AIBridge

Lecture 5

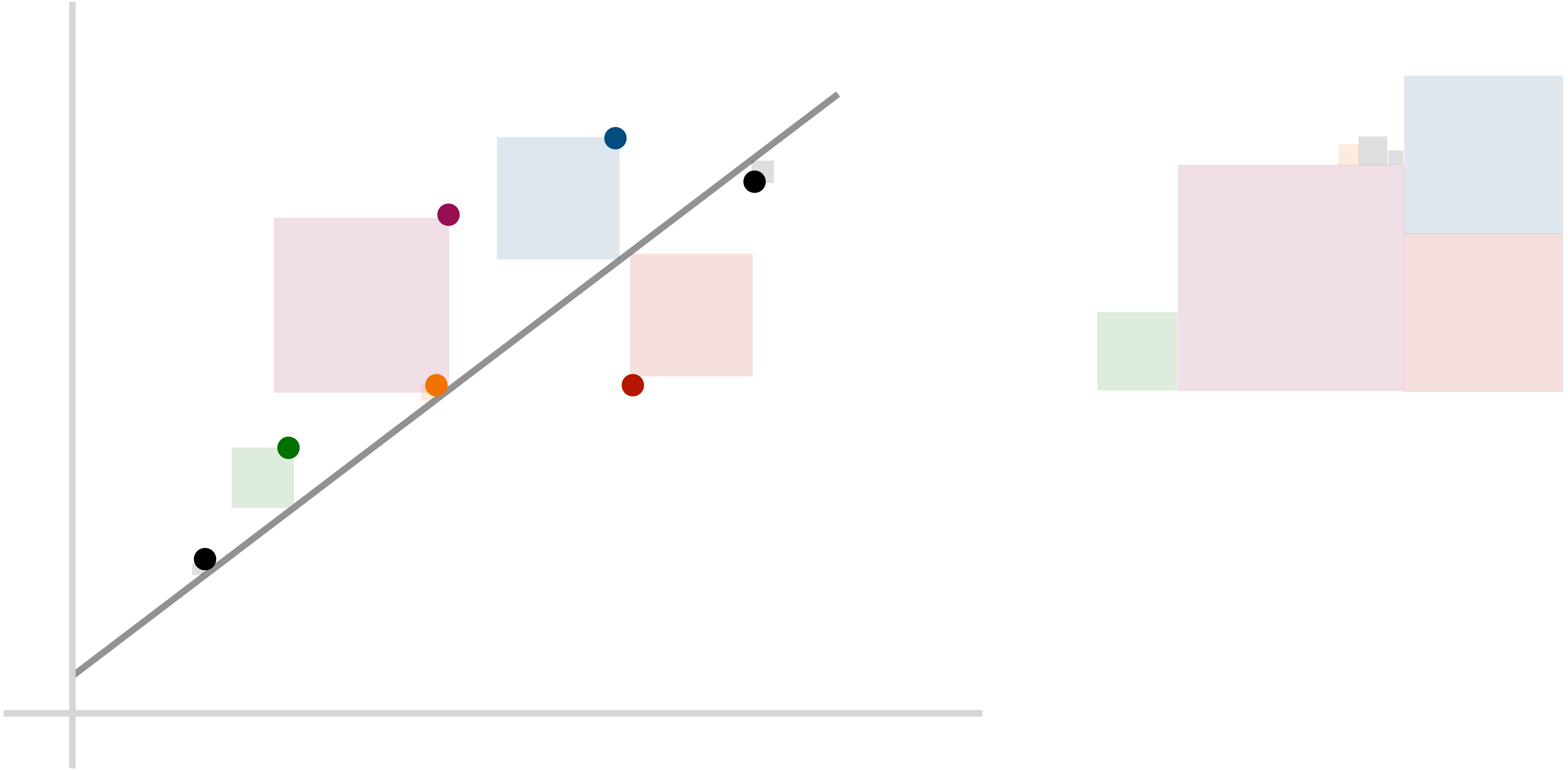
Let’s talk about last week’s lab!

Let's talk about last week's lab!

what does this even mean?

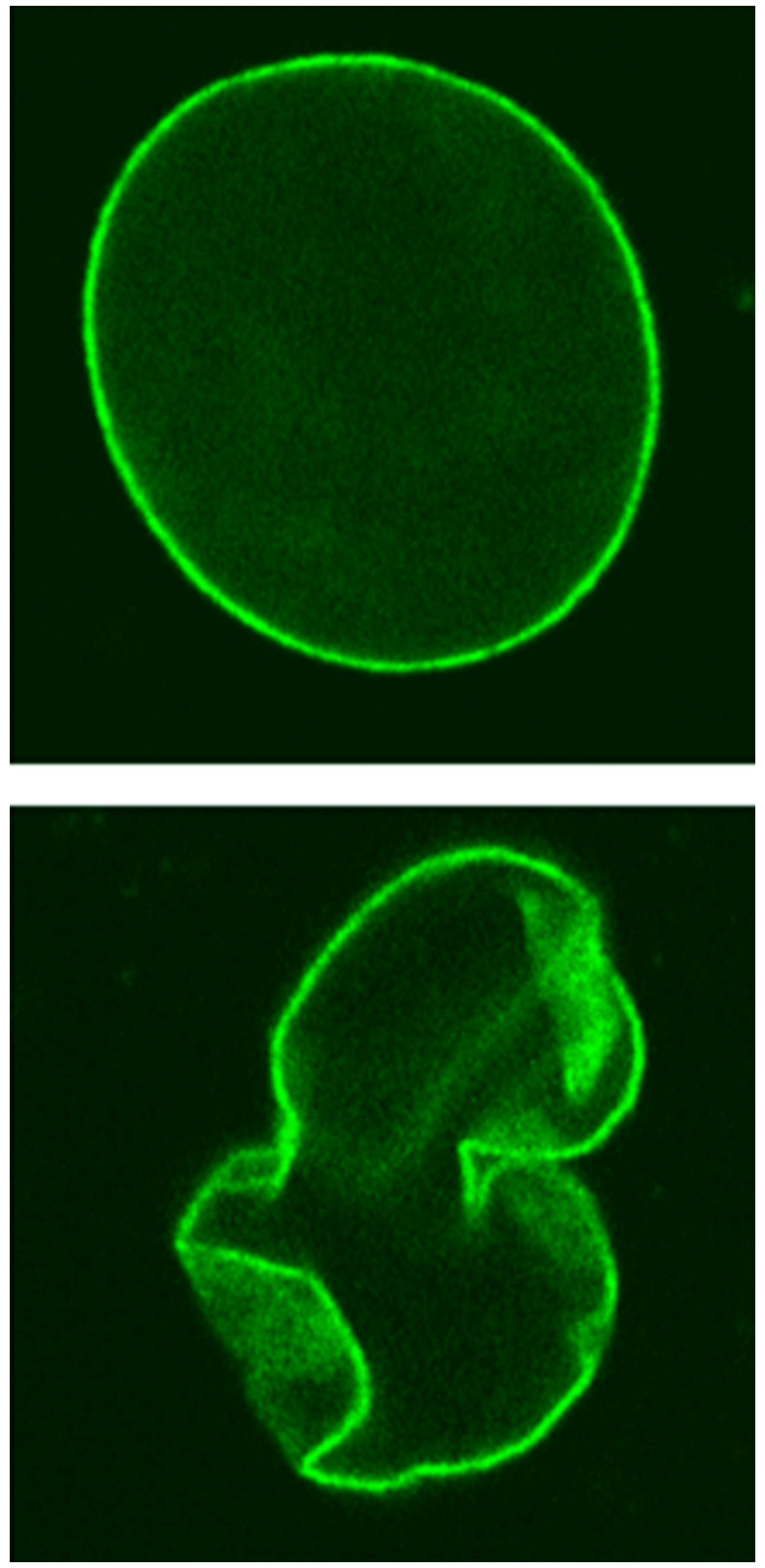
What circumstances made the model fit better?

worse?



Accuracy

“Why is it not enough?”

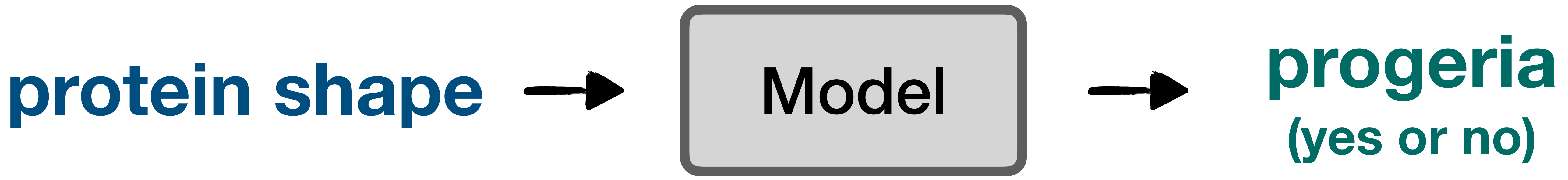


Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

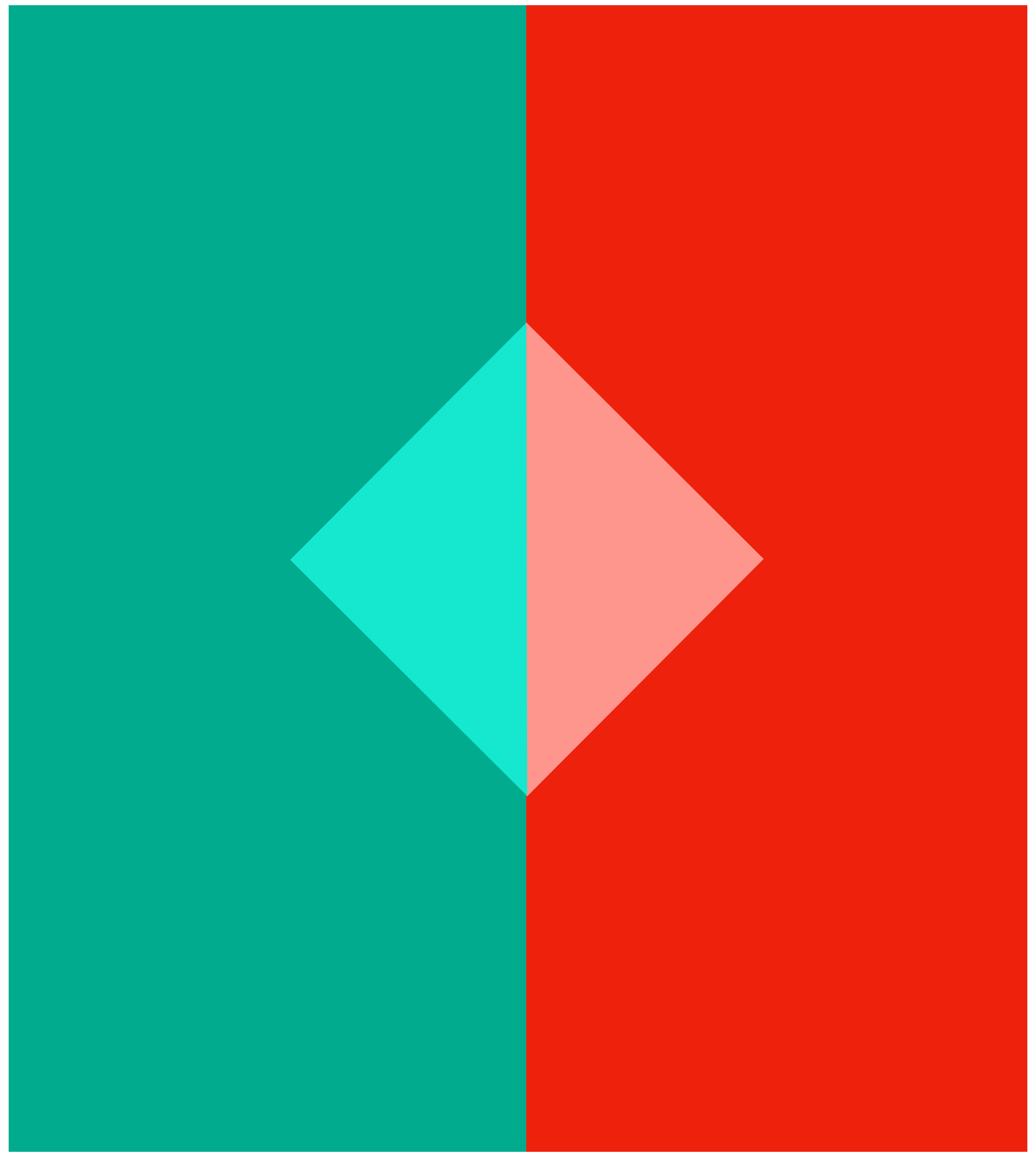
Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

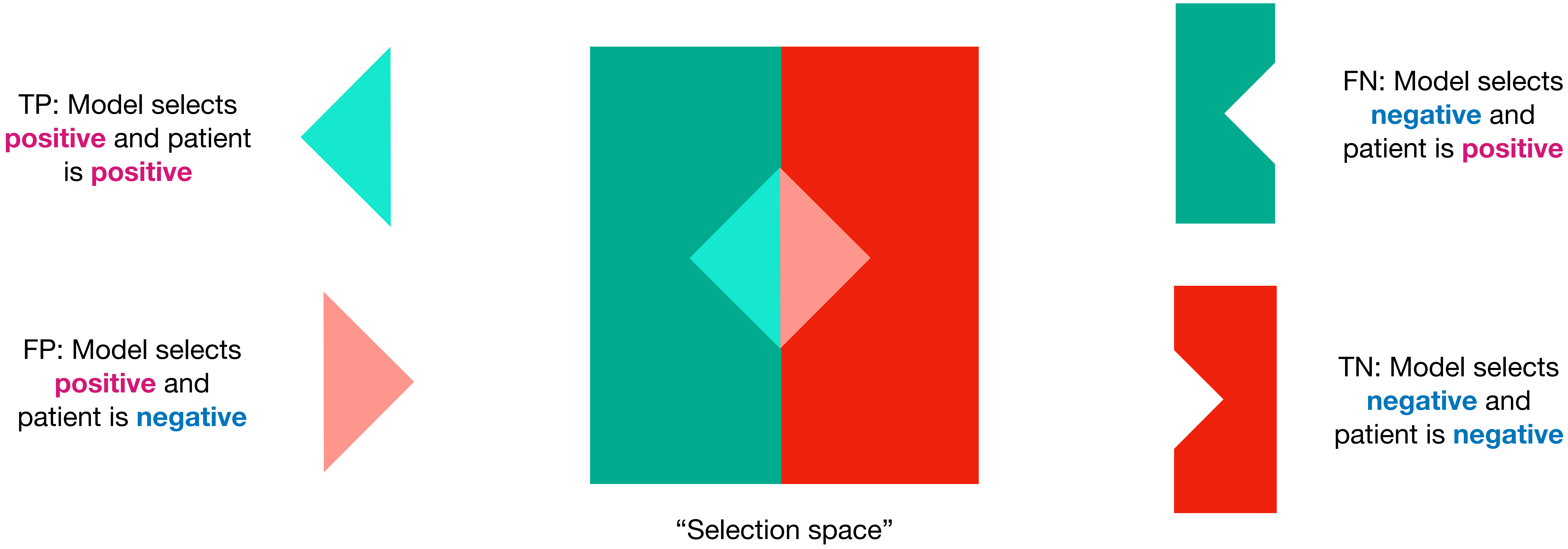


Q: If my model predicts with 99.99% accuracy, is it good enough?

