

# Unit III

Taxonomy of Visualization

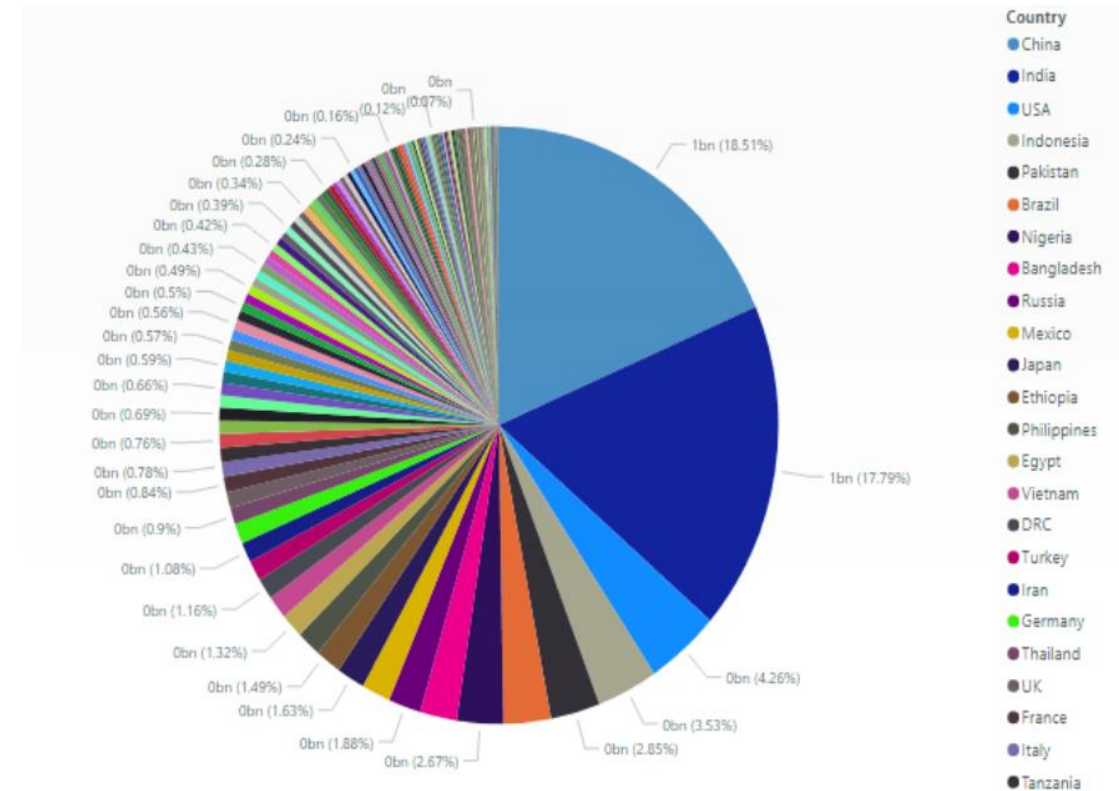
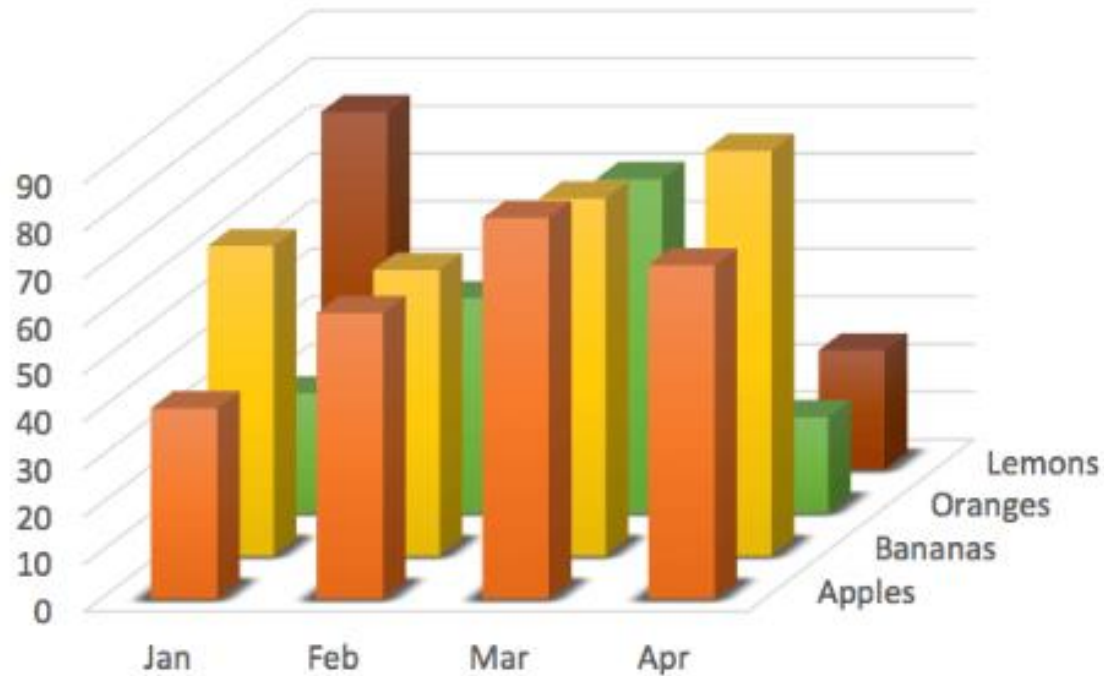
# Importance of Visualization

- Data Visualization is a way of representing data graphically to help people easily understand the information.
- It can be used to convey complex relationships between different variables or to analyze trends over time.
- Data visualization can be done using charts, graphs, maps, histograms, scatter plots, and other visuals.
- By using colors, shapes, and other visual elements, data visualization can make it easier for people to comprehend large amounts of data quickly and accurately.

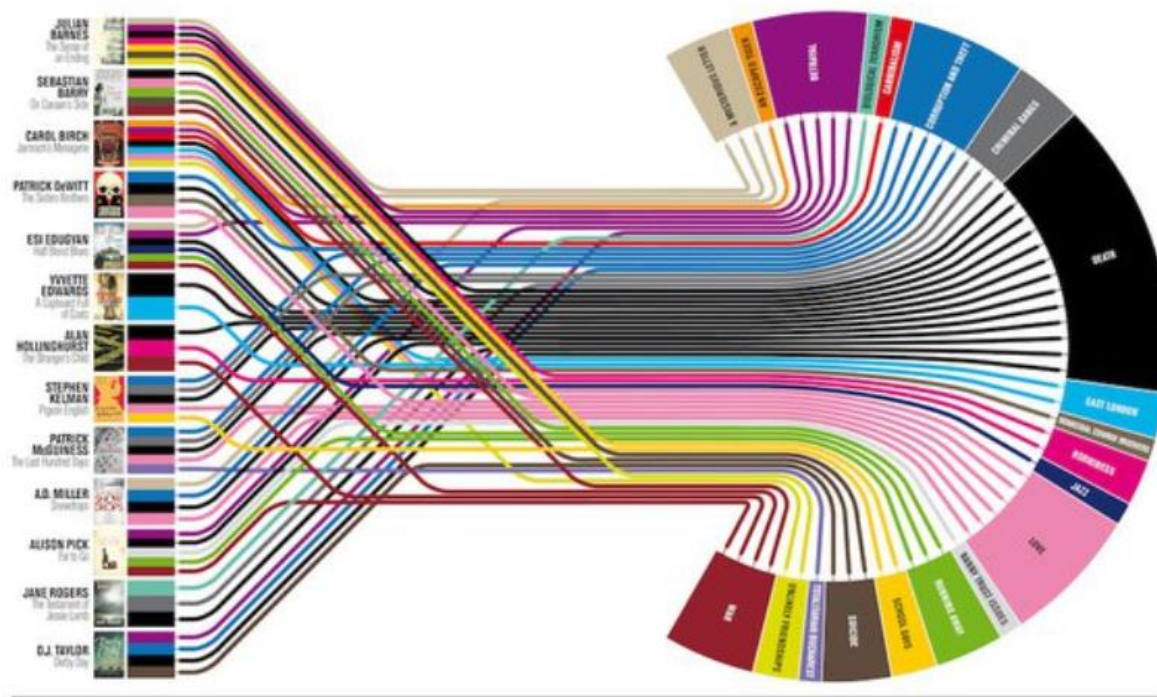
# Importance of Visualization continued..

- Data Visualization is a powerful tool for exploring data, identifying patterns and trends, and communicating insights.
- It can provide insight into correlations and trends that may otherwise be difficult to detect from examining raw data alone.
- It is why data visualization tools and methods are often used in business, financial analysis, project management, scientific studies, and just about everywhere else – as long as there is some data to visualize.

# Data Visualization Gone Wrong!



# Data Visualization Gone Wrong!



## Plot lines

What makes a prize-winning novel? As Julian Barnes wins the Booker Prize, Delayed Gratification's Johanna Kamradt charts the themes of this year's longlisters.

Illustration: Christian Tate



# The four basic Presentation Types

- Comparison
- Composition
- Distribution
- Relationship

# How to Select the Right Chart?

- answer a few questions:
  - How many variables do you want to show in a single chart?
    - One, two, three, many?
  - How many items (data points) will you display for each variable?
    - Only a few or many?
  - Will you display values over a period of time, or among items or groups?

