

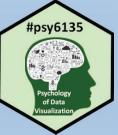


Varieties of information visualization

Michael Friendly

Psych 6135

 <https://friendly.github.io/6135>



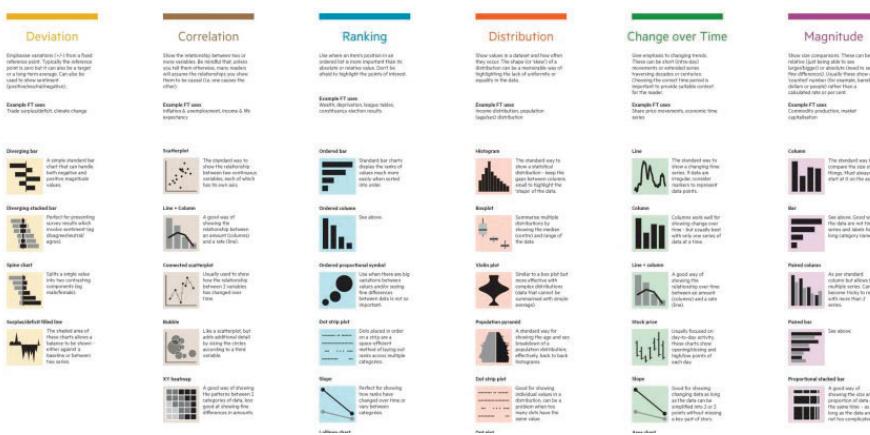
@datvisFriendly



How to classify ???

For purposes of “What kind of graph should I use?” usually most useful to think:

“What do I want to show?”

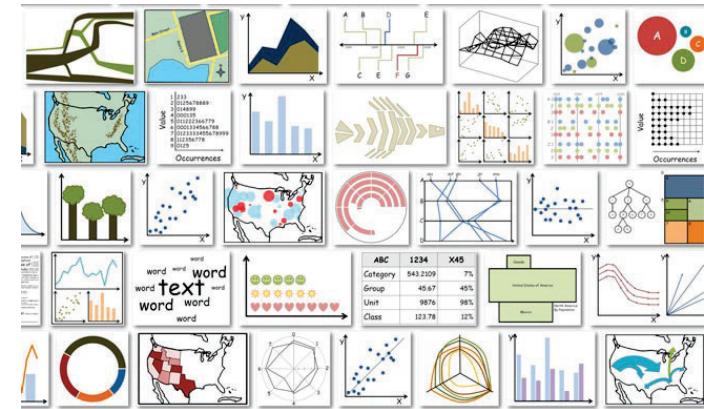


Financial Times, Chart Doctor, <https://github.com/ft-interactive/chart-doctor>

So many types

There are so many kinds of charts, diagrams, graphs, maps

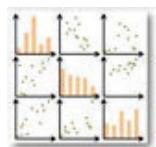
- What are their **features**?
 - What **tasks** are they **good for**? – Accuracy or speed of judgment? Memorability?

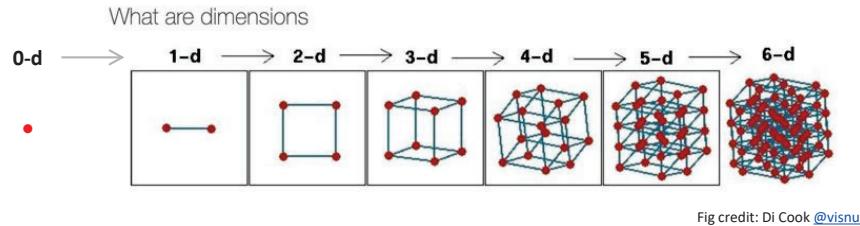


2

Topics, by graph type: Content & form

- Statistical data graphs
 - 1D: dotplot, boxplot, violin plot
 - 1.5D: time-series plot, density plot, bar chart, pie chart
 - 2D: scatterplot, ridgeline plot
 - 3D: contour plot, 3D scatterplot, surface plot
 - Thematic maps
 - Choropleth map
 - Anamorphic map
 - Flow maps
 - Network & tree visualization
 - Animation & interactive graphics





1 D
1.5 D
2 D
3 D
n D ?

Data graphs can be roughly ordered by the number of variables, data dimensions shown in a given graph

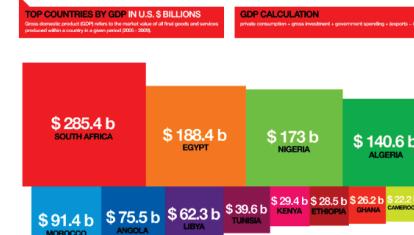
Data graphs

5

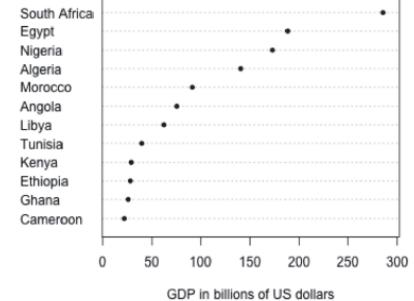
1D: Infographic vs. Data graphic

The same data can be shown in different forms, for different purposes

African Countries by GDP



African Countries by GDP



One might argue that this infographic has **greater impact** in showing the relative size of GDP

One might argue that this statistical graph **makes comparisons easier**

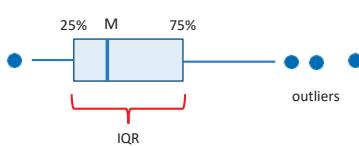
6

1.5D: Dotplots & boxplots

What number do you give to a probability phrase?

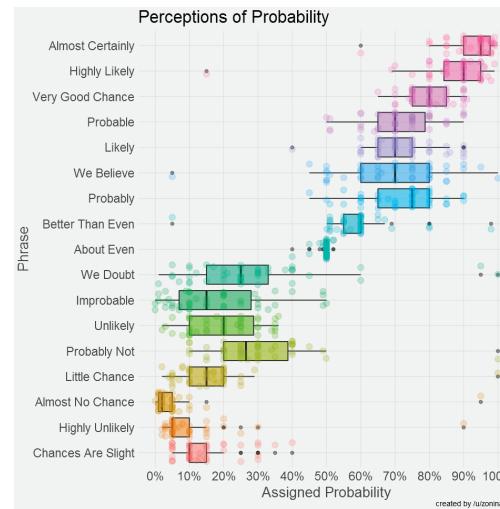
Boxplots summarize the important characteristics of a univariate data distribution:

- center (median)
- spread (IQR)
- shape (symmetric? skewed?)
- outliers?



This example overlays the boxplot with a jittered dotplot, so we can also see the individual observations

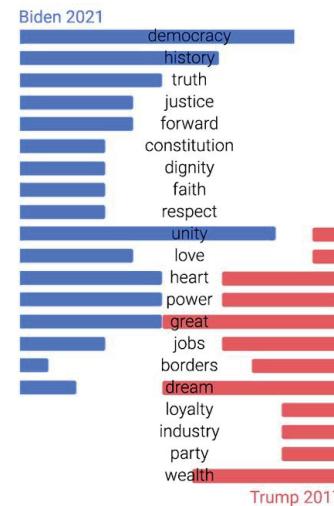
This visualization made the longlist for the 2015 Kantar Information is beautiful award. Data & R code:
<https://github.com/zonination/perceptions>



7

1.5D: Text bar charts

Inaugural Address: Word Frequency



- Text can be analyzed as data also, most often in frequency counts.
- This chart uses a novel design to compare the most frequent words by Biden (2021) & Trump (2017) in their inaugural addresses.
- The contrast is striking!
 - democracy, unity vs. great, dream

From:
https://www.reddit.com/r/dataisbeautiful/comments/l7k0f0/us_inauguration_address_word_frequency_biden_vs/

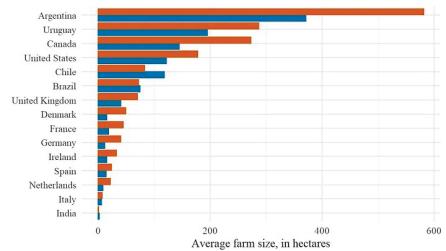
9

Bar chart variations

Bar charts are often used to make comparisons between two series

But your eyes must move around to compare

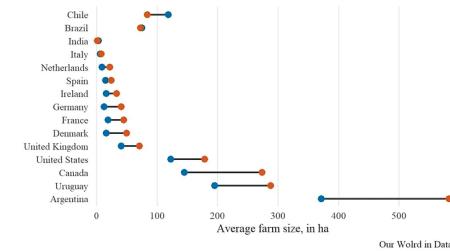
Comparison of average farm size between 1960 and 2000



An alternative is a [dumbbell](#) chart---dot plot w/ connecting lines---making comparisons direct.

Effect ordering: Countries have also been sorted to show increase / decrease

Comparison of average farm size between 1960 and 2000



Which countries gained the most from 1960 – 2000?

From: <https://medium.com/@amorimfranchi/an-easier-to-interpret-alternative-to-paired-barplots-in-r-490d4472e8cd>

10

1.5D: Time series line graphs

William Playfair (1786), *The Commercial and Political Atlas*, invented the time series line graph as a way to show data on England's trade with other countries

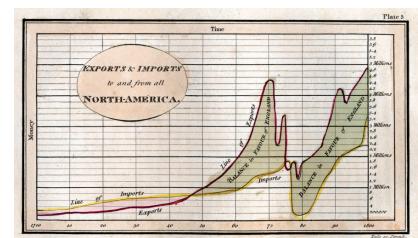
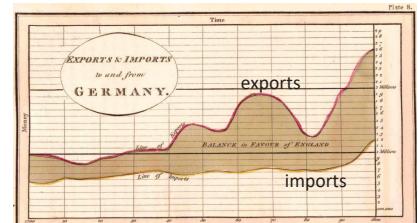
One curve for imports, one for exports

The balance of trade could be seen as the difference between the curves

Trade with Germany was consistently in favor of England

With North America, the balance changed back and forth over time

Economic 'history' could now be visualized and explained

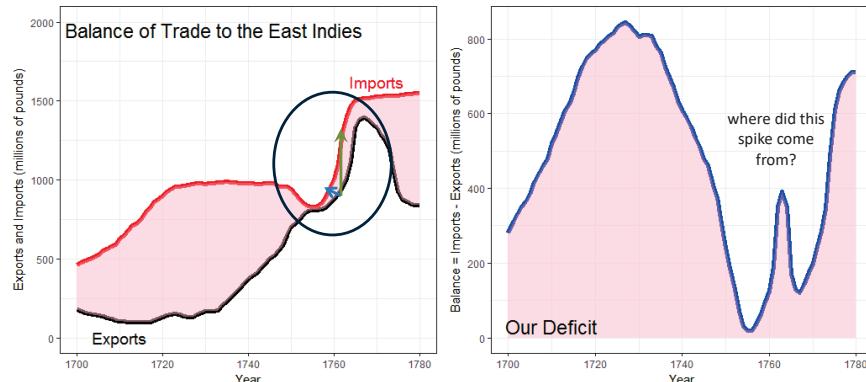


11

Psychology: Distances between curves

What Playfair didn't know is that judgments of **distance** between curves are **biased**. We tend to see the **perpendicular distance** rather than the **vertical distance**

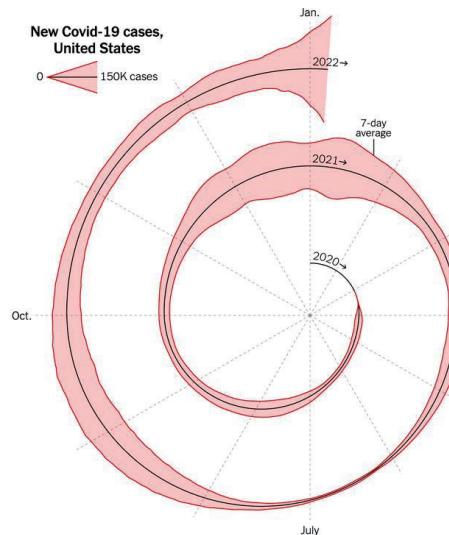
Plotting balance of trade directly



12

The New York Times

Jan. 6, 2022



1.5D COVID spiral graph

This graph shows a 1.5D time series of COVID-19 cases over the entire span of the pandemic

It is wrapped into a spiral to compare **months** over years.

Does this work for you, or is it too weird?
Is it the "tapeworm of doom"?

Hey – wait
Isn't this just a time-series in polar coordinates?

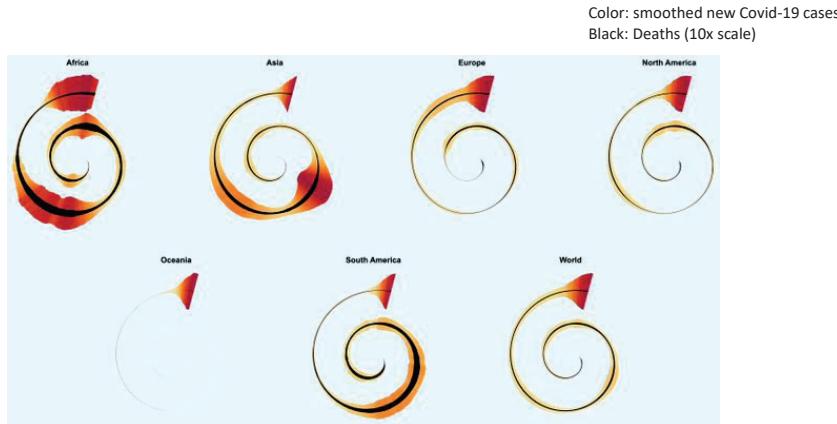
See commentary: <https://www.youtube.com/watch?v=YWOWF9Vi4Mw>

13

Multi-spirals

Multiple spirals: Allow comparison of the patterns for different geographic units

- Overlaid time-series probably would not work here
 - How could this be improved?

