

Data Visualization

Design principles

Sheetal Pereira

Assistant Professor

Department of Computer Engineering

K. J. Somaiya College of Engineering

Somaiya Vidyavihar University

Design principles Categorical, time series, and statistical data graphics

- Unity/Harmony
- Balance
- Hierarchy
- Scale/Proportion
- Dominance/Emphasis
- Similarity and contrast

Design principles Categorical, time series, and statistical data graphics

- Unity/Harmony

Unity / Harmony



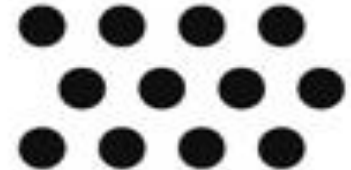
PROXIMITY

A sense of the distance between elements



SIMILARITY

Ability to seem repeatable with other elements



CONTINUATION

The sense of having a line or pattern extend



REPETITION

Elements being copied or mimicked numerous times



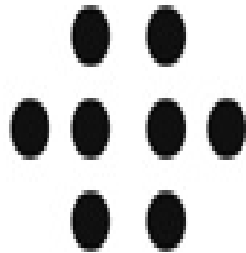
RHYTHM

Achieved when recurring position, size, color, and use of a graphic element has a focal point interruption

Design principles Categorical, time series, and statistical data graphics

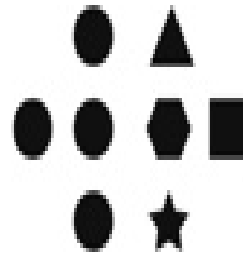
- Balance:

Balance



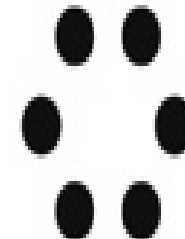
SYMMETRY

Elements on either side of the axis are arranged similarly



ASYMMETRY

Elements on each side differ in shape but still are in visual equilibrium

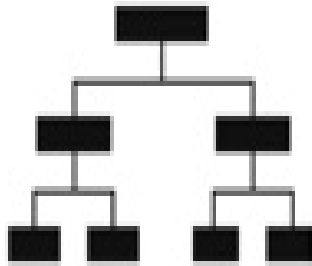


RADIAL

Elements are arranged around a circular form

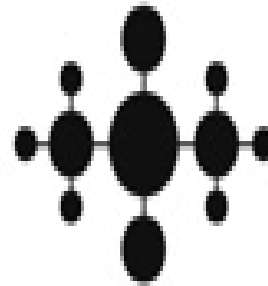
Design principles Categorical, time series, and statistical data graphics

Hierarchy



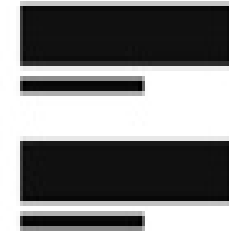
TREES

Elements arranged in the order of a tree with a trunk, branches & sub branches



NESTS

Elements mapped on to each other as parents, children & grand children



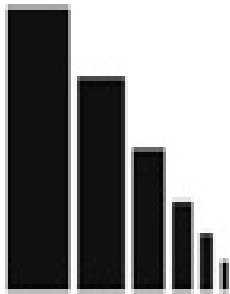
WEIGHT

Elements of the same weight belong to the same class of hierarchical positions

Design principles Categorical, time series, and statistical data graphics

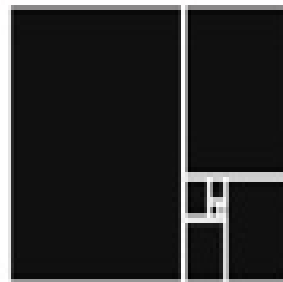
Scale/Proportion:

Scale / Proportion



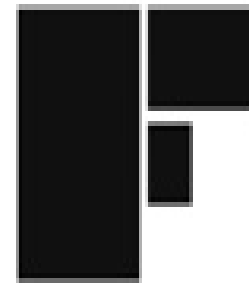
SIZE

Elements of different sizes in relationships with each other



RATIO

Elements related to each other in a ratio appear together in visual harmony



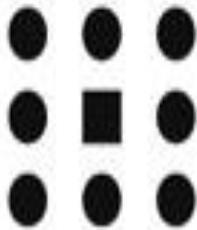
DIVISIONS

These create focal points that automatically give a sense of the relationships

Design principles Categorical, time series, and statistical data graphics

- I

Dominance / Emphasis



HIGHLIGHT

Breaking the visual hierarchy using form to lay emphasis



COLOUR

To distinguish between elements in a series of similar forms



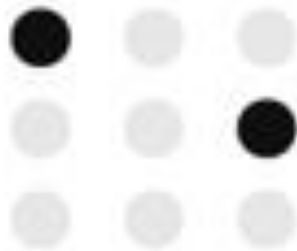
SIZE

Elements of different sizes focus the viewers attention accordingly

Design principles Categorical, time series, and statistical data graphics

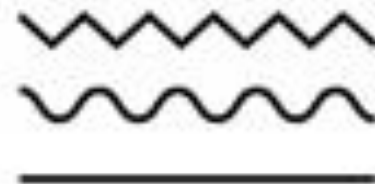
- Similarity & Contrast:

Similarity & Contrast



LIGHT & DARK

Clear foreground & background separation lend contrast between elements



LINE

Elements of varying textures & forms bring about a contrasting effect

Design principles Categorical, time series, and statistical data graphics

Follow these five principles to create compelling and competent visualizations:



Design principles Categorical, time series, and statistical data graphics

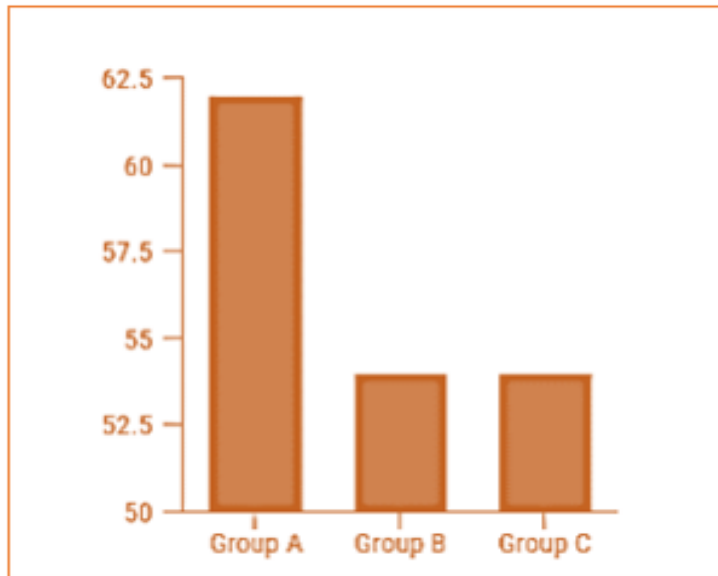
- **1. Tell the truth**
- There are a plethora of graphs that misguide the reader by showcasing skewed data and projecting false narratives.
- *The job of the presenter is to inform the audience, not influence them.*
- Avoid these three misleading methods to ensure that your graphs are clear and honest.

Design principles Categorical, time series, and statistical data graphics

- **1. Tell the truth**
- **A. Omitting the baseline**
- Generally, the baseline for a graph should start from zero unless specified otherwise. By starting the baseline from a different number, it can bias the perception of data. This technique is used to make the difference between data points seem to be greater than actual.

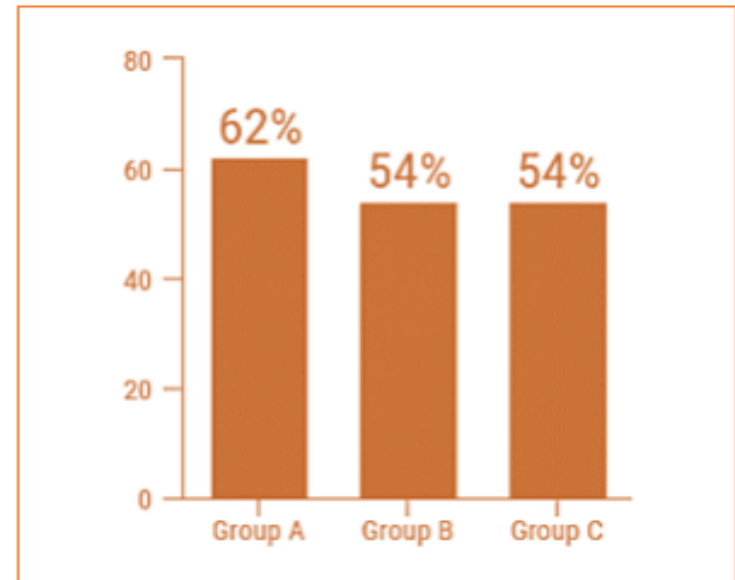
Design principles Categorical, time series, and statistical data graphics

- 1. Tell the truth
- A. Omitting the baseline



Starting the baseline from 50 makes a small difference between the groups seem massive

Group A seems to be much larger than the other two groups



Starting the vertical axis from 0 offers a more accurate depiction of data

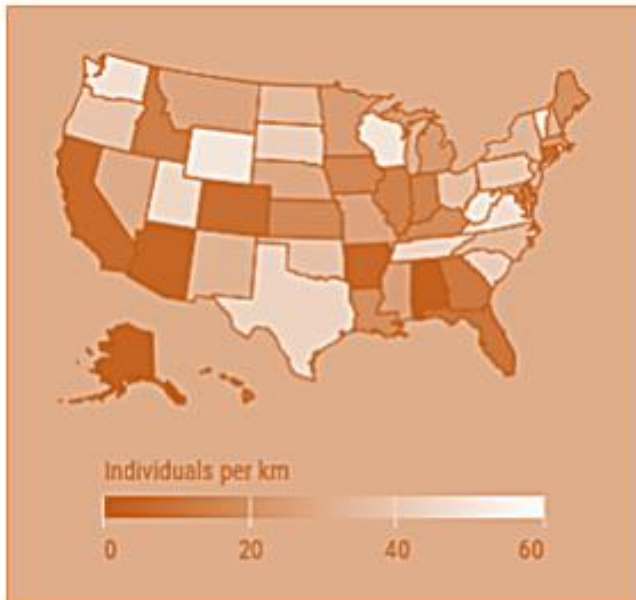
The difference between the groups is not as dramatic when visualized truthfully

Design principles Categorical, time series, and statistical data graphics

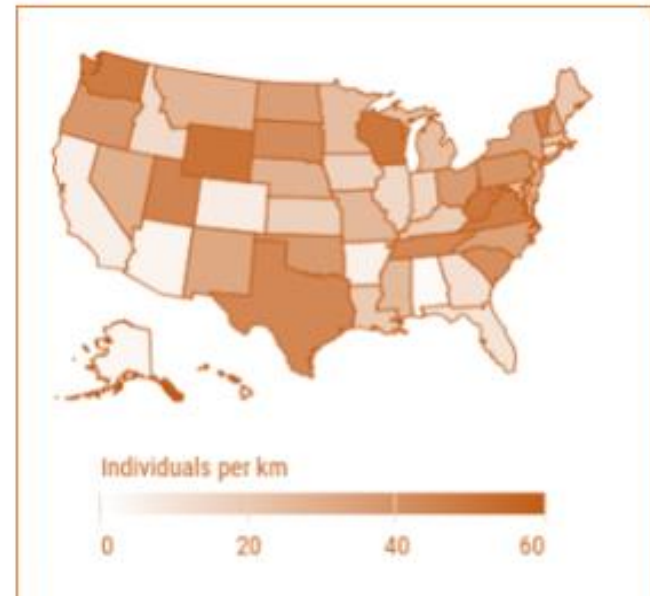
- **1. Tell the truth**
- **B. Going against the conventions**
- There are certain conventions when plotting data. A larger bar indicates a greater amount, and a bigger area represents a higher number of values. By going against the convention, the viewer is unwittingly led to a wrong inference.