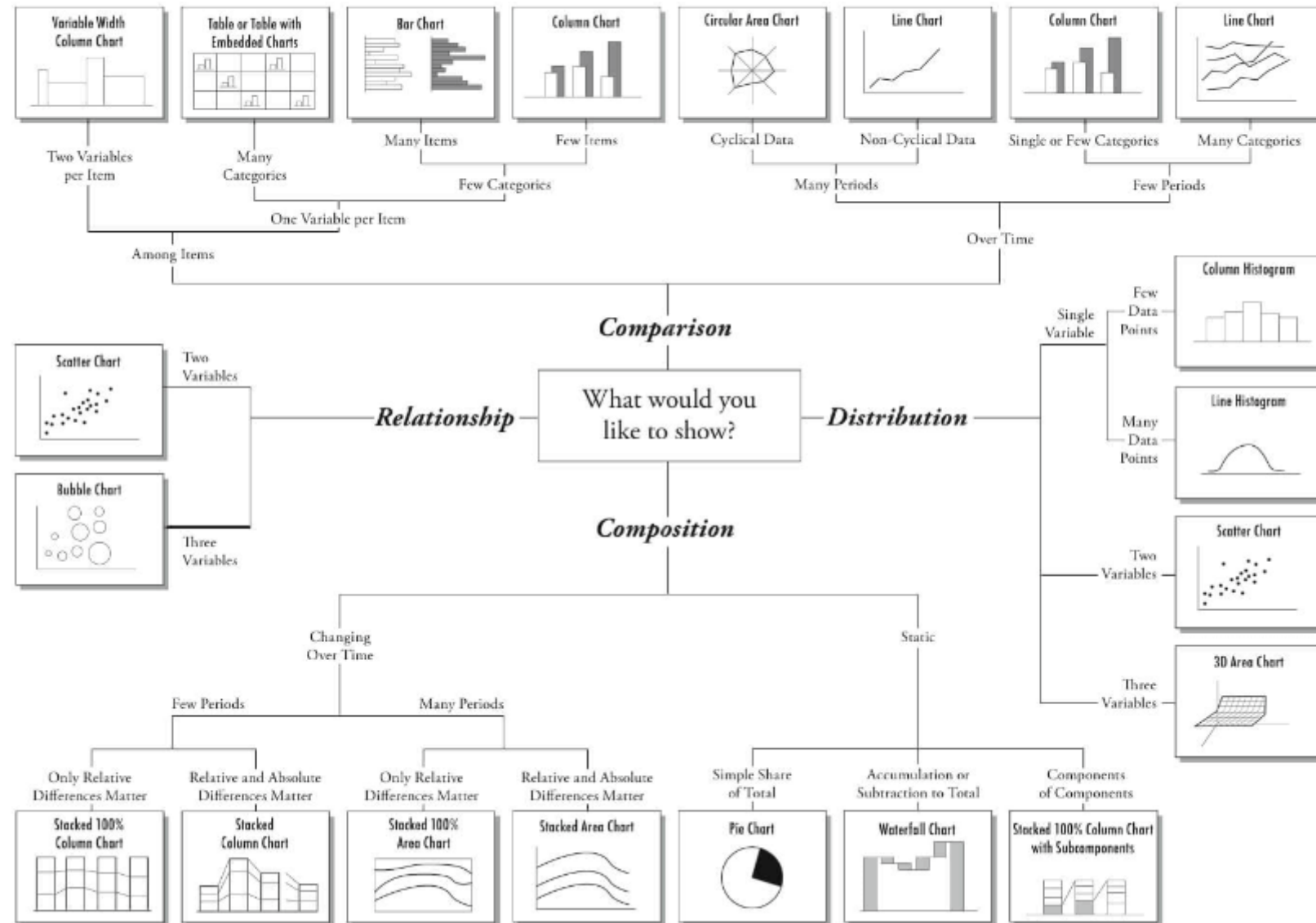
# Example of Selecting the Right!

- Bar charts are good for comparisons, while line charts work better for trends.
- Scatter plot charts are good for relationships and distributions,
- Pie charts should be used only for simple compositions — never for comparisons or distributions.



# Chart Suggestions—A Thought-Starter

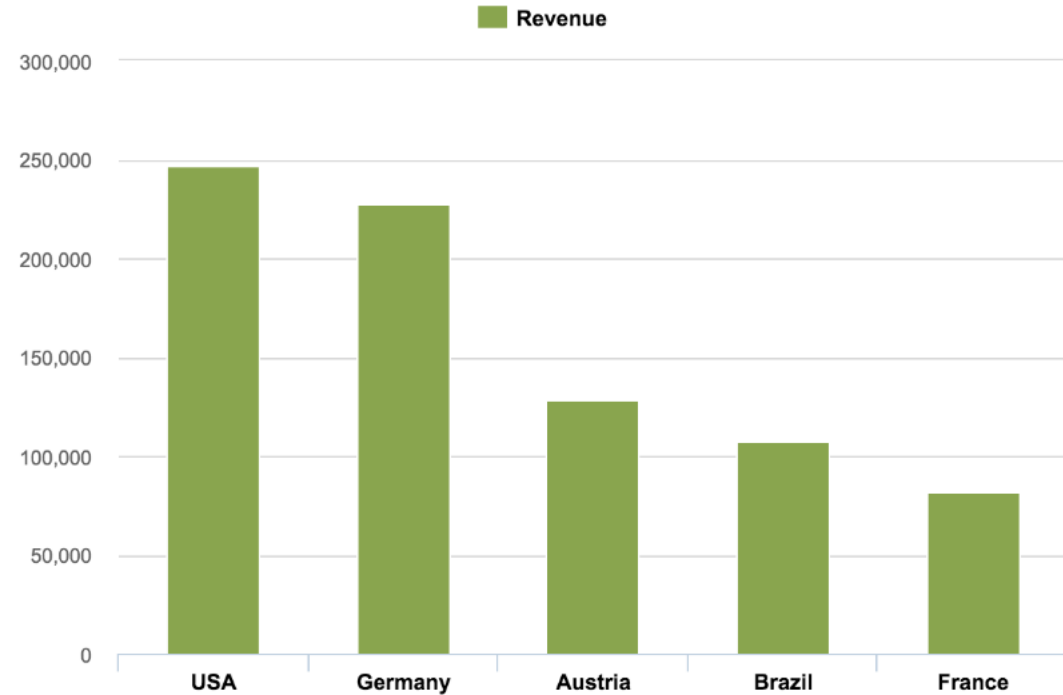
www.ExtremePresentation.com  
© 2009 A. Abela — a.v.abela@gmail.com



# Column Chart

This chart is best used to compare different values when specific values are important, and it is expected that users will look up and compare individual values between each column.

With column charts, one can compare values for different categories or compare value changes over a period of time for a single category.

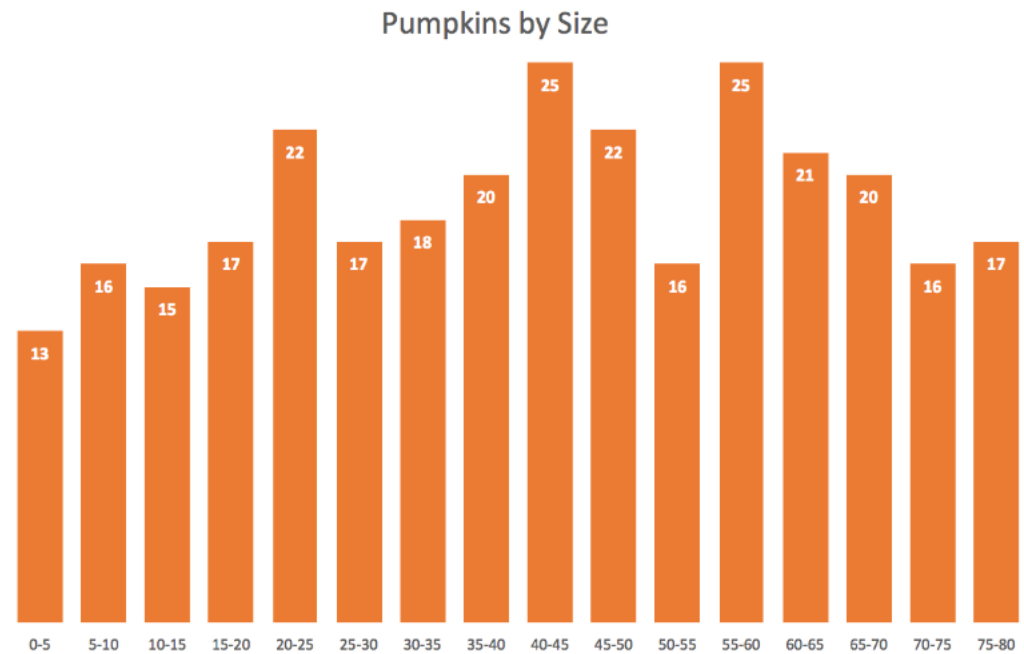


# Best Practice of Column Chart

- Use column charts for comparison if the number of categories is quite small — up to five, but not more than seven categories.
- If one of your data dimensions is time — including years, quarters, months, weeks, days, or hours — you should always set the time dimension on the horizontal axis.
- In charts, time should always run from left to right, never from top to bottom.
- For column charts, the numerical axis must start at zero. Our eyes are very sensitive to the height of columns, and we can draw inaccurate conclusions when those bars are truncated.
- Avoid using pattern lines or fills. Use a border only for highlights.
- Only use column charts to show trends if there are a reasonably-low number of data points (less than 20) and if every data point has a clearly-visible value.

# Column Histograms

A histogram is a common variation of column charts used to present the distribution and relationships of a single variable over a set of categories. A good example of a histogram would be a distribution of grades on a school exam or the sizes of pumpkins, divided by size group, in a pumpkin festival.



