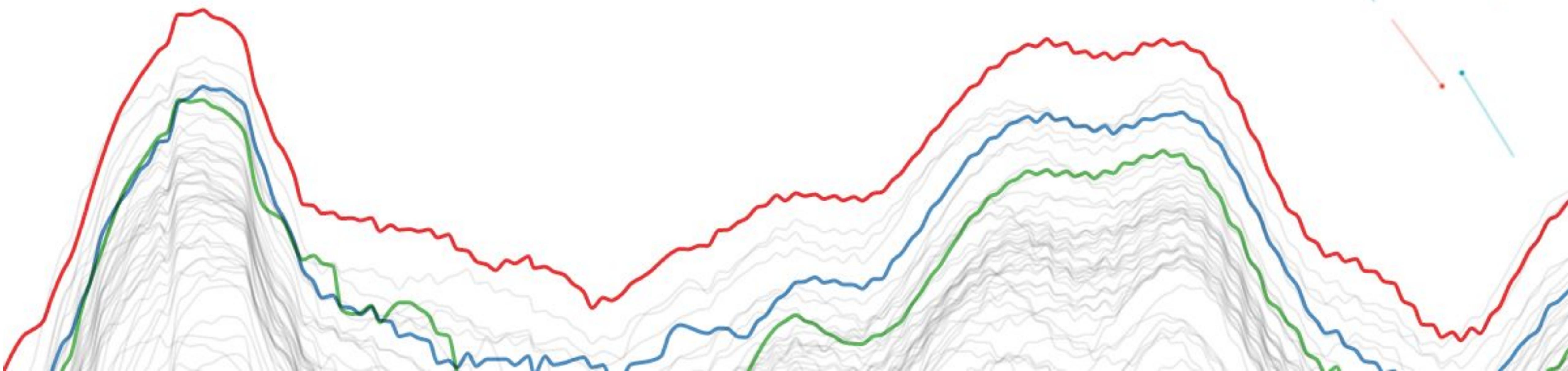


Visualización de la información

Fernando Becerra

f.becerra@udd.cl | [@fdo_becerra](https://twitter.com/fdo_becerra) | www.fernandobecerra.com

Diplomado y Magíster en Data Science, Universidad del Desarrollo



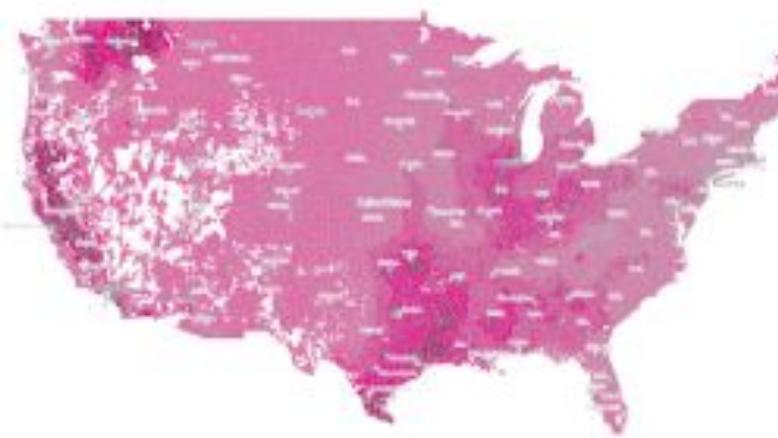
¿Quién soy?

FERNANDO BECERRA

DATA ANALYSIS + VISUALIZATION

I'm an astrophysicist turned data analyst and data visualization developer. My work is a mix of research, data science, analysis, and visualization and spans a wide range of topics such as sports, science, environment, education, government, and public policy. If you are interested in working together, feel free to [reach out](#).

 TWITTER  GITHUB  LINKEDIN  RESUME  EMAIL



Head Start Environmental Exposure

Where are children in Head Start exposed to environmental hazards?

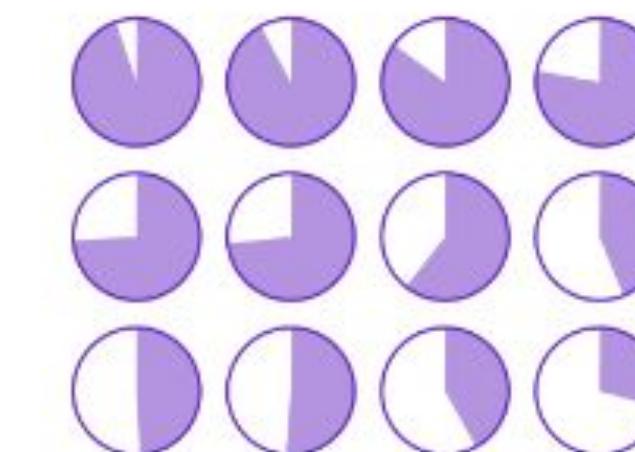
D3.JS MAPBOX DASHBOARD



Inflation for Americans at each age

How does inflation affect Americans at different ages?

D3.JS R REACT



Housing costs for Americans

How much do Americans spend on housing?

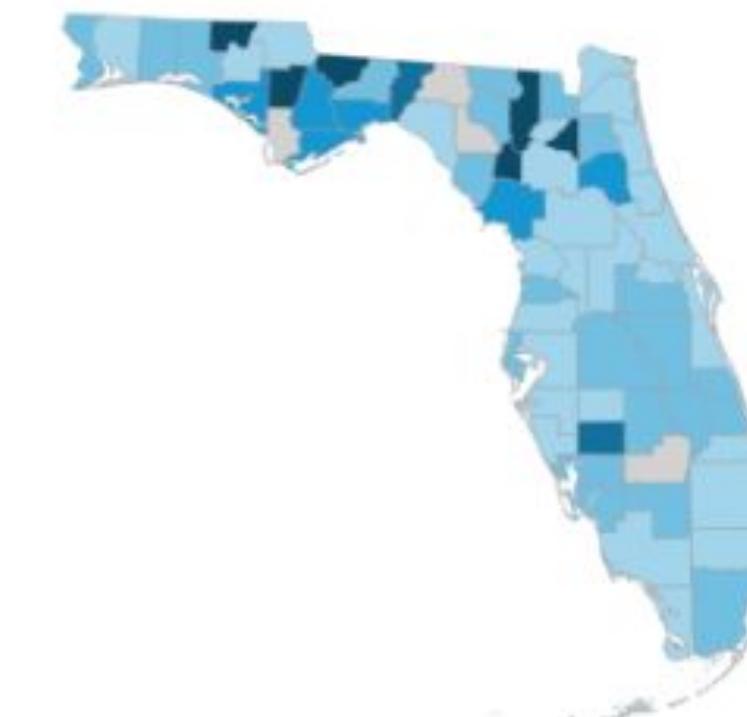
D3.JS PYTHON SVELTE



Epic Data Explorer

A web-based tool to navigate through the Positive Disruption model output.

D3.JS DASHBOARD DATA VISUALIZATION



Beyond education outcomes

How equitably distributed are educational resources?

D3.JS INTERACTIVE DATA VISUALIZATION

<https://www.fernandobecerra.com>

Este curso

Objetivos

1. Entender los conceptos fundamentales dentro del campo de visualización de información
2. Aprender las técnicas y mejores prácticas para diseñar una visualización de datos
3. Desarrollar un pensamiento crítico frente a los gráficos que se ocupan día a día
4. Crear e implementar visualizaciones de datos en Python
5. Usar la visualización de datos como una forma de comunicar y contar una historia



python

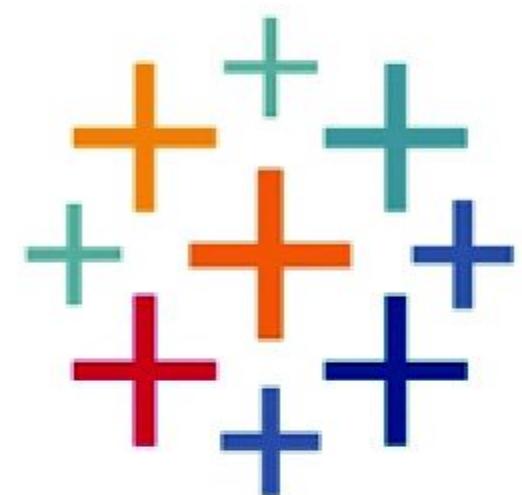




Power BI



Looker



+a b|e a u

Calendario

Fecha	Evento	Tema
viernes 23 de agosto	clase #1	Cátedra: Tipos de visualizaciones Workshop: Gráficos básicos en Python
viernes 30 de agosto	clase #2	Cátedra: Qué, por qué y cómo visualizar Workshop: Gráficos avanzados en Python
viernes 6 de septiembre	clase #3	Cátedra: Elementos de composición Workshop: Herramientas adicionales
jueves 12 de septiembre	entrega tarea #1	Crítica de visualizaciones
viernes 13 de septiembre	clase #4	Workshop: Llevando los gráficos al próximo nivel
jueves 26 de septiembre	entrega tarea #2	Los buenos ejemplos se copian
viernes 27 de septiembre	clase #5	Cátedra: Storytelling Workshop: Contando historias con gráficos
viernes 4 de octubre	clase #6	Workshop: Ensayos visuales
jueves 10 de octubre	entrega tarea #3	Creando visualizaciones de datos
viernes 18 de octubre	clase #7	Cátedra: Mas allá de Python
jueves 24 de octubre	entrega tarea #4	Ensayo explicativo visual

Tarea #1: Crítica de visualizaciones

1. Individual
2. Crítica de tres visualizaciones publicadas este año. Deben al menos ser diversas en el tipo de gráfico que muestran.
3. Comentar cosas que la visualización hace bien, cosas que hace mal y como la mejorarían.
4. La rúbrica puede ser encontrada [acá](#). (Requiere correo UDD)

Tarea #2: Los buenos ejemplos se copian

1. Individual
2. Dadas tres visualizaciones, tratar de reproducirlas en Python.
3. La rúbrica puede ser encontrada [acá](#). (Requiere correo UDD)

Tarea #3: Creando visualizaciones de datos

1. Individual
2. Dado un grupo de datos, crear 3 visualizaciones distintas: dos clásicas (líneas, barras, dispersión, mapa de calor, torta, área) y una menos convencional.
3. No necesariamente relacionadas al mismo tema.
4. La rúbrica puede ser encontrada [acá](#). (Requiere correo UDD)

Tarea #4: Ensayo explicativo visual

1. Grupo de 3 ó 4 personas
2. Jupyter notebook explicativo.
3. Ensayo visual: contar una historia apoyándose en gráficos.
4. La rúbrica puede ser encontrada [acá](#). (Requiere correo UDD)

Advertencias

1. Este curso es una mezcla de análisis y visualización de datos, así es que habrá una parte importante de procesamiento y análisis con pandas y geopandas.
2. Este curso tiene una componente altamente subjetiva, por lo tanto es difícil definir una pauta totalmente objetiva para las evaluaciones.

Materiales

En el repositorio de GitHub:

<https://github.com/fbecerra/curso-infovis-python>

(Si no saben que es git, esta guía les puede servir:

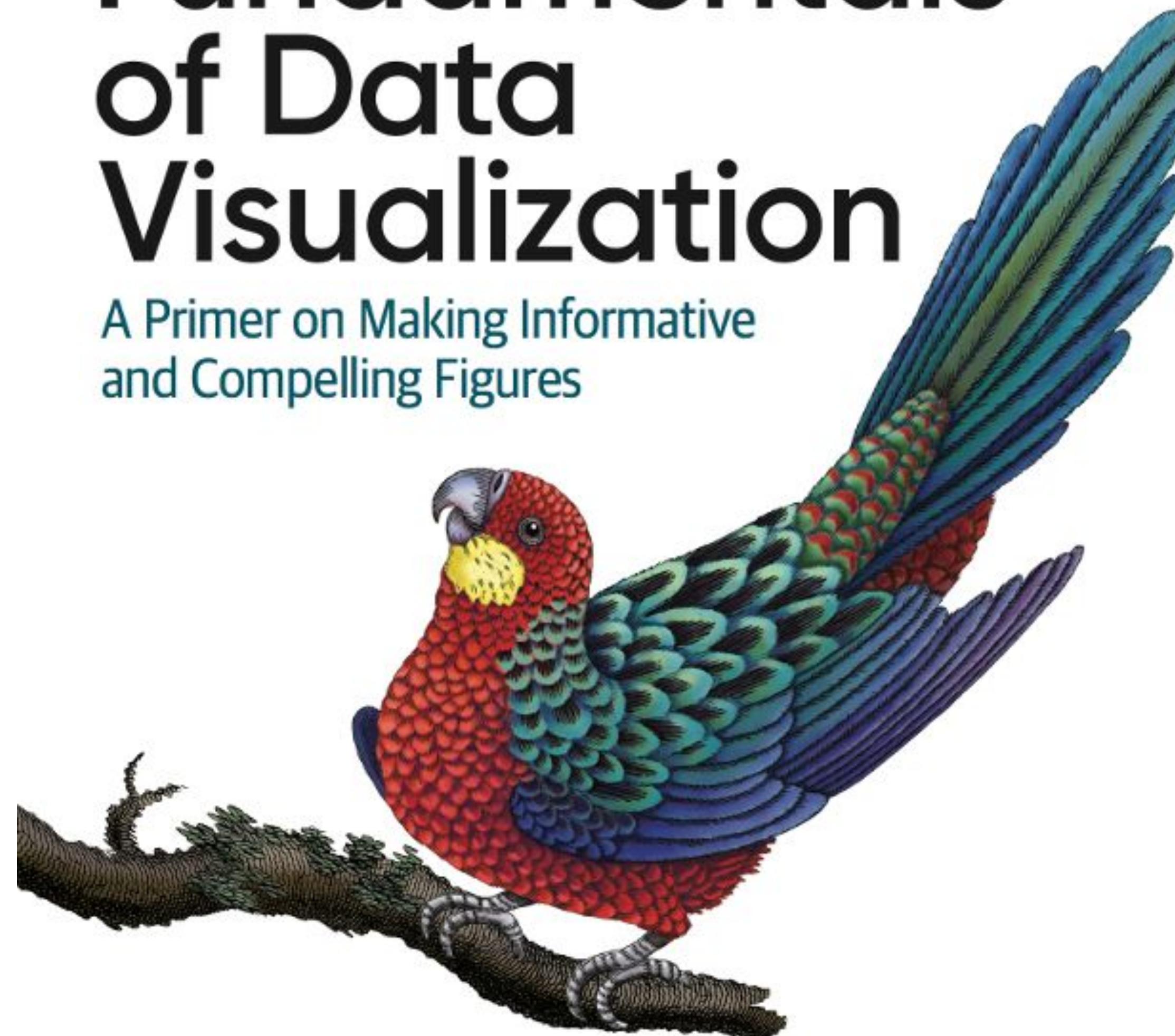
<https://rogerdudler.github.io/git-guide/index.es.html>

Bibliografía

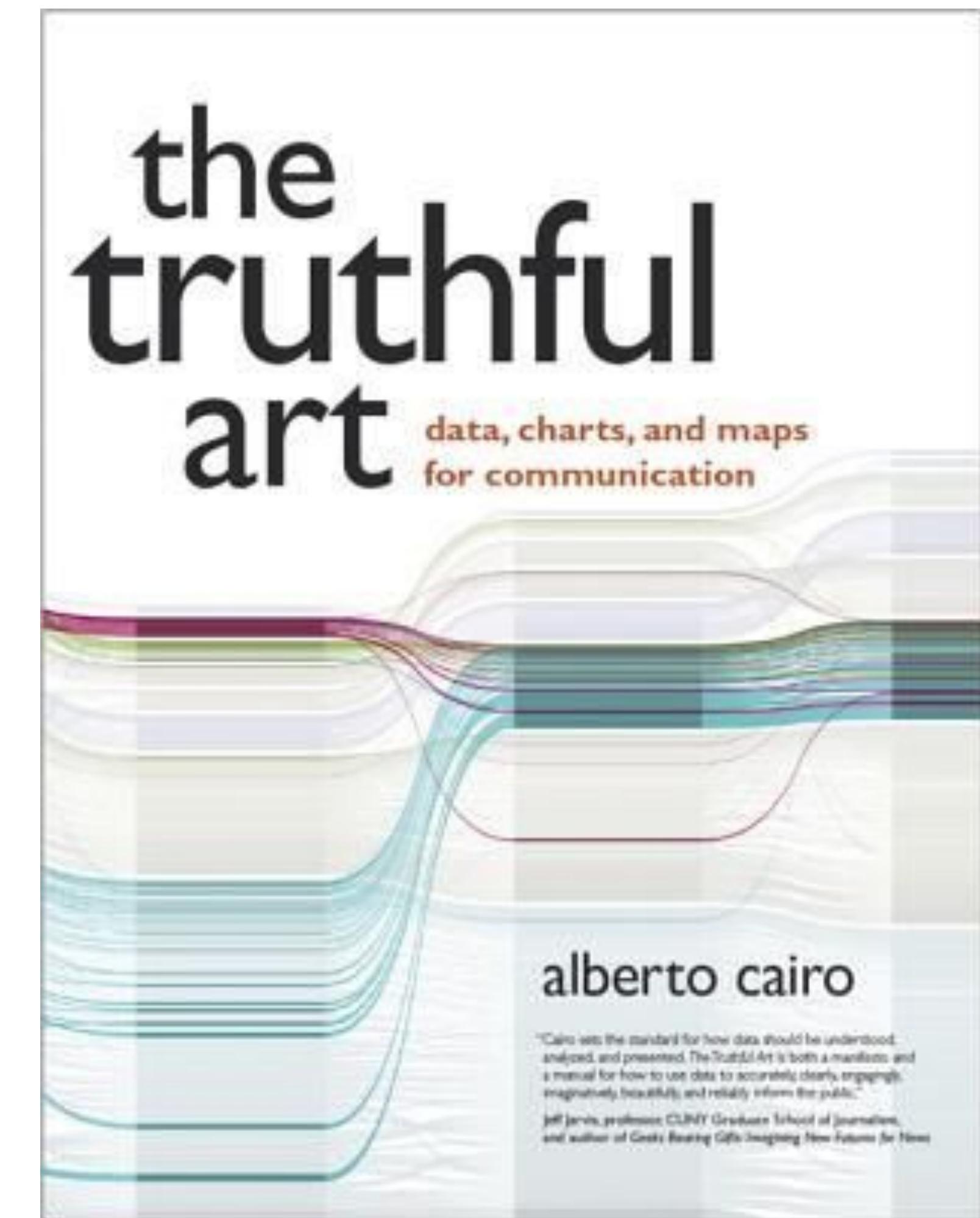
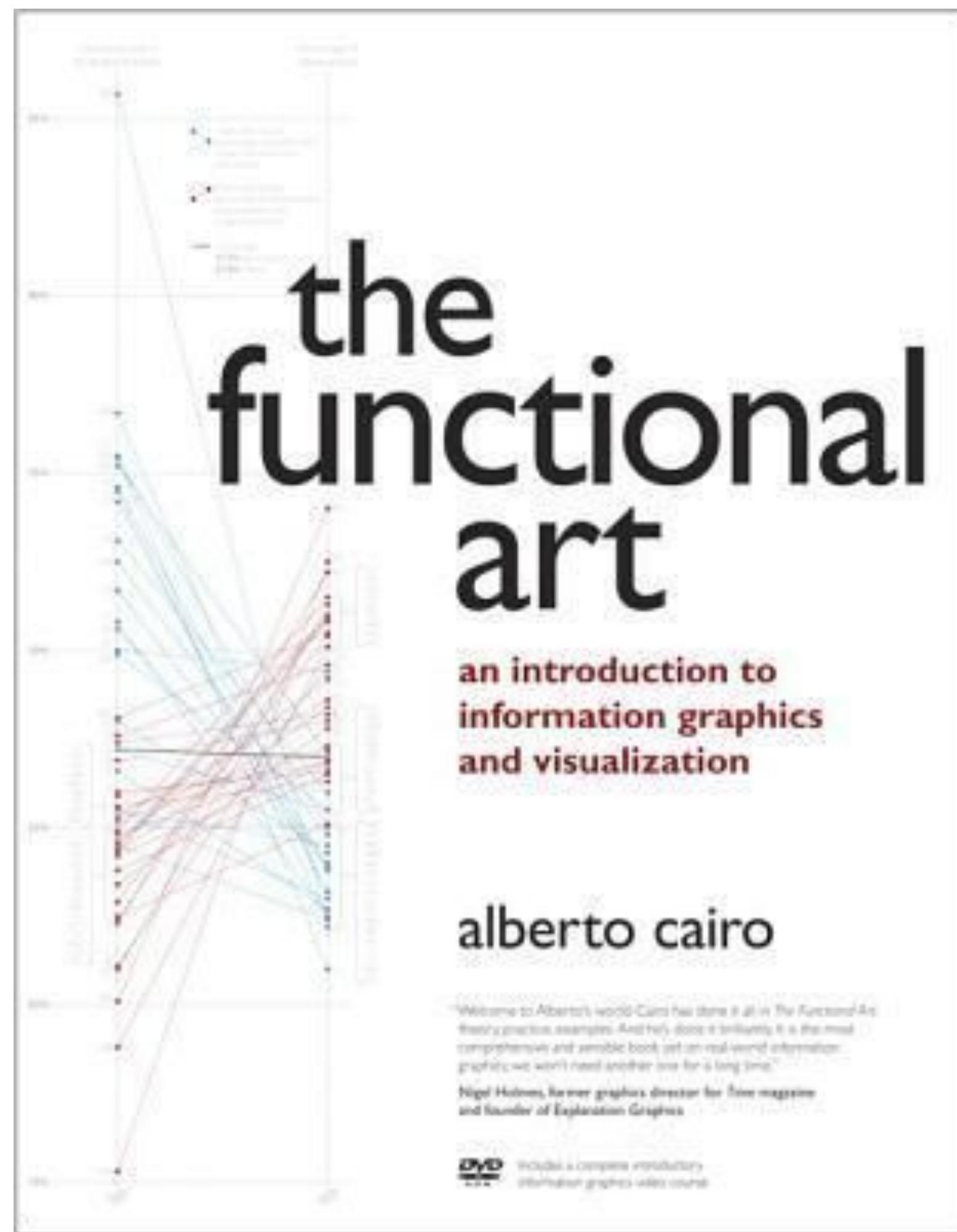
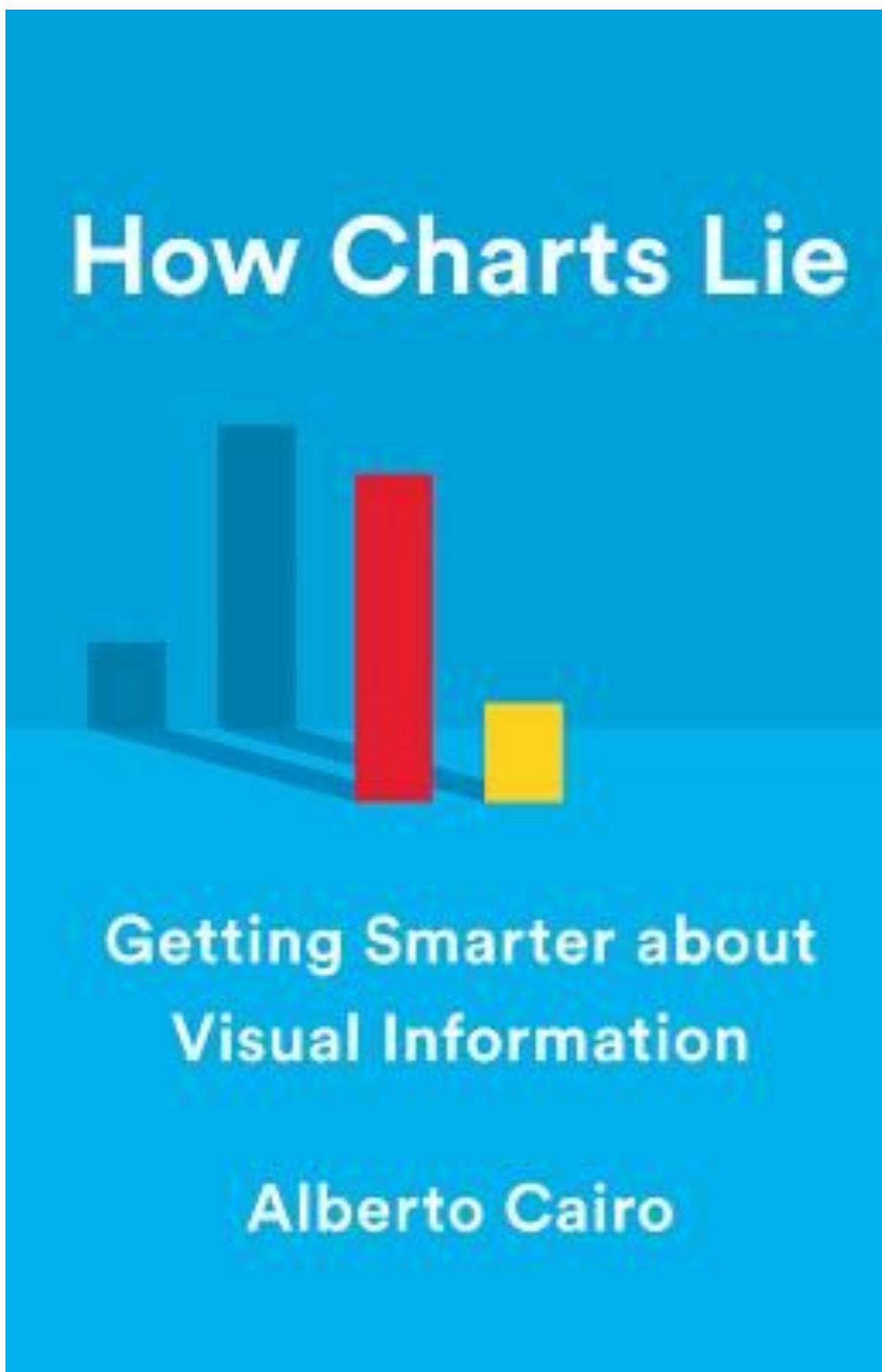
O'REILLY®

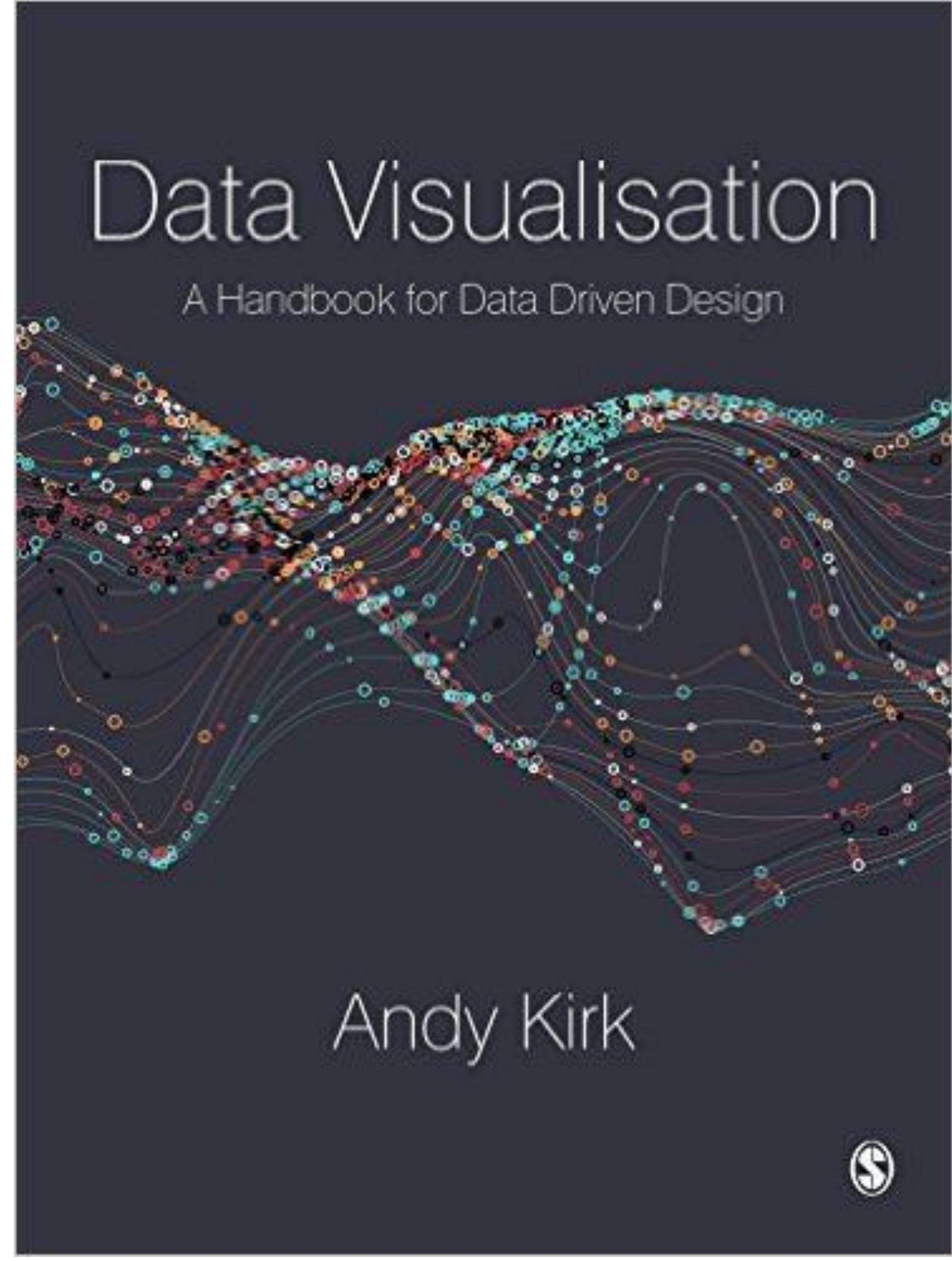
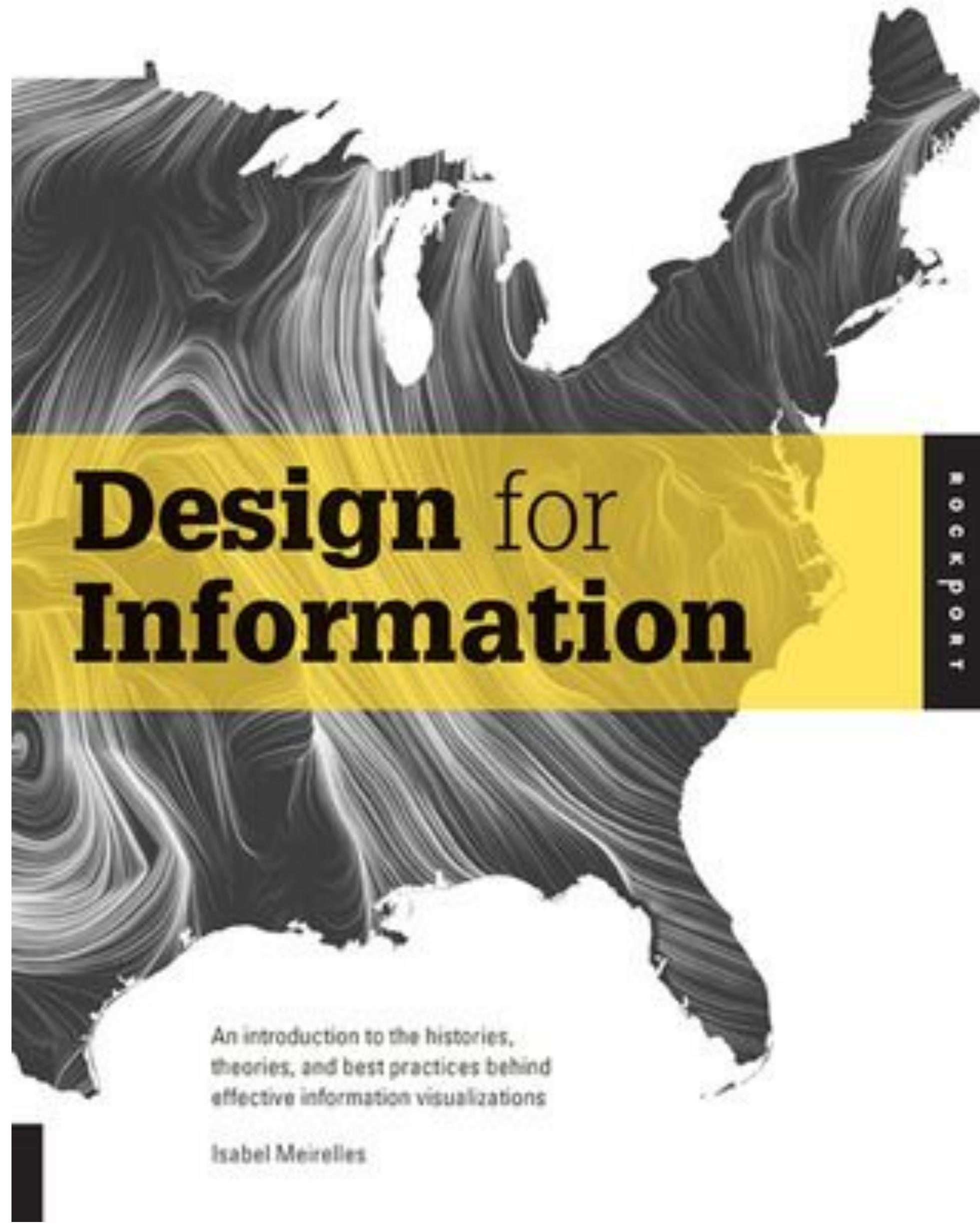
Fundamentals of Data Visualization

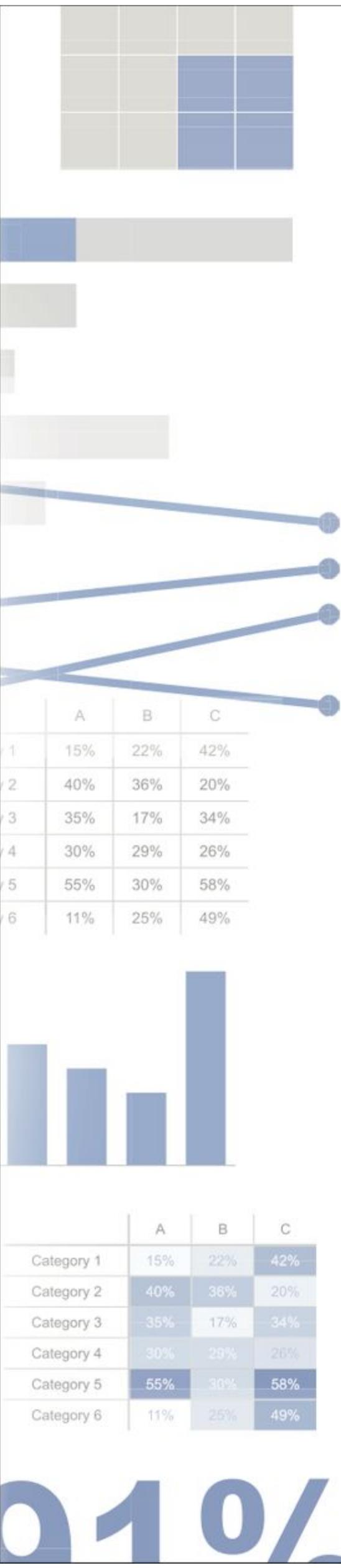
A Primer on Making Informative
and Compelling Figures



