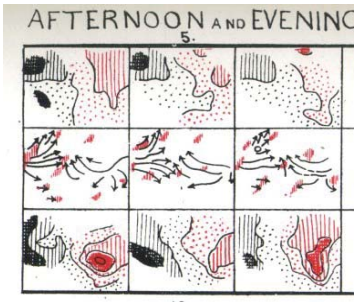



Visions of the Past, Present and Future of Data Visualization

Graphical Successes from the Golden Age

Michael Friendly, York University

Rostock Retreat, June 2017

Slides: http://datavis.ca/papers/Rostock-2x2.pdf

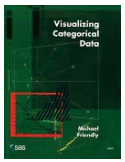
Outline

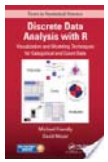
- Introduction
  - Visualization and scientific discovery?
  - The Milestones Project
- The Golden Age of Statistical Graphics
  - A.-M. Guerry & the rise of social science
  - Visual thinking: C.J. Minard
  - Francis Galton's graphical discoveries
  - Graphical excellence: Albums de Statistique Graphique
- Today?
- Tomorrow?

Introducing: me

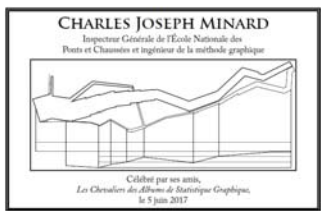
I wear two hats, both reflected on my license plate

Statistical graphics developer (categorical & multivariate data analysis)



History of data visualization: *Les Chevaliers*; *The Origins of Graphical Species* (2018)

A secret: How I got to be so smart

Much of the progress in the history of statistics and data visualization can be thought of as an expansion from

univariate → bivariate → multivariate problems

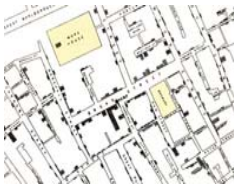
When I was younger …

… Now, I've discovered trinocular vision

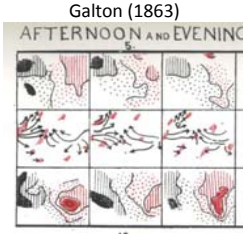
Orienting Q: Visualization-based discoveries ??

- When have graphics led to discoveries that might not have been achieved otherwise?
  - Snow (1854): cholera as a water-borne disease
  - Galton (1883): anti-cyclonic weather patterns
  - E.W. Maunder (1904): 11-year sunspot cycle
  - Hertzsprung/Russell (1911): spectral classes of stars

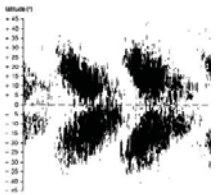
Snow (1854)



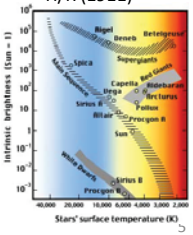
Galton (1863)



Maunder (1904)



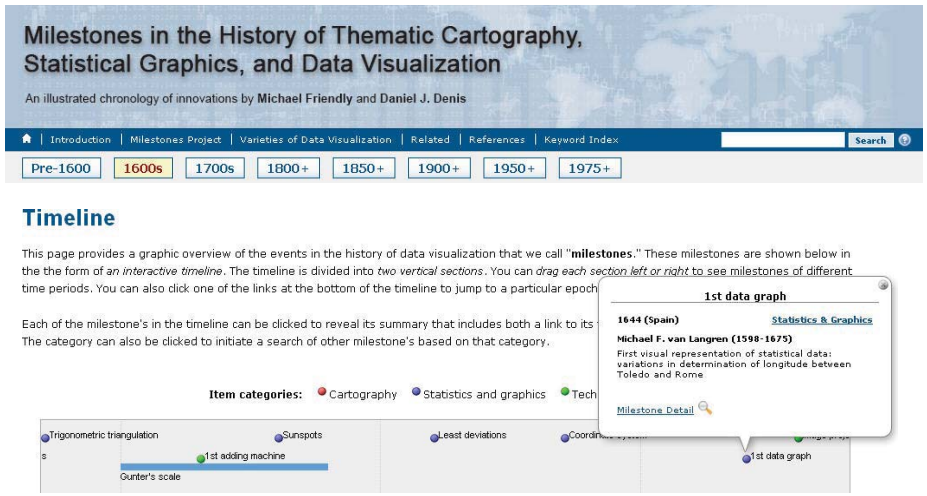
H/R (1911)



Orienting Q: Visualization-based discoveries ??

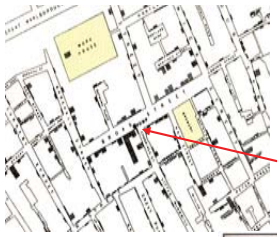
- In the history of graphs, what features and data led to such discoveries?
  - What visual ideas/representations were available?
  - What was needed to see/understand something new?
- As we go forward, are there any lessons?
  - What are the Big Questions for today?
  - How can data visualization help?

Context: Milestones Project



Web site: http://datavis.ca/milestones

1850-1900: Golden Age

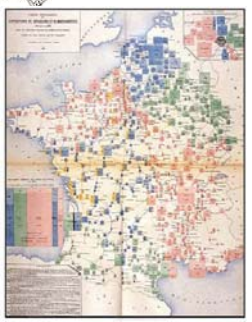


1855: Dot map of disease data (cholera)- John Snow

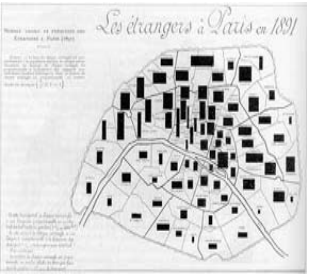
Broad St. pump

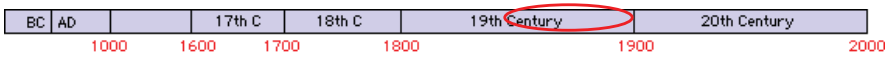
1879: Stereogram (3D population pyramid)- Luigi Perozzo

1884: Recursive multi-mosaic on a map- Emile Cheysson

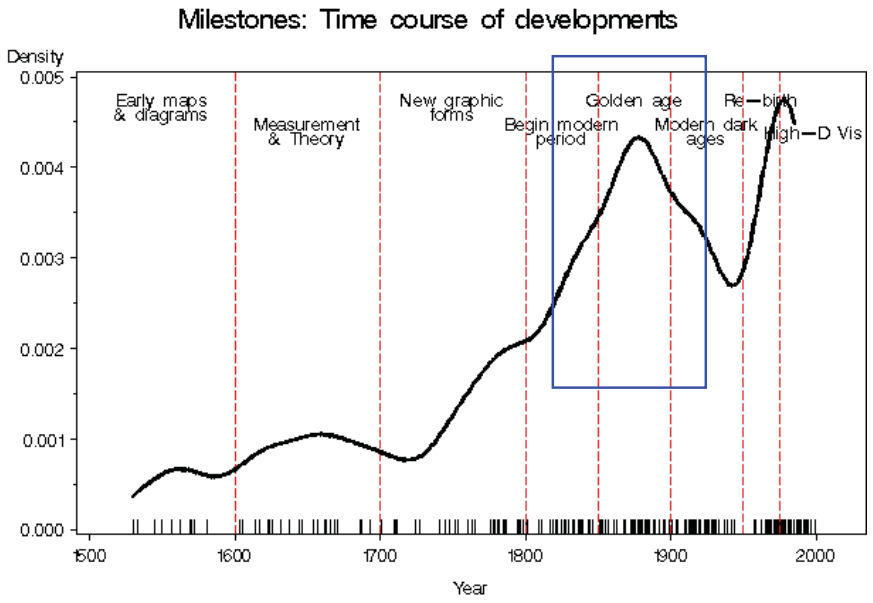


1896: Area rectangles on a map to display two variables and their product- Jacques Bertillon





Why a golden age?