Design principles Categorical, time series,

- 1. Tell the truth and statistical data graphics
- B. Going against the conventions



Darker shades indicate lower density in this chart which is against the convention and can confuse the reader



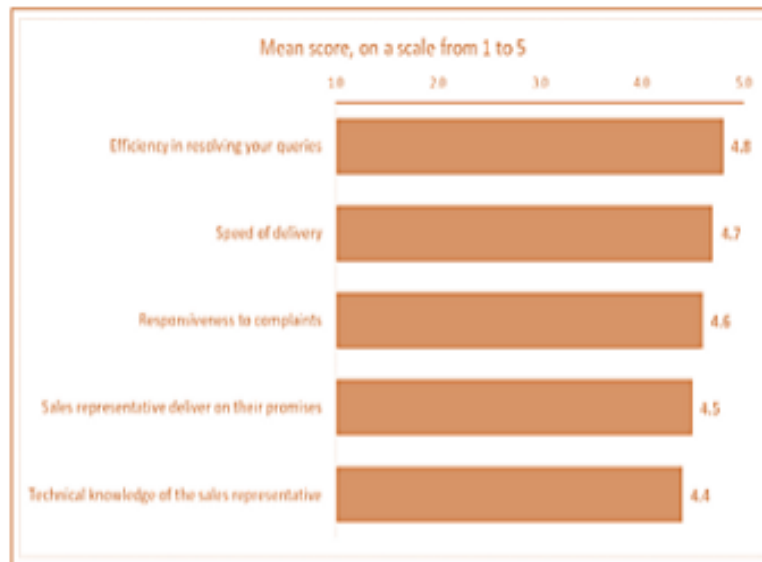
Conventional representation of darker shades representing higher density leads to correct conclusion

Design principles Categorical, time series, and statistical data graphics

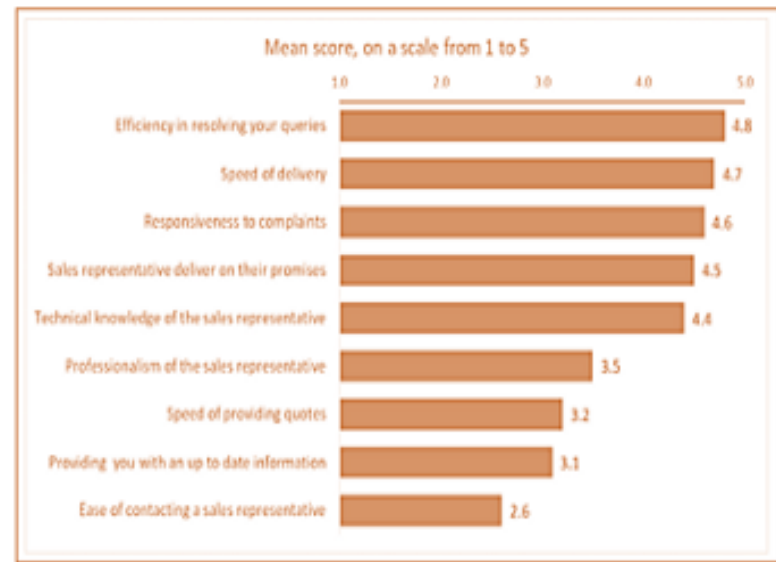
- **1. Tell the truth**
- **C. Cherry-picking data**
- Cherry-picking is when only a few data points are plotted to show a misleading trend. This is one of the most common tactics used to mislead or deceive the audience.

Design principles Categorical, time series,

- 1. Tell the truth and statistical data graphics
- C. Cherry-picking data



Only the top 5 parameters are plotted to show the highest scoring data points and make the overall results look good.



The correct graph shows all the data points which includes some low scoring parameters and gives the actual picture.

Design principles Categorical, time series, and statistical data graphics

- **2. Know your audience**
- More often than not, you will be creating data visualizations to communicate information to an audience. The whole point of creating the viz is lost if the message is not conveyed.
- Show what you have in a way they want to see.
- A good data visualization should resonate with the audience and to make sure that it does, follow these guidelines:-
- Display data according to the job role.
- Take into consideration the education and expertise of the audience in relation to the topic of the dataviz.
- Be sensitive to ethnicity and cultural values.
- Focus on the literacy of the audience in terms of technical terms, statistics, language.

Design principles Categorical, time series, and statistical data graphics

- **2. Know your audience**
- More often than not, you will be creating data visualizations to communicate information to an audience. The whole point of creating the viz is lost if the message is not conveyed.
- Show what you have in a way they want to see.
- A good data visualization should resonate with the audience and to make sure that it does, follow these guidelines:-
- Display data according to the job role.
- Take into consideration the education and expertise of the audience in relation to the topic of the dataviz.
- Be sensitive to ethnicity and cultural values.
- Focus on the literacy of the audience in terms of technical terms, statistics, language.

Design principles Categorical, time series, and statistical data graphics

- four scenarios where you should choose your data viz according to the audience you are presenting for.
- **A. Based on the job roles**
- You should present high-level data without too many details to the top-level management, while the various departments should be shown an overall picture with more details of their specific department.

Design principles Categorical, time series, and statistical data graphics



The chart for top management can be as represented showing the profits of the company over time



A dashboard with parameters like cost breakup, profits of various segments, employee performance, comparison of different products etc.

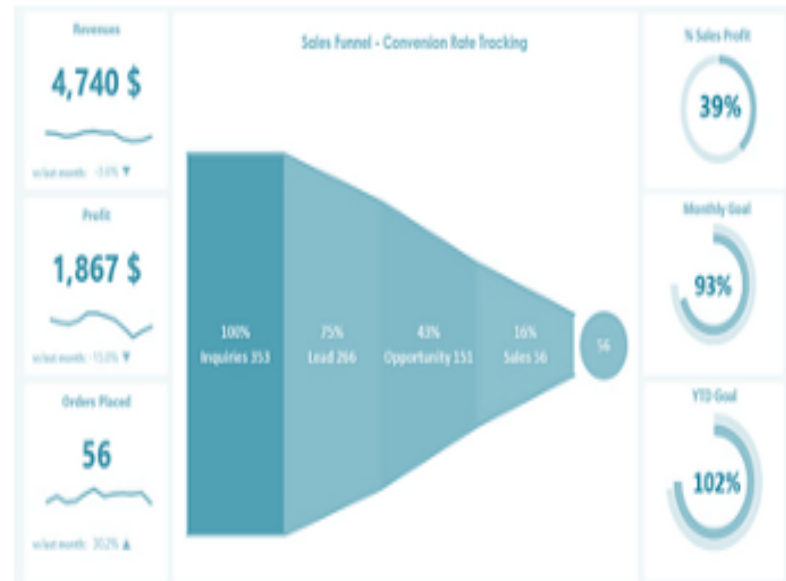
Design principles Categorical, time series, and statistical data graphics

- **B. Based on technical literacy**
- Use charts and figures based on the audience's knowledge of statistics and visualization. You can show more or less complicated information and graphs according to the end-user.

Design principles Categorical, time series,



For an informed audience, a detailed dashboard with various parameters and charts is very useful



Show basic figures and a simpler chart for audience with less technical exposure

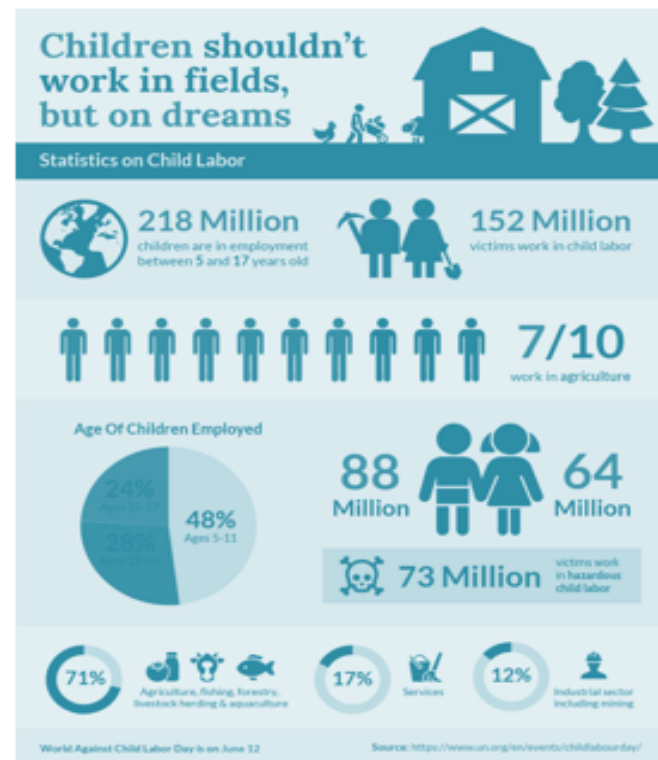
Design principles Categorical, time series, and statistical data graphics

- **C. Based on background knowledge**
- You must focus on the facts and numbers in the case of an audience with no prior perception of an issue. Show more details and peripheral information when the audience needs more knowledge about an issue that they are already aware of.

Design principles Categorical, time series, and statistical data graphics



Big numbers and facts presented to an audience not familiar with the child labor problem



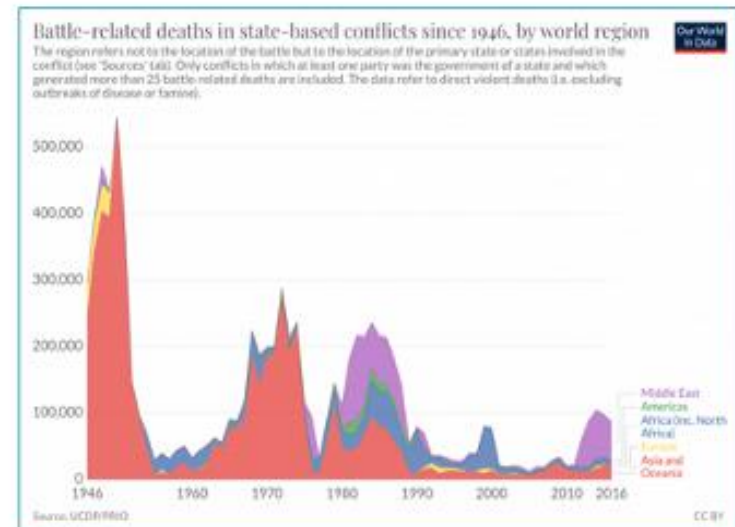
More informative visualization to enlighten audience about the issue of child labor

Design principles Categorical, time series, and statistical data graphics

- **D. Based on emotional disposition**
- It is a good idea to use emotional appeal with a temperamental audience while focussing more on factual representation with a pragmatic audience.



