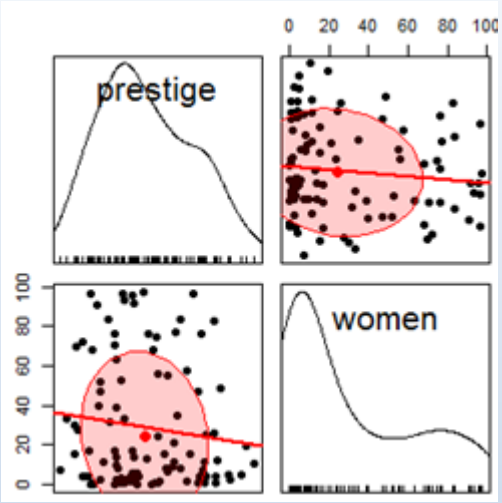
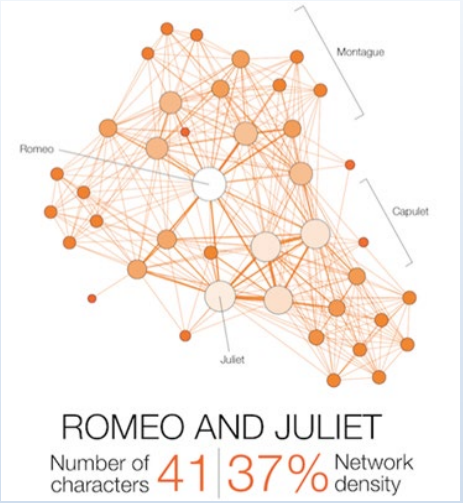




Varieties of information visualization

Michael Friendly
Psych 6135

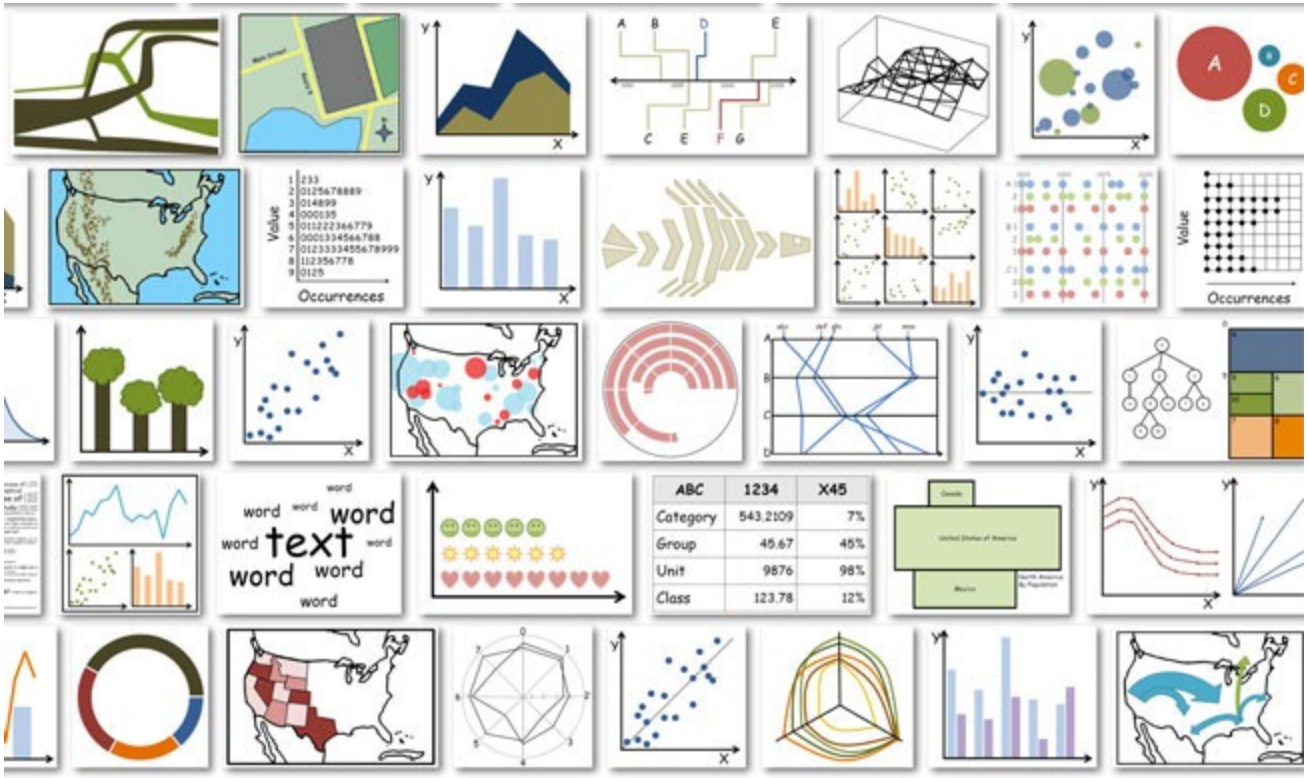
https://friendly.github.io/6135

@datvisFriendly

So many types

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

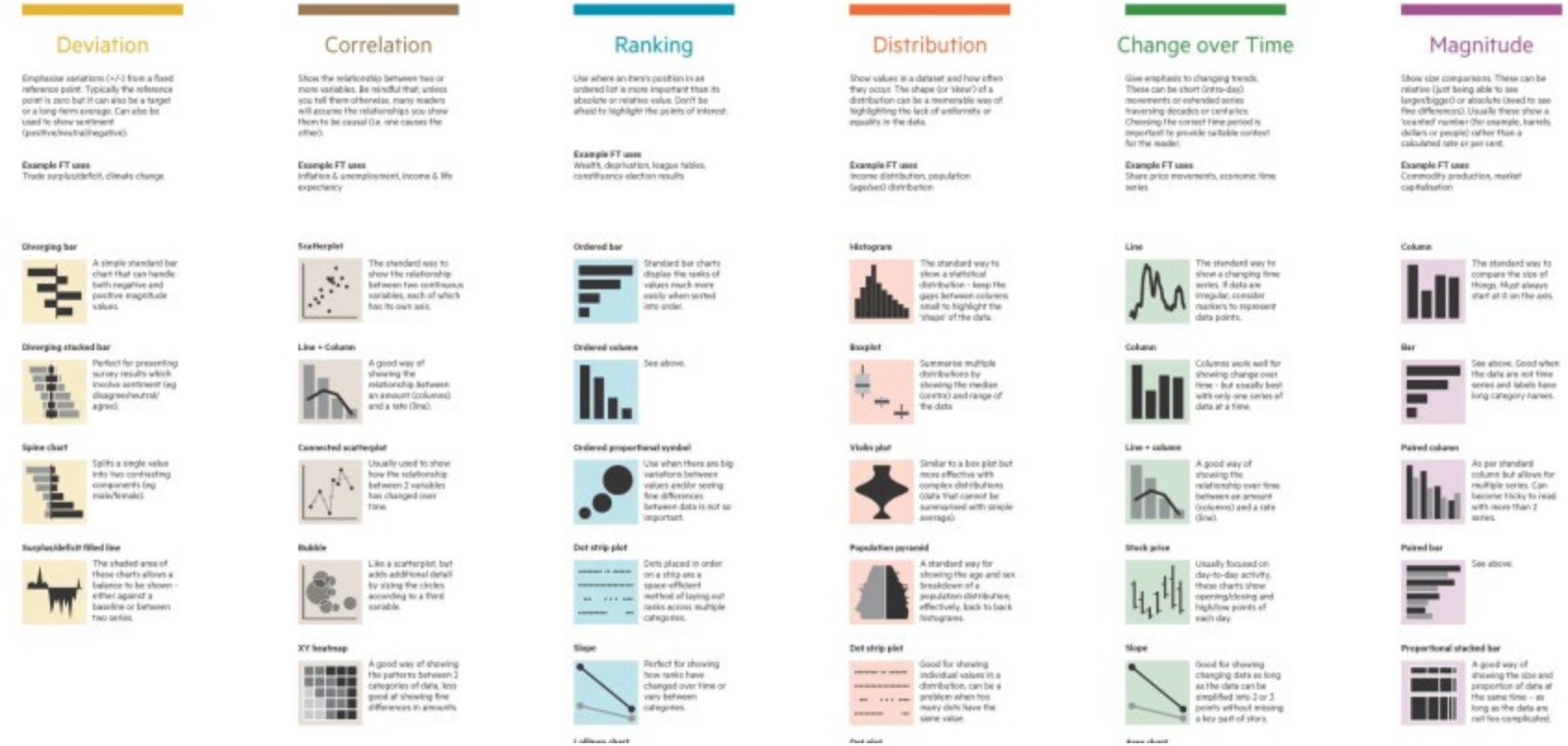
- What are their features?
- What tasks are they good for? – Accuracy or speed of judgment? Memorability?



How to classify ???

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

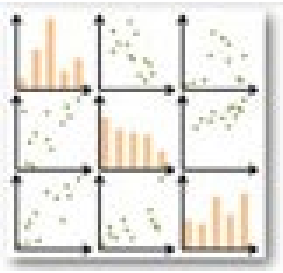
“What do I want to show?”

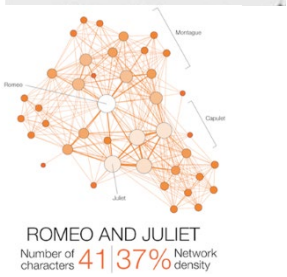


Financial Times, Chart Doctor, https://github.com/ft-interactive/chart-doctor

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





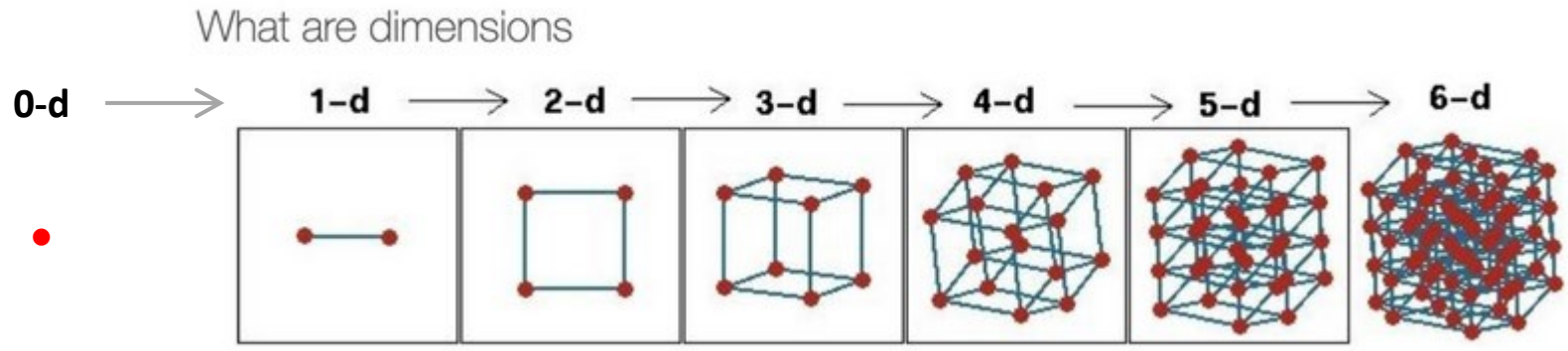


Fig credit: Di Cook @visnut

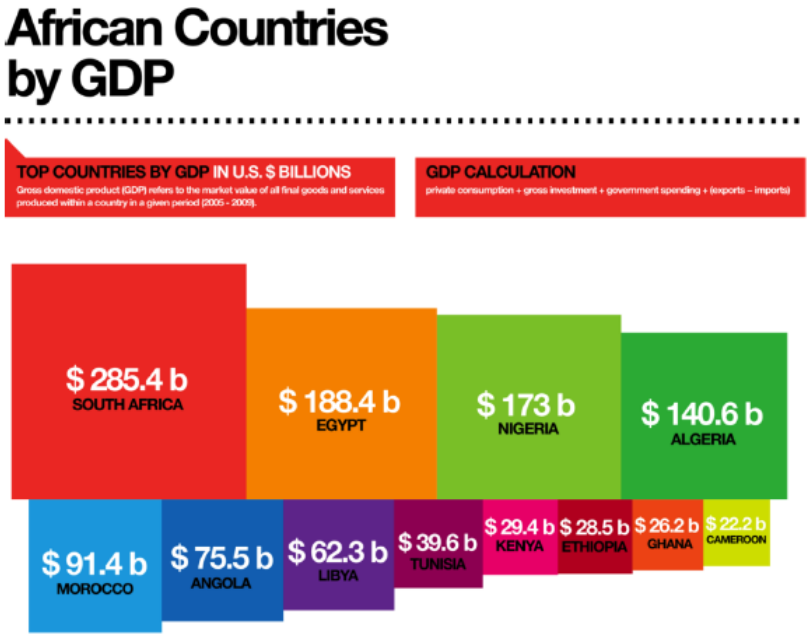
1 D
1.5 D
2 D
3 D
n D ?

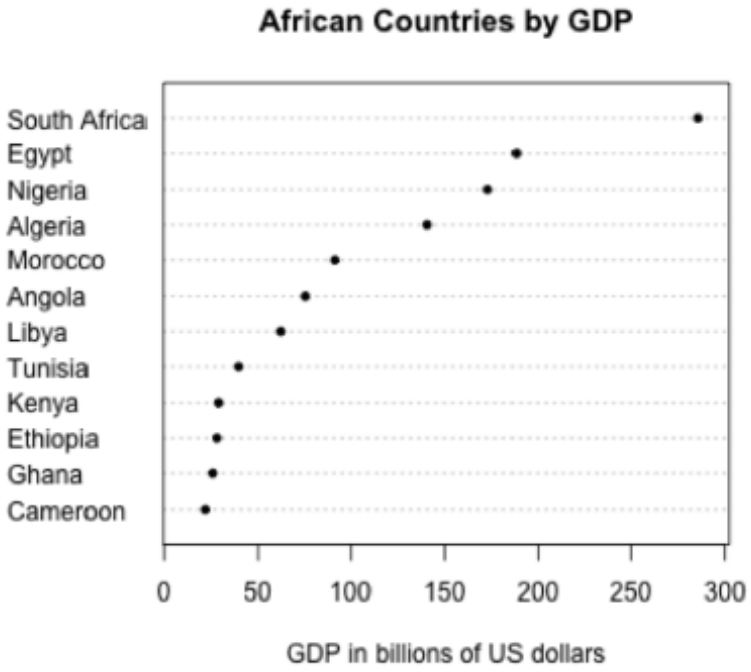
Data graphs can be roughly ordered by the number of variables, data dimensions shown in a given graph

Data graphs

1D: Infographic vs. Data graphic

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





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

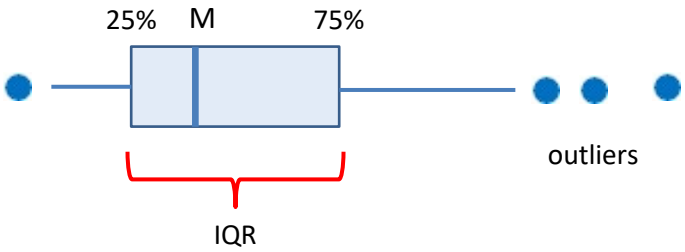
One might argue that this statistical graph makes comparisons easier

1.5D: Dotplots & boxplots

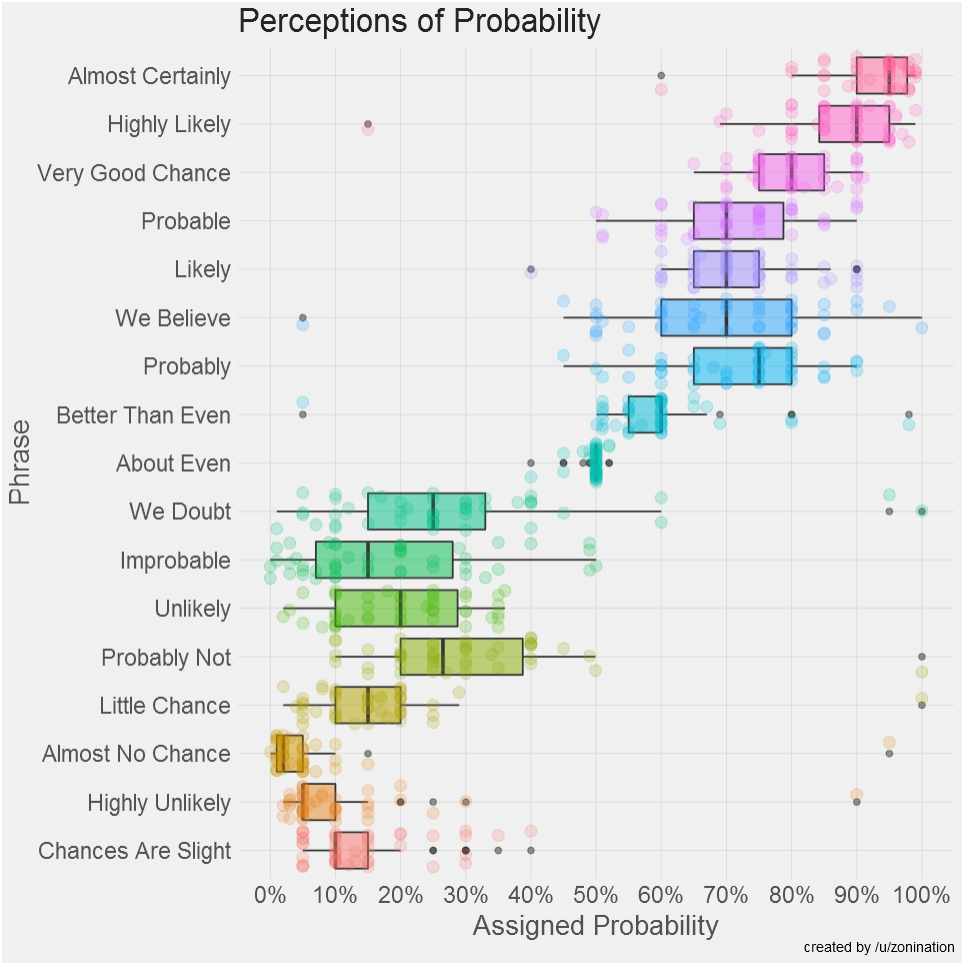
What number do you give to a probability phrase?

Boxplots summarize the important characteristics of a univariate data distribution:

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

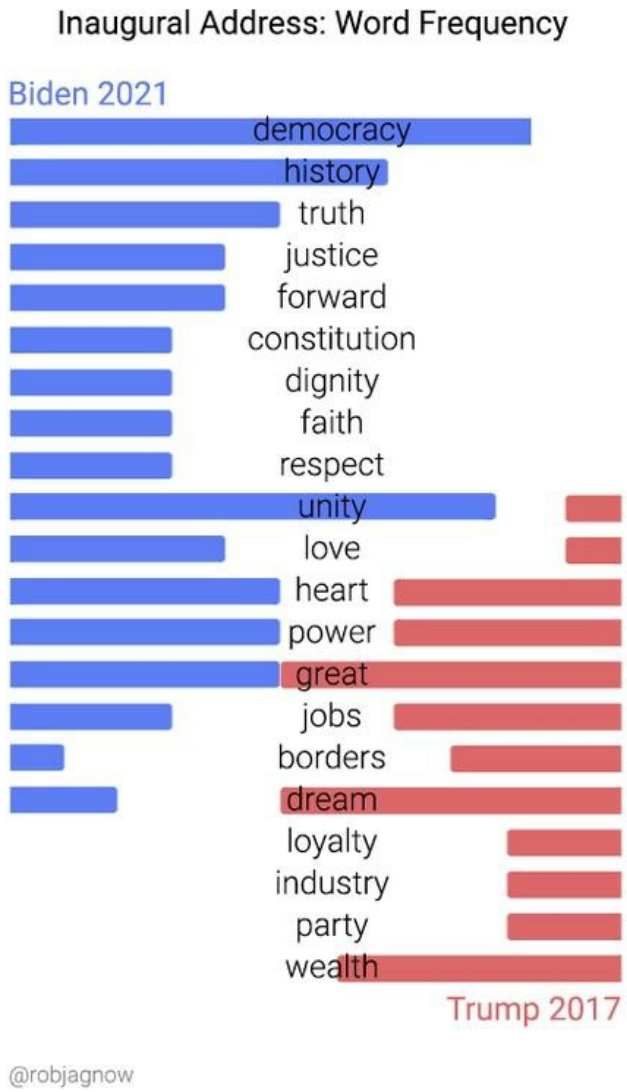


This example overlays the boxplot with a jittered dotplot, so we can also see the individual observations



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

1.5D: Text bar charts



- 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/

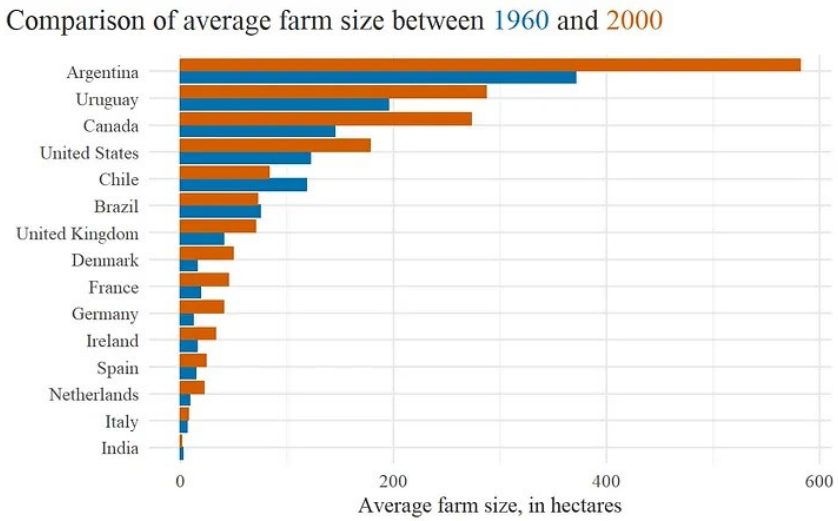
Bar chart variations

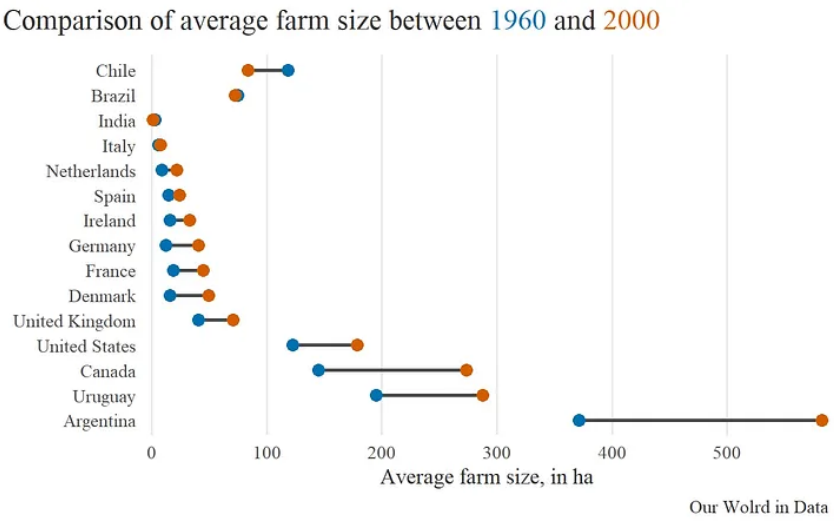
Bar charts are often used to make comparisons between two series

But your eyes must move around to compare

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





Which countries gained the most from 1960 – 2000?

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

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

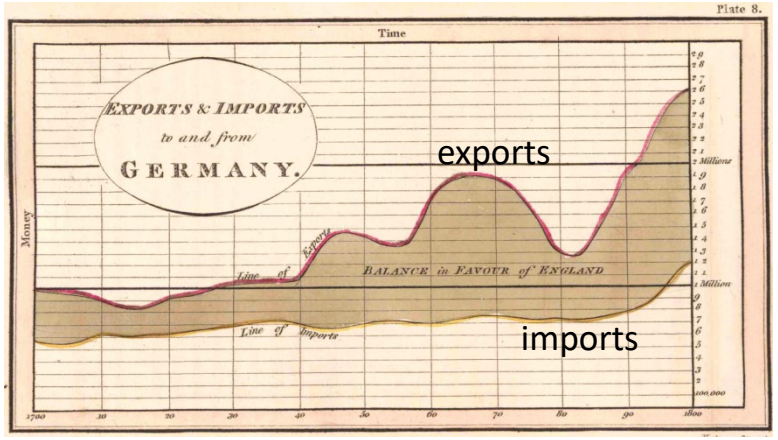
One curve for imports, one for exports

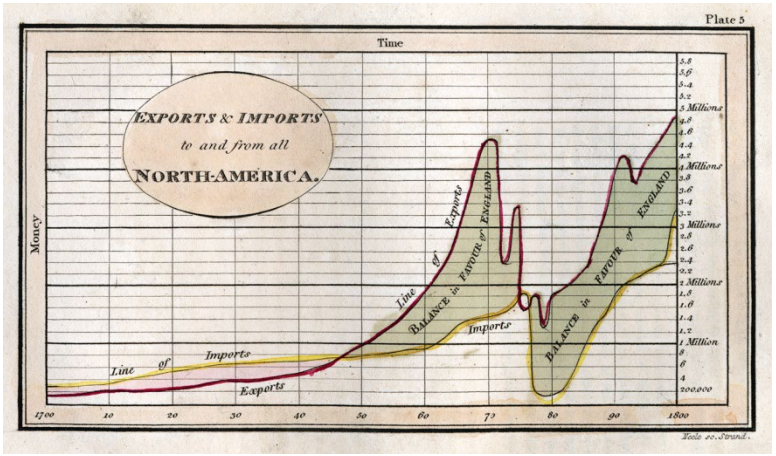
The balance of trade could be seen as the difference between the curves

Trade with Germany was consistently in favor of England

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

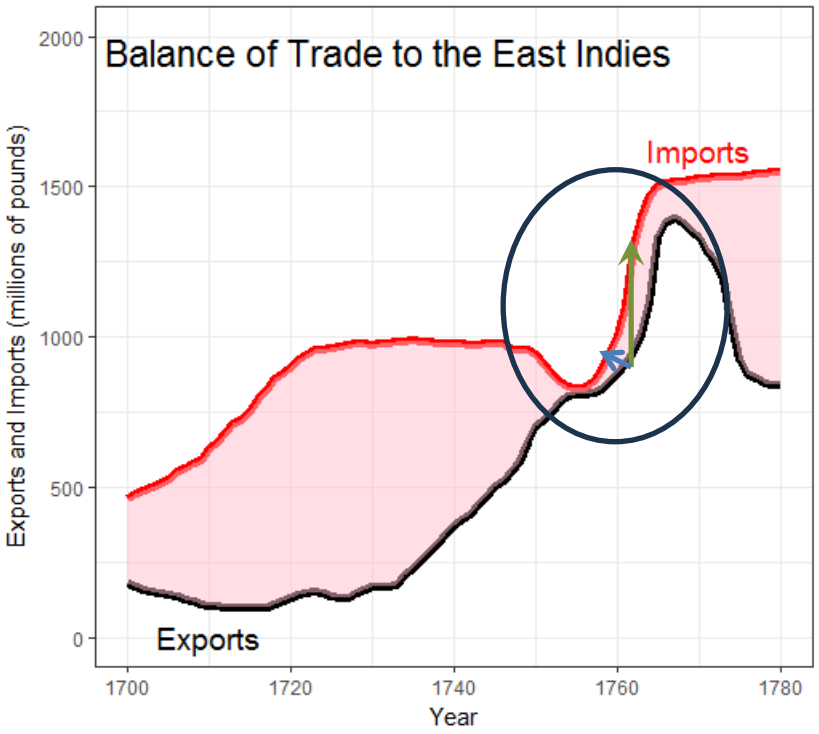
Economic 'history' could now be visualized and explained



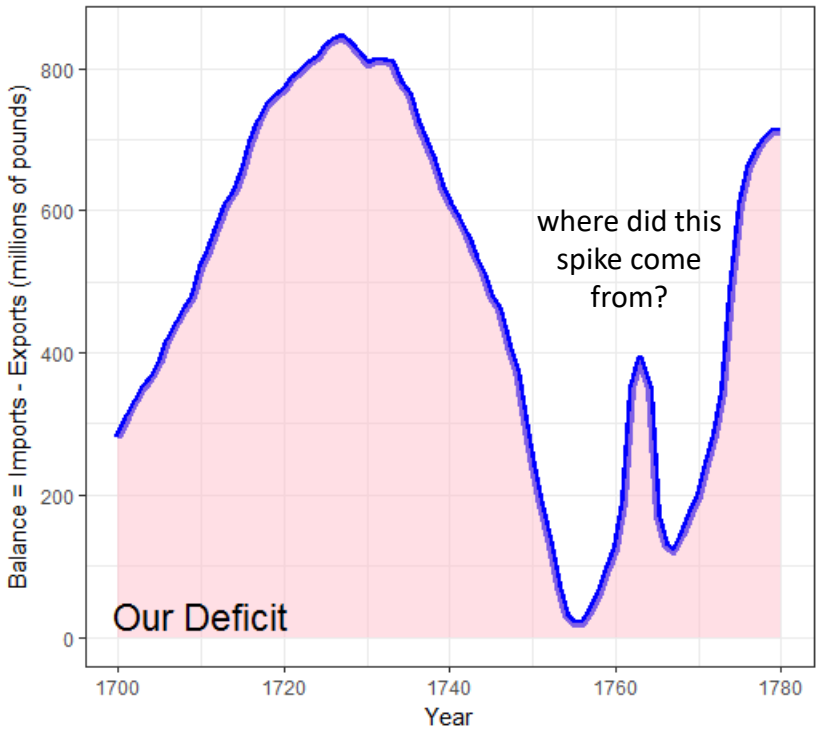


Psychology: Distances between curves

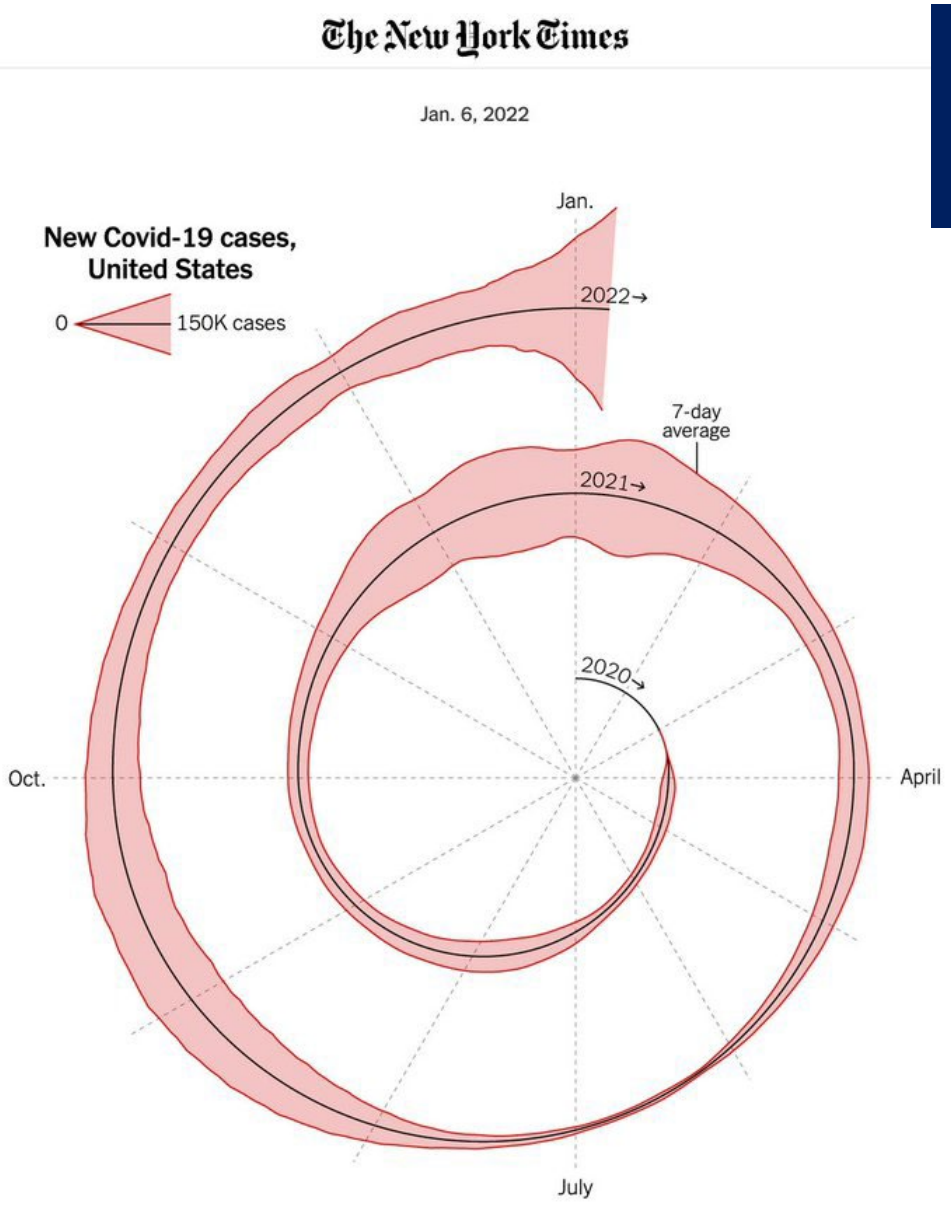
What Playfair didn't know is that judgments of **distance** between curves are biased
We tend to see the perpendicular distance rather than the vertical distance



Plotting balance of trade directly



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?

