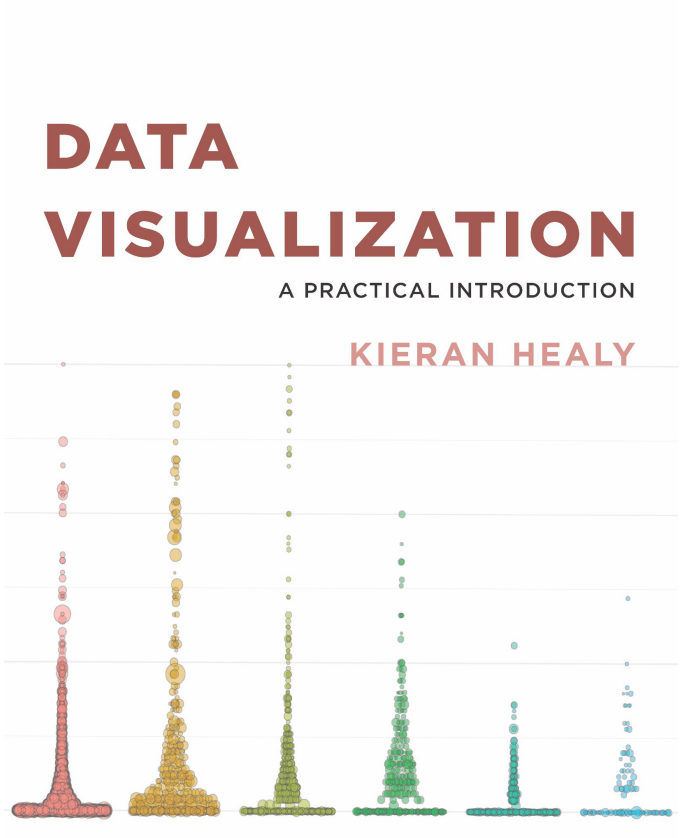
FALL 2023
DSE 12700
VISUAL ANALYTICS

Professor
Madeline Blount
she/her

Week 1

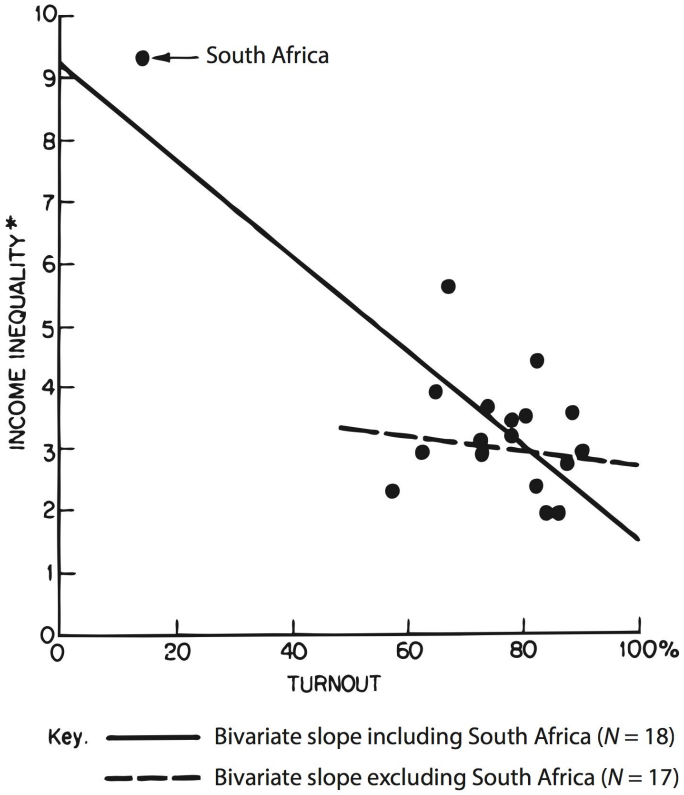
Dall-E 2: a blue jay creating neon data visualizations

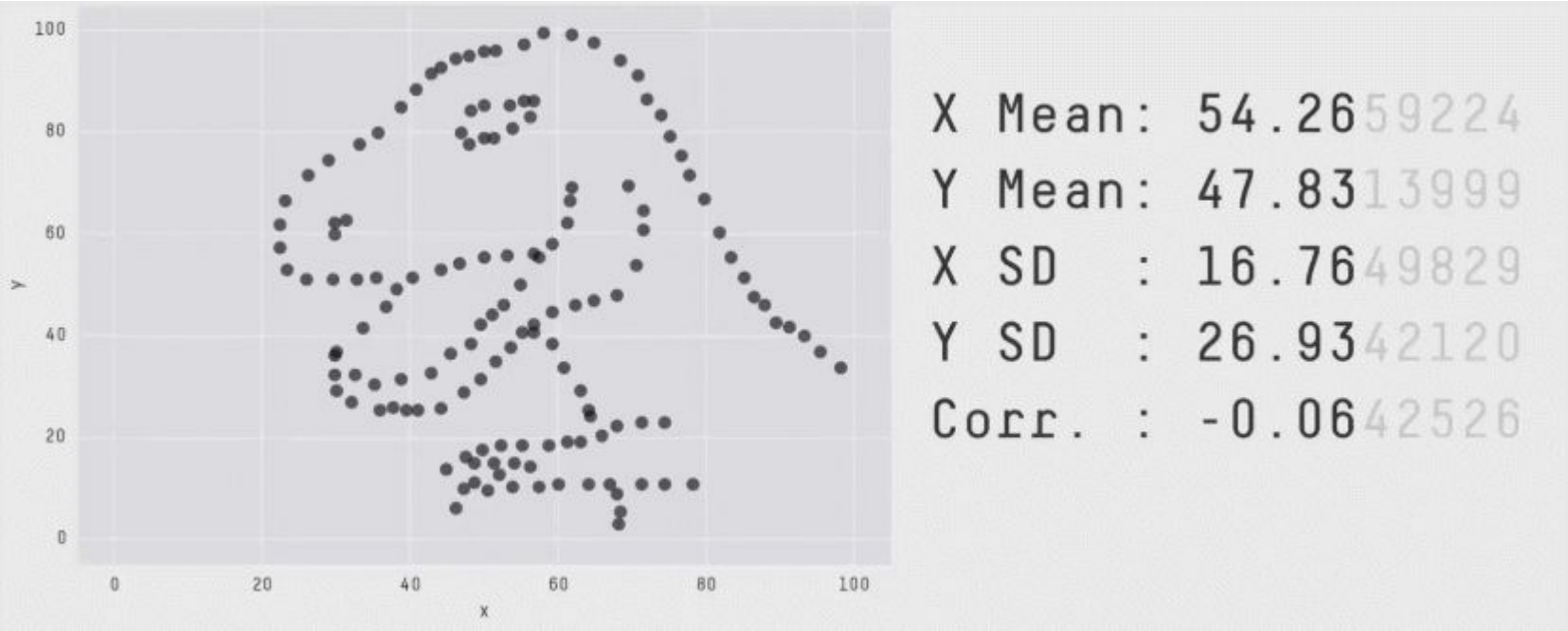
Kieran Healy
(R)



I am Professor of Sociology at Duke University. Much of my research has been about the social organization of exchange in human blood and organs, cultural goods, software, and ideas. My current work focuses on the moral order of market society, the effect of models and measurement on social classification, and the link between those two topics. I also work on techniques and methods for data visualization, and some problems in social theory.

