History of Data Visualization

Michael Friendly
Psych 6135

http://euclid.psych.yorku.ca/www/psy6135
Outline

• Overview:
  ▪ Roles of graphics in scientific discovery
  ▪ Visualizing history: The *Milestones Project*

• Milestones tour of the history of data vis
  ▪ Pre-history of visualization
  ▪ The first statistical graph
  ▪ The Big Bang: William Playfair
  ▪ Influence of data, technology & visual thinking

• Other topics (later):
  ▪ Moral statistics: the birth of social science
  ▪ Graphs in the public interest: Nightingale, Farr and Snow
  ▪ The Golden Age of statistical graphics
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
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?
Dreams and snakes
August Kekulé (1862) discovers the structure of benzene in a dream

Tree of evolution
Darwin (1859) imagines generations of species

Solitaire and the periodic table
Mendeleev (1869) organized chemical elements after a mental image of cards on a table.

See: https://medium.com/@michael.friendly/visual-thinking-graphic-discoveries-128468677592
How to visualize travel? A route map!

In 1675, chartmaker John Ogliby told a graphic story of what you would see on a travel from London to Land’s End.
How to visualize history? A route map!

In 2017, graphic storyteller RJ Andrews adopted Ogilby’s form to show the history of data visualization.

The online version, https://history.infowetrust.com/ is fully interactive, with details about the images on this journey.
The web site: [http://datavis.ca/milestones](http://datavis.ca/milestones) has an interactive timeline, allowing different kinds of search
Milestones: Content Overview

*Every picture has a story – Rod Stewart*

- **c. 550 BC**: The first world map? (Anaximander of Miletus)
- **1669**: First graph of a continuous distribution function (Gaunt's life table) – Christiaan Huygens.
- **1701**: First contour map - Edmund Halley
- **1801**: Pie chart, circle graph - William Playfair
- **1896**: Bivariate map - Jacques Bertillon
- **1924**: Pictograms - Otto Neurath
- **1991-1996**: Interactive data visualization systems (Xgobi, ViSta)
Milestones Tour: Epochs

New graphic forms

Beginning of modern graphics

Golden Age of data graphics

Modern Dark Ages

Re-birth

High-D data vis

Early maps & diagrams

Measurement & Theory

Beginning of modern graphics

Modern Dark Ages

Re-birth

High-D data vis
Historical information, suitably organized can be treated as data, and analyzed. This plot shows a smoothed frequency distribution of 248 milestones items over time, in relation to the named time periods.
Prehistory of visualization

Lascaux Cave, ~ 15000 BCE, the “Sistine Chapel of pre-historic art”
Lascaux: What were they thinking?

- Visual features:
  - show perspective, a sense of motion, rich use of color & texture

- What was the purpose?
  - Hunting success? NO (they hunted reindeer)
  - mostly symbolic – visual language, story of communal myths

- How to understand them?
  - A cognitive revolution: evidence for the modern human mind in Cro Magnon man
  - inner vision, visual thinking, mental imagery– a gleam in the mind’s eye

- Other cave art [20000BC – 10000BC]: Altamira (Spain); Chauvet (France), Cueva de las Manos (Argentina), ...
Prehistory: Diagrams, graphic stories

Early Egyptian animated graphic diagram
Wrestling scene on east wall, tomb of Baqt at Beni Hasan (ca. 2000 BCE).

A visual explanation of a wrestling match

Anticipates modern graphic novels

Why? Perhaps Baqt’s last lesson as a wrestler in his youth and later as a coach.
Pre 17th C.: Early maps & diagrams

c. 550 BC: The first world map? (Anaximander of Miletus)

1305: Mechanical diagram of knowledge - Ramon Llull, Spain

1350: Bar graph of theoretical function N. Oresme, France

1375: Catalan Atlas, an exquisitely beautiful visual cosmography, perpetual calendar, and thematic representation of the known world - Abraham Cresques, Spain
1305: Mechanical diagram of knowledge - Ramon Llull, Spain

*Tree of porphyry:* Aristotle’s categories of knowledge (center)

- Left: questions
- Right: rotating disks → answers
1375: Catalan Atlas, an exquisitely beautiful visual cosmography, perpetual calendar, and thematic representation of the known world- Abraham Cresques, Majorca, Spain [BNF: ESP 30]
1600-1699: Measurement and Theory

- The 17th century saw growth in theory and the dawn of attempts at visualization.
- Featured in this were:
  - the rise of analytic geometry: \((x, y)\) coordinates (Descartes),
  - theories of errors of measurement: astronomical observations (Laplace)
  - the birth of probability theory-- games of chance, annuities (Fermat, DeMoivre, ...),
  - automatic graphic recording (Scheiner)
  - the first graphical representations of **statistical** data (van Langren)
1600-1699: Measurement and Theory

1626: Visual representations used to chart the changes in sunspots over time—Christopher Scheiner

1669: First graph of a continuous distribution function (Gaunt's life table)—Christiaan Huygens.