Claus O. Wilke







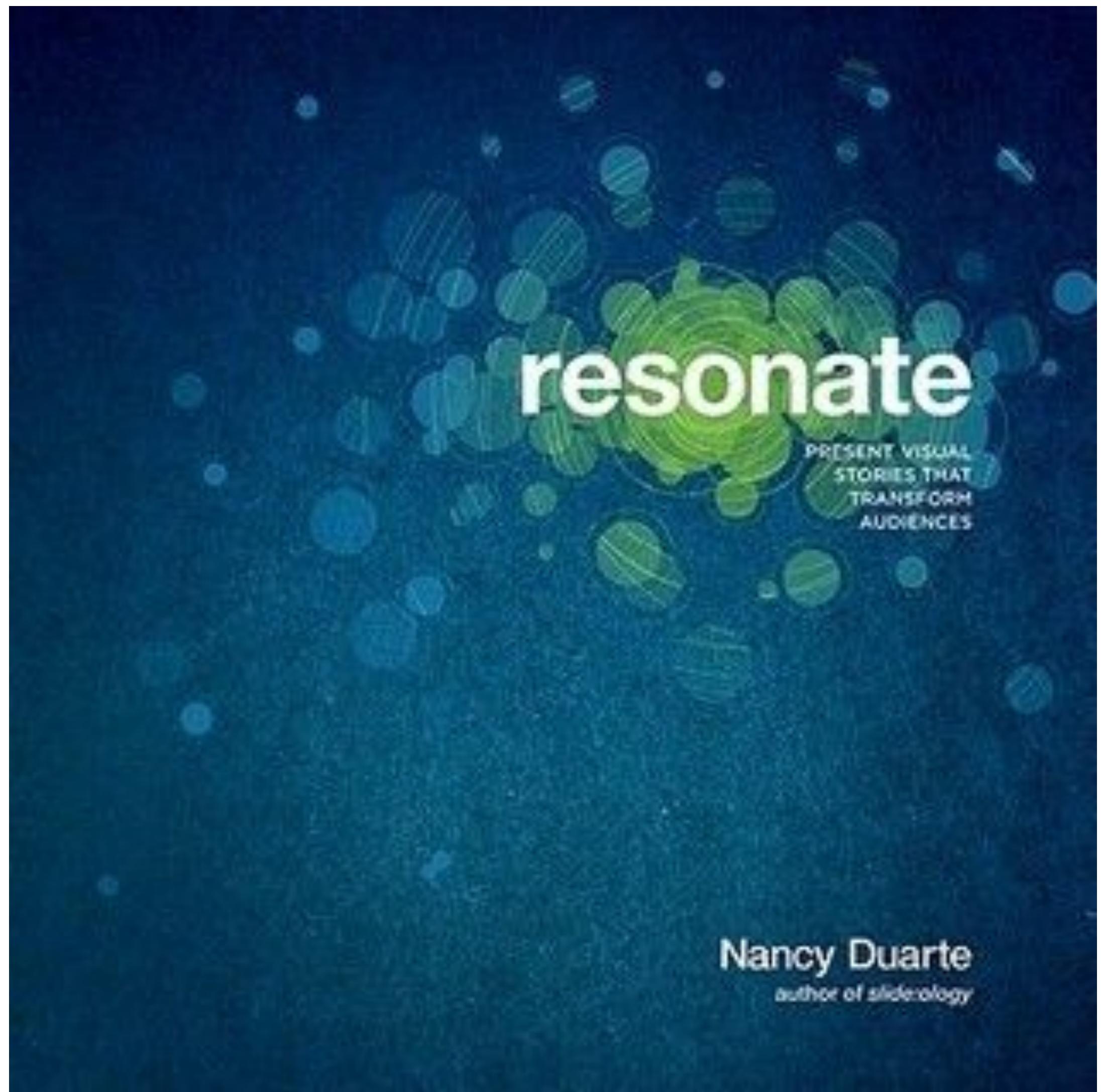
cole nussbaumer knaflic

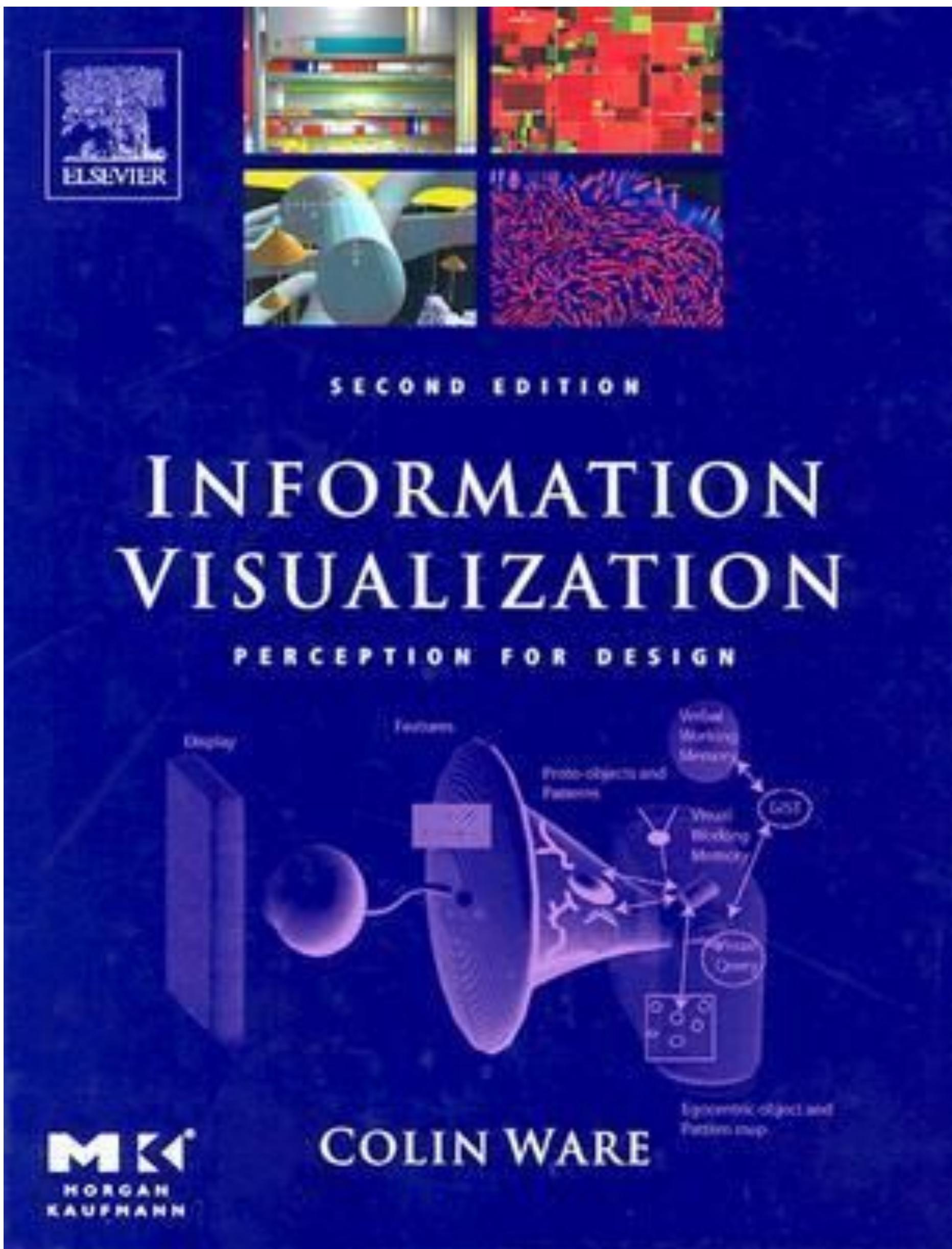
storytelling with data

a data
visualization
guide for
business
professionals

WILEY

010%





Narrative Visualization: Telling Stories with Data

Edward Segel and Jeffrey Heer

Abstract—Data visualization is regularly promoted for its ability to reveal stories within data, yet these “data stories” differ in important ways from traditional forms of storytelling. Storytellers, especially online journalists, have increasingly been integrating visualizations into their narratives, in some cases allowing the visualization to function in place of a written story. In this paper, we systematically review the design space of this emerging class of visualizations. Drawing on case studies from news media to visualization research, we identify distinct genres of narrative visualization. We characterize these design differences, together with interactivity and messaging, in terms of the balance between the narrative flow intended by the author (imposed by graphical elements and the interface) and story discovery on the part of the reader (often through interactive exploration). Our framework suggests design strategies for narrative visualization, including promising under-explored approaches to journalistic storytelling and educational media.

Index Terms—Narrative visualization, storytelling, design methods, case study, journalism, social data analysis.

1 INTRODUCTION

In recent years, many have commented on the storytelling potential of data visualization. News organizations including the New York Times, Washington Post, and the Guardian regularly incorporate dynamic graphics into their journalism. Politicians, activists, and television reporters use interactive visualizations as a backdrop for stories about global health and economics [10] and election results [9]. A recent feature in The Economist [6] explores the proliferation of digital data and notes that visualization designers are “*melding the skills of computer science, statistics, artistic design and storytelling.*”

Static visualizations have long been used to support storytelling, usually in the form of diagrams and charts embedded in a larger body of text. In this format, the text conveys the story, and the image typically provides supporting evidence or related details. An emerging class of visualizations attempts to combine narratives with interactive graphics. Storytellers, especially online journalists, are increasingly integrating complex visualizations into their narratives.

Crafting successful “data stories” requires a diverse set of skills. Gershon and Page [12] note that effective story-telling “*require[s] skills like those familiar to movie directors, beyond a technical expert’s knowledge of computer engineering and science.*” While techniques from oration, prose, comic books, video games, and film production are applicable to narrative visualization, we should also expect this emerging medium to possess unique attributes. Data stories differ in important ways from traditional storytelling. Stories in text and film typically present a set of events in a tightly controlled progression. While tours through visualized data similarly can be organized in a linear sequence, they can also be interactive, inviting verification, new questions, and alternative explanations.

Currently most sophisticated visualization tools focus on data ex-

In this paper, we investigate the design of narrative visualizations and identify techniques for telling stories with data graphics. We take an empirical approach, analyzing visualizations from online journalism, blogs, instructional videos, and visualization research. After reviewing related work, we share five selected case studies which highlight varied design strategies and illustrate our analytic approach. We then formulate a design space constructed from an analysis of 58 examples. Our analysis identifies salient dimensions of visual storytelling, including how graphical techniques and interactivity can enforce various levels of structure and narrative flow. We describe seven genres of narrative visualization: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and video. These genres can be combined with interactivity and messaging to produce varying balances of author-driven and reader-driven experiences. Finally, we discuss the implications of our framework, noting recurring design strategies, promising yet under-utilized approaches to integrating visualization with other media, and the potential for improved user interfaces for crafting data stories. By focusing on the graphical and interactive elements of narrative visualization, our approach gives less attention to the cognitive and emotional experience of the reader. We recognize the importance of these elements, however, and describe directions for future reader-centric research in our conclusion.

2 RELATED WORK

Storytelling and visual expression are integral parts of human culture; storytelling has even been referred to as “*the world’s second-oldest profession*” [12]. Without summarizing millennia of achievement, we describe a few of the key concepts informing narrative visualization.

Recursos

The Python Graph Gallery



Welcome to the Python Graph Gallery, a collection of hundreds of charts made with [Python](#). Charts are organized in about 40 sections and always come with their associated reproducible code. They are mostly made with [Matplotlib](#) and [Seaborn](#) but other library like [Plotly](#) are sometimes used. If you're new to python, this [online course](#) can be a good starting point.

Distribution



Violin



Density



Histogram



Boxplot



Ridgeline

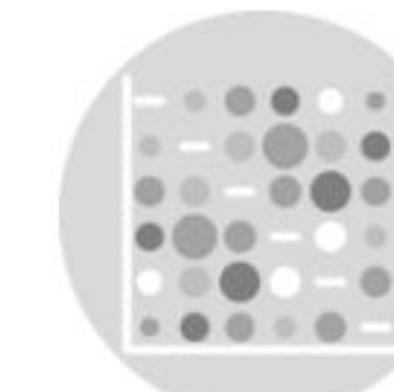
Correlation



Scatterplot



Heatmap



Correlogram



Bubble

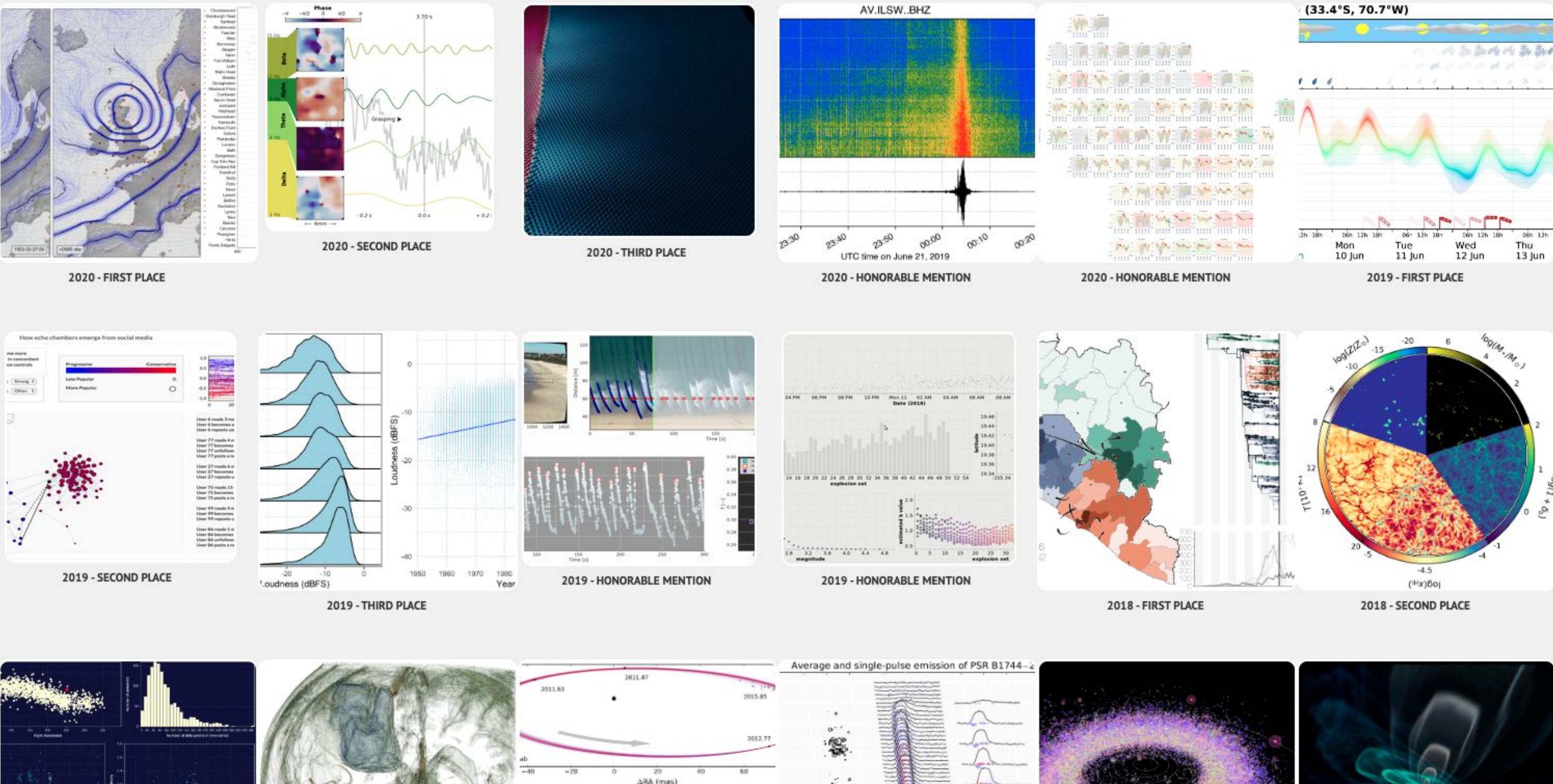


Connected Scatter



2D Density

Gallery



matplotlib

Cheat sheet Version 3.5.0

Quick start

```
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt

X = np.linspace(0, 2*np.pi, 100)
Y = np.cos(X)

fig, ax = plt.subplots()
ax.plot(X, Y, color='green')

fig.savefig("figure.pdf")
fig.show()
```

Anatomy of a figure

Subplots layout

```
subplot[s](rows,cols,...)
fig, axs = plt.subplots(3, 3)

G = gridspec(rows,cols,...)
ax = G[0,:]

ax.inset_axes(extent)

d=make_axes_locatable(ax)
ax = d.new_horizontal('10%')
```

Getting help

- [matplotlib.org](#)
- [github.com/matplotlib/matplotlib/issues](#)
- [discourse.matplotlib.org](#)
- [stackoverflow.com/questions/tagged/matplotlib](#)
- [gitter.im/matplotlib](#)
- [twitter.com/matplotlib](#)
- [Matplotlib users mailing list](#)

Basic plots

```
plot([X], Y, [fmt]...)
X, Y, fmt, color, marker, linestyle
scatter(X,Y,...)
X, Y, [sizes, [c]olors, marker, cmap]
bar[h](x,height,...)
x, height, width, bottom, align, color
imshow(Z,...)
Z, cmap, interpolation, extent, origin
contour[f]([X],[Y],Z,...)
X, Y, Z, levels, colors, extent, origin
pcolormesh([X],[Y],Z,...)
X, Y, Z, vmin, vmax, cmap
```

```
quiver([X],[Y],U,V,...)
X, Y, U, V, C, units, angles
pie(X,...)
Z, explode, labels, colors, radius
text(x,y,text,...)
x,y,text,va,ha,size,weight,transform
fill_[between][x]...
X, Y1, Y2, color, where
```

Advanced plots

```
step(X,Y,[fmt]...)
X, Y, fmt, color, marker, where
boxplot(X,...)
X, notch, sym, bootstrap, widths
errorbar(X,Y,xerr,yerr,...)
X, Y, xerr, yerr, fmt
hist(X, bins, ...)
X, bins, range, density, weights
violinplot(D,...)
D, positions, widths, vert
barbs([X],[Y], U, V, ...)
X, Y, U, V, C, length, pivot, sizes
eventplot(positions,...)
positions, orientation, lineoffsets
hexbin(X,Y,C,...)
X, Y, C, gridsize, bins
```

Scales

```
ax.set_[xy]scale(scale,...)
linear any values
symlog any values
log values > 0
logit 0 < values < 1
```

Projections

```
subplot(...,projection=p)
p='polar' p='3d'
p=Orthographic()
from cartopy.crs import Cartographic
```

Lines

```
linestyle or ls
----- ..- .-. : -.
capstyle or dash_capstyle
"butt" "round" "projecting"
```

Markers

o	O	s	p	x	*	p	D	<	>	^	v
.'	'o'	's'	'p'	'x'	'*'	'p'	'D'	'<'	'>'	'^'	'v'
'1'	'2'	'3'	'4'	'5'	'6'	'7'					
\$\\$	\$\\$	\$\\$	\$\\$	\$\\$	\$\\$	\$\\$	\$\\$	\$\\$	\$\\$	\$\\$	\$\\$

markervery
10 [0,-1] (25,5) [0,25,-1]

Colors

C0	C1	C2	C3	C4	C5	C6	C7	C8	C9
b	g	r	c	m	y	k	w		
darkRed	firebrick	crimson	indianred	salmon					
(1,0,0)	(1,0,0,0.75)	(1,0,0,0.5)	(1,0,0,0.75)						
#FF00008B	#FF00008B	#FF00008B	#FF00008B	#FF00008A					
0.6	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	1.0
X, Y									

Colormaps

```
plt.get_cmap(name)

Uniform
viridis
magma
plasma

Sequential
Greys
YlOrBr
Wistia

Diverging
Spectral
coolwarm
RdGy

Qualitative
tab10
tab20

Cyclic
twilight
```

Tick locators

```
from matplotlib import ticker
ax.[xy]axis.set_[minor|major]_locator(locator)
ticker.NullLocator()

ticker.MultipleLocator(0.5)
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0

ticker.FixedLocator([0, 1, 5])
0 1 5

ticker.IndexLocator(base=0.5, offset=0.25)
0.25 0.75 1.25 1.75 2.25 2.75 3.25 3.75 4.25 4.75

ticker.AutoLocator()
0 1 2 3 4 5

ticker.MaxNLocator(n=4)
0.0 1.5 3.0 4.5

ticker.LogLocator(base=10, numticks=15)
10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9 10^10
```

Tick formatters

```
from matplotlib import ticker
ax.[xy]axis.set_[minor|major]_formatter(formatter)
ticker.NullFormatter()

ticker.FixedFormatter(['zero', 'one', 'two', '-'])
zero one two three four five

ticker.FuncFormatter(lambda x, pos: "%2.2f" % x)
0.00 1.00 2.00 3.00 4.00 5.00

ticker.FormatStrFormatter('>%d<')
>0< >1< >2< >3< >4< >5<

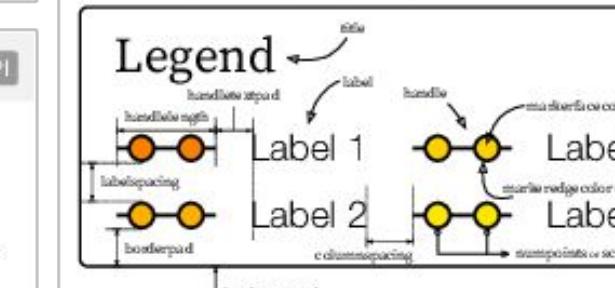
ticker.ScalarFormatter()
0 1 2 3 4 5

ticker.StrMethodFormatter('{x}')
0.0 1.0 2.0 3.0 4.0 5.0

ticker.PercentFormatter(xmax=5)
0% 20% 40% 60% 80% 100%
```

Ornaments

```
ax.legend(...)
handles, labels, loc, title, frameon
```



Colorbar

```
ax.colorbar(...)
mappable, ax, cax, orientation
```

```
ax.annotate(...)
text, xy, xytext, xycoords, textcoords, arrowprops
```

```
text
xytext
textcoords
xy
xycoords
```

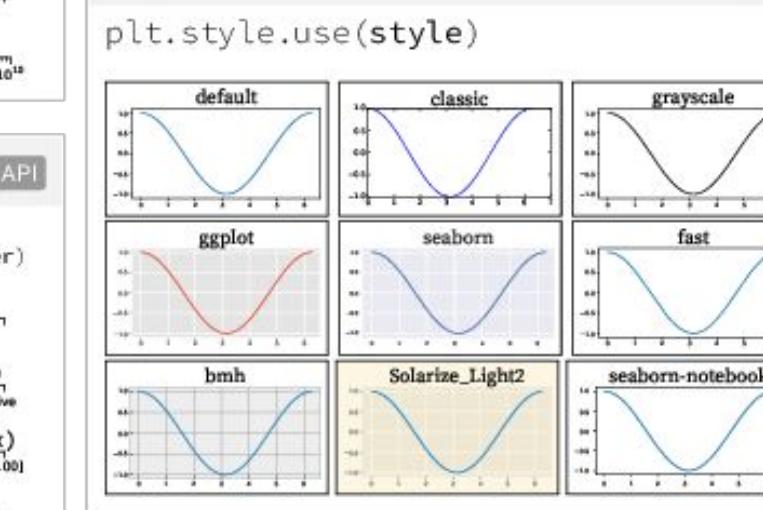
Event handling

```
fig, ax = plt.subplots()
def on_click(event):
    print(event)
fig.canvas.mpl_connect('button_press_event', on_click)
```

Animation

```
import matplotlib.animation as mpl
T = np.linspace(0, 2*np.pi, 100)
S = np.sin(T)
line, = plt.plot(T, S)
def animate(i):
    line.set_ydata(np.sin(T+i/50))
anim = mpl.FuncAnimation(
    plt.gcf(), animate, interval=5)
plt.show()
```

Styles



Quick reminder

```
ax.grid()
ax.set_[xy]lim(vmin, vmax)
ax.set_[xy]label(label)
ax.set_[xy]ticks(ticks, [labels])
ax.set_[xy]ticklabels(labels)
ax.set_title(title)
ax.tick_params(width=10, ...)
ax.set_axis_[on|off]()
```

```
fig.suptitle(title)
fig.tight_layout()
plt.gcf(), plt.gca()
mpl.rc('axes', linewidth=1, ...)
[fig|ax].patch.set_alpha(0)
text=r'$\frac{-e^{i\pi}}{2^n}$'
```

Keyboard shortcuts

ctrl+s	Save	ctrl+w	Close plot
r	Reset view	f	Fullscreen 0/1
f	View forward	b	View back
p	Pan view	o	Zoom to rect
x	X pan/zoom	y	Y pan/zoom
g	Minor grid 0/1	G	Major grid 0/1
l	X axis log/linear	L	Y axis log/linear

Ten simple rules

1. Know Your Audience
2. Identify Your Message
3. Adapt the Figure
4. Captions Are Not Optional
5. Do Not Trust the Defaults
6. Use Color Effectively
7. Do Not Mislead the Reader
8. Avoid "Chartjunk"
9. Message Trumps Beauty
10. Get the Right Tool

¿Dudas?

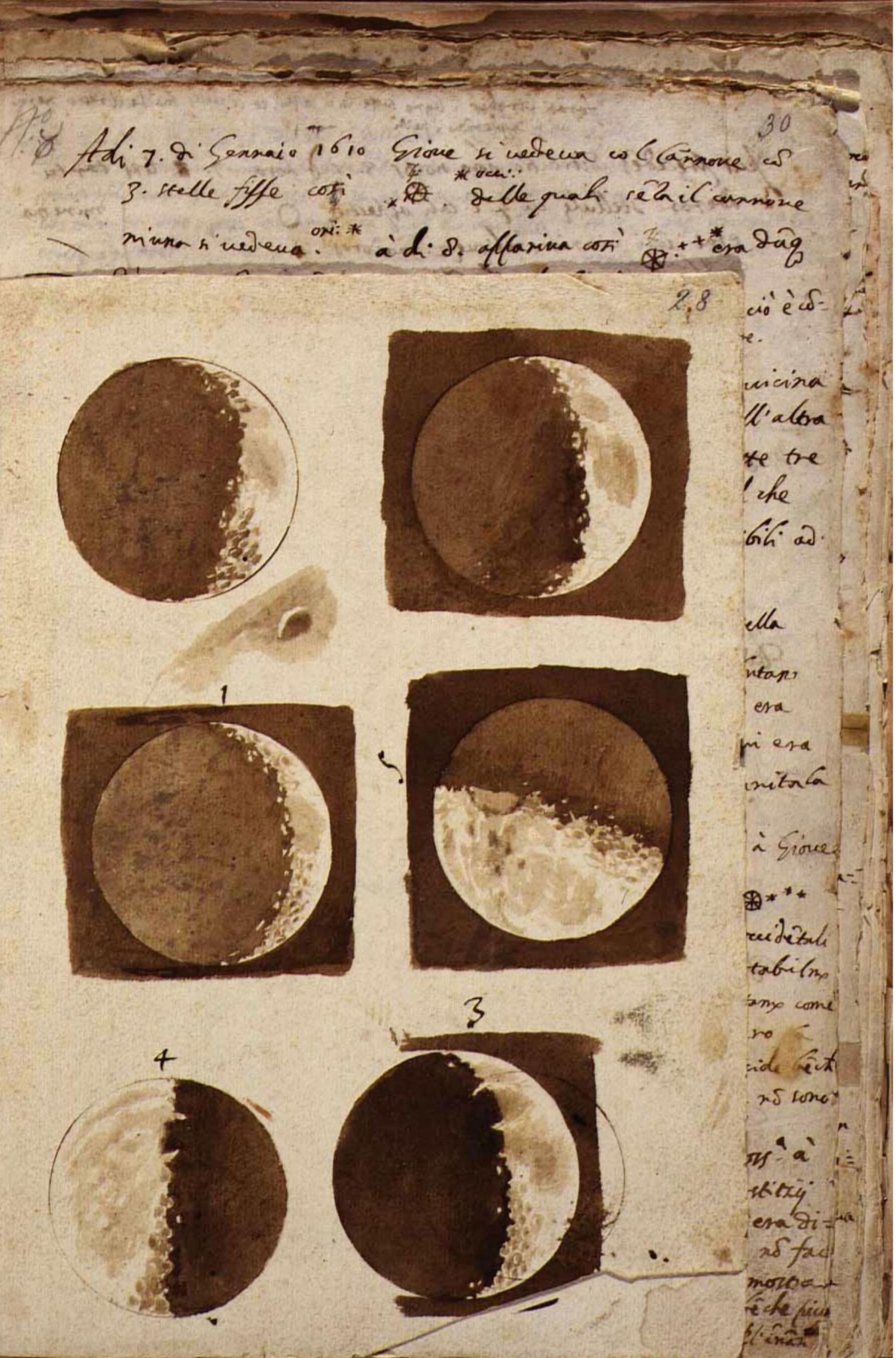
Visualización de datos

Visualización

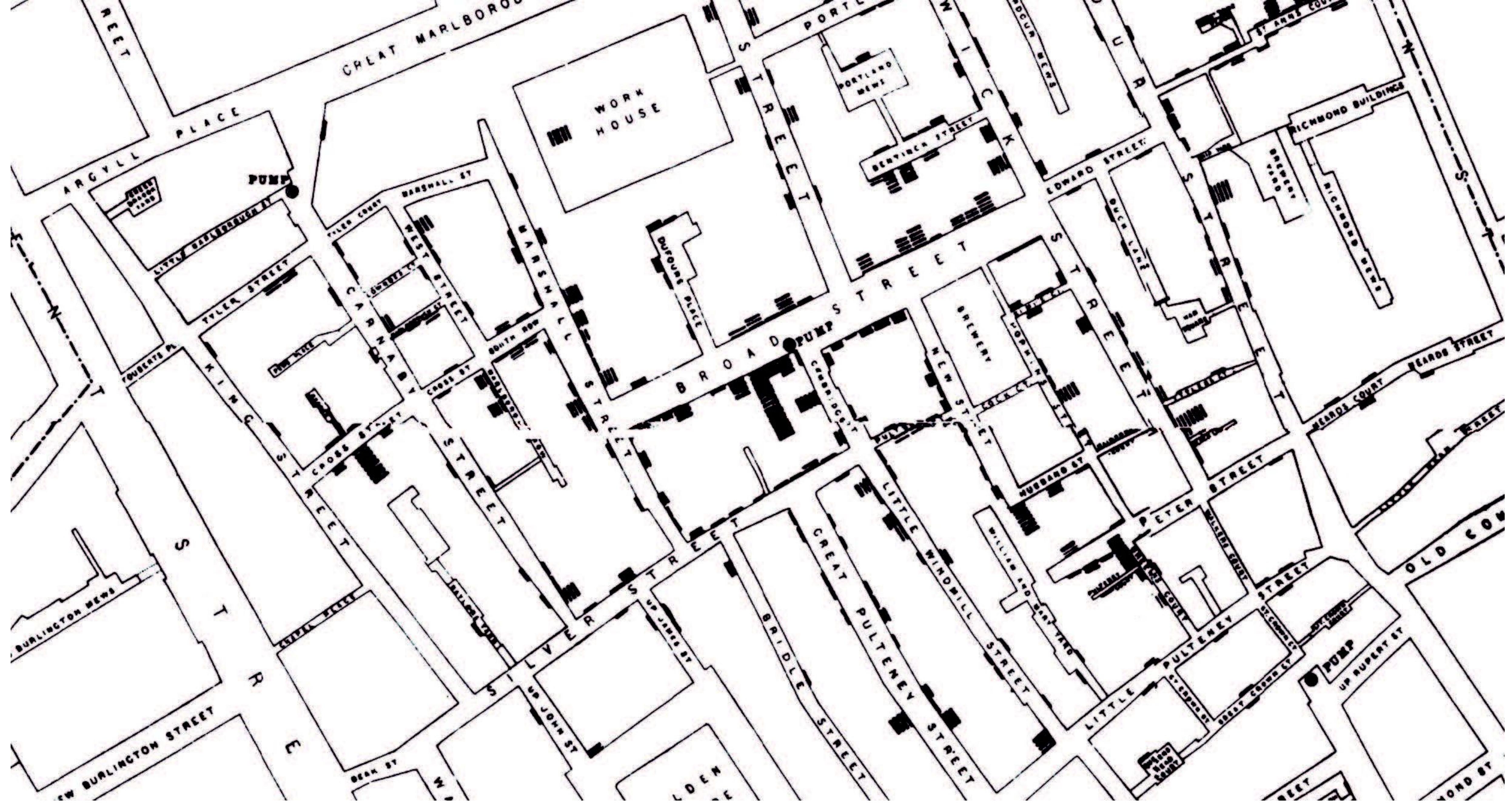
1. Formación de imágenes mentales de forma visual
2. El acto o proceso de interpretar en términos visuales o poner en forma visible para transmitir información a través de la representación gráfica de los datos

Objetivos de la visualización

1. *Guardar o grabar* información
2. *Analizar (explorar)* datos
3. *Comunicar (explicar)* a otr@s



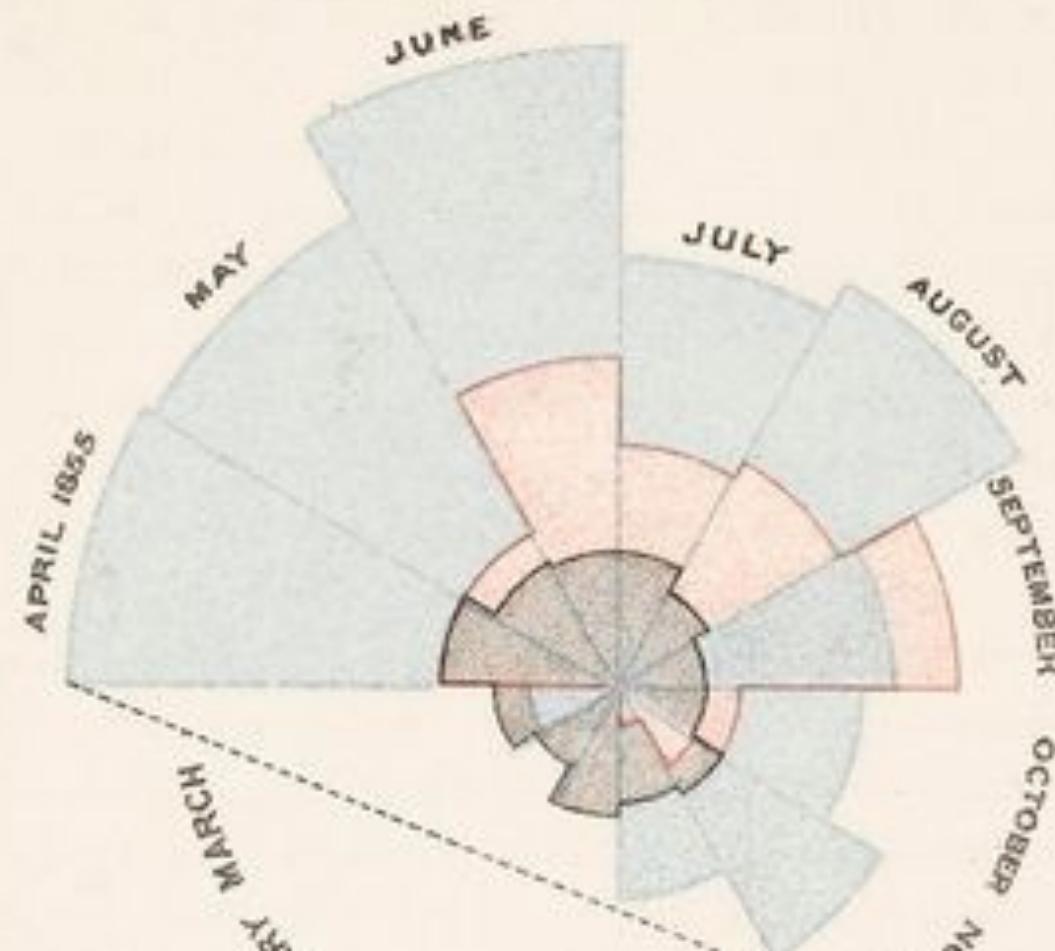
Galileo Galilei, 1610



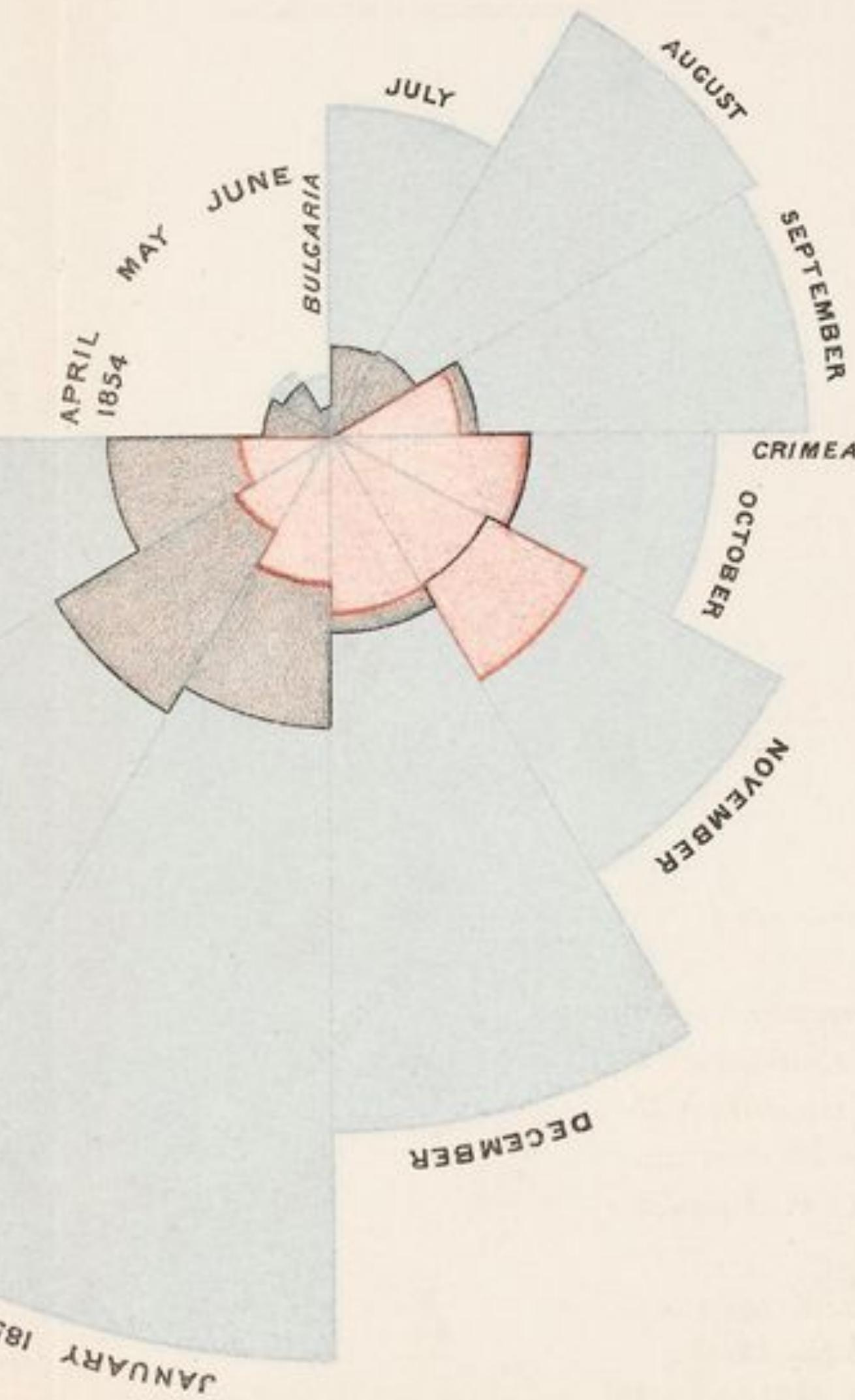
John Snow, 1854

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1855, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.
Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite

Paris, le 20 Novembre 1869

Les nombres d'hommes perdus sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui ont été en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dessiner la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séguir, de Fezenac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent Oroscha en Witlobek, avaient toujours marché avec l'armée.

