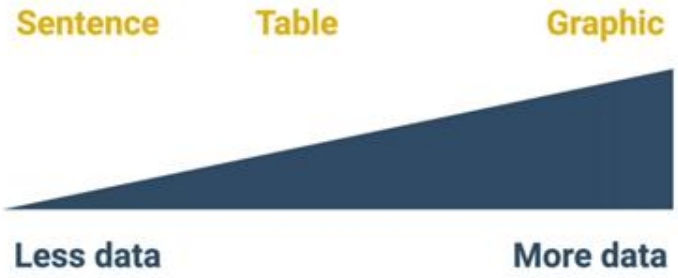
Section 5

29- Do you really need a plot

29. Wait, do you really need a plot?

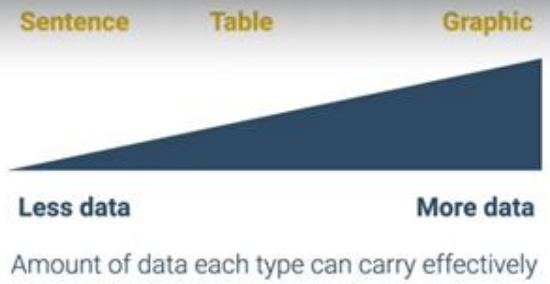
How much data?



Amount of data each type can carry effectively

29. Wait, do you really need a plot?

How much data?

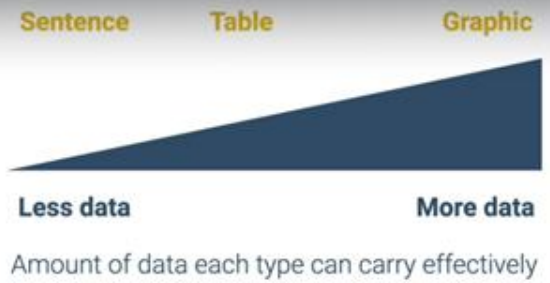


Amount of data each type can carry effectively

In the summer holiday period, 43% of children go on vacation, while the rest stays in town.

29. Wait, do you really need a plot?

How much data?



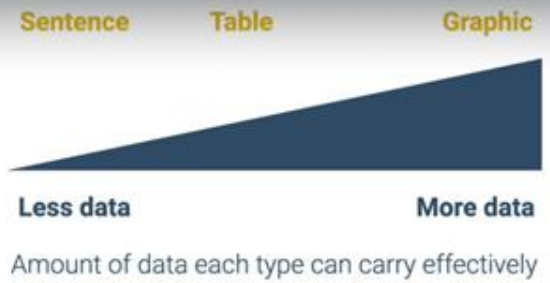
Amount of data each type can carry effectively

In the summer holiday period, 43% of children go on vacation, while the rest stays in town.

Children that go on vacation	43 %
Children that stay in town	57 %

29. Wait, do you really need a plot?

How much data?



Amount of data each type can carry effectively

In the summer holiday period, 43% of children go on vacation, while the rest stays in town.

Children that go on vacation	43 %
Children that stay in town	57 %

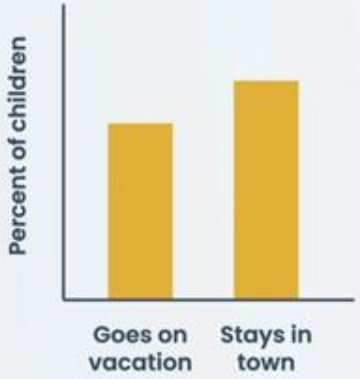
29. Wait, do you really need a plot?

How much data?

Amount of data each type can carry effectively

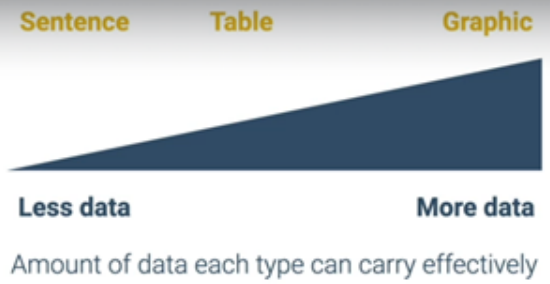
In the summer holiday period, 43% of children go on vacation, while the rest stays in town.

Children that go on vacation	43 %
Children that stay in town	57 %



29. Wait, do you really need a plot?

How much data?

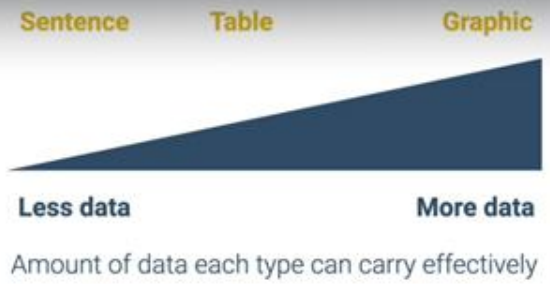


Amount of data each type can carry effectively

In the summer holiday period,
43% of children go on vacation,
15% of children enroll in summer camps,
20% spend their time with their grandparents,
20% watch TV all day long and
only a 2% study for next year.

29. Wait, do you really need a plot?

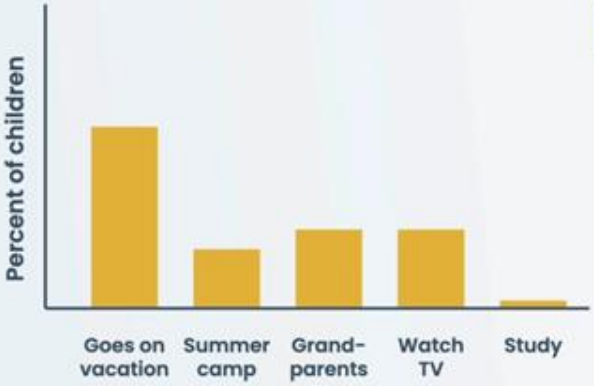
How much data?



Amount of data each type can carry effectively

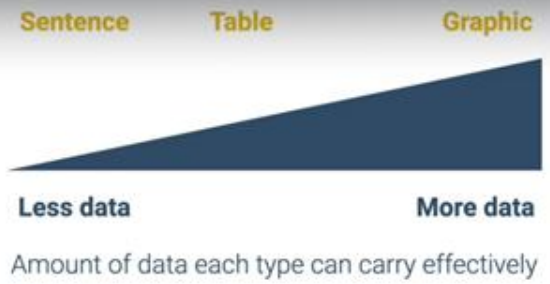
In the summer holiday period,
43% of children go on vacation,
15% of children enroll in summer camps,
20% spend their time with their grandparents,
20% watch TV all day long and
only a 2% study for next year.

Go on vacation	43 %
Enroll in a summer camp	15 %
Go to grandparent's	20 %
Watch TV	20 %
Study for next year	2 %



29. Wait, do you really need a plot?

How much data?

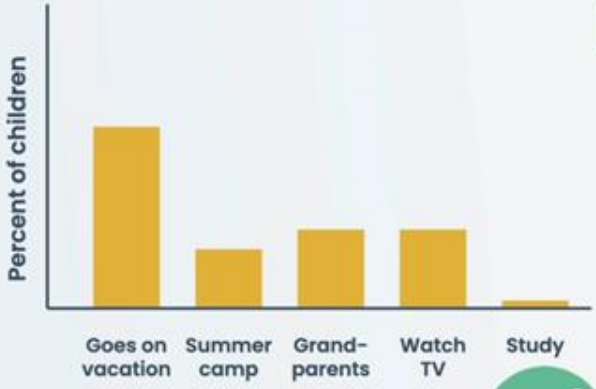


Amount of data each type can carry effectively

In the summer holiday period,
43% of children go on vacation,
15% of children enroll in summer camps,
20% spend their time with their grandparents,
20% watch TV all day long and
only a 2% study for next year.

Go on vacation	43 %
Enroll in a summer camp	15 %
Go to grandparent's	20 %
Watch TV	20 %
Study for next year	2 %

Quantitative



1/28/2025 8:22:16 PM

29. Wait, do you really need a plot?

How much data?

Sentences

Are an effective way to display numerical information, but not if there are more than two values, because it does not allow for comparisons.

29. Wait, do you really need a plot?

How much data?

Sentences

Are an effective way to display numerical information, but not if there are more than two values, because it does not allow for comparisons.

Tables

Are the best way to show exact numerical values.
But the entries should be arranged in semi-graphical form, helping the user locate the data.

29. Wait, do you really need a plot?

How much data?

Sentences

Are an effective way to display numerical information, but not if there are more than two values, because it does not allow for comparisons.

Tables

Are the best way to show exact numerical values.
But the entries should be arranged in semi-graphical form, helping the user locate the data.

Broadband and/or smartphone adoption

% of adults in each group who have a home broadband connection or a smartphone

	White	Black	Difference
Total 18+	80%	72%	-8
Gender			
Male	81	70	-11
Female	80	73	-7
Age			
18-29	95	98	not sig
30-49	91	83	-8
50-64	79	59	-20
65+	54	34	-20
Education			
High school grad or less	64	48	-16
Some college	87	89	not sig
College+	93	93	not sig
Household income			
<$30,000	63	64	not sig
$30,000-$74,999	84	81	not sig
$75,000+	96	96	not sig
Other demographics			
Parents	91	85	-6
Students	97	96	not sig

Pew Research Center's Internet Project July 18-September 30, 2013 tracking survey. N=6010 adults ages 18+. For results based on all adults, n=4,223 for whites and n=664 for African Americans.

PEW RESEARCH CENTER

29. Wait, do you really need a plot?

How much data?

Sentences

Are an effective way to display numerical information, but not if there are more than two values, because it does not allow for comparisons.

Tables

Are the best way to show exact numerical values.
But the entries should be arranged in semi-graphical form, helping the user locate the data.

Graphics

Are best used when:

(I) You are not interested in the particular numerical values, but on trends

(II) When you want to compare data

30- Types of Plots

Mastering Data Visualization: Theory & Foundations

PLOTS

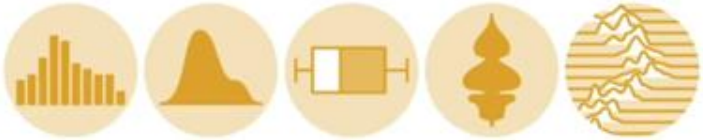
Types of Plots

The Data Ink Lab

What is the purpose?

Distributions

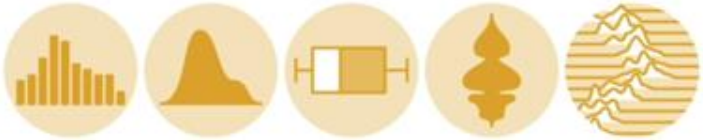
When you have a lot of observations of one variable.
When summary statistics are not informative.



What is the purpose?

Distributions

When you have a lot of observations of one variable.
When summary statistics are not informative.



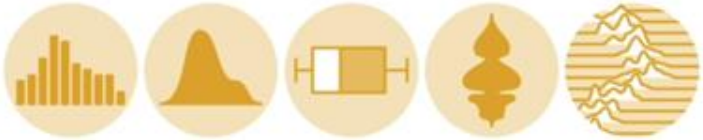
Relationship

When you want to compare two (maximum three) variables, to see what is the dependence of one versus the other.

What is the purpose?

Distributions

When you have a lot of observations of one variable. When summary statistics are not informative.



Relationship

When you want to compare two (maximum three) variables, to see what is the dependence of one versus the other.

Ranking

You want to compare several variables, and you are interested in which one is bigger than the other, but not necessarily *how much* bigger.

What is the purpose?

Part to whole

When a variable is divided into categories, and you want to know what is the predominant category. Not so much for an exact comparison of the proportions of categories.

What is the purpose?

Part to whole

When a variable is divided into categories, and you want to know what is the predominant category. Not so much for an exact comparison of the proportions of categories.

What is the purpose?

Part to whole

When a variable is divided into categories, and you want to know what is the predominant category. Not so much for an exact comparison of the proportions of categories.

Map

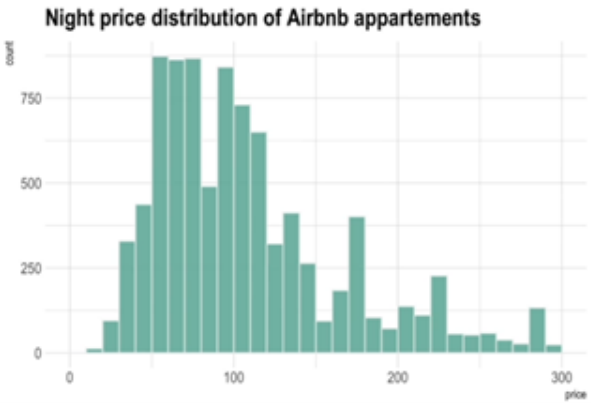
When your data is geographical, and you want to plot some variable on top.

31- Plotting Distribution

Plotting distributions

Histogram

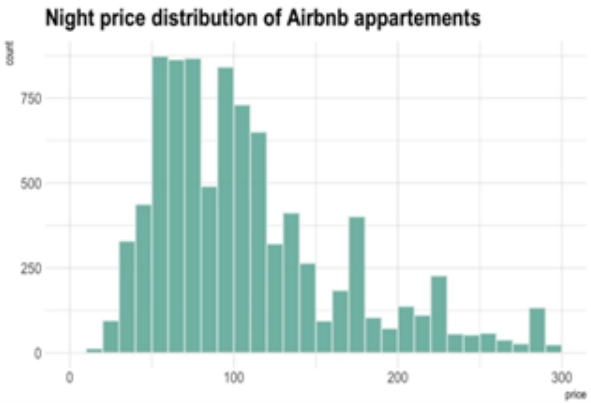
A **histogram** is the graphical representation of the distribution of a numeric variable.



Histogram

A **histogram** is the graphical representation of the distribution of a numeric variable.

The variable is cut into several bins, and the number of observations per bin is represented by the height of the bar.

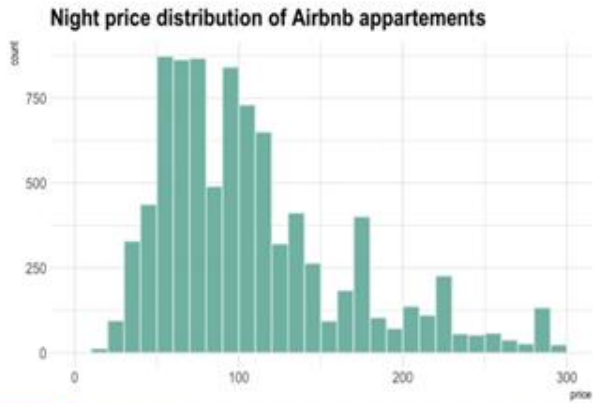


Histogram

A **histogram** is the graphical representation of the distribution of a numeric variable.

The variable is cut into several bins, and the number of observations per bin is represented by the height of the bar.

Try several bin sizes. It can lead to graphs that look very different.

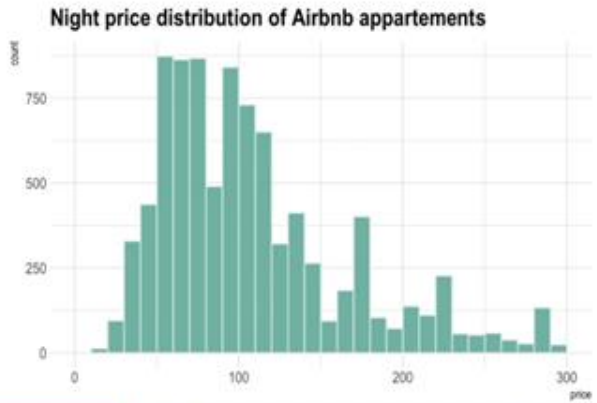


Histogram

A **histogram** is the graphical representation of the distribution of a numeric variable.

The variable is cut into several bins, and the number of observations per bin is represented by the height of the bar.

- *Try several bin sizes. It can lead to graphs that look very different.*
- *It takes as input numeric variables only.*

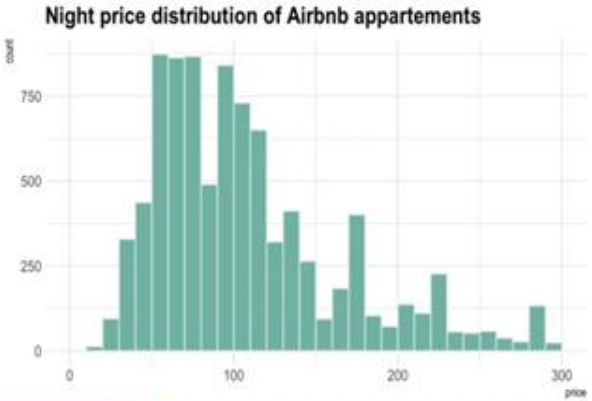


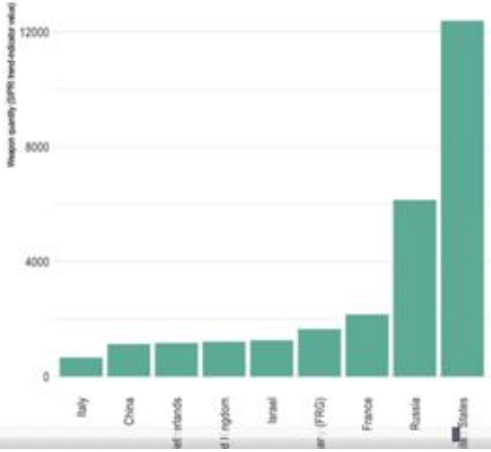
Histogram

A **histogram** is the graphical representation of the distribution of a numeric variable.

The variable is cut into several bins, and the number of observations per bin is represented by the height of the bar.

- *Try several bin sizes. It can lead to graphs that look very different.*
- *It takes as input numeric variables only.*



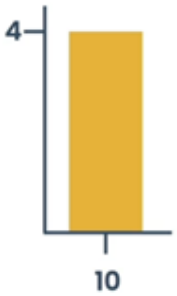


Histogram

A **histogram** is the graphical representation of the distribution of a numeric variable.

The variable is cut into several bins, and the number of observations per bin is represented by the height of the bar.

It takes as input numeric variables only.



Histogram

A **histogram** is the graphical representation of the distribution of a numeric variable.

The variable is cut into several bins, and the number of observations per bin is represented by the height of the bar.

It takes as input numeric variables only.

Bar plot

1 observation
where x=10 and y=4

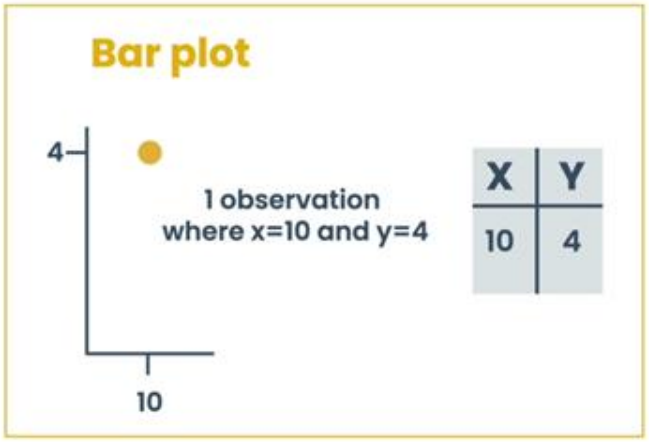
X	Y
10	4

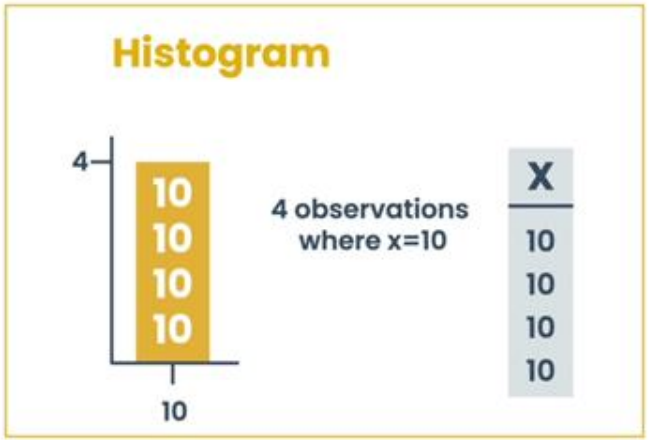
Histogram

A **histogram** is the graphical representation of the distribution of a numeric variable.

The variable is cut into several bins, and the number of observations per bin is represented by the height of the bar.

It takes as input numeric variables only.





Histogram

A **histogram** is the graphical representation of the distribution of a numeric variable.

The variable is cut into several bins, and the number of observations per bin is represented by the height of the bar.

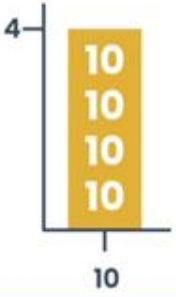
It takes as input numeric variables only.

Bar plot

1 observation
where x=10 and y=4

X	Y
10	4

Histogram

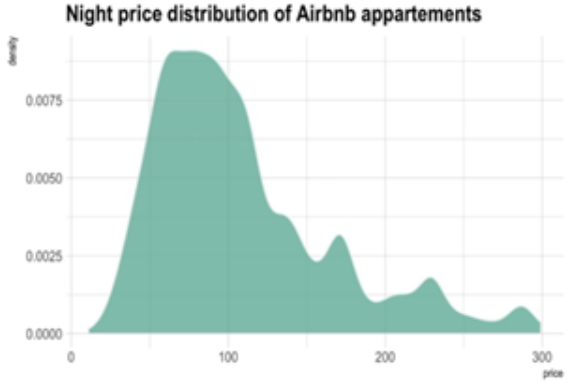


4 observations
where x=10

X
10
10
10
10

Density plot

Representation of the distribution of a numeric variable (histogram for continuous variables).



Density plot

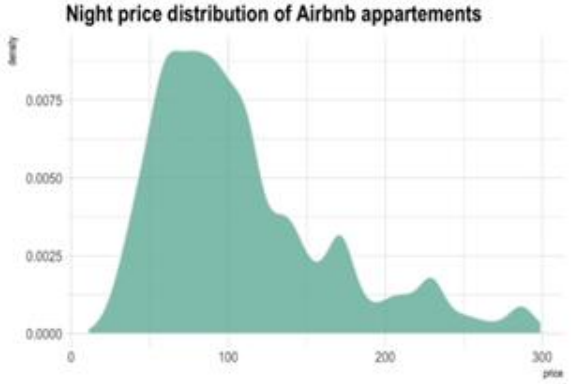
Representation of the distribution of a numeric variable (histogram for continuous variables).

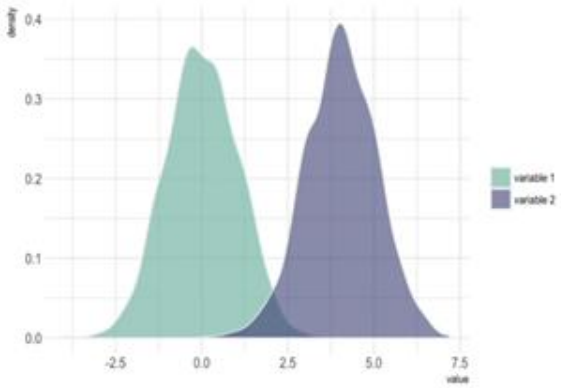
It takes as input numeric variables only.

Density plot

Representation of the distribution of a numeric variable (histogram for continuous variables).

It takes as input numeric variables only.





A good way to compare distributions

Boxplot

Also known as **Box and Whiskers plot**.

Boxplot

Also known as **Box and Whiskers plot**.

Boxplot

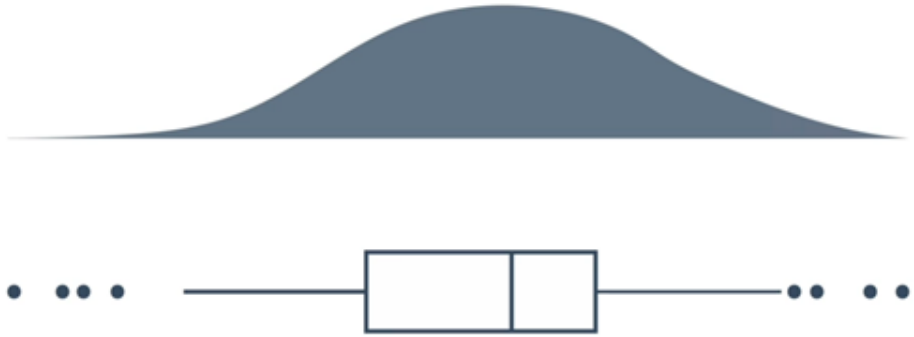
Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

Boxplot

Also known as **Box and Whiskers plot**.

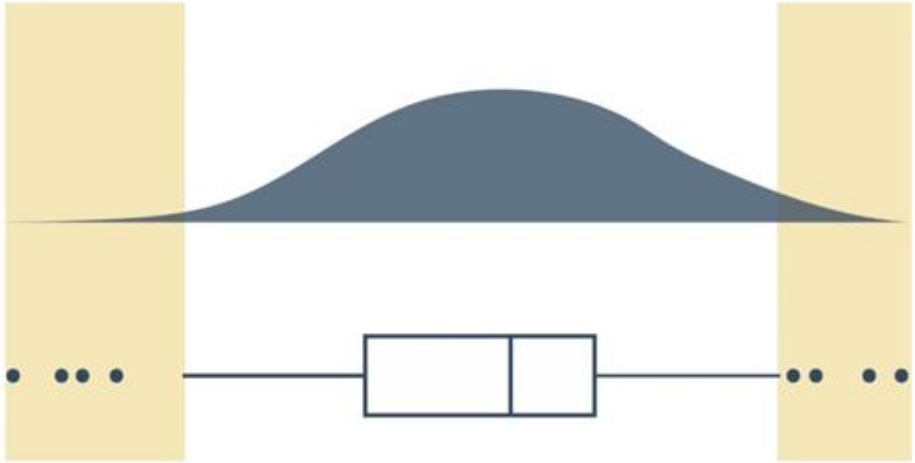
A boxplot summarizes the distribution of a numeric variable for several groups.



Boxplot

Also known as **Box and Whiskers plot**.

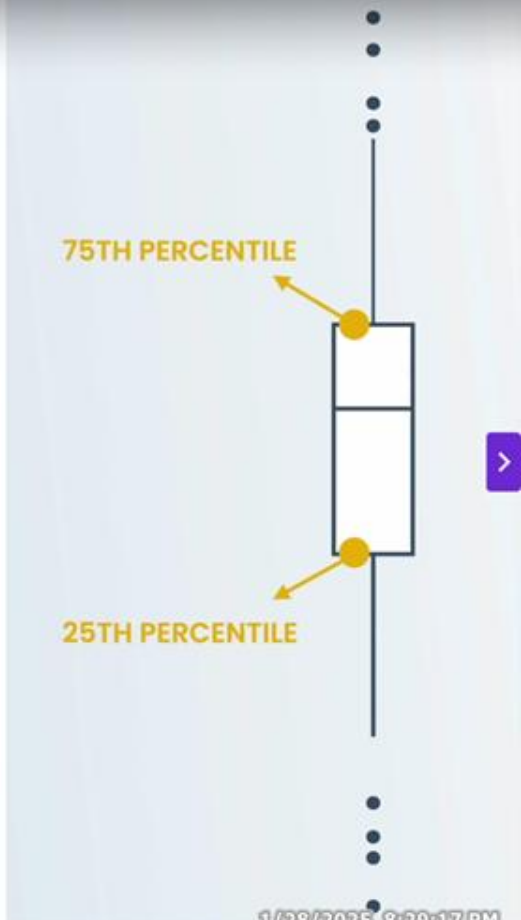
A boxplot summarizes the distribution of a numeric variable for several groups.



Boxplot

Also known as **Box and Whiskers plot**.

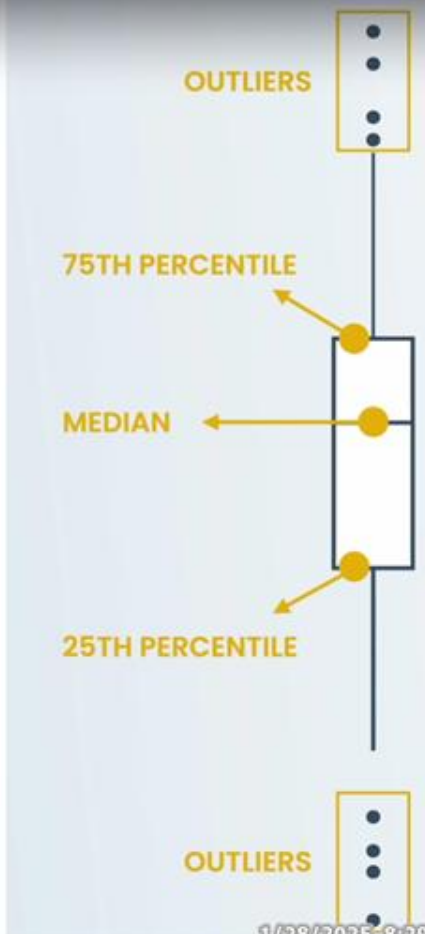
A boxplot summarizes the distribution of a numeric variable for several groups.



Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

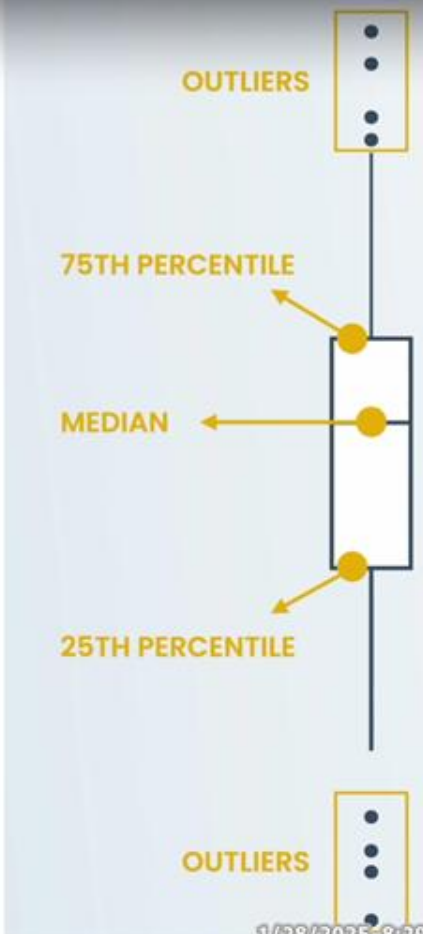


Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

The problem is that summarizing also means losing information.

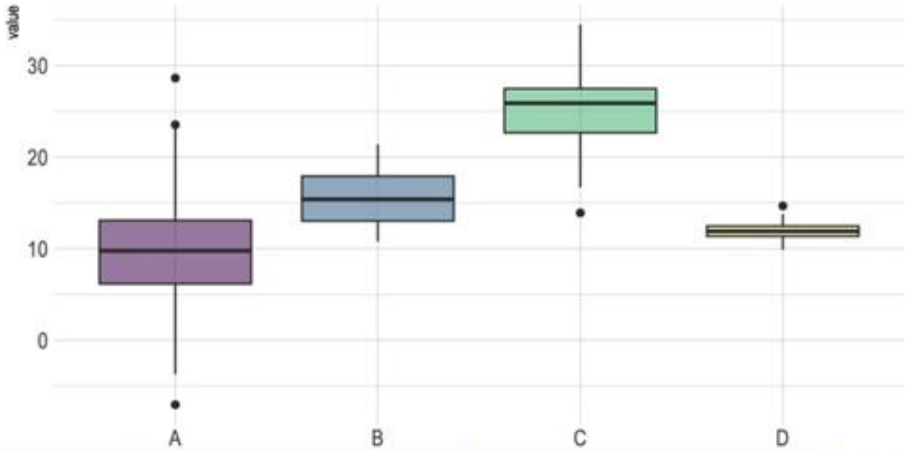


Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

The problem is that summarizing also means losing information.



