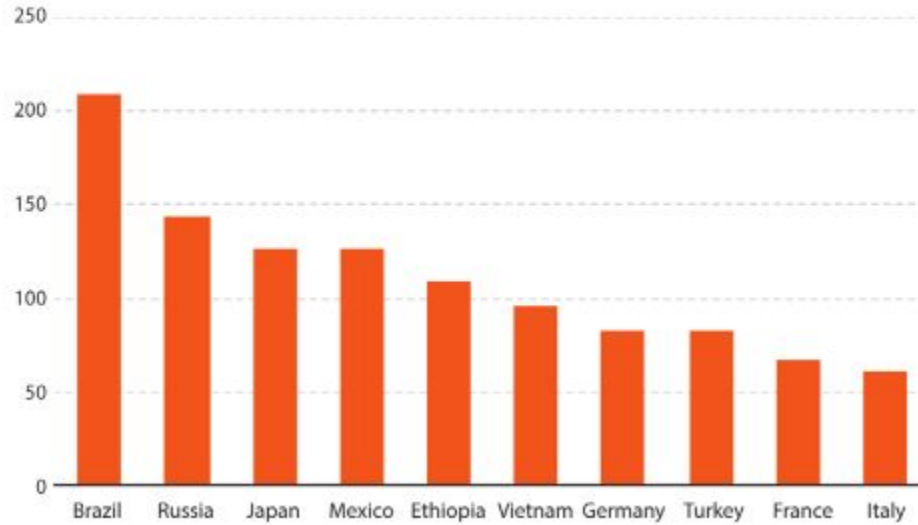


A gallery of plots

Comparing categories

The total population in Brazil exceeds that of other countries

(Millions of people)

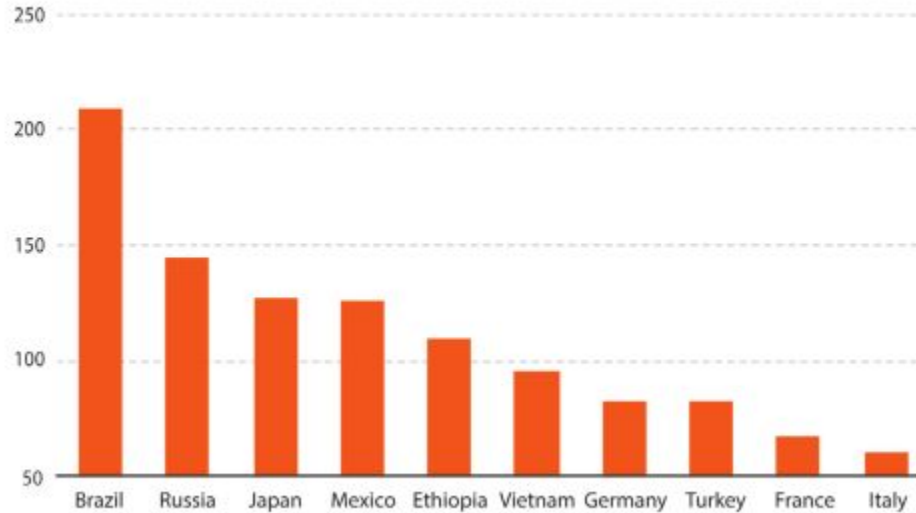


Source: The World Bank

Sorting categories by the displayed value is a good idea

The total population in Brazil exceeds that of other countries

(Millions of people)

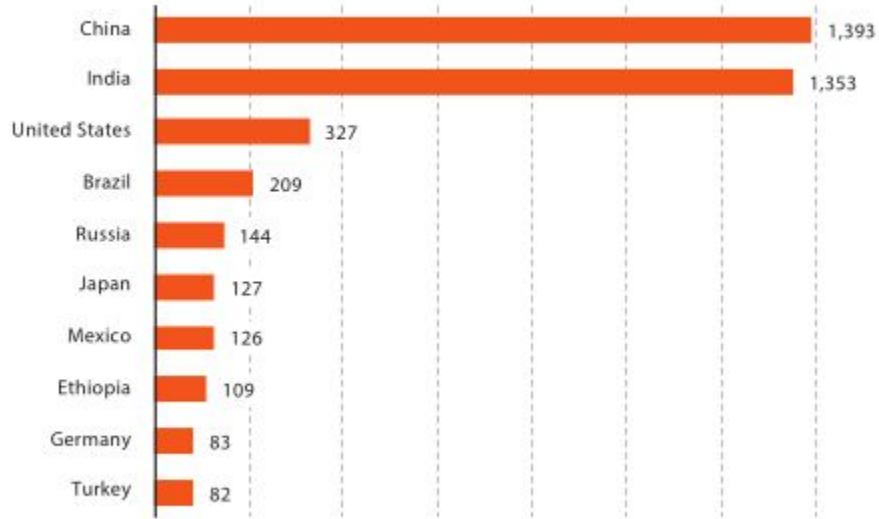


Source: The World Bank

The axis should *always* start at 0!

China and India are the most populous countries in the world

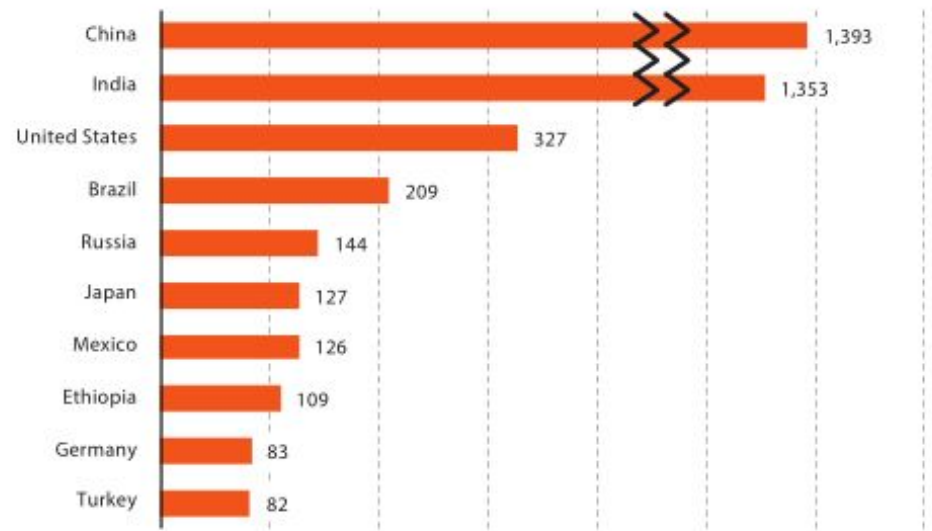
(Millions of people)



Source: The World Bank

China and India are the most populous countries in the world

(Millions of people)

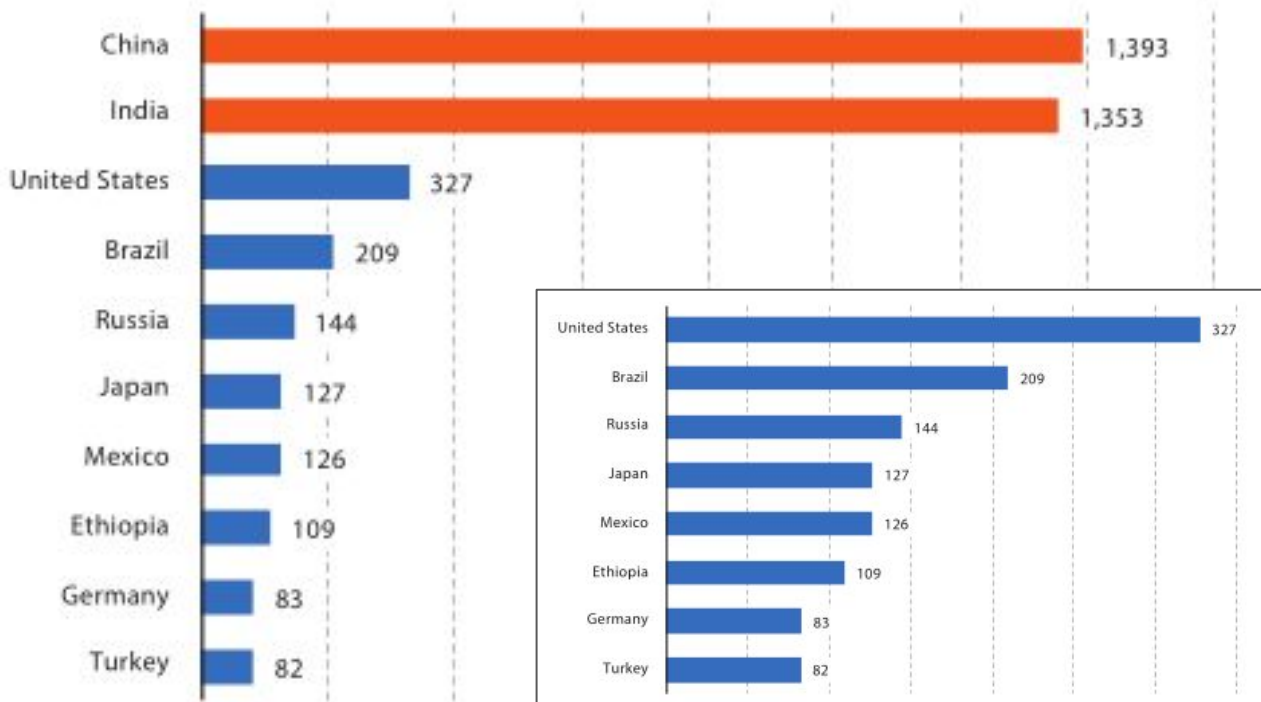


Source: The World Bank

Breaking the bars is never a good idea

China and India are the most populous countries in the world

(Millions of people)

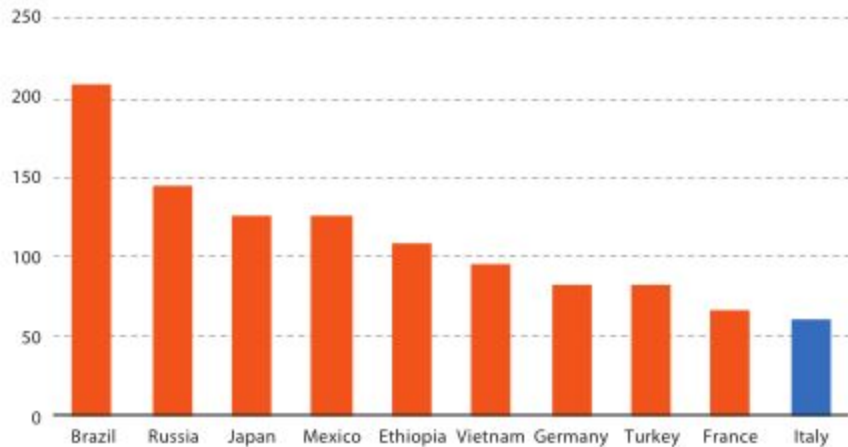


Source: The World Bank

In alternativa, usate due grafici

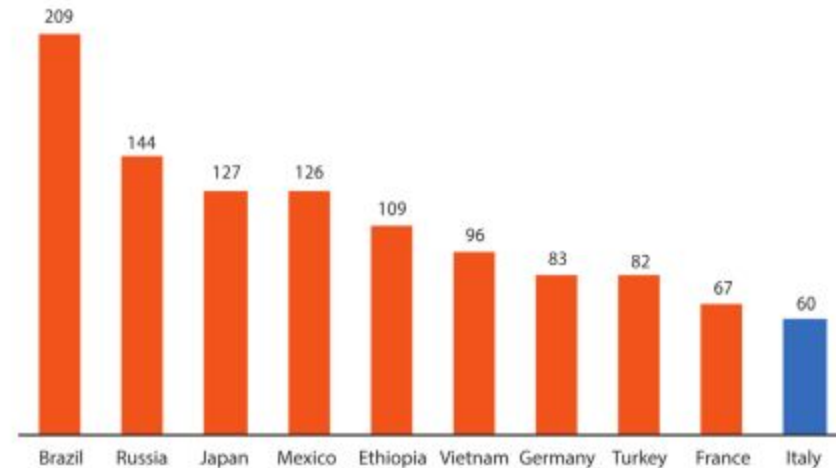
The total population in Brazil exceeds that of other countries

(Millions of people)



The total population in Italy is one-third that of Brazil

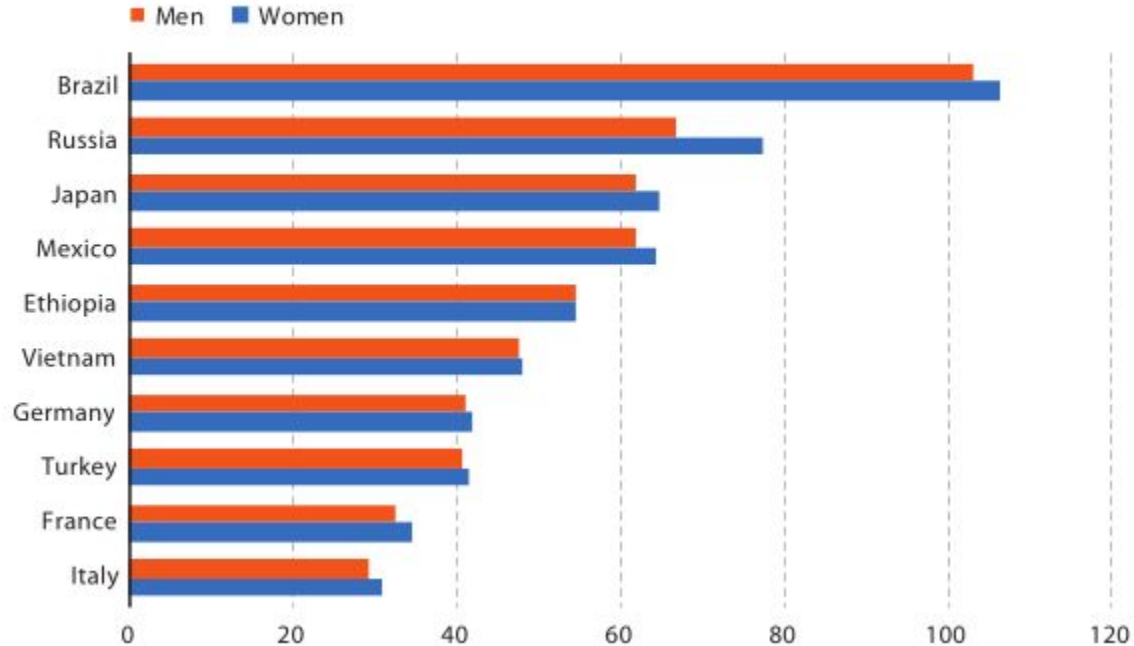
(Millions of people)



Consider replacing the grid lines with direct value labeling

There are more women than men in each country except for Ethiopia

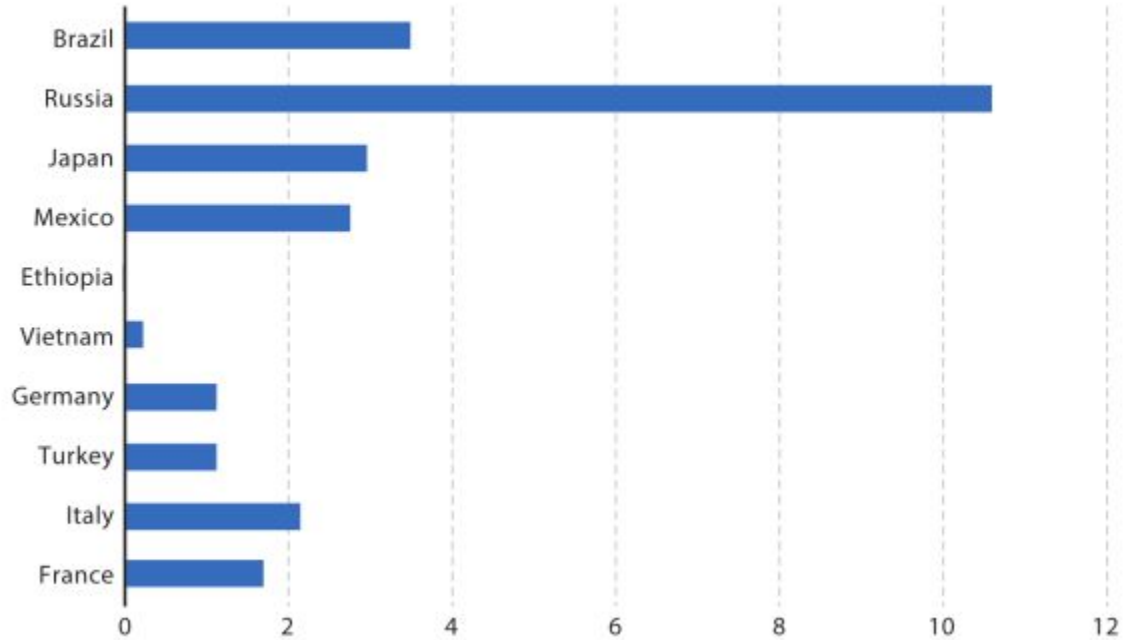
(Millions of people)



Paired bar charts favor comparing values within and between groups, but hide the total

