AIBridge

Lecture 5

Let’s talk about the last lab!

Let’s talk about the last lab!

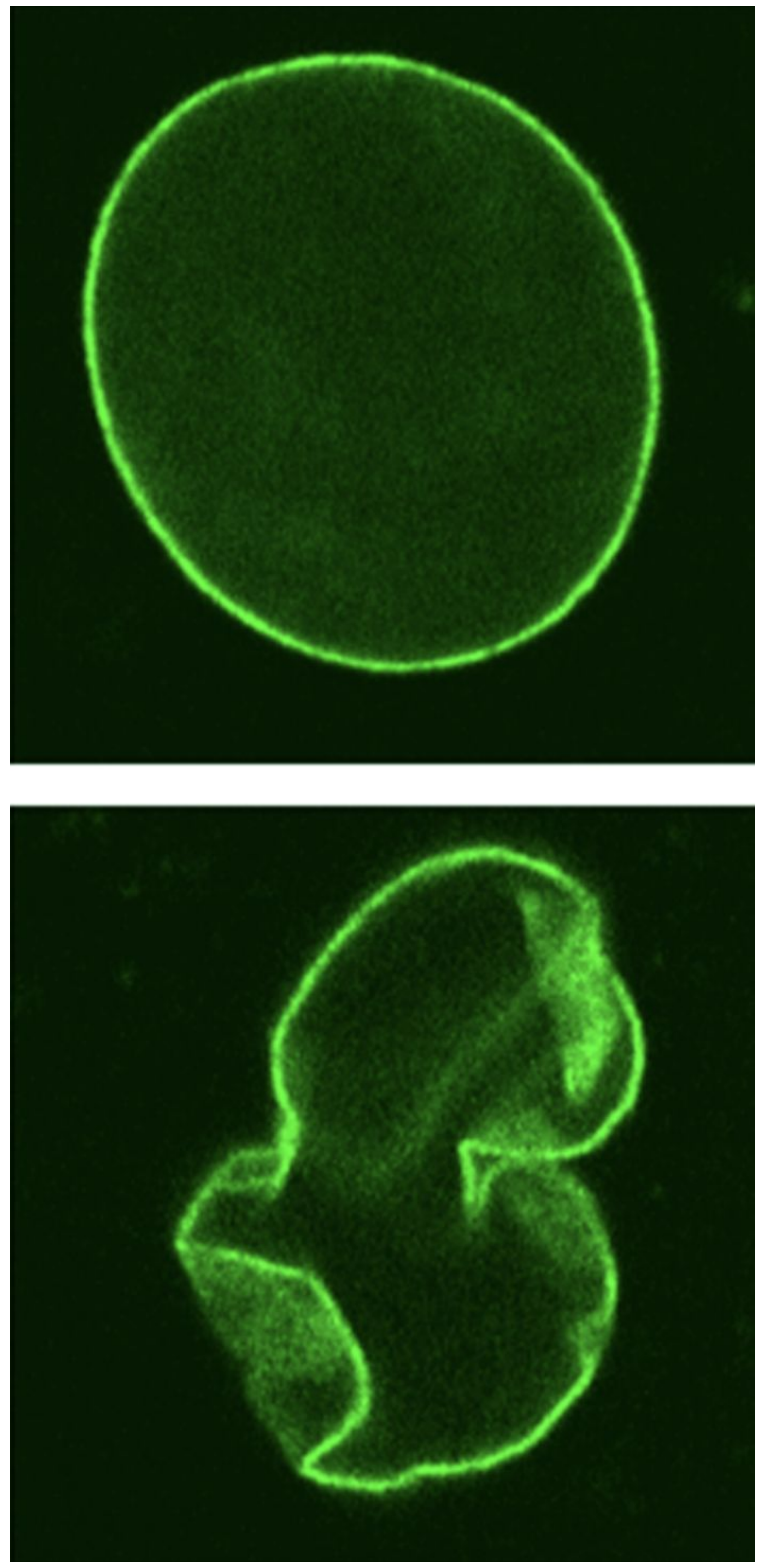
what does this even mean?

What circumstances made the model fit better?

worse?

Accuracy

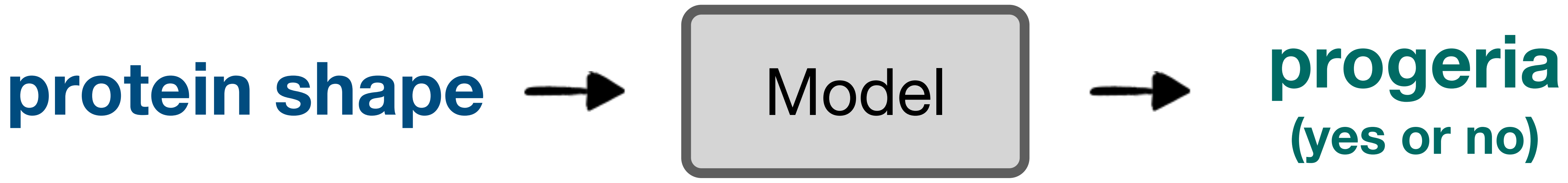
“Why is it not enough?”



Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

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



protein shape

Model

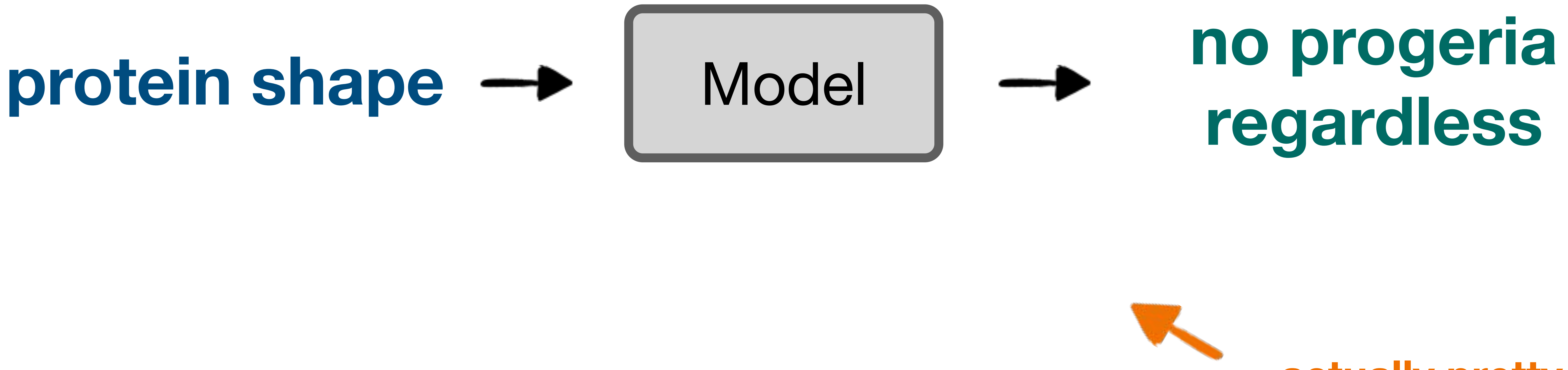
progeria

(yes or no)

Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

a proposed model:



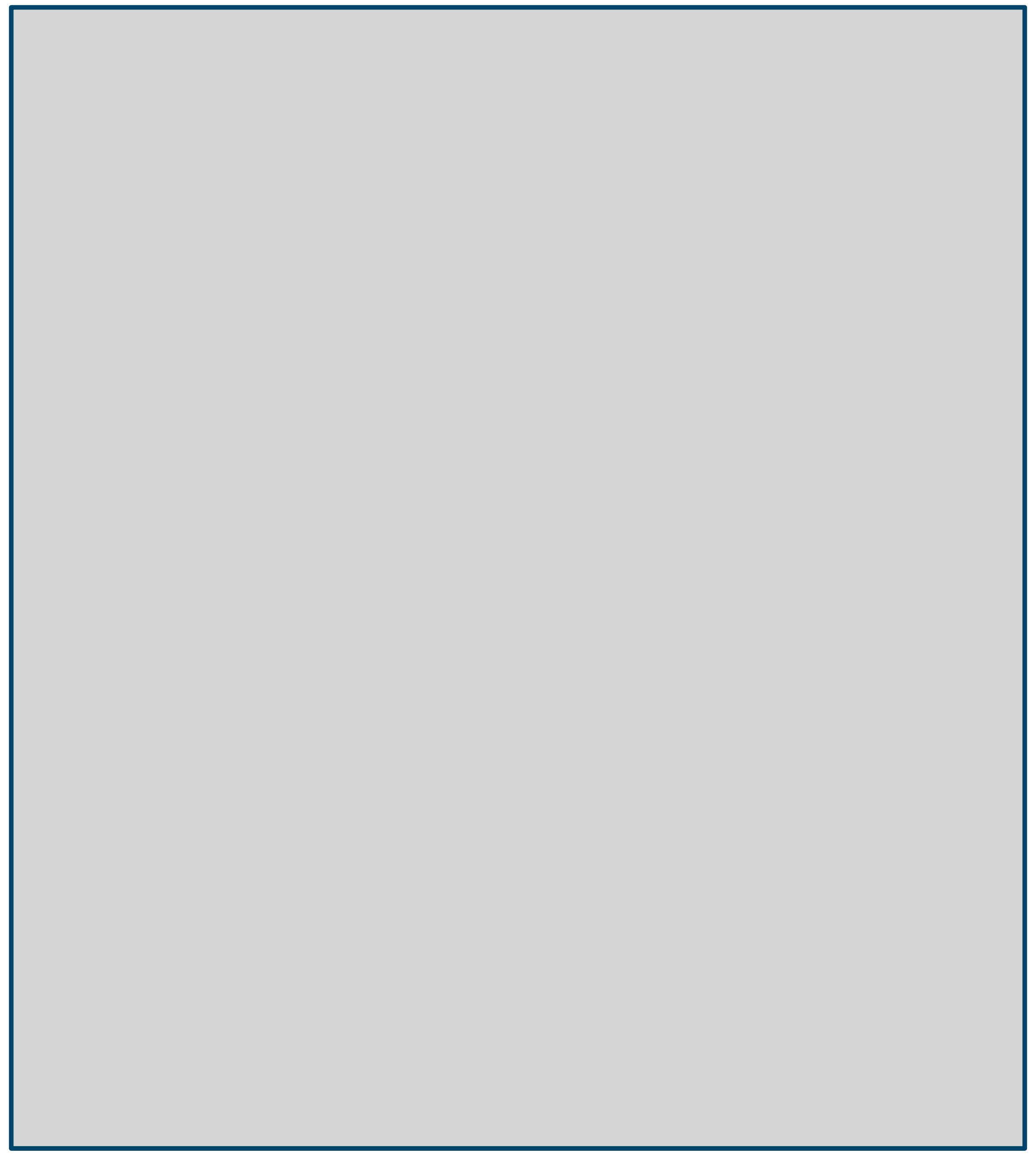
protein shape → Model → no progeria regardless

actually pretty accurate!

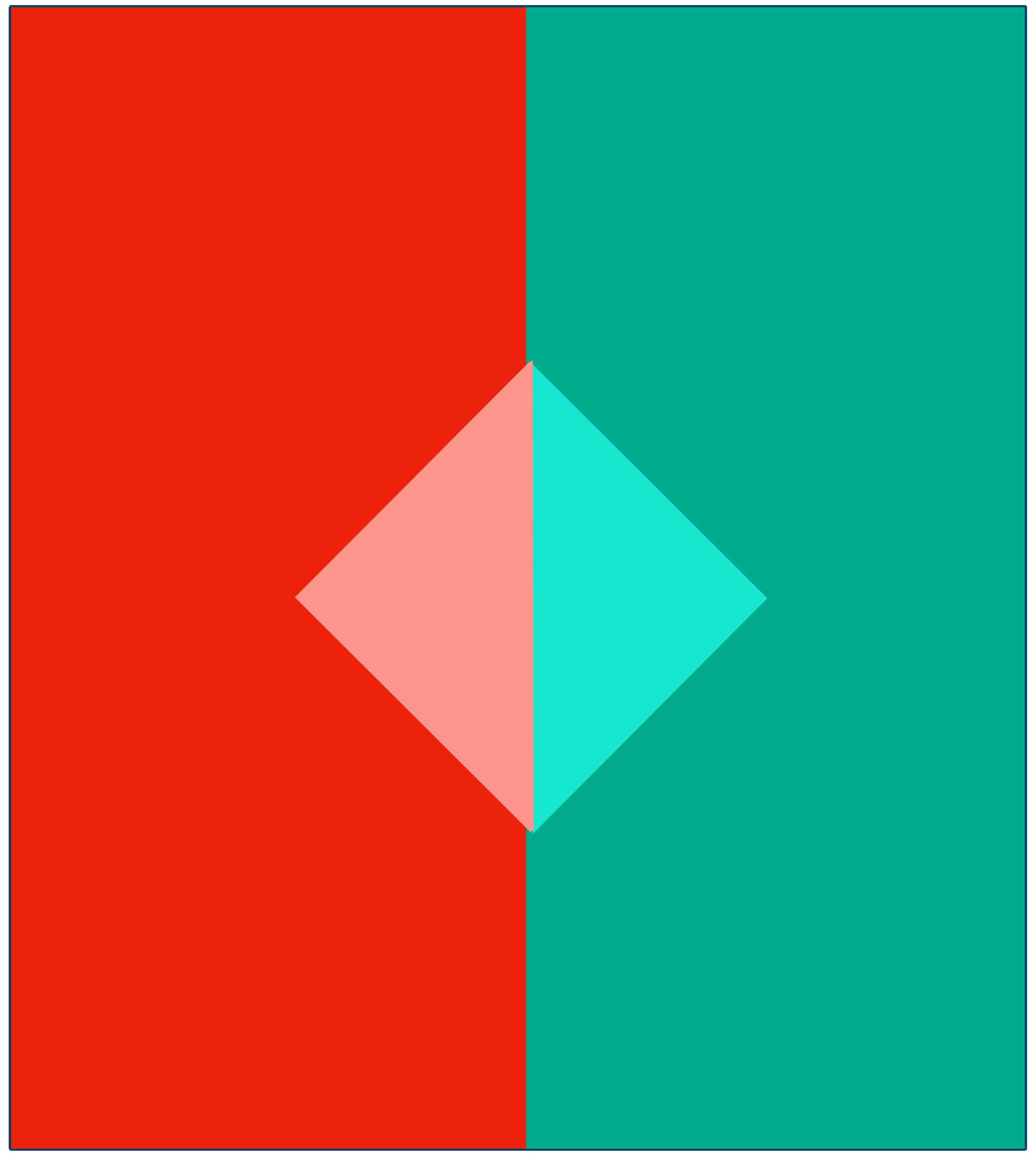
Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

Accuracy , Precision, and Recall

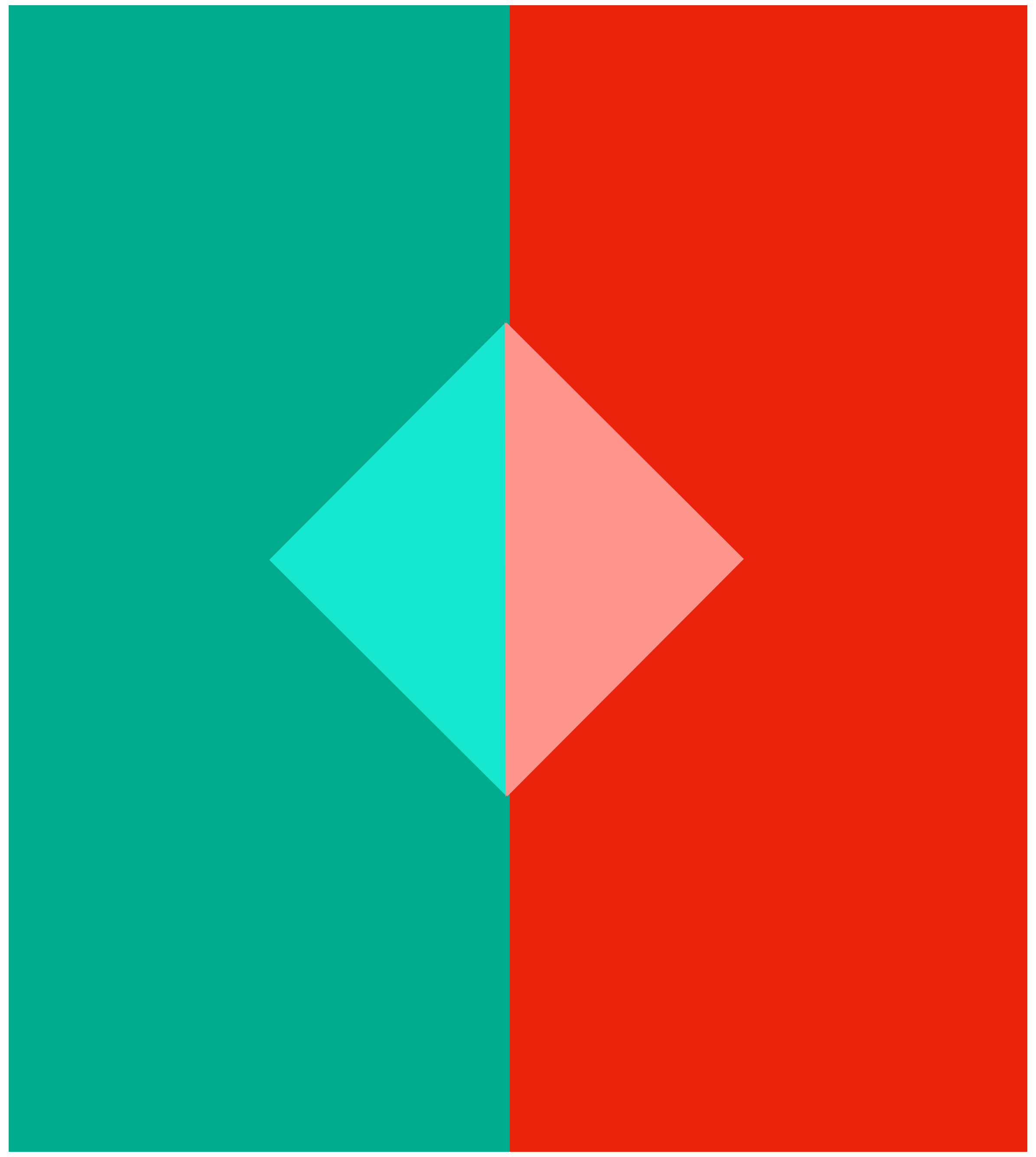


“Selection space”



“Selection space”

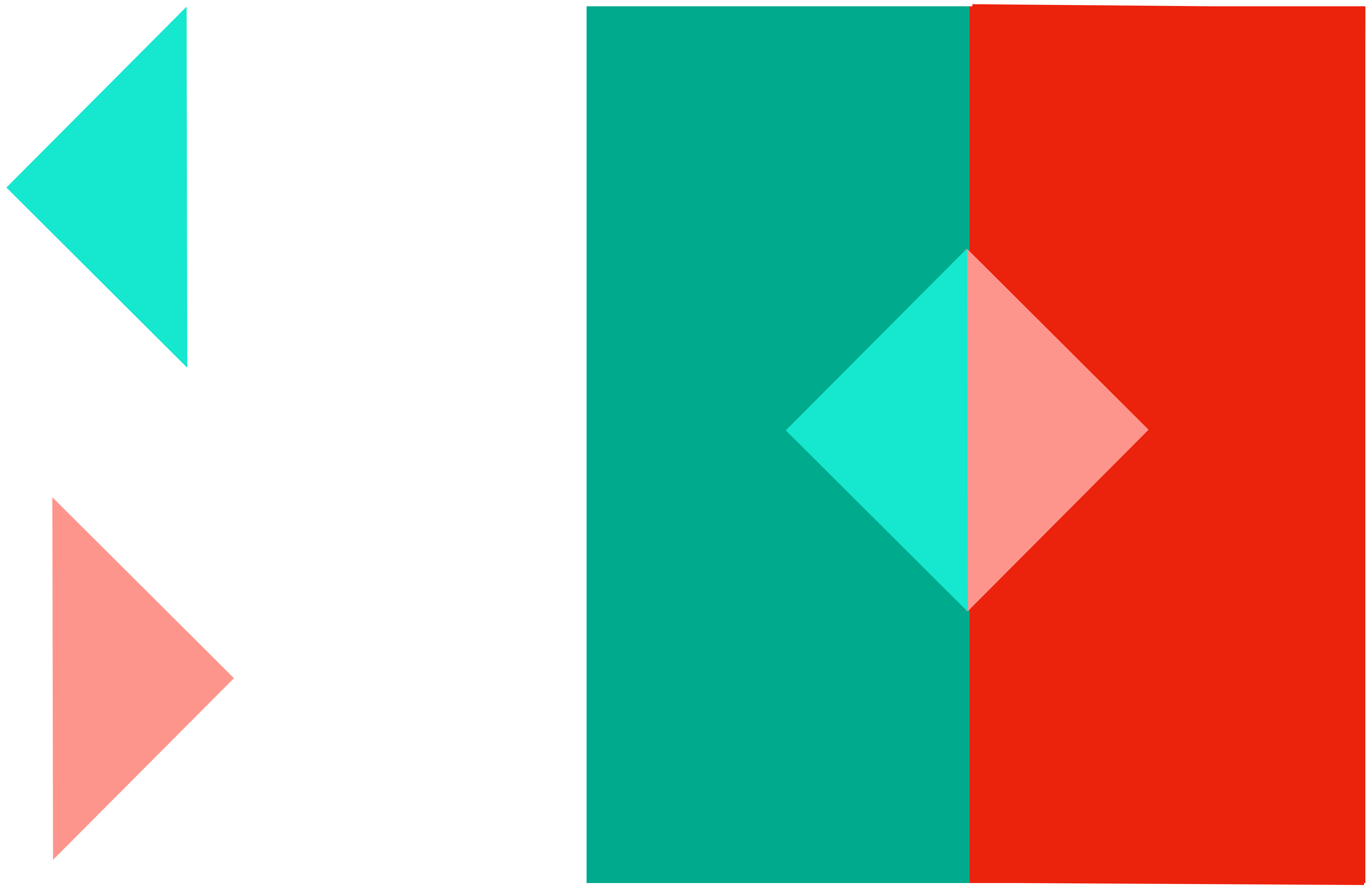
Model selects **positive** and patient is **positive** ◀



“Selection space”

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

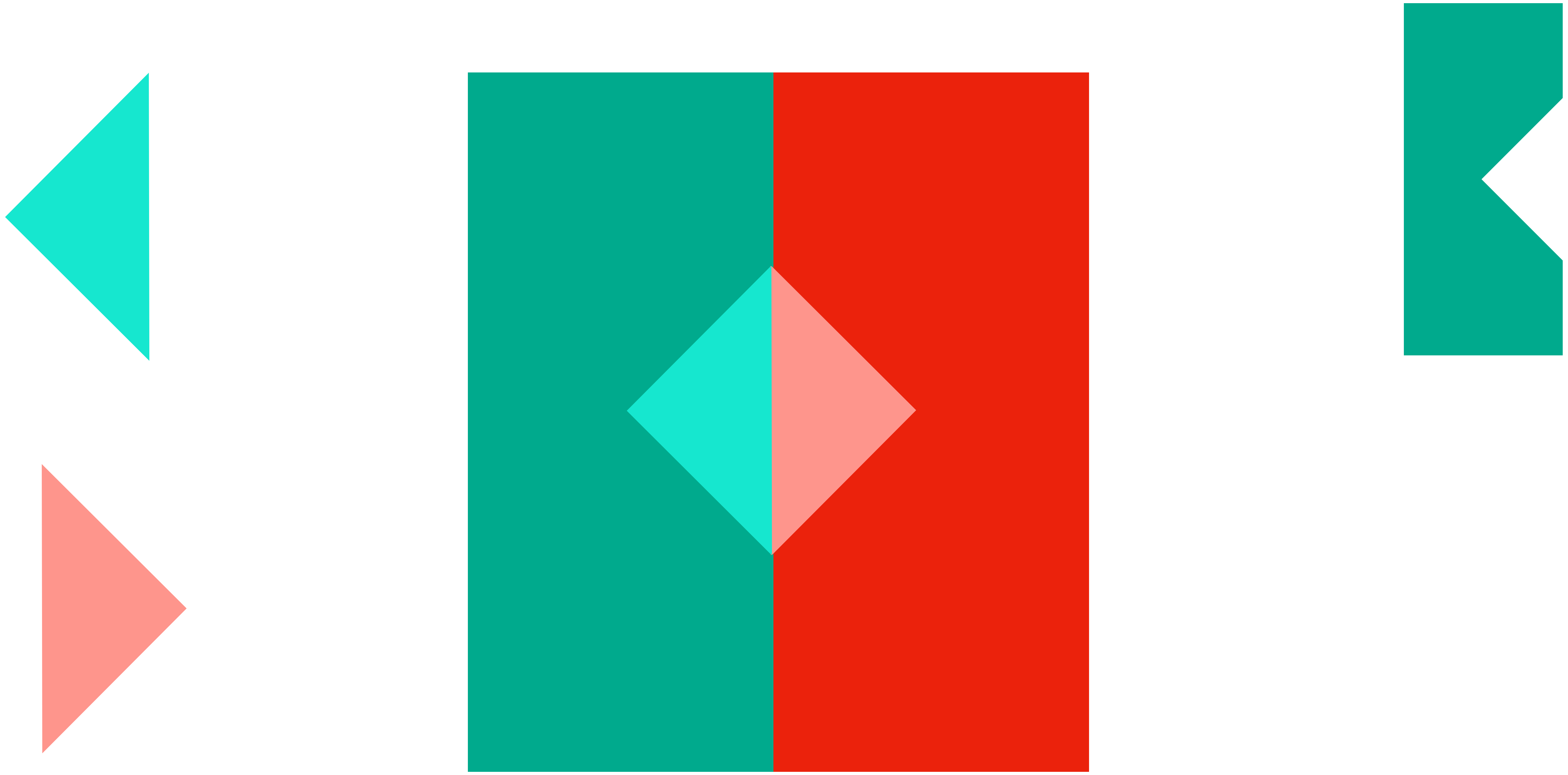
Model selects **positive** and patient is **negative**



“Selection space”

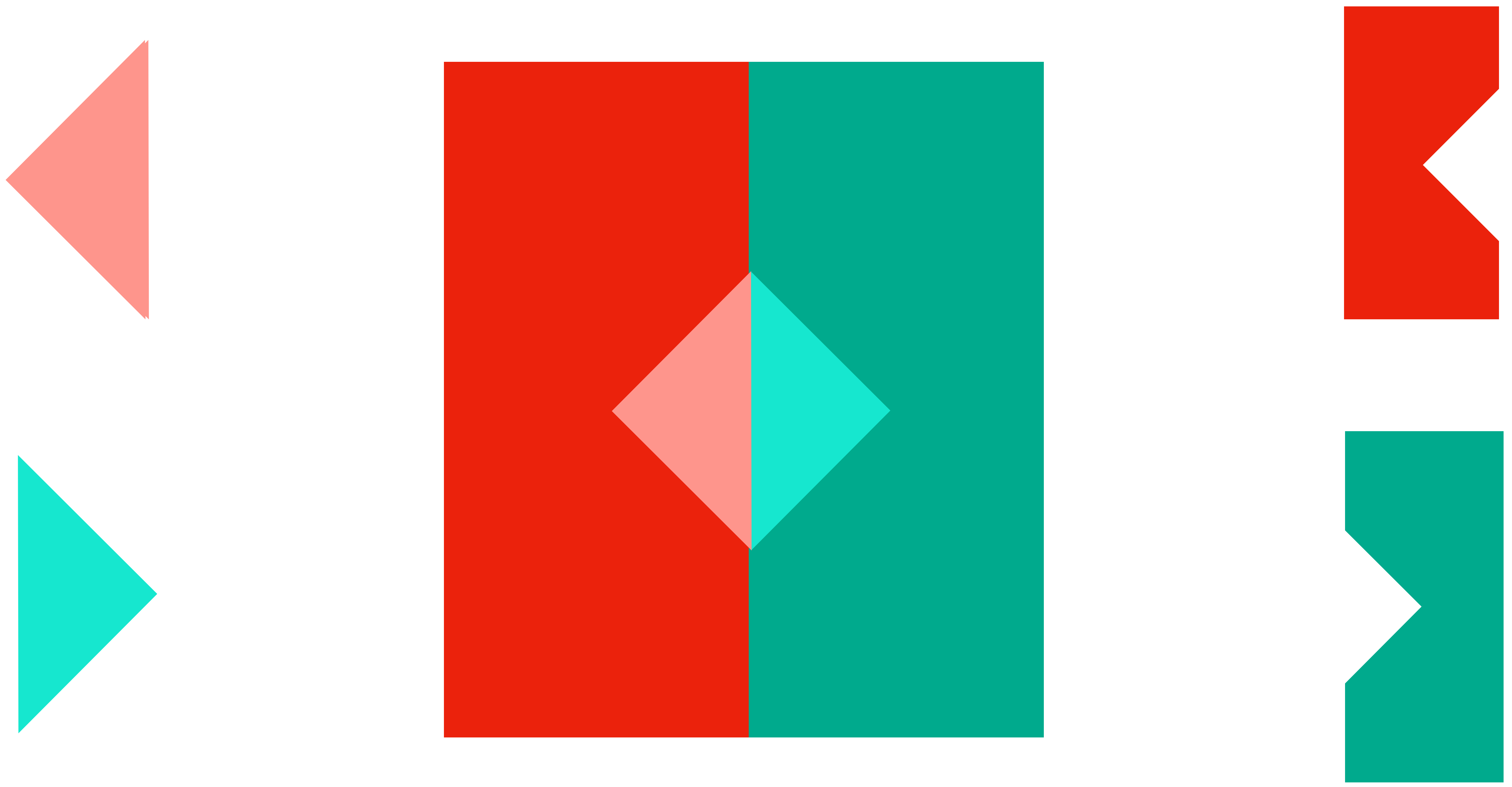
Model selects **positive** and patient is **positive**

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



“Selection space”

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



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

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

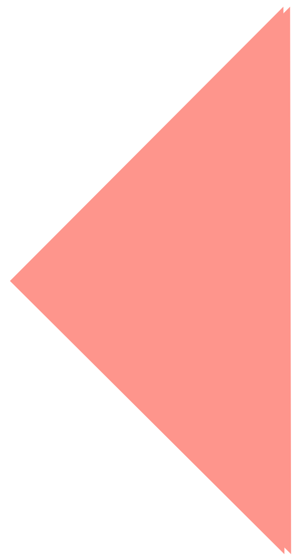
“Selection space”

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

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

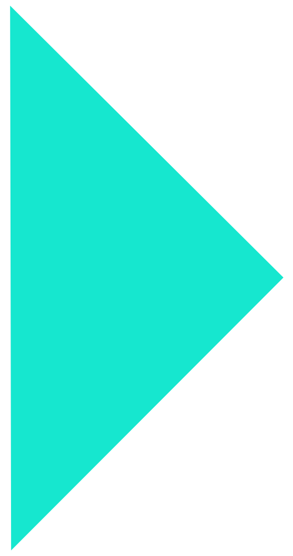
TRUE **P**OSITIVE

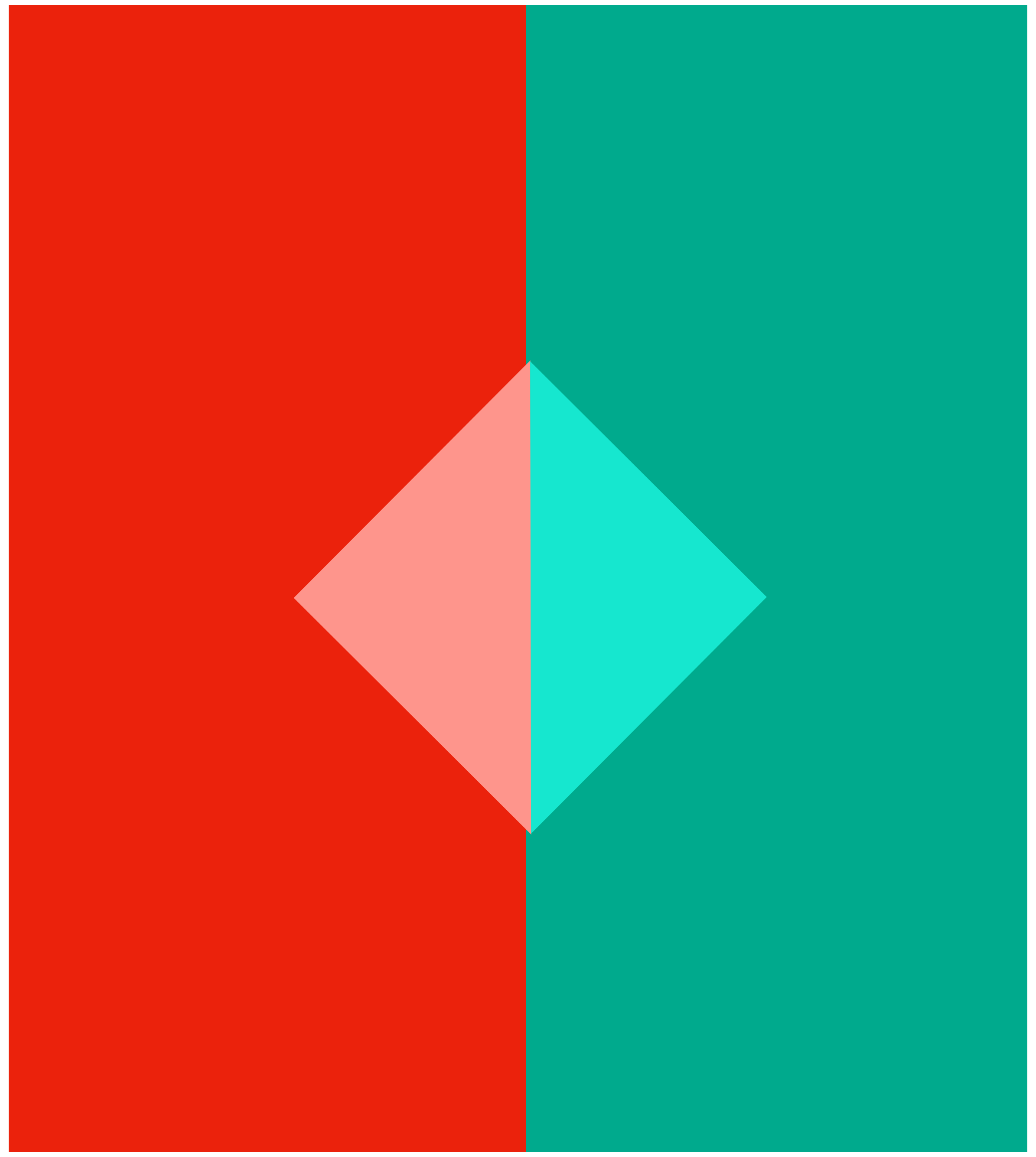
TP: Model selects **positive** and patient is **positive**



