D3 Scales

dataviz rant

linear scales

square root scales

Analyzing Bad Dataviz

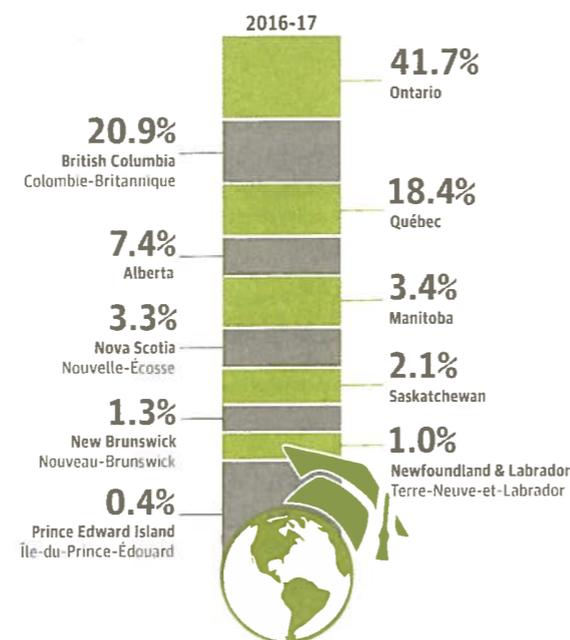## By the numbers
Statistiques sous la loupe

International post-secondary students in Canada
Portrait des étudiants internationaux au Canada

International and domestic post-secondary enrolment and tuition growth, 2006-07 to 2016-17
Croissance des inscriptions nationales et internationales de 2006-07 à 2016-17

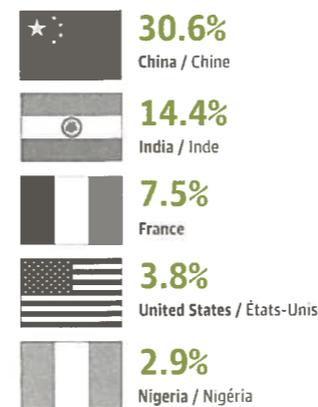■ International student population / Nombre total d'étudiants internationaux
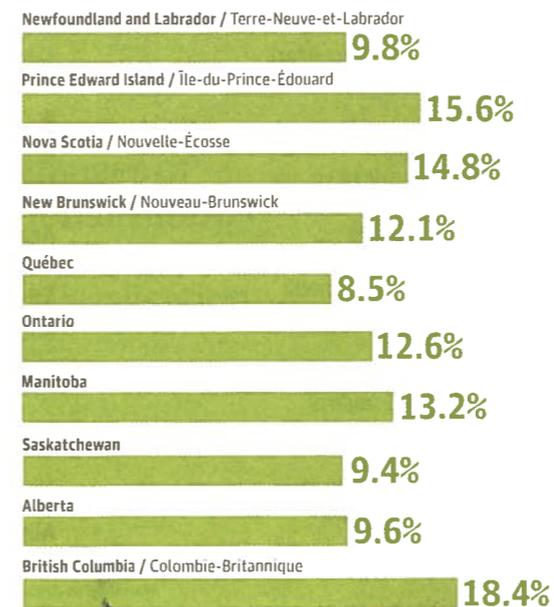■ Canadian student enrolment / Nombre total d'étudiants canadiens

134%

12%

Distribution of post-secondary international student enrolment across Canada
Répartition des étudiants internationaux à travers le pays

2016-17

41.7% Ontario

20.9% British Columbia Colombie-Britannique

18.4% Québec

7.4% Alberta

3.4% Manitoba

3.3% Nova Scotia Nouvelle-Écosse

2.1% Saskatchewan

1.3% New Brunswick Nouveau-Brunswick

1.0% Newfoundland & Labrador Terre-Neuve-et-Labrador

0.4% Prince Edward Island Île-du-Prince-Édouard

International post-secondary students in Canada, top 5 countries of citizenship in 2016-17
Nombre d'étudiants internationaux au Canada, top 5 selon la citoyenneté en 2016-17

30.6% China / Chine

14.4% India / Inde

7.5% France

3.8% United States / États-Unis

2.9% Nigeria / Nigéria

Post-secondary international enrolment as a share of total enrolment within each province in 2016-17
Effectifs et diplômés postsecondaires par rapport aux inscriptions totales dans chaque province en 2016-17

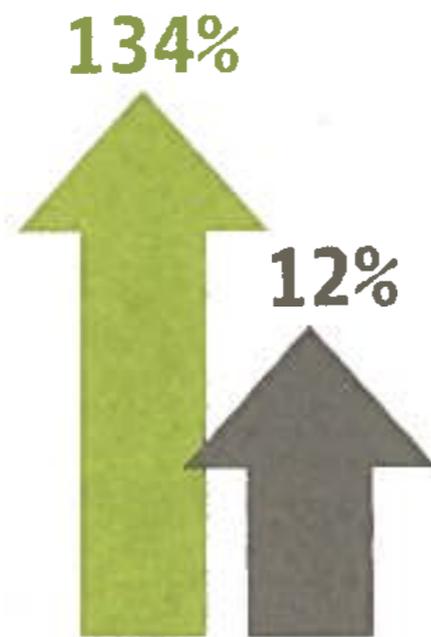| Province	Share
Newfoundland and Labrador / Terre-Neuve-et-Labrador	9.8%
Prince Edward Island / Île-du-Prince-Édouard	15.6%
Nova Scotia / Nouvelle-Écosse	14.8%
New Brunswick / Nouveau-Brunswick	12.1%
Québec	8.5%
Ontario	12.6%
Manitoba	13.2%
Saskatchewan	9.4%
Alberta	9.6%
British Columbia / Colombie-Britannique	18.4%

Source: Statistics Canada, PSIS / Statistique Canada, EDPC

International and domestic post-secondary enrolment and tuition growth, 2006-07 to 2016-17
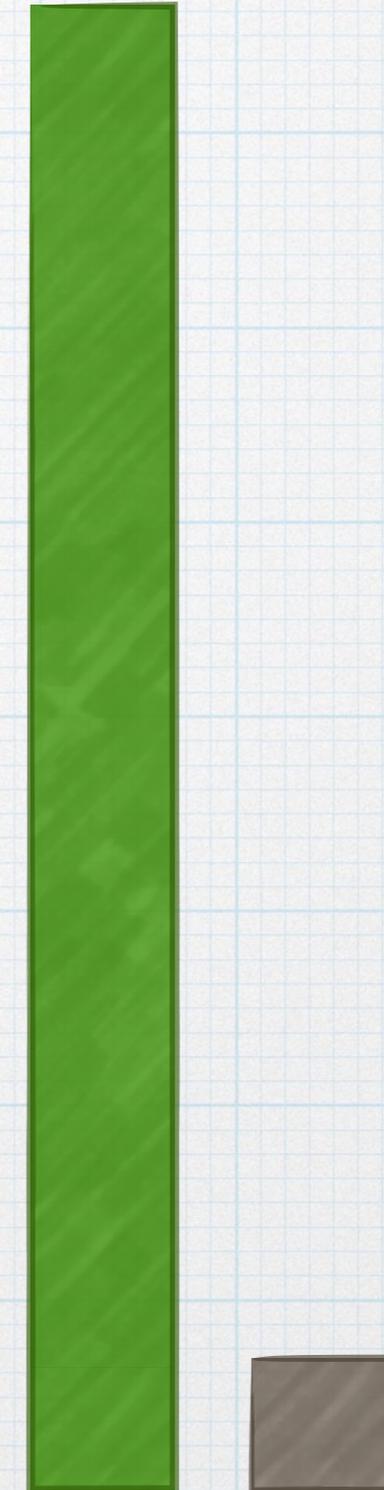Croissance des inscriptions nationales et internationales de 2006-07 à 2016-17

