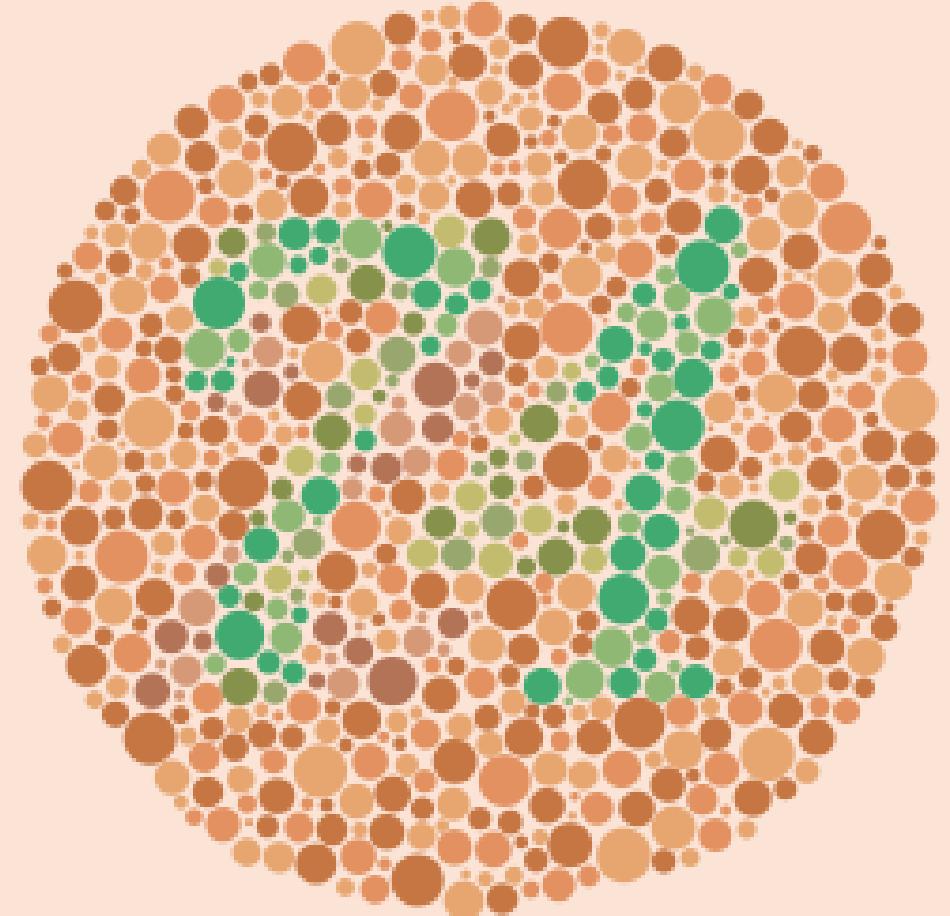


A blue pen with a silver clip and a translucent barrel lies diagonally across a bar chart. The chart consists of several vertical bars of varying heights, all colored blue. The bars are set against a background of light blue horizontal grid lines. The pen's tip points towards the bottom-left corner of the image.

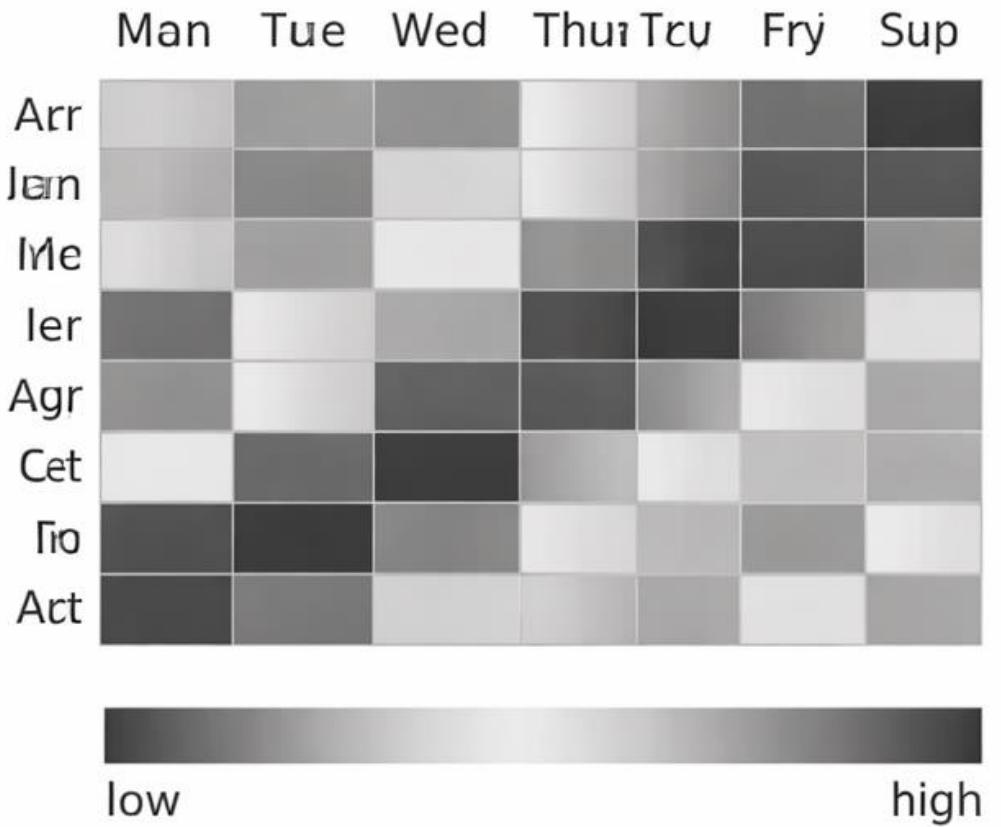
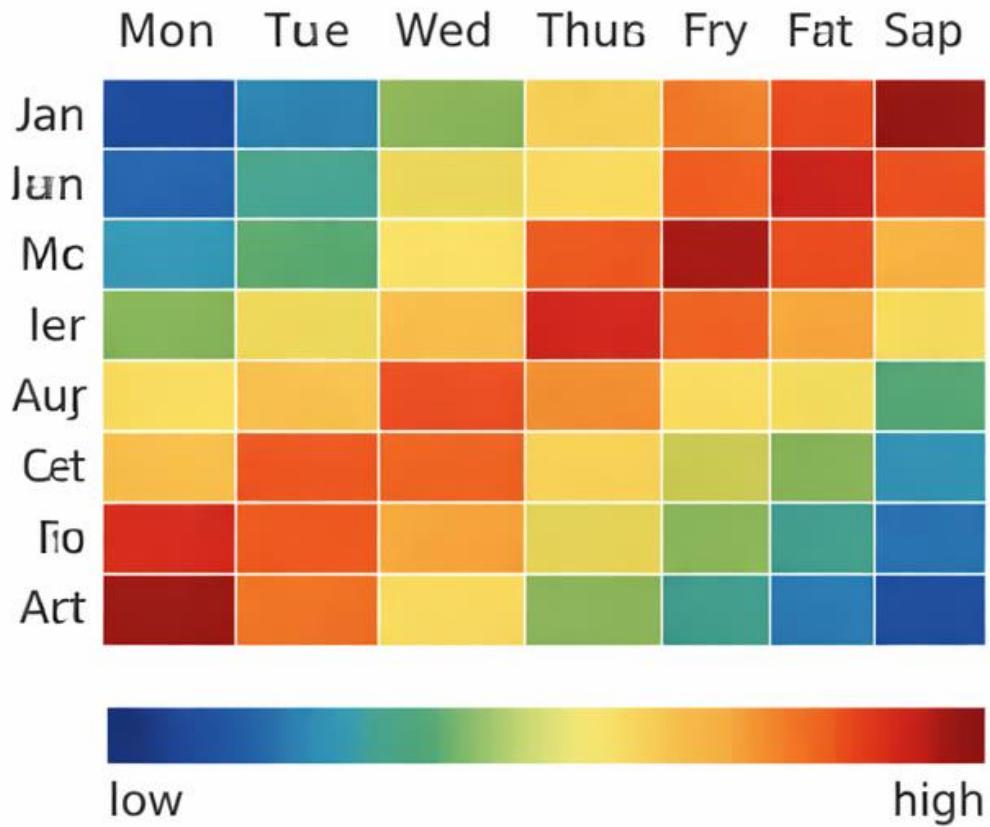
Charts

Excursion: Data Visualization & visual impairments



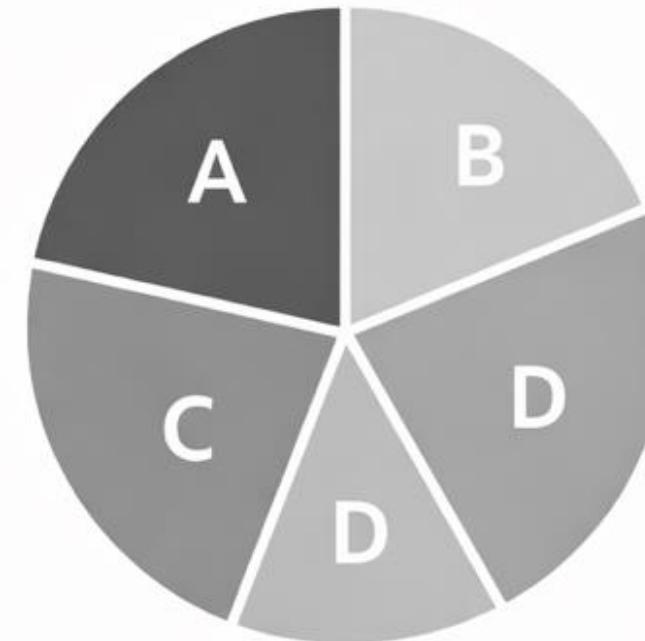
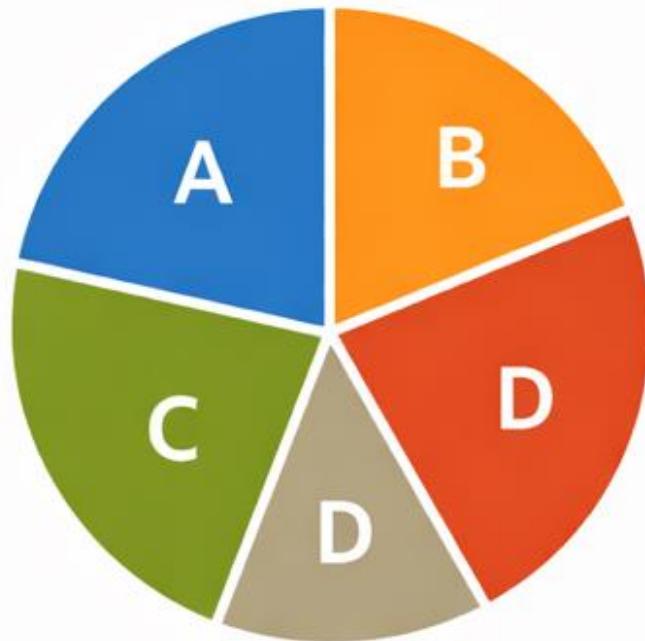
Charts may fail when color is lost

Heatmaps

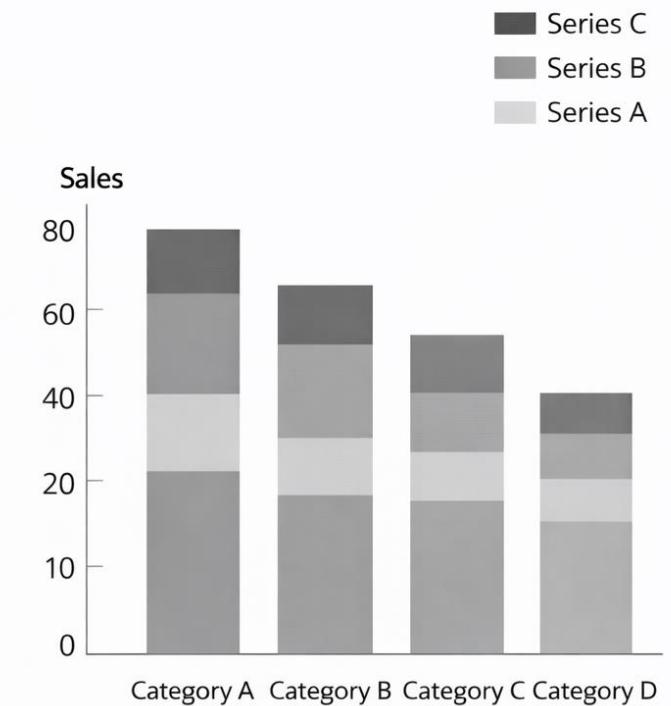
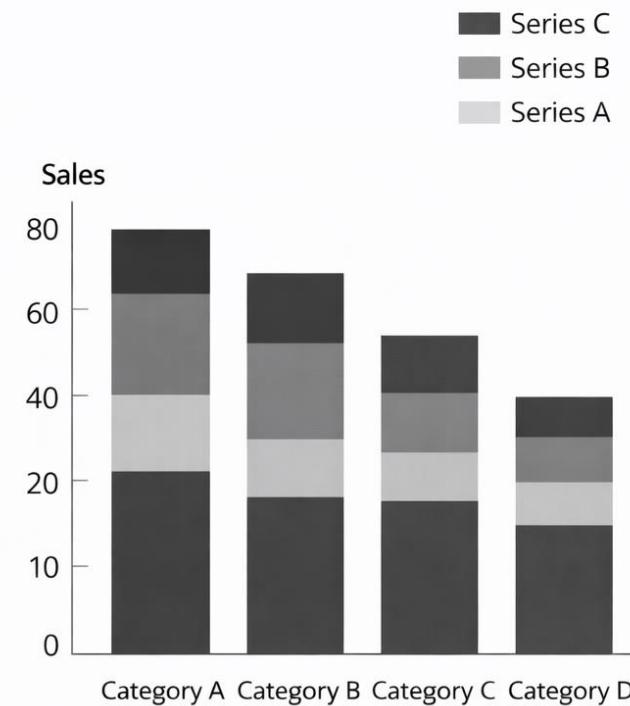
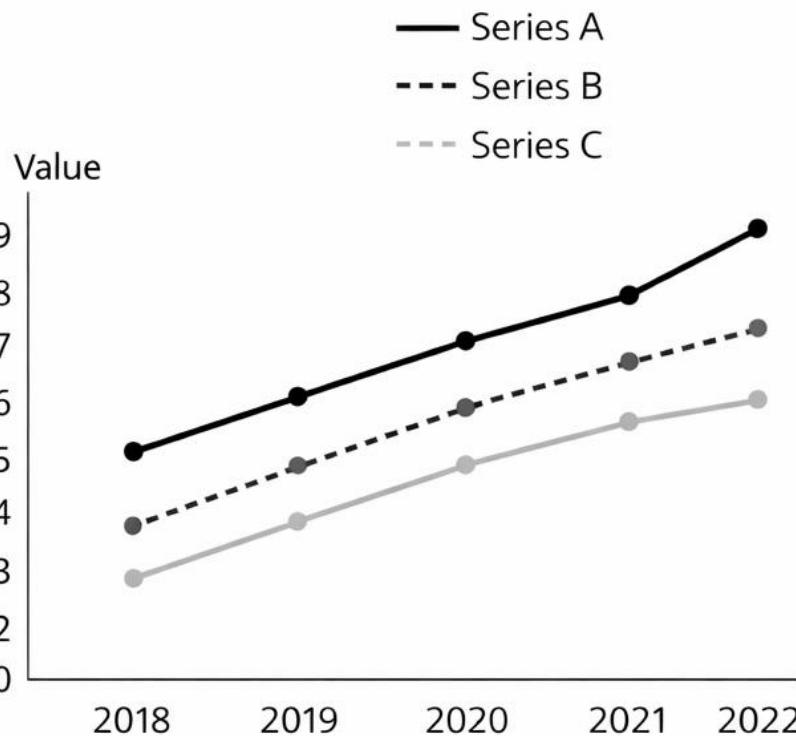


