data visualization

THESE THREE QUETSTION HELP TO UNDERSTAND DATA

1) WHO 2)WHAT 3)HOW

most used to tools to cleaning the and visualization data : excel , power bi, tableau, python ,Hadoop

WHO

you are data scientist or data analysts . first you need to know whose data is this .it was bank atm data are it was medical data etc like this .

but this help to better understand your brain ok this is the some data we need some columns and rows

WHAT

ACTION

everything under this action like most of the time we deal with action data analysts role

what was the sources of the data ? , where all data was present ?, what was the client requirements ? ,what was cleaning process of data ? and then it was a numeric or categorical data either both numerical and categorical

clean up the empty cells ,columns and convert unstructured data into useful data as you like it was Jason files ,csv,excel .....

Here are some action words to help act as thought starters as you determine what you are asking of your audience:

accept | agree | begin | believe | change | collaborate | commence | create | defend | desire | differentiate | do | empathize | empower | encourage | engage | establish | examine | facilitate | familiarize | form | implement | include | influence | invest | invigorate | know | learn | like | persuade | plan | promote | pursue | recommend | receive | remember | report | respond | secure | support | simplify | start | try | understand | validate

How

understanding the data

you are the data analysts

how is the data connected ? how process the data ? how data are influences other columns, rows   ?
then you have the idea what are column and rows required. as show the figure most of time we
visualization data
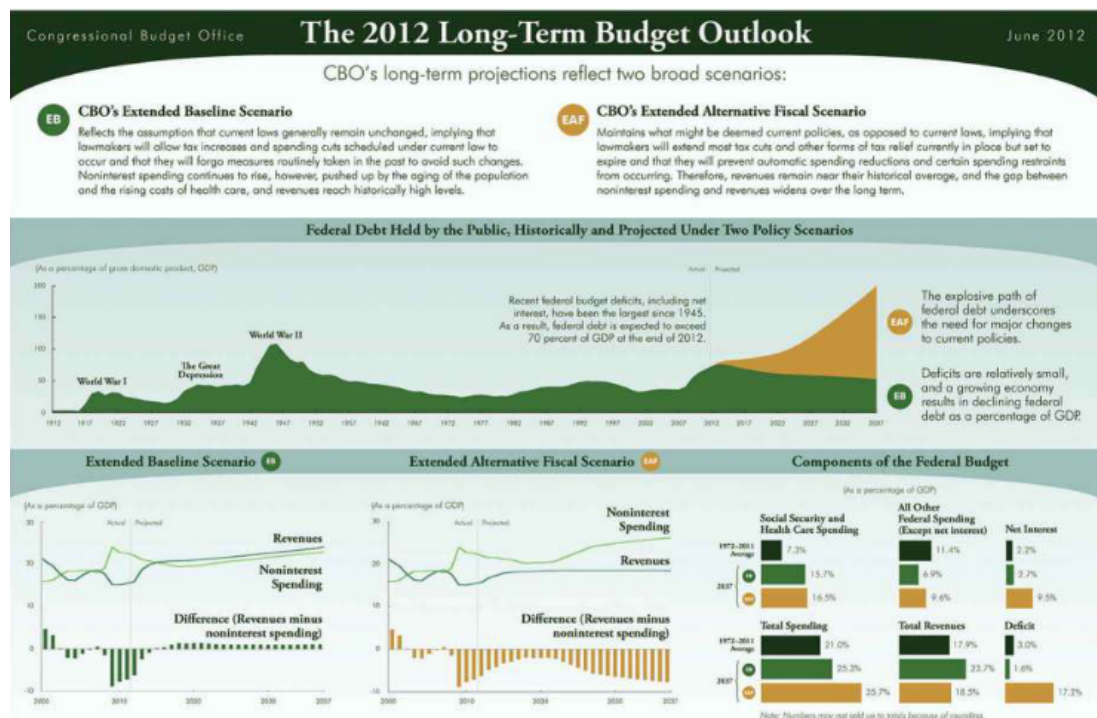
you are data scientist

how is the data was interacting with other data are any influences ? . first we need to understand what
are independent variables and what are the depended variables (you already have the client
requirements what data they want outcomes predict values) finding the correlation is there any
independent variables and depended variables   what was the mean, standard deviation , outliers ,
where all data lies in like first quartile(the 25th,57 percentile then we understand data was in between
25th to 57 percentile   ) it data was normal distribution are left, right skewed data using box plots ,
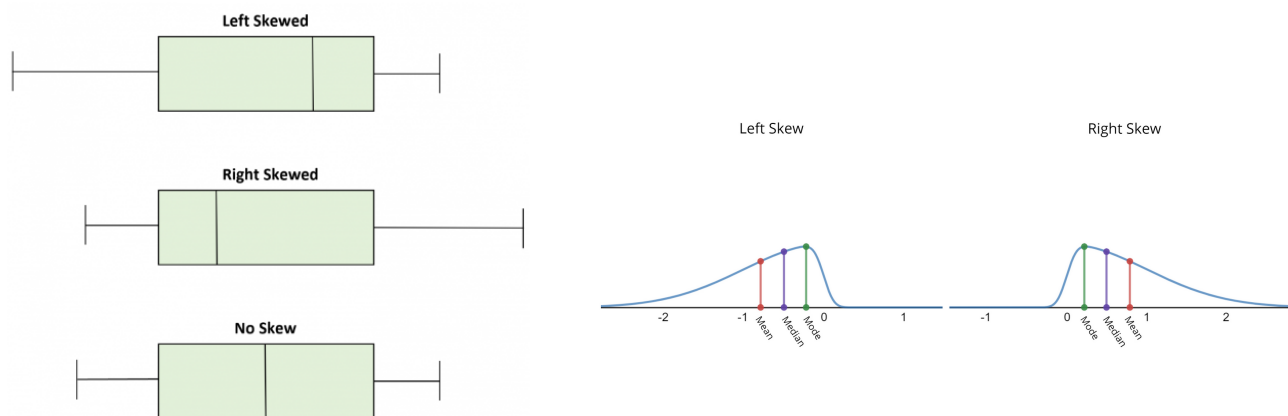
finally you have the idea which modal is better use in machine learning   : example : use linear
regression ,logistic regression , multi linear regression ....

choosing an effective visual for data analyst

these are frequently help in data visualization

91%

Simple text

Scatterplot

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Table

Line

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap
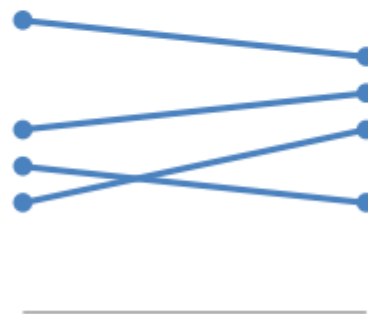
Slopegraph
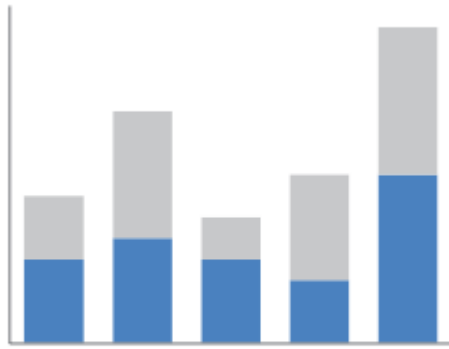
Vertical bar   Horizontal bar
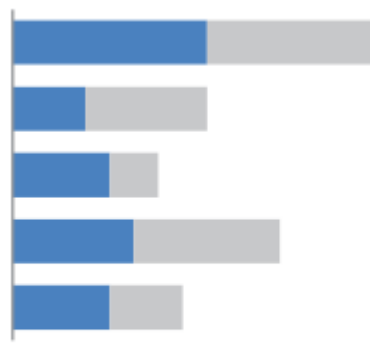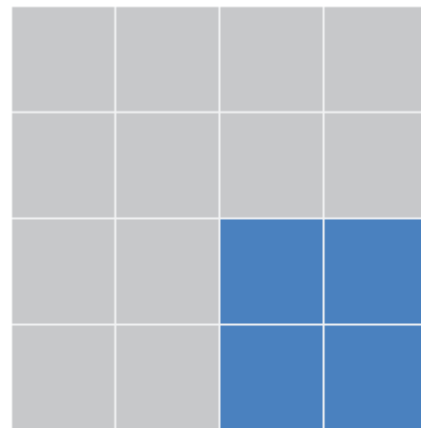
Stacked vertical bar   Stacked horizontal bar
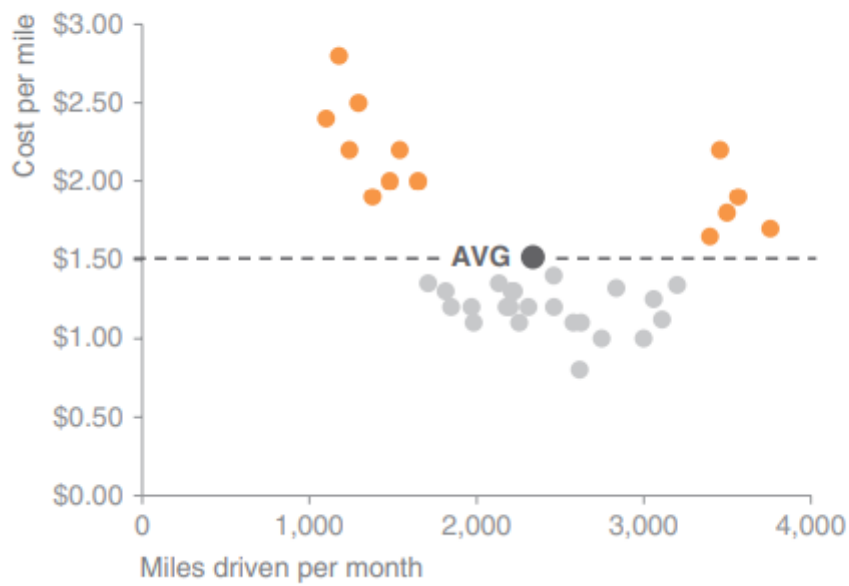
Waterfall   Square area

Scatterplot

Scatterplots can be useful for showing the relationship between two things, because they allow you to encode data simultaneously on a horizontal x-axis and vertical y-axis to see whether and what relation   ship exist

Cost per mile by miles driven

line graphs

when we pick up the line graph as show in the first figure all countries very confusing we can't understand

then we highlighted only few countries adding the names on the end of the line easy way to understand just in few sec