- International student population / Nombre total d'étudiants internationaux
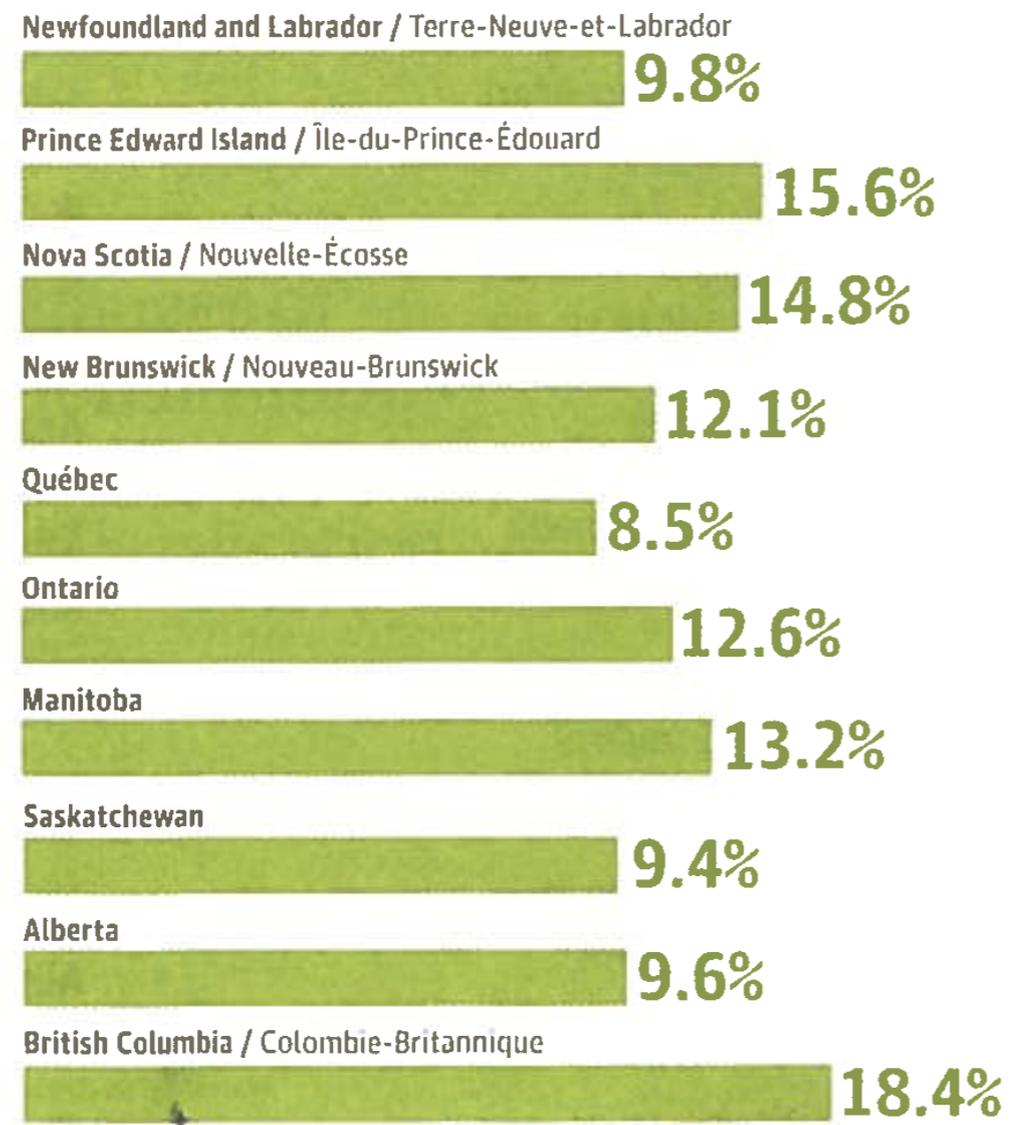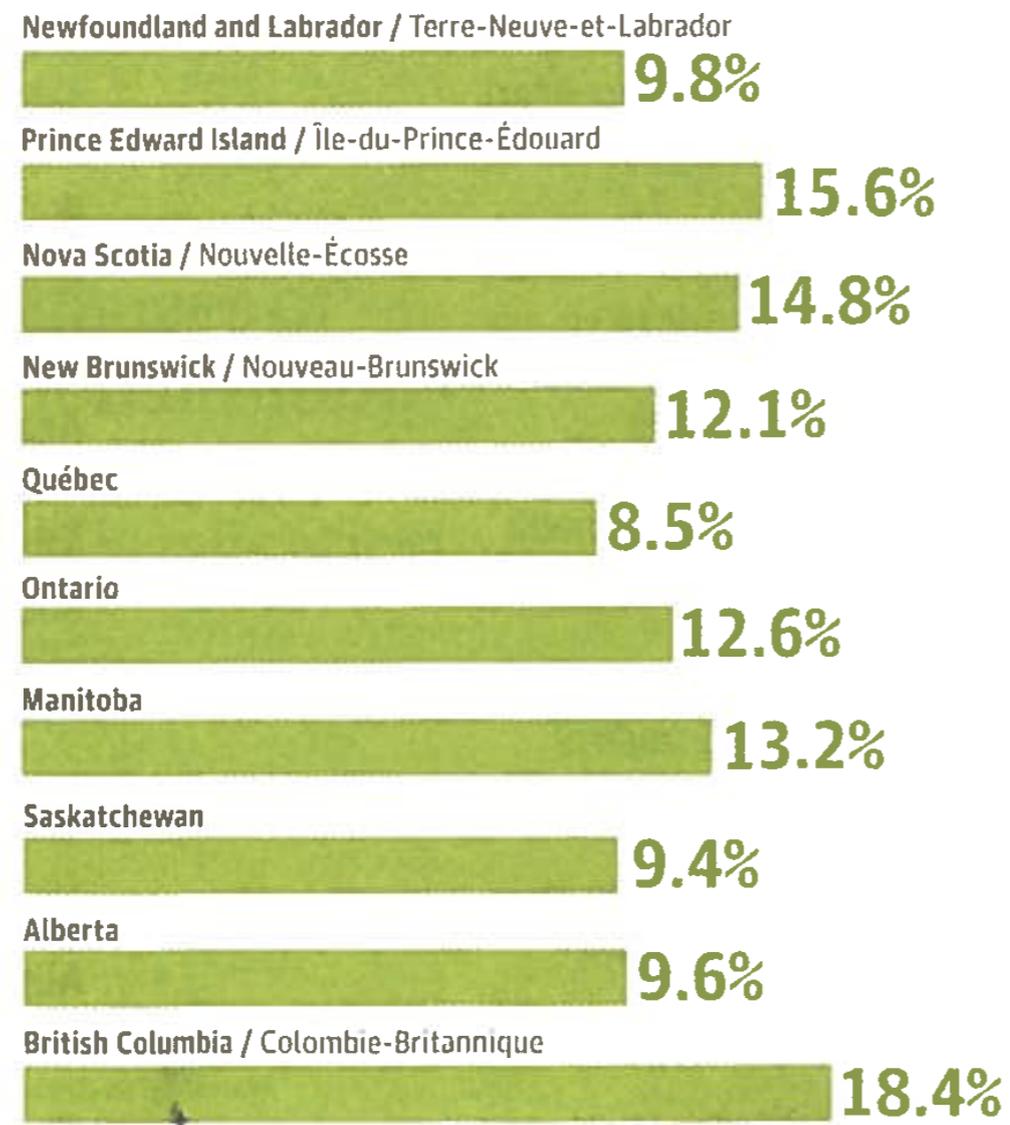- Canadian student enrolment / Nombre total d'étudiants canadiens

134%

12%

International and domestic post-secondary enrolment and tuition growth, 2006-07 to 2016-17
Croissance des inscriptions nationales et internationales de 2006-07 à 2016-17

■ International student population / Nombre total d'étudiants internationaux
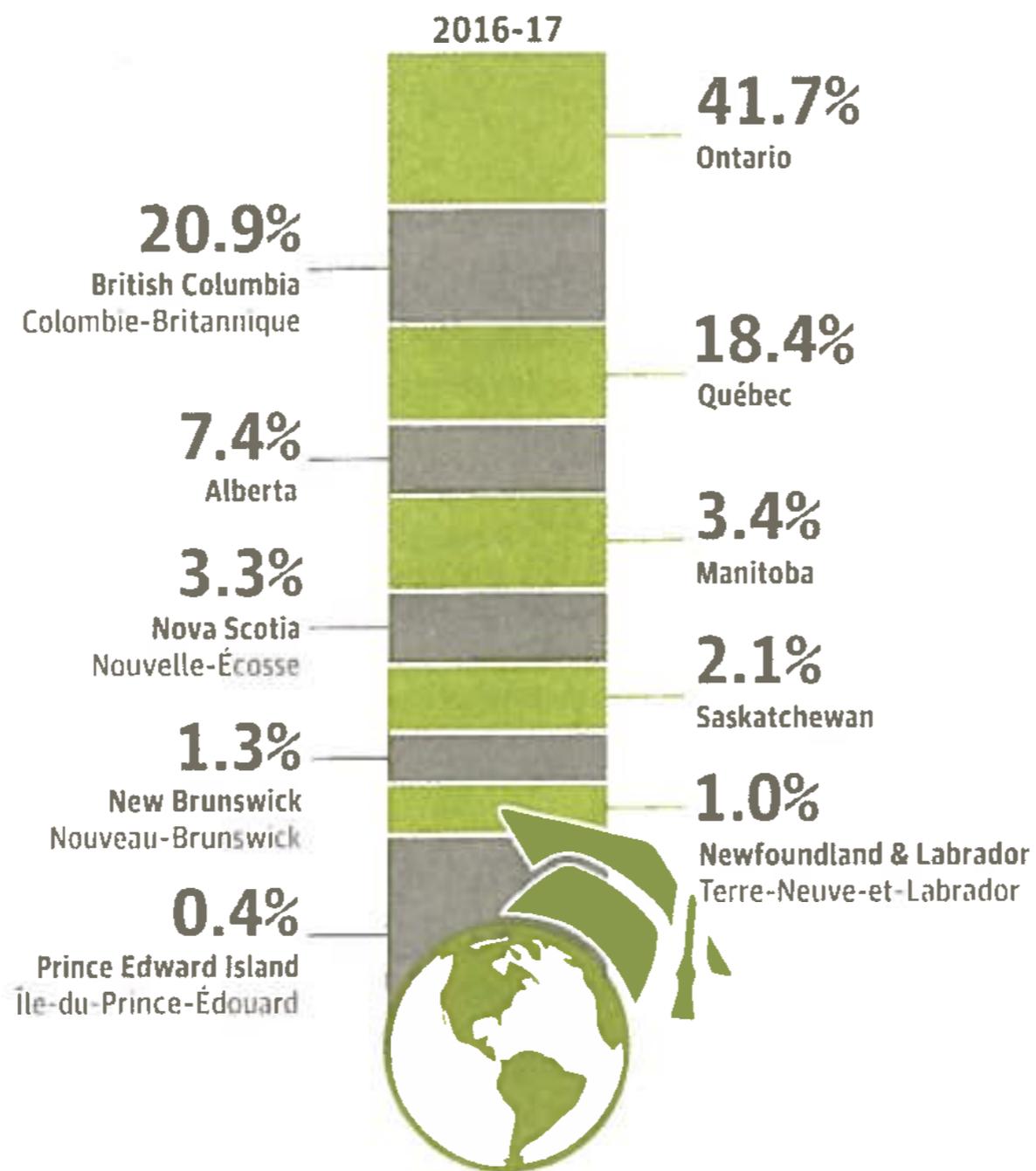■ Canadian student enrolment / Nombre total d'étudiants canadiens

