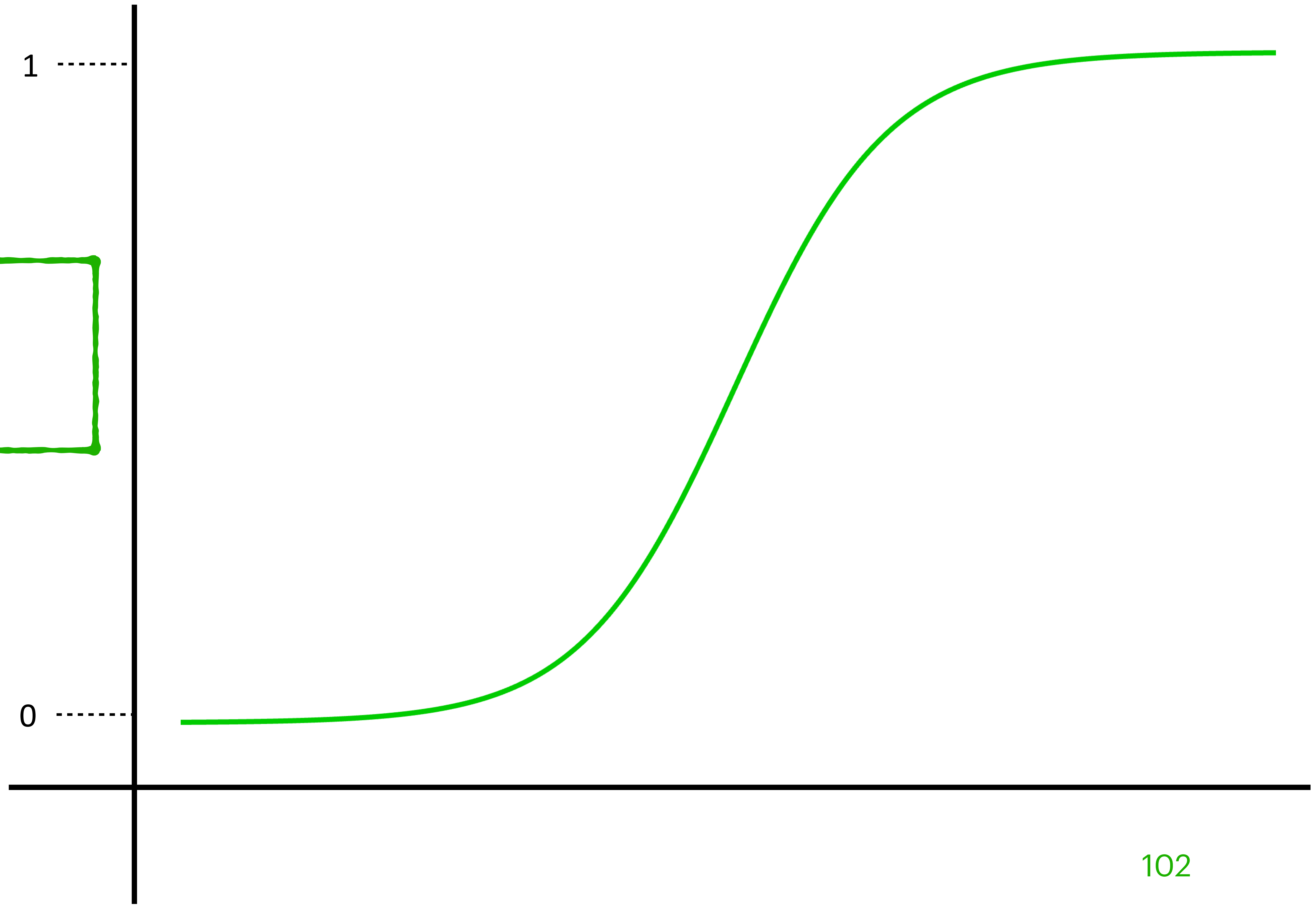
Generalizing to k independent variables and $k+1$ parameters

Curve of best fit is…

$$\hat{Y} = \sigma(W^T X + \beta)$$

The Problem Statement:

Logistic Regression: Find the values of β and W for the curve of best fit.

