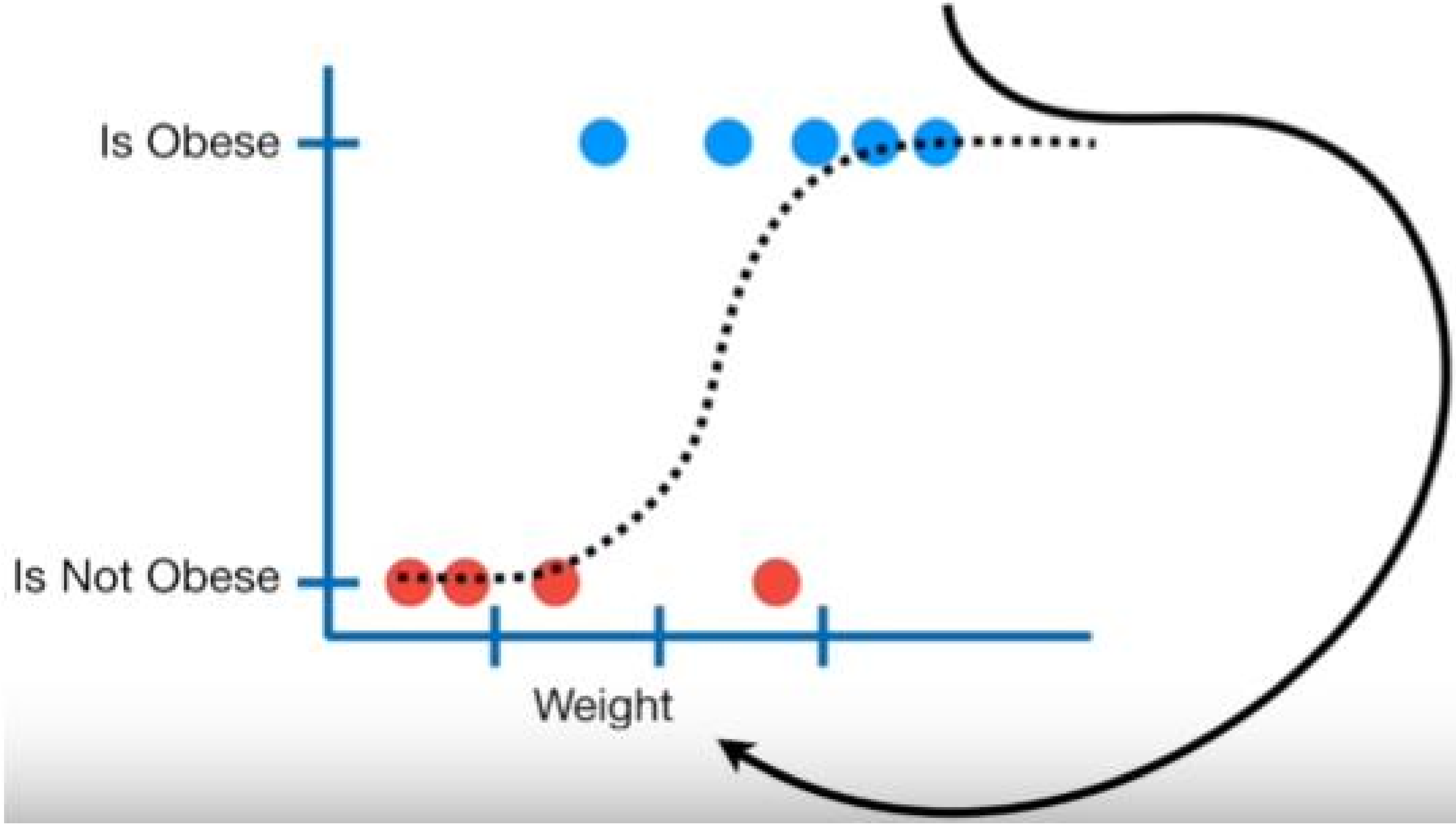
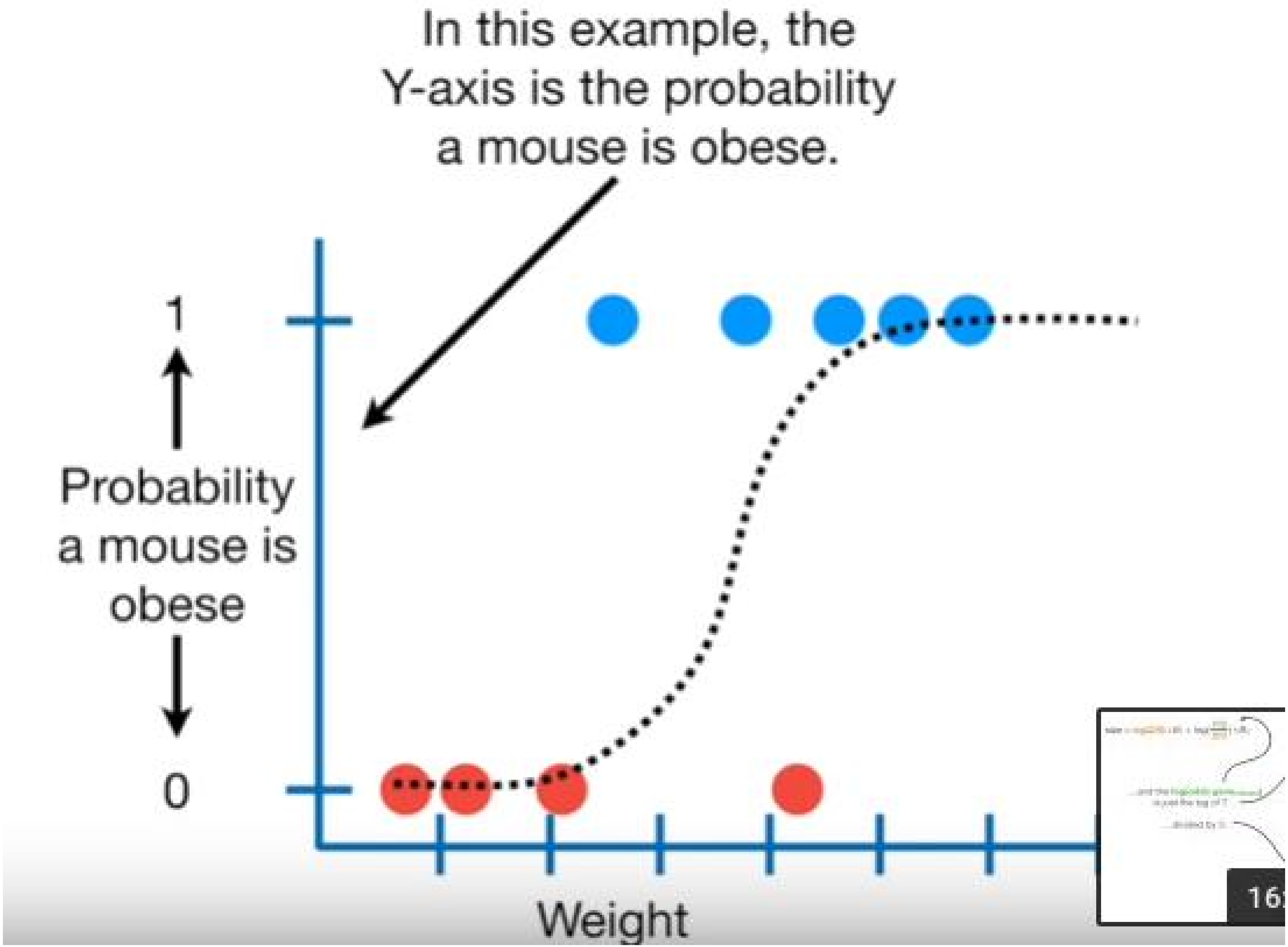
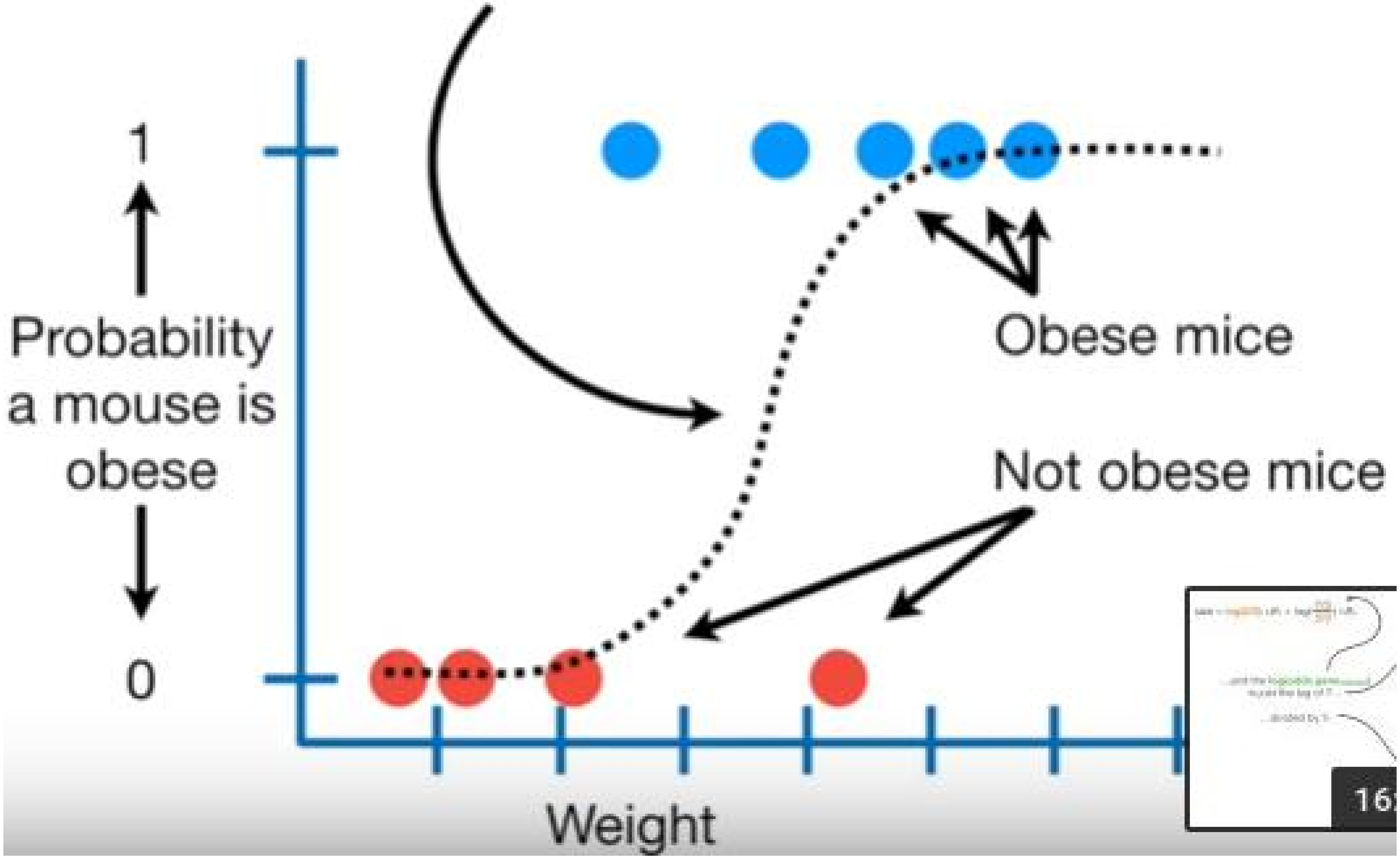
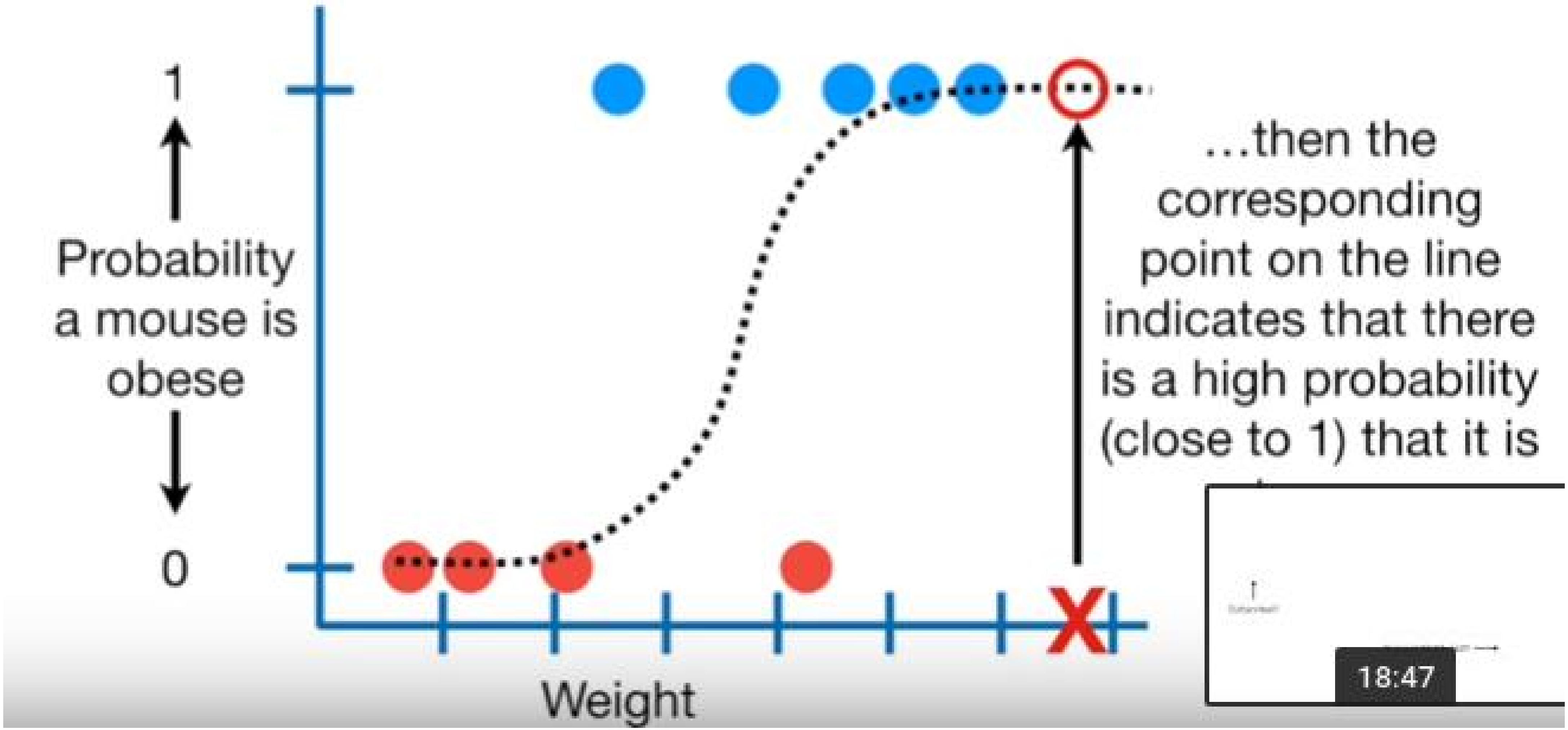
We'll talk about the coefficients in the context of using a continuous variable like "weight" to predict obesity...

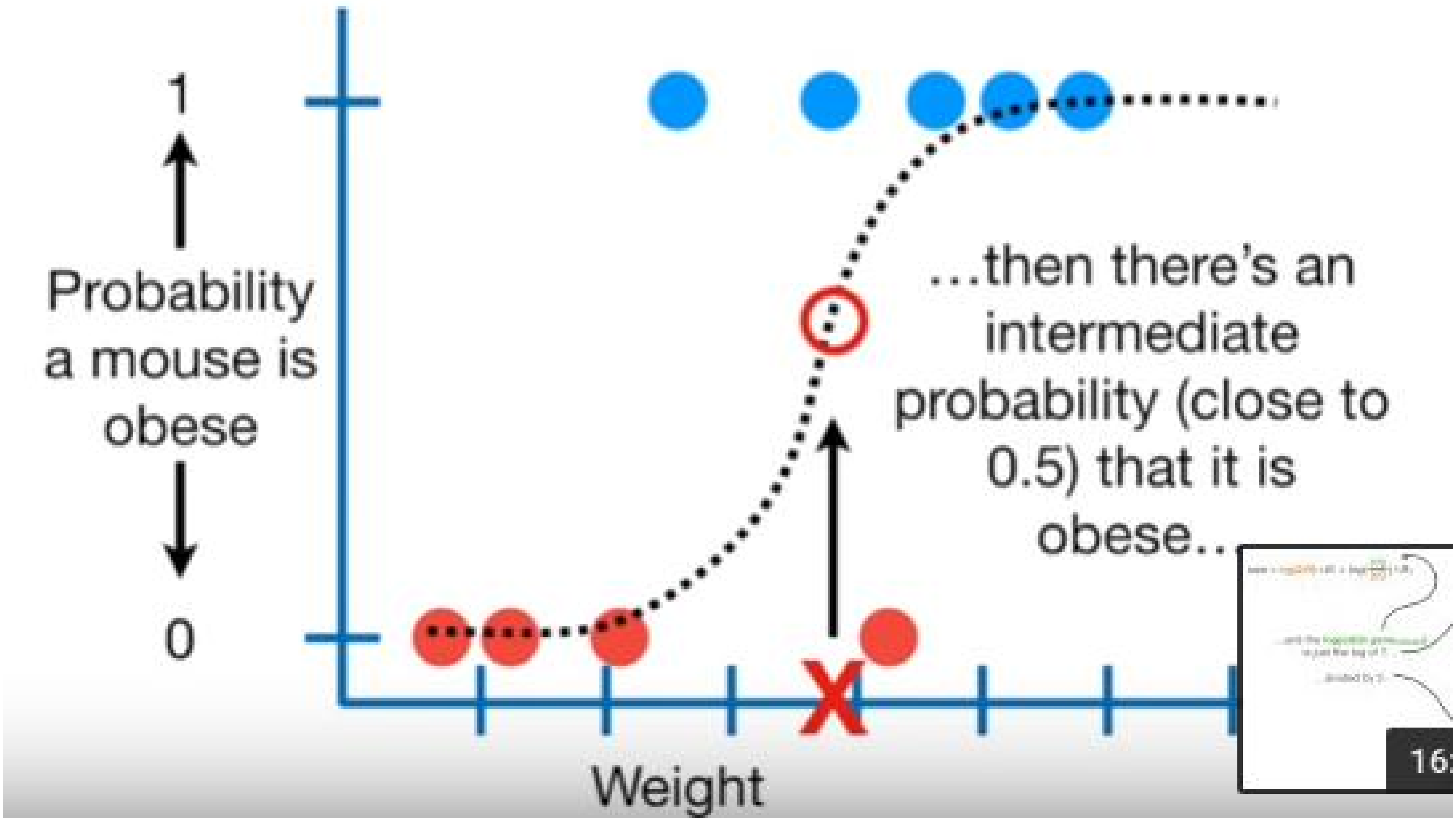


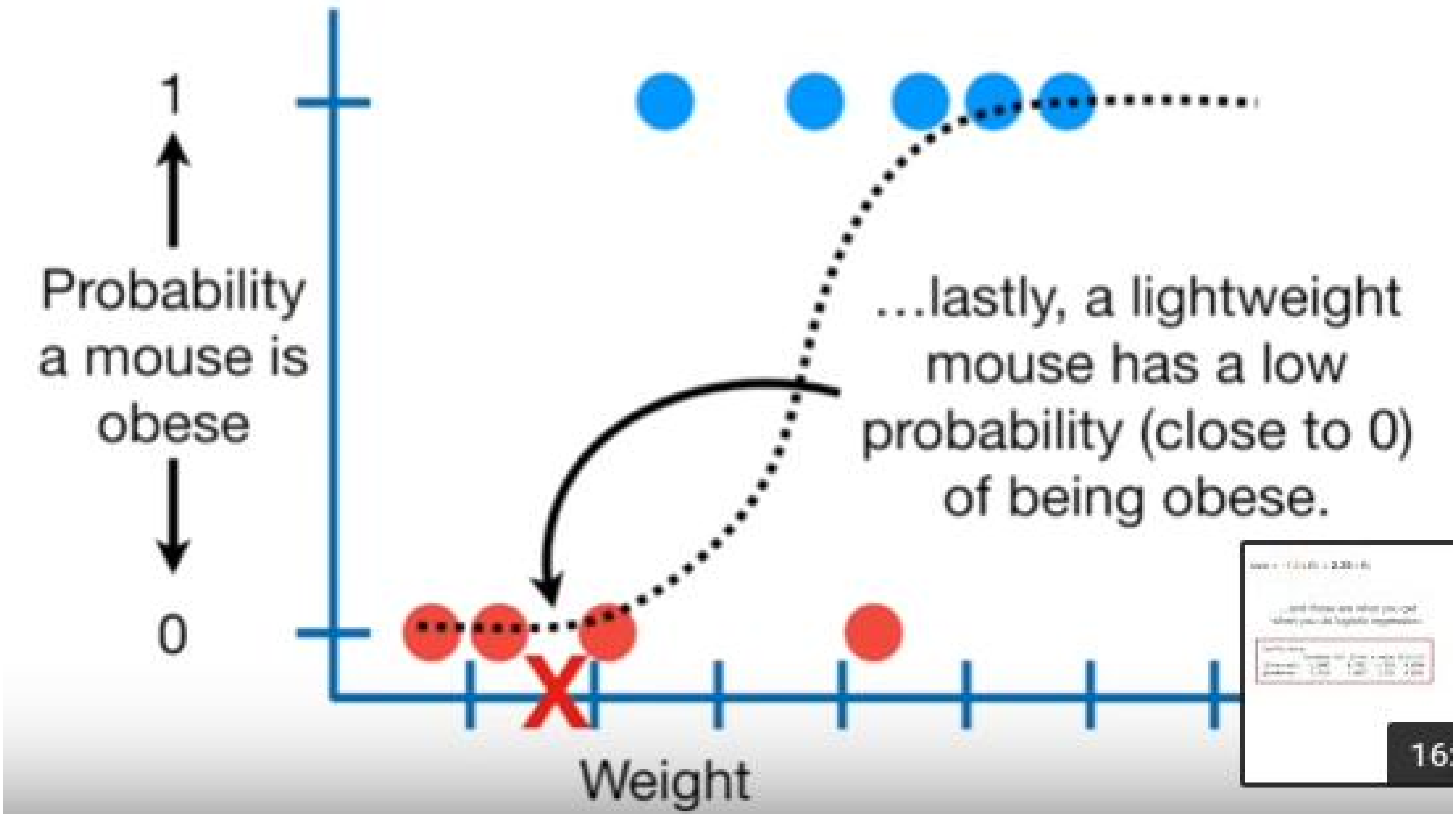

In this example, the
Y-axis is the probability
a mouse is obese.
1
0
Probability
a mouse is
obese
Weight

The dotted line is fit to the data to predict the probability a mouse is obese given its weight.

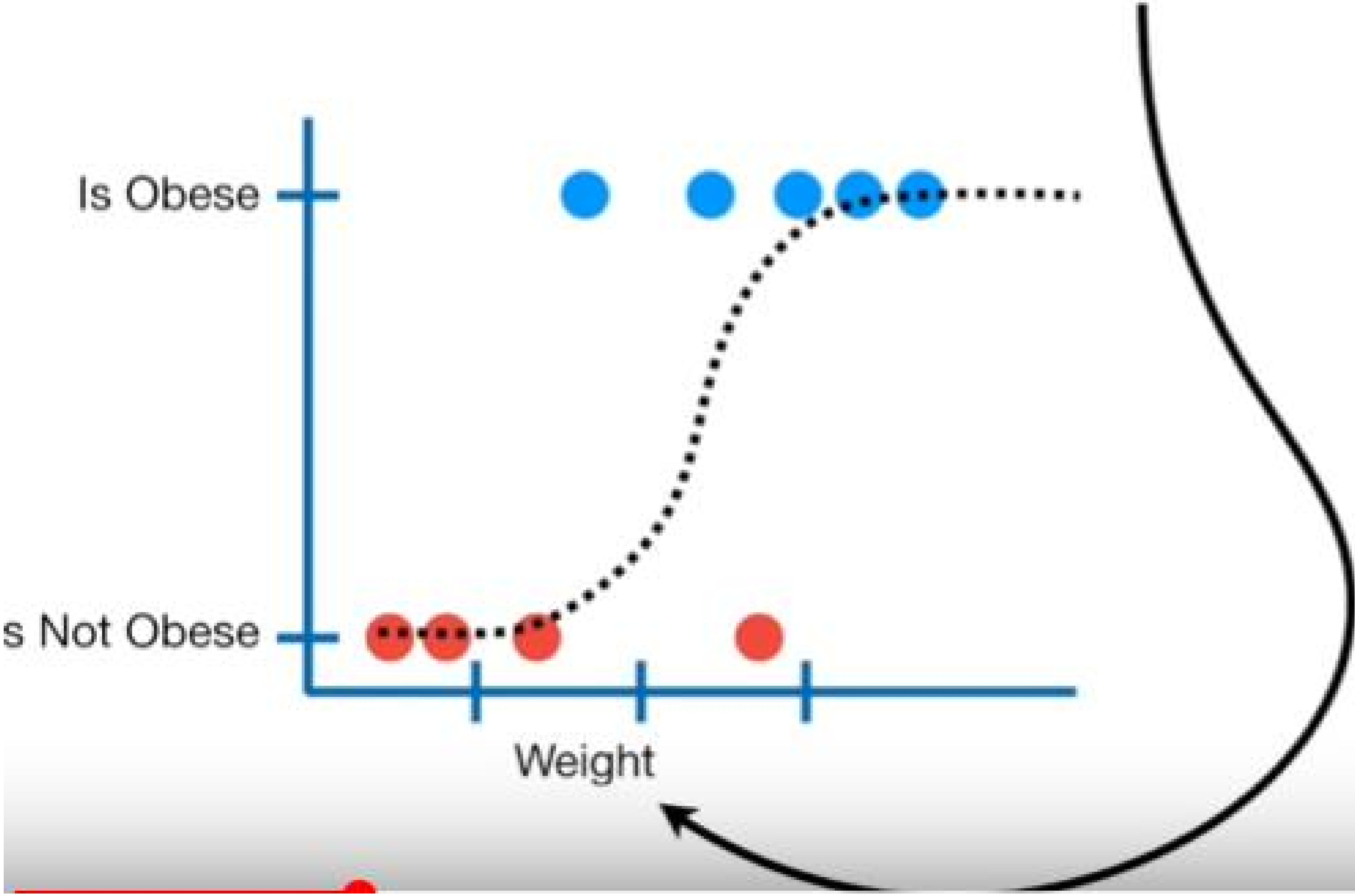


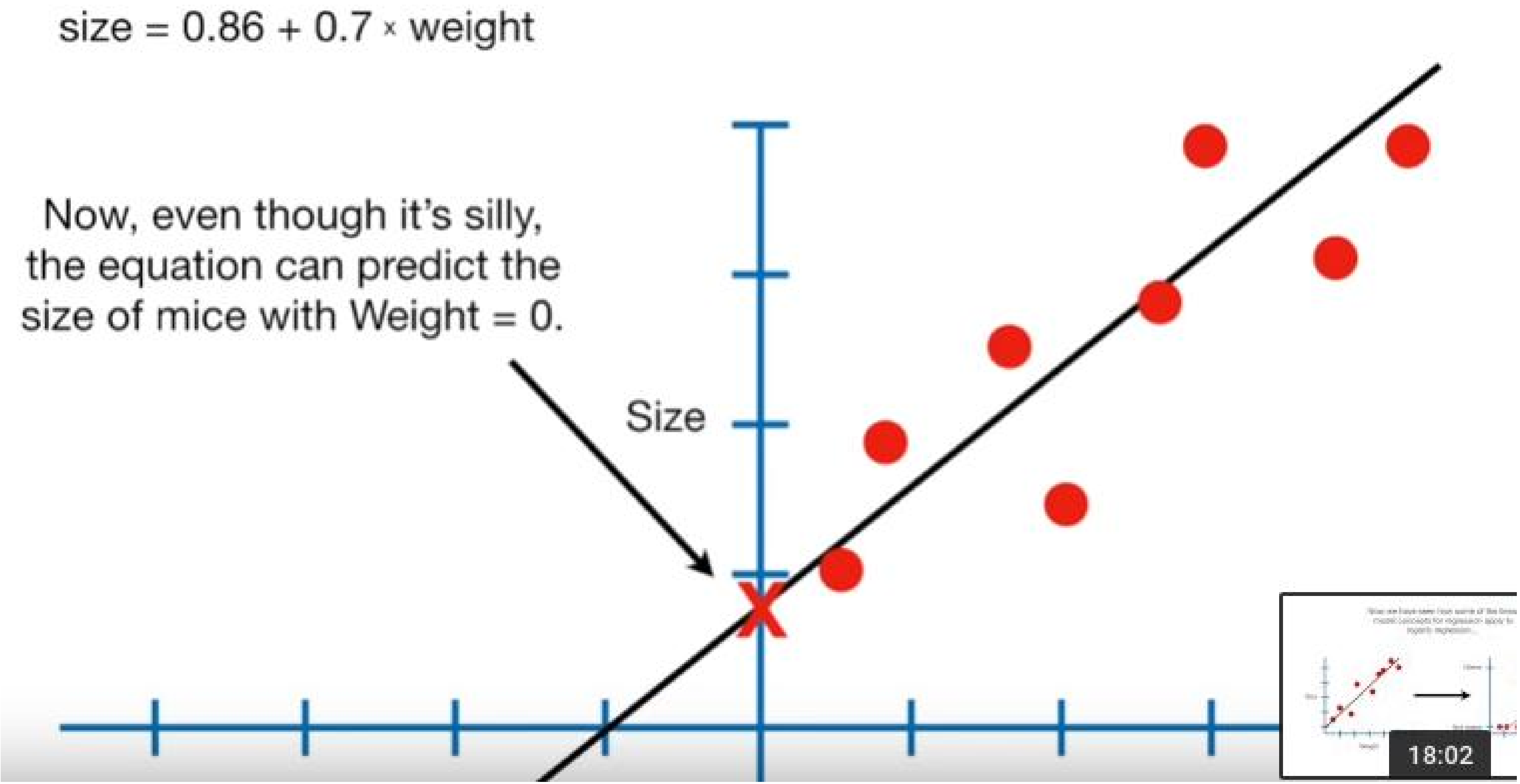

1
Probability
a mouse is
obese
0
Weight
…then the
corresponding
point on the line
indicates that there
is a high probability
(close to 1) that it is
18:47

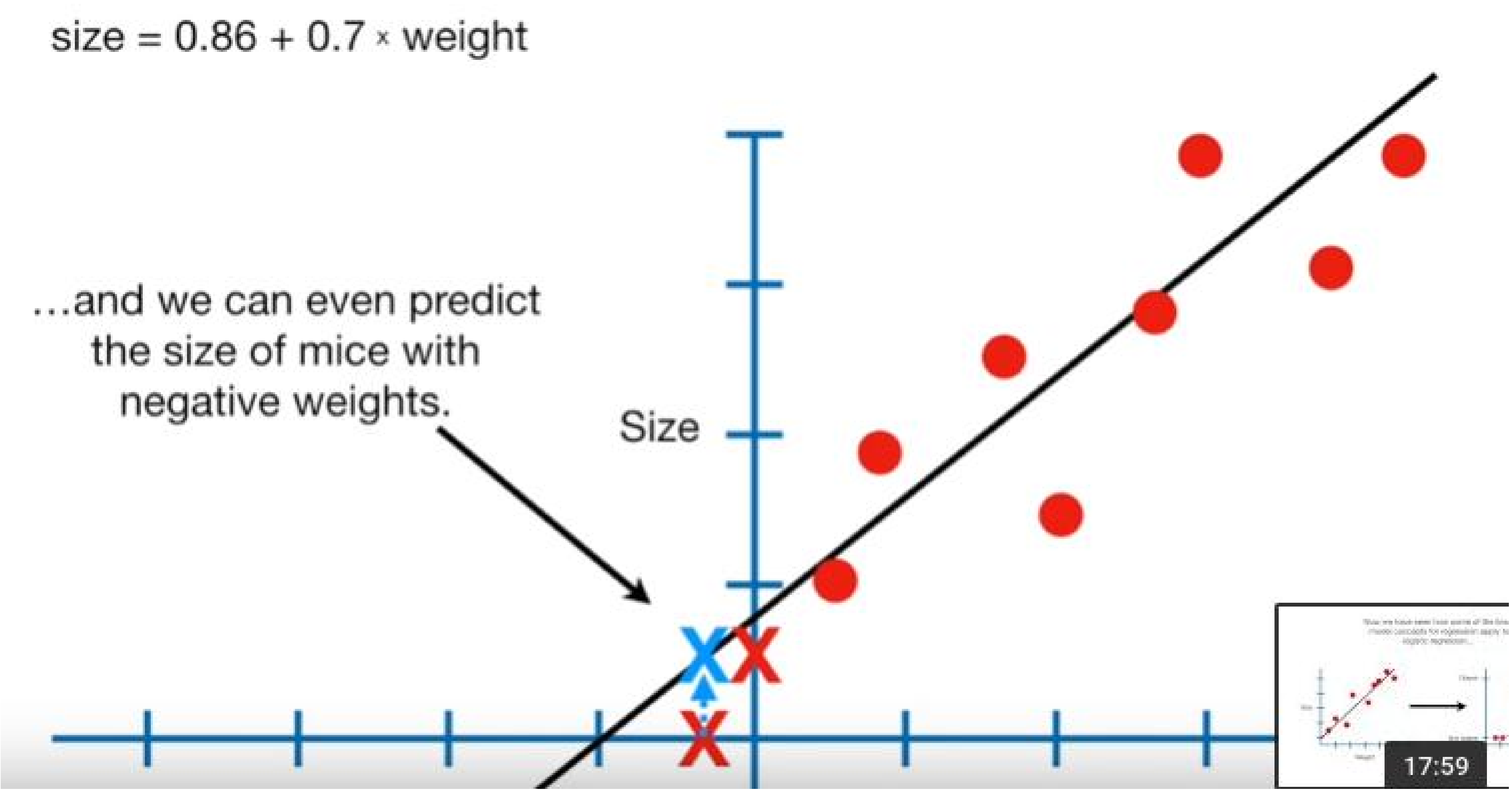

1
0
Probability a mouse is obese
Weight
...then there's an intermediate probability (close to 0.5) that it is obese...


