

# **Data Visualizations & Graphical Storytelling for Data Analysts and Data Scientists**

## **Introduction to Data Visualization Design**

**Dr Cédric Scherer**

Hello Heart // March 2024



# *Introduction to Data Visualization Design*



# *Data Visualization*

is any graphical representation  
of information and data.



# *Data Visualization*

helps to amplify cognition, gain insights,  
discover, explain, and make decisions.



# *Data Visualization*

**converts information into visual forms  
as quantifiable features.**



# *Data Visualization*

is part art and part science.



# Typology of Information Graphics

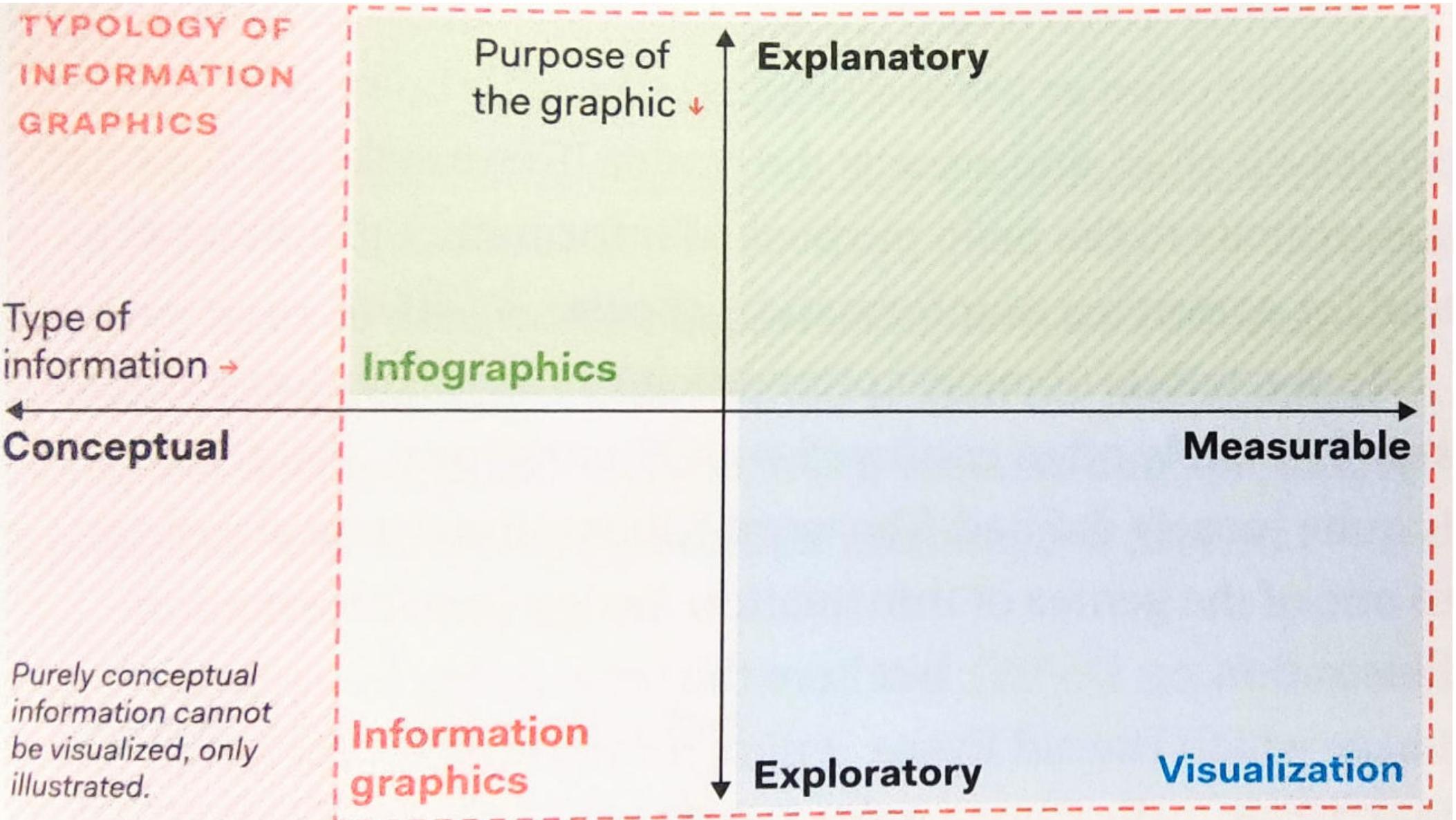
Is the information **conceptual** or **measurable**?

→ Type of information: depict information schematically <> convert information into visual forms

Is the aim to **explore** or to **explain** the information?

→ Purpose of the graphic: facilitate discovery <> communicate information





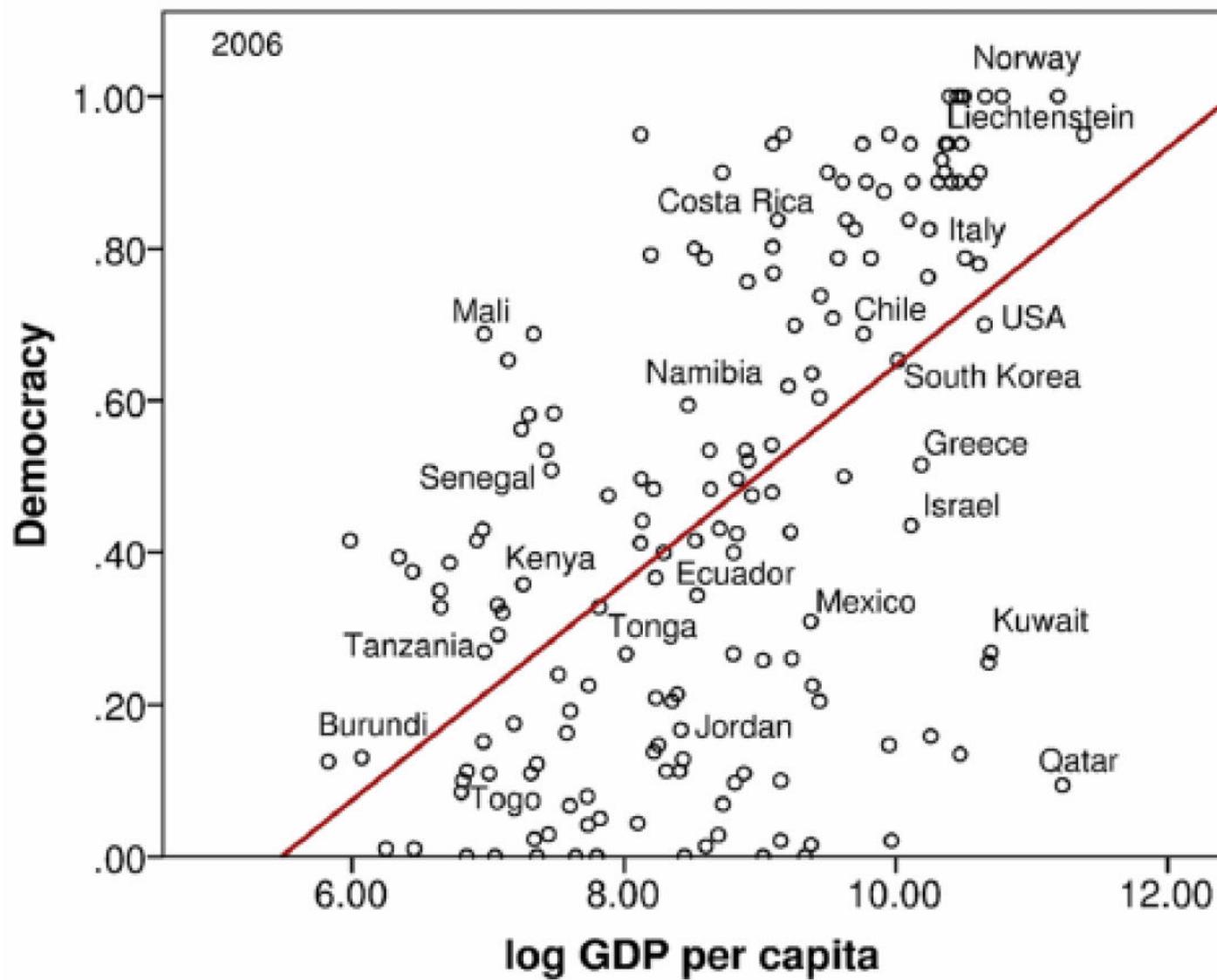
Source: Koponen & Hildén, "Data Visualization Handbook" (2020), page 25