Stories from the Golden Age (1850-1900)

Stories:

- A.-M. Guerry & the rise of social science
- Graphic vision of C. J. Minard
- Galton's graphical discoveries
- Statistical albums

Themes:

- Statistics: numbers of the state
- Rise of visual thinking
- Escaping flatland: 2D → 3D
- Visualization → Theory (graphic discovery)
- Data → Theory → Practice
- Graphical excellence

Big questions of the early 1800s

- Issues for European states
  - Demography: taxes, raising an army (Süssmilch, 1741)
  - "Statistik": Numbers of the state (Achenwall, 1748)
  - Social problems: crime, suicide, literacy, etc.
  - Disease epidemics, e.g., cholera
- Anthropometry: the measure of Man
  - Distributions of human characteristics
  - Birth, mortality, lifespan
- Beginnings of statistical theory and application
  - Normal dist^{n} (de Moivre, 1733)
  - *L'homme moyen* (Quetelet, 1835)

Big data of the early 1800s:

"An avalanche of social numbers"

- J.-B.J. Fourier: *Recherches statistique sur la ville de Paris* (1821-1829)
  - Massive tabulations: births, deaths (by cause), admission to insane asylums (age, sex, affliction)
- Ministry of Justice: *Compte generale (1825--)*
  - First national compilation of criminal justice data
  - All charges & dispositions, quarterly, 86 departments
- Other sources:
  - Bureau de Longitudes (illegitimate births)
  - Parent-Duchatelet (prostitution); Min. of War (desertions)
  - Suicide notes in Paris collected and analyzed for motives
- Social issues could now be addressed with **DATA**

1. A. M. Guerry and the rise of social science

Essai sur la statistique moral de la France

The launching pad of modern social science

- Presented to Academie des Sciences Français July 2, 1832
- First systematic analysis of comprehensive data on crime, suicide, and other social variables.
- Along with Quetelet (1831, 1835), established the study of "moral statistics" ↦ modern social science, criminology, sociology


ESSAI
LA STATISTIQUE
MORALE
DE LA FRANCE,
A. M. GUERRY,

Social context of crime in 1820s France

- Crime a serious concern:
  - Explosive growth in Paris
  - Widespread unemployment,
  - Emergence of "dangerous classes"
- Liberal ("philanthrope") view
  - Increase education
  - Better prison conditions, diet (bread ***and*** soup)
  - Religious instruction
- Conservative view
  - Build more prisons
  - Harsher treatment of recidivists
- Now, there was finally some DATA!

The discovery of "social facts"

Stability and Variation

Guerry's results were both compelling and startling:

- Rates of crime and suicide remained remarkably invariant over time, yet varied sytematically by region, sex of accused, type of crime, etc.
- In any given French city or department, almost the same number committed suicide, stole, gave birth out of wedlock, etc.

Year	1826	1827	1828	1829	1830	Avg
Sex			All accused (%)			
Male	79	79	78	77	78	78
Female	21	21	22	23	22	22
Age			Accused of Theft (%)			
16–25	37	35	38	37	37	37
25–25	31	32	30	31	32	31
Crime			Committed in summer (%)			
Indecent assault	.	36	36	35	38	36
Assault & battery	.	28	27	27	27	28

The discovery of "social facts"

Social laws á la physical laws

Do crime and other moral variables represent:

- structural, lawful characteristics of society, or are they
- simply indicants of individual behaviour?

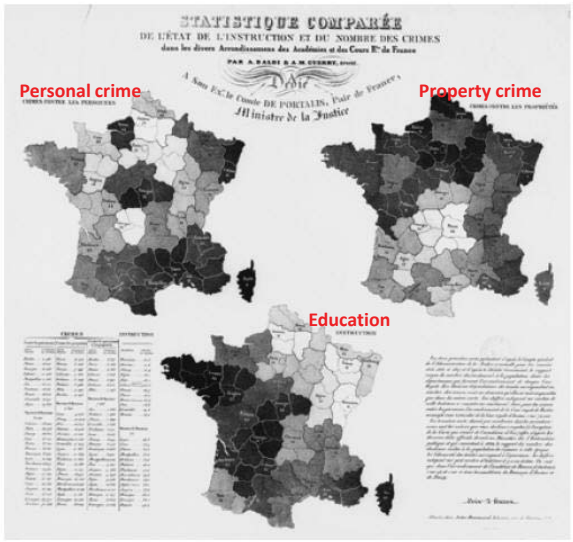
Guerry argued:

Each year sees the same number of crimes of the same degree reproduced in the same regions.(Guerry, 1833, p.10)

... We are forced to recognize that the facts of the moral order are subject, like those of the physical order, to invariable laws (Guerry, 1833, p14)

1829: *Statistique comparée de l'état de l'instruction...*

- First shaded thematic maps of crime data
- First comparative maps of social data
- ↦ crime against persons seemed inversely related to crime against property!
- Instruction: ↦ *France obscure* and *France éclairée* (Dupin, 1826)
- North of France highest in education, but also in property crime!



1833: *Essai sur la statistique morale de la France*

- Divided the 86 departments into 5 regions
- Supplemented data from the *Compte général* with:
  - Suicides in Paris, 1794–1832
  - Prostitutes in Paris (Parent-Duchâtelet)
  - Wealth (taxes per inhabitant)
  - Distribution of clergy
  - ...
- First study to use crime data to 'test' hypotheses
- Attracted widespread interest in Europe

Guerry's 1833 map of literacy in France

1833: Semi-graphic tables

Crimes against persons

- Indecent assault on adults (*viol sur des adultes*) decreases with age
- Indecent assault on children increases with age (top for 70+)
- Paricide rises to max at age 60–70


INFLUENCE DE L'AGE.

DISTRIBUTION DES CRIMES AUX DIFFÉRENS AGES, PAR PÉRIODE DE DIX ANNÉES. IV.

CRIMES CONTRE LES PERSONNES.

Figure: Ranking of crimes against persons at different ages

1864: *Statistique morale de l'Angleterre comparée...*

Dayenu!

- Proposes to replace simple "moral statistics" (tables) with "analytical statistics"
  - calculation, graphic display
  - ↦ general, abstract results
- 17 large color plates (56 × 39 cm):
  - data for France (1825–1855), England (1834–1855)
  - crimes against persons and property decomposed in various ways
  - first attempt to delineate multivariate relations among moral variables
- Voluminous data:
  - 85,564 suicide records (1836–1860), classified by motive
  - 226,224 accused of personal crime
  - numbers, in a line → 1170 meters!


