**My Visualization Project - Topical Visualizations**

**Cali State Fire Dashboard  
  
A close up of words

Description automatically generated  
A screenshot of a computer

Description automatically generated**

**Part 1: Categorical/Linguistic Analysis**

**Stakeholder Category/Text Analysis Needs**

1. **Stakeholders and Needs:**
   * **Firefighters and Emergency Responders:**
     + Insights: Require patterns in fire incidents by type, year, and geographic region to plan resource allocation effectively.
     + Interaction: Analyze categorical patterns of fire names, states, and fire years for quick response prioritization.
   * **Policy Makers and Environmental Agencies:**
     + Insights: Need data on trends in fire occurrences by type and region to develop preventive measures and allocate funding.
     + Interaction: Explore categorical data distributions and recurring patterns (e.g., most common fire types or highest-risk years).
   * **Public Awareness Campaigners:**
     + Insights: Use text data on fire precautions to craft effective messaging for public safety campaigns.
     + Interaction: Analyze broad themes in fire-related precautions to identify communication priorities.
2. **Type of Text Data:**
   * **Structured Data:** Predefined categories (e.g., fire names, states, years).
   * **Unstructured Text:** Free-form text (e.g., fire precautions and safety messages).
   * **Analysis Focus:**
     + Broad themes: Identify common fire-related safety measures and prevention strategies.
     + Specific topics: Extract insights from categorical data like state-wise fire occurrences.

**Data Assessment**

1. **Representation of Categories/Text:**
   * **Structured Categories:** Fire data includes fire names, state-wise distributions, and fire years, represented in predefined classification systems.
   * **Unstructured Text:** Includes fire precautionary measures, providing rich insights for awareness campaigns.
   * **Mixed Data:** Combines structured numerical and categorical data with unstructured textual inputs.
2. **Quality and Consistency:**
   * **Categorical Data:** Well-structured and consistent, enabling clear visualization of patterns.
   * **Unstructured Text:** Varied in length and complexity but relevant for qualitative insights.
   * **Preprocessing Needs:**
     + For unstructured text: Tokenization, stop-word removal, and lemmatization for text analysis.
     + For categorical data: Standardization of labels (e.g., fire names and states).
3. **Additional Context Needed:**
   * Incorporating temporal trends in fire patterns.
   * Adding demographic or environmental context for better interpretation (e.g., population density or vegetation data).

**Initial Design Exploration**

1. **Visualization 1: Sankey Diagram of Fire Data**
   * **Design:**
     + Displays flow between fire names, states, and years.
     + Visualizes how fire incidents are distributed across states and years.
   * **Purpose:**
     + Helps stakeholders like policymakers and emergency planners understand categorical relationships in fire data.
     + Enables trend analysis to prioritize regions or years with frequent incidents.
   * **Insights:**
     + Shows high-frequency states and fire names, aiding in efficient resource allocation.
2. **Visualization 2: Word Cloud of Fire Precautions**
   * **Design:**
     + Highlights frequently mentioned fire precautionary measures.
     + Displays terms like "evacuation," "fire-resistant," and "emergency" prominently.
   * **Purpose:**
     + Educates the public by identifying key safety themes for awareness campaigns.
   * **Insights:**
     + Identifies recurring safety measures, assisting campaigners in crafting targeted messages.

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

1. **AI Tools Used:**
   * **Model:** ChatGPT (OpenAI v4.0).
   * **Prompts Used:**
     + “Suggest visualizations for categorical data in fire incidents.”
     + “How can I represent fire precautions effectively in a word cloud?”
     + “What are best practices for preprocessing unstructured text for NLP?”
2. **Why These Prompts Were Structured:**
   * Open-ended prompts ensured AI provided a range of visualization options.
   * Specific prompts focused on improving data preprocessing and visualization clarity.
3. **Implementation Plan:**
   * **Data Preparation:**
     + Clean unstructured text: Remove irrelevant data, tokenize, and lemmatize for word cloud generation.
     + Standardize categorical labels: Ensure consistent naming conventions for states, fire names, etc.
   * **Tools Used:**
     + **Pandas and NLTK:** For preprocessing structured and unstructured data.
     + **Matplotlib and Plotly:** For creating the Sankey diagram and word cloud.
     + **Reason for Selection:** These tools offer precision and flexibility in handling both structured and textual data.
   * **Interactive Features:**
     + Enable filtering by state, fire year, or fire name in the Sankey diagram.
     + Interactive word cloud allowing deeper exploration of safety themes.
4. **Evaluation of AI Suggestions:**
   * **Helpful Suggestions:**
     + AI effectively recommended Sankey diagrams and word clouds, matching stakeholder needs.
     + Suggested using NLP for better insights from textual data.
   * **Limitations:**
     + AI did not address potential overlap issues in the Sankey diagram visualization.
     + Required manual adjustments to align word cloud themes with stakeholder priorities.
   * **Best Practices Missed:**
     + AI didn’t emphasize integrating visualizations with temporal trends, which was manually incorporated.
     + Missed suggesting tooltip features for Sankey diagram interactivity.