Choropleth Maps

A choropleth map is a type of thematic map that shows data through color variations on predefined geographic areas, such as countries, states, or provinces.

Usually, a color scale or gradient is employed to represent the range of values, with lighter or darker colors indicating lower or higher values.

Choropleth maps are widely used in various fields, such as geography, economics, demography, and public health, to illustrate and analyze spatial data.

They can help identify trends and concentrations of specific characteristics across different regions.

our AI-powered location intelligence platform, each shade of color represents a province in Switzerland and the population of restaurants and cafes in that province.

Key insights can immediately be interpreted and extracted from the map, even without a technical background.

Use Cases for Choropleth Maps

Creating a choropleth map

Creating a choropleth map is easy. Follow the step by step guide below.

Step 1 – Gather your data

Gather the data you need present. Next, find the range of your values and develop a shading scale. Between 4 and 8 shading bands should be appropriate. Ensure the shading bands get darker as values increase.

Step 2 – info

Source a base map of the area(s) or region(s) you are presenting data for. Include the key and title on your map.

Step 3 – info

Shade the areas/regions according to where they fit on the scale.

You can also create choropleth maps in a range of software including Google Sheets

Describing a choropleth map

TEA is a great acronym to use when describing patterns on a choropleth map. TEA stands for trend, example and anomaly.

Trend

What is the general pattern shown on the map?

- Is there an even distribution (spread)?
- Is there an uneven distribution (spread)?

Examples

Discuss the pattern on the map including examples. You could consider:

- What continents/countries are most of the feature contained in?
- Which have the least?
- Are these HICs or LICs?
- Are they near the equator or further away?
- Are they inland or coastal?

Case Study: Patient Health Outcomes Analysis

You are analyzing a dataset from a healthcare study with the following variables:

Patient ID	Age	Gender	Treatment Type	Duration of Treatment (weeks)	Health Outcome (1-5)	Medication Adherence (Yes/No)	Follow-up Visits
001	45	Female	Medication	12	4	Yes	3
002	60	Male	Therapy	16	3	No	2
003	30	Female	Medication	8	5	Yes	4
004	55	Male	Therapy	10	2	Yes	1
005	40	Female	Medication	20	4	No	3
006	50	Male	Therapy	14	3	Yes	2
007	35	Female	Medication	6	5	Yes	5
008	65	Male	Therapy	18	1	No	1
009	48	Female	Medication	10	4	Yes	3
010	52	Male	Therapy	15	2	Yes	2

Questions:

1. **Use a scatter plot matrix to represent the relationships among the following four variables: Age, Duration of Treatment, Health Outcome, and Follow-up Visits. Justify the effectiveness of this technique in illustrating the correlations and distributions among these variables. (5 Marks)**
2. **Apply a hyperbox method to analyze relationships among Treatment Type, Gender, and Medication Adherence. Discuss how these findings can enhance understanding of treatment effectiveness and patient engagement in the healthcare system. (5 Marks)**

1. Scatter Plot Matrix Analysis

Variables to Analyze:

- Age
- Duration of Treatment
- Health Outcome
- Follow-up Visits

Effectiveness of the Scatter Plot Matrix:

- **Correlation Identification**: The scatter plot matrix allows for visualization of pairwise relationships between the four variables. For example, we can explore if older patients tend to have different health outcomes or if a longer duration of treatment correlates with more follow-up visits.
- **Distribution Visualization**: Each variable's distribution can be analyzed. For instance, we might see that health outcomes are skewed towards higher scores for younger patients or those with longer treatment durations.
- **Multi-dimensional Insight**: It enables a comprehensive examination of interactions. For example, we might observe that patients with higher adherence to medication also have better health outcomes, regardless of age.
- **Outlier Detection**: Any outliers, such as a patient with very high follow-up visits but low health outcomes, can be identified, indicating the need for further investigation into their treatment or engagement.

- **Age vs. Health Outcome**: There may be a negative correlation, suggesting that older age is associated with lower health outcomes.
- **Duration of Treatment vs. Health Outcome**: A positive correlation could indicate that longer treatments lead to better health outcomes.
- **Medication Adherence vs. Health Outcome**: Patients with high adherence may show significantly better outcomes, suggesting the importance of adherence in treatment success.

2. Hyperbox Method Analysis

Analysis of Relationships:

1. **Treatment Type vs. Gender**: Hyperboxes can show how many patients of each gender received different treatment types. For instance, we might find that more females are treated with medication compared to males.
2. **Treatment Type vs. Medication Adherence**: We can analyze how adherence varies by treatment type. If therapy patients show lower adherence rates, it might indicate that therapy requires more engagement or effort than medication.
3. **Combining Insights from SPLOM**: The scatter plot insights can be overlaid on the hyperbox findings. For instance, if it's found that older patients have lower health outcomes with therapy but higher adherence with medication, this could inform treatment recommendations.