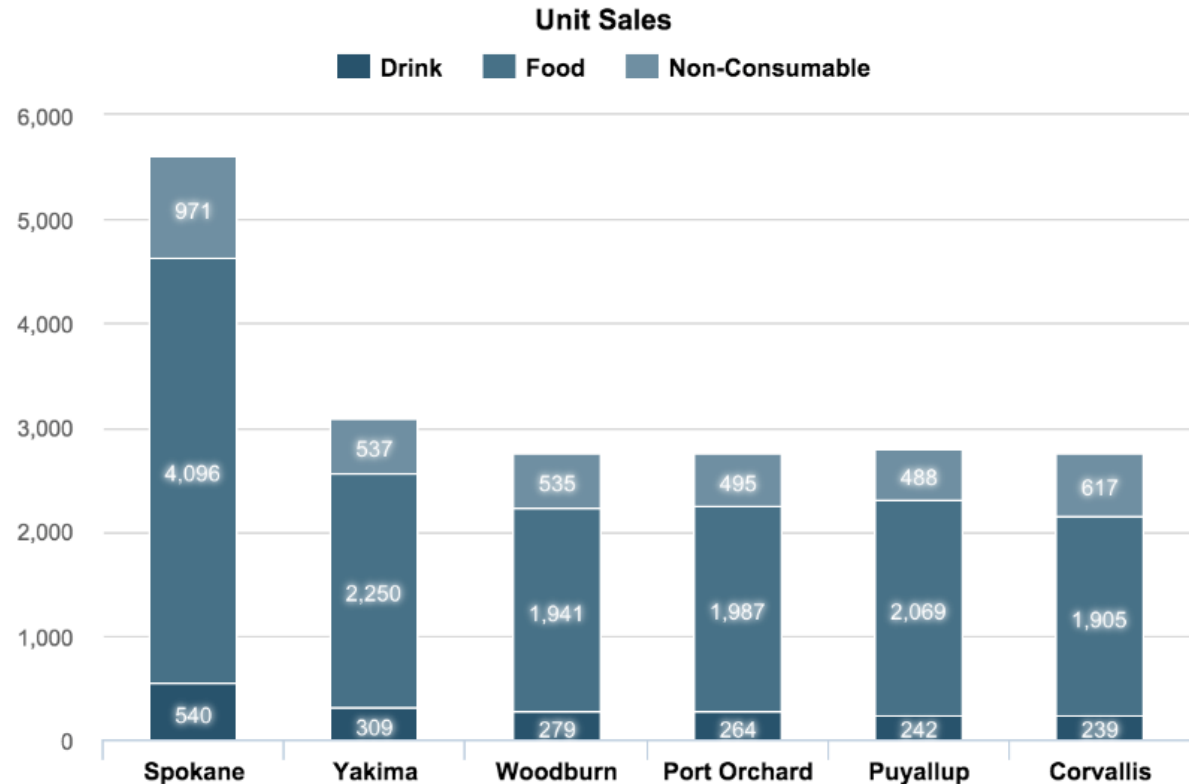
# Stacked Column Chart

Use stacked column charts to show a composition.

Do not use too many composition items (not more than three or four).

Make sure the composing parts are relatively similar in size.

It can get messy very quickly.

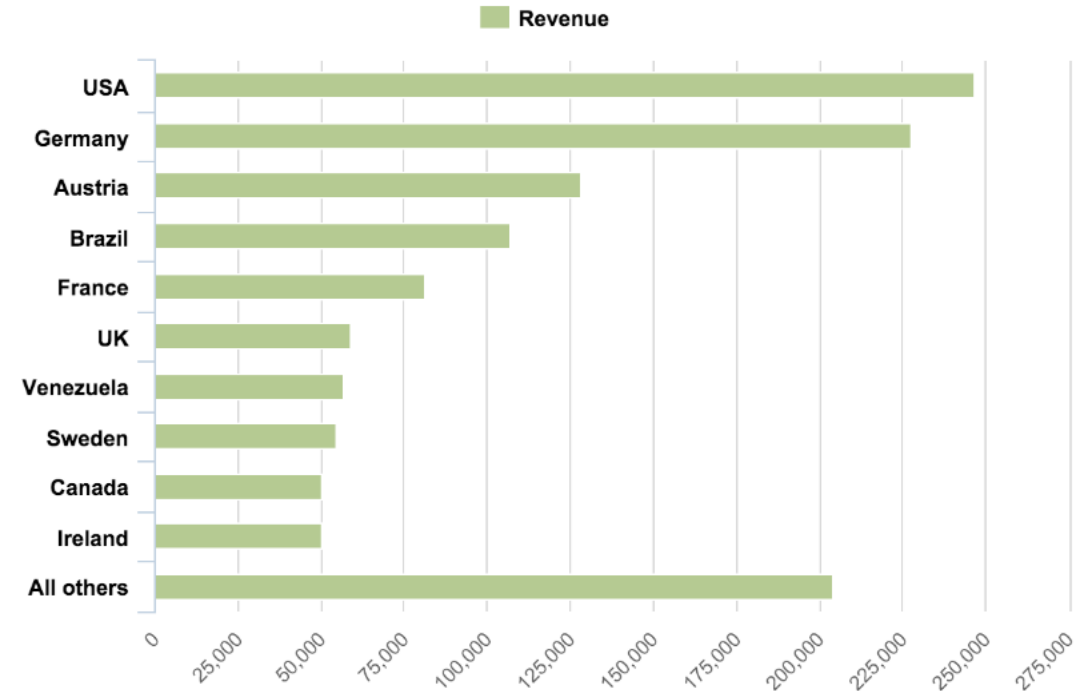


# Bar Charts

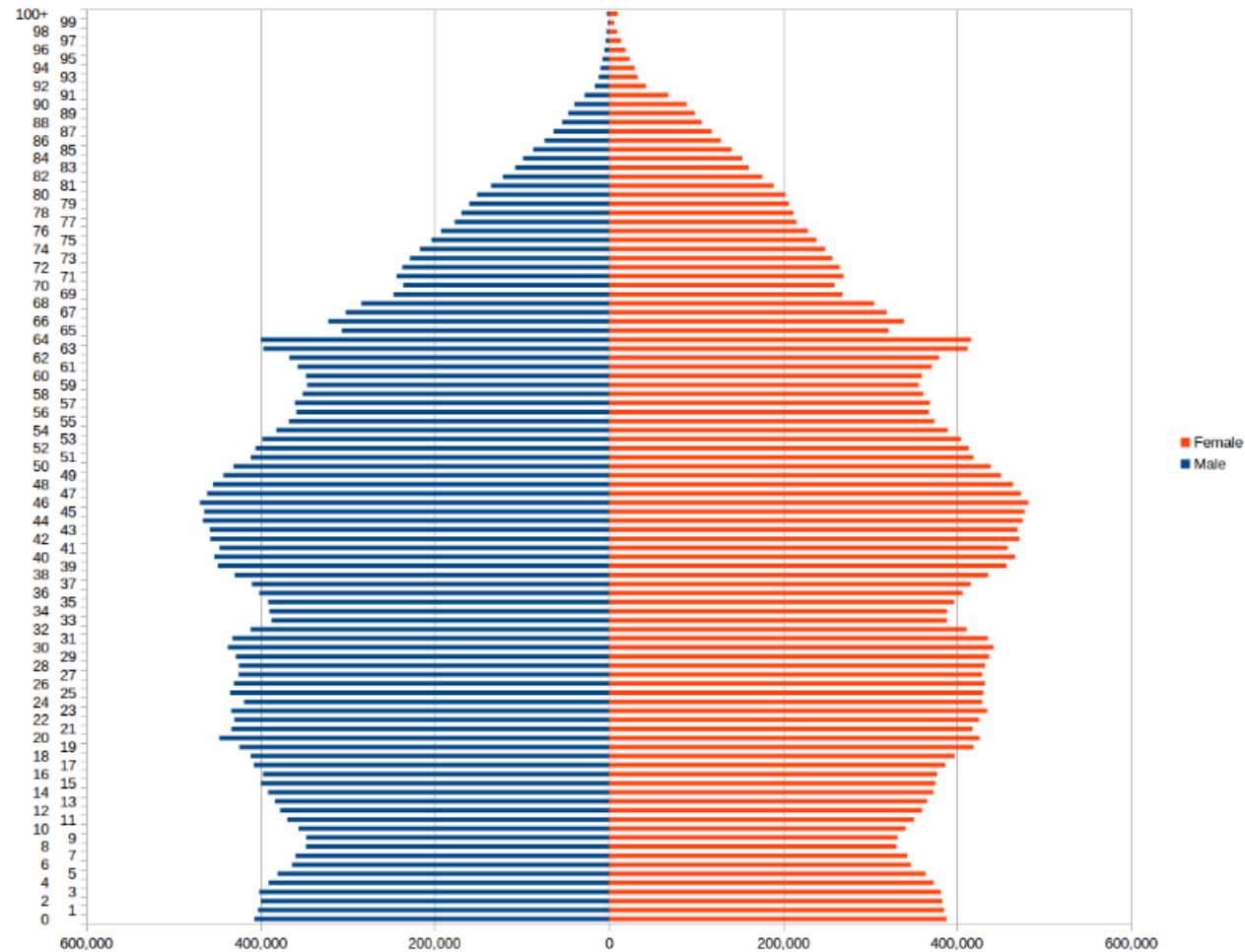
Bar charts are essentially horizontal column charts.

For a long category names, it is best to use bar charts because they give more space for long text.

