

lecture 02:

Visualization Fundamentals

September 13, 2017

Data

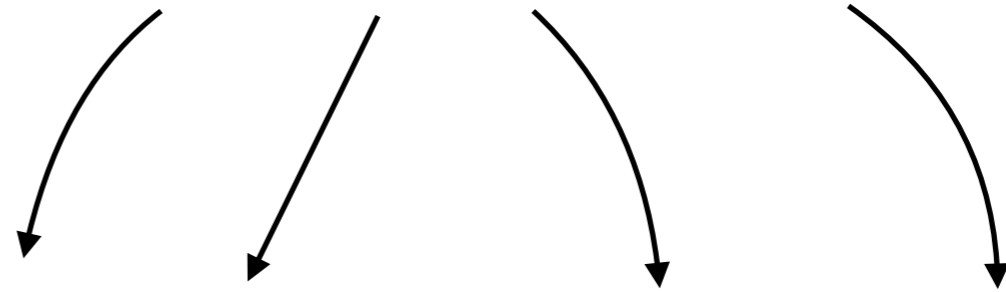
Last class we talked about “tidy data” but there are other ways data can be represented

Hadley Wickham (author of ggplot2, an R package for data viz)

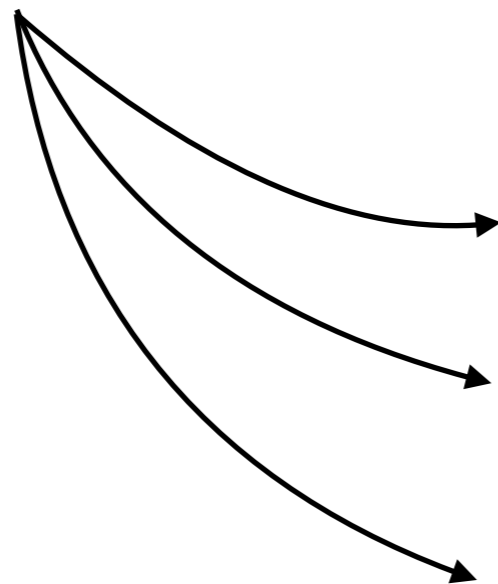
has an idea of

Tidy Data

Columns are variables



Rows are observations

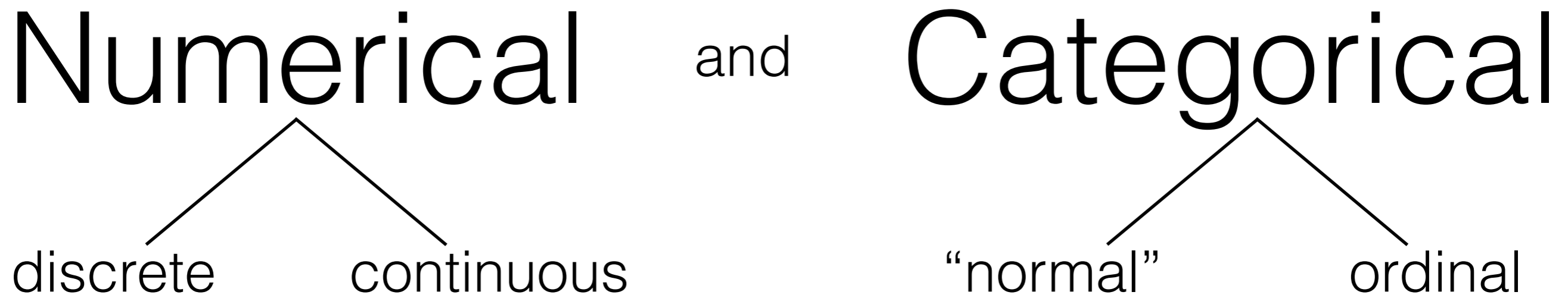


There are other ways to format and represent data

Another common method is a hierarchical, or list-based structure










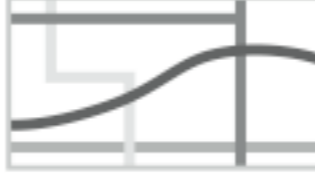





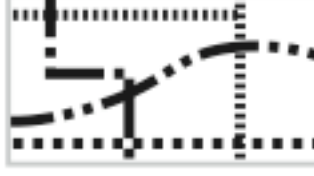
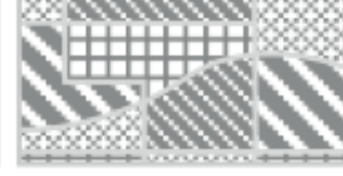
We often make the
distinction between



Let's brainstorm some variables that could be recorded about you that are numerical and categorical

A big question for this course is how to best map variables to visual attributes

Some (all?) of the visual attributes we have to play with

	<i>Points</i>	<i>Lines</i>	<i>Areas</i>	<i>Best to show</i>
<i>Shape</i>		<i>possible, but too weird to show</i>	<i>cartogram</i>	<i>qualitative differences</i>
<i>Size</i>			<i>cartogram</i>	<i>quantitative differences</i>
<i>Color Hue</i>				<i>qualitative differences</i>
<i>Color Value</i>				<i>quantitative differences</i>
<i>Color Intensity</i>				<i>qualitative differences</i>
<i>Texture</i>				<i>qualitative & quantitative differences</i>

Jacques Bertin, *Semiology of Graphics*. 1967

Some history

(old white guy alert)

Laxcaux, France

cave paintings

15,000 B.C.



(via Jordan Crouser)

900s

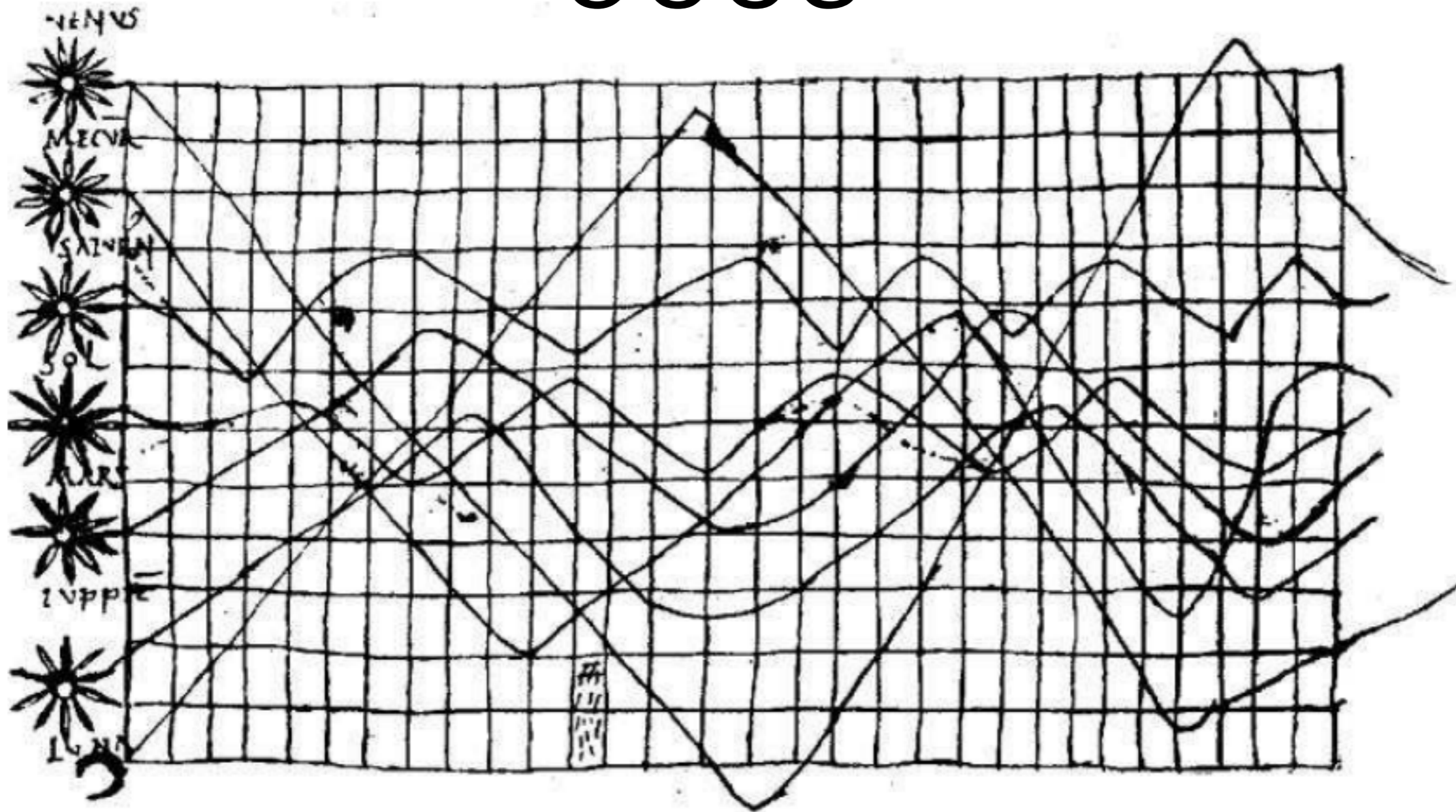
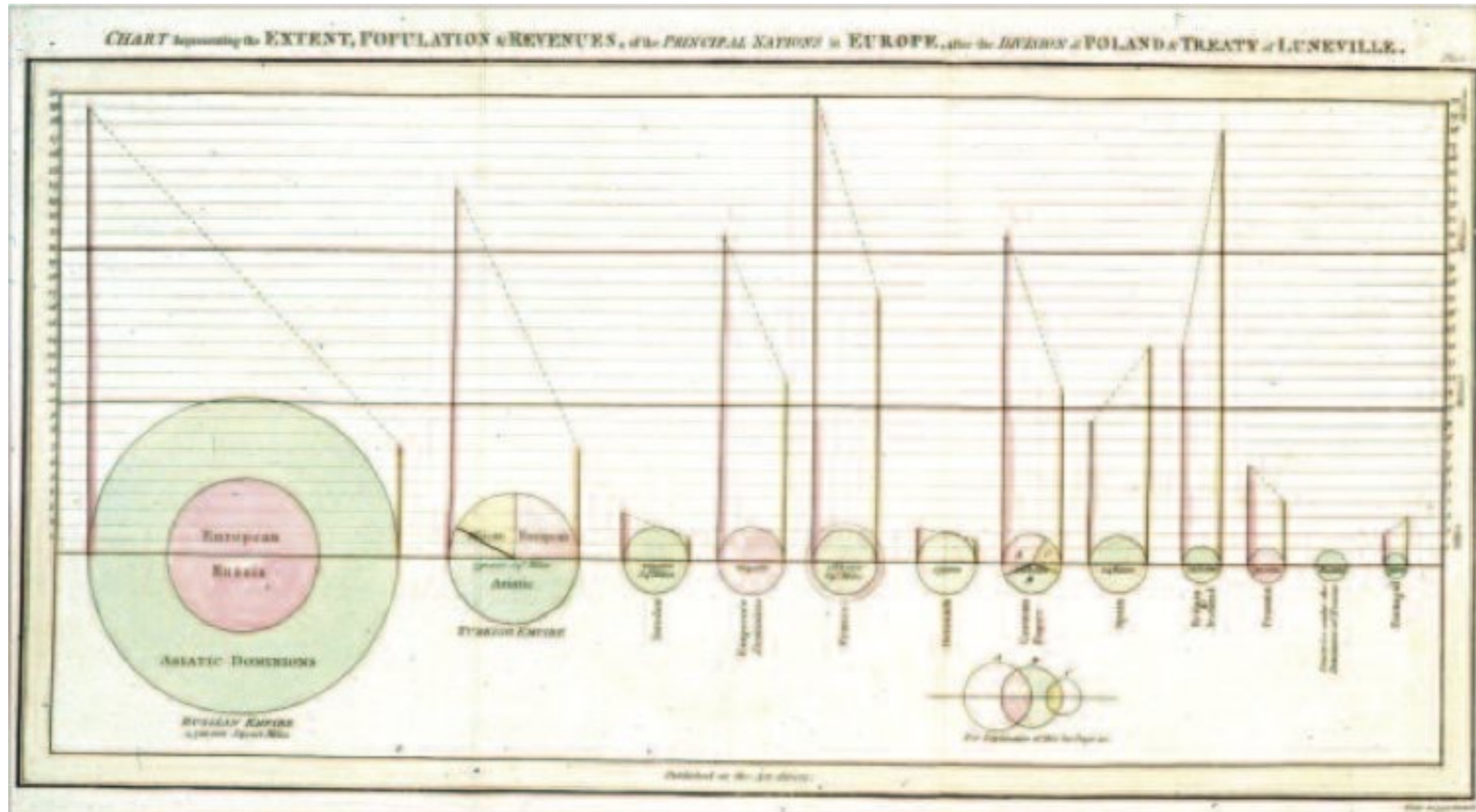