FALSE **P**OSITIVE

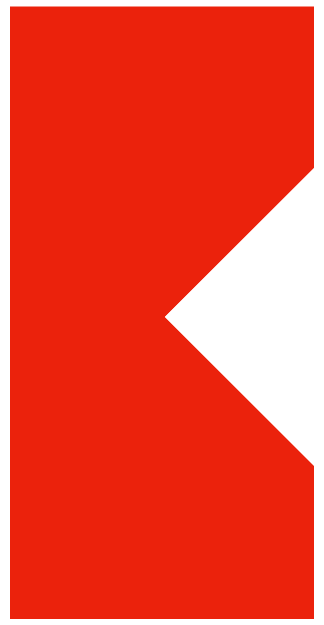
FP: Model selects **positive** and patient is **negative**





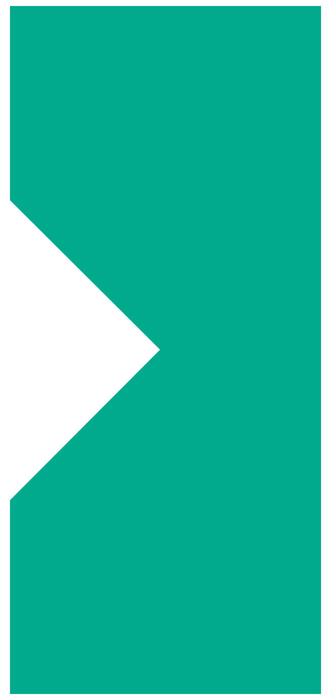
“Selection space”

FALSE **N**EGATIVE



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

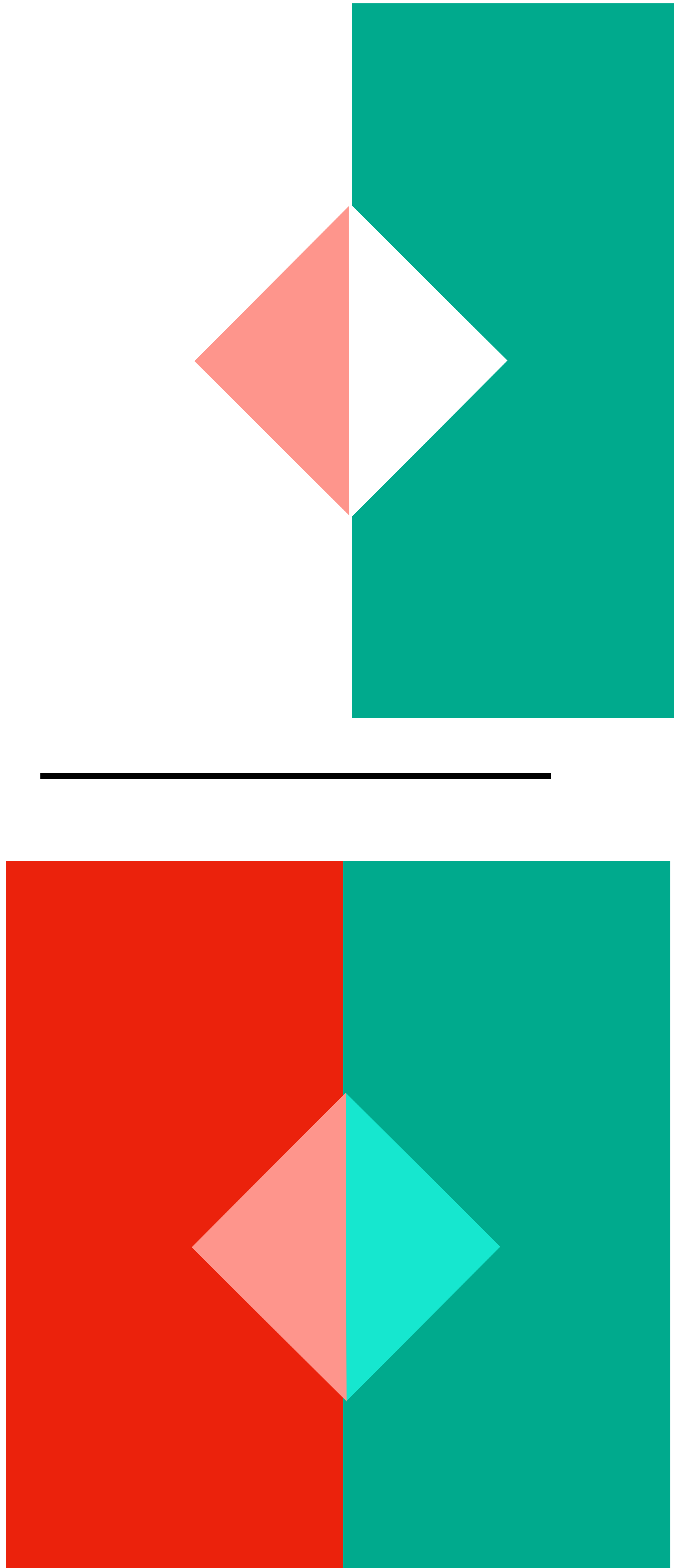
TRUE **N**EGATIVE



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

Accuracy

Overall ability of model

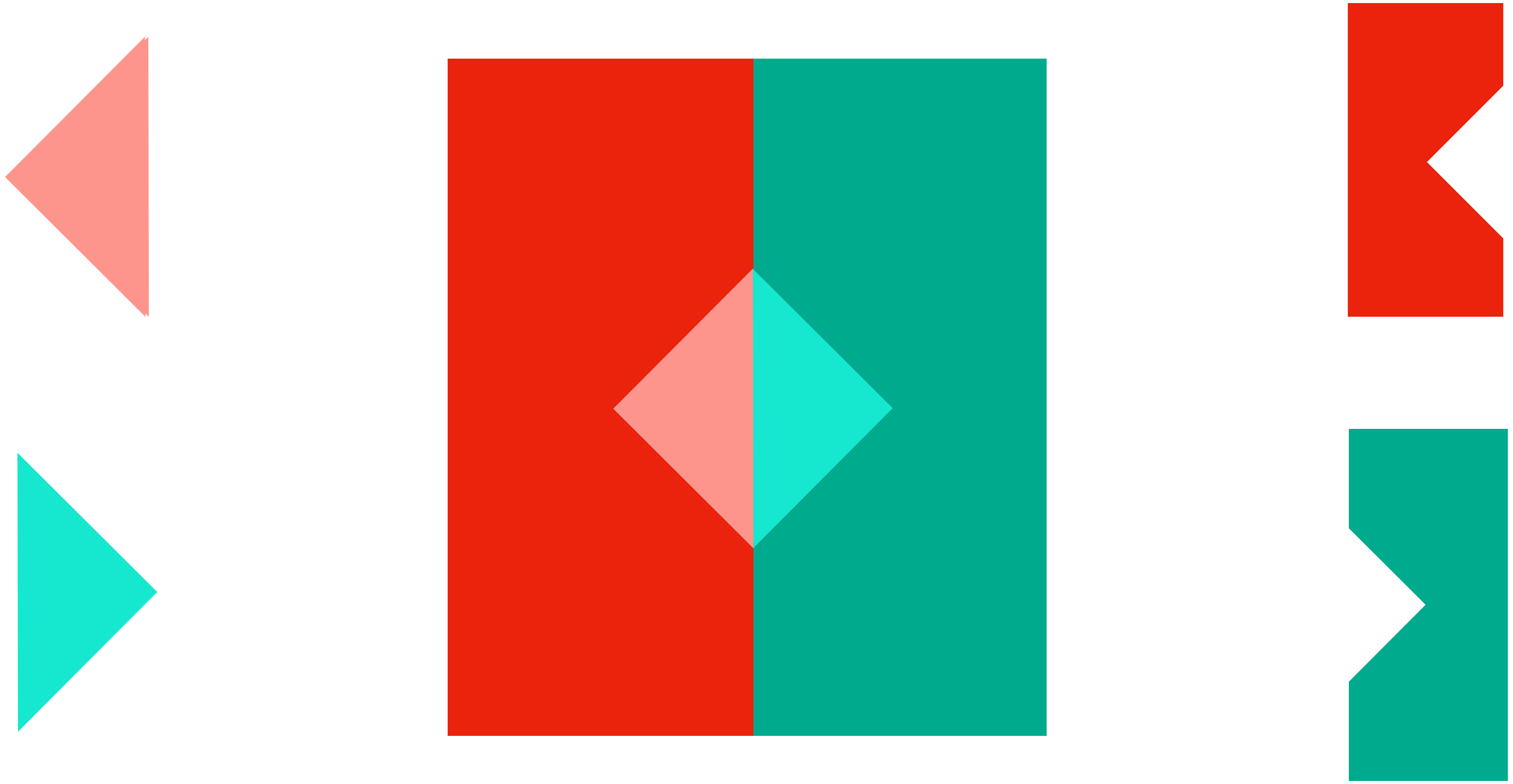


“Number of cases where **we chose** **positive** **when patient is** **positive**

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**

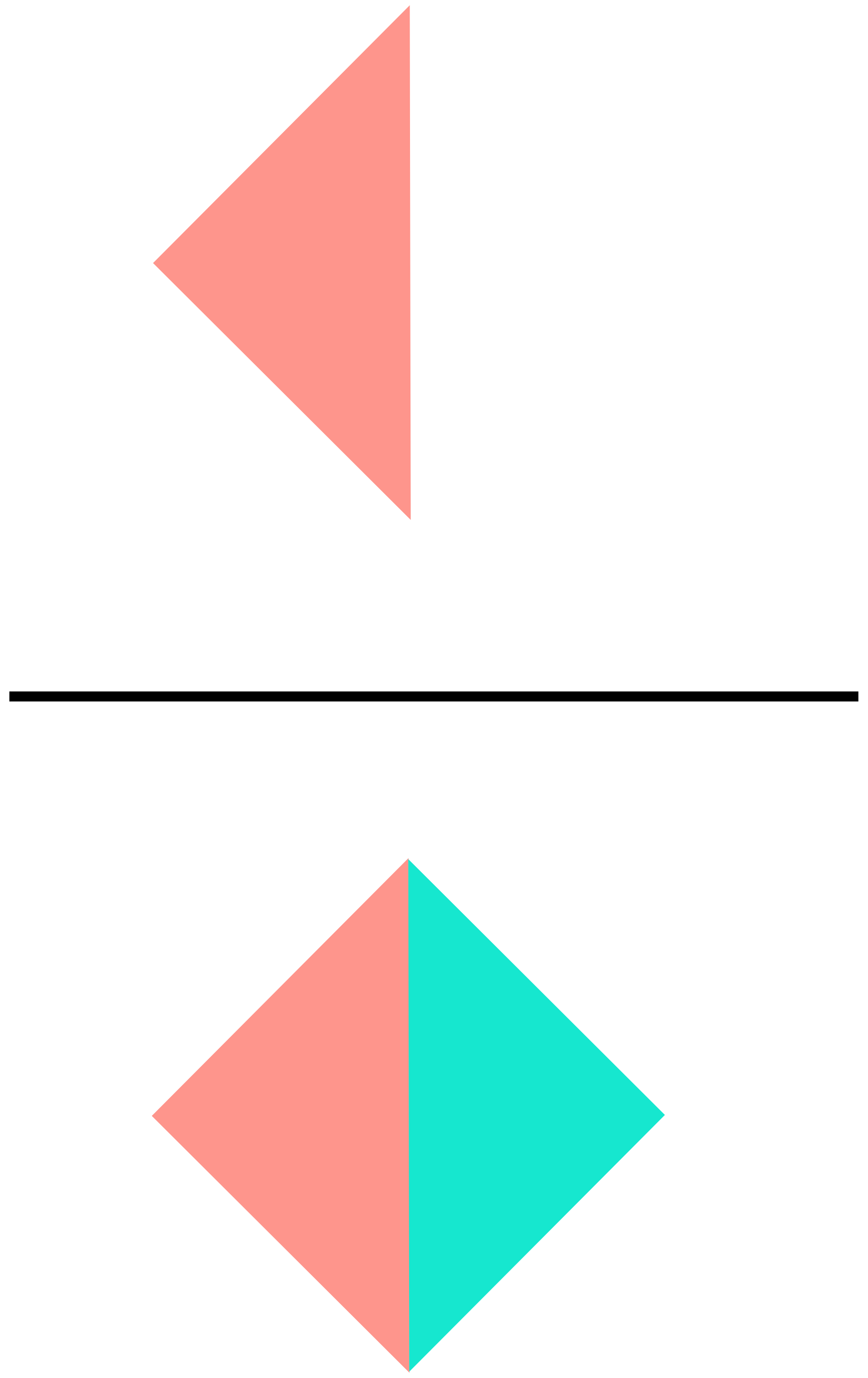
TN: Model selects **negative** and patient is **negative**

Accuracy

Overall ability of model

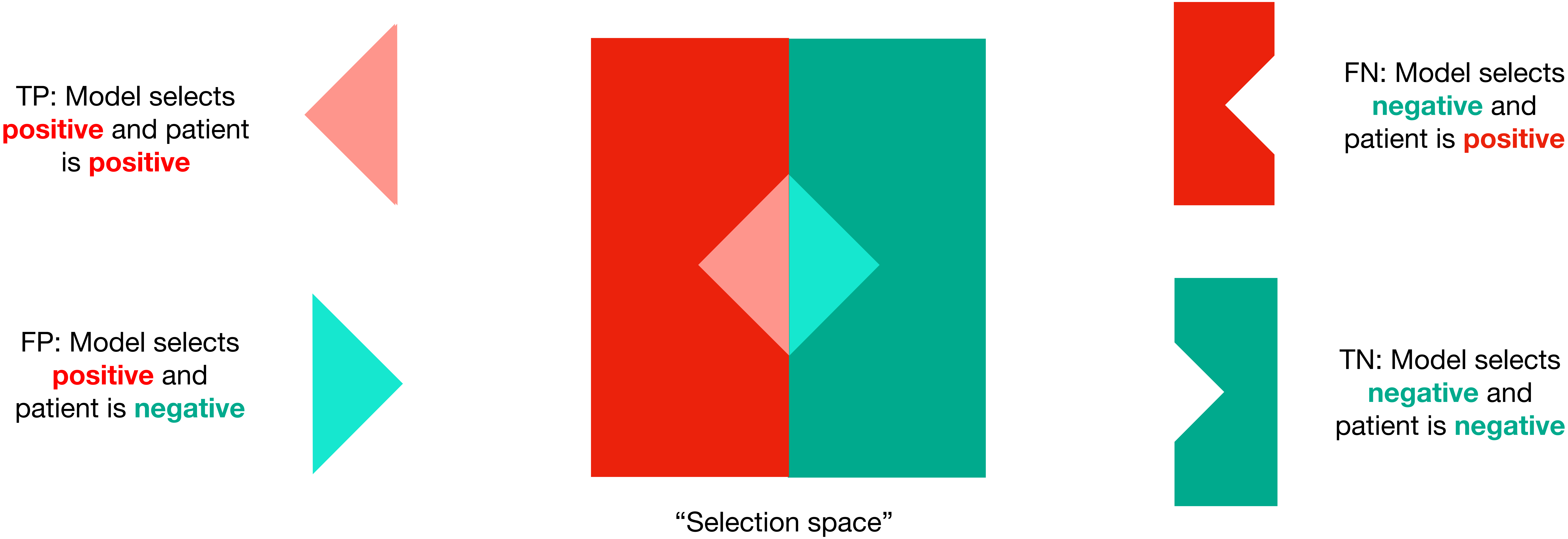
Precision

Amount of selection that's actually correct.



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

"All selected **positive** by the model"



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

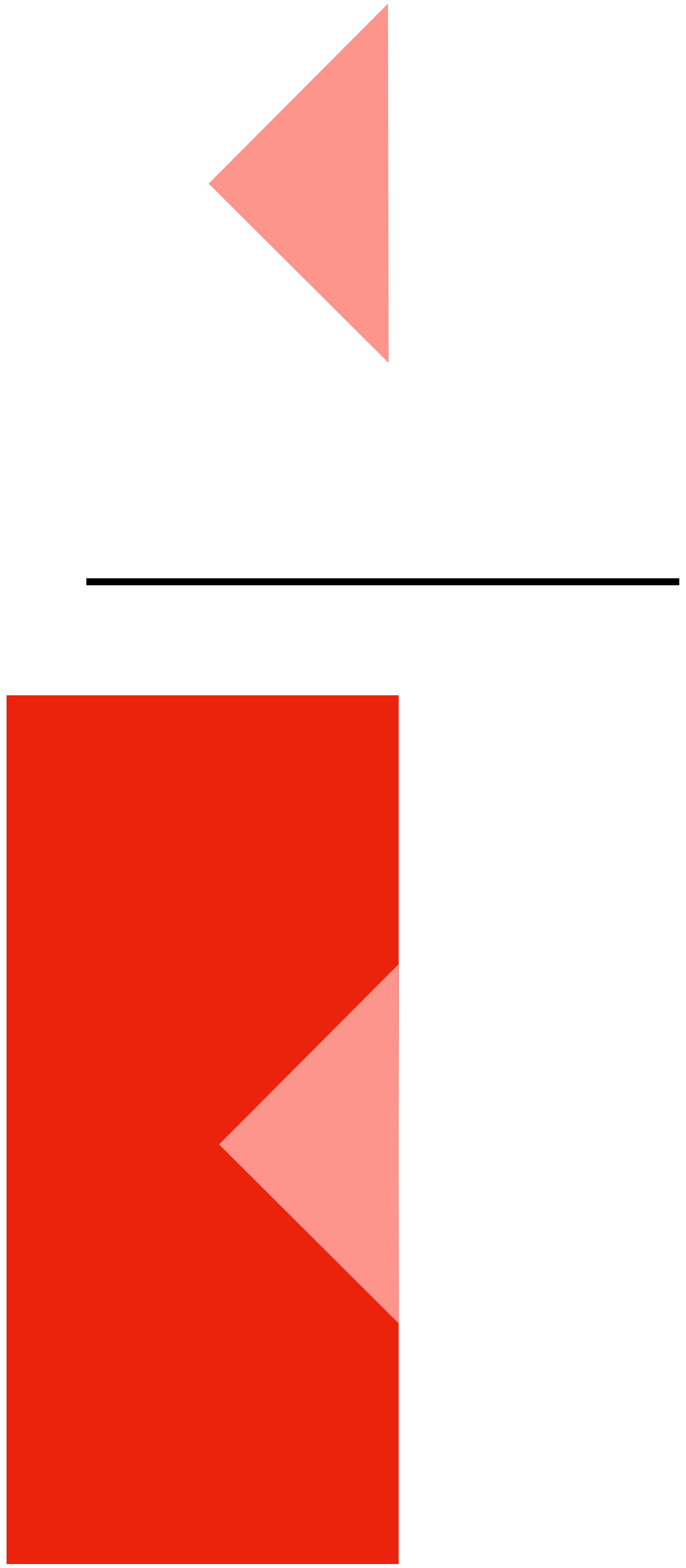
Overall ability of model

Precision

Amount of selection that’s actually correct.

Recall

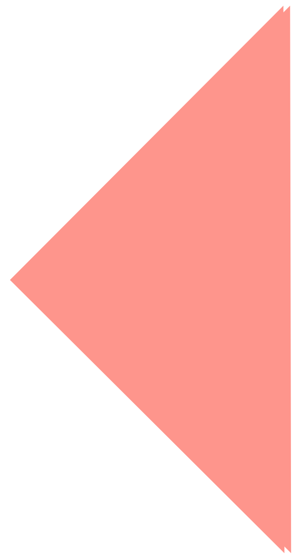
Amount of what needs to be selected that is selected



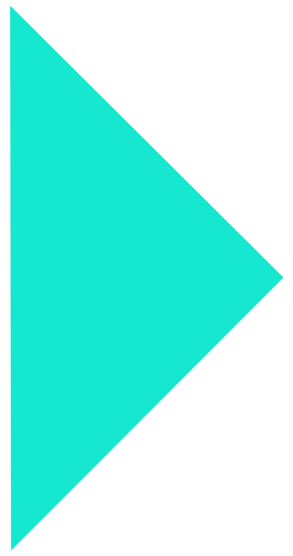
“Number of cases where **we chose** **positive** **when patient is** **positive**”

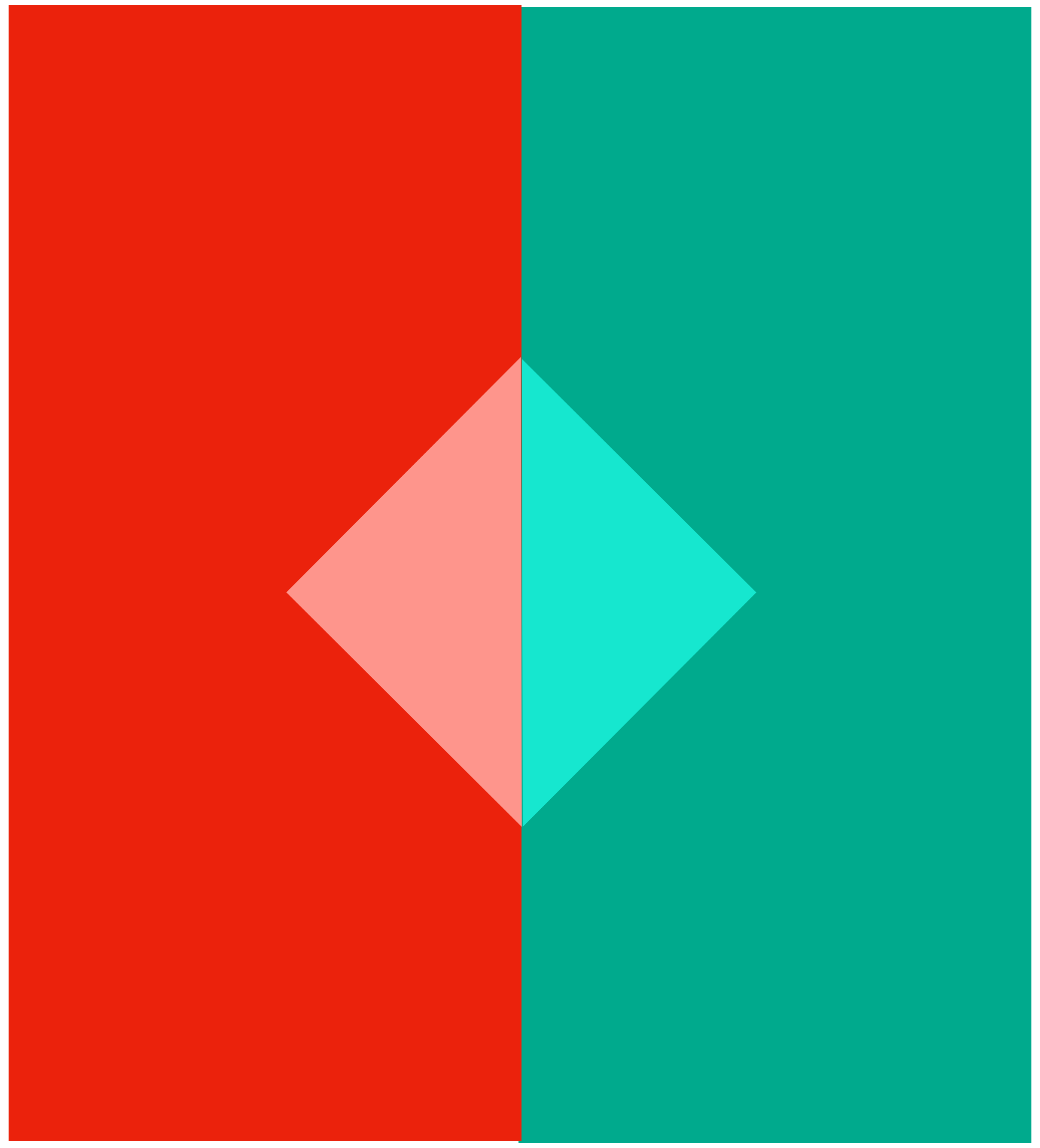
“All cases that the patients are **positive**”

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

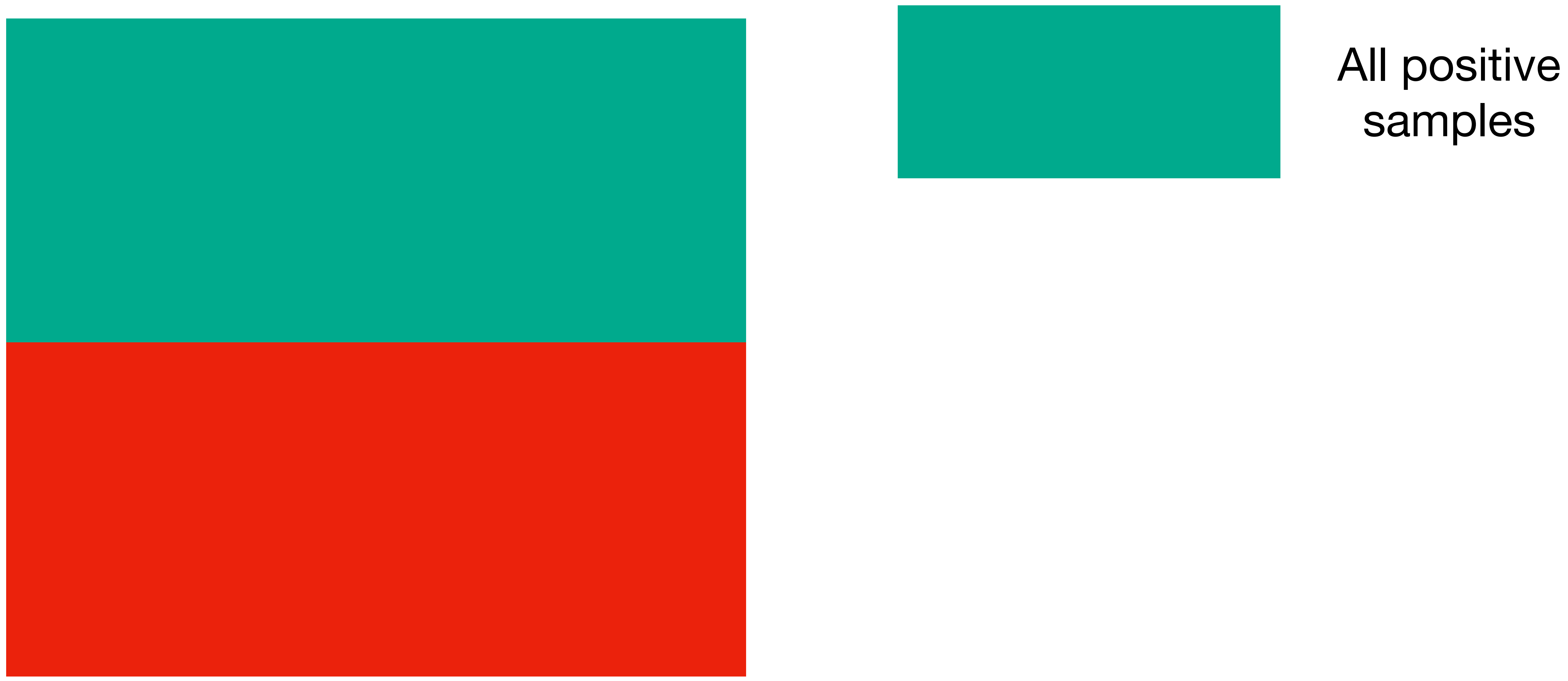
Precision

Amount of selection that’s actually correct.

Recall

Amount of what needs to be selected that is selected

“Selection space”

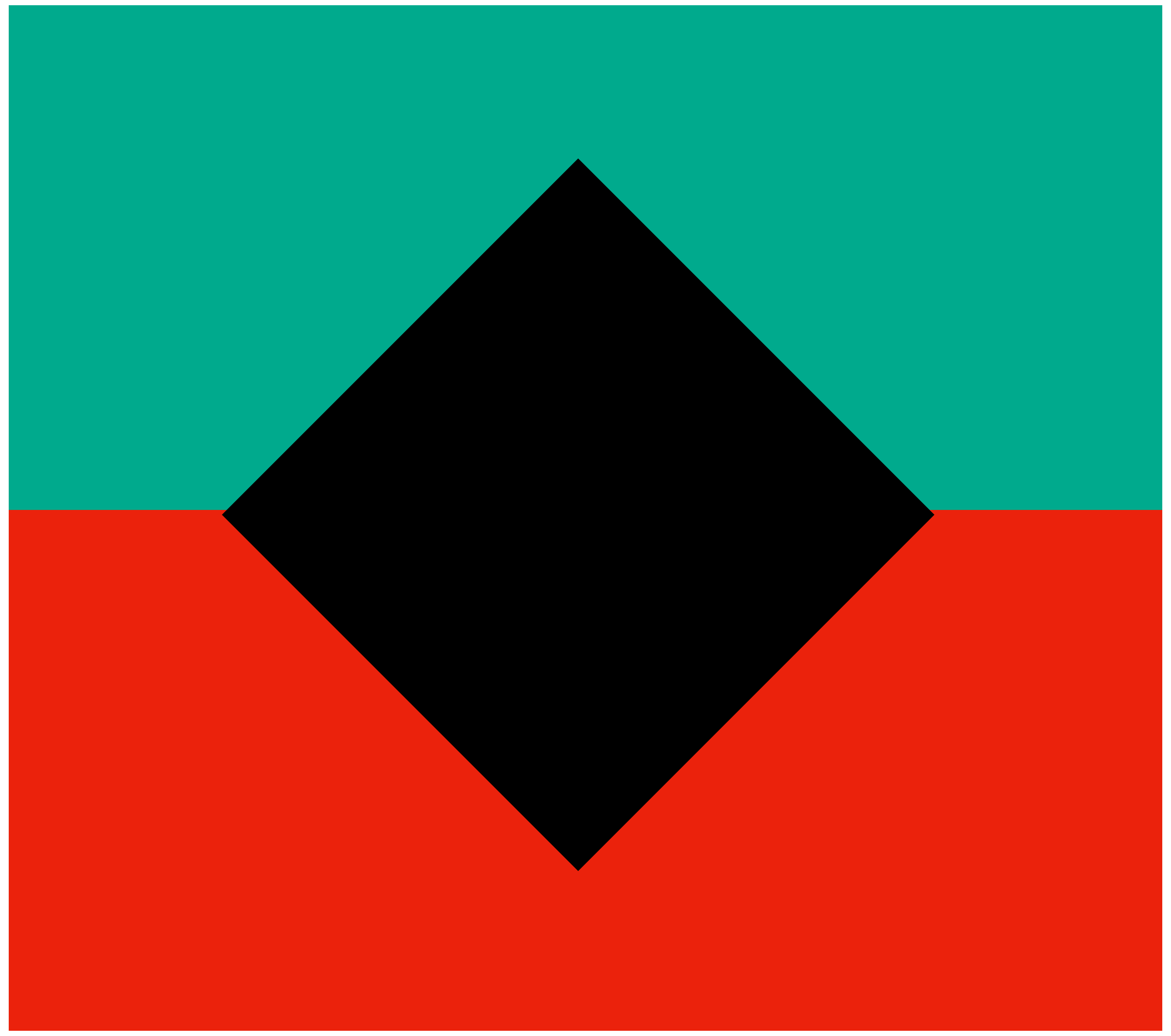


“Selection space”

🟩 All positive samples

🟥 All negative samples

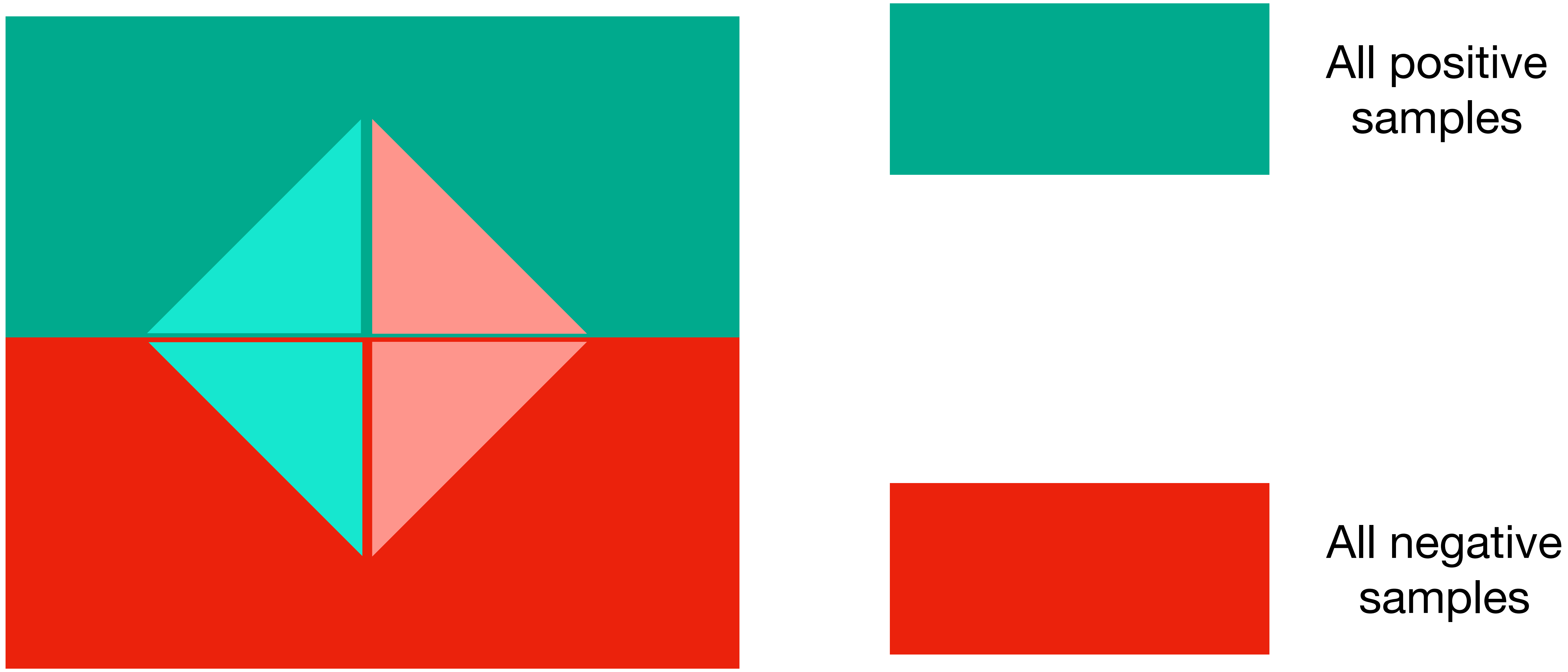
“Selection space”