Accuracy, Precision, and Recall

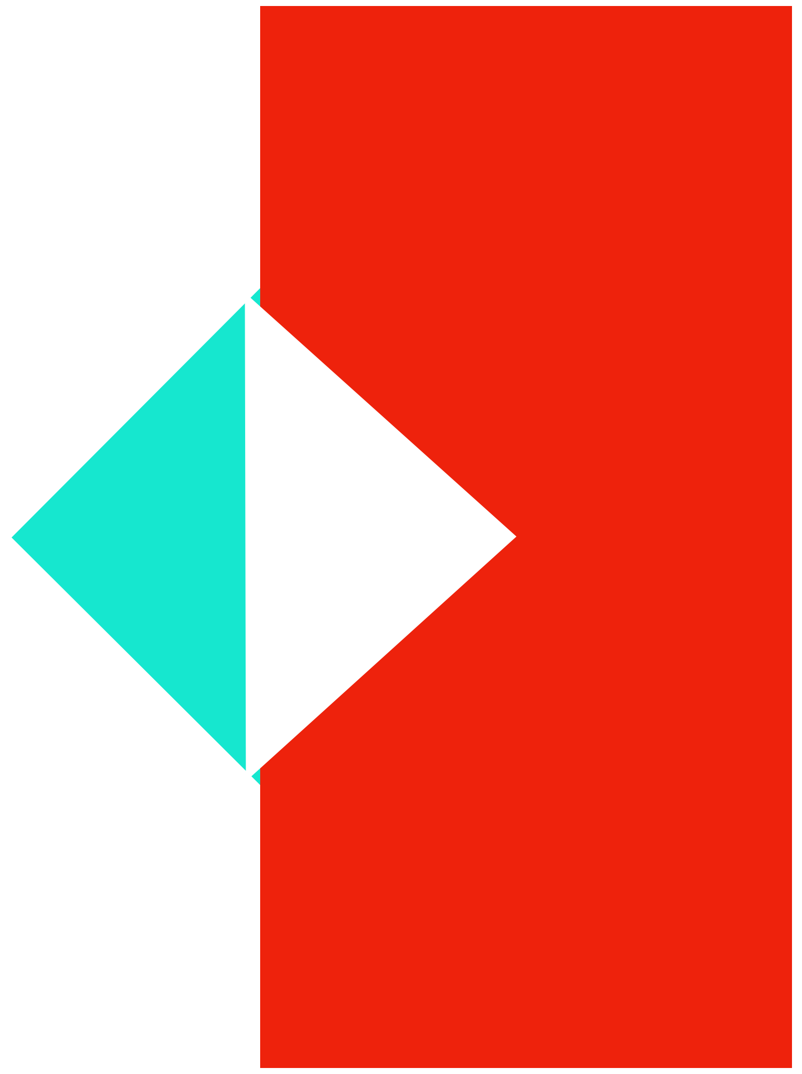


“Selection space”


TP: Model selects positive and patient is positive
FP: Model selects positive and patient is negative
"Selection space"
FN: Model selects negative and patient is positive
TN: Model selects negative and patient is negative

Accuracy

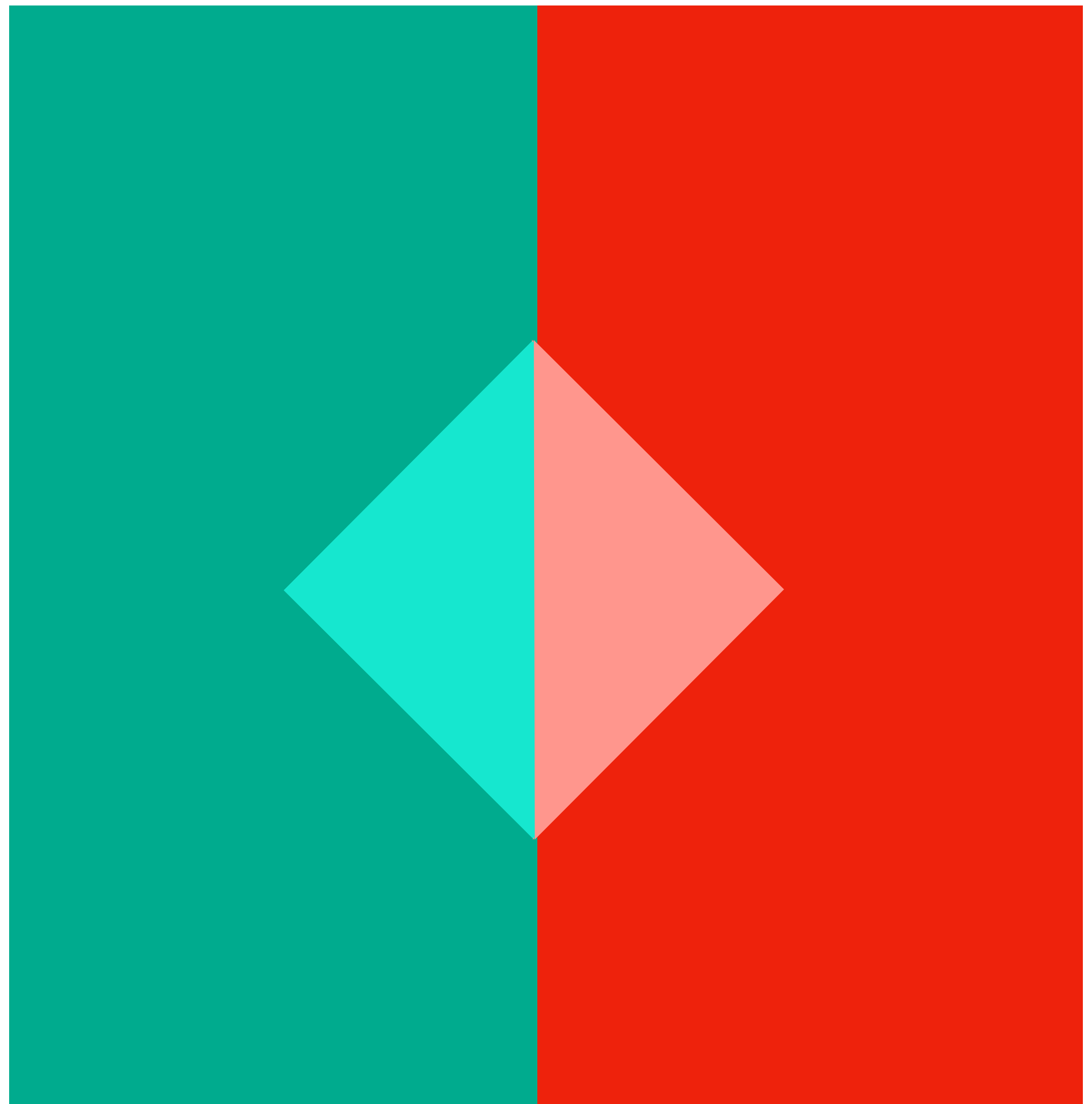
Overall ability of model



"Number of cases where we chose **positive** when patient is **positive**

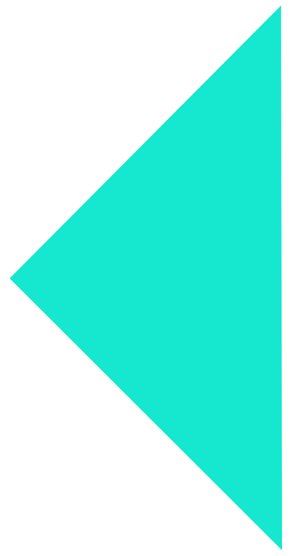
and

Number of cases where we chose **negative** when patient is **negative**"

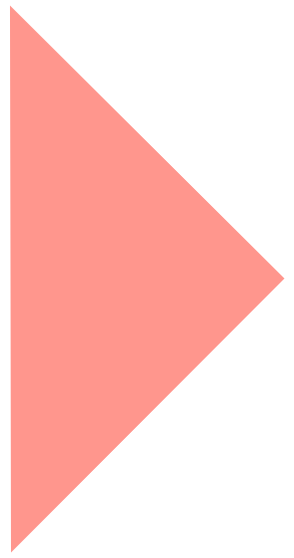


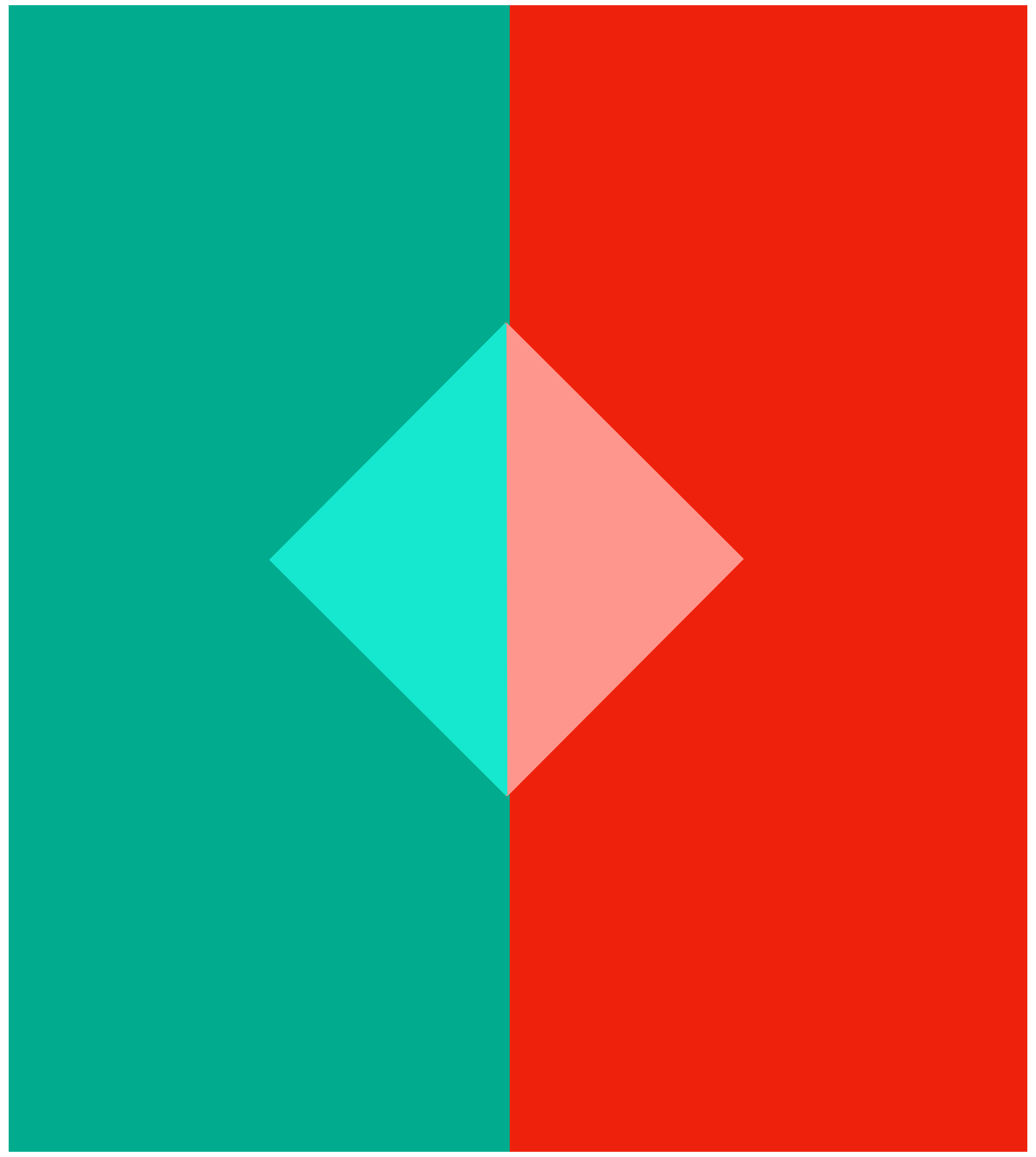
"Everything"

TP: Model selects **positive** and patient is **positive**



FP: Model selects **positive** and patient is **negative**





"Selection space"

