…but if you don’t like silly songs…

…that’s okay!!!

StatQuest!!!

…that’s okay!!!

StatQuest!!!

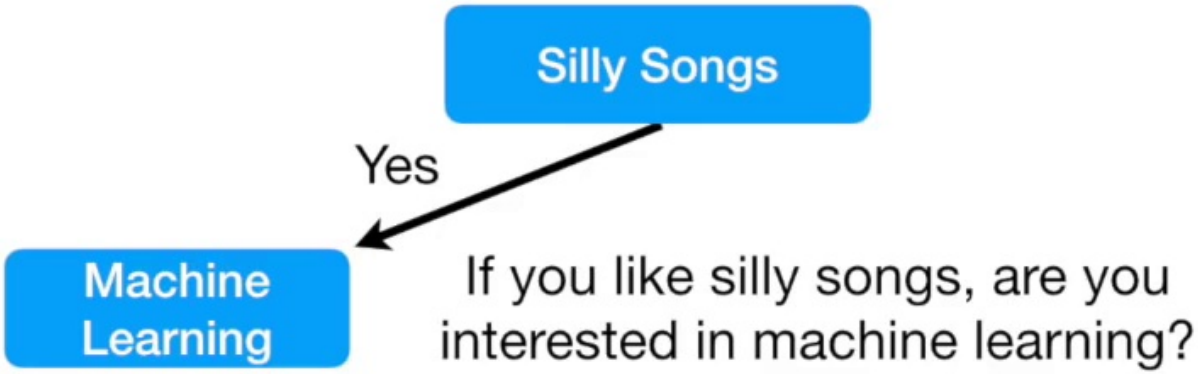
NOTE: This StatQuest was originally prepared for and presented at the **Society for Scientific Advancement's** annual conference…

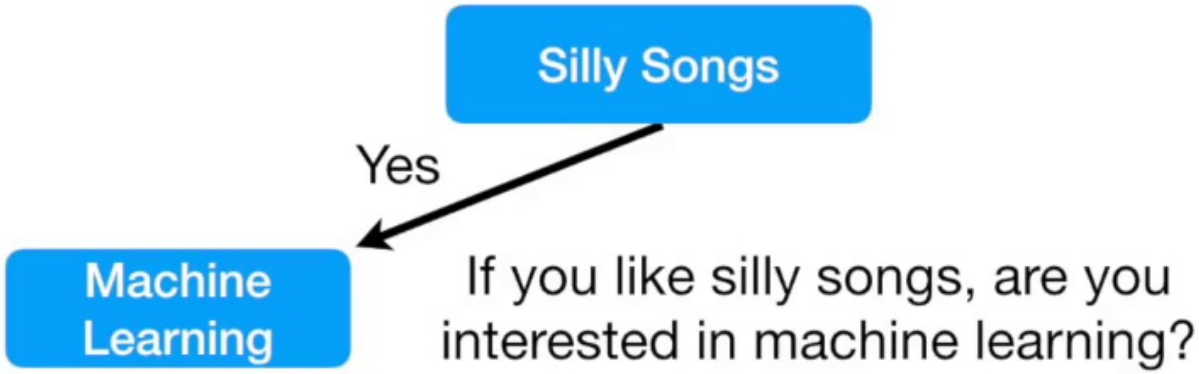
SOSA | SOCIETY FOR SCIENTIFIC ADVANCEMENT

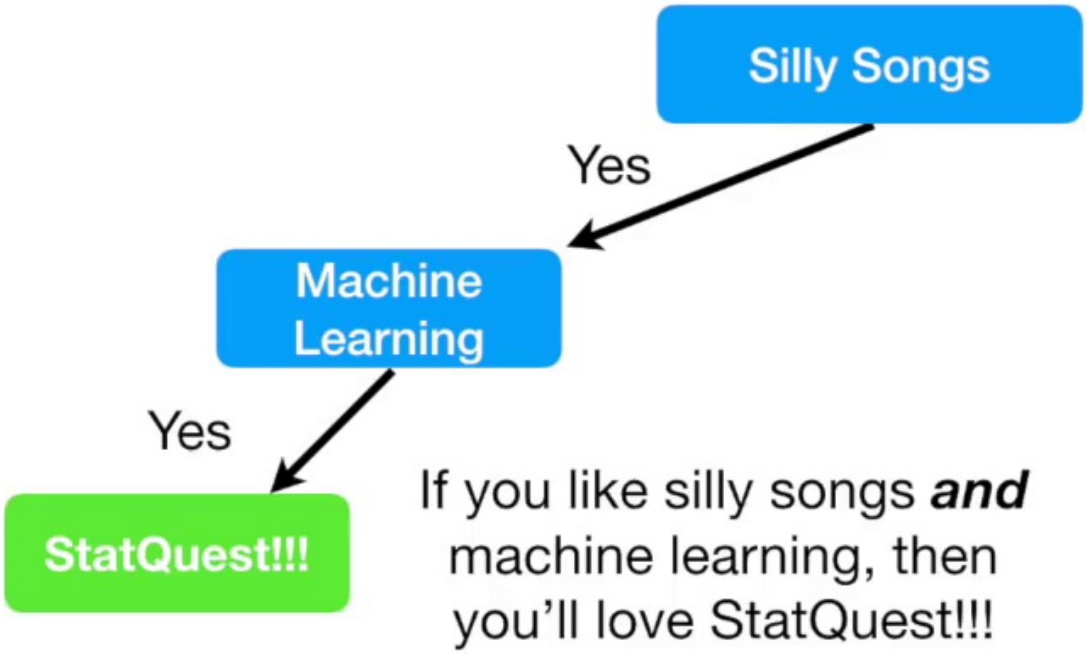
www.thesosa.org

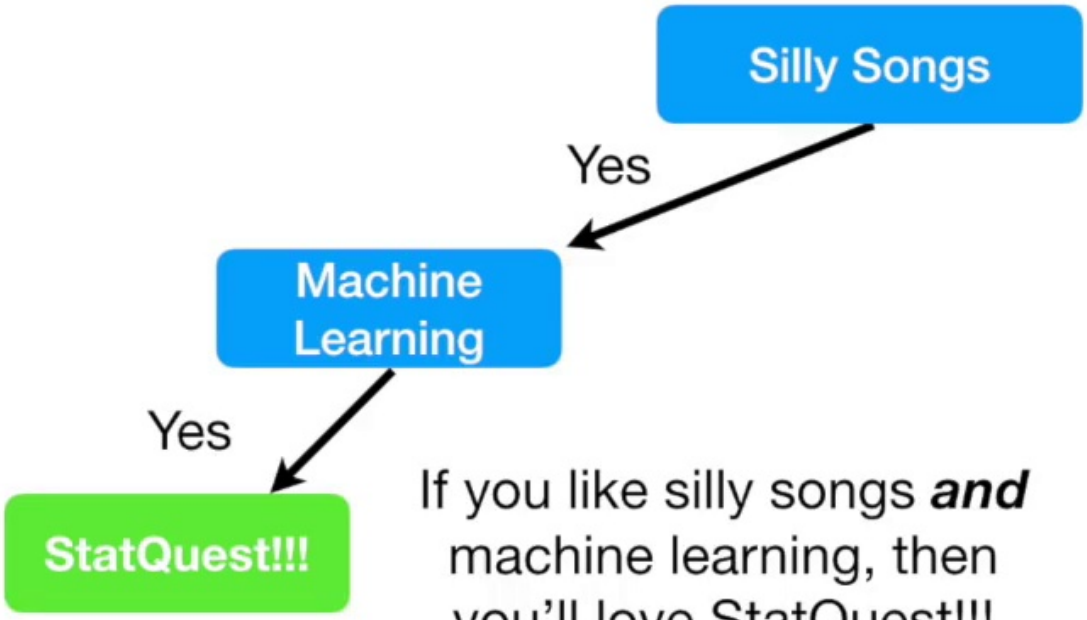
One of the things **SoSA** does is promote Science and Technology in **Jamaica**…

One of the things **SoSA** does is promote Science and Technology in **Jamaica**…

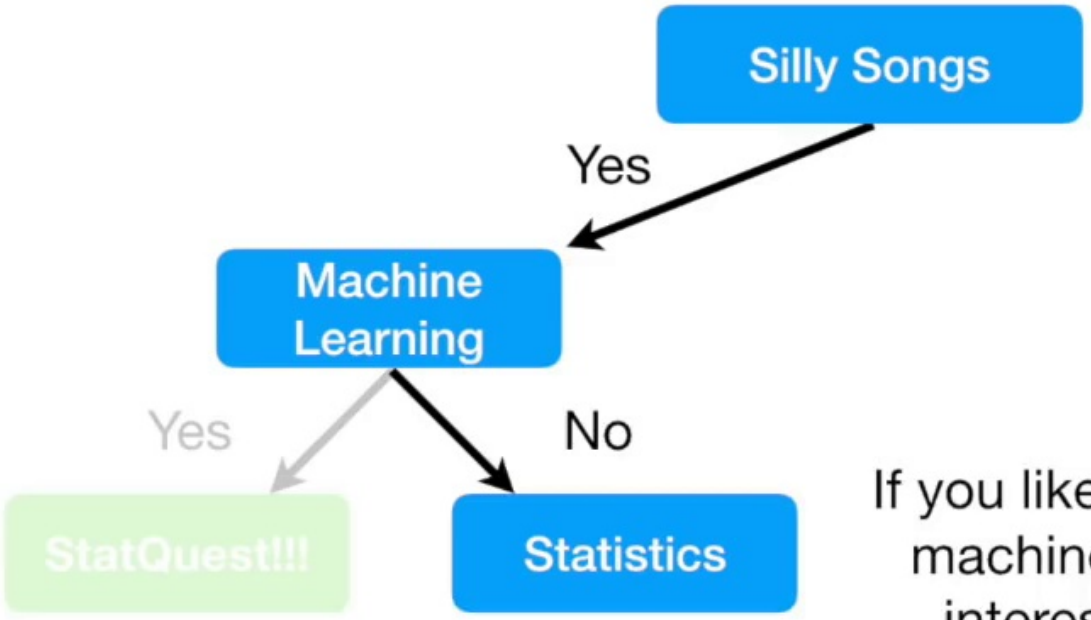

Silly Songs
Yes
Machine Learning
If you like silly songs, are you interested in machine learning?


Silly Songs
Yes
Machine Learning
If you like silly songs, are you interested in machine learning?

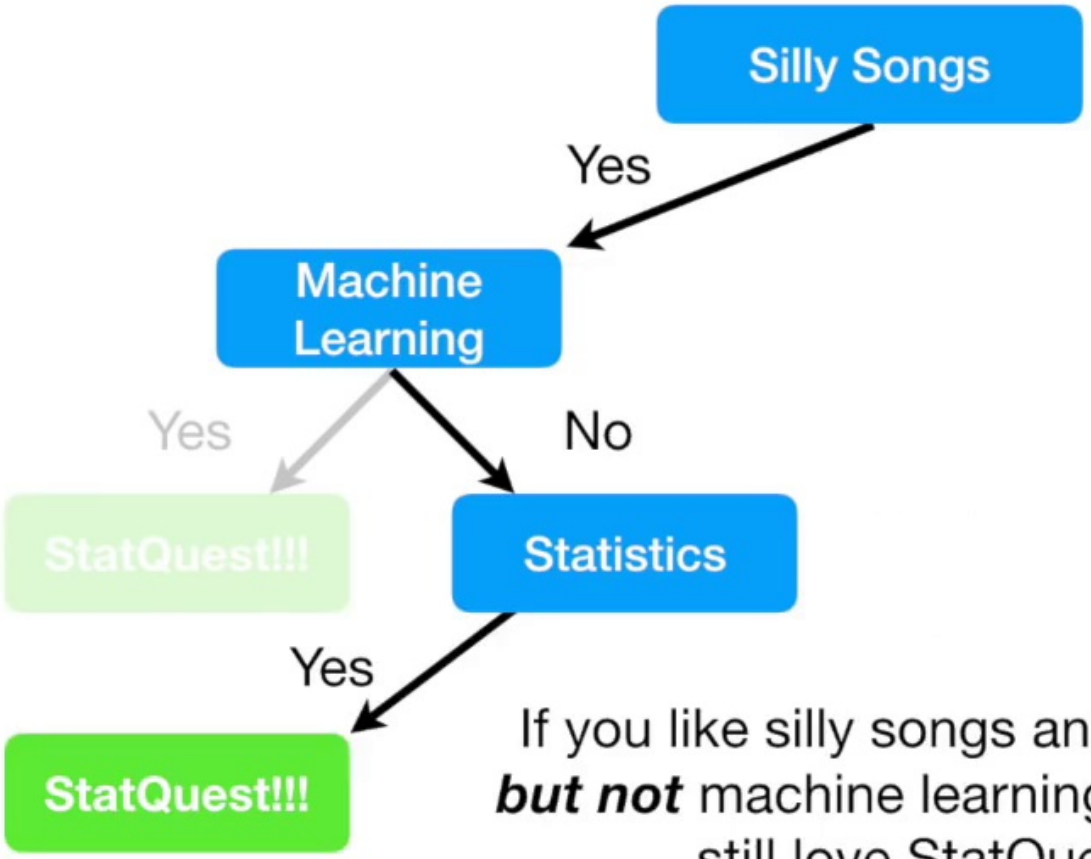

Silly Songs
Yes
Machine Learning
Yes
StatQuest!!!
If you like silly songs ***and*** machine learning, then you'll love StatQuest!!!



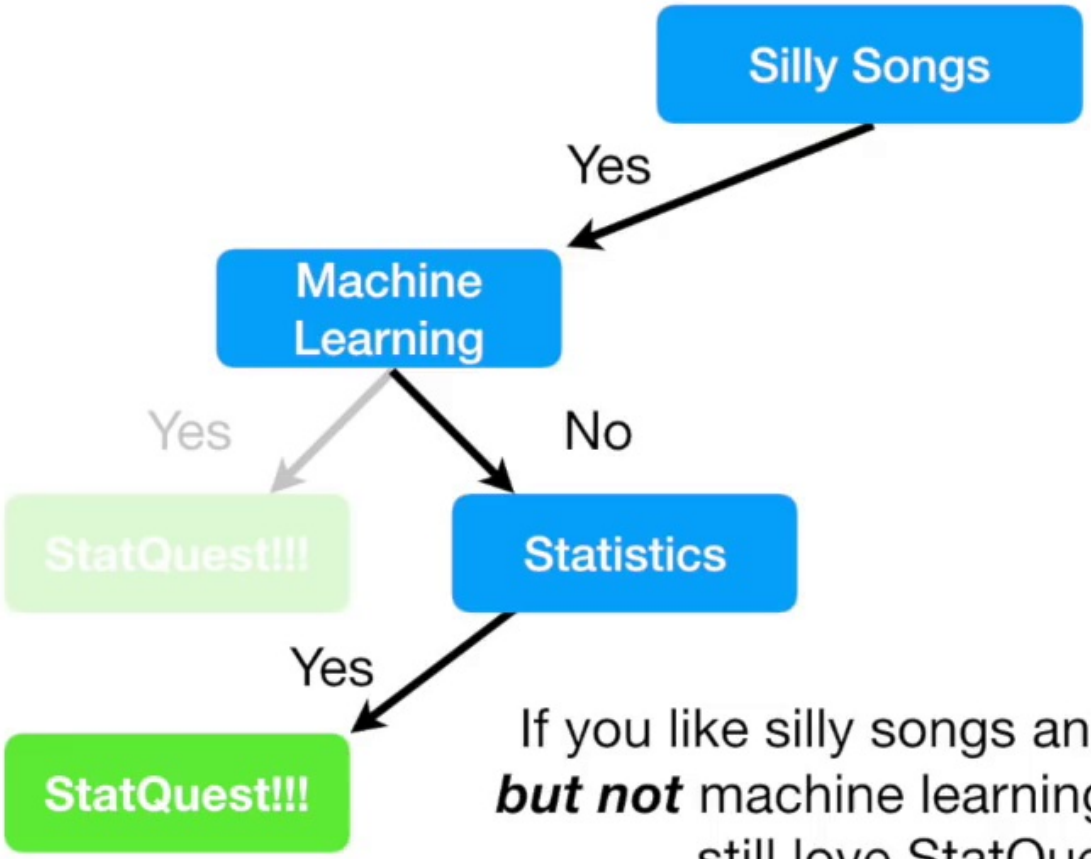
If you like silly songs ***and*** machine learning, then you'll love StatQuest!!!



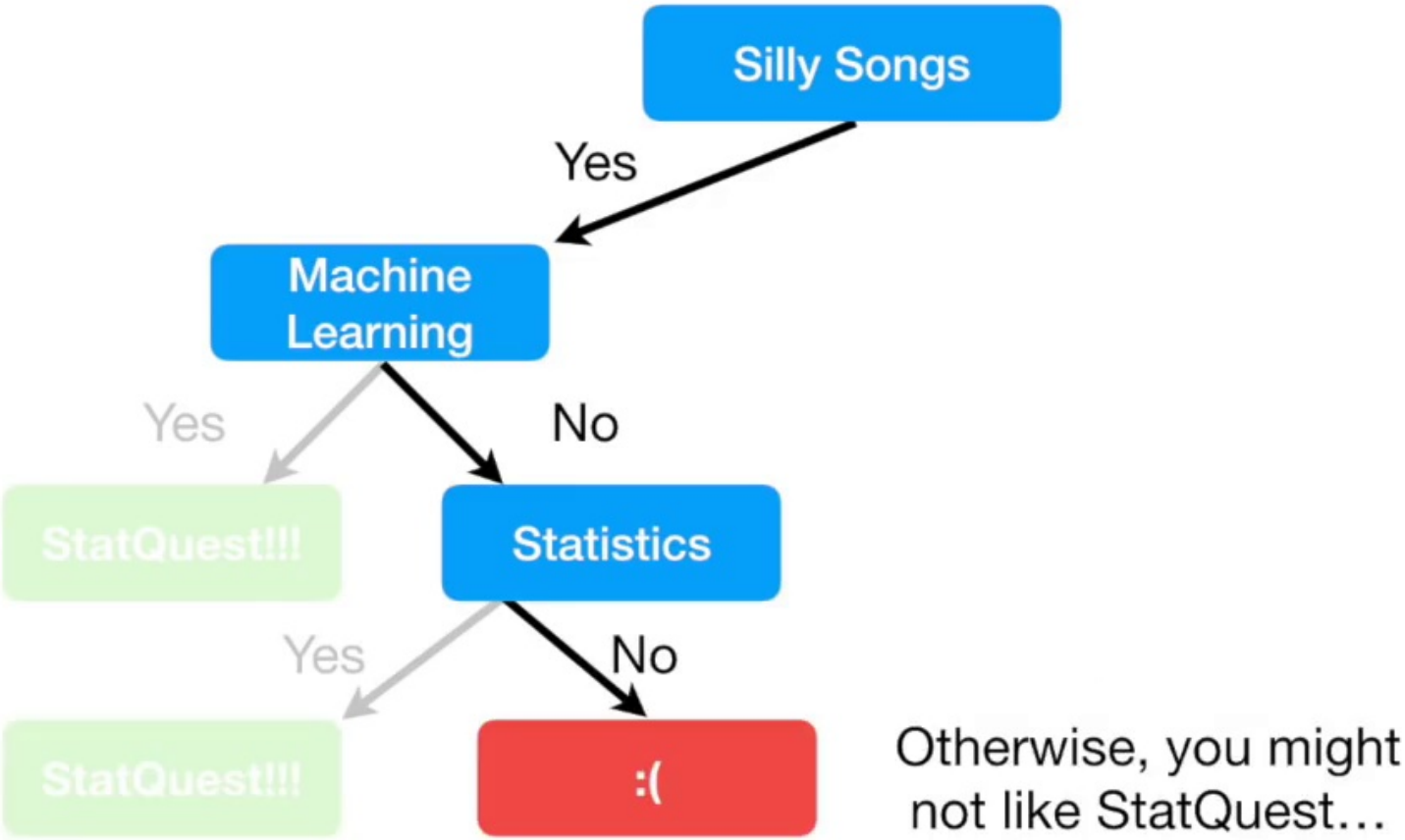
If you like silly songs ***but not*** machine learning, are you interested in statistics?

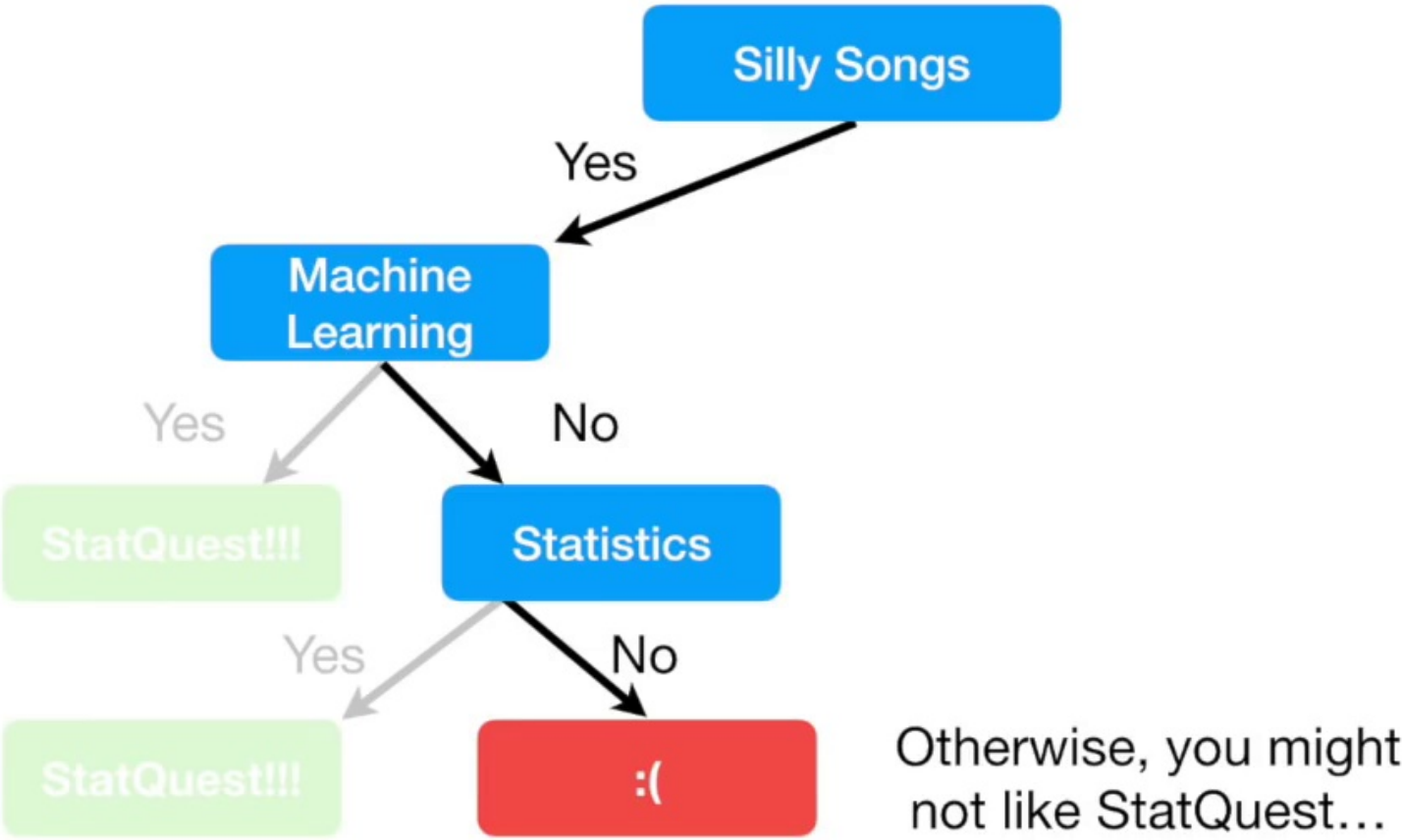


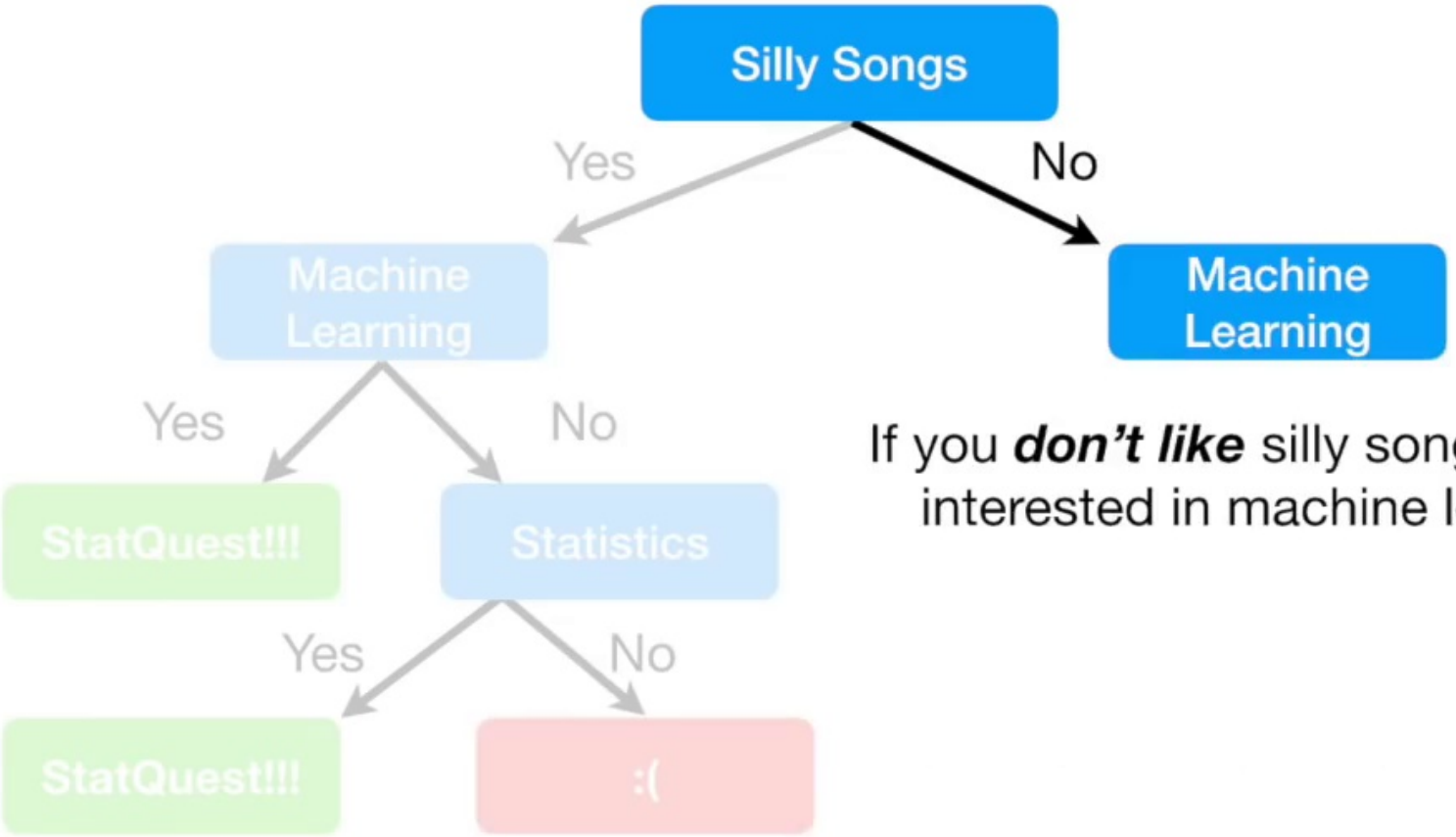
If you like silly songs and statistics, ***but not*** machine learning, then you'll still love StatQuest!!!



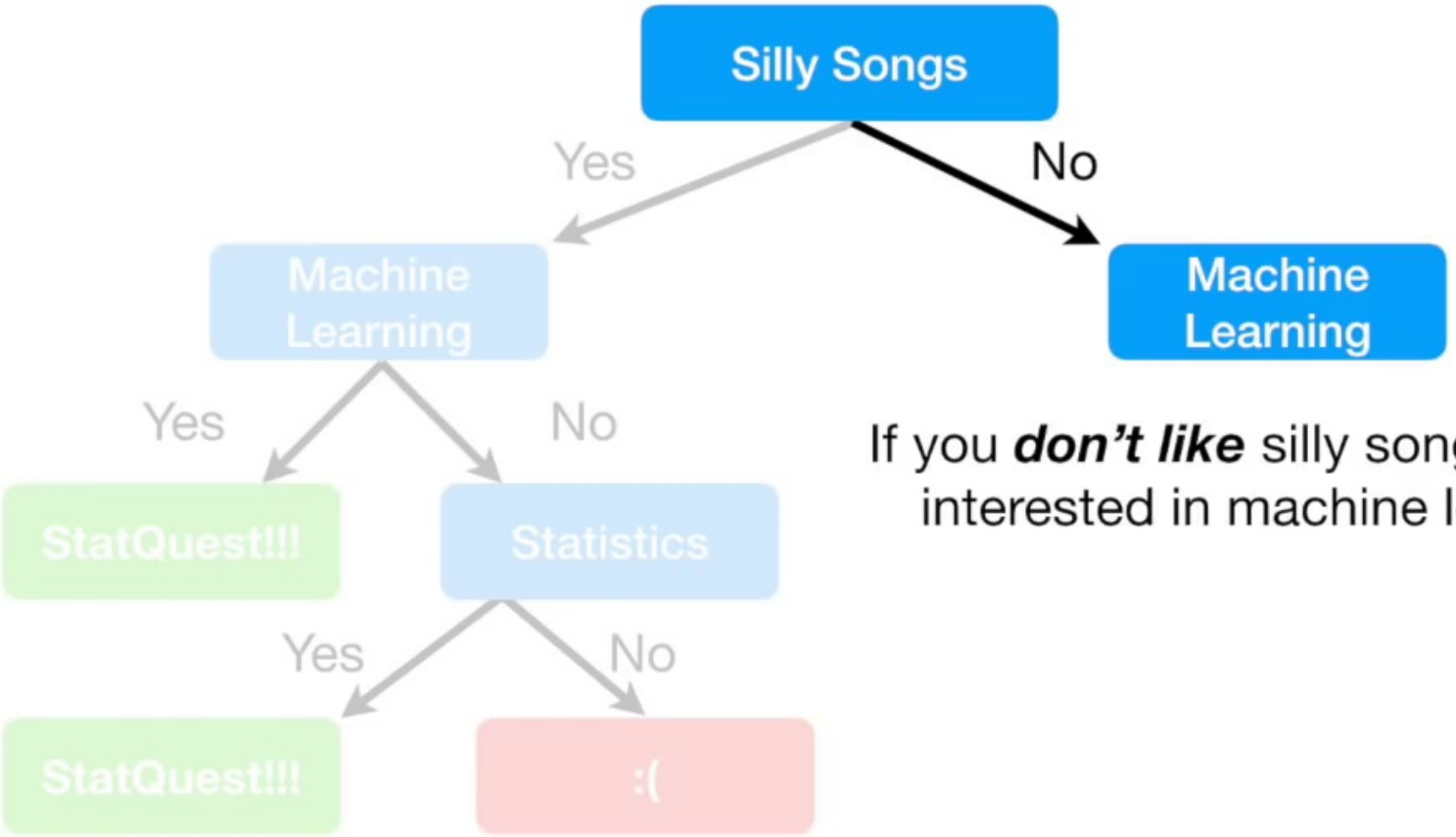
If you like silly songs and statistics, ***but not*** machine learning, then you'll still love StatQuest!!!


Silly Songs
Yes
Machine Learning
Yes
No
StatQuest!!!
Statistics
Yes
No
StatQuest!!!
:(
Otherwise, you might not like StatQuest…


Silly Songs
Yes
Machine Learning
Yes
No
StatQuest!!!
Statistics
Yes
No
StatQuest!!!
:(
Otherwise, you might not like StatQuest…

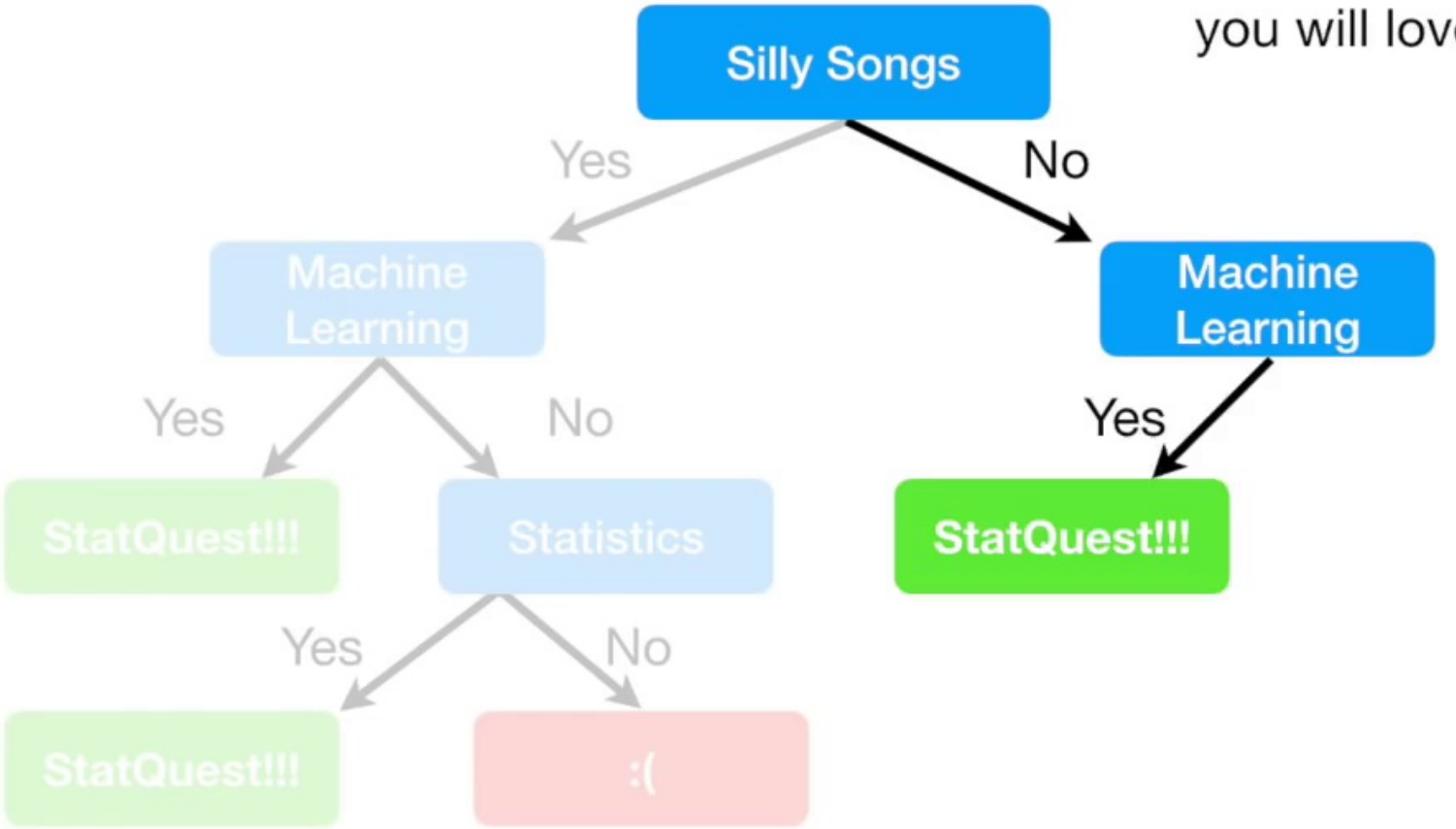


If you ***don't like*** silly songs, are you interested in machine learning?

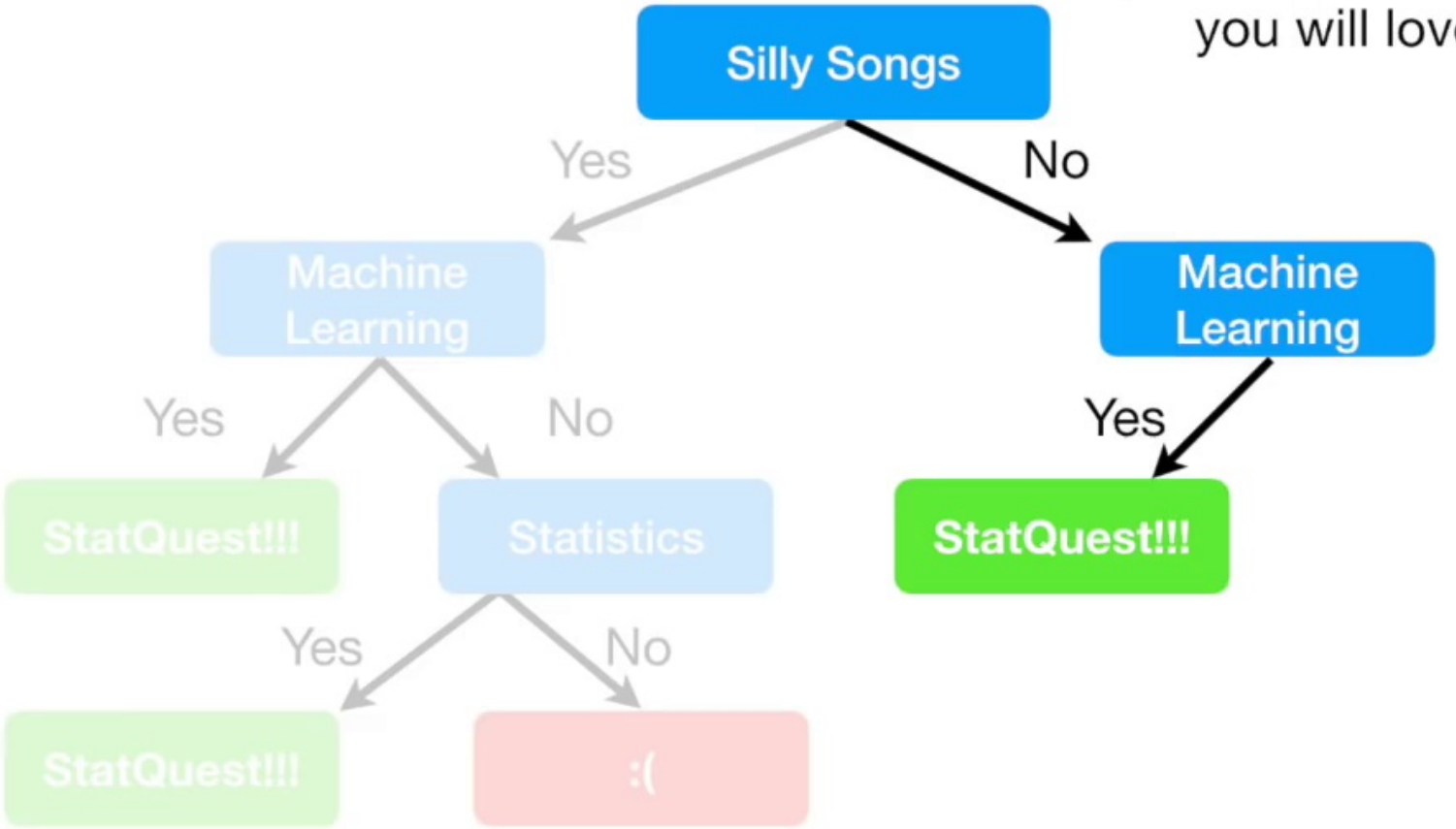


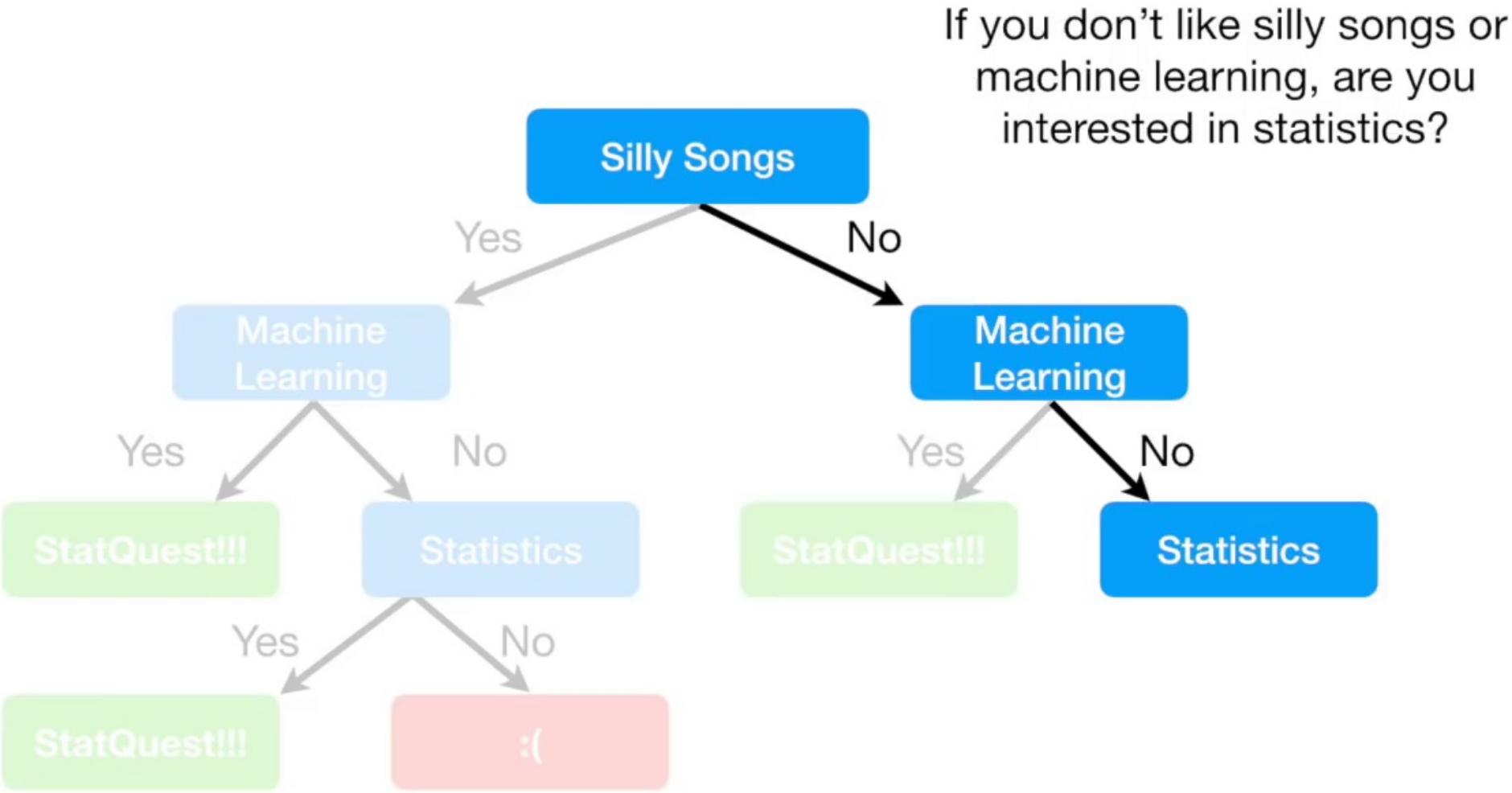
If you ***don't like*** silly songs, are you interested in machine learning?

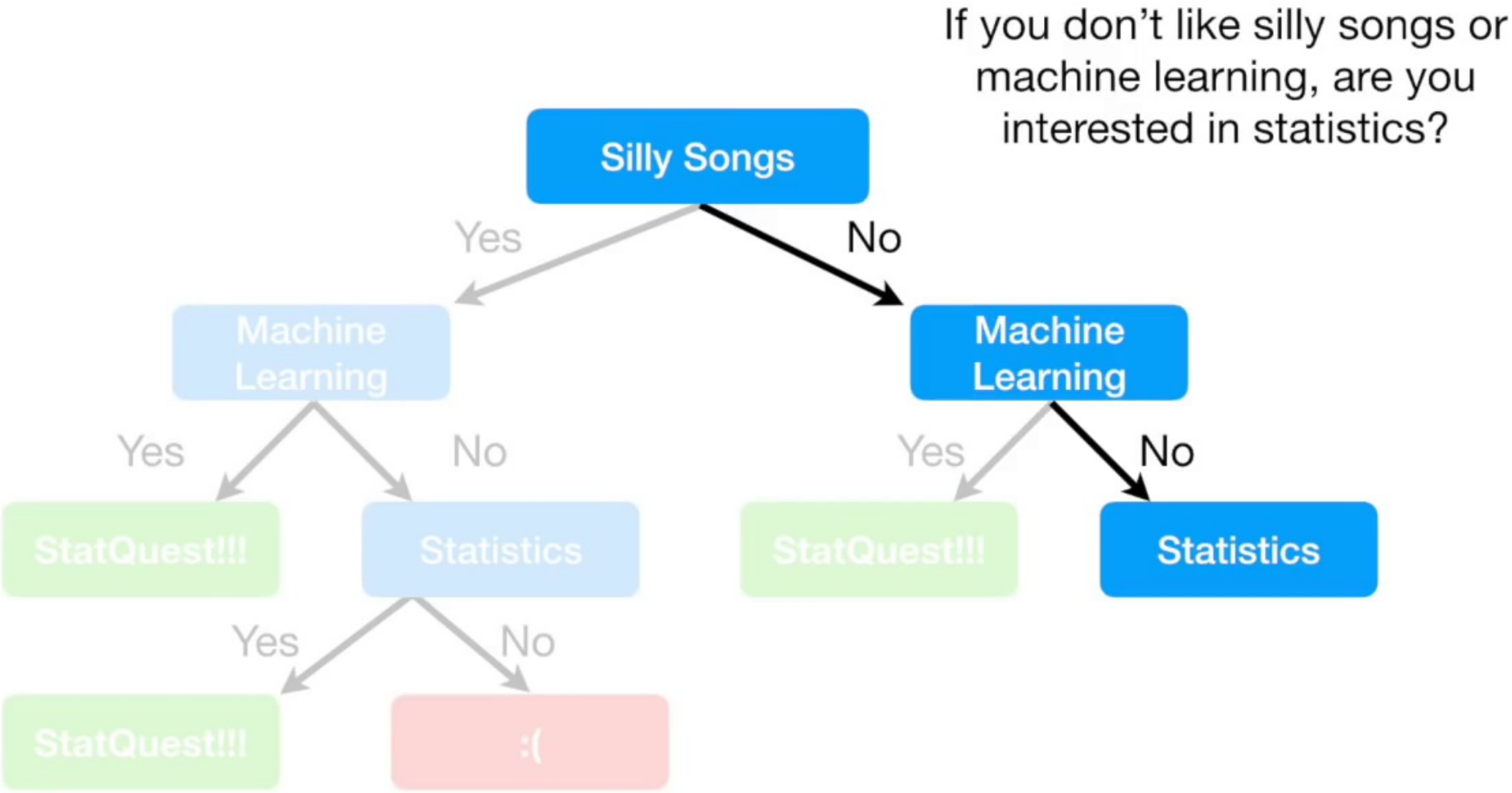
If you ***don't like*** silly songs, but you like machine learning, then you will love StatQuest!!!



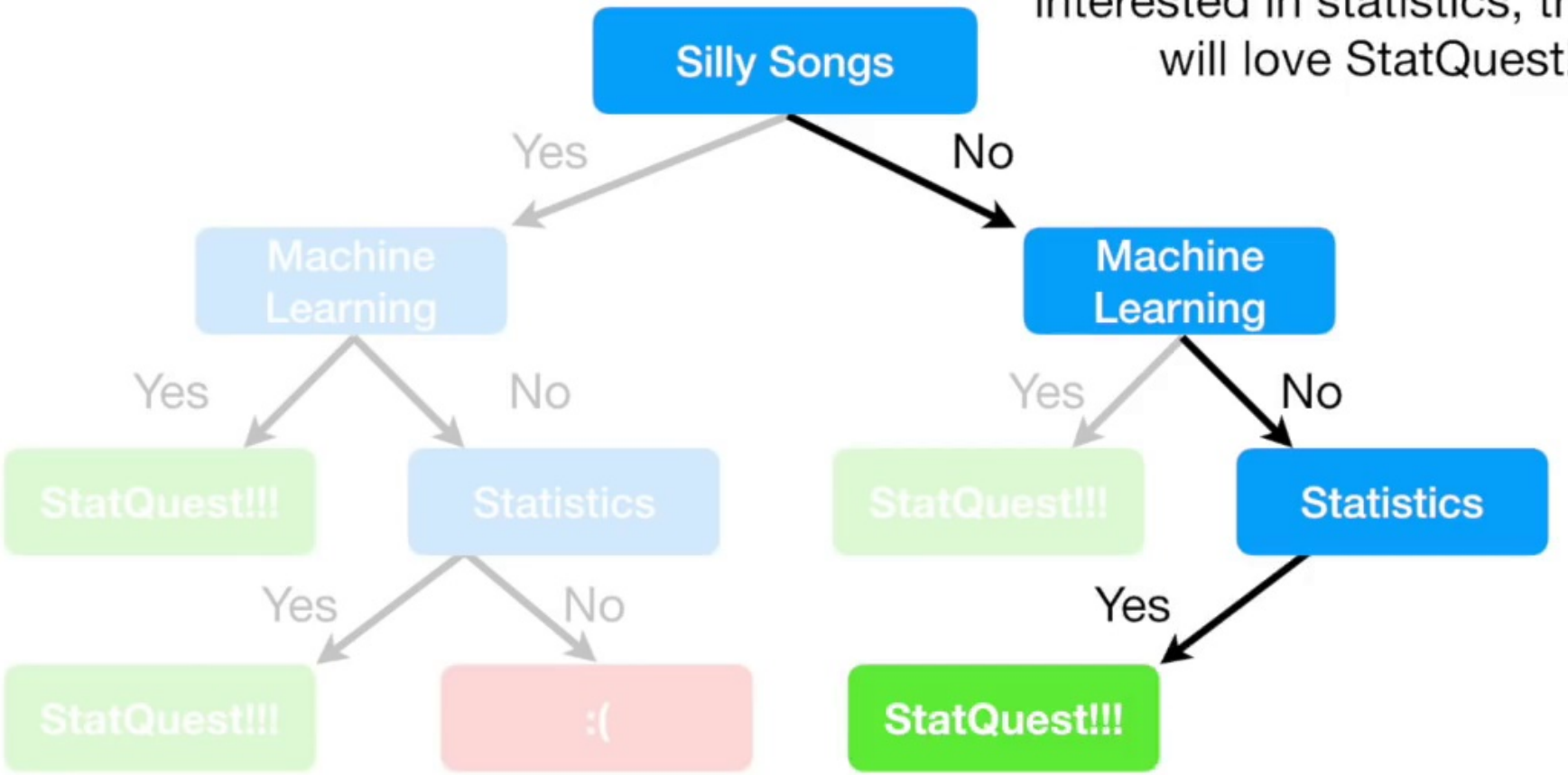
If you ***don't like*** silly songs, but you like machine learning, then you will love StatQuest!!!



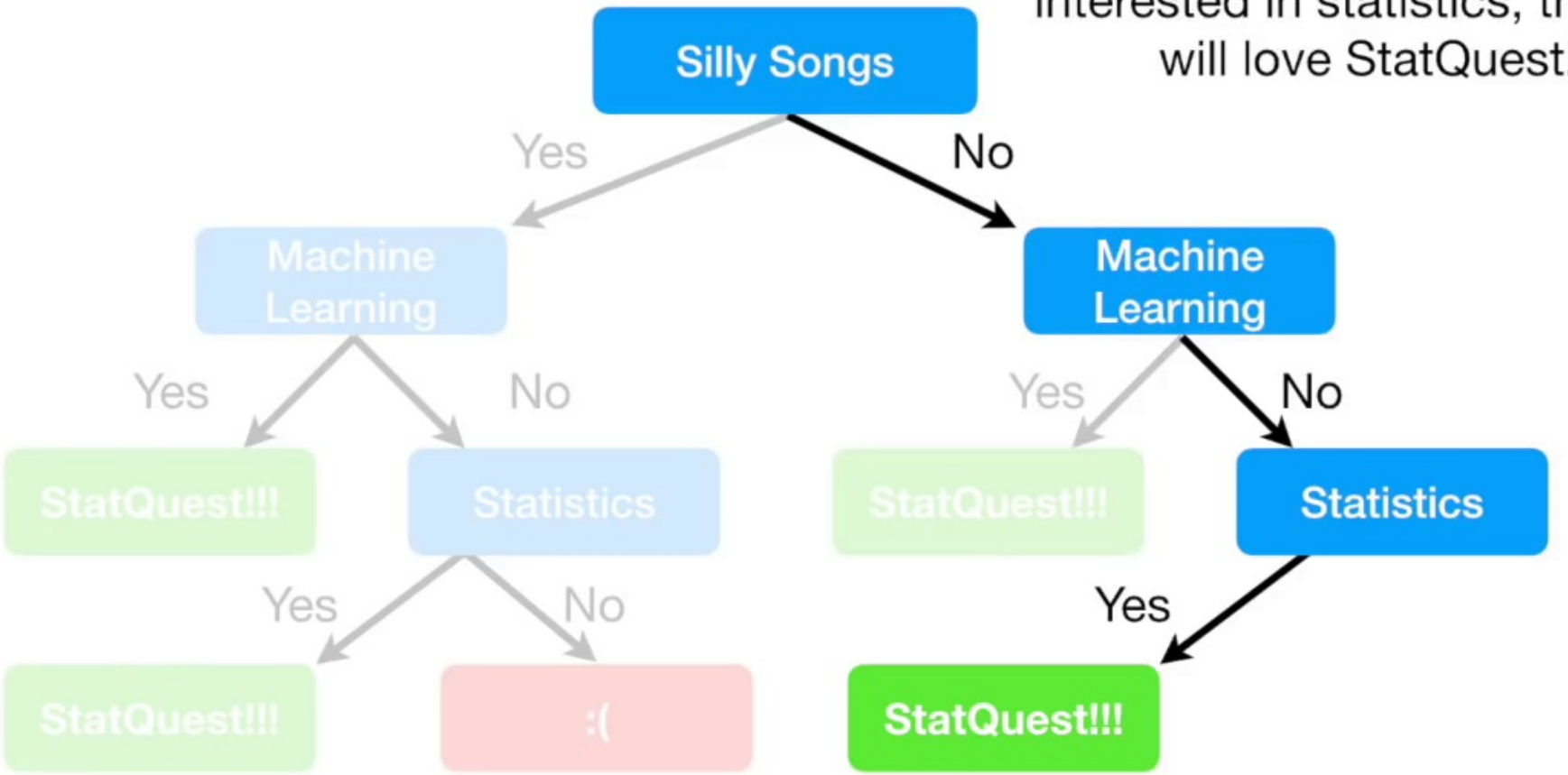

If you don't like silly songs or machine learning, are you interested in statistics?
Silly Songs
Yes
No
Machine Learning
Machine Learning
Yes
No
Yes
No
StatQuest!!!
Statistics
StatQuest!!!
Statistics
Yes
No
StatQuest!!!
:(


If you don't like silly songs or machine learning, are you interested in statistics?
Silly Songs
Yes
No
Machine Learning
Machine Learning
Yes
No
Yes
No
StatQuest!!!
Statistics
StatQuest!!!
Statistics
Yes
No
StatQuest!!!
:(

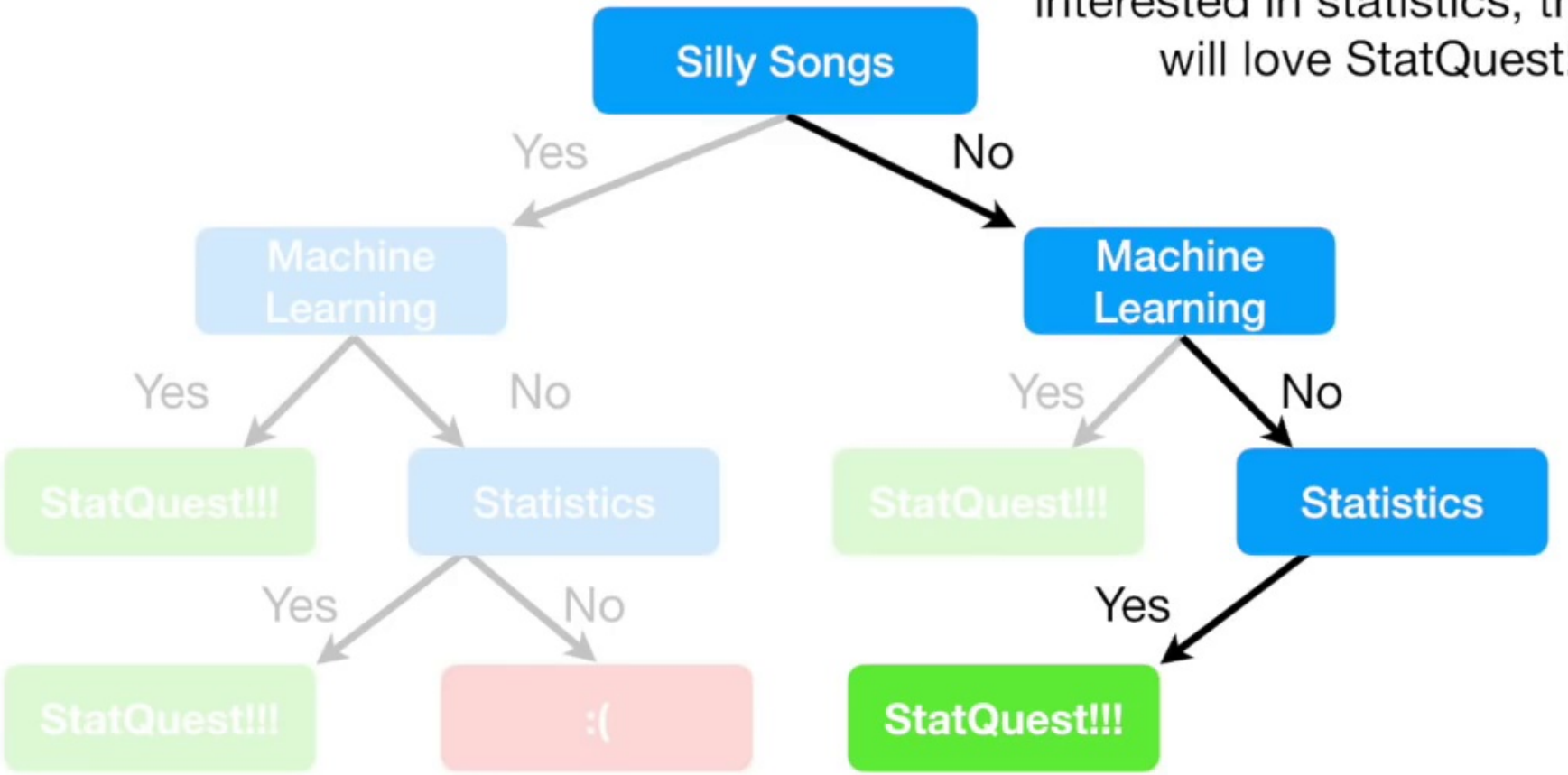
If you ***don't like*** silly songs or machine learning, but you're interested in statistics, then you will love StatQuest!!!

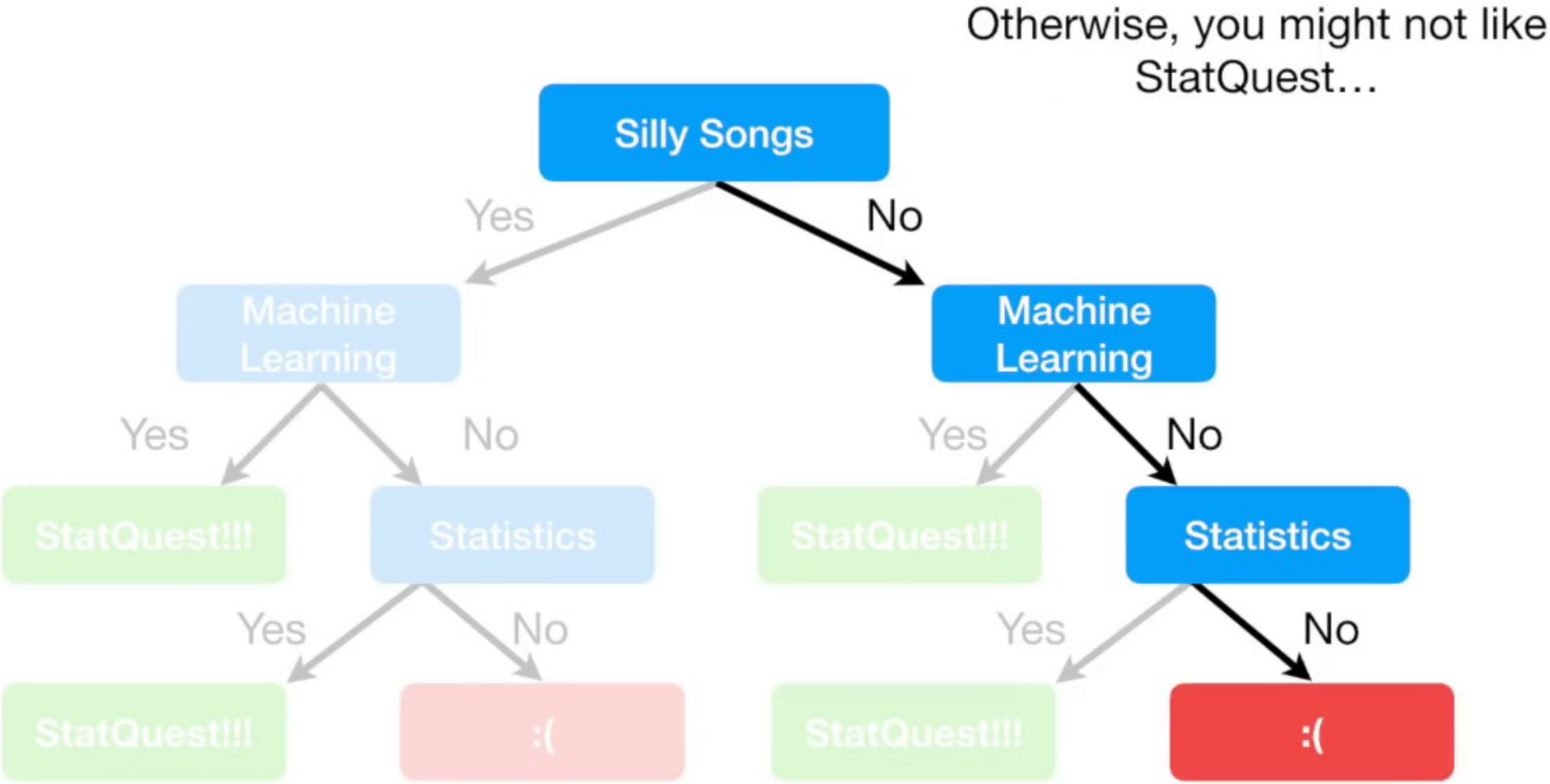


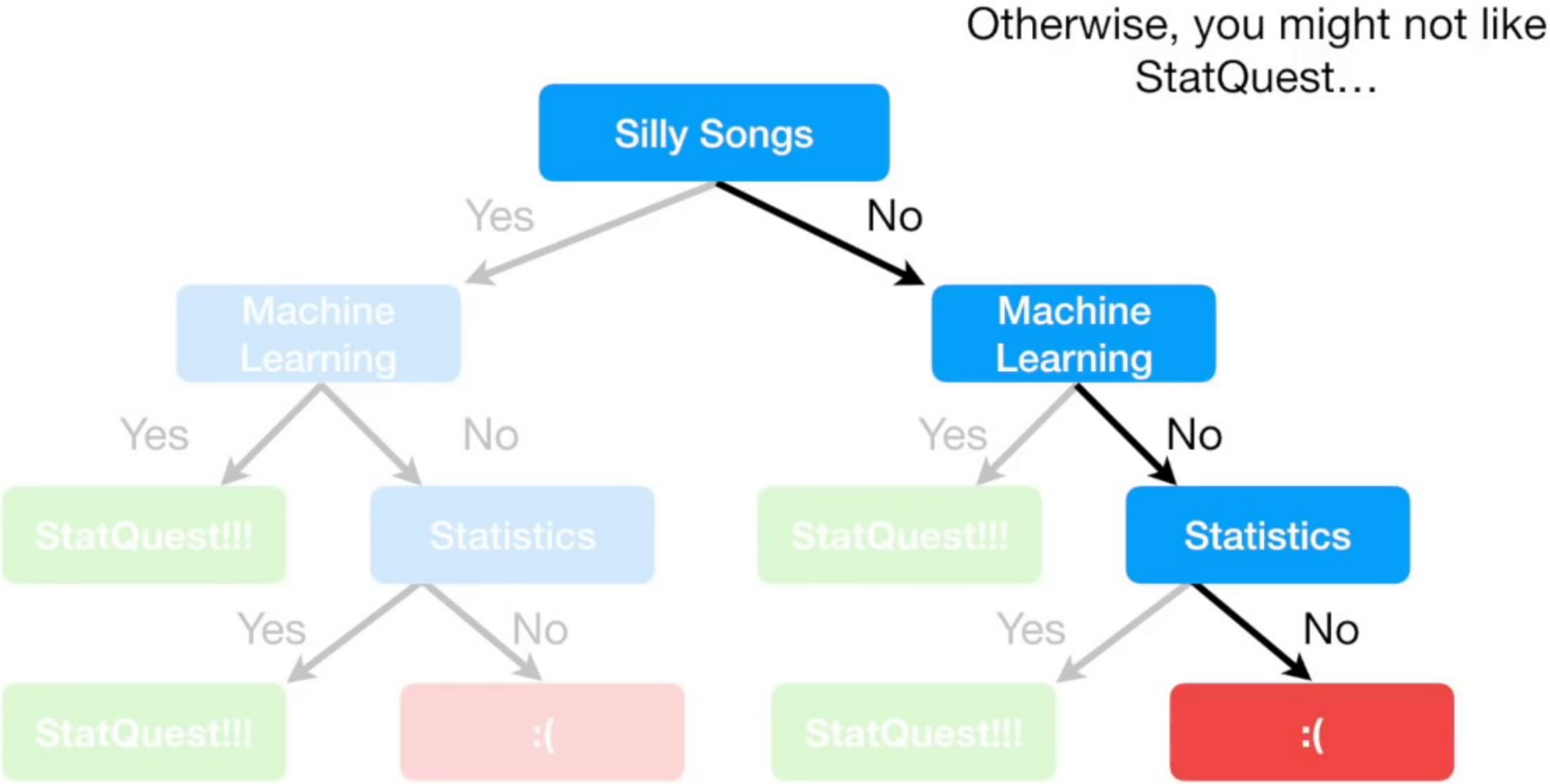
If you ***don't like*** silly songs or machine learning, but you're interested in statistics, then you will love StatQuest!!!



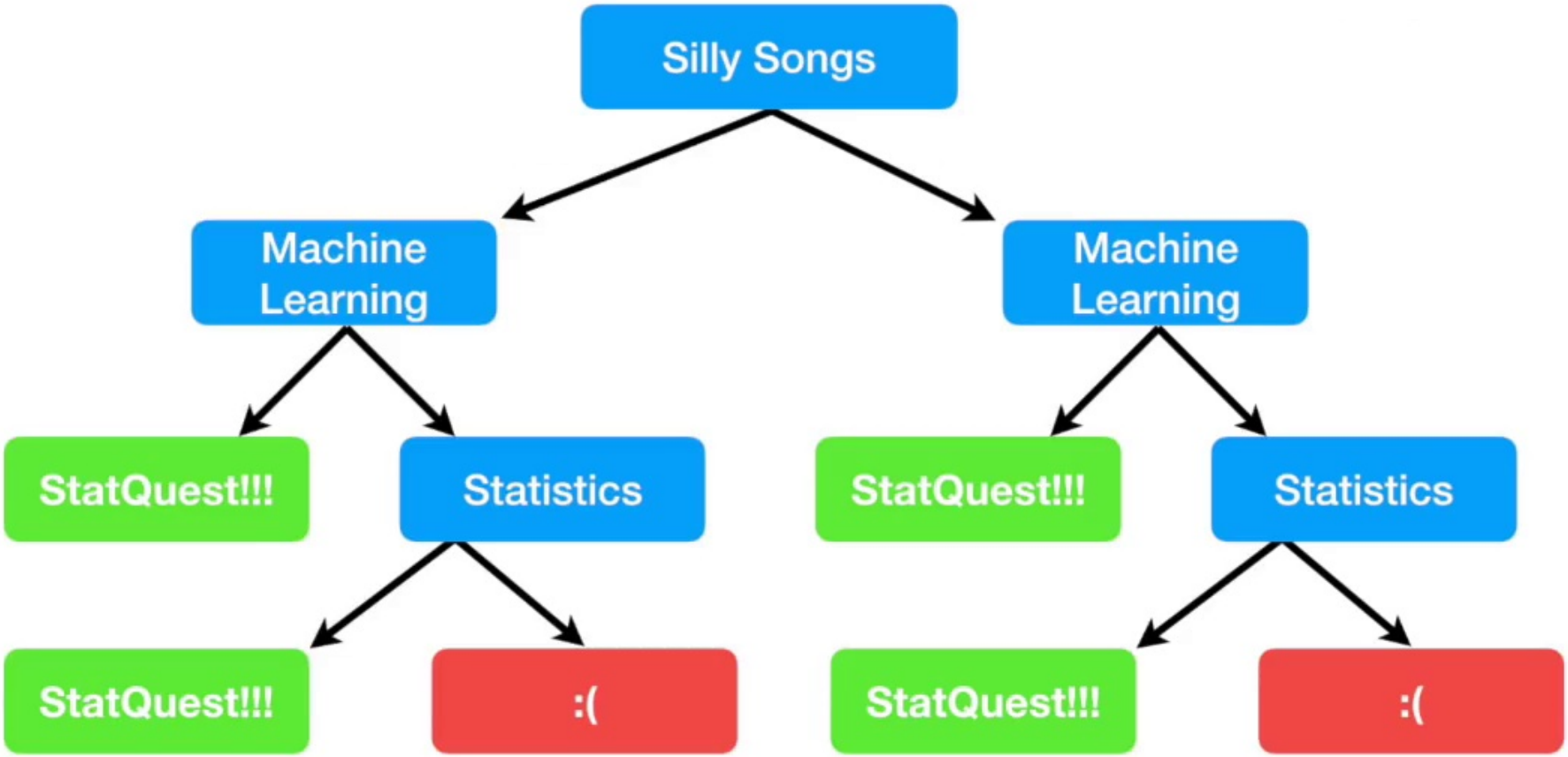
If you ***don't like*** silly songs or machine learning, but you're interested in statistics, then you will love StatQuest!!!



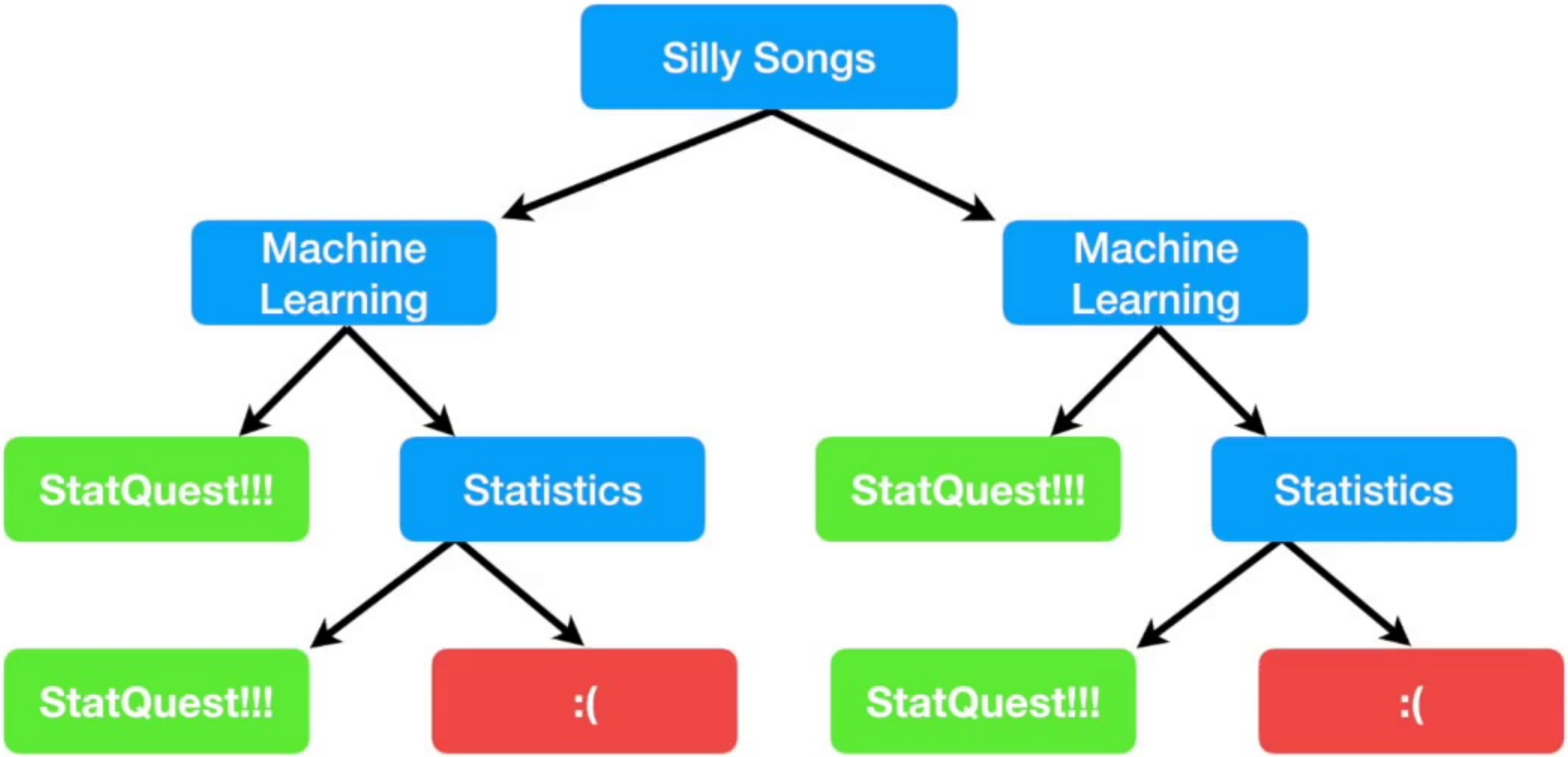

Otherwise, you might not like StatQuest…
Silly Songs
Yes
No
Machine Learning
Machine Learning
Yes
No
Yes
No
StatQuest!!!
Statistics
StatQuest!!!
Statistics
Yes
No
Yes
No
StatQuest!!!
:(
StatQuest!!!
:(


Otherwise, you might not like StatQuest…
Silly Songs
Yes
No
Machine Learning
Machine Learning
Yes
No
Yes
No
StatQuest!!!
Statistics
StatQuest!!!
Statistics
Yes
No
Yes
No
StatQuest!!!
:(
StatQuest!!!
:(

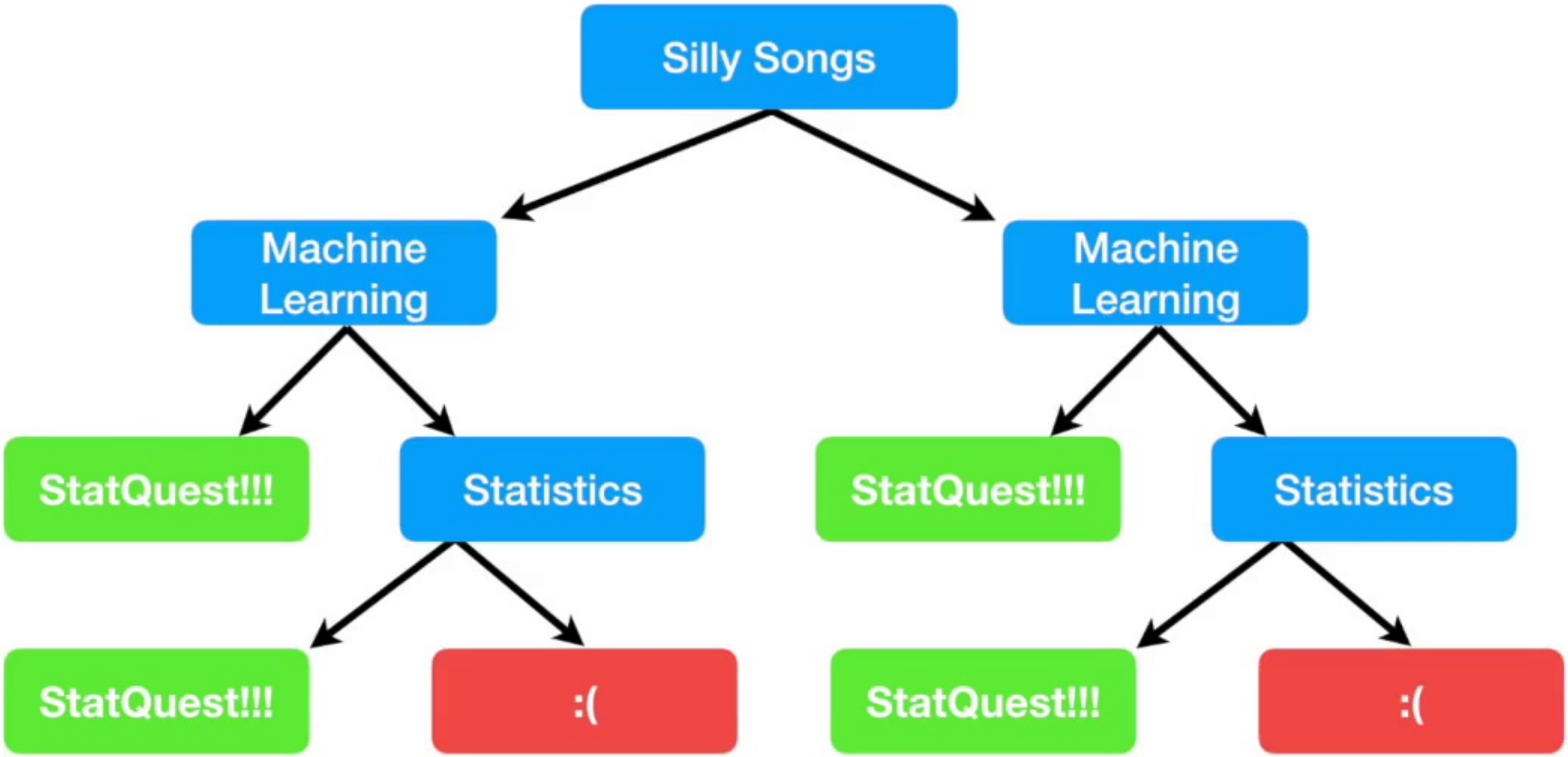
This is a silly example but it illustrates a **Decision Tree**, a simple machine learning method.



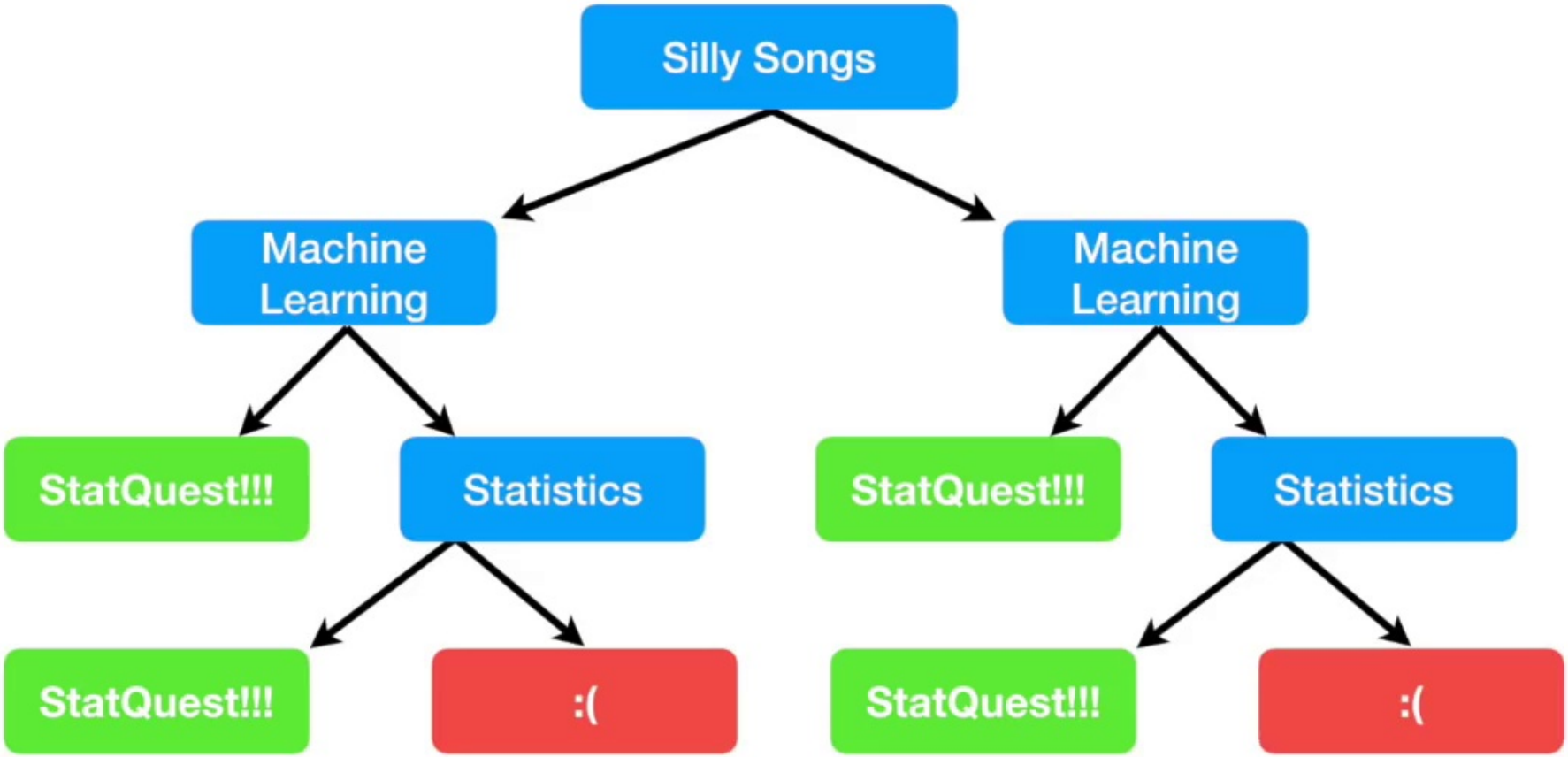
This is a silly example but it illustrates a **Decision Tree**, a simple machine learning method.



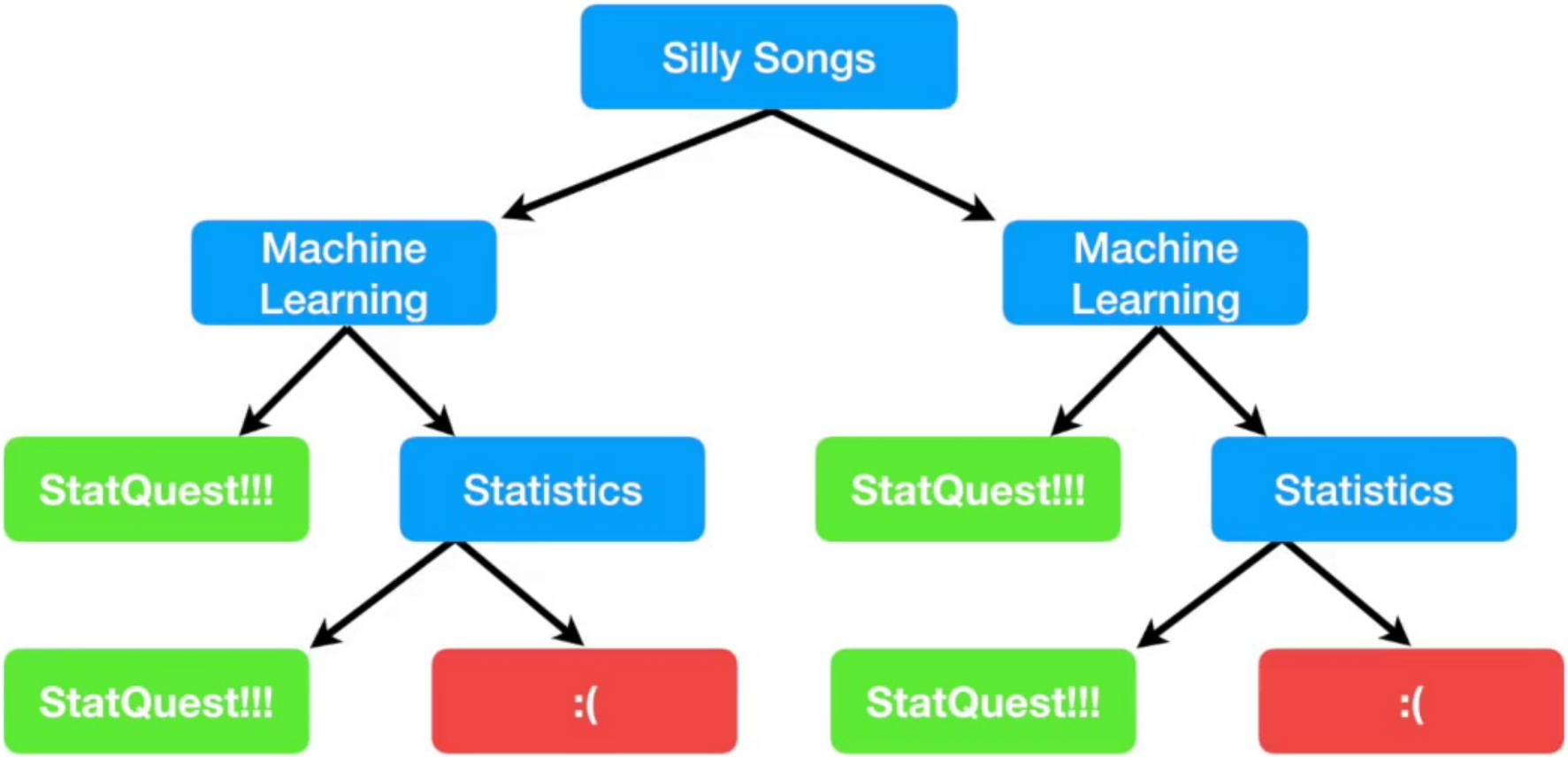
The purpose of this particular **Decision Tree** is to ***predict*** whether or not someone will love StatQuest.