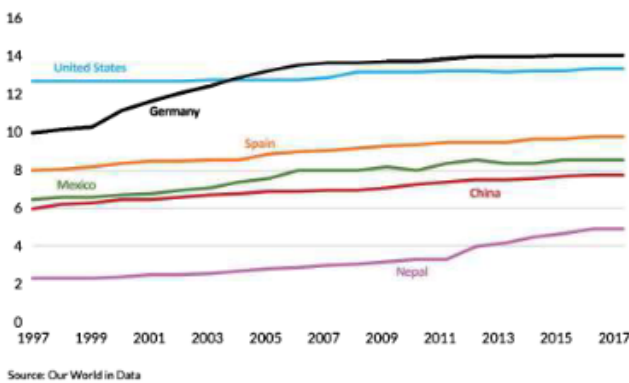* AT THAT WE USE VERTICAL BARS TO BETTER VIEW

Average years of schooling has increased around the world
(Number of years)
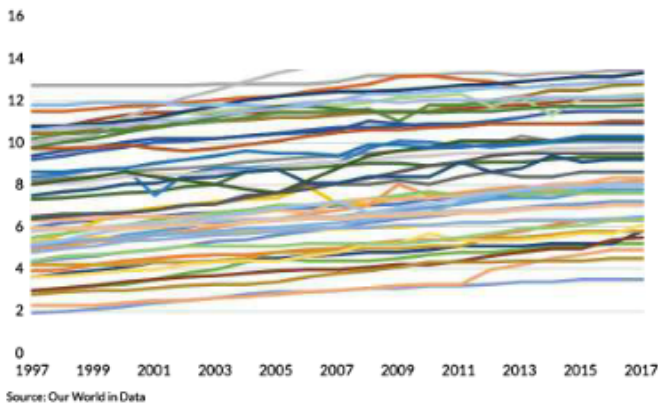
Source: Our World in Data

Average years of schooling has increased around the world
(Number of years)

Source: Our World in Data

Average years of schooling has increased around the world
(Number of years)

Source: Our World in Data
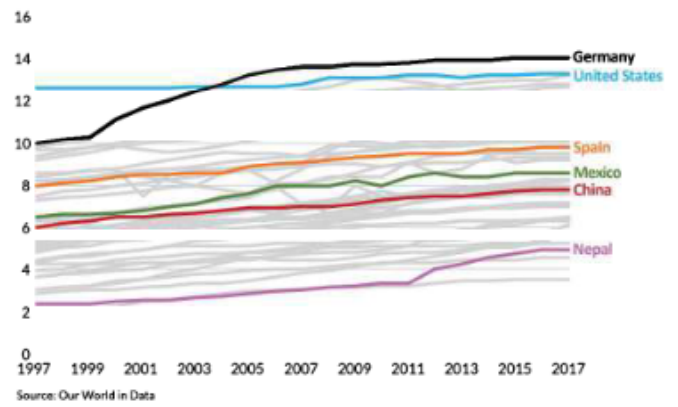
Average years of schooling has increased around the world
(Number of years)
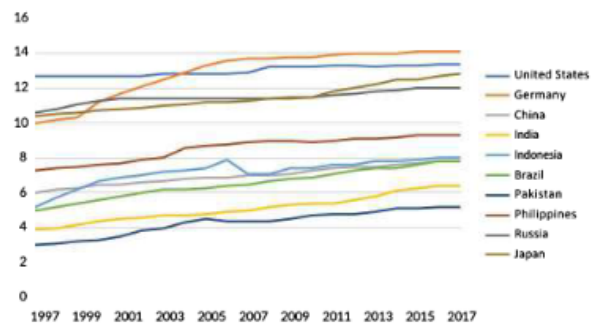
DI and SSI allowance rates have generally moved in tandem over the past 25 years
(Percent)
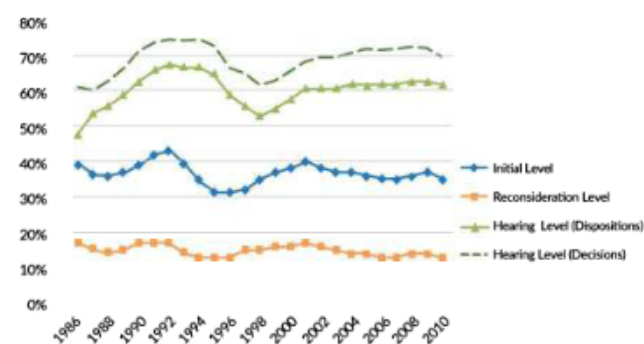
DI and SSI allowance rates have generally moved in tandem over the past 25 years
(Percent)

3. ADD EXPLAINES

THIS BETTER UNDERSTAND WHERE IN SINGLE GRAPH WHY GROWS HIGHS AND WHAT WAS DOWN FALLS WHAT NEXT DECSION TO INCREASE THE COMPNAY

Rise and Fall of the name Neil in the USA
Births 1912-2015
Source: data.gov
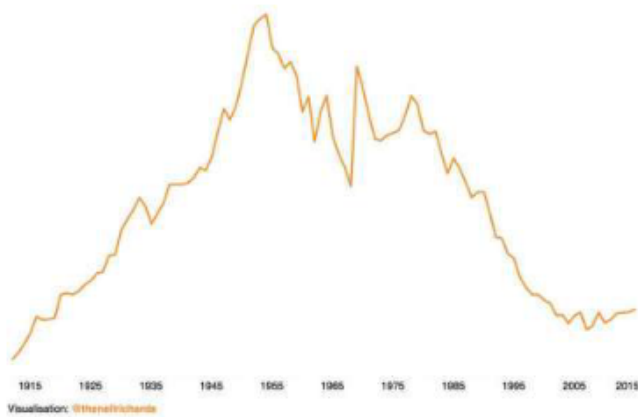
Rise and Fall of the name Neil in the USA
Births 1912-2015
Source: data.gov

Slope graph

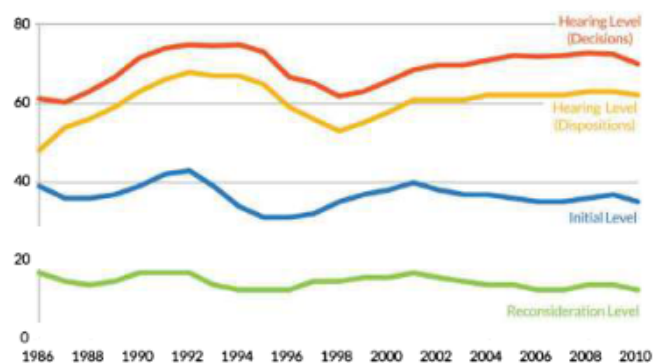Slopegraphs can be useful when you have two time periods or points of comparison and want to quickly show relative increases and decreases or differences across various categories between the two data points

Employee feedback over time

Survey category | Percent favorable

	2014	2015
		96%
		91%
Peers	85%	
Culture	80%	
Work environment	76%	75%
Leadership	59%	62%
Career development	49%	
Rewards & recognition	41%	45%
		42%
Perf management	33%	33%