FN: Model selects **negative** and patient is **positive**

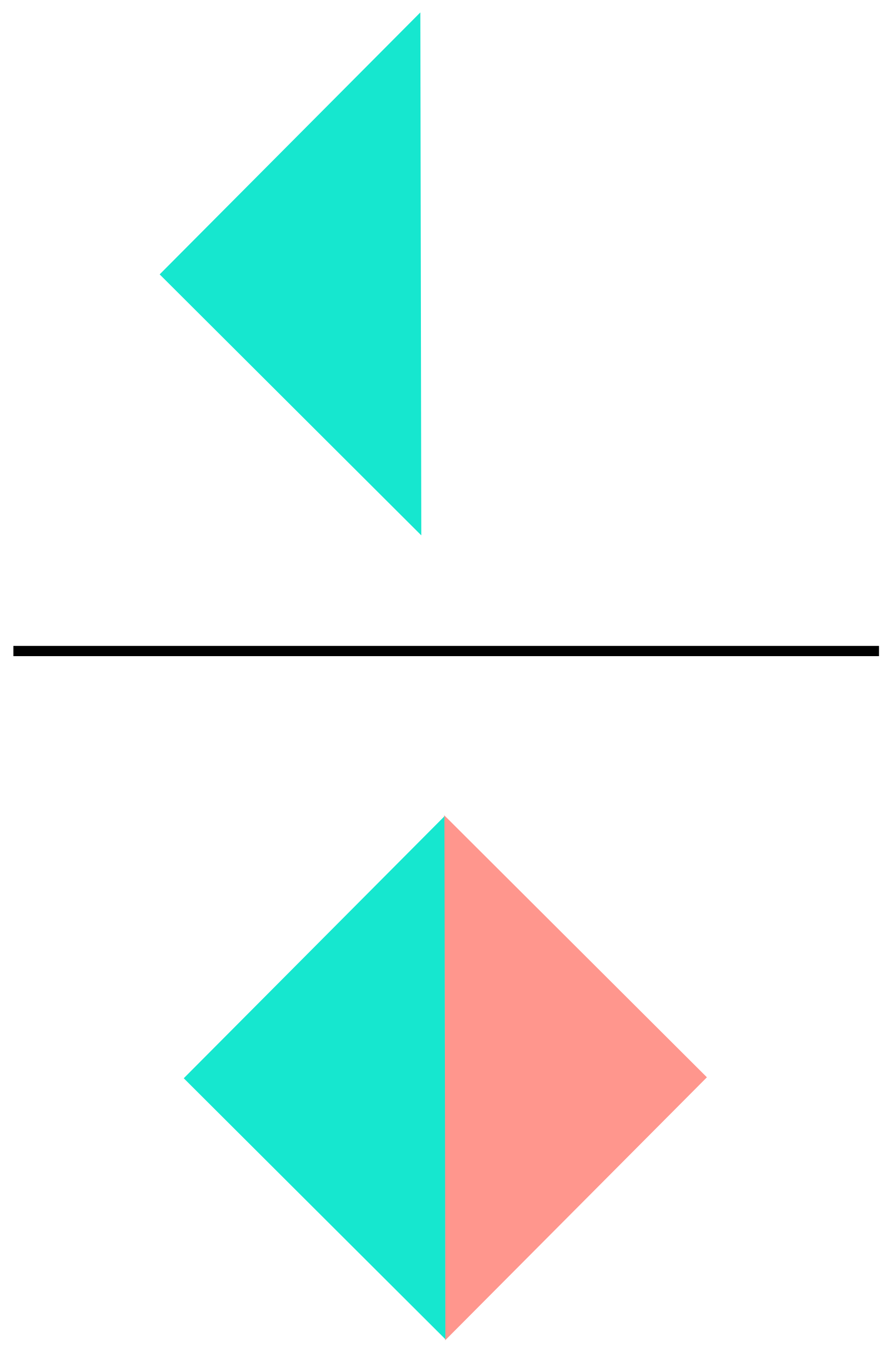
TN: Model selects **negative** and patient is **negative**

Accuracy

Overall ability of model

Precision

Amount of selection that’s actually correct.

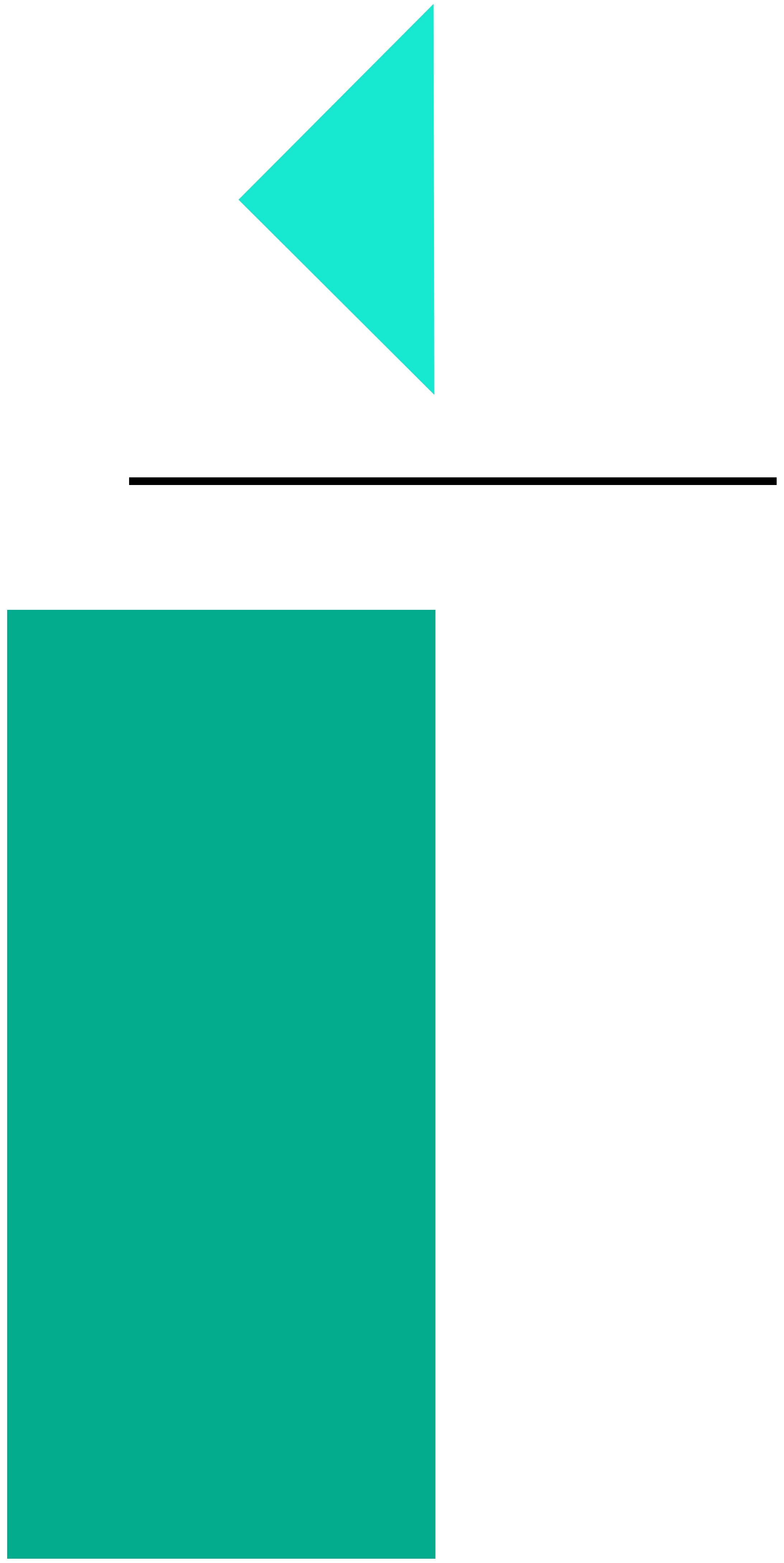


“Number of cases where we chose **positive** when patient is **positive**”

“All selected **positive** by the model”

Recall

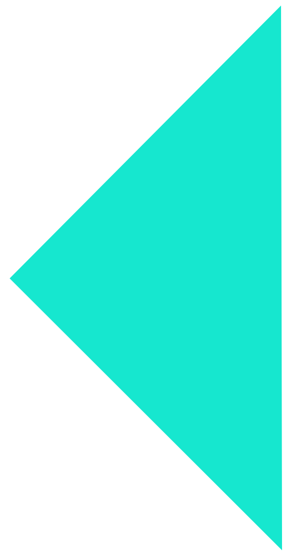
Amount of what needs to be selected that is selected



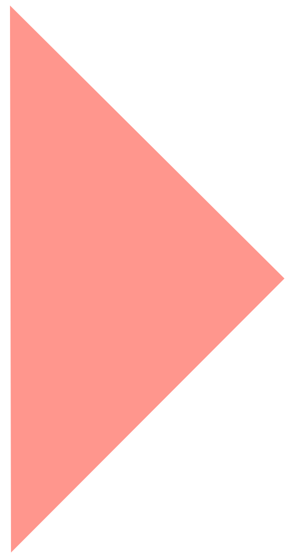
“Number of cases where we chose **positive** when patient is **positive**”

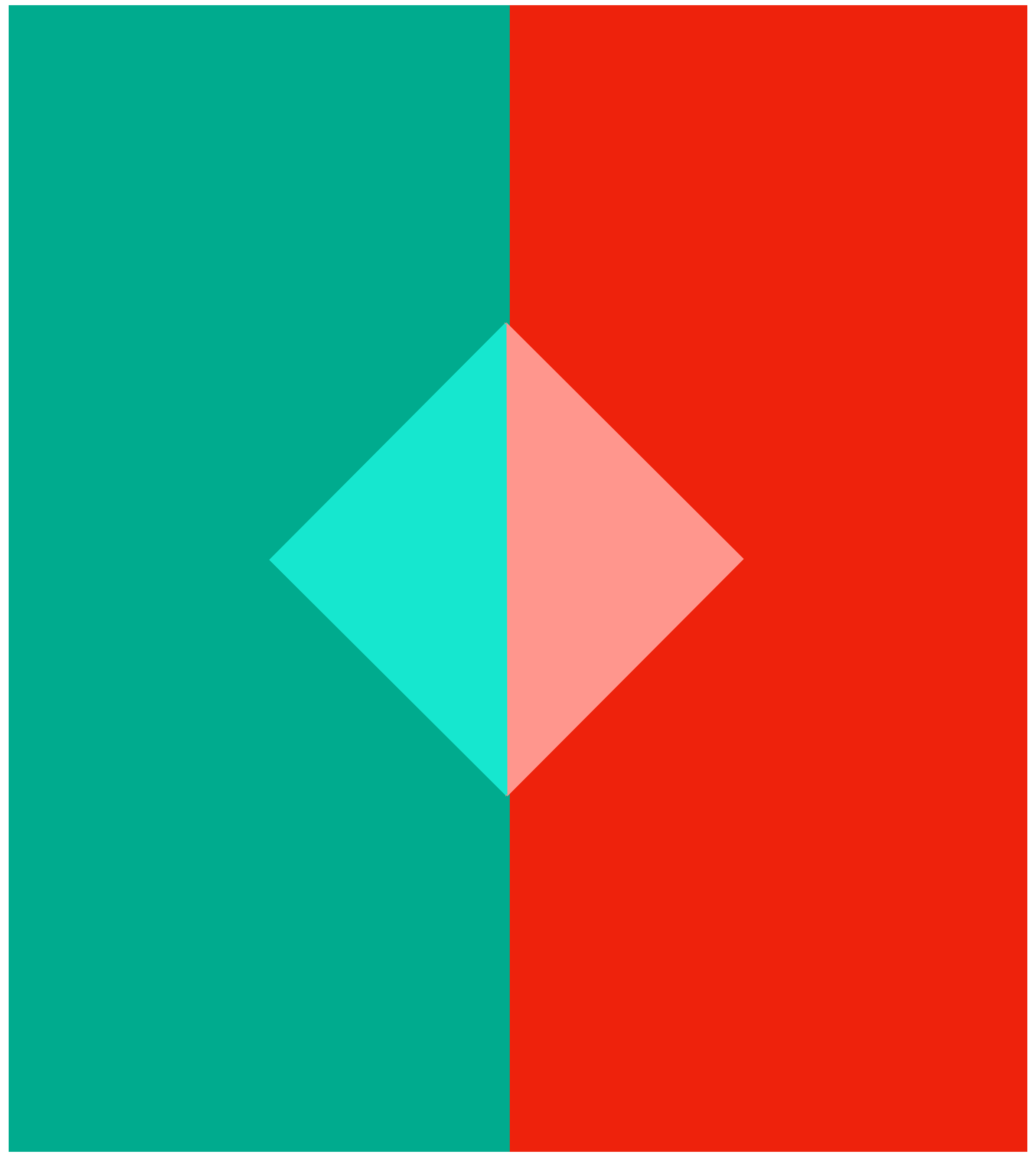
“All cases that are **positive**”

Model selects **positive** and patient is **positive**



Model selects **positive** and patient is **negative**





“Selection space”