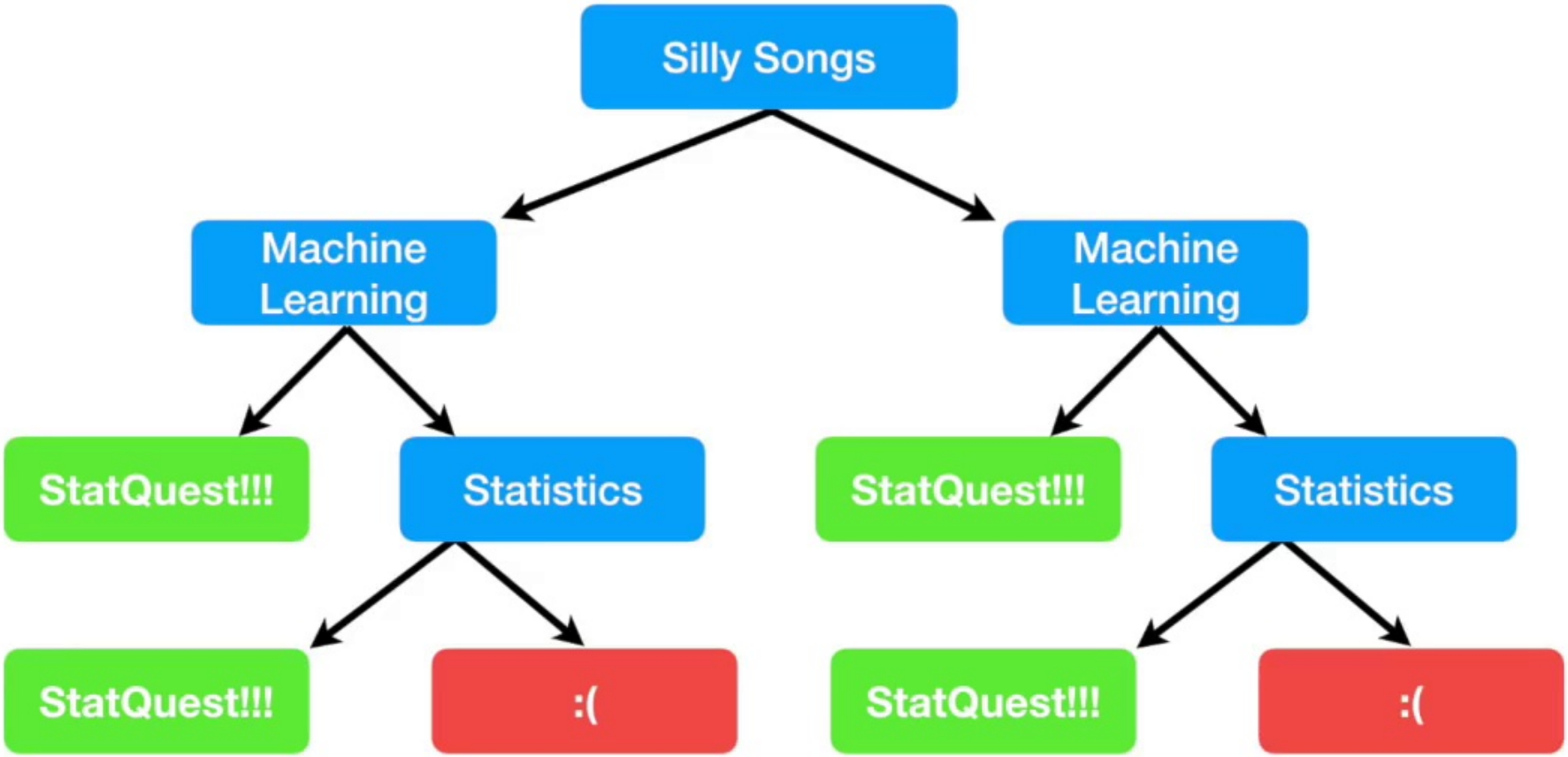
The purpose of this particular **Decision Tree** is to ***predict*** whether or not someone will love StatQuest.



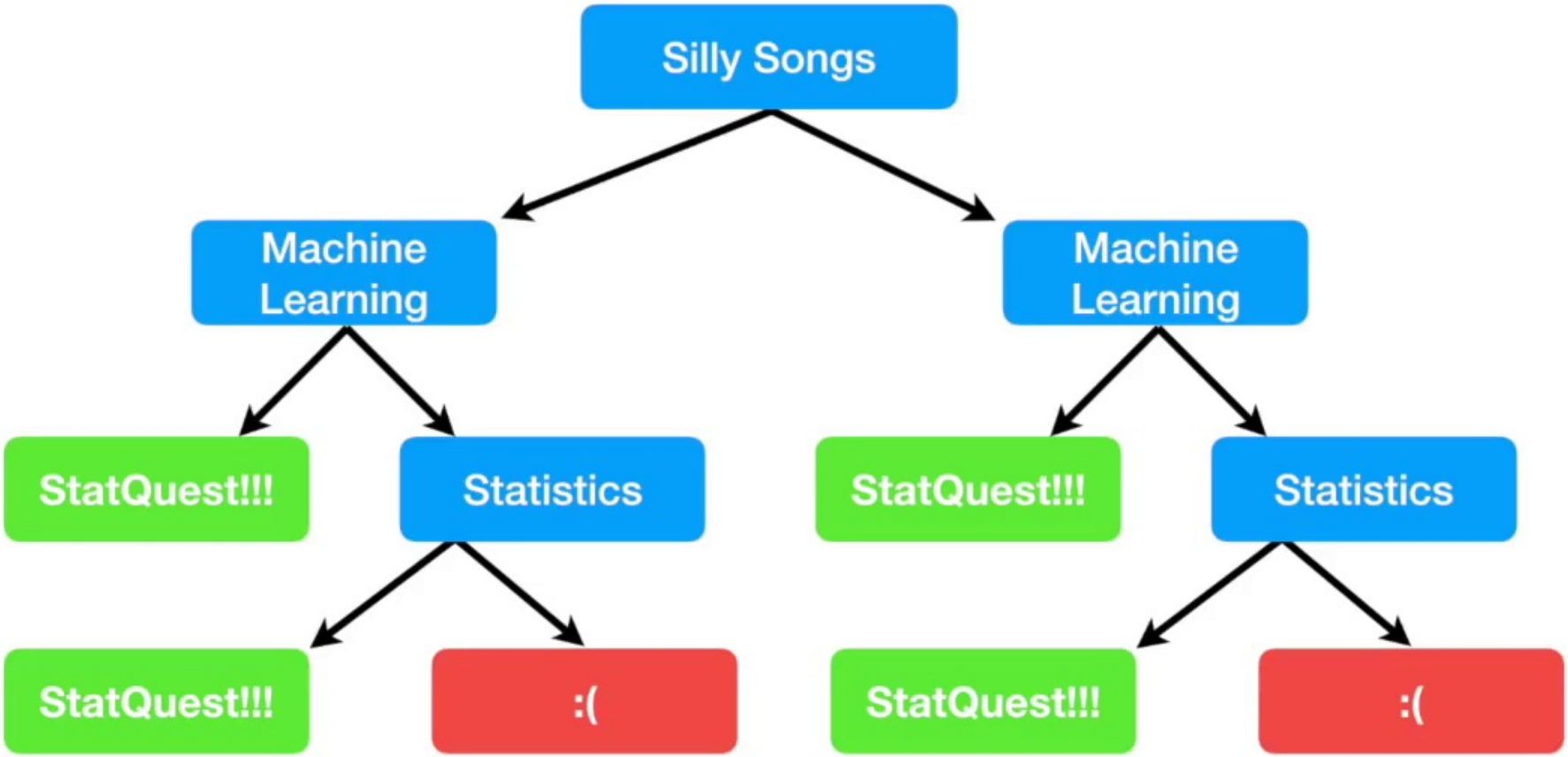
The purpose of this particular **Decision Tree** is to ***predict*** whether or not someone will love StatQuest.



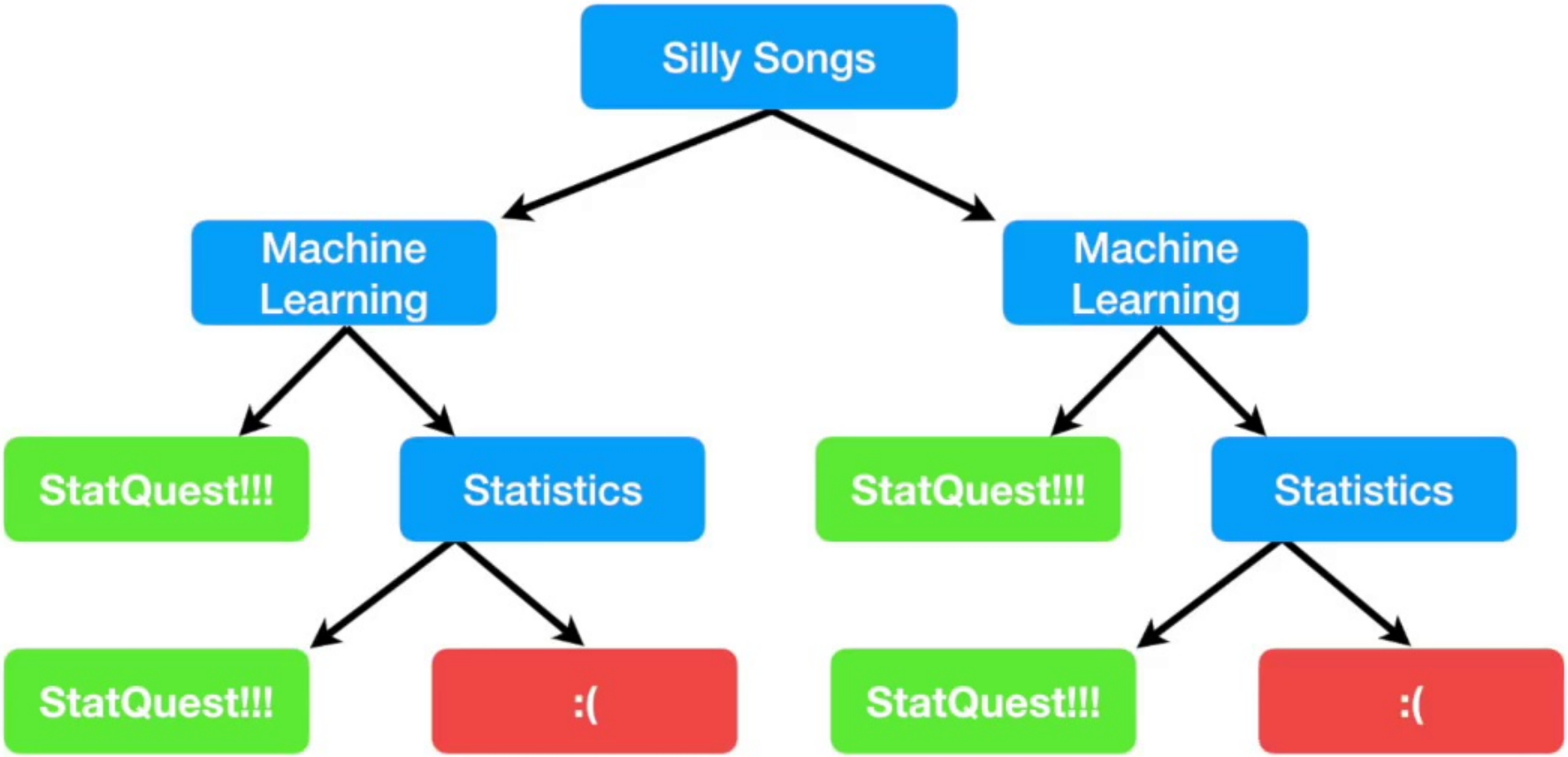
Alternatively, we could say that this **Decision Tree** ***classifies*** a person as either someone who loves StatQuest or someone who doesn't.



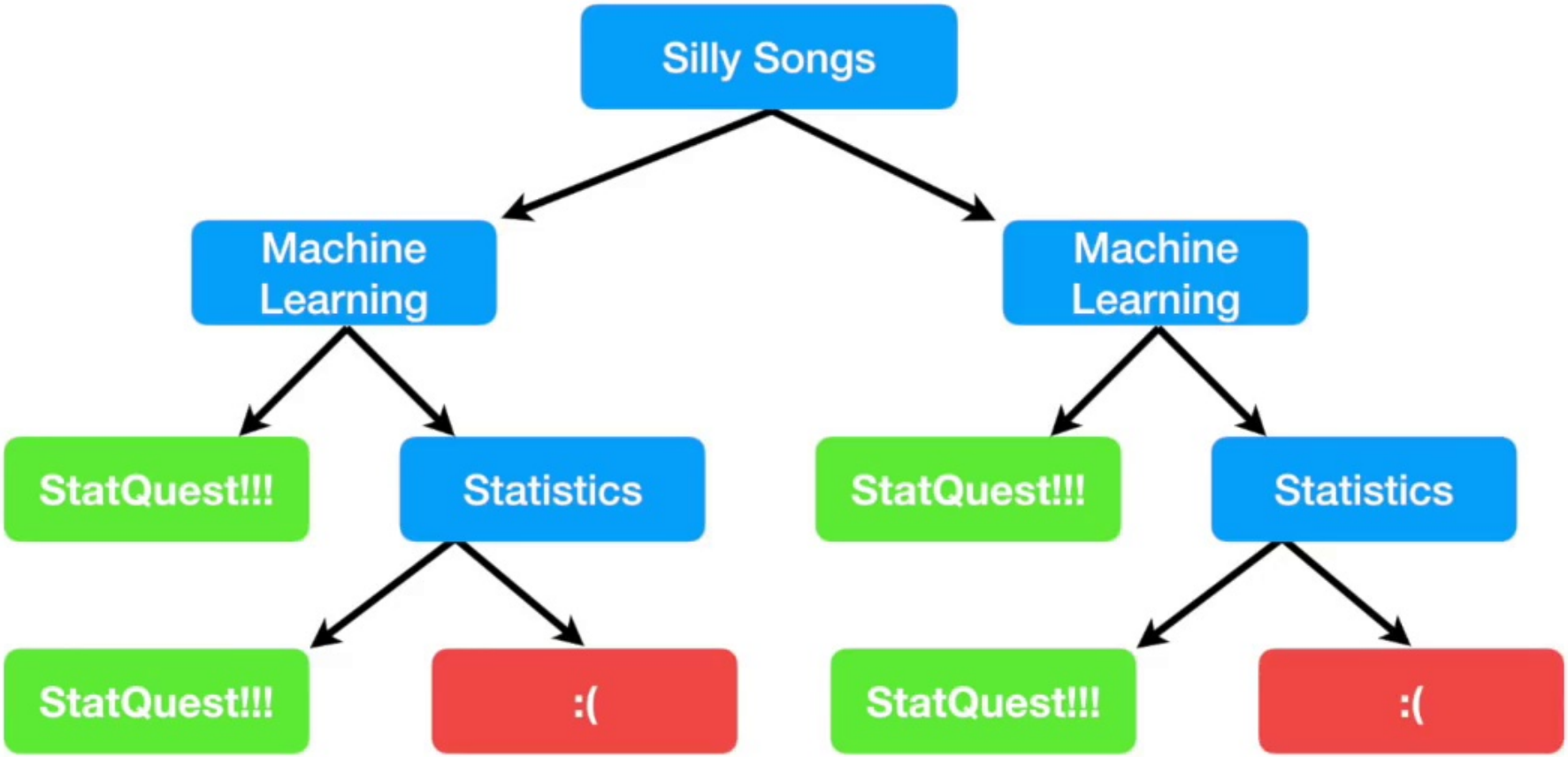
Alternatively, we could say that this **Decision Tree** ***classifies*** a person as either someone who loves StatQuest or someone who doesn't.



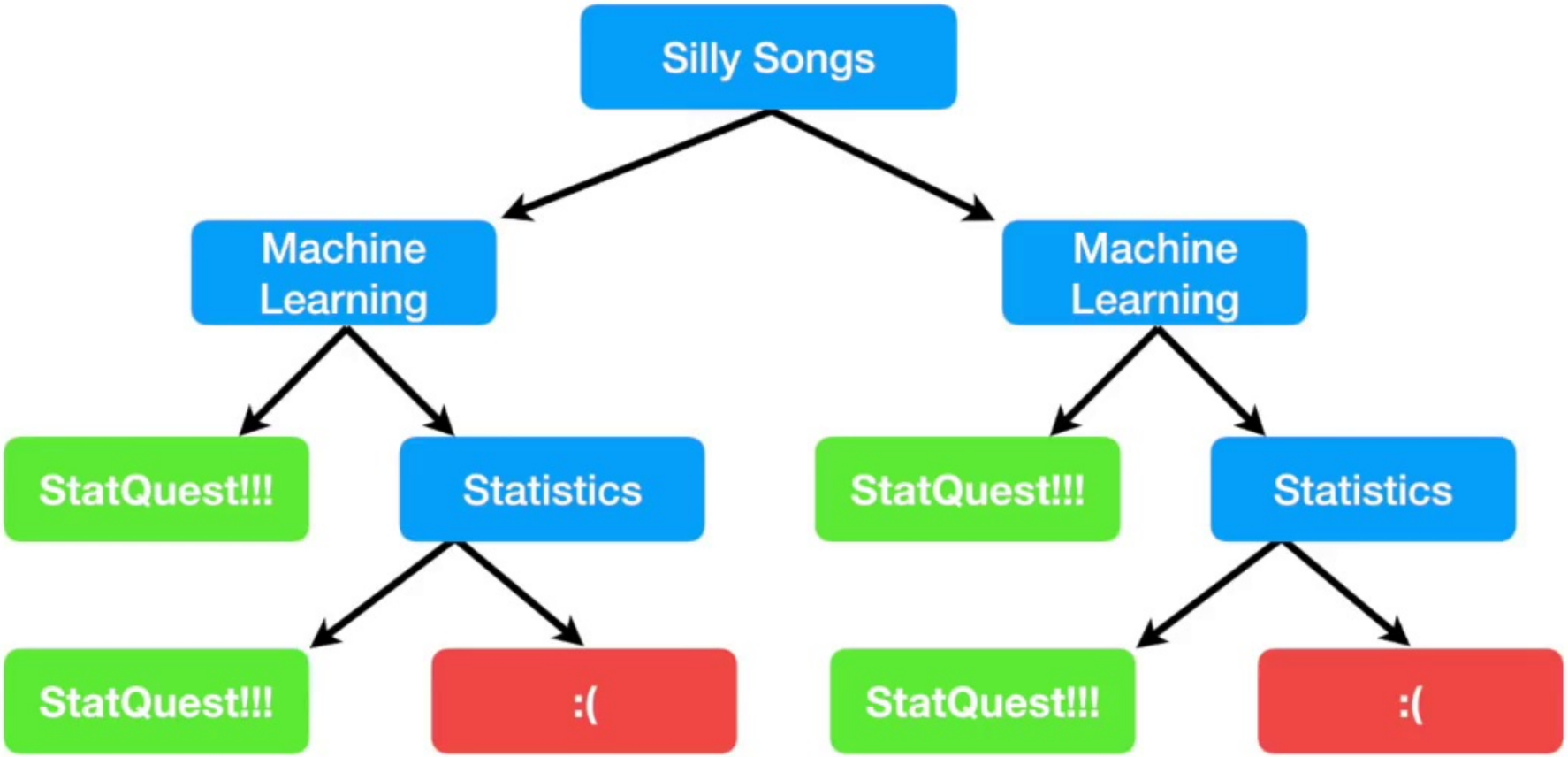
Since **Decision Trees** are type of
Machine Learning…

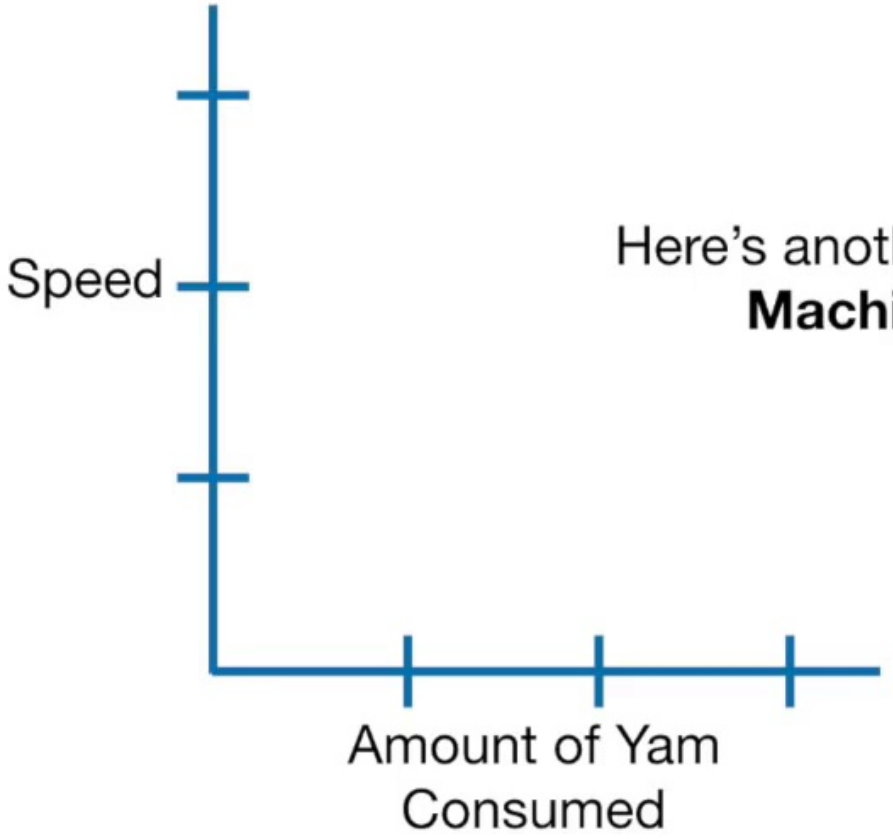


Since **Decision Trees** are type of **Machine Learning**…

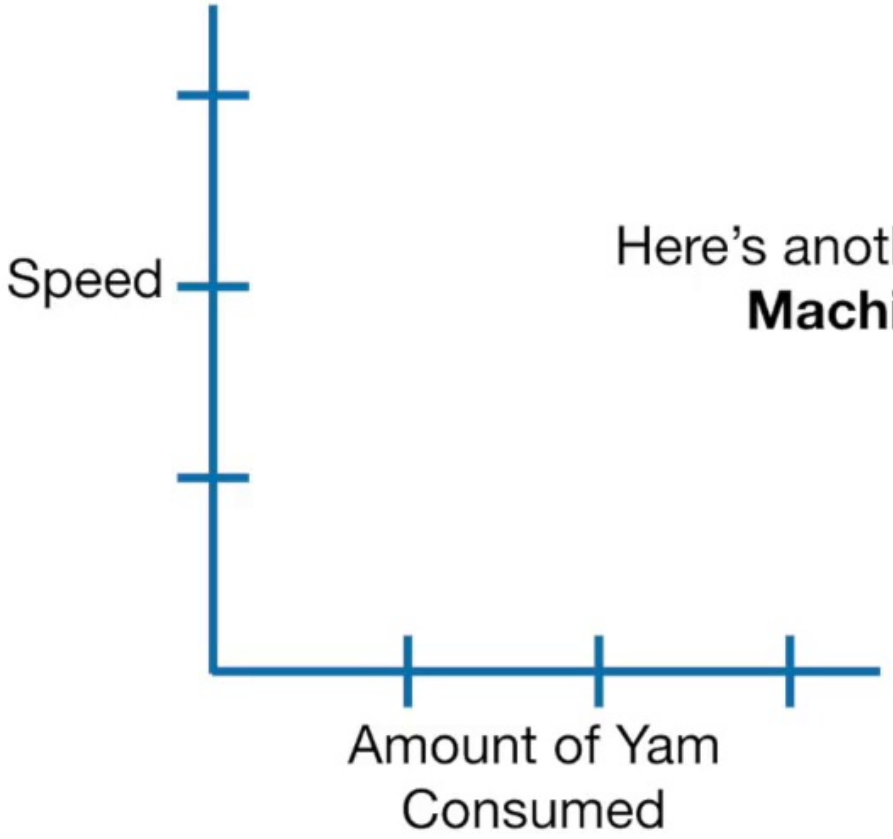


…then if you understand how we used this tree to ***predict*** or ***classify*** if someone would love StatQuest, you are well on your way to understanding **Machine Learning**.

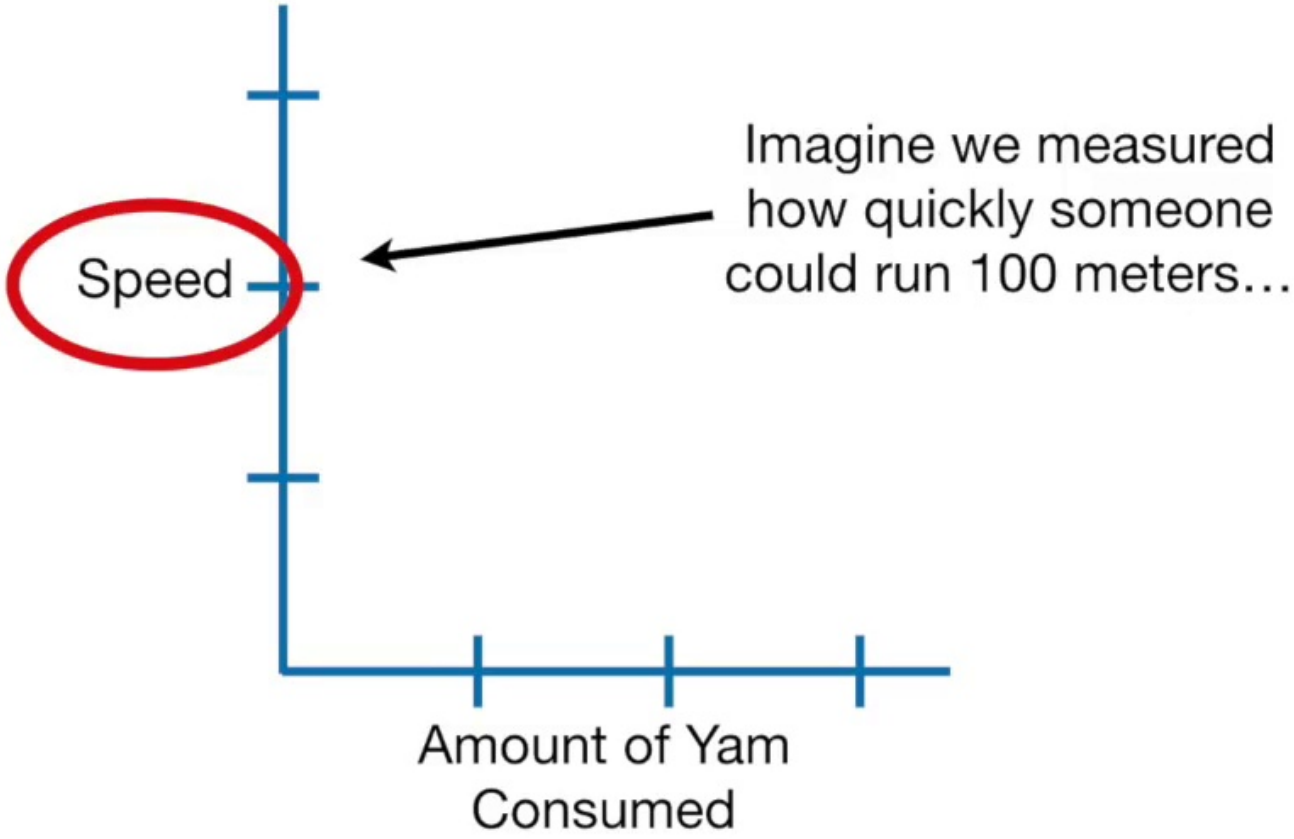


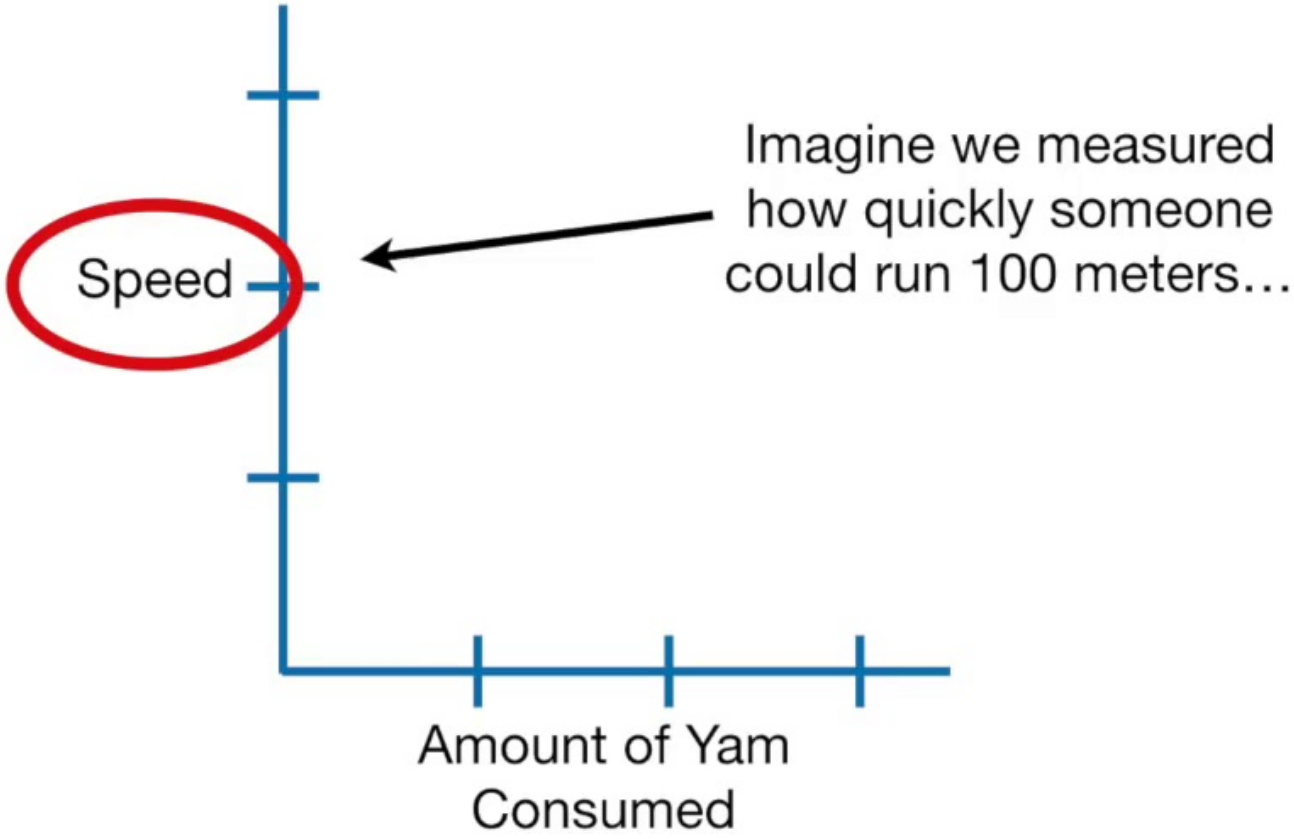


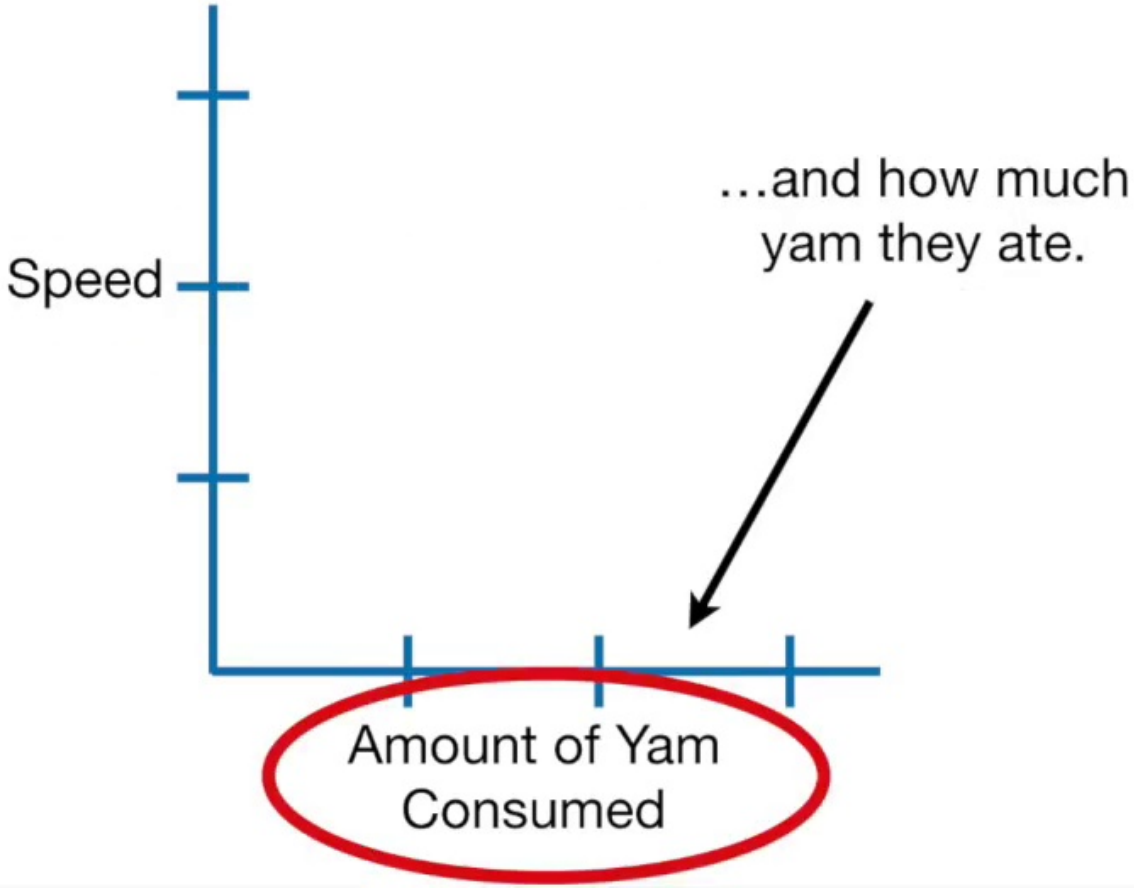
Here's another silly example of
Machine Learning.

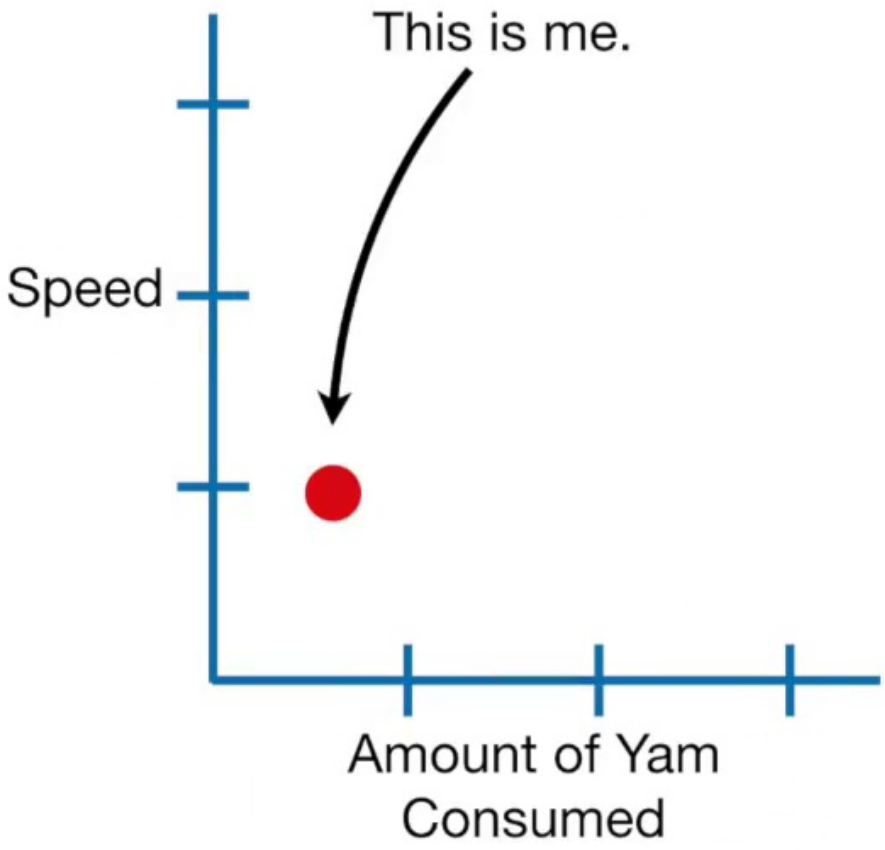


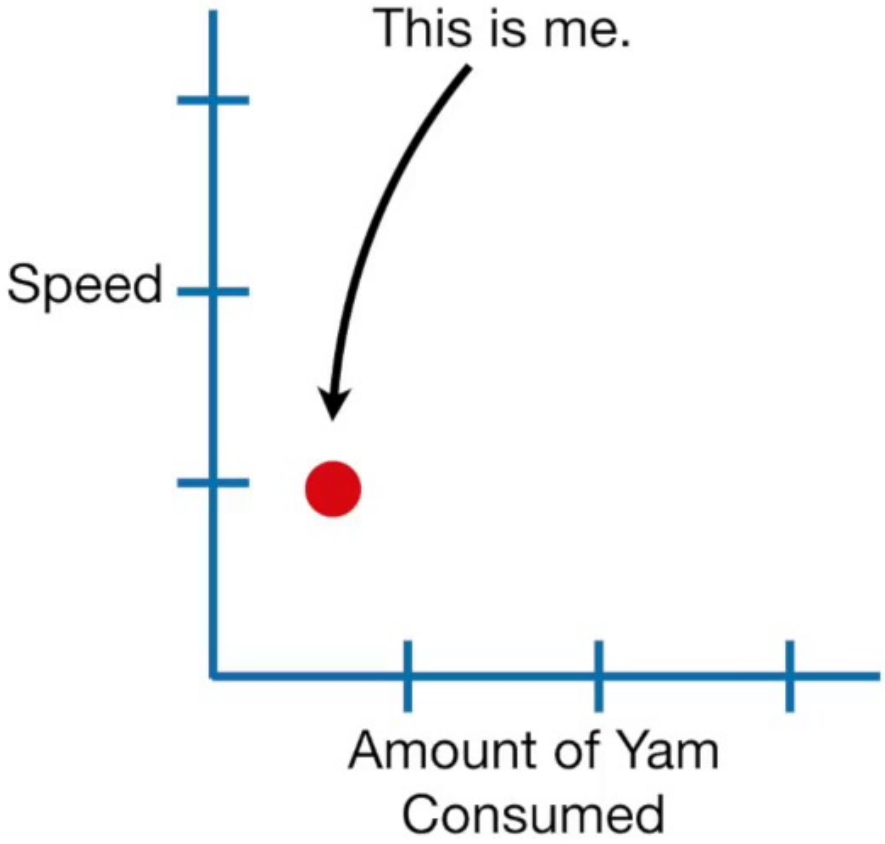
Here's another silly example of
Machine Learning.

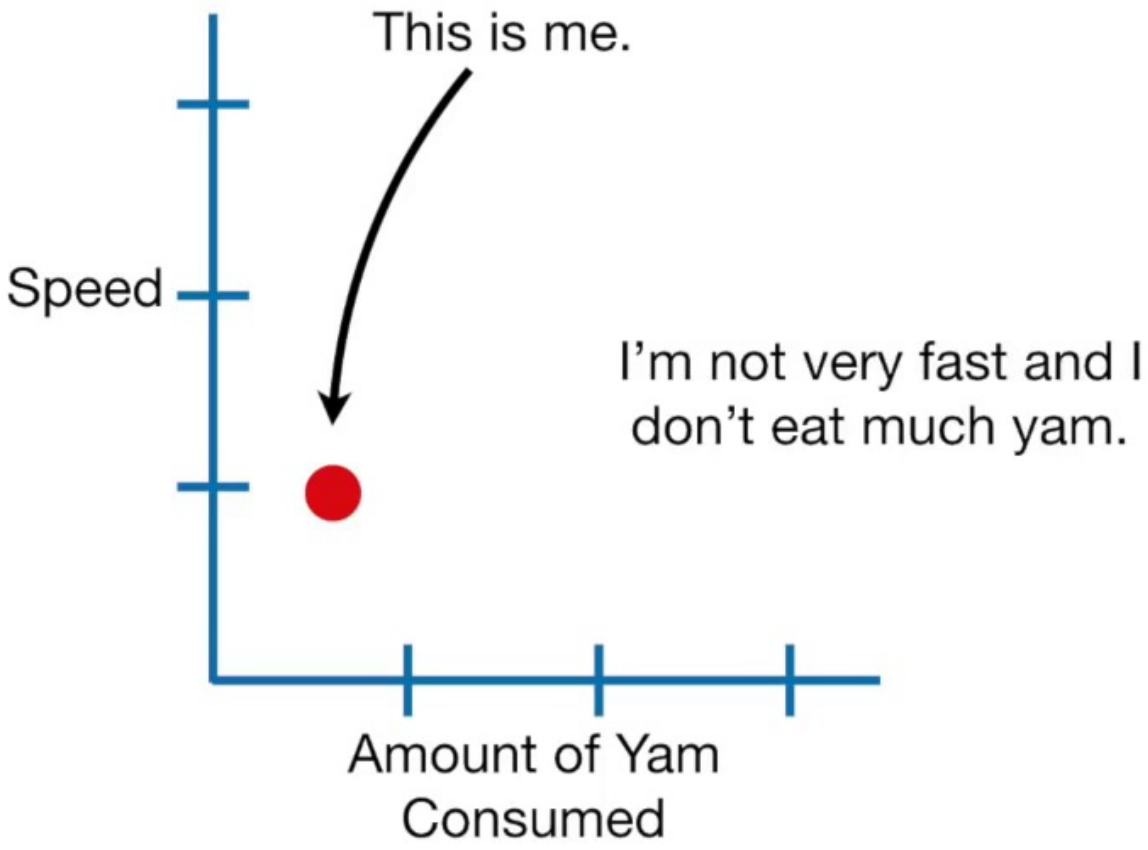

Speed
Imagine we measured how quickly someone could run 100 meters…
Amount of Yam Consumed

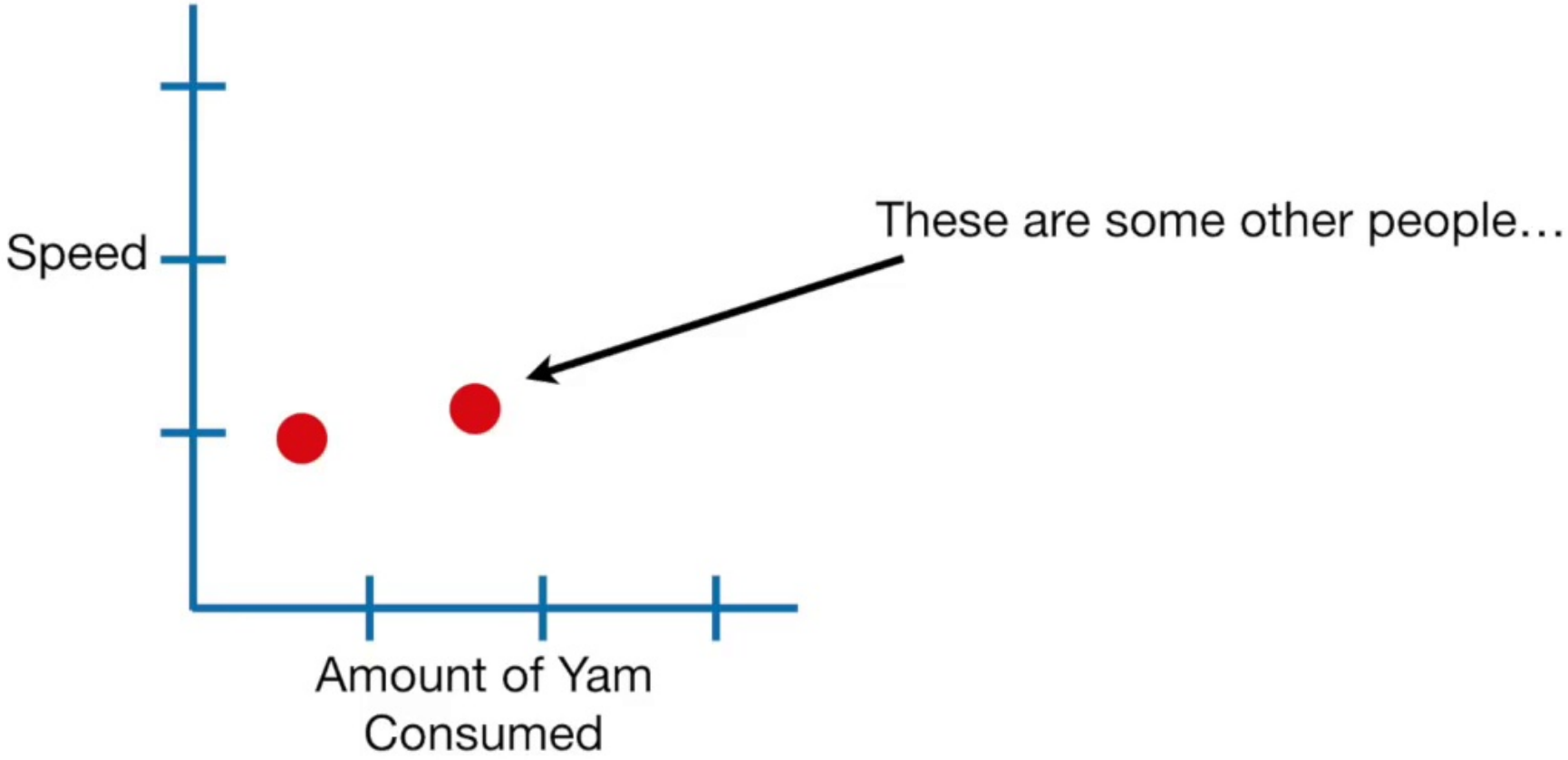

Speed
Imagine we measured how quickly someone could run 100 meters…
Amount of Yam Consumed

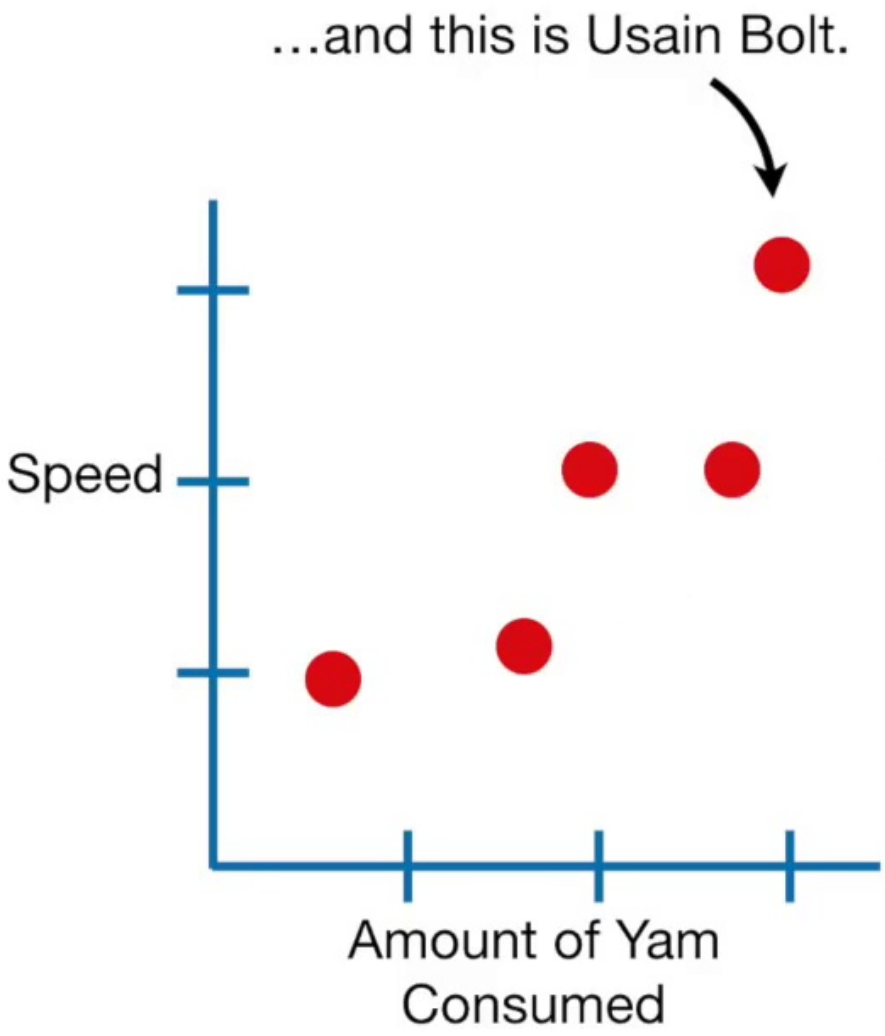

Speed
…and how much
yam they ate.
Amount of Yam
Consumed

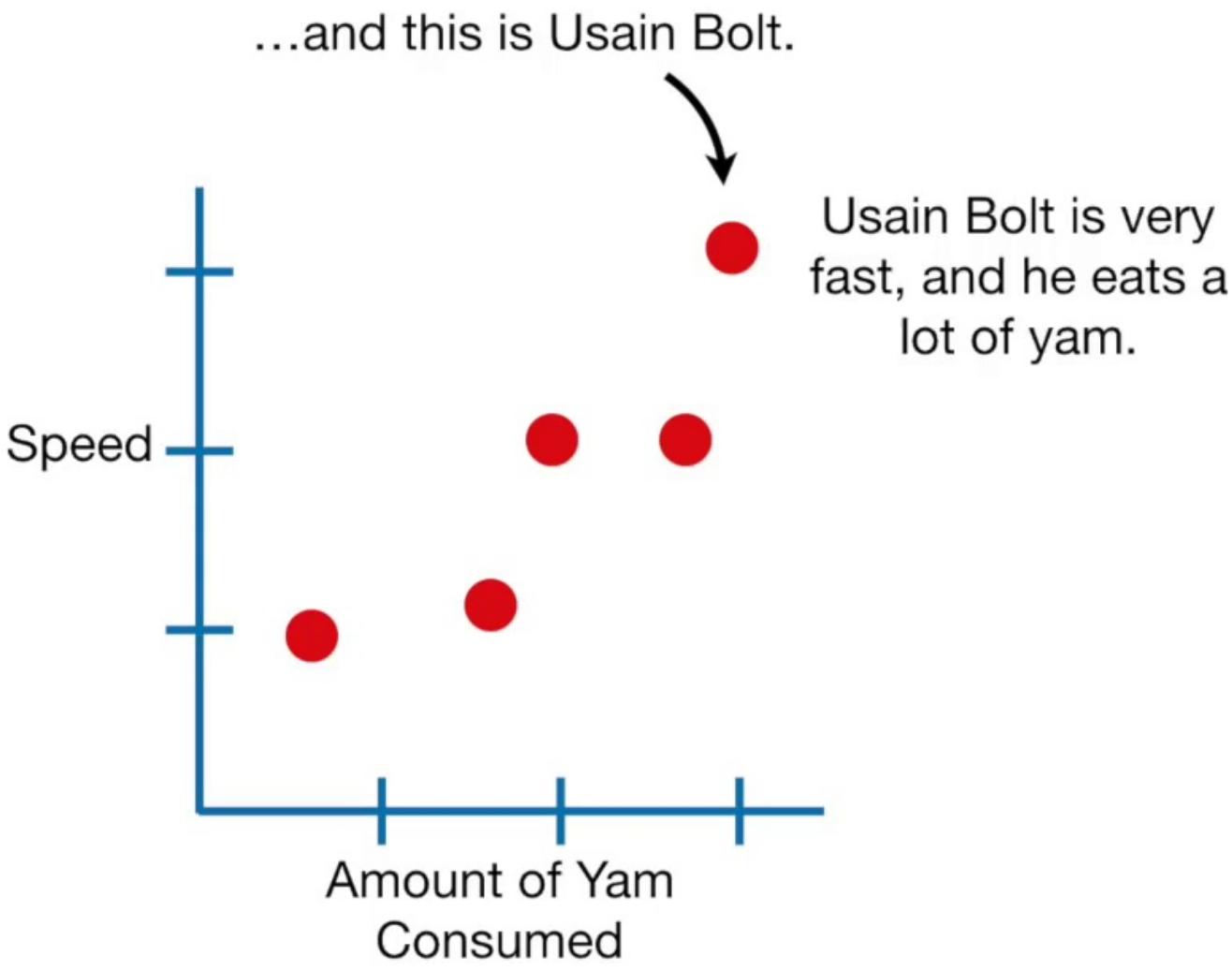

This is me.
Speed
Amount of Yam
Consumed

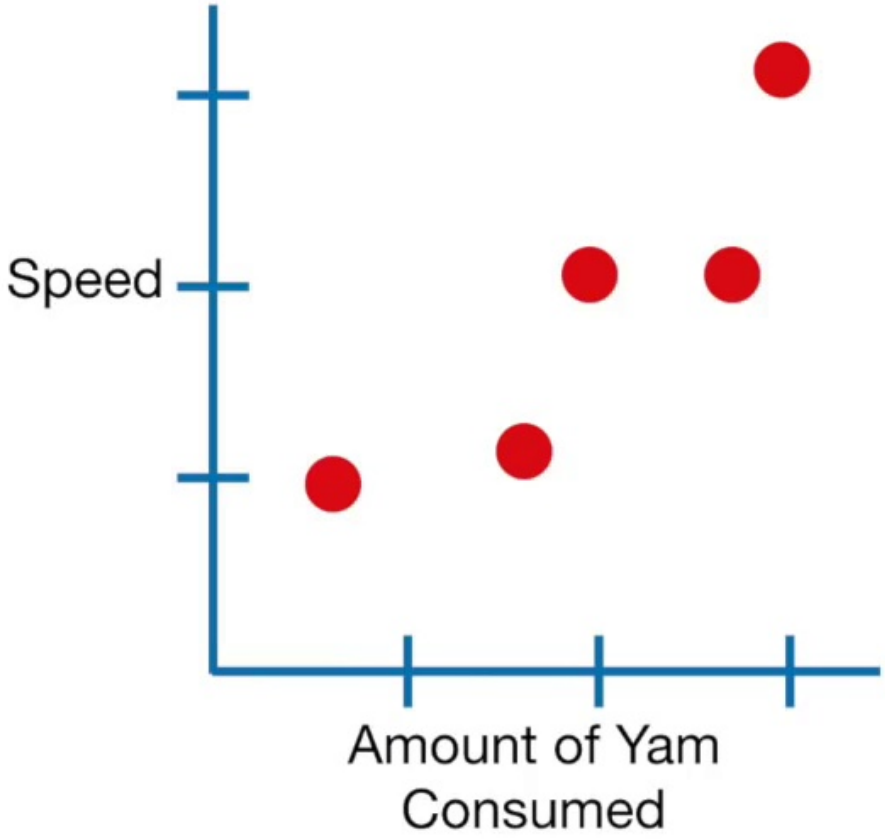

This is me.
Speed
Amount of Yam
Consumed


This is me.
Speed
I'm not very fast and I don't eat much yam.
Amount of Yam Consumed

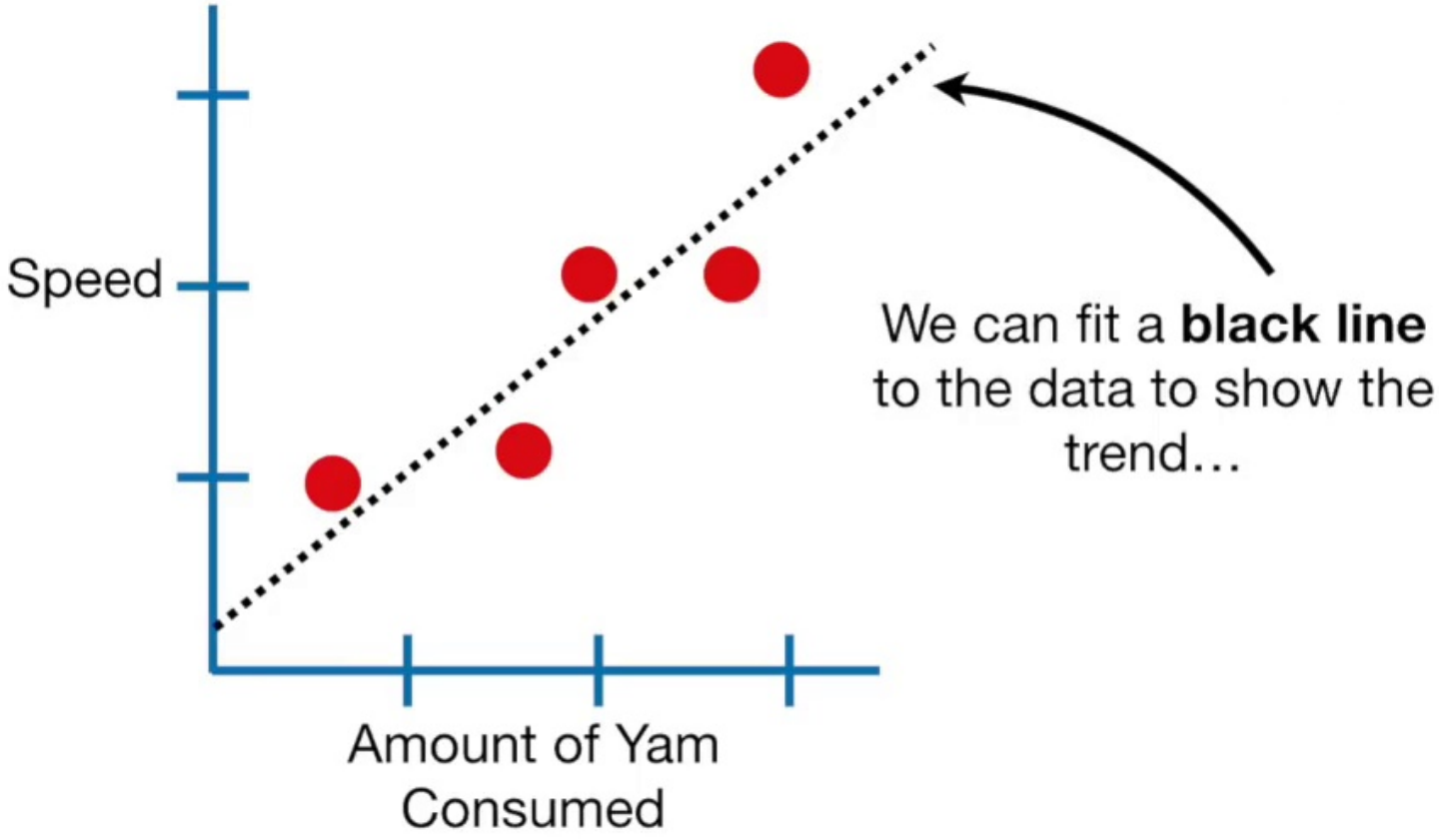

Speed
These are some other people…
Amount of Yam
Consumed

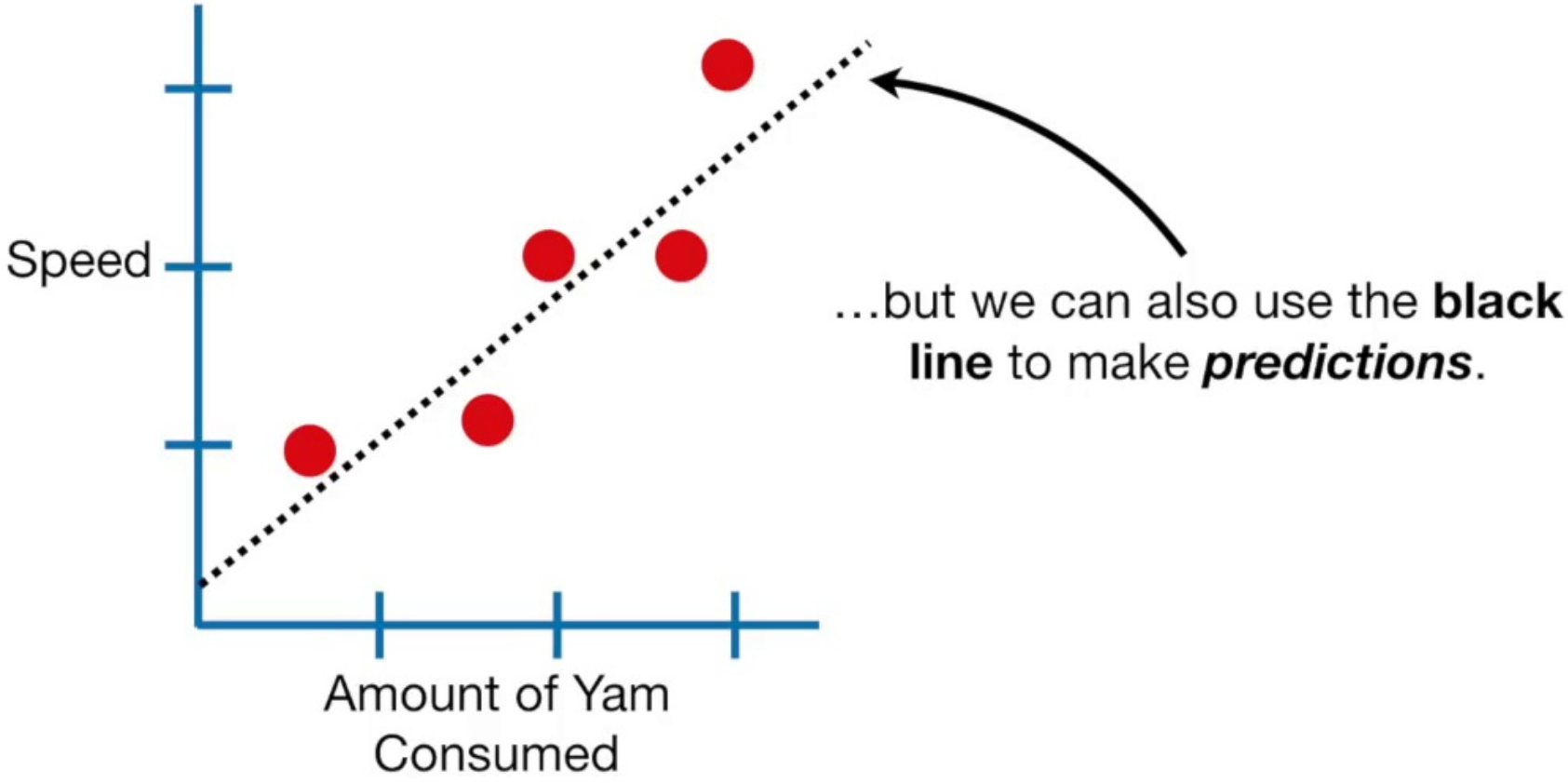

…and this is Usain Bolt.
Speed
Amount of Yam
Consumed

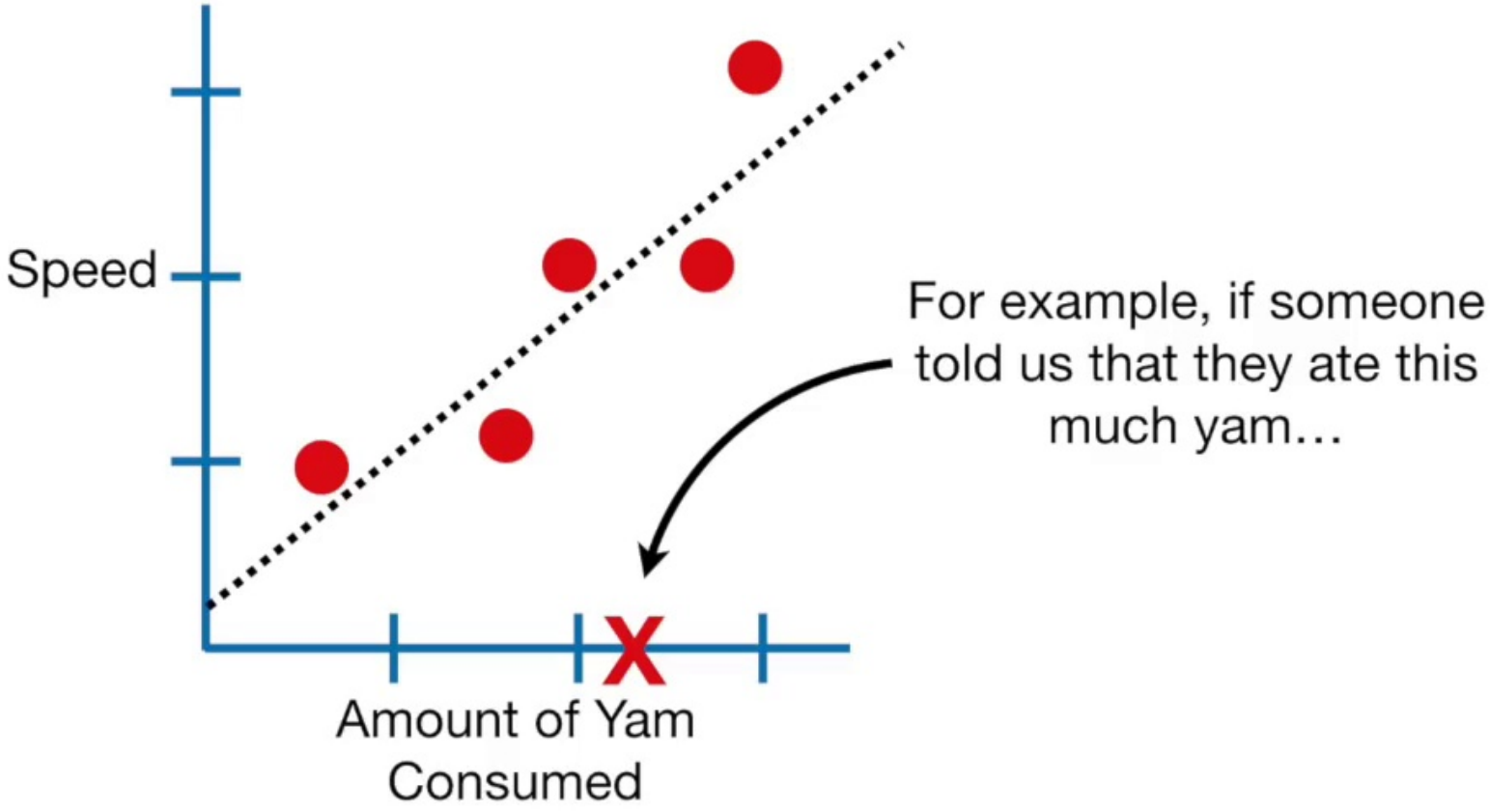

…and this is Usain Bolt.
Usain Bolt is very fast, and he eats a lot of yam.
Speed
Amount of Yam Consumed

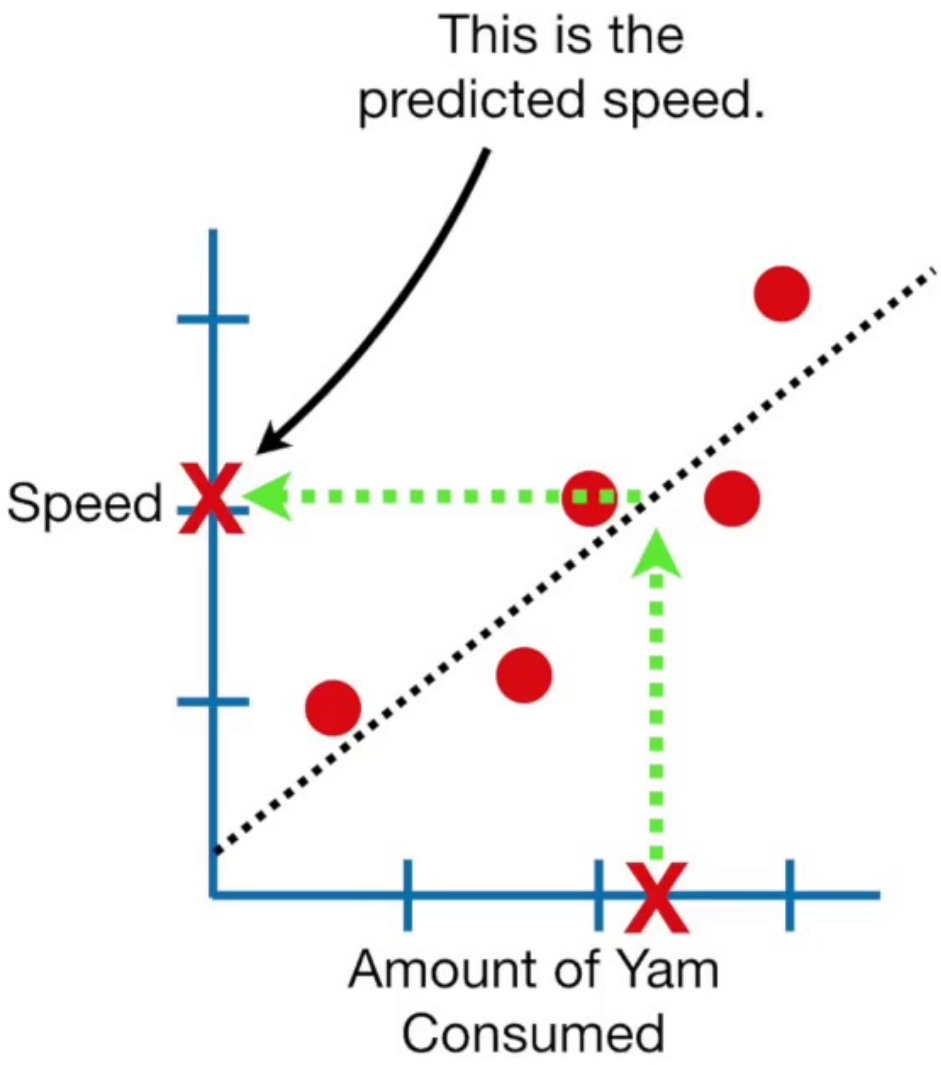


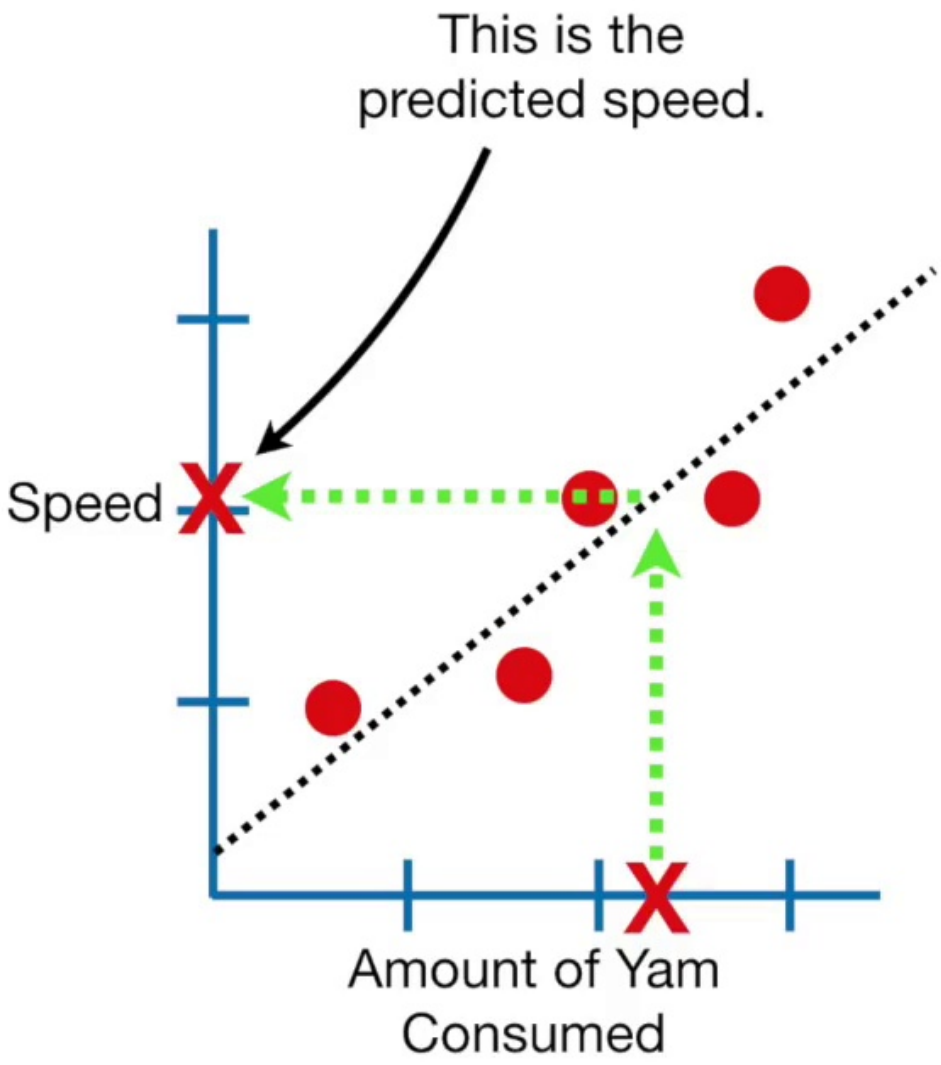
Given this pretend data, we see that the more yam someone eats, the faster they run the 100 meter dash.

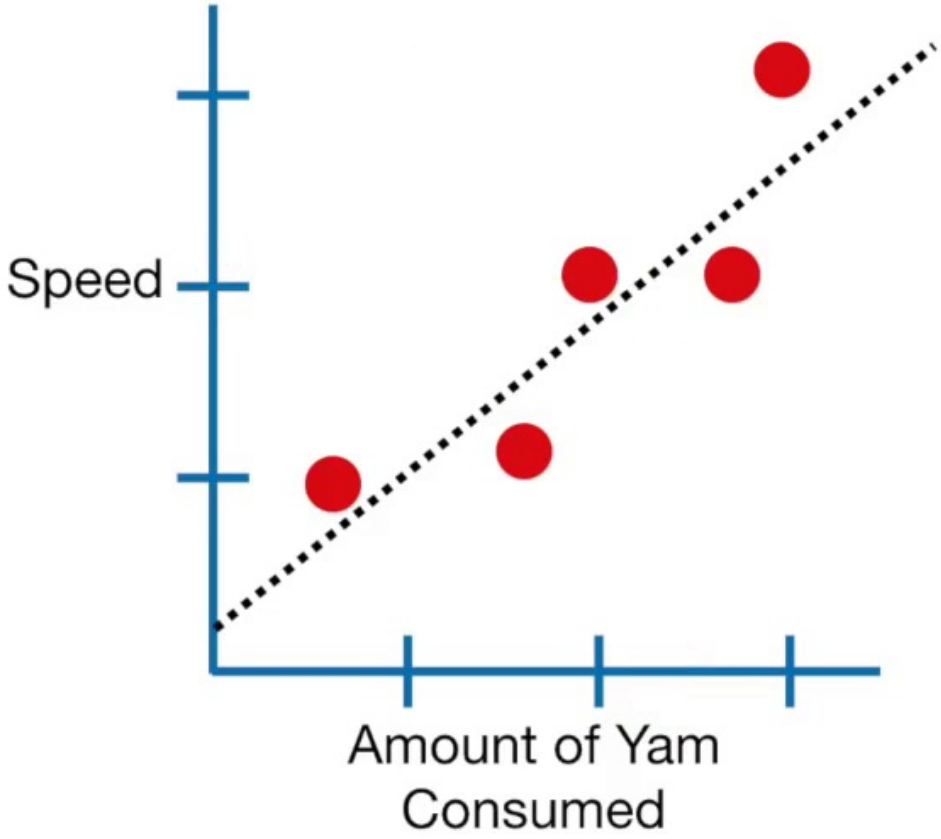

Speed
Amount of Yam Consumed
We can fit a **black line** to the data to show the trend…


Speed
Amount of Yam Consumed
…but we can also use the **black line** to make ***predictions***.

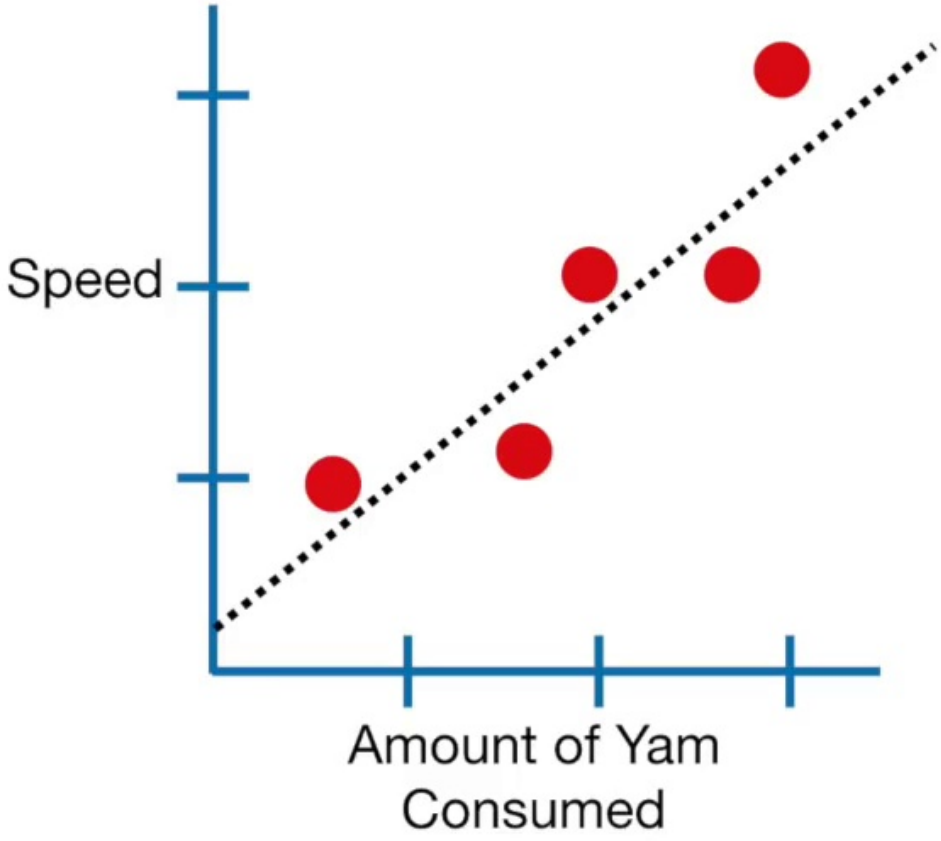

Speed
For example, if someone told us that they ate this much yam…
Amount of Yam Consumed


This is the
predicted speed.
Speed
Amount of Yam
Consumed


This is the
predicted speed.
Speed
Amount of Yam
Consumed

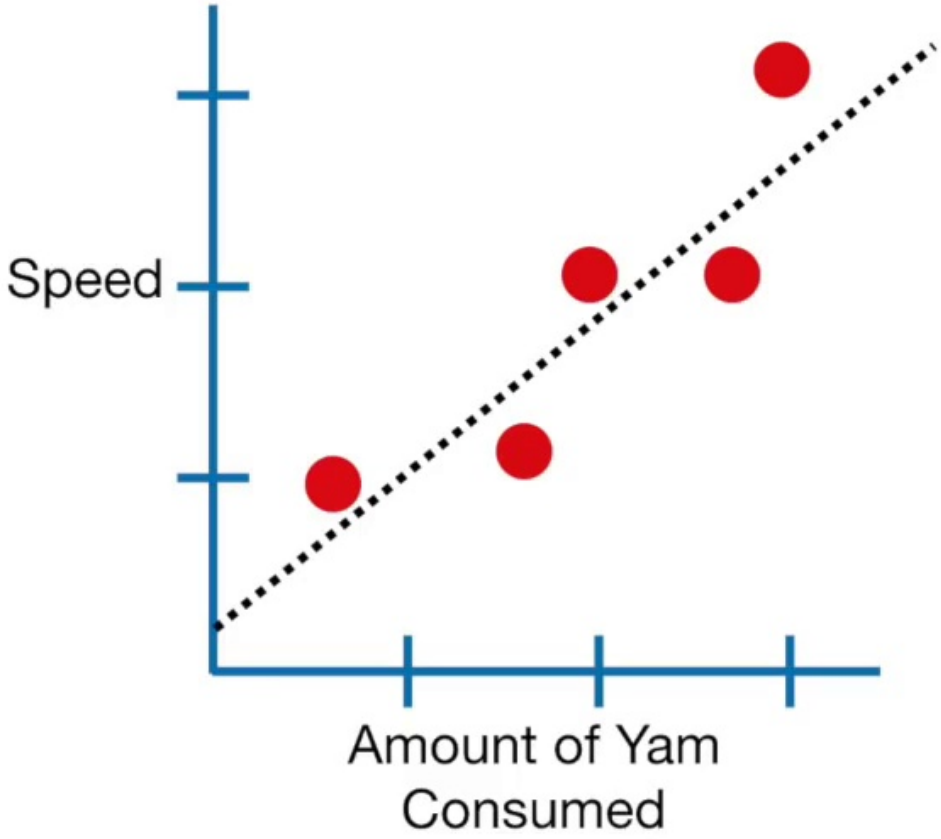


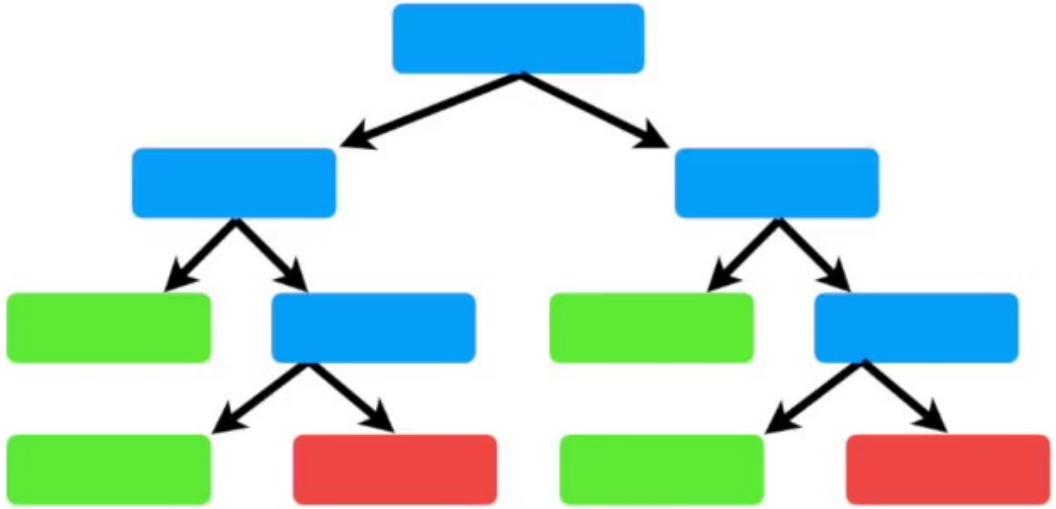
The **black line** is a type of **Machine Learning** because we can use it to make ***predictions***.



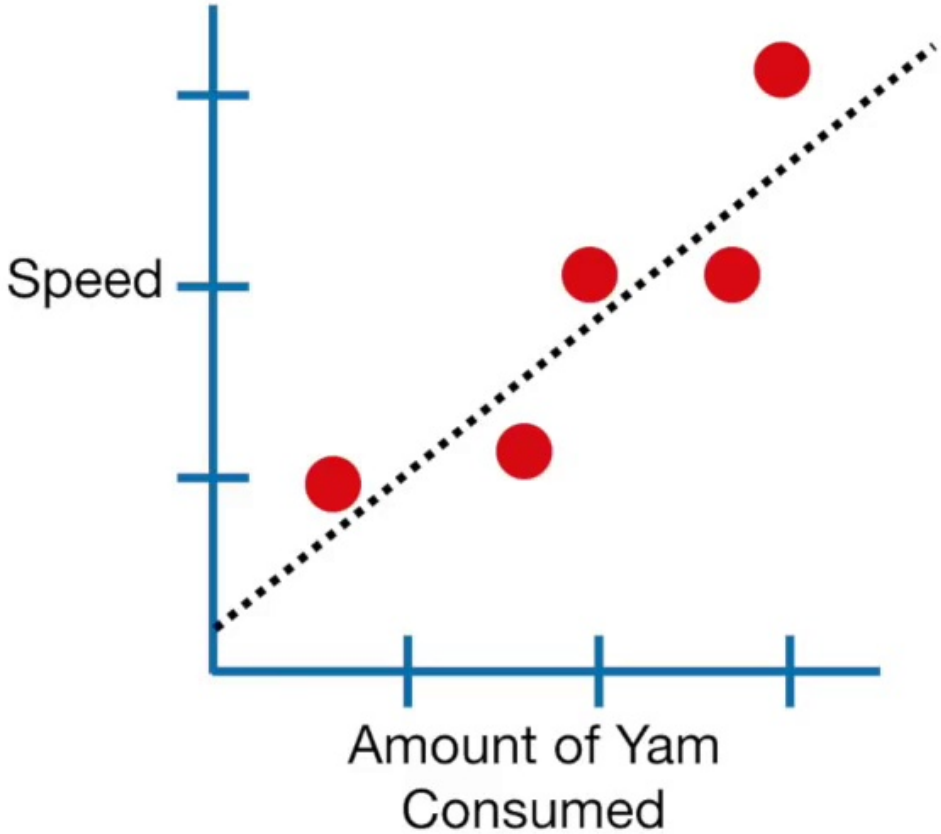
The **black line** is a type of **Machine Learning** because we can use it to make ***predictions***.

