DATA SCIENCE

Introduction to Data Visualization

Basic Charts



UNC GREENSBORO

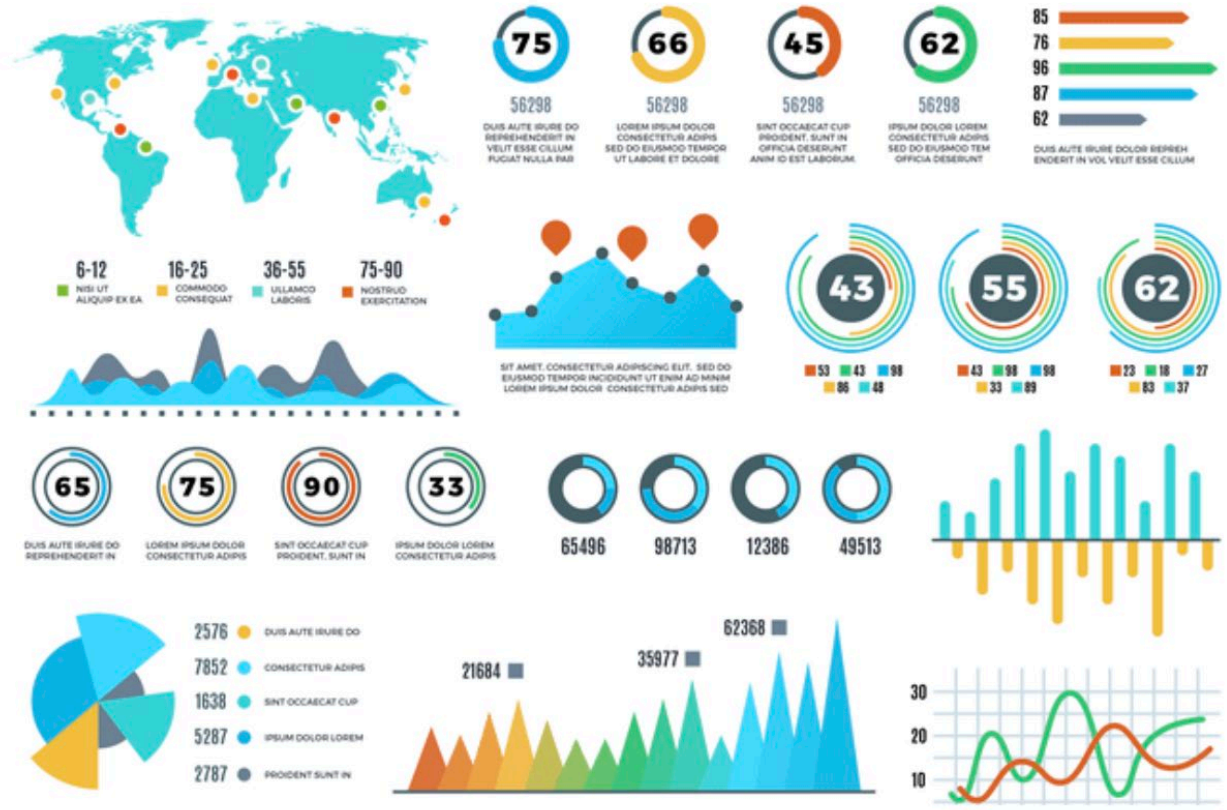
Visualization

What is data visualization?

Example. Consider a case where you are asked to illustrate crucial sales aspects (like sales performance, target, revenue, acquisition cost, etc.) from huge amounts of sales data, which one would you prefer:

1. Exploring the data using Excel (or spreadsheets) and keeping track of each sales aspect manually.

2. Exploring the data using different types of sales graphs and charts.



Obviously, you would prefer graphs and charts.
So, data visualization plays a key role in
data exploration and data analysis.

UNC GREENSBORO

Visualization

What is data visualization?

"A picture is worth a thousand words."

Data visualization represents data through visual forms such as charts, graphs, colors, shapes and maps.

> It helps people see patterns, trends, relationships and differences that may be difficult to notice in raw values.

Key idea: A chart should make data easier to understand and interpret.

Visualization

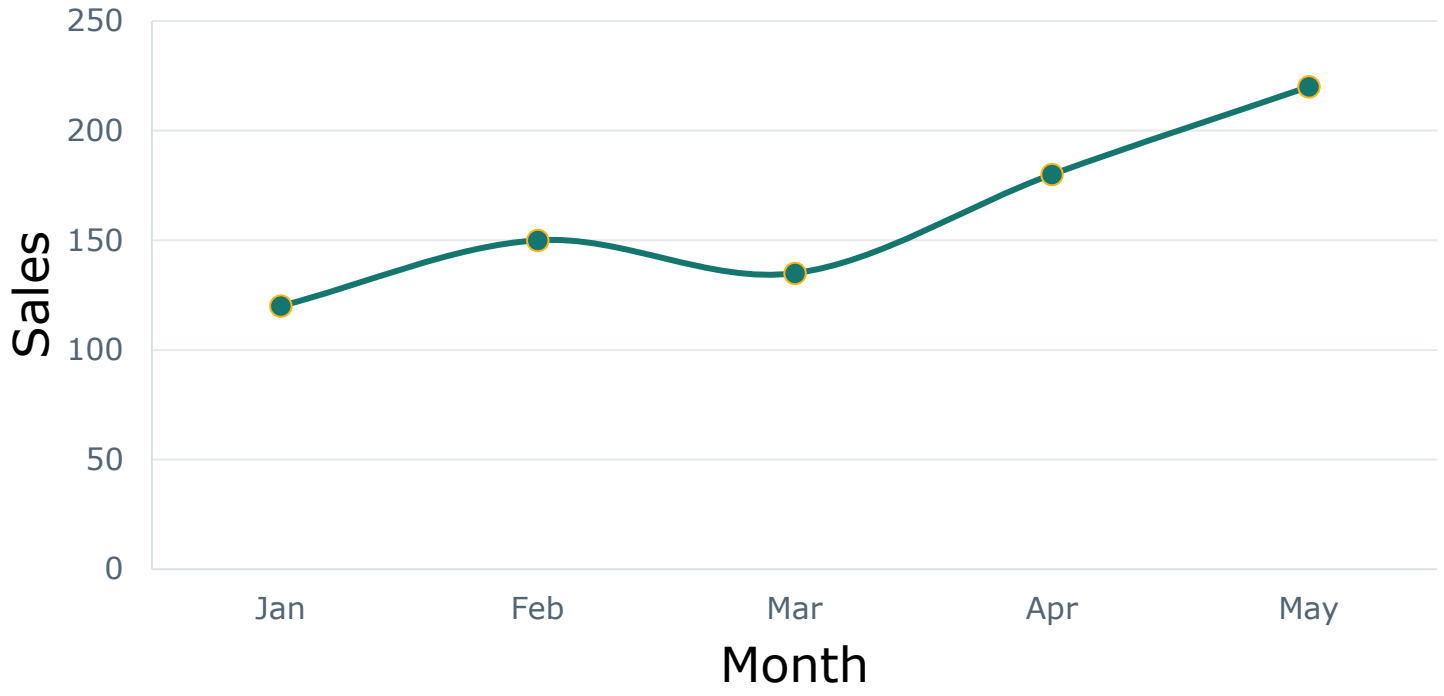
Four reasons to visualize data

01	**Find patterns**	See trends, clusters or unusual values. Example: seasonal sales.
02	**Compare values**	Compare categories, groups or periods. Example: product sales.
03	**Explore relationships**	Look for associations and possible outliers. Example: advertising and revenue.
04	**Communicate findings**	Explain evidence so a team can make a decision. Example: business performance.

Visualization

Monthly sales show an overall rise

Month	Sales
January	120
February	150
March	135
April	180
May	220



May is highest. Sales rise overall, with a dip in March.

UNC GREENSBORO

Visualization

Common Types of Data Visualization

- **Charts and Graphs**: They are used to visualize data, with charts comparing data points across categories or showing trends over time and graphs analyzing relationships between variables to identify correlations, trends and outliers.
  Examples: Bar Charts, Line Charts, Pie Charts, Scatter Plots, Histograms, Box Plots.
- **Maps**: They are used to display geographical data which provides spatial context to trends and patterns.
  Examples: Geographic Maps, Heat Maps
- **Dashboards**: They combine multiple visualizations into a single interface which provides real-time insights and interactive features for users to explore data.

Chart families follow analysis goals

Analysis goal	Common chart families	Typical question
Compare categories	Column, bar	Which category is larger?
Show change over time	Line, area	How does a value change?
Show composition	Stacked column, pie	How does a total break down?
Explore relationships	Scatter, bubble	Do two or three measures vary together?
Explain change or progress	Waterfall, funnel, bullet	How did a total change, or where do we lose users?
Find patterns in a matrix	Heat map	Where are values unusually high or low?

Visualization

Visualization connects to statistical methods

Previously learned concept	Related visualization	What it helps reveal
Variable types	Bar chart, histogram	Categories or numeric distribution
Descriptive statistics	Bar chart, box plot	Group differences, center and spread
Probability distributions	Histogram, density plot, Q–Q plot	Shape, skew and distribution fit
Missing values and outliers	Missingness heat map, box plot, scatter plot	Where gaps or extreme values appear
Correlation	Scatter plot, correlation heat map	Pairwise patterns among numeric variables
Hypothesis testing	Group box plot, violin plot, confidence interval plot	Difference size and uncertainty
Regression and classification	Regression/residual plot, confusion matrix, ROC curve	Fit, error patterns and class performance

Visualization

Statistics and visuals answer complementary questions

VISUALIZATION

Shows shape, variation and individual observations.

STATISTICAL METHOD

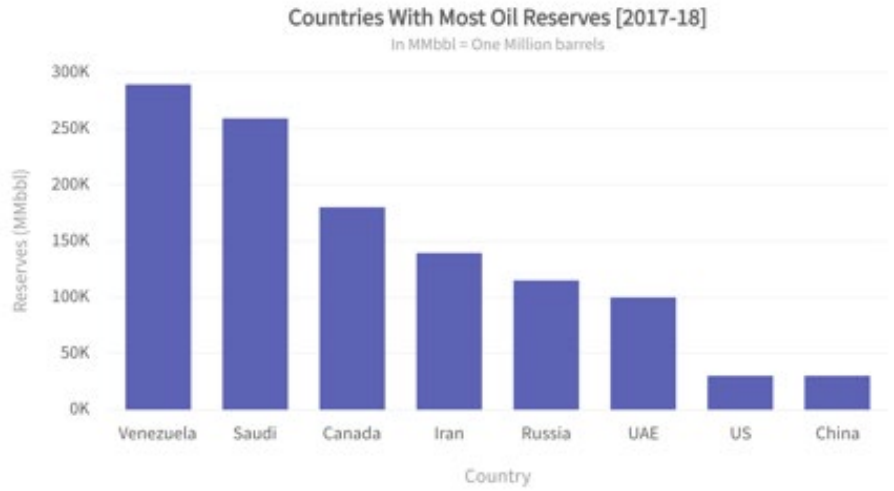
Quantifies patterns, estimates uncertainty and tests claims.

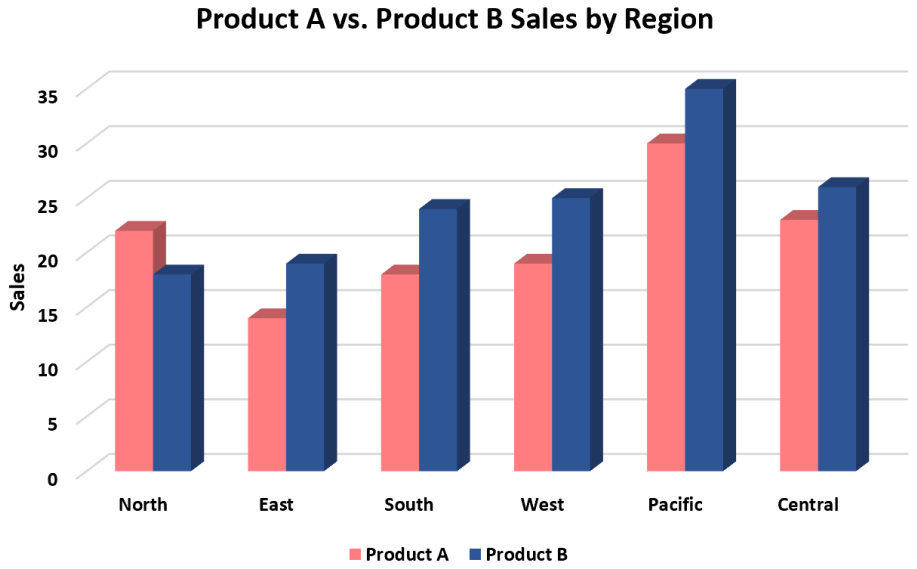
The chart shows what the summary can miss. The method quantifies the evidence.

Different Types of Basic Charts

1) Column Chart

A column chart is a vertical bar graph used to compare discrete categories or show data changes over time, with categories on the horizontal axis and values on the vertical axis.



