**Part 1: Geospatial Analysis**

**1. Stakeholder Spatial Needs**

**a. Key Geographic Patterns**

* Law Enforcement: Focus on crash hotspots and high-risk intersections to optimize patrols.
* City Planners: Analyze crash density and road infrastructure for safer designs.
* Public Health Officials: Identify areas with severe injuries to allocate emergency resources.
* Insurance Companies: Assess risk clusters for pricing and claims analysis.
* General Public: Need safe route maps highlighting accident-prone areas.

**b. Interaction with Spatial Data**

* Law Enforcement: Real-time maps for quick decision-making.
* City Planners: Compare regions and model infrastructure impacts.
* Public Health: Use heatmaps to track injury trends.
* Insurance: Identify risk clusters and overlay policy data.
* Public: Interactive maps for safe route planning.

**c. Relevant Geographic Scales**

* Neighborhood: For local safety issues (schools, intersections).
* City: For strategic resource allocation and urban planning.
* Regional: To understand broader trends and coordinate policies.

2. **Data Assessment:**

* **Location Representation:** Point data using latitude and longitude.
* **Geographic Coverage:** Monroe County from 2003 to 2015.
* **Complementary Data Needed:** Population density, traffic volume, road conditions.

3. **Mapping Approaches:**

**a. Visualizations**

A. **Cluster Map:** This approach groups crash data points based on geographic proximity. As users zoom in, clusters break down into individual crash markers, revealing detailed information about each incident. This helps in identifying crash hotspots without overwhelming the map with overlapping markers.

A map with pins on it

Description automatically generated

B. **Heatmap:** The heatmap visualizes crash density using color gradients, where warmer colors (e.g., red) indicate higher concentrations of crashes. This approach provides an intuitive overview of accident-prone areas, highlighting patterns that may not be immediately visible in raw data.

A map with a map and a map

Description automatically generated with medium confidence

**b. Interaction with Other Data Dimensions**

* **Temporal Data:** Both mapping approaches can be enhanced by integrating temporal data such as time of day, day of the week, or season. This helps in identifying trends like peak accident hours or seasonal variations in crash frequency.
* **Injury Severity:** Layering crash severity data on top of the cluster map allows for quick identification of areas with high numbers of severe accidents. In the heatmap, adjusting the weight of each data point based on injury severity can emphasize more critical hotspots.
* **Demographic Data:** Integrating demographic information (e.g., population density, age distribution) can help uncover correlations between crash frequency and socio-economic factors, aiding in targeted interventions.

**c. Proposed Visualization Sketches**

* **Cluster Map:** Features dynamic clusters with interactive popups displaying crash details (e.g., injury type, primary factor). Clusters are color-coded based on the number of incidents, with larger clusters indicating higher crash densities.
* **Heatmap:** Displays smooth gradients from green (low density) to red (high density), representing crash concentrations across the map. The heatmap can be layered with road networks to identify hazardous intersections and corridors.
* **Choropleth Map:** Shows crash density by regions.
* **Time-Animated Heatmap**: Reveals trends over time.
* **Injury Severity Layers**: Toggle between severity levels for targeted insights.

**d. Design Rationale**

* **Cluster Map:** Selected for its ability to manage large datasets effectively, providing both macro and micro-level insights. It is particularly useful for law enforcement and city planners who need to analyze specific crash locations and patterns.
* **Heatmap:** Chosen for its simplicity and effectiveness in highlighting broad spatial patterns. It offers an immediate visual representation of crash density, making it easy to identify high-risk areas for quick decision-making.

**Part 2: AI-Assisted Design Process**

**1. Documenting AI Interactions**

* AI Model Used: ChatGPT (OpenAI, GPT-4)
* Prompts Used:
  + "Explain geospatial analysis concepts relevant to car crash data."
  + "Suggest effective approaches for visualizing spatial data."
  + "Describe techniques for identifying geographic patterns in accident datasets."
  + "Help organize concise documentation for geospatial data analysis projects."
  + "Outline key stakeholder needs in spatial data projects."
* Prompt Structuring: Prompts were structured to focus on conceptual understanding, data visualization strategies, and documentation guidance. The goal was to encourage comprehensive, high-level responses that could be adapted to specific project requirements without relying on direct code outputs.

**2. Implementation Plan**

* **Data Preparation Steps:**
  + Cleaning data to handle missing or inconsistent latitude/longitude values.
  + Standardizing date and time formats for temporal analysis.
  + Removing duplicate or erroneous records to maintain data quality.
* **Data Analysis & Visualization Tools:**
  + **Pandas:** For efficient data manipulation and preprocessing.
  + **Folium & Leaflet:** To create interactive maps like heatmaps and cluster maps.
  + **Seaborn & Matplotlib:** For statistical data visualization.
  + **GeoPandas:** For spatial operations when dealing with geographic boundaries.
* **Tool Selection Rationale:** Chosen for their versatility, ease of use, and robust support for geospatial data analysis in Python.
* **Interactive Features:**
  + Zoomable, dynamic maps with cluster breakdowns.
  + Popups showing detailed crash information.
  + Layer controls to toggle between different data visualizations, such as injury severity.
* **Handling Data Quality Issues:**
  + Imputing missing values where feasible.
  + Removing outliers and inconsistent geographic points.
  + Ensuring data consistency through validation checks.

**3. Evaluating AI Suggestions**

* **Helpful Contributions:**
  + Offered effective Python code templates for heatmaps and clustering.
  + Suggested additional visualizations like choropleth maps and time-animated heatmaps.
  + Helped structure concise, well-organized documentation.
* **Limitations Encountered:**
  + Some code required modifications for dataset compatibility.
  + Limited ability to address complex data-specific issues without manual adjustments.
* **Modifications:**
  + Refined AI-generated code to fix syntax issues and align with specific dataset requirements.
  + Integrated personal insights to enhance stakeholder analysis and visualization techniques.
* **Missed Best Practices:**
  + Initial outputs lacked proper legends, titles, and error handling in code.
  + Addressed these gaps by adding clear legends, improving interactivity, and enhancing visual aesthetics manually.