1
Probability
a mouse is
obese
0
...lastly, a lightweight
mouse has a low
probability (close to 0)
of being obese.
X
Weight
16:

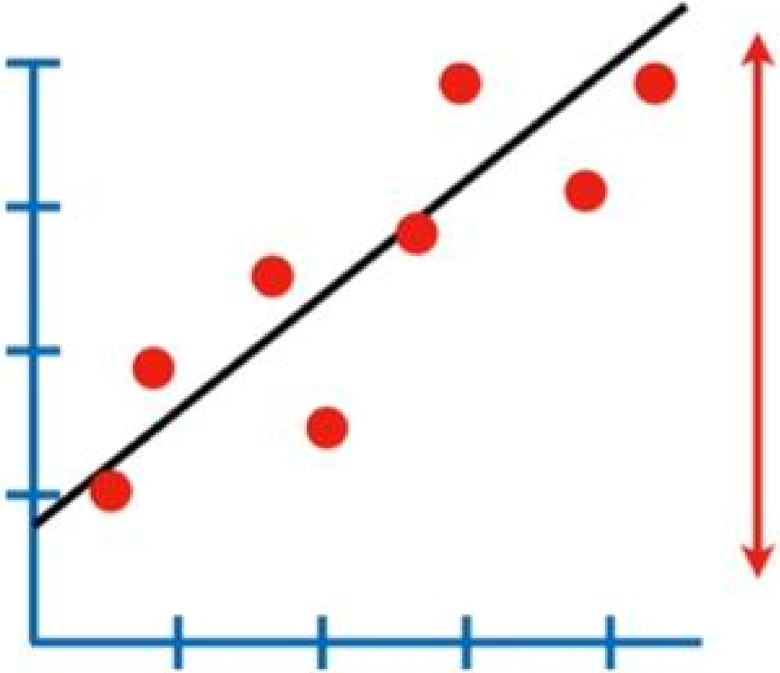
We'll start by talking about Logistic Regression when we use a continuous variable (like weight) to predict obesity.

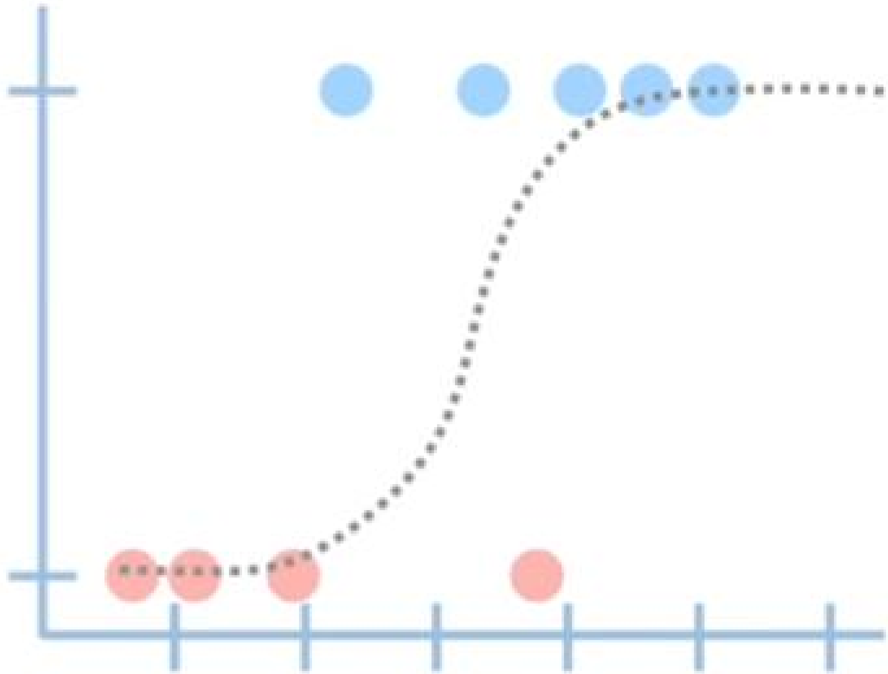



size = 0.86 + 0.7 × weight
Now, even though it's silly,
the equation can predict the
size of mice with Weight = 0.
Size
X
18:02

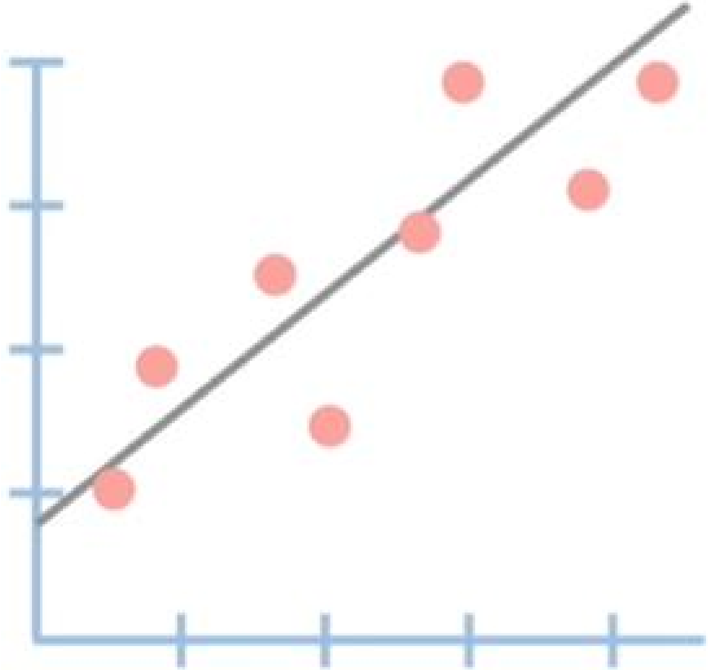

size = 0.86 + 0.7 × weight
…and we can even predict
the size of mice with
negative weights.
Size
17:59

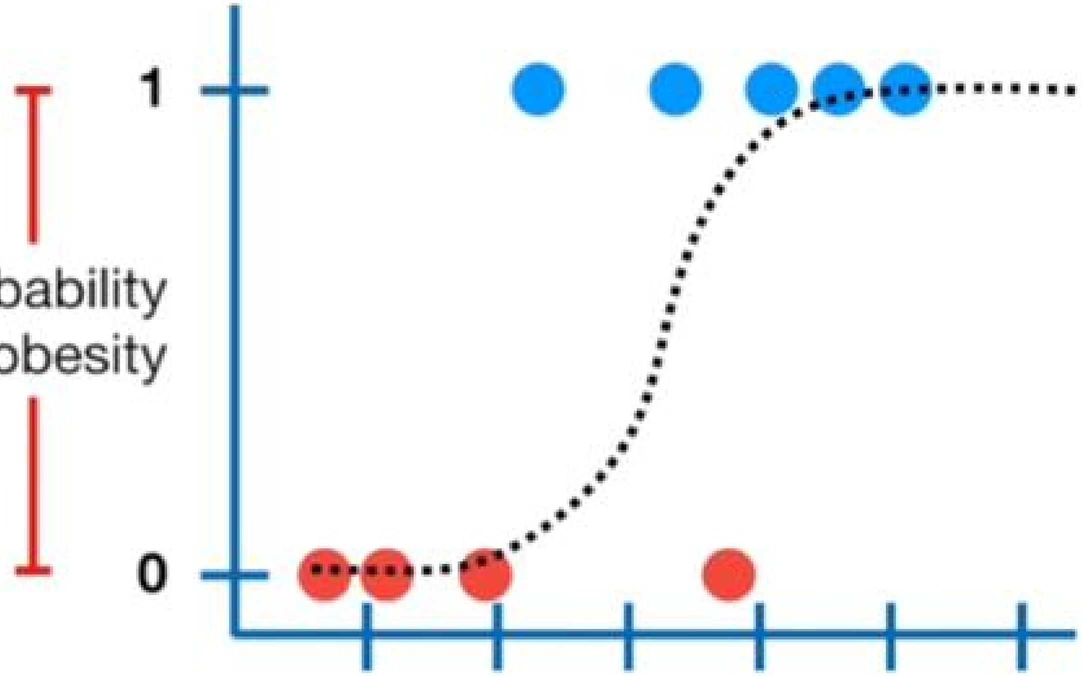
With linear regression, the values on the y-axis can, in theory, be any number...



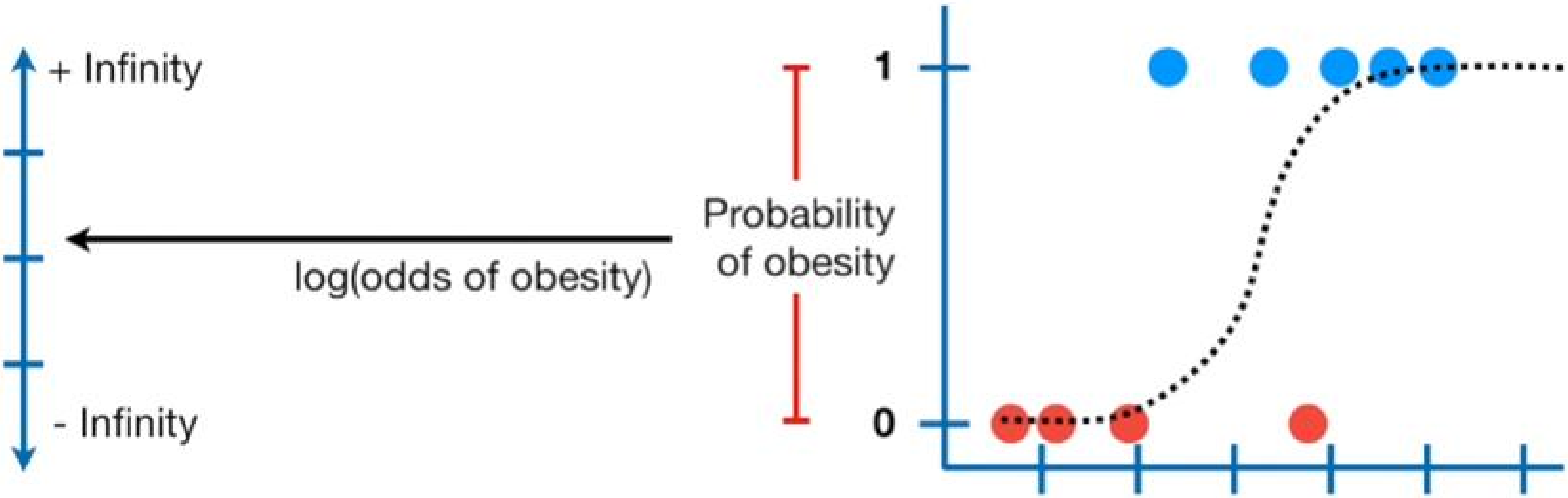


...unfortunately, with logistic regression, the y-axis is confined to probability values between 0 and 1.

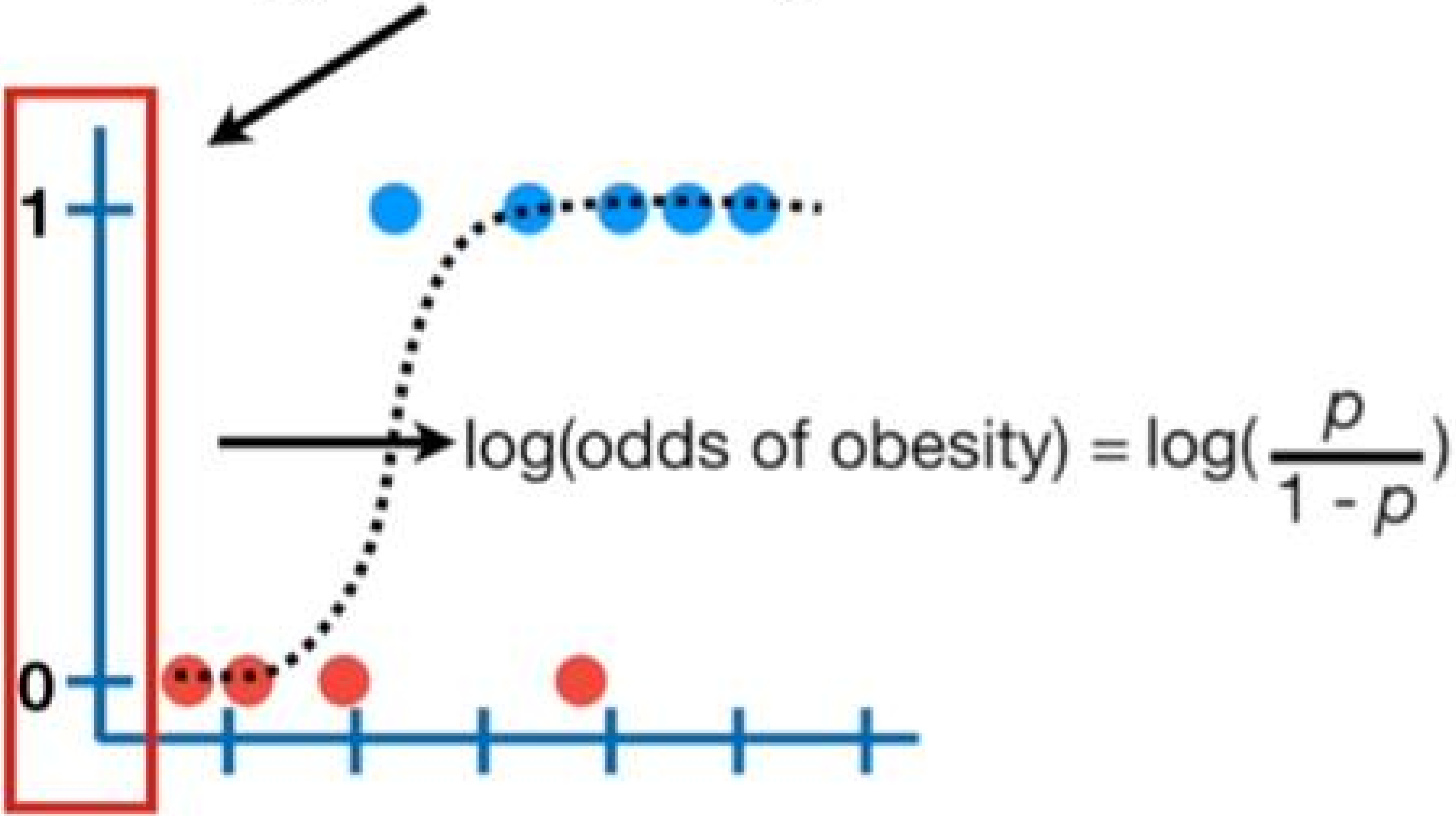


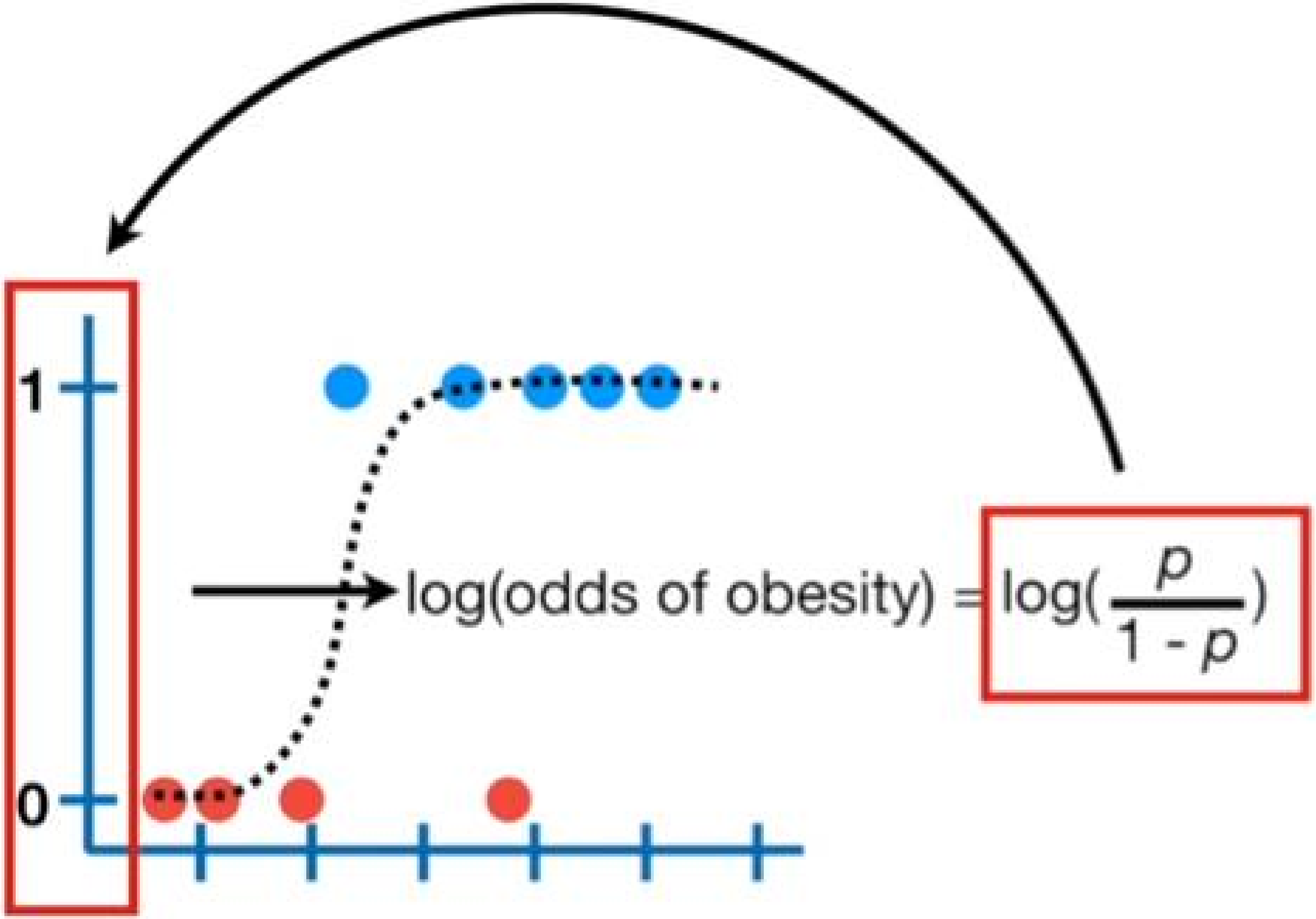


To solve this problem, the y-axis in logistic regression is transformed from the “probability of obesity” to the “log(odds of obesity)” so, just like the y-axis in linear regression, it can go from -infinity to +infinity.

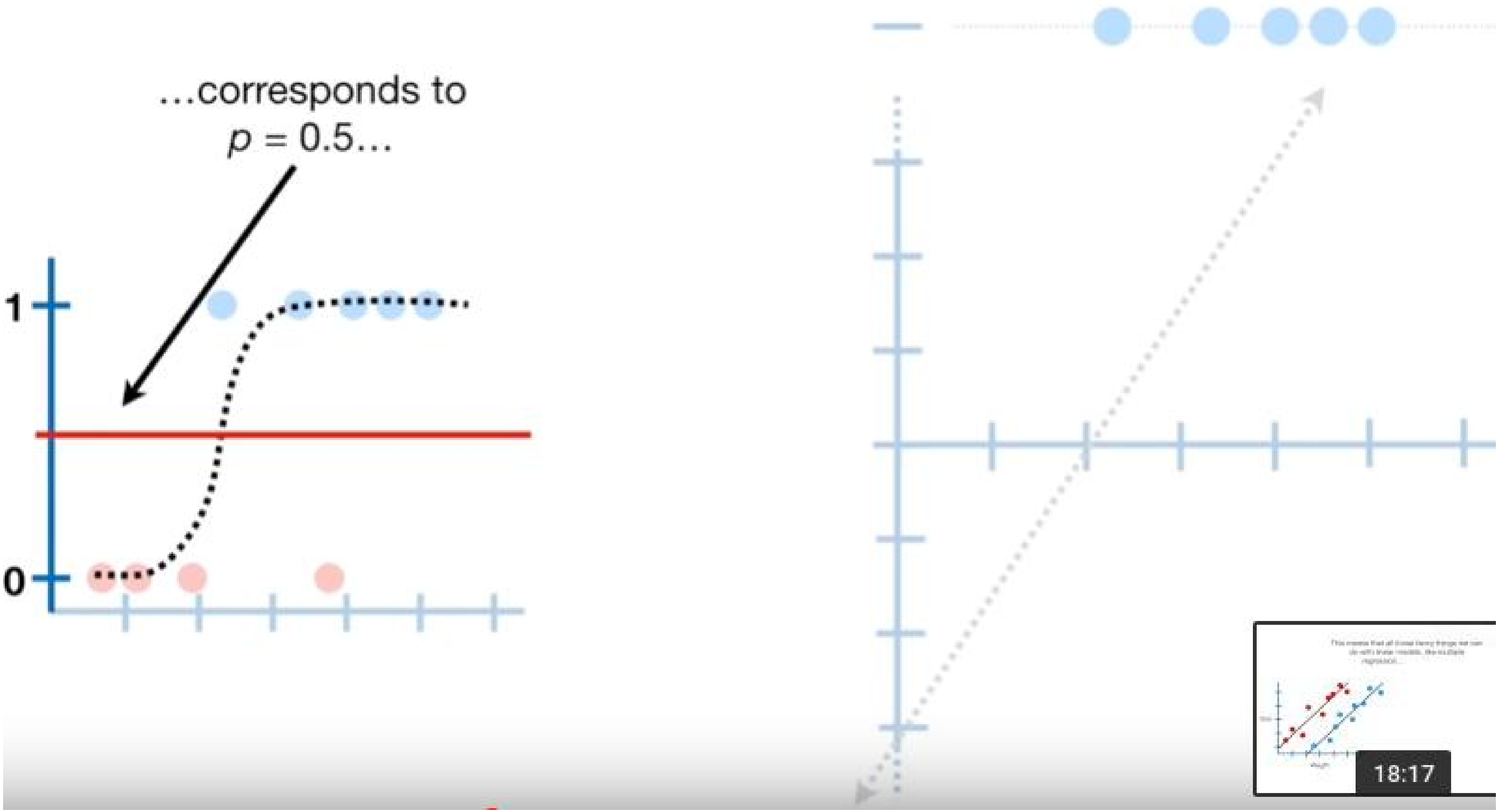


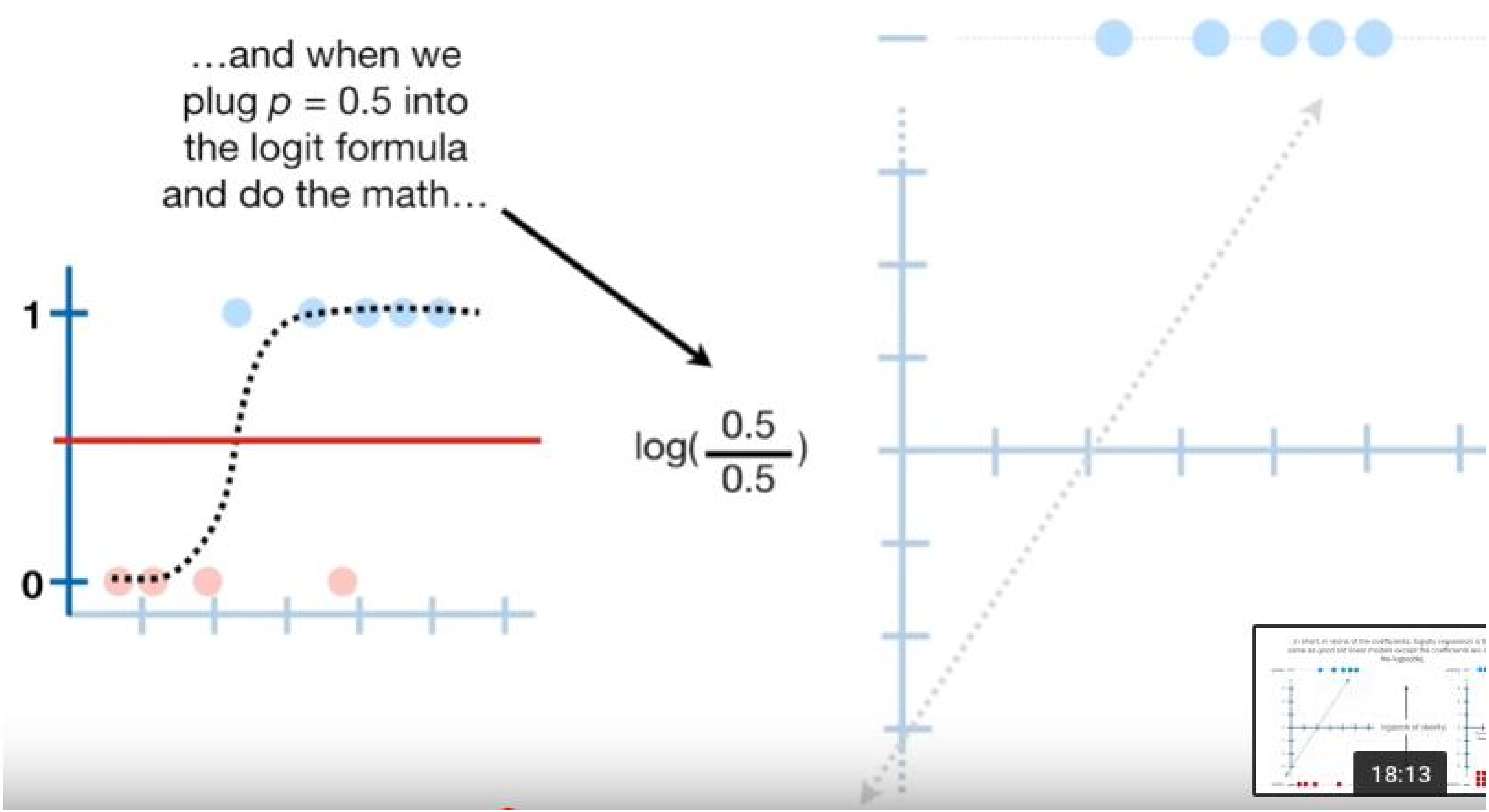
Now let's transform this y-axis from a "probability of obesity" scale to a "log(odds of obesity)" scale.