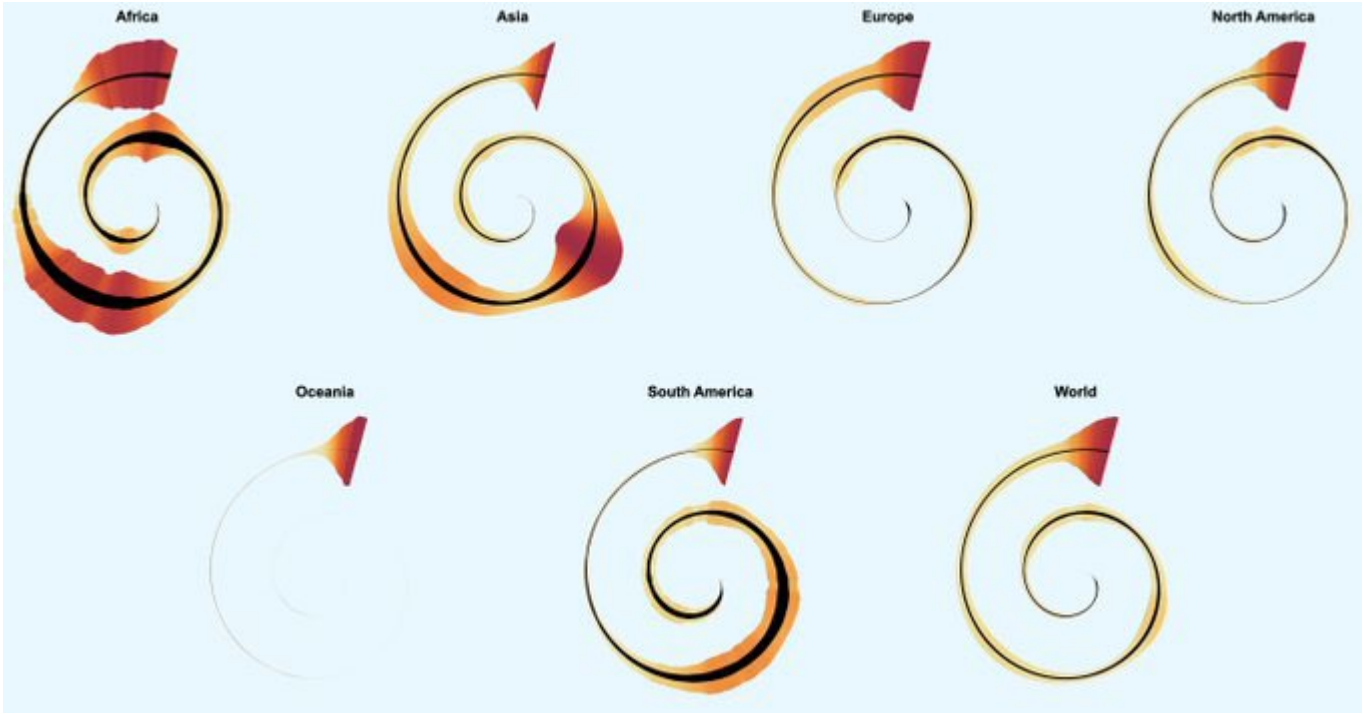
See commentary: https://www.youtube.com/watch?v=YWOWF9Vi4Mw

Multi-sprials

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

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

Color: smoothed new Covid-19 cases
Black: Deaths (10x scale)

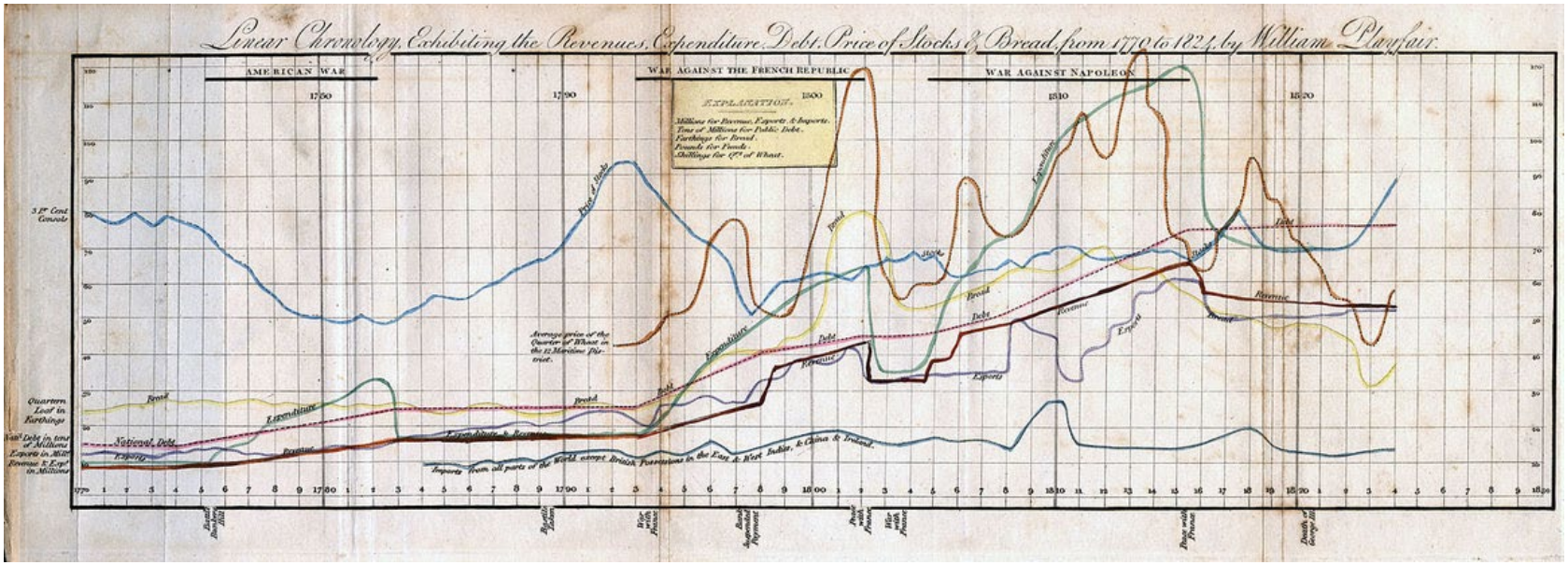


https://twitter.com/h_i_g_s_c_h/status/1483195230404947968

Multiple time series graphs

Things get messy when there are many series to be compared

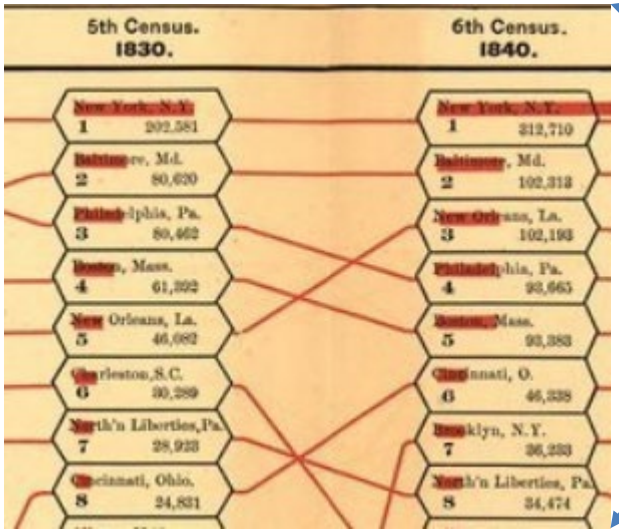
- To be fair, this was designed as timeline of history– a visual story of economics (prices, wages, imports, exports, debt)
- History shown as a strip-chart recording (e.g., EKG)
- It was Playfair's last graph.
- Perhaps the earliest example of **visual storytelling**



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

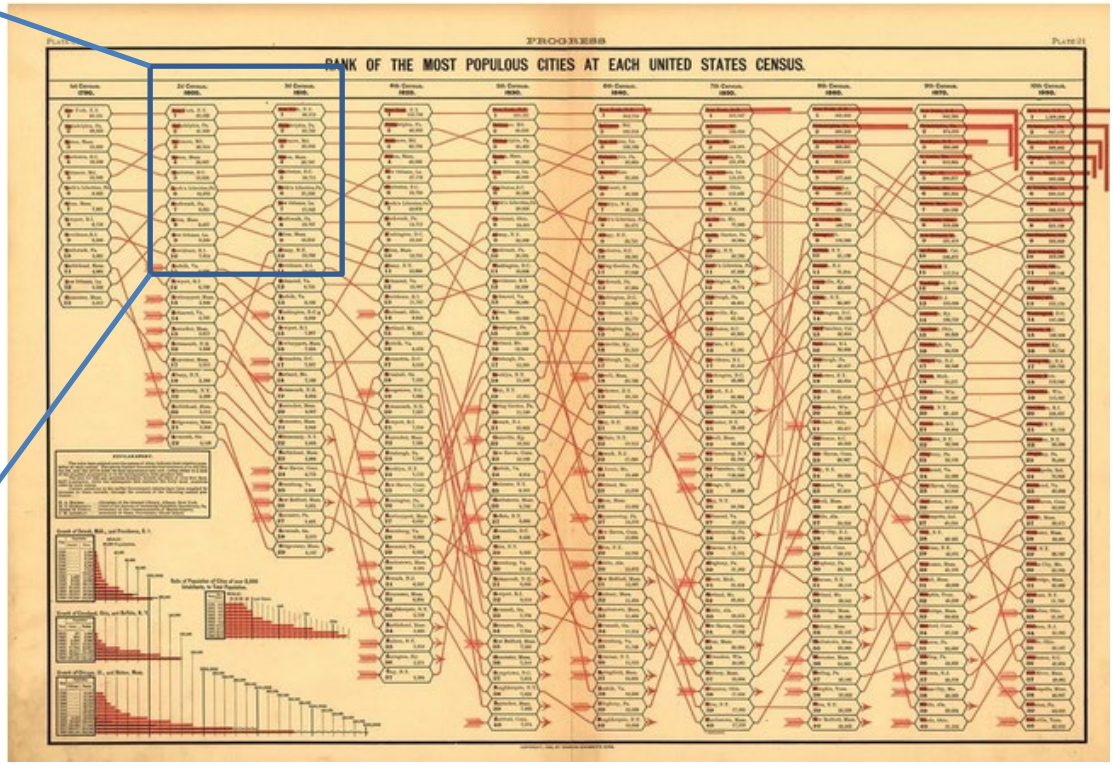
Parallel ranked list charts

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

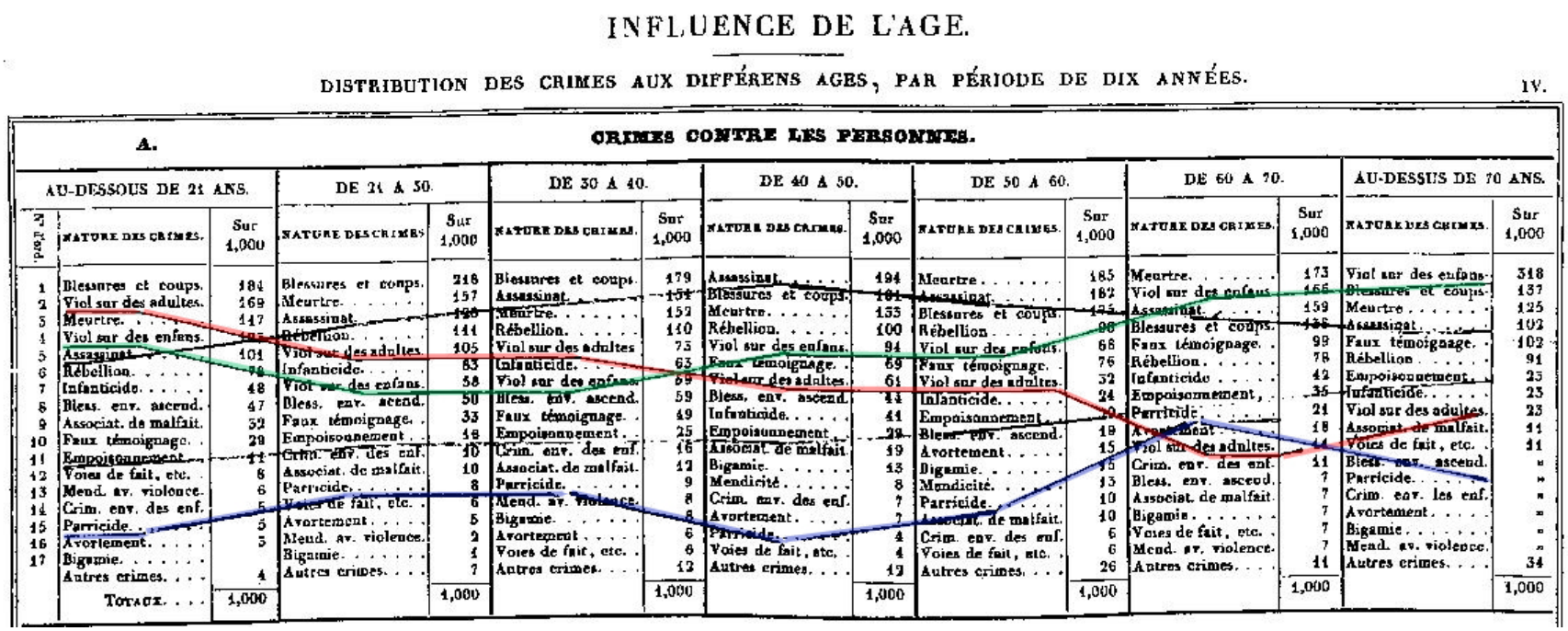


Ranks of populations of US cities across census years
From: *Statistical Atlas of the United States* (1880)

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



INFLUENCE DE L'AGE.

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

A. CRIMES CONTRE LES PERSONNES.

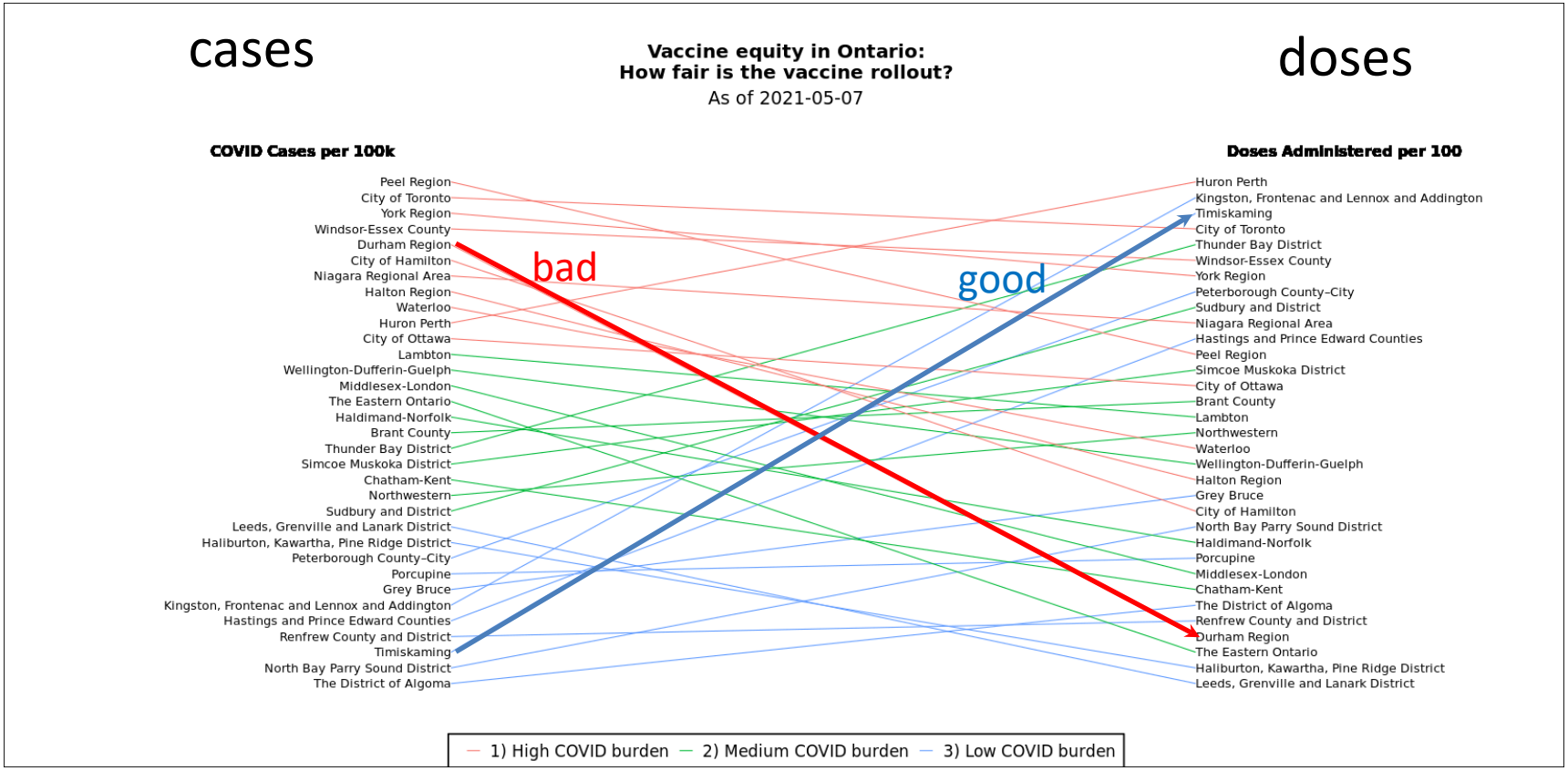
	AU-DESSOUS DE 21 ANS.		DE 21 A 30.		DE 30 A 40.		DE 40 A 50.		DE 50 A 60.		DE 60 A 70.		AU-DESSUS DE 70 ANS.	
Rang	Nature des crimes.	Sur 1,000	Nature des crimes.	Sur 1,000	Nature des crimes.	Sur 1,000	Nature des crimes.	Sur 1,000	Nature des crimes.	Sur 1,000	Nature des crimes.	Sur 1,000	Nature des crimes.	Sur 1,000
1	Blessures et coups.	184	Blessures et coups.	216	Blessures et coups.	179	Assassinat.	194	Meurtre.	185	Meurtre.	173	Viol sur des enfans.	318
2	Viol sur des adultes.	169	Assassinat.	157	Assassinat.	154	Blessures et coups.	181	Assassinat.	182	Viol sur des enfans.	166	Blessures et coups.	137
3	Meurtre.	147	Meurtre.	156	Meurtre.	152	Meurtre.	155	Blessures et coups.	125	Assassinat.	159	Meurtre.	125
4	Viol sur des enfans.	125	Rébellion.	111	Rébellion.	110	Rébellion.	100	Rébellion.	98	Blessures et coups.	136	Assassinat.	102
5	Assassinat.	101	Viol sur des adultes.	105	Viol sur des adultes.	73	Viol sur des enfans.	94	Viol sur des enfans.	68	Faux témoignage.	99	Faux témoignage.	102
6	Rébellion.	78	Infanticide.	83	Infanticide.	63	Faux témoignage.	69	Faux témoignage.	76	Rébellion.	78	Rébellion.	91
7	Infanticide.	48	Viol sur des enfans.	58	Viol sur des enfans.	59	Viol sur des adultes.	61	Viol sur des adultes.	52	Infanticide.	42	Empoisonnement.	25
8	Bless. env. ascend.	47	Bless. env. ascend.	50	Bless. env. ascend.	59	Bless. env. ascend.	44	Infanticide.	24	Empoisonnement.	35	Infanticide.	23
9	Associat. de malfait.	32	Faux témoignage.	33	Faux témoignage.	49	Infanticide.	41	Empoisonnement.	20	Parricide.	21	Viol sur des adultes.	23
10	Faux témoignage.	29	Empoisonnement.	18	Empoisonnement.	25	Empoisonnement.	29	Bless. env. ascend.	19	Avortement.	18	Assassinat. de malfait.	11
11	Empoisonnement.	11	Crim. env. des enf.	10	Crim. env. des enf.	16	Associat. de malfait.	19	Avortement.	15	Viol sur des adultes.	14	Voies de fait, etc.	11
12	Voies de fait, etc.	8	Associat. de malfait.	10	Associat. de malfait.	12	Bigamie.	13	Bigamie.	13	Crim. env. des enf.	11	Bless. env. ascend.	»
13	Mend. av. violence.	6	Parricide.	8	Parricide.	9	Mendicité.	8	Mendicité.	13	Bless. env. ascend.	7	Parricide.	»
14	Crim. env. des enf.	5	Voies de fait, etc.	6	Mend. av. violence.	8	Crim. env. des enf.	7	Parricide.	10	Associat. de malfait.	7	Crim. env. les enf.	»
15	Parricide.	5	Avortement.	5	Bigamie.	8	Avortement.	7	Associat. de malfait.	10	Bigamie.	7	Avortement.	»
16	Avortement.	5	Mend. av. violence.	2	Avortement.	6	Parricide.	4	Crim. env. des enf.	6	Voies de fait, etc.	7	Bigamie.	»
17	Bigamie.		Bigamie.	1	Voies de fait, etc.	6	Voies de fait, etc.	4	Voies de fait, etc.	6	Mend. av. violence.	7	Mend. av. violence.	»
	Autres crimes.	4	Autres crimes.	7	Autres crimes.	12	Autres crimes.	13	Autres crimes.	26	Autres crimes.	11	Autres crimes.	34
	Totaux.	1,000		1,000		1,000		1,000		1,000		1,000		1,000

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

Slope graph of COVID: Cases vs. Doses

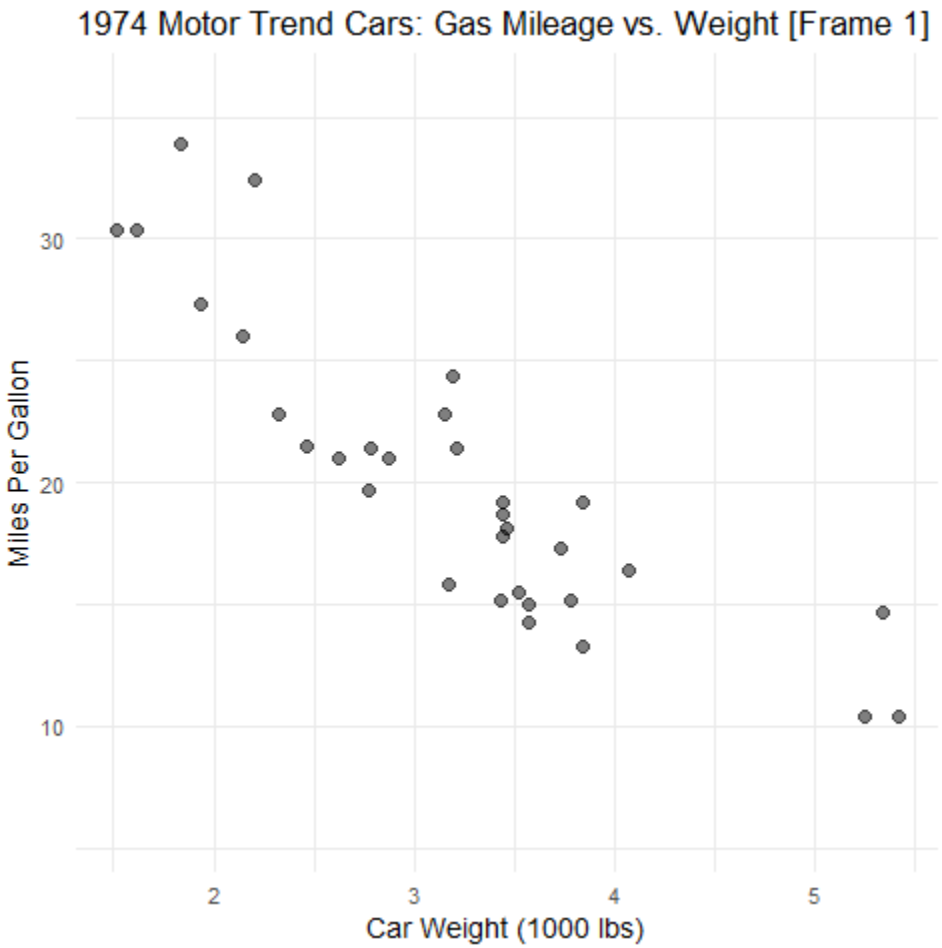
Vaccine equity → all lines should be ≈ flat

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



Graphic by Matthew Dubins, @MDubins

2D: Scatterplots



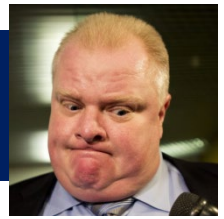
Goals: Want to see--

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

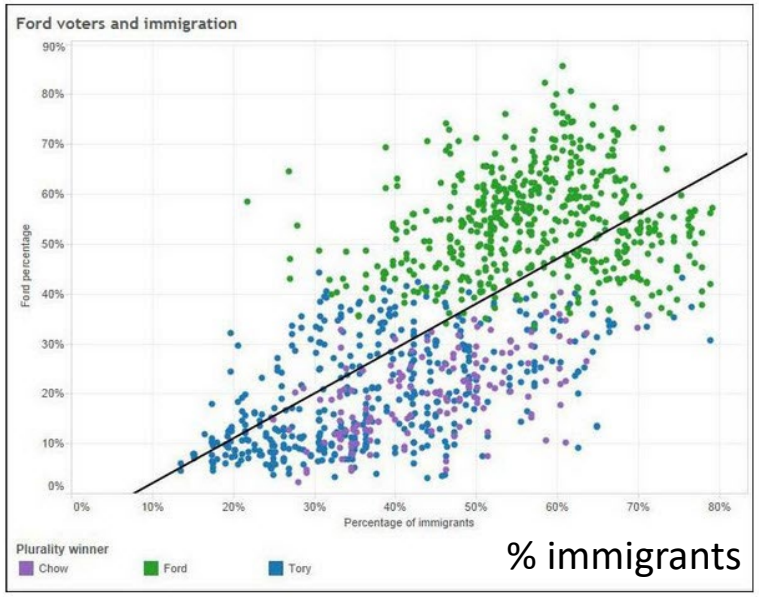
Annotations can help:

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

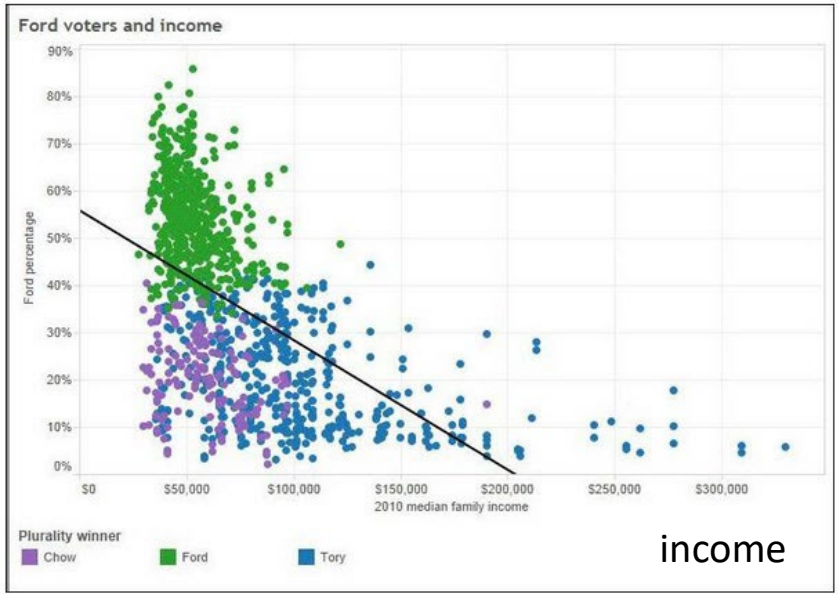
2D: Scatterplots: Ford Nation



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



Ford percentage

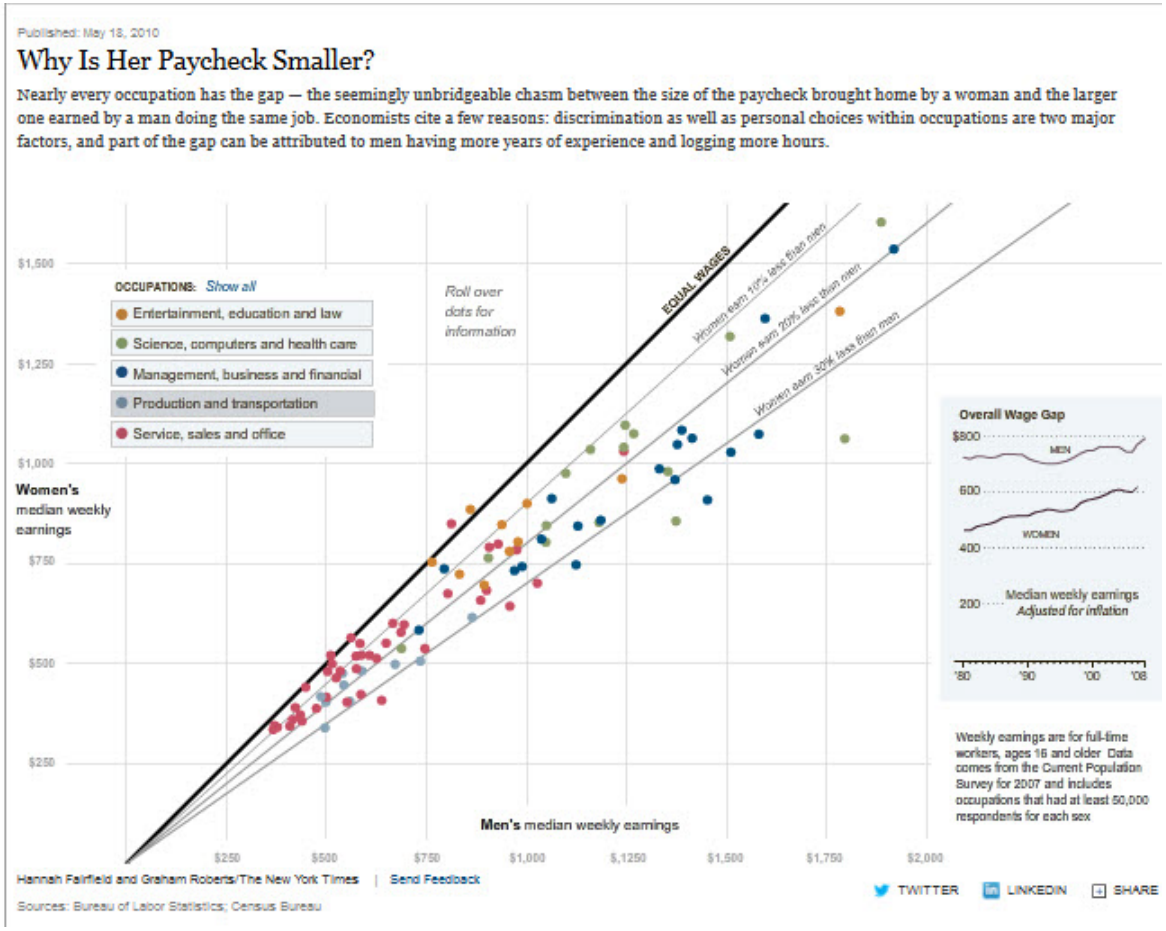


These simple scatterplots by data journalist Patrick Cain use simple enhancements:

- 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/

Scatterplots: Wage gap



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 45º 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

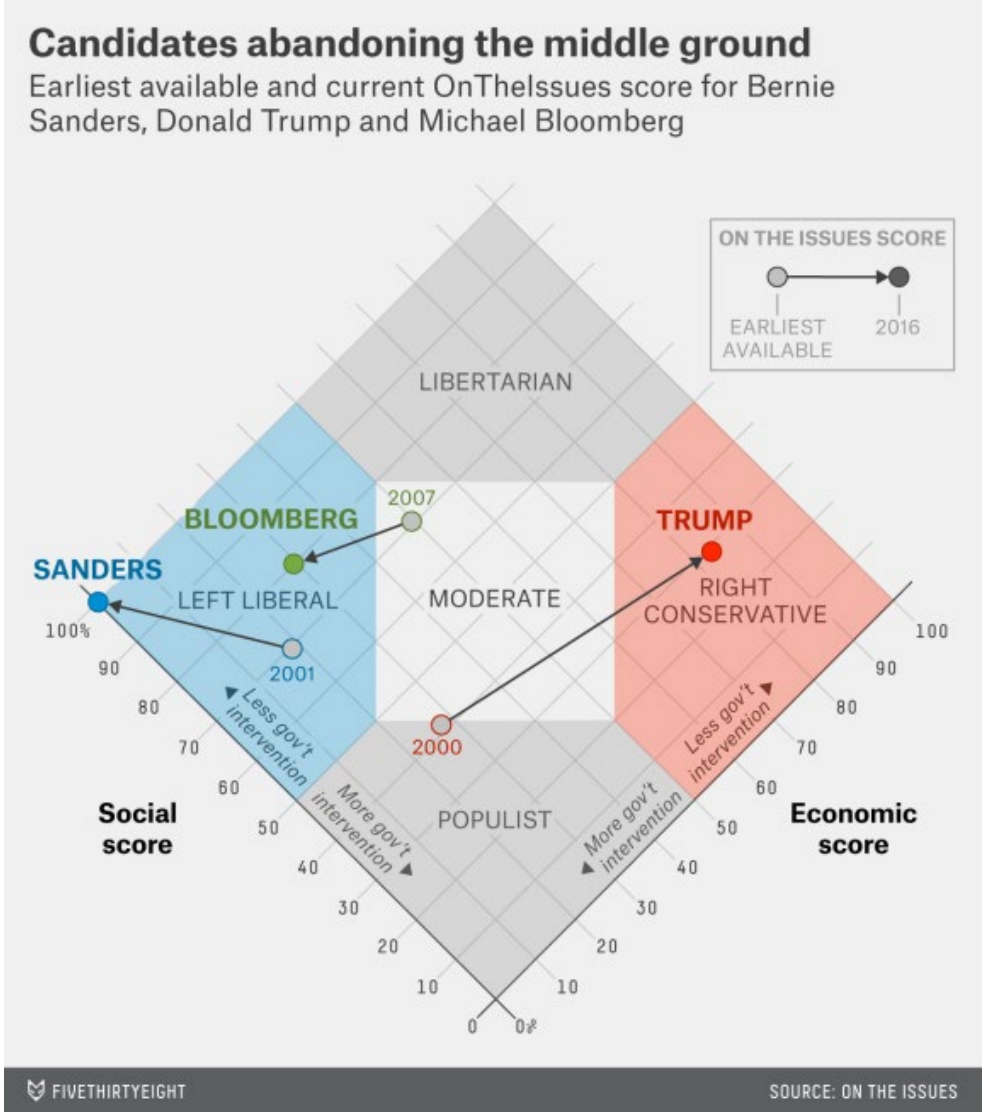
Scatterplots: InfoVis

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



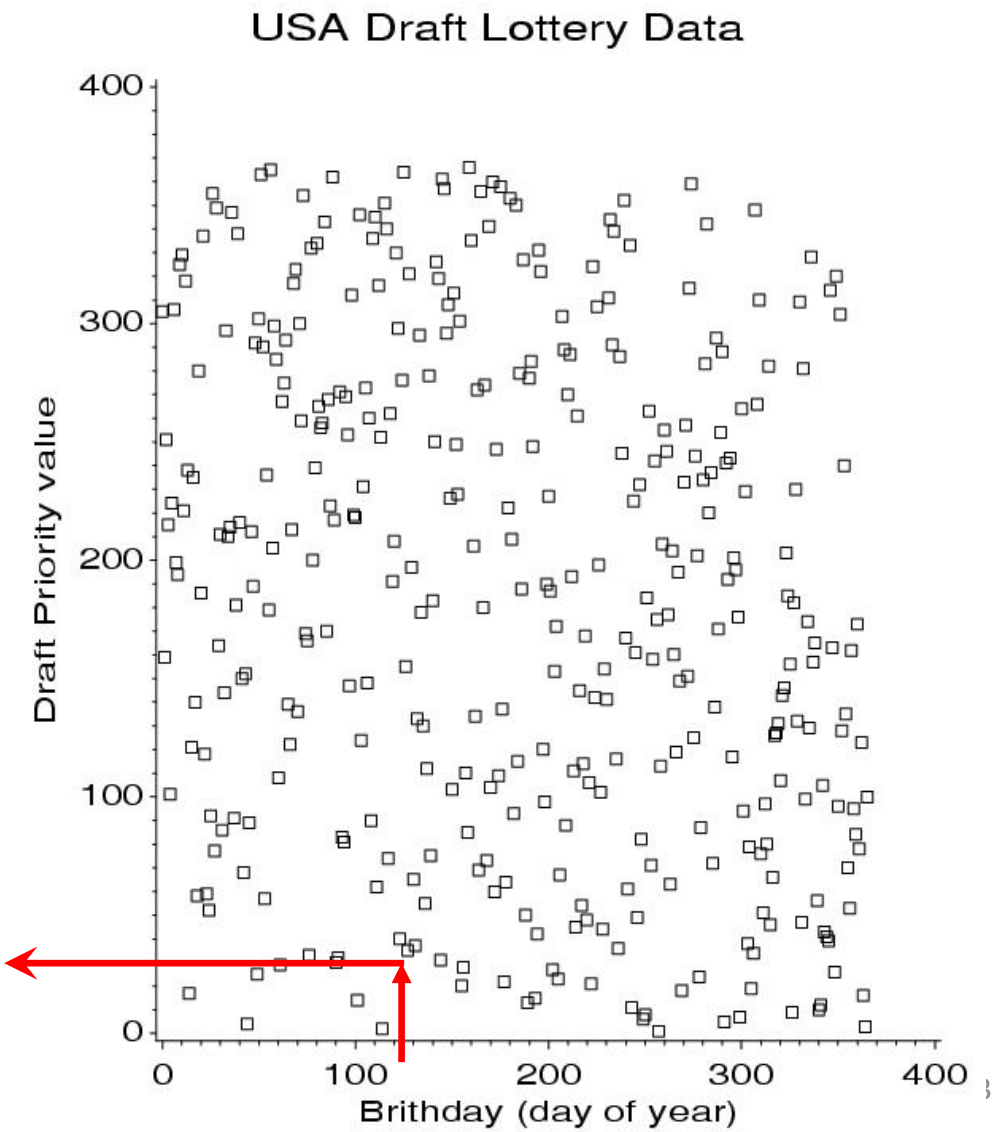
Scatterplots: Annotations enhance perception