Charts may fail when color is lost

Pie Charts



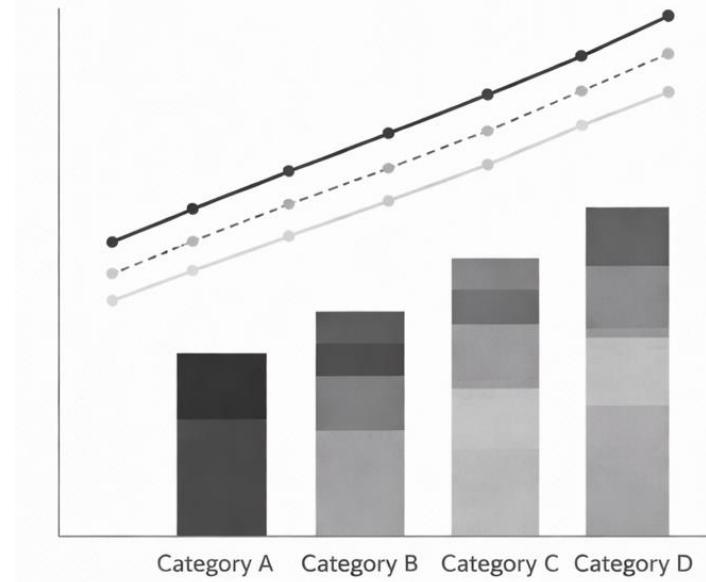
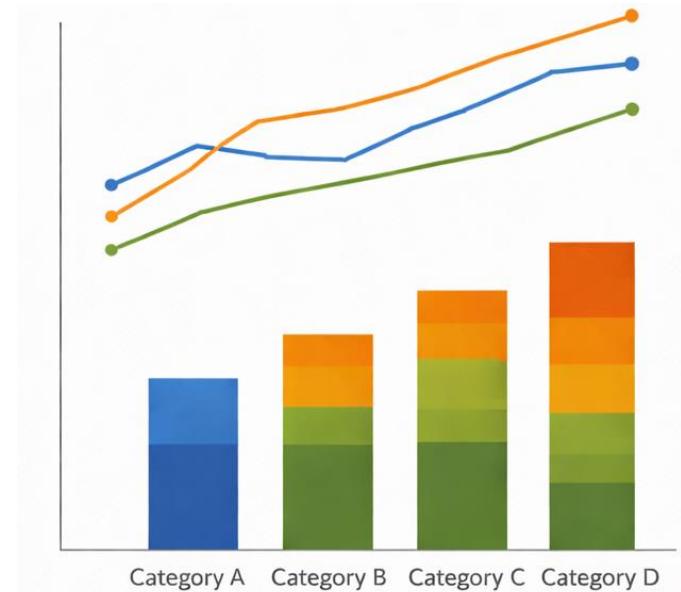
Charts may fail when color is lost



Line charts and bar charts are the safest options

Best practices

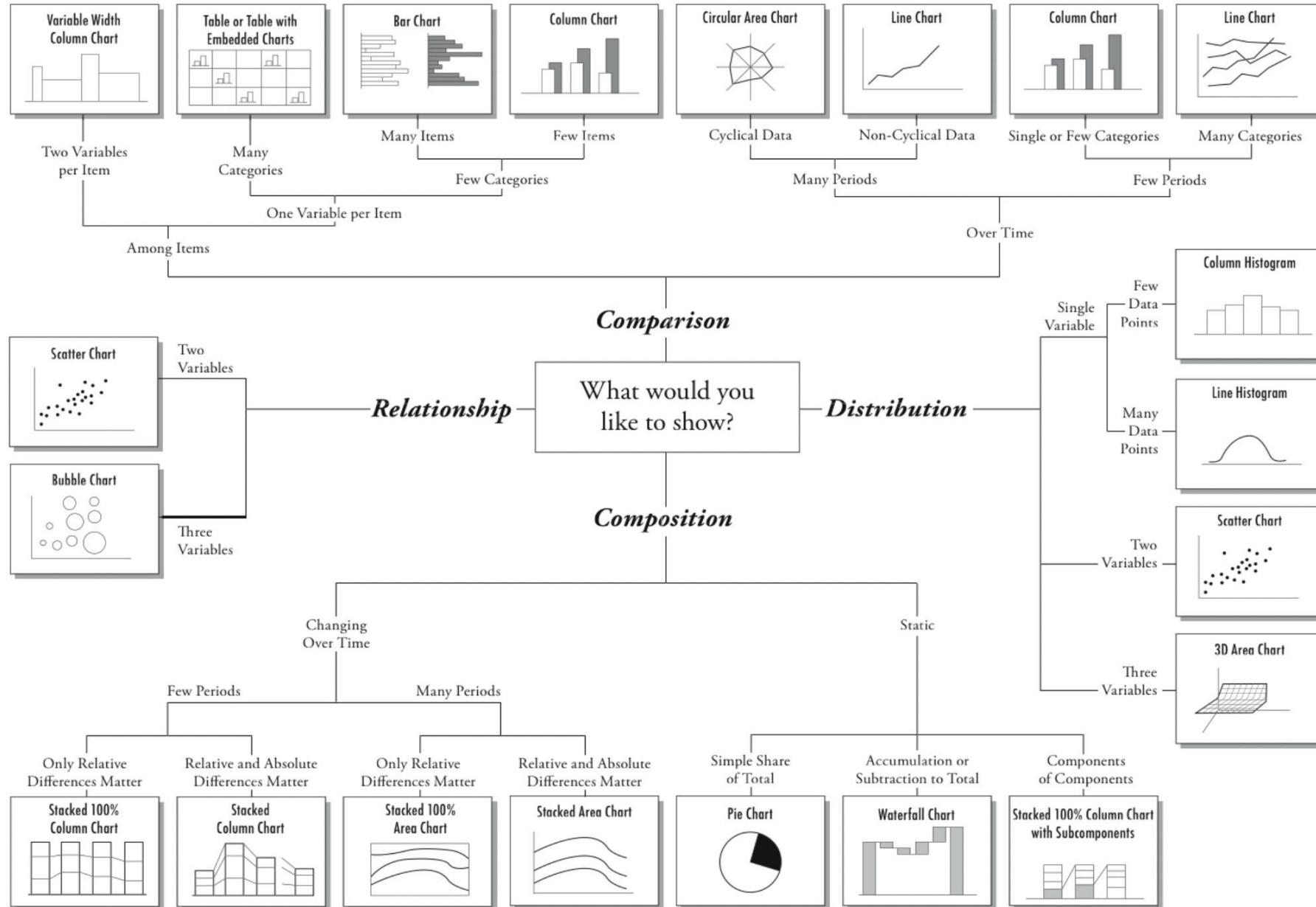
- Color should **not** be the **only carrier of information**
- Color **emphasizes** — it should not **explain**
- **Contrast** beats aesthetics
- Contrasting visuals **improve robustness**



How do you pick the right chart or graph for your data?

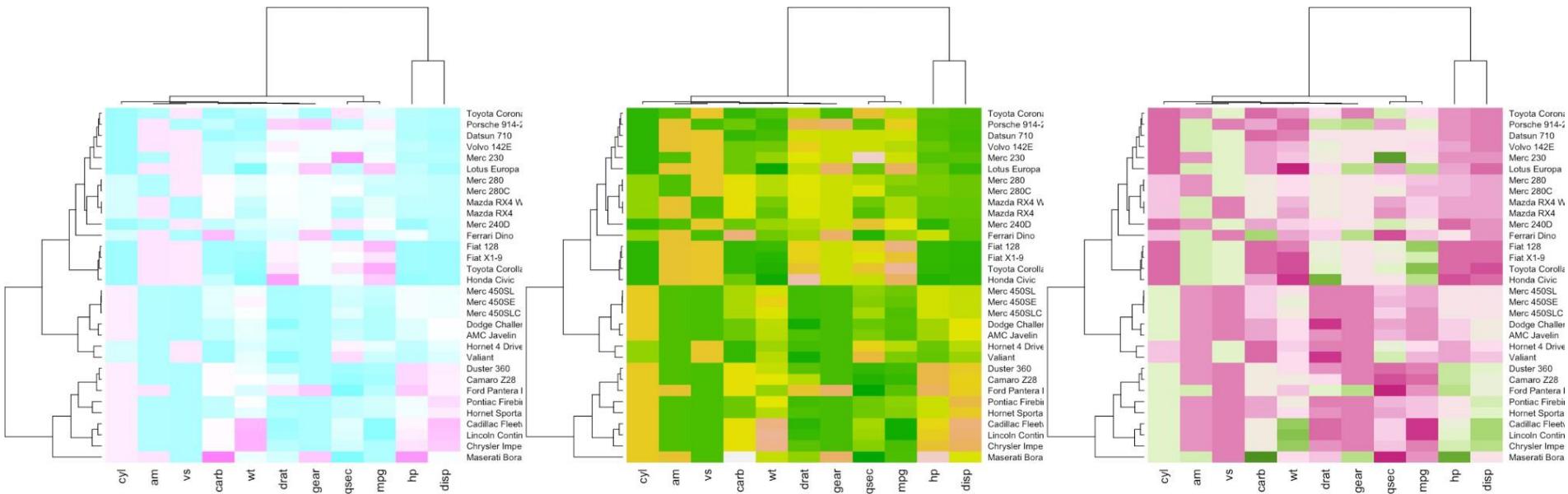
- Before making a chart it's important to understand why you need one.
- Charts help people understand complicated data, find patterns, identify trends, and tell a story.
- Think about the message you want to share with your audience.
 - How many variables do you want to show?
 - How many data points you want to display?
 - How you want to scale your axis?

Chart Suggestions – A Thought Starter



Heat Map

A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colors.



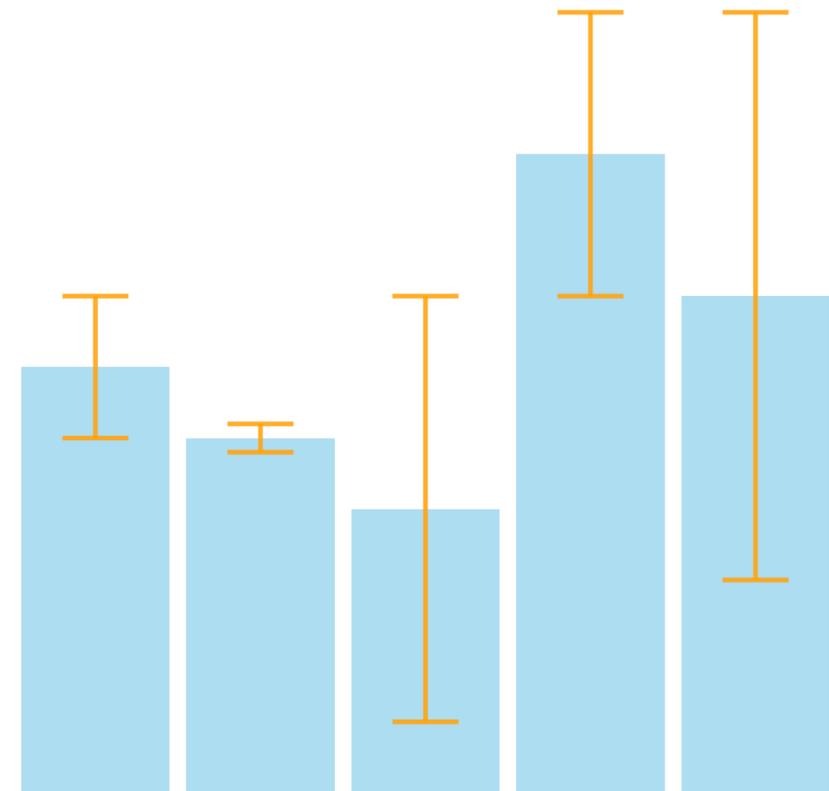
Heatmaps are built thanks to the `heat map()` function, natively provided in R.

Bar plot

One of the most common types of graphic. It shows the relationship between a numeric and a categoric variable.

`ggplot2` allows to build barplot thanks to the `geom_bar()` function

Basic R can build quality barplots thanks to the `barplot()` function



```
# Most basic error bar
ggplot(data) +
  geom_bar( aes(x=name, y=value), stat="identity", fill="skyblue", alpha=0.7) +
  geom_errorbar( aes(x=name, ymin=value-sd, ymax=value+sd), width=0.4, colour="orange", alpha=0.9, size=1.3)
```

For a better visualization, barplots can be replaced by...

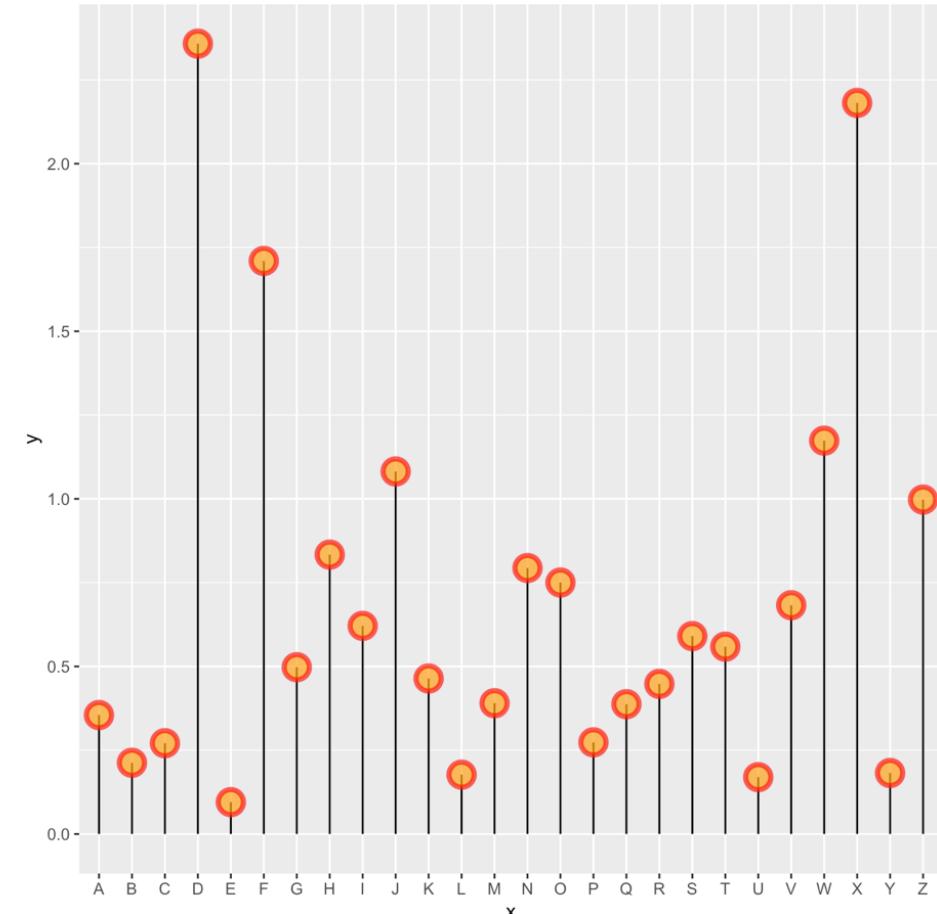
- Lollipop plot that replaces the bars with a segment and a dot.

```
# Plot
ggplot(data, aes(x=x, y=y)) +
  geom_point() +
  geom_segment( aes(x=x, xend=x, y=0, yend=y))

# Library
library(tidyverse)

# Create data
data <- data.frame(
  x=LETTERS[1:26],
  y=abs(rnorm(26))
)

# plot
ggplot(data, aes(x=x, y=y)) +
  geom_segment( aes(x=x, xend=x, y=0, yend=y)) +
  geom_point( size=5, color="red", fill=alpha("orange", 0.3), alpha=0.7, shape=21, stroke=2)
```

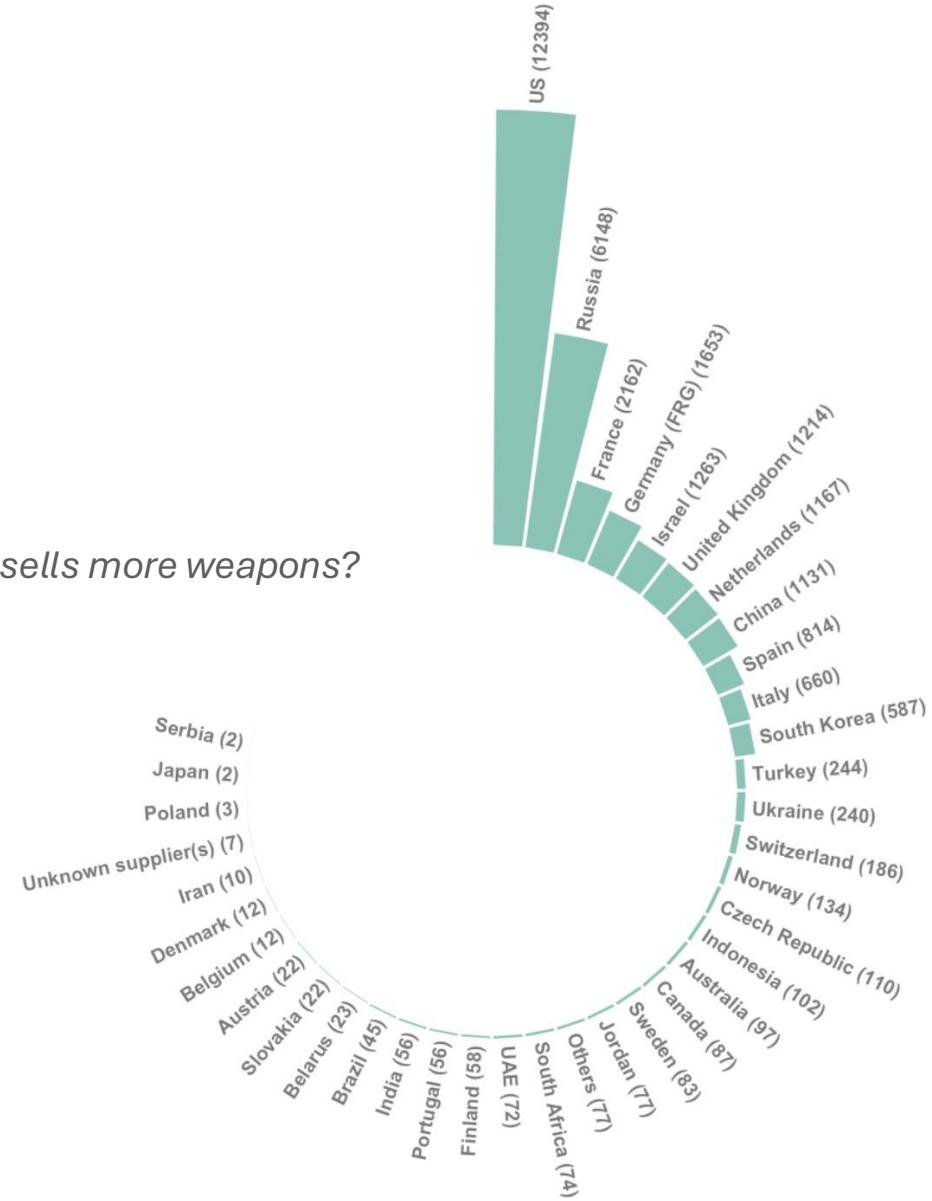


Circular barplot can also be considered if you have many groups.

