Developer Student Clubs

Tree Based Regressions

Decision Trees & Random Forest Regression

Muhammad Huzaifa Shahbaz
DSC Lead @ Google Developers,
Machine Learning Engineer
@mhuzaifadev

Credits:
StatQuest with Josh Starmer

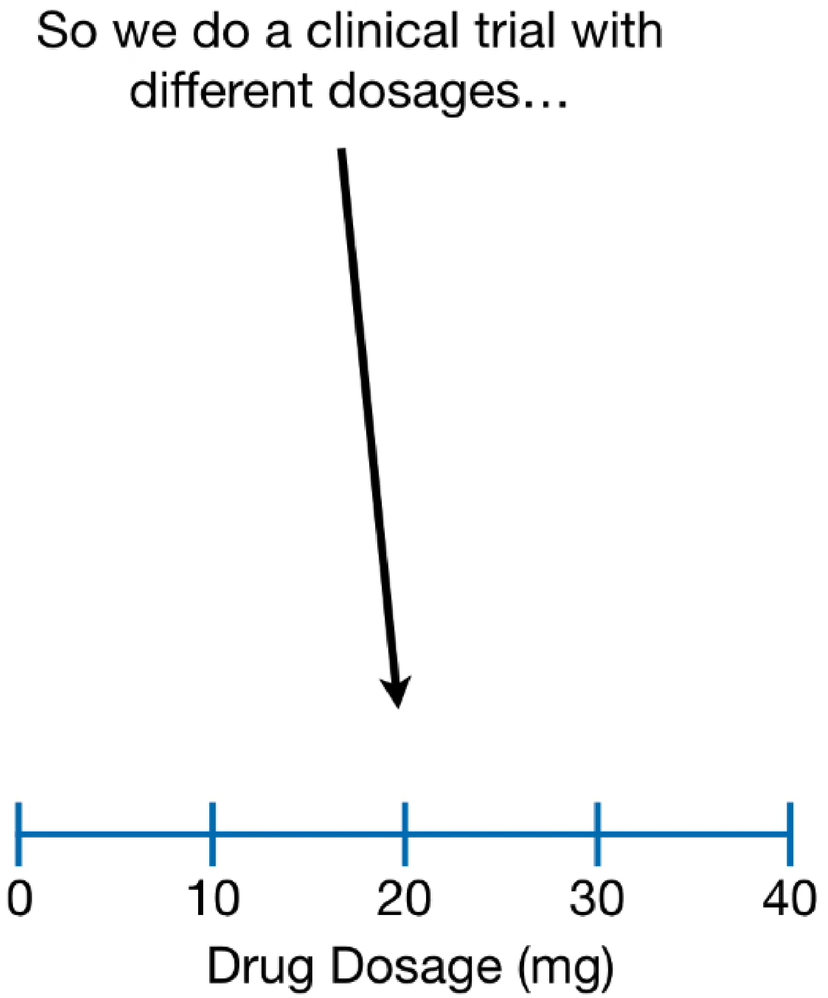
Google Developers

Credits:
StatQuest with Josh Starmer

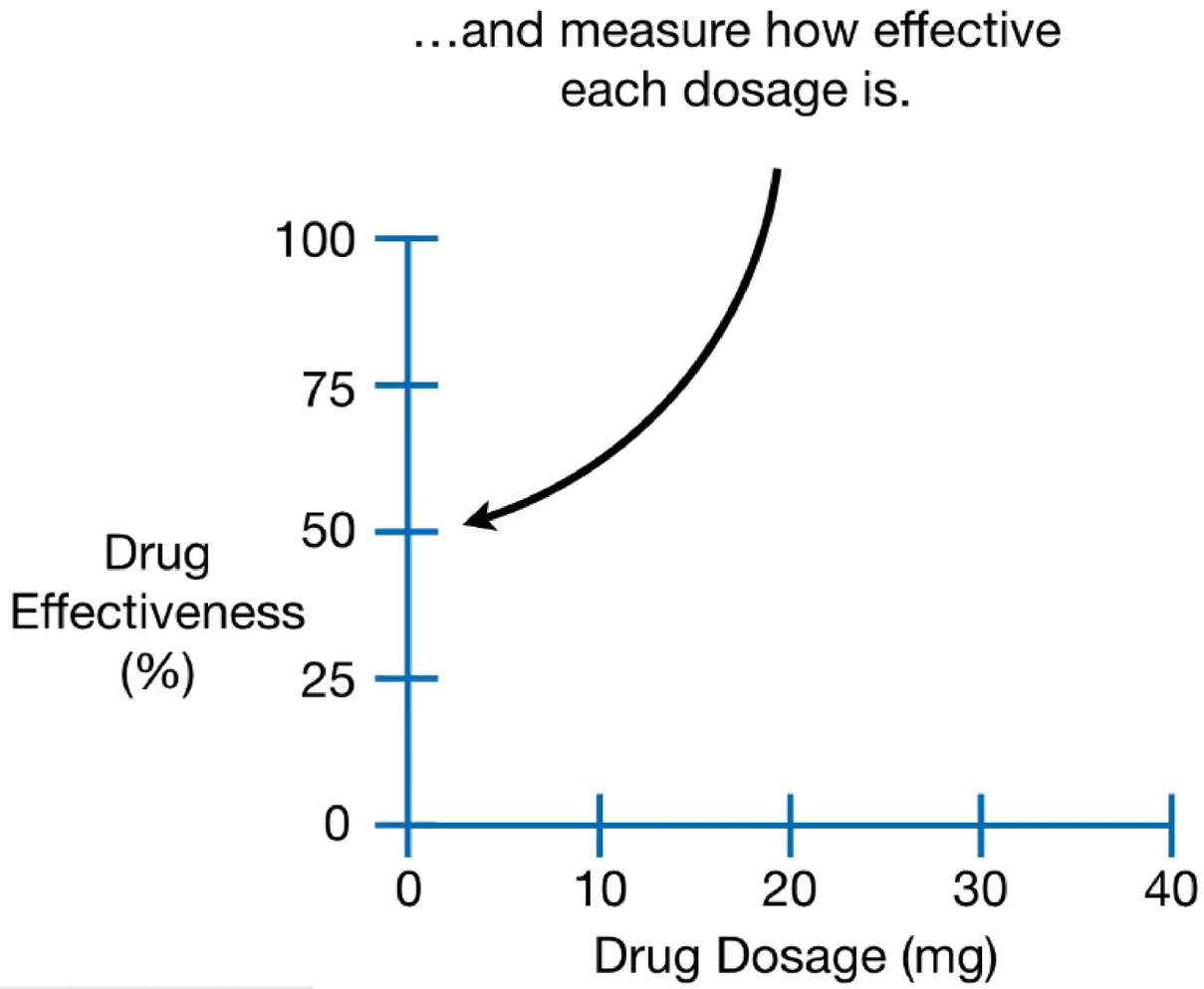
However, we don't know the optimal dosage to give to patients.

vs.

