Statistics Report
Wildfires of Portugal (2015)

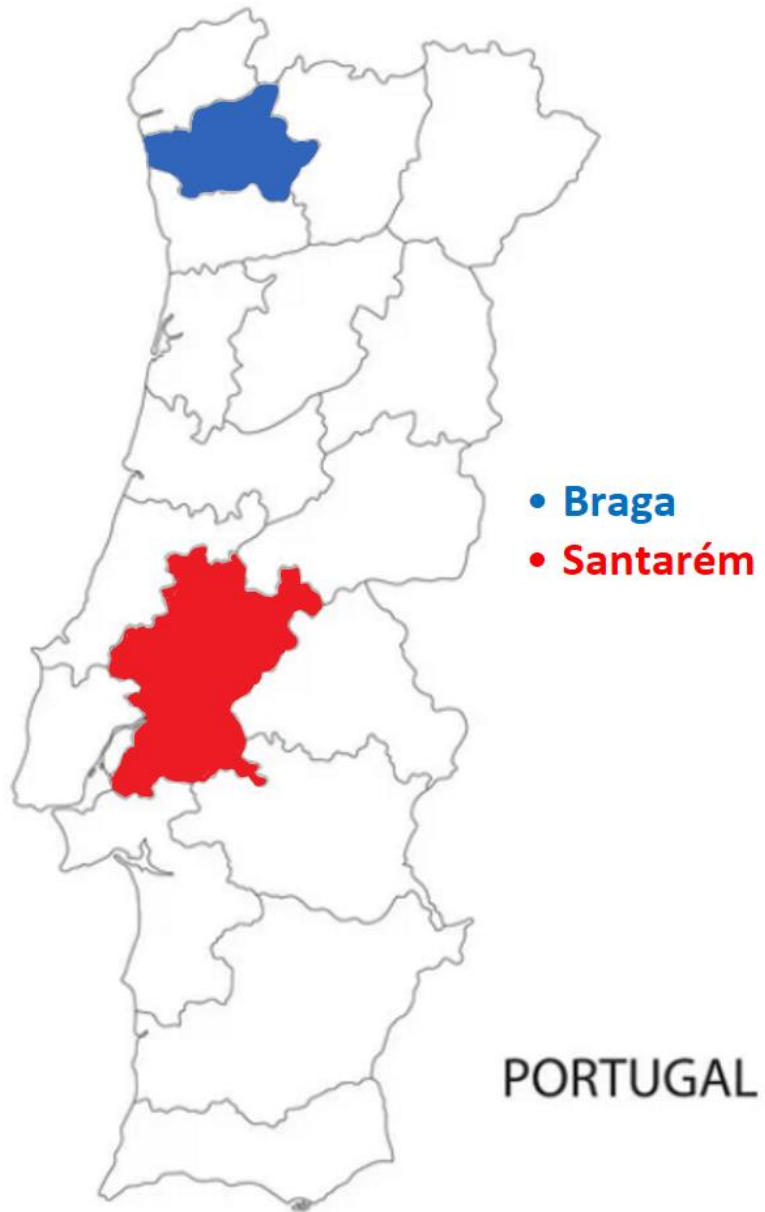Master's degree in Data Science & Engineering - FEUP

Fundamentos de Ciência e Engenharia dos Dados

October 2022

Group 6

Henrique Ribeiro, Rojan Aslani, Sónia Ferreira

Professor: António Miguel Gomes

Introduction

- Braga
- Santarém

PORTUGAL

- Wildfires can have immeasurable consequences for both structures and living beings. For this reason, studying and managing the response to wildfires is important for maintaining resources, protecting people and ecosystems, and reducing air pollution.

- Thus, this work presents an explanatory analysis for the wildfires of 2015, in the districts of Braga and Santarém, through the reported data from Instituto da Conservação da Natureza e das Florestas – ICNF.

- Method used on this study: Six-Steps Statistical Investigation Method

- **Research Question:**

Compare response, extinction time and cause of the wildfires, to see the influence on the burn areas for districts of Braga and Santarém.