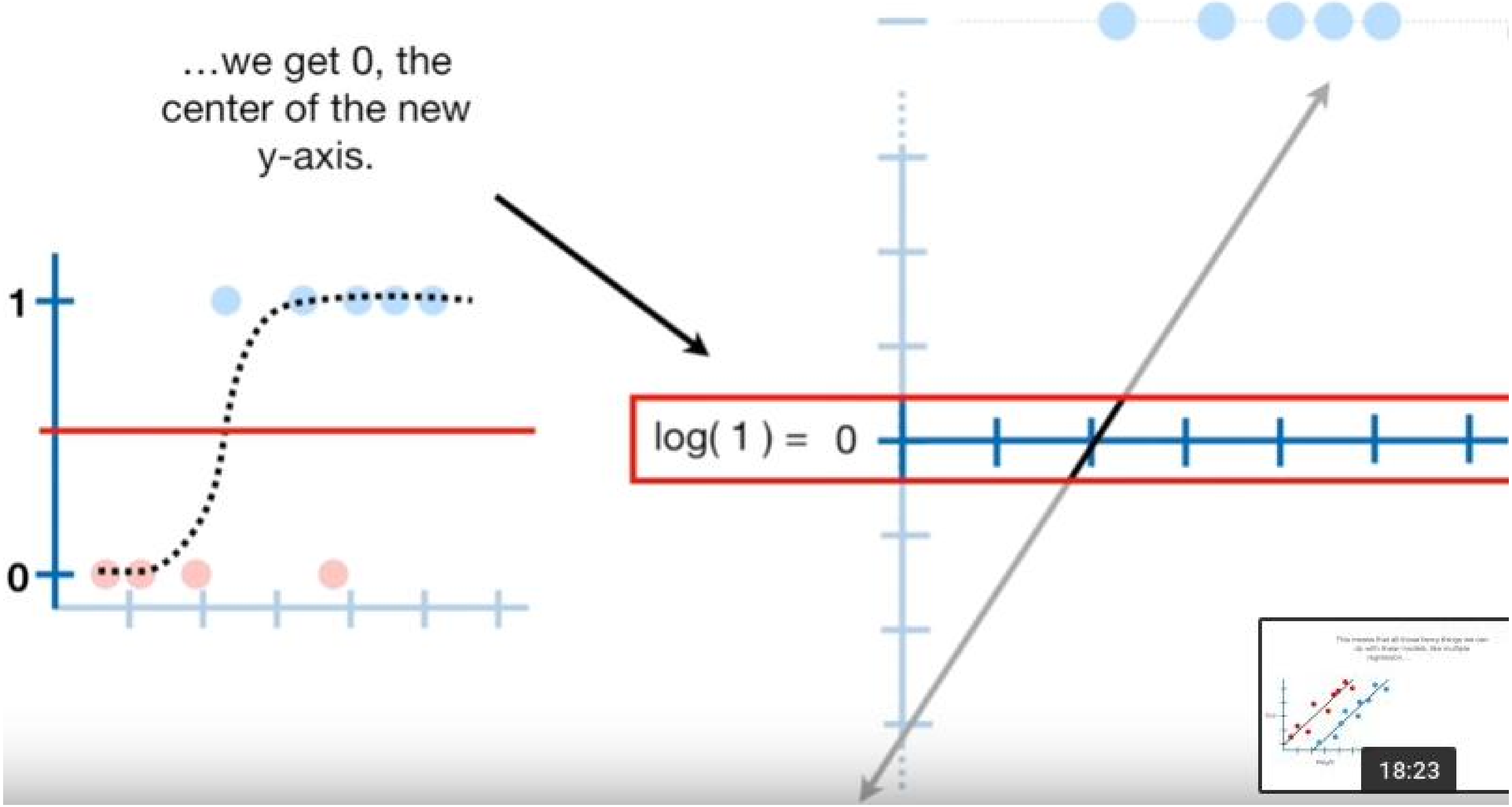


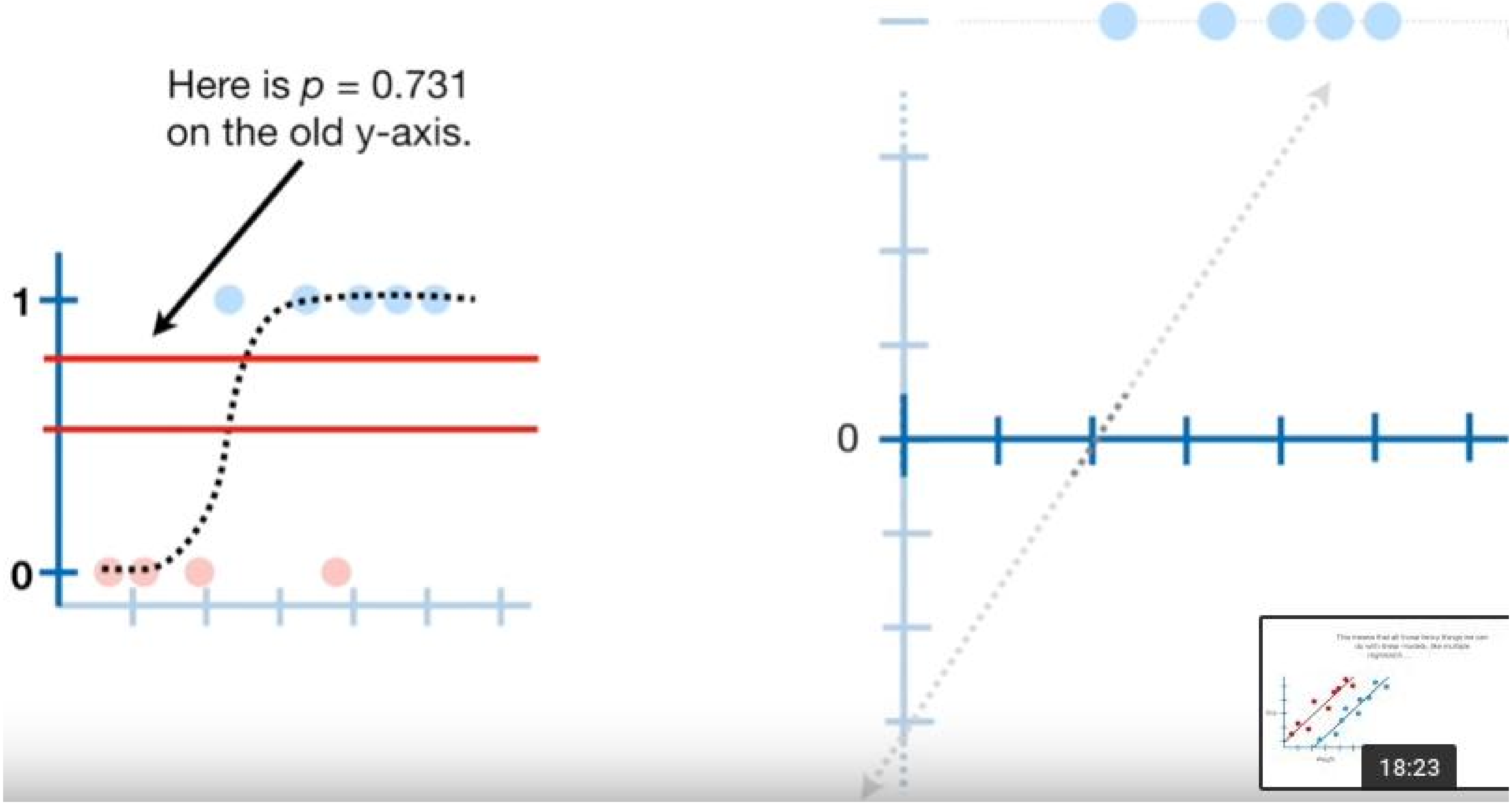
p, in this case, is the probability of a mouse being obese, and corresponds a value on the old y-axis between 0 and 1.

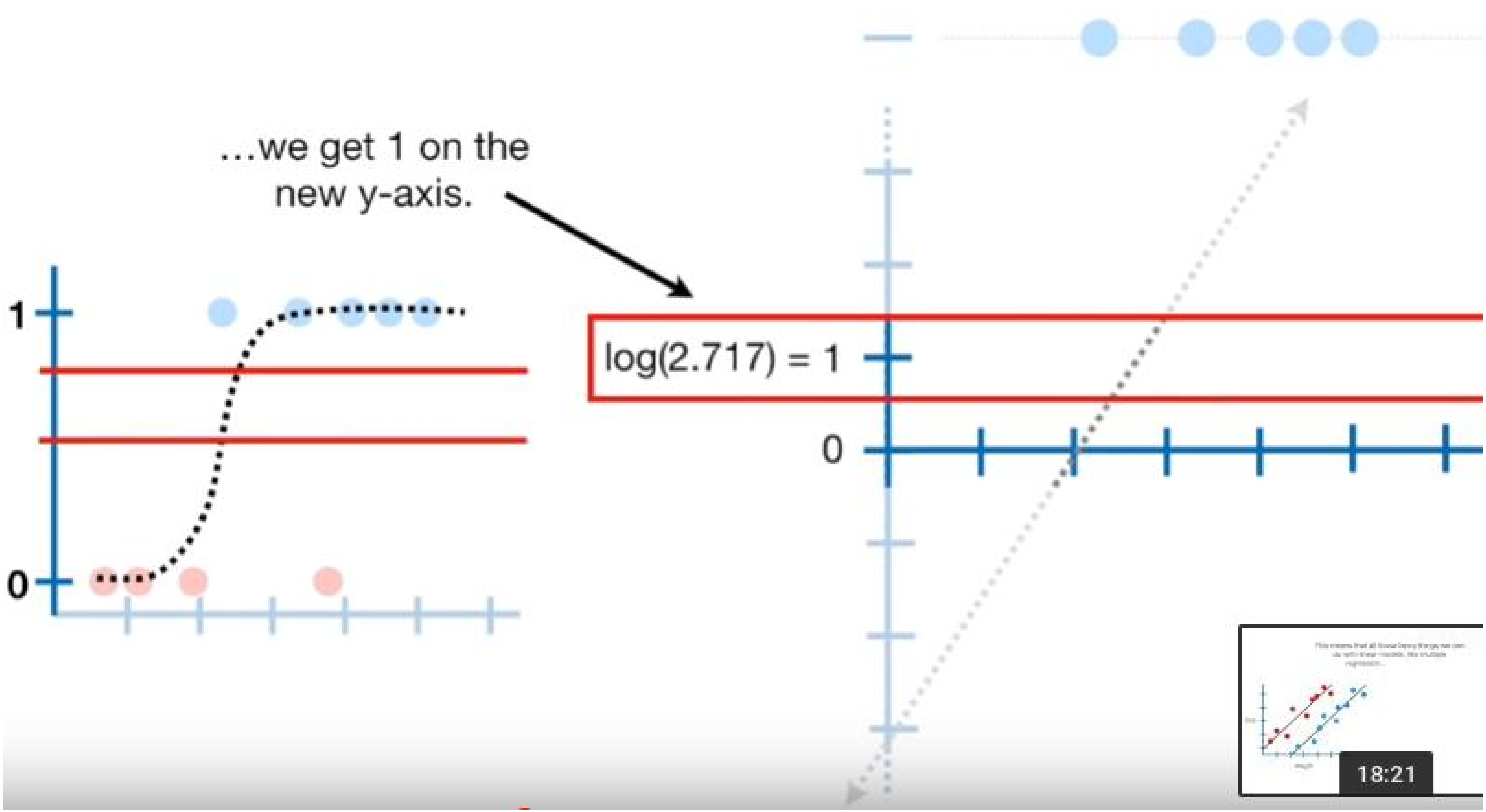

...corresponds to
$p = 0.5$...
1
0
18:17



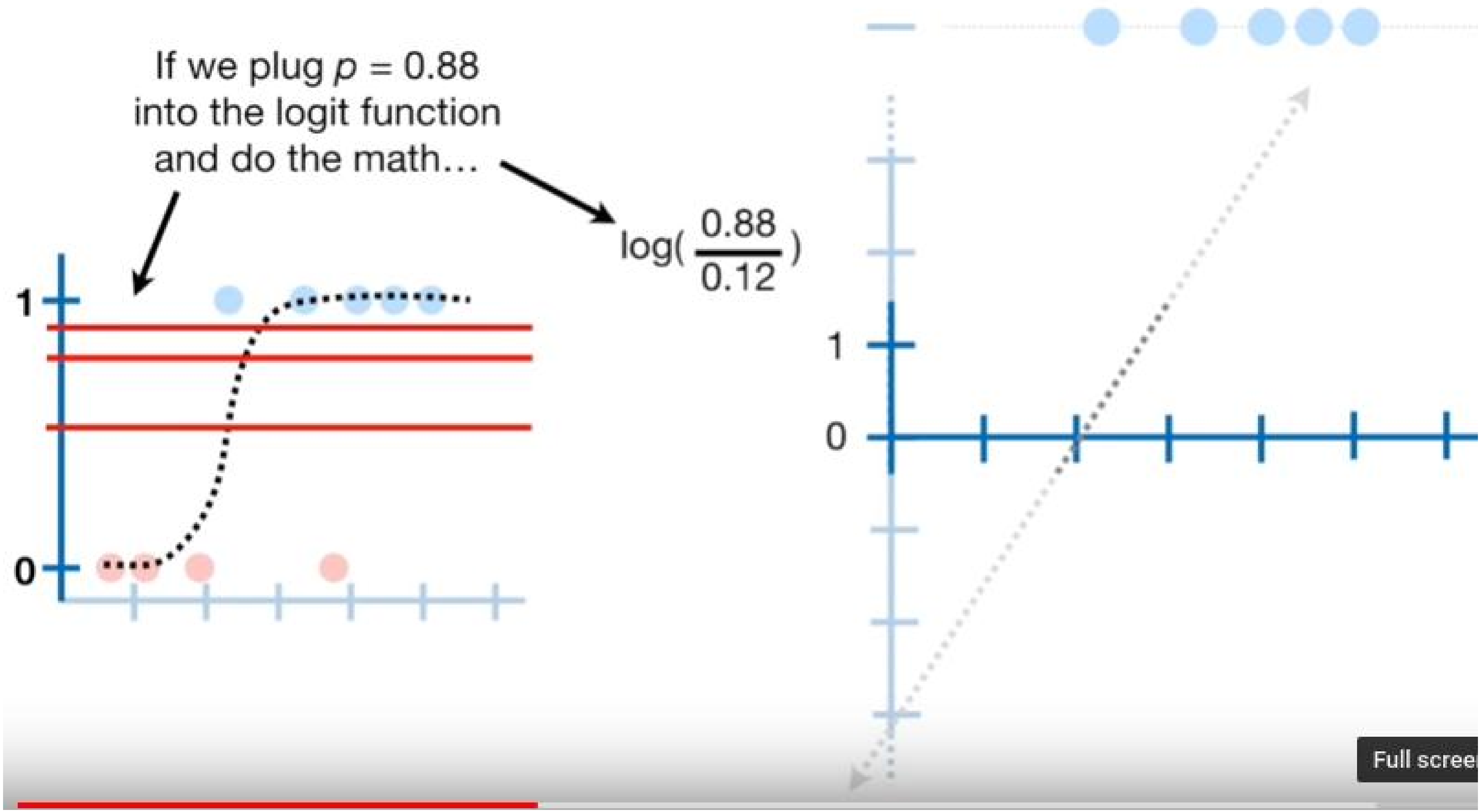
...and when we plug p = 0.5 into the logit formula and do the math...
1
0
log(0.5/0.5)
18:13

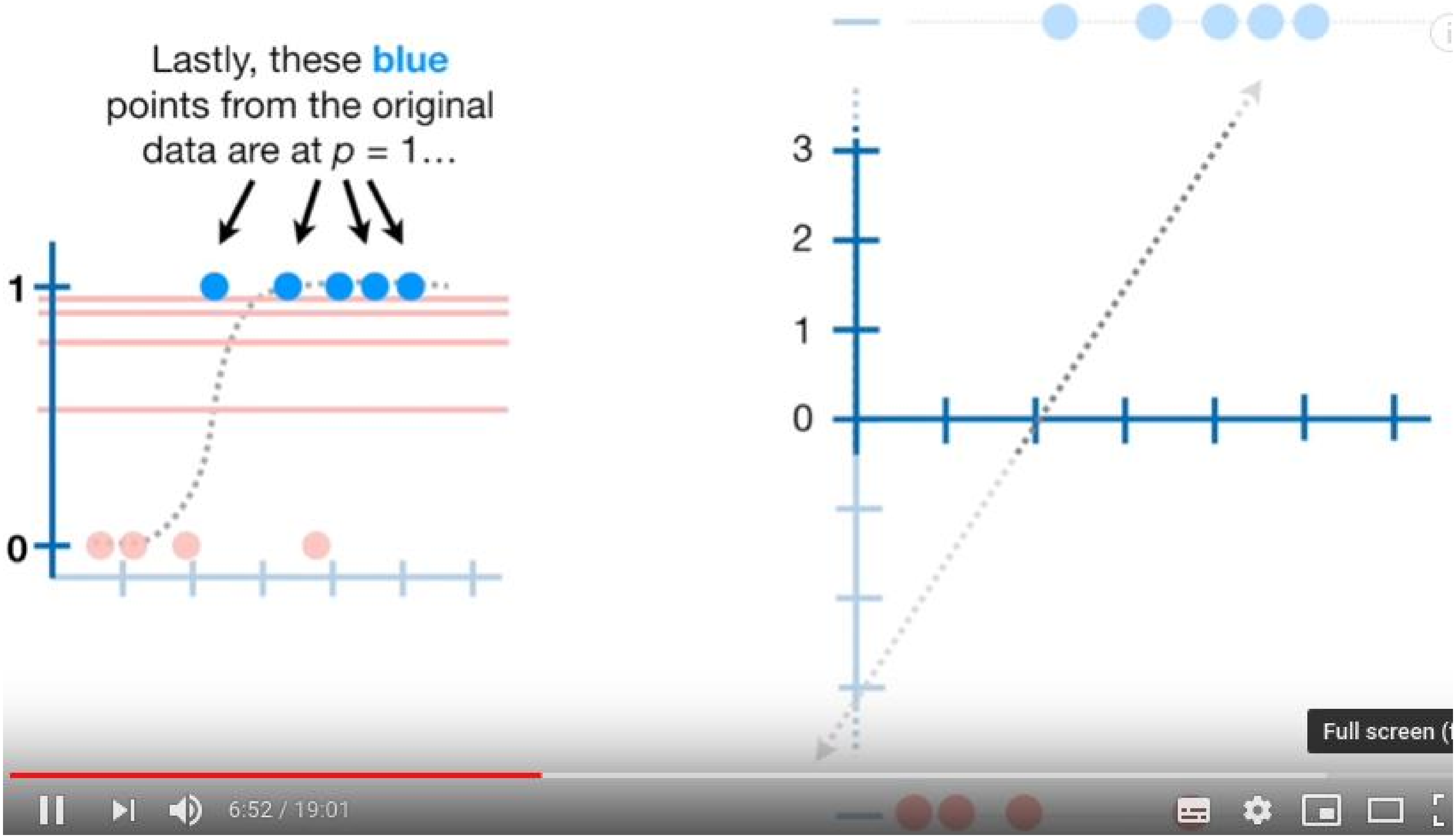

…we get 0, the
center of the new
y-axis.
1
0
log(1) = 0
18:23


Here is p = 0.731
on the old y-axis.
1
0
0
18:23

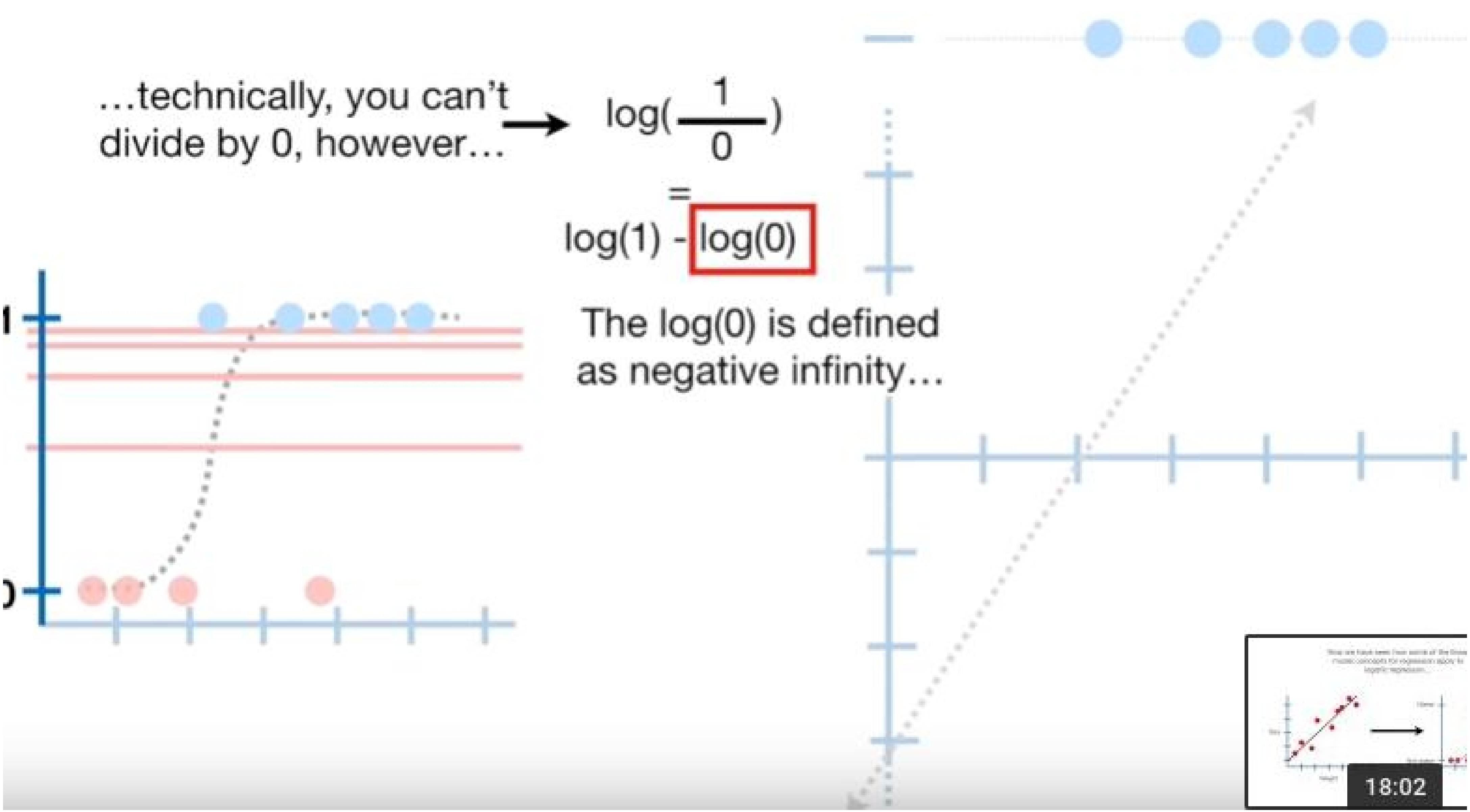


...we get 1 on the new y-axis.
1
0
log(2.717) = 1
0
18:21


If we plug $p = 0.88$ into the logit function and do the math…
$\log(\frac{0.88}{0.12})$
1
0
1
0
Full scree



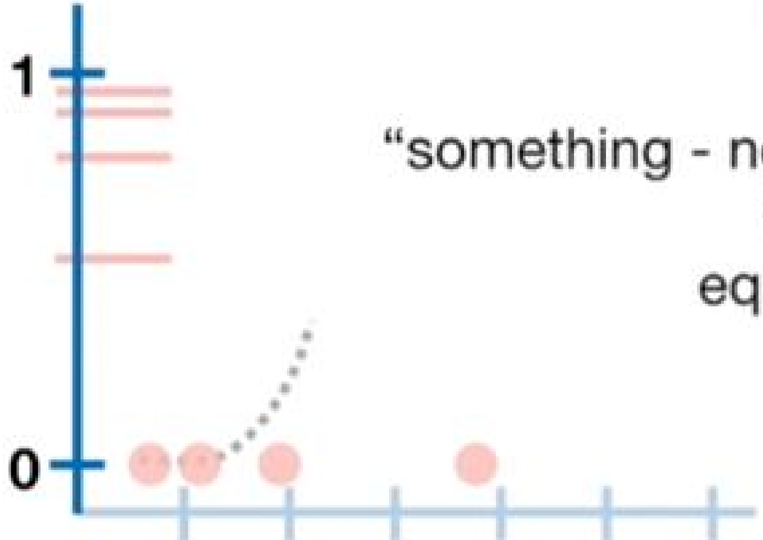
Lastly, these blue points from the original data are at p = 1…
1
0
3
2
1
0
Full screen (f
6:52 / 19:01


…technically, you can't divide by 0, however…
log($\frac{1}{0}$)
=
log(1) - $\boxed{\text{log(0)}}$
The log(0) is defined as negative infinity…
18:02

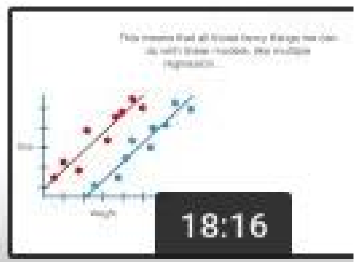
…technically, you can't divide by 0, however… → $\log(\frac{1}{0})$

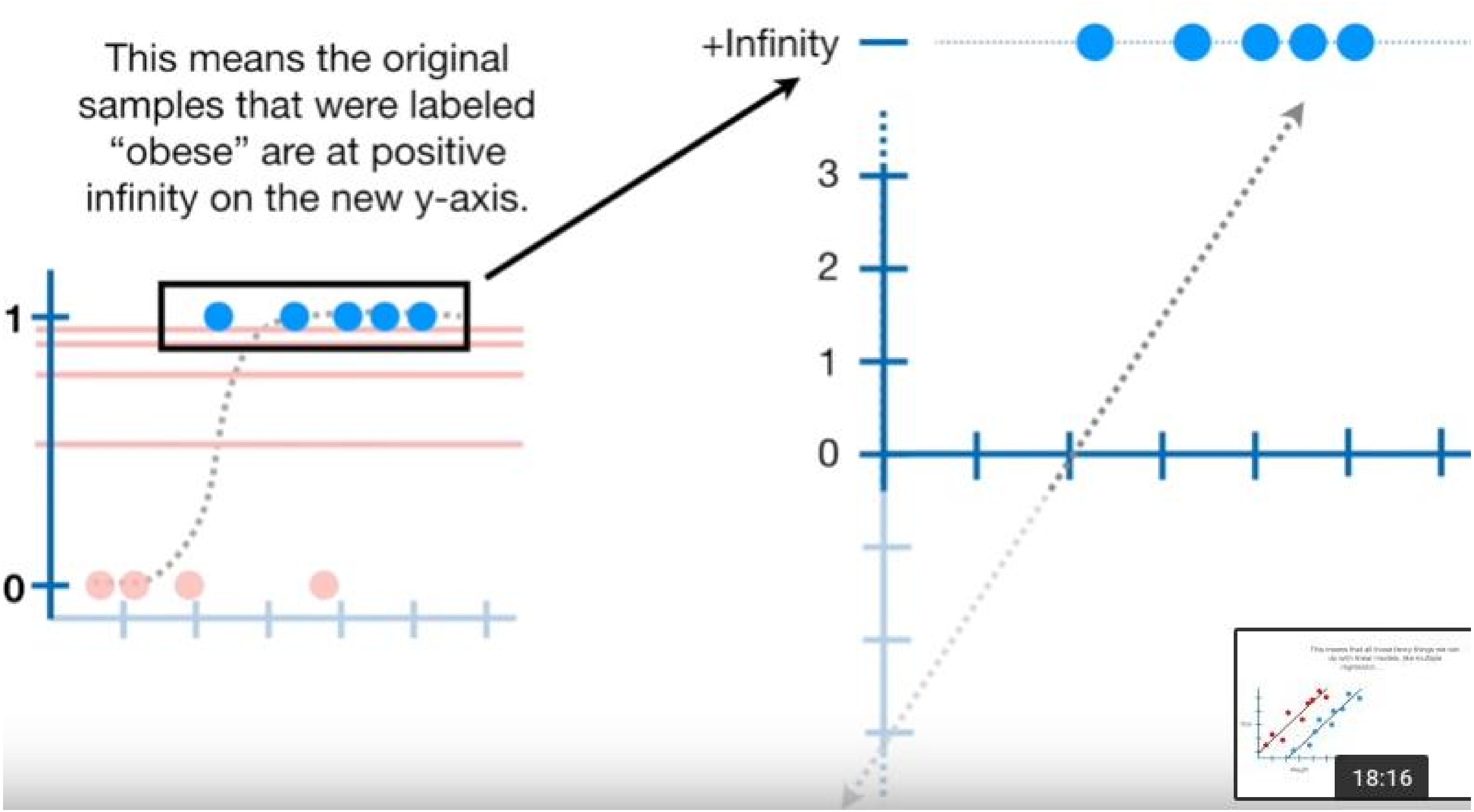
$=$

$\boxed{\log(1) - \log(0)}$

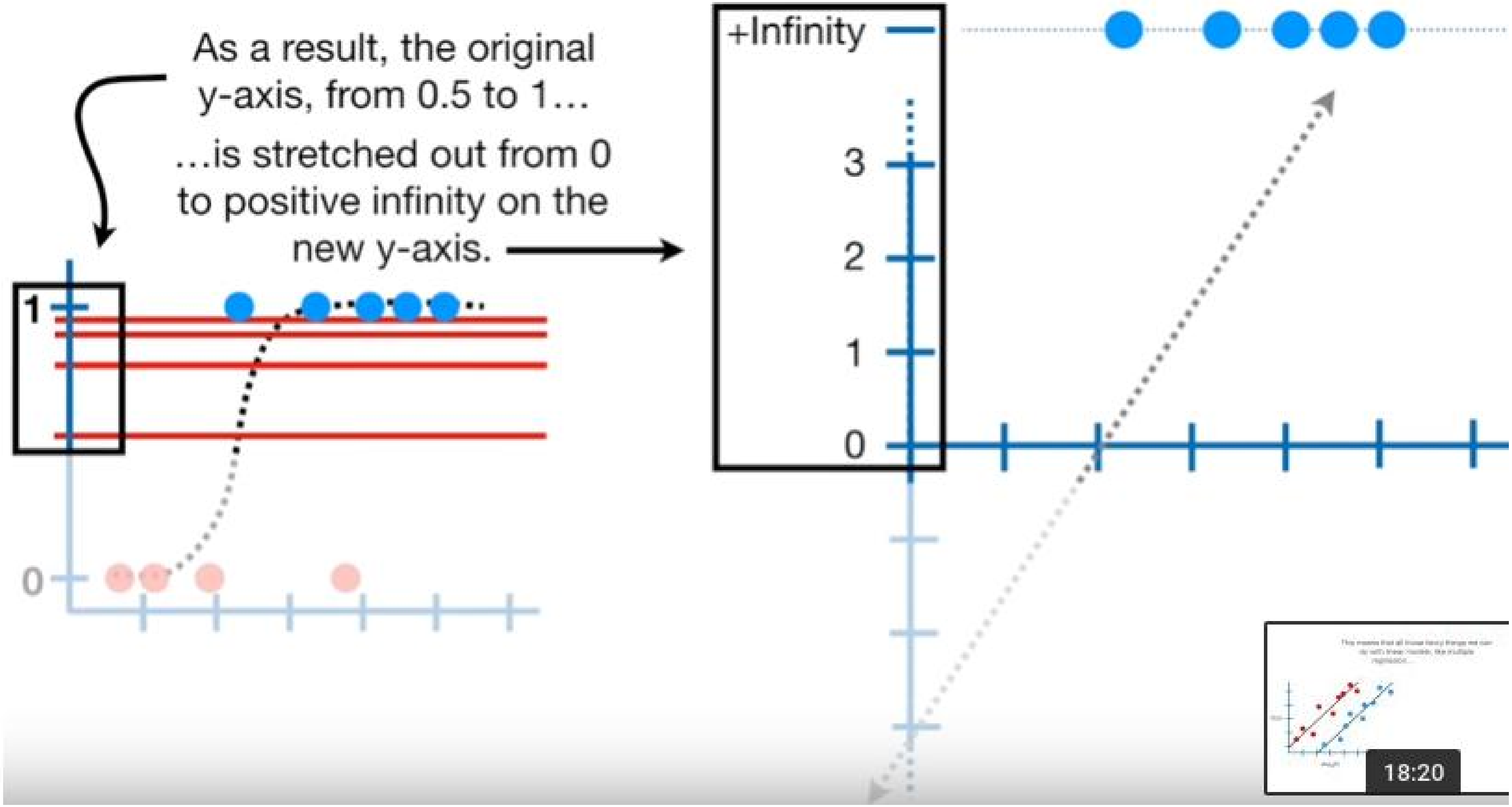


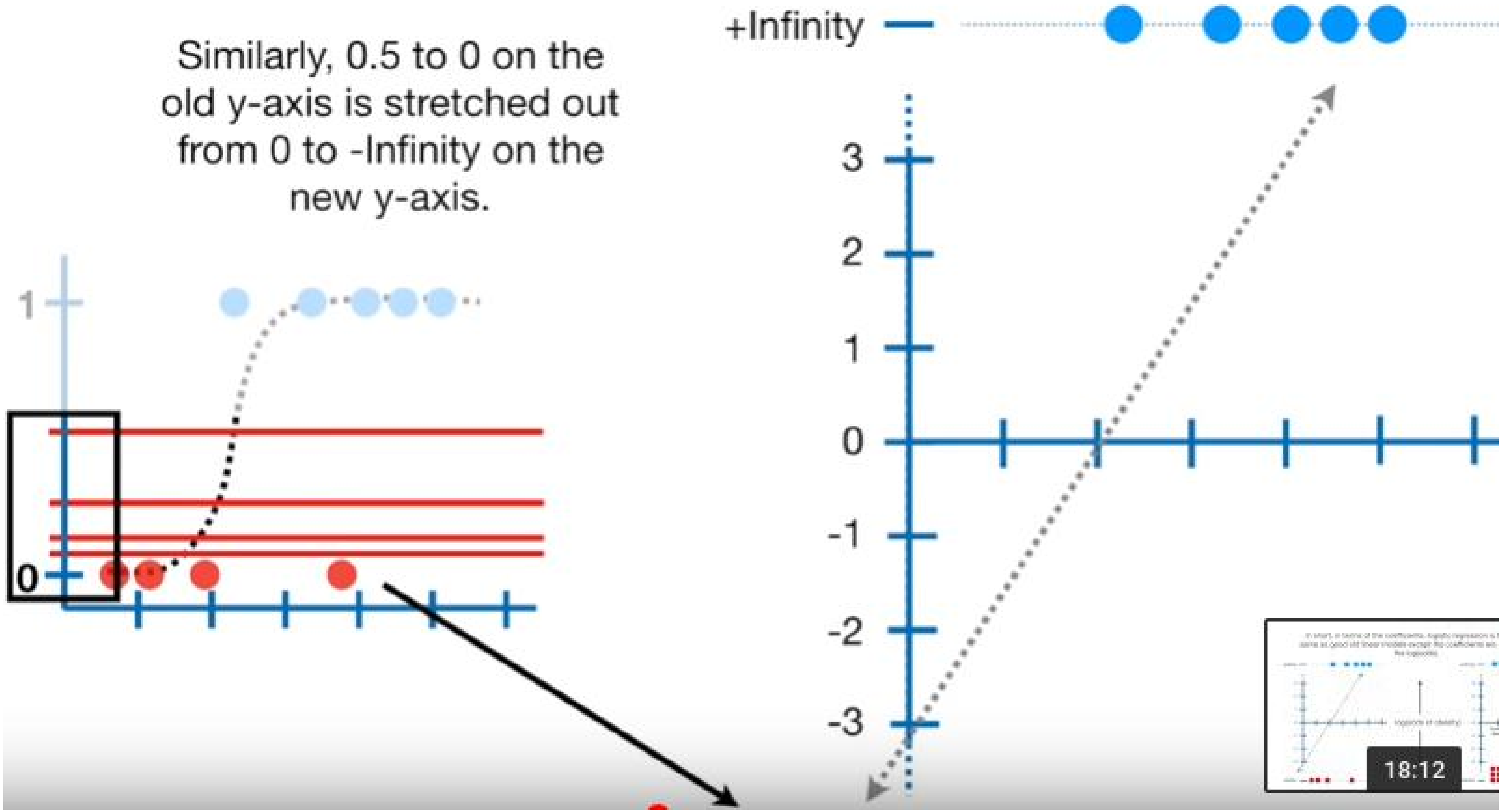
…and since
"something - negative infinity = positive infinity",
this whole thing is
equal to positive infinity.