134%

12%

Post-secondary international enrolment as a share of total enrolment within each province in 2016-17
Effectifs et diplômés postsecondaires par rapport aux inscriptions totales dans chaque province en 2016-17

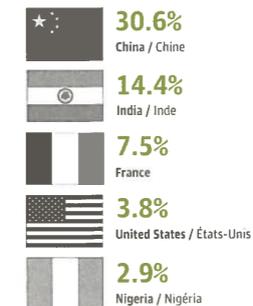**Newfoundland and Labrador / Terre-Neuve-et-Labrador**
9.8%

Prince Edward Island / Île-du-Prince-Édouard
15.6%

Nova Scotia / Nouvelle-Écosse
14.8%

New Brunswick / Nouveau-Brunswick
12.1%

Québec
8.5%

Ontario
12.6%

Manitoba
13.2%

Saskatchewan
9.4%

Alberta
9.6%

British Columbia / Colombie-Britannique
18.4%

Post-secondary international enrolment as a share of total enrolment within each province in 2016-17
Effectifs et diplômés postsecondaires par rapport aux inscriptions totales dans chaque province en 2016-17

Newfoundland and Labrador / Terre-Neuve-et-Labrador
9.8%

Prince Edward Island / Île-du-Prince-Édouard
15.6%

Nova Scotia / Nouvelle-Écosse
14.8%

New Brunswick / Nouveau-Brunswick
12.1%

Québec
8.5%

Ontario
12.6%

Manitoba
13.2%

Saskatchewan
9.4%

Alberta
9.6%

British Columbia / Colombie-Britannique
18.4%
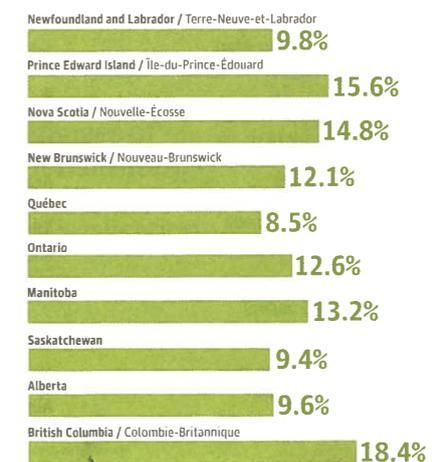
Source: Statistics Canada, PSIS / Statistique Canada, EDPC

Distribution of post-secondary international student enrolment across Canada
Répartition des étudiants internationaux à travers le pays

2016-17

41.7%
Ontario

20.9%
British Columbia
Colombie-Britannique

18.4%
Québec

7.4%
Alberta

3.4%
Manitoba

3.3%
Nova Scotia
Nouvelle-Écosse

2.1%
Saskatchewan

1.3%
New Brunswick
Nouveau-Brunswick

1.0%
Newfoundland & Labrador
Terre-Neuve-et-Labrador

0.4%
Prince Edward Island
Île-du-Prince-Édouard

The problem with all these visualizations is that the *scales* are wrong. The proportions of the SVG elements do not match the numbers.

D3 has a number of built-in scales that help us avoid this situation. They are easy to use and help us tell more truthful stories with our dataviz.

Scales

TERMS

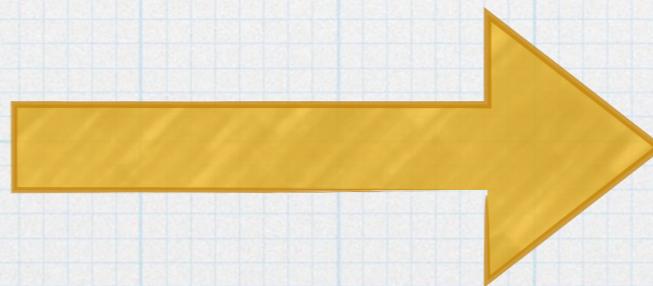A *scale* is a way that we can map or correlate some input (called a *domain*) to some output (called a *range*).

domain
values we plug in (input)

range
values we get out (output)

amount of food in my refrigerator → days until I have to go to the grocery store

TERMS

A *scale* is a way that we can map or correlate some input (called a *domain*) to some output (called a *range*).
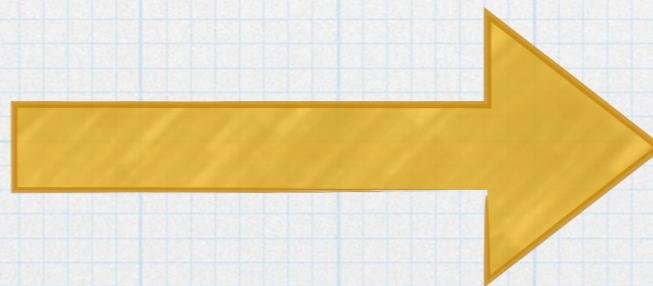
domain
values we plug
in (input)

range
values we get
out (output)

amount of
time I have to
study

⟹

my test scores
[0% to 100%]

TERMS

A *scale* is a way that we can map or correlate some input (called a *domain*) to some output (called a *range*).
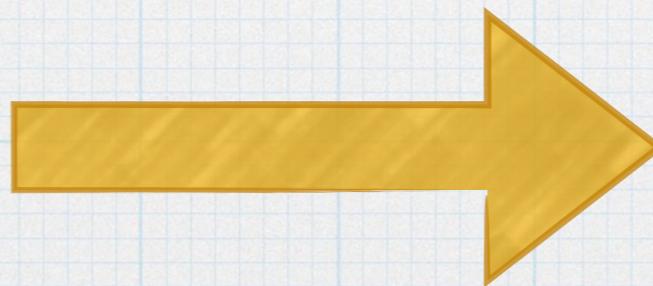
domain
values we plug
in (input)

range
values we get
out (output)

how much gas
is in my car

⟶

how far I can
drive

TERMS

A *scale* is a way that we can map or correlate some input (called a *domain*) to some output (called a *range*).
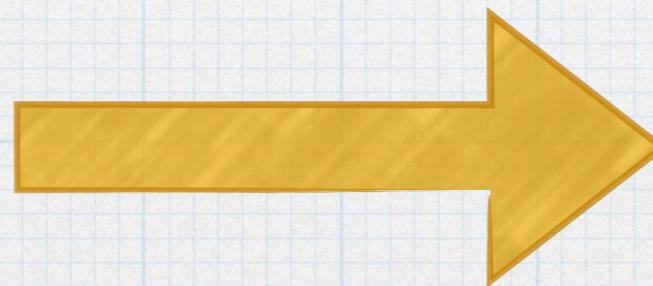
domain