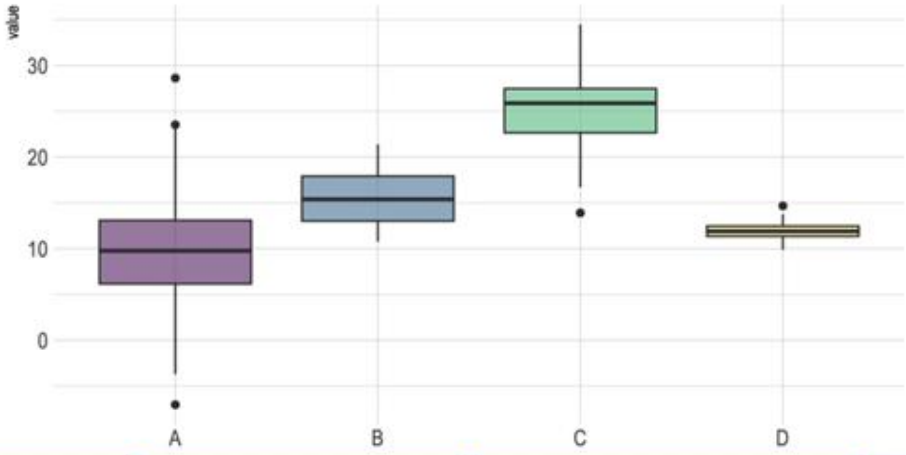
A boxplot hides the sample size

Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

The problem is that summarizing also means losing information.



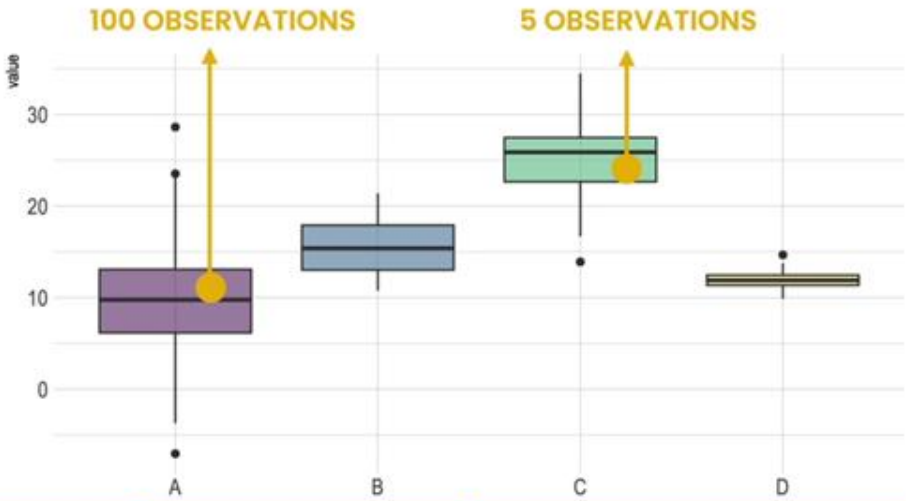
A boxplot hides the sample size

Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

The problem is that summarizing also means losing information.



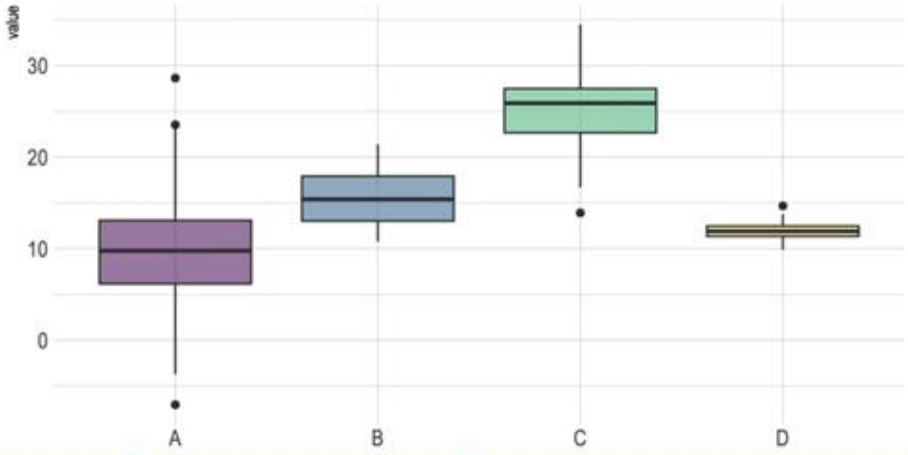
A boxplot hides the sample size

Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

The problem is that summarizing also means losing information.



A boxplot hides the sample size

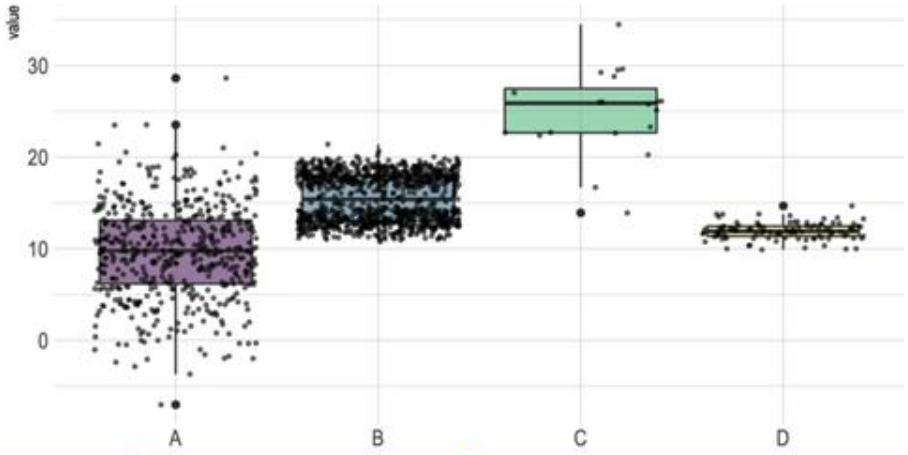
Plot your data points on top

Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

The problem is that summarizing also means losing information.



A boxplot hides the sample size

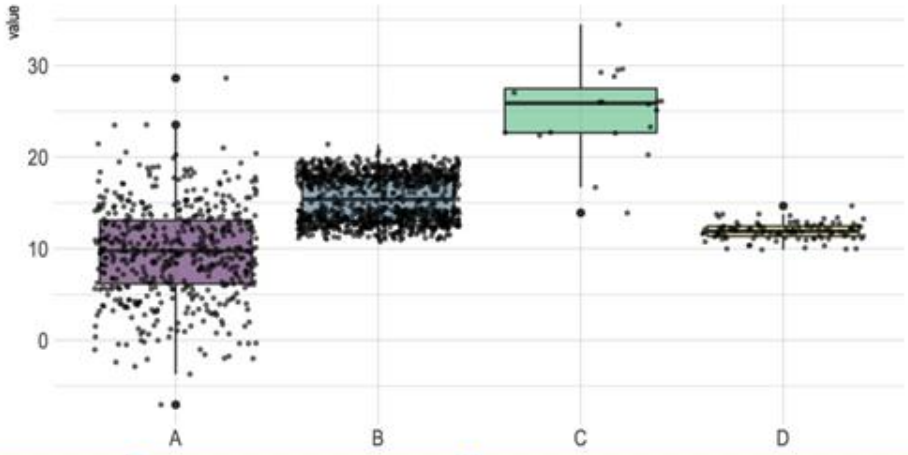
Plot your data points on top

Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

The problem is that summarizing also means losing information.



A boxplot hides the sample size

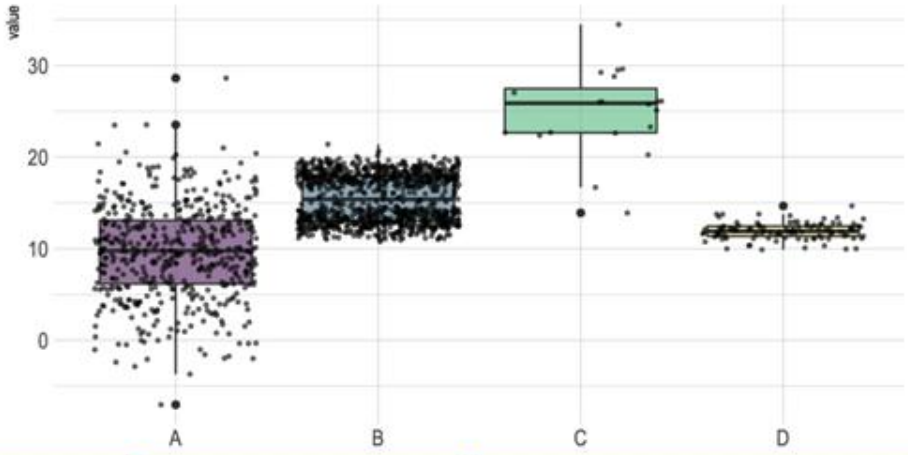
Plot your data points on top

Boxplot

Also known as **Box and Whiskers plot**.

A boxplot summarizes the distribution of a numeric variable for several groups.

The problem is that summarizing also means losing information.



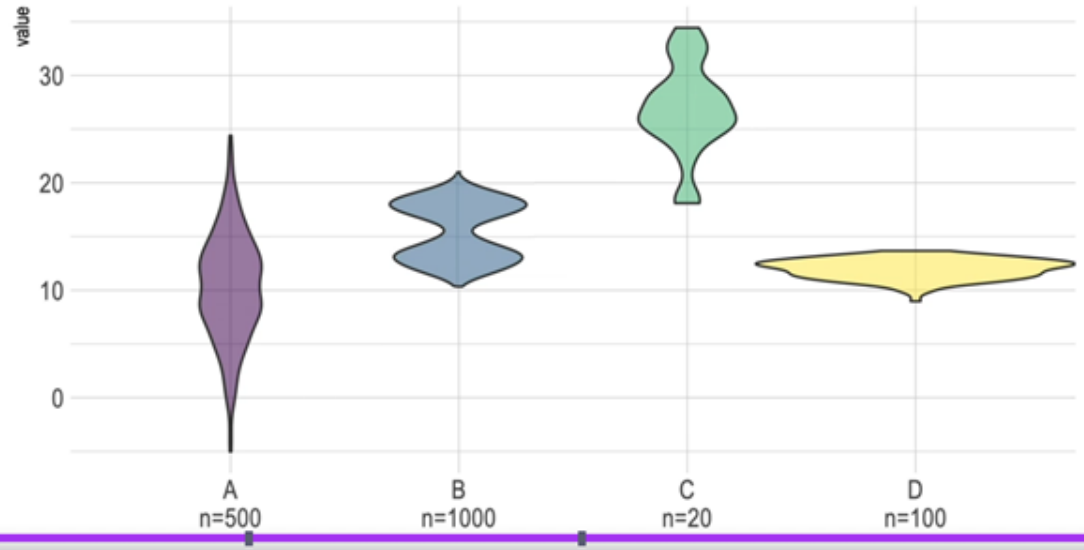
A boxplot hides the sample size

Plot your data points on top

Use a violin plot

Violin plot

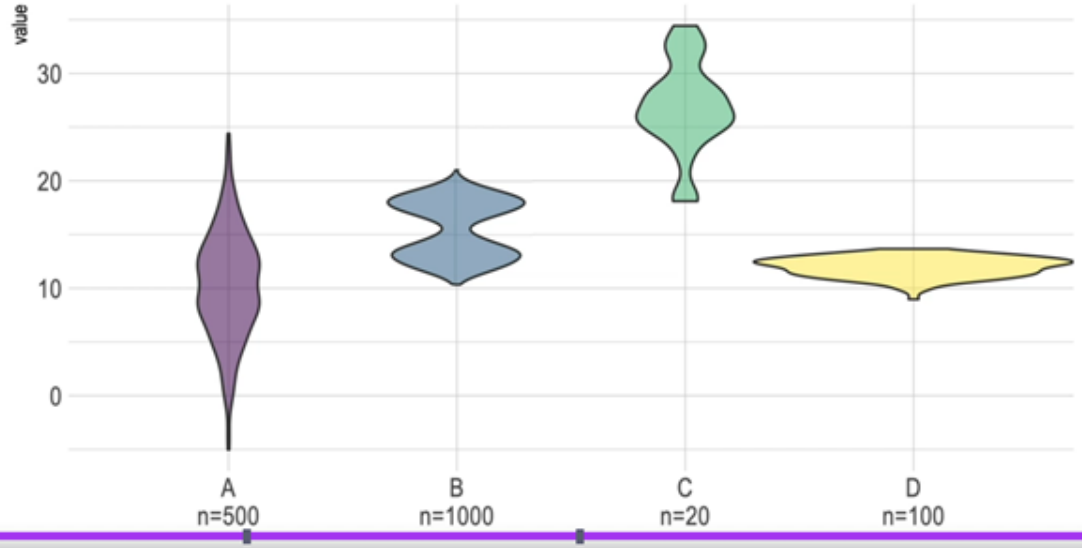
Alternative to a boxplot.



Violin plot

Alternative to a boxplot.

Plot of the distribution of each variable in a symmetric way.

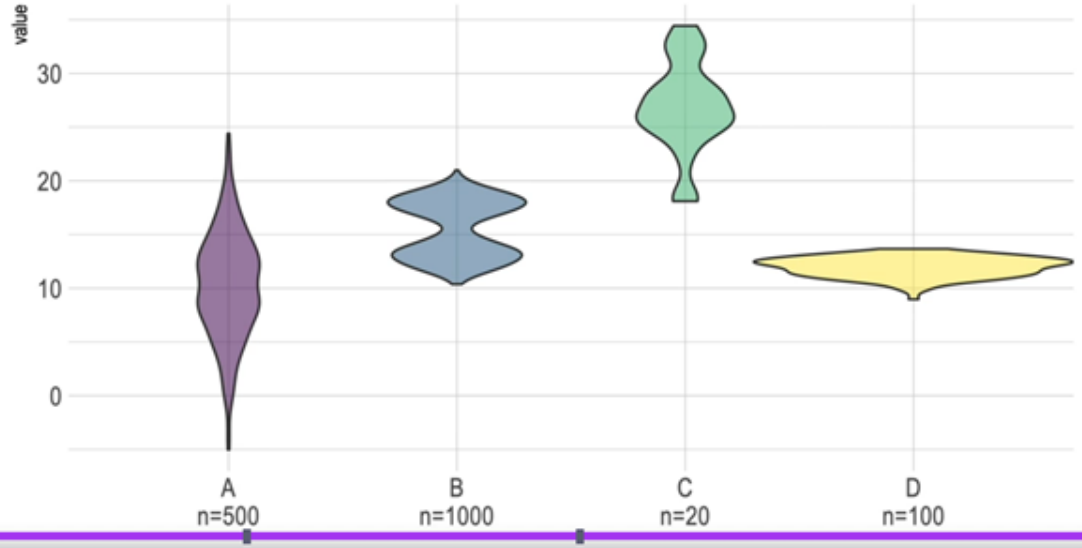


Violin plot

Alternative to a boxplot.

Plot of the distribution of each variable in a symmetric way.

Good when we have lots of data and cannot show individual observations.

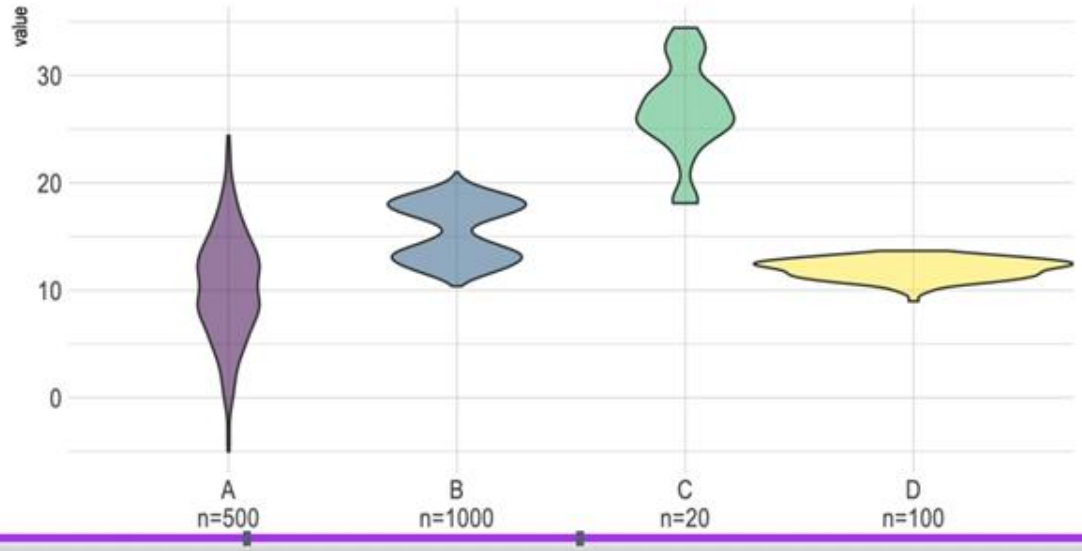


Violin plot

Alternative to a boxplot.

Plot of the distribution of each variable in a symmetric way.

Good when we have lots of data and cannot show individual observations.

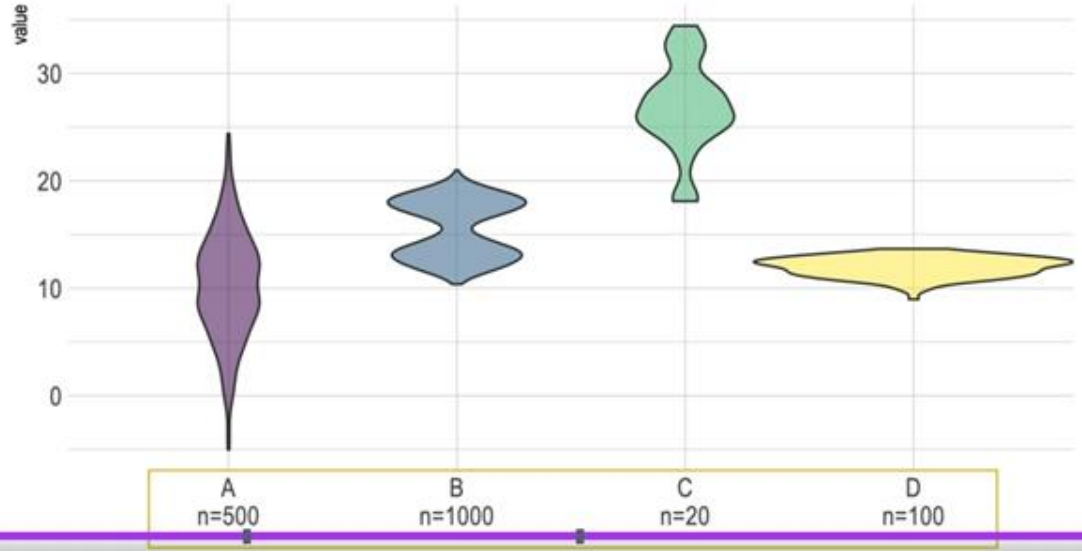


Violin plot

Alternative to a boxplot.

Plot of the distribution of each variable in a symmetric way.

Good when we have lots of data and cannot show individual observations.



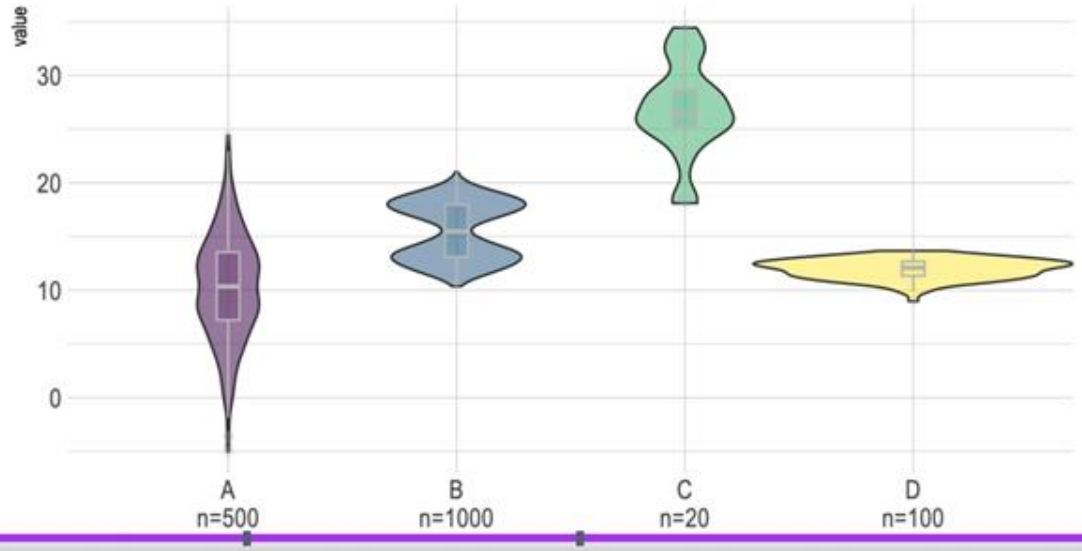
If sample sizes are very different, show it

Violin plot

Alternative to a boxplot.

Plot of the distribution of each variable in a symmetric way.

Good when we have lots of data and cannot show individual observations.



If sample sizes are very different, show it

Can be used together with a boxplot

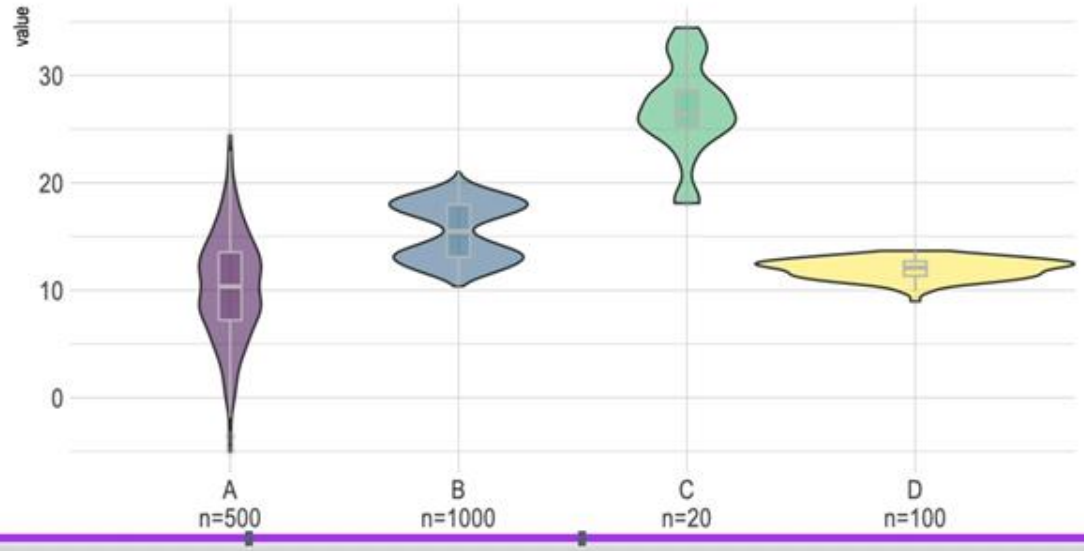
Unless your variables have an implicit order, order them by median value

Violin plot

Alternative to a boxplot.

Plot of the distribution of each variable in a symmetric way.

Good when we have lots of data and cannot show individual observations.



If sample sizes are very different, show it

Can be used together with a boxplot

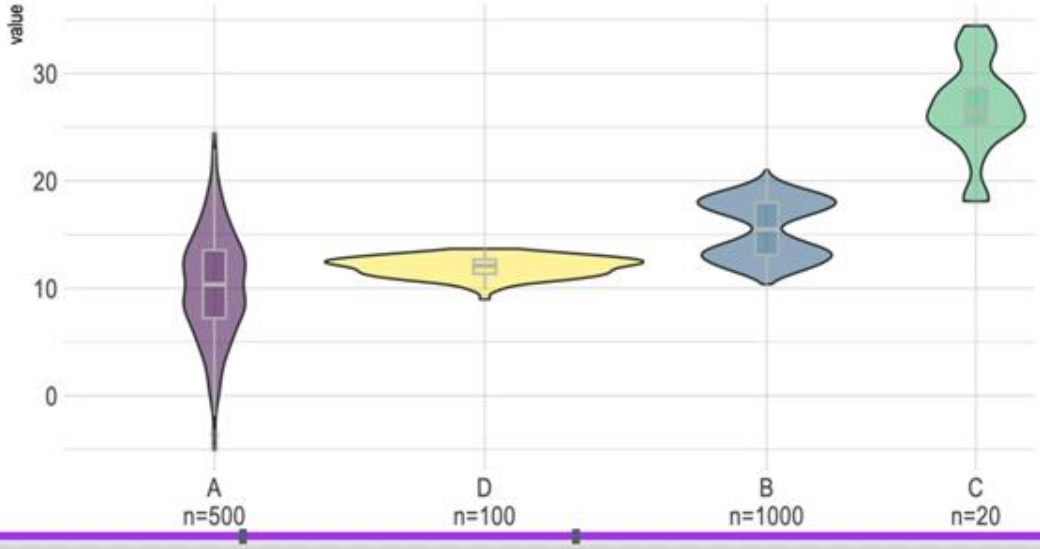
Unless your variables have an implicit order, order them by median value

Violin plot

Alternative to a boxplot.

Plot of the distribution of each variable in a symmetric way.

Good when we have lots of data and cannot show individual observations.

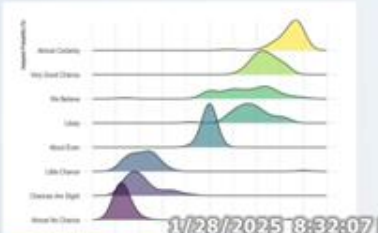


If sample sizes are very different, show it

Can be used together with a boxplot

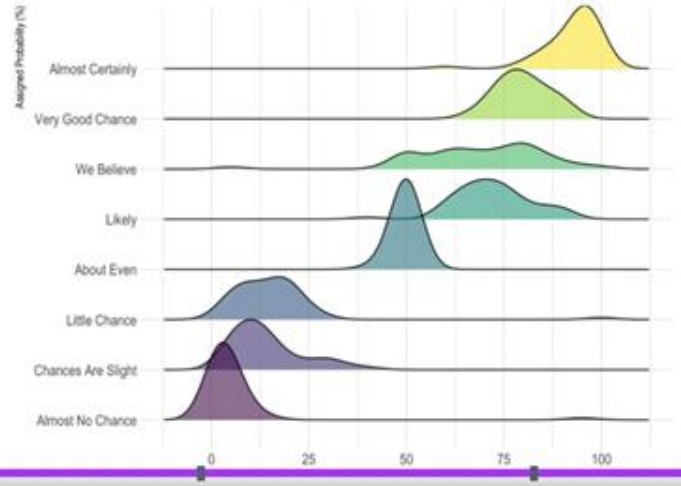
Unless your variables have an implicit order, order them by median value

For few variables: consider a ridgeline plot.



Ridgeline plot

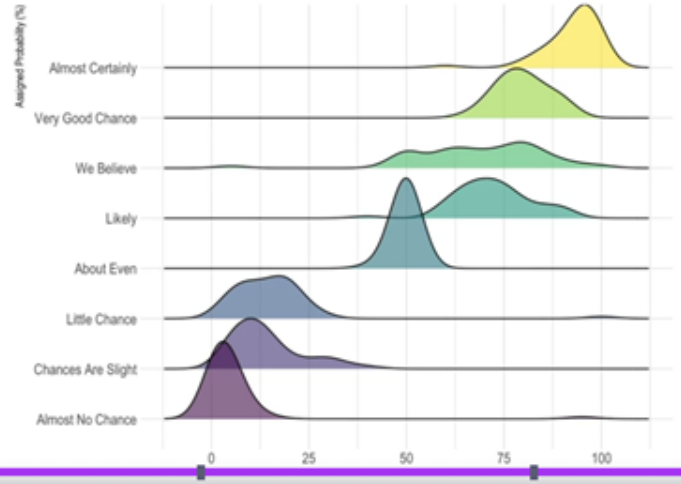
Ridgeline plots show the distribution of a numerical value for several groups.

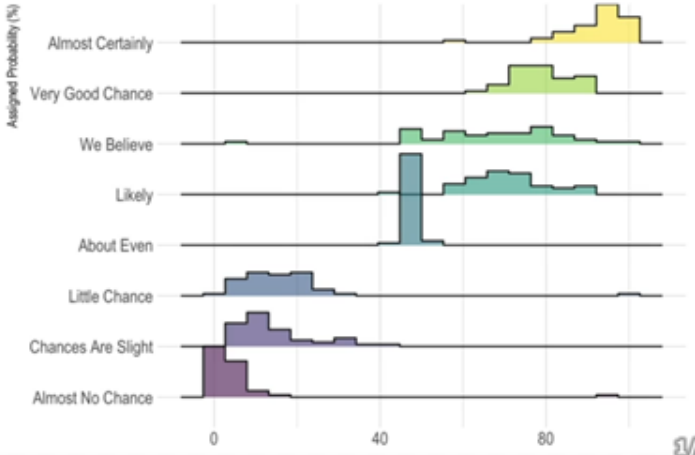


Ridgeline plot

Ridgeline plots show the distribution of a numerical value for several groups.

The distribution can be shown using density plots or histograms.

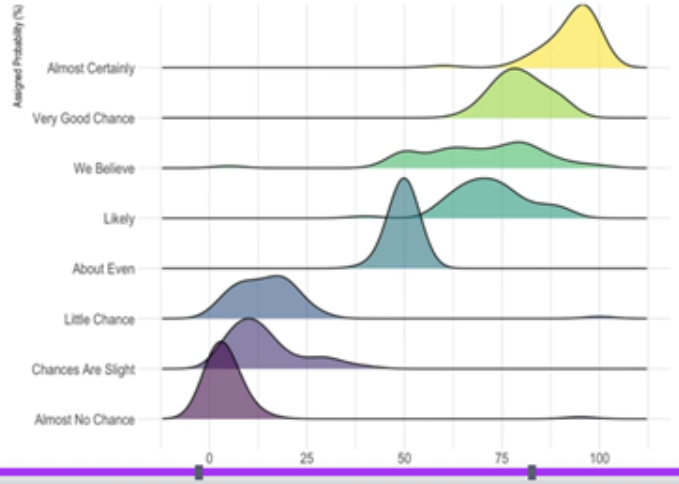


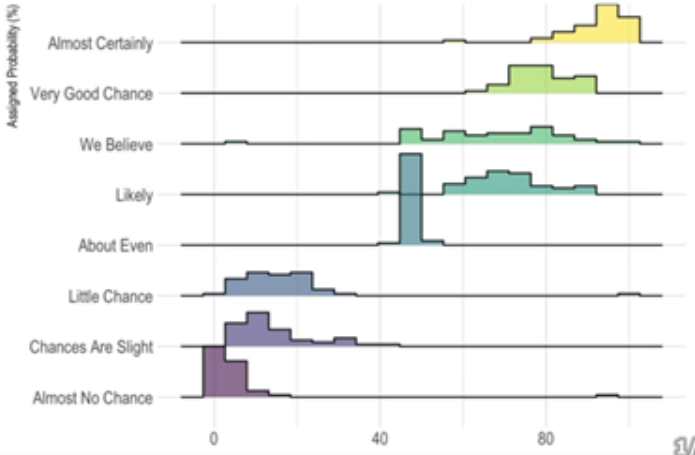


Ridgeline plot

Ridgeline plots show the distribution of a numerical value for several groups.