Materials and Methods

```
#dataset size
print(df.shape)    (23175, 38)
```


Missing values in the dataset
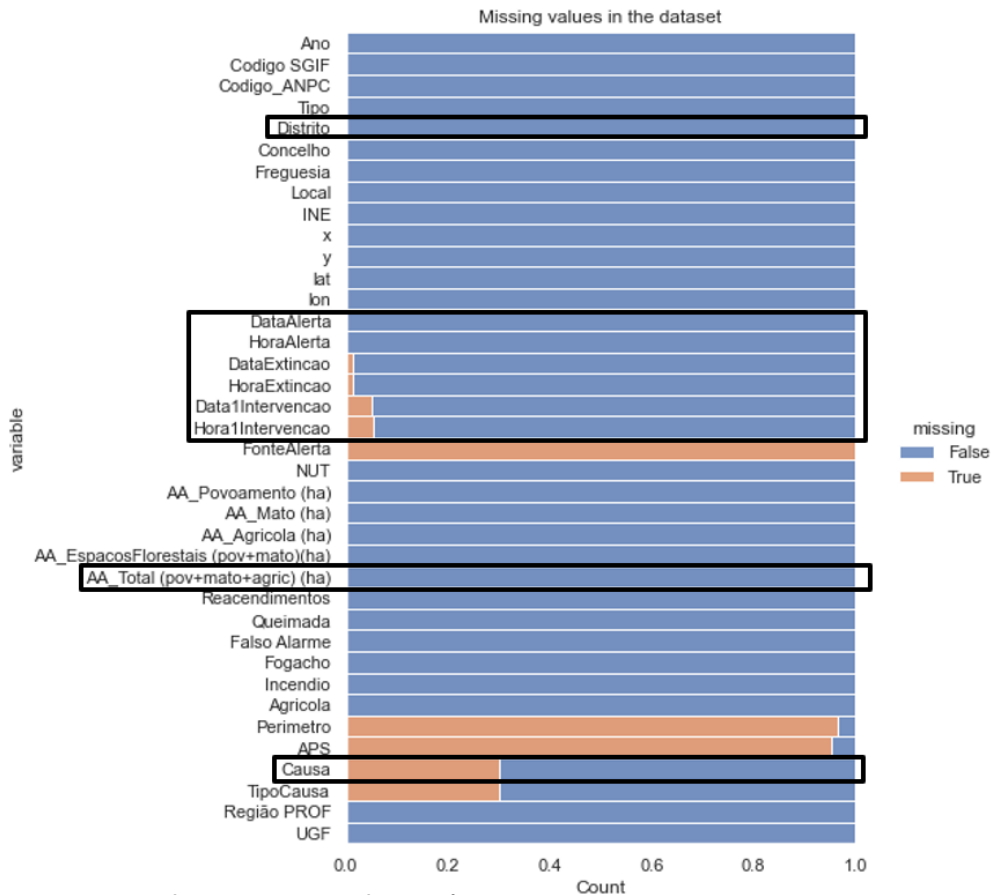
New datasets data:

- **District name** (Categorical ariable – Braga and Santarém)
- **Code of cause of fire** (Categorical variable – 1 and 4*)
- **Total Burned area** (Numerical continuous variable)
- **Intervention time** (subtraction of Intervention time from Alert time; Numerical discrete variable)
- **Extinction time** (subtraction of Extinction time from intervention time; Numerical discrete variable)

For intervention and extinction time, all values under 2 minutes were considered input error and were removed, as this time of intervention is unrealistic.

Henrique Ribeiro, Rojan Aslani, Sónia Ferreira

*Cause 1: Use of fire (codes 1-199) | Cause 4: Incendiarism (code 400-499)

STEP 3: Explore the Data

Numerical Summary and Statistics

```
print(district1 , dataset1.shape)
Braga (952, 4)
```