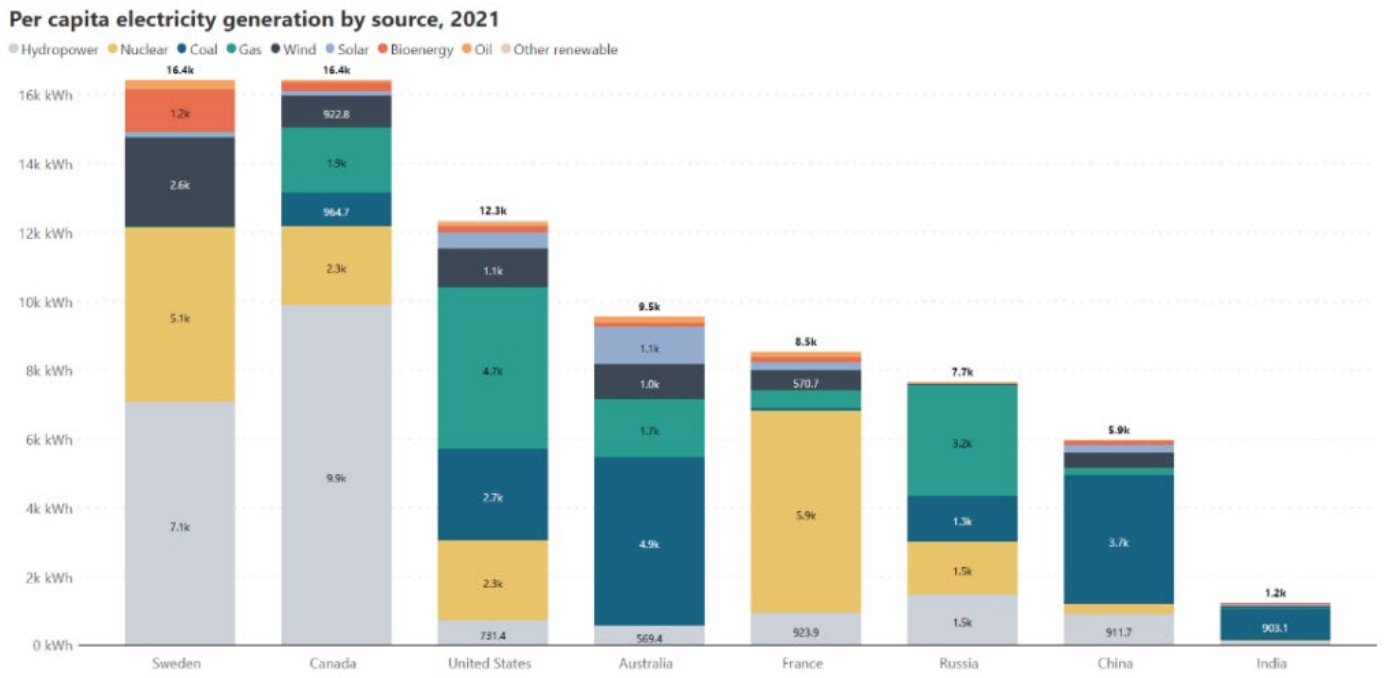


Different Types of Basic Charts

1) Column Chart

A column chart is a vertical bar graph used to compare discrete categories or show data changes over time, with categories on the horizontal axis and values on the vertical axis.

Stacked column charts

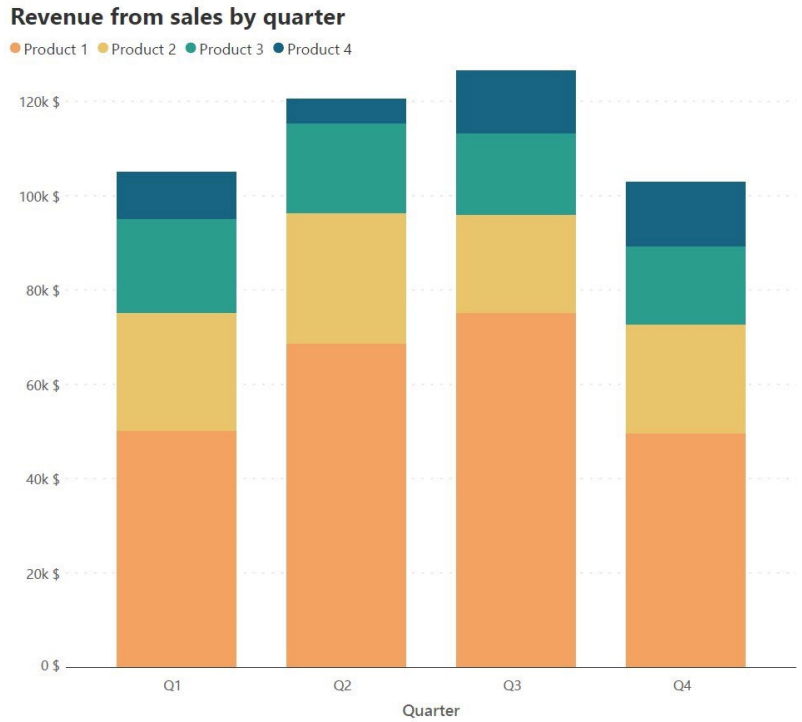


UNC GREENSBORO

Different Types of Basic Charts

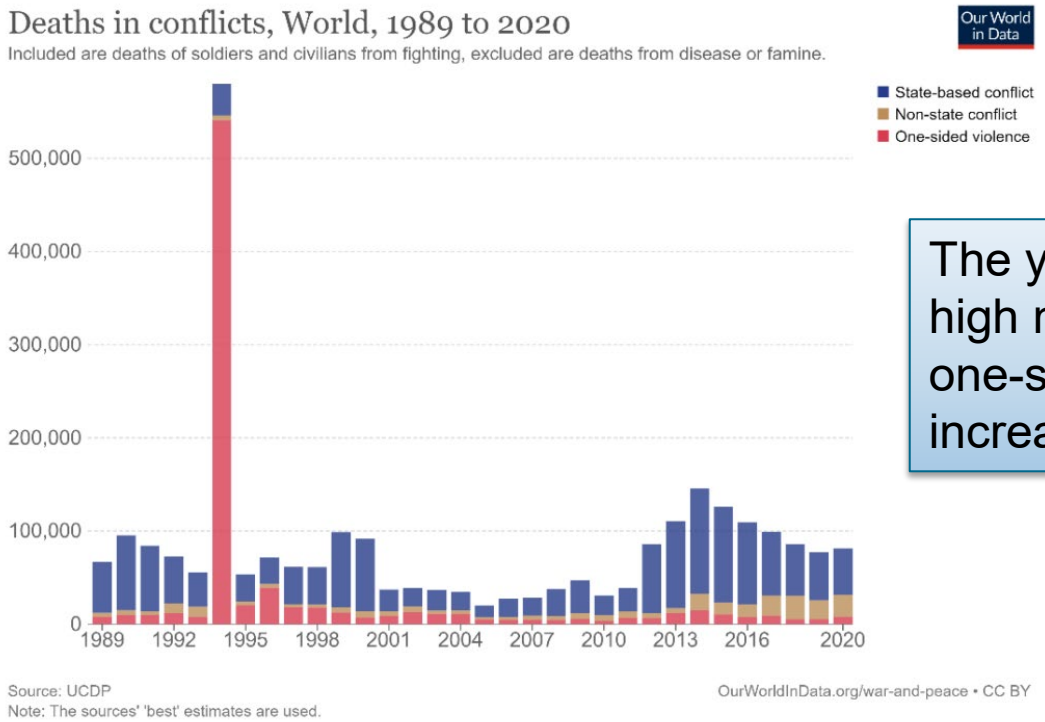
1) Column Chart: Stacked column charts: the primary use for stacked column charts is in representing data where values are divided into contributing categories that sum to the total. This means that the data naturally has two categorical variables, one of which we designate as the primary variable to plot as columns, and the second which we use to divide the columns into segments.

Quarter	Product 1 ($)	Product 2 ($)	Product 3 ($)	Product 4 ($)
Q1	50000	25000	20000	10000
Q2	68500	27750	19000	5300
Q3	75000	20750	17290	13500
Q4	49500	23000	16600	13800



Different Types of Basic Charts

1) Column Chart: Stacked column charts: the primary use for stacked column charts is in representing data where values are divided into contributing categories that sum to the total. This means that the data naturally has two categorical variables, one of which we designate as the primary variable to plot as columns, and the second which we use to divide the columns into segments.

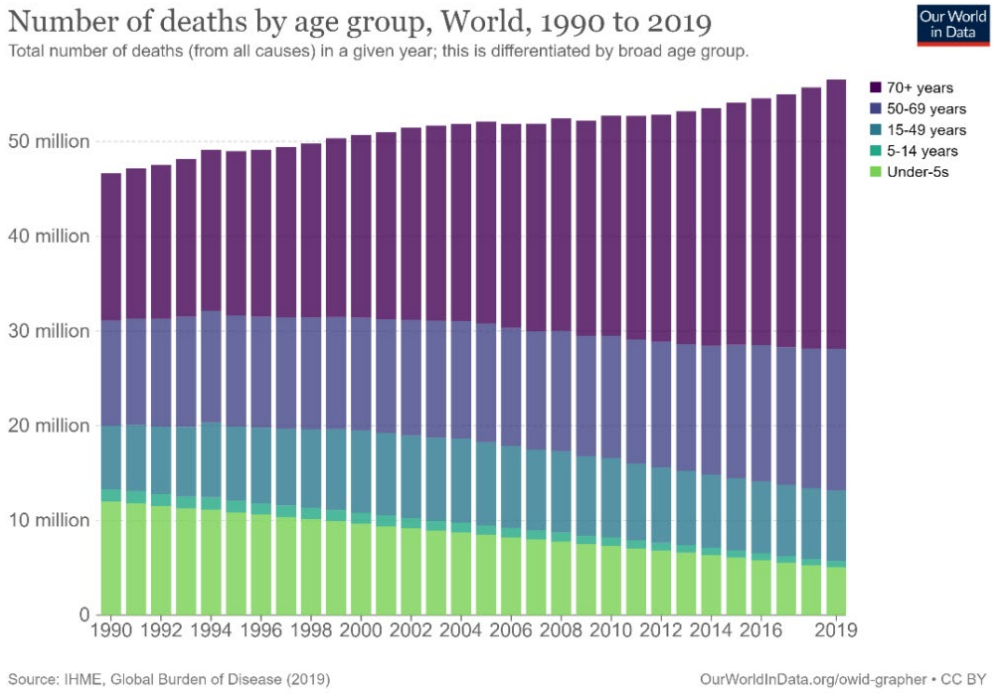


The year 1994 jumps out as being one with a disproportionately high number of deaths, and it is apparent that the escalation in one-sided violence is the main cause for this tremendous increase.

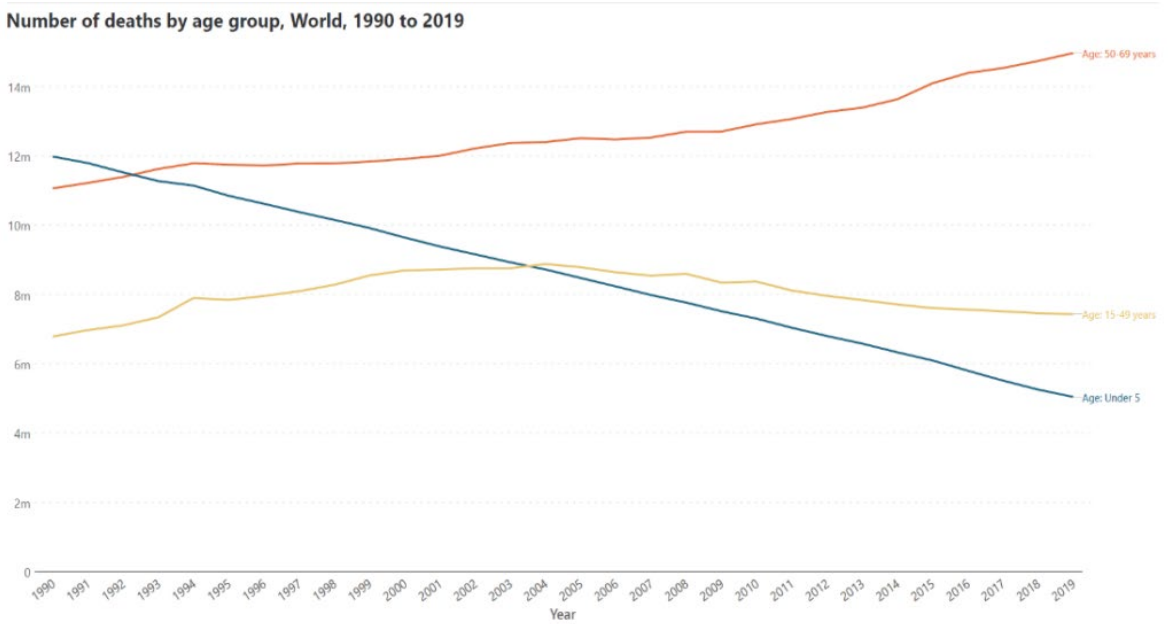
UNC GREEN

Different Types of Basic Charts

1) Column Chart: Stacked column charts: the primary use for stacked column charts is in representing data where values are divided into contributing categories that sum to the total. This means that the data naturally has two categorical variables, one of which we designate as the primary variable to plot as columns, and the second which we use to divide the columns into segments.



