

RStudio analogie



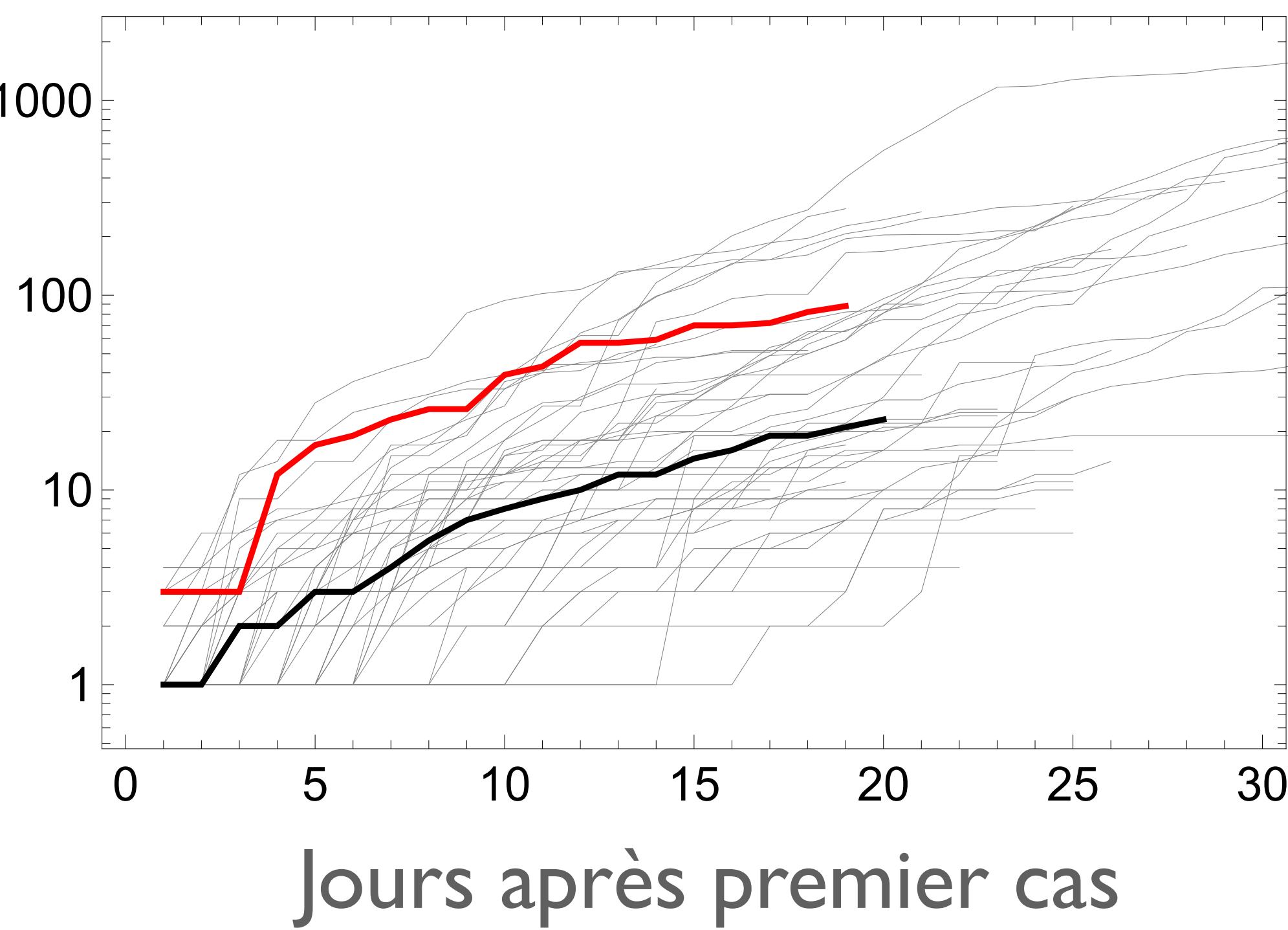
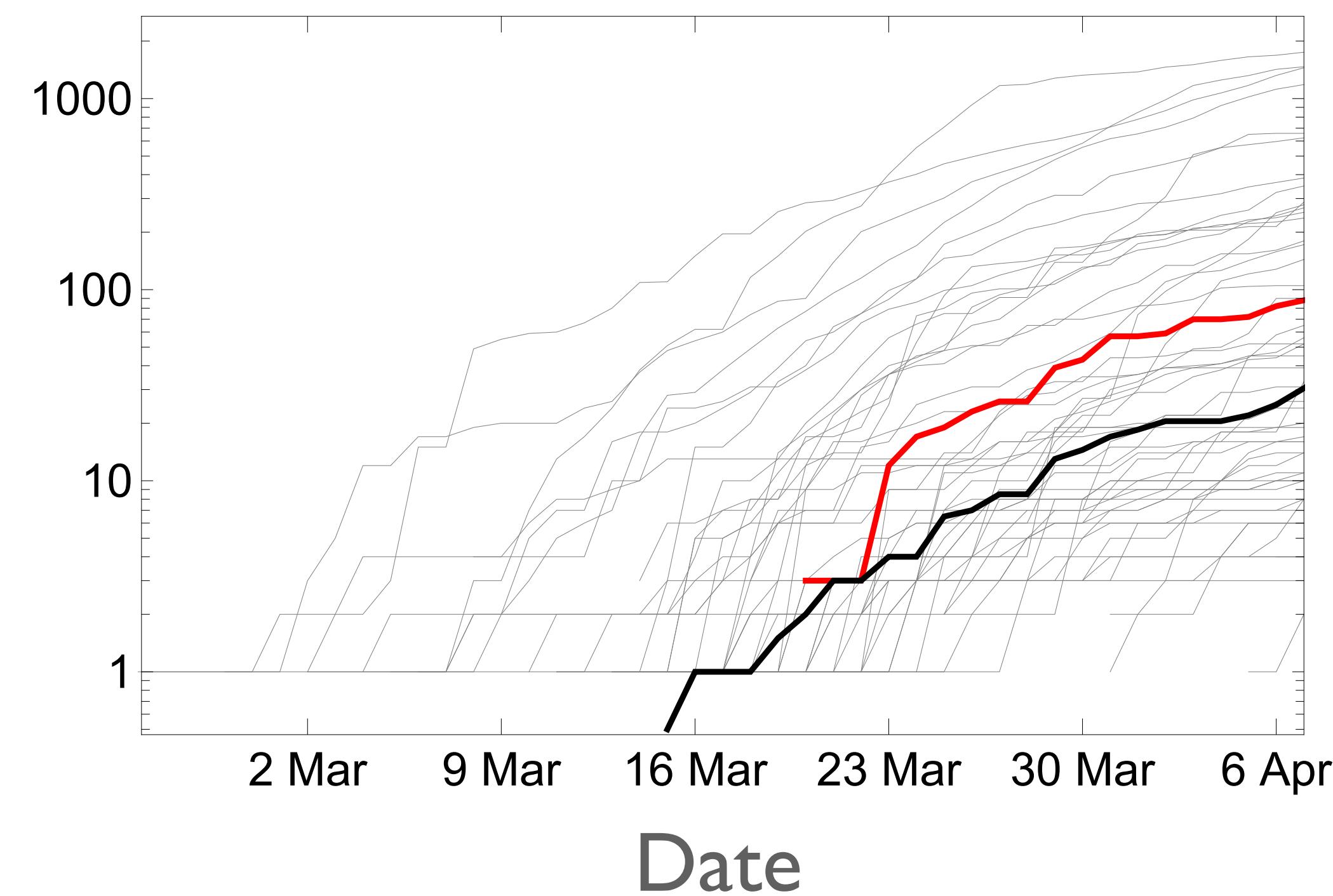
Visualization scientifique avec R et ggplot

Tanjona Ramiadantsoa

Un conseil: avant de faire une figure sur R, essayer de faire une caricature

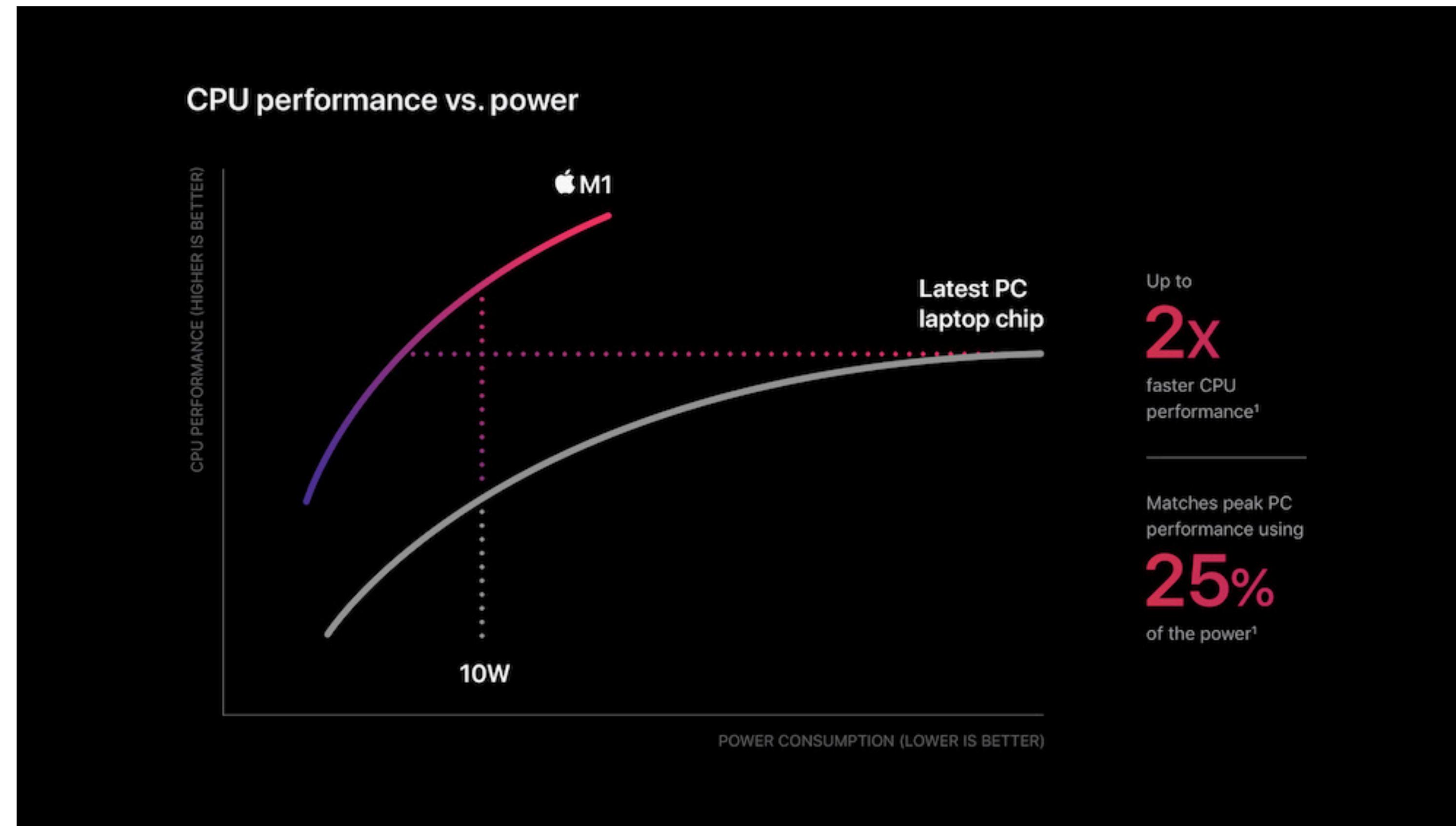
Nombre de cas cumulés

— Madagascar — Median — Other African countries

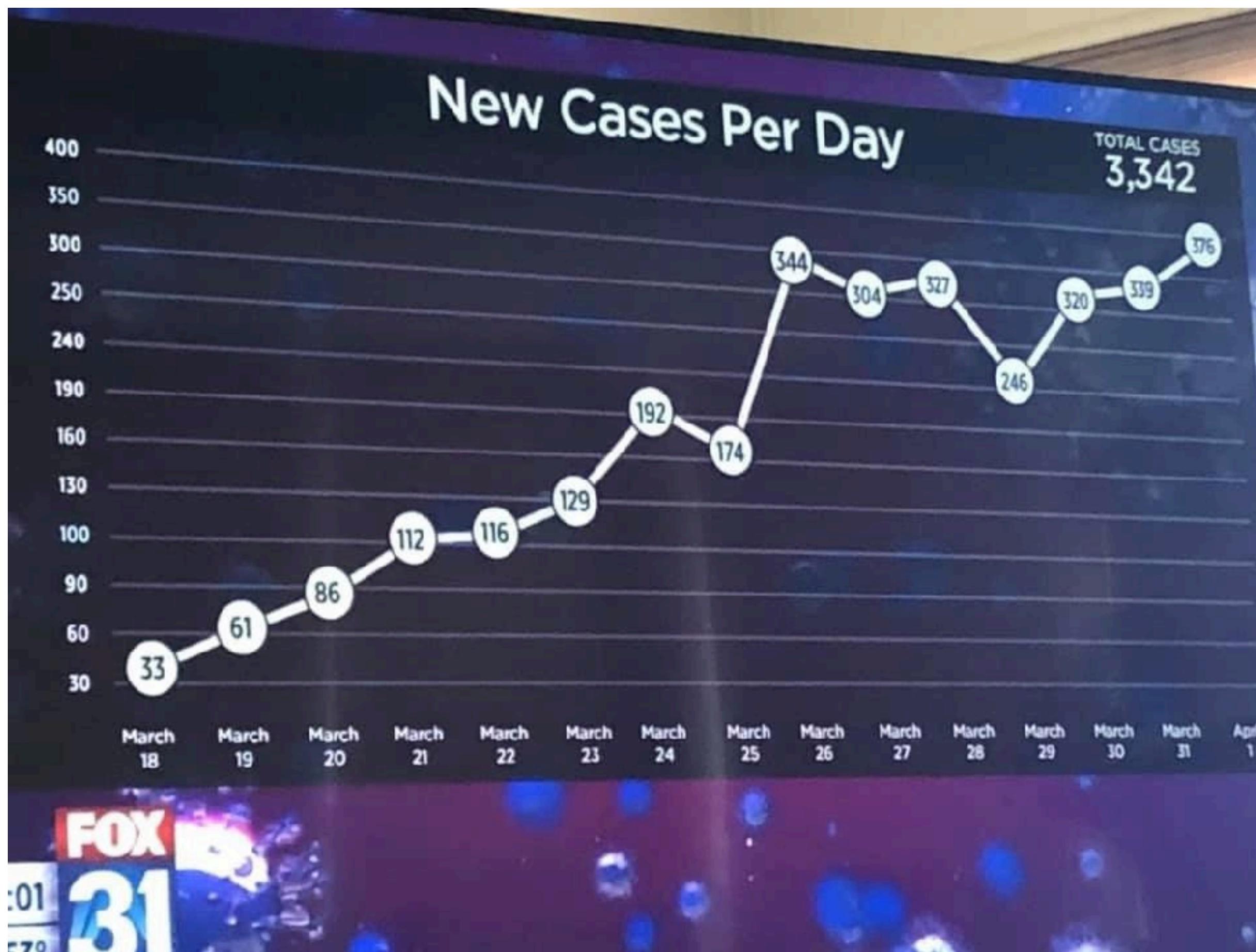


Une Image Vaut Mille Mots

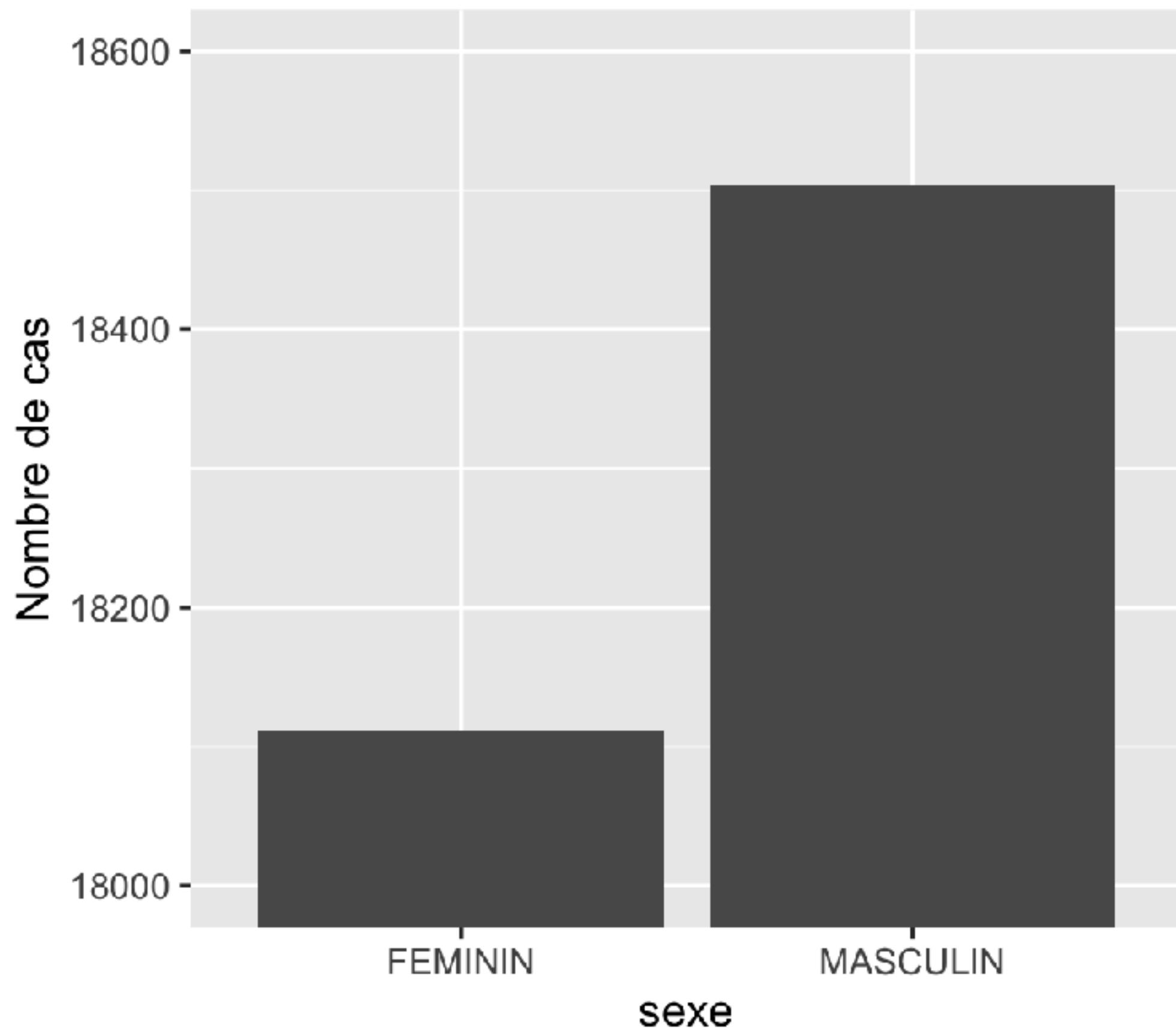
Péché capital: axes sans valeurs et graduation



Péché capital: incompétence ou manipulation?