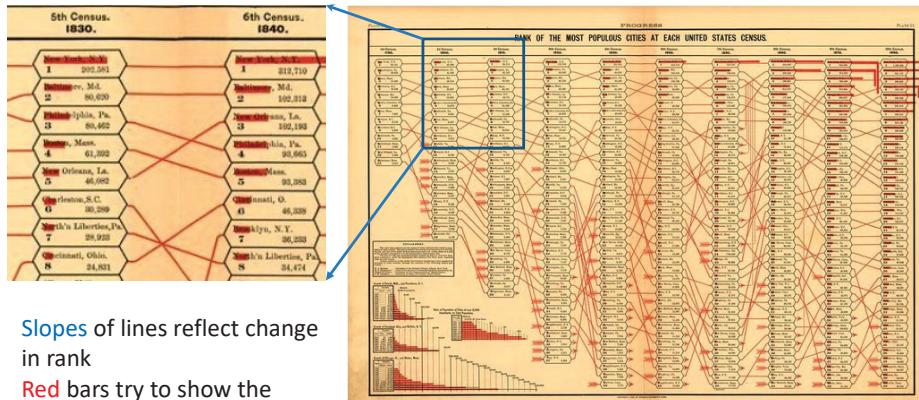


<https://twitter.com/higsc/status/1483195230404947968>

14

Parallel ranked list charts

Another solution for multiple time series is to chart the **ranks** of observations and connect them with lines to show changes in relative position.



Slopes of lines reflect change in rank

Red bars try to show the numbers

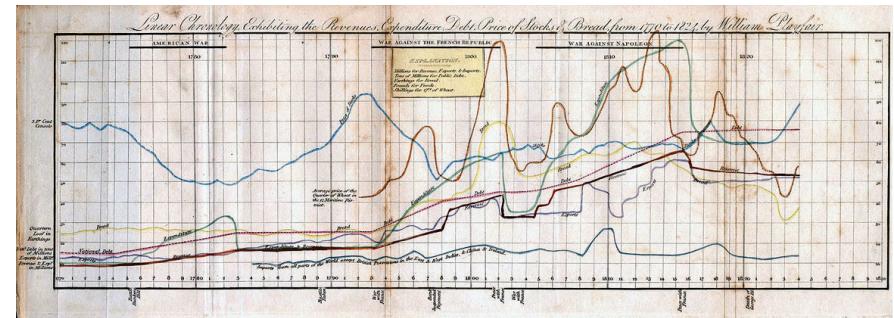
Ranks of populations of US cities across census years
From: *Statistical Atlas of the United States* (1880)

16

Multiple time series graphs

Things get messy when there are many series to be compared

- To be fair, this was designed as **timeline of history**— a visual story of economics (prices, wages, imports, exports, debt)
 - History shown as a **strip-chart recording** (e.g., EKG)
 - It was Playfair's last graph.
 - Perhaps the earliest example of **visual storytelling**



Playfair, W. (1824) *Chronology of Public Events and Remarkable Occurrences.*

15

Propensity of crimes across age

André-Michel Guerry (1833), rank order of crimes at different ages

- sexual assault **against adults** decreases with age
 - sexual assault **against children** increases with age
 - **parricide** relatively low until 60-70



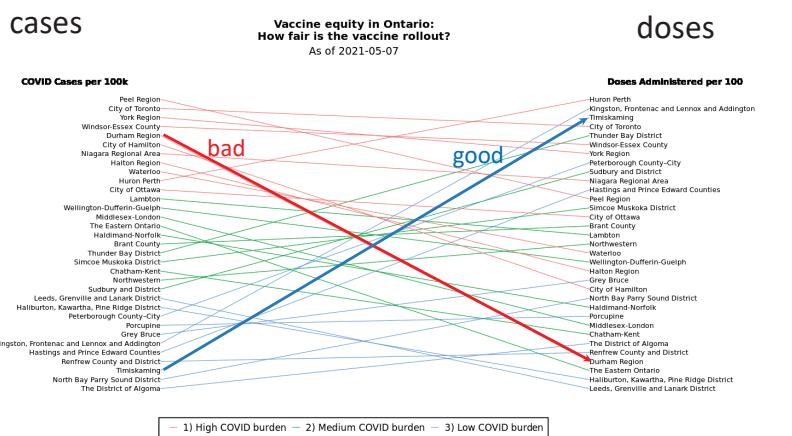
Friendly, (2007). "A.-M. Guerry's Moral Statistics of France: Challenges for Multivariable Spatial Analysis," *Statistical Science*, vol. 22, no. 3. <https://www.datavis.ca/papers/guerry-STS241.pdf>

17

Slope graph of COVID: Cases vs. Doses

Vaccine equity → all lines should be ≈ flat

Which health regions stand out?
How could this graph be better?

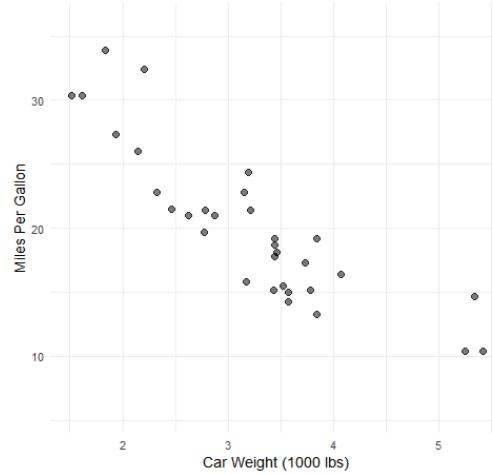


Graphic by Matthew Dubins, @MDubins

18

2D: Scatterplots

1974 Motor Trend Cars: Gas Mileage vs. Weight [Frame 1]



Goals: Want to see--

- Overall pattern:
 - direction
 - form
 - strength
- Deviations from pattern:
 - residuals
 - outliers

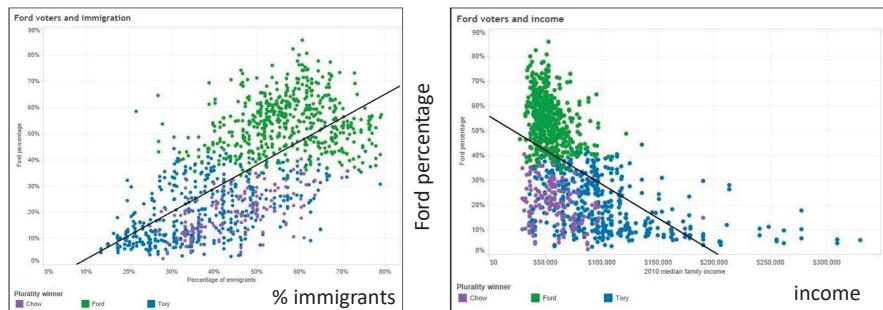
Annotations can help:

- regression line
- confidence band
- smoothed curve (nonlinear?)
- residuals

2D: Scatterplots: Ford Nation



Who voted for Rob Ford in the 2014 Toronto mayoral election?



These simple scatterplots by data journalist Patrick Cain use simple enhancements:

- Color, for candidate (Chow, Ford, Tory)
- Overall regression line

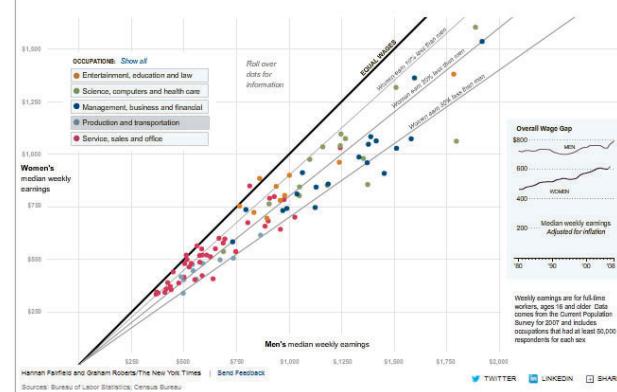
Source: <https://globalnews.ca/news/1652571/ford-nation-2014-15-things-demographics-tell-us-about-toronto-voters/>

20

Scatterplots: Wage gap

Published May 16, 2010
Why Is Her Paycheck Smaller?

Nearly every occupation has the gap — the seemingly unbridgeable chasm between the size of the paycheck brought home by a woman and the larger one earned by a man doing the same job. Economists cite a few reasons: discrimination as well as personal choices within occupations are two major factors, and part of the gap can be attributed to men having more years of experience and logging more hours.



How to compare salaries of men & women in different occupations?

The NYT chose to plot median salaries for women against those for men, in different occupational groups

The 45° line represents wage parity
Other lines show 10, 20, 30% less for women

How else to show this?

Alberto Cairo, *The Truthful Art*, Fig 9.19, from:
http://www.nytimes.com/interactive/2009/03/01/business/20090301_WageGap.html

21

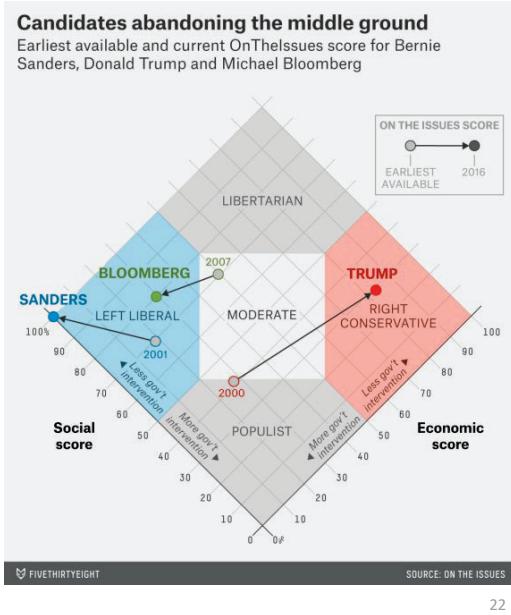
Scatterplots: InfoVis

This graph, from fivethirtyeight.com, was designed to show how some presidential candidates had shifted positions before the 2016 election.

The axes are a score on **social** and **economic** policy, but they rotate the axes by 45° to create zones related to political thought.

This info graphic is **eye-catching** and **self-explanatory**:

- colored/labeled zones
interpretive labels on axes
arrows showing movement to extremes

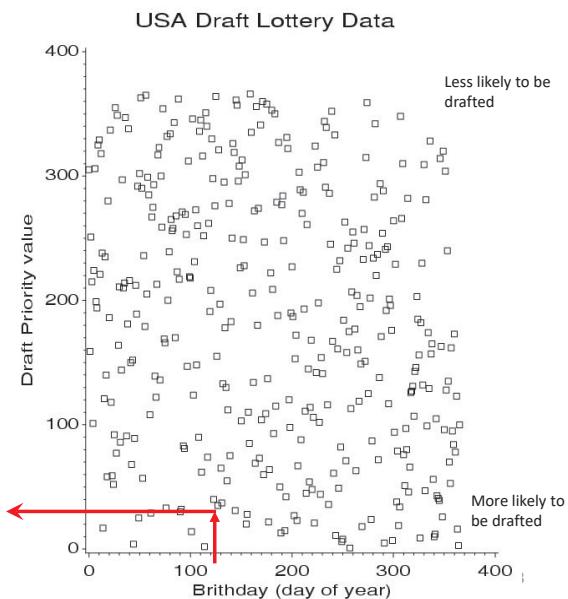


Scatterplots: Annotations enhance perception

Data from the US draft lottery, 1970

- Birth dates were drawn at random to assign a “draft priority value” (1=bad)
 - Can you see any pattern or trend?

This is an example of data with a weak signal and a lot of noise

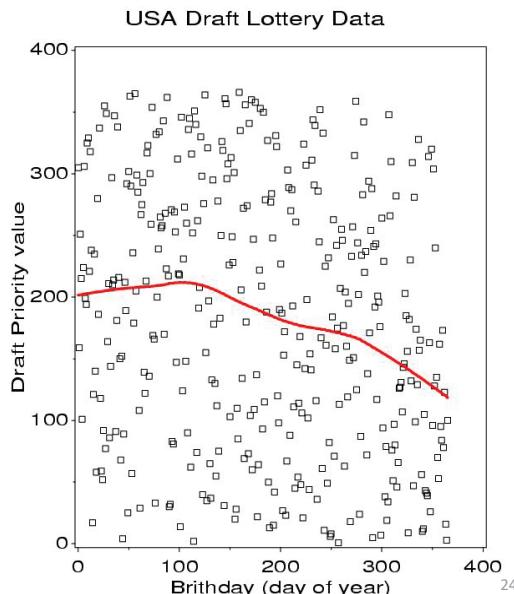


Scatterplots: Smoothing enhances perception

Drawing a smooth curve shows a systematic decrease toward the end of the year.

- The smooth curve is fit by **loess**, a form of non-parametric regression.