Figure 2: Planetary movements shown as cyclic inclinations over time, by an unknown astronomer, appearing in a 10th century appendix to commentaries by A. T. Macrobius on Cicero's *In Somnium Scipionus*. Source: [Funkhouser \(1936, p. 261\)](#).

William Playfair

1759-1823

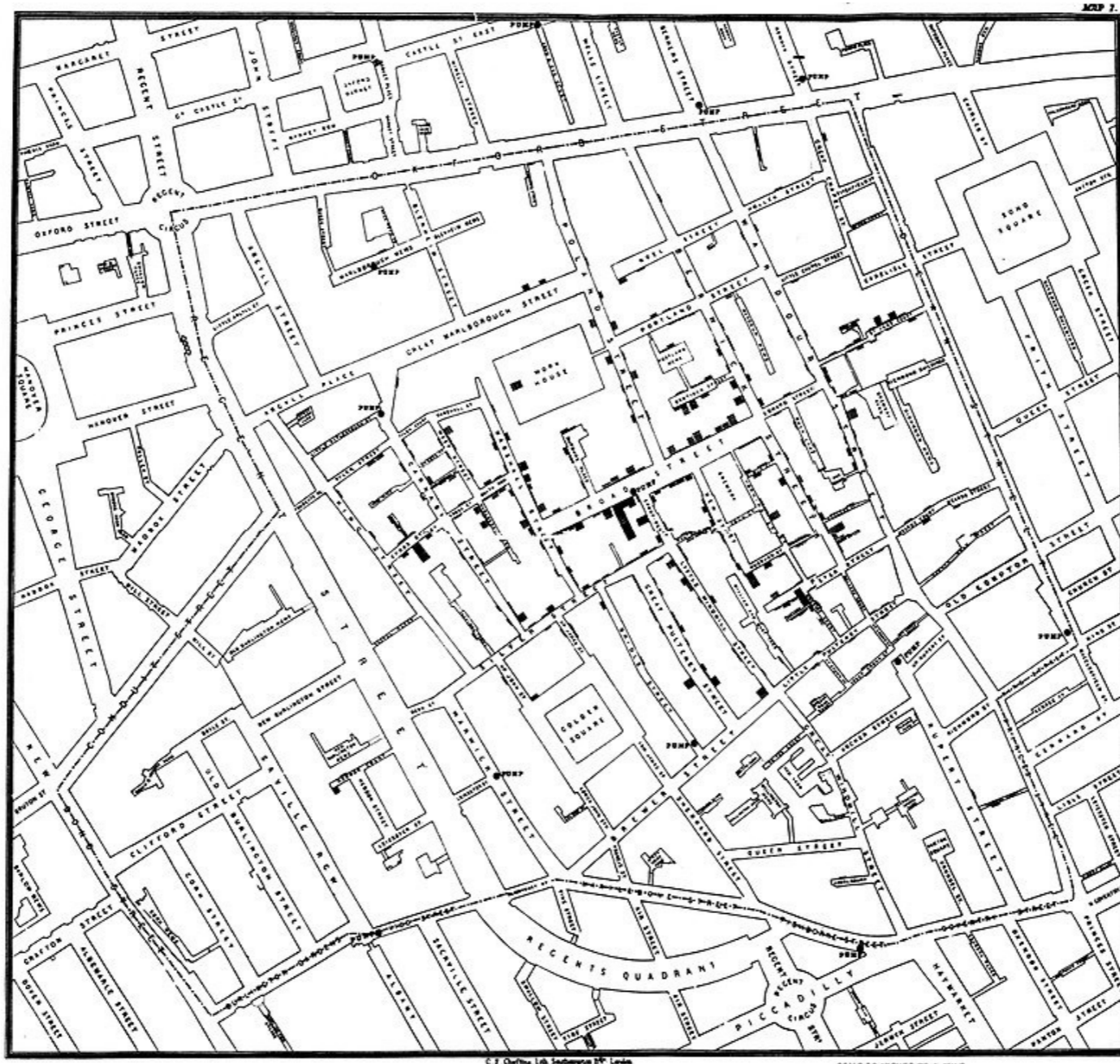
Credited with the invention of many common data visualizations: the pie chart, the bar chart, the line and area chart



John Snow

(no, not the one you're thinking about)

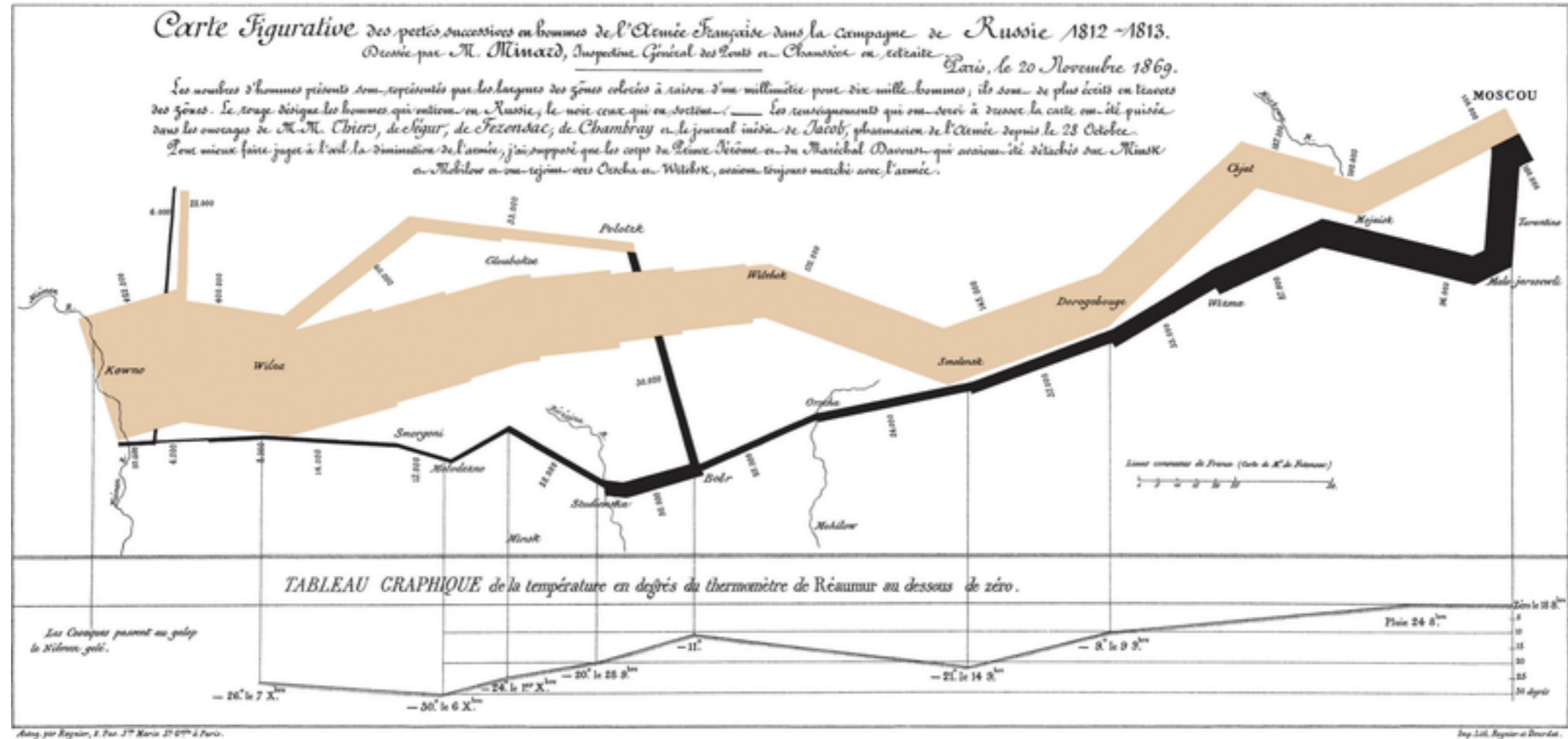
Used mapping to solve a cholera epidemic in London



(poor Charles Cheffins, who drew this graphic and gets no credit)