The distribution can be shown using density plots or histograms.



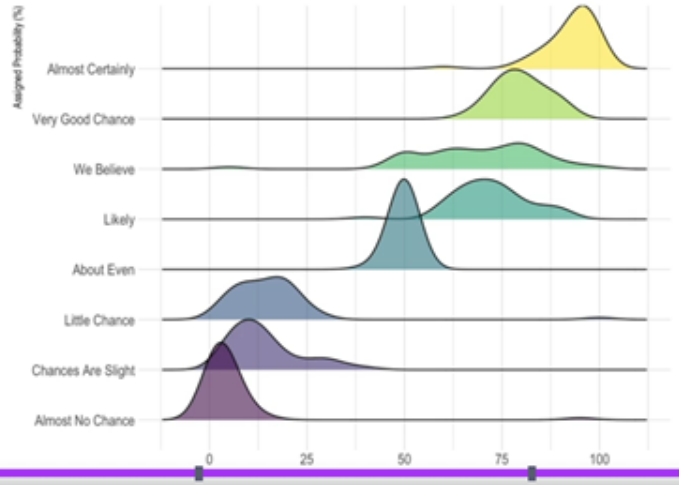


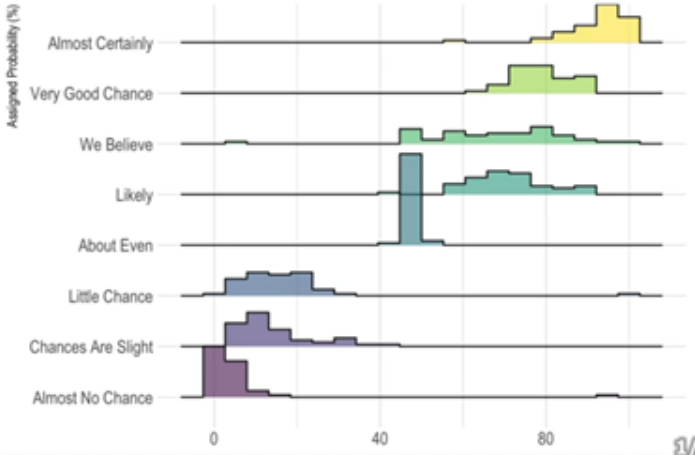
Ridgeline plot

Ridgeline plots show the distribution of a numerical value for several groups.

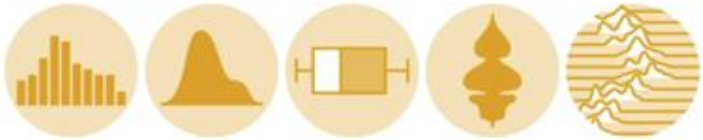
The distribution can be shown using density plots or histograms.

Only good when: you have few variables & there is a clear pattern (not too many overlaps)





Plotting distributions

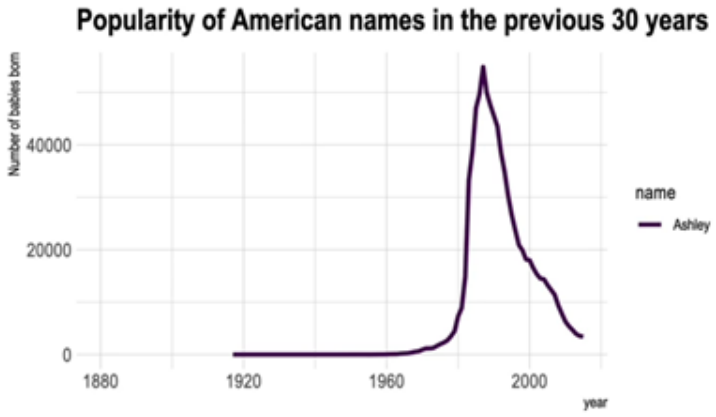


32- Plotting Relationship Between Variables

Relationships between variables

Line plot

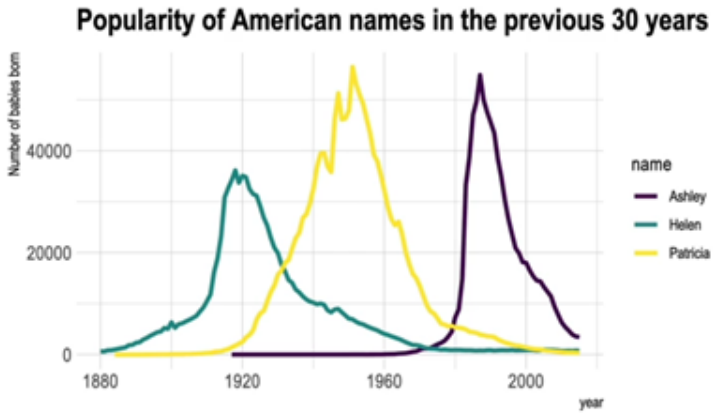
A line chart graph displays the evolution of one or several numeric variables.



Line plot

A line chart graph displays the evolution of one or several numeric variables.

Often used to visualize a trend in data over intervals of time – a time series.

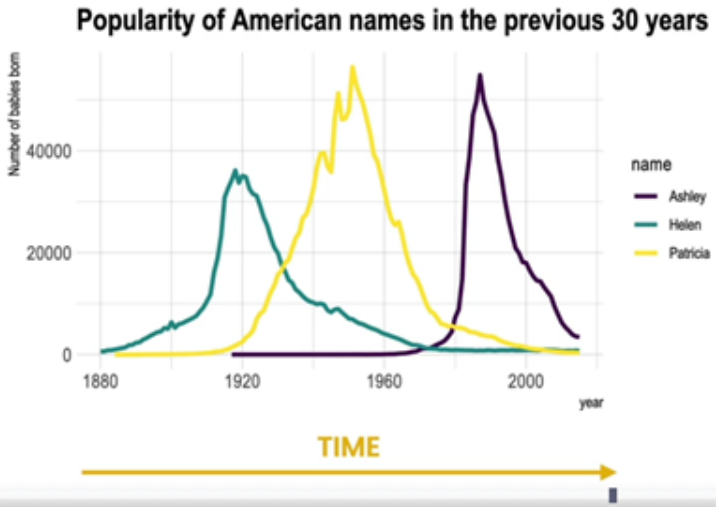


Line plot

A line chart graph displays the evolution of one or several numeric variables.

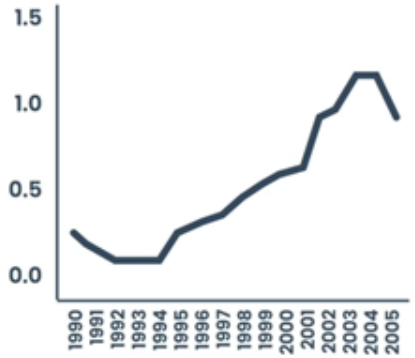
Often used to visualize a trend in data over intervals of time – a time series.

When we see a line plot, we assume that variable x is continuous.



When we see a line plot, we assume that variable x is continuous.

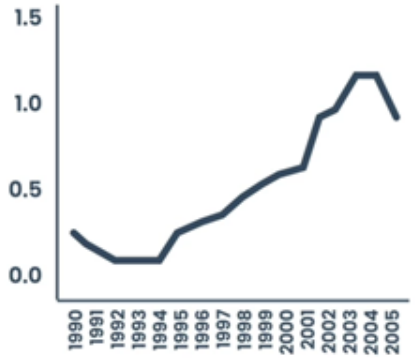
Numerical and continuous
or
Numerical and discrete
with very high resolution



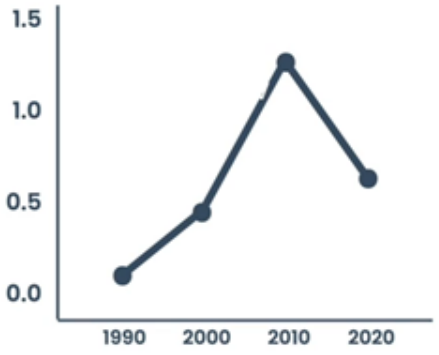
32. Plotting Relationships between variables

When we see a line plot, we assume that variable x is continuous.

Numerical and continuous
or
Numerical and discrete
with very high resolution

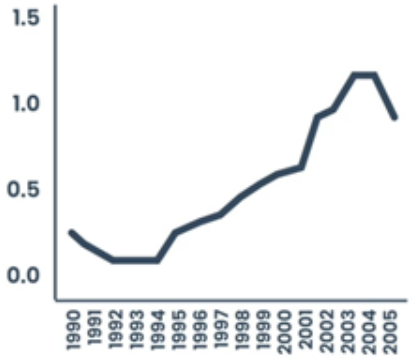


Numerical and discrete
with low resolution

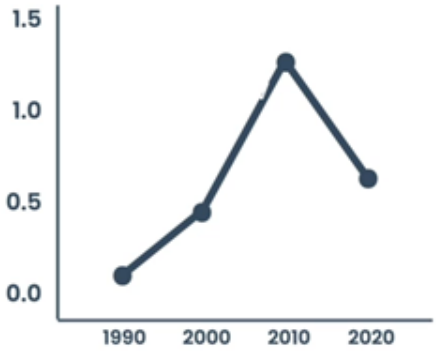


When we see a line plot, we assume that variable x is continuous.

Numerical and continuous
or
Numerical and discrete
with very high resolution



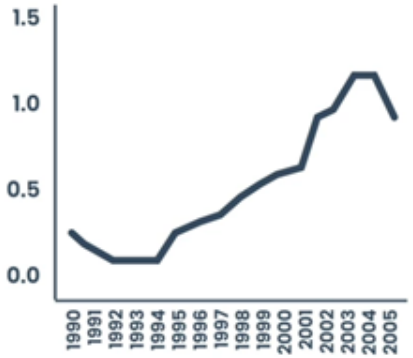
Numerical and discrete
with low resolution



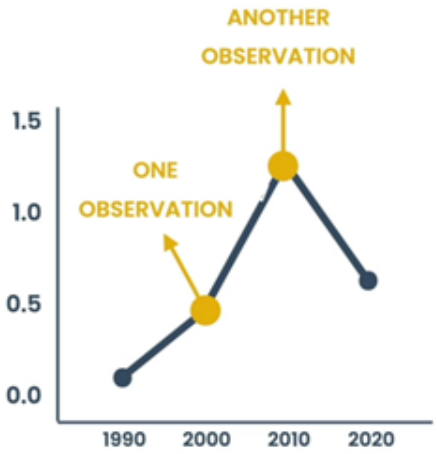
Connected scatterplot

When we see a line plot, we assume that variable x is continuous.

Numerical and continuous
or
Numerical and discrete
with very high resolution



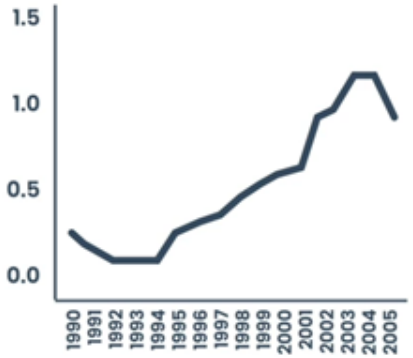
Numerical and discrete
with low resolution



Connected scatterplot

When we see a line plot, we assume that variable x is continuous.

Numerical and continuous
or
Numerical and discrete
with very high resolution

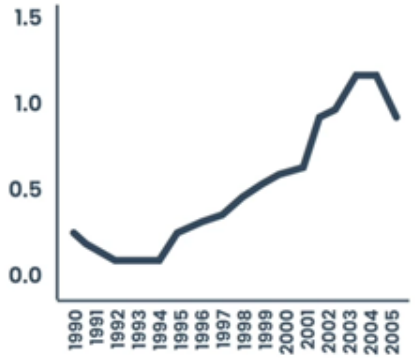


Numerical and discrete
with low resolution

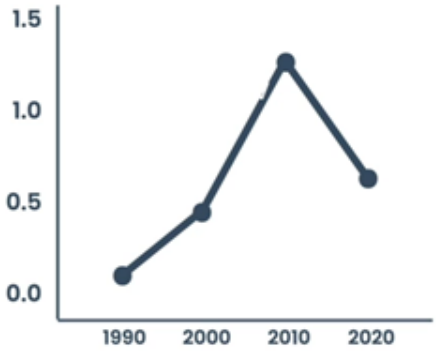
Connected scatterplot

When we see a line plot, we assume that variable x is continuous.

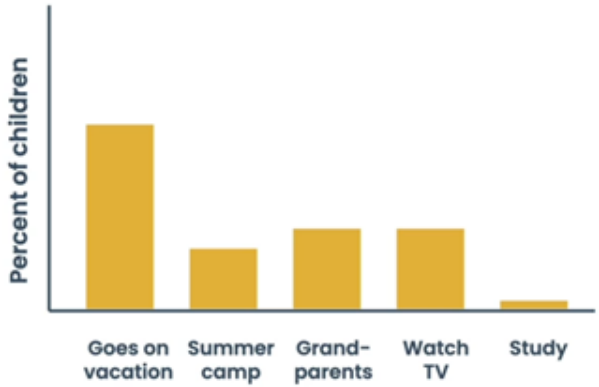
Numerical and continuous or Numerical and discrete with very high resolution



Numerical and discrete with low resolution



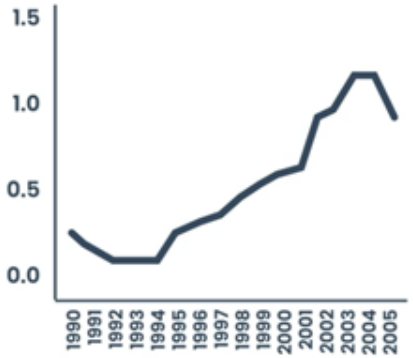
Categorical



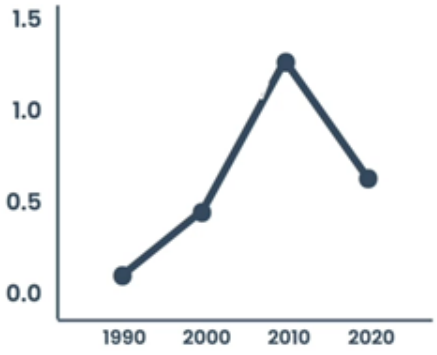
32. Plotting Relationships between variables

When we see a line plot, we assume that variable x is continuous.

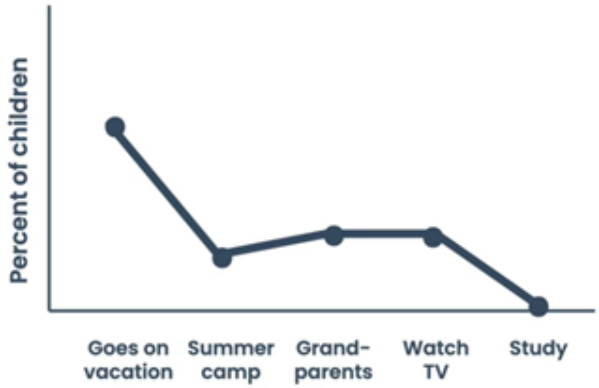
Numerical and continuous or Numerical and discrete with very high resolution



Numerical and discrete with low resolution



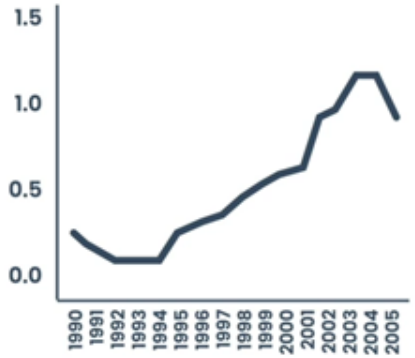
Categorical



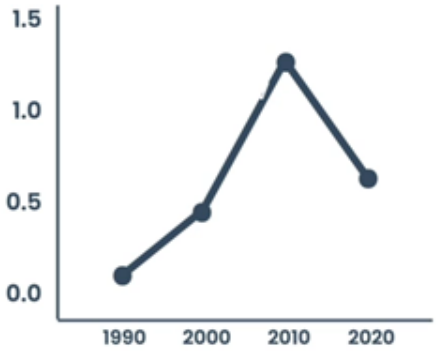
1/28/2025 8:35:39 PM

When we see a line plot, we assume that variable x is continuous.

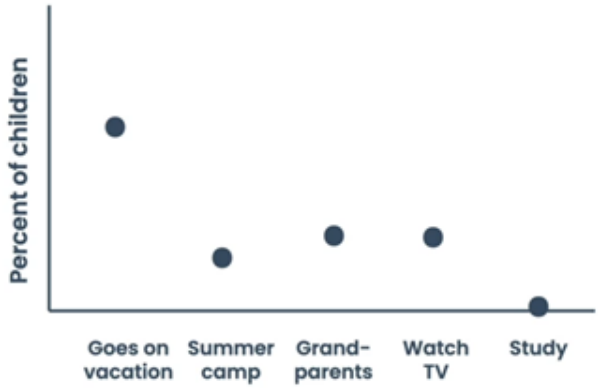
Numerical and continuous or Numerical and discrete with very high resolution



Numerical and discrete with low resolution



Categorical



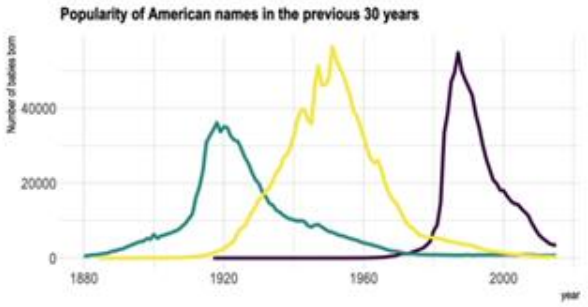
Scatterplot

Line plot > Variation: the area plot

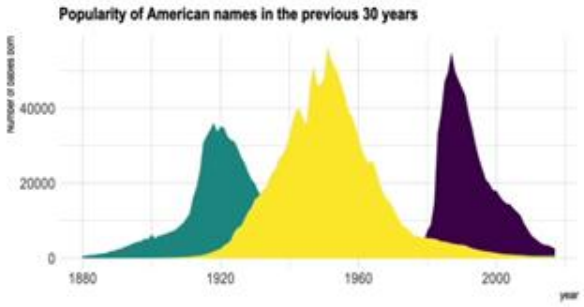
Same line plot, but the area beneath the line is shaded.

Line plot › Variation: the area plot

Same line plot, but the area beneath the line is shaded.



Line plot

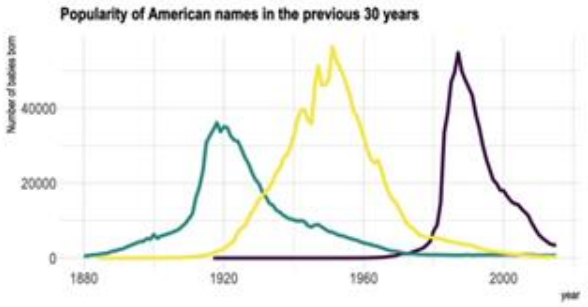


Area plot

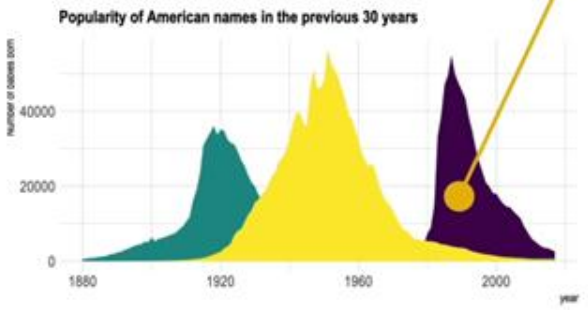
Line plot › Variation: the area plot

Same line plot, but the area beneath the line is shaded.

DOES NOT COMPLY WITH
THE PRINCIPLE OF MINIMUM INK



Line plot



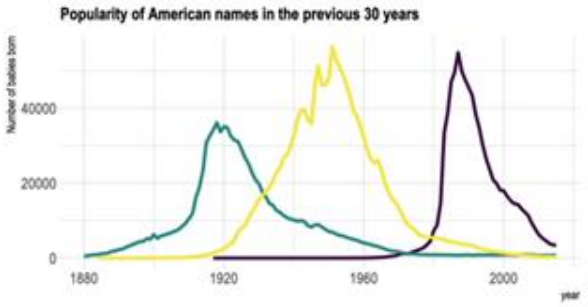
Area plot

Line plot › Variation: the area plot

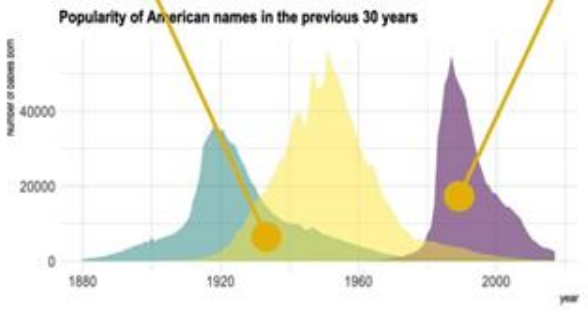
Same line plot, but the area beneath the line is shaded.

DOES NOT COMPLY WITH
THE PRINCIPLE OF MINIMUM INK

ONE AREA COVERS THE OTHER



Line plot



Area plot

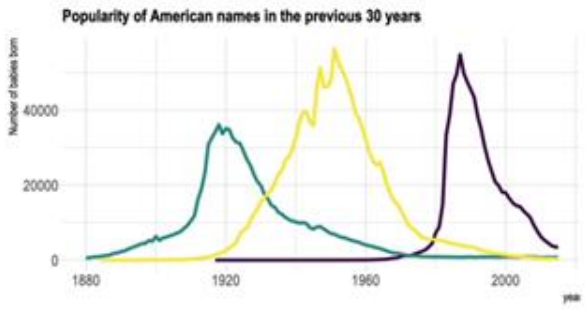
Line plot > Variation: the stacked area plot

Line plot > Variation: the stacked area plot

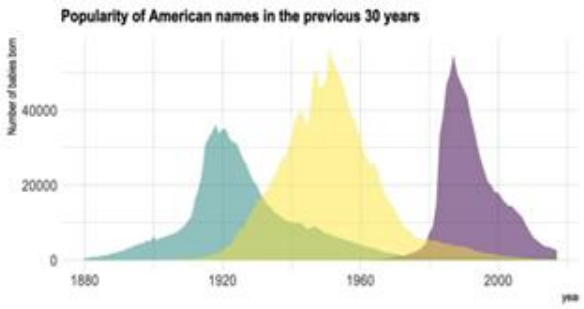
Displays the **distributions** of variables so as to avoid overlap between variables.

Line plot › Variation: the stacked area plot

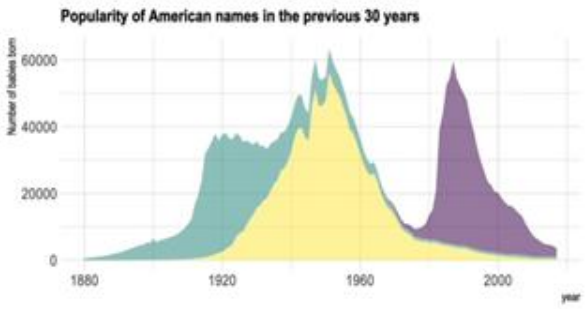
Displays the **distributions** of variables so as to avoid overlap between variables.



Line plot



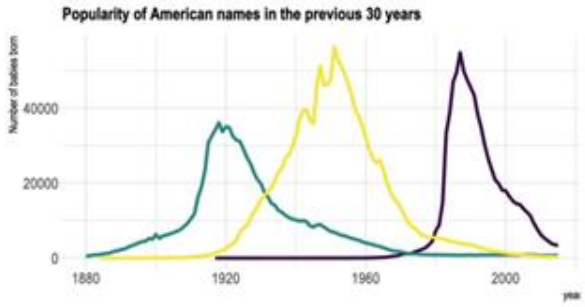
Area plot



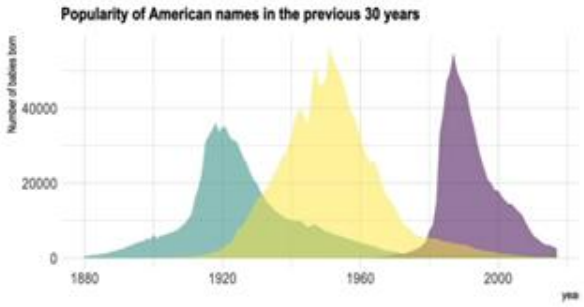
Stacked area plot

Line plot > Variation: the stacked area plot

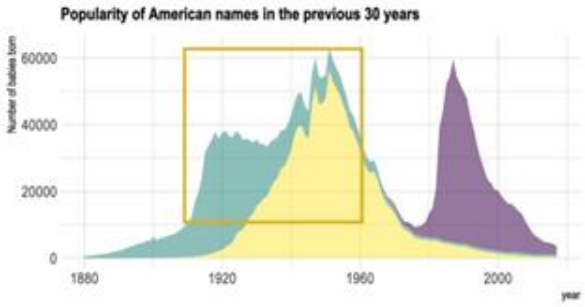
Displays the **distributions** of variables so as to avoid overlap between variables.



Line plot



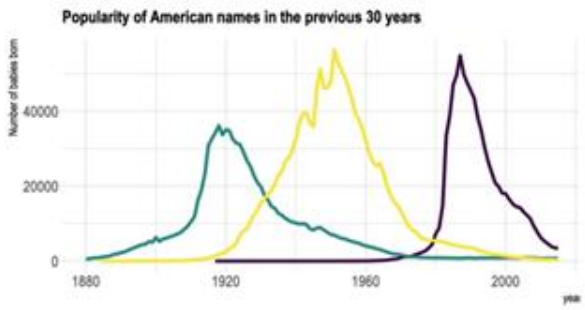
Area plot



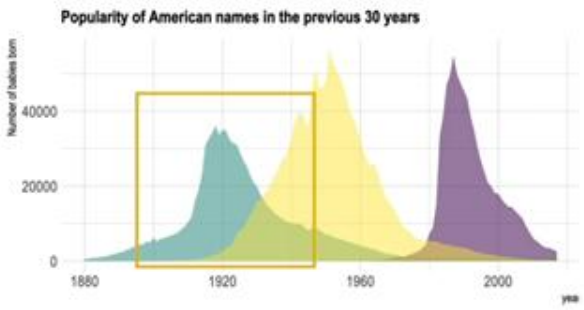
Stacked area plot

Line plot › Variation: the stacked area plot

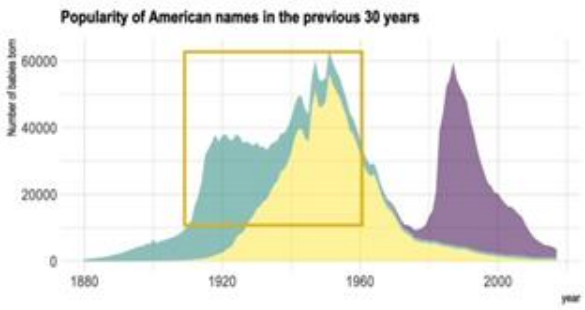
Displays the **distributions** of variables so as to avoid overlap between variables.



Line plot



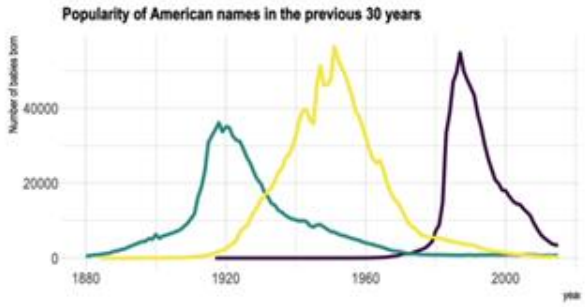
Area plot



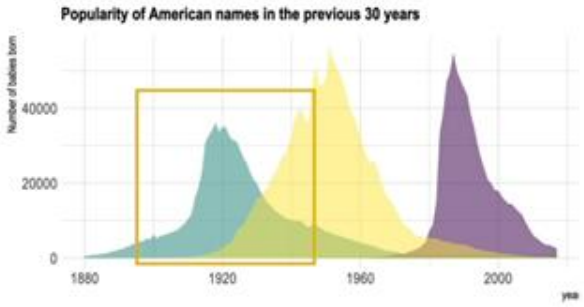
Stacked area plot

Line plot › Variation: the stacked area plot

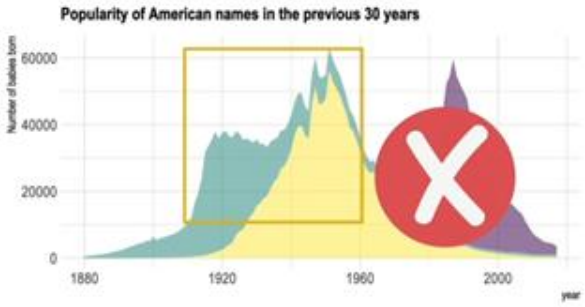
Displays the **distributions** of variables so as to avoid overlap between variables.



Line plot



Area plot



Stacked area plot

32. Plotting Relationships between variables

Scatter plot

1/28/2025 8:38:39 PM

32. Plotting Relationships between variables

Scatter plot

A scatterplot displays the relationship between two numeric variables.

1/28/2025 8:38:49 PM

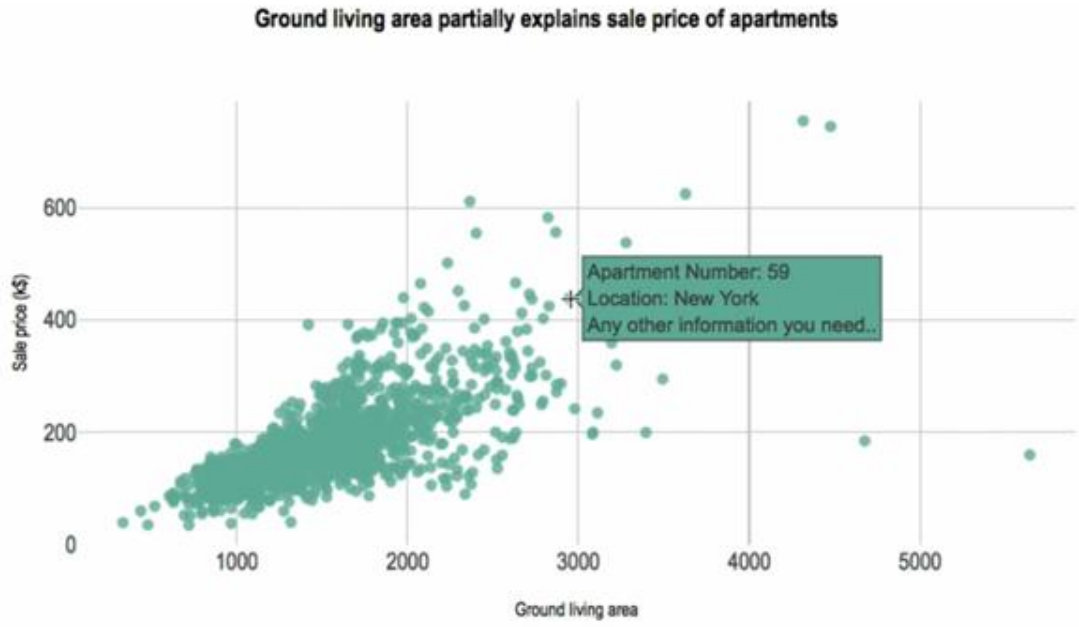
Scatter plot

A scatterplot displays the relationship between two numeric variables.