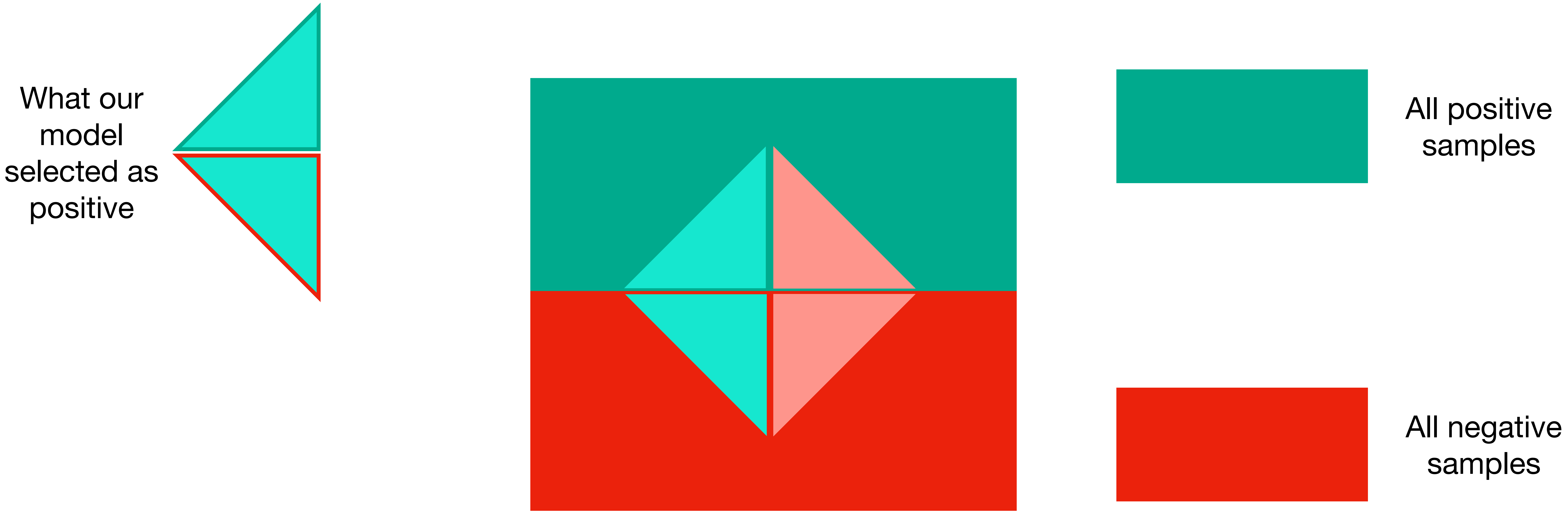
🟩 All positive samples

🟥 All negative samples

“Selection space”



“Selection space”

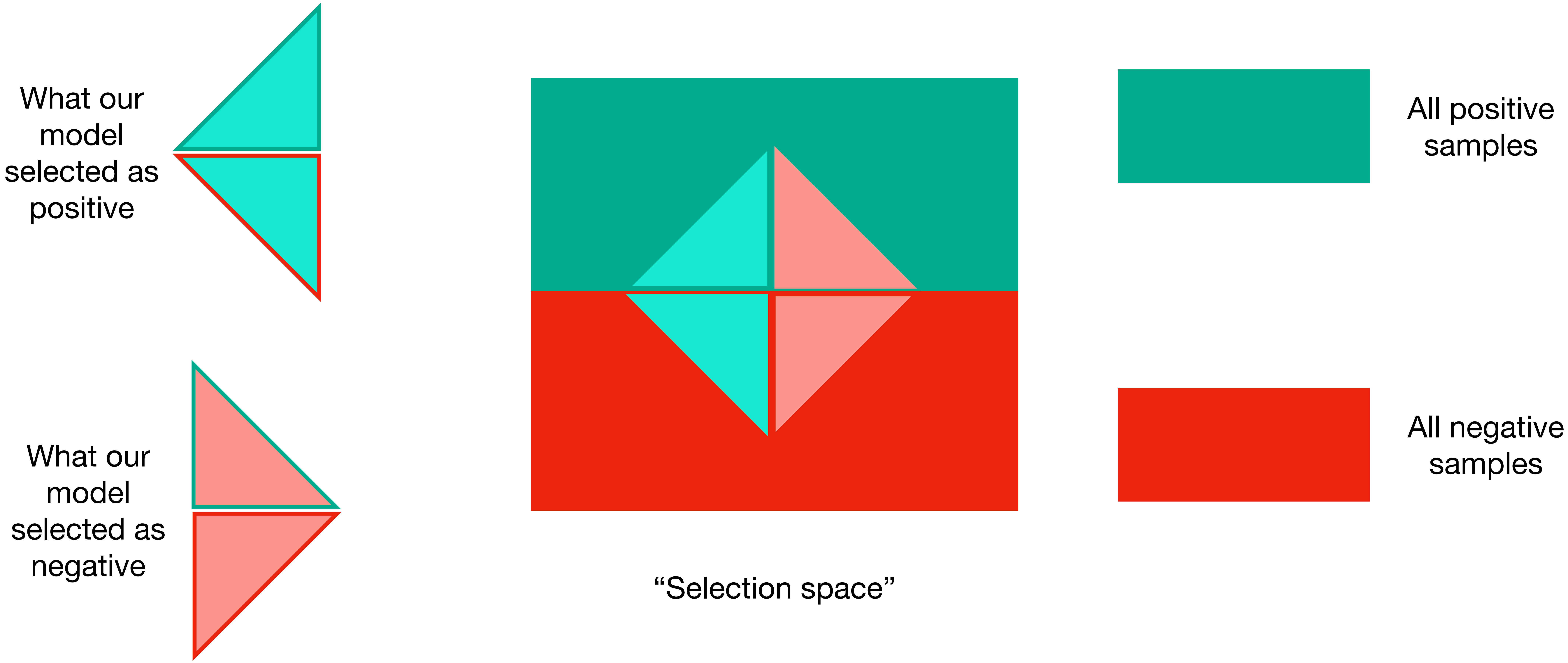


What our model selected as positive

All positive samples

All negative samples

“Selection space”



What our model selected as positive

What our model selected as negative

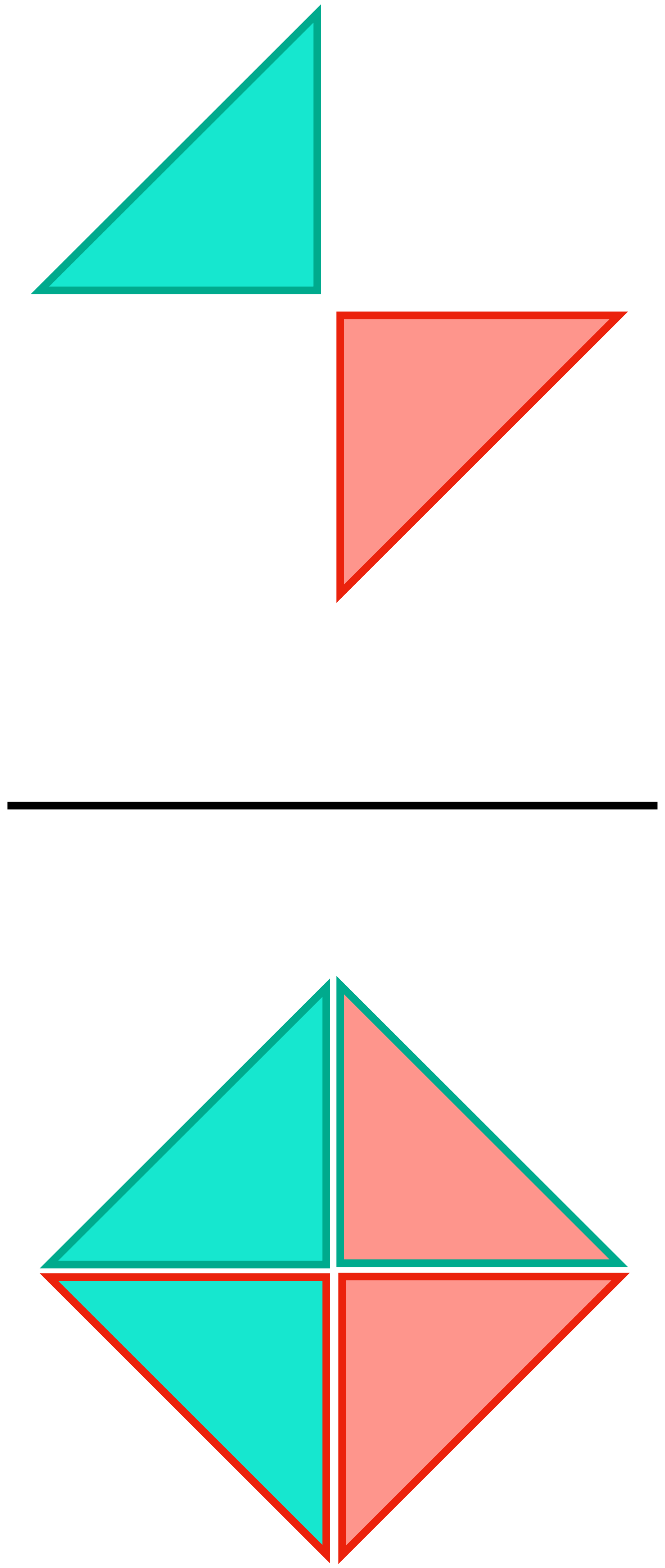
“Selection space”

All positive samples

All negative samples

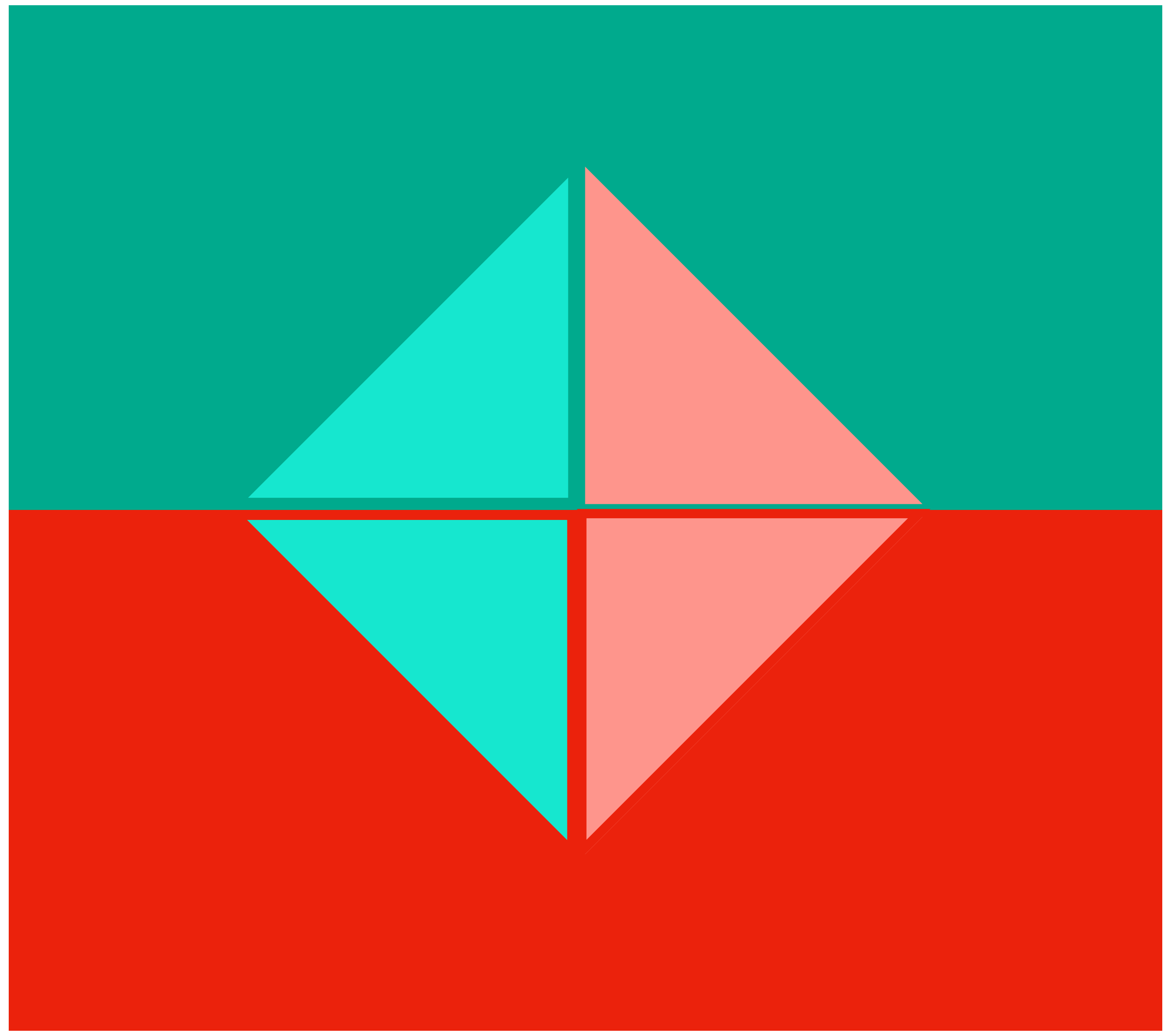
Accuracy

Overall ability of model



“Number of things that we **should** select that we ***did*** select and the number of things that we **shouldn’t** select that we ***didn’t*** select.”

“Out of everything”



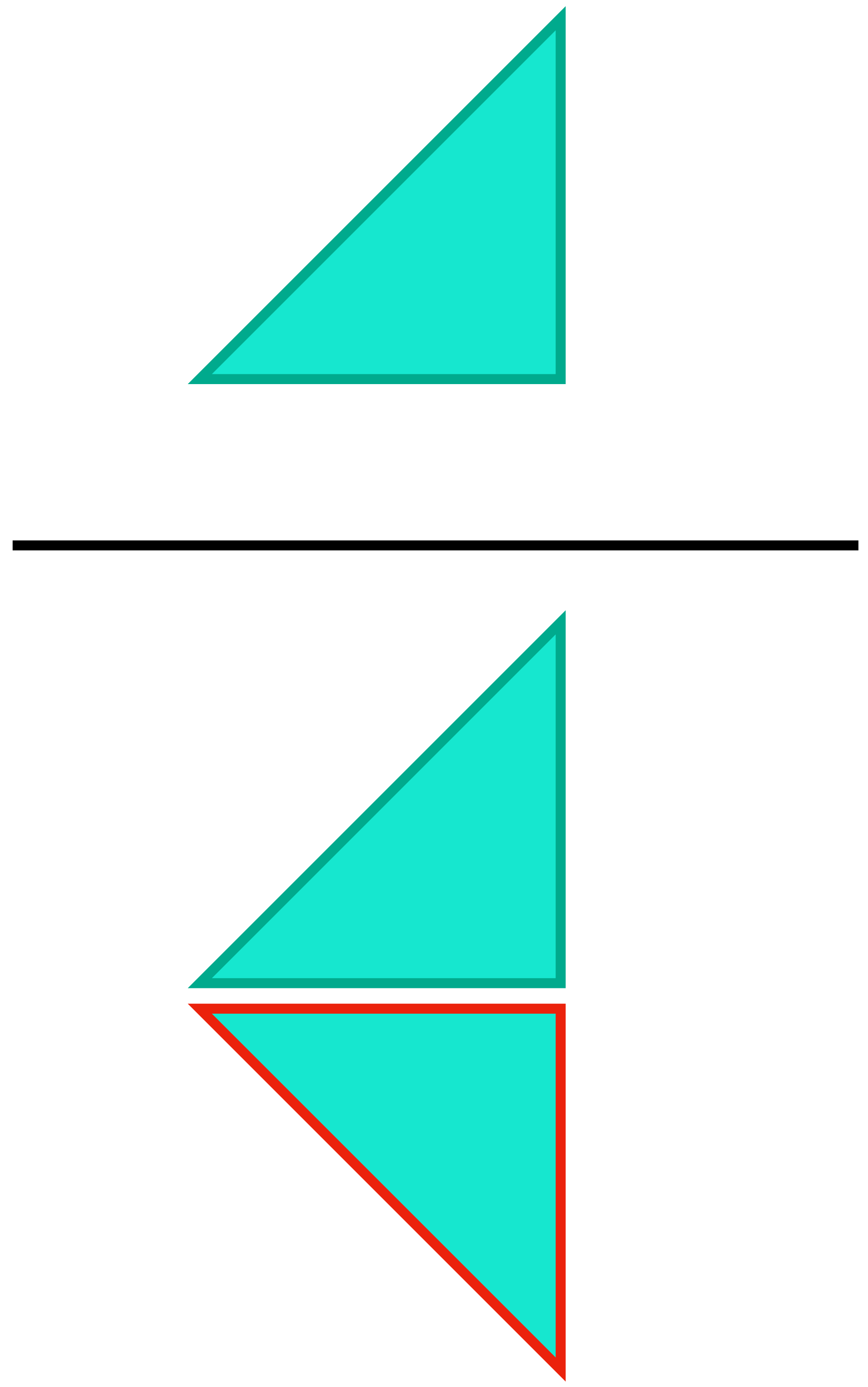
“Selection space”

Accuracy

Overall ability of model

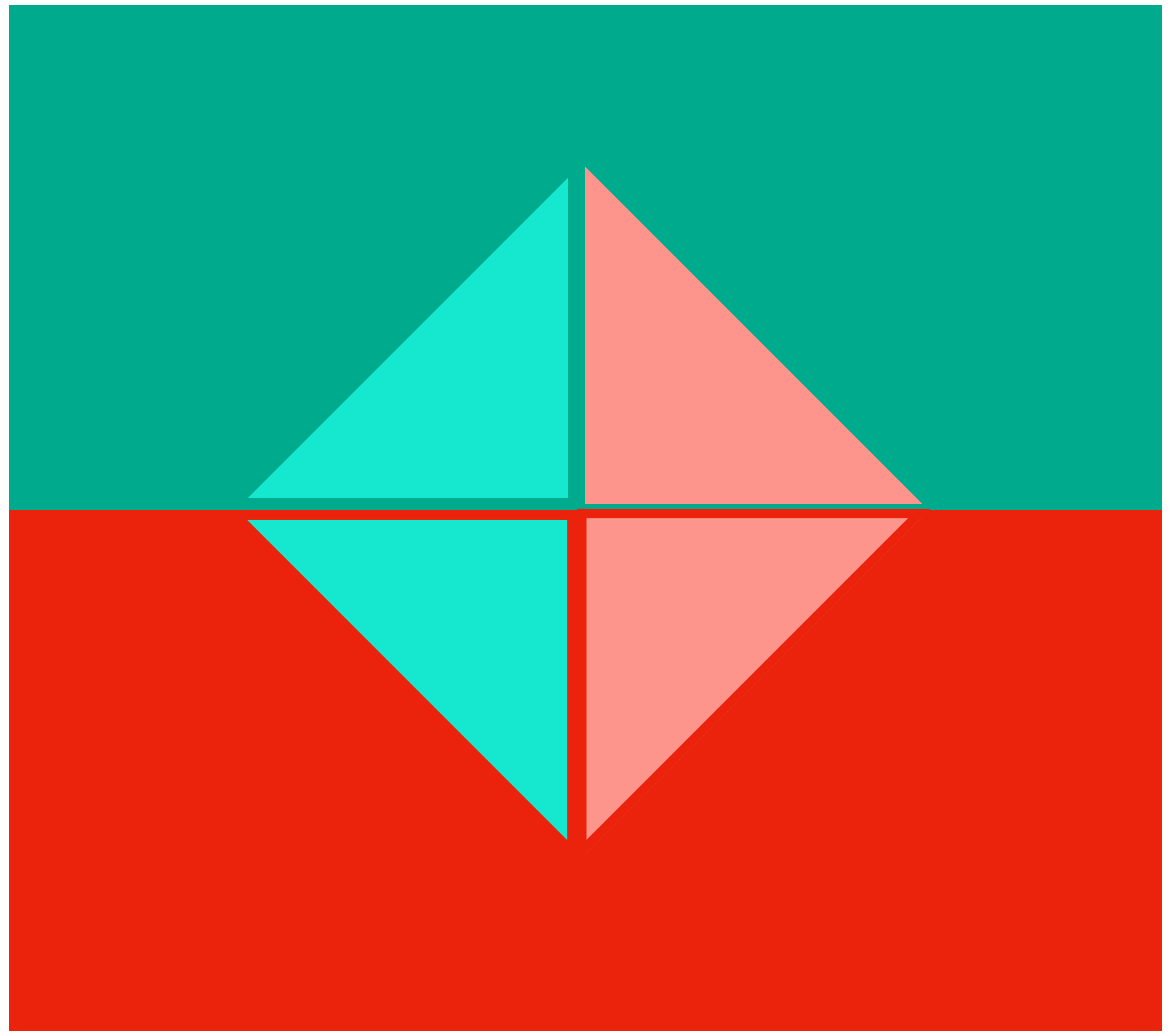
Precision

Amount of selection that’s actually correct.



“Number of things that we **should** select that we ***did*** select”

“Number of things that we **should** select that we ***did*** select and the number of things that we **shouldn’t** select that we ***did*** select.”



“Selection space”

Accuracy

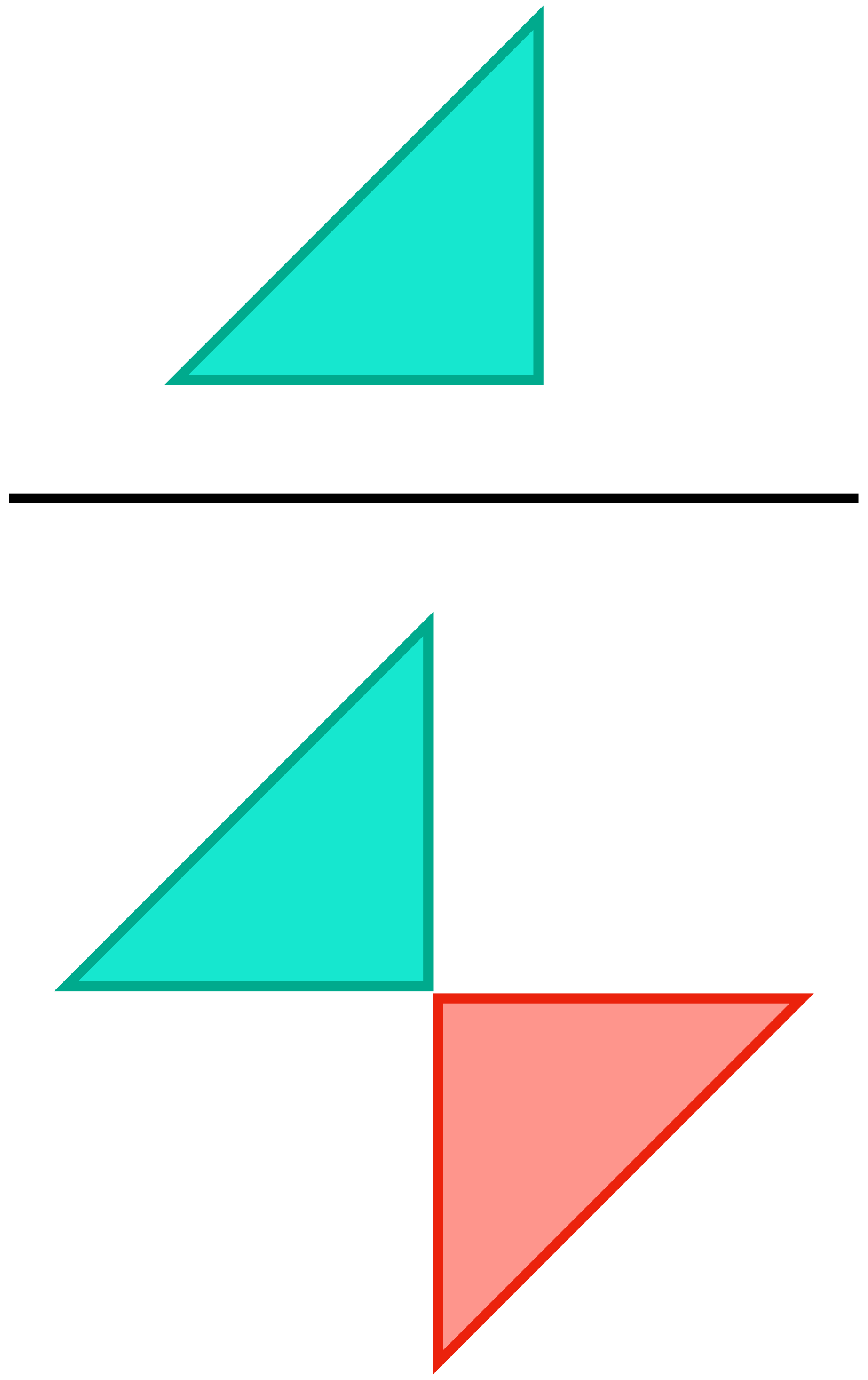
Overall ability of model

Precision

Amount of selection that’s actually correct.

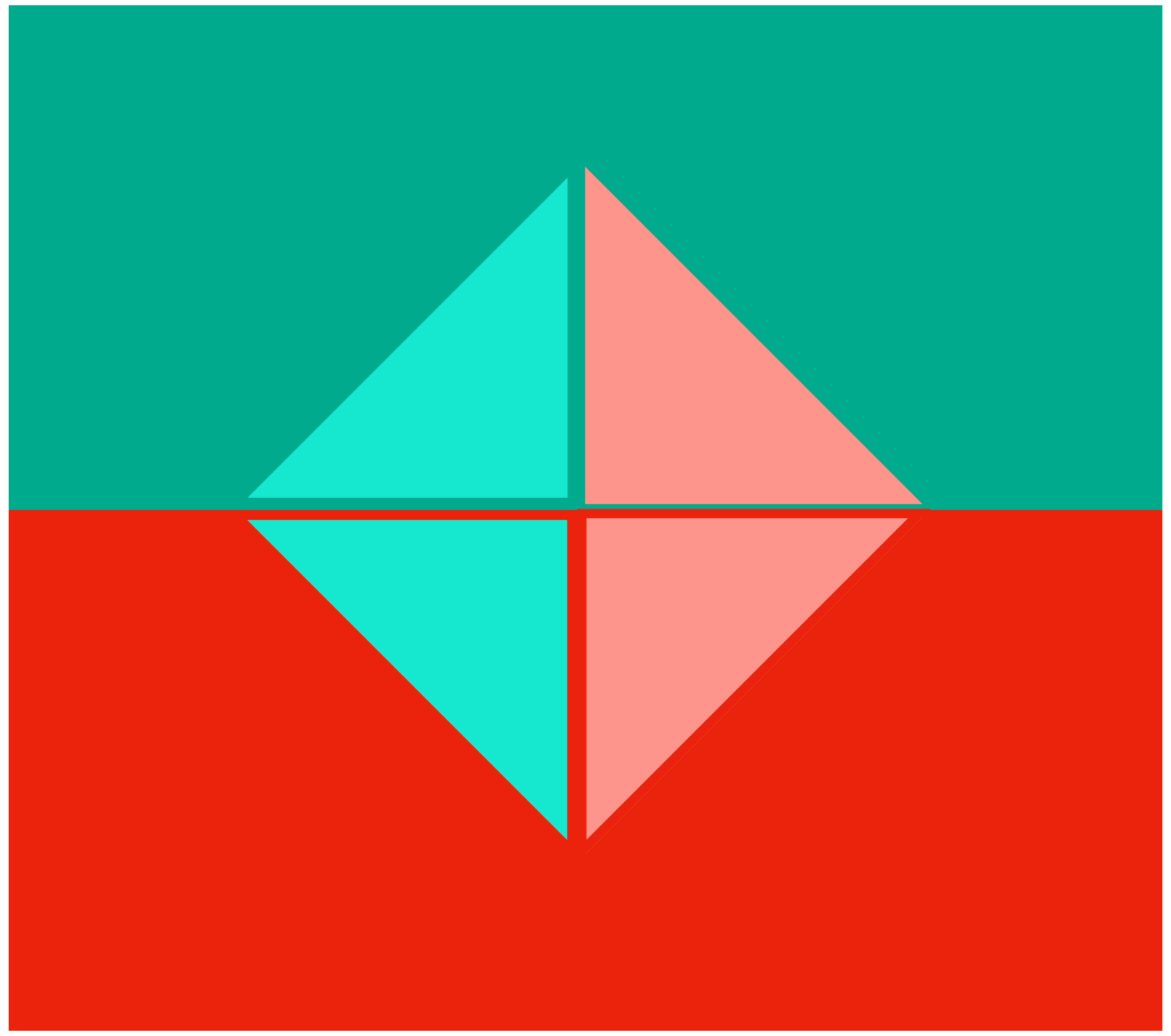
Recall

Amount of what needs to be selected that is selected



“Number of things that we **should** select that we ***did*** select”

“Number of things that we **should** select in total”



“Selection space”

Accuracy

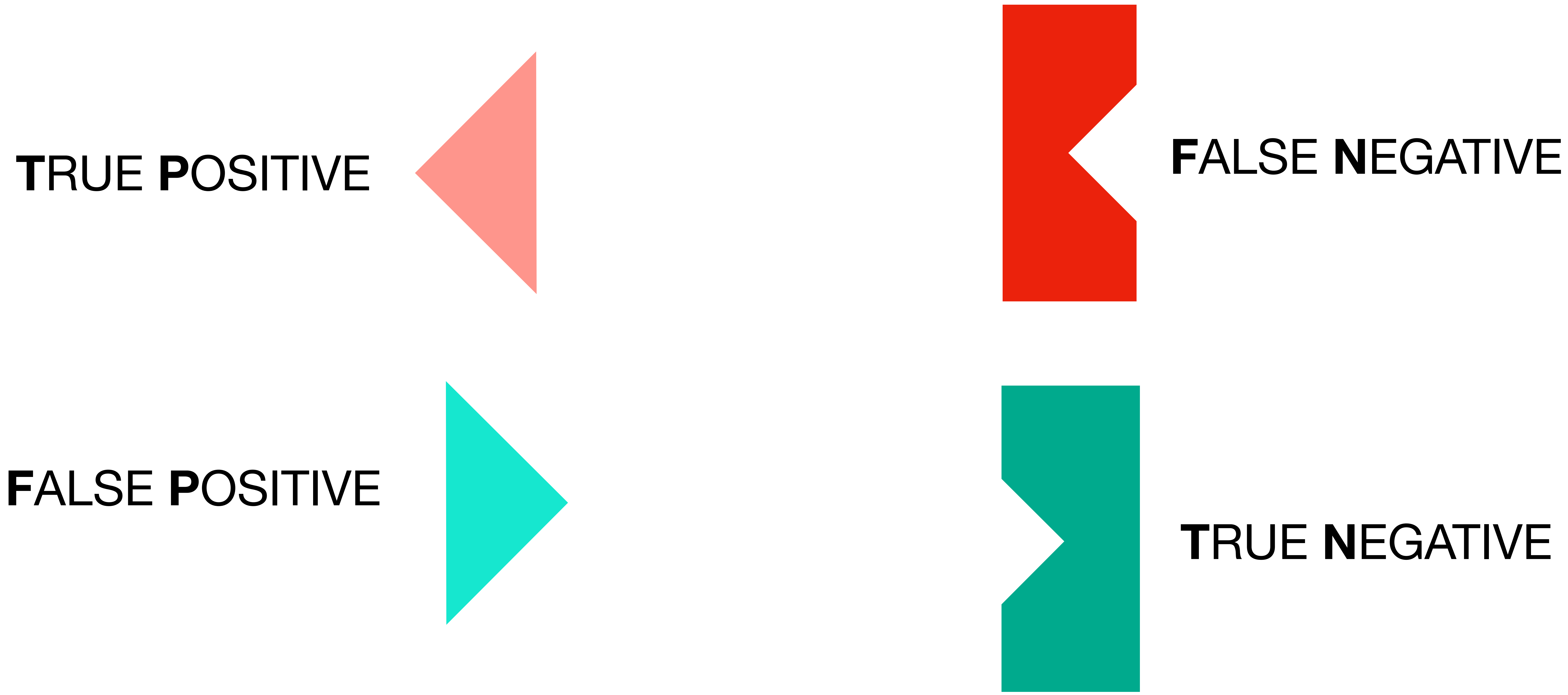
Overall ability of model

Precision

Amount of selection that’s actually correct.