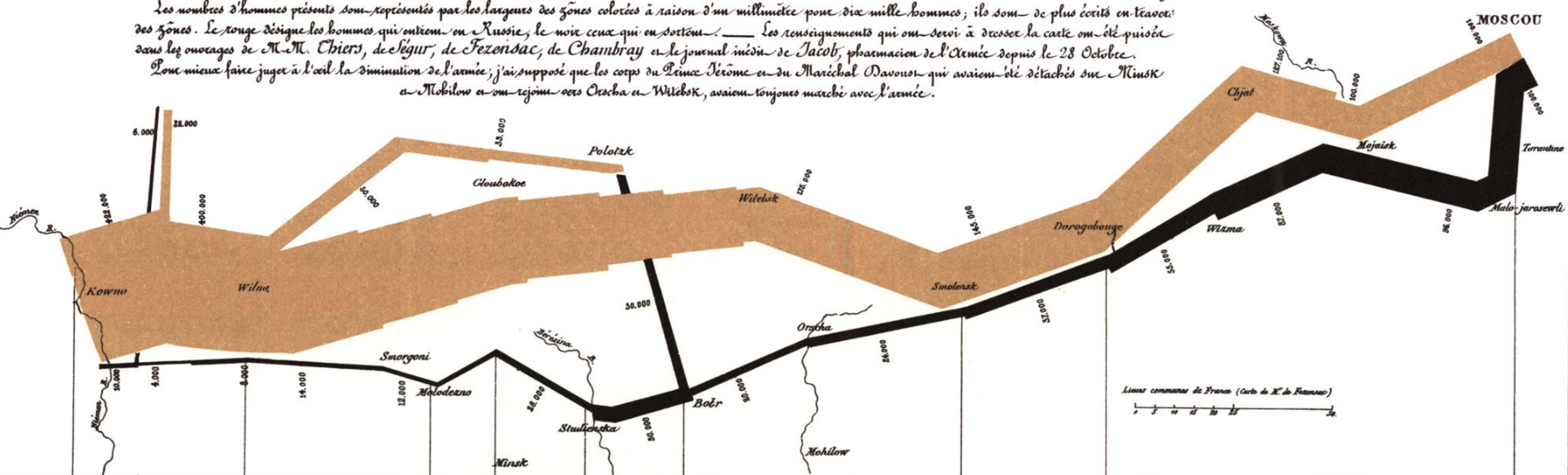
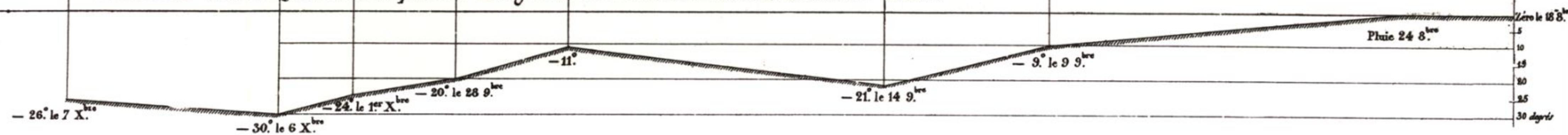


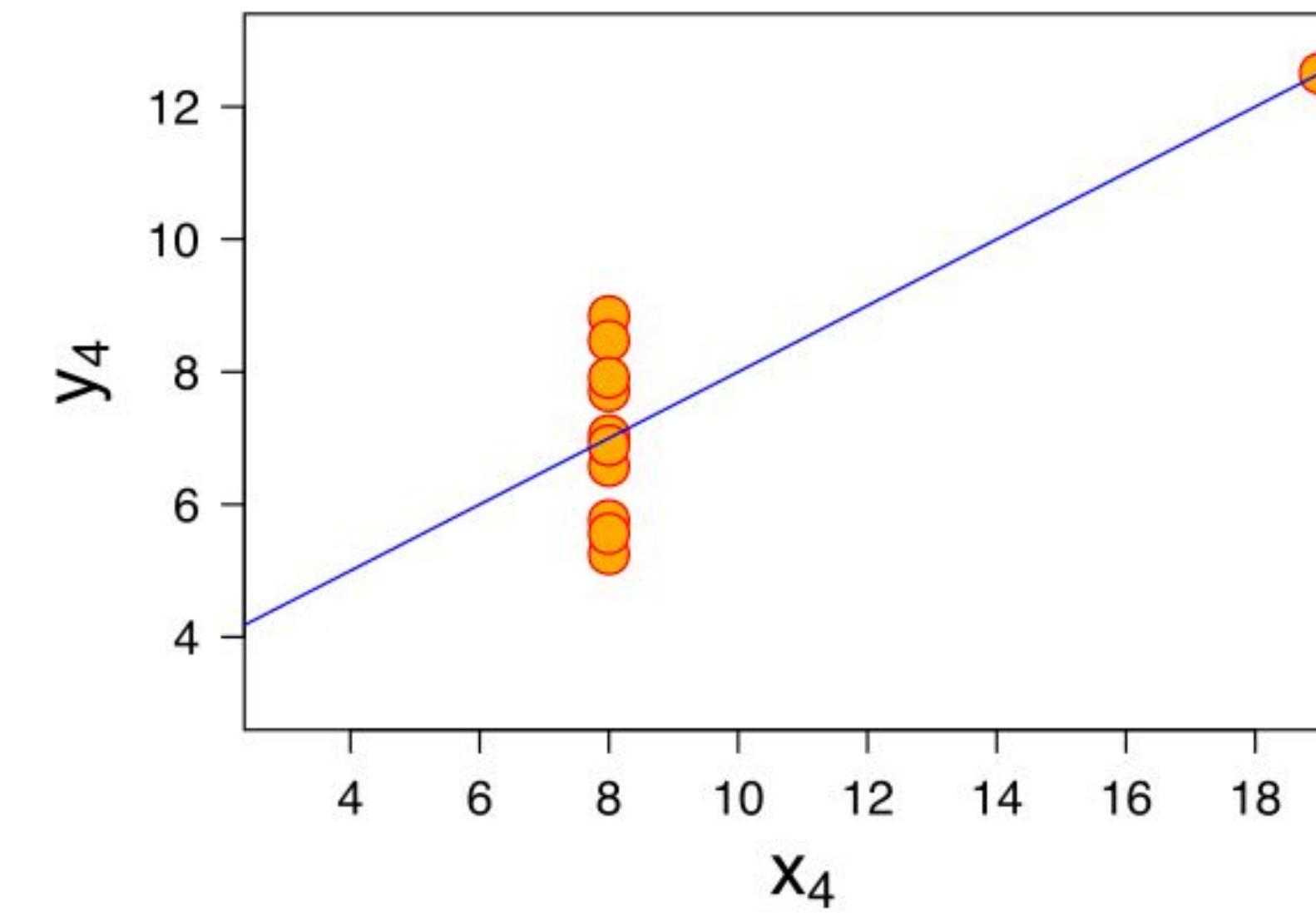
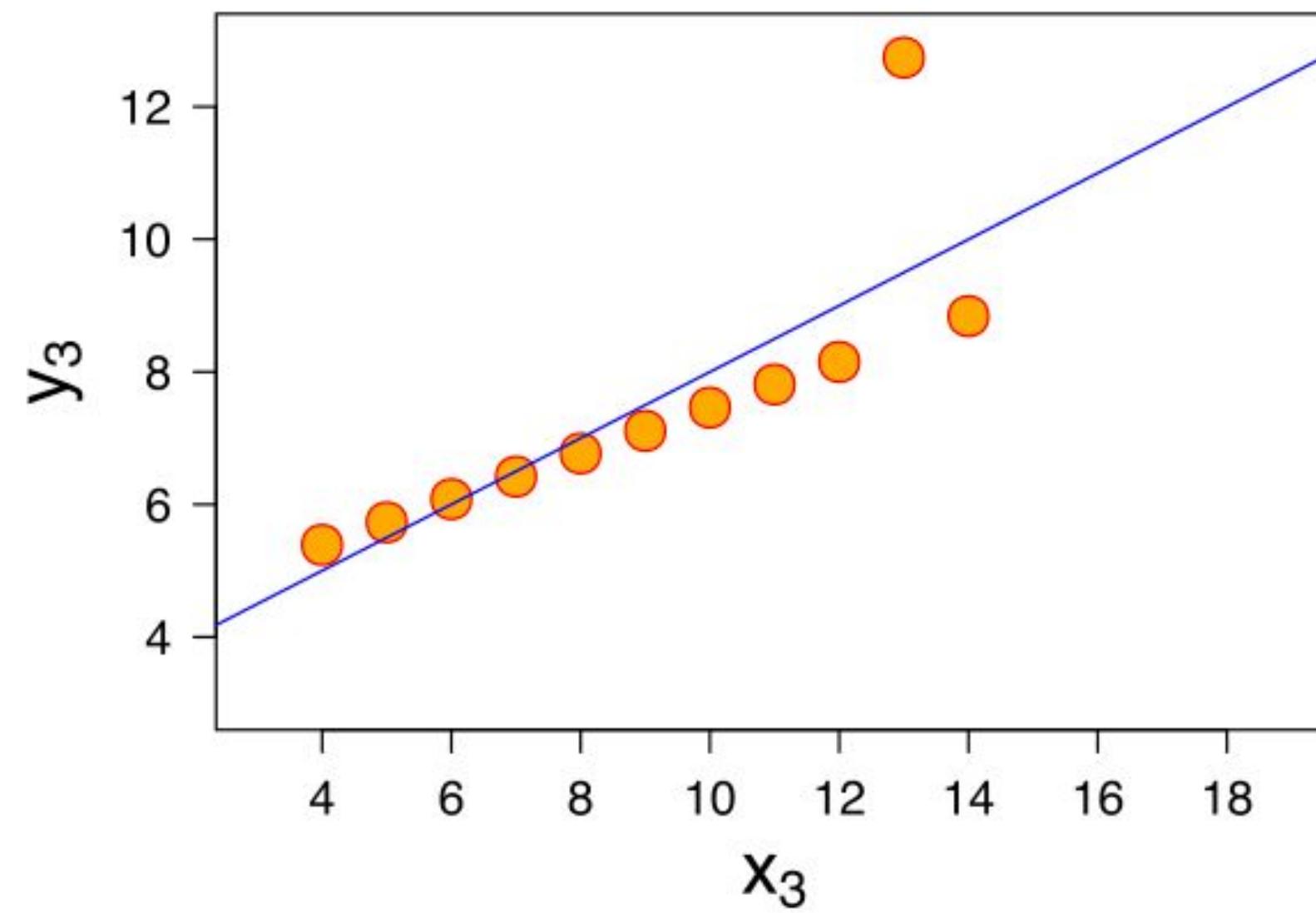
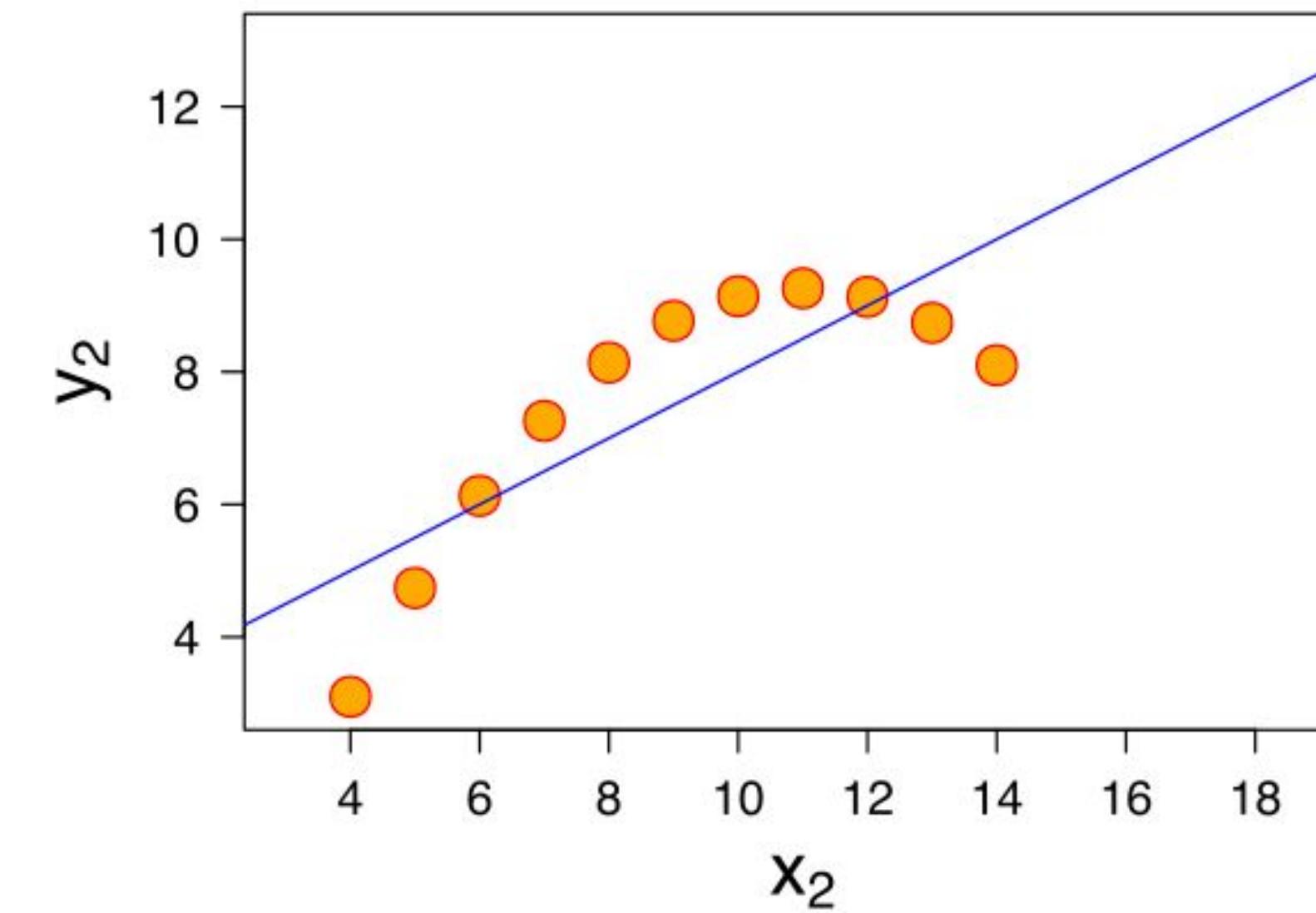
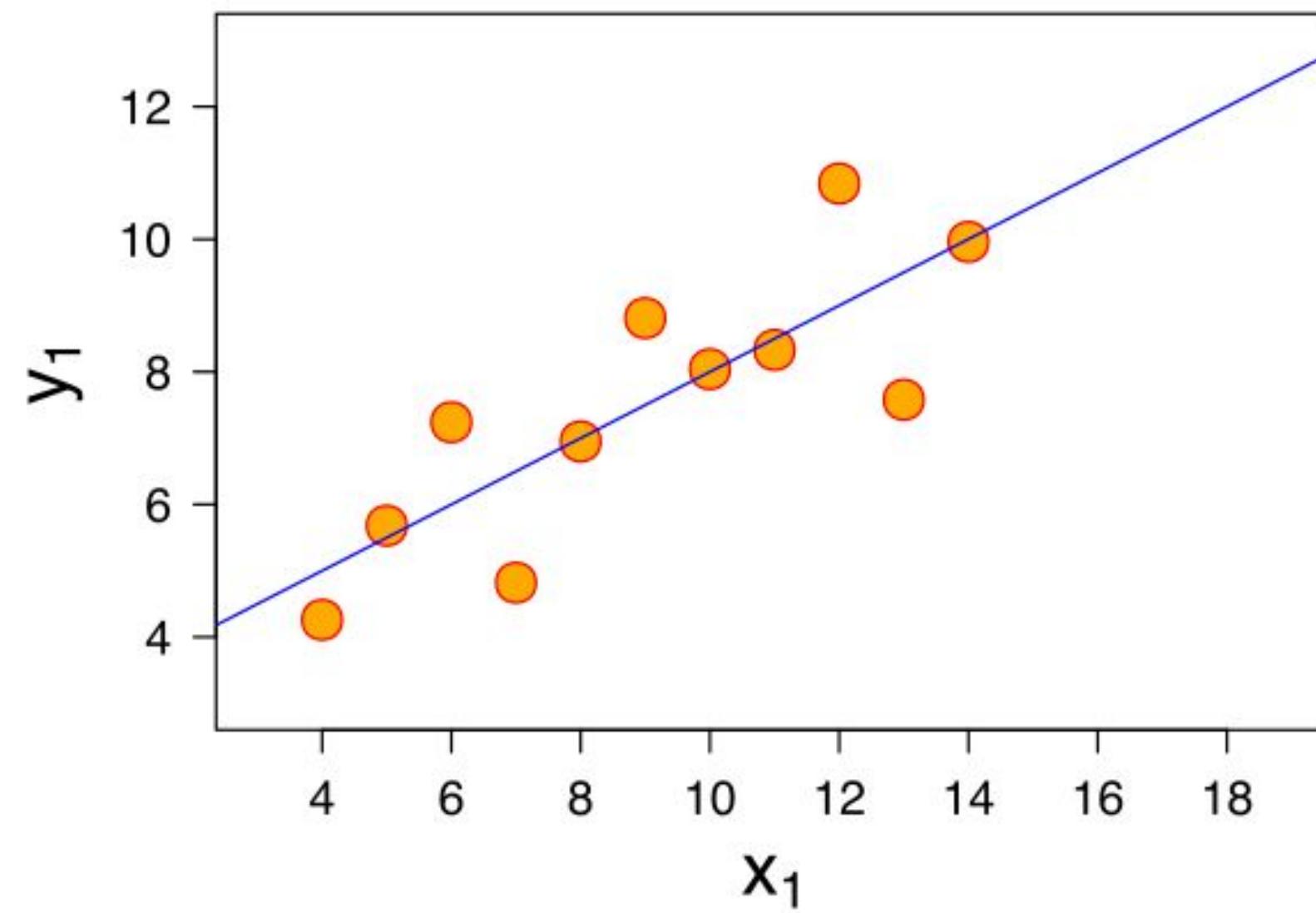
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

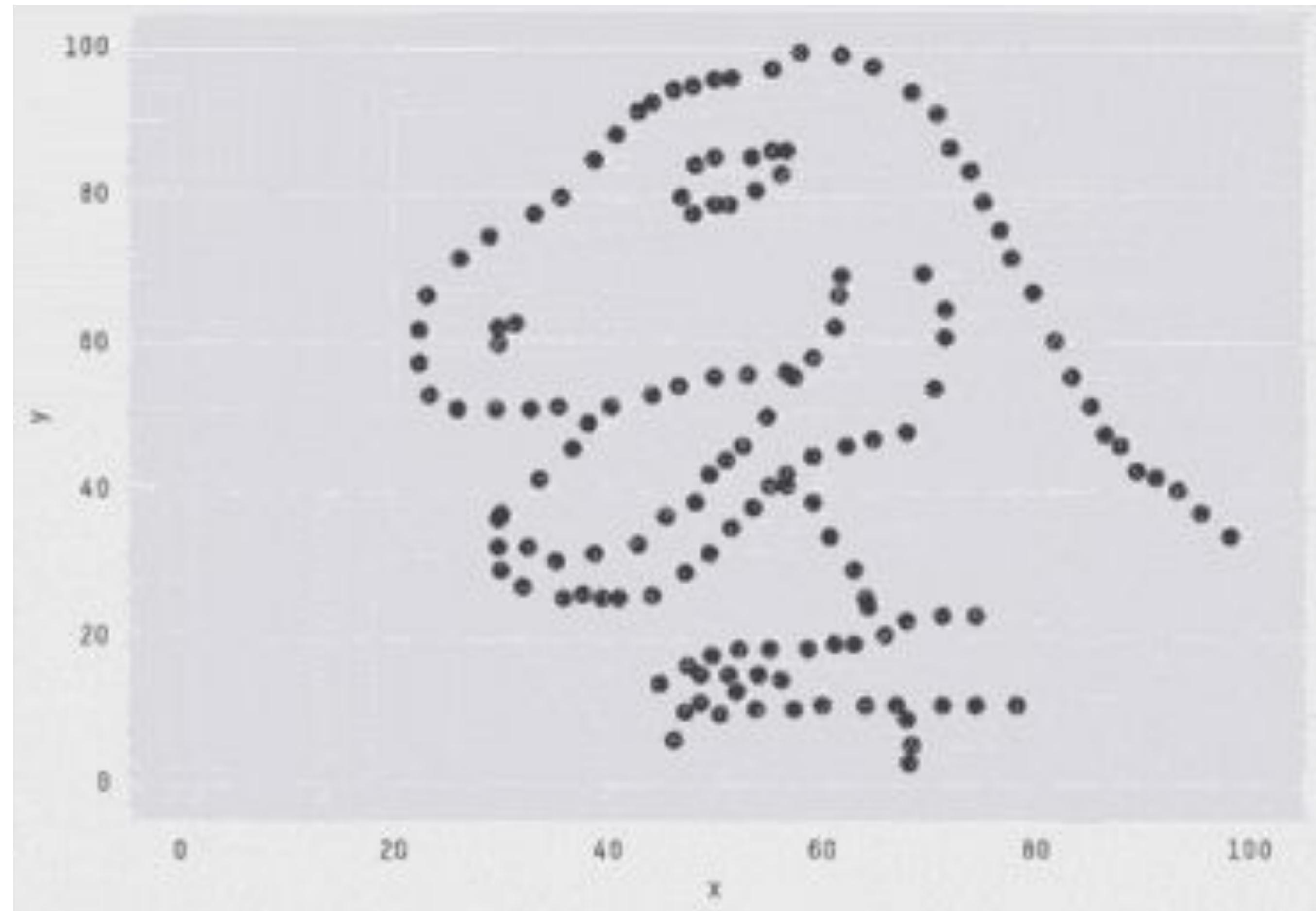
Les Cosaques passent au galop
le Niemen gelé.



¿Por qué visualizar los datos?

	Set A		Set B		Set C		Set D	
	X	Y	X	Y	X	Y	X	Y
0	10	8.04	10	9.14	10	7.46	8	6.58
1	8	6.95	8	8.14	8	6.77	8	5.76
2	13	7.58	13	8.74	13	12.74	8	7.71
3	9	8.81	9	8.77	9	7.11	8	8.84
4	11	8.33	11	9.26	11	7.81	8	8.47
5	14	9.96	14	8.10	14	8.84	8	7.04
6	6	7.24	6	6.13	6	6.08	8	5.25
7	4	4.26	4	3.10	4	5.39	19	12.50
8	12	10.84	12	9.13	12	8.15	8	5.56
9	7	4.82	7	7.26	7	6.42	8	7.91
10	5	5.68	5	4.74	5	5.73	8	6.89
mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
std	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
corr	0.82		0.82		0.82		0.82	
lin. reg.	$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$	





X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Diseño

Actividad

Crea al menos 3 formas de visualizar estas dos cantidades:

42

23



42
23

23
42

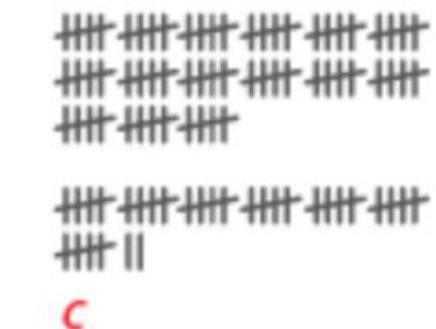
↓ ↓ →

Fijación de diseño

Adherencia ciega a un conjunto de ideas o conceptos

1. writing, number notation

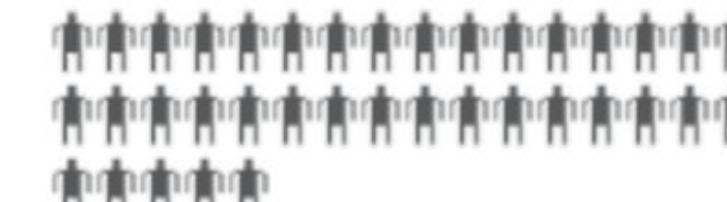
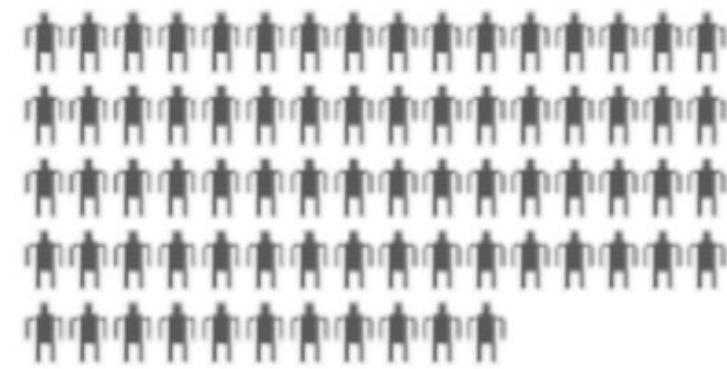
75, 37  



2. squares



3. repeated icon



4. hundreds, tens, units, decimals...

represented by squares



5. bars



6. line graph

28. gray tones



29. color scale



30.

geometric proportions

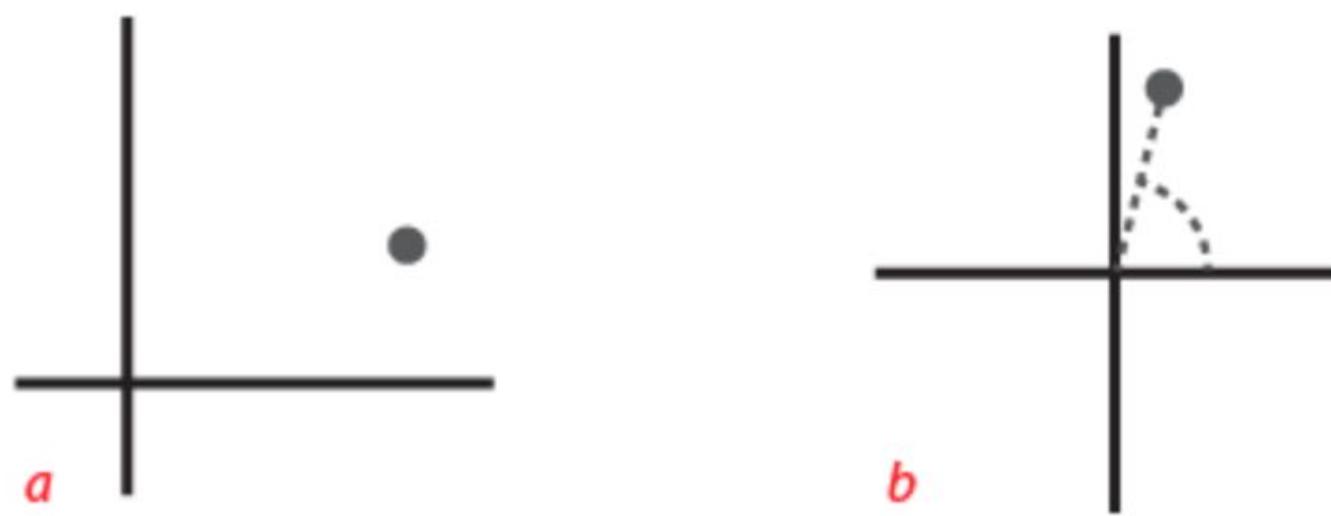


31. horizontal/vertical proportions



Useful if the first value represents a horizontal distance, and the second value a vertical one.

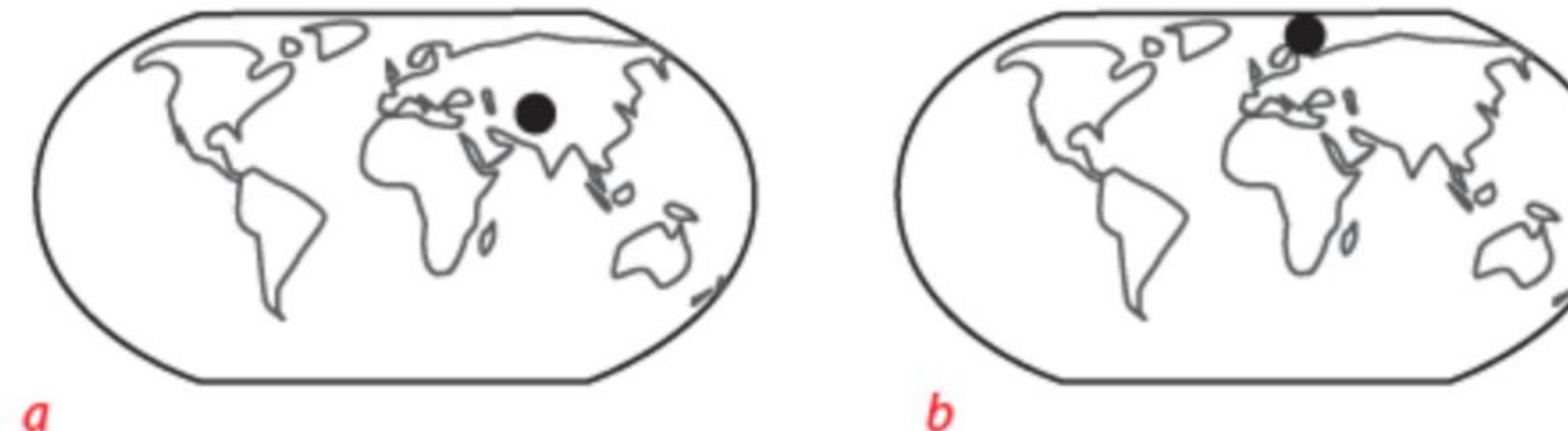
32. coordinates



a cartesian and polar

<https://rockcontent.com/blog/45-ways-to-communicate-two-quantities/>

34. geographic coordinates

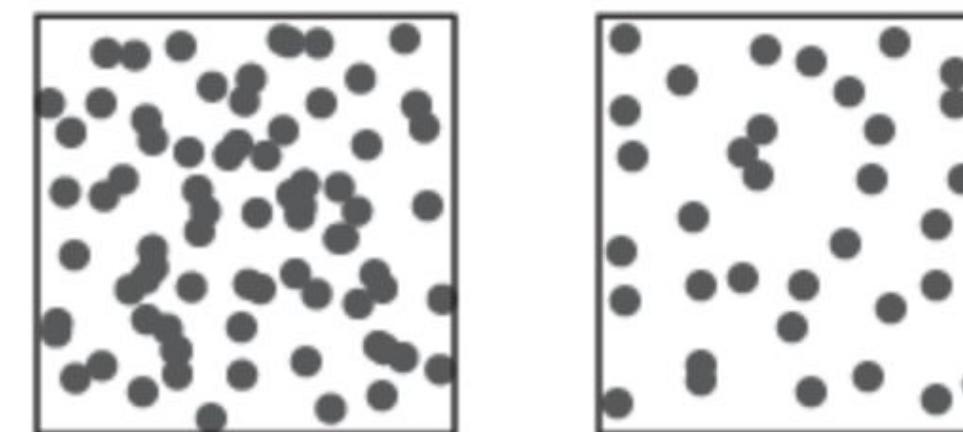


a longitude, latitude *b* latitude, longitude

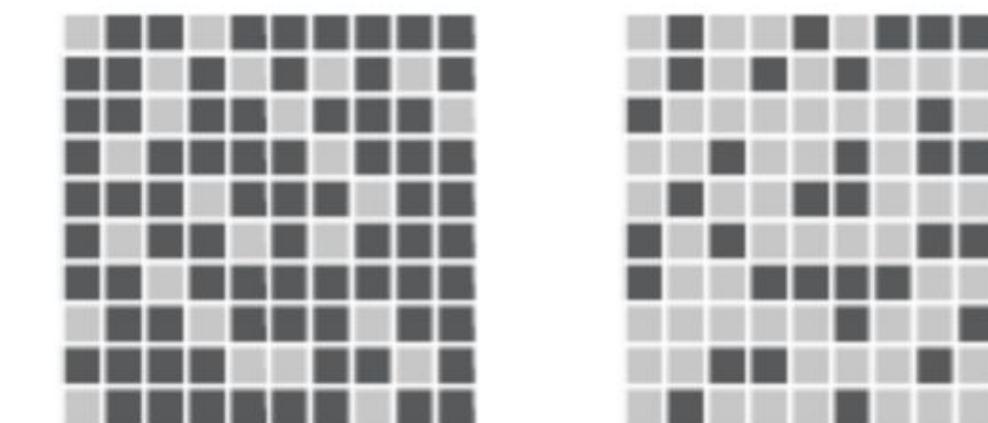
35. values associated to countries



36. density



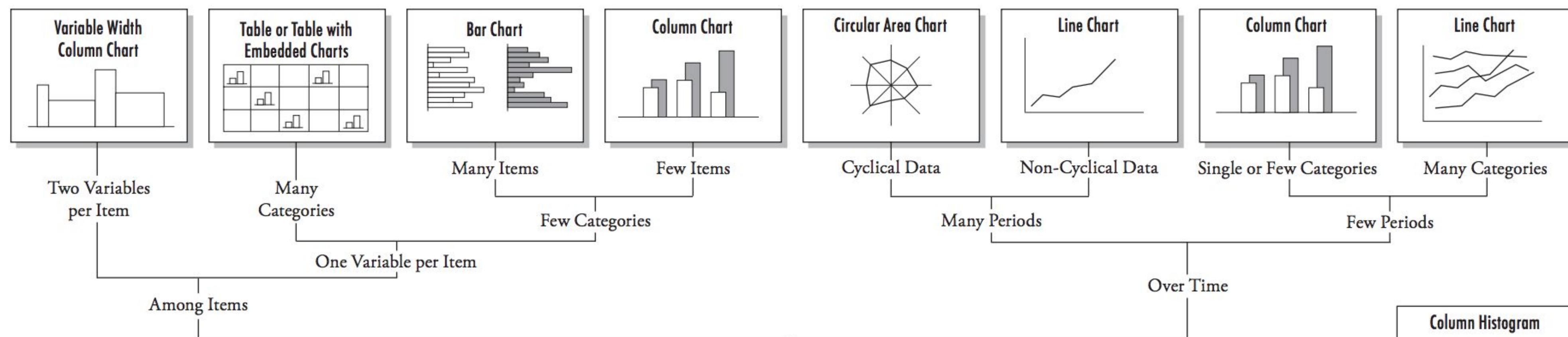
37. percentages / density



Combination of values

<https://rockcontent.com/blog/45-ways-to-communicate-two-quantities/>

Tipos de gráficos



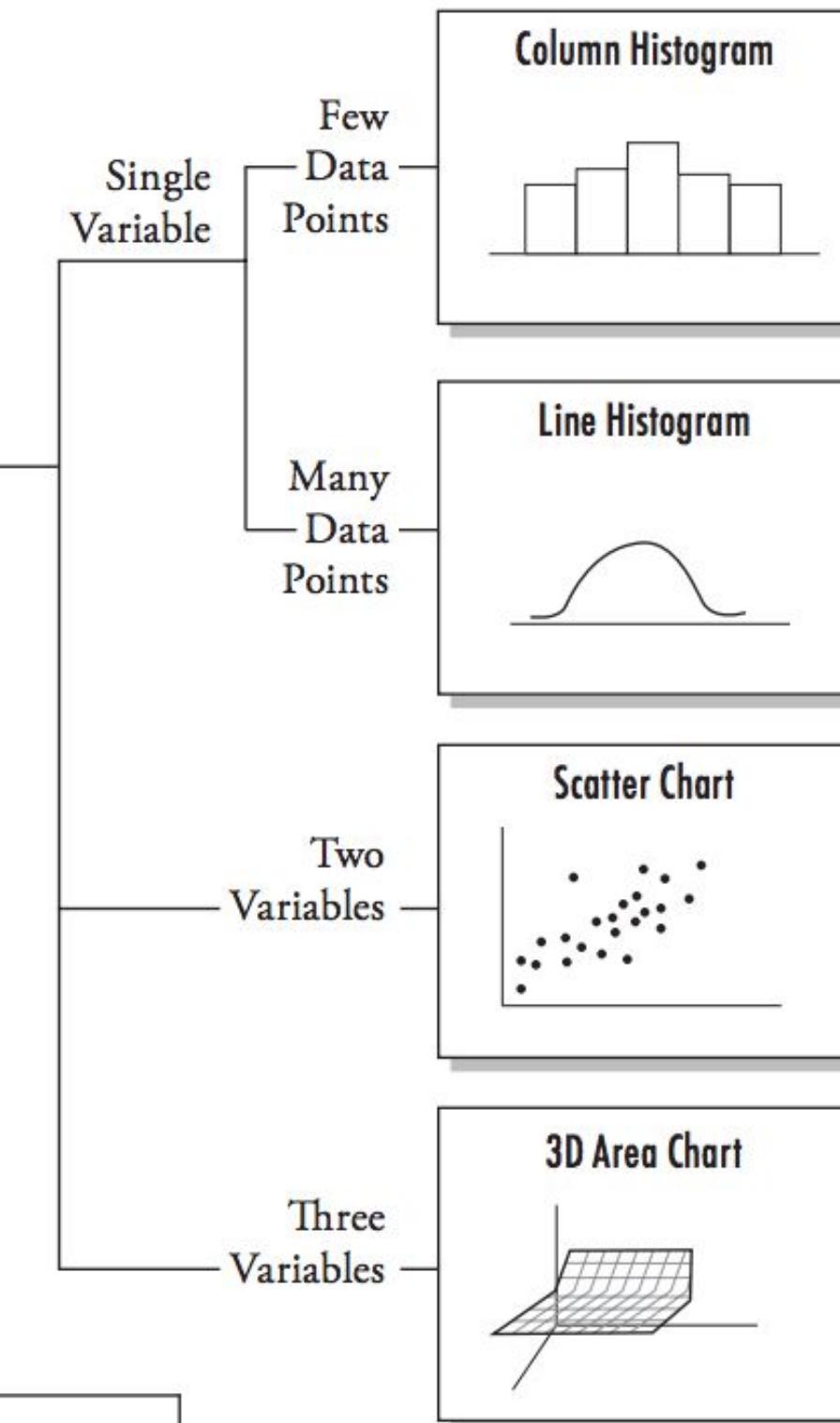
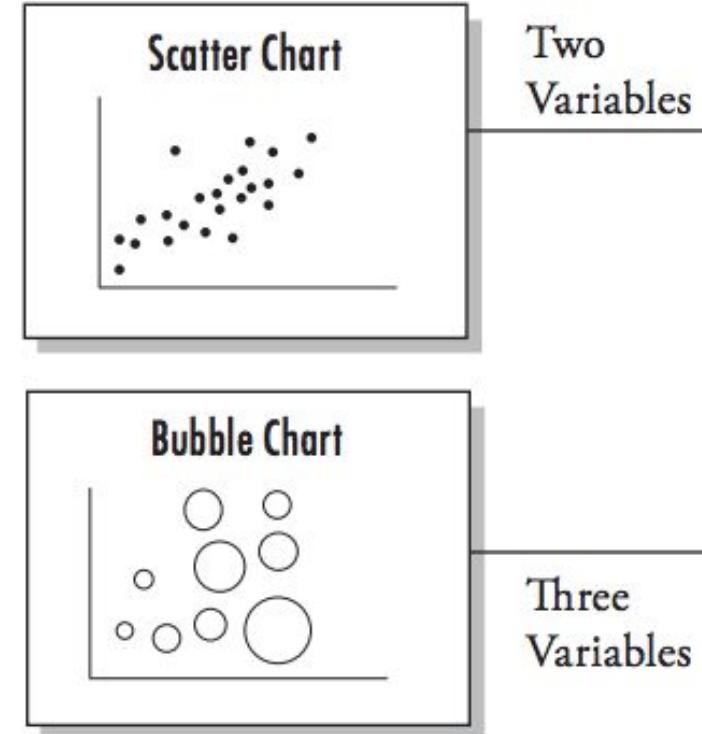
Comparison

Relationship

What would you like to show?