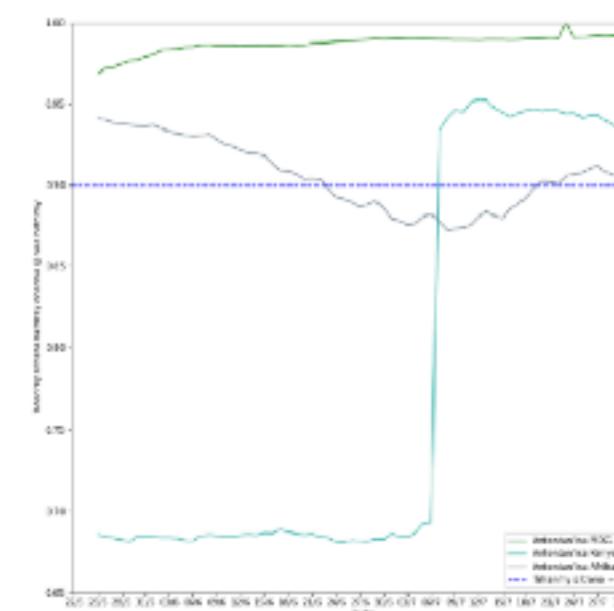
Péché capital: supprimer l'origine



Péché mortel: axes illisibles

Naoty fanadihadina hafa.

Covid-19 Mankaiza?



Tabilao ankapobeany sy modely ijerena ny fivoaran'ny Covid-19 ao Madagasikara.

[View the Project on GitHub](#)
tabilaocov/ady_cov.github.io

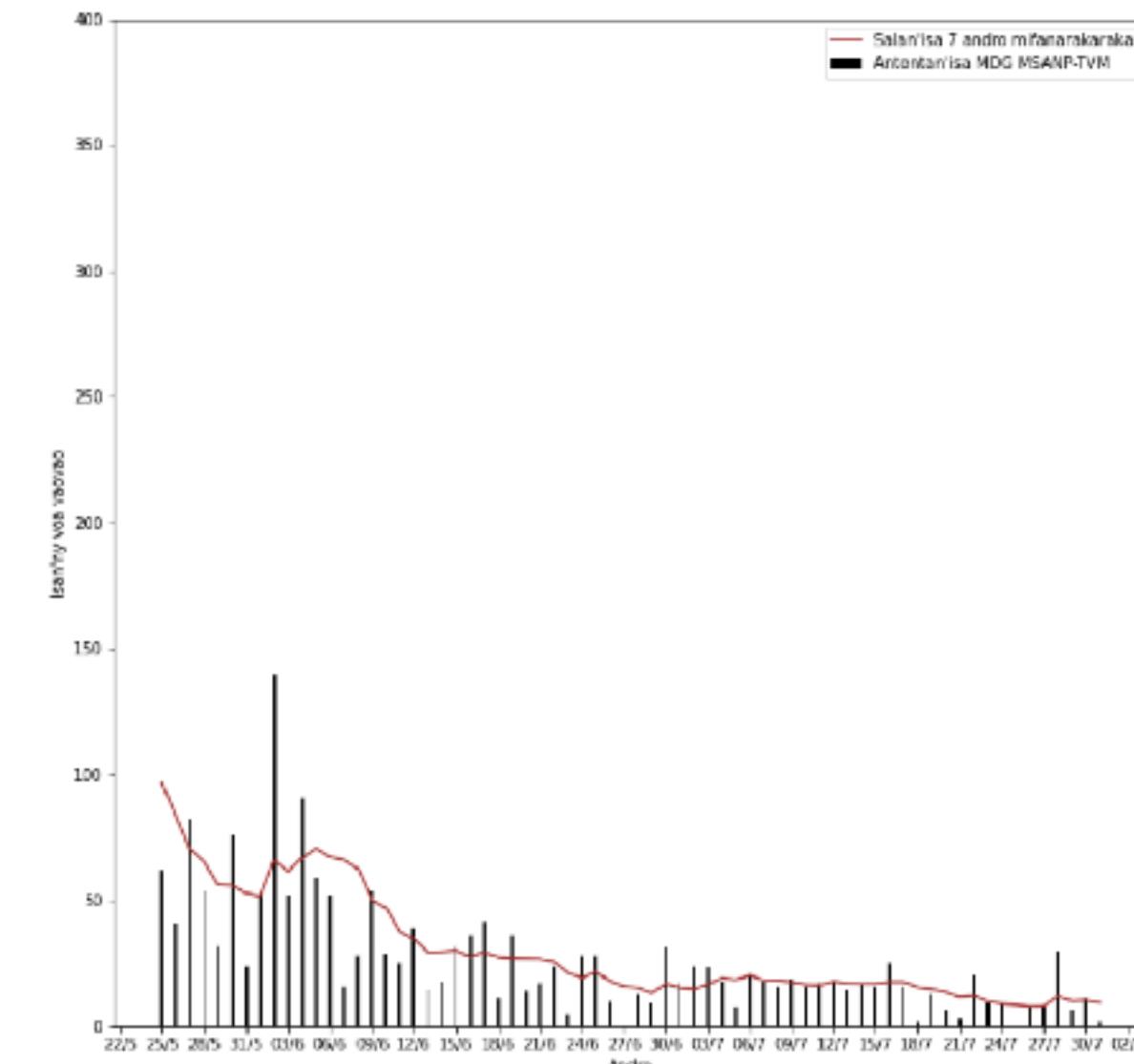
This project is maintained by [tabilaocov](#)

Hosted on GitHub Pages — Theme by [orderedlist](#)

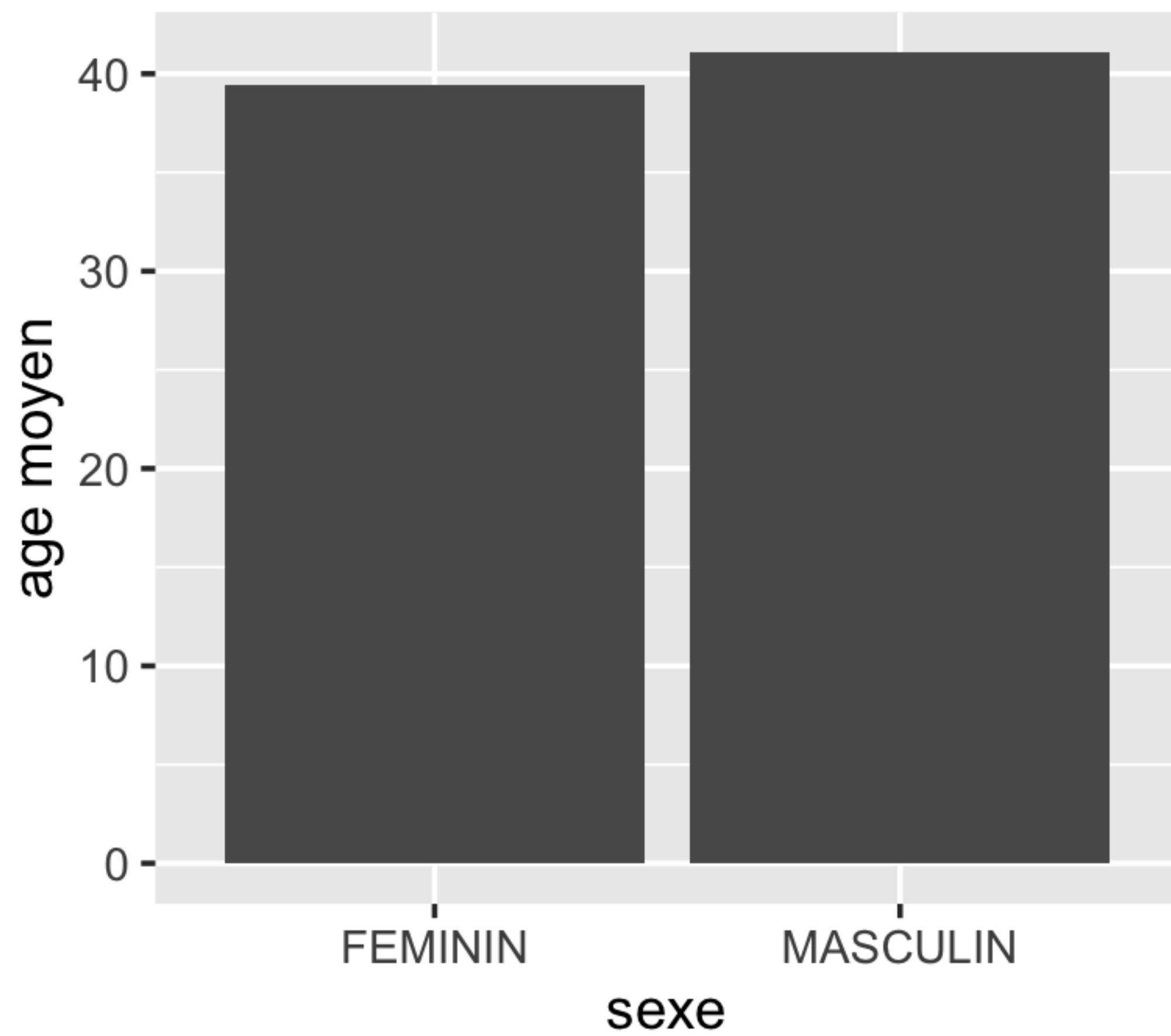
Tabilao ankapobeny sy modely ijerena ny fivoaran'ny Covid-19 ao Madagasikara (31/07/2021)

Fanadihadiana an-tsary hoan'Analamanga, Matsiatra Ambony sy Vakinakaratra.

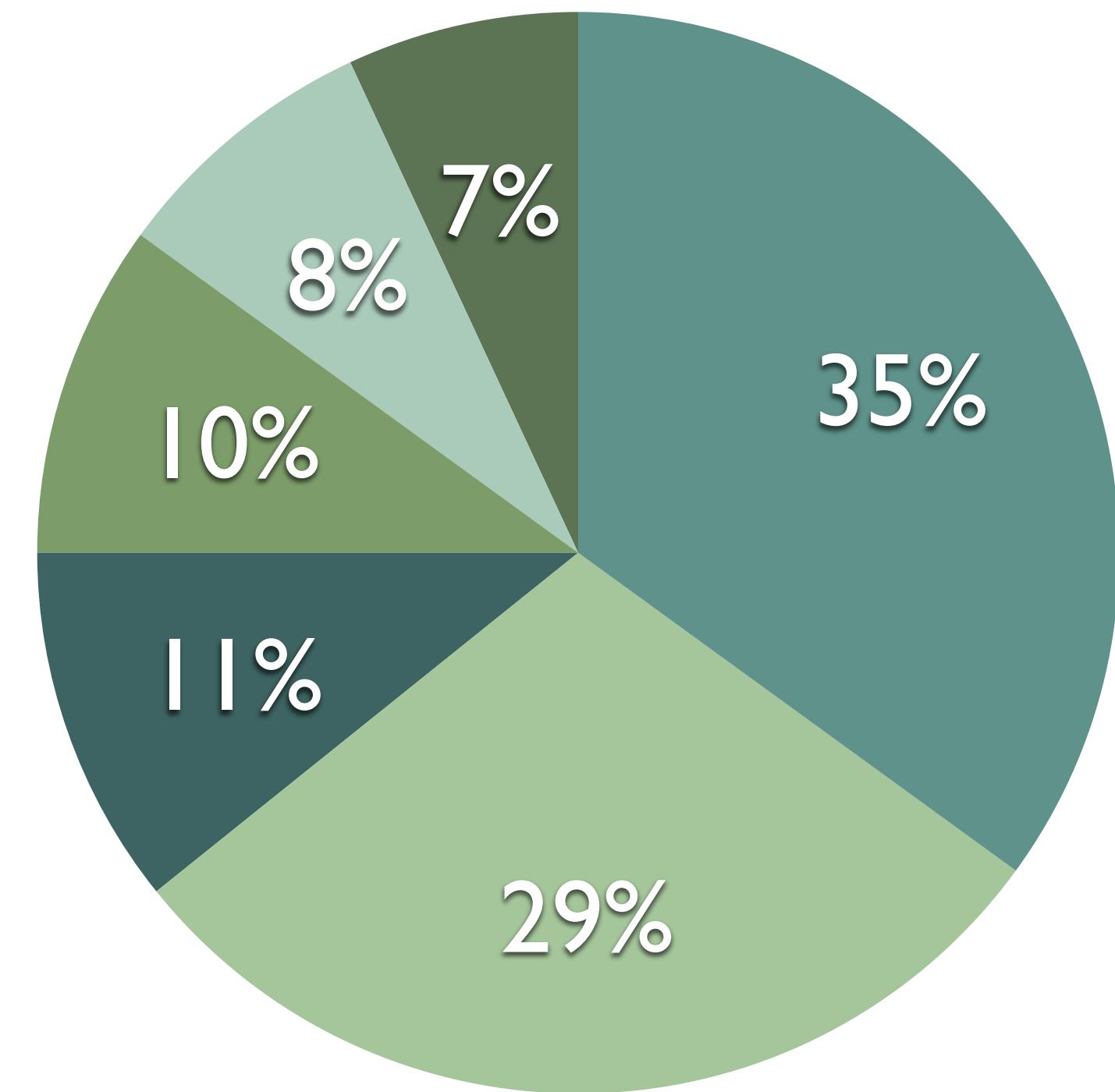
Antontanin'isa ao Madagasikara



Péché mortel: ignorer la distribution



Péché vénial: utiliser les pie charts



Péché vénial: sans serif vs serif

■ A faire

- Utiliser des polices sans serif telle que Arial, Helvetica, Calibri
- Pourquoi? Elles sont propres et faciles à lire

■ A ne pas faire

- Utiliser des polices avec serif telle que Times, Times New Roman, ou *des polices sophistiquées*
- Pourquoi? Elles sont chargées et peu lisibles

Péché vénial: capitalisation

- **A faire**
 - Utiliser les majuscules et minuscules
 - Soyez consistent avec les capitalisations
- **A NE PAS FAIRE**
 - EVITER DE METTRE TOUT EN MAJUSCULES
 - AU DELA DE L'IMPRESSION D'HURLEMENT, C'EST DIFFICILE A LIRE

Péché vénial: contraste

- **A faire**

- Choisir des couleurs qui un contraste fort
- **Vous pouvez voir ceci!**
- **Et aussi ceci**

- **A ne pas faire**

- Avoir des couleurs qui ne ressortent pas
- Est-ce que vous pouvez lire ce texte?
- Et aussi ceci

Péché vénial: rouge et vert



Normal



Red Color Blind



Green Color Blind

Il existe un choix innombrable de type de figures

e.g. <https://gramener.github.io/visual-vocabulary-vega/#>

Partage: Récept x | OpenABM-Covi x | r [CSBQ-QCBS x | Atelier 1: Intro x | DeepL Translate x | Computational x | Launch Meeting x | Visual Vocabulary x +

gramener.github.io/visual-vocabulary-vega/ ← → C Update :

Apps YouTube Maps News IRD_webmail Gather covid19_shiny BES Reading List

Visual Vocabulary Deviation Correlation Ranking Distribution Change-over-Time Magnitude Part-to-whole Spatial Flow FullList

Visual Vocabulary - Vega Edition



Inspired by Financial Times's Visual Vocabulary & Andy Kriebel's ft.