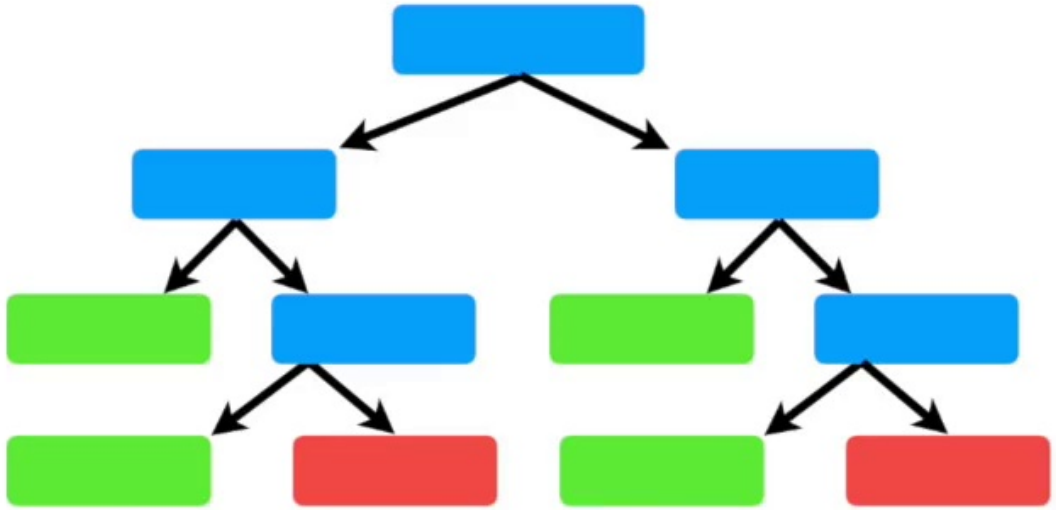
In general, **Machine Learning** is all about making ***predictions*** and ***classifications***.



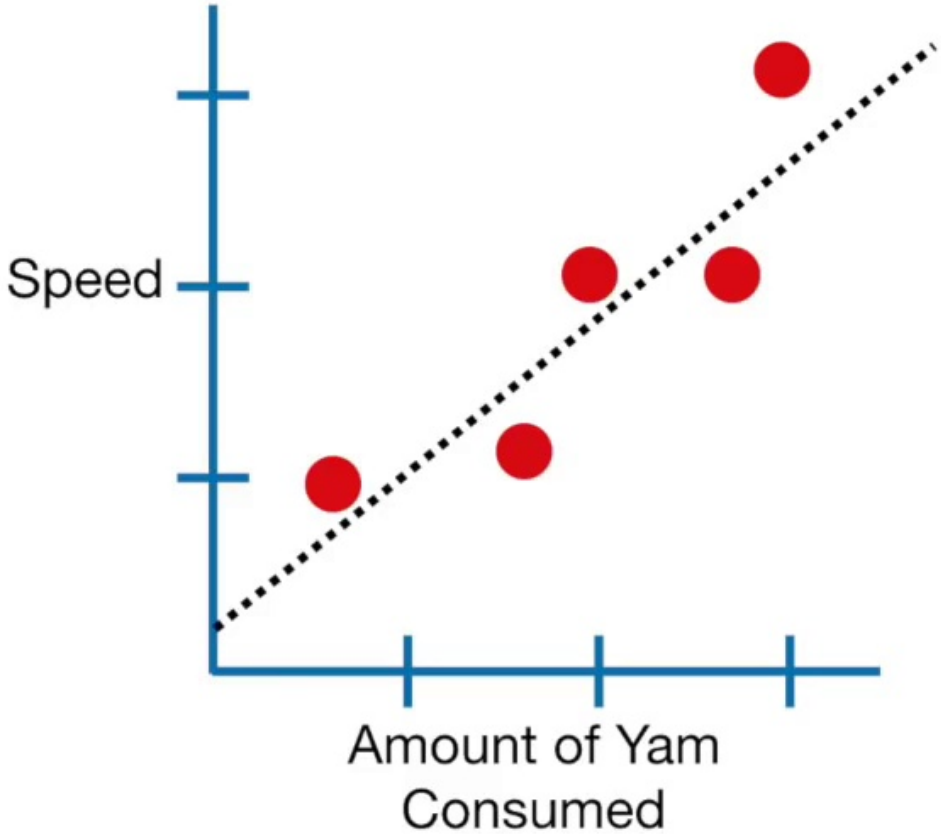


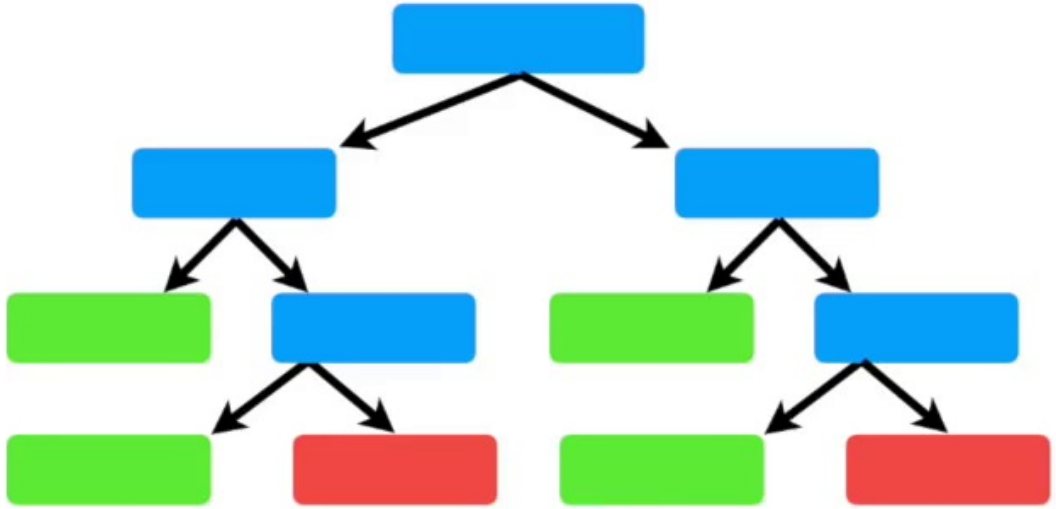
In general, **Machine Learning** is all about making ***predictions*** and ***classifications***.



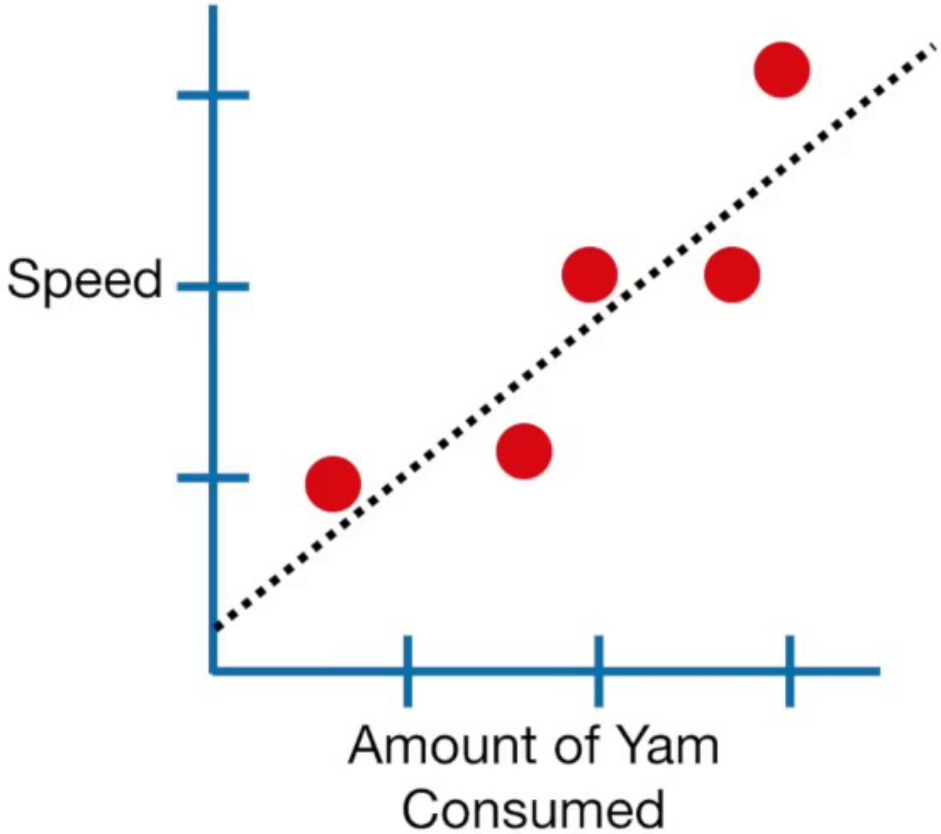


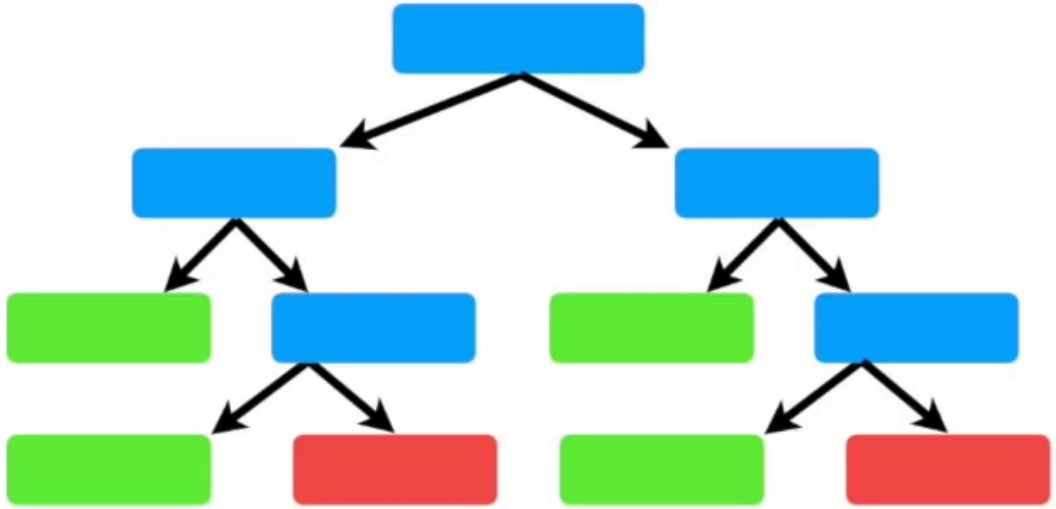
BAM!!!



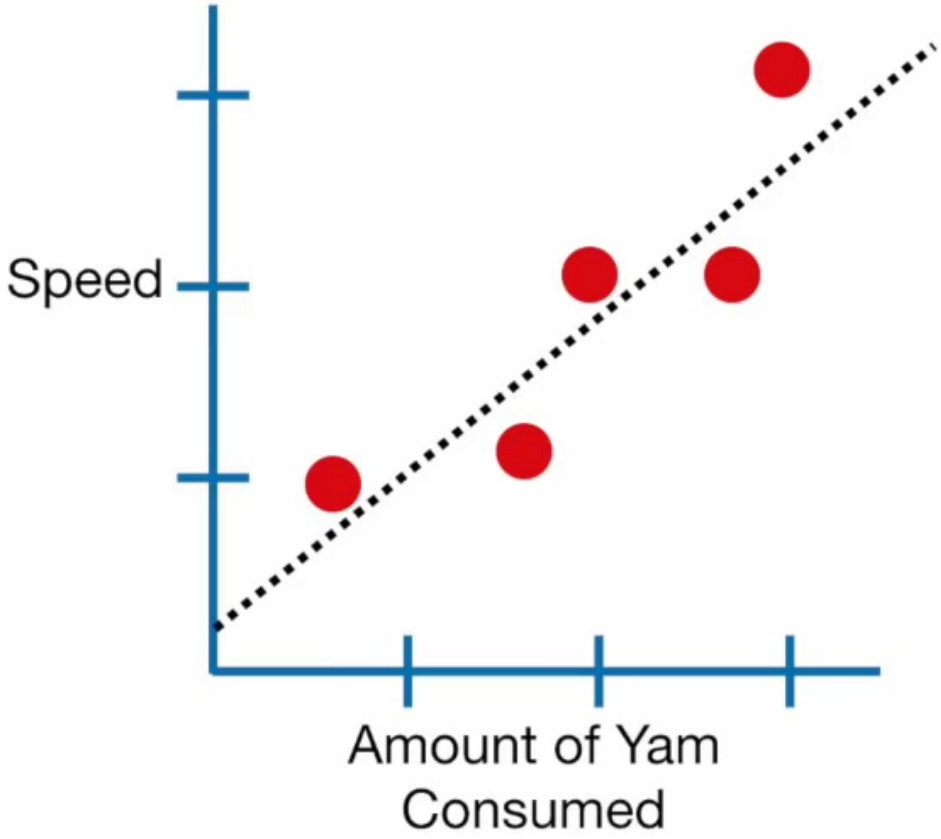


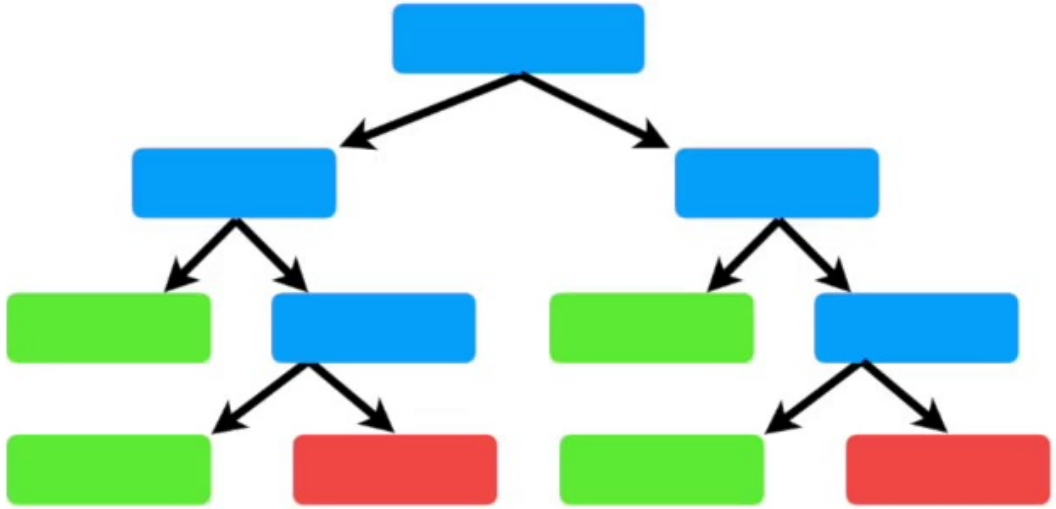
BAM!!!



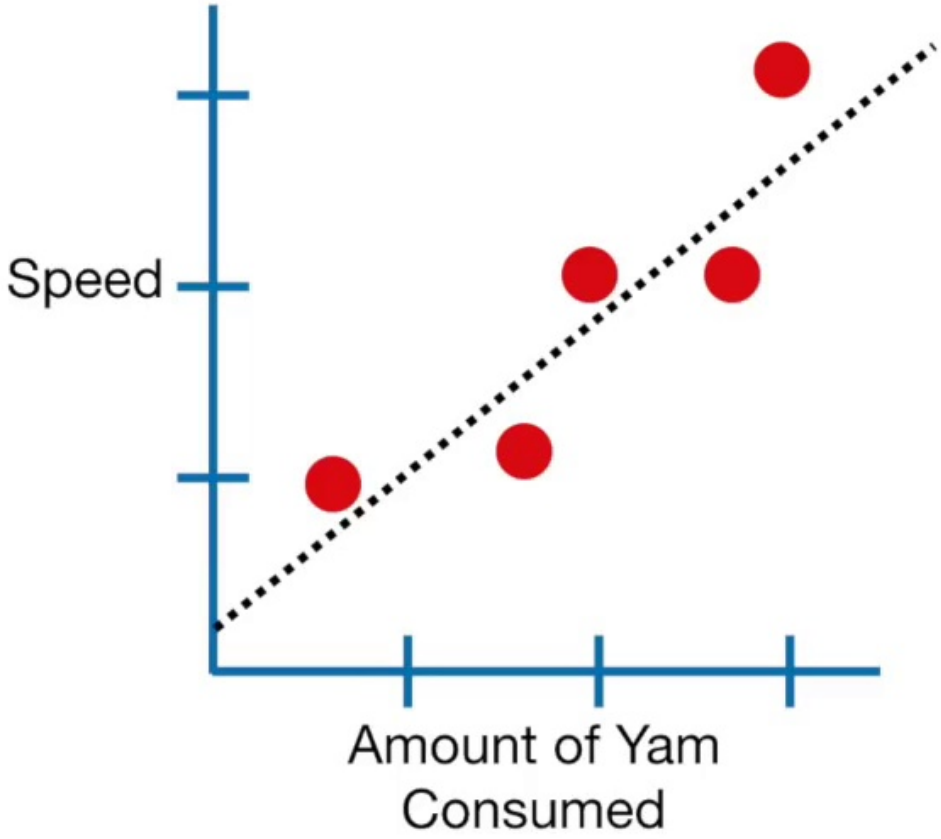


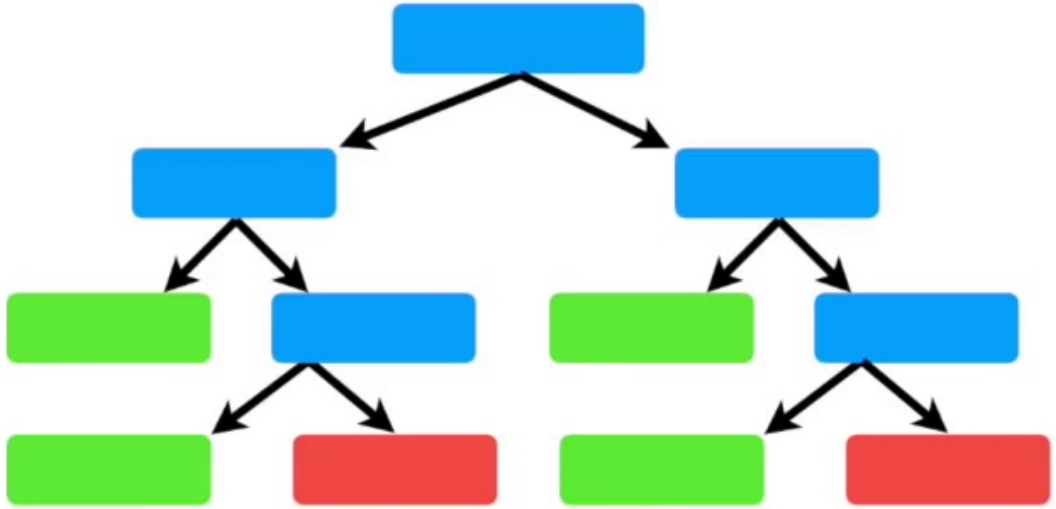
Now that we can make ***predictions*** and ***classifications***, let's talk about some of the main ideas in **Machine Learning**.

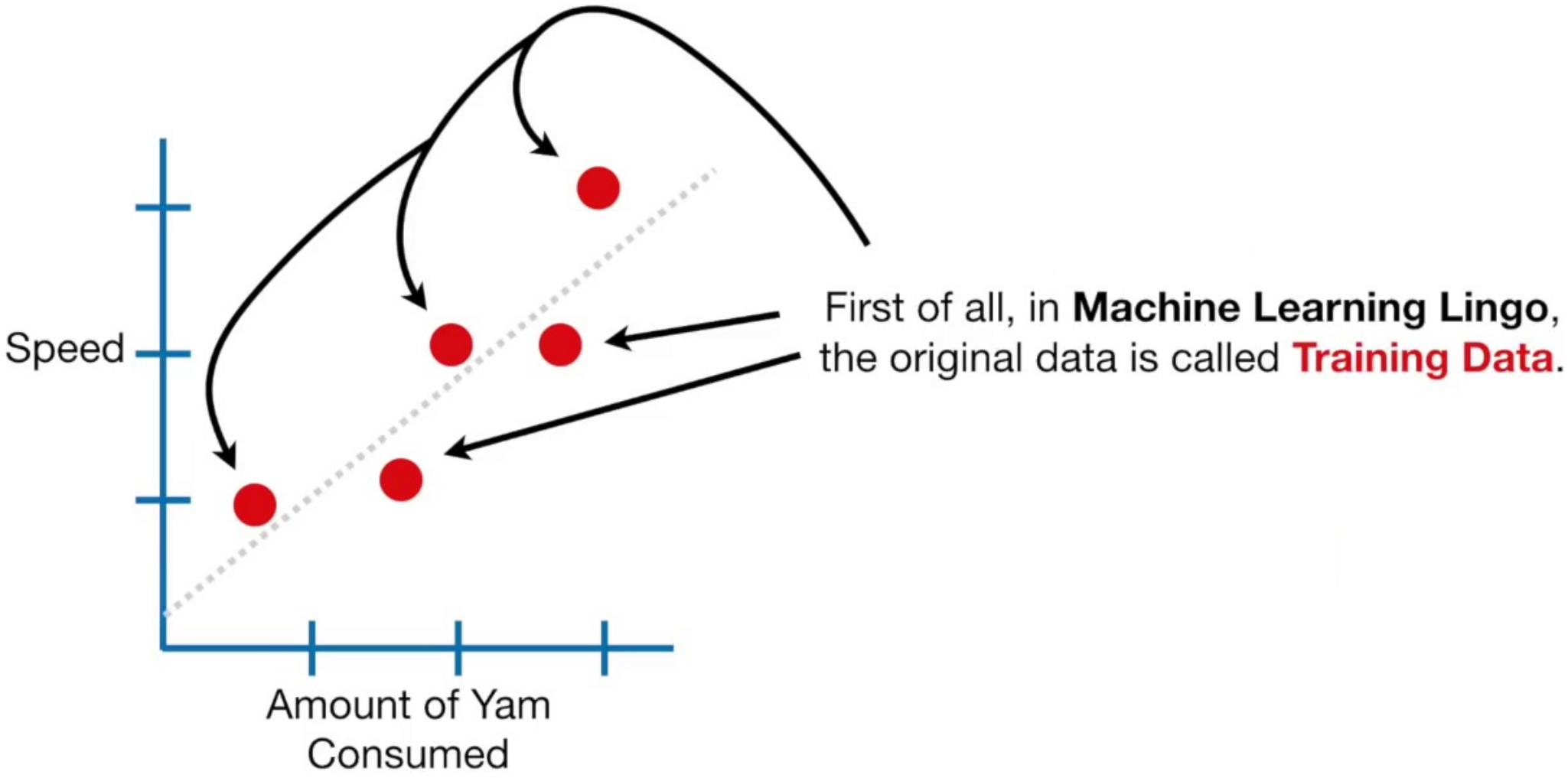


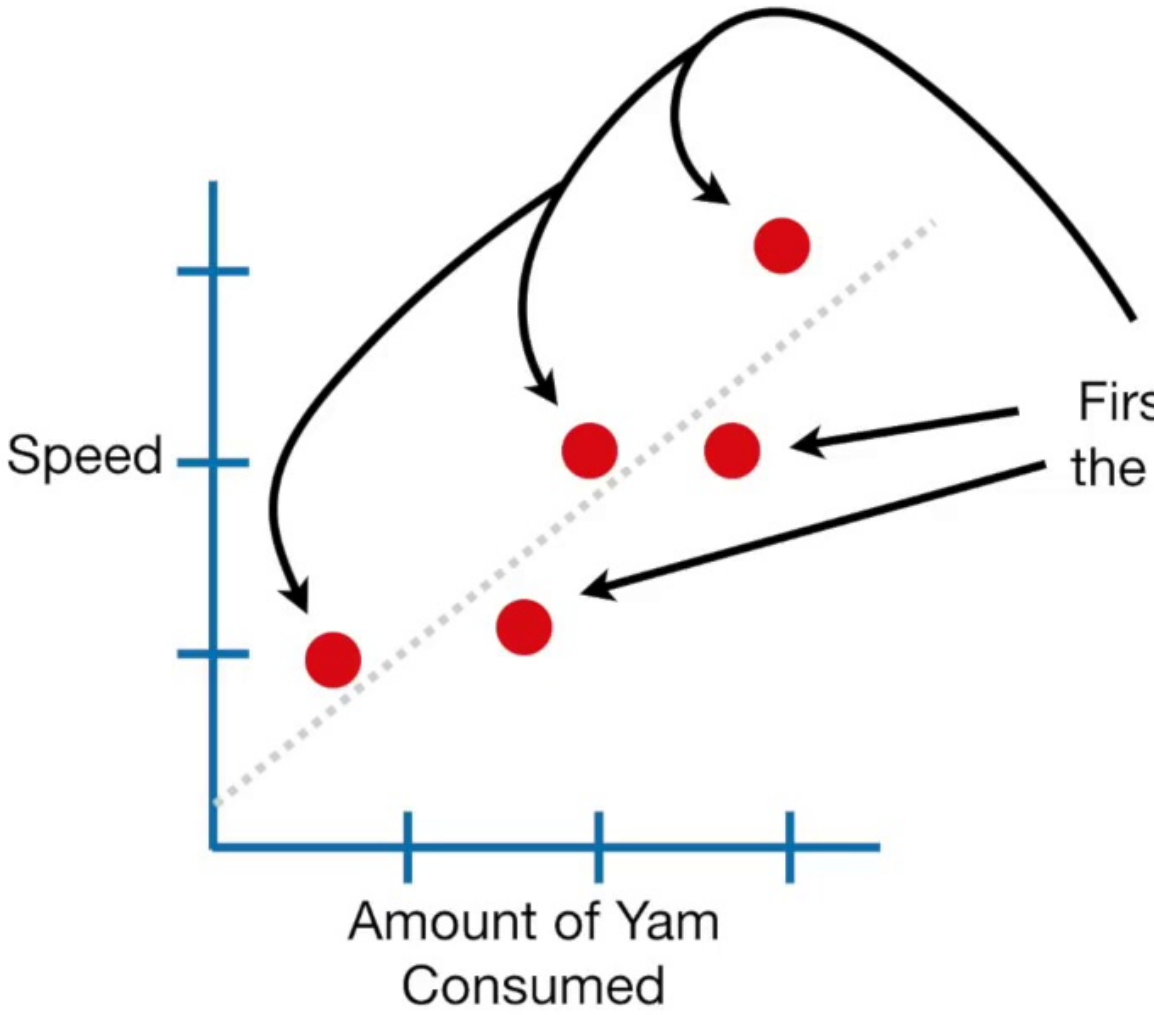


Now that we can make ***predictions*** and ***classifications***, let's talk about some of the main ideas in **Machine Learning**.

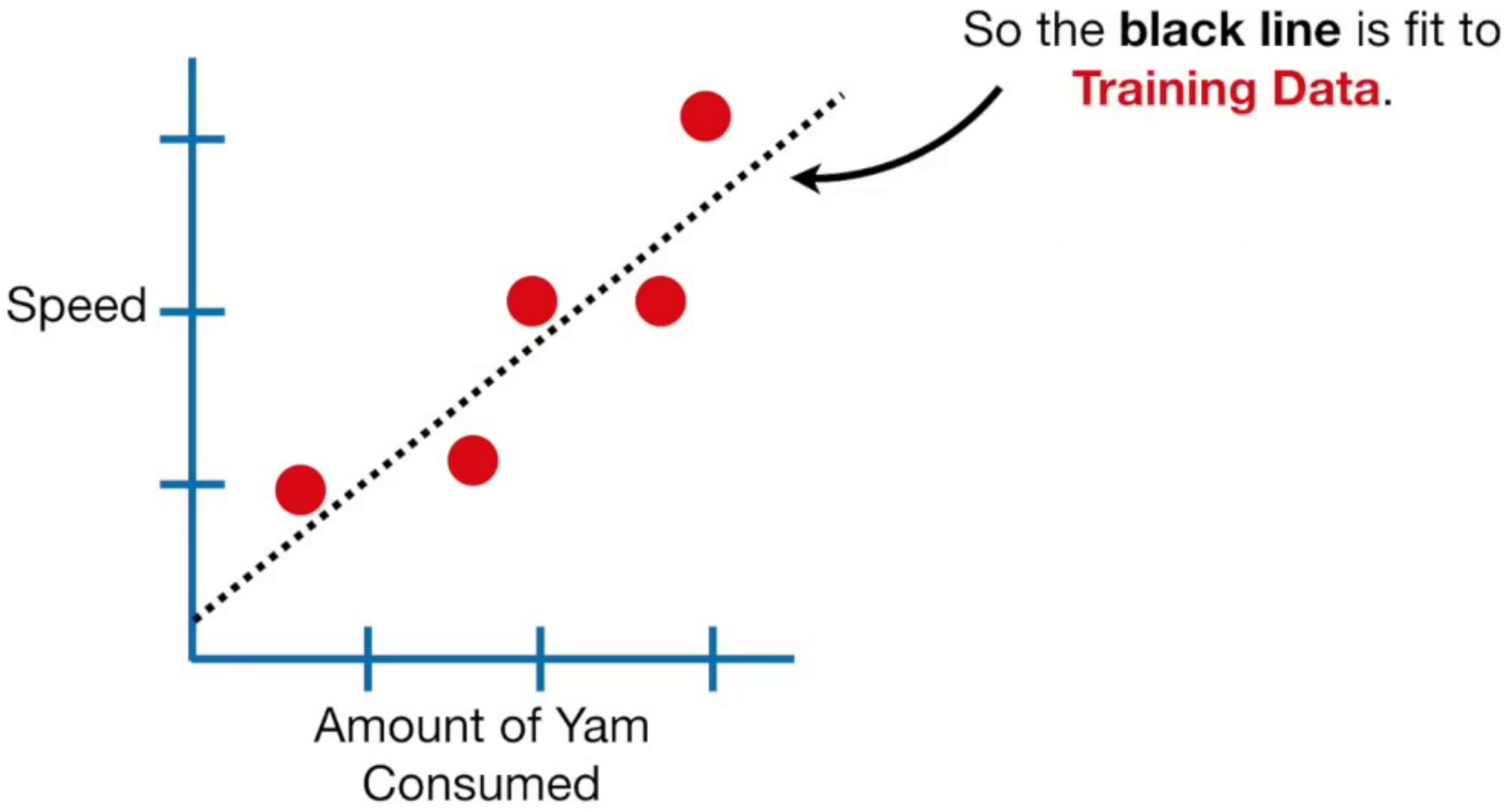


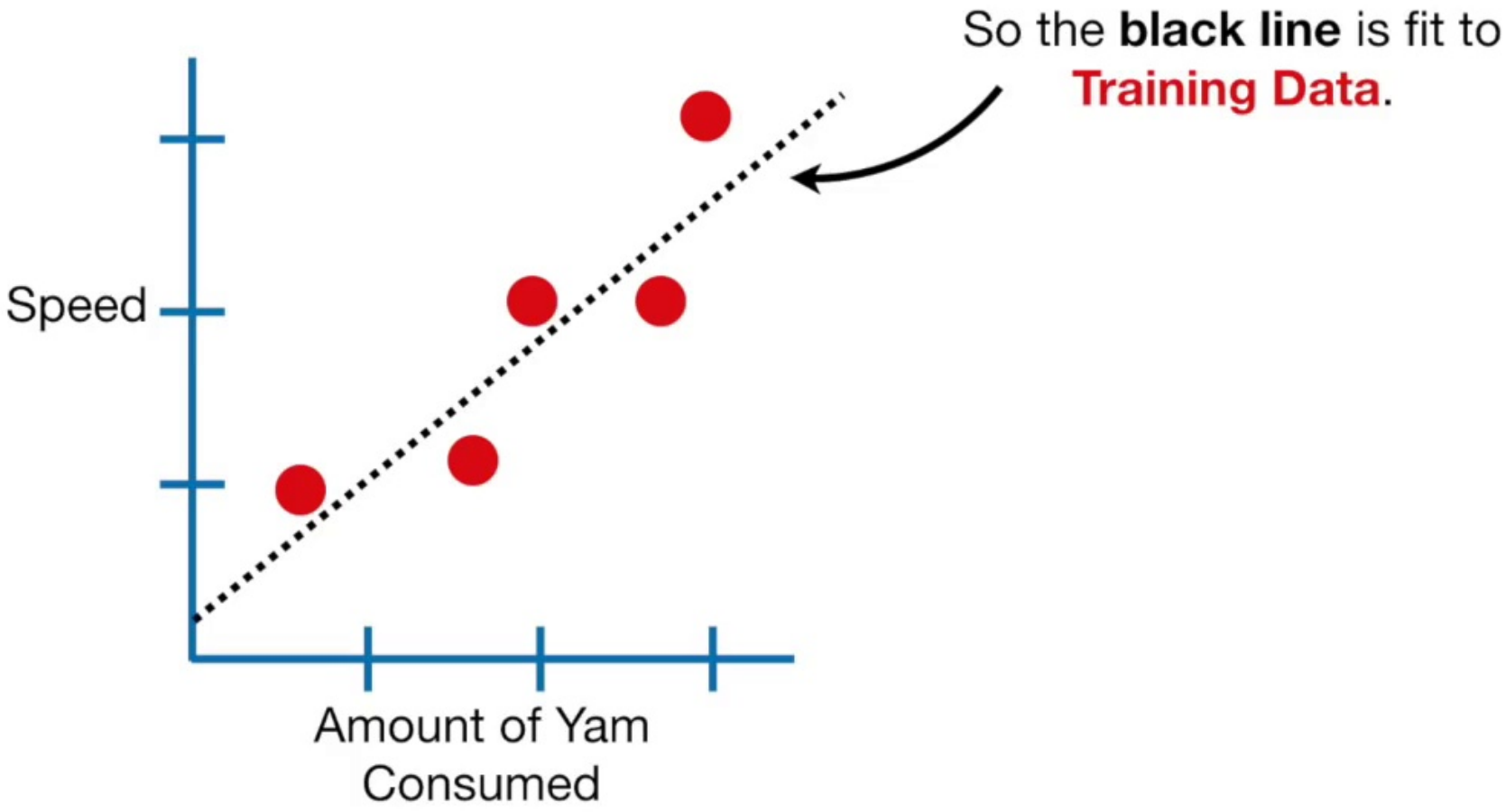


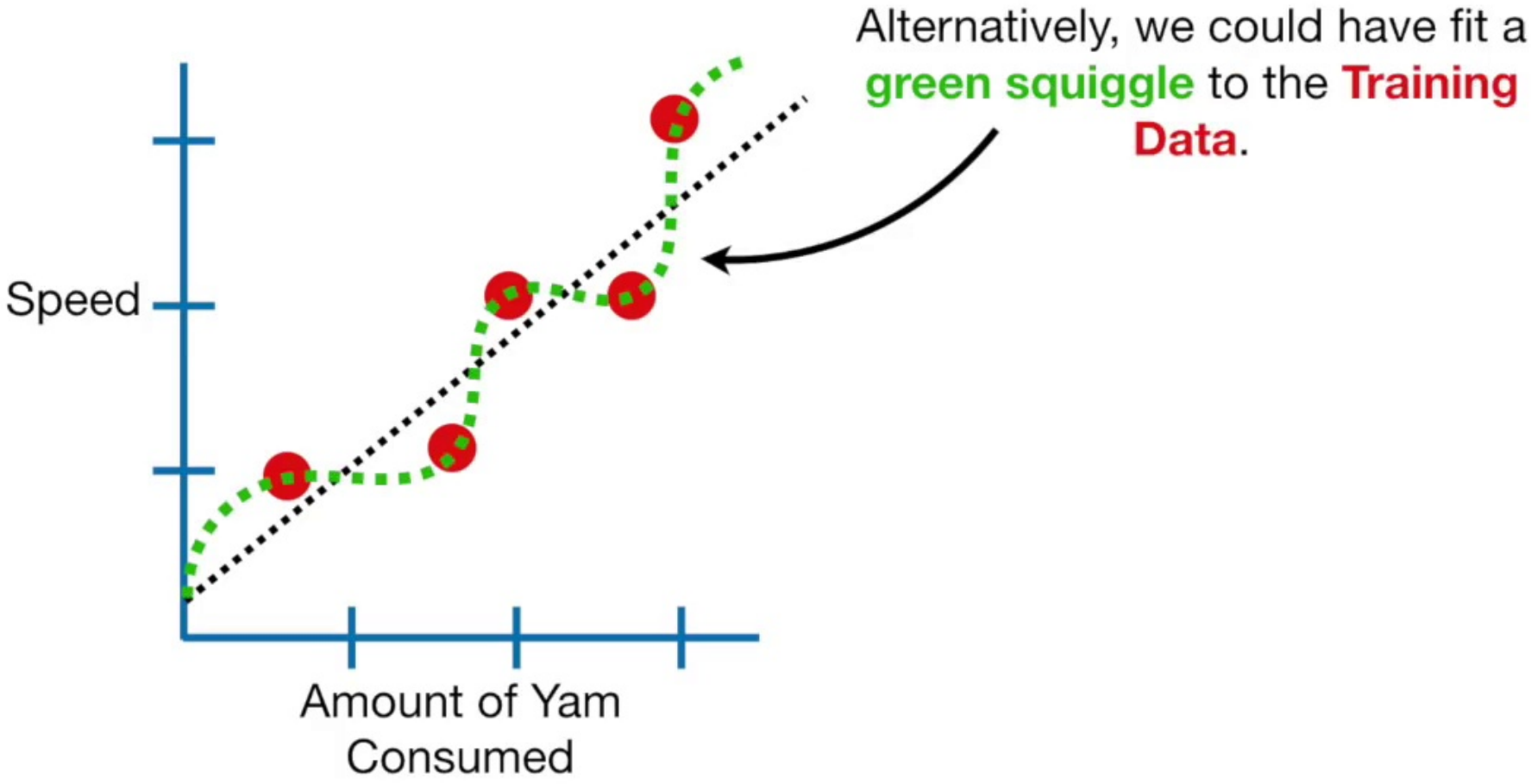

Speed
Amount of Yam Consumed
First of all, in **Machine Learning Lingo**, the original data is called **Training Data**.

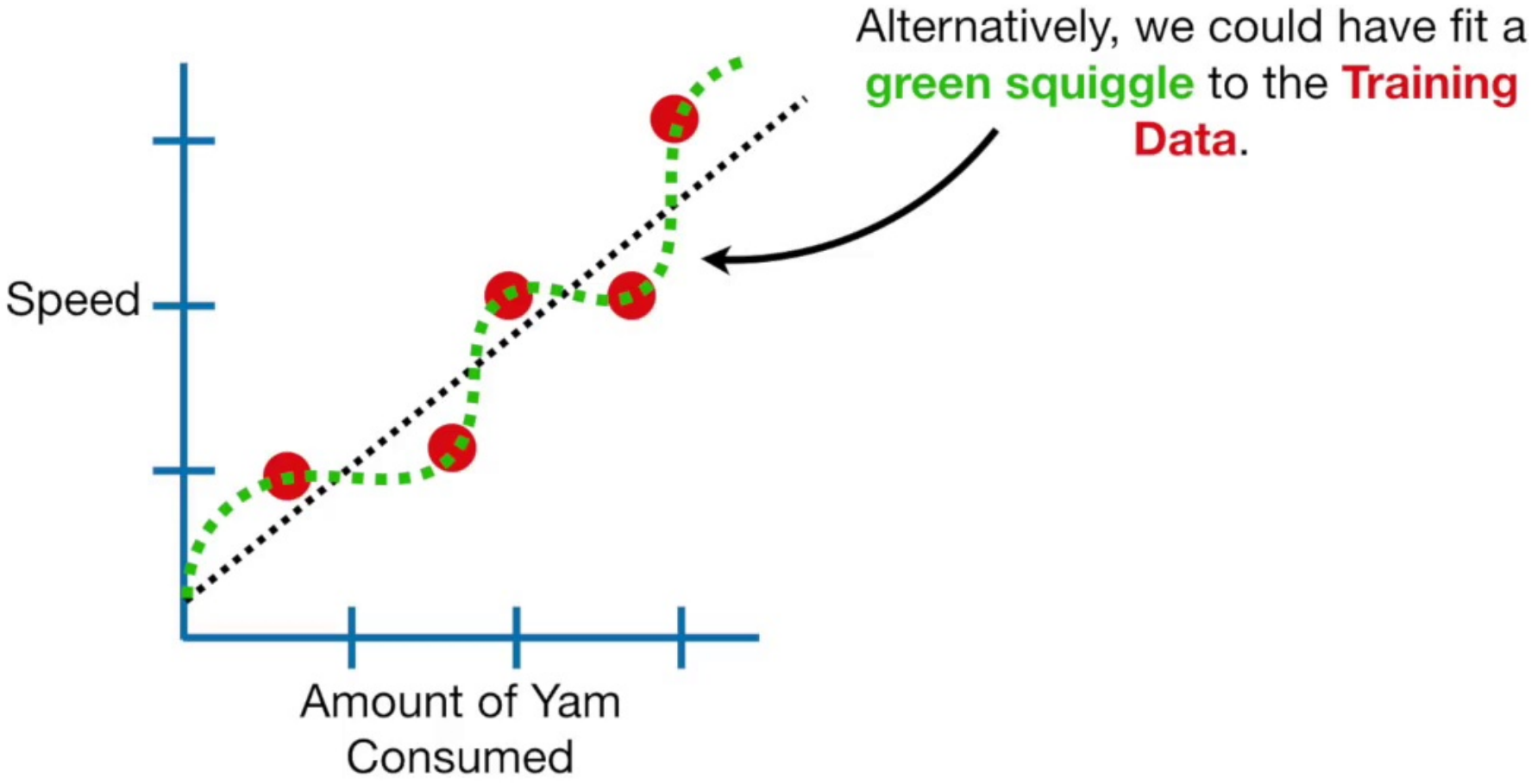


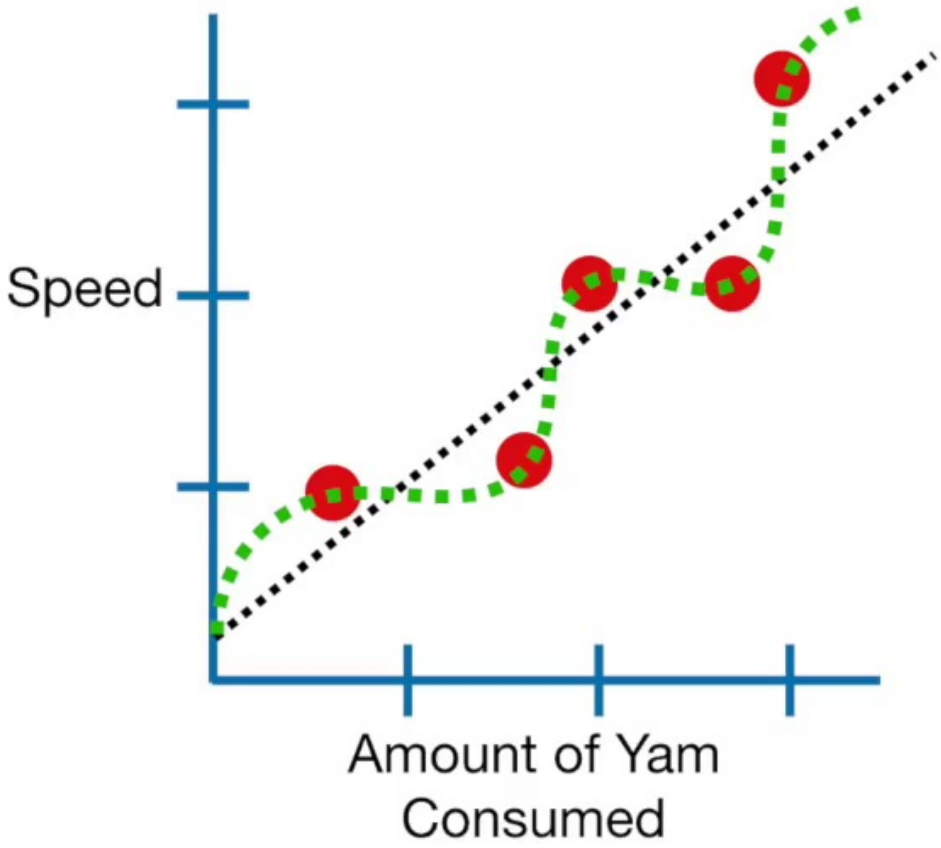
First of all, in **Machine Learning Lingo**, the original data is called **Training Data**.


So the **black line** is fit to **Training Data**.
Speed
Amount of Yam Consumed

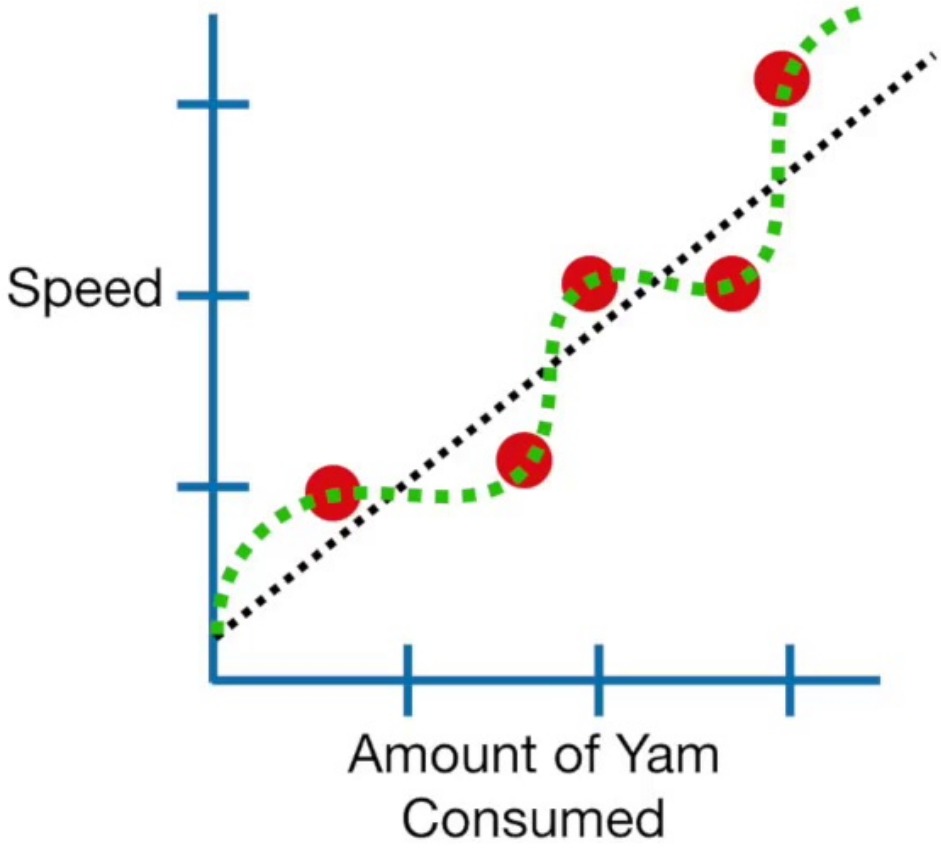

So the **black line** is fit to **Training Data**.
Speed
Amount of Yam Consumed


Alternatively, we could have fit a green squiggle to the Training Data.
Speed
Amount of Yam Consumed

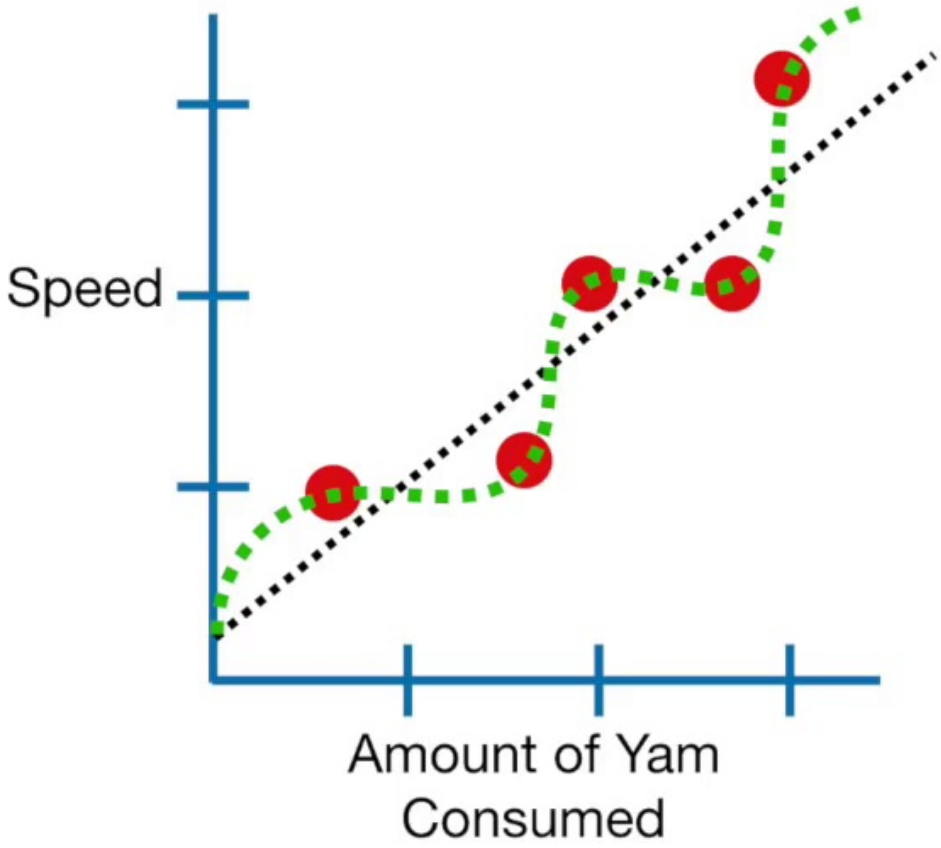

Alternatively, we could have fit a green squiggle to the Training Data.
Speed
Amount of Yam Consumed



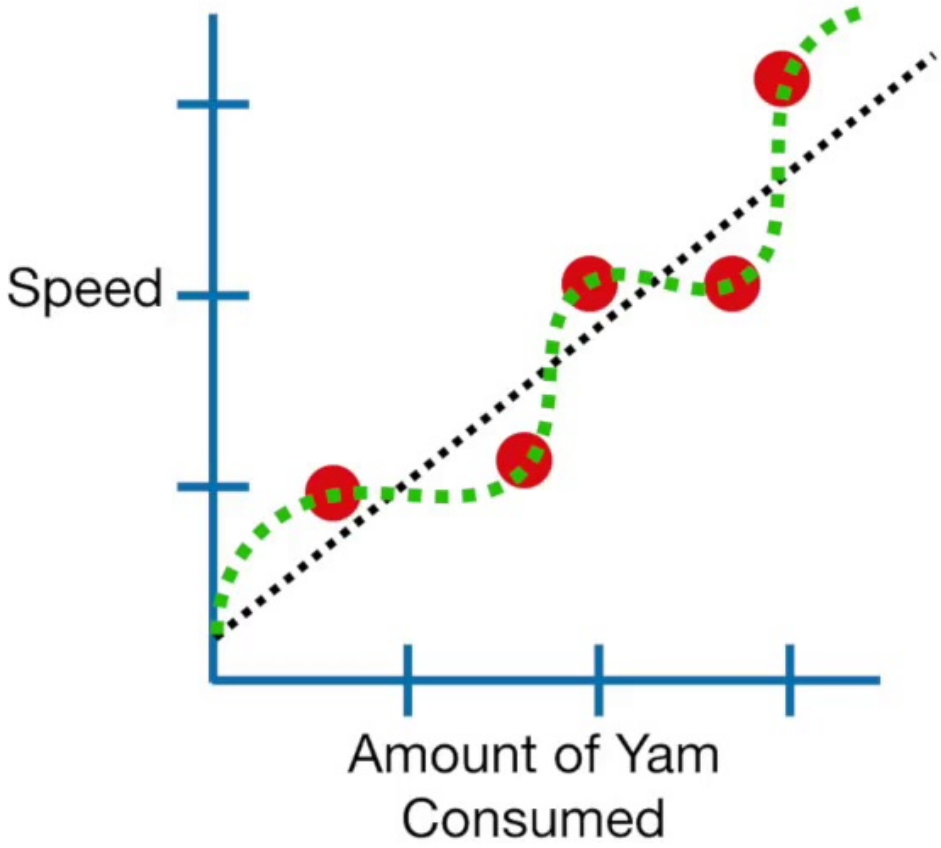
The **green squiggle** fits the **Training Data** better than the **black line**, but remember, the goal of **Machine Learning** is to make ***predictions***…



The **green squiggle** fits the **Training Data** better than the **black line**, but remember, the goal of **Machine Learning** is to make ***predictions***…

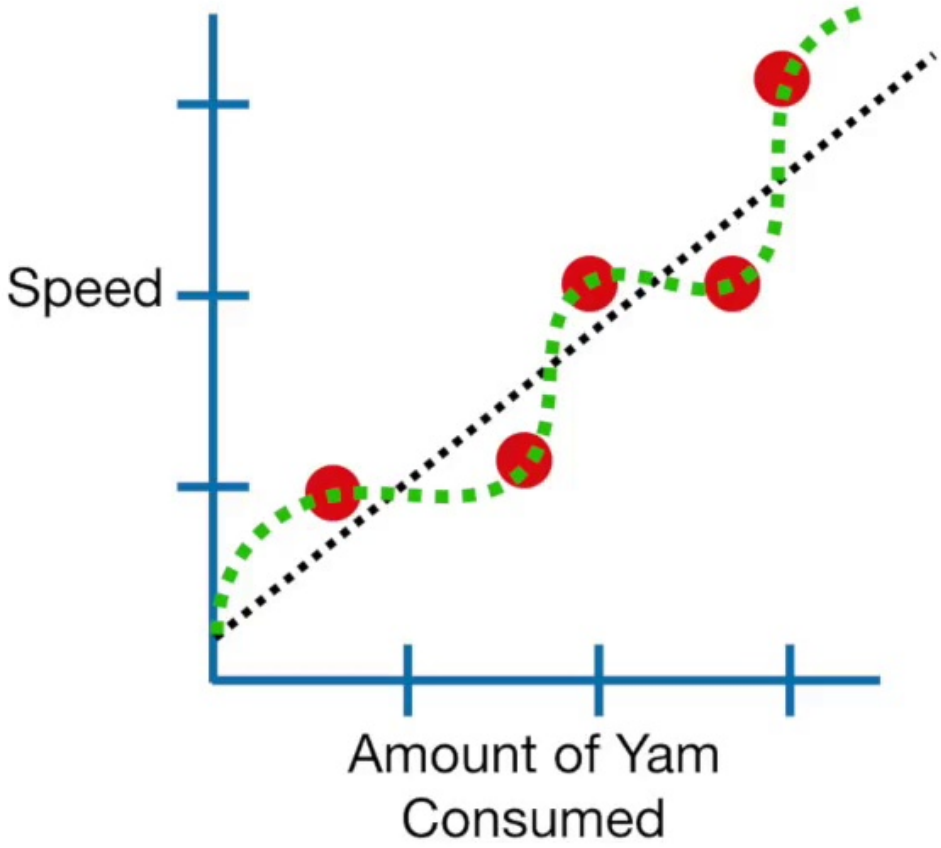


The **green squiggle** fits the **Training Data** better than the **black line**, but remember, the goal of **Machine Learning** is to make ***predictions***…



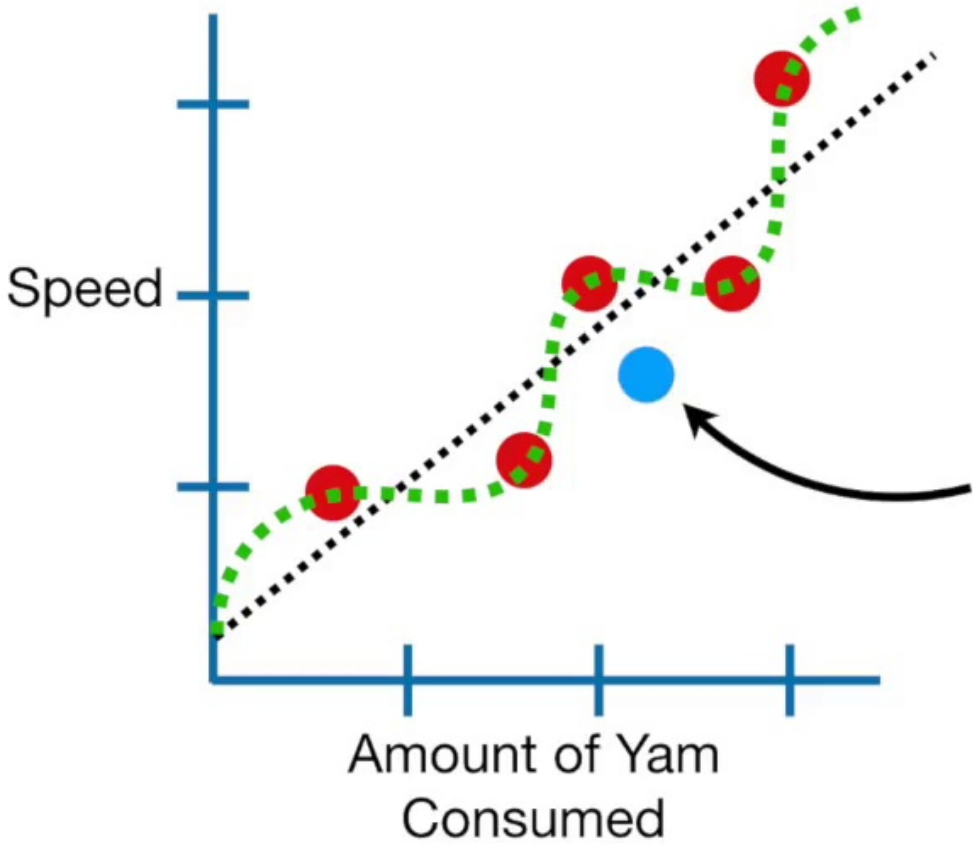
The **green squiggle** fits the **Training Data** better than the **black line**, but remember, the goal of **Machine Learning** is to make ***predictions***…

…so we need a way to decide if the **green squiggle** is better or worse than the **black line** at making ***predictions***.

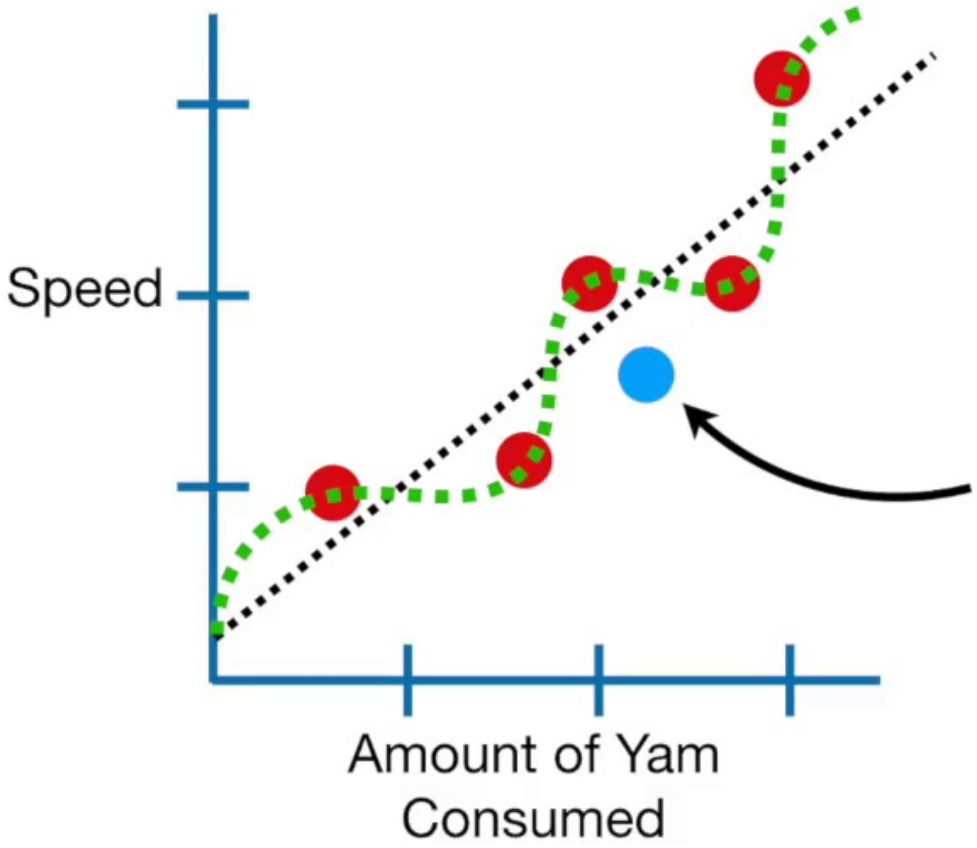


The **green squiggle** fits the **Training Data** better than the **black line**, but remember, the goal of **Machine Learning** is to make ***predictions***…

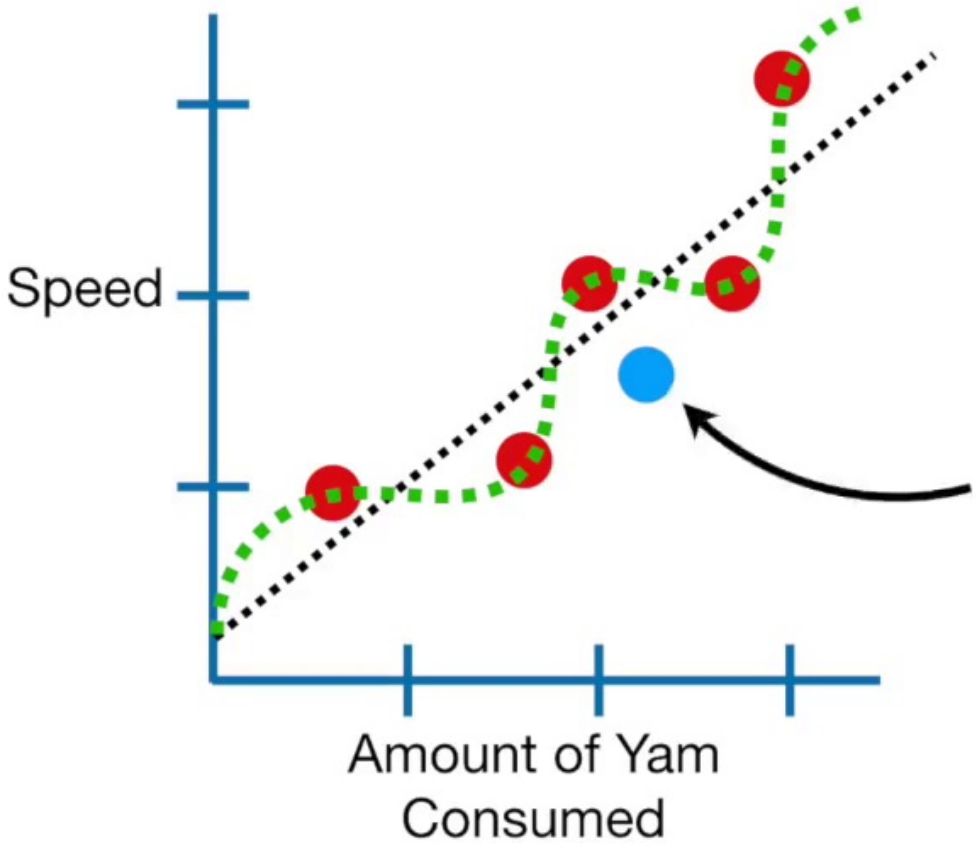
…so we need a way to decide if the **green squiggle** is better or worse than the **black line** at making ***predictions***.



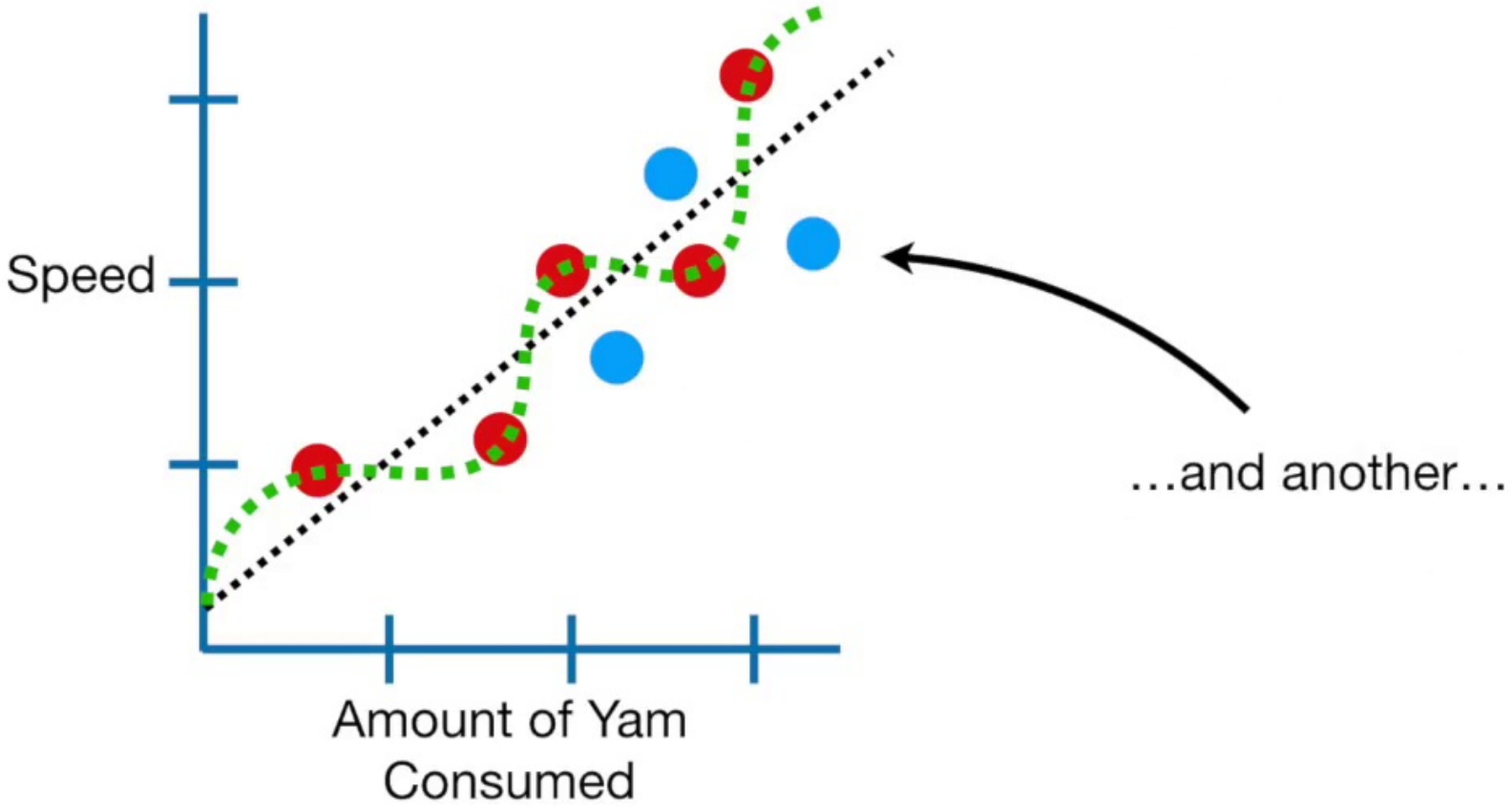
So we find a new person, and measure how fast they run and how much yam they eat…

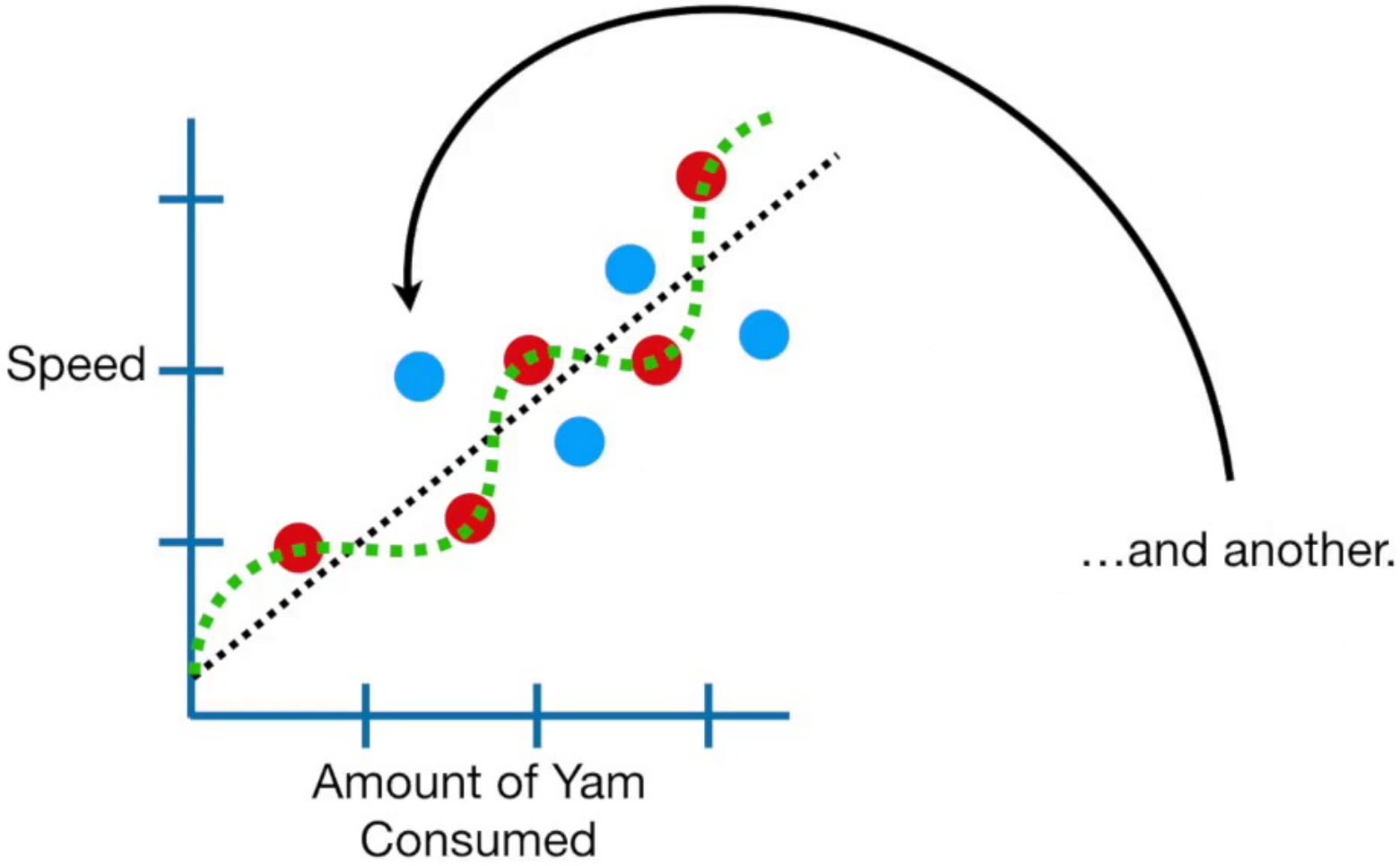


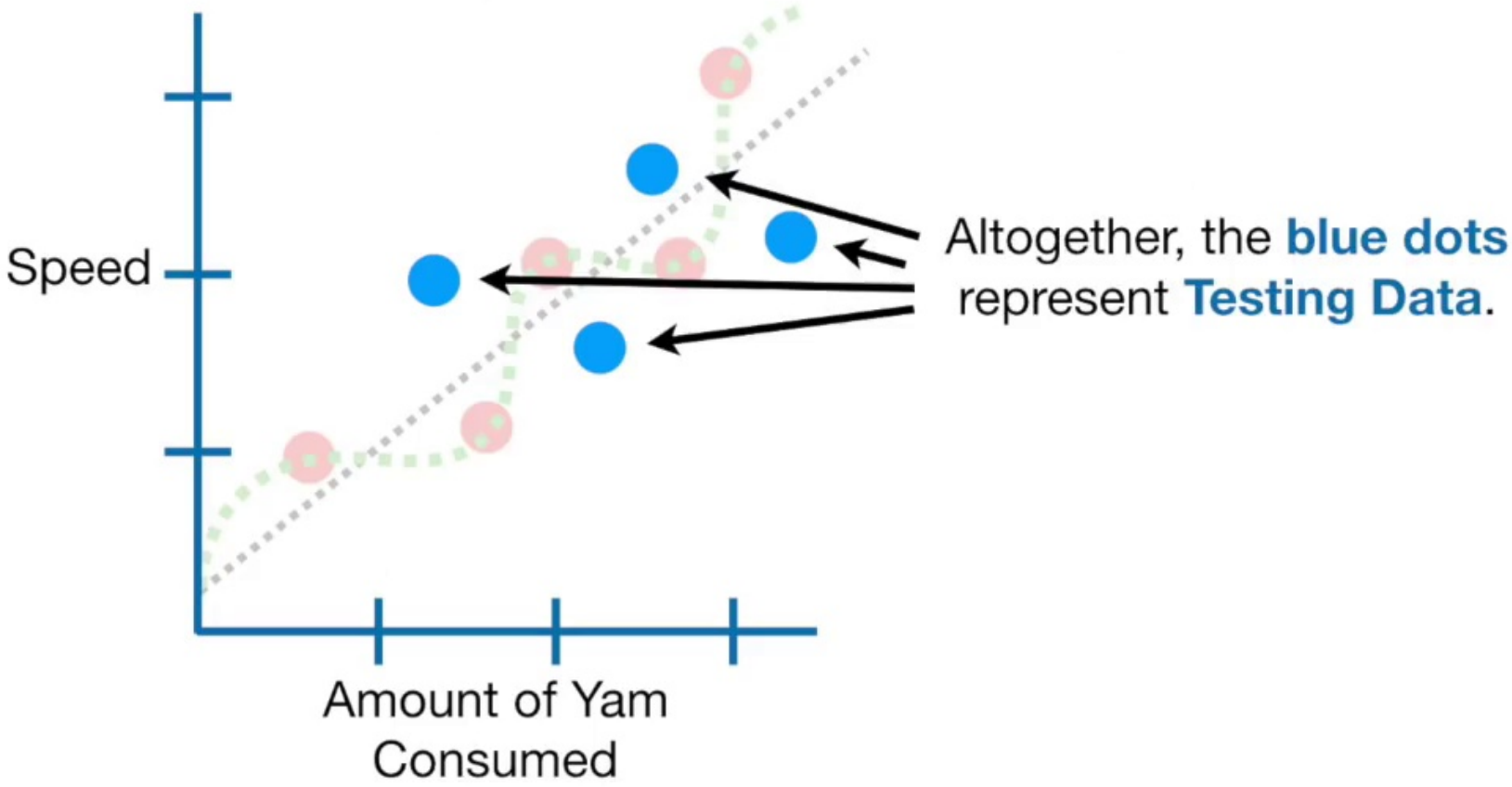
So we find a new person, and measure how fast they run and how much yam they eat…

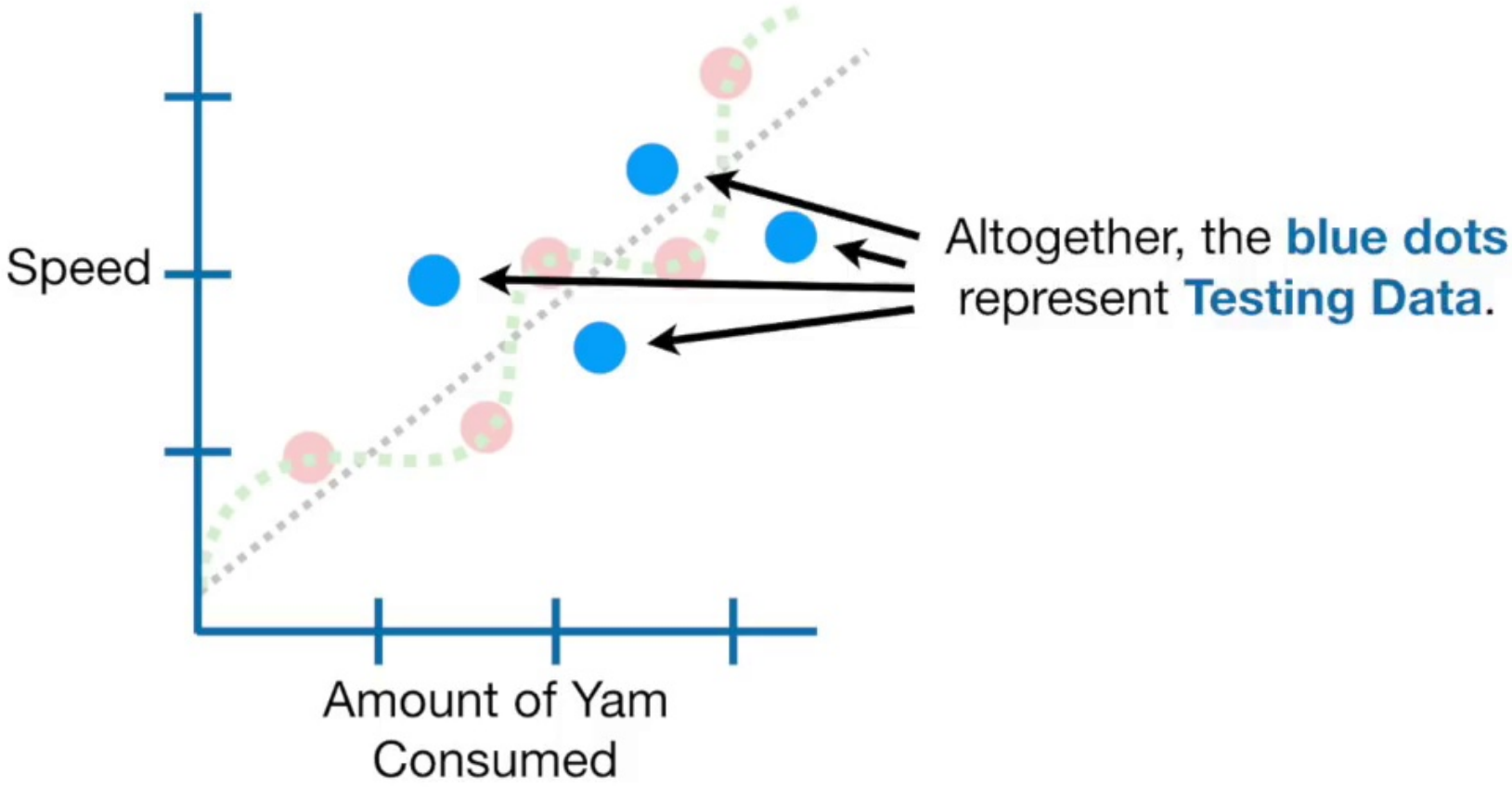


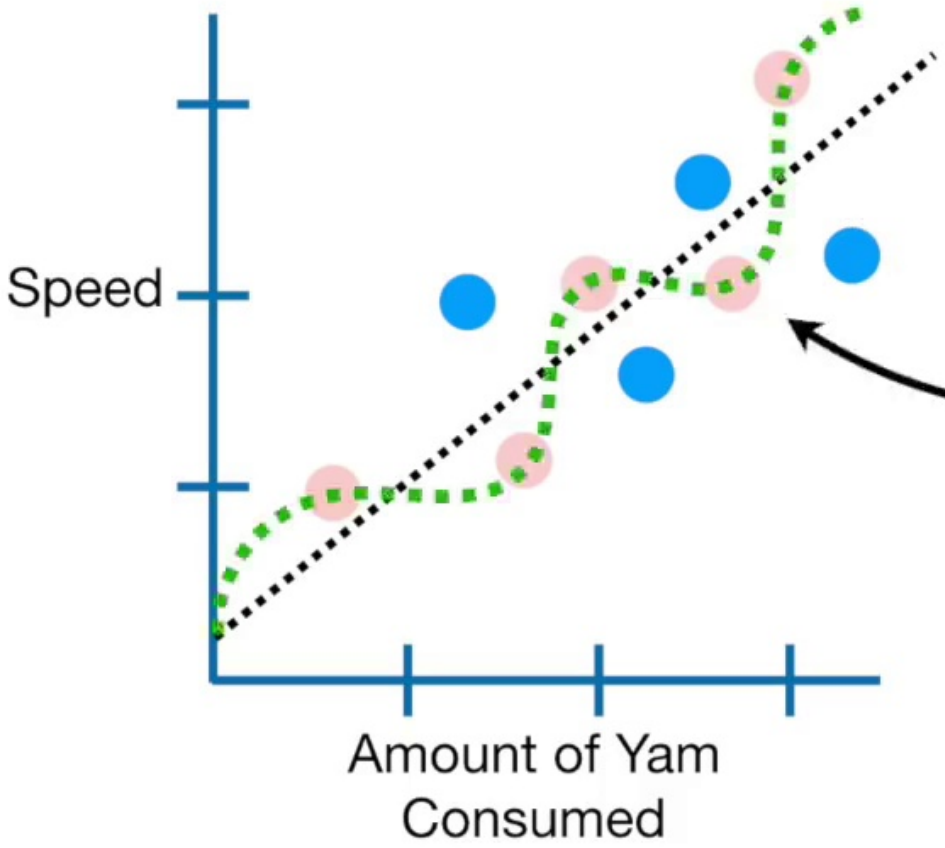
So we find a new person, and measure how fast they run and how much yam they eat…


Speed
Amount of Yam
Consumed
…and another…


Speed
Amount of Yam
Consumed
…and another.

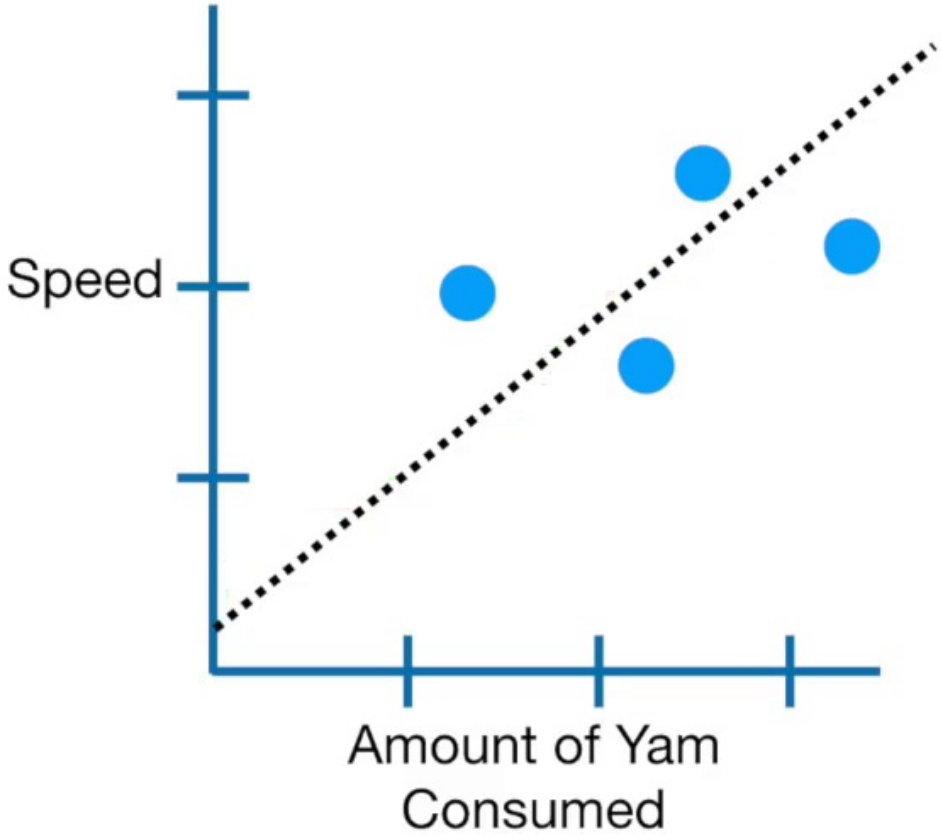

Speed
Amount of Yam Consumed
Altogether, the **blue dots** represent **Testing Data**.


Speed
Amount of Yam Consumed
Altogether, the **blue dots** represent **Testing Data**.