Distribution

Composition



Deviation

Emphasise deviations from a fixed reference point. Typically, a reference point is zero but it can also be a target or a long-term average. Can also be used to show asymmetries (posit/negatively).

Example FT uses:
Trade surplus/deficit; climate change

Correlation

Show the relationship between two or more variables. In general, if two tell them other than that, most readers will assume that one causes the other.

Example FT uses:
Inflation and unemployment; income and life expectancy

Ranking

Show where an item's position in an ordered list is most important than its absolute or relative values. Don't be afraid to highlight the points of interest.

Example FT uses:
World cup/league tables; constituency election results

Distribution

Show values in a dataset and how often they occur. The closer the data distribution can be to a normal curve, the less 'outliers' or inequality there is in the data.

Example FT uses:
Income distribution; population (age/gender) distribution; revealing inequality

Change over Time

Show emphasis to changing trends. There can be short-term (e.g. daily) movements or extended series (increasing/decreasing over time). Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses:
Share price movements; economic time series; seasonal changes in a market

Magnitude

Show size comparisons. These can be relative (not being able to see large/larger or absolute to need to see the differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or percent.

Example FT uses:
Company production, market capitalisation, volumes in general

Part-to-whole

Adds up one box or map only used when parts/elements are geographically close. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses:
Fiscal budgets; company structures; national/sectoral results

Spatial

Show the vector volumes of流量 of things that move across states or countries. These might be goods, services or geographical locations.

Example FT uses:
Population density; natural resources; locations; nature; cluster; dissimilarity; catchment areas; migration; election results

Flow

Show the vector volumes of流量 of things that move across states or countries. These might be goods, services or geographical locations.

Example FT uses:
Migrants; trade; migrants; locations; information; relationships

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Ordered bar

Standard bar chart showing the rank order of values much more easily when sorting into order.

Histogram

The standard way to show a statistical distribution - keep the gaps between columns small to highlight the shape of the curve.

Line

The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column

The standard way to compare the size of three items. Must always start at 0 on the axis.

Stacked column/bar

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Basic choropleth (ratio)

The standard approach for ratio data on a map - should always be ratios rather than totals and use a sensible base geography.

Sankey

Show changes in flow from one condition to another and also for tracing the overall outcome of a complex process.

Diverging stacked bar

Perfect for presenting survey results which involve sentiment (e.g. disagree/really agreed).

Column + line timeline

A good way of showing the relationship between an amount (column) and a rate (line).

Ordered column

See above.

Dot plot

A simple way of showing the change or range (minimum or maximum) of data across multiple categories.

Dot strip plot

Good for showing individual values, e.g. distribution can be a problem when too many dots have the same value.

Column + line timeline

A good way of showing change over time, especially for rates, between an amount (column) and a rate (line).

Bar

See above. Good when the data are not time series but usually best with only one series of data at a time.

Marimekko

A good way of showing the size and proportion of data at the same time - as long as the data are not too complicated.

Proportional symbol (count/magnitude)

Use for totals rather than rates - be wary that small differences in data will be hard to see.

Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Spine

Splits a single value into two contrasting components (e.g. profit/loss).

Connected scatterplot

Usually used to show how the relationship between two variables has changed over time.

Ordered proportional symbol

See above.

Dot strip plot

Good for showing individual values, e.g. distribution can be a problem when too many dots have the same value.

Barcode plot

Like dot strip plots, good for displaying all the data in a table. They work best when highlighting individual values.

Slope

Good for showing changing patterns over time as long as the data can be split into 2 or 3 points without missing a key part of the trend.

Paired column

As per standard column but allows for multiple series. Can become messy to read with more than 3 series.

Pie

A non-linear way of showing parts of whole data - but be aware that it's difficult to accurately compare the size of the segments.

Flow map

For showing ambiguous movement across areas.

Chord

A complex but powerful diagram which can illustrate 2-way flows (and net values) in a metric.

Surplus/deficit filled line

The shaded area of these dots shows a balance to be shown either against a baseline or between two series.

Bubble

Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY heatmap

A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in contours.

Slope

Perfect for showing how marks have changed over time or vary between categories.

Lollipop

Lollipops draw more attention to the data value than standard boxplots and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates. For large datasets, consider grouping first using colour.

Boxplot

Summarise multiple distributions by showing the median (center) and range of the data.

Violin plot

Similar to a box plot but more effective with complex distributions (data cannot be summarised with simple averages).

Candlestick

Leads to focus on day-to-day activity; these charts show opening/closing and high/low points of each day.

Area chart

Use with care - these are good at showing changes to total, but poor at showing components as they are very difficult.

Marimekko

A good way of showing the size and proportion of data at the same time - as long as the data are not too complicated.

Treemap

Use for hierarchical data - good for showing changes to total shape - good for representing varying regions with equal values.

Equalised cartogram

Converting each unit on a map to a regular and equal-sized shape - good for representing varying regions with equal values.

XY heatmap

A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in contours.

Slope

Perfect for showing how marks have changed over time or vary between categories.

Population pyramid

A standard way for showing the age and sex breakdown of a population distribution effectively, back to back histograms.

Cumulative curve

A good way of showing how much distribution a y-axis is always cumulative. Frequency x-axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Fan chart (projections)

Use to show the uncertainty in future projections - this grows the further towards projection.

Connected scatterplot

A good way of showing changing data for two categories whenever there is a relatively clear pattern of progression.

Lollipop

Lollipops draw more attention to the data value than standard boxplots - doesn't have to start at zero (but preferable).

Radar

A space-efficient way of showing values of multiple variables but make sure they are organised in a way that makes sense to the reader.

Parallel coordinates

An alternative to radar charts - again, the organisation of the variables is important. Usually benefits from highlighting values.

Beeswarm

Used to emphasise individual points in a dataset. Points can be sized to an additional variable. Best with medium-sized datasets.

Calendar heatmap

A great way of showing temporal patterns (daily, weekly, monthly) at the expense of showing precision in detail.

Priestley timeline

Great when date and duration are key elements of the story in the data.

Circle timeline

Good for showing discrete sizes of varying sizes across multiple categories (e.g. households by continent).

Vertical timeline

Places time on the Y axis, good for displaying detailed time series that work especially well when scaling on both axes.

Sedismogram

Another alternative to the circle timeline for showing timelines where there are big variations in the data.

Streamgraph

A type of area chart; use when seeing changes in proportions over time is more important than individual values.

Gridplot

Good for showing 2 dimensions, the grid best when used on a grid-based layout form.

Venn

Generally only used for schematic representation.

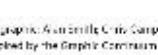
Heat map

Grid-based data values mapped to an area or colour scale. As choropleth maps - can be shaped to an administrative unit.

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.



ft.com/vocabulary

FT

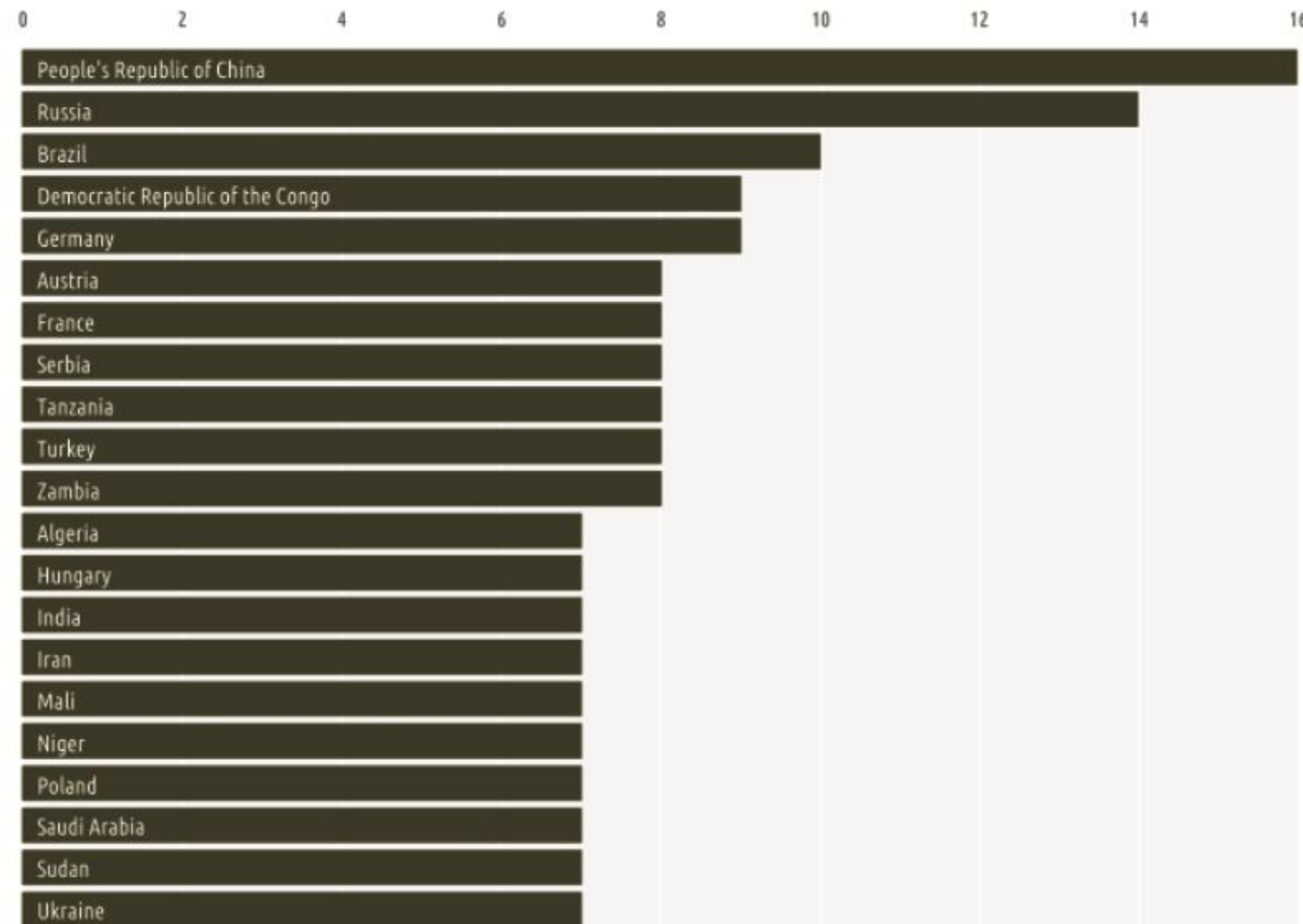
© Financial Times 2020-2021
This is a free resource for anyone to use.
Attribution-NonCommercial-International License

Tipos de gráficos dependiendo del objetivo

Bar chart

Comparación

THE COUNTRIES WITH THE MOST LAND NEIGHBOURS



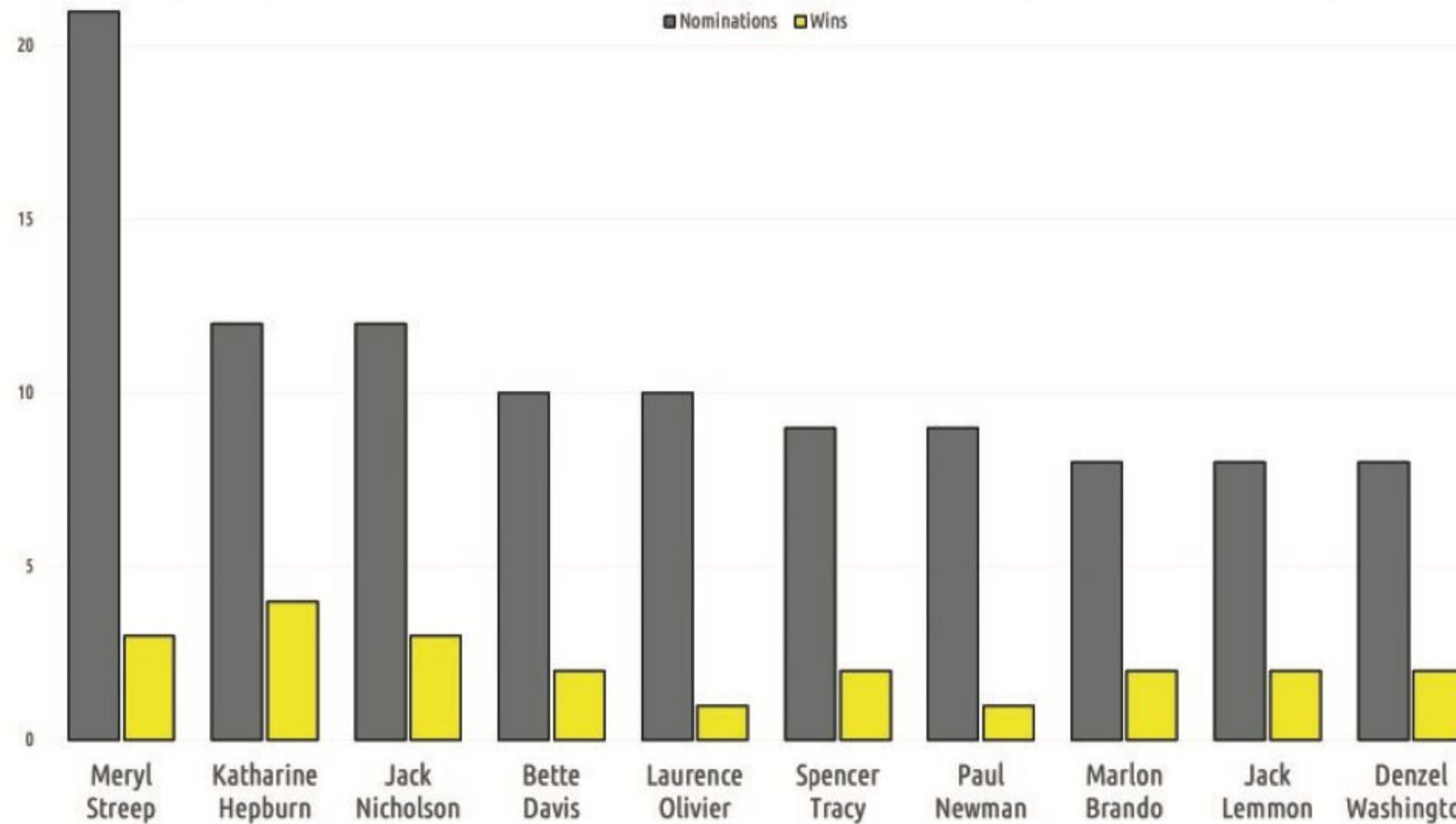
Source: https://en.wikipedia.org/wiki/List_of_countries_and_territories_by_land_borders (as at December 2018).

Notes: Minimum 7 neighbouring countries. France's figure does not include French overseas departments, collectivities, and territories.

Clustered bar chart

Comparación

The 10 actors who have received the most Oscar nominations for acting



Source: www.filmsite.org (as at December 2018)

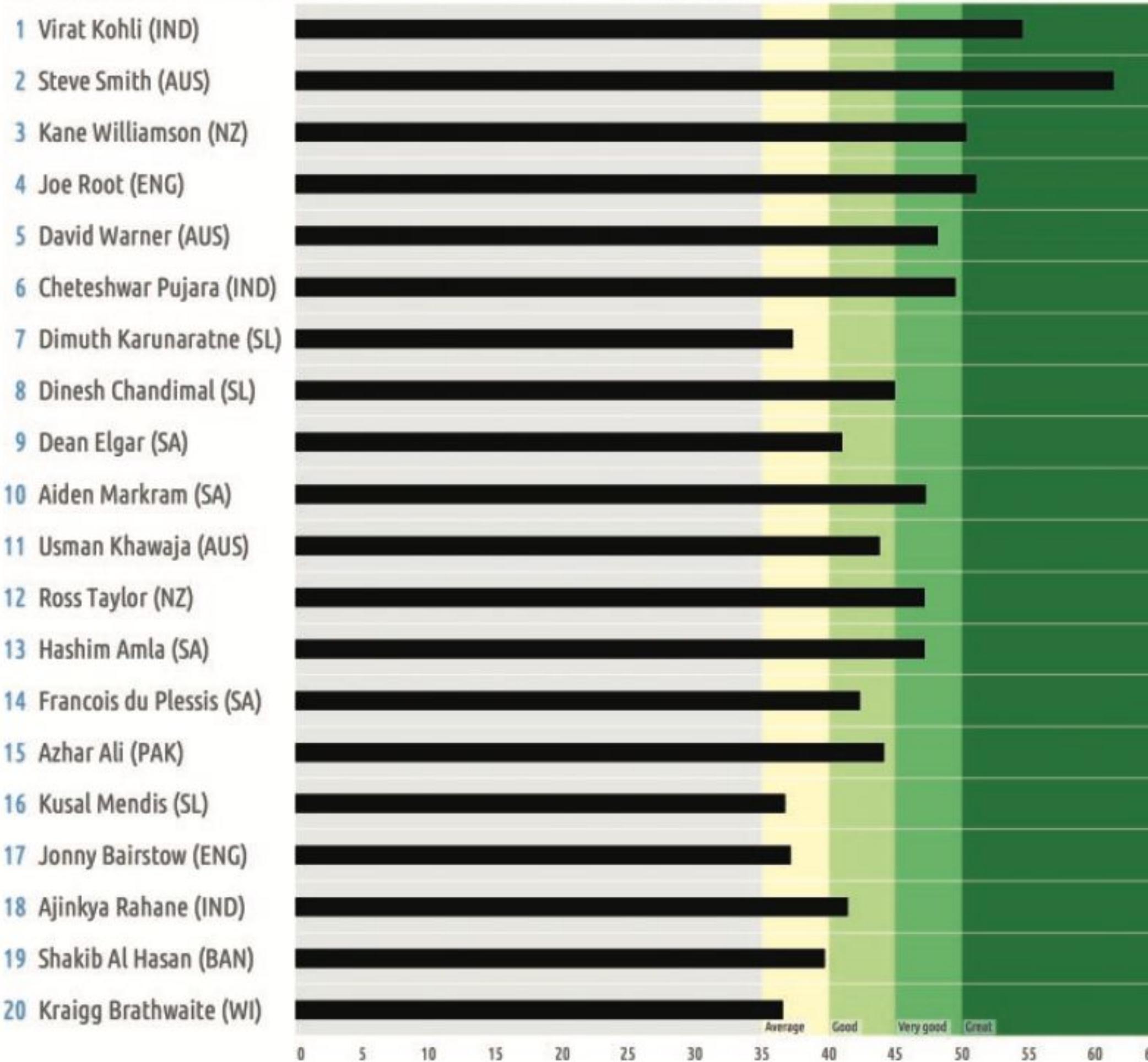
Notes: Al Pacino and Geraldine Page each have 8 nominations, 1 win. Peter O'Toole has 8 nominations but no win.

Bullet chart

Comparación

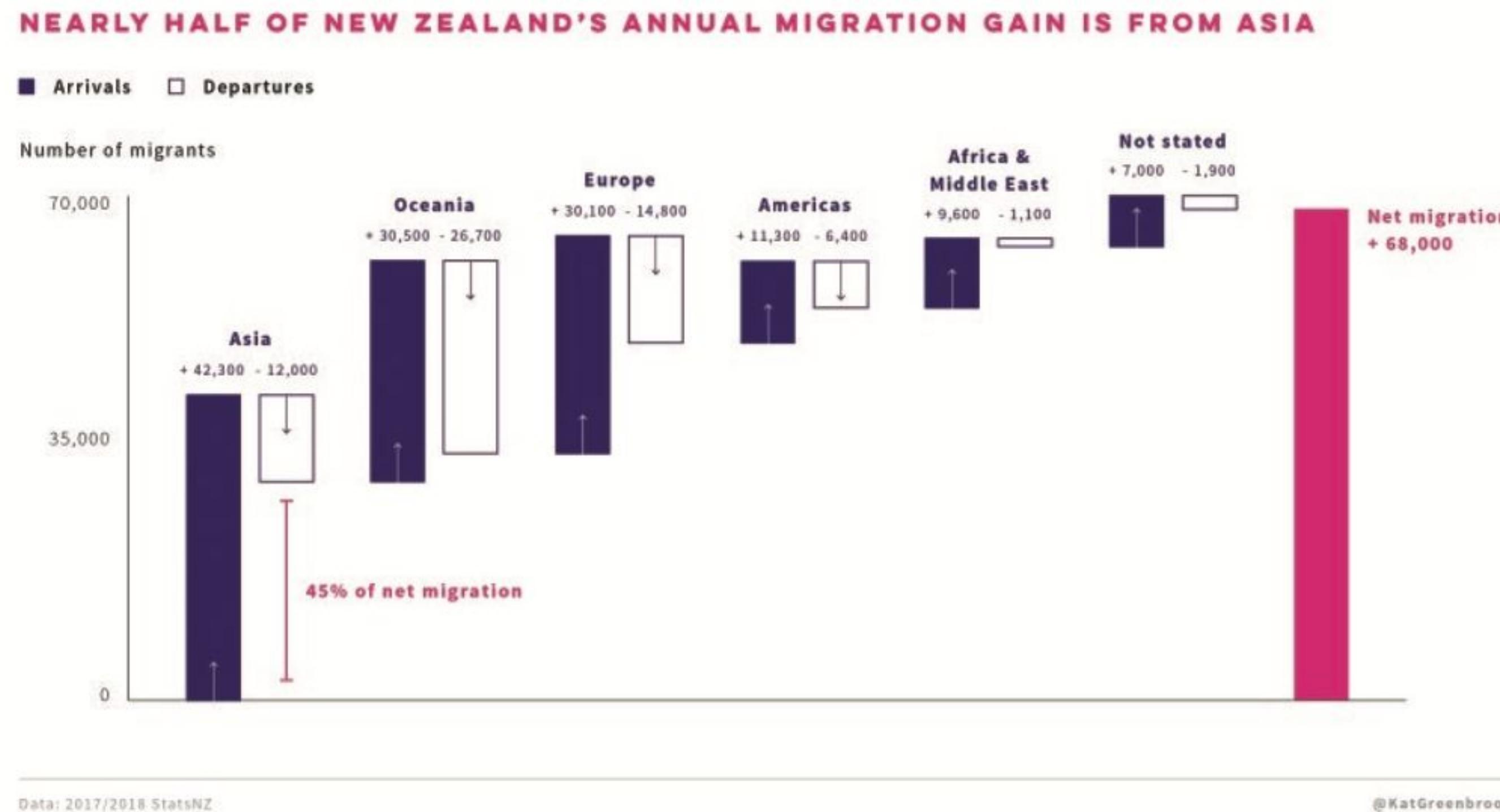
Top 20 Ranked Batters in Men's Test Cricket

SOURCE: ICC Rankings <https://www.icc-cricket.com/rankings/mens/player-rankings/test/batting> | Batting averages <http://www.espncricinfo.com/> as at October 2018



Waterfall chart

Comparación



Radar chart

Comparación

United Kingdom

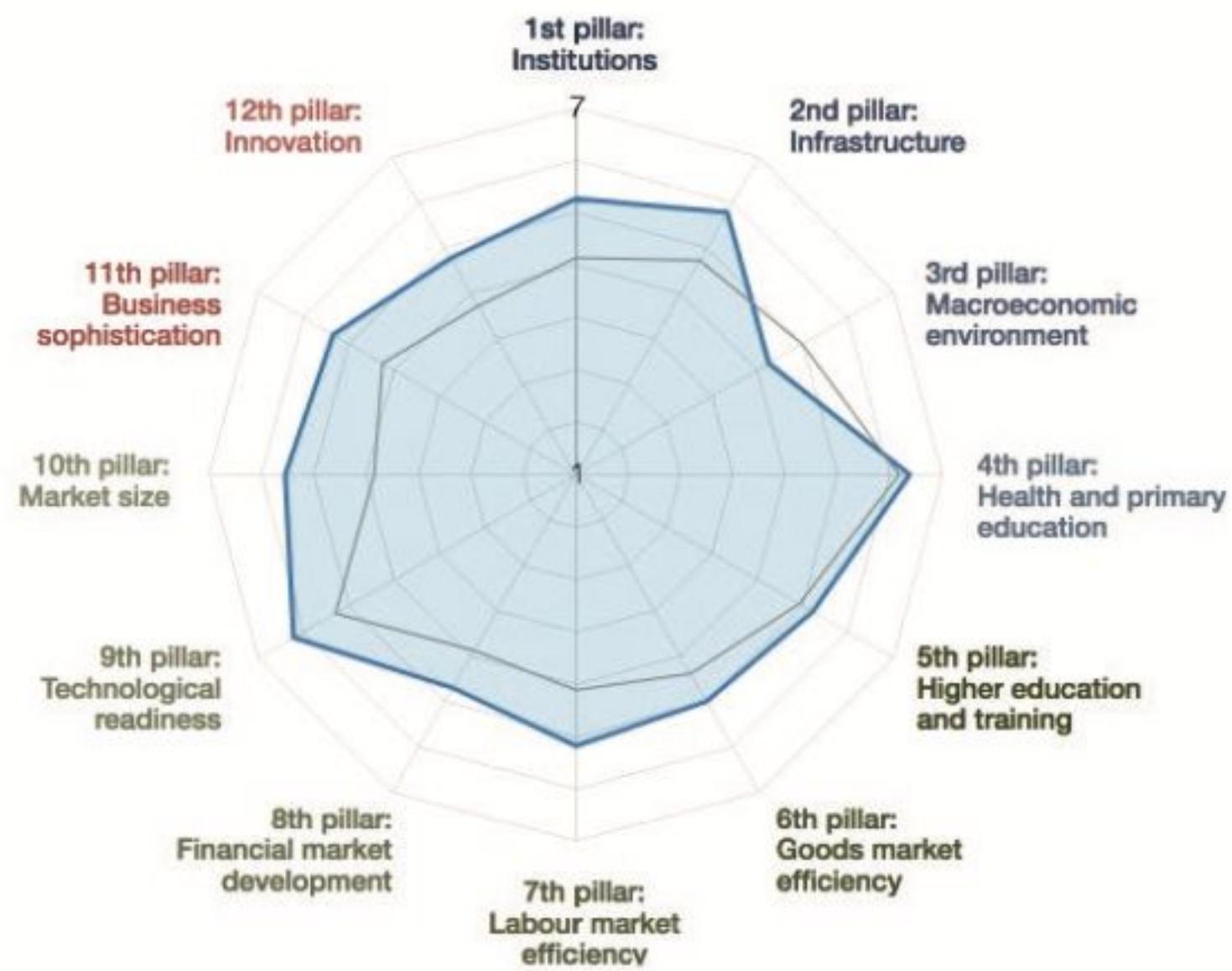
Global Competitiveness Index 2017-2018 edition

8th
/137

Key indicators, 2016

Source: International Monetary Fund; World Economic Outlook Database (April 2017)

Population millions	66.1	GDP per capita US\$	39,734.6
GDP US\$ billions	2,624.5	GDP (PPP) % world GDP	2.29