Model selects **negative** and patient is **negative**

Model selects **negative** and patient is **positive**

Accuracy

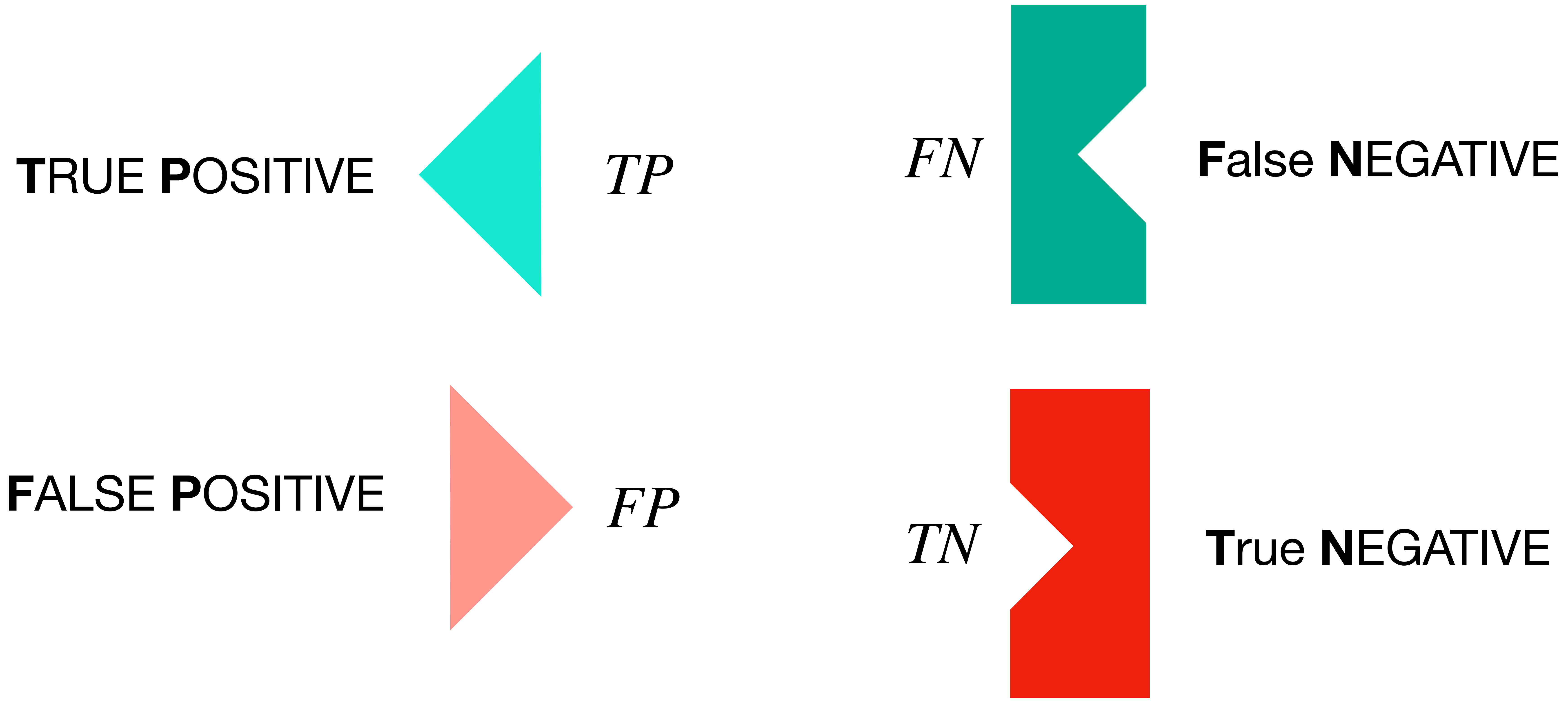
Overall ability of model

Precision

Amount of selection that’s actually correct.

Recall

Amount of what needs to be selected that is selected



Accuracy

Overall ability of model

$$\frac{TP + TN}{Total}$$

Precision

Amount of selection that's actually correct.

$$\frac{TP}{TP + FP}$$

Recall

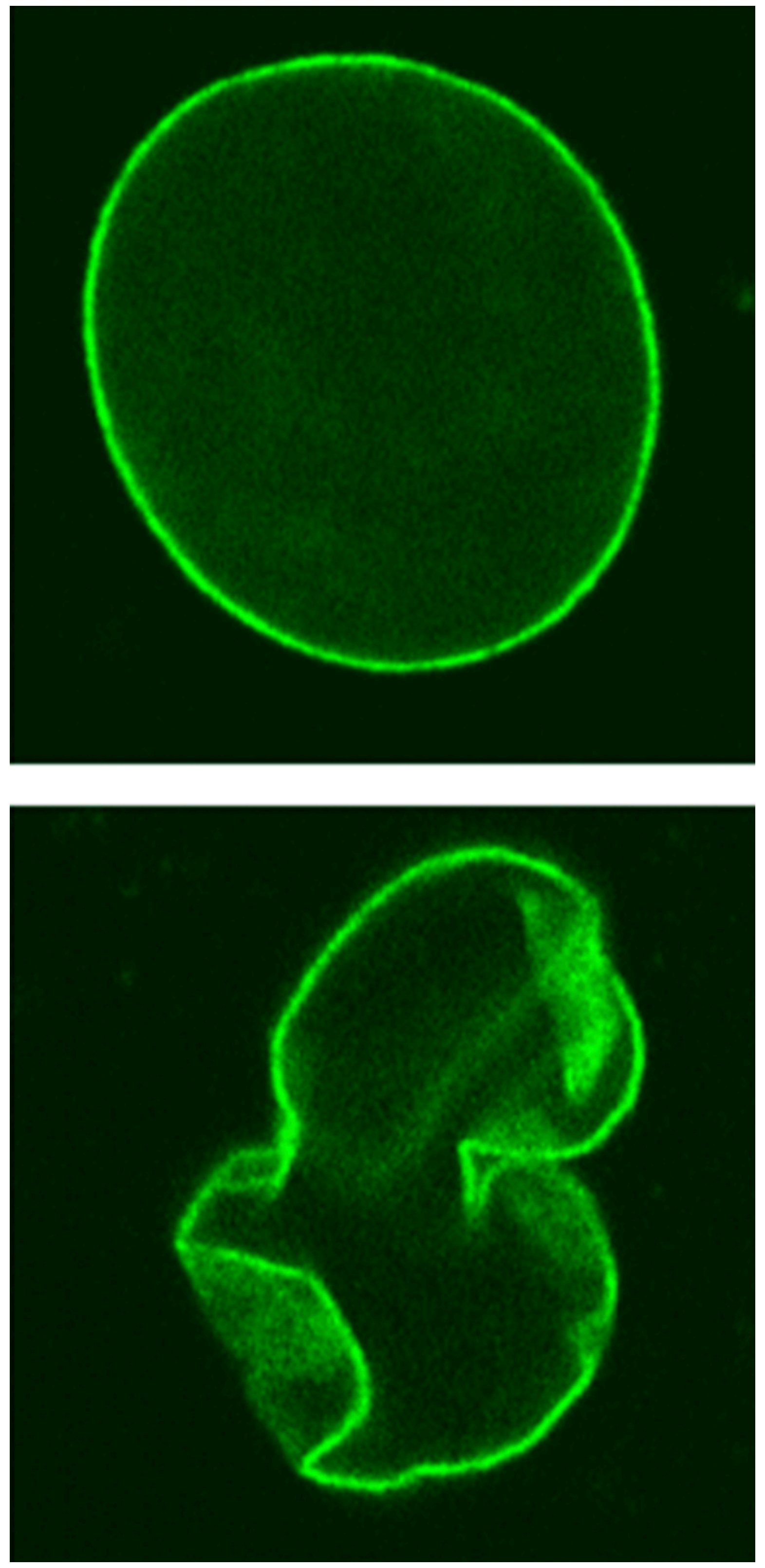
Amount of what needs to be selected that is selected

$$\frac{TP}{TP + FN}$$

Sources: [6][7][8][9][10][11][12][13][14] view · talk · edit

		Predicted condition			
	Total population $= P + N$	**Positive (PP)**	**Negative (PN)**	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	**Positive (P)**	**True positive (TP), hit**	**False negative (FN), type II error, miss, underestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
	Negative (N)	**False positive (FP), type I error, false alarm, overestimation**	**True negative (TN), correct rejection**	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN} = 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2 PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

https://en.wikipedia.org/wiki/Precision_and_recall



Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

Q: If my model predicts with 99.99% accuracy, is it good enough?

protein shape → Model → progeria (yes or no)

Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

a proposed model:

protein shape → Model → no progeria regardless

actually pretty accurate!

Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

Accuracy

Overall ability of model

$$\frac{TP + TN}{Total}$$

exactly zero

Precision

Amount of selection that's actually correct.

$$\frac{TP}{TP + FP}$$

Recall

Amount of what needs to be selected that is selected

$$\frac{TP}{TP + FN}$$

← scaled properly!

Model → no progeria regardless

Progeria affects ~159 patients in the US

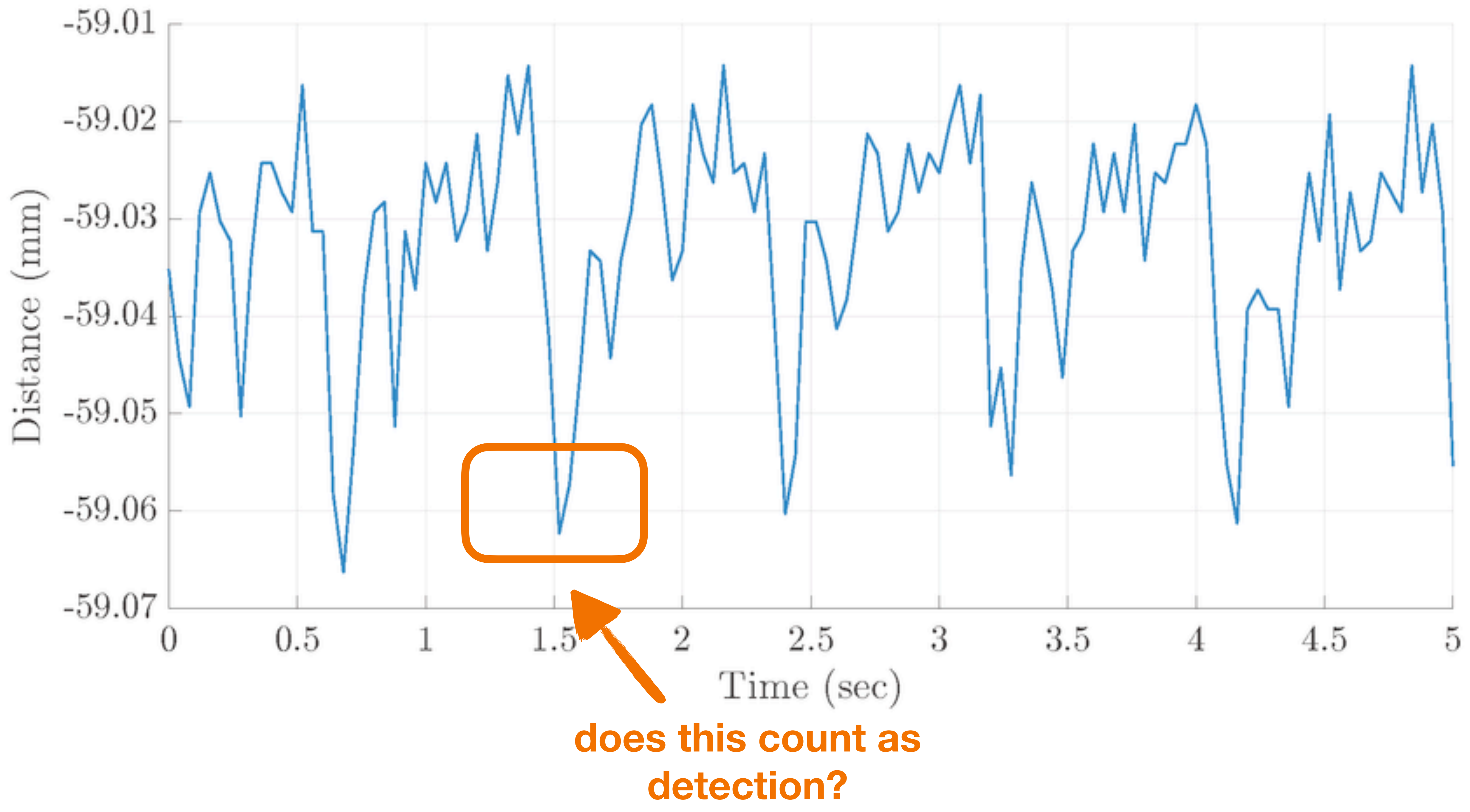
we have a dataset of all American pediatric patients

storytime!

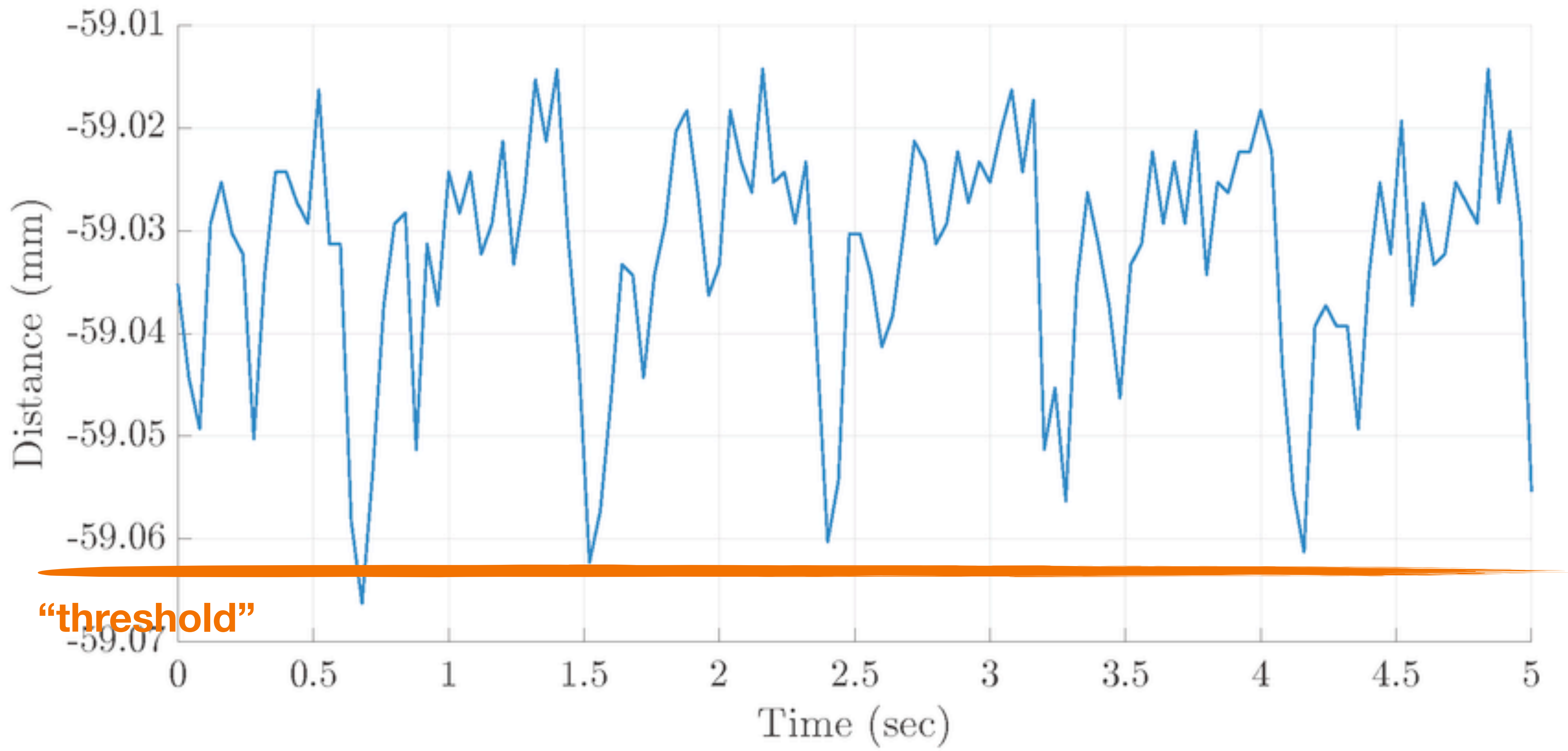
storytime!

storytime!