Click any section below to view the charts

There are so many ways to visualise data - how do we know which one to pick? Click on a category below to decide which data relationship is most important in your story, then look at the different types of charts within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative)

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other)

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Distribution

Show values in a dataset and how often they occur. The shape (or skew) of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data

Change-over-Time

Give emphasis to changing trends. These can be short (intraday) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead

Spatial

Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations

CREATED BY

Pratap Vardhan | @PratapVardhan

CREDITS

Vega Authors & Community, Gramener

<https://gramener.github.io/visual-vocabulary-vega/>

<https://github.com/gramener/visual-vocabulary-vega/>

INSPIRED BY

Andy Kriebel | @VizWizBI (design / data)

FT Graphics: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg-

Shannon; Paul McCallum; Martin Stabe

Visual Vocabulary: ft.com/vocabulary

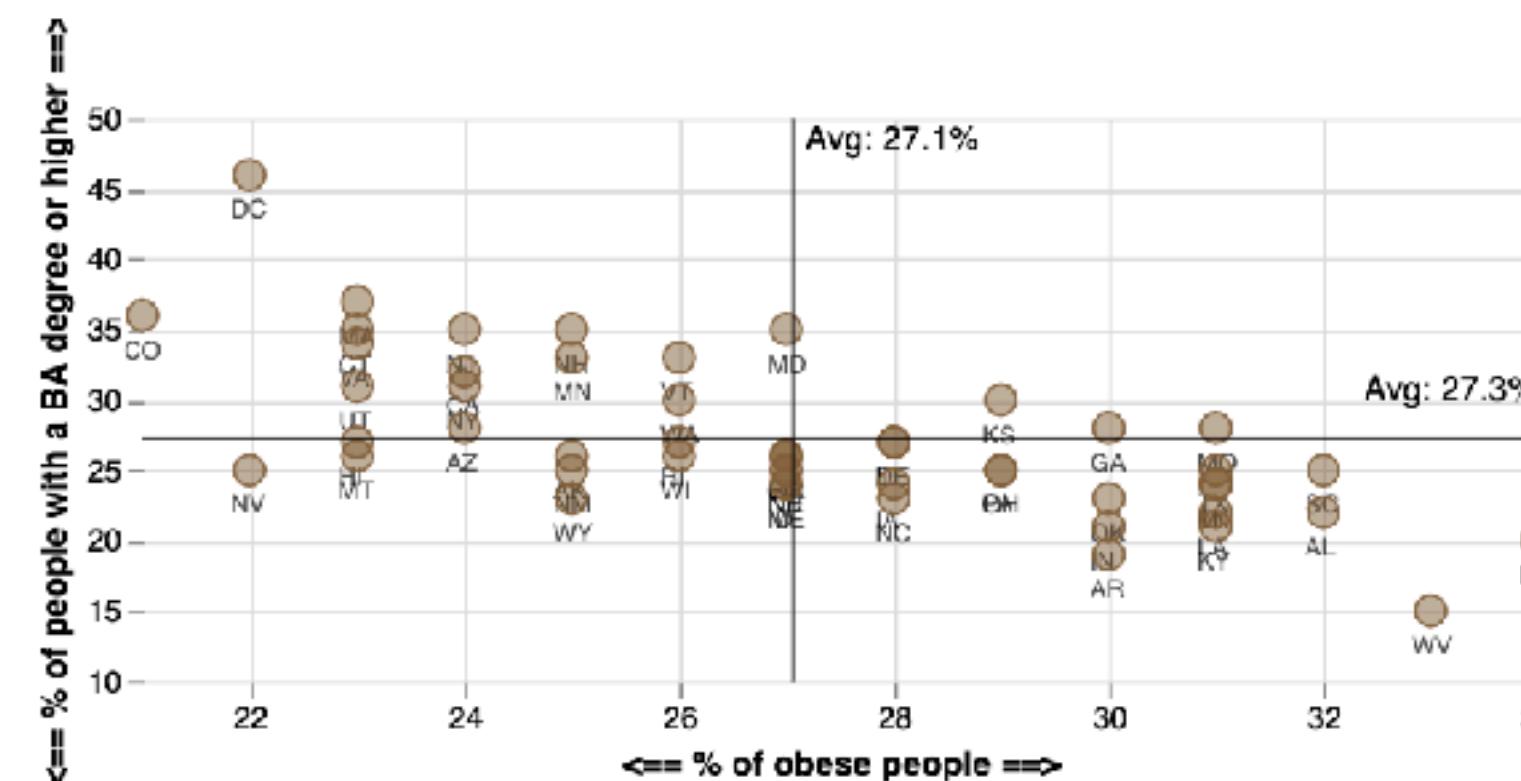


Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other)

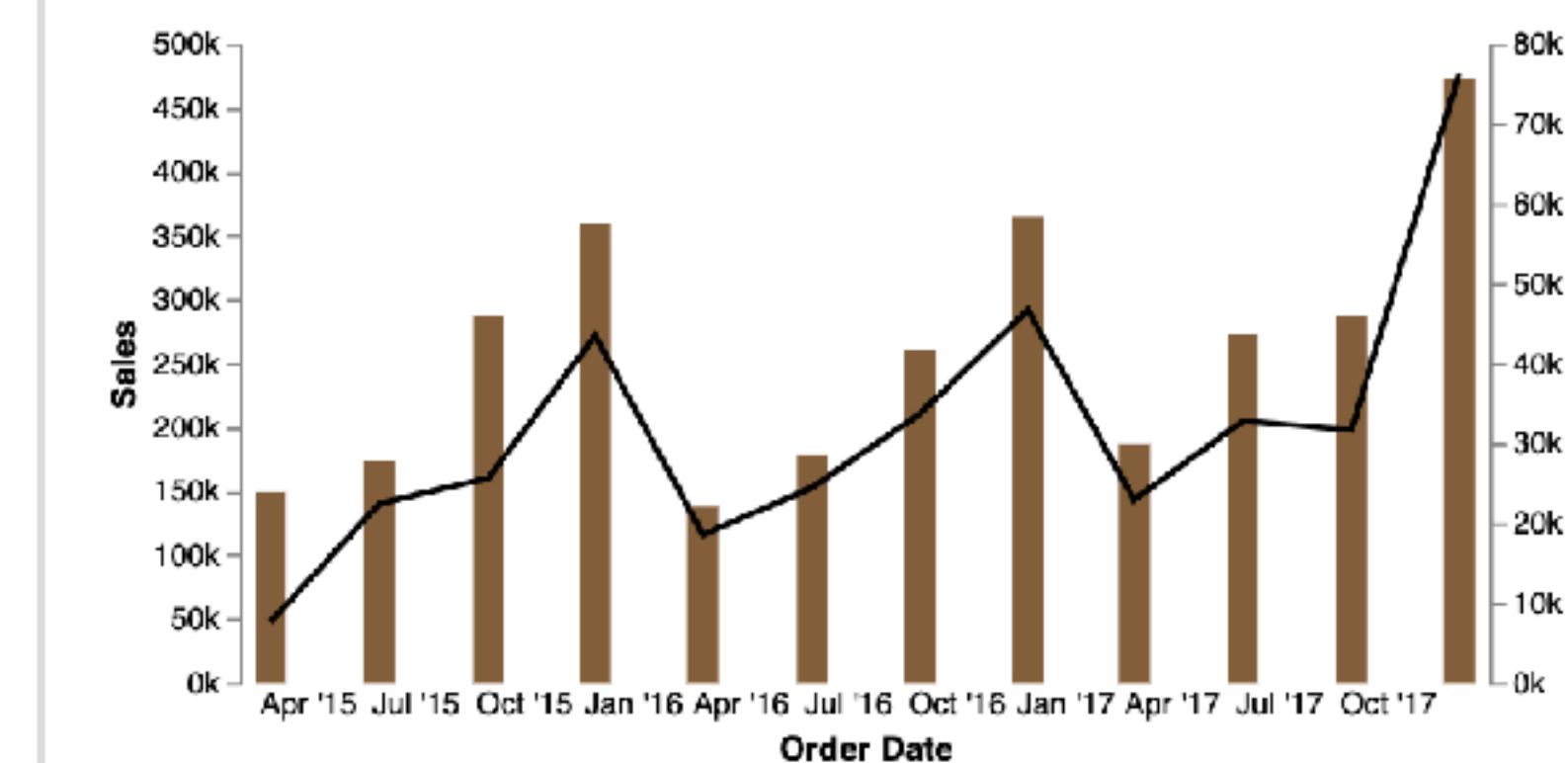
Scatterplot

The standard way to show the relationship between two variables, each of which has its own axis



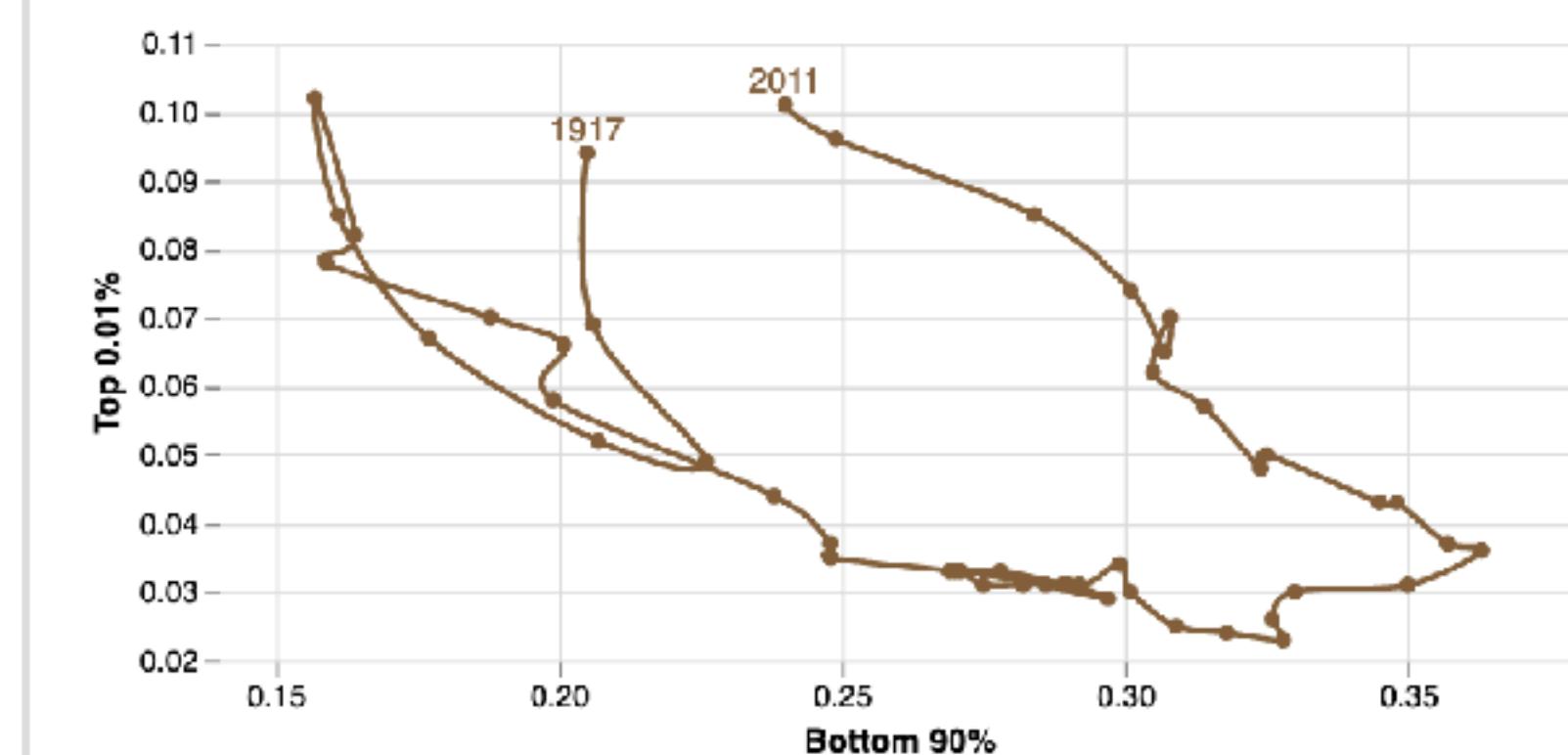
Column + line timeline

A good way of showing the relationship over time between an amount (columns) and a rate (line)



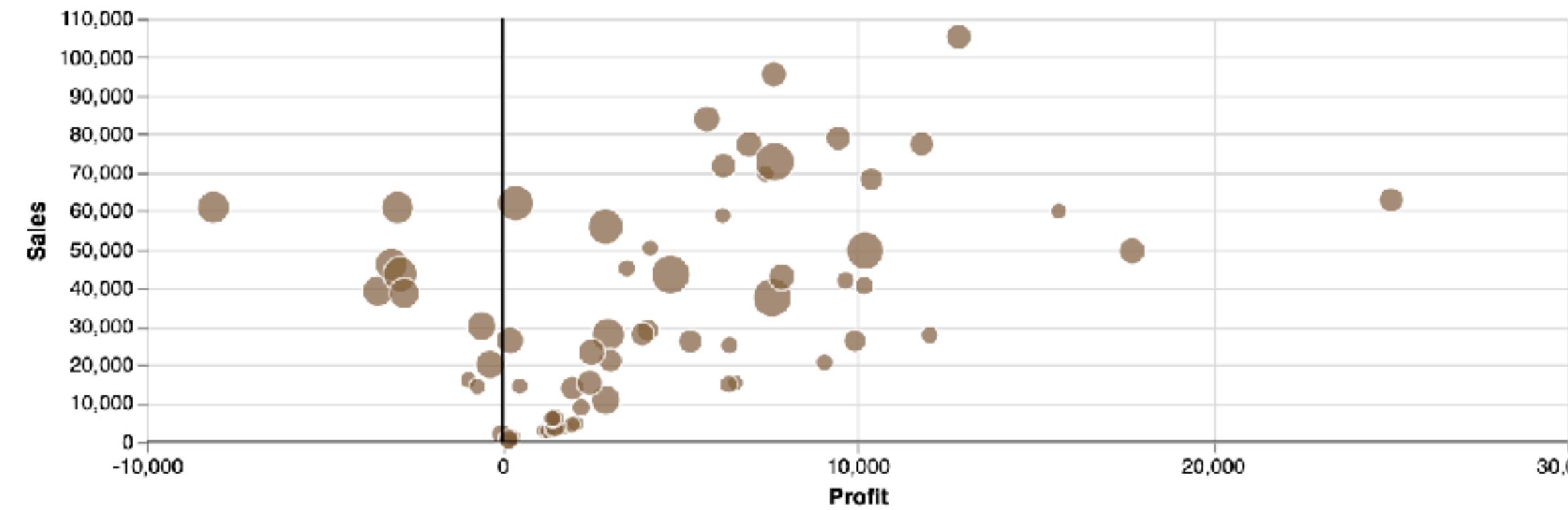
Connected scatterplot

Usually used to show how the relationship between 2 variables has changed over time



Bubble

Like a scatterplot, but adds additional detail by sizing the circles according to a third variable



Edit

X Y Heatmap

A good way of showing the patterns between 2 categories of data, less good at showing fine differences in amounts