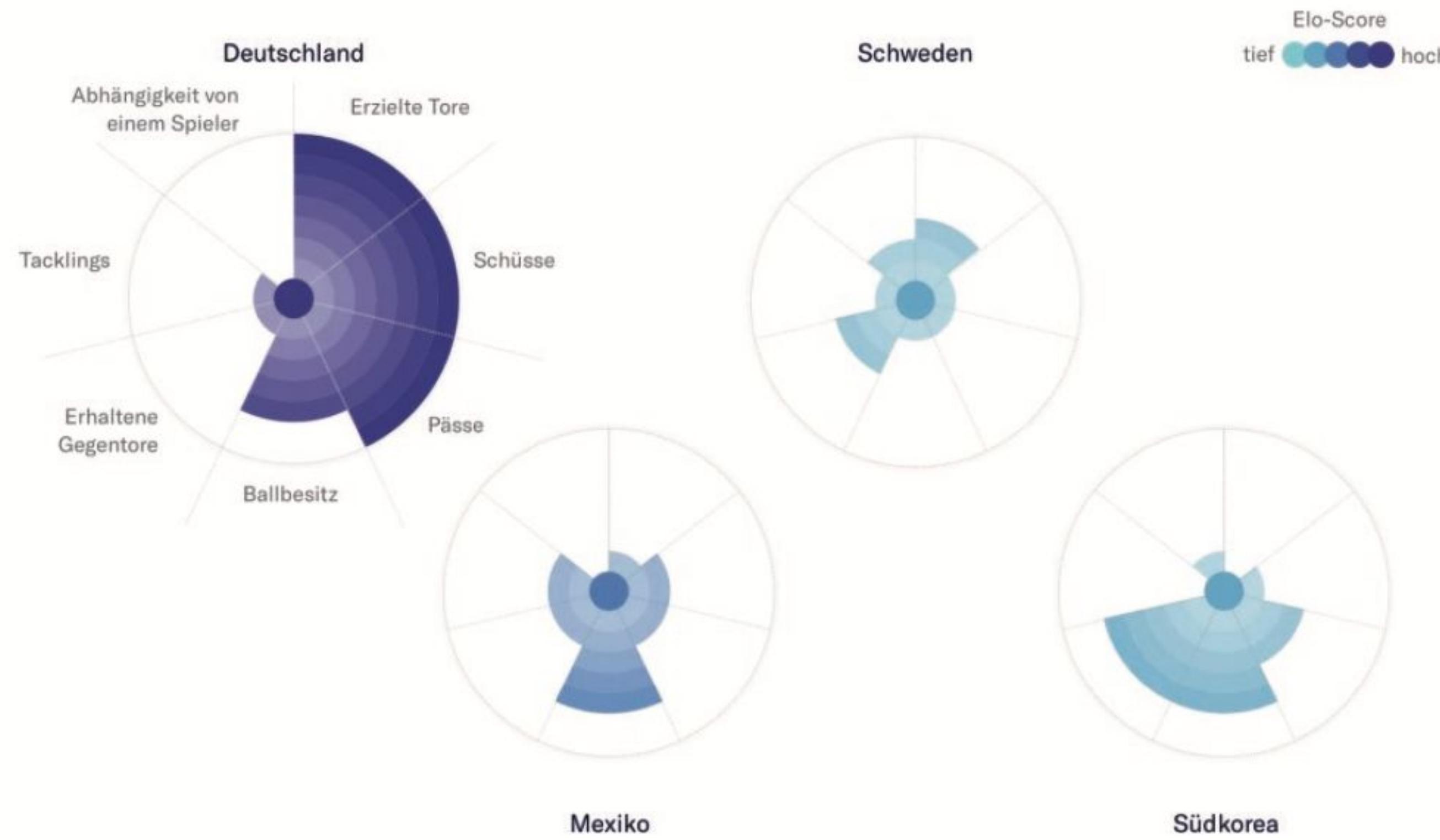


■ United Kingdom ■ Europe and North America

Polar chart

Comparación

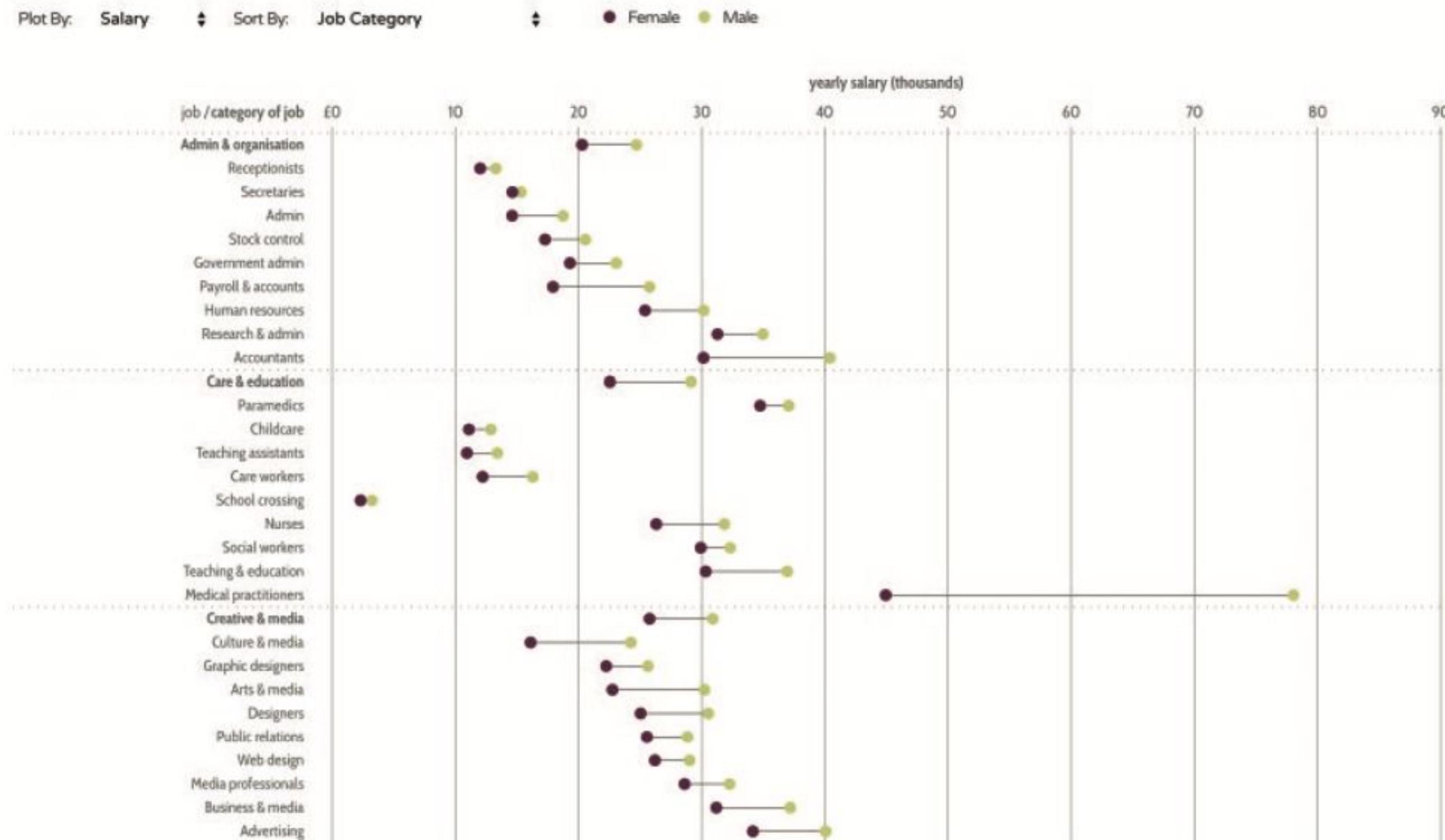
Gruppe F: Breit abgestützte Offensiven



Connected dot plot

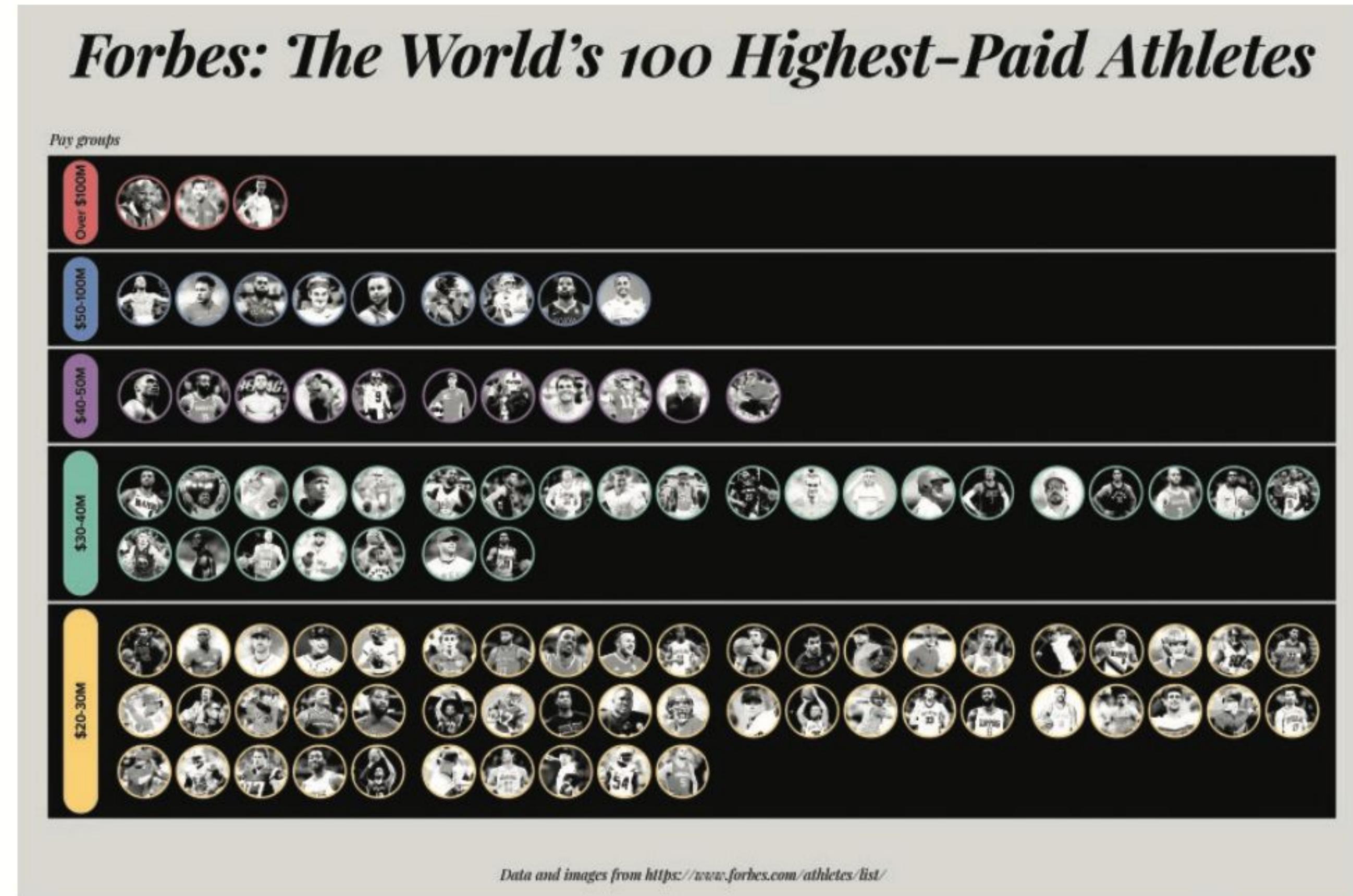
Comparación

Gender Pay Gap US | UK



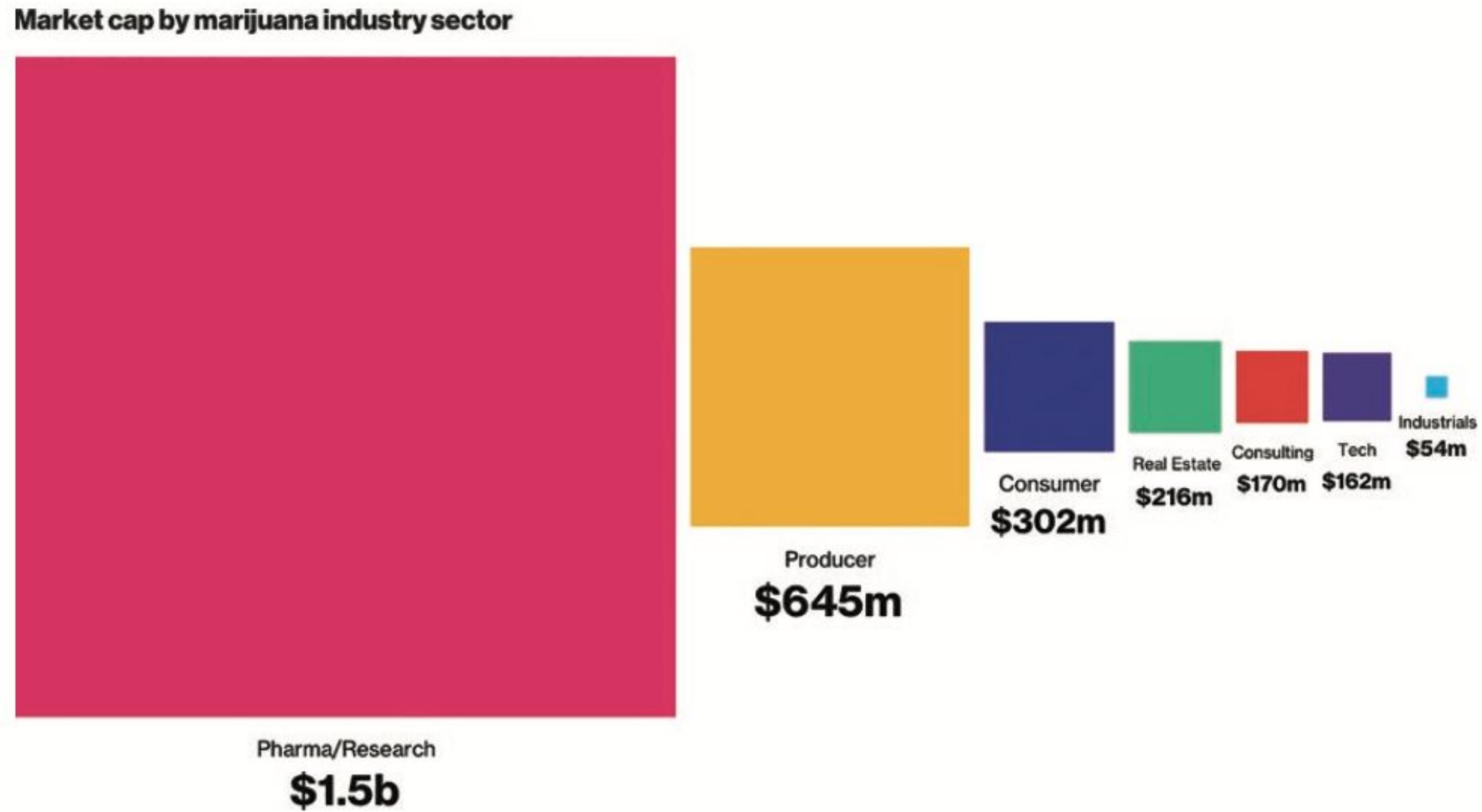
Pictogram

Comparación



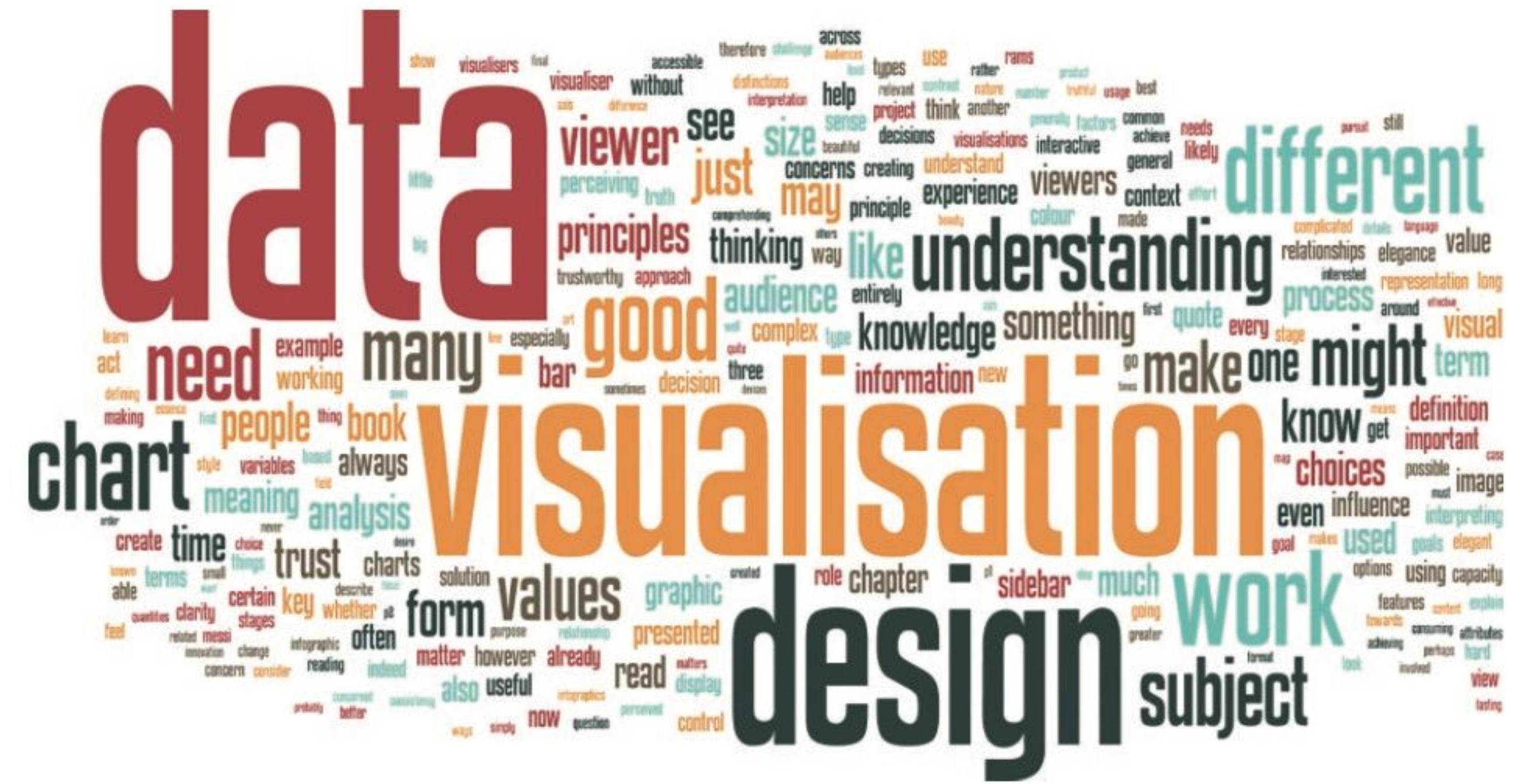
Proportional symbol chart

Comparación



Word cloud

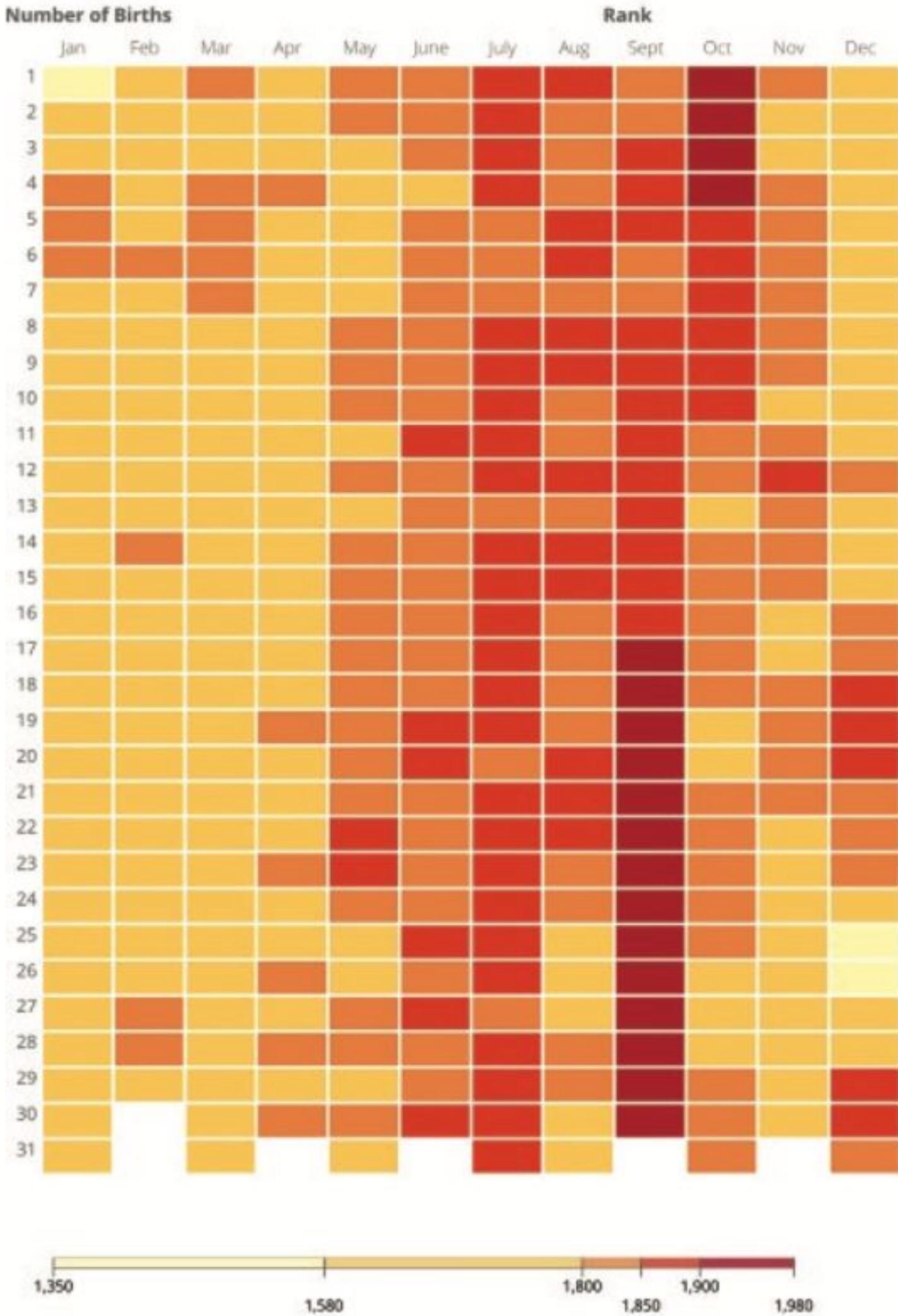
Comparación



Heat map

How popular do you think your own birthday is? Find out with our interactive graphic

Average daily births, England and Wales, 1995 to 2014

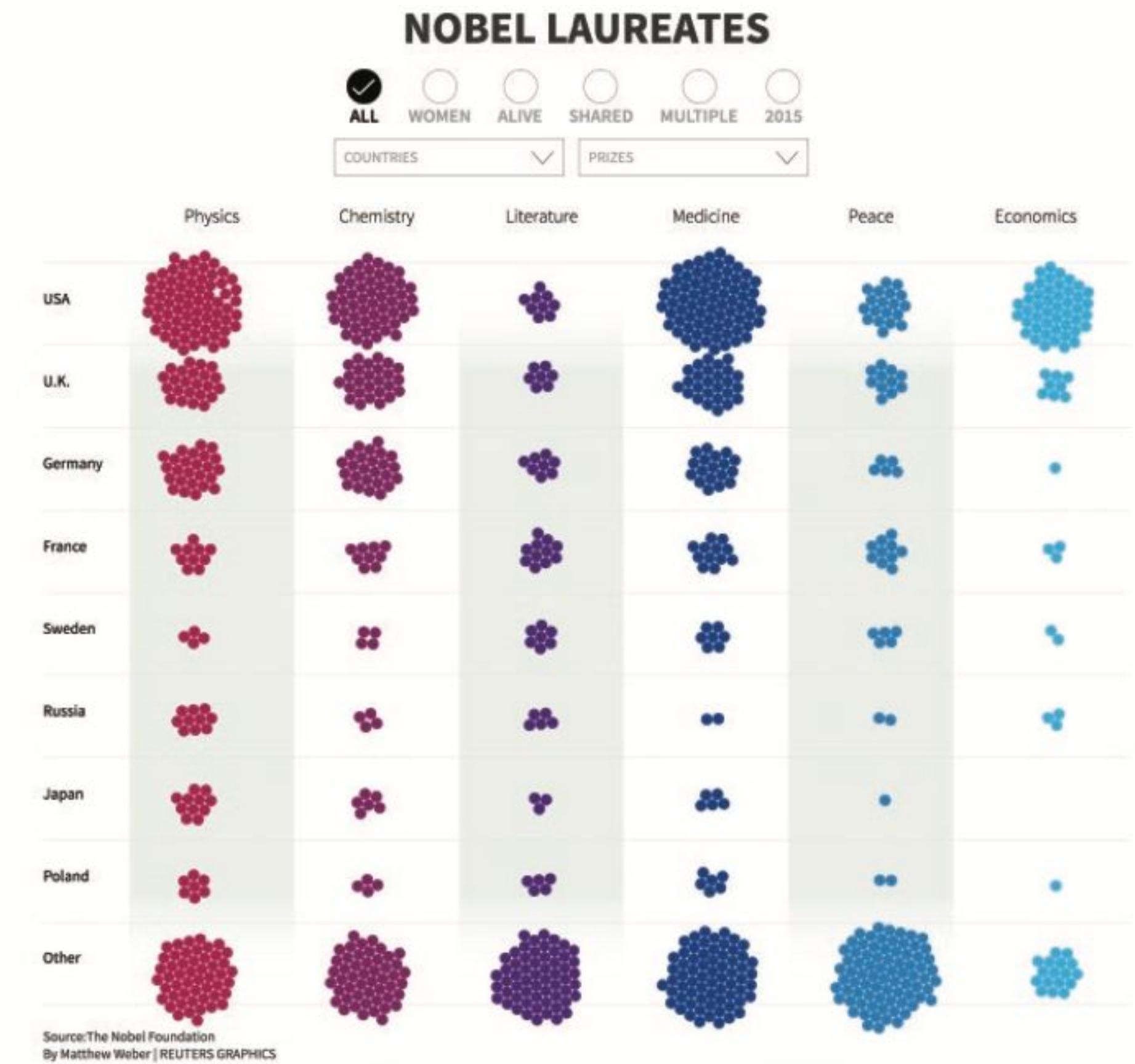


Source: Birth registrations in England and Wales

Comparación

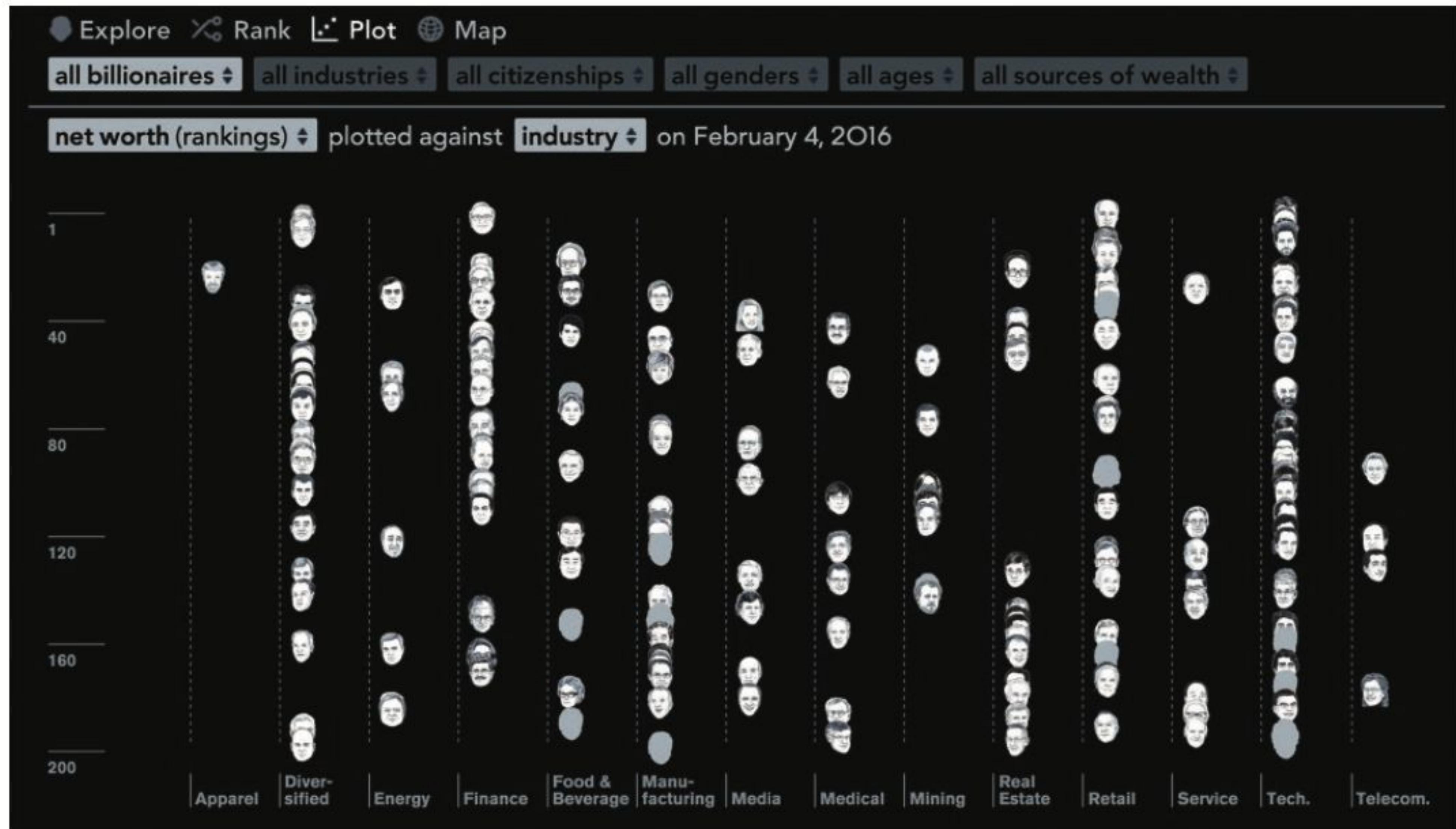
Matrix chart

Comparación



Dot plot

Distribución



Beeswarm plot

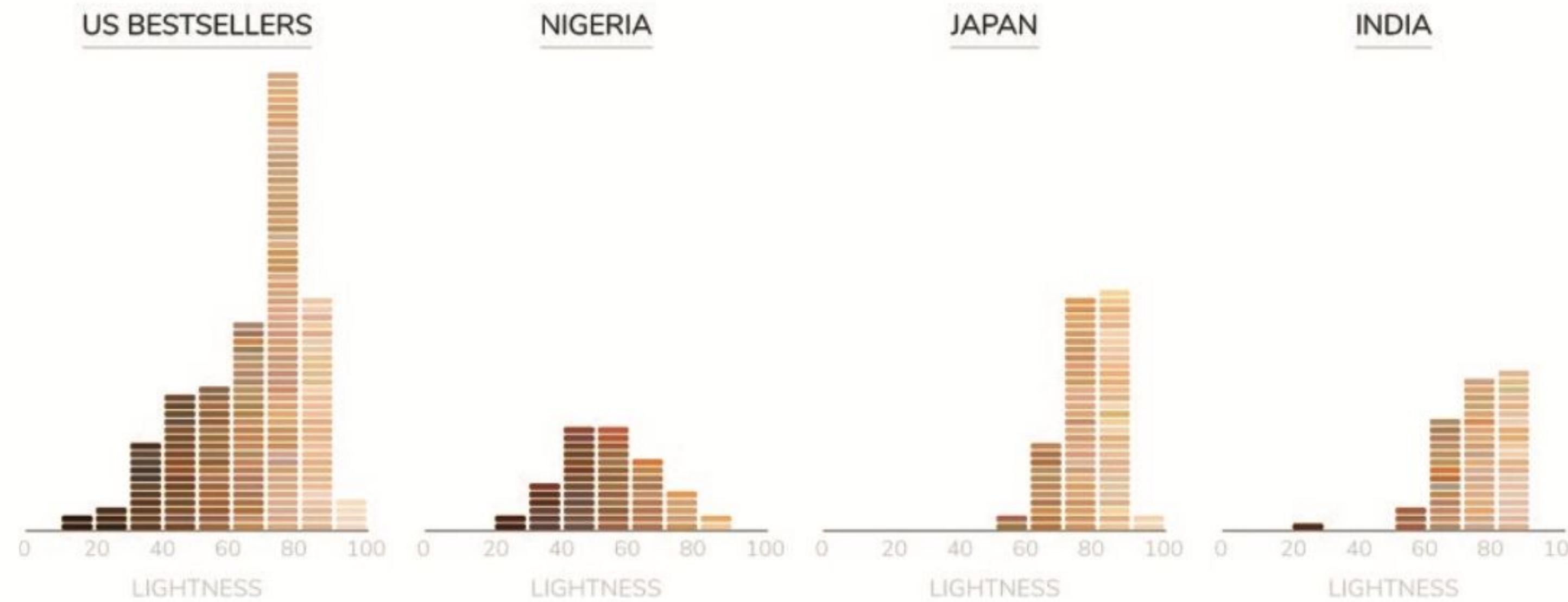
Distribución



Histogram

Distribución

Foundation Lightness Around the World

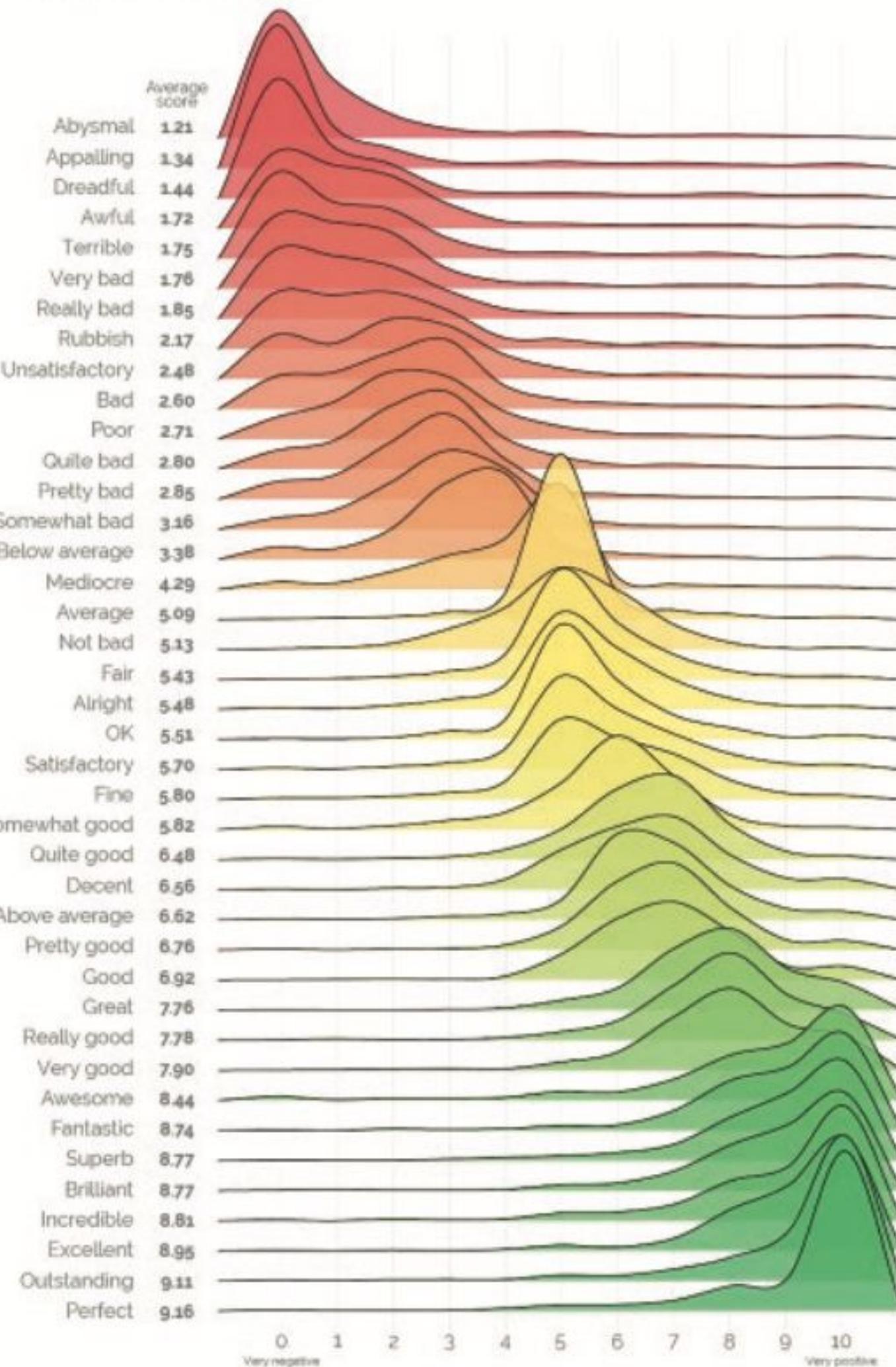


Density plot

Distribución

How good is "good"? Now with even more words!

On a scale of 0 to 10, where 0 is 'very negative' and 10 is 'very positive', in general, how positive or negative would the following word/phrase be to someone when you used it to describe something?

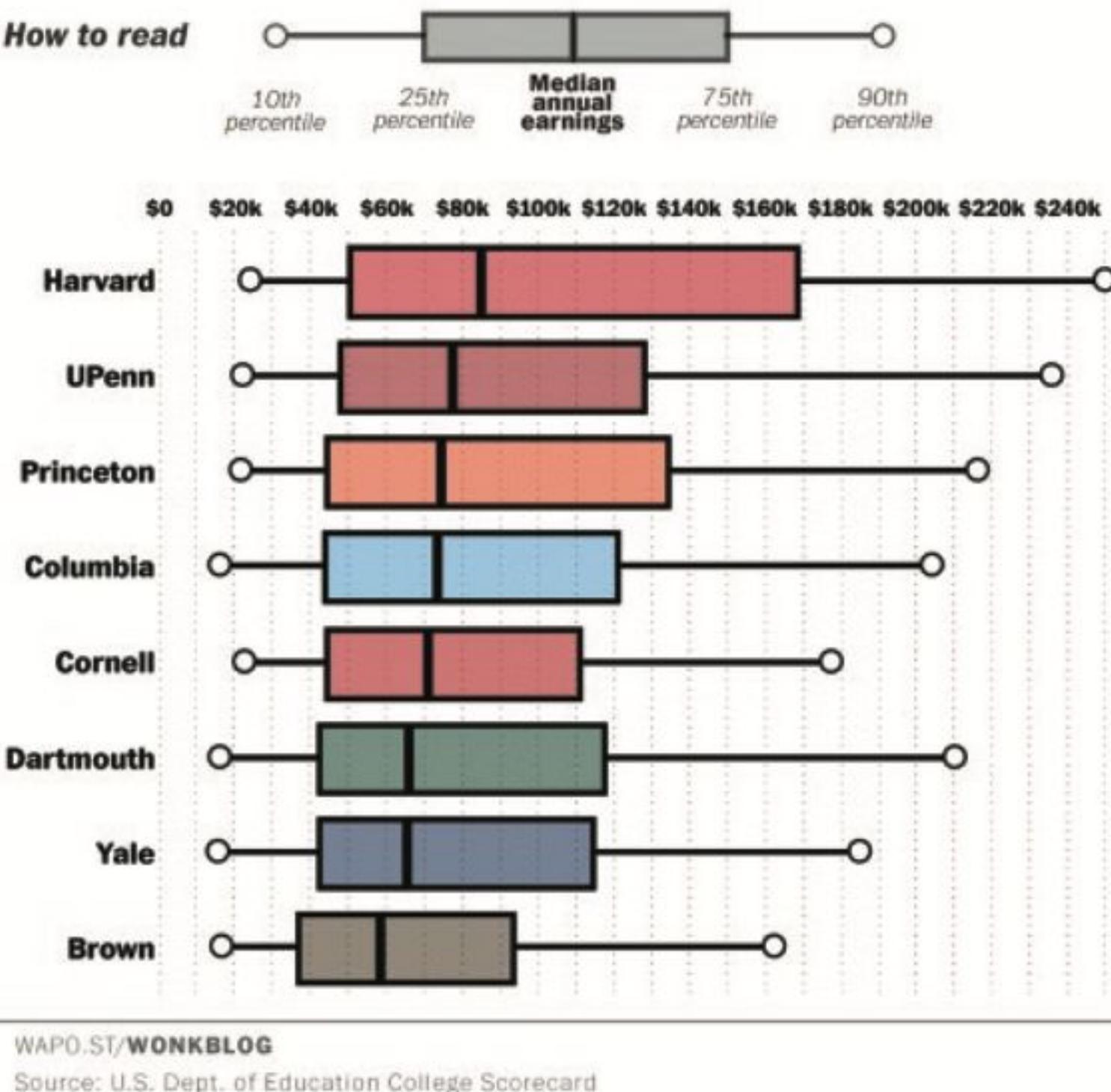


Box-and-whisker plot

Distribución

Ranking the Ivies

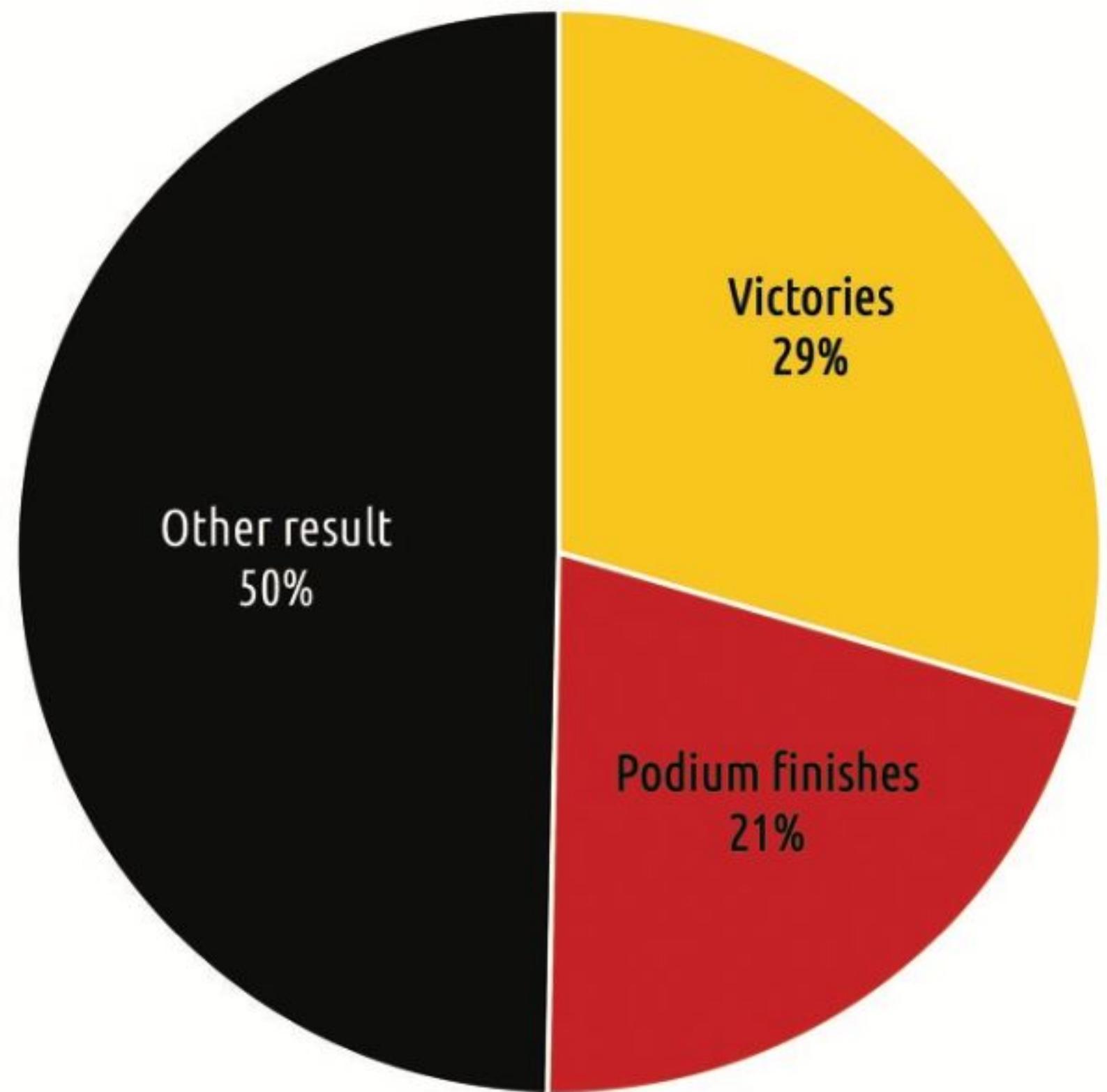
Annual earnings distributions, 10 years after starting school



Pie chart

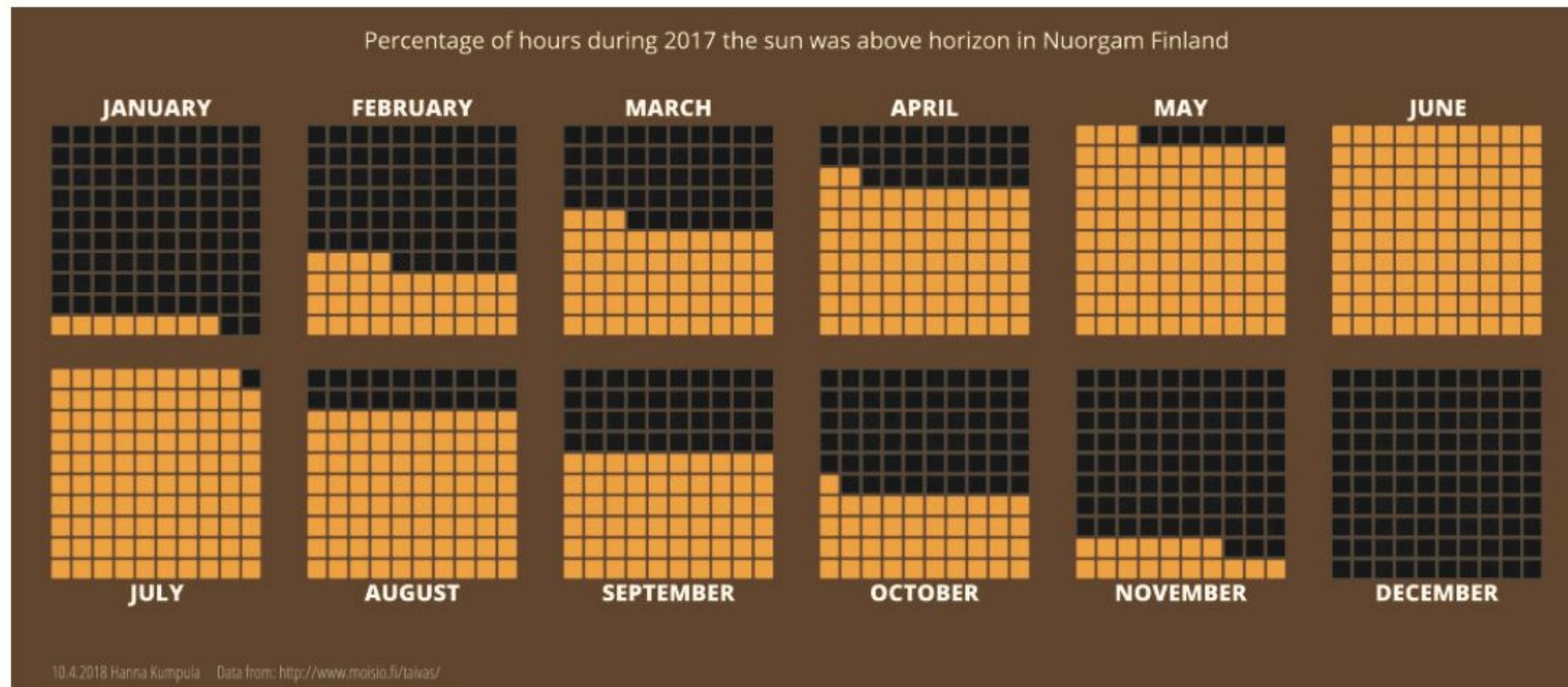
Part-to-whole

Breakdown of Michael Schumacher's F1 Career Over 308 Races



Waffle chart

Part-to-whole



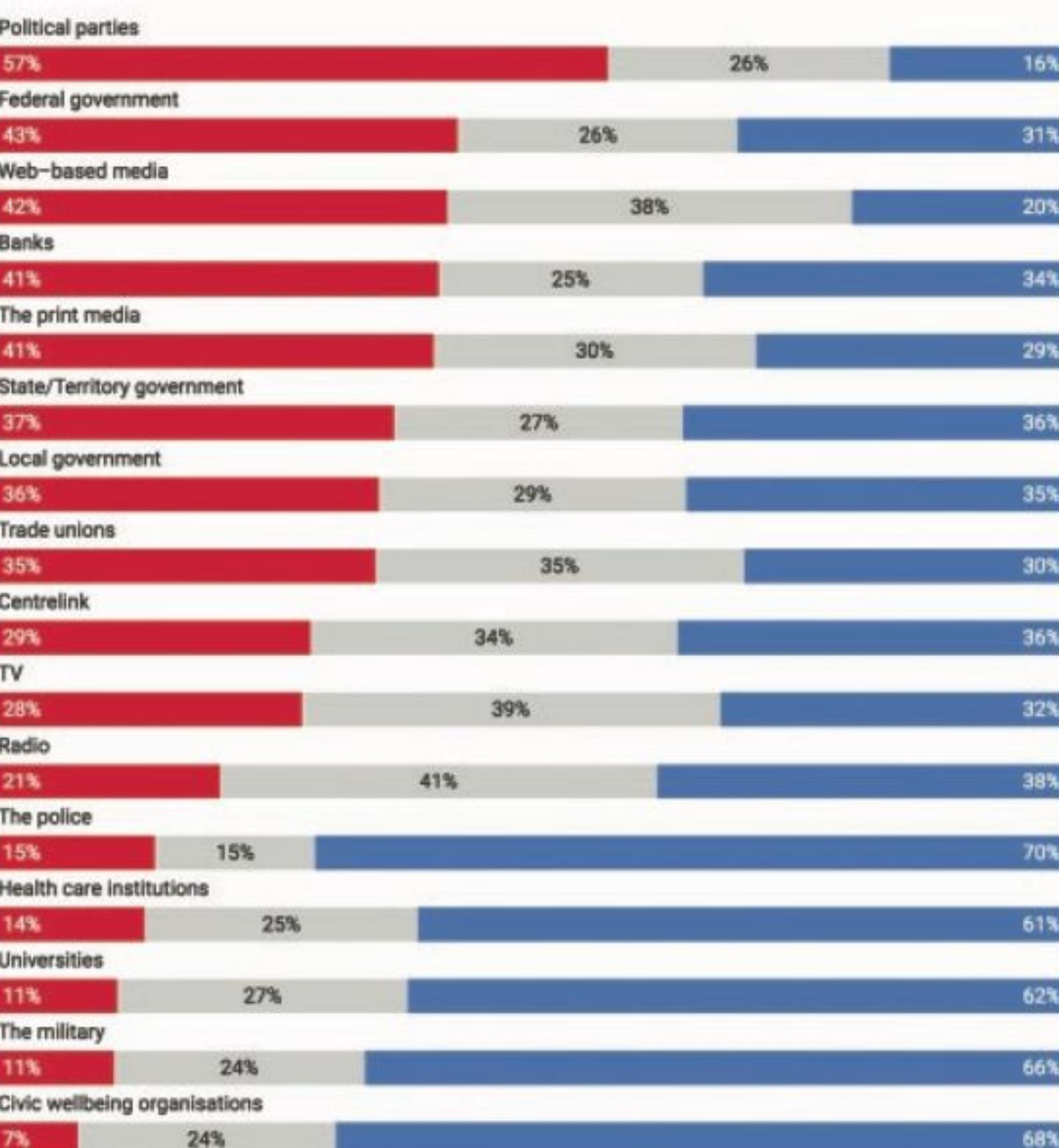
Stacked bar chart

Part-to-whole

Only 16 per cent of Australians trust political parties, compared to 30 per cent for trade unions and 70 per cent for police.