Difference between the number of women and men

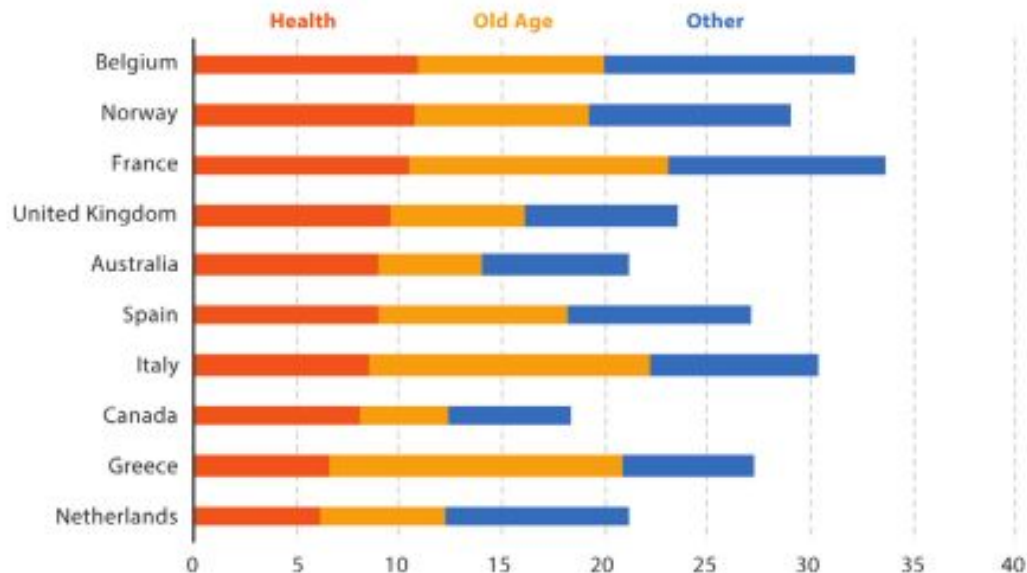
(Millions of people)



If the focus is on differences, consider showing those directly

Social expenditures for 10 OECD countries

(Percent of GDP)

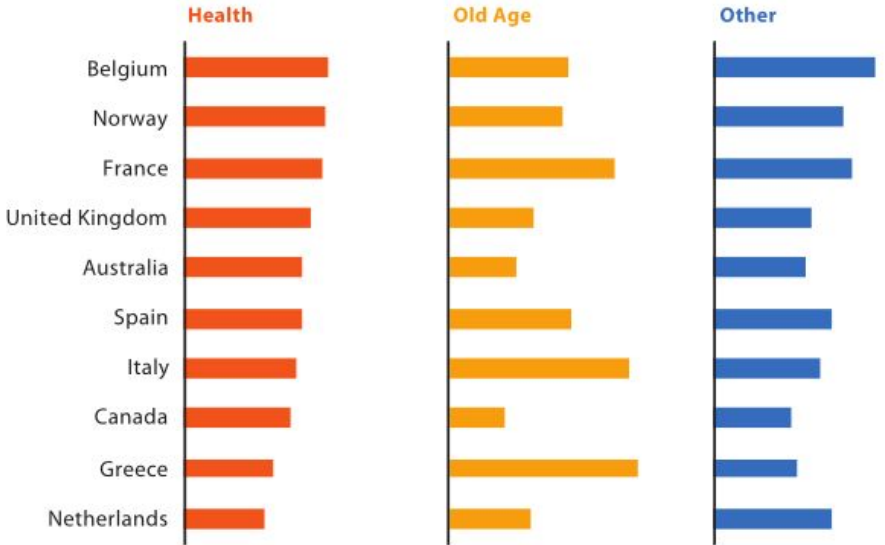


Source: Organisation for Economic Co-Operation and Development

Stacked bar charts favor comparing the totals

Social expenditures for 10 OECD countries

(Percent of GDP)



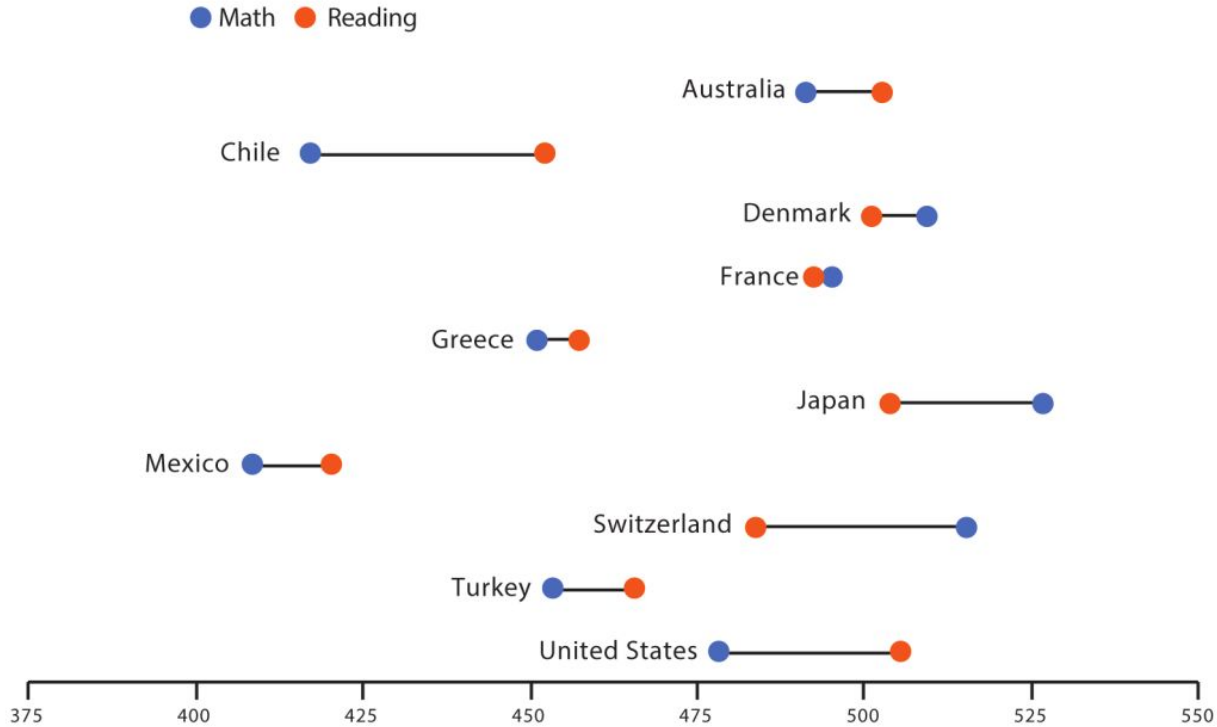
Social expenditures for 10 OECD countries

(Percent of GDP)



Faceting the components can help comparing the segments

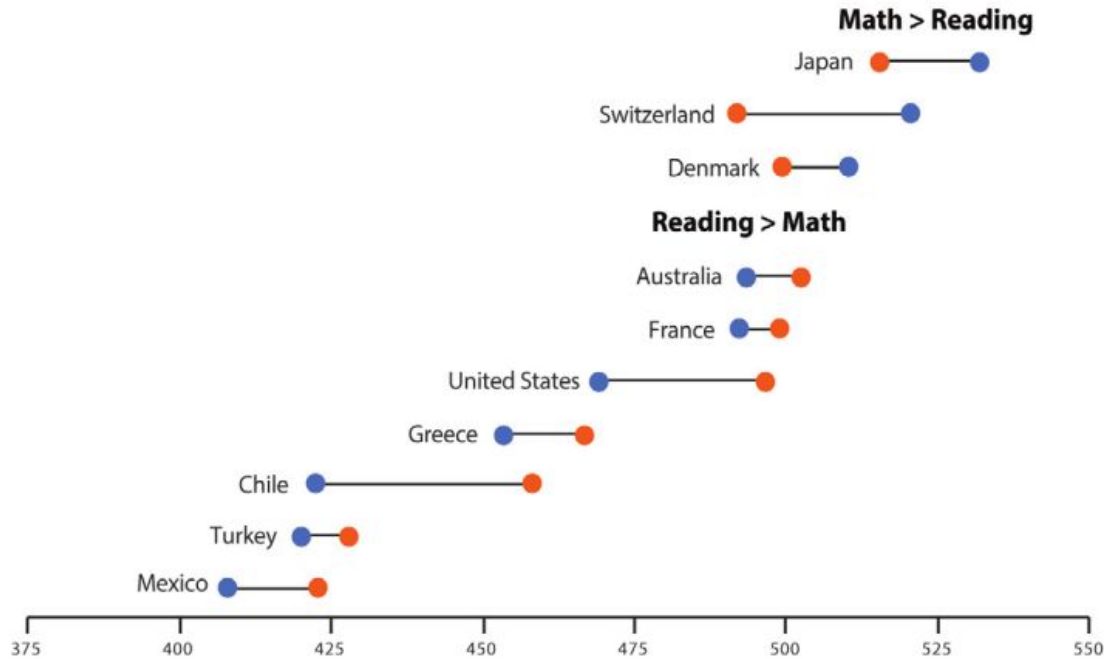
PISA scores for math and reading among 10 OECD countries



Source: Programme for International Student Assessment

Dot plots allow to represent differences between two categories

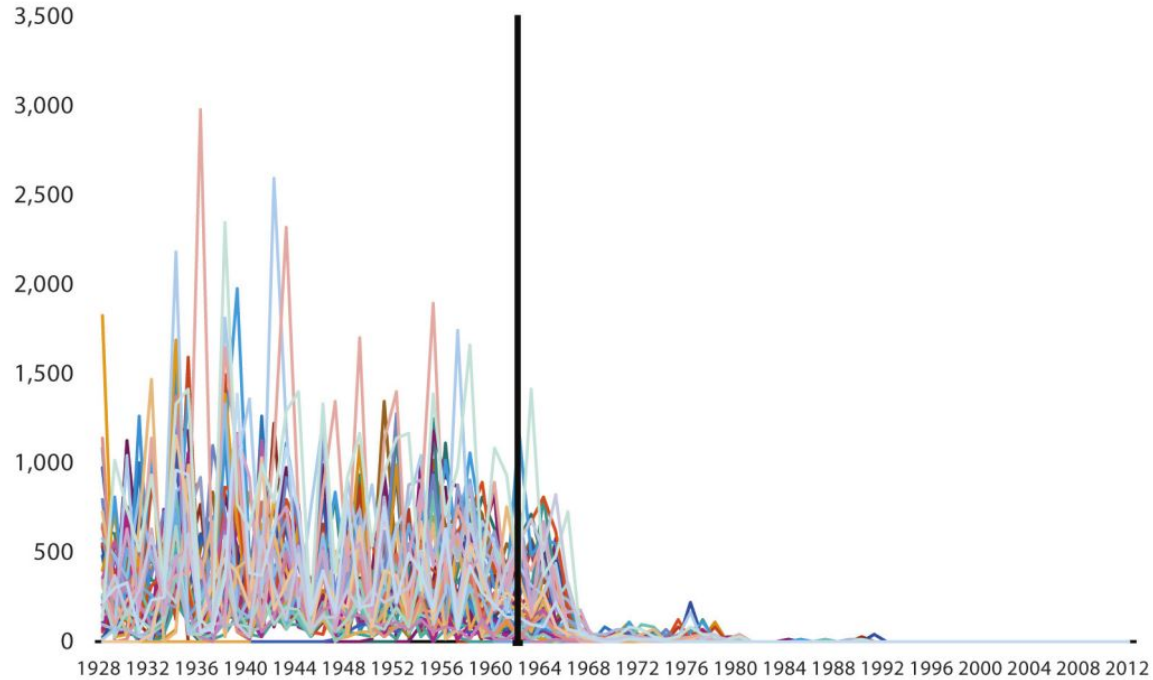
PISA scores for math and reading among 10 OECD countries



Source: Programme for International Student Assessment

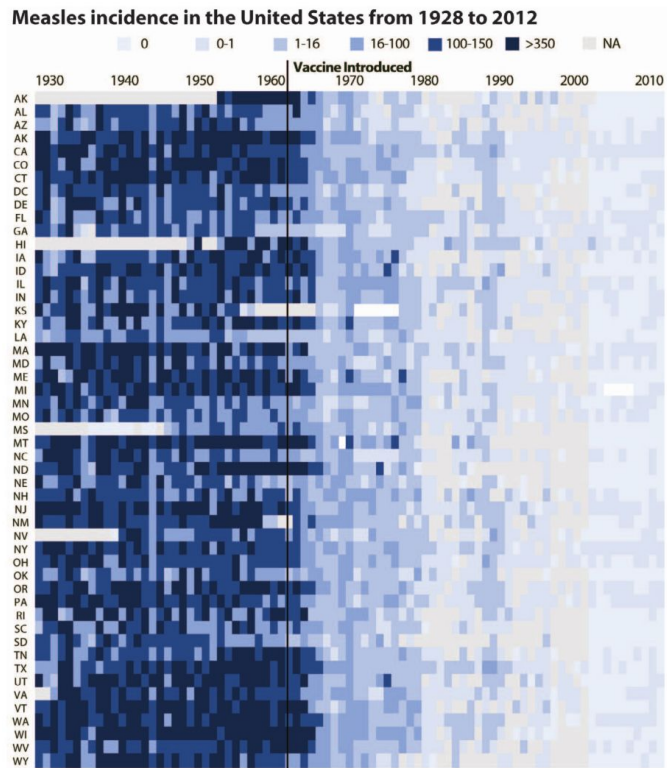
Labels and sorting further highlight the data

Measles incidence in the United States from 1928 to 2012



Source: Data from Project Tycho, <https://www.tycho.pitt.edu/data>

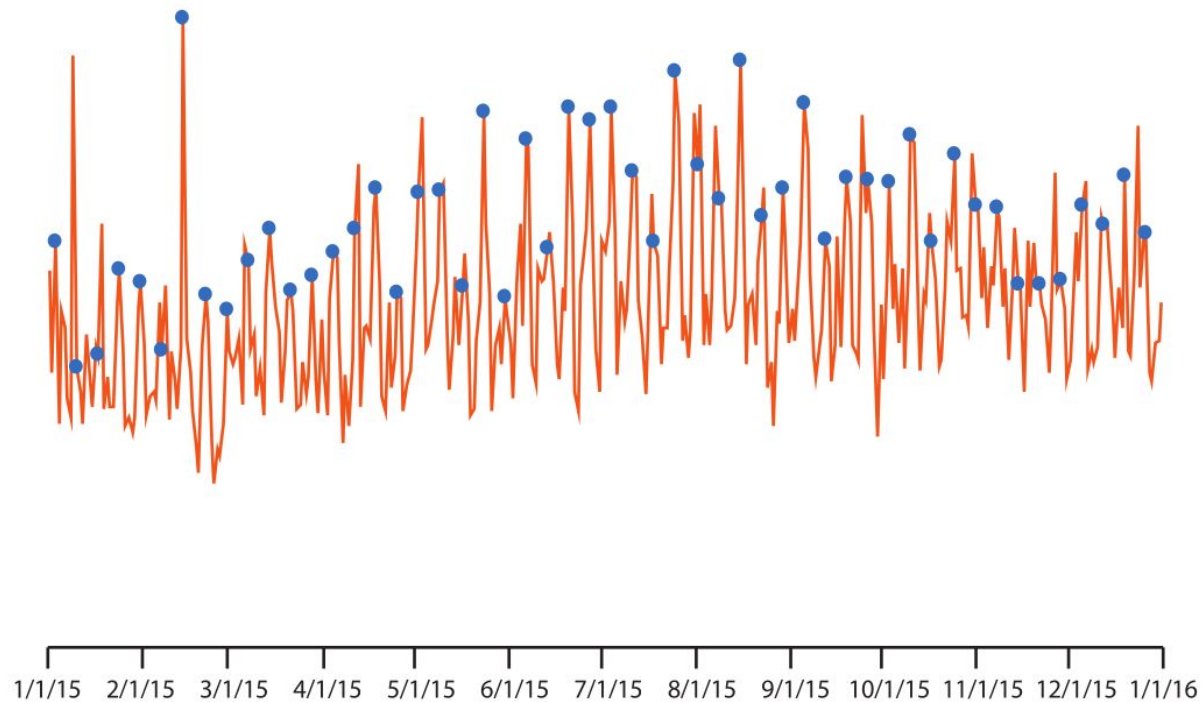
How can we improve this spaghetti mess?



Source: Project Tycho, <https://www.tycho.pitt.edu/data>

A heatmap can go a long way in showing patterns across many categories

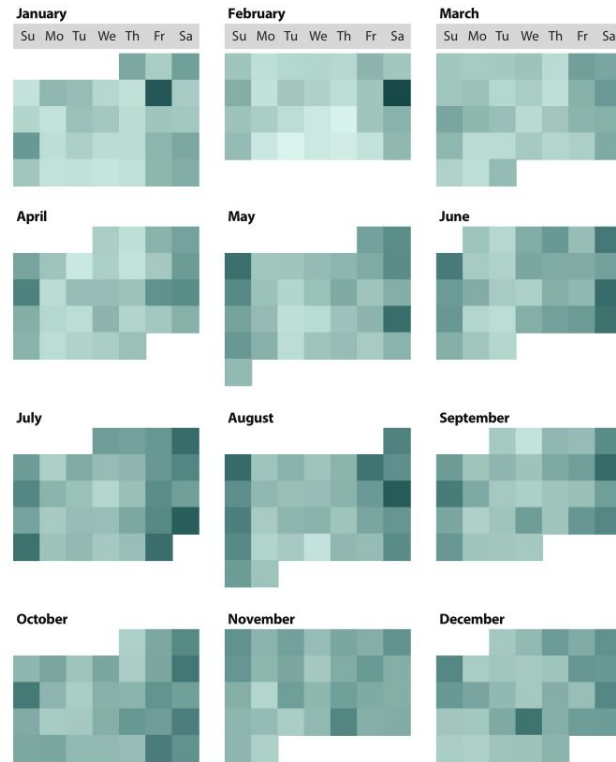
Vehicle fatalities in 2015



Source: National Highway Traffic Safety Administration

When are most fatalities happening?