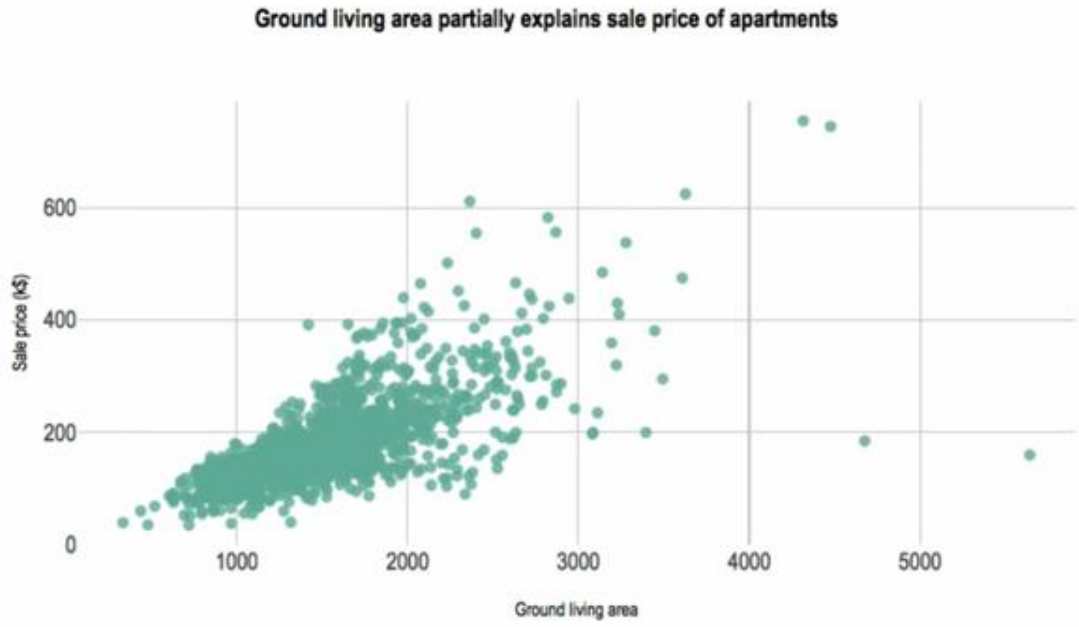
For each data point, the value of its first variable is represented on the X axis, the second on the Y axis.

Very useful to reveal patterns of your data

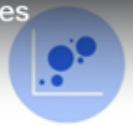
Scatter plot > Variation: interactive scatter plot



Scatter plot > Variation: interactive scatter plot

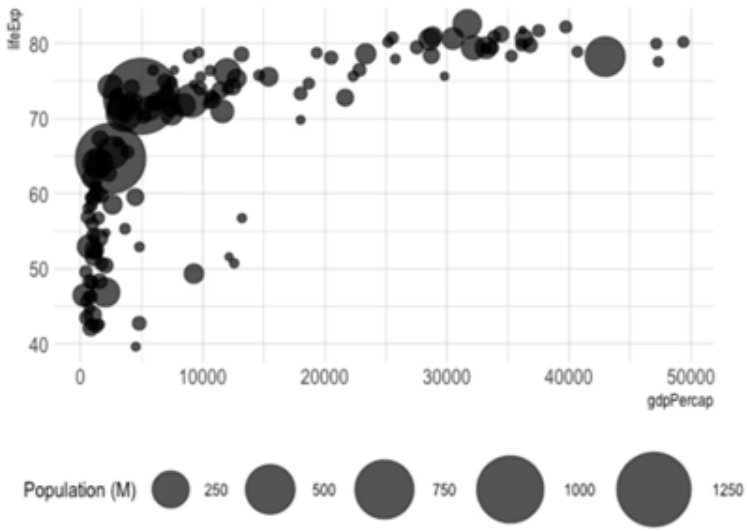


Scatter plot > Variation: the bubble plot



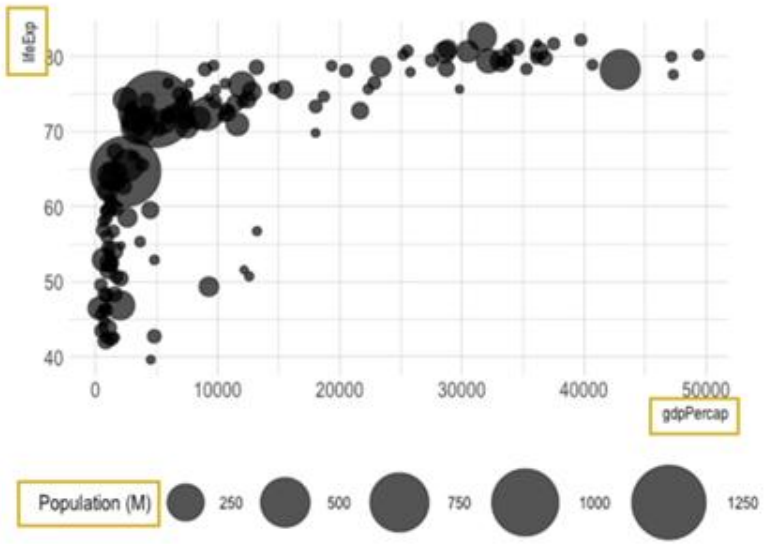
Scatter plot > Variation: the bubble plot

A bubble plot is a scatterplot where a third dimension is added:
the value of an additional numeric variable is represented through the size of the dots.



Scatter plot > Variation: the bubble plot

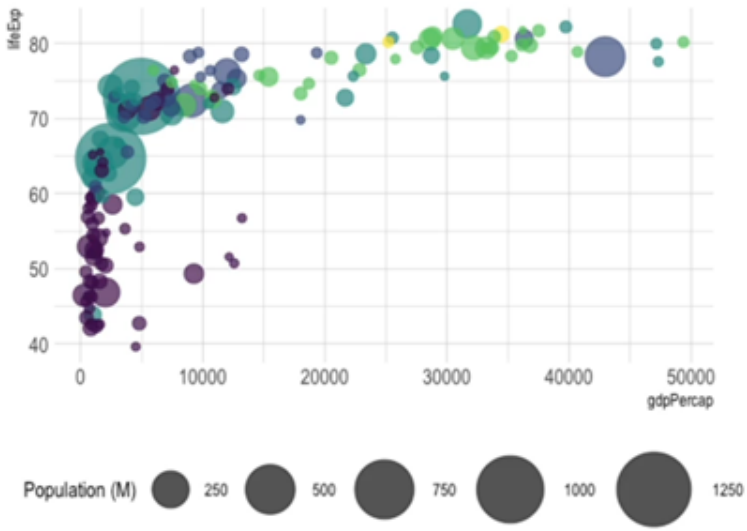
A bubble plot is a scatterplot where a third dimension is added:
the value of an additional numeric variable is represented through the size of the dots.



3rd dimension must be proportional to the **area**, not diameter.

Scatter plot › Variation: the bubble plot

A bubble plot is a scatterplot where a third dimension is added:
the value of an additional numeric variable is represented through the size of the dots.



3rd dimension must be proportional to the **area**, not diameter.

The relationship between the variable of the X and Y axis is much more obvious than the relationship with the third variable. Prioritize.

#1 #4

Adding color means adding a 4th variable.

Heat map

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

Heat map

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

Like a table but without caring for the exact values

Heat map

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

Like a table but without caring for the exact values

Very useful to reveal patterns of your data

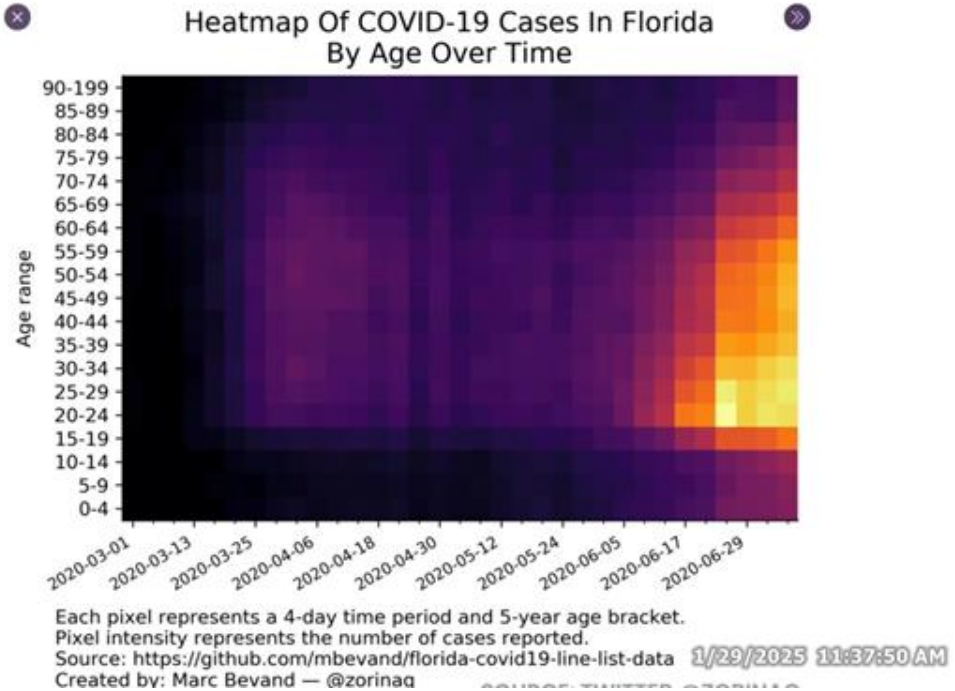
1/29/2025 11:37:30 AM

Heat map

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

Like a table but without caring for the exact values

Very useful to reveal patterns of your data

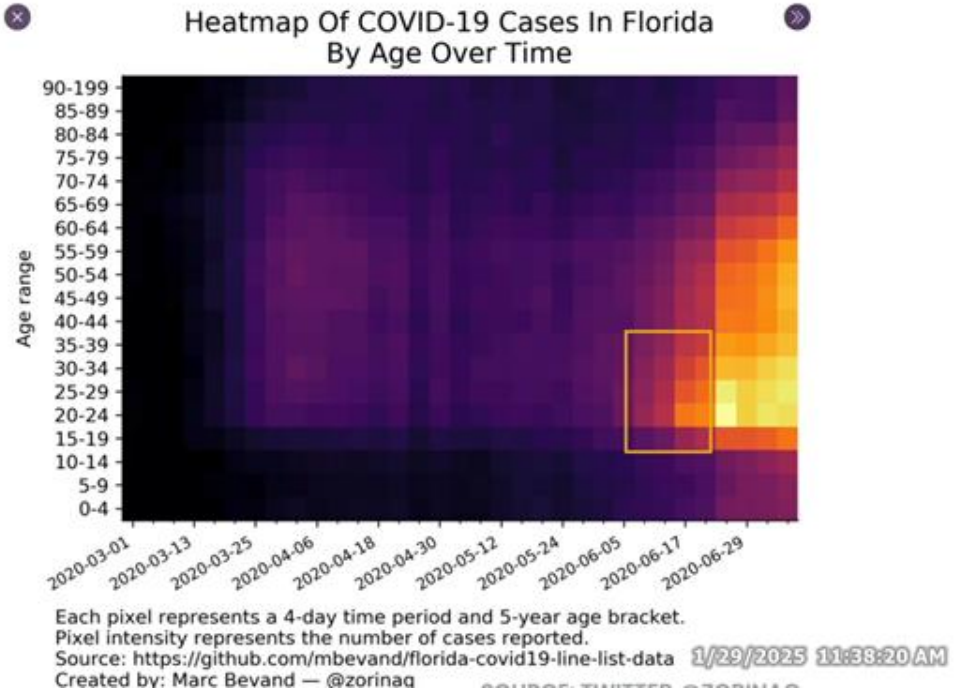


Heat map

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

Like a table but without caring for the exact values

Very useful to reveal patterns of your data

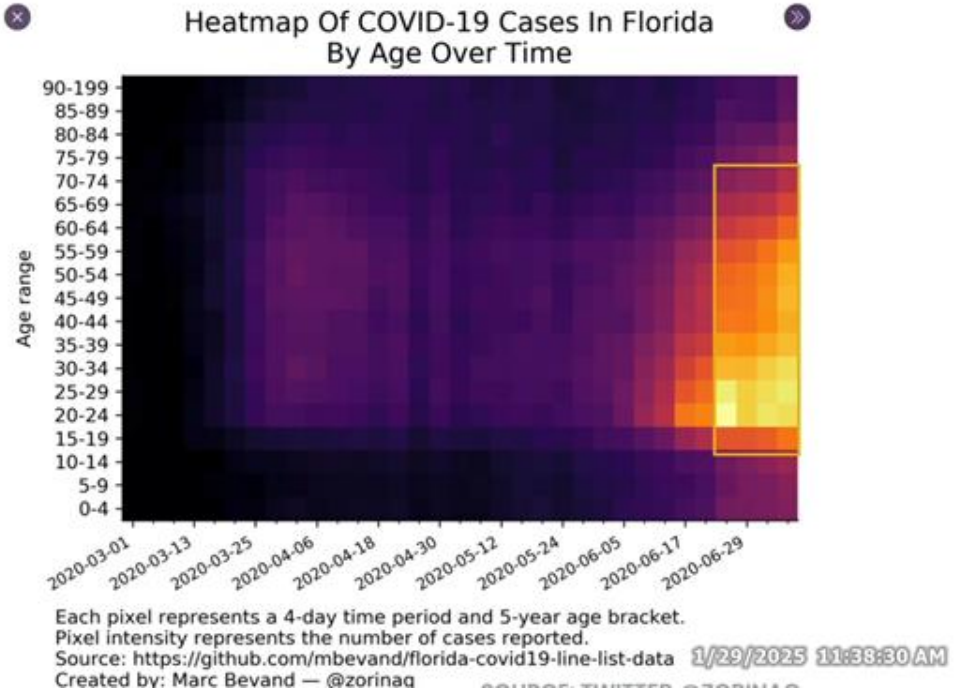


Heat map

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

Like a table but without caring for the exact values

Very useful to reveal patterns of your data



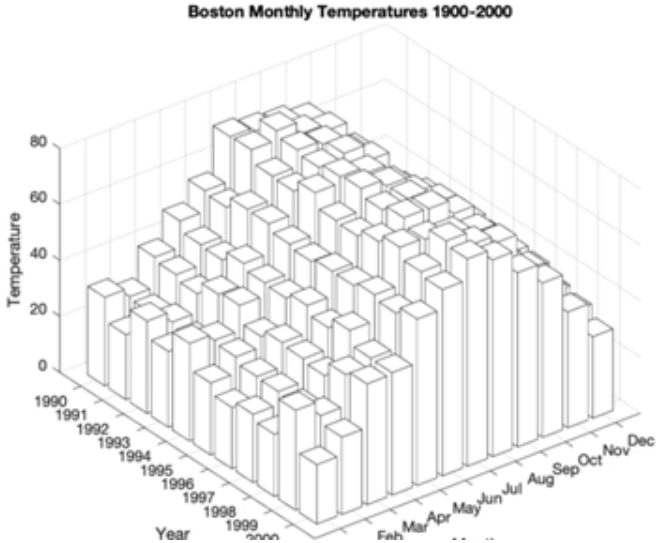
Heat map

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

Like a table but without caring for the exact values

Very useful to reveal patterns of your data

Depicts the relationship between 3 variables.



1/29/2025 11:38:40 AM

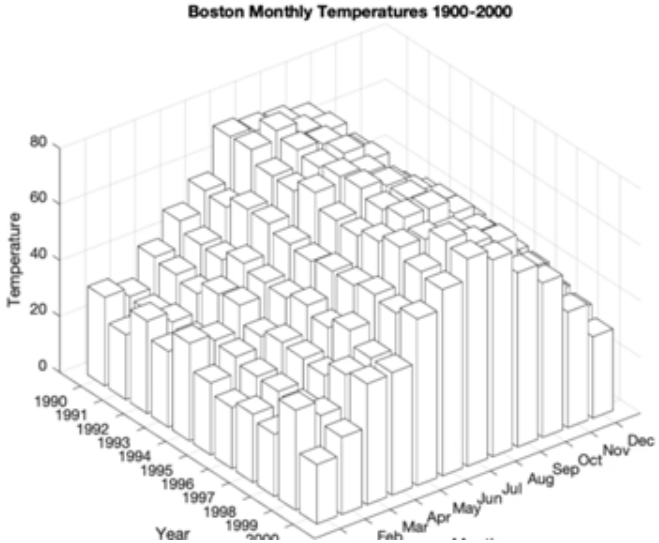
Heat map

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

Like a table but without caring for the exact values

Very useful to reveal patterns of your data

Depicts the relationship between 3 variables.



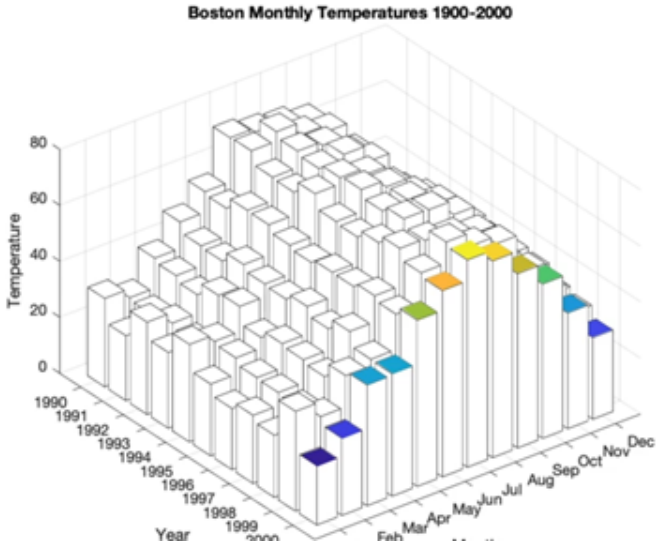
Heat map

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

Like a table but without caring for the exact values

Very useful to reveal patterns of your data

Depicts the relationship between 3 variables.



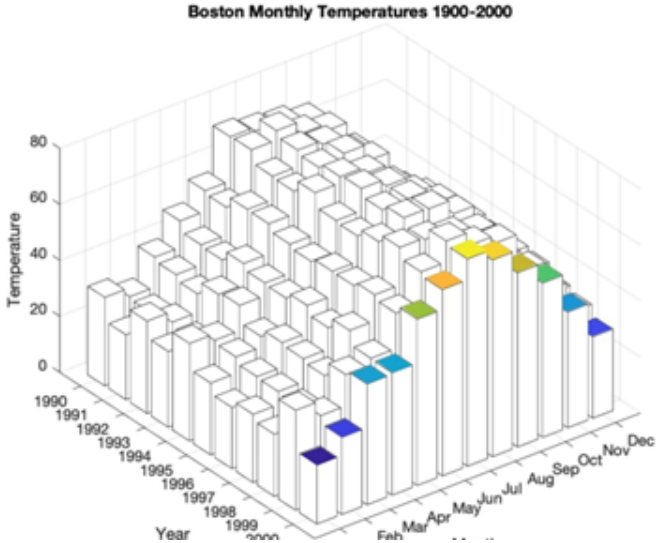
Heat map

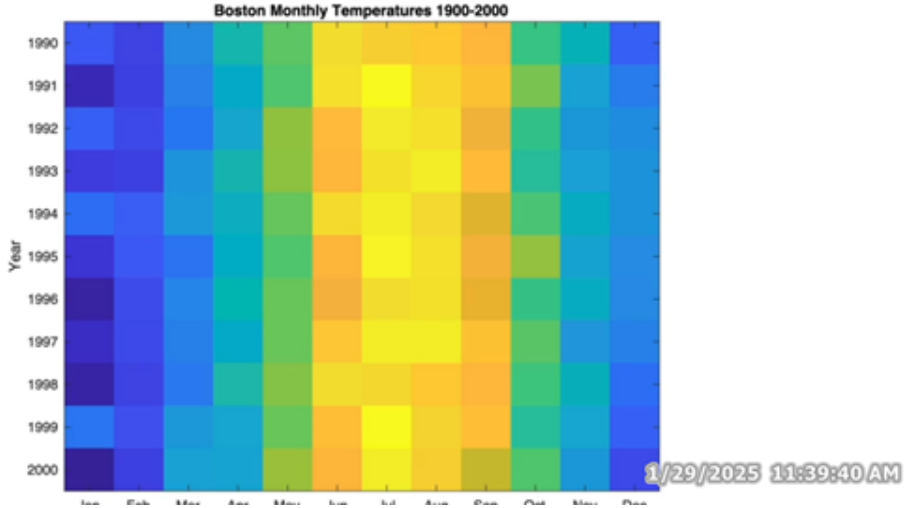
A heatmap is a graphical representation of data where the values contained in a matrix are represented as colors.

Like a table but without caring for the exact values

Very useful to reveal patterns of your data

Depicts the relationship between 3 variables.





Relationships between variables

Relationships between variables

Relationships between variables

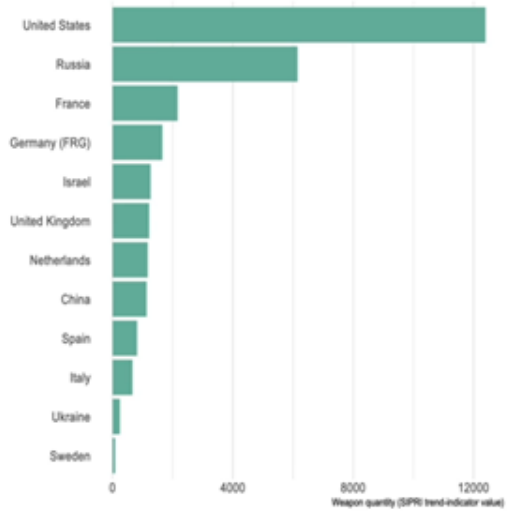

1/29/2025 11:40:21 AM

33- Plotting Ranking

Plotting rankings

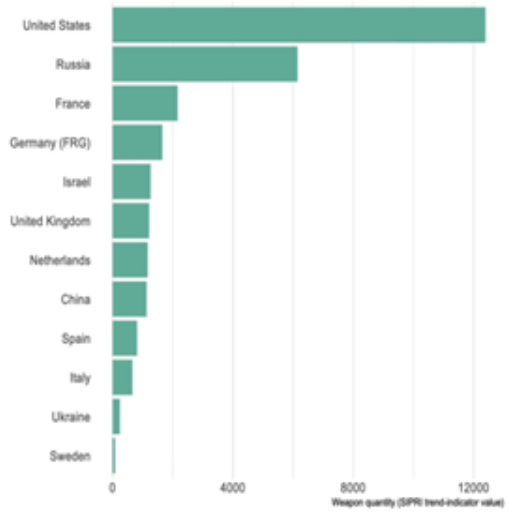
Barplot

A barplot shows the relationship between a numeric and a categoric variable.



Barplot

A barplot shows the relationship between a numeric and a categoric variable.

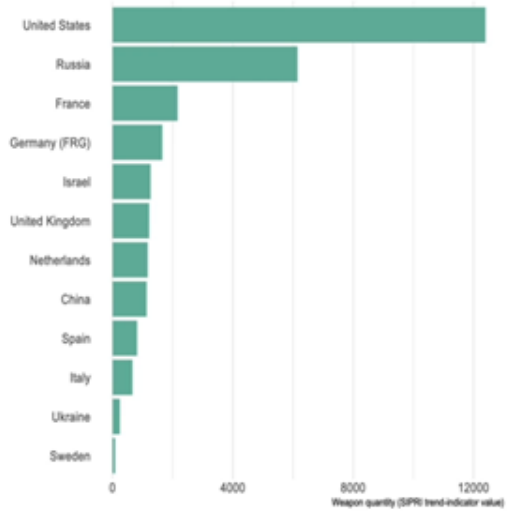


Barplot

A barplot shows the relationship between a numeric and a categoric variable.

Each categoric variable = bar

Numeric value = length of the bar

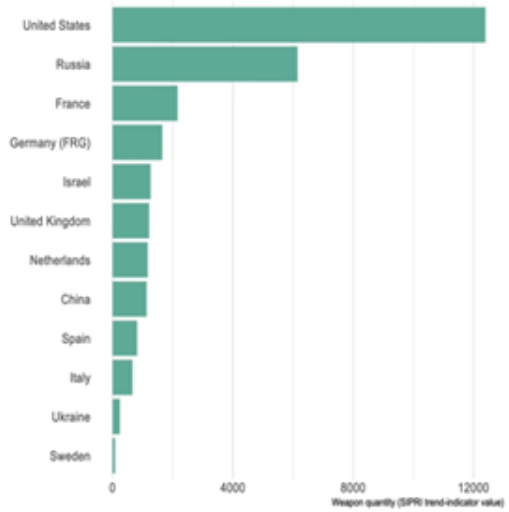


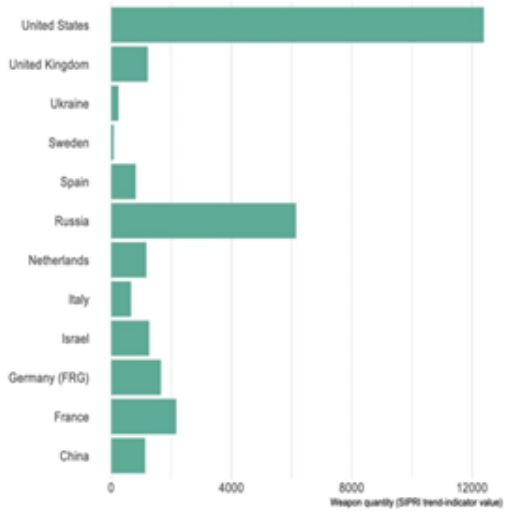
Barplot

A barplot shows the relationship between a numeric and a categoric variable.

Each categoric variable = bar

Numeric value = length of the bar





Order your barplots

1/29/2025 11:41:21 AM

Barplot

A barplot shows the relationship between a numeric and a categoric variable.

Each categoric variable = bar

Numeric value = length of the bar

Order your barplots

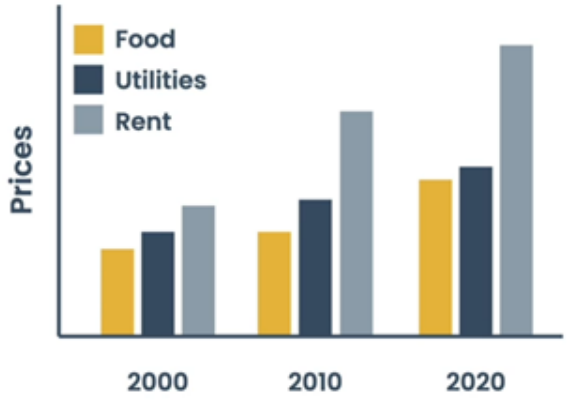
If several observations per categorical variable, use a grouped bar chart

Barplot

A barplot shows the relationship between a numeric and a categoric variable.

Each categoric variable = bar

Numeric value = length of the bar



Grouped barplot

Order your barplots

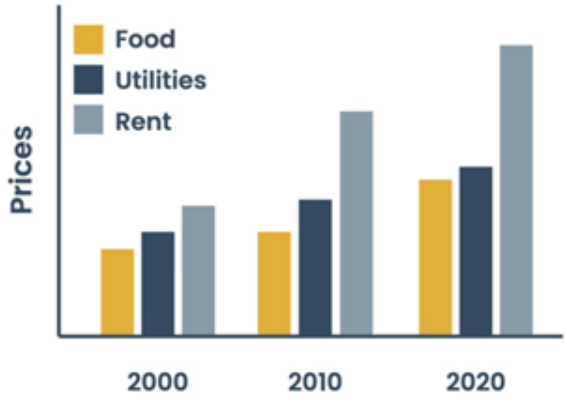
If several observations per categorical variable, use a grouped bar chart

Barplot

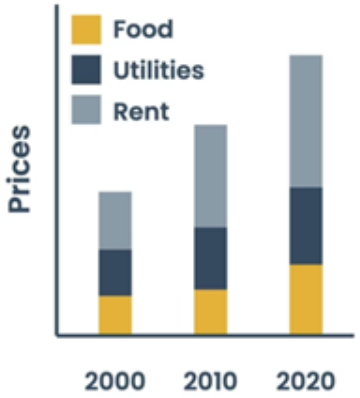
A barplot shows the relationship between a numeric and a categoric variable.

Each categoric variable = bar

Numeric value = length of the bar



Grouped barplot



Stacked barplot

Order your barplots

If several observations per categorical variable, use a grouped bar chart

Barplot > Variation: the lollipop plot

Barplot > Variation: the lollipop plot

The bar is replaced with a line ending with a dot.

Barplot > Variation: the lollipop plot

The bar is replaced with a line ending with a dot.

Good idea because

Good for dense barplots with similar values

Barplot > Variation: the lollipop plot

The bar is replaced with a line ending with a dot.

Good idea because

- Good for dense barplots with similar values
- Principle of minimum ink
- It is clearer that we do not have to decode area

1/29/2025 11:42:41 AM

Barplot › Variation: the lollipop plot

The bar is replaced with a line ending with a dot.