How much do you trust ...?

■ Distrust ■ Neither ■ Trust

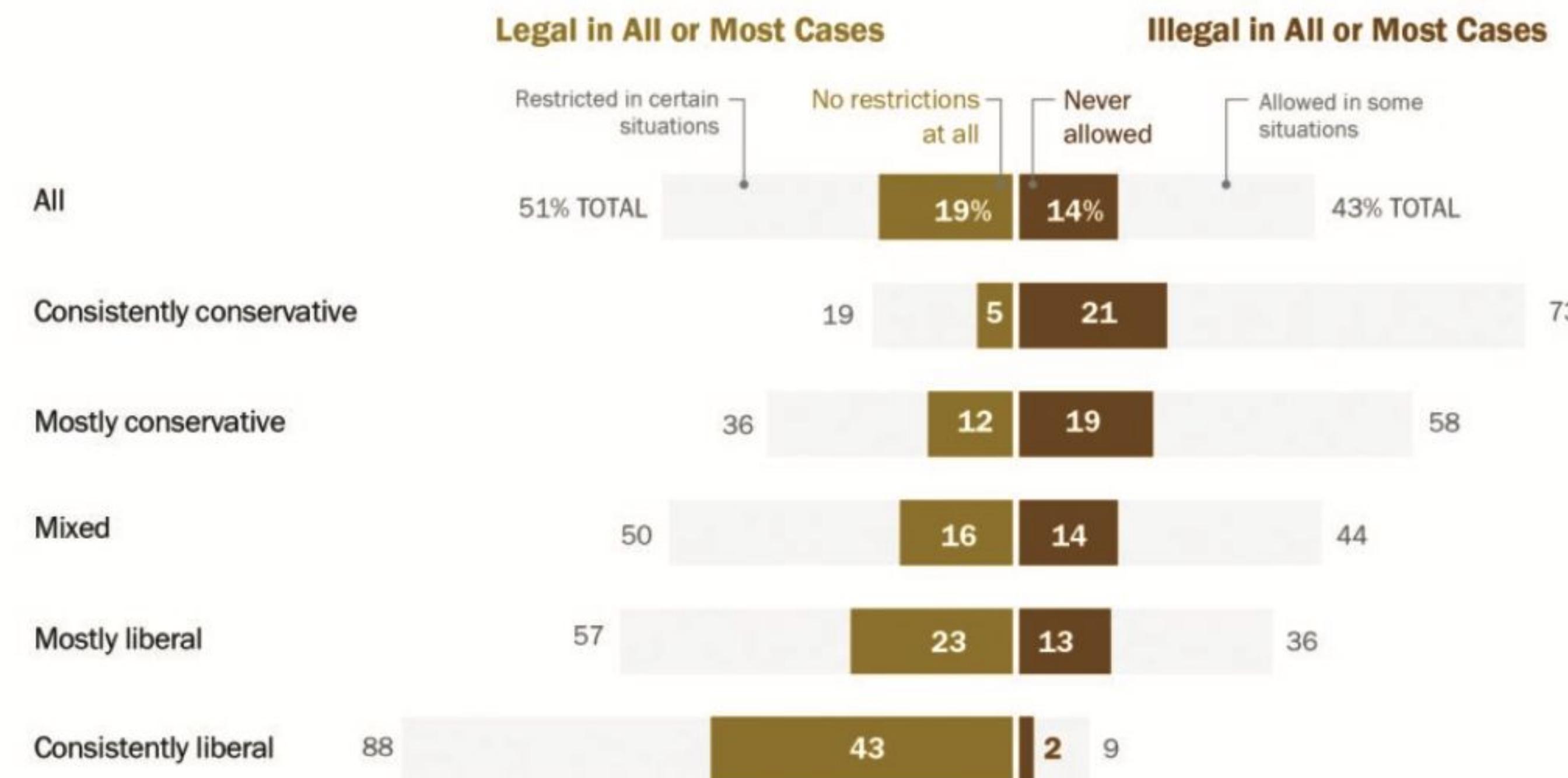


Diverging bar chart

Part-to-whole

Liberals Most Likely to Favor No Restrictions on Abortion

Abortion should be ...



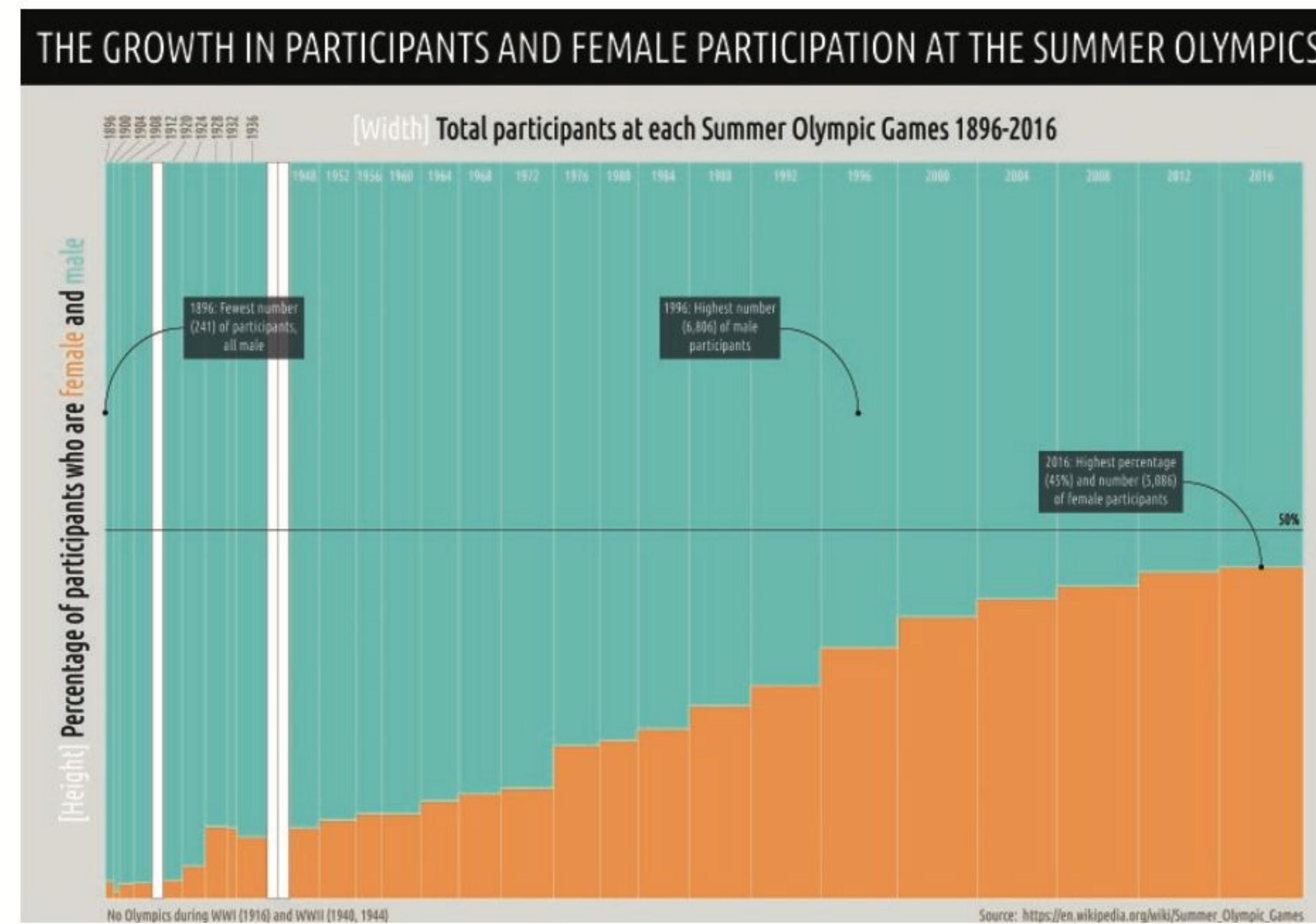
Source: 2014 Political Polarization in the American Public

Notes: "Don't know" responses not shown. Ideological consistency based on a scale of 10 political values questions (see Appendix A)

PEW RESEARCH CENTER

Marimekko chart (proportional stacked bar)

Part-to-whole



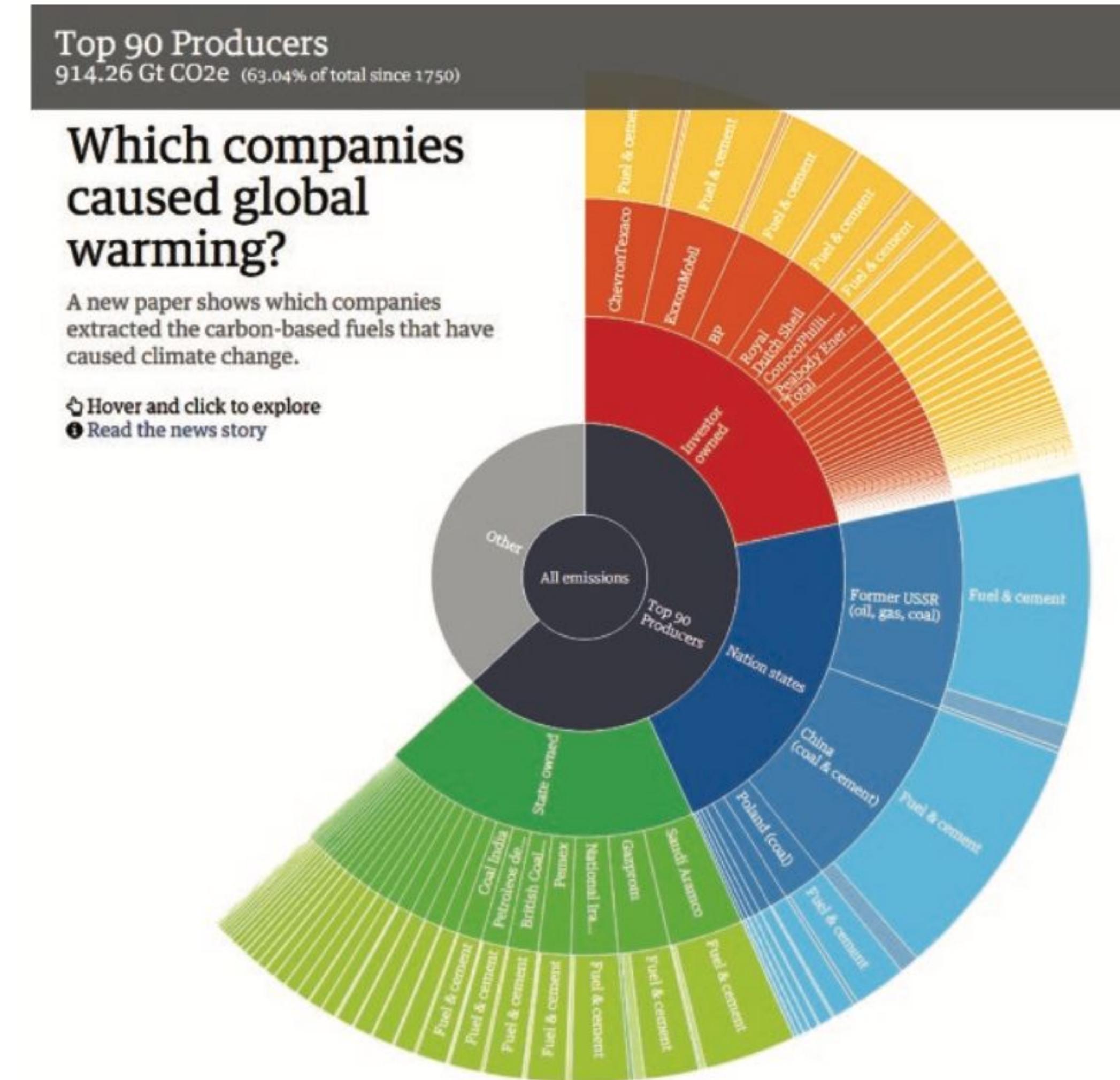
Treemap

Jerarquía



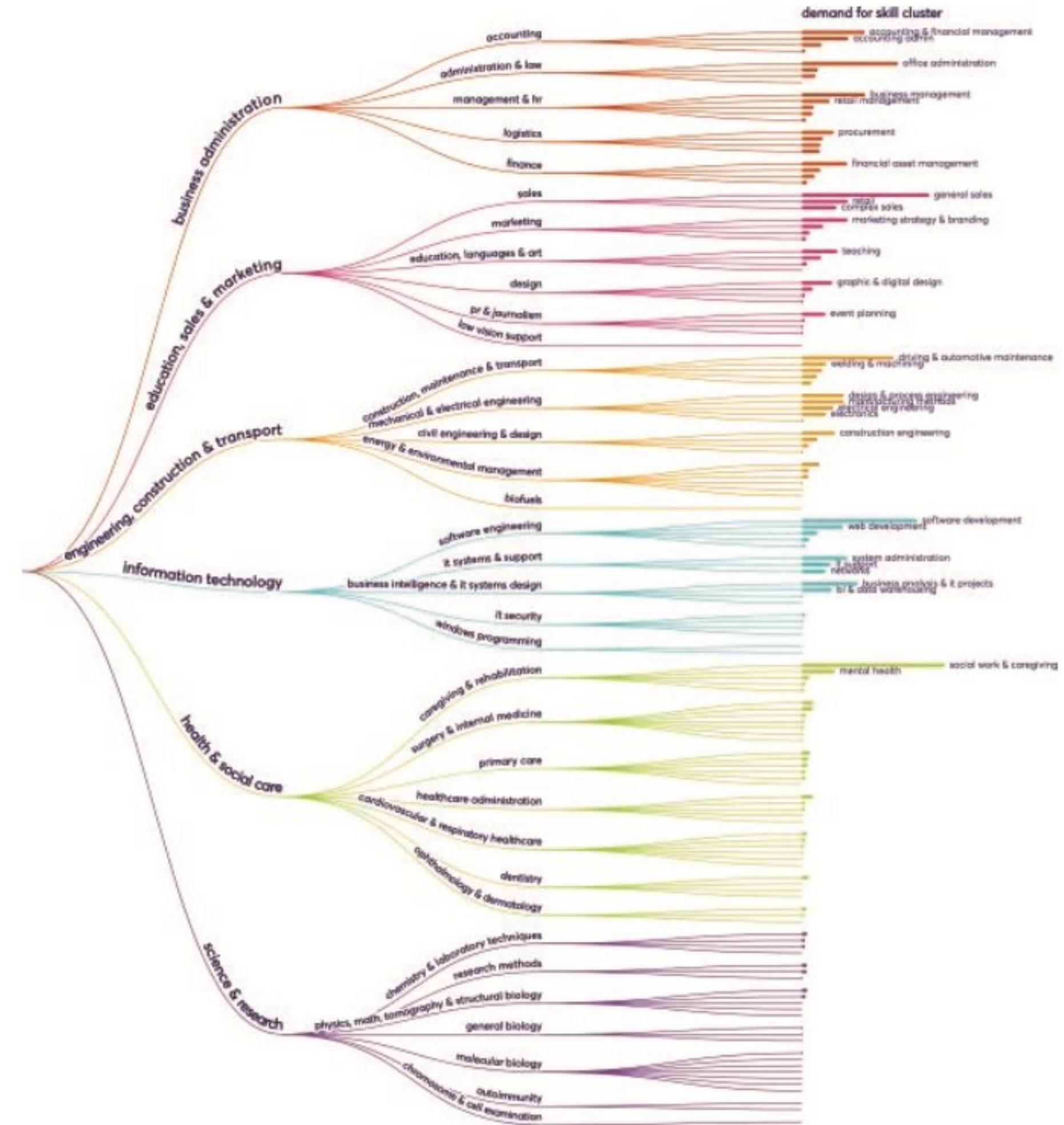
Sunburst chart

Jerarquía



Dendrogram

Jerarquía



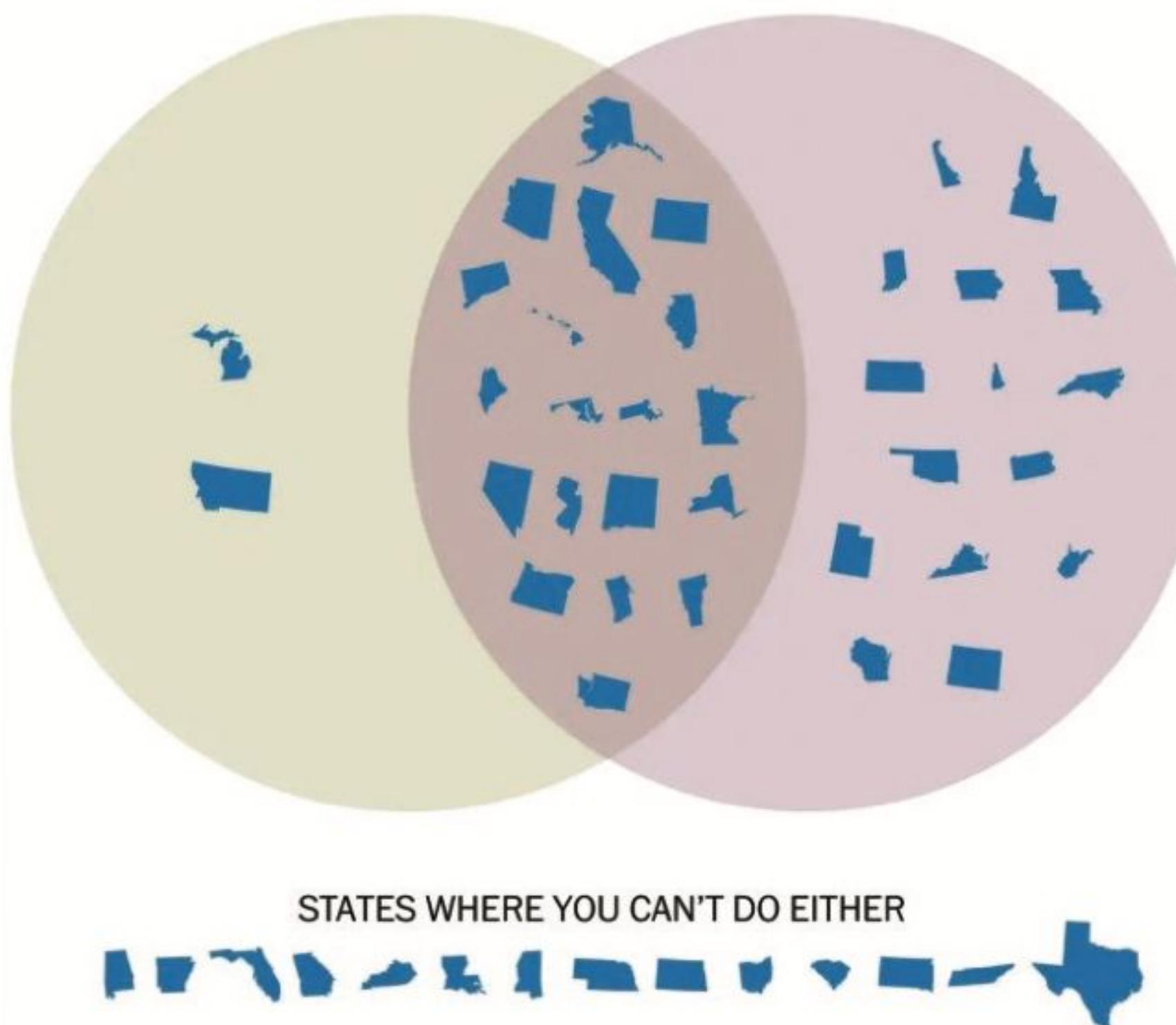
Venn diagram

Jerarquía

The Venn diagram of cultural politics

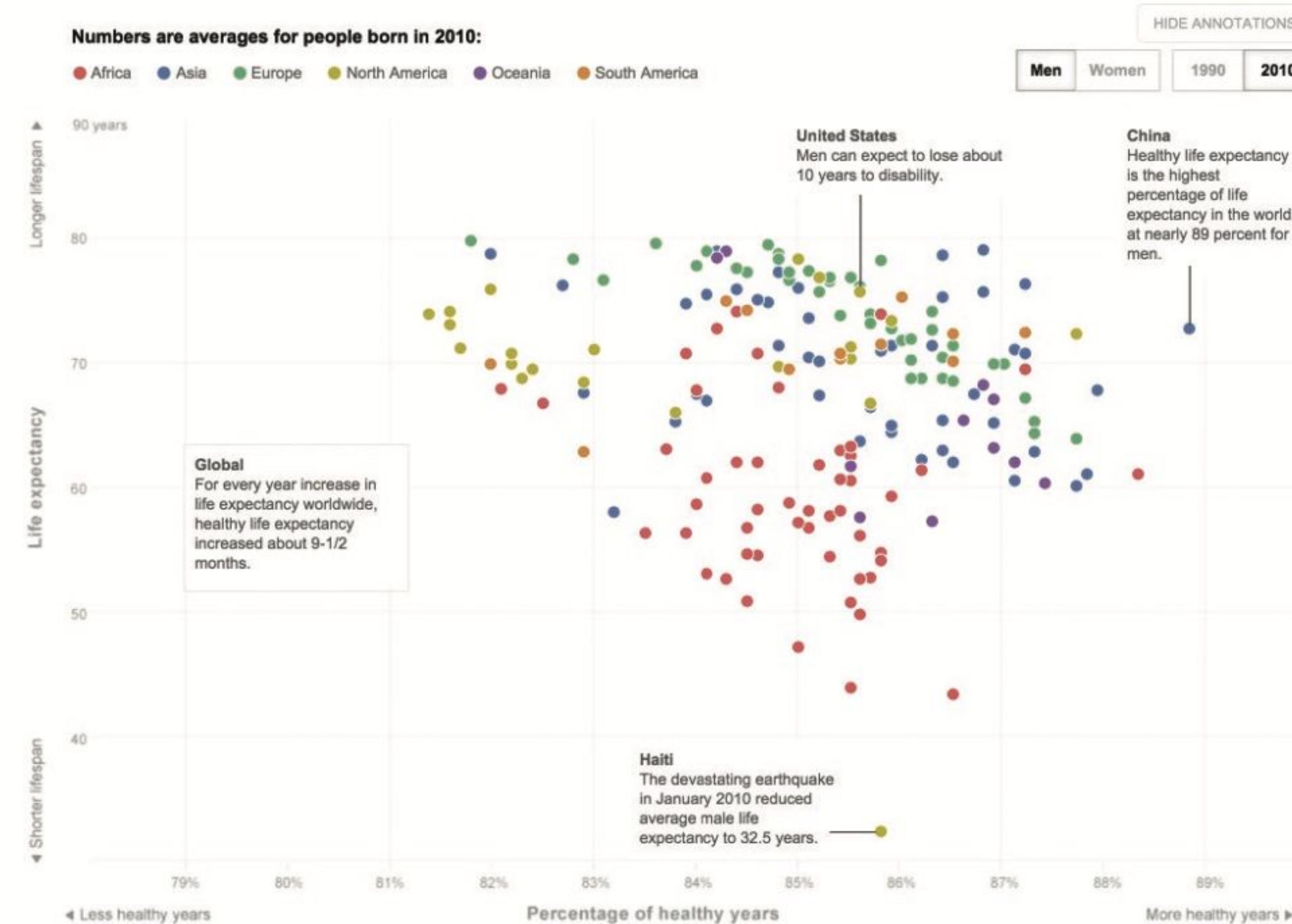
STATES WHERE YOU CAN
LEGALLY USE MARIJUANA