Charles Joseph Minard

1871-1870



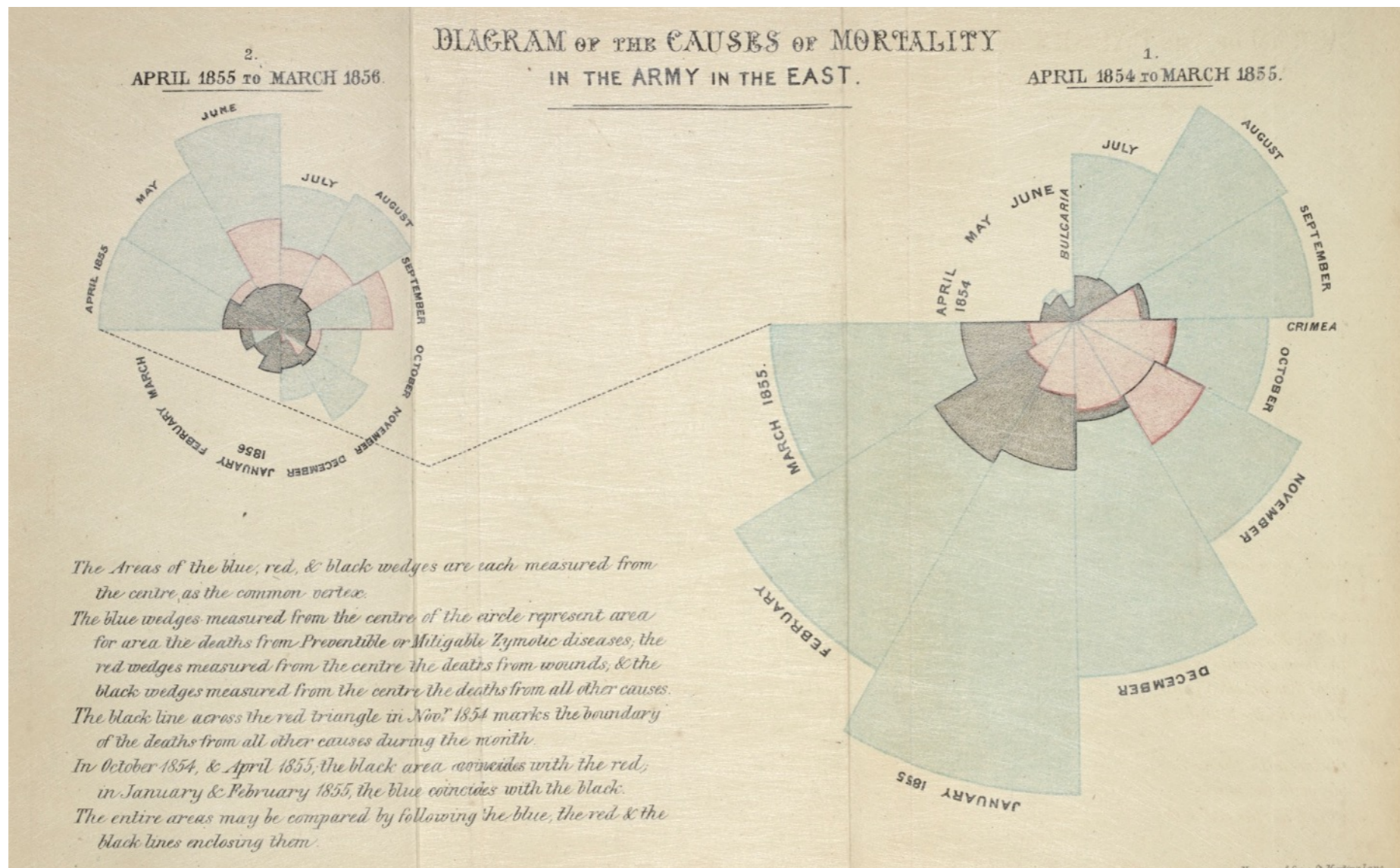
Napoleon's march on Russia

“The best statistical graphic ever drawn”?

Florence Nightingale

1820-1910

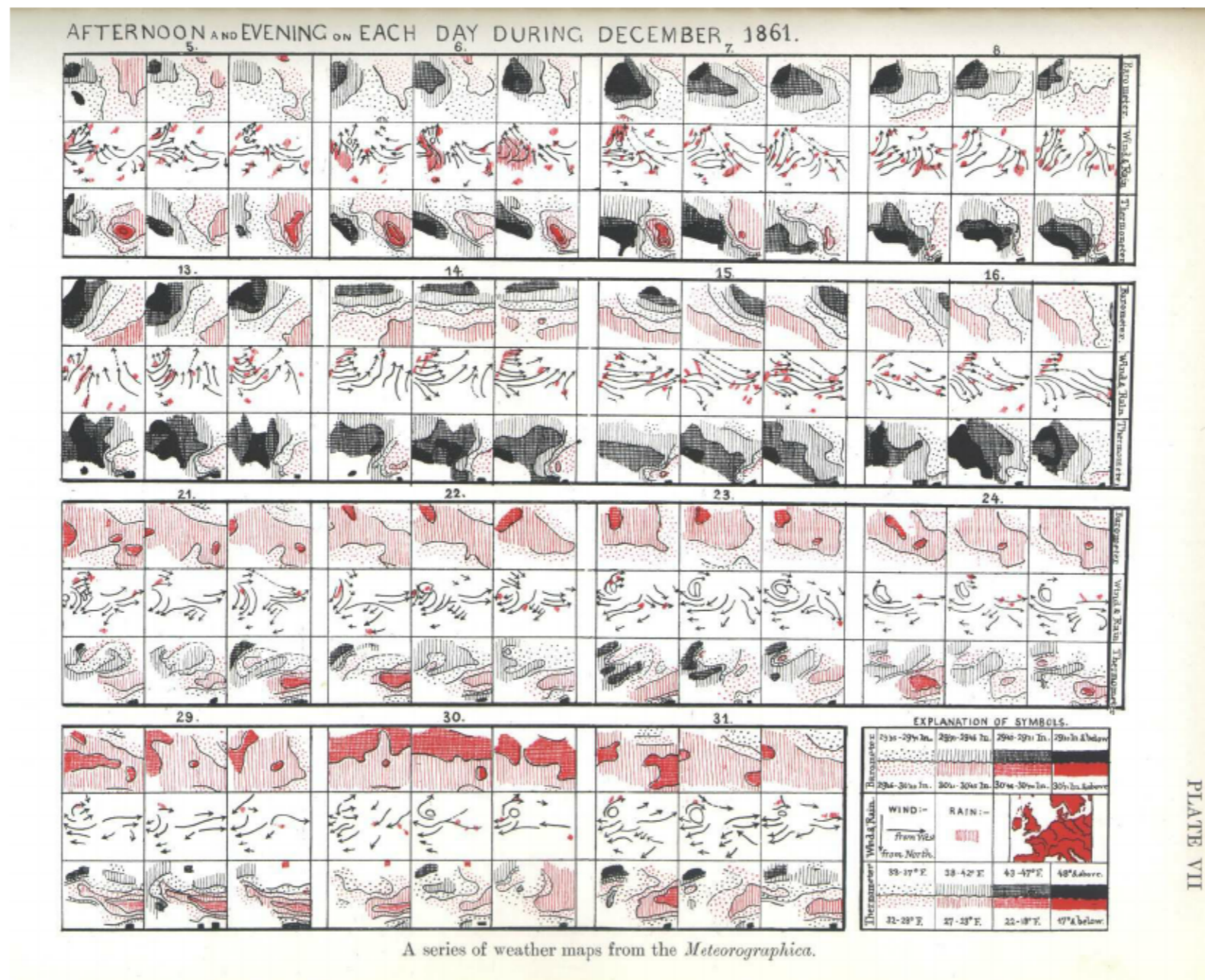
In addition to her work as a nurse, Nightingale was a statistician and invented the “coxcomb,” a variation on the pie chart



Francis Galton

1822-1911

Super-famous statistician 😊
and eugenicist 👎



Statistical atlases

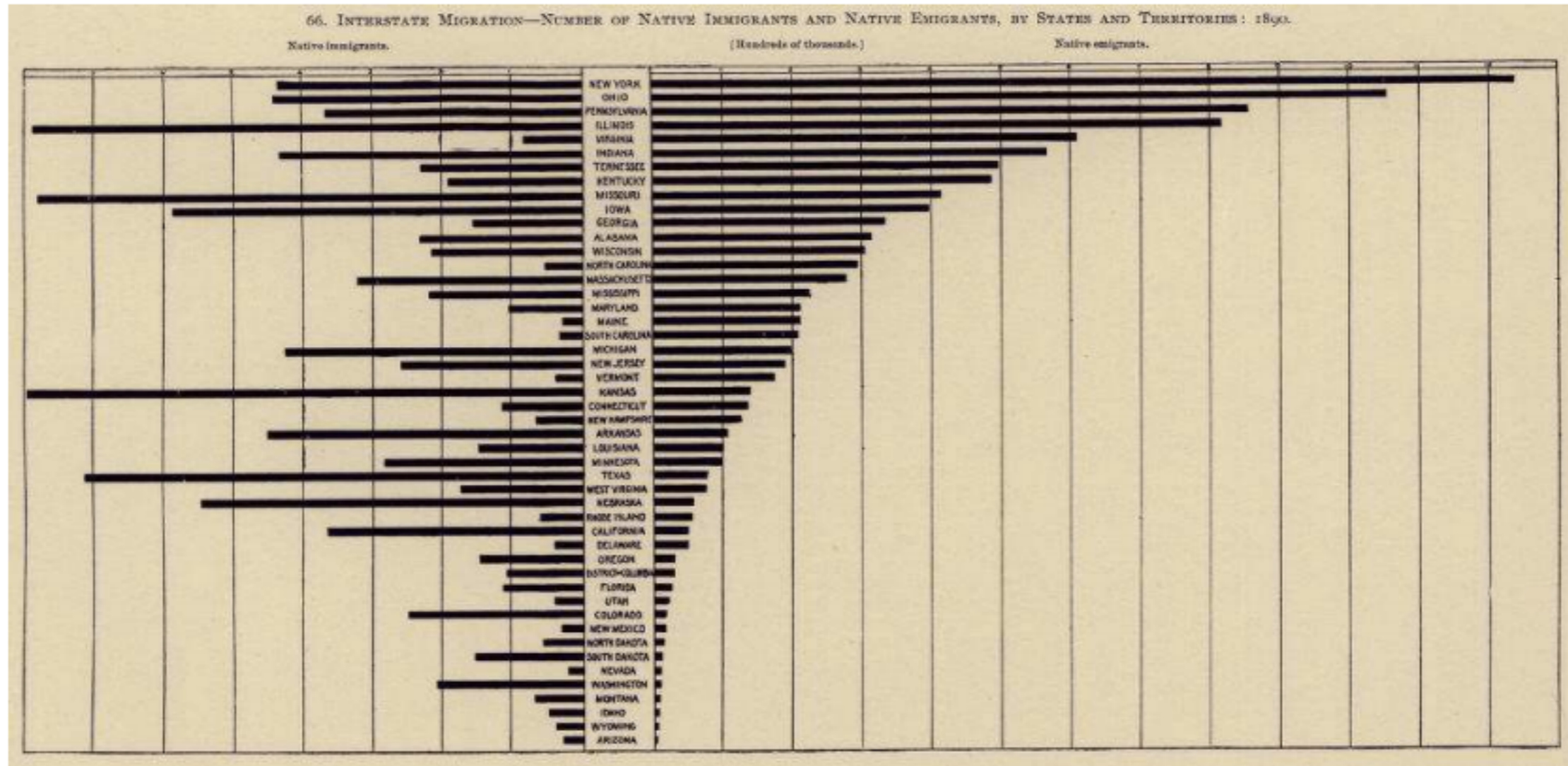
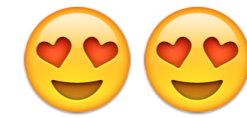


Figure 13: Interstate migration shown by back-to-back bar charts, sorted by emigration. *Source:* Statistical Atlas of the Eleventh Census, 1890, diagram 66, p. 23 (author's collection).