US AIR FORCE

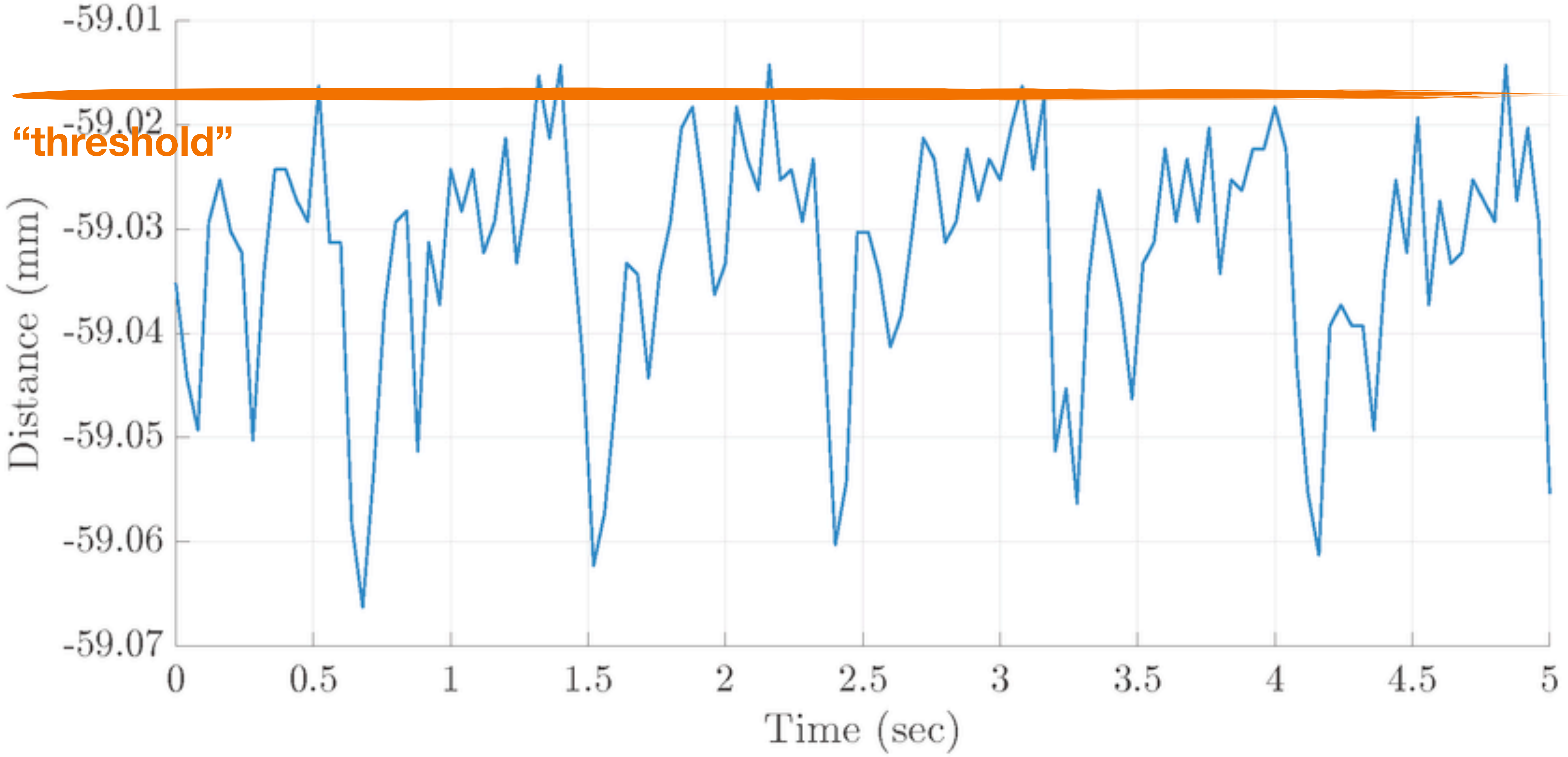

-59.01
-59.02
-59.03
-59.04
-59.05
-59.06
-59.07
Distance (mm)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Time (sec)
does this count as detection?

storytime!



high recall, low precision

storytime!

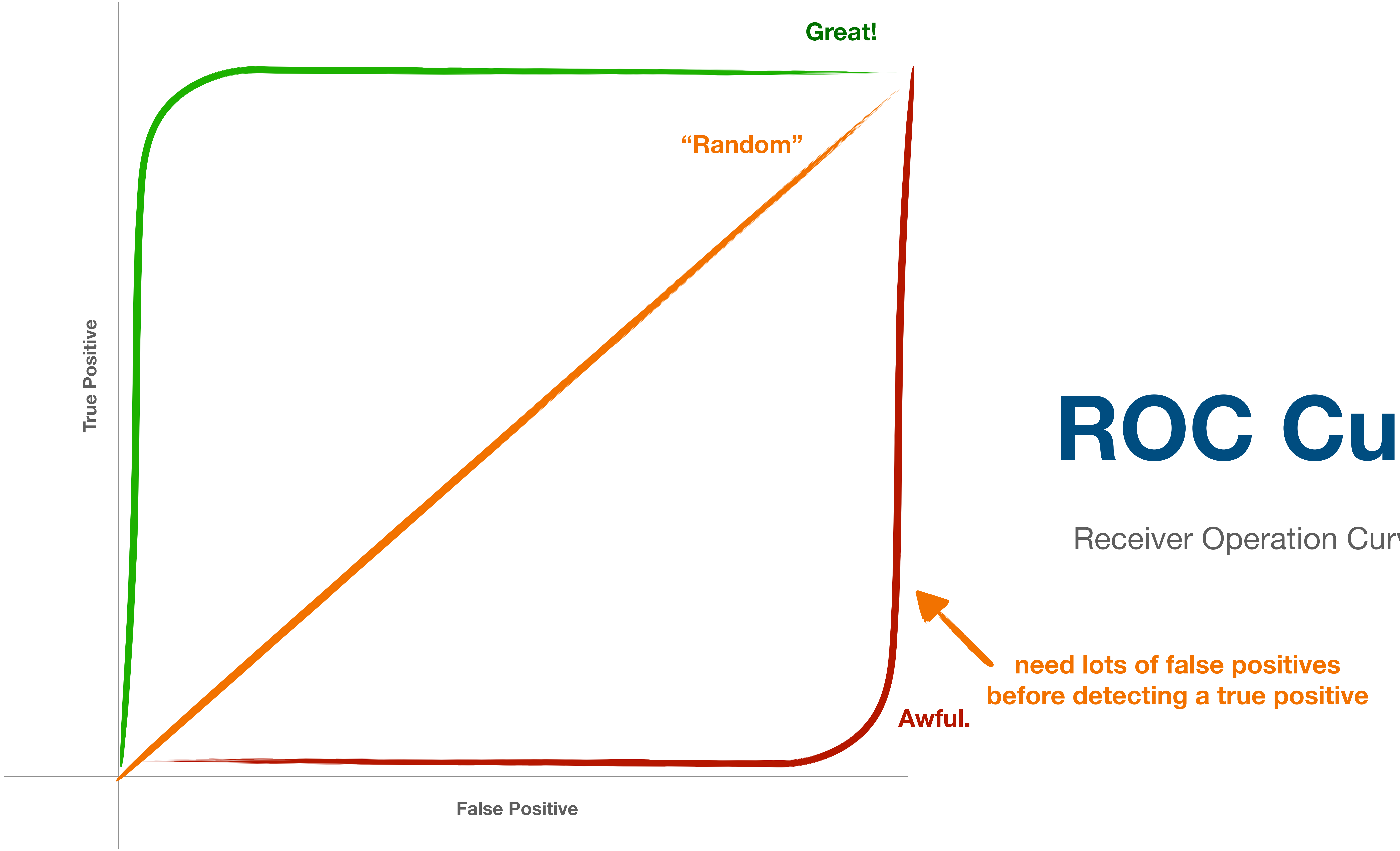


high precision, low recall

quantifying “threshold”

quantifying “threshold”

ROC Curve!



ROC Curve!

Receiver Operation Curve

- **ROC Curve** quantify the amount of "error"/noise that is necessary for a classifier to make a good prediction

AUC

area under [the ROC] curve

- **AUC** and also Precision-Recall Area Under Curve (PR AUC).