2014   2015

Survey year

BAR CHATS

Sometimes bar charts are avoided because they are common. This is a mistake. Rather, bar charts should be leveraged because they are common, as this means less of a learning curve for your audience. Instead of using their brain power to try to understand how to read the graph, your audience spends it figuring out what information to take away from the visual
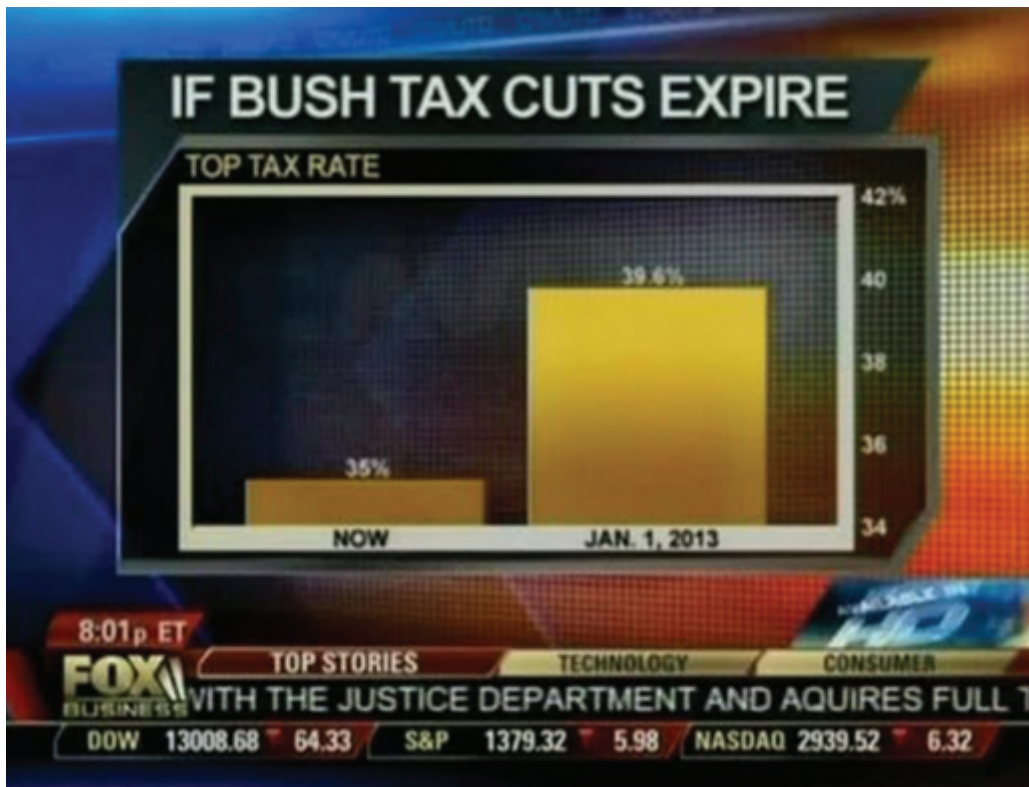
(Millions of people)
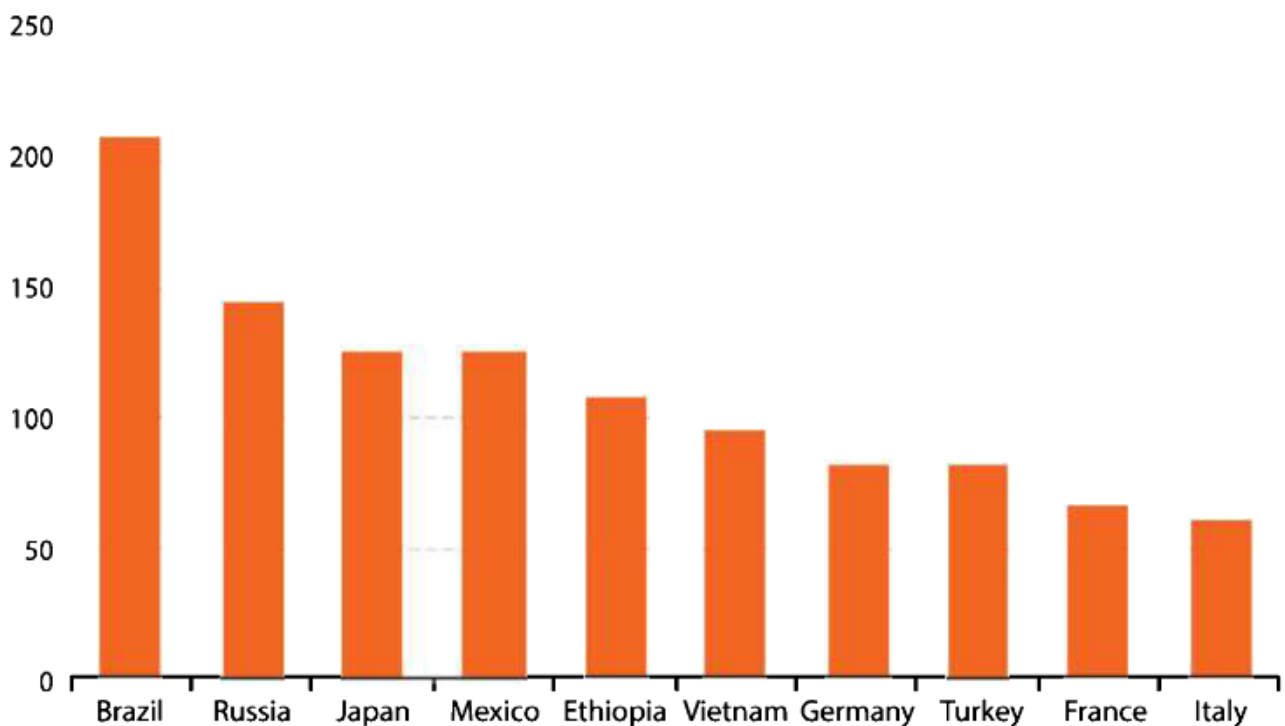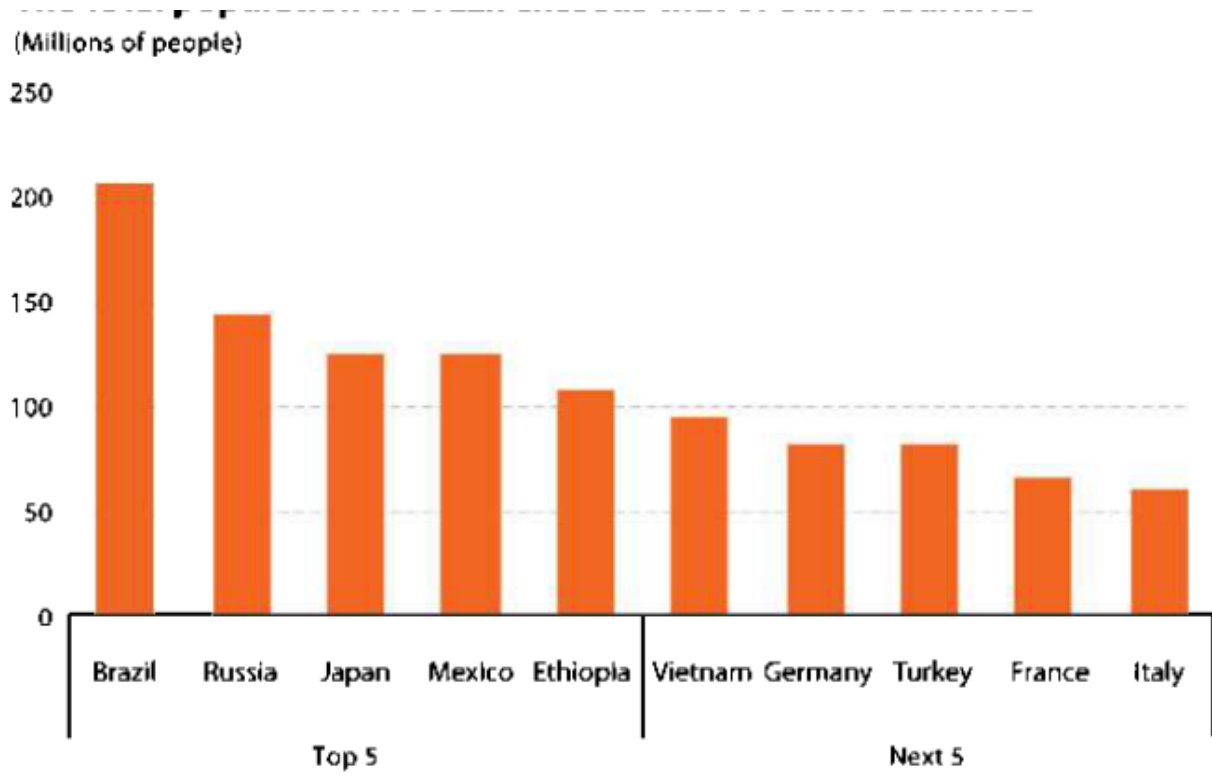
IF ANY SPECIAL CHANGE CHANGE THAT COLOUR IT WILL HIGLETS
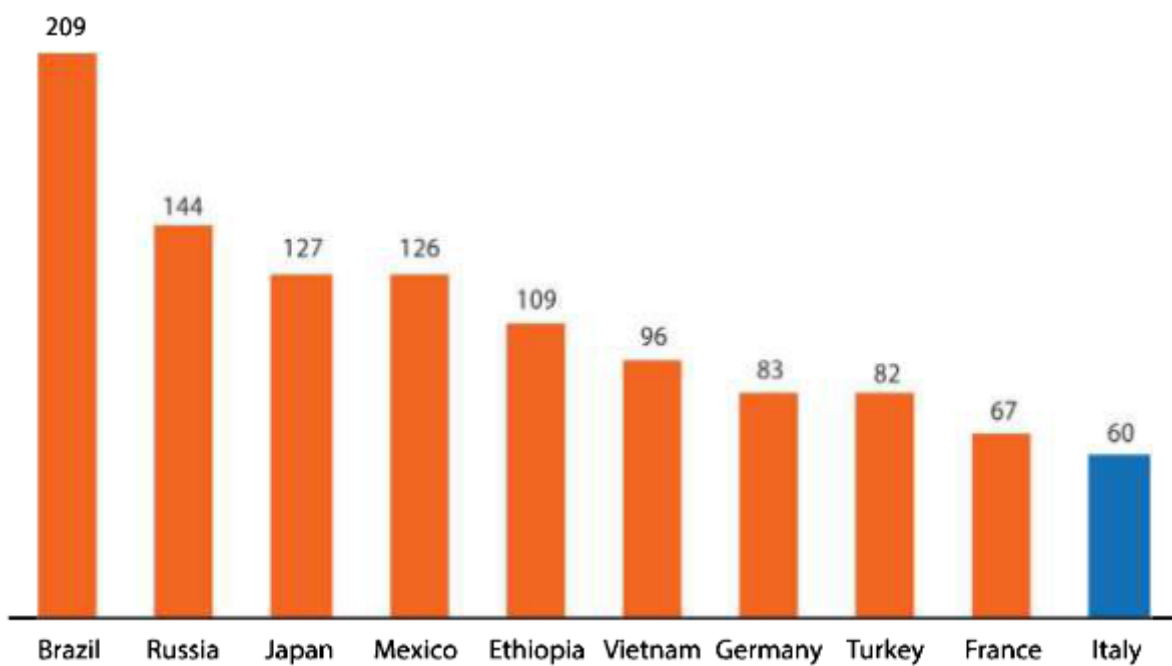
VERTICAL BAR CHARTS

The total population in Brazil exceeds that of other countries

(Millions of people)