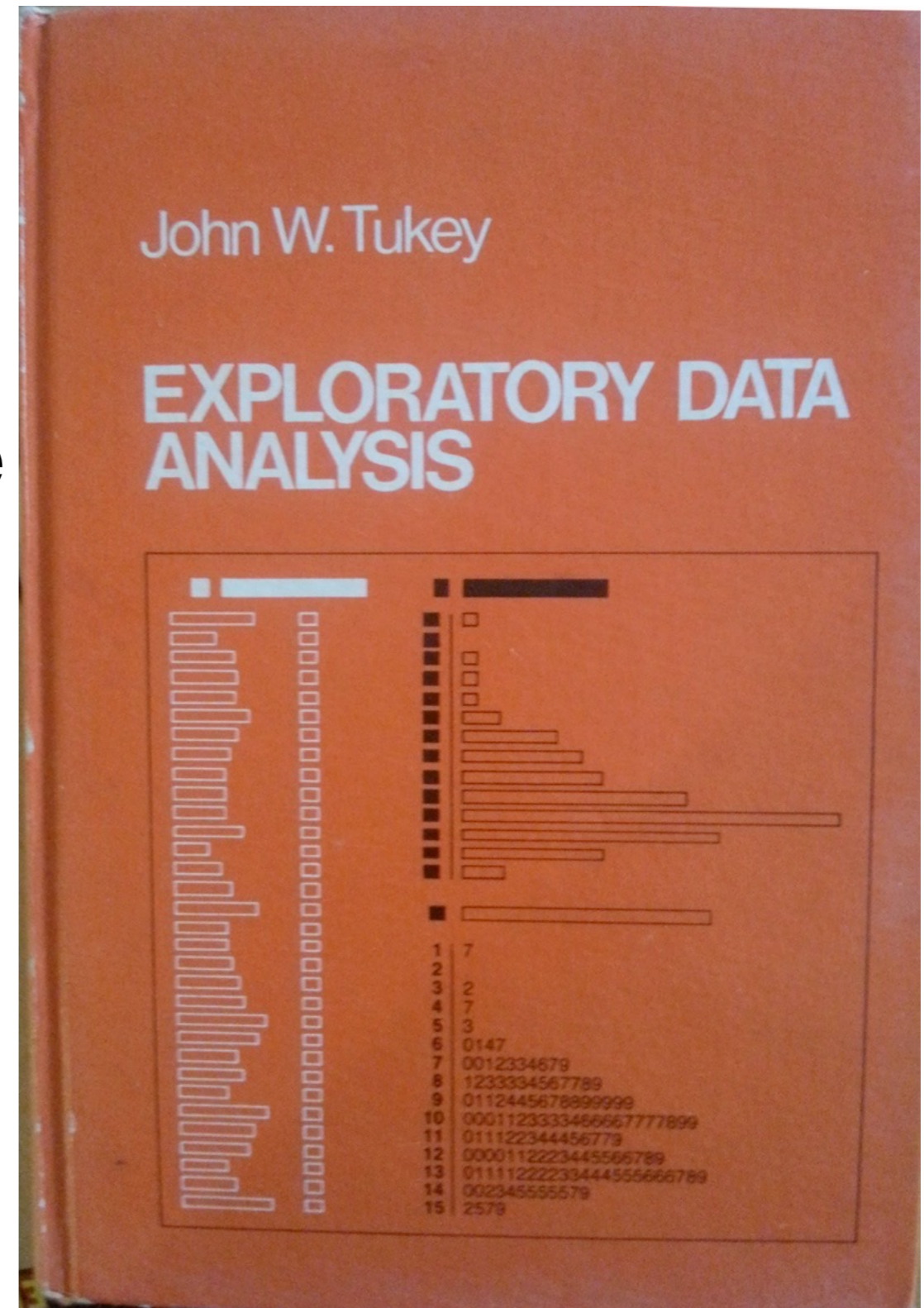
John Tukey



1915–2000

Statistician who rocked the boat

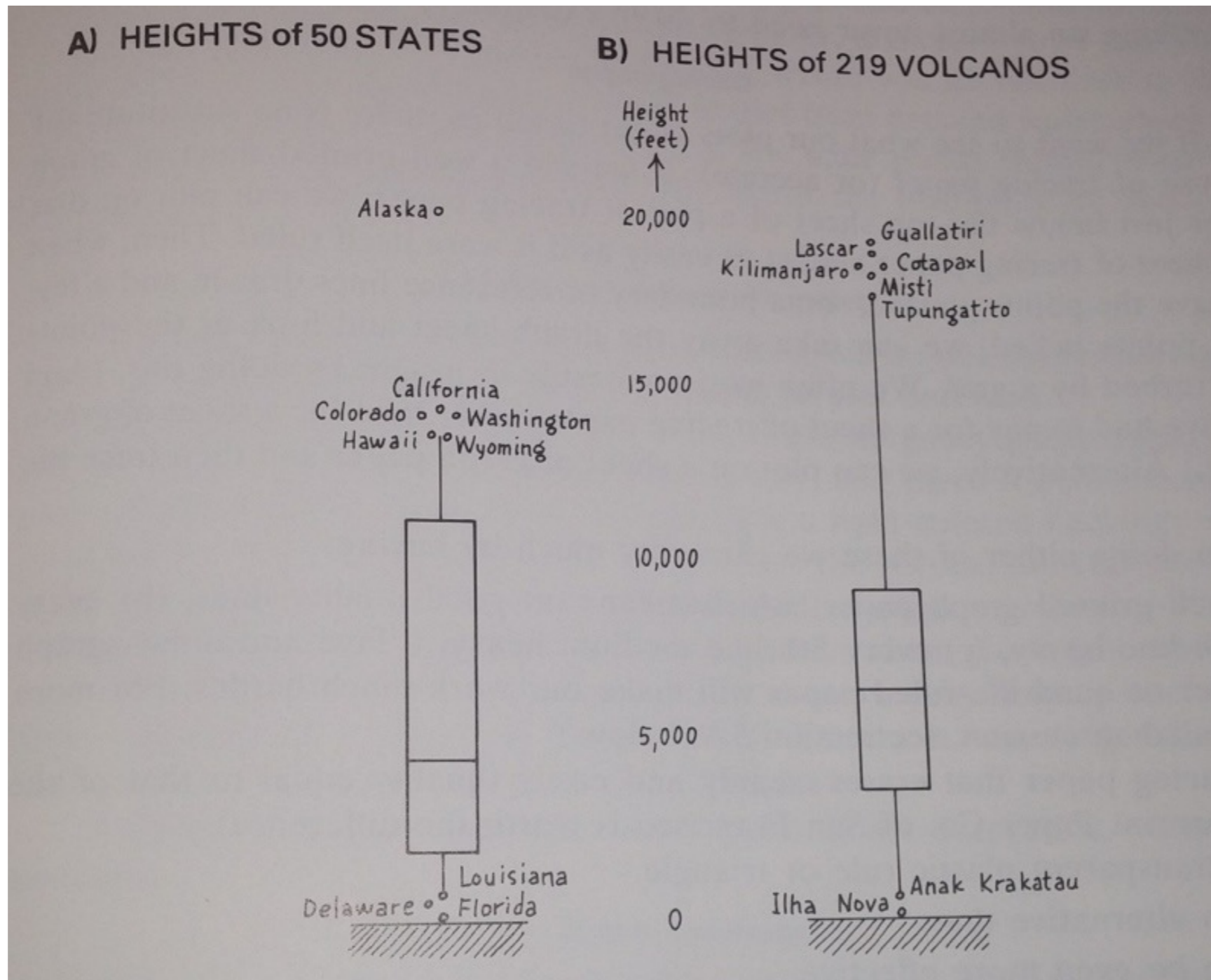
Proposed a method called Exploratory Data Analysis (EDA), which involves making many simple graphs to understand data