Size of flower indicates the causality in the war and the reddish colors symbolize the bloodshed.



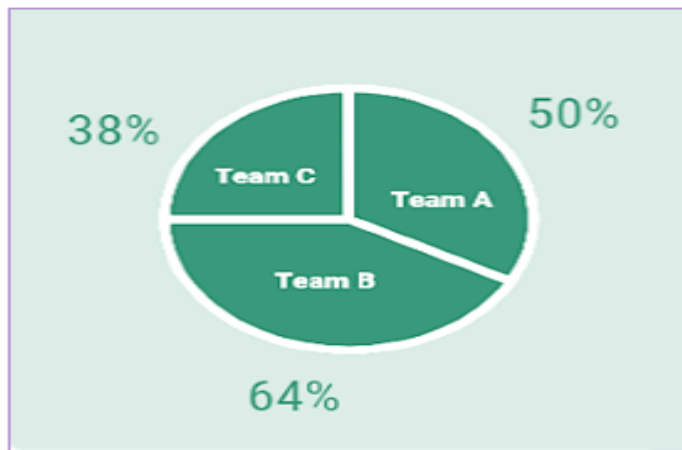
A more practical representation of deaths in war and distinct colors to show the different regions of world.

Design principles Categorical, time series, and statistical data graphics

- **3. Choose the right chart**
- Using the wrong chart is like having good intentions but poor execution.
- Choose the graph based on the kind of data and the message to be conveyed.
- Do not use different graphs just for variety, as specific graphs convey certain types of information more effectively than others.
- If not required, do not use any chart — show only numbers.
- The right chart enhances information, the wrong one conceals insights.

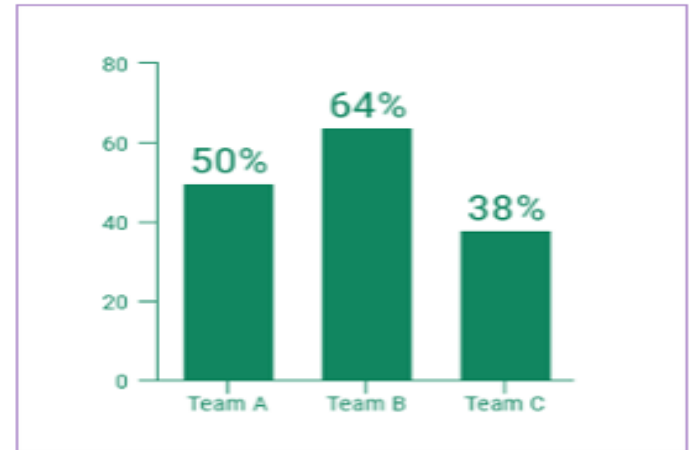
Design principles Categorical, time series, and statistical data graphics

- **3. Choose the right chart**
- **A. Bar graph instead of a pie chart**
- When showing comparisons, a bar chart is generally better than a pie chart. It is easier to make out the difference in length of bars than the size of pie segments.



Pie charts are useful to compare parts of a whole.

It is difficult to gauge proportion difference in pies.

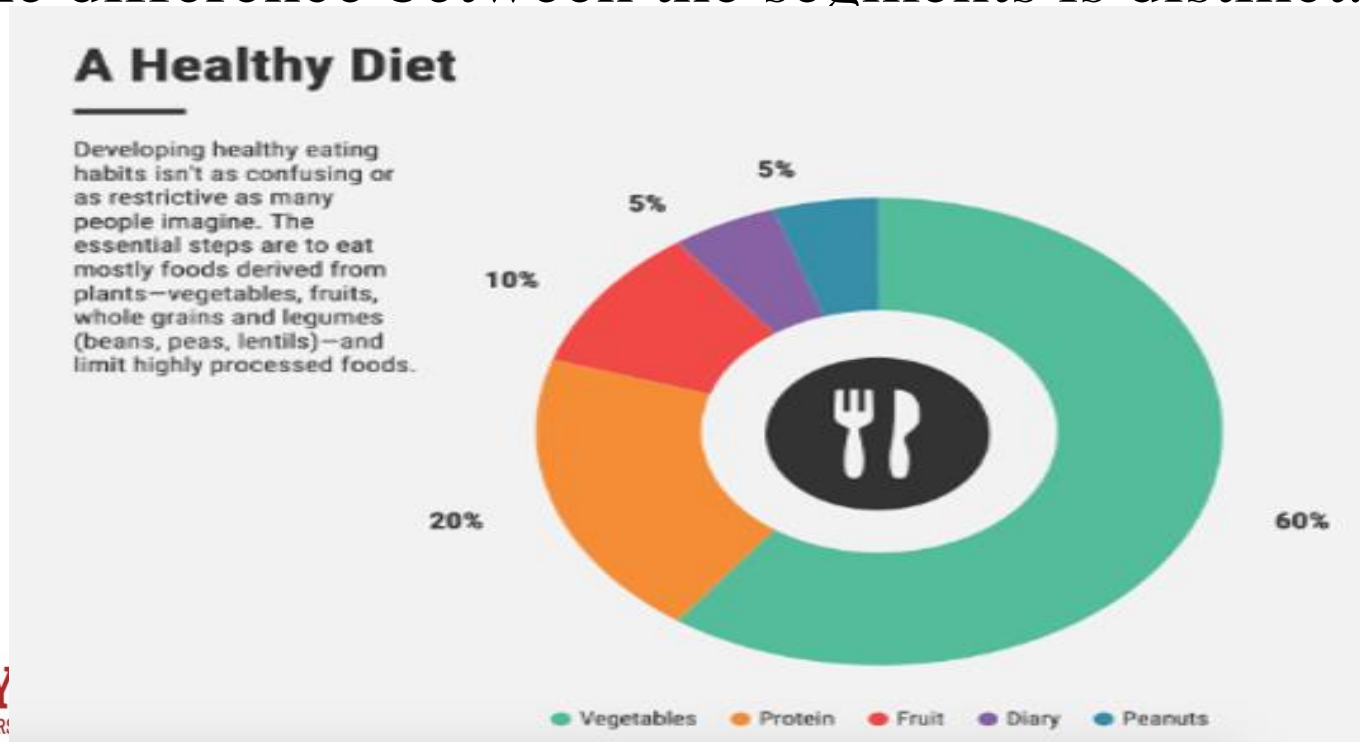


Bar graphs are great for showing difference between groups.

Bars show a distinct difference between the data points.

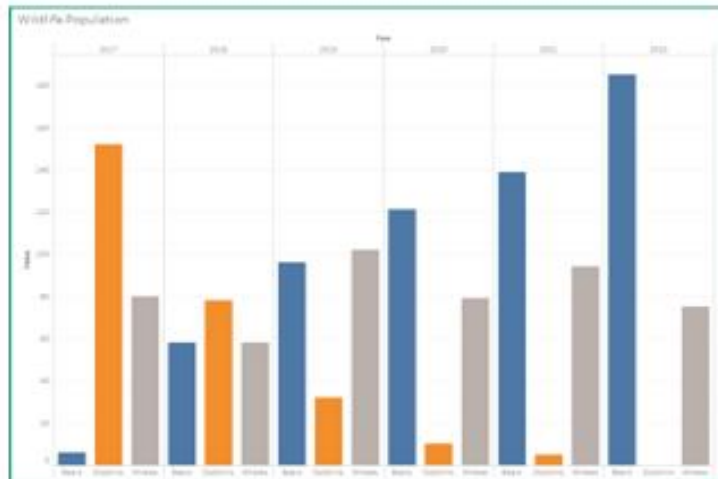
Design principles Categorical, time series, and statistical data graphics

- **3. Choose the right chart**
- **B. When pie charts are actually useful**
- Pie charts are great for showing parts of a whole, but only when there are a small number of segments and the difference between the segments is distinct.

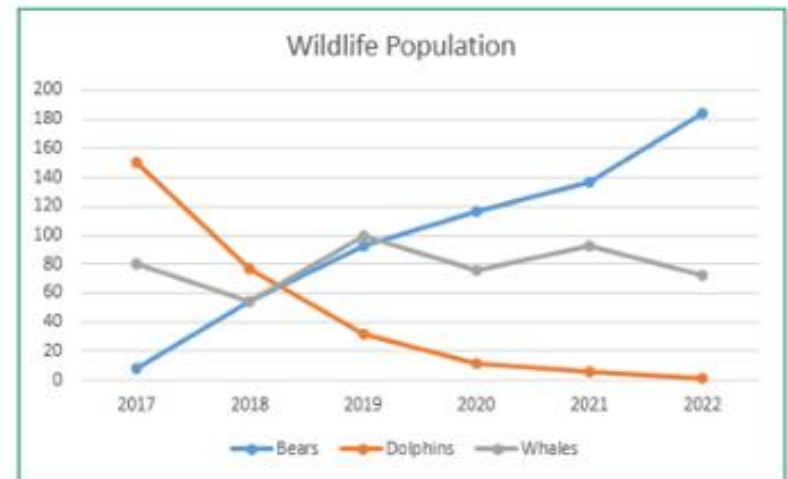


Design principles Categorical, time series, and statistical data graphics

- **3. Choose the right chart**
- **C. Line graph instead of a bar graph**
- Line graphs are typically better suited than bar charts to show comparison over time. Also, line charts are very useful to indicate trends or patterns.



It is difficult to compare the various data points over time by briefly looking at the bar graph



It can be easily observed that there is an upward trend in population of bears while dolphins are becoming endangered

Design principles Categorical, time series,

- **3. Choose the right chart** and statistical data graphics
- **D. Numbers only instead of any graph**
- When there is a specific figure to be highlighted, it can be effective to use only numbers with no charts. One or two numbers are easy to grasp and can make a large impact.



Design principles Categorical, time series, and statistical data graphics

- **4. Emphasize the most important facts**
- A data visualization can encode many data points, so highlight the most important facts to convey the message faster and with more impact.
- Omit the insignificant to highlight the essential!
- You can help direct attention on the most important facts by removing noise, such as unnecessary gridlines, axes and labels. Use color, size and pattern to emphasize specific data points or a focus area.
- *“There is no such thing as information overload. There is only bad design.” — Edward Tufte*
- The following three cases indicate how to accentuate what you want to communicate to the viewer.

Design principles Categorical, time series, and statistical data graphics

- **4. Emphasize the most important facts**
- **A. Eliminate distractions**
- Along with using the right chart, it is important to use colors and text strategically. Moreover, if there is a conclusive result, you should enhance that in the visualization.