Visual explanation:

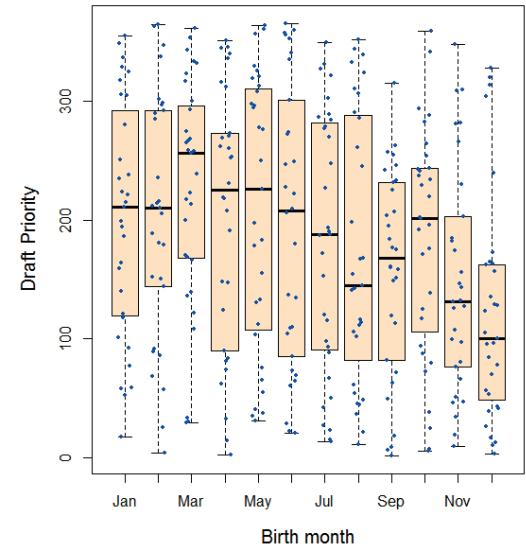


Smoothing by grouping and summarization

Another form of smoothing is to make one variable discrete & show a graphical summary – here a boxplot

The decrease in later months becomes apparent

Perception: the boxplots form the foreground; the jittered points show the data



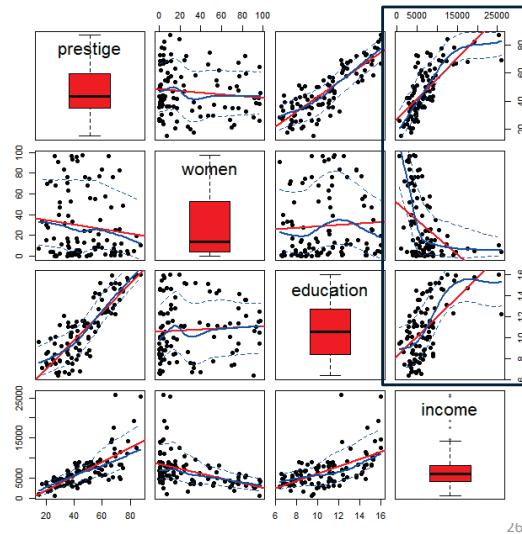
Scatterplot matrices

A scatterplot matrix shows the bivariate relation between all **pairs** of variables. Seeing these all together is more useful than a collection of separate plots.

Q: How does occupational prestige depend on %women, education and income?

The individual plots are enhanced with linear regression lines and non-parametric smoothes to show non-linearity

This figure uses `scatterplotMatrix()` in the [car](#) package. There are many options.



Larger data sets

Scatterplot matrices hold up reasonably well with a larger number of variables

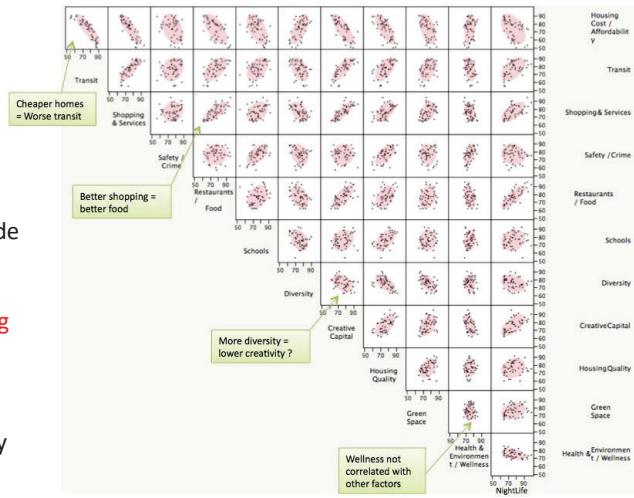
Where to live in NYC?

This SPM shows 12 variables on ~ 60 neighborhoods

The data **ellipses** provide a visual summary

I call this **visual thinning** – reducing details in a larger picture

In an interactive display we can **zoom** in/out



From: http://junkcharts.typepad.com/junk_charts/2010/06/the-scatterplot-matrix-a-great-tool.html

Scatterplot matrices

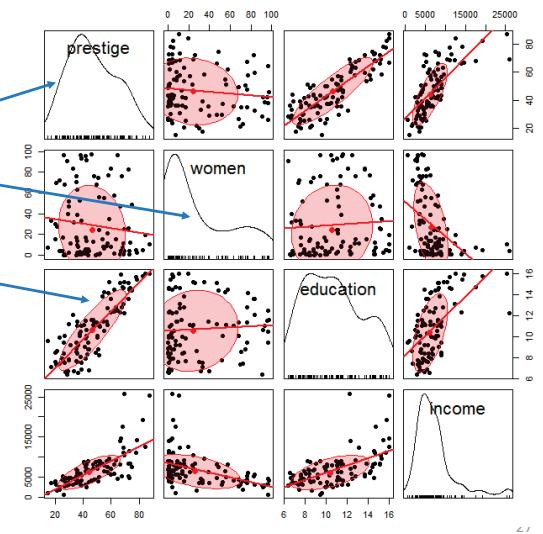
Essential idea: You can choose different methods to render the diagonal & off-diagonal panels

Density plots are often more useful for showing the shapes of distributions

- prestige: ~ symmetric
- women: bimodal
- income: highly skewed

A **data ellipse** gives a visual summary of the direction and strength of the relationship

Again, graphical annotation aids interpretation.



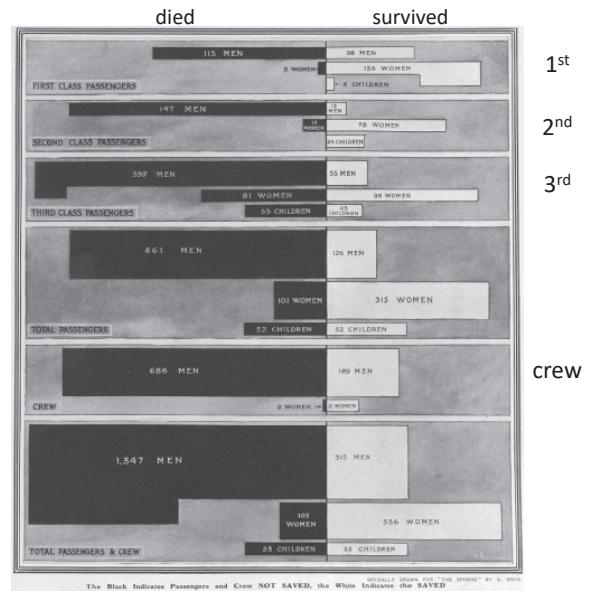
Categorical data

This remarkable chart shows survival on the *Titanic*, by Class for passengers and Gender and Age.

It was drawn by G. Bron, a graphic artist, and published in *The Sphere*, one month after the *Titanic* sank.

It uses back-to-back bar charts, with area ~ frequency

See our web page:
<http://datavis.ca/papers/titanic/>



Categorical data: Mosaic plots

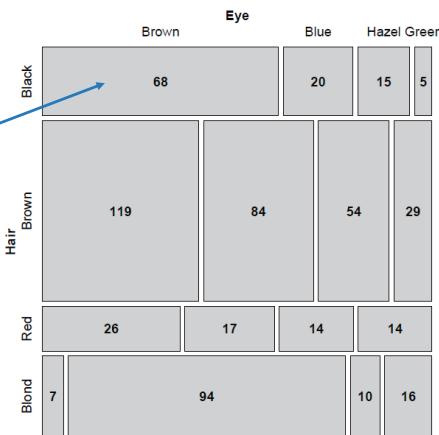
Similar to a grouped bar chart
Shows a frequency table with tiles,
area \sim frequency

```
> data(HairEyeColor)
> HEC <- margin.table(HairEyeColor, 1:2)
> HEC
  Eye
  Hair Brown Blue Hazel Green
    Hair Black 68 20 15 5
    Brown 119 84 54 29
    Red 26 17 14 14
    Blond 7 94 10 16
> chisq.test(HEC)

Pearson's Chi-squared test

data: HEC
X-squared = 140, df = 9, p-value <2e-16
```

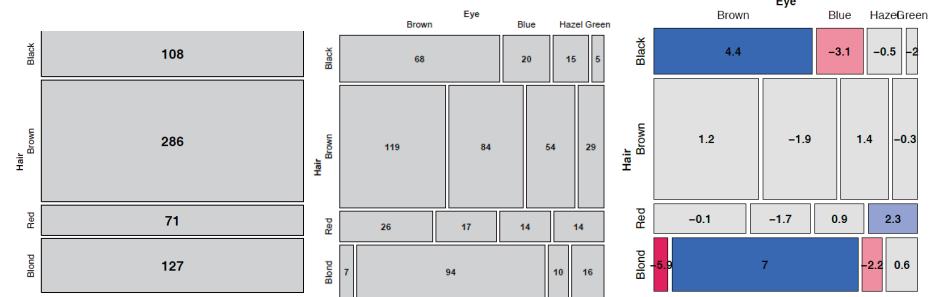
How to understand the association
between hair color and eye color?



30

Mosaic plots

Area proportional display of frequencies in n -way table
Shaded according to deviation (residual) from independence



Divide unit square $\sim V_1$
marginal frequencies

Subdivide each $\sim V_2 | V_1$
conditional frequencies

$$d_{ij} = \frac{(O_{ij} - E_{ij})}{\sqrt{E_{ij}}} \quad \chi^2 = \sum d_{ij}^2$$

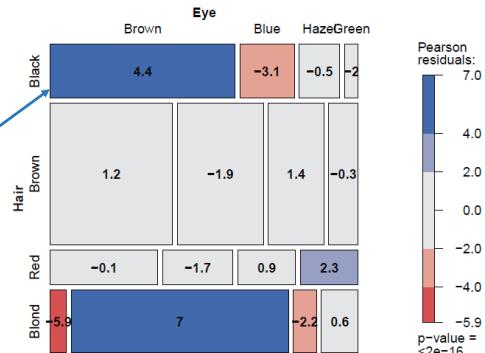
31

Mosaic plots

Shade each tile in relation to the
contribution to the Pearson χ^2
statistic

$$\chi^2 = \sum d_{ij}^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

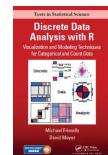
```
> round(residuals(chisq.test(HEC)), 2)
  Eye
  Hair Brown Blue Hazel Green
    Hair Black 4.4 -3.1 -0.5 -2
    Brown 1.2 -1.9 1.4 -0.3
    Red -0.1 -1.7 0.9 2.3
    Blond -5.9 7 -2.2 0.6
```



32

Mosaic plots extend readily to 3-way + tables

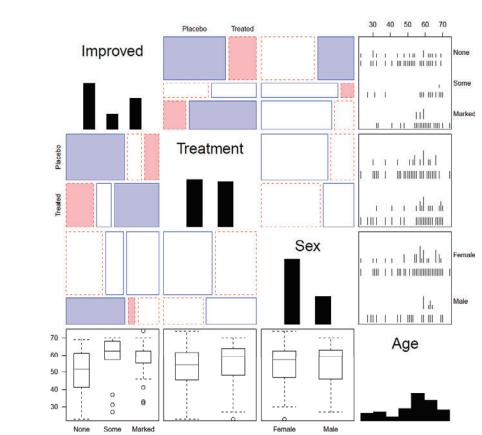
They are intimately connected with loglinear models
See: Friendly & Meyer (2016), Discrete Data Analysis with R, <http://ddar.datavis.ca/>



Generalized pairs plots

Generalized pairs plots from the `gpairs`
package handle both categorical (**C**) and
quantitative (**Q**) variables in sensible ways

x	y	plot
Q	Q	scatterplot
C	Q	boxplot
Q	C	barcode
C	C	mosaic

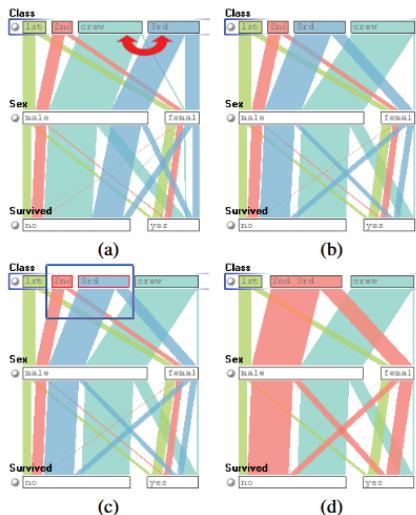


```
library(gpairs)
data(Arthritis)
gpairs(Arthritis[, c(5, 2:5)], ...)
```

33

Parallel Sets

Titanic data: Who survived?



From: Kosara et al. (2006), https://kosara.net/papers/2006/Kosara_TVCG_2006.pdf

34

Parallel sets use **parallel coordinate axes** to show the relations among categorical variables.

The frequencies of one variable (Class) are sub-divided according to the joint frequencies in the next (Sex) and shown by the width of the connecting line.