- **Treatment Effectiveness**: Understanding which treatment types lead to better outcomes can help tailor treatment plans. For instance, if medication shows higher effectiveness for younger patients, this can guide healthcare providers in making recommendations based on age.
- **Patient Engagement**: If a specific treatment type has lower adherence, healthcare providers can investigate further and develop strategies to improve patient engagement, such as more follow-up visits or educational resources.
- **Gender Dynamics**: Analyzing gender differences in treatment response and adherence can inform more personalized healthcare strategies, addressing barriers specific to different patient demographics.

You are analyzing spatial data for a coastal region to aid in resource management and environmental protection. The dataset includes various coastal features and activities:

Feature Type	Location (X, Y)	Area (sq km)	Pollution Level (1-5)	Fishing Activity (High/Medium/Low)	Protected Status (Yes/No)
Coral Reef -	(10, 20)	5	2	High	Yes
Mangrove Forest	(15, 25)	10	1	Medium	Yes
Fishing Zone	(20, 30)	15	3	High	No
Urban Development	(25, 15)	8	4	Low	No
Conservation Area	(30, 20)	12	1	Low	Yes

Questions:

1. **Propose a mapping technique for visualizing the spatial distribution of coastal features. Specify color schemes for pollution levels and protected status, and discuss how this design can enhance understanding of coastal resource management.**
2. **Discuss the role of icon based analysis in monitoring fishing activities and pollution levels. Explain how these insights can inform policy decisions and promote sustainable resource management in the coastal region.**

Suggested Solutions

1.

Color Schemes:

- **Pollution Levels**:
  - Level 1: Green (low pollution)
  - Level 2: Yellow (moderate pollution)
  - Level 3: Orange (high pollution)
  - Level 4: Red (very high pollution)
  - Level 5: Dark Red (extreme pollution)
- **Protected Status**:
  - Protected: black
  - Not Protected: Gray

Mapping Design:

- Each feature type (coral reefs, mangrove forests, etc.) can be represented as polygons on a map.
- The color of each polygon will indicate its pollution level, while an overlay will show protected areas with distinct coloring for easy identification.

Enhancement of Understanding:

- This color-coded mapping allows stakeholders to quickly visualize areas that are environmentally at risk due to pollution.
- It provides an immediate understanding of which coastal features are protected and highlights areas needing conservation efforts.
- Decision-makers can easily identify high pollution zones that may require intervention or policy changes.

2. Role of Spatial Analysis

Monitoring Fishing Activities:

- Spatial analysis can reveal patterns in fishing activities by mapping fishing zones against pollution levels. High fishing activity in polluted areas could indicate unsustainable practices and potential health risks for consumers.

Pollution Level Insights:

- By analyzing the spatial distribution of pollution levels, authorities can identify trends over time, such as increases in pollution linked to urban development or fishing activities.
- This information can be used to monitor compliance with environmental regulations and assess the effectiveness of existing conservation measures.

Informing Policy Decisions:

- Insights gained from spatial analysis can lead to targeted policies, such as the establishment of additional protected areas in regions facing high pollution or intensive fishing activities.

- Data can inform zoning regulations for new developments to minimize environmental impact and guide sustainable fishing practices that balance economic needs with ecological health.

Promoting Sustainable Resource Management:

- The combination of monitoring fishing activities and pollution levels supports a holistic approach to coastal management.
- Engaging stakeholders, including local fishing communities and conservation groups, in the decision-making process can enhance compliance and promote community-led conservation efforts.

Case Study: Student Performance Analysis in a School District

You are analyzing a comprehensive dataset that includes various student performance metrics across different schools within a school district. The dataset includes the following variables:

Student ID	School	Grade Level	Subject	Score	Attendance (%)	Extracurricular Participation	Parent Involvement (Yes/No)
001	Lincoln High	11	Math	85	95	Yes	Yes
002	Washington High	10	Science	78	88	No	No
003	Lincoln High	12	English	92	90	Yes	Yes
004	Jefferson Middle	7	Math	70	85	Yes	No
005	Washington High	11	Math	60	80	No	Yes
006	Jefferson Middle	8	Science	82	92	Yes	Yes
007	Lincoln High	10	English	75	85	No	Yes
008	Washington High	12	Science	88	94	Yes	No
009	Jefferson Middle	9	English	90	95	Yes	Yes
010	Lincoln High	11	Science	95	98	Yes	Yes

Questions:

1. **In Tableau, create a detailed dashboard that visualizes the following: (6 Marks)**
    - Average Score and Attendance by School and Grade Level.
    - Distribution of Extracurricular Participation and Parent Involvement across different subjects.

    - Specify how you would organize the layout (e.g., charts, filters) to facilitate analysis and comparison. Justify your choices in terms of usability and clarity.
2. **Analyze the dashboard to derive insights regarding: (4 Marks)**
    - The performance trends of different schools and grade levels, focusing on scores, attendance, and involvement.
    - Recommendations for improving student performance, including strategies to enhance parental involvement and extracurricular participation.

1. Creating a Dashboard in Tableau

Dashboard Layout:

- **Charts**:
    - **Bar Chart**: Average Score by School and Grade Level.
    - **Line Chart**: Attendance Percentage by School and Grade Level.
    - **Pie Chart**: Distribution of Extracurricular Participation (Yes/No) and Parent Involvement (Yes/No) for each subject.