Vehicle fatalities in 2015



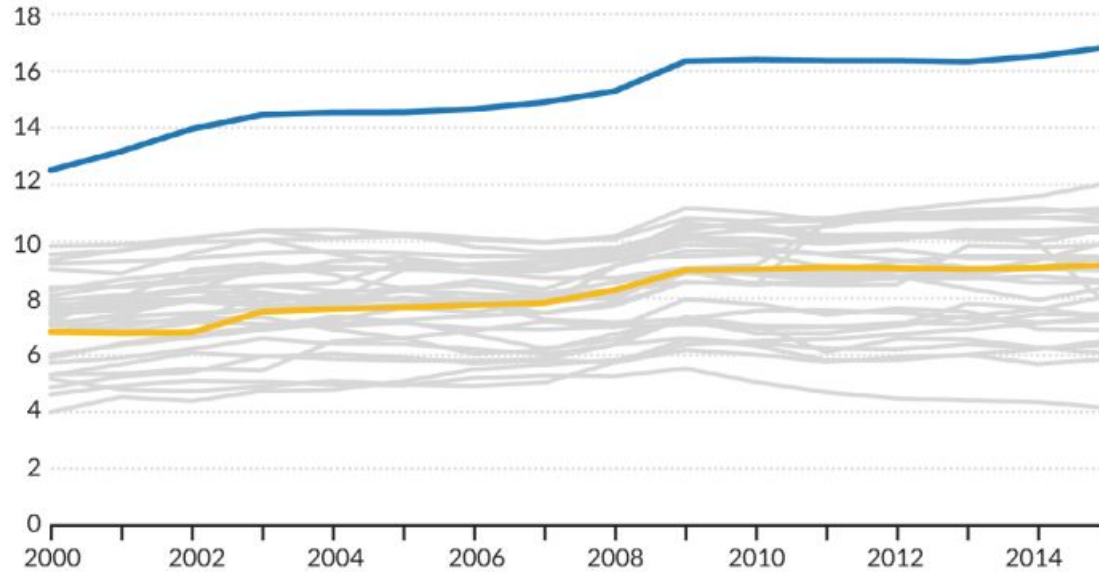
Source: National Highway Traffic Safety Administration
Note: Inspired by Nathan Yau at FlowingData.com

A heatmap arranged as a calendar shows that weekends see more fatalities.

Time-evolving data

Total health care spending in the United States and Germany increased between 2000 and 2015

(Percent of GDP)

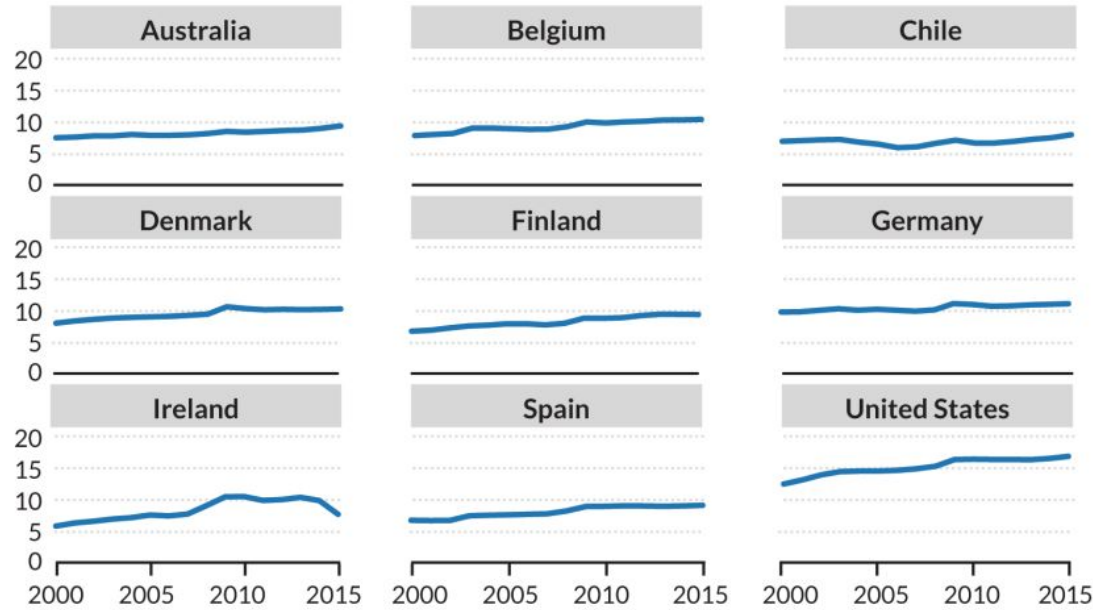


Source: The World Bank

Many lines can provide the background for a few highlighted ones

Health care spending across major countries has largely increased since 2000

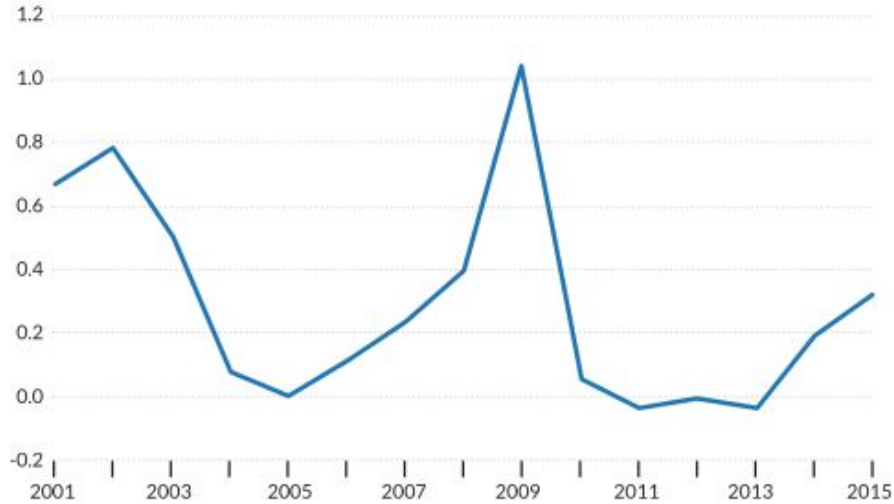
(Percent of GDP)



Source: The World Bank

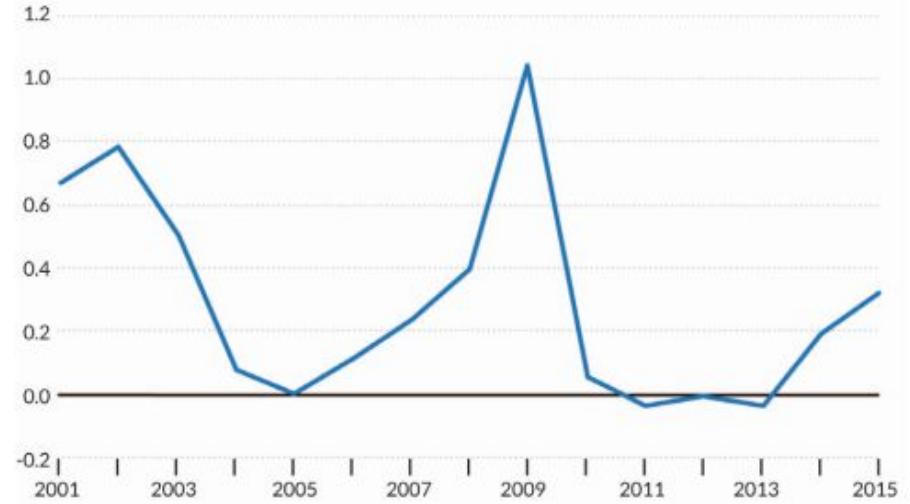
Faceting helps in breaking down dense line charts

Year-to-year change in U.S. health care spending: Zero axis not marked
(Percent of GDP)



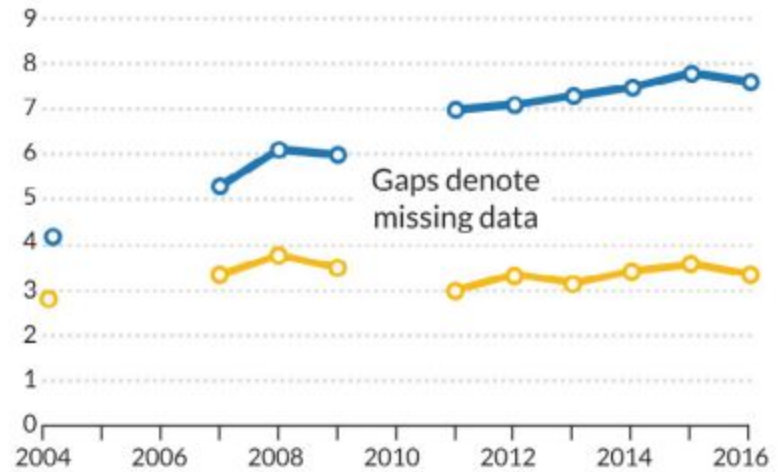
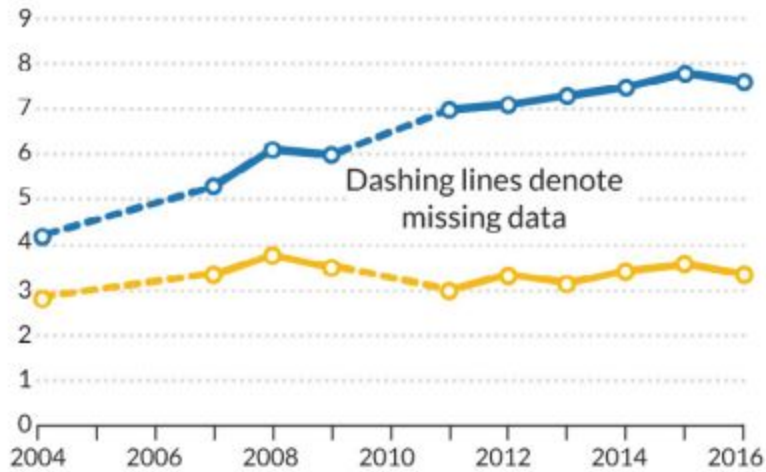
Source: The World Bank

Year-to-year change in U.S. health care spending: Zero axis marked
(Percent of GDP)



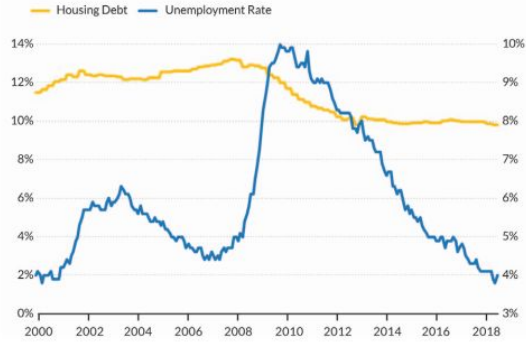
Source: The World Bank

We are inclined in thinking the bottom is the zero:
so mark explicitly when this is not the case

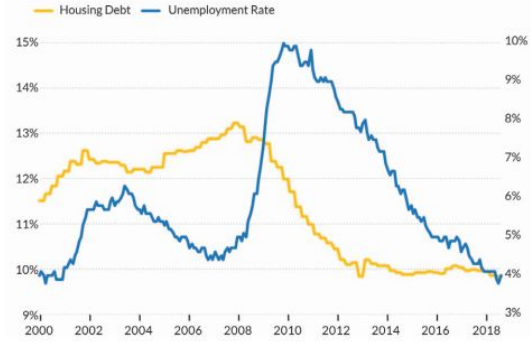


Use either gaps or other line styles to denote missing data

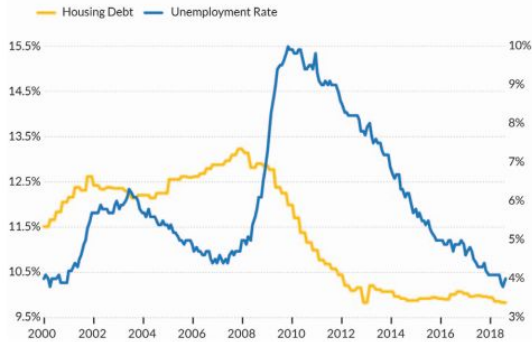
Economic climate for consumers in 2017 and 2018 was quite good



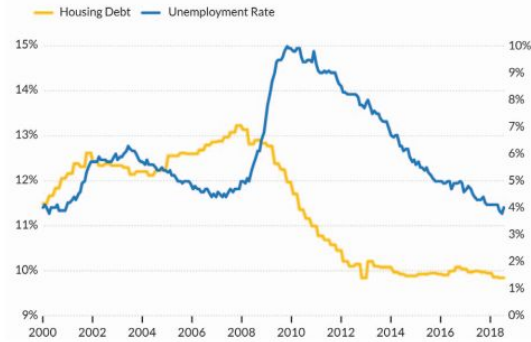
Economic climate for consumers in 2017 and 2018 was quite good



Economic climate for consumers in 2017 and 2018 was quite good

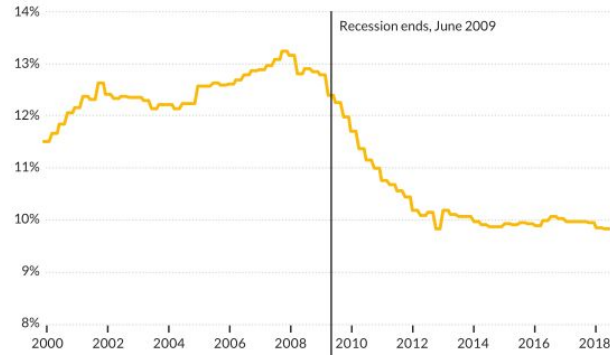


Economic climate for consumers in 2017 and 2018 was quite good

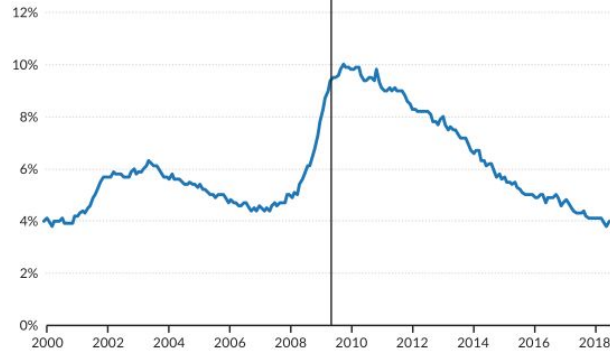


Another issues with dual-axis plots is that we are drawn to the intersection of the line, whose position is arbitrary, depending on the scales.

Housing debt in the United States has declined since 2008



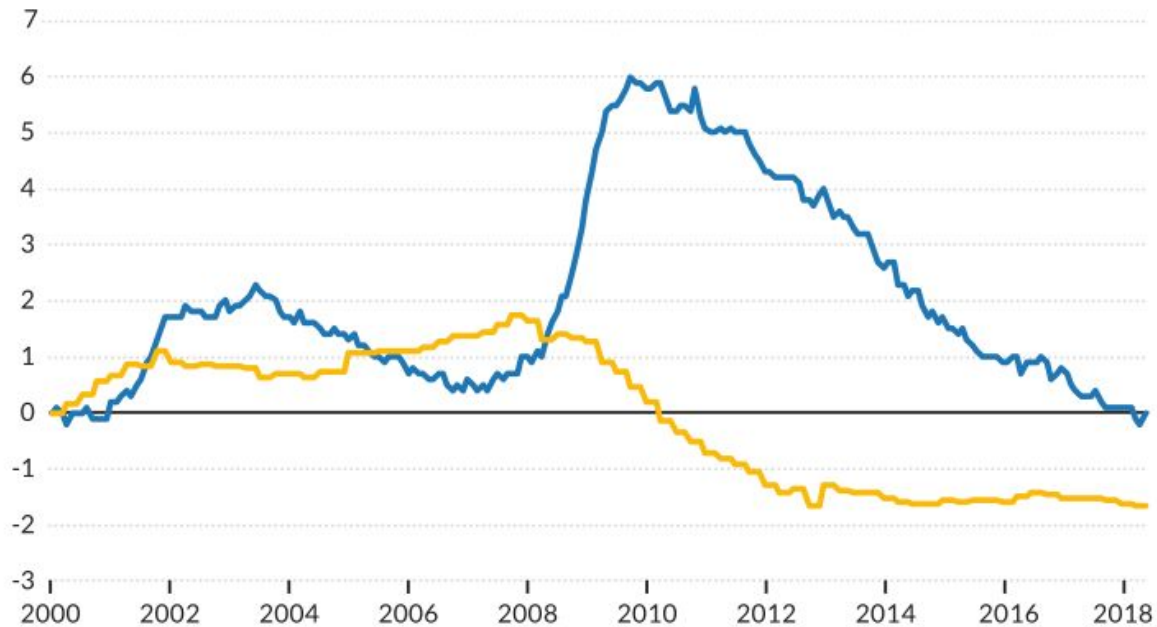
The unemployment rate has declined since about 2010



Solution: produce two plots aligned vertically,
which allows to put annotation markers

The economic climate for consumers in 2017 and 2018 was quite good

(Percent pointchange since 2000)

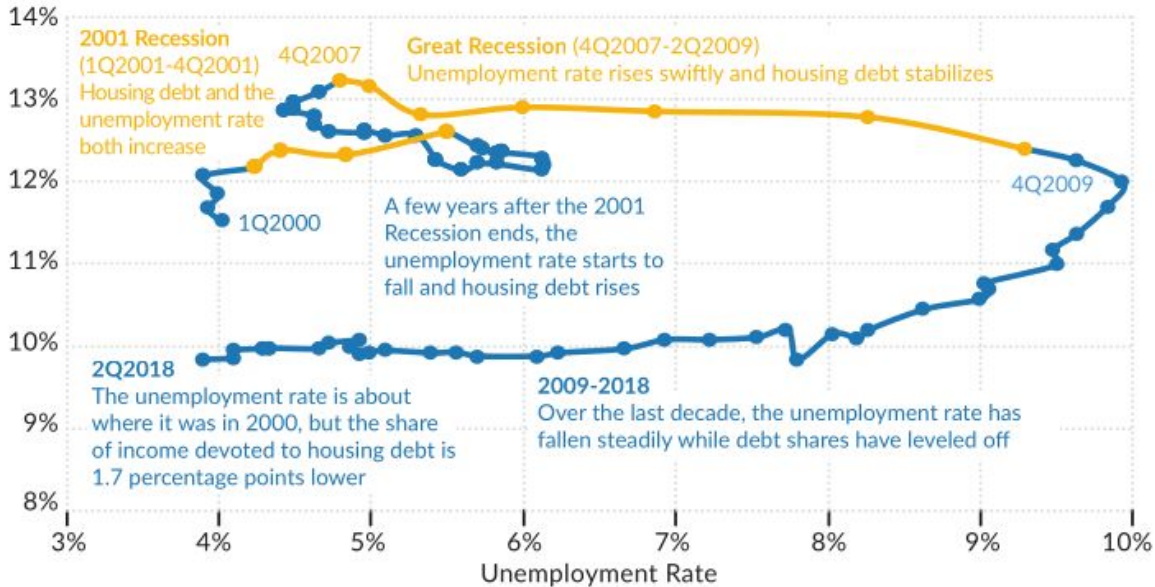


Source: Household debt service ratio, Federal Reserve Board of Governors; Unemployment rate, Bureau of Labor Statistics. Unemployment rate averaged to quarters.