"You should look at your data."





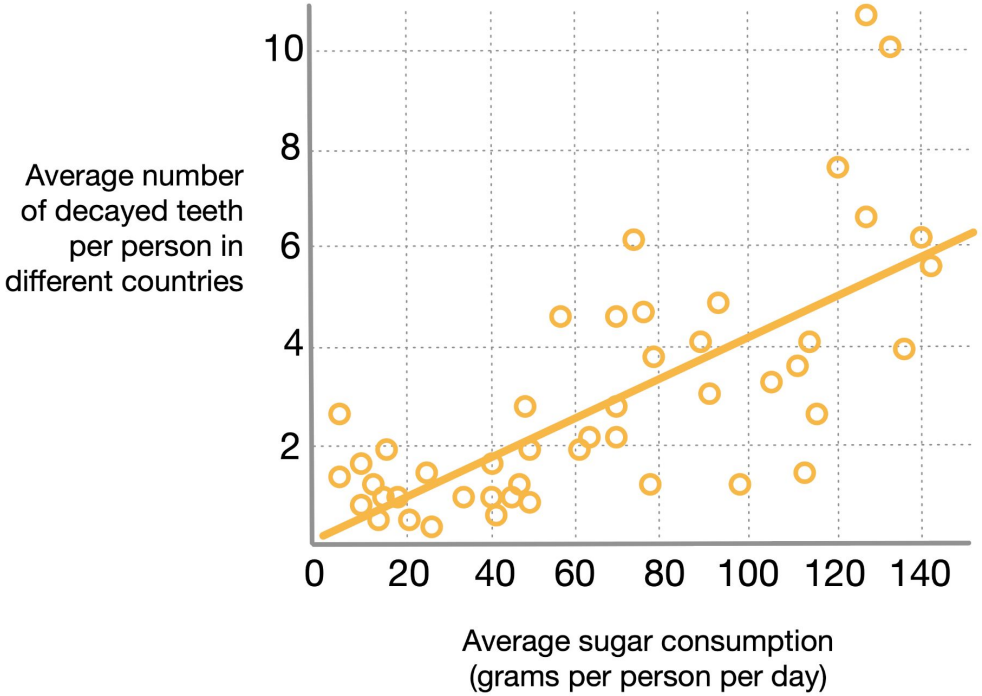
Cairo; Matejka & Fitzmaurice; found via Healy

Anscombe's quartet, revisited

“Graphics *reveal* data.” - Tufte

“[Graphics] are not some sort of magical means of seeing the world as it really is.” - Healy

Which of the following statements best describes the data in the graph below?



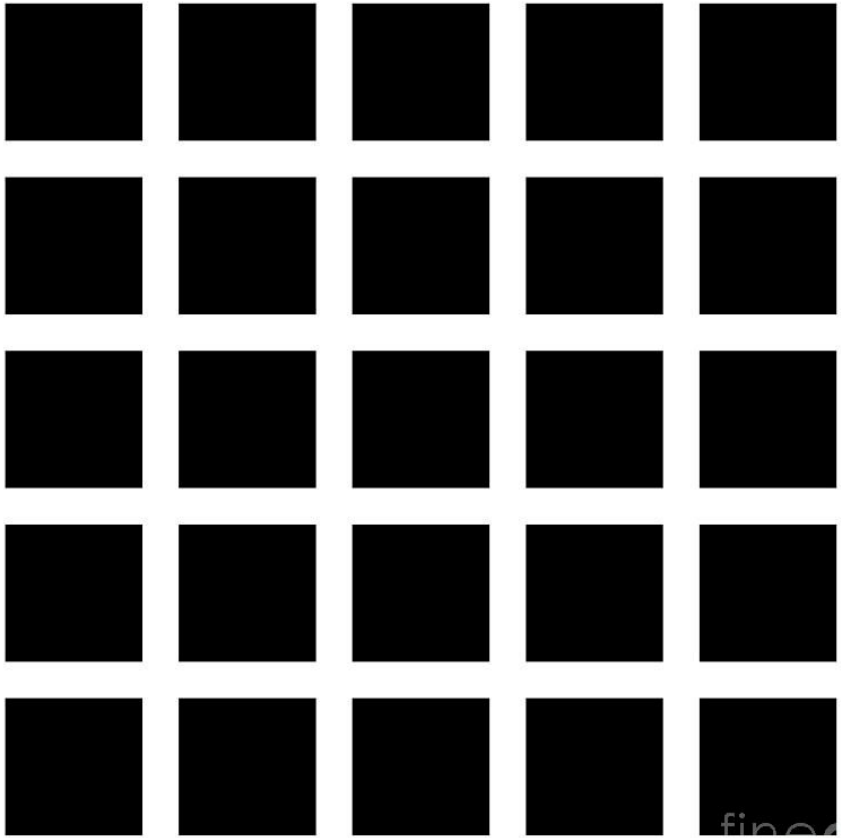
A. In recent years, the rate of cavities has increased in many countries
B. In some countries, people brush their teeth more frequently than in other countries
C. The more sugar people eat, the more likely they are to get cavities
D. In recent years, the consumption of sugar has increased in many countries