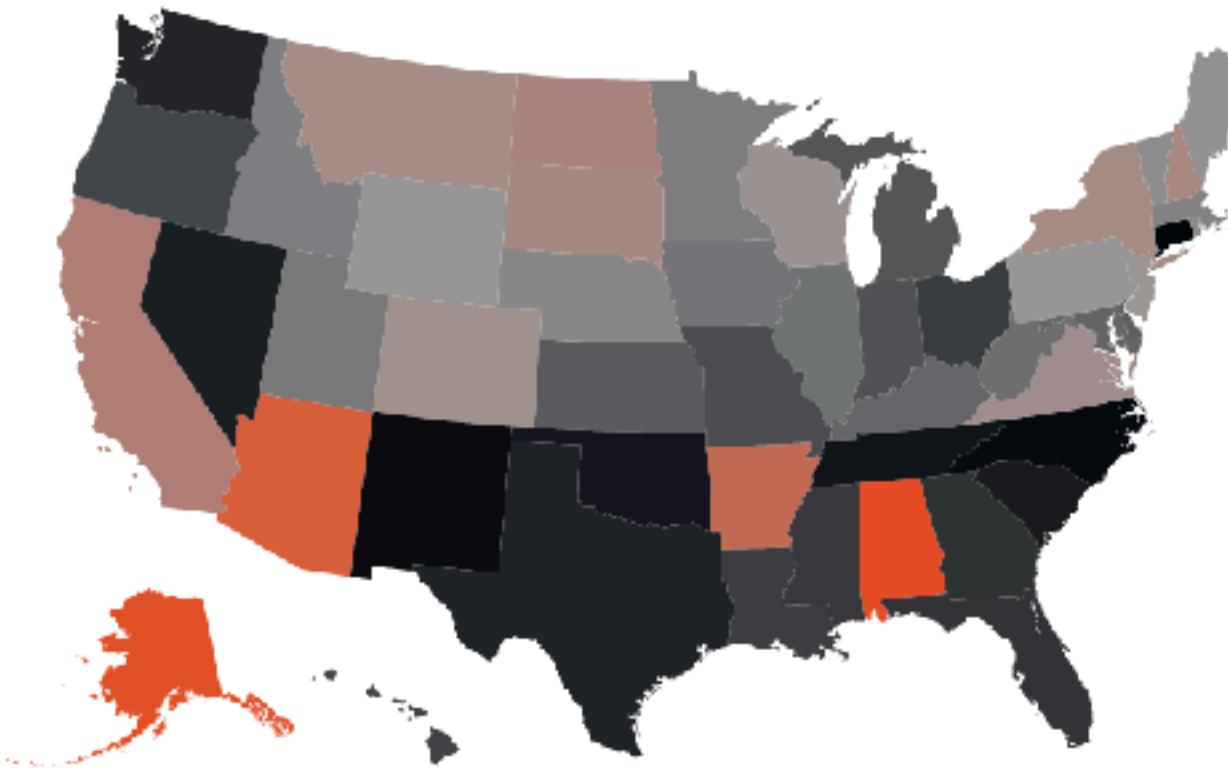
Savings Amount	65+	55-64	45-54	35-44	25-34	18-24	Overall
\$10,000 or more	20	16.8	16.2	16	12.1	7.5	14
\$5,000-\$9,999	4.7	4.8	5.2	5.6	5.4	4.7	5
\$1,000-\$4,999	7.2	8	7.5	9.8	12.5	14.7	10
Less than \$1,000	8.2	10.7	10.9	11.6	15.2	19.1	13
Just the minimum balance required	10.7	8.4	7.7	6.6	10.6	9.7	9
\$0	27.6	28.4	30.8	31.6	26.3	21.8	28
I don't have a savings account	21.6	22.8	21.6	18.9	18	22.4	21

Spatial

Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

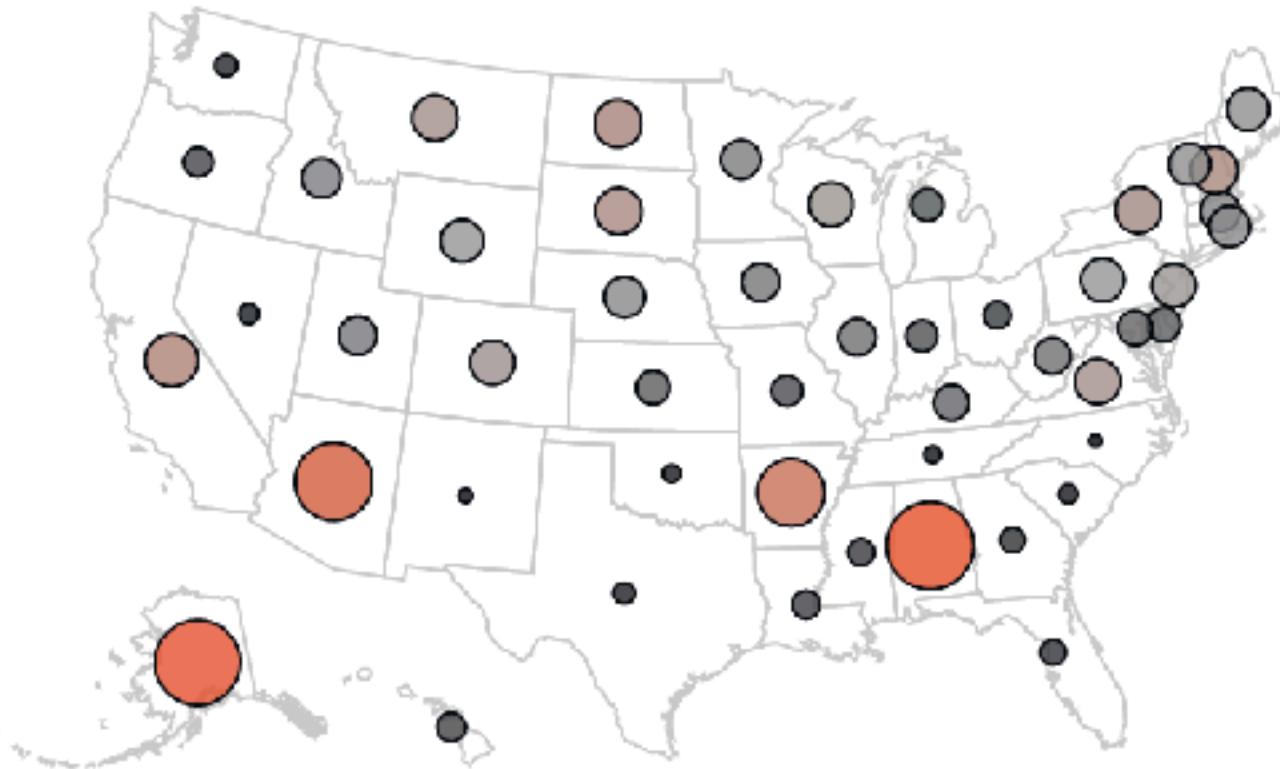
Basic choropleth (rate/ratio)

The standard approach for putting data on a map - should always be rates rather than totals and use a sensible base geography.



Proportional symbol (count/magnitude)

Use for totals rather than rates - be wary that small differences in data will be hard to see.



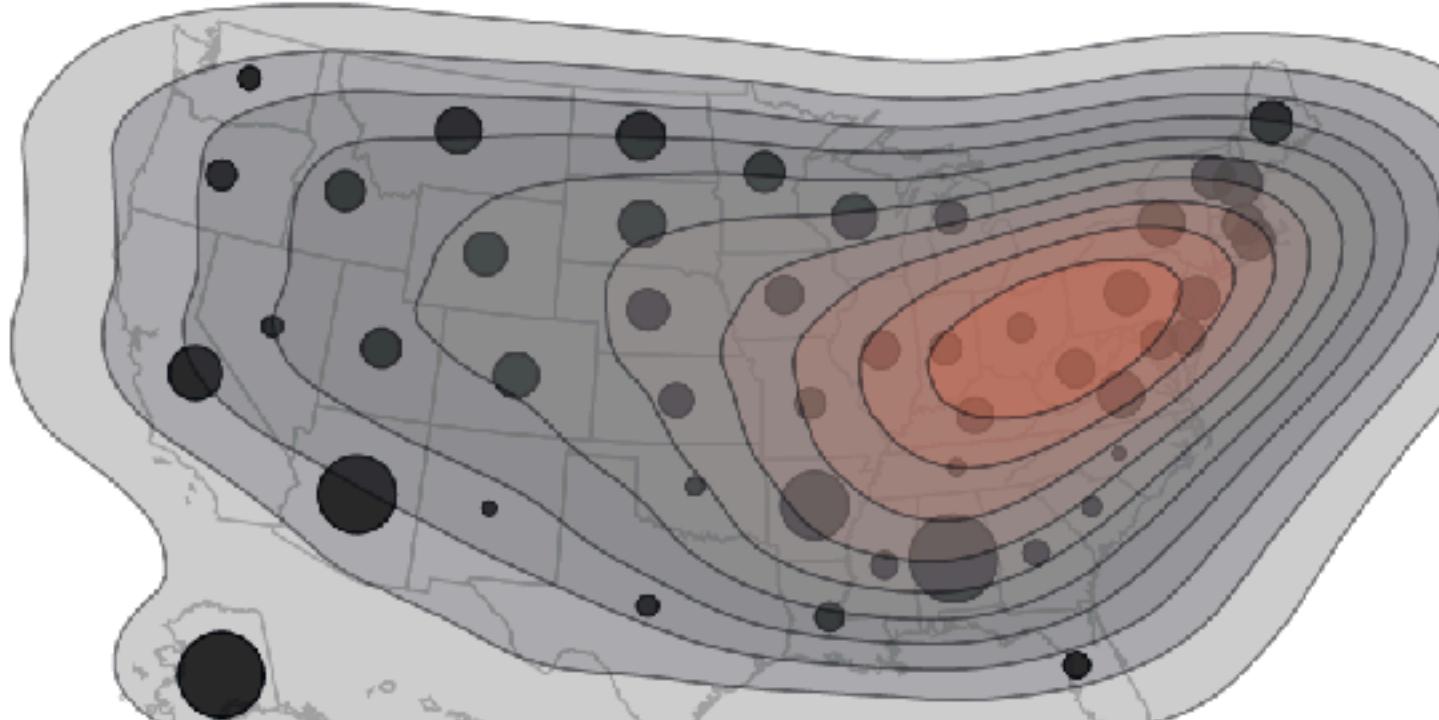
Flow map

For showing unambiguous movement across a map



Contour map

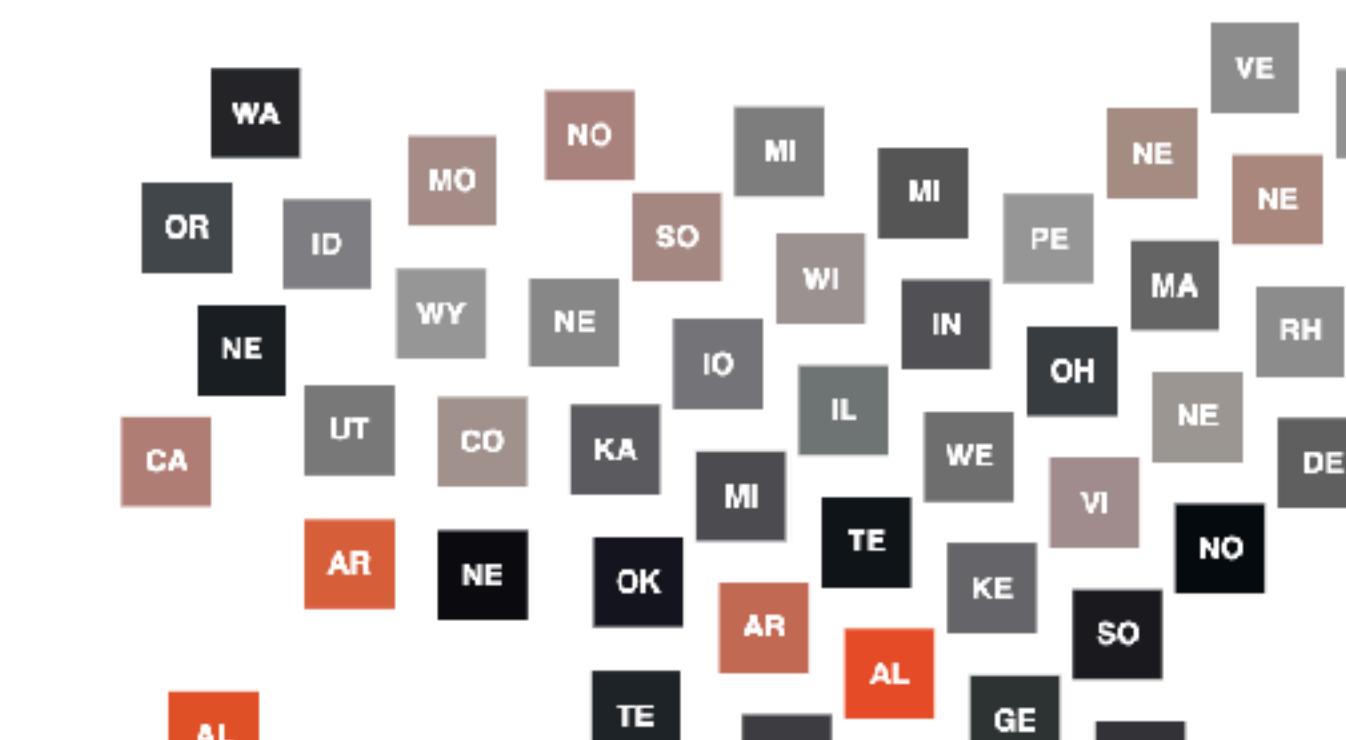
For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values



Edit

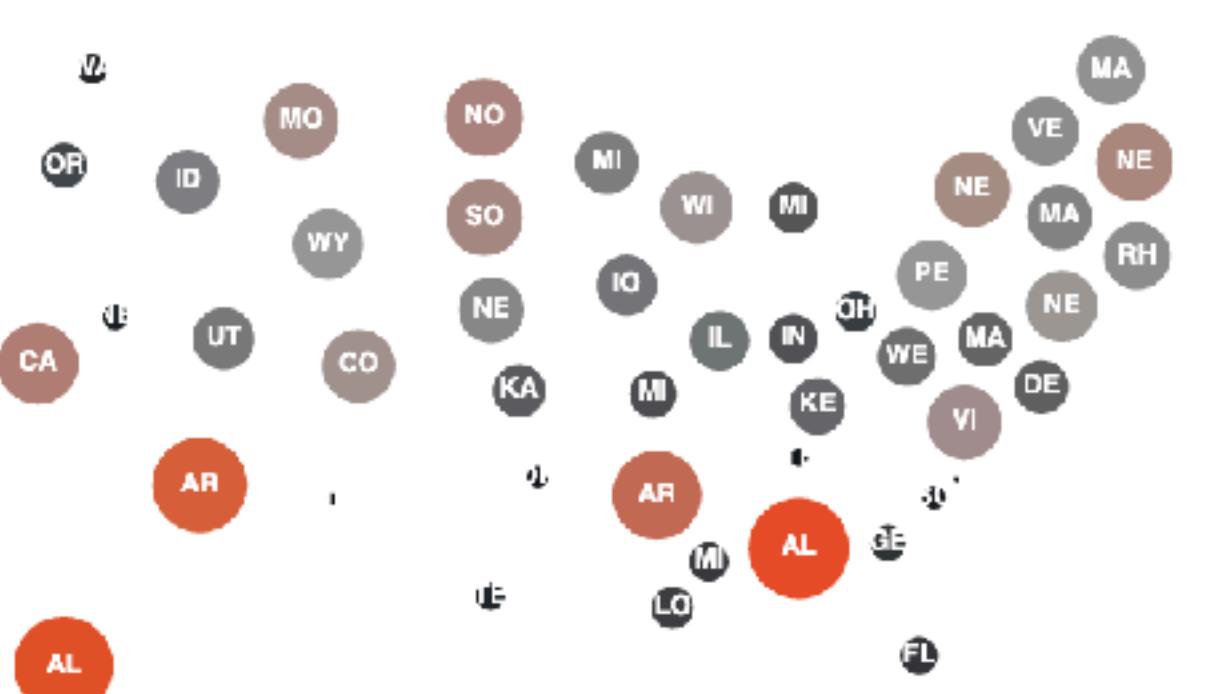
Equalised cartogram

Converting each unit on a map to a regular and equally-sized shape - good for representing voting regions with equal share.