Source: eazybi

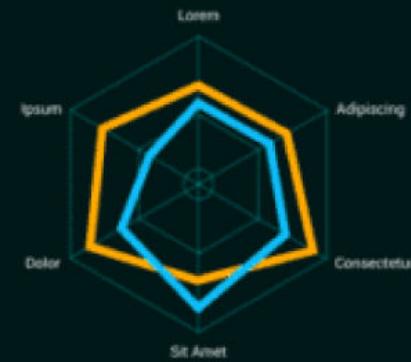
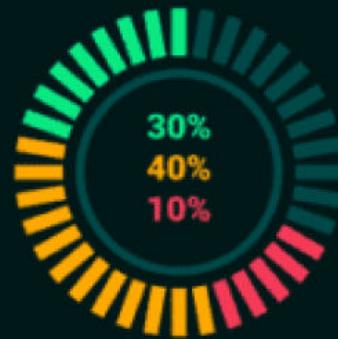
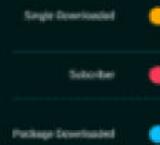
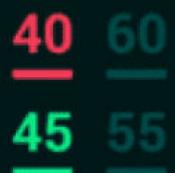
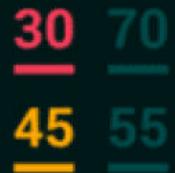
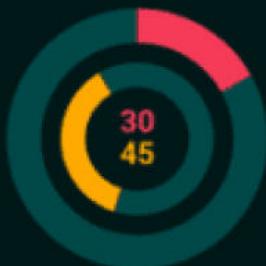




Source: Ranganathan et al. 2014

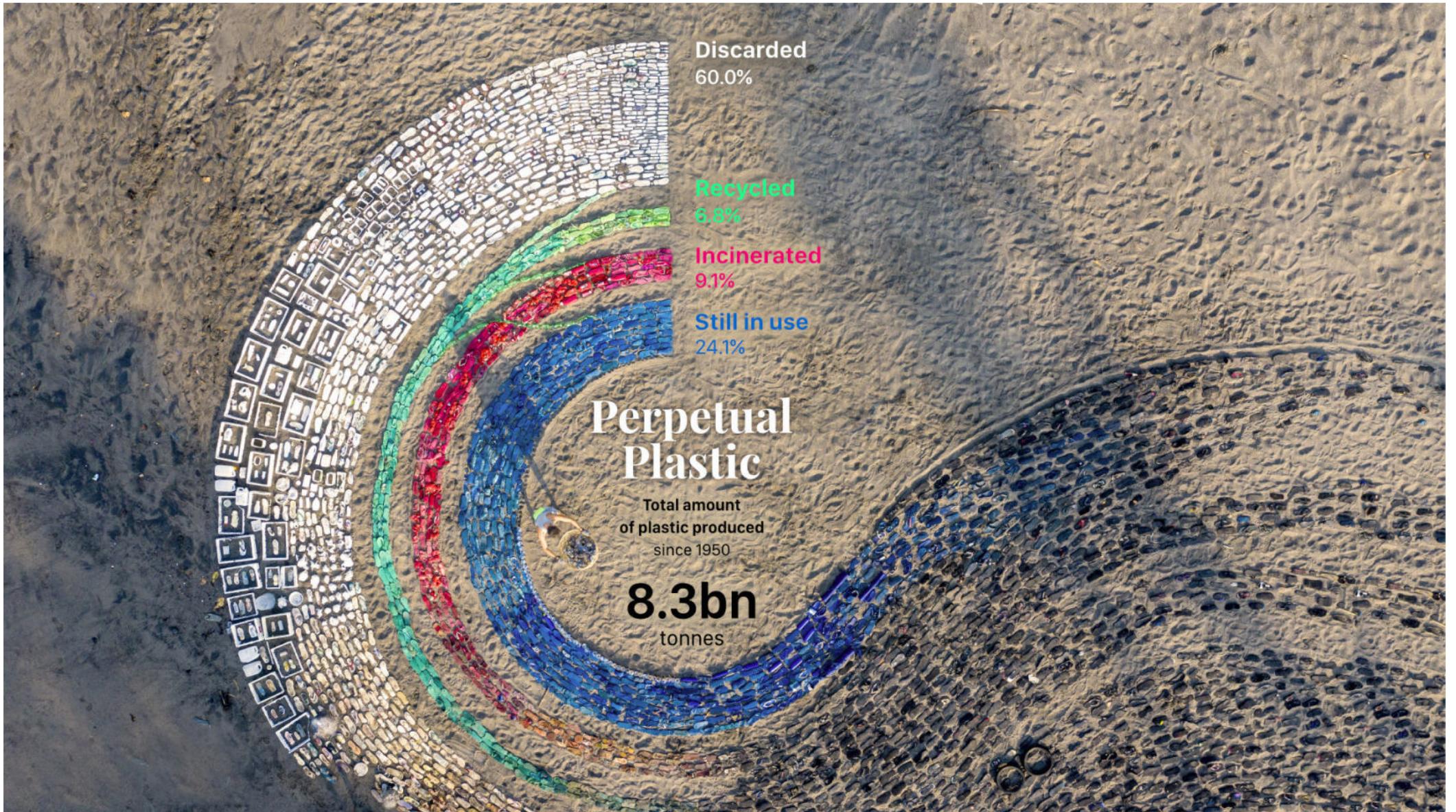
Cédric Scherer @ Hello Heart // Data Visualization & Information Design





Source: [datameer.com](http://datameer.com)





Source: "Perpetual Plastic" by Liina Klauss, Skye Morét & Moritz Stefaner

Cédric Scherer @ Hello Heart // Data Visualization & Information Design





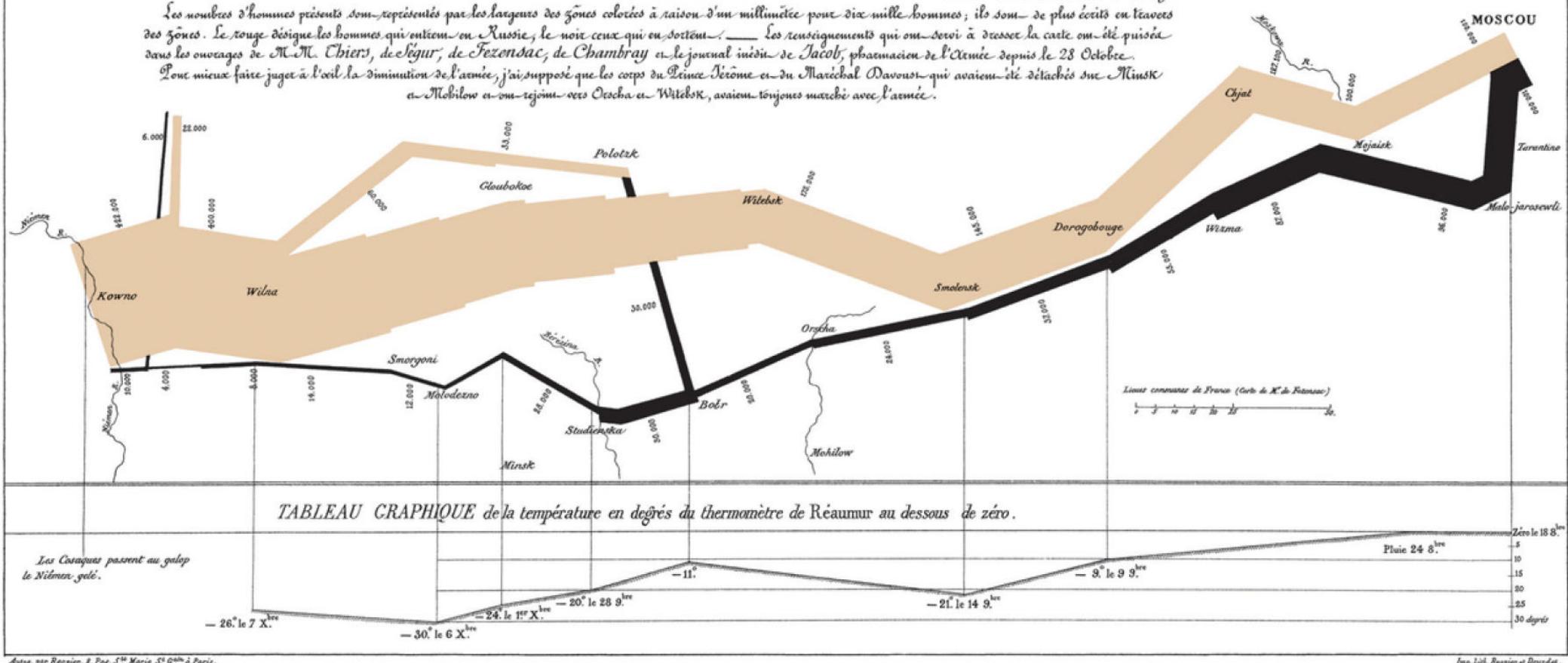
Source: "Patchwork Kingdoms" by Nadieh Bremer



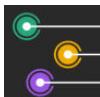
*Carte Figurative* des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.  
Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