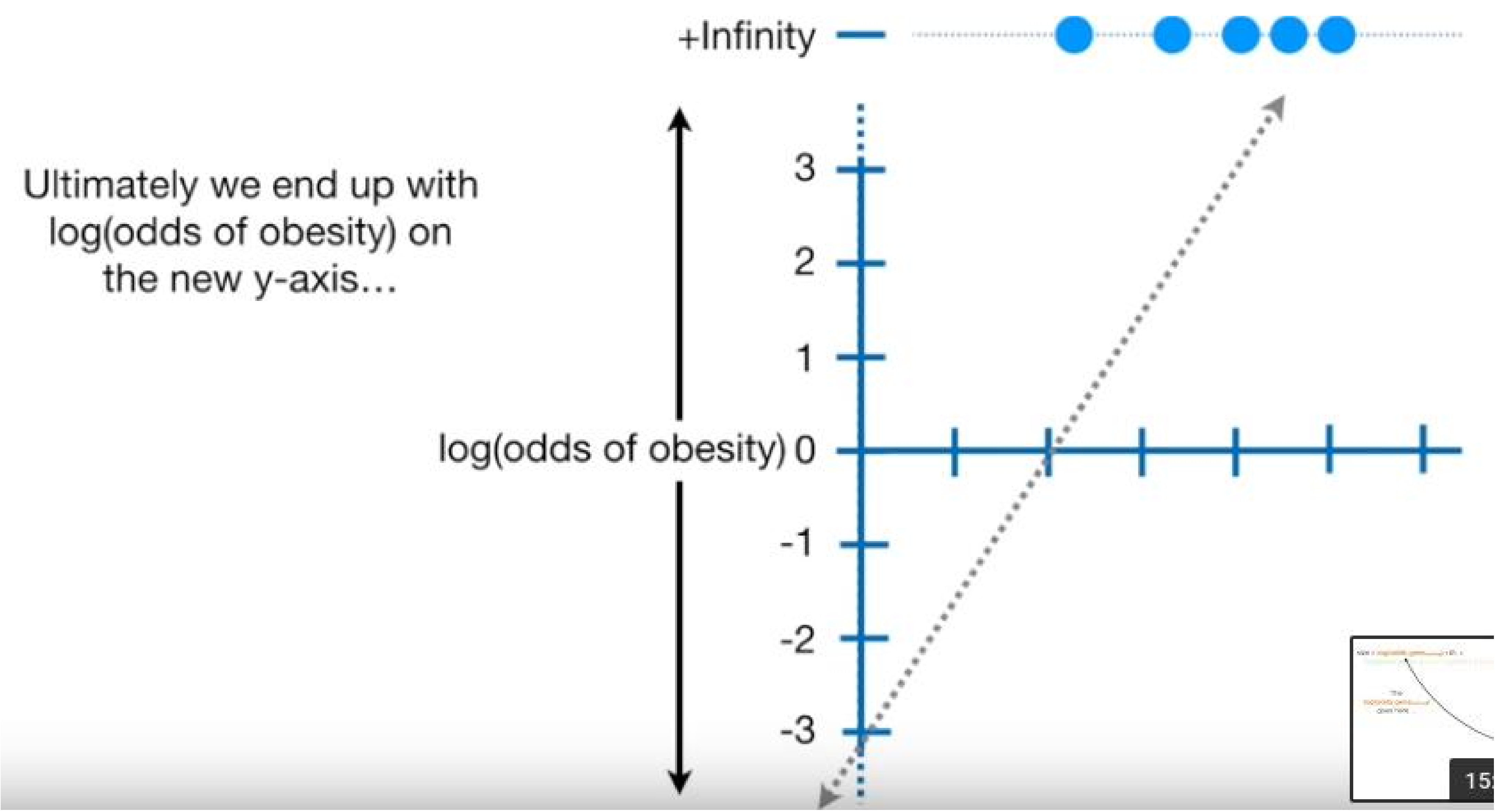


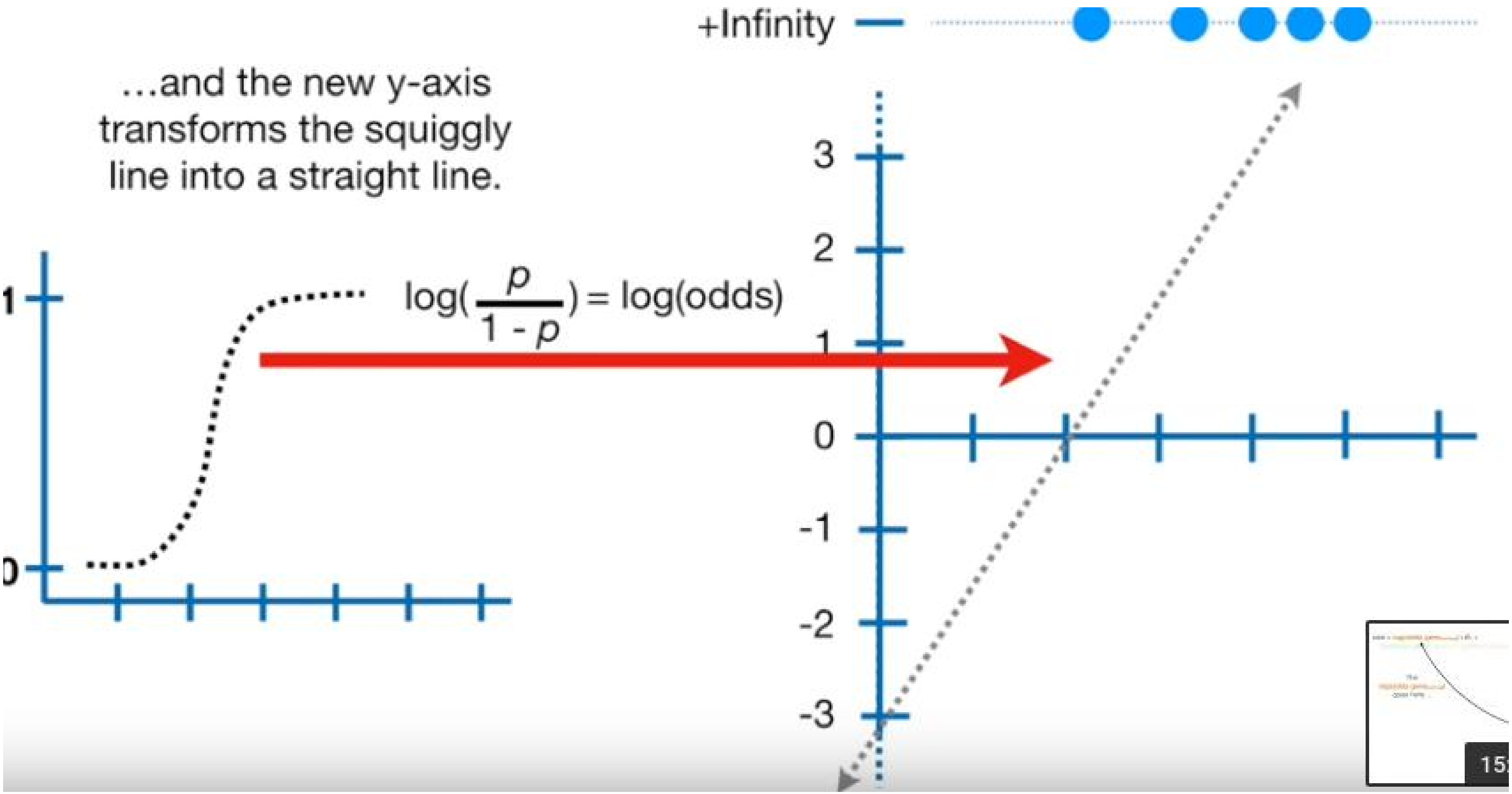
This means the original samples that were labeled "obese" are at positive infinity on the new y-axis.
+Infinity
3
2
1
0
1
0
18:16


As a result, the original y-axis, from 0.5 to 1…
…is stretched out from 0 to positive infinity on the new y-axis.
1
0
+Infinity
3
2
1
0
18:20



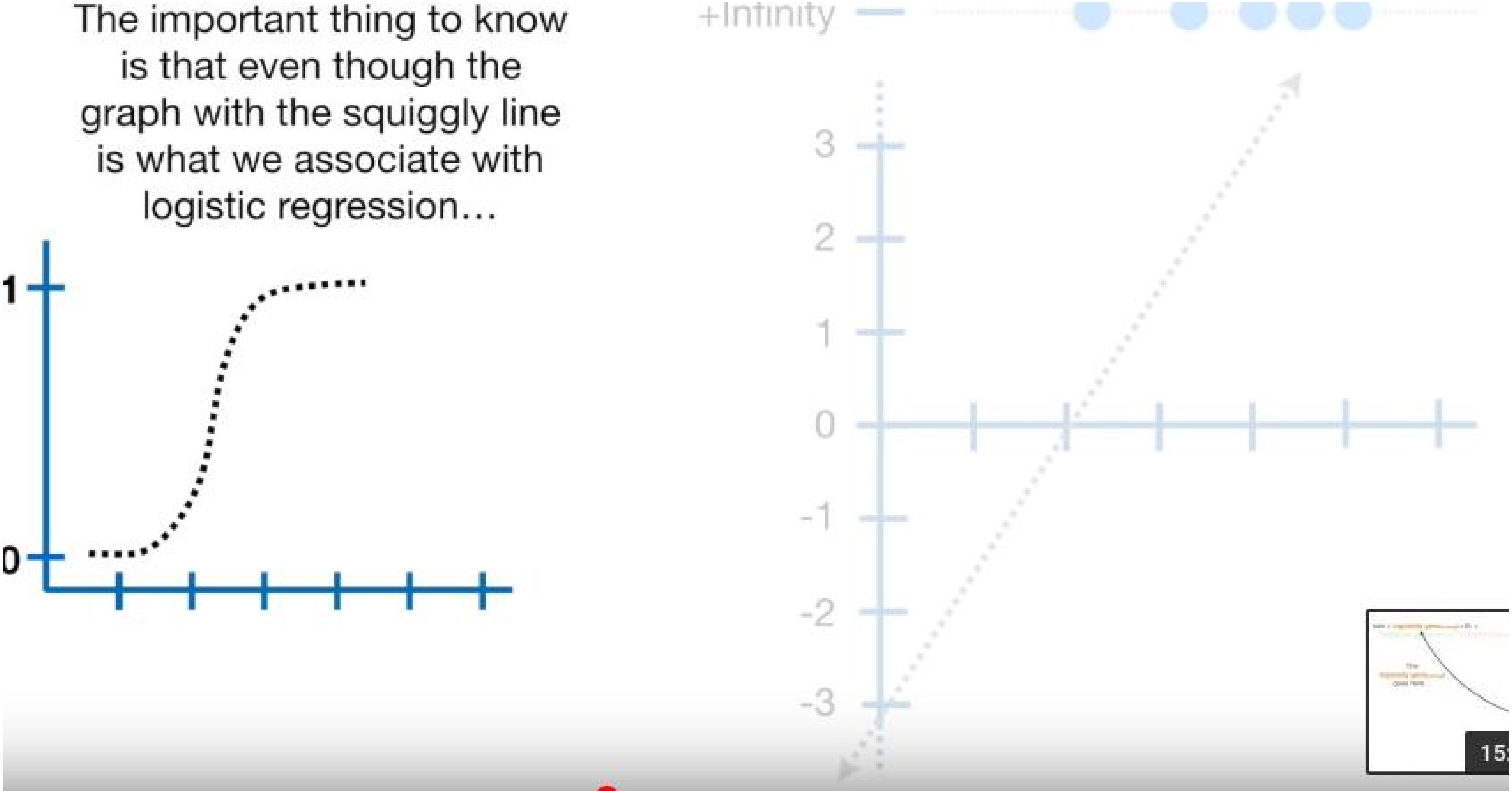
Similarly, 0.5 to 0 on the old y-axis is stretched out from 0 to -Infinity on the new y-axis.
+Infinity
3
2
1
0
-1
-2
-3
1
0
18:12

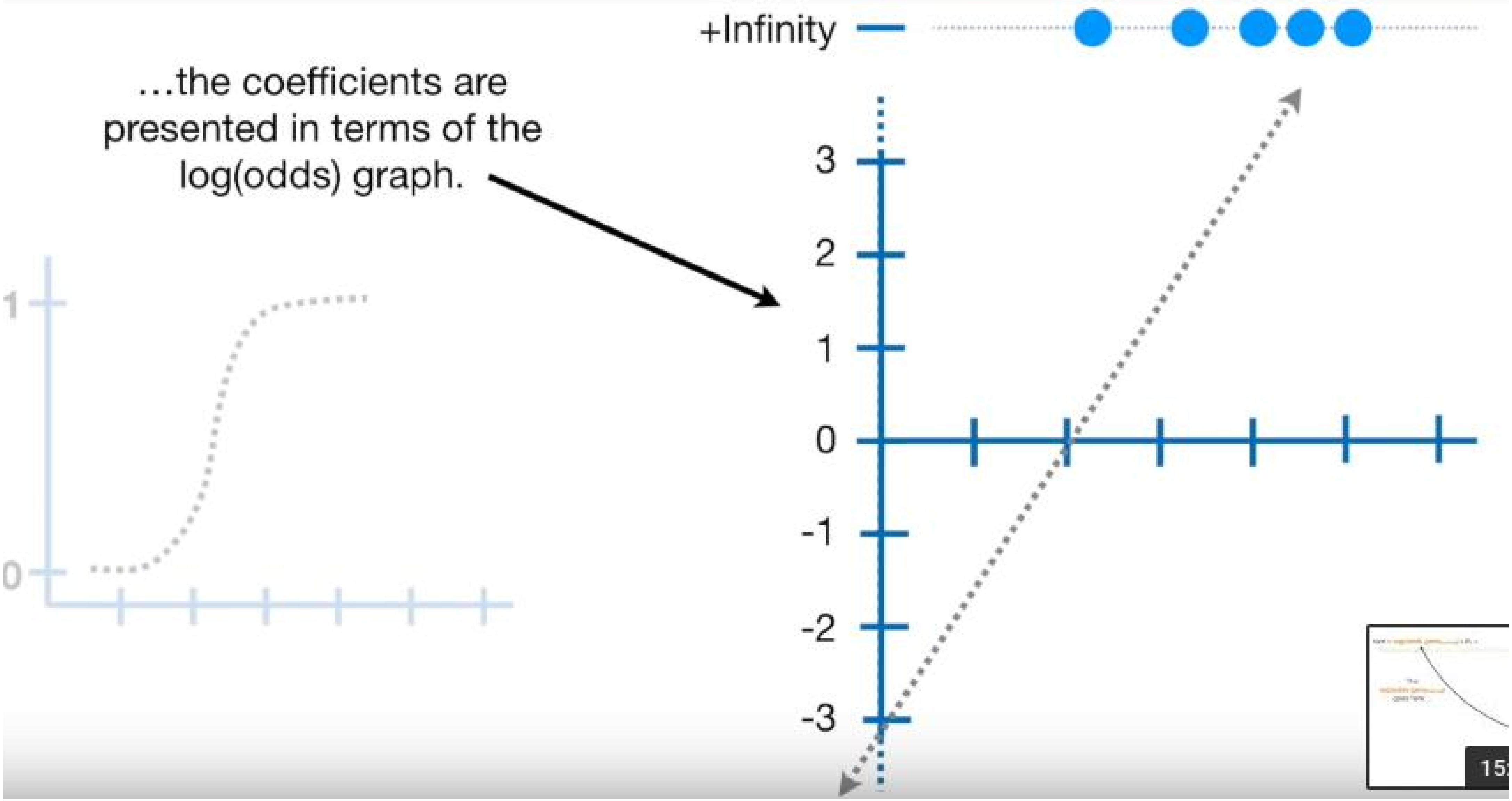

Ultimately we end up with log(odds of obesity) on the new y-axis...
+Infinity
3
2
1
log(odds of obesity) 0
-1
-2
-3



...and the new y-axis transforms the squiggly line into a straight line.
+Infinity
$\log(\frac{p}{1-p}) = \log(\text{odds})$
3
2
1
0
-1
-2
-3
1
0

The important thing to know is that even though the graph with the squiggly line is what we associate with logistic regression…

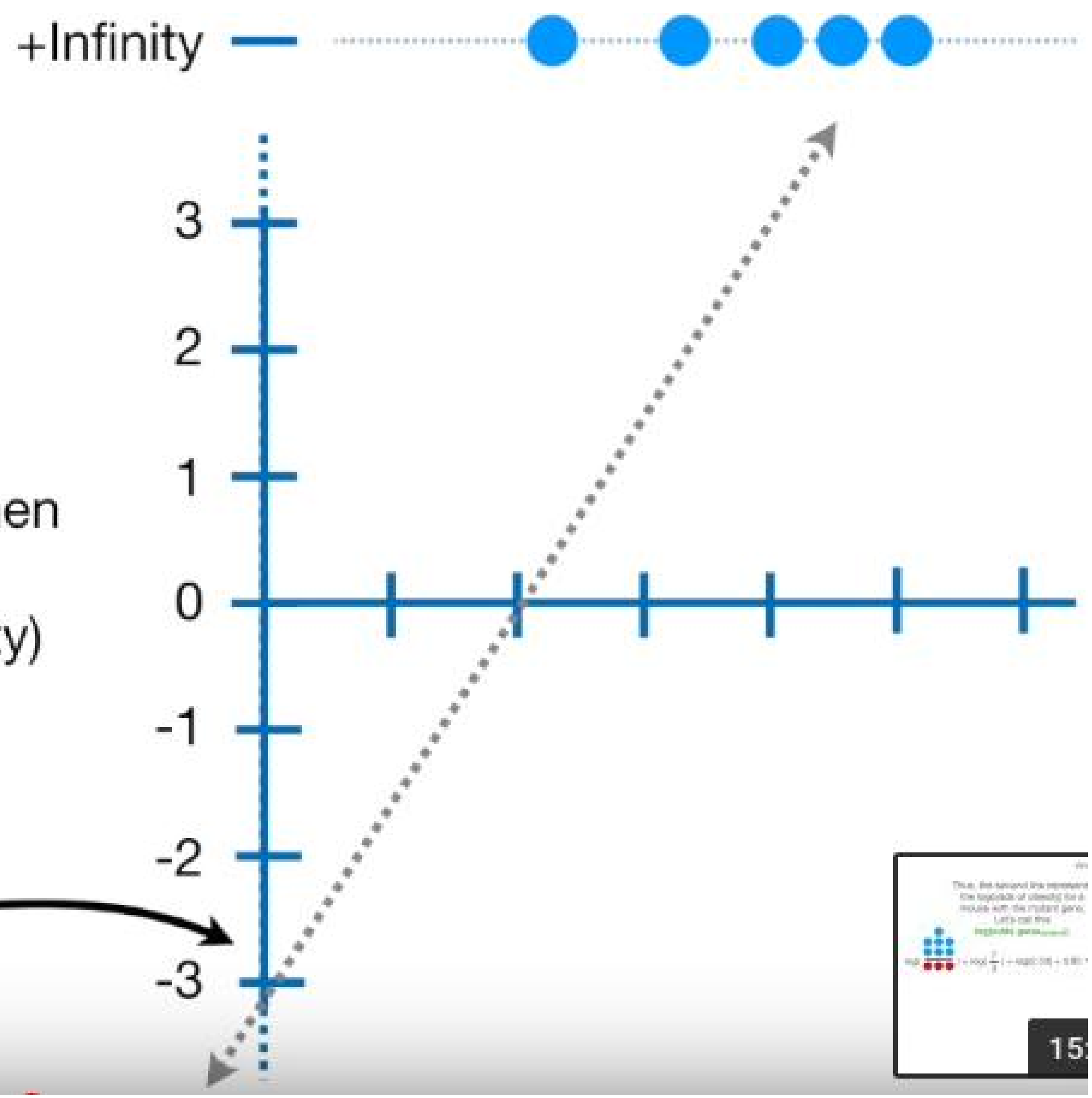


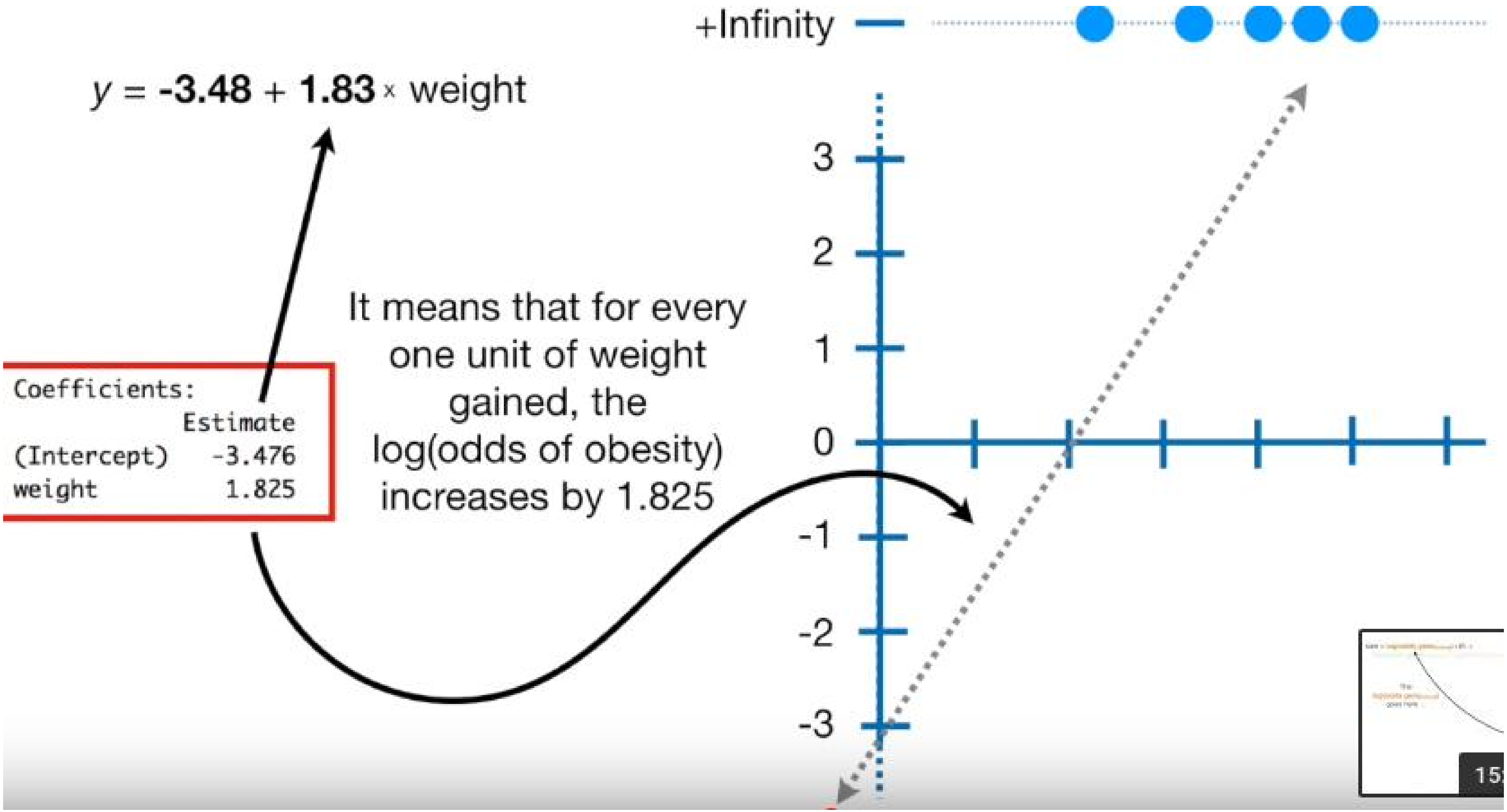

+Infinity
…the coefficients are presented in terms of the log(odds) graph.
1
0
3
2
1
0
-1
-2
-3

$y = \mathbf{-3.48} + \mathbf{1.83} \times \text{weight}$

```
Coefficients:
            Estimate
(Intercept)   -3.476
```

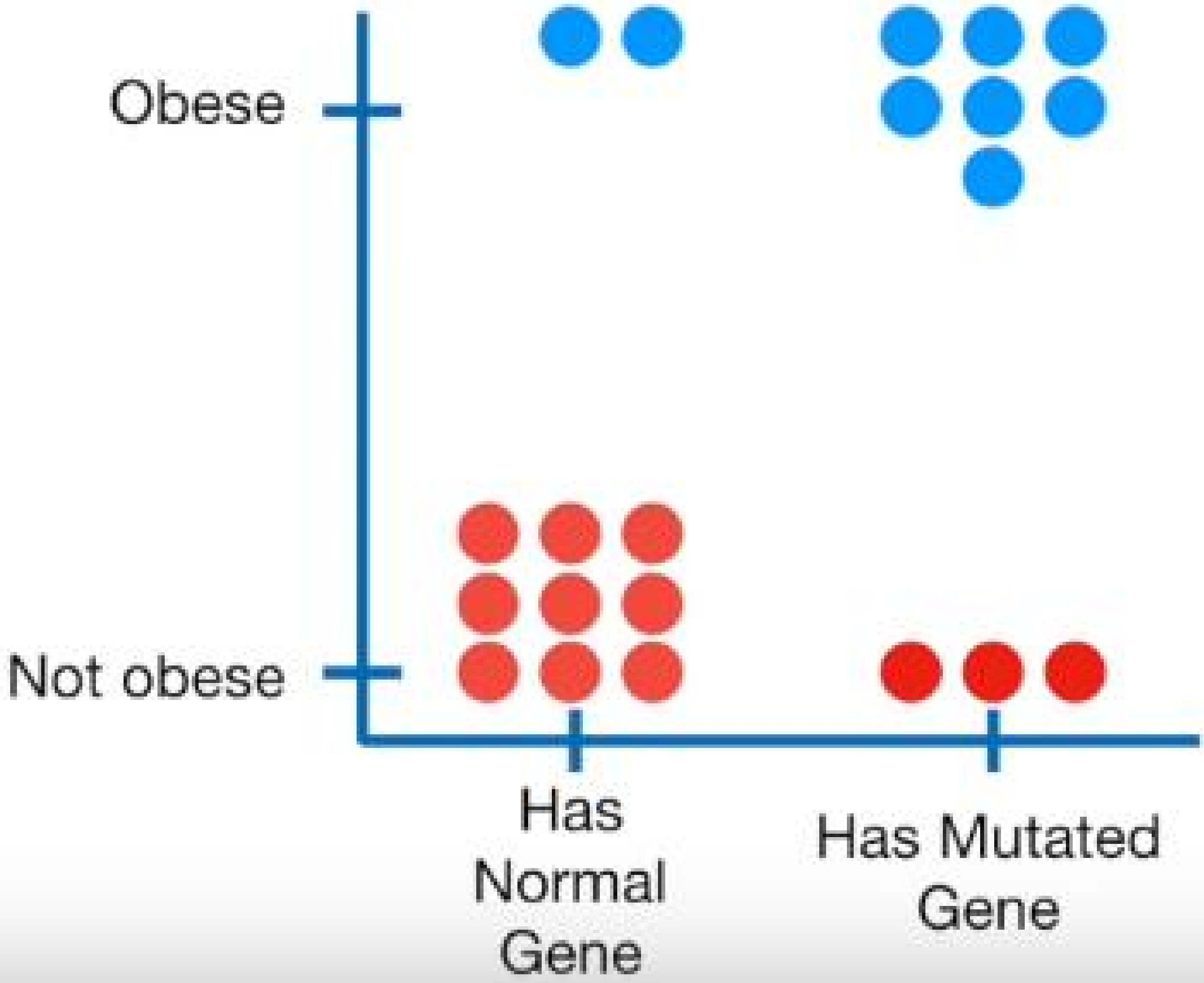
...it means that when weight = 0, the log(odds of obesity) are -3.476.