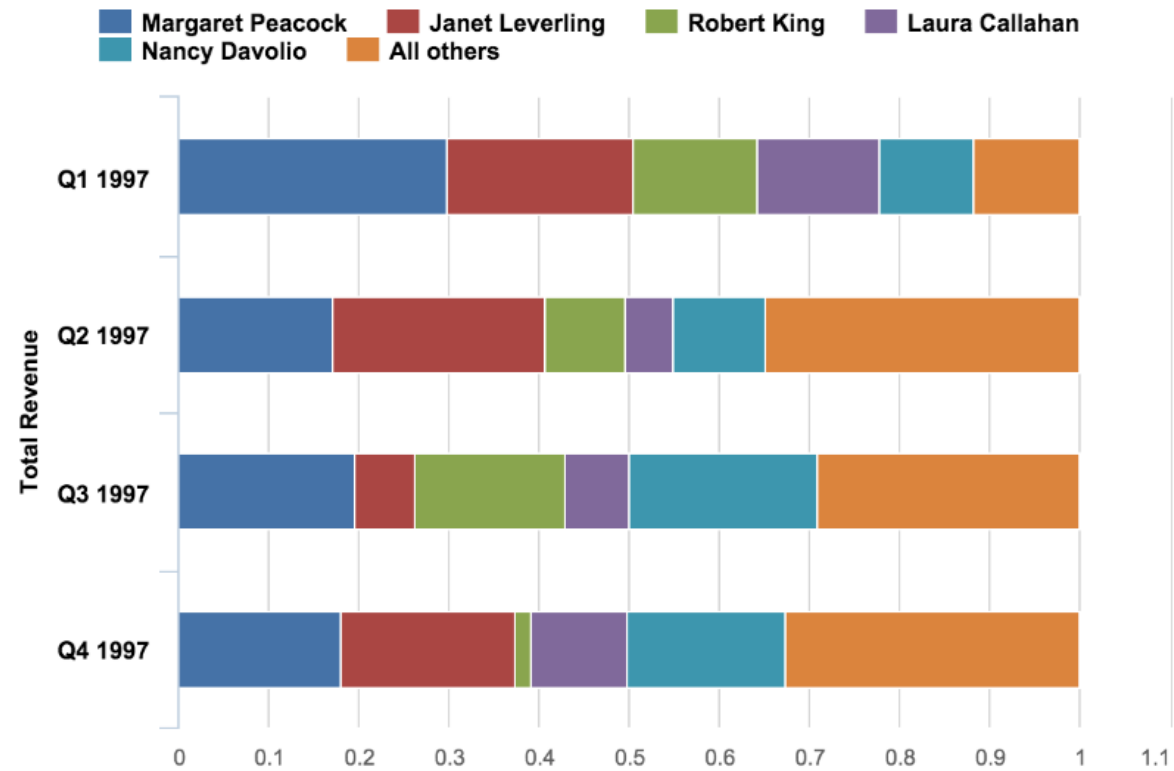
You should also use bar charts, instead of column charts, when the number of categories is greater than seven (but not more than fifteen) or for displaying a set with negative numbers.



# Bar Histogram Charts



# Stacked Bar Chart





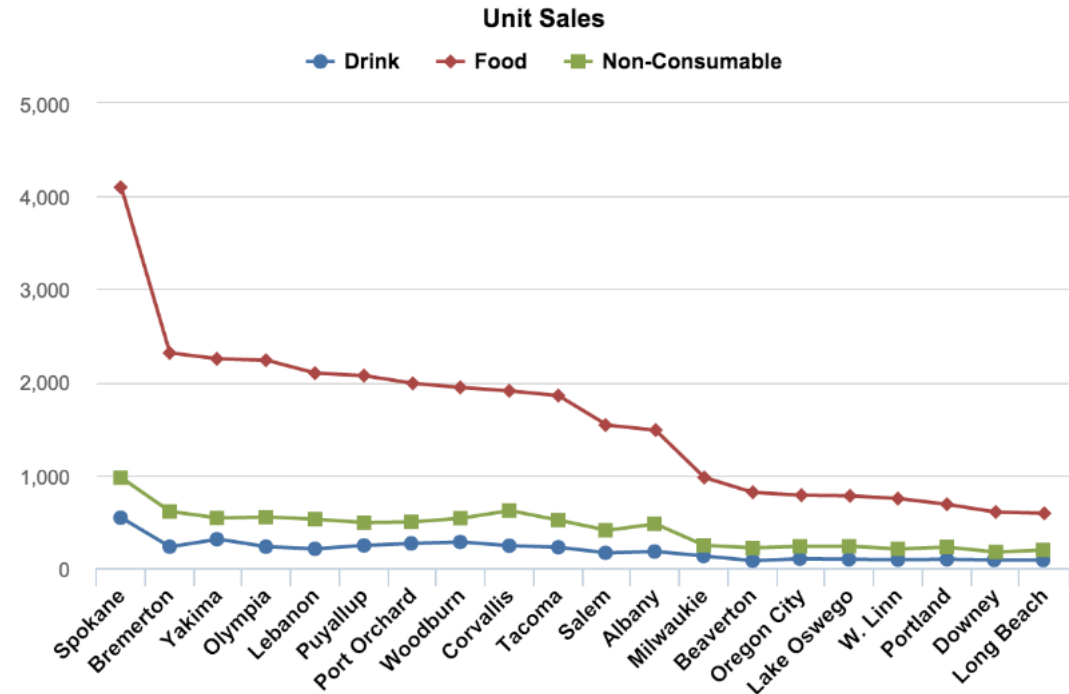
# Line Charts

Line charts are among the most frequently used chart types. Use lines when you have a continuous data set.

These are best suited for trend-based visualizations of data over a period of time, when the number of data points is very high (more than 20).

With line charts, the emphasis is on the continuation or the flow of the values (a trend), but there is still some support for single value comparisons, using data markers (only with less than 20 data points.)

A line chart is also a good alternative to column charts when the chart is small.

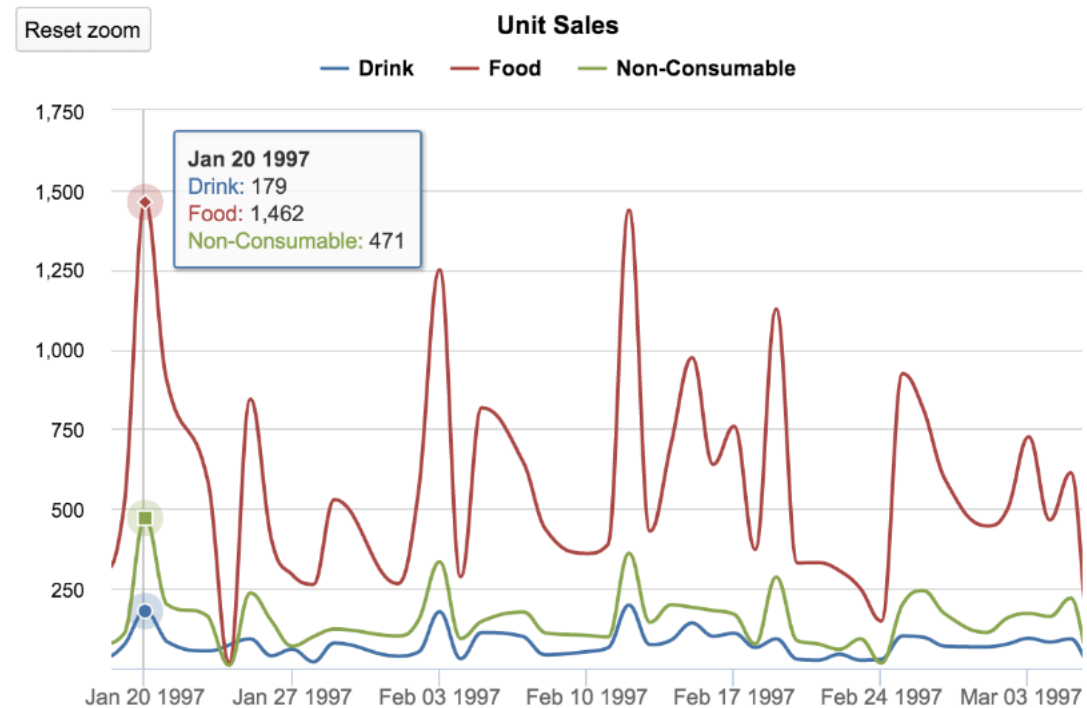


# Timeline Chart

The timeline chart is a variation of line charts.

Any line chart that shows values over a period of time is a timeline chart.

The only difference is in functionality — most timeline charts will let you zoom in and out and compress or stretch the time axis to see more details or overall trends.