WE HAVE LARGE DATA POVERTY EASY TO SEE IN VERTICAL BARS IT TAKES ONE PAGE

Country

India 266.7
Nigeria 84.8
Congo, Dem. Rep. 55.2
Ethiopia 27.6
Tanzania 24.3
Bangladesh 24.3
Kenya 17.8
Uganda 17.3
Madagascar 17.3
Mozambique 17.1
Indonesia 15.1
Malawi 10.8
South Africa 10.4
China 10.0
Zambia 9.3
Niger 8.5
Pakistan 7.7
Burkina Faso 7.7
Mali 7.5
...and Central Afric 7.1
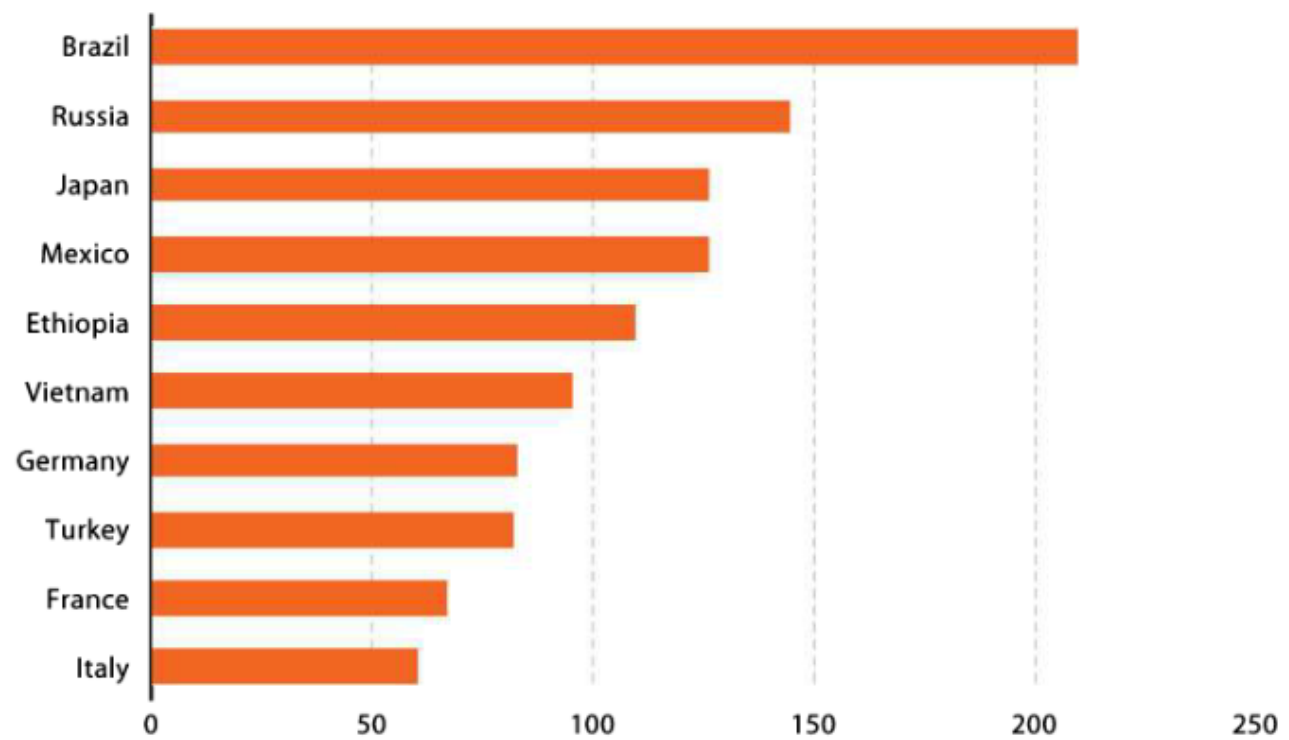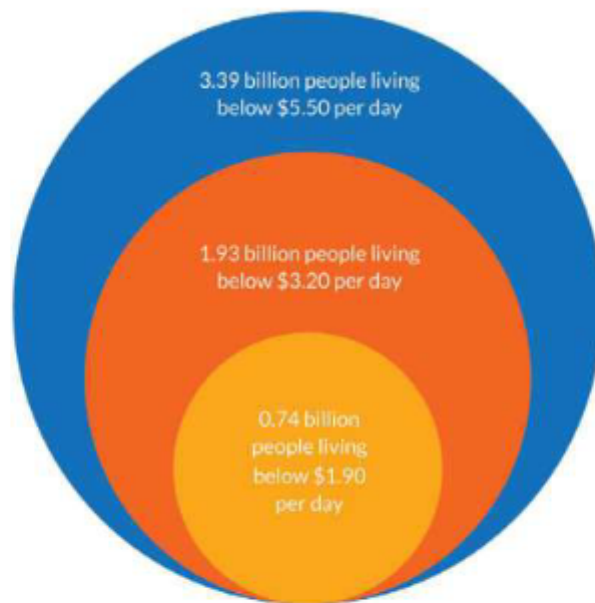Burundi 7.1
Brazil 6.9
Angola 6.6
Cote d'Ivoire 6.5
Rwanda 6.4
Cameroon 5.3
Benin 5.2
Senegal 5.1
Sudan 5.0
Yemen, Rep. 4.9
Chad 4.7
South Sudan 4.3
United States 4.0
Nepal 4.0
Guinea 4.0
Togo 3.6
Sierra Leone 3.5
Myanmar 3.3
Mexico 3.2
Ghana 3.2
Zimbabwe 3.1
...al African Republic 2.9
Papua New Guinea 2.7
Haiti 2.4
Colombia 2.2
Vietnam 1.8
Liberia 1.7
Congo, Rep. 1.7
Lao PDR 1.5
Honduras 1.5
Guatemala 1.4
Egypt, Arab Rep. 1.3
Lesotho 1.2
Italy 1.2
Romania 1.1
Peru 1.1
Guinea-Bissau 1.0
Iraq 0.8
Bolivia 0.8
Rwanda 0.6
Spain 0.5
Ecuador 0.5
Timor-Leste 0.4
Tajikistan 0.4
Morocco 0.4
Botswana 0.4
Namibia 0.3
Japan 0.3
Turkey 0.2
Tunisia 0.2
Sri Lanka 0.2
Nicaragua 0.2
Mauritania 0.2
Iran, Islamic Rep. 0.2
Greece 0.2
Georgia 0.2
Gambia, The 0.2
Dominican Republic 0.2
Djibouti 0.2
Chile 0.2
Canada 0.2
Argentina 0.2
Algeria 0.2
United Kingdom 0.1
Solomon Islands 0.1
Sao Tome and Principe 0.1
Portugal 0.1
Paraguay 0.1
Panama 0.1
Macedonia, FYR 0.1
Kyrgyz Republic 0.1
Korea, Rep. 0.1
Gabon 0.1
El Salvador 0.1
Costa Rica 0.1
Comoros 0.1
Bulgaria 0.1
Aruba 0.1
Australia 0.1
Armenia 0.1

0   50   100   150   200   250   300

I HAVE A BETTER OPTION FOR THAT NESTED BUBBLE CHART

3.39 billion people living below $5.50 per day

1.93 billion people living below $3.20 per day

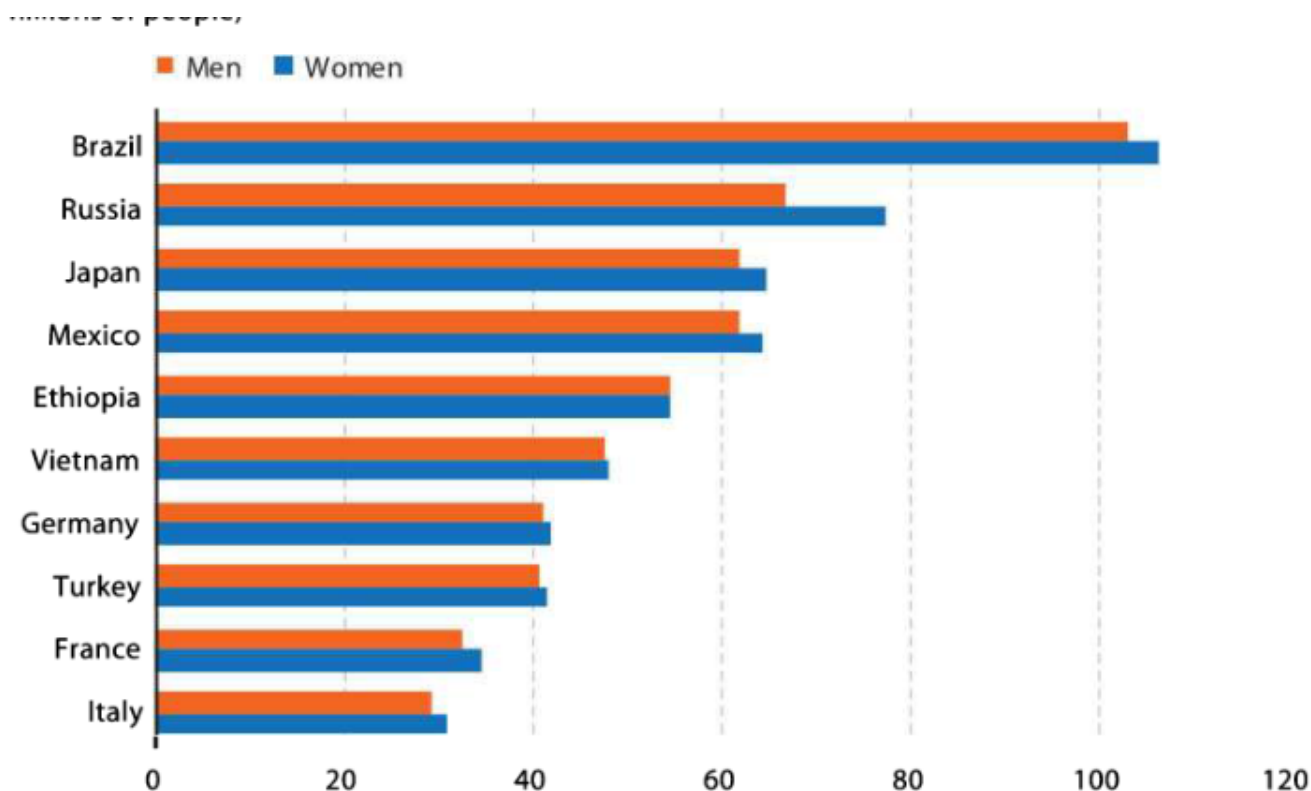0.74 billion people living below $1.90 per day

Source: The World Bank

The *nested bubble chart* can sometimes mask circles in the back—but it can also make for easier comparisons.

The total population in Brazil exceeds that of other countries

(Millions of people)
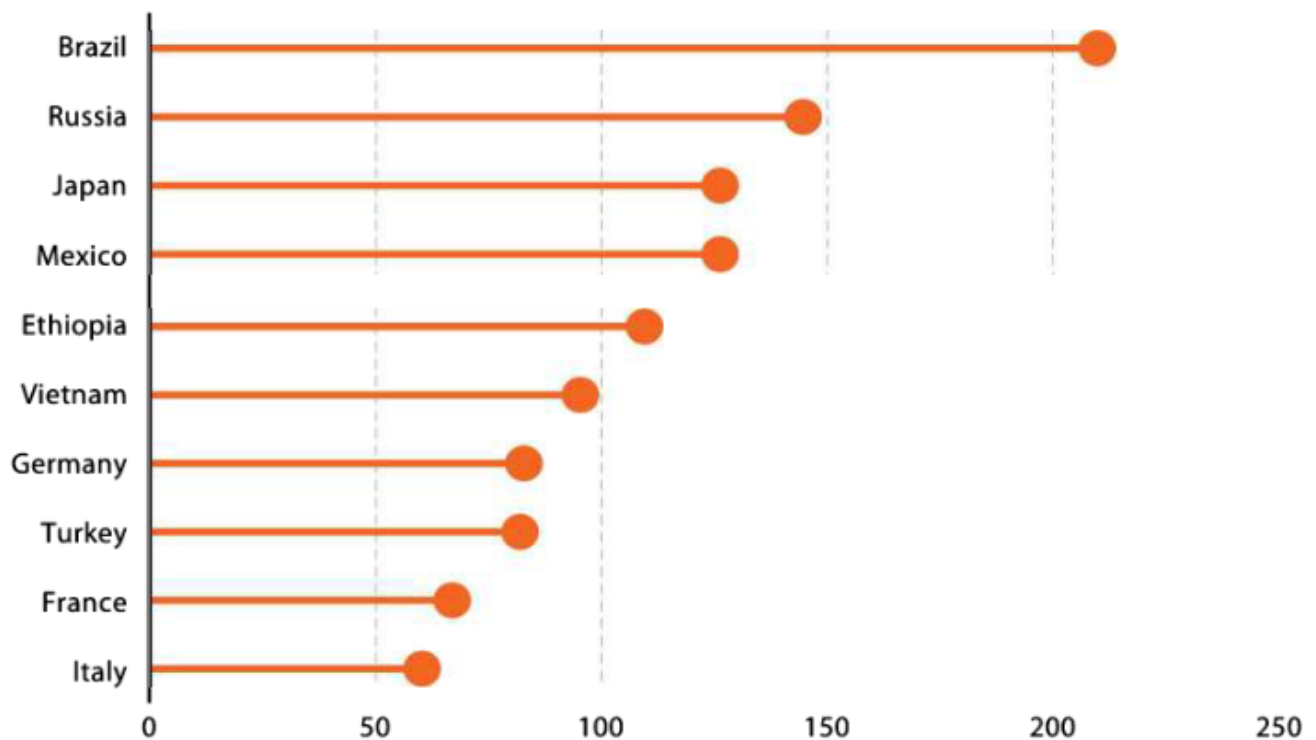
Stacked vertical bar chart