John Tukey



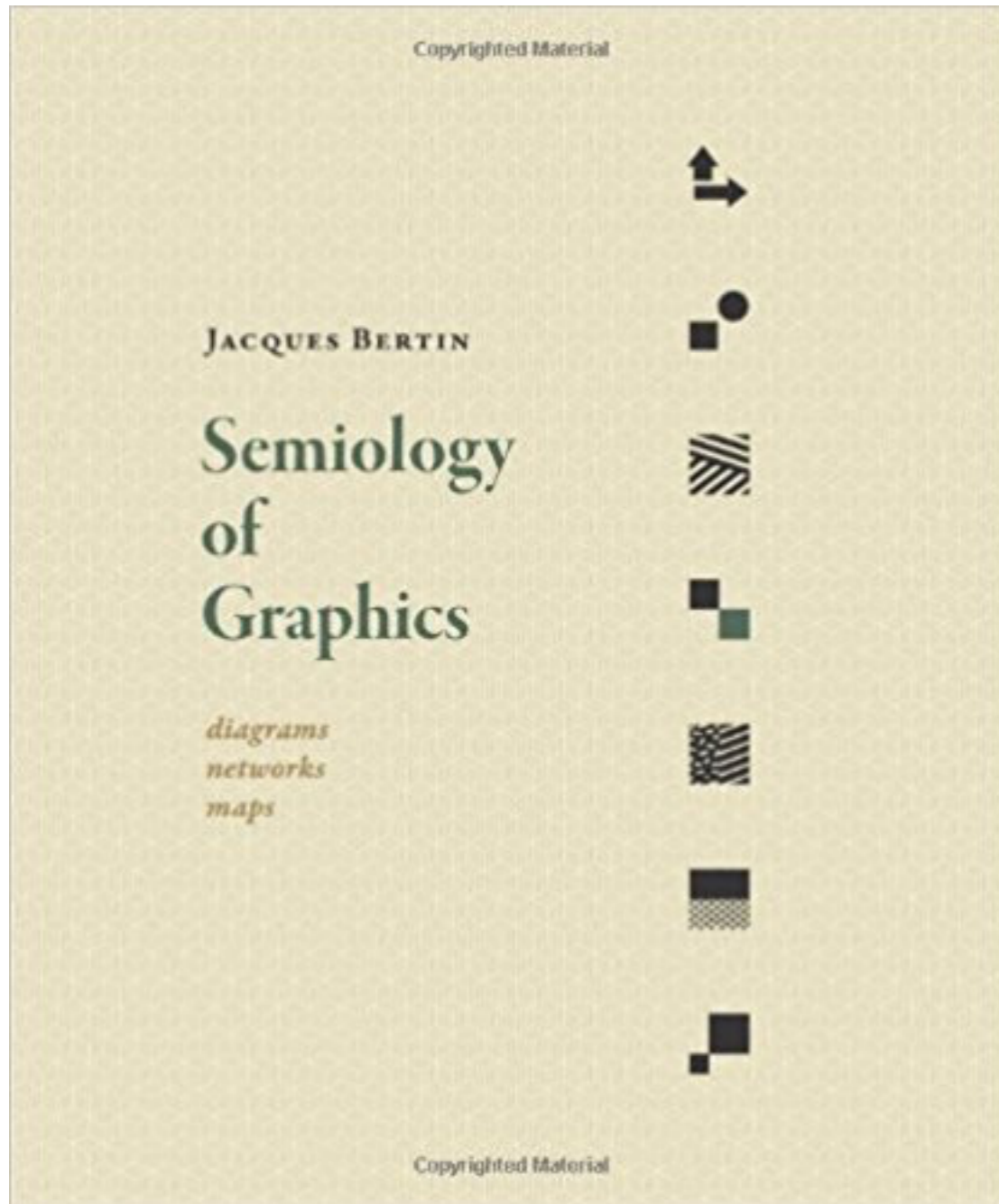
1915–2000



Jacques Bertin

1918-2010

“Cartographer and theorist”



	<i>Points</i>	<i>Lines</i>	<i>Areas</i>	<i>Best to show</i>
<i>Shape</i>		<i>possible, but too weird to show</i>	<i>cartogram</i>	<i>qualitative differences</i>
<i>Size</i>			<i>cartogram</i>	<i>quantitative differences</i>
<i>Color Hue</i>				<i>qualitative differences</i>
<i>Color Value</i>				<i>quantitative differences</i>
<i>Color Intensity</i>				<i>qualitative differences</i>
<i>Texture</i>				<i>qualitative & quantitative differences</i>

Jacques Bertin

1918-2010

les variables rétiniennes

Dans le troisième composant d'une information et en cartographie, c'est la troisième, la représentation graphique, dont fait appel une variable rétinienne.

LES VARIATIONS VISUELLES DISPONIBLES EN 3^e DIMENSION

La psychologie expérimentale définit la perception de la profondeur comme le résultat de facteurs multiples :

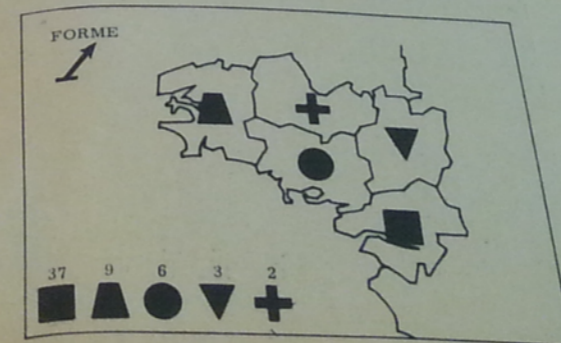
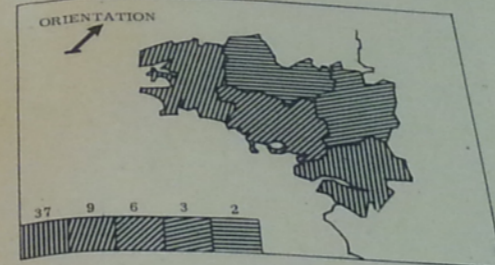
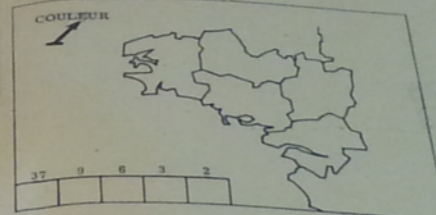
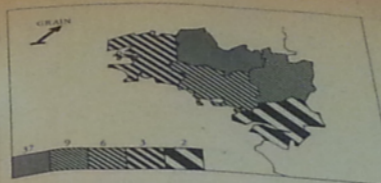
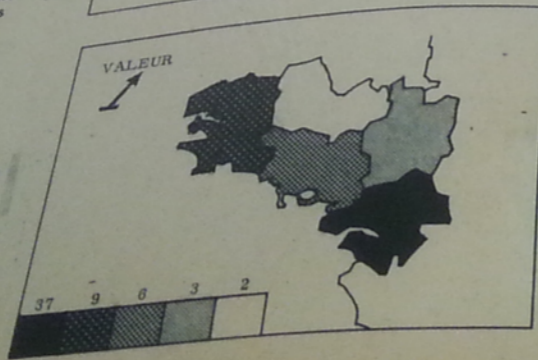
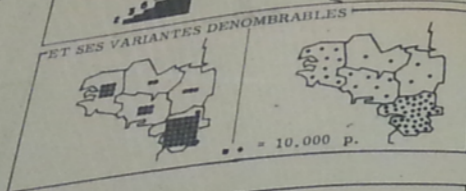
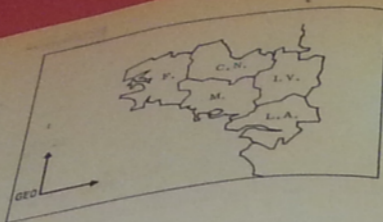
- La vision binoculaire, dans une limite de quelques mètres.
- La diminution de la taille d'un objet connu.
- Le déplacement apparent des objets lorsque l'observateur se déplace.
- La diminution des valeurs d'un contraste connu.
- La réduction du grain d'une rugosité connue.
- La déformation des couleurs d'objets connus.
- Les déformations d'orientation et de forme (perspective).

Toutes ces variations à l'exclusion des deux premières sont à la disposition du rédacteur graphique, qui peut les utiliser pour représenter une composante de 3^e dimension.

Il peut faire correspondre les catégories de la composante de 3^e dimension avec chacune de ces variables.

Avec des catégories de TAILLE : hauteur d'une colonne, surface d'un signe, quantité de signes égaux.

Avec des catégories de VALEUR, qu'il est possible de créer, entre le blanc et le noir.



Avec des catégories de GRAIN, c'est-à-dire avec la variation de finesse des constituants d'une plage de valeur donnée et qui s'échelonnent de la plus grossière à la plus fine. On obtient cette variation par exemple en augmentant ou en réduisant photographiquement une trame.

Avec des catégories de COULEUR, en utilisant le répertoire des sensations colorées, que l'on peut rendre indépendant de la variation de valeur.

Avec des catégories d'ORIENTATION, orientations diverses d'un baguettage ou d'une ligne, qui s'échelonnent de la verticale à l'horizontale et ont un sens.

Avec des catégories de FORME puisqu'une tache, de surface constante peut néanmoins avoir une infinité de formes différentes.

Ainsi toute variable visuelle de 3^e dimension peut être employée à la représentation de toute composante. Mais il est évident que toute variable n'est pas adaptée à toute composante. C'est la notion de niveau d'organisation qui fournit la clef de ce problème.