The group comparison is less accurate; thus, this option must be selected if there is an obvious pattern to show in the data.

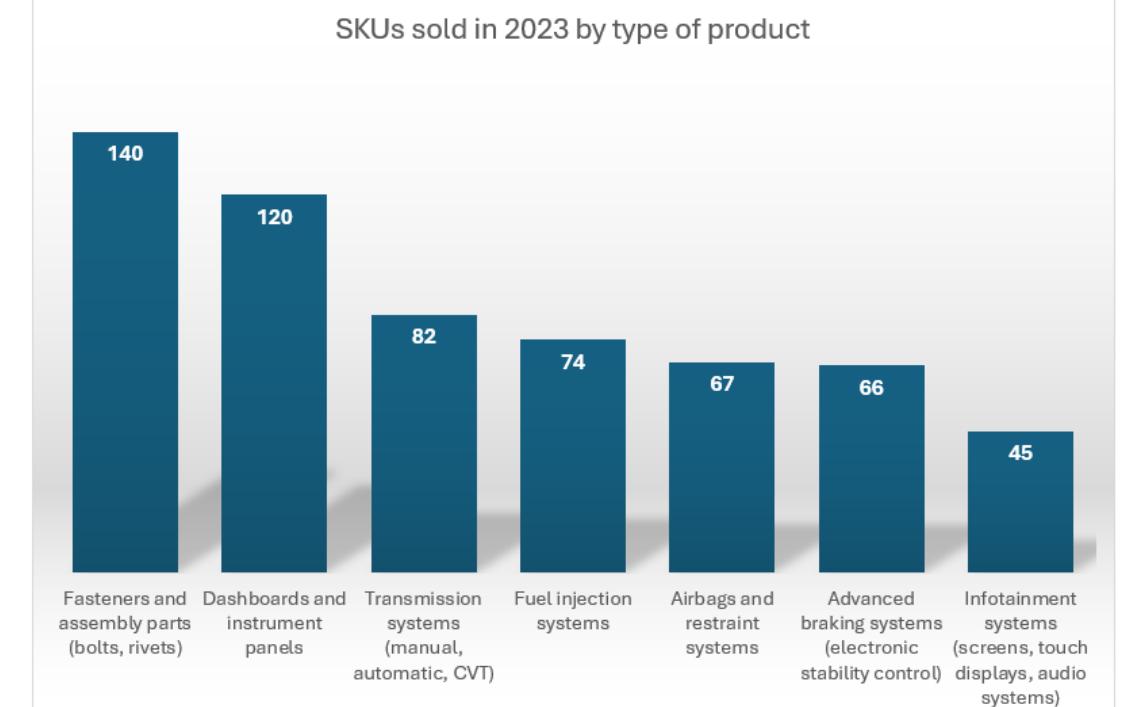
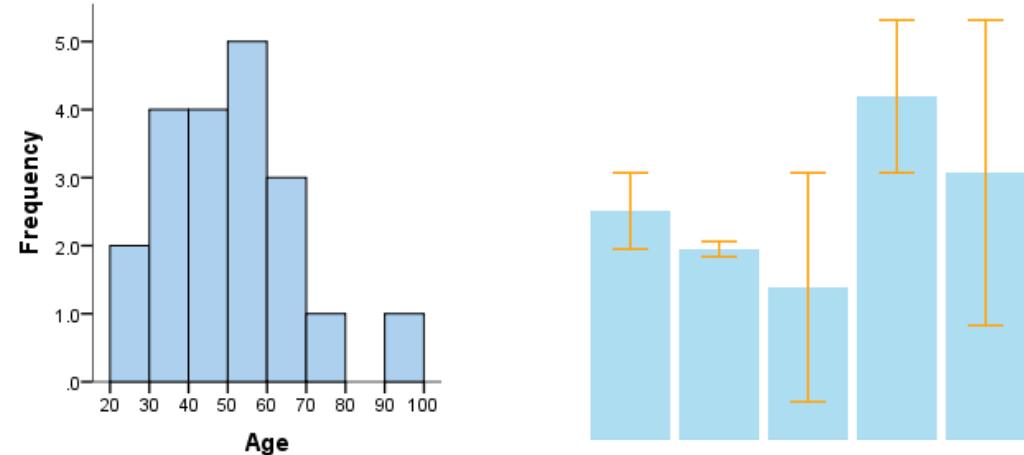
`coord_polar()`

Who sells more weapons?



Common mistakes: Barplots

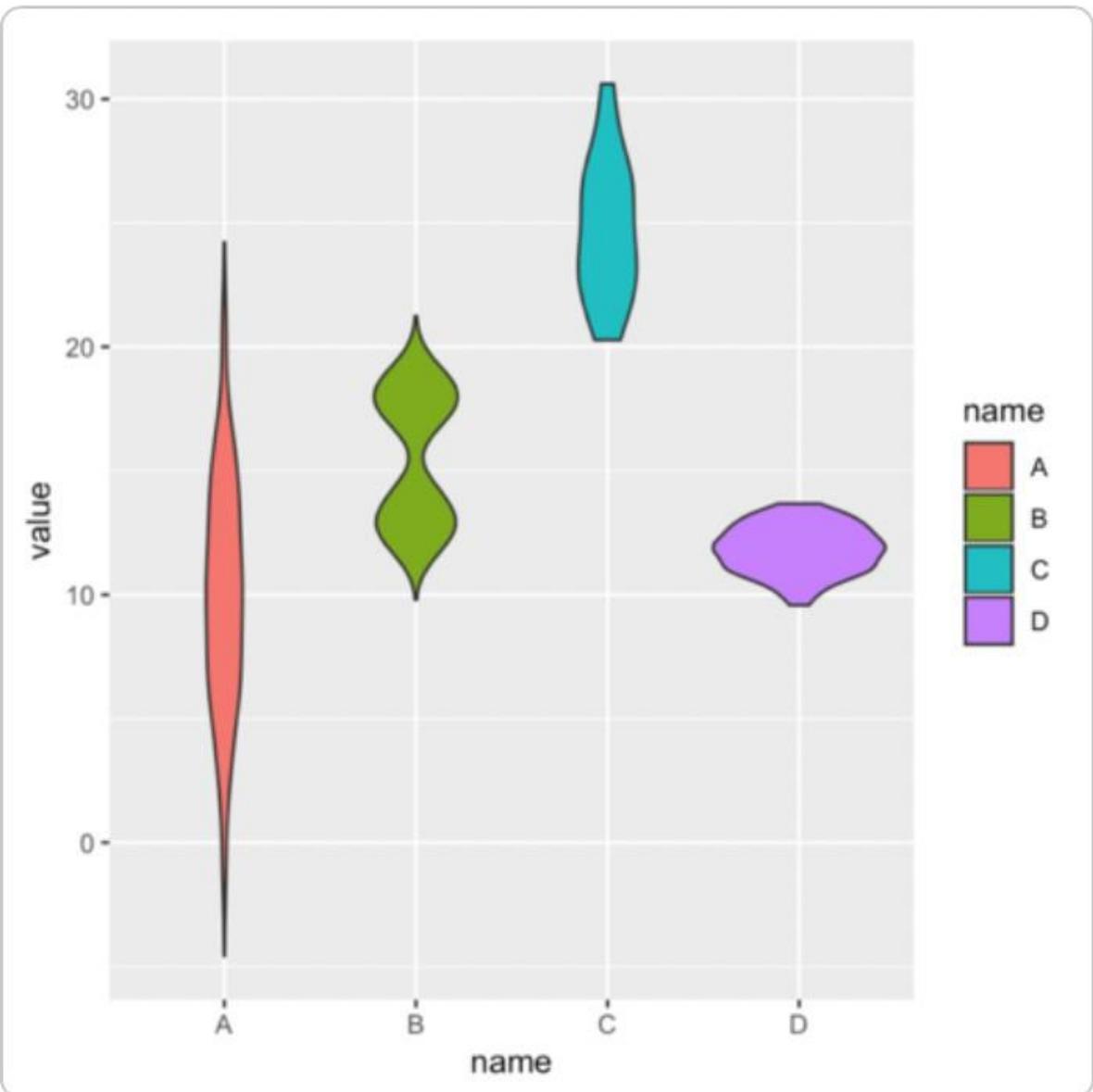
- Do not confound bar chart with histogram. A histogram has only a numeric variable as input and shows its distribution.
- Order your bars. If the levels of your categoric variable have no obvious order, order the bars following their values.
- Several values per group? Don't use a barplot.





Violin

Use it when the amount of data is vast and showing individual observations gets impossible.





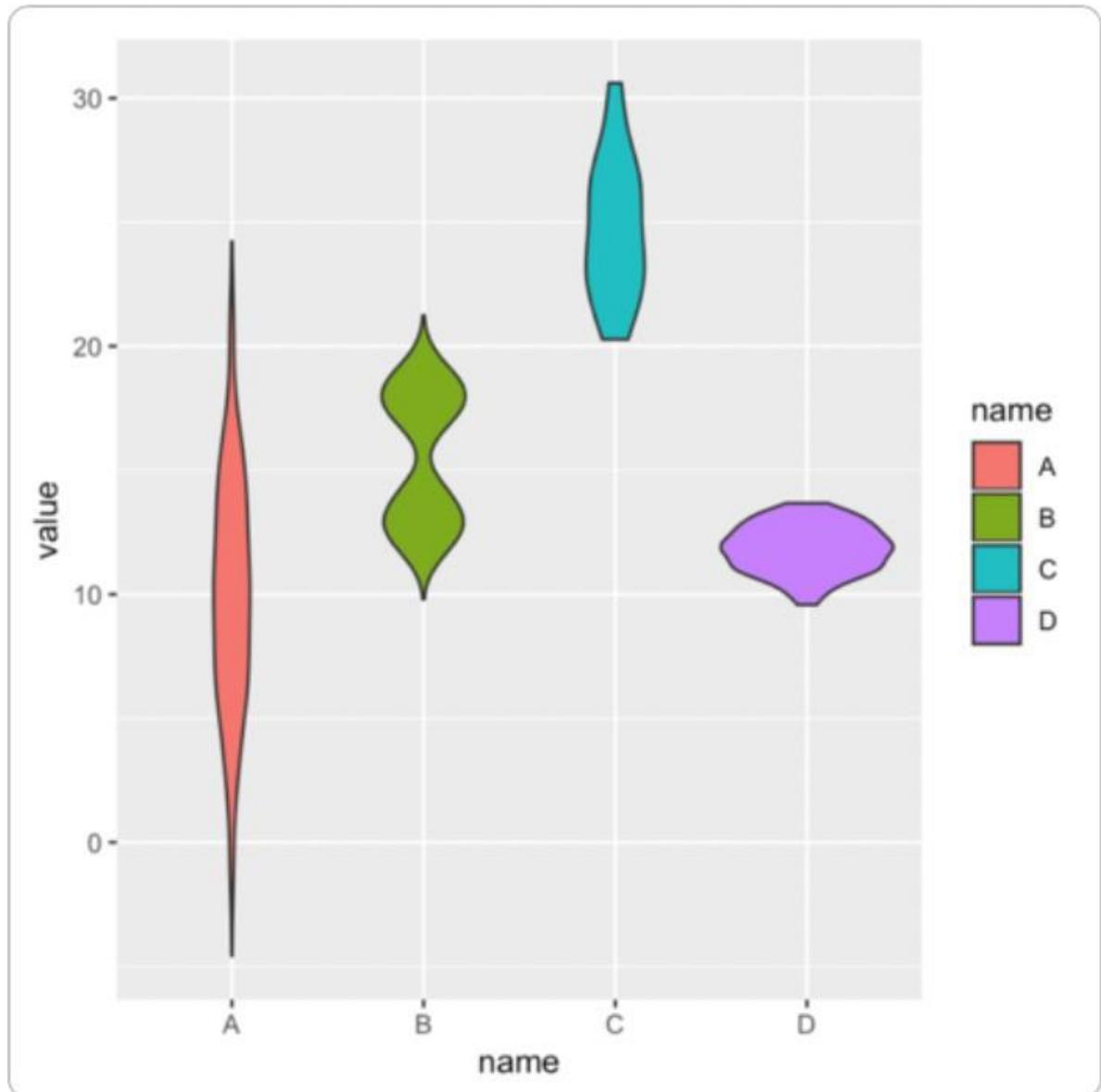
Violin

Allows to visualize the distribution of a numeric variable for one or several groups.

Each ‘violin’ represents a group or a variable.

The shape = the density estimate of the variable:
the more data points in a specific range, the
larger the violin is for that range

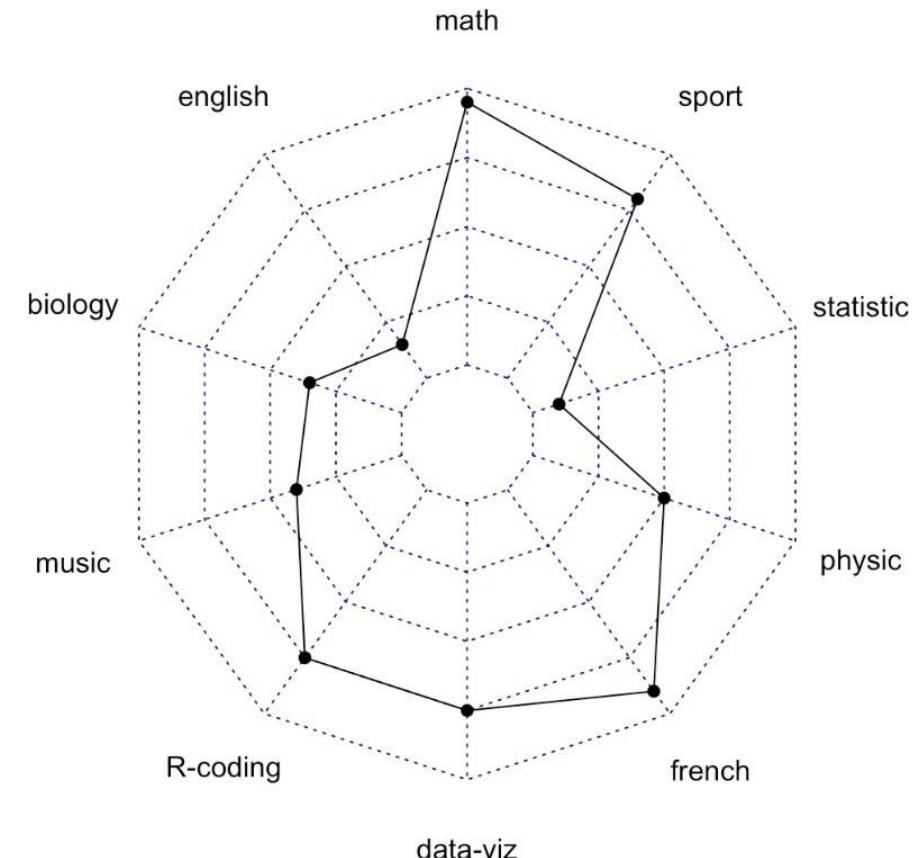
Violin charts can be produced with `ggplot2`
thanks to the `geom_violin()` function



Radar plot

A **radar** or **spider** or **web** chart is a two-dimensional chart type designed to plot one or more series of values over multiple quantitative variables.