how much gas
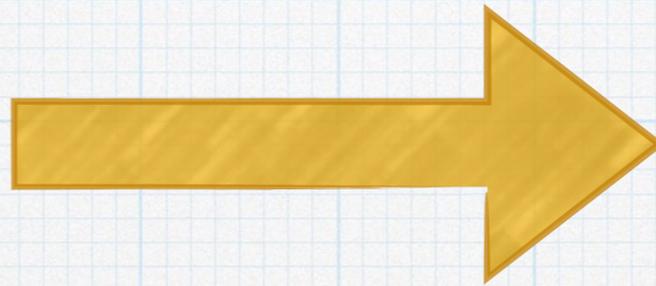is in my car

[0 to 50 liters]

range

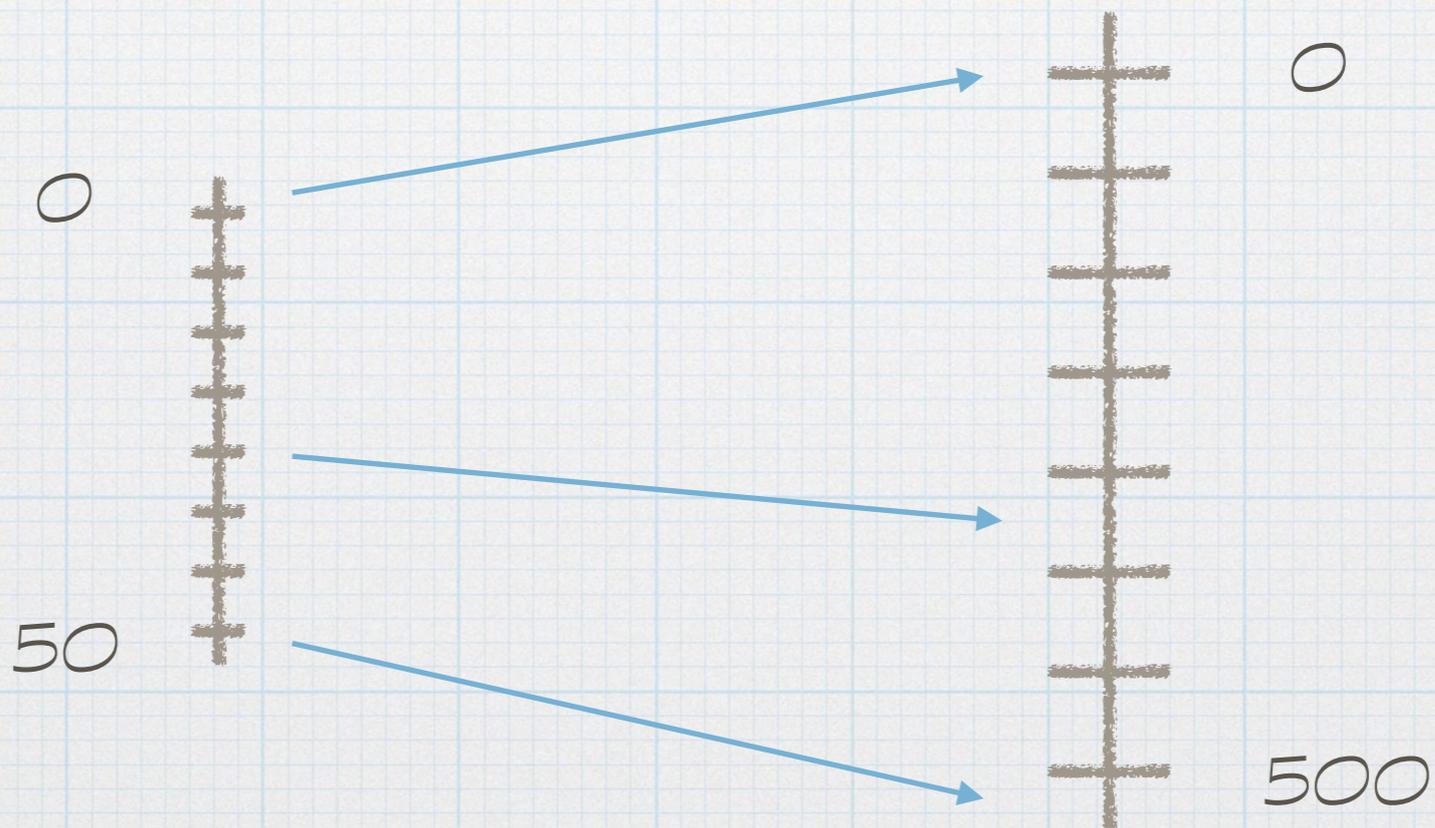how far I can
drive

[0 to 500 km]

TERMS

A *scale* is a way that we can map or correlate some input (called a *domain*) to some output (called a *range*).

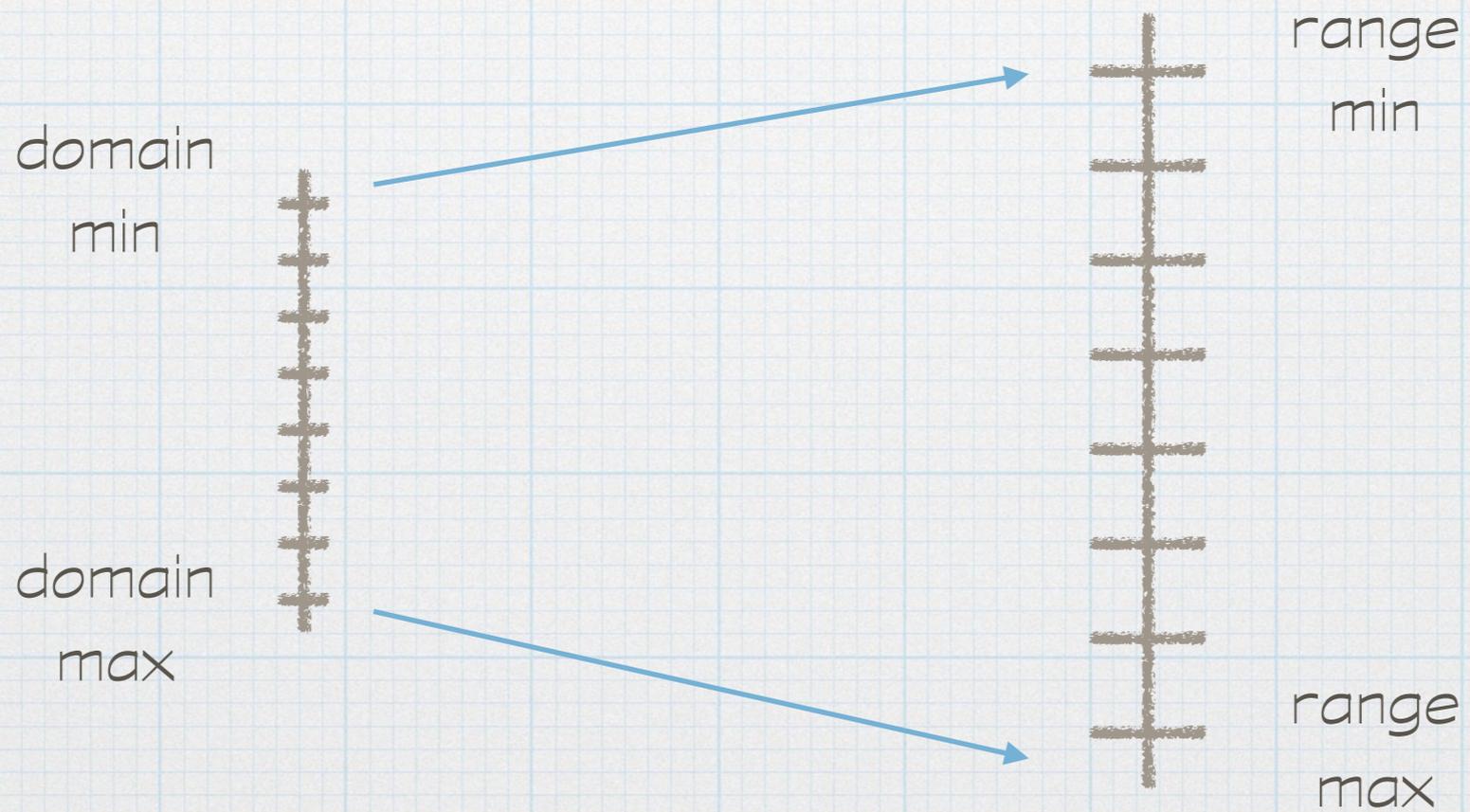[0 to 50 liters] → [0 to 500 km]

KINDS OF SCALES

scale	description
scaleLinear()	both domain and range are continuous
scaleSqrt()	square root (useful for the areas of circles!)
scalePow()	exponential curves
scaleLog()	logarithmic curves (Richter scale, decibels)
scaleQuantize()	putting continuous data into defined "buckets"
scaleQuantile()	calculates "buckets" based on equal-sized groups
scaleOrdinal()	old "buckets" to new "buckets"
scaleTime()	establishing timelines

SCALES

```
d3.scaleLinear()                    d3.scaleSqrt()
   .domain( [ , ] )                    .domain( [ , ] )
   .range( [ , ] );                    .range( [ , ] );
```

range
min

domain
min

domain
max

range
max

MINIMUM AND MAXIMUM

domain
min

d3.min(*‹dataset›* , *‹callback function›*)

d3.max(*‹dataset›* , *‹callback function›*)

domain
max

{score: -0.2988240538638829, retweets: 115687}

max_sentiment = d3.max(global_dataset, d => d.score)

max_retweets = d3.max(global_dataset, d => d.retweets)

MINIMUM AND MAXIMUM

domain
min

domain
max

d3.min(*‹dataset›* , *‹callback function›*)

d3.max(*‹dataset›* , *‹callback function›*)

```
{score: -0.2988240538638829, retweets: 115687}
```

```
min_sentiment = d3.min(global_dataset, d => d.score)
```

```
min_retweets = d3.min(global_dataset, d => d.retweets)
```

Example: Sentiment Analysis Scores

SENTIMENT ANALYSIS

Sentiment analysis is an algorithmic calculation that deciphers a text's attitude or tone: it's either positive, negative, or neutral toward its subject. Sentiment scores range from totally negative (-1) through neutral (0) to totally positive (+1).