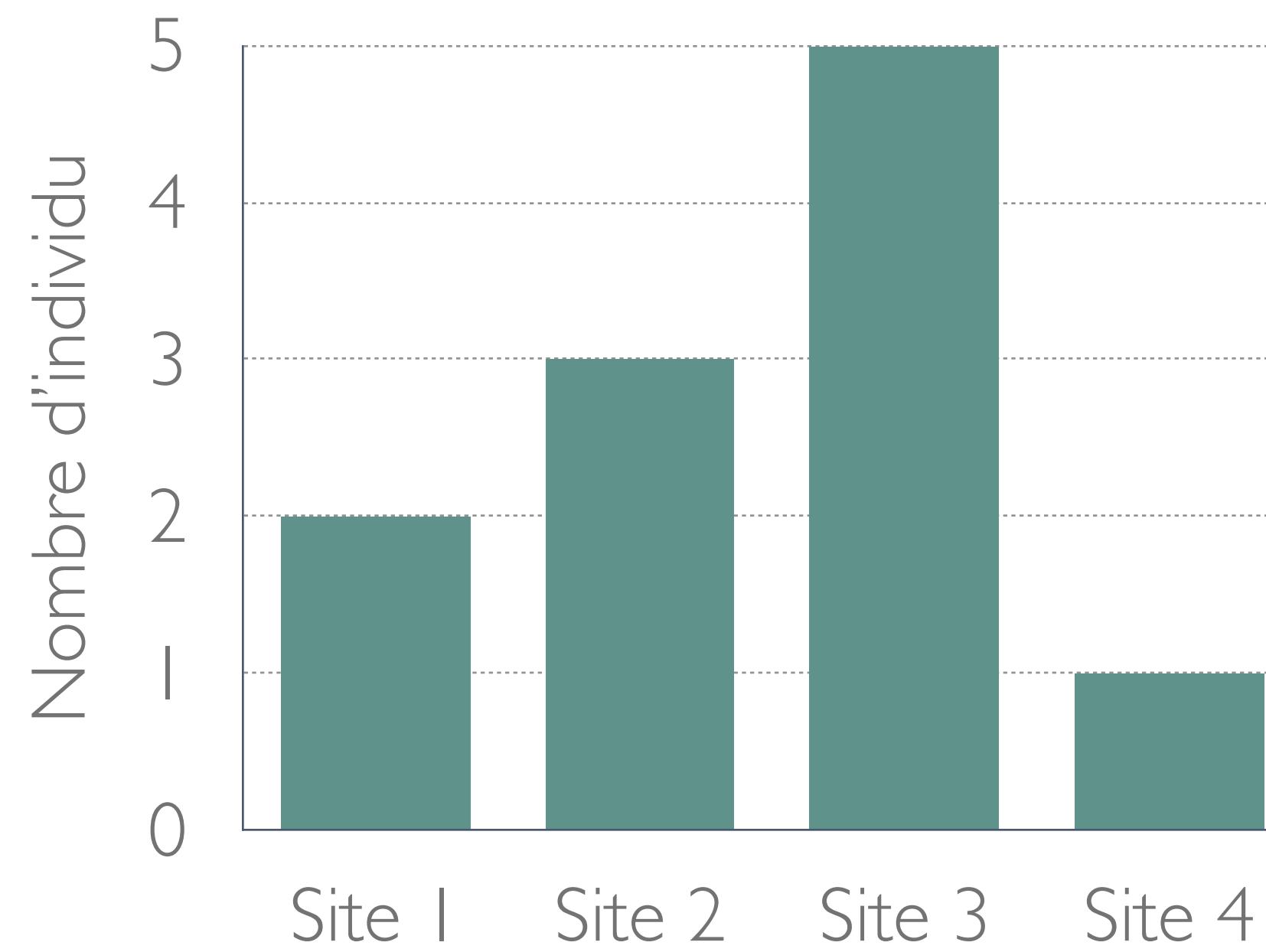
Scaled cartogram (value)

Stretching and shrinking a map so that each area is sized according to a particular value.



Un bar plot comme exemple

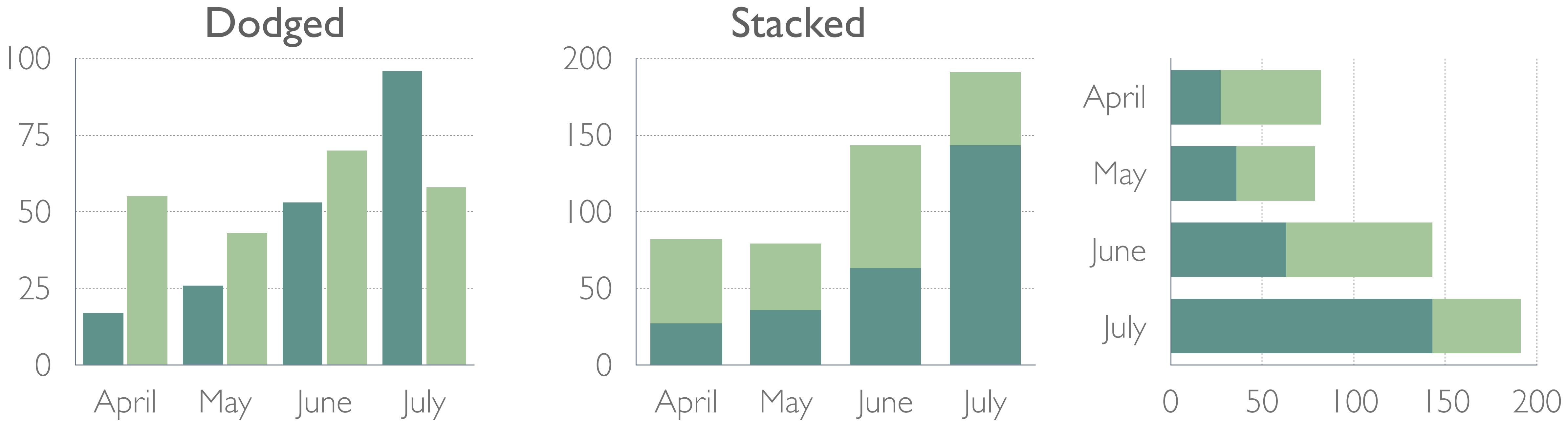
	Site 1	Site 2	Site 3	Site 4
Nombre	2	3	5	1



Id	Site	Sex
1	1	M
2	1	M
3	2	M
4	2	F
5	2	M
6	3	M
7	3	F
8	3	F
9	3	F
10	3	M
11	4	M

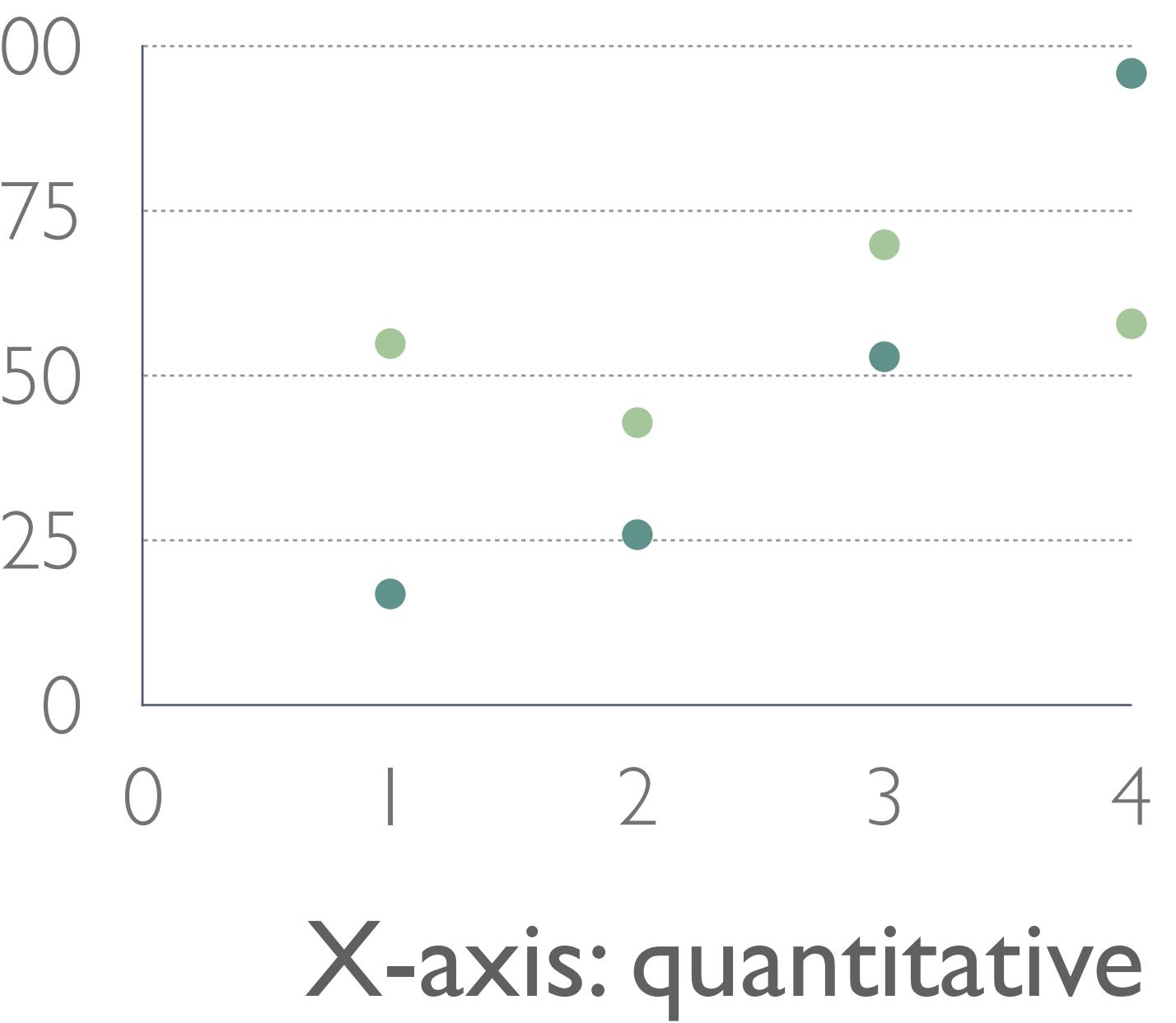
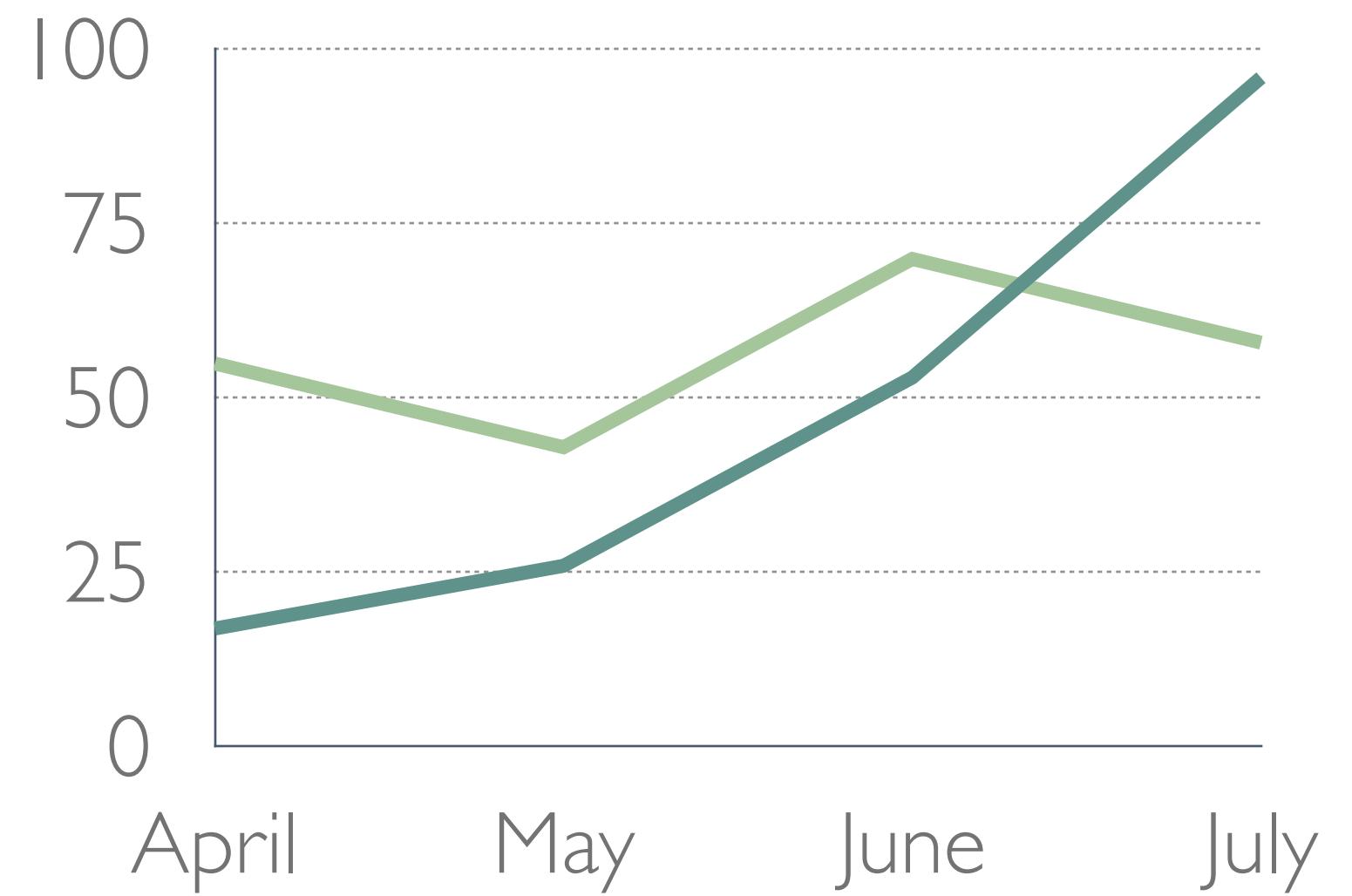
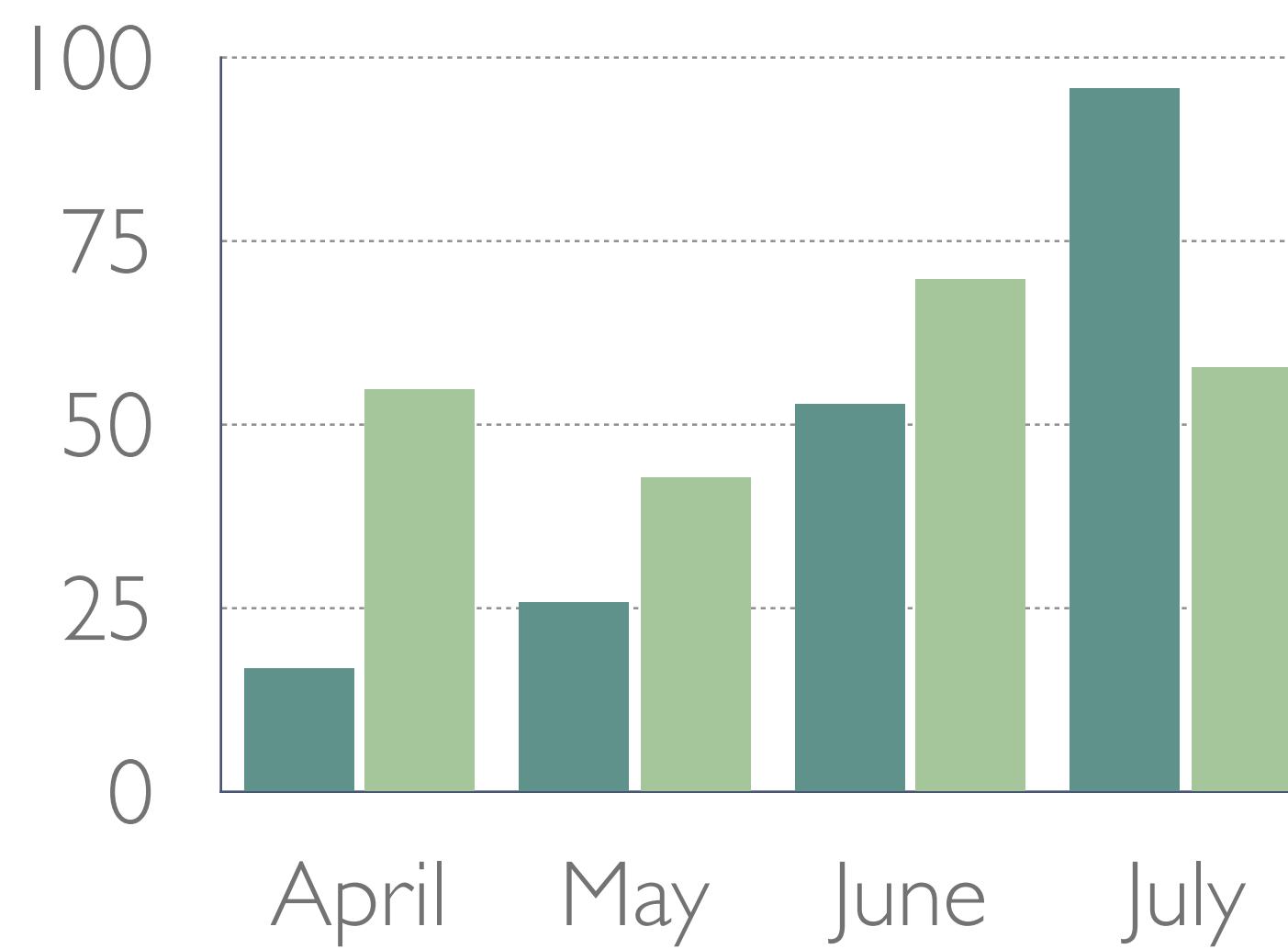
- Data format
- X-axis: Qualitative
- Y-axis: Quantitative
- Y-axis
 - Explicite
 - A calculer