Solution: normalize the values, or consider the percent point change

The U.S. economy appears supportive of the consumer with low-unemployment rate and housing debt

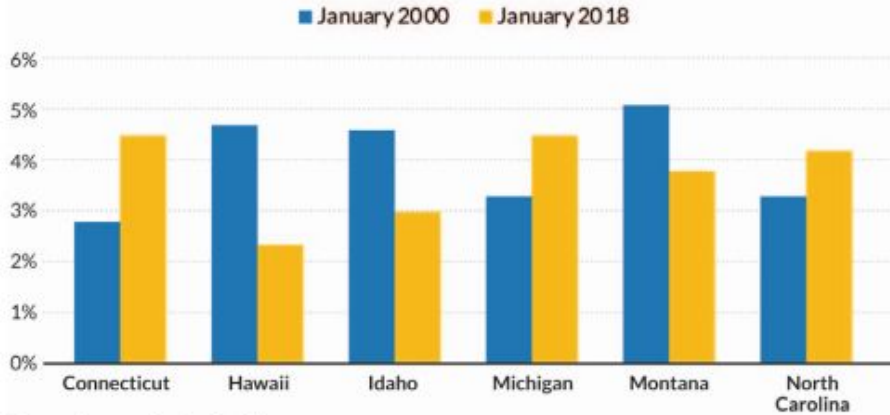
(Household debt service ratio)



Source: Household debt service ratio, Federal Reserve Board of Governors; Unemployment rate, Bureau of Labor Statistics. Unemployment rate averaged to quarters.

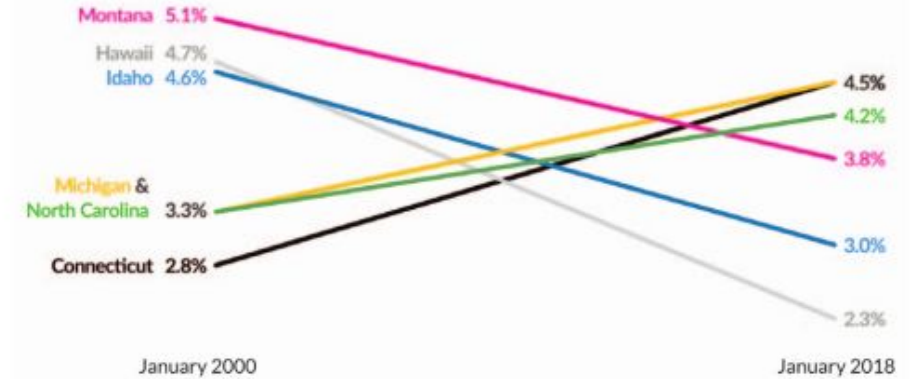
Solution: connected scatter plot

Biggest changes in the unemployment rate between, January 2000 and January 2018



Source: Bureau of Labor Statistics

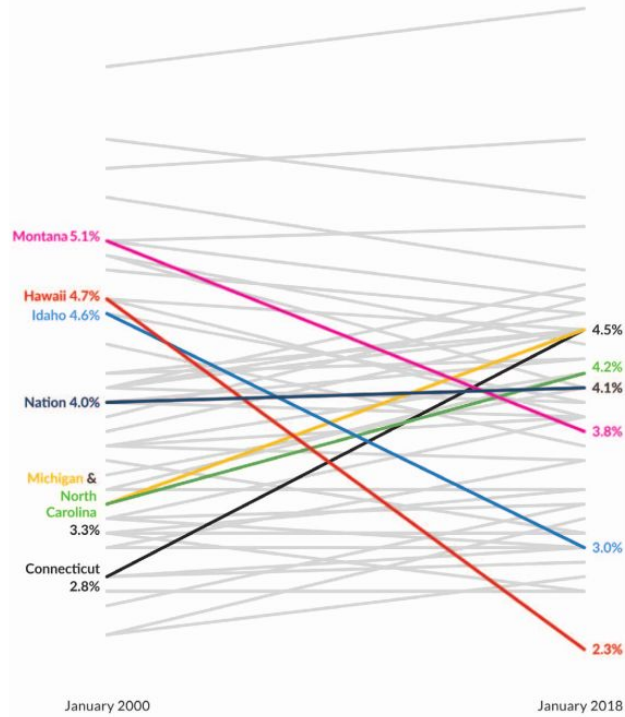
Biggest changes in the unemployment rate between, January 2000 and January 2018



Source: Bureau of Labor Statistics

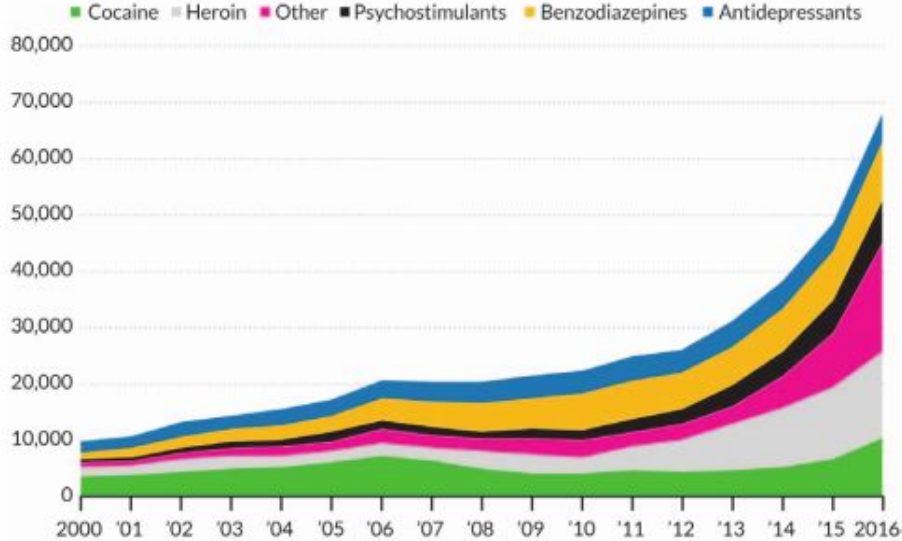
A slope graph allows to make several comparisons simultaneously

Biggest changes in the unemployment rate
between January 2000 and January 2018



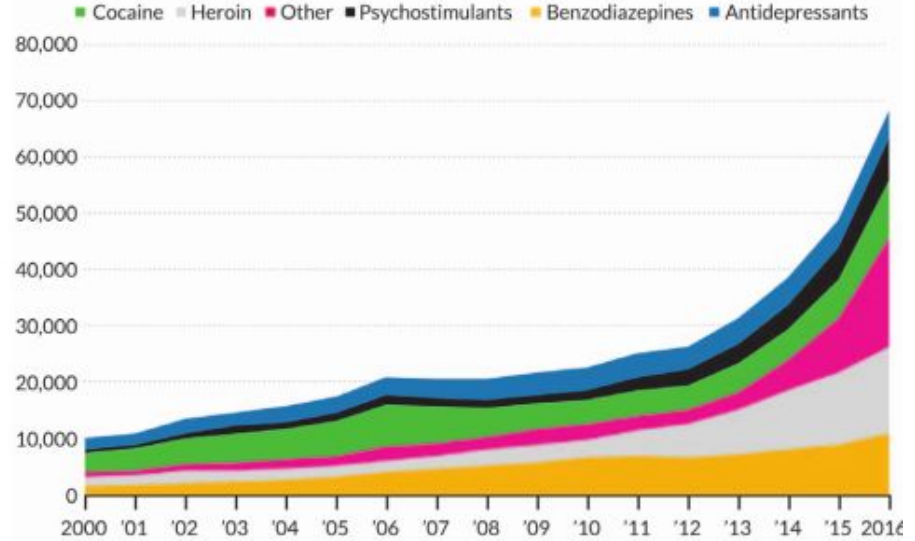
A slope graph can include many gray lines as a background

More than 60,000 people died from drug overdoses in 2016



Source: National Institute on Drug Abuse

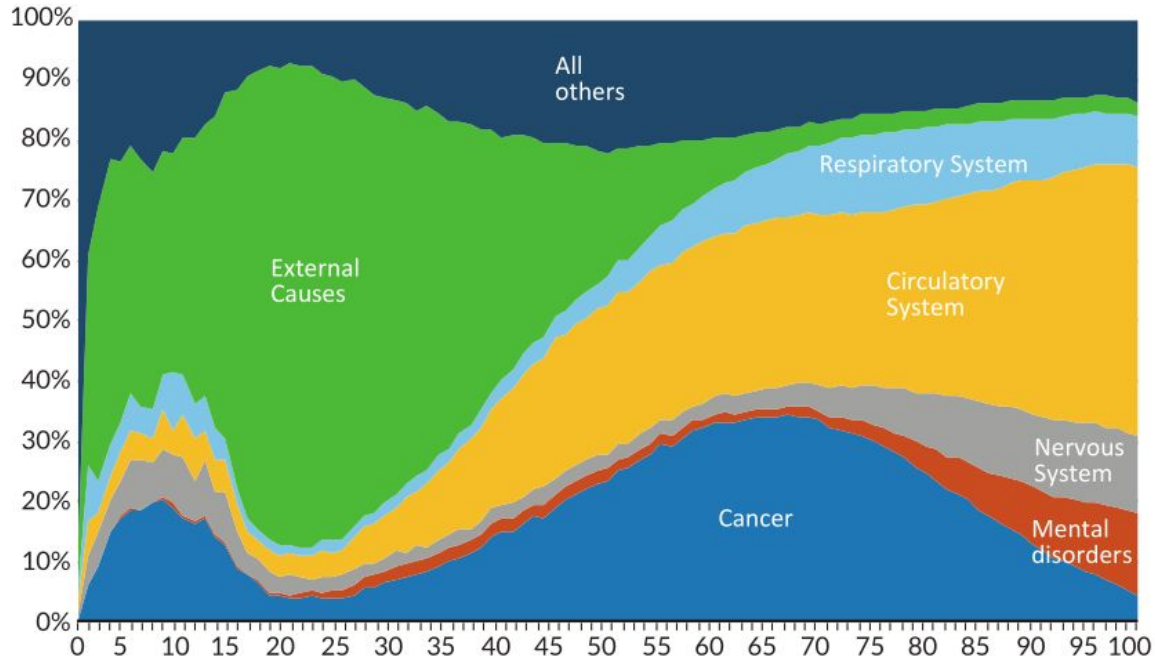
More than 60,000 people died from drug overdoses in 2016



Source: National Institute on Drug Abuse

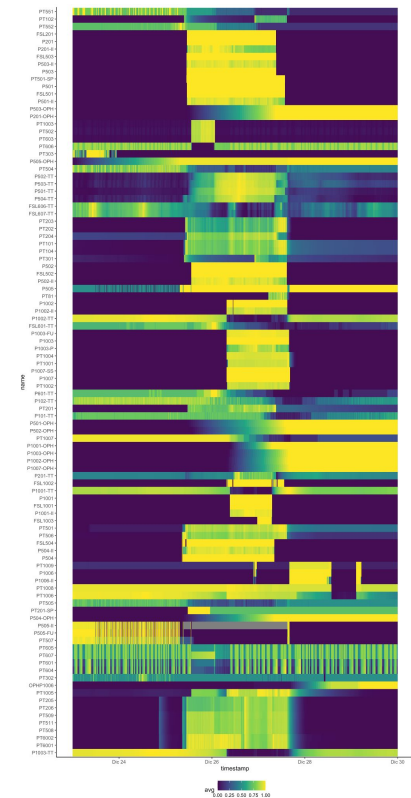
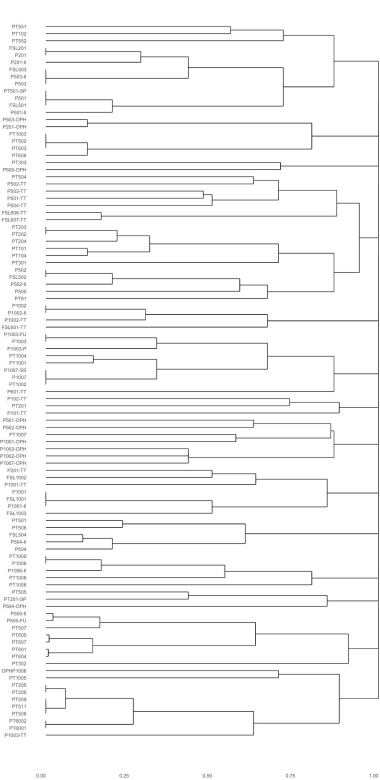
Area charts allow to see the evolution of the sum of several categories.
Beware: the bottom area is the most readable

Causes of death by age in the United States in 2017



Source: Centers for Disease Control and Prevention

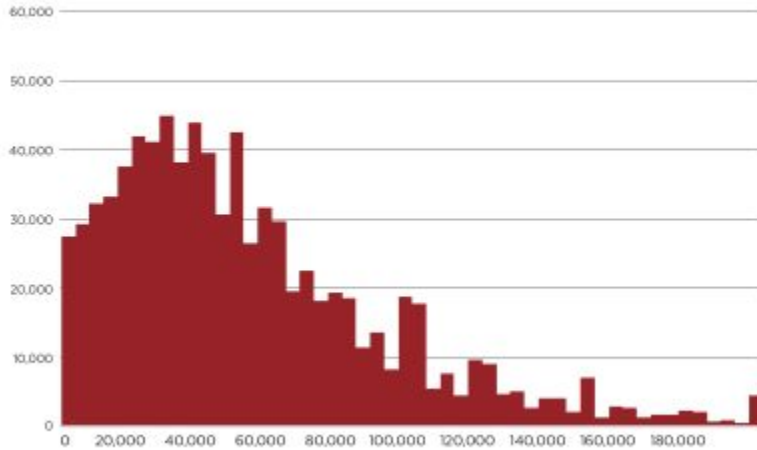
When normalized, stacked area charts allow to see the changes in distribution



Using the order of the leaves of a hierarchical clustering of data allows to effectively group time series by similarity.

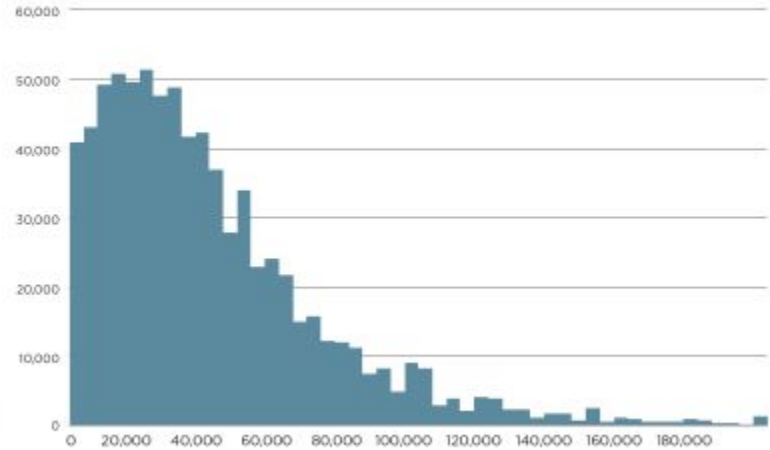
Distributions

MEN'S EARNINGS DISTRIBUTION IN 2016



Source: U.S. Census Bureau

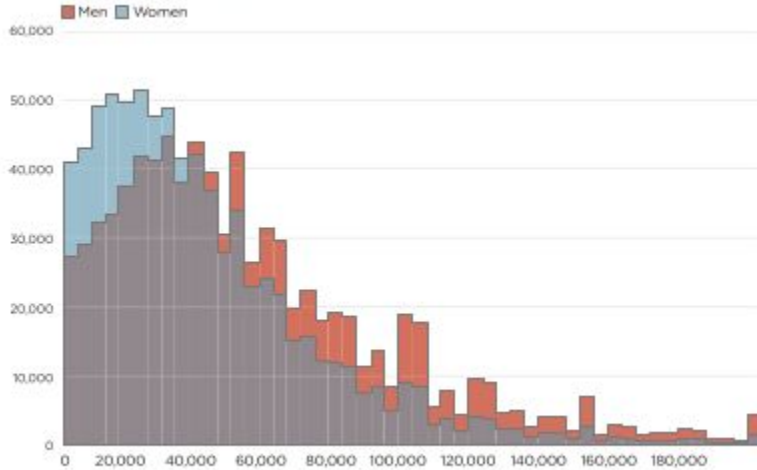
WOMEN'S EARNINGS DISTRIBUTION IN 2016



Source: U.S. Census Bureau

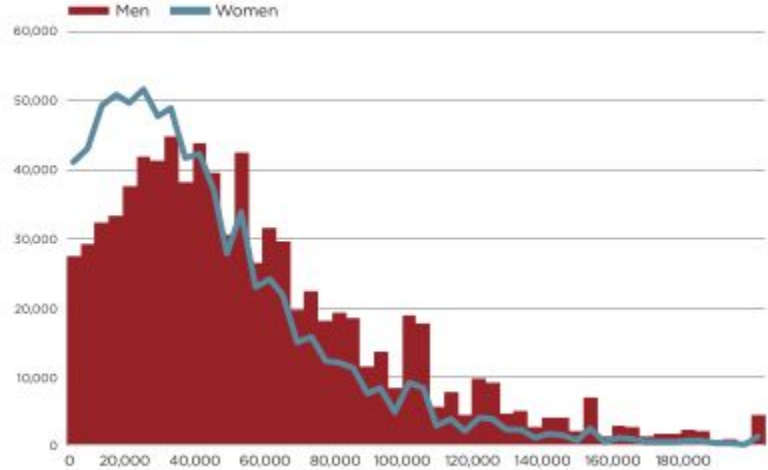
Histograms are one of the basic tools to show distributions