Pew Research; found via Healy

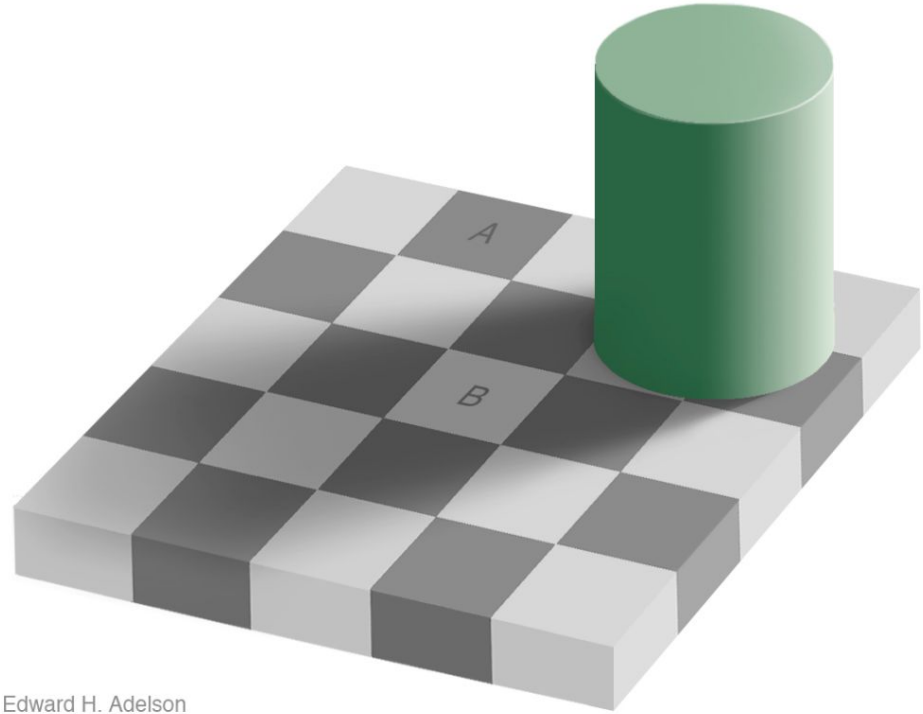
Hermann grid effect, 1870

Perception as **active**

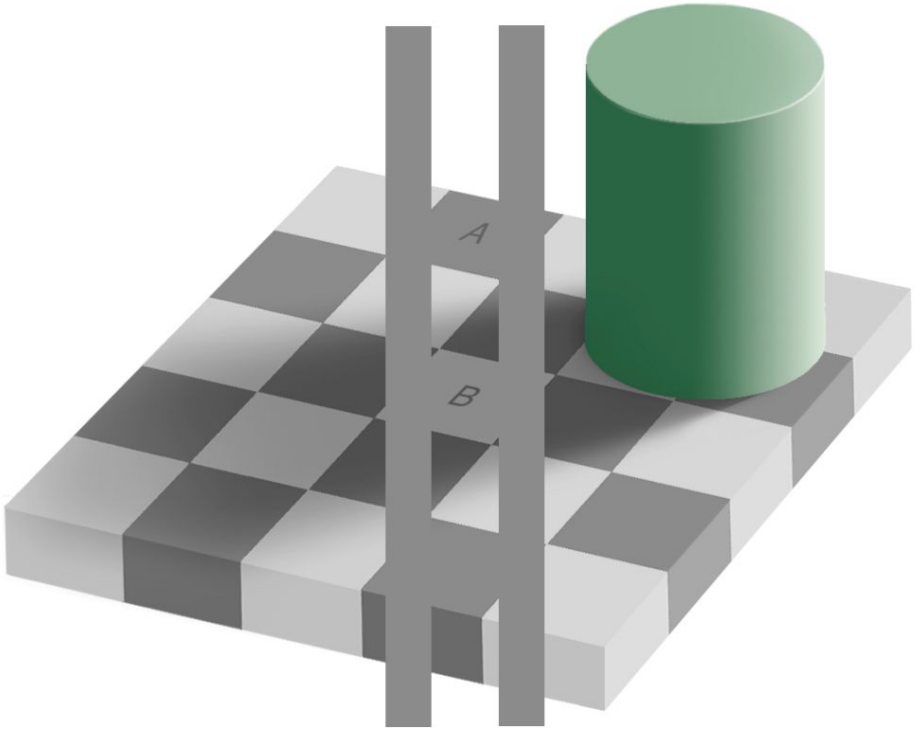
Vision drawn to the edges



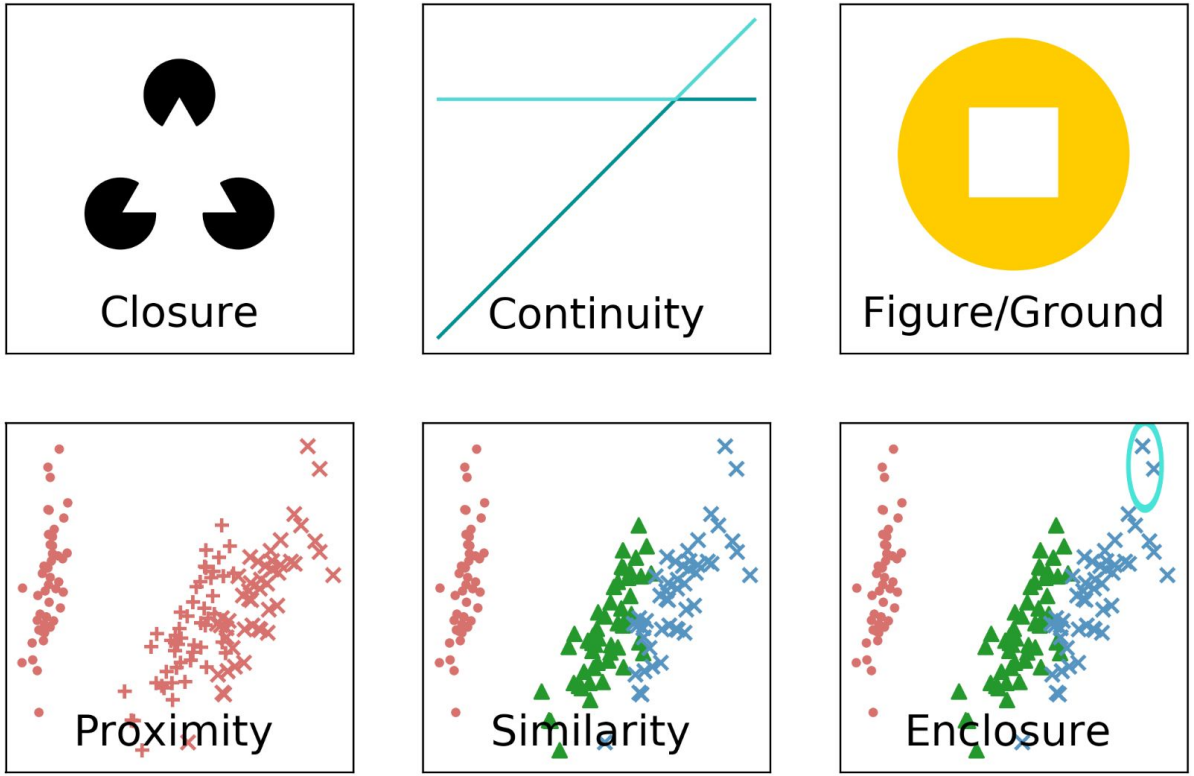
Healy



Edward H. Adelson

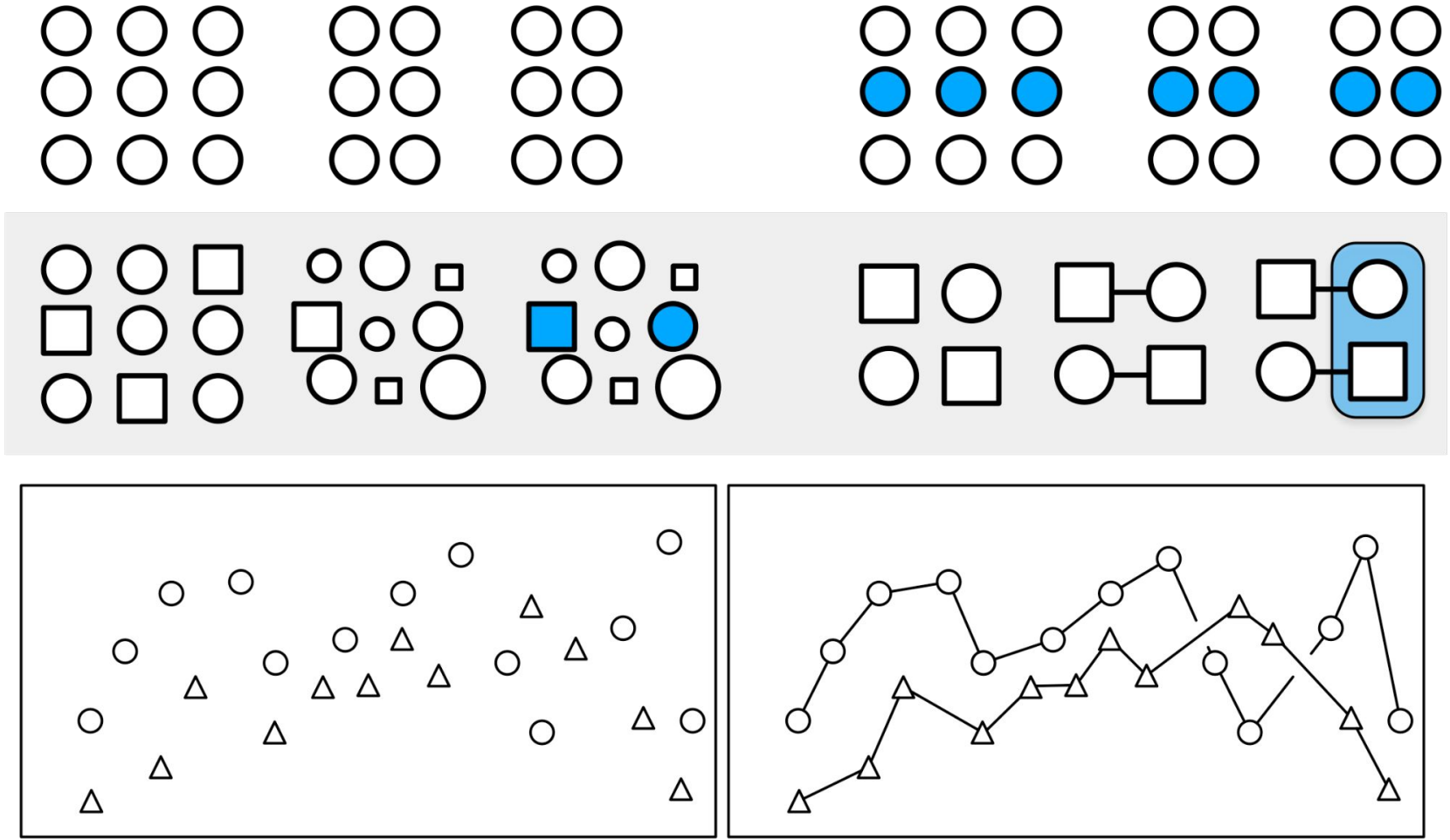


Healy



GESTALT PRINCIPLES: whole *other* than parts

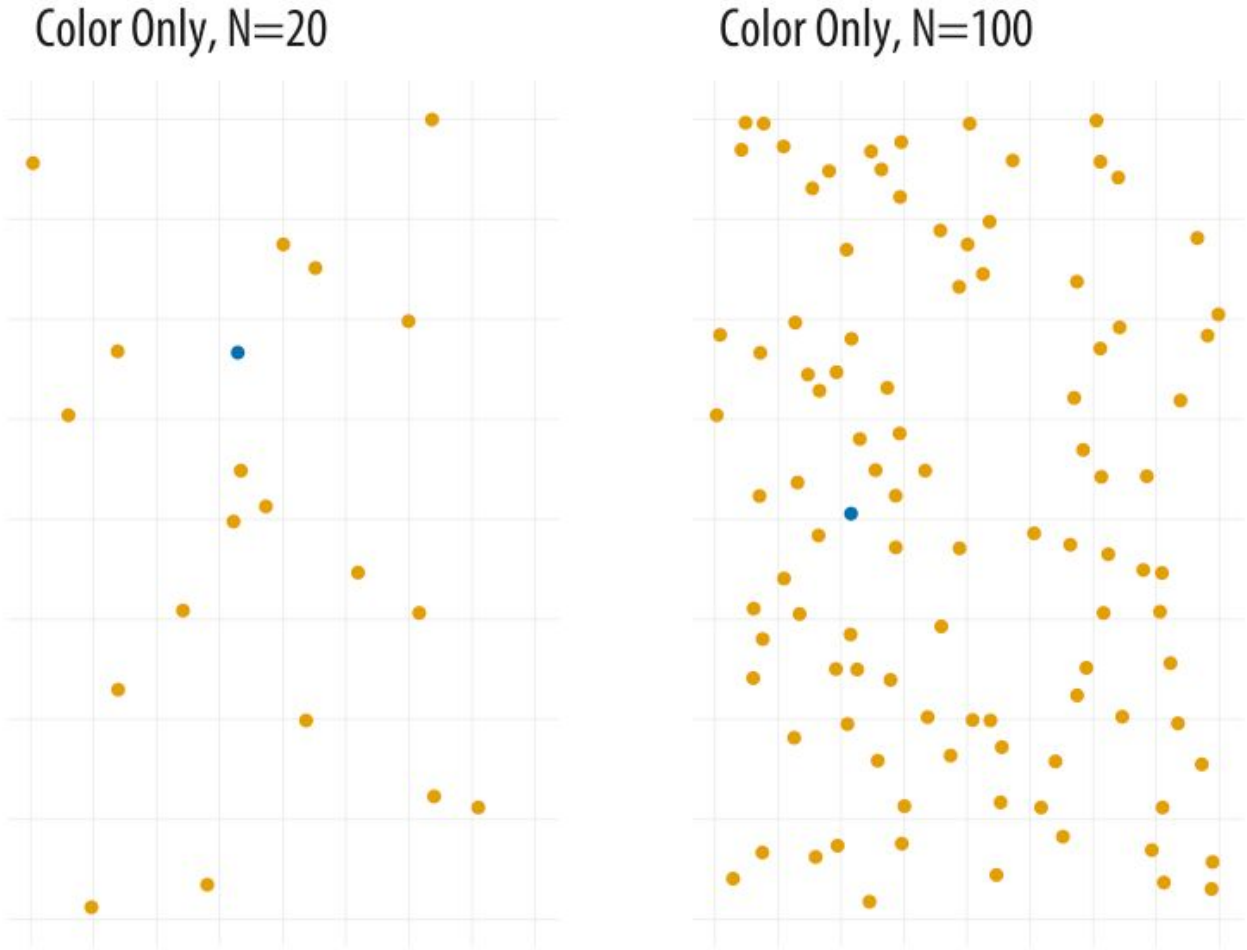
O'Reilly



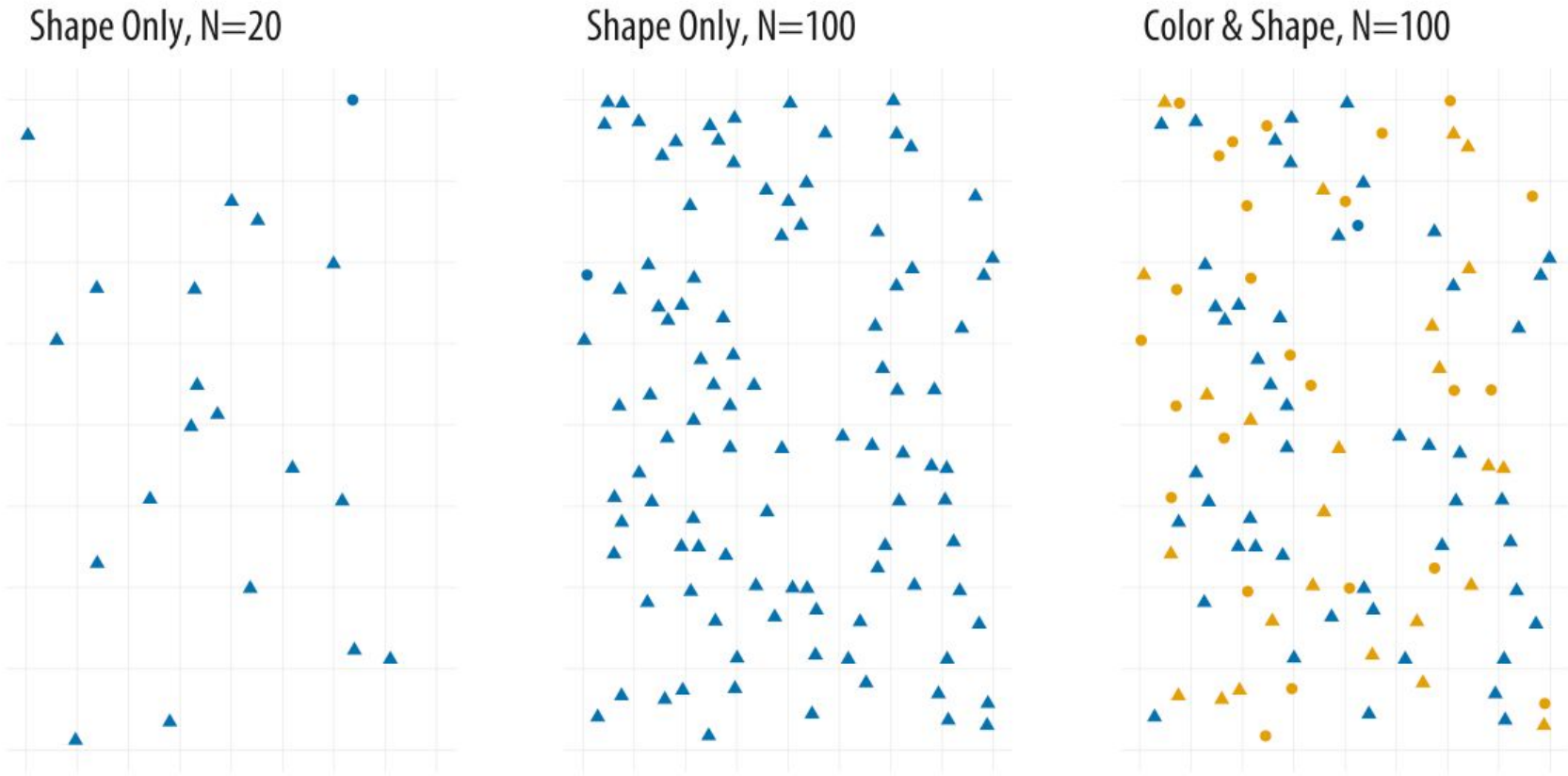
Healy

PRE-ATTENTIVE POP-OUT