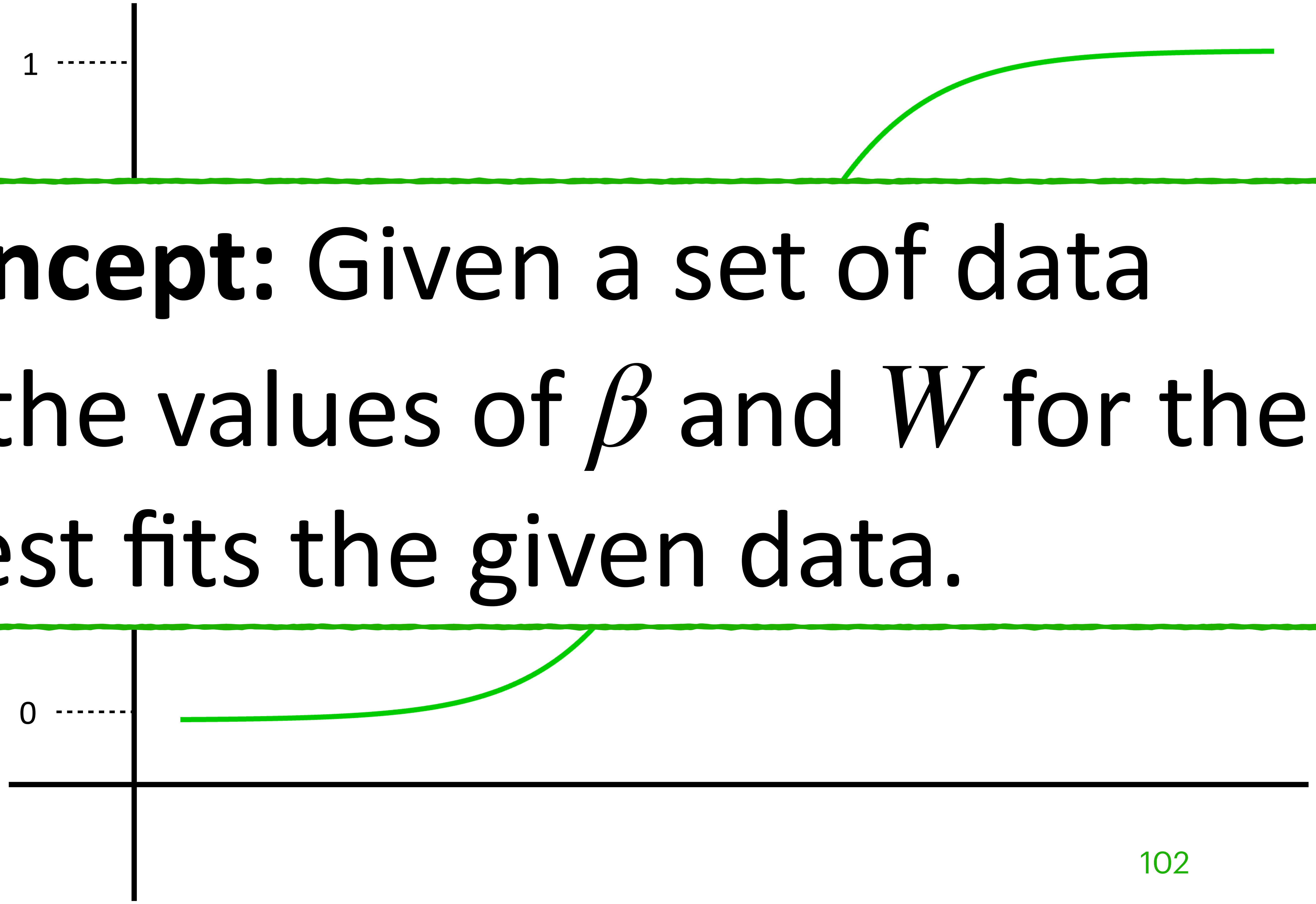


Logistic Regression

Generalizing to k independent variables and $k+1$ parameters

Curve of best fit is…

$$\hat{Y} = \sigma(W^T X + \beta)$$



Fundamental Concept: Given a set of data (observations), find the values of β and W for the curve that best fits the given data.