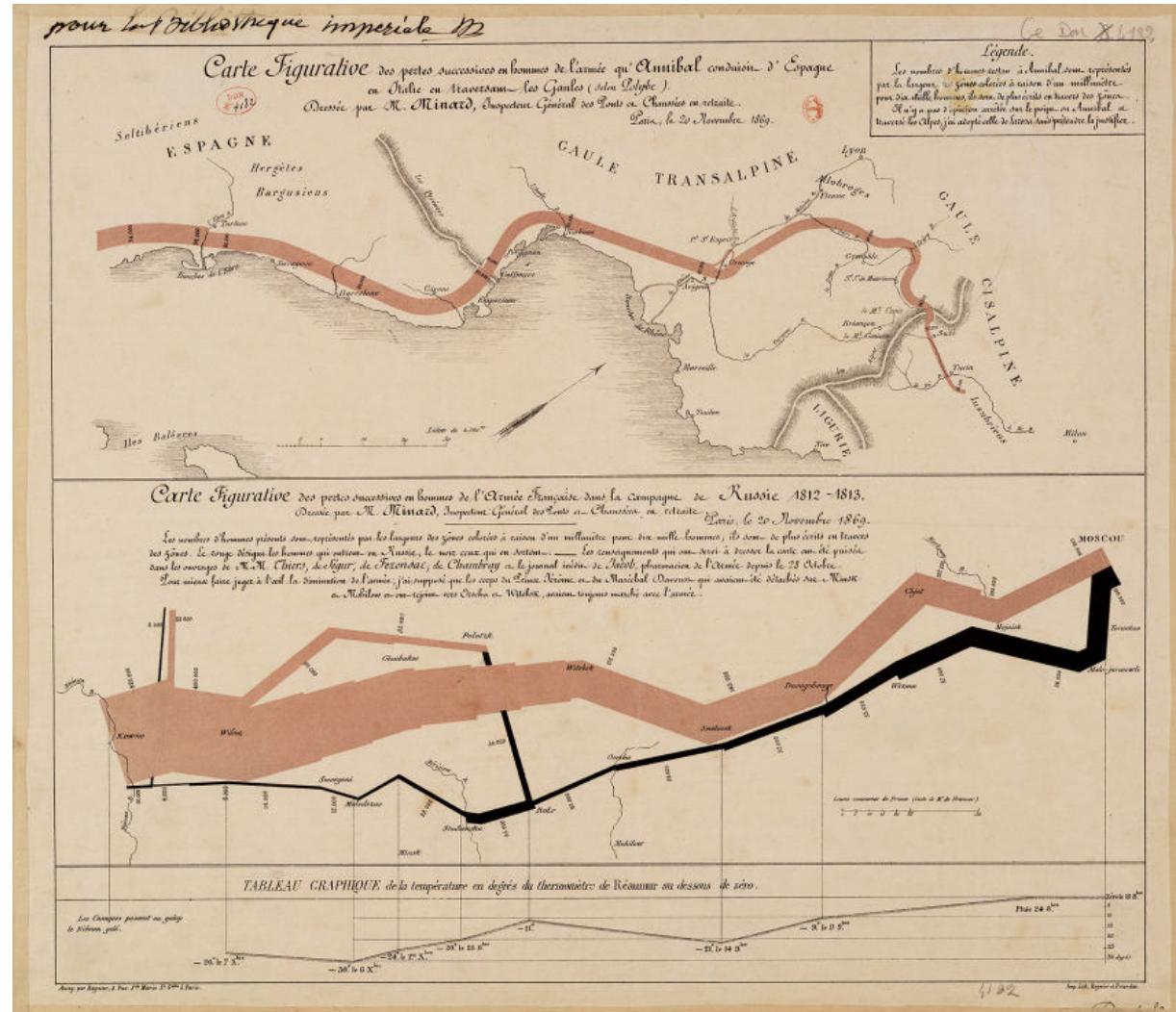
Les nombres d'hommes perdus sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été pris dans les ouvrages de M. Chiers, de Séguir, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Napoléon et du Maréchal Davout, qui avaient été détachés une à Minsk et à Mohilow et se rejoignaient à Ochta en Witelsk, avaient toujours marché avec l'armée.



"Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813" by Charles Joseph Minard (1869)

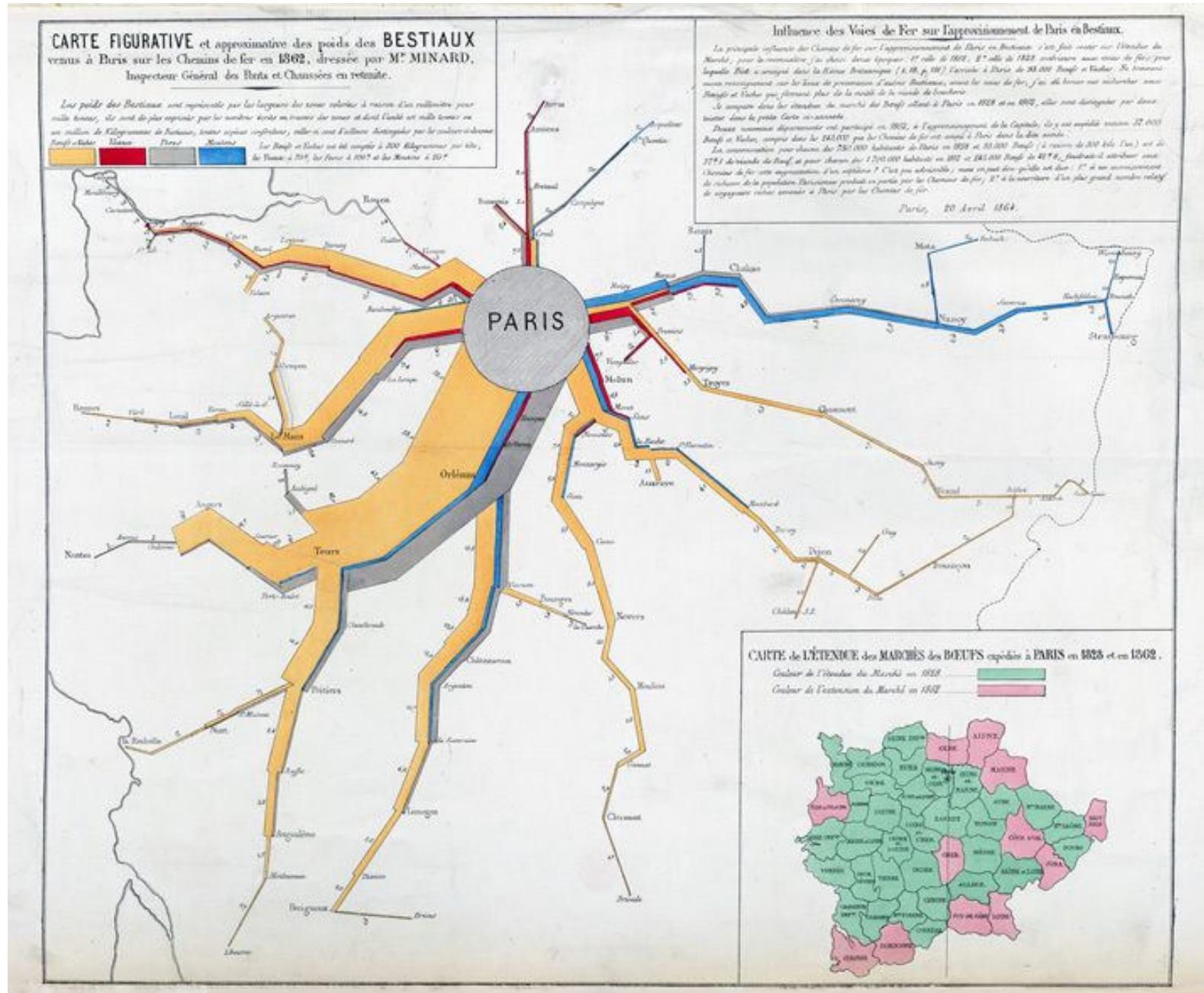




"Carte figurative des pertes successives en hommes de l'Armée qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules (selon Polybe)" (top) and "Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813" (bottom) by Charles Joseph Minard (1869)

- shows the force levels of the armies of Hannibal (218 BC) and Napoleon (1812-1813), respectively
- some data visualization practitioners call it (one of) **the best statistical drawings ever created**





"Carte figurative et approximative des poids des bestiaux venus à Paris sur les chemins de fer en 1862" by Charles Joseph Minard (1864)