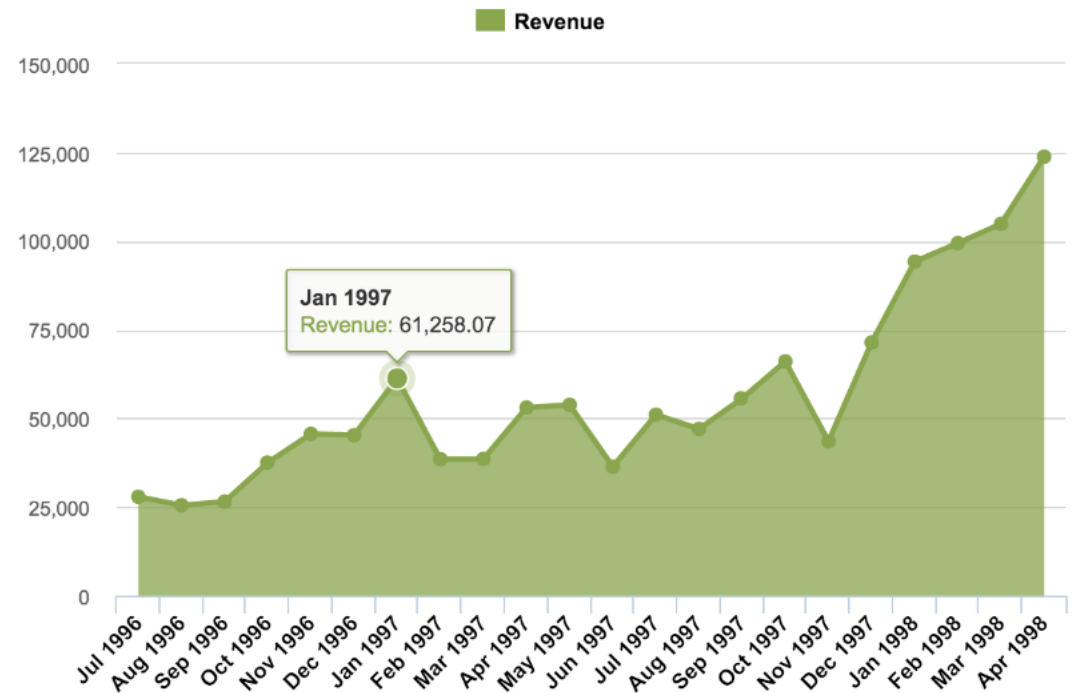


# Dos and Don'ts for Line Charts

- Use lines to present continuous data in an interval scale, where intervals are equal in size.
- For line charts, the axis may not start from zero if the intended message of the chart is the rate of change or overall trend, not exact values or comparison.
- It's best to start the axis with zero for wide audiences because some people may otherwise interpret the chart incorrectly.
- In line charts, time should always run from left to right.
- Do not skip values for consistent data intervals presenting trend information, for example, certain days with zero values.
- Remove guidelines to emphasize the trend, rate of change, and to reduce distraction.

# Area Chart

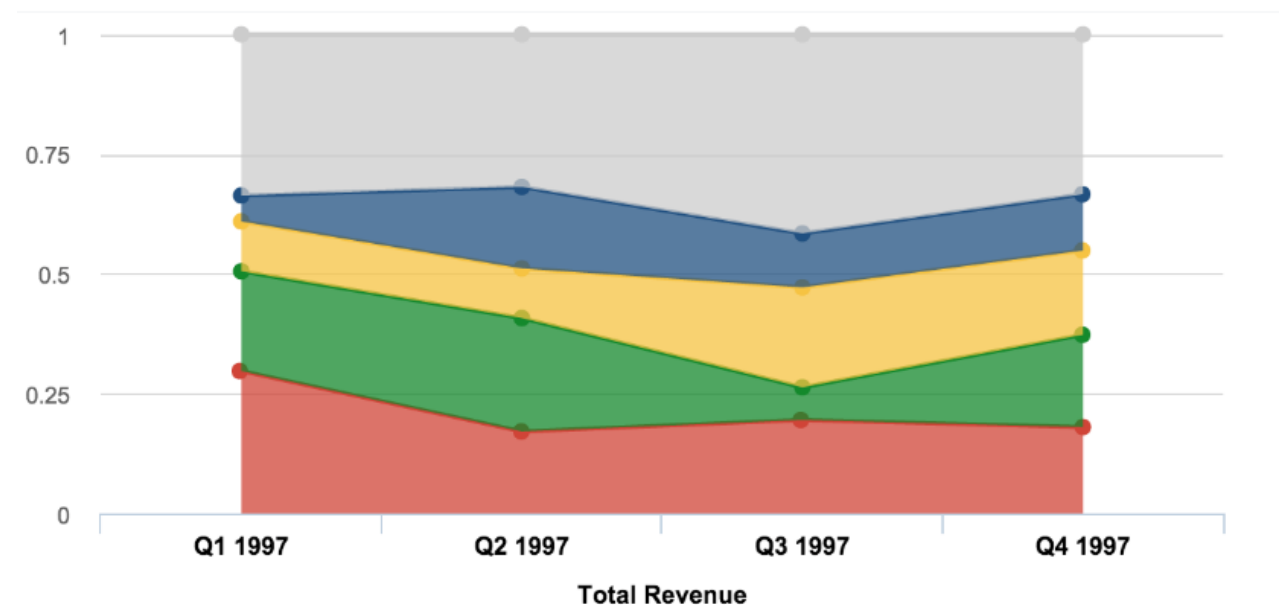
An area chart is essentially a line chart — good for trends and some comparisons. Area charts will fill up the area below the line, so the best use for this type of chart is for presenting accumulative value changes over time, like item stock, number of employees, or a savings account. Do not use area charts to present fluctuating values, like the stock market or prices changes.



# Stacked Area Chart

Stacked area charts are best used to show changes in composition over time.

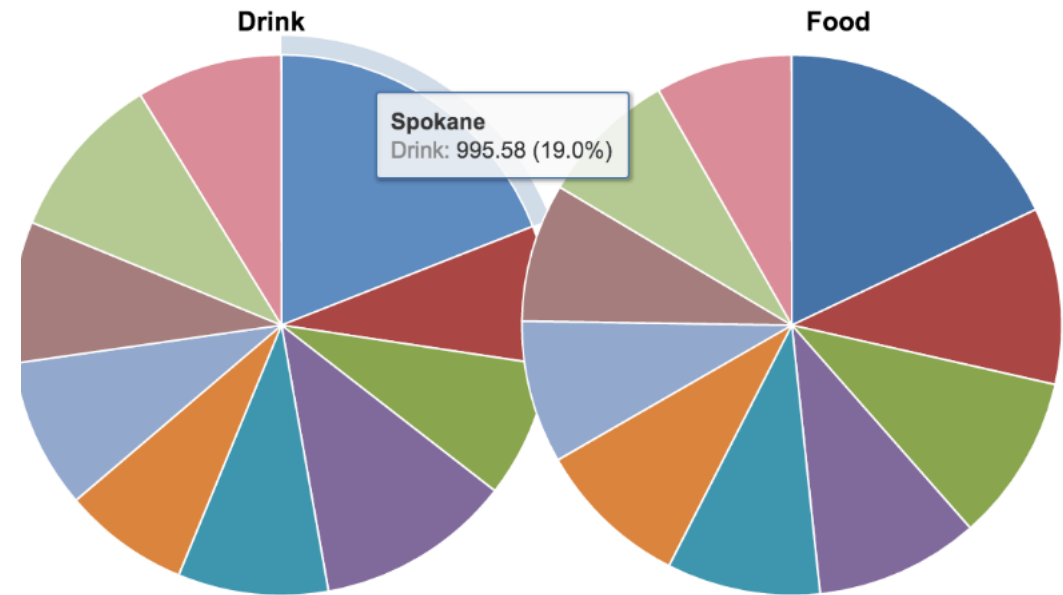
Stacked area charts might be colorful and fun, but you should use them with caution, because they can quickly become a mess. Don't use them if you need an exact comparison and don't stack together more than three to five categories.



# Pie Chart

A pie chart typically represents numbers in percentages, used to visualize a part to whole relationship or a composition.

Pie charts are not meant to compare individual sections to each other or to represent exact values (you should use a bar chart for that).



*Above is a good example of a terrible, useless pie chart - too many components, very similar values.*

# Dos and Don'ts for Pie charts

- Make sure that the total sum of all segments equals 100 percent.
- Use pie charts only if you have less than six categories, unless there's a clear winner you want to focus on.
- Ideally, there should be only two categories, like men and women visiting your website, or only one category, like a market share of your company, compared to the whole market.
- Don't use a pie chart if the category values are almost identical or completely different. You could add labels, but that's a patch, not an improvement.
- Don't use 3D or blow apart effects — they reduce comprehension and show incorrect proportions.

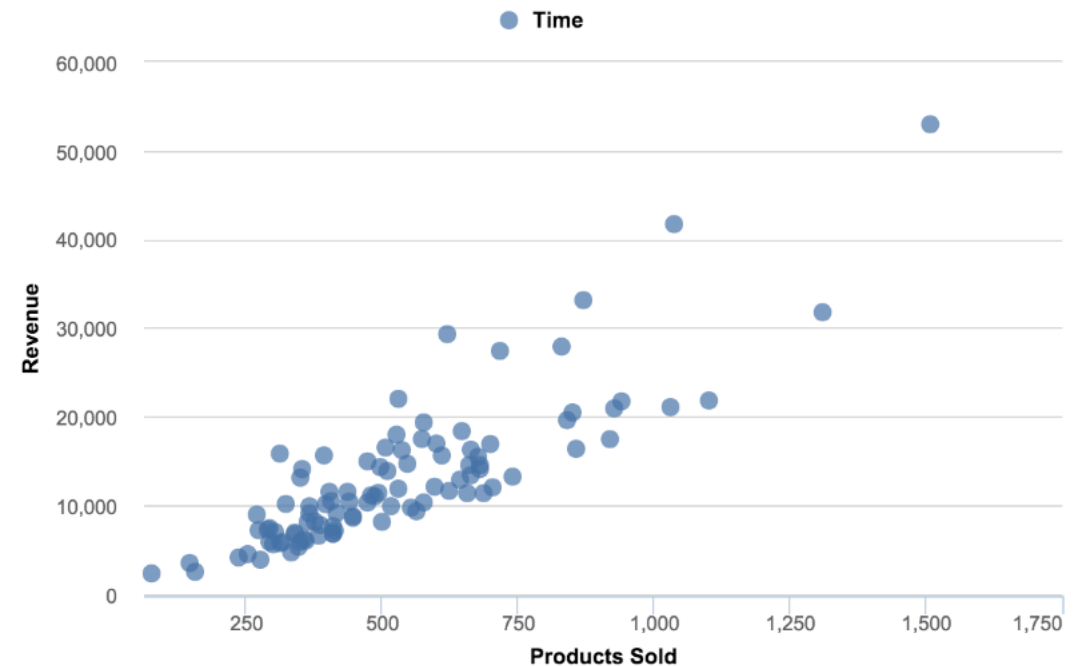
# Scatter Charts

Scatter charts are primarily used for correlation and distribution analysis.