Good idea because

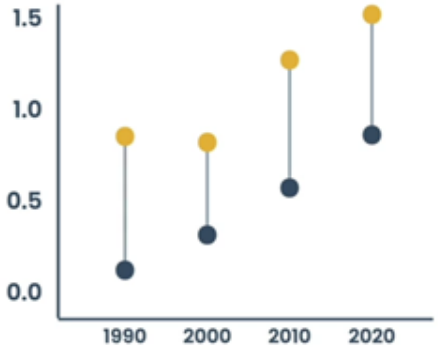
- Good for dense barplots with similar values
- Principle of minimum ink
- It is clearer that we do not have to decode area
- Decode position in a common scale

#1 #3

1/29/2025 11:42:51 AM

Barplot › Variation: the lollipop plot

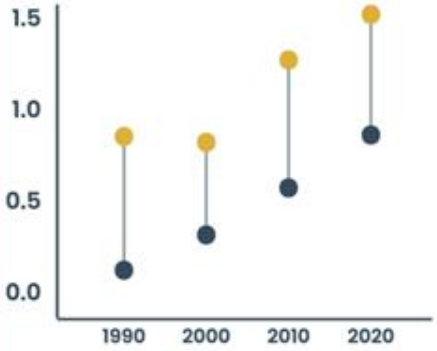
The Cleveland dot plot -or- dumbbell dot plot



1/29/2025 11:43:01 AM

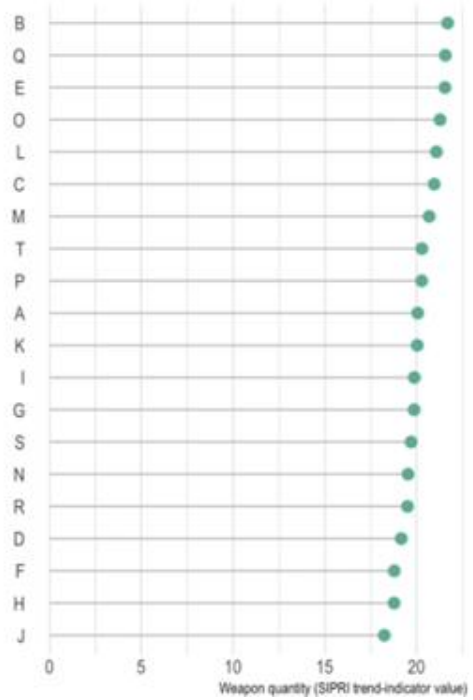
Barplot › Variation: the lollipop plot

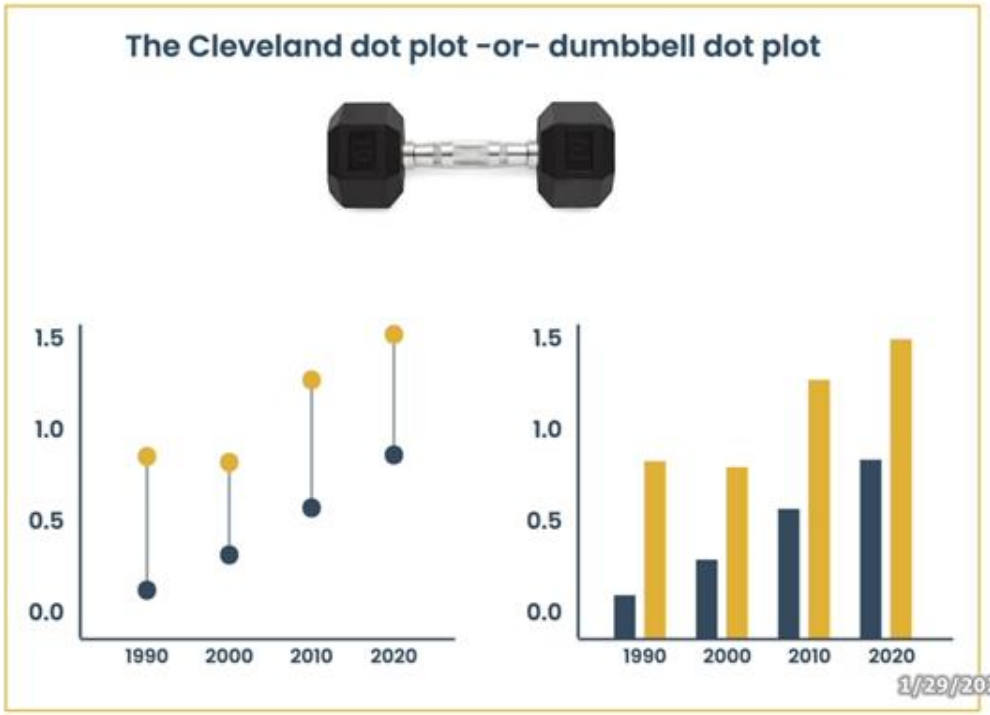
The Cleveland dot plot -or- dumbbell dot plot



1/29/2025 11:43:11 AM

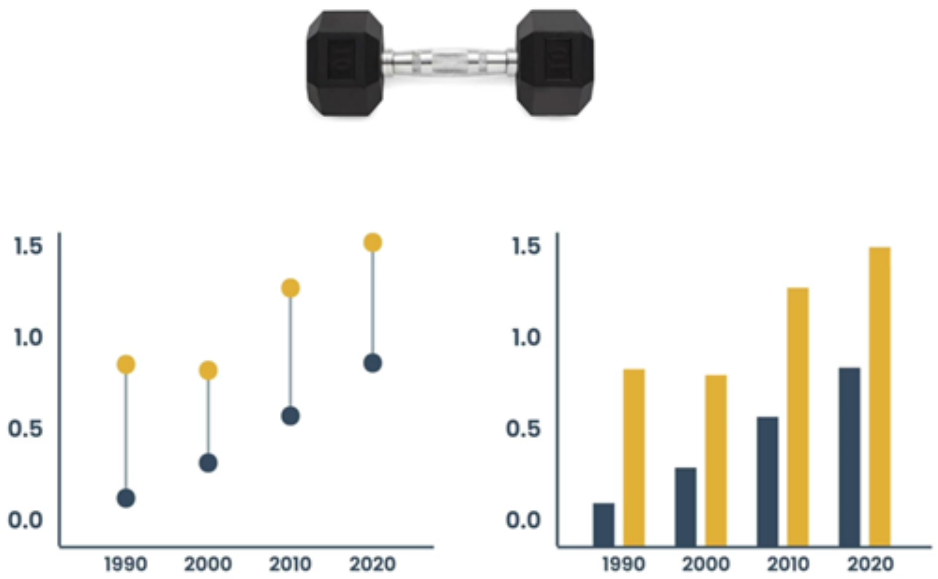
Barplot › Variation: the lollipop plot





Barplot > Variation: the lollipop plot

The Cleveland dot plot -or- dumbbell dot plot



Total Revenue by City and Gender

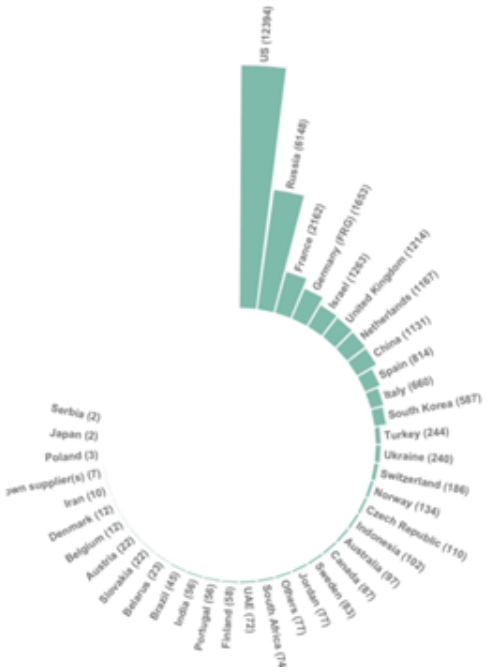
Out of 23 cities, eight locations experience a 20% or greater difference in revenue generated by males versus females. Hidalgo experiences the greatest difference with females generating 86% more revenue than males.

Barplot >  Variation: the circular barplot

Barplot > Variation: the circular barplot

A bar plot where the x-axis is wrapped around a circle.

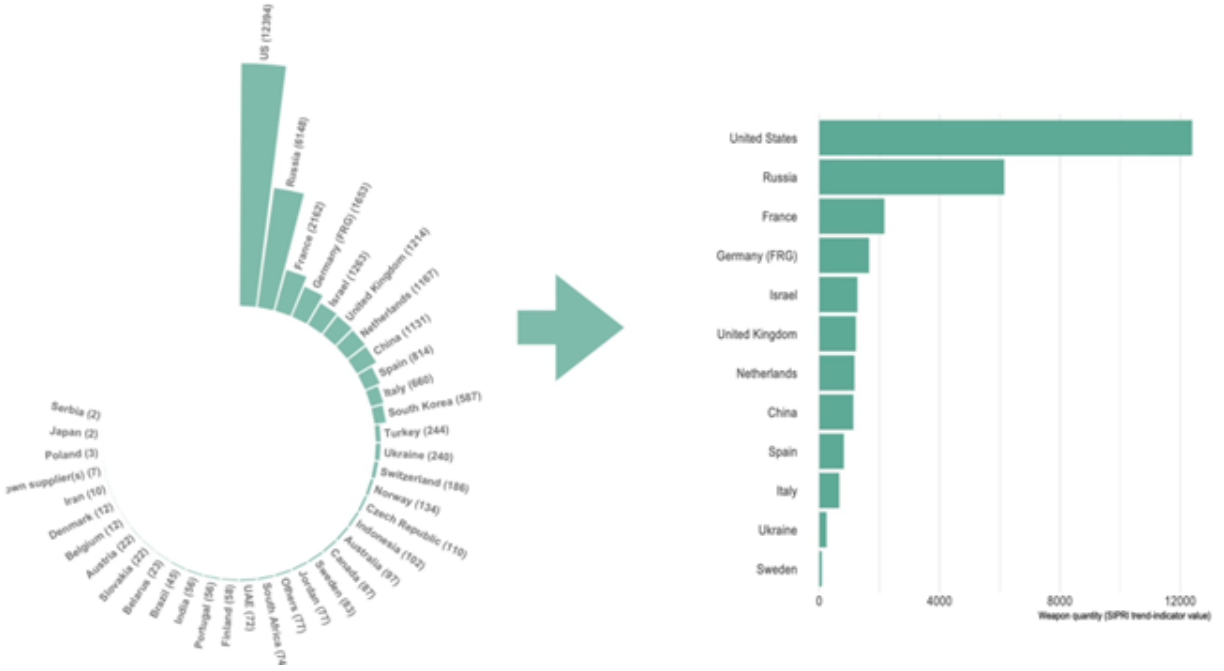
Bad idea: the user has to mentally "undo" the coordinates transformation.



Barplot › Variation: the circular barplot

A bar plot where the x-axis is wrapped around a circle.

Bad idea: the user has to mentally "undo" the coordinates transformation.



#2

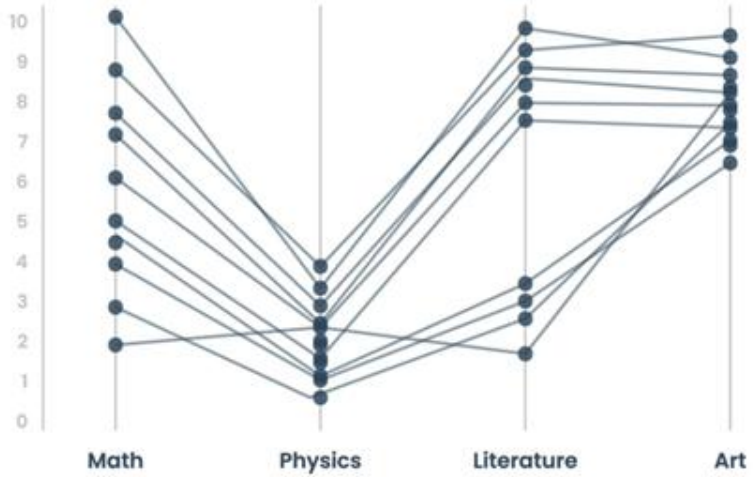
#3

#5

1/29/2025 11:44:12 AM

Parallel plot

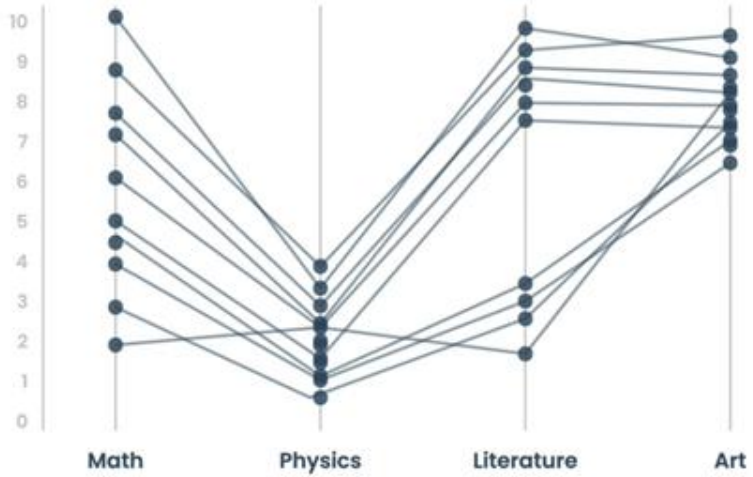
Useful when we want to compare multivariate data.



Parallel plot

Useful when we want to compare multivariate data.

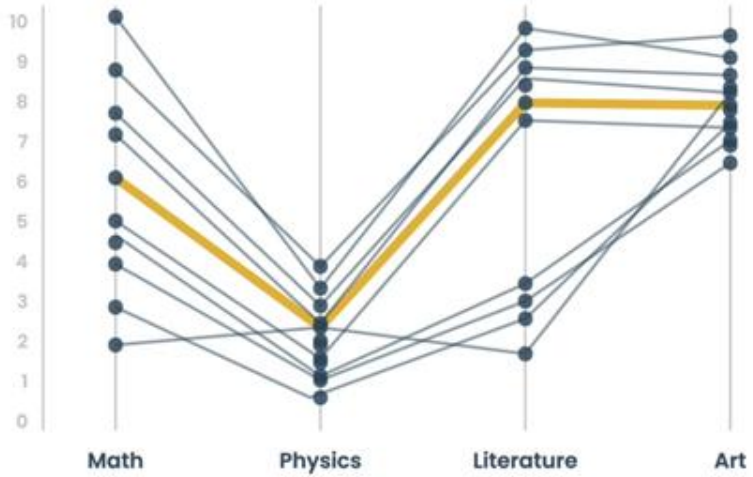
Each vertical bar represents a variable and usually has its own scale. (The units can even be different).
Values are plotted as series of lines connected across each axis.



Parallel plot

Useful when we want to compare multivariate data.

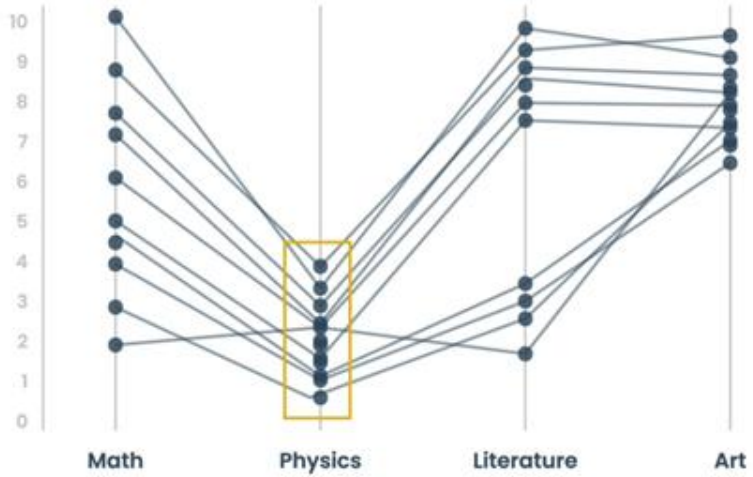
Each vertical bar represents a variable and usually has its own scale. (The units can even be different).
Values are plotted as series of lines connected across each axis.



Parallel plot

Useful when we want to compare multivariate data.

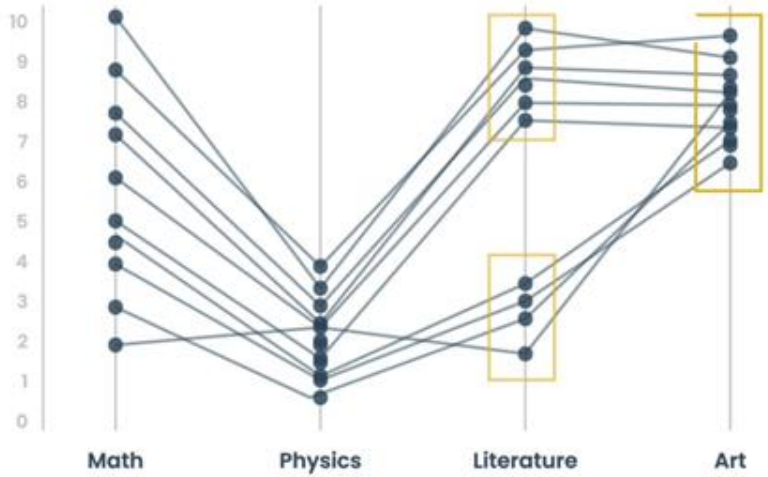
Each vertical bar represents a variable and usually has its own scale. (The units can even be different).
Values are plotted as series of lines connected across each axis.



Parallel plot

Useful when we want to compare multivariate data.

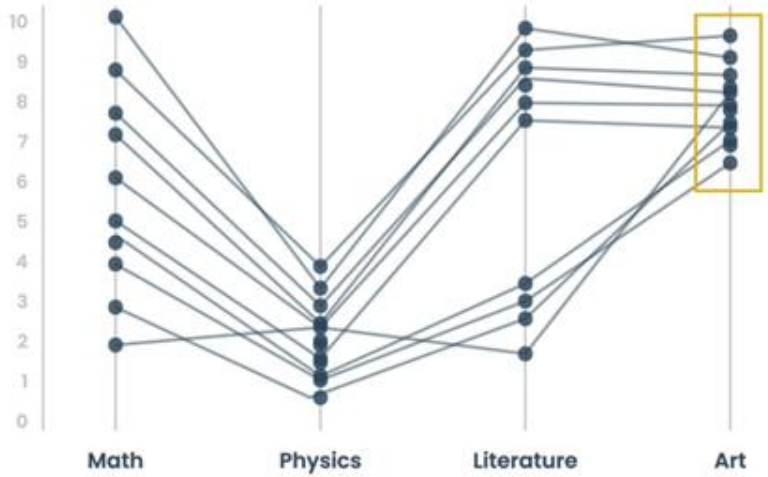
Each vertical bar represents a variable and usually has its own scale. (The units can even be different).
Values are plotted as series of lines connected across each axis.



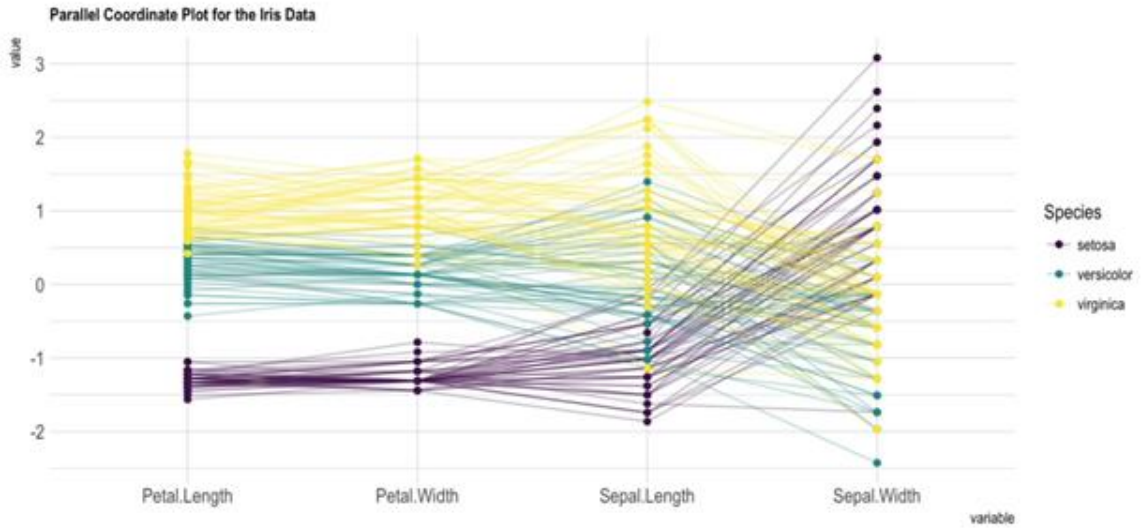
Parallel plot

Useful when we want to compare multivariate data.

Each vertical bar represents a variable and usually has its own scale. (The units can even be different).
Values are plotted as series of lines connected across each axis.



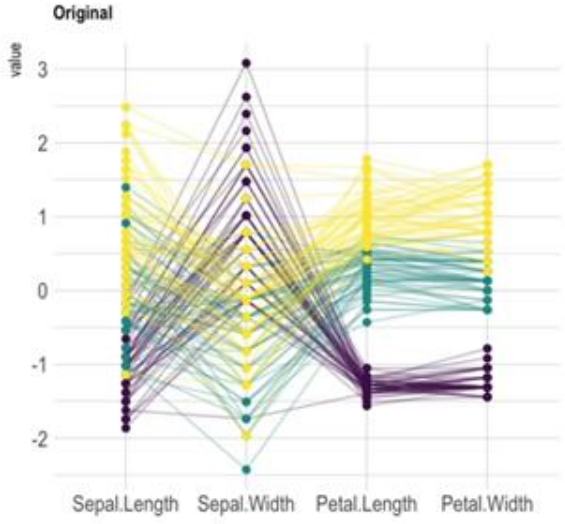
Parallel plot

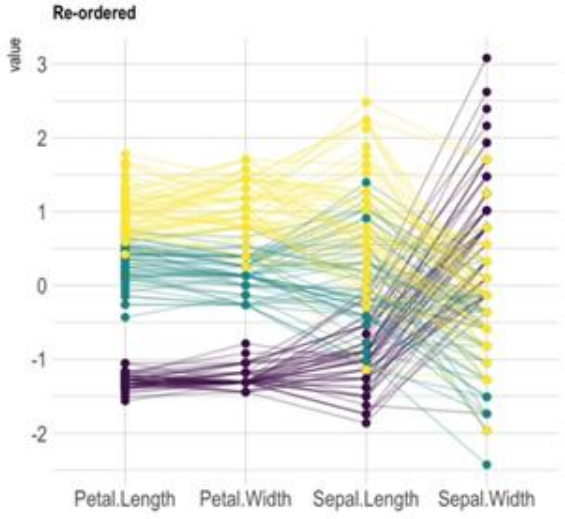


To achieve a good result:

Use thin lines if there are many samples

Parallel plot



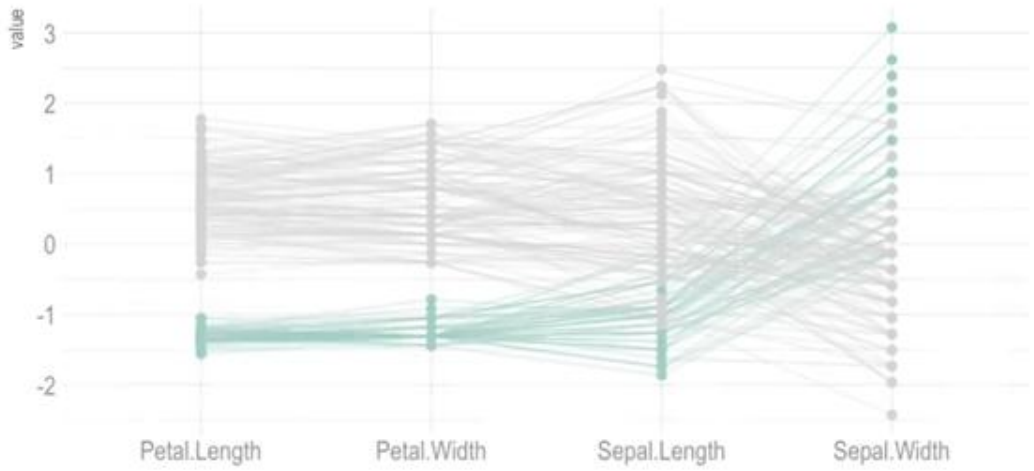


To achieve a good result:

Use thin lines if there are many samples

Sort the variables on the X axis

Parallel plot



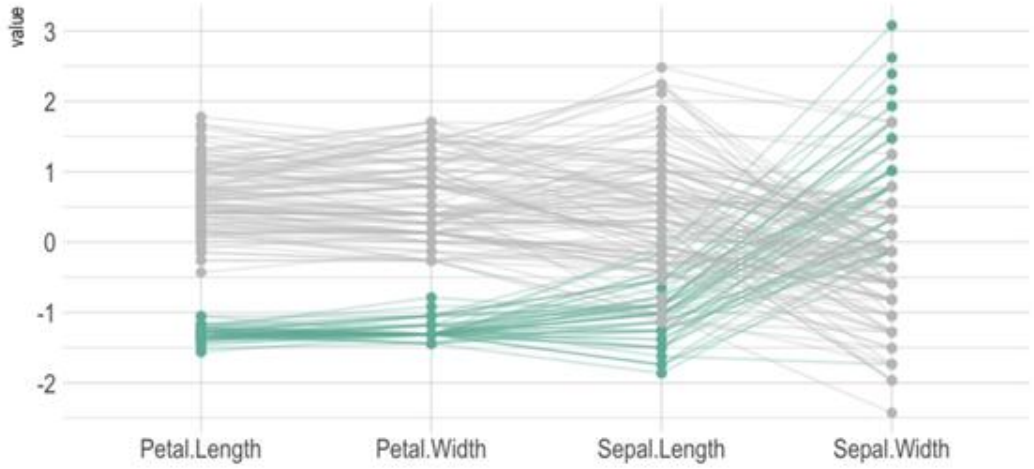
To achieve a good result:

Use thin lines if there are many samples

Sort the variables on the X axis

If there are too many lines, use highlighting of groups or individual samples

Parallel plot



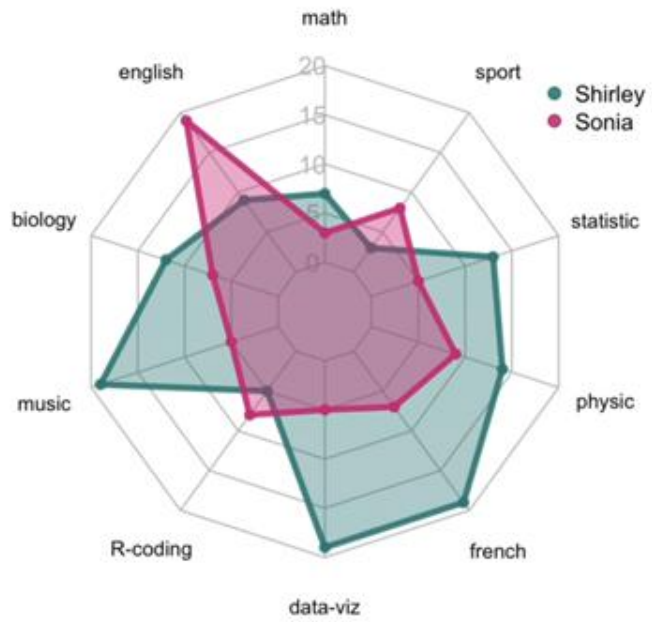
To achieve a good result:

Use thin lines if there are many samples

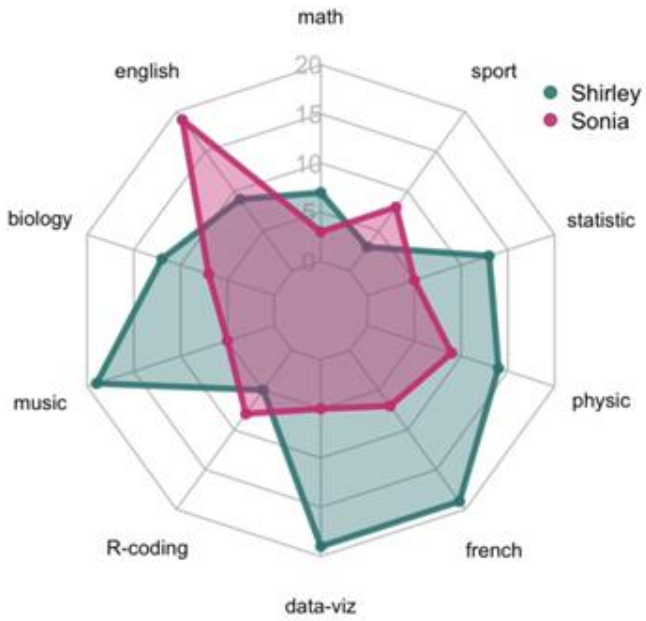
Sort the variables on the X axis

If there are too many lines, use highlighting of groups or individual samples

Parallel plot > Variation: the radar/spider plot



Parallel plot > Variation: the radar/spider plot



A bad idea because

The circular layout makes it more difficult to read

Parallel plot > Variation: the radar/spider plot

A bad idea because

The circular layout makes it more difficult to read

The reader focuses too much on areas

The shapes of the areas vary depending on the ordering of the variables

Wordcloud

Wordcloud

1/29/2025 11:47:53 AM

Wordcloud

A horrible idea b

1/29/2025 11:48:03 AM

Wordcloud

A horrible idea because

Area is hard to decode

Longer words appear bigger

1/29/2025 11:48:13 AM

Wordcloud

A horrible idea because

Area is hard to decode

Longer words appear bigger

Different angles make the perception of words different

1/29/2025 11:48:23 AM

Wordcloud

A horrible idea because

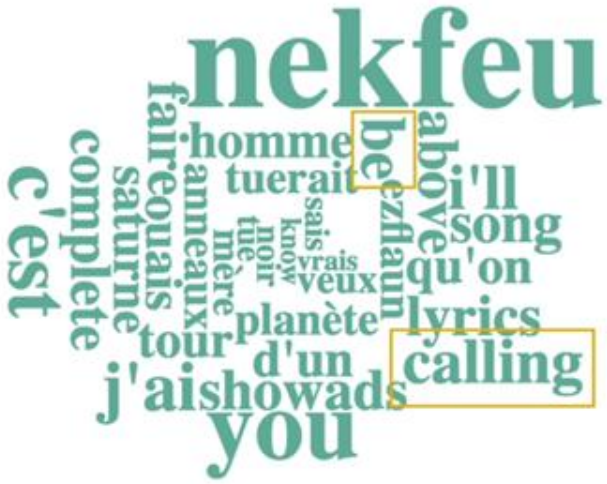
Area is hard to decode

Longer words appear bigger

Different angles make the perception of words different

1/29/2025 11:48:33 AM

Wordcloud



A horrible idea because

Area is hard to decode

Longer words appear bigger

Different angles make the perception of words different

1/29/2025 11:48:43 AM

Wordcloud

A horrible idea because

Area is hard to decode

Longer words appear bigger

Different angles make the perception of words different

1/29/2025 11:48:53 AM

Plotting rankings

Plotting rankings

34- Comparing Part to Whole

Comparing part to whole

1/29/2025 11:49:33 AM

Pie chart

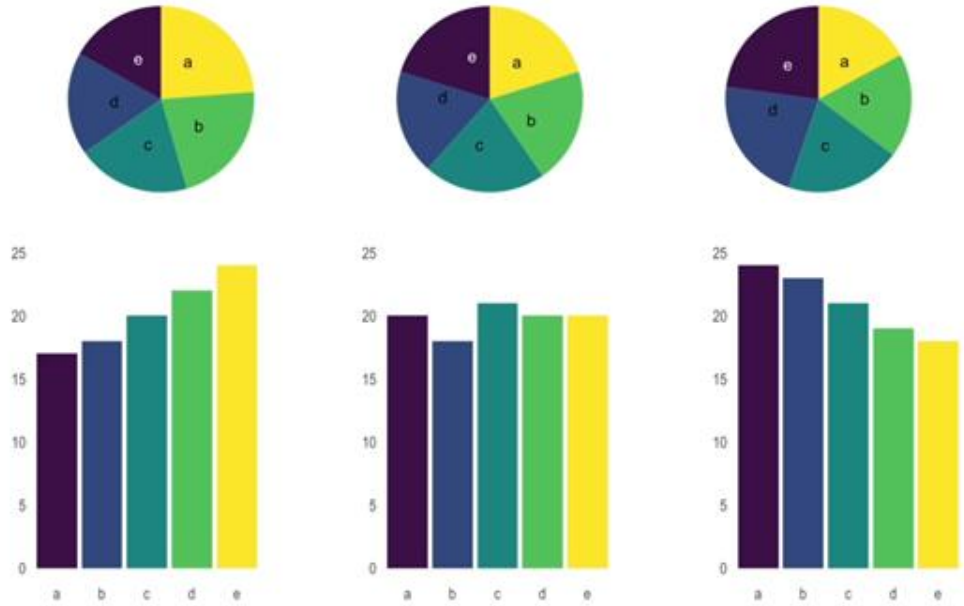
1/29/2025 11:49:43 AM

Pie chart

A pie chart is a circle divided into sectors that each represent a proportion of the whole.
It is often used to show proportion, where the sum of the sectors equal 100%.

Pie chart

A pie chart is a circle divided into sectors that each represent a proportion of the whole.
It is often used to show proportion, where the sum of the sectors equal 100%.



Pie chart

A pie chart is a circle divided into sectors that each represent a proportion of the whole.
It is often used to show proportion, where the sum of the sectors equal 100%.

Pie chart

A pie chart is a circle divided into sectors that each represent a proportion of the whole.
It is often used to show proportion, where the sum of the sectors equal 100%.

The only acceptable use of a pie chart

Pie chart

A pie chart is a circle divided into sectors that each represent a proportion of the whole.
It is often used to show proportion, where the sum of the sectors equal 100%.