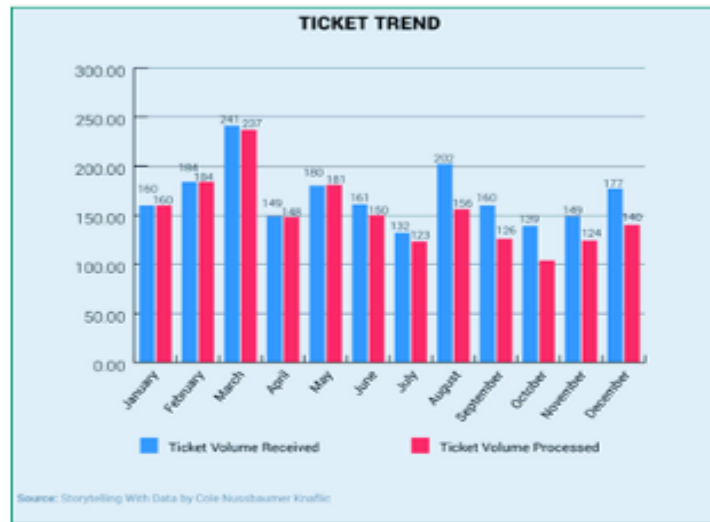
There are too many colors without any purpose and no conclusive result is easily understood



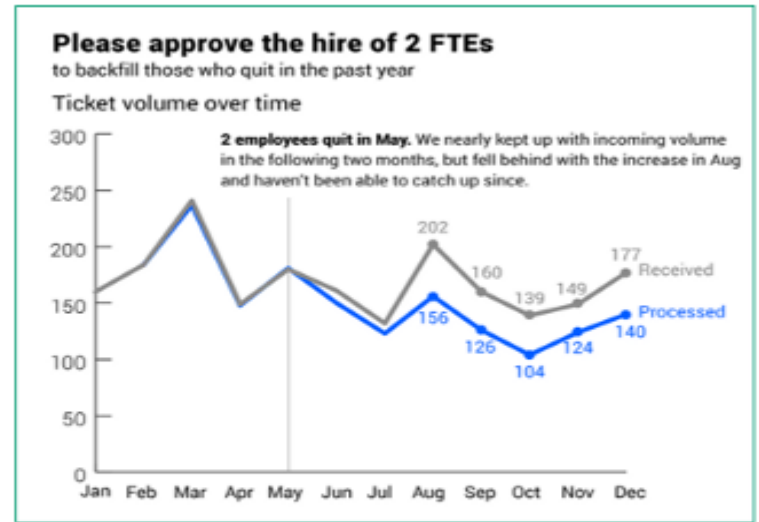
The color scheme and the highlighted bars with text convey the conclusion clearly

Design principles Categorical, time series,

- **4. Emphasize the most important facts**
- **B. Enhance the essential**
- It is helpful to use the title to convey the crux of the viz and remove any irrelevant grids, labels and bright colors. You can also use reference lines and text to draw attention to specific points.



The graph provides no helpful information and does not aid in any decision making



A clear trend is observed, and the line chart supports the suggestion thereby helping in decision making.

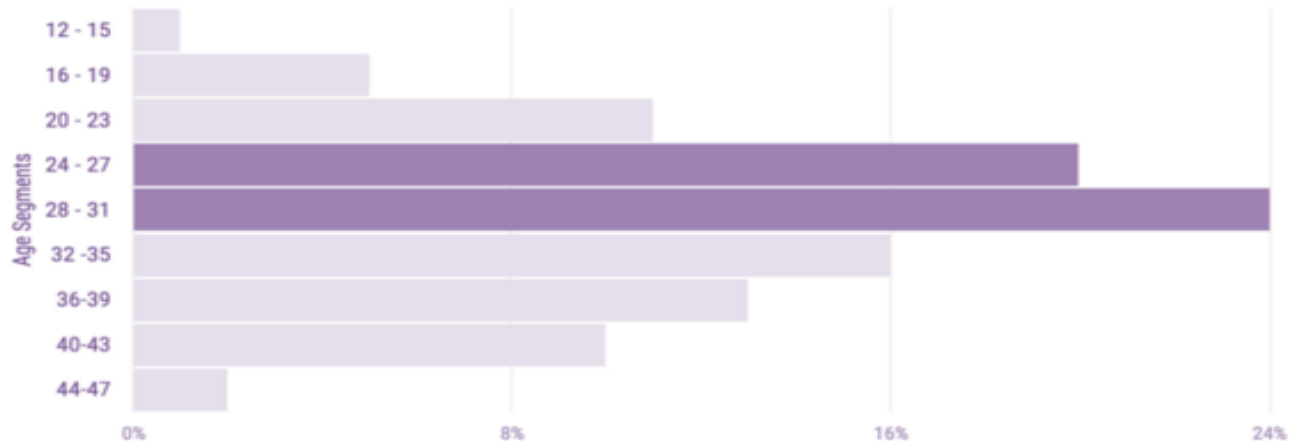
Design principles Categorical, time series,

- **4. Emphasize the most important facts**
- **C. Highlight the significant**
- A distinct color or even different hue/shade can be used to emphasize important data points and highlight areas for attention

Market Segmentation Report

Consumer Profile: Age Demographics

Primary consumer segments are ages 24-27 and 28-31

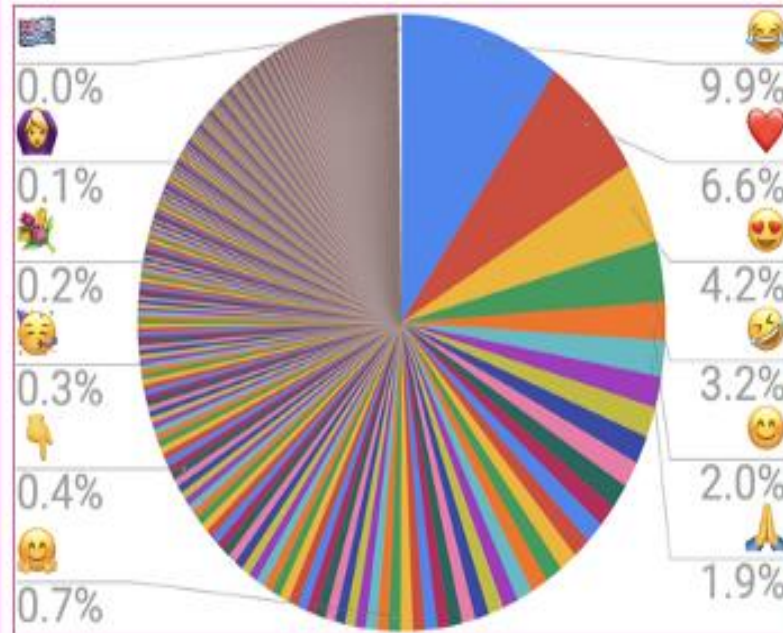


Design principles Categorical, time series, and statistical data graphics

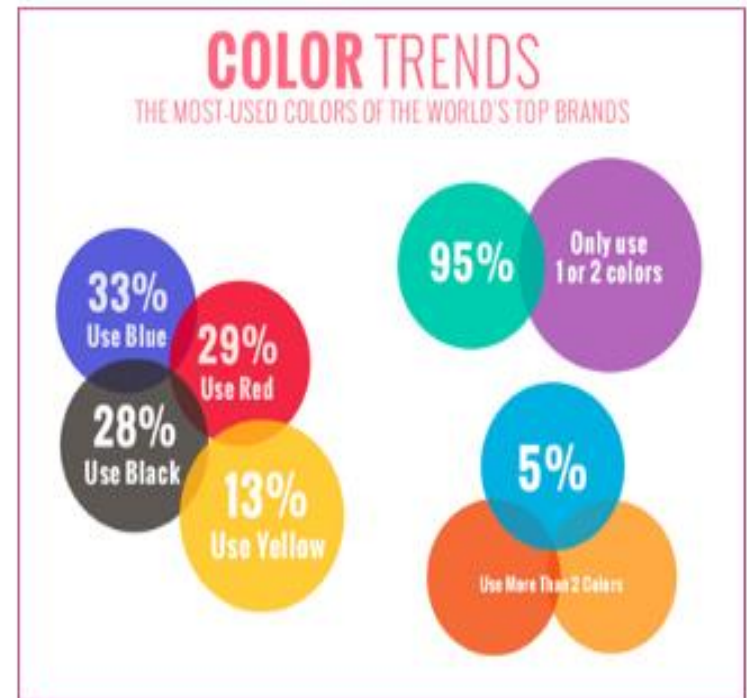
- **5. Form should follow function**
- Aesthetics are important, but should not come at the cost of missing the point. It is more important to communicate the message clearly than to have an eye-candy graph offering no value.
- Purpose of data visualization is insights, not pictures.
- An intuitive design is more important than appealing charts, and graphs should convey the meaning of data in an easy-to-understand manner.
- *“Confusion and clutter are failures of design, not attributes of information.” — Edward Tufte*
- There are many such graphs — here are two that may look fun and pretty but don't successfully impart value to the viewer.

Design principles Categorical, time series, and statistical data graphics

- **5. Form should follow function**



There are too many data points and colors with no apparent information about most of them. Also, the colors have no usefulness.



The size of the circles is disproportionate to the value inside them and the colors of the circles on the right have no purpose

Design principles Categorical, time series, and statistical data graphics

- **5. Form should follow function**
- Data is an invaluable resource and data visualization is one of the most effective tools for analysing and communicating interesting ideas and insights from the data. But badly conceived, incorrectly created or downright untruthful visualizations miss the whole point of visualizing data. Remember these basic principles to make your visualizations impactful and prolific!

Design principles of categorical data graphics

- Graphical methods for quantitative data are easily generalized; for example, the scatterplot for two variables provides the basis for visualizing any number of variables in a scatterplot matrix; available graphical methods for categorical data tend to be more specialized

Design principles of categorical data graphics

- For categorical data, a count is more naturally displayed by an area or by the visual density of an area.

Design principles of categorical data graphics

- **Goals and Design Principles for Visual Data Display:**
- Designing good graphics is surely an art, but equally surely, it is one that ought to be informed by science.
- In constructing a graph, quantitative and qualitative information is encoded by visual features, such as position, size, texture, symbols and color. This translation is reversed when a person studies a graph.
- The representation of numerical magnitude and categorical grouping, and the aperception of patterns and their *meaning* must be extracted from the visual display.

Design principles of categorical data graphics

- **Goals and Design Principles for Visual Data Display:**
- There are many views of graphs, of graphical perception, and of the roles of data visualization in discovering and communicating information.
- On the one hand, one may regard a graphical display as a 'stimulus' – a package of information to be conveyed to an idealized observer.
- From this perspective certain questions are of interest: which form or graphic aspect promotes greater accuracy or speed of judgment (for a particular task or question)? What aspects lead to greatest memorability or impact?