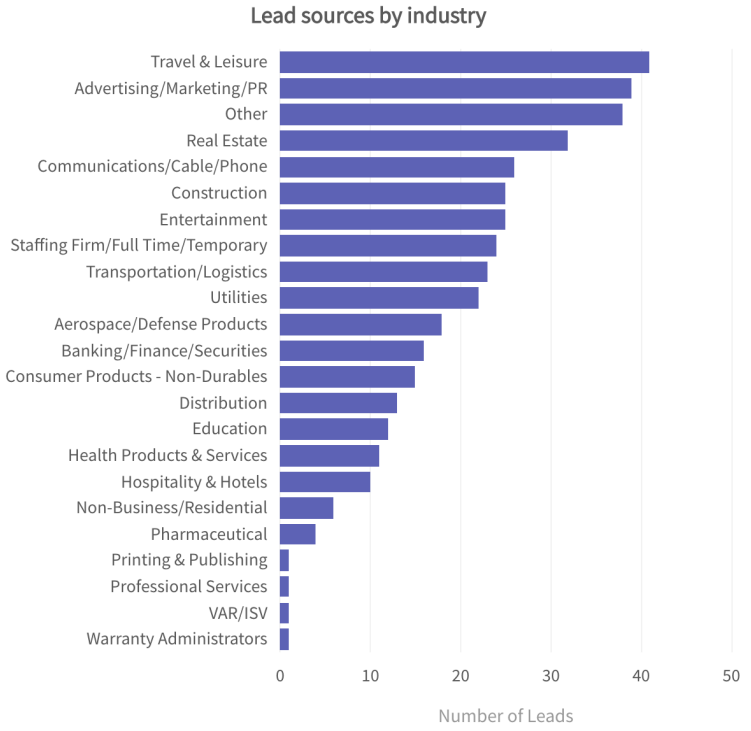
Stacked column charts are not always a good option

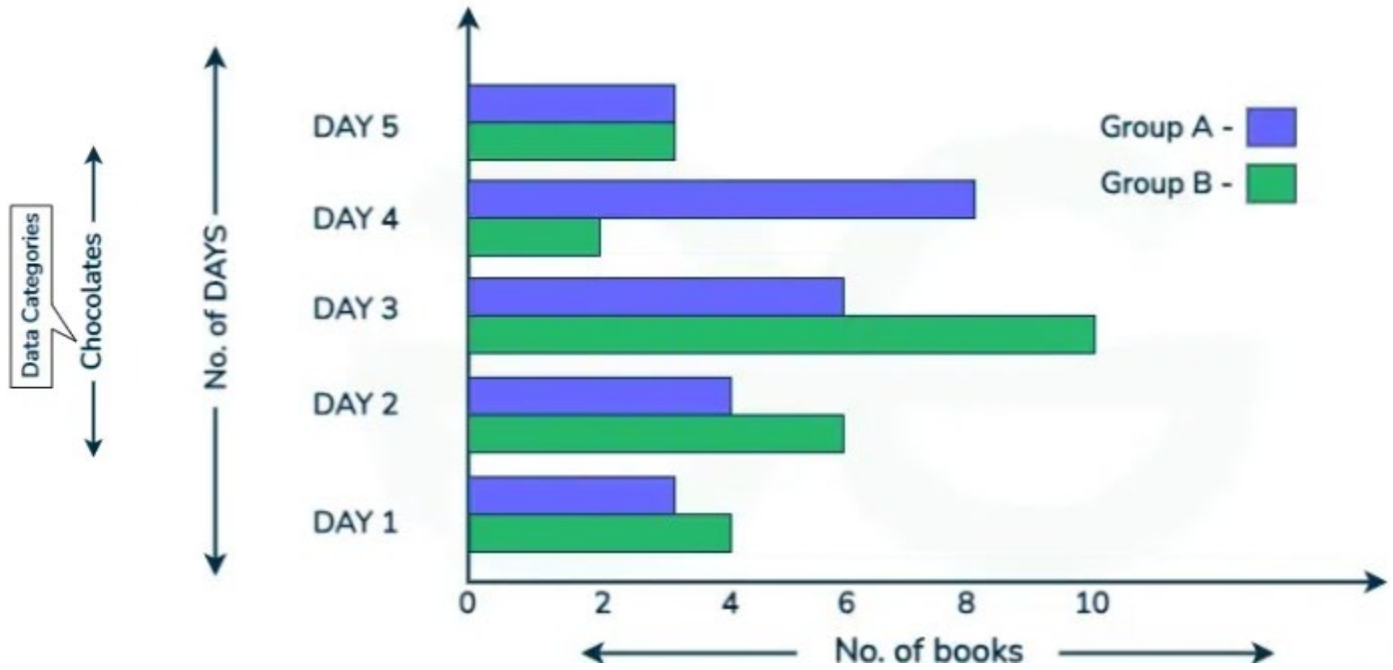


Different Types of Basic Charts

2) Bar Chart

A bar chart (a horizontal column chart), should be used to avoid clutter when one data label is long or if you have more than 10 items to compare. This type of visualization can also be used to display negative numbers.

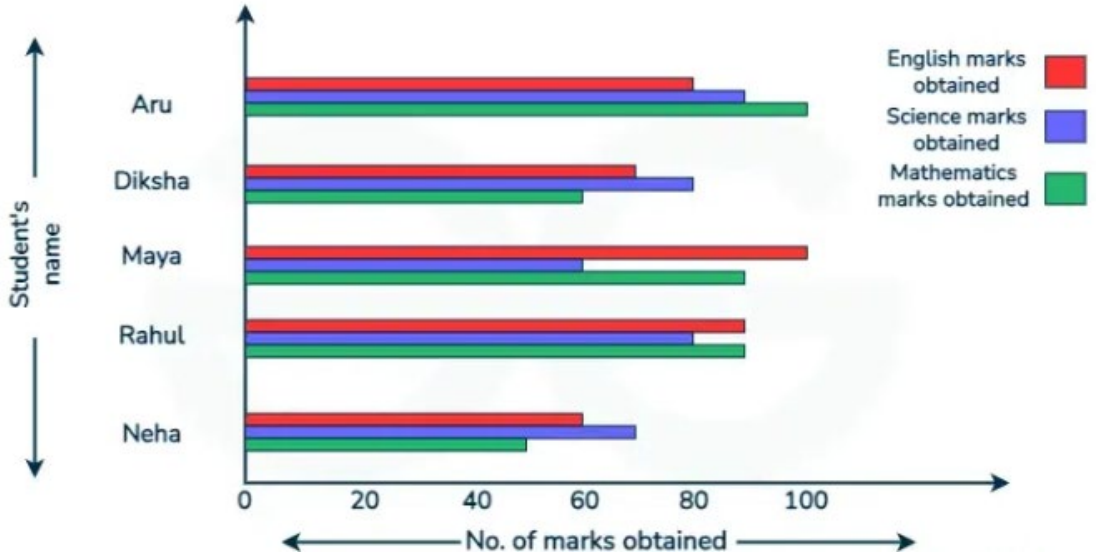




Different Types of Basic Charts

2) Bar Chart

A bar chart (a horizontal column chart), should be used to avoid clutter when one data label is long or if you have more than 10 items to compare. This type of visualization can also be used to display negative numbers.

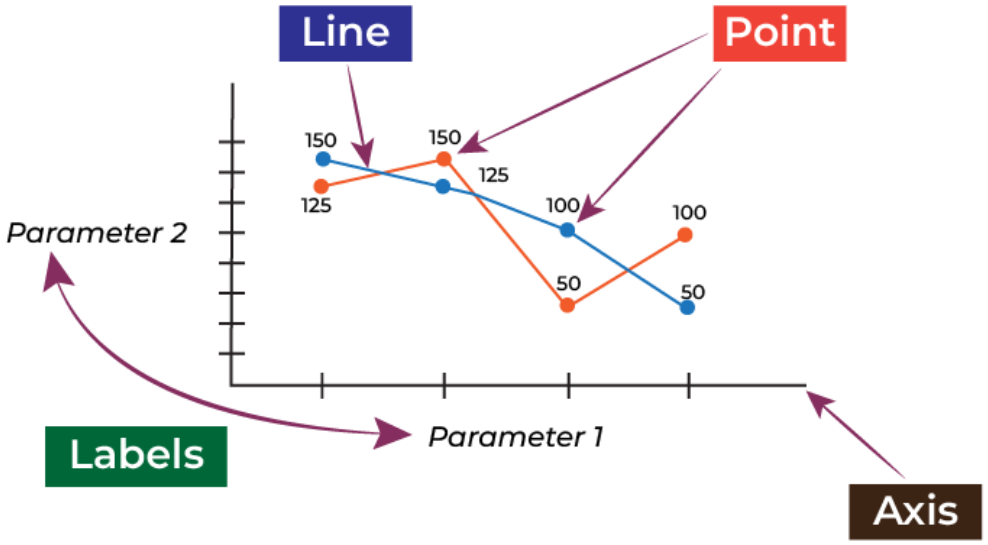


- How many marks were obtained out of 100 by Neha in English?
- Who scored the highest marks in Mathematics?
- Name the students who obtained less than 80 in Science.

Different Types of Basic Charts

3) Line Chart: is a graphical representation of the data that displays the relationship between two or more variables concerning time. It is made by connecting data points with straight-line segments.

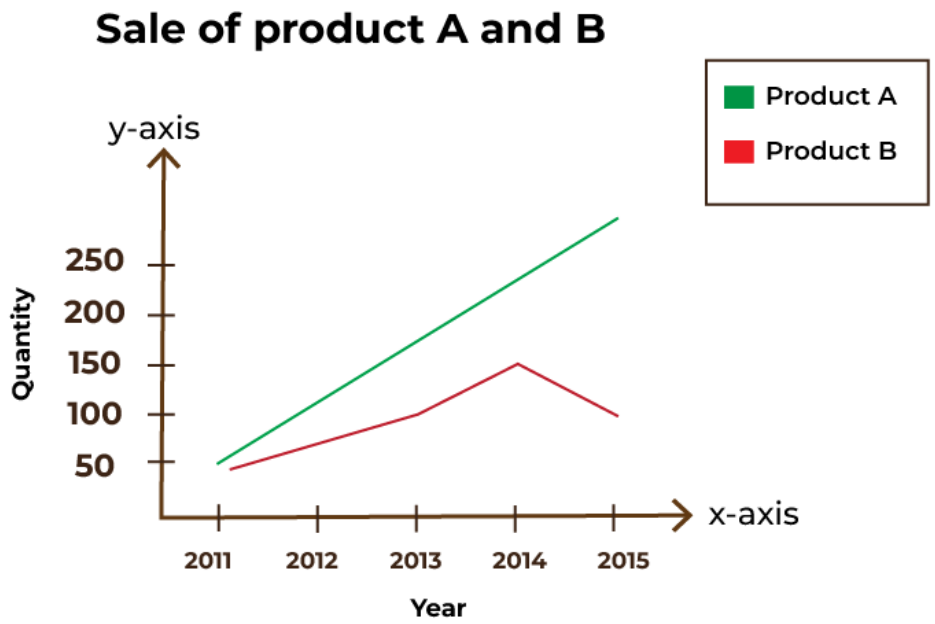
A line chart reveals trends or progress over time and can be used to show many different categories of data. You should use it when you chart a continuous data set.

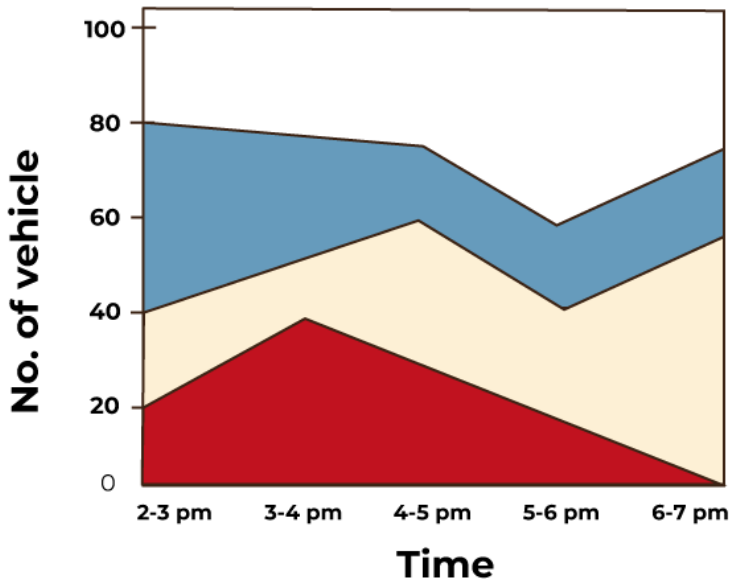


Different Types of Basic Charts

3) Line Chart: is a graphical representation of the data that displays the relationship between two or more variables concerning time. It is made by connecting data points with straight-line segments.

A line chart reveals trends or progress over time and can be used to show many different categories of data. You should use it when you chart a continuous data set.

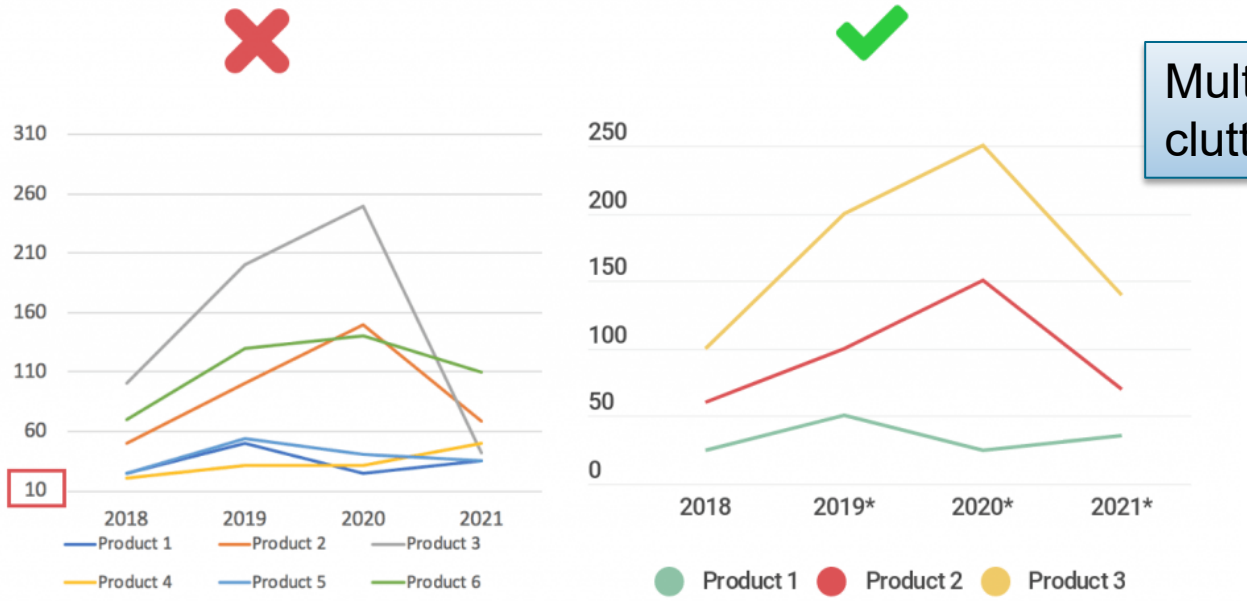




Different Types of Basic Charts

3) Line Chart

Some tips:

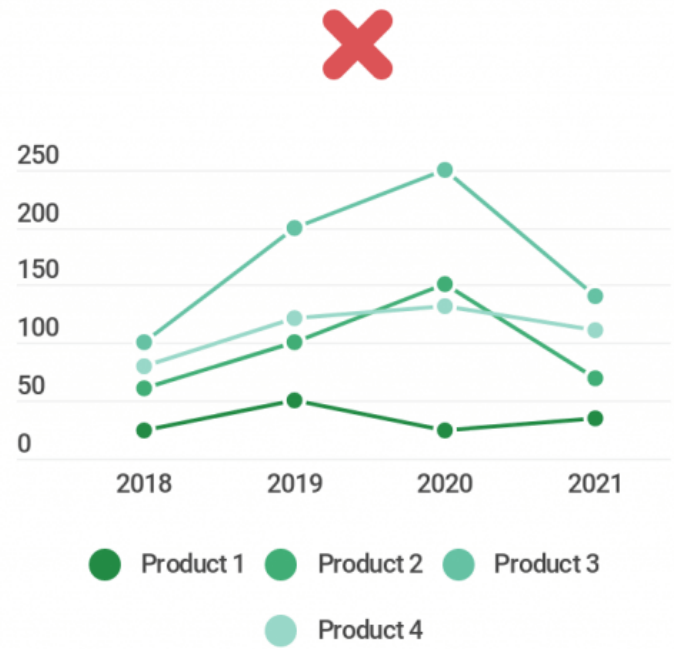


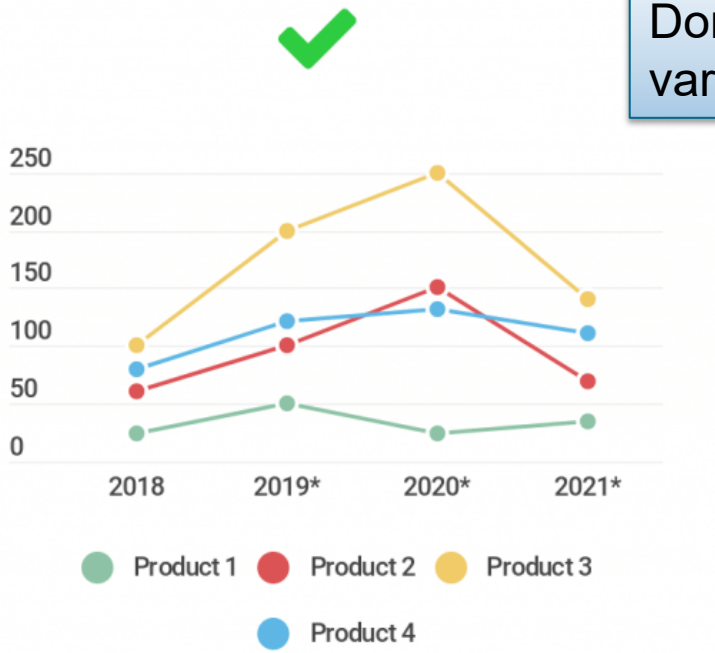
Multiple-line charts with more than 5 lines can become cluttered and very difficult to read.

Different Types of Basic Charts

3) Line Chart

Some tips:





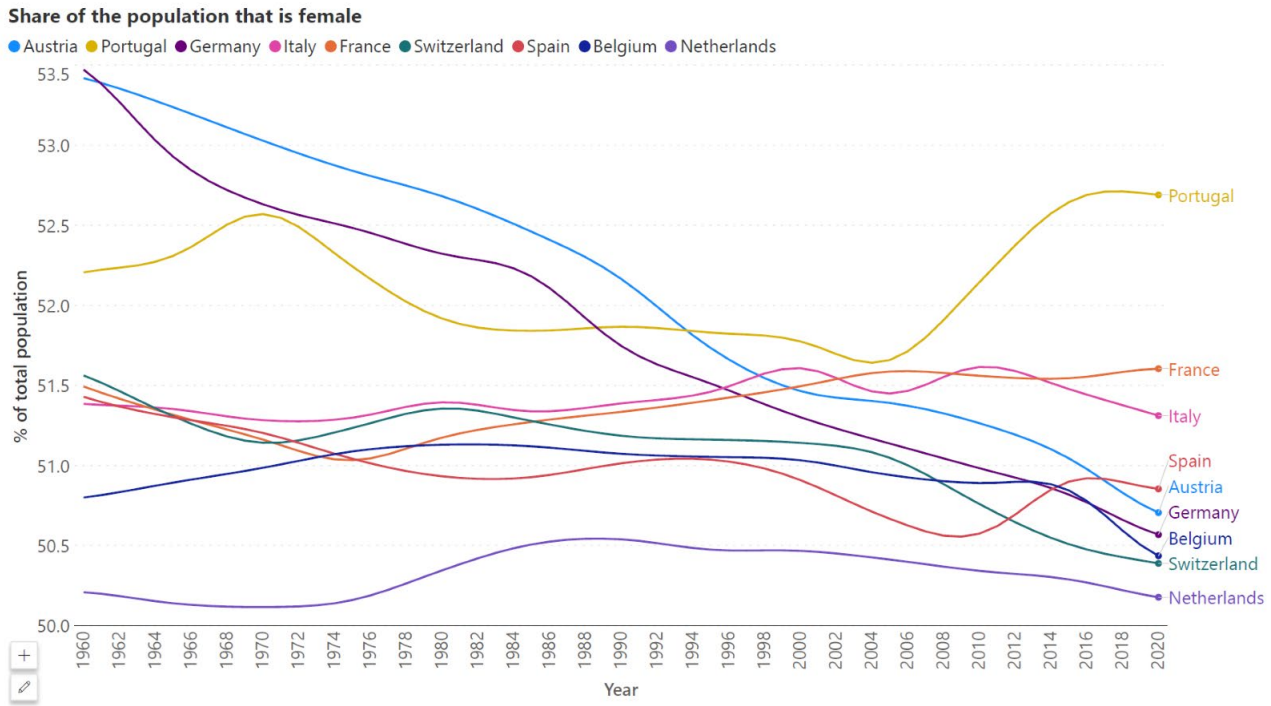
Don’t make your lines all the same color, or variations of the same shade.

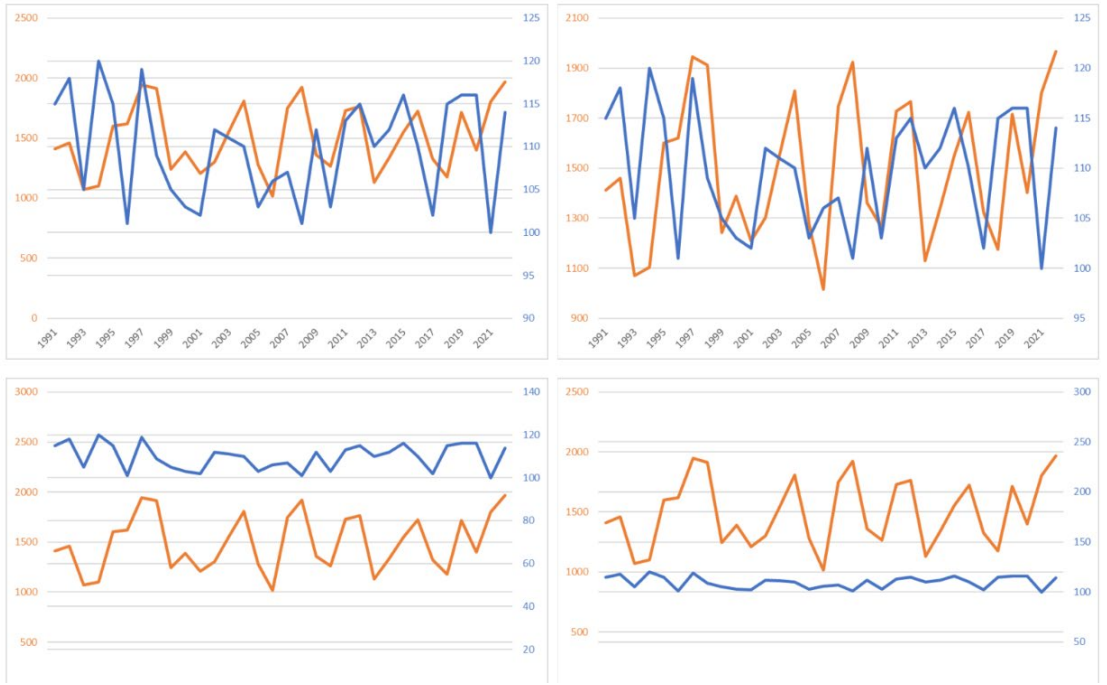
Visualization

Different Types of Basic Charts

3) Line Chart

Some tips:



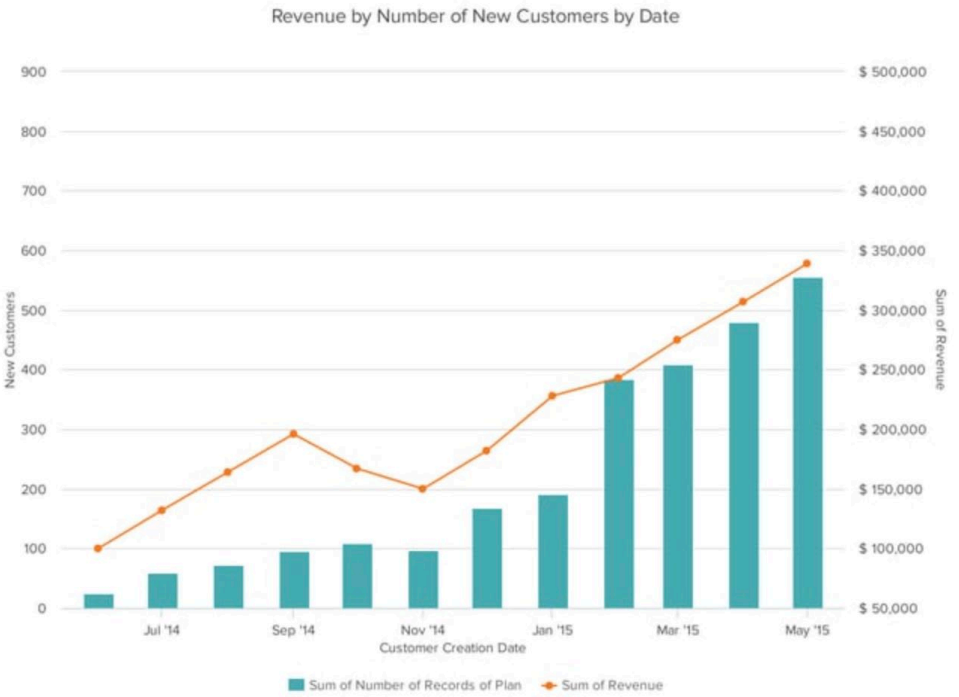


Different Types of Basic Charts

4) Dual Axis Chart

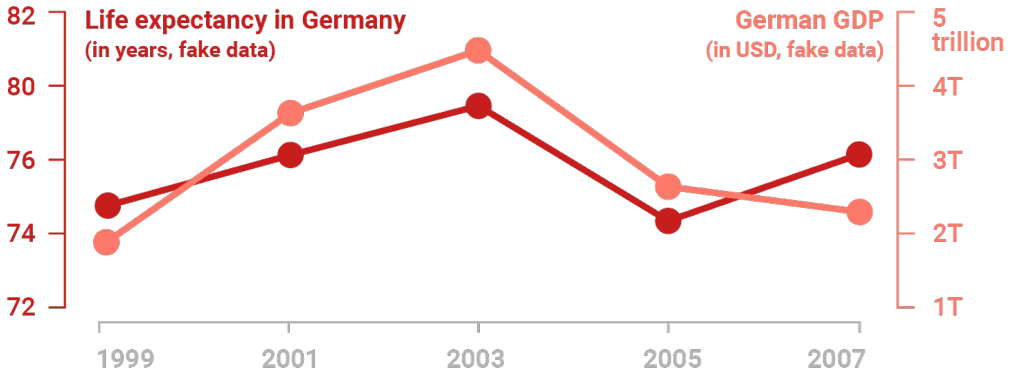
A dual axis chart allows you to plot data using two y-axes and a shared x-axis. It's used with three data sets, one of which is based on a continuous set of data and another which is better suited to being grouped by category. This should be used to visualize a correlation or the lack thereof between these three data sets.

- Use the y-axis on the left side for the primary variable because brains are naturally inclined to look left first.
- Use different graphing styles to illustrate the two data sets, as illustrated above.
- Choose contrasting colors for the two data sets.