This was the greatest movie I've ever seen! ⟶ +1
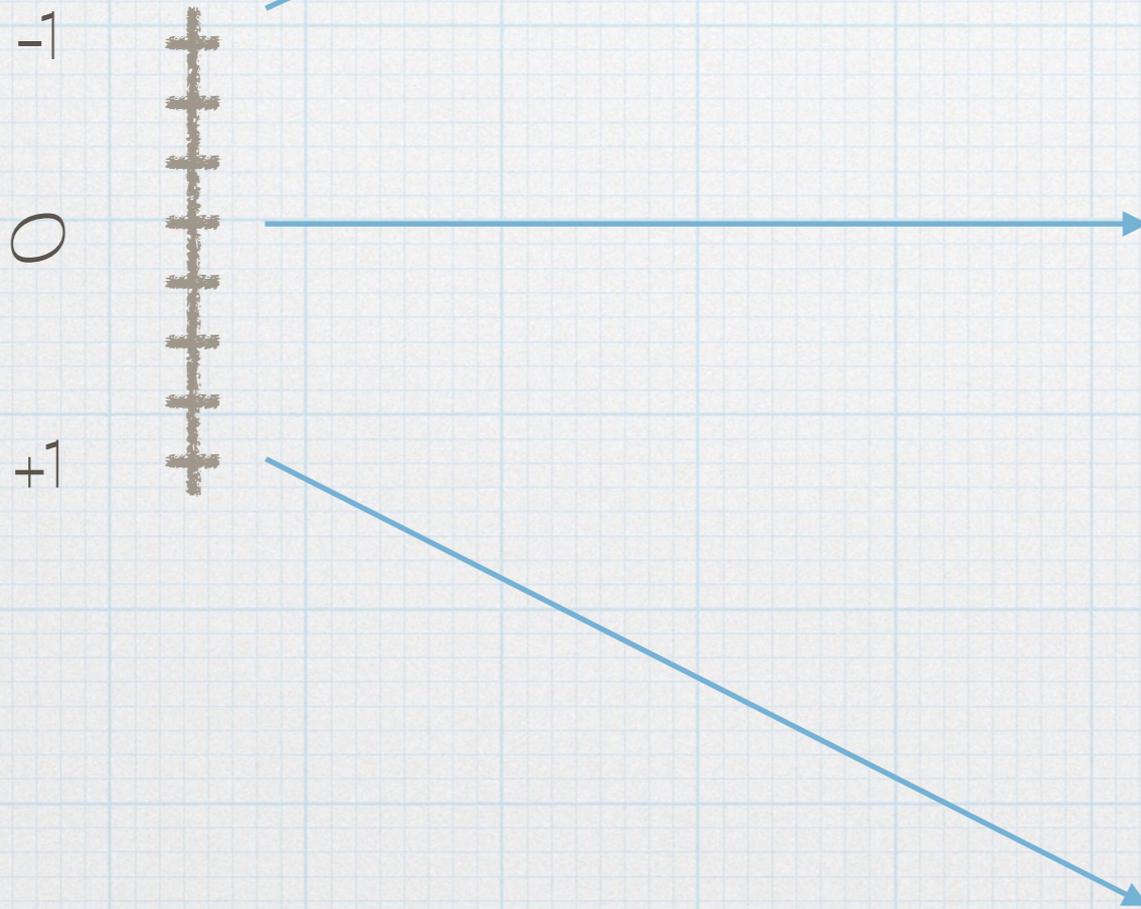
It was OK. Nothing special, but entertaining. ⟶ 0
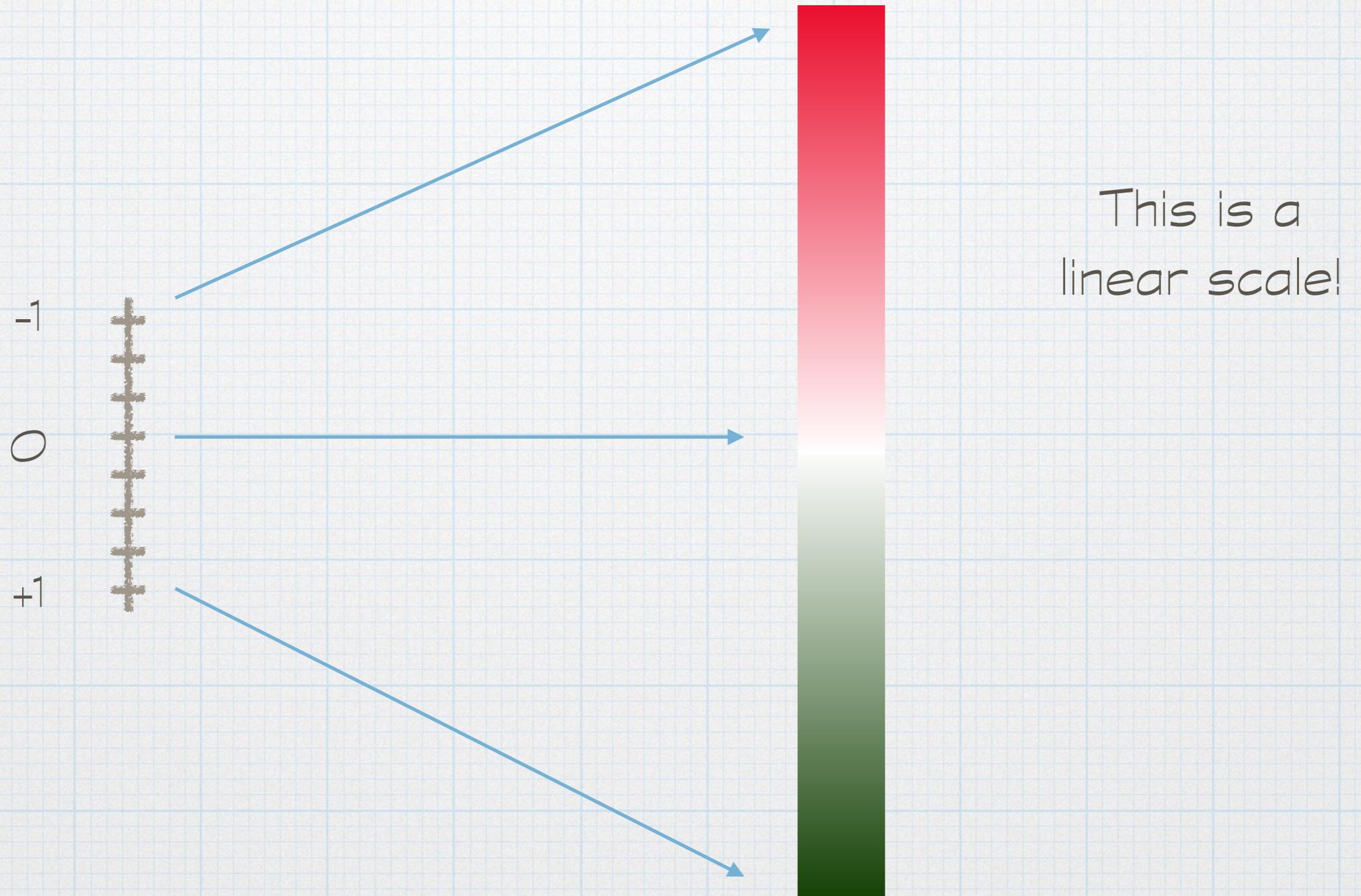
Worst piece of garbage ever. DO NOT WATCH! ⟶ -1

SENTIMENT ANALYSIS SCALE

domain

range of
some kind

-1

0

+1

SENTIMENT ANALYSIS SCALE

-1

0

+1

This is a
linear scale!