Braga

	Cause	InterventionTimeMin	ExtinctionTimeMin	AA_Total (pov+mato+agric) (ha)	Distrito no	Distrito Nr.
count	952.000000	952.000000	952.000000	952.000000	952.0	952.0
mean	1.633403	13.153361	156.433824	3.341241	1.0	1.0
std	1.224984	8.890360	210.972516	11.873399	0.0	0.0
min	1.000000	2.000000	10.000000	0.000000	1.0	1.0
25%	1.000000	8.000000	70.000000	0.050000	1.0	1.0
50%	1.000000	12.000000	110.000000	0.450000	1.0	1.0
75%	1.000000	16.000000	175.000000	1.757500	1.0	1.0
max	4.000000	152.000000	3260.000000	143.000000	1.0	1.0

```
print(district2 , dataset2.shape)
Santarém (689, 4)
```

Santarém

	Cause	InterventionTimeMin	ExtinctionTimeMin	AA_Total (pov+mato+agric) (ha)	Distrito no	Distrito Nr.
count	689.000000	689.000000	689.000000	689.000000	689.0	689.0
mean	2.258345	12.288824	93.483309	3.629433	2.0	2.0
std	1.481482	7.610382	122.317615	60.780020	0.0	0.0
min	1.000000	2.000000	7.000000	0.000100	2.0	2.0
25%	1.000000	7.000000	42.000000	0.015800	2.0	2.0
50%	1.000000	11.000000	68.000000	0.067000	2.0	2.0
75%	4.000000	16.000000	109.000000	0.313200	2.0	2.0
max	4.000000	44.000000	2096.000000	1580.000000	2.0	2.0

STEP 4: Draw inferences

- **Research Question:** Compare response, extinction time and cause of the wildfires, to see the influence on the burn areas for districts of Braga and Santarém.

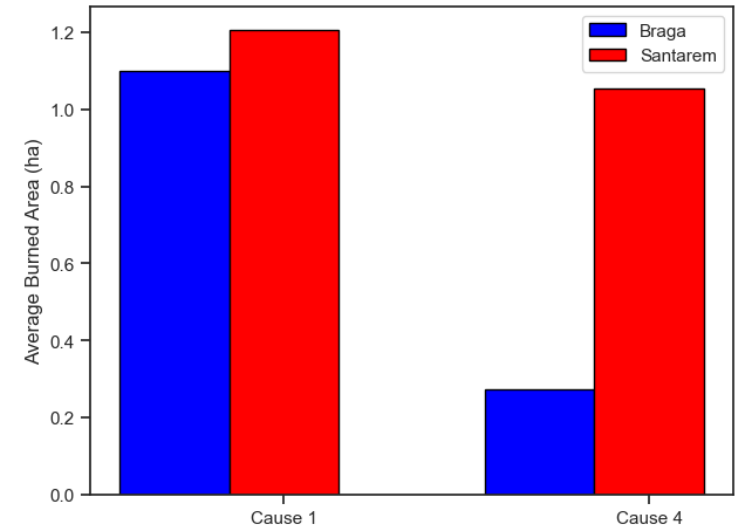**Question 1.** Does the cause of fire impact the burnt area?

Question 2. Does the intervention time impact the burnt area of the district?