STATES WHERE SAME-SEX
COUPLES CAN GET MARRIED



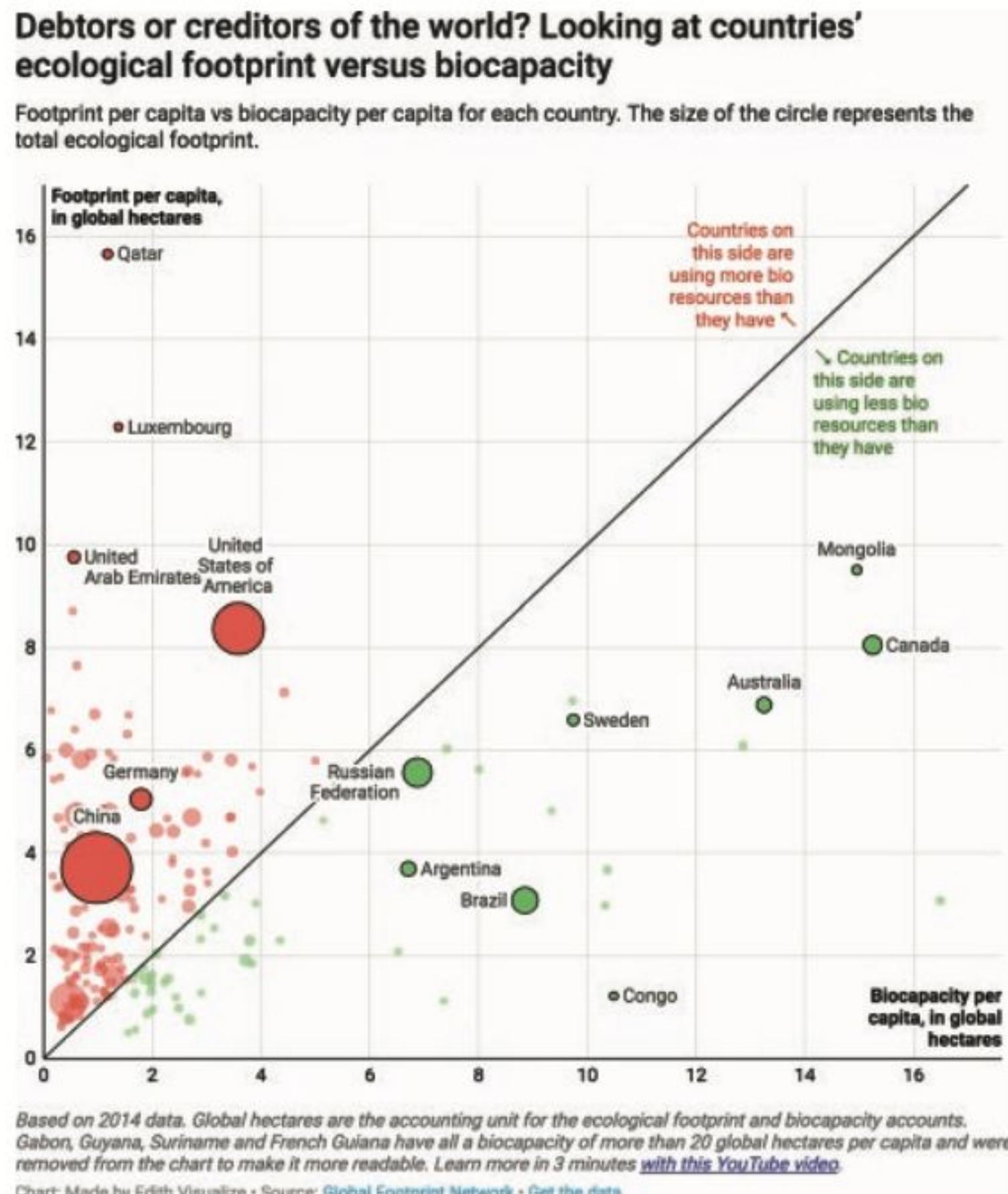
Scatter plot

Correlaciones



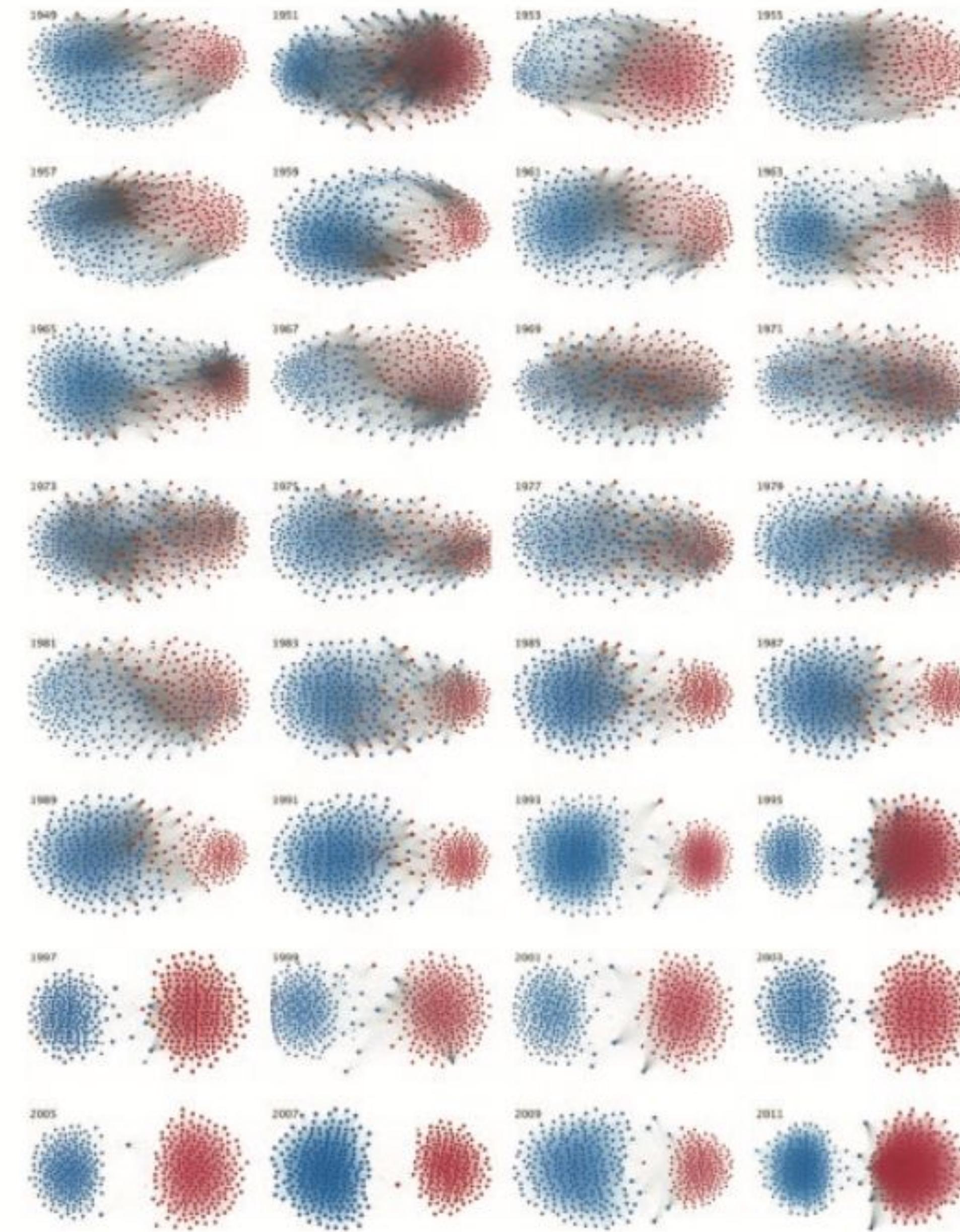
Bubble plot

Correlaciones



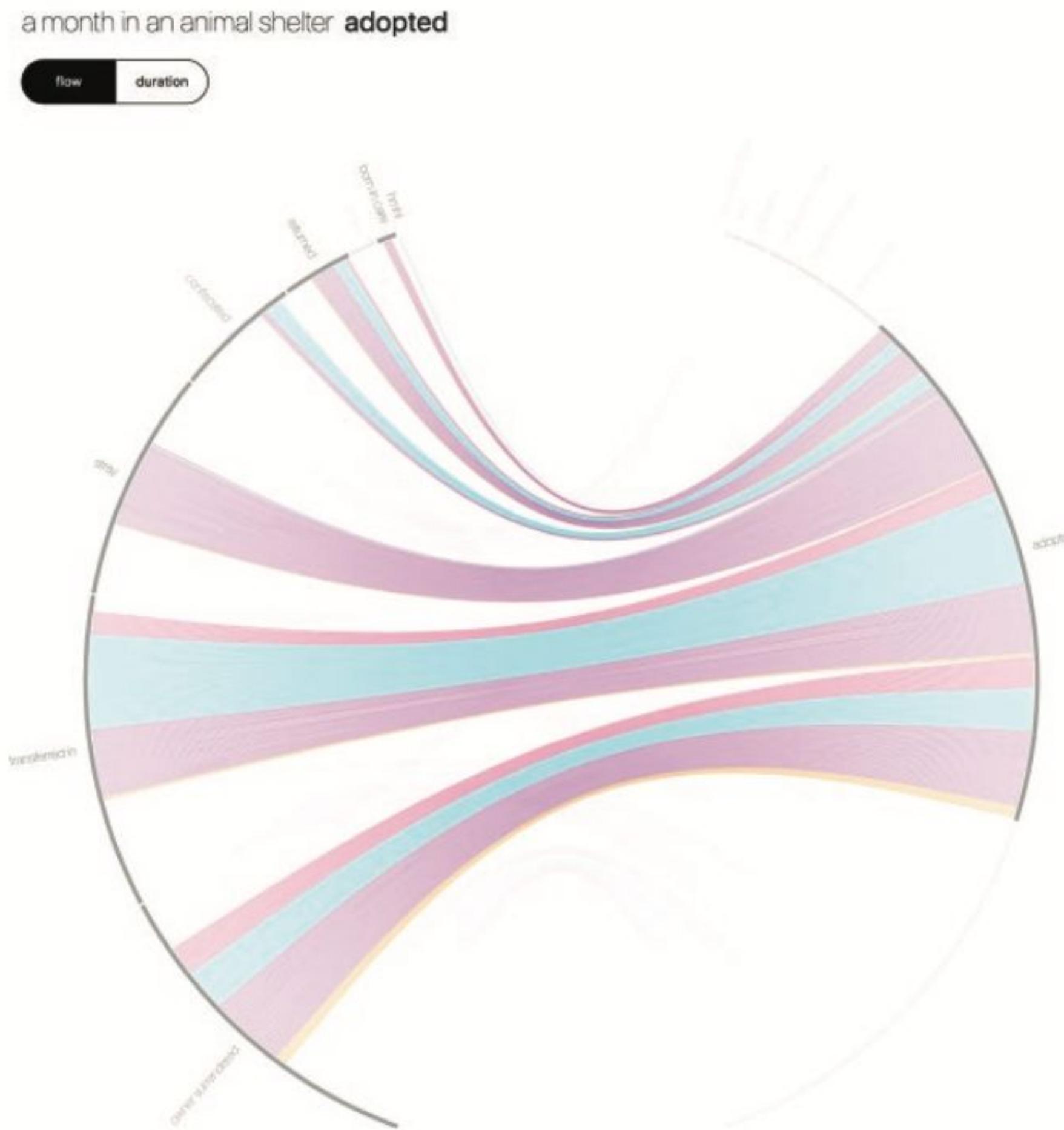
Network diagram

Conexiones



Sankey diagram

Conexiones



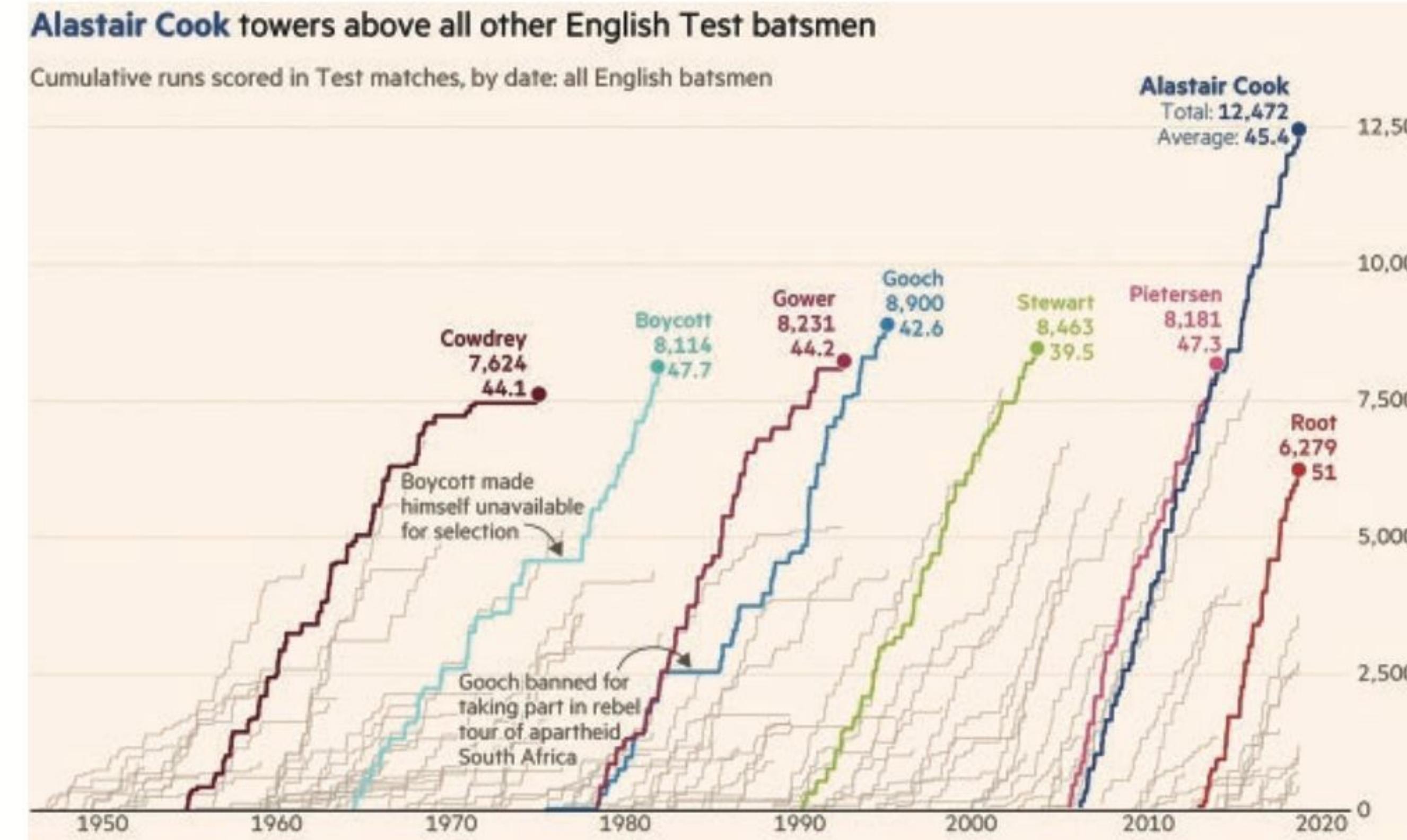
Chord diagram

Conexiones



Line chart

Tendencias

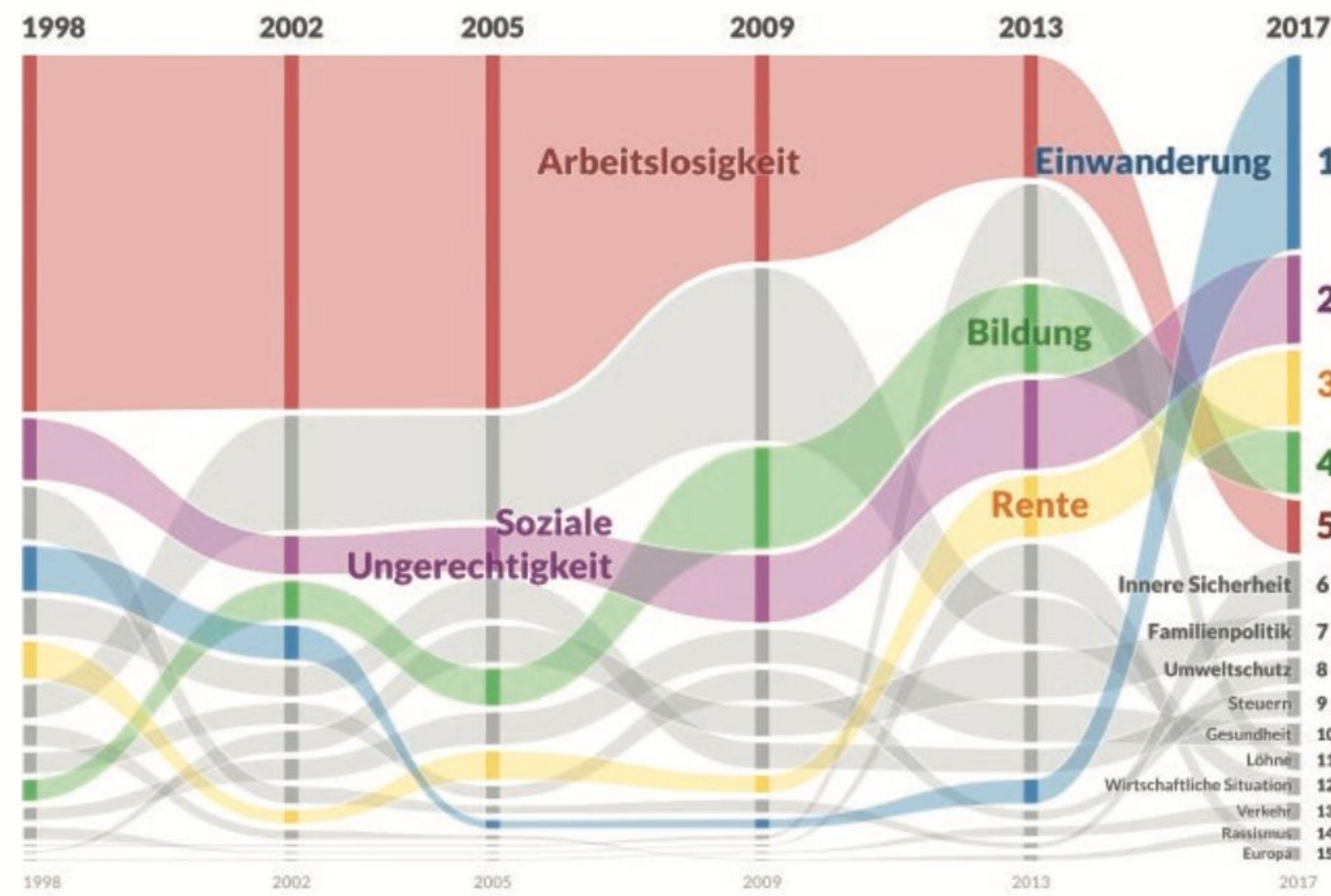


Bump chart (rank chart)

Tendencias

These are the 15 most important political problems in Germany

The chart shows which topics Germans are the most active in this general election and what significance they had in previous elections.

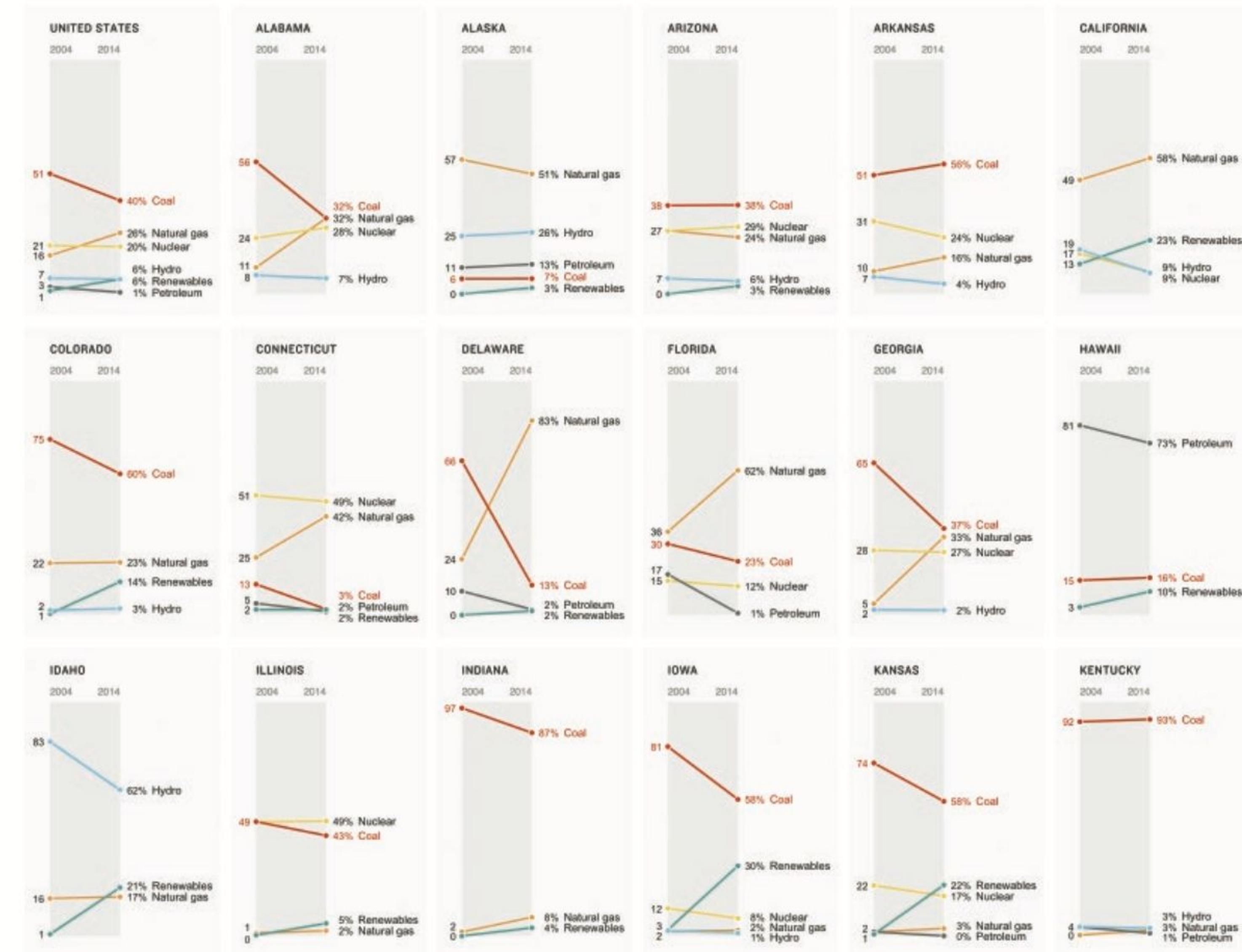


Source: Infratest dimap

Slope graph

How Each State Generates Electric Power (2004-2014)

Tendencias

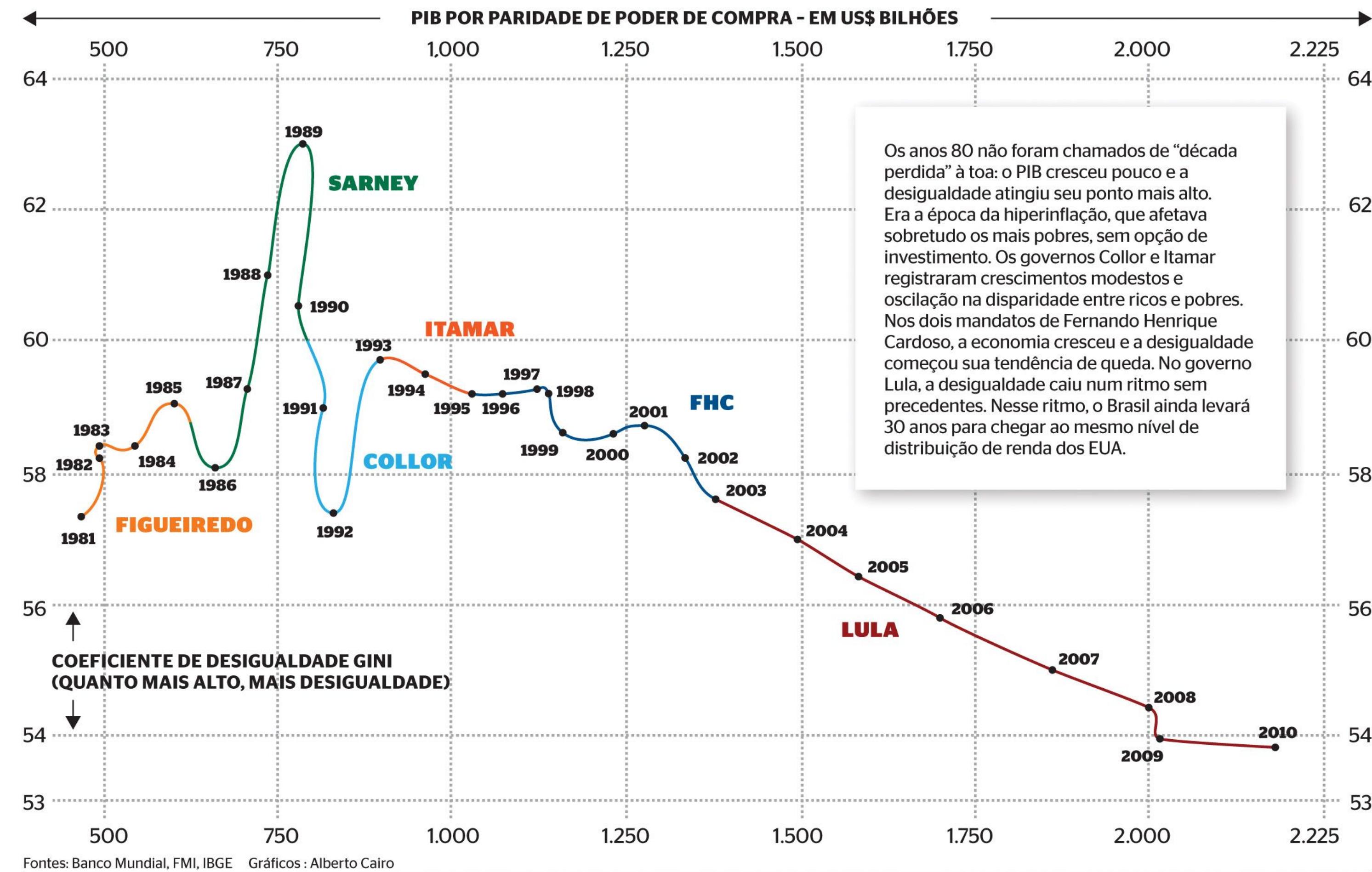


Connected scatter plot

Tendencias

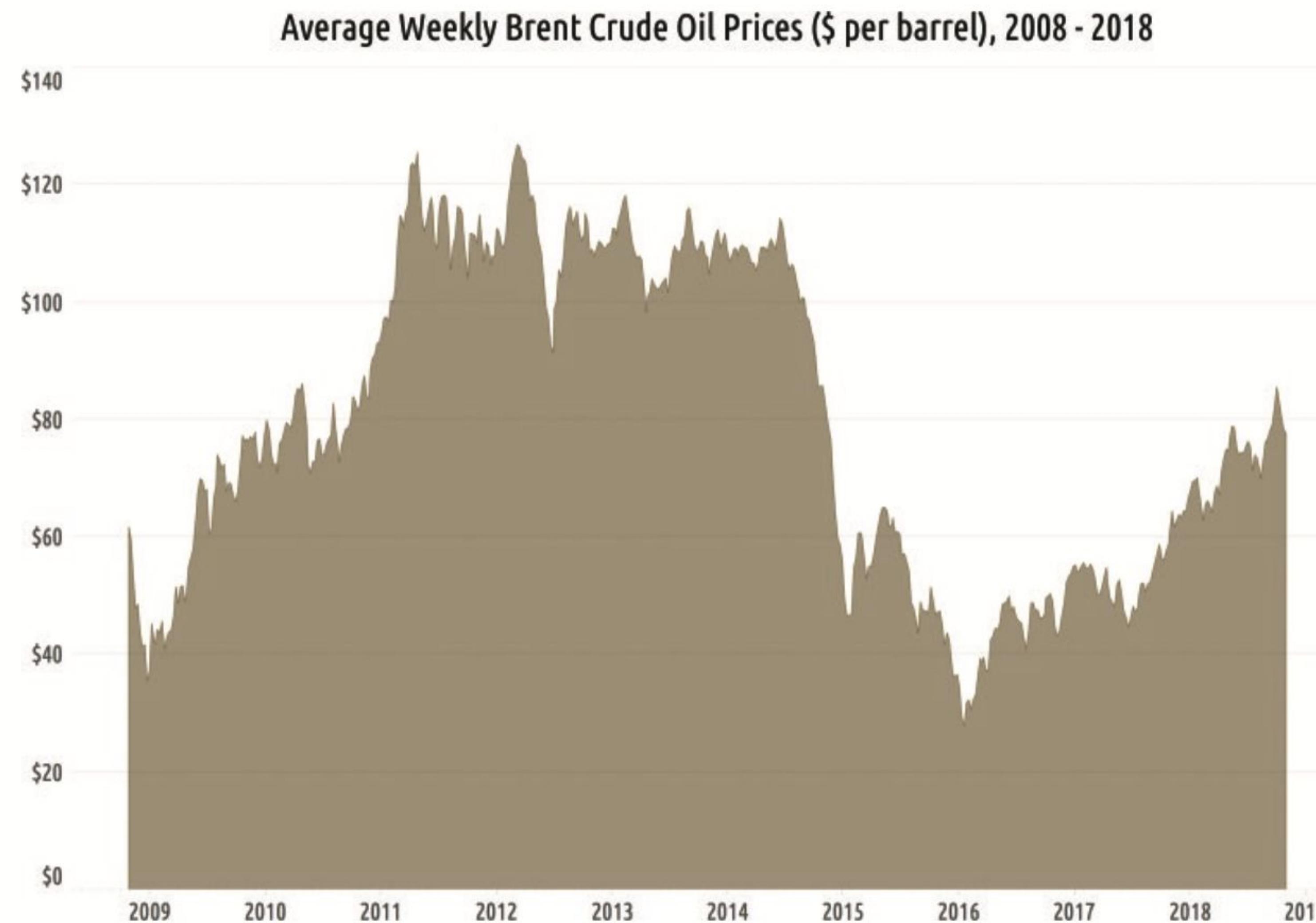
Quando o PIB cresce, nem sempre a desigualdade cai

O gráfico abaixo mostra o avanço do PIB comparado à evolução da desigualdade no Brasil desde 1980. Nem sempre o crescimento econômico levou a uma redução proporcional na disparidade de renda entre os mais pobres e os mais ricos



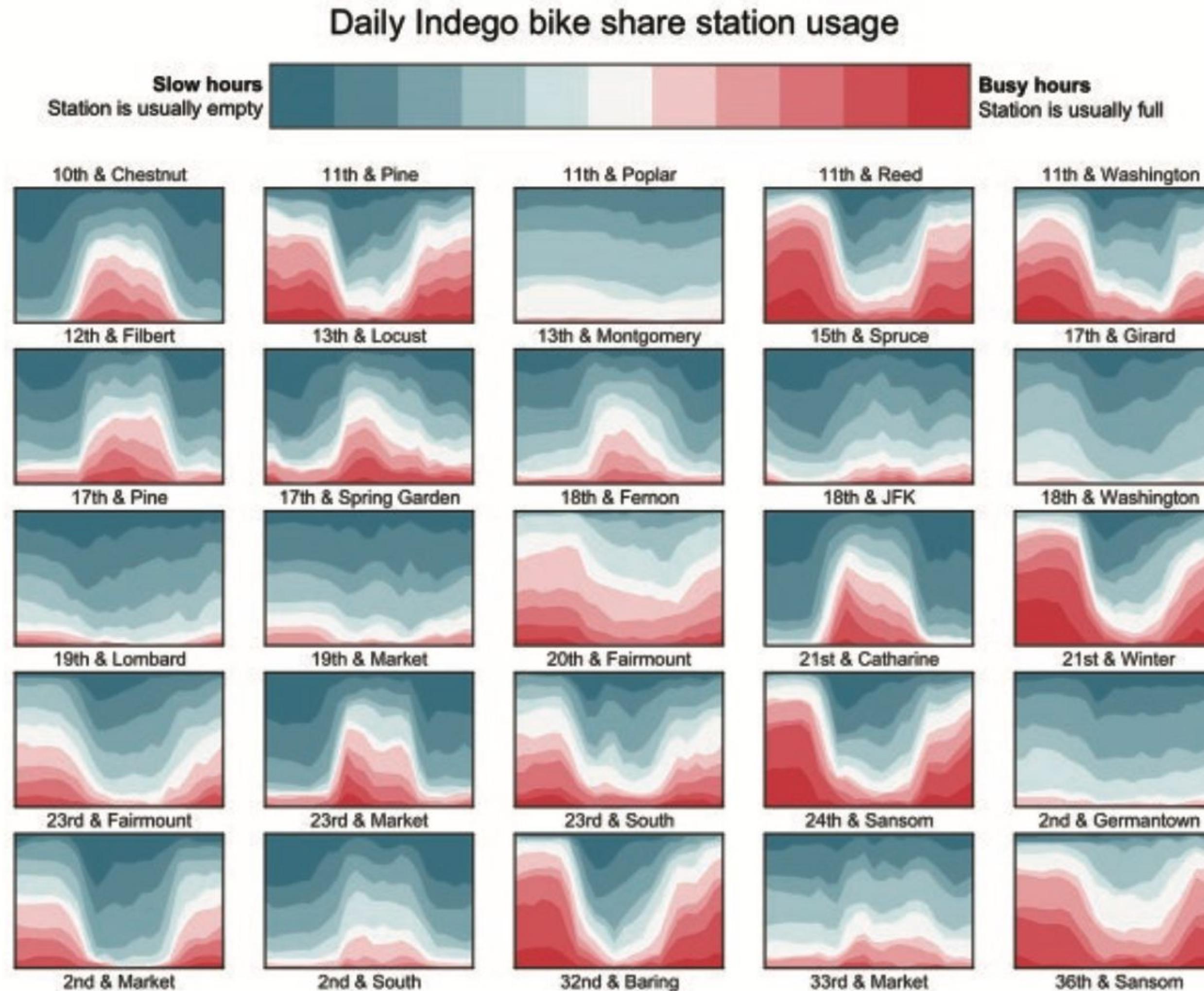
Area chart

Tendencias



Stacked area chart

Tendencias

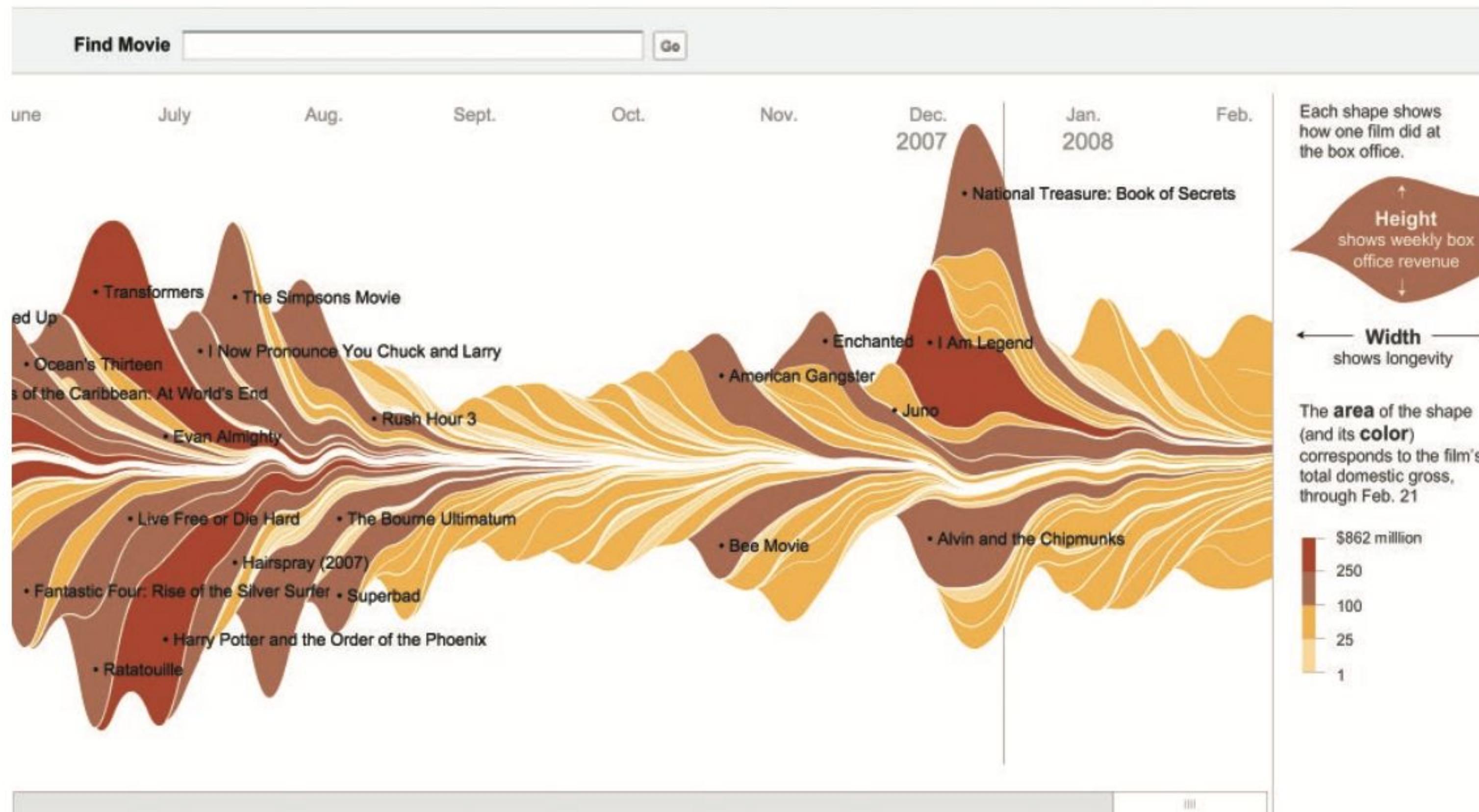


Stream graph

Tendencias

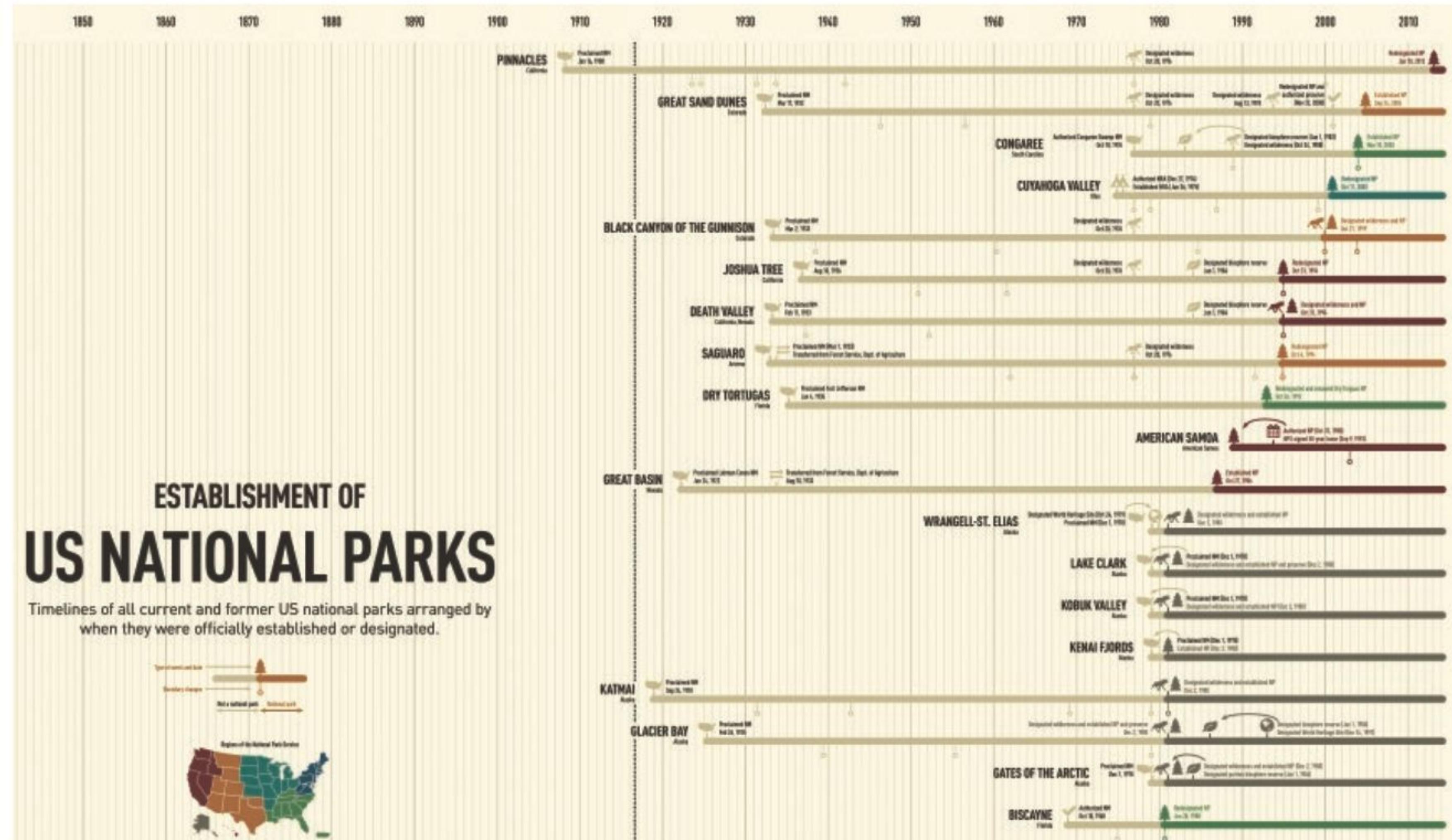
The Ebb and Flow of Movies: Box Office Receipts 1986 – 2008

Summer blockbusters and holiday hits make up the bulk of box office revenue each year, while contenders for the Oscars tend to attract smaller audiences that build over time. Here's a look at how movies have fared at the box office, after adjusting for inflation.



Gantt chart

Intervals

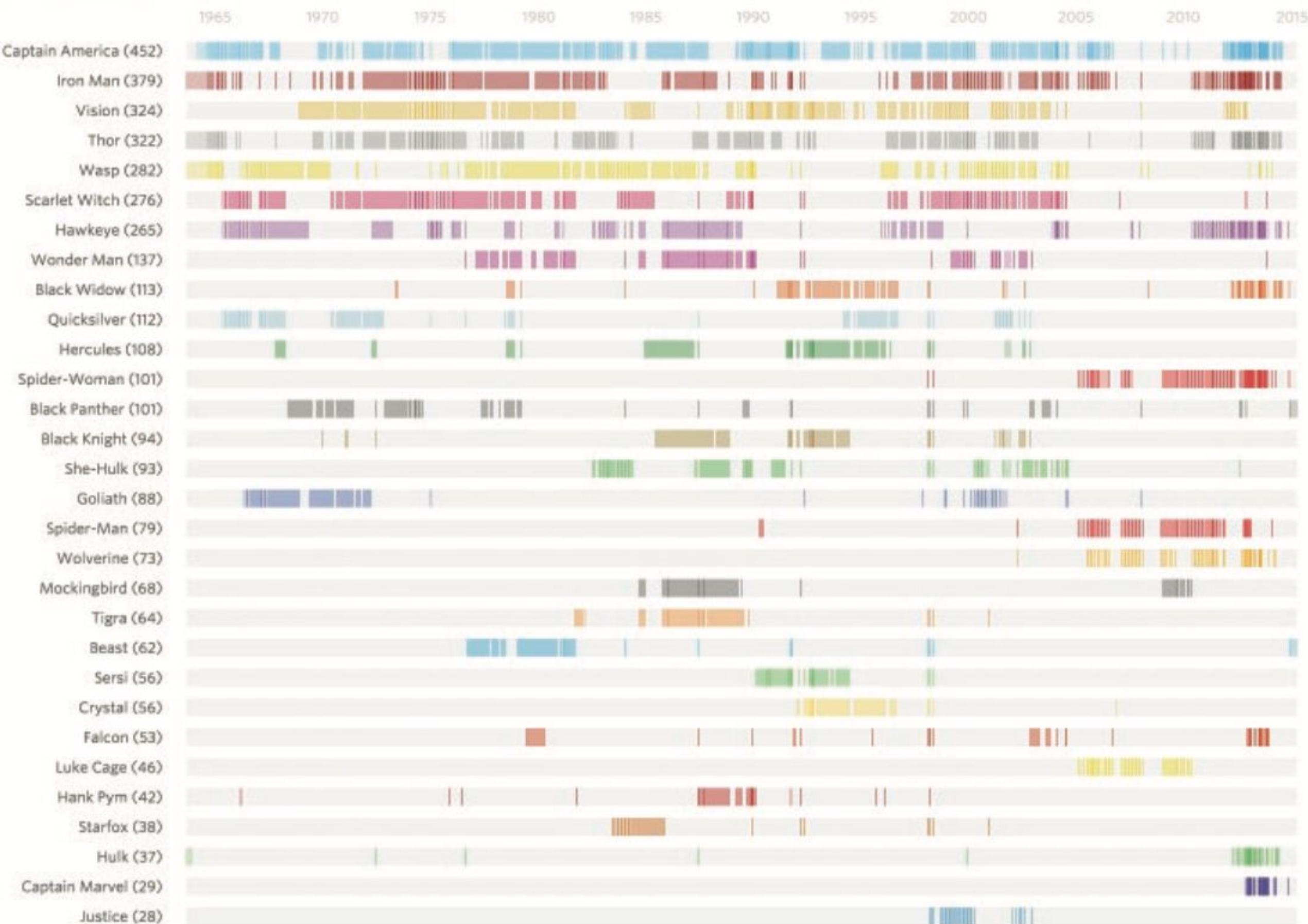


Instance chart

Actividades

'Avengers' characters' appearances over time

Avengers team members sorted by most number of appearances, across the 'Avengers' comic book titles in our analysis*. Each colored vertical stripe is an appearance in one of the issues as an Avenger.