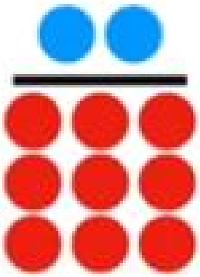


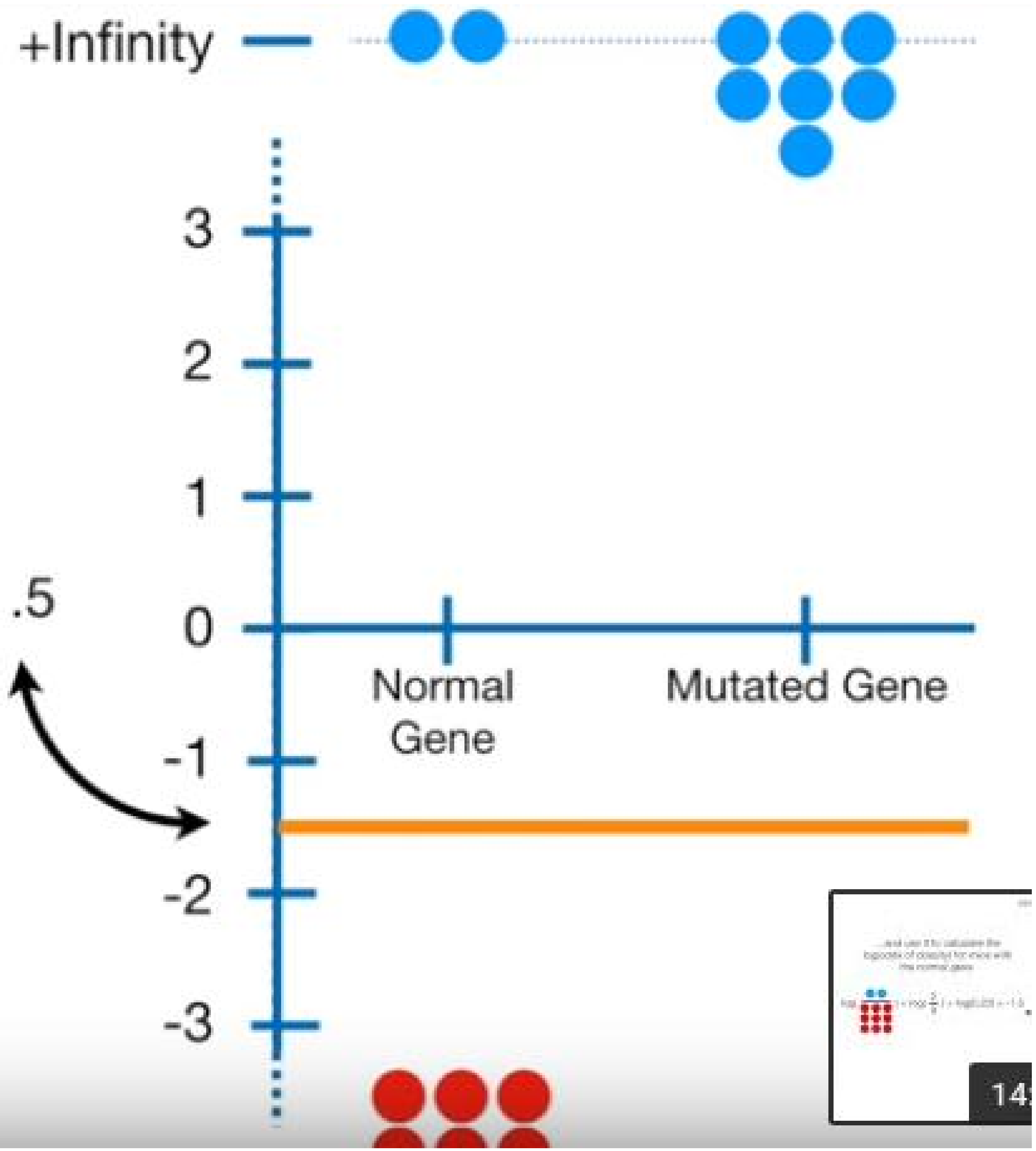
+Infinity
y = -3.48 + 1.83 × weight
Coefficients:
Estimate
(Intercept) -3.476
weight 1.825
It means that for every one unit of weight gained, the log(odds of obesity) increases by 1.825
3
2
1
0
-1
-2
-3

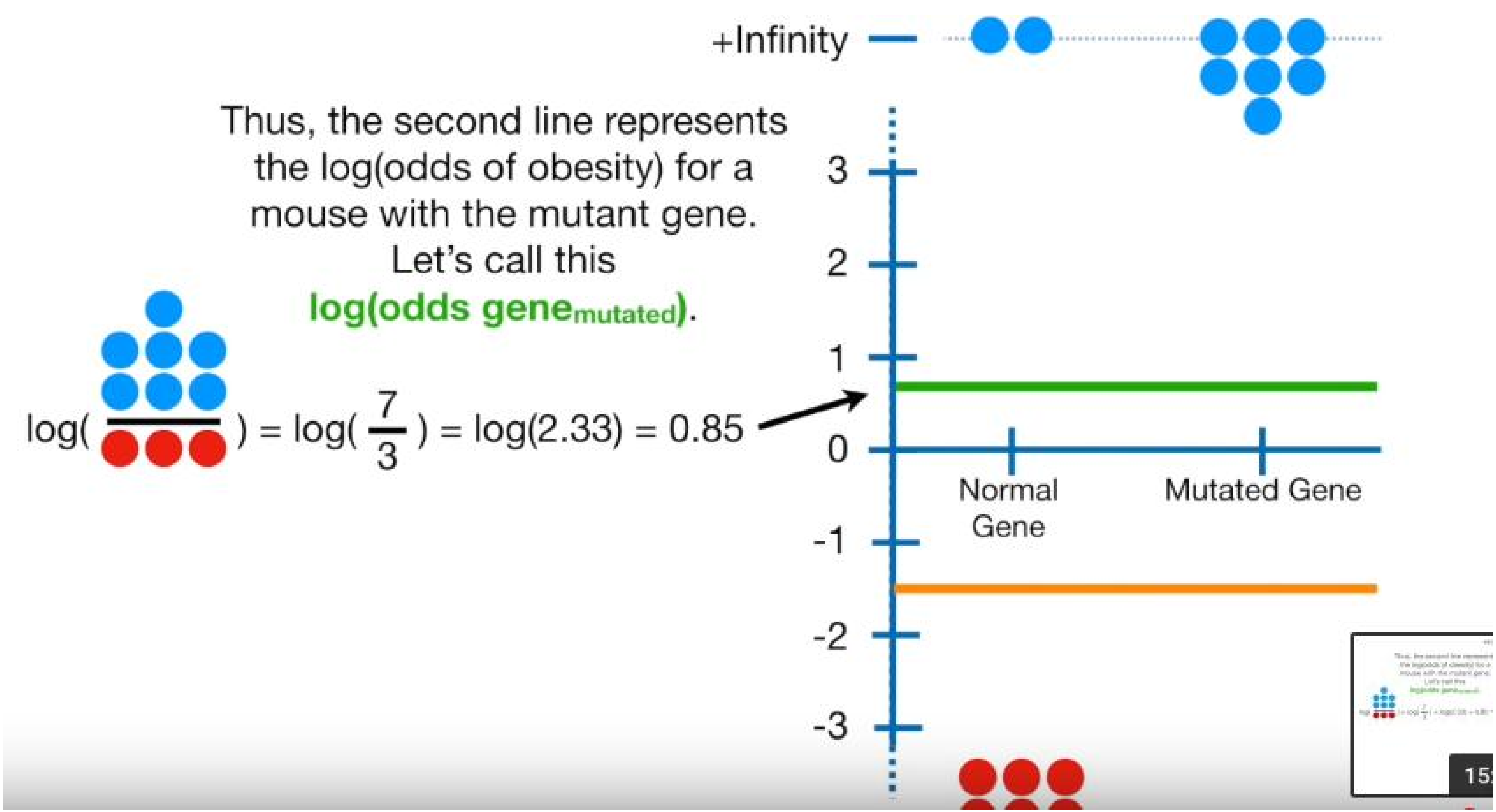
Now let's talk about logistic regression coefficients in the context of testing if a discrete variable like "whether or not a mouse has a mutated gene" is related to obesity.



Thus, the first line represents the log(odds of obesity) for the mice with the normal gene. Let's call this the **log(odds $gene_{normal}$)**.

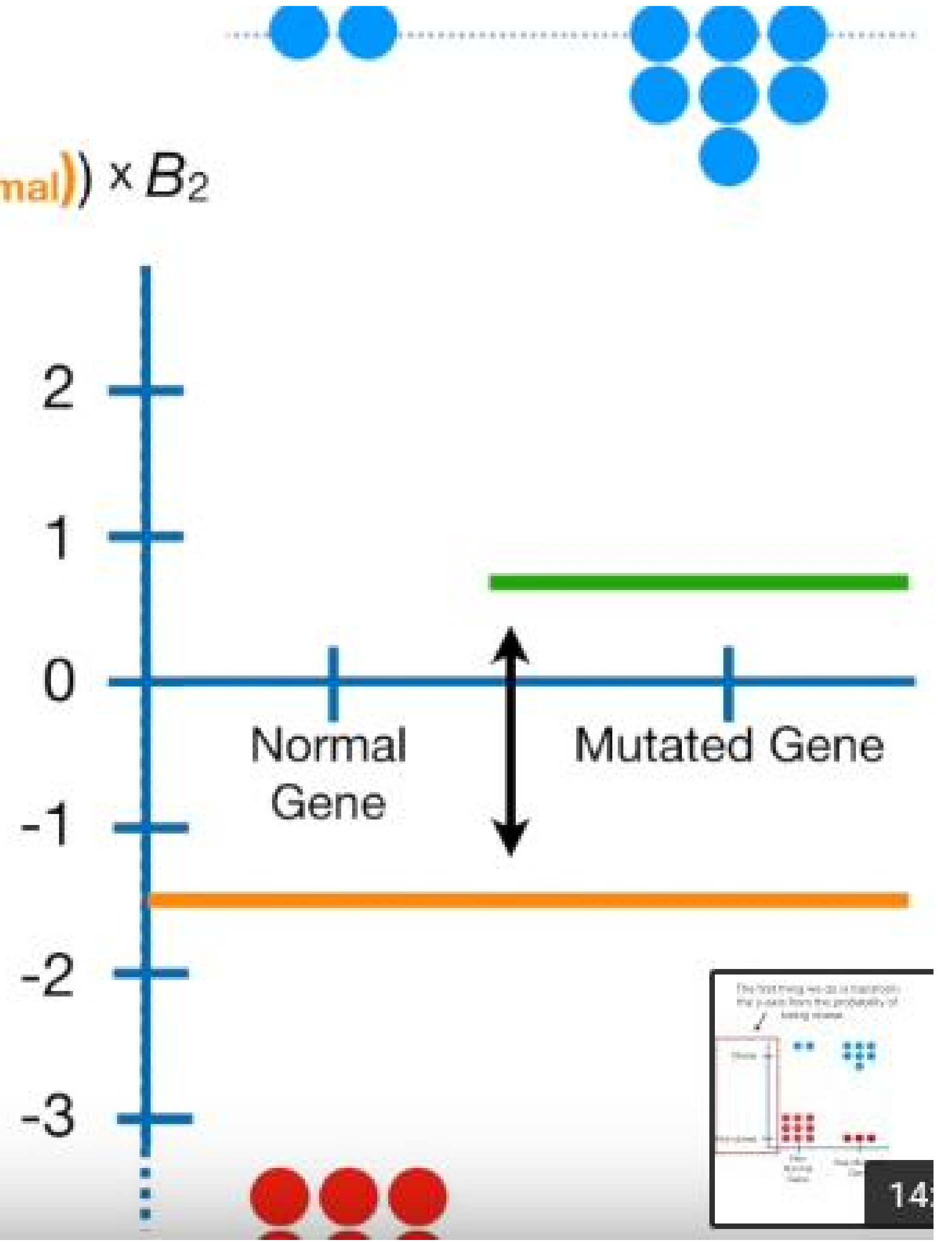
log() $= \log(\frac{2}{9}) = \log(0.22) = -1.5$




+Infinity
Thus, the second line represents
the log(odds of obesity) for a
mouse with the mutant gene.
Let's call this
log(odds gene$_{mutated}$).
log( ) = log(7/3) = log(2.33) = 0.85
3
2
1
0
-1
-2
-3
Normal
Gene
Mutated Gene

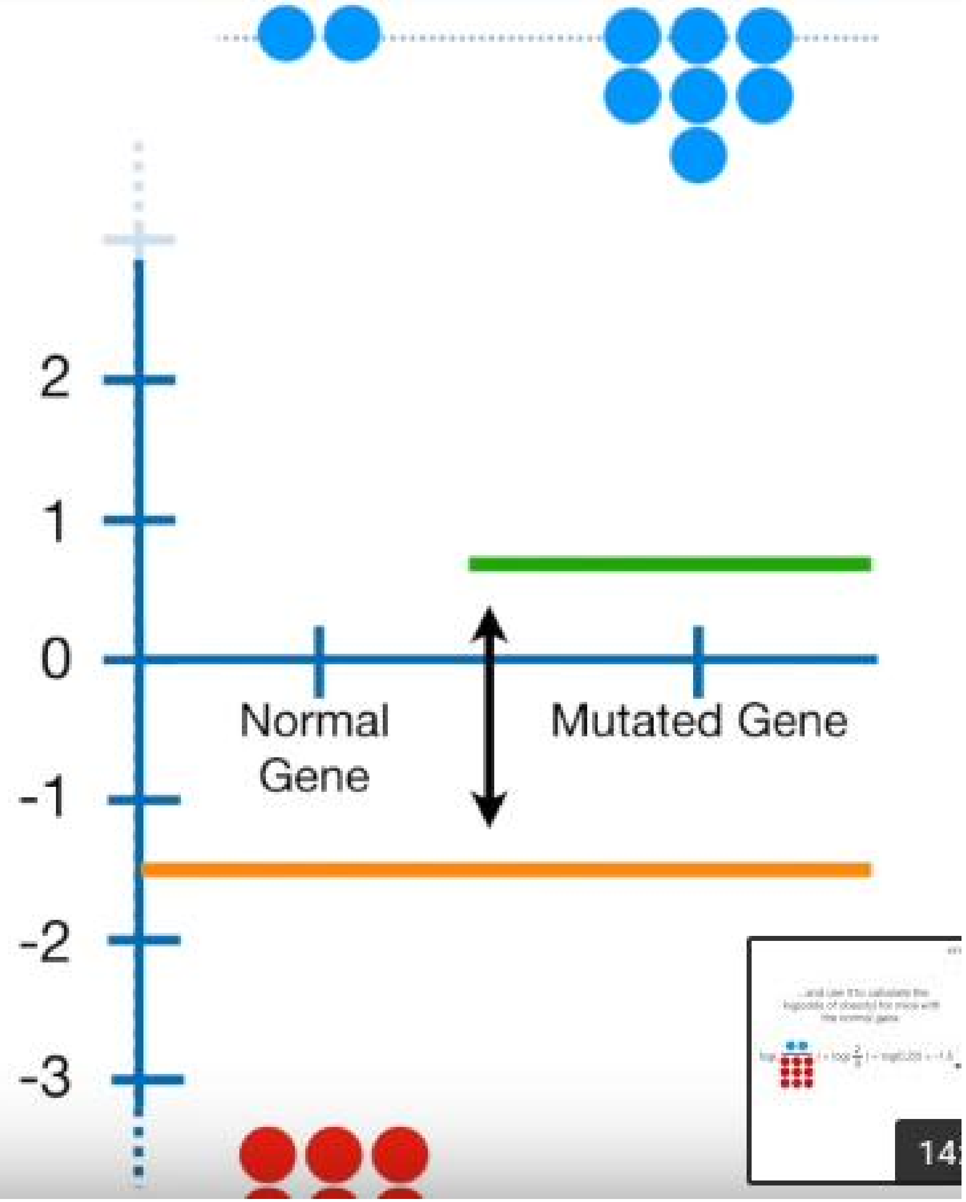
$$\text{size} = \log(\text{odds gene}_{normal}) \times B_1 + (\log(\text{odds gene}_{mutated}) - \log(\text{odds gene}_{normal})) \times B_2$$

…and since subtracting one log from another…



$$\text{size} = \log(\text{odds gene}_{\text{normal}}) \times B_1 + \log\left(\frac{\text{odds gene}_{\text{mutated}}}{\text{odds gene}_{\text{normal}}}\right) \times B_2$$

...can be converted into division, this term is a **log(odds ratio)**.



$$\text{size} = \log(\text{odds gene}_{\text{normal}}) \times B_1 + \log\left(\frac{\text{odds gene}_{\text{mutated}}}{\text{odds gene}_{\text{normal}}}\right) \times B_2$$

...can be converted into division, this term is a **log(odds ratio)**.

It tells us, on a log scale, how much having the mutated gene increases (or decreases) the odds of a mouse being obese.

$$\text{size} = \log(2/9) \times B_1 + \log\left(\frac{7/3}{2/9}\right) \times B_2$$

$$\text{size} = -1.5 \times B_1 + 2.35 \times B_2$$

...and those are what you get when you do logistic regression.

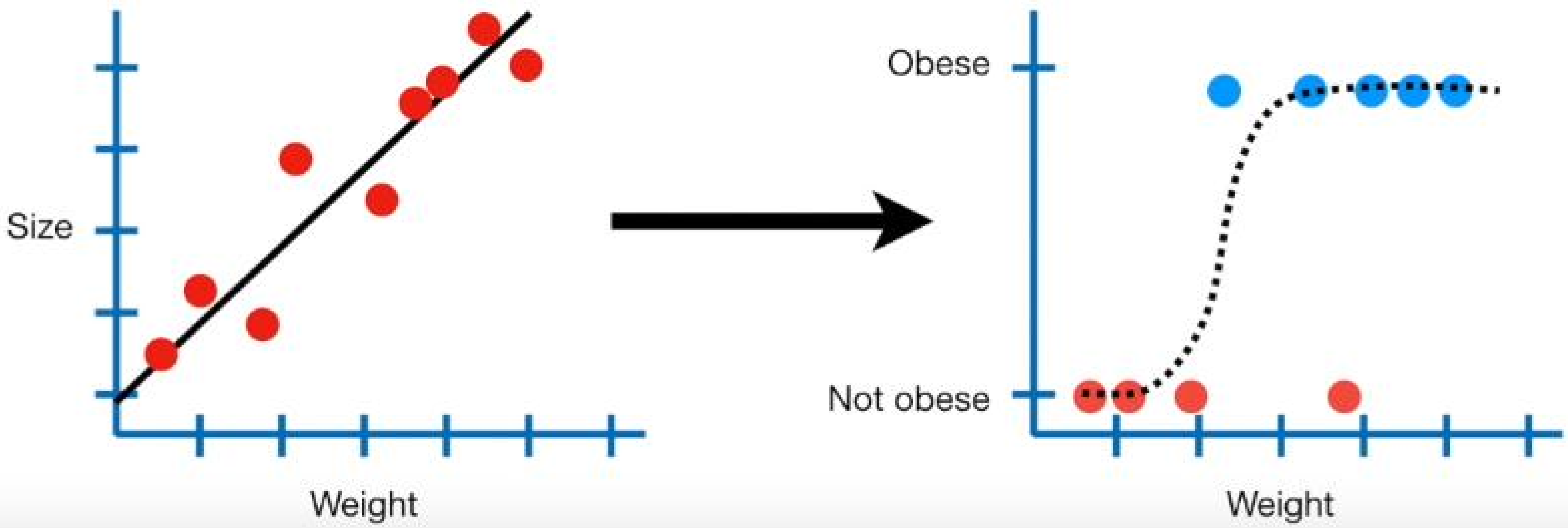
```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.5041     0.7817  -1.924   0.0544
geneMutant    2.3514     1.0427   2.255   0.0241
```

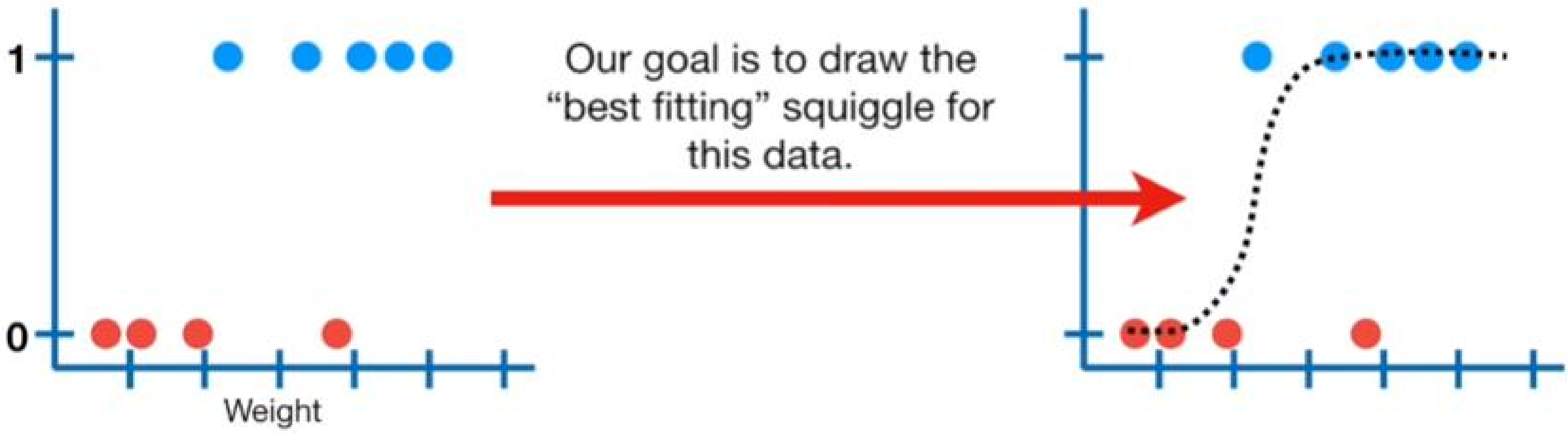
$$\text{size} = \mathbf{-1.5} \times B_1 + \mathbf{2.35} \times B_2$$

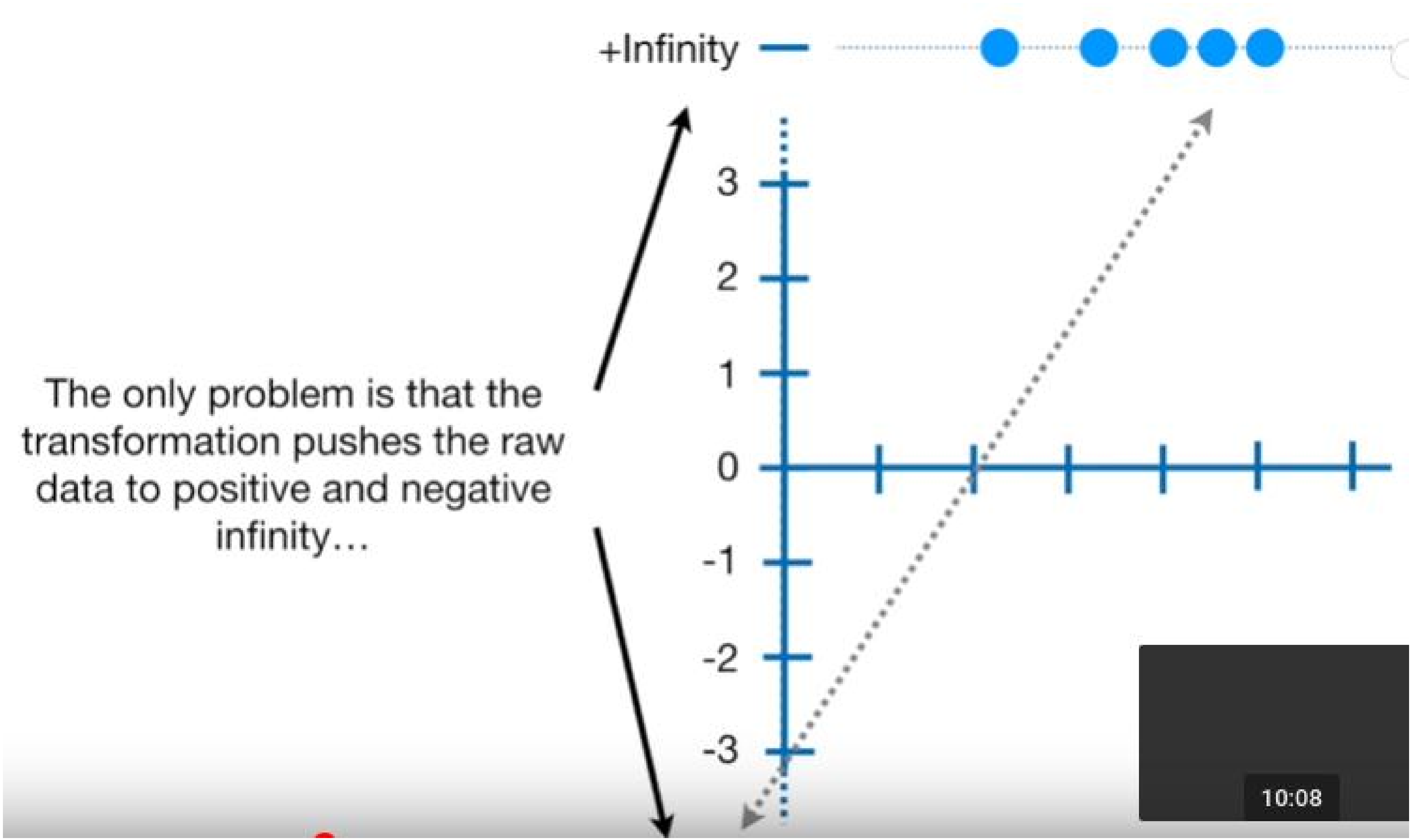
```
Coefficients:
            Estimate
(Intercept)  -1.5041
geneMutant    2.3514
```

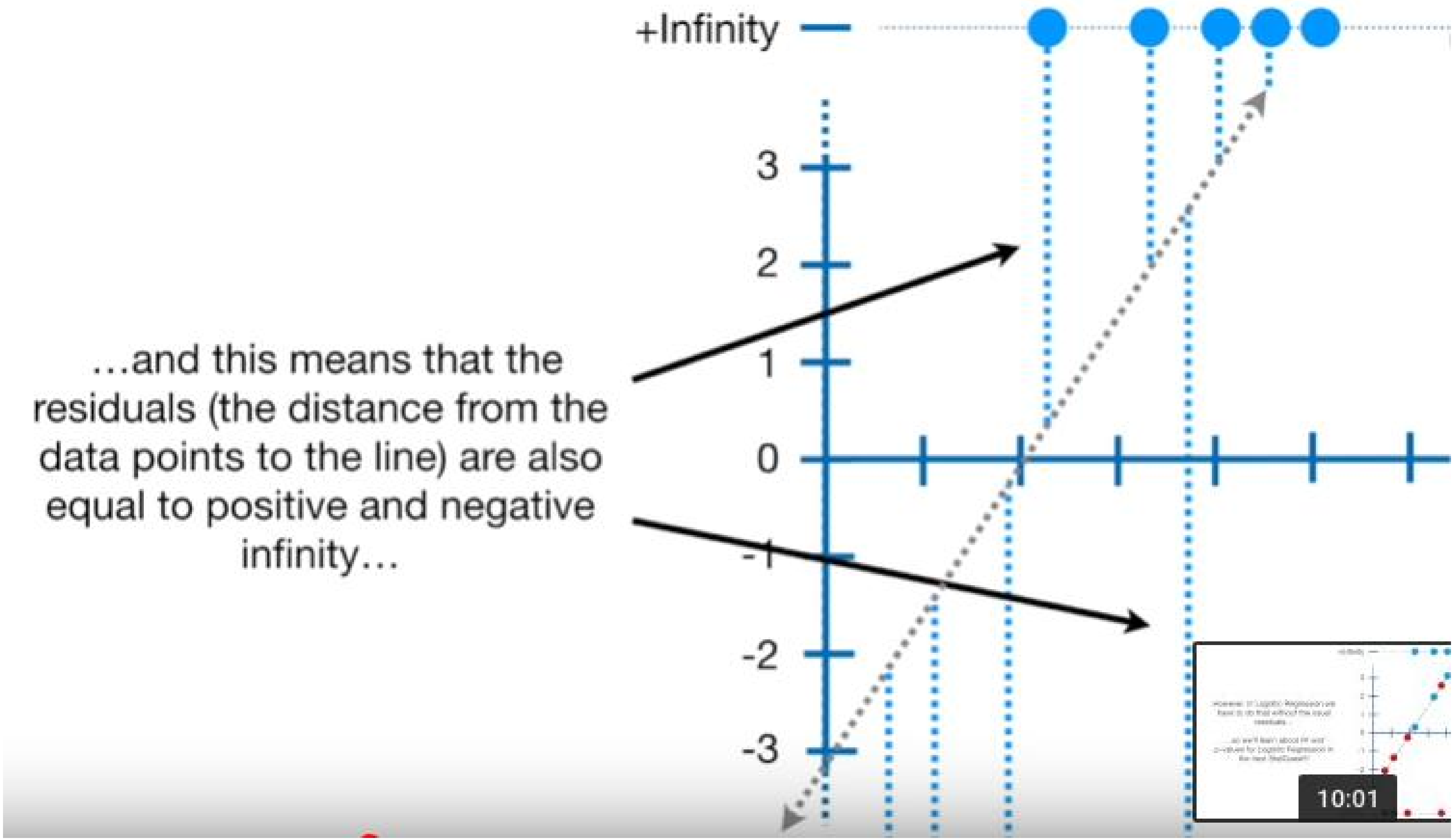
…and the “geneMutant” term is the **log(odds ratio)** that tells you, on a log scale, how much having the mutated gene increases or decreases the odds of being obese.