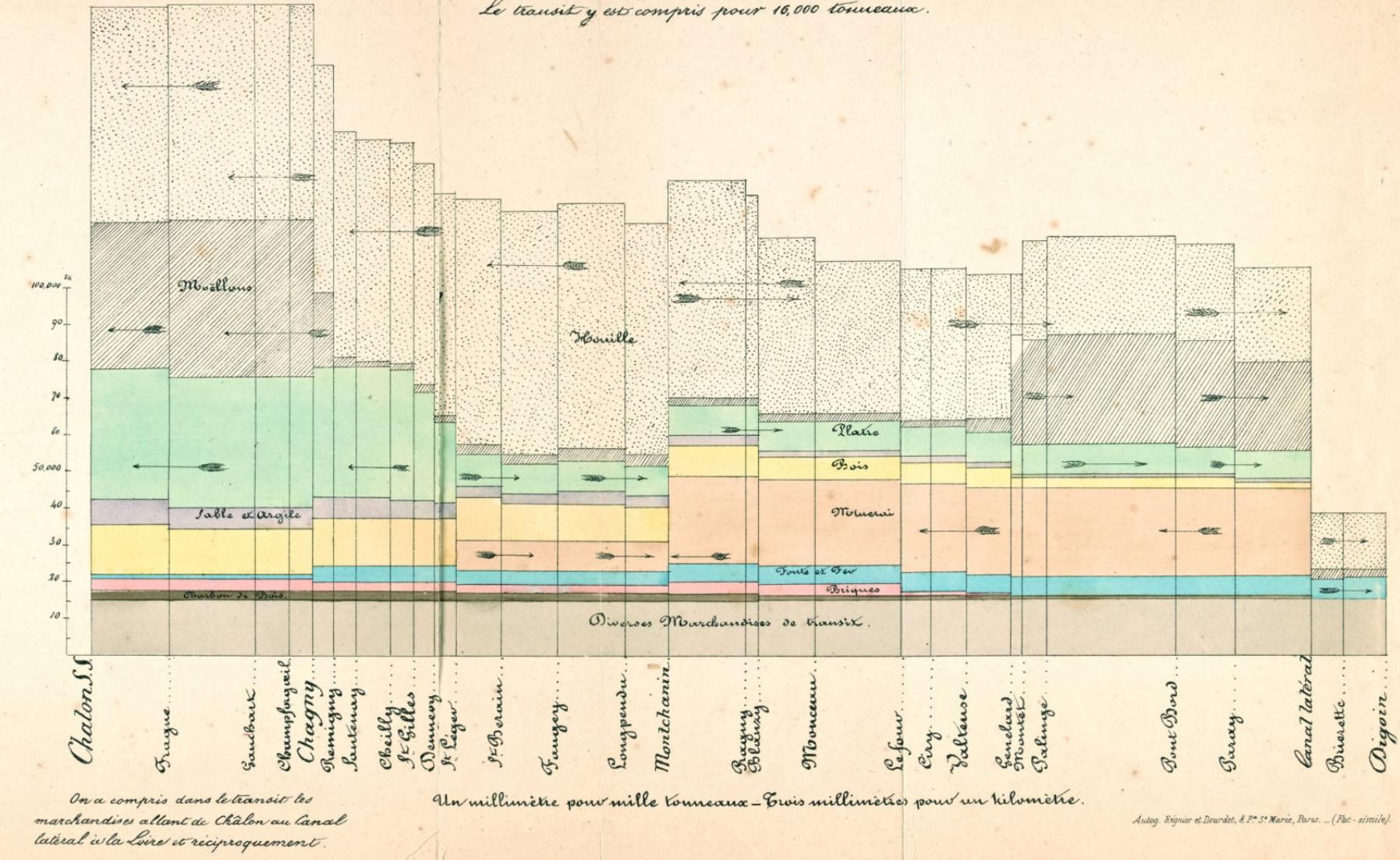


# Tableau figuratif du mouvement commercial du Canal du Centre en 1844

dressé par M<sup>e</sup> Minard sur les renseignements de M<sup>e</sup> Comoy

Le mouvement total équivaut à 131,000 tonnes parcourant la longueur du Canal ou 117 kilomètres

Le transit y est compris pour 16,000 tonnes.



"Tableau figuratif du mouvement commercial du Canal du Centre en 1844" by Charles Joseph Minard (1845)

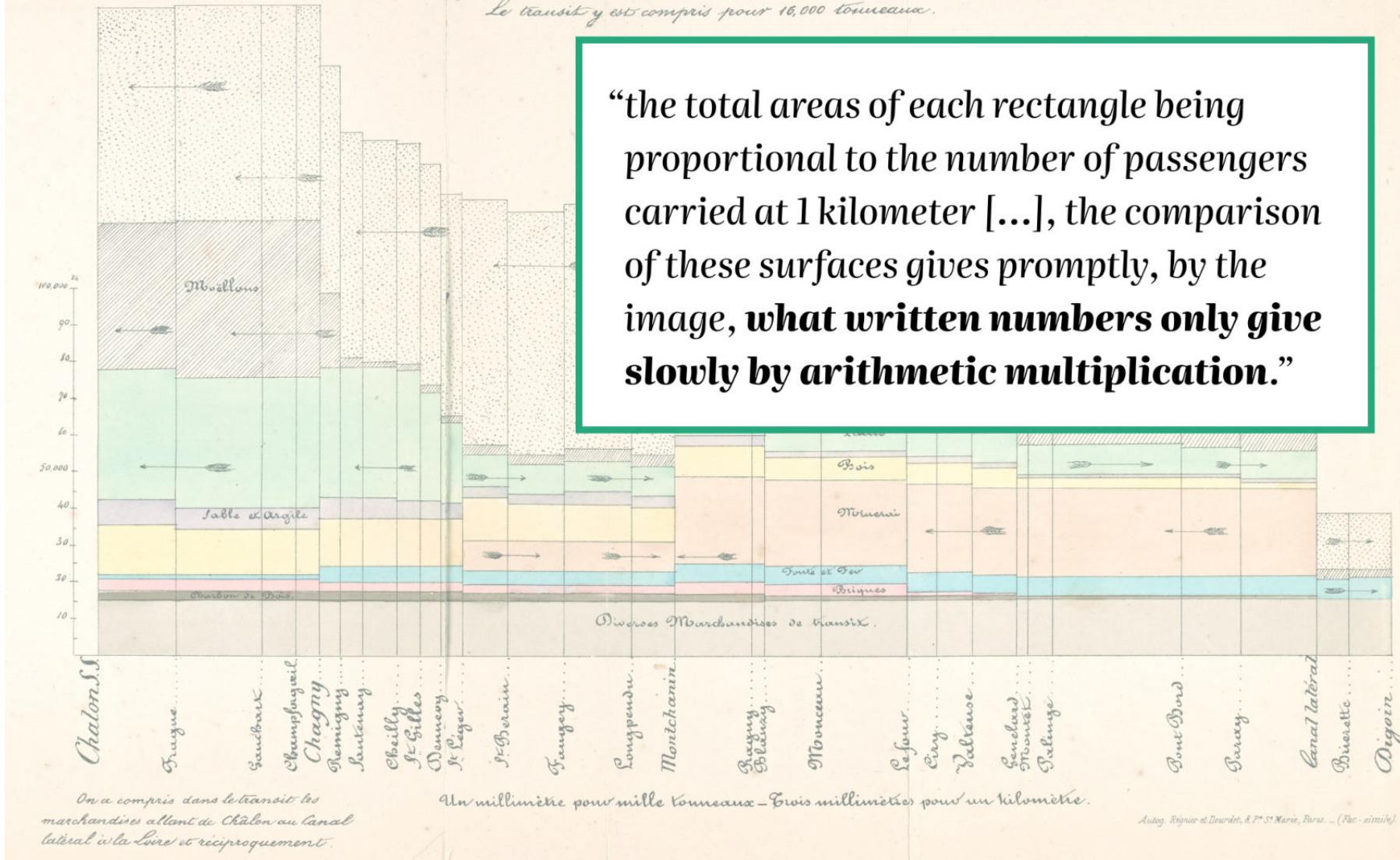


# Tableau figuratif du mouvement commercial du Canal du Centre en 1844

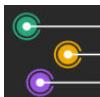
dressé par M<sup>e</sup> Minard sur les renseignements de M<sup>e</sup> Comoy

Le mouvement total équivaut à 131,000 tonnes parcourant la longueur du Canal ou 117 kilomètres

Le transit y est compris pour 16,000 tonnes.