Demographic data, France from Bertin; found in Fry



Healy



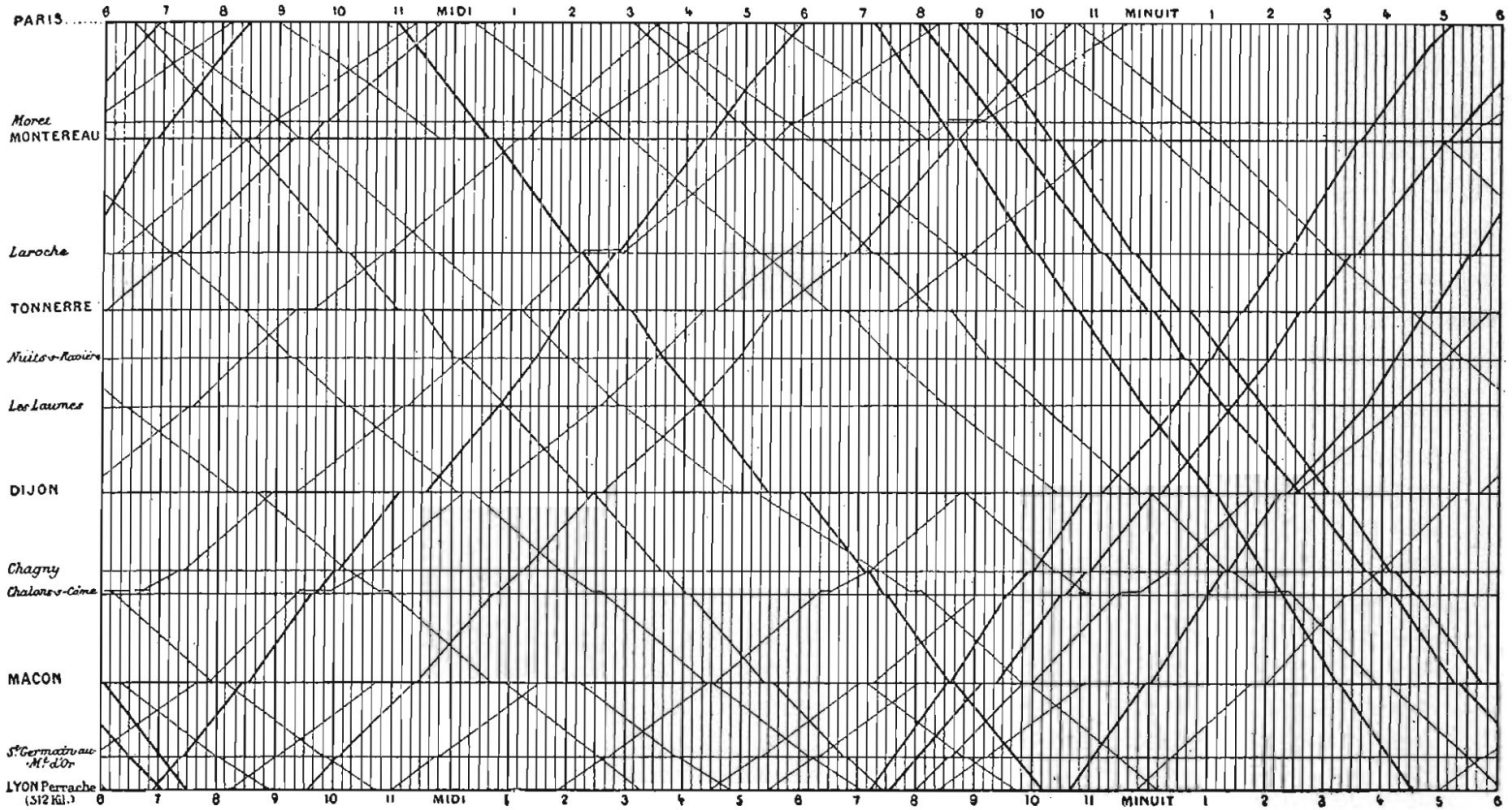
CHANNELS

Healy

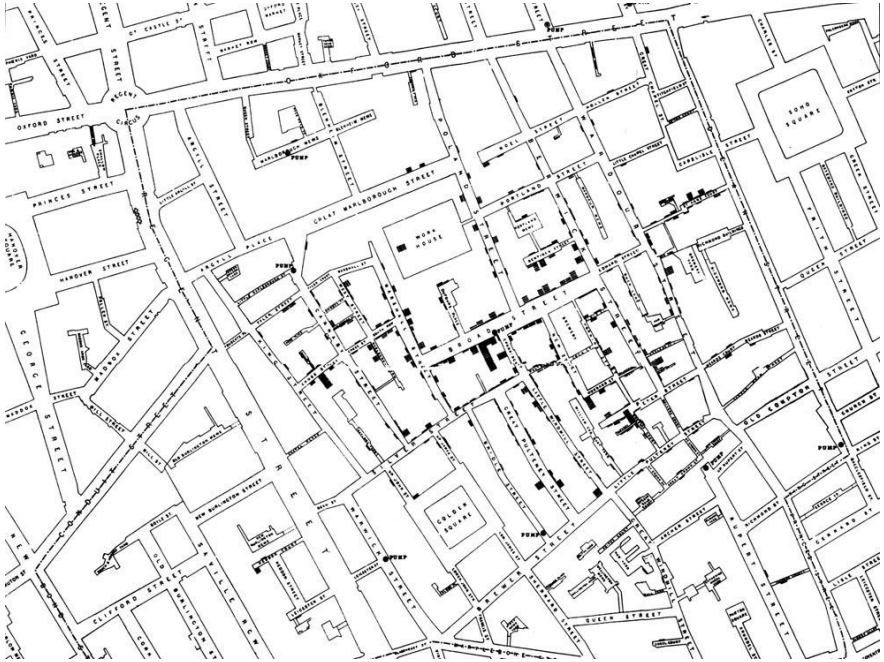

THE ART OF
DATA VISUALIZATION

Tufte: “the maps repay careful study.”

Is **SIMPLICITY** a “virtue” of graphic excellence?



E. J. Marey, *La méthode graphique* (Paris, 1885), p. 20. The method is attributed to the French engineer, Ibry.



ORIGINAL John Snow: https://wellcomecollection.org/works/uxgfjt62/items