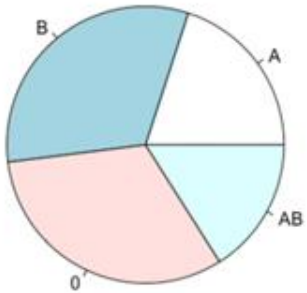
The only acceptable use of a pie chart

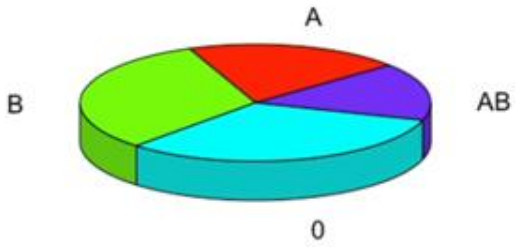
If you're comparing one part to the whole = only 2 slices

The exact values are irrelevant

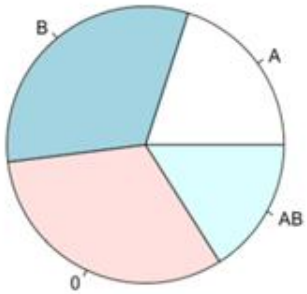
The slices are easily compared at a glance

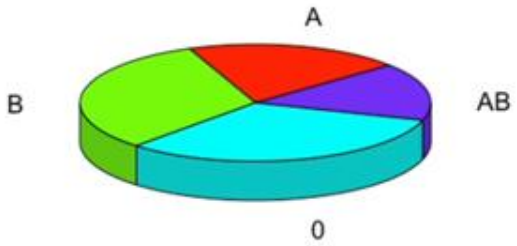
Pie chart

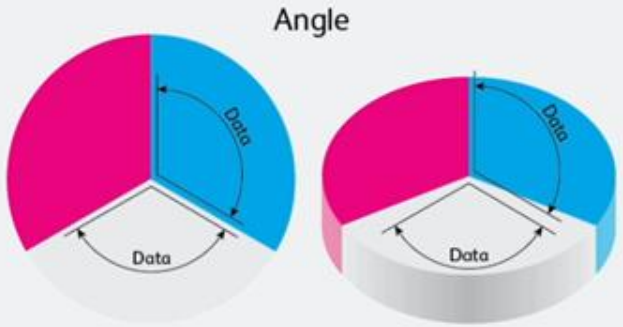




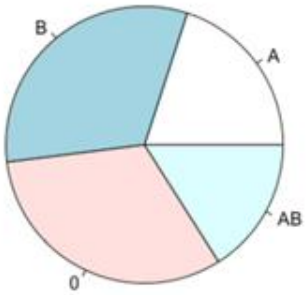
Pie chart

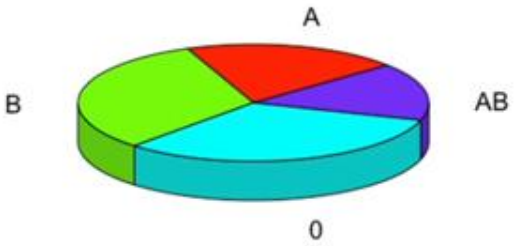


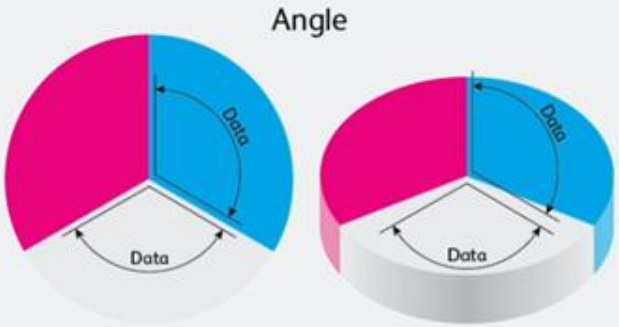




Pie chart

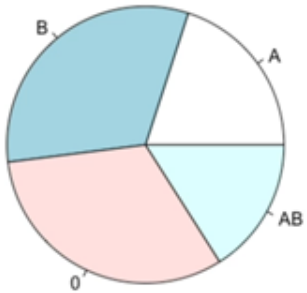

B
A
AB
0

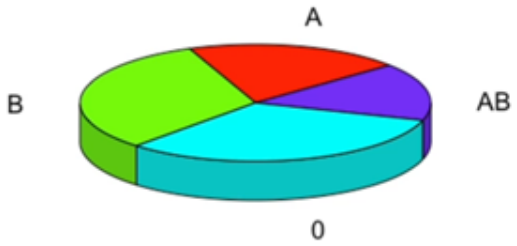

A
B
AB
0

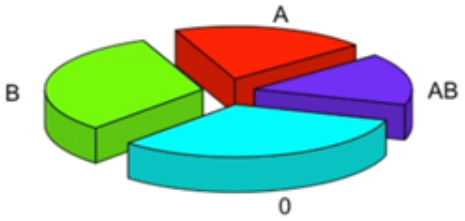

Angle
Data
Data
Data
Data


U.S. SmartPhone Marketshare
RIM
Apple
Palm
Motorola
Nokia
Other
21.2%
39.0%
3.1%
7.4%
9.8%
19.5%
1/29/2025 11:51:33 AM

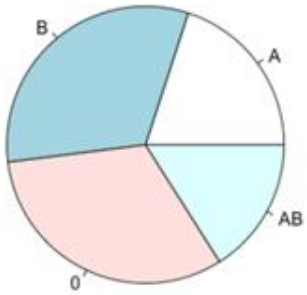
Pie chart

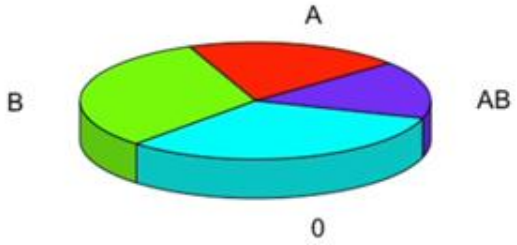


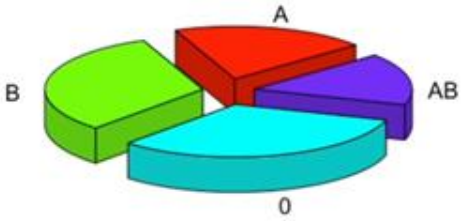




Pie chart





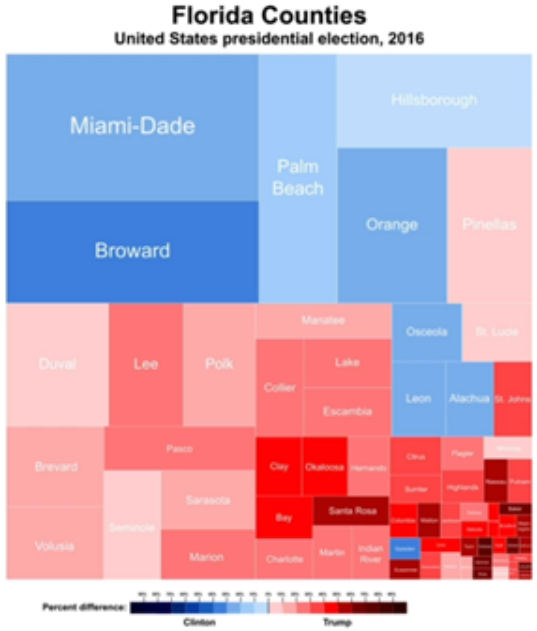


Used
Available

A doughnut plot

Treemap

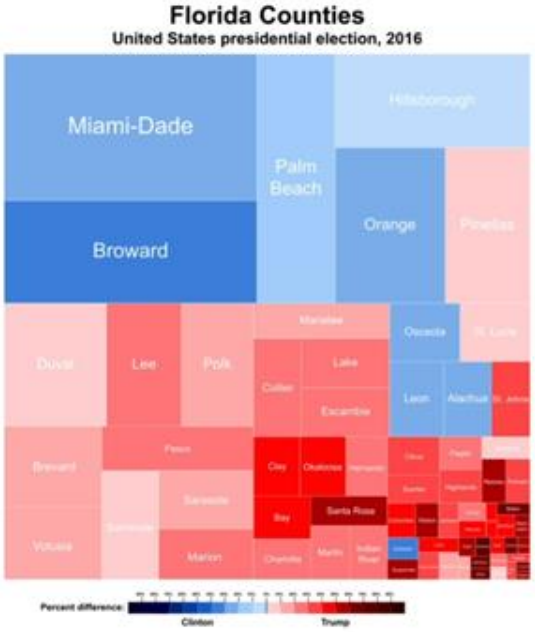
A Treemap displays hierarchical data as a set of nested rectangles.
Each group is represented by a rectangle whose area is proportional to its value.



1/29/2025 11:52:33 AM

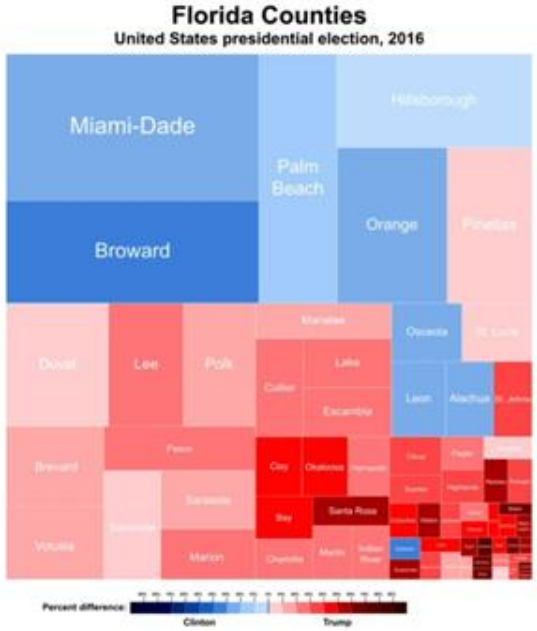
Treemap

A Treemap displays hierarchical data as a set of nested rectangles.
Each group is represented by a rectangle whose area is proportional to its value.



Treemap

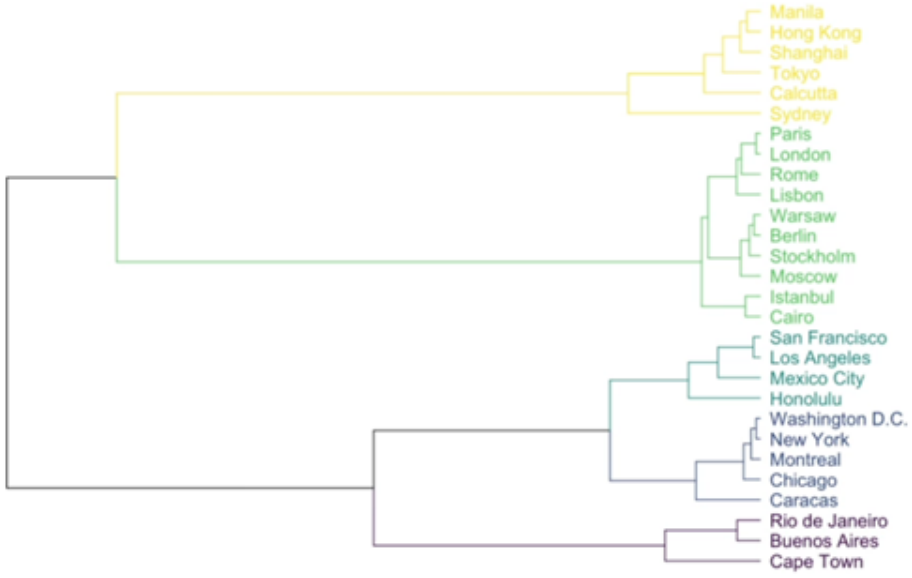
A Treemap displays hierarchical data as a set of nested rectangles.
Each group is represented by a rectangle whose area is proportional to its value.



1/29/2025 11:53:03 AM

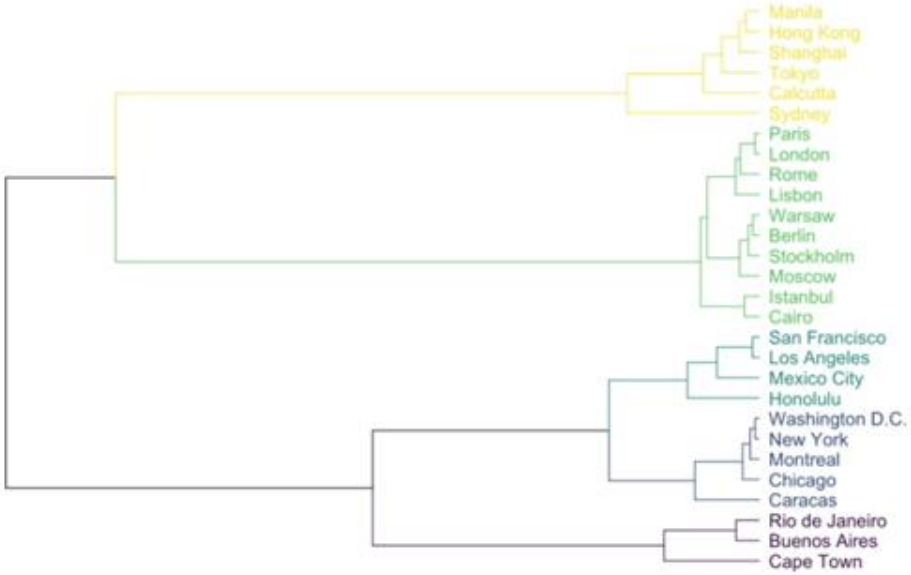
Dendrogram

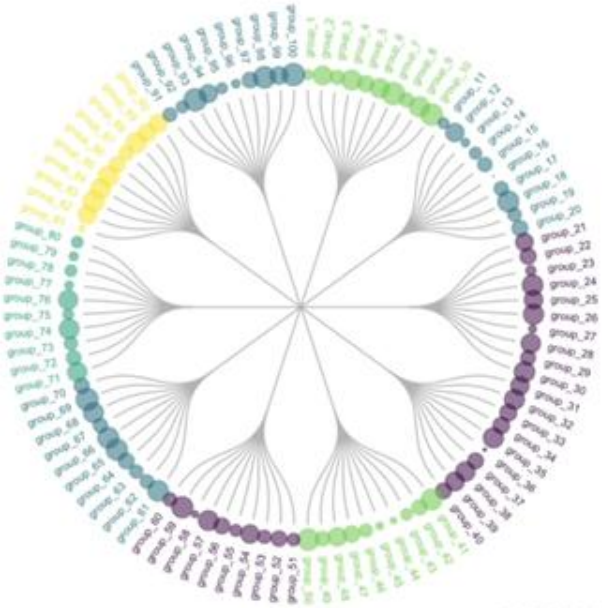
A dendrogram is a diagram that depicts hierarchical relationships between categories.



Dendrogram

A dendrogram is a diagram that depicts hierarchical relationships between categories.



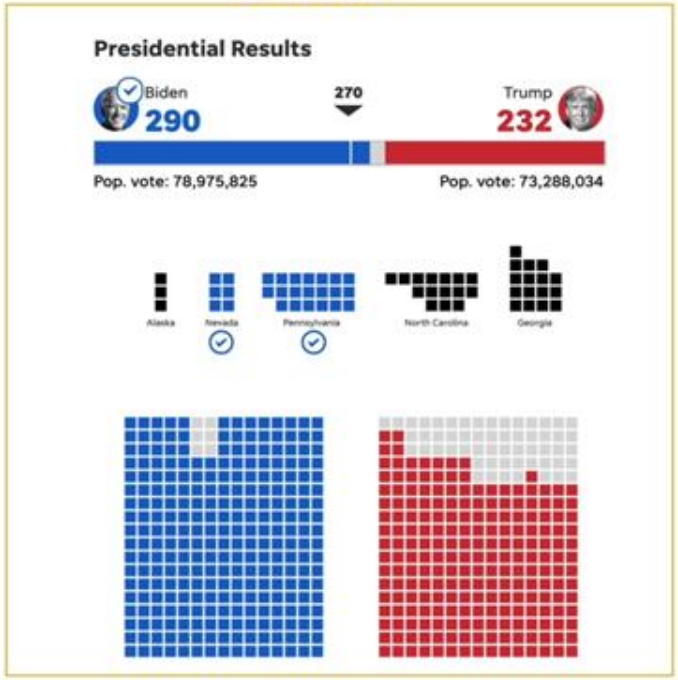


1/29/2025 11:53:34 AM

Mixed modern approaches > The 2020 US Election

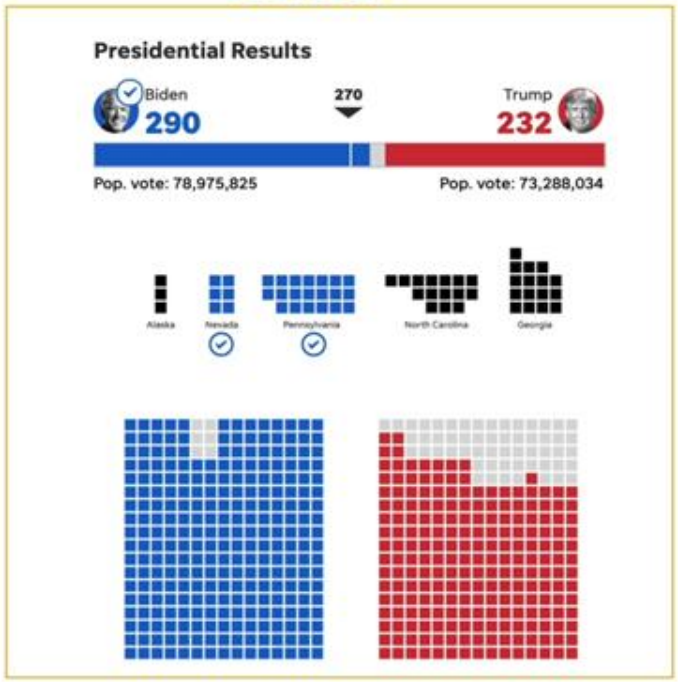
Mixed modern approaches › The 2020 US Election

USA TODAY



Mixed modern approaches › The 2020 US Election

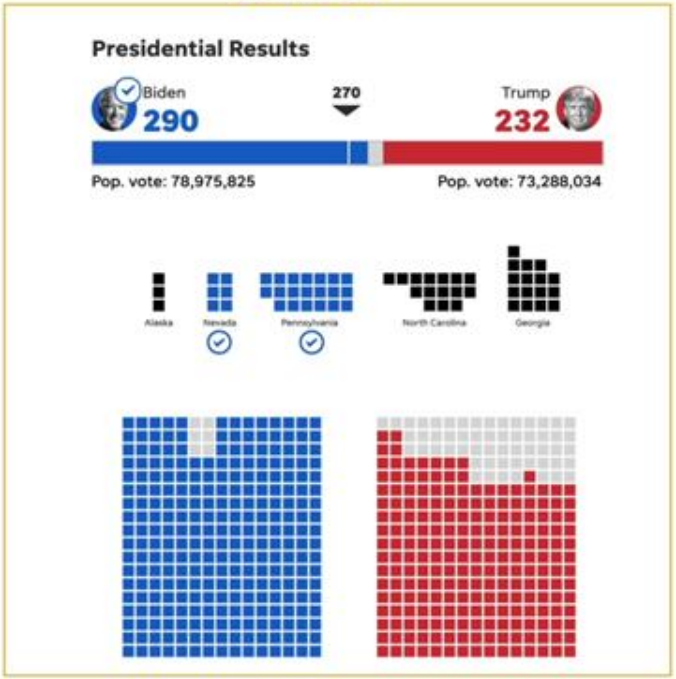
USA TODAY



BBC

Mixed modern approaches › The 2020 US Election

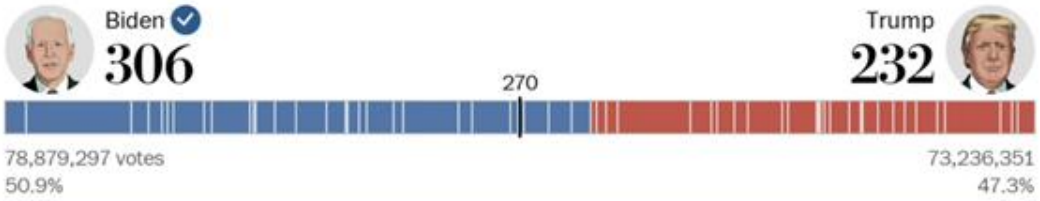
USA TODAY



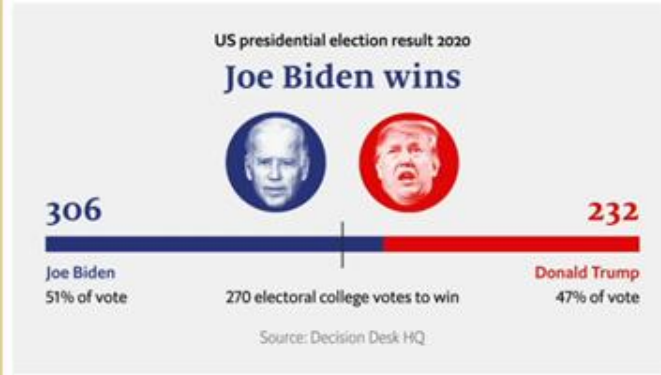
BBC

Mixed modern approaches > The 2020 US Election

THE WASHINGTON POST



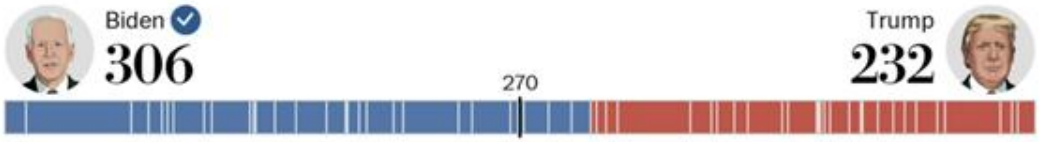
THE ECONOMIST



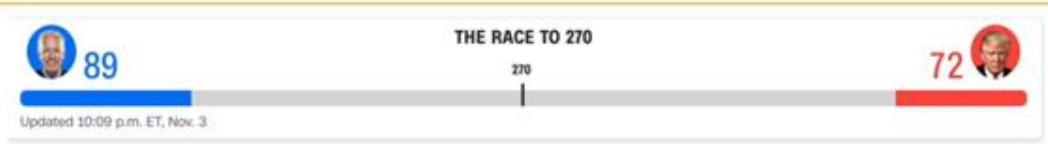
1/29/2025 11:55:34 AM

Mixed modern approaches > The 2020 US Election

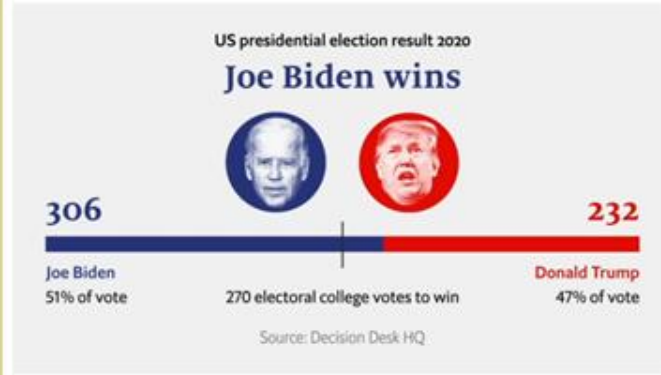
THE WASHINGTON POST



CNN



THE ECONOMIST



Comparing part to whole

Comparing part to whole

36- Plotting Spatial Data: Maps

Drawing maps

Background map

The starting point of any geospatial visualization.

(1) Find spatial data

Shape file (*.shp)
GeoJSON
R Libraries
Google/OpenStreet map

Background map

The starting point of any geospatial visualization.

(1) Find spatial data

Shape file (*.shp)
GeoJSON
R Libraries
Google/OpenStreet map

(2) Plot it

maps package in R
ggplot2 package in R
ggmap package in R
rgdal(), rgeos()

Background map

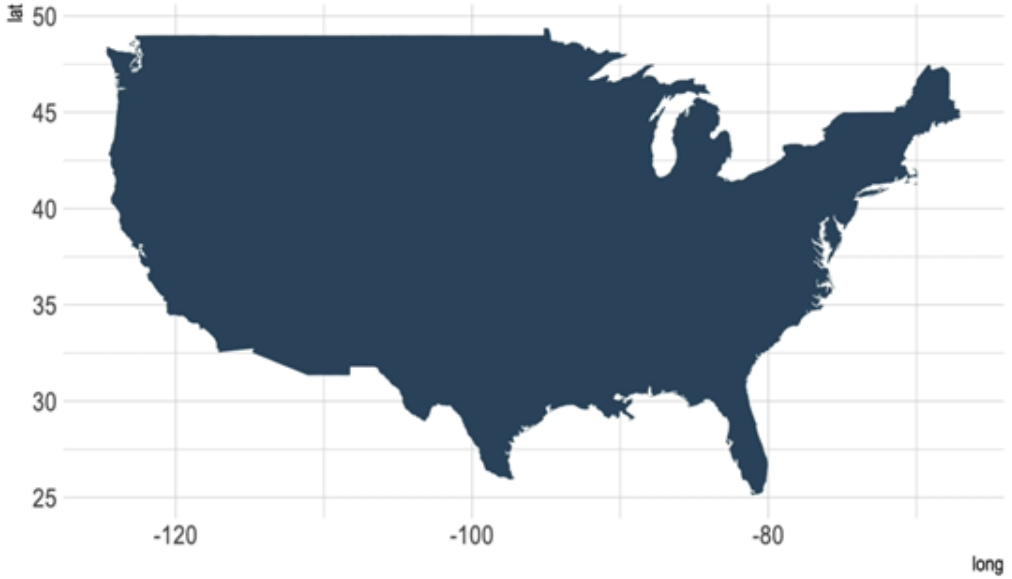
The starting point of any geospatial visualization.

(1) Find spatial data

Shape file (*.shp)
GeoJSON
R Libraries
Google/OpenStreet map

(2) Plot it

maps package in R
ggplot2 package in R
ggmap package in R
rgdal(), rgeos()



Background map

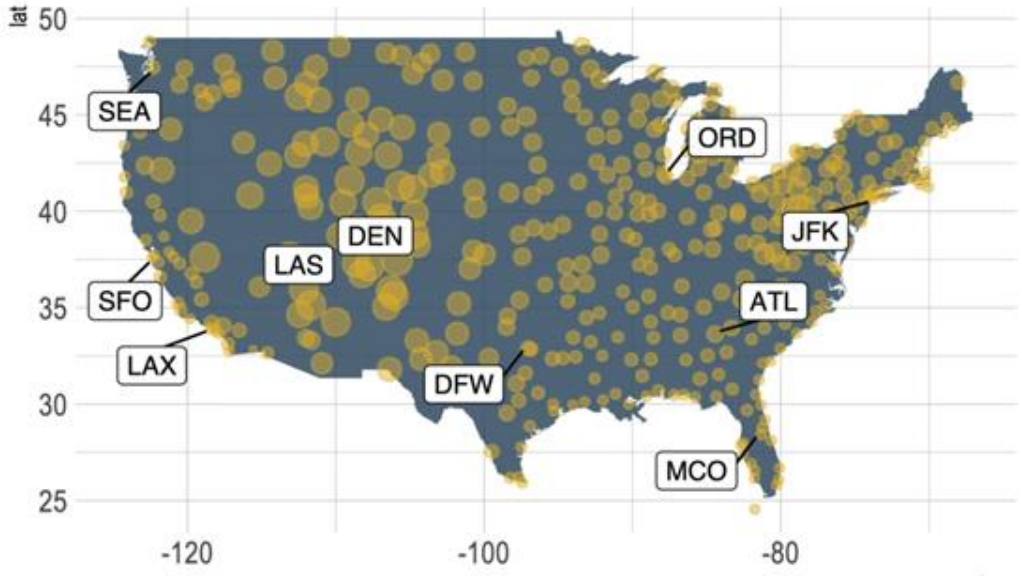
The starting point of any geospatial visualization.

(1) Find spatial data

Shape file (*.shp)
GeoJSON
R Libraries
Google/OpenStreet map

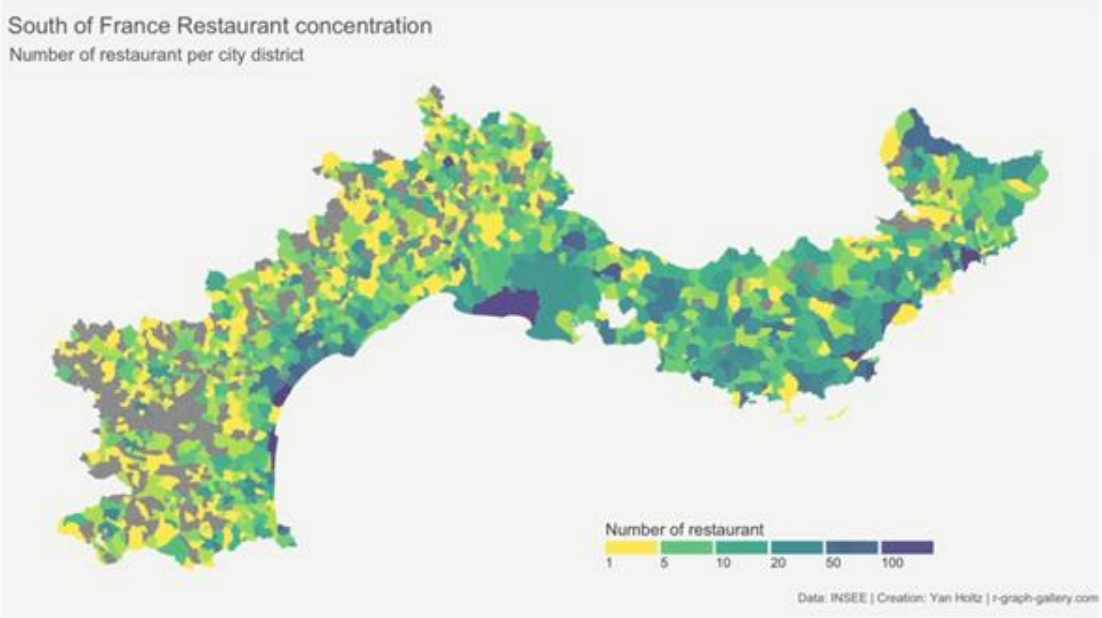
(2) Plot it

maps package in R
ggplot2 package in R
ggmap package in R
rgdal(), rgeos()



Background map > Variation: the choropleth map

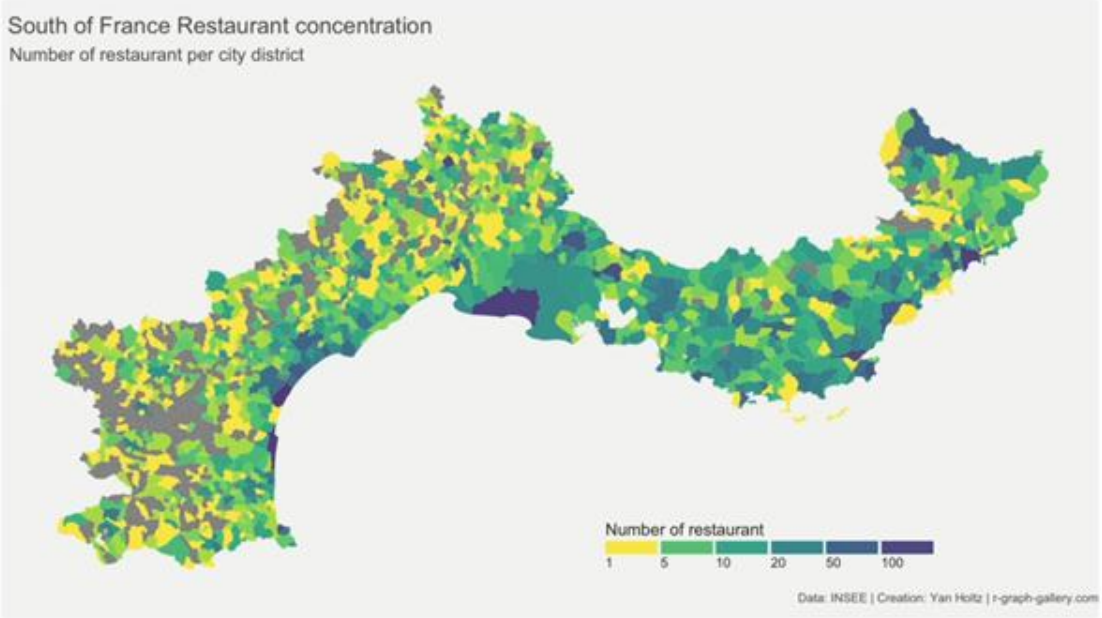
A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.