Bar plot: X-axis à deux dimensions

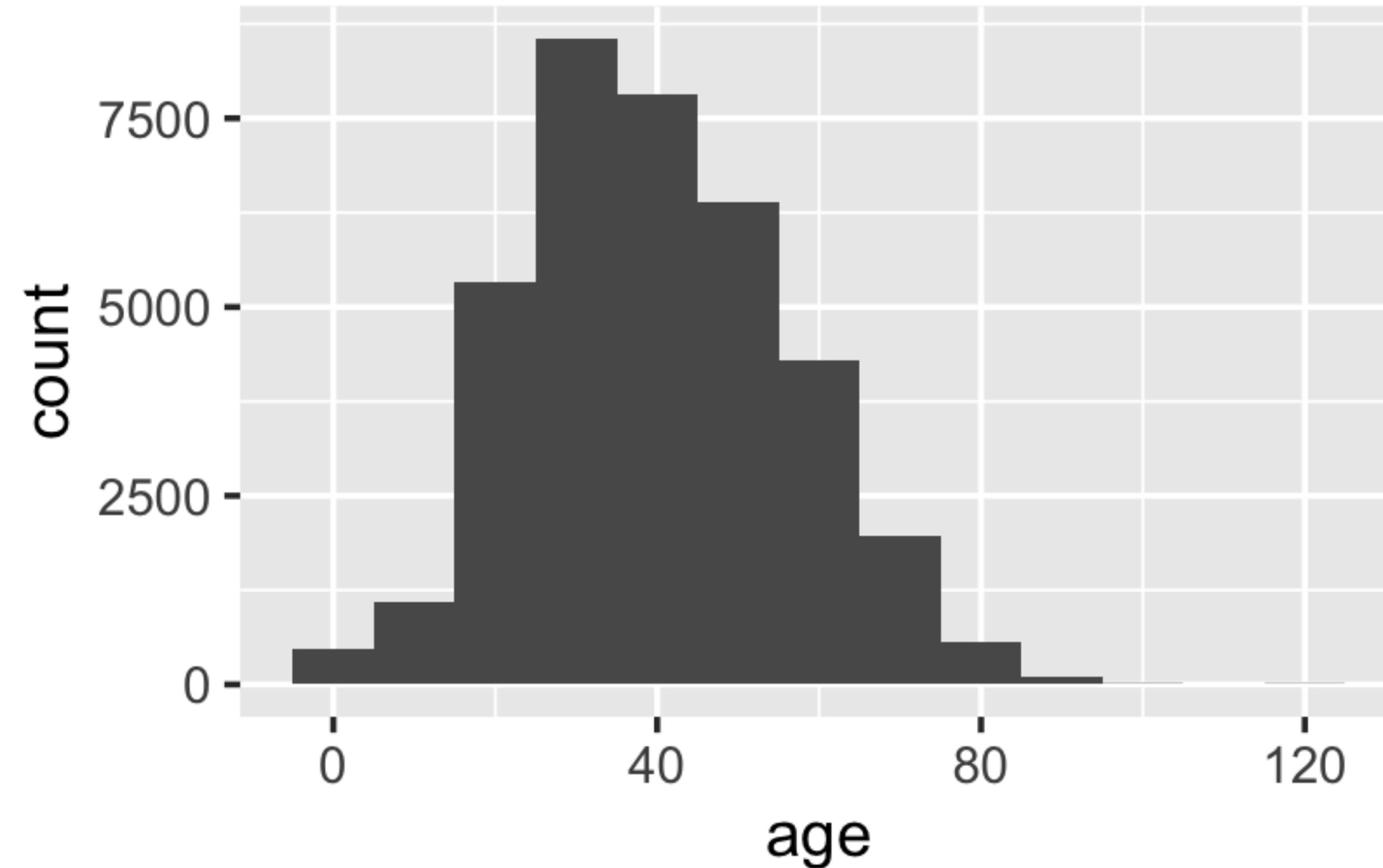
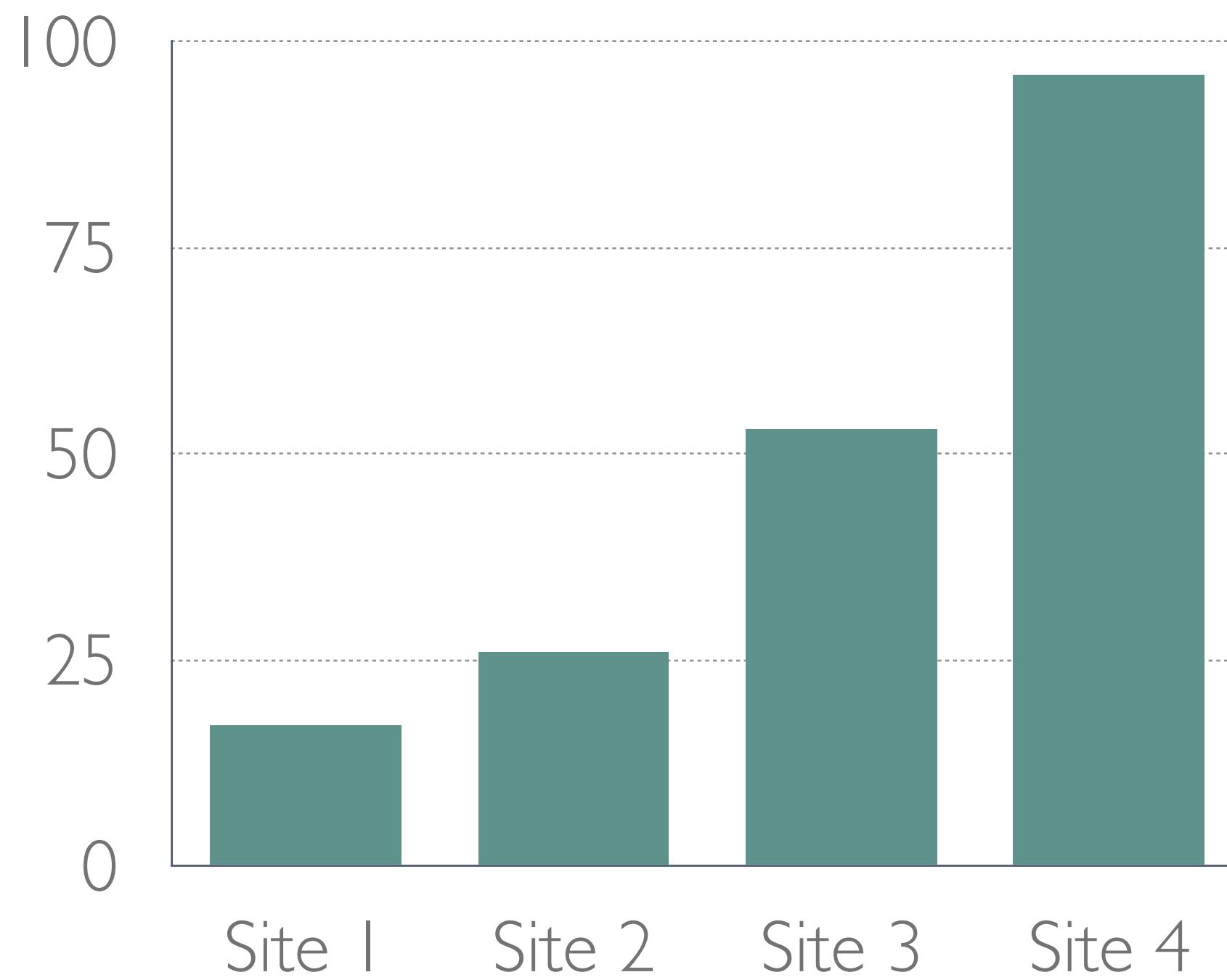


Quel est le problème avec ces figures?

Bar plot vs. line plot vs. scatter plot

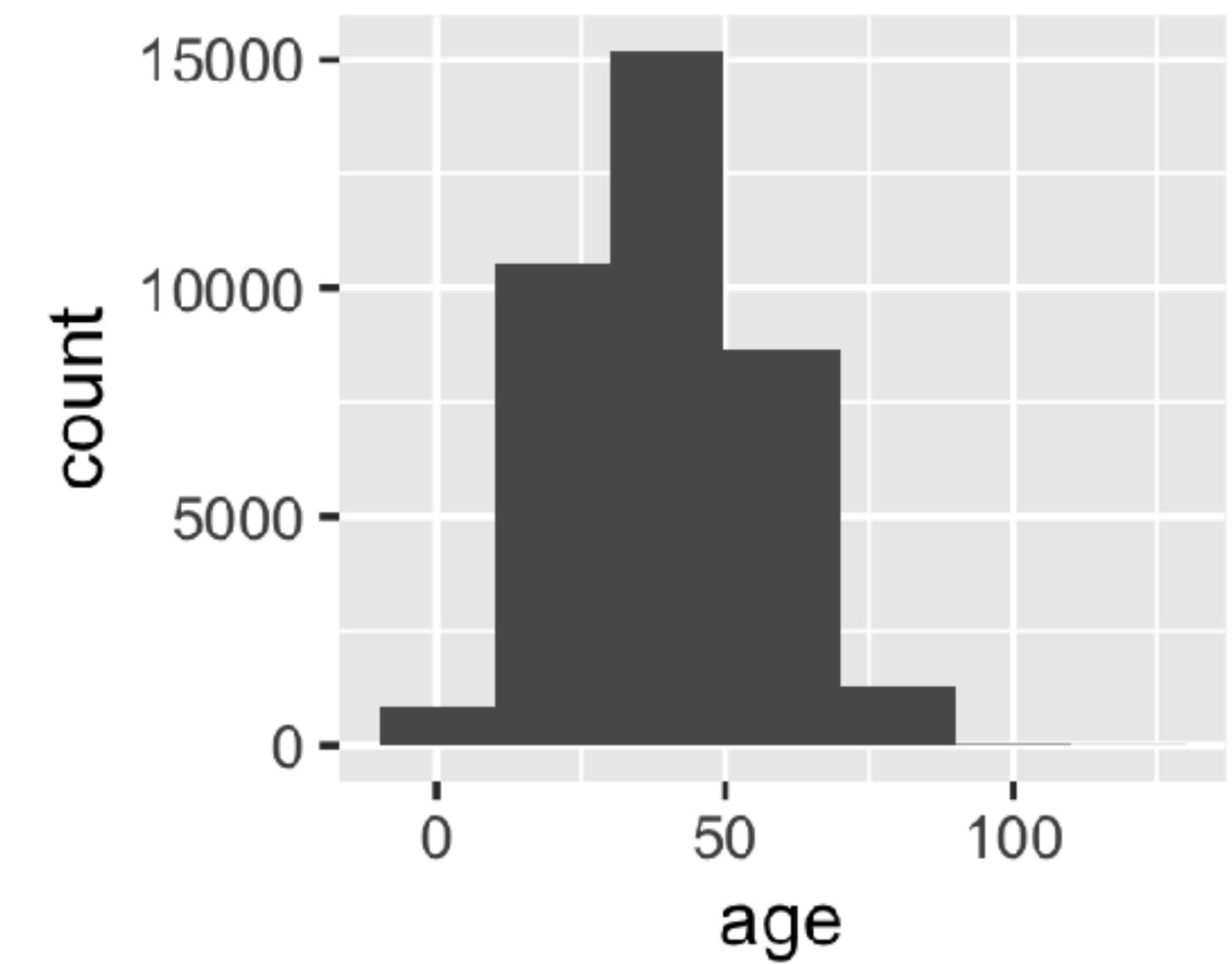
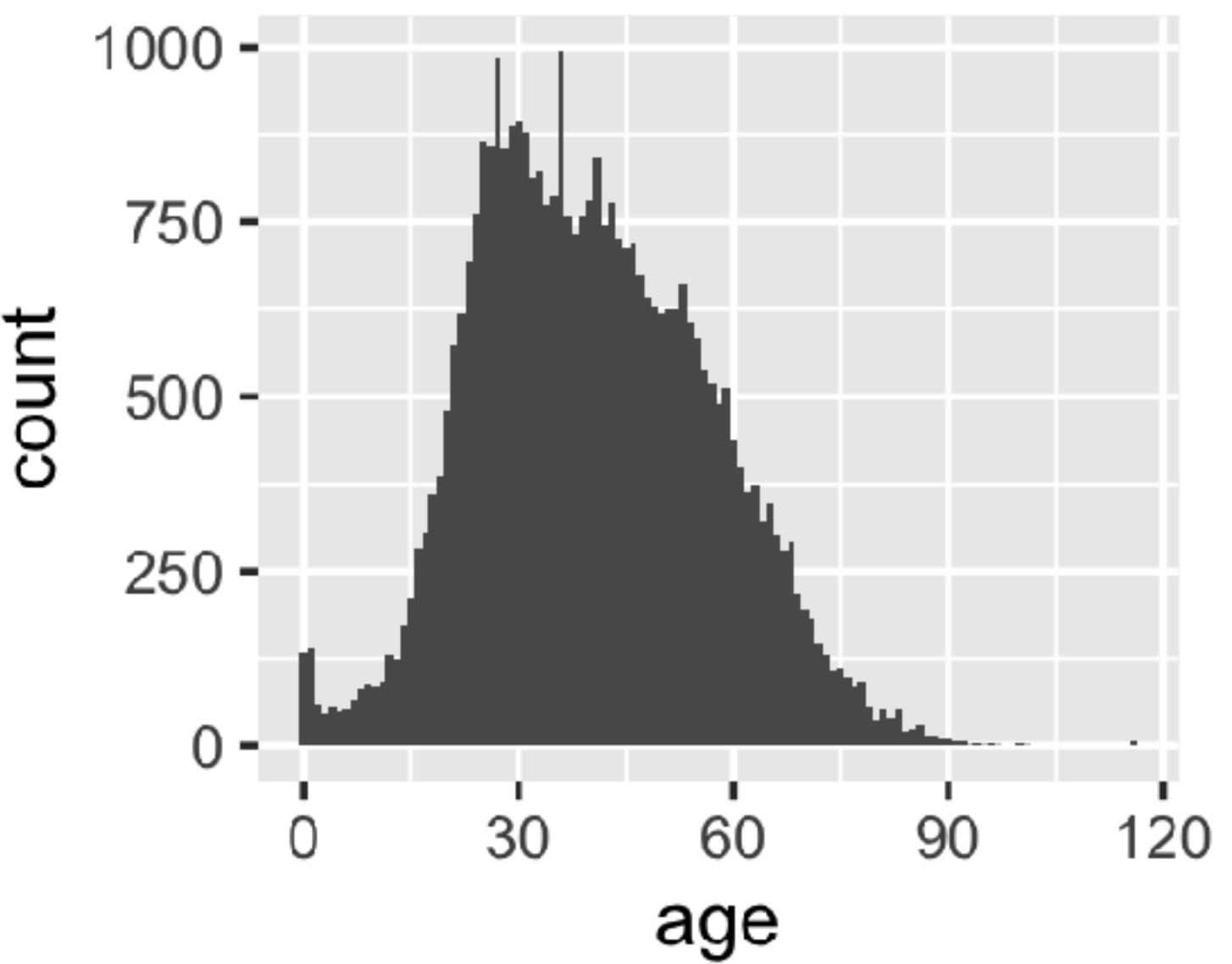
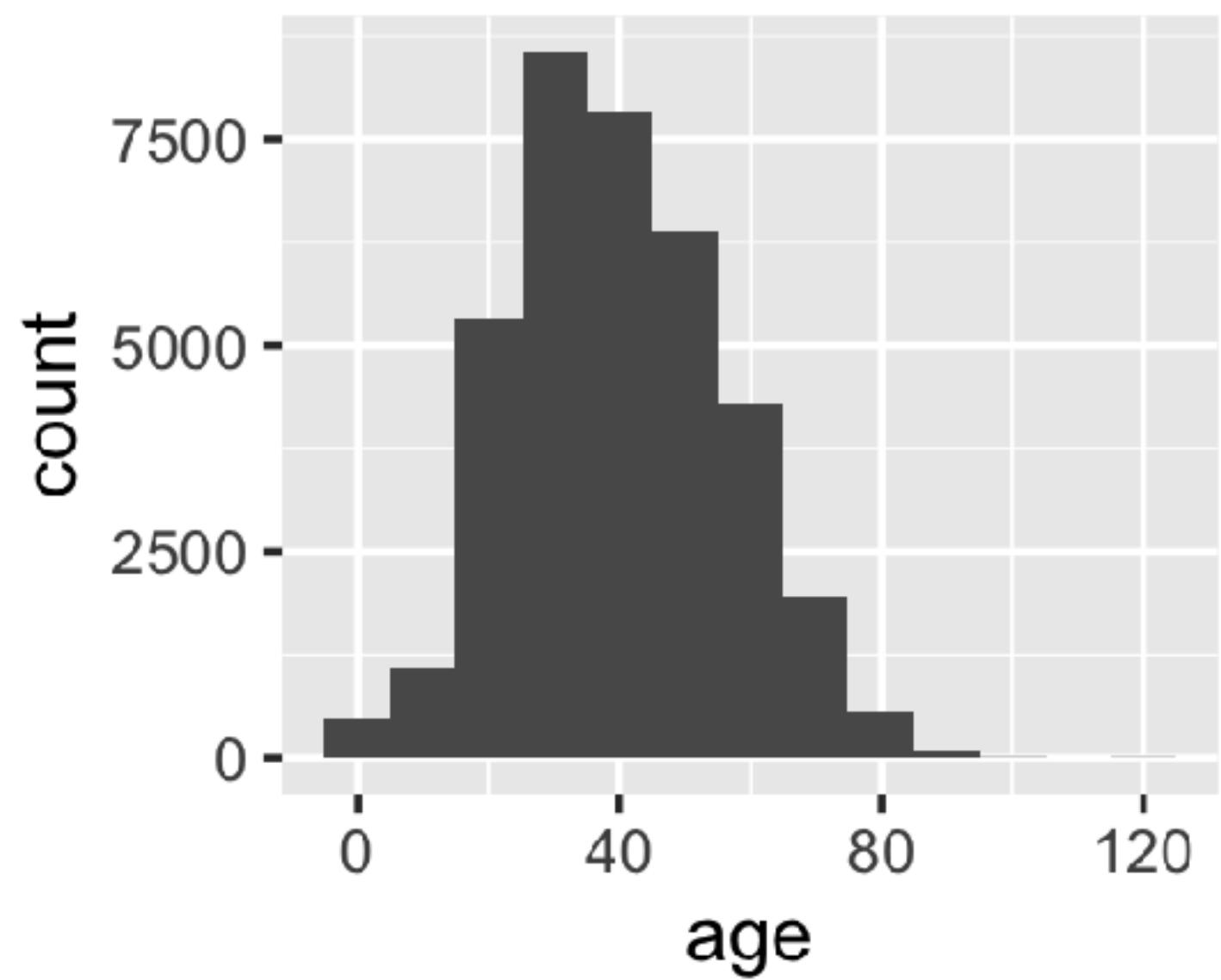