1644: First visual representation of statistical data—M.F. van Langren, Spain

1693: First use of areas of rectangles to display probabilities of independent binary events—Edmund Halley, England
1611: Galileo records **movement** of sunspots over time (*Three letters on sunspots*, 1613)

**Visual ideas:**
- Animated graphic
- “Small multiples”
- Allows comparison
- Self-explaining diagram

A+ for info design!
1626: Christoph Scheiner invents helioscope & camera obscura to record sunspots
(*Rosa Ursina sive Sol*, 1626-1630)
Sunspots: Great graph, wrong theory

1626: Christopher Scheiner’s graph of changes in sunspots over time.
• “small multiples”
• allows comparison
• multiple legends
• A+ for info design!

He argued (incorrectly) that these were evidence for solar satellites.

The idea of graphs for visualizing phenomena had arrived.
Why the 1st statistical graph got it right

1644: First visual representation of statistical data: determination of longitude between Toledo and Rome- Michael Florent van Langren, Spain

Actual distance=16°30’

Estimated distance
What else could he have done?

- What would occur to men of his time to convey a message to the king?
- ... he could used a table have sorted by year to establish priority (or show change).
• ... he could have sorted by *name*, to show *authority*.

• ... he could have sorted by *longitude* to show the *range*.

### Sorted by Authority

<table>
<thead>
<tr>
<th>Name</th>
<th>Longitude</th>
<th>Year</th>
<th>Where</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argelius, A.</td>
<td>28.0</td>
<td>1541</td>
<td></td>
</tr>
<tr>
<td>Brahe, T.</td>
<td>21.5</td>
<td>1578</td>
<td>Denmark</td>
</tr>
<tr>
<td>Clavius, C.</td>
<td>26.5</td>
<td>1567</td>
<td>Germany</td>
</tr>
<tr>
<td>Ianfonius, G.</td>
<td>17.7</td>
<td>1501</td>
<td></td>
</tr>
<tr>
<td>Lantsbergius, P.</td>
<td>21.1</td>
<td>1530</td>
<td></td>
</tr>
<tr>
<td>Maginus, A.</td>
<td>29.8</td>
<td>1582</td>
<td>Italy</td>
</tr>
<tr>
<td>Mercator, G.</td>
<td>19.6</td>
<td>1567</td>
<td>Flanders</td>
</tr>
<tr>
<td>Organus, D.</td>
<td>30.1</td>
<td>1601</td>
<td></td>
</tr>
<tr>
<td>Ortonius</td>
<td>26.0</td>
<td>1542</td>
<td>France</td>
</tr>
<tr>
<td>Ptolomeus, C.</td>
<td>27.7</td>
<td>150</td>
<td>Alexandria</td>
</tr>
<tr>
<td>Regiomontanus, I.</td>
<td>25.4</td>
<td>1471</td>
<td>Germany</td>
</tr>
<tr>
<td>Schonerus, I.</td>
<td>20.8</td>
<td>1536</td>
<td>Germany</td>
</tr>
</tbody>
</table>

### Sorted by Longitude

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Name</th>
<th>Year</th>
<th>Where</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.7</td>
<td>G. Ianfonius</td>
<td>1501</td>
<td></td>
</tr>
<tr>
<td>19.6</td>
<td>G. Mercator</td>
<td>1567</td>
<td>Flanders</td>
</tr>
<tr>
<td>20.8</td>
<td>I. Schonerus</td>
<td>1536</td>
<td>Germany</td>
</tr>
<tr>
<td>21.1</td>
<td>P. Lantsbergius</td>
<td>1530</td>
<td></td>
</tr>
<tr>
<td>21.5</td>
<td>T. Brahe</td>
<td>1578</td>
<td>Denmark</td>
</tr>
<tr>
<td>25.4</td>
<td>I. Regiomontanus</td>
<td>1471</td>
<td>Germany</td>
</tr>
<tr>
<td>26.0</td>
<td>Orontius</td>
<td>1542</td>
<td>France</td>
</tr>
<tr>
<td>26.5</td>
<td>C. Clavius</td>
<td>1567</td>
<td>Germany</td>
</tr>
<tr>
<td>27.7</td>
<td>C. Ptolomeus</td>
<td>150</td>
<td>Egypt</td>
</tr>
<tr>
<td>28.0</td>
<td>A. Argelius</td>
<td>1541</td>
<td></td>
</tr>
<tr>
<td>29.8</td>
<td>A. Maginus</td>
<td>1582</td>
<td>Italy</td>
</tr>
<tr>
<td>30.1</td>
<td>D. Organus</td>
<td>1601</td>
<td></td>
</tr>
</tbody>
</table>
Only a graph shows...