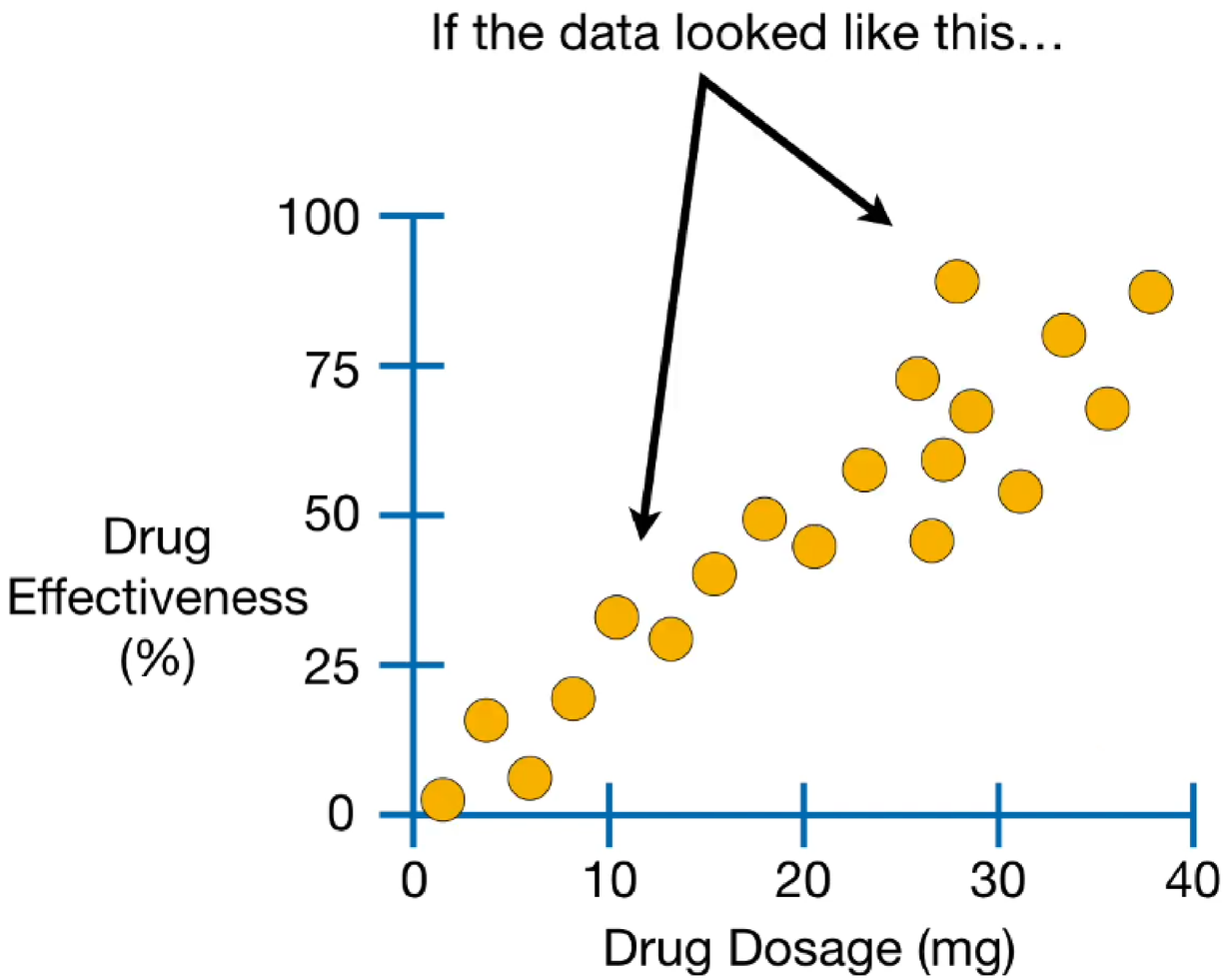
Credits:
StatQuest with Josh Starmer



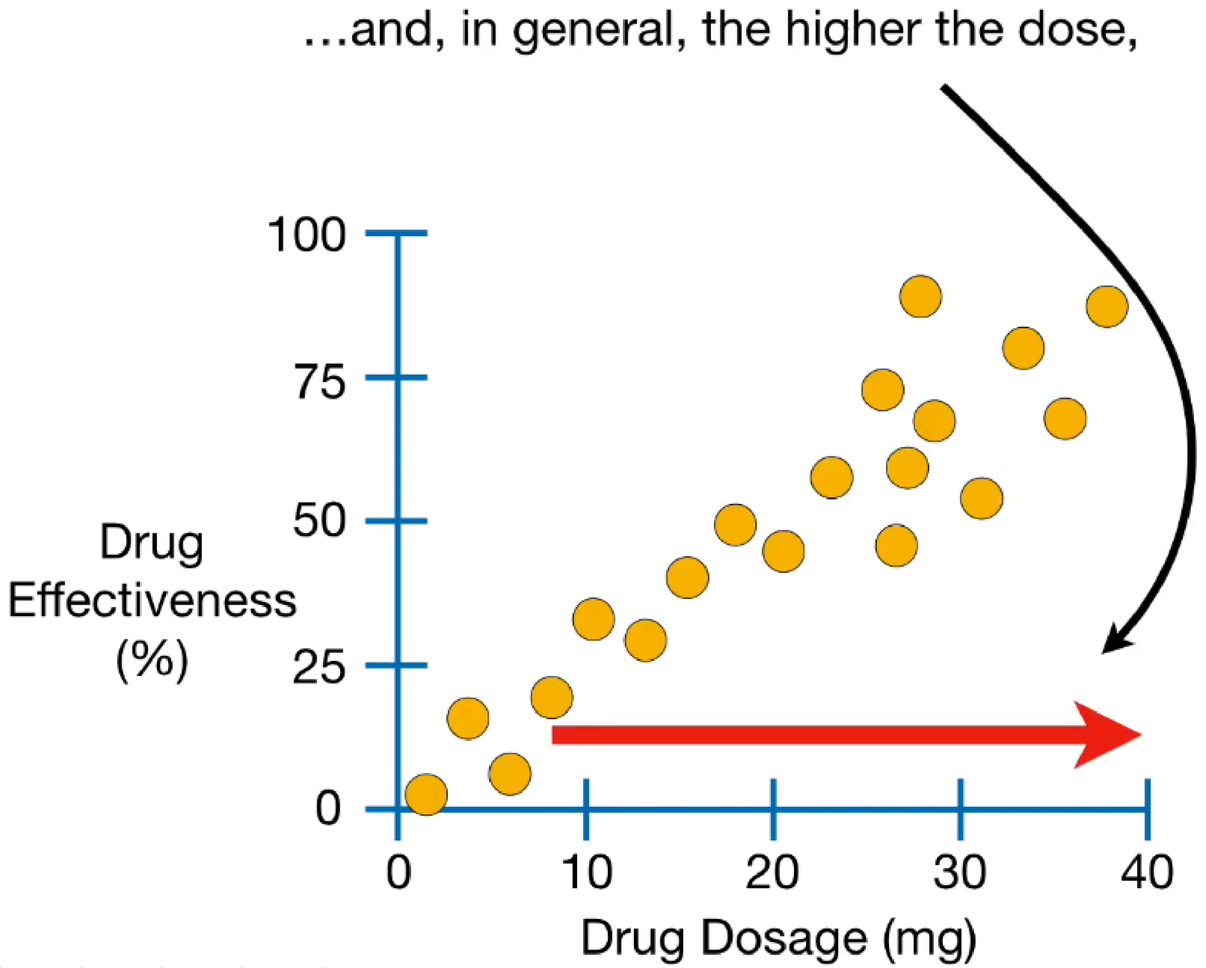
Credits:
StatQuest with Josh Starmer



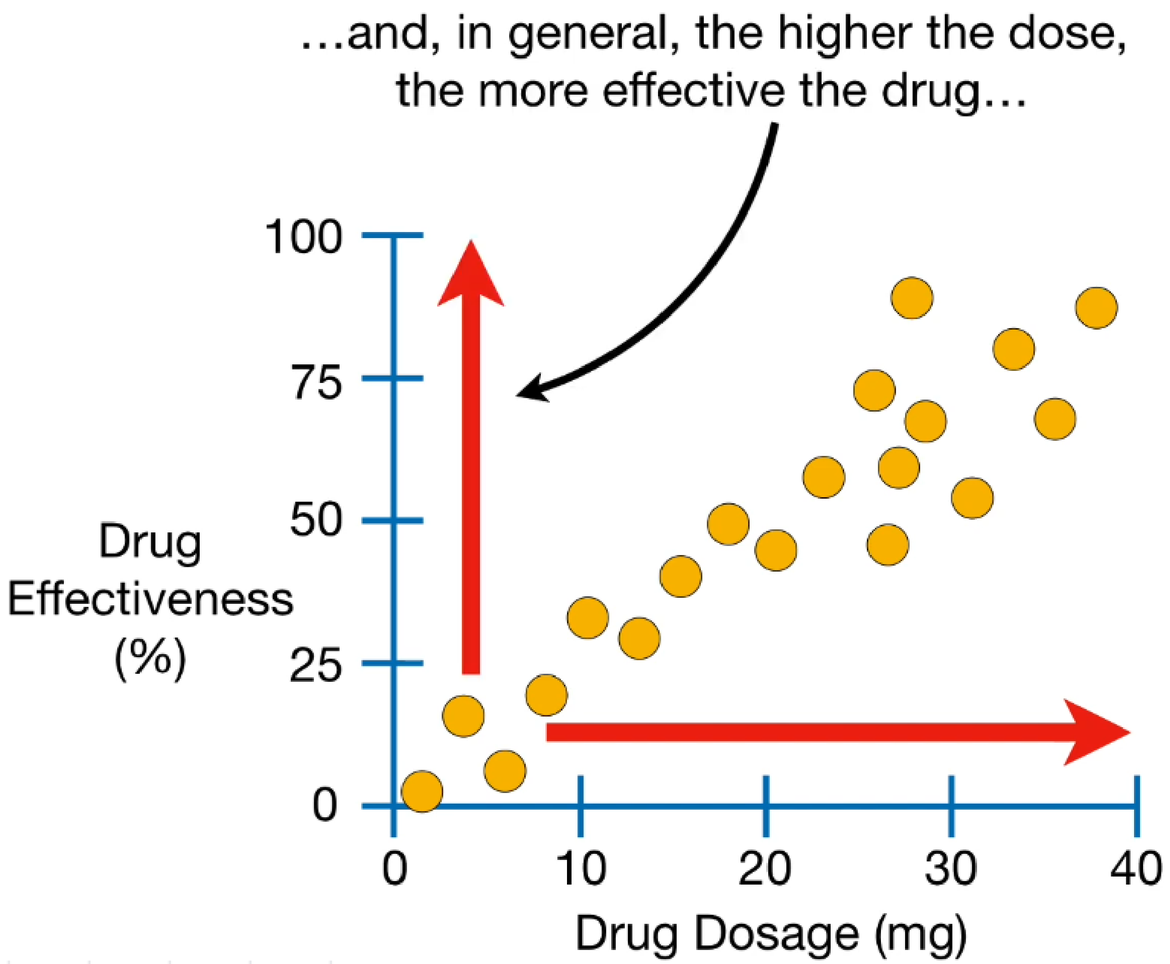
Credits:
StatQuest with Josh Starmer



Credits:
StatQuest with Josh Starmer

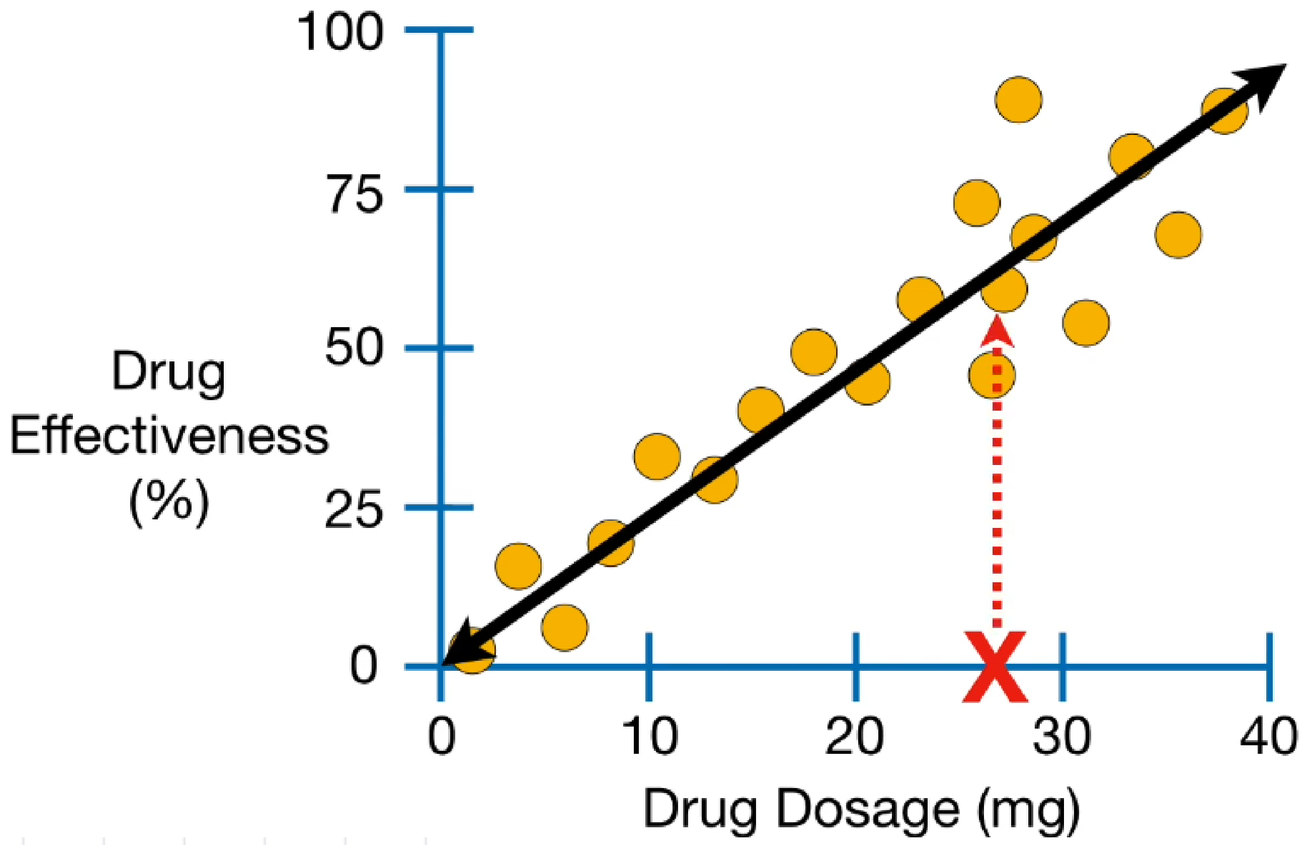


Credits:
StatQuest with Josh Starmer



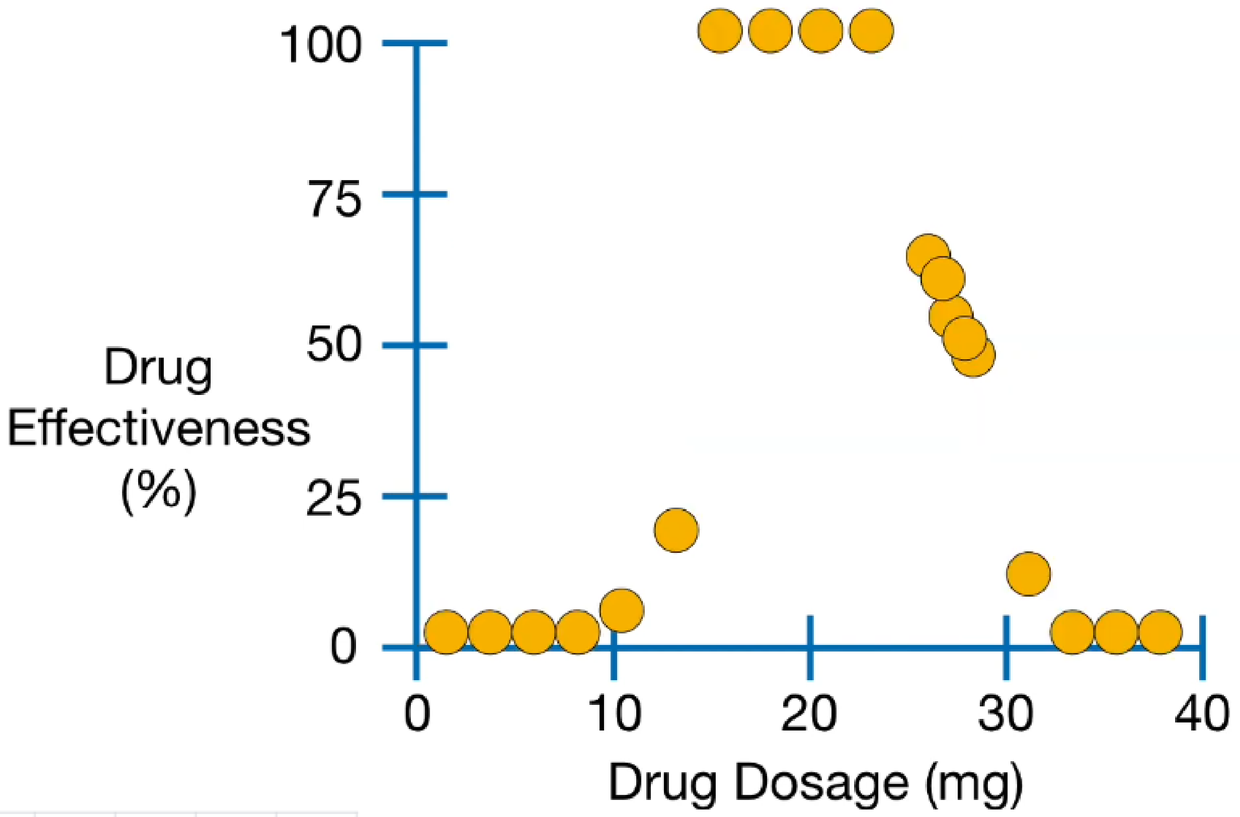
Credits:
StatQuest with Josh Starmer

…we could use the line to predict that a **27 mg Dose** should be **62% Effective**.