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

Me (May 7):
127 → priority = 35

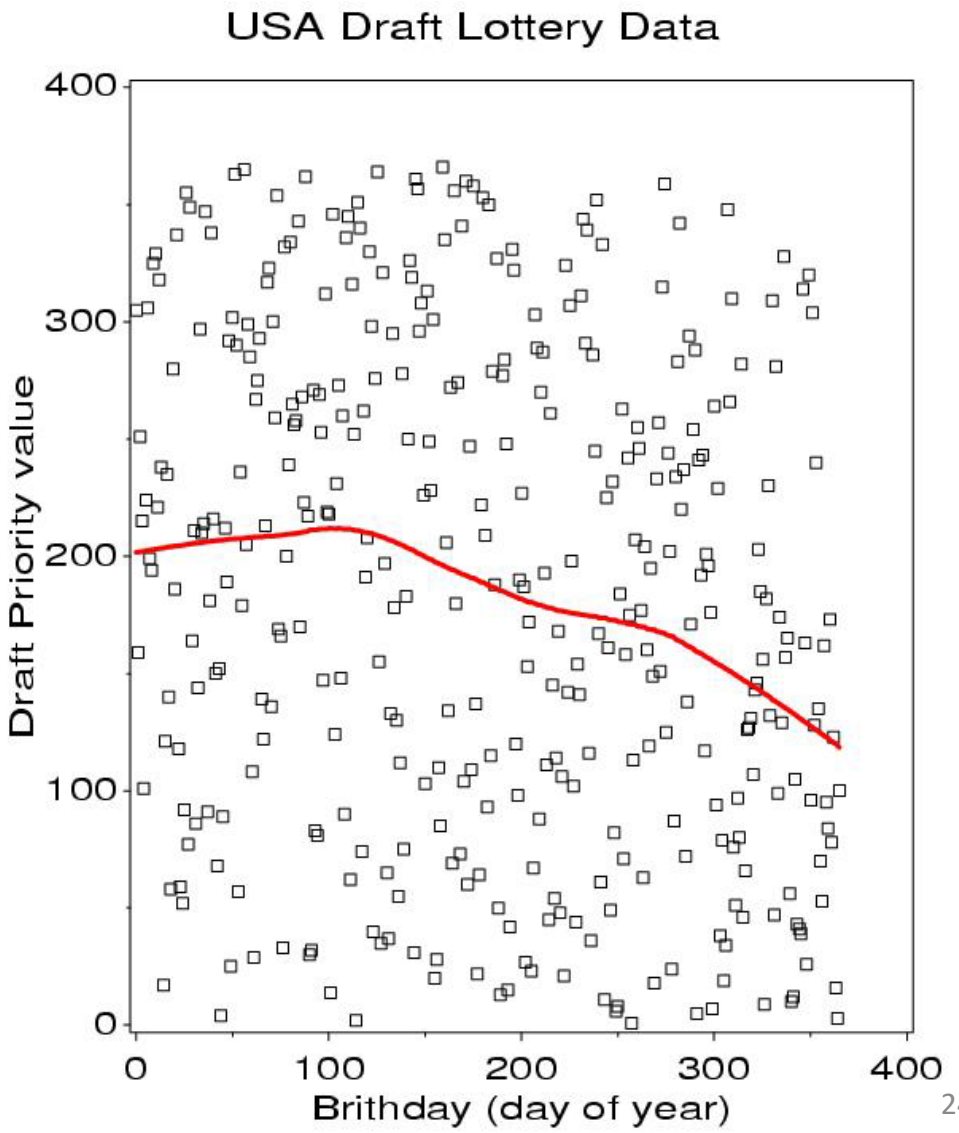


Scatterplots: Smoothing enhances perception

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:

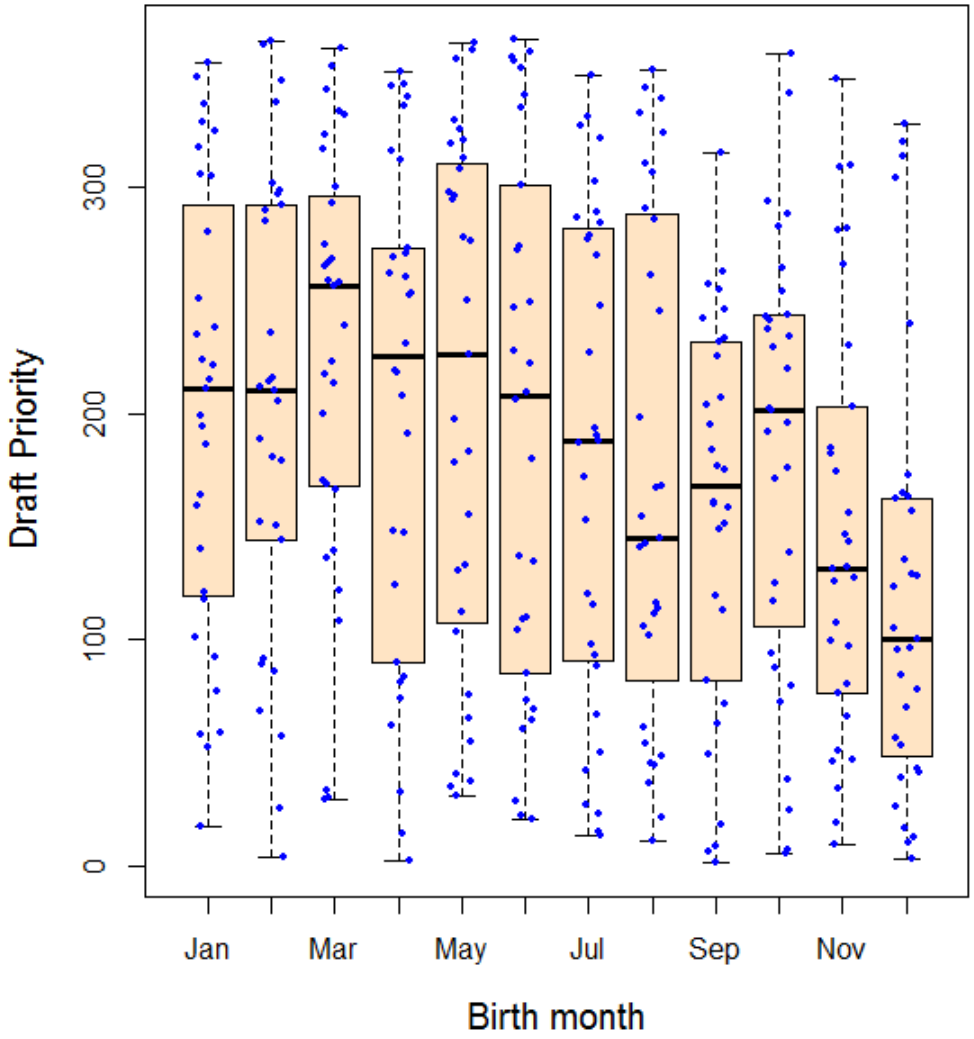


Smoothing by grouping and summarization

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



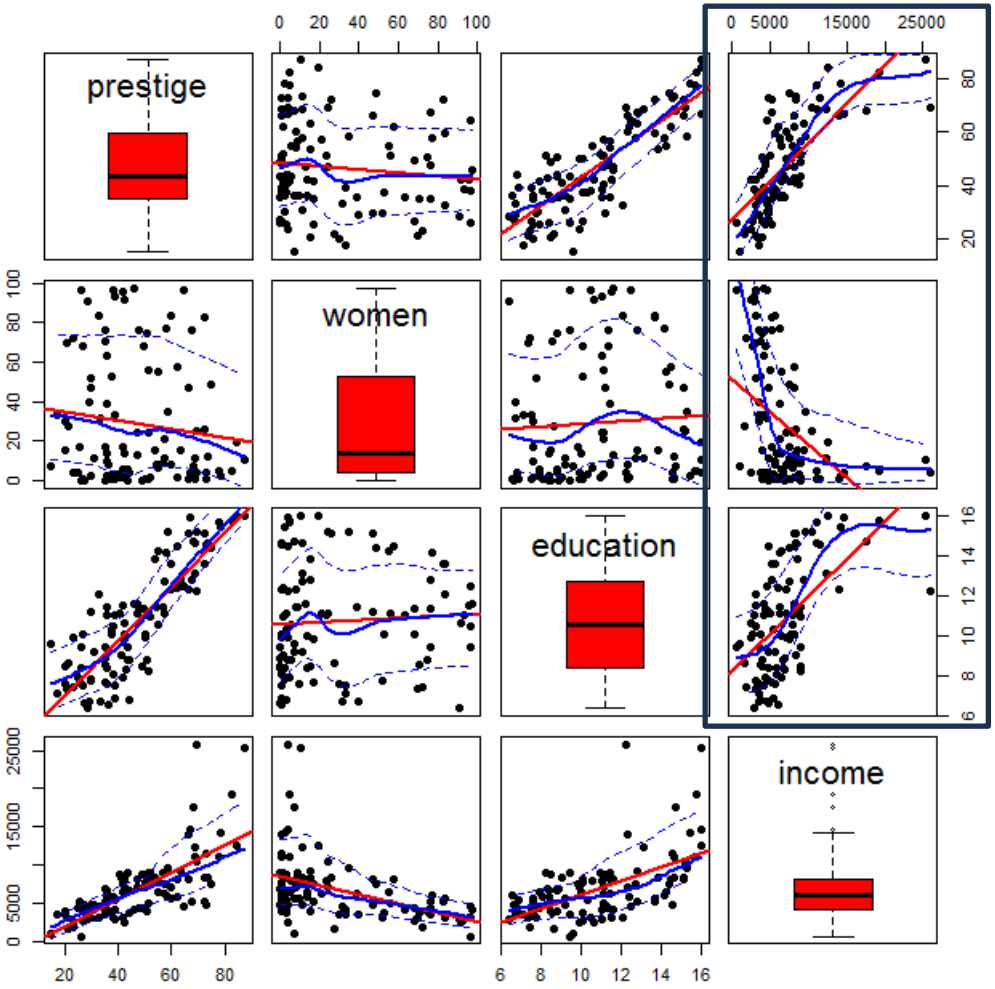
Scatterplot matrices

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 smooths to show non-linearity

This figure uses scatterplotMatrix() in the car package. There are many options.



Scatterplot matrices

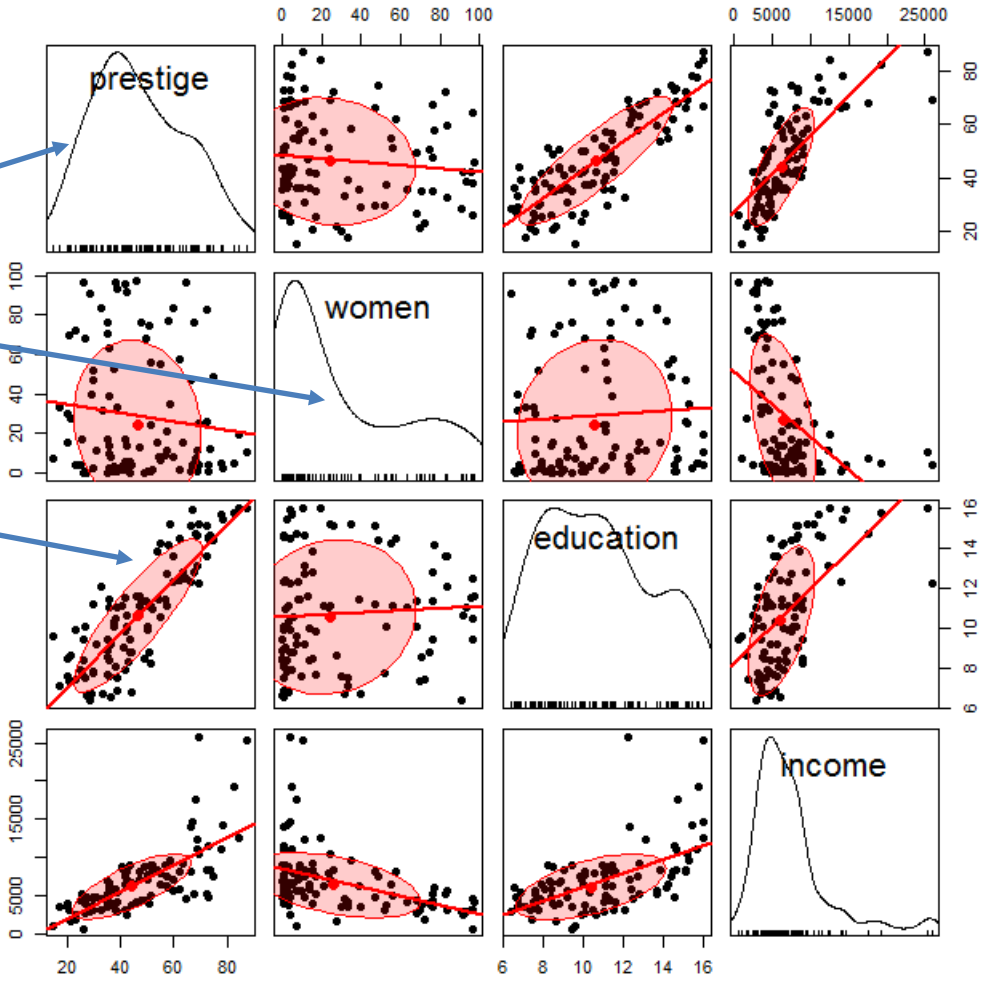
Essential idea: You can choose different methods to render the diagonal & off-diagonal panels

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.



Larger data sets

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

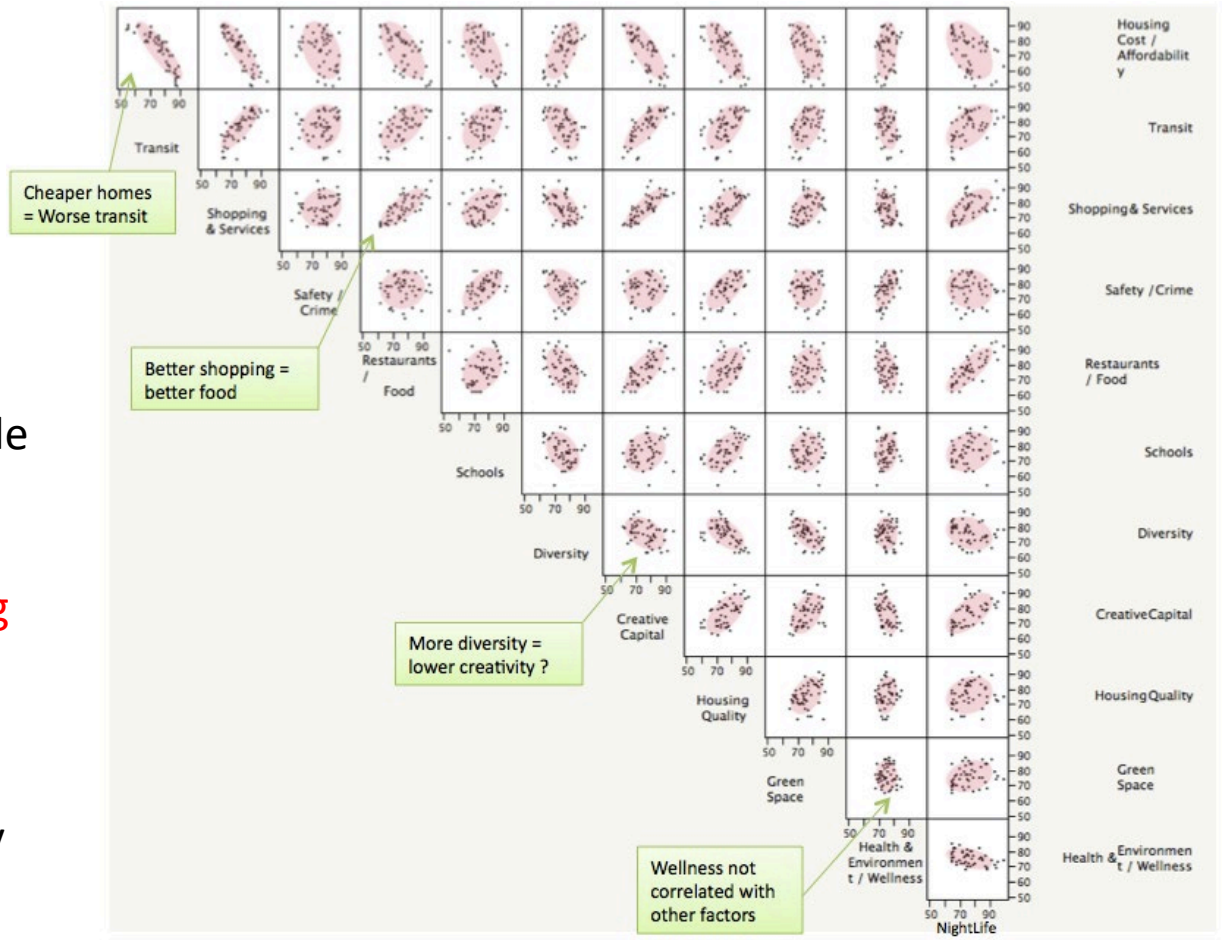
Where to live in NYC?

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

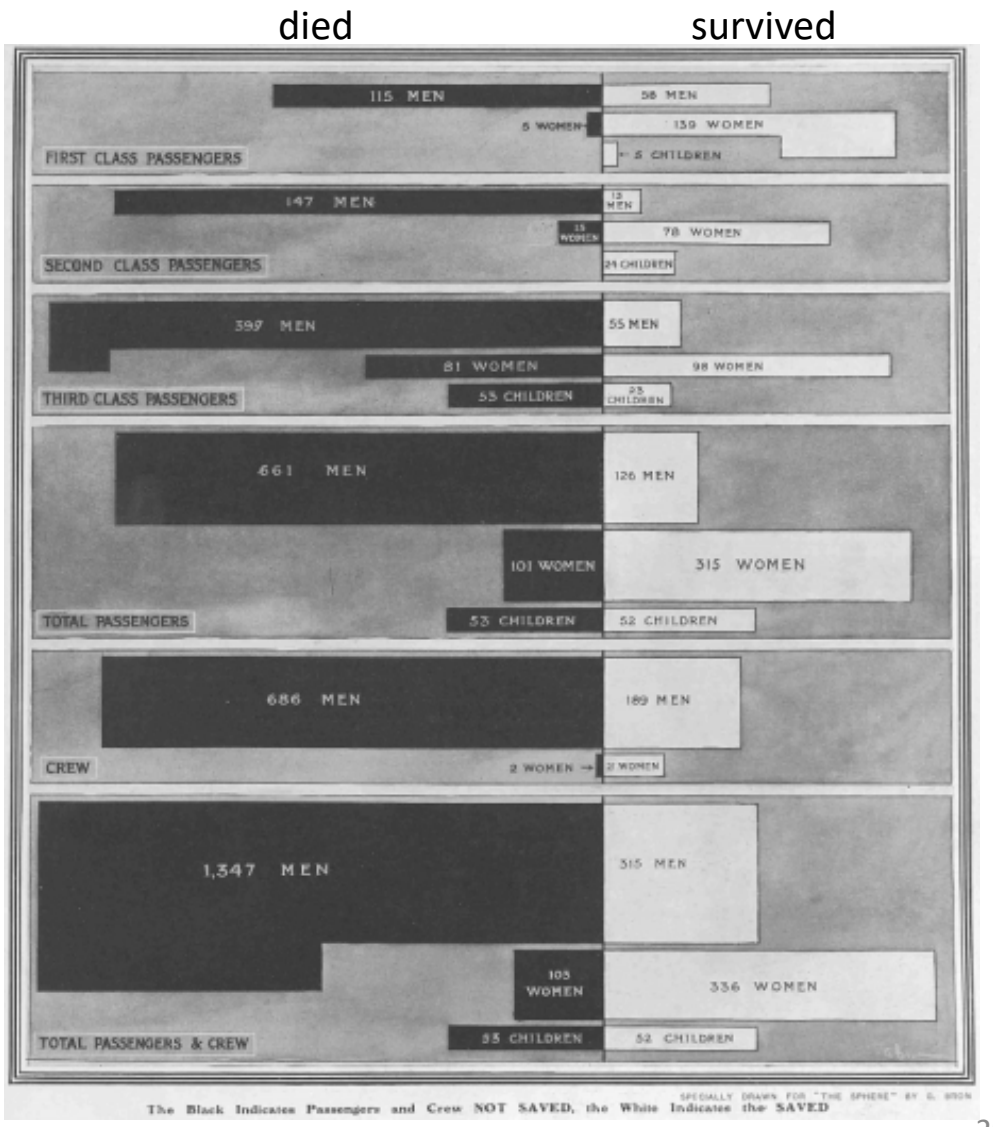
Categorical data

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/



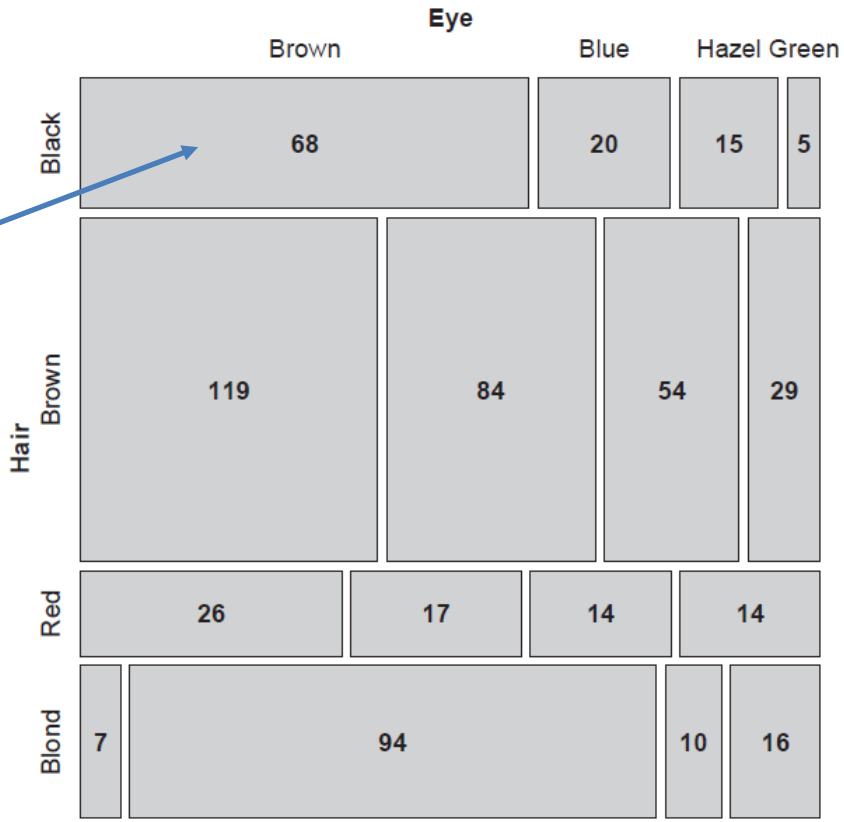
Categorical data: Mosaic plots

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

```
> data(HairEyeColor)
> HEC <- margin.table(HairEyeColor, 1:2)
> HEC
       Eye
Hair     Brown Blue Hazel Green
  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
```

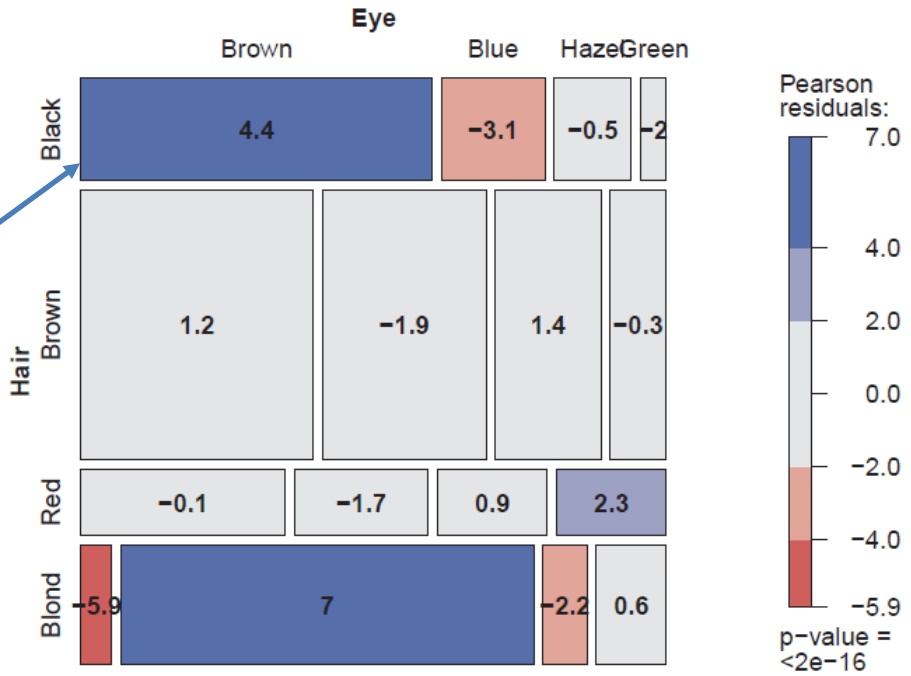


How to understand the association between hair color and eye color?

Mosaic plots

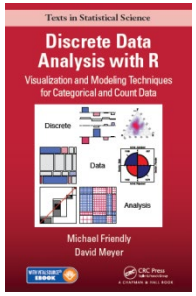
Shade each tile in relation to the contribution to the Pearson χ^2 statistic

$$\chi^2 = \sum r_{ij}^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$



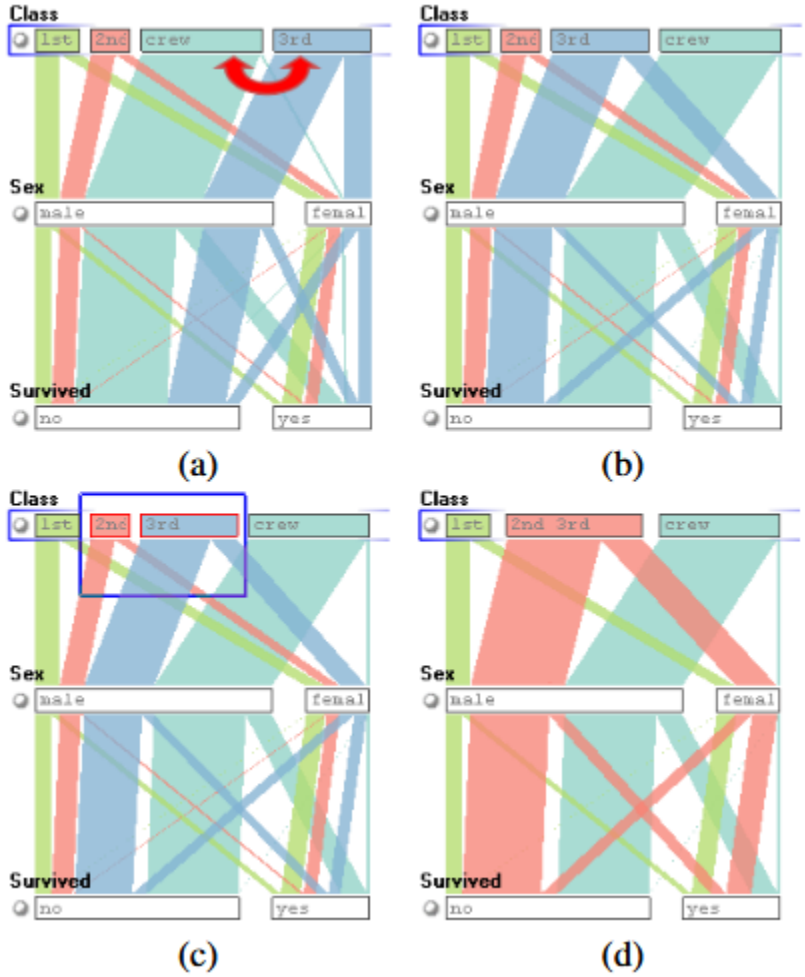

```
> round(residuals(chisq.test(HEC)),2)
       Eye
Hair     Brown  Blue Hazel Green
  Black   4.40 -3.07 -0.48 -1.95
  Brown   1.23 -1.95  1.35 -0.35
  Red    -0.07 -1.73  0.85  2.28
  Blond  -5.85  7.05 -2.23  0.61
```

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/



Parallel Sets

Titanic data: Who survived?



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)

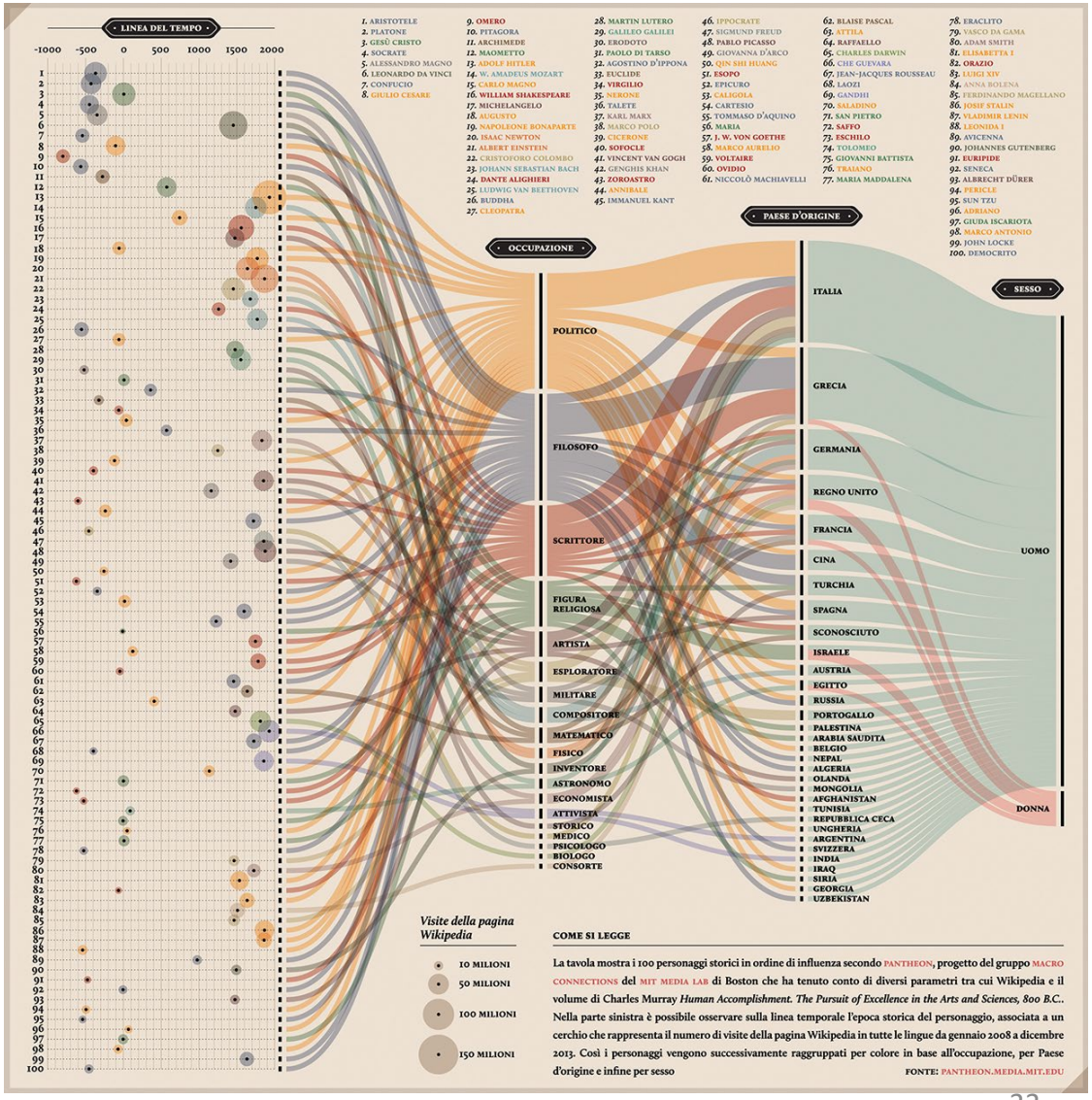
From: Kosera et al. (2006), https://kosara.net/papers/2006/Kosara_TVCG_2006.pdf

Sankey diagram

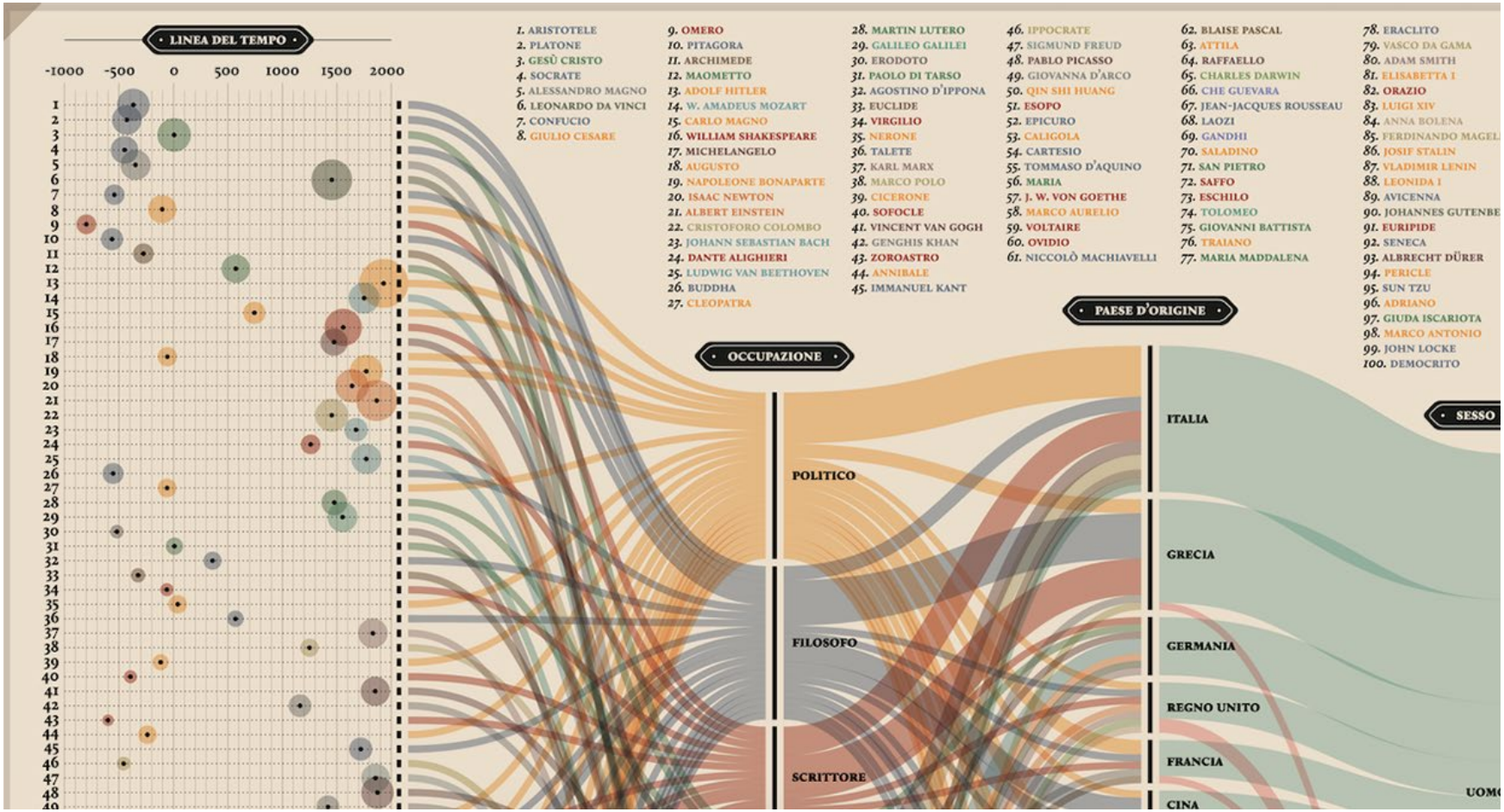
Pantheon, by Valerio Pellegrini
Visualizing the 100 most influential figures in History (Wikipedia visits)

