Data Visualization & Storytelling

2.1: Intro to Data Visualization

Learning Goals 🎯

- **Understand the role and importance of data visualization** for data analysts.
- **Learn how visualizations aid analysis, communication, and the discovery of new perspectives.**

1️⃣ Introduction to Data Visualization 🌟

- **Goal**: Move from numeric data analysis to **visual analysis** to interpret spatial patterns (e.g., trends across regions or demographics).
- **Purpose**: Visuals help us spot patterns, see correlations, and communicate complex data quickly and clearly.

Example: You've completed descriptive (mean, standard deviation) and inferential (hypothesis testing) analyses, and now you're adding visuals to make your findings more understandable and compelling for stakeholders.

What is Data Visualization? 📊

- **Definition**: Data visualization is the representation of data through **charts, maps, diagrams,** etc., to simplify and enhance data interpretation.
- **Why Visualize?**: Visuals communicate data much faster and are easier to understand than tables. Humans process images quickly, making visual data representations like maps and graphs invaluable for effective communication.

Examples:

- **Maps**: A map showing black bear and Bigfoot sightings in Ohio reveals that the sightings overlap in wooded areas, suggesting many "Bigfoot" sightings might actually be black bears.
- **Bar Charts**: Instead of reading a table of sales numbers, a bar chart shows at a glance which months had the highest and lowest sales.

The Evolution of Data Visualization 📜

1. Mapping (Ancient Times) 🗺️

- **Importance**: Ancient civilizations used maps to show trade routes and navigate land.
- **Oldest Map**: The first known map dates back to 6200 B.C. in Babylon.
- **Example Today**: We still use maps to navigate, like GPS on phones, which provides both **verbal directions and visual maps** to make complex routes easy to follow.

2. Coordinate System (1600s) 📐

- **Created by**: René Descartes, a French mathematician.
- **Contribution**: Developed the **Cartesian coordinate system** (x- and y-axes), which lets us plot values in 2D space, helping us see relationships in data.
- **Example Today**: Scatter plots and line graphs are based on this system and help us analyze relationships between two variables.

3. Proportional Representation (1782) 📏

- **Created by**: Charles Louis de Fourcroy.
- **Contribution**: Used squares to represent the sizes of European cities proportionally, allowing viewers to quickly compare city sizes.
- **Modern Example**: Population density maps where size or color intensity represents population—larger or darker areas mean more people.

4. Introduction of Charts and Graphs (1800s) 📈

- **Created by**: William Playfair, a Scottish engineer.
- **Contribution**: Developed **bar charts, line charts, and pie charts** to help explain economic data.
- **Example Today**: Line charts show trends over time, like a company's revenue growth or a stock's performance over the years.

5. Thematic and Density Maps (1800s) 🌍

- **Thematic Maps**: Show data based on a theme (e.g., literacy rates across regions).
- **Density Maps**: Use symbols to show event clusters (e.g., John Snow's 1854 cholera map).
- **Example**: John Snow mapped cholera deaths in London, revealing that cases clustered around a single water pump, helping him identify it as the source of the outbreak.

6. Florence Nightingale and Polar Area Charts (1850s) 🌹

- **Who**: Florence Nightingale, a nurse during the Crimean War.

- **Contribution**: Created **polar area charts** to show that most British soldiers died from preventable diseases rather than combat wounds, pushing for better sanitation.
- **Example Today**: Nightingale's "rose diagram" inspired pie charts and other circular visuals, which are still widely used.

7. Simplification and Modern Mapping (1900s) 👩‍💻

- **Notable Figure**: Harry Beck, who simplified London's subway map in 1933, making it easier to navigate.
- **Standardization**: Jacques Bertin and Edward Tufte further standardized visualizations, emphasizing **clarity** and **simple, effective data representation.**
- **Example Today**: Modern subway maps and Tufte's principles guide analysts to create uncluttered, meaningful visuals.

Data Visualization with Computers 💻

Visualization Software 🔧

- **Popular Tools**: Excel, Tableau, PowerBI, open-source tools like ggplot2 in R and matplotlib in Python.
- **Benefits**:
    - **Excel**: Great for basic charts and accessible to everyone.
    - **Tableau**: Industry-standard for interactive and advanced visuals, which you'll be learning in this course.
    - **Python/R**: Highly customizable but requires coding knowledge.

Interactivity 🔄

- **Interactive Visuals**: Allow users to explore data by changing parameters themselves.
- **Static vs. Interactive**:
    - **Static**: Fixed visuals (like a PDF or PowerPoint slide).
    - **Interactive**: Users can change parameters (e.g., year intervals, data segments), making exploration more accessible without needing an analyst to create new charts each time.

Example: Instead of a fixed 10-year line graph, an interactive chart lets users view data for 5-year intervals or different regions, providing a deeper understanding of trends.

The Power & Purpose of Visualization 💪

1. Visualization for Analysis 🔍

- Visuals help analysts identify trends, outliers, and relationships that raw data alone might miss.
- **Anscombe's Quartet**:
    - Statistician Francis Anscombe created four data sets with identical descriptive stats (mean, variance, etc.), but when visualized, they revealed distinct patterns, highlighting how visuals expose data nuances.
- **Core Insight**: Visualizations are essential in exploratory data analysis to truly understand data.

2. Visualization for Communication 🗣️

- Visuals democratize data, making complex information more accessible to non-analysts.
- **Why It Matters**: Tables and numbers require more effort to interpret, while charts and maps convey key insights instantly.
- **Role of the Analyst**: Acts as a bridge between stakeholders and data by using visuals to make findings clear.

Example: In your analysis project, creating a storyboard with visuals for each key finding will make it easier for stakeholders to understand your results without needing technical knowledge.

3. Visualization for New Perspectives 💡

- Visuals can challenge assumptions and spark new ways of thinking.
- **Notable Examples**:
    - **Florence Nightingale**: Used data visuals to shift public perception and improve British army sanitation.
    - **W.E.B. Du Bois**: Developed compelling visuals to dispel stereotypes about African Americans, highlighting disparities and progress in areas like education and employment.
    - **Hans Rosling**: Used animated bubble charts to show child survival rates and economic changes, challenging outdated views on "developing" countries.

Example: Rosling's work demonstrated that many countries now have small family sizes and high child survival rates, breaking the stereotype of "developing" vs. "developed" countries.

Summary 🏆

- **Data Visualization** turns raw data into visuals to communicate insights more effectively.
- **Uses of Visualization**:

1. **Analysis**: Helps analysts see patterns, outliers, and correlations not visible in raw data.
2. **Communication**: Makes data accessible to stakeholders, allowing them to quickly understand complex information.
3. **New Perspectives**: Challenges assumptions and offers fresh insights, potentially influencing decisions and policy.

Quiz: 2.1: Intro to Data Visualization

1. Why do analysts use data visualization as part of the analysis process?
    a. To identify trends that are unclear in table format
        i. By creating data visualizations during an analysis, an analyst can discover trends in the data that may be unclear in a table format. Unlike measures of central tendency or standard deviation, trends can't be easily identified without a data visualization.

2. Harry Beck's 1933 London Underground map is an example of how modern day data visualization evolved in what way?
    a. Increased simplification and refinement
        i. When Harry Beck took a complex and cluttered map of the London Underground and created a new version, he demonstrated how modern-day data visualization resulted in increased simplification and refinement. Many visualization methods that are still in use today originated from early data visualizations, but the modern era brought more simple and effective visualizations.

3. Kim wants to convince her company's chief marketing officer (CMO) to increase her department's marketing budget in order to increase product sales. Her data shows that previous increases in marketing spending have resulted in increased sales. How could Kim use a visualization to help her case?
    a. Create a line chart that plots historical marketing spending and sales to show how increased spending has resulted in more sales.
        i. To help convince her CMO that a change is needed in marketing budget, Kim should plot historical marketing spend and sales in a line chart to show how an increase in marketing spending led to more sales. Presenting this compelling data in an easy-to-understand visualization may help change the perspective of her CMO.

4. What's the core function of data visualization?
    a. To communicate with others about data and an analysis
        i. While data visualization is used by analysts to identify trends and to prompt new understandings, the core function of data visualization is to communicate with others about data. This

can be as simple as adding charts to a document or presentation that shares analysis project results with stakeholders.

5. Elizabeth is analyzing voter data for an upcoming state election in California. Elizabeth visualizes the number of voters in each district in a diagram where the size of each visualized district is based on the number of voters, ordered from largest to smallest. What concept is Elizabeth demonstrating?
    a. **Proportional representation**
        i. Because Elizabeth visualizes voting districts in size based on the number of voters within them, she's demonstrating proportional representation. Districts with more voters are displayed with larger areas, while those with fewer voters are displayed with smaller areas. This visualization bridges the gap between the geographical size of districts and the number of voters within, which isn't always correlative.

2.2: Visual Design Basics & Tableau

Learning Goals 🎯

- **Understand styling and design principles** for effective visualizations.
- **Connect to data using Tableau** to create visualizations.

1️⃣ Introduction: Why Visual Design Matters 🖌️

- **Goal**: Design impactful visuals that are clear, appealing, and informative.
- **Importance**: Good visuals make complex data easy to understand and share.
- **Tool Focus**: We'll be using **Tableau**, a powerful software for data visualization.

Elements of Good Data Visualization 🌈

Napoleon's Campaign by Charles Minard 🎨

- **Historical Context**: Minard's map of Napoleon's 1812 Russian campaign is a masterpiece of simplicity. It shows **troop movement, temperature, direction, geography, time, and army size**.

- **Key Point**: This map balances **multiple data points** without overwhelming the viewer, making it a classic example of effective visualization.

1. **Simplicity** ✂️

- **Definition**: Minimize elements; use only essential details.
- **Example**: Minard's map didn't include a full map of Russia—just key cities and routes.
- **Bad Example**: H-1B Visa Circle Graph (Figure 2) shows too many employers with unreadable text, making it confusing.
- **Improvement Tip**: Simplify by showing top employers only and grouping categories with labels.

2. **Text** ✍️

- **Labels**: Titles, axis labels, and legends clarify the visualization. For example, Minard labeled temperature to help understand the harsh conditions of Napoleon's retreat.
- **Callouts**: Explain significant points in the data, like spikes in BrandX social mentions in Figure 3.
- **Legend**: Shows color meanings, sizes, and symbols. A clear legend helps users understand visuals at a glance.
- **Example of Overuse**: Map of minimum wage workers (Figure 5) shows redundant labels for percentages—color alone could convey this information clearly.