The ParSets application is interactive:

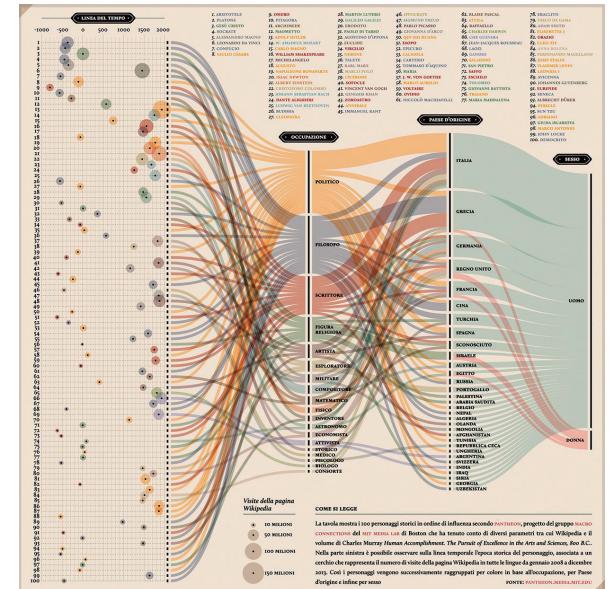
- categories can be reordered (a, b)
- categories can be grouped (c, d)

Sankey diagram

Pantheon, by Valerio Pellegrini
Visualizing the 100 most influential figures in History
(Wikipedia visits)

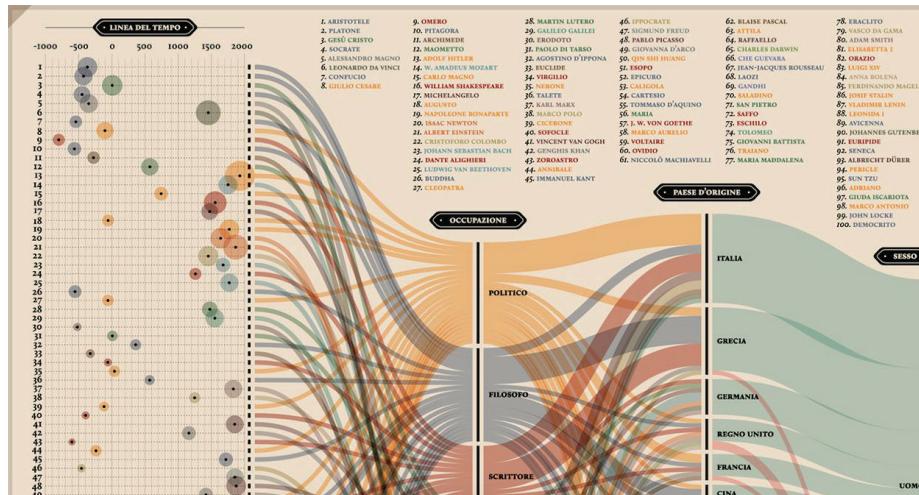
Columns show **occupation**,
country of origin and
gender

Flow lines link individuals to the column variables, width ~ influence



35

Sankey diagram



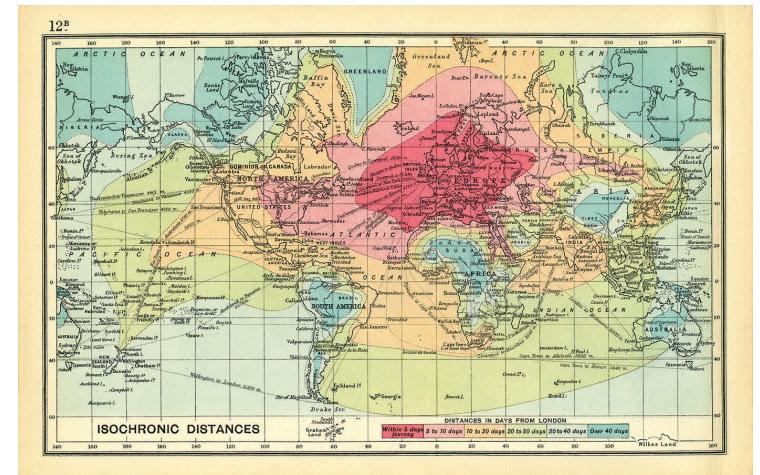
Multiple dimensions of the most influential people in history

From: <http://visualoop.com/blog/83382/pantheon-by-valerio-pellegrini>

36

3D: Iso-contour maps

Early attempts to show 3D data used **contours of equal value** on a map
The data was actually very thin; the contours the result of imaginative smoothing

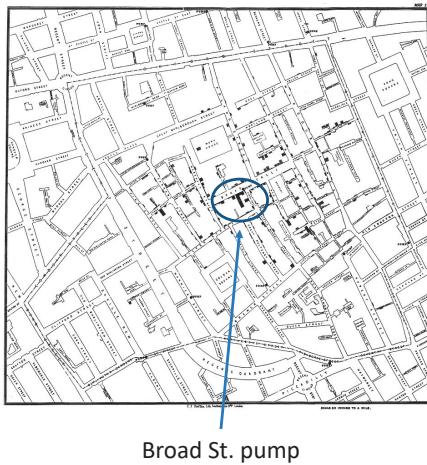


Francis Galton, Isochronic chart of travel time, 1881

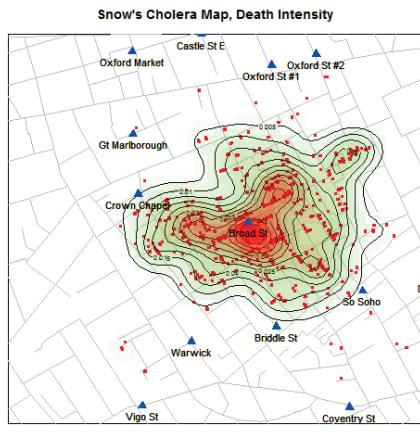
37

3D: Bivariate density estimation

John Snow's map of cholera deaths in London, 1854



Modern statistical techniques can compute contours of constant density



Data: [HistData](#) package for R



38

3D: population pyramid

Italian demographer Luigi Perozzo (1880) developed the first true 3D diagram showing the population of Sweden over years and age groups as a 3D surface

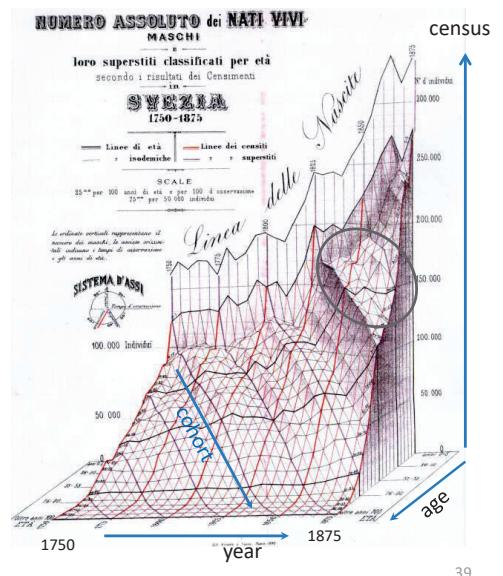
Census counts for a given **year** are shown by the **red lines**

Survival of a given **age** are shown by **black lines**

Cohorts are shown by diagonal lines down & to the right

These 3 variables are primary in demography.

A mystery here: what caused the decline at the upper right?



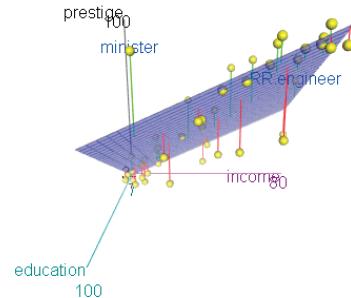
39

3D: scatterplot & regression surface

How does occupational prestige depend on income & education?

This plot shows the data and a fitted multiple regression surface, connecting the points to the regression plane

It is hard to see in a static view, but easier when the plot is rotated dynamically



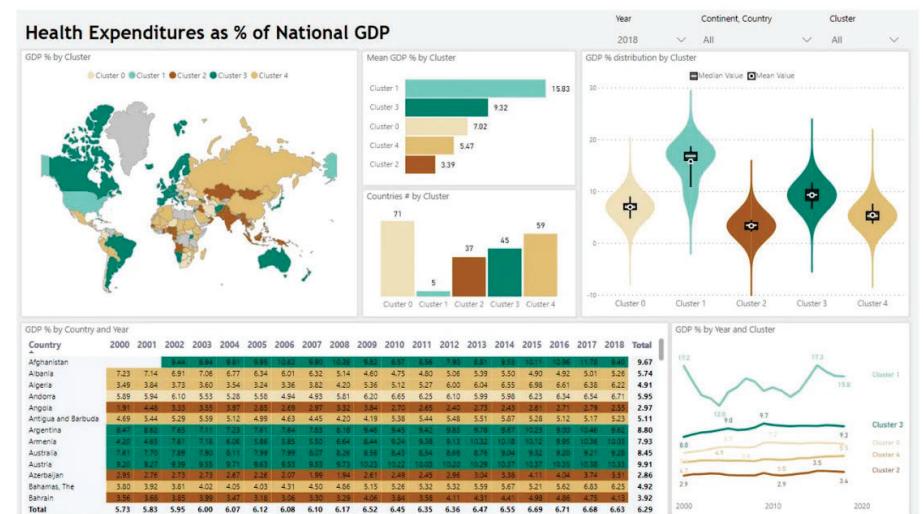
This plot is produced in R, using the [car](#) and [rgl](#) packages

```
data("Duncan", package="car")
scatter3d(prestige ~ income + education, data=Duncan, id.n=2)
movie3d(spin3d(c(0,1,0), rpm=6), duration=6, movie="duncan-reg3d")
```

40

Dashboards

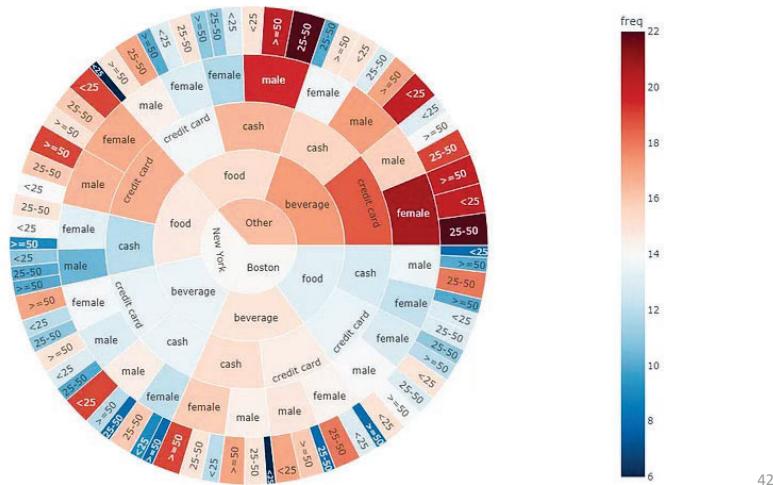
Dashboards provide multiple images of different aspects, often linked interactively



Exercise

What is this SUNBURST chart trying to show?

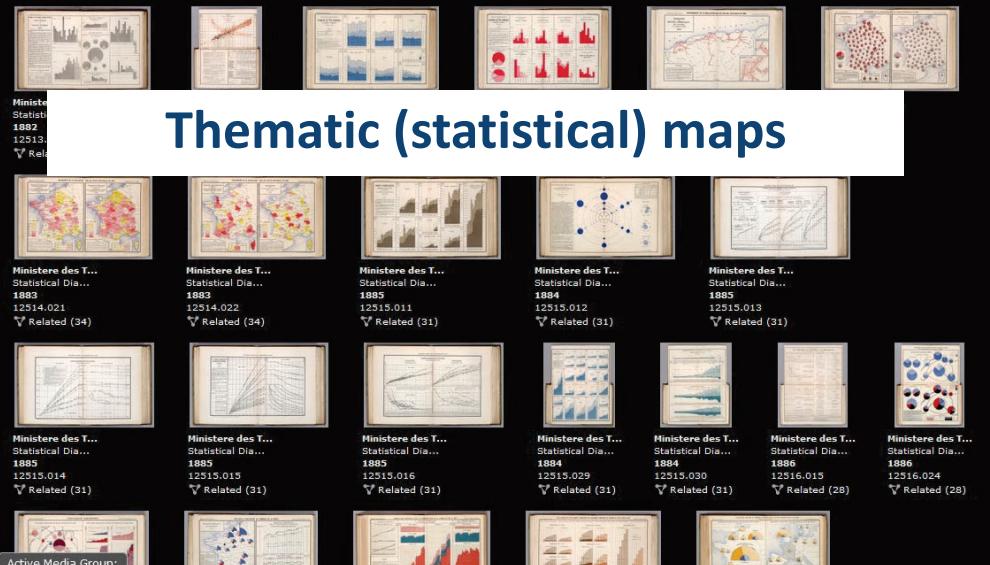
- How many dimensions of data?
- What do the colors mean?



42

Search Results: All Fields similar to 'Statistical and Albums'

51-100 of 666 1 2 3 4 ► ►



Thematic (statistical) maps

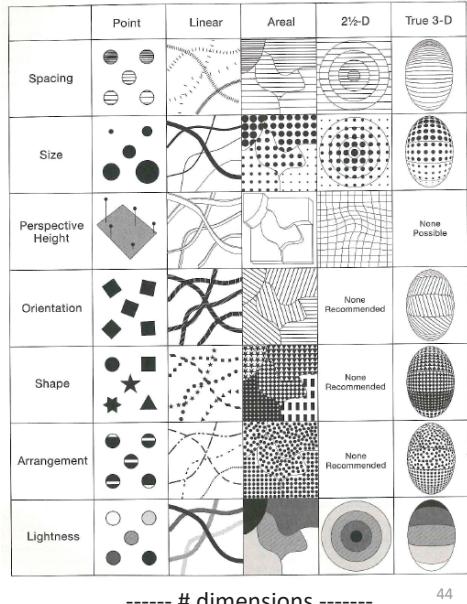
Image from: David Rumsey Map Collection, <https://www.davidrumsey.com/>

43

Thematic maps & Spatial visualization

Thematic maps use a wide variety of techniques to display quantitative or qualitative variables on the geographic framework of a map

Once the domain of cartographers, these ideas are now being developed as an area of [geospatial visualization](#) and geospatial statistical methods

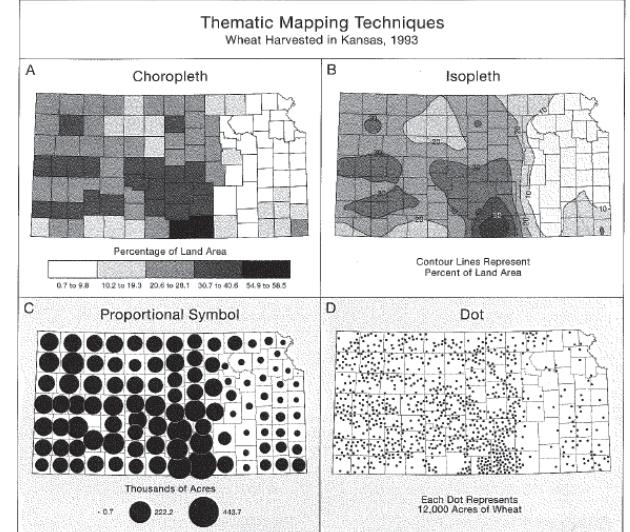


44

Basic types of thematic maps

Most are [direct](#) mappings of numbers to visual variables

- **Choropleth** maps shade by geographic unit
- **Isopleth** maps use contours of equal value



From: Slocum et al., *Thematic cartography and geographical visualization*, Fig 4.9

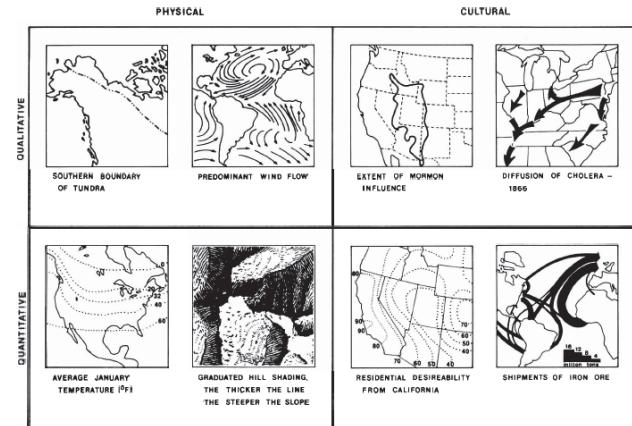
45

Thematic maps: Theory

Alan MacEachern (1979) classifies point, line and area symbols on thematic maps according to whether they depict **quantitative** or **qualitative** phenomena, in the **physical** or **cultural** domain.