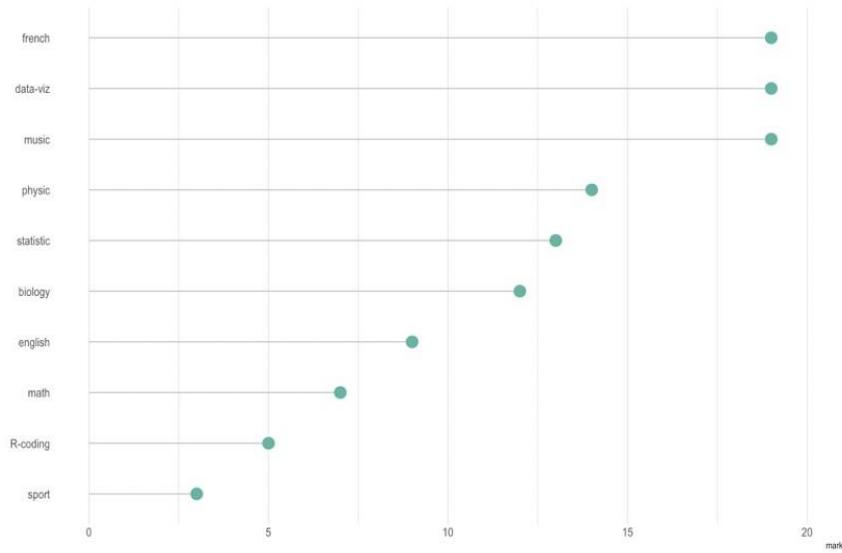
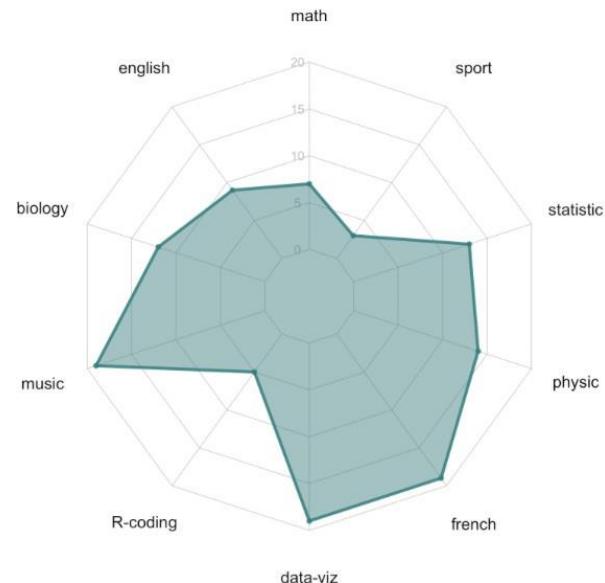
The package **fmsb** allows to build radar charts in R `radarchart(data)`



Common mistakes

Radar plots

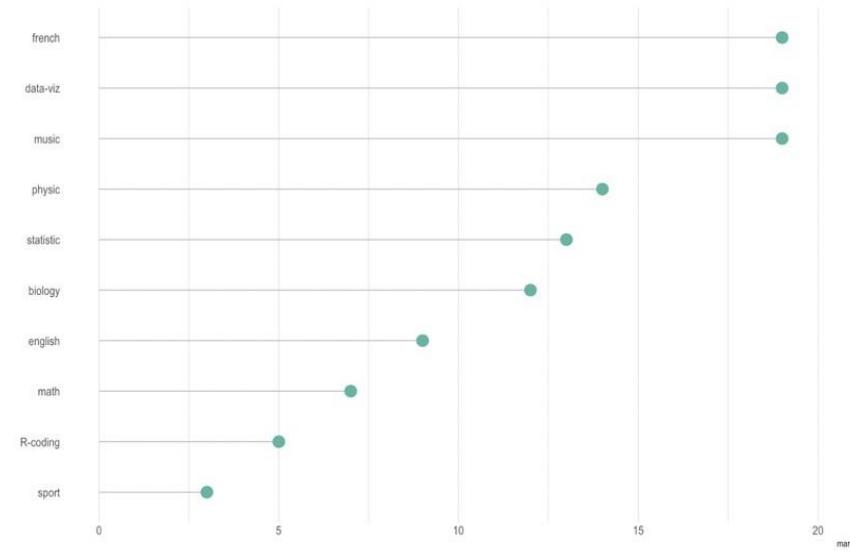
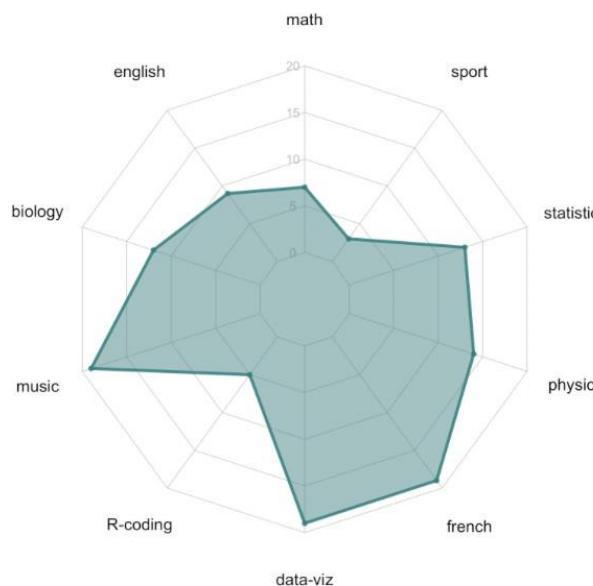
- *Circular layout = harder to read. Quantitative values are much easier to compare when they are laid out along a single vertical or horizontal axis.* This is a general weakness made of circular layouts.
- *Supporting the ranking* - The lollipop plot is ordered in the above example. It allows you to see instantly which topic had the best mark and what the ranking of each topic was.



Common mistakes

Radar plots

- *Category order has a huge impact* - Radar chart readers will probably focus on the shape observed. This can be misleading since this shape highly depends on the ordering of categories around the plot.



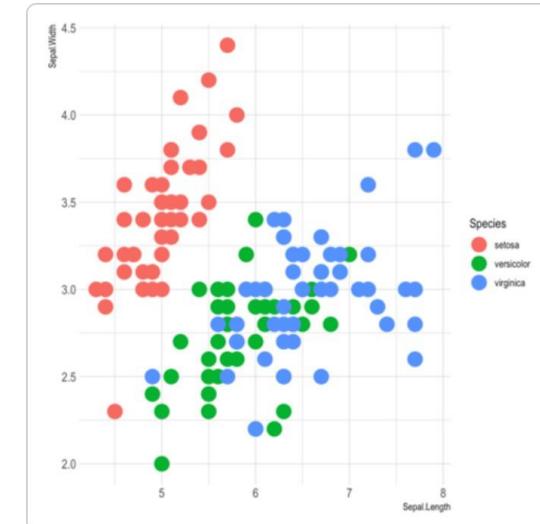
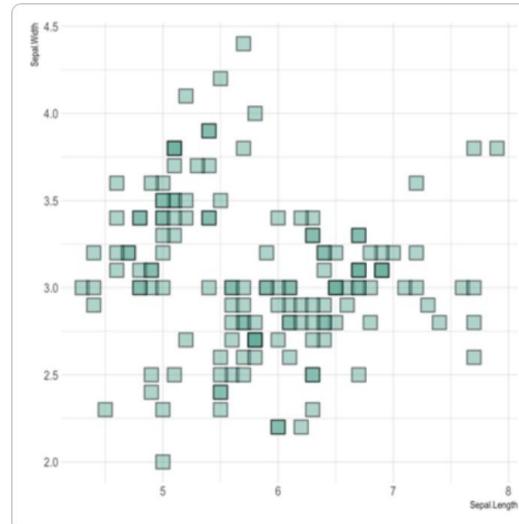
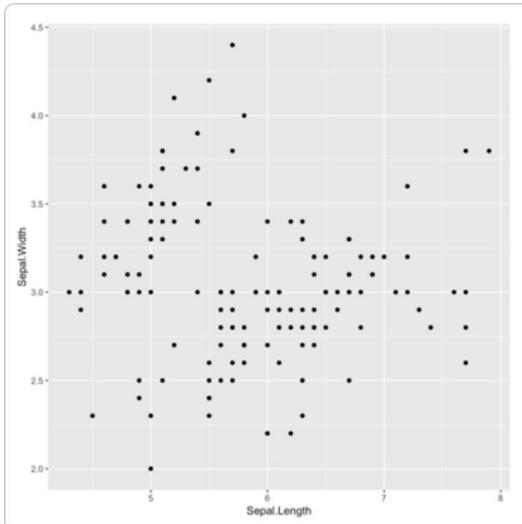


Scatter

Scatter

A **Scatterplot** displays the relationship between 2 numeric variables. Each dot represents an observation. Their position on the X (horizontal) and Y (vertical) axis represents the values of the 2 variables.

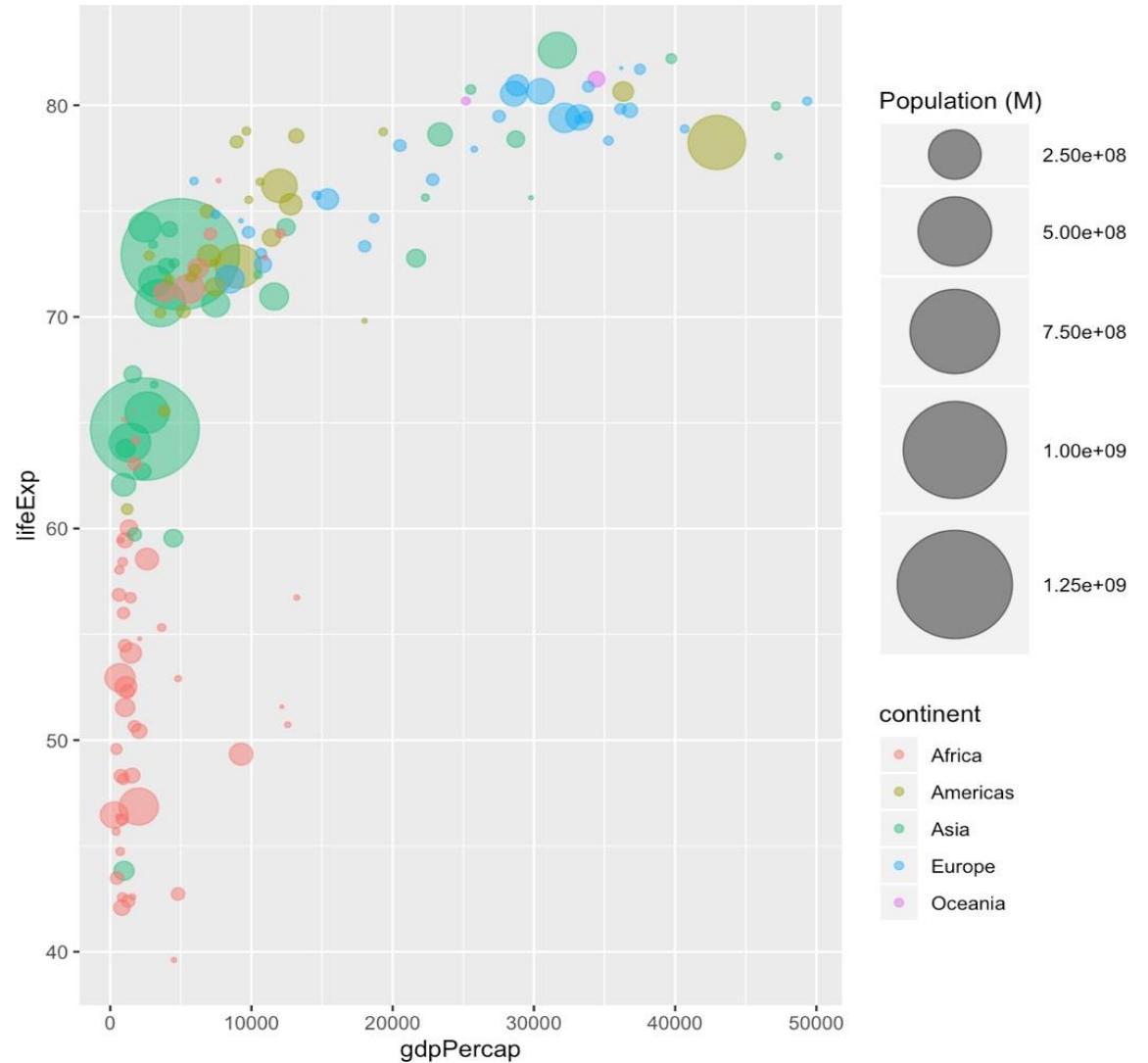
Using **ggplot2**, scatterplots are built thanks to the `geom_point` function.



Bubble Plot

A **bubble plot** is a **scatter plot** with a third numerical variable used for circle **size**.

Ggplot2: `geom_point()`

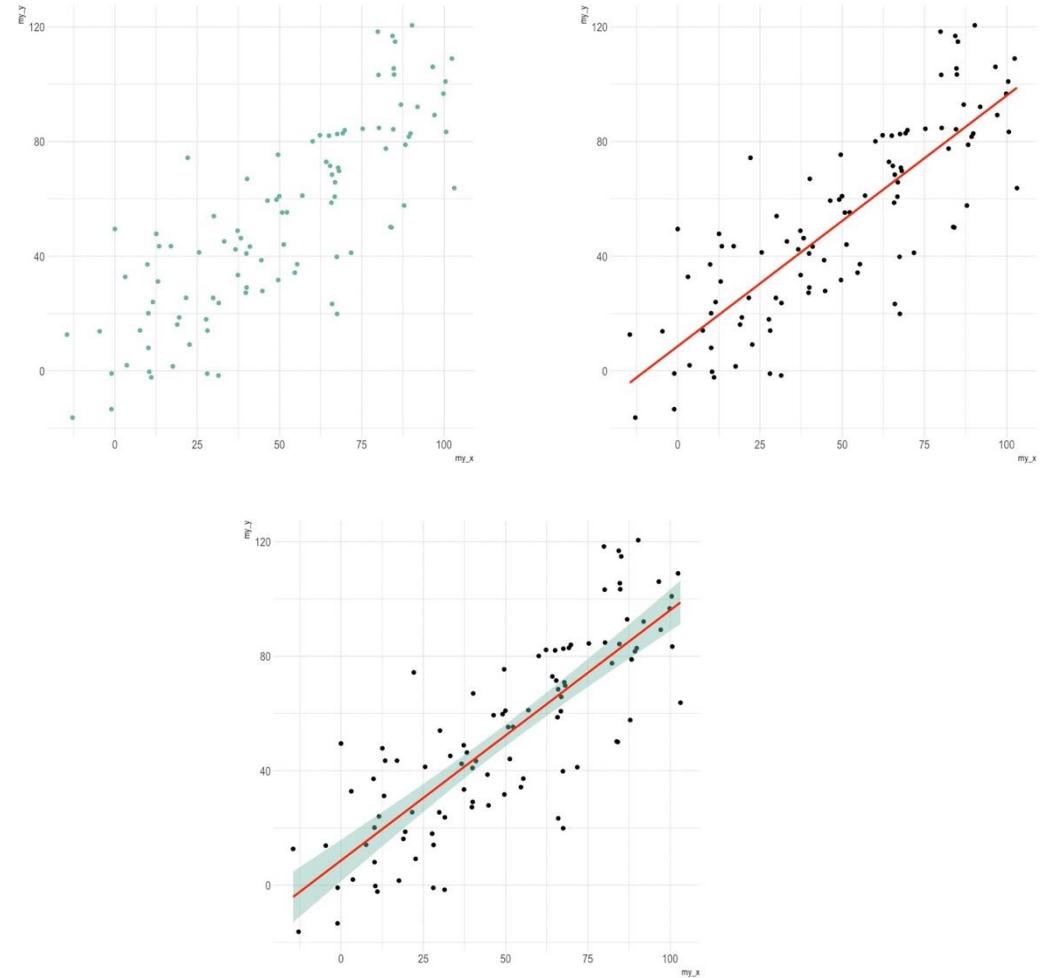


Linear Trend with Scatter plot

Adding a linear trend to a **scatterplot** helps the reader in seeing patterns.

`ggplot` provides `geom_smooth` function.
2
`h()`

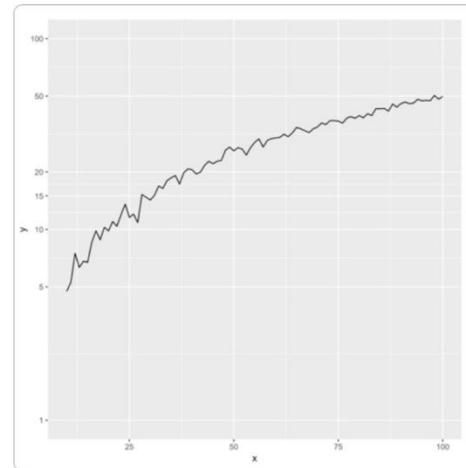
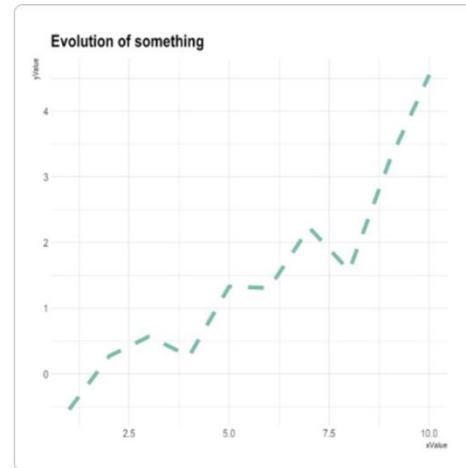
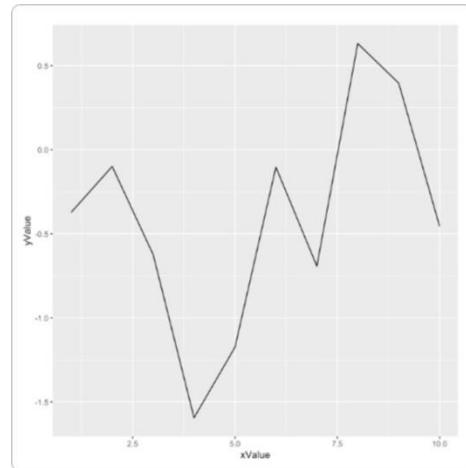
Allows to add the linear trend and the confidence interval around it if needed (option `se=TRUE`).