102

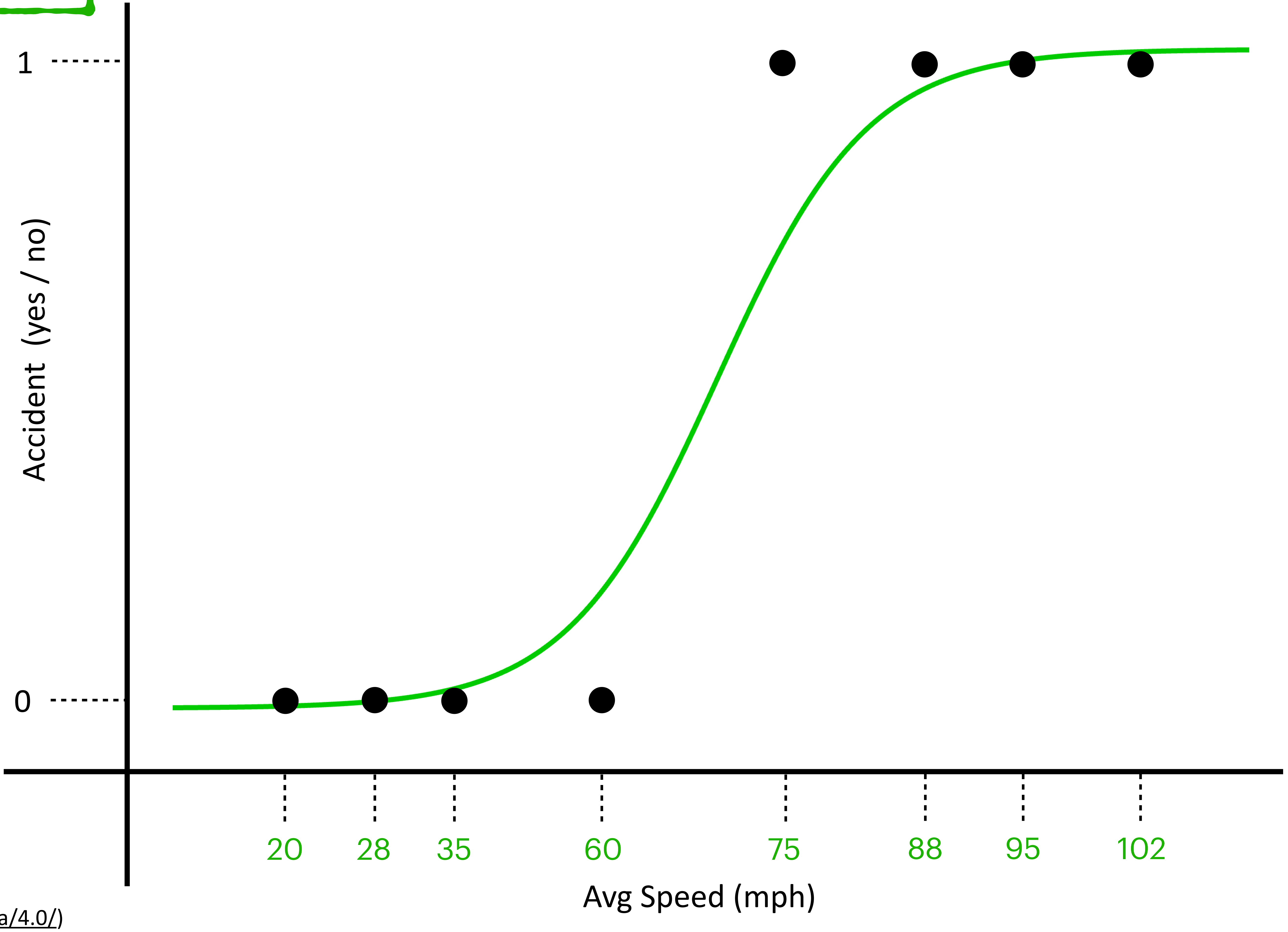
The Problem Statement:

Logistic Regression: Find the values of β and W for the curve of best fit.

Curve of best fit is…

$$\hat{Y} = \sigma(W^T X + \beta)$$

We can use **Gradient Descent** to find the optimal values of β and W



Related Tutorials & Textbooks

Multiple Regression

Multiple regression extends the two dimensional linear model introduced in Simple Linear Regression to k + 1 dimensions with one dependent variable, k independent variables and k+1 parameters.

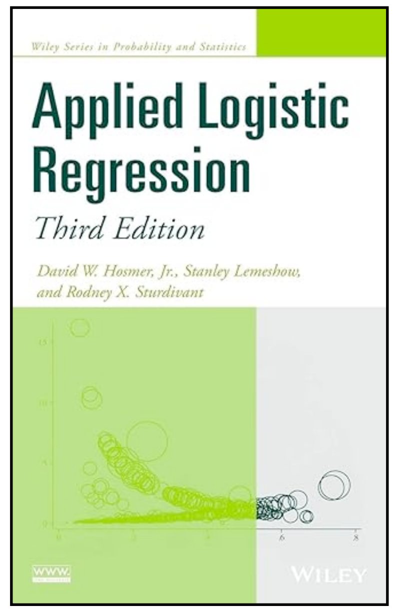
Gradient Descent for Simple Linear Regression

Gradient Descent algorithm for multiple regression and how it can be used to optimize k + 1 parameters for a Linear model in multiple dimensions.

Cost Function & Gradient Descent for Logistic Regression

An introduction to the Cost function for Logistic Regression long with its partial derivative (the gradient vector). The model parameters (B & W) are then optimized using Maximum Likelihood Estimation and Gradient Descent.

Recommended Textbooks



Applied Logistic Regression

by David W. Hosmer Jr., Stanley Lemeshow, Rodney X. Sturdivant

For a complete list of tutorials see:
https://arrsingh.com/ai-tutorials