This is a coarse classification.

Theories, ideas, and methods have advanced considerably since this time.



MacEachern, A. (1979). The Evolution Of Thematic Cartography / A Research Methodology and Historical Review, *The Canadian Cartographer* 16(1) June 1979, p. 17-33

46

Anamorphic maps

- Anamorph:** Deforming a spatial size or shape to show a quantitative variable
- Émile Cheysson used this to show the decrease in travel time from Paris to anywhere in France over 200 years



Album de Statistique Graphique, 1888, plate 8

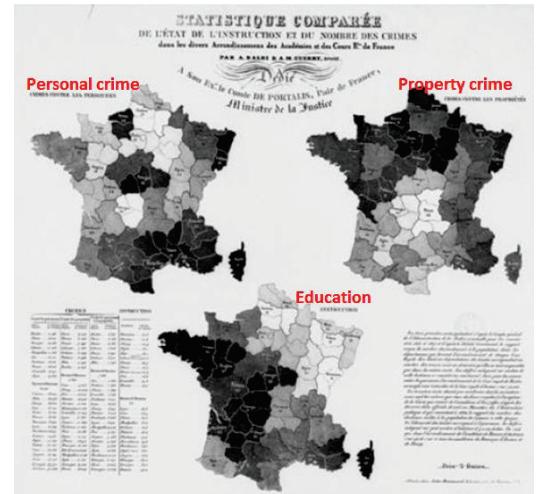
48

Choropleth maps

Balbi & Guerry (1829)

- First thematic maps of crime data
- First comparative maps ("small multiples")
- Crime against persons inversely related to crime against property
- Education: *France obscure* & *France éclairée*
- N. of France highest in education & also property crime

This was the beginning of modern criminology & social science --- relations among social variables made visible



See: Friendly, M. (2022). [The life and works of André-Michel Guerry, revisited](#). *Sociological Spectrum*, 42, 233–259

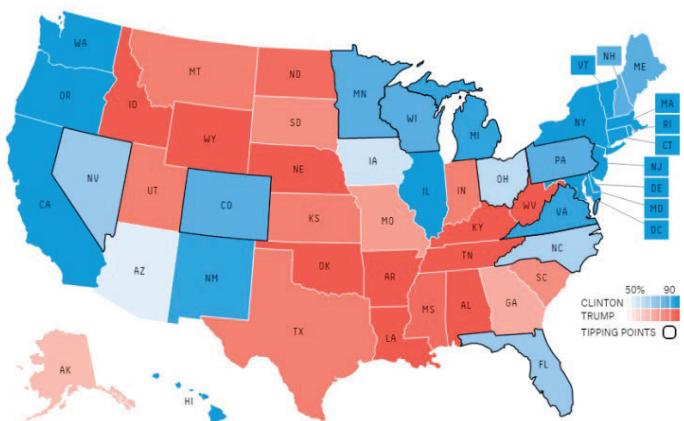
47

What's wrong with choropleth maps?

Choropleth maps are misleading because size (area) of units dominates perception. This is particularly true for maps of the US & Canada. Not so for France (why?)

Montana looks bigger than Washington

Note use of labels for small NE states



fivethirtyeight.com election predictions, Oct. 13, 2017

49

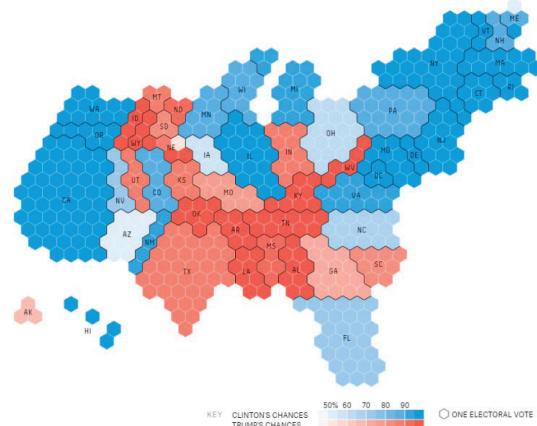
Cartogram (tilegrams)

A tilegram uses hexagonal tiles to make area proportional to a given variable

Here, the size of each state is made ~ number of electoral college votes

Now, it is easy to see the impact of states

Take-away: Area doesn't vote; People do!



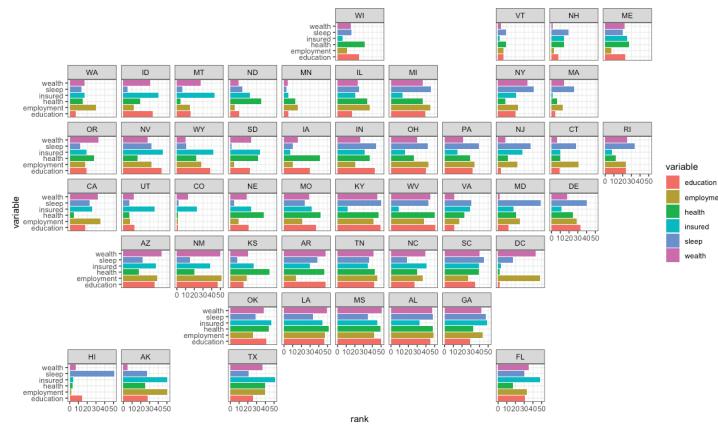
fivethirtyeight.com election predictions, Oct. 13, 2017

50

Geo-faceting

More general idea:

- A grid of (x, y) locations defines geographic units
- Each grid cell can be composed of any kind of plot conceivable with ggplot2

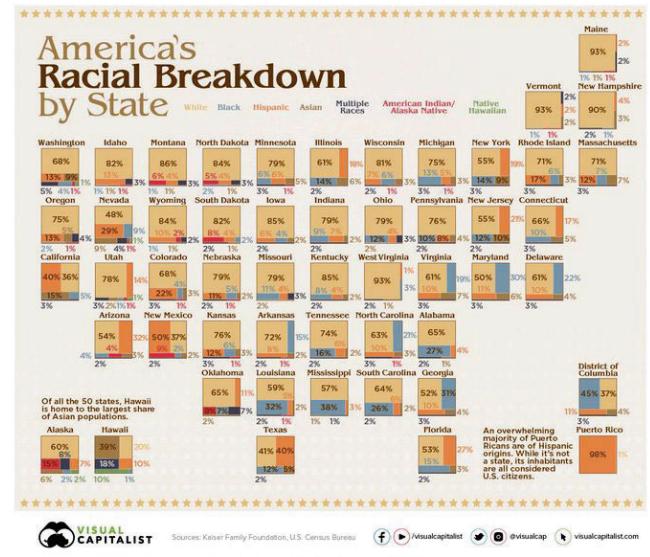


From: <https://cran.r-project.org/web/packages/geofacet/vignettes/geofacet.html>

Mosaic cartograms

US map provides a spatial framework for showing the distribution of categorical data

Each tile is positioned as in a schematic US map



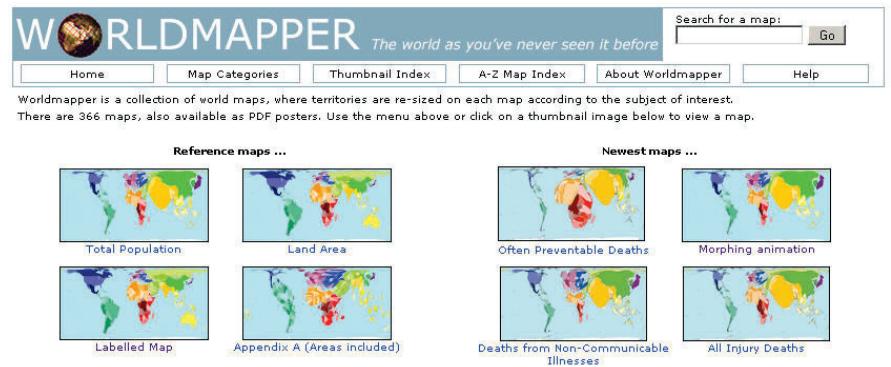
SOURCES: Kaiser Family Foundation, US Census Bureau | COLLABORATORS: RESEARCH + WRITING Anupriya Imran Ghosh, Reuel Amoros | DESIGN Zack Abrahams | ART DIRECTION Melissa Havivis | visualcapitalist.com

51

Worldmapper: The world in cartograms

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

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

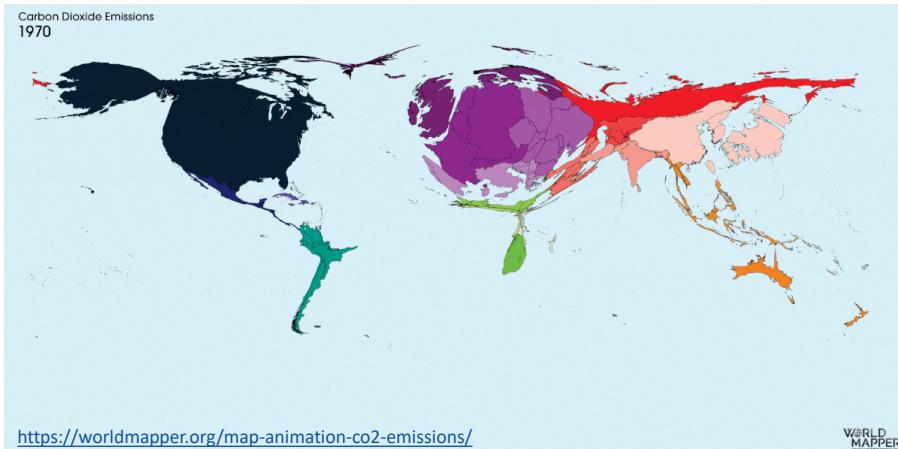


52

53

Worldmapper: Carbon emissions

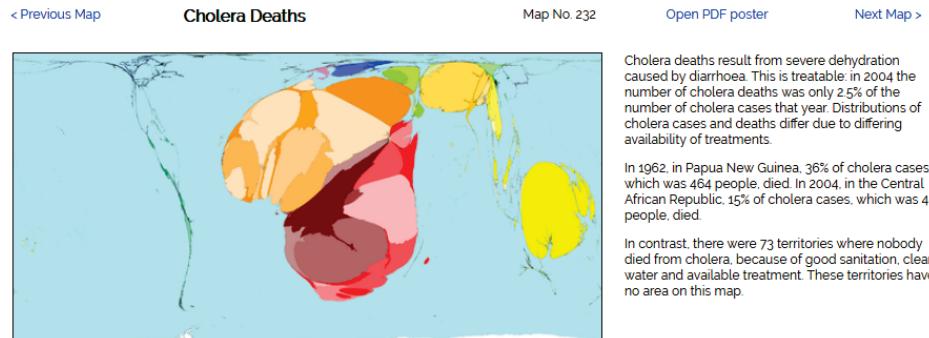
This map animation shows the growth in CO₂ emissions globally: the map keeps growing in size from 1970 - 2016 also shows the changing shares of countries contributing to the increase in emissions



54

Worldmapper: Cholera deaths

Deaths from cholera in 2004. Territory size ~ proportion of worldwide deaths



Territory size shows the proportion of worldwide deaths from cholera that occurred there in 2004 or most recent year available.

<http://www.worldmapper.org/display.php?selected=232>

55

Bivariate maps

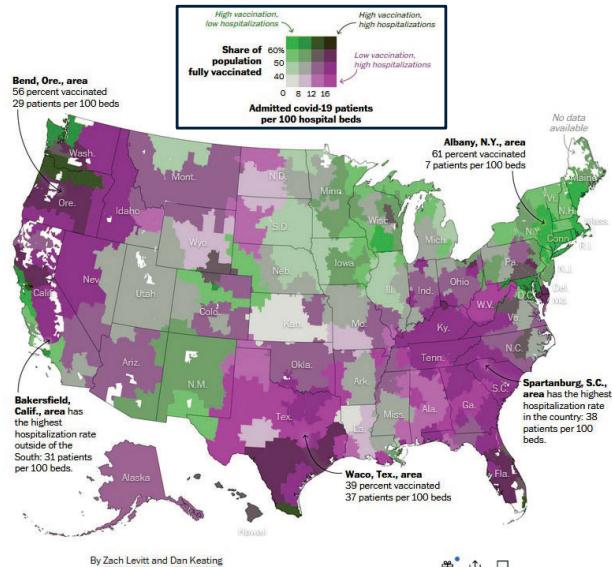
How to show relation between two variables on a map?

Bivariate maps combine two colors with degrees of saturation

Claim:
Regions with more vaccinations have fewer hospitalizations

Can you see this?