Line Chart

A **line chart** is often used to visualize a trend in data over intervals of time – a time series – thus the line is often drawn chronologically.



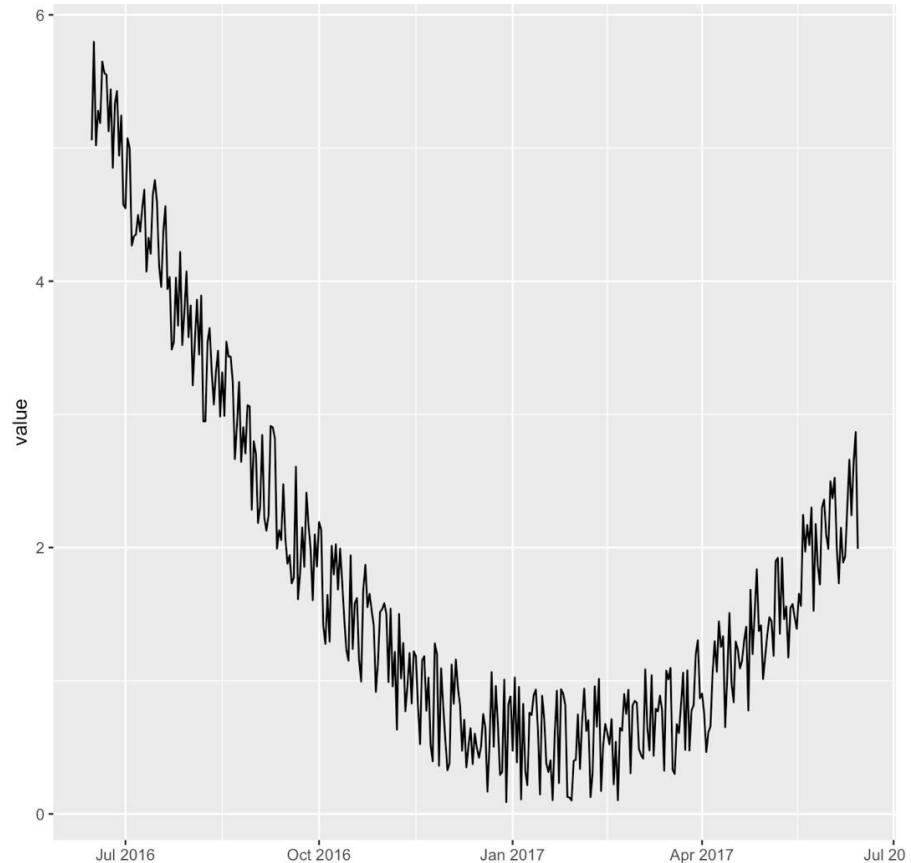
`ggplot` allows to draw line charts thanks to the `geom_line` function. It expects as input a data frame with **2** numeric variables, one displayed on each axis.

Time Series

The `ggplot2` package recognizes `date` format and automatically uses a specific type of X axis. If the time variable isn't at the `date` format, this won't work.

Always check with `str(data)` how variables are understood by R. If not read as a date, use `lubridate` to convert it.

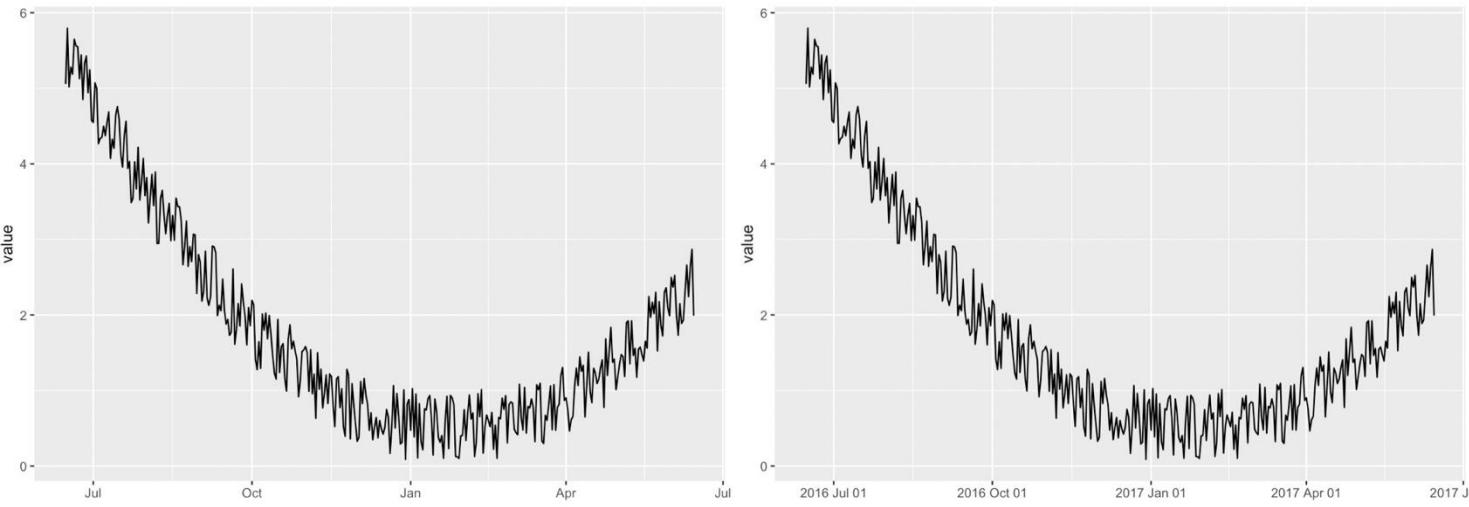
The `date` format will be recognized automatically, resulting in neat X-axis labels.



Time Series

As soon as the time variable is recognized as a **date**, you can use the **scale_x_date** function to choose the **format** displayed on the X axis.

```
p+scale_x_date(date_labels = "%b")  
p+scale_x_date(date_labels = "%Y %b %d")  
p+scale_x_date(date_labels = "%W")  
p+scale_x_date(date_labels = "%m-%Y")
```

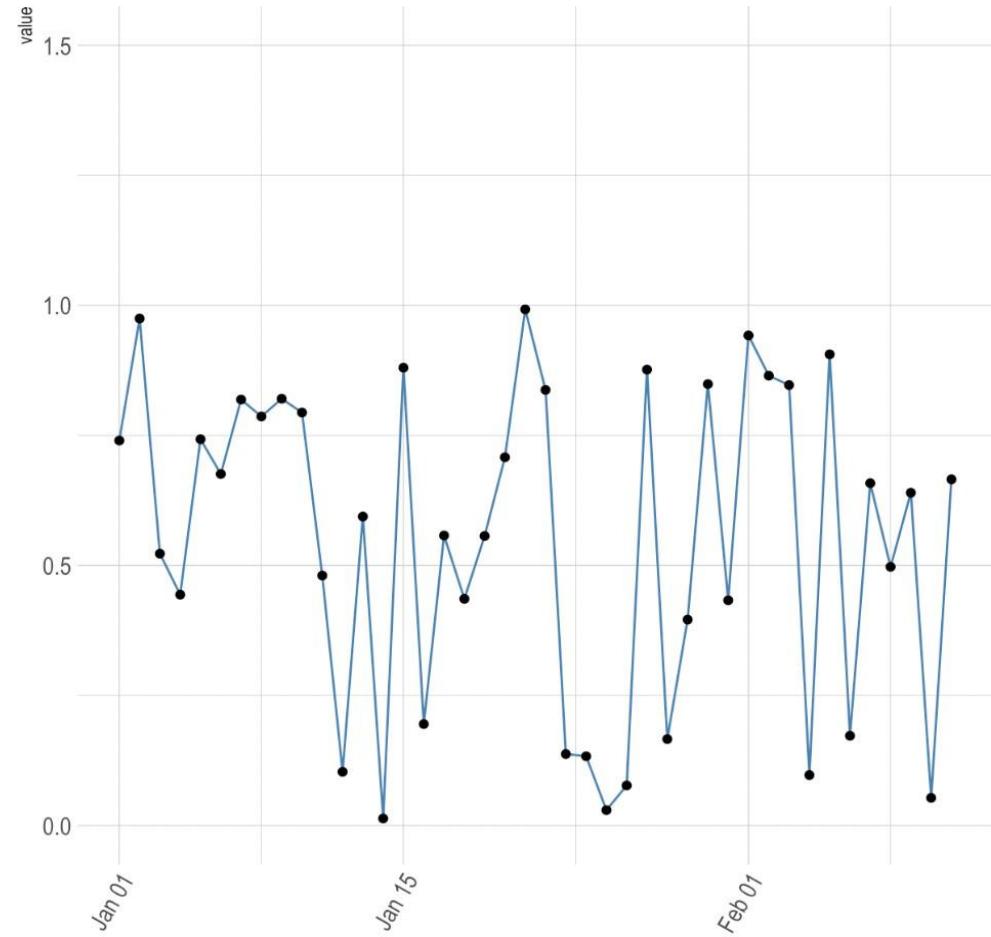


Symbol	Meaning	Example
%d	day as a number (0-31)	01-31
%a	abbreviated weekday	Mon
%A	unabbreviated weekday	Monday
%m	month (00-12)	00-12
%b	abbreviated month	Jan
%B	unabbreviated month	January
%y	2-digit year	07
%Y	4-digit year	2007

Select time frame

Use the `limit` option of the `scale_x_date()` function to select a time frame in the data:

```
p <- ggplot(data, aes(x=day, y=value)) +  
  geom_line( color="steelblue") +  
  geom_point() +  
  xlab("") +  
  theme_ipsum() +  
  theme(axis.text.x=element_text(angle=60, hjust=1)) +  
  scale_x_date(limit=c(as.Date("2017-01-01"),as.Date("2017-02-11"))) +  
  ylim(0,1.5)
```

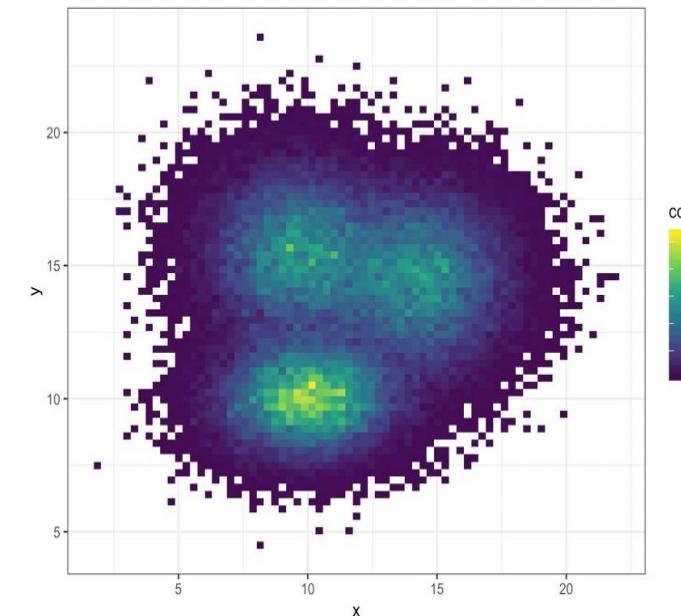
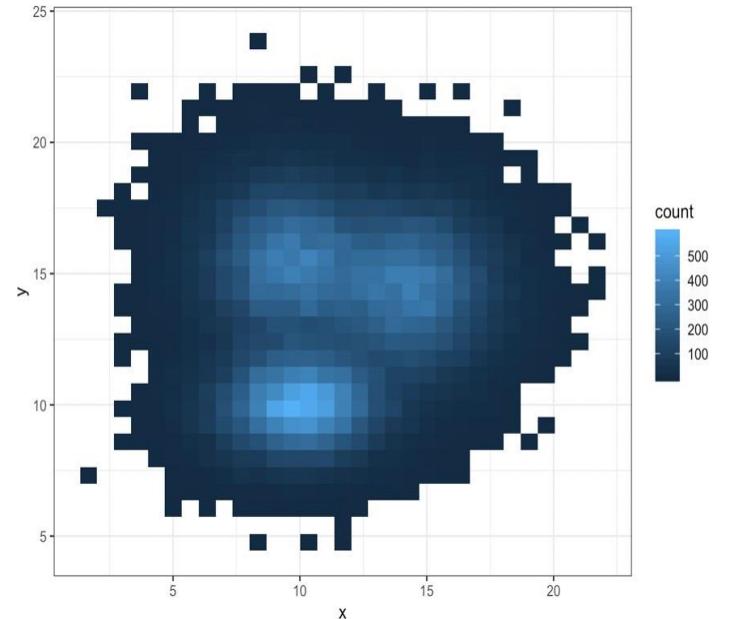


2d Histogram

The plot area is divided in a multitude of squares

Use the `geom_bin_2d()` function. This function offers a `bins` argument that controls the number of bins you want to display.

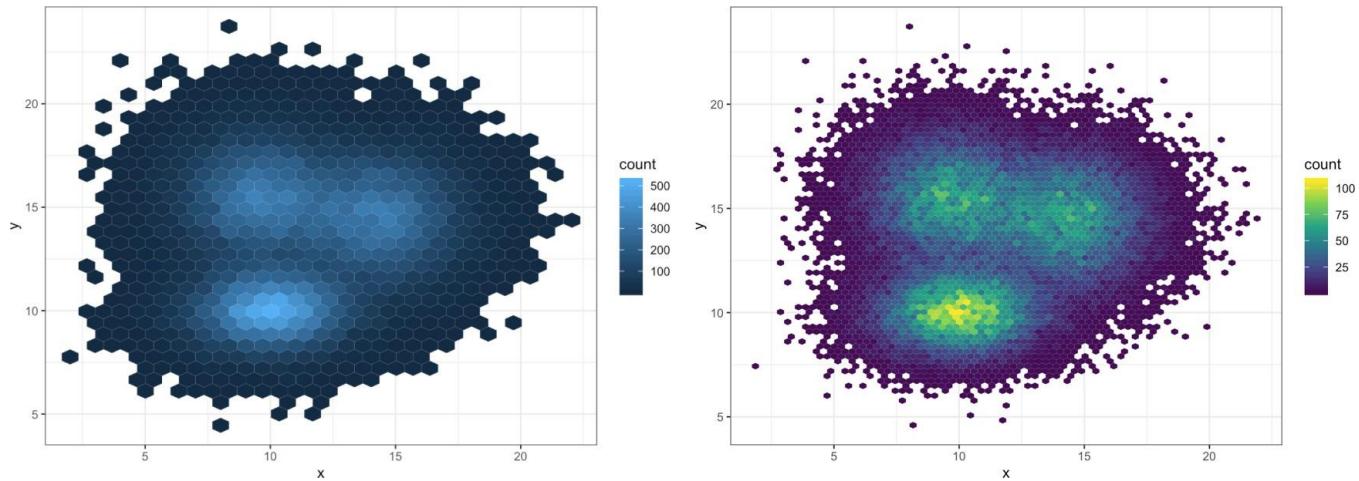
```
# Bin size control + color palette
ggplot(data, aes(x=x, y=y) ) +
  geom_bin2d(bins = 70) +
  scale_fill_continuous(type = "viridis") +
  theme_bw()
```



Hexbin chart

Another alternative is to divide the plot area in a multitude of hexagons: it is thus called a hexbin made using `geom_hex()` function.