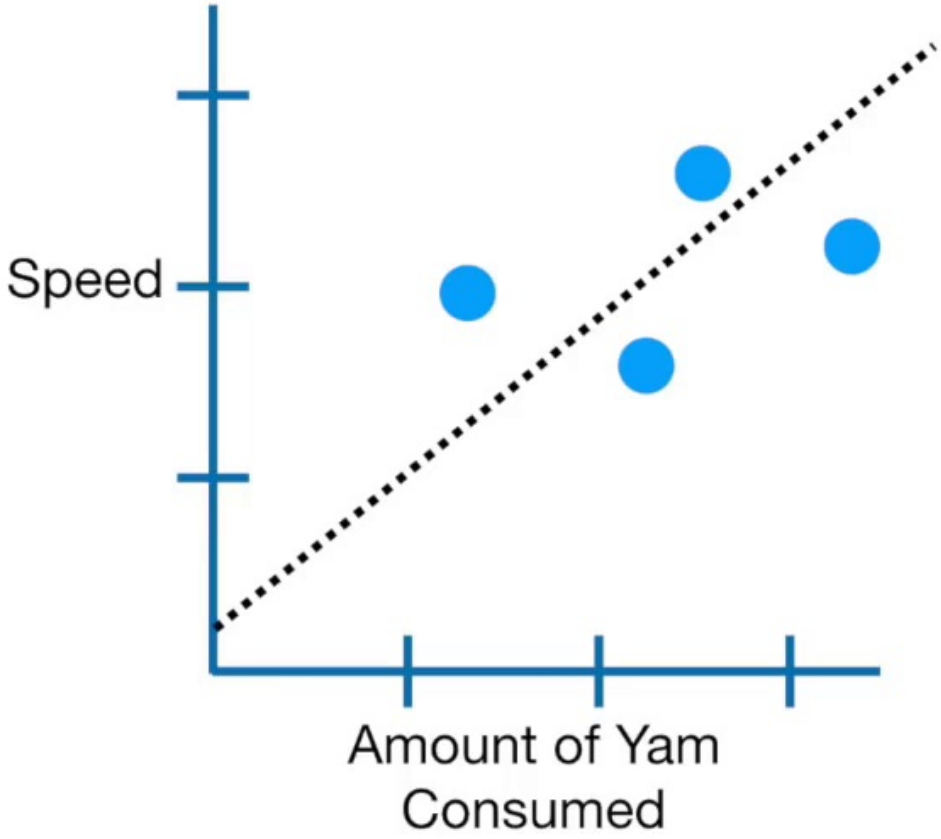


We use the **Testing Data** to compare the ***predictions*** made by the **black line**…

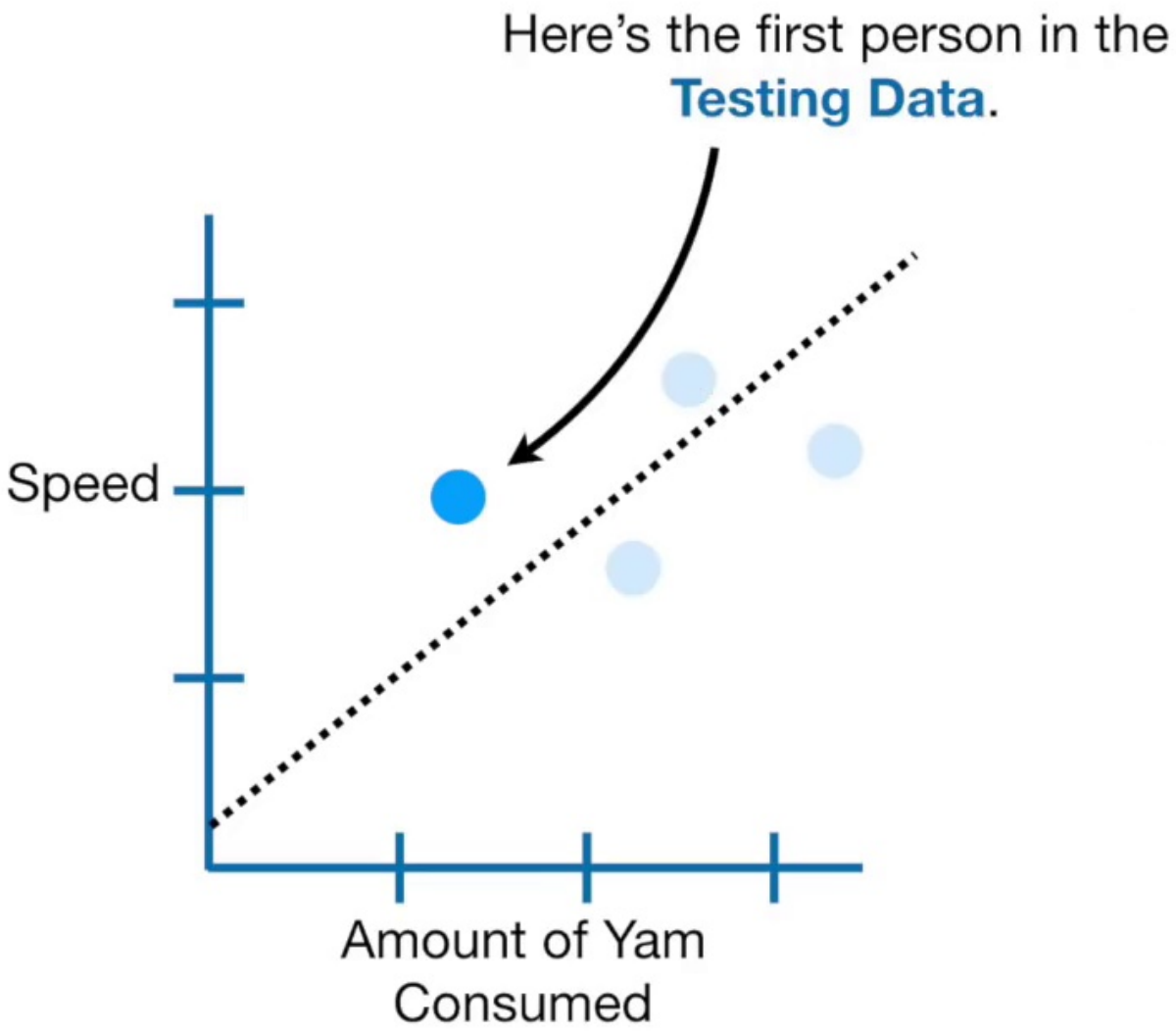
…to the ***predictions*** made by the **green squiggle**.

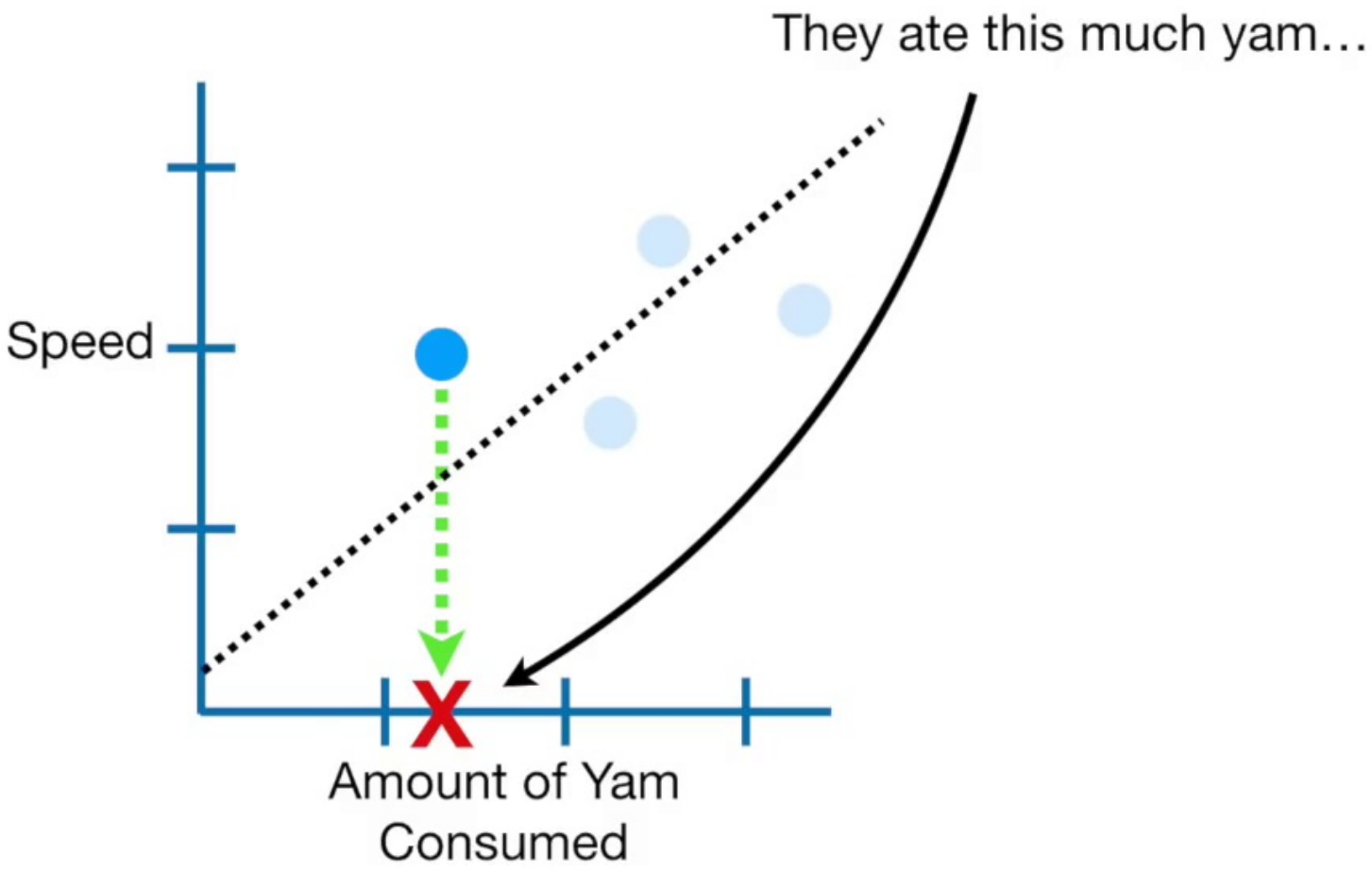


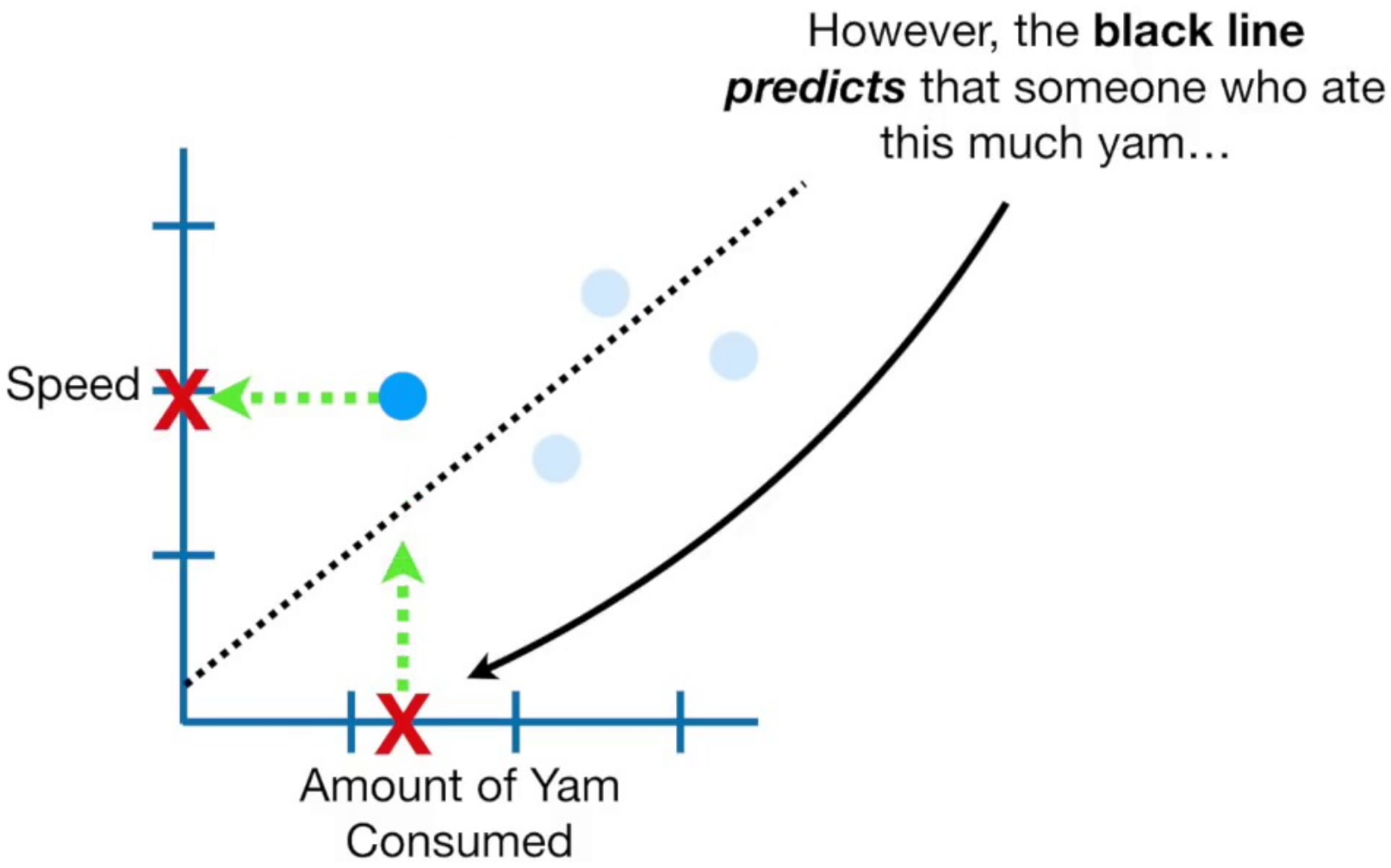
Let's start by seeing how well the **black line** predicts the speed of each person in the **Testing Data**.

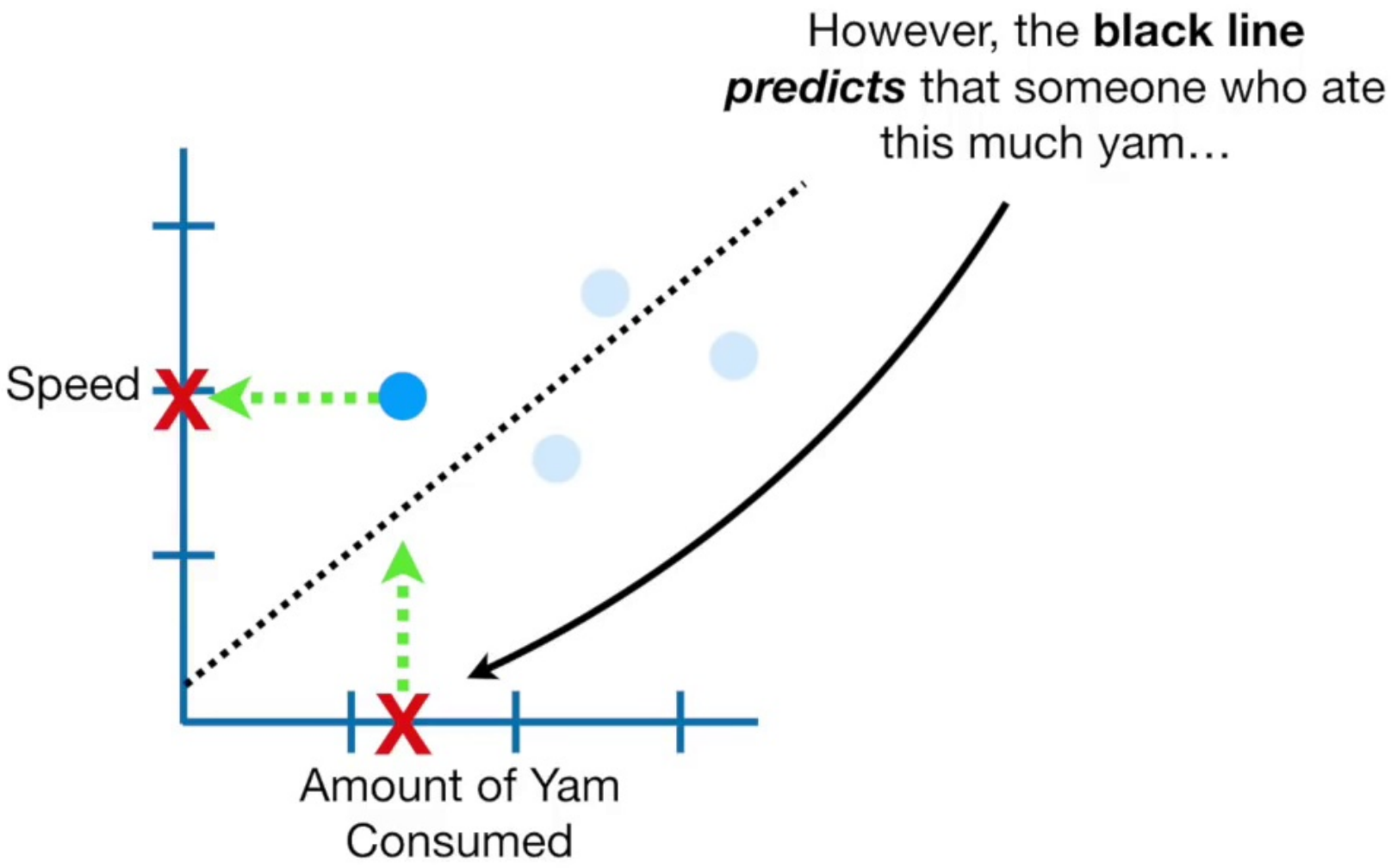


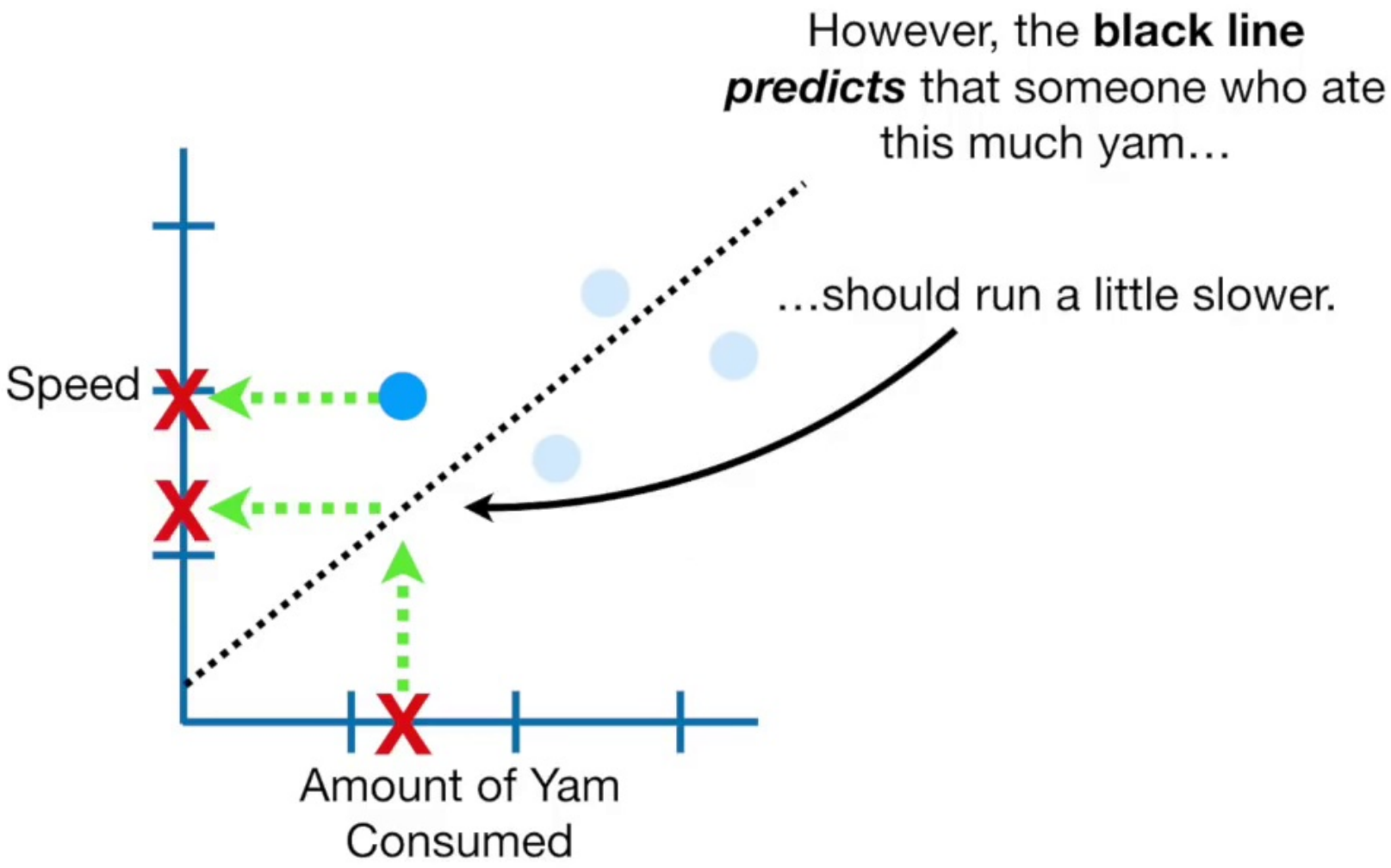
Let's start by seeing how well the **black line** predicts the speed of each person in the **Testing Data**.

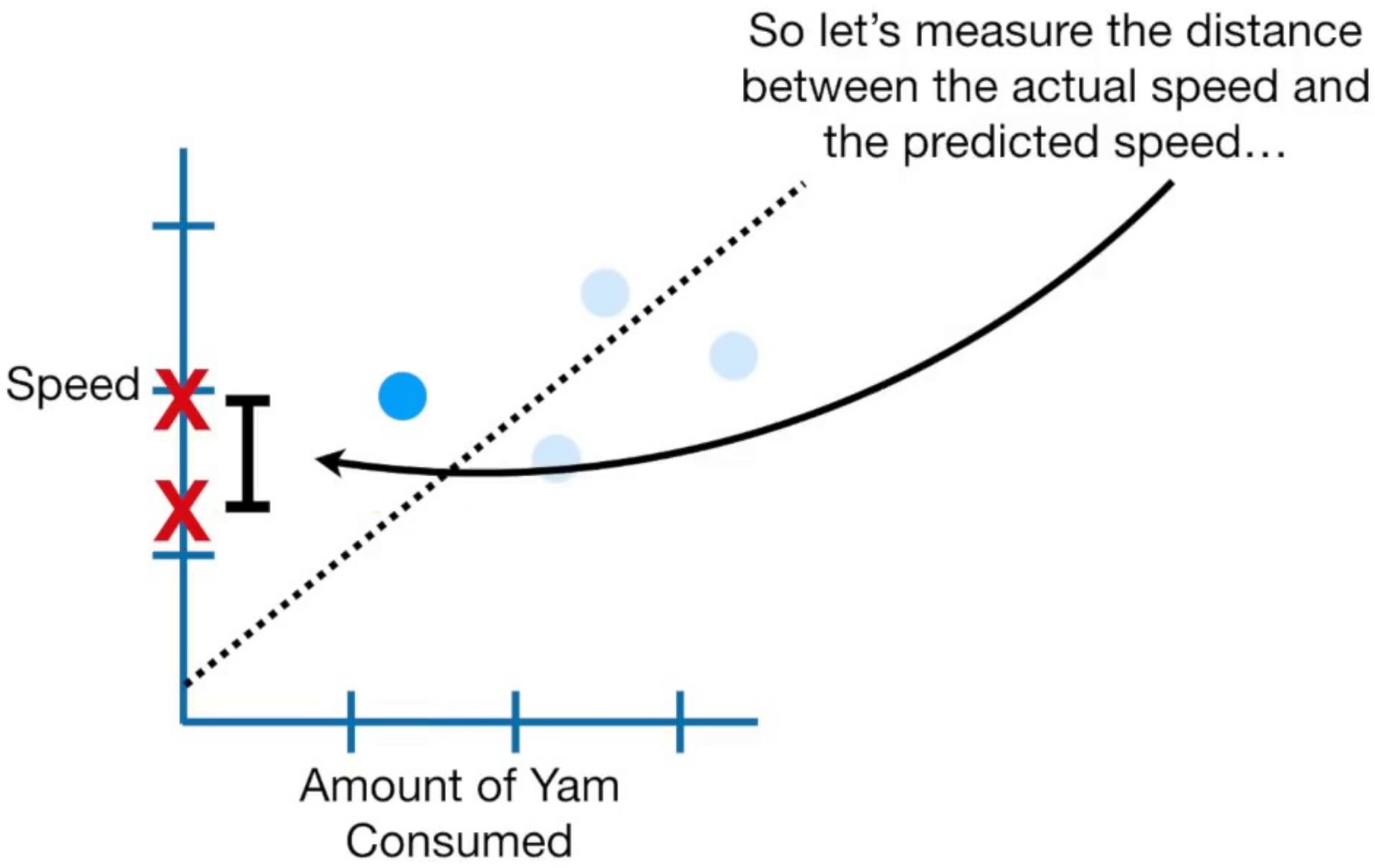

Here's the first person in the
Testing Data.
Speed
Amount of Yam
Consumed

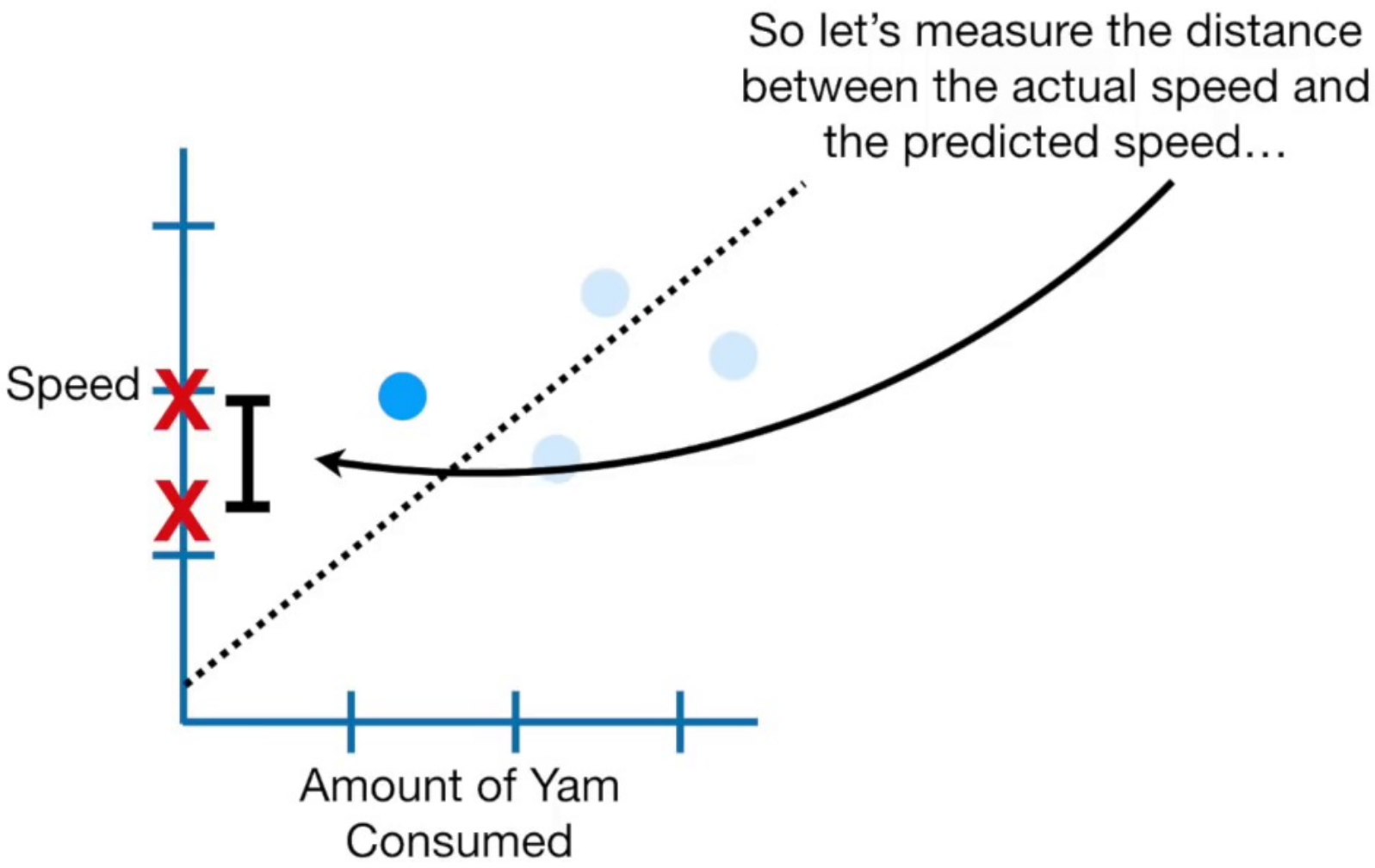

They ate this much yam…
Speed
Amount of Yam
Consumed

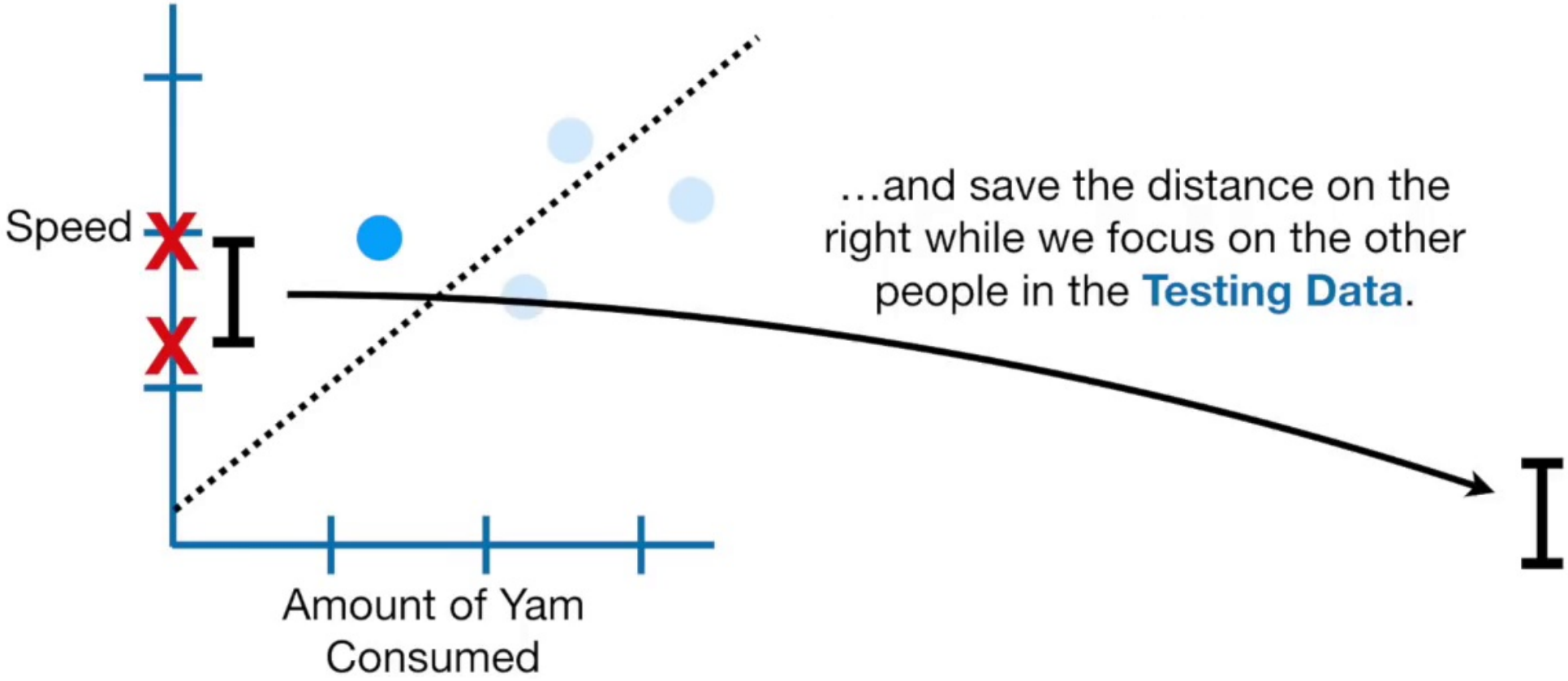

However, the **black line** ***predicts*** that someone who ate this much yam…
Speed
Amount of Yam Consumed

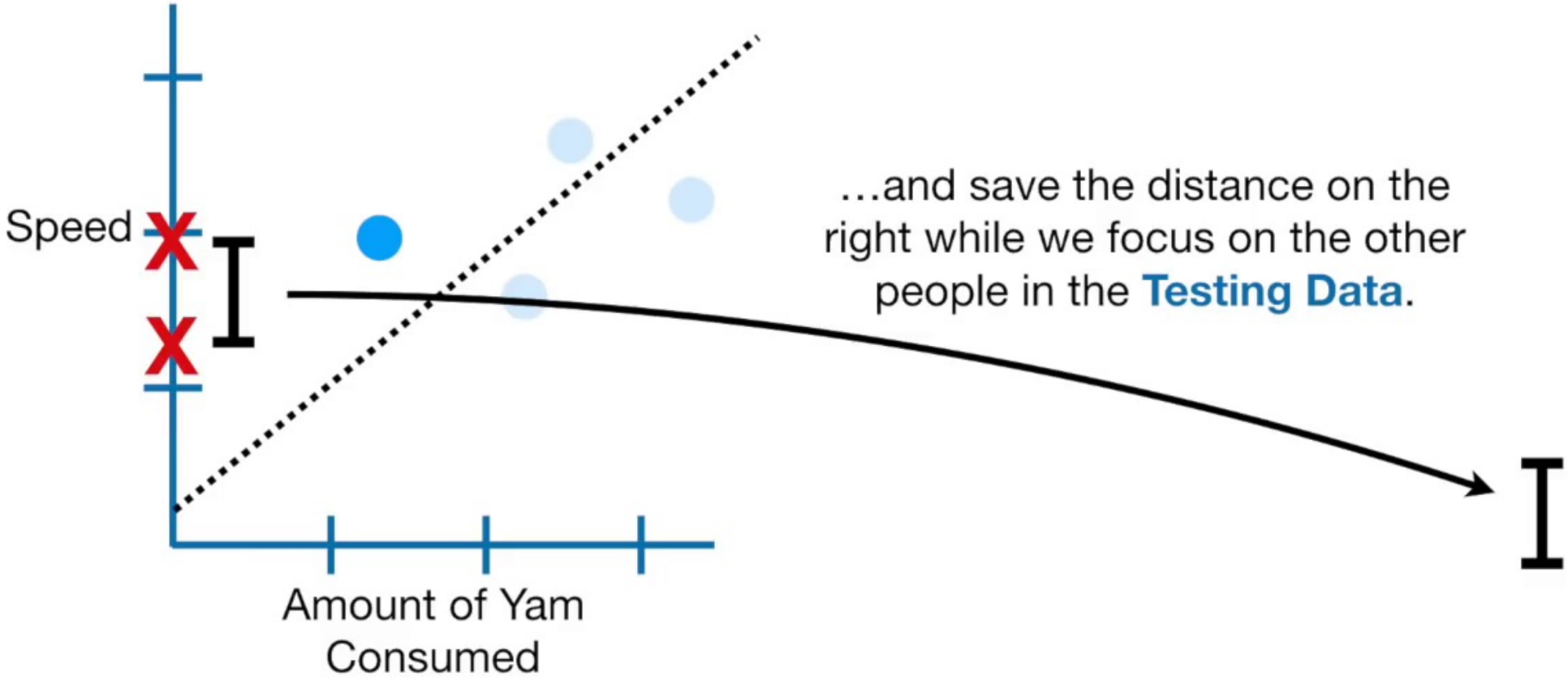

However, the **black line** ***predicts*** that someone who ate this much yam…
Speed
Amount of Yam Consumed

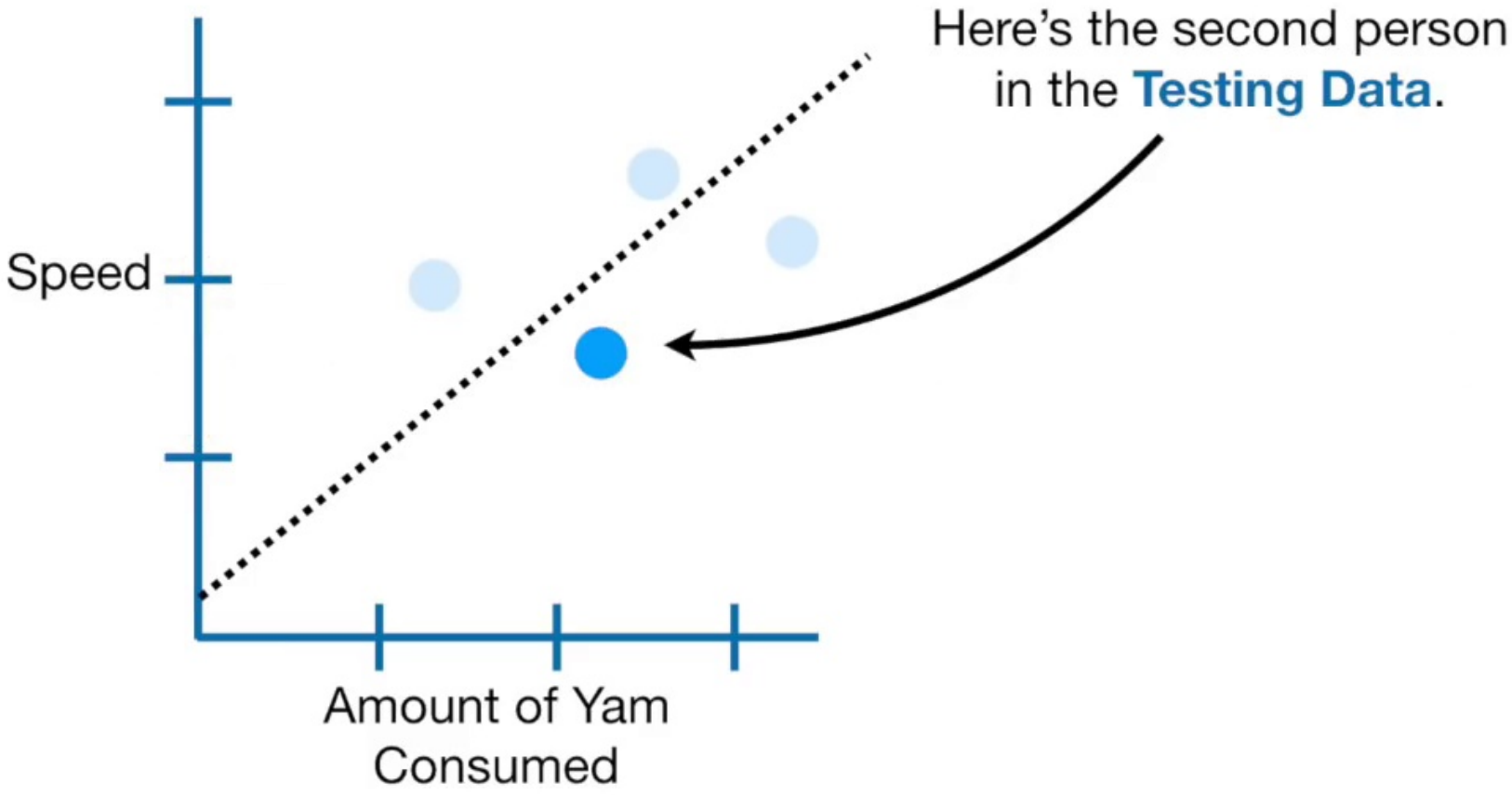

However, the **black line** ***predicts*** that someone who ate this much yam…
…should run a little slower.
Speed
Amount of Yam Consumed

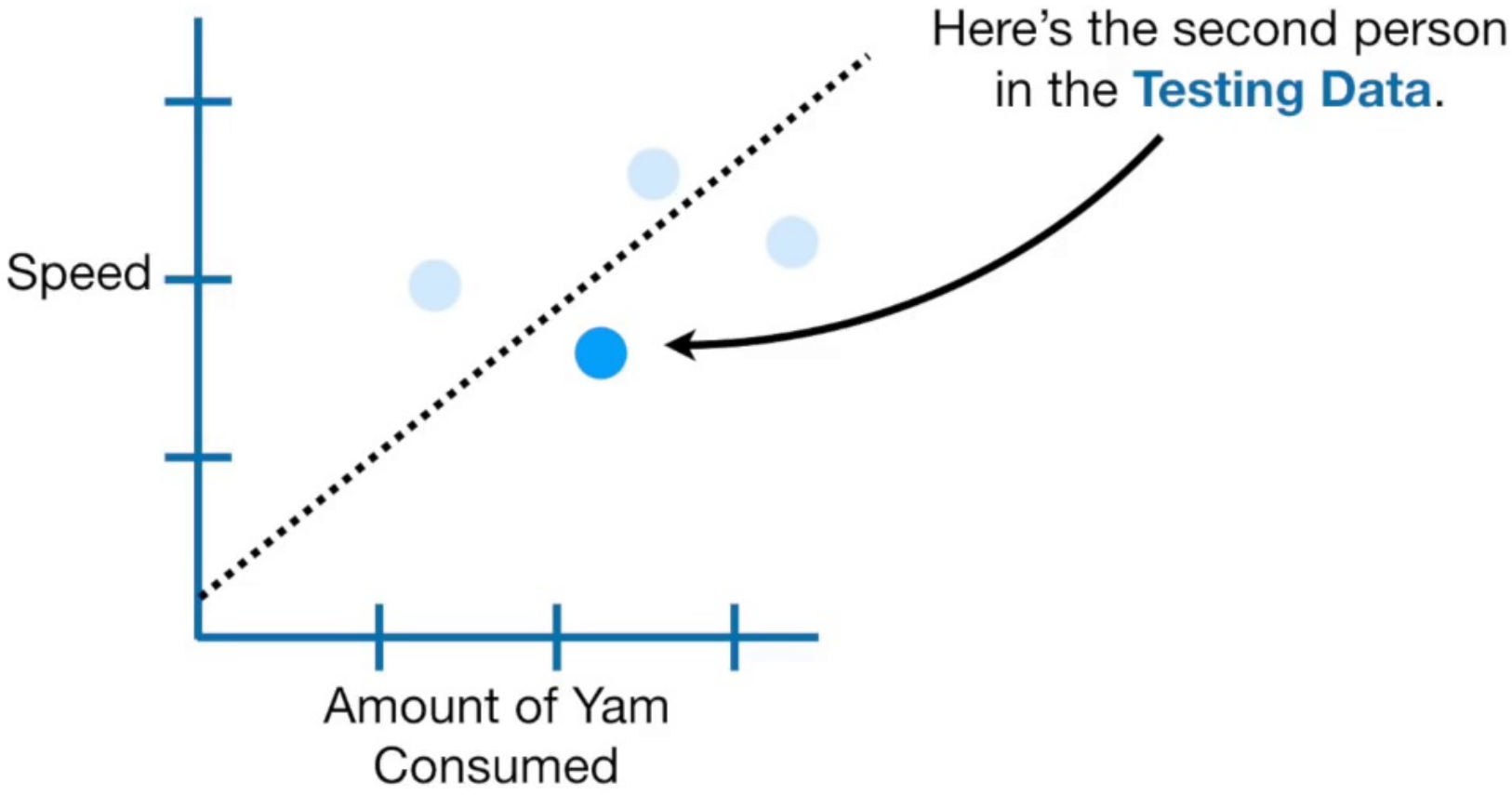

So let's measure the distance between the actual speed and the predicted speed…
Speed
Amount of Yam Consumed

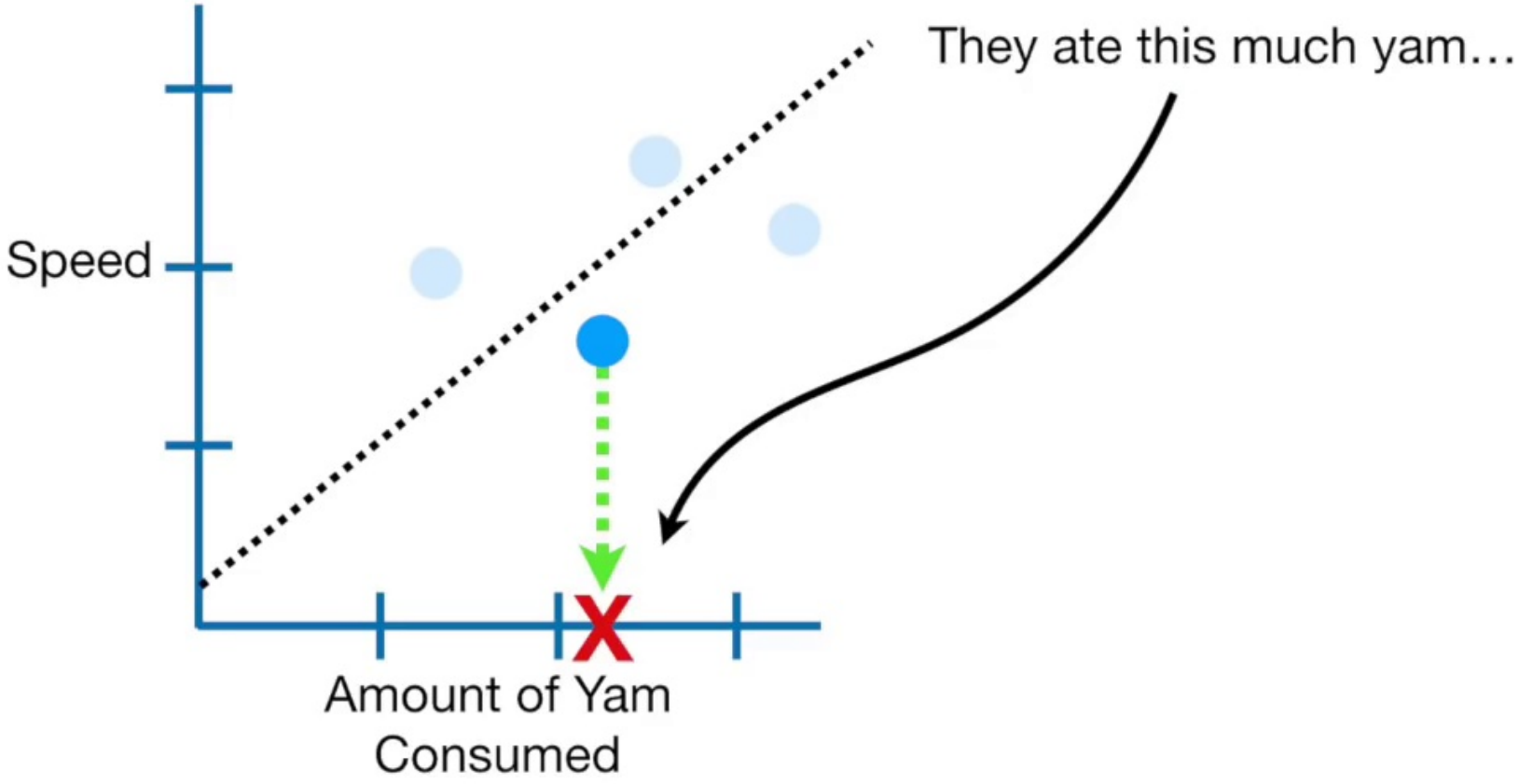

So let's measure the distance between the actual speed and the predicted speed…
Speed
Amount of Yam Consumed

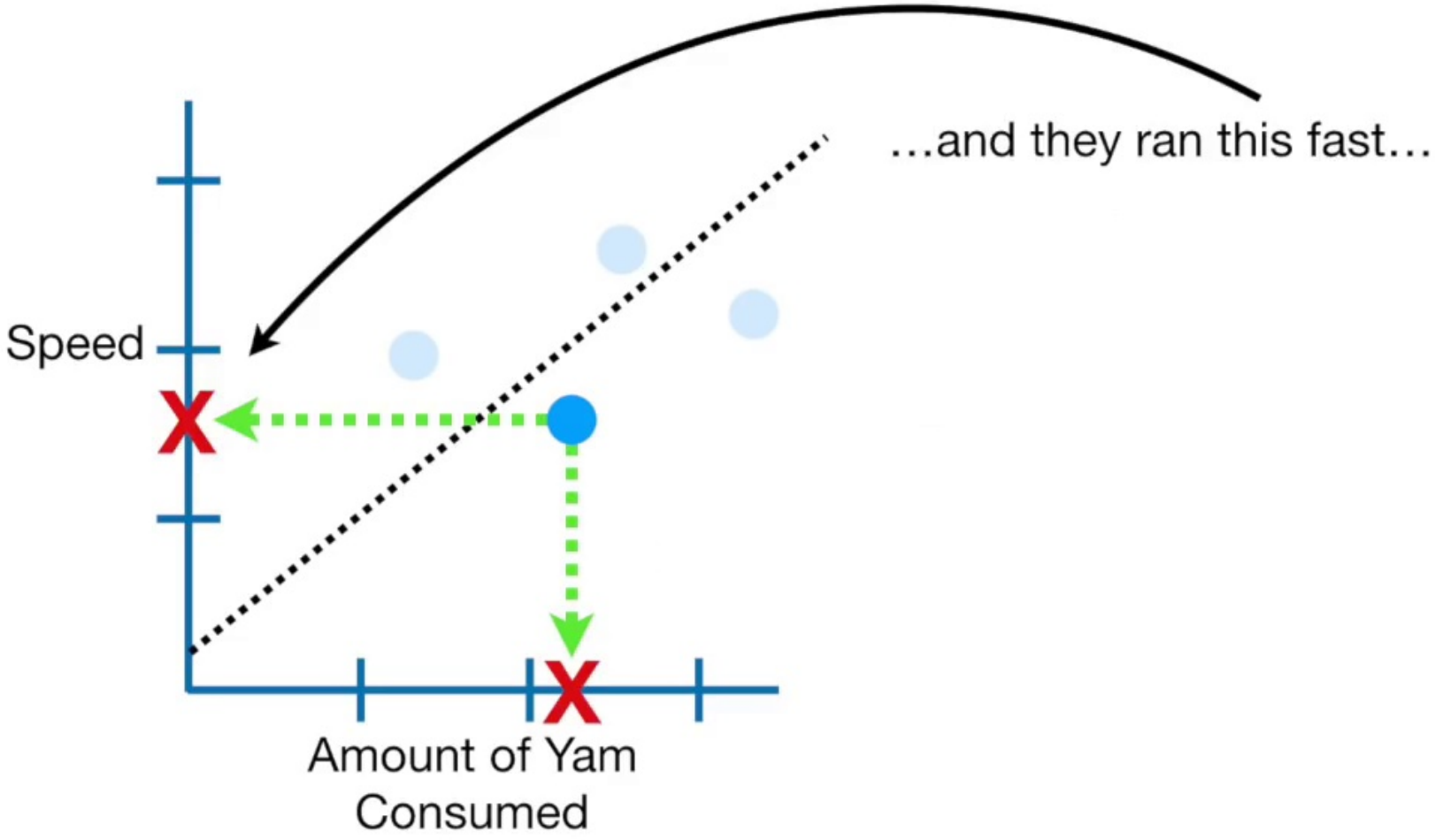

Speed
Amount of Yam Consumed
…and save the distance on the right while we focus on the other people in the **Testing Data**.

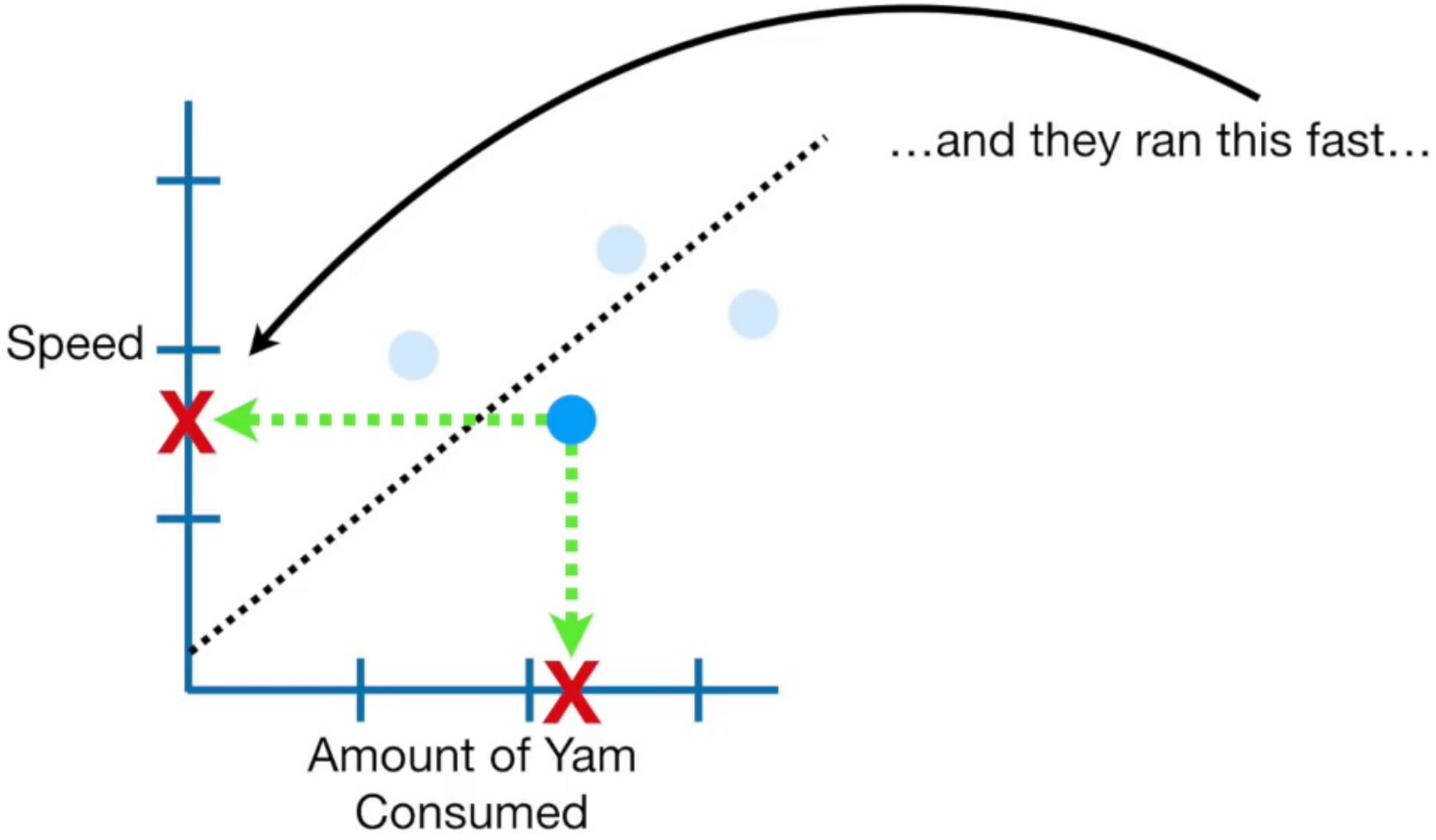

Speed
Amount of Yam Consumed
…and save the distance on the right while we focus on the other people in the **Testing Data**.

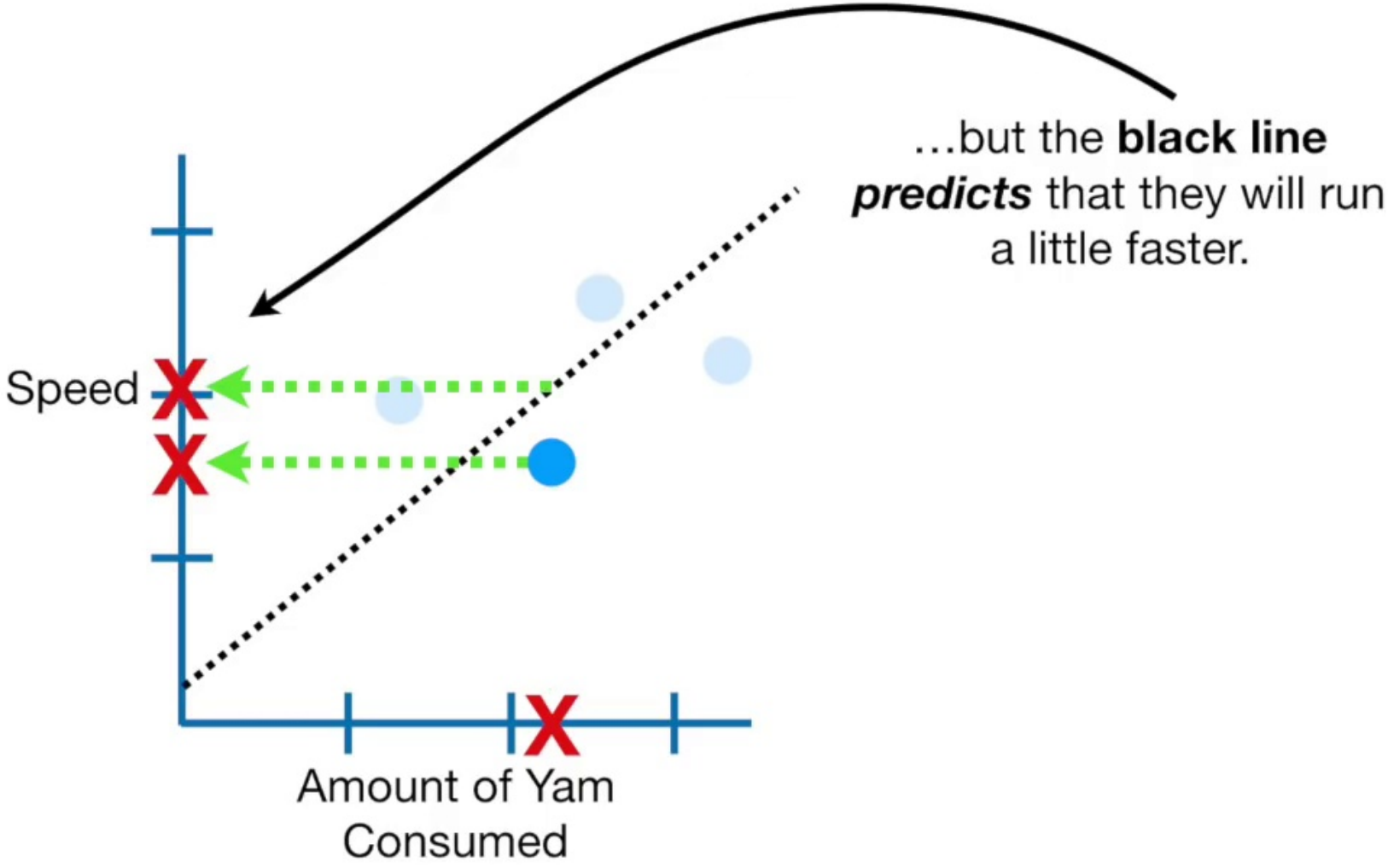

Here's the second person
in the Testing Data.
Speed
Amount of Yam
Consumed

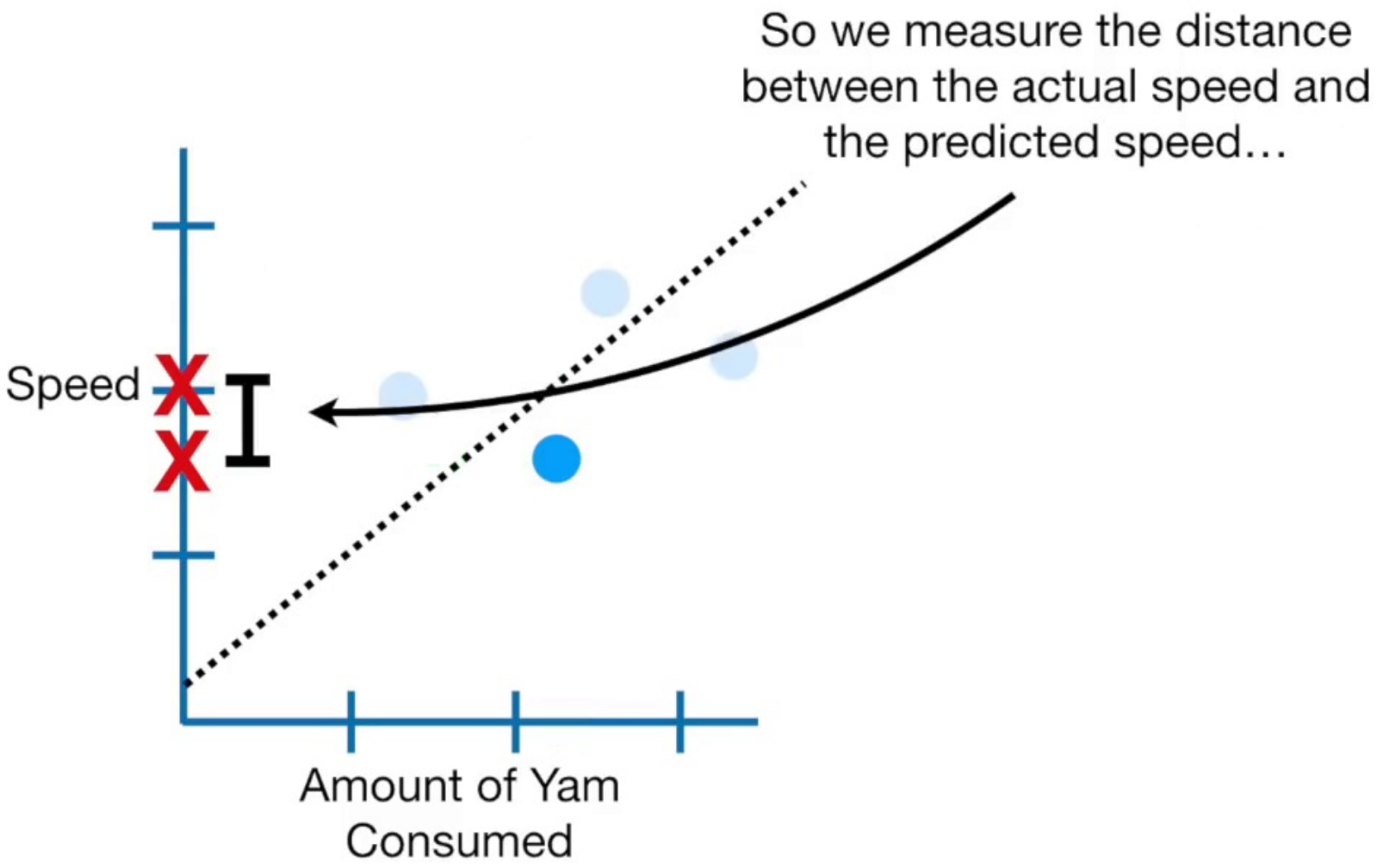

Here's the second person
in the Testing Data.
Speed
Amount of Yam
Consumed

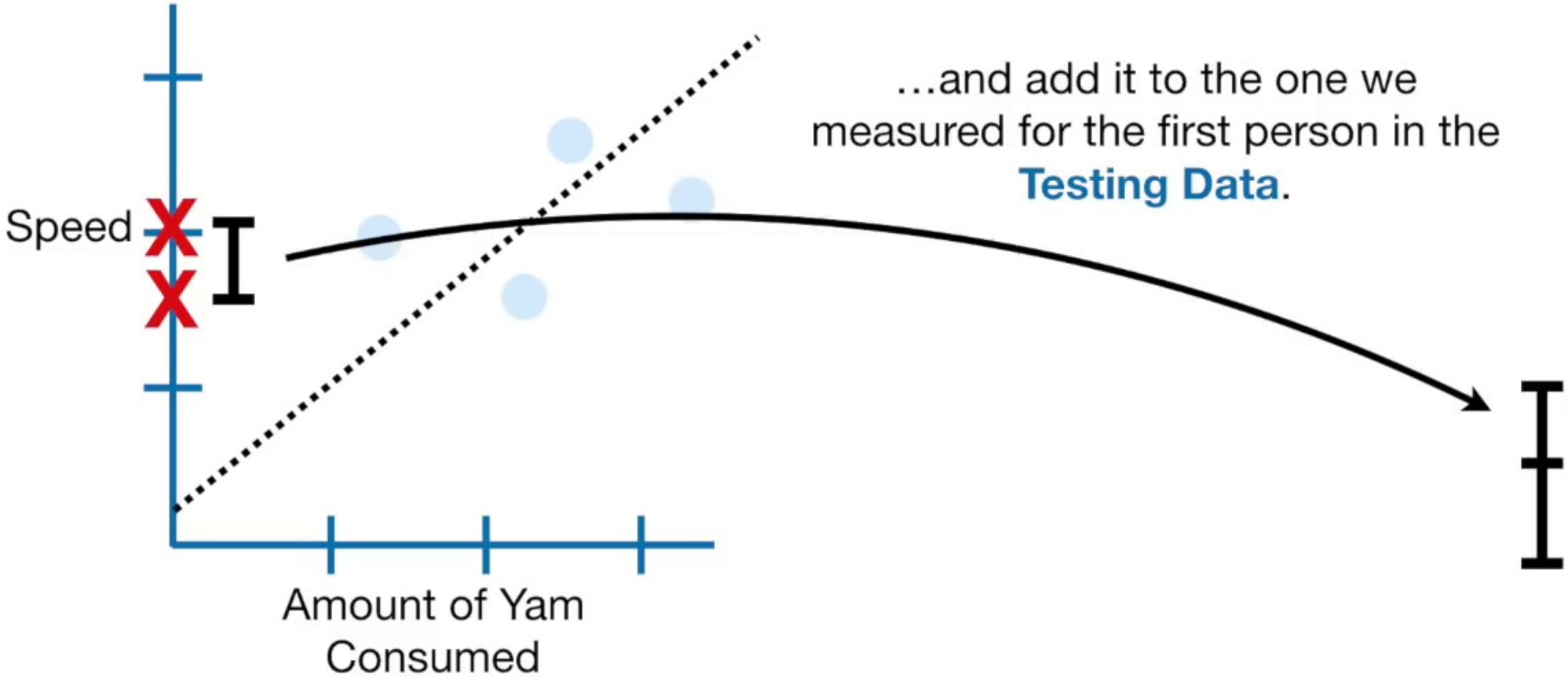

They ate this much yam…
Speed
Amount of Yam
Consumed

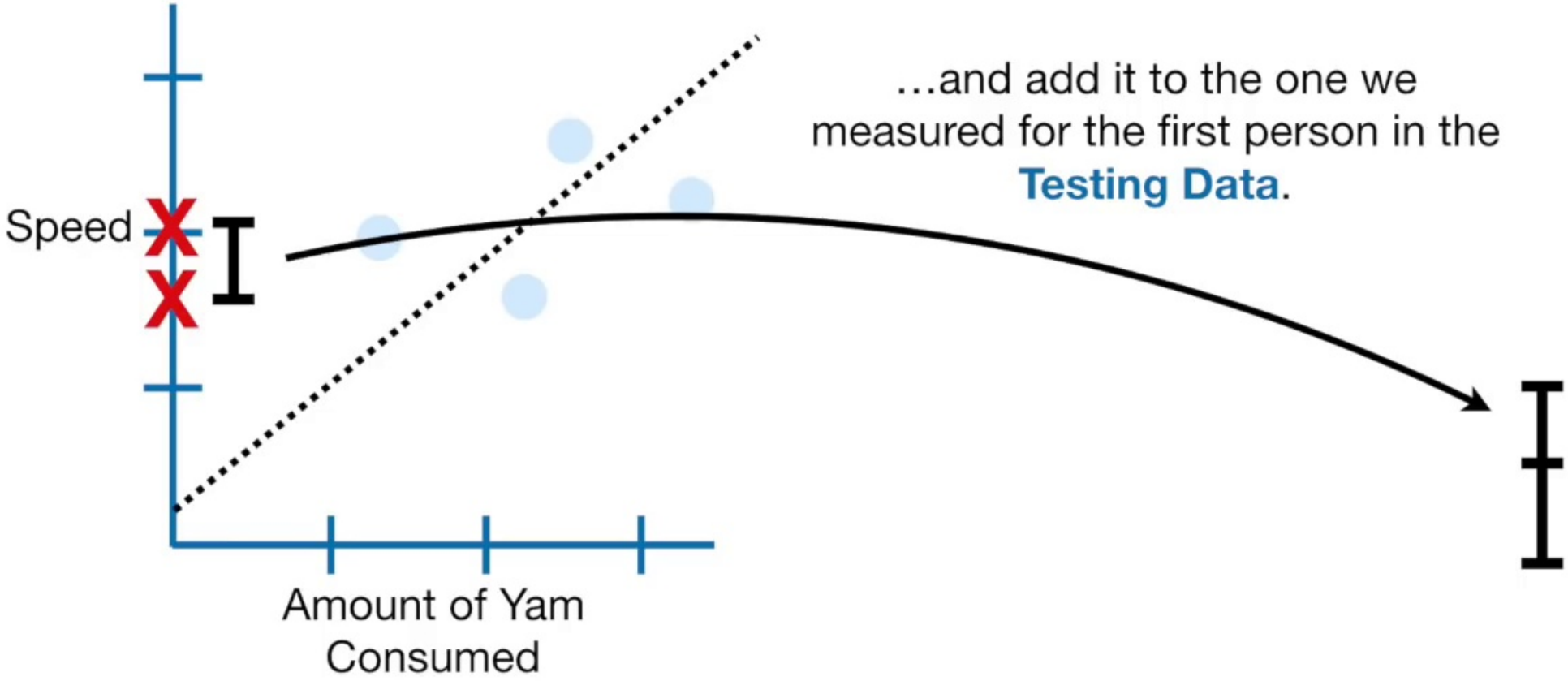

…and they ran this fast…
Speed
Amount of Yam
Consumed

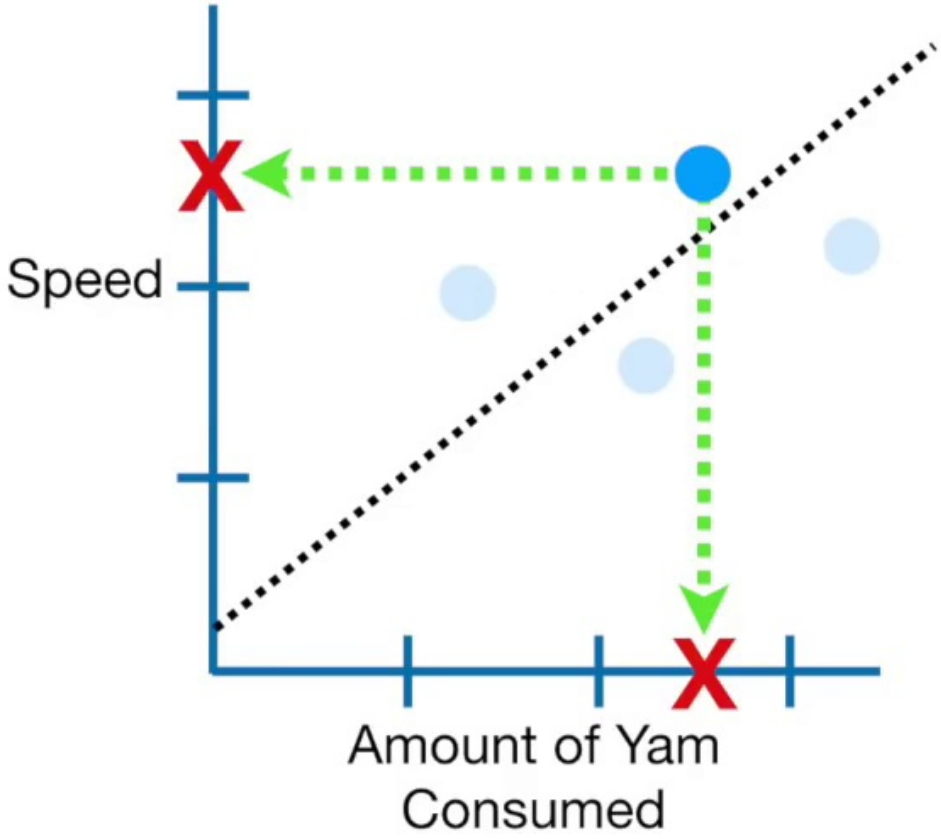

…and they ran this fast…
Speed
Amount of Yam
Consumed


…but the **black line** ***predicts*** that they will run a little faster.
Speed
Amount of Yam Consumed

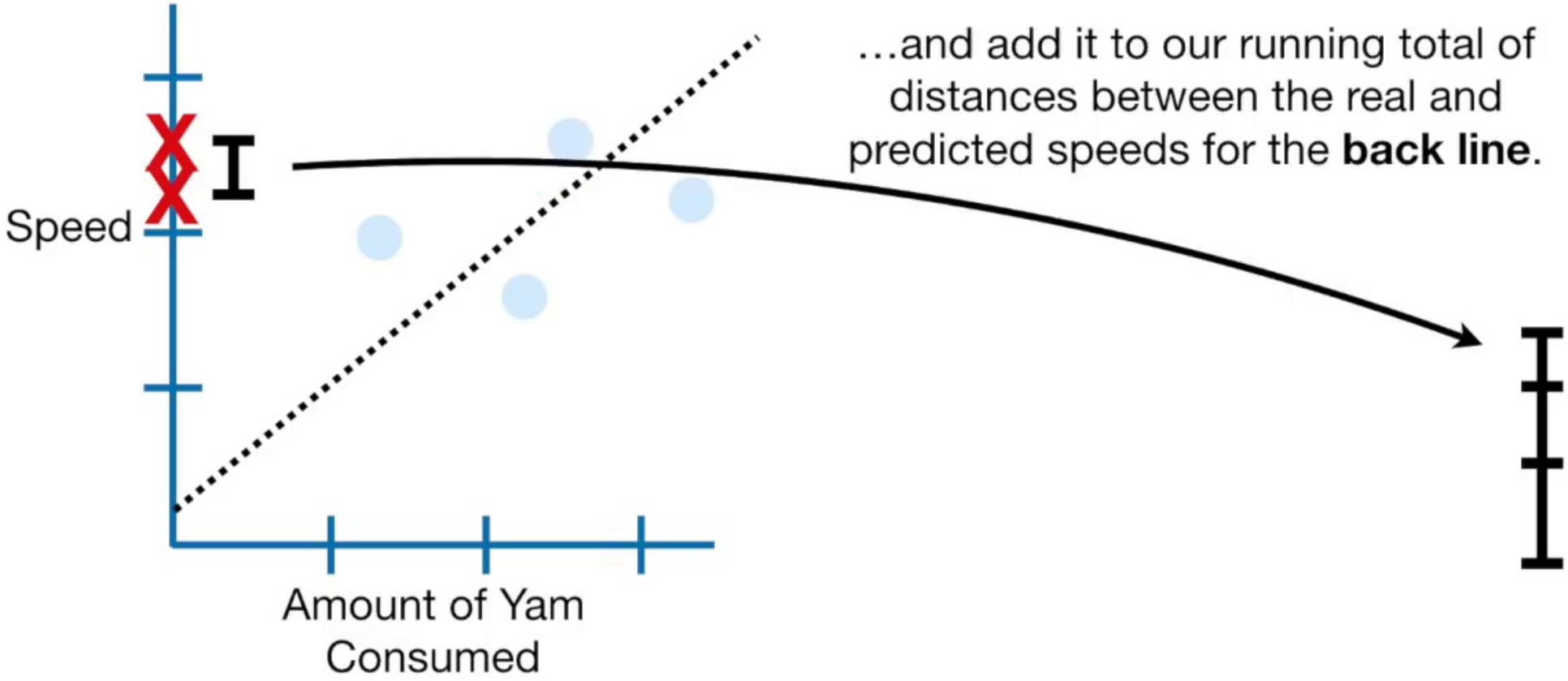

So we measure the distance between the actual speed and the predicted speed…
Speed
Amount of Yam Consumed

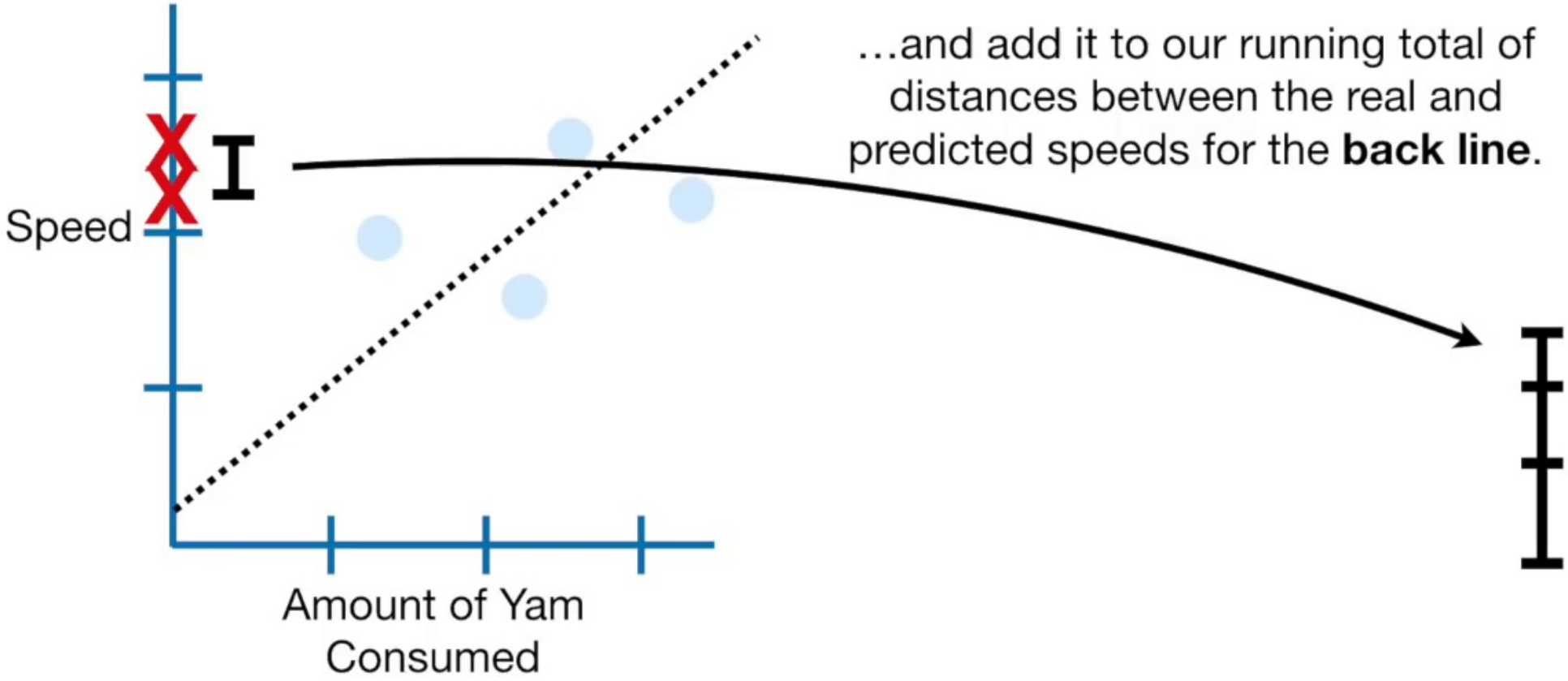

Speed
Amount of Yam Consumed
…and add it to the one we measured for the first person in the **Testing Data**.

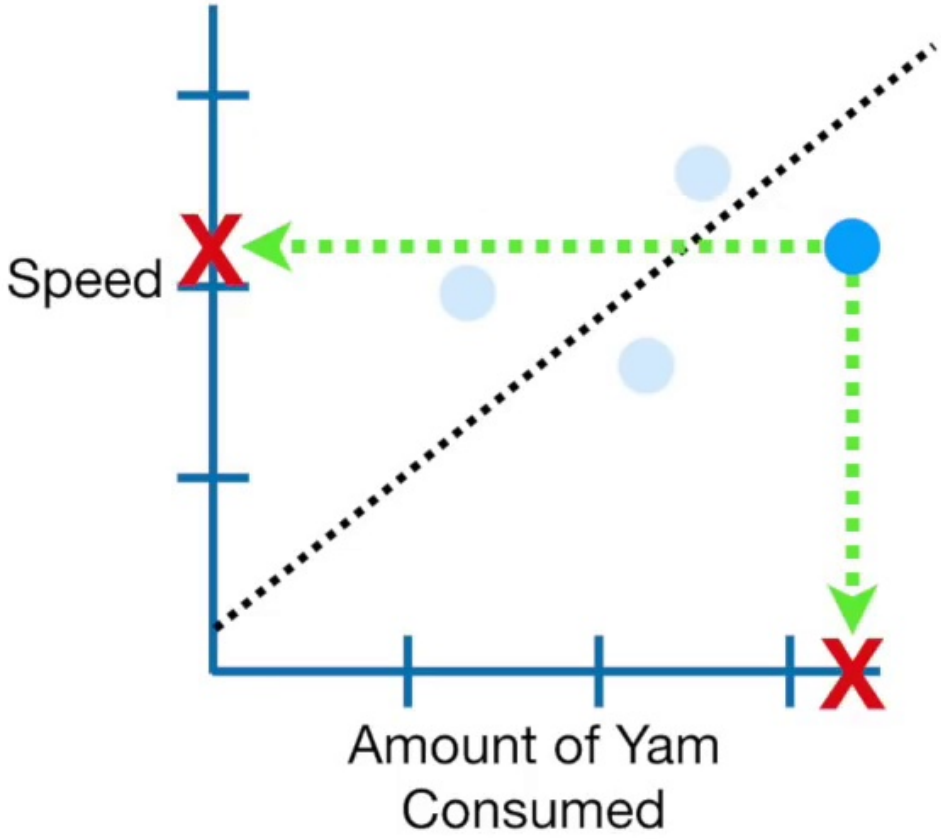

…and add it to the one we measured for the first person in the **Testing Data**.
Speed
Amount of Yam Consumed



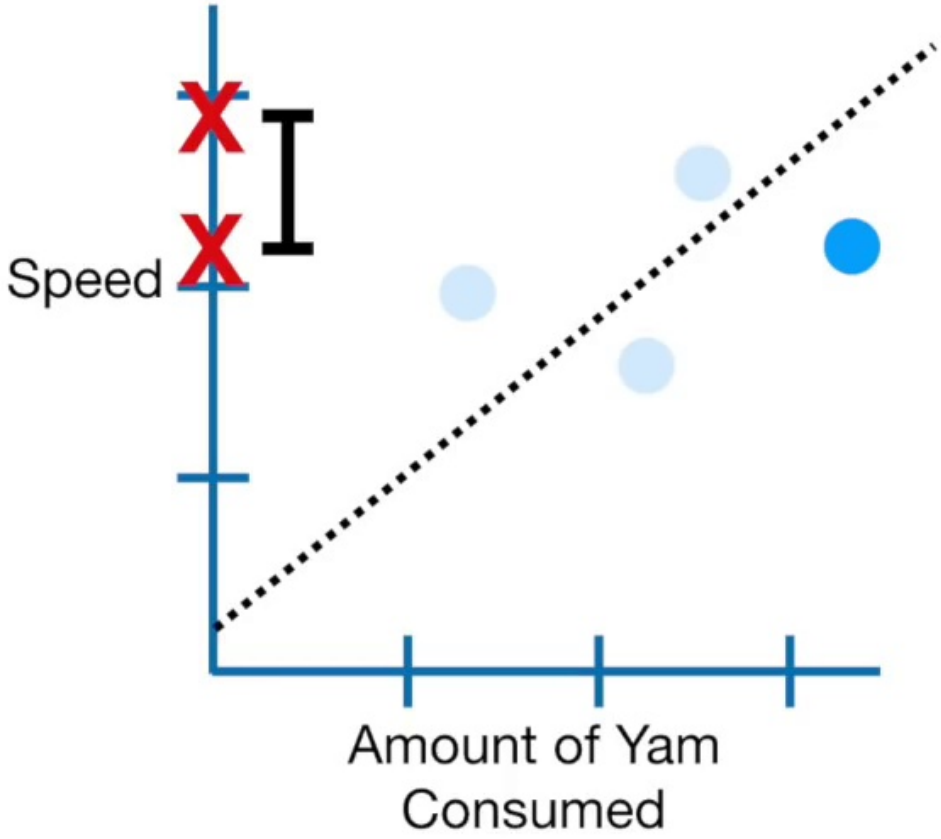
Then we measure the distance between the real and the predicted speed for the third person in the **Testing Data**…


…and add it to our running total of distances between the real and predicted speeds for the **back line**.
Speed
Amount of Yam Consumed

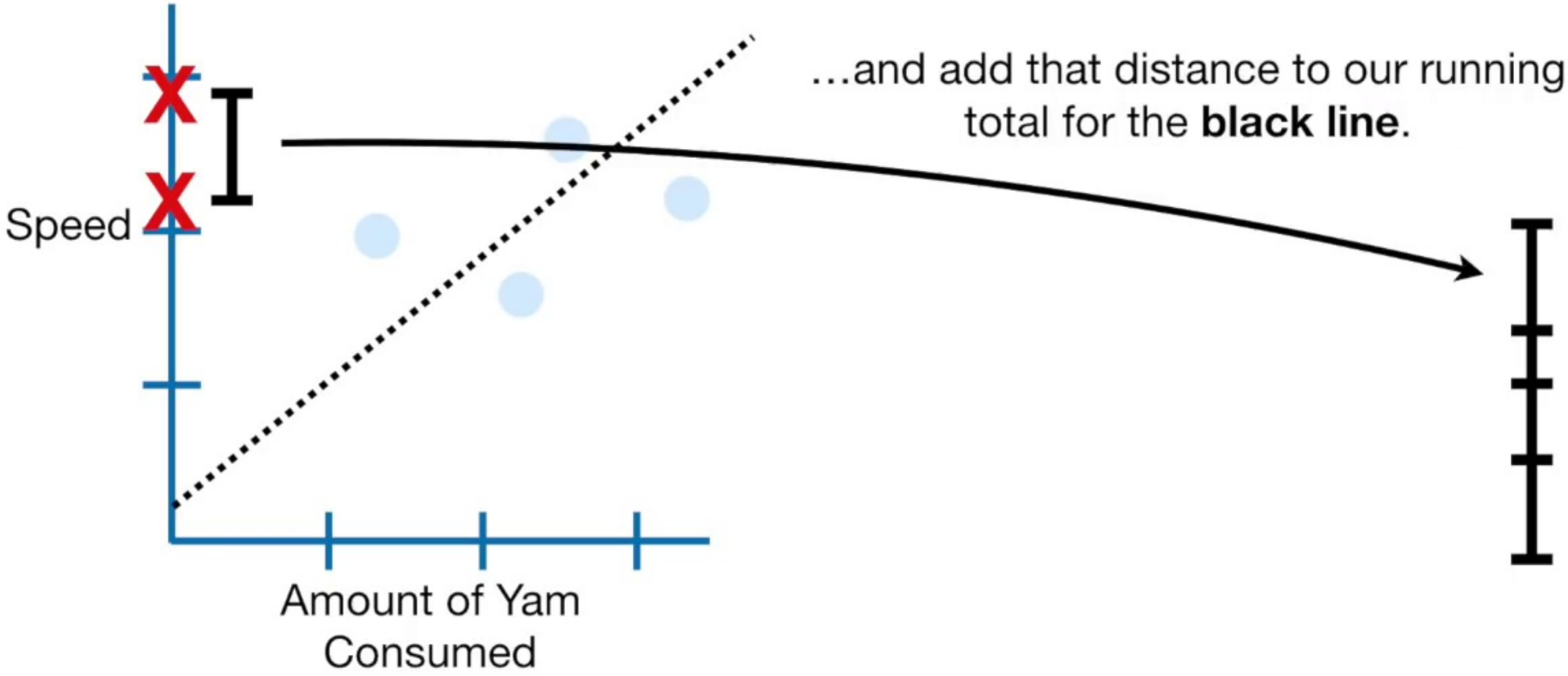

…and add it to our running total of distances between the real and predicted speeds for the **back line**.
Speed
Amount of Yam Consumed

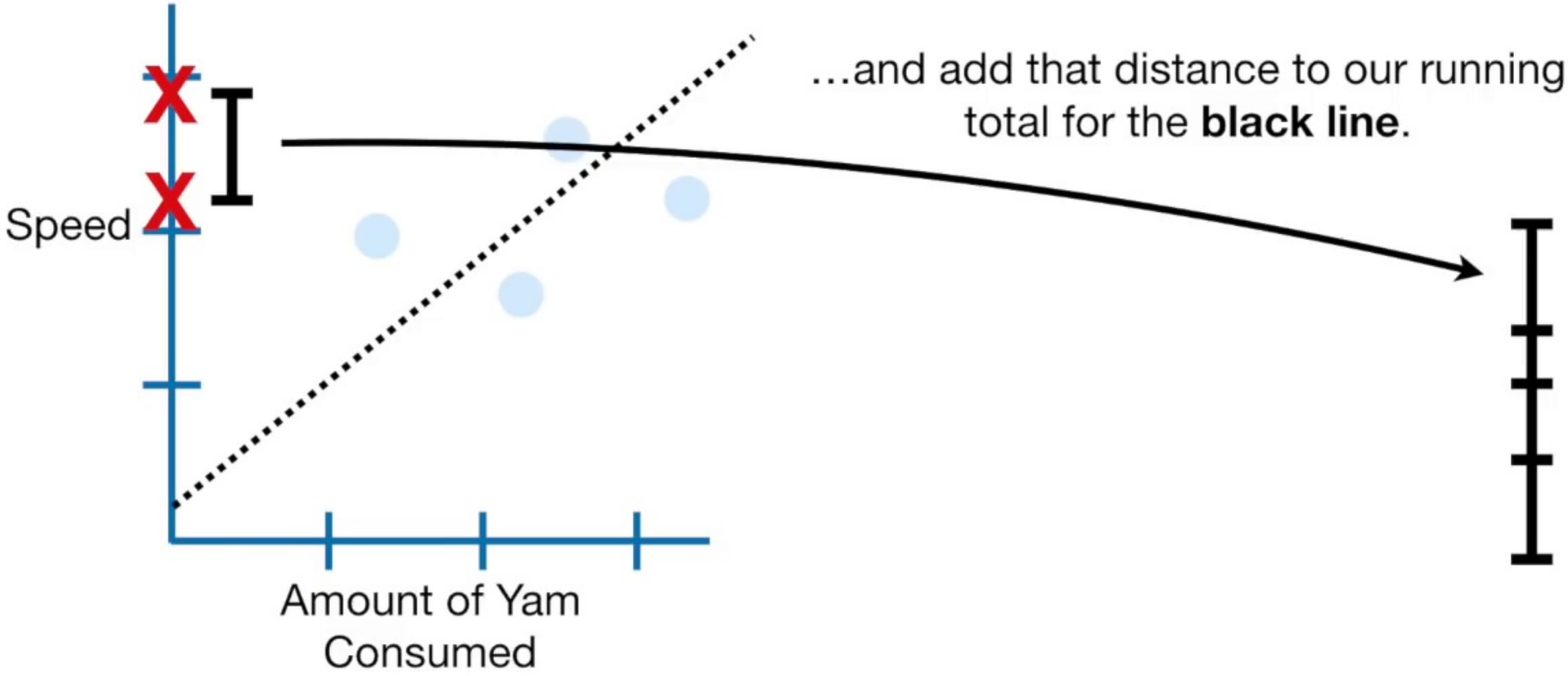


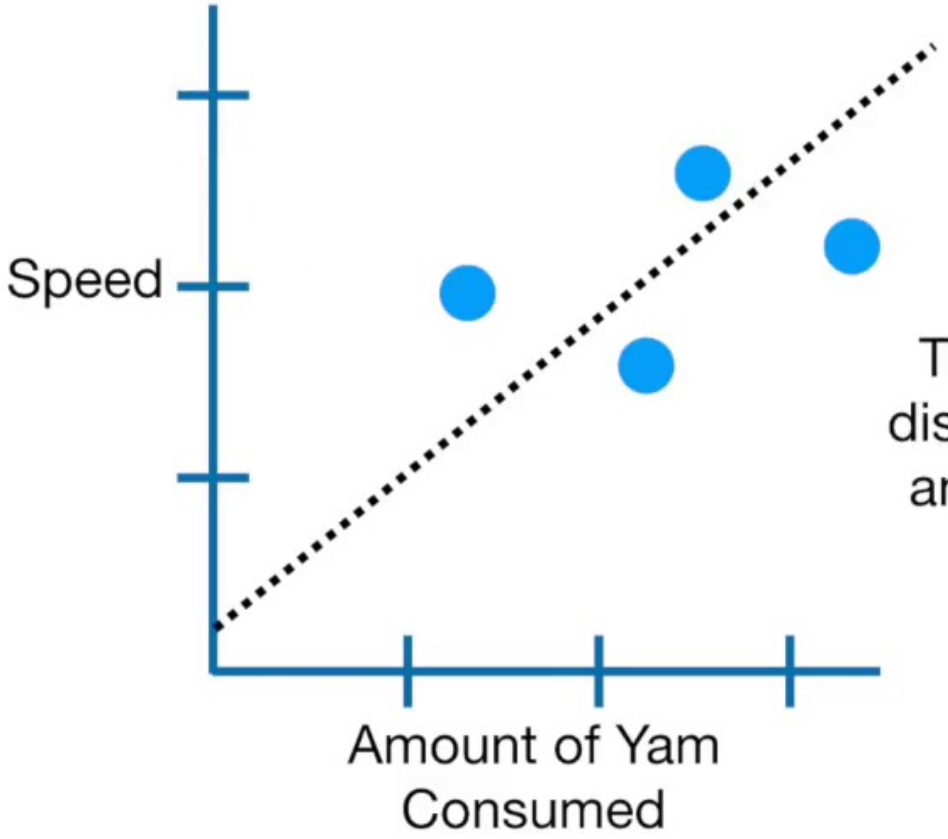
Then we do the same thing for the fourth person in the **Testing Data**.