Bar plot vs Histogram



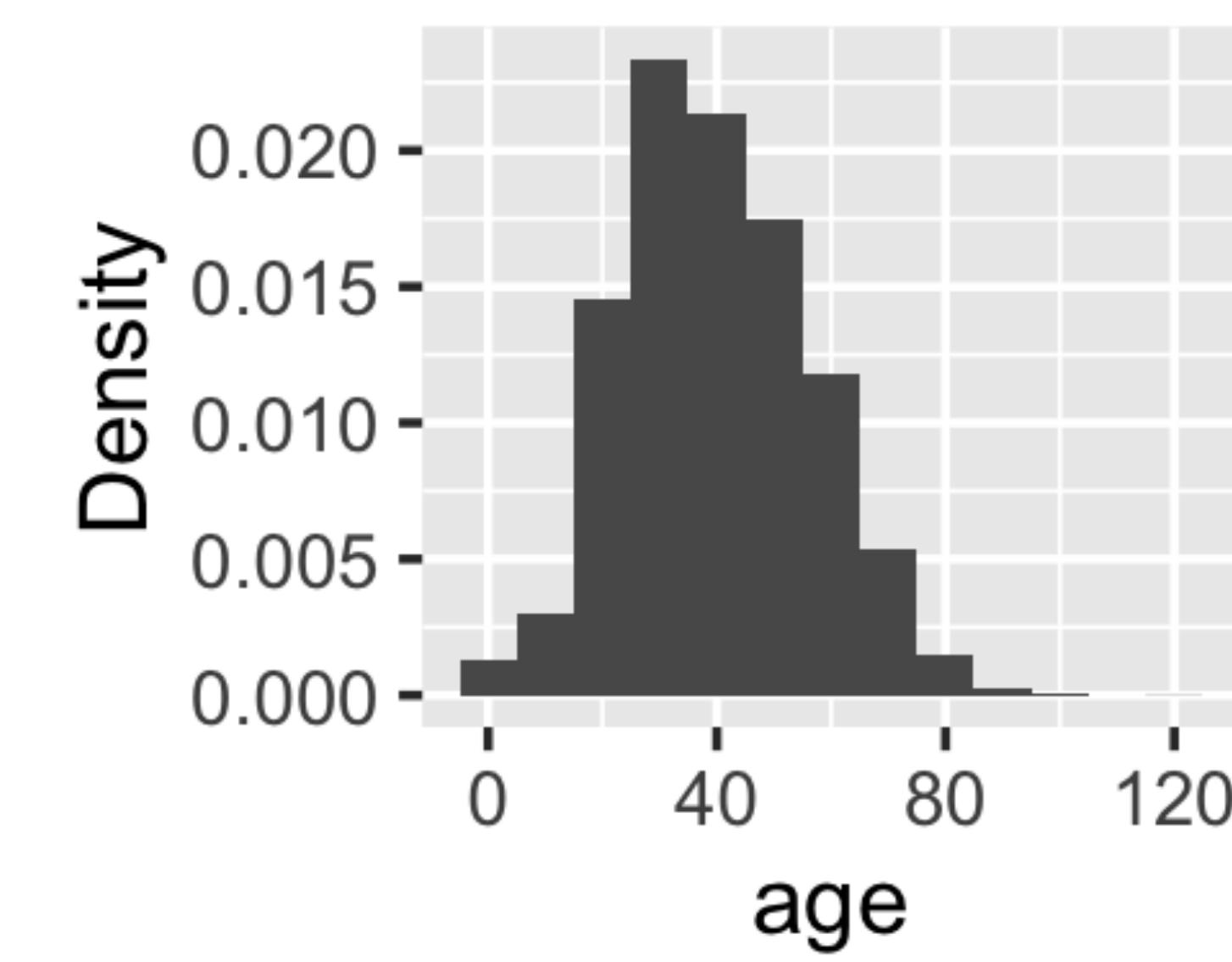
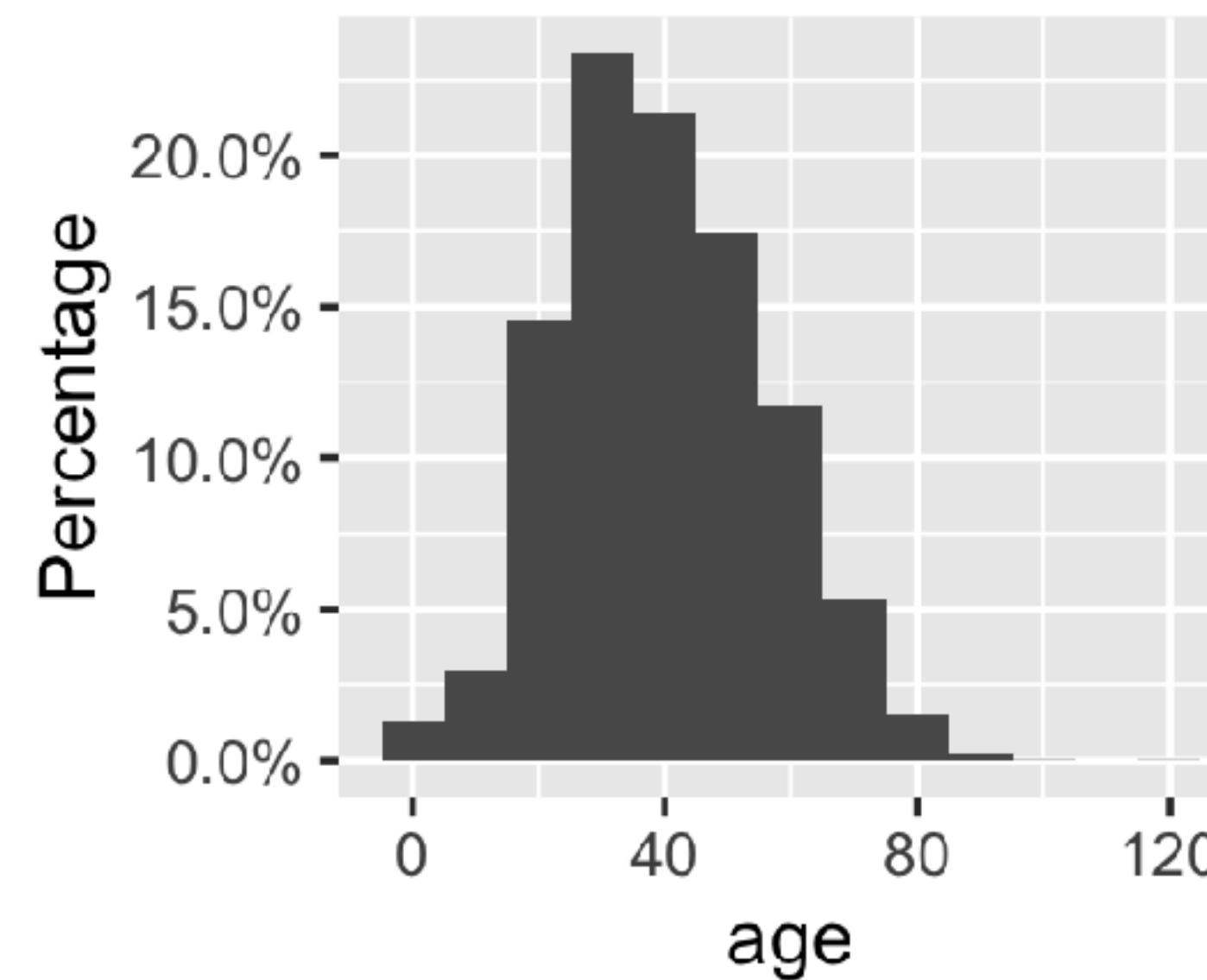
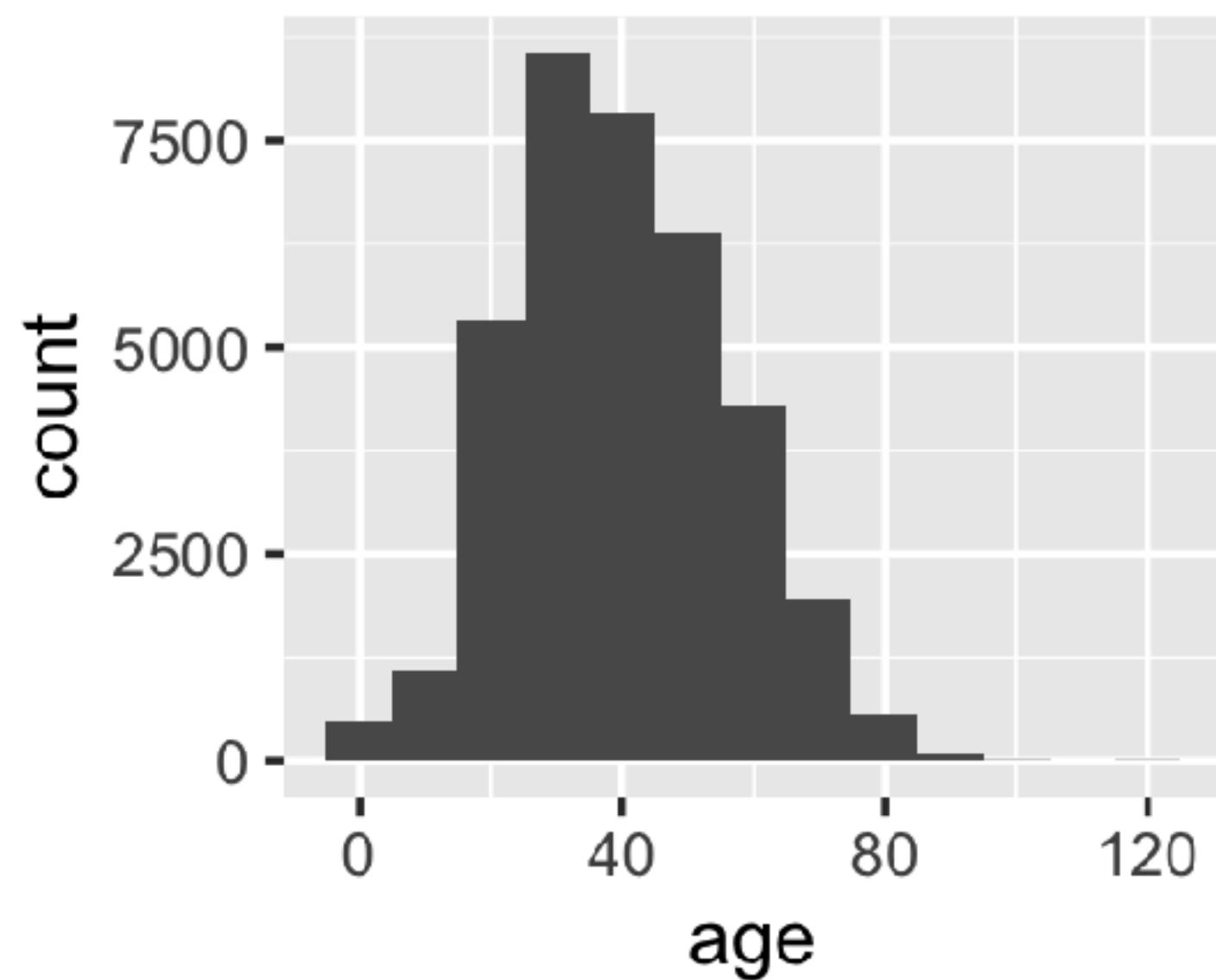
Quel est la différence fondamentale entre les deux types de figures?