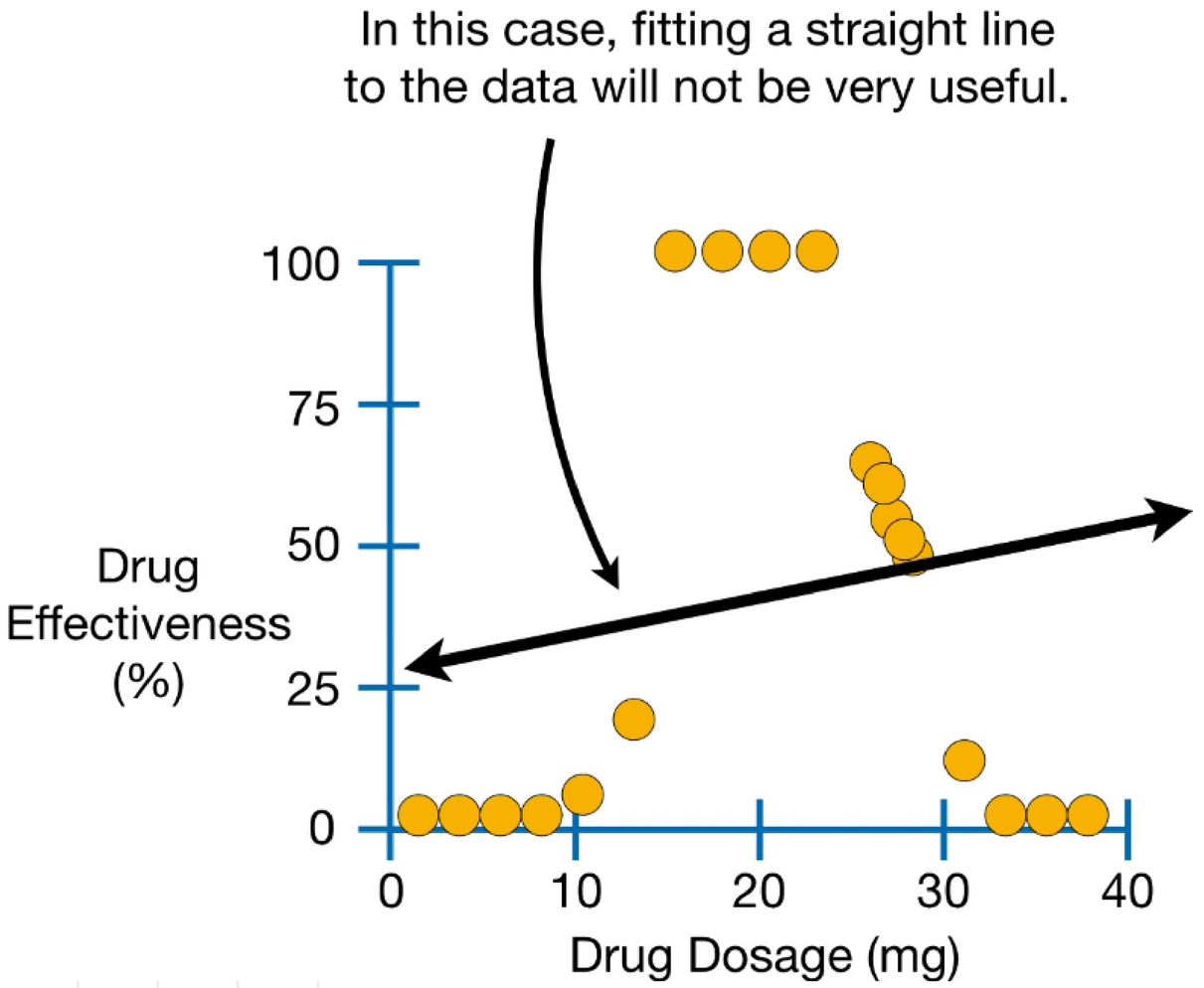


Credits:
StatQuest with Josh Starmer

However, what if the data looked like this?