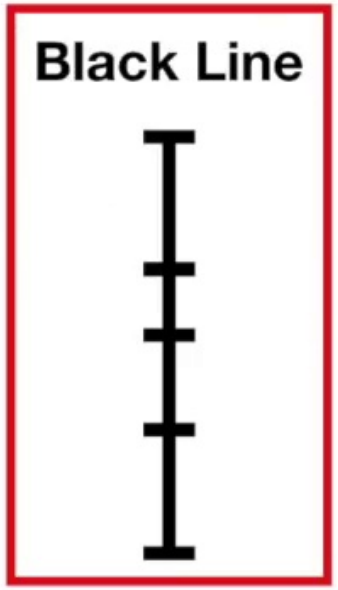
Then we do the same thing for the fourth person in the **Testing Data**.

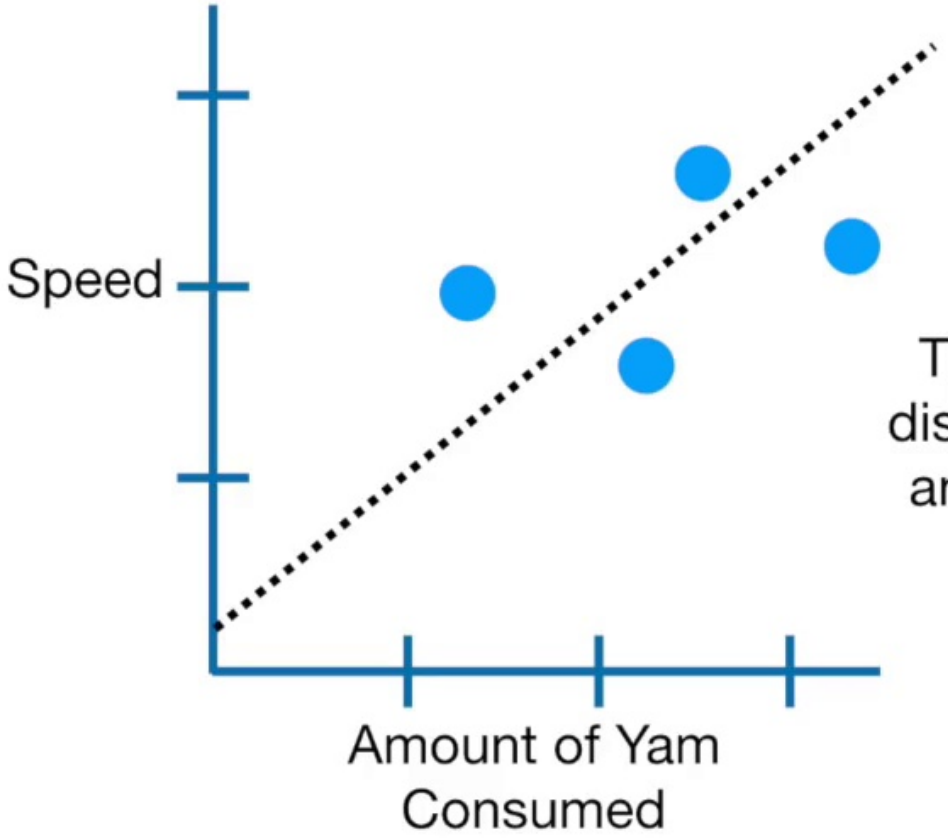

…and add that distance to our running total for the **black line**.
Speed
Amount of Yam Consumed


…and add that distance to our running total for the **black line**.
Speed
Amount of Yam Consumed

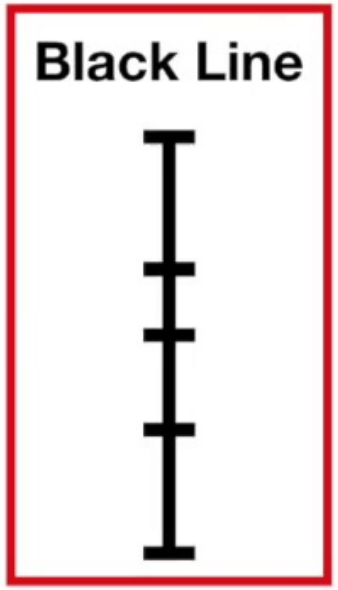


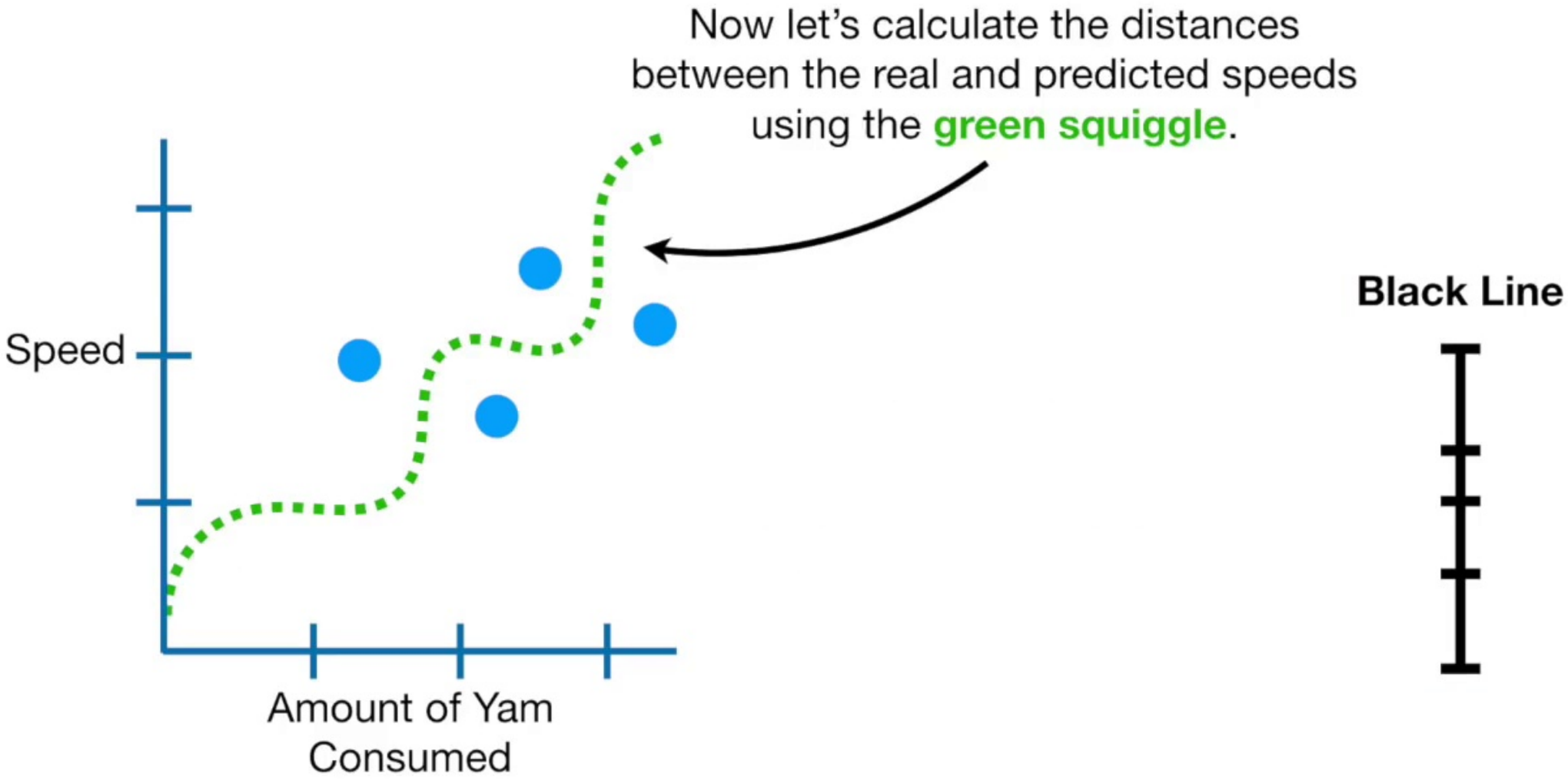
This is the sum of all the distances between the real and predicted speeds for the **black line**.

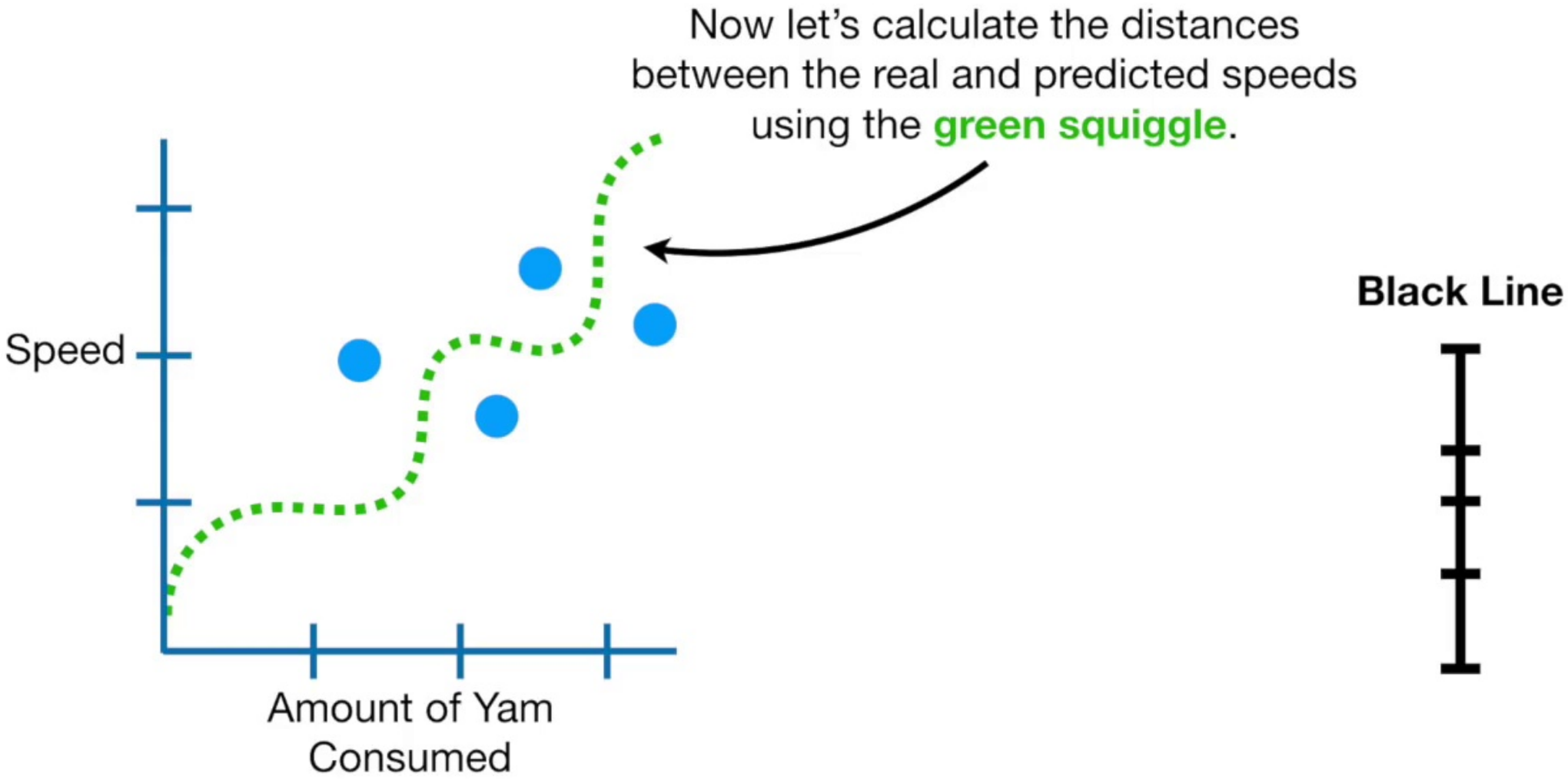


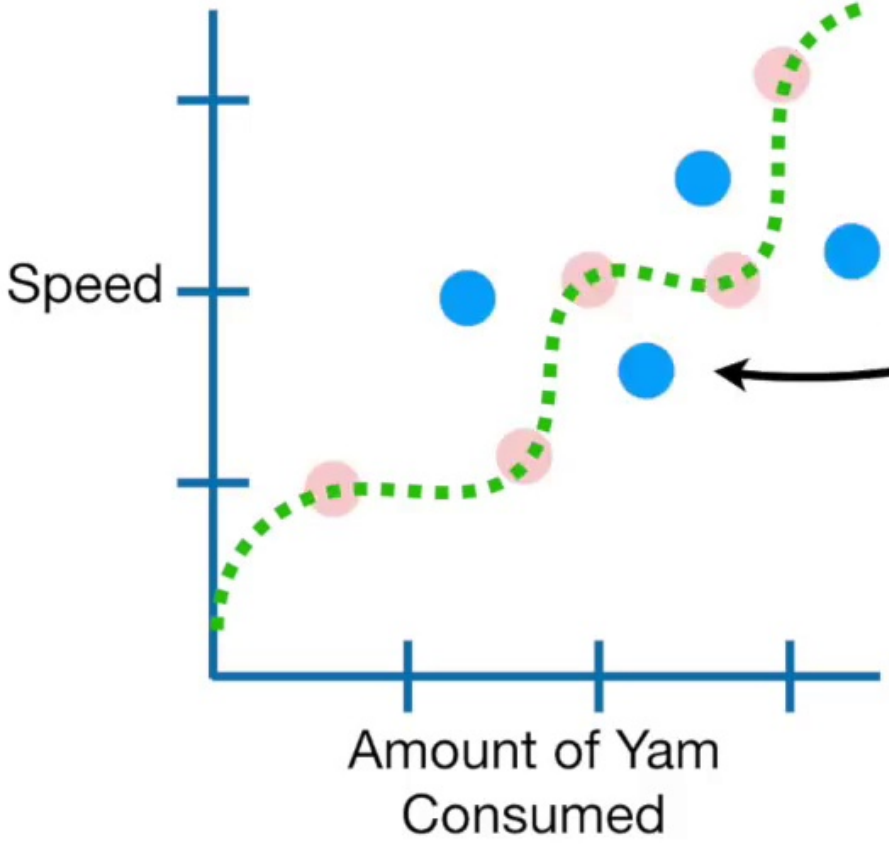


This is the sum of all the distances between the real and predicted speeds for the **black line**.



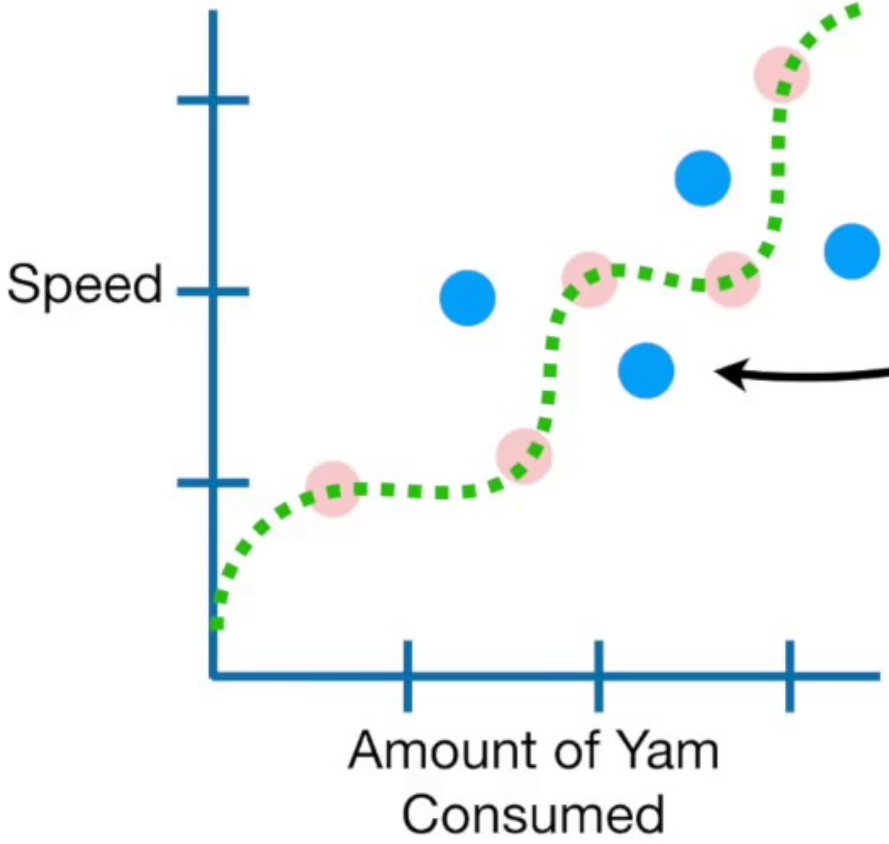

Now let's calculate the distances between the real and predicted speeds using the green squiggle.
Speed
Amount of Yam Consumed
Black Line


Now let's calculate the distances between the real and predicted speeds using the green squiggle.
Speed
Amount of Yam Consumed
Black Line



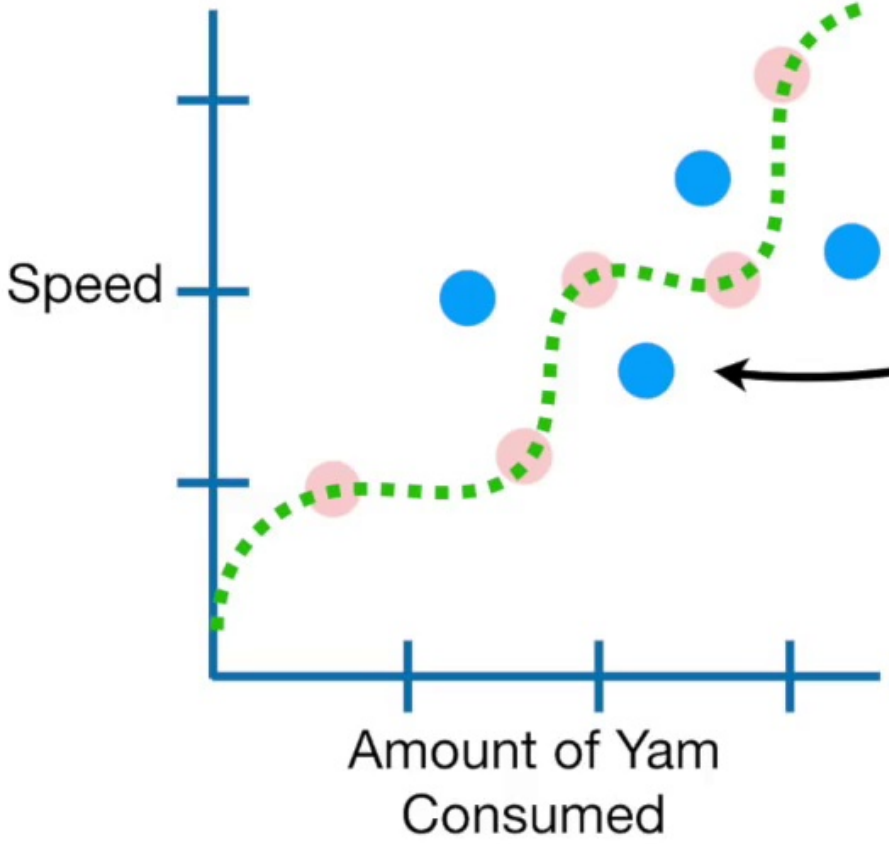
…but when we are doing **Machine Learning**, we're more interested in how well the **green squiggle** can make ***predictions*** with new data.

Black Line



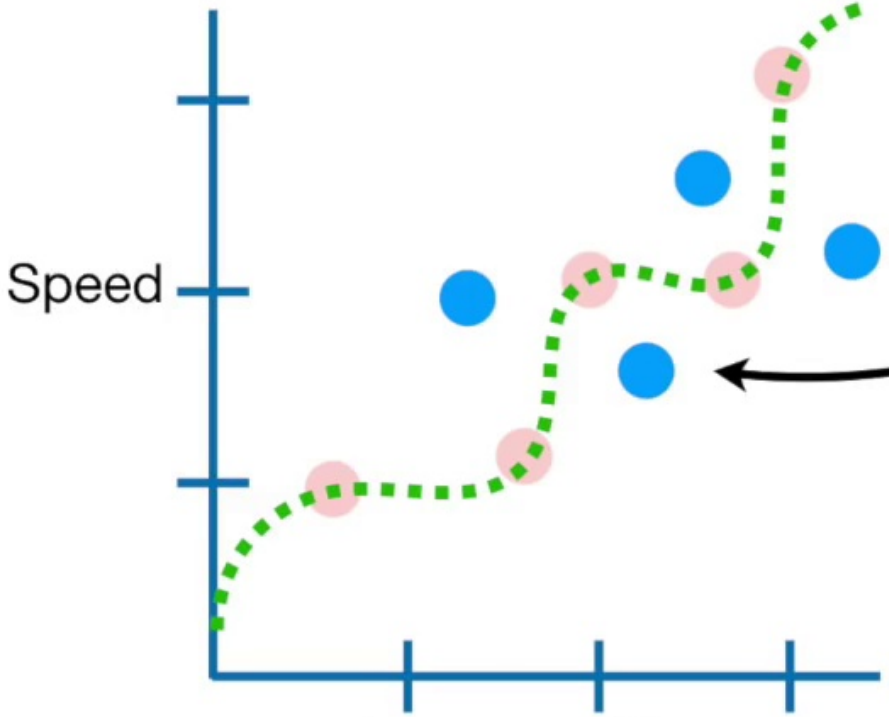
…but when we are doing **Machine Learning**, we're more interested in how well the **green squiggle** can make ***predictions*** with new data.

Black Line



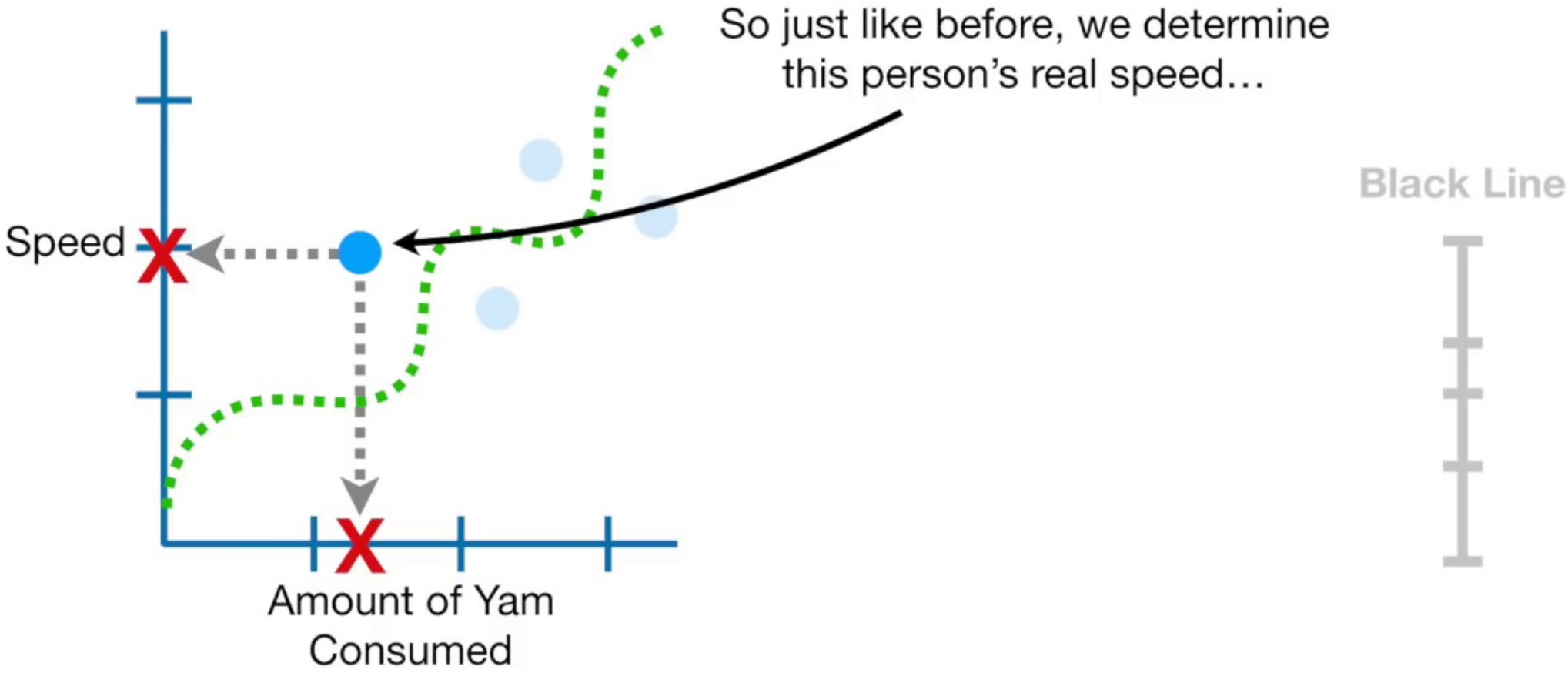
…but when we are doing **Machine Learning**, we're more interested in how well the **green squiggle** can make ***predictions*** with new data.

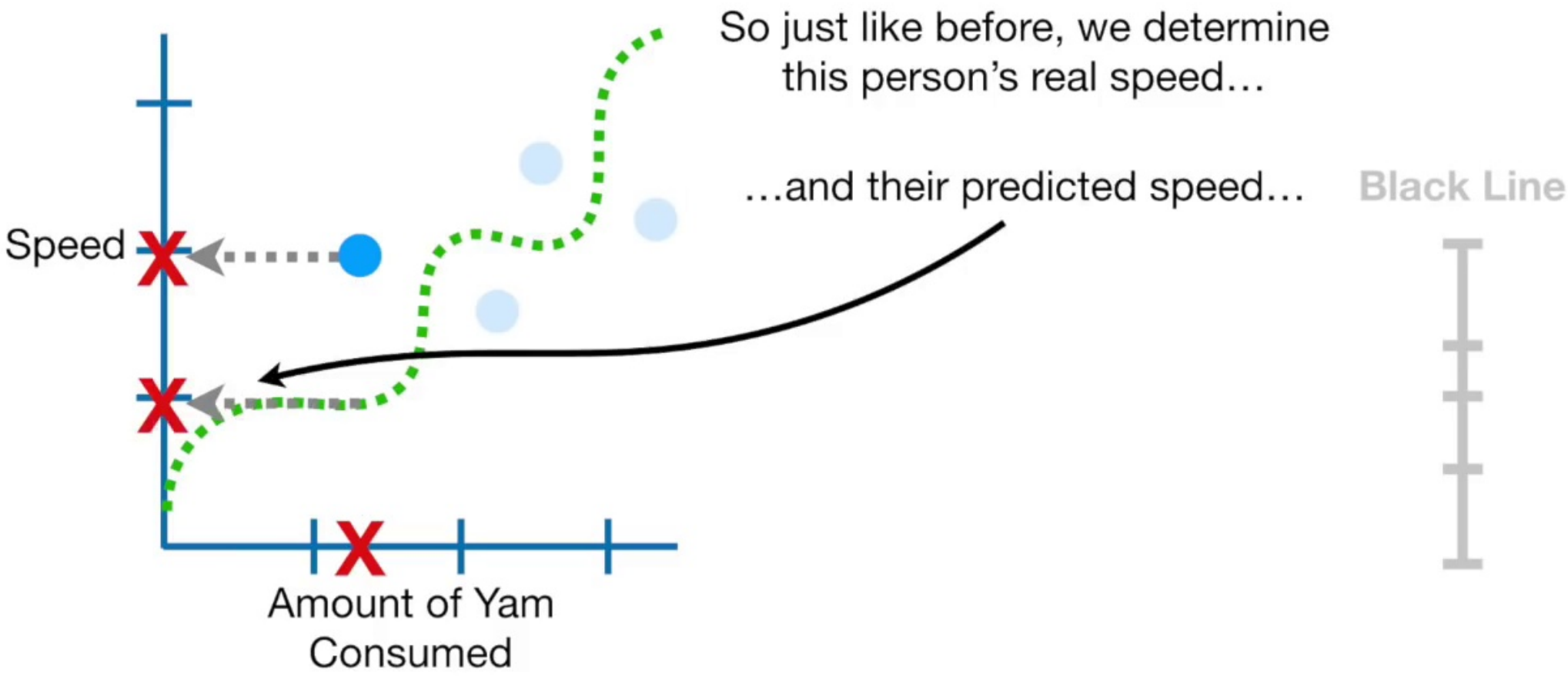
Black Line

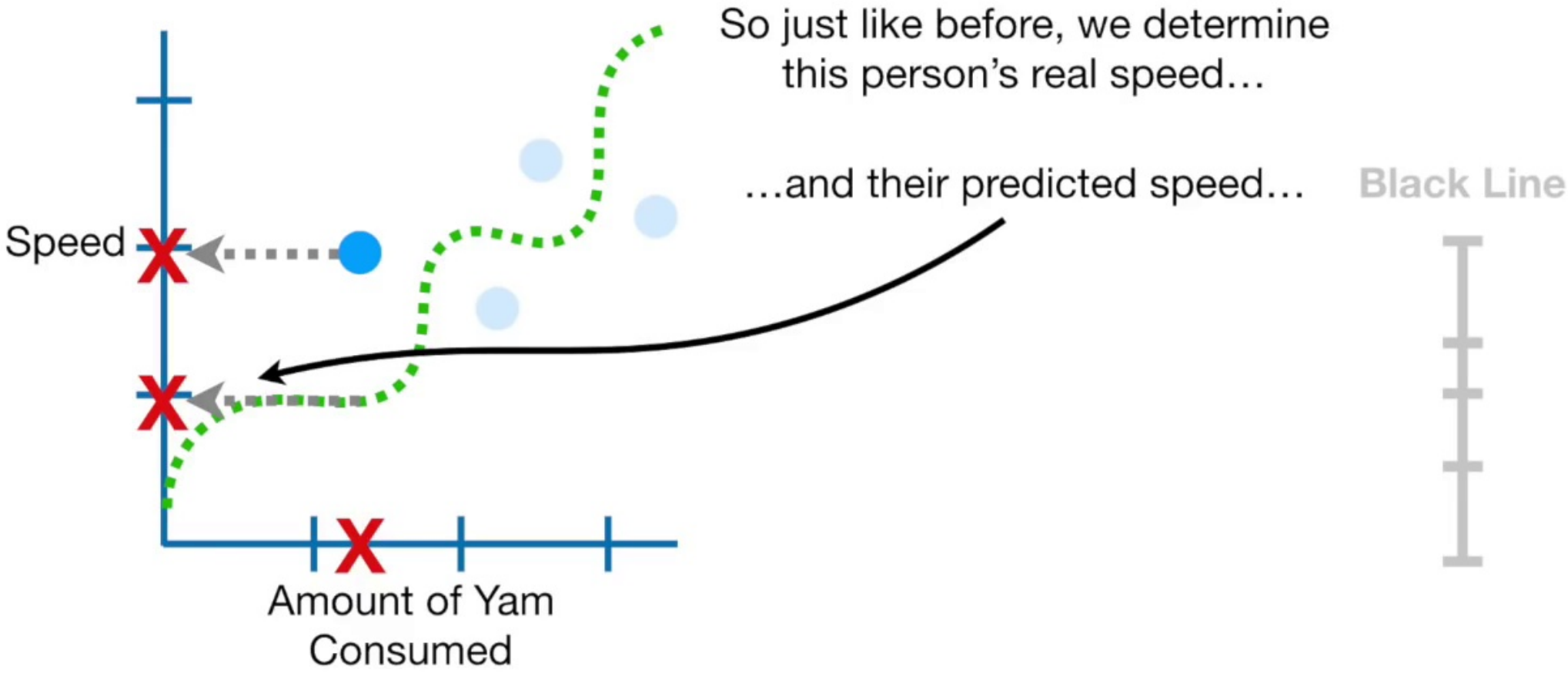


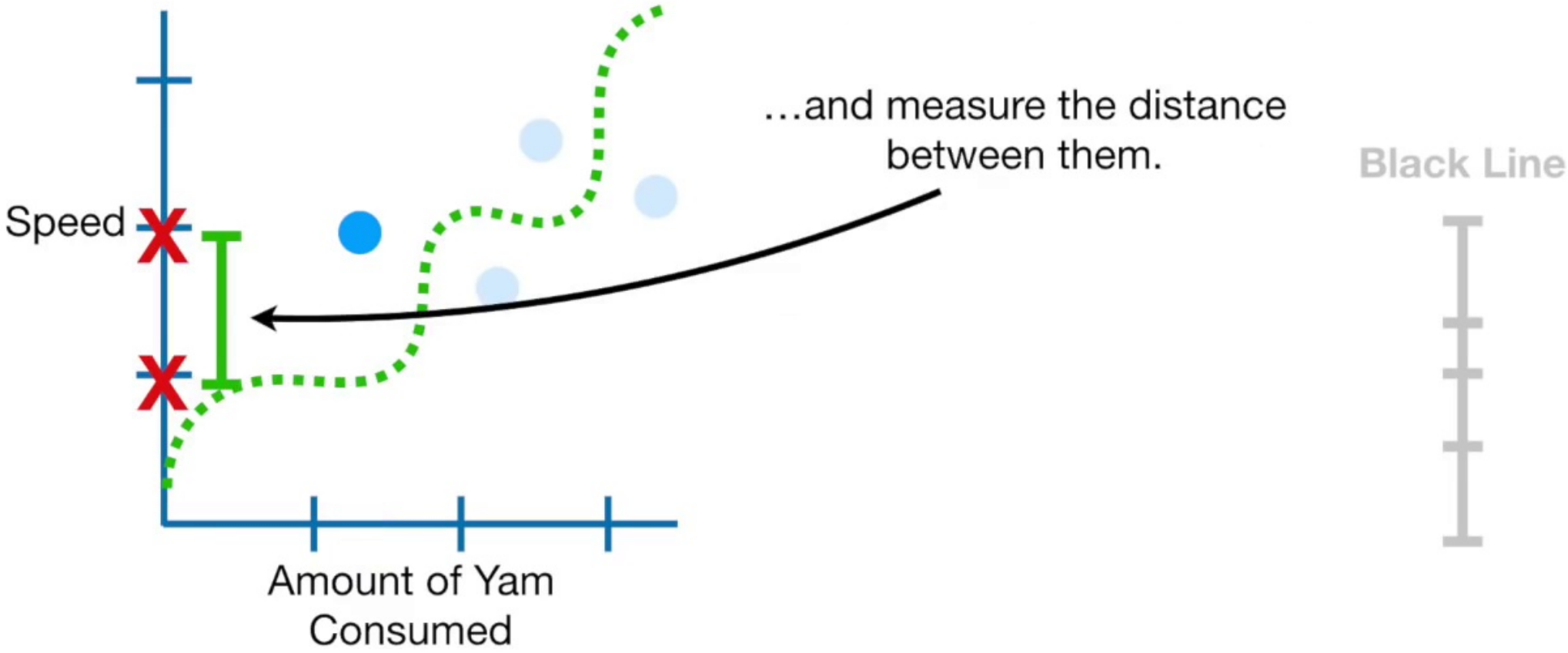
…but when we are doing **Machine Learning**, we're more interested in how well the **green squiggle** can make ***predictions*** with new data.

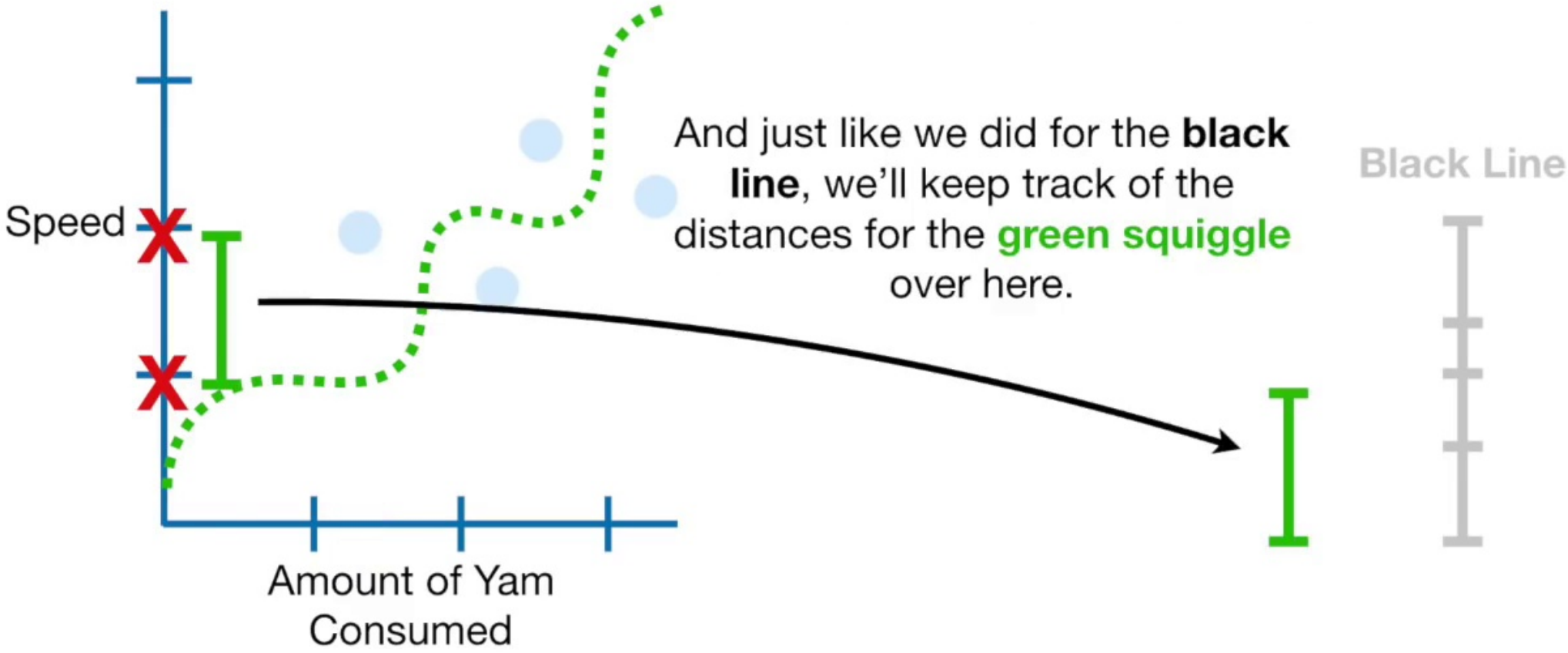
Black Line

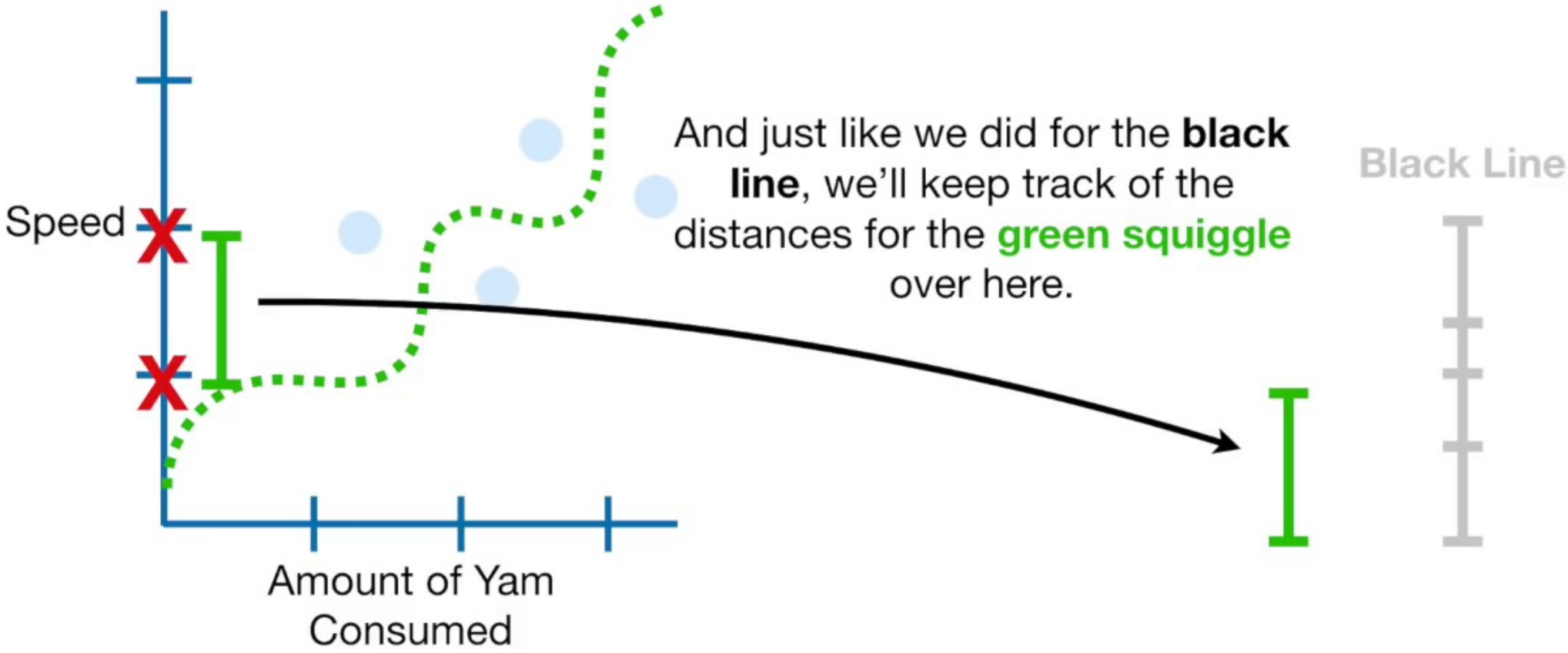

So just like before, we determine this person's real speed…
Speed
Amount of Yam Consumed
Black Line

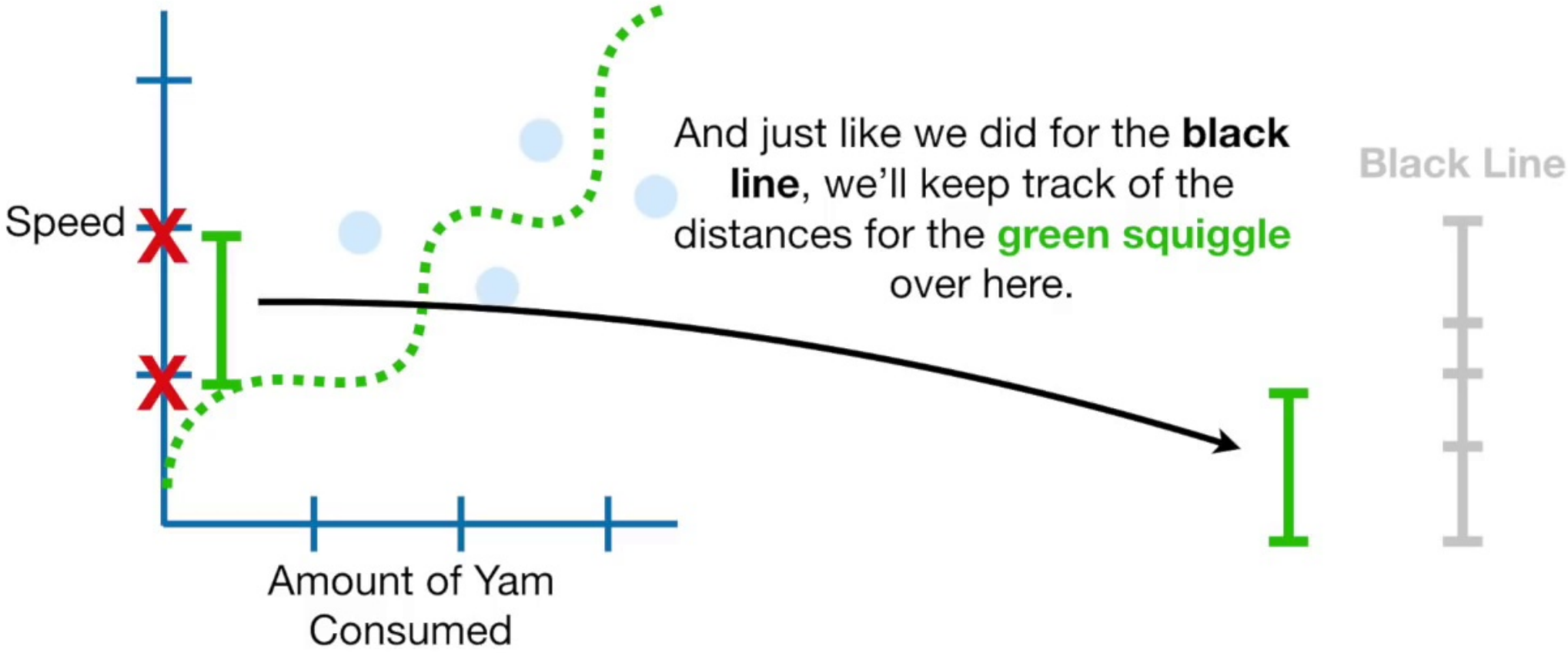

So just like before, we determine this person's real speed…
…and their predicted speed…
Black Line
Speed
Amount of Yam Consumed

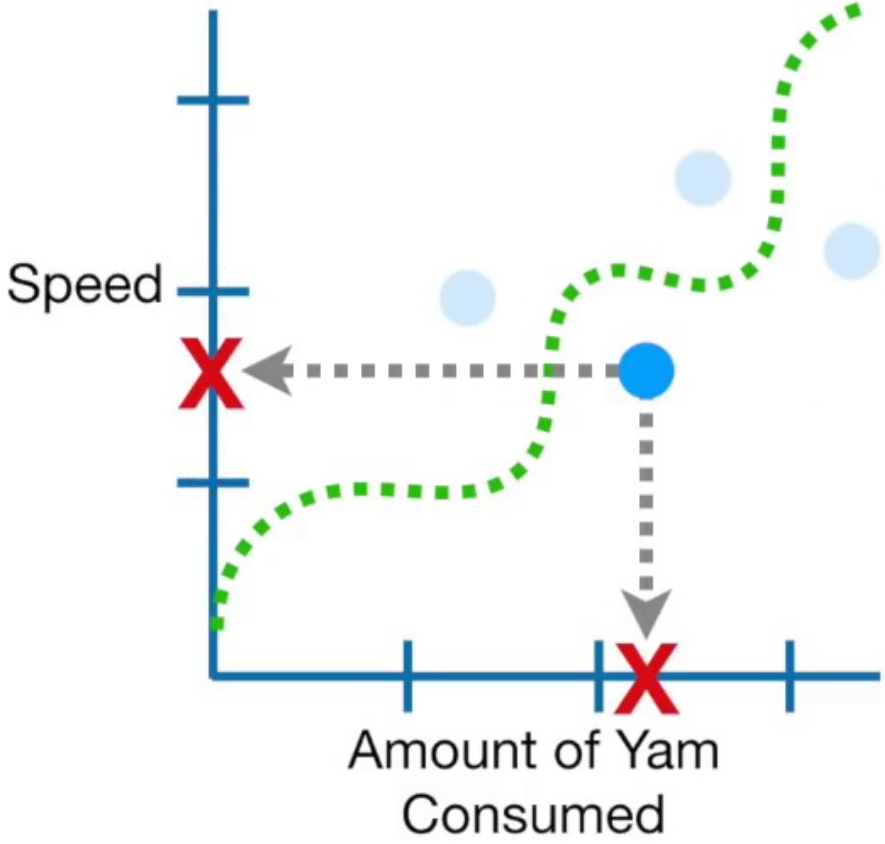

So just like before, we determine this person's real speed…
…and their predicted speed…
Black Line
Speed
Amount of Yam Consumed


…and measure the distance between them.
Black Line
Speed
Amount of Yam Consumed

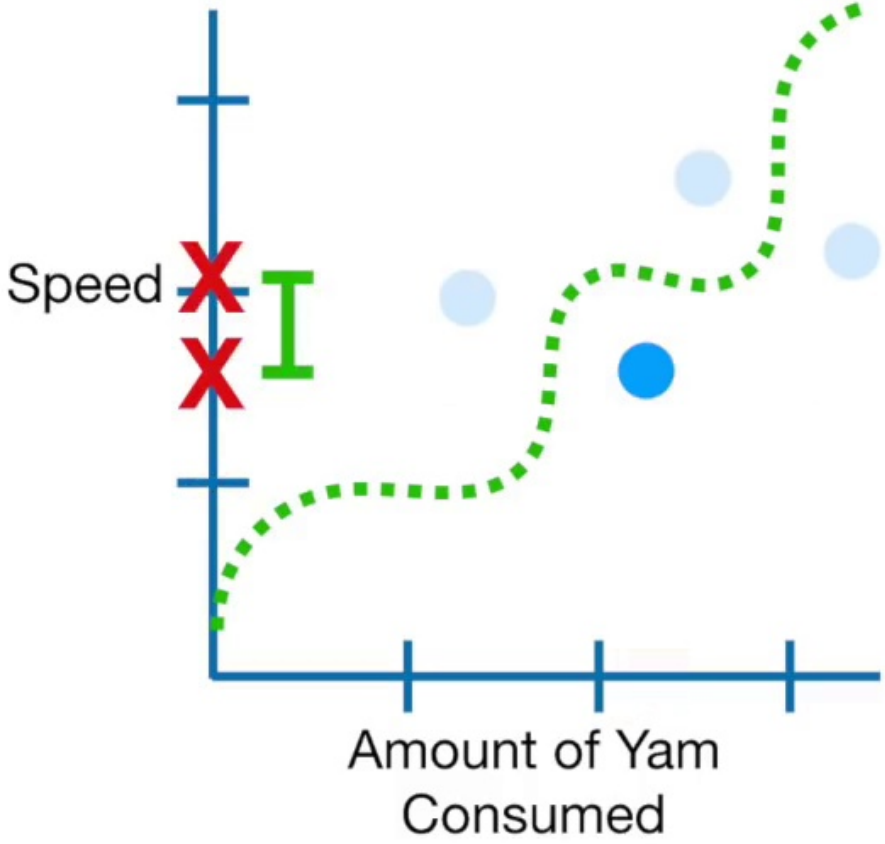

Speed
Amount of Yam Consumed
And just like we did for the **black line**, we'll keep track of the distances for the **green squiggle** over here.
Black Line


Speed
Amount of Yam Consumed
And just like we did for the **black line**, we'll keep track of the distances for the **green squiggle** over here.
Black Line

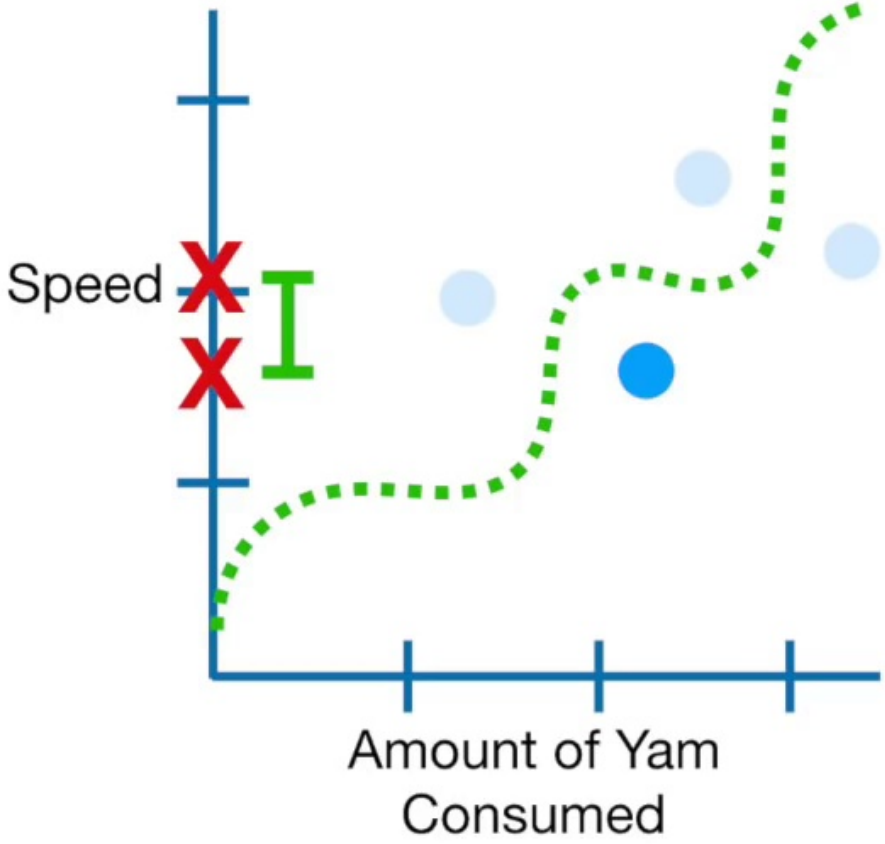

Speed
Amount of Yam Consumed
And just like we did for the **black line**, we'll keep track of the distances for the **green squiggle** over here.
Black Line



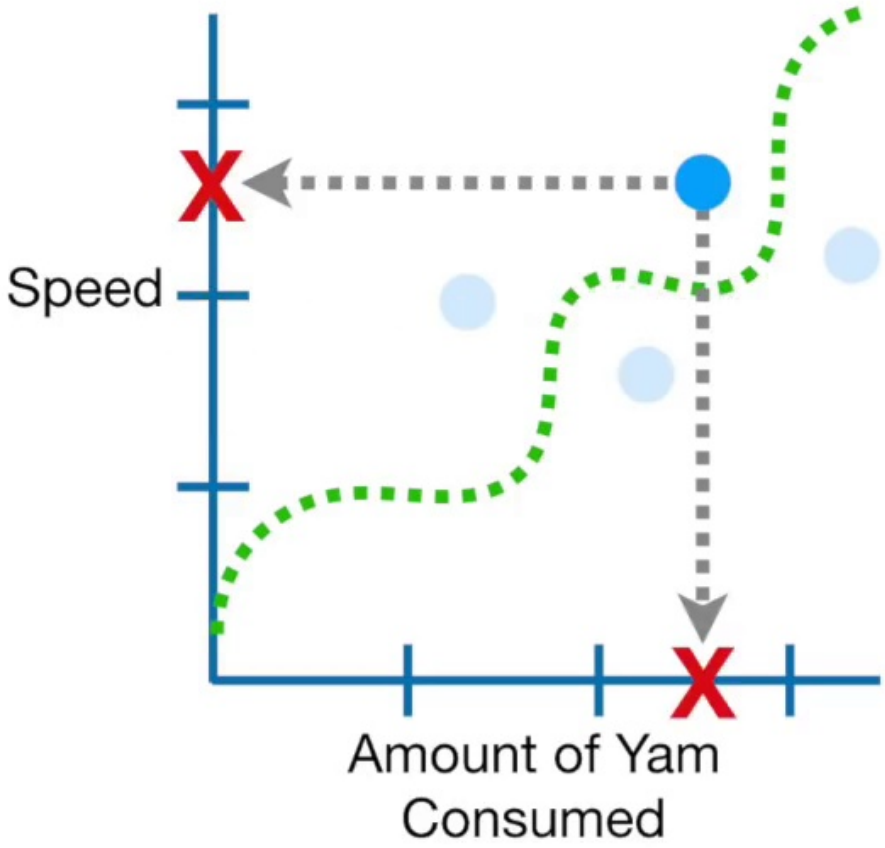
Then we do the same thing for the second person in the **Testing Data**…



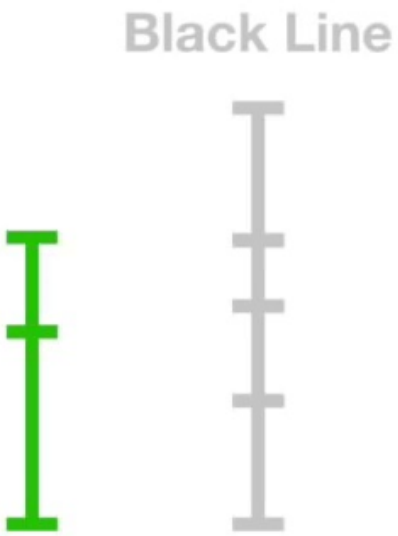
Then we do the same thing for the second person in the **Testing Data**…

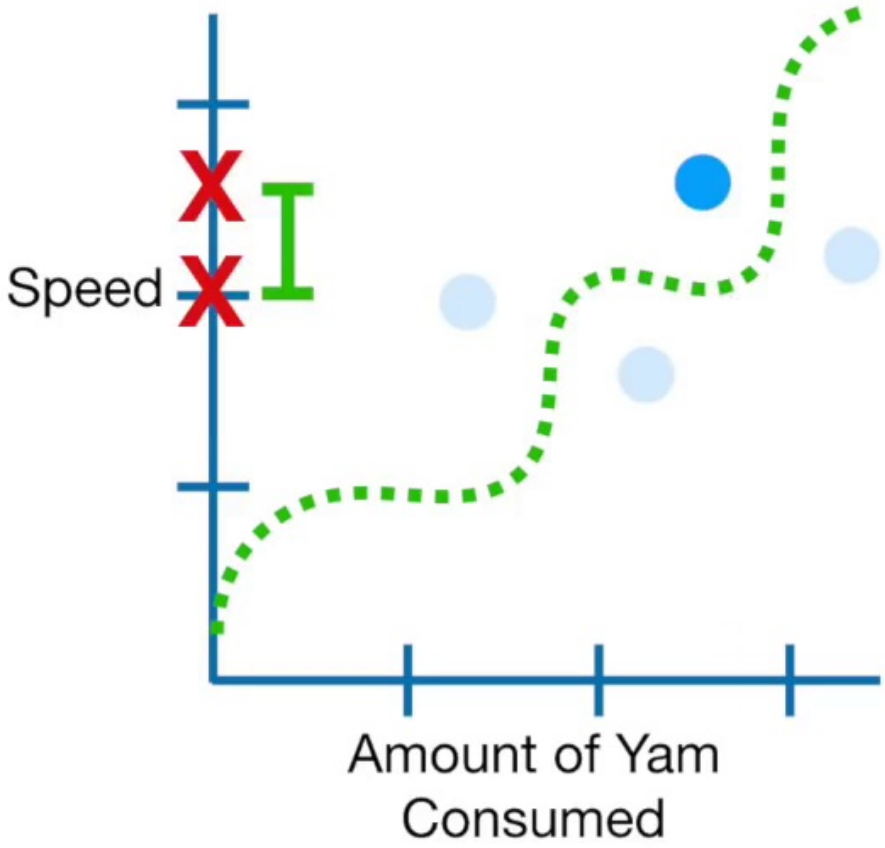


Then we do the same thing for the second person in the **Testing Data**…

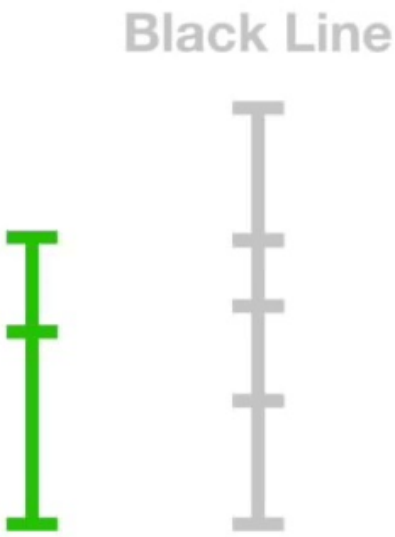


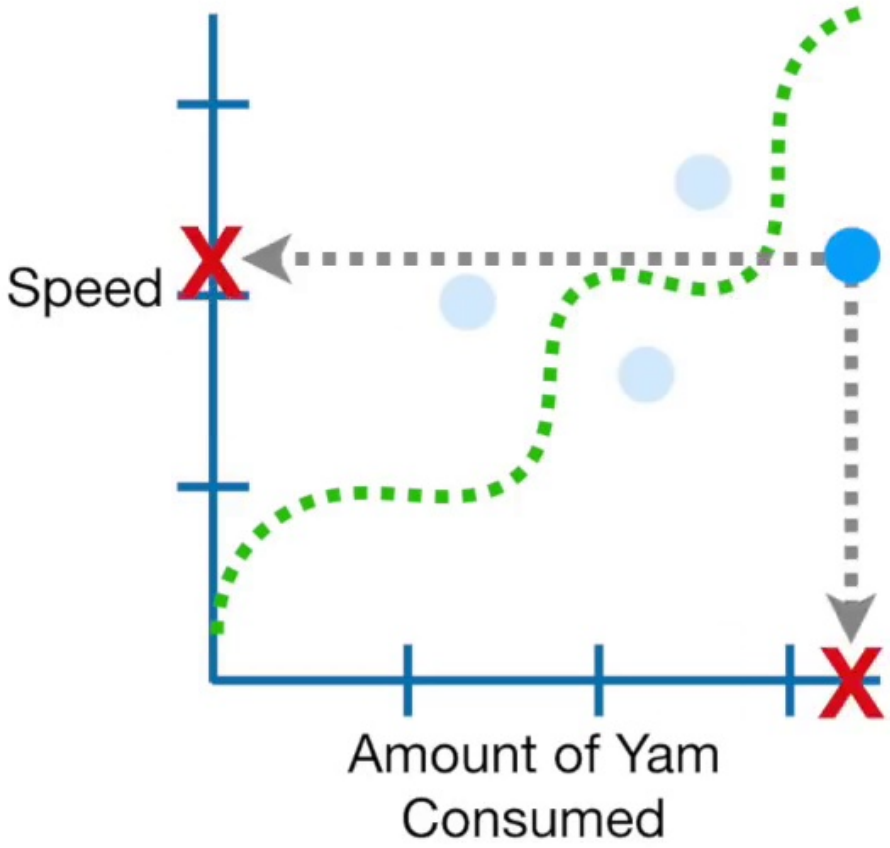
…and the third person…



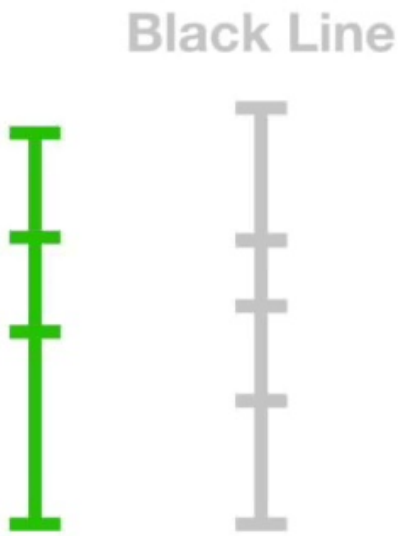


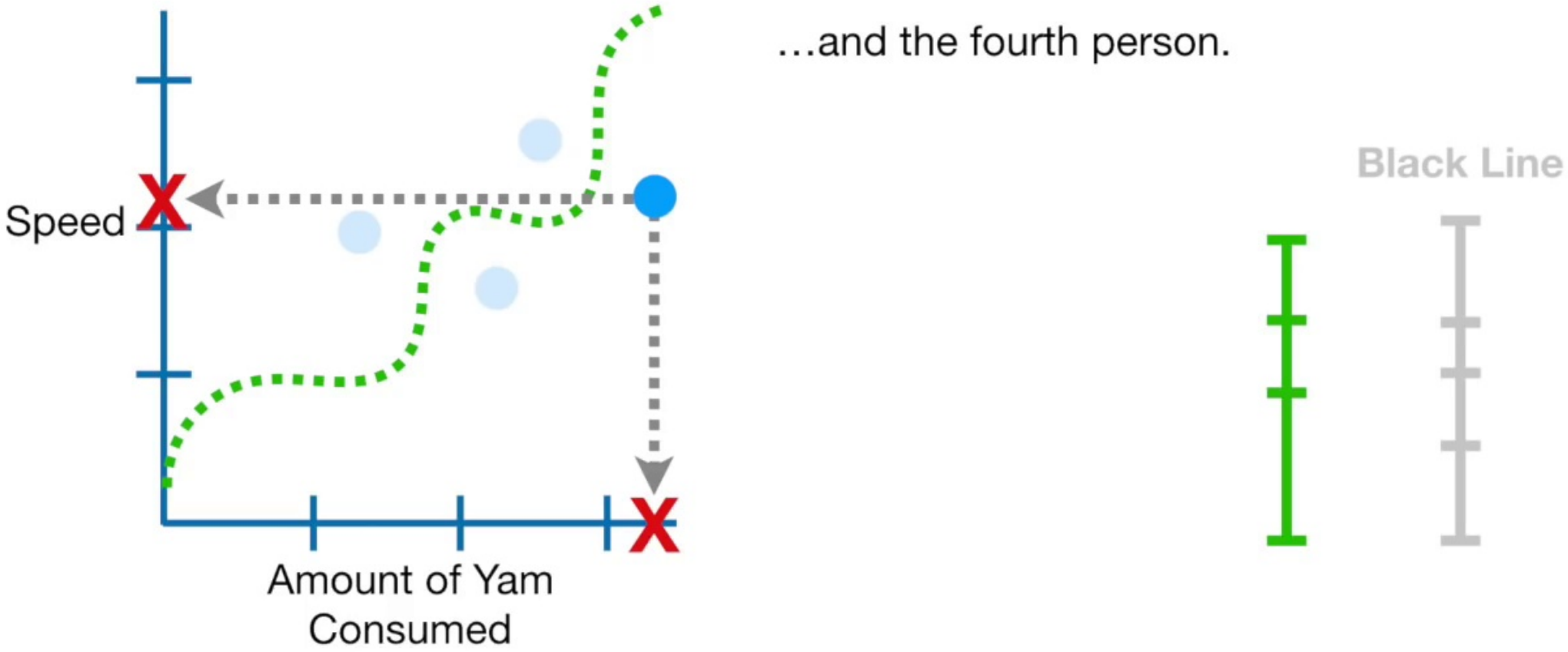
…and the third person…



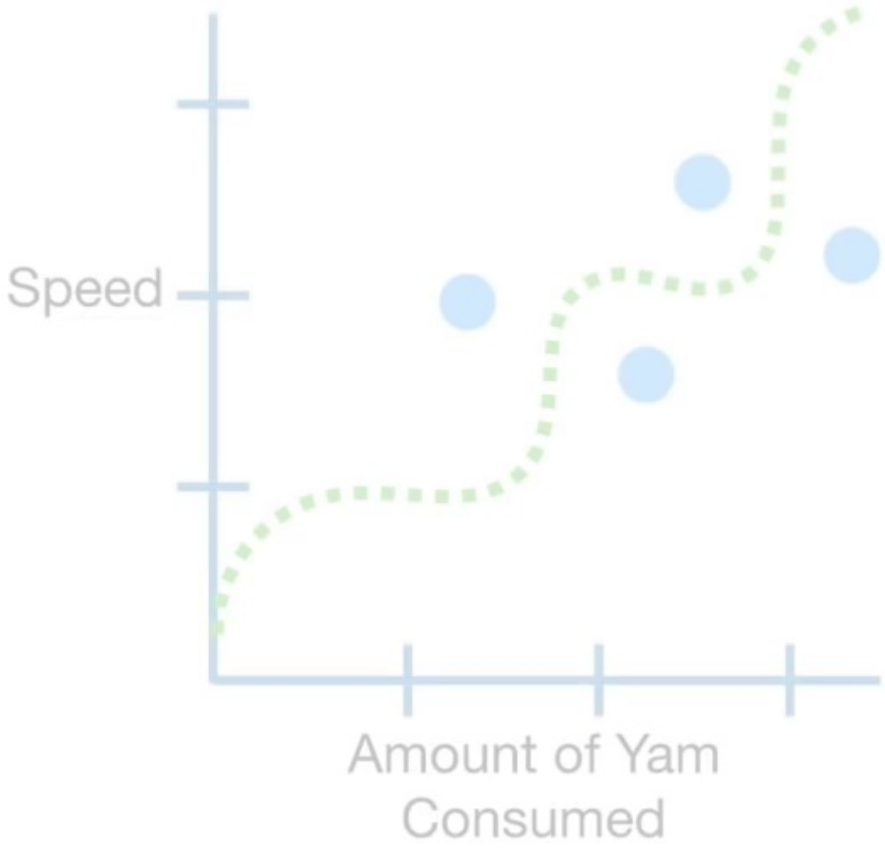


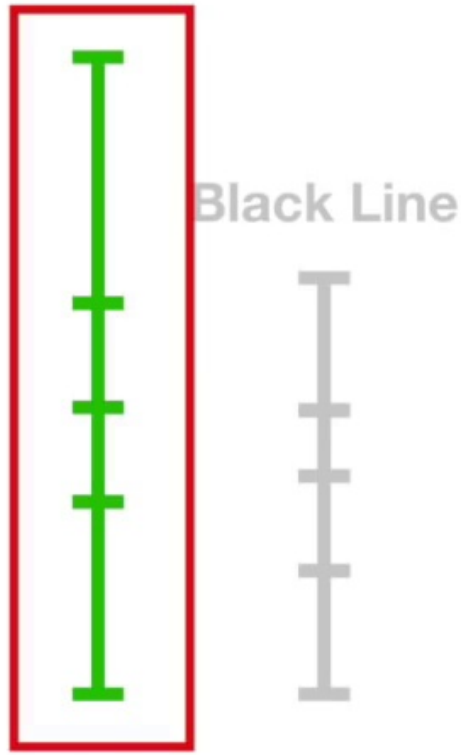
…and the fourth person.

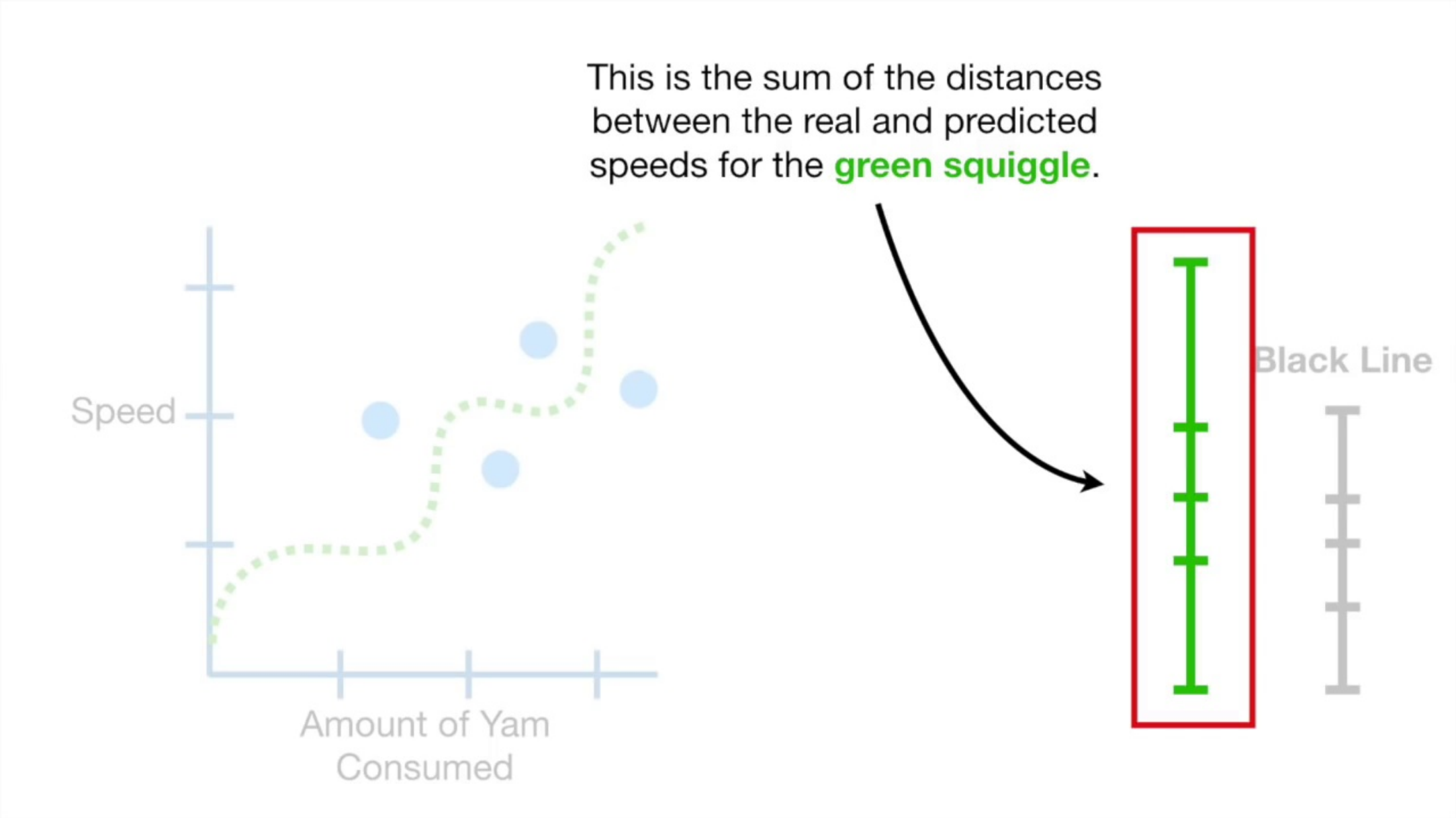



…and the fourth person.
Speed
Amount of Yam Consumed
Black Line

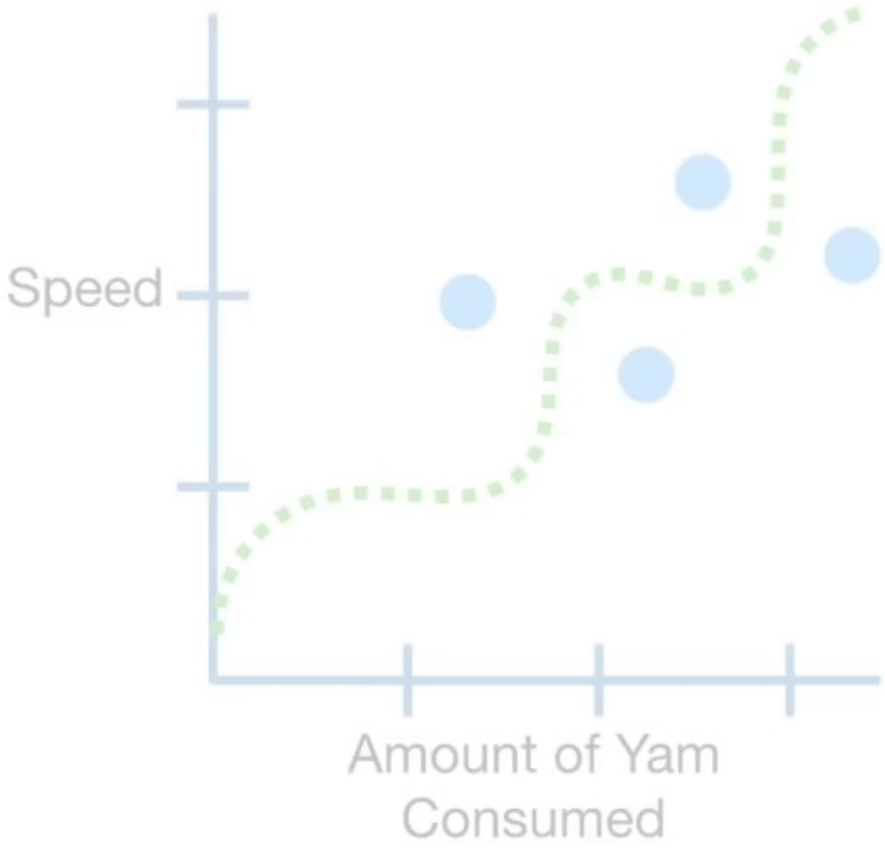
This is the sum of the distances between the real and predicted speeds for the **green squiggle**.