Background map > Variation: the choropleth map

A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.

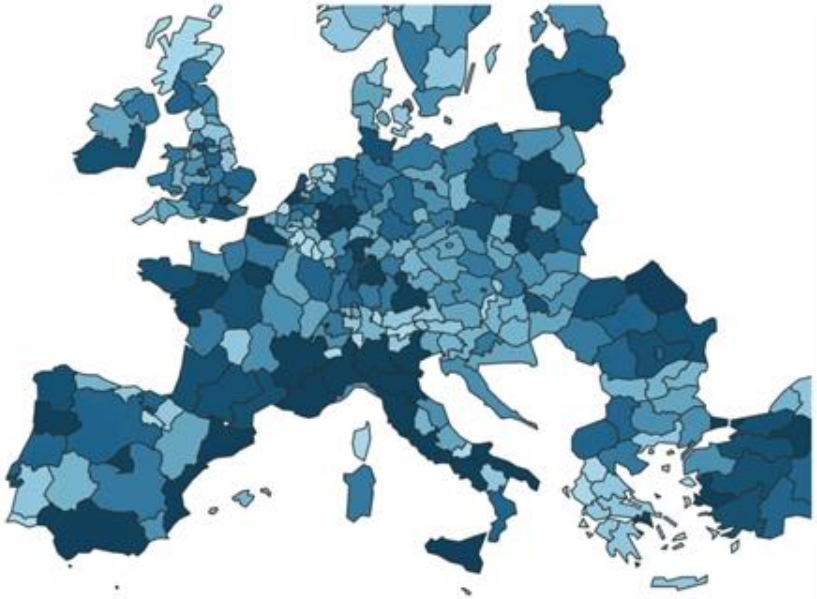
It allows to study how a variable evolves along a territory.



Background map › Variation: the choropleth map

A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.

It allows to study how a variable evolves along a territory.

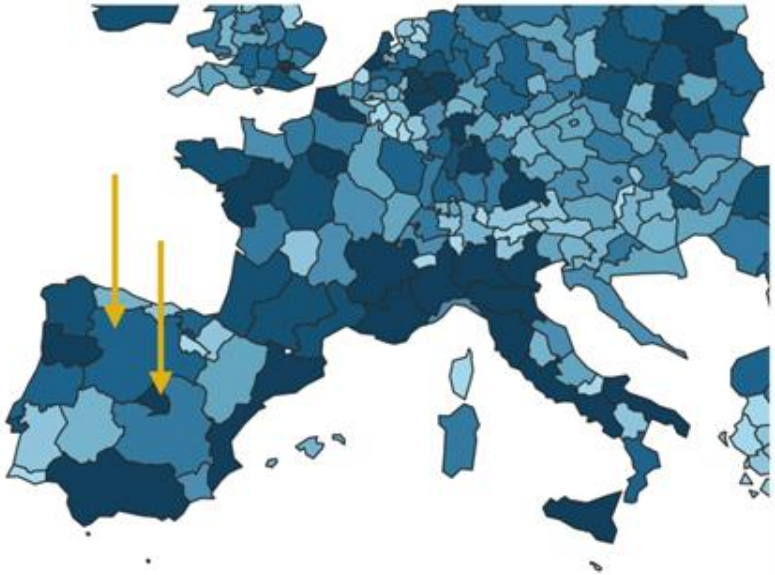


Bigger regions attract more attention

Background map › Variation: the choropleth map

A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.

It allows to study how a variable evolves along a territory.

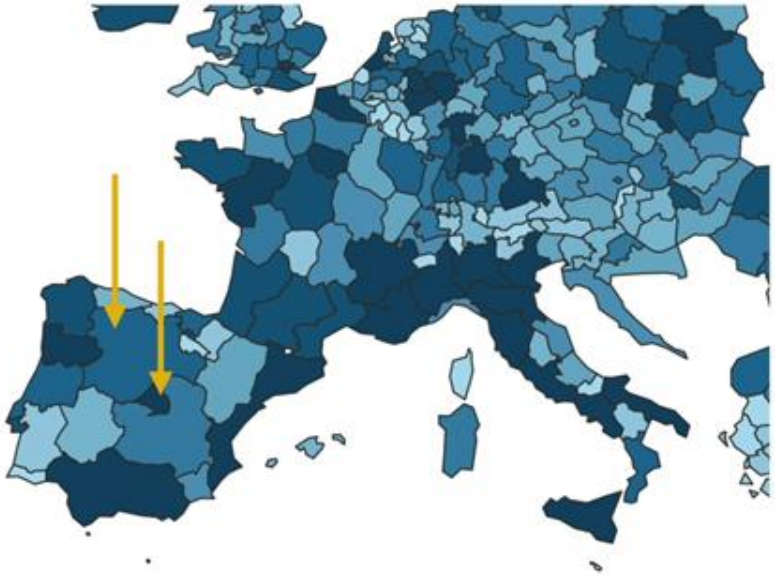


Bigger regions attract more attention

Background map › Variation: the choropleth map

A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.

It allows to study how a variable evolves along a territory.



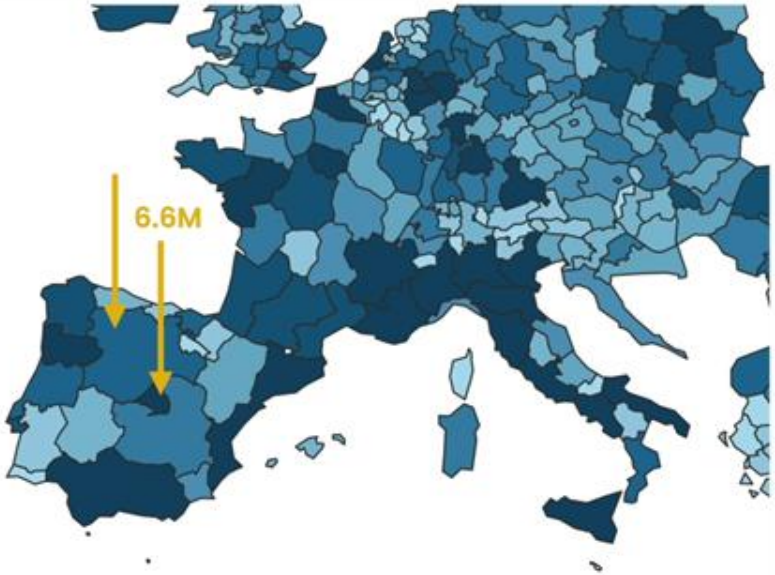
Bigger regions attract more attention

Your variable should be normalized

Background map › Variation: the choropleth map

A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.

It allows to study how a variable evolves along a territory.



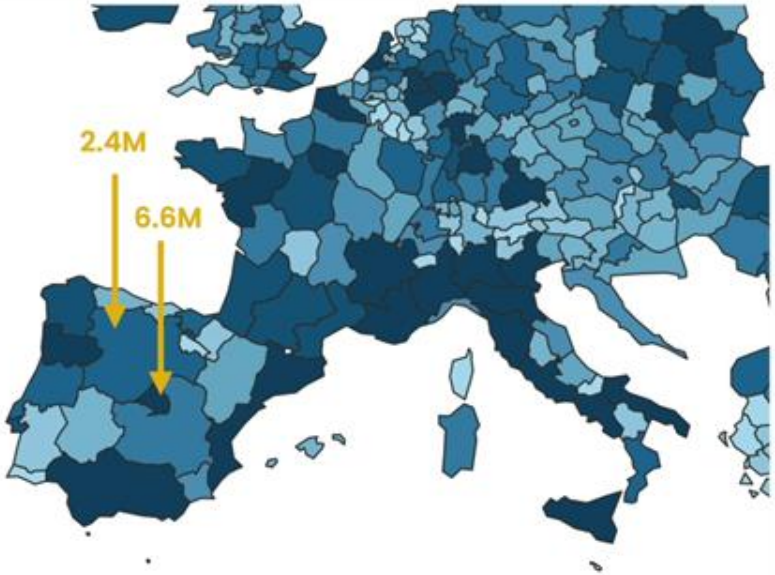
Bigger regions attract more attention

Your variable should be normalized

Background map › Variation: the choropleth map

A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.

It allows to study how a variable evolves along a territory.



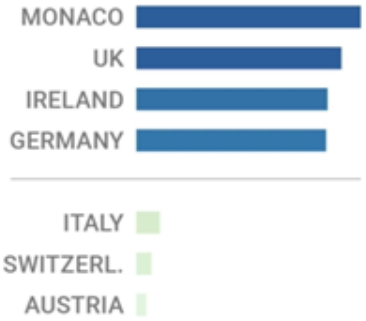
Bigger regions attract more attention

Your variable should be normalized

Background map › Variation: the choropleth map

A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.

It allows to study how a variable evolves along a territory.



Bigger regions attract more attention

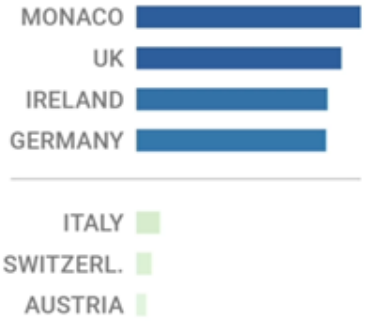
Your variable should be normalized

Good to see the big picture, but not for subtle differences

Background map > Variation: the choropleth map

A choropleth map displays divided geographical areas or regions that are coloured in relation to a numeric variable.

It allows to study how a variable evolves along a territory.



Bigger regions attract more attention

Your variable should be normalized

Good to see the big picture, but not for subtle differences

1/29/2025 12:00:35 PM

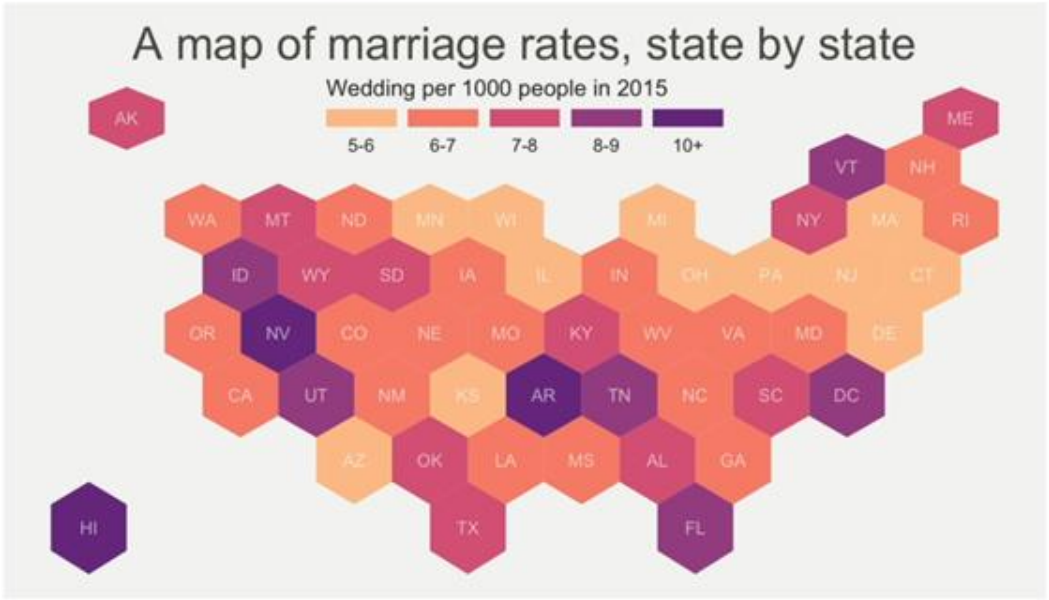
Connection map

SOURCE: R-GRAPH-GALLERY.DE

Connection map

SOURCE: R-GRAPH-GALLERY.DE

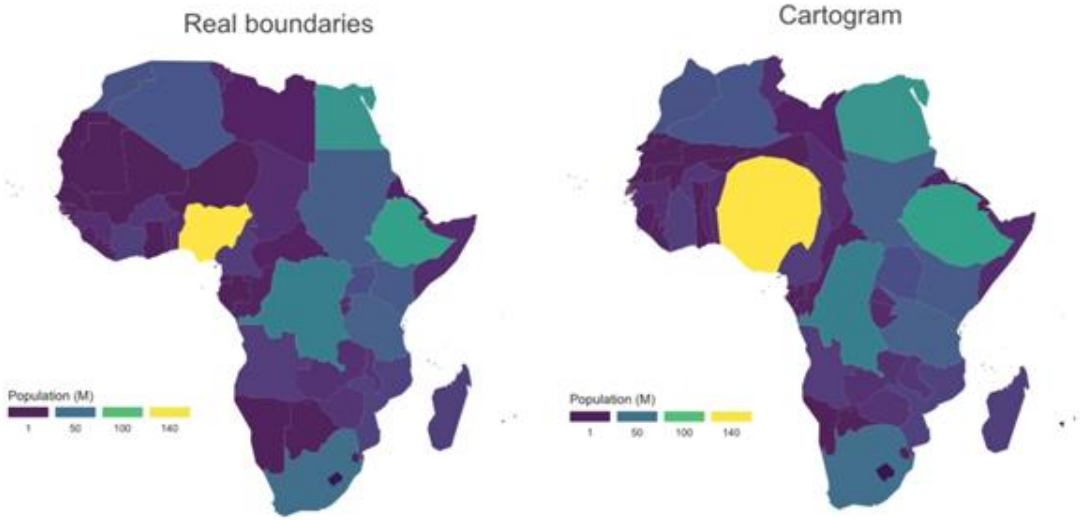
Hexbin map



SOURCE: R-GRAPH-GALLERY.DE

Connection map

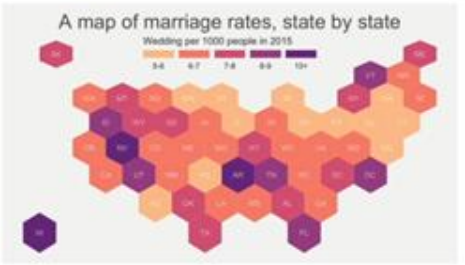
Cartogram



SOURCE: DATA-TO-VIZ

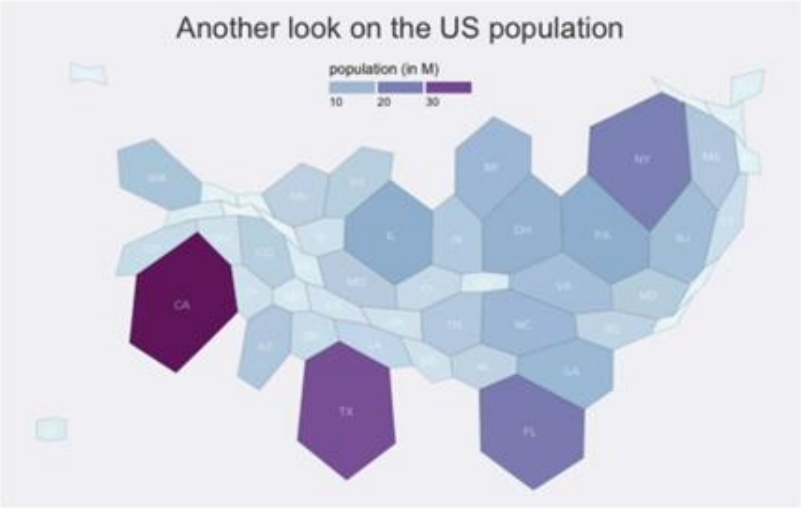
Connection map

Hexbin map



1/29/2025 12:02:06 PM

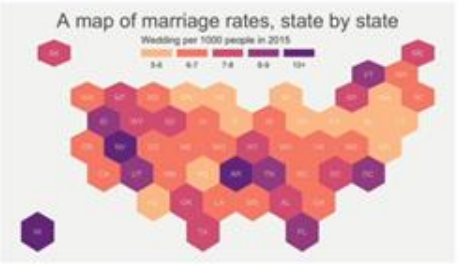
Hexbin map + Cartogram



SOURCE: DATA-TO-VIZ

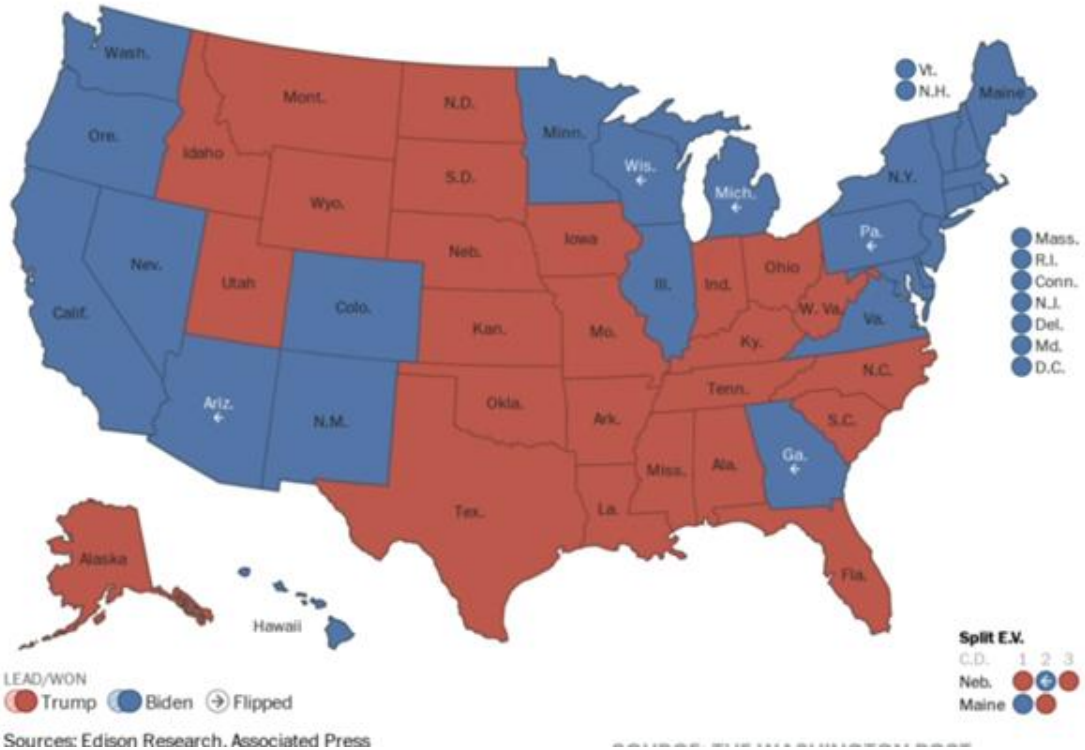
Connection map

Hexbin map



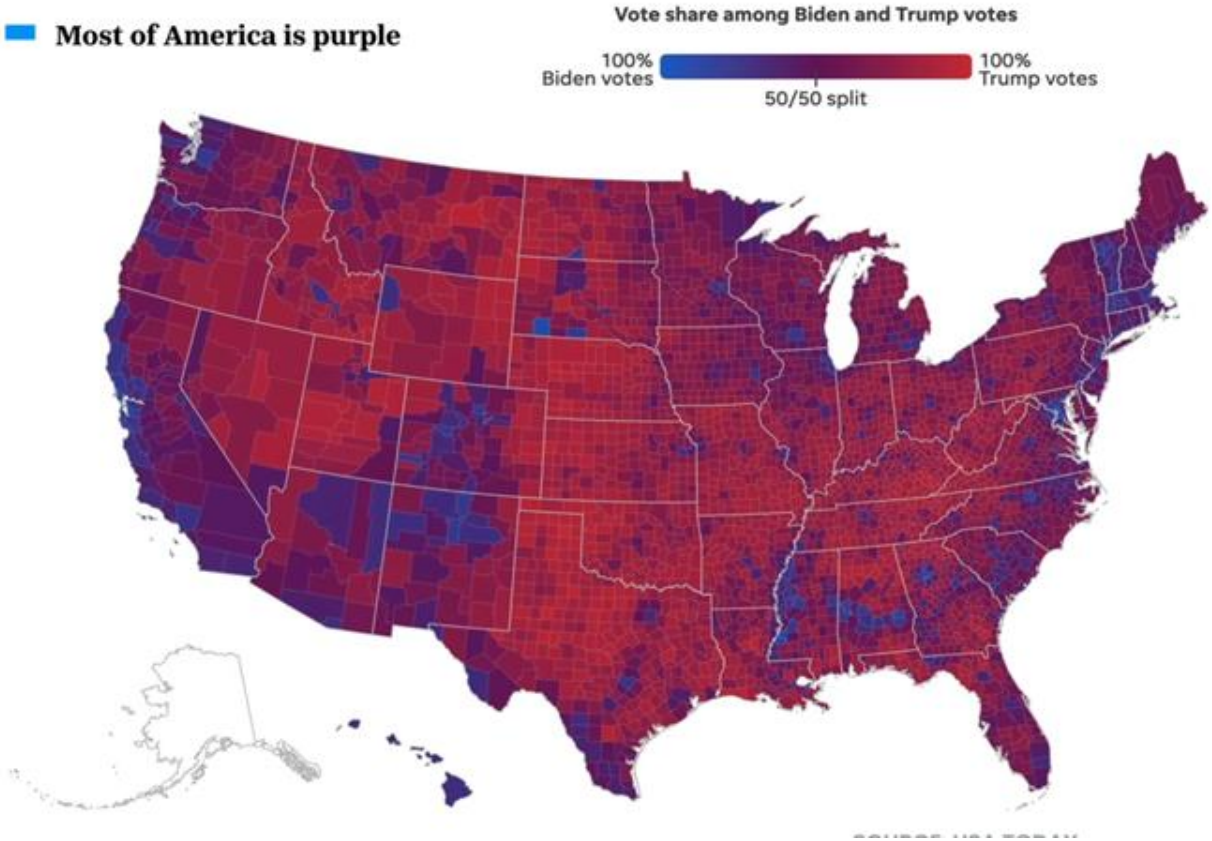
Cartogram

Modern approaches to maps

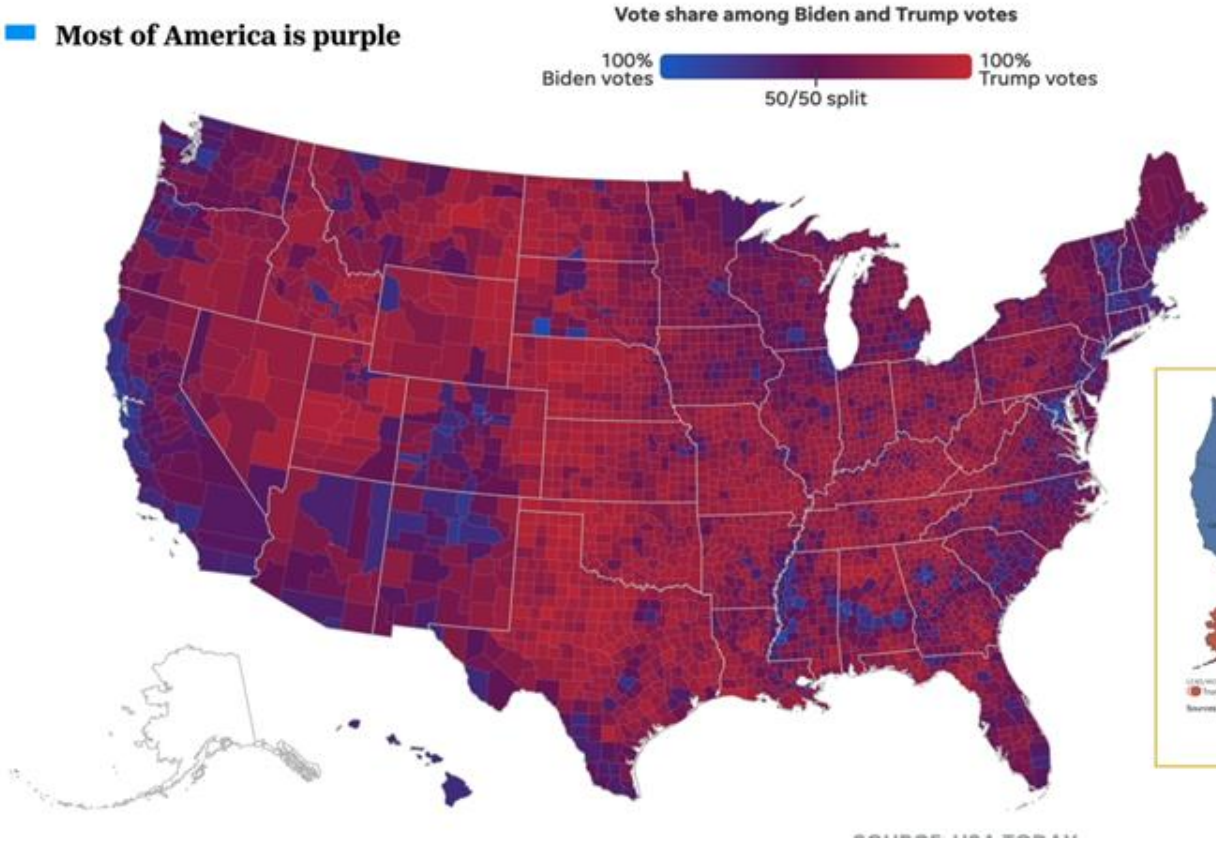


1/29/2025 12:03:36 PM

Modern approaches to maps

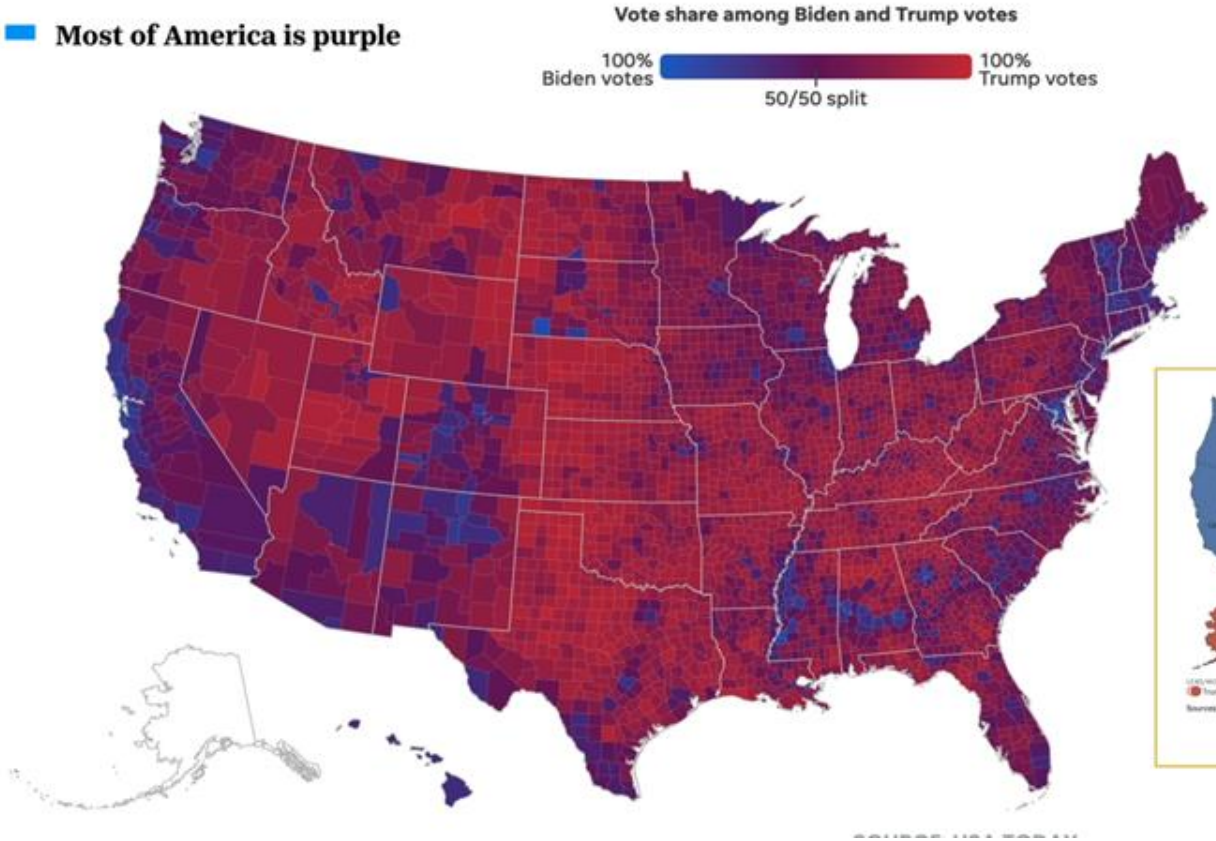


Modern approaches to maps



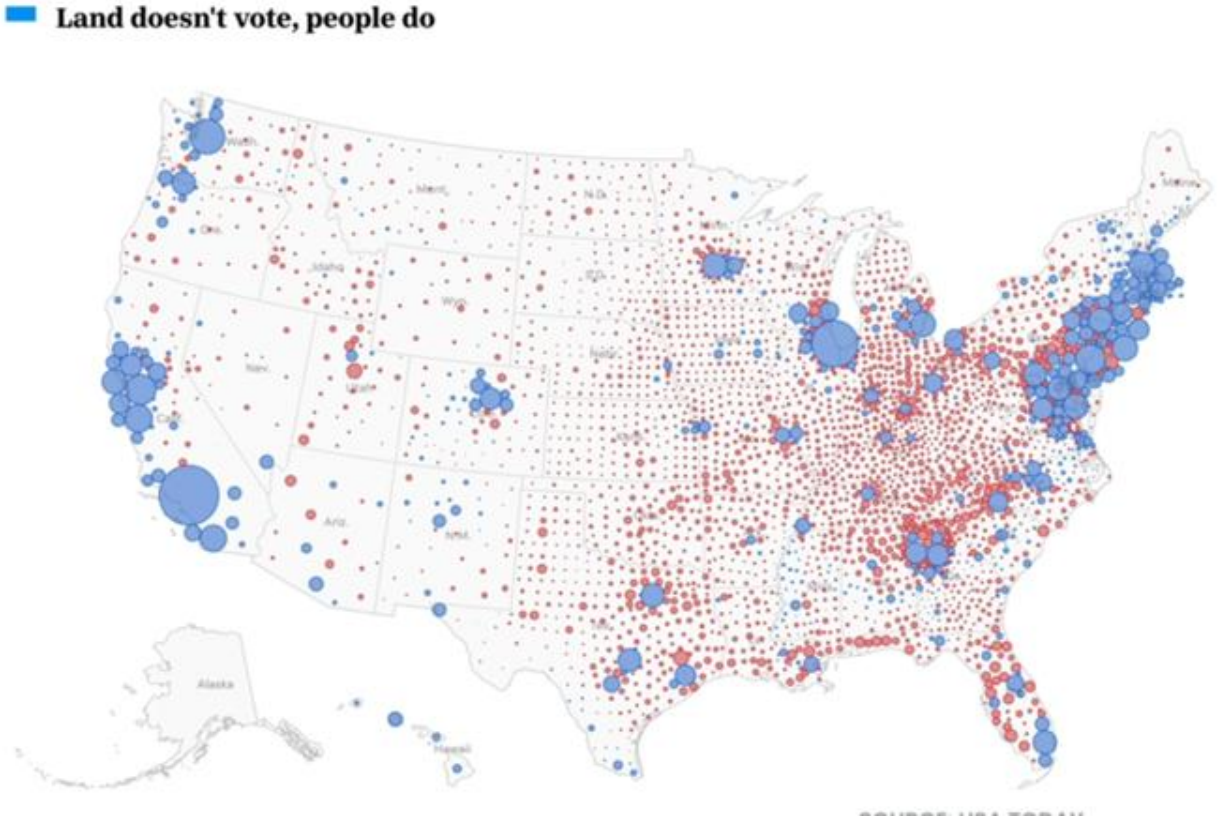
1/29/2025 12:04:46 PM

Modern approaches to maps



1/29/2025 12:05:06 PM

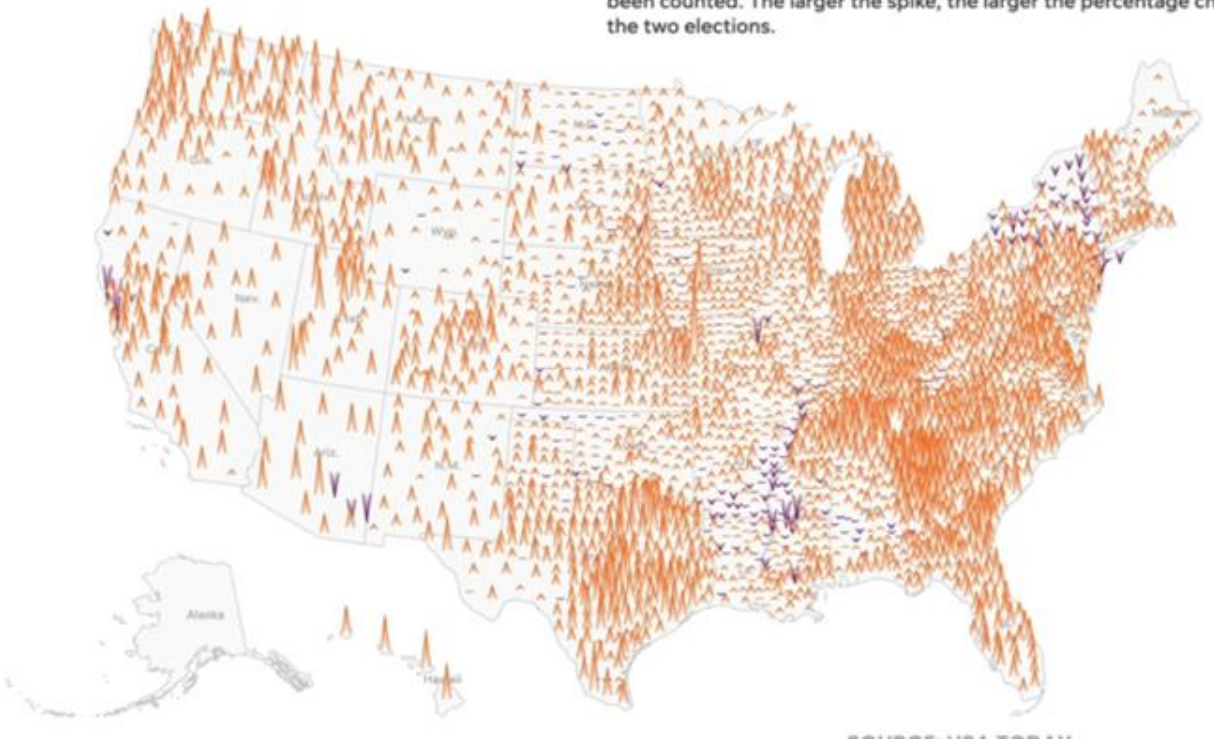
Modern approaches to maps



Modern approaches to maps

Voters turned out in record numbers

This map shows which counties saw the number of ballots cast **increase** and **decrease** compared to 2016, where at least an estimated 70% of the vote has been counted. The larger the spike, the larger the percentage change between the two elections.

