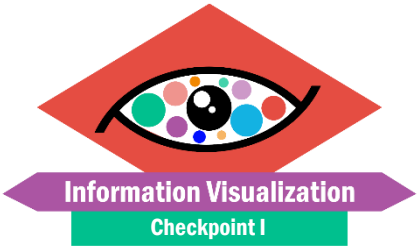

Checkpoint I: Project Proposal

Group: G43
Date: 2025/09/12

Problem Domain

The problem domain of this project focuses on **environmental hazards, public safety, and prevention**, specifically analyzing rural fires in Portugal. Rural fires threaten human life, infrastructure, and ecosystems, with socioeconomic impacts including economic losses, community disruption, and biodiversity degradation. Portugal's Mediterranean climate, extensive forests, rural depopulation, and human activities such as negligence or intentional fire-setting make the country particularly vulnerable. Understanding patterns of fire occurrence, causes, and fire sizes is essential for developing effective prevention strategies and mitigating their impact.

Prevention relies not only on understanding fire dynamics but also on the strategic allocation of firefighting resources. The number and distribution of **firefighters** is a key factor in reducing fire damage and responding efficiently to emergencies. By analyzing data on fires by year, region, cause, size, and firefighting personnel, this project aims to uncover trends, identify high-risk regions, and explore correlations between fire occurrence and resource availability. This data-driven approach supports proactive planning, resource optimization, and more effective fire prevention measures, highlighting the intersection of environmental management, public safety, and regional planning in Portugal.

Task Abstraction

The purpose of this project is to analyze rural fires in Portugal by integrating **temporal, spatial, categorical, and resource-related dimensions**. The study examines trends over time, regional patterns, causes, fire sizes, and their relationship with available firefighting personnel, with a particular focus on how these factors can inform strategies for fire prevention.

1. Temporal Trend Analysis

- **Abstract form:** How does X change over time?
- **Specific question:** How has the total number of fires in Alentejo Central evolved over the past fourteen years?
- **Task type:** Trends / Ordinal time

2. Regional Impact Comparison

- **Abstract form:** How does X differ across Y?
- **Specific question:** Which regions have the highest number of fires in 2024, and how does this compare with the number of firefighters in each region?
- **Task type:** Comparison / Geolocation / Correlation

3. Cause Distribution Analysis

- **Abstract form:** What is the distribution of X?

- **Specific question:** What is the distribution of fire causes within Algarve in 2024, and does it correlate with the number of firefighters available?
- **Task type:** Distribution / Proportion / Correlation

4. Multi-dimensional Relationships

- **Abstract form:** How does X relate to Y across multiple dimensions?
- **Specific question:** How do fire causes and dimensions correlate with the number of firefighters in Alto Minho in 2023?
- **Task type:** Correlation / Relationships / Multi-dimensional

5. Resource Proportional Analysis

- **Abstract form:** How does X proportionally compare across Y?
- **Specific question:** Which regions have the most or fewest firefighters relative to the number of fires, and does this reveal patterns of high fire risk?
- **Task type:** Proportion / Geolocation / Comparison

6. Multi-dimensional Relationships

- **Abstract form:** How does X relate to Y across multiple regions?
- **Specific question:** Is there a relationship between the number of fires and the number of firefighters in each region in 2020?
- **Task type:** Correlation / Relationship / Geolocation

7. Resource Proportional Analysis

- **Abstract form:** How does X distribute across Y?
- **Specific question:** How is the total burned area distributed across regions and does it correlate with the number of fires by dimension?
- **Task type:** Distribution / Proportion / Multi-dimensional

This set of tasks ensures a **diverse and integrated analytical experience**, enabling the exploration of trends, spatial differences, distributions, correlations, and proportionalities. Together, they provide a comprehensive view of rural fire patterns and their influencing factors, helping to uncover insights across multiple dimensions of the phenomenon.