STATISTIQUE MORALE

L'ANGLETERRE

LA STATISTIQUE MORALE

LA FRANCE

PAR A.-M. GUERRY

ATLAS

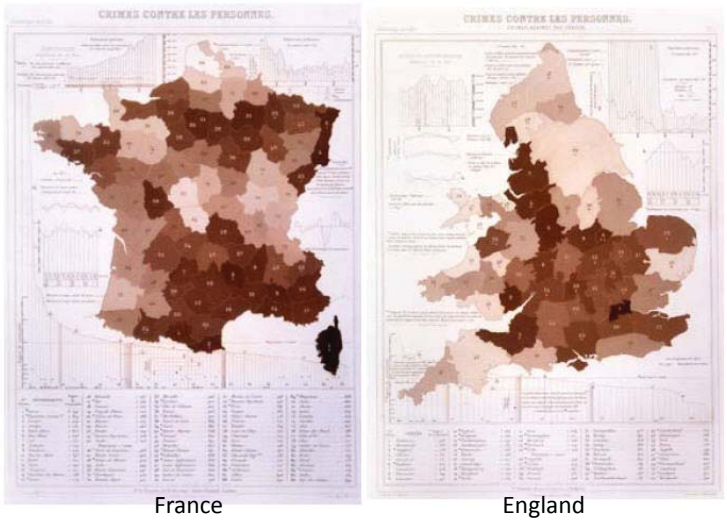
PROSPECTUS

PARIS

Further details: Friendly, M. (2007). A.-M. Guerry's *Moral Statistics of France:* Challenges for Multivariable Spatial Analysis, *Statistical Science, 22*, 368-399

1864: *Statistique morale de l'Angleterre comparée...*

Comparing France and England



France England

Crimes against persons

Statistique analytique: General causes of crime

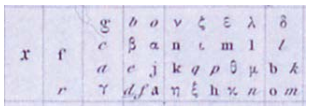
Plate XVII: M. Guerry's magnum opus

Goal:

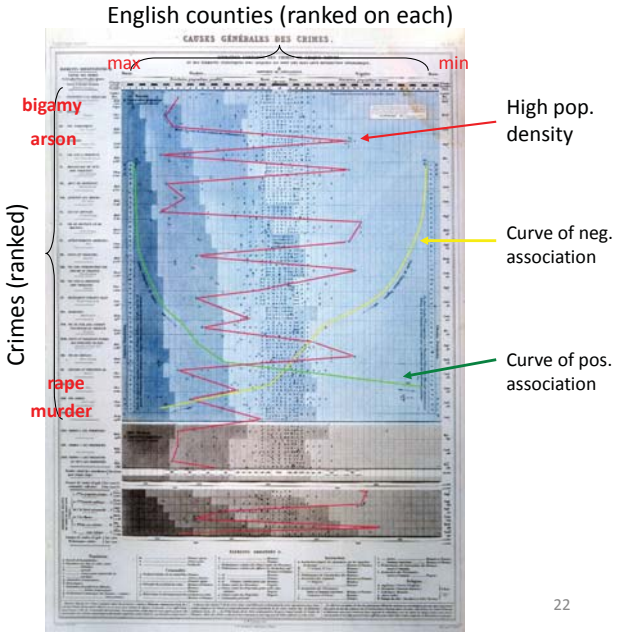
- Show multivariate factors associated with distribution of crime
- Before invention of correlation

Entries: Codes for factors

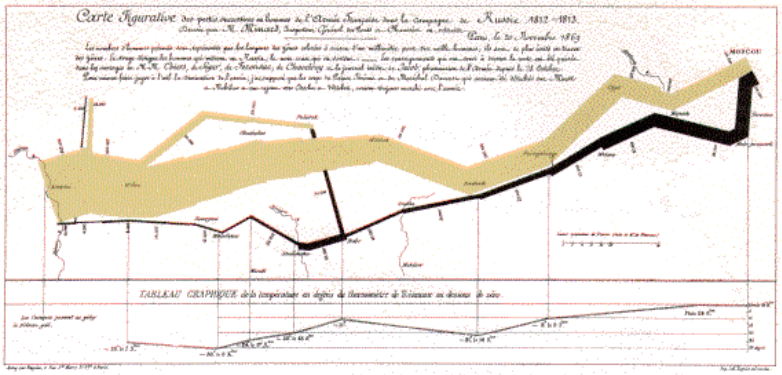
- Pop: (% Irish, domestics, …)
- Criminality: (male, young, …)
- Religion (Anglicans, dissenters, …)



English counties (ranked on each)



2. The graphic vision of C. J. Minard

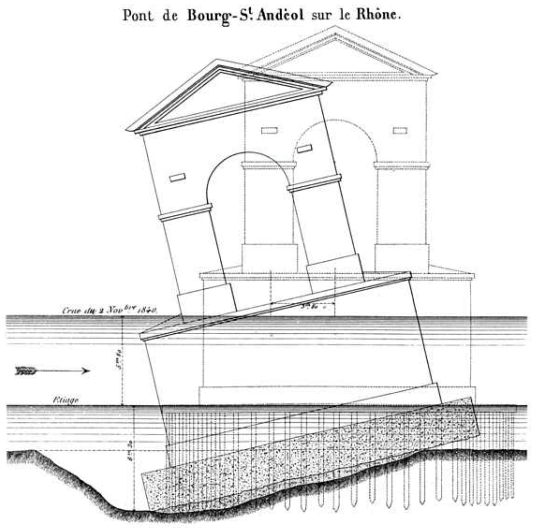


- Marey (1878): *"defies the pen of the historian in its brutal eloquence"*
- Tufte (1983): *"the best statistical graphic ever produced"*

Visual thinking, visual explanation

1840: Why did the bridge at Bourg-St. Andèol collapse?

Minard's report consisted essentially of this self-explaining diagram.



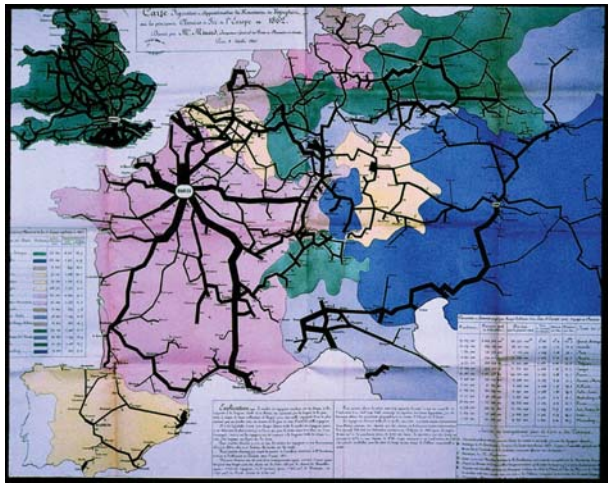
Big questions of the mid 1800s

- 1830—1860: emergence of modern French state, dawn of globalization
- Trade, commerce, transportation:
  - Where to build railroads, canals?
  - How to compete with imports/exports?
  - Visualizing changes over time, differences over space
  - → Flow maps and other graphical innovations
- These questions motivated the "Golden Age" of statistical graphics.

See: Friendly, M. (2008). The Golden Age of Statistical Graphics, *Statistical Science, 23*, 502-535

Flow maps as visual tools

Transport of passengers on the principal railroads in Europe in 1862



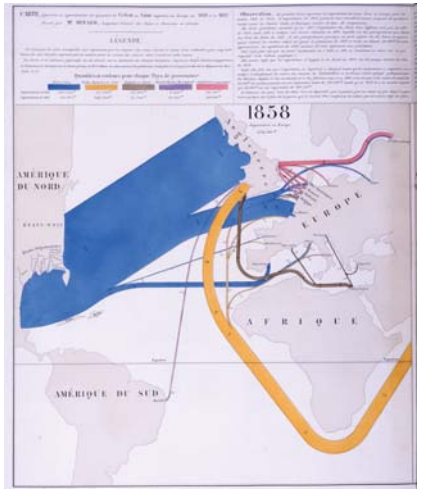
The dominant principle which characterizes my graphic tables and my figurative maps is to make immediately appreciable to the eye, as much as possible, the proportions of numeric results.