2.3 - Le

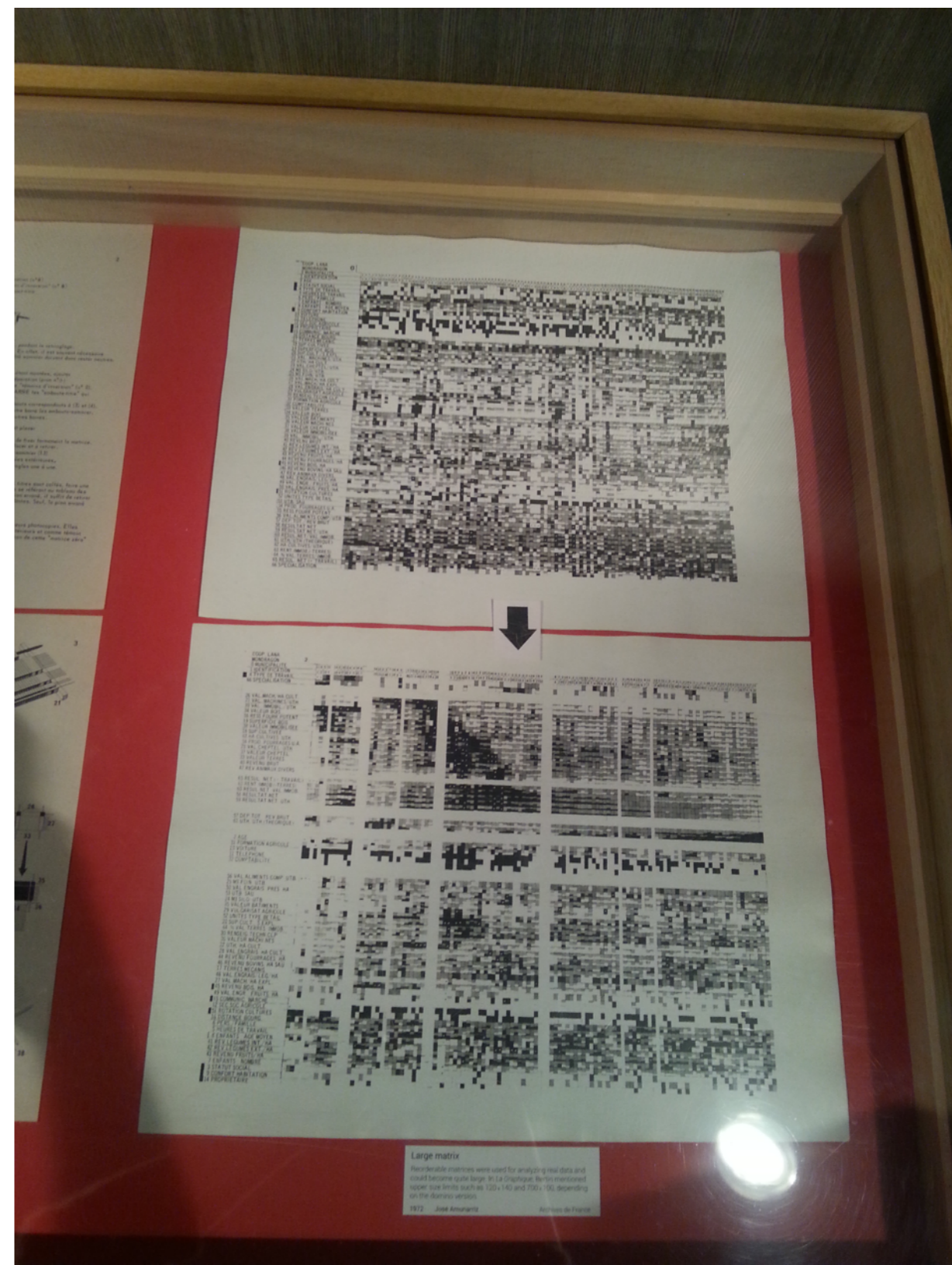
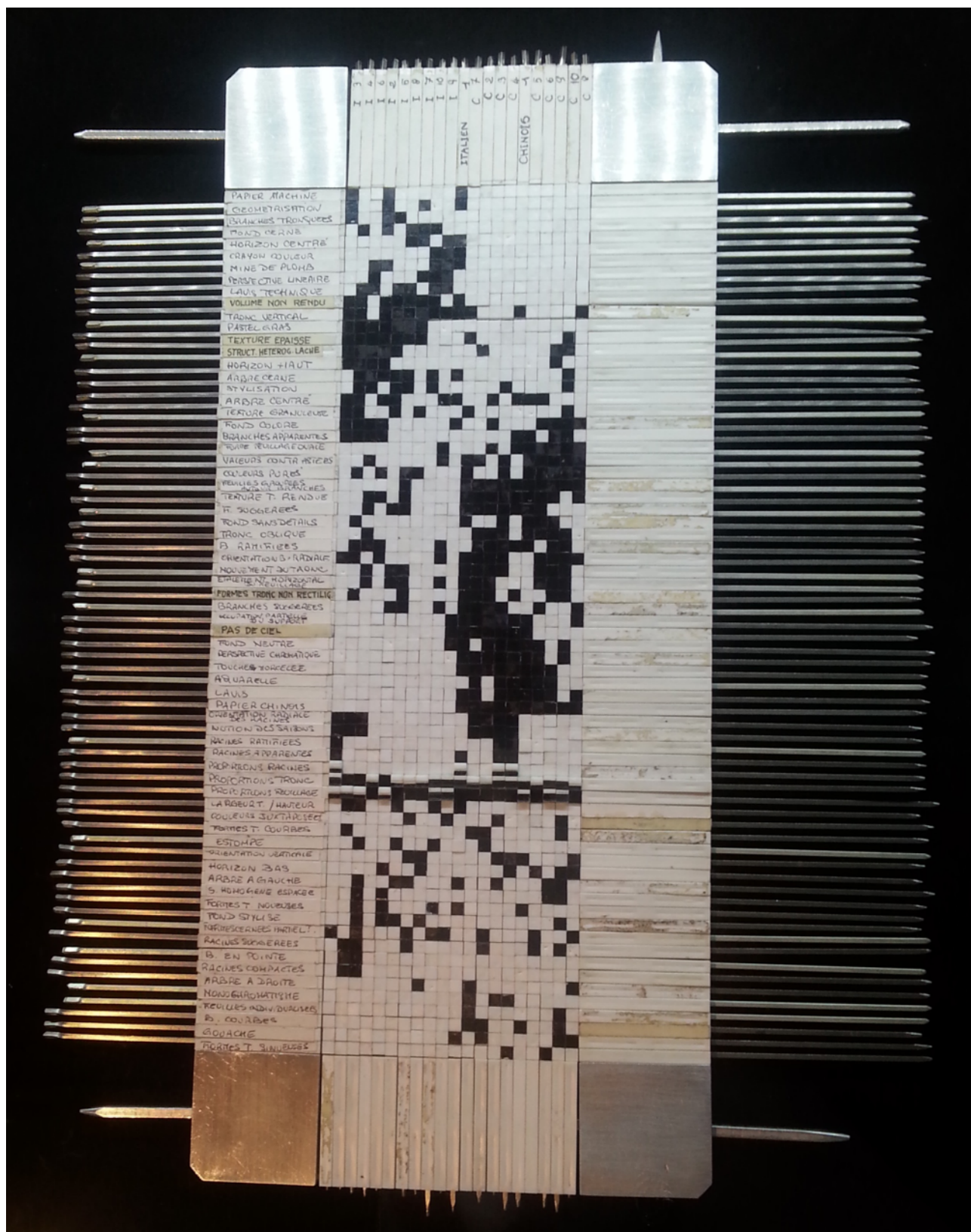
2.3.1. Le

Les fl
On cor
puis
gnée
Ré
Ré
fi

51

Jacques Bertin

Bertin matrices



William Cleveland

1943-