Use cases for stacked vertical bar charts are more limited. They are meant to allow you to compare totals across categories and also see the subcomponent pieces within a given category. This can quickly become visually overwhelming, however—especially given the var ied default color schemes in most graphing applications
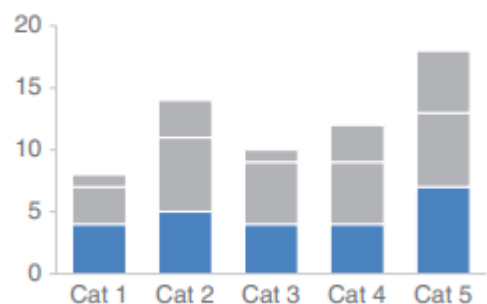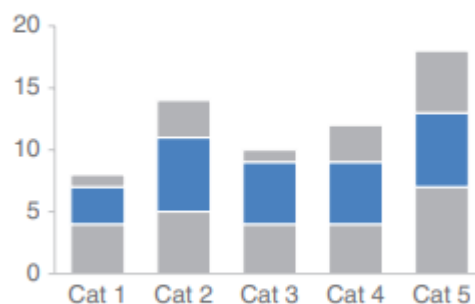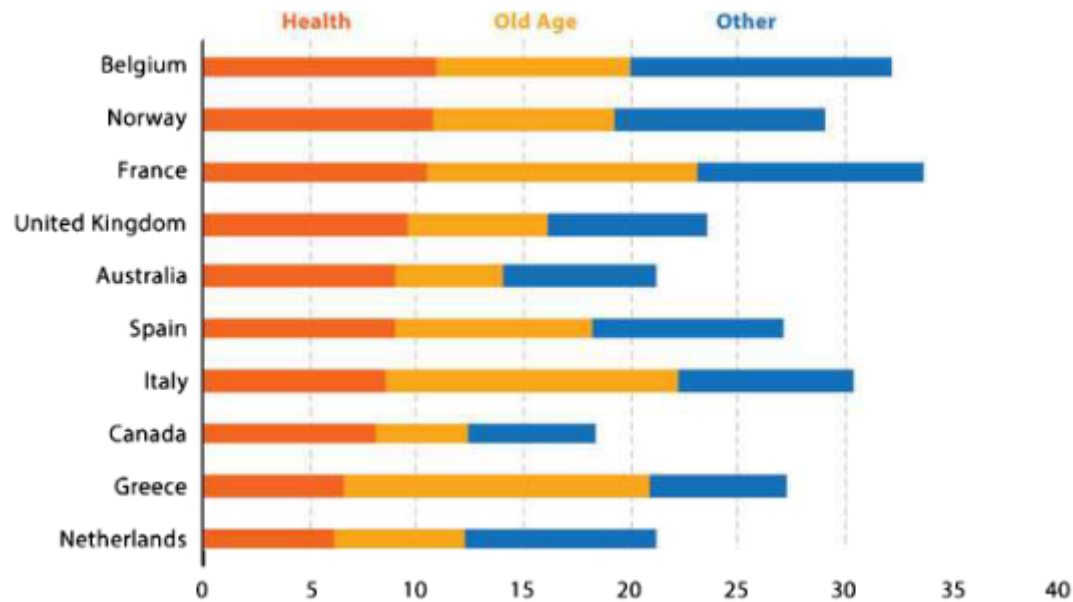

FIGURE 2.16   Comparing series with stacked bar charts
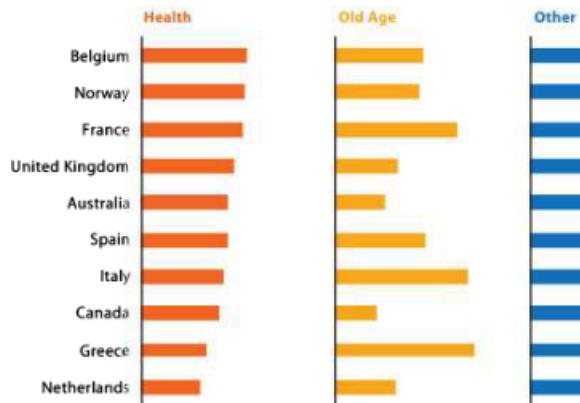
STACKED BARS

(Percent of GDP)

Source: Organization for Economic Co-Operation and Development

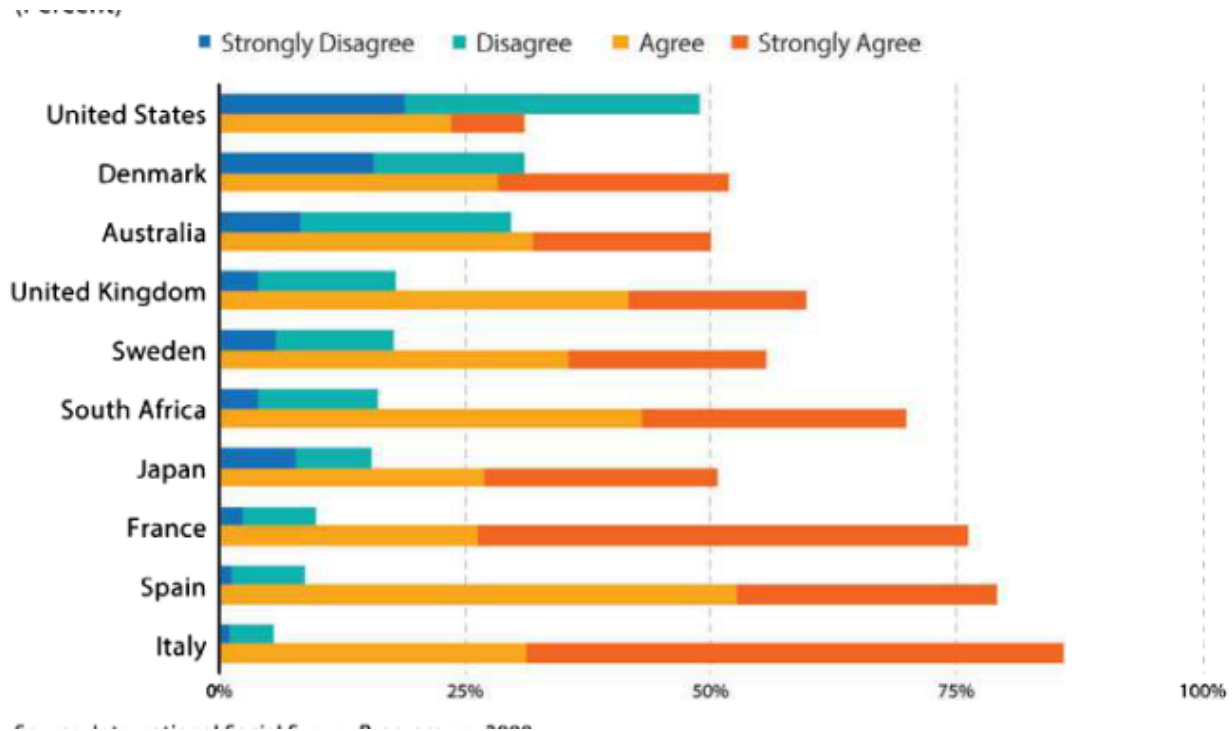**Social expenditures for 10 OECD countries**
(Percent of GDP)

Social expenditures for 10 OECD countries
(Percent of GDP)

Instead of stacking all the data series together, we can break them up (either with or without the totals) to create a sort of small-multiples approach. Here, we move up to the top of the perceptual ranking list because each series sits on its own baseline.

Source: International Social Survey Programme 2000

DOT PLOTS

Because the data spike in the last year, a dot plot is probably not a good way to show the change from 2015 to 2020.