...Not only do my maps speak, but even more, they count, they calculate by the eye.

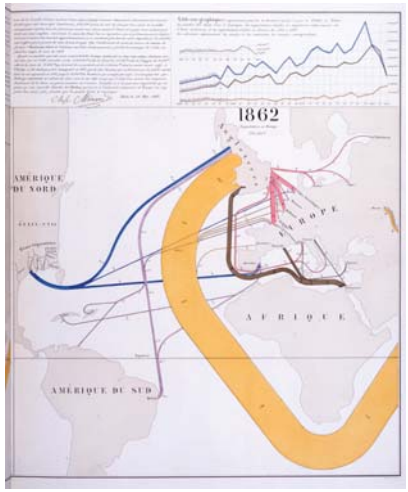
-- Minard (1862)

Effect of US civil war on cotton trade

Before



After



3. Galton's visual discoveries-

Bivariate normal correlation surface (1886)

Table 9.1 One of Galton's correlation tables

Height of the mid-parent in inches	Height of the adult child													
	<61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	>73.7
>73.0	—	—	—	—	—	—	—	—	—	—	—	1	3	—
72.5	—	—	—	—	—	—	—	1	2	1	2	7	2	4
71.5	—	—	—	—	1	3	4	3	5	10	4	9	2	2
70.5	1	—	1	—	1	1	3	12	18	14	7	4	3	3
69.5	—	—	1	16	4	17	27	20	33	25	20	11	4	5
68.5	1	—	7	11	16	25	31	34	48	21	18	4	3	—
67.5	—	3	5	14	15	36	38	28	38	19	11	4	—	—
66.5	—	3	3	5	2	17	17	14	13	4	—	—	—	—
65.5	1	—	9	5	7	11	11	7	7	5	2	1	—	—
64.5	1	1	4	4	1	5	5	—	2	—	—	—	—	—
<64.0	1	—	2	4	1	2	2	1	1	—	—	—	—	—
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14
Medians	—	—	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0	—	—

Source: Galton (1886), p. 68.

Visual smoothing → Insight

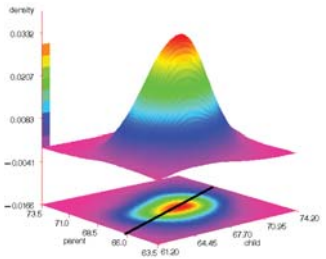
Table 9.1 One of Galton's correlation tables

Height of the mid-parent in inches	Height of the adult child <61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	>73.7
>73.0	—	—	—	—	—	—	—	—	—	—	1	3	—	—
72.5	—	—	—	—	—	—	—	1	2	7	2	4	—	—
71.5	—	—	—	—	—	1	3	4	3	5	4	2	—	—
70.5	—	—	1	—	1	5	4	3	14	7	4	1	—	—
69.5	—	—	1	5	4	3	14	14	28	11	3	5	—	—
68.5	1	—	3	3	5	10	5	27	20	14	5	—	—	—
67.5	—	3	3	5	2	12	14	15	36	8	3	1	—	—
66.5	—	3	3	5	7	11	14	29	21	7	1	—	—	—
65.5	1	9	5	7	11	11	24	17	5	2	1	—	—	—
64.5	1	1	4	4	1	5	5	—	2	—	—	—	—	—
<64.0	1	—	2	4	1	2	2	1	1	—	—	—	—	—
Totals	5	7	32	59	48	117	138	120	187	99	41	17	14	
Medians	—	—	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0	—	—

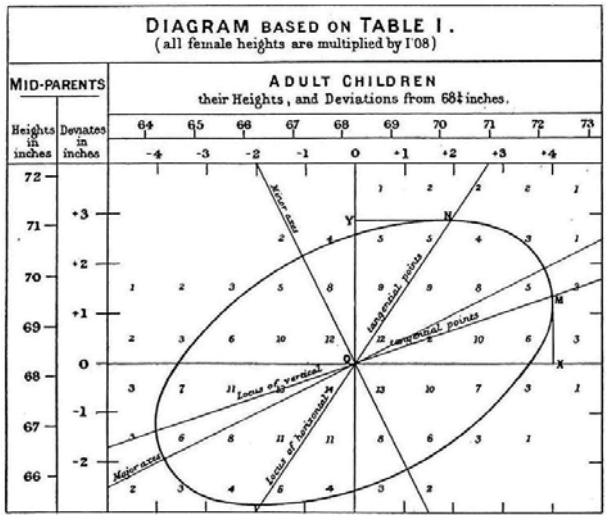
Source: Galton (1886), p. 68.

Visual insight → Theory

- Level curves are **ellipses**
- Regression lines are loci of conjugate **tangents**

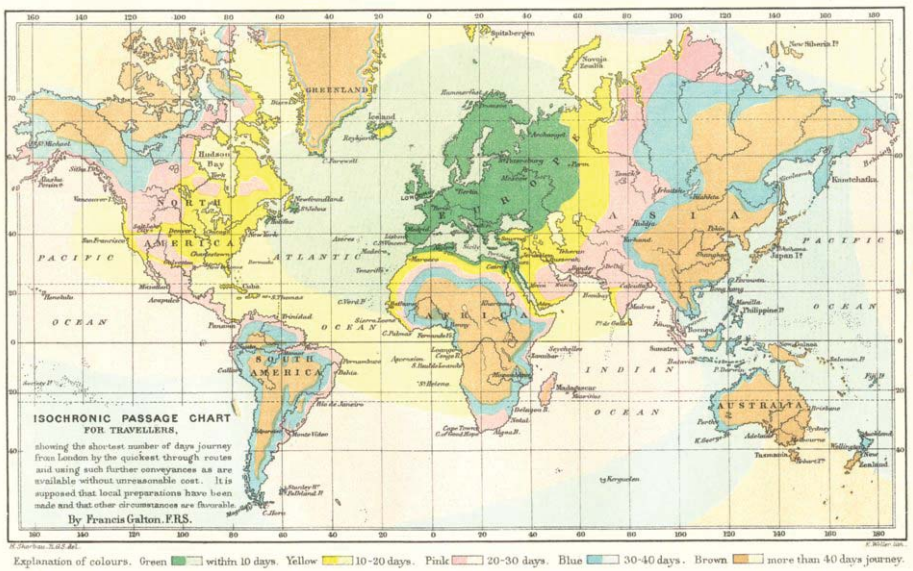


… that Galton should have evolved all this … is to my mind one of the most note-worthy scientific discoveries arising from analysis of pure observation (Pearson 1920, p37)



Galton (1886, Pl X): Smoothed contours of heights of parents and children

Galton's big data : Isochronic chart (1881)



Galton's discovery of weather patterns-

Perhaps the most notable *purely graphic* discovery ever!

METEOROGRAPHICA,

OR

METHODS OF MAPPING THE WEATHER;

ILLUSTRATED BY UPWARDS OF 600 PRINTED AND LITHOGRAPHED DIAGRAMS

REFERRING TO

THE WEATHER OF A LARGE PART OF EUROPE,

During the Month of December 1861.