Columns show occupation, country of origin and gender

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



Sankey diagram



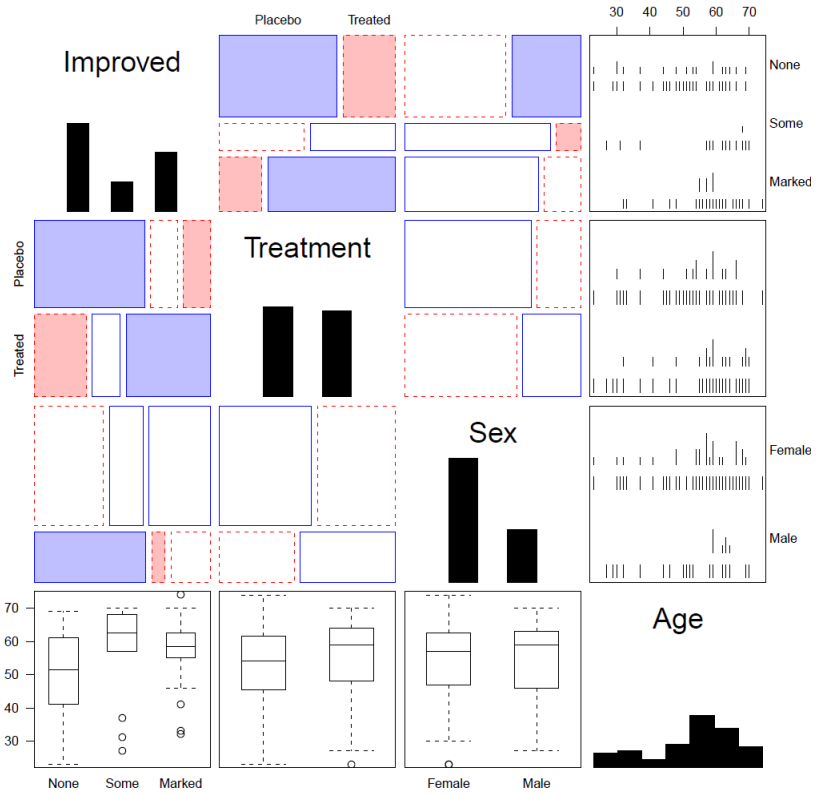
Multiple dimensions of the most influential people in history

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

Generalized pairs plots

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

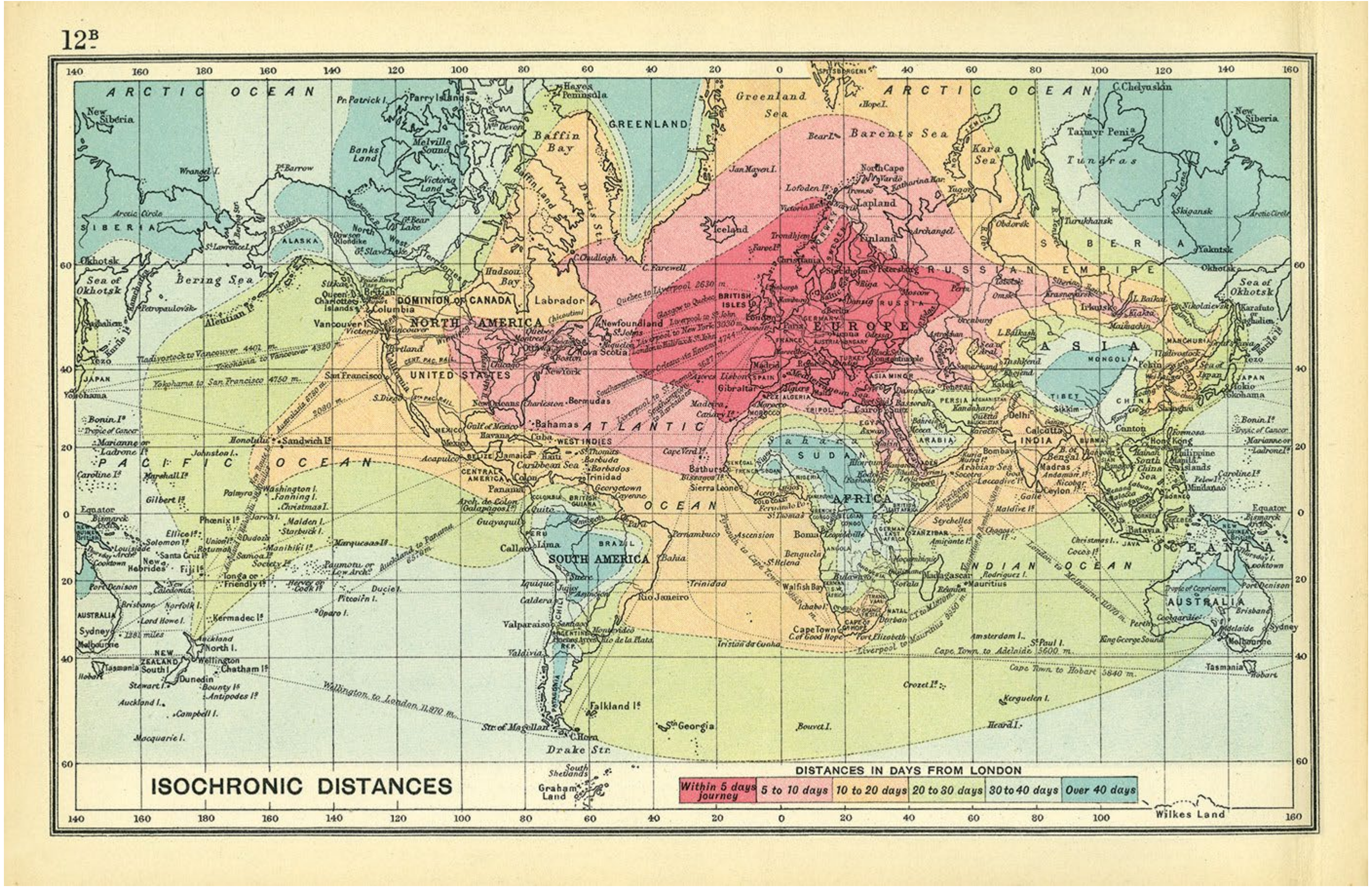
x	y	plot
Q	Q	scatterplot
C	Q	boxplot
Q	C	barcode
C	C	mosaic




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

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



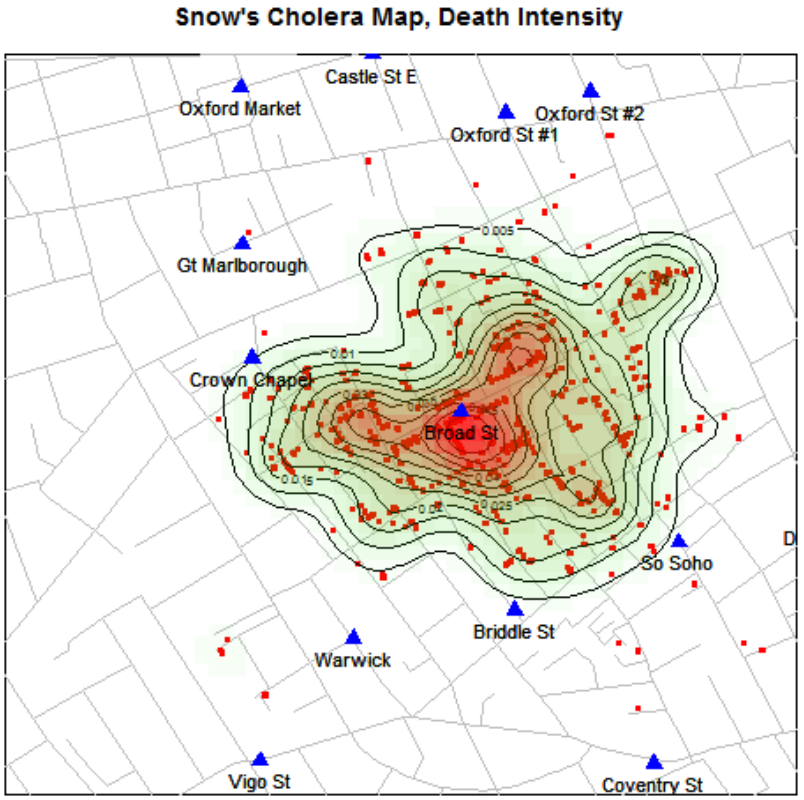
Francis Galton, *Isochronic chart of travel time*, 1881

3D: Bivariate density estimation

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

Broad St. pump

Modern statistical techniques can compute contours of constant density



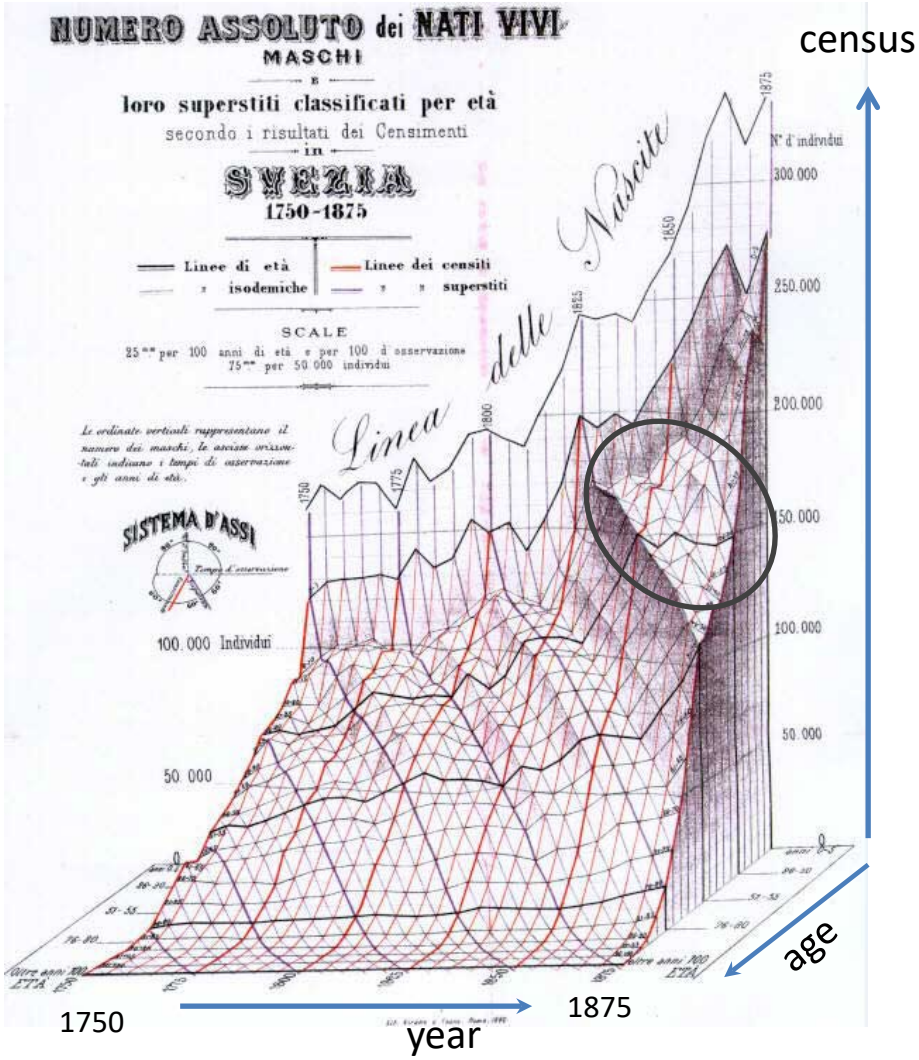
Data: HistData package for R

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 blue diagonal lines down & to the right
These 3 variables are primary in demography.

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

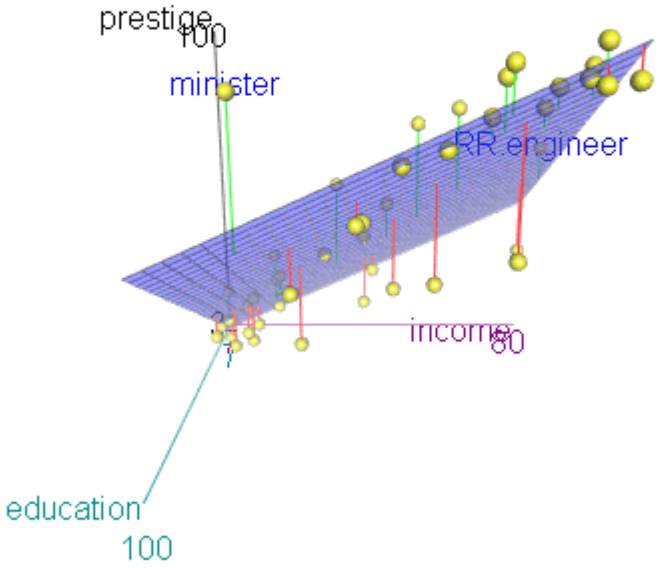


3D: scatterplot & regression surface

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

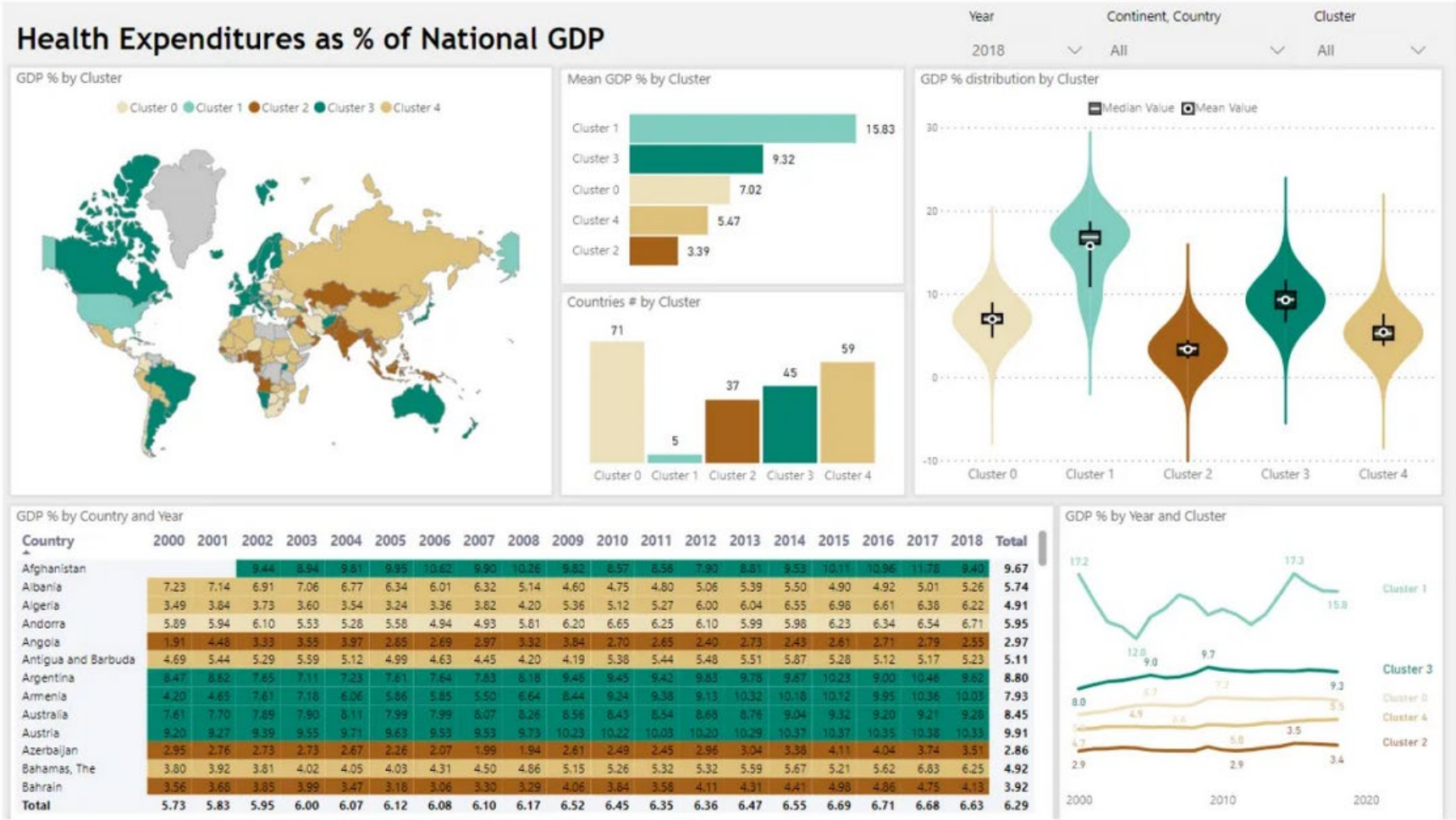


This plot is produced in R, using the car and rgl packages

```
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")
```

Dashboards

Dashboards provide multiple images of different aspects, often linked interactively

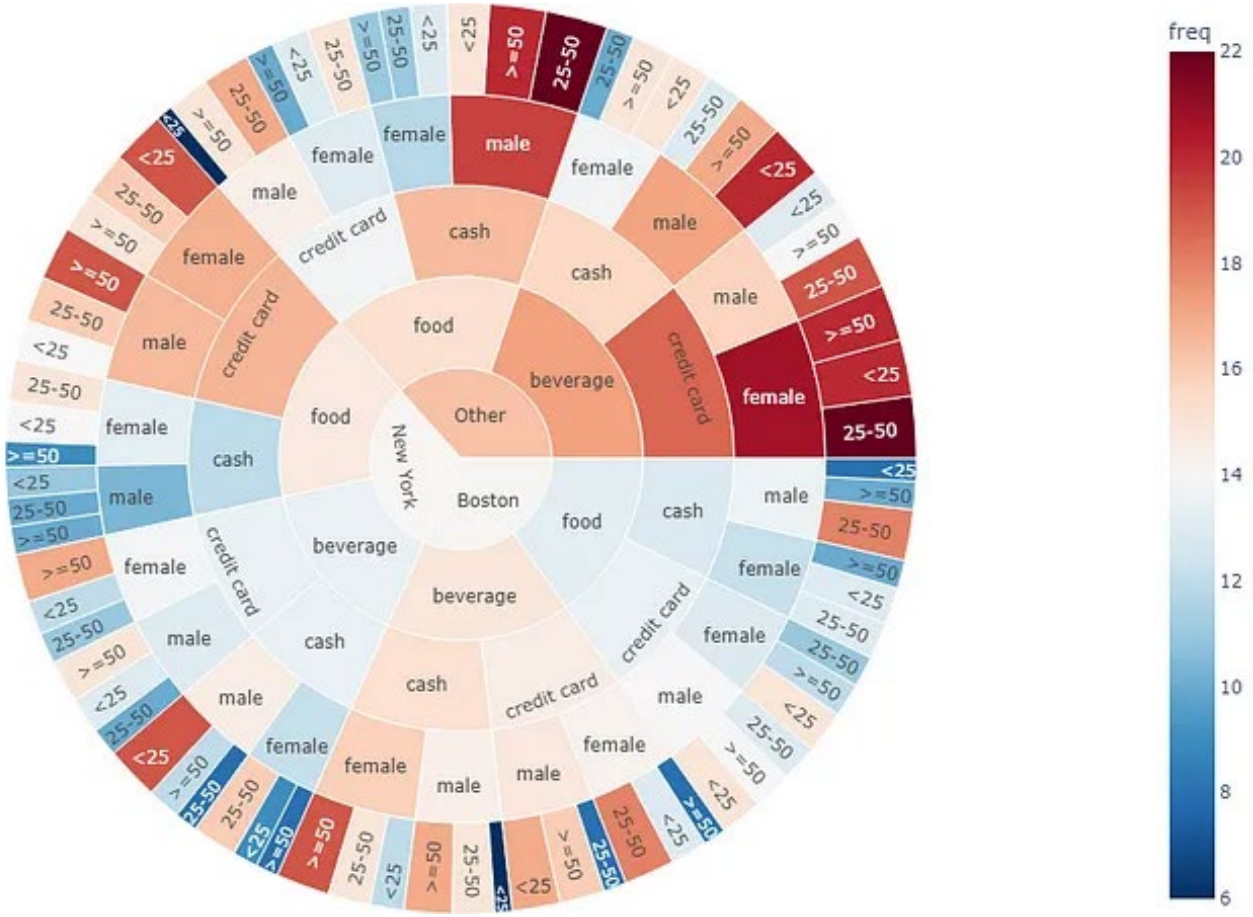


Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
Afghanistan			9.44	8.94	9.61	9.95	10.62	9.90	10.26	9.82	8.57	8.56	7.90	8.81	9.53	10.11	10.96	11.78	9.40	9.67
Albania	7.23	7.14	6.91	7.06	6.77	6.34	6.01	6.32	5.14	4.60	4.75	4.80	5.06	5.39	5.50	4.90	4.92	5.01	5.26	5.74
Algeria	3.49	3.84	3.73	3.60	3.54	3.24	3.36	3.82	4.20	5.36	5.12	5.27	6.00	6.04	6.55	6.98	6.61	6.38	6.22	4.91
Andorra	5.89	5.94	6.10	5.53	5.28	5.58	4.94	4.93	5.81	6.20	6.65	6.25	6.10	5.99	5.98	6.23	6.34	6.54	6.71	5.95
Angola	1.91	4.48	3.33	3.55	3.97	2.65	2.69	2.97	3.32	3.84	2.70	2.65	2.40	2.73	2.43	2.61	2.71	2.79	2.55	2.97
Antigua and Barbuda	4.69	5.44	5.29	5.59	5.12	4.99	4.63	4.45	4.20	4.19	5.38	5.44	5.48	5.51	5.87	5.28	5.12	5.17	5.23	5.11
Argentina	8.47	8.62	7.65	7.11	7.23	7.61	7.64	7.83	8.18	9.45	9.45	9.42	9.83	9.78	9.87	10.23	9.00	10.46	9.62	8.80
Armenia	4.20	4.65	7.61	7.18	6.06	5.86	5.85	5.50	6.64	8.44	9.24	9.38	9.13	10.32	10.18	10.12	9.95	10.36	10.03	7.93
Australia	7.61	7.70	7.89	7.90	8.11	7.99	7.99	8.07	8.26	8.56	8.43	8.54	8.68	8.76	9.04	9.32	9.20	9.21	9.28	8.45
Austria	9.20	9.27	9.39	9.55	9.71	9.63	9.53	9.53	9.73	10.23	10.22	10.03	10.20	10.29	10.37	10.37	10.35	10.38	10.33	9.91
Azerbaijan	2.95	2.76	2.73	2.73	2.67	2.26	2.07	1.99	1.94	2.61	2.49	2.45	2.96	3.04	3.38	4.11	4.04	3.74	3.51	2.86
Bahamas, The	3.80	3.92	3.81	4.02	4.05	4.03	4.31	4.50	4.86	5.15	5.26	5.32	5.32	5.59	5.67	5.21	5.62	6.83	6.25	4.92
Bahrain	3.56	3.68	3.85	3.99	3.47	3.18	3.06	3.30	3.29	4.06	3.84	3.58	4.11	4.31	4.41	4.98	4.86	4.75	4.13	3.92
Total	**5.73**	**5.83**	**5.95**	**6.00**	**6.07**	**6.12**	**6.08**	**6.10**	**6.17**	**6.52**	**6.45**	**6.35**	**6.36**	**6.47**	**6.55**	**6.69**	**6.71**	**6.68**	**6.63**	**6.29**

Exercise

What is this SUNBURST chart trying to show?

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



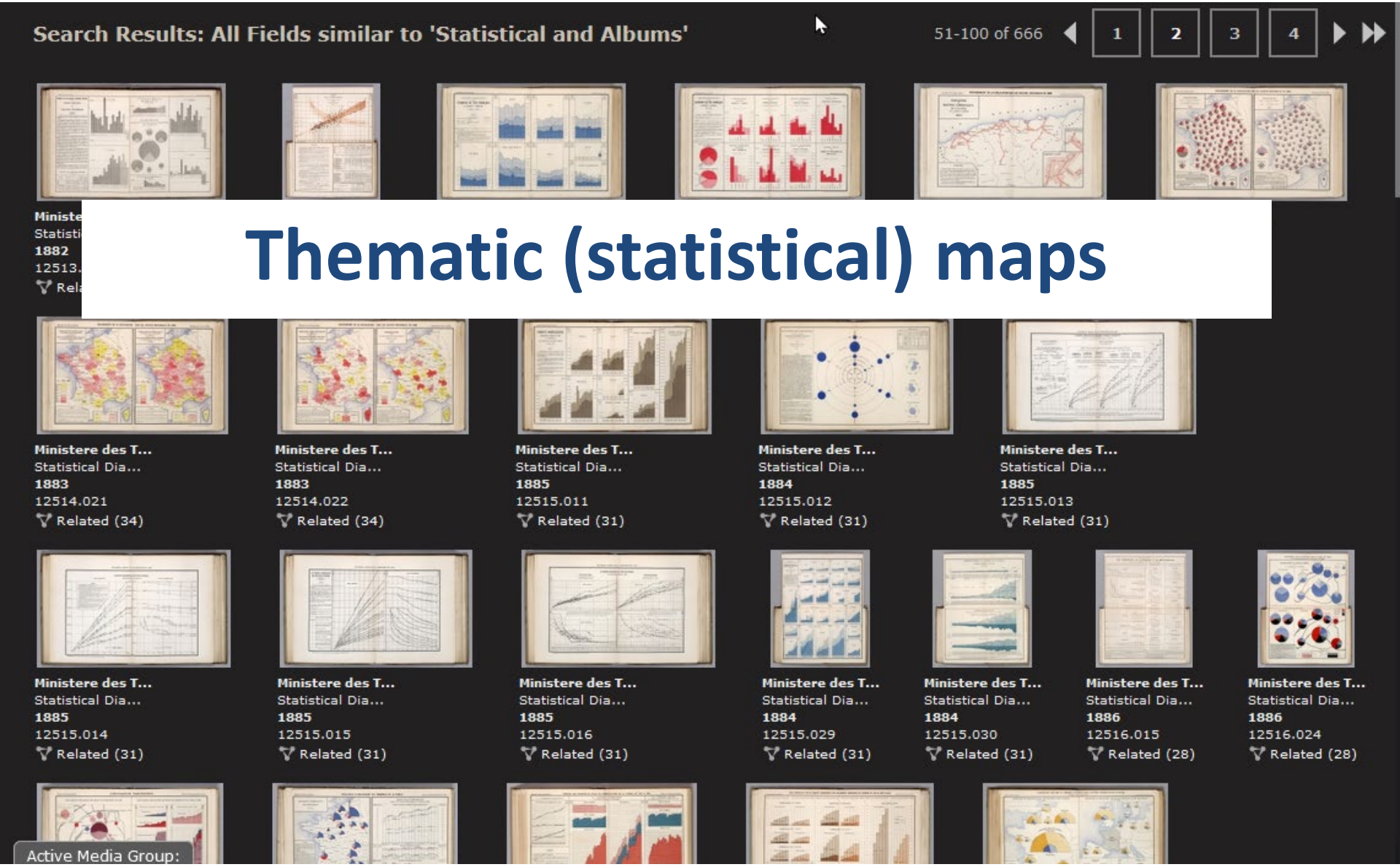
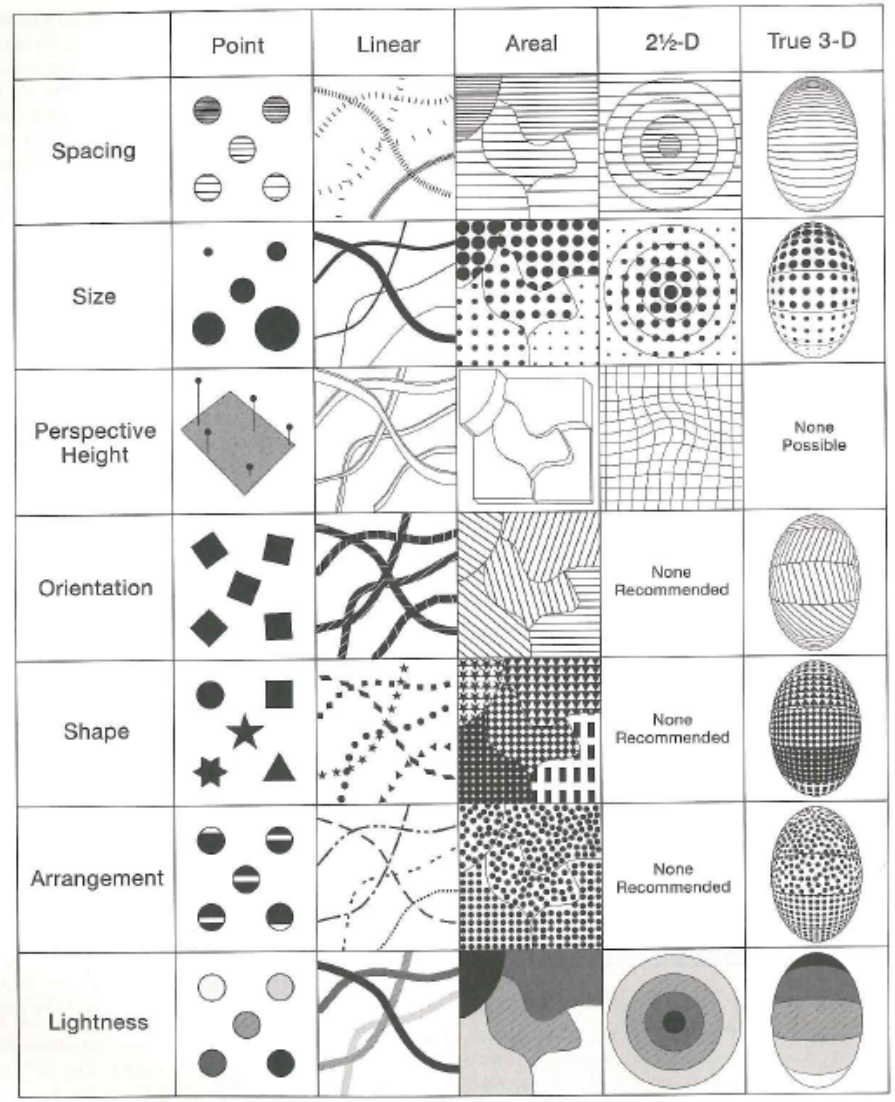


Image from: David Rumsey Map Collection, https://www.davidrumsey.com/

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



------ # dimensions -------

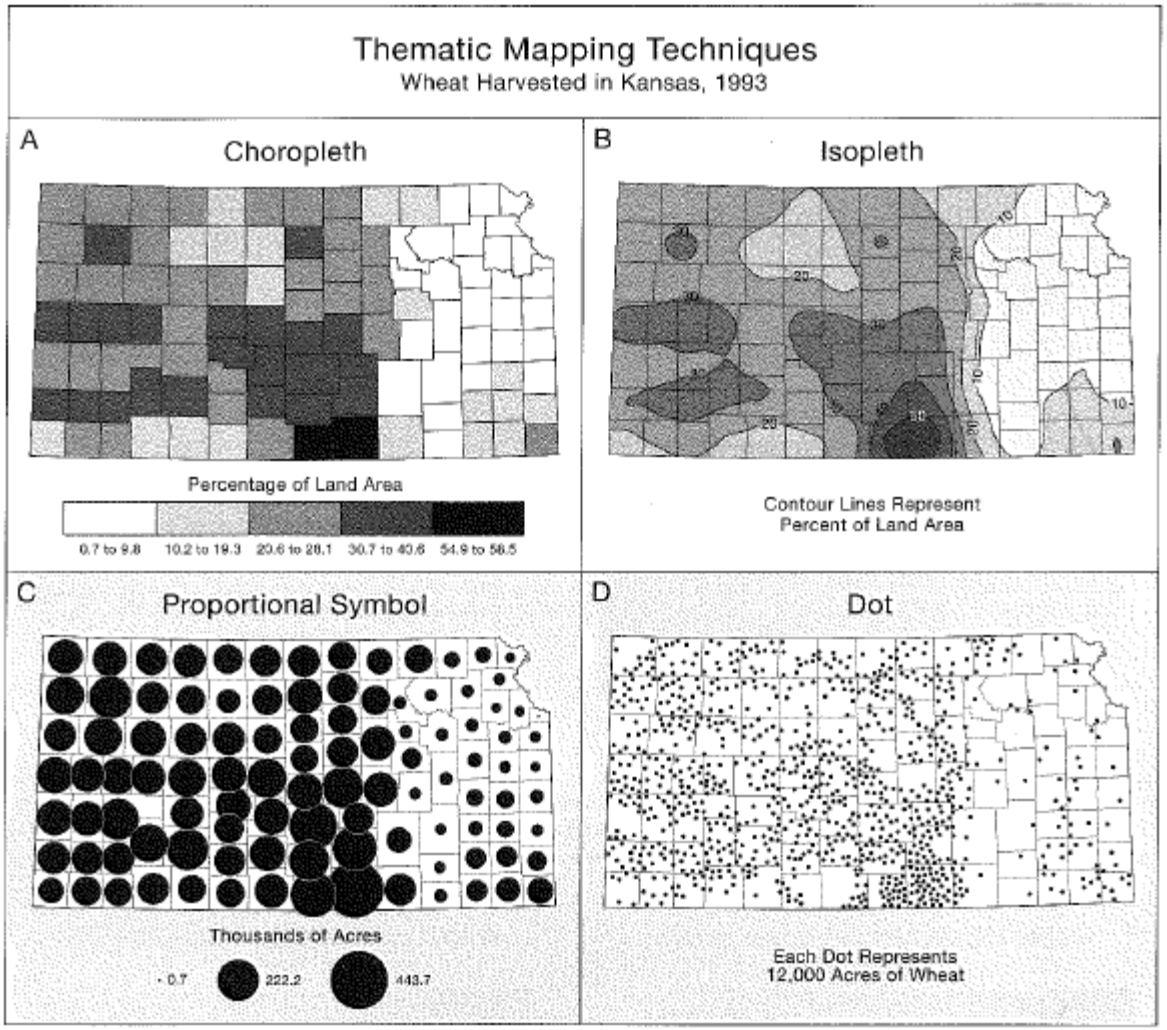
From: Slocum et al., *Thematic cartography and geographical visualization*, Fig 4.3

Thematic maps: Types

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



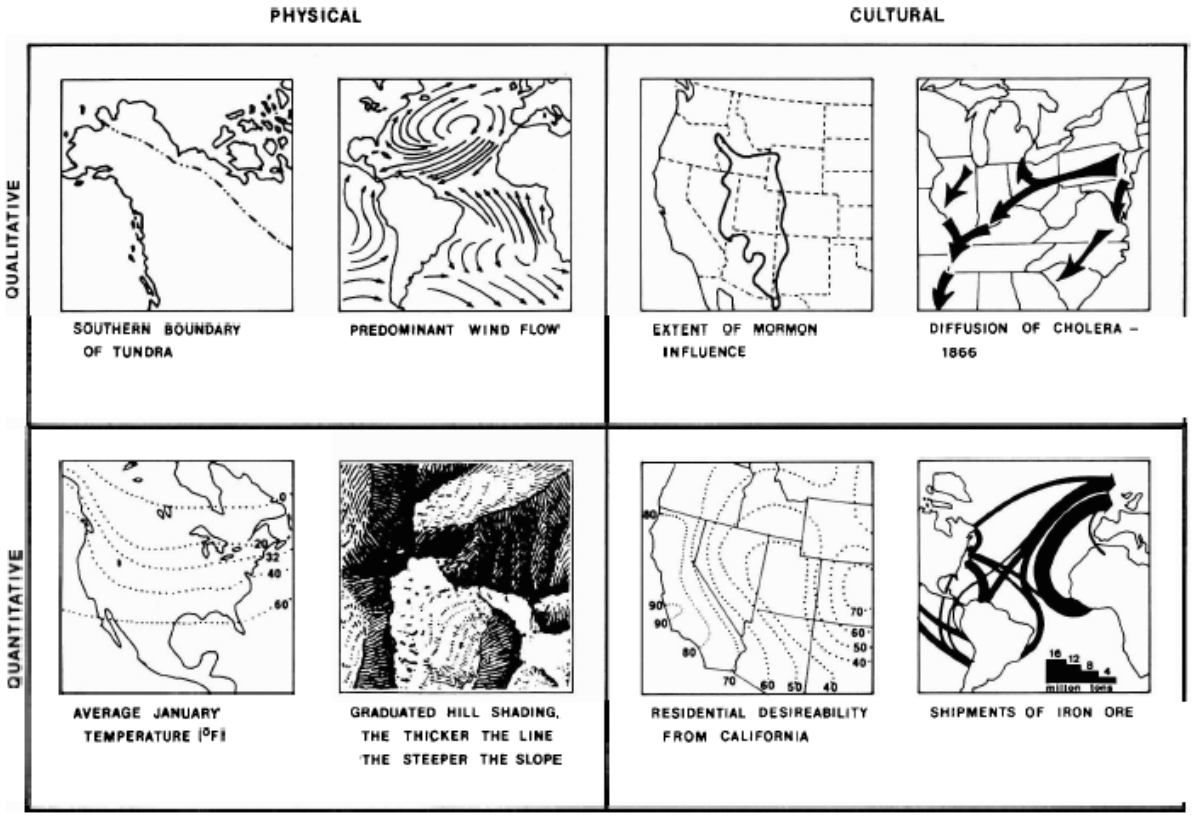
From: Slocum et al., *Thematic cartography and geographical visualization*, Fig 4.9

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.