Recall

Amount of what needs to be selected that is selected



TRUE **P**OSITIVE

FALSE **N**EGATIVE

FALSE **P**OSITIVE

TRUE **N**EGATIVE

Accuracy

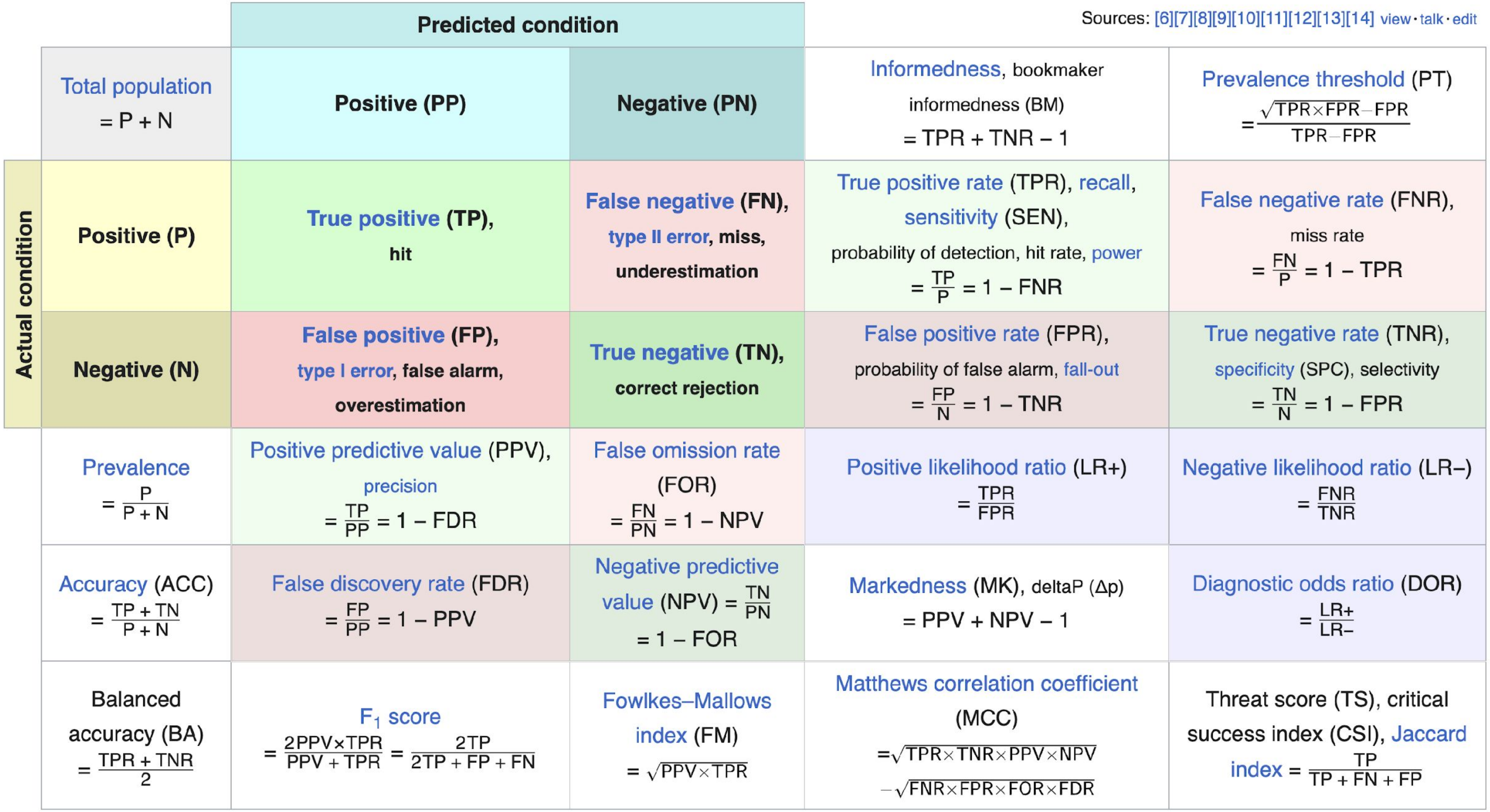
Overall ability of model

Precision

Amount of selection
that’s actually correct.

Recall

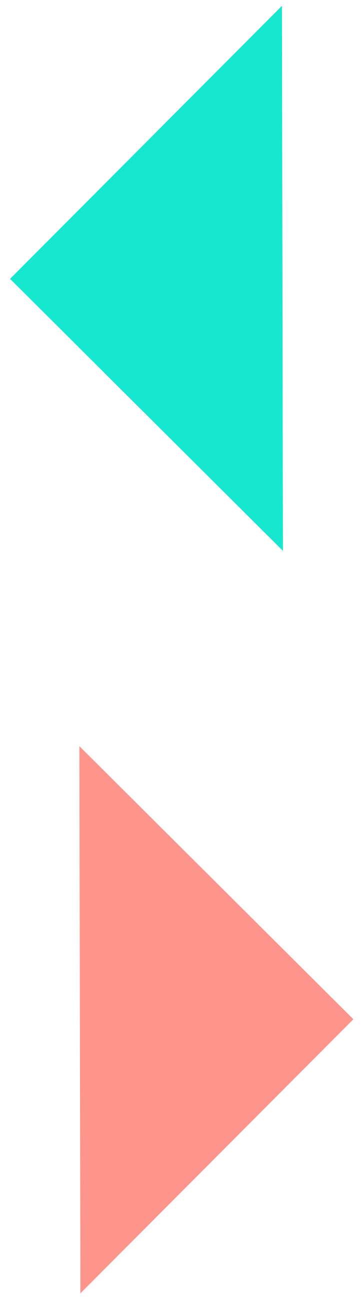
Amount of what needs to
be selected that is selected



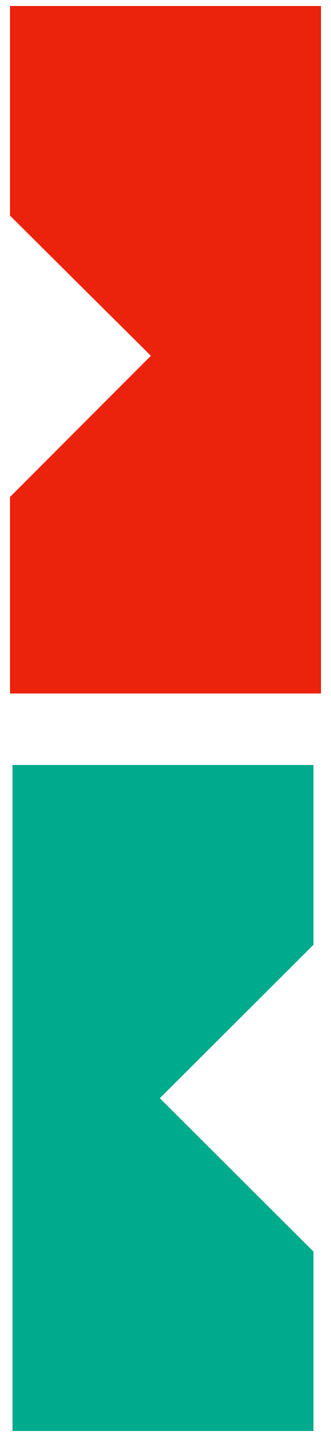
Sources: [6][7][8][9][10][11][12][13][14] view · talk · edit

		Predicted condition			
	Total population $= \mathrm{P} + \mathrm{N}$	**Positive (PP)**	**Negative (PN)**	Informedness, bookmaker informedness (BM) $= \mathrm{TPR} + \mathrm{TNR} - 1$	Prevalence threshold (PT) $= \frac{\sqrt{\mathrm{TPR} \times \mathrm{FPR}} - \mathrm{FPR}}{\mathrm{TPR} - \mathrm{FPR}}$
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 $= \frac{\mathrm{TP}}{\mathrm{P}} = 1 - \mathrm{FNR}$	False negative rate (FNR), miss rate $= \frac{\mathrm{FN}}{\mathrm{P}} = 1 - \mathrm{TPR}$
	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 $= \frac{\mathrm{FP}}{\mathrm{N}} = 1 - \mathrm{TNR}$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{\mathrm{TN}}{\mathrm{N}} = 1 - \mathrm{FPR}$
	Prevalence $= \frac{\mathrm{P}}{\mathrm{P} + \mathrm{N}}$	Positive predictive value (PPV), precision $= \frac{\mathrm{TP}}{\mathrm{PP}} = 1 - \mathrm{FDR}$	False omission rate (FOR) $= \frac{\mathrm{FN}}{\mathrm{PN}} = 1 - \mathrm{NPV}$	Positive likelihood ratio (LR+) $= \frac{\mathrm{TPR}}{\mathrm{FPR}}$	Negative likelihood ratio (LR–) $= \frac{\mathrm{FNR}}{\mathrm{TNR}}$
	Accuracy (ACC) $= \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}}$	False discovery rate (FDR) $= \frac{\mathrm{FP}}{\mathrm{PP}} = 1 - \mathrm{PPV}$	Negative predictive value (NPV) $= \frac{\mathrm{TN}}{\mathrm{PN}}$ $= 1 - \mathrm{FOR}$	Markedness (MK), deltaP (Δp) $= \mathrm{PPV} + \mathrm{NPV} - 1$	Diagnostic odds ratio (DOR) $= \frac{\mathrm{LR+}}{\mathrm{LR-}}$
	Balanced accuracy (BA) $= \frac{\mathrm{TPR} + \mathrm{TNR}}{2}$	F_1 score $= \frac{2\mathrm{PPV} \times \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$	Fowlkes–Mallows index (FM) $= \sqrt{\mathrm{PPV} \times \mathrm{TPR}}$	Matthews correlation coefficient (MCC) $= \sqrt{\mathrm{TPR} \times \mathrm{TNR} \times \mathrm{PPV} \times \mathrm{NPV}} - \sqrt{\mathrm{FNR} \times \mathrm{FPR} \times \mathrm{FOR} \times \mathrm{FDR}}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}$

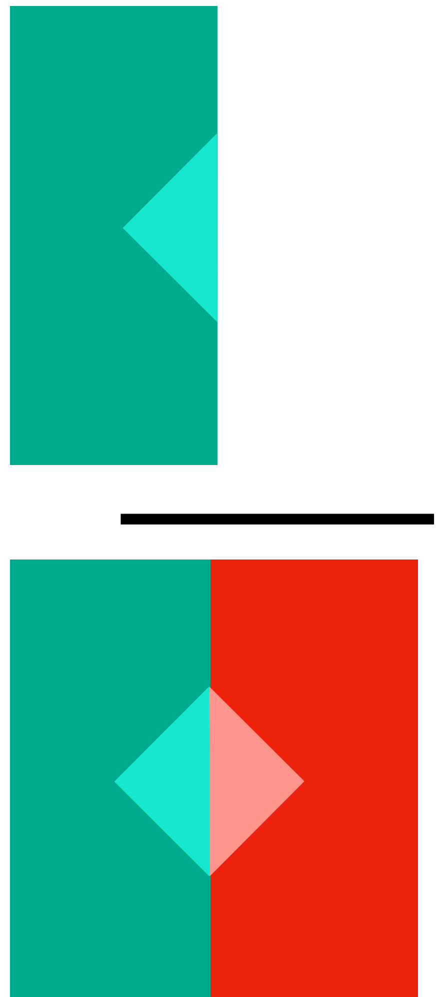
https://en.wikipedia.org/wiki/Precision_and_recall



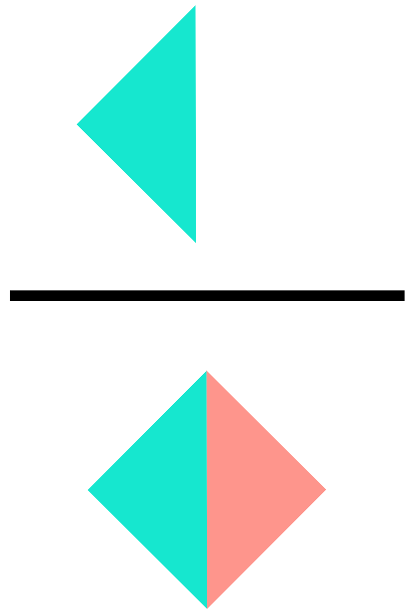
	Predicted Positive	Predicted Negative
Actually Positive	TP	FN
Actually Negative	FP	TN



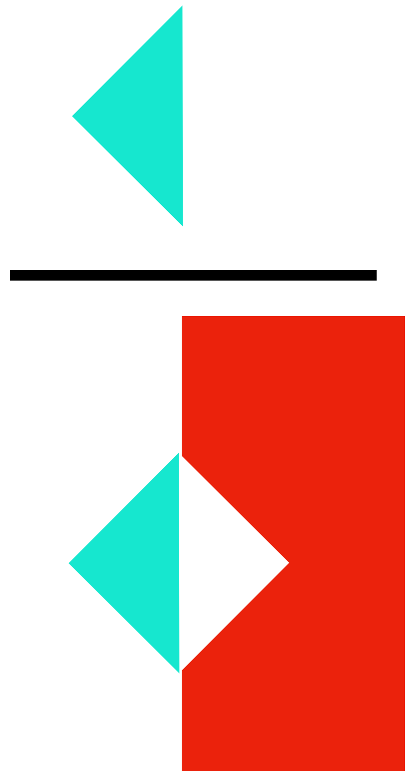
Accuracy



Precision



Recall



Accuracy

Overall ability of model

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

exactly zero

Precision

Amount of selection that’s actually correct.

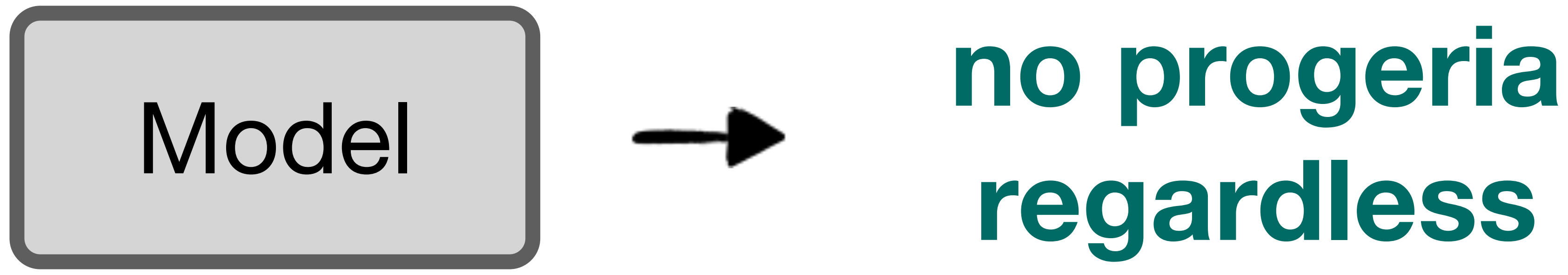
$$\frac{TP}{TP + FP}$$

Recall

Amount of what needs to be selected that is selected

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

scaled properly!



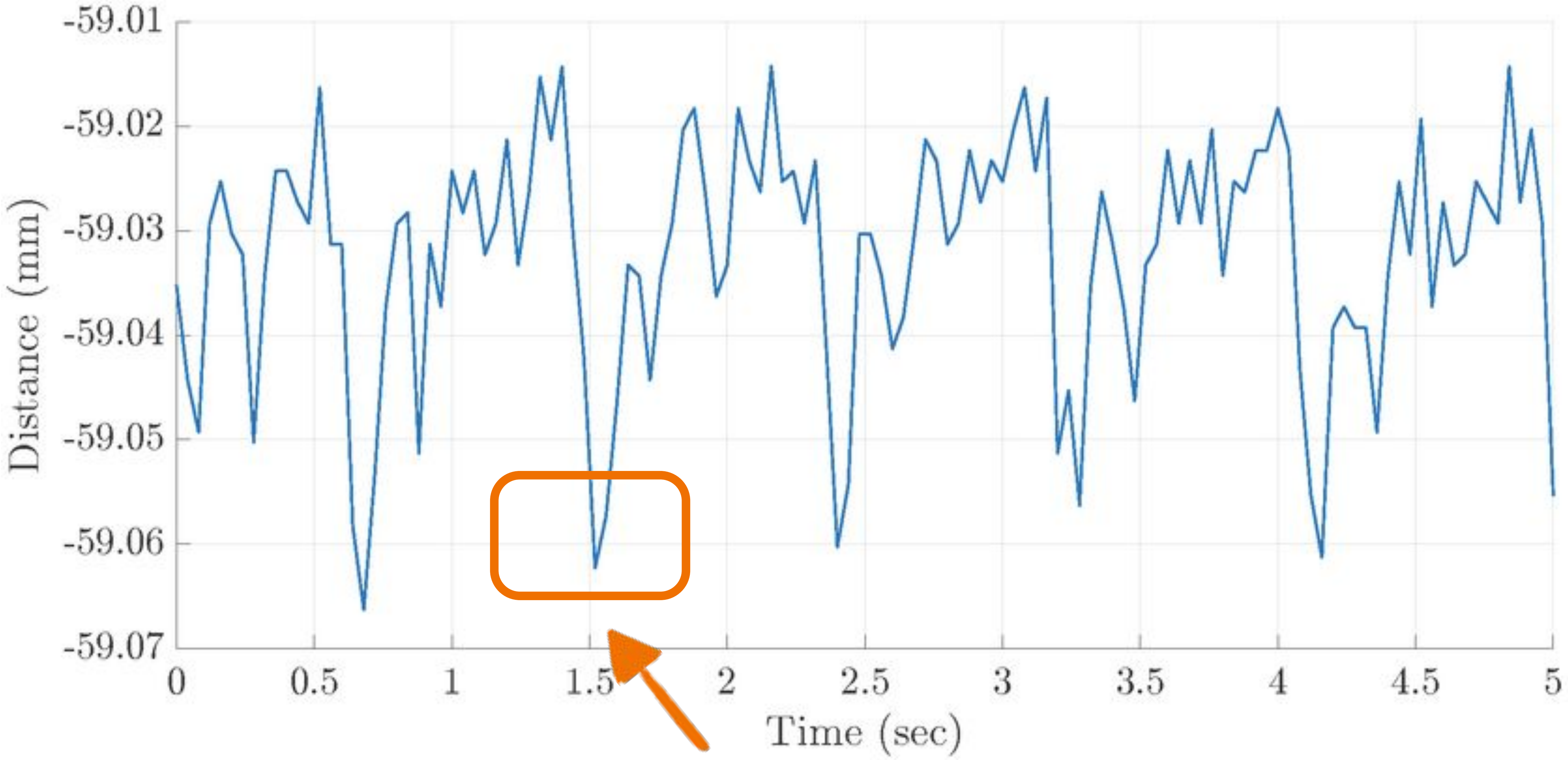
Progeria affects ~159 patients in the US

we have a dataset of all American pediatric patients

storytime!

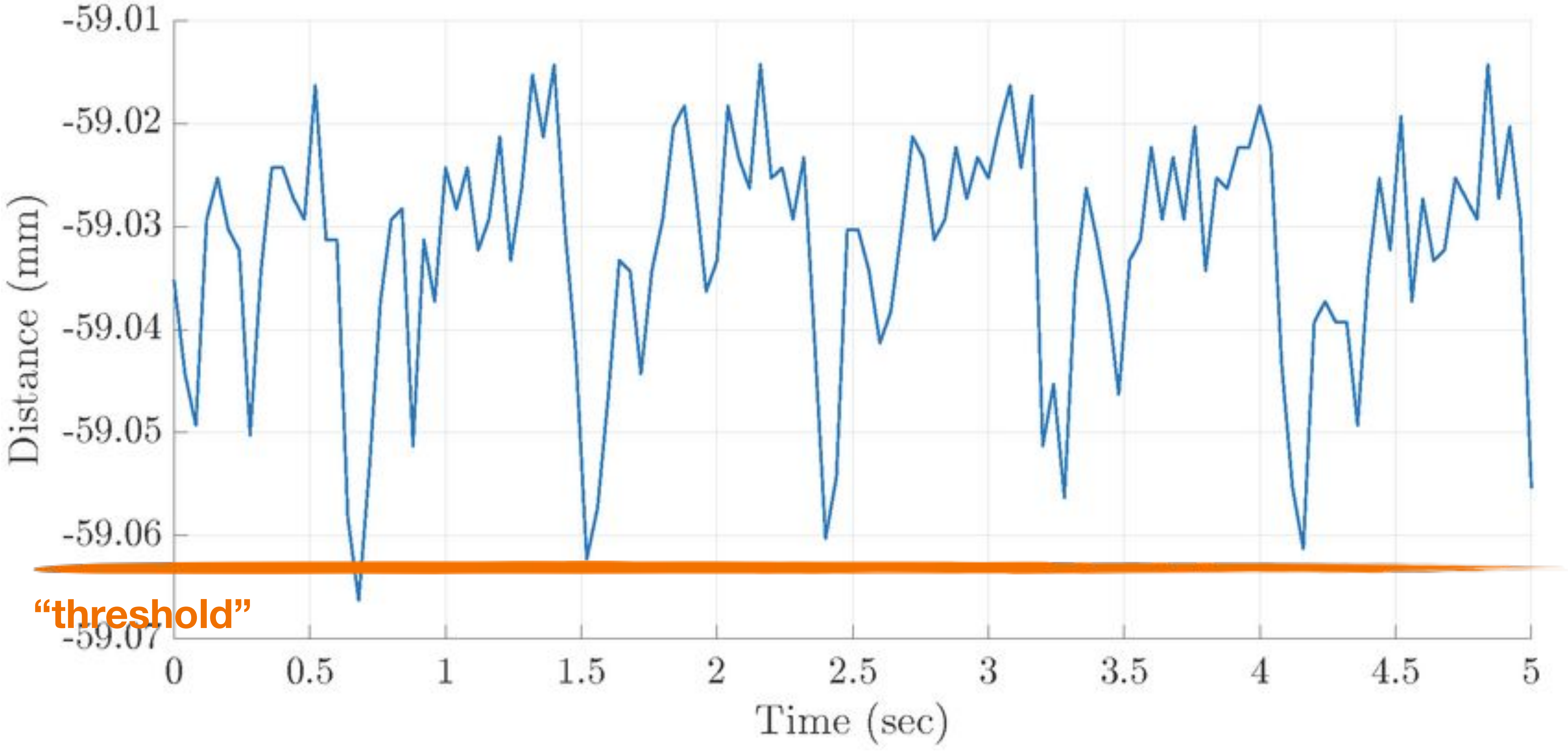
storytime!

storytime!



does this count as detection?

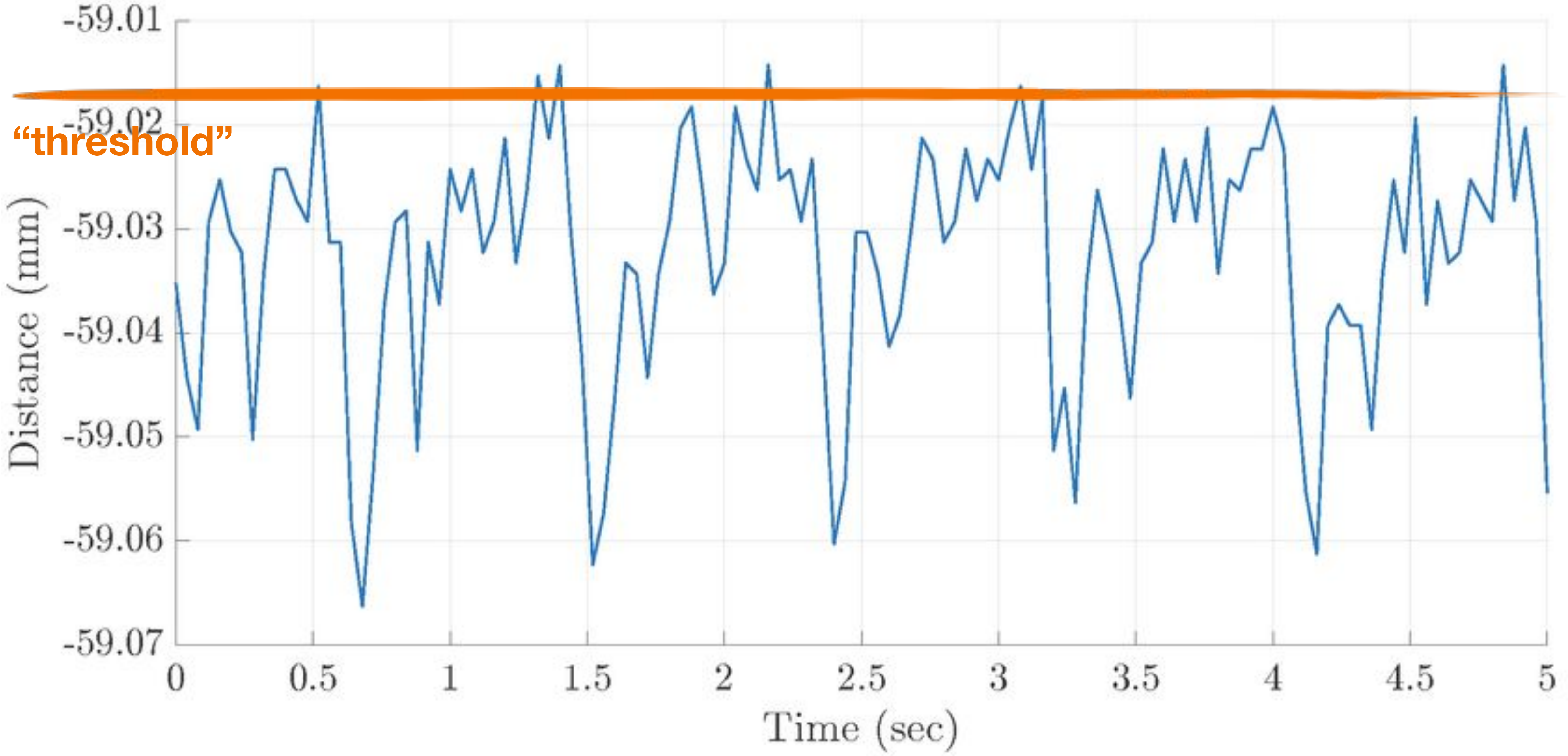
storytime!



“threshold”

high recall, low precision

storytime!



“threshold”

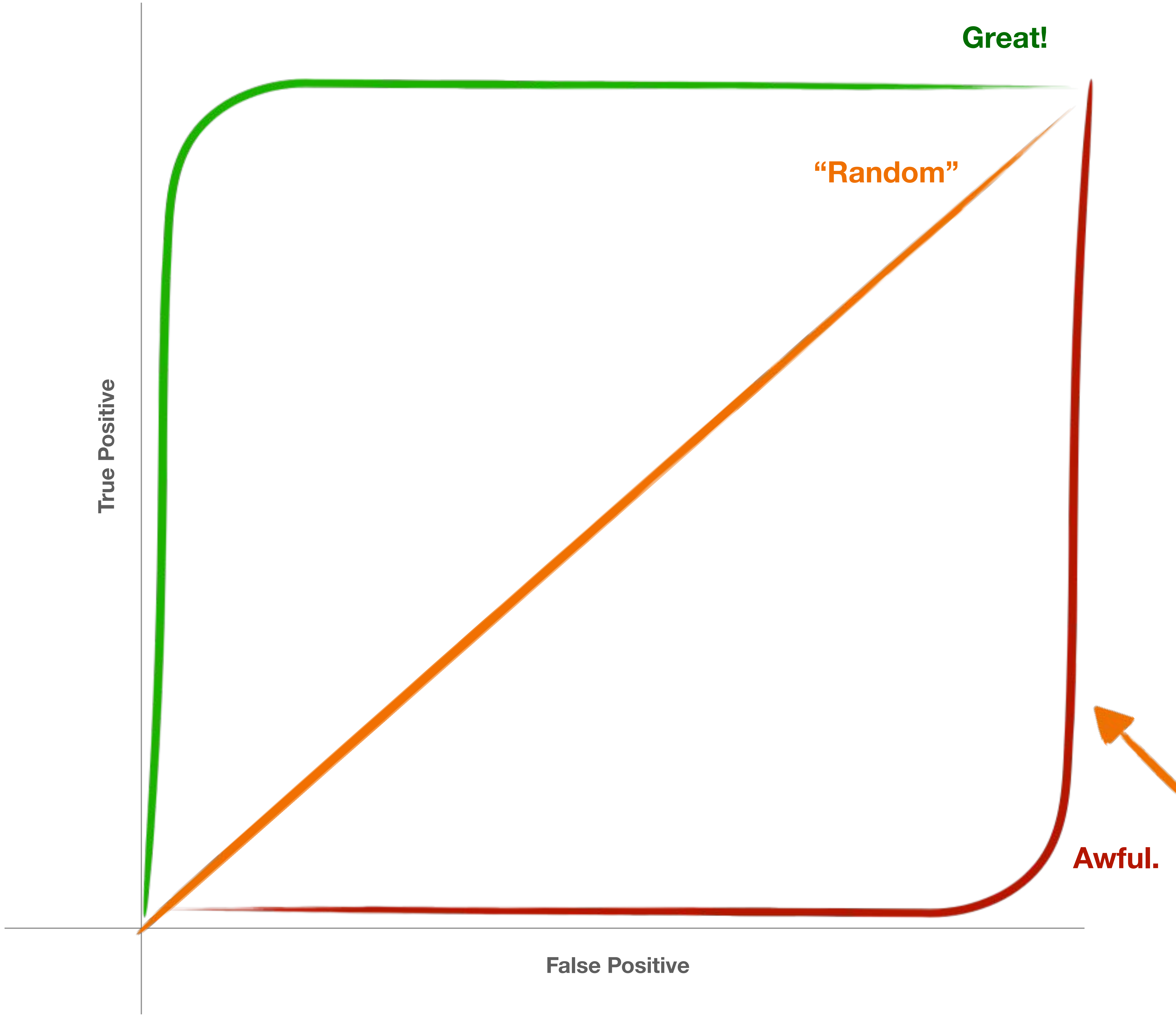
high precision, low recall

quantifying “threshold”

quantifying

“threshold”

ROC Curve!

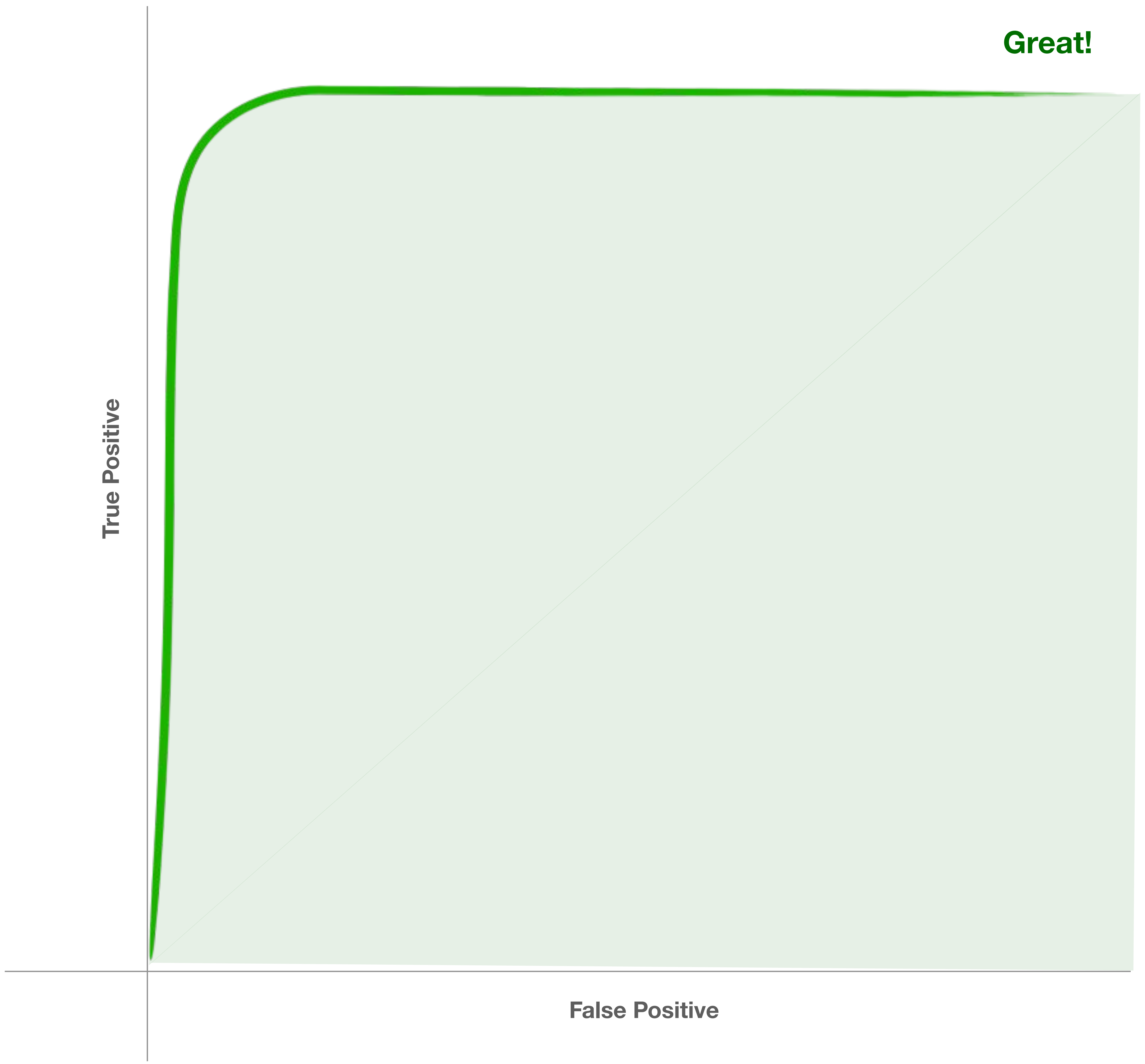


need lots of false positives
before detecting a true positive

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

Q: how do you compare these points

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

what makes models fit better

more data

balanced data

normalized data

quality data

more data

balanced data

normalized data

quality data

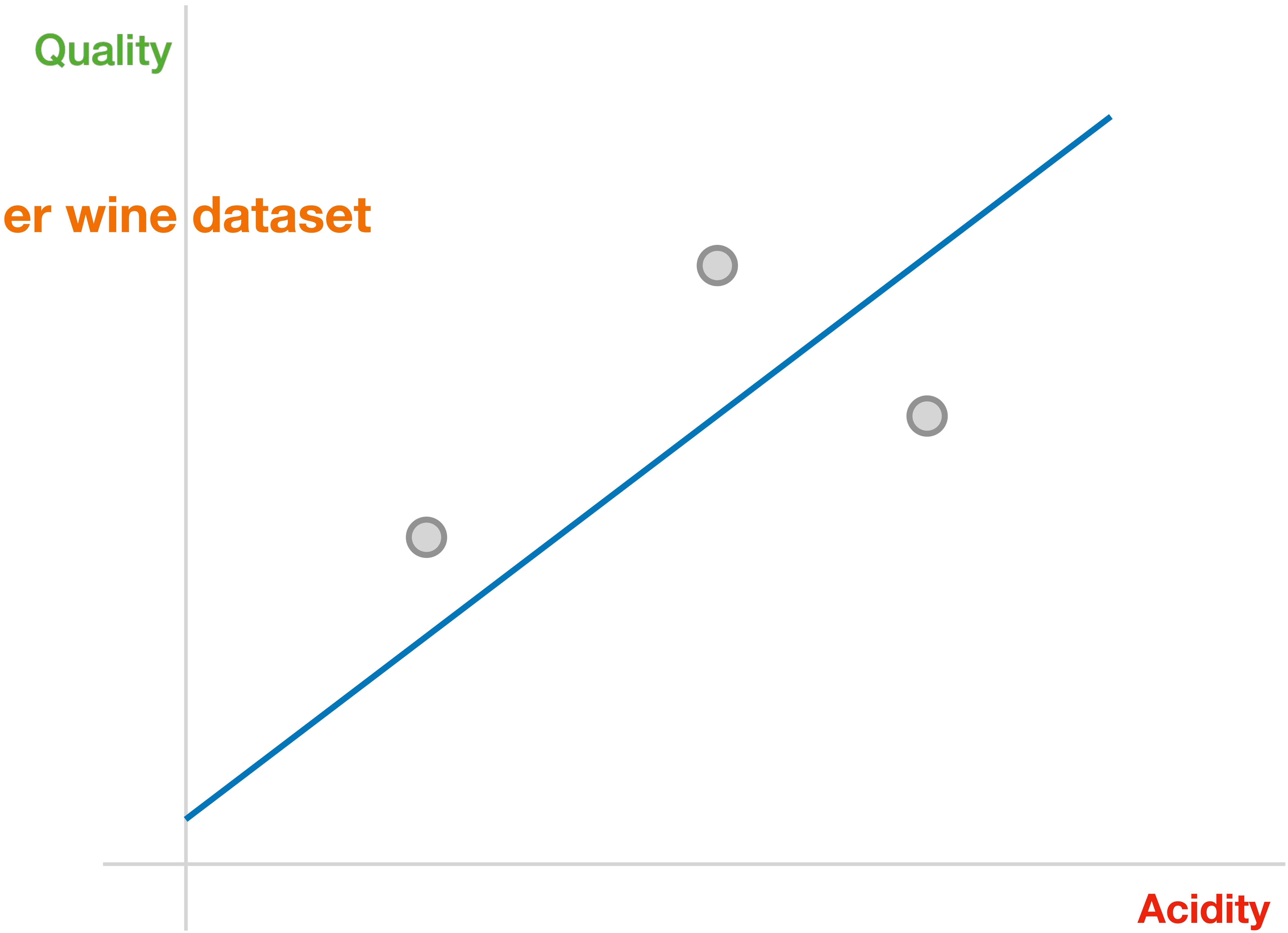
more data **balanced** data **normalized** data **quality** data