Because the data move in the same direction, a dot plot is probably sufficient to show the change from 2015 to 2020.
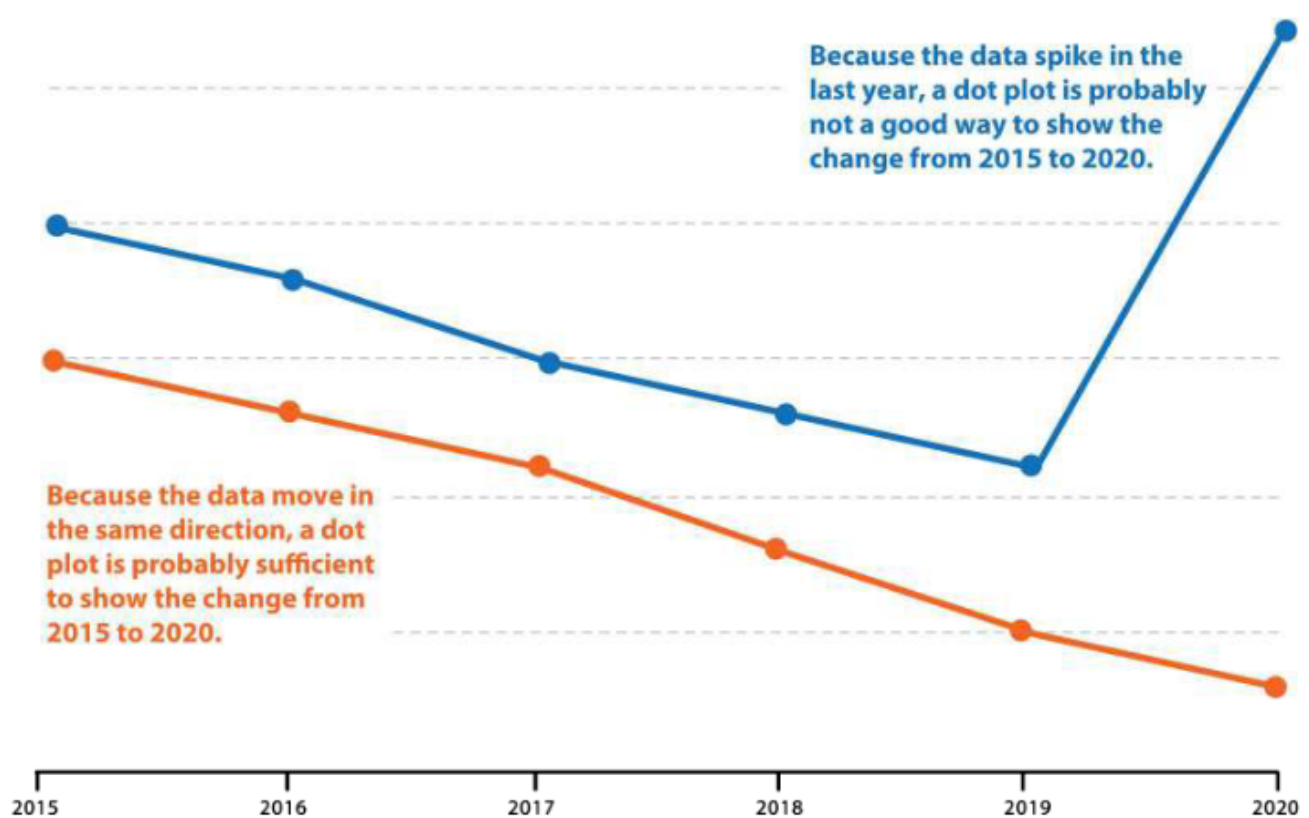
FIGURE 4.14 Data labels used sparingly help draw attention

GAGUGE AND BULLET CHART

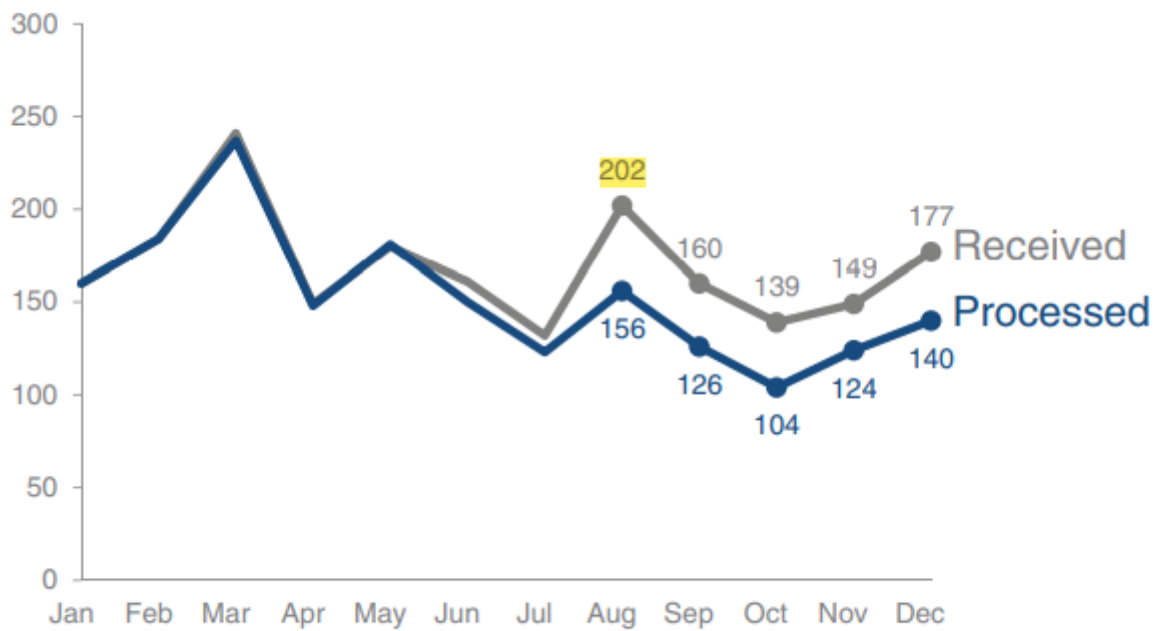BETTER VIEW TAGET IS REACHED ARE NOT

Gauge charts are familiar and easy to read.

NORTHERN VIRGINIA MARKET STATISTICS

Northern Virginia Closed Sales
- Nov 2019 (2,497)
- 5-Yr Nov Avg (2,486)

© 2019 MarketStats by ShowingTime. Data provided by Bright MLS as of Nov 6, 2019

Northern Virginia New Pendings
- Nov 2019 (2,541)
- 5-Yr Nov Avg (2,627)

© 2019 MarketStats by ShowingTime. Data provided by Bright MLS as of Nov 6, 2019

Northern Virginia Active Listings
- Nov 2019 (3,748)
- 5-Yr Nov Avg (6,006)

© 2019 MarketStats by ShowingTime. Data provided by Bright MLS as of Nov 6, 2019

Northern Virginia Months of Supply
- Nov 2019 (1.2)
- 5-Yr Nov Avg (2.0)

© 2019 MarketStats by ShowingTime. Data provided by Bright MLS as of Nov 6, 2019
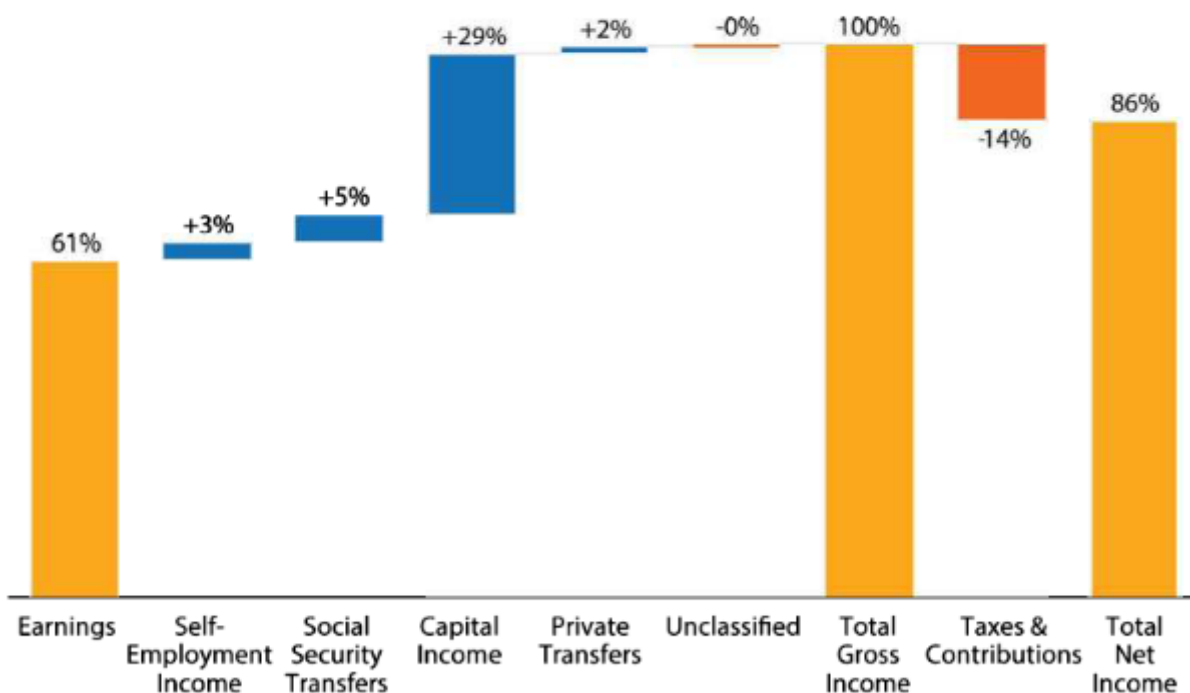
WATERFALL CHARTS

a waterfall charts show basic mathematical equation : adding or subtracting values from some initial values to produce a final amount


Income composition in Australia in 2016
(Percent of total gross income)

	+29%	+2%	-0%	100%		
+5%					-14%	86%
+3%						
61%						

Earnings | Self-Employment Income | Social Security Transfers | Capital Income | Private Transfers | Unclassified | Total Gross Income | Taxes & Contributions | Total Net Income
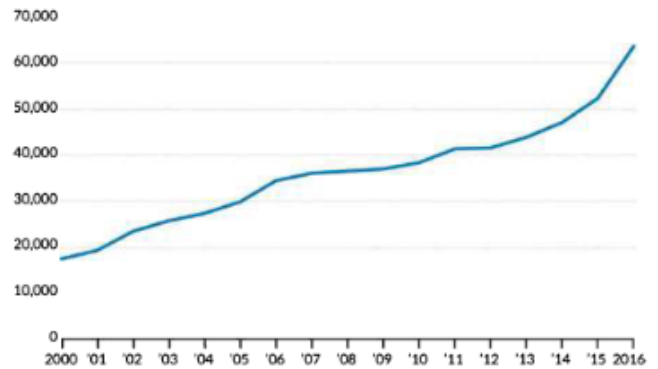
AREA CHARTS