MEN'S AND WOMEN'S EARNINGS DISTRIBUTIONS IN 2016



Source: U.S. Census Bureau

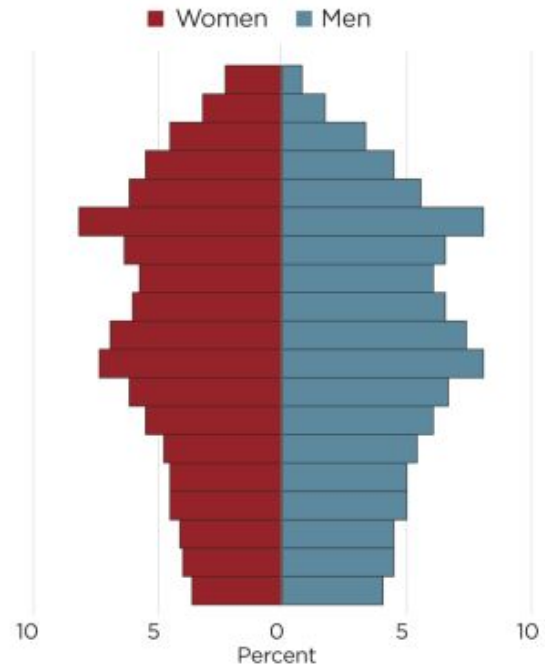
MEN'S AND WOMEN'S EARNINGS DISTRIBUTIONS IN 2016



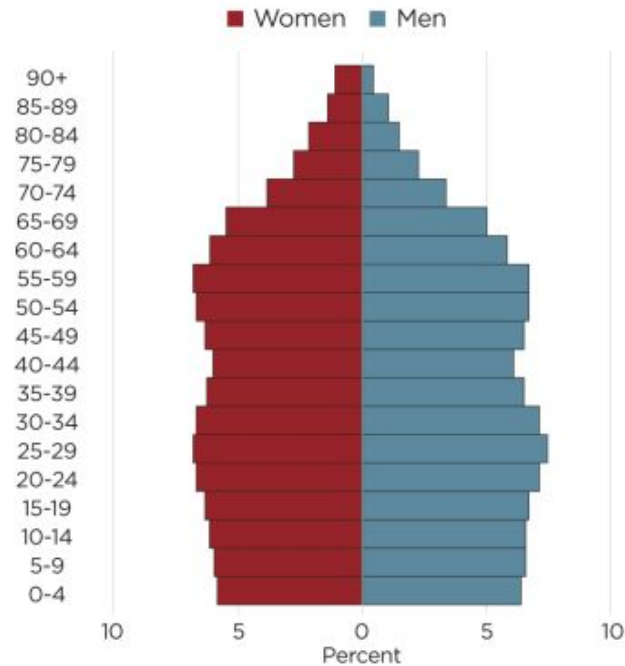
Source: U.S. Census Bureau

Overlay (with transparency) or use different aesthetics to save space

AGE DISTRIBUTION OF MEN AND WOMEN JAPAN, 2016

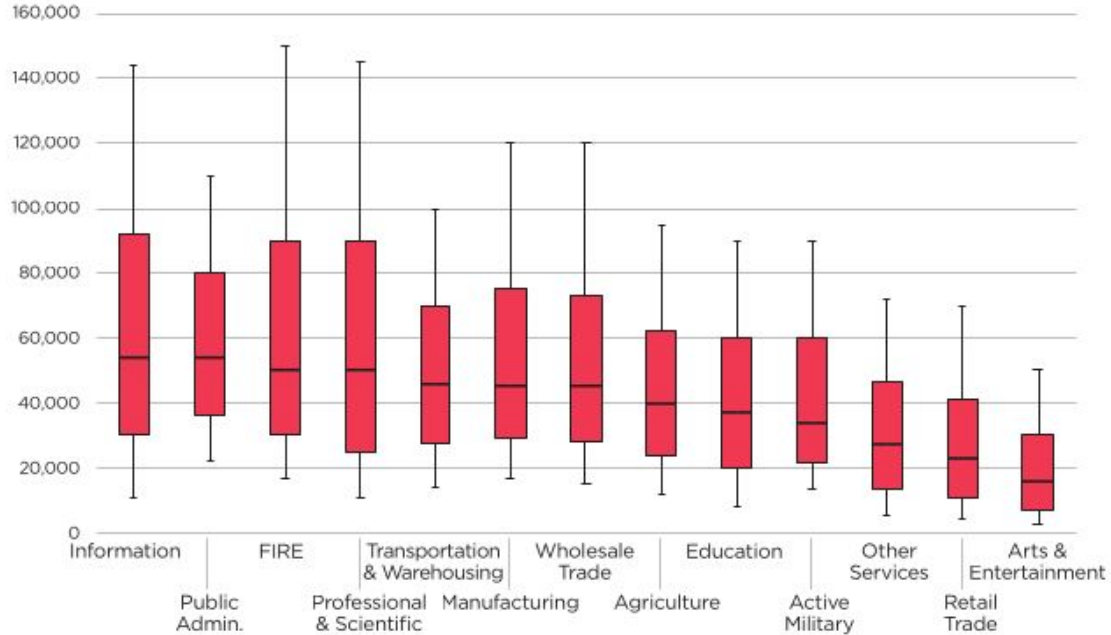


AGE DISTRIBUTION OF MEN AND WOMEN UNITED STATES, 2016



Pyramid charts are often used to compare the distribution of two categories

EARNINGS DISTRIBUTION IN U.S. INDUSTRIES

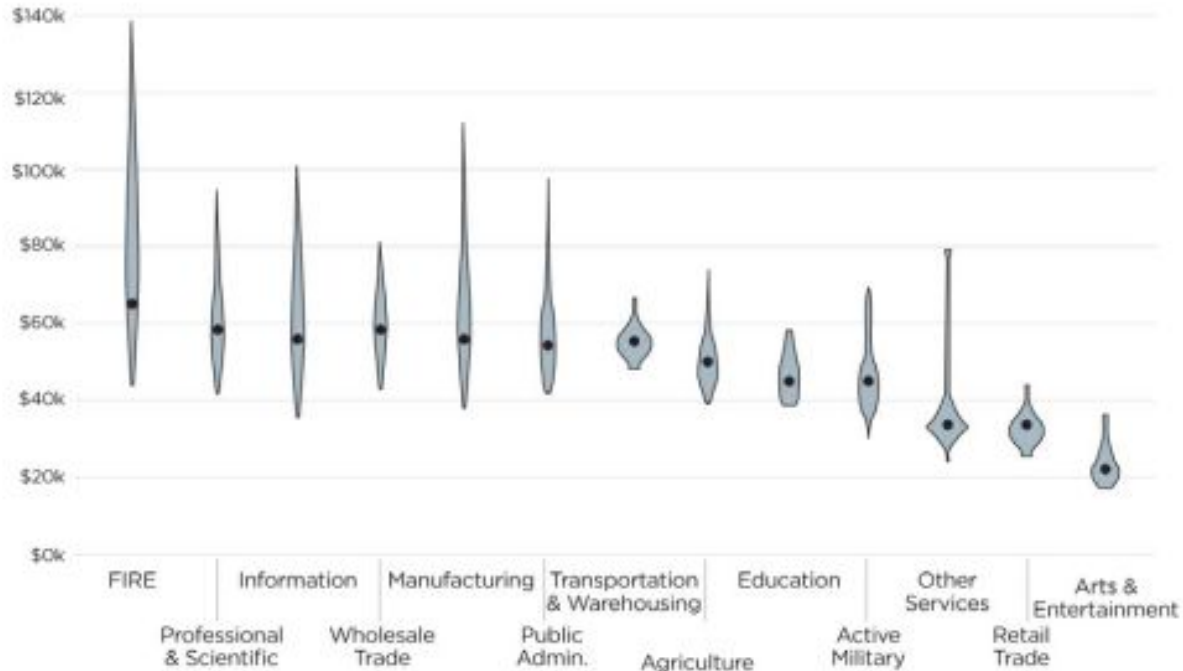


Source: U.S. Census Bureau

Note: FIRE = Finance, Insurance, and Real Estate

Box plots allow to compare the (rough) distribution of different populations

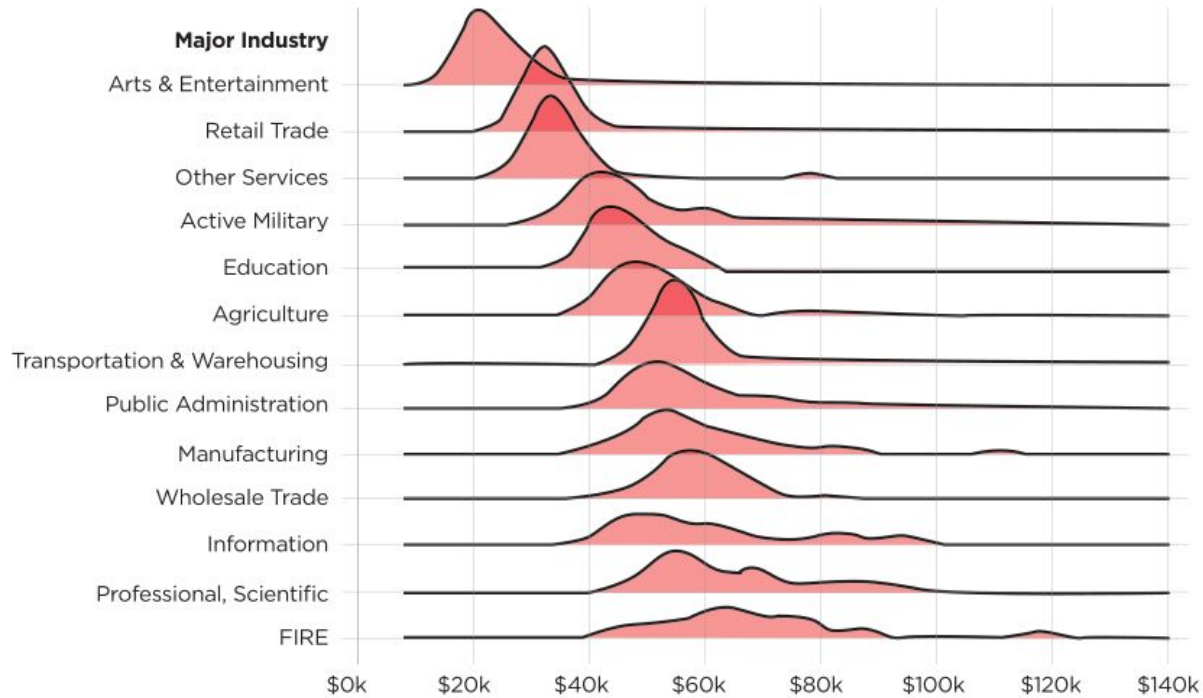
EARNINGS DISTRIBUTION IN U.S. INDUSTRIES



Source: U.S. Census Bureau
Note: FIRE = Finance, Insurance, and Real Estate

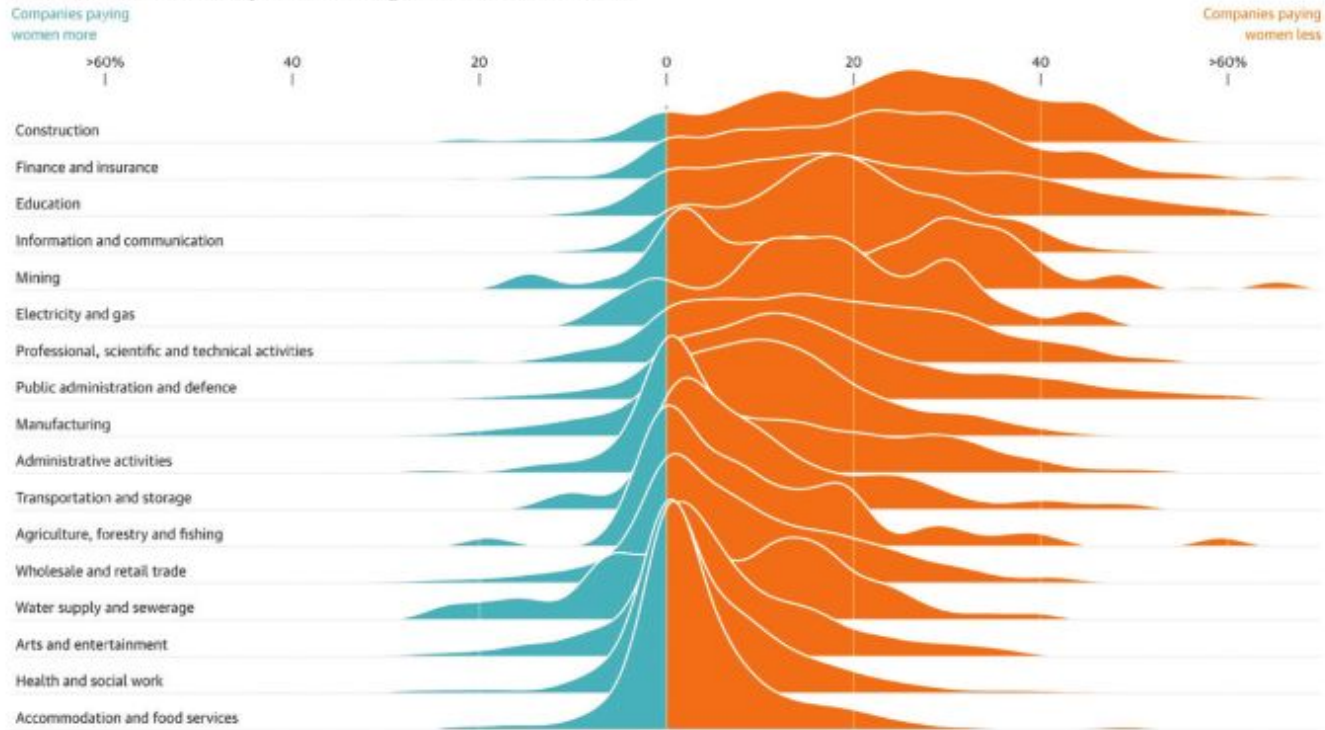
Violin plots show an estimate of the entire distribution of the data

EARNINGS DISTRIBUTION IN U.S. INDUSTRIES



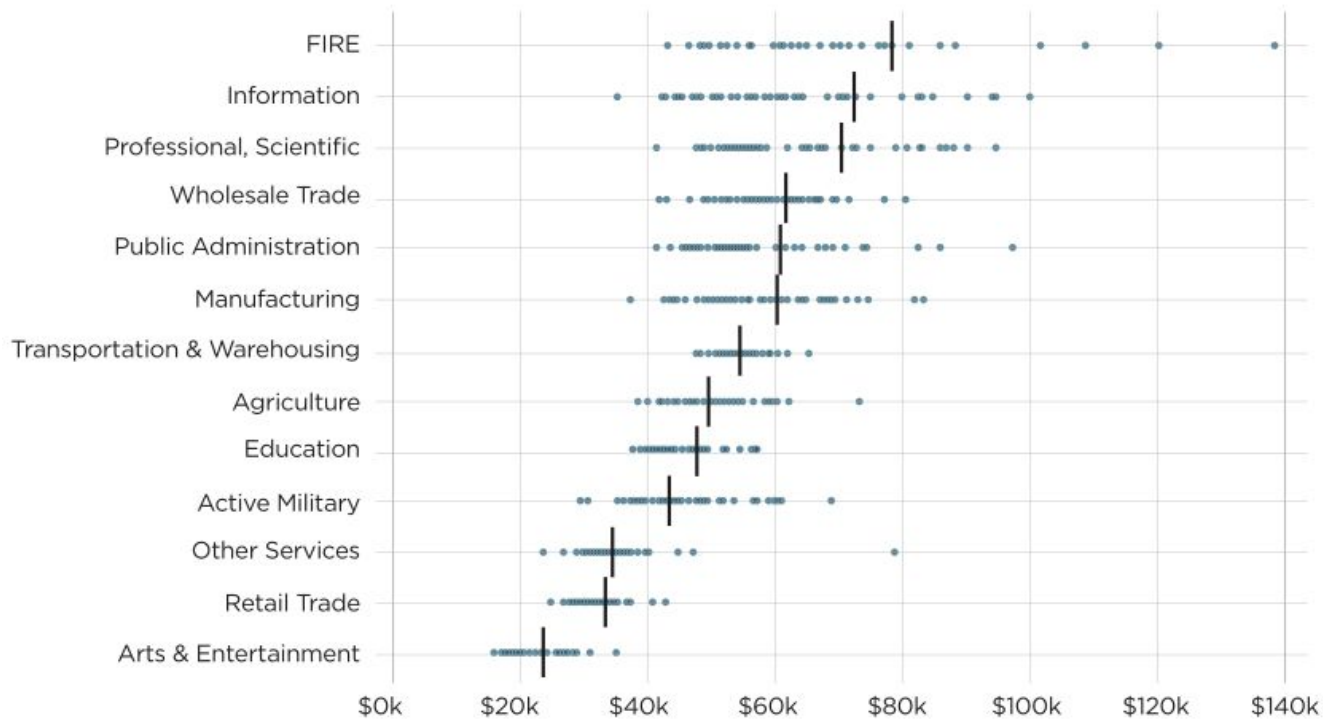
Ridgeline plots achieve the same, by using a more compact display