3. **Whitespace** 🌌

- **Definition**: Empty space around elements to reduce clutter.
- **Examples**: Apple and Dropbox websites (Figures 6 & 7) use whitespace for a clean look.
- **Bad Example**: xkcd's Money Visualization (Figure 8) is overloaded with information, making it hard to focus.
- **Tip**: If the visualization feels crowded, consider splitting it into multiple visuals.

4. **Size** 📏

- **Purpose**: Larger elements show higher magnitude, like bigger circles for larger values.
- **Example**: Minard's map uses line thickness to represent army size, with each millimeter equaling 10,000 men.
- **Best Practice**: Use size sparingly and with 5 or fewer categories to avoid confusion.

5. **Color** 🎨

- **Role**: Color adds meaning; it's not just decorative.

- **Good Example**: Traffic Light Scheme (Figure 12) uses green for “good,” yellow for “caution,” and red for “bad.”
- **Bad Example**: Avoid stereotypes (e.g., pink for women and blue for men in gender-based data).
- **Color Schemes**:
    - **Monochromatic**: One hue with shades/tones (good for accessibility).
    - **Complementary**: Opposite colors on the wheel (e.g., blue/orange).
    - **Analogous**: Colors next to each other (e.g., red/orange/yellow).
- **Tips for Color-Blind Accessibility**:
    - Use monochromatic schemes or color-blind-friendly palettes (e.g., ColorBrewer).
    - Avoid using red and green together.

Gestalt Psychology 👥

- **Principles**: Gestalt psychology explains how we naturally perceive patterns.
    - **Similarity**: Elements with similar shapes, colors, or sizes are seen as grouped.
    - **Proximity**: Elements close to each other are seen as related.
- **Example**: Using the same color to show related items (like “plastic” waste in one color and other items in gray) helps the viewer see groups and patterns easily.

Using Gestalt Principles in Data Visualization

Gestalt psychology can enhance your design by helping viewers quickly recognize patterns. By strategically using proximity, similarity, and continuity, you can make visuals more intuitive and engaging.

Accessibility 🔓

- **Goal**: Make visuals usable for everyone, including those with disabilities.
    - **For Color Blindness**: Use monochromatic schemes or accessible palettes (ColorBrewer).
    - **Alt Text**: Describe visuals for screen readers.
    - **Readable Labels**: Ensure font size and contrast are clear.

Visual Design Checklist ✅

Before finalizing, ask yourself these questions:

Text 📑

- Are the title and text clear?
- Are labels and legends used effectively?
- Is any text redundant or cluttering?

Color 🌈

- Does the color scheme make sense (analogous, complementary, monochromatic)?
- Is color used to highlight key info with darker hues for focus?
- Are there too many colors (limit to 2-5)?

Other Elements 🎨

- Are sizes used meaningfully (e.g., bigger shapes for larger values)?
- Are items grouped effectively by color, size, or position?
- Is there enough whitespace?

Bottom Line: Does the visualization teach you something clearly? If not, simplify and adjust.

Style Guides 📕

- **Purpose**: Create a consistent, effective look for all visuals.
- **Contents**: Company colors, logo placements, when to use certain visuals, etc.

Getting Started with Tableau 🖥️

What is Tableau?

- **Software for Data Visualization**: It's a tool specifically for creating complex, interactive visuals.
- **Versions**: **Tableau Public** (free but data is public) and **Tableau Desktop** (professional with license fees).
- **Benefit**: Excel is great for calculations; Tableau excels at interactive and polished visuals.

Installing Tableau Public 🚀

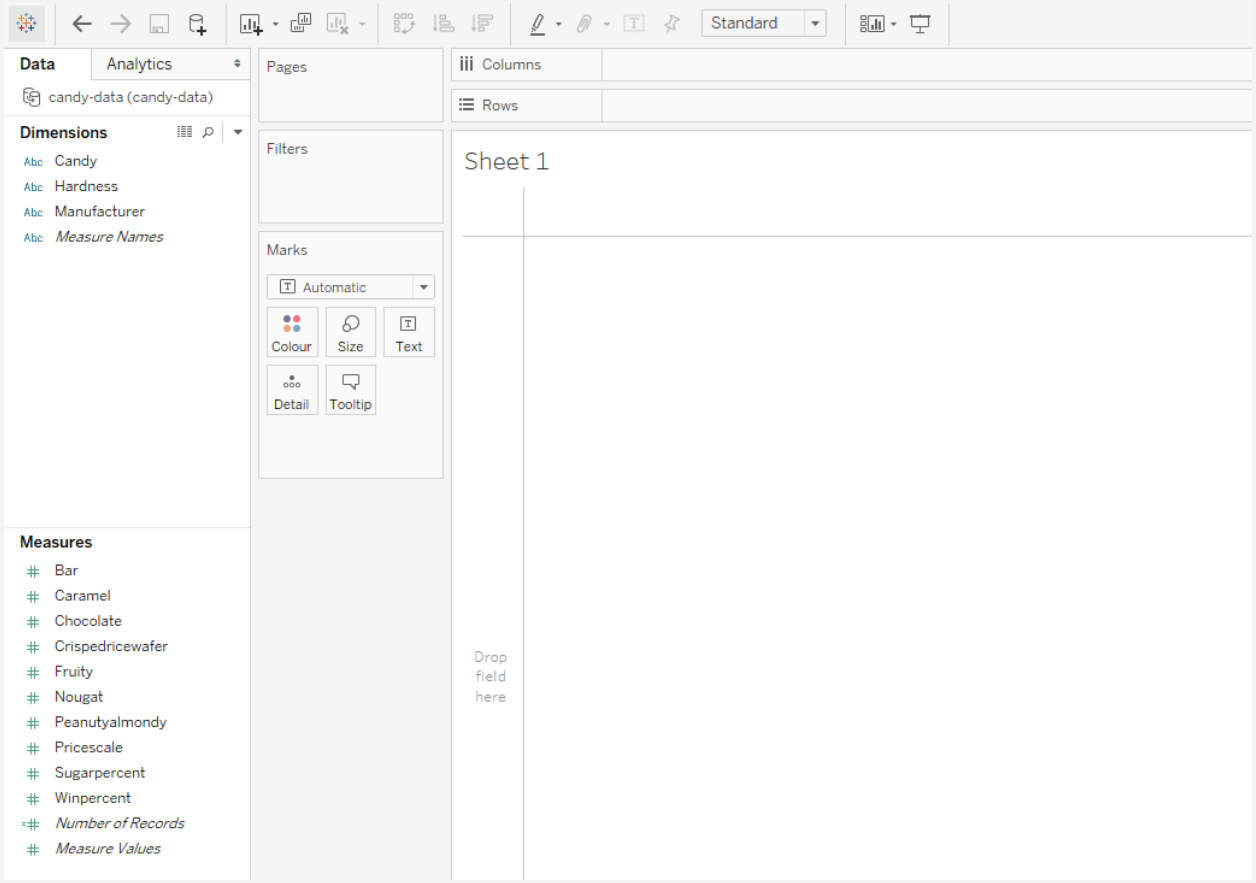
1. **Download Tableau Public**: Go to Tableau Public's site and follow the instructions.
2. **Create an Account**: Required for saving and sharing work online.

Exploring Tableau Interface 📊

1. **Connecting Data**:

    - Go to **Microsoft Excel** in the left panel.
    - Select your **Excel file** to import data.
2. **Data Source Screen**:
    - Displays file name and data sheets.
    - Double-click a sheet to load a preview.
3. **Main Dashboard**:
    - **Dimensions (Categorical)** and **Measures (Numerical)** are on the left.
    - **Blue**: Discrete (e.g., names, categories).
    - **Green**: Continuous (e.g., age, salary).

Dimensions & Measures



Summary 🏆

- **Design Principles**: Simplicity, text clarity, color, size, and accessibility are crucial.
- **Tableau**: A powerful tool for creating effective visualizations.
    - **Remember**: Use Excel for data prep and calculations, Tableau for visualization.

Quiz: 2.2: Visual Design Basics & Tableau

1. Which of the following is an element of a good data visualization?
    a. Simplicity in design

i. Because data visualization is all about presenting complex data in an easy-to-understand pictorial manner, simplicity in design is essential for a good data visualization. This includes minimizing the design so as not to overwhelm the reader and conveying information using the least amount of text or graphics possible.

2. A style guide is a list of rules or guidance for how a company formats visualizations. Which of the following items would a style guide typically include?
    a. A specific color scheme
        i. A company's style guide will usually include a color scheme. These are often branded colors used in the company logo or other marketing materials to bring uniformity to company deliverables. The major unit for axes and the date range to cover will always vary depending on the data, while grammar tips aren't generally covered by a style guide.

3. Under the principle of proportional representation, the size of elements in a visualization should correspond to magnitude. What's the maximum number of sizes or categories a visualization should contain?
    a. 5
        i. When symbols represent discrete size, a rule of thumb is to limit sizes or categories to five or fewer. In visualizations with more than five discrete size categories, it may be hard to tell the difference between each category.

4. When assessing visualizations for accessibility, which of the following factors should be considered?
    a. If the color scheme is color-blind safe
        i. When assessing visualizations for accessibility, the analyst should check that the color scheme is color-blind safe. Monochromatic colors, or those that come from a single base hue, are often color-blind friendly, as are those which vary in tone, shade, and tint.

5. Which terms does Tableau use to differentiate between textual and numerical data?
    a. Textual data is a dimension and numerical data is a measure
        i. In Tableau, textual data (categorical data) is a dimension, and numerical data (data on which you'd perform math) is a measure. While Tableau automatically classifies data into dimensions and measures, it's important to check that these auto-classifications are correct before building data visualizations.

2.3: Composition & Comparison Charts

Learning Goals 🎯

- Understand the use cases for **composition** and **comparison charts** in data visualization.
- Create basic composition and comparison charts such as **pie charts**, **bar/column charts**, and **treemaps** using **Tableau**.
- Practice effectively labeling, organizing, and adjusting charts to enhance clarity and storytelling.

Introduction 🌐

You're building on visual storytelling principles by moving from broad design concepts into specific types of charts. **Composition and comparison charts** are foundational tools that show how different data segments relate to a whole or differ across categories. You'll use Tableau for hands-on creation, focusing on how to adjust, color, and label charts to make data insights stand out.