what makes models fit better

- **more** data
- **balanced** data
- **normalized** data

more data

balanced data

normalized data

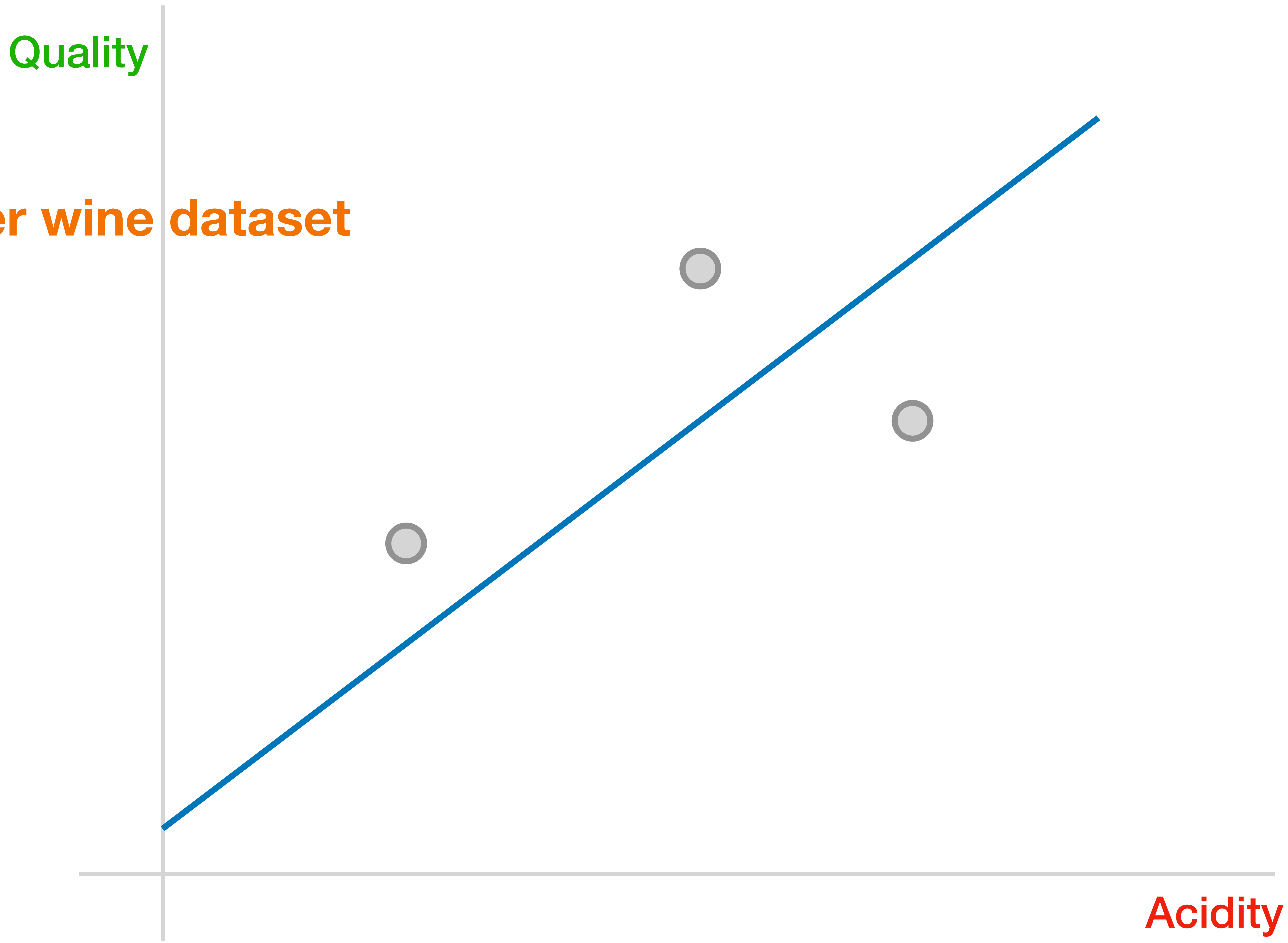
more data balanced data normalized data

more data

let’s say we have a simpler wine dataset

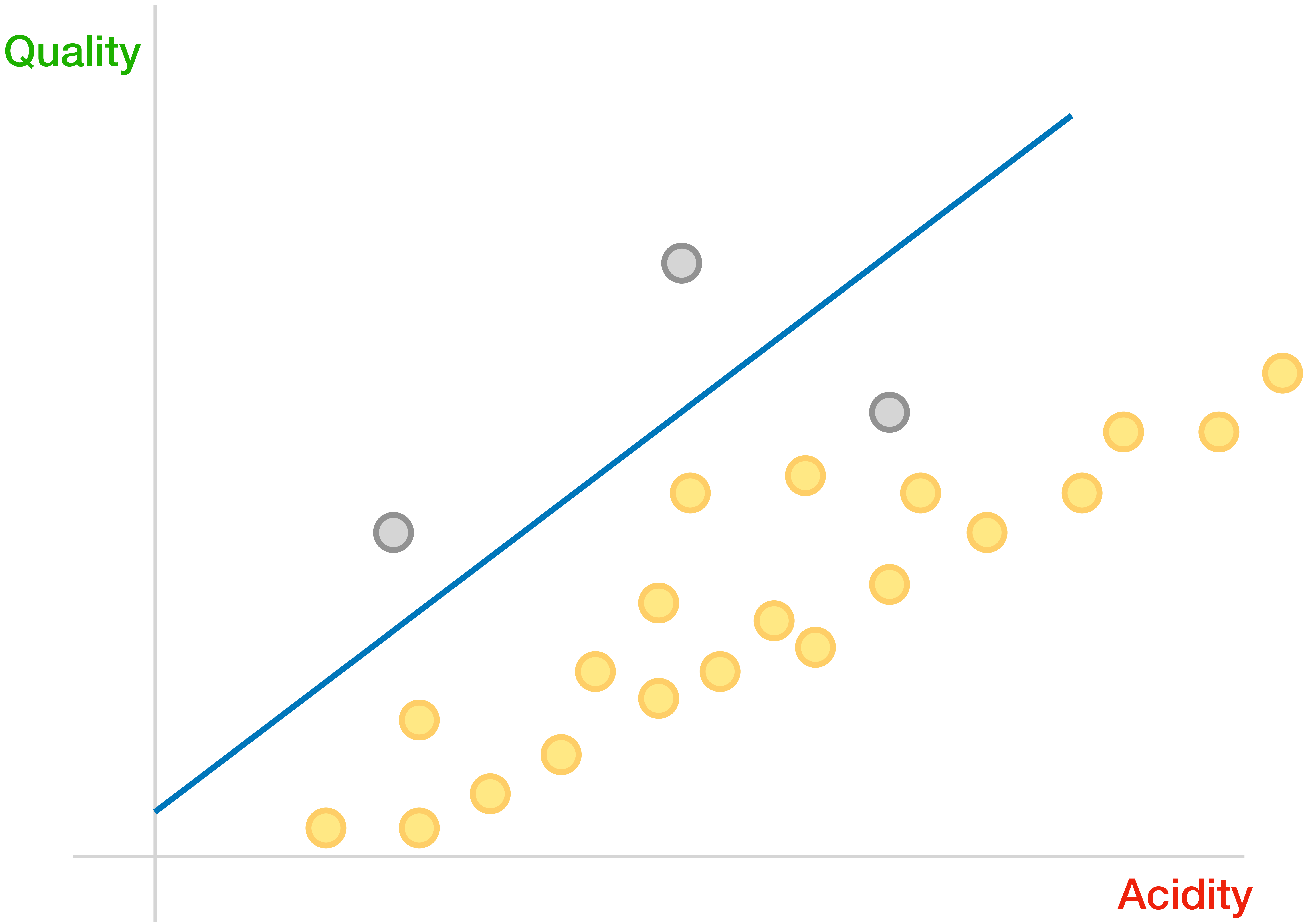
Quality on the y axis

Acidity on the x axis



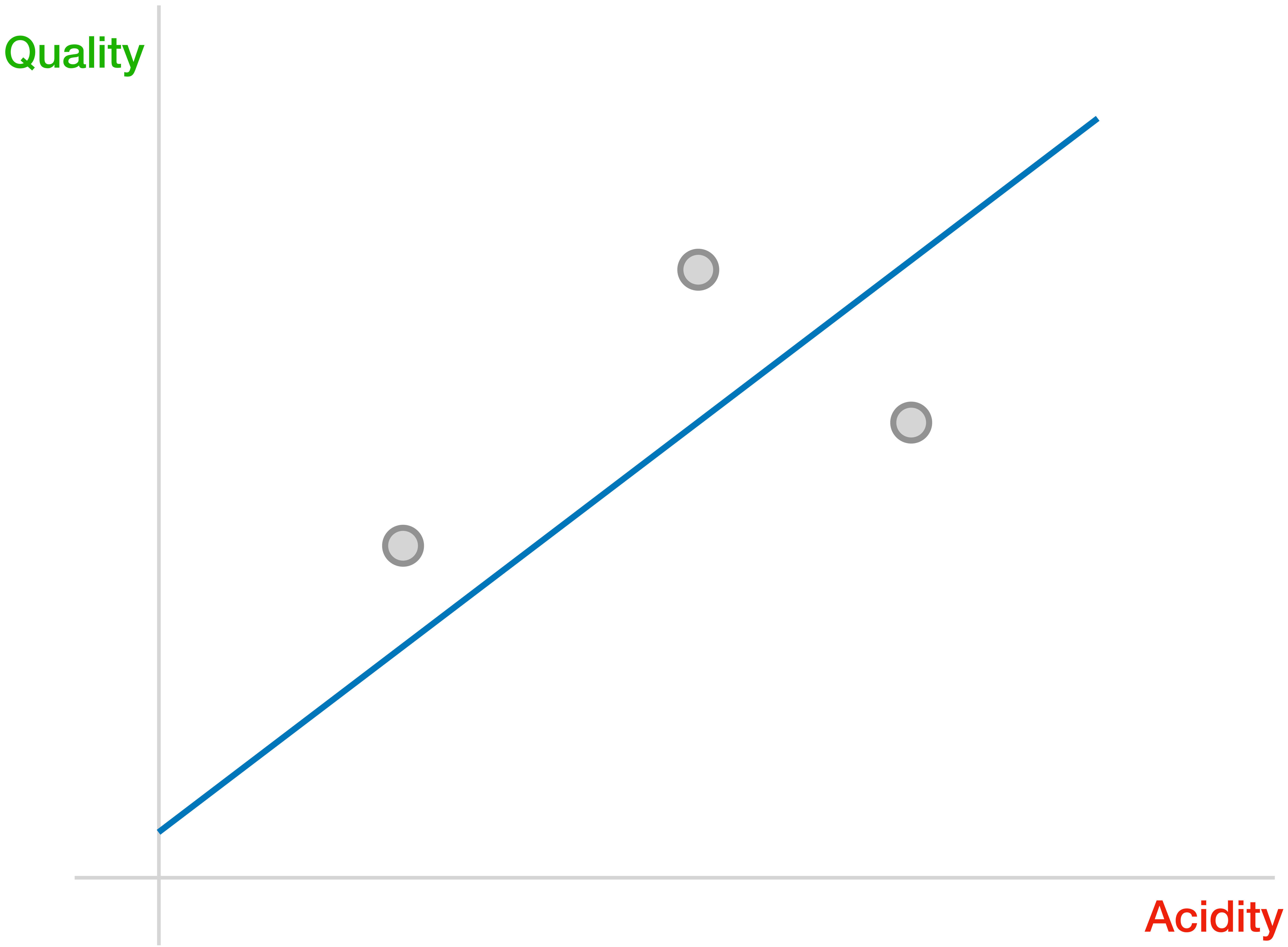
more data

Quality on the y axis
Acidity on the x axis

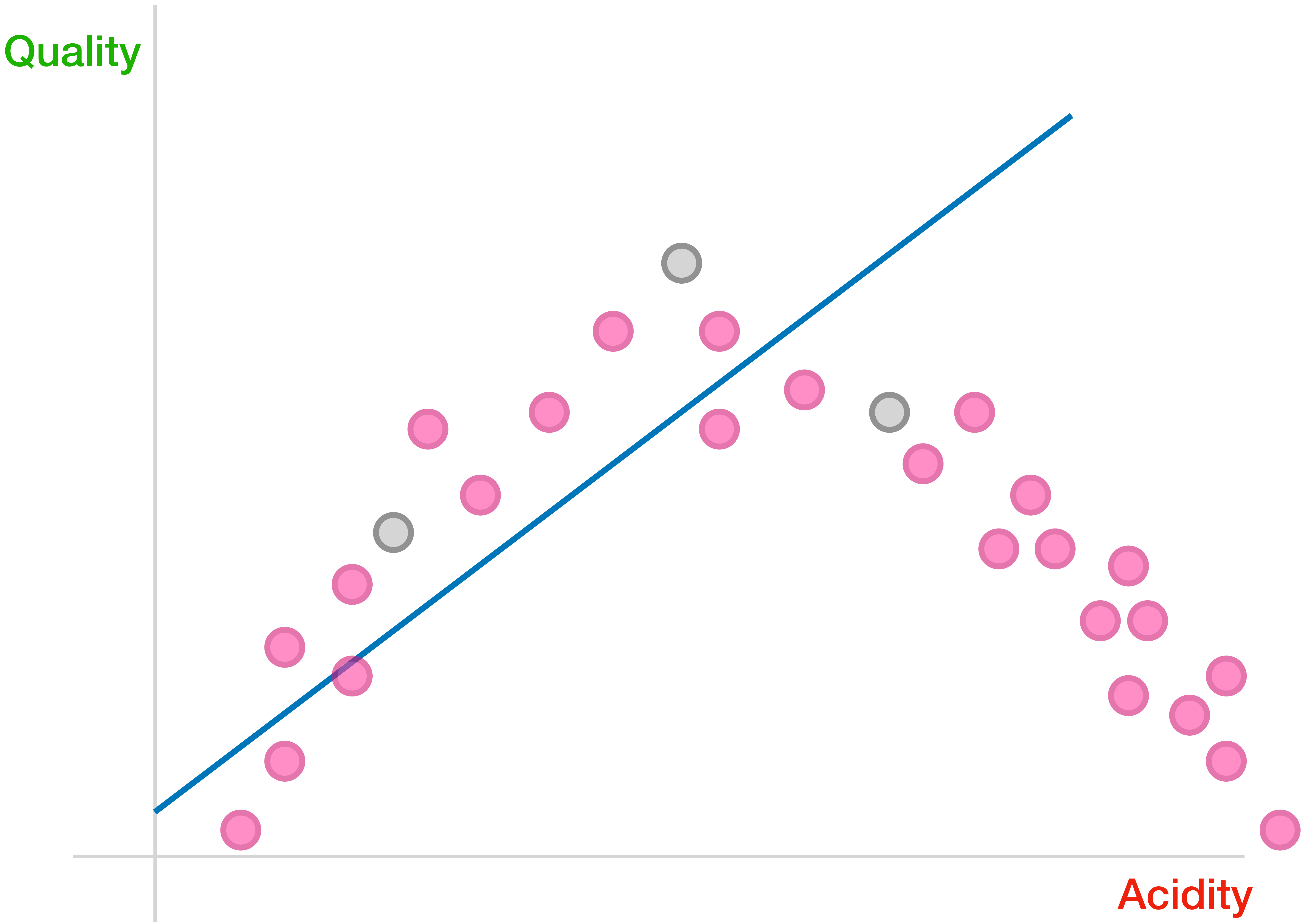


more data

Quality on the y axis
Acidity on the x axis

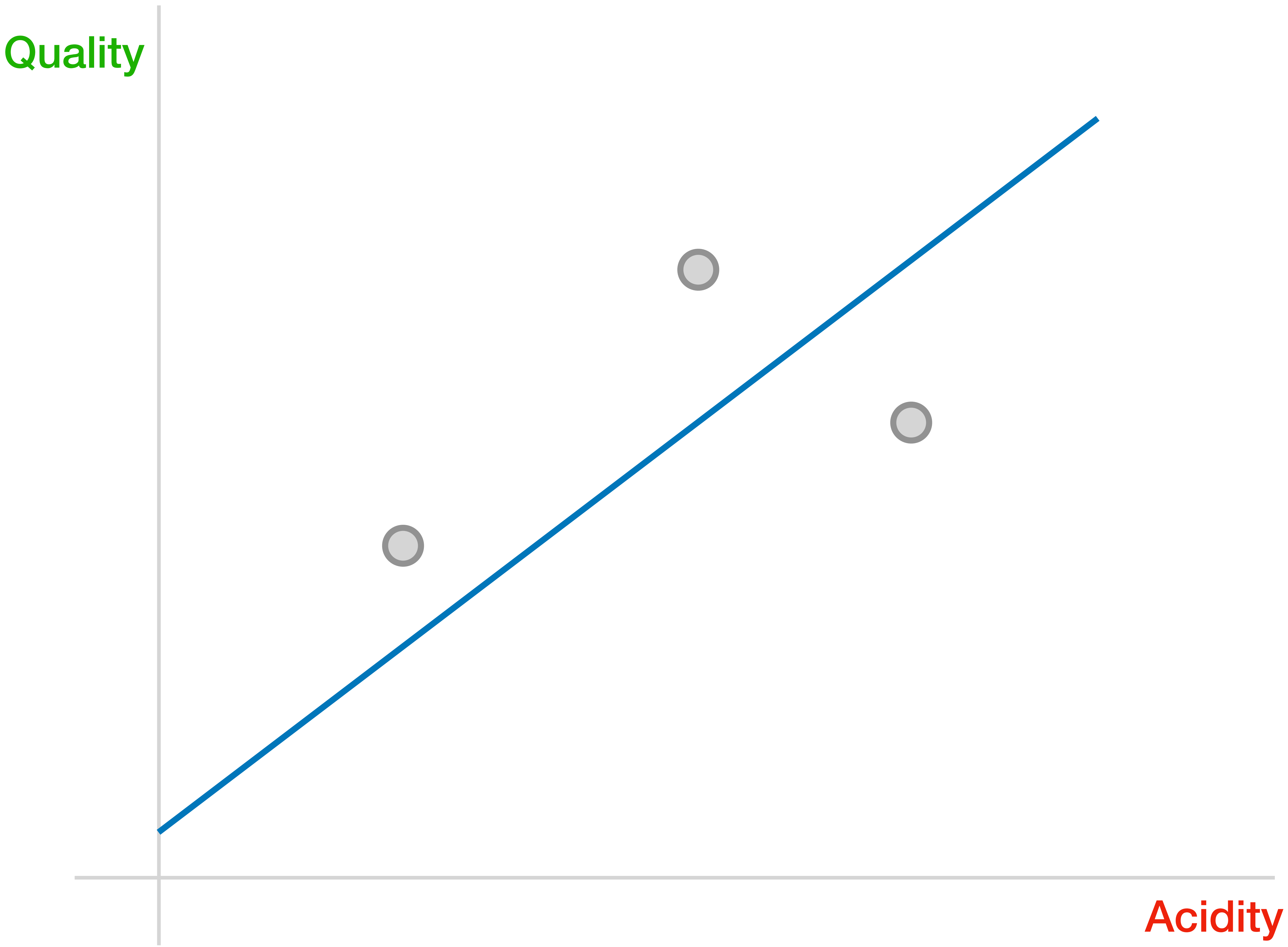