Professor of statistics at
Purdue

Did famous research about
effectiveness of visualizations

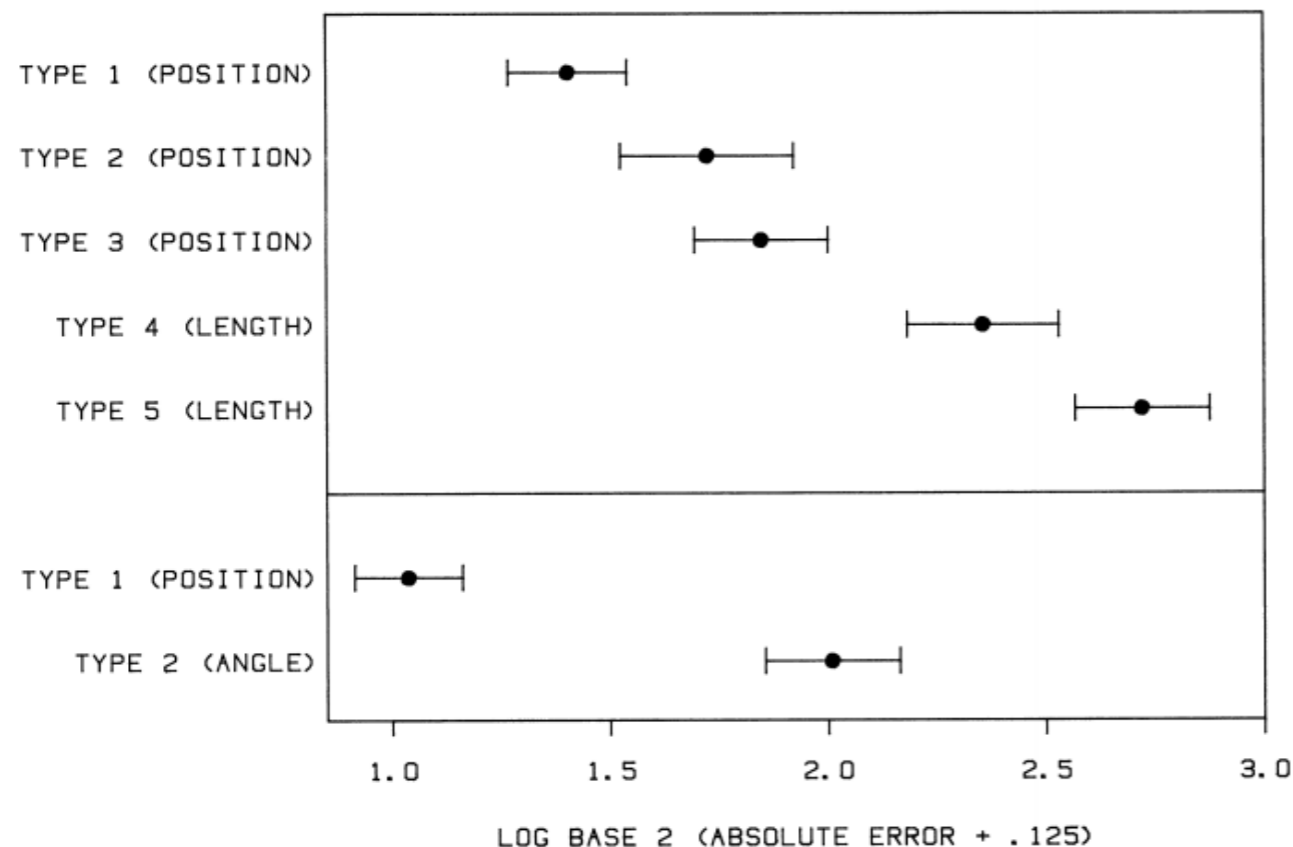
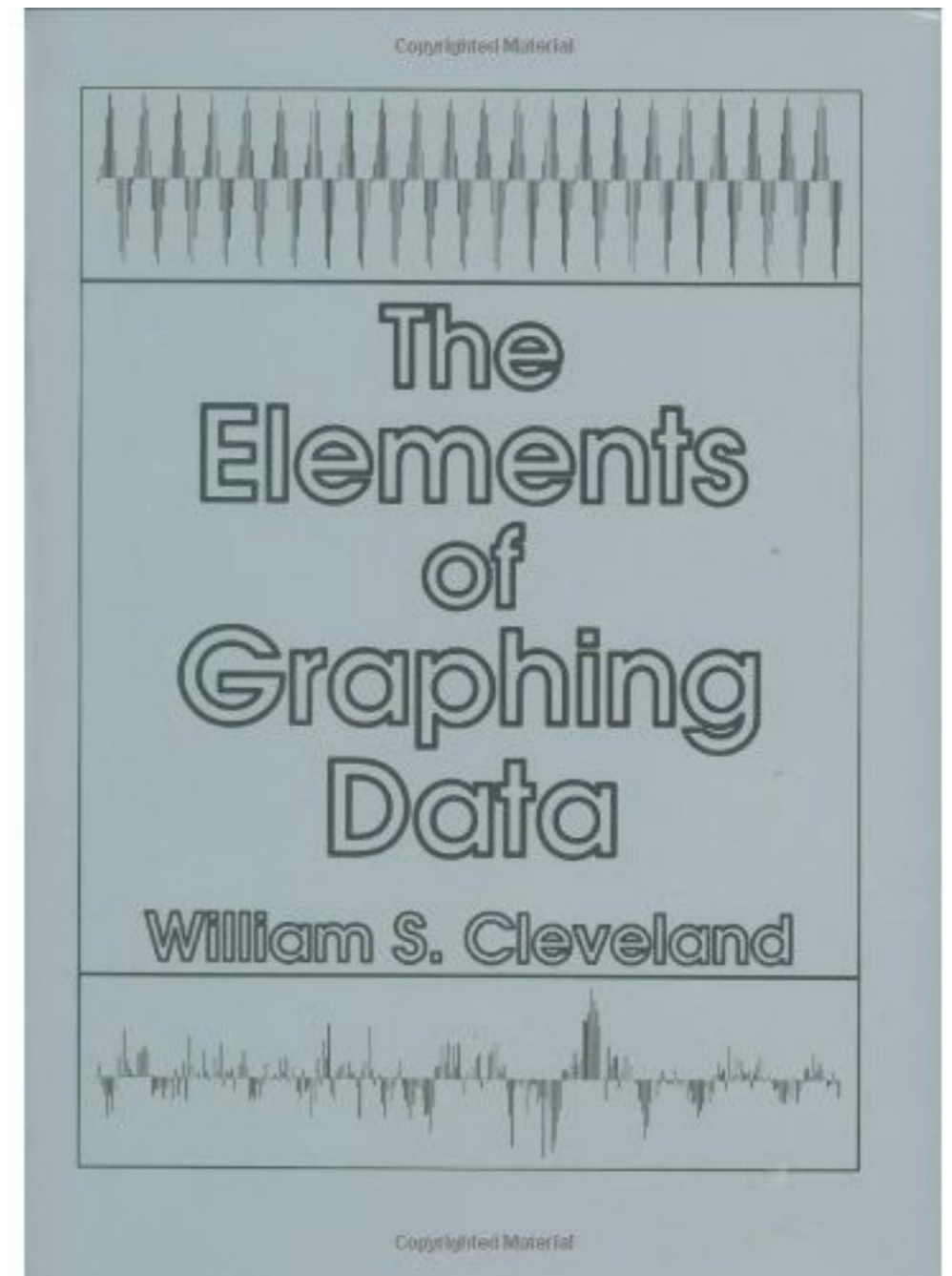
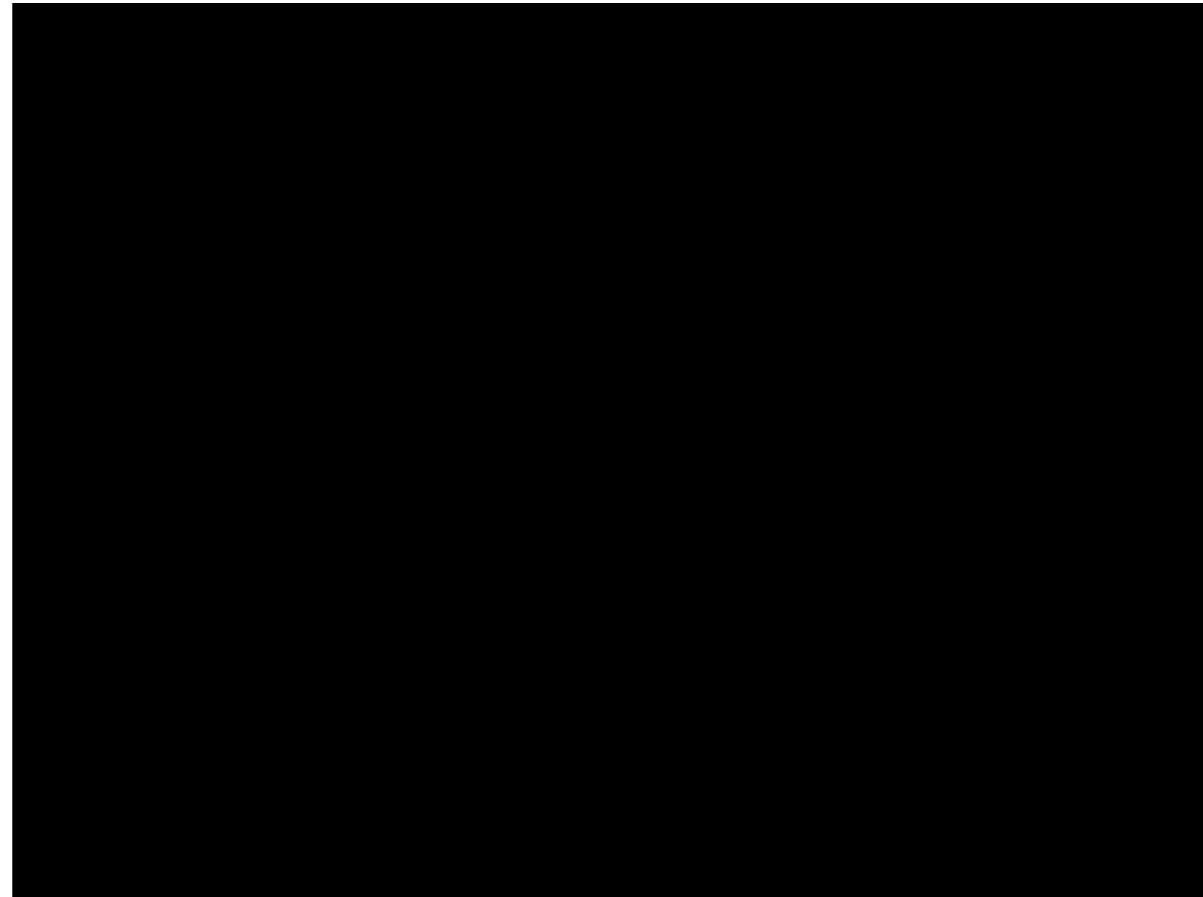