Now we have seen how some of the linear model concepts for regression apply to logistic regression…

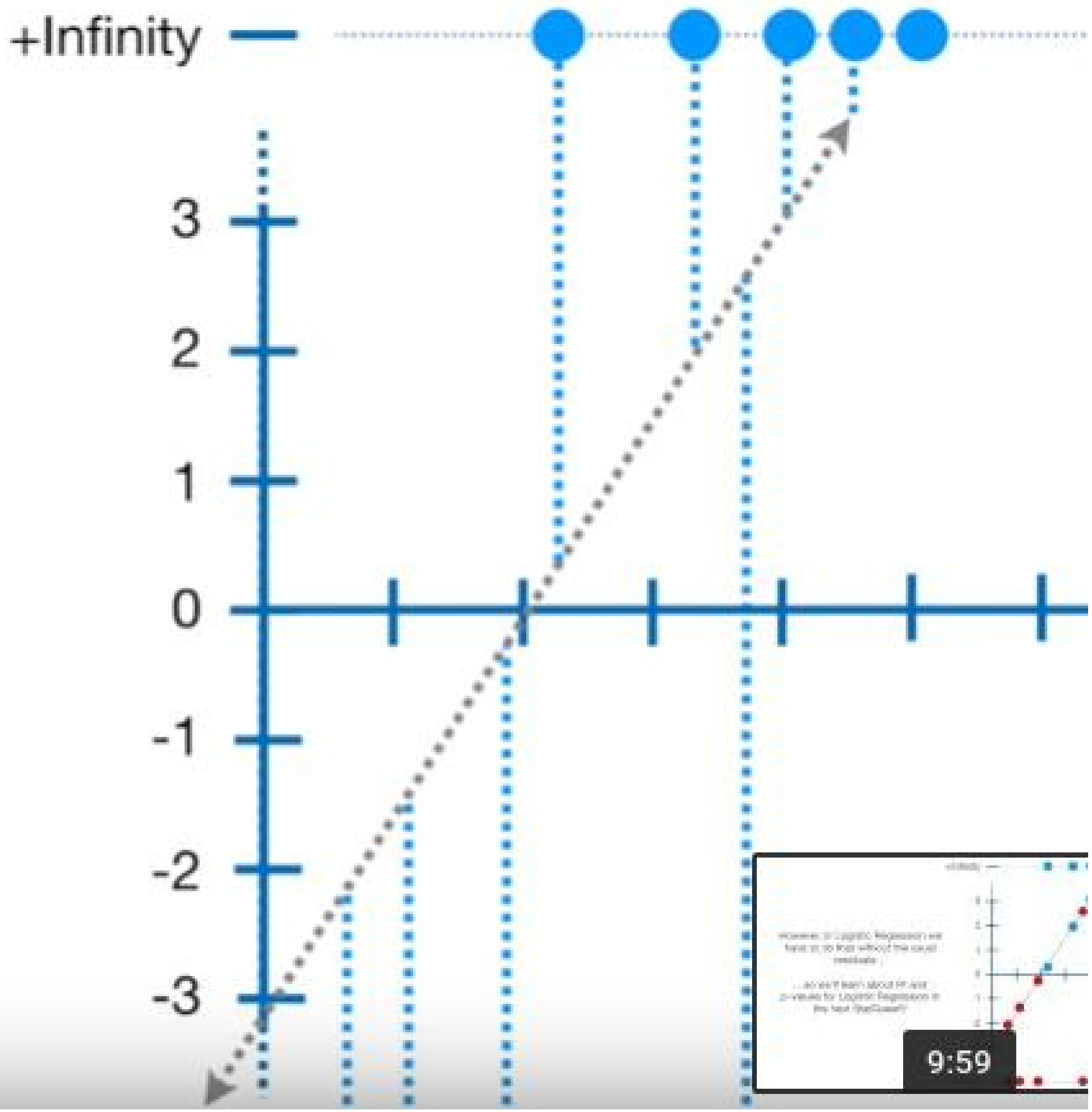



1
0
Weight
Our goal is to draw the
"best fitting" squiggle for
this data.

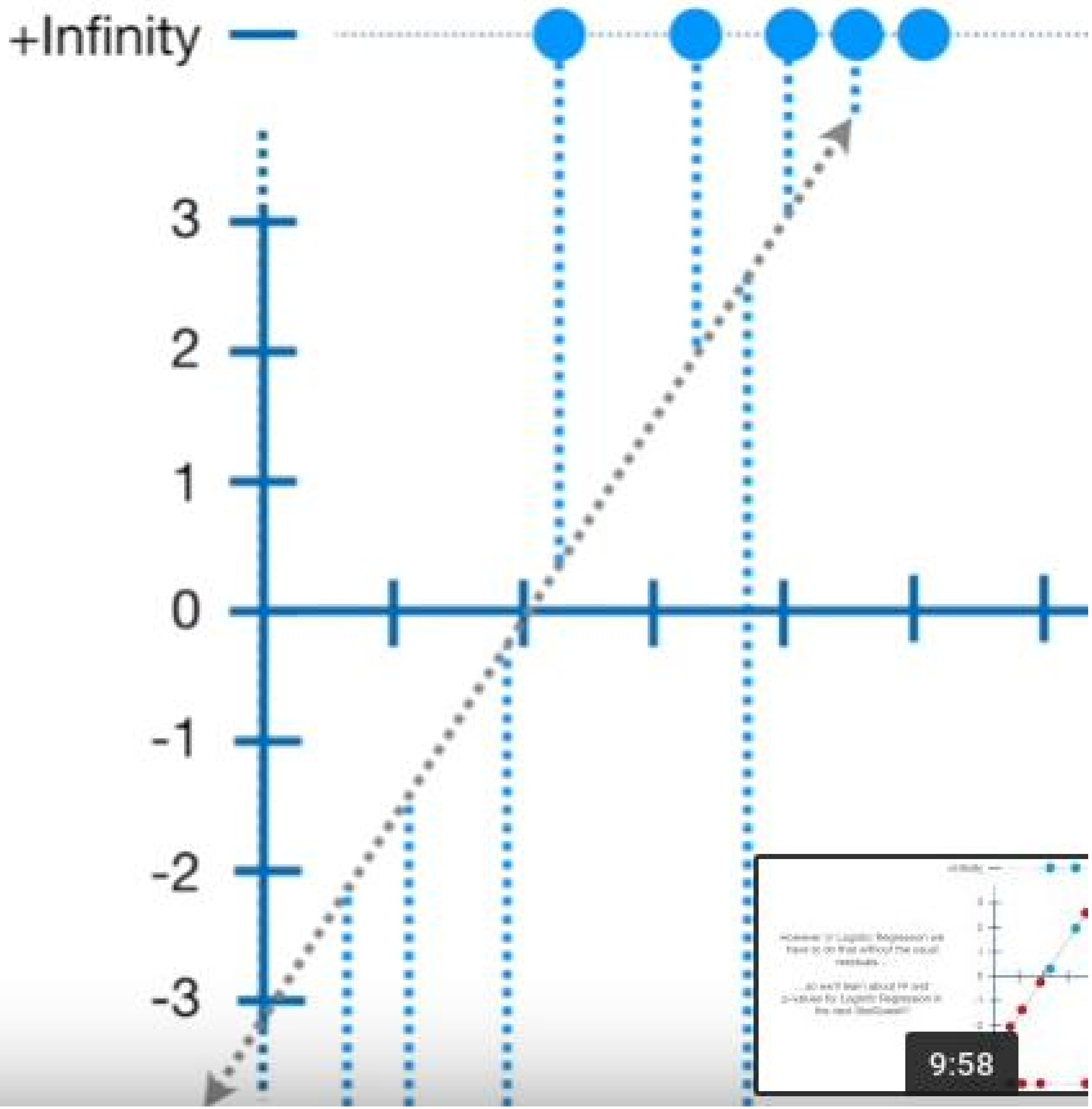

+Infinity
3
2
1
0
-1
-2
-3
The only problem is that the transformation pushes the raw data to positive and negative infinity…
10:08


+Infinity
3
2
1
0
-1
-2
-3
...and this means that the residuals (the distance from the data points to the line) are also equal to positive and negative infinity...
10:01

…and this means we can't use least-squares to find the best fitting line.

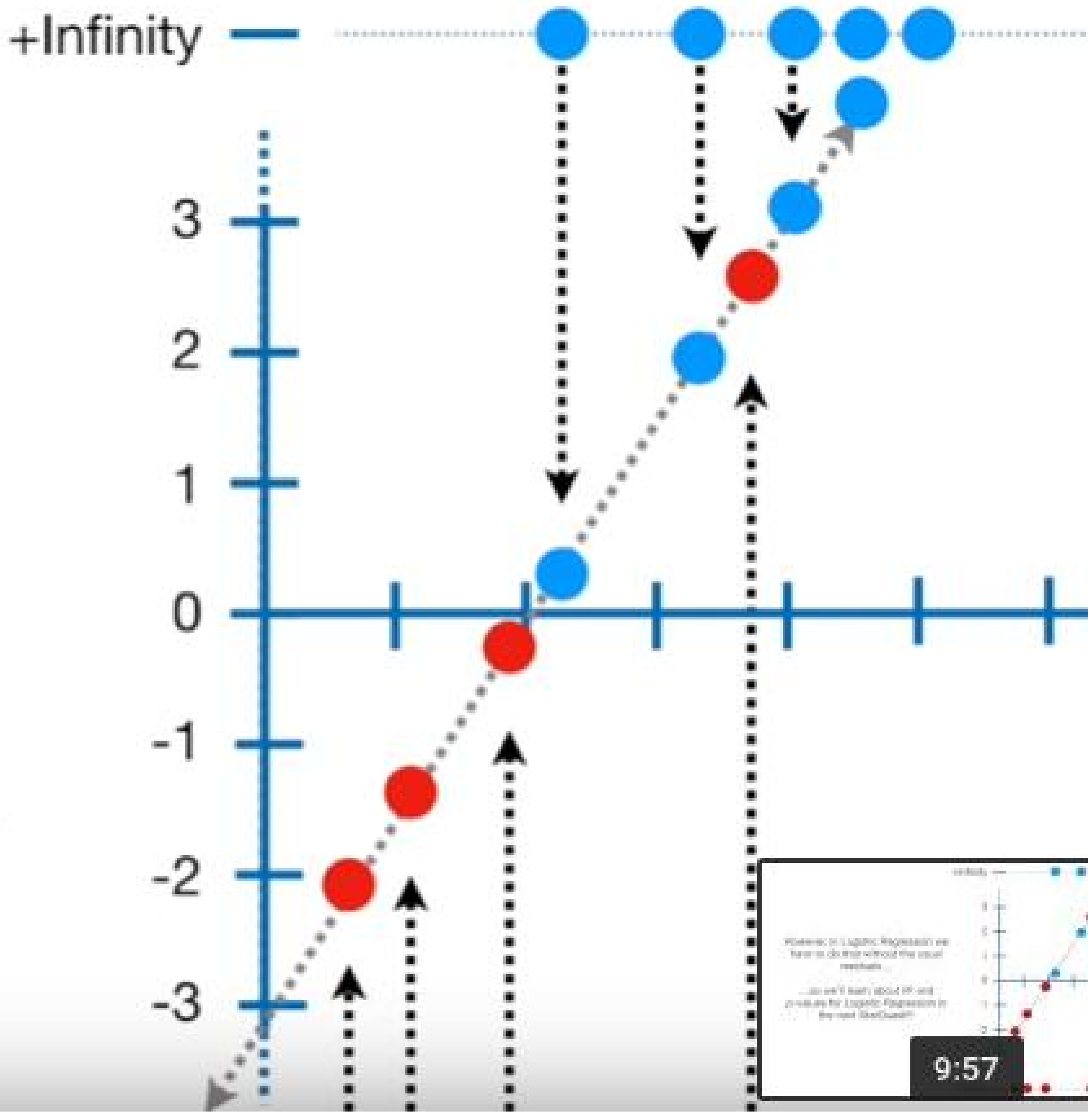


Instead, we use maximum likelihood…

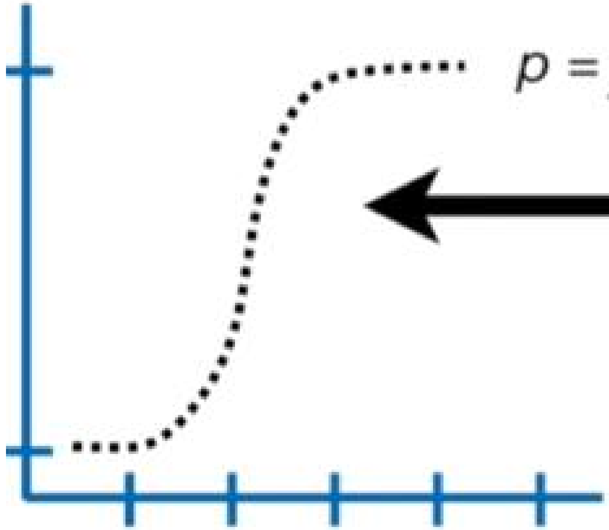


The first thing we do is project the original data points onto the candidate line.

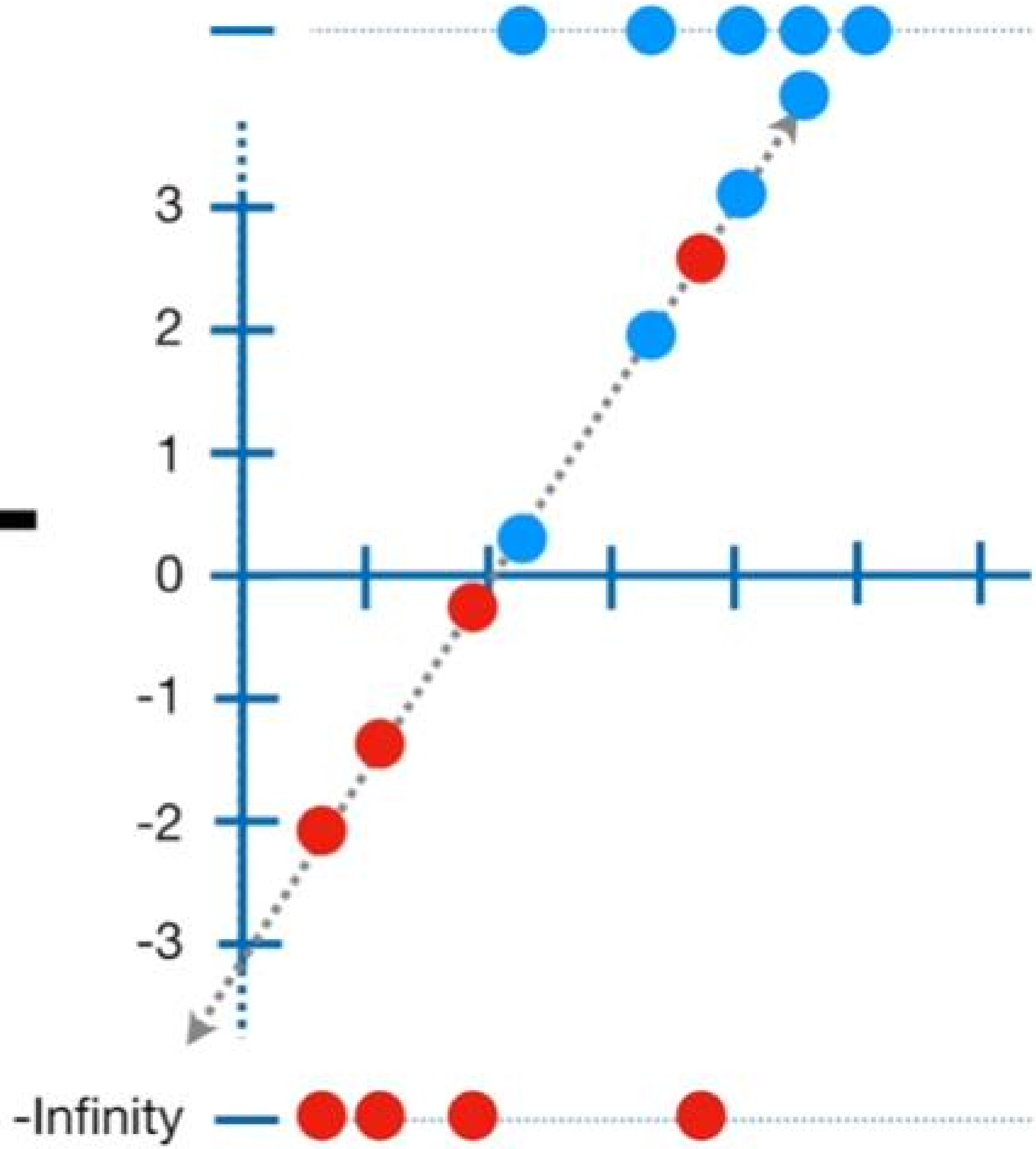
This gives each sample a candidate log(odds) value.

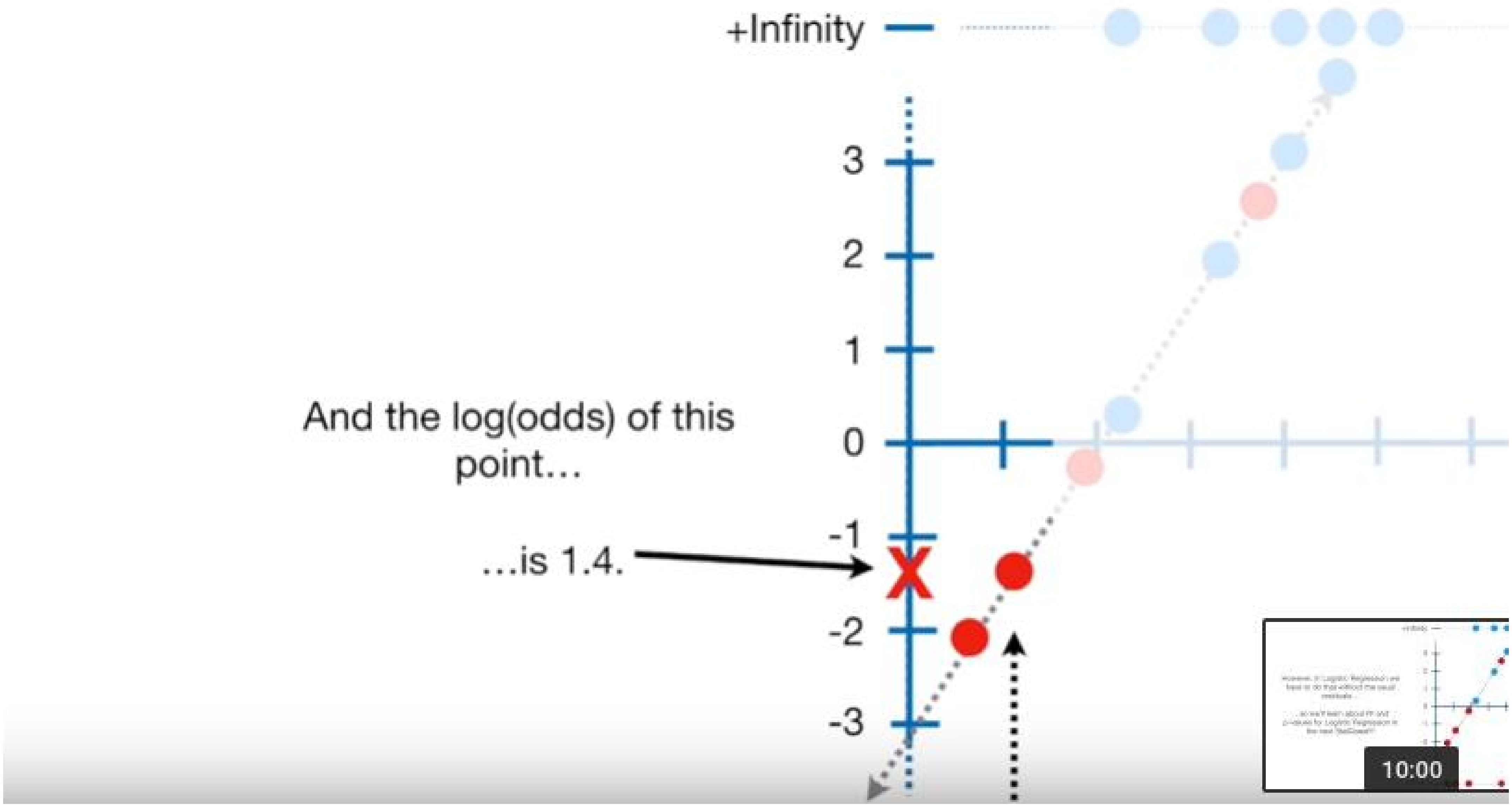


Then we transform the candidate log(odds) to candidate probabilities using this fancy looking formula…



$$p = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$$

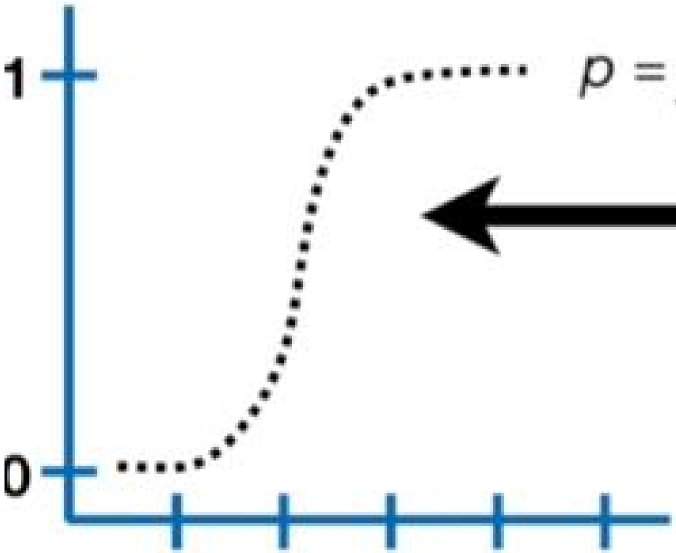


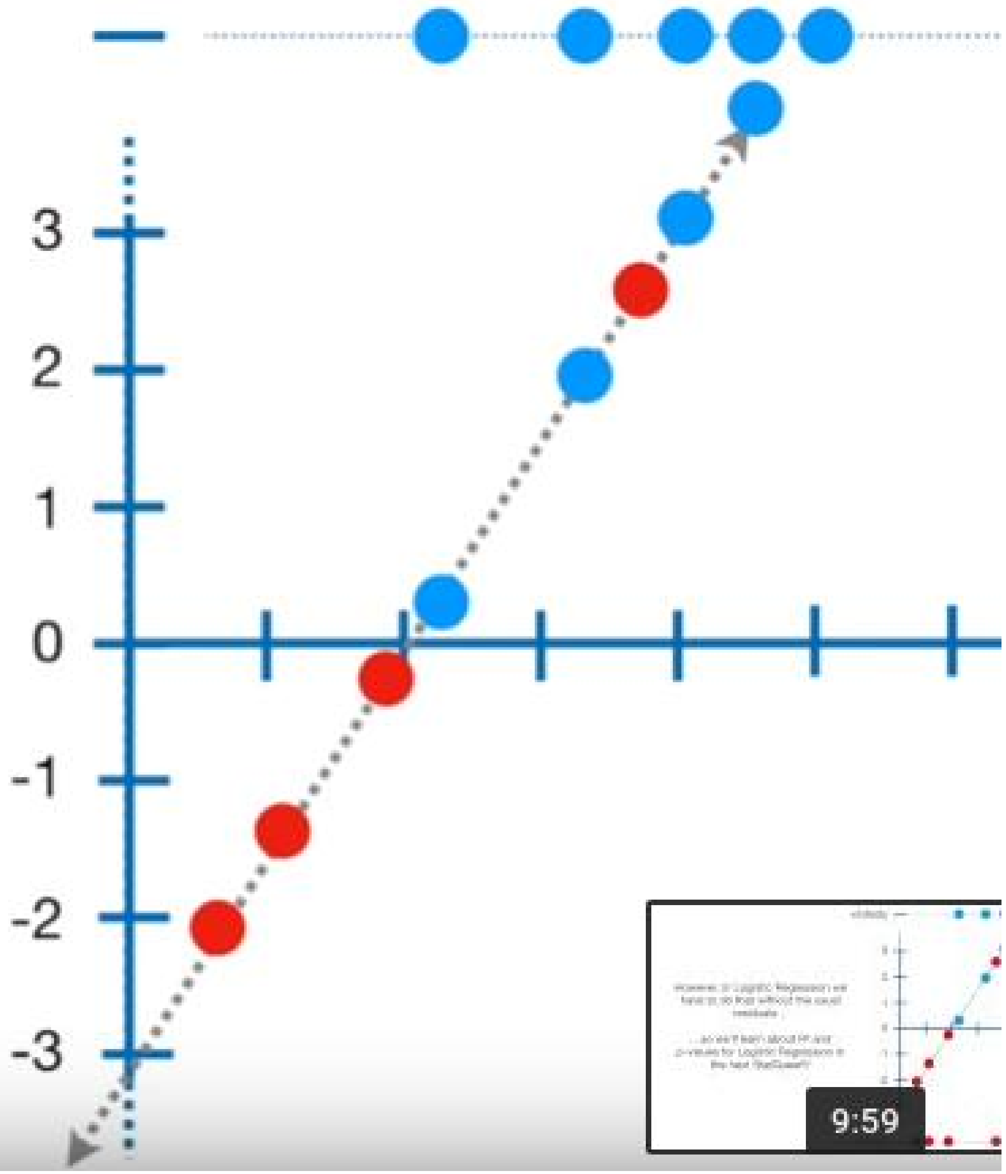


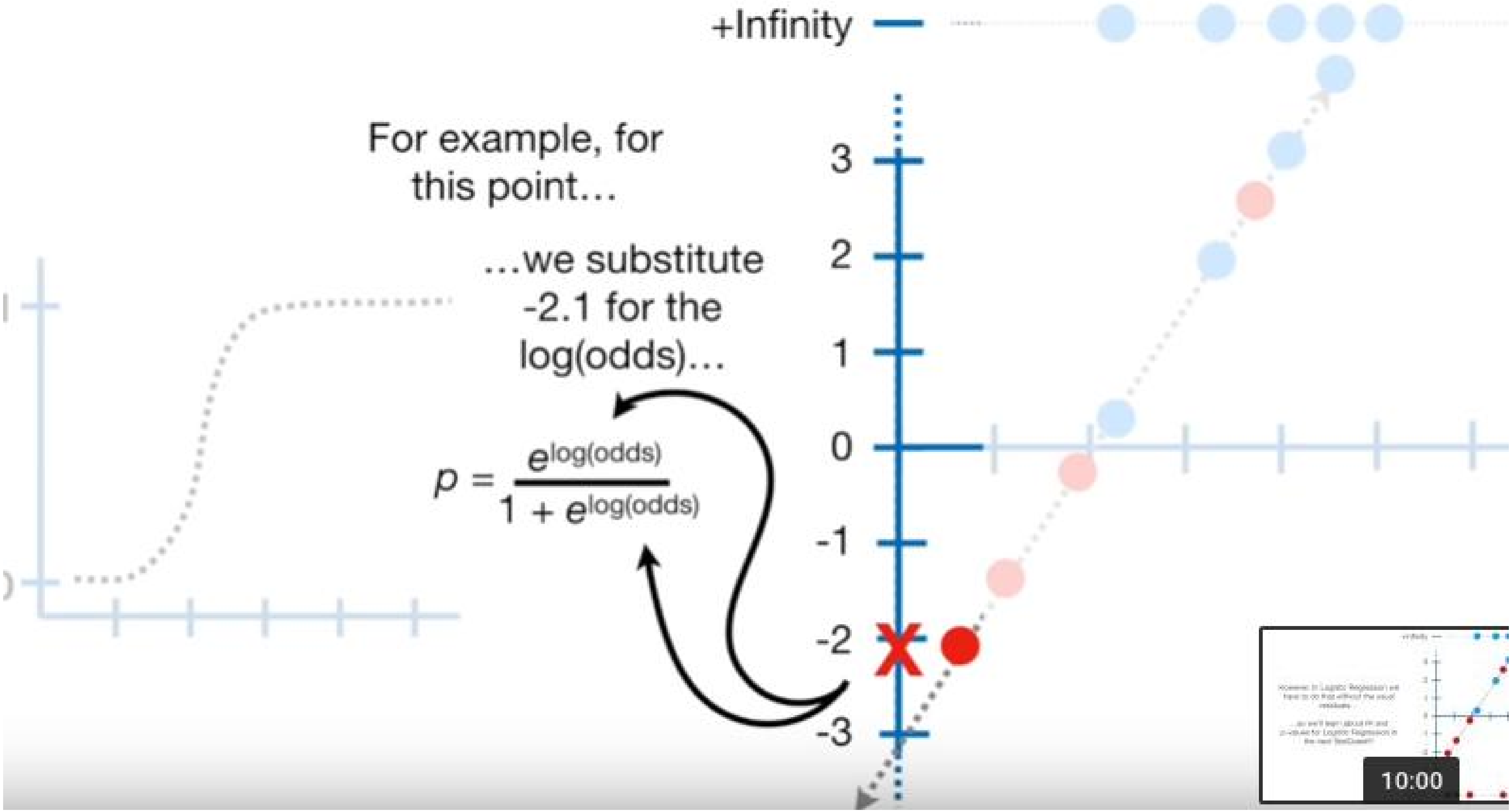
+Infinity
3
2
1
0
-1
-2
-3
And the log(odds) of this point…
…is 1.4.
10:00

