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

Context: Milestones Project

Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization
An illustrated chronology of innovations by Michael Friendly and Daniel J. Denis

Timeline
This page provides a graphic overview of the events in the history of data visualization that one call "milestones." These milestones are shown below in the form of an interactive timeline. The timeline is divided into two vertical sections. You can skip each section left or right to see milestones of different time periods. You can also click one of the links at the bottom of the timeline to jump to a particular epoch.

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

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

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

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?
Why a golden age?

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 distn (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

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
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 Académie des Sciences Françaises 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

---

**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 systematically 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.

<table>
<thead>
<tr>
<th>Year</th>
<th>1826</th>
<th>1827</th>
<th>1828</th>
<th>1829</th>
<th>1830</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>79</td>
<td>79</td>
<td>78</td>
<td>77</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>Female</td>
<td>21</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-25</td>
<td>37</td>
<td>35</td>
<td>38</td>
<td>37</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>25-25</td>
<td>31</td>
<td>32</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>Crime</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indecent assault</td>
<td>.</td>
<td>36</td>
<td>36</td>
<td>35</td>
<td>38</td>
<td>36</td>
</tr>
<tr>
<td>Assault &amp; battery</td>
<td>28</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>28</td>
</tr>
</tbody>
</table>

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, p.14)
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

1864: Statistique morale de l’Angleterre comparée...

- 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!

1864: *Statistique morale de l'Angleterre comparée*...
Comparing France and England

Plate XVII: M. Guerry’s magnum opus

**Crimes against persons**

<table>
<thead>
<tr>
<th>Crime</th>
<th>France</th>
<th>England</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bigamy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arson</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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, …)

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”

1840: Why did the bridge at Bourg-St. Andéol 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.


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

<table>
<thead>
<tr>
<th>Height of the mid-parent in inches</th>
<th>Height of the adult child</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;61.7</td>
<td>62.2</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>&gt;73.0</td>
<td>—</td>
</tr>
<tr>
<td>72.5</td>
<td>—</td>
</tr>
<tr>
<td>71.5</td>
<td>—</td>
</tr>
<tr>
<td>70.5</td>
<td>1</td>
</tr>
<tr>
<td>69.5</td>
<td>—</td>
</tr>
<tr>
<td>68.5</td>
<td>1</td>
</tr>
<tr>
<td>67.5</td>
<td>—</td>
</tr>
<tr>
<td>66.5</td>
<td>—</td>
</tr>
<tr>
<td>65.5</td>
<td>1</td>
</tr>
<tr>
<td>64.5</td>
<td>1</td>
</tr>
<tr>
<td>&lt;64.0</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>5</td>
</tr>
<tr>
<td>Medians</td>
<td>—</td>
</tr>
</tbody>
</table>

Source: Galton (1886), p. 68.
Table 9.1  One of Galton’s correlation tables

<table>
<thead>
<tr>
<th>Height of the mid-parent in inches</th>
<th>Height of the adult child</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;73.7</td>
<td></td>
</tr>
<tr>
<td>72.5</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td></td>
</tr>
<tr>
<td>71.5</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td></td>
</tr>
<tr>
<td>70.5</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
</tr>
<tr>
<td>69.5</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
</tr>
<tr>
<td>68.5</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td></td>
</tr>
<tr>
<td>67.5</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
</tr>
<tr>
<td>66.5</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
</tr>
<tr>
<td>65.5</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td></td>
</tr>
<tr>
<td>64.5</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
</tr>
<tr>
<td>&lt;64</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td></td>
</tr>
<tr>
<td>Medians</td>
<td></td>
</tr>
</tbody>
</table>

Source: Galton (1886), p. 68.

Galton’s big data : Isochronic chart (1881)

Perhaps the most notable *purely graphic* discovery ever!

Galton’s discovery of weather patterns-

---

`Visual smoothing → Insight`

Table 9.1  One of Galton’s correlation tables

<table>
<thead>
<tr>
<th>Height of the mid-parent in inches</th>
<th>Height of the adult child</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;73.7</td>
<td></td>
</tr>
<tr>
<td>72.5</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td></td>
</tr>
<tr>
<td>71.5</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td></td>
</tr>
<tr>
<td>70.5</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
</tr>
<tr>
<td>69.5</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
</tr>
<tr>
<td>68.5</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td></td>
</tr>
<tr>
<td>67.5</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
</tr>
<tr>
<td>66.5</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
</tr>
<tr>
<td>65.5</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td></td>
</tr>
<tr>
<td>64.5</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
</tr>
<tr>
<td>&lt;64</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td></td>
</tr>
<tr>
<td>Medians</td>
<td></td>
</tr>
</tbody>
</table>

Source: Galton (1886), p. 68.

Galton’s big data : Isochronic chart (1881)

Perhaps the most notable *purely graphic* discovery ever!

---

`Visual insight → Theory`

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

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

... 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)
**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

**The large picture → Insight**

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

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)

(Data for Dec 5, 1861)

[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!
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: Wookieepedia, [http://starwars.wikia.com](http://starwars.wikia.com)

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.


Making the message more explicit

What you should want to show here in an Info graphic is the overall impact of vaccination on measles

Ed Tufte did this by adding a histogram at the top showing total # of cases by year

Migration patterns of birds in N/S America

- massive, crowd-sourced eBird database
- 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.

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.

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.

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)

---

Carbon Emissions 2000

Carbon dioxide causes roughly 60% of the ‘human-induced’ greenhouse effect or global warming resulting from various gases emitted by human activities. It is the most important contributor to global warming, causing about 23 billion tons of carbon dioxide annually worldwide. 50%+ comes from North American territories; 0.6% come from Central & South American territories.

Emissions of carbon dioxide vary greatly between places, due to differences in lifestyle and energy consumption patterns. For example, living in 60 territories emitted less than 1 tonne per person in 2000. More than 10 tonnes per person were emitted by people living in the highest polluting 21 tall nations that year.

---

HIV Prevalence

HIV or Human Immunodeficiency Virus infection, affects the immune system and can lead to AIDS, which stands for Acquired Immunodeficiency Syndrome. AIDS is caused by the human immunodeficiency virus (HIV). In the United States in 1991, AIDS was one of the 10 leading causes of death. In 2001, the highest HIV prevalence was found in South Africa, where 3.3% of the population is infected. The highest prevalence of HIV in the world is in Eastern and South-Eastern Africa. Transmission of HIV through sexual activity, injection drug use and in mother to child transmission are not shown here. HIV/AIDS infection has an advanced social stigma. The virus also shows the contagion of people, aged 15 through 49, who have stress and anxiety from HIV/AIDS diagnosis.
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/](http://www.martingrandjean.ch/connected-world-air-traffic-network/)
See more: [https://flowingdata.com/2016/05/31/air-transportation-network/](https://flowingdata.com/2016/05/31/air-transportation-network/)