By FRANCIS GALTON, F.R.S.

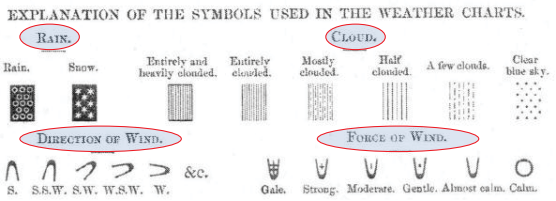
(Galton, 1863)

Method: All weather stations across Europe asked to record data 3x/day for all of Dec., 1861

Data: recordings of barometric pressure, wind dir/speed, rain, temp., cloud: 3x/day, 50 weather stations in Europe.

Graphic analysis: 3x31=93 maps, each with multivariate glyphs showing all variables

Visual ideas:
- Iconic symbols
- Multivariate glyphs (stamps!)



Visual abstraction → Patterns

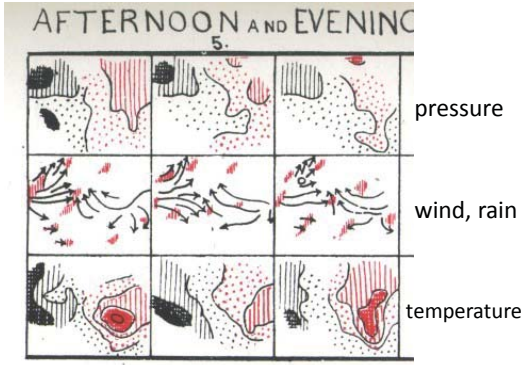
How to see patterns of geographical variation over time?
- Iconic symbols on a geographical grid
- "Small multiples:" separate graphs laid out for direct comparison

Symbols in Barometrical Charts.

Black	-	Inches. 29.95 to 29.71	Inches.	Inches. 29.70 to 29.46	Inches.	Inches. 29.45 to 29.21	Inches.	Inches. 29.20 and below.
Red	-	29.96 to 30.20		30.21 to 30.45		30.46 to 30.70		30.71 and above.

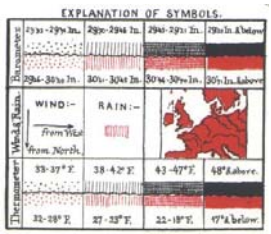
Visual abstraction → Patterns

What varies with what, over time and space?
- mini, abstract maps: vars x TOD
- iso-contours, shading to show equivalence
- arrows to show wind direction



AM noon PM

Data for Dec 5, 1861

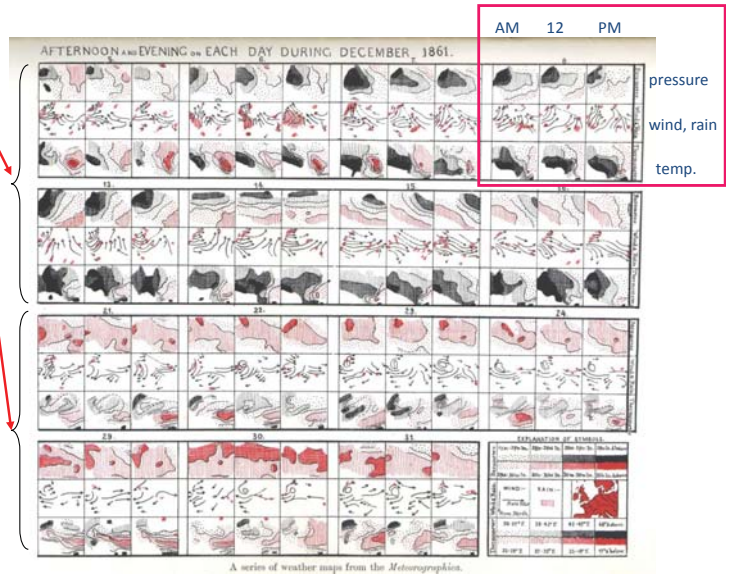


The large picture → Insight

Pattern:
Low pressure (black) in early Dec. → CCW wind
High pressure (red) in late Dec. → CW wind

Graphic: 3x3x31 grid, mapping {pressure, wind/ rain, temperature} x {AM, 12, PM} x day {1:31}

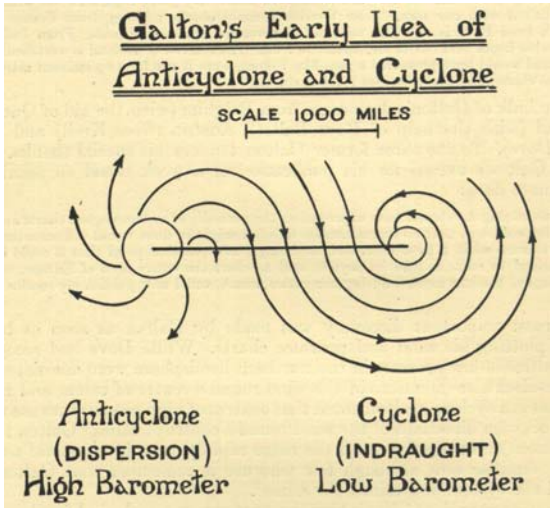
(try this with your software!)



Visual insight → Theory

Visual insight from 93 (3x31) high-D graphs:
- Changes in wind dir w/ pressure over time
- → Winds revolve inwardly (CCW) in low pressure areas– as in a cyclone;
- → revolve outwardly (CW) in high pressure areas– "anti-cyclone"

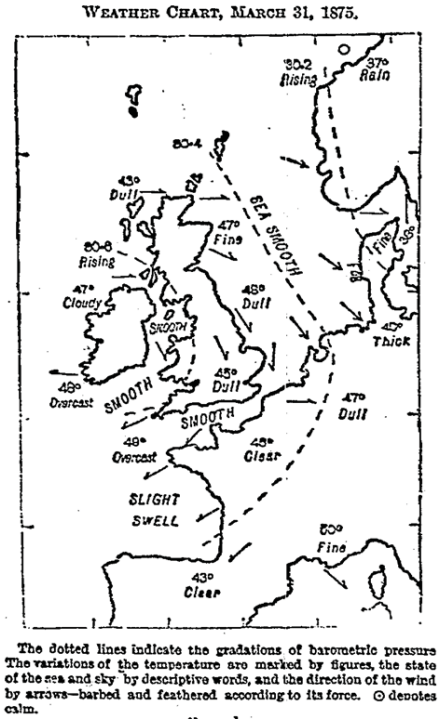
Theory:
- Explained by Dove's 'Law of Gyration'
- Prediction: reversed pattern (CW/CCW) in southern hemisphere – confirmed!