Women are more likely to be underpaid in certain sectors



Clever use of color and annotations allows to convey more information

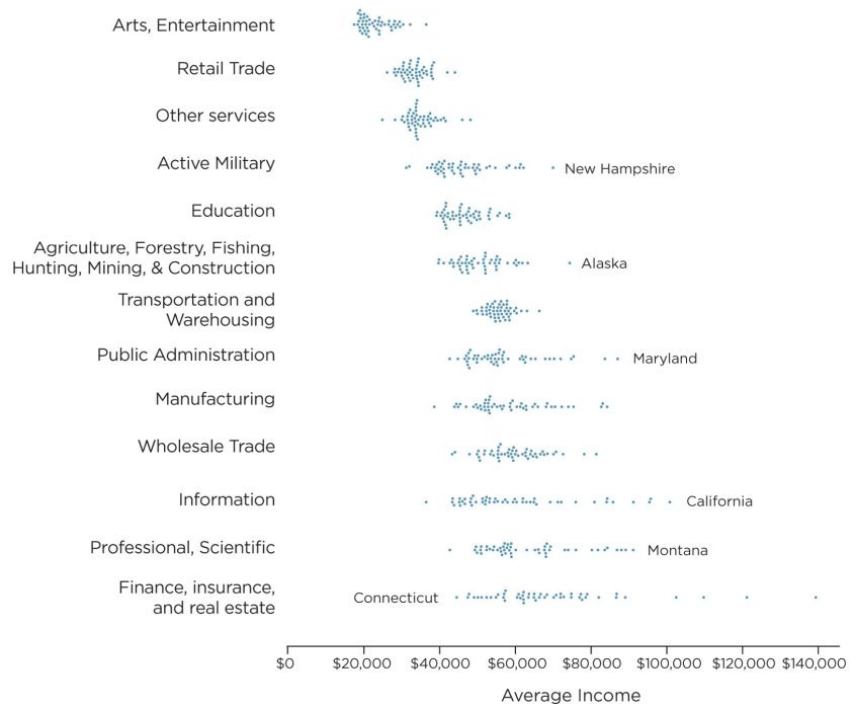
• Each dot represents a different state



You can show the distribution of data by showing the data itself

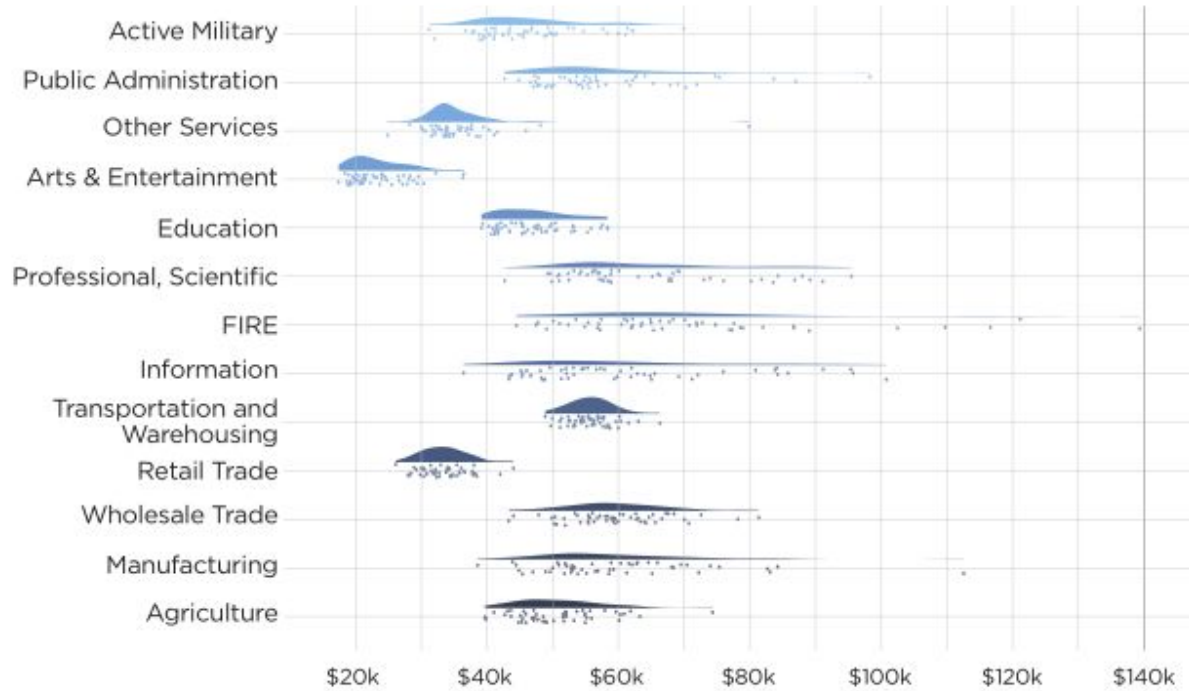
EARNINGS DISTRIBUTION IN U.S. INDUSTRIES

(Major industries by state)



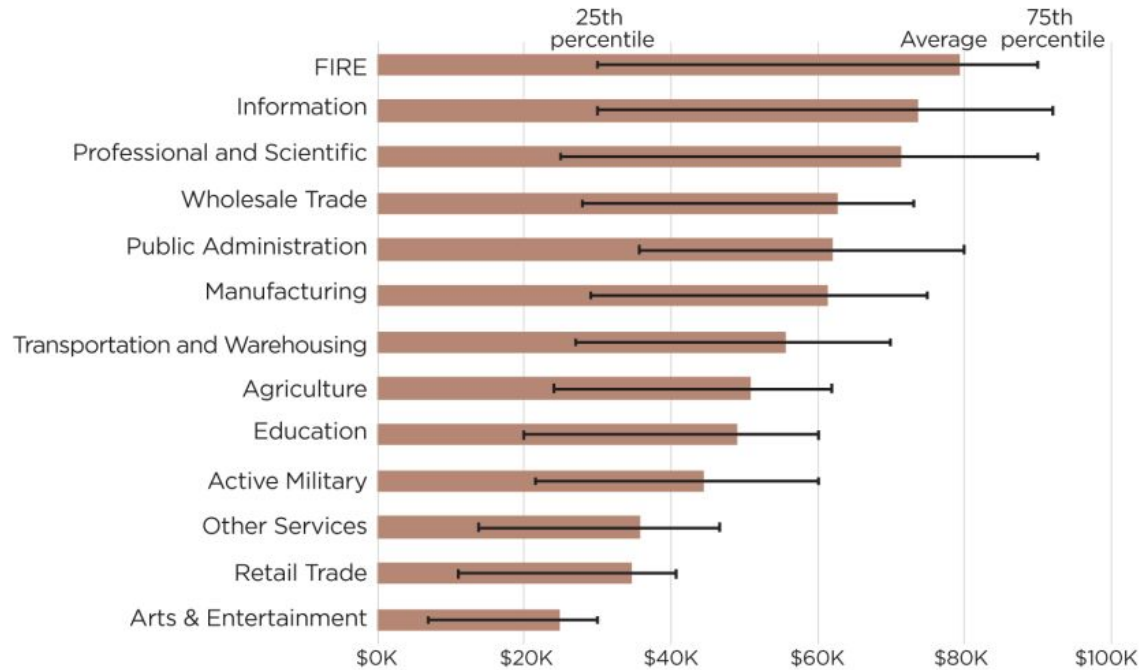
A beeswarm plot displaces points to give a cleaner visualization

EARNINGS DISTRIBUTION IN U.S. INDUSTRIES



A raincloud plot combines all of the above

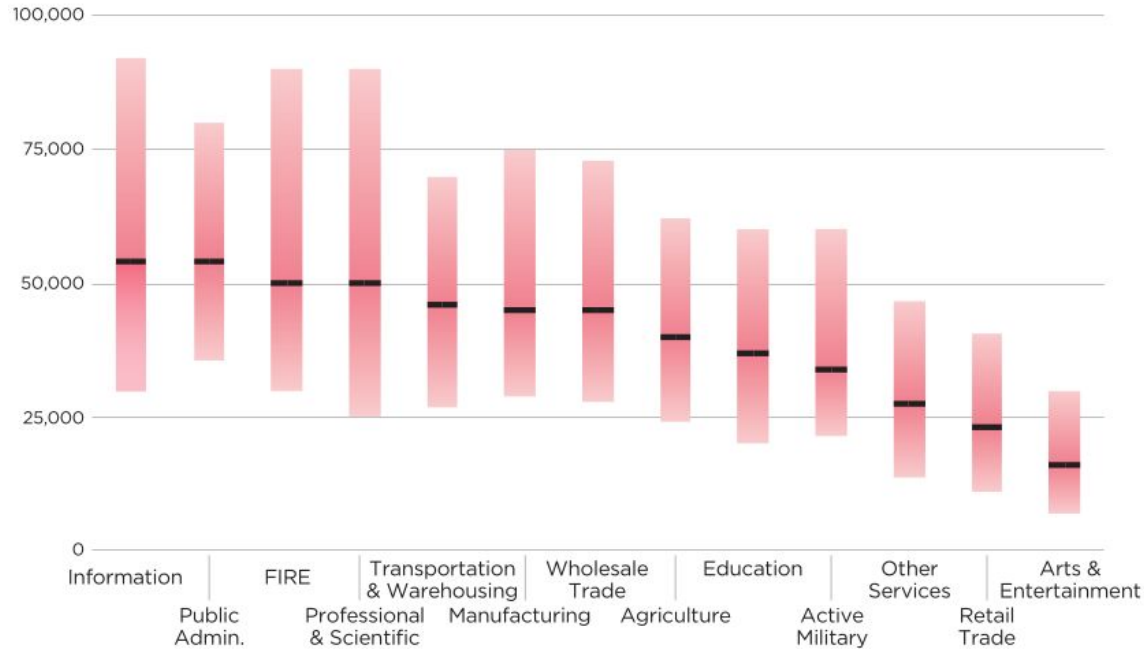
AVERAGE EARNINGS IN U.S. INDUSTRIES IN 2016



Source: U.S. Census Bureau

Error bars are a classic: just make sure to state what they mark

MEDIAN INCOME FOR DIFFERENT U.S. INDUSTRIES IN 2016

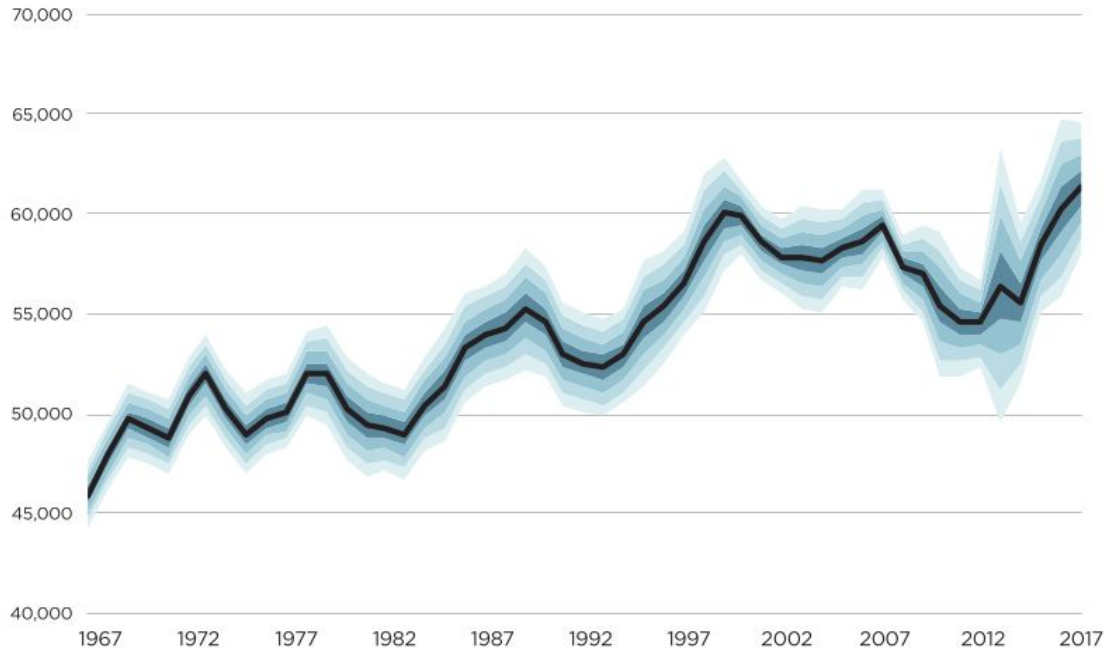


Source: U.S. Census Bureau

Note: FIRE = Finance, Insurance, and Real Estate

Gradient charts provide a more nuanced representation of the error distribution

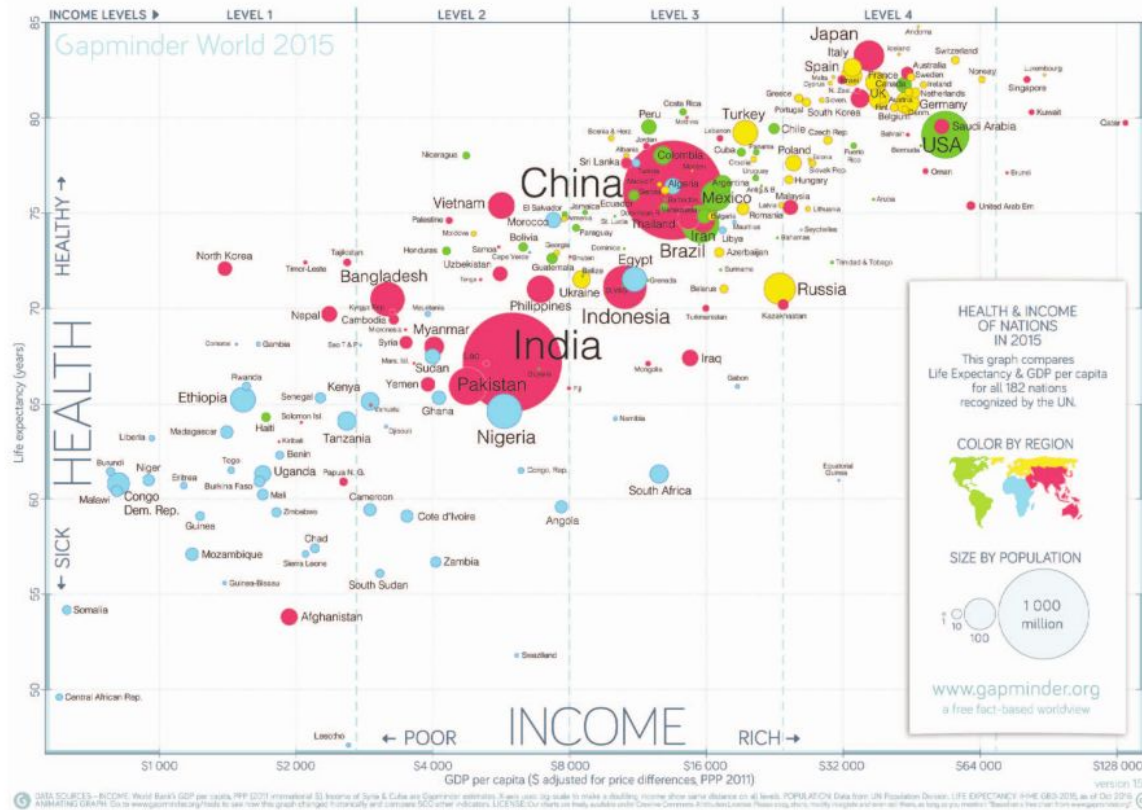
MEDIAN INCOME IN THE UNITED STATES FROM 1967 TO 2017



Source: U.S. Census Bureau

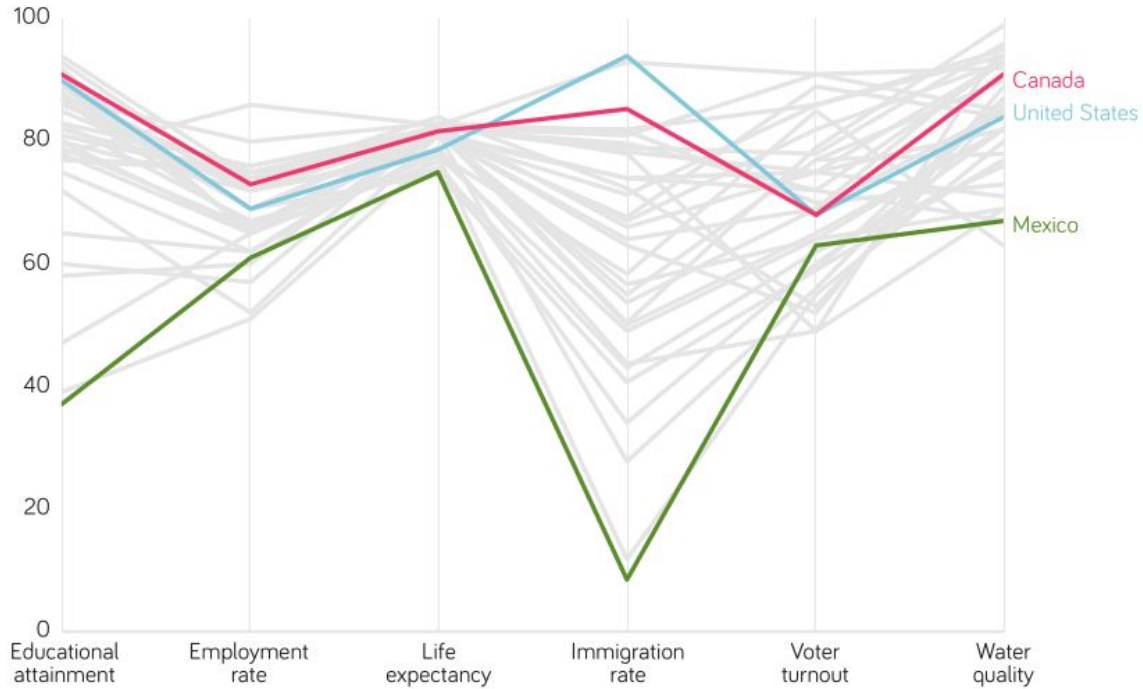
A more readable alternative shows discretized steps
(in this case: standard error)