Good for showing the relationship between two different variables where one correlates to another (or doesn't).

Scatter charts can also show the data distribution or clustering trends and help you spot anomalies or outliers.

A good example of scatter charts would be a chart showing marketing spending vs. revenue.



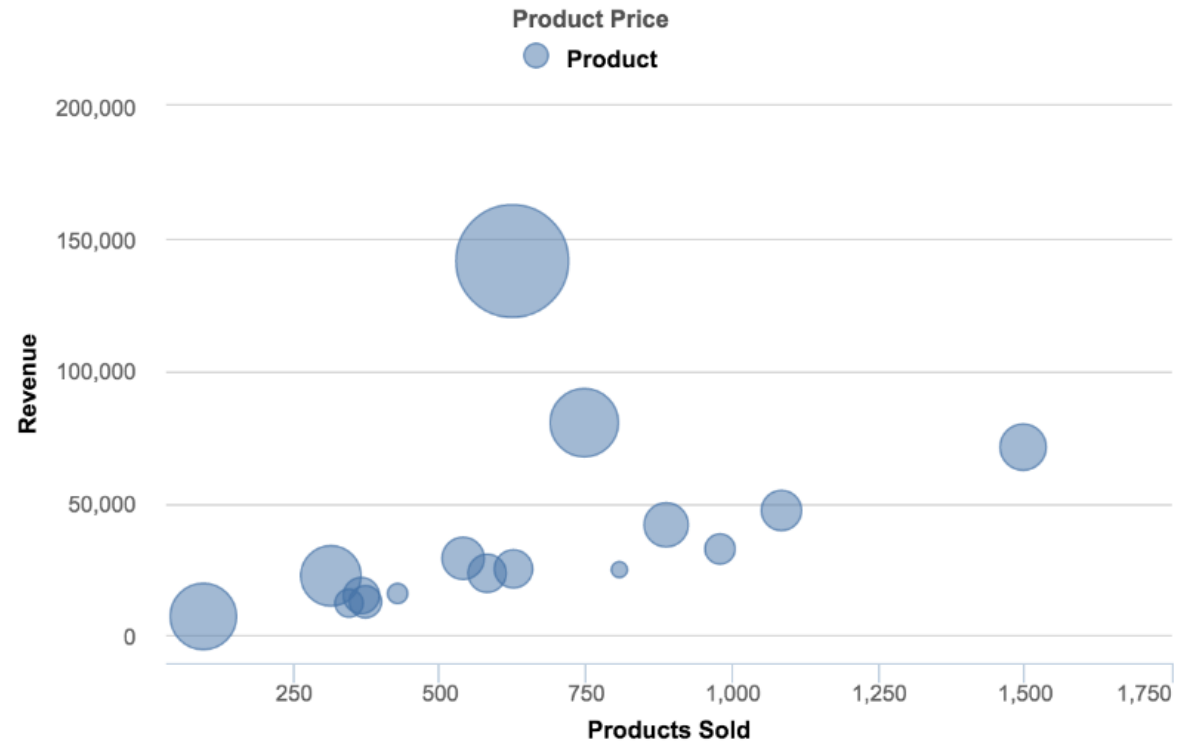


# Bubble Charts

A bubble chart is a great option if you need to add another dimension to a scatter plot chart.

Scatter plots compare two values, but you can add bubble size as the third variable and thus enable comparison.

If the bubbles are very similar in size, use labels.



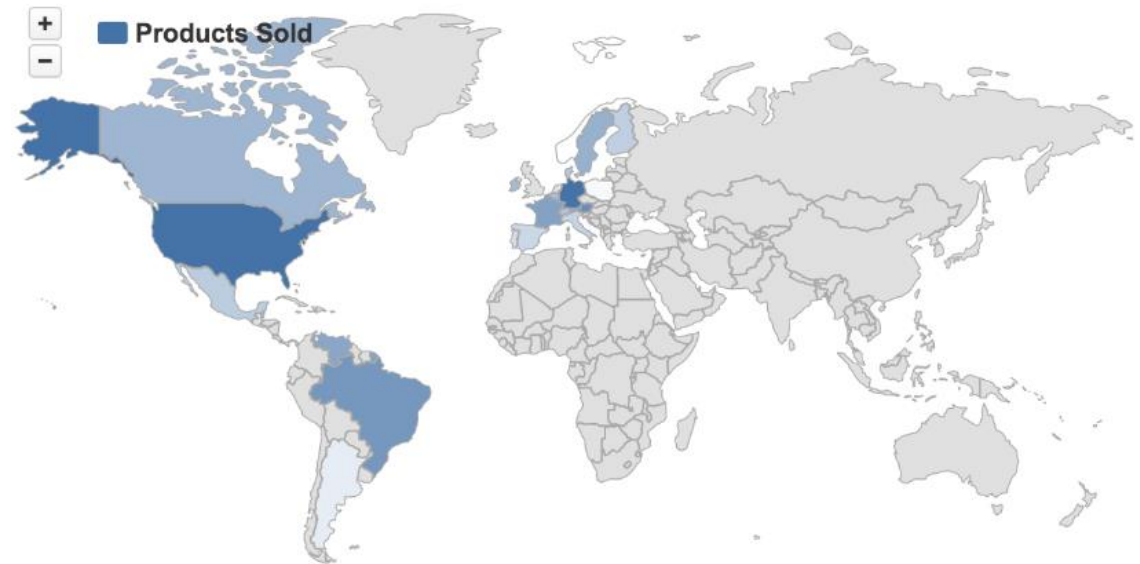
*We could in fact add the fourth variable by color-grading those bubbles*

# Dos and Donts in Bubble Chart

- Present relationships between two (scatter) or three (bubble) numerical variables,
- Plot two or three sets of variables on one x-y coordinate plane,
- Turn the horizontal axis into a logarithmic scale, thus showing the relationships between more widely distributed elements.
- Present patterns in large sets of data, linear or non-linear trends, correlations, clusters, or outliers.
- Compare large number of data points without regard to time. The more data you include in a scatter chart, the better comparisons you can make.
- Present relationships, but not exact values for comparisons.

# Map Charts

Map charts are good for giving your numbers a geographical context to quickly spot best and worst performing areas, trends, and outliers. If you have any kind of location data like coordinates, country names, state names or abbreviations, or addresses, you can plot related data on a map.



# Gantt Chart

A Gantt chart illustrates project progress and overview at a glance. The chart presents activities in a bar chart where separate bars represent each activity. A bar's length and positioning on the chart indicate when the activity began, its duration, and its planned completion date.



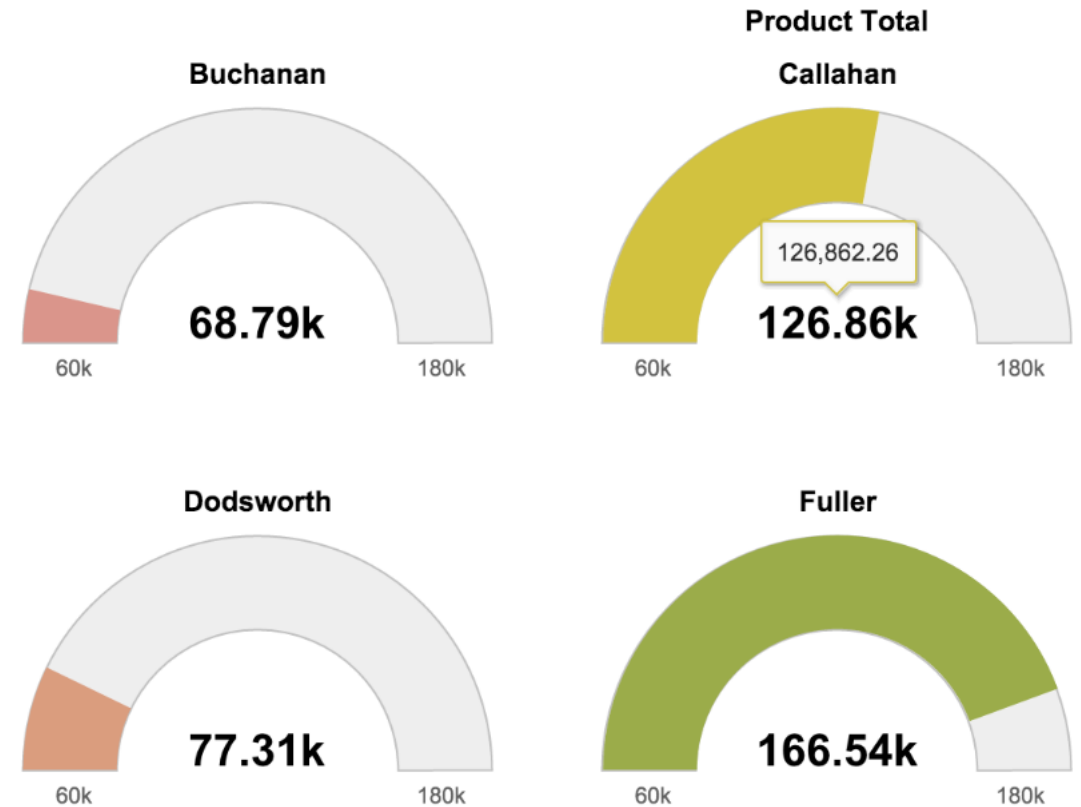
# Gauge Charts

Gauge charts are good for displaying KPIs (Key Performance Indicators). They typically display a single key value, comparing it to a color-coded performance level indicator, typically showing green for “good” and red for “trouble.”