“Tableau figuratif du mouvement commercial du Canal du Centre en 1844” by Charles Joseph Minard (1845)



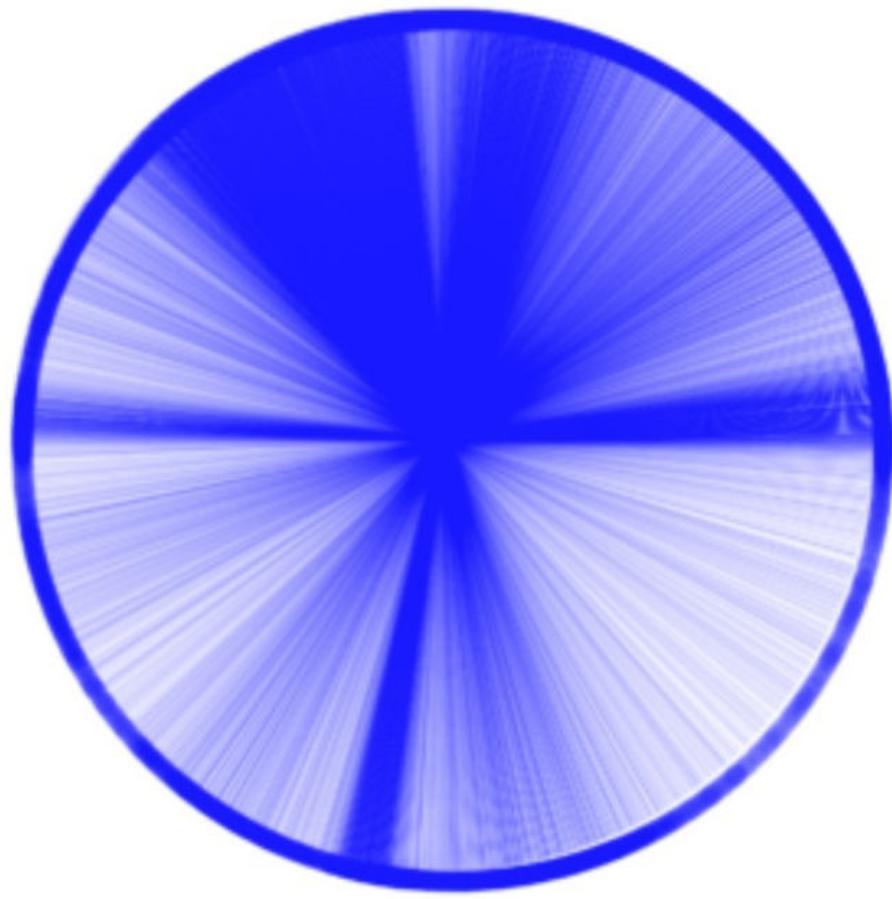
# Exercise



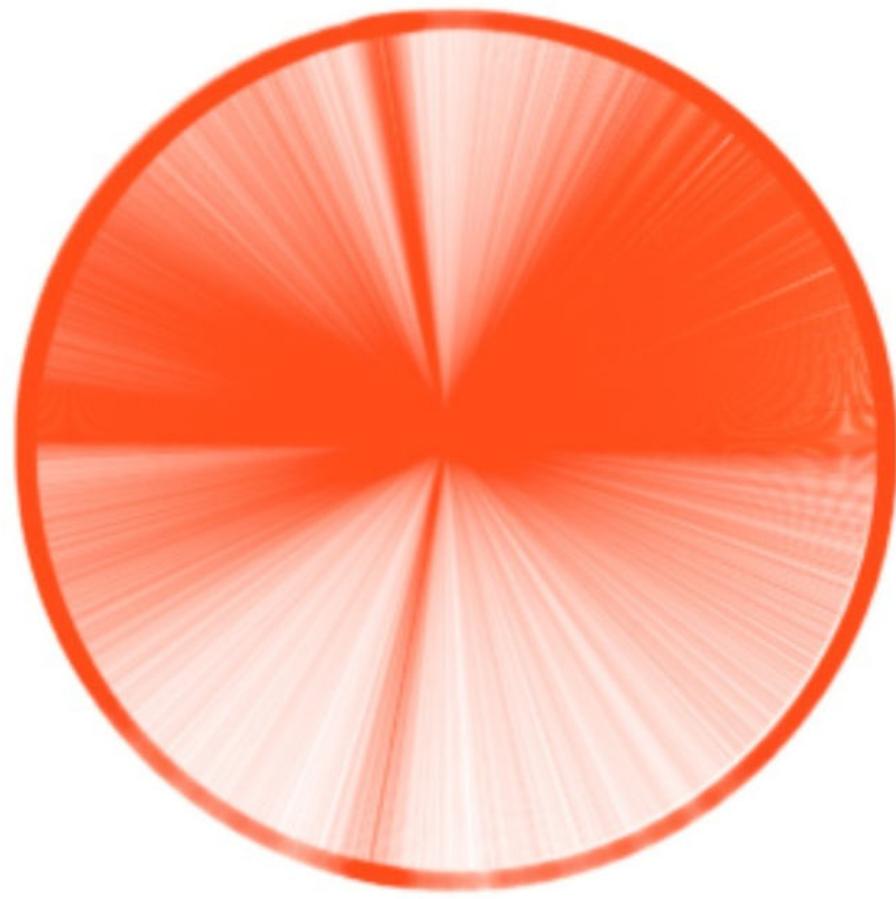
# *If the year is a circle—where's March and December in your mind?*

- Imagine you had to create a polar representation of the months.
- Draw a circle and indicate the position of March and December.
- Use an arrow to illustrate the direction of time.





# December

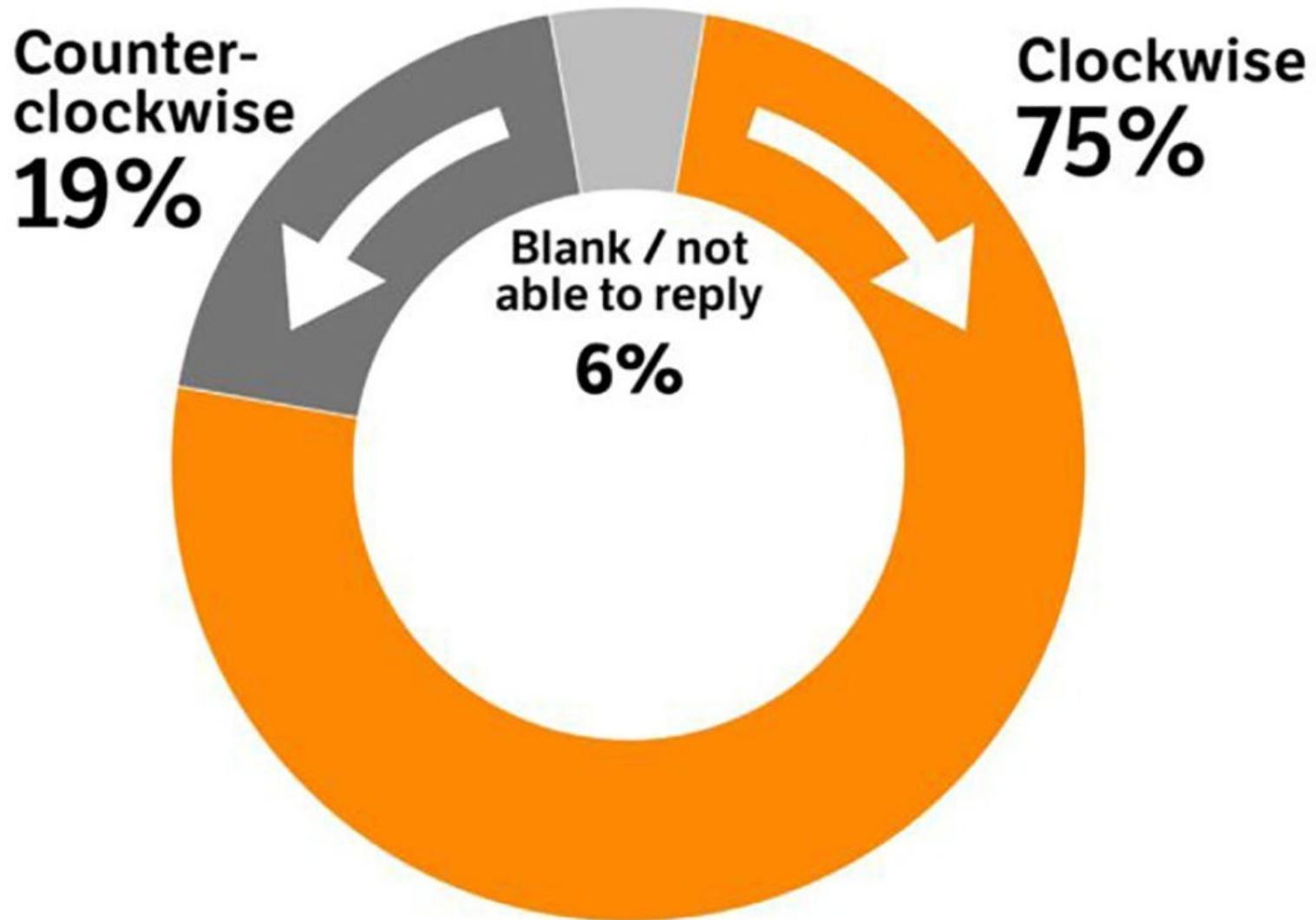


# March

Wheel diagram of 76,922 placements of the months of December and March on the circumference of an empty circle.