Relationship



A scatterplot (or bubble plot) visualizes the relationship between two variables

Economic well-being in the OECD

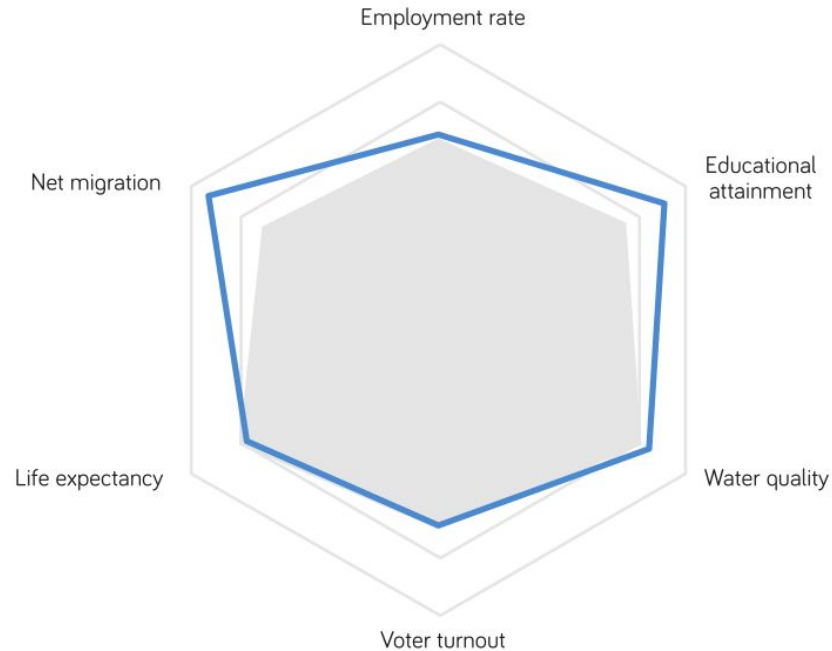


Source: Organisation for Economic Co-Operation and Development

Parallel coordinates compare multiple attributes

Economic well-being in the United States

(Gray area denotes average among the OECD)

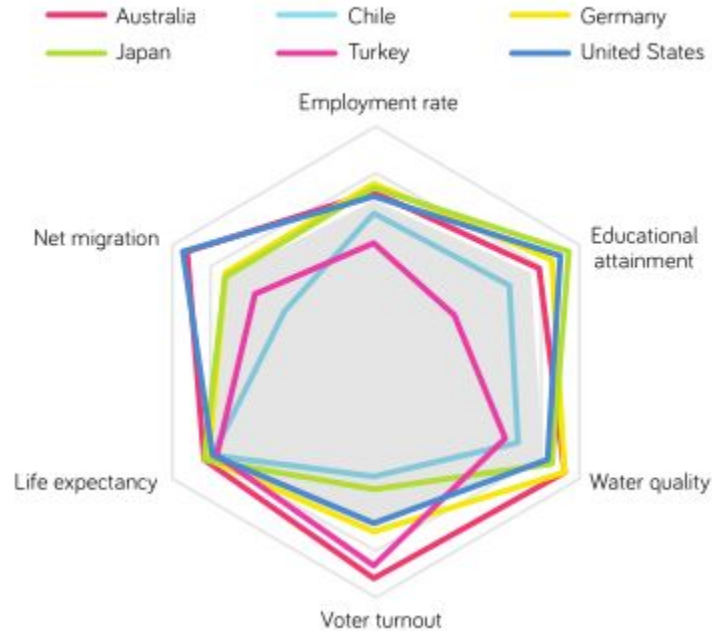


Source: Organisation for Economic Co-Operation and Development

Radar charts are parallel coordinates plots in polar coordinates

Economic well-being in the OECD

(Gray area denotes overall average)

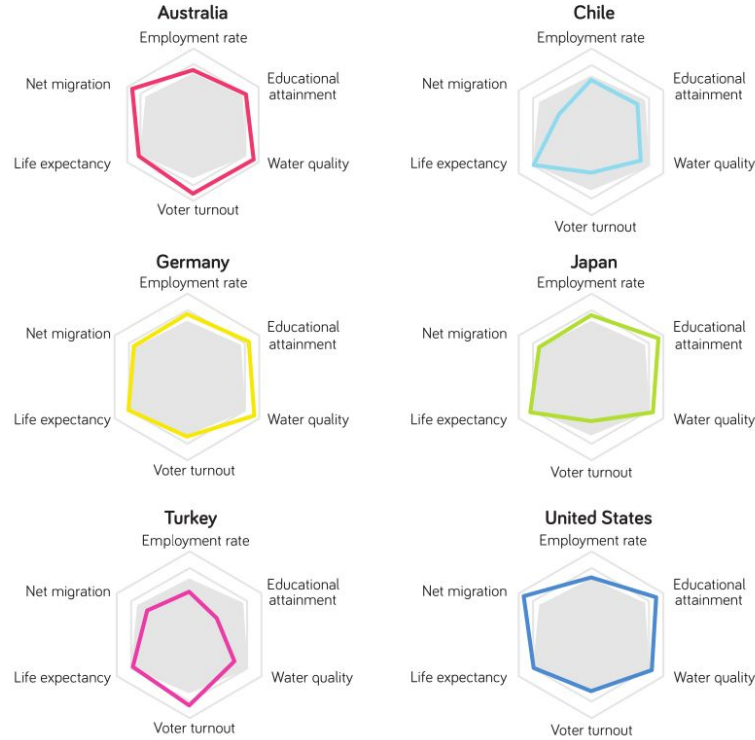


Source: Organisation for Economic Co-Operation and Development

The comparison of multiple lines can be confusing

Economic well-being in the OECD

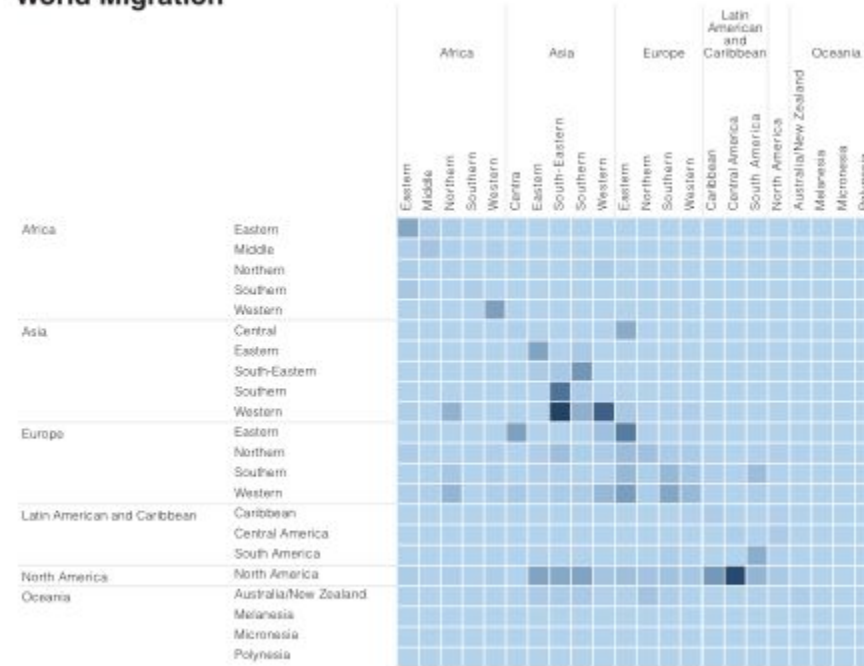
(Gray area denotes overall average)



Source: Organisation for Economic Co-Operation and Development

Faceting by multiple variables, with a common reference, helps comparing multiple variables

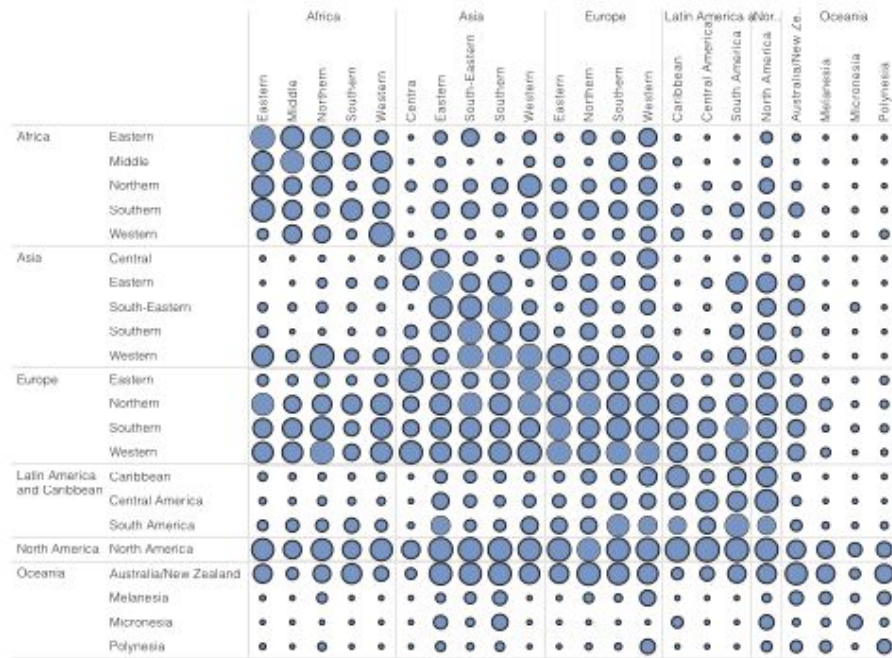
World Migration



Source: Organisation for Economic Co-Operation and Development.
Note: Data limited to a minimum of 200,000 immigrants or emigrants

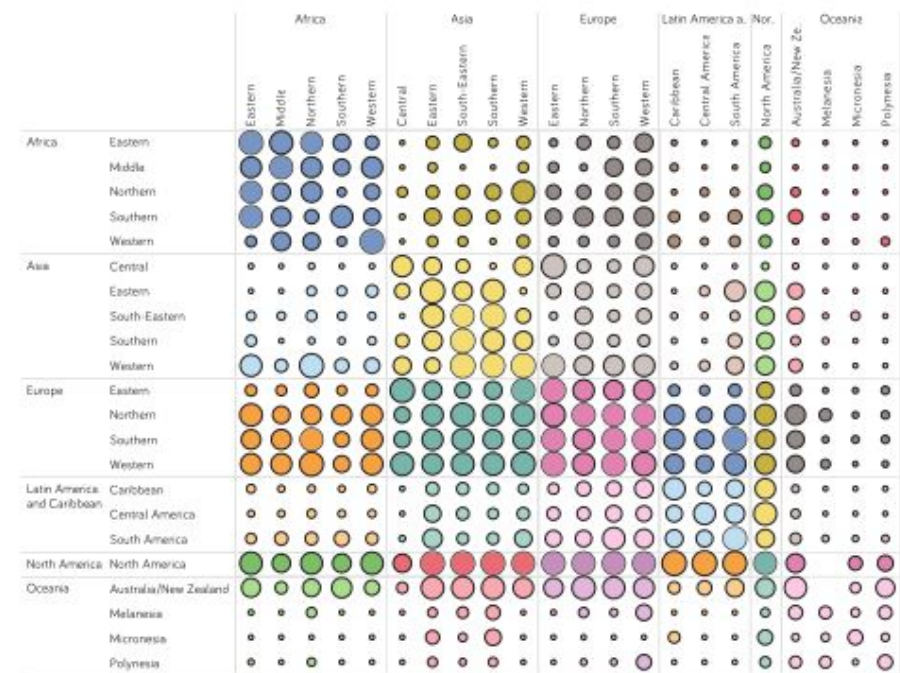
A correlation matrix shows the pairwise relationship (not necessarily the correlation) of multiple categories, using color to encode the values

World Migration



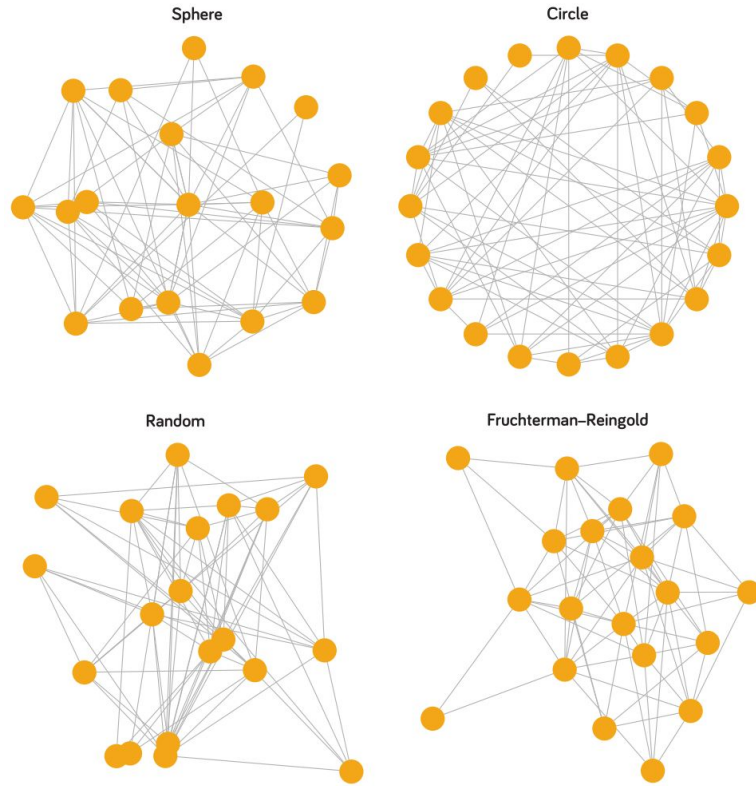
Source: Organisation for Economic Co-Operation and Development

World Migration

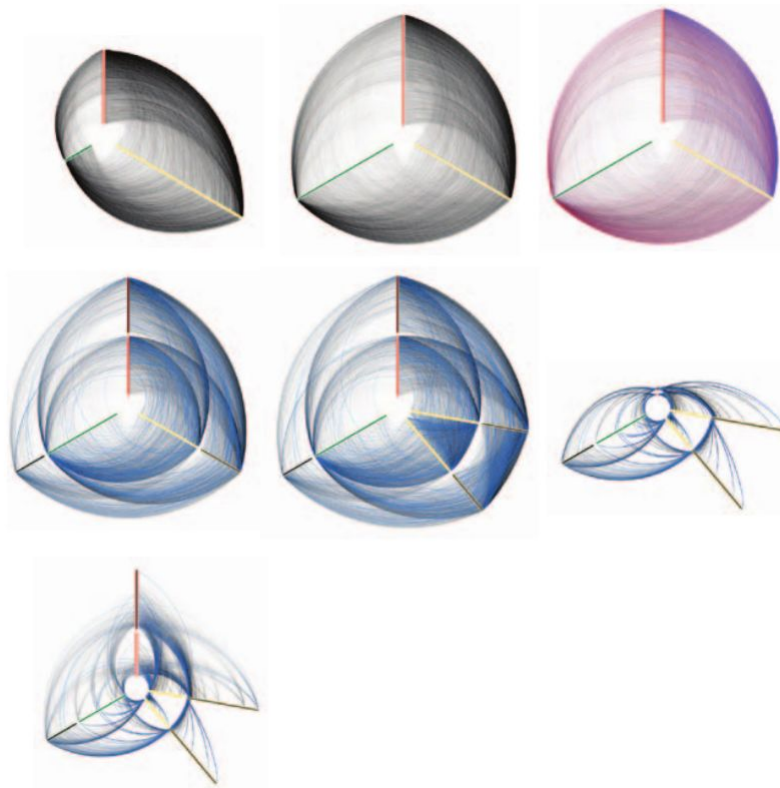


Source: Organisation for Economic Co-Operation and Development

Other symbols and aesthetics can be used,
thus freeing color for grouping purposes