Vaccination rates and COVID hospitalizations

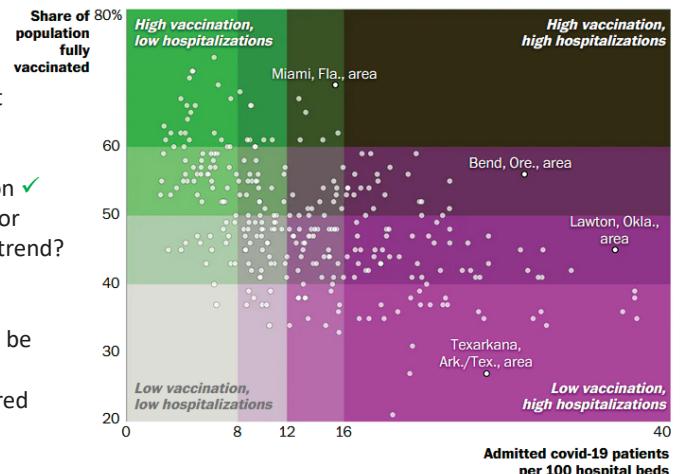


56

Scatterplot to the rescue

Distribution of vaccination and hospitalization rates

Most hospital regions with high vaccination rates have lower hospitalization rates – and places with lower vaccination rates tend to see higher rates of admitted patients.



This lovely scatterplot shows:

- Yes, inverse relation ✓
- NB: Classing of color
- Whoa: non-linear trend?

But:

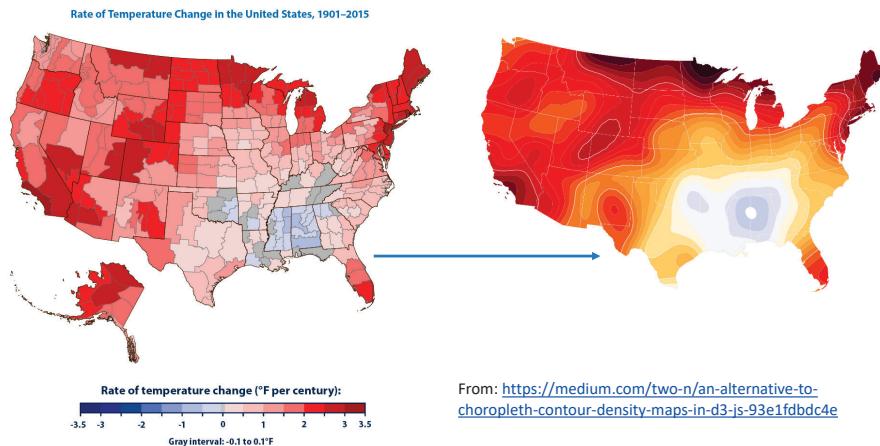
- Shouldn't the axes be reversed?
- Geography is ignored

<https://www.washingtonpost.com/health/2021/09/23/covid-vaccination-hospitalization-map/>

57

Contour maps

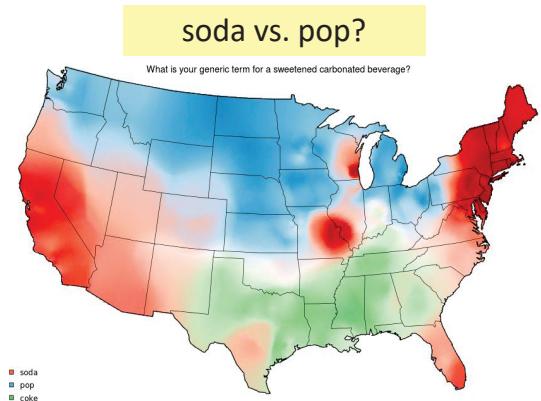
Contour maps ignore region boundaries and estimate constant contours of a phenomenon over geographical space. This is a form of **geo-smoothing**.



58

Spatial visualization: Analysis + maps

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?

What is your generic term for a sweetened carbonated beverage?

Take the survey: http://www.tekstlab.uio.no/cambridge_survey

Programming in R: <http://blog.revolutionanalytics.com/2013/06/r-and-language.html>

59

Spatial visualization: Analysis + maps

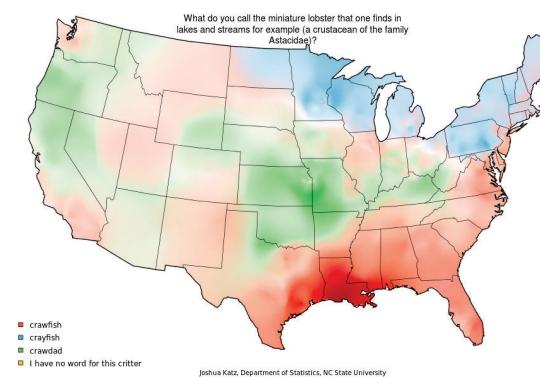
Linguistics: Food dialect maps– visualizing how people speak

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.

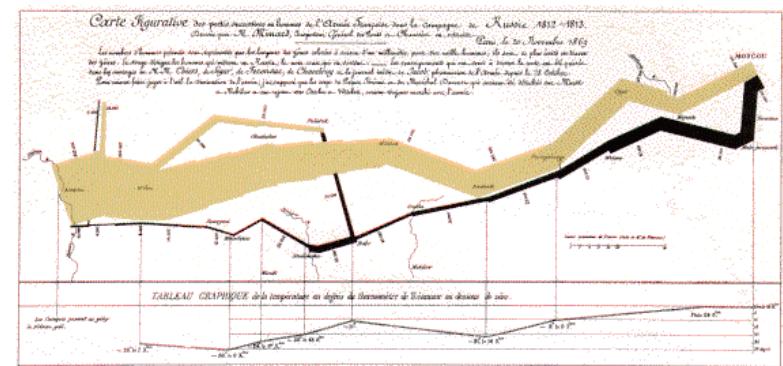
The use of color combines discrete categories with intensity to give a meaningful display



60

Flow maps

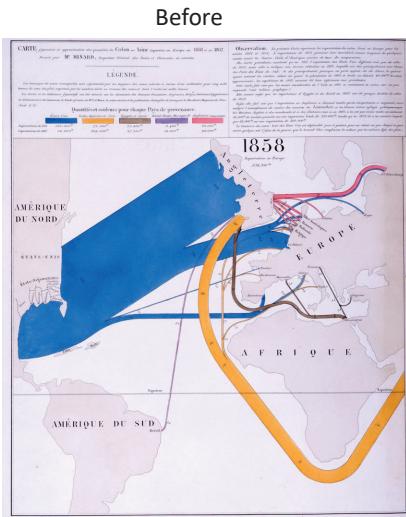
Flow maps show movement or change in a geographic framework
The master work is this image by Charles-Joseph Minard (1869)



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

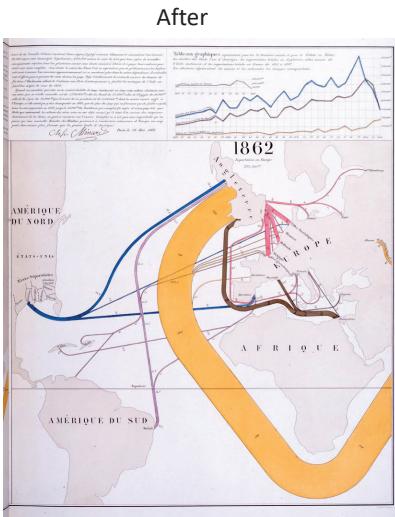
61

Effect of US civil war on cotton trade



Note the deformation of the map to accommodate the data

62



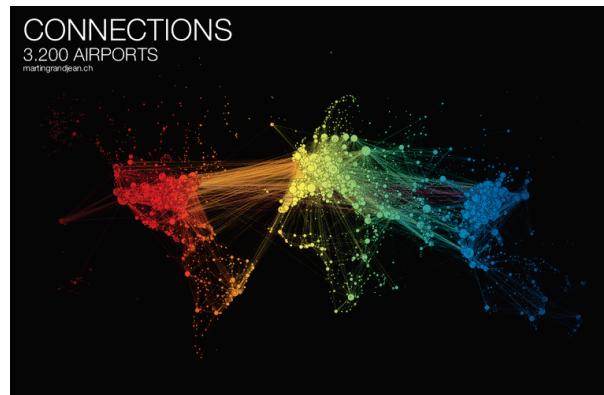
Network visualization



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

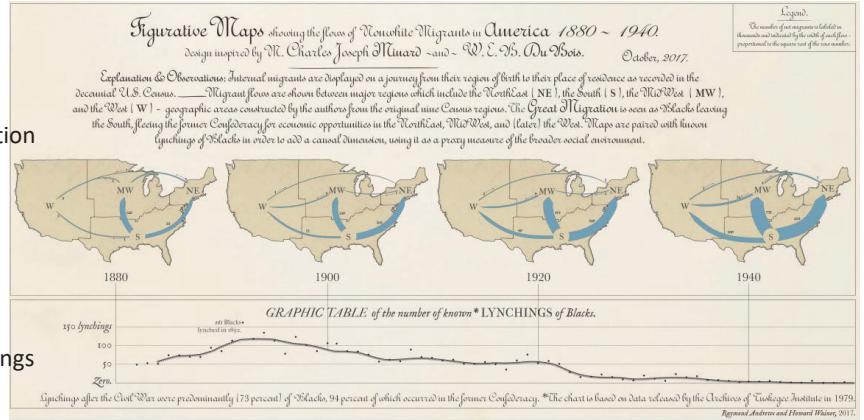


From: <http://www.martingrandjean.ch/connected-world-air-traffic-network/>
See more: <https://flowingdata.com/2016/05/31/air-transportation-network/>

64

The Great Migration

In a graphic tribute to C.-J. Minard and W. E. B. Du Bois, RJ Andrews & Howard Wainer tell the story of the migration of blacks from the southern US after freedom from slavery.



Andrews, R. J. & Wainer, H. The Great Migration: A Graphics Novel Featuring the Contributions of W. E. B. Du Bois and C. J. Minard. *Significance*, 2017, 14, 14-19. See also: <http://infowetrust.com/picturing-the-great-migration/> for the story of this graphic

63

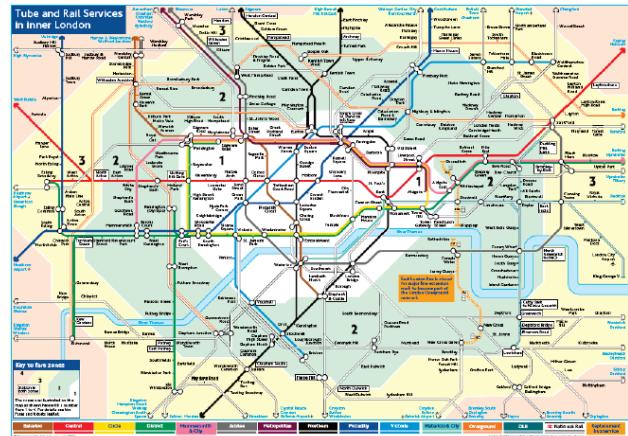
Network visualization: Transport maps

How do I get from Chigwell to Charing Cross?
How much will it cost?

This route map shows the connections and fare zones

The first one was designed by Henry Beck in 1931.

The modern version is zoomable and available on your phone.



See: <https://tfl.gov.uk/maps/travel>

65

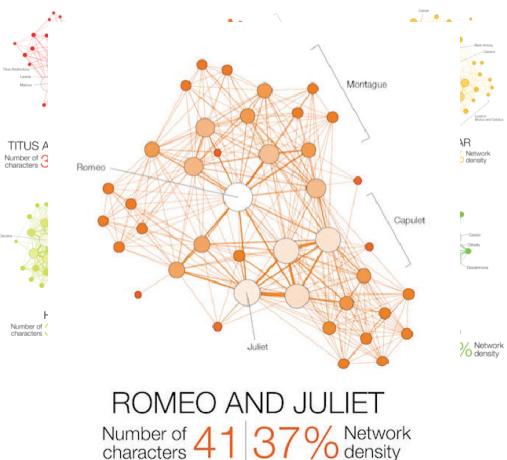
Network visualization: Shakespeare tragedies

A new form of literary criticism?

Martin Grandjean looked at the structure of Shakespeare tragedies through character interactions.

Each circle (node) represents a character, and an edge represents two characters who appeared in the same scene.

The structural characteristics of the graphs have meaningful interpretations.



From: <https://flowingdata.com/2015/12/30/shakespeare-tragedies-as-network-graphs/>

66

Semantic memory: Cognitive structure

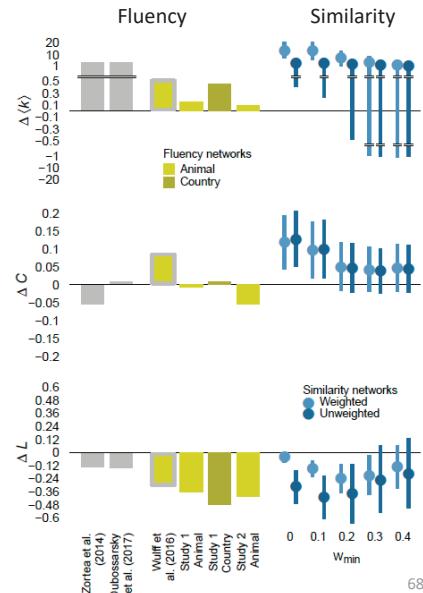
Do younger and older adults differ on measures calculated from their network diagrams?

$\langle k \rangle$: Average “degree” # of connections

C : average local clustering

L : average path length in network

$\Delta()$: young – old difference



68

IMHO, this graph tries to do too much.

The fluency data is most important to their argument.