Graphics by Henrik Lied at NRKbeta. [Laeng & Hofseth, \*Front Psychol.\* 2019](#)



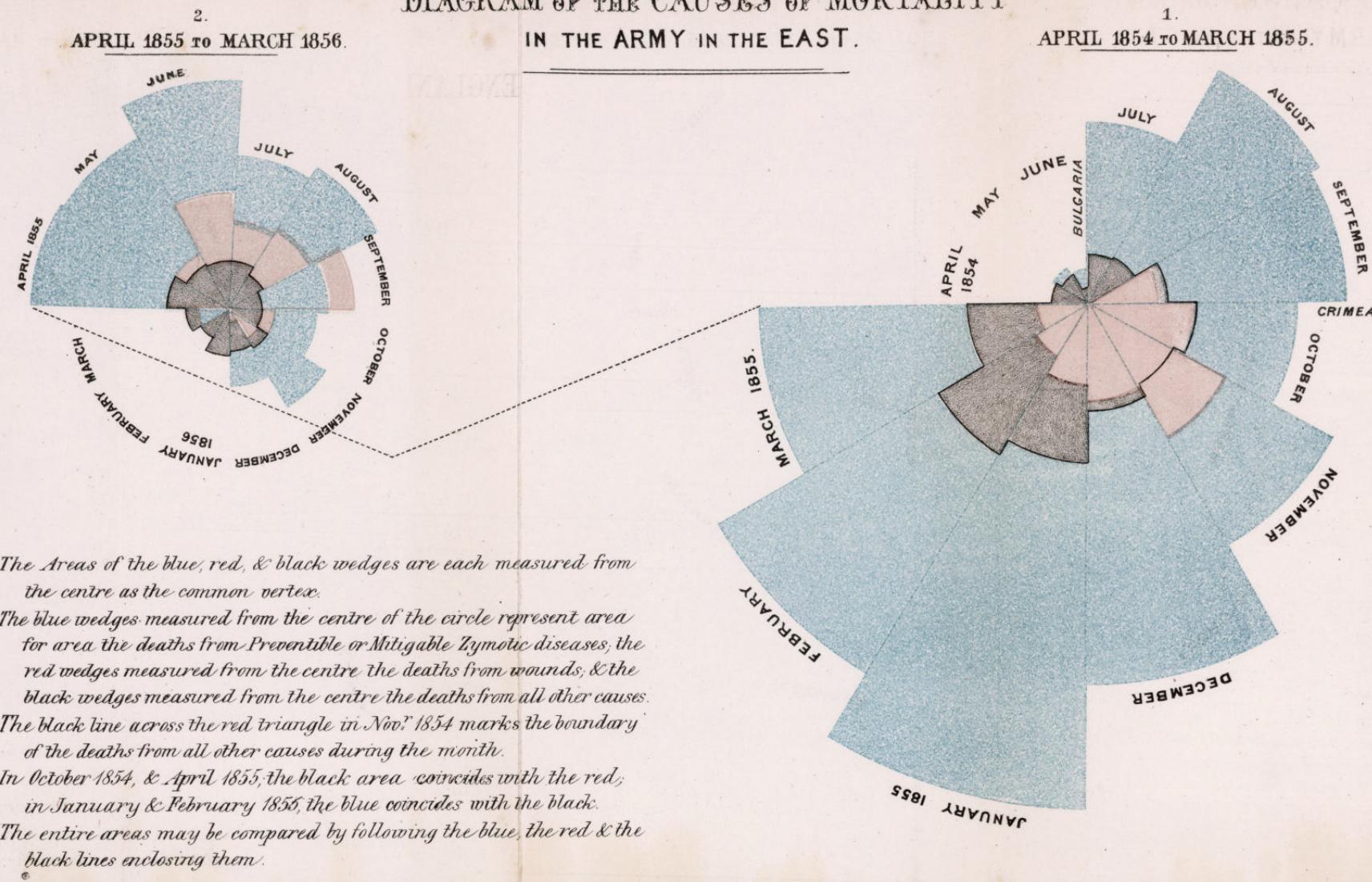


Proportion of respondents choosing opposite direction of time on the year's wheel.

Graphics by Vidar Kvien, NRK. [Laeng & Hofseth, \*Front Psychol.\* 2019](#)



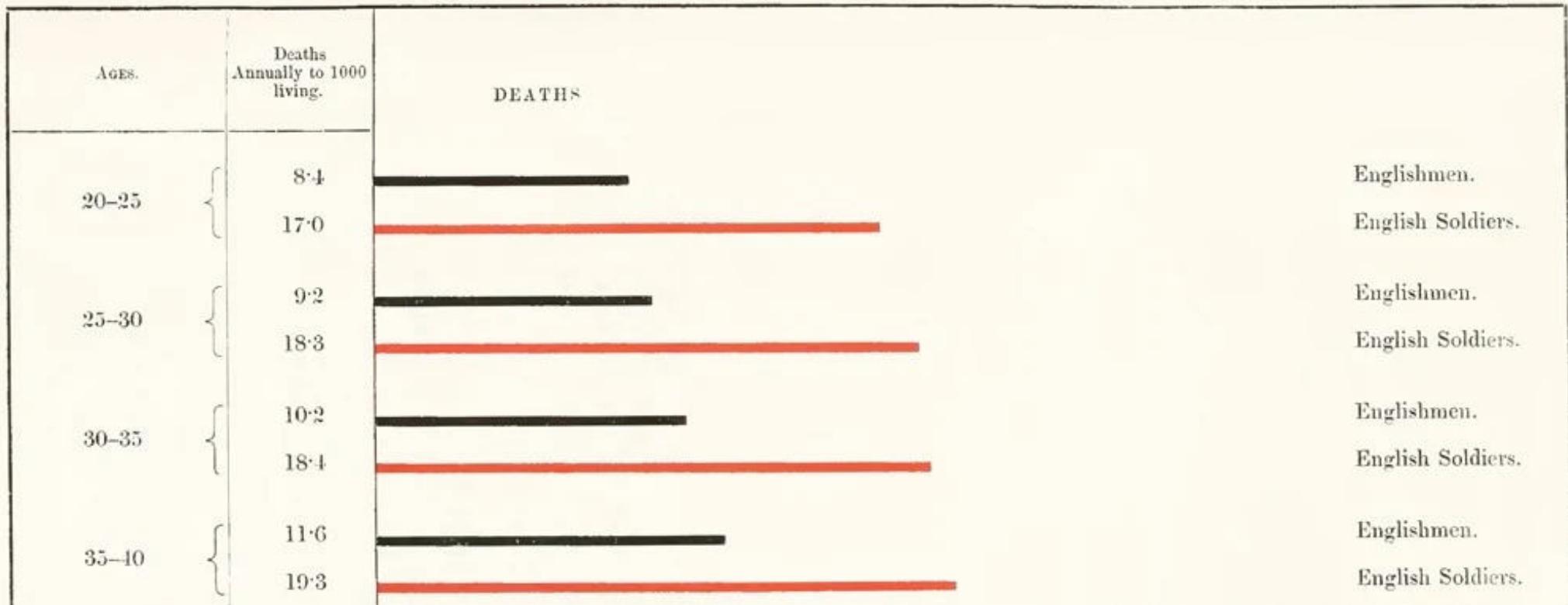
DIAGRAM OF THE CAUSES OF MORTALITY  
IN THE ARMY IN THE EAST.



“Diagram of the causes and mortality in the army in the East” (a so-called coxcomb diagram) by Florence Nightingale (1858)



*Representing the Relative Mortality of the Army at Home and of the English Male Population at corresponding Ages.*



JAMES LEWIS, del.

"Relative mortality of the army at home and of the English male population at corresponding ages" by Florence Nightingale (1858)



# Visualize Your Data



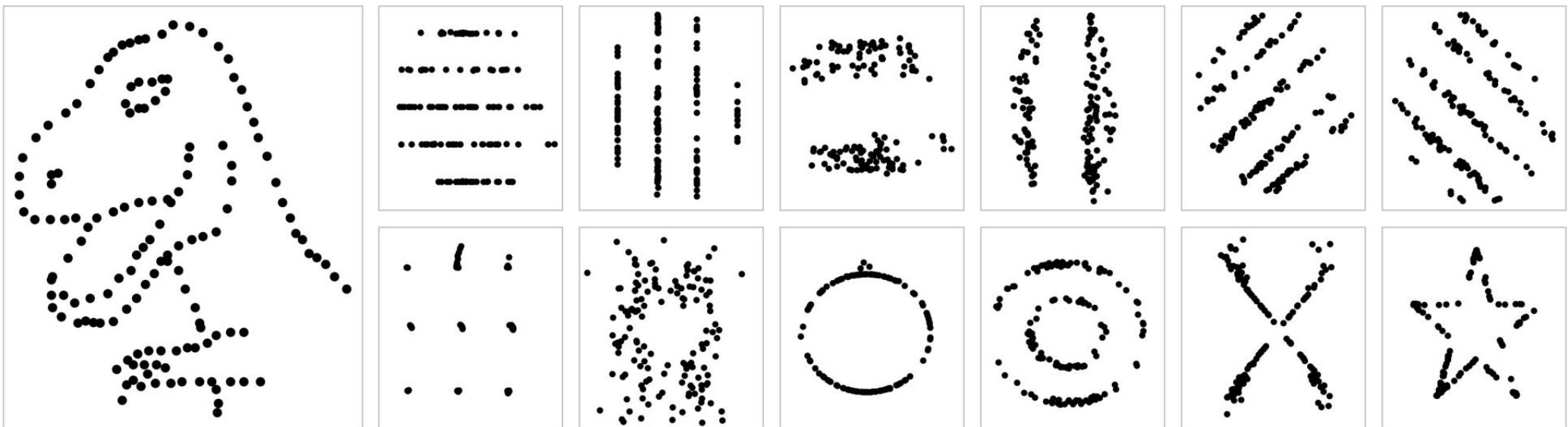
*“... make both **calculations** and **graphs**.  
Both sorts of output should be studied;  
each will **contribute** to **understanding**.”*

F. J. Anscombe (1973)



# The Datasaurus (*Anscombosaurus spec.*) Dozen

is a set of 13 different datasets with nigh-identical summary statistics, which could lead one to believe the datasets are quite similar. After visualizing the data, it becomes clear that the datasets are markedly different.