Question 3. Does the extinction time impact the burnt area of the district?
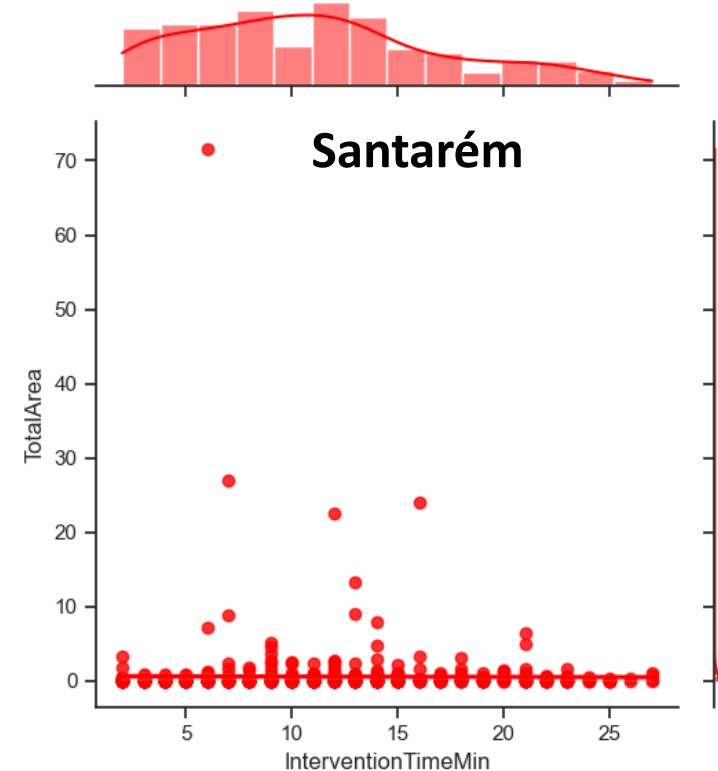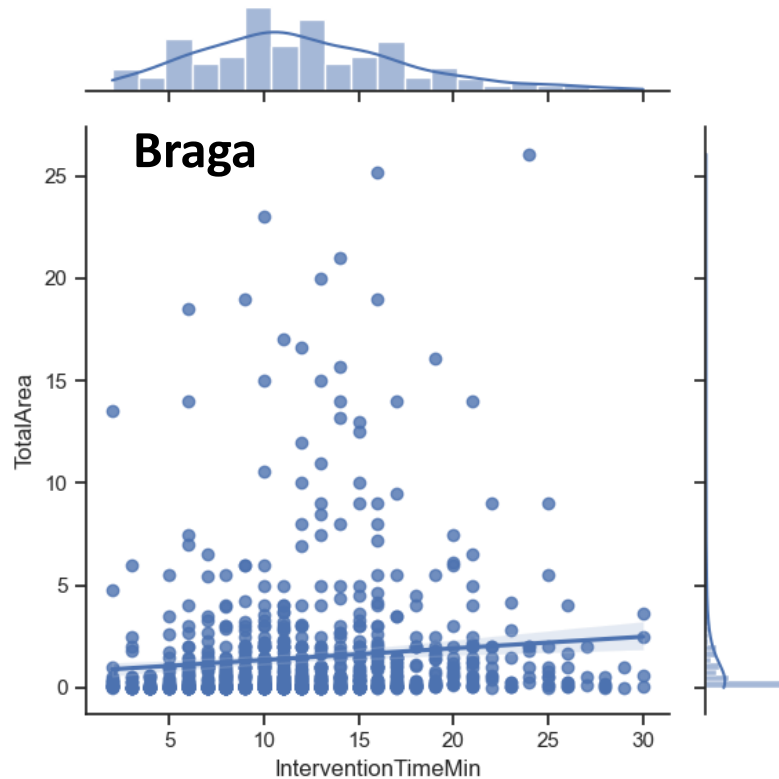
Inference 1

- **Question 1:** Does the cause of fire impact the burnt area?

    - Hypothesis test: T-test

    - H_0 = There is no association between mean burnt area and the cause of fire.

    - H_a = there is an association between mean burnt area and the cause of fire.

    - P-value (Braga) = 0.1 $\longrightarrow$ **Failed to reject** H_0 (0.10 > 0.05)

    - P-value (Santarém) = 0.0 $\longrightarrow$ H_0 is **rejected** (0.00 < 0.05)