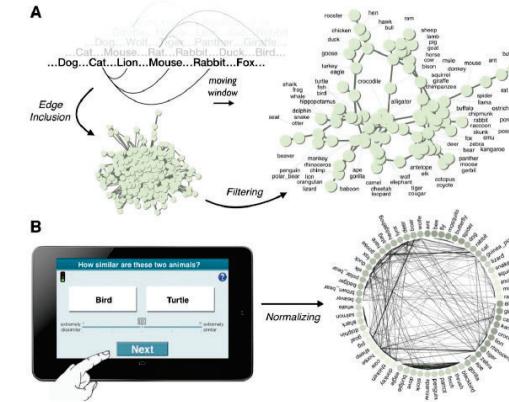
ΔL & $\Delta \langle k \rangle$ show consistent differences between young & old

Semantic memory: Cognitive structure

Various tasks can be used to assess the relations among words/concepts in our semantic memory

The data can be used to calculate measures of **similarity**, and be shown in network or other diagrams

Verbal fluency task: Say/write all the names of [animals, countries, ...] you can in 1 minute.



From: Wulff et al. (2018), Structural differences in the semantic networks of younger and older adults, <https://www.nature.com/articles/s41598-022-11698-4>

67

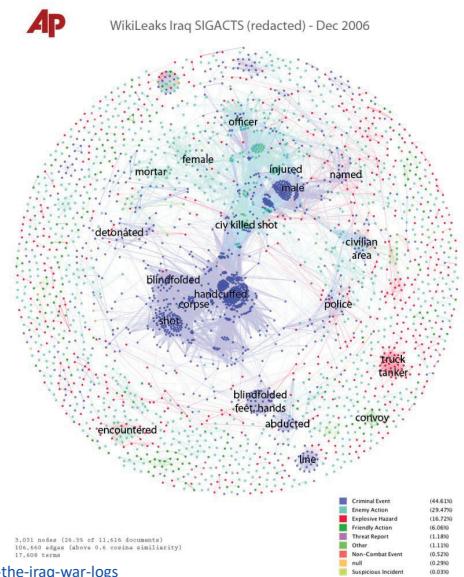
WikiLeaks Iraq war logs

Johnathan Stray & Julian Burgess analyzed > 11,000 documents for SIGACT (“significant action”) reports from the 2006 Iraqi civil war made available by WikiLeaks.

Each report is a dot. Each dot is labelled by the three most “characteristic” words in that report.

Documents that are “similar” have edges drawn between them, width ~ similarity

The graph-drawing algorithm placed similar nodes together



From: <http://jonathanstray.com/a-full-text-visualization-of-the-iraq-war-logs>

70

WikiLeaks Iraq war logs

Certain themes became clear, and could be studied in rich detail
The underlying methods use “term frequency–inverse document frequency”
measures of [text-mining](#).

Murder cluster. All contain the word “corpse”



<http://jonathanstray.com/wp-content/uploads/2010/12/Murders.png>



<http://jonathanstray.com/wp-content/uploads/2010/12/Torture-abduction.png>

71

er
s

Who do I most often interact with?

Three rings to show my twitter world

One ring to rule them all:
@datavisFriendly

Other rings: #datavis,
#maps, #rstats, #psy6135

A collage of 30 circular profile pictures of various people, some with text overlays like "Maarten", "8 POSTS", "TODAY HAVE BEEN", "MT MICHIGANTOUCH.COM", "wp", and "chirpty.com".

Generated by <http://chirpty.com>

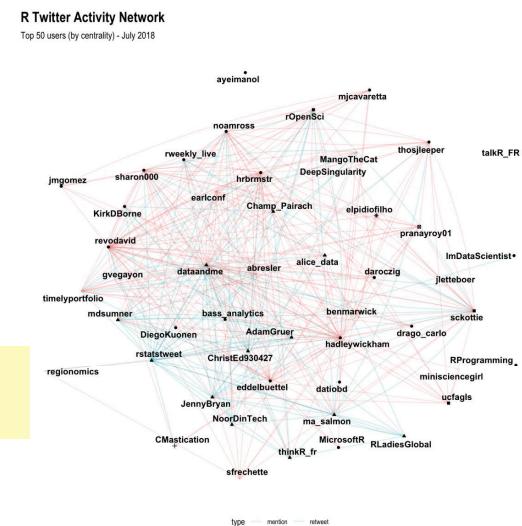
73

Twitter network of R users

Perry Stephenson explores the connections among the top 50 R users on Twitter

The [rtweet](#) package provides access to Twitter info

```
library(rtweet)  
followers <-  
get_followers("datavisFriendly")
```



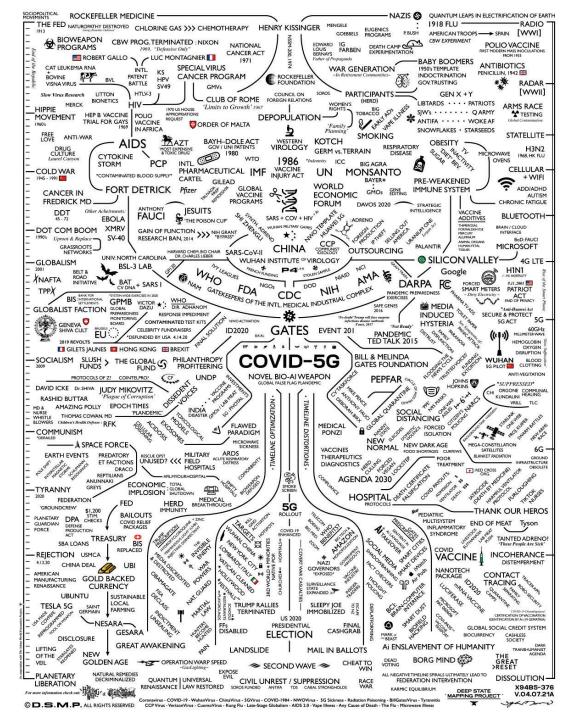
From: <https://perrystephenson.me/2018/09/29/the-r-twitter-network/>

72

Pathological networks

Conspiracy theories (QANON) are aided by diagrams like this suggesting that everything is connected to everything else

- “labyrinthic chartjunk”
 - Information overload
 - The world is chaotic
 - Something MUST BE DONE!

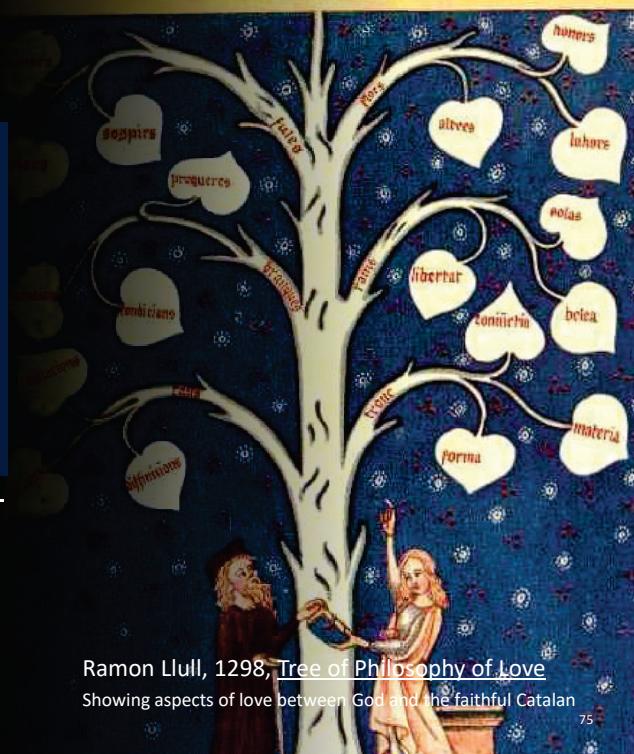


From: <https://deepstatemappingproject.com/>

73

Tree-based Visualization

Branching patterns
History as a tree
Treemaps



Ramon Llull, 1298, [Tree of Philosophy of Love](#)
Showing aspects of love between God and the faithful Catalan

History as a Tree: Geschichtsbaum Europa (2003)

The entire history of Europe in one diagram

- space-filling design: resolution \sim time 2
- natural metaphors for roots, branches

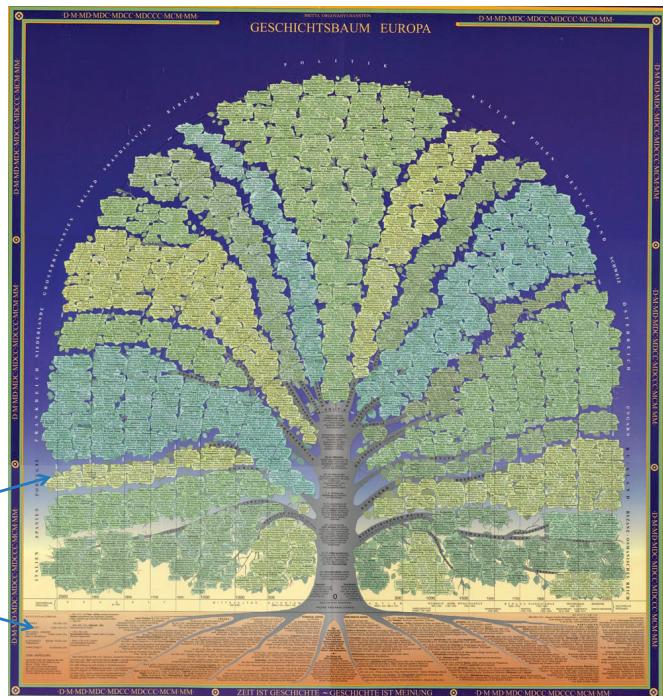


Image: David Rumsey map collection, <https://bit.ly/3GYZ2iw>

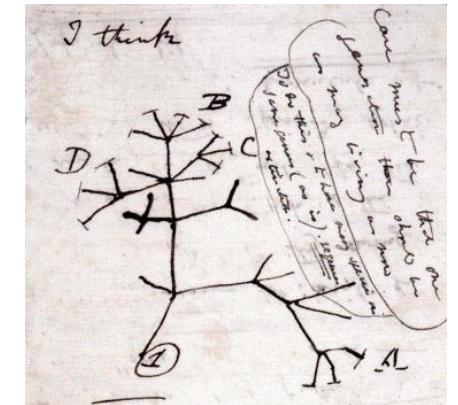
77

Tree diagrams

Trees are natural, organic visual metaphors for [branching processes](#) and [space-filling designs](#).



[Ramon Llull's tree of science](#), showing roots and branches of knowledge

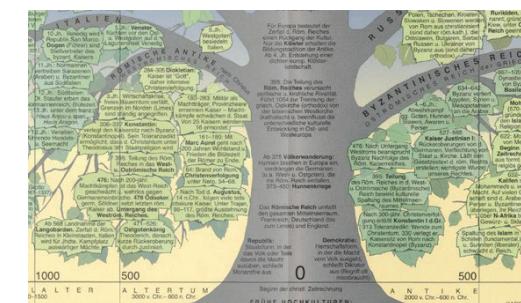
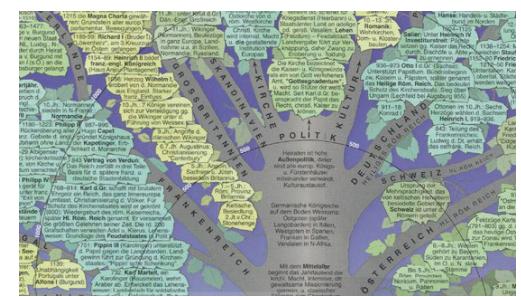


[Charles Darwin's first visual sketch of the evolution of species](#)

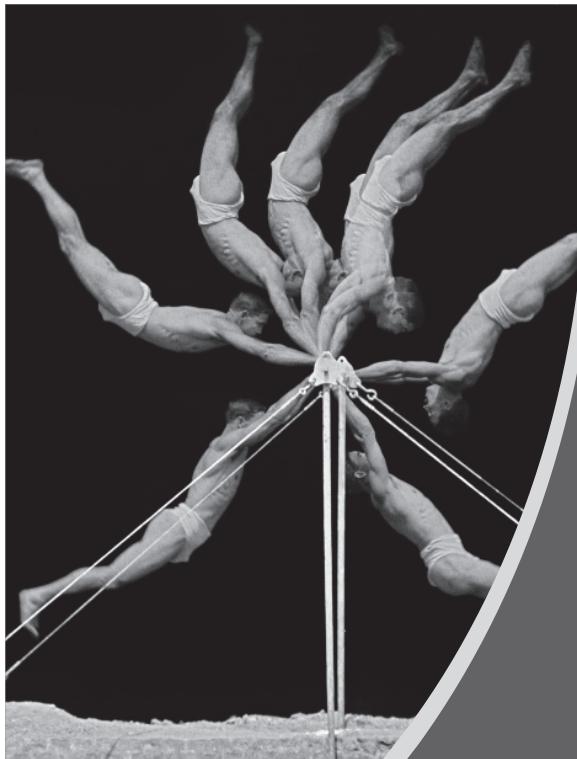
76

History as a Tree

- Branches for countries & domains of thought
- Leaves for all the details



78



Animation & Interactive Graphics

8
4

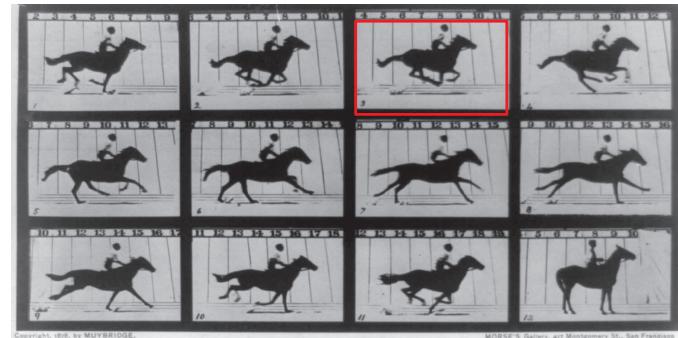
Origins: Visualizing motion
Animated graphics
Dynamically updated graphics
Linking views
Interactive application development frameworks

A wager about a horse in motion

In the late 1800s, a popular quasi-scientific question was: [Does a horse, in a trot, cantor or gallop ever have all four feet off the ground?](#)

This came to be called the [Hypothesis of Unsupported Transit](#)

Eadweard Muybridge solved the problem by automating multiple photographs



Copyright, 1881, by MUYSBRIDGE.
THE HORSE IN MOTION.