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. 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 puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit 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 Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

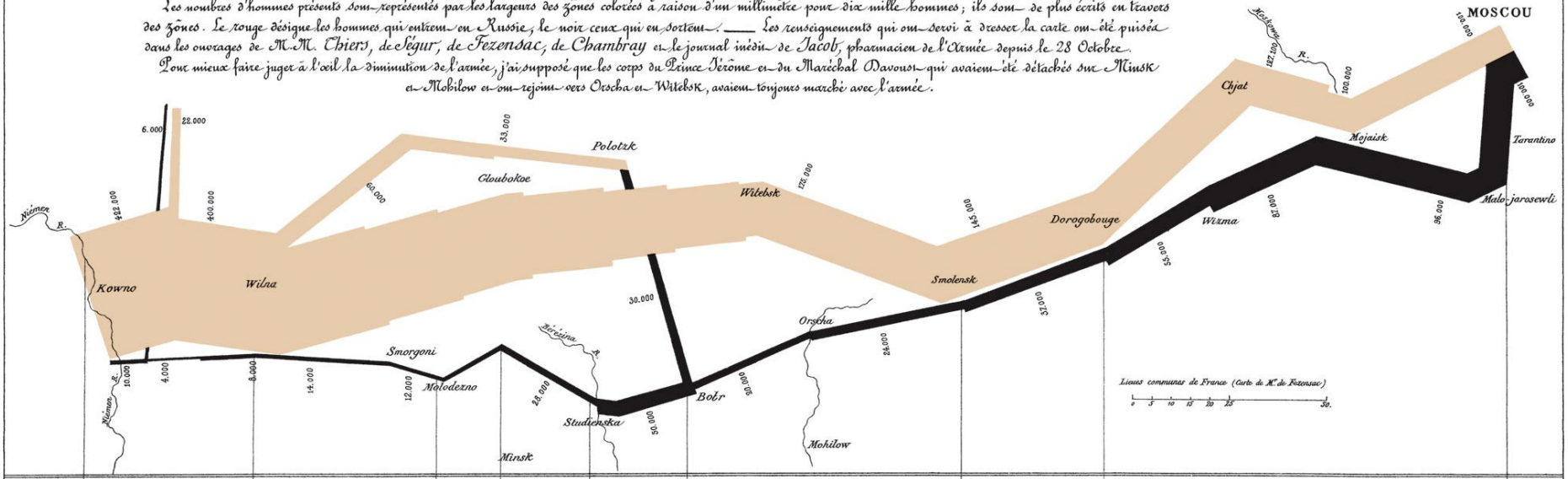
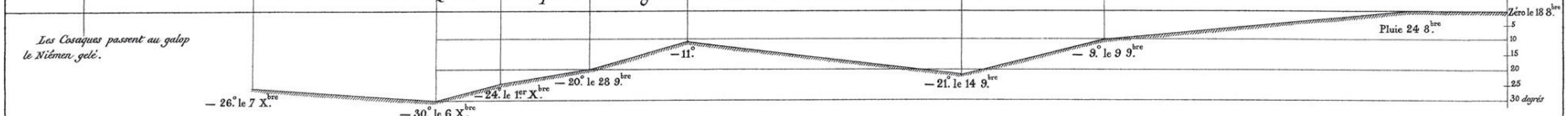


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



Autog. par Regnier, 8. Pas. Ste Marie St Gloire à Paris.

Imp. Lith. Regnier et Dourdet.