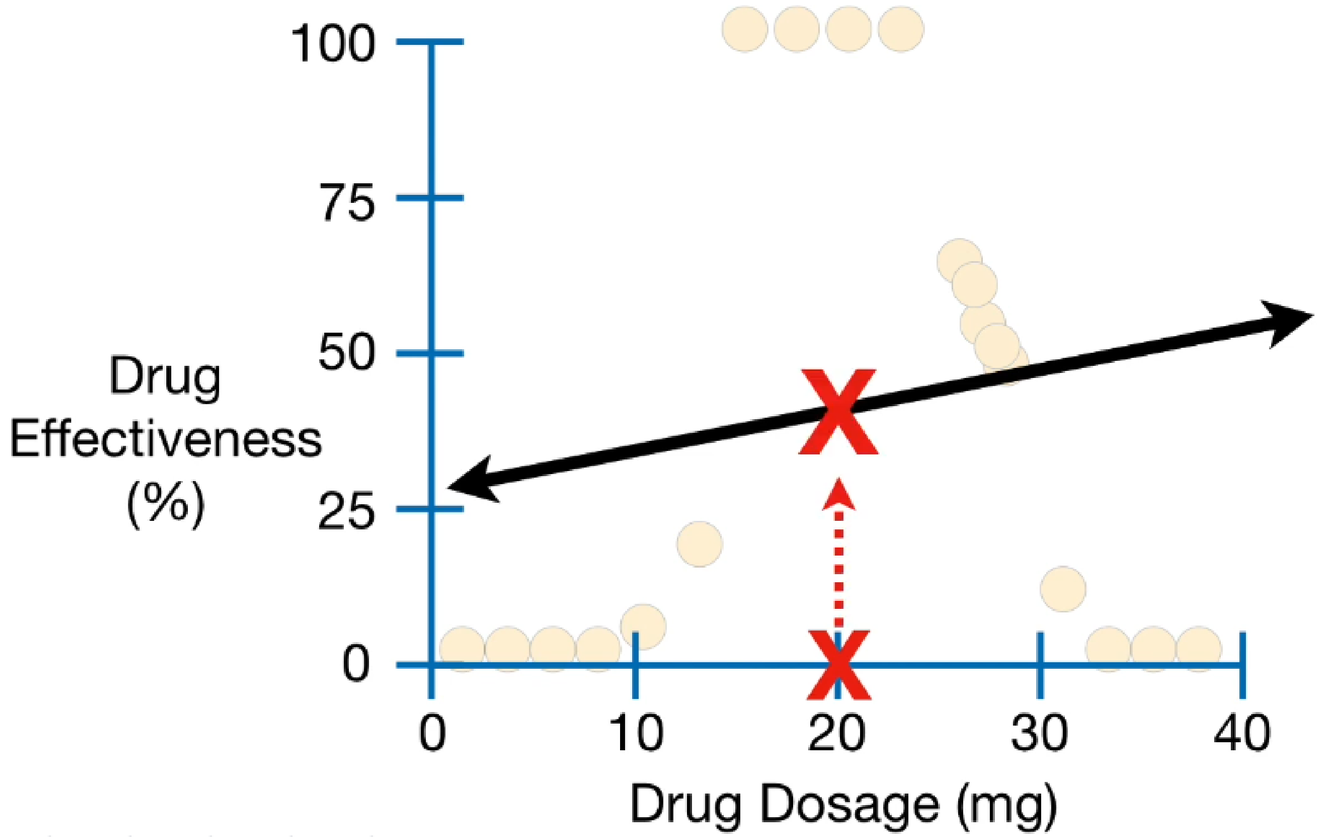


Credits:
StatQuest with Josh Starmer

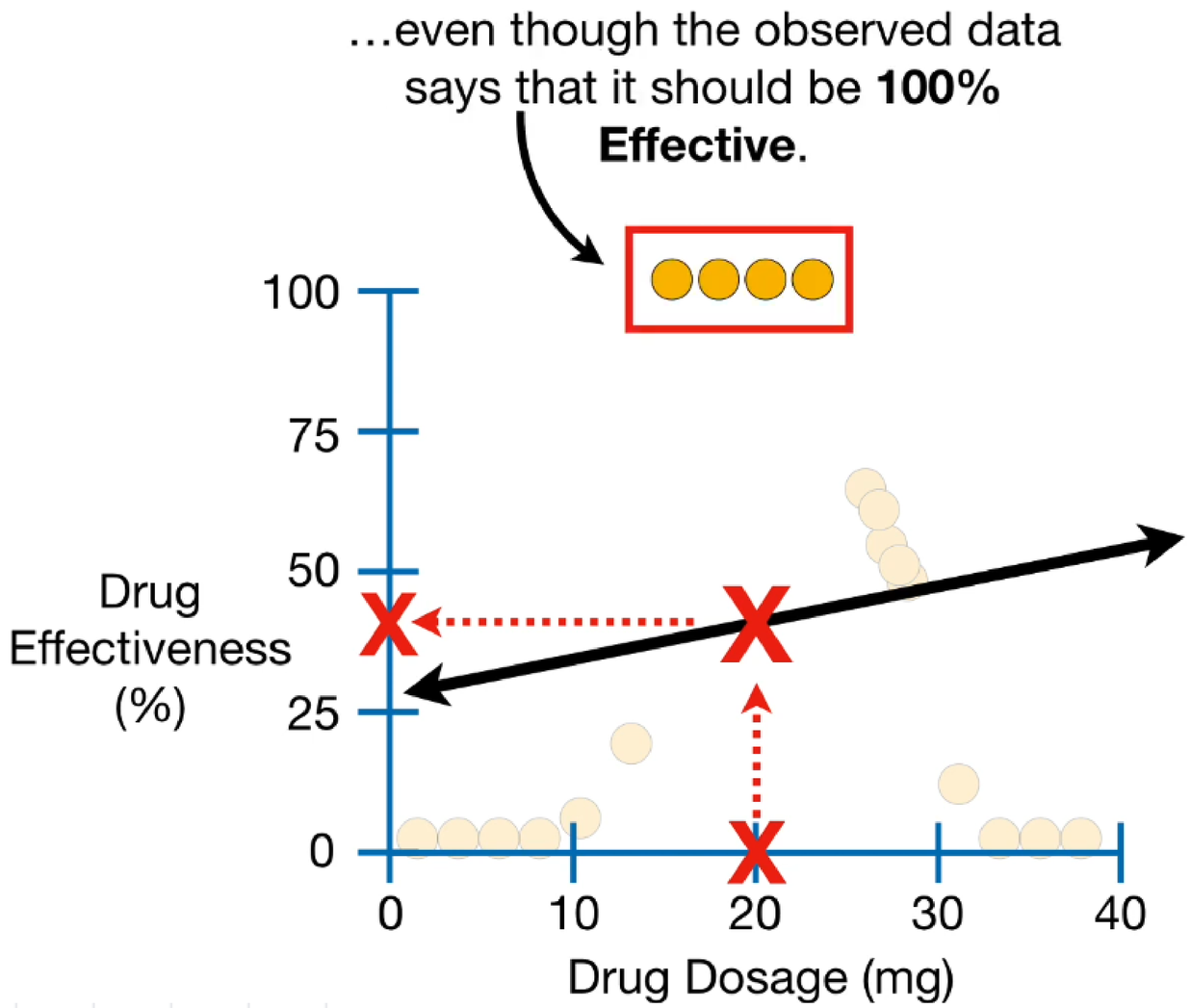


Credits:
StatQuest with Josh Starmer

…then we would predict that a **20 mg Dose** should be **45% Effective**…

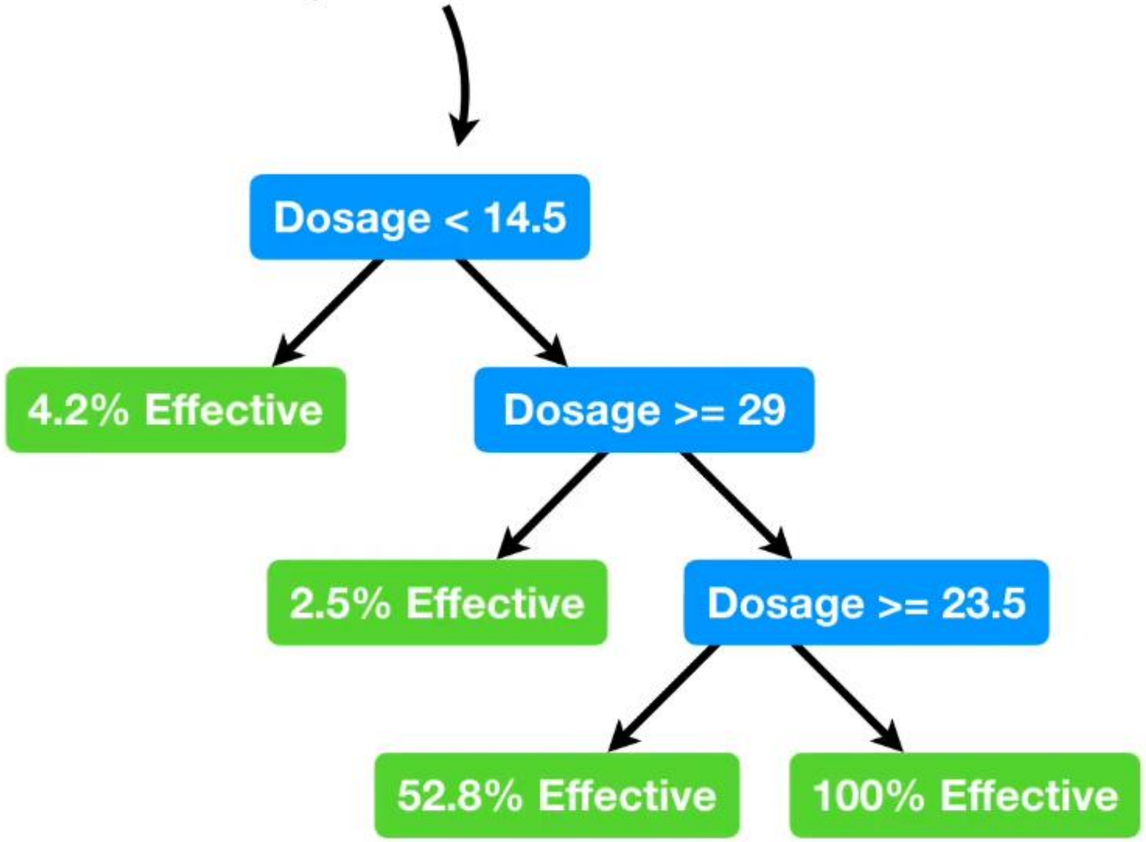


Credits:
StatQuest with Josh Starmer



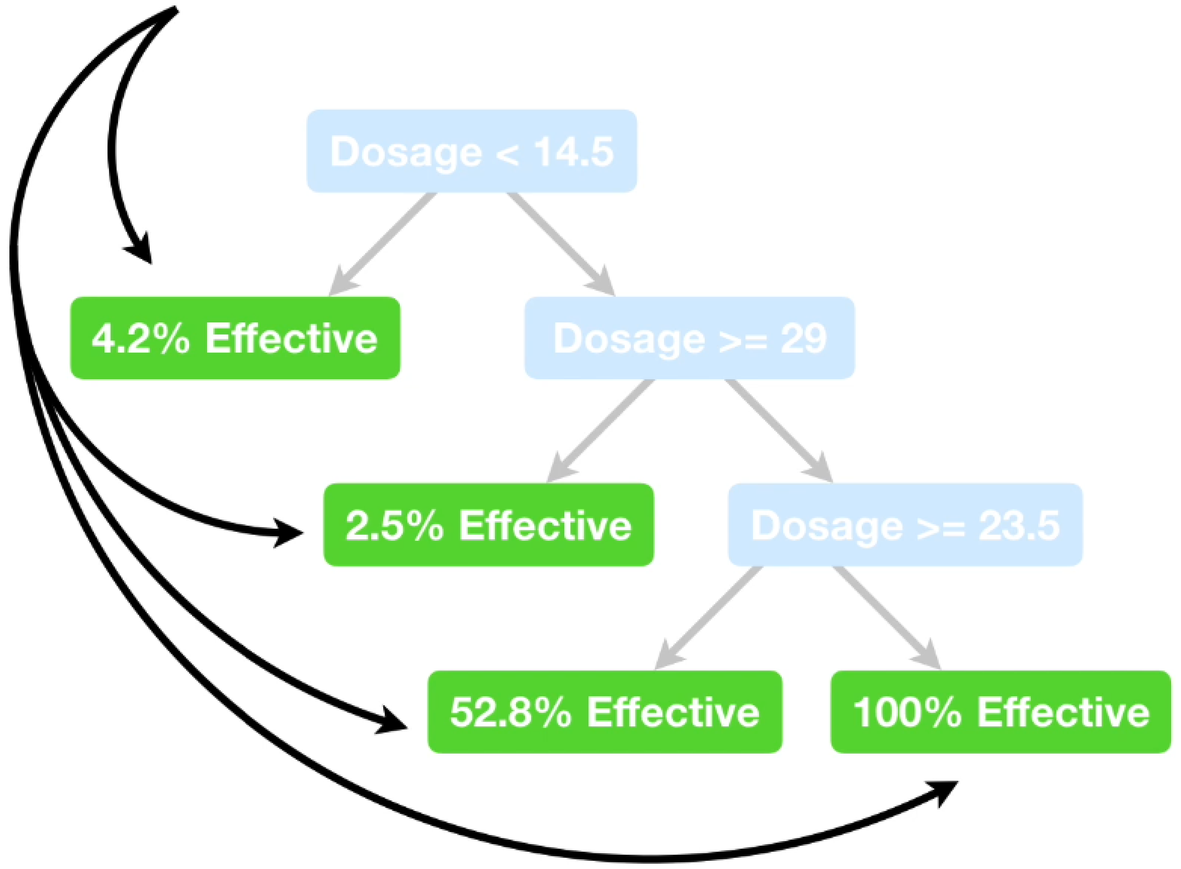
Credits:
StatQuest with Josh Starmer

One option is to use a **Regression Tree**.