Types of Visualizations & Key Use Cases 📊

1. Composition Charts

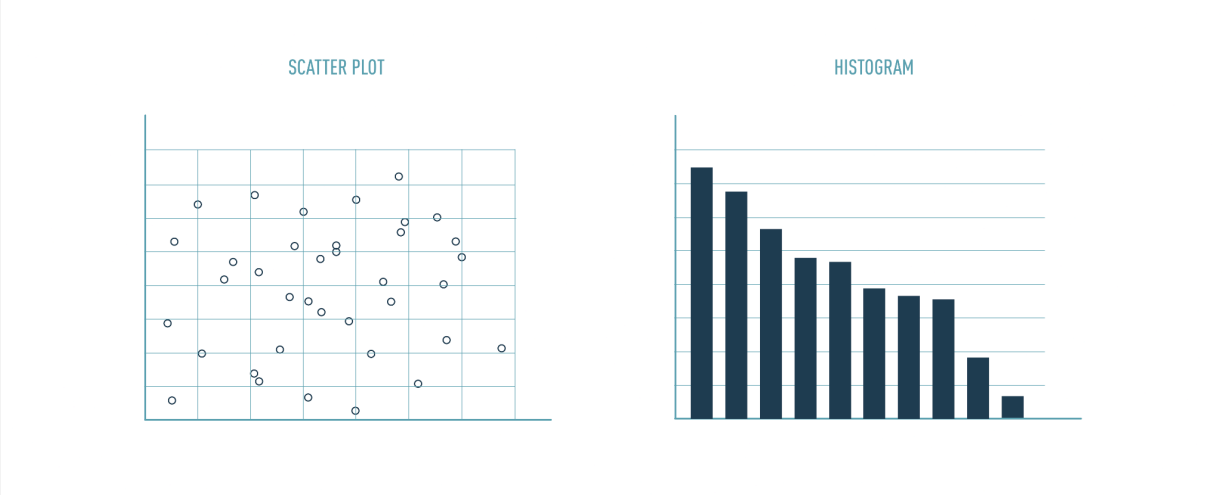
- **Purpose**: Show proportions of a whole and compare group sizes within a single dataset.
- **Common Types**:
    - **Bar Charts**: Ideal for larger datasets and many categories, allowing for direct size comparison.
    - **Pie Charts**: Best for displaying simple proportions with 2-3 clear segments.
    - **Treemaps**: Useful for many small categories or when comparing hierarchical data, like product sales by category.

2. Temporal Charts

- **Purpose**: Show data trends over time, capturing patterns and changes.
- **Common Chart**: **Line chart** (with time on the x-axis).
- **Ideal for**: Visualizing trends, forecasting, and observing seasonal or annual changes.

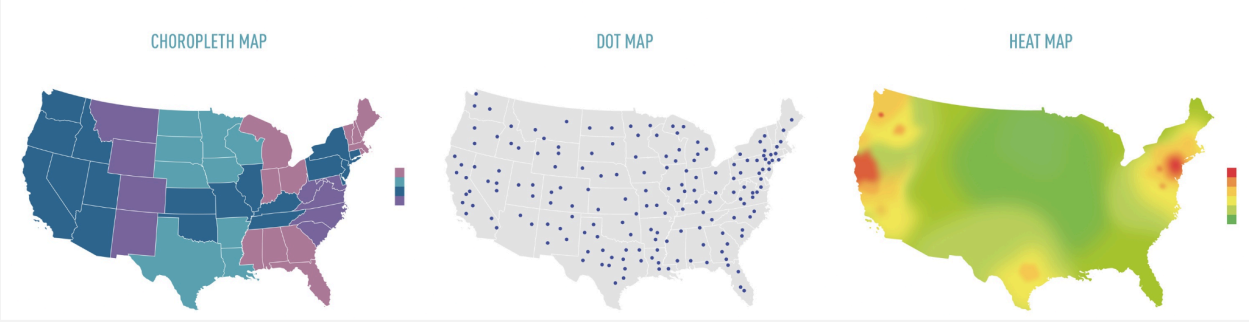
3. Statistical Charts

- **Purpose**: Show data distribution, variability, or relationships between variables.
- **Common Types**:
  - **Histograms**: Visualize frequency distribution across categories.
  - **Scatterplots**: Show correlations between two variables, helping identify patterns or clusters.
- **Use Case**: Statistical charts are ideal for visualizing and interpreting data patterns, distributions, and relationships.



4. Geospatial Charts

- **Purpose**: Map data across regions or locations, highlighting spatial patterns.
- **Common Types**:
  - **Choropleth maps**: Color-coded regions (e.g., by vaccination rates).
  - **Dot maps**: Track events across specific locations.
  - **Heat maps**: Show intensity or density in areas, ideal for “hotspot” analysis.
- **Ideal for**: Regional comparisons, identifying geographic trends, or observing density distributions.



5. Textual Analysis Charts

- **Purpose**: Visualize non-numeric, qualitative data.
- **Common Type**: **Word cloud** (based on frequency).
- **Use Case**: Display text data like survey feedback, where the frequency of certain words or themes indicates areas of focus.

Converting Dimensions & Measures in Tableau 📊

Understanding Dimensions and Measures:

- In Tableau, data is categorized as either **dimensions** (qualitative) or **measures** (quantitative).
- **Continuous (green)** and **discrete (blue)** values help distinguish whether data represents a finite count (discrete) or an infinite range (continuous).
- For example:
    - **Dimensions**: Textual or categorical data such as names, dates, or indicators.
    - **Measures**: Numerical data, like amounts or prices, where values can be aggregated.

Key Concepts for Classifying Data:

1. **Check Indicator Variables** (e.g., columns with 1s and 0s):
    - Indicator variables (like “1” for chocolate-containing candy and “0” otherwise) should be **dimensions**. Although they are numeric, they classify rather than measure.
    - Common example: ID numbers should be dimensions, as adding or aggregating them is meaningless.

2. **Reclassify as Needed**:
    - Right-click the variable > **Convert to Dimension** to move it from the Measures area to Dimensions.
    - Dragging variables up to **Dimensions** will achieve the same result.

Example:

- In a candy dataset, variables like **Chocolate**, **Caramel**, **Peanutyalmondy** are indicator dimensions and should be in the **Dimensions** section, not **Measures**.

Final Tips:

- **Measures** are typically aggregated in visualizations (e.g., total sales).
- **Dimensions** help define categories and groups.
- **Continuous measures** (green) allow detailed plotting, while **discrete dimensions** (blue) categorize data.

Creating Composition Charts in Tableau 🖌️

Composition charts help you to visually break down data by categories to see how parts compare to the whole or with each other. Here's a guide to creating three core types of composition charts:

1. Bar & Column Charts

Definition: **Bar charts** (horizontal) and **column charts** (vertical) display data using rectangular bars. Bar charts are often preferred for longer category labels and can be ordered by value for easy comparison.

Steps to Create a Bar Chart in Tableau:

1. **Select Data**: Choose the category variable (dimension) for the y-axis and a measure (numeric count) for the x-axis.
2. **Convert Data**: Right-click your selected variable and choose "Count" to turn it into a measure if needed.
3. **Adjust Orientation**: For longer category names, select a **bar chart**; for short names, a **column chart**.
4. **Add Color & Labels**:
    - Add a **color dimension** to visualize a second attribute (e.g., candy hardness).
    - Label bars with category names and counts to enhance clarity.
5. **Sort Bars**: Arrange bars by value (ascending or descending) for easy interpretation.

Use Case: Bar charts are ideal for comparing values across categories (e.g., types of candies by manufacturer). They work well for datasets with many categories, especially when arranged by size for visual order.

Tip: Use **stacked bar charts** to add a third variable by stacking bars with color-coded segments, helping viewers understand multiple layers of information within one chart.

2. Pie Charts 🥧

Definition: A pie chart is a circular chart divided into slices, each representing a portion of the whole. Each slice's size reflects the percentage of each category within a dataset.

Steps to Create a Pie Chart in Tableau:

1. **Set Chart Type**: Select "Pie" in the **Marks** dropdown.
2. **Add Data**: Drag your category variable to **Color** (e.g., candy hardness) and your measure variable to **Size** (e.g., count of candy types).
3. **Filter**: Add only relevant measures (e.g., count of "hard" and "soft" candy).
4. **Label**: Add category names and percentages/values to each slice for clarity.

Use Case: Pie charts work best with 2-3 distinct categories that represent clear parts of a whole. They are less effective with many similar-sized segments, as subtle differences in slice sizes can be hard to distinguish.

Tip: Avoid using pie charts for more than 3 categories or where values are close in size, as they can be hard to interpret.

3. Treemaps 🌳

Definition: Treemaps display hierarchical data using nested rectangles, where each box's size represents its proportion of the total. They are ideal for showcasing data with many categories within a limited space.

Steps to Create a Treemap in Tableau:

1. **Select Data**: Choose a category (e.g., candy manufacturer) and drag it to the **Rows.**
2. **Add Measure**: Drag a measure variable (e.g., candy count) to the **Columns** to show size.
3. **Set Chart Type**: Use **Show Me** to select the Treemap chart type.
4. **Color Adjustments**: Use **stepped color** to differentiate between segments, ideally with 3-5 color steps.
5. **Label**: Include category names and counts on boxes to add clarity.

Use Case: Treemaps work well for displaying large datasets with many categories (e.g., candy types by manufacturer) when screen space is limited.

Tip: Stepped color helps differentiate categories, and treemaps are useful for visualizing the entire data hierarchy at a glance.

Publishing with Tableau Public 🌐

Since Tableau Public is a free, web-based tool, your work must be published online rather than saved locally. Here's how to save your visualizations:

1. **Go to File** > **Save to Tableau Public**.
2. **Log In**: Sign in with your Tableau Public account.
3. **Name & Publish**: Give your project a descriptive name, and it will be accessible with a public link.

Summary of Key Composition Chart Insights

- **Bar charts** work best for comparing many categories, especially when sorted by value.
- **Pie charts** are limited to a few categories and best when proportions are distinct.
- **Treemaps** offer a compact way to display many categories but may sacrifice some clarity for larger datasets.
- **Color & Labels**: Always use labels and legends to make charts understandable without additional context.
- **Publication on Tableau Public**: Tableau Public is a valuable platform for sharing work, particularly for building an online portfolio.

Quiz: 2.3: Composition & Comparison Charts

1. Where in Tableau can the analyst change the appearance of a chart, such as the size, color, and text?
    a. Marks
        i. In Tableau, the marks card is where an analyst can alter the size, color, text, detail, and tooltips for their charts. Using the pulldown menu within the marks card, the analyst can quickly change the chart type, for example, switching between a pie chart and a column chart.

2. Distribution, histogram, and correlation all fall under what type of chart?
    a. Statistical
        i. Distribution, histogram, and correlation are all types of statistical charts in that they plot some statistical aspect of the data such as frequency distributions.

3. Bar charts are useful for displaying different categories of data. When there are too many categories to display, however, a bar chart becomes unreadable. Which of the following charts would be a suitable alternative?
    a. Treemap chart
        i. Treemaps are good for displaying data sets with lots of categories. When there are too many categories, data in a column, bar, or pie chart can be hard to read, while histograms typically display only a single variable or category. Treemaps, on the other hand, make it easier to display many categories in a smaller area.