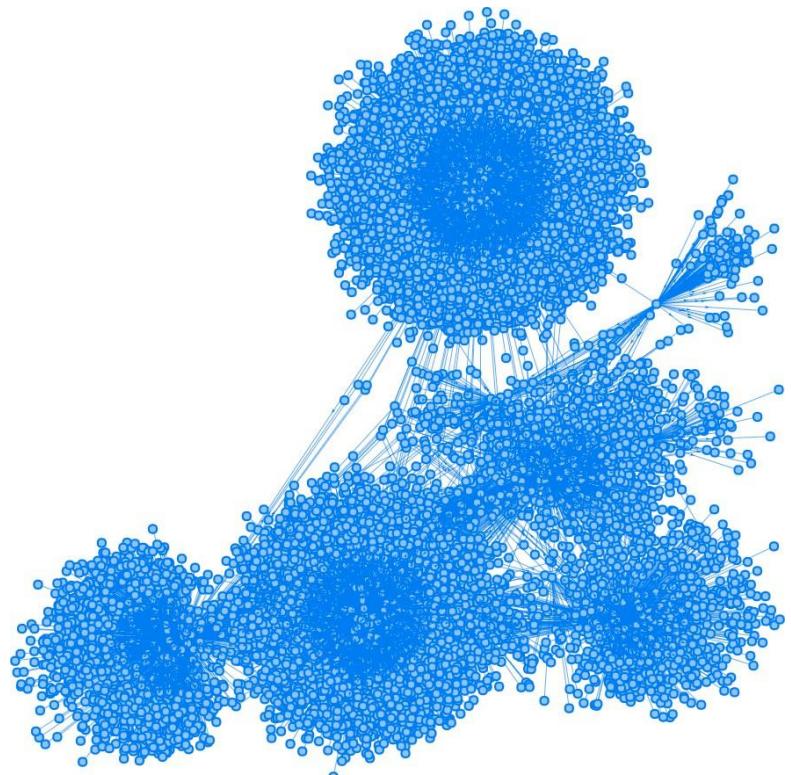
Provides `bins` argument as well, to control the number of divisions per axis.



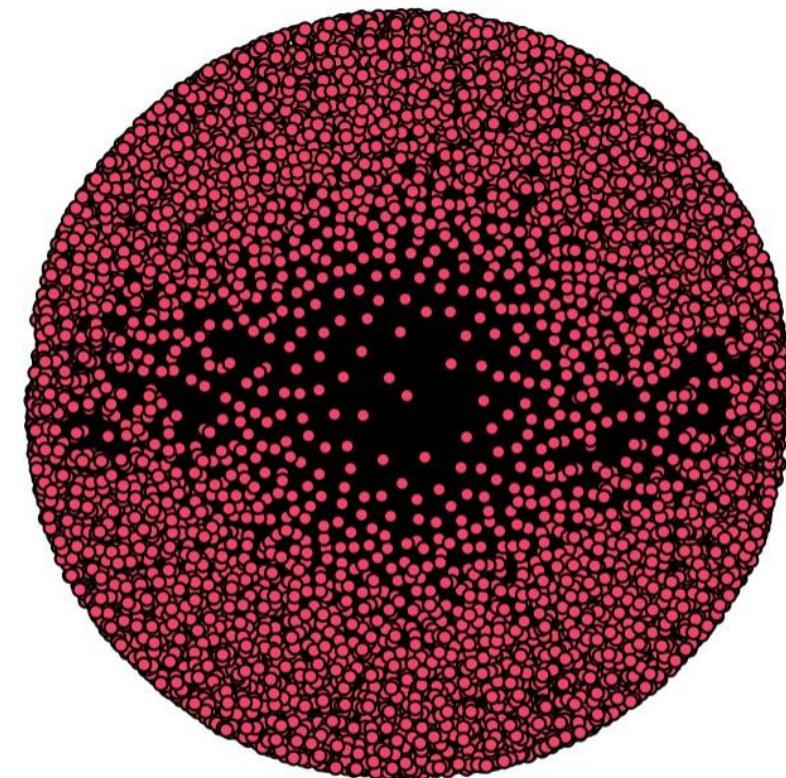
```
# Bin size control + color palette
ggplot(data, aes(x=x, y=y) ) +
  geom_hex(bins = 70) +
  scale_fill_continuous(type = "viridis") +
  theme_bw()
```

Social Network - Entities are represented as nodes, and the relationships between entities are represented as lines that are called links.

library(visNetwork)



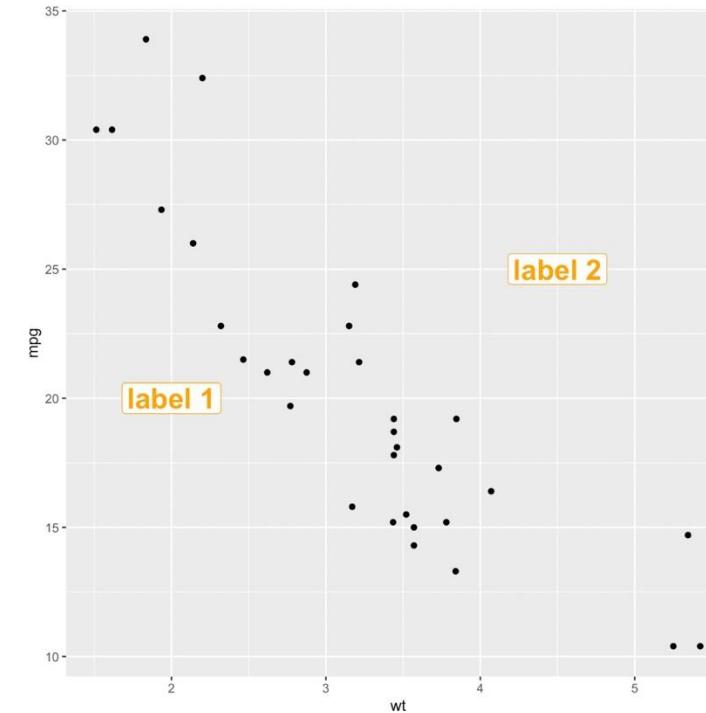
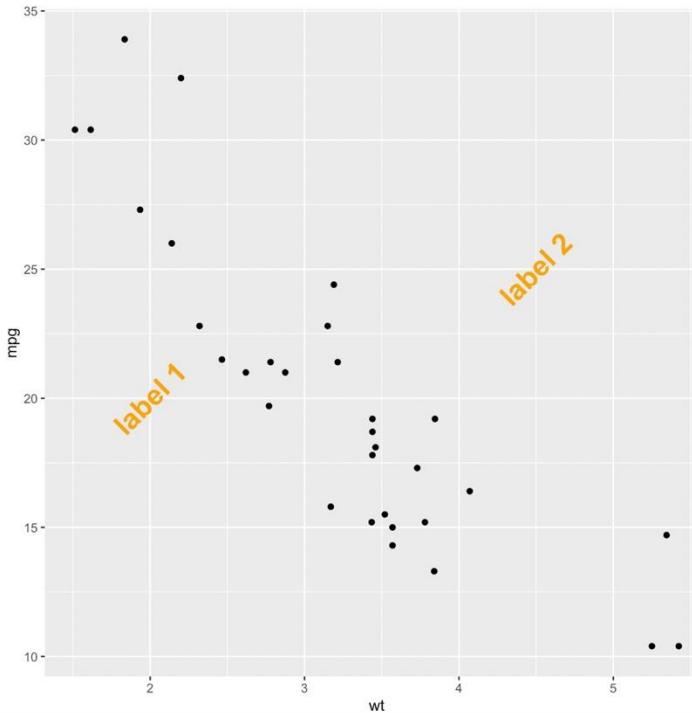
library(Network)



Annotation with `geom_text()` or `geom_label()`

Text is the most common kind of annotation. It allows to give more information on the most important part of the chart.

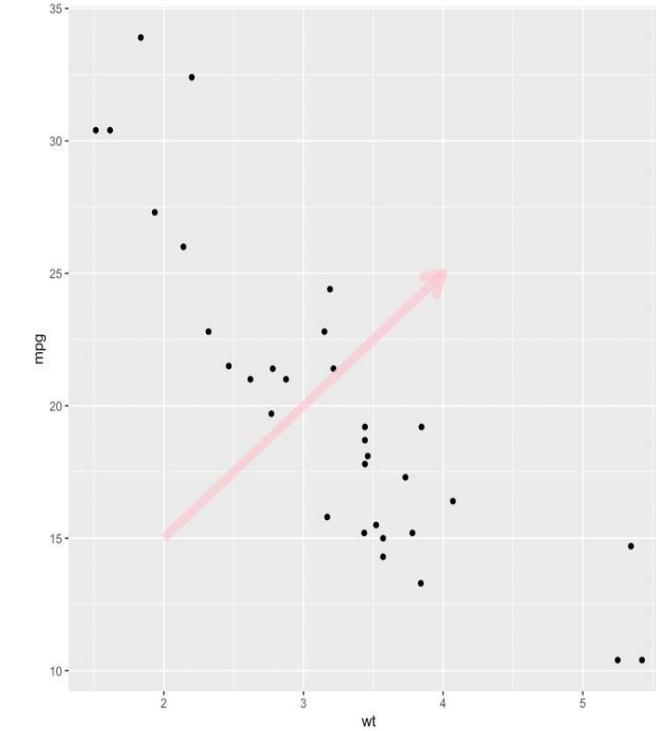
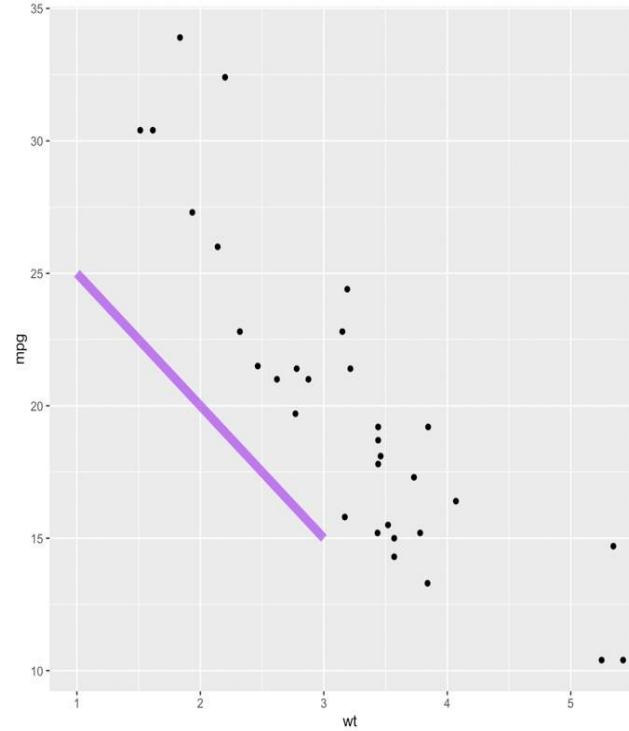
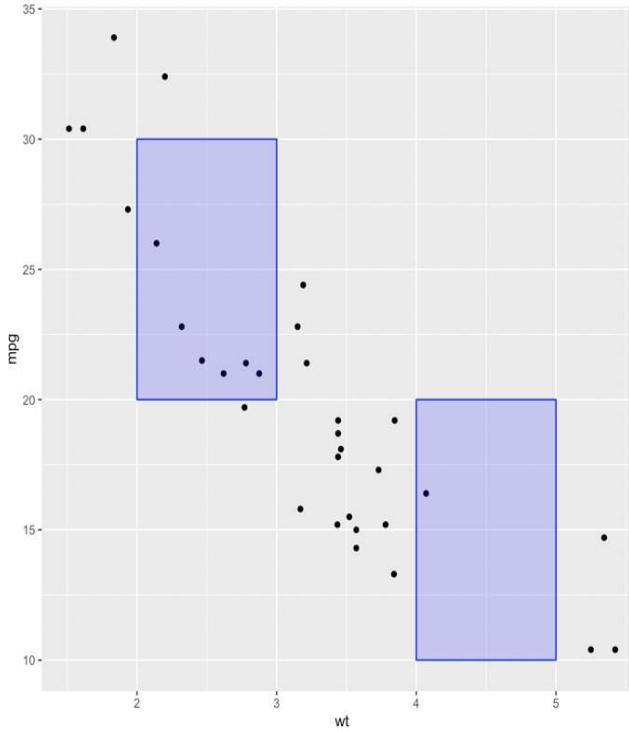
Using `ggplot2`, 2 main functions are available for that kind of annotation:



`geom_text` to add a simple piece of text
`geom_label` to add a label: framed text
el

Adding text with `geom_text()` or `geom_label()`

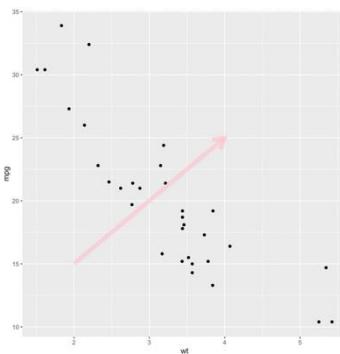
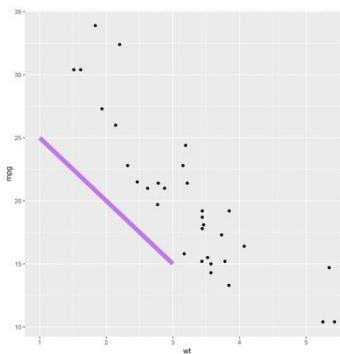
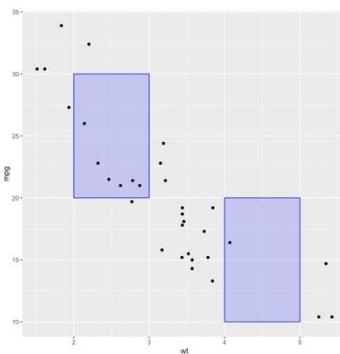
Add shapes with `annotate()`



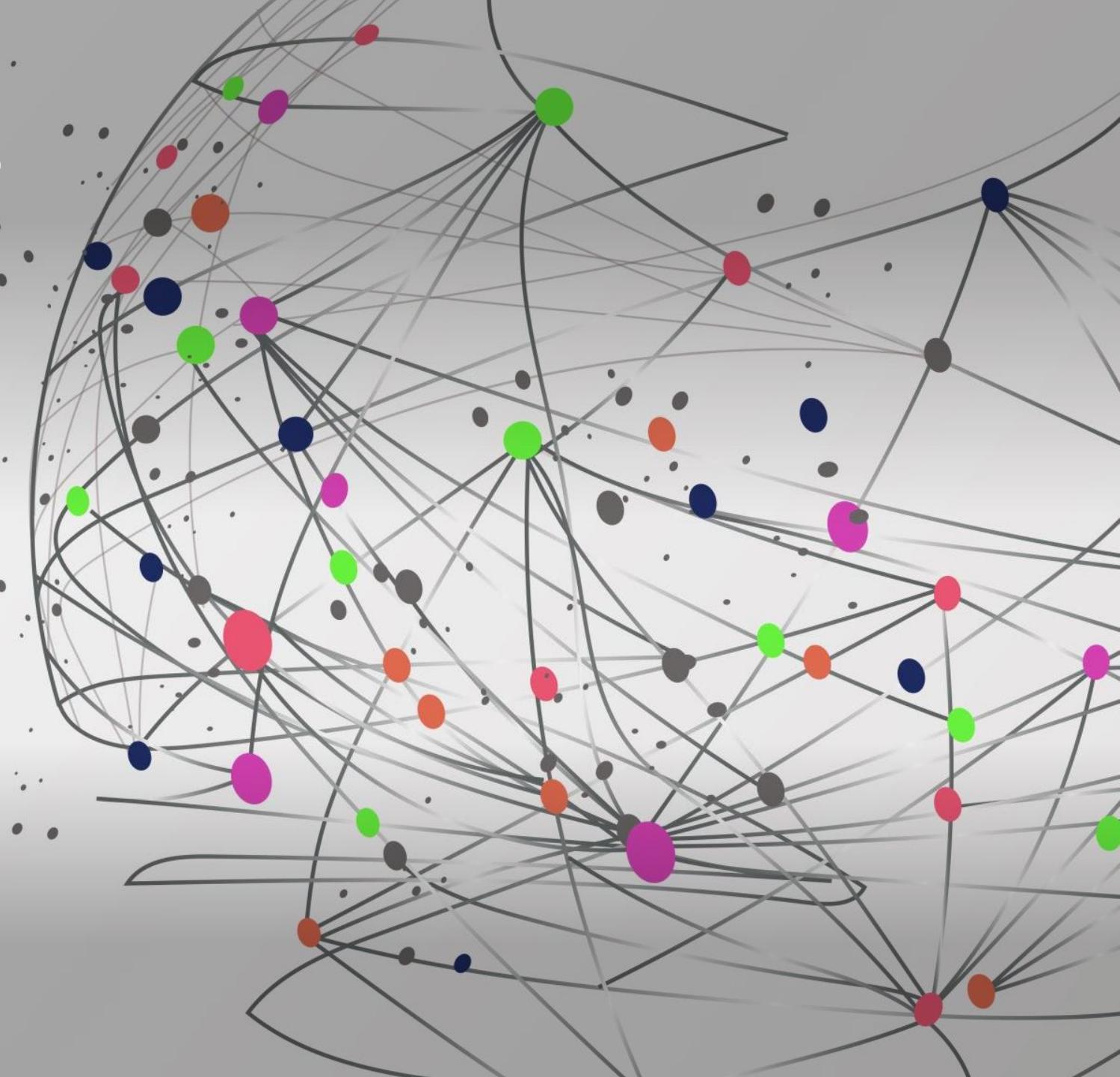
Adding text with `geom_text()` or `geom_label()`

Add shapes with `annotate()`

```
# Add rectangles  
p + annotate("rect", xmin=c(2,4), xmax=c(3,5), ymin=c(20,10) , ymax=c(30,20), alpha=0.2, color="blue", fill="blue")  
  
# Add segments  
p + annotate("segment", x = 1, xend = 3, y = 25, yend = 15, colour = "purple", size=3, alpha=0.6)  
  
# Add arrow  
p + annotate("segment", x = 2, xend = 4, y = 15, yend = 25, colour = "pink", size=3, alpha=0.6, arrow=arrow())
```



dataviz resources



Data Visualization Tools

Our world data mission: “research and data to make progress against the world’s largest problems”.