more data



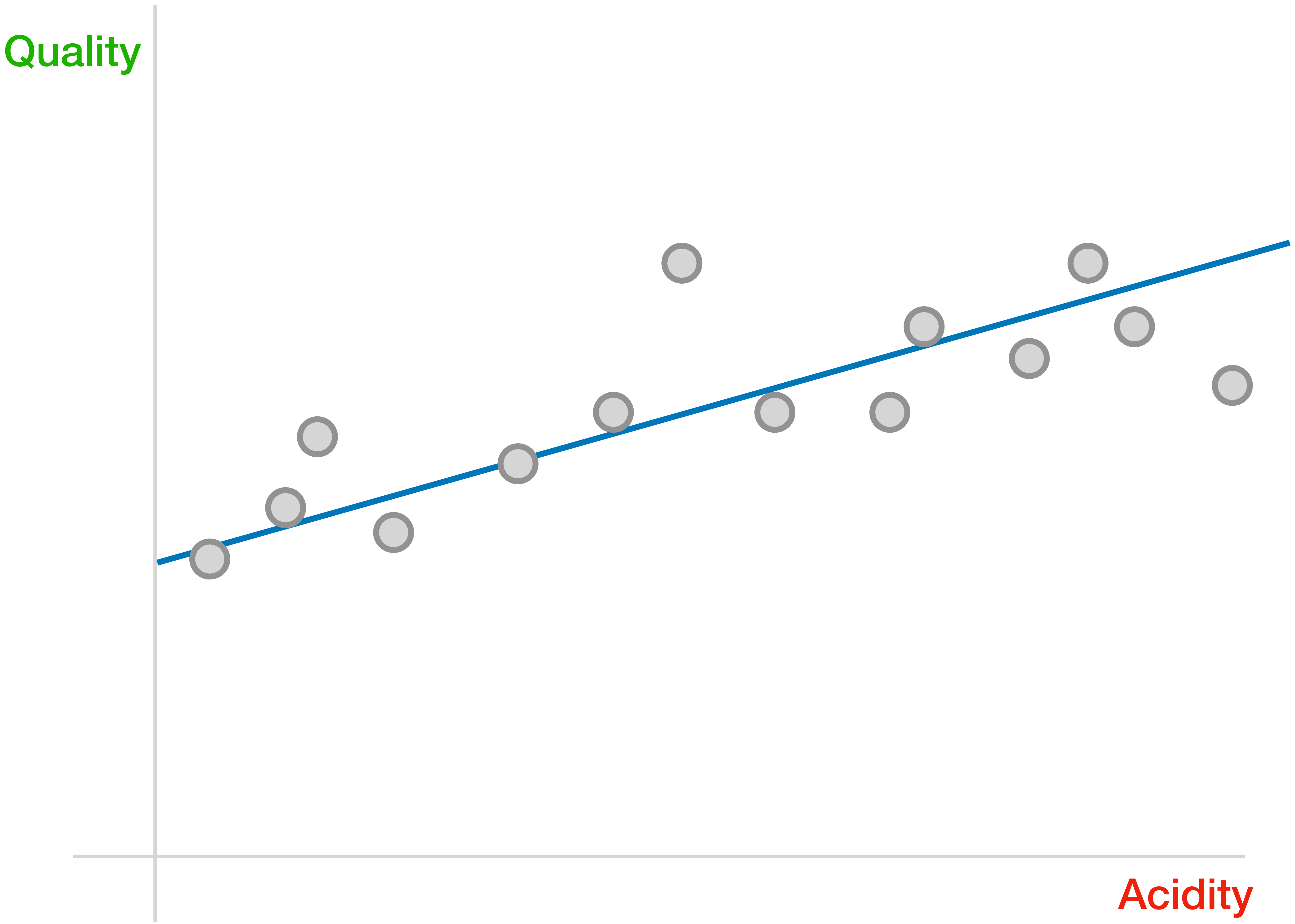
more data

Quality on the y axis
Acidity on the x axis



more data

Quality on the y axis
Acidity on the x axis



- use more data, get more accurate results

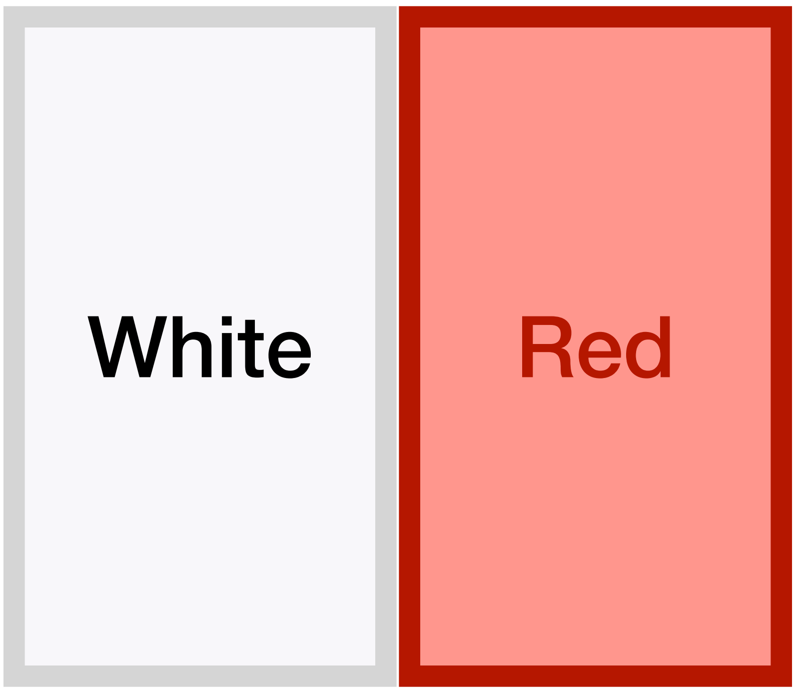
more data **balanced** data **normalized** data more data

balanced data

normalized data

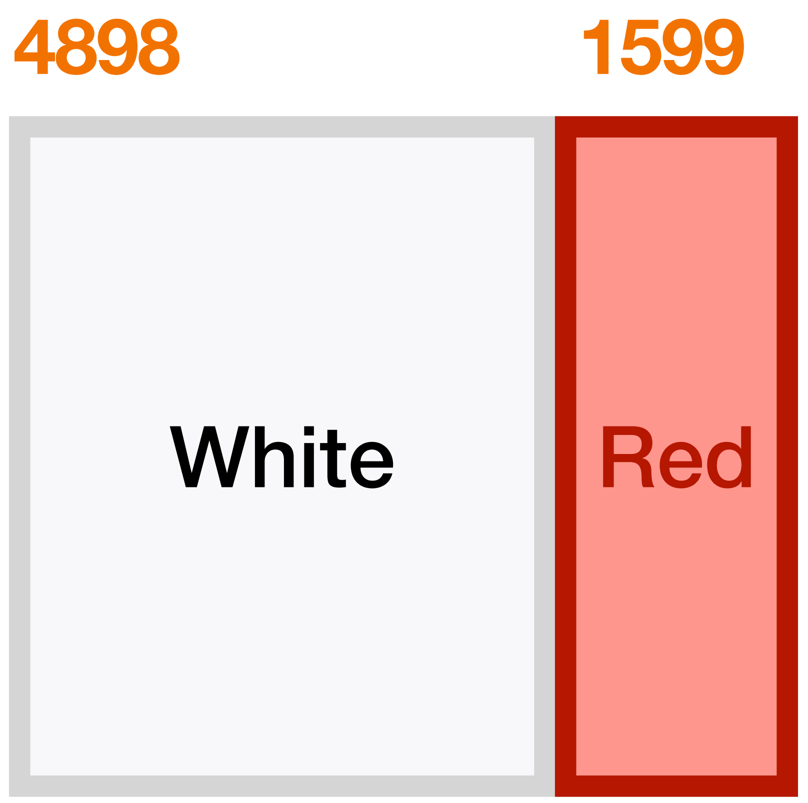
Let's think about logistic functions!

balanced data



in an ideal world

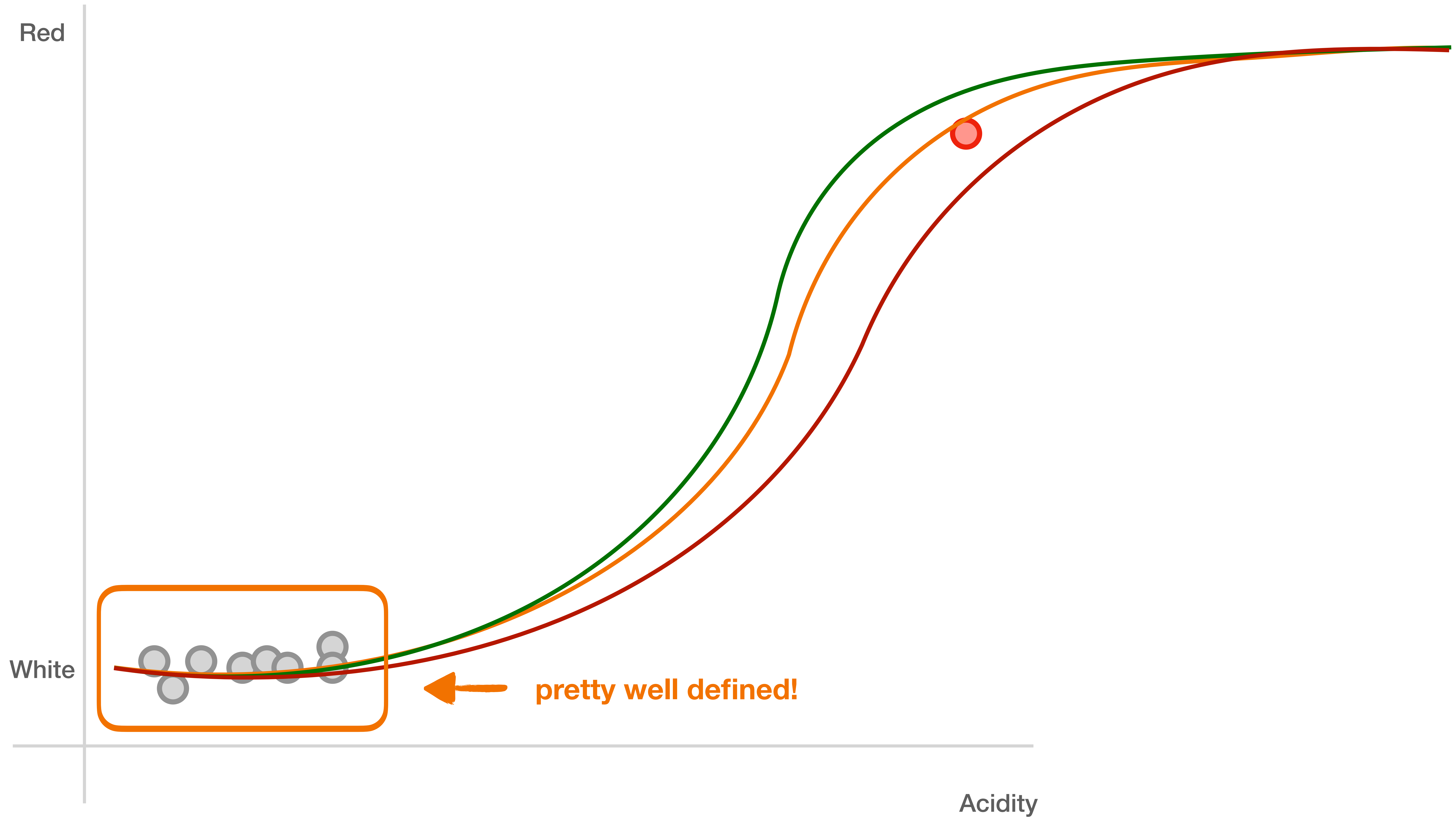
…but no



What happens when we fit this dataset entirely?