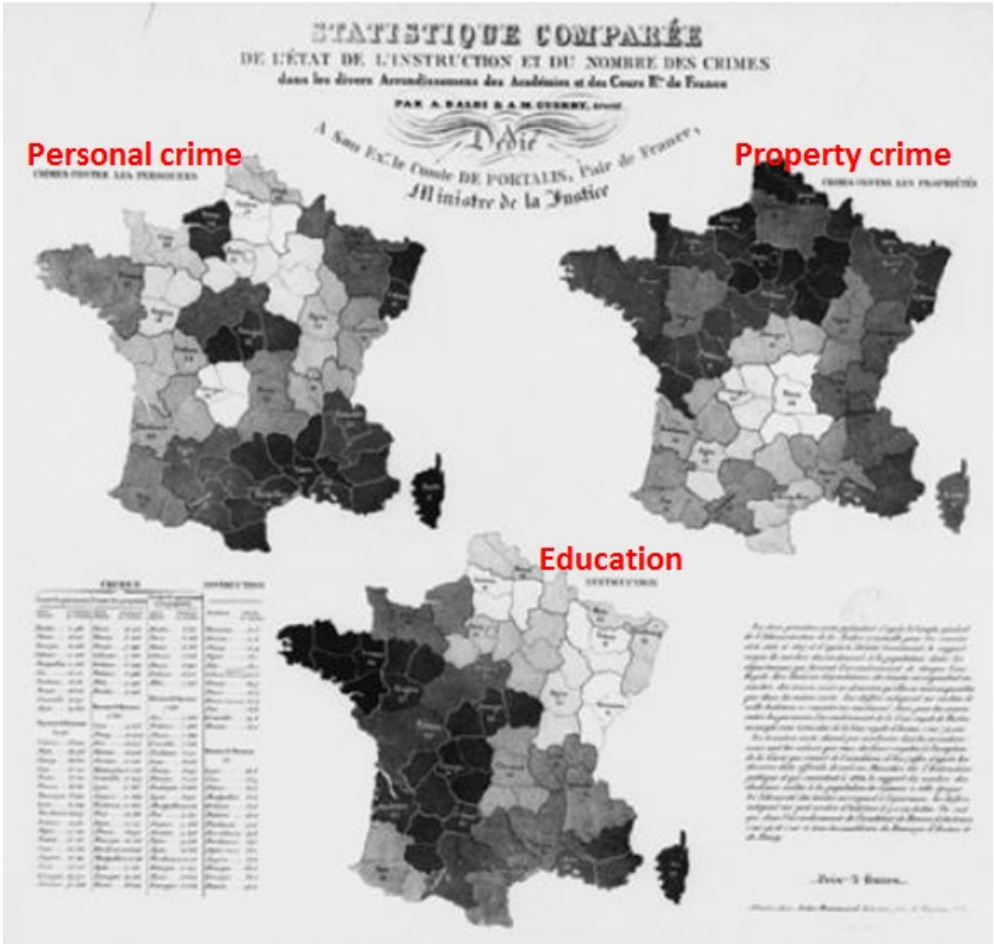
MacEachern, A. (1979). The Evolution Of Thematic Cartography / A Research Methodology and Historical Review, *The Canadian Cartographer* 16(1) June 1979, p. 17-33

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

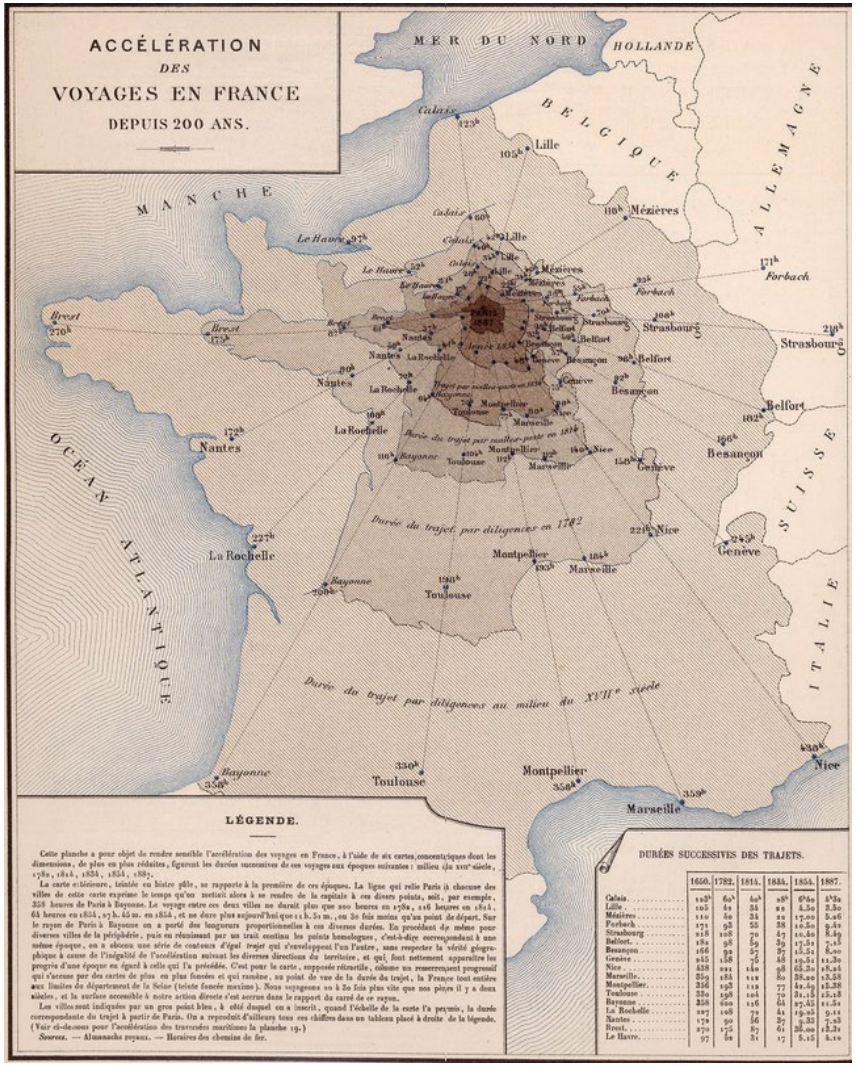


See: Friendly, M. (2022). The life and works of André-Michel Guerry, revisited. *Sociological Spectrum*, 42, 233–259

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



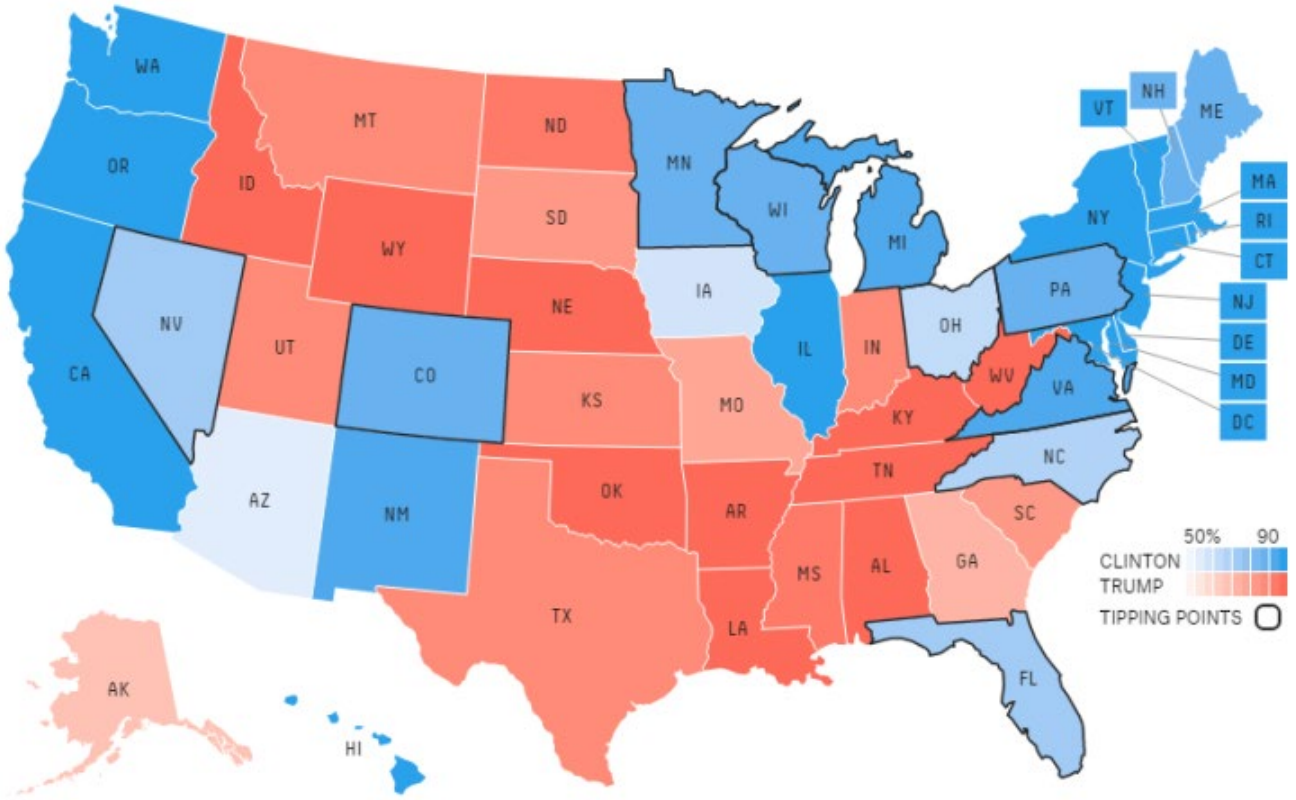
Album de Statistique Graphique, 1888, plate 8

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

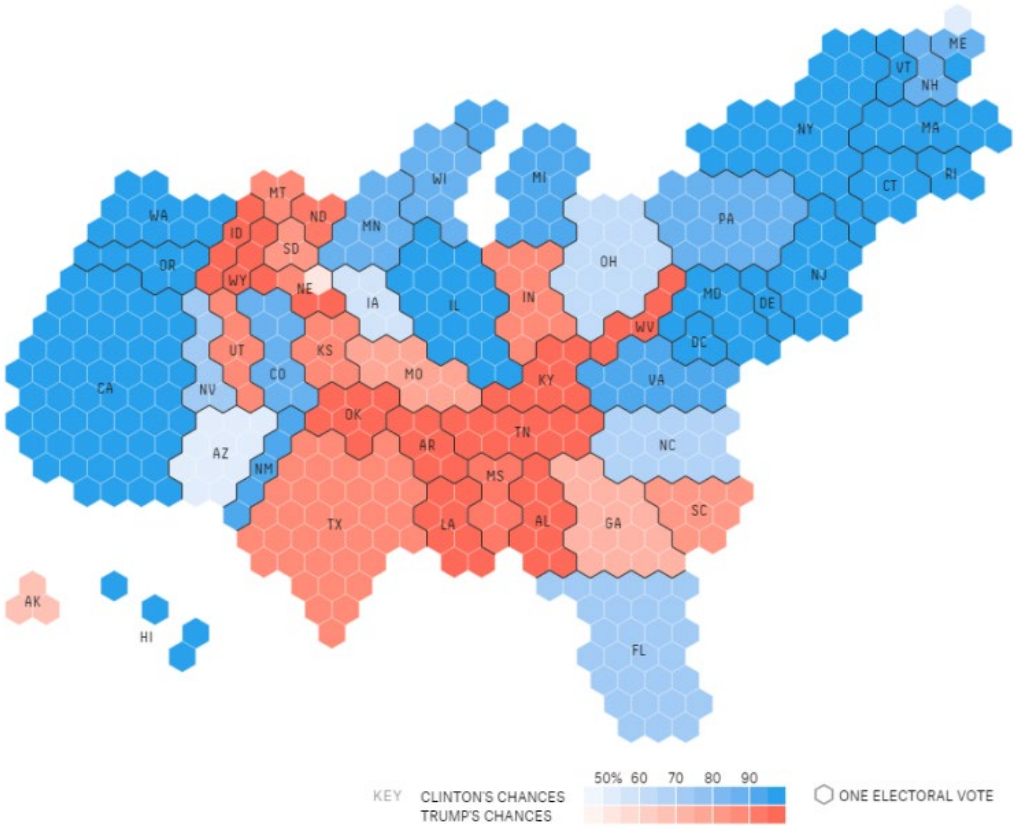
Cartogram (tilegrams)

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!

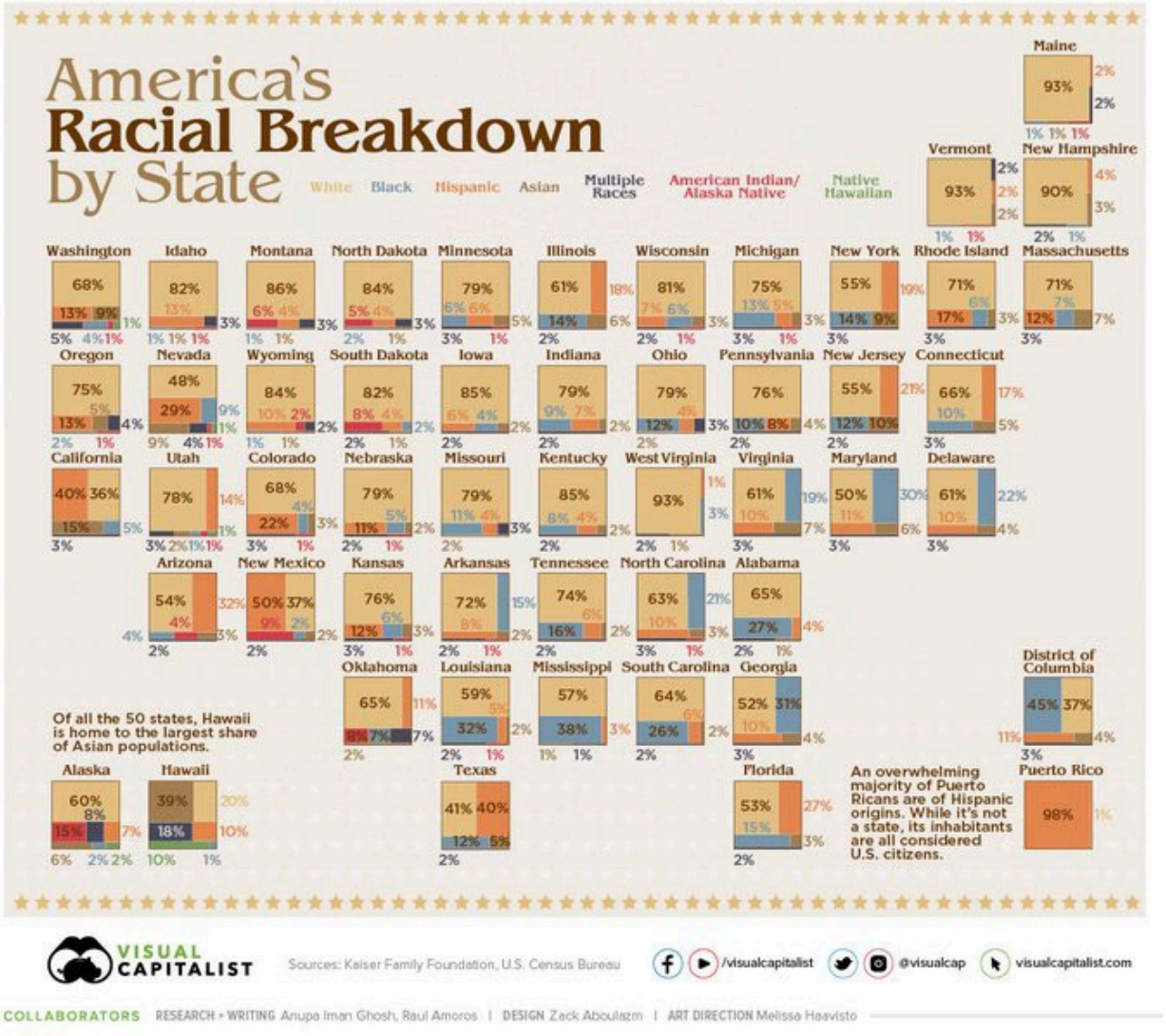


fivethirtyeight.com election predictions, Oct. 13, 2017

Mosaic cartograms

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

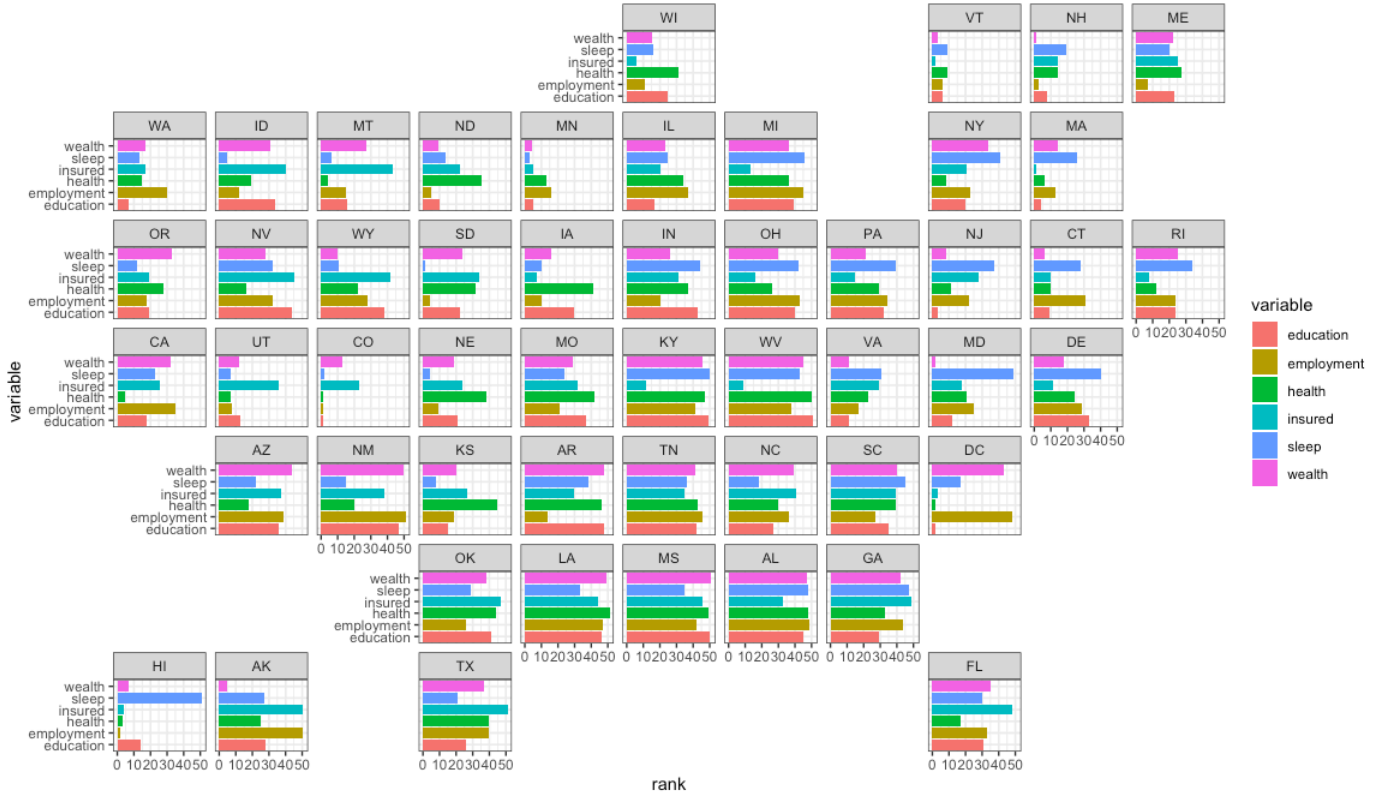
Each tile is positioned as in a schematic US map



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

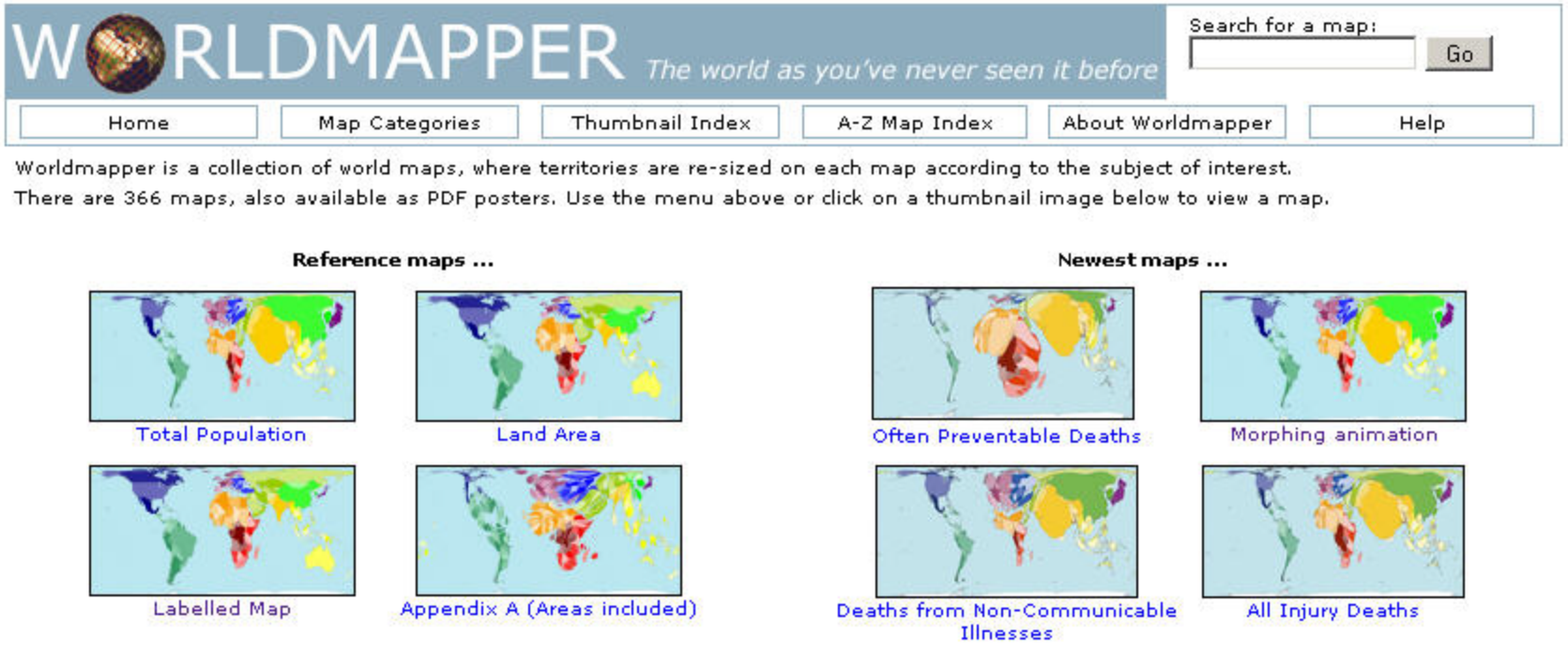


From: https://cran.r-project.org/web/packages/geofacet/vignettes/geofacet.html

Worldmapper: The world in cartograms

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

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

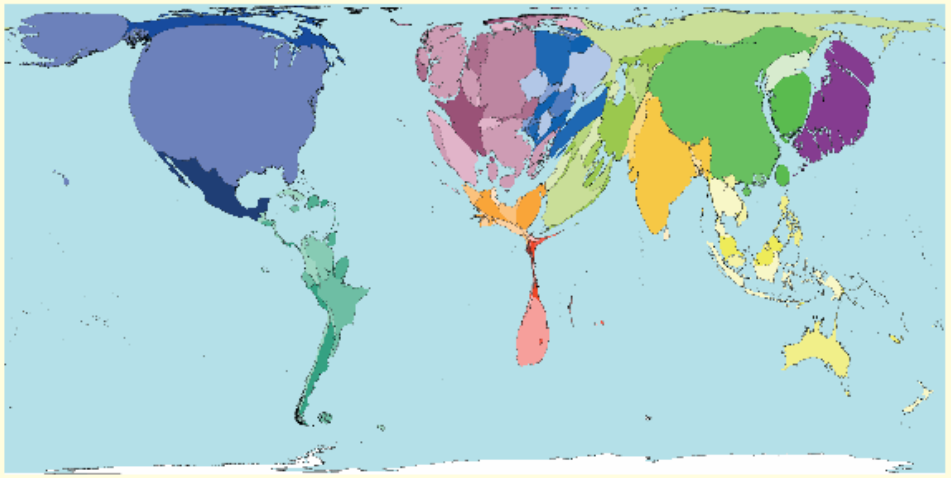


Worldmapper: Carbon emissions

These pages are well-designed according to data vis. Ideas: high impact graph + interpretive details & explanation

Carbon Emissions 2000

Produced by the SASI group (Sheffield) and Mark Newman (Michigan)



Carbon dioxide causes roughly 60% of the 'enhanced greenhouse effect' or global warming resulting from certain gases emitted by human activities. In 2000 there were almost 23 billion tonnes of carbon dioxide emitted worldwide. Of this, 28% came from North American territories; 0.09% came from Central African territories.

Emissions of carbon dioxide vary hugely between places, due to differences in lifestyle and ways of producing energy. Whilst people living in 66 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 territories that year.

Territory size shows the proportion of carbon dioxide emissions in 2000 that were directly from there.

Land area

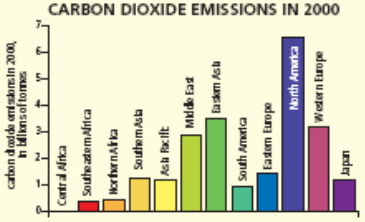
Technical notes
- Data are from the United Nations Development Programme's 2004 Human Development Report.
- *The denominator used is population in 2002.
- See website for further information.

MOST AND LEAST CARBON DIOXIDE EMISSIONS IN 2000

Rank	Territory	Value	Rank	Territory	Value
1	Qatar	64	191	Democratic Republic Congo	0.095
2	Bahrain	27	192	United Republic of Tanzania	0.094
3	Brunei Darussalam	21	193	Malawi	0.094
4	Kuwait	21	194	Uganda	0.094
5	Trinidad & Tobago	20	195	Comoros	0.094
6	Luxembourg	19	196	Niger	0.094
7	United States	19	197	Burundi	0.048
8	United Arab Emirates	18	198	Cambodia	0.048
9	Australia	18	199	Chad	0.047
10	Saudi Arabia	17	200	Afghanistan	0.040

tonnes of carbon dioxide emitted in 2000 per person living in that territory*

CARBON DIOXIDE EMISSIONS IN 2000



"If the world does not learn now to show respect to nature, what kind of future will the new generations have?"

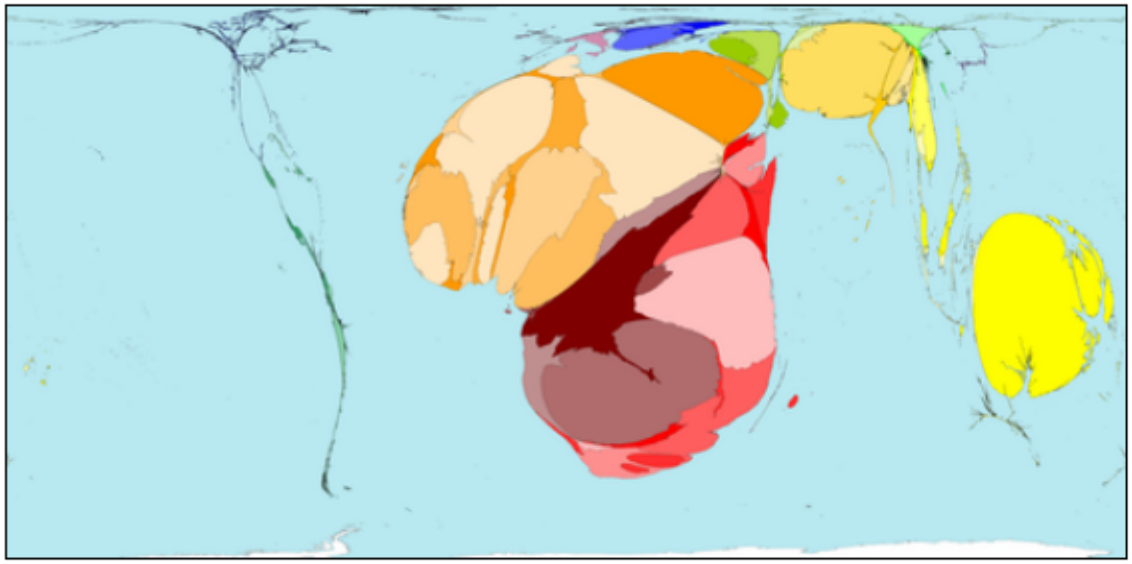
Rigoberta Menchú Tum, 1992

Map 295

Worldmapper: Cholera deaths

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

< Previous Map | **Cholera Deaths** | Map No. 232 | Open PDF poster | Next Map >



Cholera deaths result from severe dehydration caused by diarrhoea. This is treatable: in 2004 the number of cholera deaths was only 2.5% of the number of cholera cases that year. Distributions of cholera cases and deaths differ due to differing availability of treatments.

In 1962, in Papua New Guinea, 36% of cholera cases, which was 464 people, died. In 2004, in the Central African Republic, 15% of cholera cases, which was 48 people, died.

In contrast, there were 73 territories where nobody died from cholera, because of good sanitation, clean water and available treatment. These territories have no area on this map.

'The cholera outbreak has continued ... water provided by the tankers is not enough and they try to boost their supply from the wells, which are not covered. The rain washes faeces and other pollutants into the wells ...' Pierre Kahozi, 2004

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

Bivariate maps

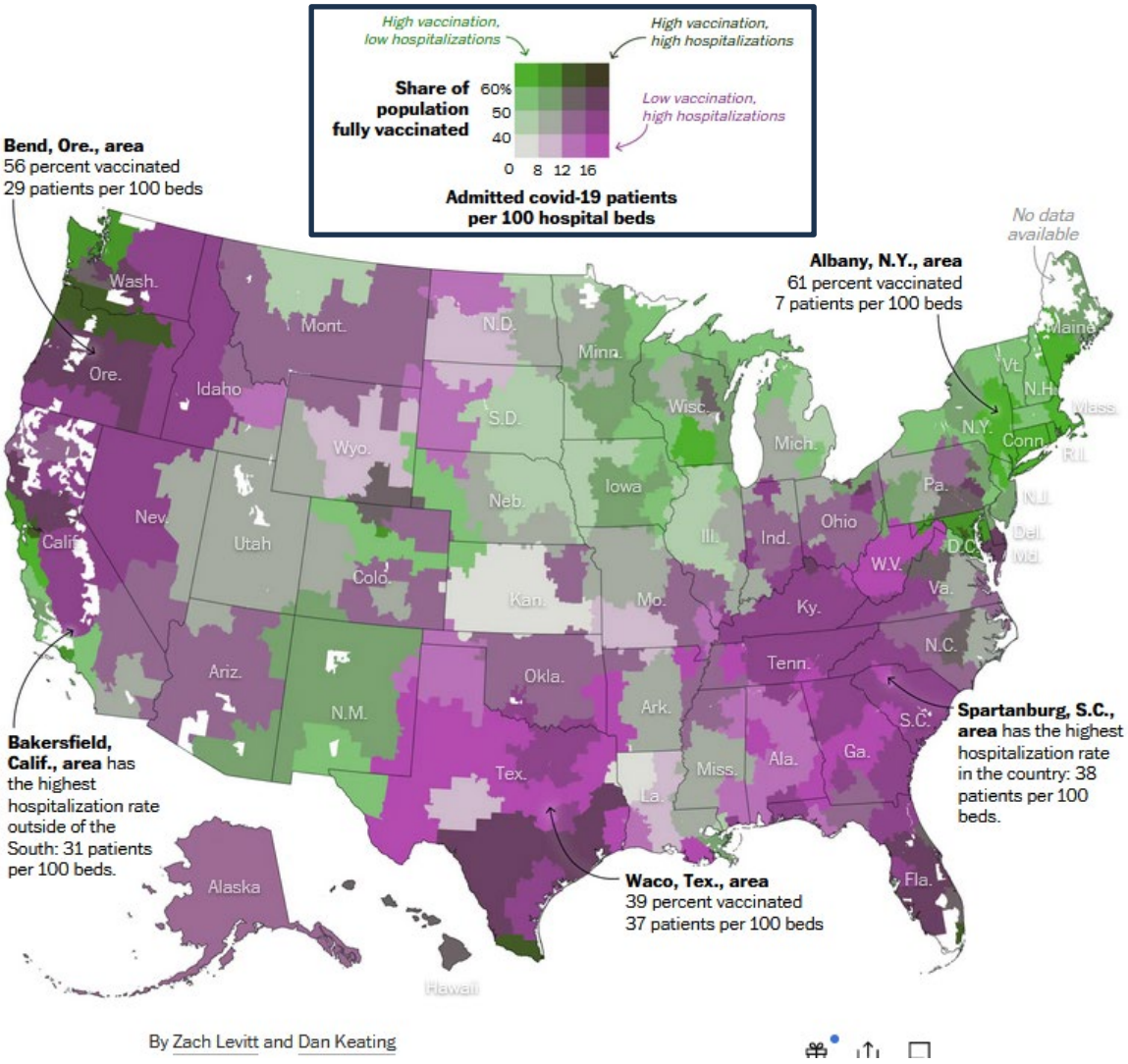
Vaccination rates and COVID hospitalizations

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?