more data

let’s say we have a simpler wine dataset

Quality on the y axis

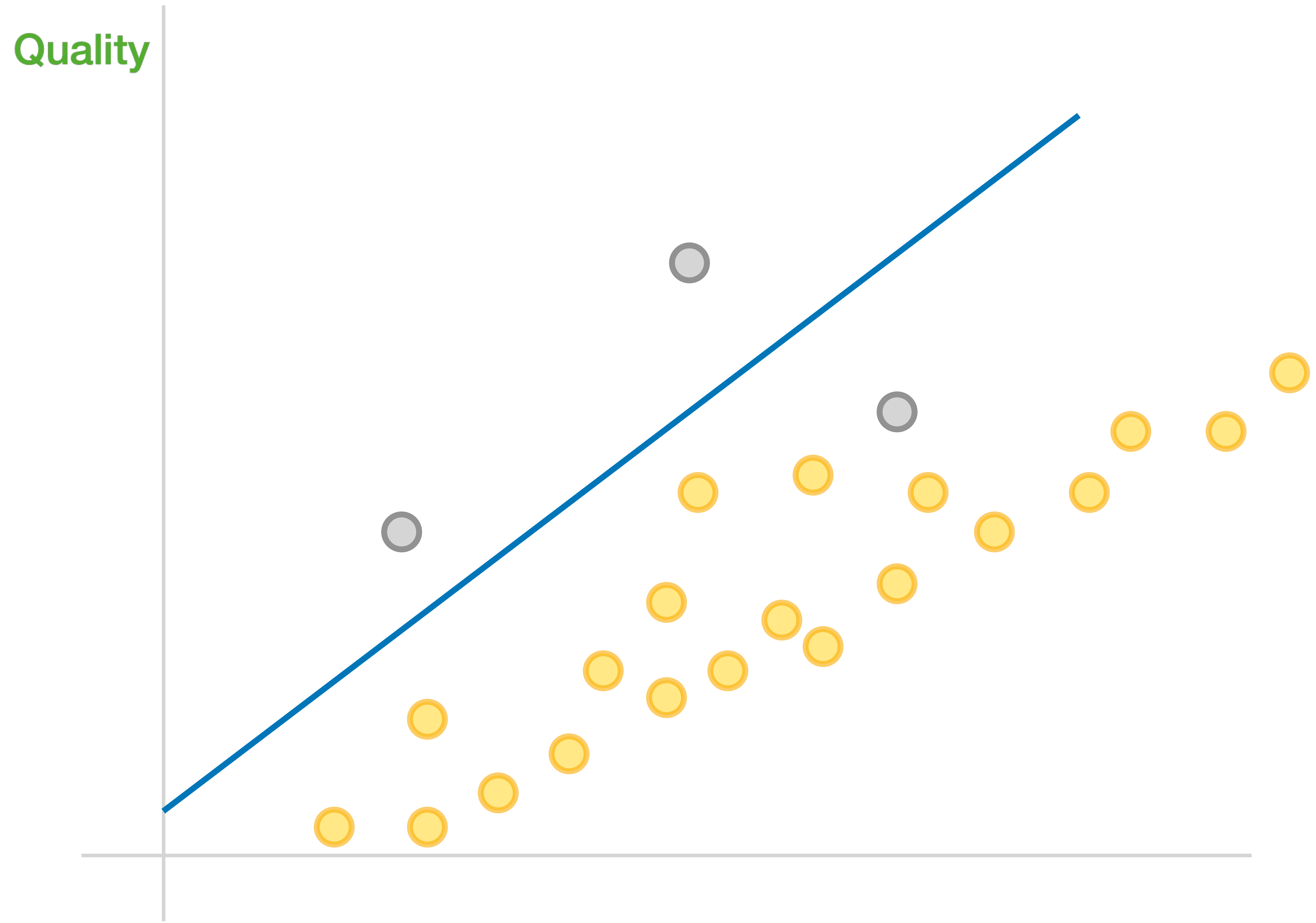
Acidity on the x axis



more data

Quality **on the y axis**

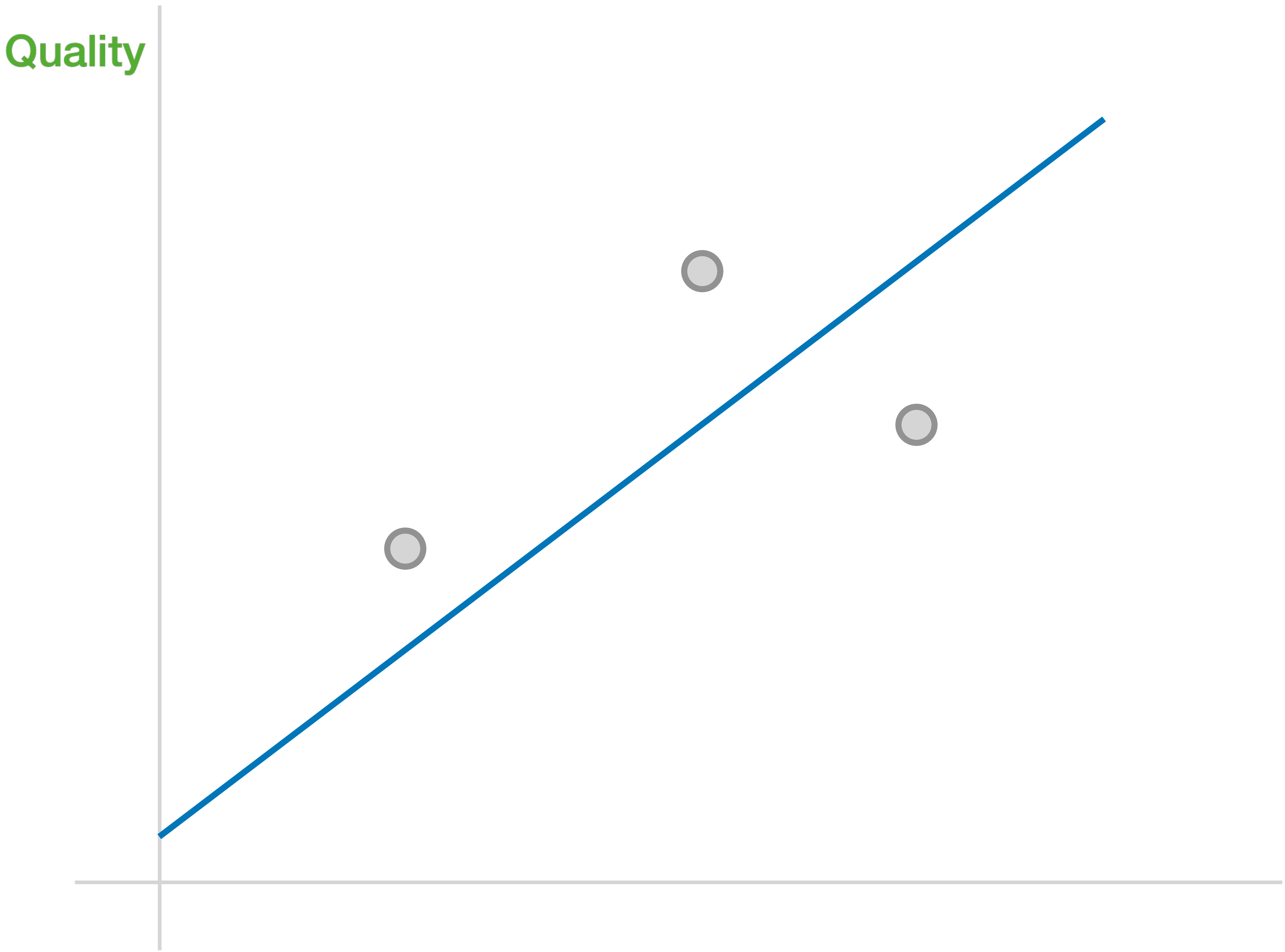
Acidity on the x axis



more data

Quality **on the y axis**

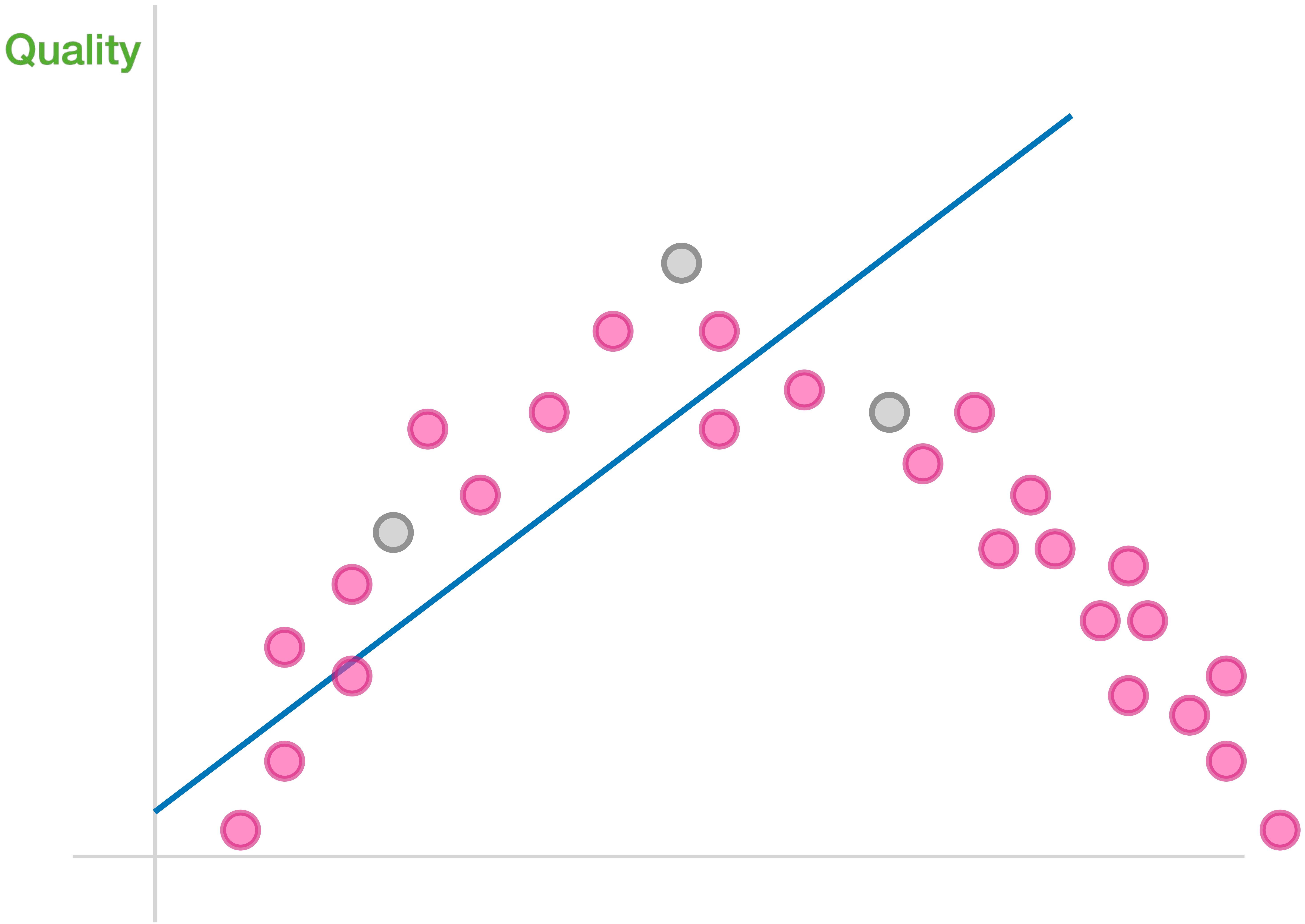
Acidity on the x axis



more data

Quality on the y axis

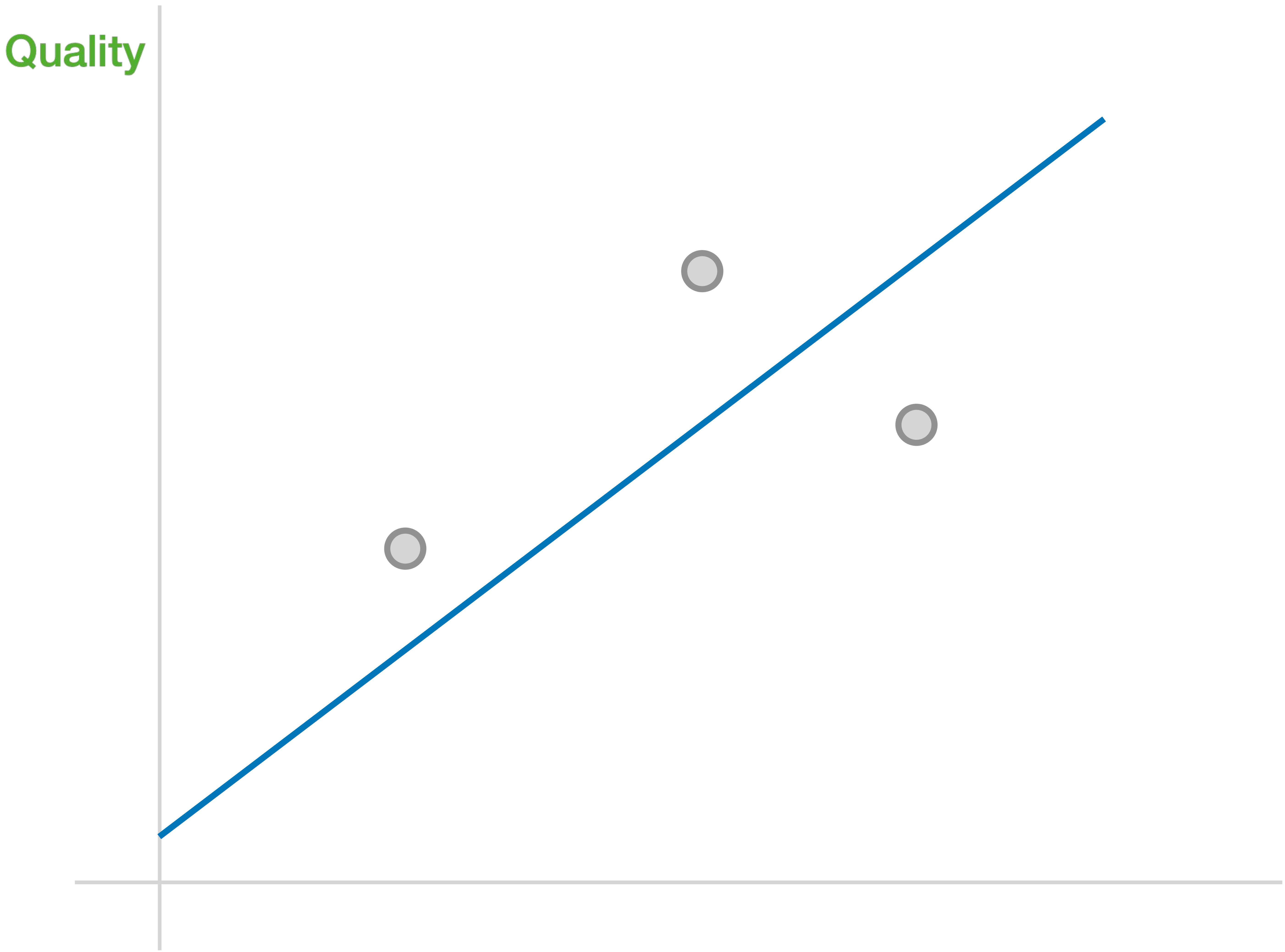
Acidity on the x axis



more data

Quality on the y axis

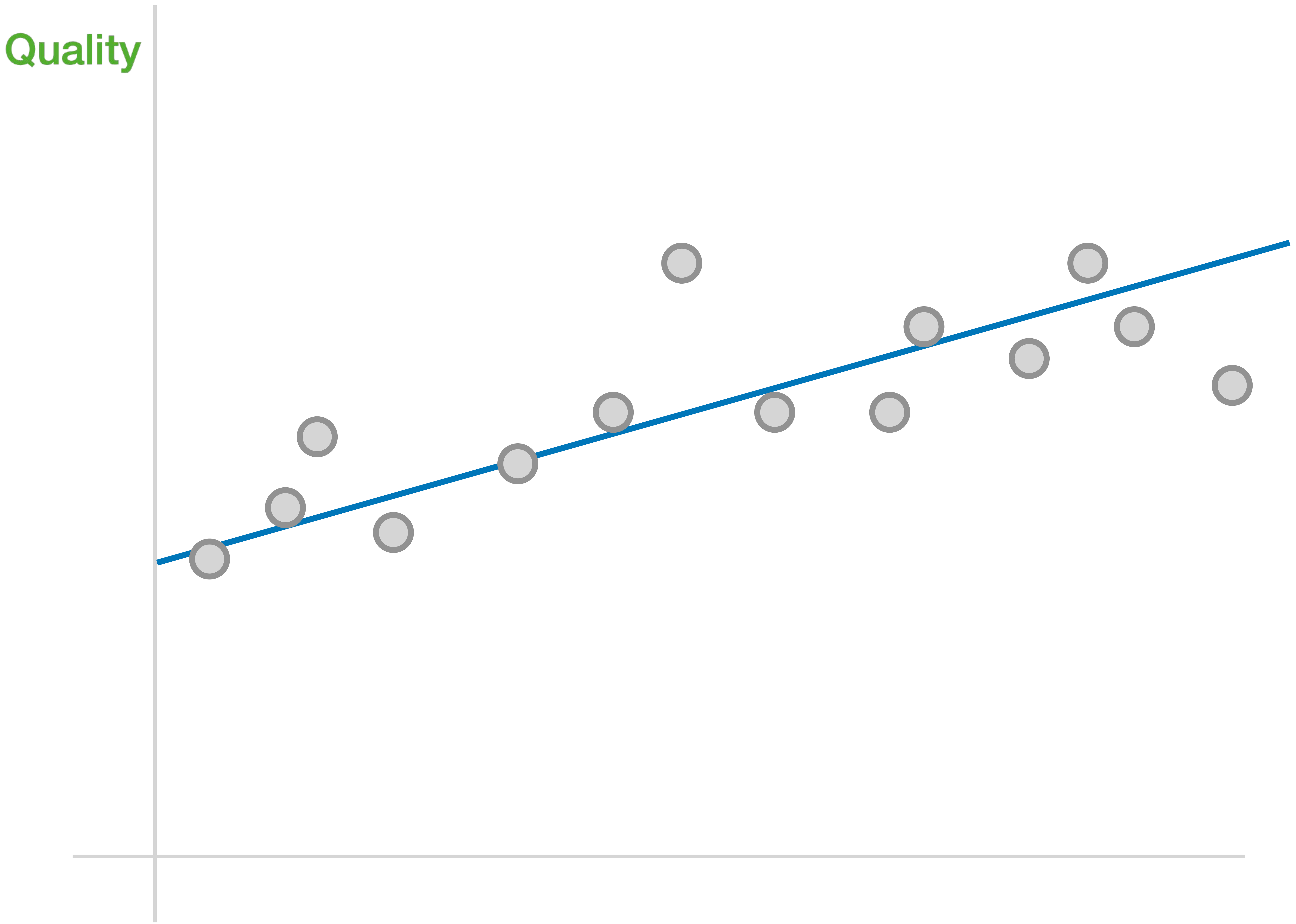
Acidity on the x axis



more data

Quality on the y axis

Acidity on the x axis



- **use more data, get more accurate results**

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

Product

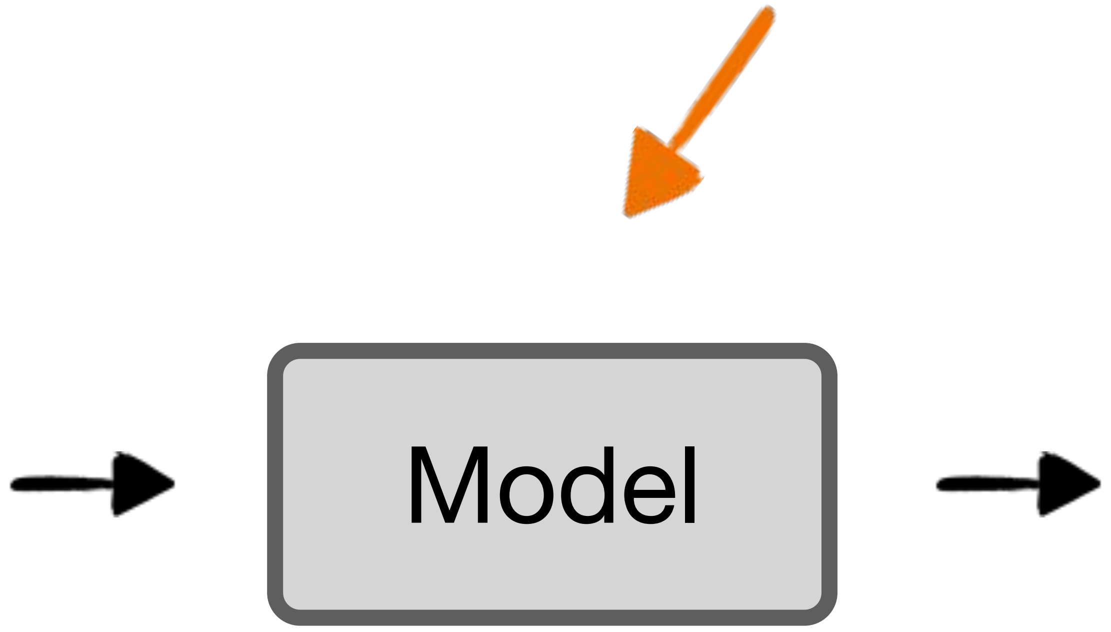
Product

Amber Colored
Angora Super Sweet
Black Ethiopian
Burbank Slicing
Dona
Sophie's Choice
White Bush
Ace 55