Design principles of categorical data graphics

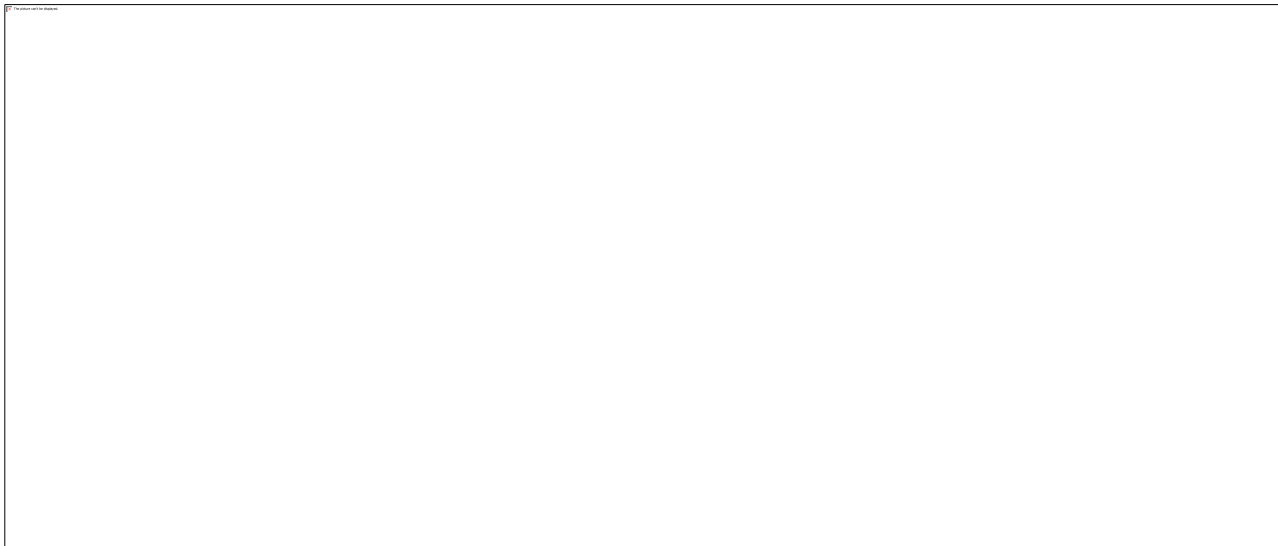
- **Balance**
- **Hierarchy**
- **Scale/Proportion**
- **Dominance/Emphasis**
- **Similarity & Contrast.**
- **Simple colors, minimal text and white space work together to tell a clear story.**

Design principles of categorical data graphics



Design principles of categorical data graphics

- **1. Balance the Design:**
- Equally distribute the visual elements like shape, colour, negative space, and texture across the plot.
- **Types of balances in design:**
- **1. Symmetrical**—Each side of the visual is the same as the other.



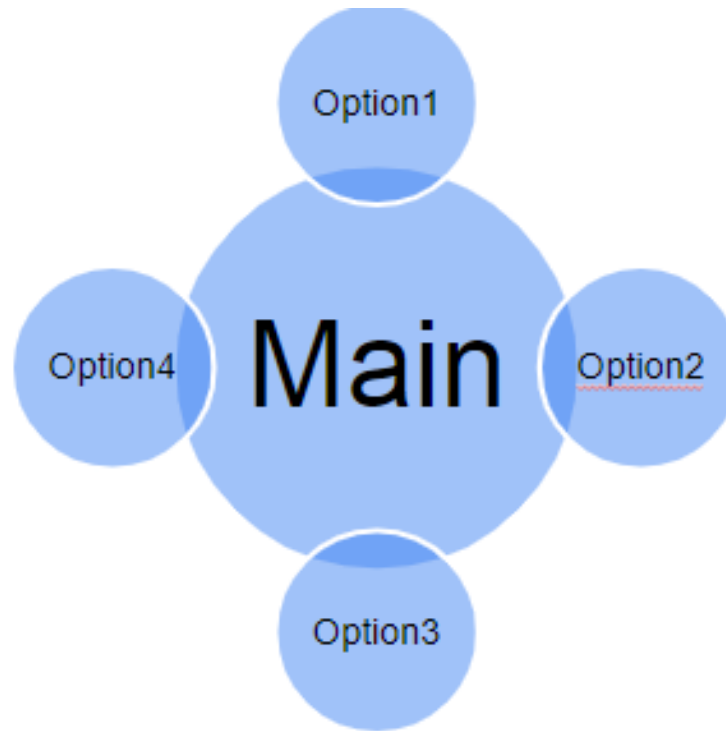
Design principles of categorical data graphics

- **1. Balance the Design:**
- **2. Asymmetrical**—Both sides are different but still have a similar visual weight



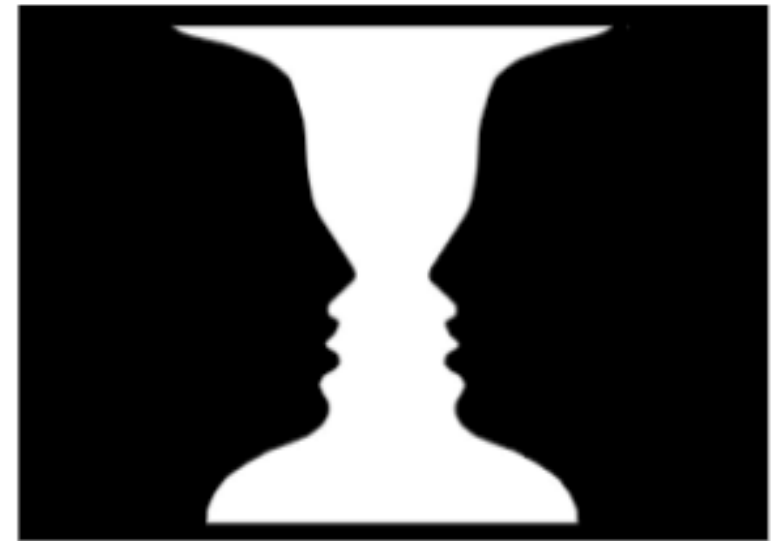
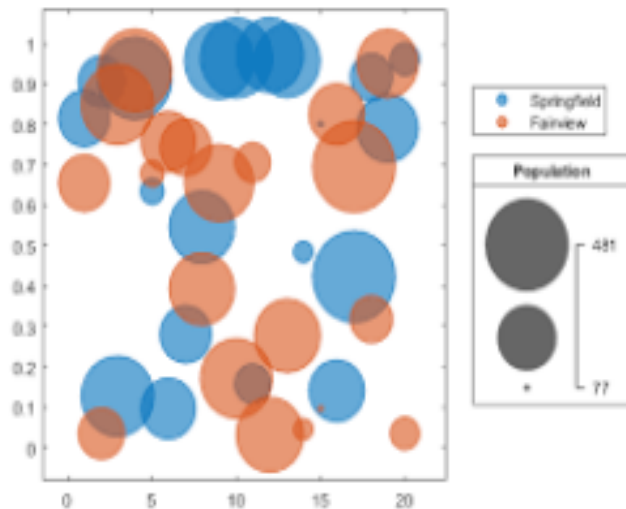
Design principles of categorical data graphics

- **1. Balance the Design:**
- **3. Radial**—Elements are placed around a central object which acts as an anchor



Design principles of categorical data graphics

- **2. Emphasise the Key Areas**
- Ensure important data will not get unnoticed
- Draw the user's attention to the right data points by carefully choosing the size, colours, contrast, and negative space.



Negative Space

Design principles of categorical data graphics

- **3. Illustrating Movement**
- Movement directs the user's attention in a certain direction
- Visual elements should mimic movement in an “F” pattern, which is how people read.
- Starting from top left to right, and gradually down the page.

1 2 3

4

5

6

Design principles of categorical data graphics

- **4. Smart Use of Patterns**
- Repeated design elements form a pattern.
- Patterns make for a great way to display similar types of information spread across the page as one.
- Using similar colours, chart types and elements are the way to go.
- Easier to communicate an anomaly, since any disruption in the pattern will naturally draw the viewer's attention and curiosity.
- Using patterns is one of the simplest and most effective design principles when it comes to data visualization.

Design principles of categorical data graphics

- **6. Proper Rhythm**
- Balanced rhythm when the design elements together create a pleasing movement to the eye

Design principles of categorical data graphics

- **7. Variety**
- keeps viewers engaged and interested
- not only eye-catching but also helps the viewer retain the information presented for longer.
- Use apt type of visuals , graphs suitable for the purpose

Design principles of categorical data graphics

- **8. Theme**
- Design is consistent and follows a standard.
- Incorporate a theme for company or based on the niche of the visualization.
- This helps connect with the user on a deeper level and augments the visual design.

Design principles of categorical data graphics

- Gestalt Principles:
- Gestalt means “unified whole” and is generally associated with the idea that the whole is greater than the sum of its parts. It refers to the patterns that you perceive when presented with a few graphical elements.
- The Gestalt Principles consist of several principles that describe how the human brain sees visual information, namely – proximity, similarity, continuity, closure, connection, and enclosure. People, especially designers who understand these principles, can develop visuals that communicate information in the most effective ways.

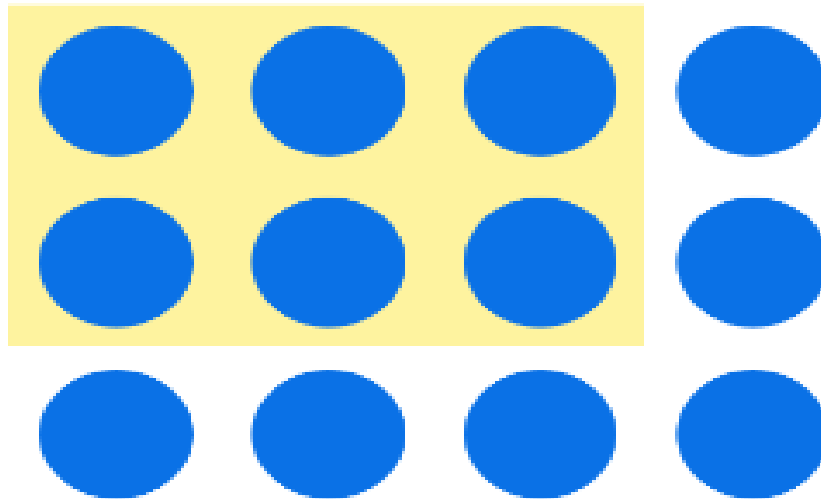
Design principles of categorical data graphics

- Gestalt Principles: **Proximity:**
- The nearer the objects to each other, the more we logically think that these objects belong to the same group
- placing visuals closer together encourages the users to think that the grouped visuals are in the same context.
- The way the objects are positioned in relation to each other can also make the user unconsciously move their eyes from left to right and/or top to bottom.



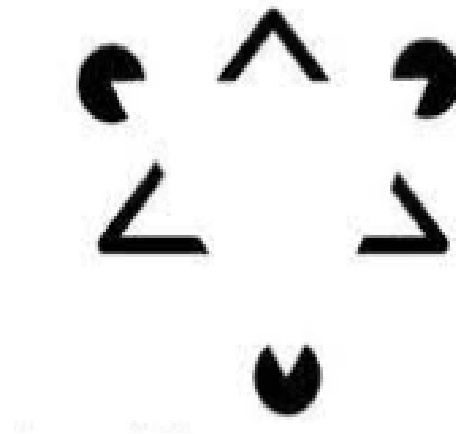
Design principles of categorical data graphics

- Gestalt Principles: Enclosure:
- A group of objects can be enclosed by anything that forms a visual border around them
- This enclosure causes the objects to appear to be set apart in a region that is distinct from the rest of what we see.