Filters:

- **Grade Level Filter**: Allow users to filter results by specific grade levels to focus on performance trends for particular grades.
- **School Filter**: Users can select specific schools to compare their performance directly.

Organizing the Layout

- **Top Section**: Place the bar chart and line chart side by side for direct comparison between average scores and attendance.
- **Bottom Section**: Position the pie chart for extracurricular participation and parental involvement to provide context for how these factors correlate with academic performance.
- **Usability and Clarity**:
    - Use consistent color coding across charts to represent schools, making it easy to follow trends.
    - Provide tooltips for detailed information on hover, enhancing user engagement without cluttering the visual space.

2. Analysis of the Dashboard

Performance Insights

- **Average Scores**:
    - Lincoln High generally shows higher average scores across grade levels, particularly in Math and Science.
    - Washington High has lower average scores, especially in Math for Grade 11, indicating a potential area for academic improvement.
- **Attendance Trends**:
    - High attendance percentages correlate with better performance, especially at Lincoln High, suggesting that consistent attendance is vital for academic success.
    - Jefferson Middle shows varied attendance rates, which may reflect differences in student engagement or support services.
- **Extracurricular Participation and Parent Involvement**:

- Schools with higher participation in extracurricular activities (e.g., Lincoln High) tend to have better academic performance and higher parent involvement rates.
- Conversely, Washington High has lower participation and scores, indicating a potential disconnect between school engagement and student performance.

Recommendations for Improving Student Performance

1. **Enhancing Parental Involvement**:
   - Develop workshops and communication strategies to engage parents more actively, especially in underperforming schools like Washington High.
   - Encourage volunteer opportunities at schools to foster a stronger school community and increase parent involvement.
2. **Promoting Extracurricular Activities**:
   - Increase awareness and access to extracurricular programs, especially in schools with lower participation rates.
   - Highlight the benefits of extracurricular involvement on academic performance to encourage more students to participate.
3. **Targeted Academic Support**:
   - Implement tutoring programs and study groups in subjects where students are struggling, particularly at Washington High.
   - Monitor attendance closely and develop initiatives to improve engagement and reduce absenteeism, which directly impacts performance.

a) Tableau Visualization Design

1. Visualization Elements:

- **Size (Exam Score)**: Use the size of the data points (bubbles) to represent the Exam Score. Larger bubbles indicate higher scores, allowing quick visual identification of top performers.
- **Color (Attendance)**: Apply a color gradient to the data points based on Attendance percentages. For instance, a gradient from red (low attendance) to green (high attendance) provides a clear visual cue about attendance levels.
- **Shape (Gender)**: Different shapes for male and female students enhance the visualization's clarity. For example, circles for females and squares for males help in distinguishing between genders at a glance.

2. Design Choices:

- **Clarity and Usability**: By using size, color, and shape, the visualization effectively conveys multiple dimensions of data without overwhelming the viewer. This approach allows for easy identification of trends and patterns.
- **Interactive Features**: Incorporate tooltips that display detailed information (Student ID, Course, Exam Score, Attendance, Gender) when hovering over data points, enhancing user engagement and data exploration.
- **Axes**: While not necessary for every visualization type, consider including axes for Attendance percentage to provide context on performance. This could be a secondary axis that complements the size and color dimensions.

b) Interpretation and Insights

i. How do exam scores correlate with attendance?

- **Analysis**: Examining the visualization, it becomes apparent that students with higher attendance percentages generally have higher exam scores. For example, the largest bubbles (highest scores) tend to cluster towards the green end of the attendance gradient.
- **Insight**: This suggests a positive correlation between attendance and exam performance, indicating that regular class attendance is likely contributing to better academic outcomes.

ii. Are there performance differences between genders across courses?

- **Analysis**: Looking at the shapes and sizes of the bubbles in the visualization, we can see that:
    - Female students (circles) tend to perform better in Literature (92) and Science (88), while male students (squares) show a wider range of scores, with lower performance in Mathematics (65).
    - In Mathematics, the male student scored significantly lower than the female student in Science.
- **Insight**: This indicates that there might be gender-specific strengths in different courses, and while female students show higher performance overall, males may struggle particularly in Mathematics.

iii. What recommendations can you provide to improve performance for lower-scoring groups?

- **Targeted Support**: For the lower-scoring male students in Mathematics, consider implementing tutoring programs or study groups that focus specifically on areas where they struggle.
- **Encourage Attendance**: Since higher attendance correlates with better scores, initiatives to improve attendance could include:
    - Incentives for students with high attendance rates.
    - Engaging teaching methods that encourage participation and attendance.
- **Tailored Interventions**: Analyze attendance data further to identify patterns—such as specific courses where attendance is particularly low—and develop targeted interventions for those courses.
- **Mentorship Programs**: Pairing lower-performing students with mentors or peers who excel in the subject could provide additional support and motivation.