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

a
reminder…

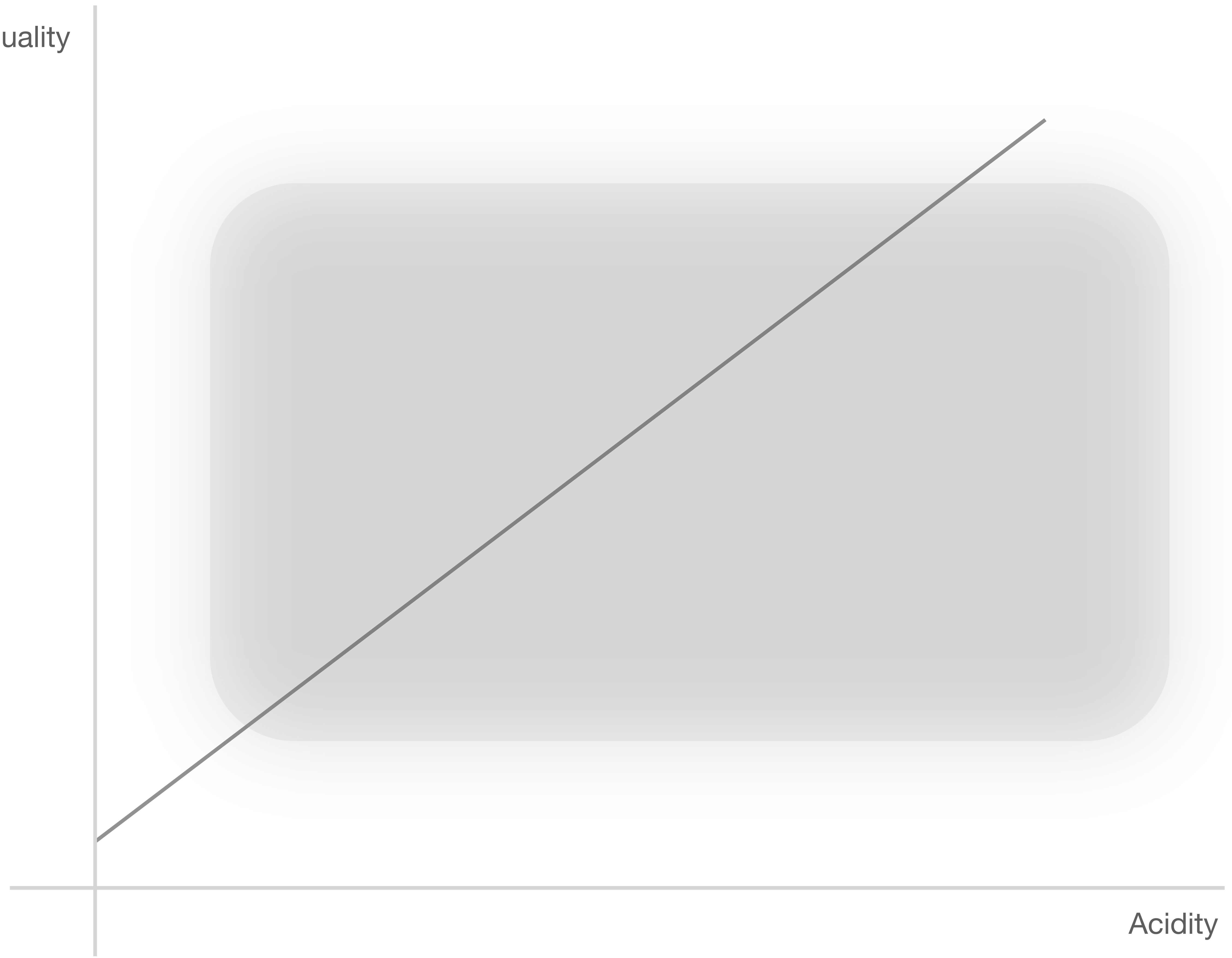
that’s linear(ish)*



models are functions

more

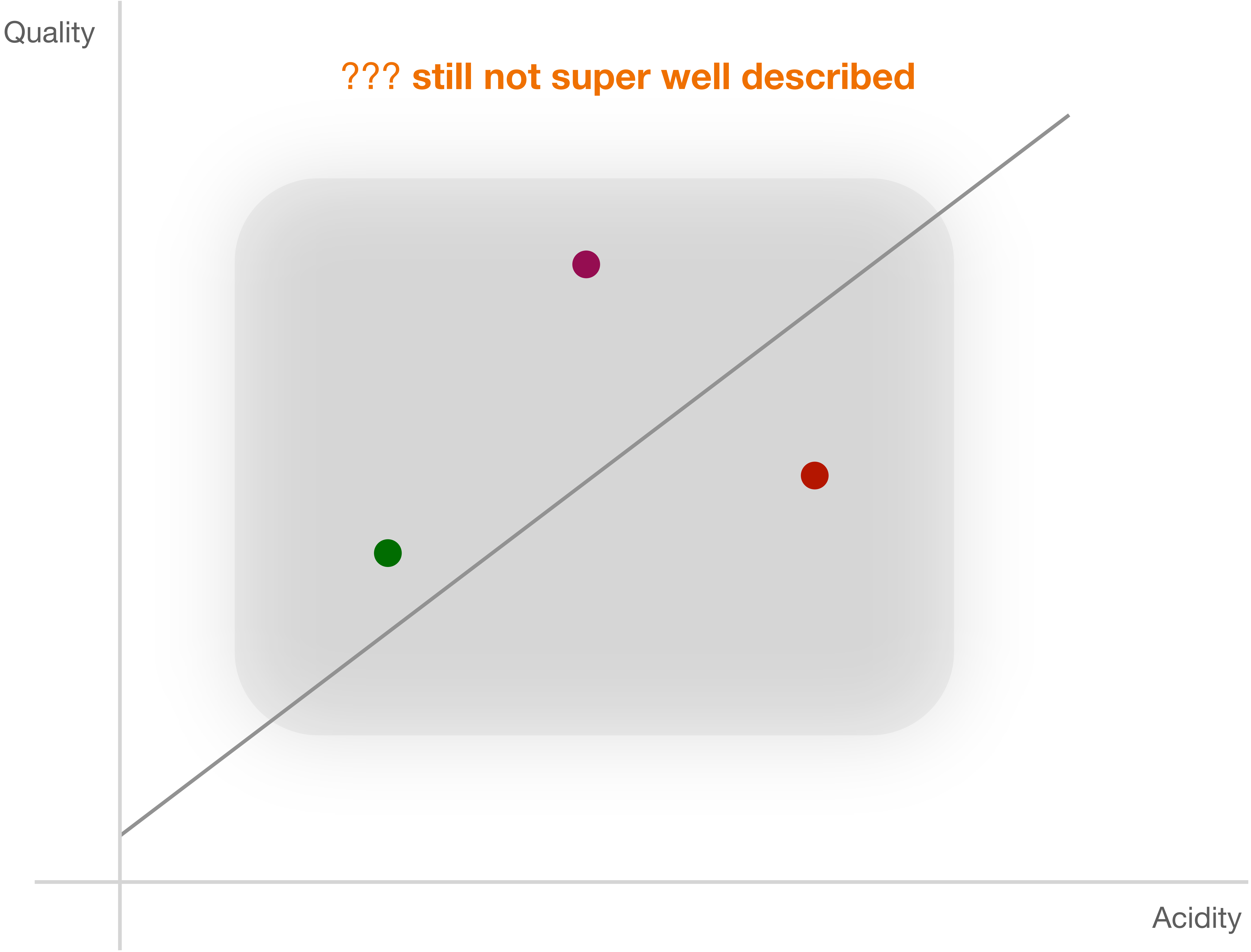
data



that’s linear(ish)*

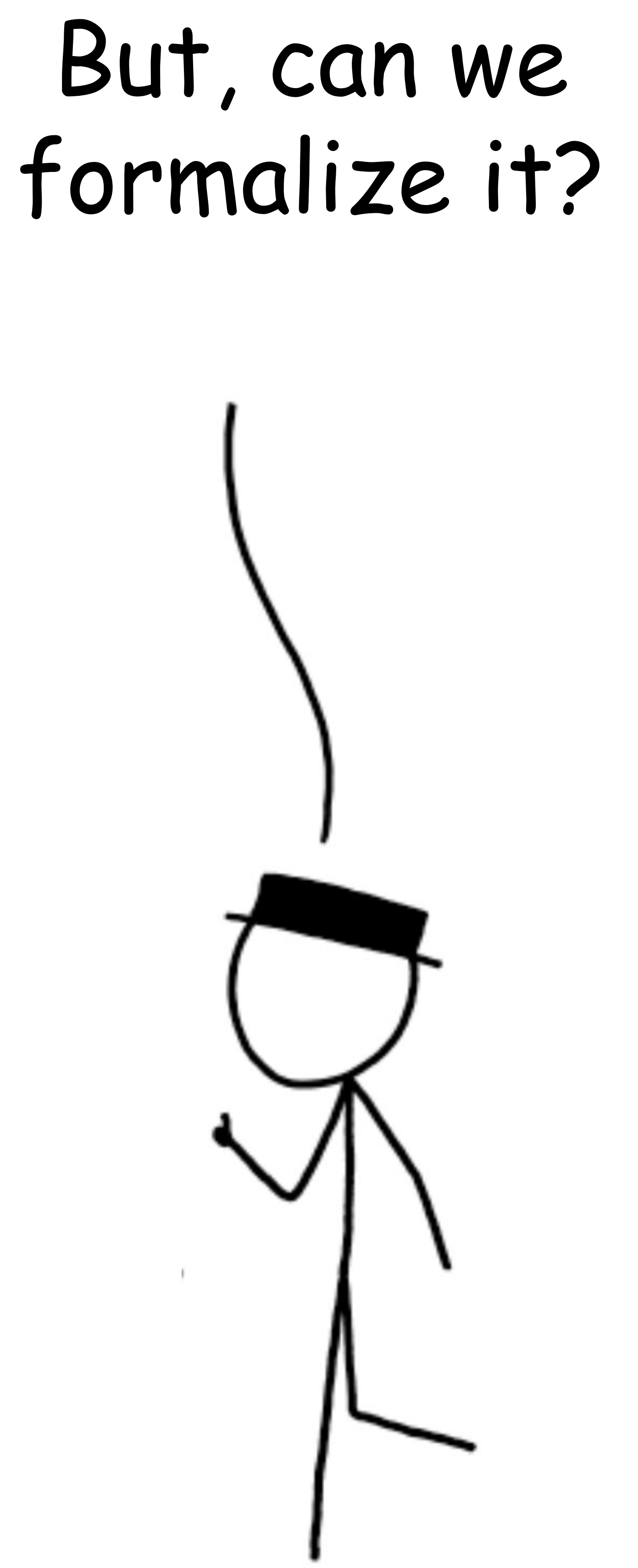
more

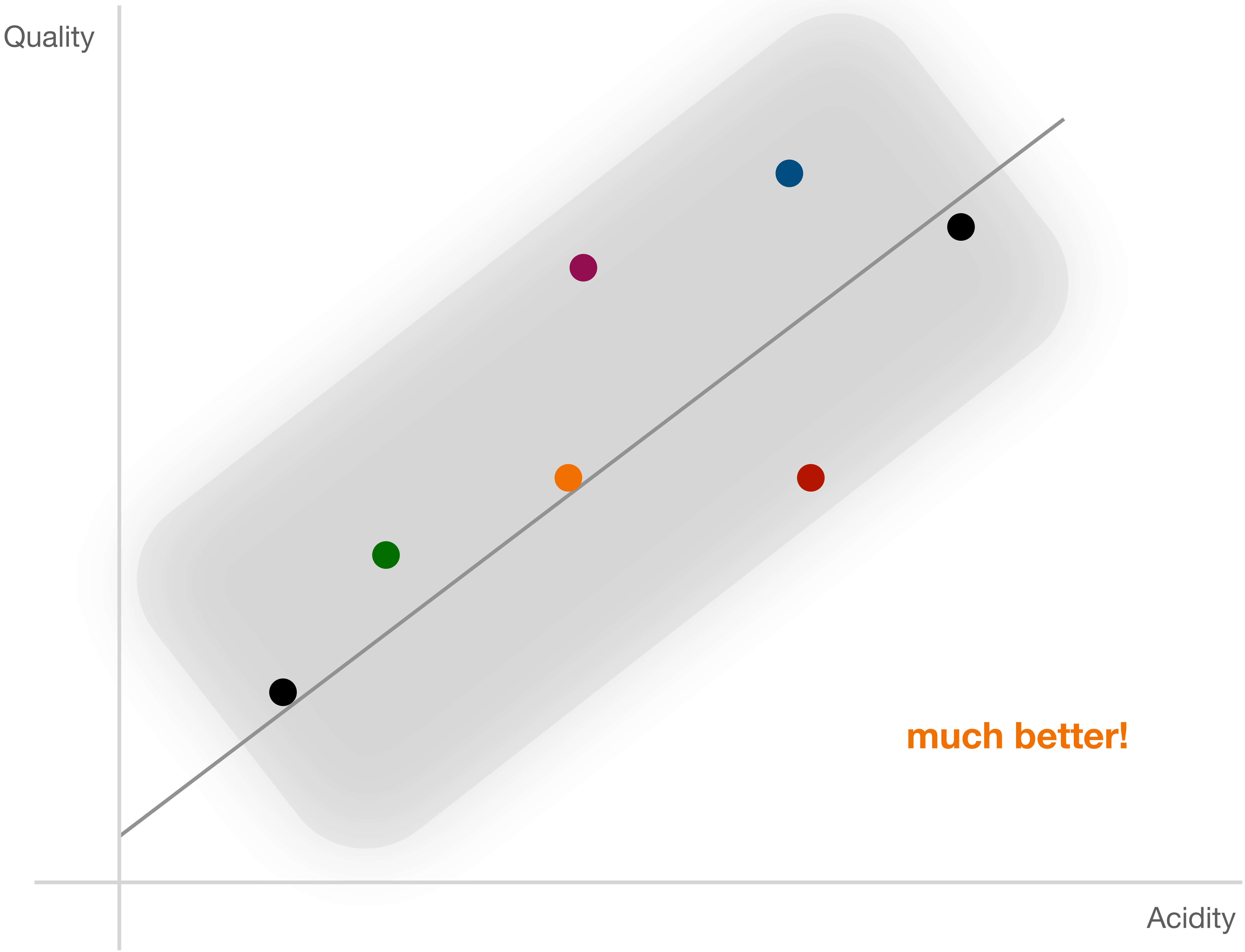
data



more

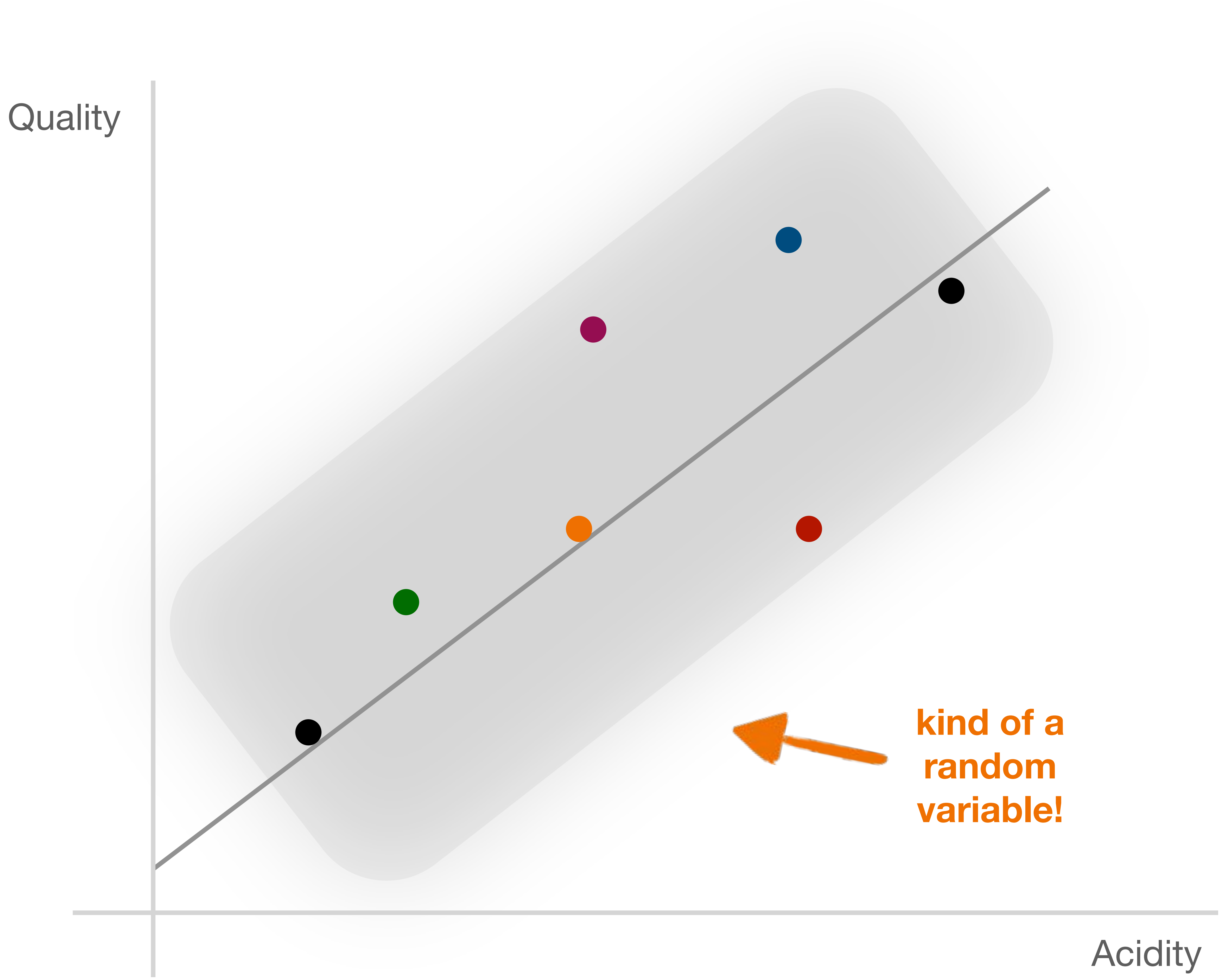
data





more

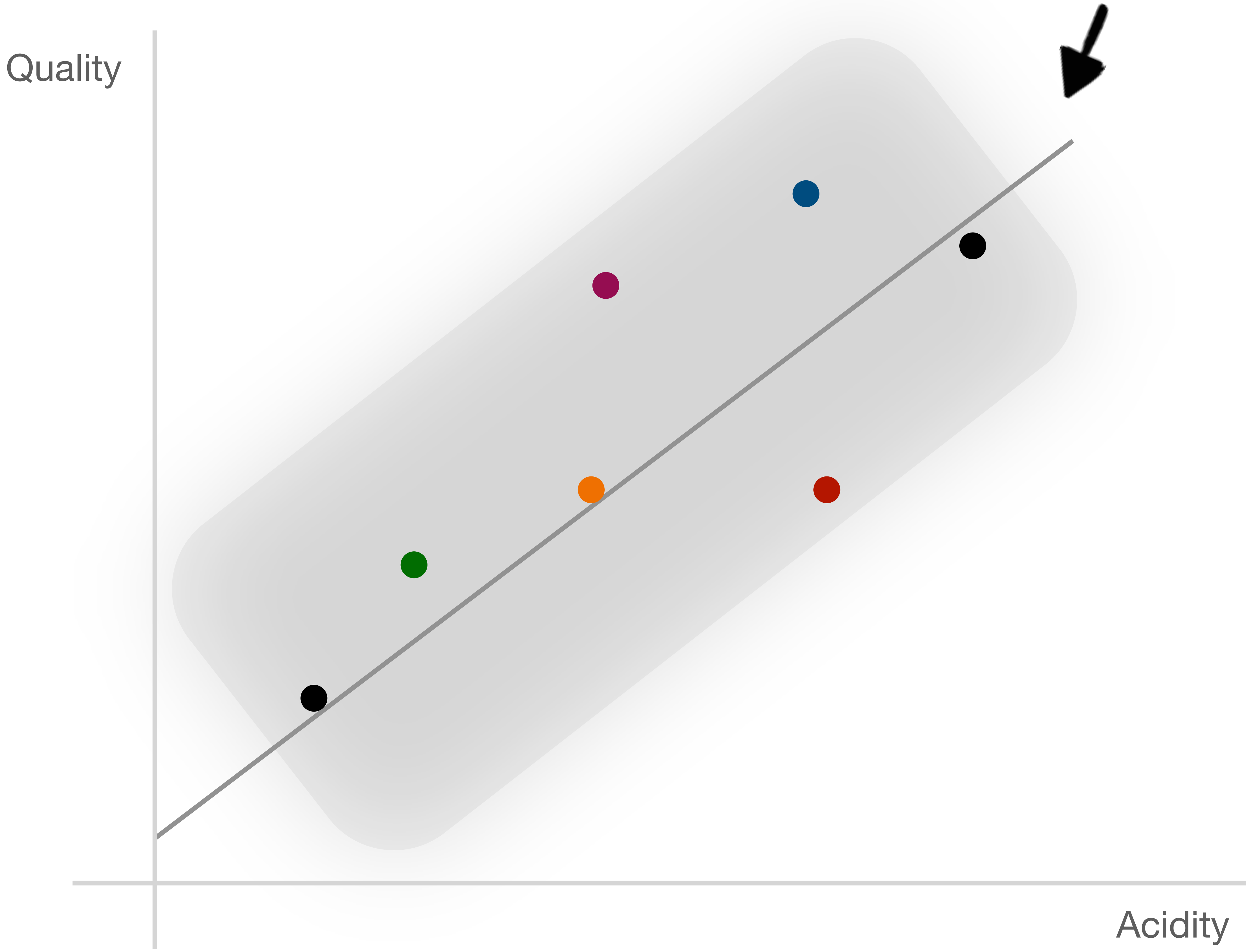
data



kind of a random variable!

more

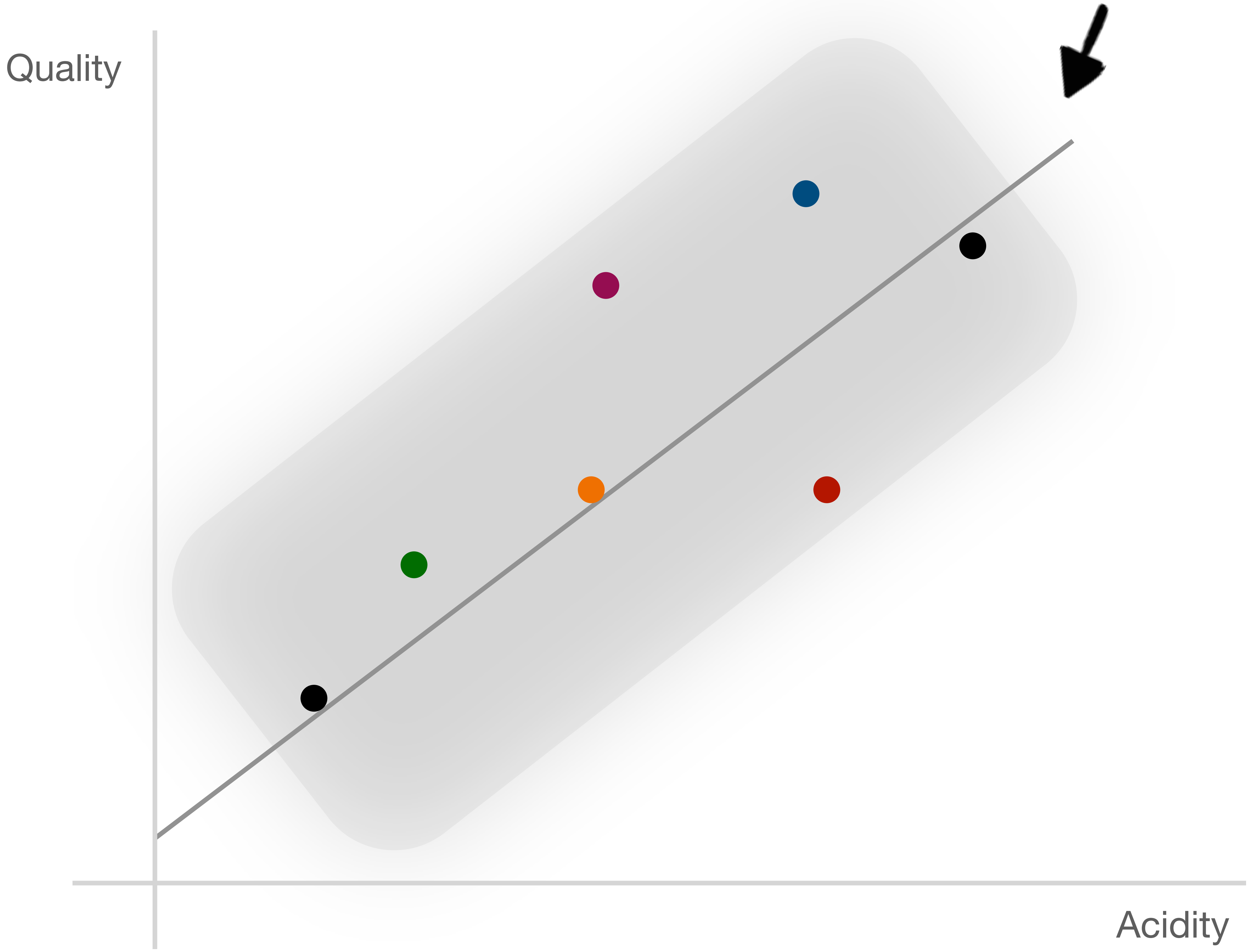
data

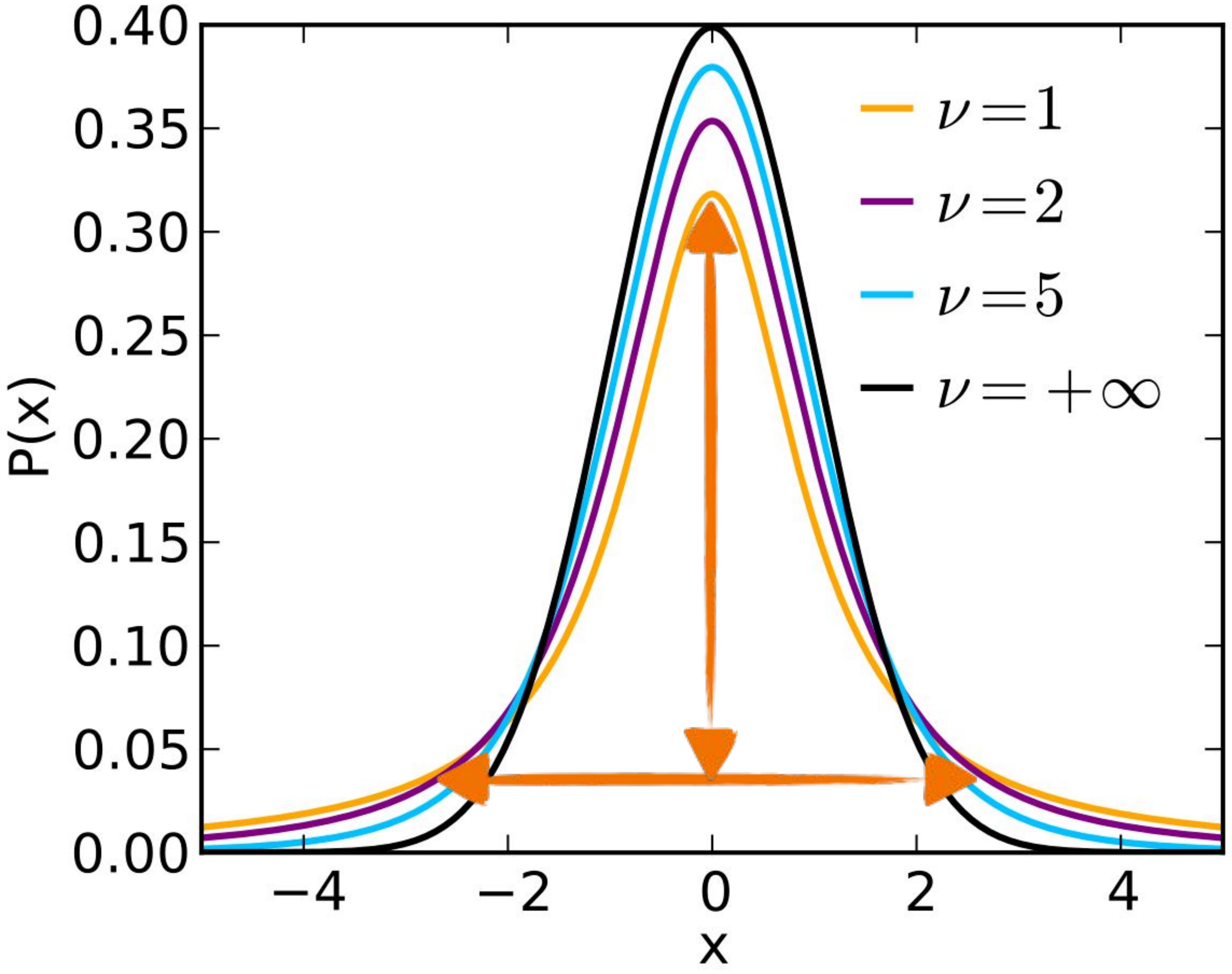


wait… this is a t-test!

more

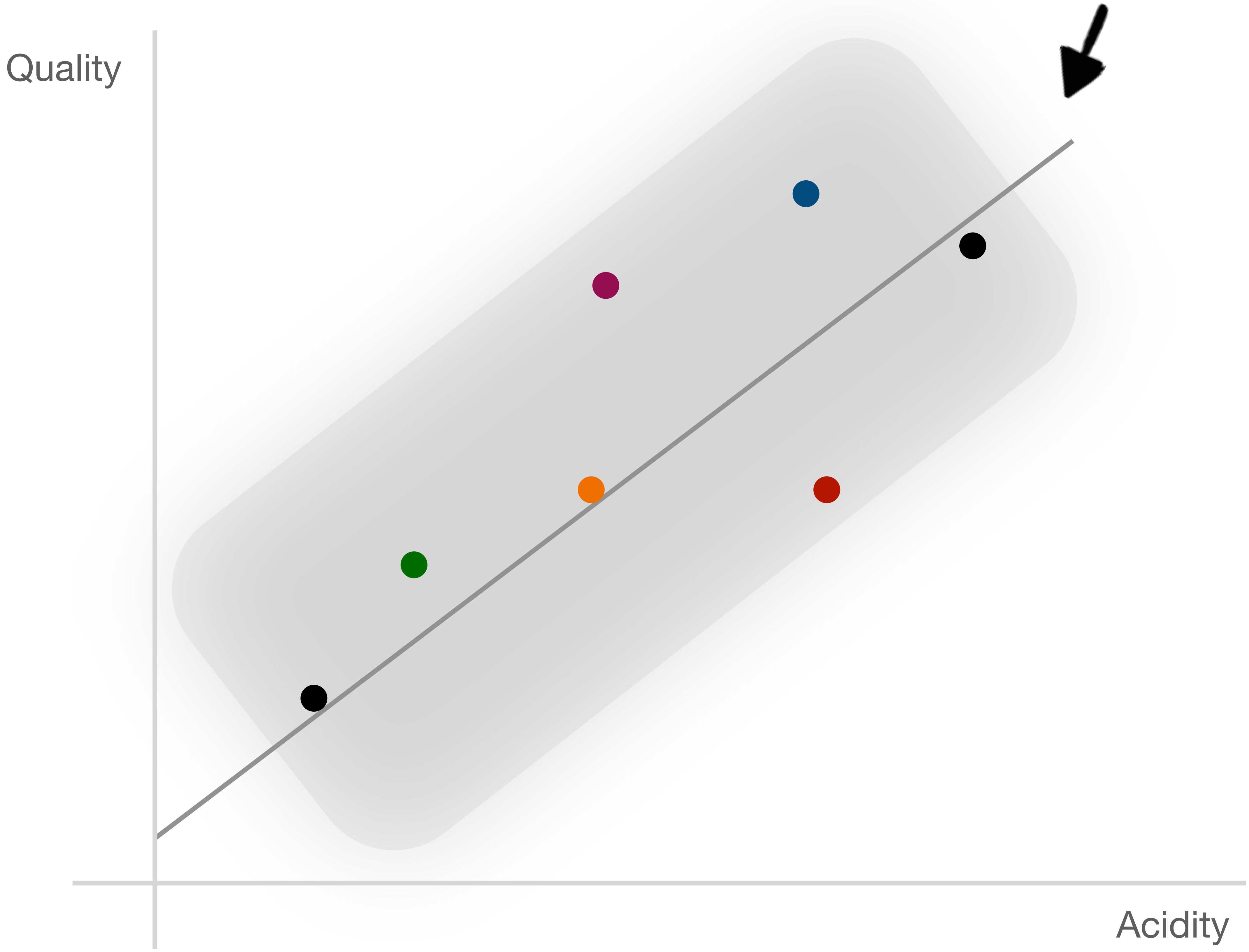
data

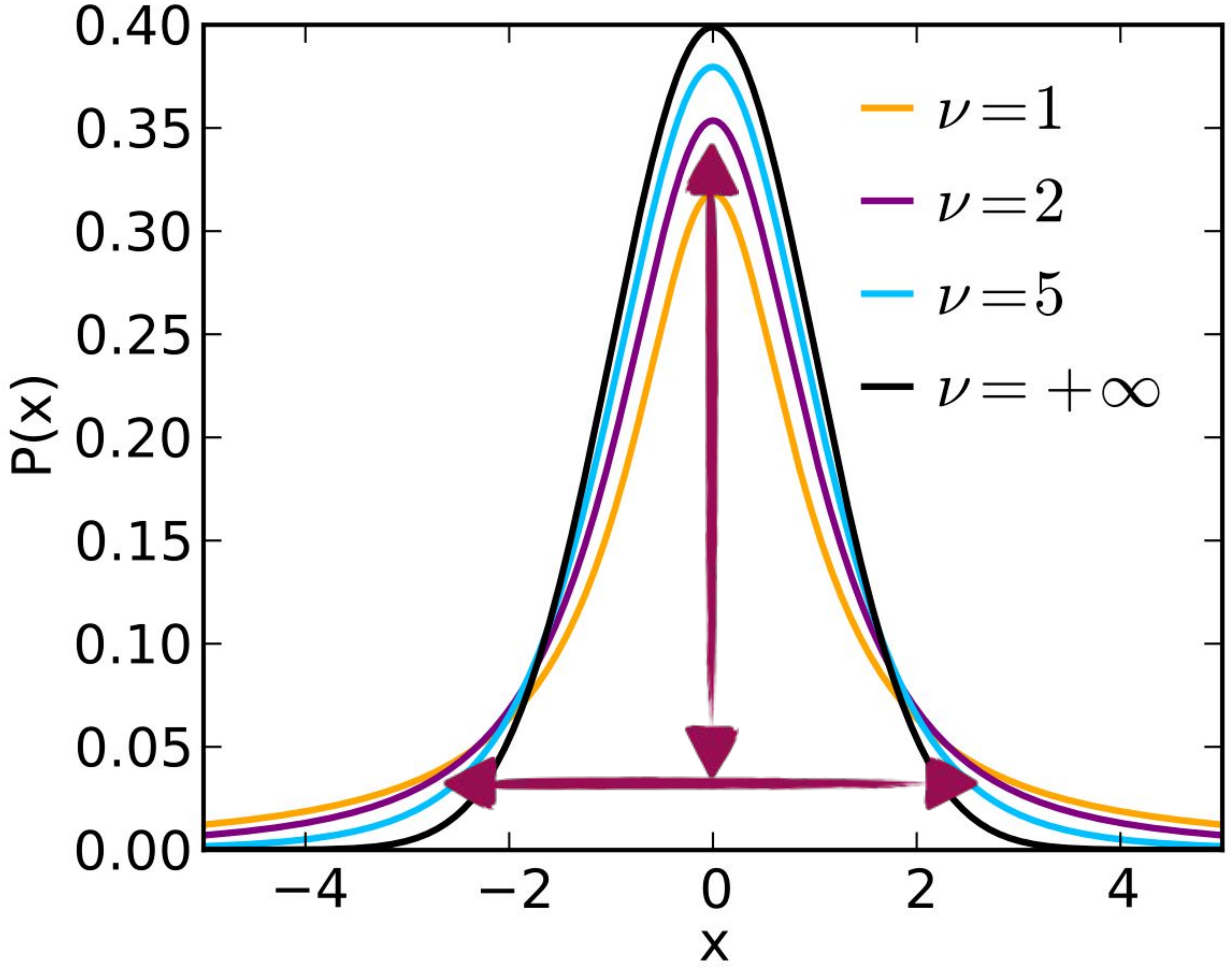

Quality
Acidity


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x)
−4
−2
0
2
4
x
$\nu=1$
$\nu=2$
$\nu=5$
$\nu=+\infty$

more

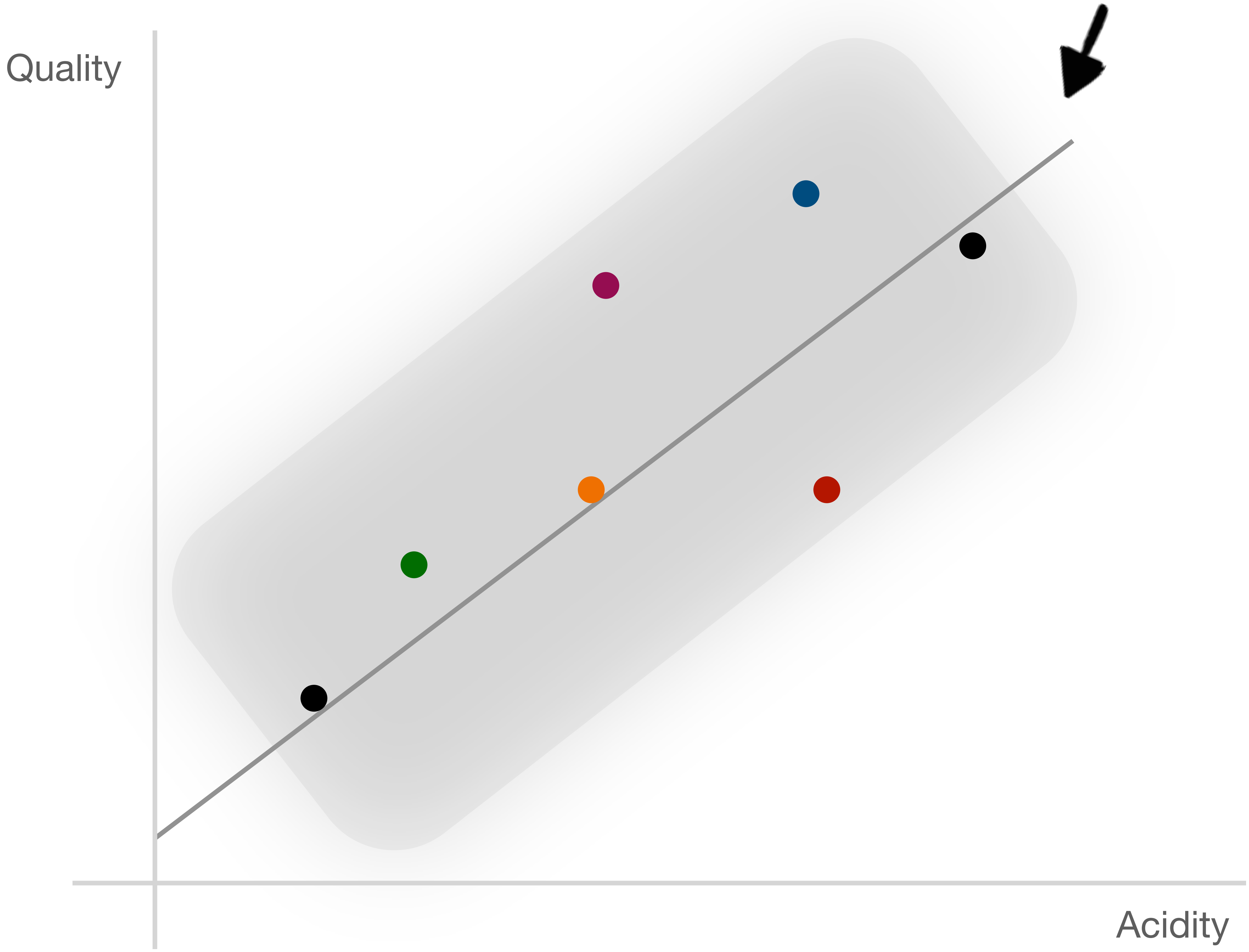
data

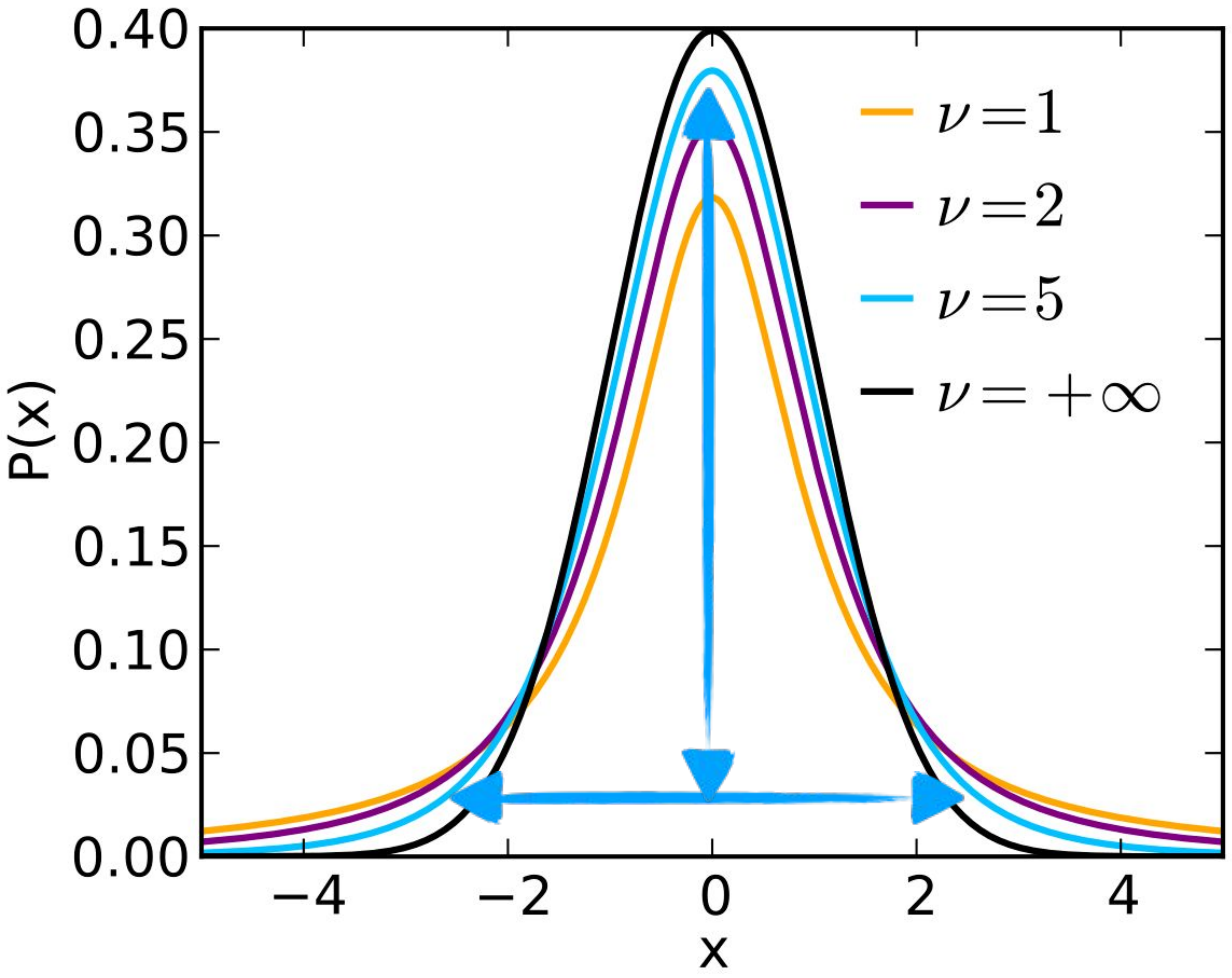

Quality
Acidity


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x)
−4
−2
0
2
4
x
$\nu = 1$
$\nu = 2$
$\nu = 5$
$\nu = +\infty$