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

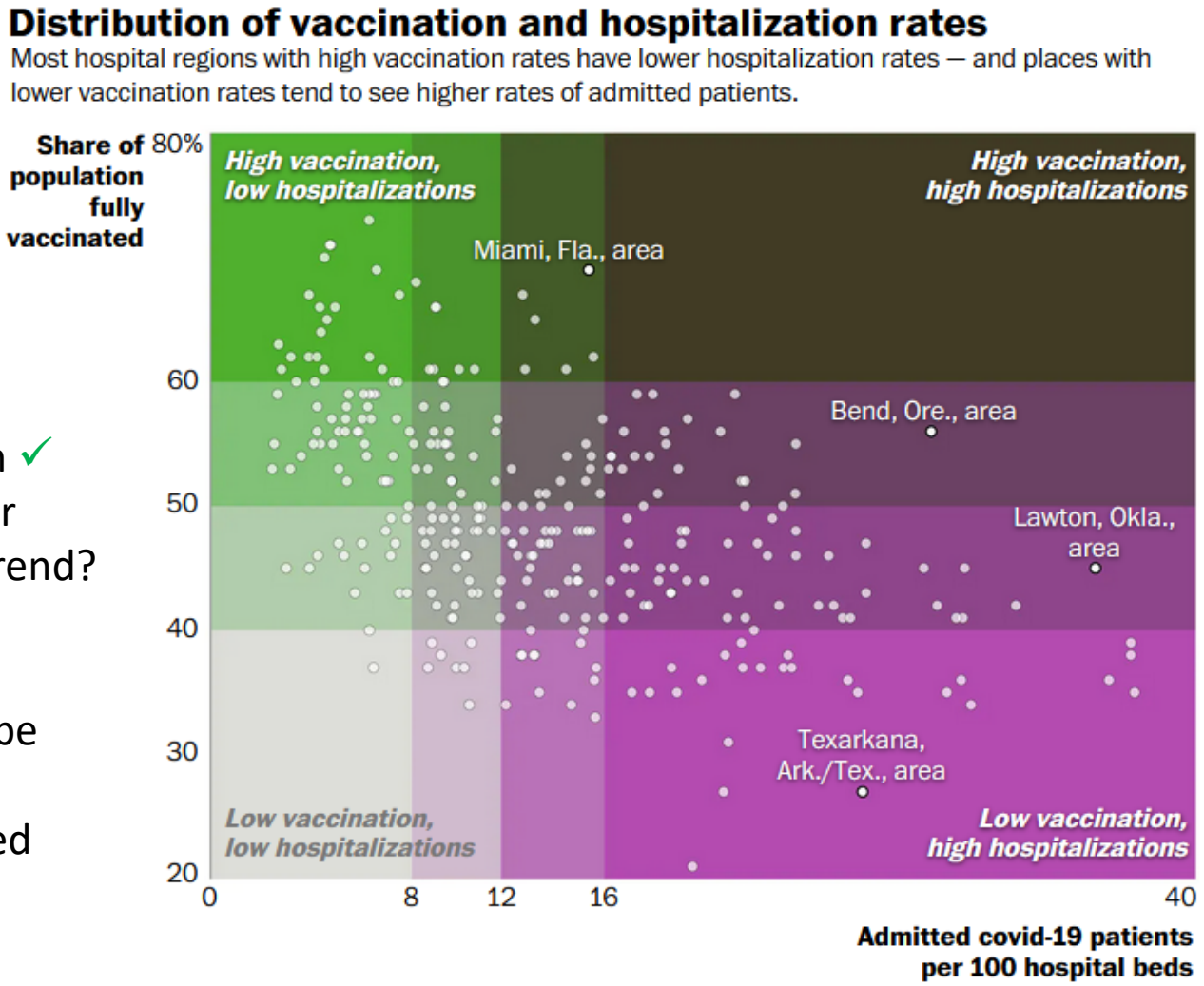
Scatterplot to the rescue

This lovely scatterplot shows:

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

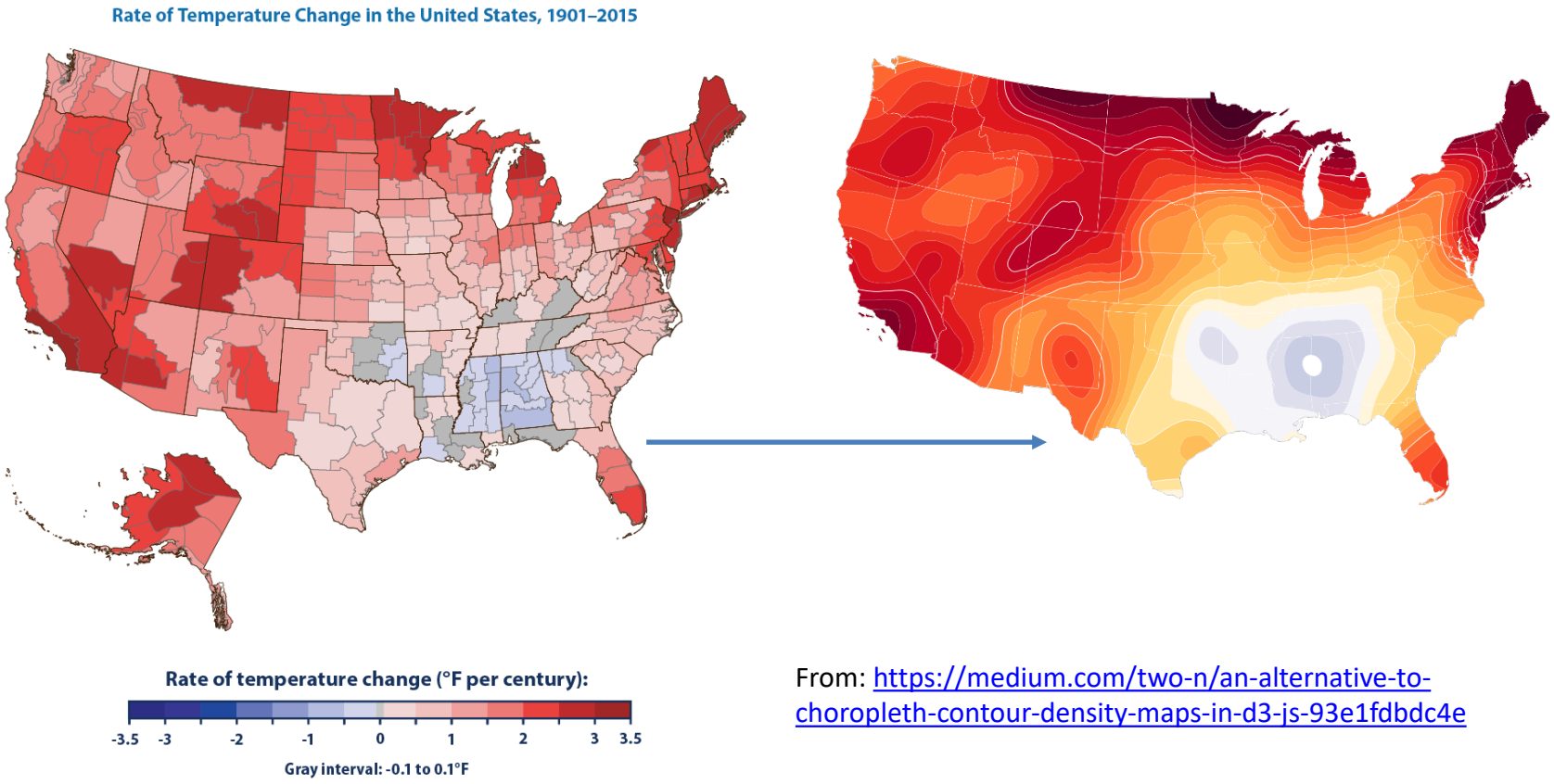
But:

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



Contour maps

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



From: https://medium.com/two-n/an-alternative-to-choropleth-contour-density-maps-in-d3-js-93e1fdbdc4e

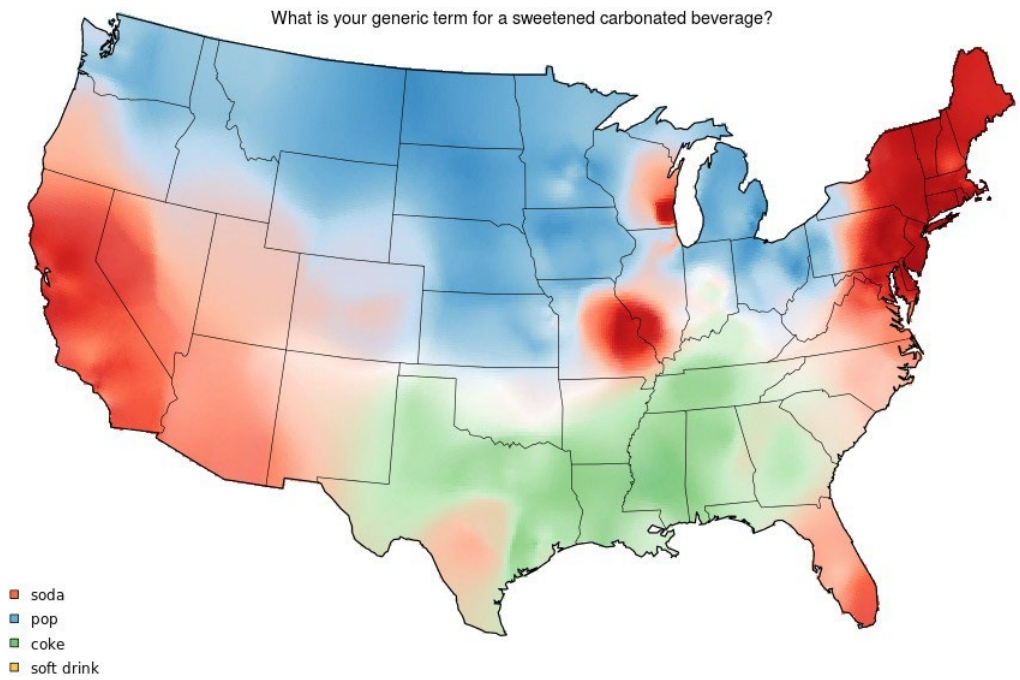
Spatial visualization: Analysis + maps

Linguistics: Food dialect maps– visualizing how people speak

soda vs. pop?

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: Analysis + maps

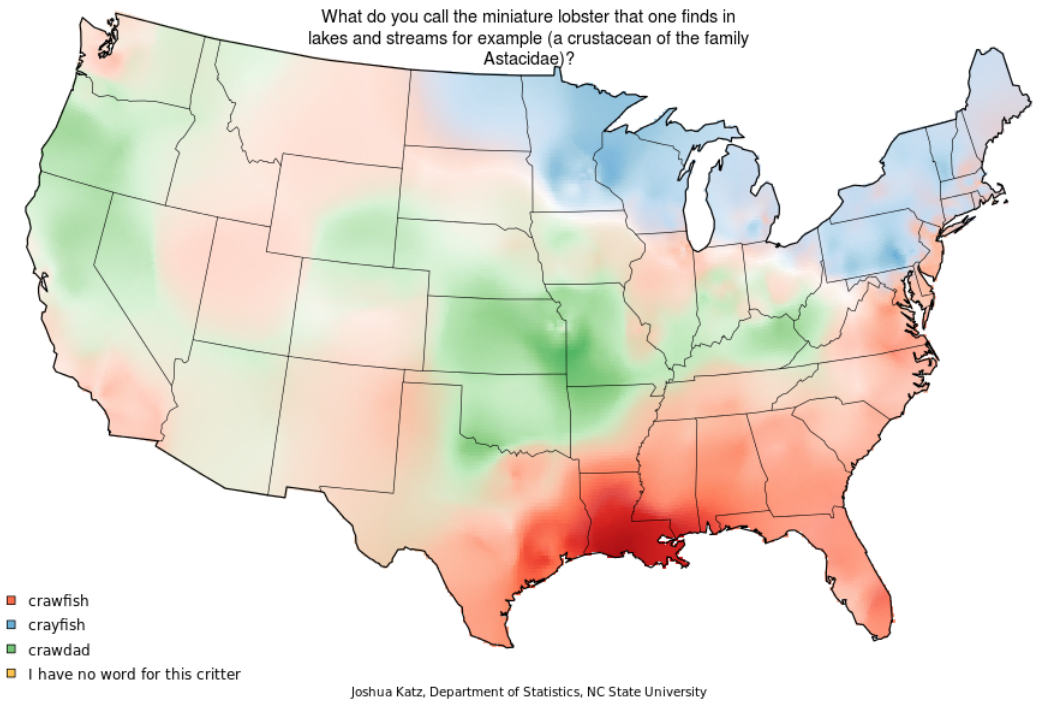
Linguistics: Food dialect maps– visualizing how people speak

crawfish, crawfish, crawdad?

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

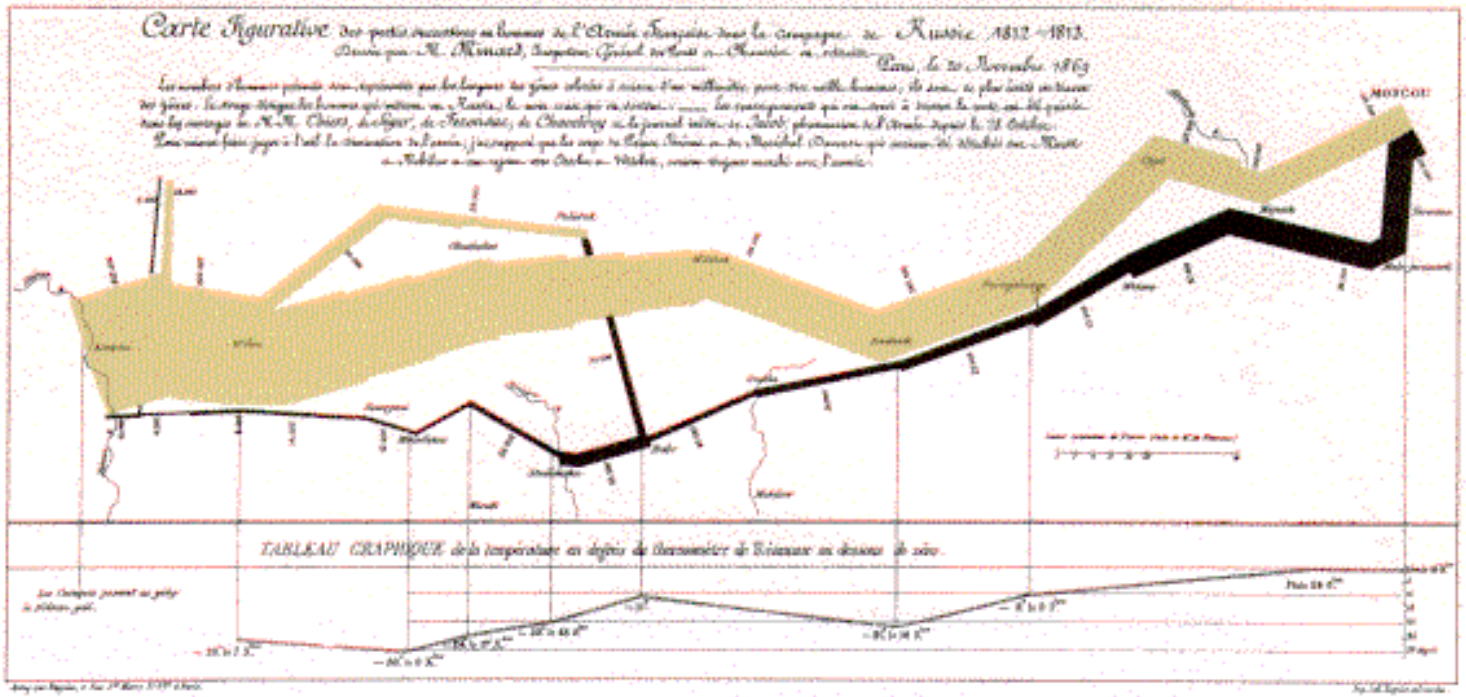
Regional differences are quite apparent.

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



Flow maps

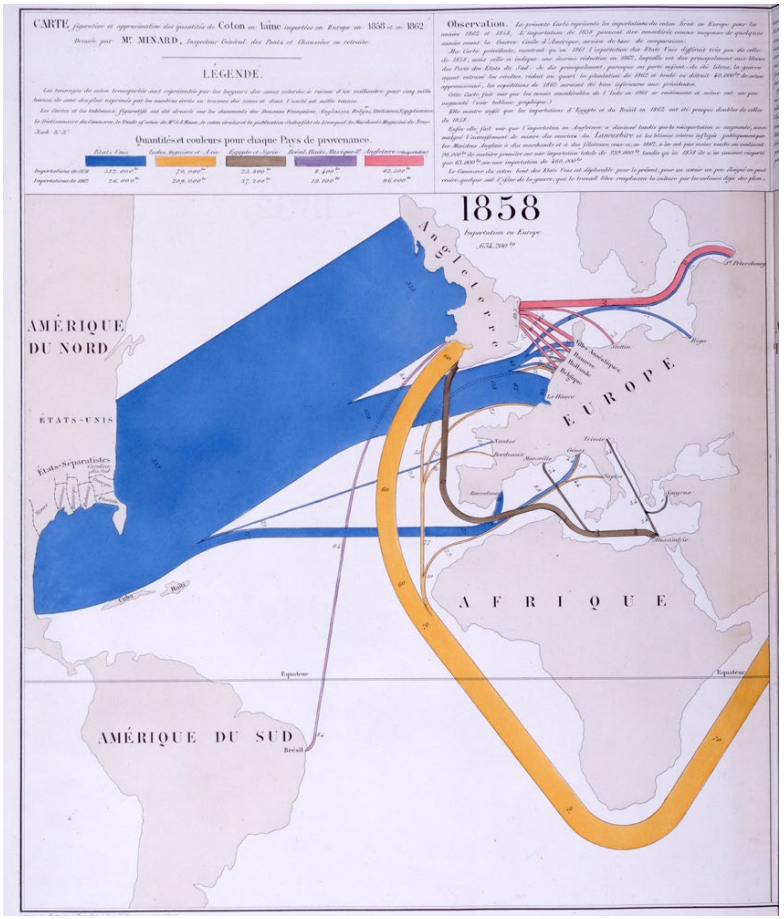
Flow maps show movement or change in a geographic framework
The master work is this image by Charles-Joseph Minard (1869)



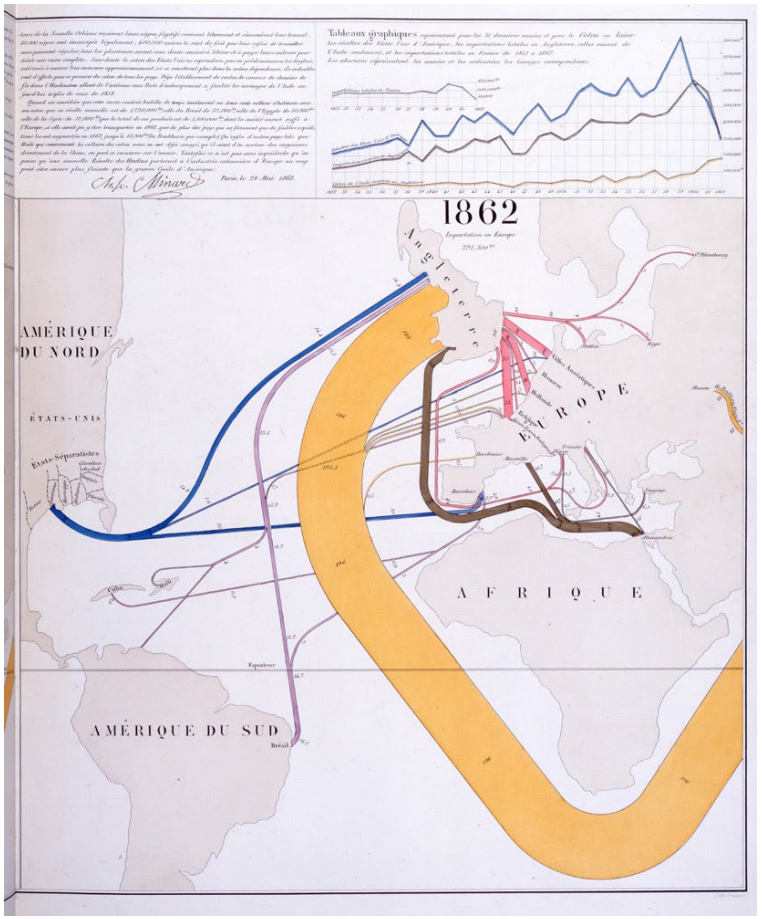
- Marey (1878): *"defies the pen of the historian in its brutal eloquence"*
- Tufte (1983): *"the best statistical graphic ever produced"*

Effect of US civil war on cotton trade

Before



After



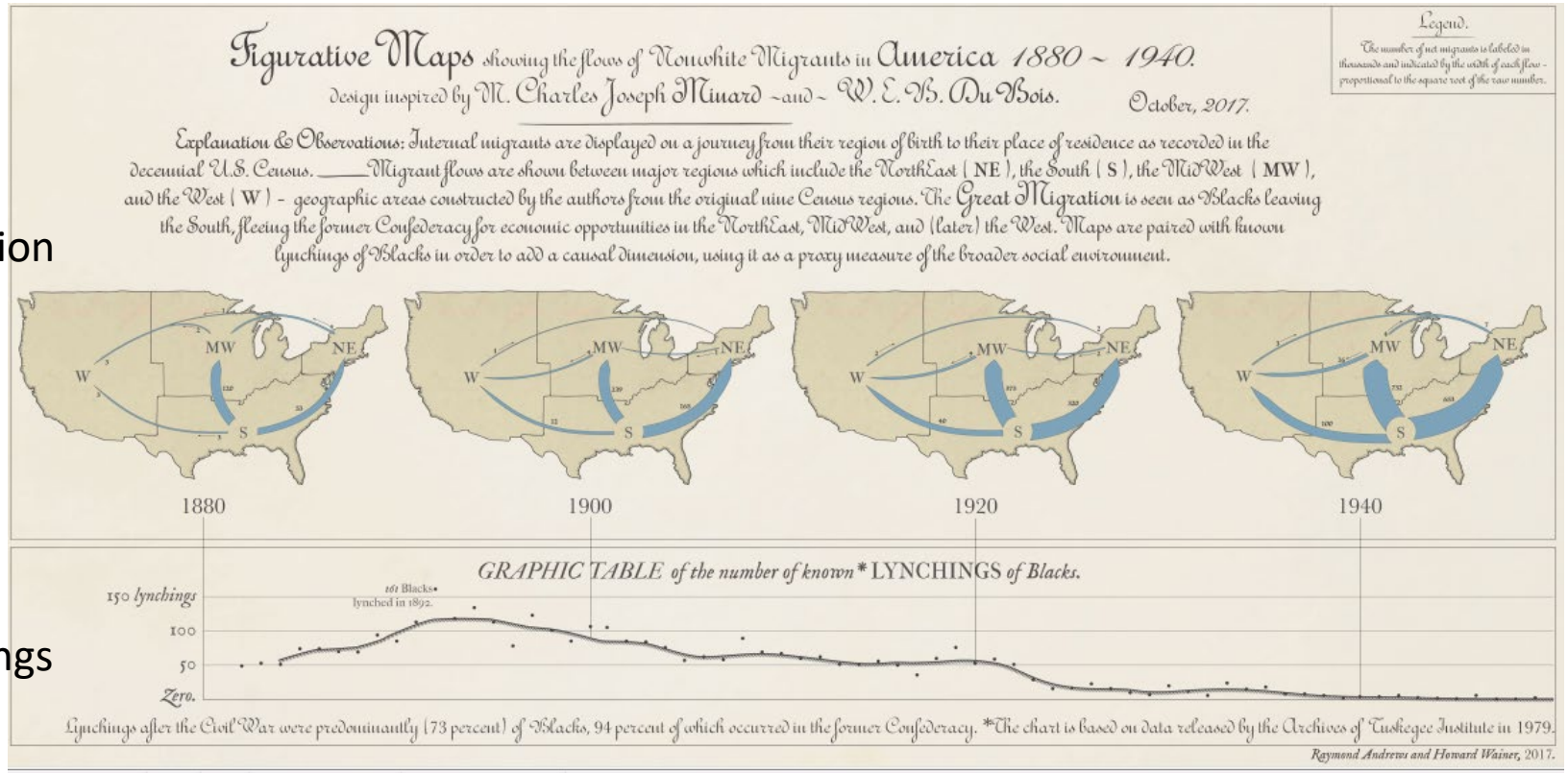
Note the deformation of the map to accommodate the data

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.

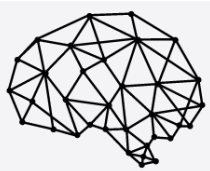
Migration

Lynchings



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

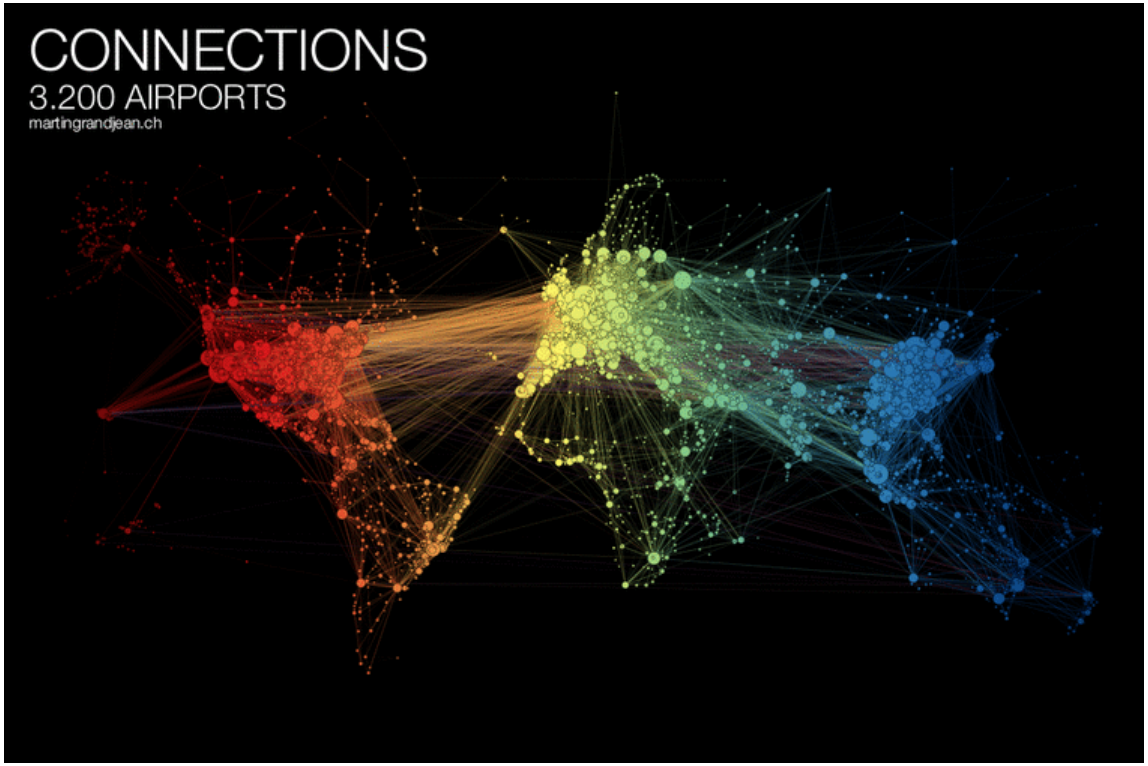
Network visualization



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/

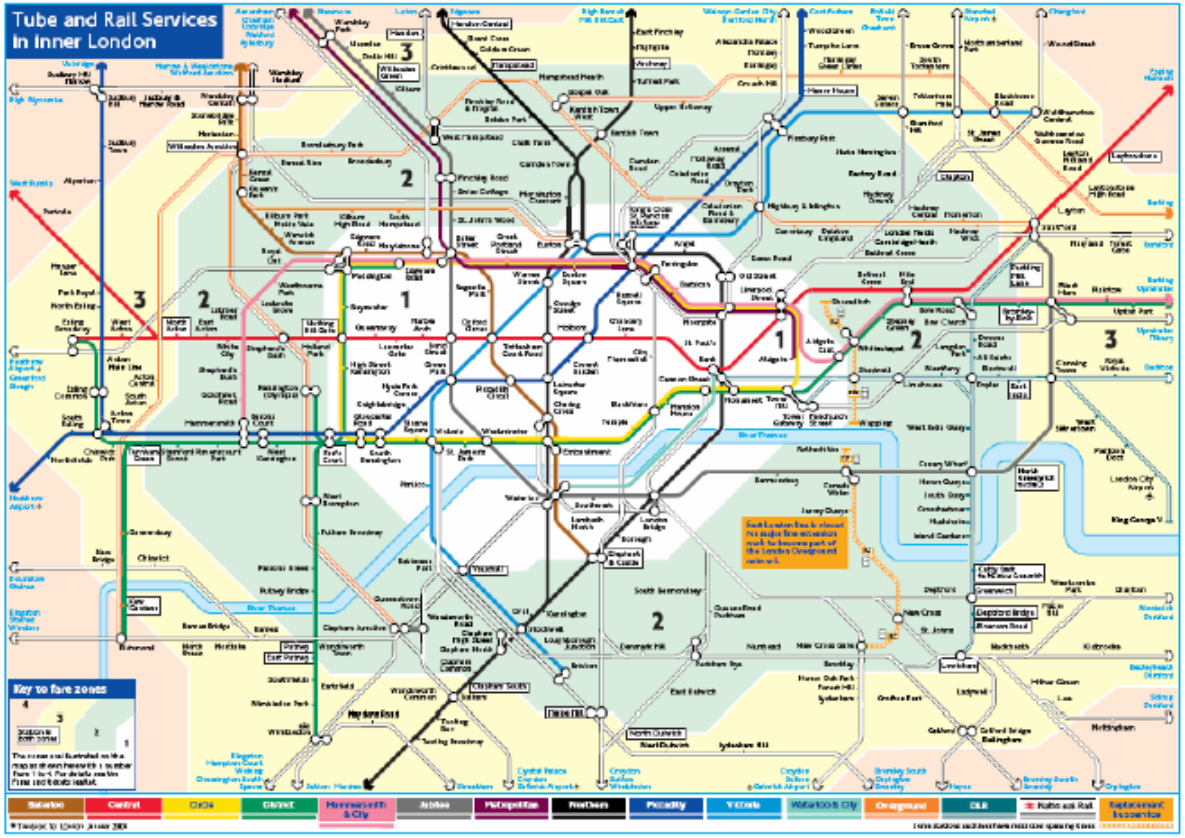
Network visualization: Transport maps

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/track

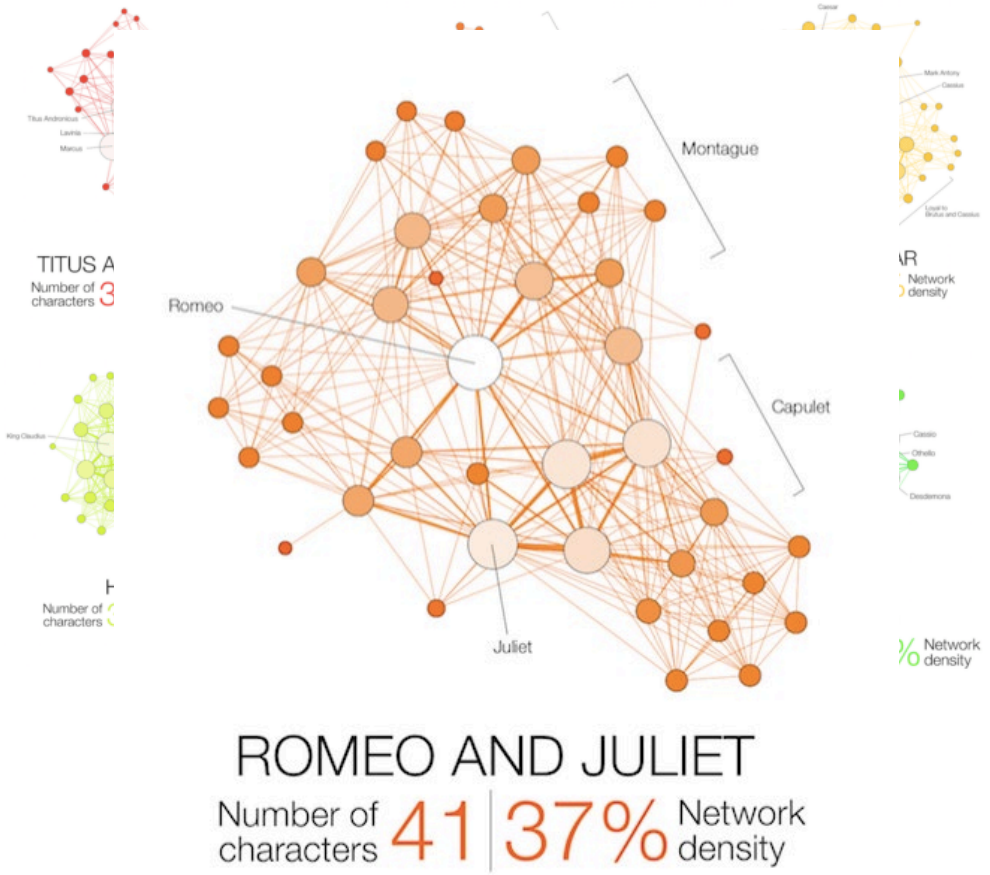
Network visualization: Shakespeare tragedies

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.



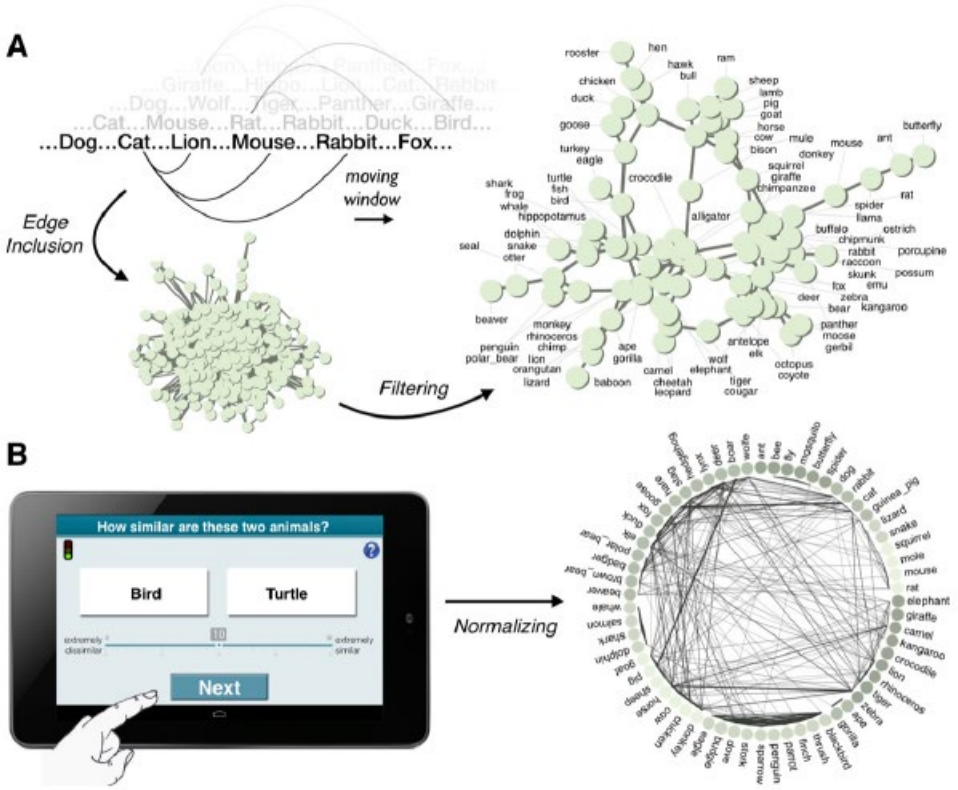
From: https://flowingdata.com/2015/12/30/shakespeare-tragedies-as-network-graphs/

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.