4. What makes temporal charts different from other chart types?
    a. Temporal charts contain a time component
        i. Temporal charts differ from other types of charts in that they contain a time component. For example, a temporal chart may plot world population by year or company sales figures by day.

5. Why is it important to use only a small number of categories in a pie chart?
    a. It's hard to decipher the relative size of each slice in pie charts with many categories
        i. It's important to keep pie charts to a small number of categories because the circular shape of the chart makes it hard to decipher the relative size of each slice. If there are too many categories in a pie chart, it's hard to see which slice corresponds to which group. As a general rule, more than three categories are too many for a pie chart.

2.4: Temporal Visualizations & Forecasting

Learning Goals

- Create visualizations that show data over time 📊.
- Discuss use cases for forecasting 🔮.

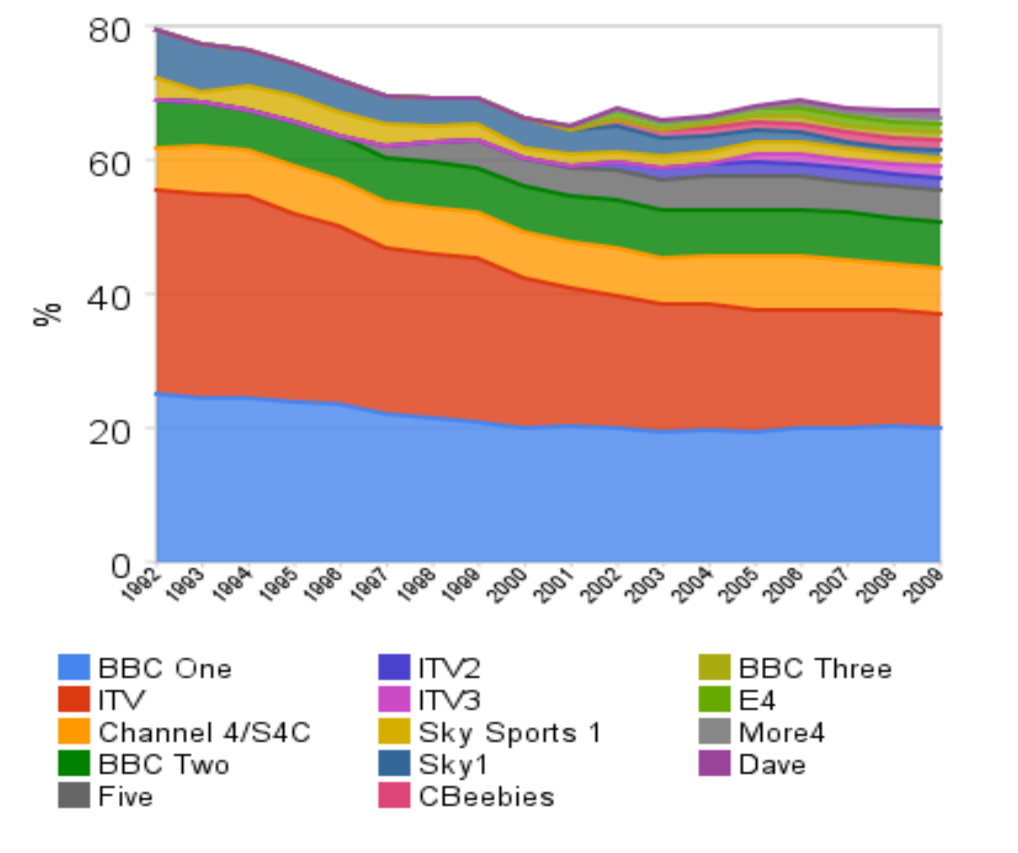
Temporal Charts

Definition

- **Temporal charts** focus on visualizing **trends across time** 🕒.
- Common use cases include tracking historical data and forecasting future values.

Chart Types for Temporal Data

1. **Bar Charts** 📊:
   - Useful for **few time points** and many categories.
2. **Line Charts** 📈:
   - Ideal for **many time points**; shows trends clearly.
3. **Area Charts** 🟩:
   - Similar to line charts but **fills the area beneath the line** for added emphasis.



Common Adjustments for Temporal Charts

- **Adjust Chart Range** 🔍:
  - Break data into smaller intervals for clarity.
- **Adjust Interval Duration (Downsampling)** ⏳:
  - Reduce the frequency of data points (e.g., from every minute to every hour).
- **Aggregate Data** 🧮:
  - Summarize by calculating totals or averages over specific intervals.

Line Charts in Tableau

Steps to Create a Line Chart

1. **Prepare Data** 📁:
    - Connect to your dataset in Tableau.
    - Change temporal data (e.g., Time) to **Date** type.
    - Rename ambiguous fields for clarity (e.g., "Time" → "Year").
2. **Filter Data** 🔧:
    - Drag filtering variables (e.g., Location) to the Filters card.
    - Select desired categories (e.g., "DEU" for Germany).
3. **Build the Chart** ✍️:
    - Add the **time variable** (e.g., Year) to the **Columns shelf.**
    - Add the **value variable** (e.g., Graduation Rate) to the **Rows shelf.**
    - Add **categorical variables** (e.g., Subject) to **Color** or **Detail.**
4. **Adjust and Refine** 🛠️:
    - Relabel ambiguous variables using **Aliases.**
    - Group related categories for clearer insights.
    - Apply meaningful colors to emphasize natural groupings (e.g., education levels).
    - Highlight key points (e.g., min/max) using **Show Mark Labels.**
5. **Format Axes and Data** 🧾:
    - Edit axis titles for clarity (e.g., "Graduation Rate").
    - Adjust the number format (e.g., display percentages).

Key Insights

- Temporal line charts effectively **visualize trends and comparisons** over time.
- Filtering and grouping data make charts more intuitive and easier to interpret.

Forecasting

Definition

- **Forecasting** predicts future trends using historical data 🔮.

Key Methods of Forecasting

1. **Linear Extrapolation** ✏️:
    - Extends a trend line into the future.
    - Best for data with **low variability.**
2. **Averaging** ⚖️:
    - Uses historical averages for future predictions.
    - **Weighted averages** emphasize recent data more heavily.

3. **Exponential Smoothing** 🚀:
    - Combines linear extrapolation and averaging.
    - **Weights recent data exponentially** higher for accuracy.
4. **Seasonality** ❄️:
    - Recognizes repeating patterns in data (e.g., flu season in winter).
    - Accounts for **frequencies** (e.g., annual, monthly).

Forecasting in Tableau

Steps to Create a Forecast

1. **Prepare Data** 📁:
    - Set up your line chart as before.
2. **Enable Forecasting** 🔧:
    - Go to **Analysis → Forecast → Show Forecast**.
    - Tableau generates future data points with prediction intervals.
3. **Adjust Forecast Options** ⚙️:
    - Set forecast length (e.g., 1 year, 2 years).
    - Choose model type (e.g., **automatic with exponential smoothing**).
    - Enable/disable seasonality based on data characteristics.
4. **Interpret Results** 🕵️:
    - **Shaded prediction intervals** represent confidence ranges.
    - Use forecasts to make informed decisions.

Key Takeaways

- Temporal visualizations reveal **trends and patterns over time**.
- Line charts are versatile for time-based data, especially with many data points.
- Forecasting aids decision-making by predicting future trends, with **exponential smoothing** being a reliable method.

Quiz: 2.4: Temporal Visualizations & Forecasting

1. In temporal charts, what type of data is plotted along the x-axis?
    a. Time or date
        i. Temporal charts display data over time and, accordingly, the time or date element is typically displayed on the x-axis of a temporal chart. This could be any measure of time based on what's in the data, such as seconds, minutes, hours, days, weeks, months, years, decades, or even centuries.

2. Similar to weighted averages, exponential smoothing puts a higher emphasis on specific data points than others. In exponential smoothing, which data points receive the higher emphasis?
    a. The most recent data points
        i. Exponential smoothing puts a higher emphasis on the most recent data points by using a scaling factor to exponentially weight historical data points. This follows the theory that data changes often and that more recent data is usually more relevant to the current and future years.

3. What's the term for the process of forecasting by drawing a line through existing data points and following this line out to a future point in time?
    a. Linear extrapolation
        i. Linear extrapolation is the process of forecasting by drawing a trend line through existing data points and following this line out to a future time. Linear extrapolation is best suited to data that has limited variability, where many of the data points intersect the trend line.

4. To create a temporal chart in Tableau, at least one time-oriented variable is required. What's the process for changing a variable, such as a column that lists the year variable, into a time-oriented dimension in Tableau?
    a. Change the data type to date
        i. Tableau treats time-oriented variables as special data types, noted by the calendar icon next to the variable name. To change a variable to a time-oriented variable, one must change the data type to "date." This is done by right-clicking the variable and selecting "date" from the change data type menu.

5. What's the purpose of forecasting?
    a. To predict a future event
        i. Forecasting is the process of predicting a future event. Historical events are used to predict what's likely to happen at a future point in time. The goal of forecasting is to predict how a variable will perform in the future.

2.5: Statistical Visualizations: Histograms & Box Plots

💡 **Learning Goals:**

- Learn how to **visualize distributions** using histograms.
- Understand how to use box plots for **comparing data distributions.**
- Discuss use cases for visualizing **statistical insights** effectively.

🧠 Introduction

Statistical visualizations quantify data insights, showing relationships, differences, and patterns. They help analysts provide concrete evidence rather than assumptions about groups or variables.

Visualization Types Covered:

1. **Histograms**: Analyze frequencies and distributions.
2. **Box Plots**: Summarize descriptive statistics.

📊 Histograms

Histograms are a type of bar chart that visualize **frequency distributions** of numeric data. Unlike bar charts, histograms group data into **bins.**

Key Characteristics

- **Bins**: Ranges along the x-axis (e.g., 1–10, 11–20).
- **Frequency**: Counts on the y-axis.
- **Adjacency**: Bins are contiguous; no gaps allowed.
- **Sorting**: Bins cannot be reordered—bins must follow their natural numeric sequence.

Benefits

- Helps identify **data patterns** (e.g., peaks or valleys in distribution).
- Useful for **business decisions**: Staffing, trends, and performance analysis.

📈 Example: Museum Ticket Sales

Suppose you analyze ticket sales at a museum:

- **Histogram Insight**:
    - Higher ticket sales in the first 50 minutes → Add staff during this peak.
    - Consistent sales between 51–100 minutes → Maintain baseline staffing.