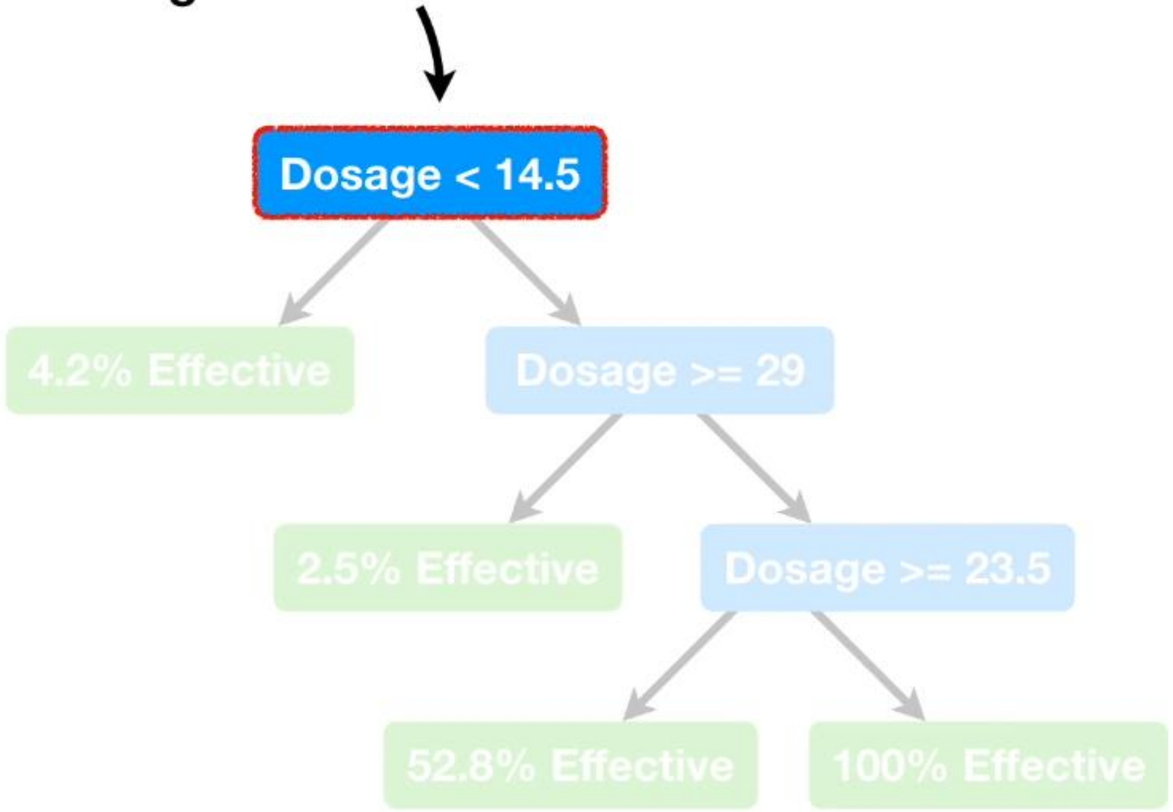
Credits:
StatQuest with Josh Starmer

In a **Regression Tree**, each leaf represents a numeric value.



Credits:
StatQuest with Josh Starmer

With *this* **Regression Tree**, we start by asking if the **Dosage** is less than **14.5**.