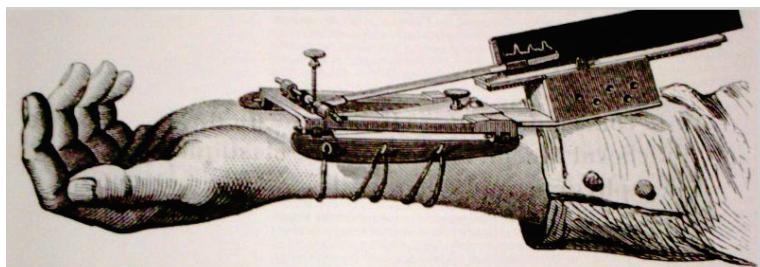
Patent for apparatus applied for.
MUYSBRIDGE.
"SALLIE GARDNER," owned by LELAND SPANFORD; ridden by G. DOMM, running at a 140 gait over the Palo Alto track, 19th June, 1878.
The apparatus consists of two sets of cameras, each set having a series of lenses, each lens being capable of taking a series of successive pictures during a single cycle of the horse. The vertical lines were twenty-seven inches apart; the horizontal lines represent distances of four inches each.

85

É.-J. Marey: A science of visualizing motion

- Physiology: How to make internal physiological processes subject to visual analysis?
 - Invented many graphic recording devices (heart rate, blood pressure, muscle contraction, etc.)
 - “Every kind of observation can be expressed by graphs”

Marey's sphygmograph, recording a visual trace of arterial blood pressure



86

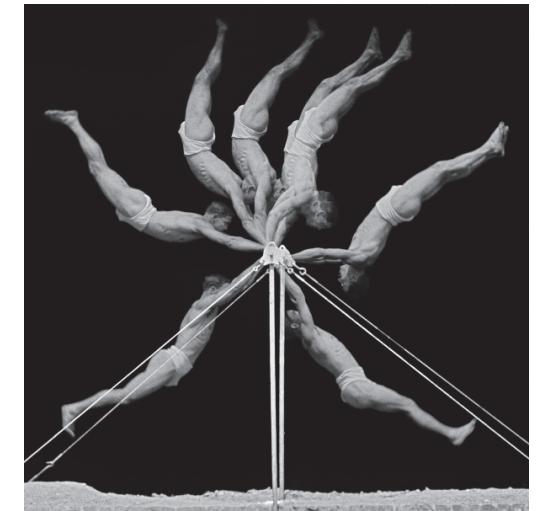
Animation: Chronophotography

Marey pioneered the study of human and animal motion photographically



Fig. 1. Maté d'œuvre de fusil photographique.

The photographic gun, allowing recording of 12 frames/sec. at intervals of 1/720 of a second



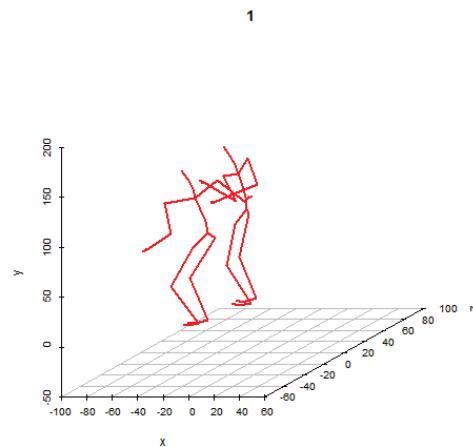
87

Animated graphics

Animated graphics, like movies are just a series of frames strung together in a sequence

The data for this animation come from human figures in motion-capture suits dancing the Charleston.

The Carnegie-Mellon Graphics Lab maintains a Motion Capture Database, <http://mocap.cs.cmu.edu/>



From: <http://blog.revolutionanalytics.com/2017/08/3-d-animations-with-r.html>

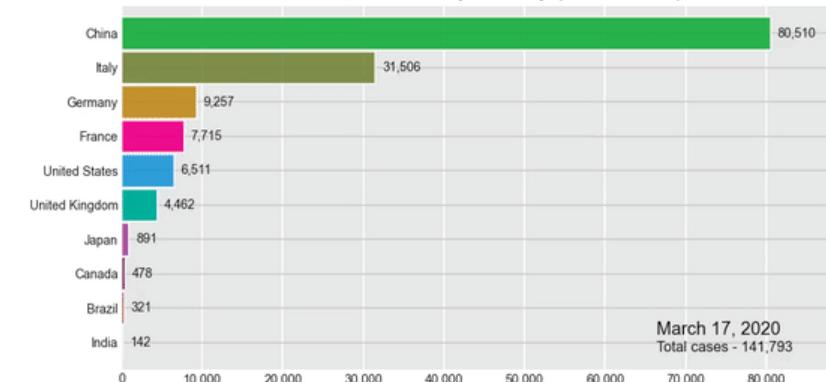
88

Bar chart races

Data that changes over time can often be shown in a simpler animated graphic

This example of a **bar chart race** shows the strengths & weaknesses of this approach.

COVID-19 Cases by Country (JAN 20 - 21)



90

Animated line graph

To show the evolution of time series over time, you can use an animation line graph



The [Data Visualization Catalog](#) shows page views of its top 10 charts
Graphs like this can be done using the [gganimate](#) package

89

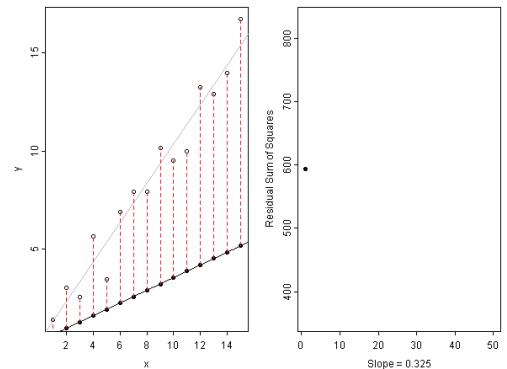
Statistical animations

Statistical concepts can often be illustrated in a dynamic plot of some process.

This example illustrates the idea of least squares fitting of a regression line.

As the slope of the line is varied, the right panel shows the residual sum of squares.

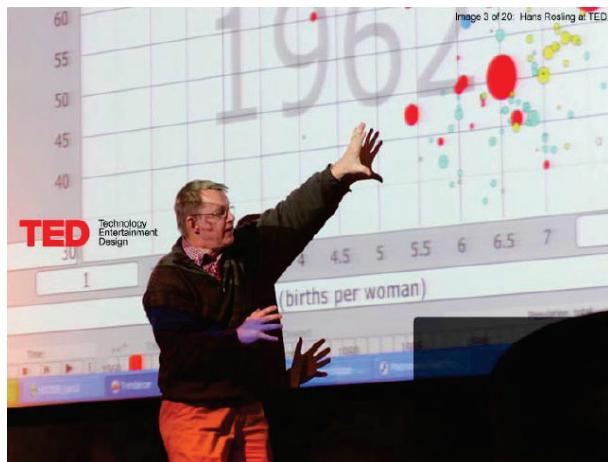
This plot was done using the [animate](#) package in R.



91

Animated graphics

Hans Rosling captivated audiences with dynamic graphics showing changes over time in world health data



Video: Hans Rosling, "The best stats you've ever seen,"
https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

92

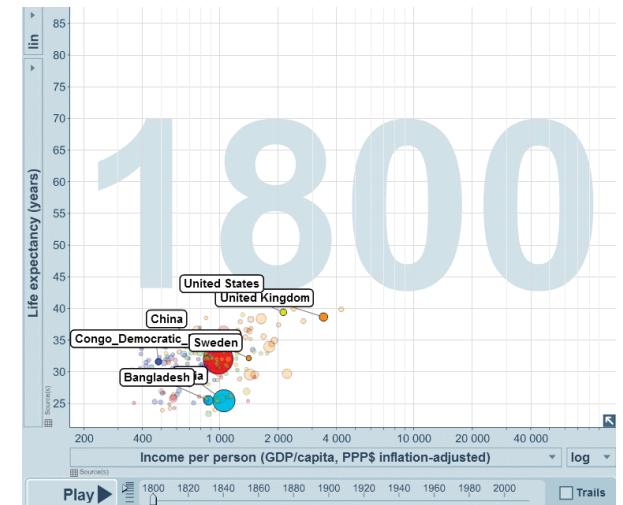
Animation & Interactivity

The Gapminder "moving bubble chart" was the vehicle.

- Choose (x, y) variables
- Choose bubble size variable
- Animate this over time

Liberating the X axis from time opened new vistas for data exploration

Software made this available as a general tool



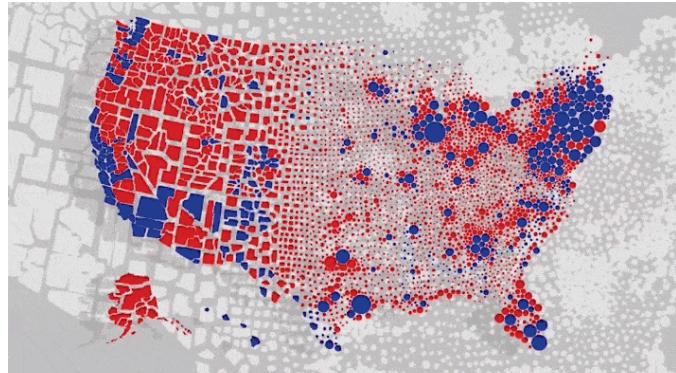
93

Animation: Interpolated views

Animation can also be used to show the difference between two views, using interpolated transitions: Current = α view₁ + (1- α) view₂

This image showed Rep vs. Dem votes in the 2016 US election, contrasting shading by area vs. shading by population.

Land doesn't vote;
people do



[Image: Karim Douieb/Jetpack.ai]

<https://www.fastcompany.com/90572489/u-s-election-maps-are-wildly-misleading-so-this-designer-fixed-them>

94

Linking animated views

This example links a dendrogram to a grand tour and map of the USArrests data to visualize a classification in 5 dimensions

The grand tour animates a series of 2D projections of the 5D data

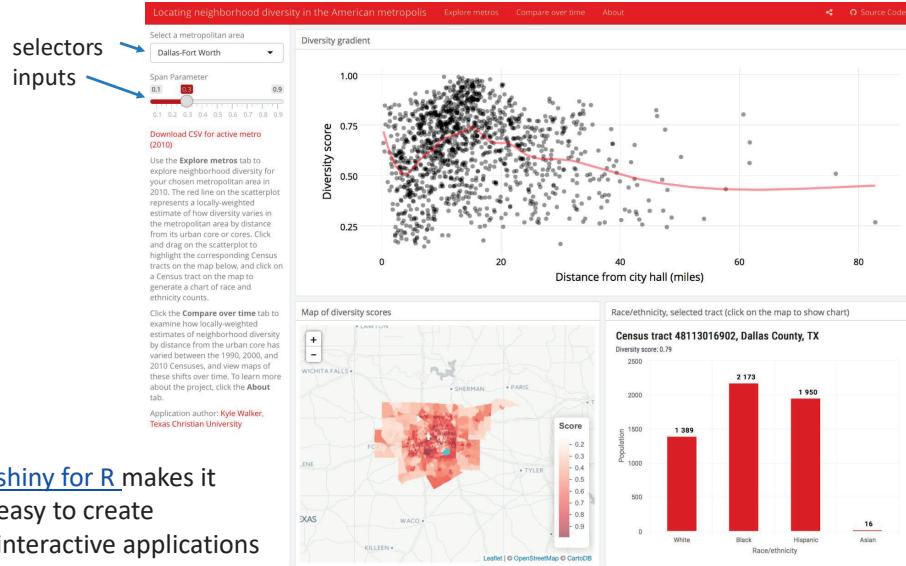
The image is recorded as a GIF



From: Carson Sievert, <https://plotly-book.cpsievert.me/linking-animated-views.html>

95

Interactive application frameworks



[shiny for R](#) makes it
easy to create
interactive applications

https://walkerke.shinyapps.io/neighborhood_diversity/

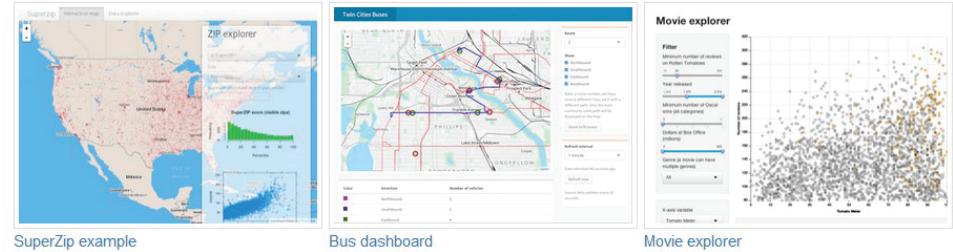
96

shiny gallery

There is now a large collection of shiny applications, <https://shiny.rstudio.com/gallery/>. These integrate other interactive web software: d3, Leaflet, Google Charts, ...

Interactive visualizations

Shiny is designed for fully interactive visualization, using JavaScript libraries like d3, Leaflet, and Google Charts.



97

Summary

- The topics here were largely about data graphs, for analysis & presentation. Mainly not Info-graphics
 - Quantitative data: different forms for 1D, 1.5D, 2D, 3+D data
 - Categorical data: often best shown as areas ~ frequency (bar plots, mosaic plots)
- Thematic maps: visualizing spatially varying data
 - Raw data with different visual encodings
 - Spatial statistical models provide some smoothings
- Networks/trees: visualizing connections
- Animation: show changes over time or space
- Interaction: allow the viewer to explore the data

98