🚀 Creating Histograms in Tableau

1. **Set up the Data**
    - Drag a **quantitative variable** to the **Rows shelf.**
    - Open **Show Me menu** and select "Histogram."
2. **Adjust Bin Sizes**
    - Tableau auto-calculates bins but allows customization.
    - **Tips for bin size**:
        - Choose intuitive bin sizes (e.g., integers like 4).
        - Avoid fractions or overly granular values like 3.13.
3. **Align the X-Axis**
    - Use **Edit Axis** to fix tick marks.
    - Ensure tick intervals match bin size.
4. **Enhance with Colors**
    - Drag a **categorical variable** (e.g., age groups) to **Color.**
    - Adjust colors for readability and intuitive groupings.
5. **Use Percentages**
    - Convert frequencies to percentages via:
        - **Quick Table Calculation → Percent of Total.**
        - Adjust the calculation to compute percentages **within bins.**
6. **Finalize Labels**
    - Add labels to display bin percentages or frequencies.
    - Format percentages to have 0 decimal places.

Key Insights from Histograms

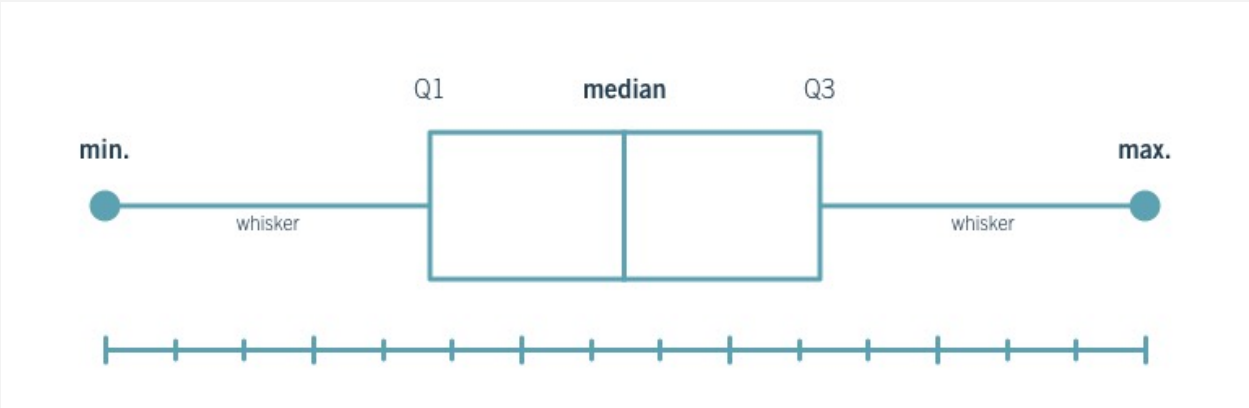
- Identify **skews** (e.g., higher frequencies at lower or higher values).
- Compare **ranges** of values for performance evaluations.
- Recognize **trends** in specific periods or groups.

📦 Box Plots

Box plots visualize **summary statistics,** providing a concise view of data spread, median, and quartiles.

Key Components

1. **Box**: Represents the middle 50% of data (Q1 to Q3).
2. **Median Line**: The middle value dividing the box.
3. **Whiskers**: Extend to minimum and maximum values (or quartile ranges).
4. **Outliers**: Shown as dots beyond whiskers.



📈 Example: Comparing Sales Regions

Box plots can reveal differences in sales performance:

- Region A: Small spread, high median → Few but high-value deals.
- Region B: Large spread, low median → Many low-value deals.

Benefits

- **Comparisons**: Ideal for analyzing distributions across groups.
- **Summary Statistics**: Quickly identify medians, ranges, and skews.
- **Outliers**: Easily spot data points outside expected ranges.

🚀 Creating Box Plots in Tableau

1. **Prepare the Data**
   - Drag the **quantitative variable** to **Rows shelf.**
   - **Uncheck Aggregate Measures** in the **Analysis menu.**
2. **Generate the Plot**
   - Open **Show Me menu** and select "Box-and-Whisker Plot."
3. **Adjust Whiskers**
   - Right-click the y-axis → Edit Reference Line.
   - Set whiskers to extend to:
     - **Maximum extent**: Include all data points.
     - **Quartile ranges**: Highlight outliers.
4. **Group Comparisons**
   - Drag a **categorical variable** (e.g., age) to the **Columns shelf** to create multiple box plots.
5. **Refine the Display**
   - Remove nulls by filtering.
   - Adjust titles, axis ranges, and labels for clarity.

🔍 Comparing Histograms & Box Plots

Feature	Histogram	Box Plot

Displays	Frequency/distribution of values.	Summary statistics (median, quartiles, etc.).
Ideal For	Finding patterns in quantitative data.	Comparing data spread and skew across groups.
Audience	More intuitive; ideal for most audiences.	More technical; better for data-savvy viewers.
Key Insight	Peaks, trends, or skews in data.	Spread, skew, and outliers in data.

📝 Summary

1. Histograms

- Visualize frequency distributions and trends in quantitative data.
- Best for **identifying patterns** or **time-based peaks**.

2. Box Plots

- Summarize descriptive statistics in a compact visual.
- Ideal for **comparing distributions** across groups or categories.

📋 Formula/Definition Chart

Feature	Histogram
Bin	A range of values in a histogram (e.g., 1–10).
Frequency	Count of data points within a bin (histogram) or category (box plot).
Median	Middle value of a data set (50% above, 50% below).
Quartile (Q1, Q3)	Divisions of data into four equal parts.
Whiskers	Lines extending from a box plot's box to show min/max values or quartile ranges.
Outliers	Data points that fall outside the expected range in a box plot.
Distribution	Spread of data across bins (histogram) or quartiles (box plot).
Skewness	A measure of asymmetry in data distribution.

📚 Quiz: 2.5: Statistical Visualizations: Histograms & Box Plots

1. What is the term for the area in a box plot between the minimum and maximum, also known as the whiskers?
    a. Answer: Range
        i. The **range** is the **difference between the minimum and maximum** values in a data set.
            - Formula: **Range = Maximum - Minimum**
            - In box plots, the whiskers extend to the minimum and maximum, showing this range.

2. What is the term for a chart with one axis displaying frequency and the other axis displaying range?
    a. Answer: Histogram
        i. **Histograms** are like bar charts but for numeric data:
            - **X-axis**: Bins (ranges of numbers, e.g., 1–10, 11–20).
            - **Y-axis**: Frequency (how many numbers fall into each bin).
            - Example: Visualizing how often grades fall within certain ranges (90–100, 80–89, etc.).

3. Evandro wants to compare the yardage of completed passes of two American football quarterbacks. He's curious to see the median yards per throw in addition to the range of yards per throw for each. Which type of chart should Evandro use?
    a. Answer: Box Plot
        i. **Why?** Box plots display:
            - **Median**: The middle value.
            - **Range**: The spread of data from minimum to maximum.
            - **Quartiles**: Show how data is divided into four equal parts.

4. What statistical measure can a box plot show that a histogram can't?
    a. Answer: Median
        i. Box plots show the **median, quartiles,** and outliers, while histograms show **frequency distributions.**
            - **Box Plot**: Good for summary statistics.
            - **Histogram**: Good for visualizing how values are distributed.

5. Which of the following is true of bins in a histogram?
    a. Bins must be adjacent and not skip any numbers

i. **Bins** in histograms represent continuous ranges of data:
   - **Adjacent**: No gaps allowed.
   - **Equal size**: All bins must cover the same range of values.
   - **Sorted numerically**: Can't reorder bins by frequency.

📝 Key Study Points

Concept	Explanation
Range (Box Plot)	Difference between maximum and minimum values (Max - Min).
Histogram	A chart where the **x-axis = range (bins)** and **y-axis = frequency** of values in each bin.
Box Plot Features	Shows **median**, **range**, and **quartiles** (not shown in histograms).
Bins (Histogram)	Adjacent, equal-sized groups of values, sorted numerically.
Median (Box Plot)	The middle value of a dataset (visible only in box plots).

2.6: Statistical Visualizations: Scatterplots & Bubble Charts

💡 Learning Goals:

- Understand **use cases** for scatterplots and bubble charts.
- Create visualizations to **demonstrate correlations** between variables.

🧠 Introduction

Scatterplots and bubble charts allow analysts to explore **relationships between variables**:

1. **Scatterplots**: Visualize the correlation between two variables.
2. **Bubble Charts**: Add additional dimensions like size and color to represent more variables.

These charts are ideal for identifying **trends, clusters, or outliers** in data.

📈 Scatterplots

What are Scatterplots?

Scatterplots graph data points along two axes:

- **X-axis**: One variable.
- **Y-axis**: Another variable.
- Each point represents an observation.

Use Case:
Visualize **correlation** (relationship) between two numeric variables.

- Example: Height vs. Weight → Taller people tend to weigh more.

Correlation in Scatterplots

Correlation Coefficient (r): Measures the strength and direction of a relationship:

- **Range**: -1 to +1.
  - **+1**: Perfect positive relationship.
  - **-1**: Perfect negative relationship.
  - **0**: No relationship.

Trend	Description
Positive Correlation	As one variable increases, so does the other.
Negative Correlation	As one variable increases, the other decreases.
No correlation	No discernible pattern between variables.

Example:

- Height & Weight: Positive correlation (taller → heavier).
- Age & Flexibility: Negative correlation (older → less flexible).

Strength of Correlation:

- Data points closer to the trend line = Stronger correlation.
- Scattered points = Weaker correlation.

Scatterplot Insights

- **Trend Lines**: Help visualize the relationship.

    - **Upward slope**: Positive relationship.
    - **Downward slope**: Negative relationship.
- **Outliers**: Points far from others can distort the trend line.

🚀 Creating Scatterplots in Tableau

1. **Set Up the Chart**:
    - Drag two numeric variables to the **Rows** and **Columns** shelves.
    - Deselect **Aggregate Measures** (in Analysis) to show individual points.
2. **Add a Trend Line**:
    - Switch to the **Analytics tab** and drag **Trend Line** to the chart.
3. **Interpret the Correlation**:
    - Hover over the trend line for **r-squared** (correlation strength).
    - Take the square root of **r-squared** to interpret as a correlation coefficient.
4. **Style the Chart**:
    - Change point colors and shapes for clarity.
    - Title: Add a descriptive name (e.g., "Correlation Between X and Y").

🔵 Bubble Charts

What are Bubble Charts?

Bubble charts are **enhanced scatterplots**:

- Add a **third variable** using **bubble size**.
- Optionally, add a **fourth variable** using **bubble color**.

Use Case:
Show relationships between multiple variables simultaneously.

- Example: Literacy Rates vs. Schooling Years, with bubble size = Income.