Data Abstraction

The data for this project comes from **multiple sources provided by INE (Instituto Nacional de Estatística)**, covering different aspects of rural fires in Portugal:

1. **Fires by Cause** – Number of fires in each region and year, broken down by cause (natural, negligence, intentional, etc.).
2. **Fires by Dimension** – Number of fires per size class for each region and year (< 1 ha, 1 - < 10 ha, …, >= 1 000 ha).
3. **Burned Area** – Percentage of area burned per region and year.
4. **Firefighters** – Number of firefighting personnel in each region and year.

All datasets are **static** and table-based, downloadable in CSV. Each dataset contains a regional (NUTS-2024) and temporal (year) component that allows integration.

Data Processing and Cleaning

1. **Filtering and selection of relevant variables**
    - Kept: Ano, Região, Percentagem, Sapadores, Total, Causas and Dimensões.
2. **Handling missing values**
    - Missing fire counts or dimension counts → treated as zero.
    - Missing causes → labeled as "Unknown."
    - Missing firefighters → labeled as "NULL."
3. **Converting the CSV files into JSON files**
    - Convert the CSV files for the causes, dimensions, burned area percentage and firefighters into JSON files.
4. **Merging datasets**
    - Cause, dimension, burned area and firefighters' data merged by Year and Region into a single unified dataset.
5. **Final export format**
    - Single JSON suitable for D3.js to power all dashboard components.

Attribute	Type	Scale	Description	Derived Measure?
Ano	Ordinal	Linear	Year of the fire occurrence	No
Região	Nominal	Categorical	NUTS-2024 region of the fire	No
Percentagem	Quantitative	Ratio	Percentage of burned area in the region	No
Sapadores	Quantitative	Ratio	Number of firefighters in the region for the year	No
Total	Quantitative	Ratio	Total number of fires in a region and year	No
Causas	Nominal	Categorical	List of causes of the fires	No
Dimensões	Ordinal	Categorical	List of the numbers of fires in each size class (<1 ha -> >= 1000 ha)	No

Mapping

- **How has the total number of fires in Alentejo Central evolved over the past fourteen years?**
  - Varied between 142 and 369 annual occurrences. Peaks in 2012 (369) and 2019 (312); minimums in 2023 (142) and 2018 (159). We can see a slight downward trend overall.
- **Which regions have the highest number of fires in 2024, and how does this compare with the number of firefighters in each region?**
  - Highest: Alto Minho (1882 fires, 115 firefighters), Alto Tâmega e Barroso (864 fires, 100 firefighters), Região de Coimbra (356 fires, 185 firefighters).
  - Relation: The most affected regions don't always have the most firefighters (e.g., Alto Minho has far more fires than Coimbra but fewer firefighters).
- **What is the distribution of fire causes within Algarve in 2024, and does it correlate with the number of firefighters available?**
  - 345 fires; 209 negligence, 14 intentional, 115 undetermined.
  - Firefighters: 40.
  - Conclusion: fires mostly due to negligence, no clear correlation with the relatively low number of firefighters.
- **How do fire causes and dimensions correlate with the number of firefighters in Alto Minho in 2023?**
  - 1163 fires. Causes: 441 negligence, 481 intentional, 214 re-ignitions.
  - Dimensions: 857 <1ha, 234 between 1–10ha, very few large fires.
  - Firefighters: 120.
  - Conclusion: despite many firefighters, most fires are human-caused and small.
- **Which regions have the most or fewest firefighters relative to the number of fires, and does this reveal patterns of high fire risk?**
  - Most firefighters /fire: Região de Coimbra (185 firefighters /356 fires ≈ 0.52).
  - Fewest firefighters /fire: Alto Minho (115/1882 ≈ 0.06).
  - Pattern: Northern regions face more fires but proportionally fewer firefighters → higher risk.
- **Is there a relationship between the number of fires and the number of firefighters in each region in 2020?**
  - We can see that overall, there is a linear relation: generally, the more fires the region has, the more firefighters it also has.
- **How is the total burned area distributed across regions and does it correlate with the number of fires by dimension?**
  - Small fires (<1ha) dominate everywhere (e.g., >70% in Algarve, >80% in Alto Minho).
  - Large fires (>100ha) are rare but account for most burned area, especially in the North and Center.
  - The distribution shows most incidents are small, but the greatest risk comes from the few large fires.