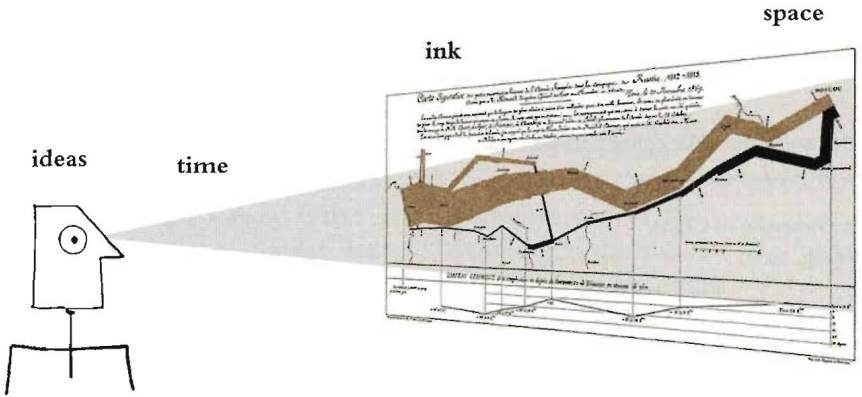

BEST CHART
EVER MADE?

Principles of Graphical Excellence

Graphical excellence is the well-designed presentation of interesting data—a matter of *substance*, of *statistics*, and of *design*.

Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency.

Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.



Graphical excellence is nearly always multivariate.

And graphical excellence requires telling the truth about the data.