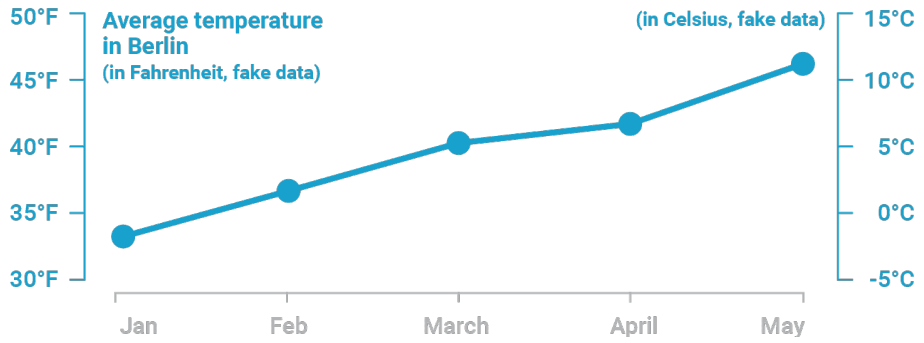


Different Types of Basic Charts

4) Dual Axis Chart



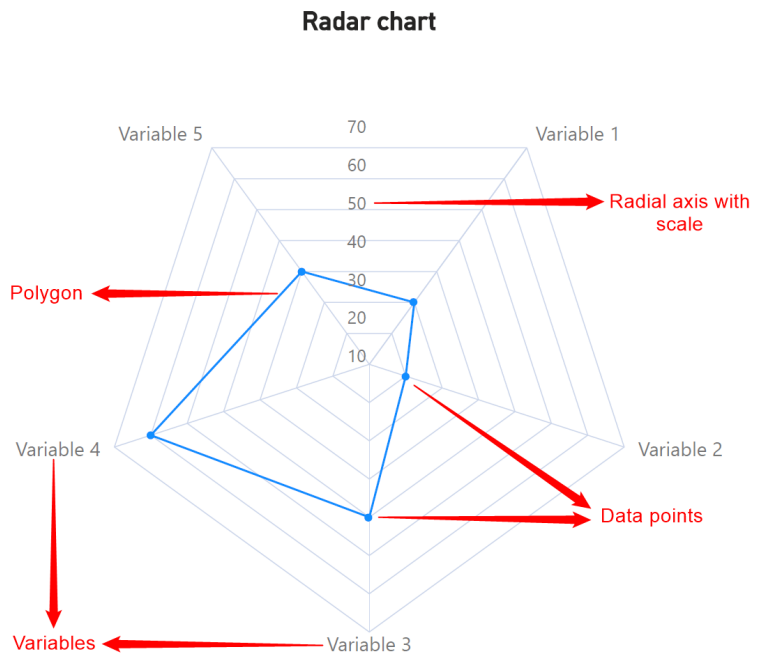




Different Types of Basic Charts

5) Radar Chart

A radar chart, sometimes called a spider chart or a star chart, is a powerful tool to examine multivariate data in a novel and engaging form.



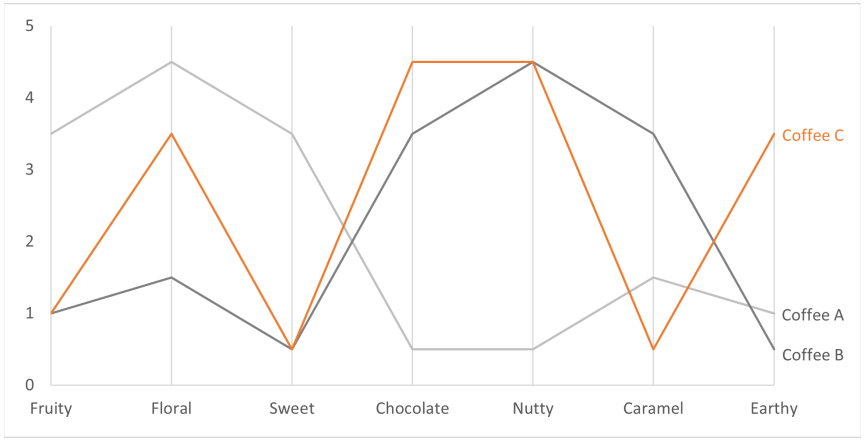


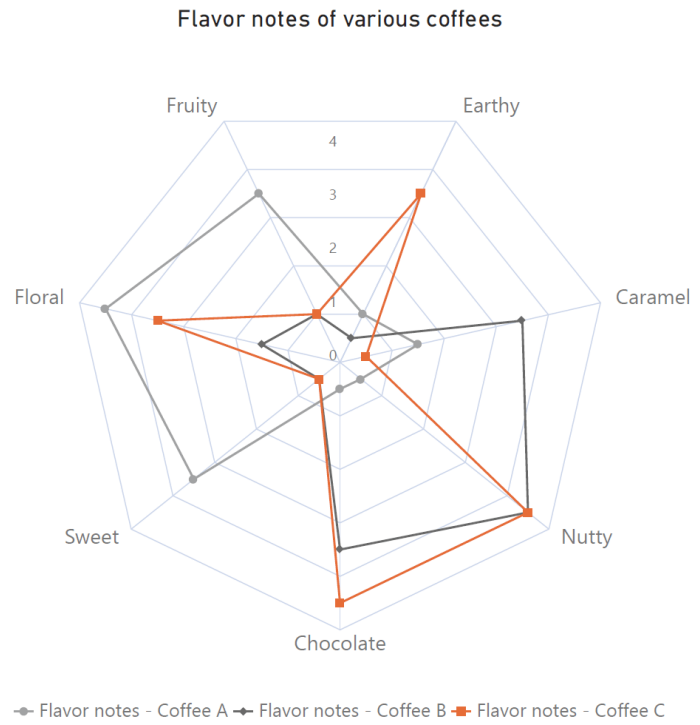
UNC GREENSBORO

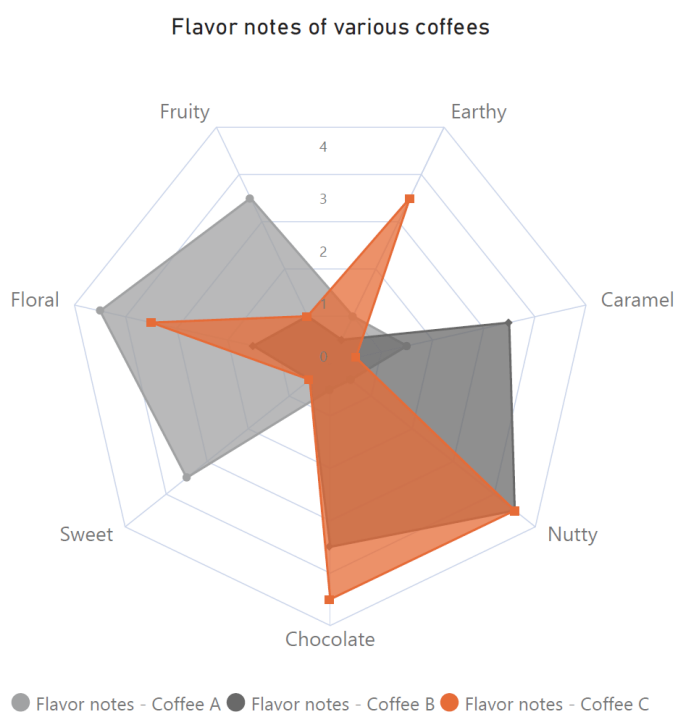
Different Types of Basic Charts

5) Radar Chart

A radar chart, sometimes called a spider chart or a star chart, is a powerful tool to examine multivariate data in a novel and engaging form.







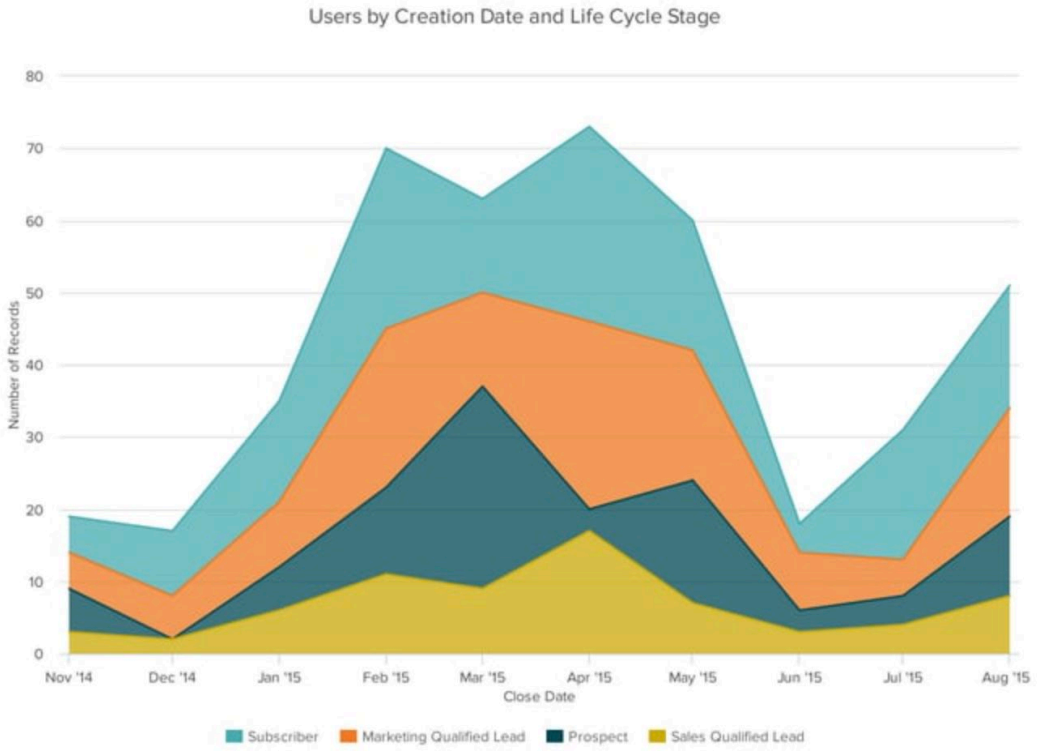
UNC GREENSBORO

Different Types of Basic Charts

6) Area Chart

Basically, a line chart, but the space between the x-axis and the line is filled with a color or pattern. It is useful for showing part-to-whole relations, such as showing individual sales contribution to total sales for a year. It helps you analyze both overall and individual trend information.

- Use transparent colors so information isn't obscured in the background.
- Don't display more than four categories to avoid clutter.
- Organize highly variable data at the top of the chart to make it easy to read

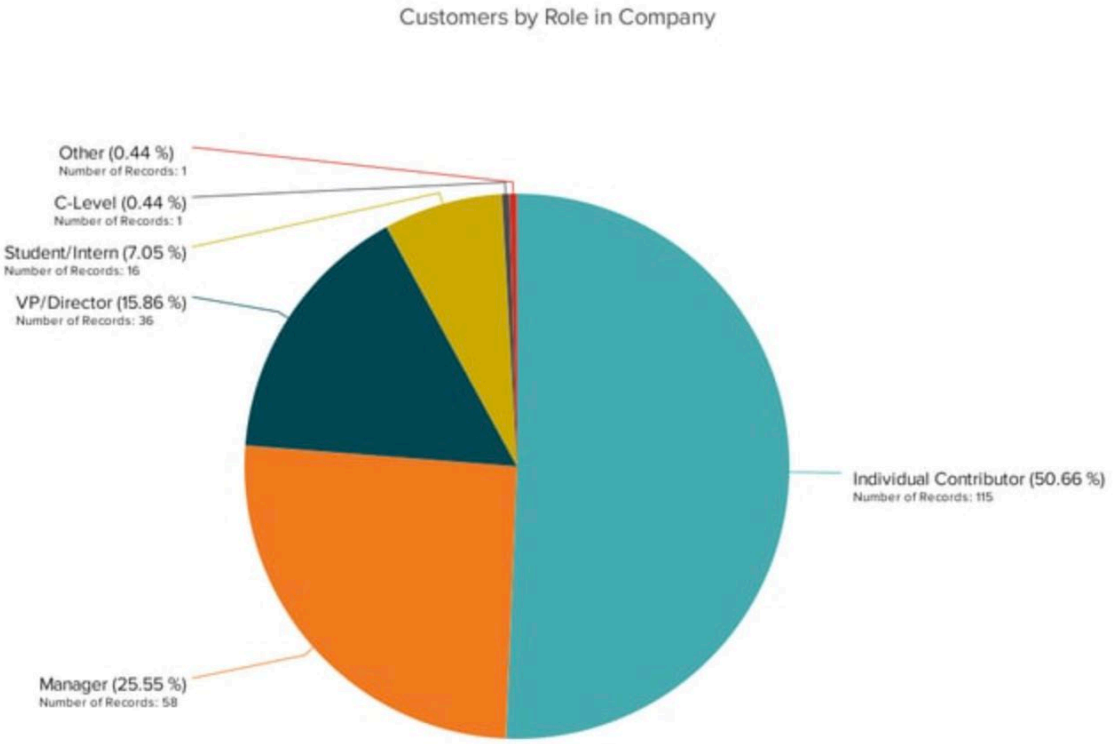


Different Types of Basic Charts

7) Pie Chart

A pie chart shows a static number and how categories represent part of a whole -- the composition of something. A pie chart represents numbers in percentages, and the total sum of all segments needs to equal 100%.

- Don't illustrate too many categories to ensure differentiation between slices.
- Ensure that the slice values add up to 100%.
- Order slices according to their size.