For graph data consider the graphviz package and the networkx library

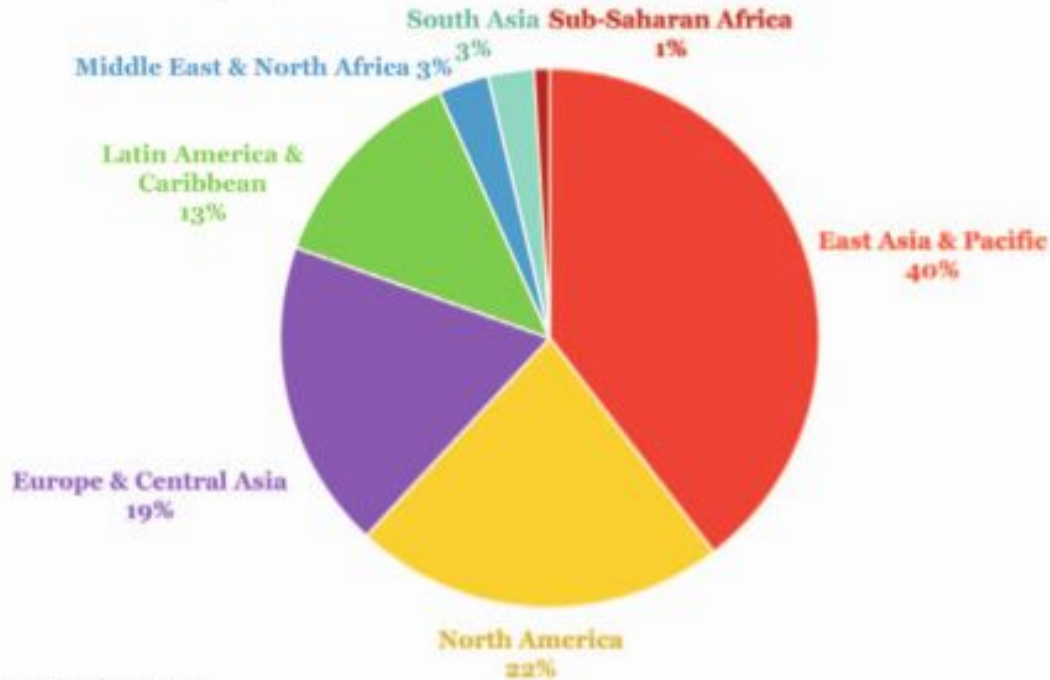


An interesting alternative for visualizing graphs are *hive charts*

Part-to-whole

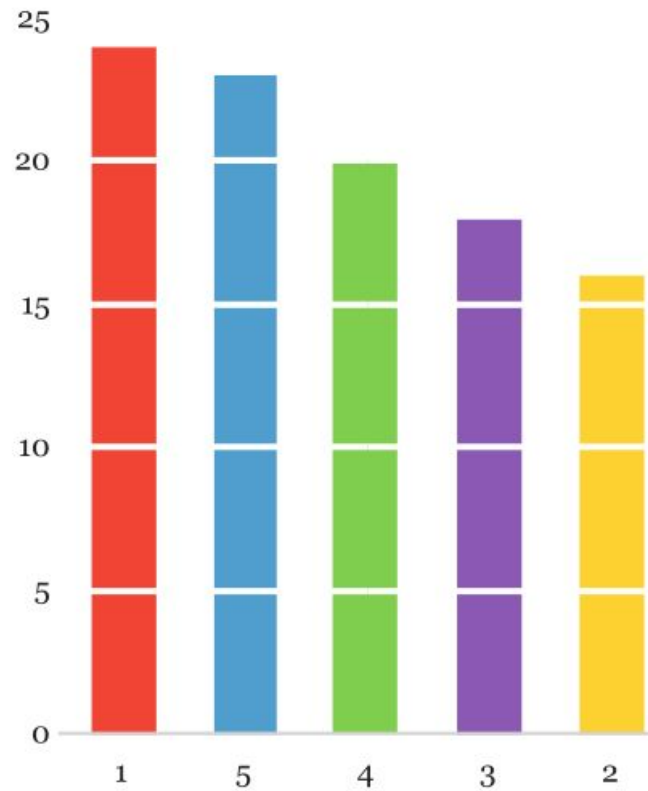
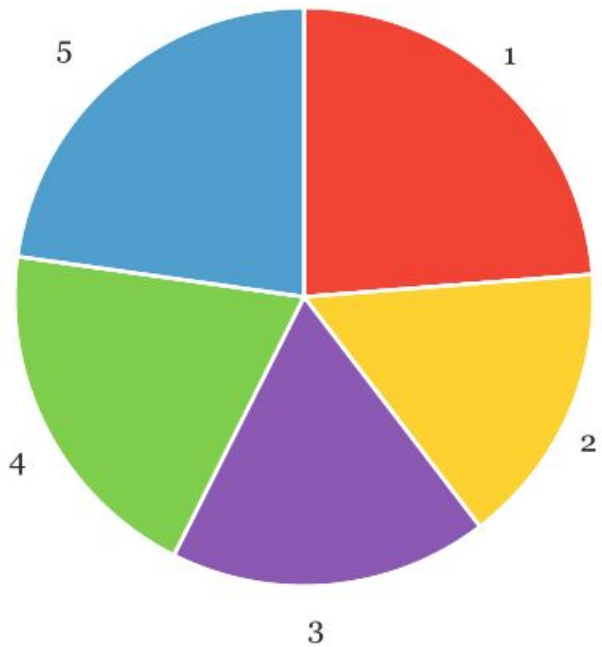
Distribution of imported goods to the United States in 2016

(Percent of total imports)



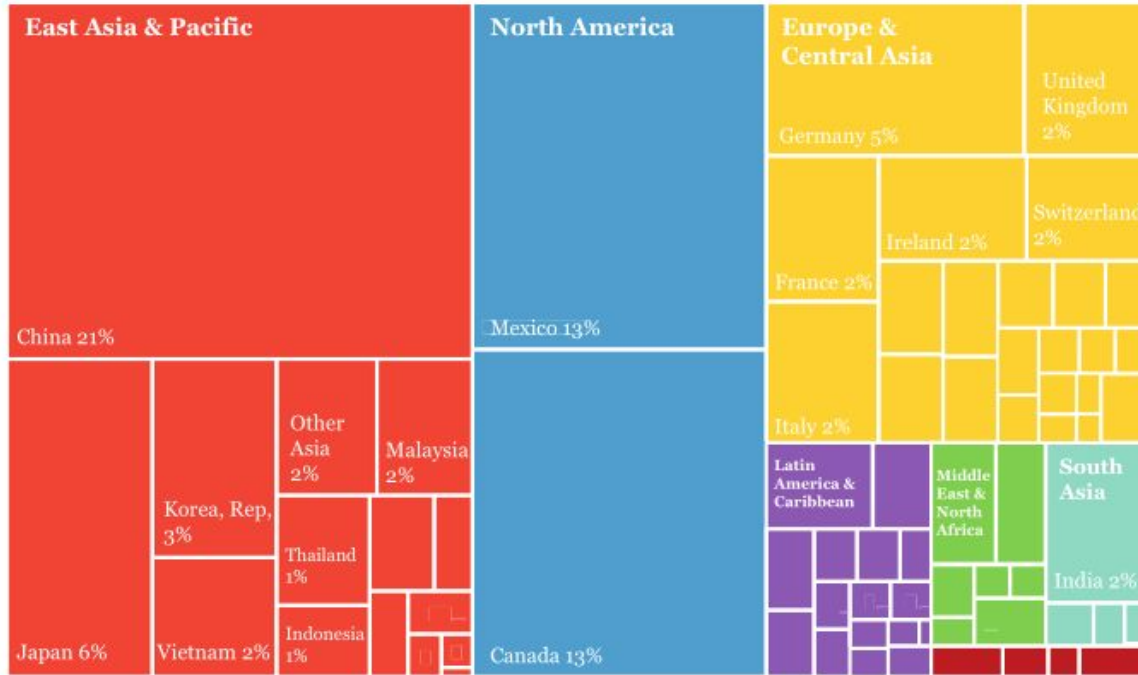
Source: The World Bank

Pie charts are the go-to tool for part-to-whole visualizations



A bar chart is often more effective

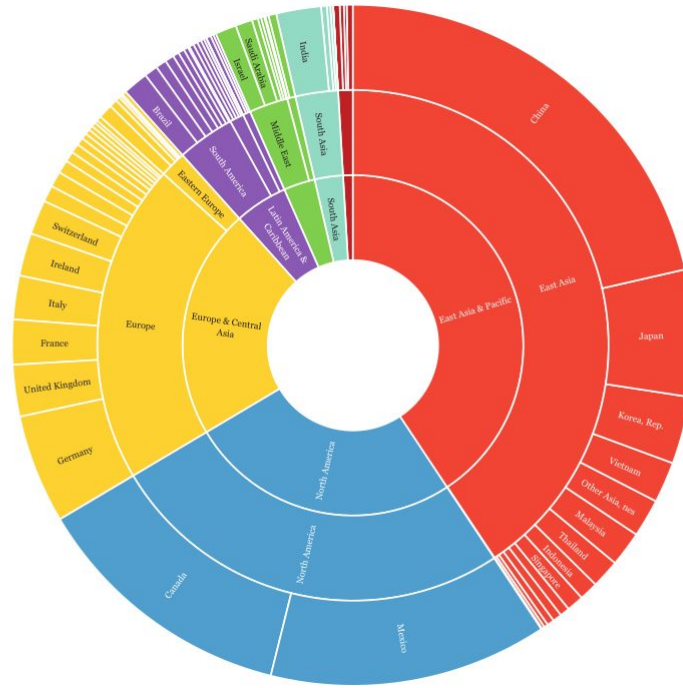
Distribution of imported goods to the United States in 2016



Source: The World Bank

Treemaps can show hierarchical data

Distribution of imported goods to the United States in 2016



Source: The World Bank

Sunburst diagrams can too, and are very similar to pie charts

Qualitative

Tables

Country	2013	2014	2015	2016	Country	2013	2014	2015	2016
China	7.23	6.76	6.36	6.12	China	7.23	6.76	6.36	6.12
India	5.10	6.14	6.90	5.89	India	5.10	6.14	6.90	5.89
United States	0.96	1.80	2.09	0.74	United States	0.96	1.80	2.09	0.74
Indonesia	4.24	3.73	3.65	3.85	Indonesia	4.24	3.73	3.65	3.85
Mexico	-0.06	1.45	1.90	1.68	Mexico	-0.06	1.45	1.90	1.68
Pakistan	2.21	2.51	2.61	3.44	Pakistan	2.21	2.51	2.61	3.44

Offset titles from body

Role	Name	ID	Start Date	Quarterly Profit	Percent Change
Operations	Waylon Dalton	A1873	May-11	5692.88	34.1
Operations	Justine Henderson	B56	Jan-10	4905.02	43.522
Operations	Abdullah Lang	J5867	Jun-14	4919.53	38
Operations	Marcu Cruz	B395	Dec-13	9877.52	37.1
Research	Thalia Cobb	C346	Apr-13	3179.49	-9
Research	Mathias Little	D401	Mar-11	5080.26	3.2
Research	Eddie Randolph	A576	Jul-18	7218.24	43.1
Contracts	Angela Walker	B31	Feb-18	6207.53	-1.788
Contracts	Lia Shelton	C840	Jan-16	1070.61	4.31
Contracts	Hadassah Hartman	D411	Nov-15	3735.96	3.01

Role	Name	ID	Start Date	Quarterly Profit	Percent Change
Operations	Waylon Dalton	A1873	May-11	\$5,693	34.1
	Justine Henderson	B56	Jan-10	4,905	43.5
	Abdullah Lang	J5867	Jun-14	4,920	38.0
	Marcu Cruz	B395	Dec-13	9,878	37.1
Research	Thalia Cobb	C346	Apr-13	3,179	-9.0
	Mathias Little	D401	Mar-11	5,080	3.2
	Eddie Randolph	A576	Jul-18	7,218	43.1
Contracts	Angela Walker	B31	Feb-18	6,208	-1.8
	Lia Shelton	C840	Jan-16	1,071	4.3
	Hadassah Hartman	D411	Nov-15	3,736	3.0

Use subtle subdividers instead of heavy grid lines

	2016	2016	2016
China	6,894.40	6,894.40	6,894.40
India	1,862.43	1,862.43	1,862.43
United States	52,319.10	52,319.10	52,319.10
Indonesia	3,974.73	3,974.73	3,974.73
Mexico	9,871.67	9,871.67	9,871.67
Pakistan	1,179.41	1,179.41	1,179.41
Average	12,683.62	12,683.62	12,683.62

Right align numbers and headers

Right-aligned and hard to read	Centered and even harder to read	Left-aligned and easiest to read
British Virgin Islands	British Virgin Islands	British Virgin Islands
Cayman Islands	Cayman Islands	Cayman Islands
Democratic Republic of Korea	Democratic Republic of Korea	Democratic Republic of Korea
Luxembourg	Luxembourg	Luxembourg
United States	United States	United States
Germany	Germany	Germany
New Zealand	New Zealand	New Zealand
Costa Rica	Costa Rica	Costa Rica
Peru	Peru	Peru

Left align text and header

Country	2014	2015	2016
China	6.76%	6.36%	6.12%
India	6.14%	6.90%	5.89%
United States	1.80%	2.09%	0.74%
Indonesia	3.73%	3.65%	3.85%
Mexico	-0.38%	-4.37%	-4.25%
Pakistan	2.51%	2.61%	3.44%
Average	3.43%	2.87%	2.63%

Country	2014	2015	2016
China	6.76%	6.36%	6.12%
India	6.14	6.90	5.89
United States	1.80	2.09	0.74
Indonesia	3.73	3.65	3.85
Mexico	-0.38	-4.37	-4.25
Pakistan	2.51	2.61	3.44
Average	3.43	2.87	2.63

Remove unit repetition

	2010	2011	2012	2013	2014	2015	2016
China	10.10	9.01	7.33	7.23	6.76	6.36	6.12
India	8.76	5.25	4.13	5.10	6.14	6.90	5.89
United States	1.68	0.85	1.46	0.96	1.80	2.09	0.74
Indonesia	4.83	4.79	4.68	4.24	3.73	3.65	3.85
Brazil	6.50	3.00	0.98	2.07	-0.38	-4.37	-4.25
Pakistan	-0.48	0.61	1.34	2.21	2.51	2.61	3.44
Nigeria	5.00	2.12	1.52	2.61	3.52	-0.02	-4.16
Bangladesh	4.40	5.25	5.28	4.77	4.84	5.37	5.96
Russia	4.46	5.20	3.48	1.57	-1.04	-3.04	-0.41
Mexico	3.49	2.12	2.15	-0.06	1.45	1.90	1.58

Highlight outliers

	2010	2011	2012	2013	2014	2015	2016
China	10.10	9.01	7.33	7.23	6.76	6.36	6.12
India	8.76	5.25	4.13	5.10	6.14	6.90	5.89
United States	1.68	0.85	1.46	0.96	1.80	2.09	0.74
Indonesia	4.83	4.79	4.68	4.24	3.73	3.65	3.85
Brazil	6.50	3.00	0.98	2.07	-0.38	-4.37	-4.25
Pakistan	-0.48	0.61	1.34	2.21	2.51	2.61	3.44
Nigeria	5.00	2.12	1.52	2.61	3.52	-0.02	-4.16
Bangladesh	4.40	5.25	5.28	4.77	4.84	5.37	5.96
Russia	4.46	5.20	3.48	1.57	-1.04	-3.04	-0.41
Mexico	3.49	2.12	2.15	-0.06	1.45	1.90	1.58

Highlight outliers

Region	Country	Per Capita GDP		Percent Change
		2015	2016	
Asia	China	6496.62	6894	6.1238
Asia	India	1758.84	1862	5.8898
North America	United States	51933.40	52319	0.7428
Asia	Indonesia	3827.55	3975	3.8453
North America	Brazil	11351.57	10869	-4.2541
Asia	Pakistan	1140.21	1179	3.4387
Africa	Nigeria	2562.52	2456	-4.1601
Asia	Bangladesh	971.64	1030	5.9627
North America	Mexico	9717.90	9872	1.5824
Asia	Japan	47163.49	47661	1.0546
Africa	Ethiopia	487.29	511	4.9041
Middle East	Egypt	2665.35	2726	2.2633
Europe	Germany	45412.56	45923	1.1240
Middle East	Iran	6007.00	6734	12.1010
Middle East	Turkey	13898.75	14117	1.5734
Europe	France	41642.31	41969	0.7845
Average		15440	15631	2.6860

Region	Country	Per Capita GDP		Percent Change
		2015	2016	
Africa	Ethiopia	487	511	4.90
	Nigeria	2,563	2,456	-4.16
Asia	Bangladesh	972	1,030	5.96
	China	6,497	6,894	6.12
	India	1,759	1,862	5.89
	Indonesia	3,838	3,975	3.85
	Japan	47,163	47,661	1.05
	Pakistan	1,140	1,179	3.44
Europe	France	41,642	41,969	0.78
	Germany	45,413	45,923	1.12
Middle East	Egypt	2,665	2,726	2.26
	Iran	6,007	6,734	12.10
	Turkey	13,899	14,117	1.57
North America	Mexico	9,718	9,872	1.58
	United States	51,933	52,319	0.74
South America	Brazil	11,352	10,869	-4.25
Average		15,440	15,631	2.69

Group similar data and increase white space

Country	2007	2016	2007-2016
China	13.64	6.12	
India	8.15	5.89	
United States	0.82	0.74	
Indonesia	4.91	3.85	
Mexico	0.70	1.58	
Pakistan	2.72	3.44	
Average	5.15	3.60	

Country	2016
China	6.12
India	5.89
United States	0.74
Indonesia	3.85
Mexico	1.58
Pakistan	3.44
Average	3.60

Add visualizations when appropriate




Country	2007	2016
China	13.64	6.12
India	8.15	5.89
United States	- 0.82	- 0.74
Indonesia	4.91	3.85
Mexico	- 0.7	- 1.58
Pakistan	2.72	3.44
Average	5.86	2.63

Add visualizations when appropriate


```

\begin{center}
  \begin{tabular}{lll}
    category & value & parameters \\
    \midrule
    A & \lollipop{200}{480} & blah blah \\
    B & \lollipop{300}{480} & blih bleh \\
    C & \lollipop{150}{480} & bluh bluh \\
    \bottomrule
  \end{tabular}
\end{center}

```

category	value	parameters
A	 200	blah blah
B	 300	blih bleh
C	 150	bluh bluh

The first argument of `\lollipop` is the value, the second is the maximum

```

\usepackage{tikz}
\usepackage{pgfplots}
\usepackage{booktabs}

\newcommand{\lollipop}[2]{
  \begin{tikzpicture}
    \begin{axis}[
      hide axis,
      height=1em,
      width=5em,
      xmax=#2,
      xmin=0,
      ymin=-0.01,
      ymax=1,
      % axis on top, major grid style={draw=black},xmajorgrids,
      scale only axis % So that we don't get size errors
    ]
      \addplot[black, thick] coordinates { (0,0.5) (#1,0.5) };
      \filldraw [red] (axis cs:#1,0.5) circle (2pt);
      \node[anchor=south west, inner xsep=0, inner ysep=.2em] at (axis cs:#1*1.1,0) {
        \small \pgfmathprintnumber[fixed,1000 sep={\,}]{#1}
      };
    \end{axis}
  \end{tikzpicture}
}

```

Here's the definition of \lollipop