SENTIMENT ANALYSIS SCALE

Code time!

Special Problems: Area and Circles

PROBLEMS WITH AREA

Human beings are notoriously bad at choosing which images have bigger or smaller areas. Here, for example, it's difficult to tell that each box covers exactly the same amount of area.
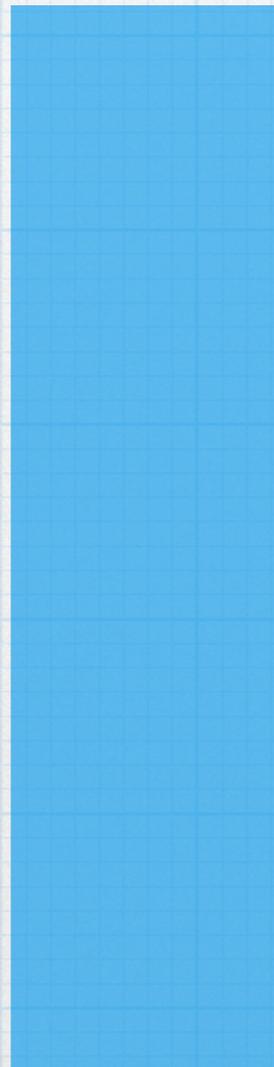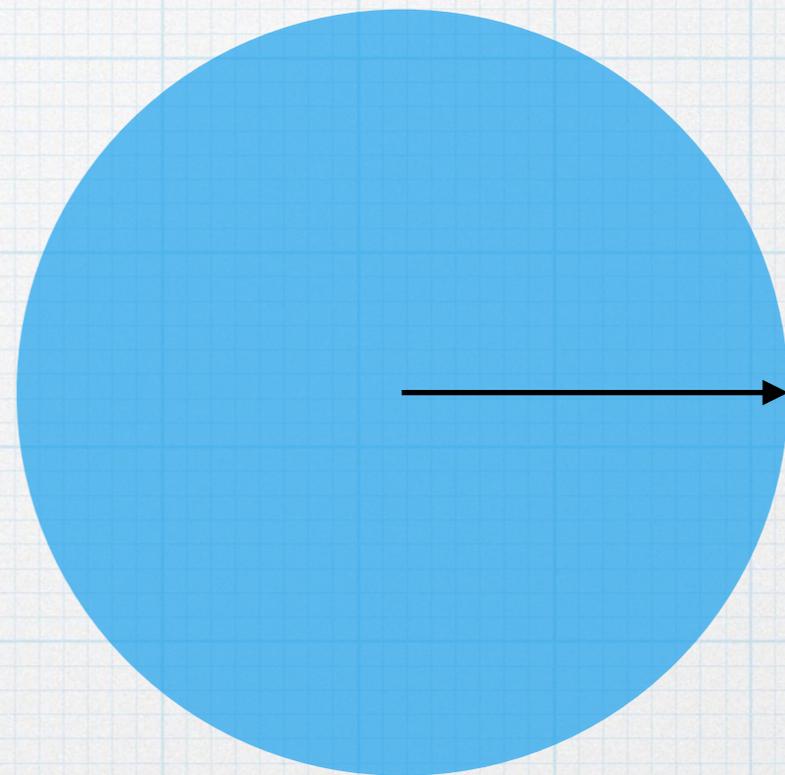
CIRCLES AND AREA

Still, area is often useful: circles especially. But circles need a bit of special attention from us. For example, let's imagine we want to create a circle twice as big as the one on the left. It's tempting to just double the radius.
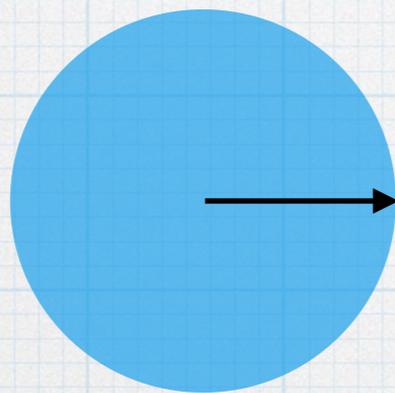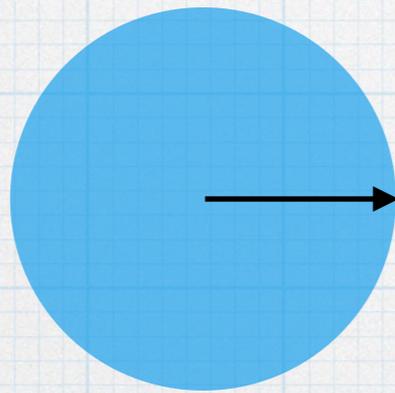
radius = 10

radius = 20

Area of Circle =
$\pi * r^2$

Circles and Area

Still, area is often useful: circles especially. But circles need a bit of special attention from us. For example, let's imagine we want to create a circle twice as big as the one on the left. It's tempting to just double the radius.
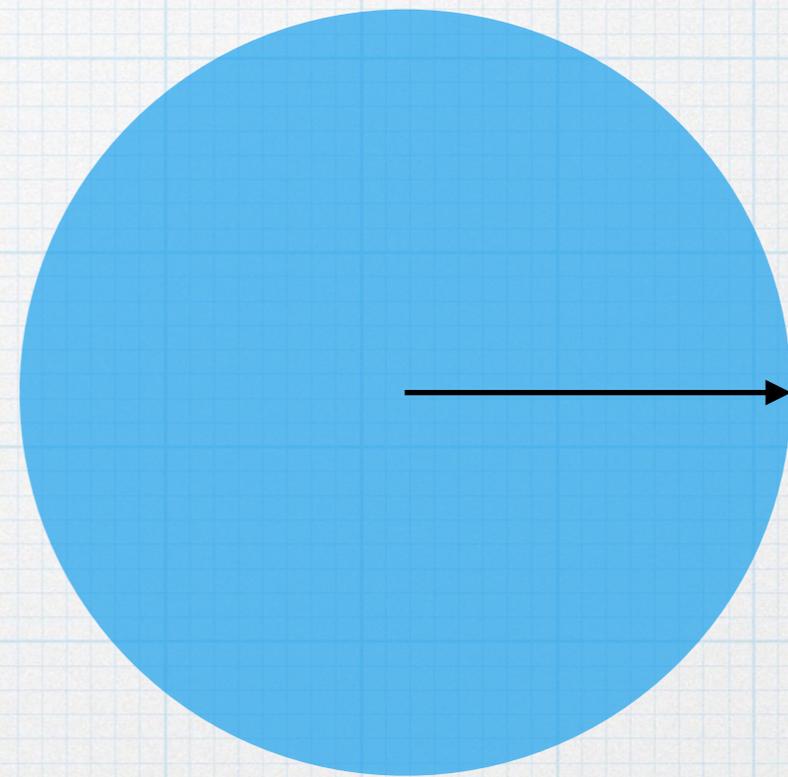
radius = 10
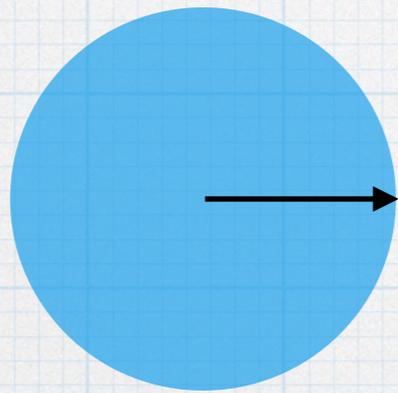
$\pi * r^2$

3.14 * 10 * 10

314 units

radius = 20

$\pi * r^2$

3.14 * 20 * 20

1256 units

CIRCLES AND AREA

Still, area is often useful: circles especially. But circles need a bit of special attention from us. For example, let's imagine we want to create a circle twice as big as the one on the left. It's tempting to just double the radius.
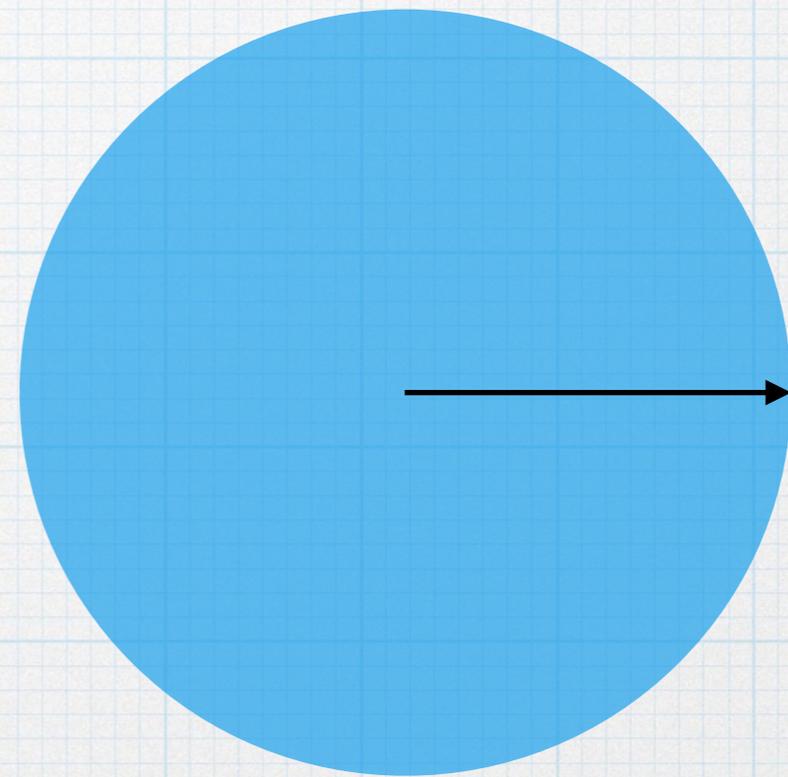
Doubling the radius makes a circle 4 times bigger! Tripling the radius would make it 9 times bigger!