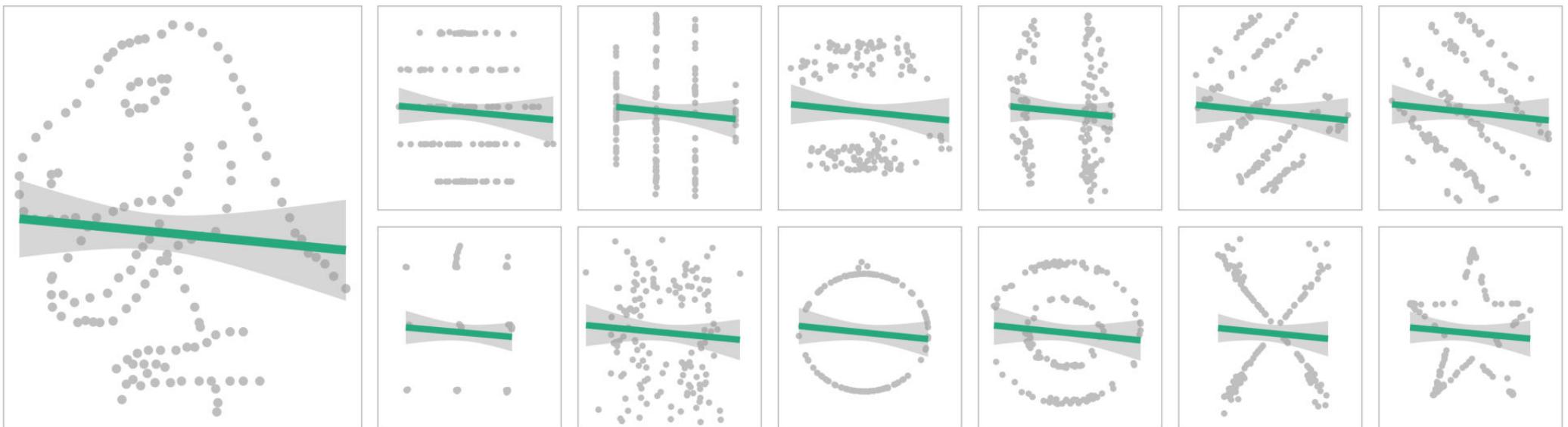


Source: "Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing" by Justin Matejka & George Fitzmaurice (2017)



# The Datasaurus (*Anscombosaurus spec.*) Dozen

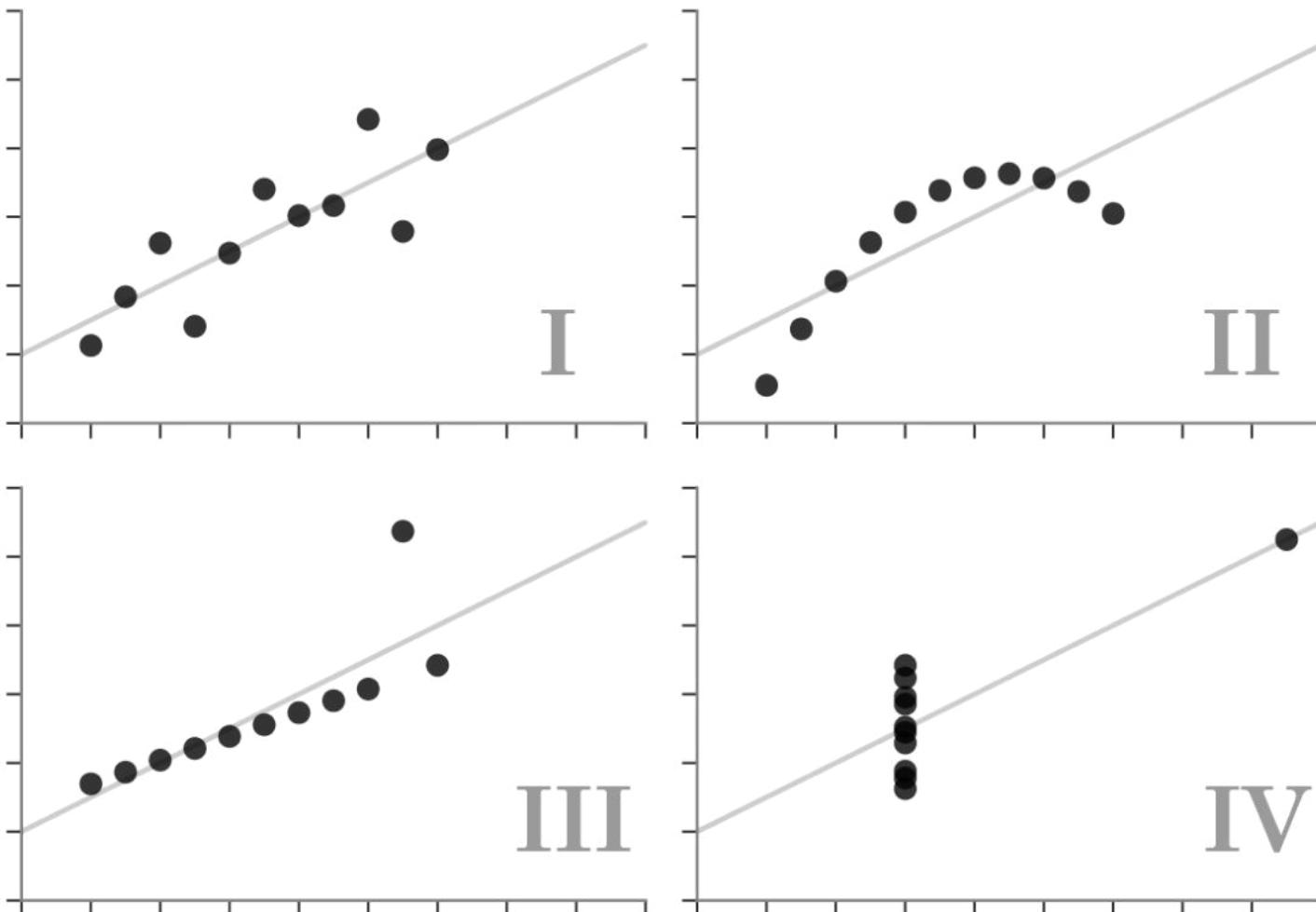
is a set of 13 different datasets with nigh-identical summary statistics, which could lead one to believe the datasets are quite similar. After visualizing the data, it becomes clear that the datasets are markedly different.



Source: "Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing" by Justin Matejka & George Fitzmaurice (2017)