Figure 16. Log absolute error means and 95% confidence intervals for judgment types in position-length experiment (top) and position-angle experiment (bottom).

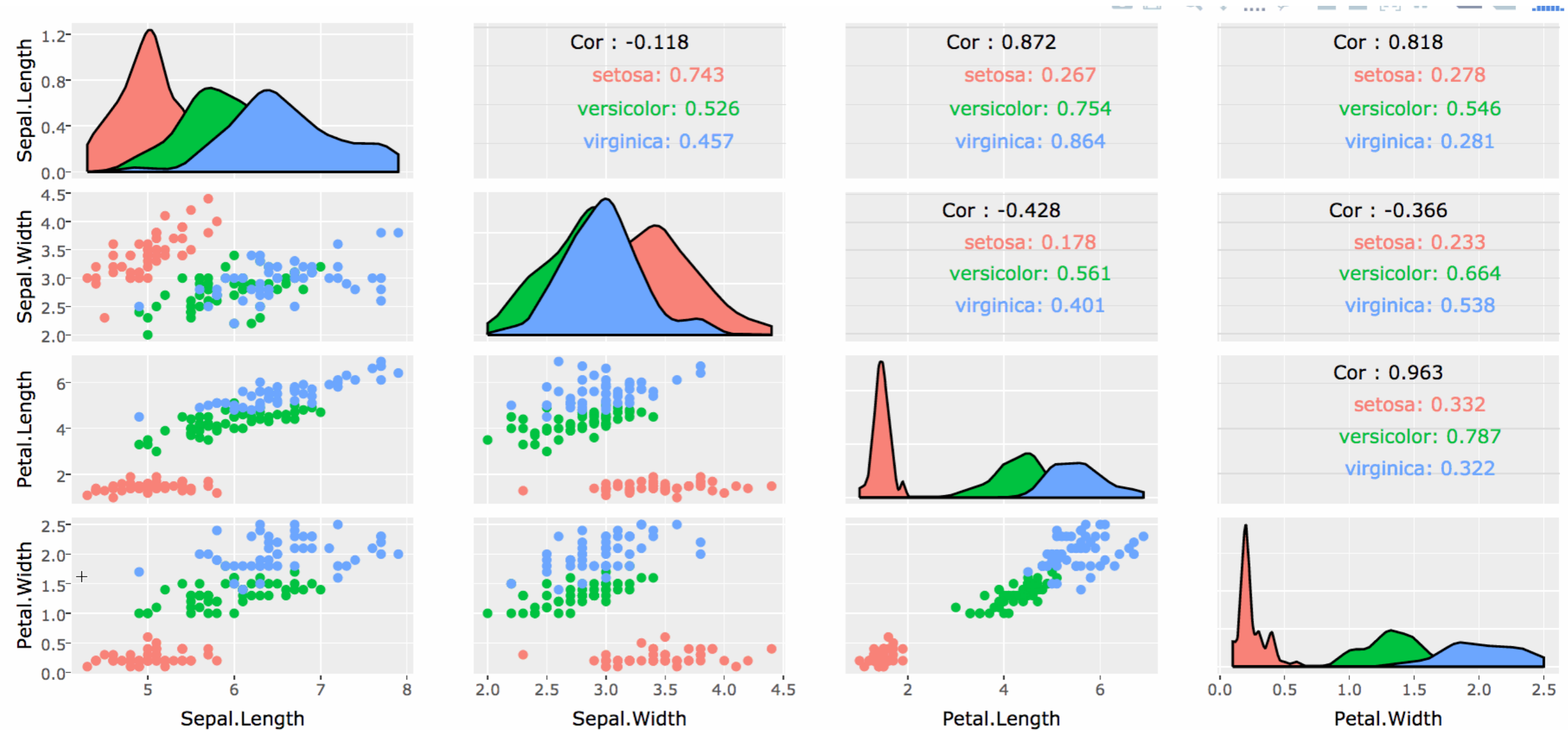


John Tukey

1915–2000



Interactivity, brushing and linking



Luke Tierney

xlisp-stat

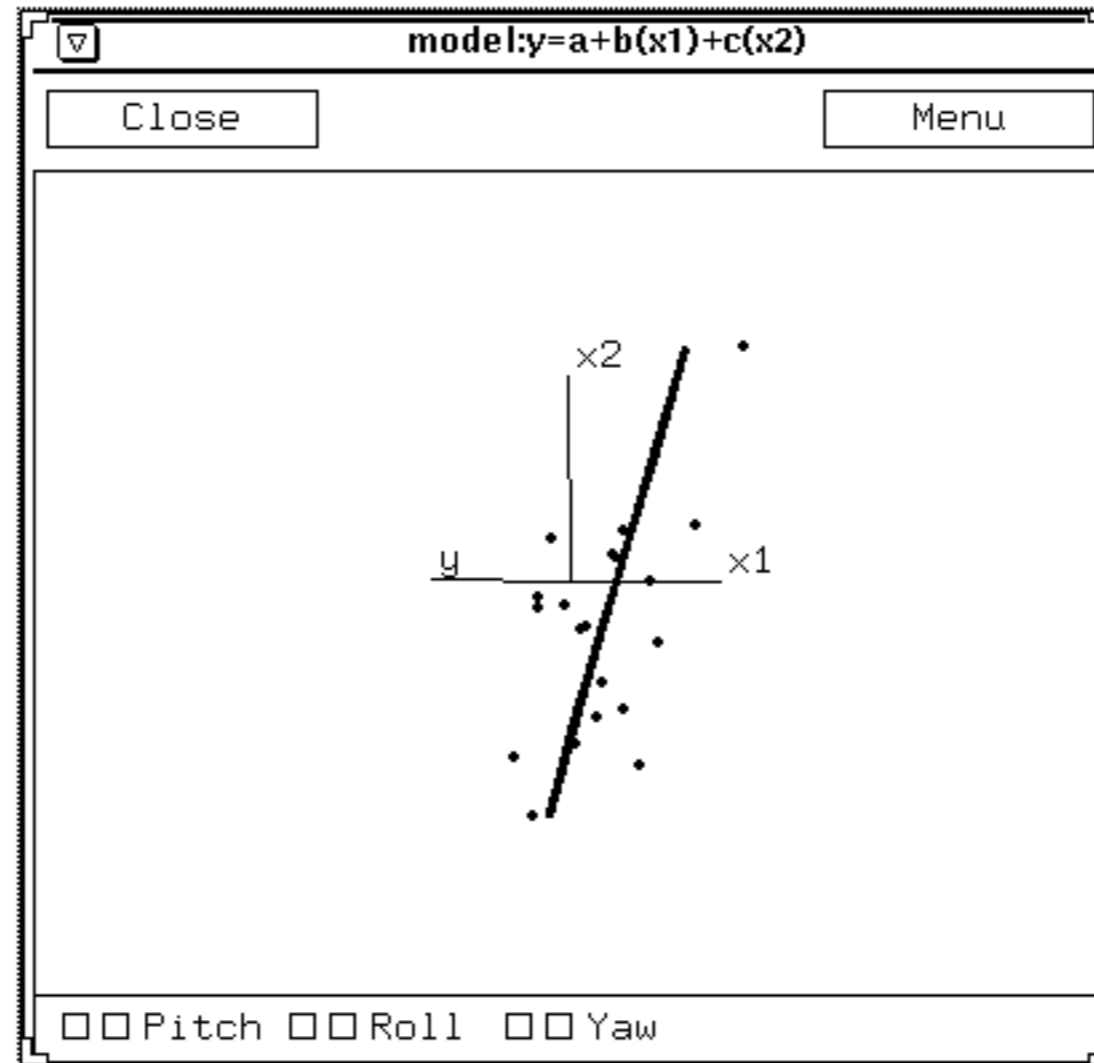
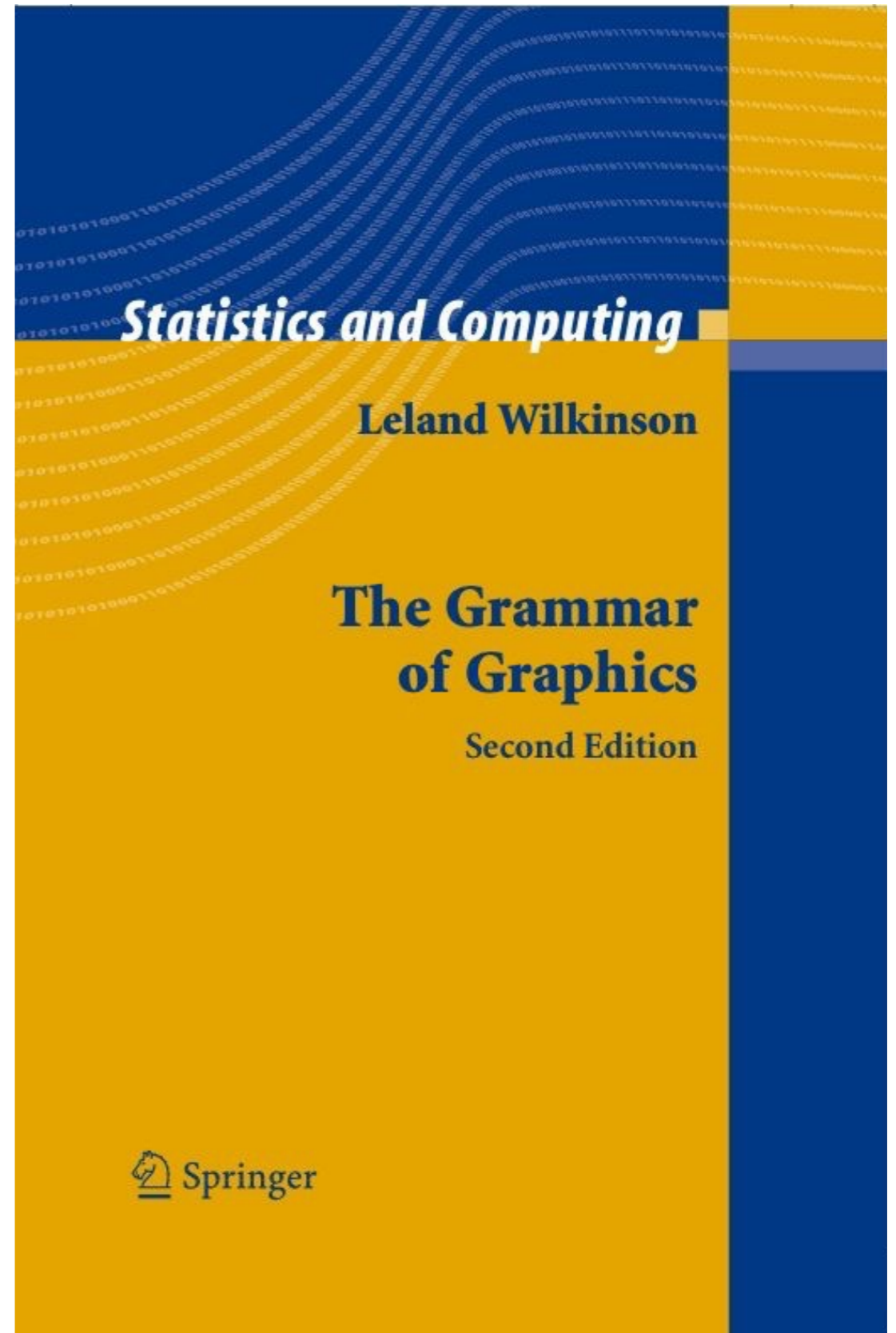


Figure 7a

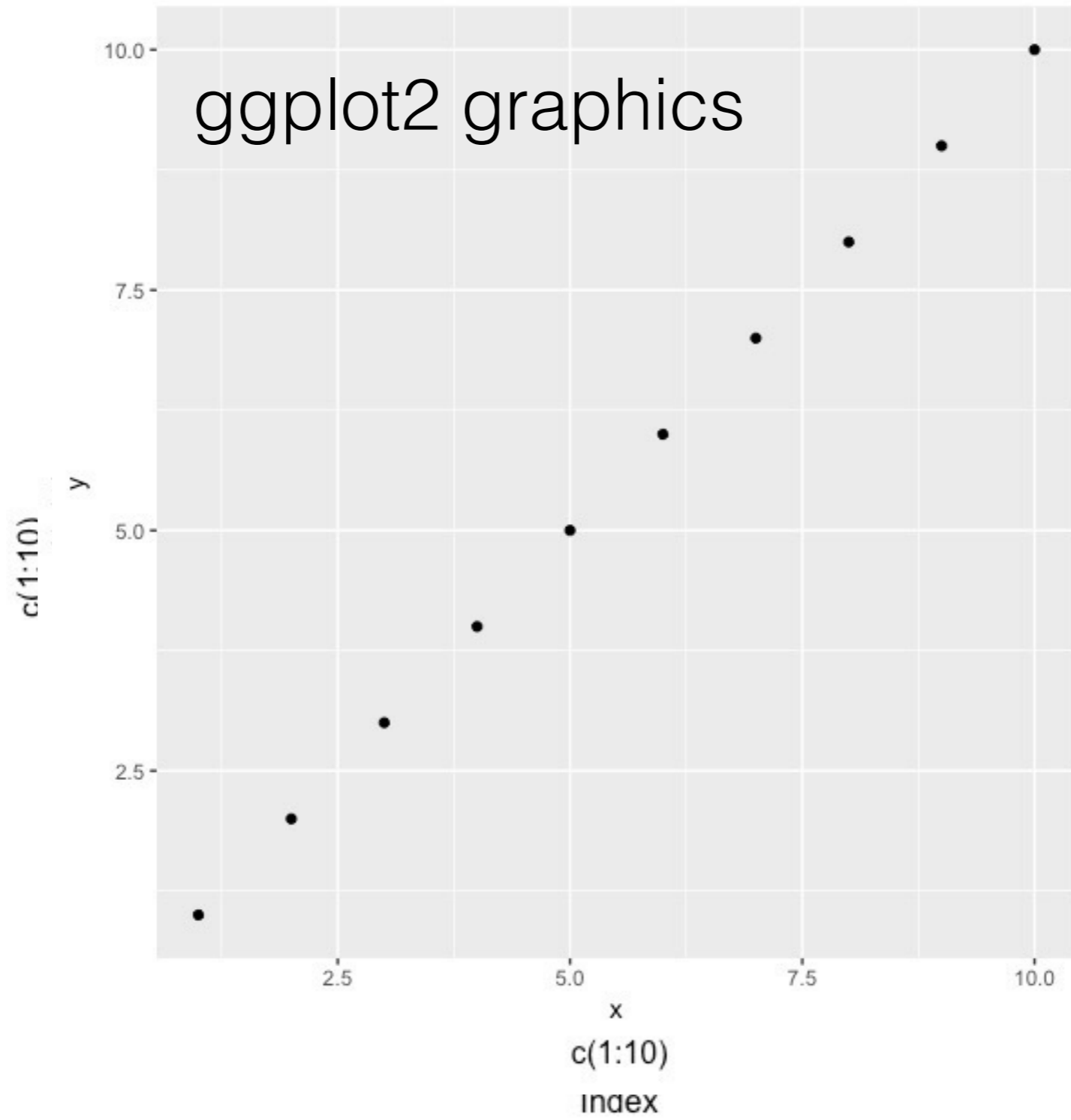
Leland Wilkinson

Statistician and software
designer

Worked on SYSTAT, SPSS,
Tableau, now H2O.ai



R



Hadley Wickham

Famous R programmer

Dissertation: "Practical tools for exploring data and models"

Advisors: Di Cook and Heike Hoffman



Mike Bostock

d3.js

[Overview](#) [Examples](#) [Documentation](#) [Source](#)

Data-Driven Documents

Fork me on GitHub



Before next class

- Read the readings posted on Moodle
 - Introduction to The Functional Art
 - Introduction to Visualize This
 - Chapter 1 of Visualize This

If you are in group A (last name beginning A-G):

- Post a response to one (or more) of the readings on Slack in the #readingresponse channel

If you are in groups B (last name beginning H through L) or C (last name beginning M through Z):

- Respond to one of your fellow classmates' responses!