Design principles of categorical data graphics

- Gestalt Principles: Closure:
- Our eyes tend to add any missing pieces of a familiar shape that is distinct from the rest of what we see.



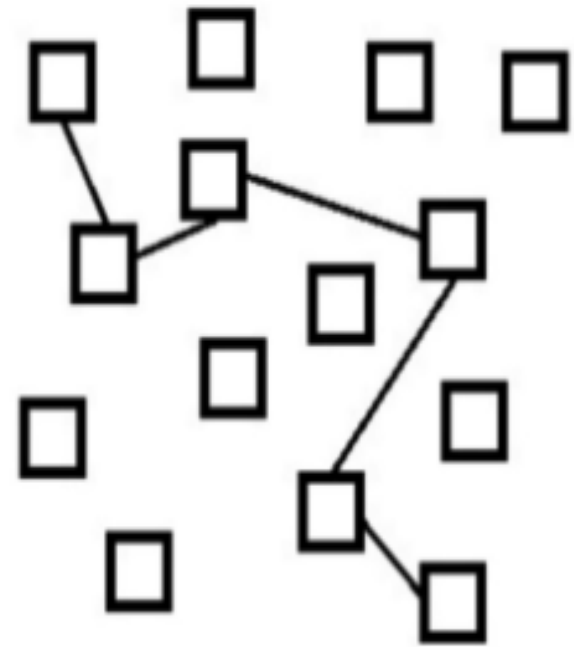
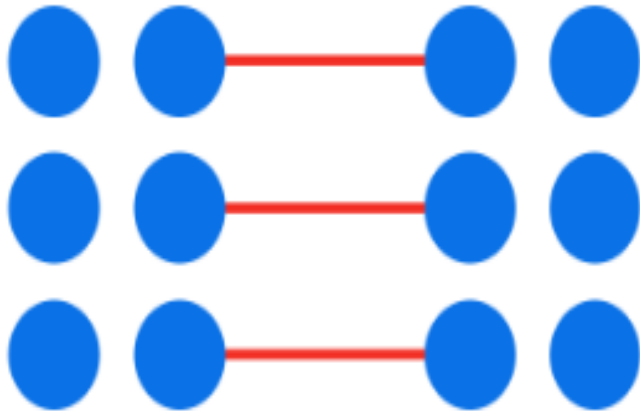
Design principles of categorical data graphics

- Gestalt Principles: Continuity:
- We perceive objects as belonging together, as part of a single whole, if they are aligned with one another or appear to form a continuation of one another.
- It's like the closure principle, but besides the visual connection to form shape, we also attach visual direction as part of the continuation.



Design principles of categorical data graphics

- Gestalt Principles: Connection:
- We perceive objects that are connected in some way, such as by a line, as part of the same group



Design principles of categorical data graphics

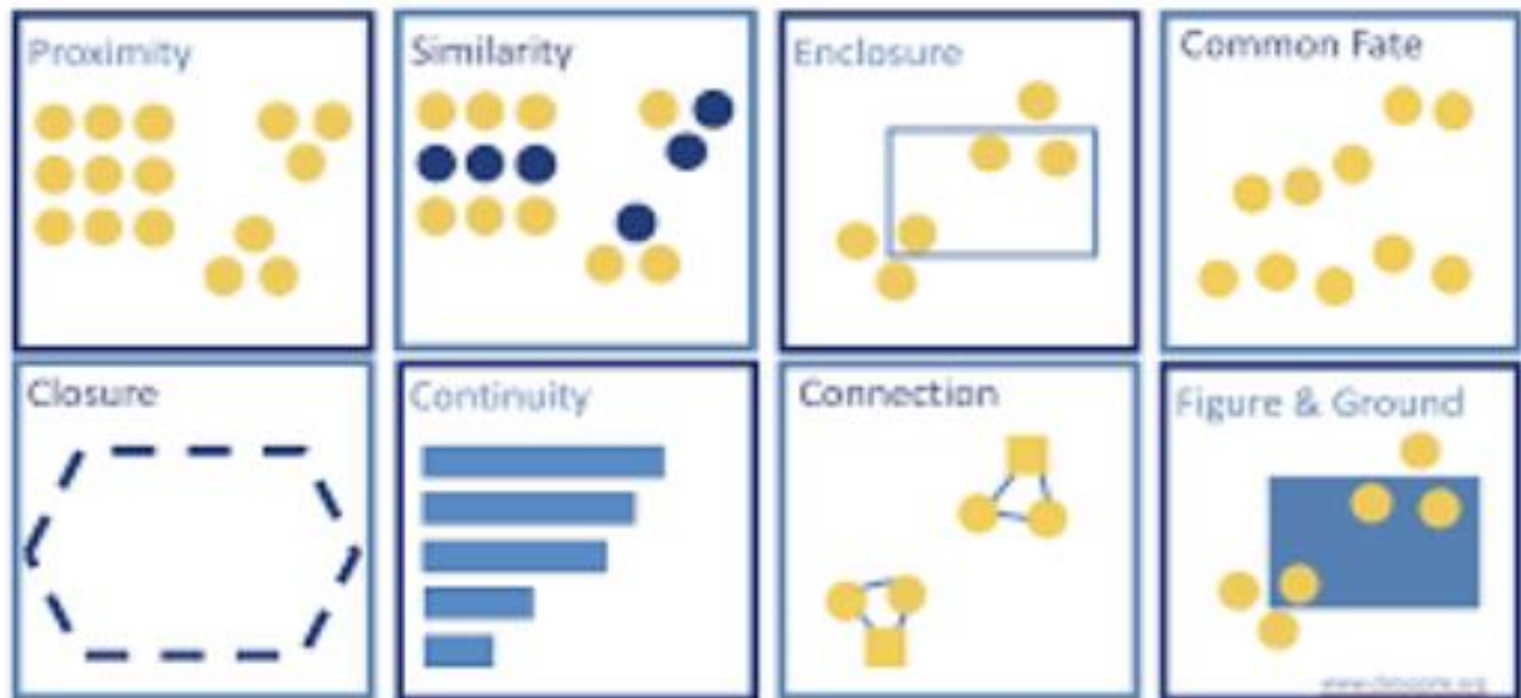
- Gestalt Principles: Figure and Ground
- We add information based on our imagination



Design principles of categorical data graphics

- Gestalt Principles: Figure and Ground
- We add information based on our imagination

Gestalt's Principles of Visual Perception



Design principles of categorical data graphics

- Here are some other principles to consider:
- **Choose Appropriate Graph Type:** Select a graph type that effectively displays categorical data. Common options include bar charts, pie charts, stacked bar charts, and grouped bar charts.
- **Clear Labels and Titles:** Clearly label the graph with a descriptive title that conveys the purpose of the visualization. Label the categorical variables on the appropriate axes, and use concise and informative labels for categories.
- **Consistency in Ordering:** Order the categories logically on the graph. You can order them by frequency, alphabetical order, chronological order, or any other meaningful criterion.
- **Color and Contrast:** Use distinct and contrasting colors for different categories to make them visually distinguishable. Ensure that the colors you choose are accessible and don't cause confusion.
- **Avoid Overcrowding:** If you have many categories, consider grouping or aggregating them to prevent overcrowding the graph. This makes the graph easier to read and interpret.
- **Legend:** If you're using colors or patterns to represent categories, include a legend that explains the meaning of each color or pattern. Place the legend where it's easily visible but doesn't obstruct the data.
- **Data Labels:** Add data labels to individual bars or segments in a pie chart to provide exact values. This helps viewers understand the proportions of each category.
- **Avoid 3D Effects:** Avoid using 3D effects or other unnecessary embellishments that can distort the data or make the graph difficult to read.

Design principles of categorical data graphics

- **Annotations:** Use annotations, such as arrows or text boxes, to highlight specific categories or explain unusual patterns.
- **Data Hierarchy:** If your categorical data has subcategories or levels, use different colors or patterns to represent these levels. This adds another layer of information.
- **Percentage Representation:** In a pie chart, consider displaying the actual percentages within or near each segment for a clearer representation of the distribution.
- **Explode Segments (Pie Chart):** If you want to emphasize a specific category in a pie chart, you can "explode" or pull out that segment slightly from the center.
- **Consistent Widths (Bar Charts):** In a bar chart, maintain consistent widths for bars to accurately represent the differences between categories.
- **Avoid Overlapping Labels:** If you have a large number of categories or long labels, rotate the labels or use alternate labeling strategies to avoid overlapping text.
- **Whitespace:** Use whitespace effectively to separate categories and make the graph easier to read.
- **Contextual Information:** Provide relevant context about the data source, time period, sample size, and any necessary explanations to aid in interpretation.
- **Accessibility:** Ensure that your graphical representation is accessible by using high-contrast colors, providing alternative text for screen readers, and avoiding excessive reliance on color alone to convey information.
- By applying these principles, you can create clear and informative visualizations that effectively communicate the distribution and patterns within categorical data.

Design principles of time series data graphics

- A time series is a sequence of data points that occur in successive order over some period of time.
- In investing, a time series tracks the movement of the chosen data points, such as a security's price, over a specified period of time with data points recorded at regular intervals.
- There is no minimum or maximum amount of time that must be included, allowing the data to be gathered in a way that provides the information being sought by the investor or analyst examining the activity.

Design principles of time series data graphics

- A time series is a data set that tracks a sample over time.
- In particular, a time series allows one to see what factors influence certain variables from period to period.
- Time series analysis can be useful to see how a given asset, security, or economic variable changes over time.
- Forecasting methods using time series are used in both fundamental and technical analysis.

Design principles of time series data graphics

- Representing time series data graphically involves applying several principles to effectively convey the information and insights present in the data. Here are some key principles:
- **Time on the X-axis:** In time series graphs, the independent variable (time) is typically plotted along the horizontal X-axis. This ensures that the chronological order of events is maintained.
- **Data on the Y-axis:** The dependent variable (the data you're tracking over time) is plotted on the vertical Y-axis. This axis represents the values or measurements being recorded.
- **Clear Labels and Titles:** Use descriptive labels and titles for both axes and the graph itself. This helps viewers understand what is being represented without confusion.

Design principles of time series data graphics

- **Consistent Intervals:** Maintain consistent intervals on the X-axis. If the data points are irregularly spaced, consider using connecting lines to indicate the time progression smoothly.
- **Scaling:** Choose appropriate scaling for both axes. Scaling should ensure that the graph uses the available space effectively without distorting the data. Common scaling methods include linear, logarithmic, and percentage scales.
- **Visual Clarity:** Use clear and easy-to-read fonts for labels, titles, and data points. Avoid overcrowding the graph with excessive information that might make it difficult to interpret.
- **Color and Contrast:** Use colors and contrasts that are visually appealing and distinguishable. If multiple lines or data series are plotted on the same graph, use different colors or line styles to differentiate them.