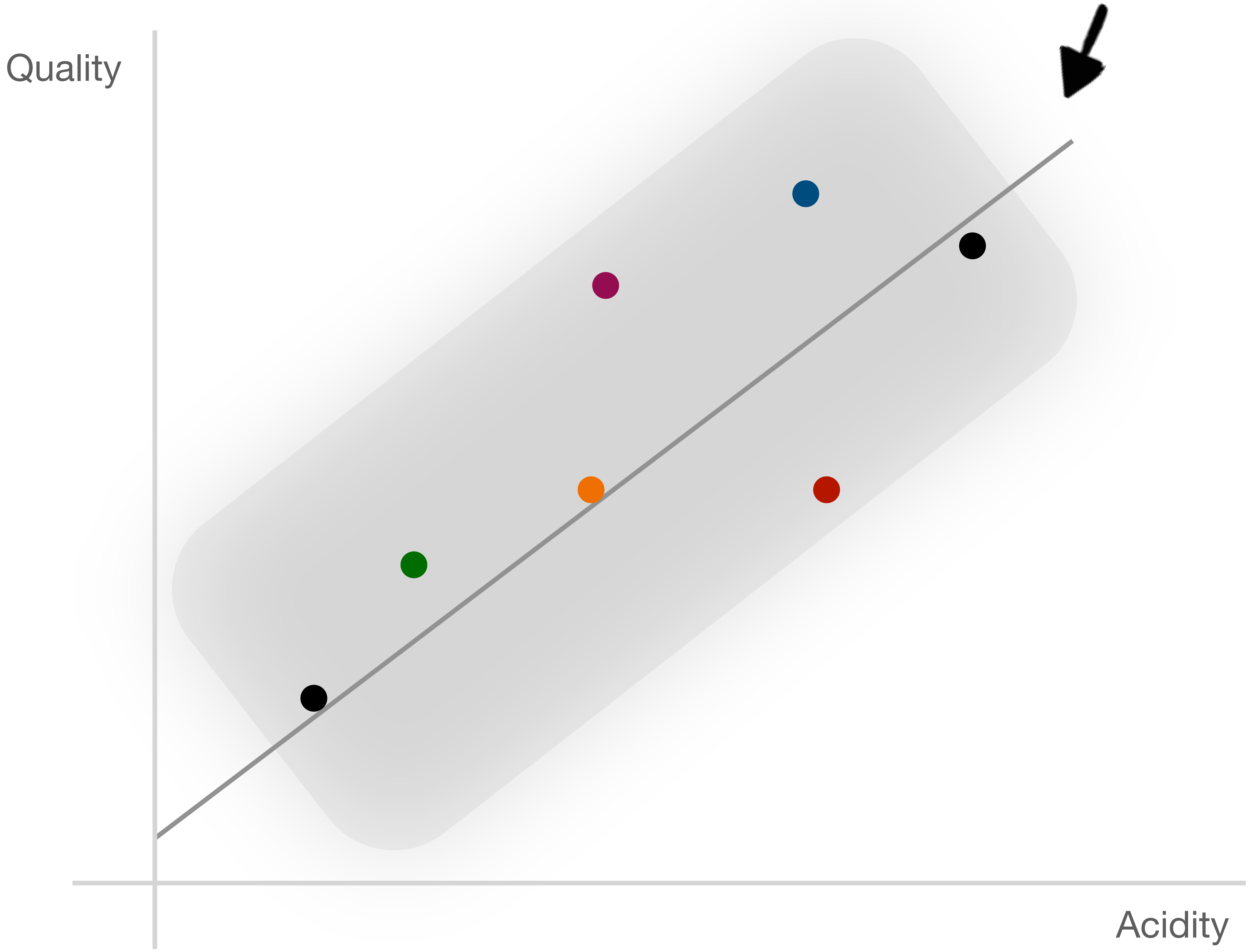
more

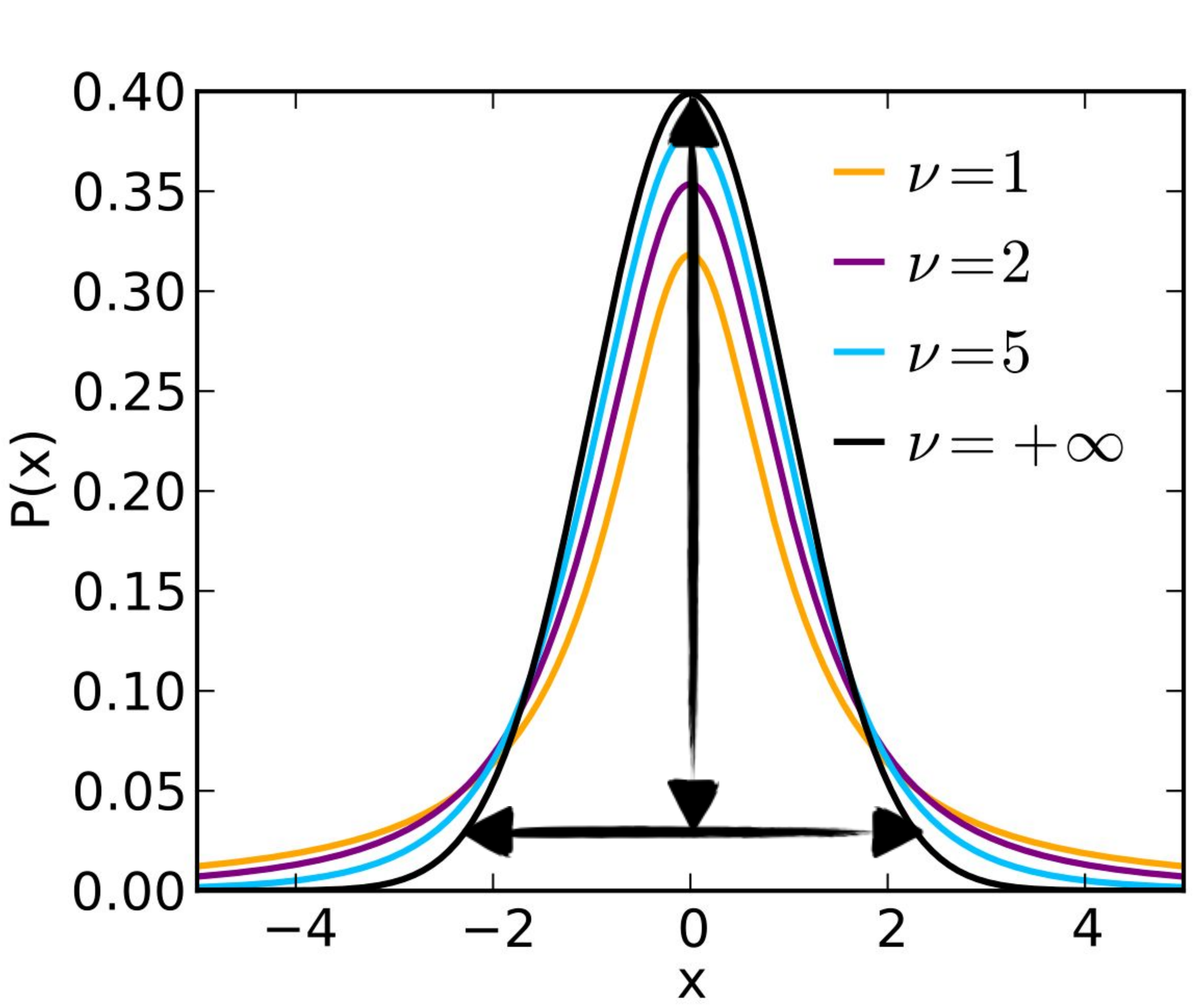
data


Quality
Acidity


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x)
−4
−2
0
2
4
x
$\nu=1$
$\nu=2$
$\nu=5$
$\nu=+\infty$

more data

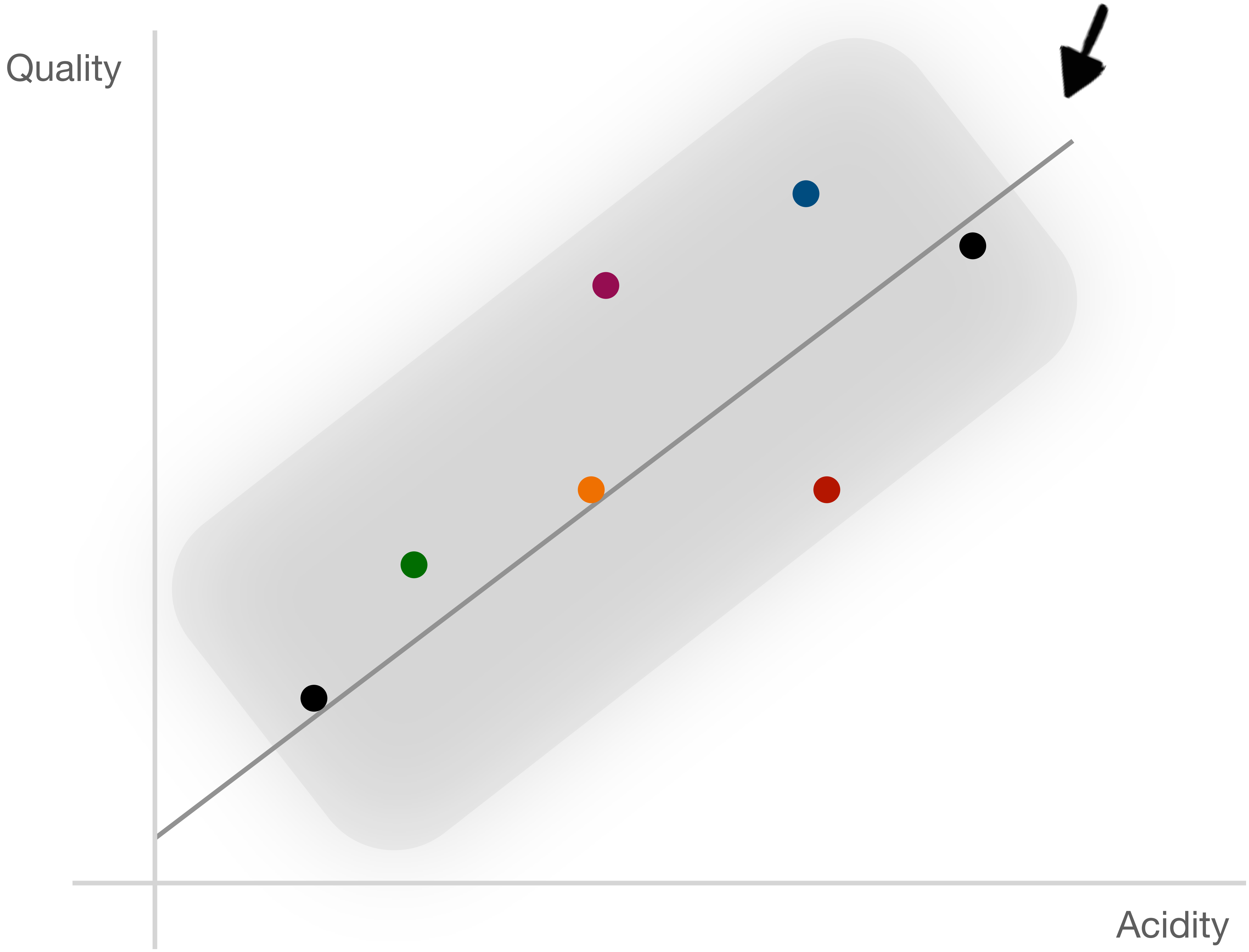


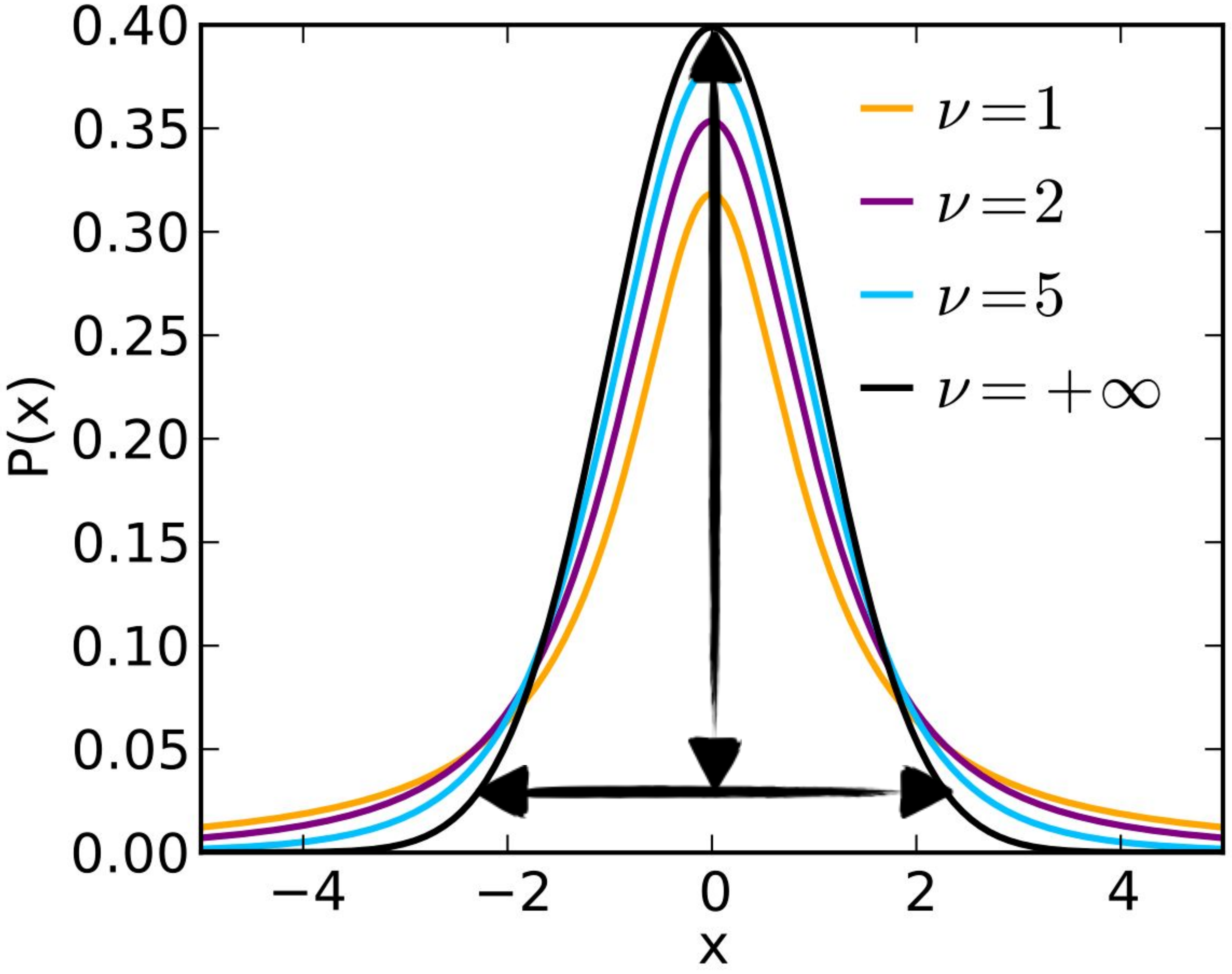


■ **increased degrees-of-freedom** increases the probability of the population equaling sample

more

data





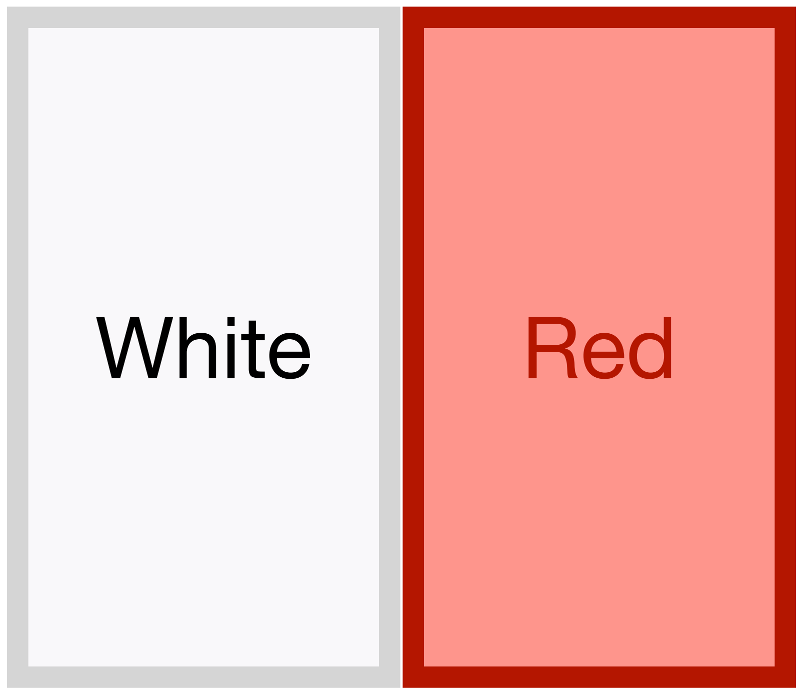
- ~~**increased degrees-of-freedom** increases the probability of the population equaling sample~~
- **more data, better line**

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

balanced data

Let’s think about logistic functions!

balanced data

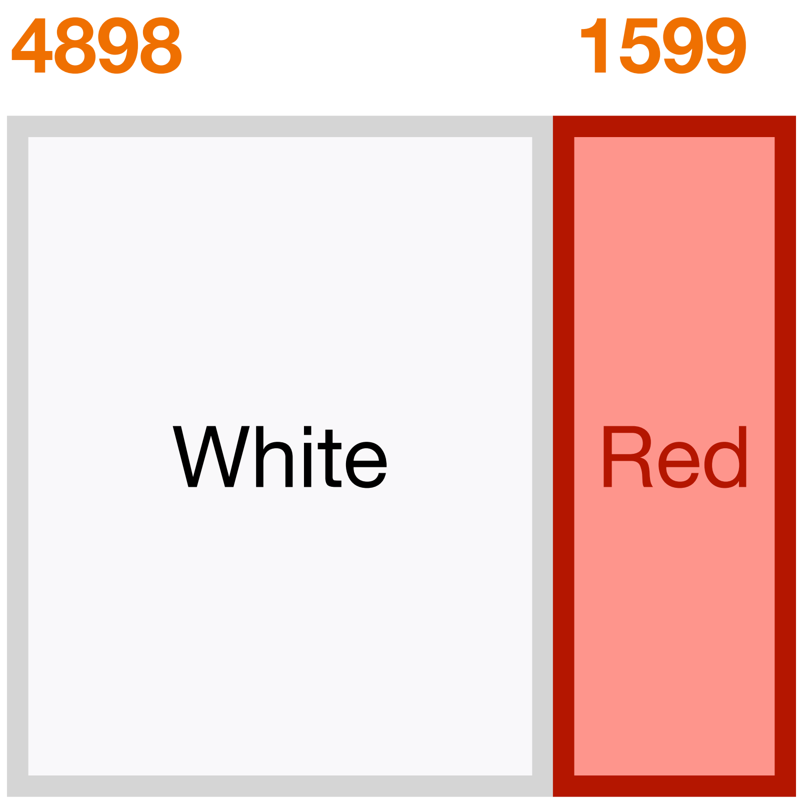

White

Red

in an ideal world

...but no

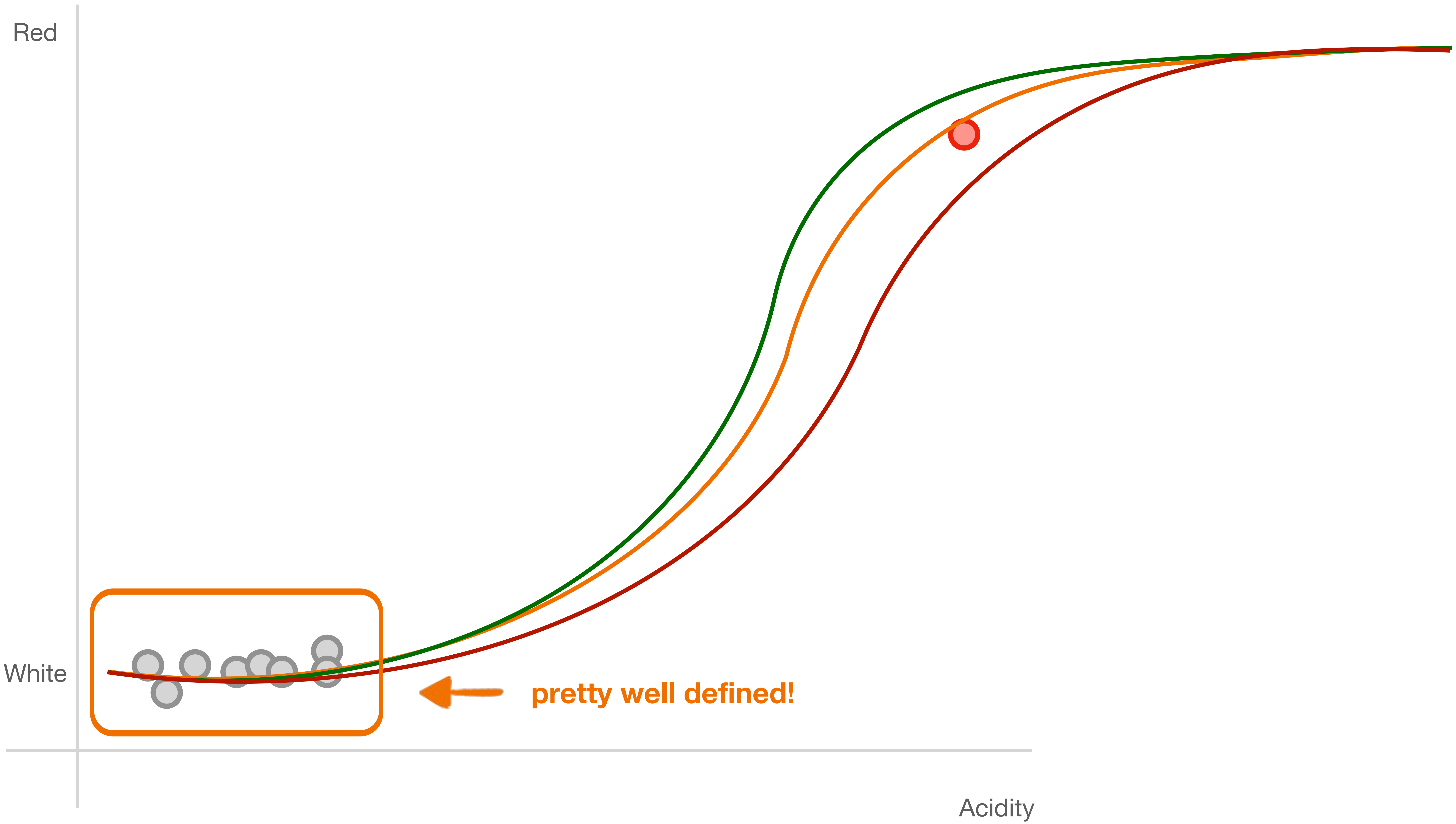
balanced data



What happens when we fit this dataset entirely?

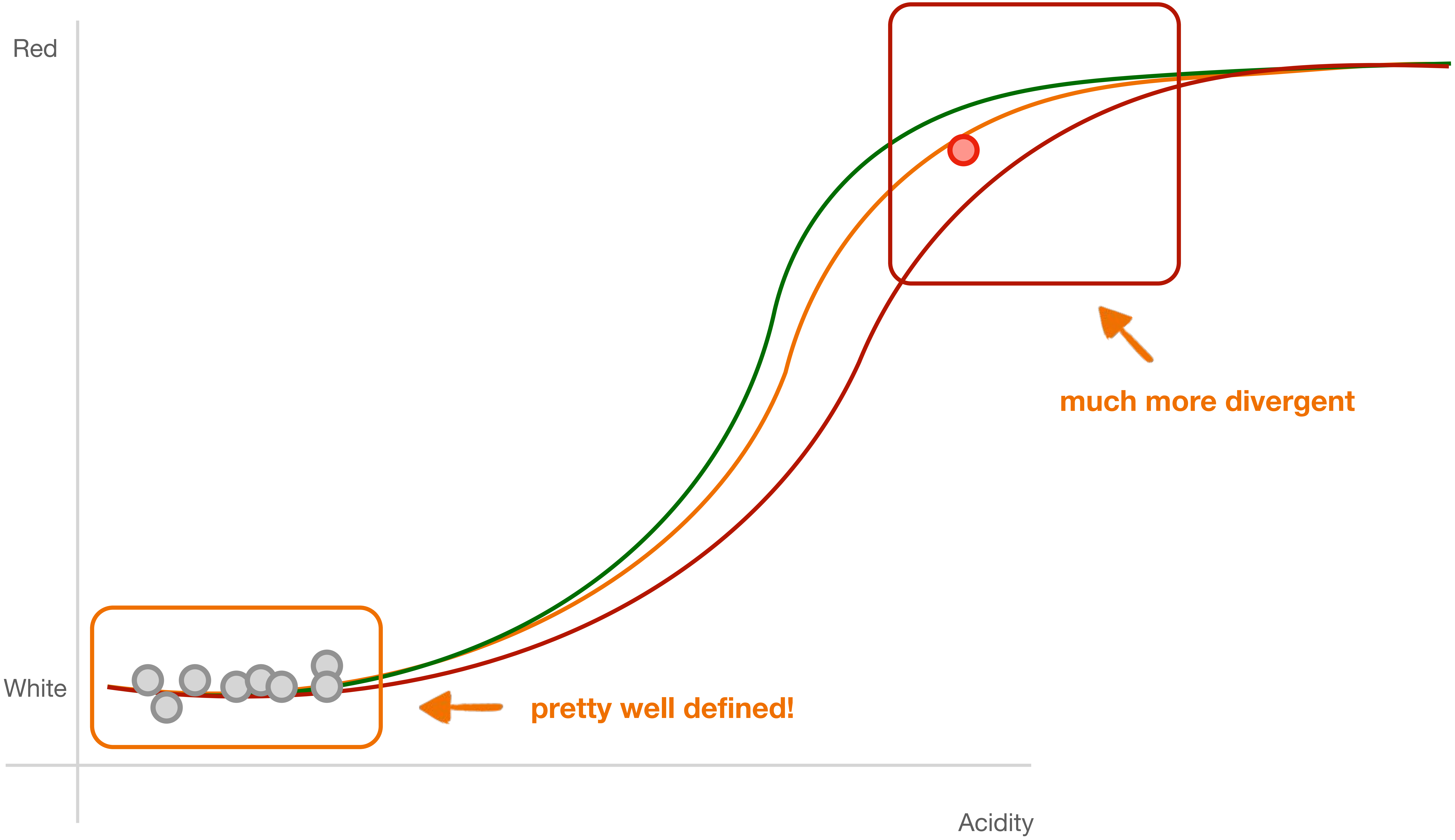
balanced data

Let’s think about logistic functions!



balanced data

Let's think about logistic functions!

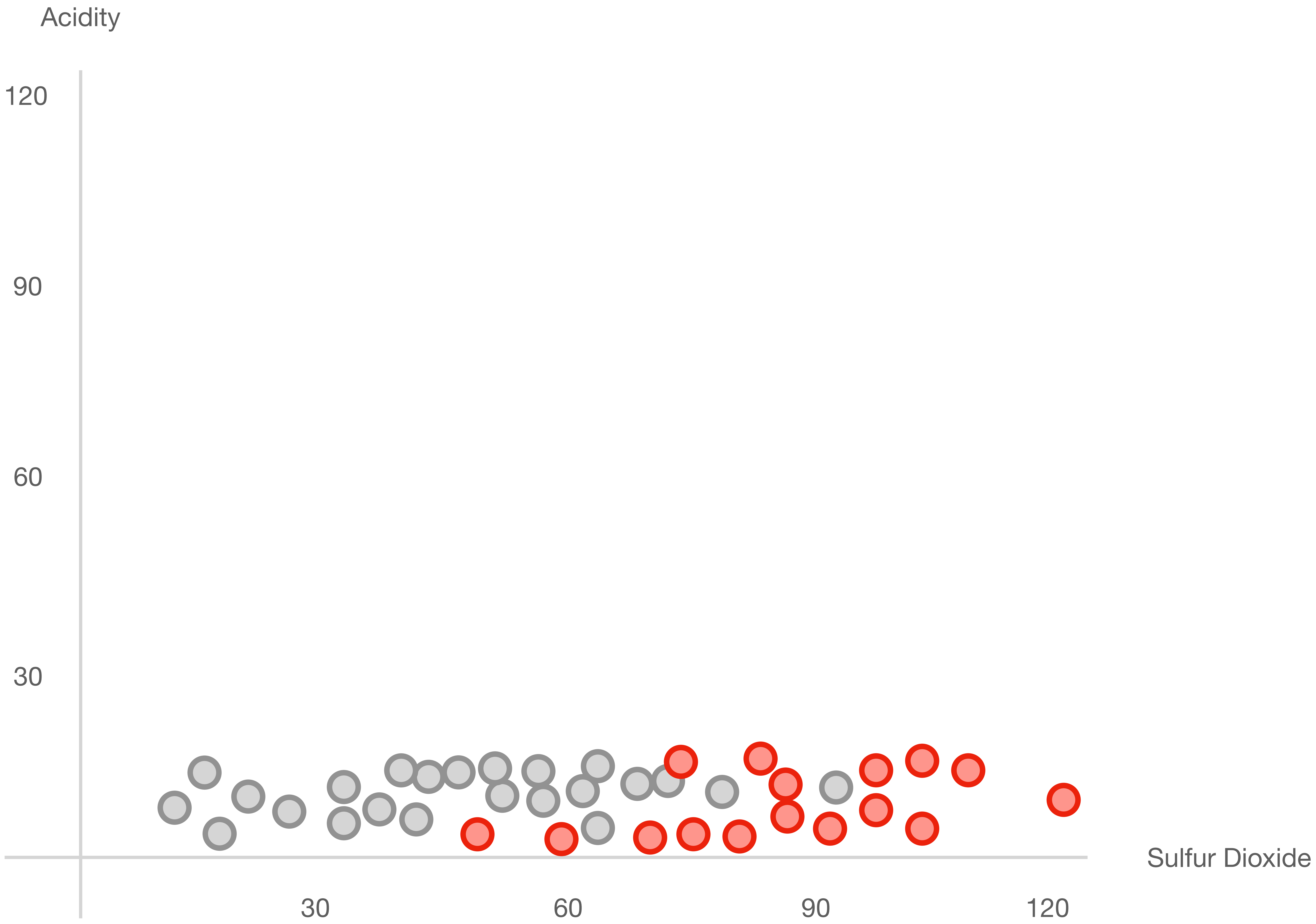


- **balanced data, more accurate results**

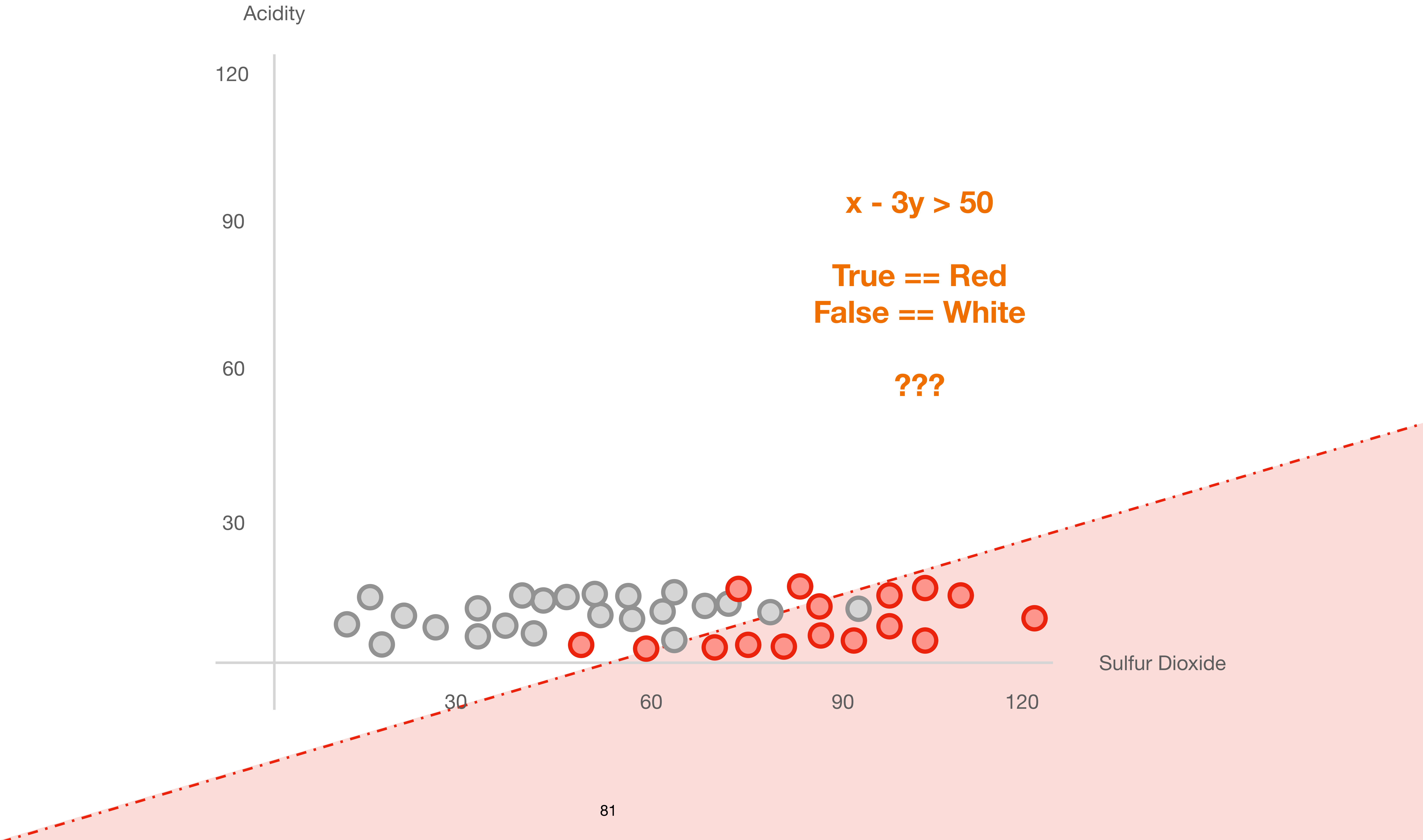
more data **balanced** data **normalized** data **quality** data