Bubble Chart Features

Dimension	How It's Represented
X-axis	First variable.

Y-axis	Second variable.
Size	Bubble size = Third variable (e.g., income, percentage).
Color	Bubble color = Fourth variable (e.g., categories like region).

🚀 Creating Bubble Charts in Tableau

1. **Duplicate Scatterplot**:
   - Right-click the scatterplot sheet → Duplicate → Rename (e.g., "Bubble Chart").
2. **Add Size Variable**:
   - Drag a numeric variable (e.g., Winpercent) to the **Size card.**
   - Adjust bubble sizes for readability.
3. **Add Color Variable**:
   - Drag a variable (e.g., Manufacturer) to the **Color card.**
4. **Filter & Refine**:
   - Add filters to show only top categories or relevant data.
   - Title: Add a descriptive name (e.g., "Relationship Between Sugar, Price, and Popularity").

Bubble Chart Insights

- Look for **clusters:** Do large bubbles appear in one part of the chart?
- Interpret relationships:
  - Example: If all large bubbles are in the top-right, higher sugar and price correlate with higher popularity.
- Detect randomness: No clear patterns mean no strong relationships between variables.

🔍 Comparing Scatterplots & Bubble Charts

Feature	Scatterplot	Bubble Chart
Purpose	Show the relationship between 2 variables.	Show the relationship between 3–4 variables.
X- & Y-Axes	Two numeric variables.	Two numeric variables.
Additional Detail	None.	Size (3rd variable), Color (4th variable).
Example Use	Correlation of Height vs. Weight.	Sugar, Price, and Popularity of Candy.

📝 Summary

1. **Scatterplots**:
    - Visualize **correlations** between two variables.
    - Use **trend lines** and **r-squared values** for interpretation.
    - Great for spotting **outliers.**
2. **Bubble Charts**:
    - Add depth to scatterplots by visualizing 3–4 variables.
    - Use **size** and **color** for added dimensions.
    - Ideal for detecting **clusters** and relationships in complex data.

📋 Formula/Definition Chart

Term	Meaning
Scatterplot	Chart for visualizing the relationship between two numeric variables.
Trend Line	A line of best fit showing the general direction of a relationship in a scatterplot.
Correlation Coefficient (r)	Measures the strength and direction of the relationship between two variables (-1 to +1).
Bubble Chart	A scatterplot with bubble size representing a third variable and color a fourth variable.
r-squared	The square of the correlation coefficient; shows the proportion of variance explained.
Outlier	A data point that is far removed from other data points and can distort trends.

📚 Quiz: 2.6: Statistical Visualizations: Scatterplots & Bubble Charts

1. Statistical Visualizations: Scatterplots & Bubble Charts
    a. Dragging a variable to the Size box on your Marks card
        i. To transform a scatterplot into a bubble chart, a new variable needs to be dragged to the Size box on your Marks card. This tells Tableau to use that new variable to change the size of each data point in the scatterplot, creating a bubble chart.

2. Which of the following is true of variables on a bubble chart?
    a. Size and color variables can be qualitative
        i. Size and color variables can be either quantitative or qualitative on a bubble chart. Variables on the x- and y-axes, however, can only be quantitative because they require number values to plot each data point as a bubble.

3. Bubble charts differ from scatterplots in that they display three variables instead of two. How does a bubble chart display this third variable?
    a. Through the size of each data point
        i. A bubble chart uses the size of each data point to indicate a third variable. For instance, a bubble chart may plot age using larger bubbles to represent older ages and smaller bubbles to represent younger ages.

4. What variables are required when creating scatterplots or bubble charts in Tableau?
    a. Two or more measures
        i. Scatterplots and bubble charts require at least two measures. These two quantitative measures form the x- and y-axes of the chart in which the data points are plotted.

5. What can be learned by examining how close data points are to a trendline in a scatterplot?
    a. If a correlation is strong
        i. In a scatterplot, the closeness of the data points to a trend line indicates the strength of the correlation between the two variables involved. Plots with data points on or very near the trendline have stronger correlations, while plots with data points far from the trendline have weaker correlations.

2.7: Spatial Analysis

💡 Key Learning Goals

- Understand the **definition and purpose** of spatial analysis.
- Learn about **different types of maps** used for spatial visualizations.
- Explore **mapping tools and capabilities in Tableau.**
- Practice creating various **map visualizations** to analyze and communicate spatial data effectively.

📌 Introduction to Spatial Analysis

Spatial analysis involves the inclusion of a **geographic component** to study data trends, patterns, or relationships. This type of analysis is crucial for understanding how location impacts data insights.

Examples of spatial questions:

- **Health Trends**: Are flu shots more common in certain states?
- **Sales Analysis**: Which cities contribute the most revenue?
- **Population Density**: Where do the most flu-related deaths occur?

Why Use Spatial Analysis?

Spatial visualizations are powerful because they:

- Highlight **geographic distributions** that tables and charts may miss.
- Reveal **regional trends** or **hotspots** (e.g., high-flu areas).
- Simplify complex datasets into **easy-to-interpret maps**.

🗺️ Types of Spatial Maps

Map Type	Description	Best Use Case
Point Map	Displays exact points using latitude/longitude coordinates.	Visualizing precise locations (e.g., flu clinics, heritage sites).
Heat Map	Uses colors to represent the density of points in an area.	Highlighting dense clusters (e.g., flu hotspots).
Choropleth Map	Shades regions (e.g., countries, states) based on a numerical variable.	Comparing aggregated statistics across regions (e.g., vaccination rates by state).
Graduated Symbol Map	Represents data counts using symbol sizes (e.g., circles) placed on geographic locations.	Showing variable magnitudes at specific locations (e.g., number of clinics per city).
Combination Map	Merges multiple types of maps to show multiple variables simultaneously.	Displaying multiple relationships (e.g., flu shot rates with vaccination infrastructure levels).

✨ Creating Maps in Tableau:

Step-by-Step Overview for Each Map Type

Point Maps: Display Exact Locations

1. **Load Spatial Data**:
    - Ensure your dataset includes **latitude and longitude coordinates** or geographic boundaries (e.g., city or state names).
2. **Create the Map**:
    - Drag **Latitude** to the Rows shelf and **Longitude** to the Columns shelf.
    - **Turn off Aggregation**: Open the **Analysis** menu and deselect **Aggregate Measures**.
3. **Add Context**:

    - Use color to categorize data points (e.g., heritage site types: cultural, natural, mixed).
4. **Customize Appearance**:
    - Go to **Map Layers** to adjust background, coastline visibility, and map style (light or dark).
    - Add a descriptive title (e.g., “Heritage Sites by Type”).

Heat Maps: Highlight Dense Clusters

1. **Start with a Point Map**:
    - Duplicate your existing point map sheet for consistency.
2. **Change the Visualization Type**:
    - Open the **Marks** card dropdown and select **Density**.
3. **Adjust Density Colors**:
    - Use intuitive heat map color schemes like **green (low)** to **red (high density).**
    - Remove unnecessary background layers (e.g., ocean, land cover).
4. **Zoom for Detail**:
    - Use the **Zoom Tool** (magnifying glass) to focus on regions with dense data.

Choropleth Maps: Compare Aggregated Data Across Regions

1. **Start with Geographic Boundaries**:
    - Use a **state/country field** to define spatial boundaries.
    - Ensure the **Geographic Role** is correctly assigned (e.g., Country/Region).
2. **Aggregate Data**:
    - Drag the **Measure** you want to analyze (e.g., flu shot counts) onto the **Color** box.
3. **Simplify the Color Scheme**:
    - Use **Stepped Colors** (5 shades maximum) to enhance clarity.
    - Stick to **monochromatic palettes** to avoid confusion.
4. **Finalize Titles and Labels**:
    - Rename the color legend (e.g., “Flu Shot Rates”).
    - Add a clear title, like “Flu Shot Rates by State.”

Graduated Symbol Maps: Use Symbol Size to Represent Data

1. **Start with a Choropleth Map**:
    - Duplicate your choropleth map for consistency.
2. **Change to a Symbol Map**:
    - Open **Show Me** and select the **Symbol Map** type.

3. **Assign Size to Variable**:
    - Drag a **Measure** (e.g., number of clinics) onto the **Size** box.
    - Adjust size ranges (4–5 categories max).
4. **Optional Enhancements**:
    - Use filled symbols (e.g., circles) for better visibility.
    - Add labels if necessary (e.g., number counts near symbols).

Combination Maps: Layer Multiple Map Types

1. **Create Base Maps**:
    - Duplicate the choropleth map as your starting point.
    - Add a second map (e.g., graduated symbol) by duplicating the **Latitude** field.
2. **Customize Both Layers**:
    - Assign color to one map (e.g., flu shot rates).
    - Assign size to the second map (e.g., healthcare access per region).
3. **Combine Maps**:
    - Select **Dual Axis** on the second **Latitude** variable to merge the layers.
4. **Refine Final Map**:
    - Adjust titles, legends, and colors for clarity.

🧭 Real-World Examples of Spatial Analysis

Use Case	Map Type	Insights Gained
Flu Shot Rates	Choropleth Map	Identify states with low vaccination coverage for targeted campaigns.
Flu Hotspots	Heat Map	Detect areas with high flu-related deaths for emergency resource allocation.
Clinic Availability	Graduated Symbol Map	Display the number of healthcare facilities in different cities for accessibility analysis.
Combined Insights	Combination Map	Show flu shot rates with healthcare spending per state to evaluate correlation.

💡 Tips for Effective Spatial Visualization

- Use **contrasting colors** to enhance readability (e.g., light backgrounds with bold colors).
- Keep color schemes **intuitive** (e.g., red for high values, blue for low).

- Avoid overloading maps with too much data—**focus on a single narrative** per map.
- Add **clear titles and legends** for easier interpretation.

📝 Summary

Spatial analysis provides invaluable insights into location-specific trends. From understanding vaccination rates to visualizing flu hotspots, maps bring geographic data to life. Tableau offers a wide range of spatial visualization tools, making it easy to create point maps, heat maps, choropleth maps, graduated symbol maps, and combination maps.

📚 Quiz: 2.7: Spatial Analysis

1. A combination map is a choropleth map combined with which other type of map?
    a. Graduated symbol map
        i. A combination map is a combination of a choropleth map and a graduated symbol map. A combination map includes shaded spatial areas, like a choropleth map, along with another variable represented through differently sized or colored symbols.
2. How can you change a point map into a heat map in Tableau?
    a. Selecting Density from the dropdown menu on your Marks card
        i. To change a point map into a heat map in Tableau, select Density from the dropdown menu on your Marks card. This tells Tableau to convert the individual points into a density aggregation to display differences in density of points by color.