Design principles of time series data graphics

- **Legend:** If you're plotting multiple data series on the same graph, include a legend to explain what each line represents. The legend should be placed in a way that it doesn't obscure the data.
- **Data Density:** Avoid overcrowding the graph with too many data points. Depending on the size of the graph and the resolution, you might need to aggregate data points for better visualization.
- **Annotations:** Use annotations, such as arrows, text boxes, or markers, to highlight important events, trends, or anomalies in the time series. This adds context and helps viewers understand key points.
- **Trend Lines and Averages:** Add trend lines or moving averages to highlight overall trends and smooth out noise in the data, making it easier to identify patterns.
- **Gridlines:** Gridlines on both axes can help viewers accurately read values from the graph. They provide a visual reference to the data points.

Design principles of time series data graphics

- **Time Periods:** If the time series covers multiple time periods (e.g., days, months, years), consider labeling the major time intervals on the X-axis for better comprehension.
- **Contextual Information:** Provide additional information about the data source, units of measurement, and any relevant contextual details that could aid in interpretation.
- **Interactivity:** In digital formats, consider adding interactivity that allows users to zoom in, pan, or toggle different data series on and off for a more customized viewing experience.

By applying these principles, you can create clear and informative time series visualizations that effectively communicate insights from your data.

Design principles of statistical data graphics

- Statistical data are the outcomes or the observations which occur in scientific experiments or an investigation.
- To conduct any analysis it is must to have some data .
- Without data we can not think about research or statistical analysis.
- In statistics, data plays a vital role in all fields and all the theories and measurement.
- Measure of central tendency(mean,median,mode), measure of dispersion (variance,mean deviation, standard deviation etc)are some statistical measure by which we find the different characteristics of the data.
- For example, In a garments factory, we want to find the female workers' height and weight. If we consider the size in feet and weight in kilograms,then we get some numerical values, which are the numerical data.

Design principles of statistical data graphics

Types of statistical data

- All statistical data may be classified into two categories:
 - **Qualitative:** Gender, Education status, marital status etc
 - **Quantitative:** Age, height, weight etc

Design principles of statistical data graphics

We can divide data into 2 categories based on data collection as,

- 1. Primary data:** collected from individuals directly and these data has never been used for any purpose earlier.
- 2. Secondary data:** The data which had been collated by some individual or agency and statistically treated to draw certain conclusions. The same data are used and analyzed to extract some other information, which is termed as secondary data.

Design principles of statistical data graphics

Methods of data collection

- **Direct personal inquiry method**
- **Indirect oral investigation**
- **By mailed questionnaires**
- **Information from local agents by old records**
- **By direct observational method**

Design principles of statistical data graphics

Requirements of reliable statistical data

- It should be complete
- It should be consistent
- It should be accurate
- It should be homogenous with respect to the unit of information.

Design principles of statistical data graphics

5 Design principles of statistical data graphics

1. Show the data clearly

- Showing the data clearly includes ensuring the data points can be seen but also providing meaningful text on the graph itself. Tips include:
- Check the data points plotted can be detected, and are not covered up or obscured.
- Don't assume the viewer is a mind-reader ... label titles and axes clearly and accurately.
- Maintain constant measurement scales and avoid distortions.

Design principles of statistical data graphics

5 Design principles of statistical data graphics

2. Use simplicity in design of the graph

- A graph with a simple design strives for a clean, uncluttered look. Simplicity in design does not mean simplicity in data however; well-designed graphs can represent rich data.
- Avoid distortions, shading, perspective, volume, unnecessary colour, decoration or pictograms, and 3D.

Design principles of statistical data graphics

5 Design principles of statistical data graphics

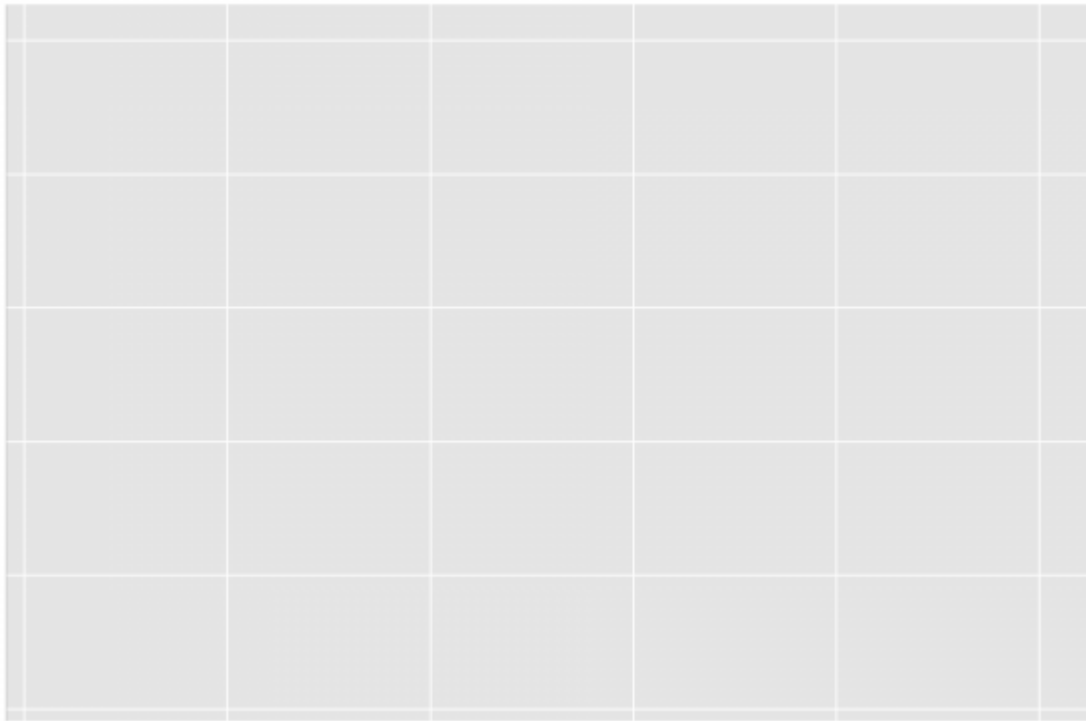
3. Use alignment on a common scale

- Good graphs support accurate estimation of the quantities represented. To estimate quantities, the reader needs to understand the scale used to represent quantity on the graph.
- Use a single linear scale whenever possible.
- Use a common scale if a single scale is not possible, for example, when using panels.
- Avoid pie charts, doughnut charts, stacked bar charts.
- Add gridlines to assist with accurate comparison and estimation; a grey background with white gridlines is often effective.

Design principles of statistical data graphics

5 Design principles of statistical data graphics

3. Use alignment on a common scale



Design principles of statistical data graphics

5 Design principles of statistical data graphics

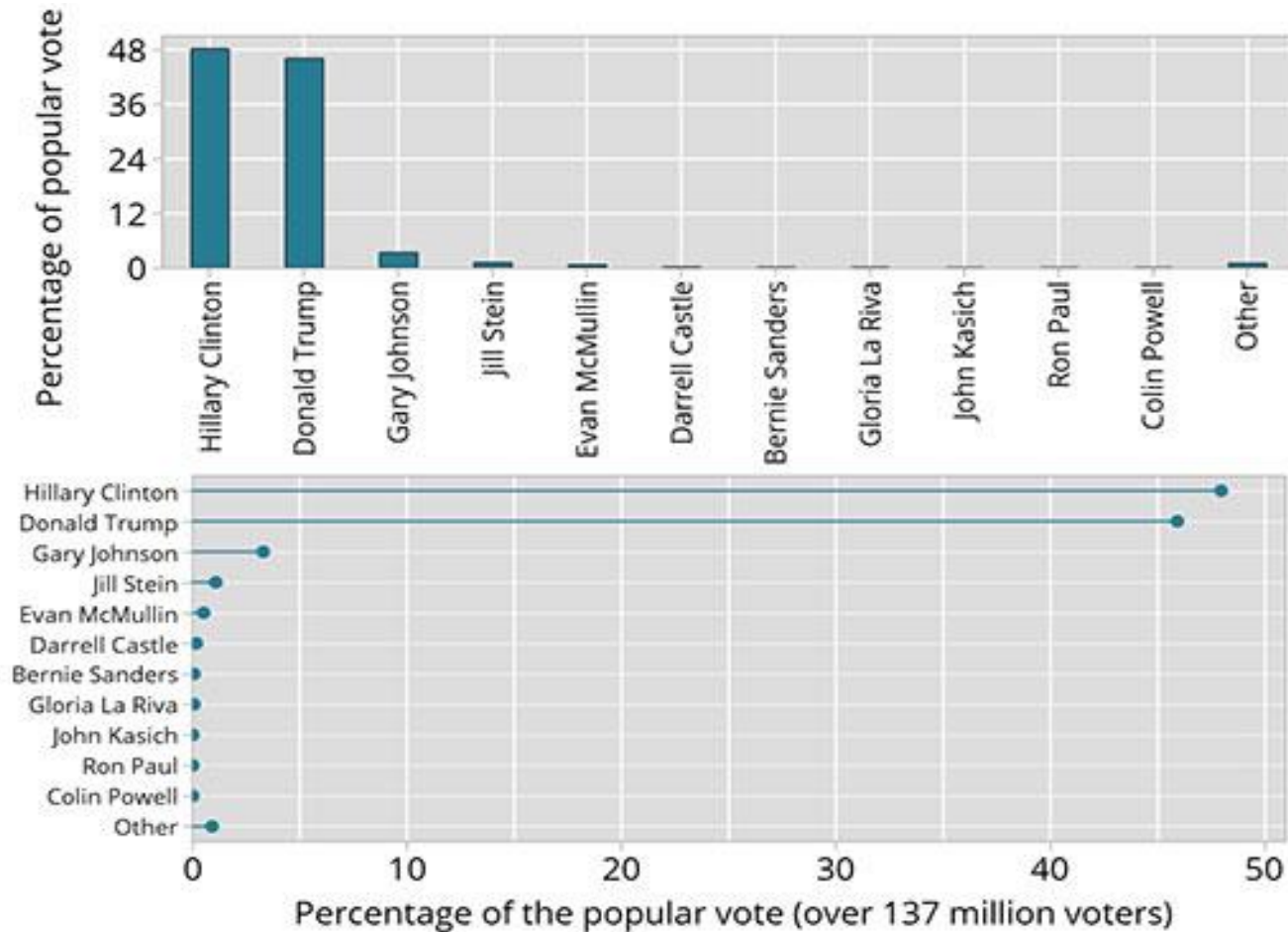
3. Use alignment on a common scale

- Consider transposing a figure to find the easiest way to display the common scale.
- The 2016 popular vote in the US Presidential election provides a simple example of a transposition:

Design principles of statistical data graphics

5 Design principles of statistical data graphics

3. Use alignment on a common scale



Design principles of statistical data graphics

5 Design principles of statistical data graphics

4. Keep the visual encoding transparent

- Graphs use visual encoding to represent data. Readers need to decode the graph. This works best when the decoding task is made easy by astute choices in the design of the graph, making the visual encoding transparent and the decoding effortless. Here are some tips:
- Follow principles 1, 2 and 3!
- Use a common scale if a single scale is not possible.
- Consider how the ordering of elements affects decoding.
- Use colour and shading with meaning.
- Solve detection problems with
 - panels
 - open circles or small points
 - jittering

Design principles of statistical data graphics

5 Design principles of statistical data graphics

5. Use standard forms that work

- Accurate interpretation of graphs relies on a shared understanding of the definition of the graphical objects used. Using standard forms for standard needs supports more reliable interpretation of data visualisations.
- Adhere to the principles that define graphical objects.
- Use standard forms
 - Dotplot
 - Histogram
 - Boxplot
 - Time series plot
 - Bar chart or dot chart
 - Line plot of estimates
 - Estimates as points, with confidence interval bars

Design principles of statistical data graphics

- Here are some principles to consider:
- **Choose the Right Graph Type:** Select a graph type that is appropriate for the type of data and the message you want to convey. Common graph types include bar charts, line graphs, scatter plots, pie charts, histograms, and box plots.
- **Simplicity:** Keep the graph simple and focused on conveying the main message. Avoid cluttering the graph with excessive data or decorative elements that don't contribute to understanding.
- **Labels and Titles:** Clearly label both axes, providing units of measurement if necessary. Include a descriptive title that succinctly conveys the purpose of the graph.
- **Consistency:** Maintain consistent formatting and design elements across graphs within the same presentation or report. This consistency makes it easier for viewers to compare and understand the data.
- **Scale Appropriately:** Choose appropriate scales for both axes. Scaling should accurately represent the data without distorting the visual perception. Avoid manipulating scales to exaggerate differences or trends.
- **Axis Intervals:** Use appropriate intervals for tick marks on the axes. Ensure that the intervals are easy to read and understand. If necessary, use labels to show cumulative data.