area charts are line graphs with the area below the line filled in , giving the series more visual weight .
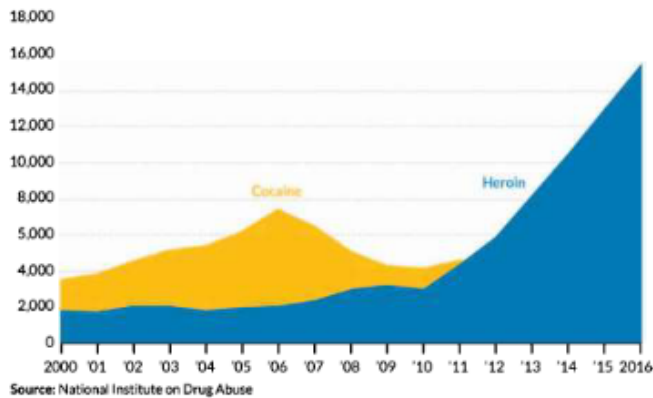
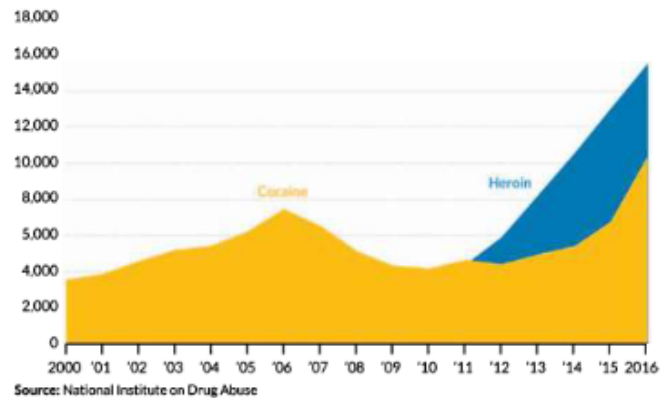## More than 60,000 people died from drug overdoses in 2016
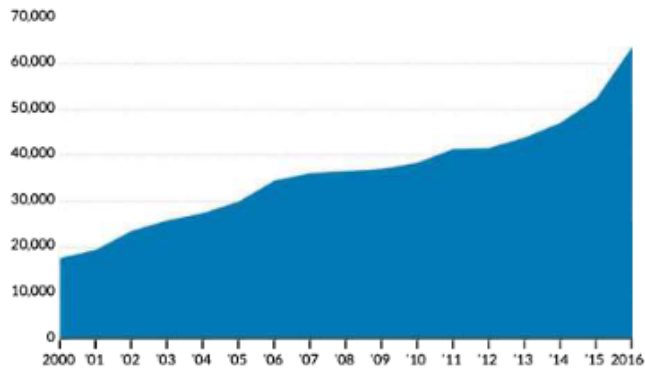
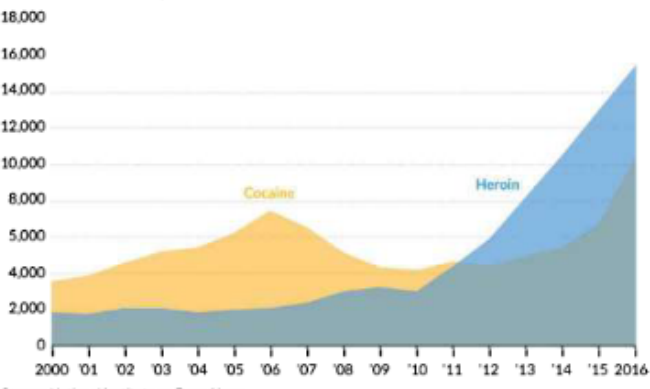## More than 60,000 people died from drug overdoses in 2016

More than 10,000 people died from cocaine drug overdoses in 2016, and more than 15,000 died from heroin overdoses in 2016
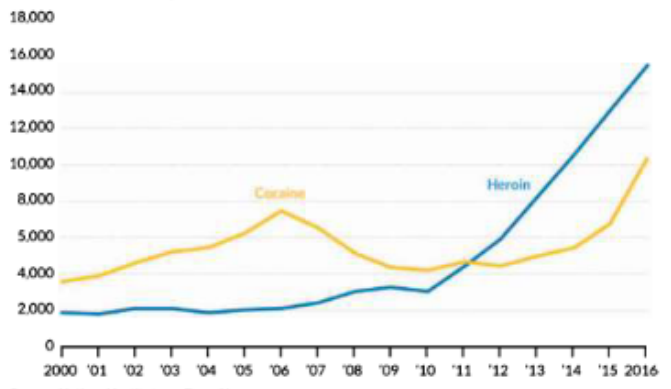

Cocaine
Heroin

Source: National Institute on Drug Abuse

More than 10,000 people died from cocaine drug overdoses in 2016, and more than 15,000 died from heroin overdoses in 2016

Cocaine
Heroin

Source: National Institute on Drug Abuse

More than 10,000 people died from cocaine drug overdoses in 2016, and more than 15,000 died from heroin overdoses in 2016

Cocaine
Heroin

More than 10,000 people died from cocaine drug overdoses in 2016, and more than 15,000 died from heroin overdoses in 2016
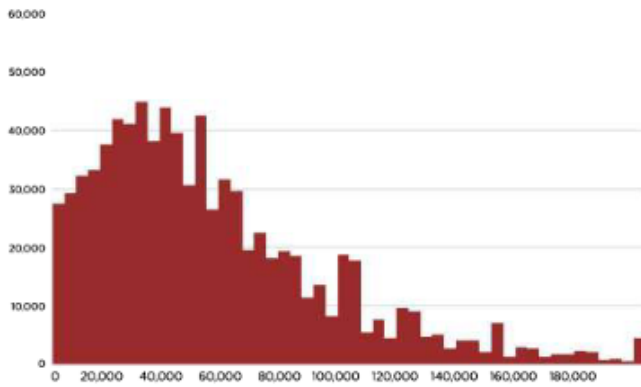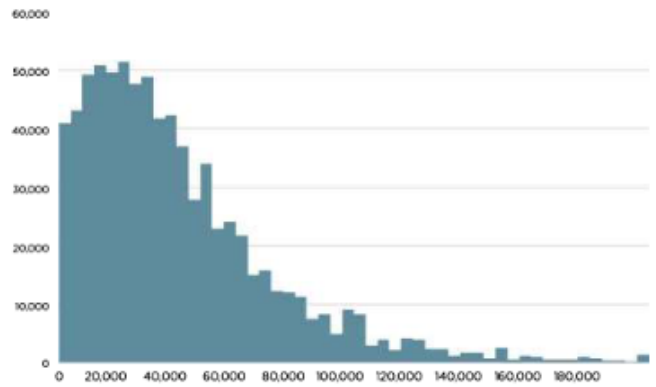

Cocaine
Heroin

HISTOGRAMS

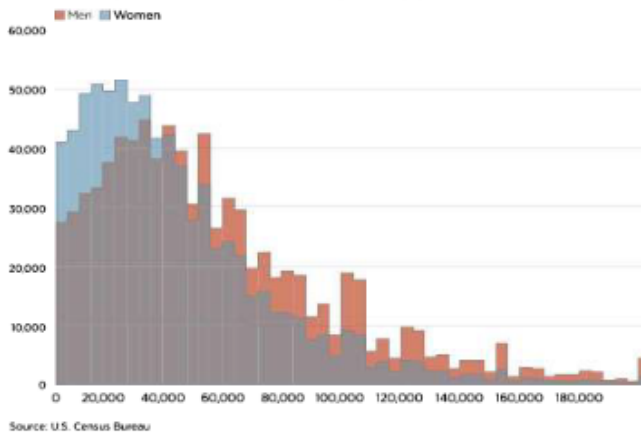HISTOGRAMS ARE MOST BASIC GRAPH TYPE FOR VISULIAZING A DISTRIBIATION

MEN'S EARNINGS DISTRIBUTION IN 2016

WOMEN'S EARNINGS DISTRIBUTION IN 2016

Source: U.S. Census Bureau

MEN'S AND WOMEN'S EARNINGS DISTRIBUTIONS IN 2016

Men   Women

MEN'S AND WOMEN'S EARNINGS DISTRIBUTIONS IN 2016
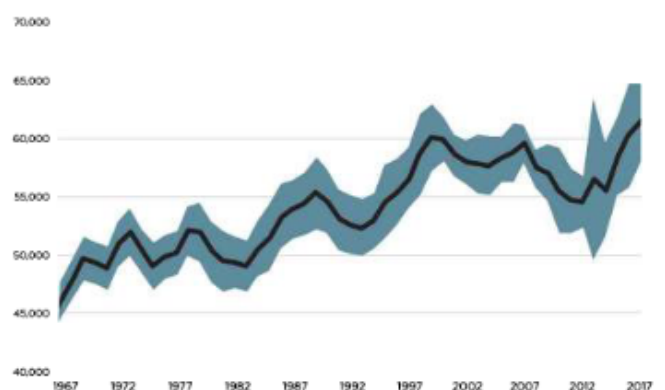
Men   Women

Source: U.S. Census Bureau

CONFINDENCE INTERVAL
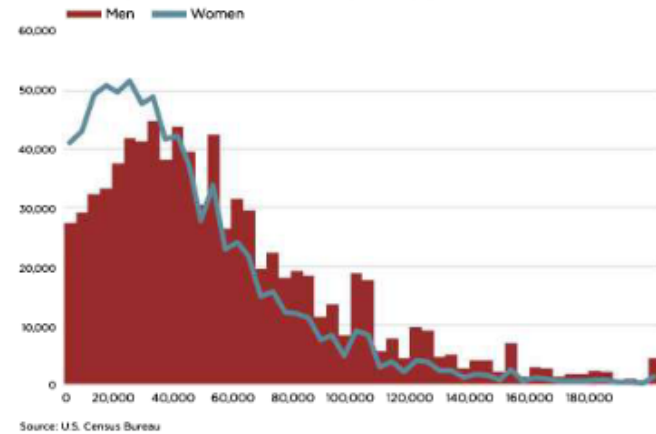
a confidence interval chart typically used linear or shaded area to depict ranges or amount of uncertainty