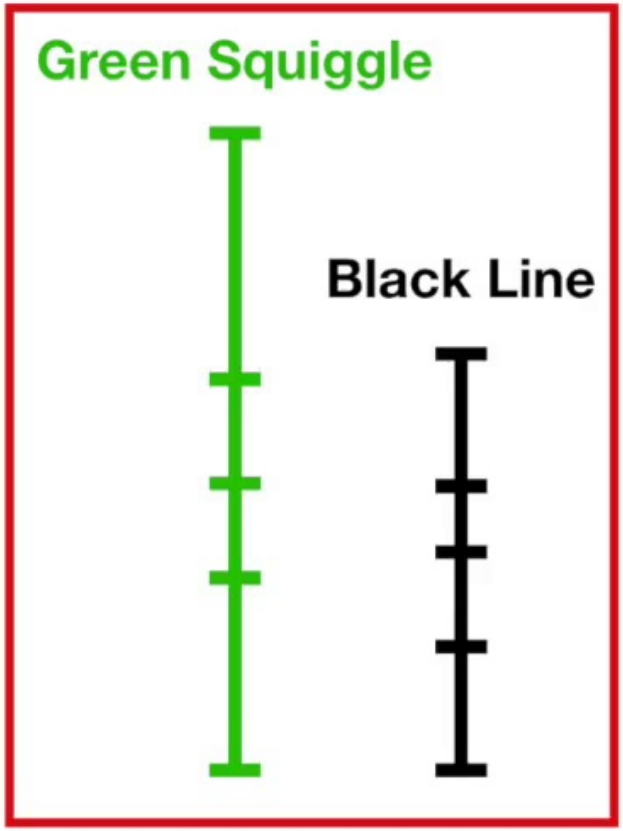




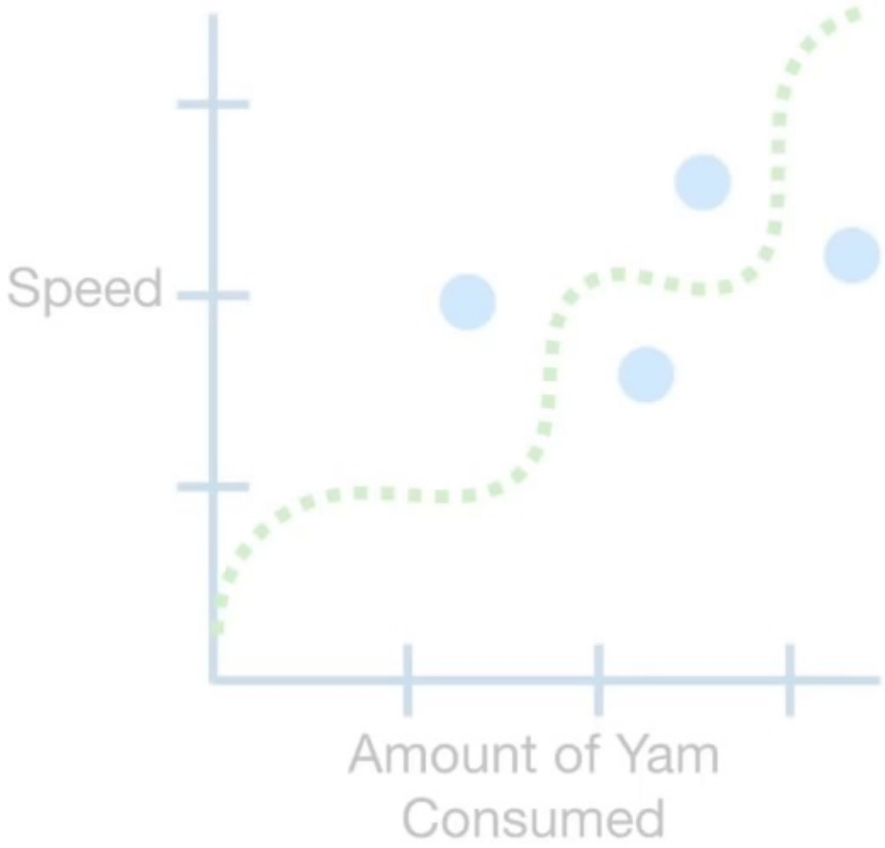

This is the sum of the distances between the real and predicted speeds for the green squiggle.
Speed
Amount of Yam Consumed
Black Line

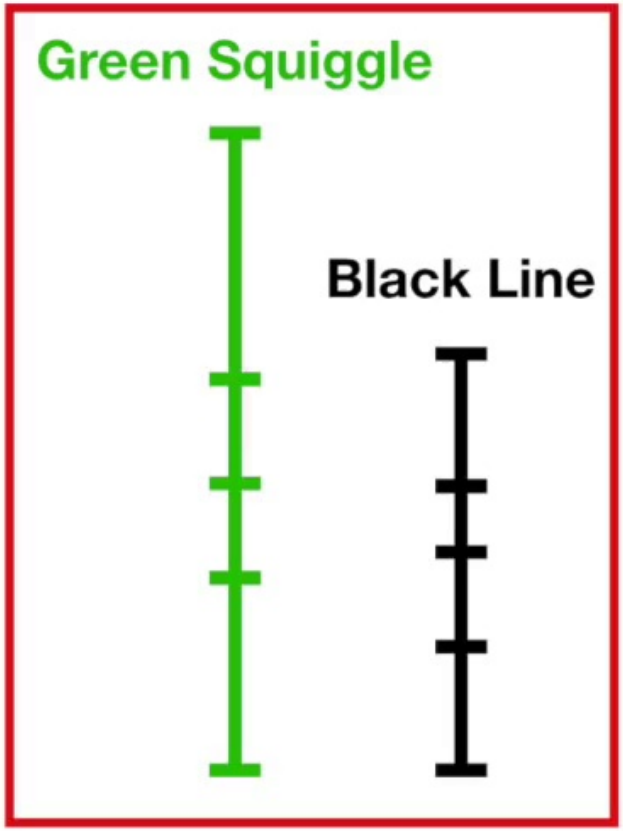
The sum of the distances is larger for the **green squiggle** then the **black line**…



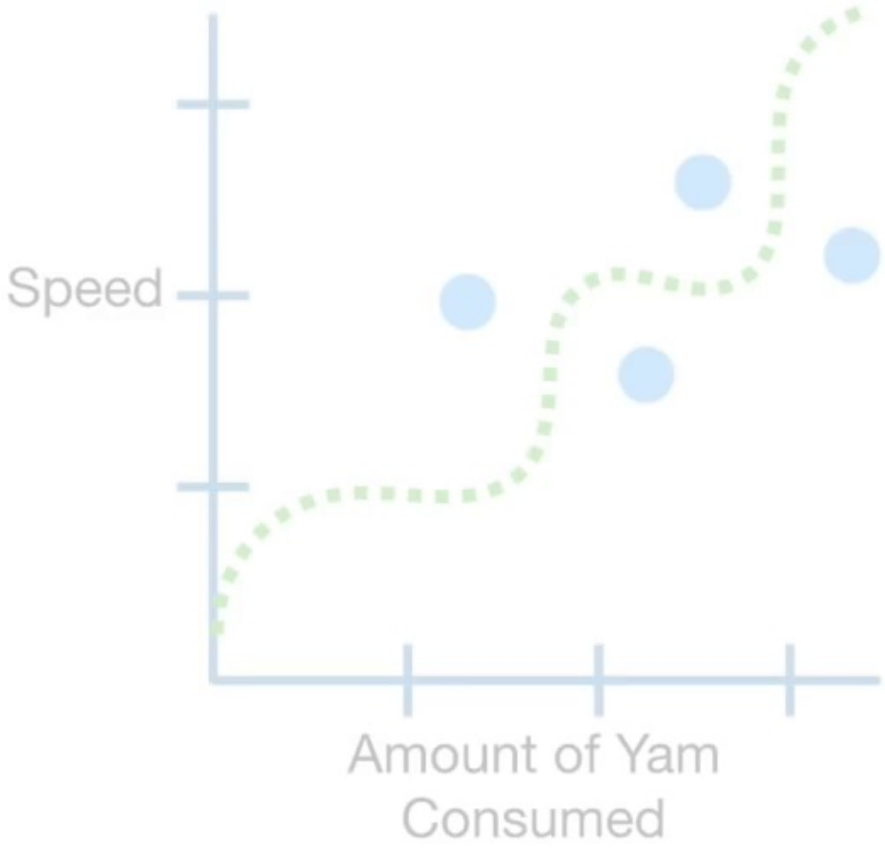


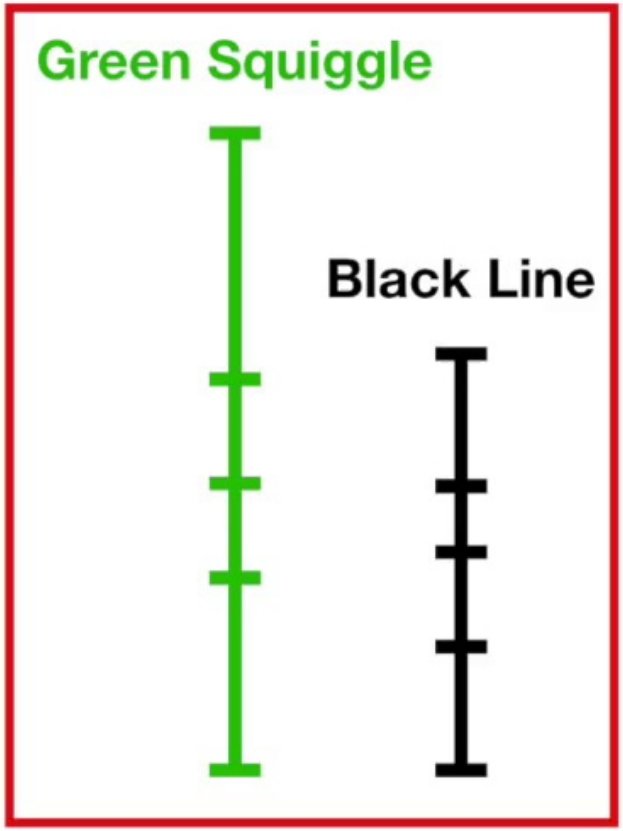
The sum of the distances is larger for the **green squiggle** then the **black line**…



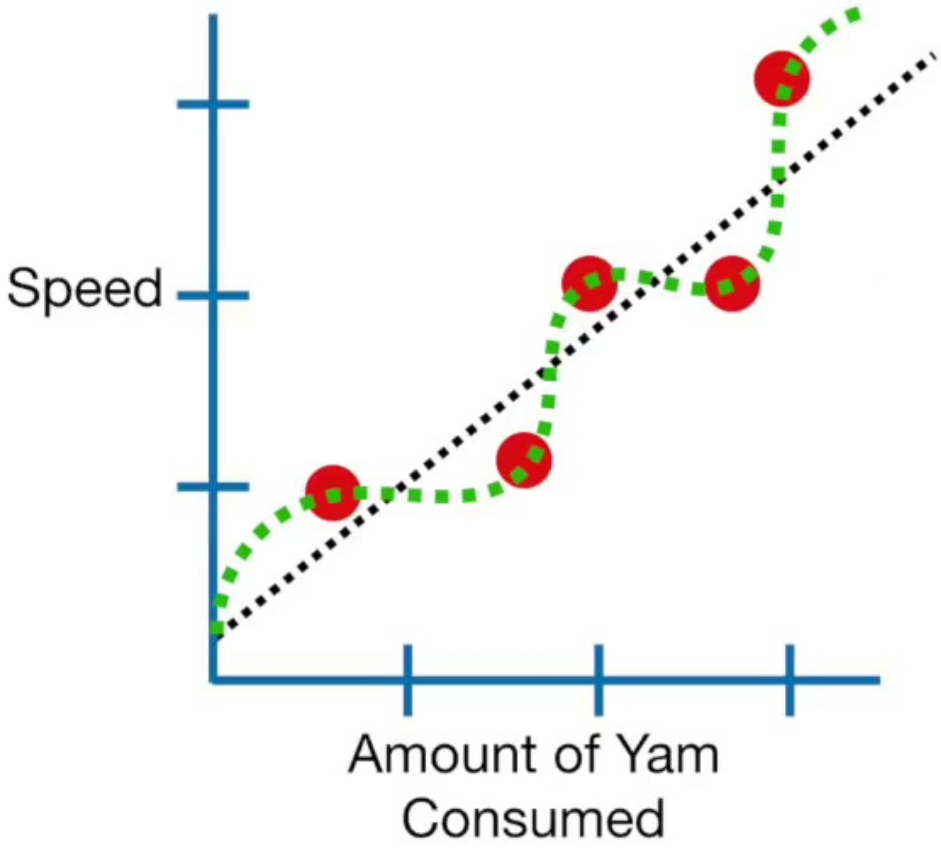


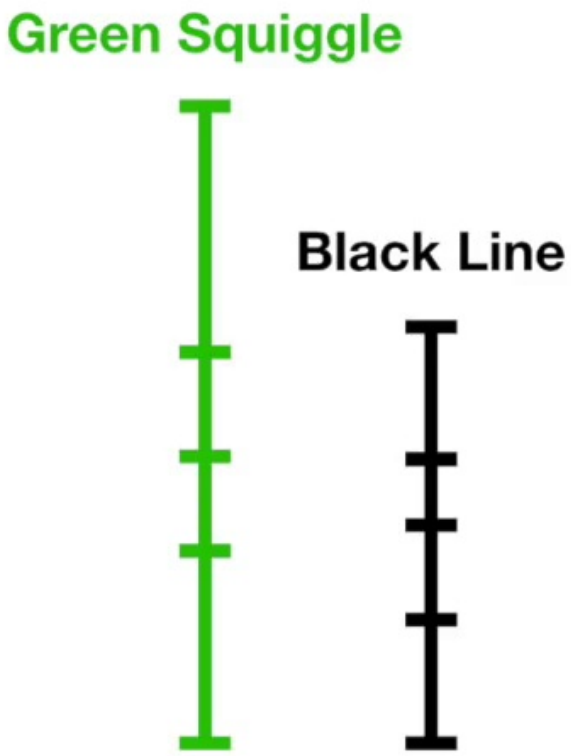
The sum of the distances is larger for the **green squiggle** then the **black line**…



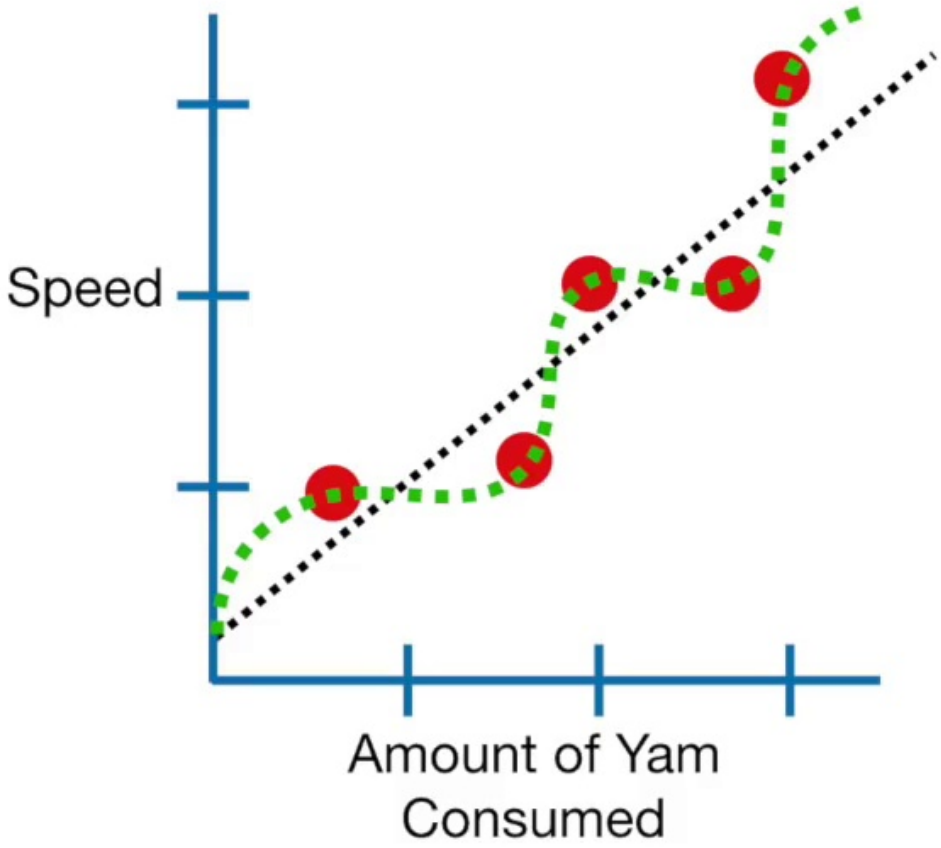


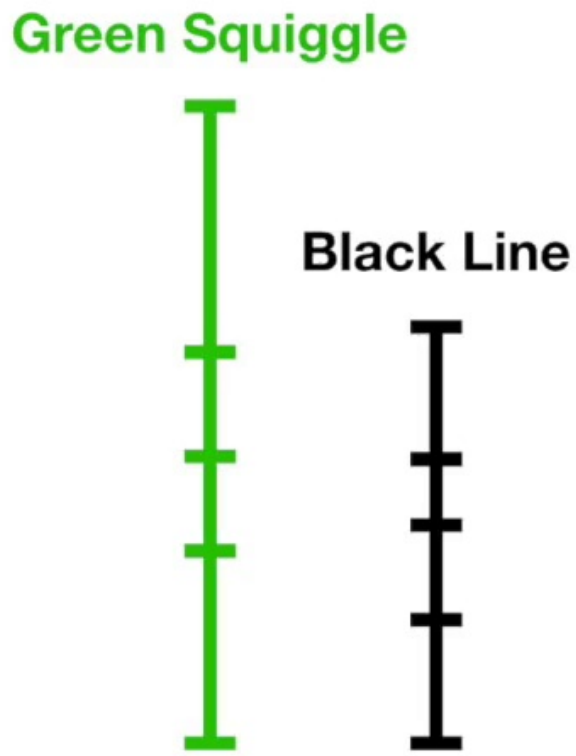
In other words, even though the **green squiggle** fit the **Training Data** way better than the **black line**…



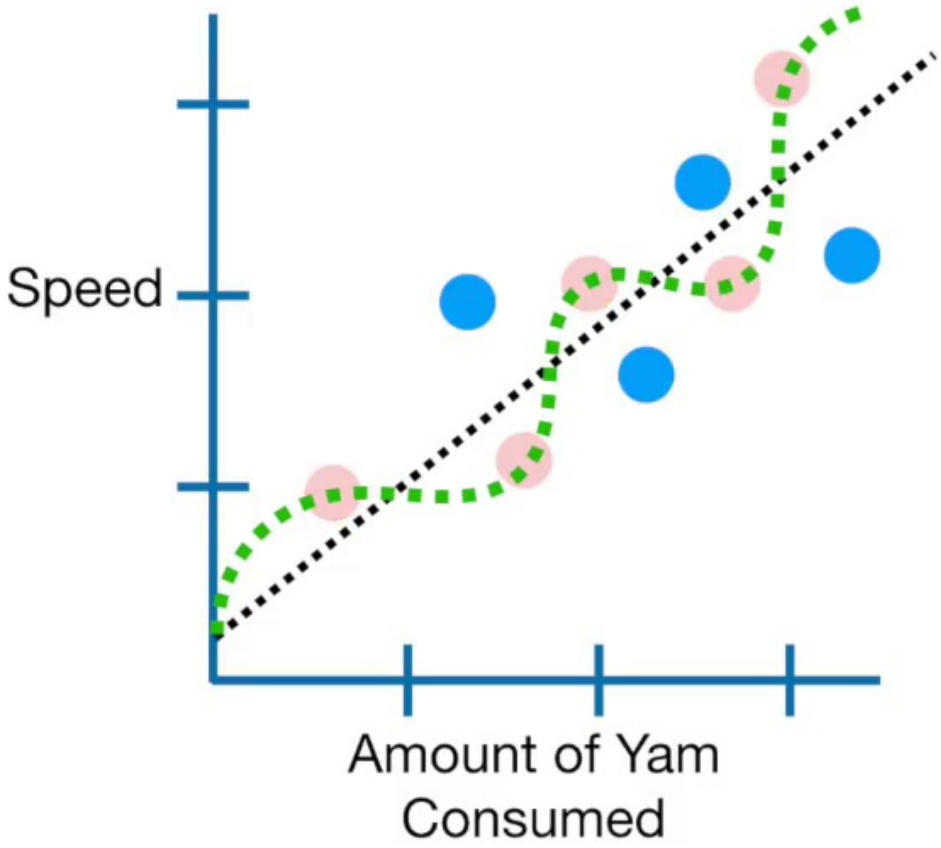


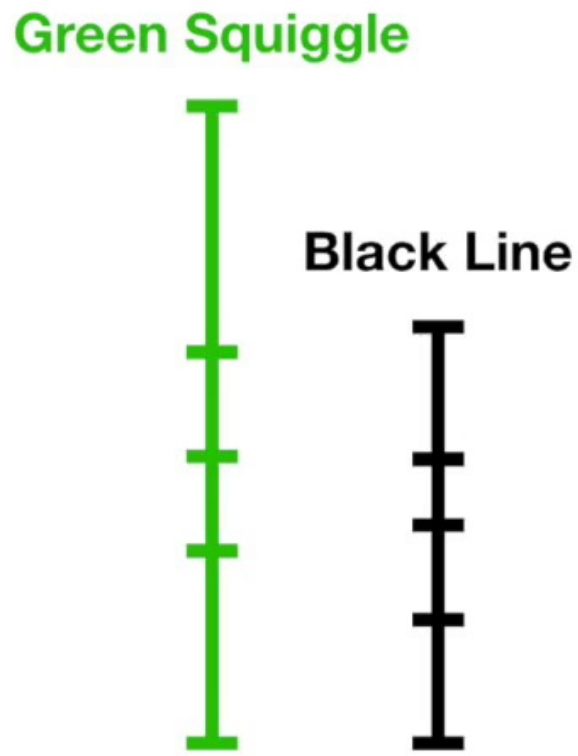
In other words, even though the **green squiggle** fit the **Training Data** way better than the **black line**…



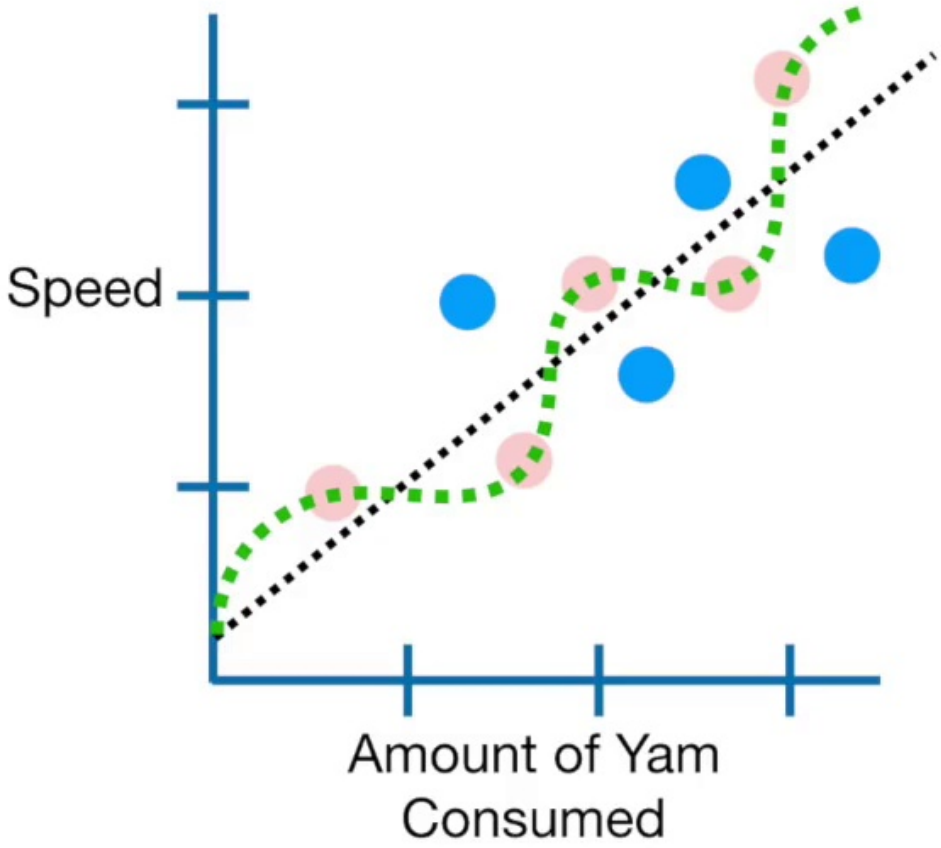


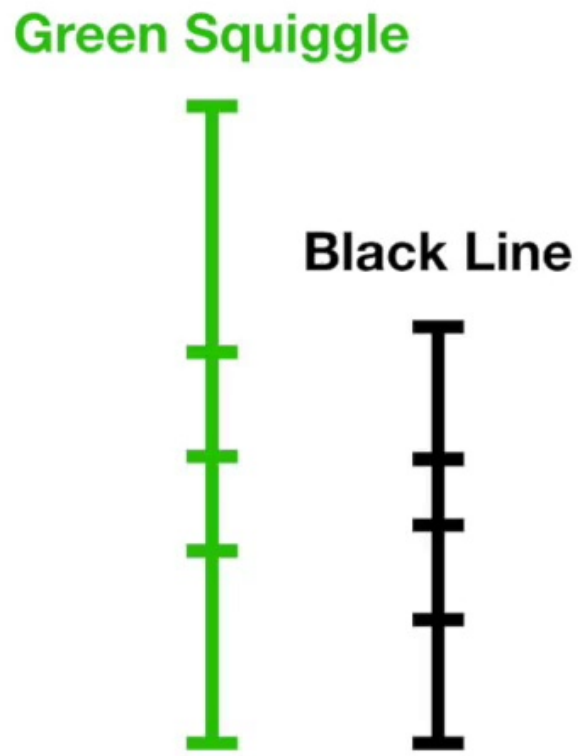
…the **black line** did a better job
predicting speeds with the **Testing Data**.



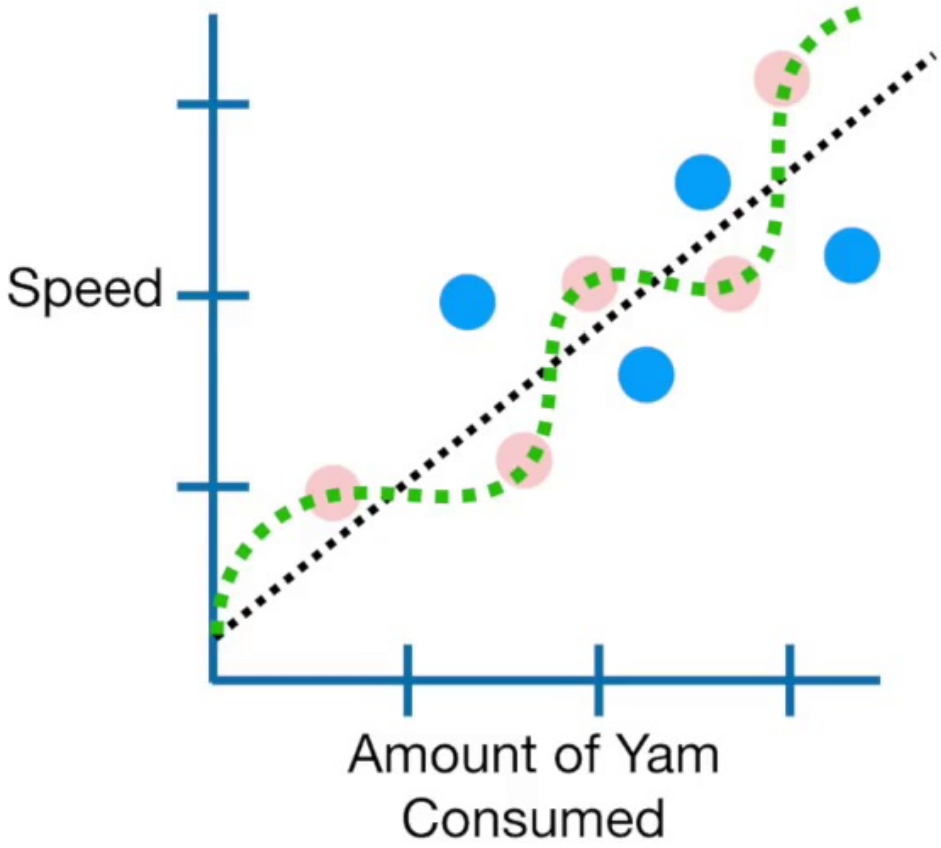


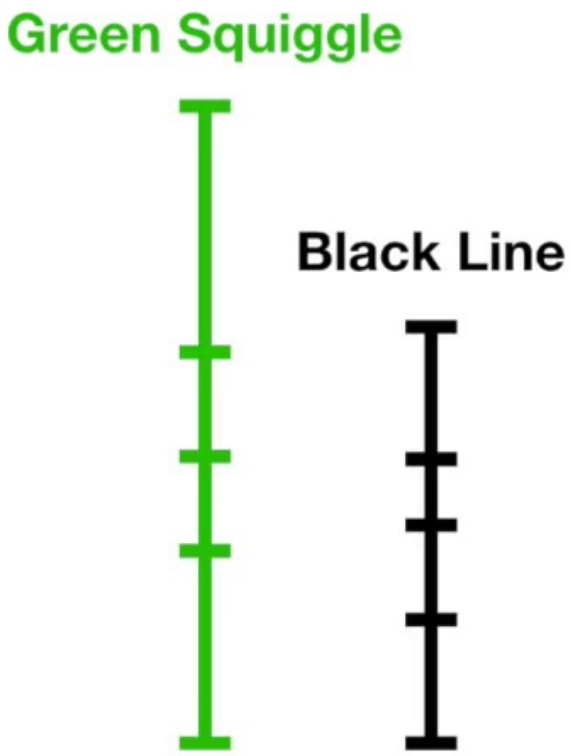
…the **black line** did a better job
predicting speeds with the **Testing Data**.



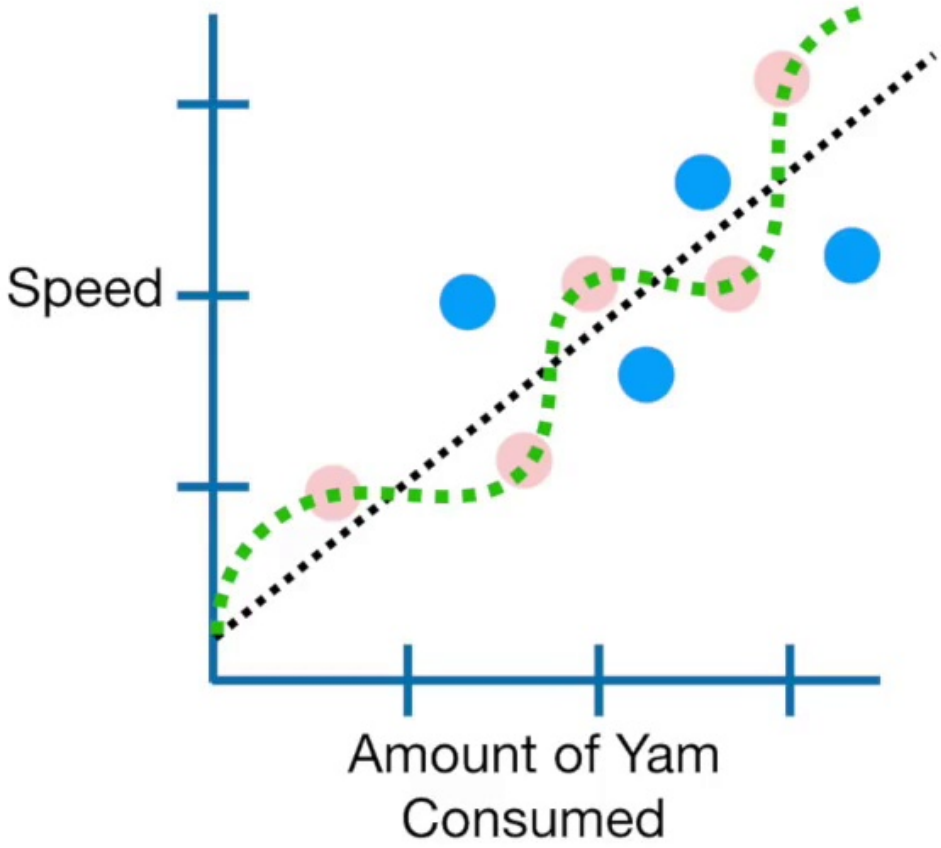


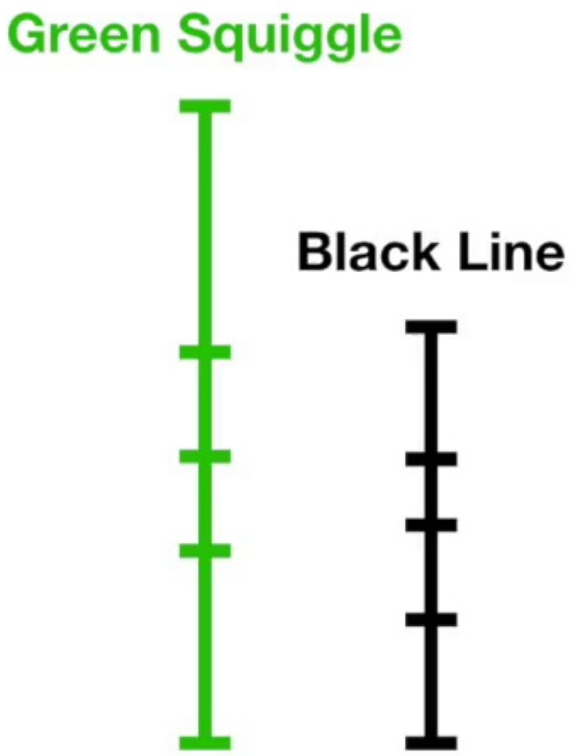
So if we had to choose between using the **black line** or the **green squiggle** to make predictions…

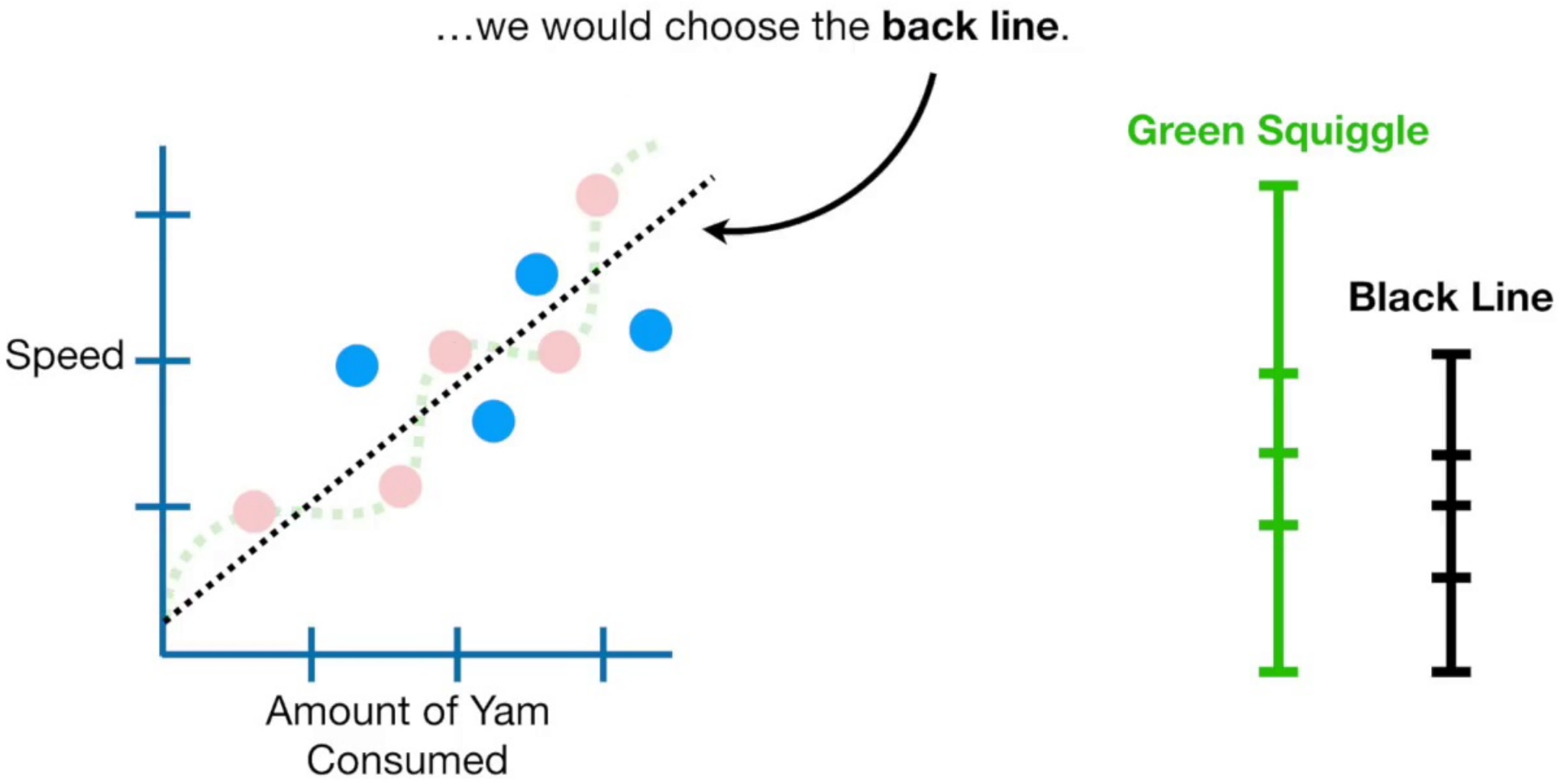




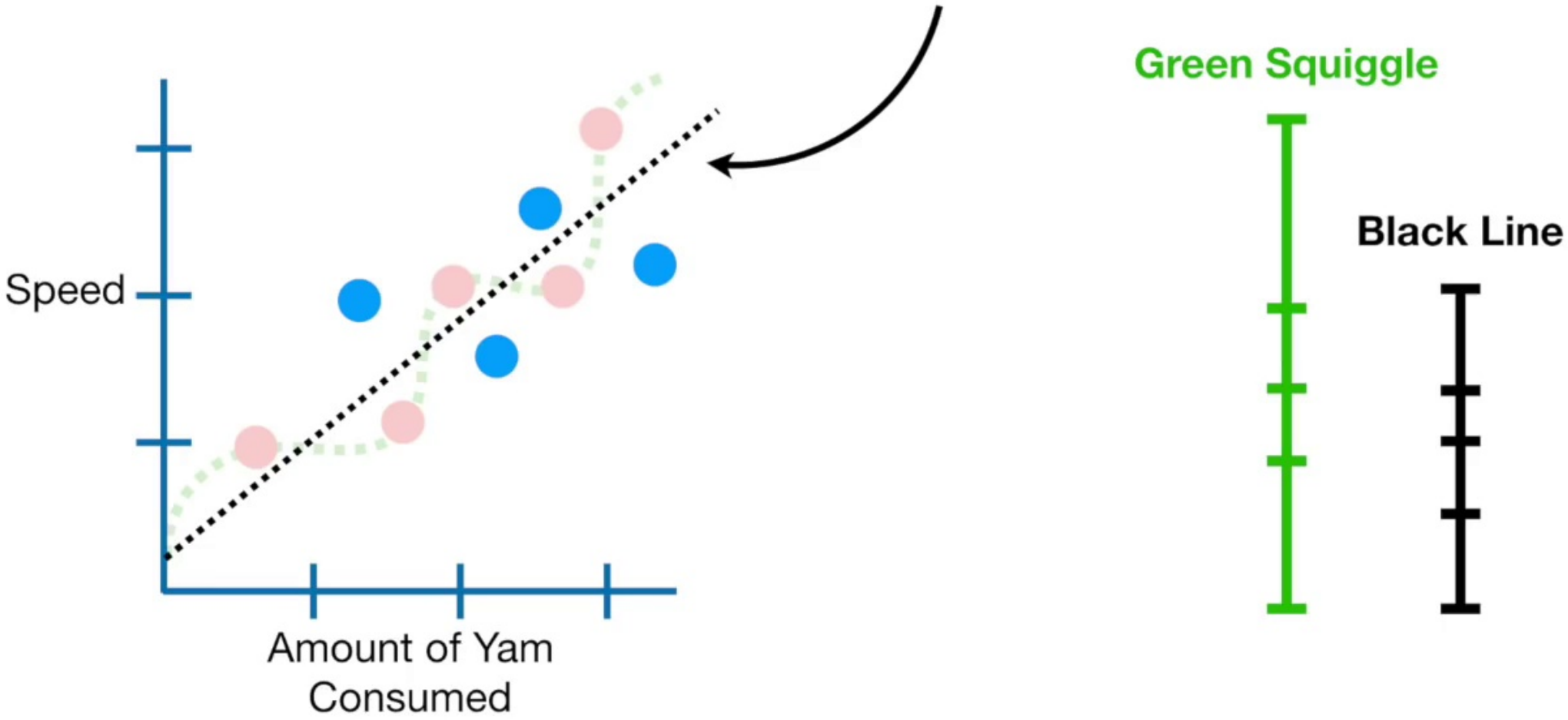
So if we had to choose between using the **black line** or the **green squiggle** to make predictions…



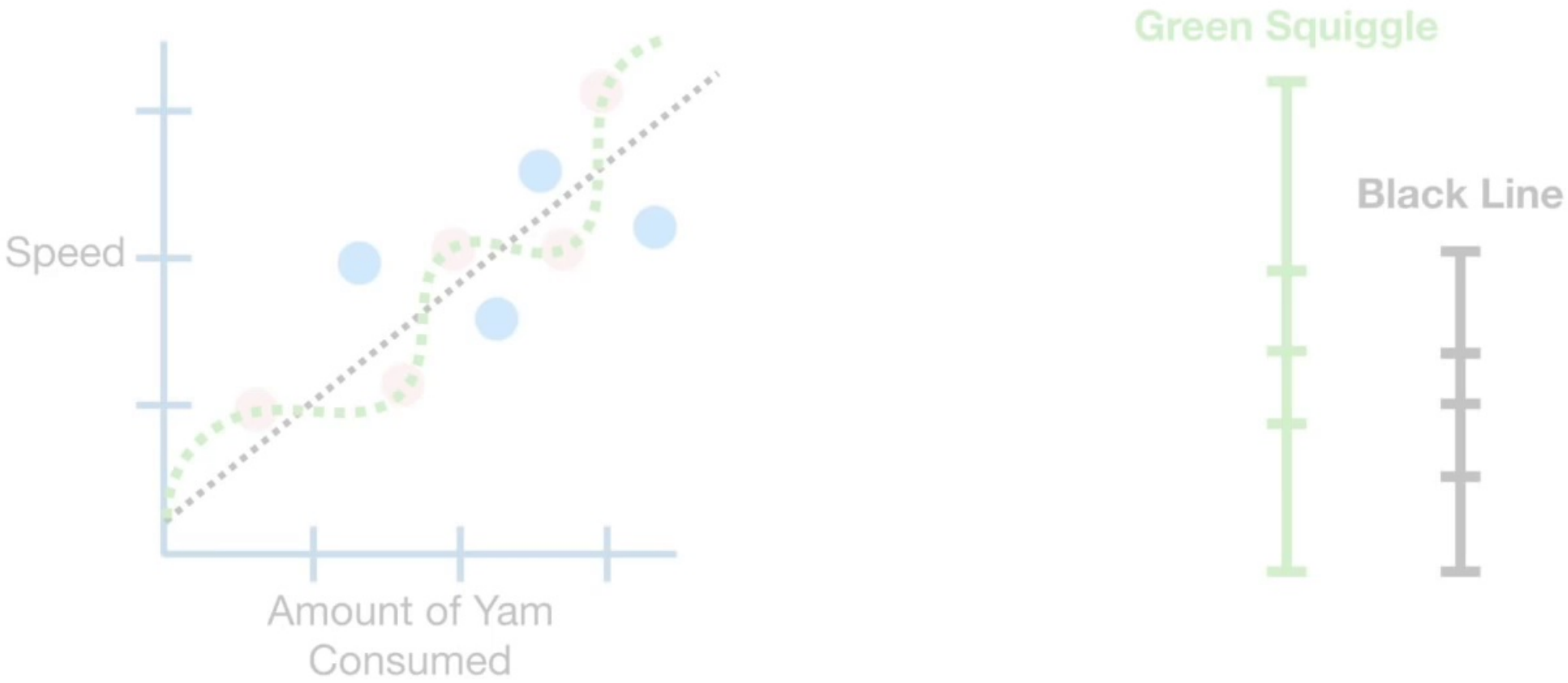



…we would choose the **back line**.
Speed
Amount of Yam Consumed
Green Squiggle
Black Line

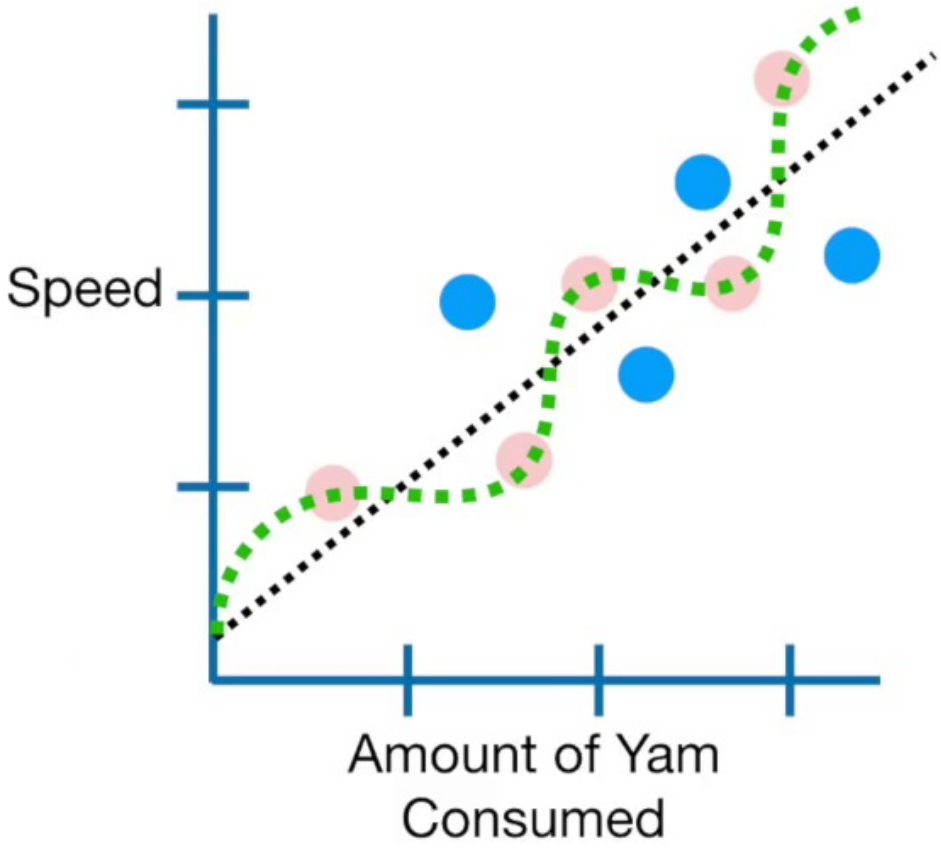
…we would choose the **back line**.

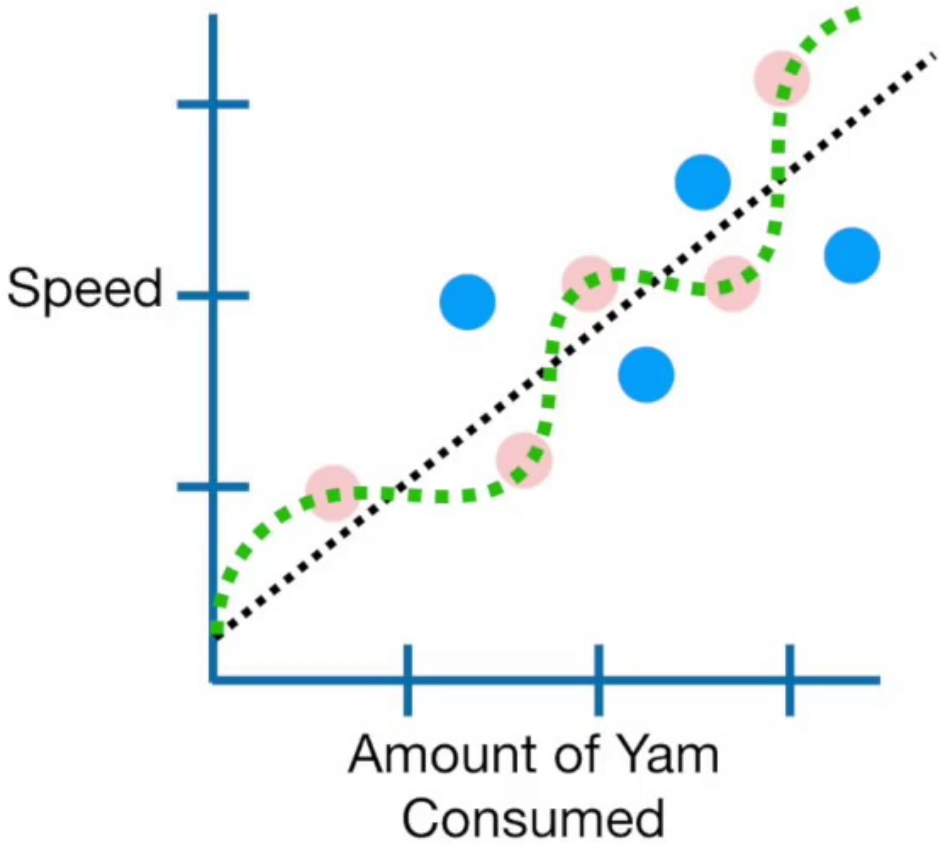


BAM!!!



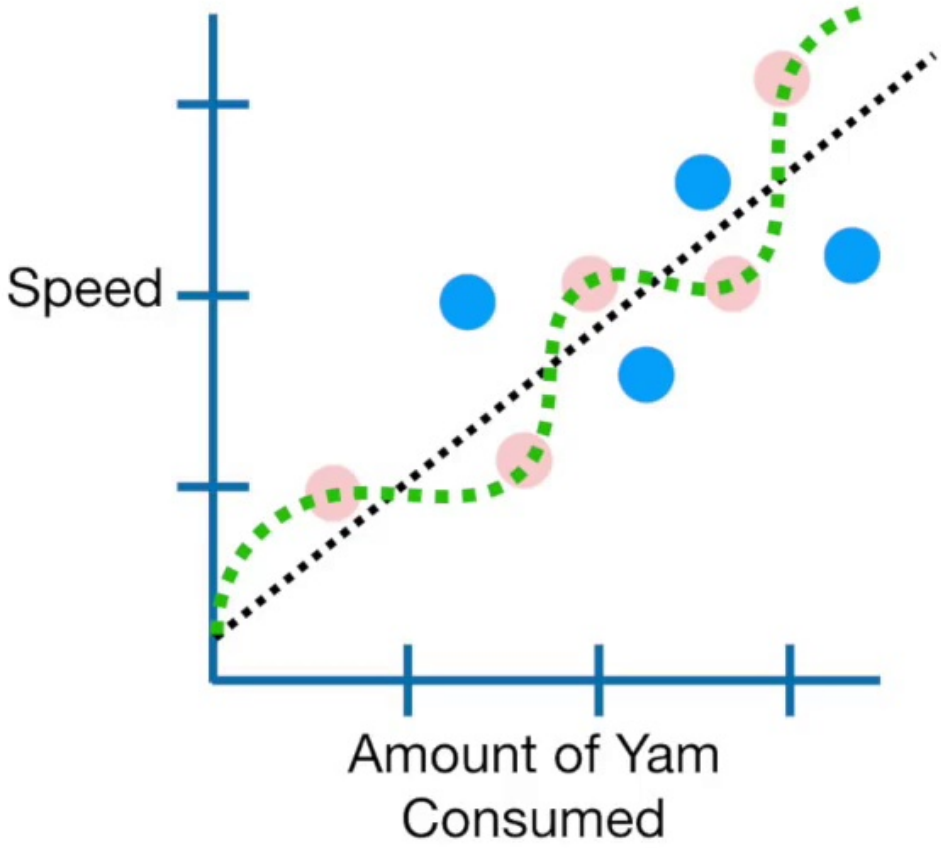
This example teaches two main ideas about **Machine Learning**.





This example teaches two main ideas about **Machine Learning**.

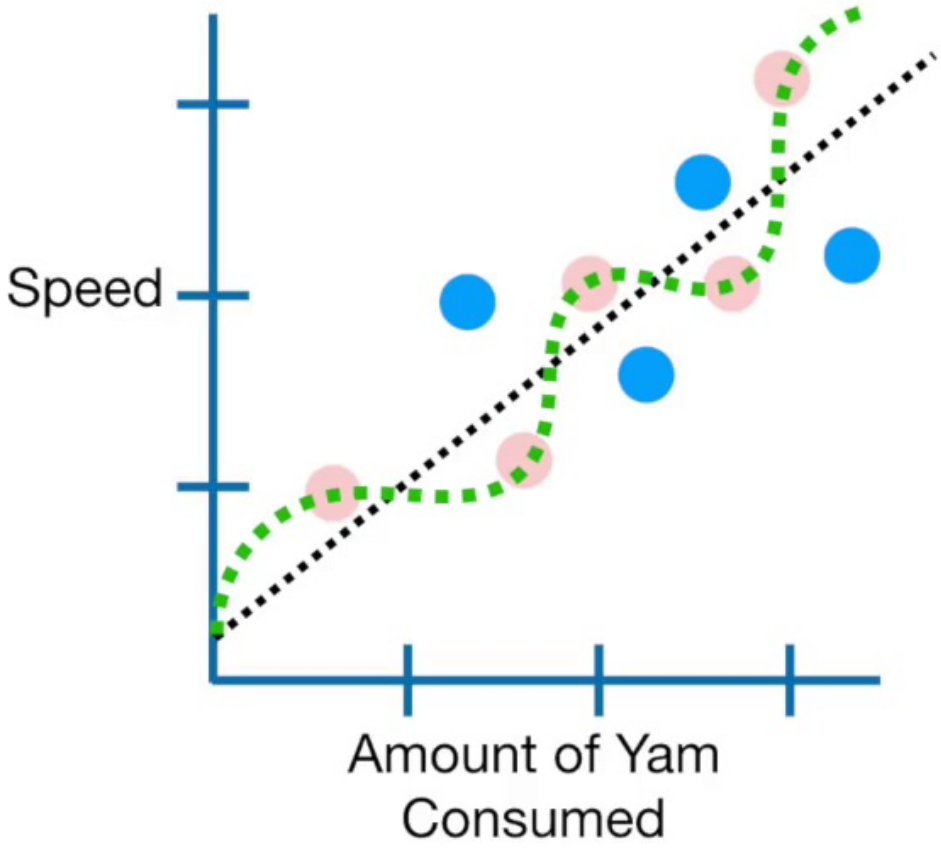
First, we use **Testing Data** to evaluate **Machine Learning** methods.



This example teaches two main ideas about **Machine Learning**.

First, we use **Testing Data** to evaluate **Machine Learning** methods.

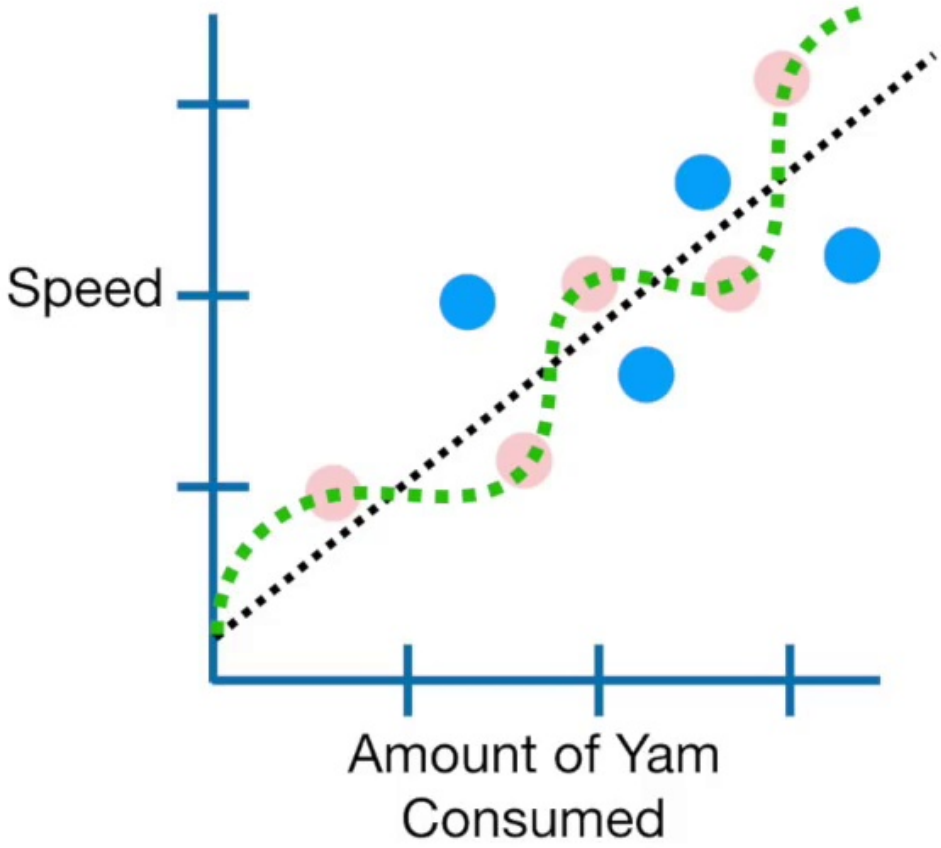
Second, don't be fooled by how well a **Machine Learning** method fits the **Training Data**.



This example teaches two main ideas about **Machine Learning**.

First, we use **Testing Data** to evaluate **Machine Learning** methods.

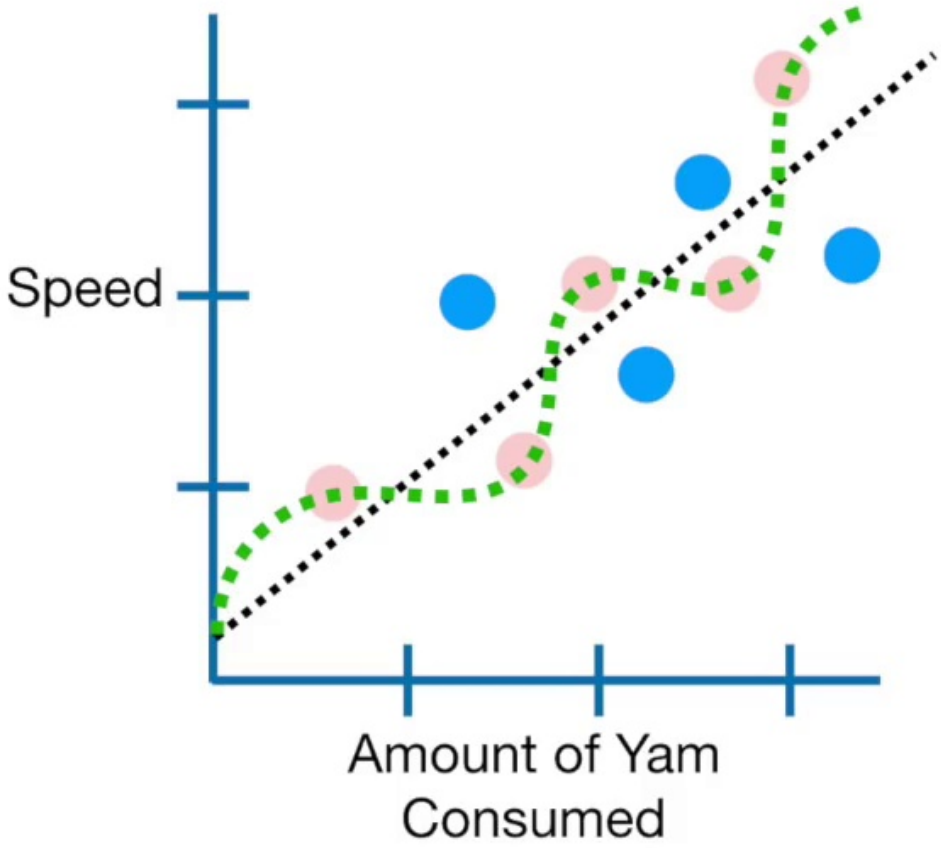
Second, don't be fooled by how well a **Machine Learning** method fits the **Training Data**.



This example teaches two main ideas about **Machine Learning**.

First, we use **Testing Data** to evaluate **Machine Learning** methods.

Second, don't be fooled by how well a **Machine Learning** method fits the **Training Data**.

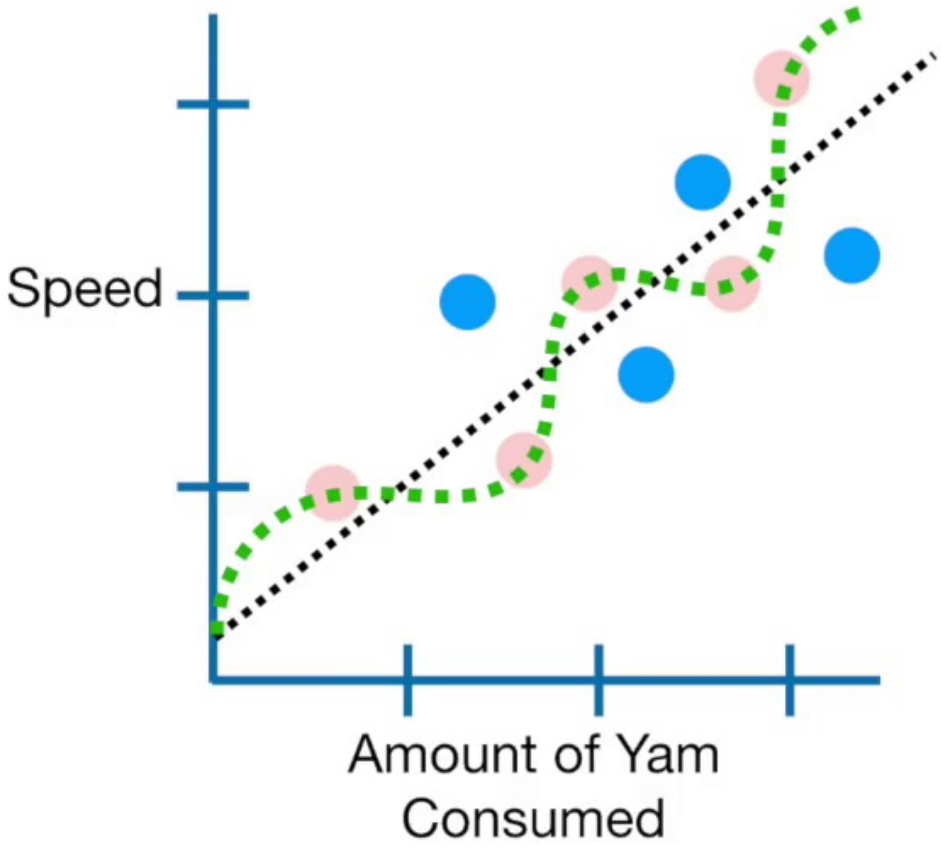


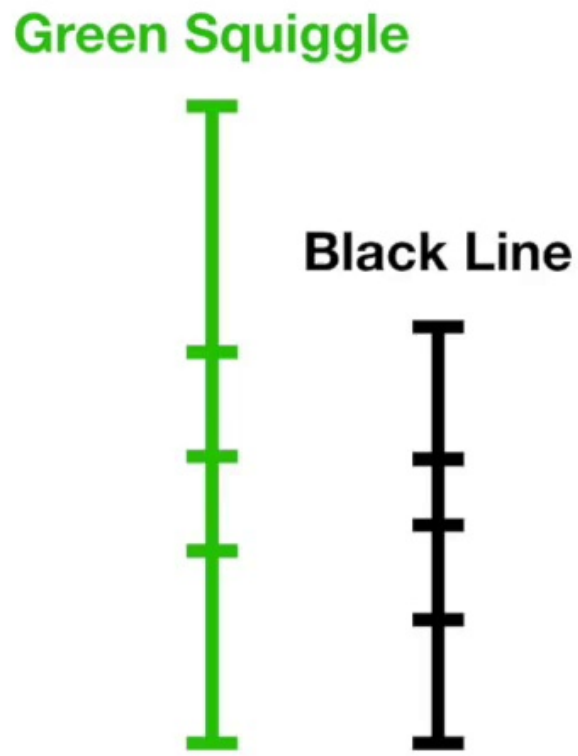
This example teaches two main ideas about **Machine Learning**.

First, we use **Testing Data** to evaluate **Machine Learning** methods.

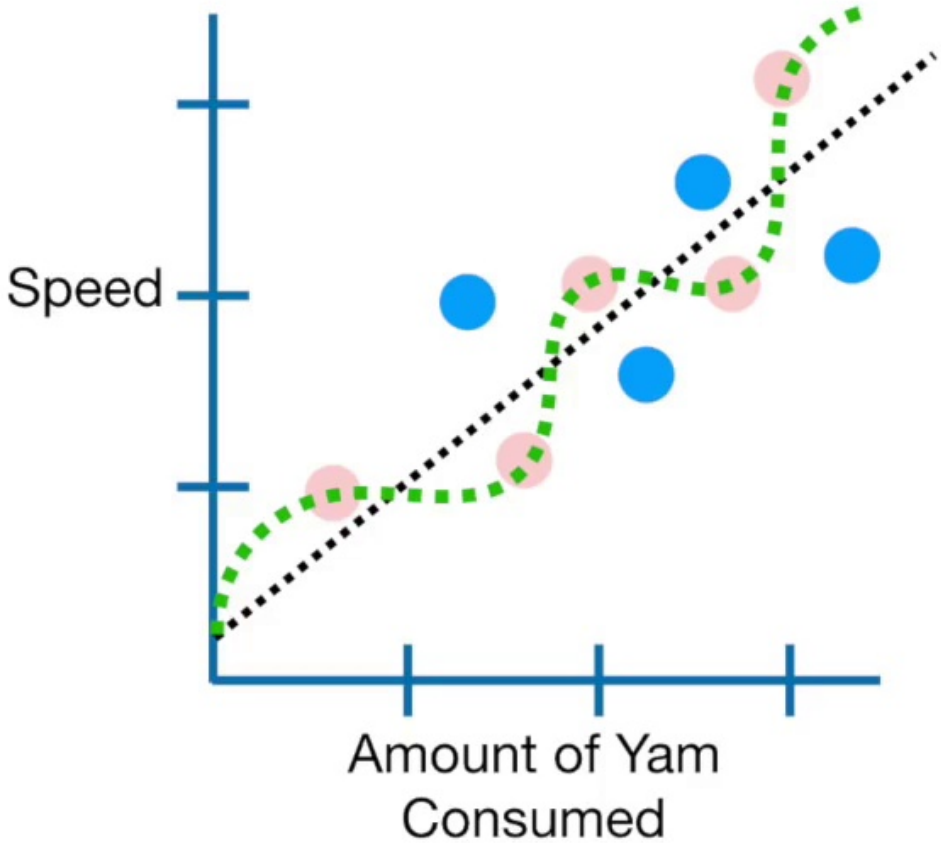
Second, don't be fooled by how well a **Machine Learning** method fits the **Training Data**.

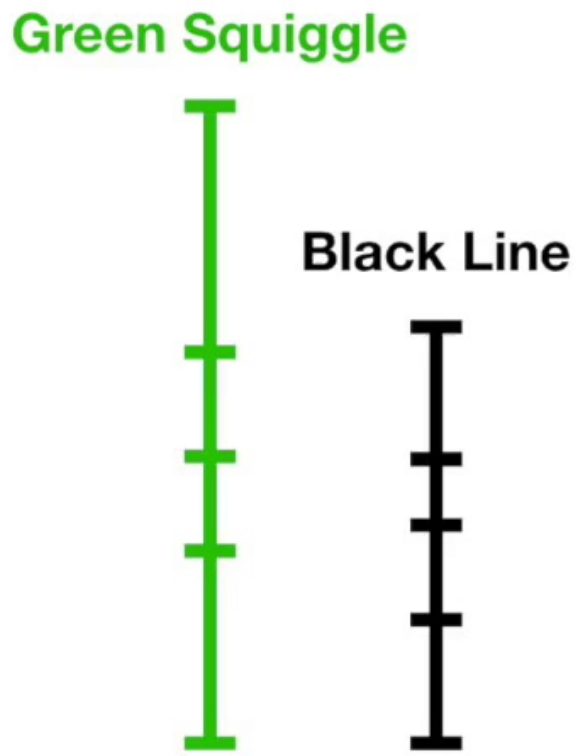
NOTE: Fitting the **Training Data** well but making poor *predictions*, is called the **Bias-Variance Tradeoff**.





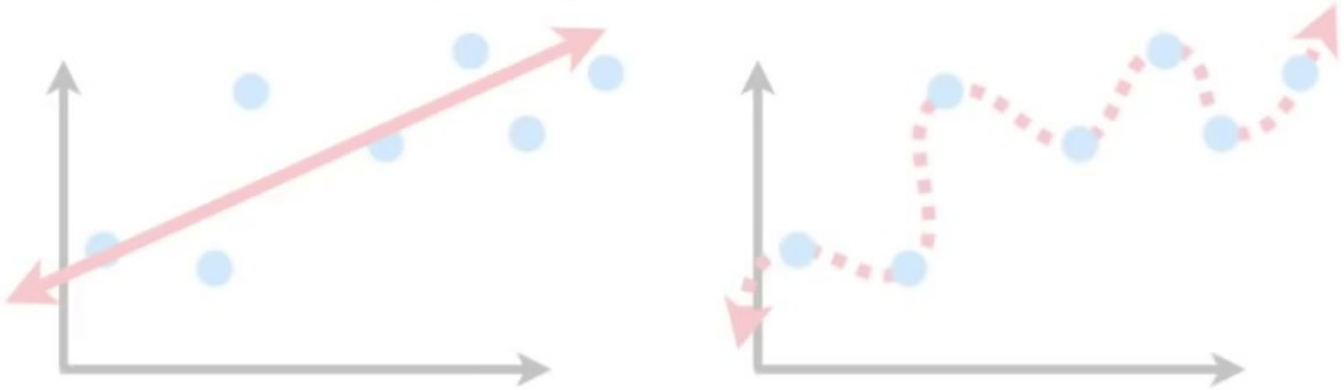
NOTE: Fitting the **Training Data** well but making poor *predictions*, is called the **Bias-Variance Tradeoff**.





SHAMELESS SELF PROMOTION

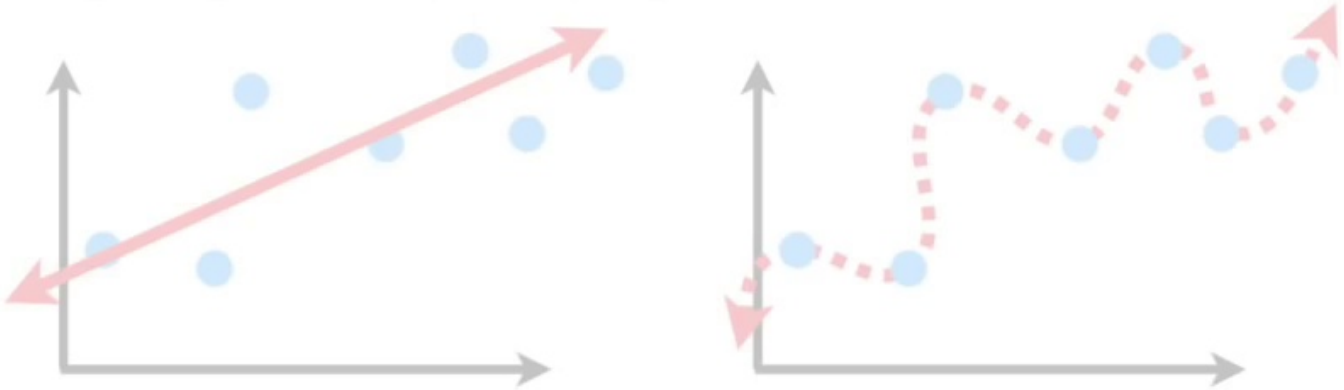
Machine Learning Fundamentals…



…Bias and Variance!!!

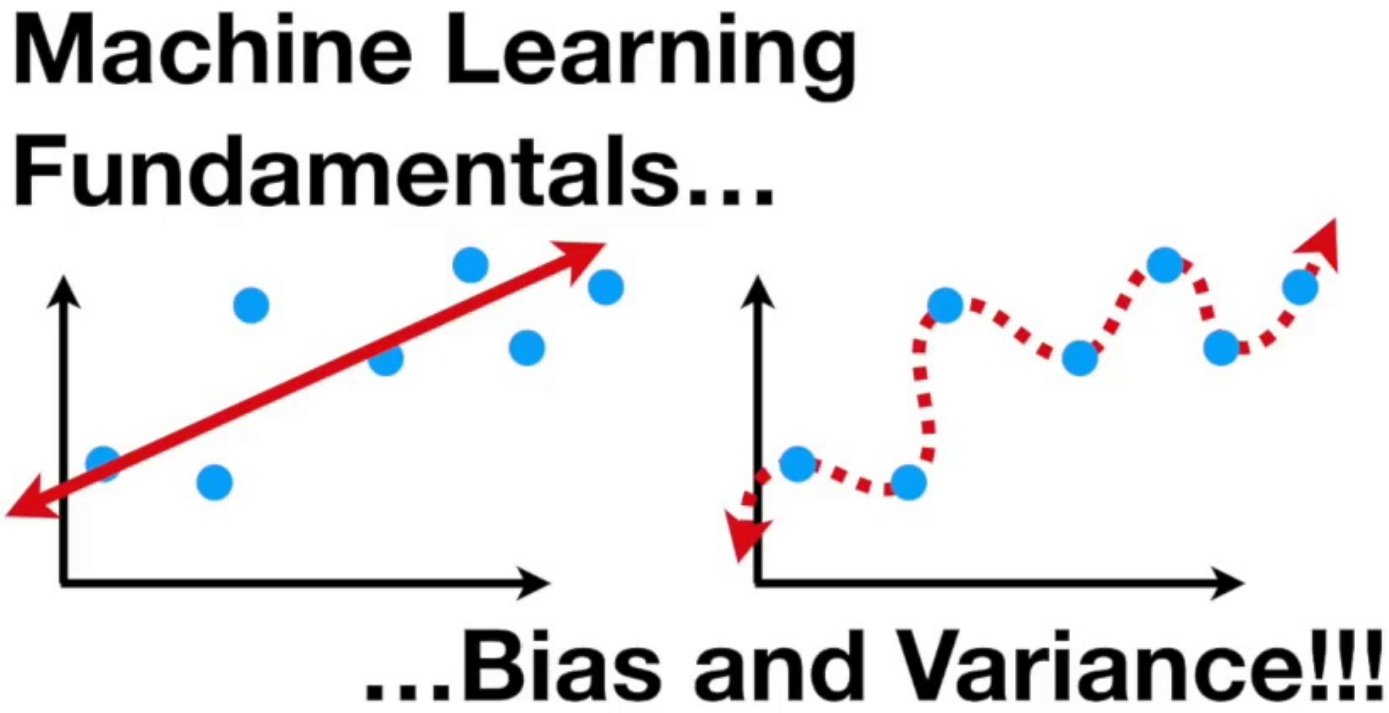
SHAMELESS SELF PROMOTION

Machine Learning Fundamentals…

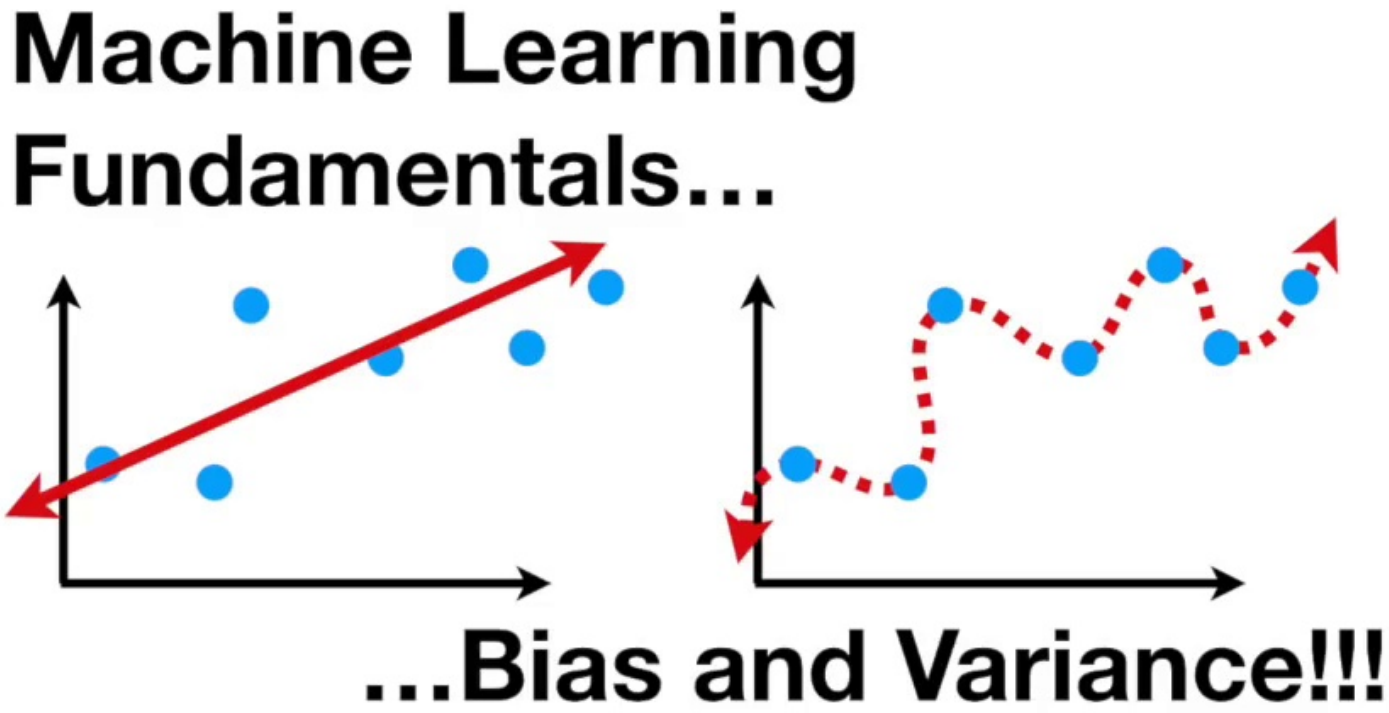


…Bias and Variance!!!

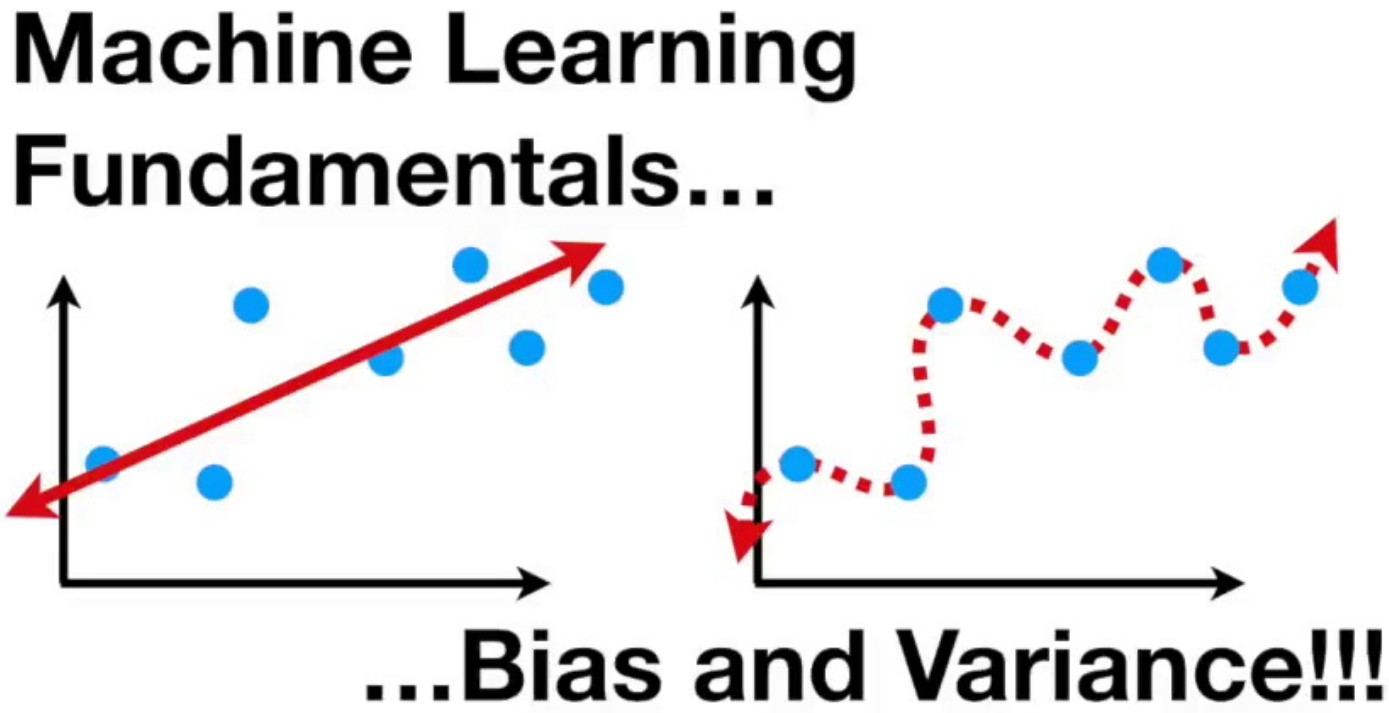
If you want to learn more about the **Bias-Variance Tradeoff**, there's a **StatQuest** that will walk you through it one step at a time.

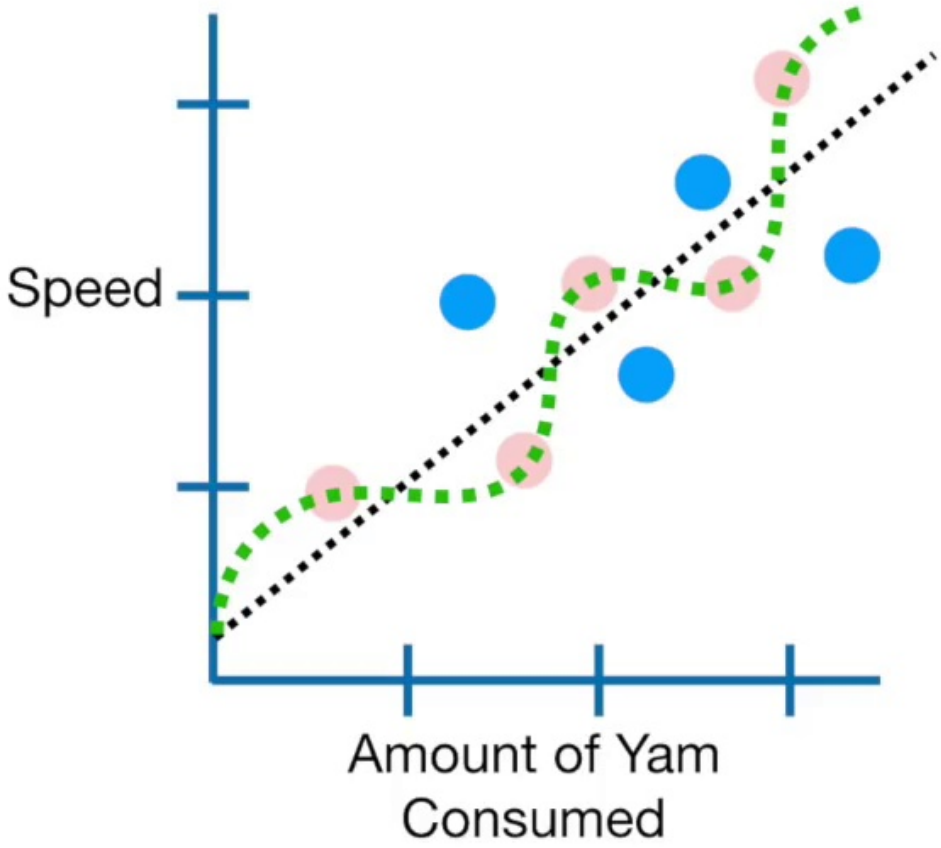


If you want to learn more about the **Bias-Variance Tradeoff**, there's a **StatQuest** that will walk you through it one step at a time.

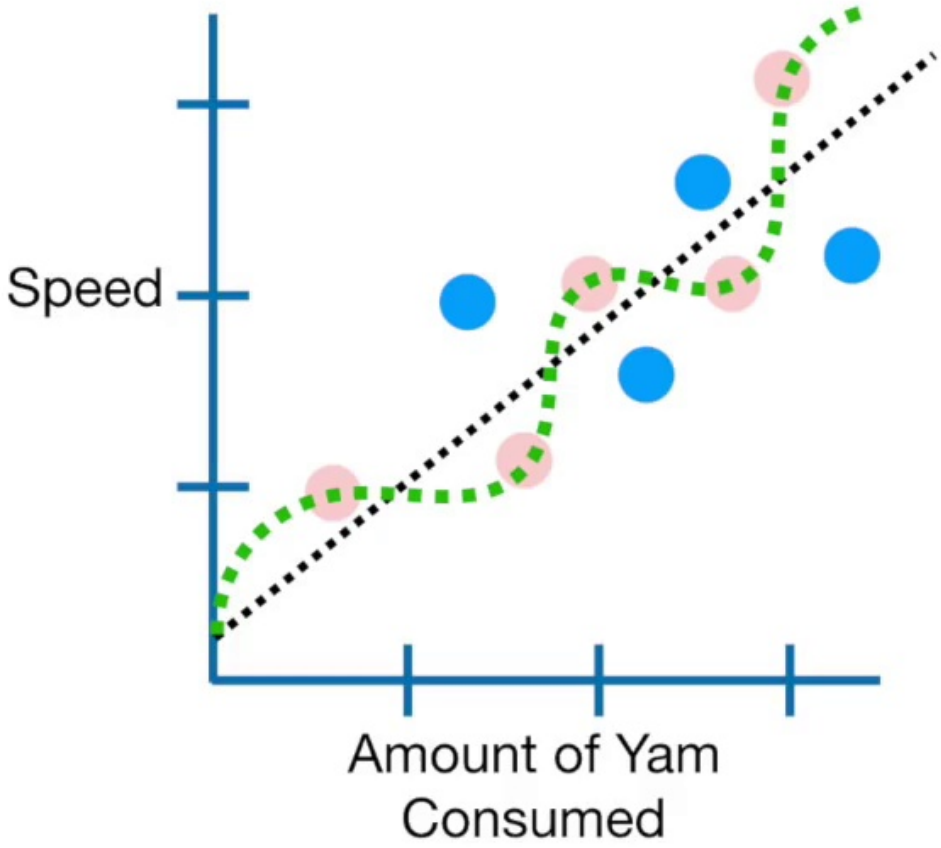


If you want to learn more about the **Bias-Variance Tradeoff**, there's a **StatQuest** that will walk you through it one step at a time.