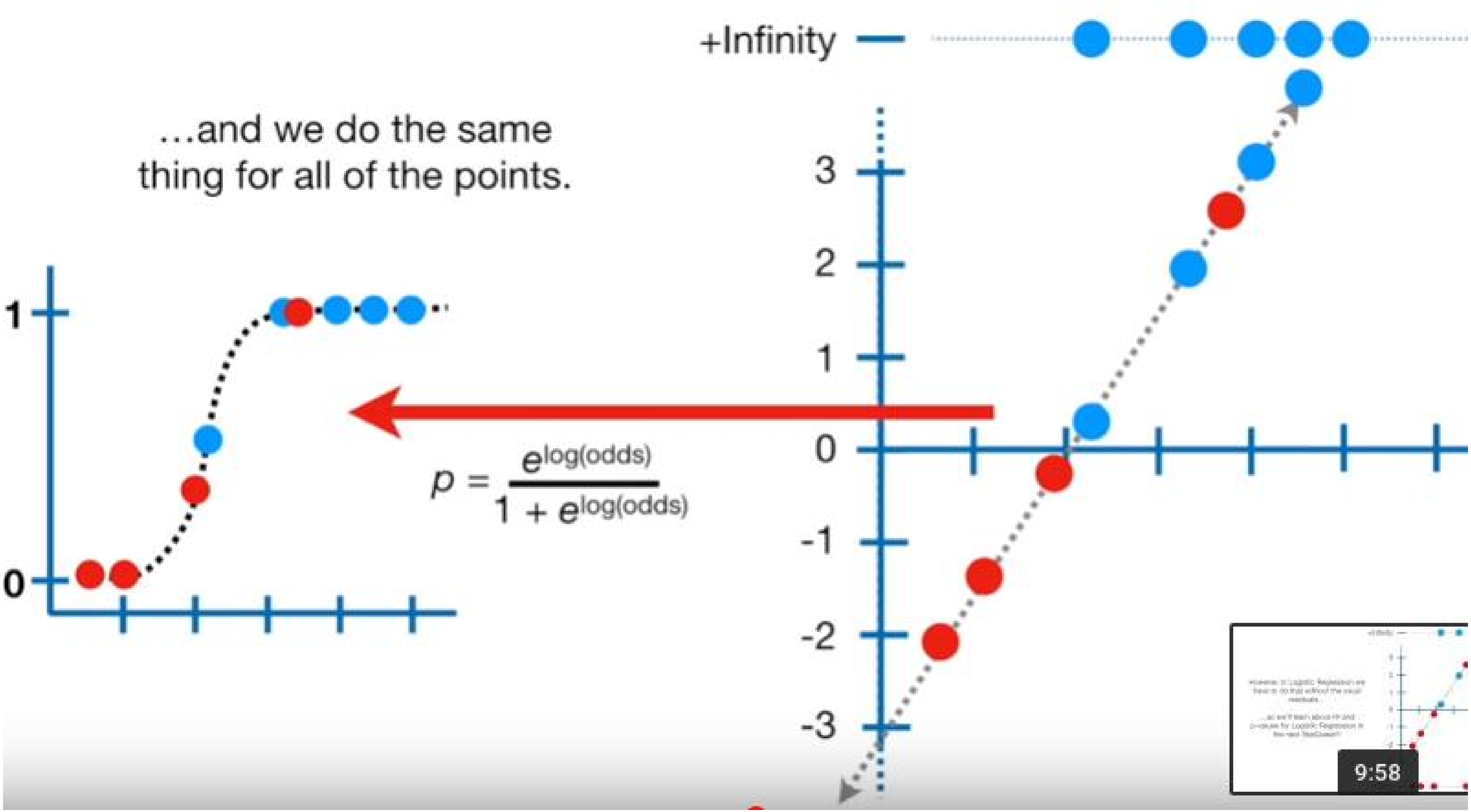
Then we transform the candidate log(odds) to candidate probabilities using this fancy looking formula…

$$p = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$$



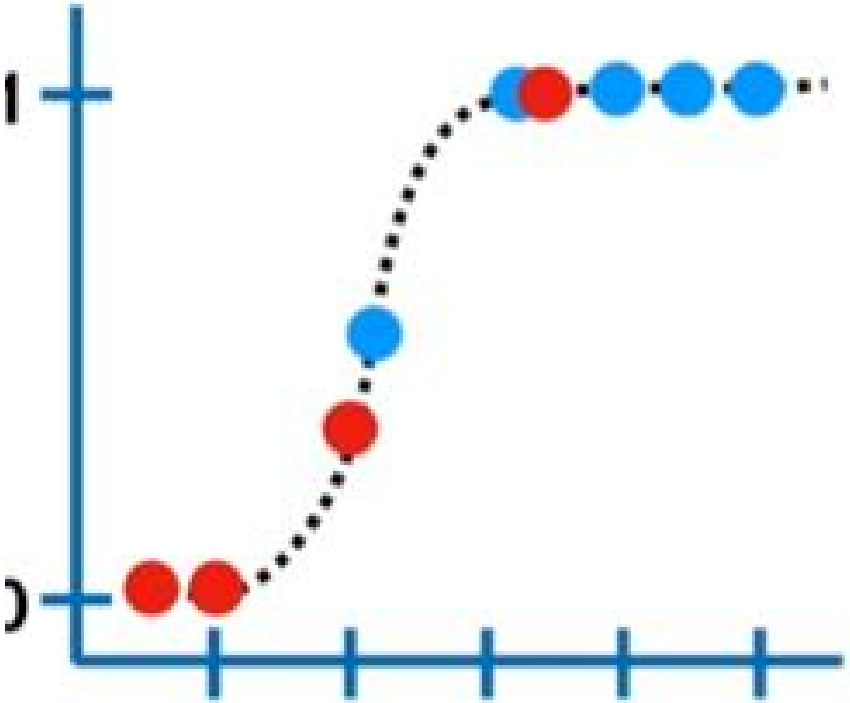


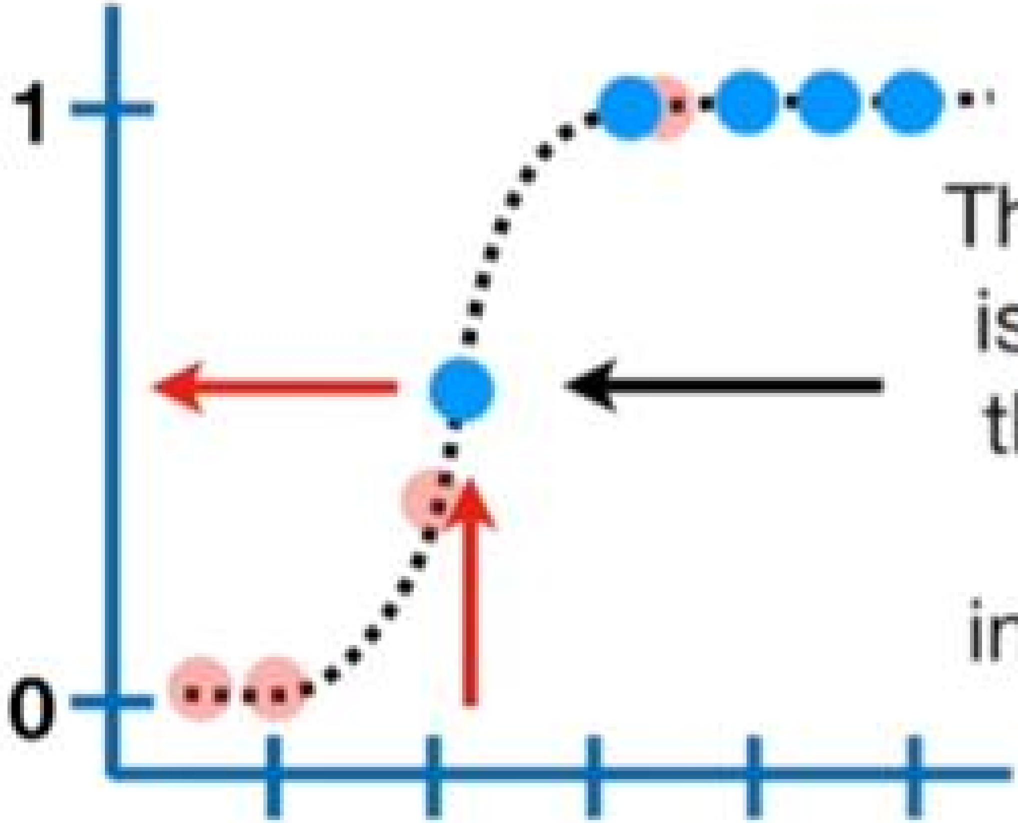

+Infinity
For example, for this point…
…we substitute -2.1 for the log(odds)…
$p = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$
3
2
1
0
-1
-2
-3
10:00



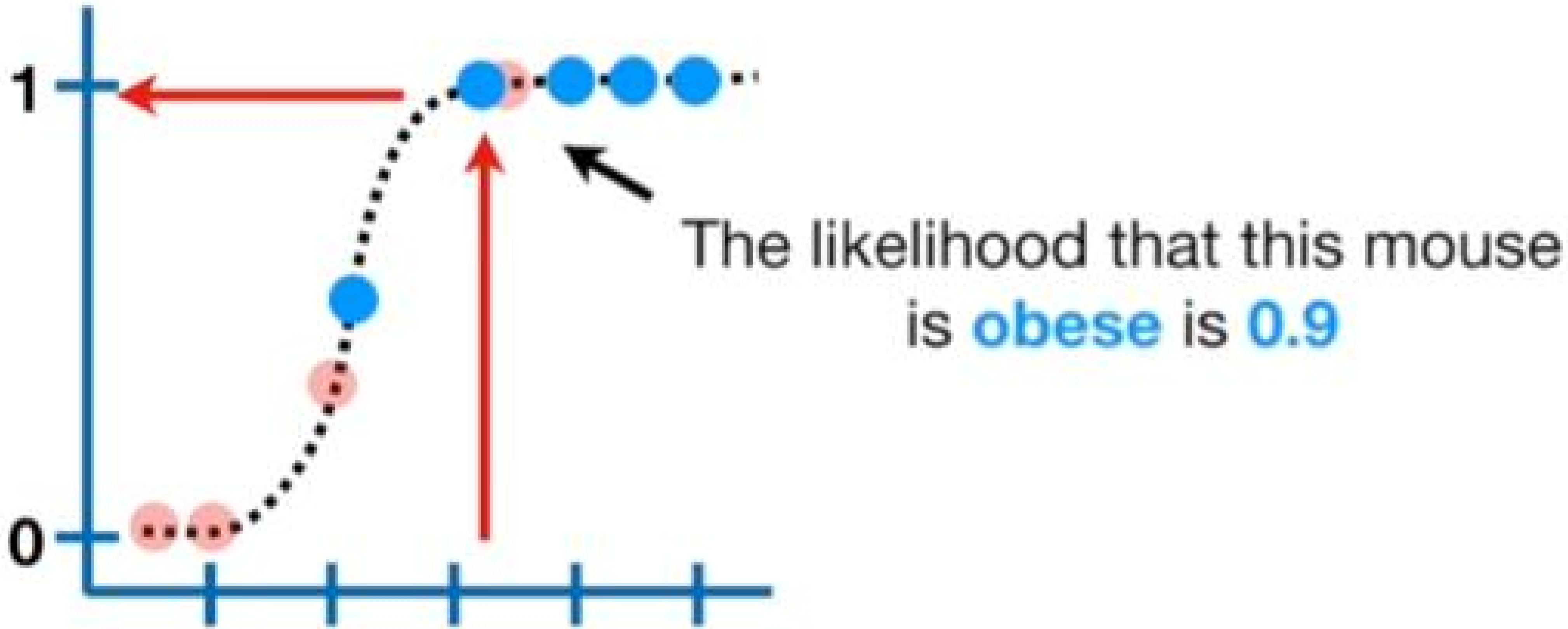
...and we do the same thing for all of the points.
+Infinity
3
2
1
0
-1
-2
-3
1
0
$p = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$
9:58

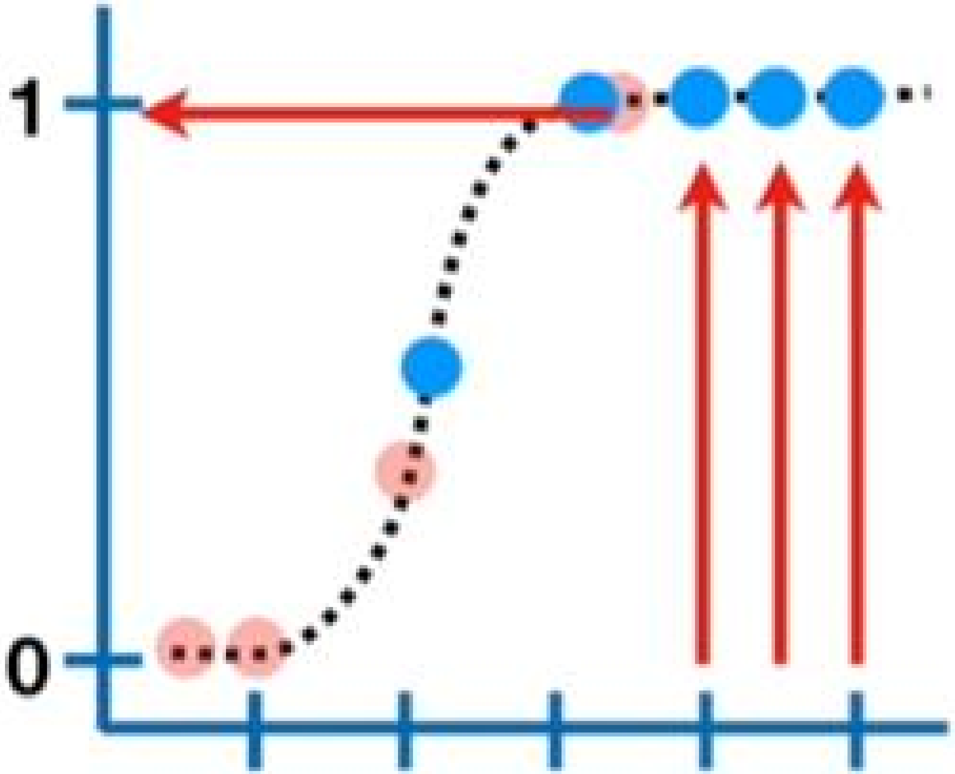
Now we use the observed status (obese or not obese) to calculate their likelihood given the shape of the squiggly line.





The likelihood that this mouse is obese, given the shape of the squiggle, is the value on the y-axis where point intersects the squiggle, 0.49.

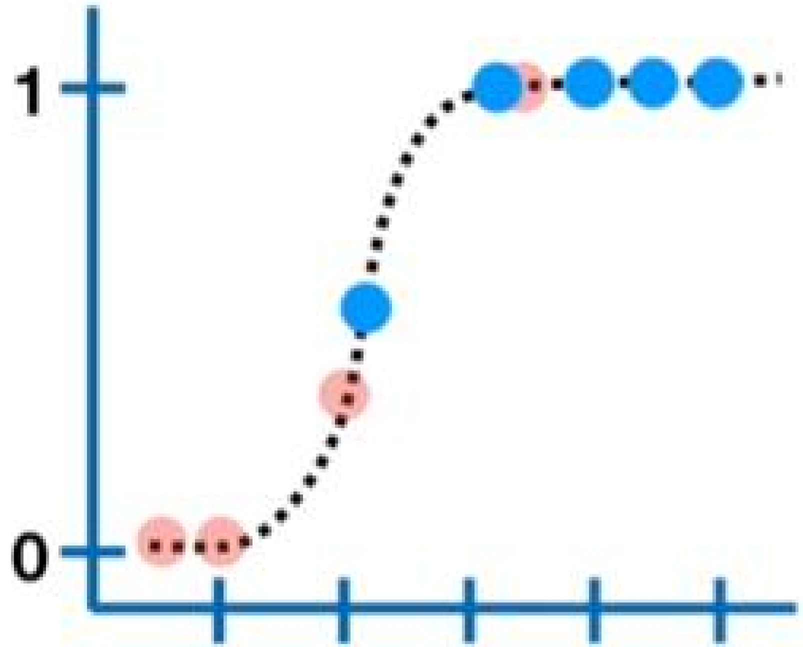

1
0
The likelihood that this mouse
is obese is 0.9



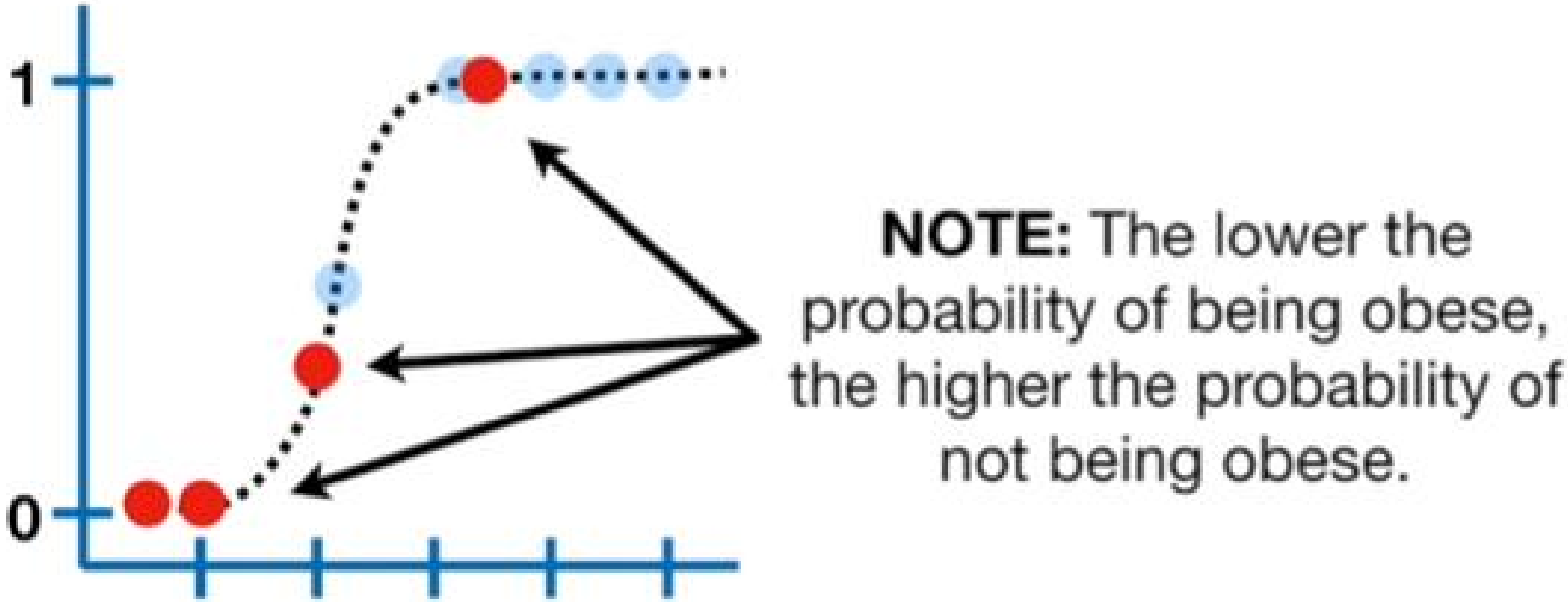
The likelihoods that these mice are obese are 0.91, 0.91 and 0.92

likelihood of data given the squiggle = 0.49 × 0.9 × 0.91 × 0.91 × 0.92 × …

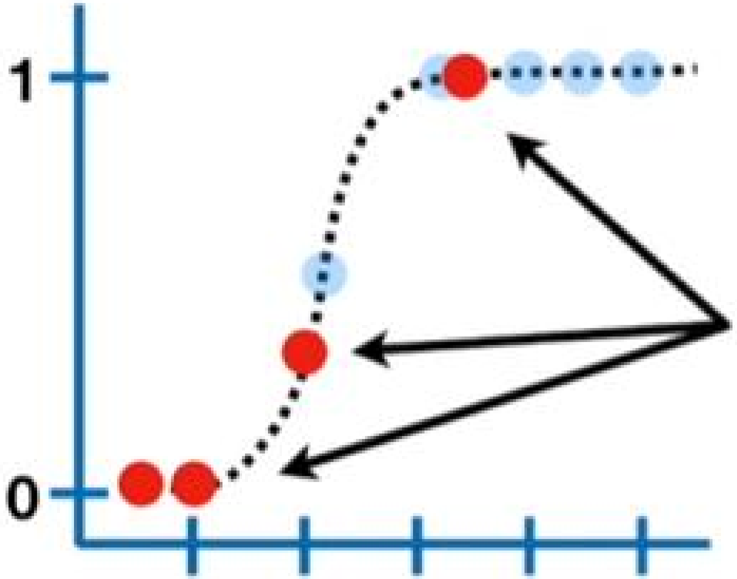
The likelihood for all of the obese mice is just the product of the individual likelihoods.



likelihood of data given the squiggle = $0.49 \times 0.9 \times 0.91 \times 0.91 \times 0.92 \times \ldots$

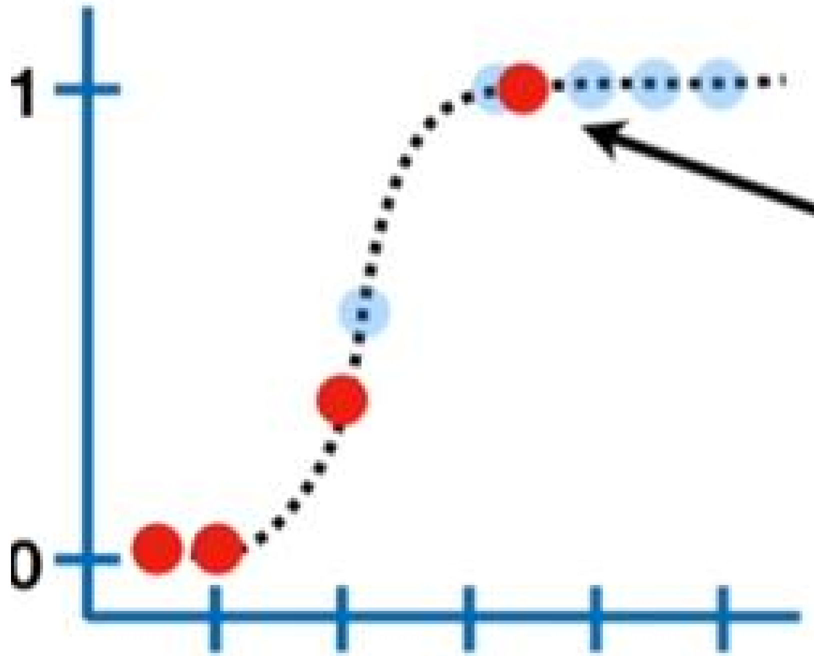


likelihood of data given the squiggle = 0.49 × 0.9 × 0.91 × 0.91 × 0.92 × …



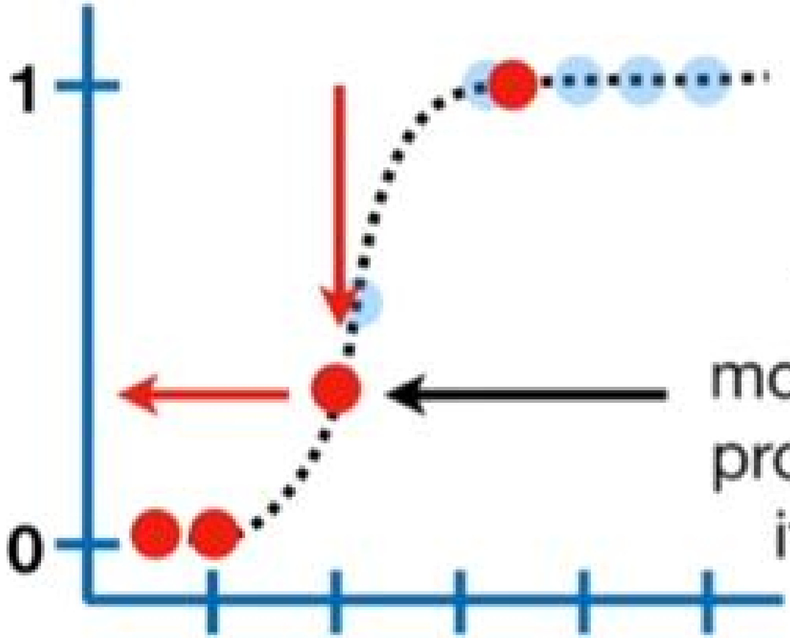
Thus, for these mice, the
likelihood = (1 - probability the mouse is obese)

likelihood of data given the squiggle = 0.49 x 0.9 x 0.91 x 0.91 x 0.92 x …



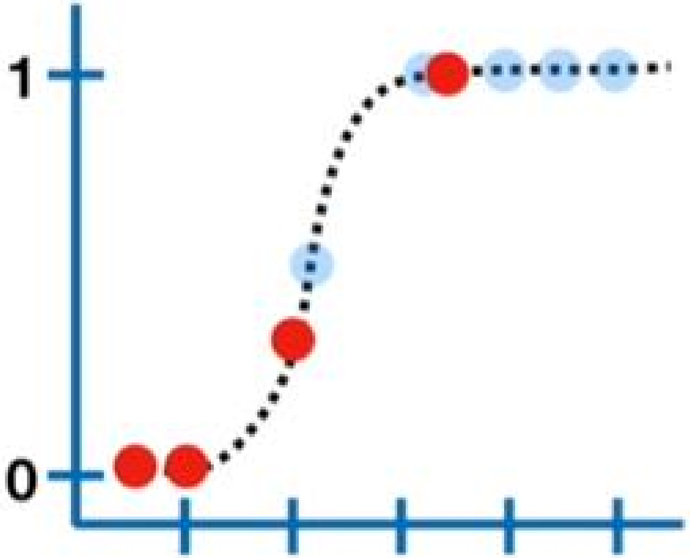
The probability that this mouse is obese is 0.9, so the probability and likelihood that it is not obese is (1 - 0.9)

likelihood of data given the squiggle = 0.49 × 0.9 × 0.91 × 0.91 × 0.92 × …



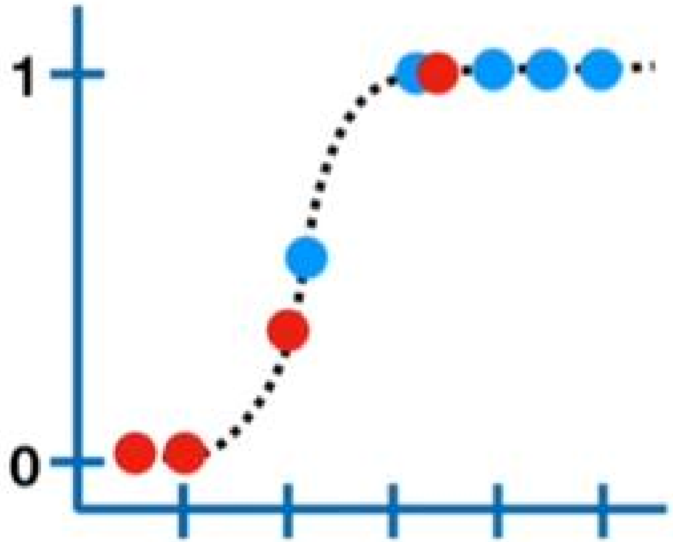
The probability that this mouse is obese is 0.3, so the probability and likelihood that it is not obese is (1 - 0.3)

likelihood of data given the squiggle = $0.49 \times 0.9 \times 0.91 \times 0.91 \times 0.92 \times$
$(1 - 0.9) \times (1 - 0.3) \times (1 - 0.01) \times (1 - 0.01)$



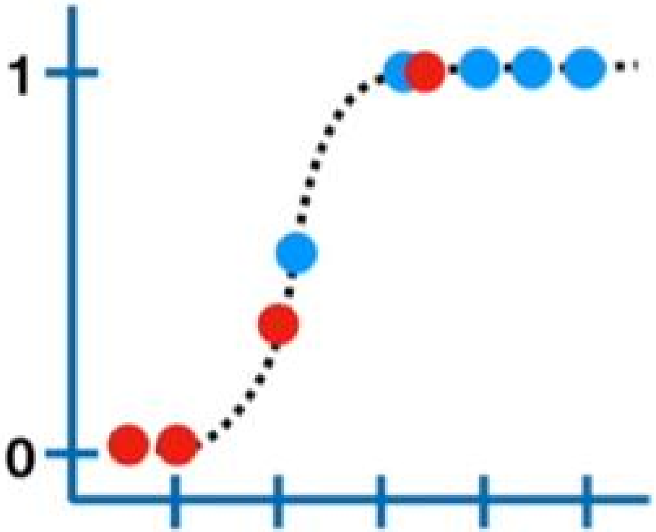
Now we can include the individual likelihoods for the mice that are **not obese** to the equation for the overall likelihood.

likelihood of data given the squiggle = $0.49 \times 0.9 \times 0.91 \times 0.91 \times 0.92 \times (1 - 0.9) \times (1 - 0.3) \times (1 - 0.01) \times (1 - 0.01)$



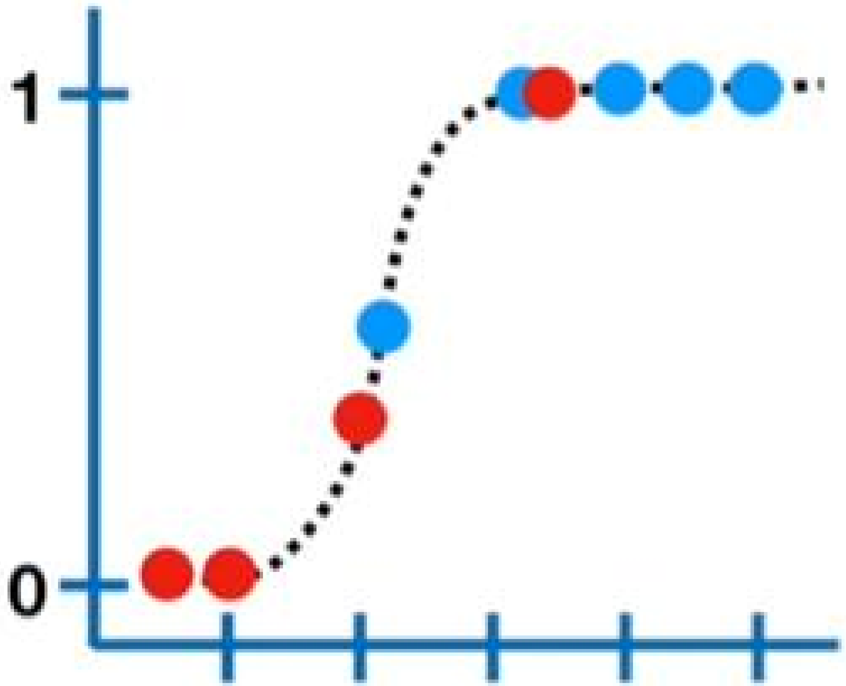
NOTE: Although it is possible to calculate the likelihood as the product of the individual likelihoods, statisticians prefer to calculate the **log of the likelihood** instead.

log(likelihood of data given the squiggle) = log(0.49) + log(0.9) + log(0.91) + log(0.91) + log(0.92) + log(1 - 0.9) + log(1 - 0.3) + log(1 - 0.01) + log(1 - 0.01)

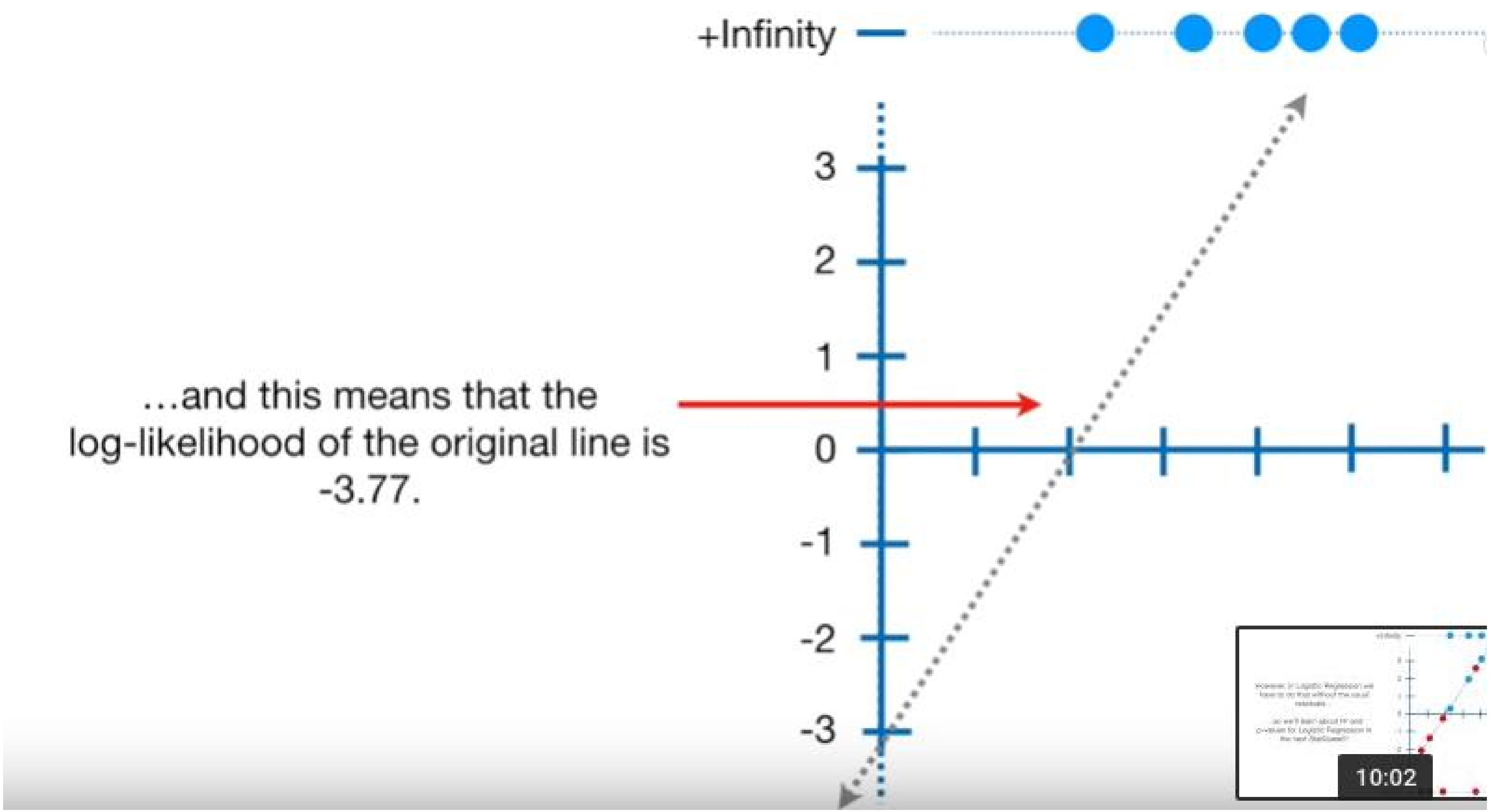


With the log of the likelihood, or "log-likelihood" to those in the know, we **add the logs of the individual likelihoods** instead of multiplying the individual likelihoods…