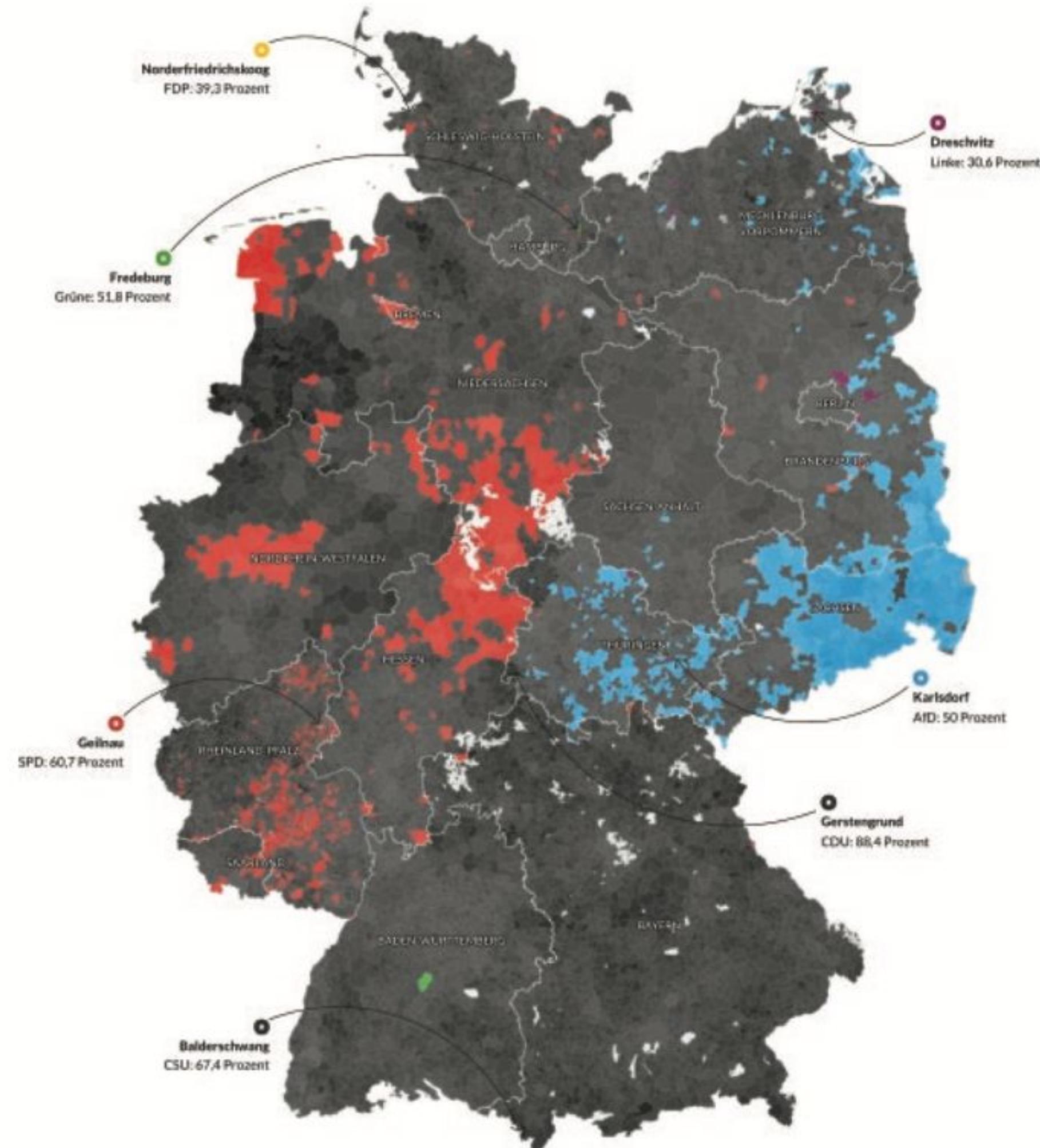
Choropleth map

BUNDESTAGSWAHL 2017 IM DETAIL

So haben die Wähler in den 11.000 Gemeinden gestimmt

Die interaktive Karte zeigt detaillierte Ergebnisse der Bundestagswahl 2017 – für jede Partei, in jedem Ort.

CDU/CSU SPD LINKE GRÜNE FDP AFD



Overlays

Isarithmic map (contour map)

Overlays

How Y'all, Youse and You Guys Talk

By JOSH KATZ and WILSON ANDREWS DEC. 21, 2013

What does the way you speak say about where you're from? Answer all the questions below to see your personal dialect map.

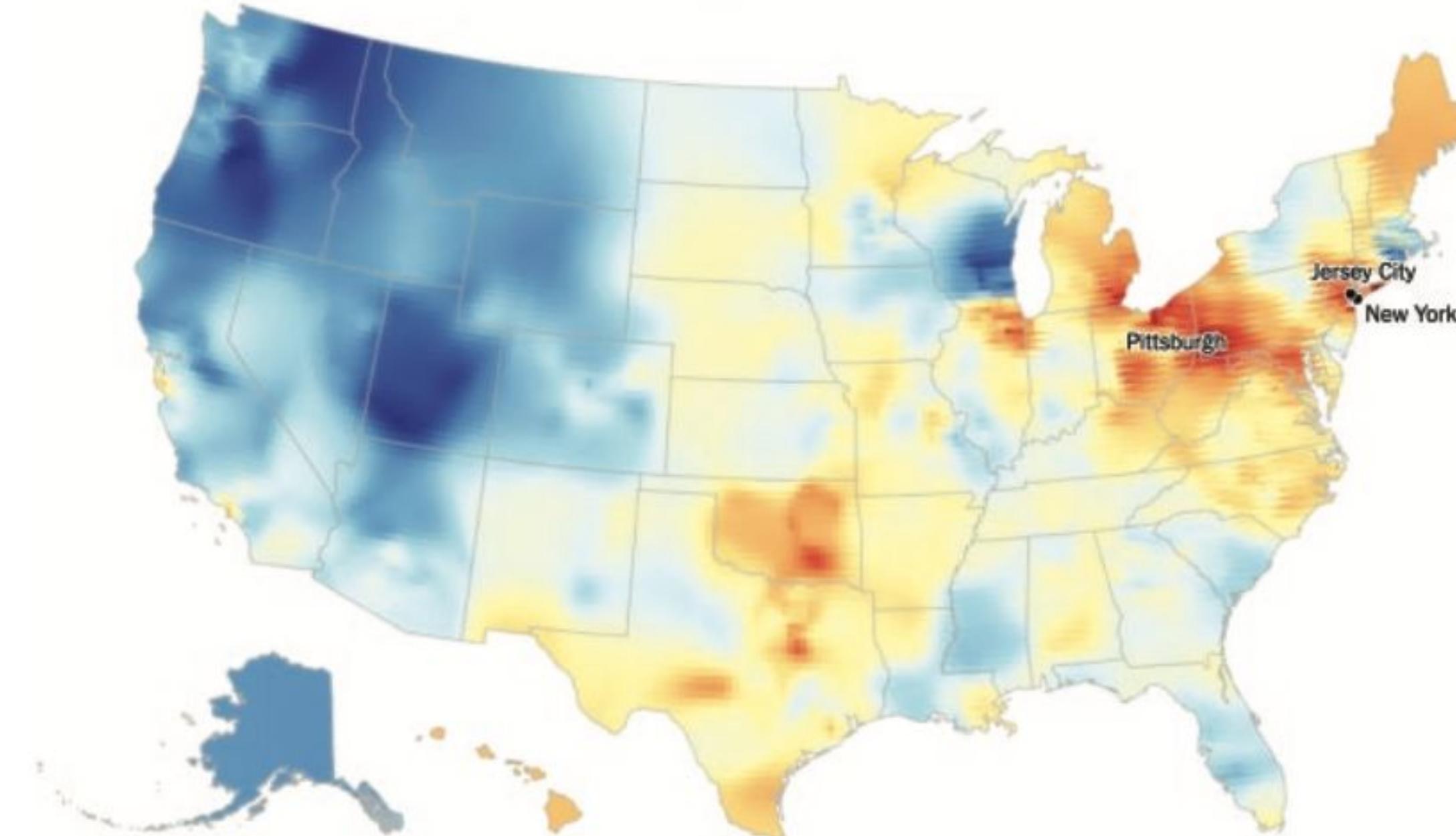
Your Map

See the pattern of your dialect in the map below. Three of the most similar cities are shown.

Least similar Most similar

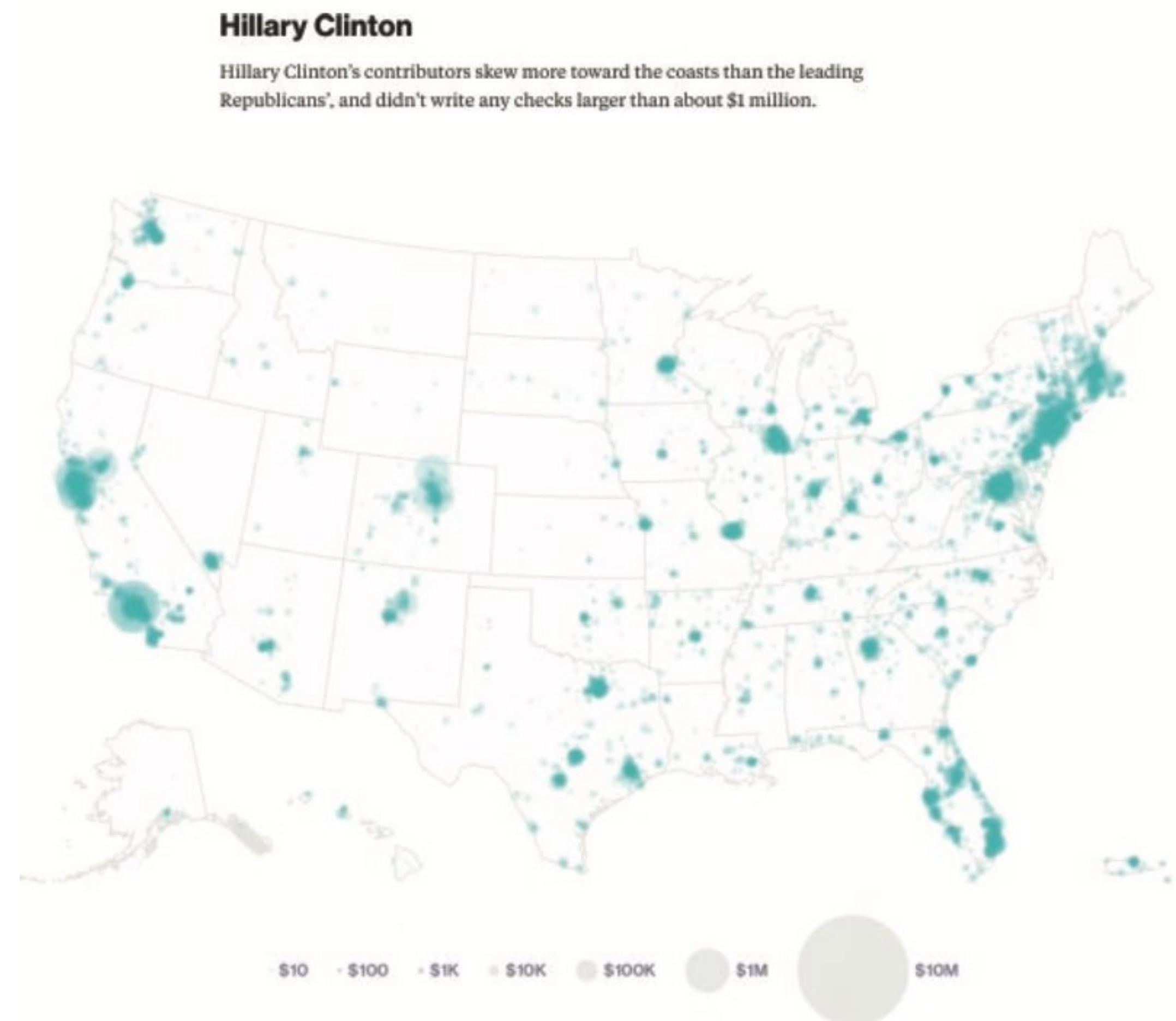
Show least similar

SHARE YOUR MAP: [f](#) [t](#) [g](#)

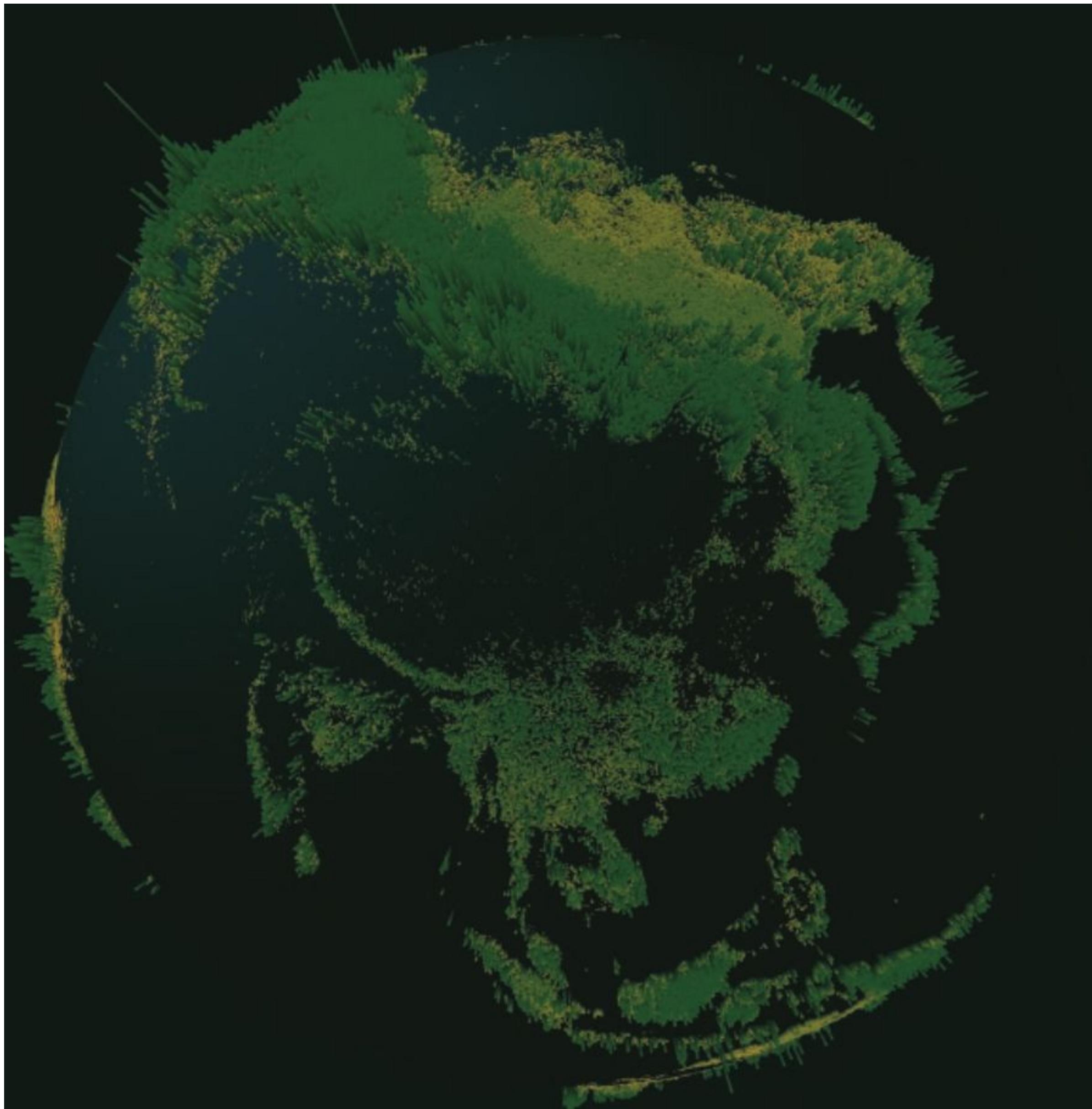


Proportional symbol map

Overlays



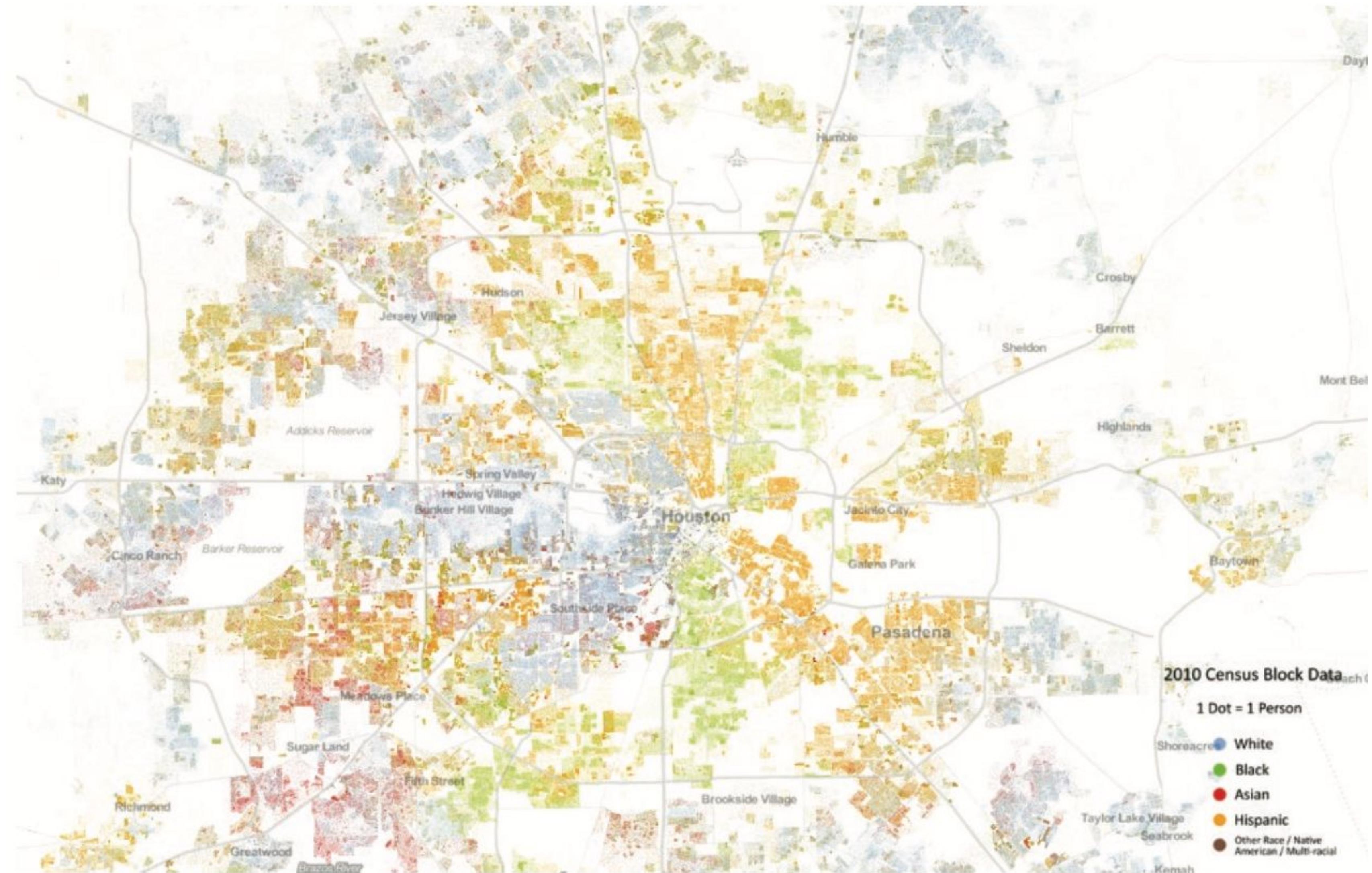
Prism map



Overlays

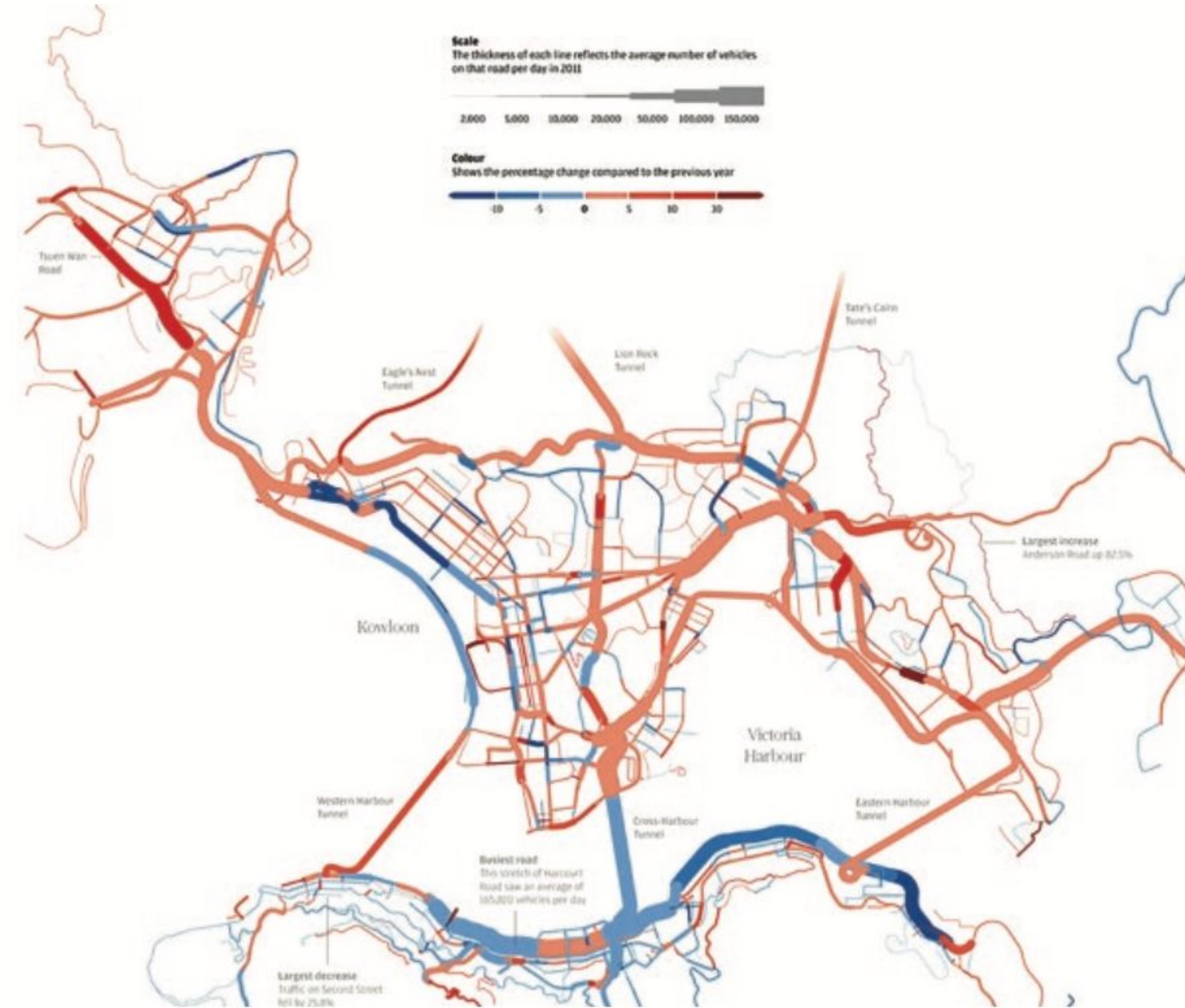
Dot map

Overlays



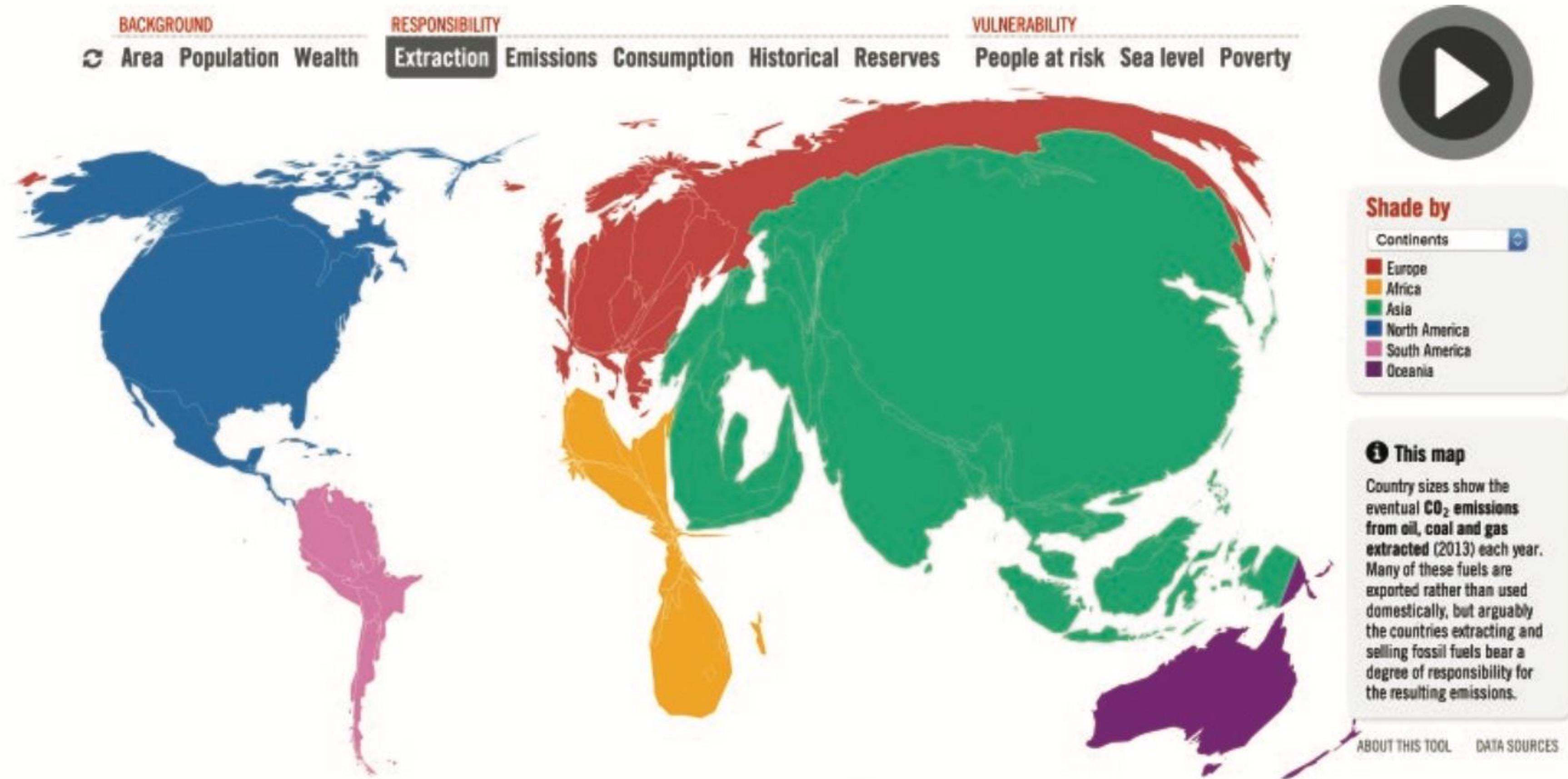
Flow map

Overlays



Area cartogram

Distorsión

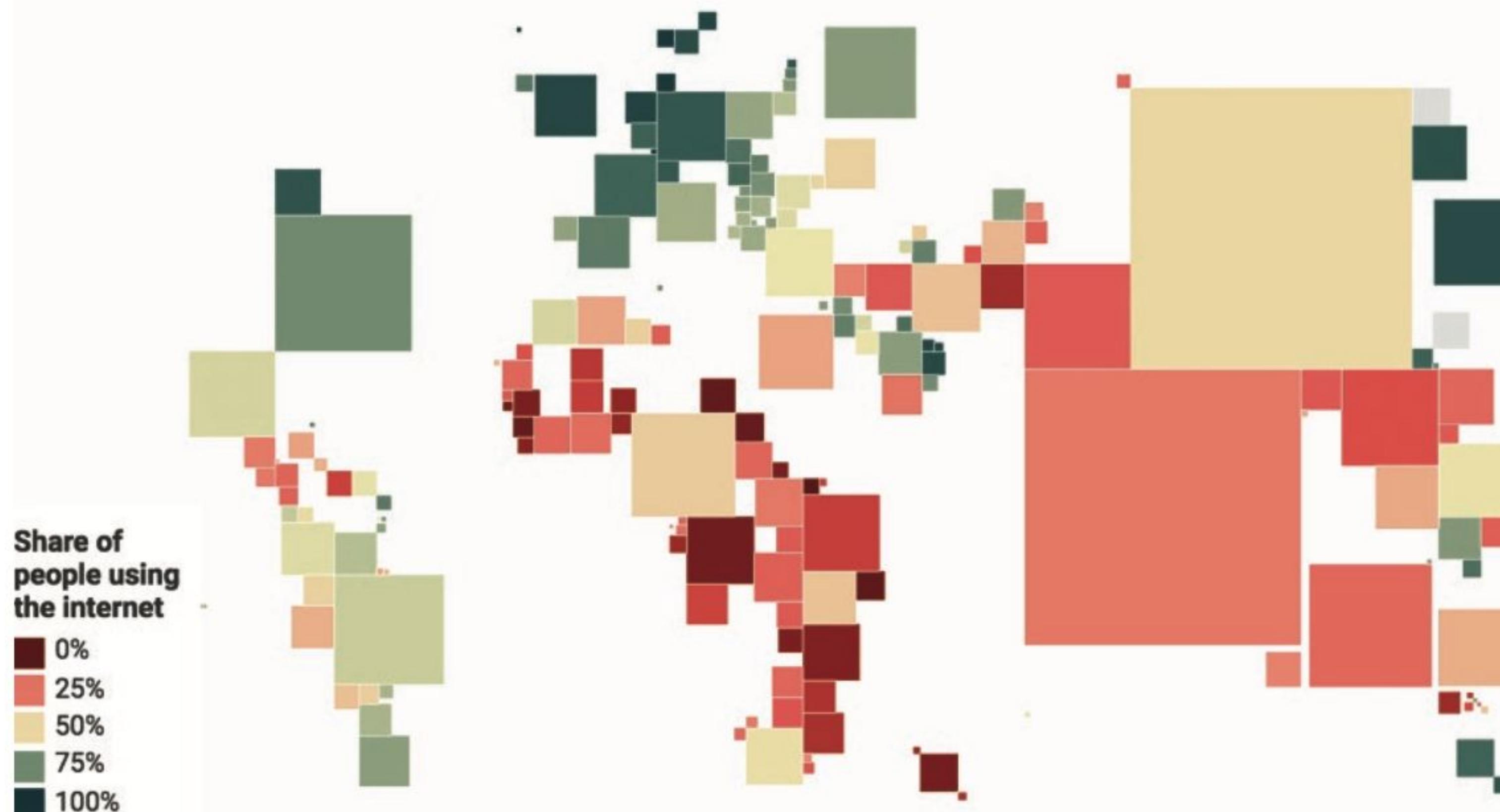


Dorling cartogram

Distorsión

Share of individuals using the internet, 2015

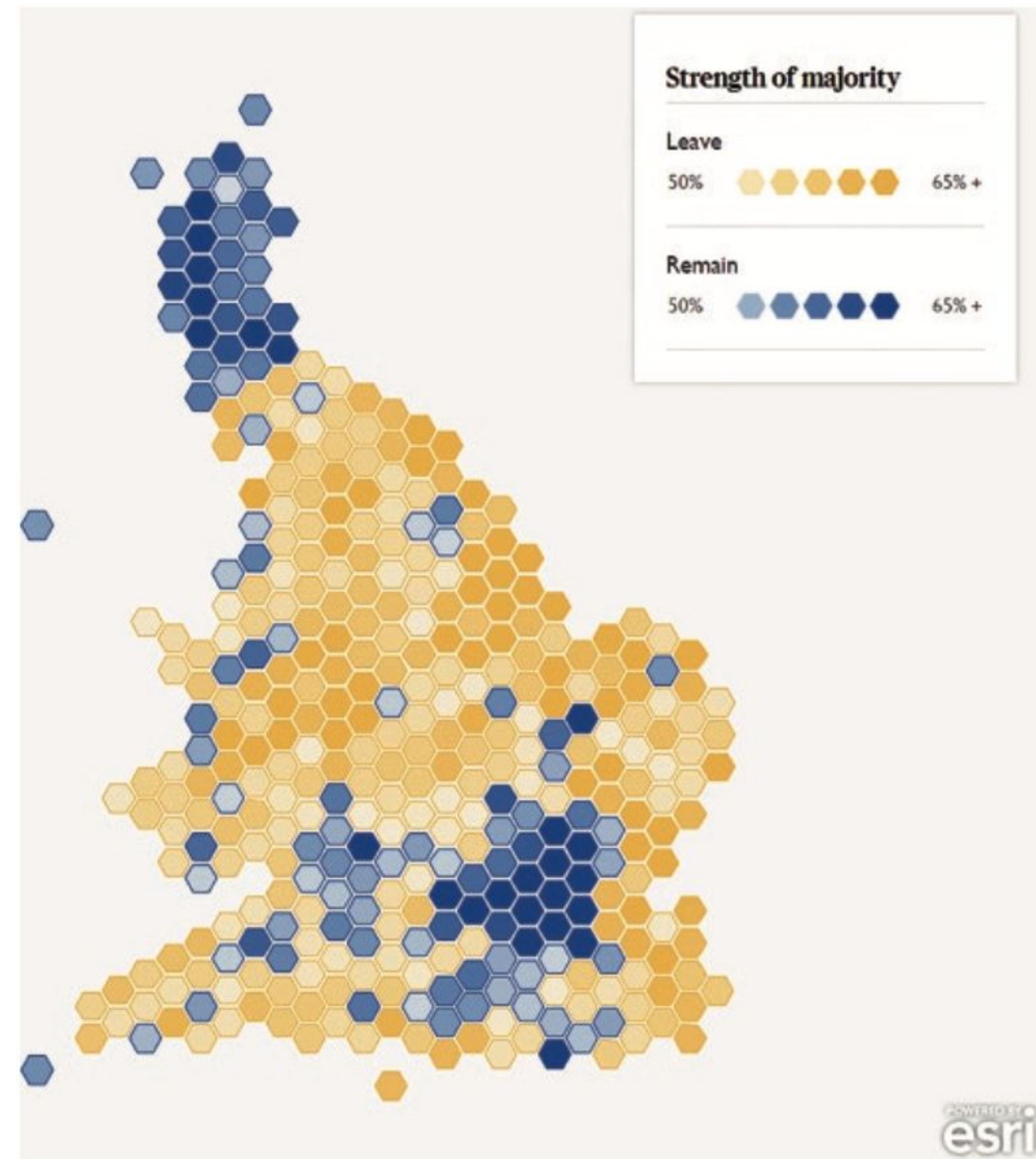
Share of individuals using the internet, measured as the percentage of the population. Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.



Source: [Our World in Data](#) • Get the data

Grid map

Distorsión



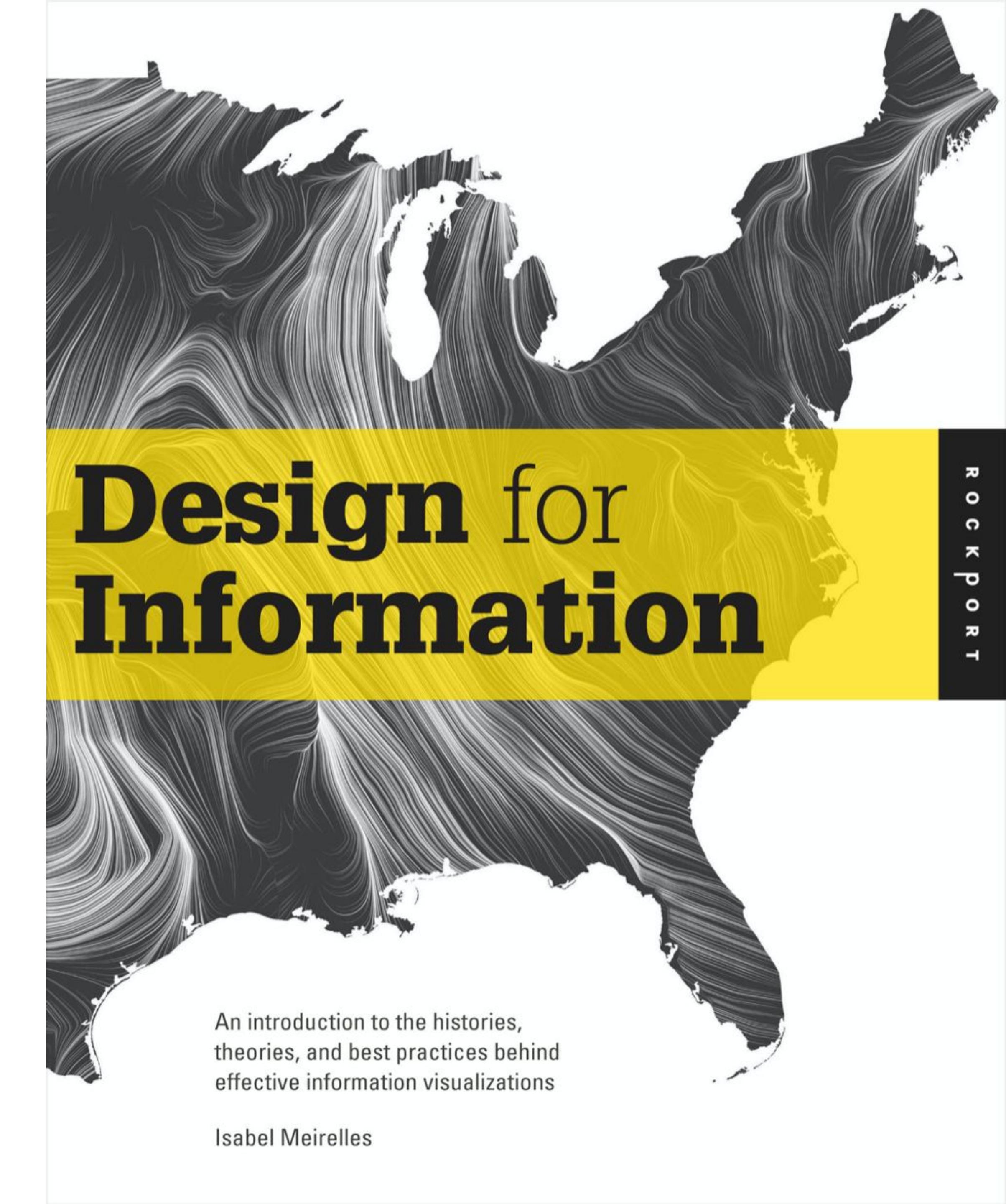
2nd Edition

Data Visualisation

A Handbook for Data Driven Design



Andy Kirk



Visualización en Python