# Anscombe's Quartet

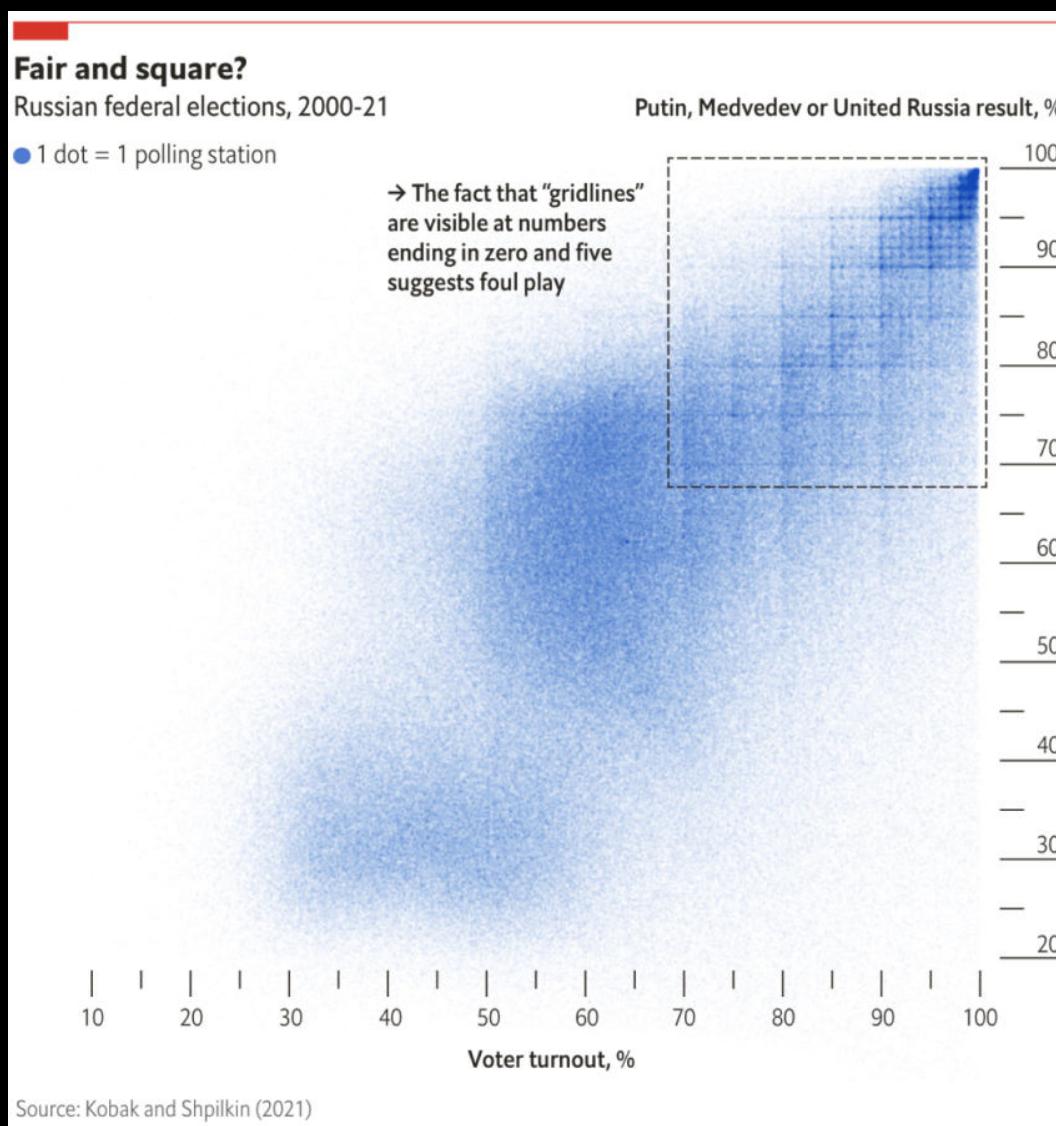


Source: Matejka & Fitzmaurice (2017)

Cédric Scherer @ Hello Heart // Data Visualization & Information Design



# Visualize Your Data!

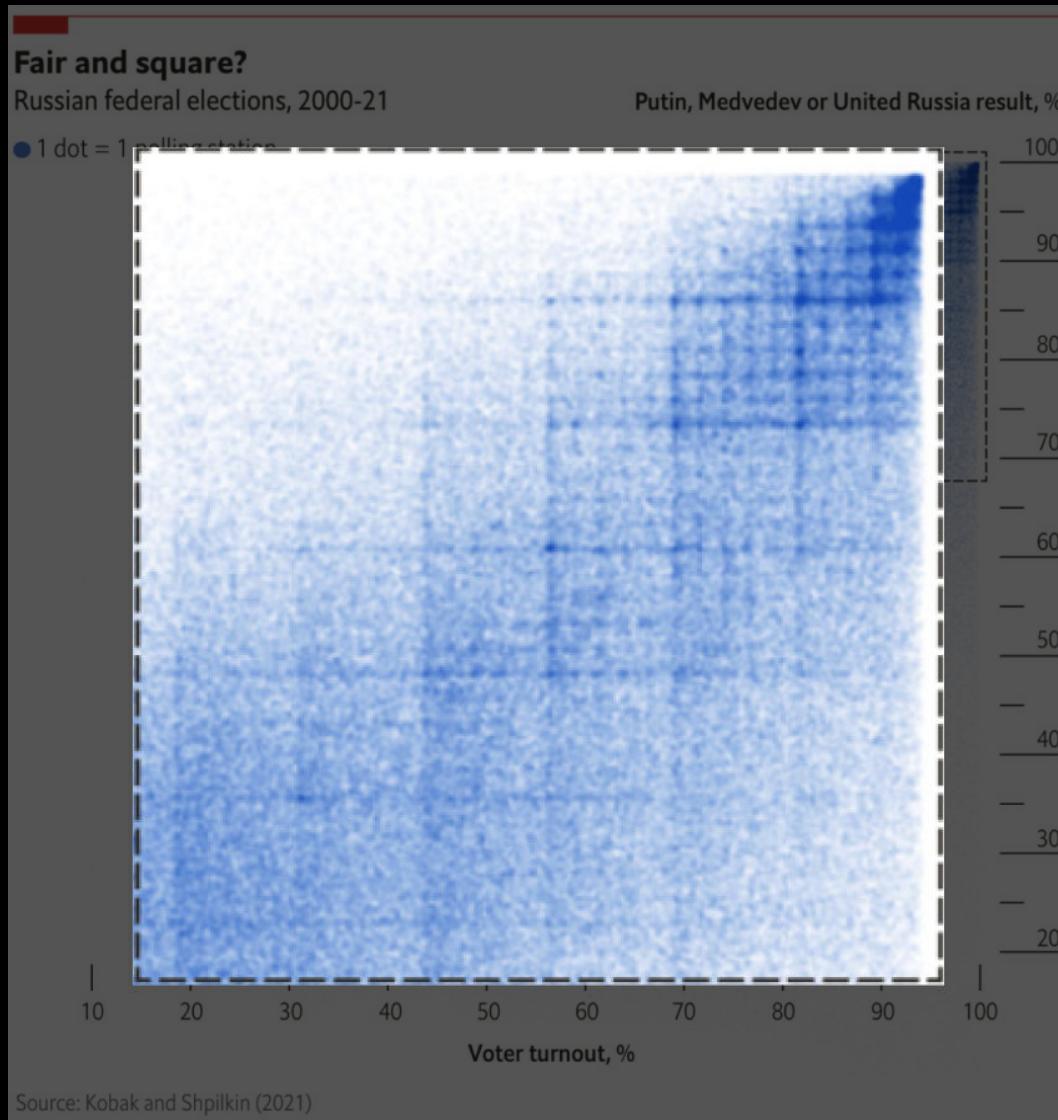


“When Dmitry Kobak and Sergey Shpilkin [...] analysed the results, they found that *an unusually high number of turnout and vote-share results were multiples of five* (eg, 50%, 55%, 60%), a tell-tale sign of manipulation.”

“Russian elections once again had a suspiciously neat result” by The Economist



# Visualize Your Data!



“When Dmitry Kobak and Sergey Shpilkin [...] analysed the results, they found that *an unusually high number of turnout and vote-share results were multiples of five* (eg, 50%, 55%, 60%), a tell-tale sign of manipulation.”

“Russian elections once again had a suspiciously neat result” by The Economist



# *A good data visualization can mean the difference between **success** and **failure**.*

- Communicating complex findings and phenomena
- Raising money for an organization, event or department
- Presenting at a board or conference
- Helping businesses and institutions to make informed decisions
- Providing guidance for improvement
- ...
- **Getting your point across!**





# Good vs Bad



# What Makes a Good Data Visualization?

- **Integrity** (information)
- **Story** (interestingness)
- **Goal** (usefulness)
- **Visual Form** (beauty)



Dashboards are  
*visual displays*



Dashboards are  
*visual displays*  
of the  
***most important information needed***



Dashboards are  
*visual displays*  
of the  
**most important information needed**  
to achieve  
**one or more objectives**



Dashboards are  
*visual displays*  
of the  
**most important information needed**  
to achieve  
**one or more objectives**  
that has been  
**consolidated on a single screen**



Dashboards are  
*visual displays*  
of the  
**most important information needed**  
to achieve  
**one or more objectives**  
that has been  
**consolidated on a single screen**  
so it can be  
**monitored at a glance.**



# Visual Displays

- combination of text, tables, and graphics
- preferably emphasis on graphical elements
  - more efficient
  - more comprehensive

→ important to **understand visual perception**



# Most Important Information Needed

- understand to whom you are communicating
- focus on the information required to take action
- investigate what data is available to provide the information
- suitability of the metrics should be reviewed from time to time

→ important to **know the audience**



# One or More Objectives

- collection of information
  - avoid too broad specification of objectives
  - avoid combination of too many data sources / subsets
- often defined as KPIs

→ important to **define meaningful objectives**



# Consolidated on a Single Screen

- all information needs to be entirely available to the viewer
- related insights must be grouped together
- focus on main objective

→ important to **condense information**



# Monitored at a Glance

- provide an overview of data
  - avoid unnecessary details
- guide the user
  - emphasize key outcomes
  - remove clutter
  - highlight required actions

→ important to **follow design guidelines**



# How to Design Good Dashboards

- understand visual perception
- know the audience
- define meaningful objectives
- condense information
- follow design guidelines



# How to Design Good Data Visualizations

- understand visual perception
- know the audience
- define meaningful objectives
- condense information
- follow design guidelines



# Exercise



# Exercise

- **Take a closer look at the following visualization.**
  - What is the main message you learn from the graphic?
  - Rate the graphics according to the four levels:  
*information, story, goal, and visual form.*
  - Collect three things you notice, no matter if positive or negative.
  - How could you fix the details you dislike?



# Annual Greenhouse Gas Emissions by Sector