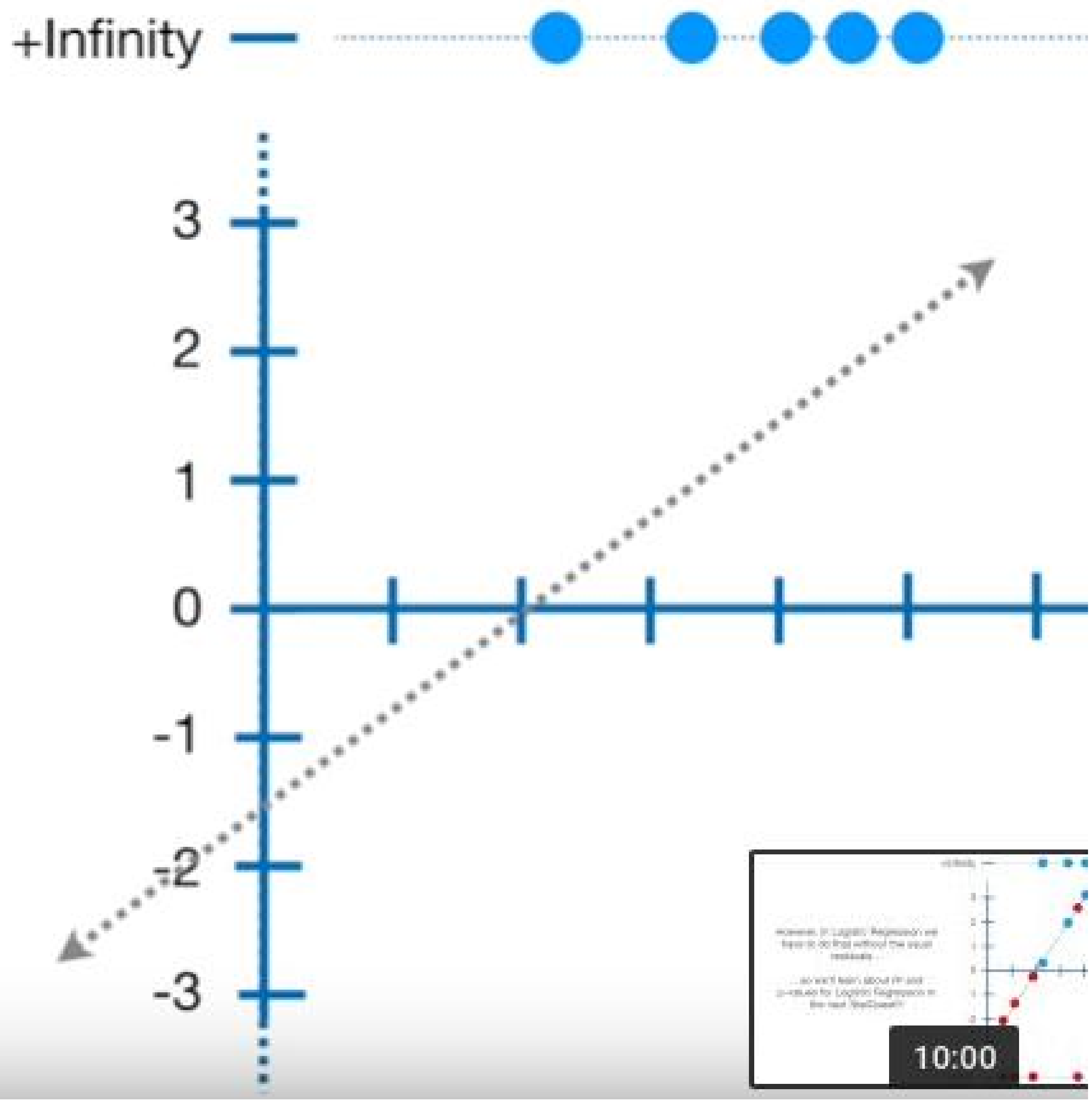
log(likelihood of data given the squiggle) = -3.77



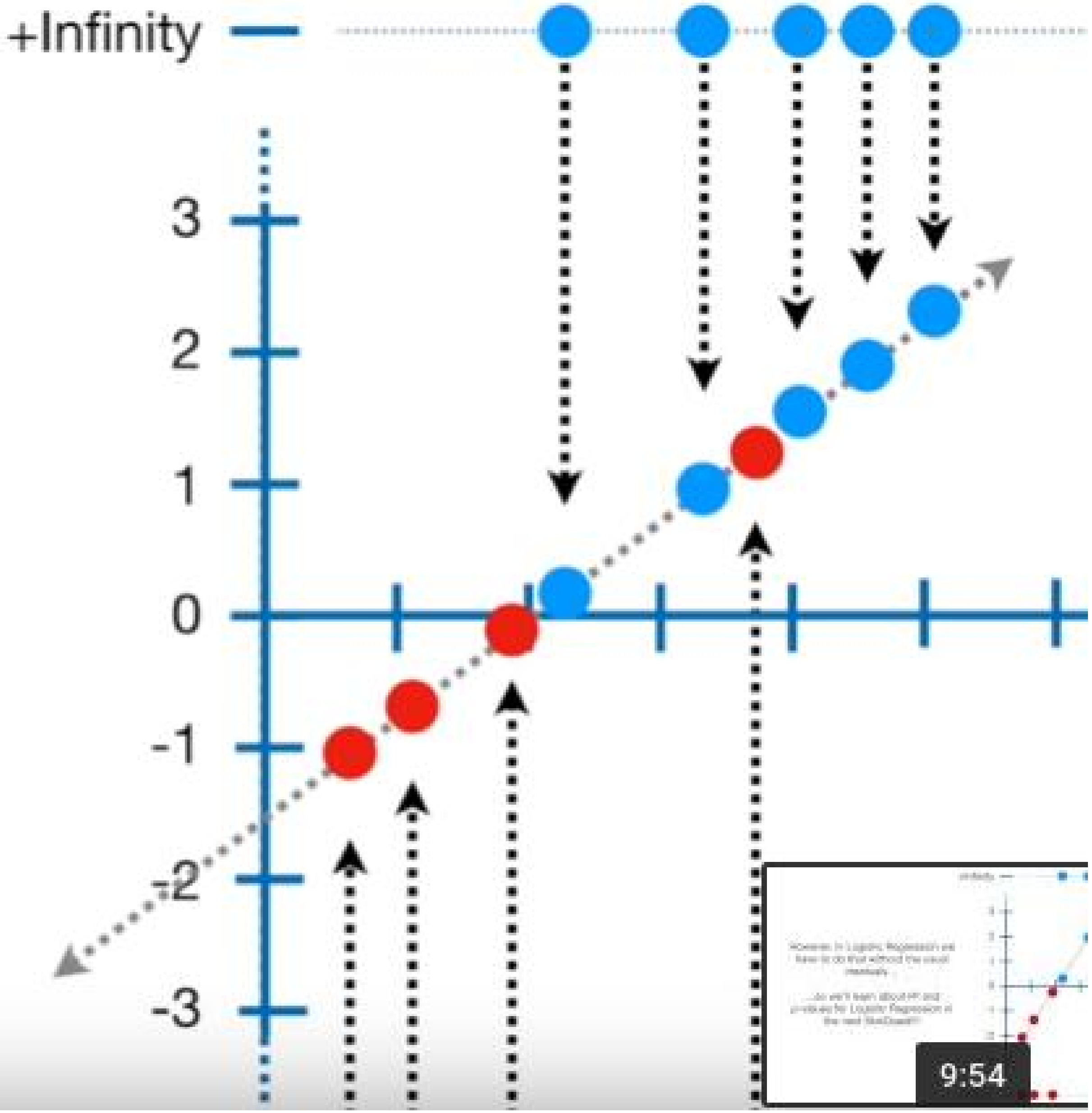
Thus, the log-likelihood of the data given the squiggle is -3.77…

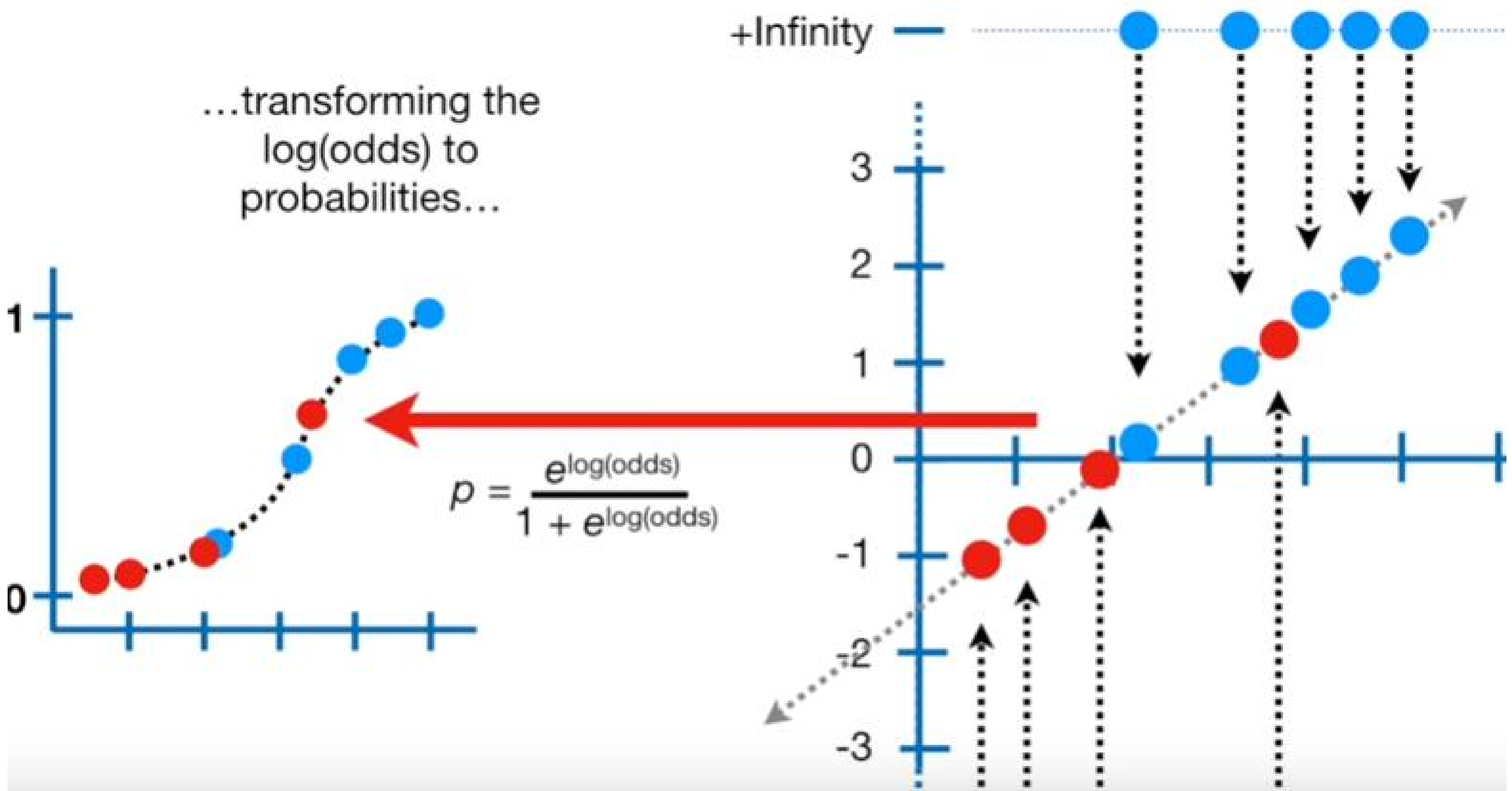

+Infinity
3
2
1
0
-1
-2
-3
...and this means that the log-likelihood of the original line is -3.77.
10:02

Now we rotate the line...



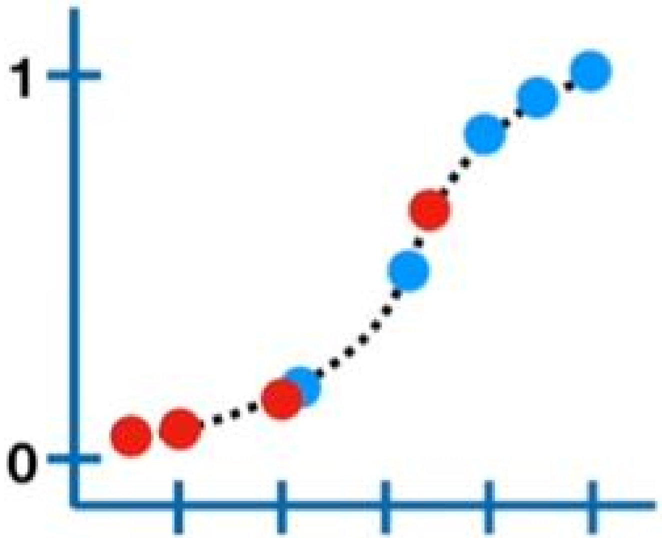
…and calculate its log-likelihood by projecting the data onto it…





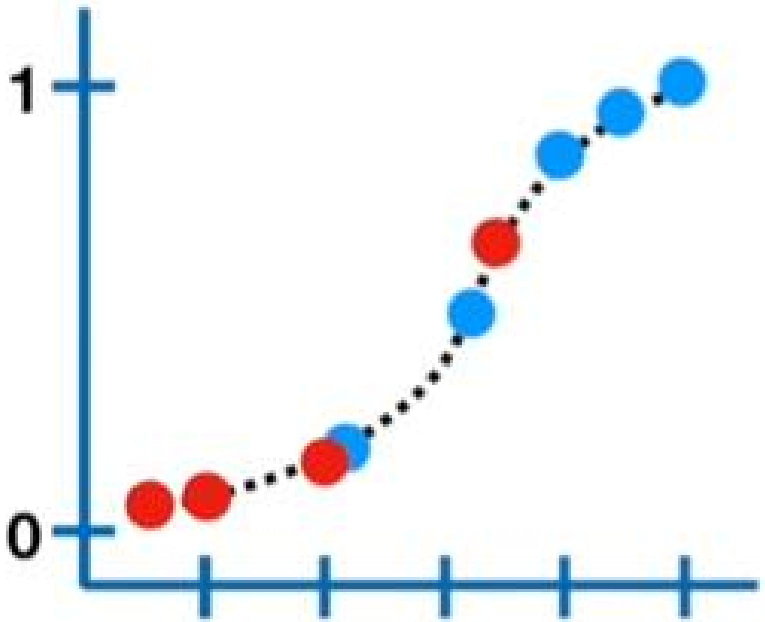
+Infinity
…transforming the log(odds) to probabilities…
$p = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$
1
0
3
2
1
0
-1
-2
-3

$$\log(\text{likelihood of data given the squiggle}) = \log(0.22) + \log(0.4) + \log(0.8) + \log(0.89) + \log(0.92) + \log(1 - 0.6) + \log(1 - 0.2) + \log(1 - 0.1) + \log(1 - 0.05)$$



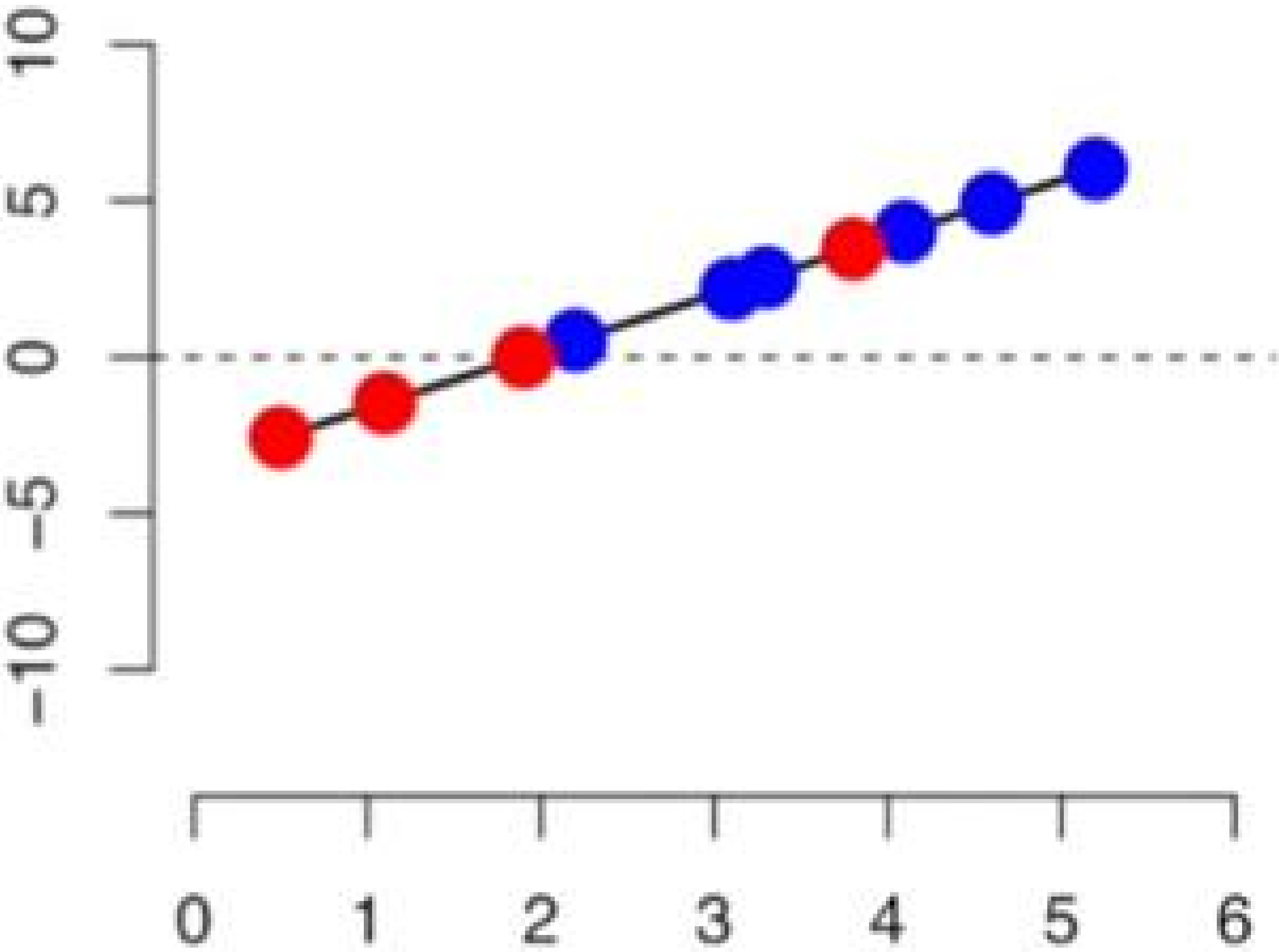
...and then calculating the log-likelihood...

log(likelihood of data given the squiggle) = -4.15

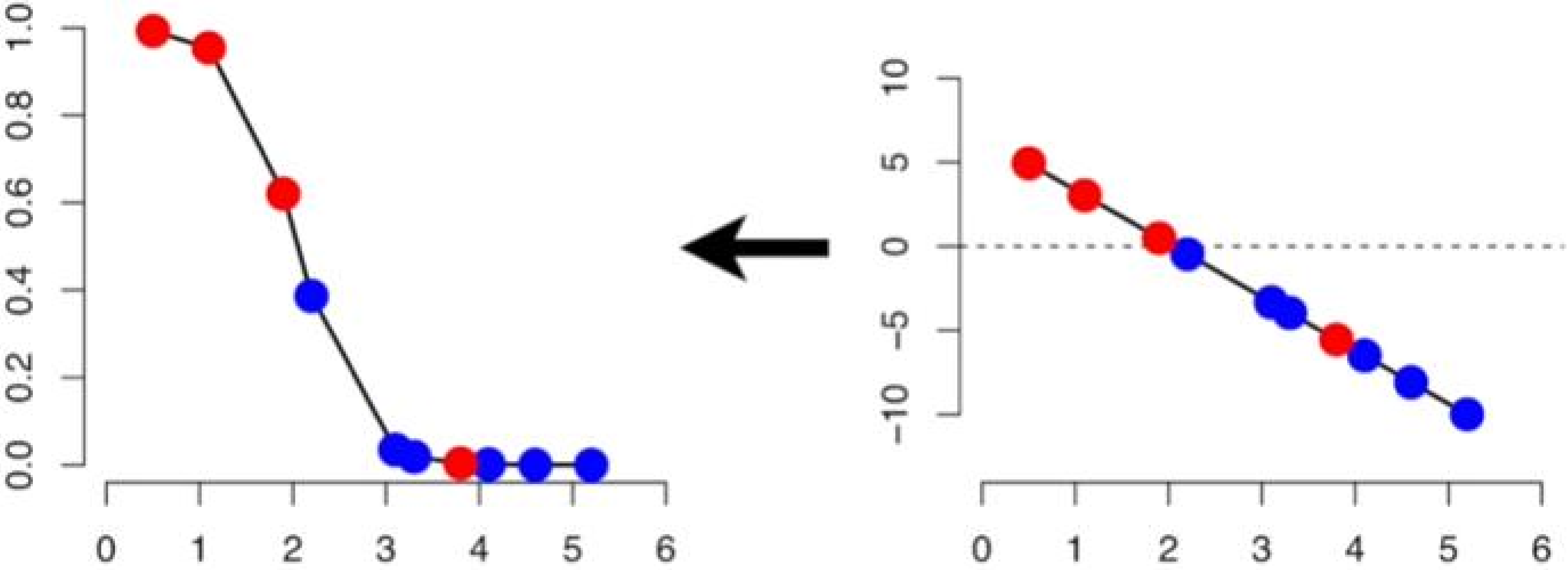


...and the final value for the log-likelihood is -4.15.

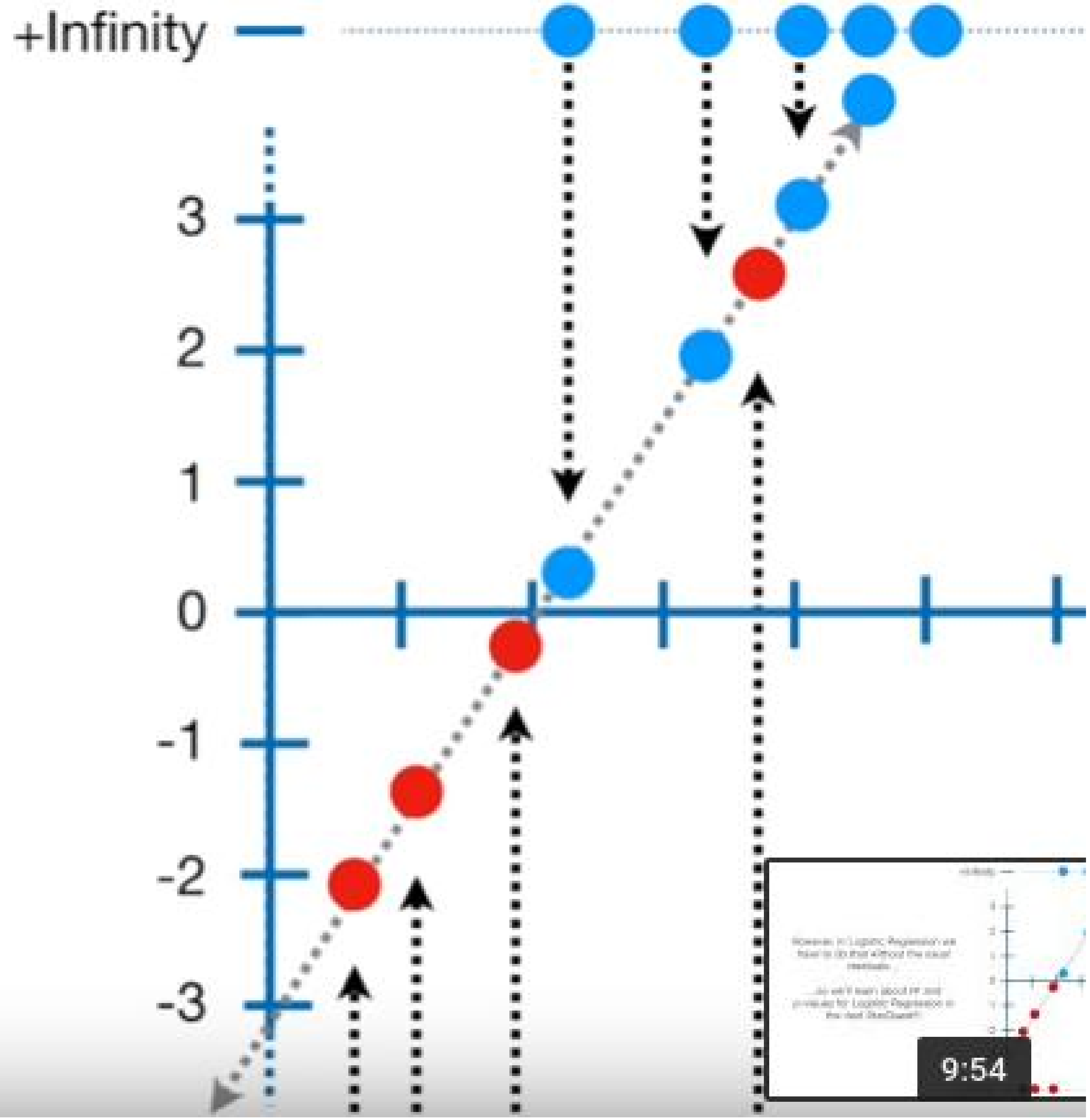
…and we just keep rotating the log(odds) line and projecting the data onto it…

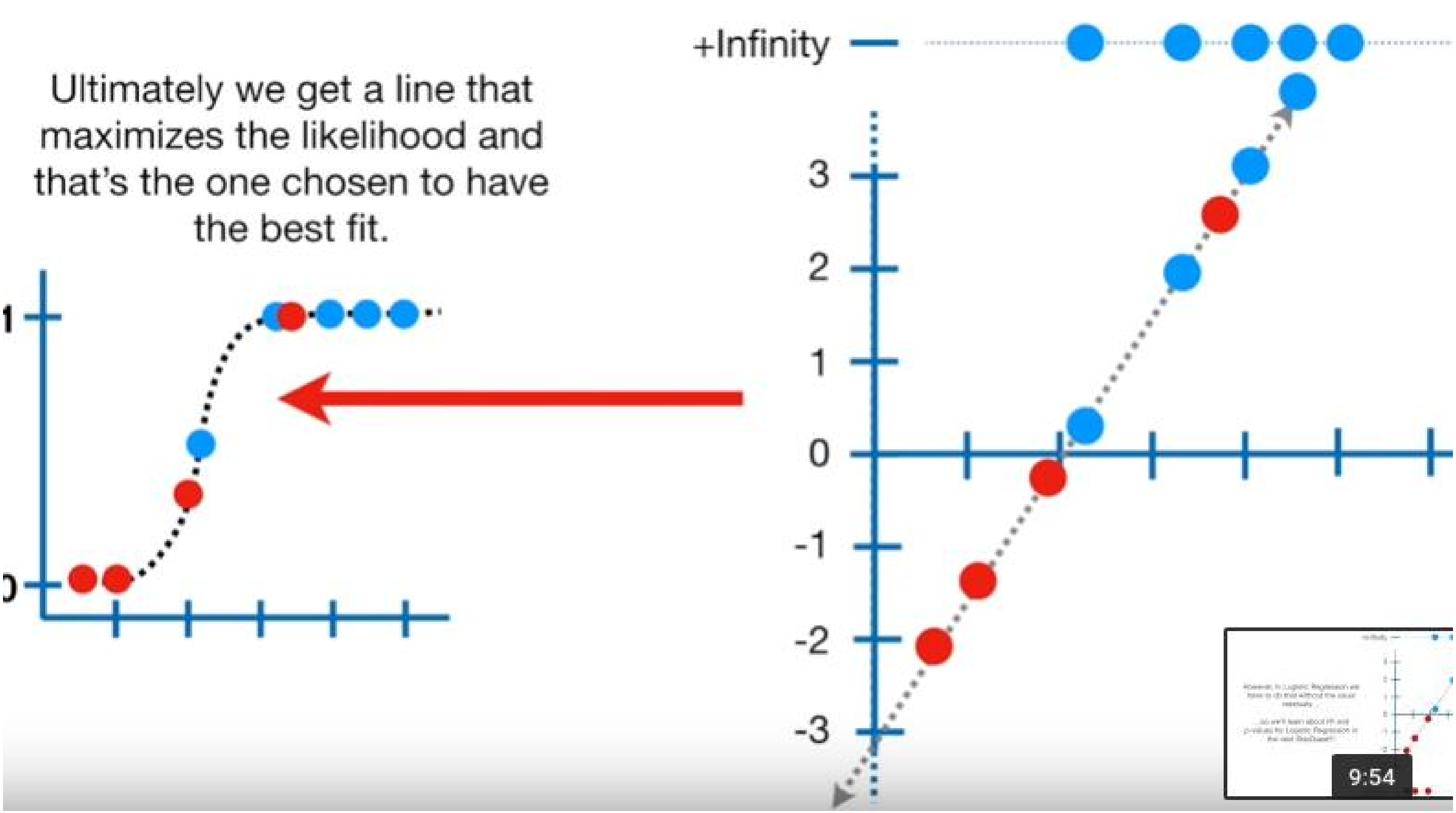


…and transforming it to
probabilities and calculating
the log-likelihood.



NOTE: The algorithm that finds the line with the maximum likelihood is pretty smart - each time it rotates the line, it does so in a way that increases the log-likelihood. Thus, the algorithm can find the optimal fit after a few rotations.




Ultimately we get a line that maximizes the likelihood and that's the one chosen to have the best fit.
+Infinity
3
2
1
0
-1
-2
-3
1
0
9:54