Theory → Practice

The first modern weather map, *London Times*, Apr. 1, 1875

Galton did for weathermen what Kepler did for Tycho Brahe. This is no small accomplishment. (Wainer 2005)



4. Statistical atlases: Data → practice, national identity & graphical excellence

- Collection of gov't statistics on pop., trade, moral & political issues widespread in Europe & US, starting ~ 1820
- Statistical albums ~ 1870—1910
  - France: *Album de Statistique Graphique*: 1879-1899
  - USA: Census atlases: 1870/80/90
  - Germany: local albums (Berlin, Frankfurt, etc.)
  - Switzerland: *Atlas graphique de la Suisse:1897, 1914*
  - *Others: Latvia, Romania, Bulgaria, etc.*

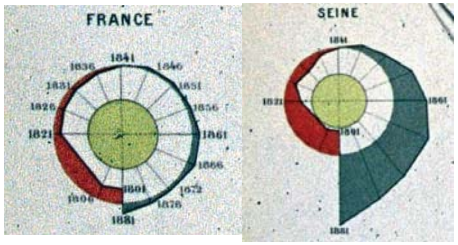
Album de statistique graphique

- Published by the *Statistical Graphics Bureau*, Ministry of Public Works, Émile Cheysson, director
- 18 volumes: 1879-1899, 12—34 plates each, ~ 11"x15" pages
- Graphic forms:
  - Flow maps (simple, double, multi)
  - Pie maps, star, radial, polar time-series, proportional circles
  - Mosaic maps, anamorphic maps, planetary diagrams
  - Choropleth, bi-polar scales
  - Charts: line, bar, time-series
- Formats: 1x1, 2x1, 2x2, 3x2, 5x3!…
- Themes:
  - Recurrent: railroads, navigation, transport--- B&B
  - Occasional: agriculture, Paris, expositions, …
- **Pinnacle of the Golden Age**: exquisite sampler of all known graphic forms!

Album de statistique graphique

Spiral time-series on a map

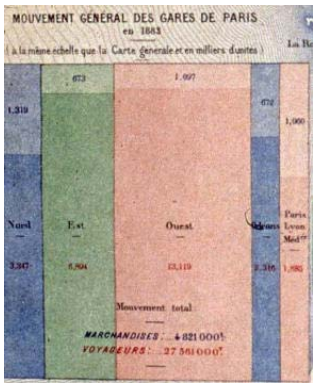
Changes in the population of France from 1801—1881, by department [Album, 1881, plate 25]

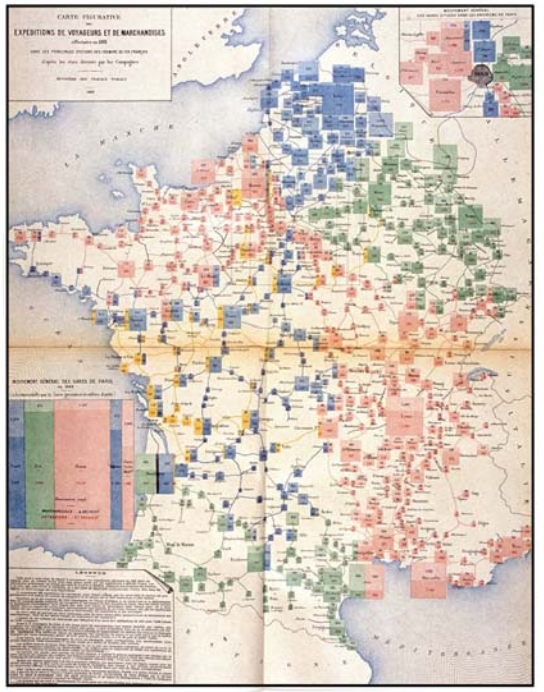


Recursive multi-mosaic map

Distribution of **passengers** and **goods** from the Paris railways to the rest of France [Album, 1884, pl. 11]

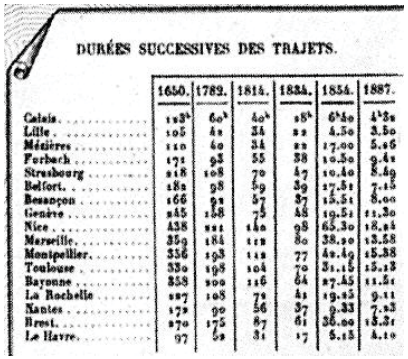
(The image that launched my interest in the history of data vis.)

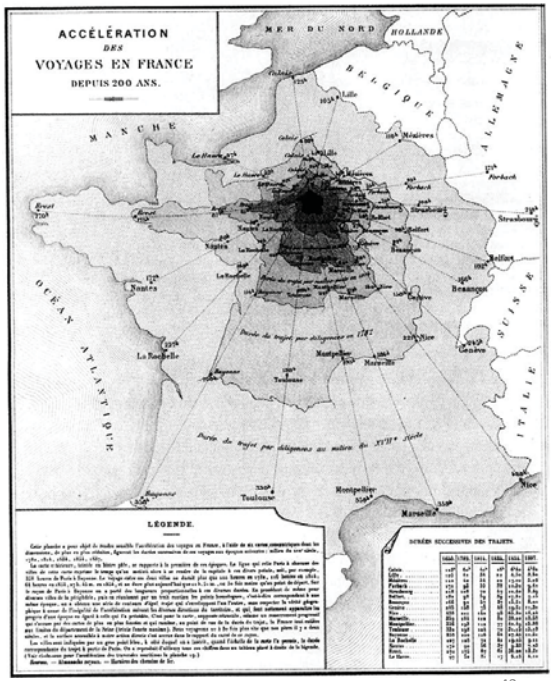




Anamorphic map

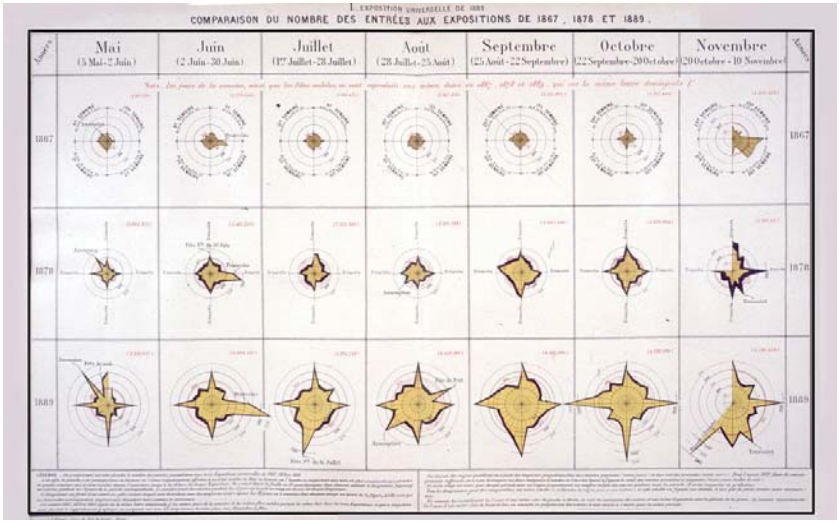
Shrinking France to show change in travel time over 200 years [Album, 1888, plate 8]





Two-way table of star/radar diagrams

Attendance at the universal expositions in 1867, 1878, 1889 (rows), by month (cols) and days (rays). [Album, 1889, plate 21]



Currently trending…

Where are we now?