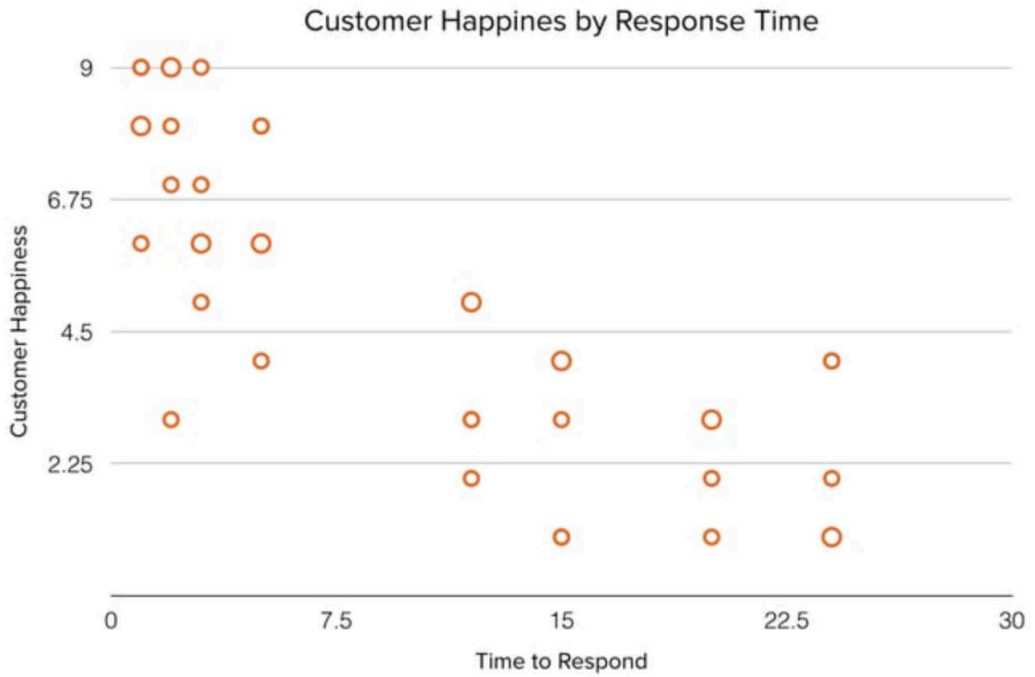
Different Types of Basic Charts

8) Scatter Plot

A scatter plot is a type of data visualization that displays individual data points on a two-dimensional graph.

Scatter plots are particularly useful for visualizing the relationship between two continuous variables and identifying patterns, trends, correlations, and outliers in the data.

- Limited to Two or Three Variables: Scatter plots are not well-suited for visualizing relationships involving more than three variables.
- Overplotting: High-density data can lead to overplotting, where points overlap excessively, obscuring patterns.

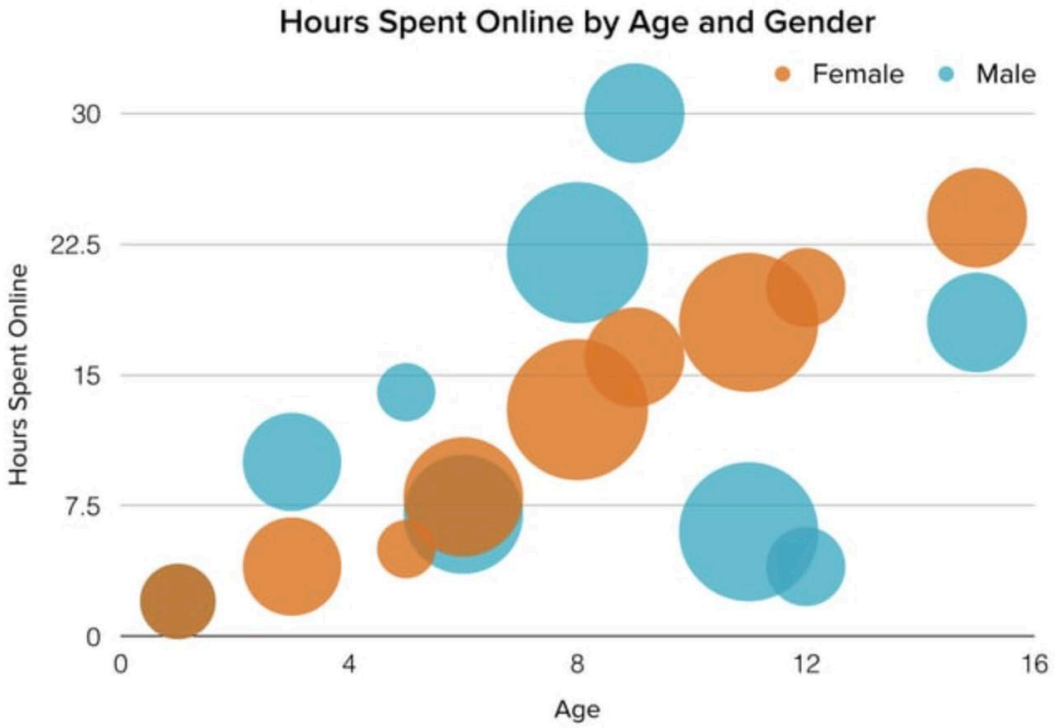


Different Types of Basic Charts

9) Bubble Chart

A bubble chart is similar to a scatter plot in that it can show distribution or relationship. There is a third data set, which is indicated by the size of the bubble or circle.

- A bubble chart is a type of graph that represents three variables using bubble sizes, colors, and positions.
- The bubble chart has an x-axis and y-axis to represent two variables, and the size of the bubbles represents the third variable. The larger the bubble, the higher the value of the third variable.



Different Types of Basic Charts

9) Bubble Chart

A bubble chart is similar to a scatter plot in that it can show distribution or relationship. There is a third data set, which is indicated by the size of the bubble or circle.

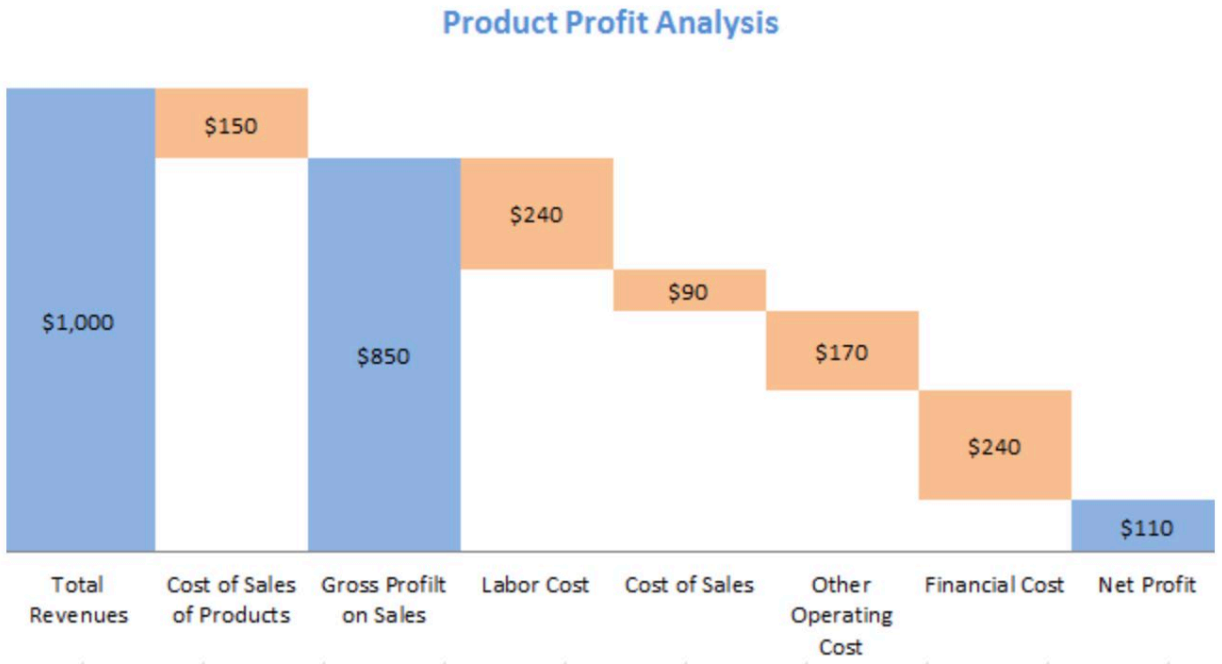


Different Types of Basic Charts

10) Waterfall Chart

also known as bridge charts, are a type of data visualization that helps in understanding the incremental changes in the value of a variable. They are commonly used in financial analysis to show how a starting value (such as net income) is affected by a series of positive and negative changes (such as revenue, expenses, taxes) to reach an ending value.

- Use contrasting colors to highlight differences
- Choose warm colors to indicate increases and cool colors to indicate decreases.

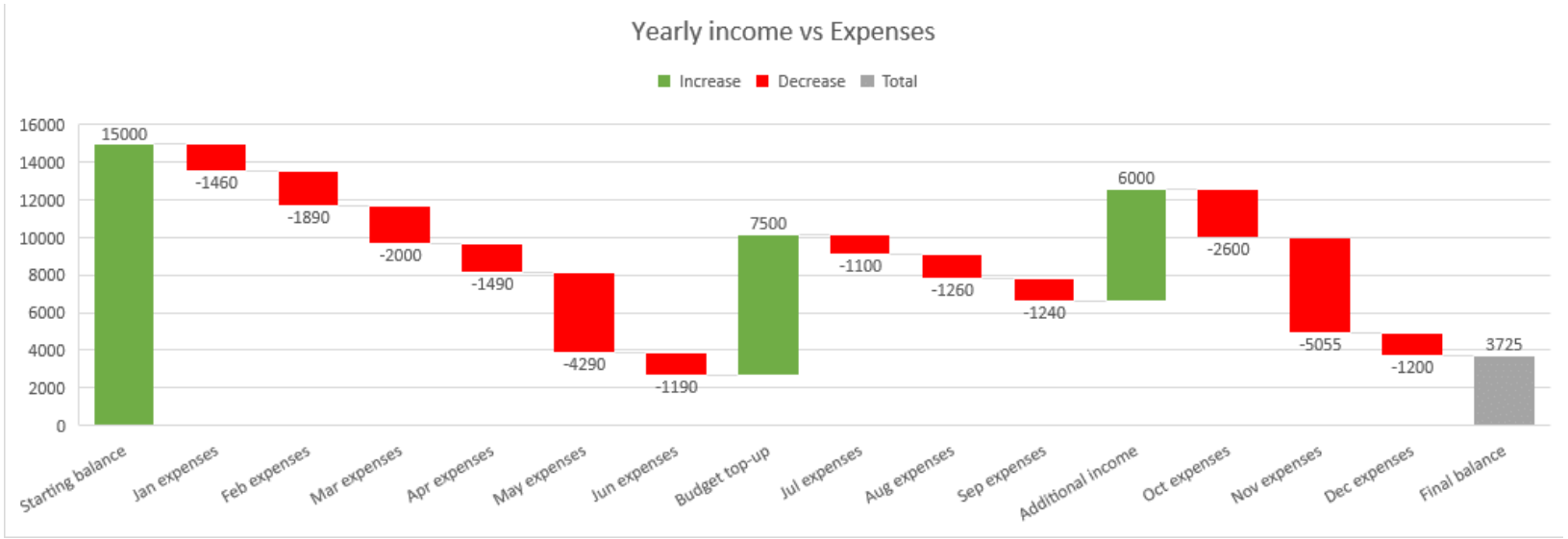


Visualization

Different Types of Basic Charts

10) Waterfall Chart

	A	B
1	**This year's budget**	
2		
3	Starting balance	15000
4	Jan expenses	-1460
5	Feb expenses	-1890
6	Mar expenses	-2000
7	Apr expenses	-1490
8	May expenses	-4290
9	Jun expenses	-1190
10	Budget top-up	7500
11	Jul expenses	-1100
12	Aug expenses	-1260
13	Sep expenses	-1240
14	Additional income	6000
15	Oct expenses	-2600
16	Nov expenses	-5055
17	Dec expenses	-1200
18	Final balance	3725



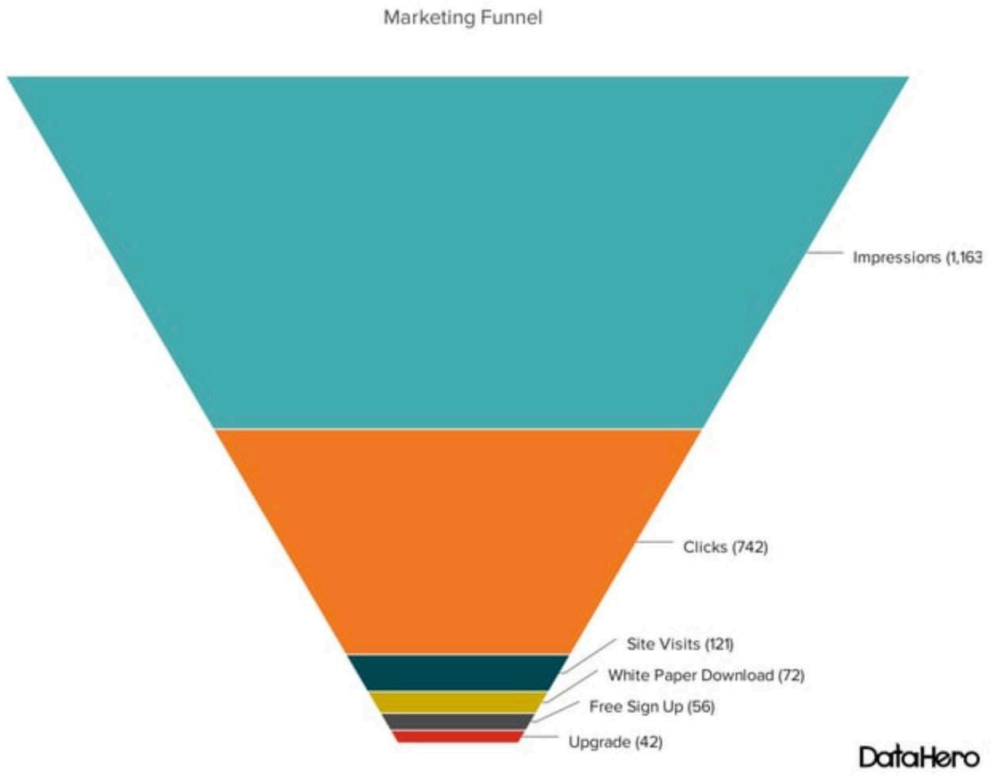
UNC GREENSBORO

Different Types of Basic Charts

11) Funnel Chart

A funnel chart shows a series of steps and the completion rate for each step. This can be used to track the sales process or the conversion rate across a series of pages or steps.