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

normalized data



normalized data

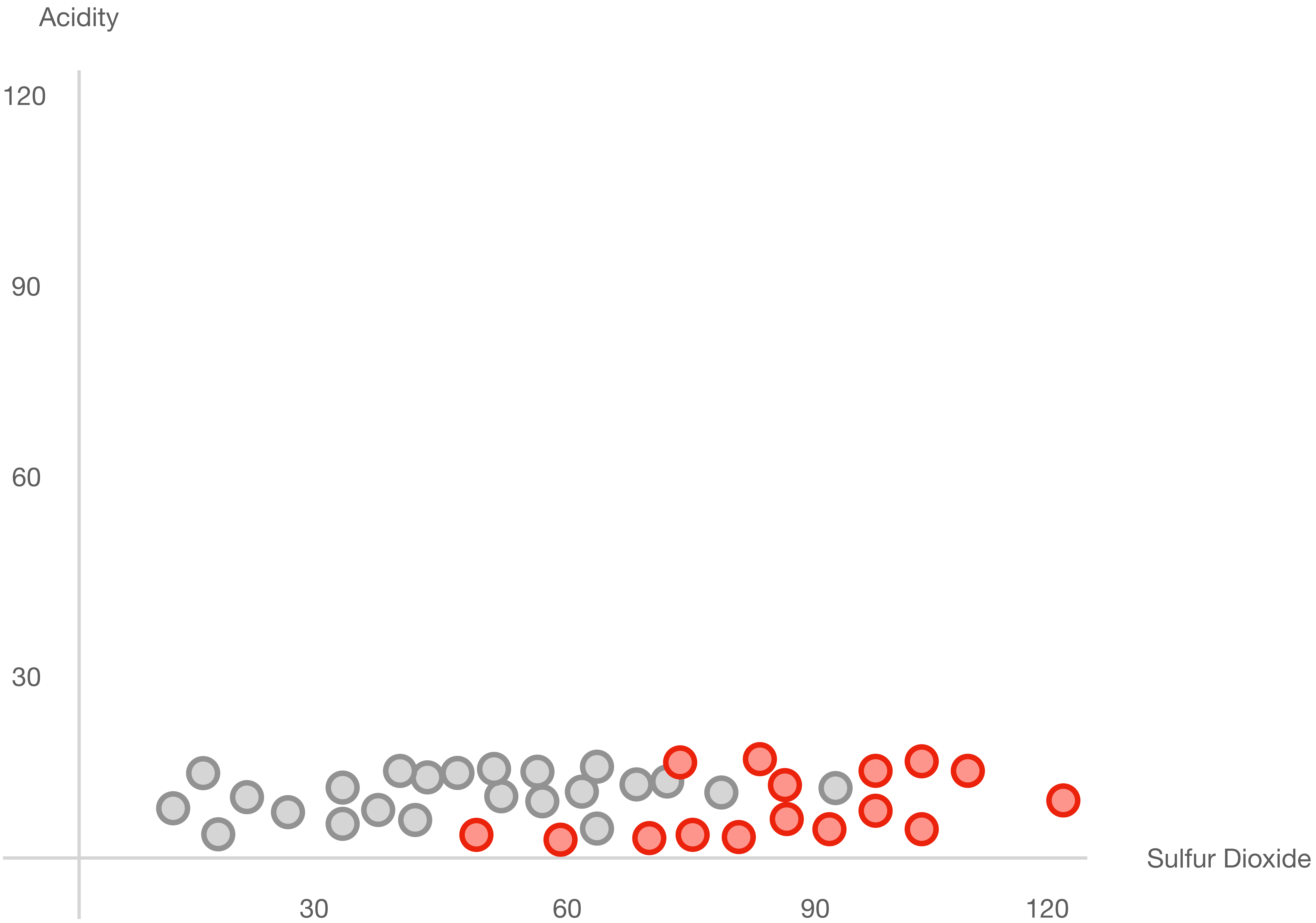


$x - 3y > 50$

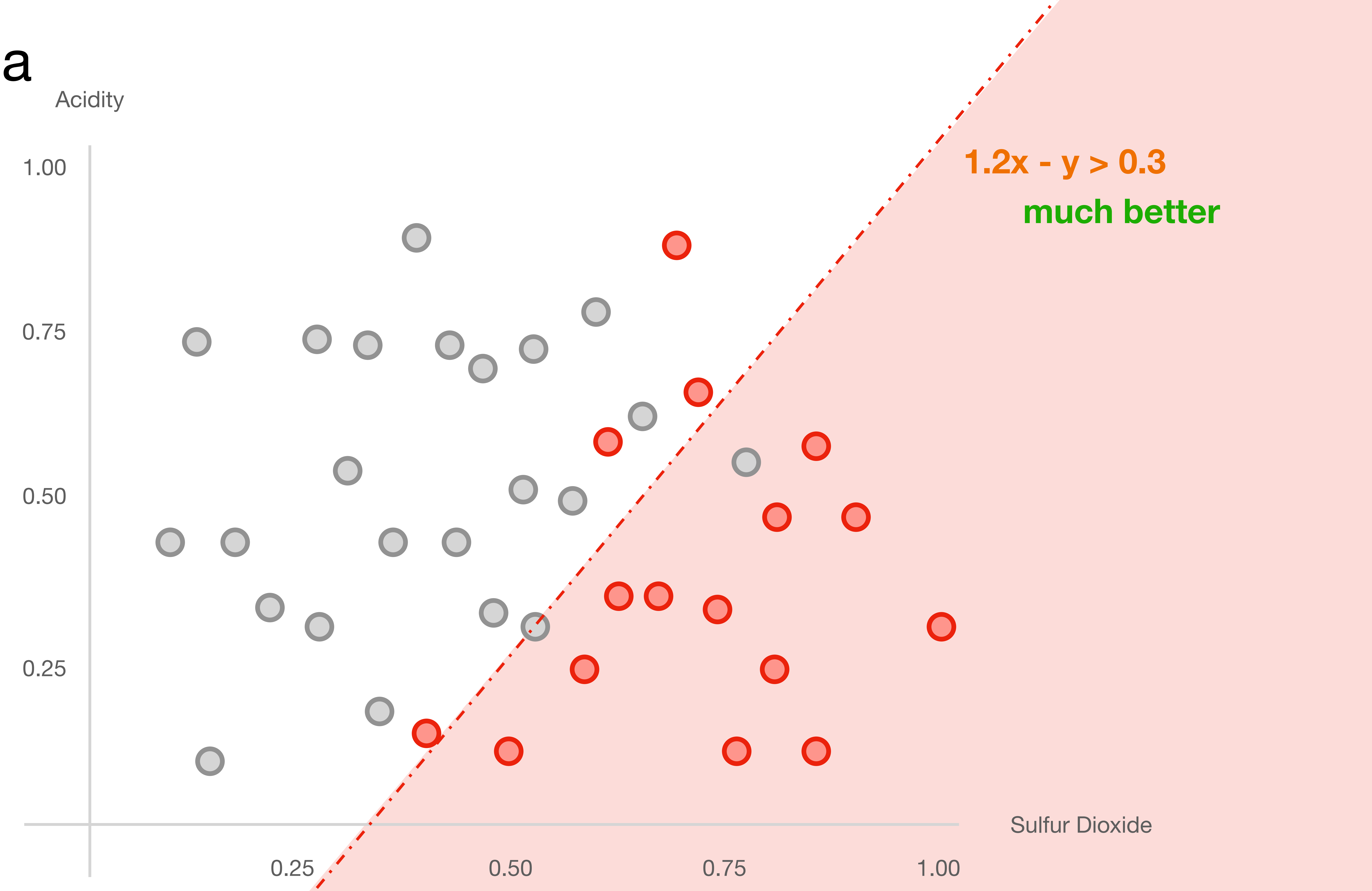
True == Red
False == White

???

normalized data



normalized data

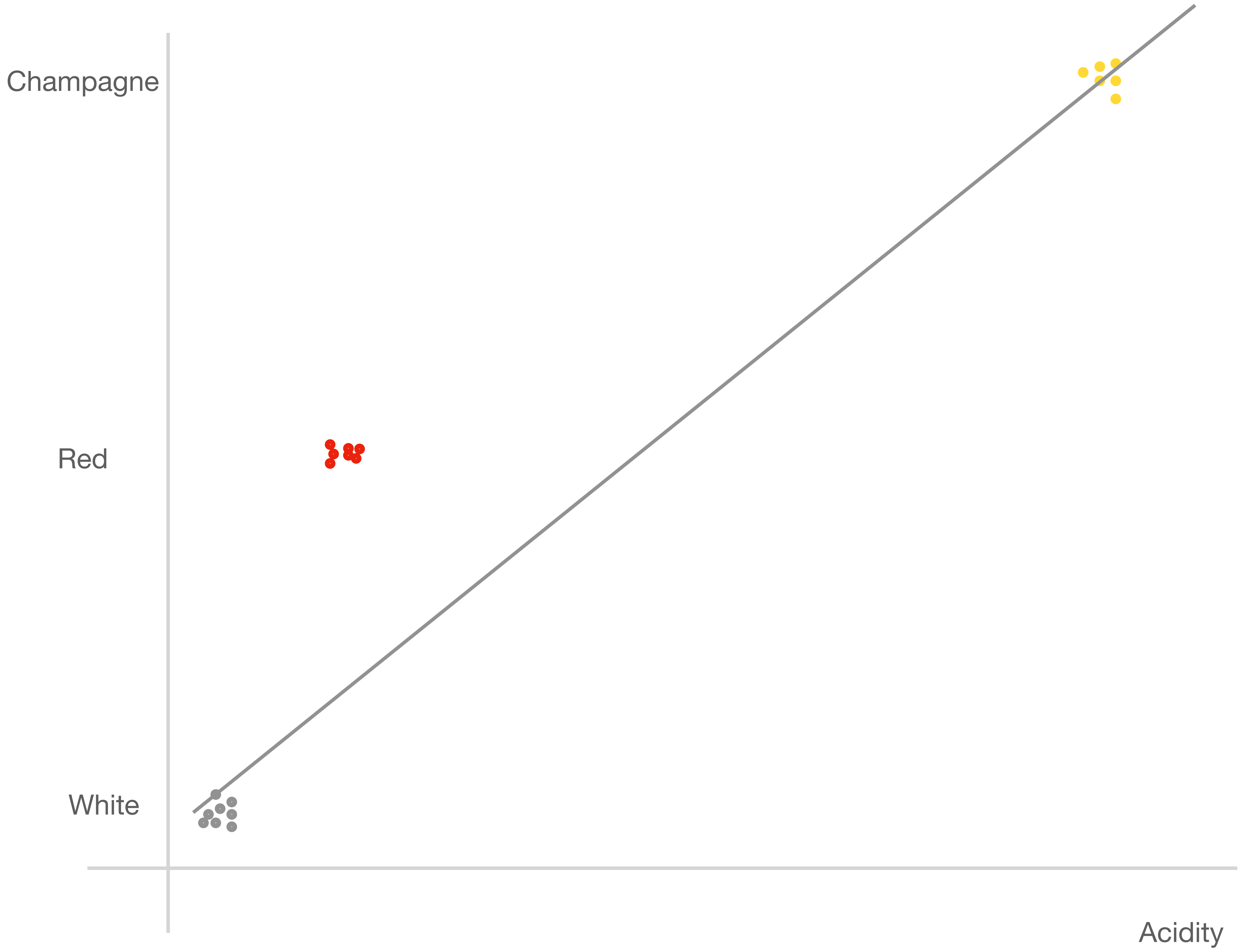


1.2x - y > 0.3

much better

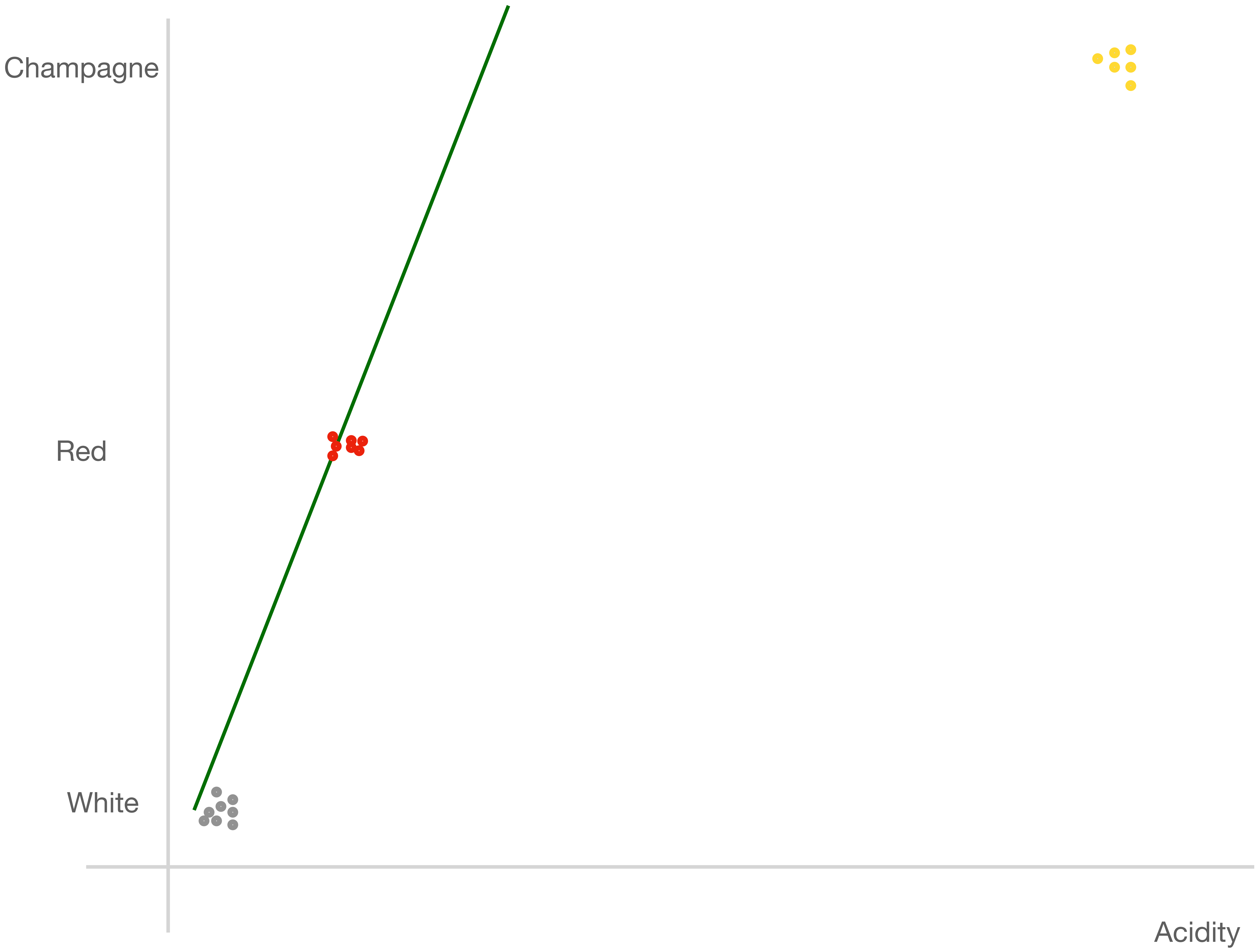
- **normalized data, better generalization, faster convergence**

normalized data



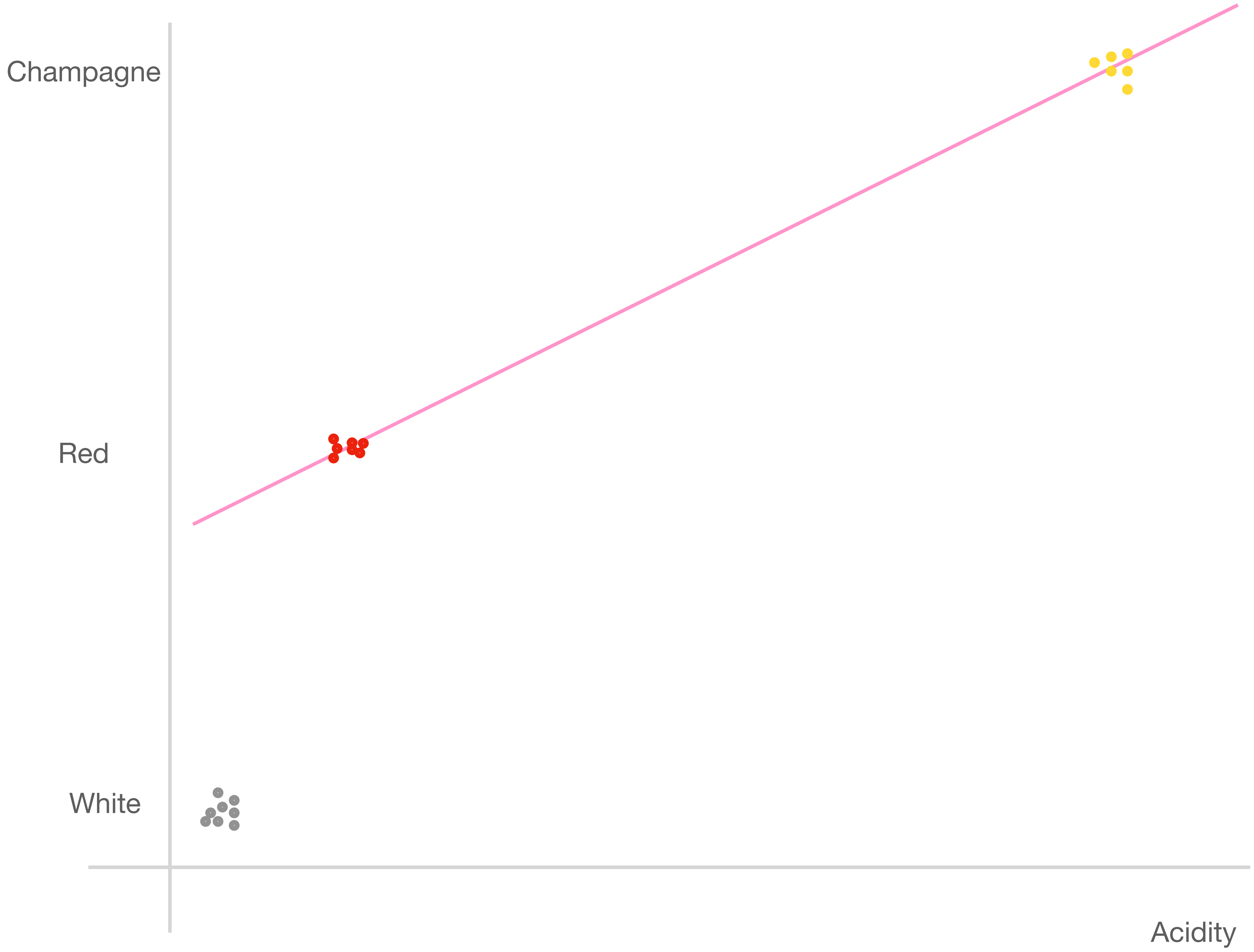
??? how to fit a line

normalized data



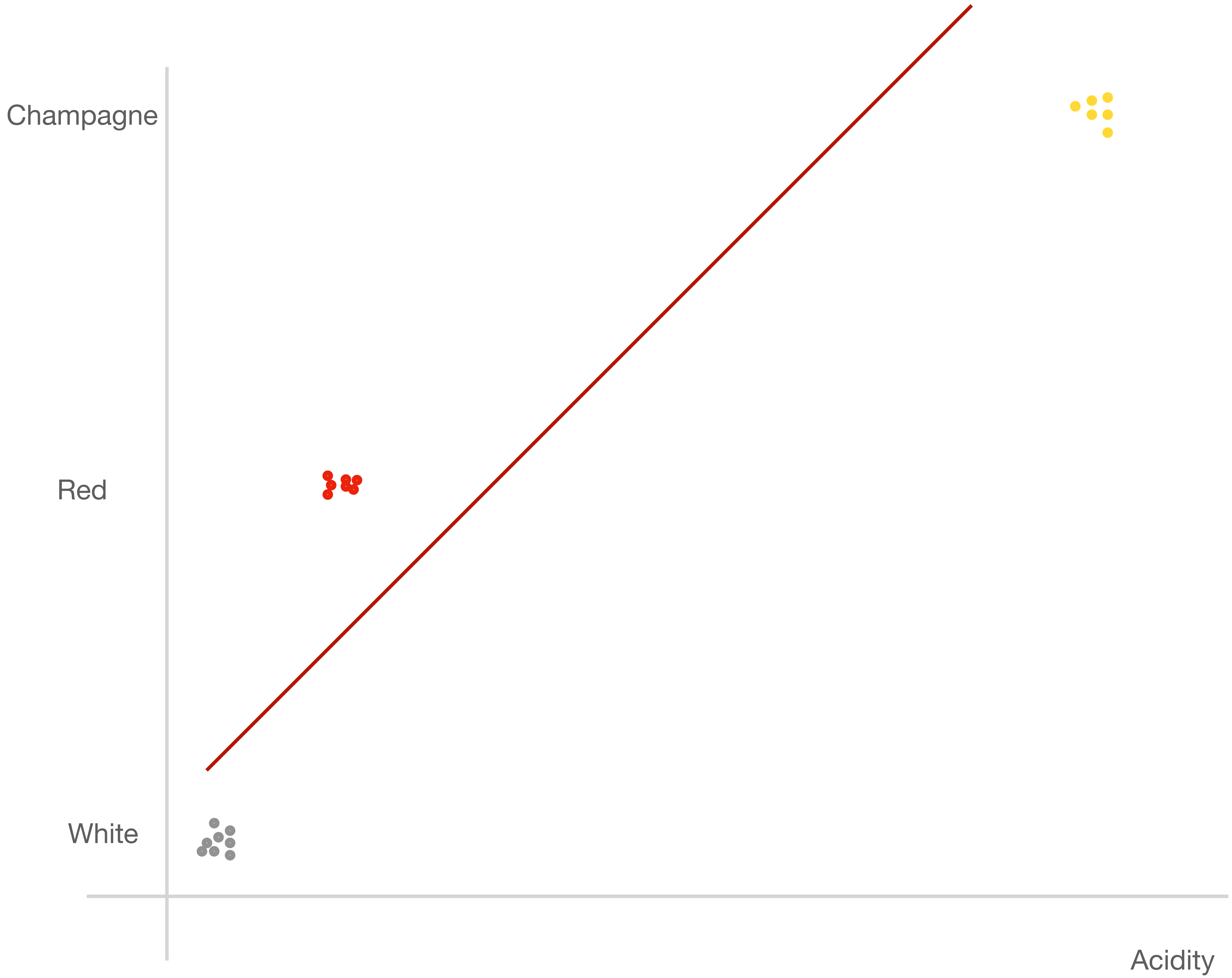
??? how to fit a line

normalized data



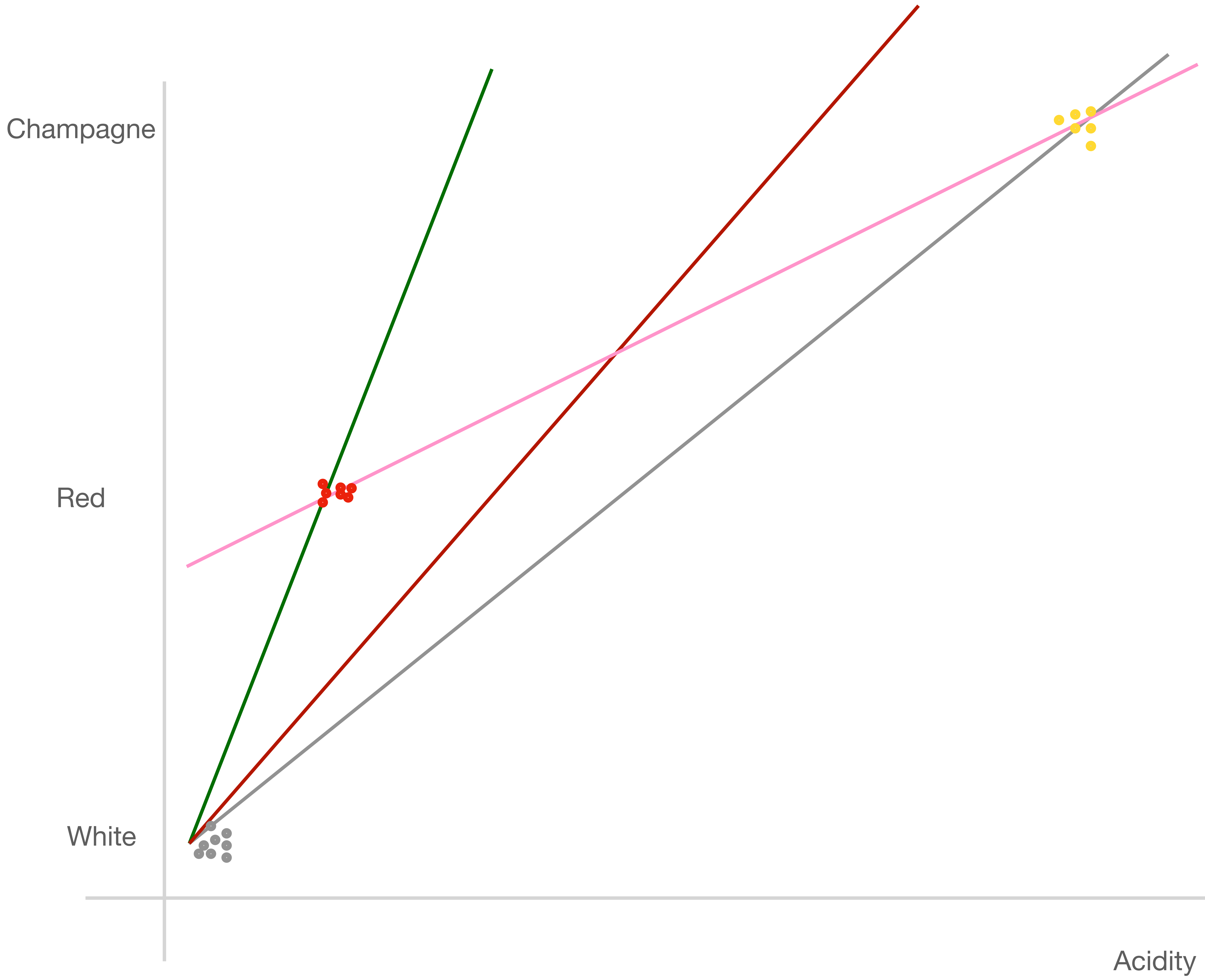
??? how to fit a line

normalized data



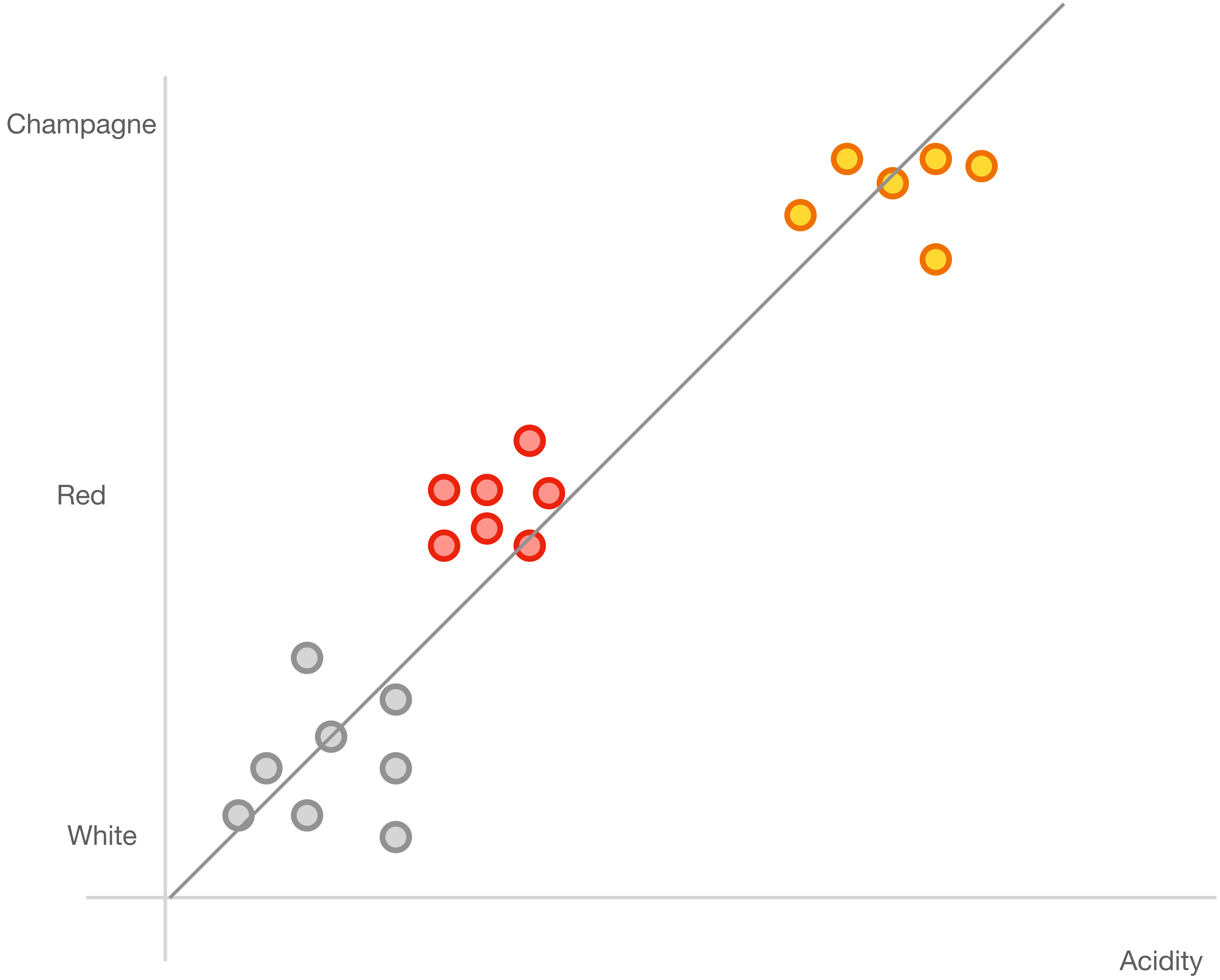
??? how to fit a line

normalized data



- non-normalized data is hard to fit

normalized data



- ensure all features are internally normalized (same order of mag.)

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

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

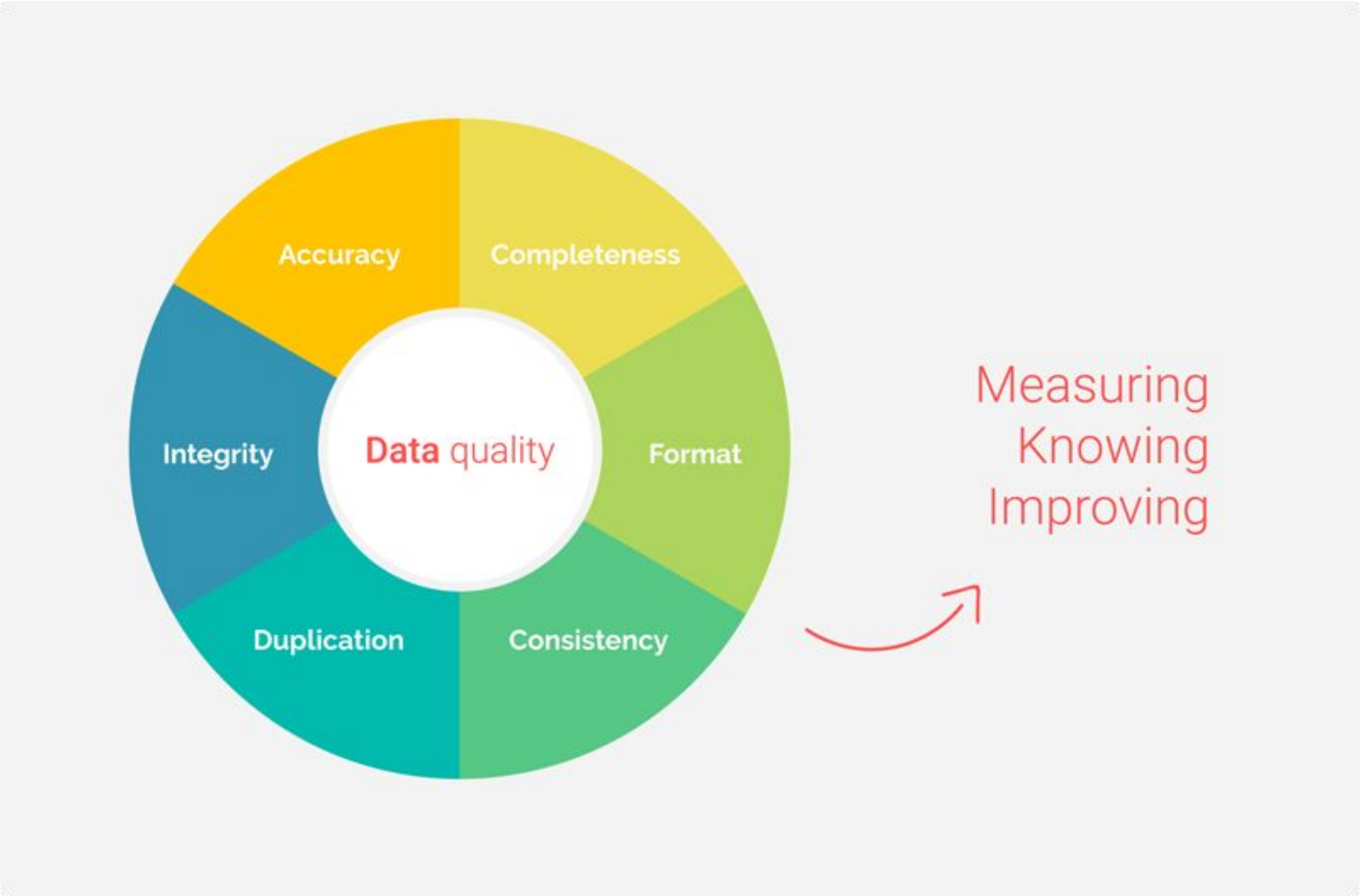


Image credit: Passionned Group

more data

balanced data

normalized data

quality data

Missing Data

Missing Data

Missing **completely at random**

Missing **at random**

Missing **not at random**

Missing Data

remove

Use mean/most often

regression

more data

balanced data

normalized data

let’s clean some data!