Credits:
StatQuest with Josh Starmer

T

S

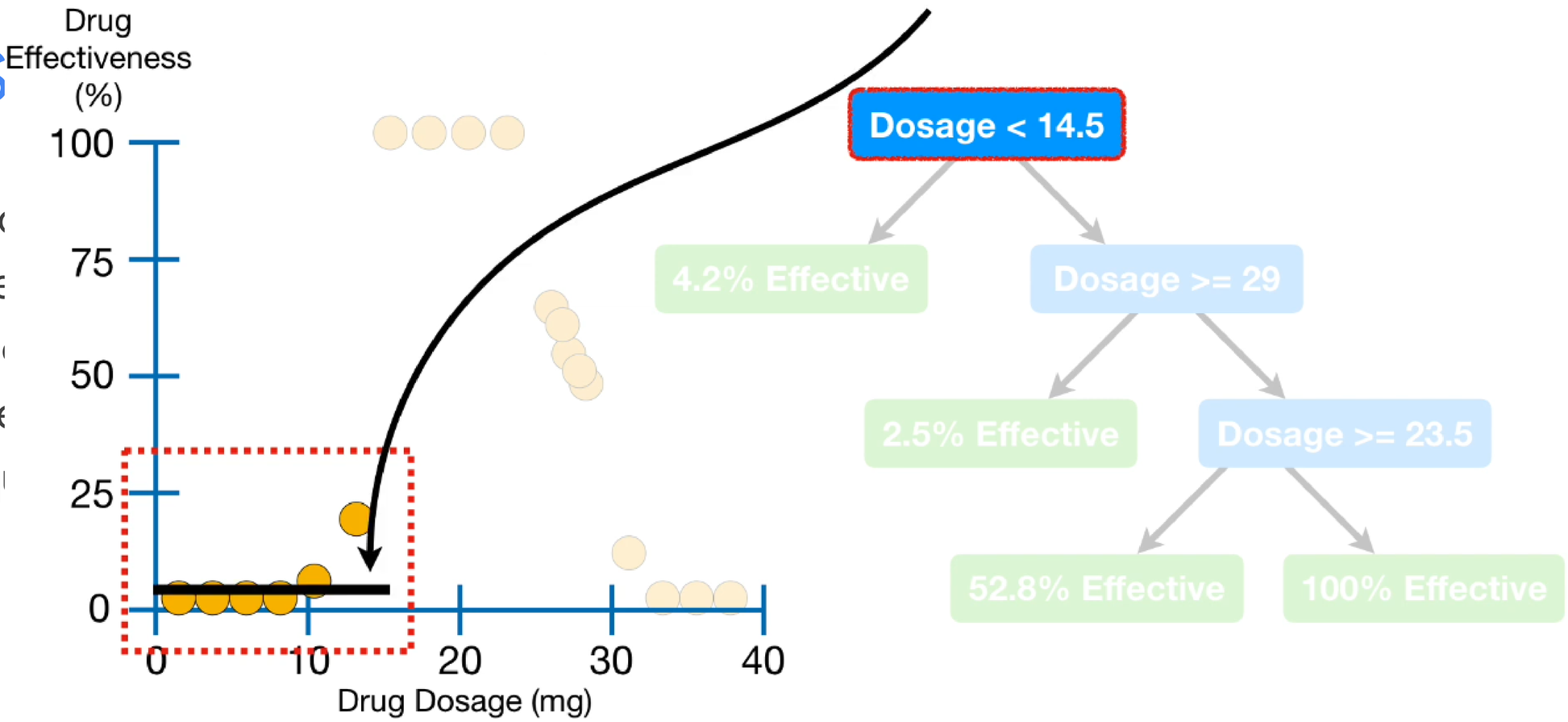
L

te

p

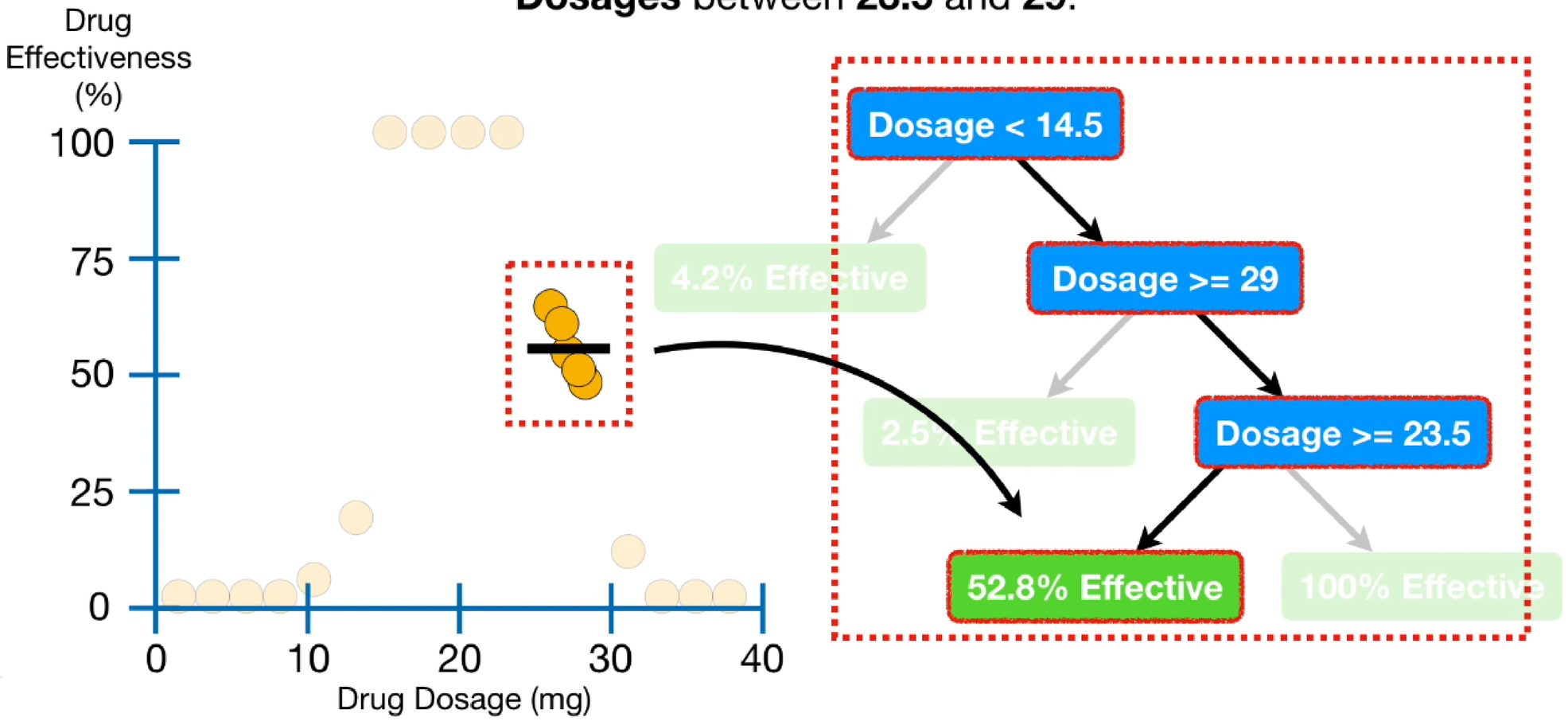
se

q



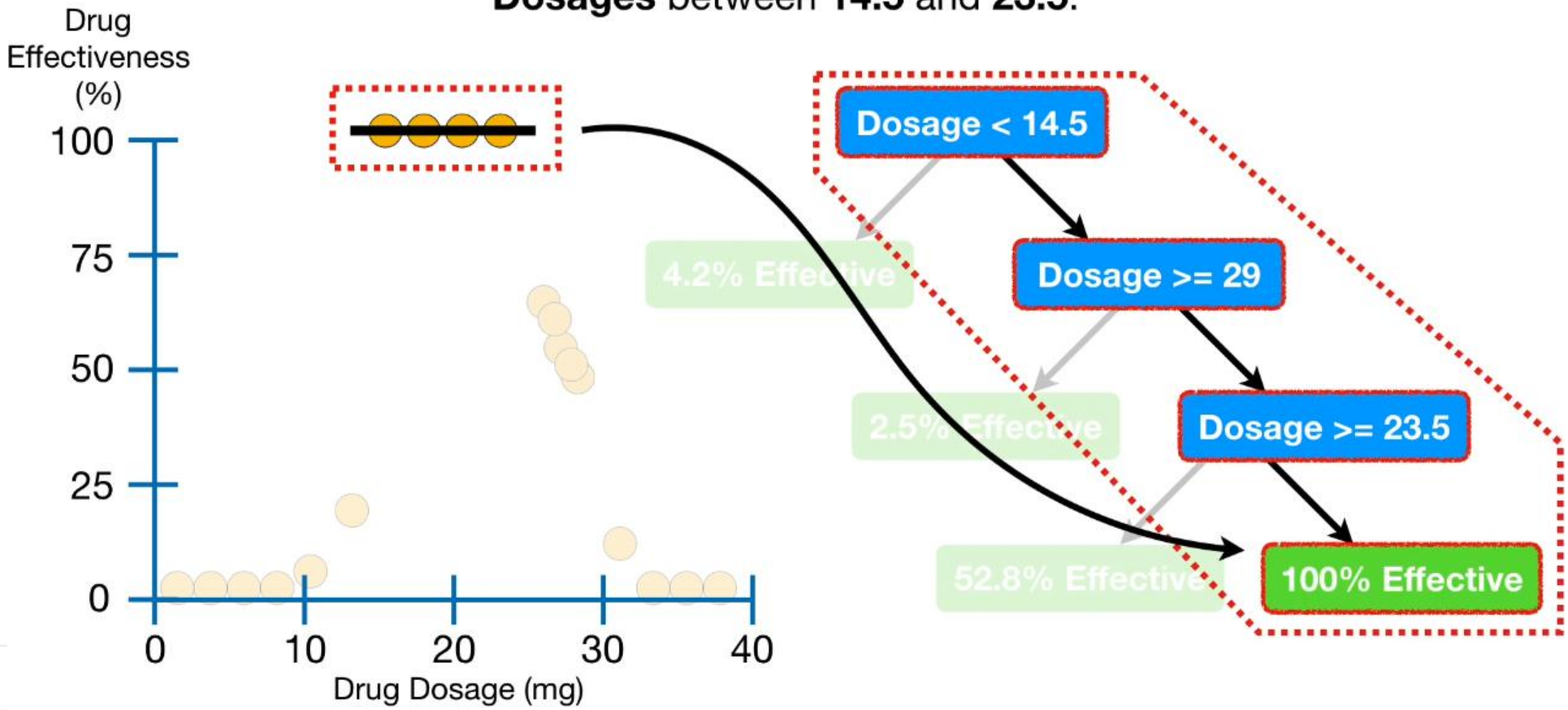
Credits:
StatQuest with Josh Starmer

…so the tree uses the average value, **52.8%**, as its prediction for people with **Dosages** between **23.5** and **29**.



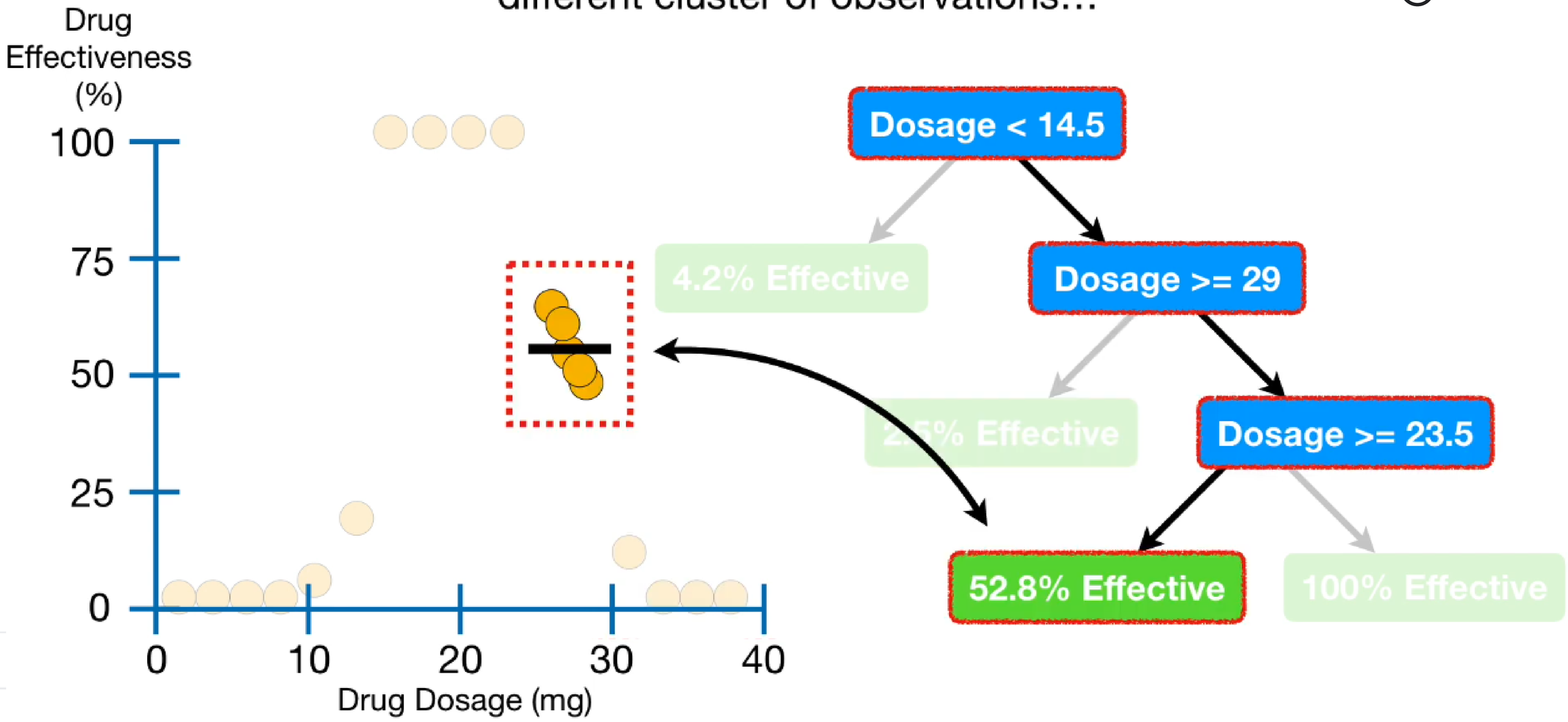
Credits:
StatQuest with Josh Starmer

…so the tree uses the average value, **100%**, as its prediction for people with **Dosages** between **14.5** and **23.5**.