Design principles of statistical data graphics

- **Color Selection:** Use a limited and cohesive color palette to distinguish different data elements. Ensure that the colors you choose are easily distinguishable for viewers with color vision deficiencies.
- **Annotations:** Include annotations, such as text labels, arrows, or callouts, to highlight specific data points, trends, or significant events in the data.
- **Data Integrity:** Avoid distorting the data by using 3D effects, exaggerated visual elements, or inappropriate scaling. Always strive to represent the data accurately.
- **Data Hierarchy:** Use graphical elements (like different colors or patterns) to represent subcategories within the data. This helps viewers understand the hierarchy and relationships.
- **Legend:** If the graph includes multiple data series, provide a legend that explains what each element represents. Place the legend where it doesn't obscure the data

Design principles of statistical data graphics

- **Gridlines:** Use gridlines sparingly to guide the viewer's eyes along the data points. Overuse of gridlines can clutter the graph.
- **Data Labels:** Add data labels, especially for individual data points on a bar chart or scatter plot. This can provide specific values and aid in interpretation.
- **Whitespace:** Utilize whitespace strategically to separate data elements and make the graph more visually appealing.
- **Contextual Information:** Provide context about the data source, time period, sample size, and any relevant explanations that help viewers interpret the graph accurately.
- **Consistency in Symbols:** If you use symbols (e.g., markers on a scatter plot), ensure that they are consistent across different graphs or data points.
- **Accessibility:** Ensure that your graphs are accessible to a diverse audience, including individuals with disabilities. Provide alternative text for screen readers and choose colors that are legible for all users.

By applying these principles, you can create graphical representations of statistical data that effectively communicate insights and support data-driven decision-making.

Microsoft excel formulas

- SUM:
- **Example**
- =SUM(A1:A5)
- Result = 41 (See Image below)

The image shows a screenshot of the Microsoft Excel interface. At the top, the formula bar displays the active cell B1 and the formula =SUM(A1:A5). Below the formula bar, a spreadsheet grid is visible. Column A contains the values 10, 5, 7, 9, and 10 in rows 1 through 5. Cell B1 is highlighted with a red border and contains the formula =SUM(A1:A5). The columns are labeled A, B, C, and D, and the rows are labeled 1 through 6.

	A	B	C	D
1	10	=SUM(A1:A5)		
2	5			
3	7			
4	9			
5	10			
6				

Microsoft excel formulas

- COUNT: This basic Excel function counts the numeric value in one or more cells or ranges.
- Ans is 4

SUM		:	✕	✓	<i>fx</i>	=COUNT(A1:A5)
	A	B	C	D		
1	10	=COUNT(A1:A5)				
2	5					
3	A					
4	9					
5	10					
6						

Microsoft excel formulas

- COUNTA: This formula counts the value in one or more cells (This will measure the cells irrespective of the number or text value).
- Ans is 5

SUM		:	✕	✓	<i>fx</i>	=COUNTA(A1:A5)
	A	B	C	D		
1	10	=COUNTA(A1:A5)				
2	5					
3	A					
4	9					
5	10					
6						

Microsoft excel formulas




- **COUNTBLANK in Excel**
- This Excel basic function counts the blank value in the range. (Note: We will not consider only space in a cell a blank cell).
- Ans is 2

The screenshot shows an Excel spreadsheet with columns A, B, and C, and rows 1 through 6. Column A contains the values 10, A, and 10 in rows 1, 3, and 5 respectively. Column B contains the formula =COUNTBLANK(A1:A5) in row 1, which is highlighted with a red border. The formula bar at the top shows the formula =COUNTBLANK(A1:A5) in a yellow box. The status bar at the bottom shows the formula =COUNTBLANK(A1:A5) in a yellow box.

	A	B	C
1	10	=COUNTBLANK(A1:A5)	
2			
3	A		
4			
5	10		
6			

Microsoft excel formulas

- **AVERAGE in Excel**
- This basic formula in Excel is used to get the value average in one or more cells or ranges.
- Ans is 4

SUM		:	  	=AVERAGE(A1:A5)	
	A	B	C	D	
1	1	=AVERAGE(A1:A5)			
2	3				
3	5				
4	1				
5	10				
6					

Microsoft excel formulas

- **MIN Formula in Excel**
- This Excel basic function is used to get the minimum value in cells or ranges.
- Ans is 2

SUM		:	X	✓	<i>f_x</i>	=MIN(A1:A5)
	A	B	C	D		
1	2	=MIN(A1:A5)				
2	5					
3	3					
4	9					
5	7					
6						
7						

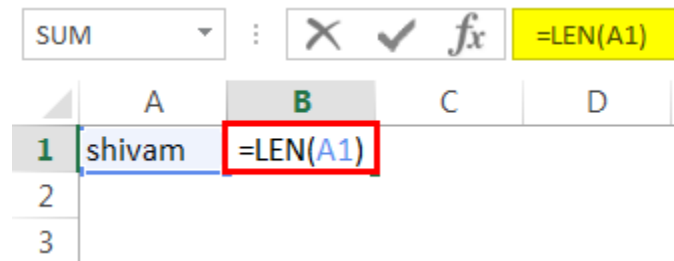
Microsoft excel formulas

- **MAX Formula in Excel**
- This basic Excel function gets the maximum value in cells or ranges.
- Ans is 9

SUM		:	X	✓	<i>fx</i>	=MAX(A1:A5)
	A	B	C	D		
1	2	=MAX(A1:A5)				
2	5					
3	3					
4	9					
5	7					
6						

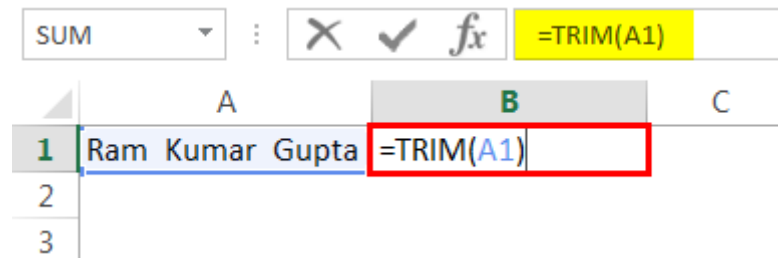
Microsoft excel formulas

- **LEN in Excel**
- This basic function of Excel calculates the number of characters in a cell or text.
- Ans is 6






Microsoft excel formulas

- **TRIM in Excel**
- This basic Excel function removes unnecessary space in a cell or text.
- Ans is: Ram Kumar Gupta



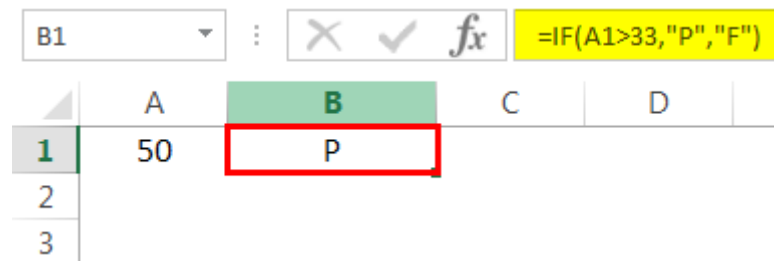
Microsoft excel formulas

- **IF in Excel**
- IF function can perform a logical test in Excel.
- =IF(A1>33,"P","F"). A value in cell A1 is 50, and a logical test is if the value is greater than 33. Then the result would be P. Else, the result would be F.

SUM		:	  		=IF(A1>33,"P","F")	
	A	B		C	D	
1	50	=IF(A1>33,"P","F")				
2						
3						

Microsoft excel formulas

- **IF in Excel**
- Since the value 50 is greater than 33, the result would be P.
(See image below).
-



The image shows a screenshot of the Microsoft Excel interface. At the top, the formula bar displays the formula `=IF(A1>33,"P","F")` in a yellow background. Below the formula bar, a spreadsheet grid is visible. Column A contains the value 50 in row 1. Column B, which is highlighted in green, contains the result 'P' in row 1. The cell B1 is outlined with a red border. The spreadsheet grid shows columns A, B, C, and D, and rows 1, 2, and 3.

	A	B	C	D
1	50	P		
2				
3				

Microsoft excel formulas

- **CONCATENATE Formula in Excel**
- It is one of the formulas used with multiple variants, which helps us join several text strings into one text string.
- For example, if we want to show “Employee ID” and “Employee Name” in a single column.

CONCATENATE...		:	X ✓ fx	=CONCATENATE(B3,C3)
	B	C	D	
1	Table A			
2	Employee ID	Employee Name	Employee Salary	
3	1	Aman Gupta	55000	
4	3	Amit Patra	84000	
5	2	Priyanshu Das	70000	
6	6	Chandan Kale	82000	
7	4	Rajesh Ved	140000	
8	6	Harsha Dangi	82000	
9				
10				=CONCATENATE(B3,C3)
11				

Microsoft excel formulas

- **ROUND Formula in Excel**

D5				=ROUND(B5,C5)	
	A	B	C	D	E
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					

ROUND function

Number	Digits	Result	
5.7845	1	5.8	// round to 1 decimal place
5.7845	2	5.78	// round to 2 decimal places
5.7845	3	5.785	// round to 3 decimal places
23542.5	0	23543	// round to nearest whole number
23542.5	-1	23540	// round to nearest 10
23542.5	-2	23500	// round to nearest 100
23542.5	-3	24000	// round to nearest 1000
23542.5	-4	20000	// round to nearest 10000

EXCELJET