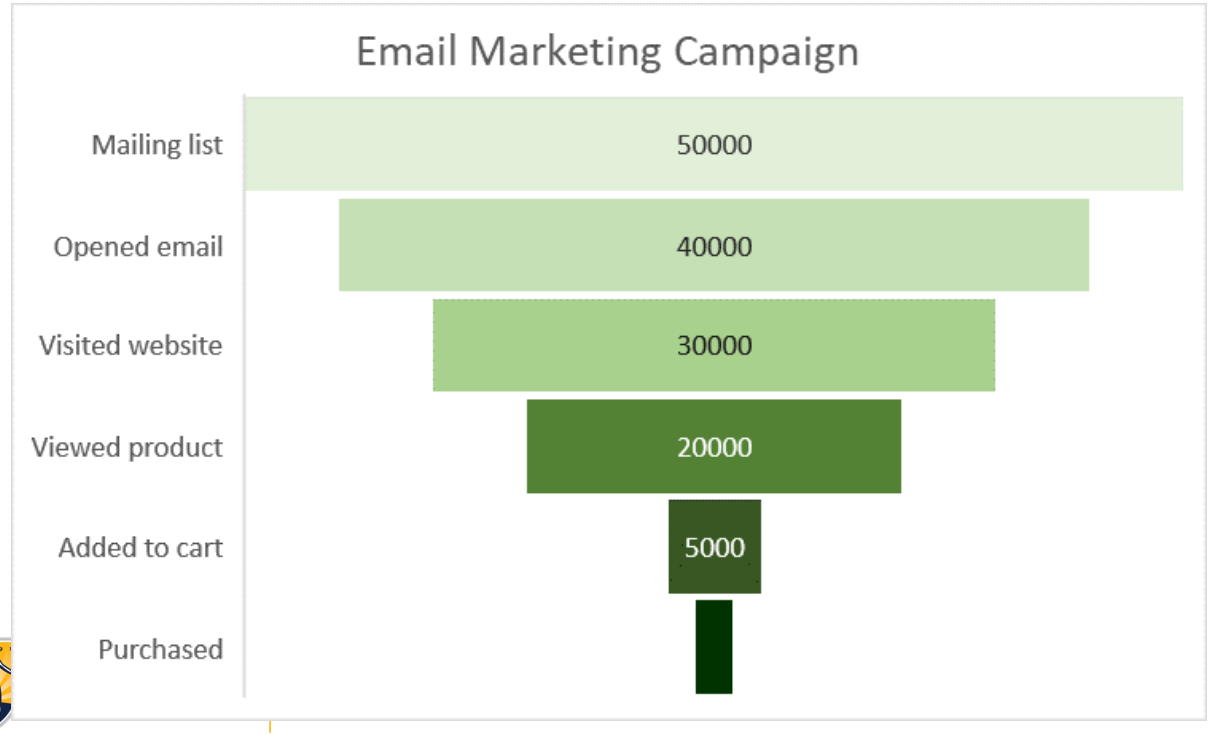
- Area-width ambiguity: In a funnel chart, people visually judge the area, but the data is usually encoded using width only.
- Use contrasting colors or one color in gradating hues, from darkest to lightest as the size of the funnel decreases.

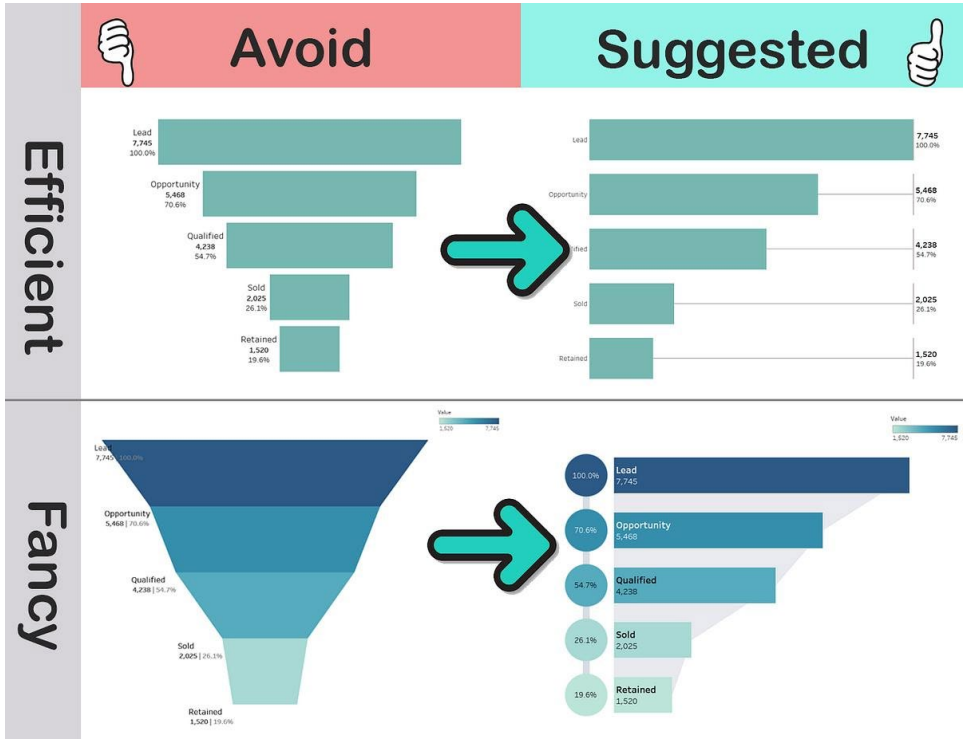


Different Types of Basic Charts

11) Funnel Chart

A funnel chart shows a series of steps and the completion rate for each step. This can be used to track the sales process or the conversion rate across a series of pages or steps.



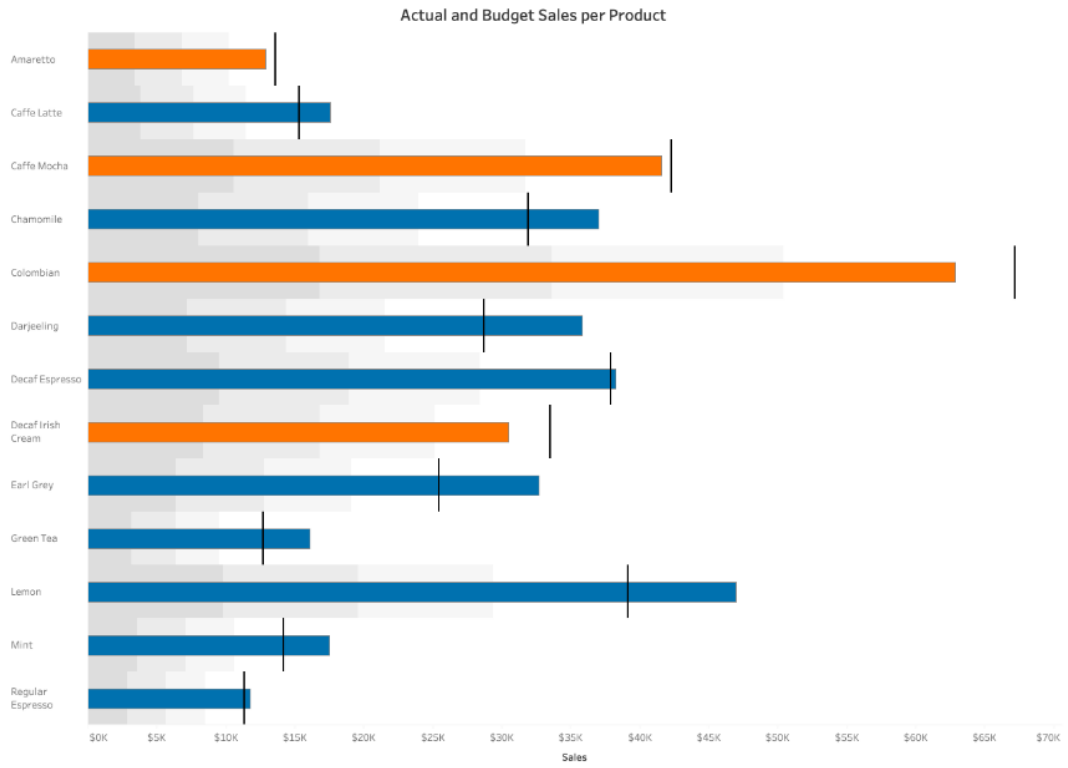


Different Types of Basic Charts

12) Bullet Chart

A bullet graph reveals progress toward a goal, compares this to another measure, and provides context in the form of a rating or performance.

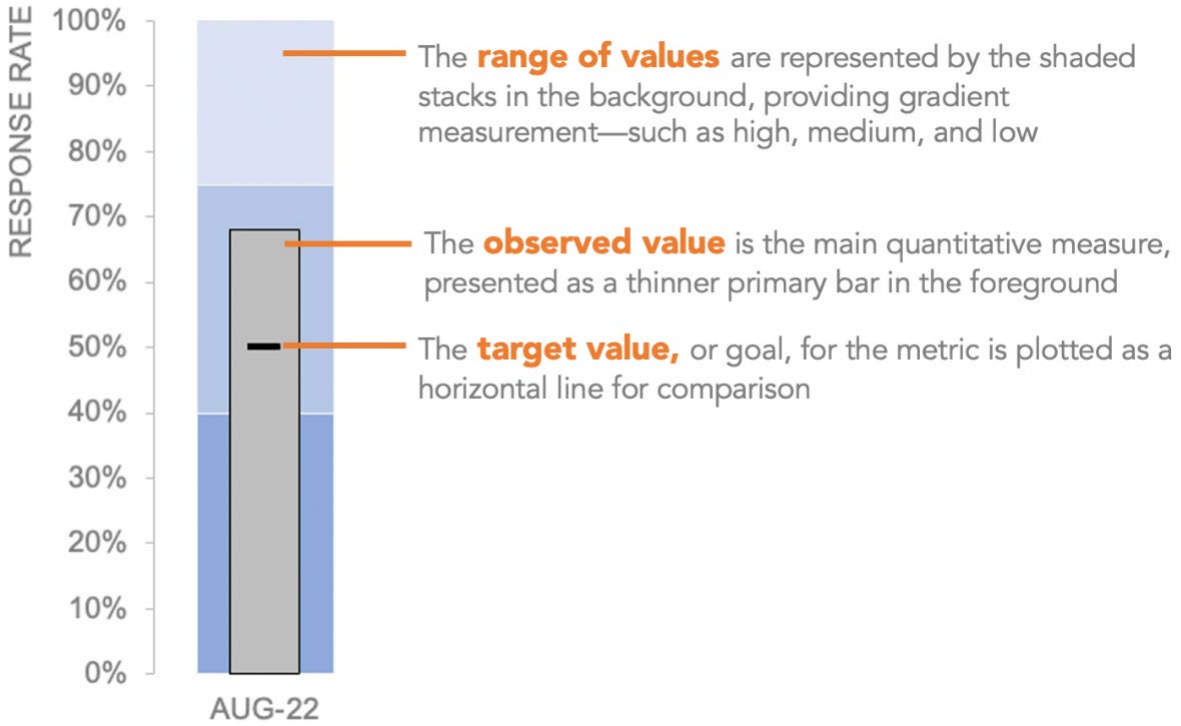
- Uses a vertical bar to show the target for each Product
- Uses colors (yellow and green) to distinguish which product met its sales goals
- Uses gray to mark the goal bar around the product bar
- Starts at a base of zero



Different Types of Basic Charts

12) Bullet Chart

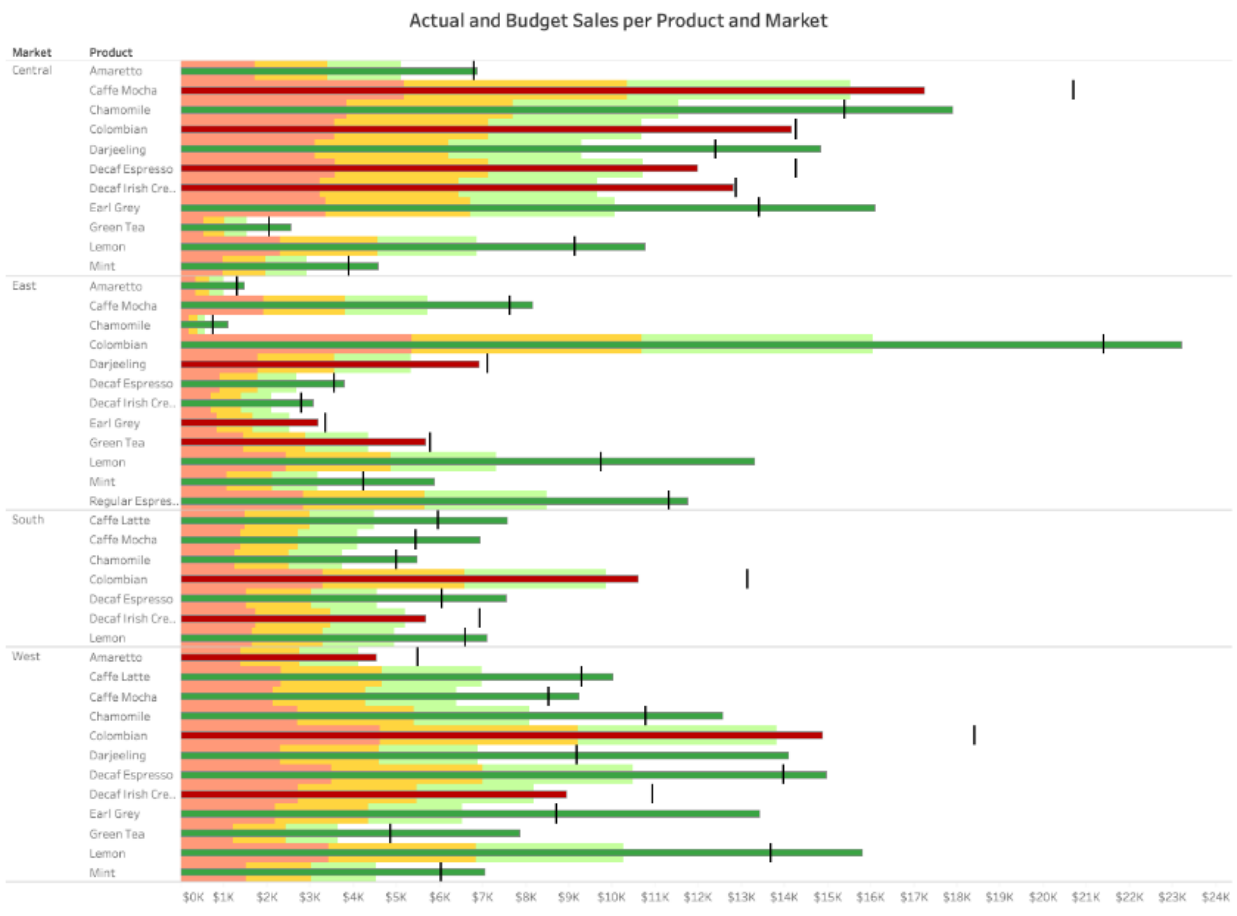
A bullet graph reveals progress toward a goal, compares this to another measure, and provides context in the form of a rating or performance.