- central location
- bias
- name labels—avoiding overplotting
- wide variability
- clustering, detached observations

The 18th century witnessed the germination of the seeds of visualization & visual thinking, planted earlier. 

Map-makers began to try to show more than just geographical position-- the beginnings of *thematic mapping* of physical quantities

- topographical maps
- iso- contour maps

New graphic forms were invented:

- bar chart,
- line chart,
- timelines

\[ \text{The Big Bang} \]
1700-1799: New graphic forms

1701: Isobar map, lines of equal magnetic declination – Edmund Halley

1765: Historical time line (life spans of famous people) Joseph Priestley

1782: First topographical map – Marcellin du Carla-Boniface

1786: Bar chart, line graphs of economic data – William Playfair
1769: Visualization of the history of civilizations & empires over ~3000 years --Joseph Priestley
1800-1849: Beginning of modern data graphics

• The first half of the 19th century witnessed an explosive growth in statistical graphics and thematic mapping
  ▪ Polar coordinates, log axes
  ▪ Shaded (choropleth) maps of social data (literacy, crime)
• The birth of data: widespread national collection of data on social and medical issues
  ▪ France: data on crime, literacy, prostitution, ... collected centrally
  ▪ England: Births, deaths, disease mortality collected by Registrar General
1800-1849: Beginning of modern data graphics

1801: Pie chart, circle graph invented- William Playfair

1819: First modern statistical map (illiteracy in France)- Charles Dupin

1843: Wind-rose (polar coordinates)- L. Lalanne

1844: variable-width, divided bars, area ~ cost of transport- C. J. Minard
1801: Pie chart, circle graph invented- William Playfair
1844: *Tableau-graphique*: variable-width, divided bars, area ~ cost of transport- Charles Joseph Minard
By the last half of the 19th century the conditions for rapid growth of visualization had been established:

- widespread data collection for planning, commerce, social theory
- the beginnings of statistical theory and visual thinking
- a wide range of graphic forms, reasonably well understood
- technology:
  - lithography and color printing
  - automatic recording devices
  - calculation: machines & graphical calculators

The result was a perfect storm-- among the most exquisite graphics ever produced.
1850-1900: Golden Age

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
E.-J. Marey: *La Méthode Graphique*

- How to make human and animal motion subject to precise scientific study?
- e.g., aerial locomotion of flying insects & birds
  - What is the frequency of wings of different species?
  - What are the mechanisms of wings to produce lift and forward motion?

A harness, designed to register the trajectory, force and speed of a bird’s wing in flight

Marey (1870) *Animal Mechanism*
E.-J. Marey: Chronophotography

Rather than separate frames, Marey’s “fusil photographique” allowed one to see motion continuously in a single static image.

This provides a visual analysis of a sprint:
- The runner takes about ½ second (7 frames) to make it to an upright position
- Successive frames alternate between power push from the hind leg to landing on the opposite leg

Source: https://lightsmellsloud.wordpress.com/tag/etienne-jules-marey/
The Falling Cat Problem

Another fundamental problem answered by chronophotography:

• How does a falling cat usually land on her feet? An OMG moment!
1896: Area rectangles on a map to display two variables and their product- Jacques Bertillon
By the 1930s, the growth of statistical methods supplanted enthusiasm for graphics

- There were few graphic innovations
- In statistics: numbers were precise; graphs were just pretty pictures

But graphical methods had entered the mainstream & were popularized

- Text books, college courses

There were several graphic-based scientific discoveries

Electronic computers were born
1900-1949: The Modern Dark Ages

1914: Brinton: *Graphic Methods for Presenting Facts*

1924: ISOTYPE method of pictorial graphics—Otto Neurath

1913: Discovery of atomic number, based on graphical analysis—H. Mosely

1911-1913: The Hertzsprung-Russell diagram & evolution of stars

1944: Harvard's Mark I, the first digital computer—Howard Aiken, Grace Hopper
1914: Willard C. Brinton publishes *Graphic Methods for Presenting Facts*, the 1st popular book on the topic
1924: Otto Neurath developed the Isotype (International System of Typographic Picture Education) method to communicate statistical information to the broad public in an intuitive, pictorial way.
1904: E.W. Maunder plots distribution of sunspots in sun’s latitude by time