Similarity ratings: For each pair, indicate how similar they are

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

Semantic memory: Cognitive structure

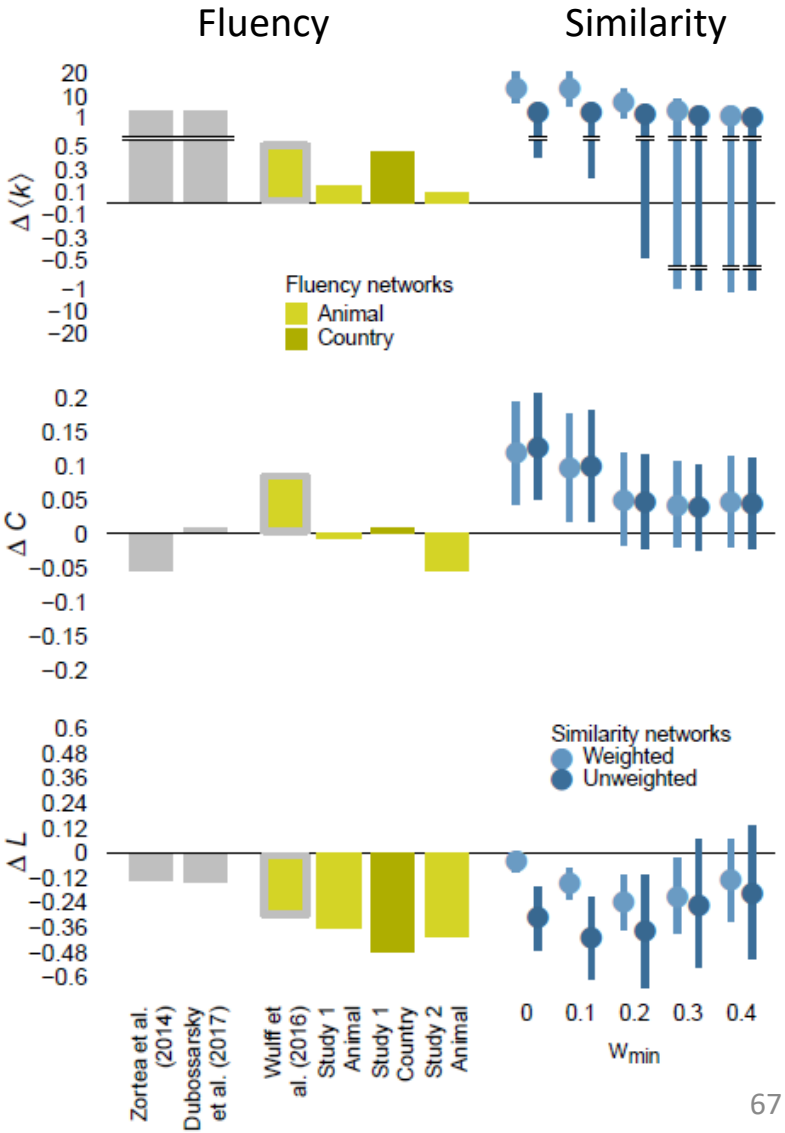
Do younger and older adults differ on measures calculated from their network diagrams?

<k> : Average "degree" # of connections
C : average local clustering
L : average path length in network

Δ() : young – old difference

IMHO, this graph tries to do too much.

The fluency data is most important to their argument.
ΔL & Δ <k> show consistent differences between young & old



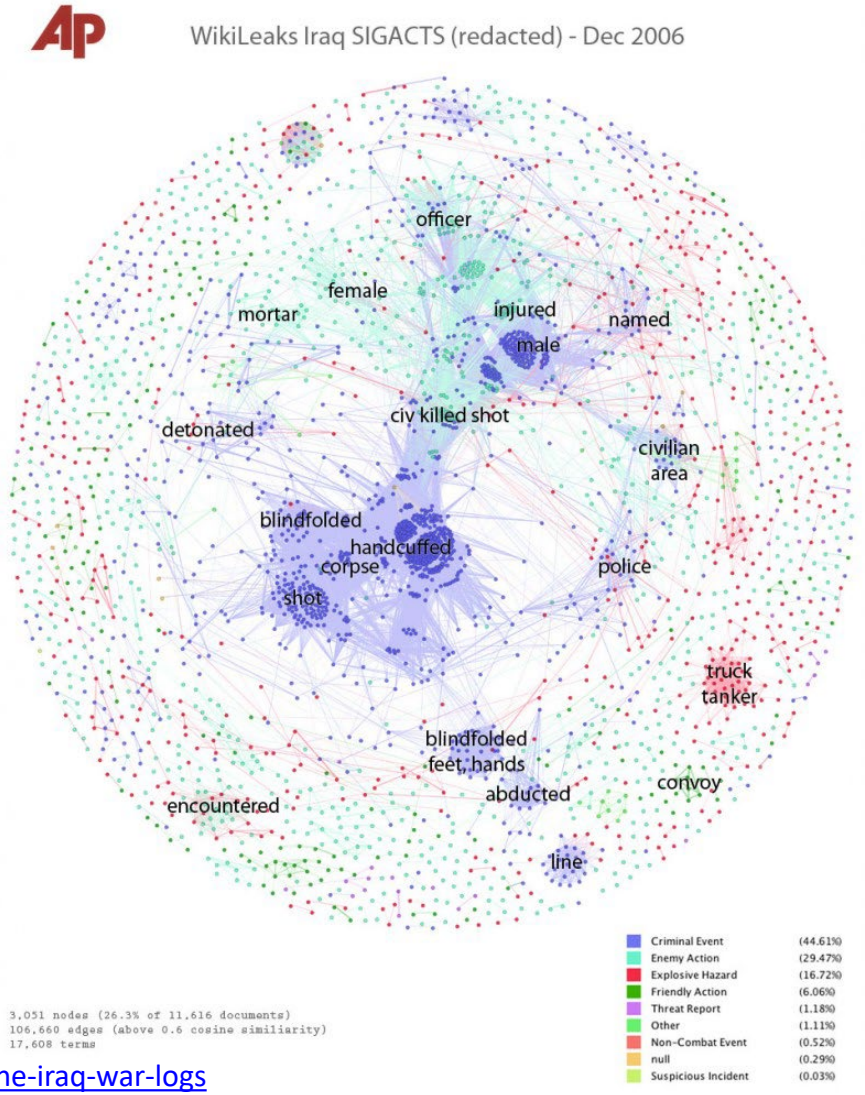
WikiLeaks Iraq war logs

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

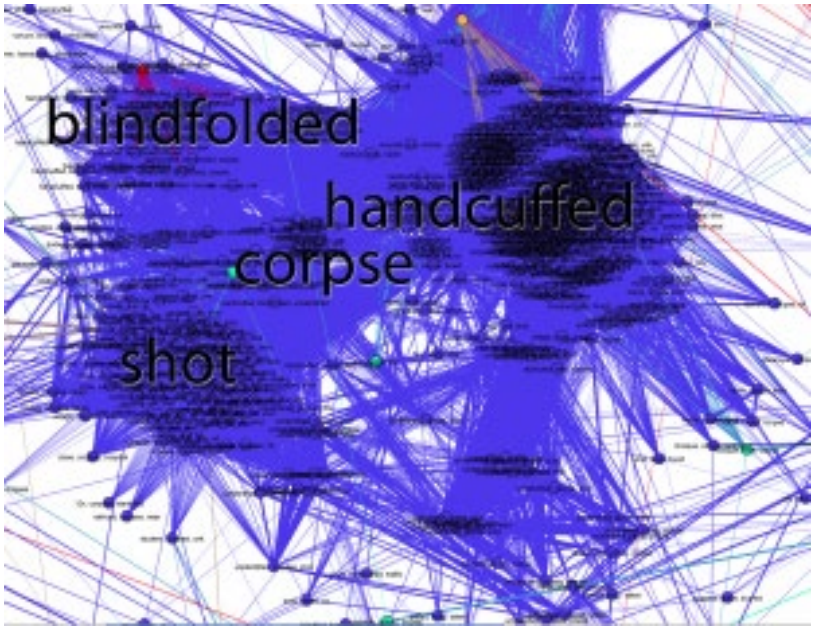


From: http://jonathanstray.com/a-full-text-visualization-of-the-iraq-war-logs

WikiLeaks Iraq war logs

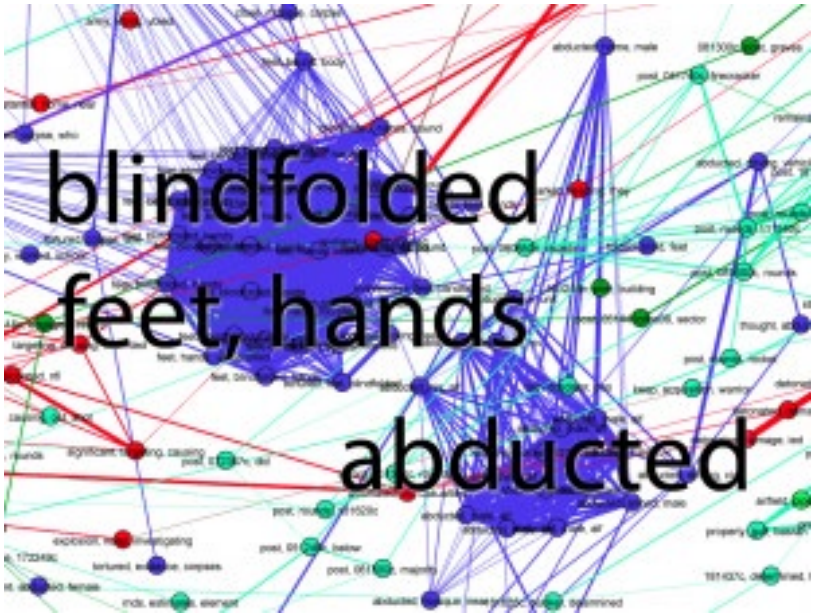
Certain themes became clear, and could be studied in rich detail
The underlying methods use “term frequency–inverse document frequency” measures of text-mining.

Murder cluster. All contain the word “corpse”



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

Torture-abduction cluster



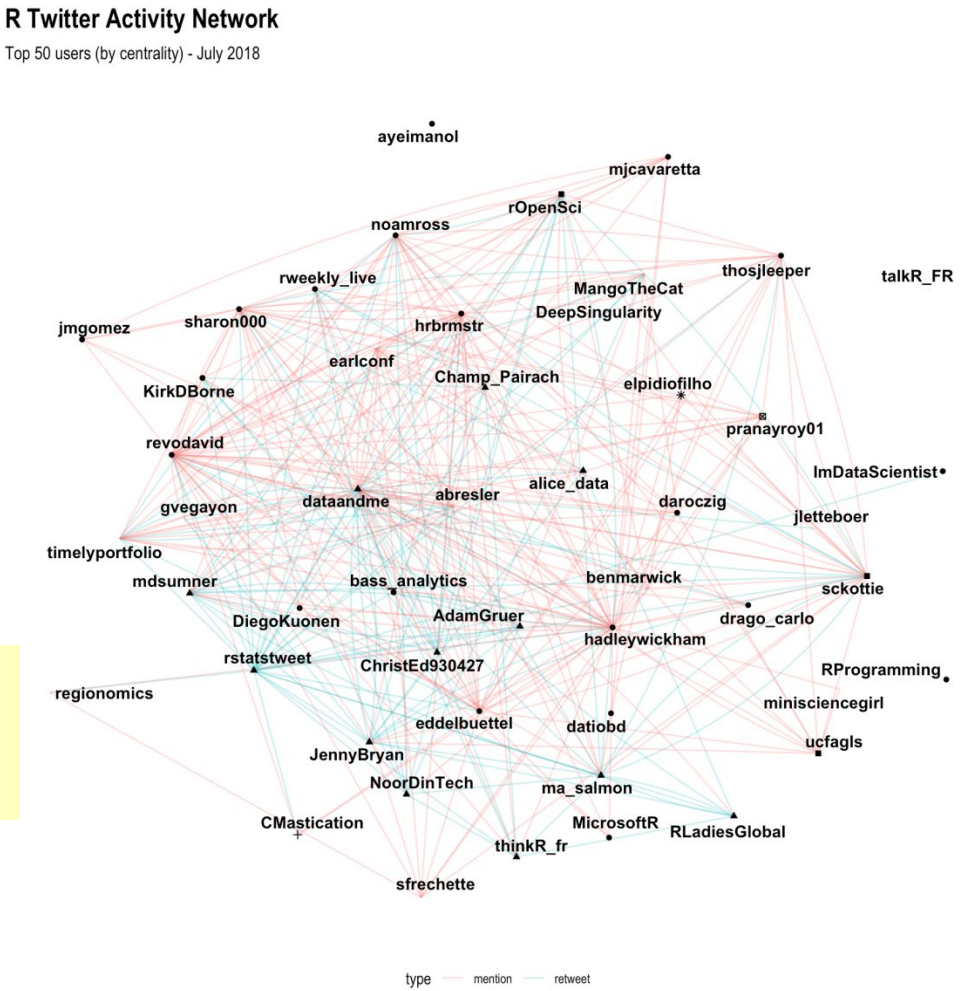
http://jonathanstray.com/wp-content/uploads/2010/12/Torture-abduction.png

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/

Twitter circles

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

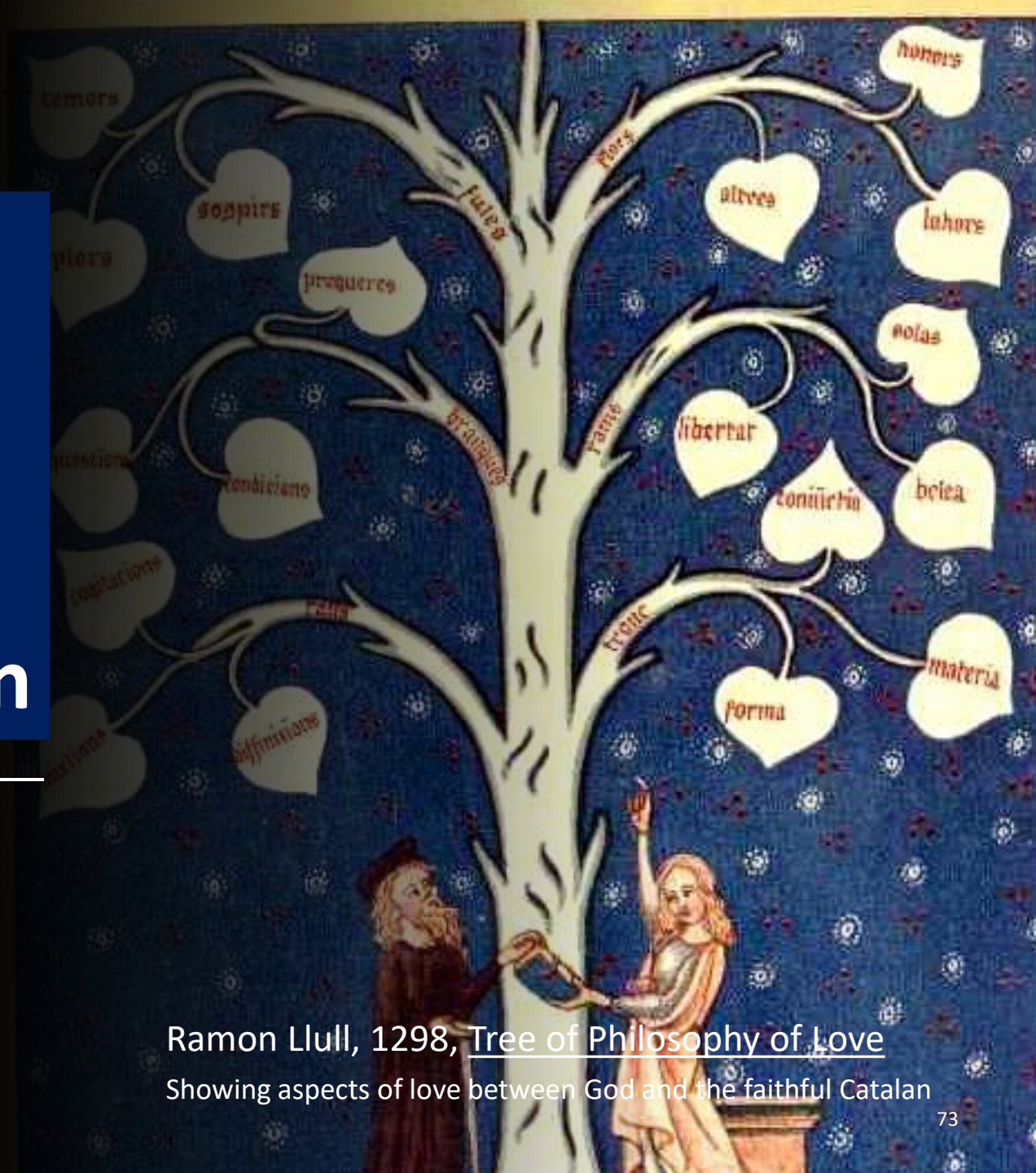
Generated by http://chirpty.com

Tree-based Visualization

Branching patterns

History as a tree

Treemaps

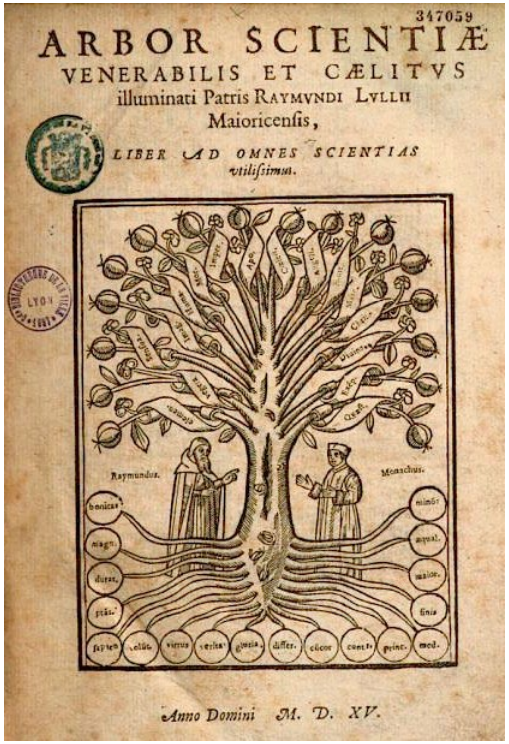


Ramon Llull, 1298, Tree of Philosophy of Love

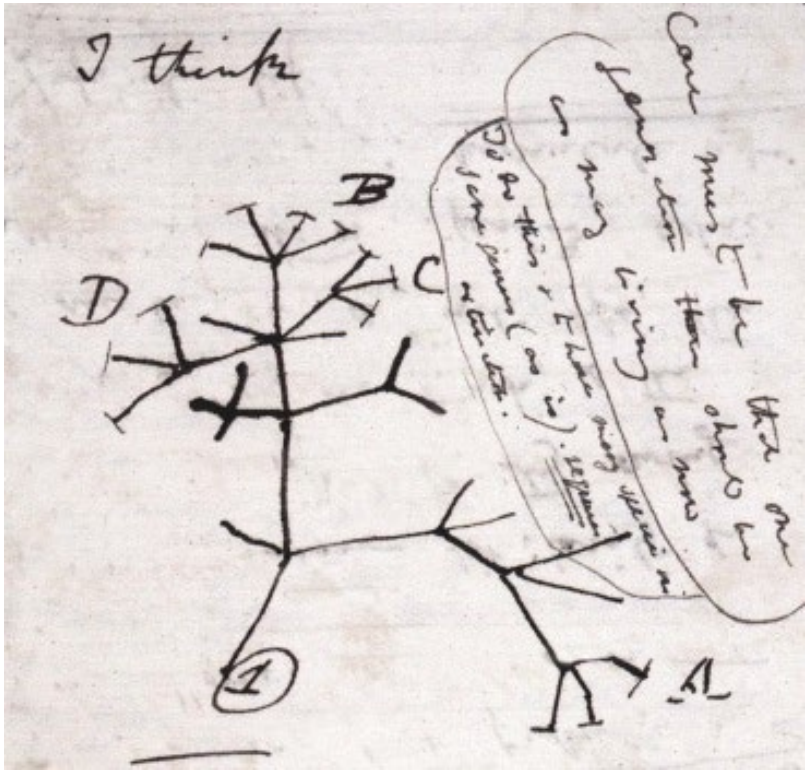
Showing aspects of love between God and the faithful Catalan

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

History as a Tree:

Geschictesbaum Europa (2003)

The entire history of Europe in one diagram

- space-filling design: resolution ~ time^2
- natural metaphors for roots, branches

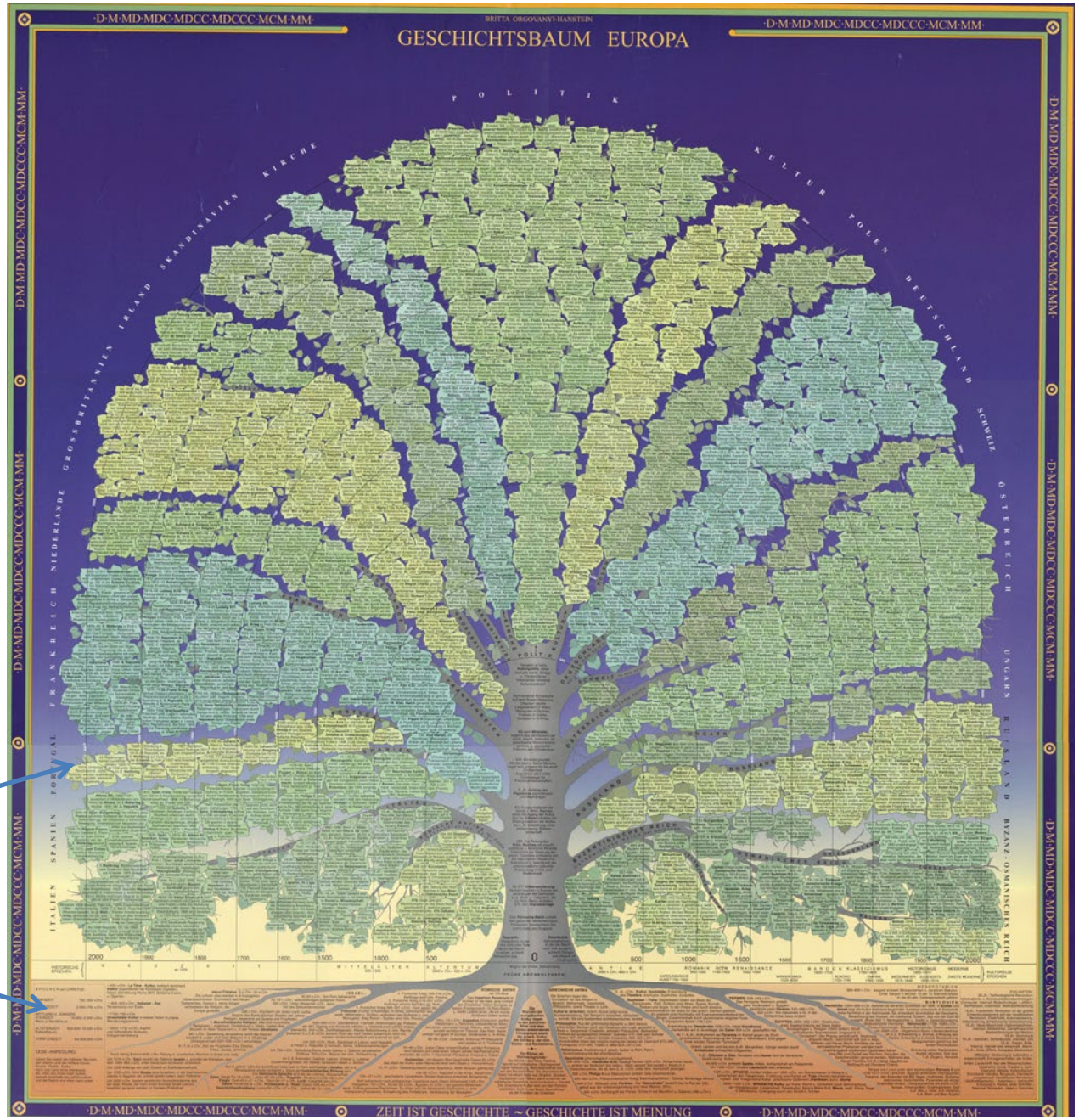
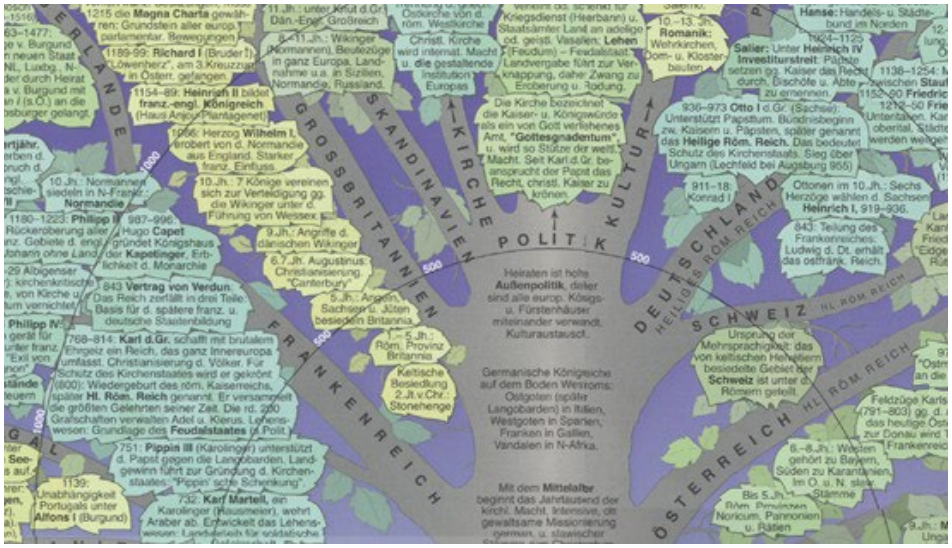
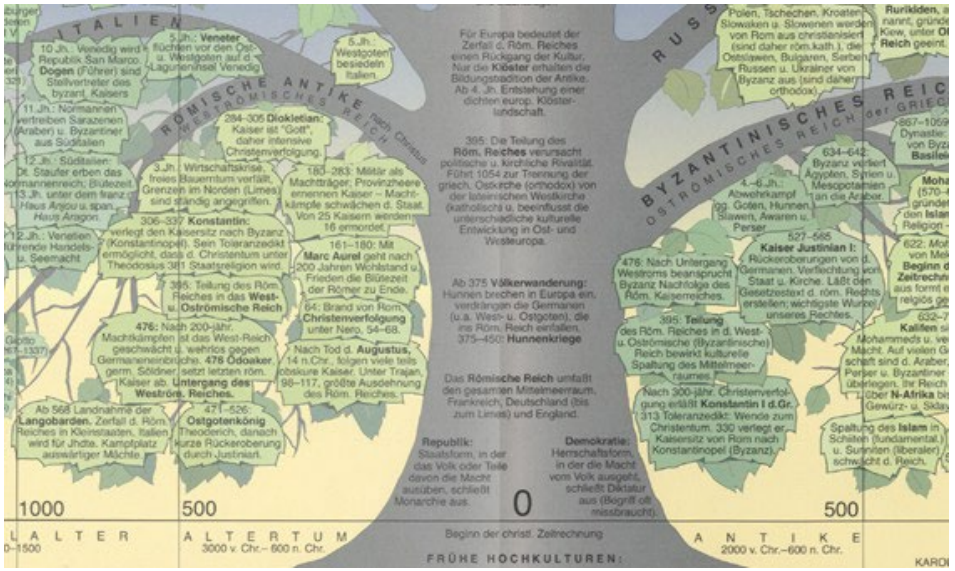


Image: David Rumsey map collection, https://bit.ly/3GYZ2iw

History as a Tree

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

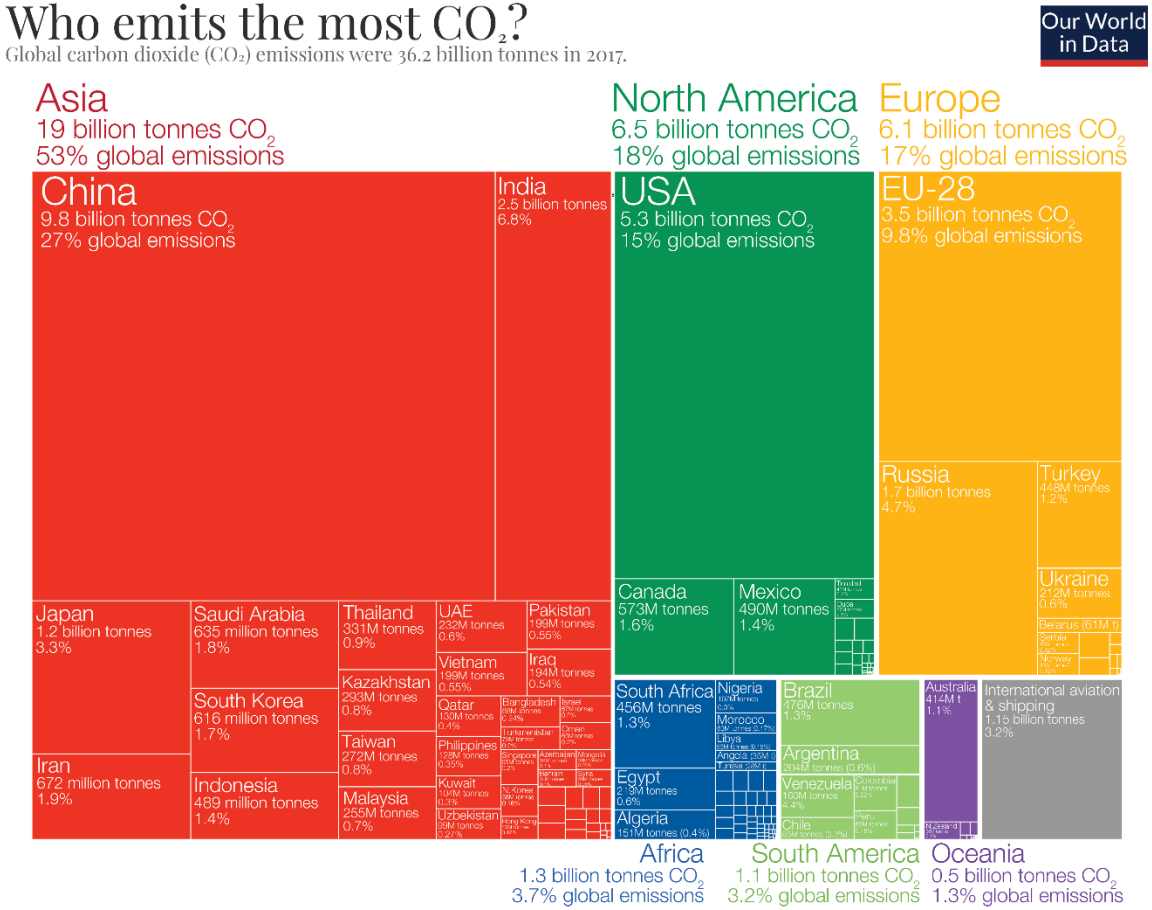




- linear horizontal scale → area ~ time^2

Treemaps

Treemaps display hierarchical data as a set of nested rectangles.
Each node (leaf) has an area ~ size (CO_2)



The construction makes efficient use of space

Nesting shows relative size at multiple levels: continent -> country

No limit to the depth of the branches

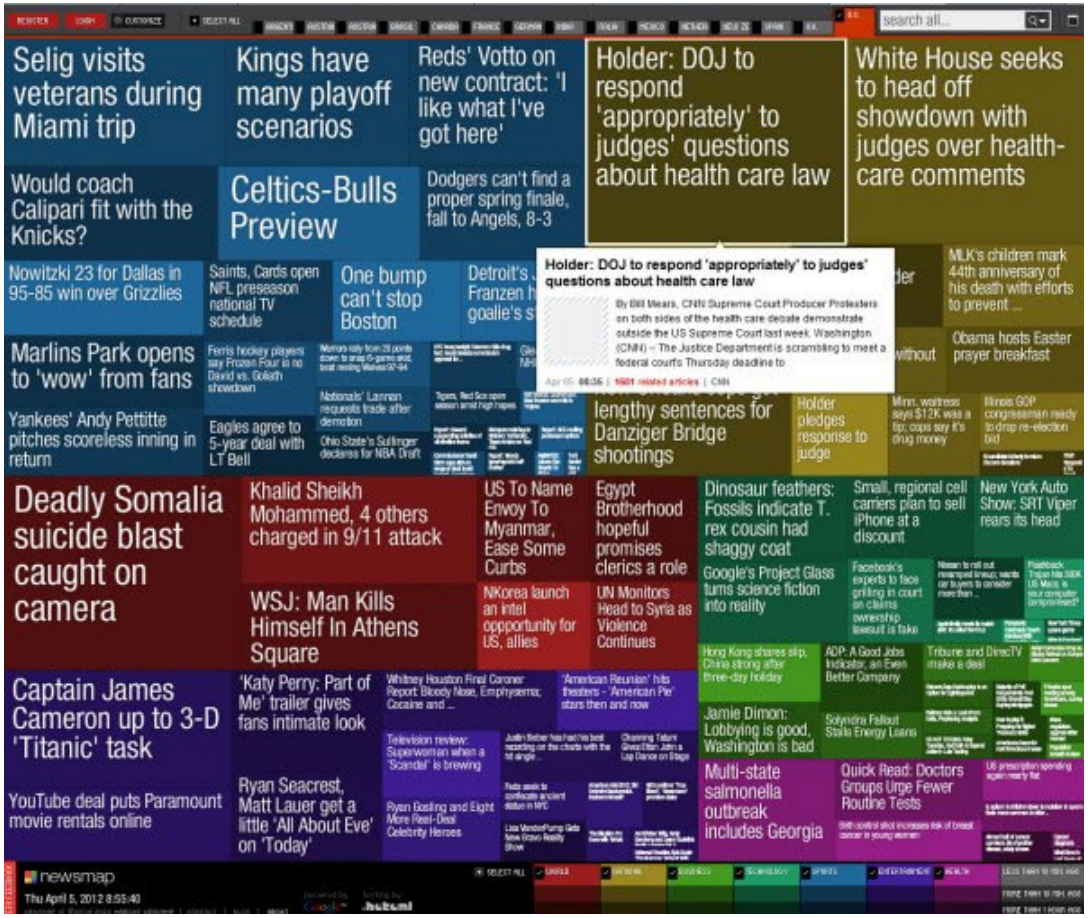
Treemaps: Google Newsmap

They turn out to be useful in a wide range of applications

Google NewsMap shows top news stories with

- Size ~ popularity
- Color: domain– world news, sports, national, …
- Shades: recency

Interactivity: Hover, click to show details



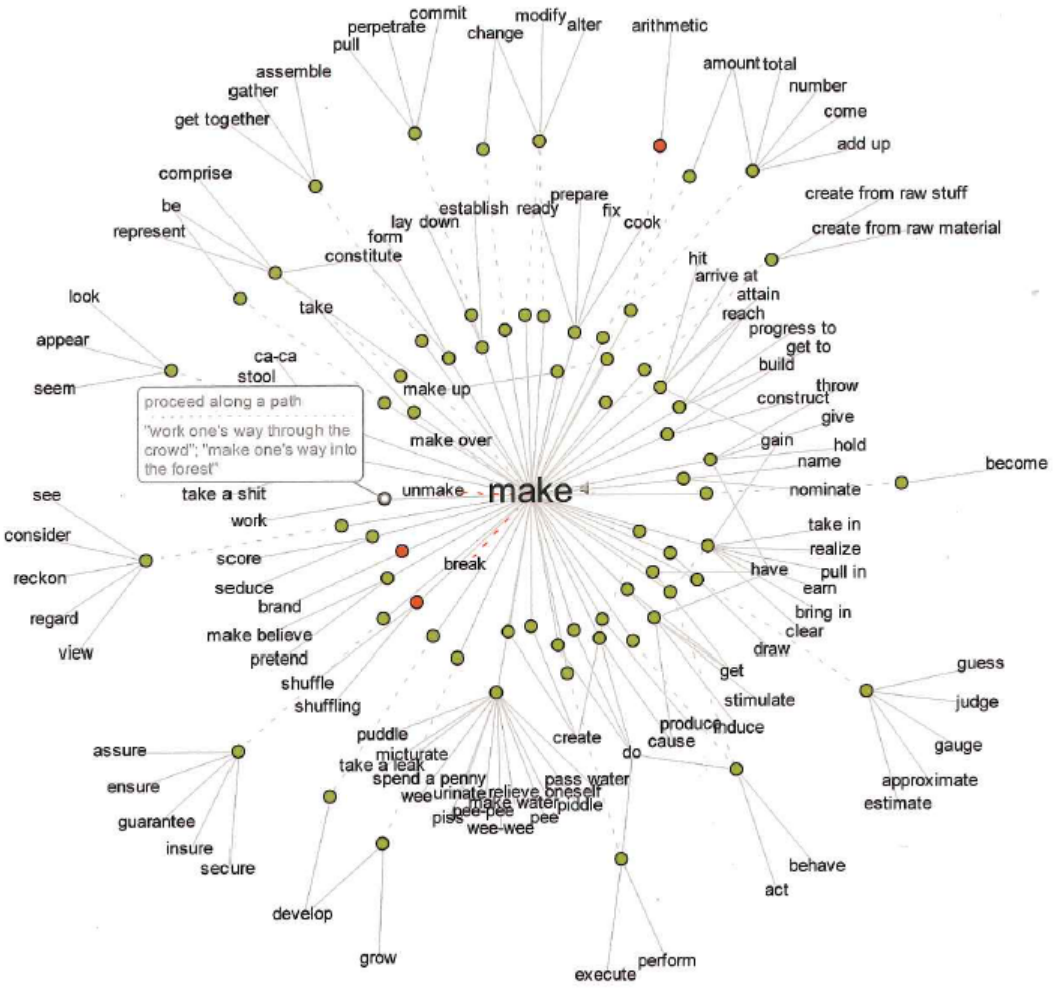
See: https://newsmap.ijmacd.com/. This uses https://github.com/ijmacd/newsmap-js

Radial trees: *Visual Thesaurus*

The *Visual Thesaurus*, from Thinkmap was the first application to make word meanings visual and interactive.