3. What's required to create a map in Tableau?
    a. One geocoded variable
        i. One geocoded variable is required to create a map in Tableau. This variable could be country names or geocoded longitude and latitude coordinates. Tableau can't convert addresses or other non-geocoded variables to data to create maps.

4. Point maps detail where events occur. What's used to map these locations?
    a. Latitude and longitude
        i. Point maps use latitude and longitude coordinates to map where events occur. A dot is placed at each set of

coordinates, allowing trends to be identified between the points.

5. Both choropleth maps and graduated symbol maps show a relationship between a geographic variable and a quantitative data element. What makes them different?
    a. Graduated symbol maps use markers to denote a quantitative data element
        i. Graduated symbol maps use symbols or markers to denote a quantitative data element, while choropleth maps use colors to denote a quantitative data element.

2.8: Textual Analysis

💡 Key Learning Goals

- Understand the **use cases for textual analysis** in data analytics.
- Learn how to analyze qualitative data and visualize it using **word clouds** and **packed bubble charts.**
- Explore the basics of **sentiment analysis** and its challenges.

📌 Introduction to Textual Analysis

Unlike quantitative data, textual analysis deals with **qualitative data**—information expressed as text rather than numbers. Examples include:

- **Survey responses**
- **Social media posts**
- **Customer reviews**
- **Emails**

Why Use Textual Analysis?

Textual analysis helps to:

- Identify key themes and patterns in unstructured data.
- Gain insights into user opinions and trends.
- Support further investigation into "why" and "how" questions.

Textual data is typically **unstructured** and harder to analyze because it allows users to express themselves freely, leading to variability in responses. For instance, social media hashtags and survey responses may include typos, synonyms, or phrases that lack consistency.

✨ Key Categories of Textual Analysis

Category	Definition	Visualization Type
Frequency Analysis	Counting how often specific words or phrases appear.	Word Clouds, Packed Bubble Charts
Sentiment Analysis	Categorizing text into **negative**, **neutral**, or **positive** sentiments to understand emotional context.	Specialized Sentiment Tools

🗒️ Frequency Analysis

Frequency analysis identifies trends by **counting occurrences** of words or phrases in text data. It is a foundational step for organizing textual data and detecting patterns.

Challenges in Frequency Analysis

- Text data often contains **unstructured formats** (e.g., free responses, hashtags).
- Typos, variations, and synonyms can lead to inflated or misleading counts.
- Single-word analysis may miss **context** (e.g., "not happy" could be read as "happy").

Frequency Visualization Options

- **Word Clouds**: Highlight word frequency with font size.
- **Packed Bubble Charts**: Combine word frequency with an added focus on precision using bubble sizes.

🗺️ Word Clouds: Visualizing Word Frequency

What is a Word Cloud?

A word cloud represents the **frequency of words** by varying font sizes. Larger words signify higher frequencies.

Strengths

- Easy to spot major themes at a glance.

- Ideal for presentations to highlight broad trends.

Weaknesses

- Single-word focus can distort meaning (e.g., “not happy” split into “not” and “happy”).
- Common filler words (e.g., “the,” “all”) may appear unless filtered.
- Lack of precise comparison between word frequencies.

Example

A word cloud of a politician’s social media feed may show larger words like "Virginia" or "coastal" based on frequency.

🔘 Packed Bubble Charts: A Word Cloud Alternative

Packed bubble charts use **bubbles** to represent word frequency, offering more clarity than font sizes.

- **Bubble Size**: Reflects frequency.
- **Bubble Color**: Adds a second variable for better differentiation (e.g., category or industry).
- **Top N Filtering**: Reduces clutter by displaying only the most frequent terms.

👩‍💻 Creating Word Clouds and Packed Bubble Charts in Tableau

Steps to Create a Word Cloud

1. **Load Data**: Open Tableau and connect your text data file.
2. **Set Up the Visualization**:
    - Drag the text variable (e.g., Job Title) to the **Text** box on the Marks card.
    - Drag the same variable to the **Size** box.
3. **Adjust Aggregation**:
    - Use the **COUNT** aggregation to size words by frequency.
4. **Filter Low Frequencies**:
    - Create a calculated field to exclude infrequent words.

Steps to Create a Packed Bubble Chart

1. **Duplicate Word Cloud**: Start from the Word Cloud sheet.
2. **Switch to Bubble Chart**: Use the **Show Me** menu to select **Packed Bubbles.**
3. **Add Labels**: Copy the COUNT variable to the **Label** box on the Marks card.

4. **Filter Top N Words**: Use the **Filter** card to display only the most frequent words.

🔍 Sentiment Analysis

What is Sentiment Analysis?

Sentiment analysis categorizes text into **positive**, **neutral**, or **negative** sentiments. It is widely used for:

- **Social Media Monitoring**: Understand user emotions about brands or events.
- **Customer Feedback**: Evaluate product/service satisfaction.

Challenges of Sentiment Analysis

- **Context Dependency**:
  - The word “bad” can have different meanings:
    - “That Starbucks is bad” (negative).
    - “I’m craving Starbucks so bad” (positive).
- **Language Complexity**:
  - Sarcasm and nuanced expressions make accurate classification difficult.
- **Algorithms Required**:
  - Advanced techniques, such as machine learning, are often needed to improve accuracy.

📌 Real-World Use Cases

Scenario	Textual Analysis	Example Insight
Survey Data	Word Cloud	Highlight the most common user suggestions for improvement.
Social Media Monitoring	Sentiment Analysis	Detect overall brand sentiment and isolate negative comments for action.
Job Market Analysis	Packed Bubble Chart	Visualize the top industries or job titles in a geographic region.

💡 Tips for Effective Textual Visualizations

- **Filter out irrelevant words**: Use stopwords to exclude common, meaningless terms (e.g., "the," "and").
- Use **color** sparingly to avoid clutter.
- Add **thresholds** (e.g., show top N values) to simplify visuals.

- Pair textual analysis with **quantitative metrics** for deeper insights (e.g., sentiment trends over time).

📝 Summary

Textual analysis is a valuable tool for analyzing qualitative data. While **word clouds** and **packed bubble charts** are excellent starting points for identifying broad trends, more advanced techniques, like **sentiment analysis,** require specialized tools.

📚 Quiz: 2.8: Textual Analysis

1. Which of the following is an example of textual data?
    a. A social media post about a restaurant experience
        i. A social media post about a restaurant experience is an example of textual data because it contains free text rather than numbers. Textual data is qualitative and generally unstructured.

2. Nancy is analyzing country-of-origin data collected by immigration surveys. She creates a list to count the number of times each survey answer is represented. What did she create?
    a. **Frequency table**
        i. Nancy has created a frequency table. A frequency table can be used to count the frequency of each answer in a survey. It can help Nancy learn which countries are mentioned most and least often in survey answers, as well as highlight any inconsistent answers such as someone answering “DR” instead of “Dominican Republic.”

3. If Tableau creates a treemap of your data instead of a word cloud, how can you change it?
    a. **Change Automatic to Text in the dropdown menu on your Marks card**
        i. If Tableau creates a treemap of your data instead of a word cloud, change Automatic to Text in the dropdown menu on your Marks card. This tells Tableau to display the text of the variable, which, in turn, changes the treemap to a word cloud.

4. What’s the term for the type of textual analysis that involves counting specific words and phrases to identify broad trends?
    a. Frequency analysis

    i. Textual analysis that involves counting specific words and phrases to identify broad trends is called frequency analysis. The resulting list can be made into a frequency table or used to create a word cloud.

5. Which of the following is true of word clouds?
    a. Words are bigger based on their frequency
        i. Words that are used more frequently appear bigger in word clouds, while words that appear less frequently appear smaller.

2.9: Storytelling with Data Presentations

💡 Learning Goals

1. Develop a compelling **data narrative** based on analytical findings.
2. Create and structure a **data storyboard** in Tableau.
3. Present your insights using a cohesive beginning, middle, and end format.

📌 Introduction to Storytelling with Data

This section explores how to combine your analyses and visualizations into a structured **data narrative.** Your goal is to create a presentation that communicates insights effectively to stakeholders using **Tableau storyboards.**

Why Storytelling Matters in Data Analytics

- **Simplifies Complexity**: Breaks down large datasets into digestible pieces.
- **Engages Audiences**: Visuals and narratives make findings more engaging.
- **Increases Retention**: Stories combine analytical insights with emotional appeal, targeting both analytical and creative thinking.

✨ Key Elements of a Data Story

Like any story, a **data narrative** has three components:

1. **Beginning**: Introduce the purpose, project motivation, and objectives.
2. **Middle**: Showcase key analyses, focusing on insights that align with the story's objectives.

3. **End**: Present conclusions, recommendations, or next steps based on your findings.

📖 Structuring the Narrative

Section	Key Points	Example (Candy Dataset)
Beginning	Frame the question or motivation.	"Why are some candies more popular than others?"
Middle	Highlight relevant analyses	Explore price, sugar content, and manufacturer.
End	Summarize insights and next steps.	Preference seems to be based on taste, not data.

🔍 Data Story Fundamentals

Simplification

A good story simplifies complex data, focusing only on the information needed to answer the central question.

- Example: Use concise visuals to show correlations between candy attributes (e.g., sugar content vs. popularity).

Engagement

Incorporate visuals like charts, maps, and interactivity to keep your audience engaged.

- Example: Allow users to click on a specific candy to see related attributes.

Retention

Blend **visuals**, **narratives**, and **interactive elements** to create a memorable experience.

- Example: Highlight winning candies with annotations explaining their success.

🎨 Creating a Storyboard in Tableau

Key Components

1. **Sheets**: Individual visualizations (e.g., scatterplots, treemaps).
2. **Dashboards**: Combine multiple sheets with text and images.
3. **Storyboards**: Sequence dashboards into a narrative.

Step-by-Step Guide

Step	Details
1. Create Visualizations (Sheets)	Design individual charts (e.g., scatterplot of price vs. popularity).
2. Build Dashboards	Arrange charts, add descriptions, and make elements interactive.
3. Sequence Story Points	Combine dashboards into a storyboard with captions to guide the audience.
4. Add Annotations	Highlight key findings or trends directly on charts to draw attention to important details.
5. Format and Test	Ensure the storyboard is accessible and visually appealing on various devices.

🗒️ Example: Candy Dataset Storyboard

Beginning

- **Goal**: Examine why certain candies are preferred.
- **Visualization**: A ranked table of candies by winning percentage.
- **Dashboard Elements**:
    - Table with candies ranked by popularity.
    - Project overview text (e.g., "82 candies, 8,371 votes").
    - Image for visual appeal.