radius = 10

$$\pi * r^2$$
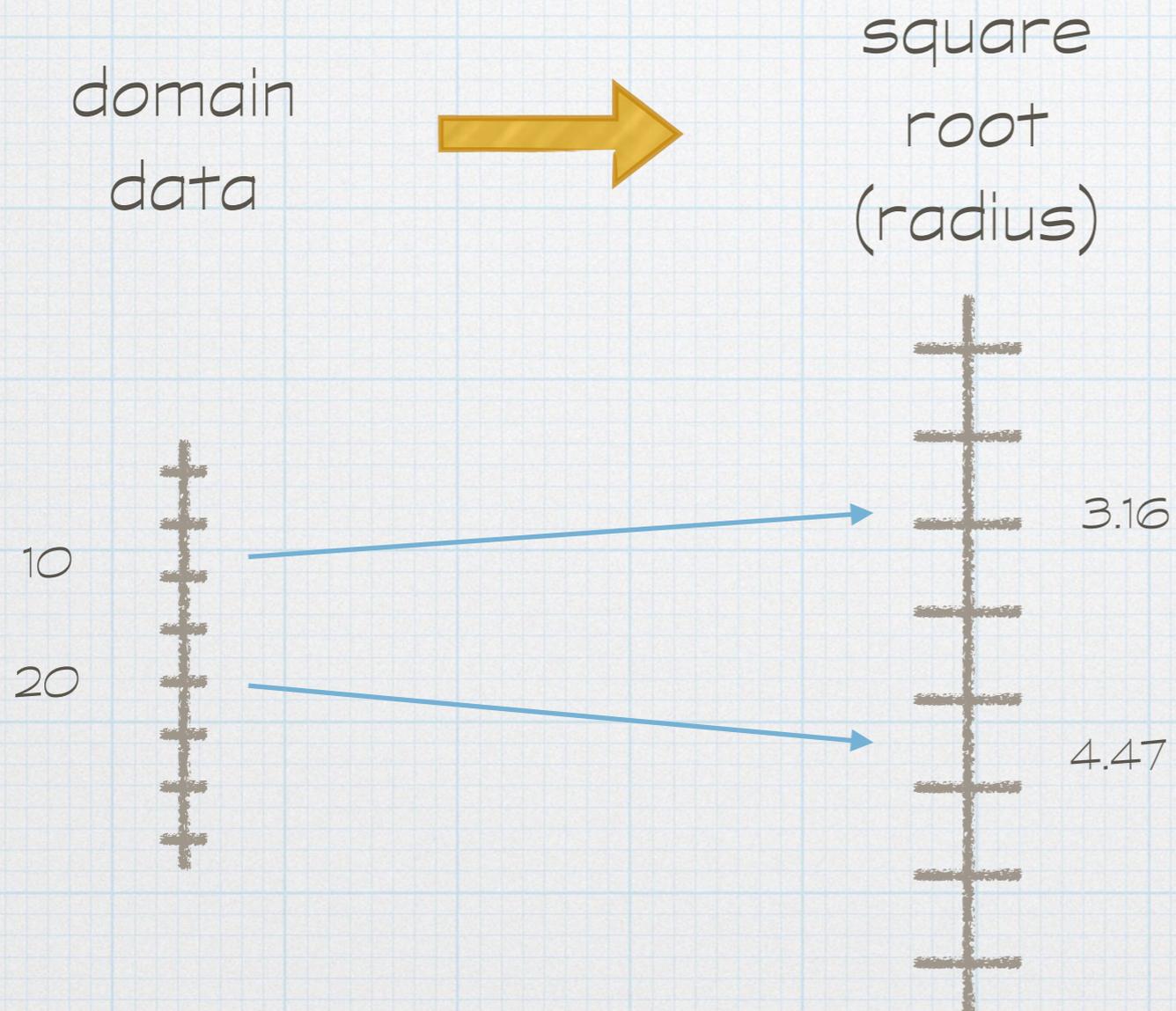
$$3.14 * 10 * 10$$

314 units

radius = 20

$$\pi * r^2$$
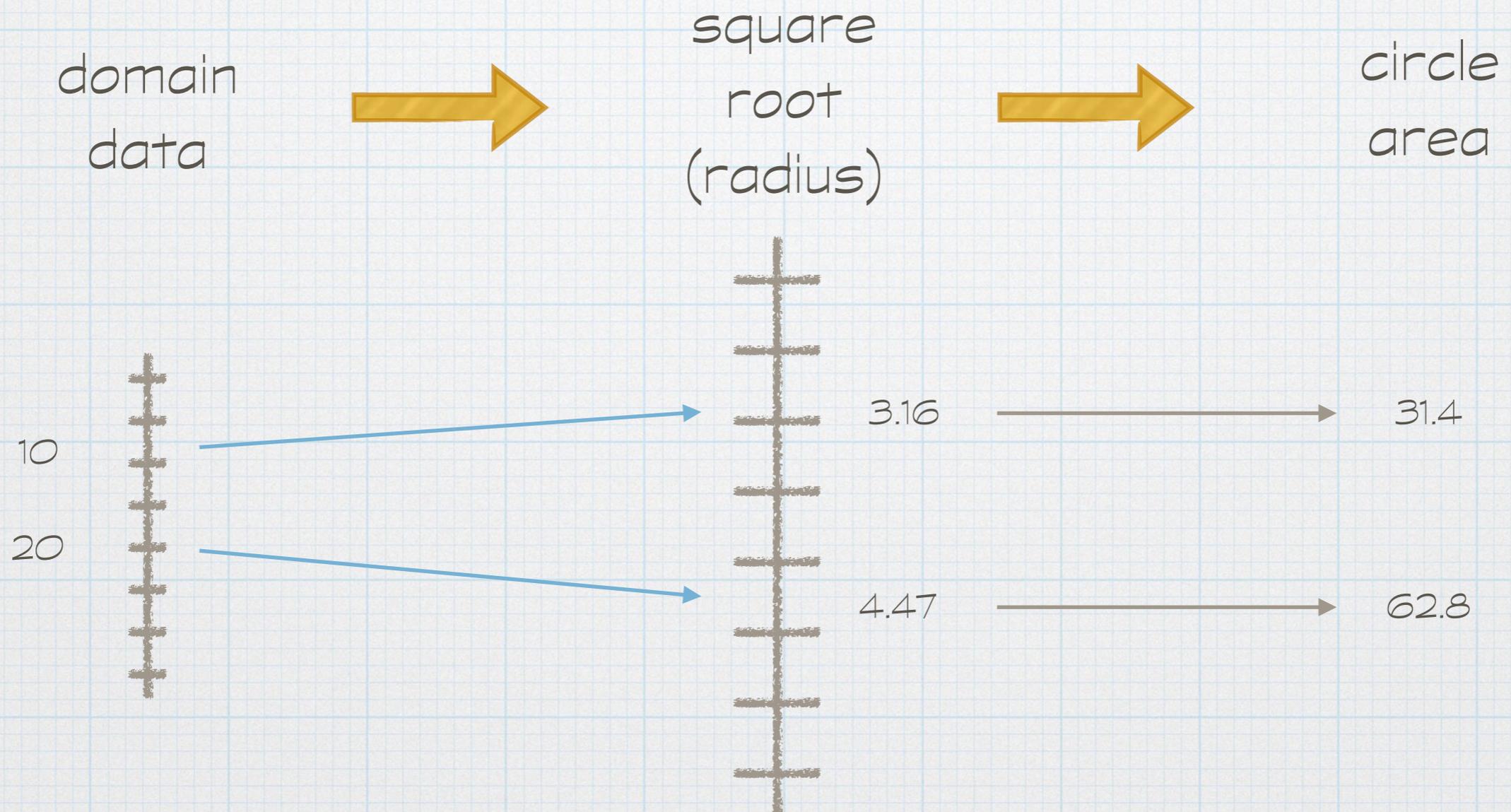
$$3.14 * 20 * 20$$

1256 units

SQUARE ROOT SCALE

The solution is to use a scale that takes the square root of our data. That way, a domain point that's twice as big will create a circle that's also twice as big.