Environmental Impacts of Food Production
by Our World in Data

We just published a redesign of our work on the Environmental Impacts of Food

We've just published a major redesign of our work on the Environmental Impacts of Food. Explore all of our data and research in one place.

By Hannah Ritchie — December 02, 2022



CO₂ and Greenhouse Gases
by Our World in Data

Data update: We've just updated all of our global CO₂ emissions data

We've just updated CO₂ emissions data for all countries in the world, including new metrics on emissions from land use change.

By Pablo Rosado, Hannah Ritchie and Edouard Mathieu — November 11, 2022

<https://ourworldindata.org/>

Data Visualization Tools

Names that are more dog or more human



<https://flowingdata.com/>

Flowing Data explores how statisticians, designers, data scientists, and others use analysis, visualization, and exploration to understand data.

Color Tools

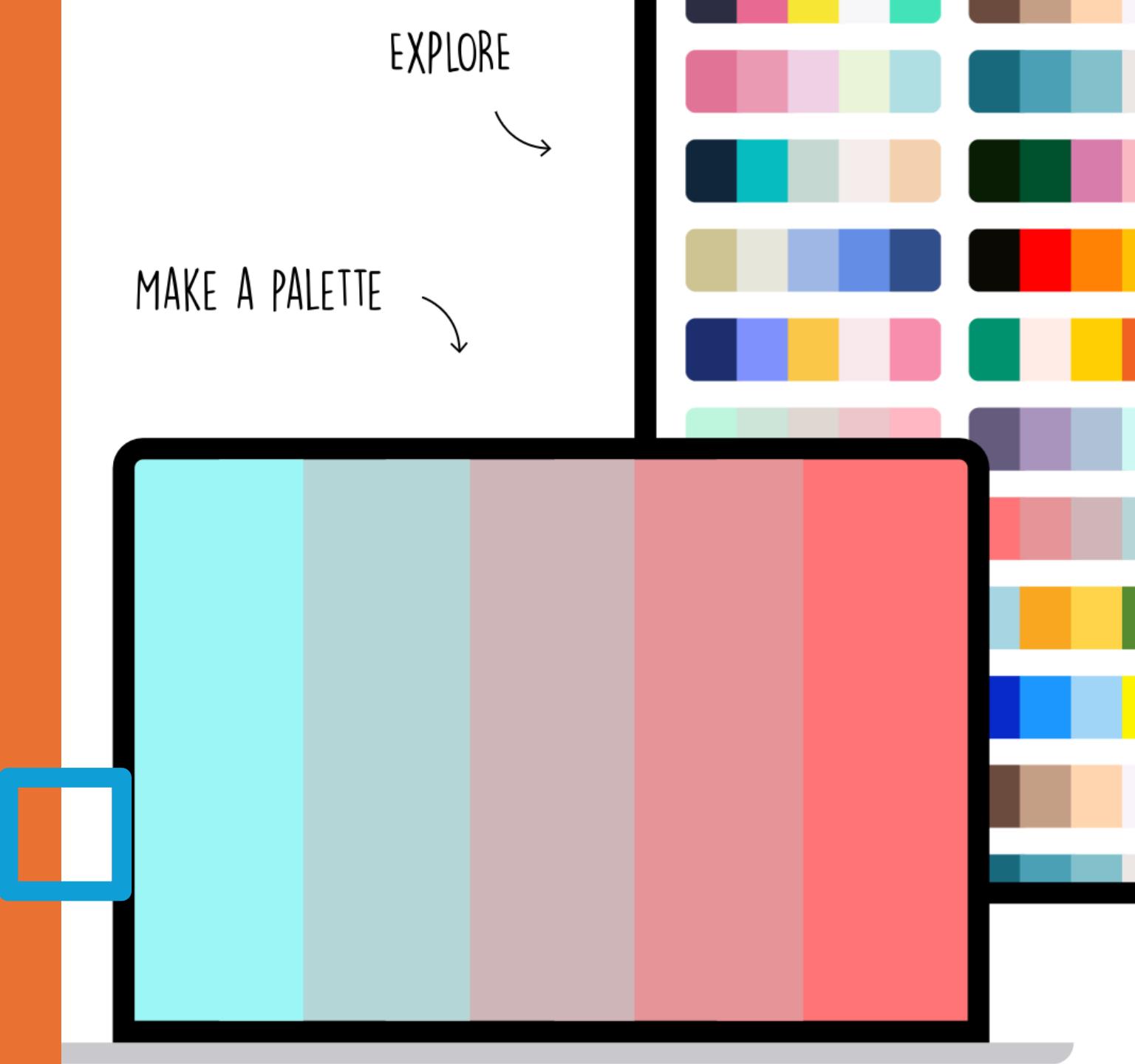
Generate cool pallets, get
inspiration!



<https://coolors.co/>
color.adobe.com

MAKE A PALETTE

EXPLORE





Copyrighted Material

cole nussbaumer knaflic

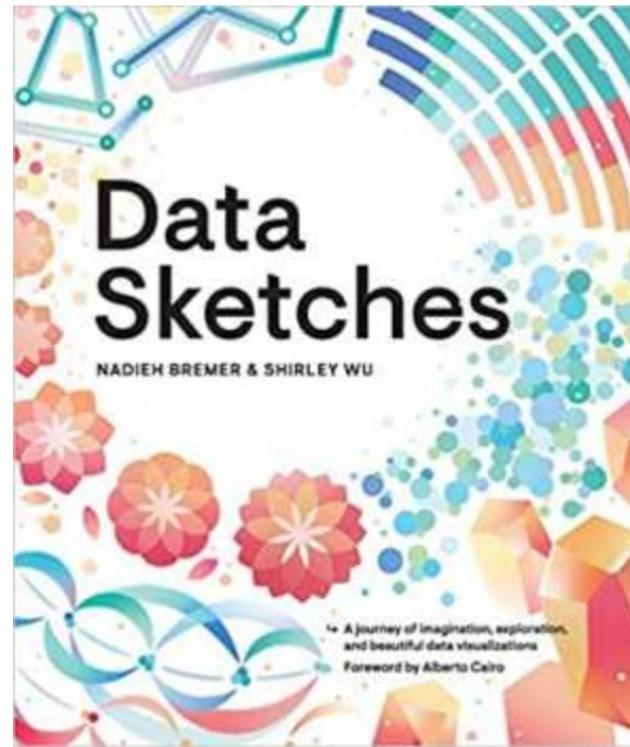
storytelling with **data**

a data
visualization
guide for
business
professionals

91%

Copyrighted Material

WILEY



Data Sketches

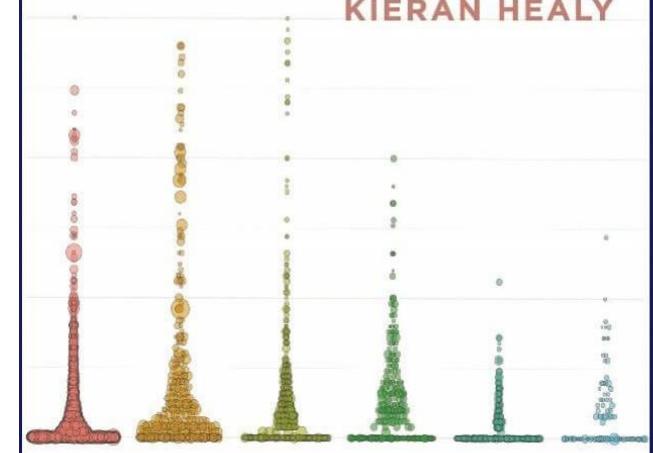
NADIEH BREMER & SHIRLEY WU

A journey of imagination, exploration,
and beautiful data visualizations
Foreword by Alberto Cairo

DATA VISUALIZATION

A PRACTICAL INTRODUCTION

KIERAN HEALY



The background of the slide features a repeating pattern of vertical bars in various shades of teal and green, creating a digital or data-oriented aesthetic.

What is used in industry?

The Best Data Visualization Software of 2024

- **Microsoft Power BI:** Best for business intelligence (BI)
- **Tableau:** Best for interactive charts
- **Qlik Sense:** Best for artificial intelligence (AI)
- **Klipfolio:** Best for custom dashboards
- **Looker:** Best for visualization options
- **Zoho Analytics:** Best for Zoho users
- **Domo:** Best for custom apps

Microsoft Power BI

BEST FOR BUSINESS INTELLIGENCE

Microsoft Power BI

4.5  Forbes Advisor

Software Integrations	Real-Time Analytics	Trend Analysis
Yes	Yes	Yes

Why We Picked It

Microsoft Power BI is a business intelligence platform that enables users to collaborate with data and track goals. It provides real-time analytics and trend analysis to help users make confident decisions. Microsoft Power BI integrates with many Microsoft products and cloud services, making it a versatile solution for businesses.

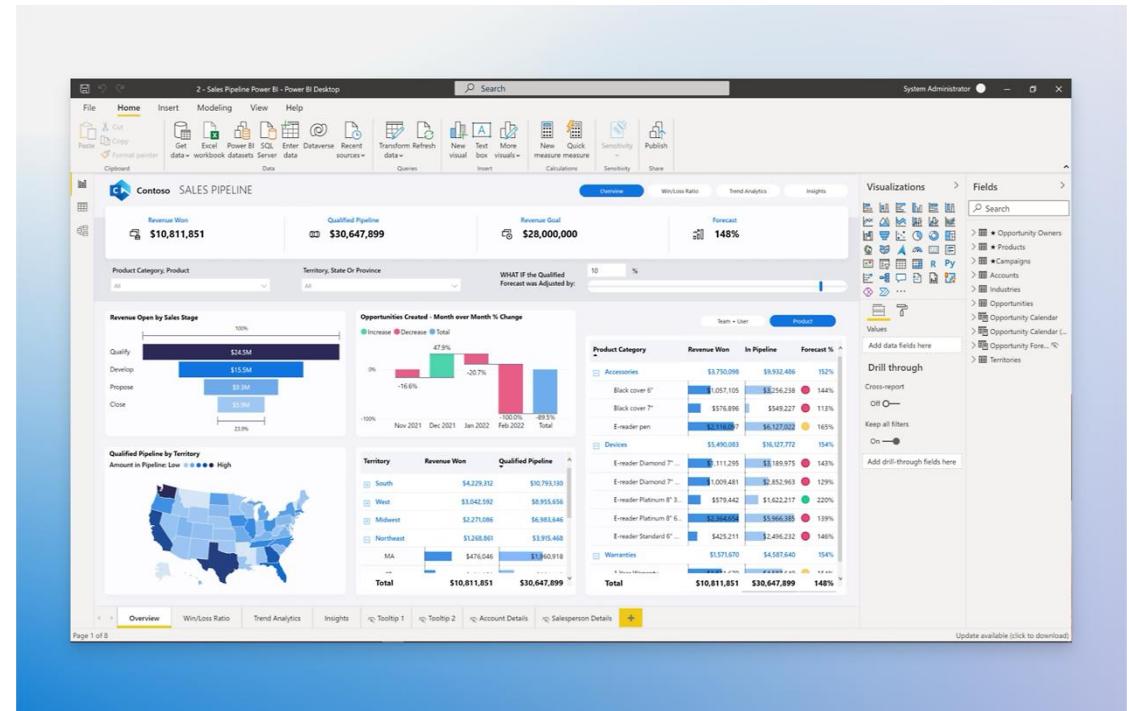
It offers data security features to protect user data. Microsoft Power BI is a powerful platform for building engaging, immersive dashboards and reports that allow you to take action.

Who should use it:

Businesses that use Microsoft products and services and need a business intelligence solution.

Pros & Cons

Pros	Cons
Has a mobile app	Heavy CPU usage
Competitive pricing	Desktop version not compatible with Mac
If you know how to use Excel, you'll have an easy time learning this	



Tableau

BEST FOR INTERACTIVE CHARTS

Tableau

4.4  Forbes Advisor

Software Integrations	Real-Time Analytics	Trend Analysis
Yes	Yes	Yes

Why We Picked It

Tableau is a data visualization and analytics platform that enables users to explore data and share insights. Data exploration is simple for anyone thanks to Tableau's flexibility and power. Users can build visualizations with drag and drop, employ AI-driven statistical modeling with a few clicks and ask questions using natural language.

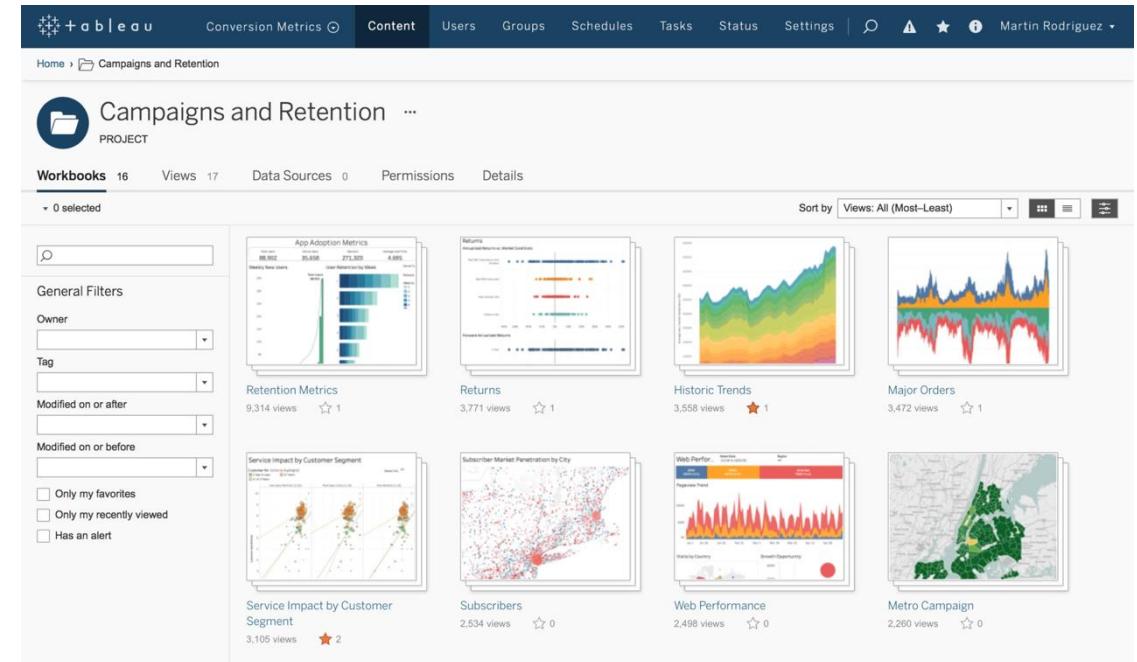
Tableau saves time and money by allowing for greater scale in the areas of governance, security, compliance, maintenance and support. Users also receive more than just an analytics tool—they get the knowledge and support they need to achieve greater scale.

Who should use it:

Businesses that need data visualization and analytics capabilities to scale their data operations.

Pros & Cons

Pros	Cons
Storytelling dashboards	Steep learning curve
It's backed by Salesforce	Prices are higher compared to similar products
Integrates with Slack	



Qlik Q

BEST FOR ARTIFICIAL INTELLIGENCE

Qlik Sense

4.4 Forbes ADVISOR

Software Integrations	Real-Time Analytics	Trend Analysis
Yes	Yes	Yes

Why We Picked It

Qlik Sense is a data visualization tool that uses artificial intelligence (AI) to help users understand and use data more effectively. It offers deeper interactivity and broader context than other data visualization tools, as well as lightning-fast calculations and the ability to connect and combine data from hundreds of data sources.