A Dashboard would be the most obvious place to use Gauge charts. There, all the KPIs will be in one place and will give a quick “health check” for your project or company.

Gauges are a great choice to:

- Show progress toward a goal.
- Represent a percentile measure, like a KPI.
- Show an exact value and meaning of a single measure.
- Display a single bit of information that can be quickly scanned and understood

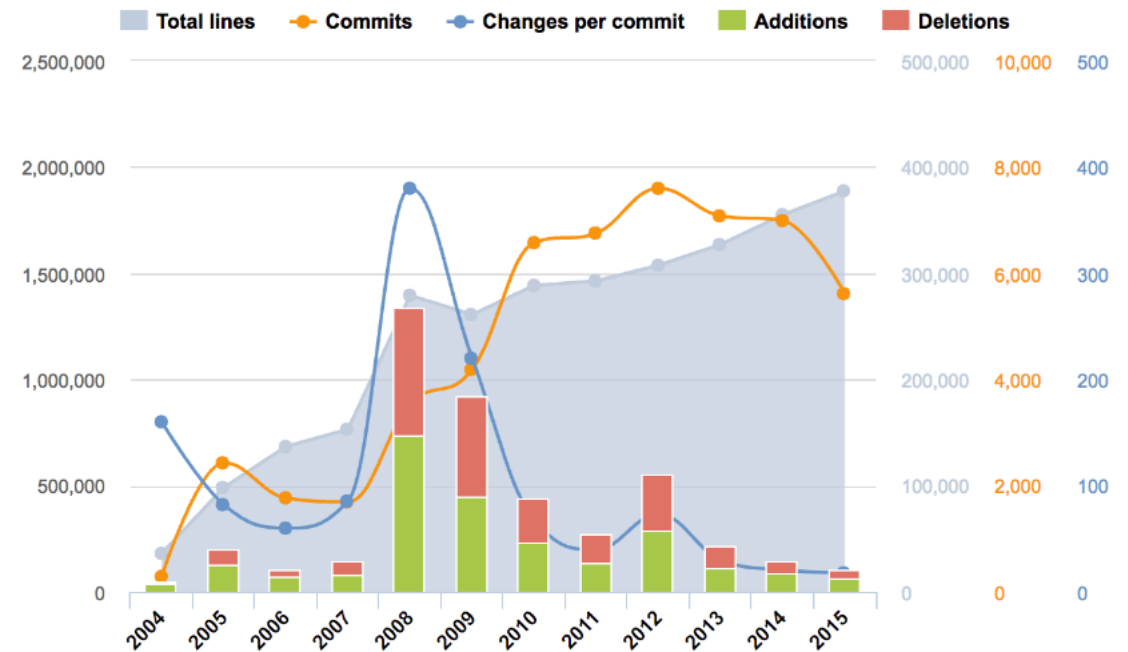


# Multi Axes Charts

There are times when a simple chart just cannot tell the whole story.

If you want to show relationships and compare variables on vastly different scales, the best option might be to have multiple axes.

A multi-axes chart will let you plot data using two or more y-axes and one shared x-axis. But it comes at a cost. That is, the charts are much more difficult to read and understand.



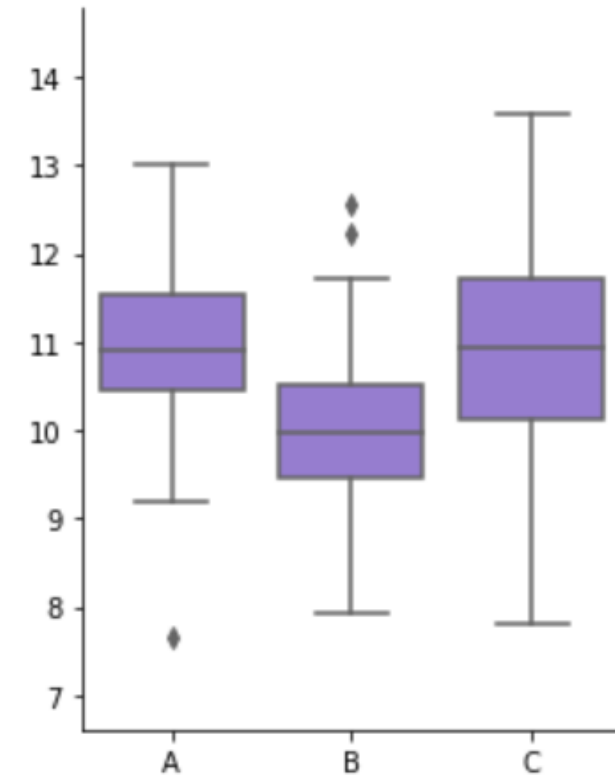
*Multi-axes charts might be good for presenting common trends, correlations (or the lack thereof) and the relationships between several data sets. But multi-axes charts are not good for exact comparisons (because of different scales) and you should not use this type if you need to show exact values.*

# Box Plot

A box plot uses boxes and whiskers to summarize the distribution of values within measured groups.

The positions of the box and whisker ends show the regions where the majority of the data lies.

We most commonly see box plots when we have multiple groups to compare to one another; other charts with more detail are preferred when we have only one group to plot.



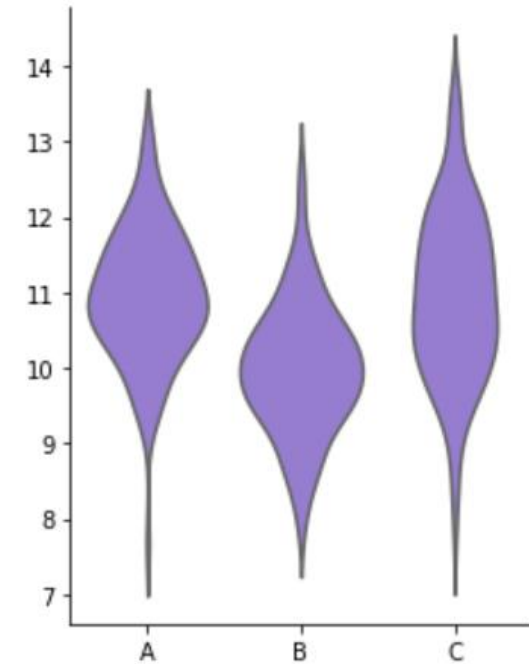
# Violin plot

An alternative to the box plot's approach to comparing value distributions between groups is the violin plot.

In a violin plot, each set of box and whiskers is replaced with a density curve built around a central baseline.

This can provide a better comparison of data shapes between groups, though this does lose out on comparisons of precise statistical values.

A frequent variation for violin plots is to include box-style markings on top of the violin plot to get the best of both worlds.

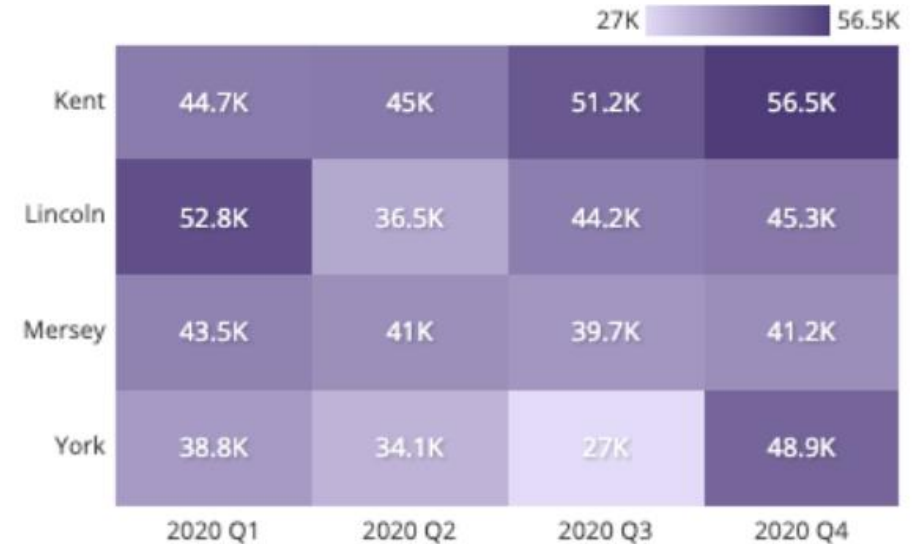




# Heatmap

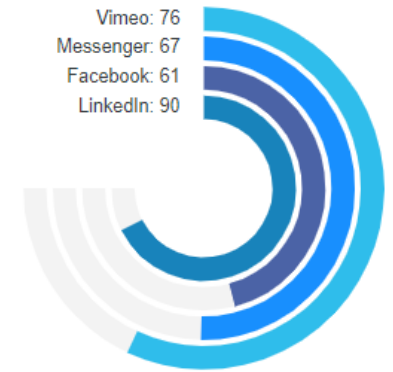
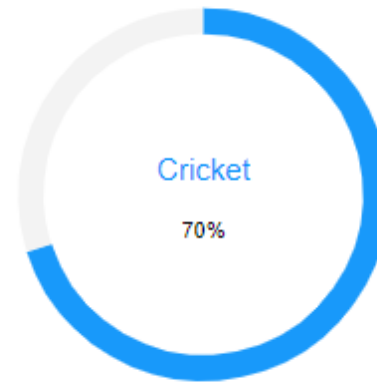
The [heatmap](#) presents a grid of values based on two variables of interest. The axis variables can be numeric or categorical; the grid is created by dividing each variable into ranges or levels like a histogram or bar chart. Grid cells are colored based on value, often with darker colors corresponding with higher values. A heatmap can be an interesting alternative to a scatter plot when there are a lot of data points to plot, but the point density makes it difficult to see the true relationship between variables.