Different Types of Basic Charts

12) Bullet Chart

Tips:



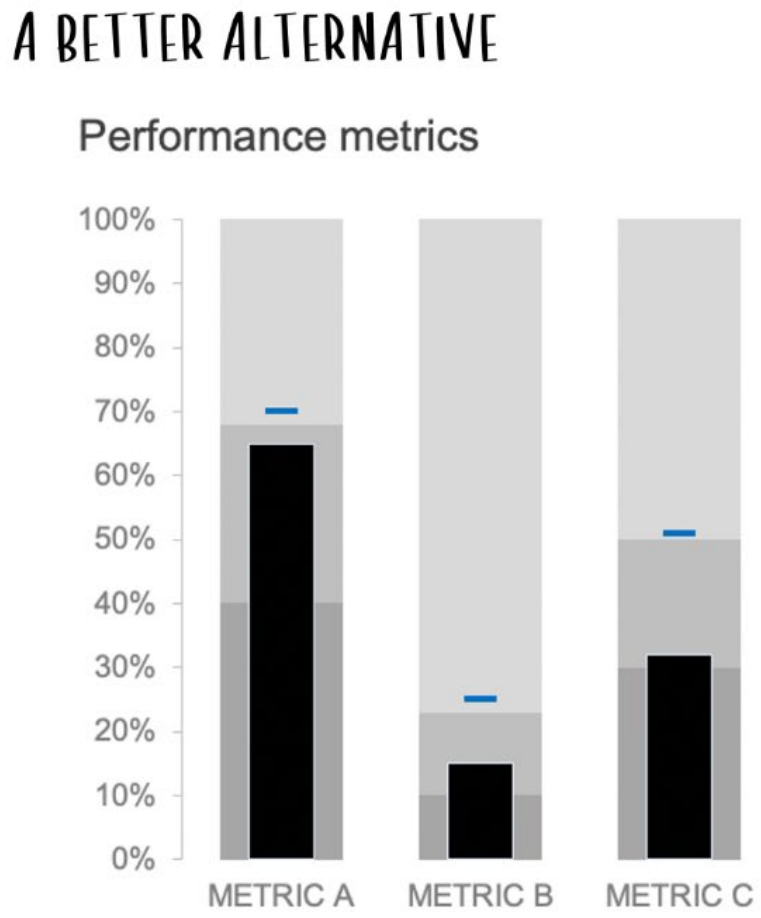
Different Types of Basic Charts

12) Bullet Chart

Tips:

Using too much color can overwhelm and confuse your audience.



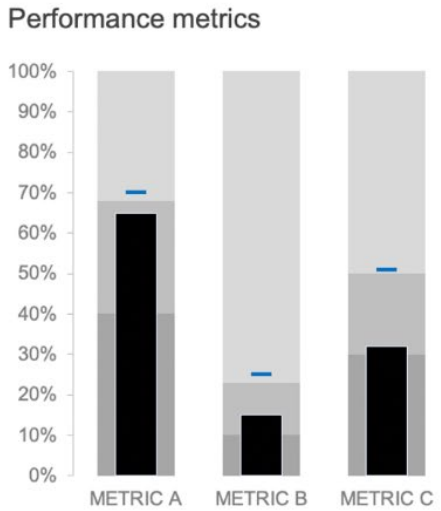


Different Types of Basic Charts

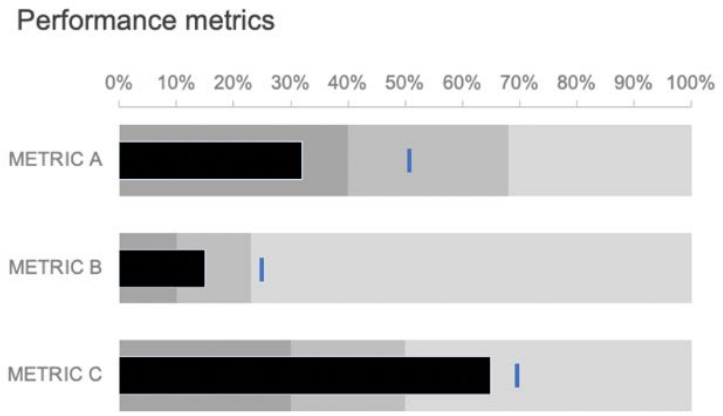
12) Bullet Chart

Tips:

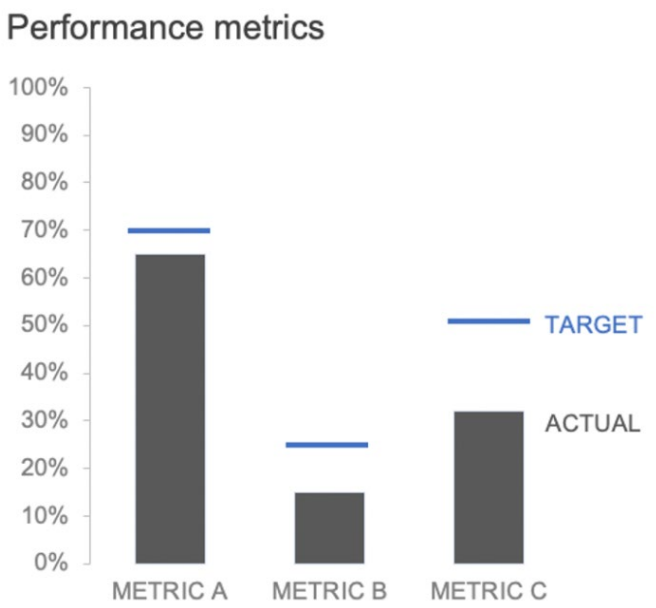
VERTICAL BULLET GRAPHS



HORIZONTAL BULLET GRAPHS



BULLET GRAPH WITHOUT GRADING



Different Types of Basic Charts

13) Heat Map

A heat map is a two-dimensional representation of data in which various values are represented by colors.

Ice cream sales - May 2023

Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

Visualization

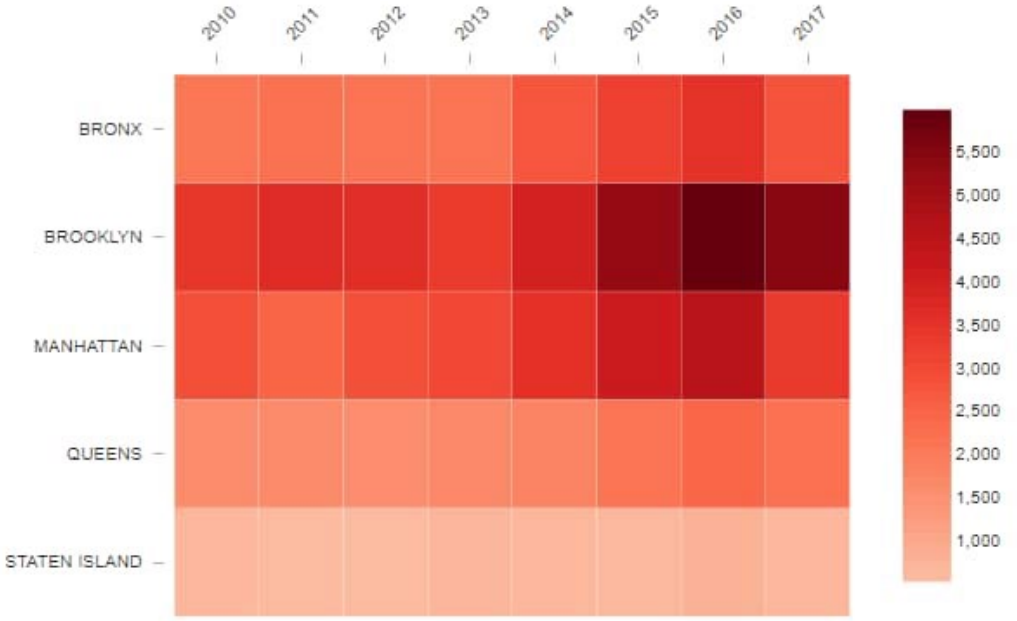
Different Types of Basic Charts

13) Heat Map

Enter or Paste Data

Enter data manually or use Ctrl-V to paste data from a spreadsheet.

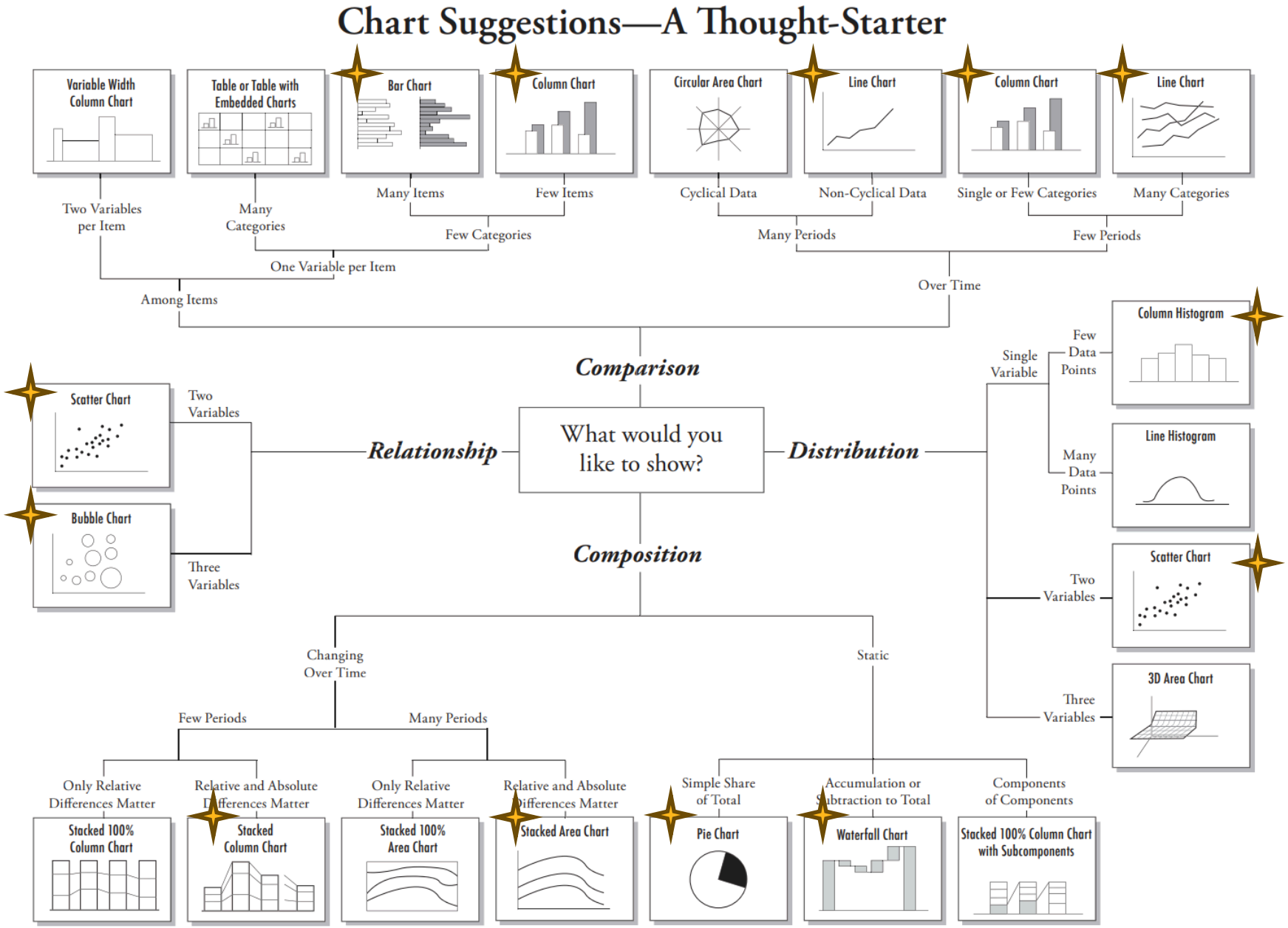
	A	B	C	D	E	F	G	H	I
1	Count	2010	2011	2012	2013	2014	2015	2016	2017
2	BRONX	2067	2174	2125	2120	2743	3189	3497	2791
3	BROOKLYN	3399	3671	3600	3300	3968	5263	5979	5493
4	MANHATTAN	2867	2461	2857	3018	3541	4123	4610	3326
5	QUEENS	1569	1600	1522	1660	1748	2103	2410	2199
6	STATEN ISLAND	632	548	538	641	617	594	734	616



This table contains the number of rats sighted in each borough of New York, from 2010 to 2017. It's not too difficult to find the biggest concentration of rats was in Brooklyn in 2016, but the heatmap makes it easy.

UNC GREENSBORO

Visualization



UNC GREENSBORO