Histogram et bin width



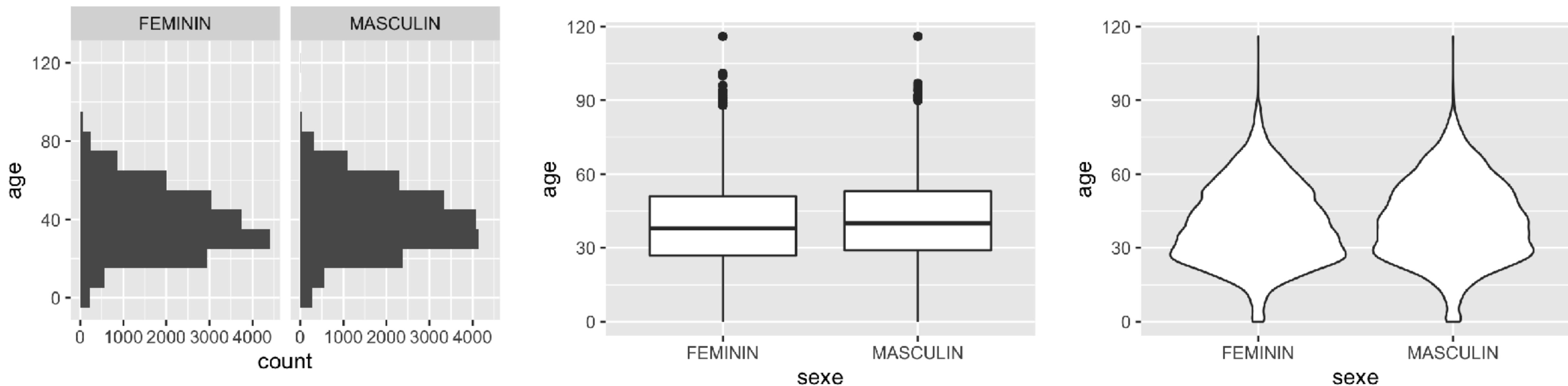
Y-axis à calculer

Histogram et échelle sur y-axis



Y-axis à calculer

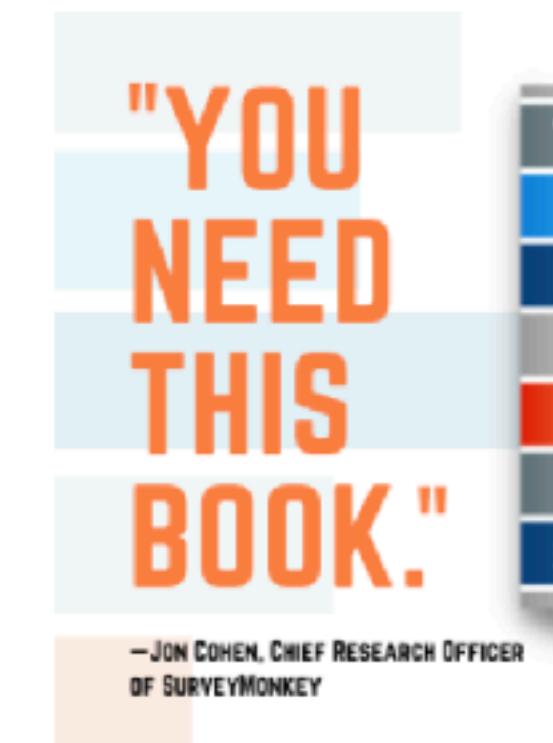
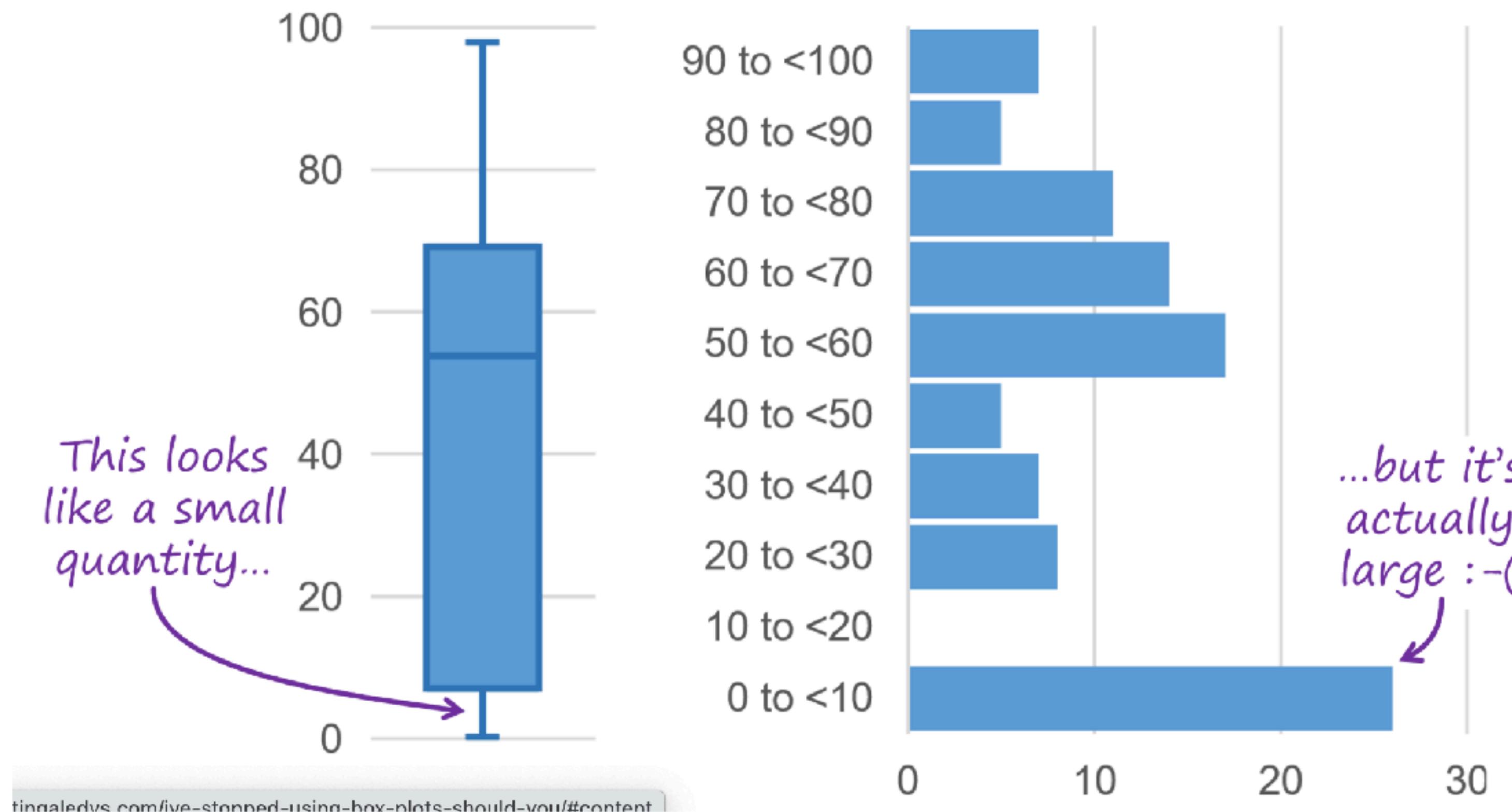
Histogram vs Box plot vs Violin plot



Y-axis à calculer

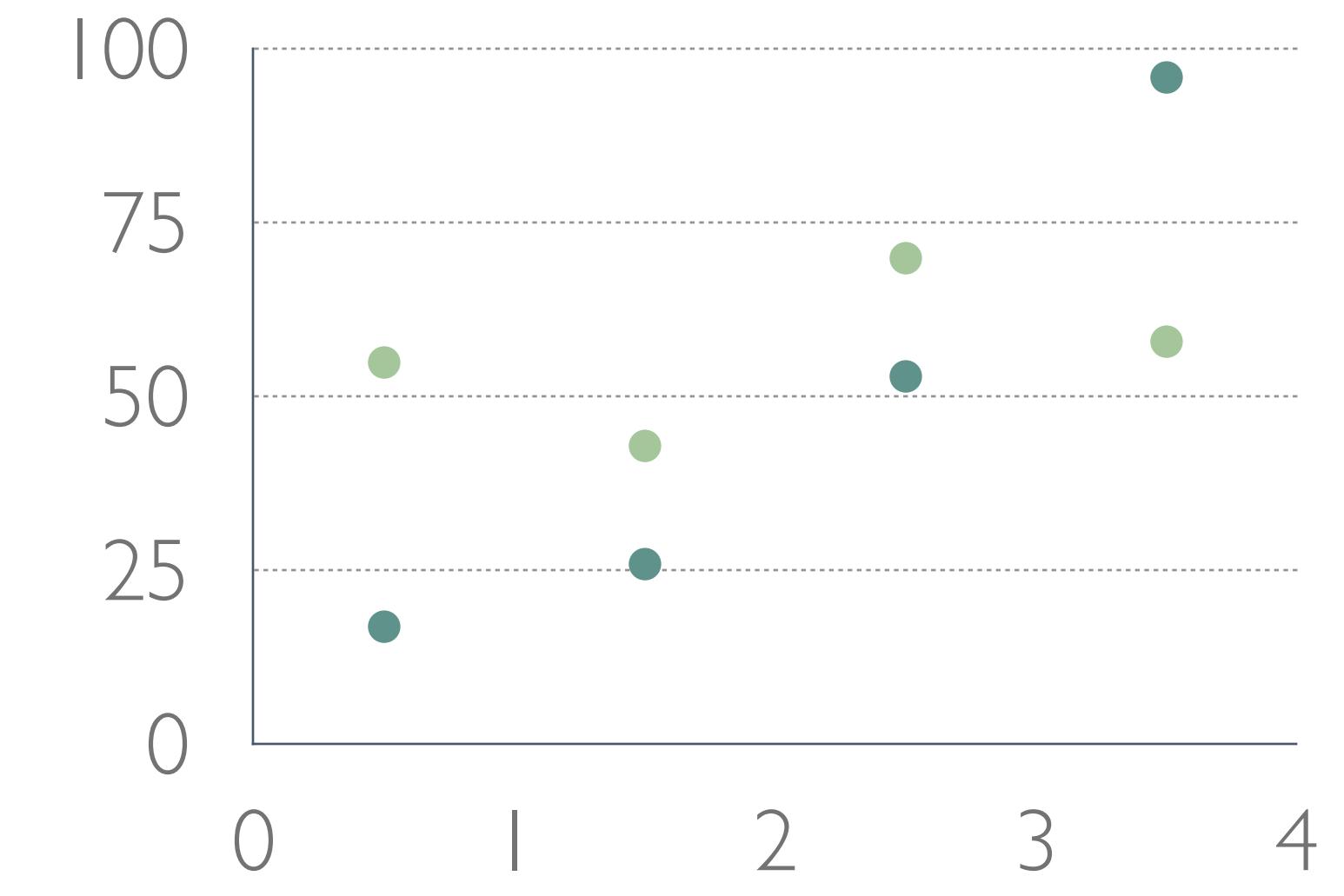
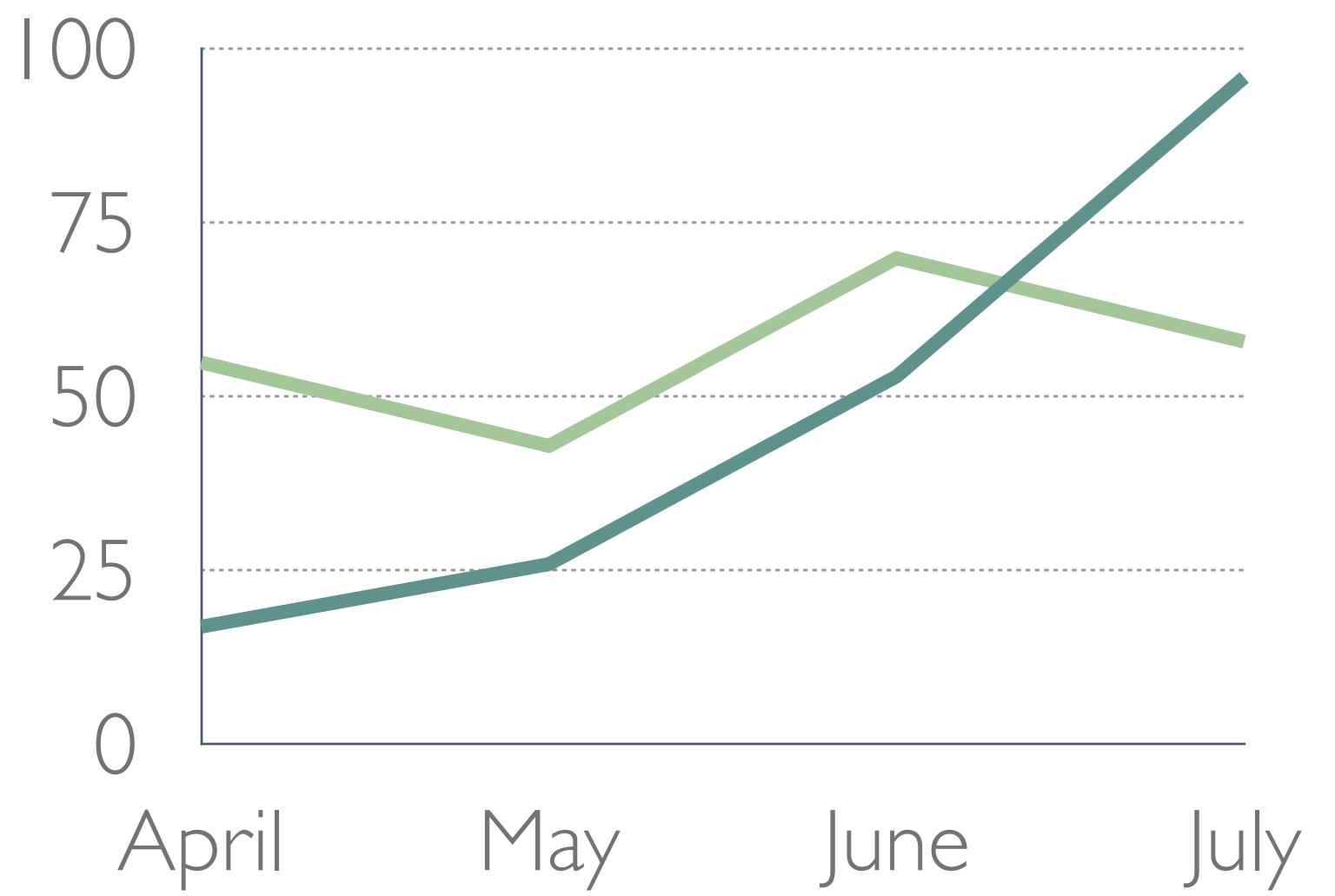
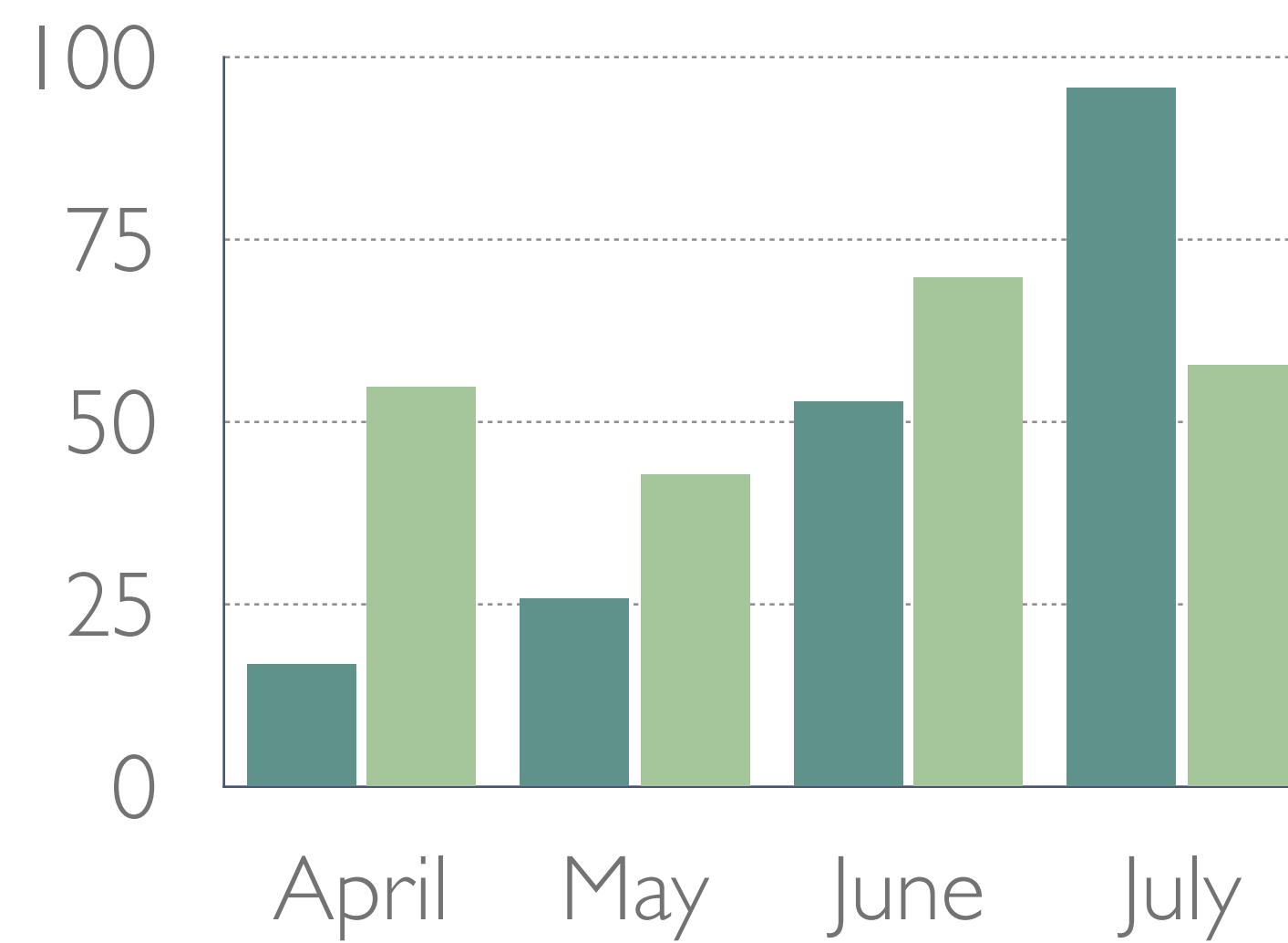
I've Stopped Using Box Plots. Should You?

Nick Desbarats • November 4, 2021



Help Nigh
get to 1,000
subscribers
Nov. 15th!

Série temporelle: x-axis = date



Data frame et long format

	...1	date	dob	age	region	District	sexe
1	1	2021-04-20	1957-07-31	64	Analamanga	ANTANANARIVO RENIVOHITRA	MASCULIN
2	2	2021-04-20	1971-06-30	50	Analamanga	ANTANANARIVO RENIVOHITRA	MASCULIN
3	3	2021-04-20	1983-07-29	38	Atsinanana	TOAMASINA I	MASCULIN
4	4	2021-04-20	1970-06-29	51	Analamanga	ANTANANARIVO RENIVOHITRA	FEMININ
5	5	2021-04-20	1998-05-29	23	Analamanga	ANTANANARIVO AVARADRANO	FEMININ
6	6	2021-04-20	1983-04-29	38	Analamanga	ANTANANARIVO RENIVOHITRA	FEMININ
7	7	2021-04-20	1961-01-28	60	Analamanga	ANTANANARIVO RENIVOHITRA	FEMININ
8	8	2021-04-20	1976-10-27	45	Analamanga	ANTANANARIVO RENIVOHITRA	FEMININ
9	9	2021-04-20	1989-05-27	32	Analamanga	ANTANANARIVO RENIVOHITRA	MASCULIN
10	10	2021-04-20	1992-10-26	29	Analamanga	ANTANANARIVO RENIVOHITRA	MASCULIN
11	11	2021-04-20	1994-12-25	26	Analamanga	ANTANANARIVO AVARADRANO	FEMININ

Le strict minimum

- Data
 - Qu'est-ce qu'il y a sur le x- et y-axis?
 - Quelle type de figure vous voulez?
-

Syntaxe minimal

Donnée

Les variables sur x-axis et y-axis

Le type de figure

```
ggplot(dataset1, aes(x = sexe, y = age))+
  geom_boxplot()
```

Exemple concret

```
ggplot(dataset1, aes(x = sexe, y = age)) +  
  geom_boxplot()
```



Un conseil: avant de faire une figure sur R, essayer de faire une caricature

Resources (with materials from Malavika Rajeev)

- Brouillon d'un livre : <https://ggplot2-book.org/index.html>
- ggplot cheatsheet: <https://nyu-cdsc.github.io/learningr/assets/data-visualization-2.l.pdf>
- Data carpentry course: <https://datacarpentry.org/R-ecology-lesson/04-visualization-ggplot2.html>
-