Inference 2

- **Question 2:** **Does the intervention time impact the burnt area of the district?**
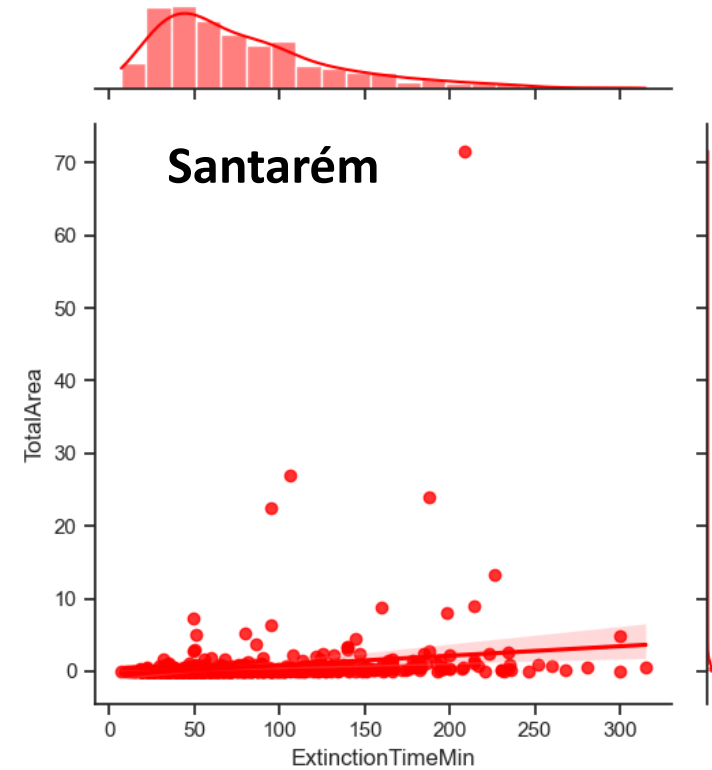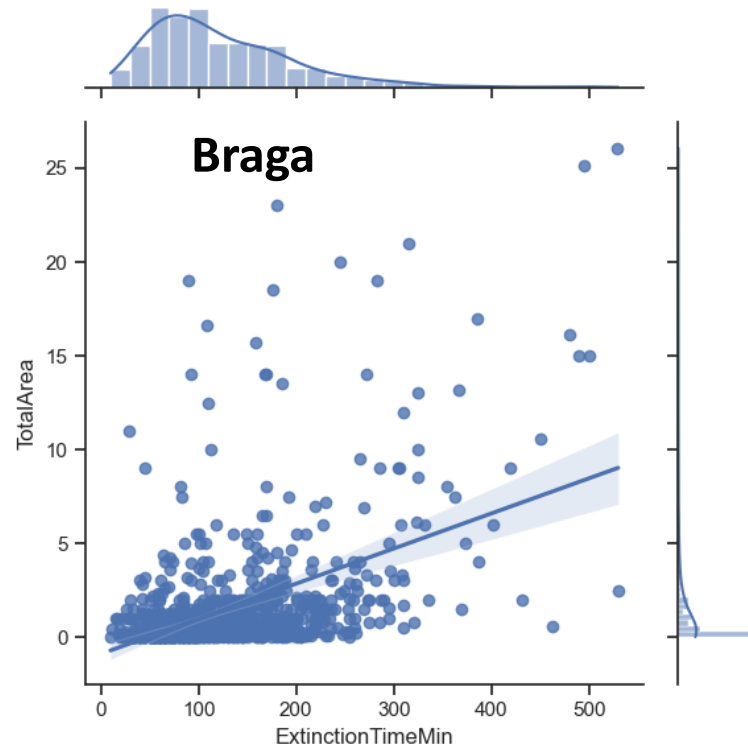
Inference 2

- **Question 2:** Does the intervention time impact the burnt area of the district?

  - Covariance & Pearson's correlation coefficient
    - how strong is the relationship between the two variables

  - Low covariance levels
    - **1.8** and - **0.2** for Braga and Santarém, respectfully

  - Low Pearson's correlation coefficient:
    - **0.1** and -**0.01** for Braga and Santarém, respectfully

  - **Braga**: minimal, extremely minute, positive correlation between Intervention Time and Burnt Area

  - **Santarém**: no correlation between Intervention Time and Burnt Area

- **Question 3: Does the extinction time impact the burnt area of the district?**

Inference 3

- **Question 3:** Does the extinction time impact the burnt area of the district?

  - Higher **covariance** levels
    - **122** - Braga
    - **38 -** Santarém

  - Higher **Pearson's correlation coefficient**:
    - **0.49** - Braga
    - **0.16** - Santarém

  - **Braga**: moderately positive correlation between Extinction Time and Burnt Area
  - **Santarém**: low positive correlation between Extinction Time and Burnt Area

Inference 3

- **Question 3:** Does the extinction time impact the burnt area of the district?

  - P-value test for the results of R value for Santarém

    - H_0 = There is <u>no</u> association between mean burnt area and extinction time.

    - H_a = There is an association between mean burnt area and extinction time.