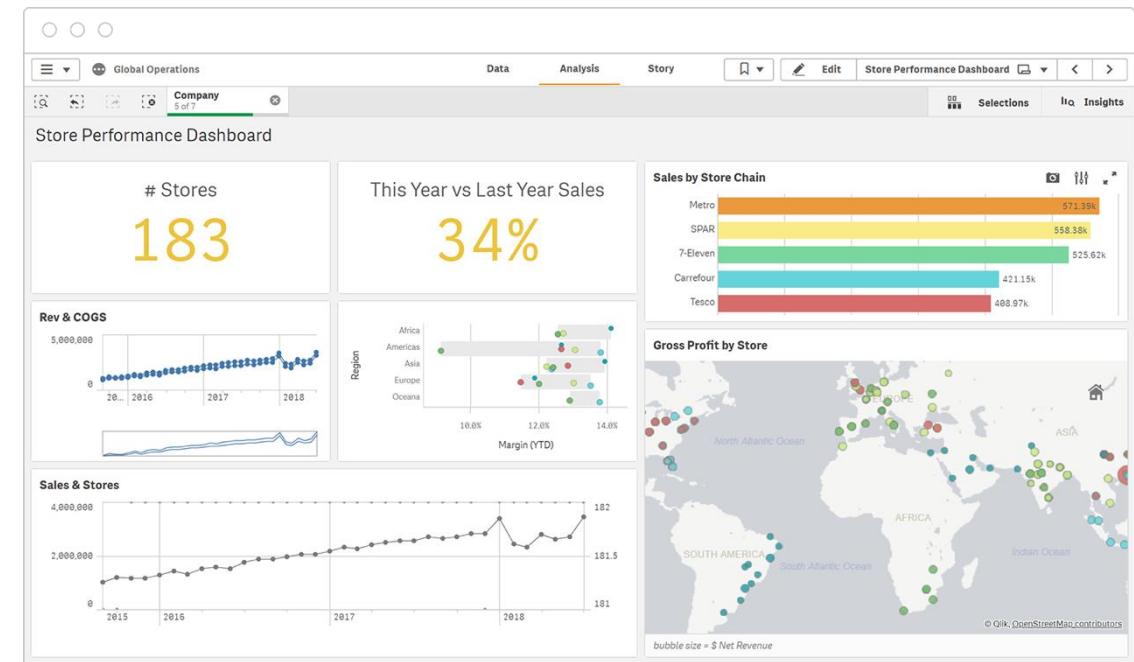
Qlik Sense is a part of the Qlik Active Intelligence Platform, which offers analytics performance and scalability to businesses of all sizes. Additionally, it's available as a software-as-a-service (SaaS) solution or as a hybrid service that extends SaaS analytics to on-premises data.

Who should use it:

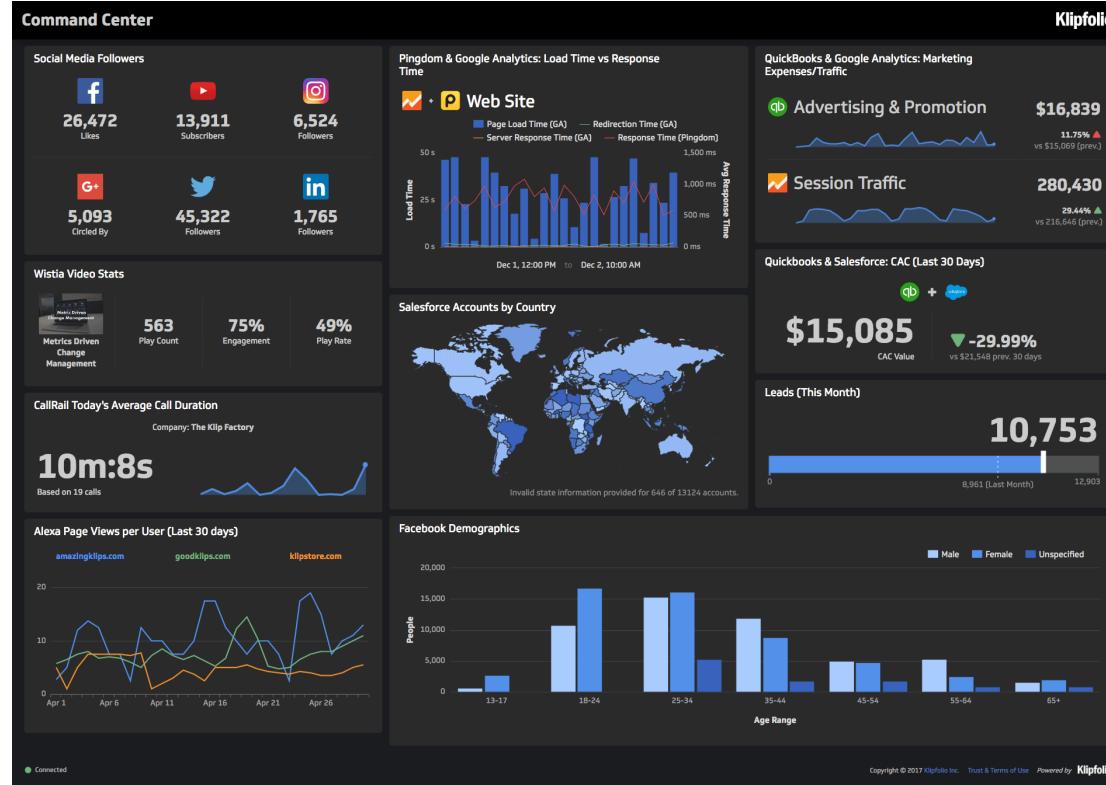
Businesses that need artificial intelligence capabilities to scale their data operations.

Pros & Cons

Pros	Cons
Automated action triggers	User interface needs work
AI insights	
Drag-and-drop interface	



Klipfolio



BEST FOR CUSTOM DASHBOARDS

Klipfolio

4.3 **Forbes ADVISOR**

Software Integrations	Real-Time Analytics	Trend Analysis
Yes	Yes	Yes

Why We Picked It

With Klipfolio, access and combine data from hundreds of services without writing any code using curated instant metrics, all of which are pre-built. With its powerful data modeler, you can leverage data in everyday decision-making. Users can import, edit and analyze data to get comprehensive and exact insight.

Its flexible payment plans require no lengthy contracts or onboarding fees. Nonprofit organizations receive a discount on plans but will need to contact Klipfolio's sales team for details.

Who should use it:

Klipfolio's unlimited dashboards with every plan make it an ideal solution for businesses requiring custom dashboards.

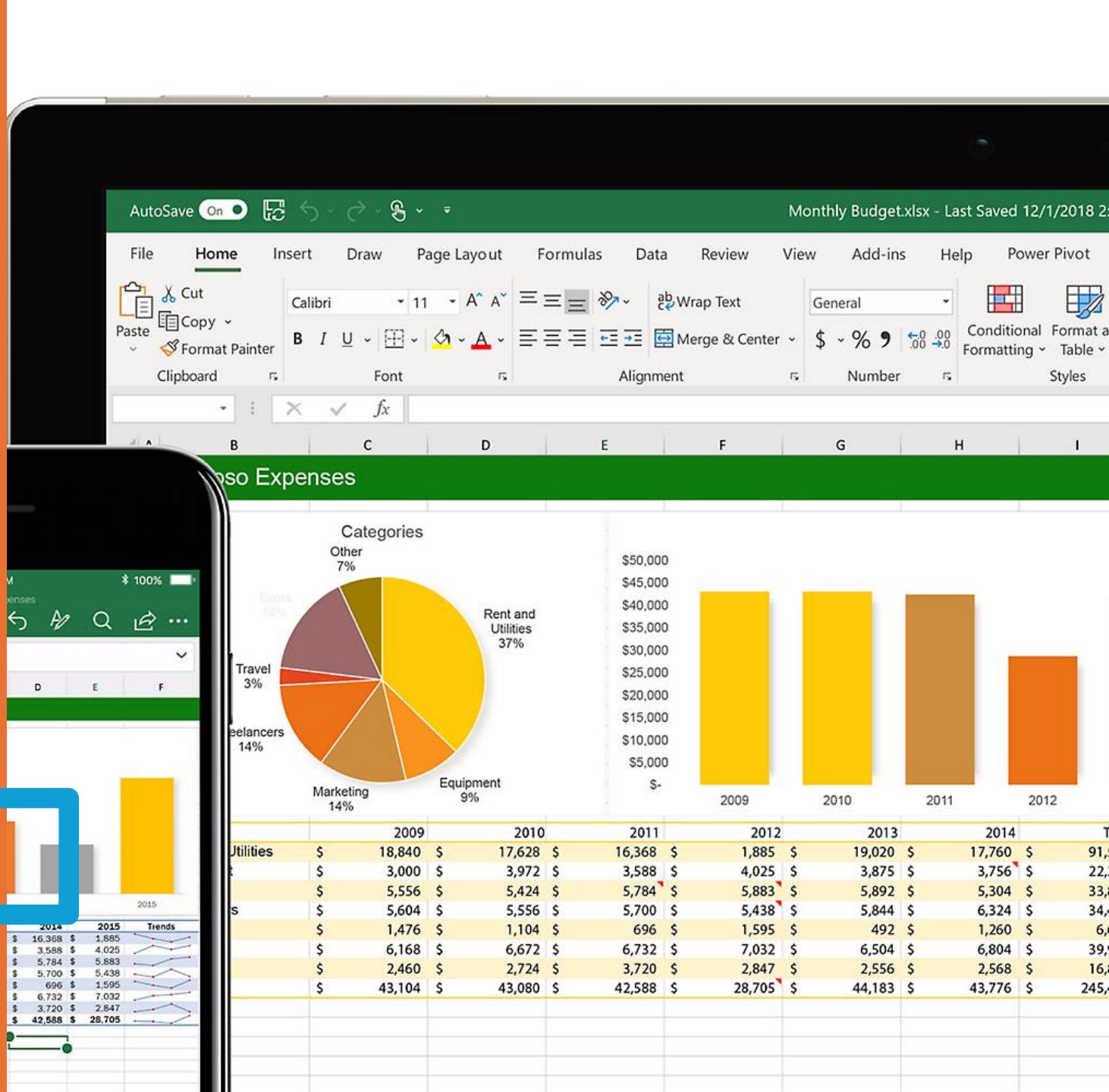
Pros & Cons

- Offers a free plan
- Has hundreds of integrations
- Each plan has unlimited dashboards, metrics and viewer users

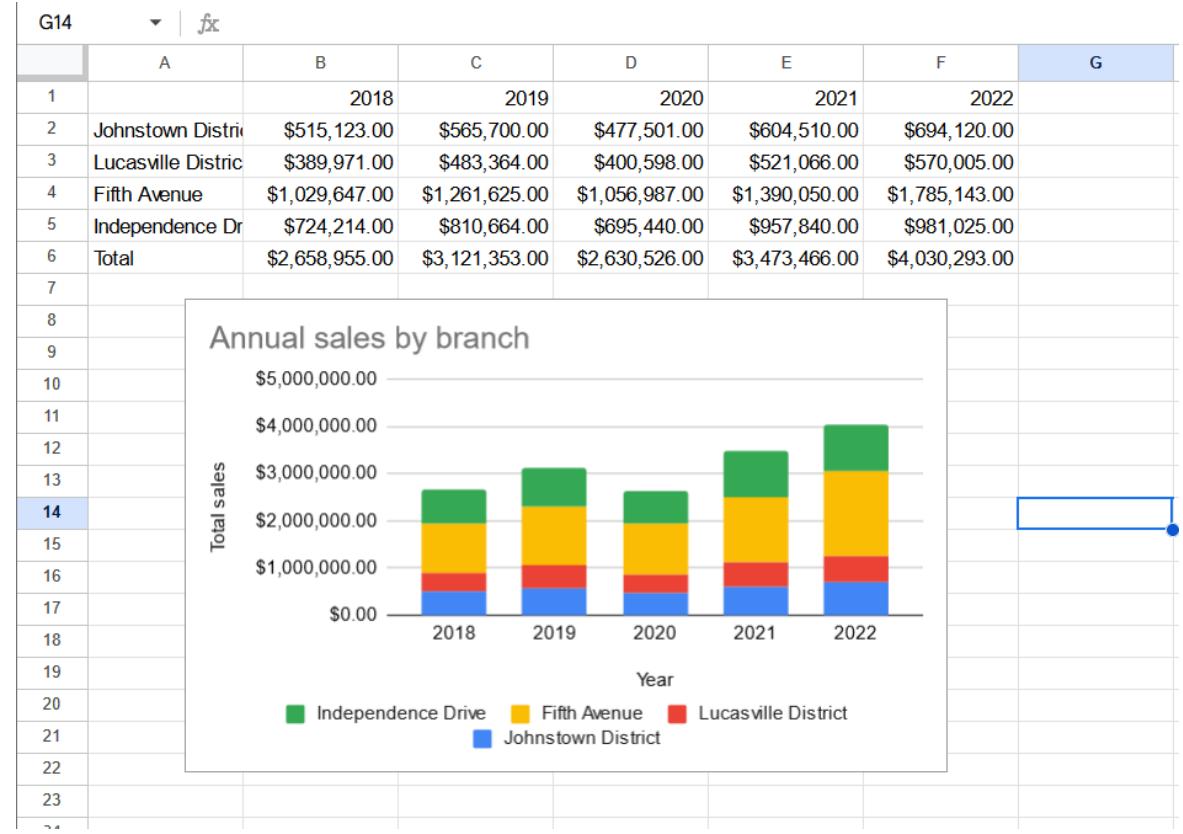
Free plan doesn't include downloadable PDF reports

<https://www.klipfolio.com>

Personal Reports:



Google Sheets

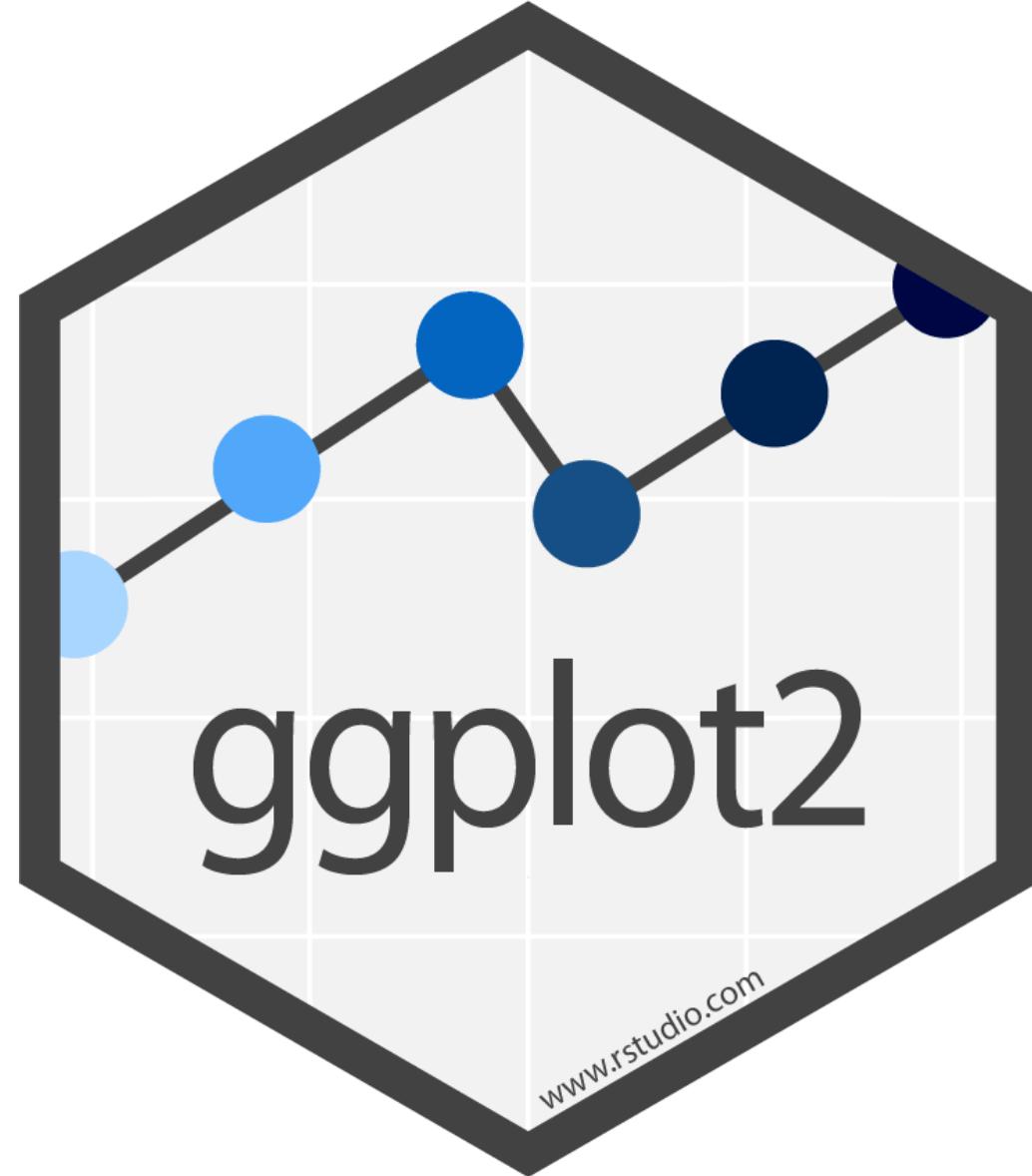




<https://www.datawrapper.de/>

Great Personal and Commercial Options

We will learn this next time





<https://multimedia.journalism.berkeley.edu/blog/free-data-visualization-tools/>

Tableau Example

<https://public.tableau.com/>



In-Class Exercise

- Objectives
 - Explore "zero-code" software
 - See required elements of various graphs
 - Learn an online software as a comparison to R
- Instructions
 - <https://public.tableau.com/> and create a free account
 - They have a web version or public desktop version you can download
 - Find and load sample data
 - <https://www.kaggle.com/datasets>
 - <https://www.tableau.com/learn/articles/free-public-data-sets>
 - Google “free data”, there is a ton out there
 - Create an interesting graph
- **On Slack, channel and post your graph(s) in the data-visualization channel, explaining what it is.**