MEDIAN INCOME IN THE UNITED STATES FROM 1967 TO 2017
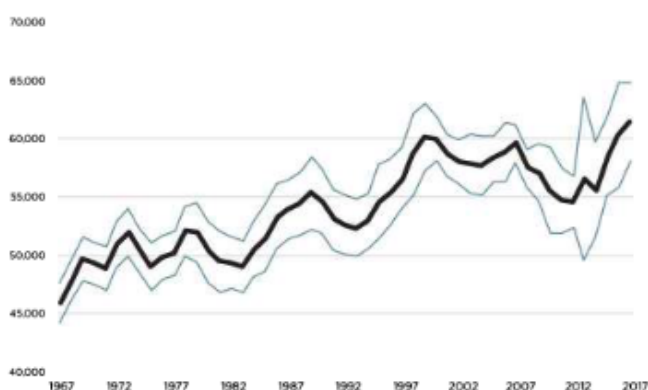
MEDIAN INCOME IN THE UNITED STATES FROM 1967 TO 2017

box - and - whisker plot
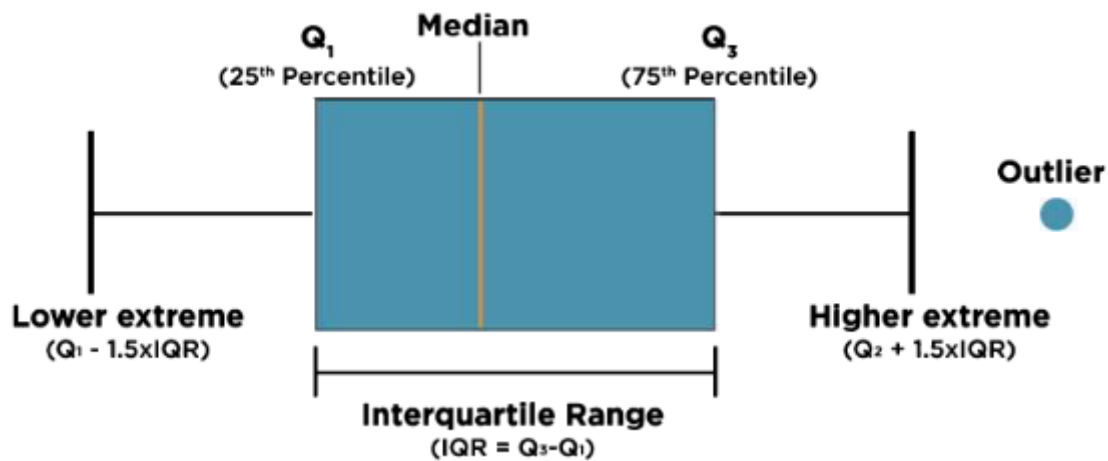
Never use 3D

One of the golden rules of data visualization goes like this: never use 3D. Repeat after me: never use 3D. The only exception is if you are actually plotting a third dimension (and even then, things get really tricky really quickly, so take care when doing this)—and you should never use 3D to plot a single dimension

Note the multiple preattentive attributes I've used in the preceding text to underscore its importance!
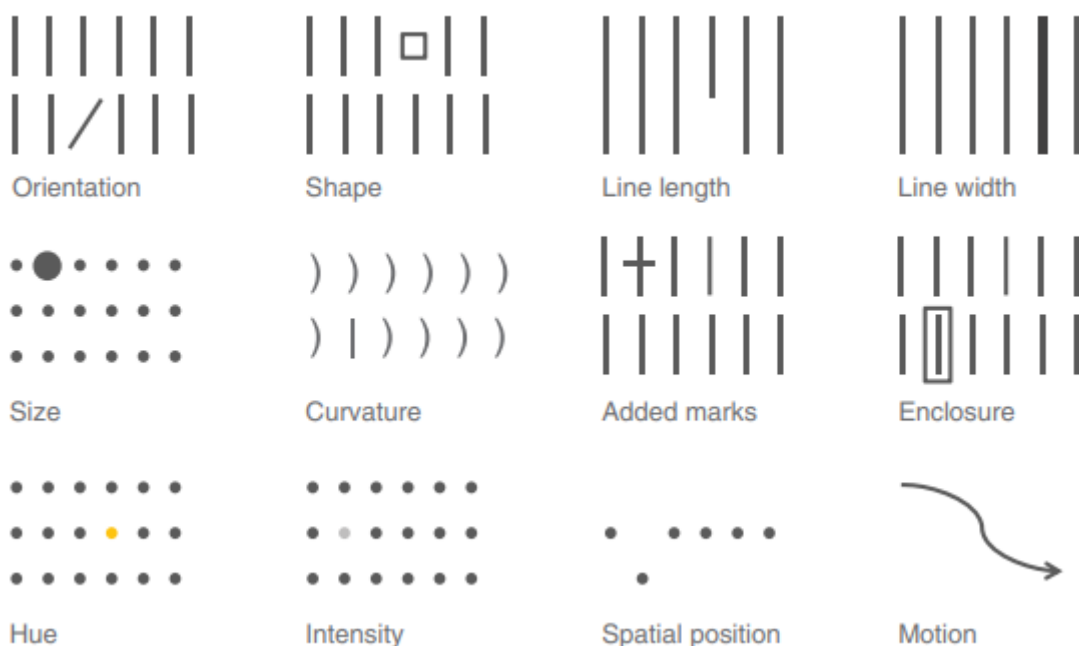
your eye is drawn to the one element within each group that is different from the rest: you don't have to look for it. That's because our brains are hardwired to quickly pick up differences we see in our environment

Preattentive attributes in text

No preattentive attributes

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Bold

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Color

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Italics

What are we doing well? Great Products. These products are clearly the best in their class. *Replacement parts are shipped when needed*. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Size

What are we doing well? Great Products. These products are the best in their class. Replacement parts are shipped when needed. You sent gaskets

without me having to
ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Separate spatially

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask.

Problems are resolved promptly.

Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Outline (enclosure)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!
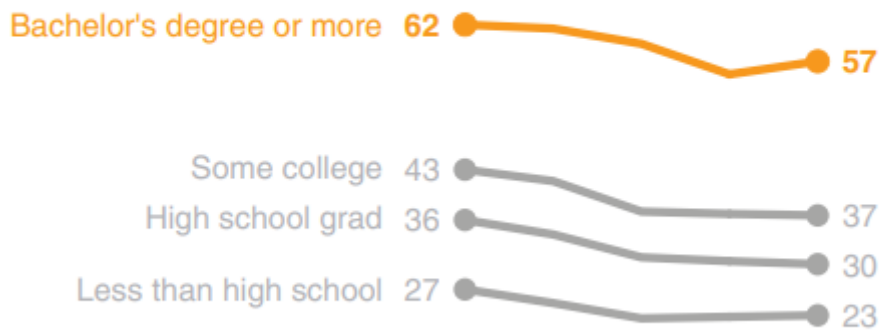
Underline (added marks)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

• Not all data are equally important. Use your space and audience's attention wisely by getting rid of noncritical data or components

New marriage rate by education

Number of newly married adults per 1,000 marriage eligible adults

Bachelor's degree or more 62 — 57
Some college 43 — 37
High school grad 36 — 30
Less than high school 27 — 23

• Make it legible: use a consistent, easy-to-read font (consider both typeface and size). •

Keep it clean: make your data visualization approachable by leveraging visual affordances.

• Use straightforward language: choose simple language over complex, choose fewer words over more words, define any specialized language with which your audience may not be familiar, and spell out acronyms (at minimum, the first time you use them or in a footnote).

• Remove unnecessary complexity: when making a choice between simple and complicated, favor simple

dissecting model visuals

Model visual #1: line graph

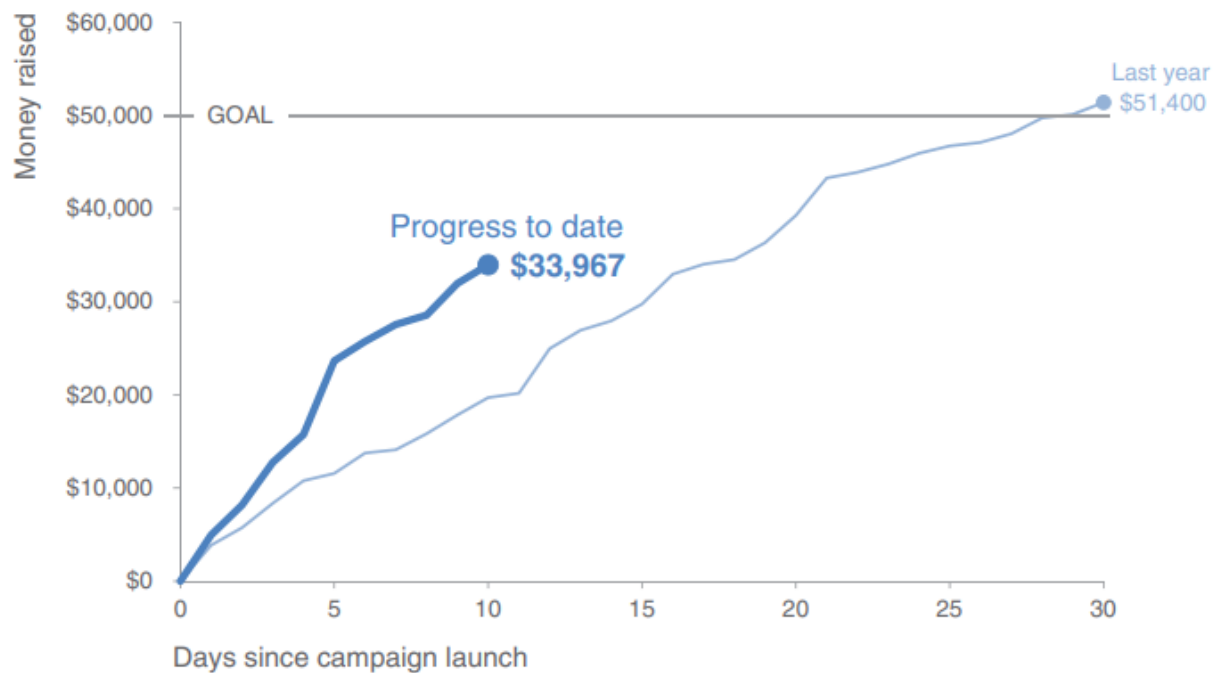## Annual giving campaign progress

FIGURE 6.1 Line graph

Company X runs an annual month-long "giving campaign" to raise money for charitable causes. Figure 6.1 shows this year's progress to date. Let's consider what makes this example good and the deliberate choices made in the course of its creation.

Model visual #2: annotated line graph with forecast

Sales over time

Sales ($Billion)

2006–09: annual sales growth of 7–8%

2010: more marked increase of 22% sales year over year, driven by a, b, and c

2011–14: another period of steady growth of 8–9% annually

2015 & beyond: assumed 10% year over year increase in sales*

$108 $119 $131 $144 $158

2006 2007 2008 2009 2010 2011 2012 2013 2014 | 2015 2016 2017 2018

ACTUAL   FORECAST