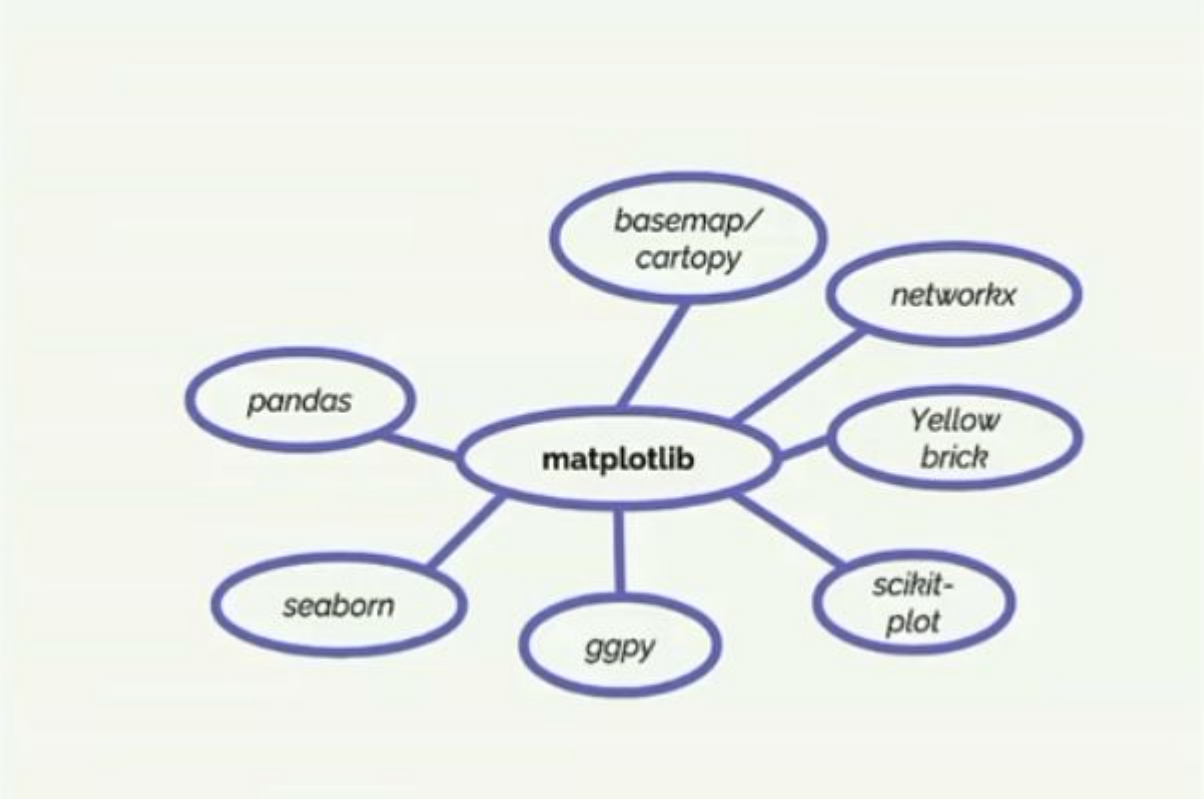
Python!

- 1989 🐍
- Netherlands, Guido van Rossum “BDFL” until 2018
- High-level programming language
- Indentations + spacing matter very much!
- *pythonic*

Zen of Python:

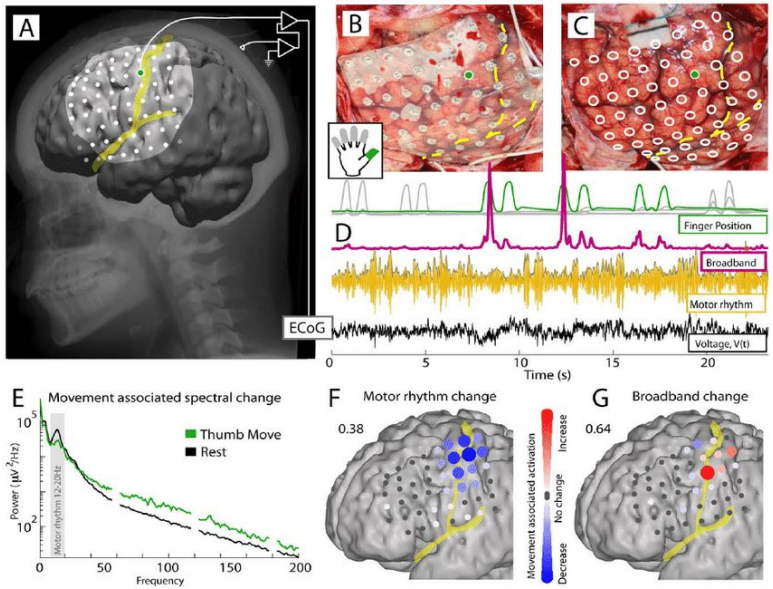
- Beautiful is better than ugly.
- Explicit is better than implicit.
- Simple is better than complex.
- Complex is better than complicated.
- Flat is better than nested.
- Sparse is better than dense.
- Readability counts.
- Special cases aren't special enough to break the rules.
- Although practicality beats purity.
- Errors should never pass silently.
- Unless explicitly silenced.
- In the face of ambiguity, refuse the temptation to guess.

…


basemap/
cartopy
networkx
pandas
matplotlib
Yellow
brick
seaborn
ggpy
scikit-
plot

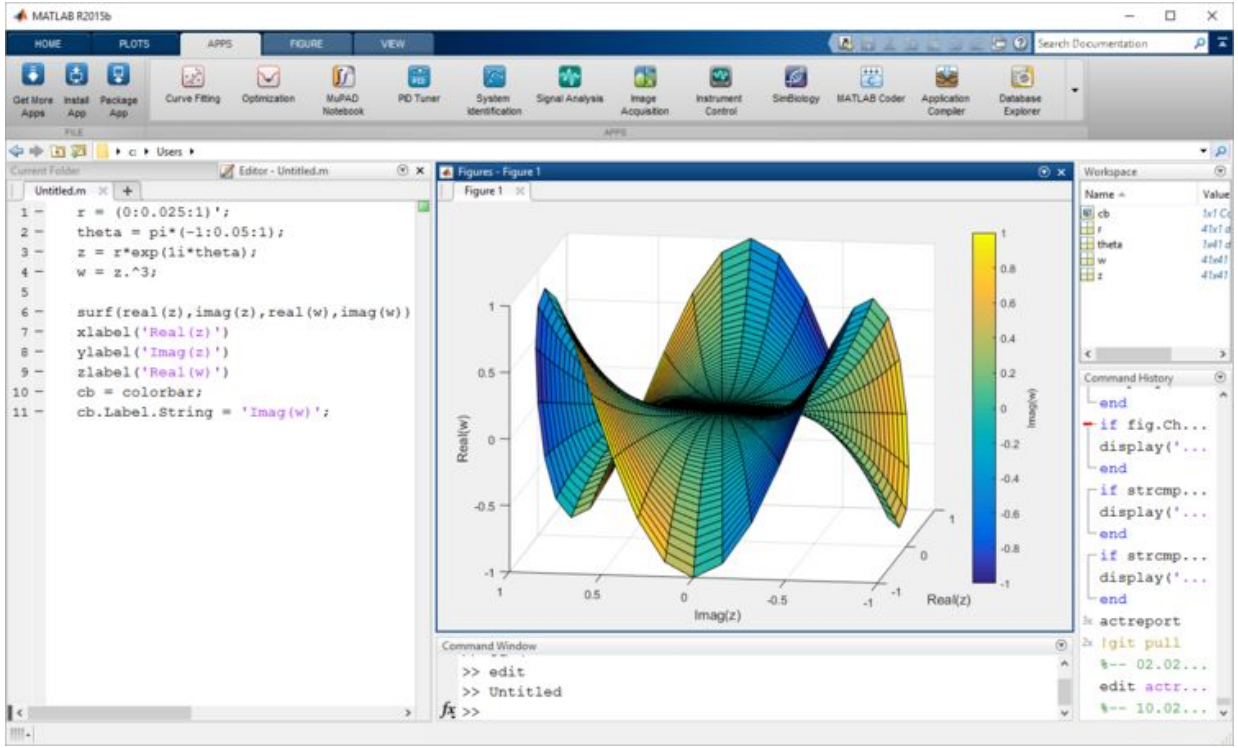
matplotlib:

- Python *library* - bundle of pre-written code, modular
- Created by John D. Hunter, originally for electrocorticography (ECoG) for PhD in neurobiology, 2003



matplotlib:

- Designed to look + feel like MATLAB - but *open source*

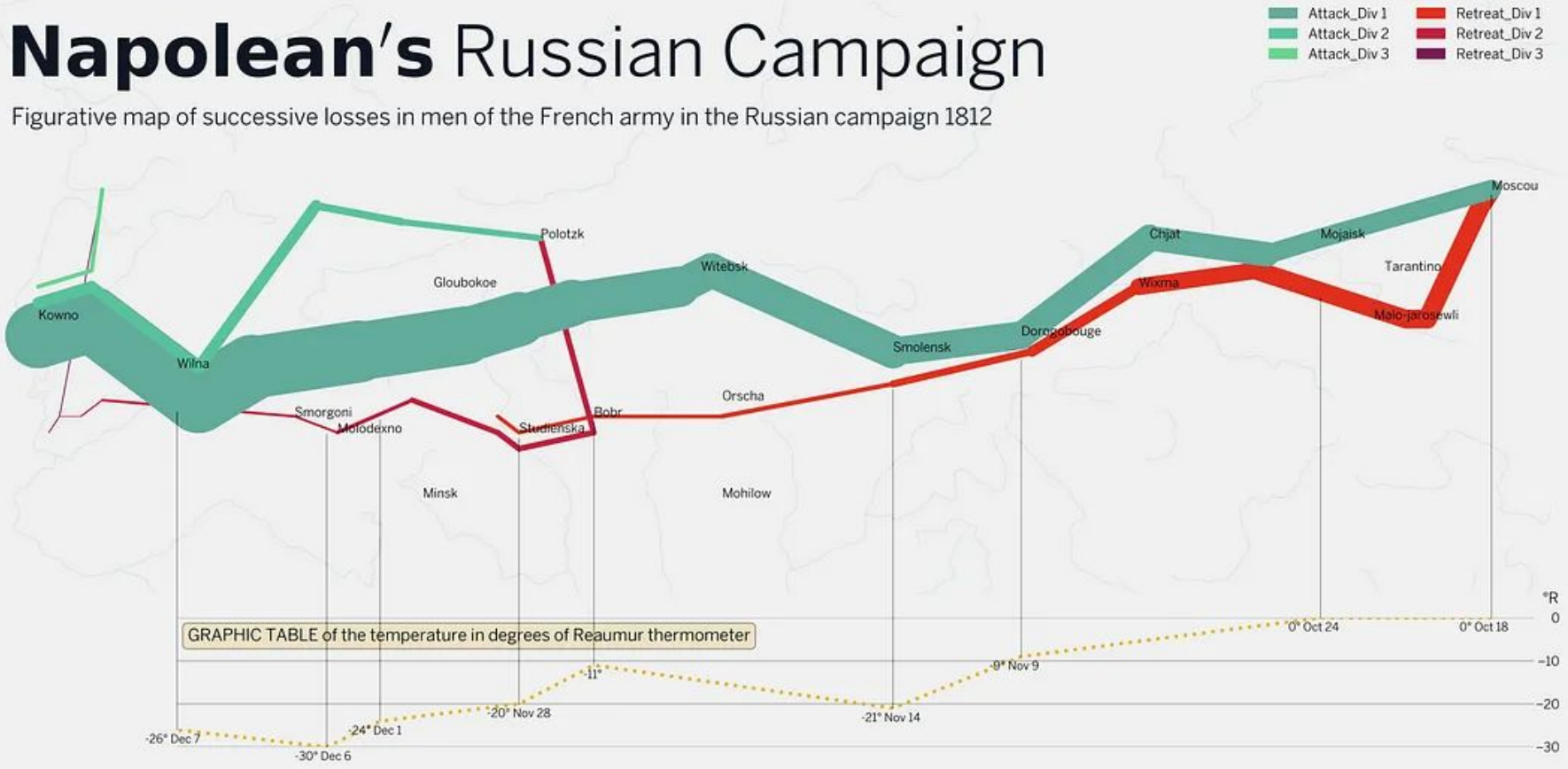


matplotlib:

- “Battle tested”
- Can do *a lot*, takes coding effort - weird, wordy syntax
- Slower with larger datasets
- Still the backbone of many other libraries!
- A lot of export possibilities

Napolean's Russian Campaign

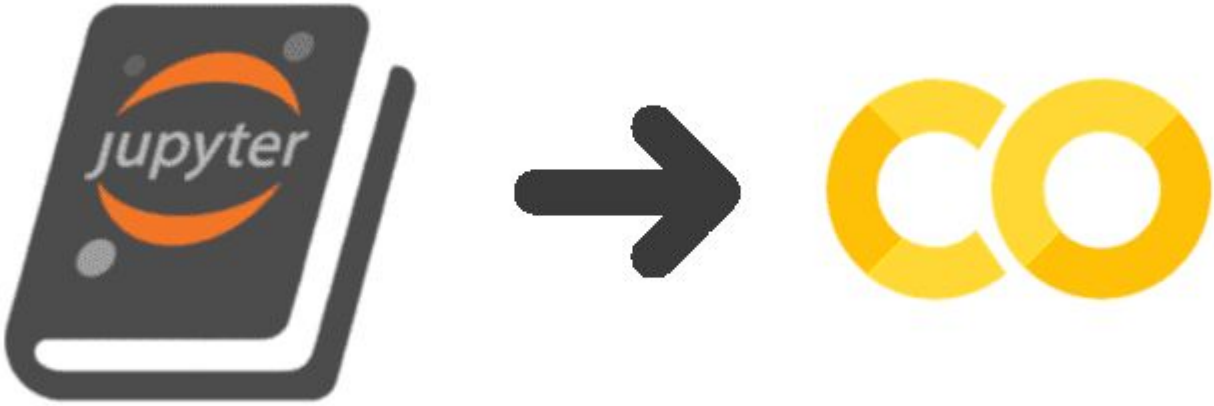
Figurative map of successive losses in men of the French army in the Russian campaign 1812



"Ghost Map"

John Snow cholera data

Infectious Diseases, NSF


jupyter