Middle

- **Goal**: Analyze the role of variables (price, sugar content, manufacturer) in candy preference.
- **Visualizations**:
    - Scatterplots: Price vs. popularity, sugar vs. popularity.
    - Treemap: Manufacturer vs. winning percentage.
- **Dashboard Elements**:
    - Interactive filters (e.g., click on a manufacturer to update scatterplots).
    - Titles and annotations explaining findings.

End

- **Goal**: Conclude with key takeaways and next steps.
- **Visualizations**:
    - Summary charts comparing most/least popular candies.

- Text-based recommendations (e.g., "Preference likely driven by individual taste. Future research could explore ingredient composition.").

⚡ Tips for Effective Storyboards

1. **Focus on Clarity**: Use descriptive titles, labels, and concise text.
2. **Engage the Audience**: Add interactive features (e.g., filters, highlighters).
3. **Highlight Key Insights**: Use annotations to emphasize important points.
4. **Test Accessibility**: Ensure visuals are clear and readable across devices.

Formula and Concept Chart 📘

Term/Formula	Definition/Usage
RANK_DENSE()	Tableau formula to rank items without skipping ranks for ties.
SUM()	Calculates the sum of values in a field.
Calculated Field	A custom field created for analysis, e.g., `RANK_DENSE(SUM([Winpercent]), 'desc')`.
Dashboard	Combines multiple visualizations (sheets) into a single view.
Storyboard	A sequence of dashboards used to tell a cohesive data story.
Annotations	Notes added to charts to explain specific data points.
Highlighting	Feature in Tableau to emphasize specific data points in a chart or across charts.
Filter	Allows interactivity by filtering data based on user selections.
Interactive Elements	Clickable visuals that dynamically update dashboards (e.g., sliders, highlights).

📌 Summary

By structuring your insights into a beginning, middle, and end, you create a compelling data story that resonates with stakeholders. With Tableau's storytelling features, you can combine interactivity, visuals, and narrative to create a memorable presentation.

📚 Quiz: 2.9: Storytelling with Data Presentations

1. Jeff is blending his visuals with text in his data presentation. Why is this important for Jeff to do?
    a. To increase information retention
        i. Blending visuals with text will activate both sides of his audience's brains. Activating both sides of the brain—the analytical left and creative right—creates a whole brain experience that increases information retention.

2. A data story has a beginning, middle, and end. In which stage are most of your findings presented?
    a. Middle
        i. The majority of your data analysis findings are presented in the middle of a data story. The beginning sets the stage with the project motivations and objectives, the middle is where the data analysis findings are presented, and the end is where the conclusions and next steps live.

3. Which of the following helps keep an audience engaged in a data story?
    a. Combining visuals with captivating text
        i. Using a combination of visuals and captivating text can keep an audience more engaged in a data story. Evidence for this comes from marketing research, which concludes that a good balance between text and visuals results in more engaging content for an audience.

4. Rob wants to add a few sheets and dashboards to a storyboard. What's the first step he should follow?
    a. Clicking the button that looks like an open book
        i. Rob should click the button in Tableau that looks like an open book. This will create a storyboard within Tableau that Rob can populate with his sheets, dashboards, and other visuals.

5. Which Tableau function creates a column of row numbers based on a variable, for instance, to denote the order of finishers in a race?
    a. RANK_DENSE()
        i. To create a column of row numbers based on a variable, use the `RANK_DENSE()` function after creating a calculated field from the Analysis menu.

2.10: Presenting Findings to Stakeholders

📝 Learning Goals

1. **Present a Data Story**
   Learn how to craft and deliver a compelling narrative based on your data insights.
   🗣️ Public speaking + storytelling = powerful presentations.
2. **Communicate Analytical Findings**
   Adjust your presentation to match the technical understanding of your audience.
   🎯 Clear, concise, and audience-focused communication is key.

📚 Introduction

- **Purpose of Presenting Data**
  You've cleaned, analyzed, and told your story. Now, it's time to share your insights in a way that drives decisions.
  🌟 Your job: transform complex data into actionable insights for stakeholders.
- **Why It Matters**
  A good presentation isn't just about sharing numbers—it's about **inspiring action** and **making an impact.**
  📊 Data without communication = no decisions.
- **Skills You'll Use:**
  - Storytelling 🎨
  - Visual design 🖼️
  - Public speaking 🎤

🎤 Oral Presentation Tips

1. Prepare 🖋️

- **Start with a Script:**
  - Outline key points to keep your presentation focused.
  - Make sure your script addresses stakeholder goals (e.g., business objectives, operational needs).
- **Refer to Project Documentation:**
  - Revisit business requirements and project goals to ensure your findings align with expectations.
  - Example: If your project aimed to improve staffing, highlight how your recommendations directly address staffing needs.
- **Avoid Reading Word-for-Word:**

    - Use your script as a guide but speak naturally to stay engaging.
      🎭 Example: Think of it like a conversation, not a textbook reading!

2. Practice 🗣️

- **Rehearse Out Loud:**
    - Identify areas where your narrative feels clunky or unclear.
    - Example: Record yourself to spot pauses, filler words, or fast pacing.
- **Build Confidence Through Repetition:**
    - The more you practice, the more natural and polished you'll feel.

3. Pacing ⏱️

- **Slow Down:**
    - Nervousness often leads to rushing.
    - Tip: Use a stopwatch to time key sections and ensure you're not too fast or slow.
    - Example: If you finish explaining a chart in 20 seconds but planned for a minute, slow down!
- **Use Pauses for Emphasis:**
    - Let key points sink in with strategic pauses.
    - Example: After showing a critical metric, pause for 2-3 seconds.

4. Body Language 🧍

- **Confident Posture:**
    - Stand naturally with feet shoulder-width apart.
    - Avoid slouching or fidgeting.
- **Smile & Breathe:**
    - Smiling can boost your mood and connect with your audience.
    - Deep breathing before speaking helps calm nerves.
- **Eye Contact:**
    - Look at your audience instead of fixating on slides or notes.

5. Passion ❤️

- **Let Enthusiasm Shine:**
    - Your excitement about the project can inspire your audience.
    - Example: Share a moment when a data insight surprised you.

🎯 Designing for Your Audience

1. Tailoring Communication 🧩

- **Know Your Audience:**
  - Match your language and detail to the audience's expertise.
  - Example: For a mixed audience, use simple terms and save technical jargon for follow-ups.

Audience Tiers:

Tier	What They Need	Example
High (Analysts)	Statistical methods, deeper analysis.	"At an alpha of 0.05, we found no significant difference between groups."
Mid (Managers)	Context + simplified results.	"We're 95% confident that the two groups show no meaningful difference in results."
Low (Non-Technical)	High-level overview and recommendations.	"Both groups received similar support, meaning no major difference in outcomes."

2. Focusing on Goals 🎯

- Highlight insights that **directly affect stakeholders.**
- Example:
  - **Analysts:** How data was processed.
  - **Frontline Workers:** How it impacts their work.
  - **Managers:** Strategic recommendations and next steps.

📹 Video Recording Tips

- **Quiet Environment:** Minimize noise and echo. Use a padded room if possible. 🔇
- **Lighting Matters:** Ensure even lighting; avoid sitting in front of windows. 💡
- **Recording Tools:** ScreenPal, Loom, QuickTime, iMovie, or Windows Movie Maker. 🛠️
- **Export Format:** Save as MP4 and upload to platforms like YouTube or Vimeo for easy sharing. 🌐

⚠️ Data Limitations

1. **Acknowledge Shortcomings:**
    - Bias, missing data, or limited sources.
    - Example: “This analysis didn’t account for seasonal trends due to incomplete data.”
2. **Privacy Concerns:**
    - Ensure compliance with laws (e.g., anonymizing personal data). 🔒
    - Example: In the school dropout project, share trends but not individual student details.

📈 Next Steps

1. Monitoring Metrics

- **What to Track:**
    - Quantitative data (metrics like graduation rates or staffing levels).
    - Example: A KPI might measure whether your staffing recommendations reduce burnout rates.
- **Define KPIs:**
    - Choose metrics tied to specific goals.
    - Example: Track accident rates monthly to achieve a 10% reduction in a year.

2. Gathering Feedback

- **Why Seek Critique?**
    - Fresh perspectives can reveal blind spots.
    - Example: Ask a colleague if your slides are clear and engaging.
- **Iterate and Improve:**
    - Incorporate feedback before finalizing your presentation.

🌟 Summary

- You’re ready to present your data story! 🎉
- Focus on **clarity, engagement, and stakeholder needs.**
- Remember: your narrative isn’t just about sharing data—it’s about inspiring decisions and driving change. 🚀

📋 Key Formula/Definition Chart

Term	Definition

KPI	Key Performance Indicator; tracks progress toward a specific goal. Example: Monthly sales growth. 📊
Data Limitations	Acknowledgment of data constraints, biases, or missing elements. ⚠️
Monitoring Metrics	Ongoing tracking of key variables (e.g., staffing levels, accident rates). 📈
Audience Tiers	Levels of expertise in your audience (high, mid, low) for tailoring content. 🎯

📚 Quiz: 2.10: Presenting Findings to Stakeholders

1. Dale is feeling anxious and nervous before his oral presentation. He only has a few minutes before he's supposed to speak. What can he do to calm his nerves?
    a. Take deep breaths for one minute
        i. Dale can spend one minute practicing deep, calming breaths. Similar to meditation, this act can help calm his nerves before a big presentation.

2. What's the term for a subcategory of metrics singled out as the most important metrics for an organization?
    a. Key performance indicators
        i. Key performance indicators (KPIs) are a subcategory of metrics that have been singled out as the most important metrics for an organization. KPIs track the progress toward a business objective or goal and are often used after an analysis.

3. When facing a mixed or unknown audience, what level of technical jargon and advanced statistical detail should you include during an oral presentation?
    a. **Include minimal to no technical jargon or advanced statistics**
        i. When an audience's technical level is unknown or the audience is mixed, it's best to err on the side of caution and avoid technical jargon and advanced statistics. While these concepts may be understood by analysts and data scientists, it's unlikely that other stakeholders would be able to follow along.

4. How fast should you speak when giving an oral presentation?
    a. Speak at or slower than your normal speaking rate

        i. When giving an oral presentation, speak at or slower than your normal speaking rate. You can additionally slow down or add pauses for emphasis.

5. There are two main goals when it comes practicing aloud for an oral presentation. The first is to find sections of the narrative that don't flow well. What's the second?
    a. Gaining confidence with the material
        i. The second goal of practicing for an oral presentation is to gain familiarity with the material. At this point, you should already know your audience and have vetted your data for bias. The act of practicing a speech aloud can be a final step to feeling comfortable presenting the material in front of others.