  - P-Value = 0 $\longrightarrow$ H_0 is **rejected** (0.00 < 0.05)

  - Possible association between mean burnt area and extinction time in Santarém
    - Confirms the previous results
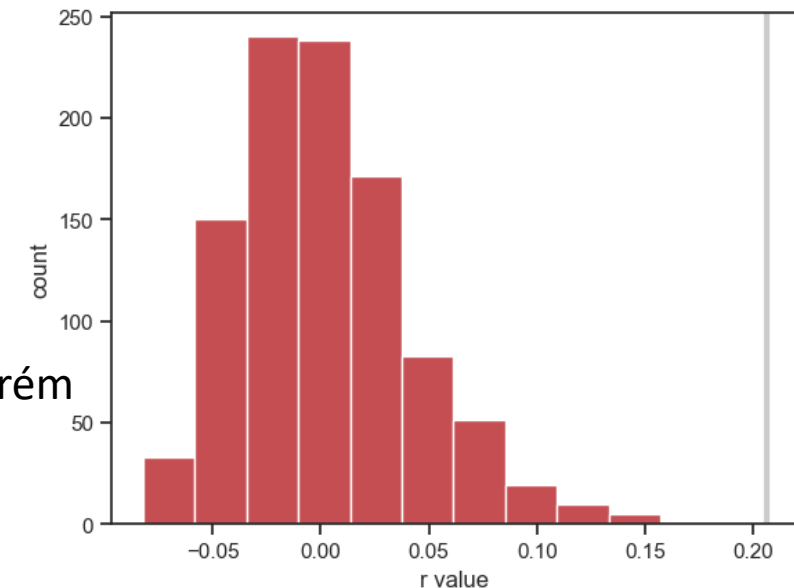
STEP 5: Conclusions
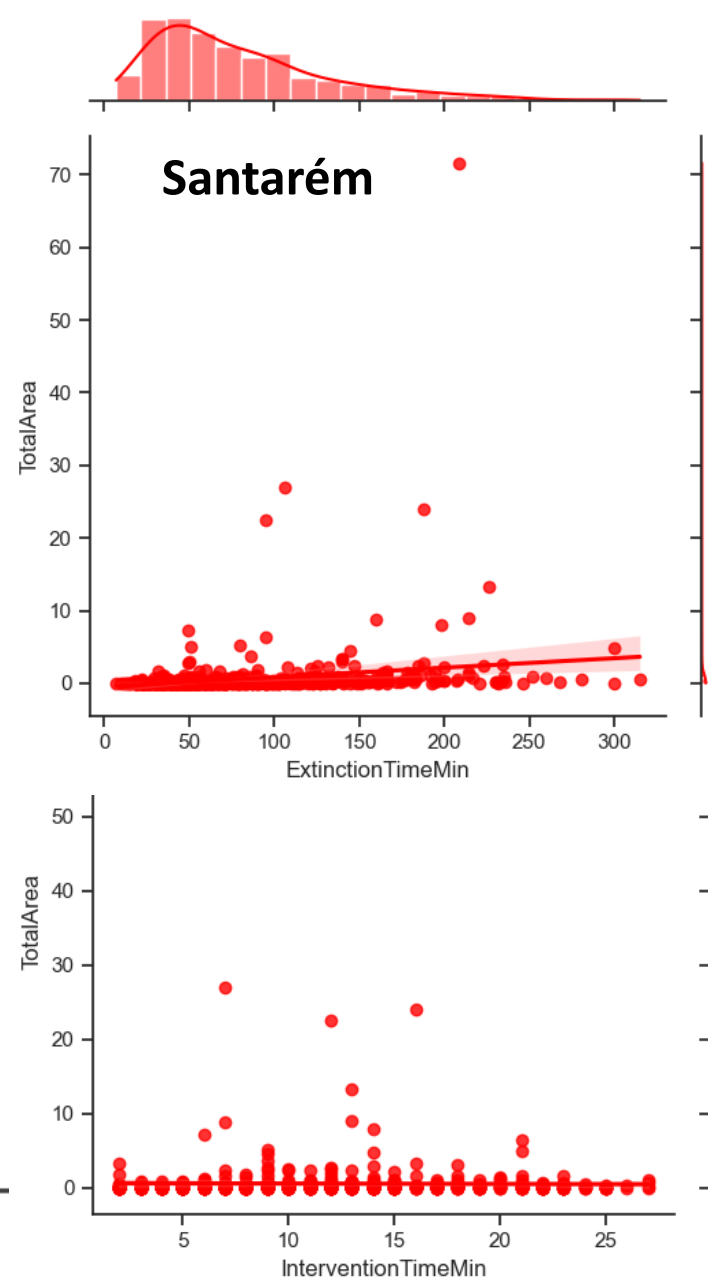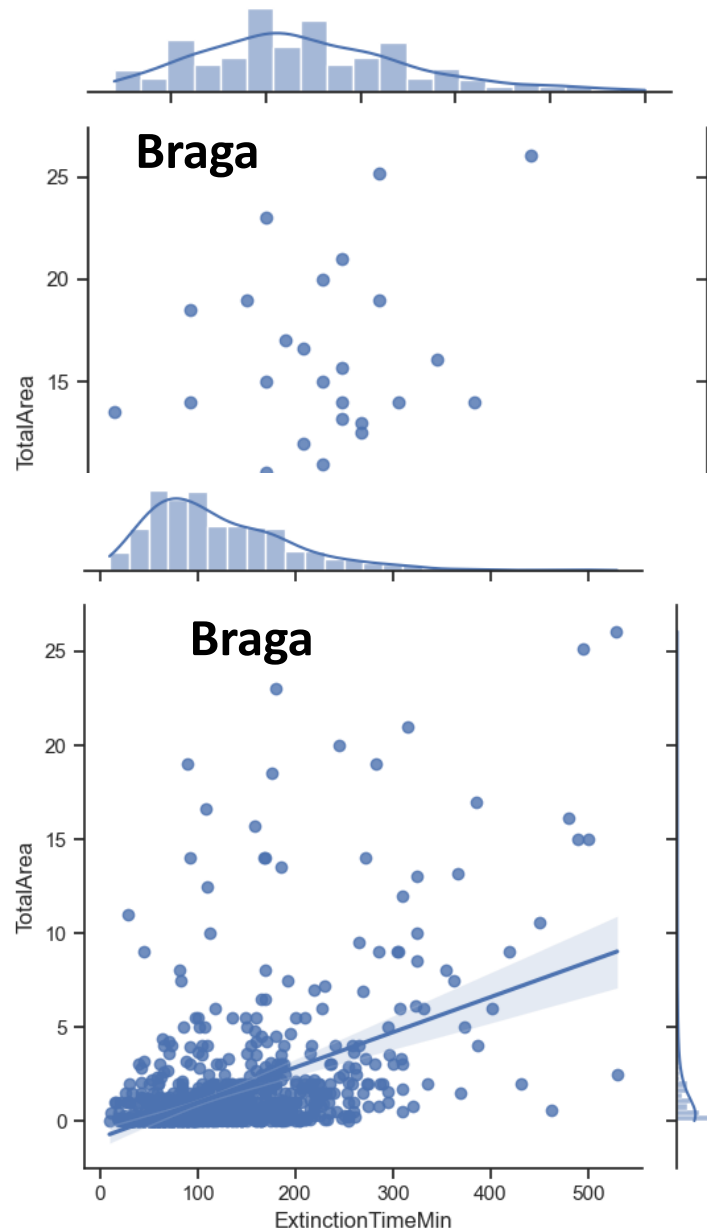
- **Burnt area vs. cause of fire:**
  - ✗ Braga
  - ☑ Santarém

- **Burnt area vs. intervention time:**
  - ✗ Braga or Santarém

- **Burnt area vs. extinction time:**
  - ☑ Braga
  - ☑ Santarém

STEP 6: Look back and ahead

- Analyzing a more detailed and accurate dataset
  - number of staff
  - Temperature, wind and humidity

- Analyzing more years
  - Draw conclusions about districts

Thank you!

- Henrique Ribeiro
- Rojan Aslani
- Sónia Ferreira