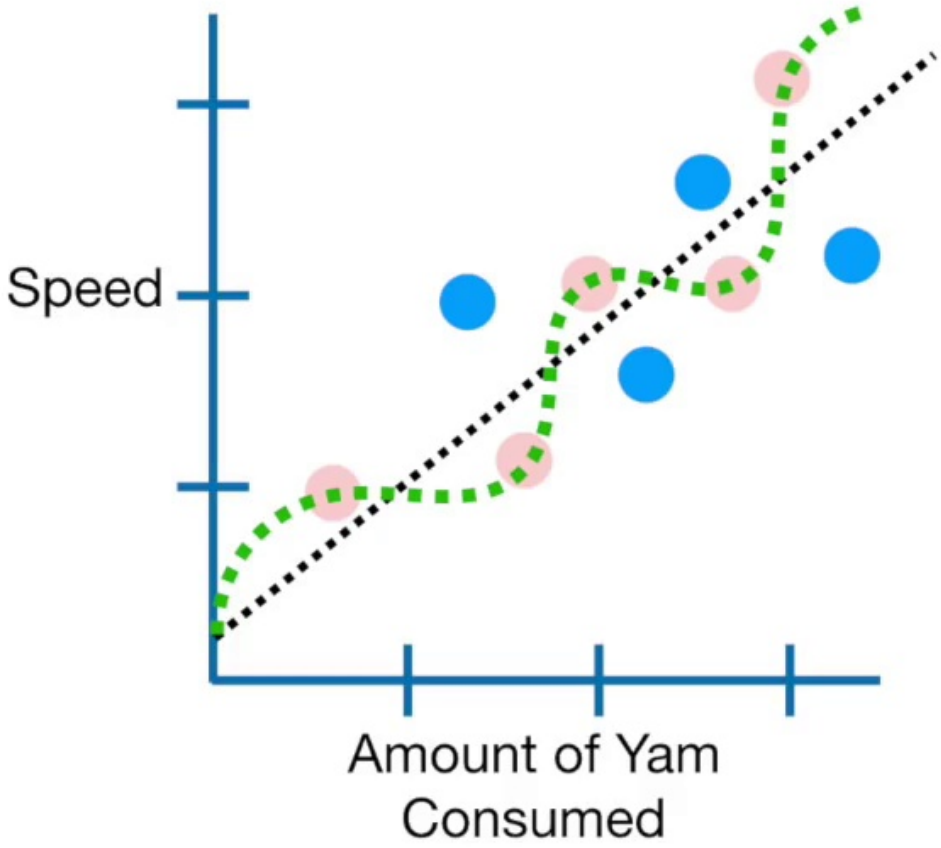




Before we move on, you may be wondering why we used a simple **black line** and a silly **green squiggle** instead of a…



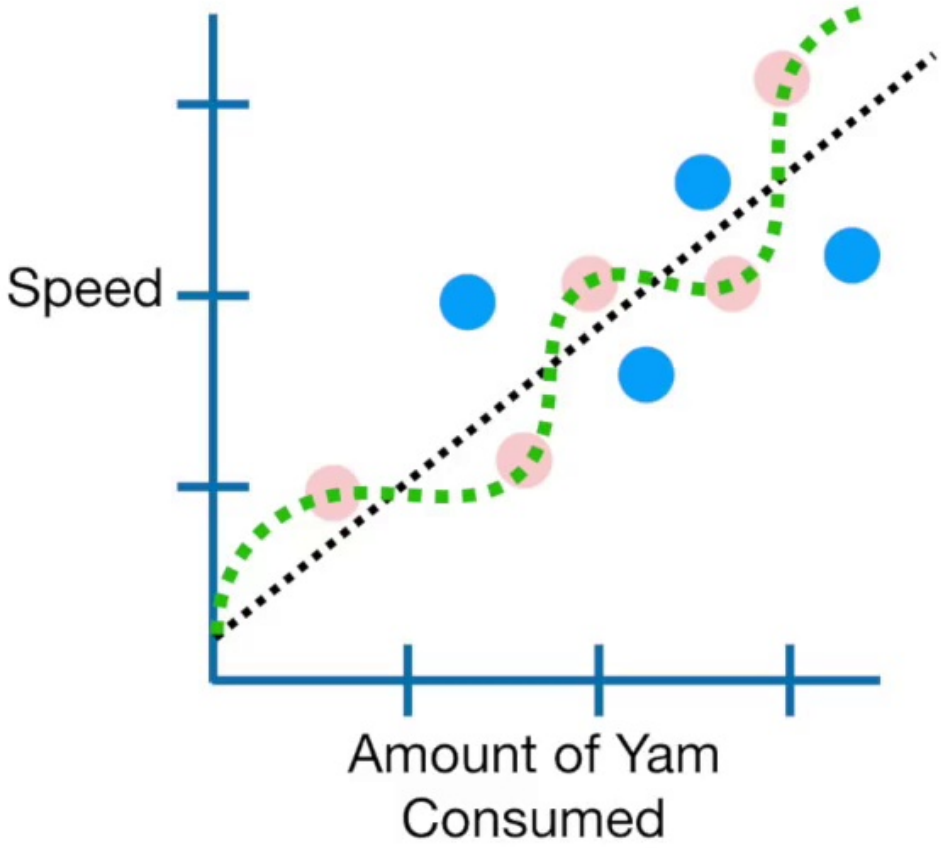
Before we move on, you may be wondering why we used a simple **black line** and a silly **green squiggle** instead of a…



Before we move on, you may be wondering why we used a simple **black line** and a silly **green squiggle** instead of a…

Deep Learning Convolutional Neural Network

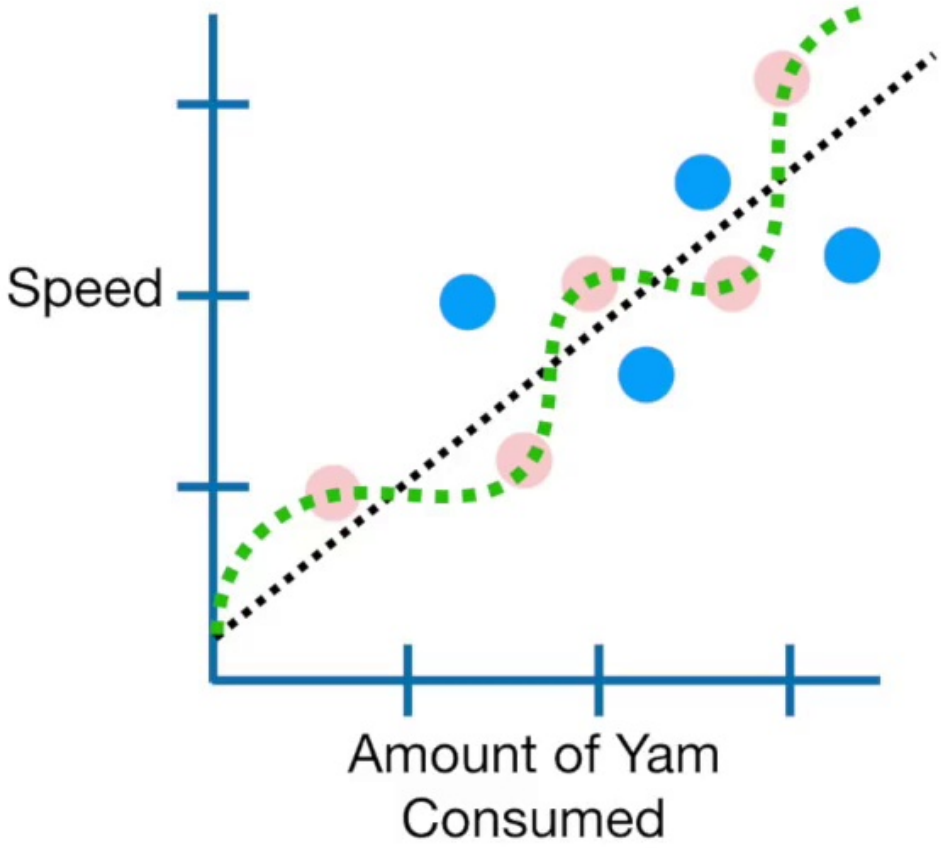
or…



Before we move on, you may be wondering why we used a simple **black line** and a silly **green squiggle** instead of a…

Deep Learning Convolutional Neural Network

or…

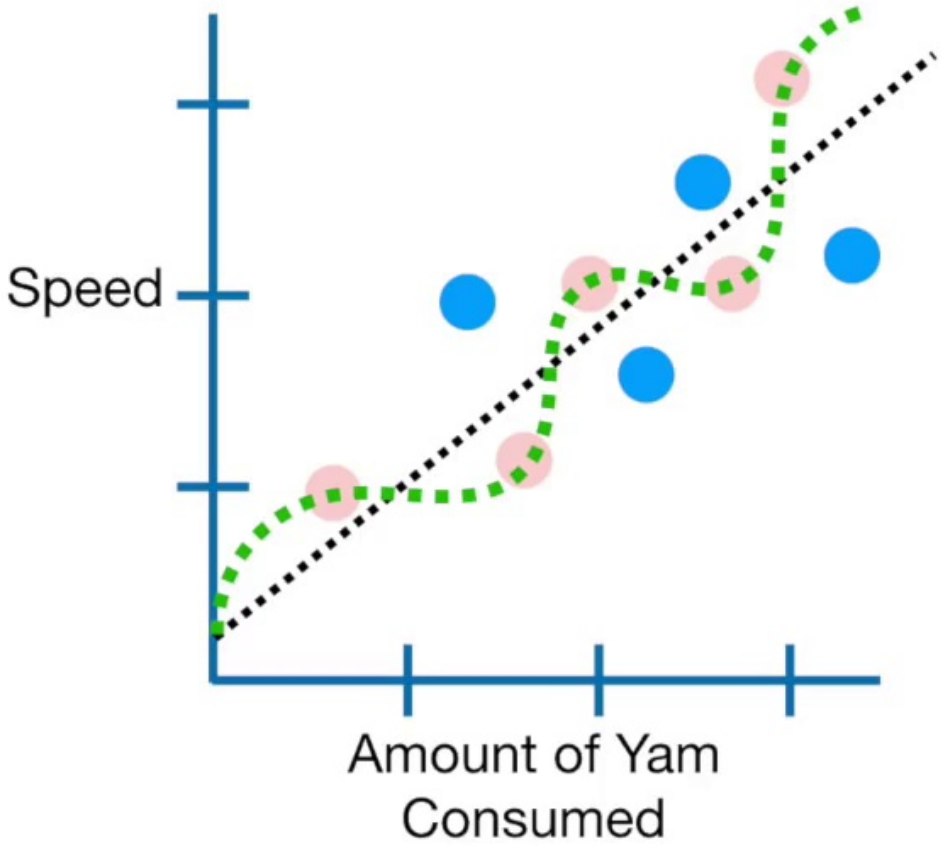


Before we move on, you may be wondering why we used a simple **black line** and a silly **green squiggle** instead of a…

Deep Learning Convolutional Neural Network

or…

[insert newest, bestest, most fancy machine learning method here].

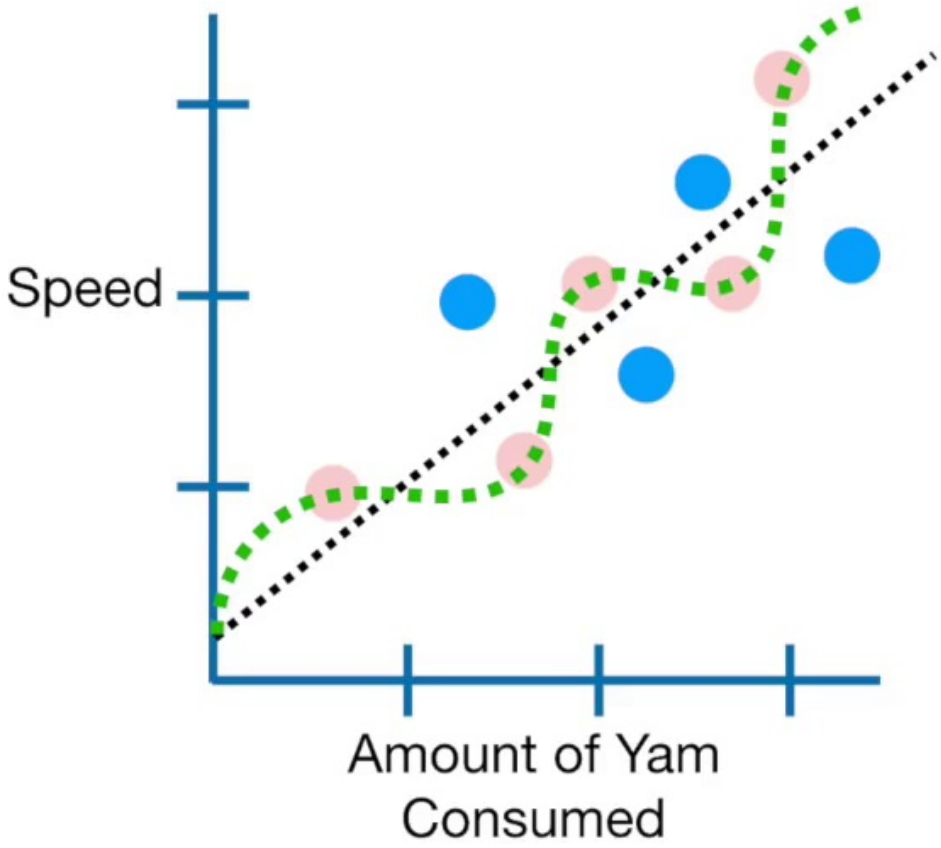


Before we move on, you may be wondering why we used a simple **black line** and a silly **green squiggle** instead of a…

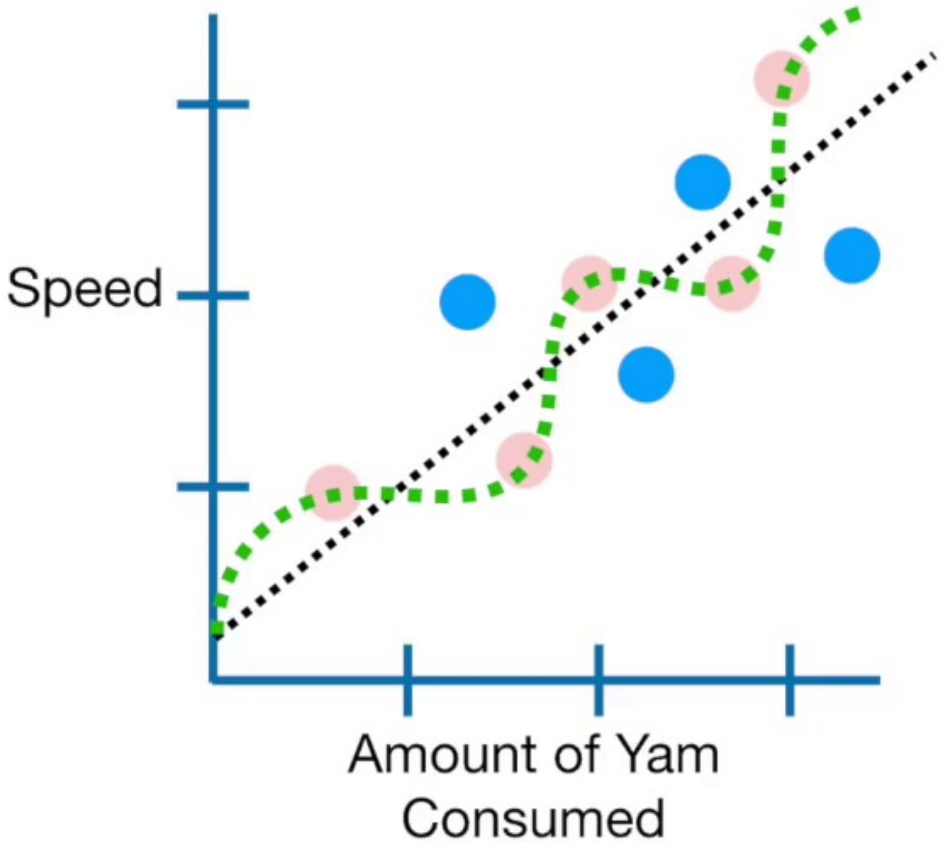
Deep Learning Convolutional Neural Network

or…

[insert newest, bestest, most fancy machine learning method here].

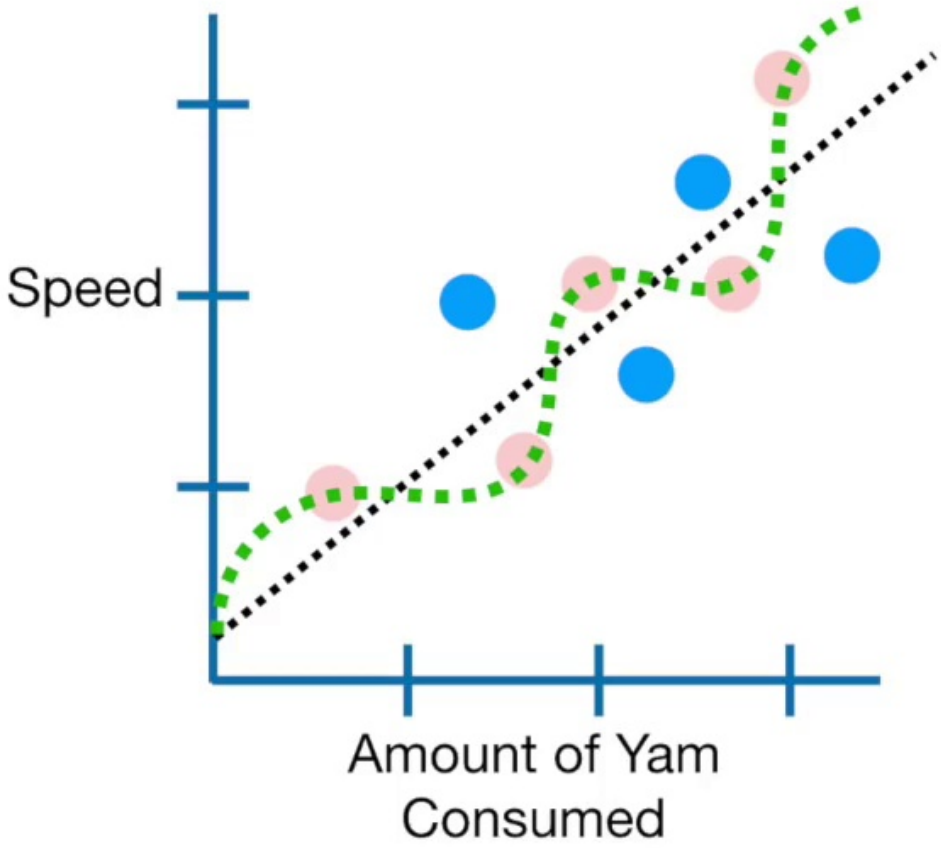


There are tons of fancy sounding **Machine Learning** methods…



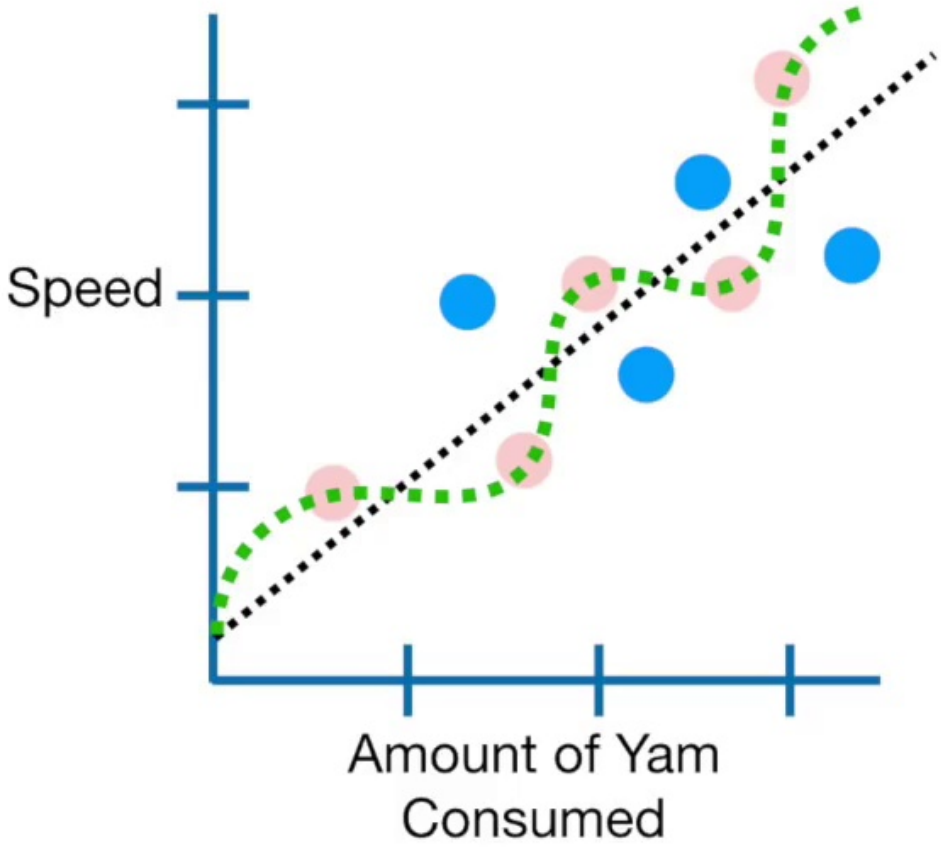
There are tons of fancy sounding **Machine Learning** methods…

…and each year something new and exciting comes on the scene…



There are tons of fancy sounding **Machine Learning** methods…

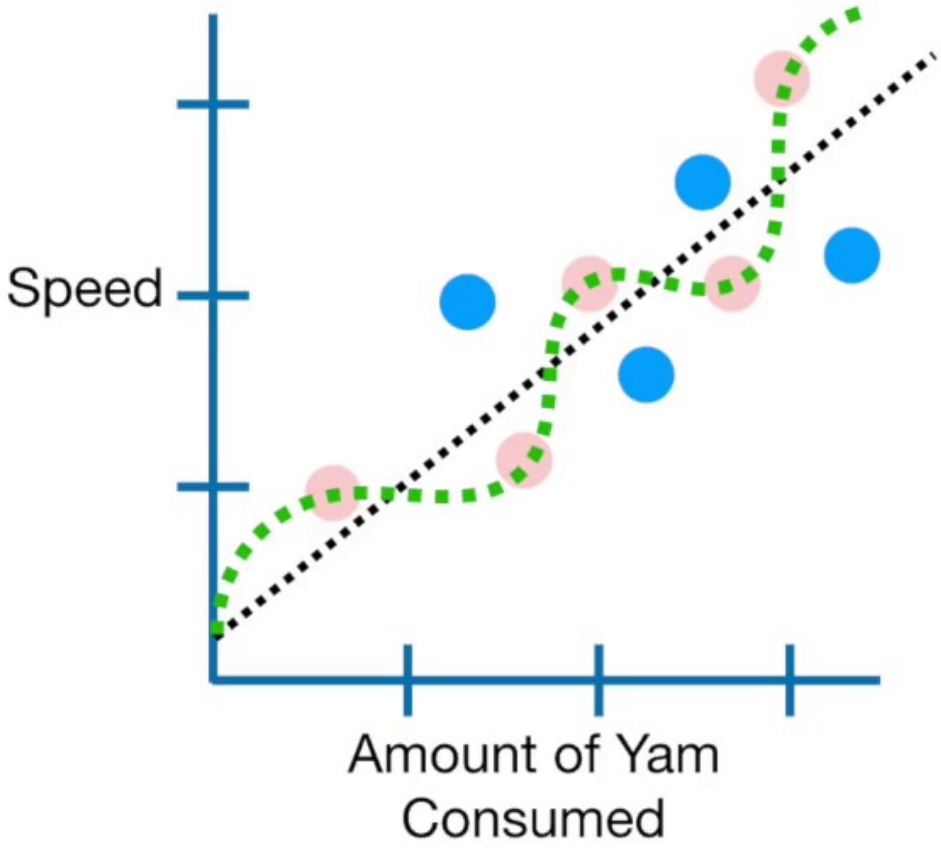
…and each year something new and exciting comes on the scene…



There are tons of fancy sounding **Machine Learning** methods…

…and each year something new and exciting comes on the scene…

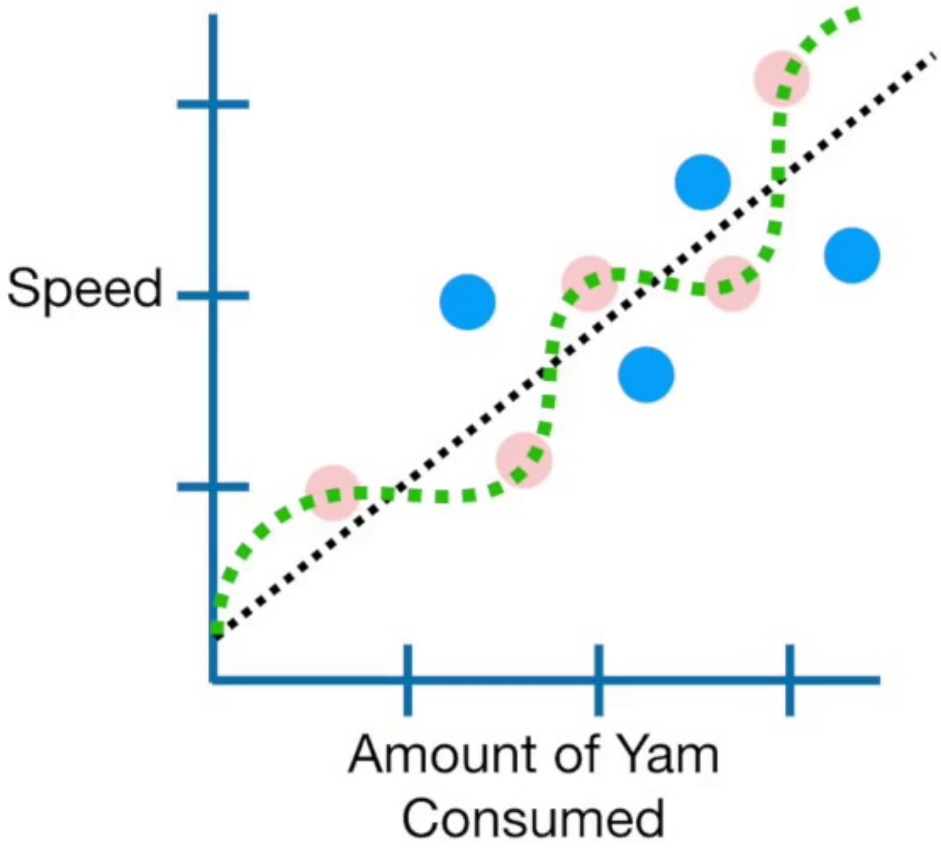
…but regardless of what you use, the most important thing isn't how fancy it is, but how it performs with **Testing Data**.



There are tons of fancy sounding **Machine Learning** methods…

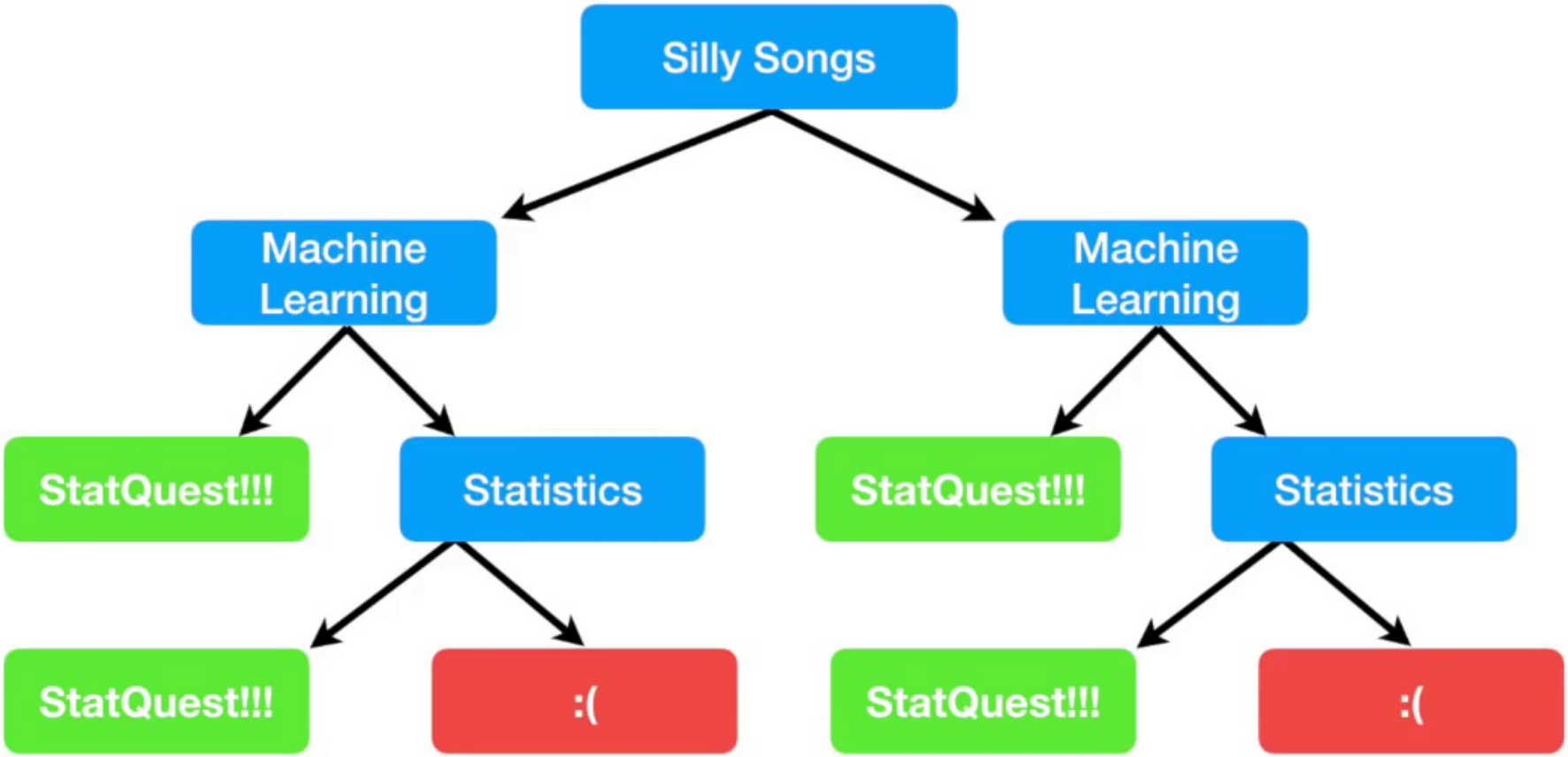
…and each year something new and exciting comes on the scene…

…but regardless of what you use, the most important thing isn't how fancy it is, but how it performs with **Testing Data**.

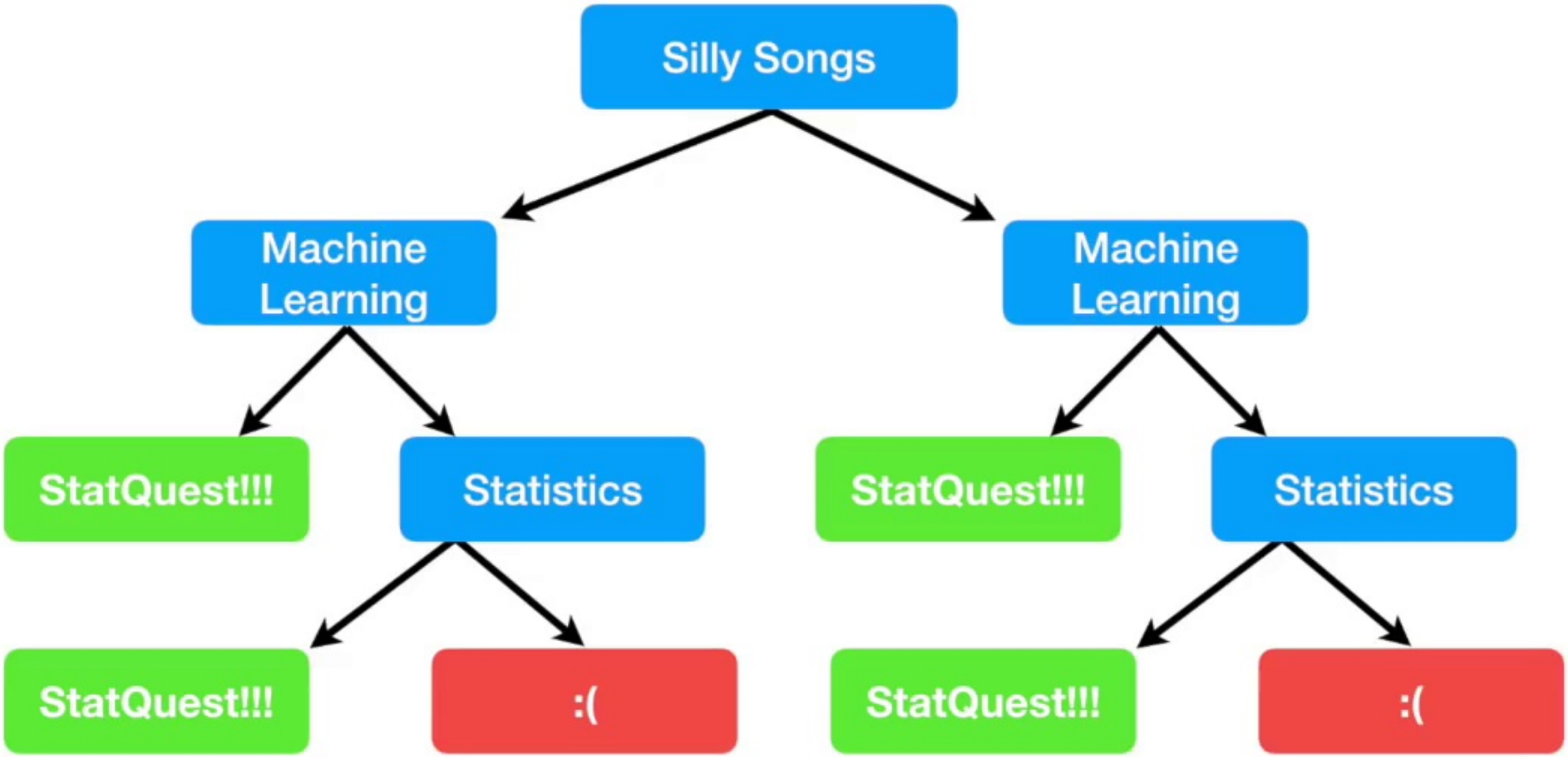


Double BAM!!!

Remember, we wanted to ***classify*** if someone loved StatQuest based on a few questions.



Remember, we wanted to ***classify*** if someone loved StatQuest based on a few questions.



Silly Songs	Machine Learning	Statistics	Love StatQuest?
Yes	No	Yes	Yes
No	No	Yes	Yes
No	Yes	No	Yes
…	…	…	…

To create the **Decision Tree**, we collected data from people who loved StatQuest…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
Yes	No	Yes	Yes
No	No	Yes	Yes
No	Yes	No	Yes
…	…	…	…

To create the **Decision Tree**, we collected data from people who loved StatQuest…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
Yes	No	Yes	Yes
No	No	Yes	Yes
No	Yes	No	Yes
…	…	…	…

No	No	No	No
No	No	No	No
No	No	No	No
…	…	…	…

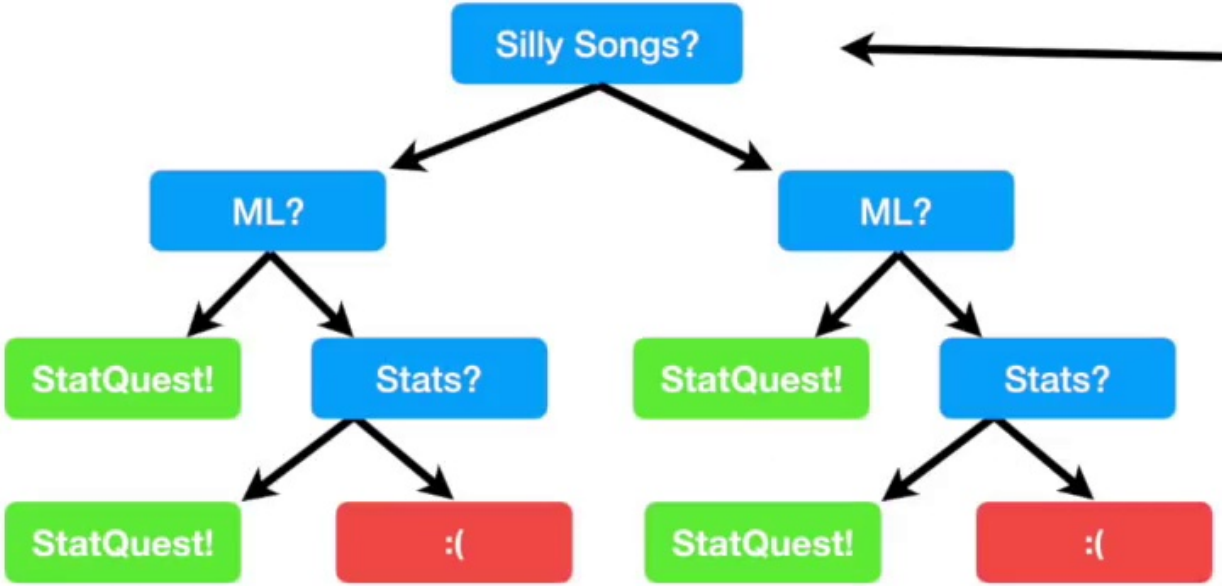
…and from people who did not love StatQuest.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
Yes	No	Yes	Yes
No	No	Yes	Yes
No	Yes	No	Yes
…	…	…	…

No	No	No	No
No	No	No	No
No	No	No	No
…	…	…	…

Altogether, this was the **Training Data**…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
Yes	No	Yes	Yes
No	No	Yes	Yes



…and we used it to build the **Decision Tree**.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
…	…	…	…

Then we got data from a few more people who loved StatQuest…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
…	…	…	…

Then we got data from a few more people who loved StatQuest…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
…	…	…	…

No	No	Yes	No
No	No	No	No
No	No	Yes	No
…	…	…	…

…and a few more people who did not love StatQuest.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
...	...	...	...

No	No	Yes	No
No	No	No	No
No	No	Yes	No
...	...	...	...

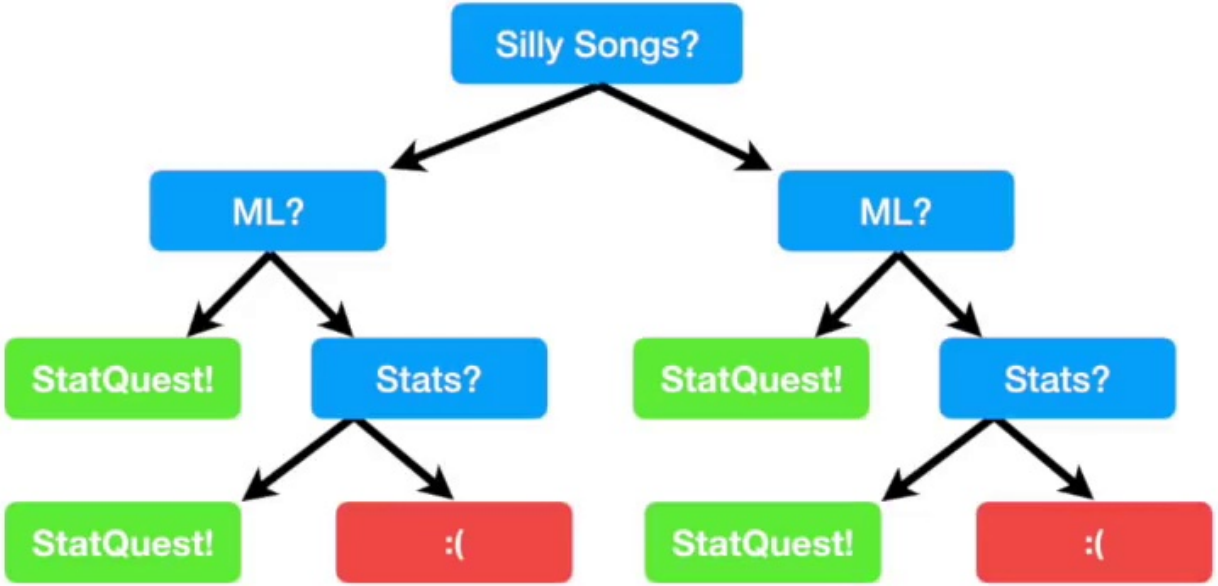
Altogether, this forms the **Testing Data**.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
...	...	...	...

No	No	Yes	No
No	No	No	No
No	No	Yes	No
...	...	...	...

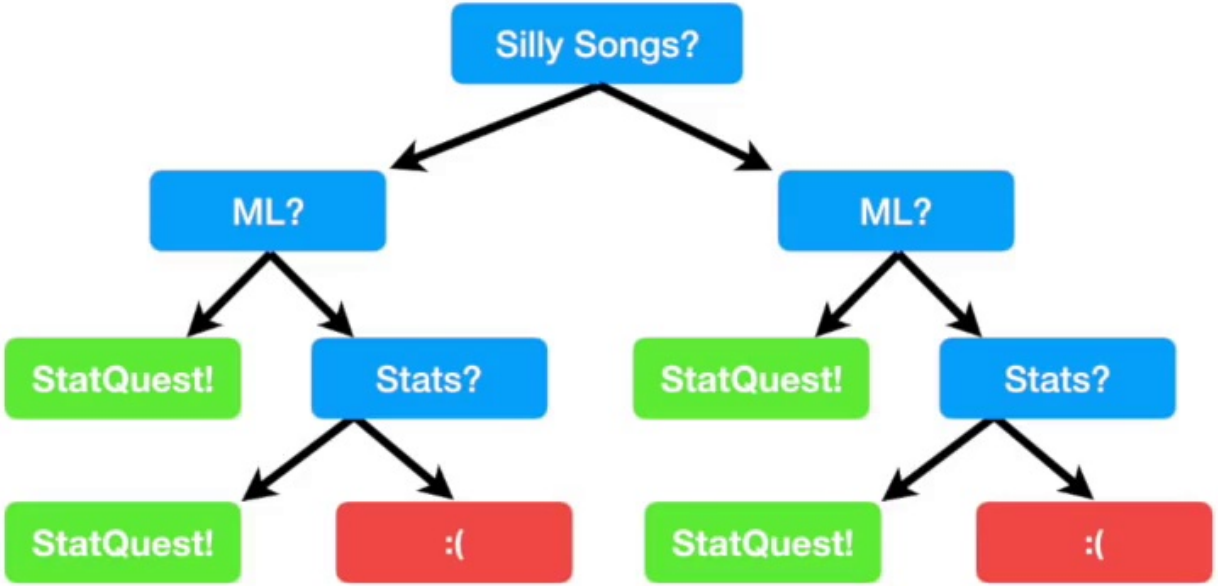
Altogether, this forms the **Testing Data**.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes



We can use the **Testing Data** to see how well our **Decision Tree** predicts if someone will love StatQuest.

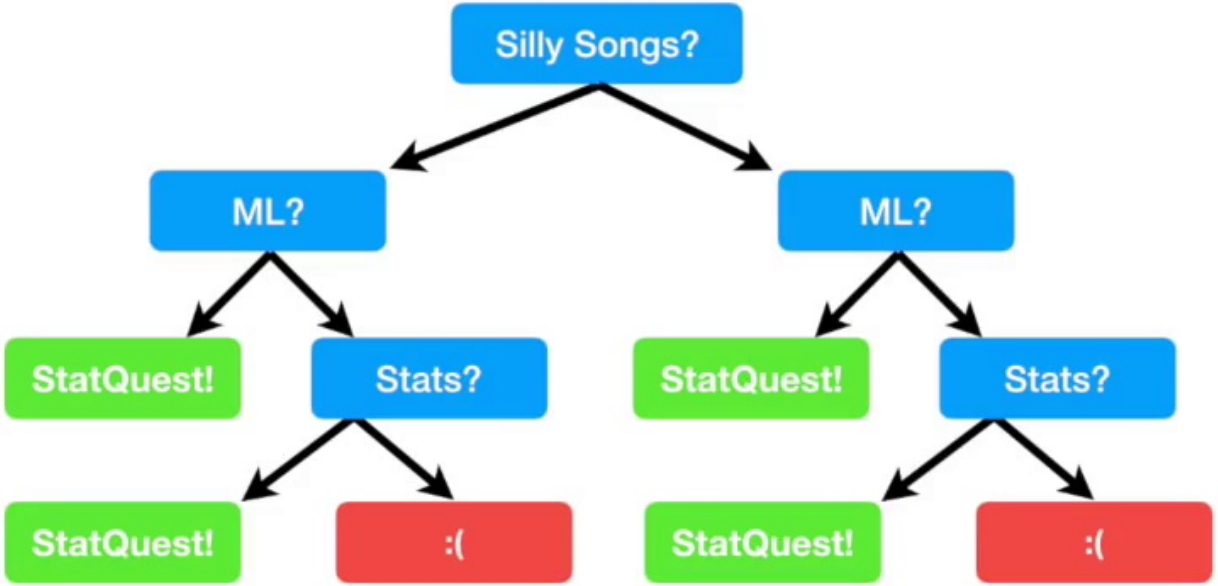
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes



We can use the **Testing Data** to see how well our **Decision Tree** predicts if someone will love StatQuest.

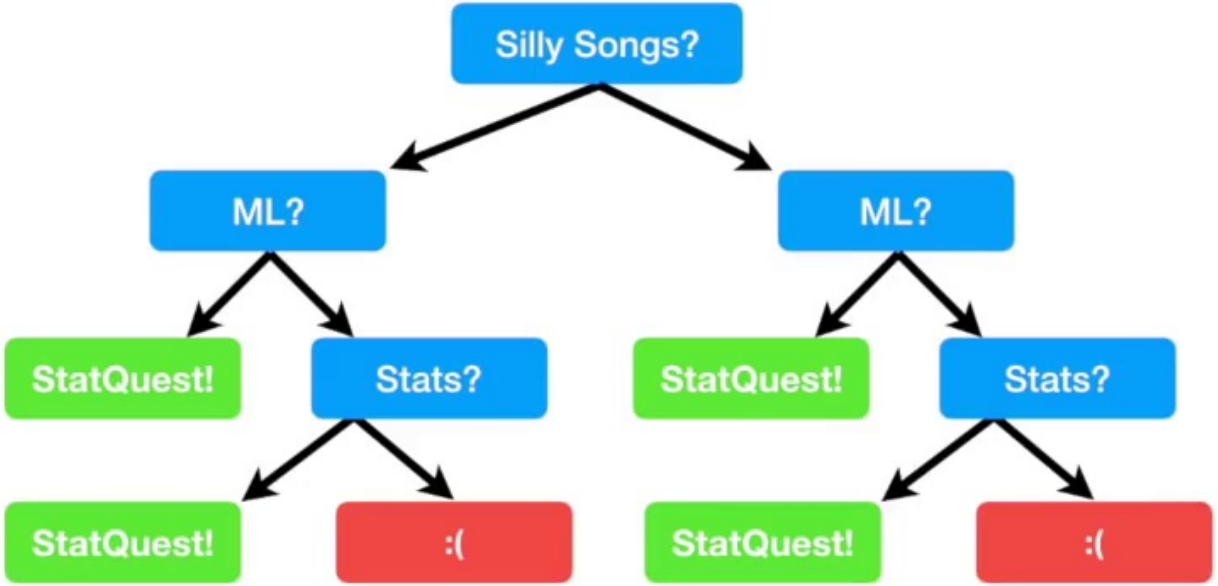
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

The first person in the **Testing Data** did not like silly songs…

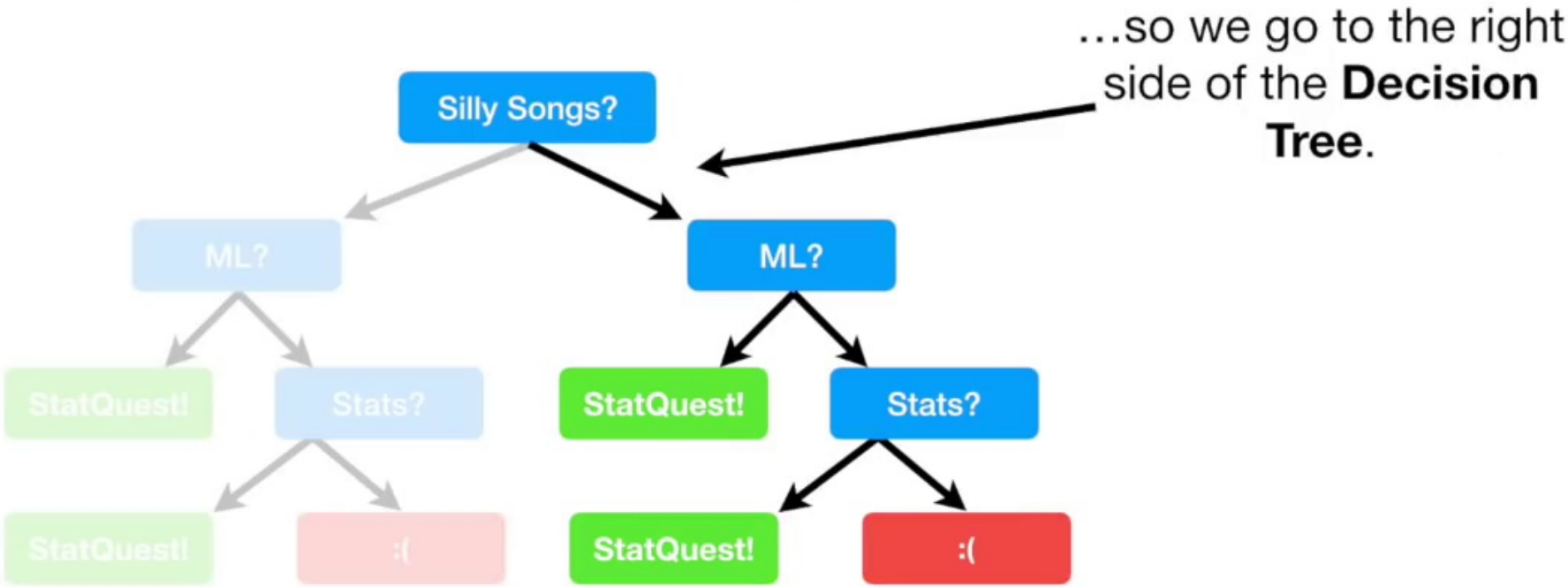


Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

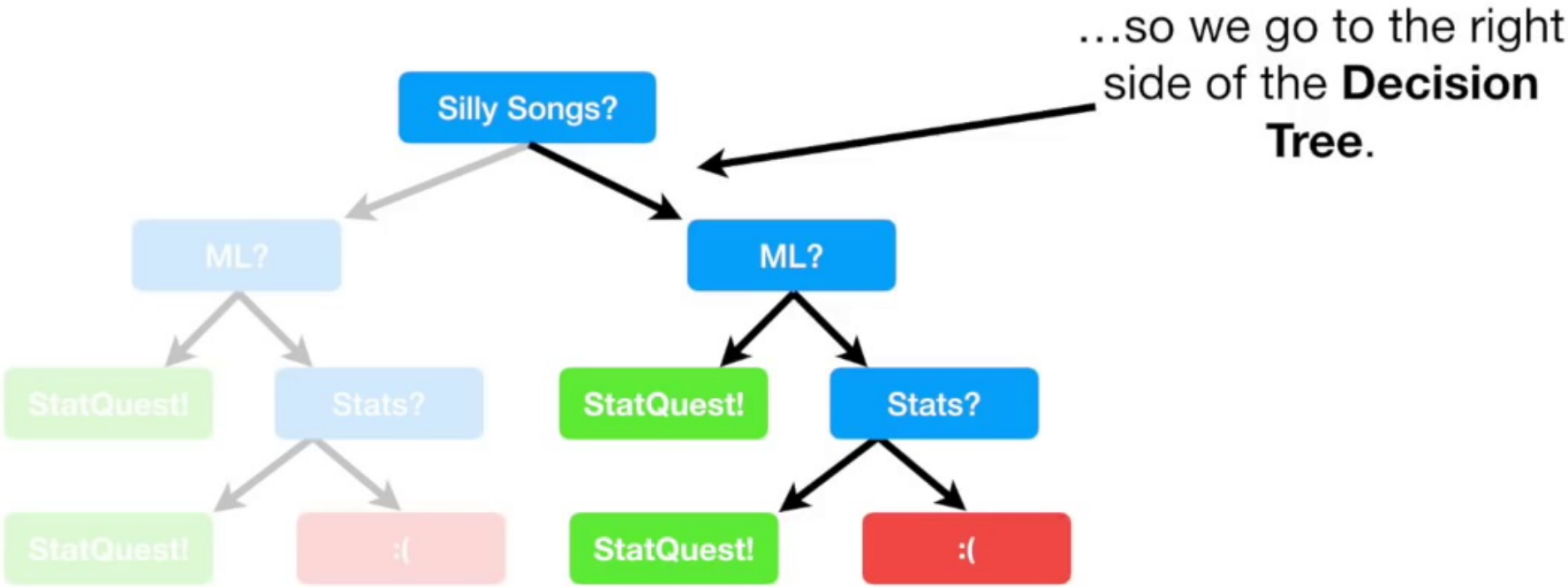
The first person in the **Testing Data** did not like silly songs…



Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

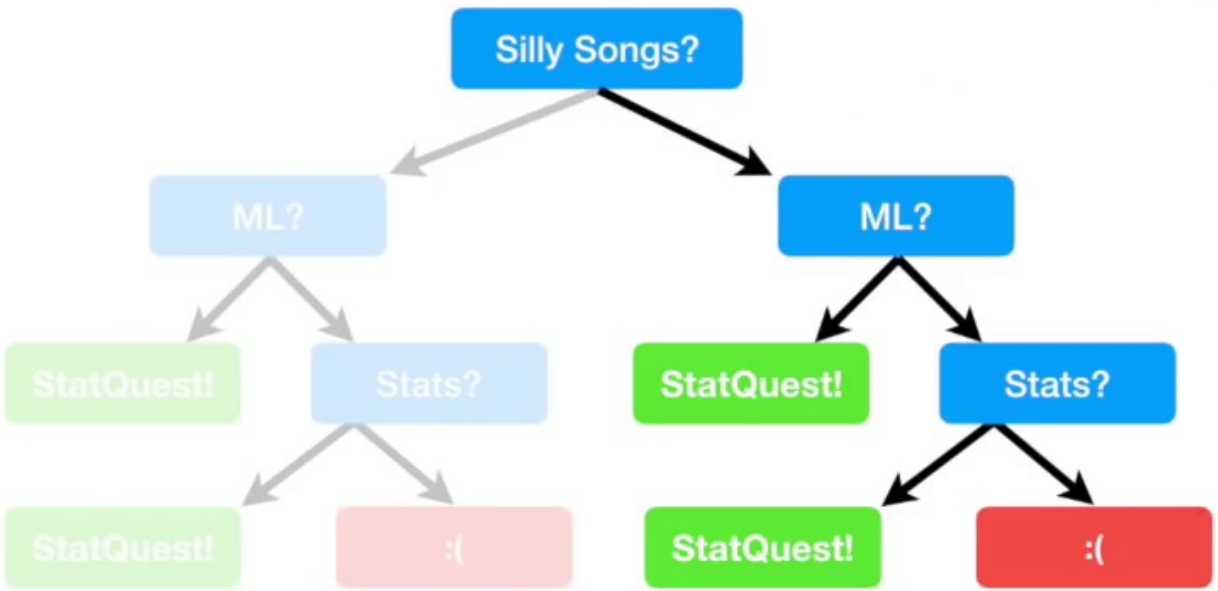


Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes



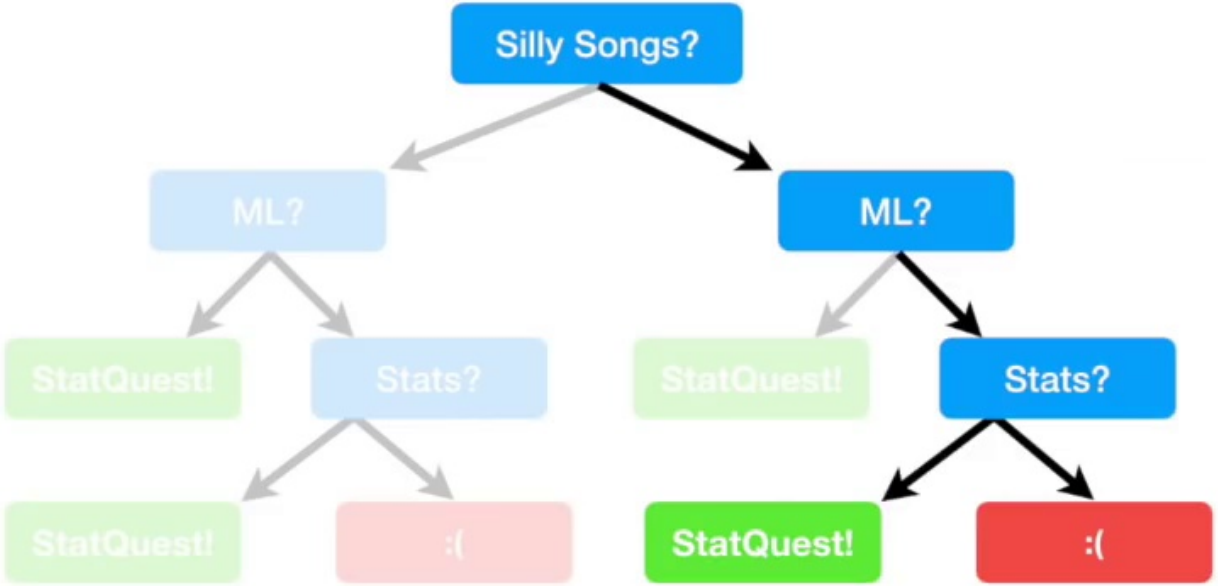
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

They didn't like Machine Learning either…



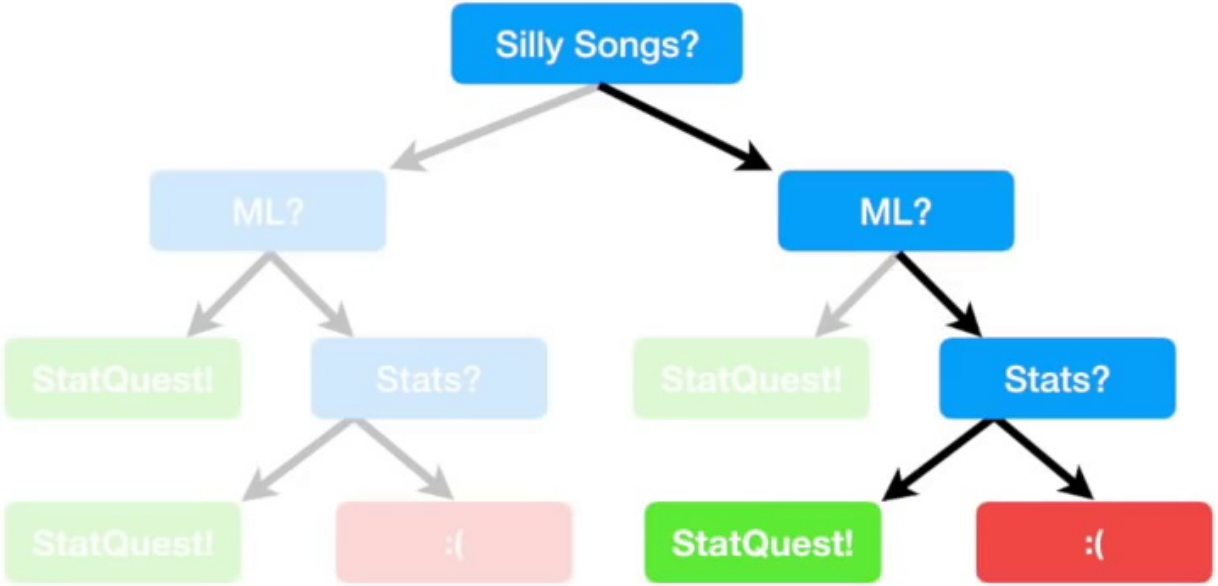
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

They didn't like Statistics either…



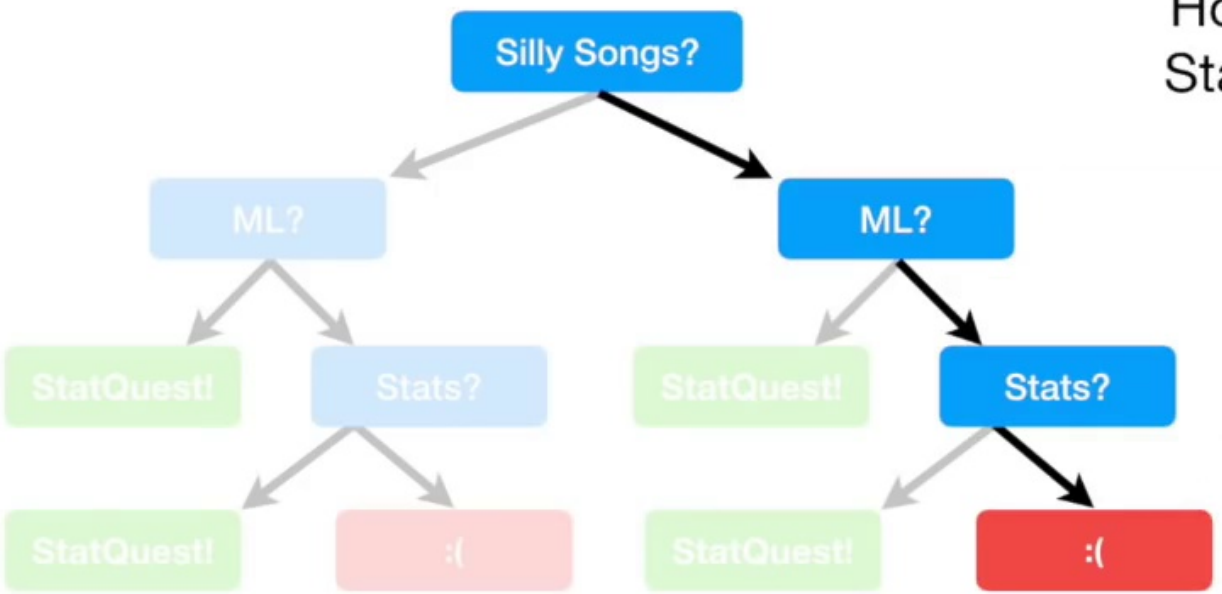
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

They didn't like Statistics either…



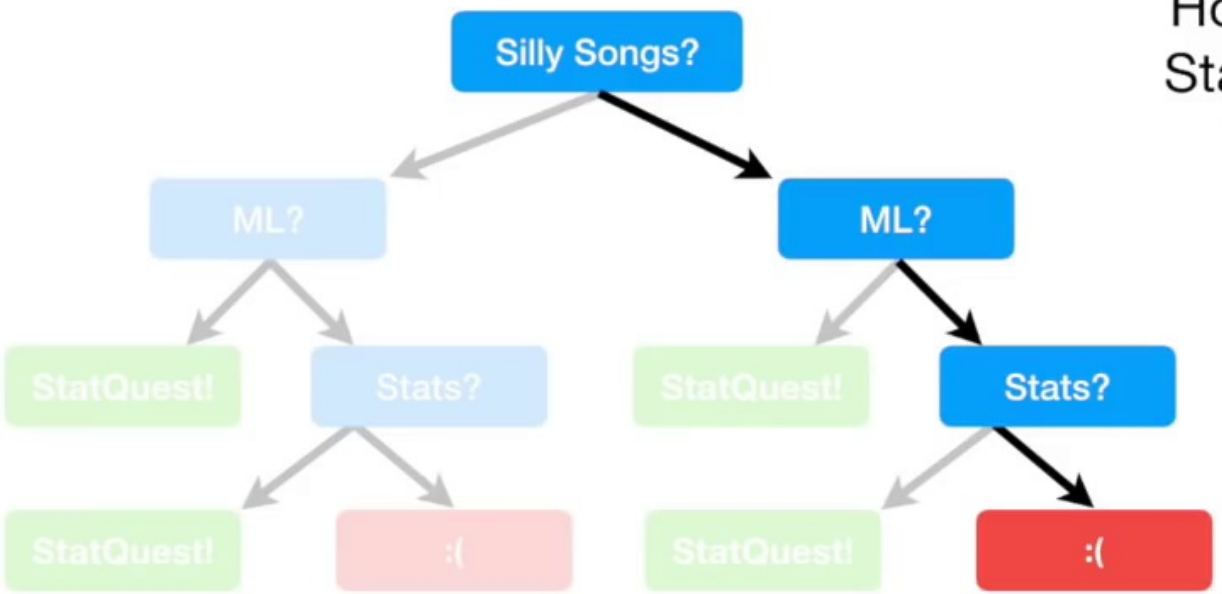
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

However, this person loves StatQuest, so the **Decision Tree** made a mistake.



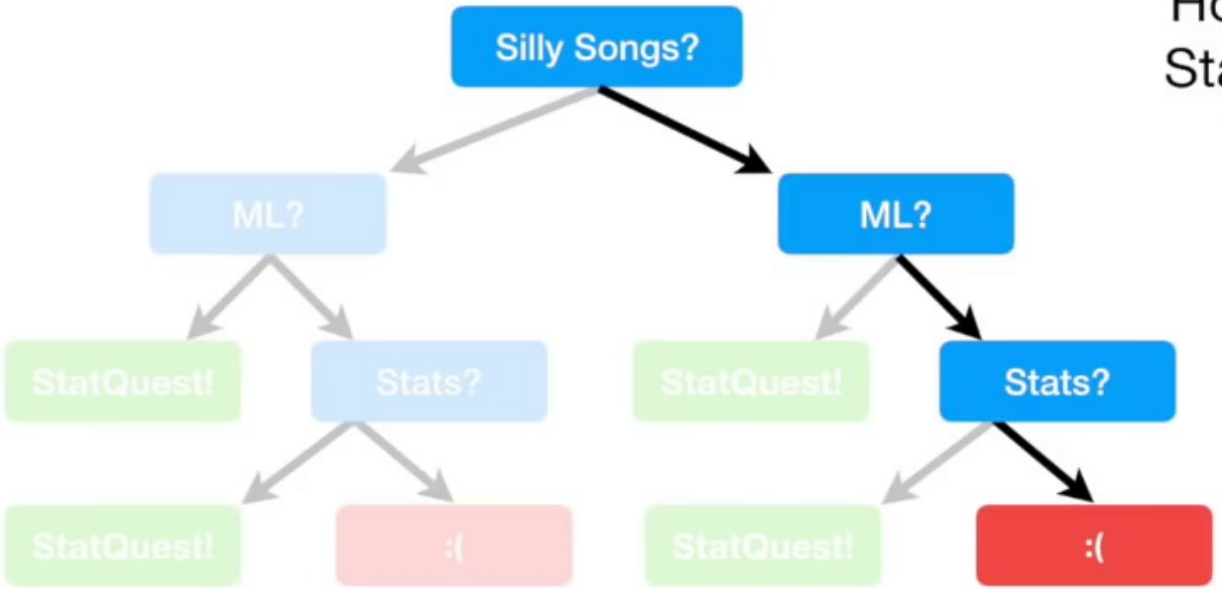
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

However, this person loves StatQuest, so the **Decision Tree** made a mistake.



Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

However, this person loves StatQuest, so the **Decision Tree** made a mistake.

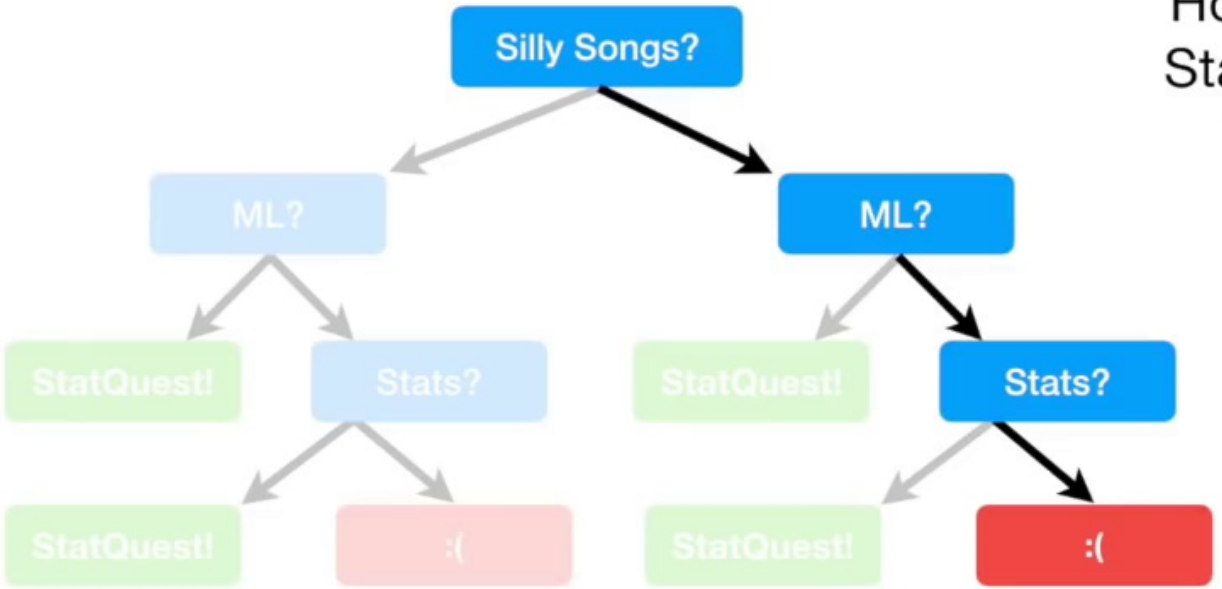


:(

wa, waa…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes

However, this person loves StatQuest, so the **Decision Tree** made a mistake.

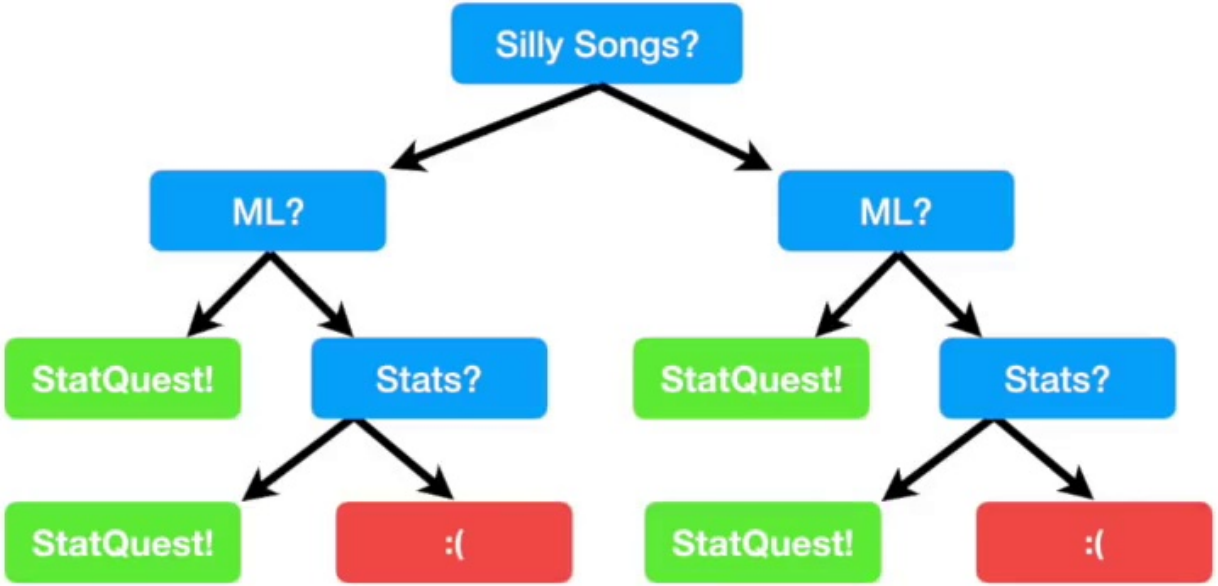


:(

wa, waa…

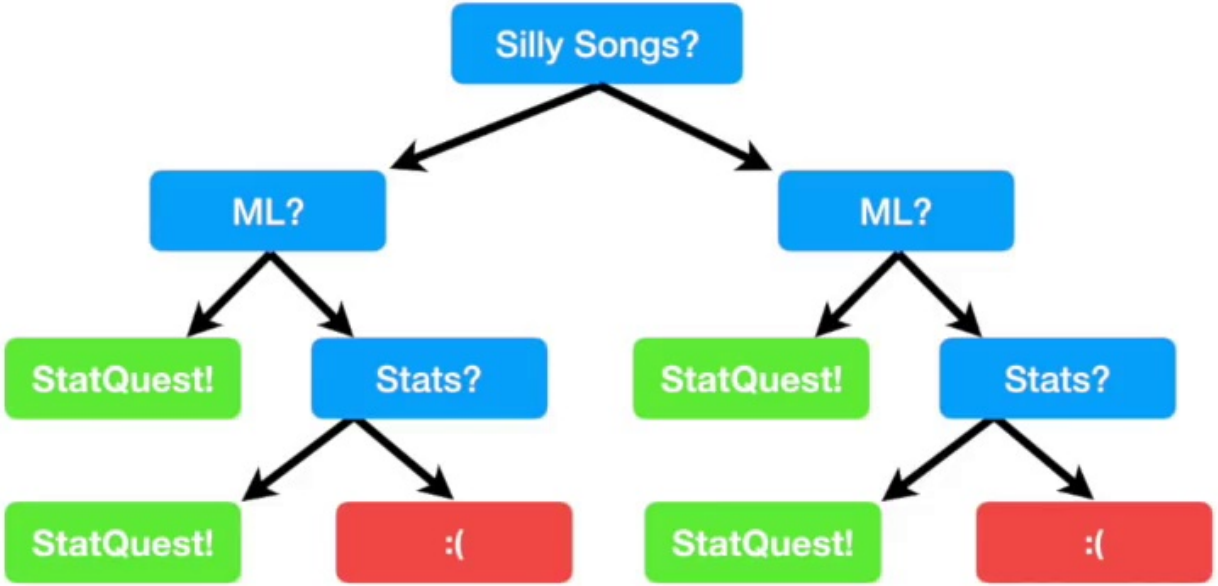
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes

The second person in the **Testing Data** liked Silly Songs…

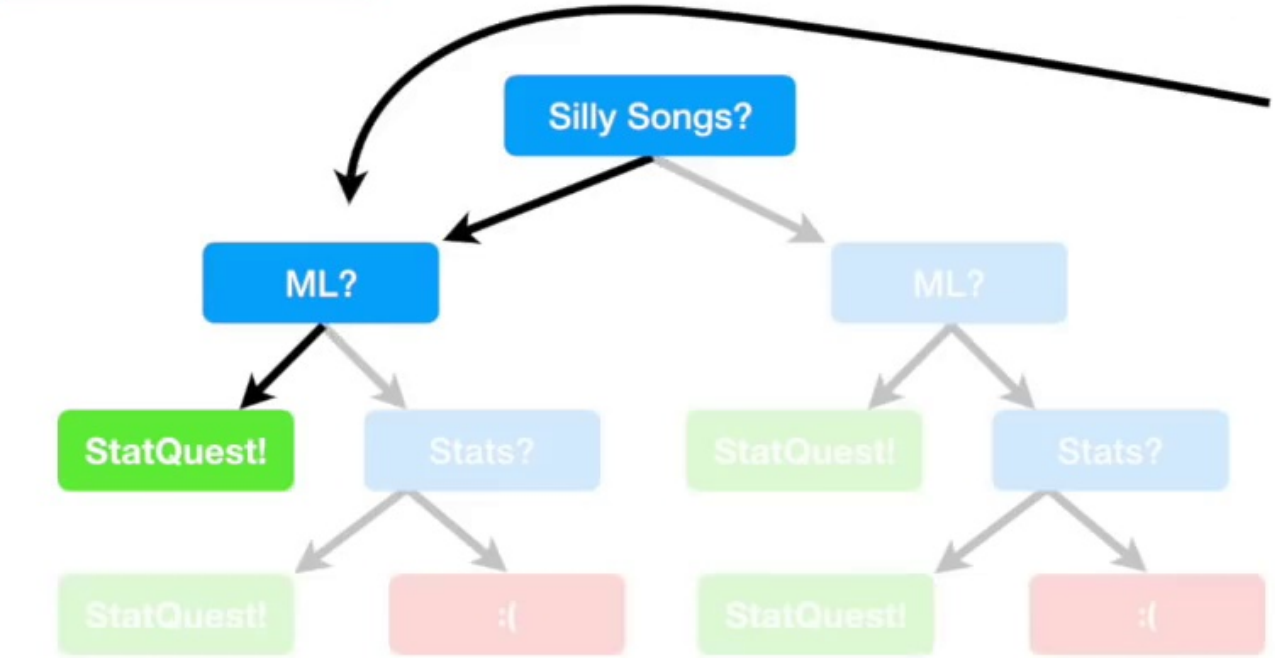


Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes

The second person in the **Testing Data** liked Silly Songs…

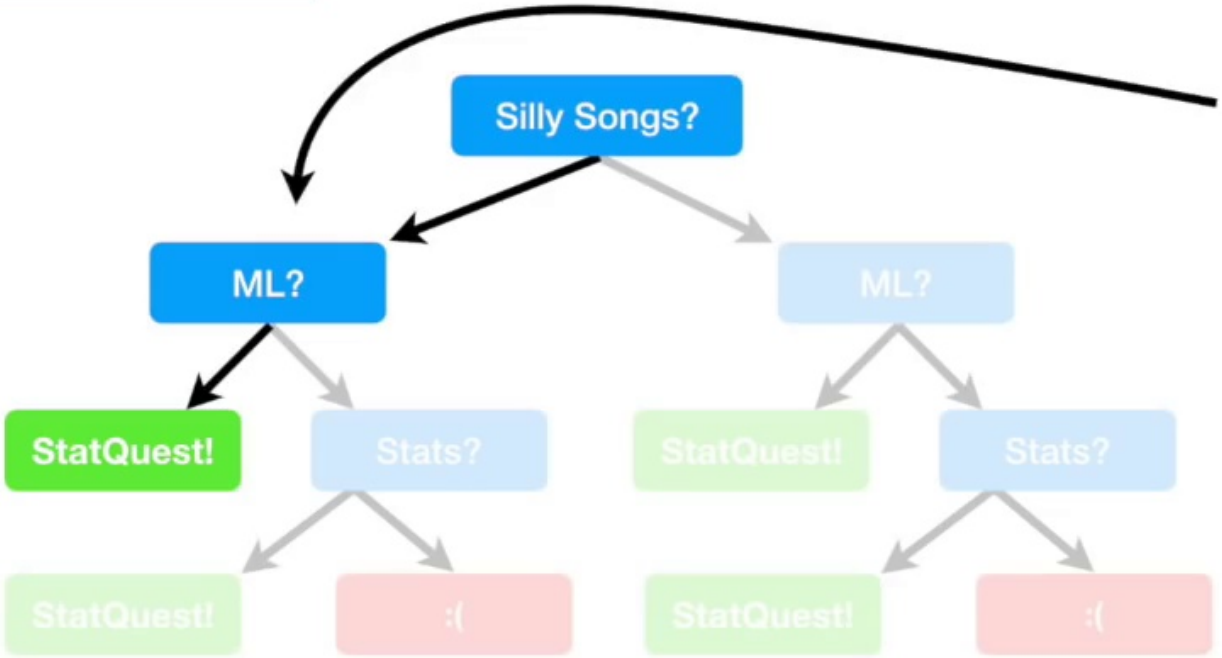


Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes



…and that takes us down the left side of the **Decision Tree**…

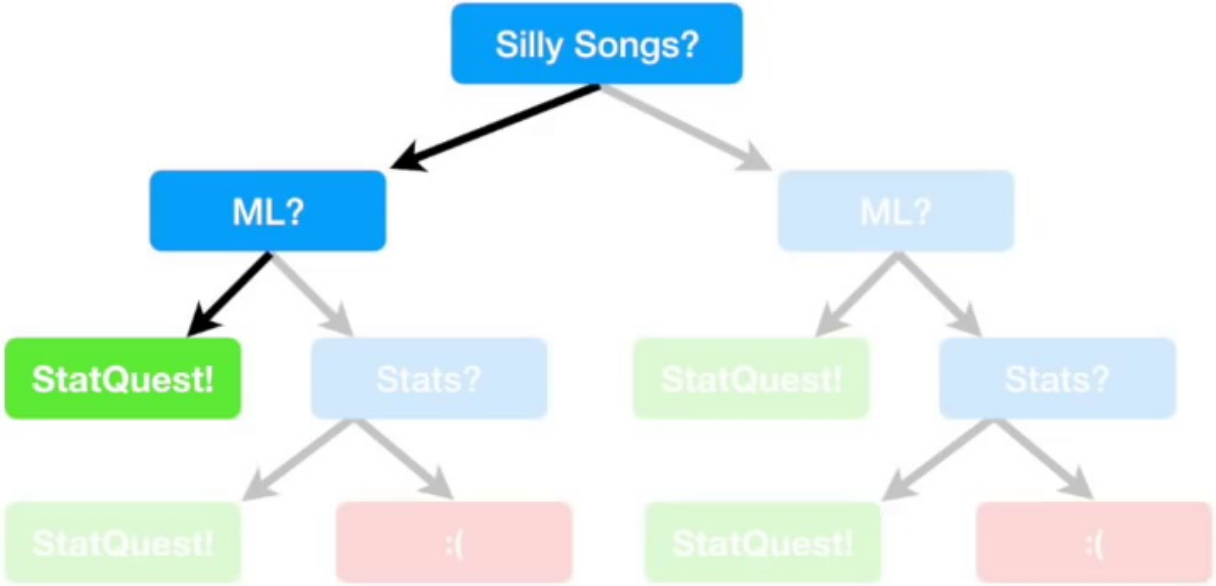
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes



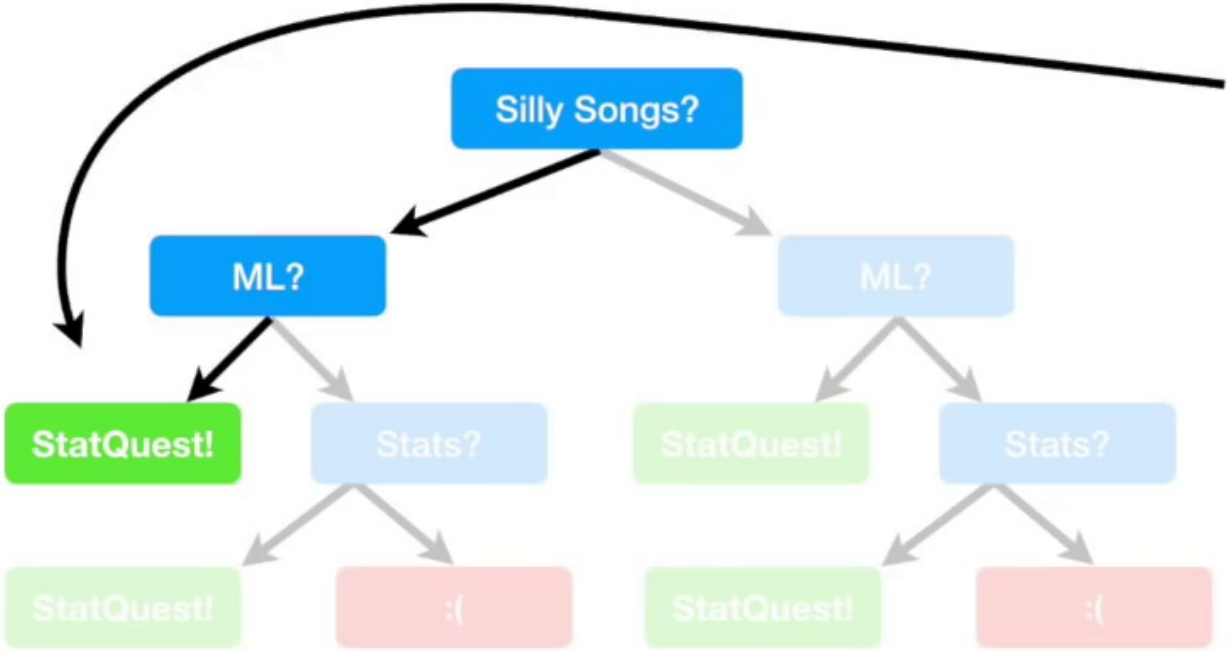
…and that takes us down the left side of the **Decision Tree**…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes

They were also interested in Machine Learning…

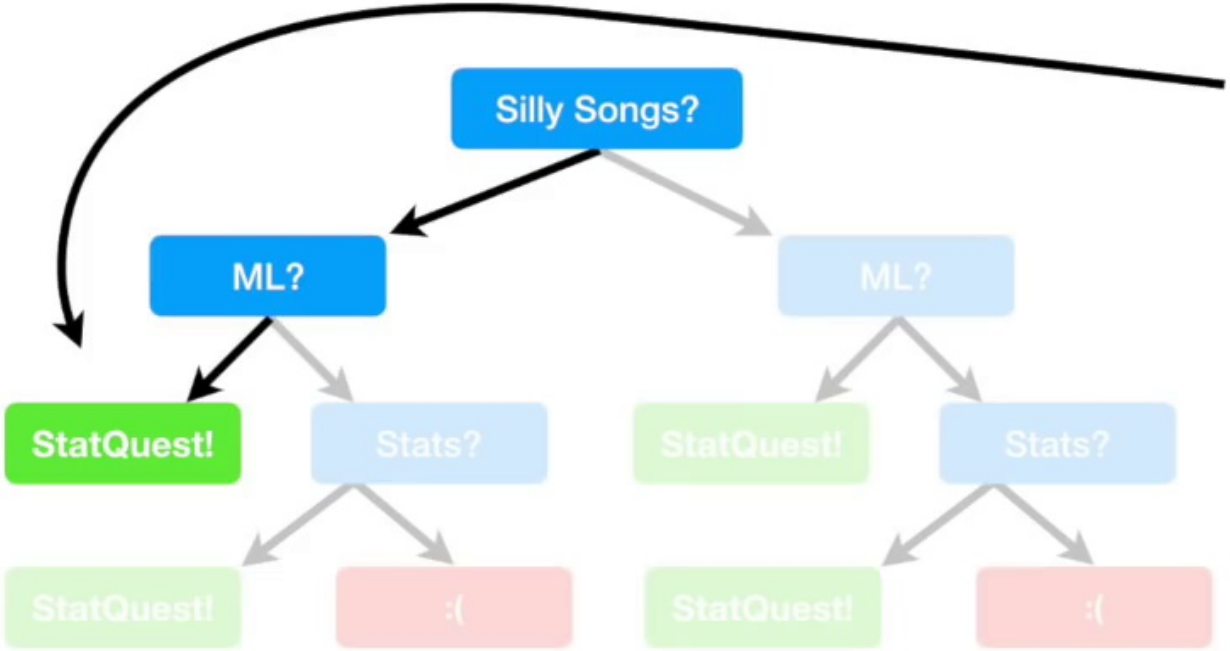


Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes



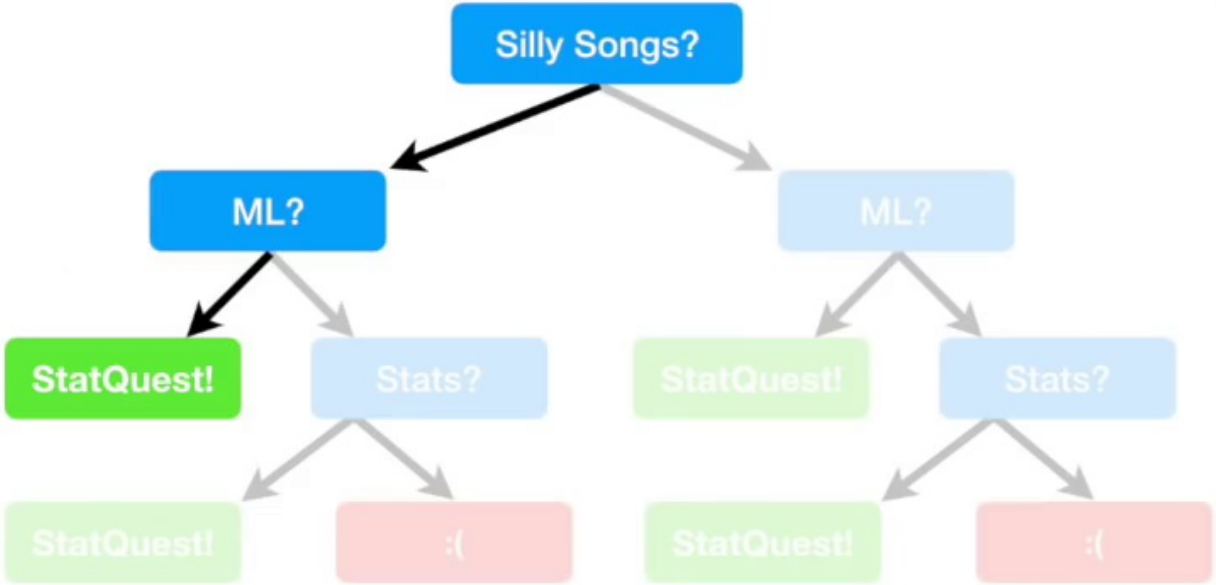
…so we predict that that person loves StatQuest…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes



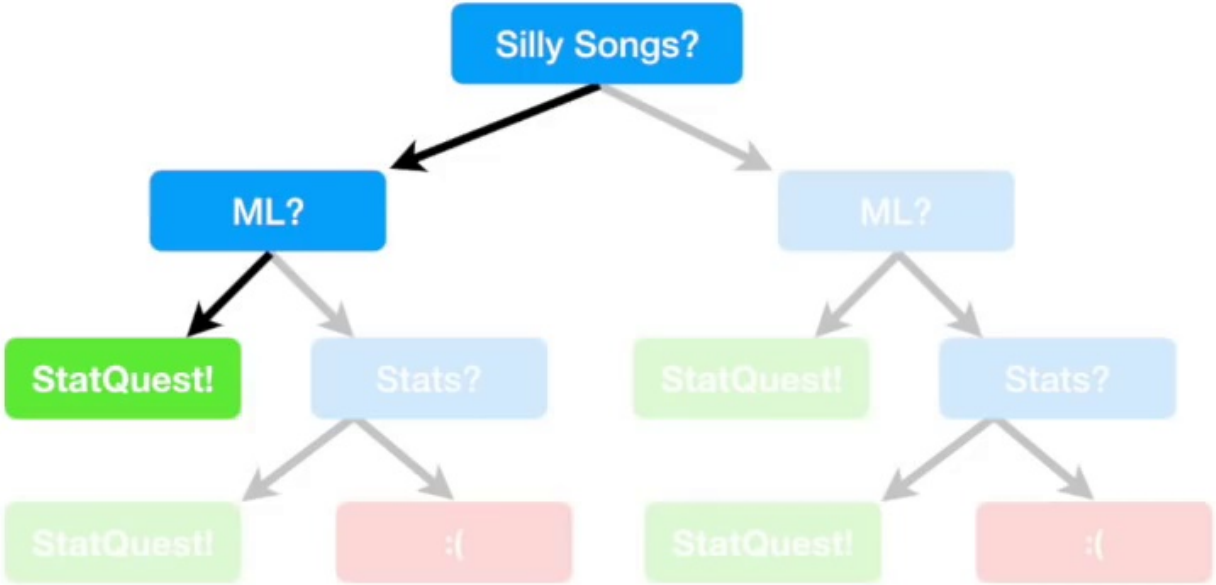
…so we predict that that person loves StatQuest…

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes



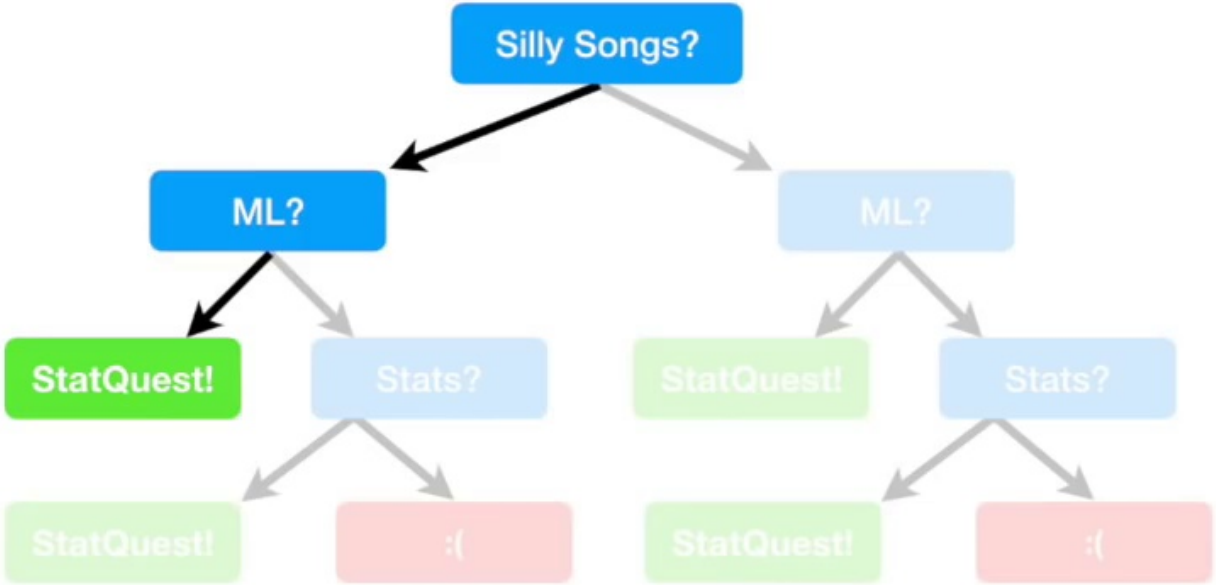
…and since this person actually loves StatQuest, the **Decision Tree** did a good job!