- Discovery of 11-year sunspot cycles (& 22-yr: reversal of sun’s magnetic field)
1904: E.W. Maunder plots distribution of sunspots in sun’s latitude by time

- Discovery of “Maunder minimum” (1645-1715): “Little Ice Age”
- Smoothing reveals other extrema
Visualization began to rise from dormancy in the mid 1960s, spurred largely by:

- J. W. Tukey’s *Exploratory Data Analysis*: The power of graphics to show the unexpected in data analysis
- Jacques Bertin’s *Semiologie Graphique*: A general theory of composing graphs and maps
- computer hardware for computation and display
- the advent of statistical and graphics software
1950-1974: Re-birth of graphics

1969: Graphical innovations for EDA (stem-and-leaf, box-plots, etc.)- J.W. Tukey

1967: Comprehensive theory of graphical symbols and modes of graphics representation- Jacques Bertin

1967: Reorderable matrix- Jacques Bertin

1971: Star plots- J. H. Siegel etal

1973: Face plots- Herman Chernoff
The biggest limitation in the early development of dynamic and interactive graphics was in graphics display devices.

Only B/W, but for the first time, dynamic displays became possible.

By the late 1950s, pen-like input devices allowed rudimentary direct interaction.
1975-present

Technology:

- Progressively more powerful computation & graphics
  - Mainframes → PCs → workstations → servers → cloud computing
  - pen plotters → CRTs → graphics hardware & firmware
  - stand-alone → client-server architecture

- Internet
  - email → file sharing (FTP) → www (HTML) → Java → javascript
  - data: open data initiatives with APIs
  - ecommerce: Amazon, Netflix, ... → BIG data

- Software
  - Statistical packages: SAS, SPSS
  - Statistical programming environments: R, matlab, Stata
  - Contributed package archives: CTAN (latex), CPAN (perl), CRAN (R)
  - Collaborative development sites: github, bitbucket, ...
1975-present

Themes in data visualization:

• high-D problems of progressively higher dimensions
  ▪ grand tour: n-D $\rightarrow$ 2D projections
  ▪ Dimension reduction methods (PCA, MDS, biplots)

• new data types:
  ▪ categorical data,
  ▪ networks, trees, ...

• interactive data vis
  ▪ linked views
  ▪ direct manipulation: select, zoom, filter
  ▪ dynamic graphics & animation
1975-present

1985: Parallel coordinates plots for high-D data - Alfred Inselberg

1991-1996: High-interaction systems for data analysis and visualization, e.g., XGobi, ViSta

1991: Mosaic display for visual analysis of log-linear models - Michael Friendly

1996: Cartographic Data Visualiser – Jason Dykes
Next steps: Hardware

• Dynamic 3D graphics was painfully slow for larger data sets.

• Specialized 3D graphics hardware:
  ▪ Early 1970s: Simple LSI graphics chips for video games
  ▪ 70s—80s: Graphics co-processors (GPUs) with increasing graphics capabilities
  ▪ 80s—90s: Silicon Graphics develops high-performance 3D graphics workstations
Next steps: Software

Paul Velleman (~1985): Data Desk provided multiple 1D, 2D, 3D views

- **Brushing**: selection of points, regions, … via mouse
- **Linking**: Any action in one plot reflected in all others
Young, Valero-Mora & Friendly (2006)

A philosophy & pedagogy for statistics based on dynamic interactive graphics

A theory of datavis software:
- objects (data, model, ...)
- methods (print, plot, )
- manipulating plot objects & dimensions
- spreadplots: dynamically linked views
- workmaps: visual record of analysis steps

Details: [https://www.uv.es/visualstats/](https://www.uv.es/visualstats/)

Conclusions

• Data Visualization has deep & wide roots:
  ▪ **Cartography**: map-making, geo-measurement, thematic cartography, GIS, geo-visualization
  ▪ **Statistics**: probability theory, distributions, estimation, models, stat-graphics, stat-visualization
  ▪ **Data**: population, economic, social, moral, medical, ...
  ▪ **Visual thinking**: geometry, functions, mechanical diagrams, EDA, ...
  ▪ **Technology**: printing, lithography, computing...

• **Problem driven**: developments often driven by practical and theoretical problems of the day

• **Communication driven**: developments often arose from a desire to communicate better