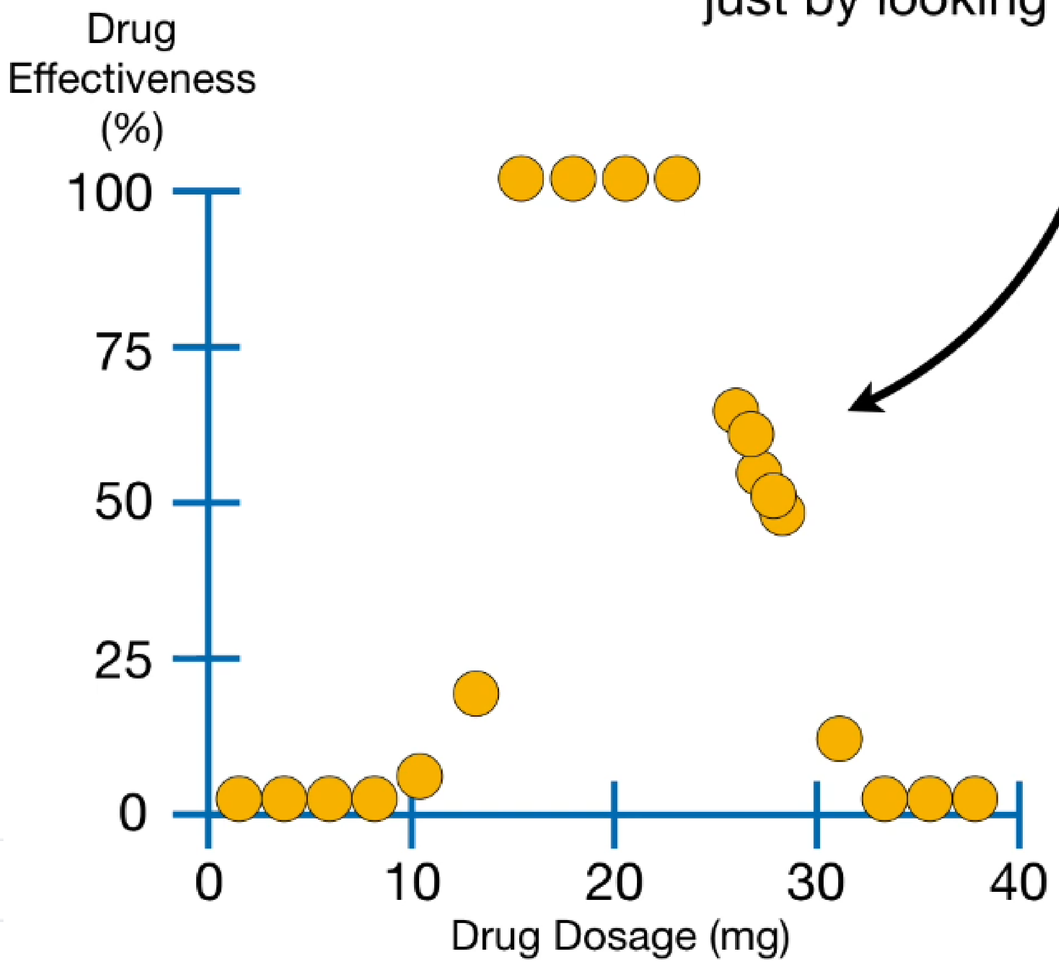
Since each leaf corresponds to the average **Drug Effectiveness** in a different cluster of observations…

Credits:
StatQuest with Josh Starmer



Credits:
StatQuest with Josh Starmer

At this point you might be thinking, “The **Regression Tree** is cool, but I can also predict **Drug Effectiveness** just by looking at the graph…”

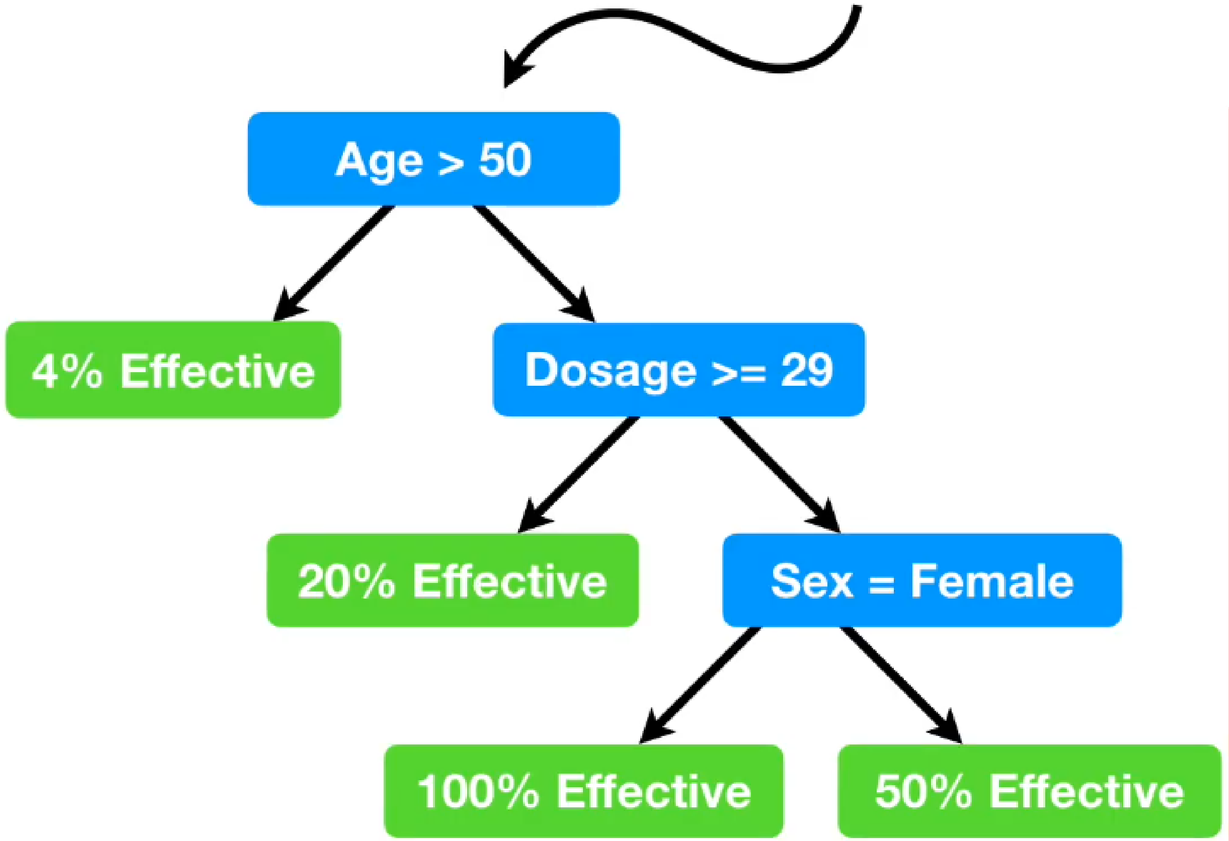


Credits:
StatQuest with Josh Starmer

But when we have **3** or more predictors, like **Dosage**, **Age** and **Sex**, to predict **Drug Effectiveness**, drawing a graph is very difficult, if not impossible.

Dosage	Age	Sex	Etc.	Drug Effect.
10	25	Female	…	98
20	73	Male	…	0
35	54	Female	…	100
5	12	Male	…	44
etc…	etc…	etc…	etc…	etc…

In contrast, a **Regression Tree** easily accommodates the additional predictors.

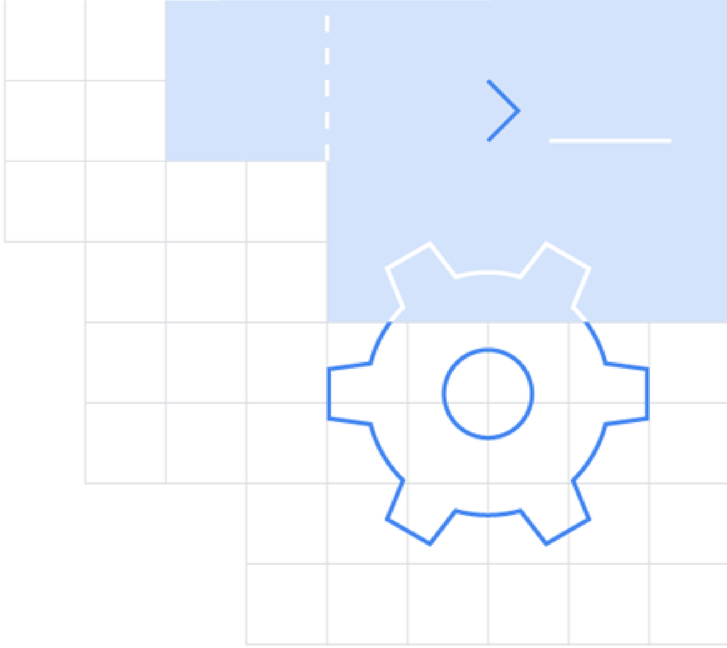


Dosage	Age	Sex	Etc.	Drug Effect.
10	25	Female	…	98
20	73	Male	…	0
35	54	Female	…	100
5	12	Male	…	44
etc…	etc…	etc…	etc…	etc…

Now you fully understand the concept

Decision Trees

Random Forest



Developer Student Clubs

Developer Student Clubs

Thank You!

Muhammad Huzaifa Shahbaz
DSC Lead at Google Developers
@mhuzaifadev

Credits:
StatQuest with Josh Starmer