- Everyone wants in on the Data Vis bandwagon
  - InfoVis: Some spectacular, mostly bad
  - High-D scientific data visualization
  - Data journalism, public-interest graphics
- Massive data bases, often crowd sourced
  - eBird: bird migration
  - genomics → "omics"
- Dynamic, interactive graphics
  - animation, time-motion charts
  - query / drill-down to detailed views
- Spatial data analysis & visualization
- Network visualization

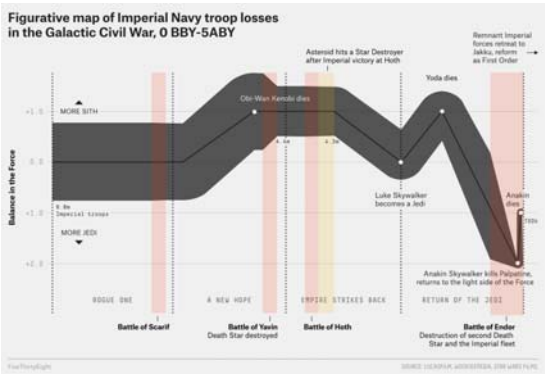
InfoVis: Minard still lives

The epic *Star Wars* saga, all in one chart

A visual history of the Galactic Civil War, pitting the Imperial Navy (sith) against the ragtag Rebel Alliance (jedi)

As the author, Walt Hickey says, "here is why you should never invade Hoth in winter".

Data sources: Wookieepidia, http://starwars.wikia.com



Source: http://fivethirtyeight.com/features/star-wars-in-one-chart/

Data journalism: Measels and vaccines

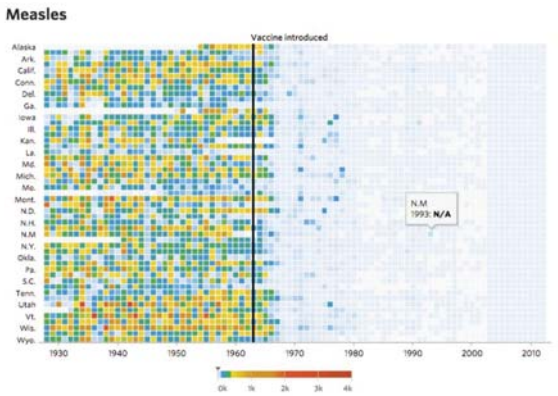
Visualizing the impact of health policy interventions

In 2015 Tynan DeBold & Dov Friedman in the *Wall Street Journal* tried to show the effect of the introduction of vaccination programs in the US states on disease incidence, using color-coded heat maps for a variety of diseases

The long time series ~70 years made this work.

The heat map color scale is not exemplary, but the message is still clear: disease incidence declined after vaccines were introduced.

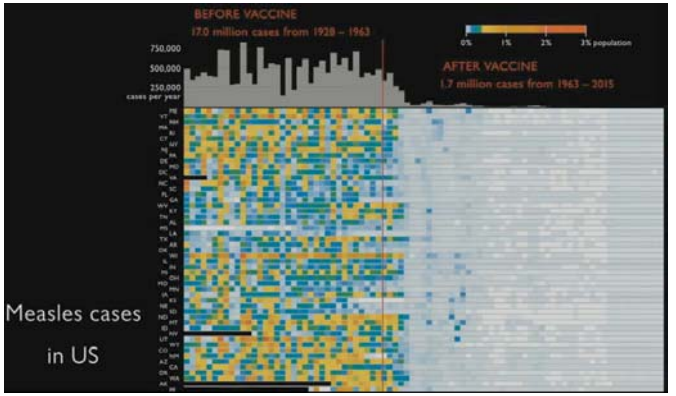
The images are "interactive," in the weak sense that tool-tips are shown on mouse movement.



Source: http://graphics.wsj.com/infectious-diseases-and-vaccines/

Making the message more explicit

What you should want to show here in an Info graphic is the overall impact of vaccination on measels
Ed Tufte did this by adding a histogram at the top showing total # of cases by year



Visualization over time and space

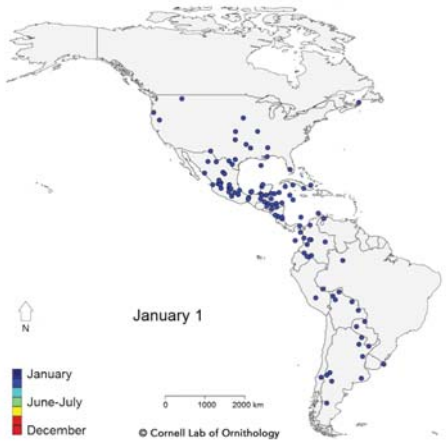
Migration patterns of birds in N/S America

- massive, crowd-sourced eBird data base
- how to visualize? Then: how to model statistically? explanation?

"We used millions of observations from the eBird citizen-science database".

"After tracing the migration routes of all these species and comparing them, we concluded that a combination of geographic features and broad-scale atmospheric conditions influence the choice of routes used during spring and fall migration."

Each dot represents a single bird species; the location represents the average of the population for each day of the year.

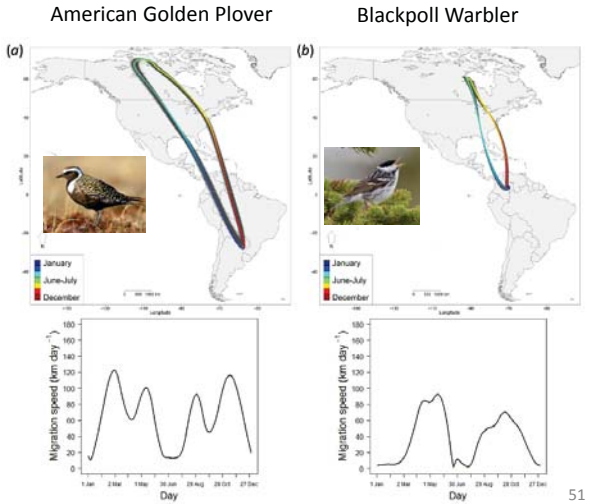


Visualization → Model

Modeling population-level daily migration trajectories shows clear differences among species

Technical note: this used a generalized linear mixed model, with species as a random effect.

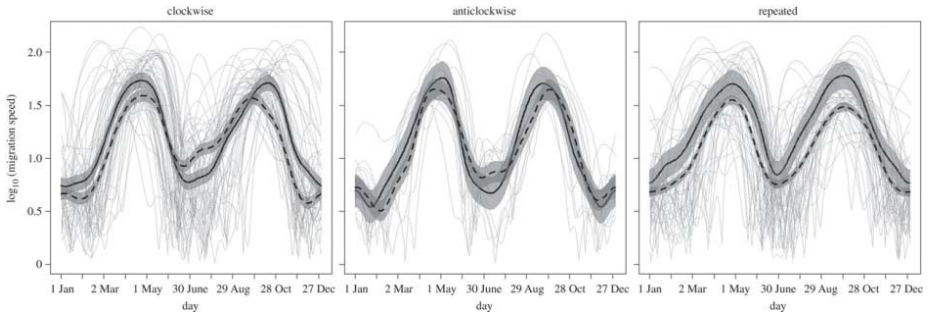
More importantly, calculating migration speed showed big differences in seasonal patterns across species.



Model → Explanation

Analysis allowed classifying species into six migration patterns

- A generalized additive mixed model for migration speed (species as a random effect) gave fitted estimates.
- These had a clear interpretation in terms of adaptive strategies to deal with greater risks from transoceanic migration, plus seasonal environmental and atmospheric constraints



Migration speeds for 118 bird species for 2002–2014 summarized within six migration categories.