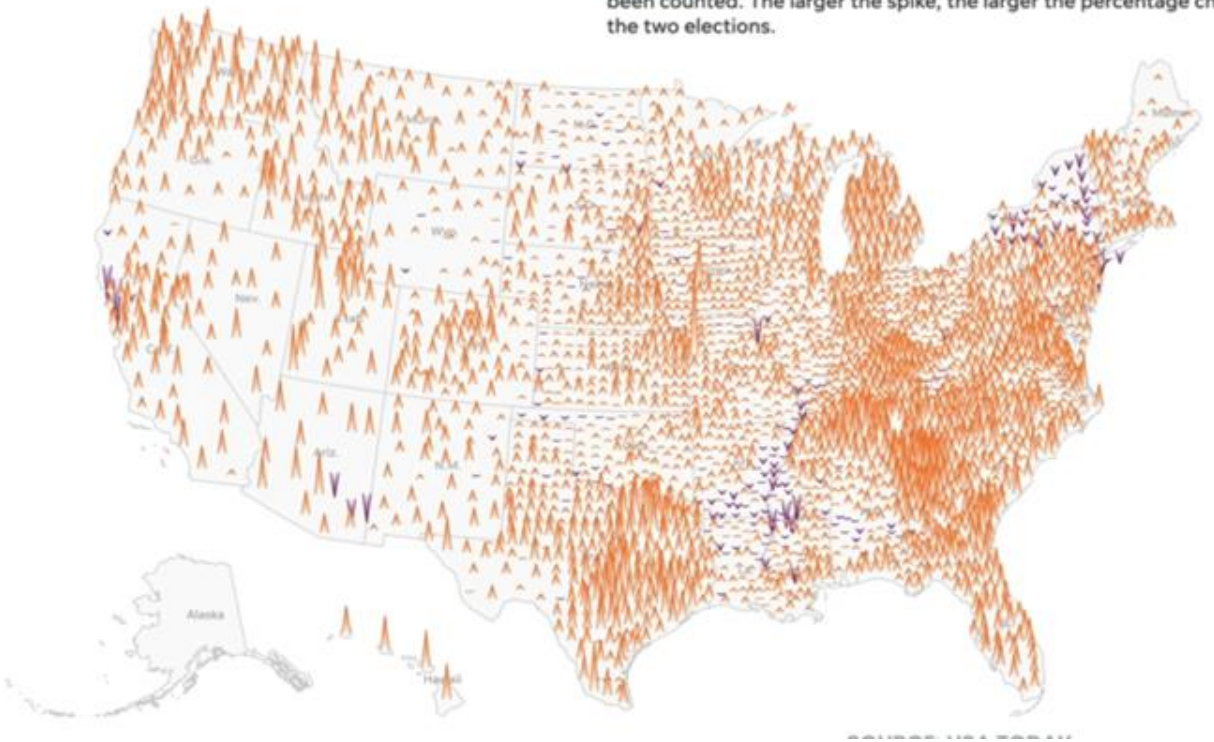


1/29/2025 12:06:16 PM

Modern approaches to maps

Voters turned out in record numbers

This map shows which counties saw the number of ballots cast **increase** and **decrease** compared to 2016, where at least an estimated 70% of the vote has been counted. The larger the spike, the larger the percentage change between the two elections.

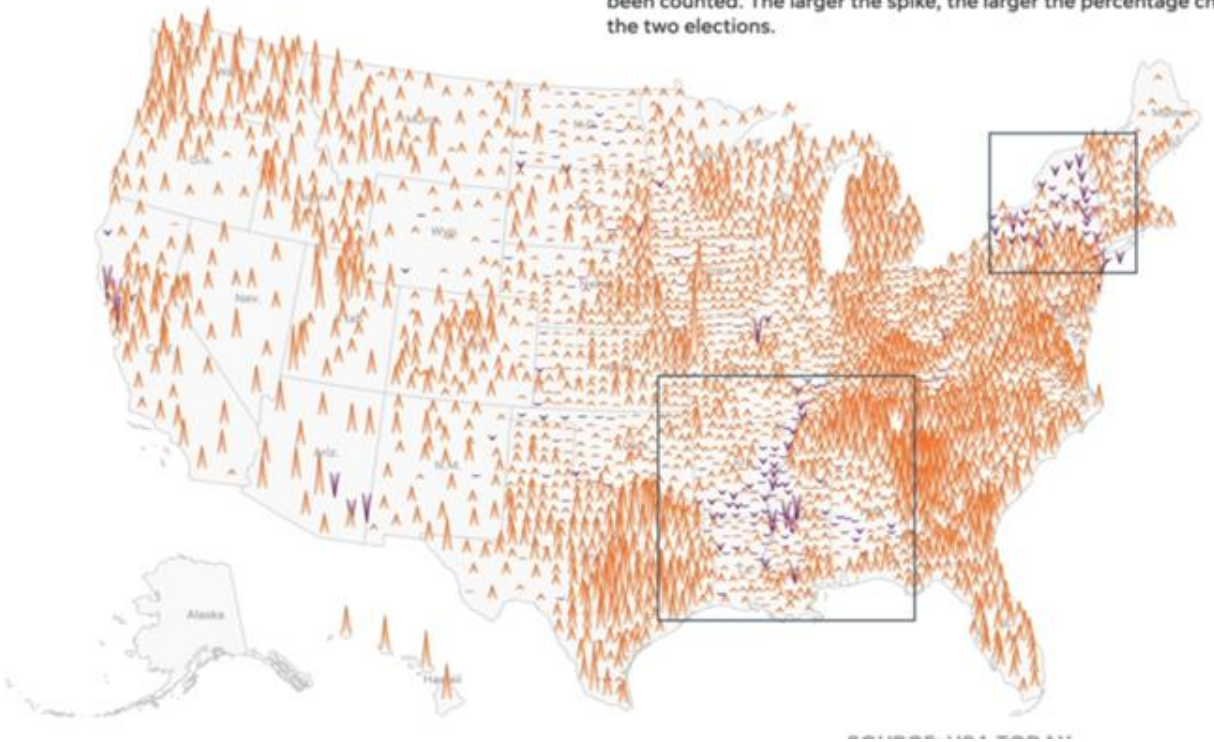


1/29/2025 12:06:56 PM

Modern approaches to maps

Voters turned out in record numbers

This map shows which counties saw the number of ballots cast **increase** and **decrease** compared to 2016, where at least an estimated 70% of the vote has been counted. The larger the spike, the larger the percentage change between the two elections.

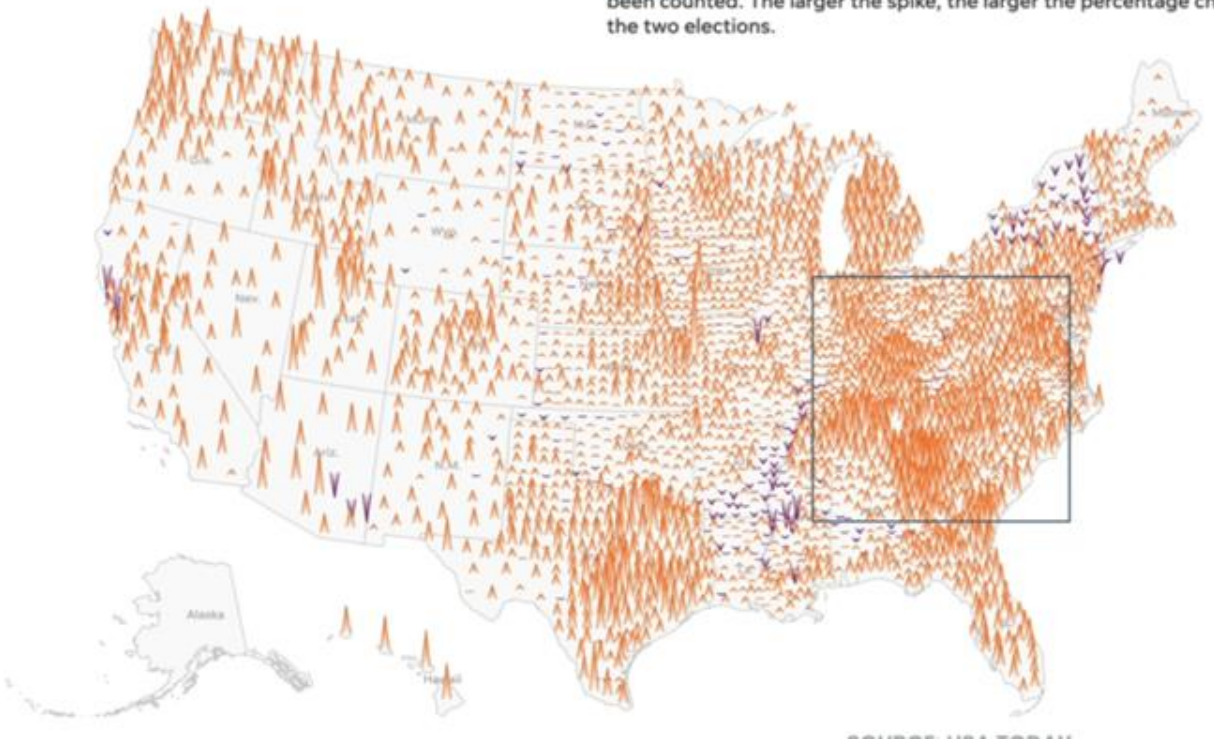


1/29/2025 12:07:06 PM

Modern approaches to maps

Voters turned out in record numbers

This map shows which counties saw the number of ballots cast **increase** and **decrease** compared to 2016, where at least an estimated 70% of the vote has been counted. The larger the spike, the larger the percentage change between the two elections.

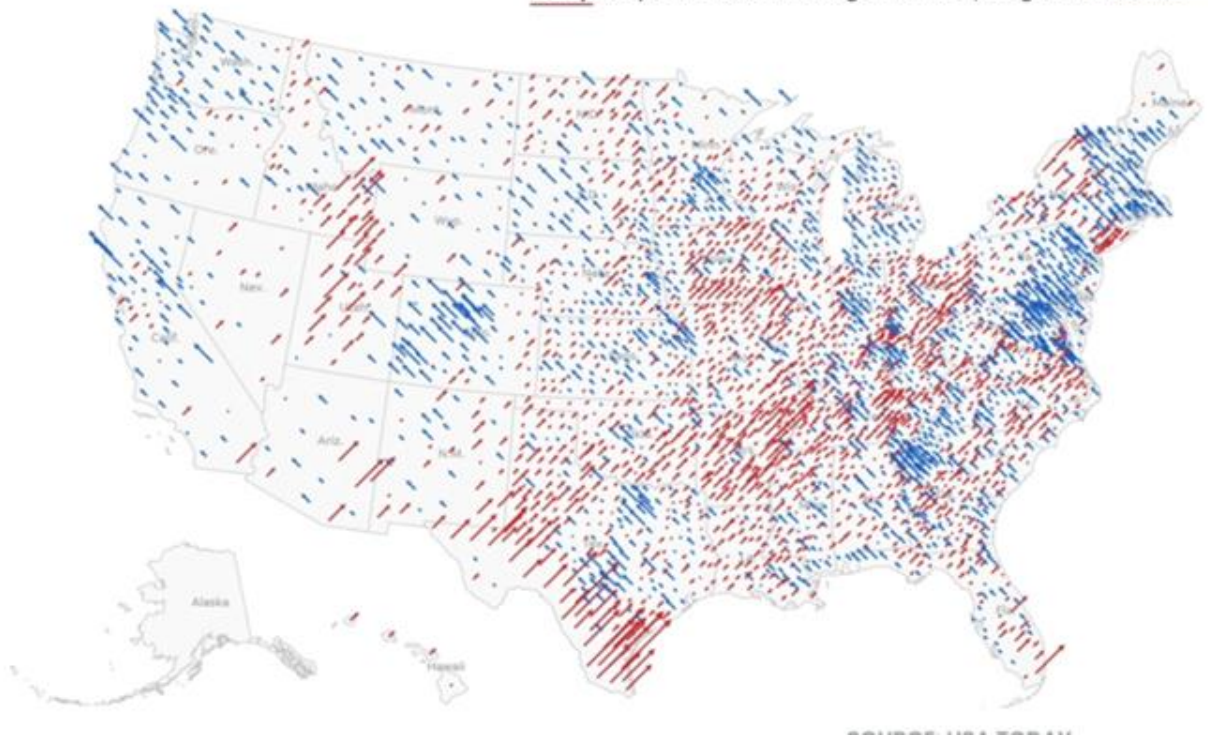


1/29/2025 12:07:16 PM

Modern approaches to maps

How counties shifted from 2016

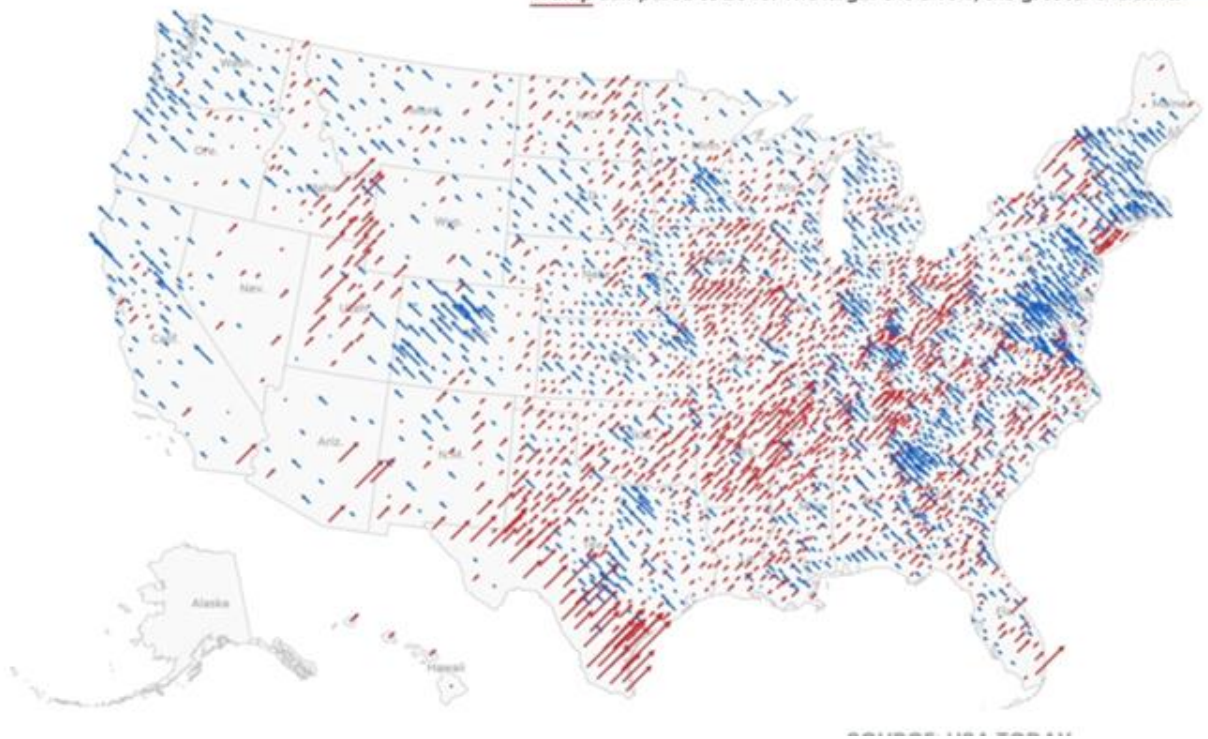
The map below depicts in which counties the margin of victory **shifted toward Biden** compared to Clinton in 2016, and in which the margin **shifted toward Trump** compared to 2016. The larger the arrow, the greater the shift.



Modern approaches to maps

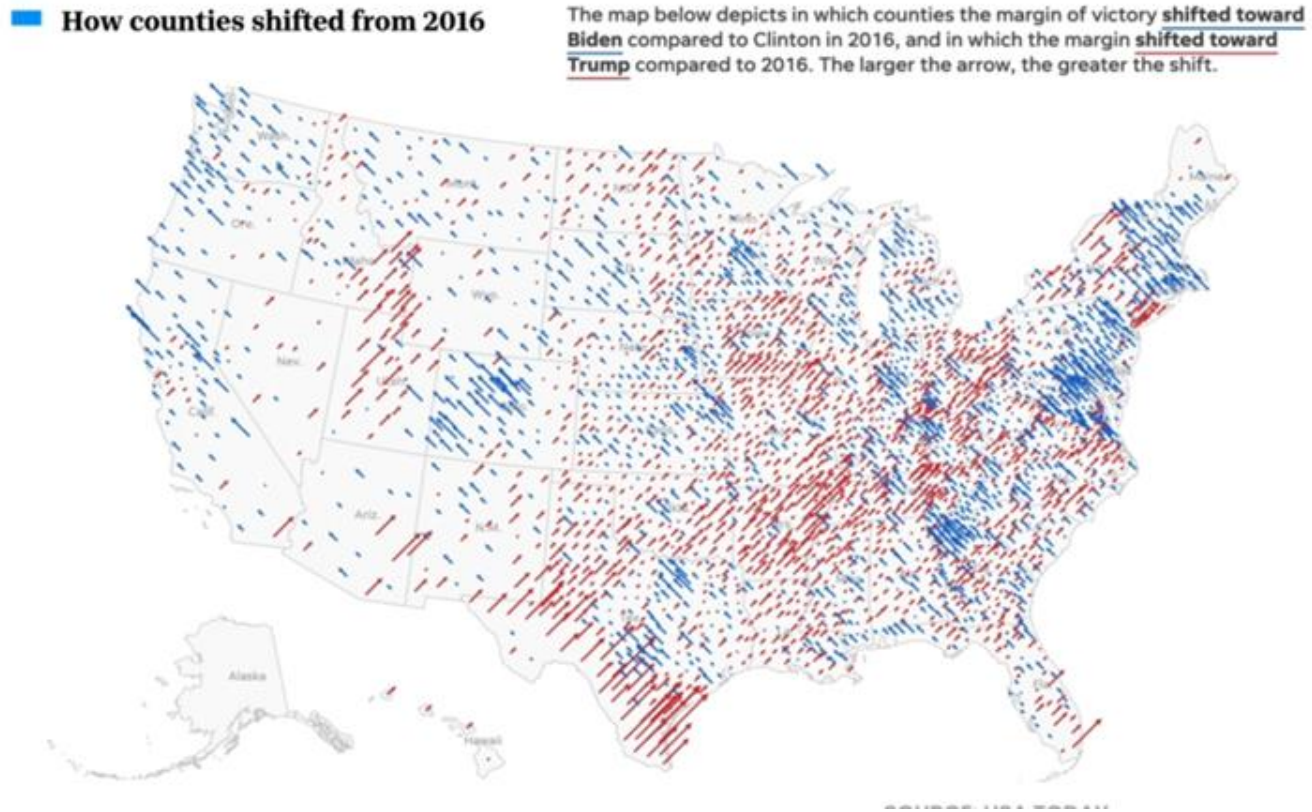
How counties shifted from 2016

The map below depicts in which counties the margin of victory **shifted toward Biden** compared to Clinton in 2016, and in which the margin **shifted toward Trump** compared to 2016. The larger the arrow, the greater the shift.

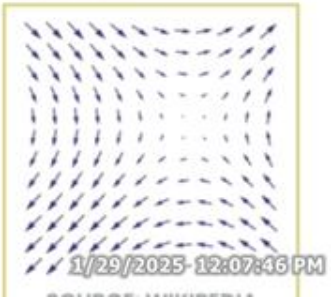


1/29/2025 12:07:36 PM

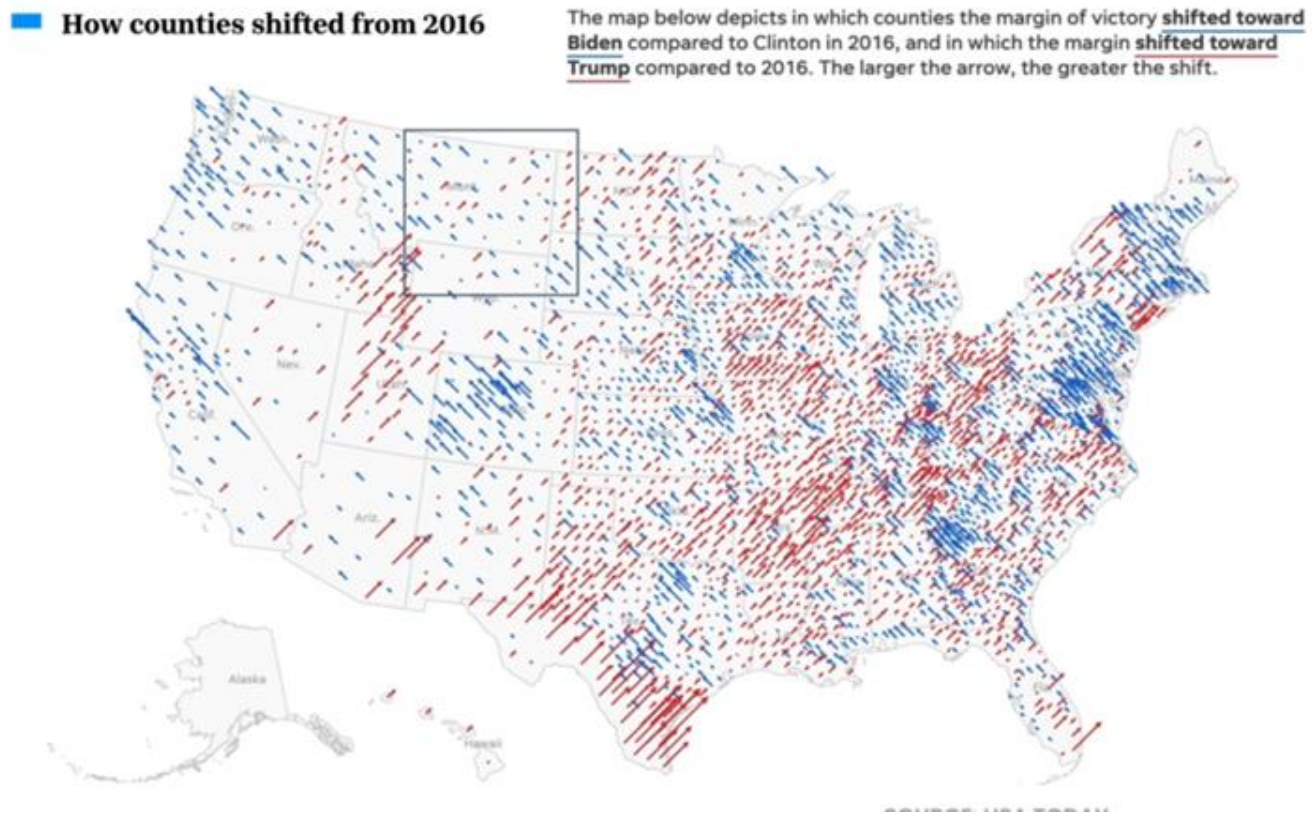
Modern approaches to maps



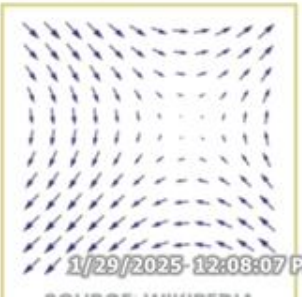
VECTOR FIELD DIAGRAM



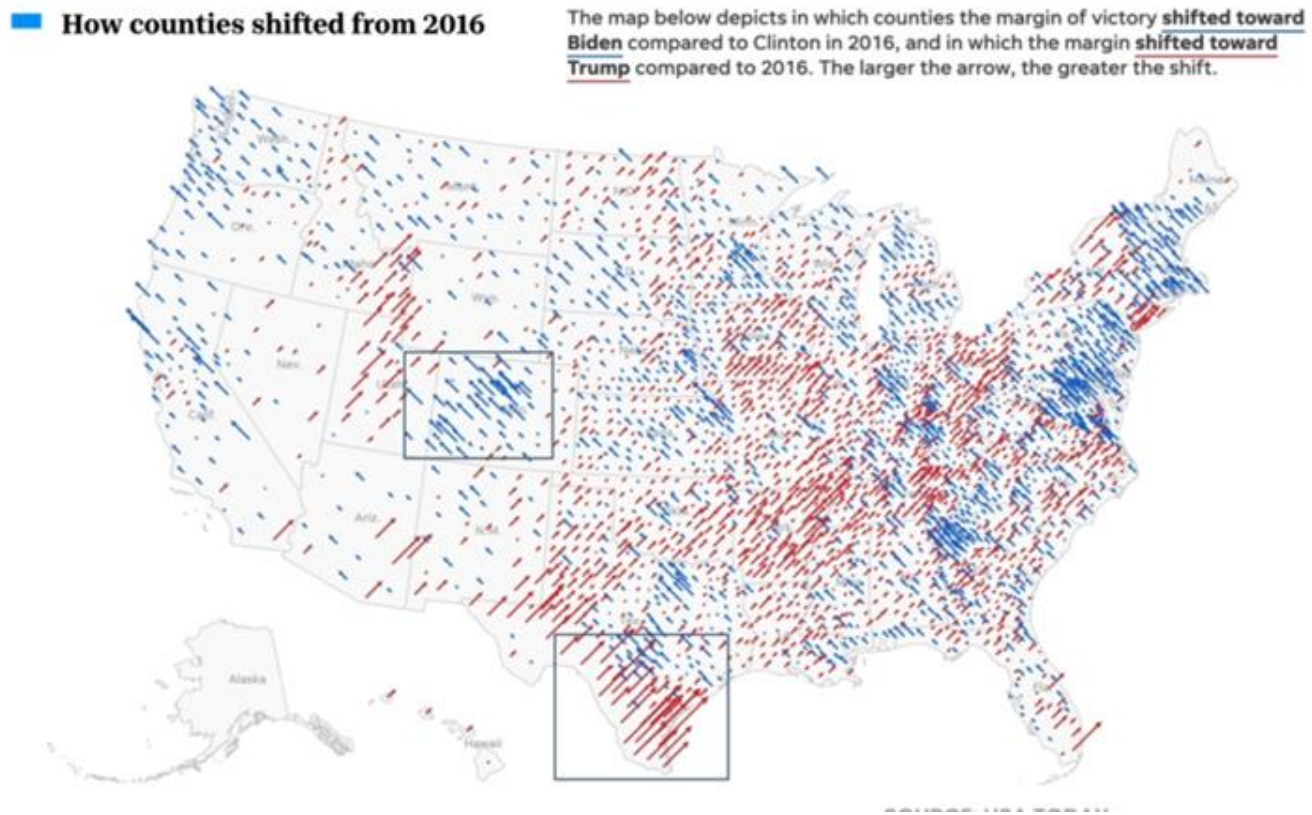
Modern approaches to maps



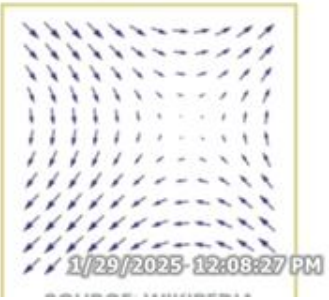
VECTOR FIELD DIAGRAM



Modern approaches to maps



VECTOR FIELD DIAGRAM



Drawing maps

Drawing maps

Drawing maps