They used a radial layout to show the various related senses of given focus word.

This application was incisive in promoting ideas of interaction with tree-based data: query, zoom, tool-tips, …



This fig from Manuel Lima, *The Book of Trees*, p. 127

Voroni treemaps

Le nouveau camembert?

A voroni diagram subdivides space into polygons.

As a treemap, this is done recursively in hierarchical categories.

The circular form makes it an easy replacement for a lowly π chart,

Image from: https://community.tableau.com/s/question/0D54T00000C6fMDSAZ/voronoi-diagram-and-other-charts

Animation & Interactive Graphics

Origins: Visualizing motion

Animated graphics

Dynamically updated graphics

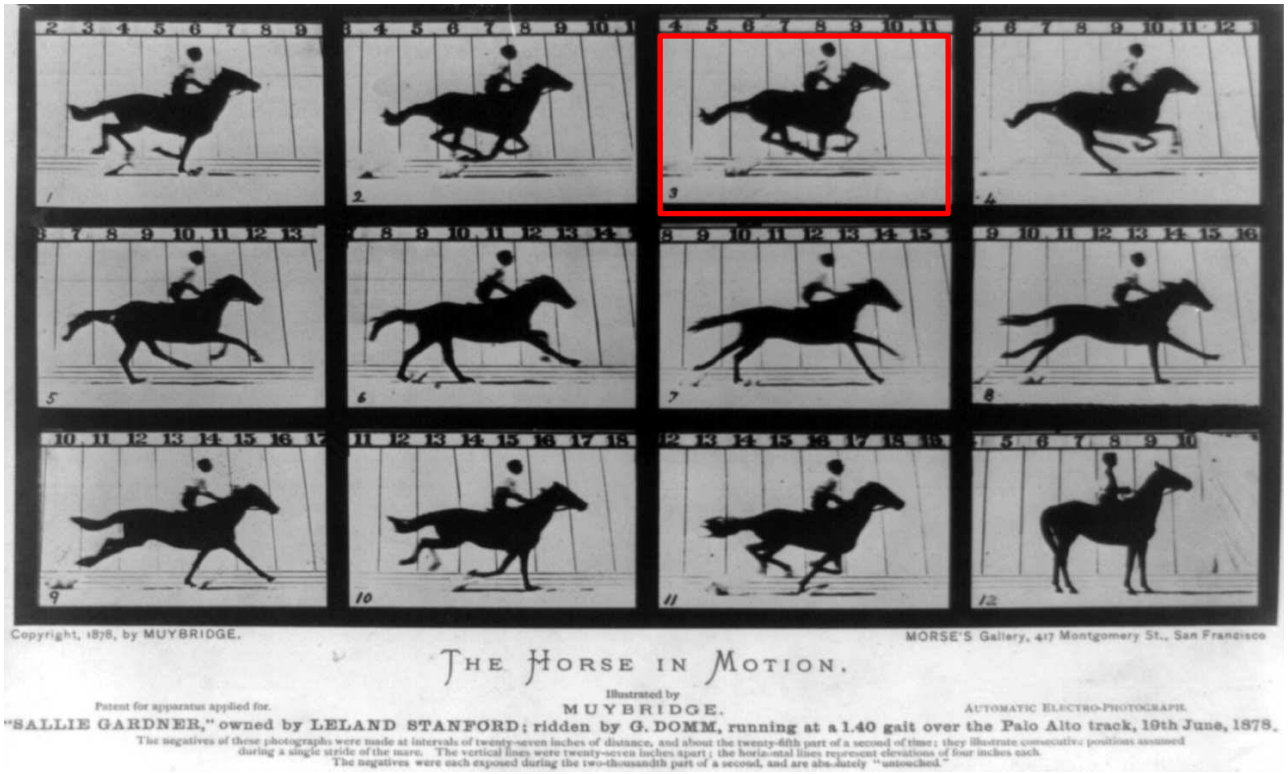
Linking views

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

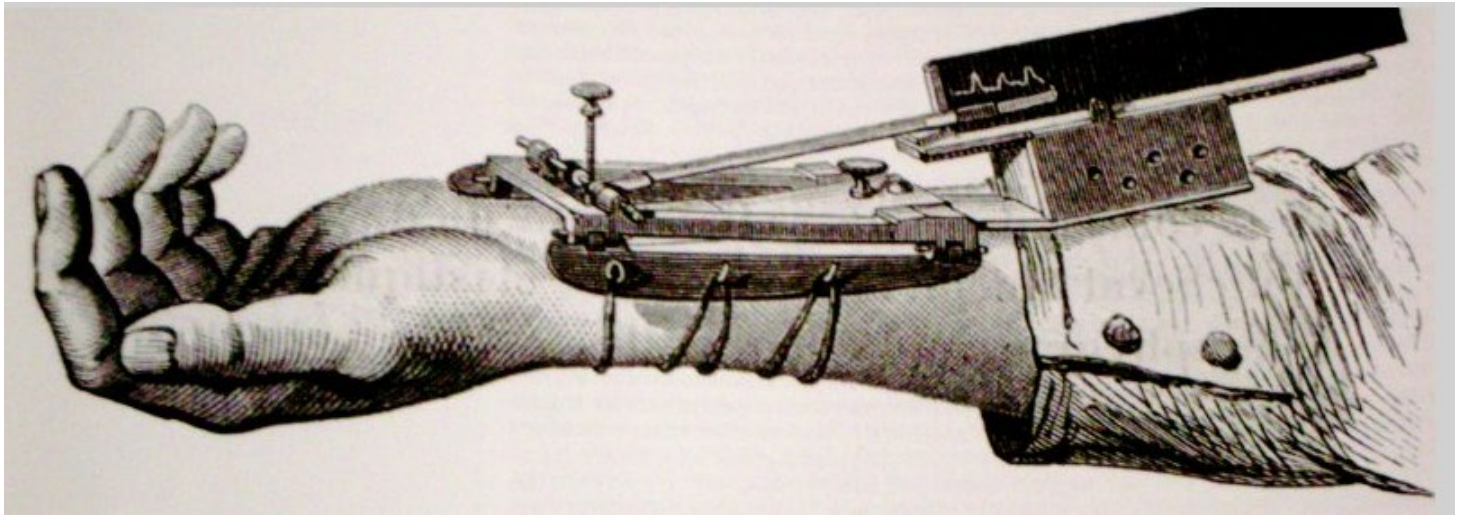
Eadweard Muybridge solved the problem by automating multiple photographs



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”

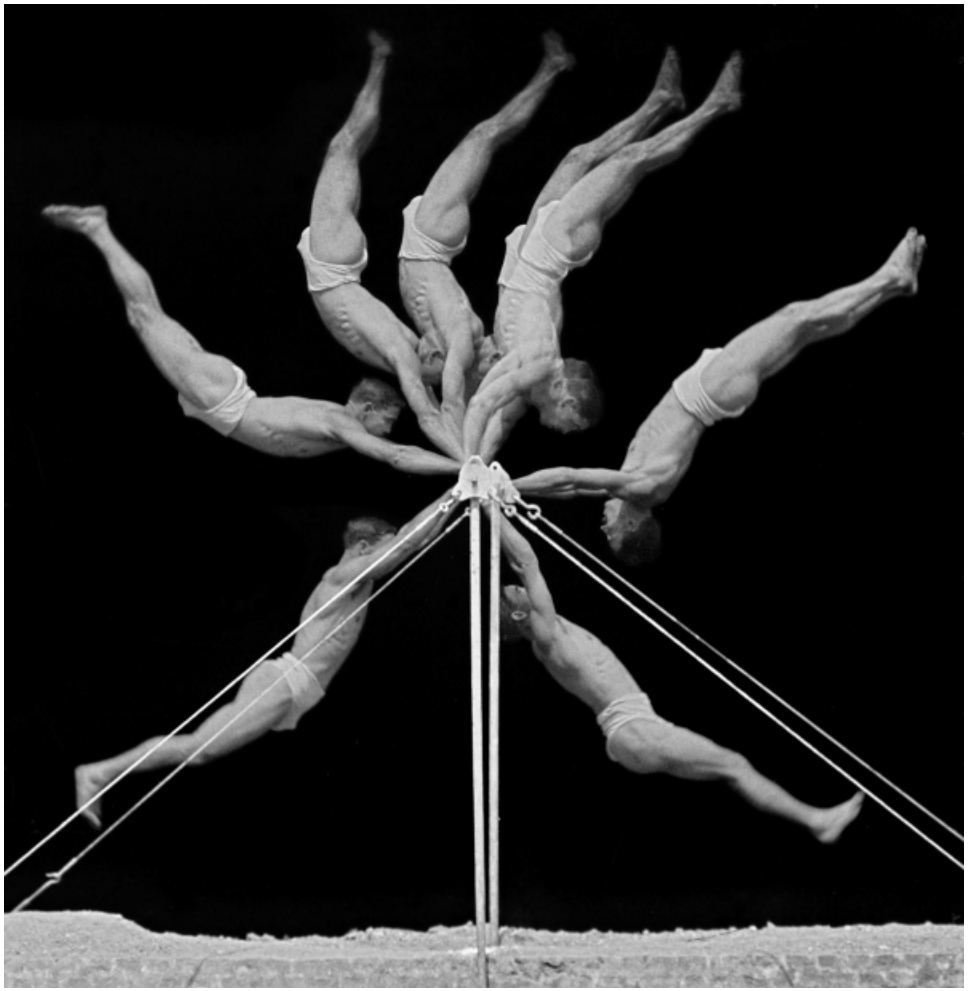
Marey’s sphygmograph, recording a visual trace of arterial blood pressure



Animation: Chronophotography

Marey pioneered the study of human and animal motion photographically

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

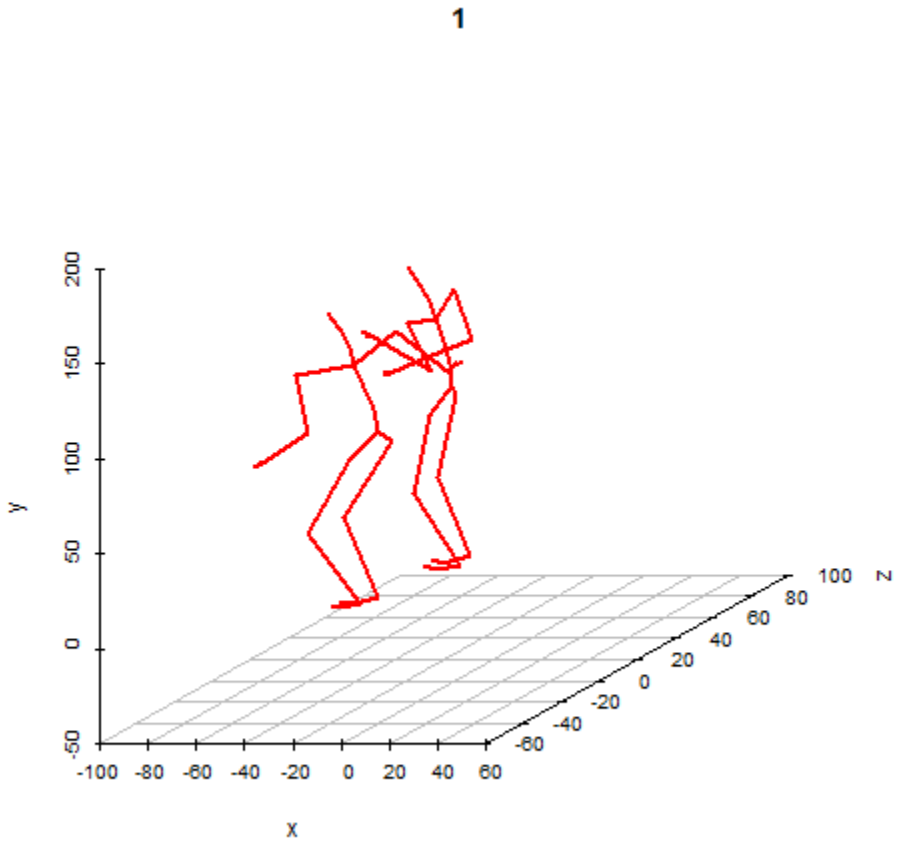


Animated graphics

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

Animated line graph

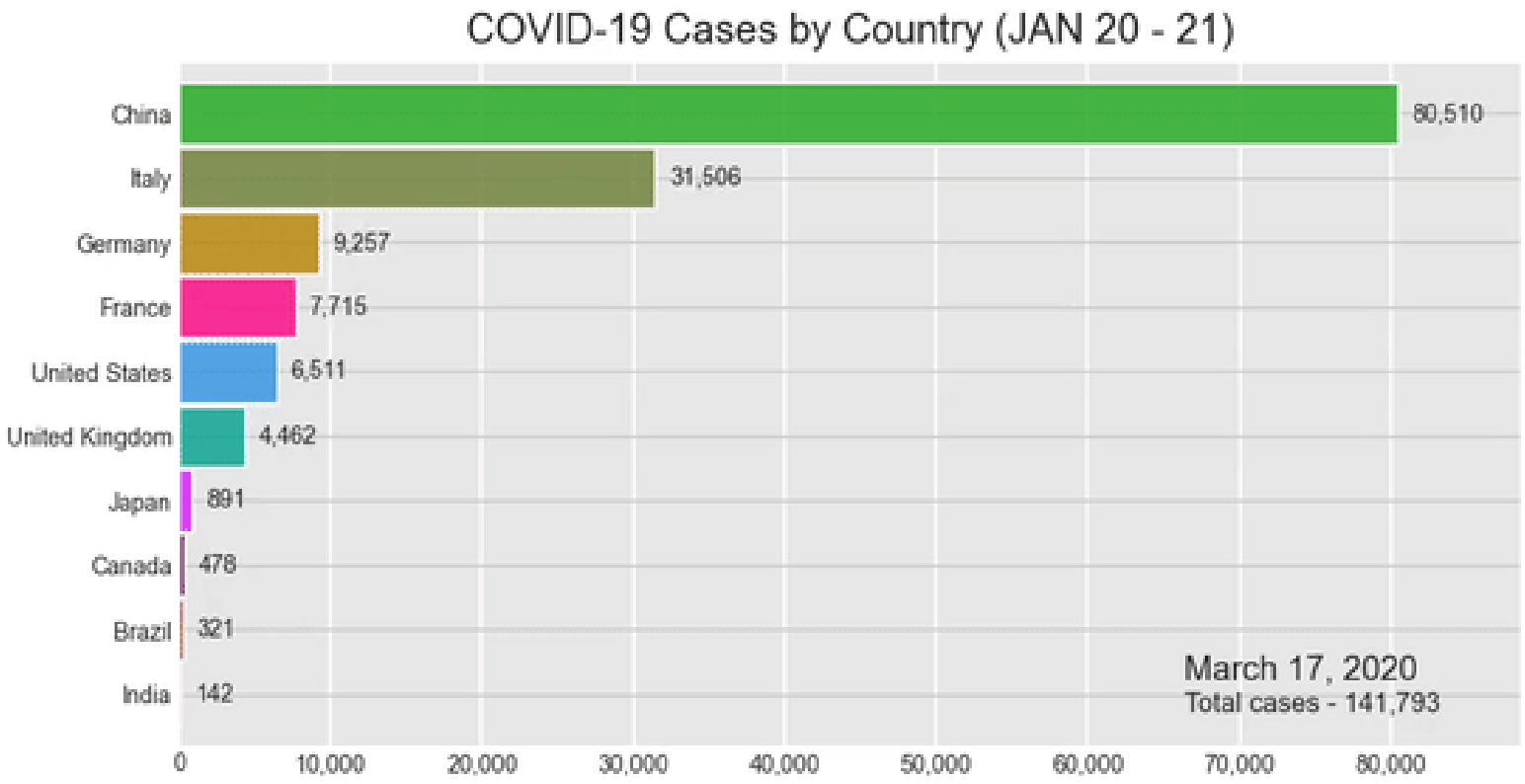
To show the evolution of time series over time, you can use an animation line graph

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

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.



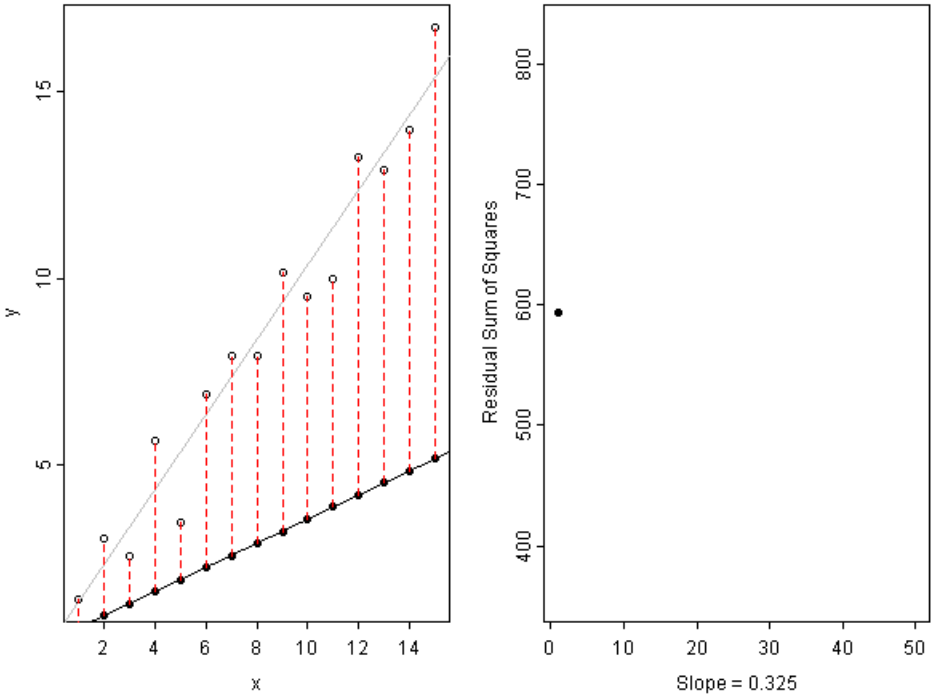
Statistical animations

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.

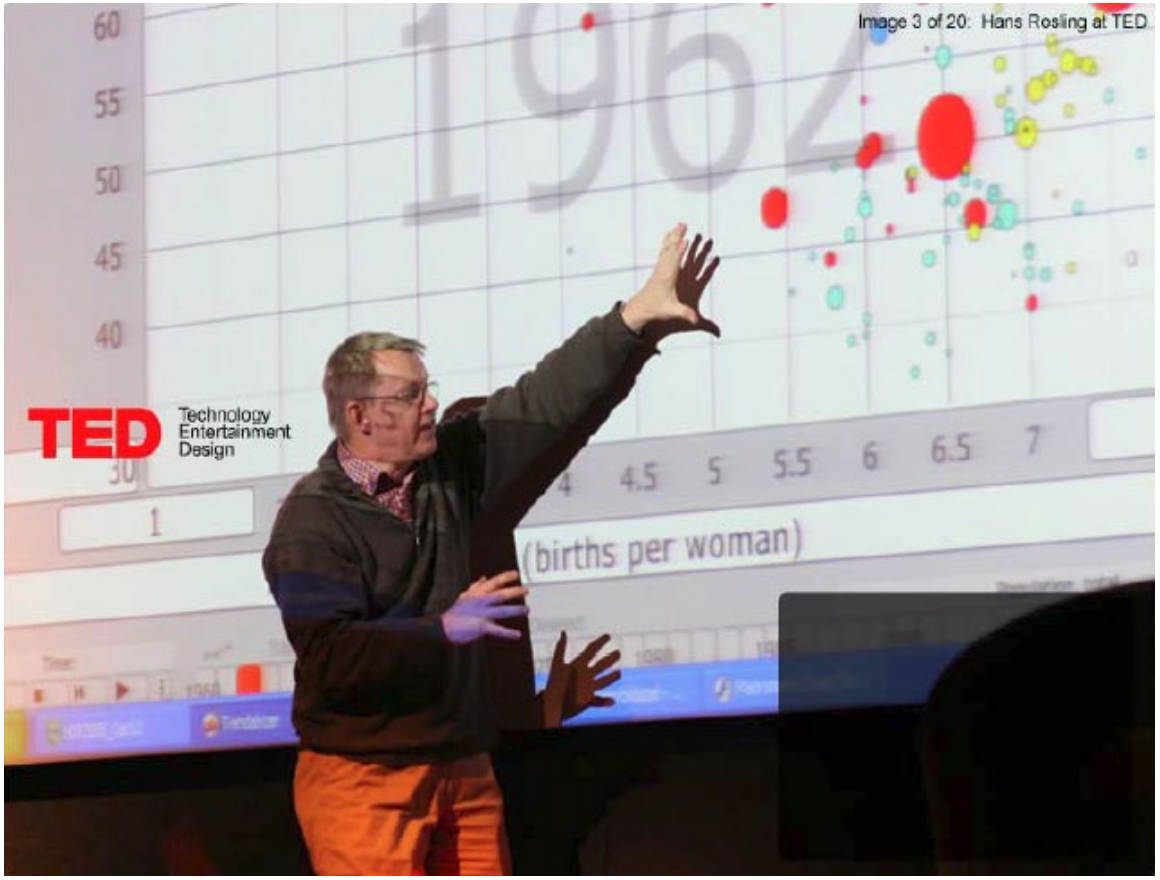
As the slope of the line is varied, the right panel shows the residual sum of squares.

This plot was done using the animate package in R.



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

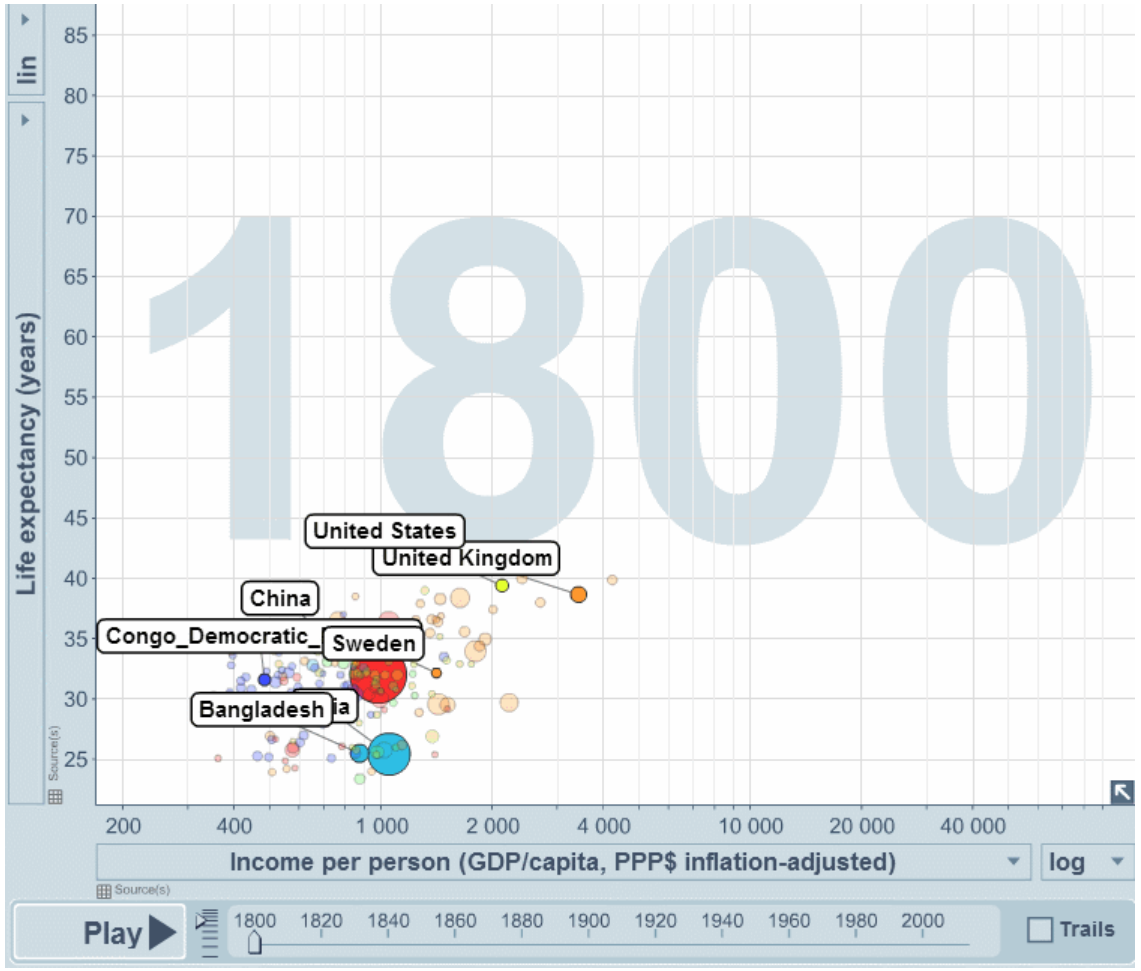
Animation & Interactivity

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

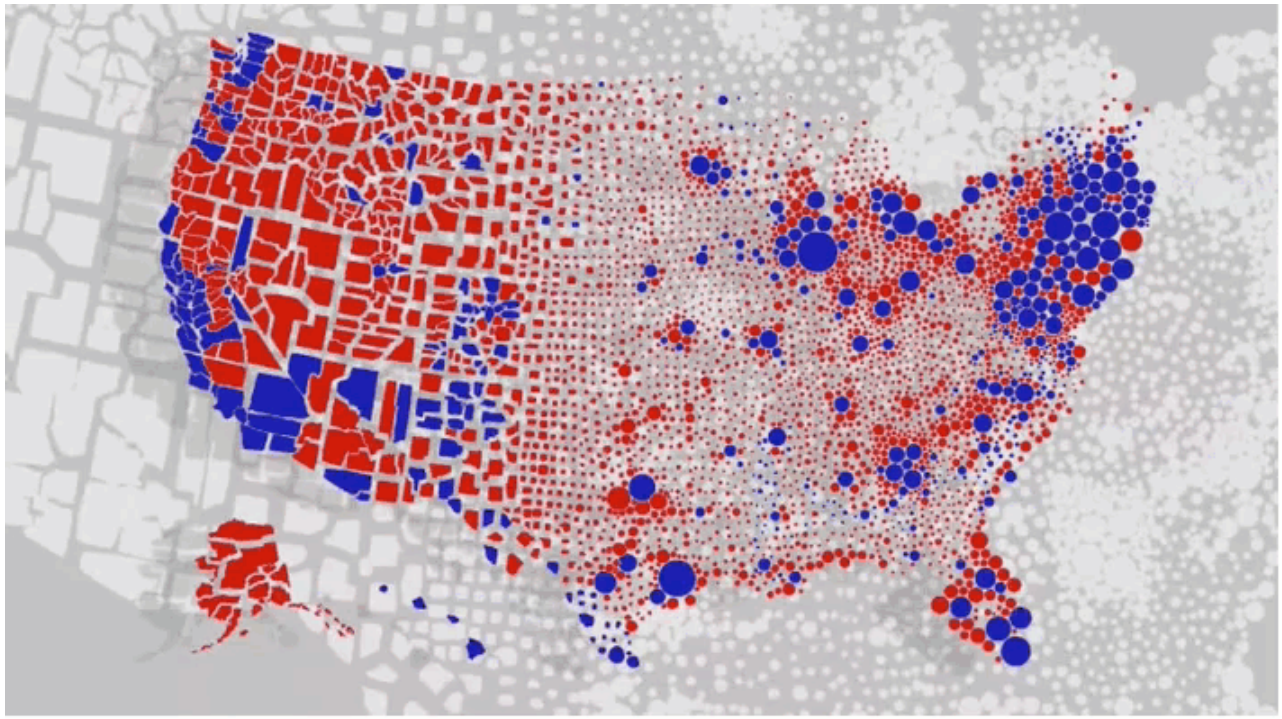


Animation: Interpolated views

Animation can also be used to show the difference between two views, using interpolated transitions: Current = α $view_1$ + (1- α) $view_2$

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 Douïeb/Jetpack.ai]

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

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

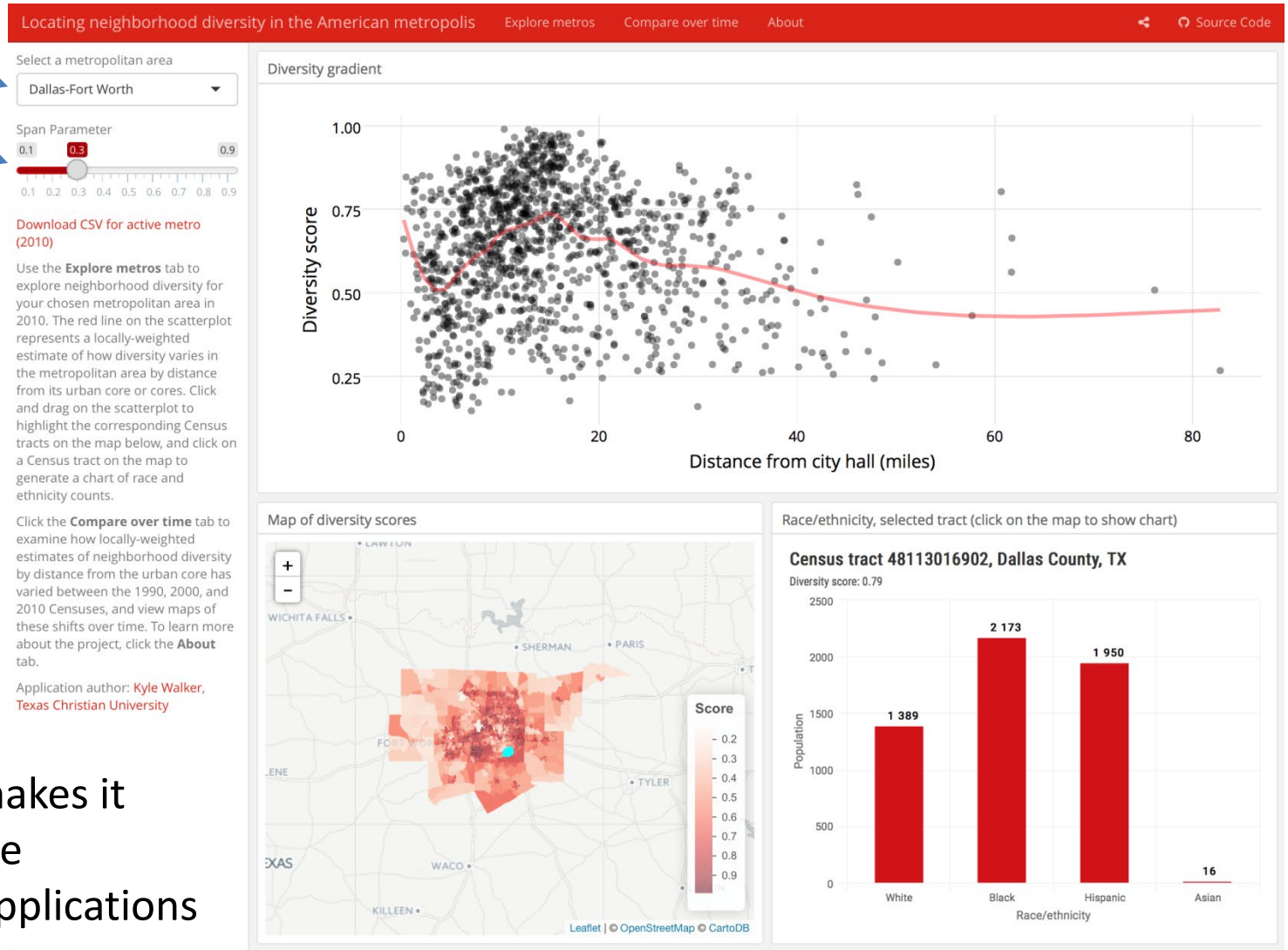
The grand tour animates a series of 2D projections of the 5D data

The image is recorded as a GIF

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

Interactive application frameworks

selectors
inputs

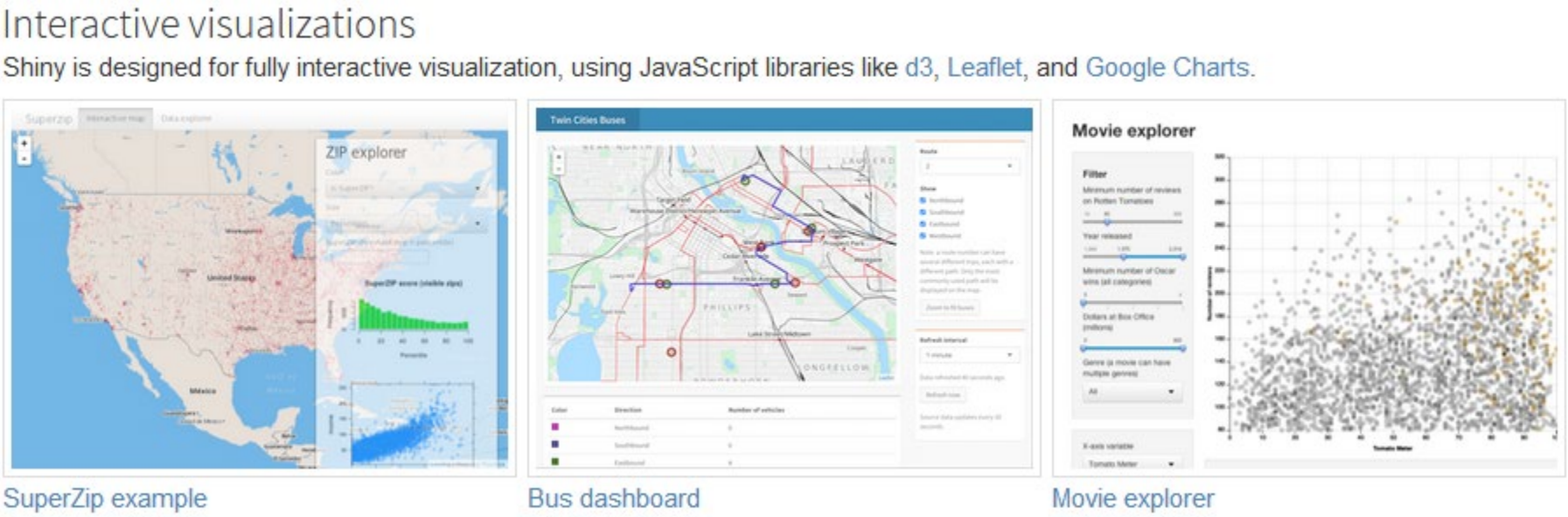


shiny for R makes it easy to create interactive applications

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

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, …



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