- Grey lines: individual species.
- Solid lines: transoceanic migrants.
- Dashed lines: terrestrial migrants

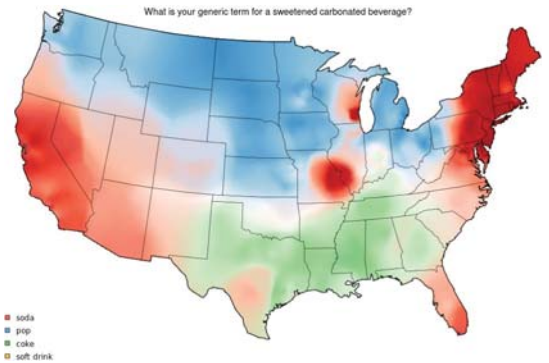
Spatial visualization

Linguistics: Food dialect maps– visualizing how people speak

In the *Cambridge Online Survey of World Englishes*, Bert Vaux and Marius L. Jøhndal surveyed 11,500 people to study the ways people use English words.

NC State Univ. student Joshua Katz turned the US data into shaded kernel density maps.

soda vs. pop?



Take the survey: http://www.tekstlab.uio.no/cambridge_survey
Programming in R: http://blog.revolutionanalytics.com/2013/06/r-and-language.html

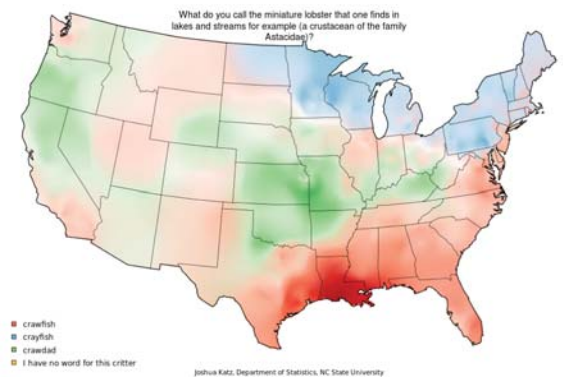
Spatial visualization

Linguistics: Food dialect maps– visualizing how people speak

crawfish, crawfish, crawdad?

A k-nearest neighbor kernel density estimate over (x,y) locations gives a smoothed & interpretable display of the choice probabilities.

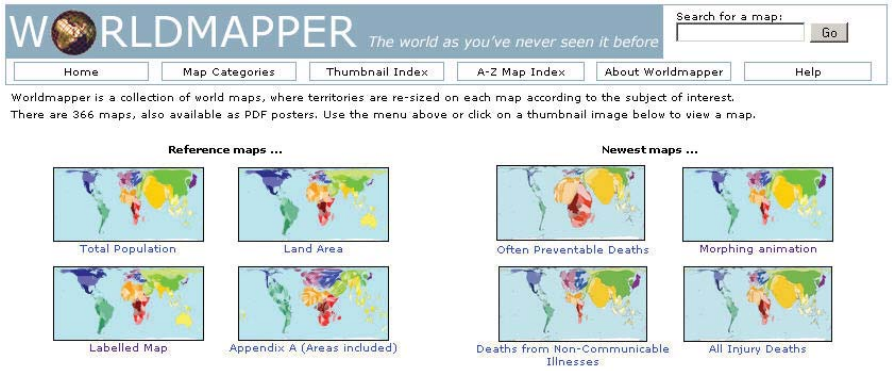
Regional differences are quite apparent.



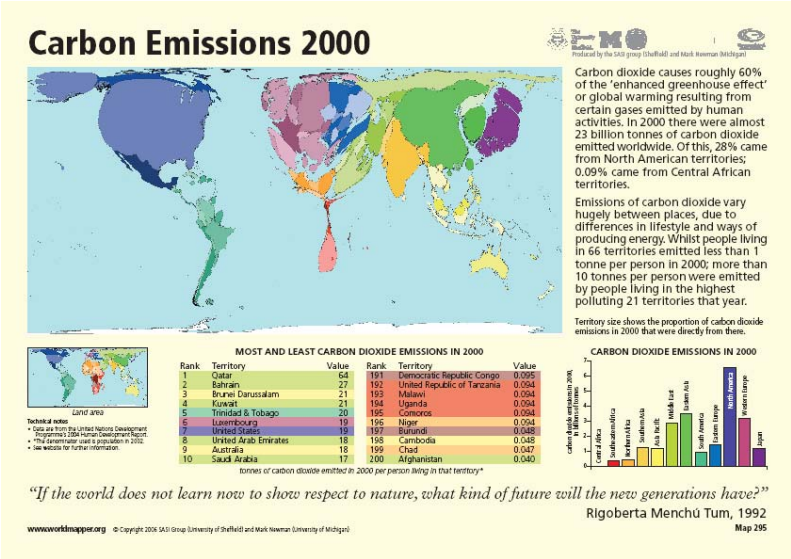
Worldmapper: The world in cartograms

How to visualize social, economic, disease, … data for geographic units?

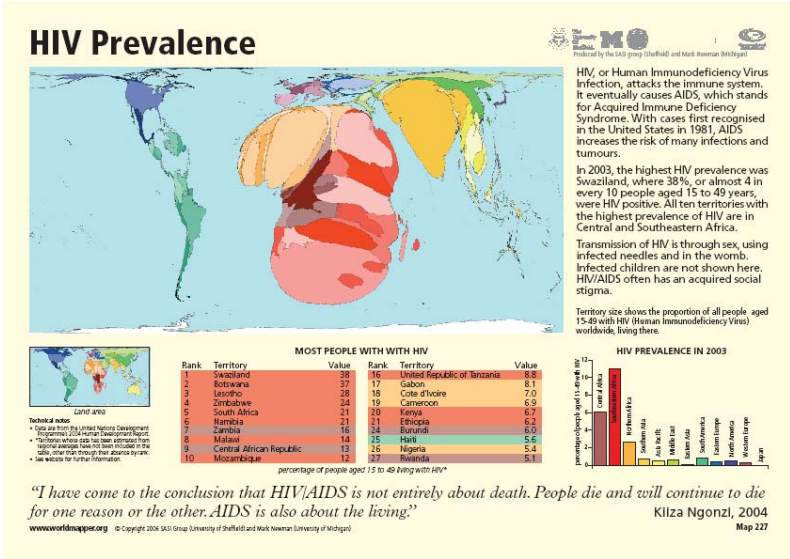
worldmapper.com : **cartograms: area ~ variable of interest** (700+ maps)



Worldmapper: The world in cartograms



Worldmapper: The world in cartograms

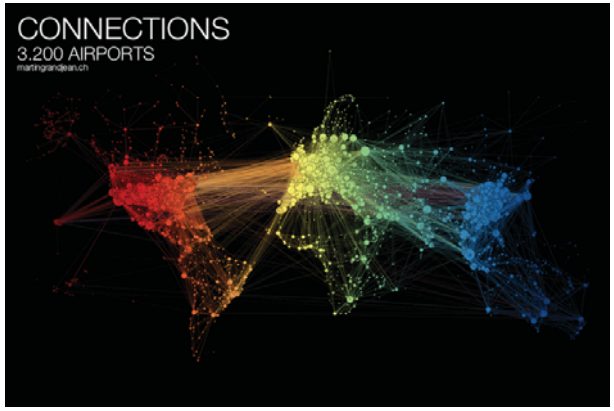


Network visualization

Once the domain of mathematicians & computer scientists, graph theory and network visualization turn out to have surprising & interesting applications.

Animated demo by Martin Granjean showing transport of passengers from/to world airports.

It illustrates the difference between geography & force-directed layout to focus on volume & connections



From: http://www.martingrandjean.ch/connected-world-air-traffic-network/
See more: https://flowingdata.com/2016/05/31/air-transportation-network/