New Revenue



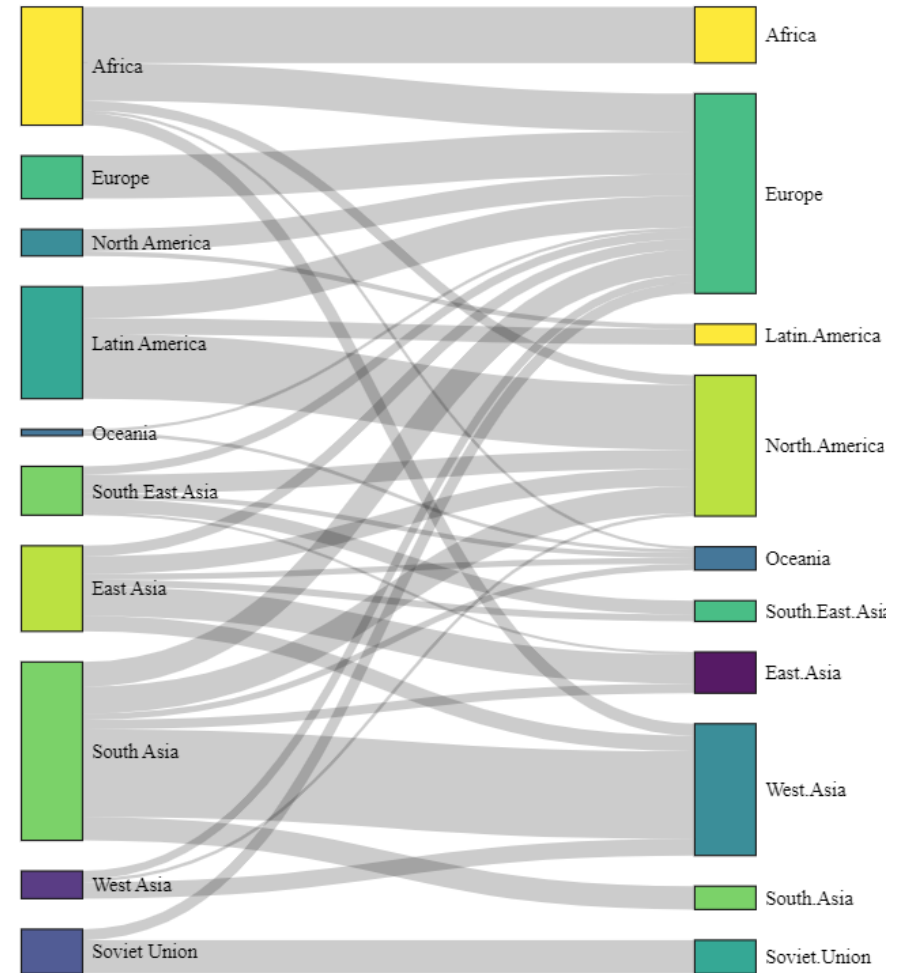
# Radial Graph

Radial Bar Charts are valuable in showing comparisons between categories by using circularly shaped bars. Also known as the circular bar chart, it is simply a typical bar chart plot represented on a polar coordinate system. JavaScript Radial Bar Charts are drawn by fixing a unique radius for each bar representing each data and each bar is judged by its angle. In ApexCharts, data can be represented on a radial bar chart in the various formats such as multiple radial bar charts, radial bar with an image, and even in semi-circular gauge forms.



# Sankey Diagram

A Sankey Diagram is a visualization technique that allows to display flows. Several entities (nodes) are represented by rectangles or text. Their links are represented with arrow or arcs that have a width proportional to the importance of the flow.



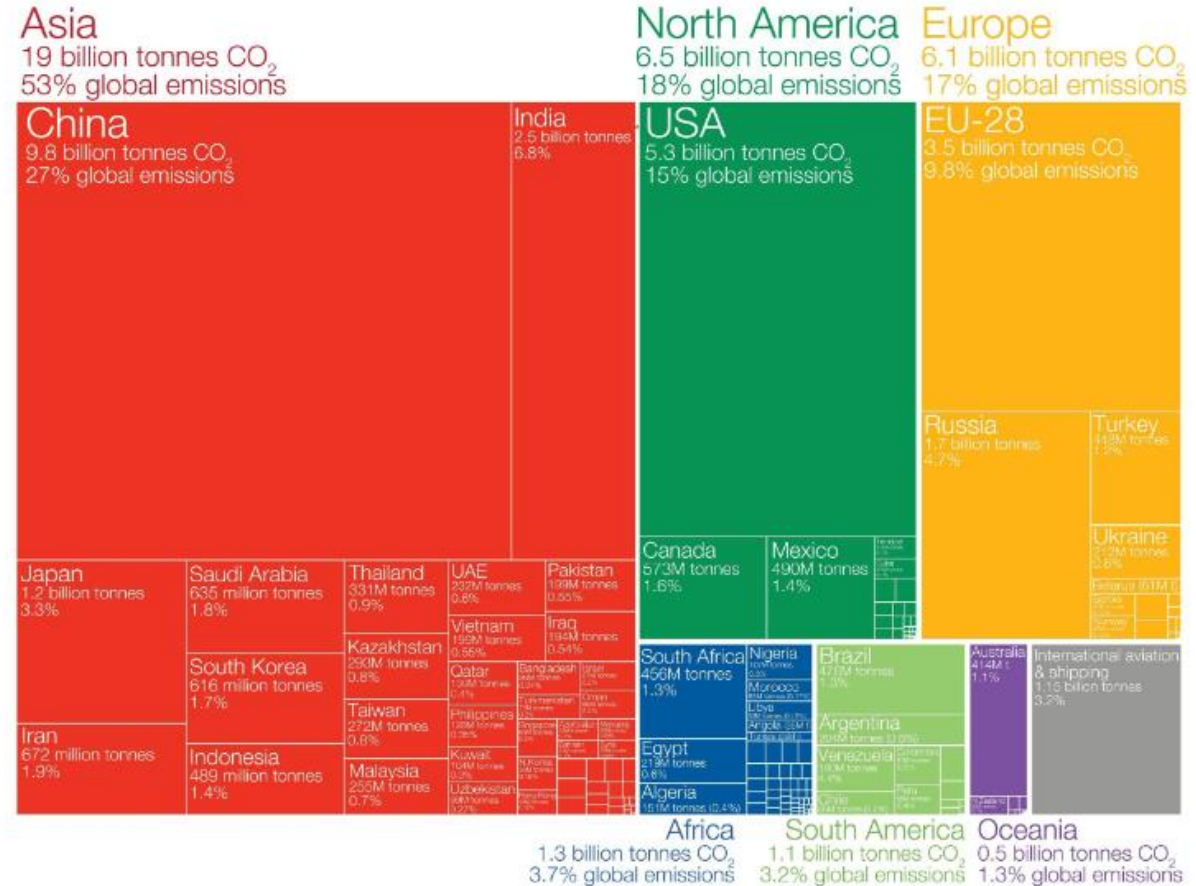
# Square Pie Chart

Square chart, also called waffle chart, are a form of pie charts that use squares instead of circles to represent percentages. Similar to basic circular pie charts, square pie chart take each percentage out of a total 100%. They are often 10 by 10 grids, where each cell represents 1%. One major benefit to square chart is that smaller percentages, difficult to see on traditional pie chart, can be easily depicted.

## Who emits the most CO<sub>2</sub>?

Global carbon dioxide (CO<sub>2</sub>) emissions were 36.2 billion tonnes in 2017.

Our World  
in Data



# Data Visualization Do's and Don'ts

- **Time axis.** When using time in charts, set it on the horizontal axis. Time should run from left to right. Do not skip values (time periods), even if there are no values.
- **Proportional values.** The numbers in a chart (displayed as bar, area, bubble, or other physically measured element in the chart) should be directly proportional to the numerical quantities presented.
- **Data-Ink Ratio.** Remove any excess information, lines, colors, and text from a chart that does not add value. More about [data-ink ratio](#)
- **Sorting.** For column and bar charts, to enable easier comparison, sort your data in ascending or descending order by the value, not alphabetically. This applies also to pie charts.
- **Legend.** You don't need a legend if you have only one data category.
- **Labels.** Use labels directly on the line, column, bar, pie, etc., whenever possible, to avoid indirect look-up.
- **Inflation adjustment.** When using monetary values in a long-term series, make sure to adjust for inflation. ([EU inflation rates](#), [US inflation rates](#))
- **Colors.** In any chart, don't use more than six colors.
- **Colors.** For comparing the same value at different time periods, use the same color in a different intensity (from light to dark).
- **Colors.** For different categories, use different colors. The most widely used colors are black, white, red, green, blue, and yellow.
- **Colors.** Keep the same color palette or style for all charts in the series, and same axes and labels for similar charts to make your charts consistent and easy to compare.
- **Colors.** Check how your charts would look when printed out in gray-scale. If you cannot distinguish color differences, you should change hue and saturation of colors.
- **Colors.** Seven to 10 percent of men have color deficiency. Keep that in mind when creating charts, ensuring they are readable for color-blind people.
- **Data Complexity.** Don't add too much information to a single chart. If necessary, split data in two charts, use highlighting, simplify colors, or change chart type.