domain
data

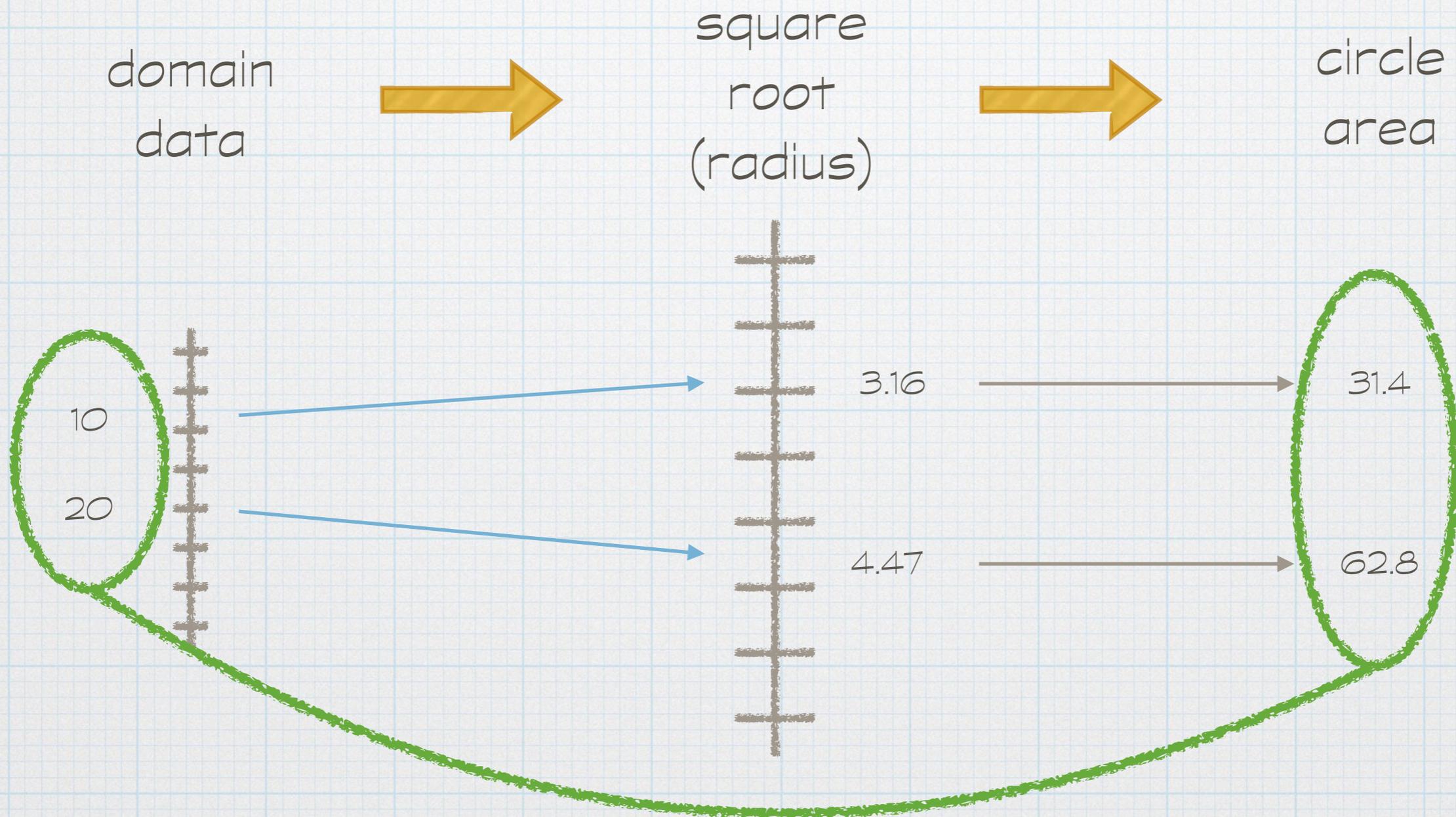square
root
(radius)

10

20

3.16

4.47

SQUARE ROOT SCALE

The solution is to use a scale that takes the square root of our data. That way, a domain point that's twice as big will also create a circle that's twice as big.

domain
data

square
root
(radius)

circle
area

10

20

3.16

4.47

31.4

62.8

SQUARE ROOT SCALE

The solution is to use a scale that takes the square root of our data. That way, a domain point that's twice as big will also create a circle that's twice as big.

SQUARE ROOT SCALE

And I can have D3 "constrain" the range as well. So not only will it take the square root, but it will keep all the conversions within a reasonable range, so I can guarantee that all my circles will still fit inside my SVG container.
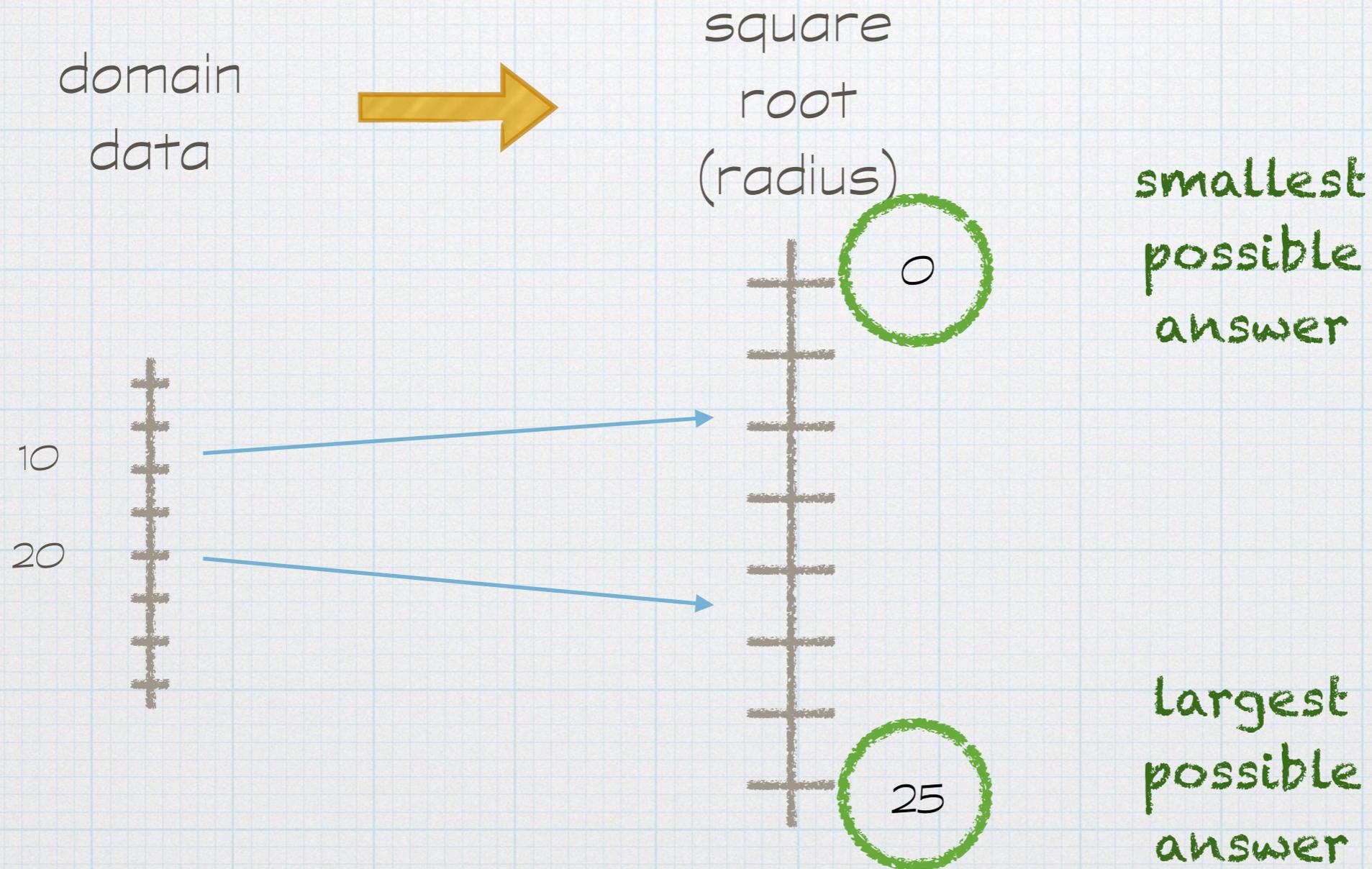
SQUARE ROOT SCALE

Code time!

D3 Scales

daraviz rant

linear scales

square root scales