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes



…and since this person actually loves StatQuest, the **Decision Tree** did a good job!

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes

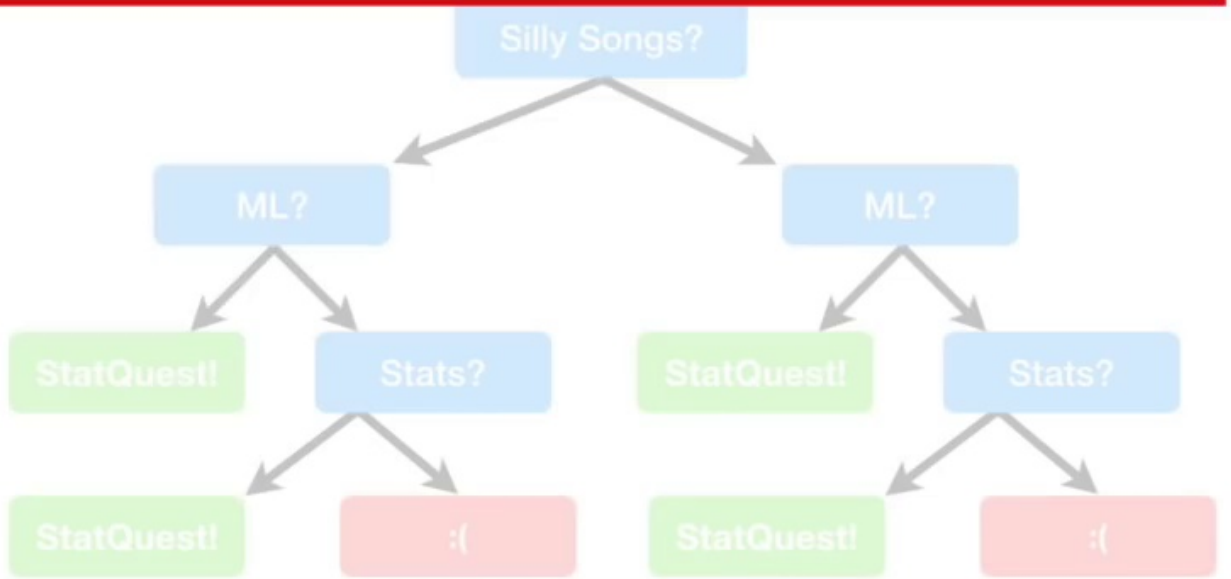


…and since this person actually loves StatQuest, the **Decision Tree** did a good job!

:)

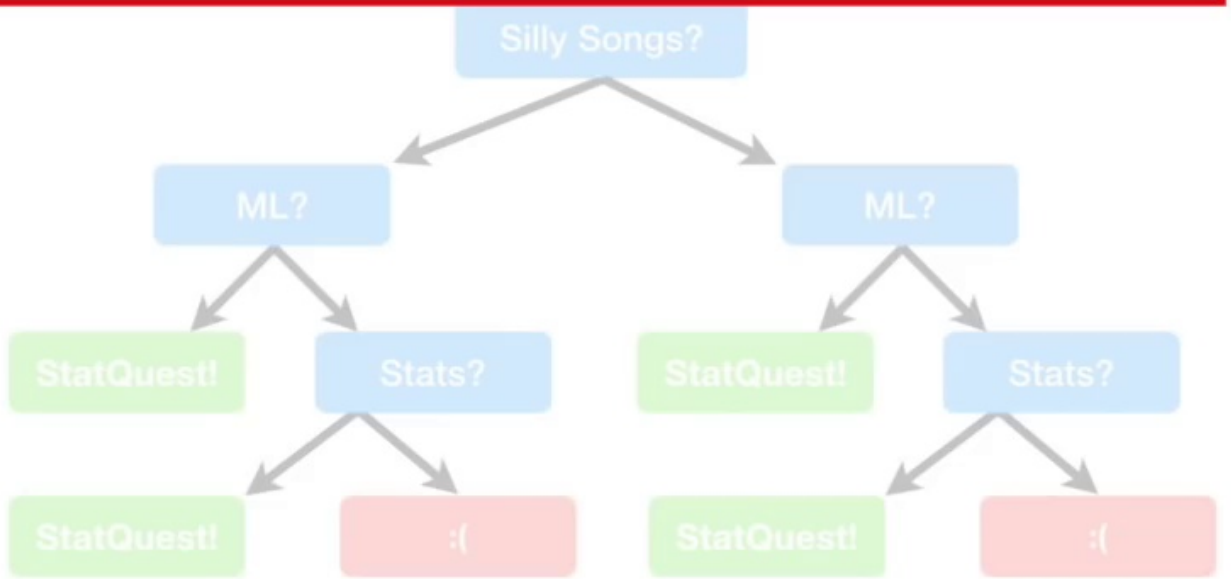
Hooray!

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes



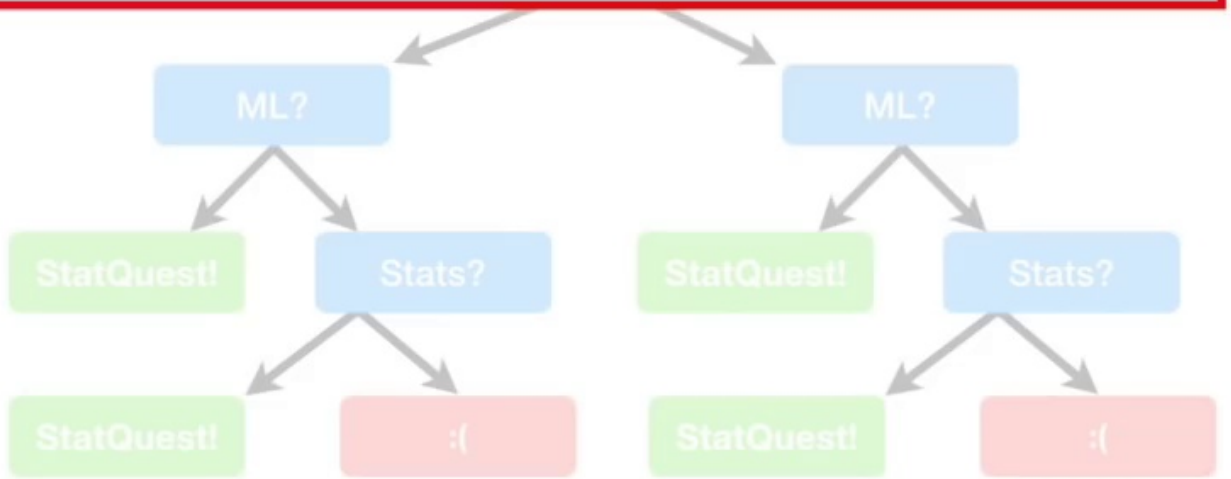
Now we just run all of the other people in the **Testing Data** down the **Decision Tree** and compare the predictions to reality.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes



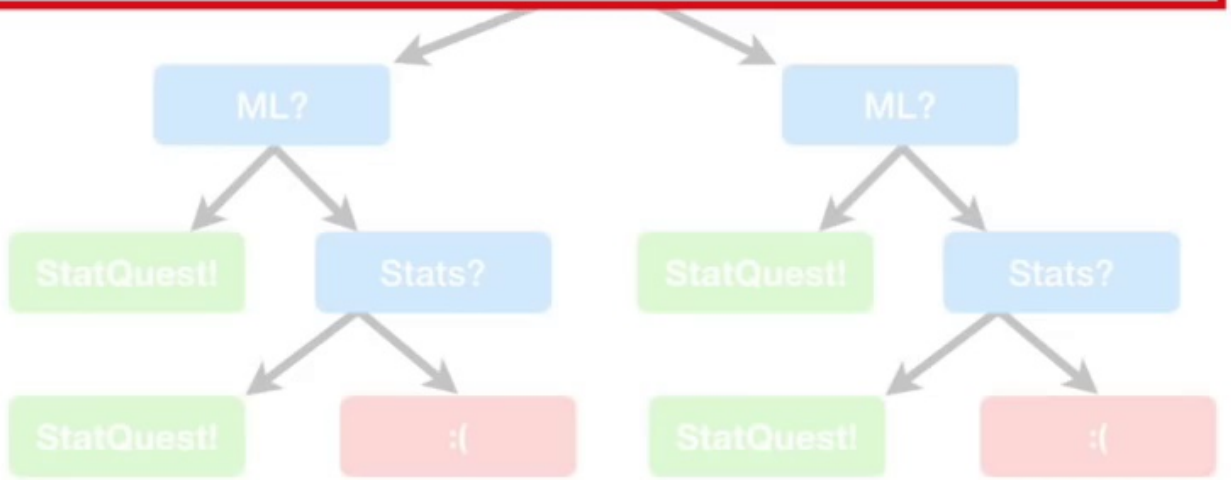
Now we just run all of the other people in the **Testing Data** down the **Decision Tree** and compare the predictions to reality.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
...	...	...	...



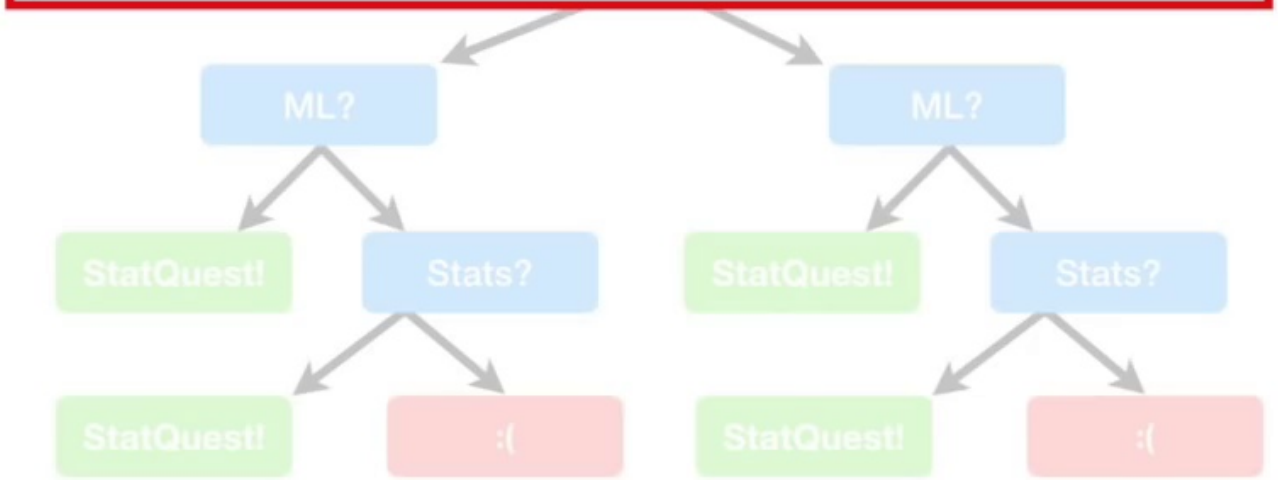
Now we just run all of the other people in the **Testing Data** down the **Decision Tree** and compare the predictions to reality.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
...	...	...	...



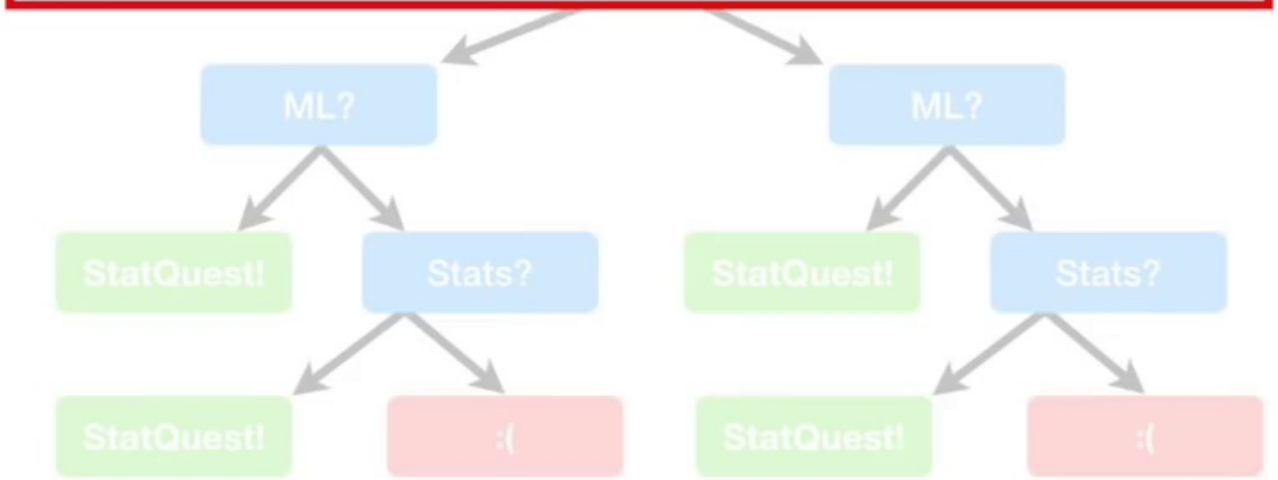
Then we can compare this **Decision Tree** to the latest, greatest **Machine Learning** method.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
...	...	...	...



Then we can compare this **Decision Tree** to the latest, greatest **Machine Learning** method.

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes	Yes	No	Yes
Yes	No	Yes	Yes
...	...	...	...



Ultimately, we'll pick the method that does the best job predicting if someone will love StatQuest or not.

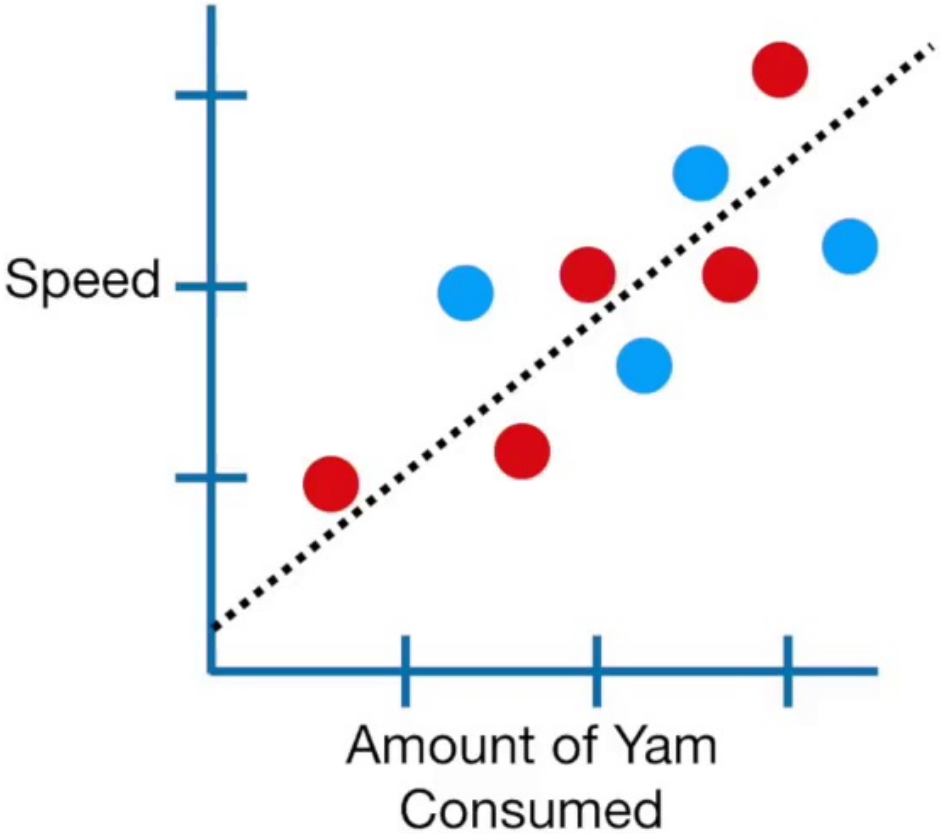
Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes			
Yes			
...			

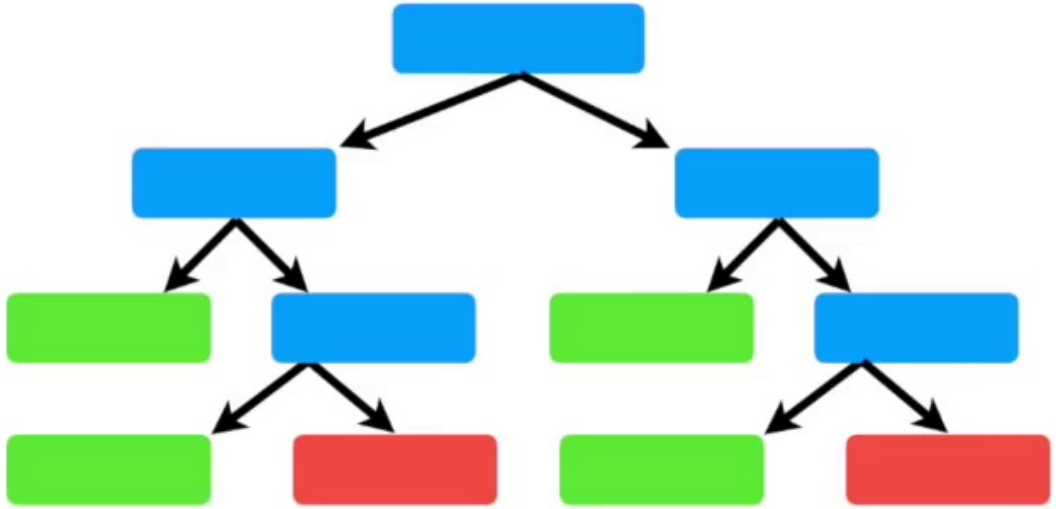
TRIPLE BAM!!!

Silly Songs	Machine Learning	Statistics	Love StatQuest?
No	No	No	Yes
Yes			
Yes			
...			

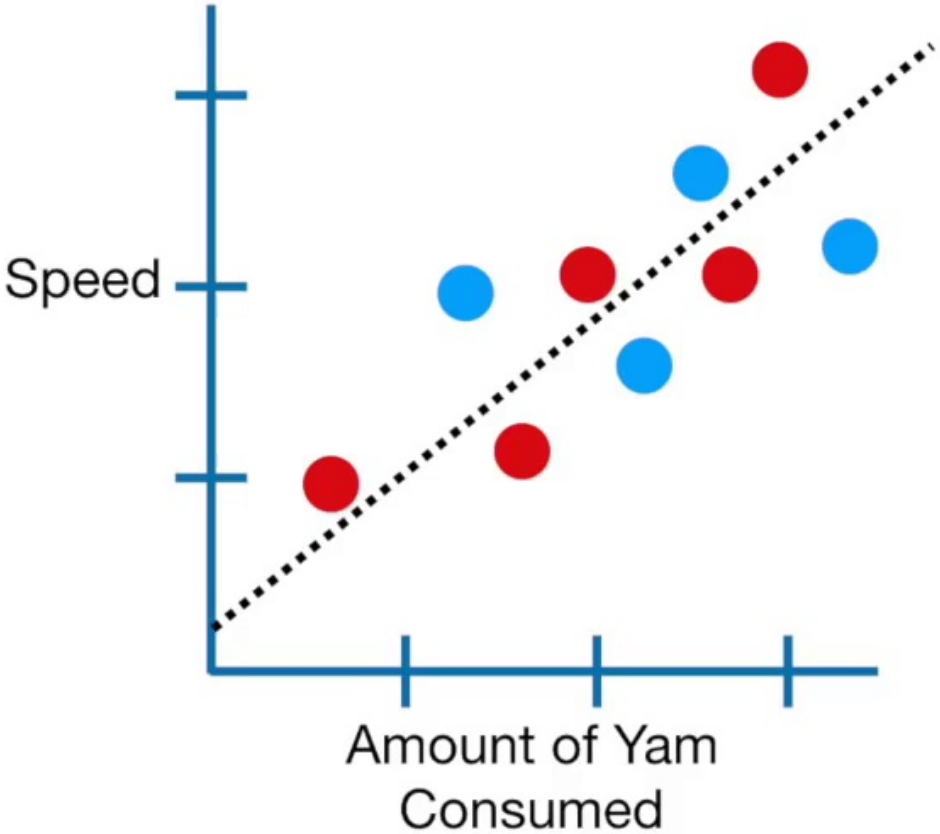
TRIPLE BAM!!!

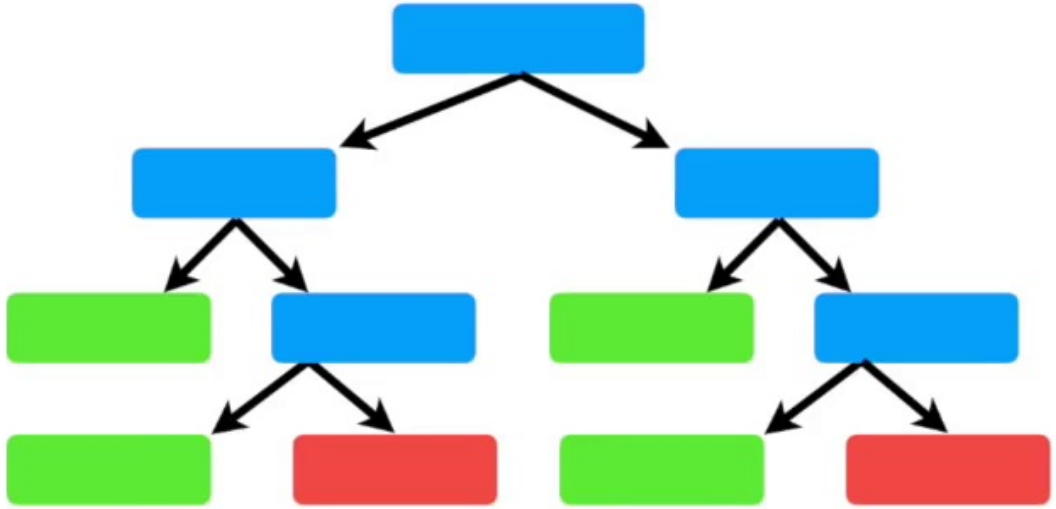
Machine Learning is all about making ***predictions*** and ***classifications***.



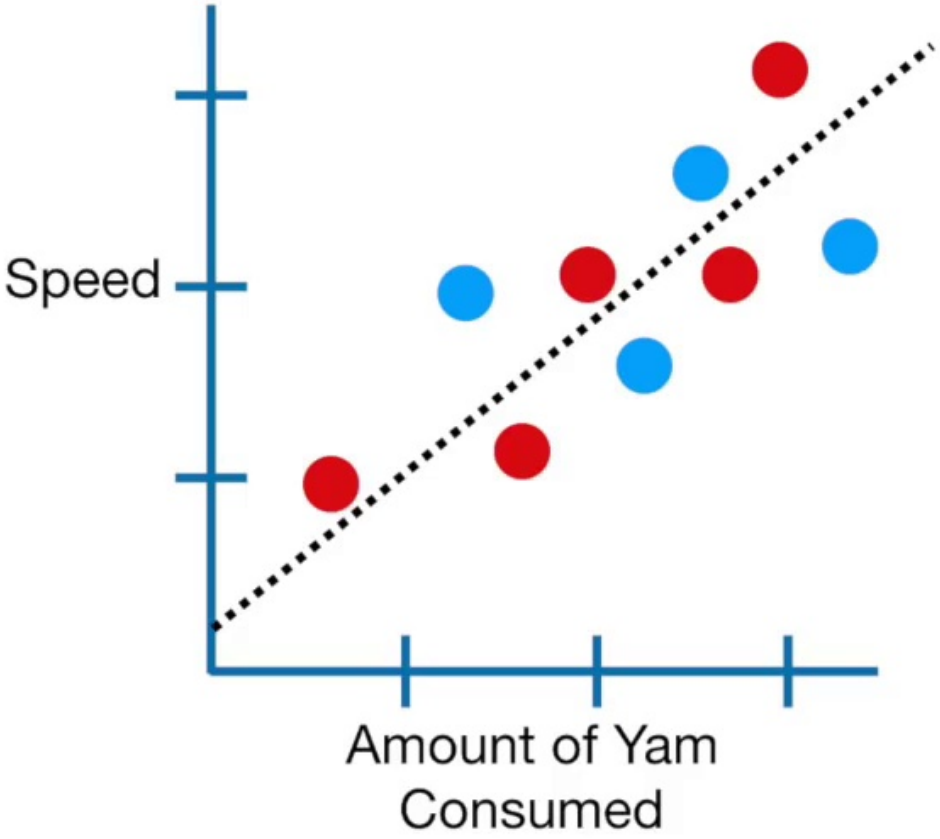


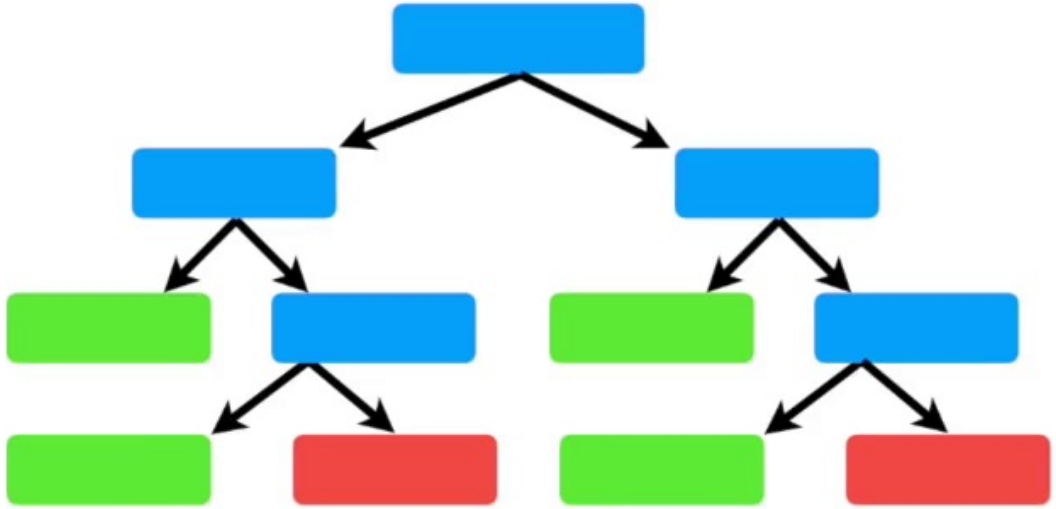
Machine Learning is all about making ***predictions*** and ***classifications***.



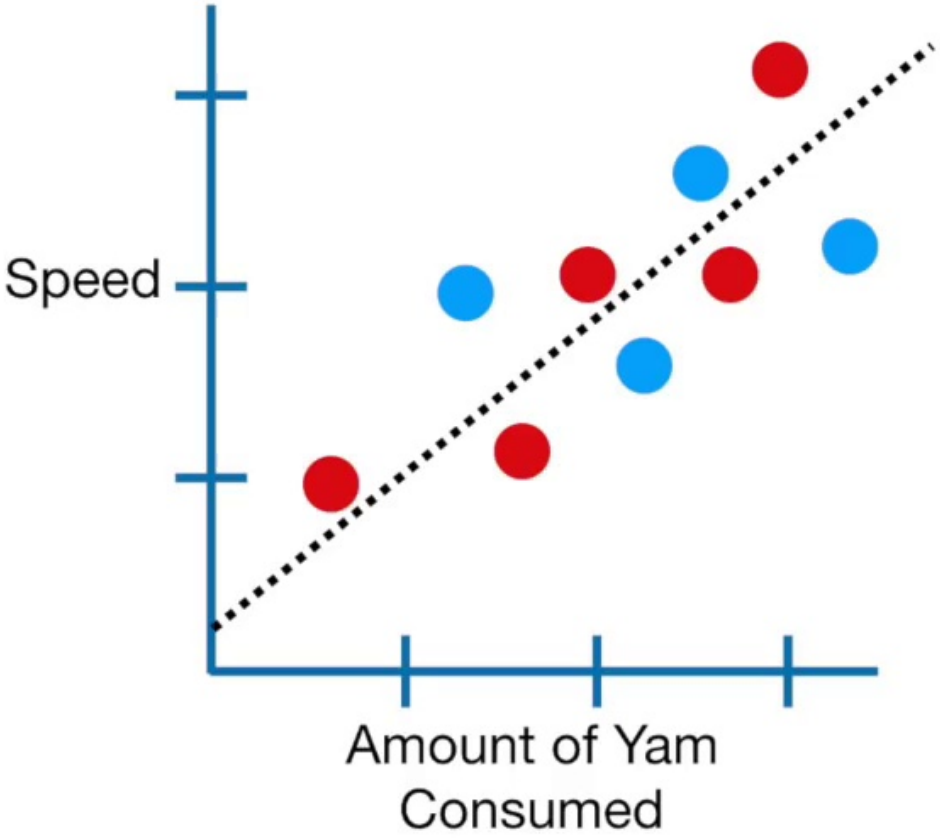


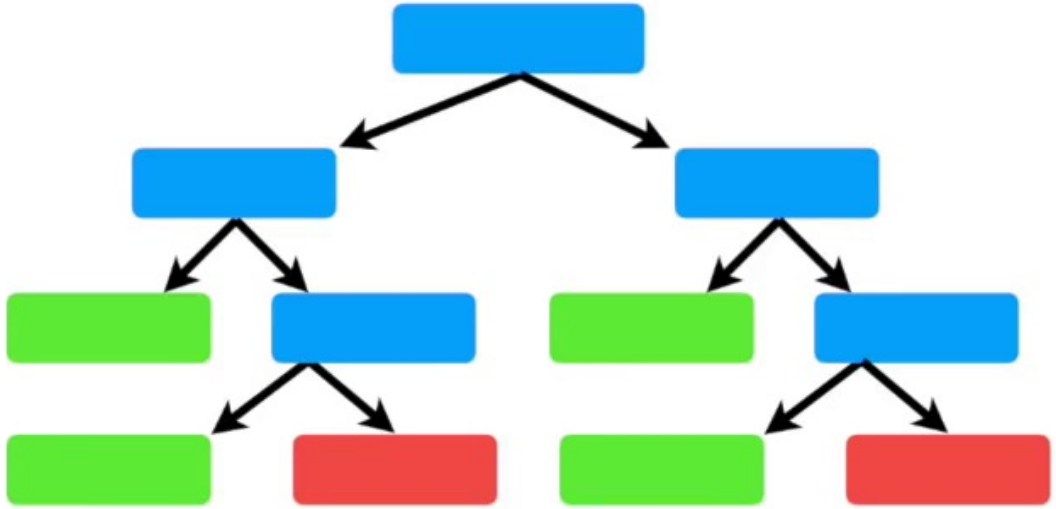
There are tons of fancy **Machine Learning** methods, but the most important thing to know about them isn't what makes them so fancy…



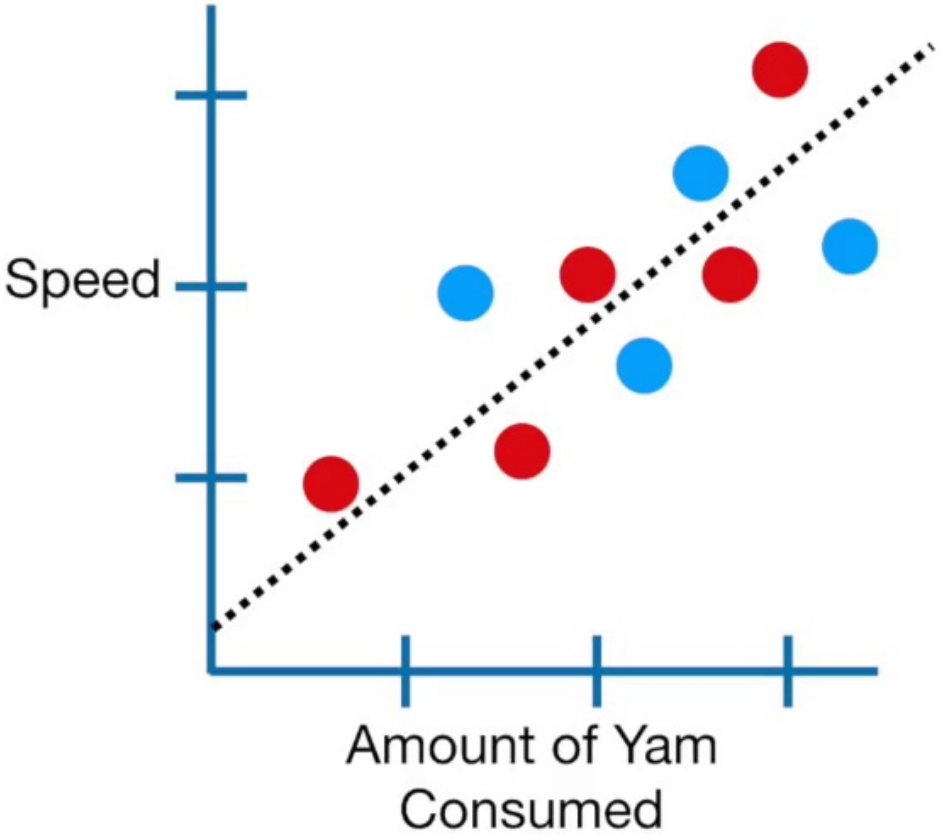


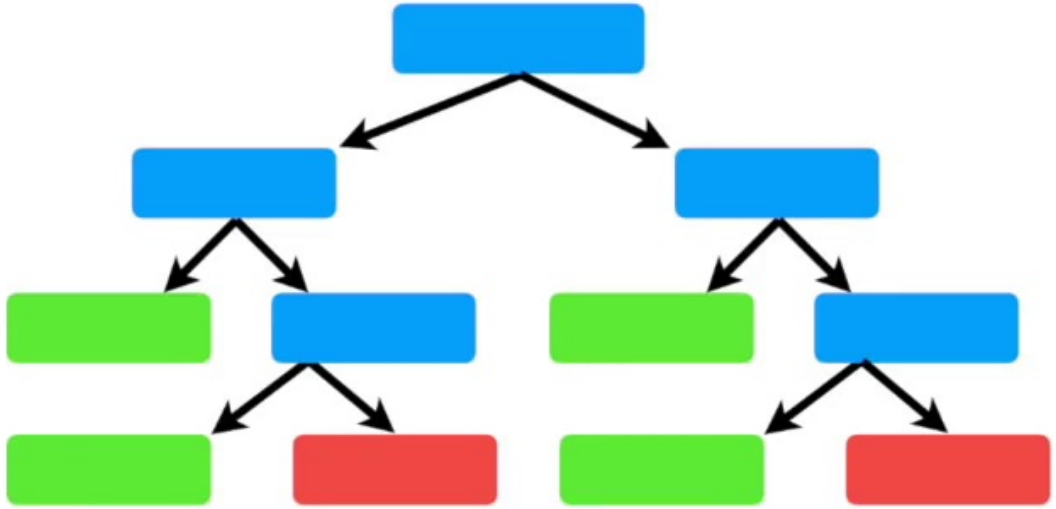
There are tons of fancy **Machine Learning** methods, but the most important thing to know about them isn't what makes them so fancy…

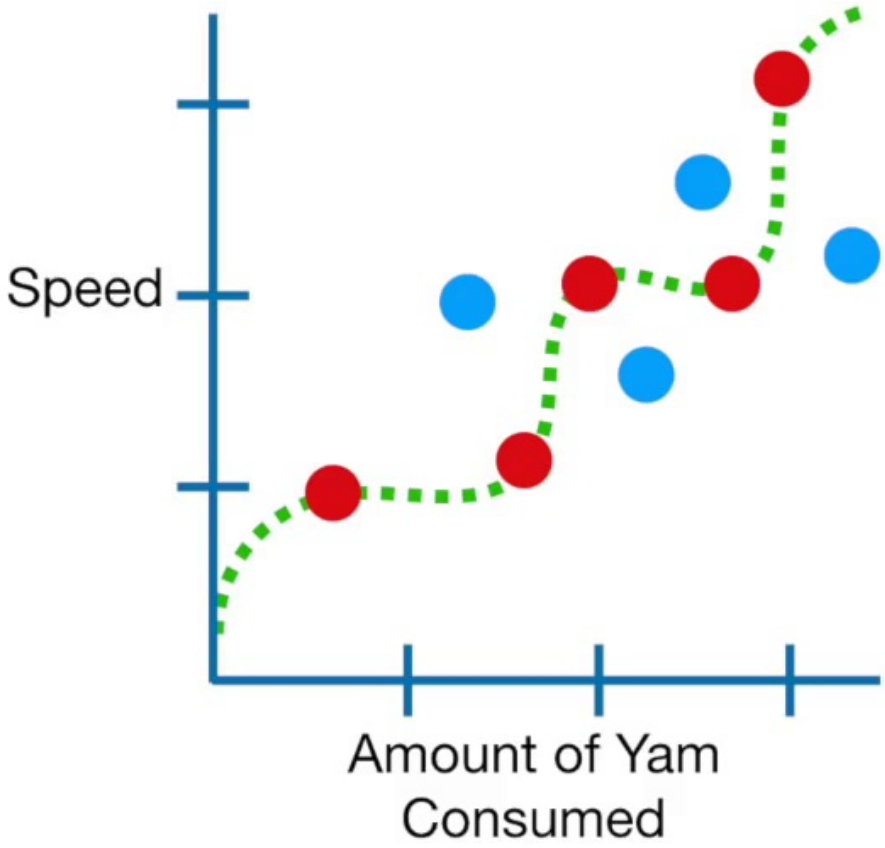




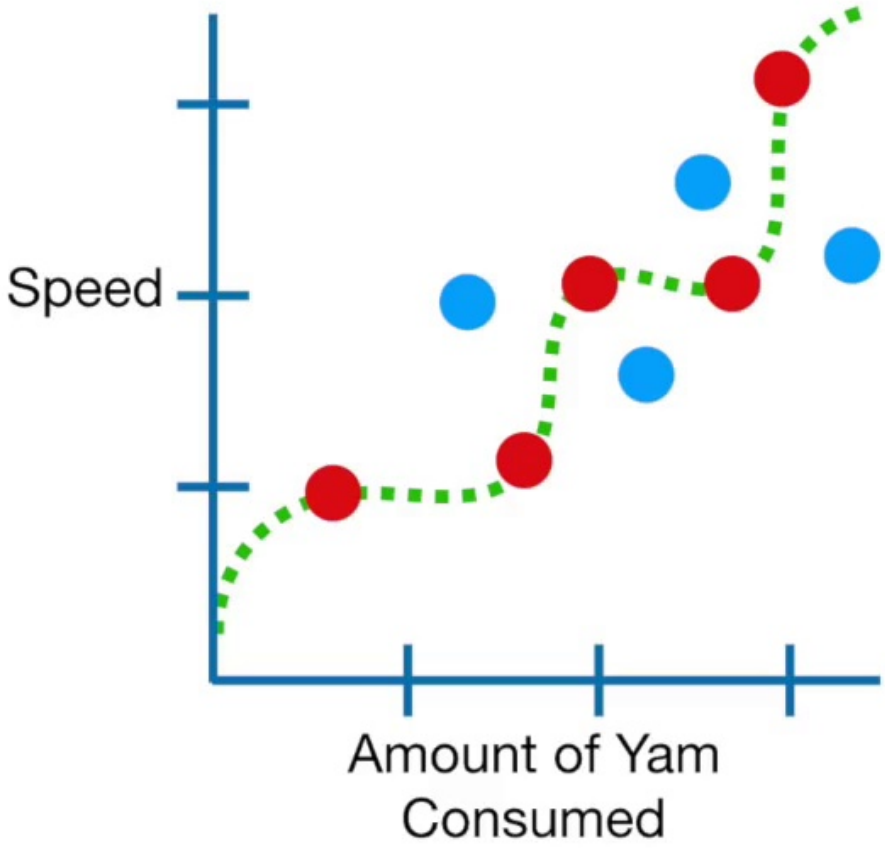
…it's that we decide which method fits
our needs the best by using
Testing Data.



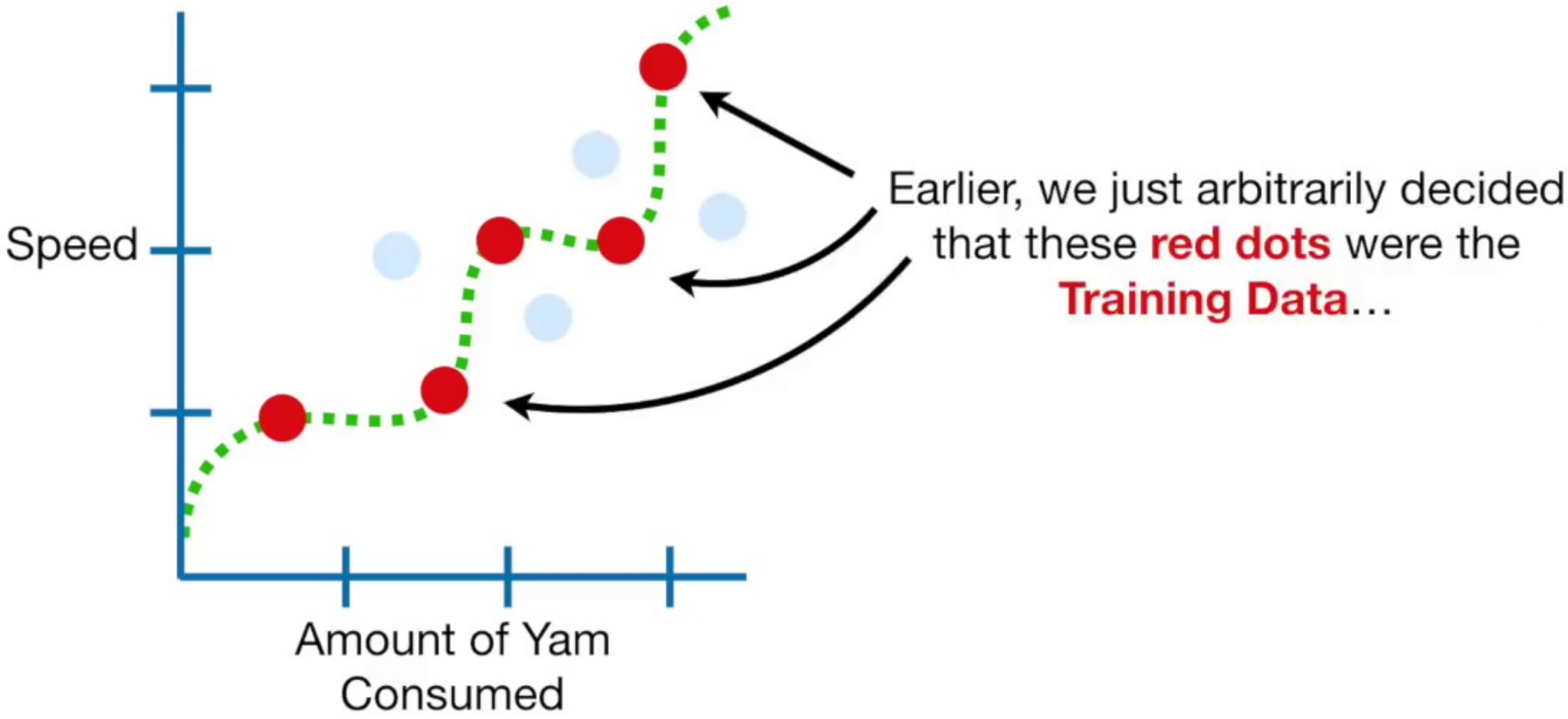


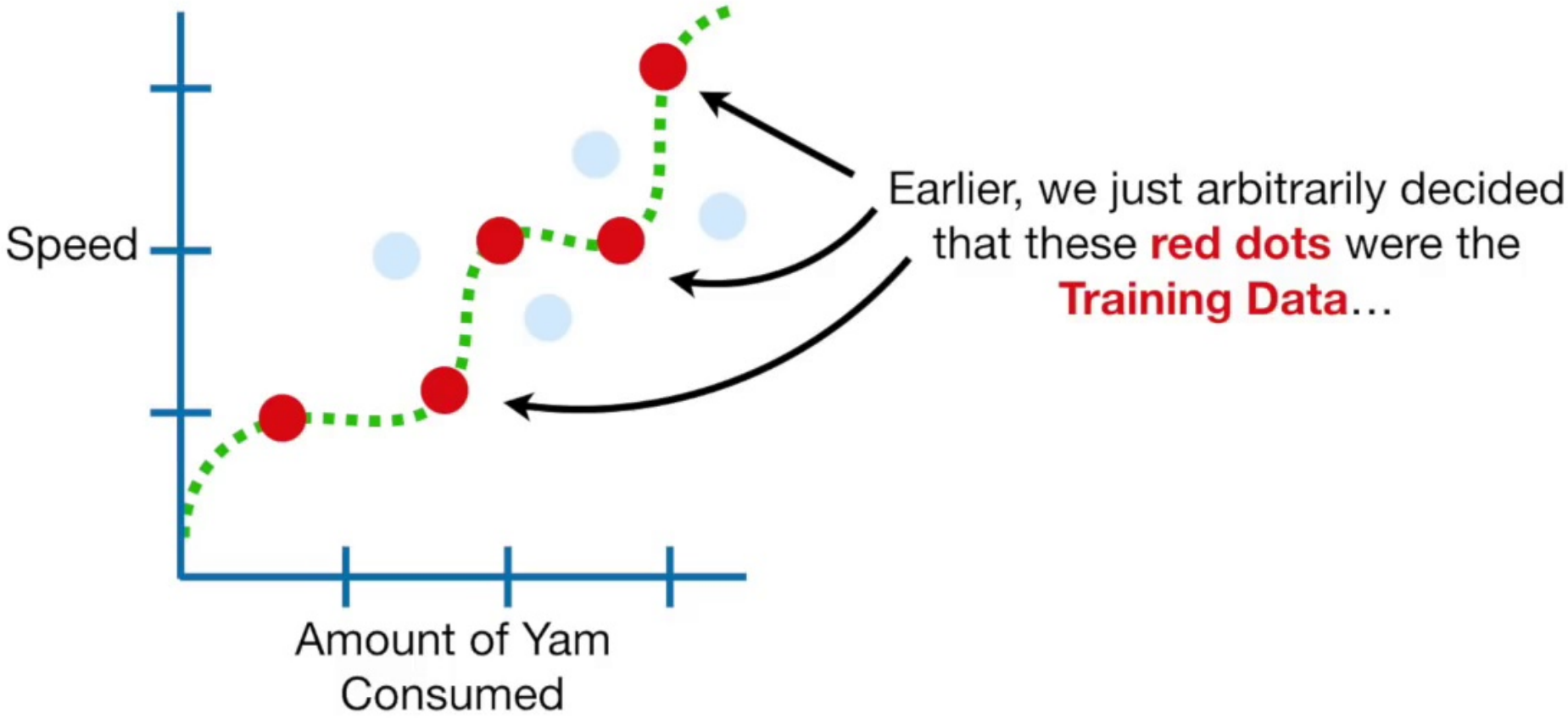


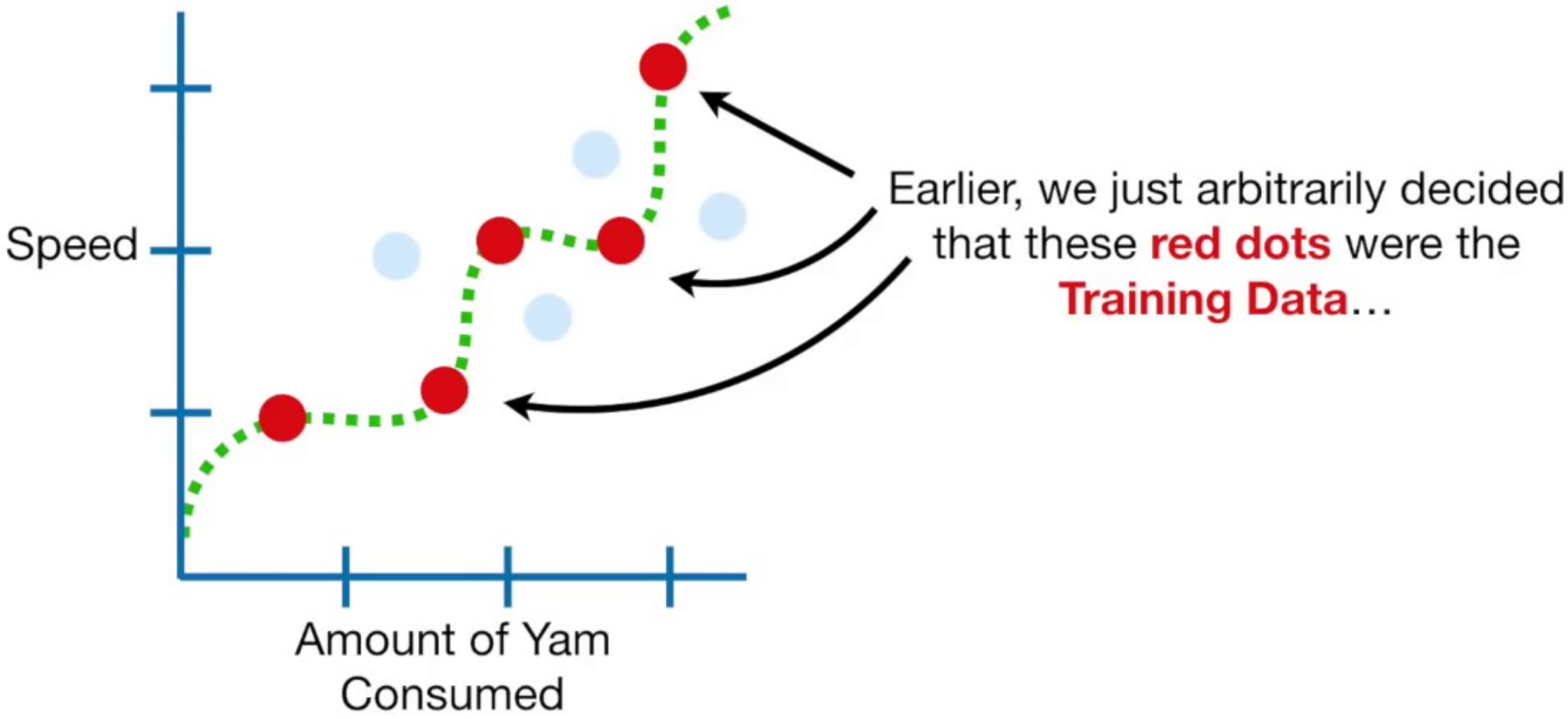
You may be wondering how we decide which data go into the **Training Set** and which data go into the **Testing Set**.

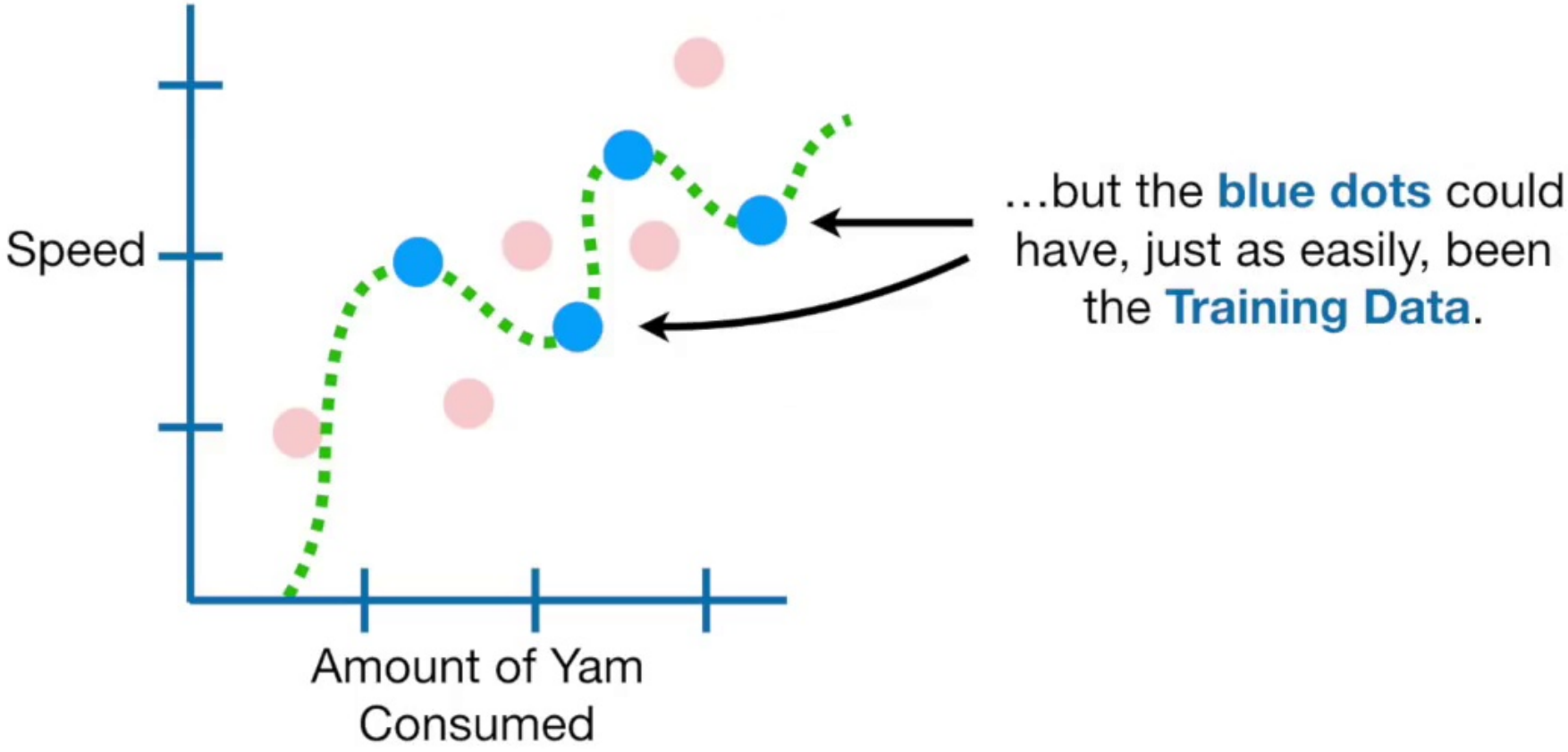


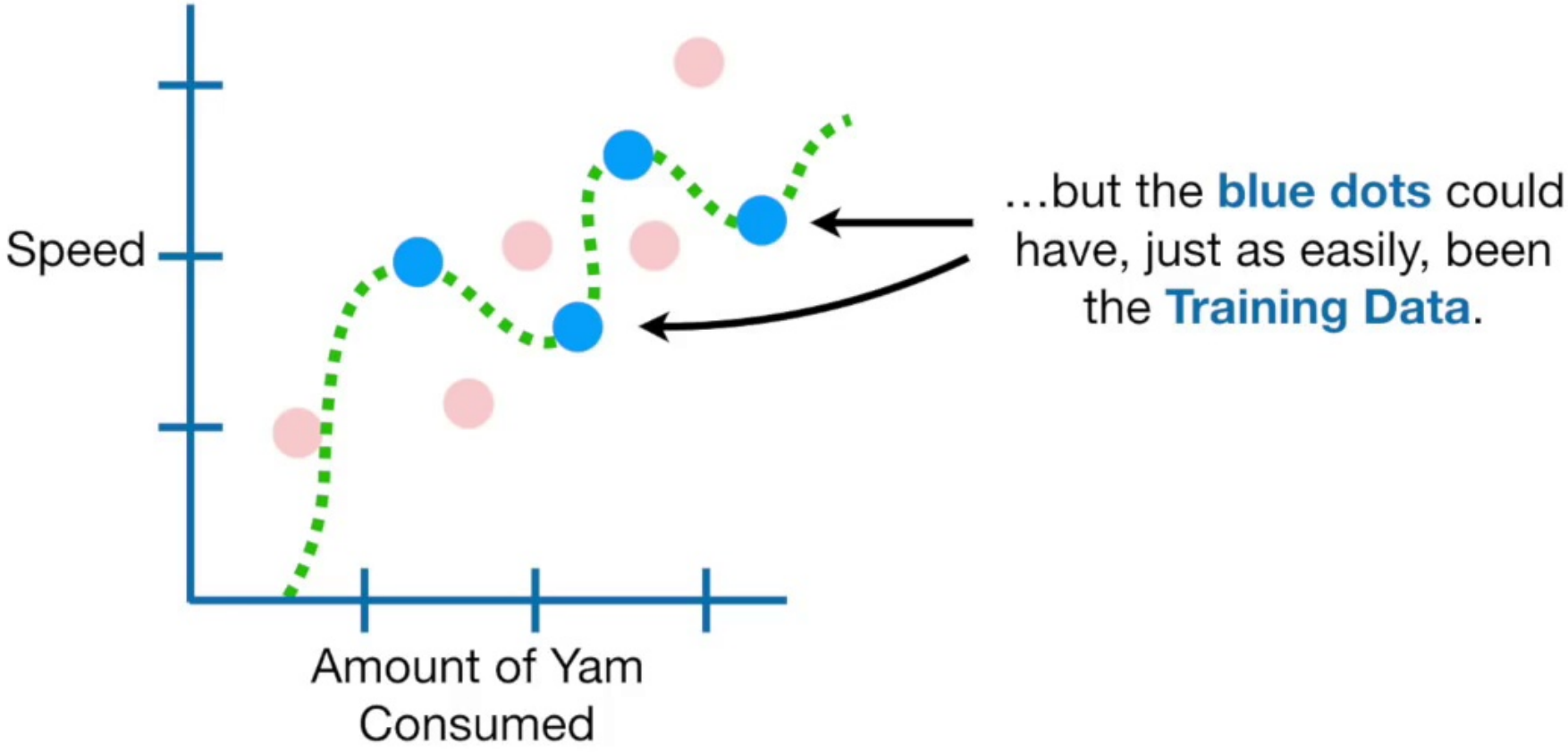
You may be wondering how we decide which data go into the **Training Set** and which data go into the **Testing Set**.

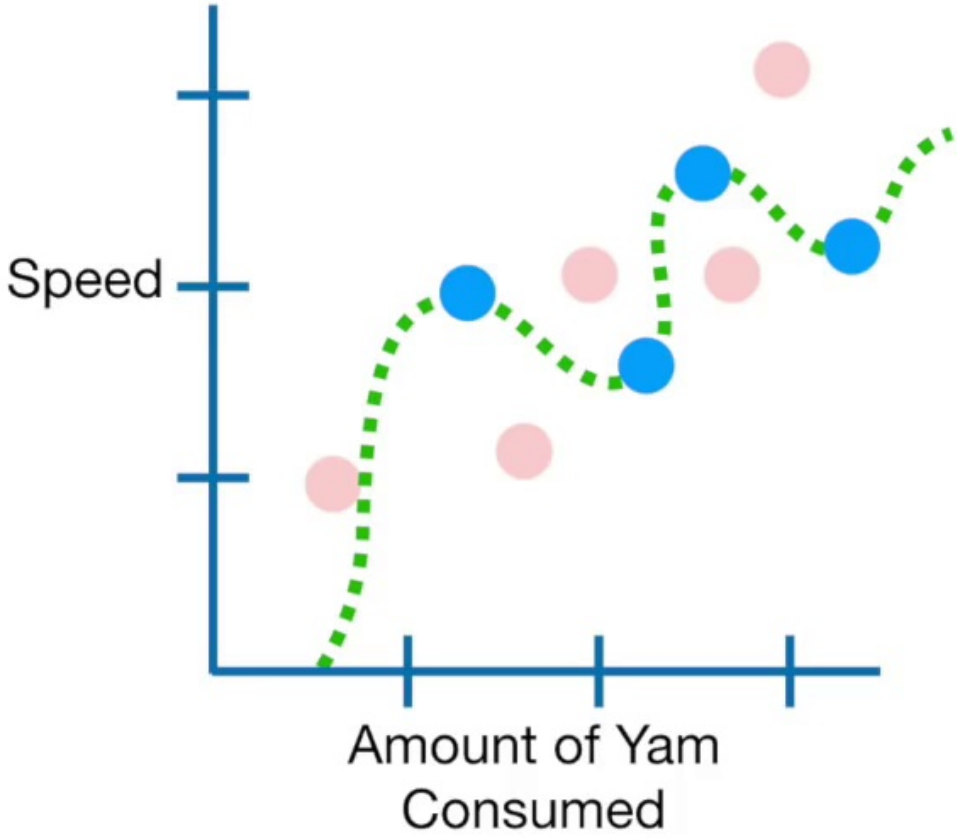

Speed
Amount of Yam Consumed
Earlier, we just arbitrarily decided that these **red dots** were the **Training Data**…


Speed
Amount of Yam Consumed
Earlier, we just arbitrarily decided that these **red dots** were the **Training Data**…

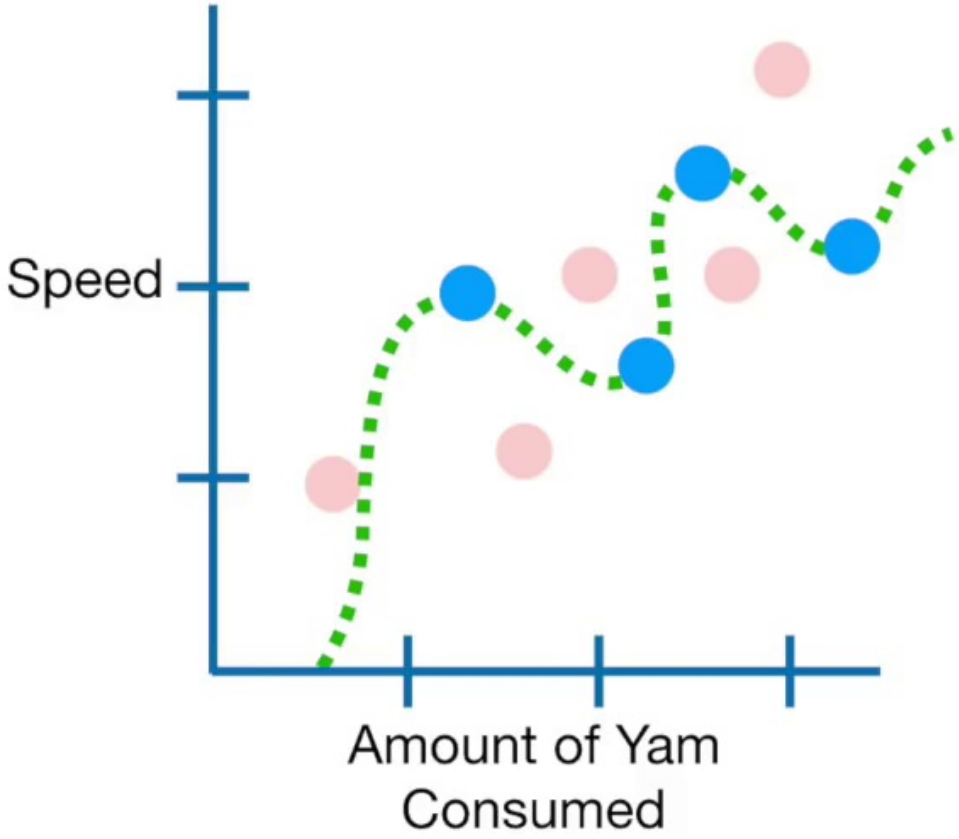

Speed
Amount of Yam Consumed
Earlier, we just arbitrarily decided that these **red dots** were the **Training Data**…


Speed
Amount of Yam Consumed
…but the **blue dots** could have, just as easily, been the **Training Data**.

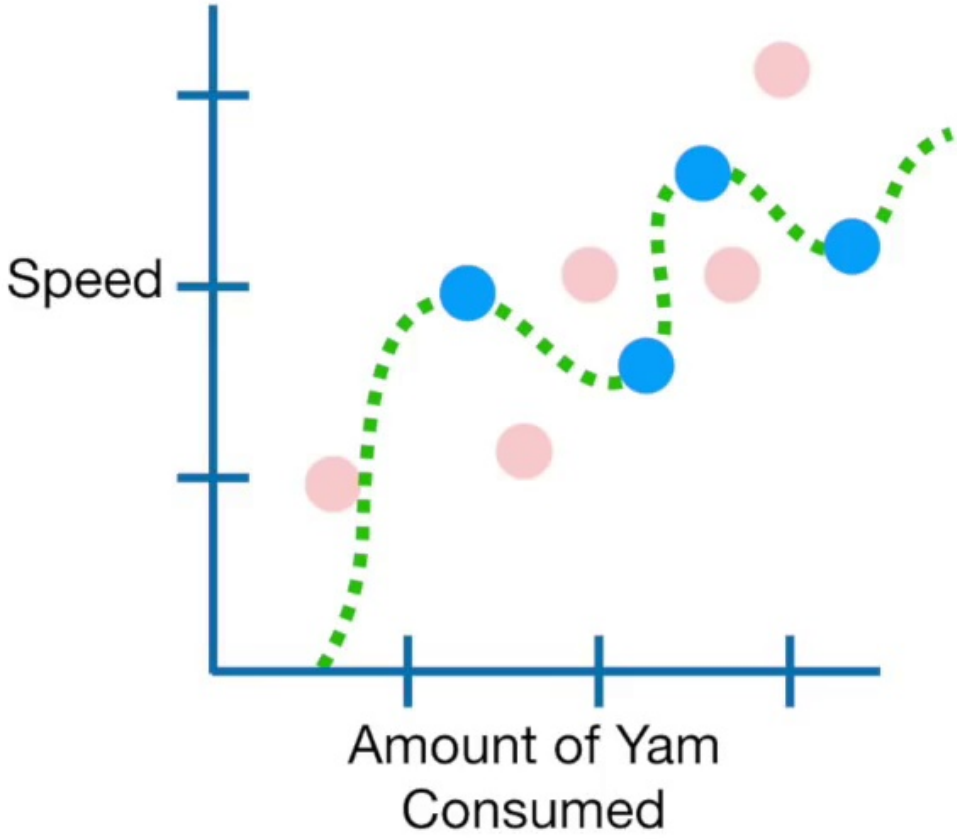

Speed
Amount of Yam Consumed
…but the **blue dots** could have, just as easily, been the **Training Data**.



The good news is that there are ways to determine which samples should be used for **Training Data** and which samples should be used for **Testing Data**…



The good news is that there are ways to determine which samples should be used for **Training Data** and which samples should be used for **Testing Data**…

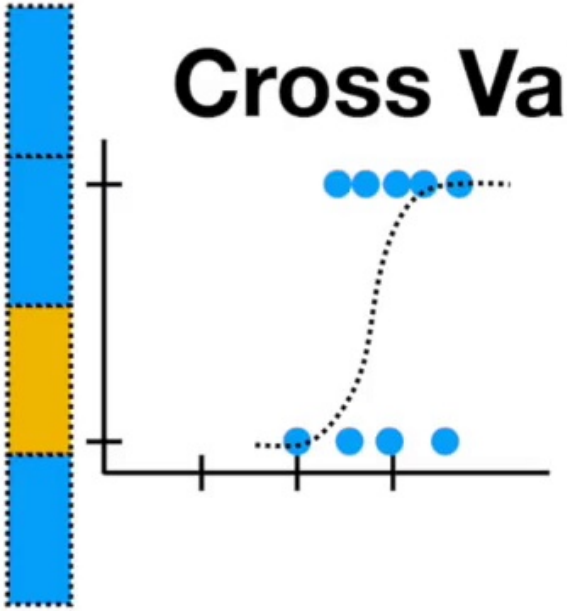


The good news is that there are ways to determine which samples should be used for **Training Data** and which samples should be used for **Testing Data**…

…and if you're interested in learning more about this, check out the StatQuest!

Shameless Self Promotion!!!!

Shameless Self Promotion!!!!



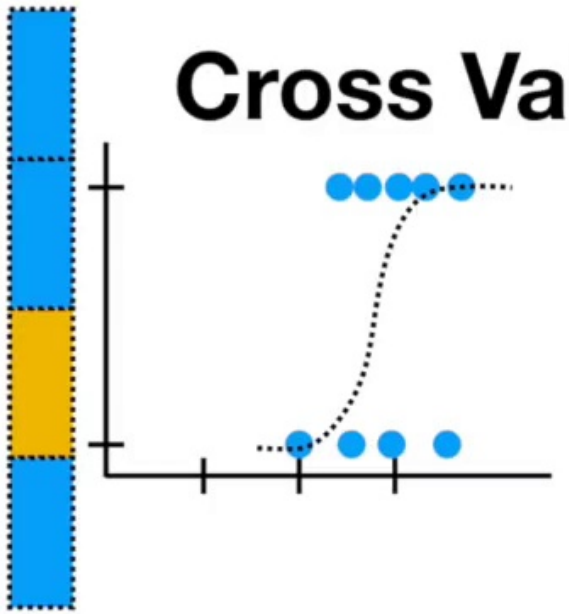
Cross Validation….

…it's no big deal!!!

…and if you're interested in learning more about this, check out the StatQuest!

Shameless Self Promotion!!!!

Shameless Self Promotion!!!!



Cross Validation….

…it's no big deal!!!

…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

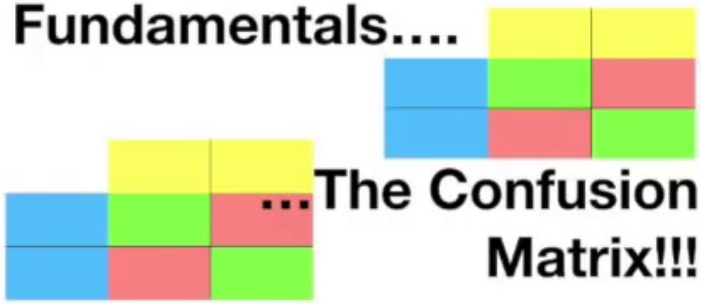
…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

Machine Learning Fundamentals….

…The Confusion Matrix!!!

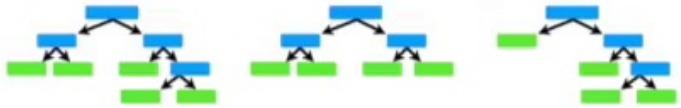
…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

Machine Learning Fundamentals….



…The Confusion Matrix!!!

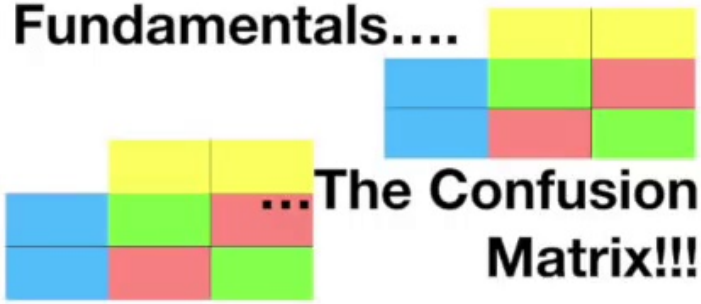
Random Forests Part 1…



Building, using and evaluating, clearly explained!!!

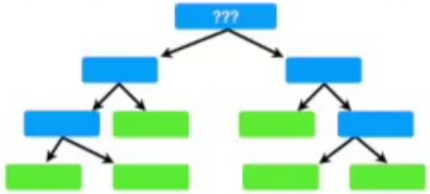
…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

Machine Learning Fundamentals….



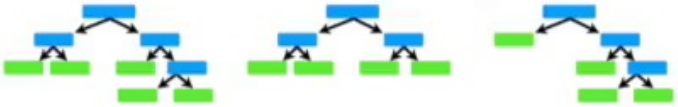
…The Confusion Matrix!!!

Decision Trees…



…clearly explained!

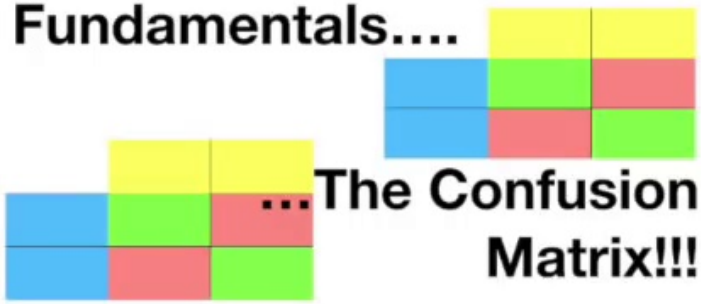
Random Forests Part 1…



Building, using and evaluating, clearly explained!!!

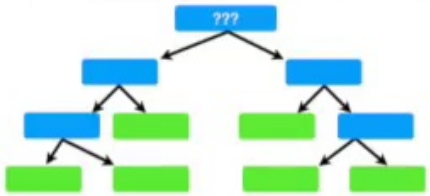
…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

Machine Learning Fundamentals….



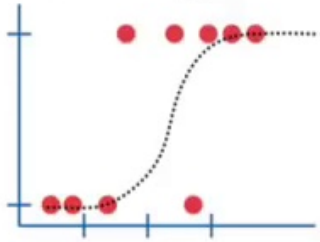
…The Confusion Matrix!!!

Decision Trees…



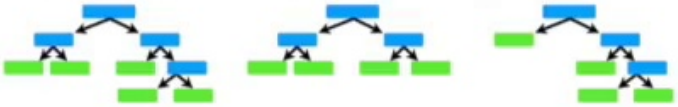
…clearly explained!

Logistic Regression…



…Clearly Explained!!!

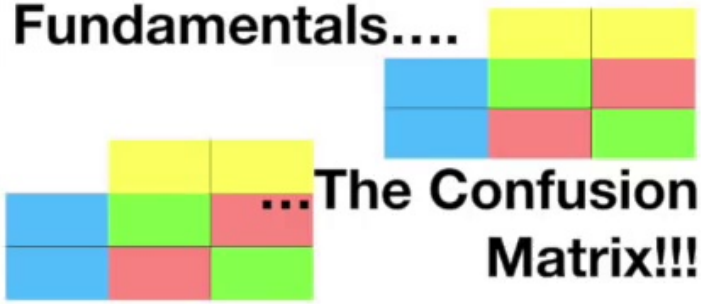
Random Forests Part 1…



Building, using and evaluating, clearly explained!!!

…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

Machine Learning Fundamentals….

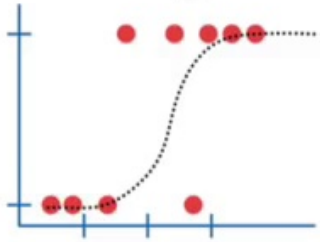


…The Confusion Matrix!!!

Decision Trees…

…clearly explained!

Logistic Regression…



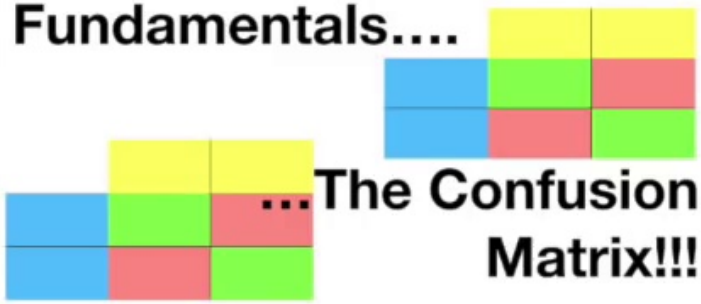
…Clearly Explained!!!

Random Forests Part 1…

Building, using and evaluating, clearly explained!!!

…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

Machine Learning Fundamentals….

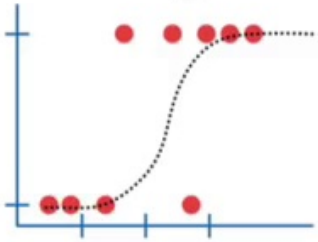


…The Confusion Matrix!!!

Decision Trees…

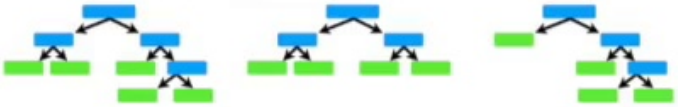
…clearly explained!

Logistic Regression…



…Clearly Explained!!!

Random Forests Part 1…



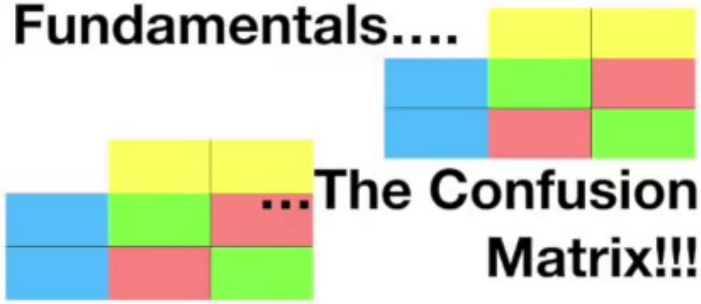
Building, using and evaluating, clearly explained!!!

Machine Learning Fundamentals….

…Sensitivity and Specificity!!!

…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

Machine Learning Fundamentals….

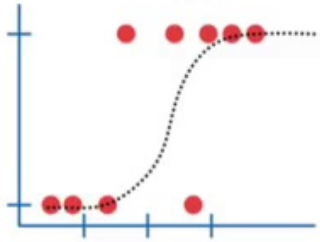


…The Confusion Matrix!!!

Decision Trees…

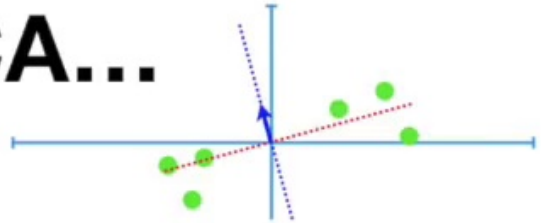
…clearly explained!

Logistic Regression…



…Clearly Explained!!!

PCA…

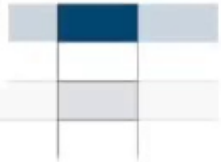


Step-by-Step!!!

Random Forests Part 1…

Building, using and evaluating, clearly explained!!!

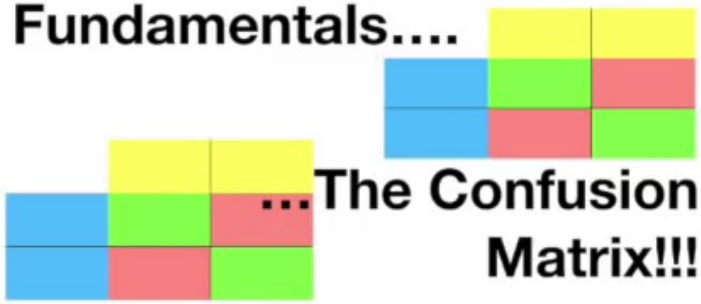
Machine Learning Fundamentals….



…Sensitivity and Specificity!!!

…and are lots more StatQuests that walk you through **Machine Learning** concepts, step-by-step, so check them out!

Machine Learning Fundamentals….

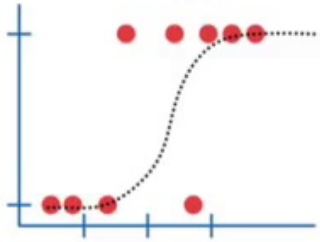


…The Confusion Matrix!!!

Decision Trees…

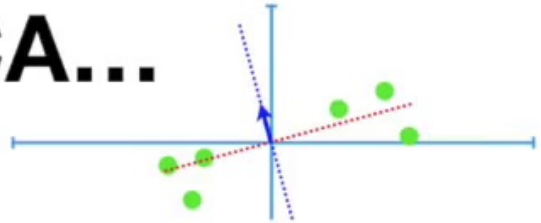
…clearly explained!

Logistic Regression…



…Clearly Explained!!!

PCA…

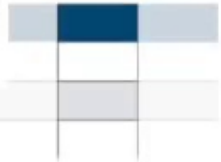


Step-by-Step!!!

Random Forests Part 1…

Building, using and evaluating, clearly explained!!!

Machine Learning Fundamentals….



…Sensitivity and Specificity!!!

The End!!!

The End!!!

Subscribe!!!

Support StatQuest!!!