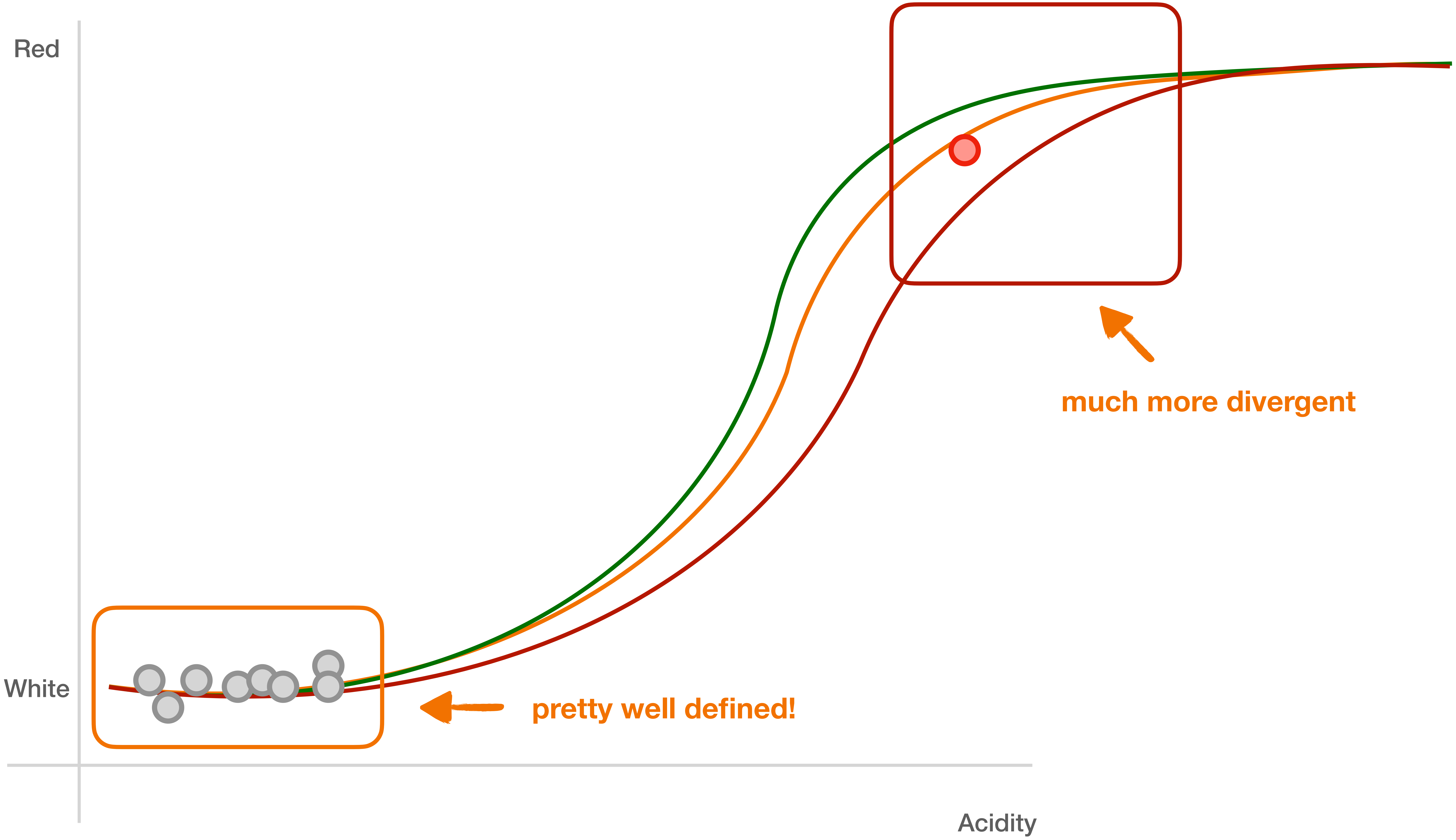
balanced data

Let's think about logistic functions!



balanced data

Let's think about logistic functions!

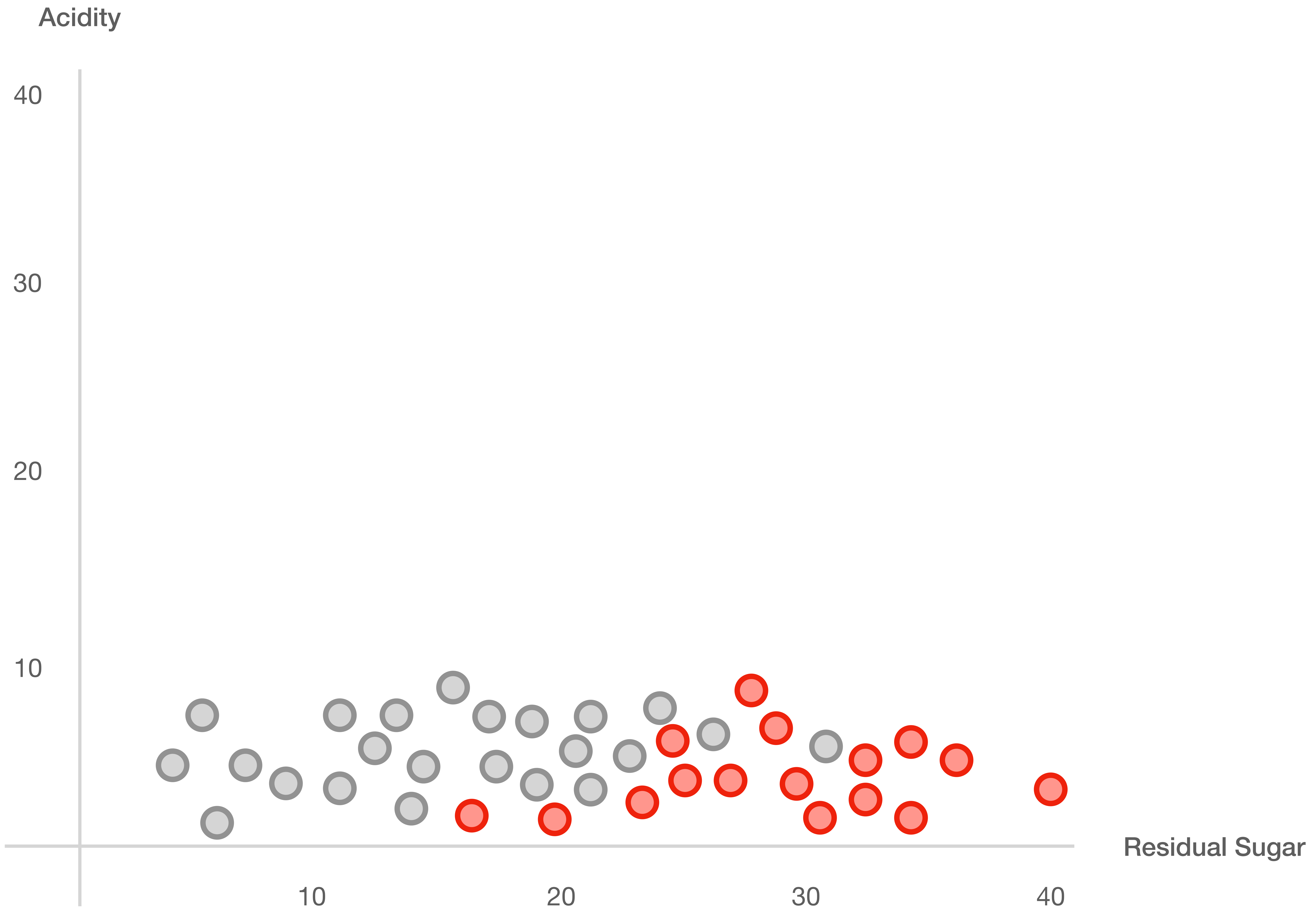


- balanced data, more accurate results

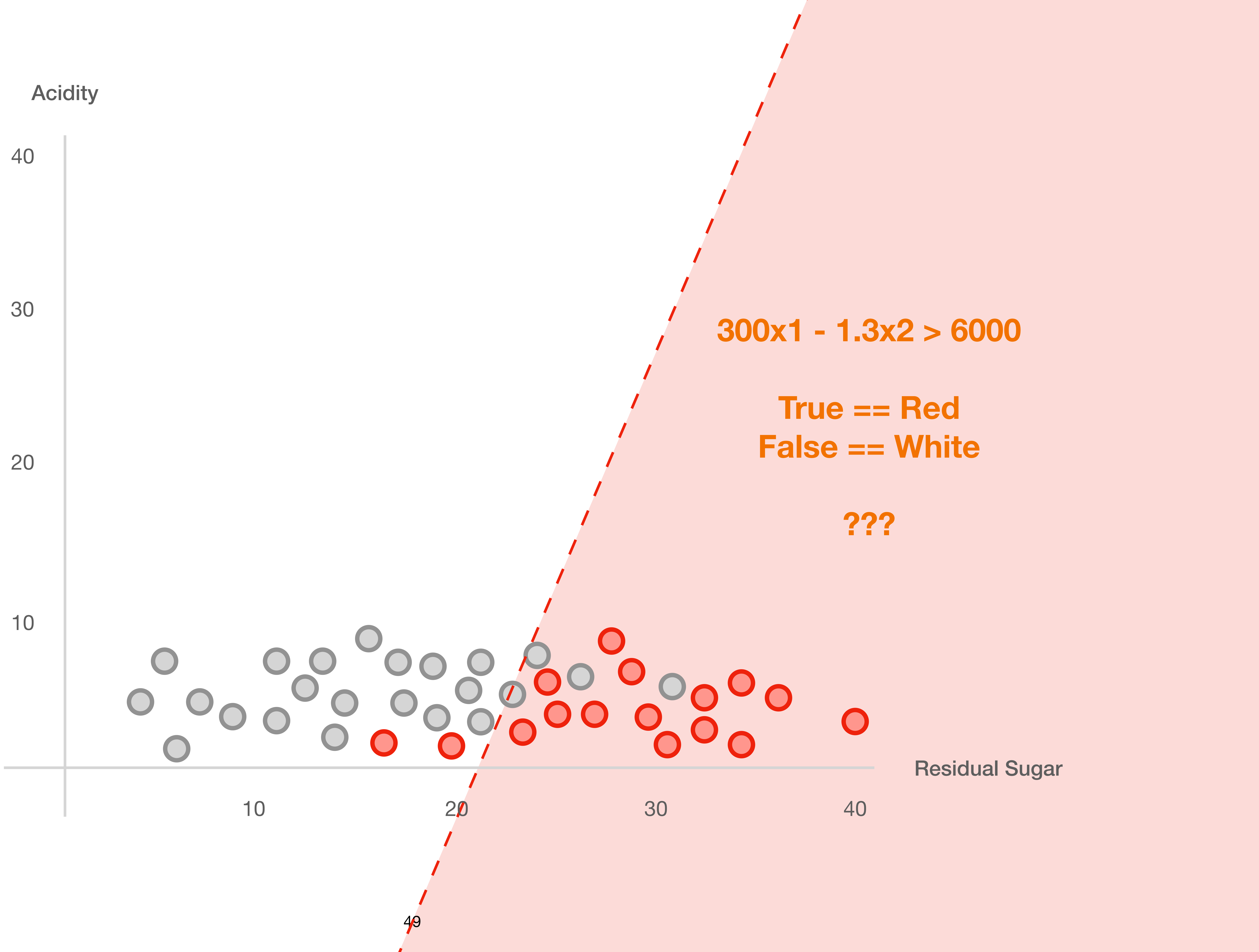
more data **balanced data** normalized data

more data **balanced data** **normalized data**

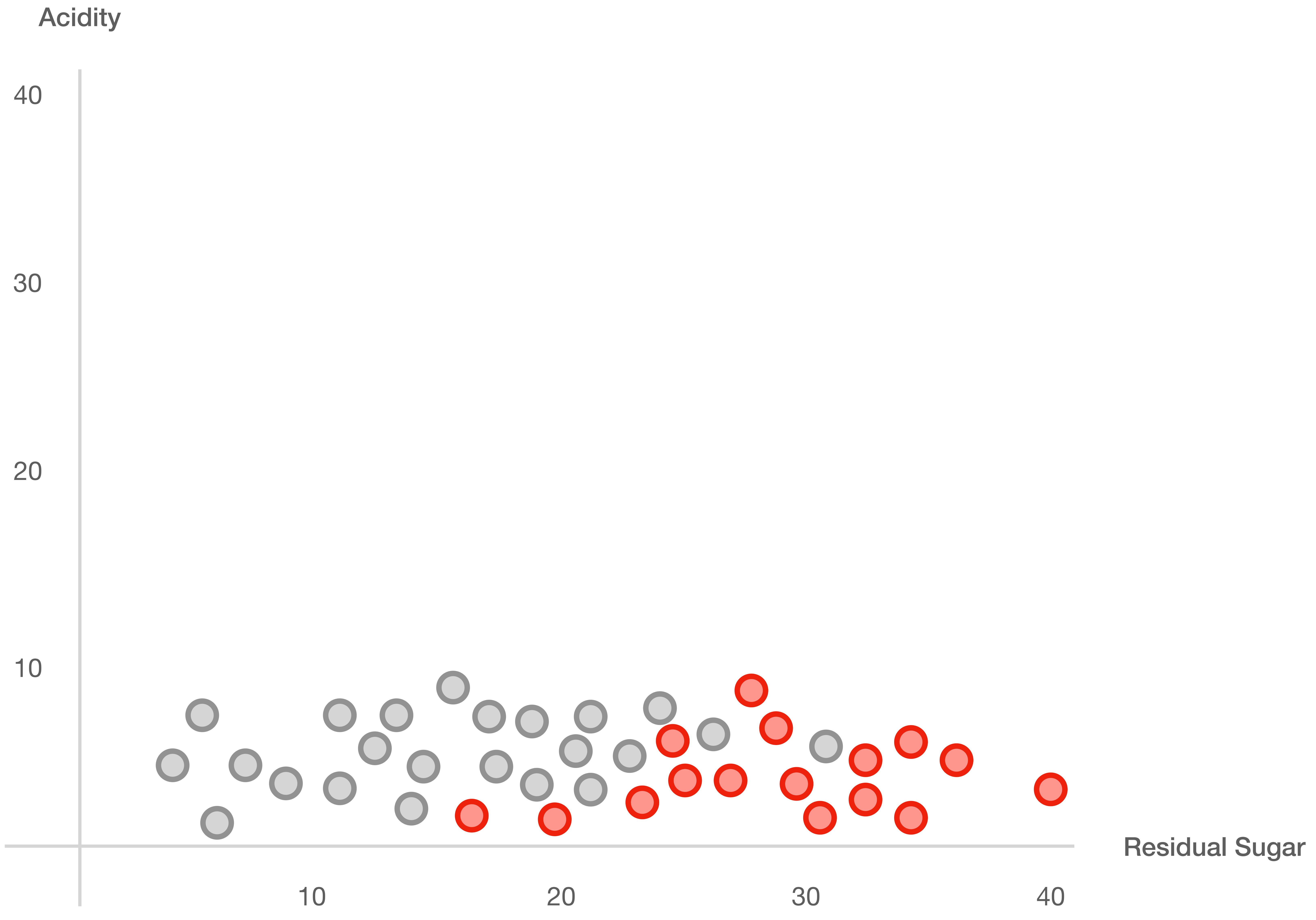
normalized data



normalized data



normalized data



normalized data

- normalized data, better generalization, faster convergence more data balanced data **normalized data**

more data

balanced data

normalized data

let's clean some data!