Msc Data Science

DATA VISUALIZATION

2nd Assignment

VOULGARI ELENI

A.M. 17005

Introduction

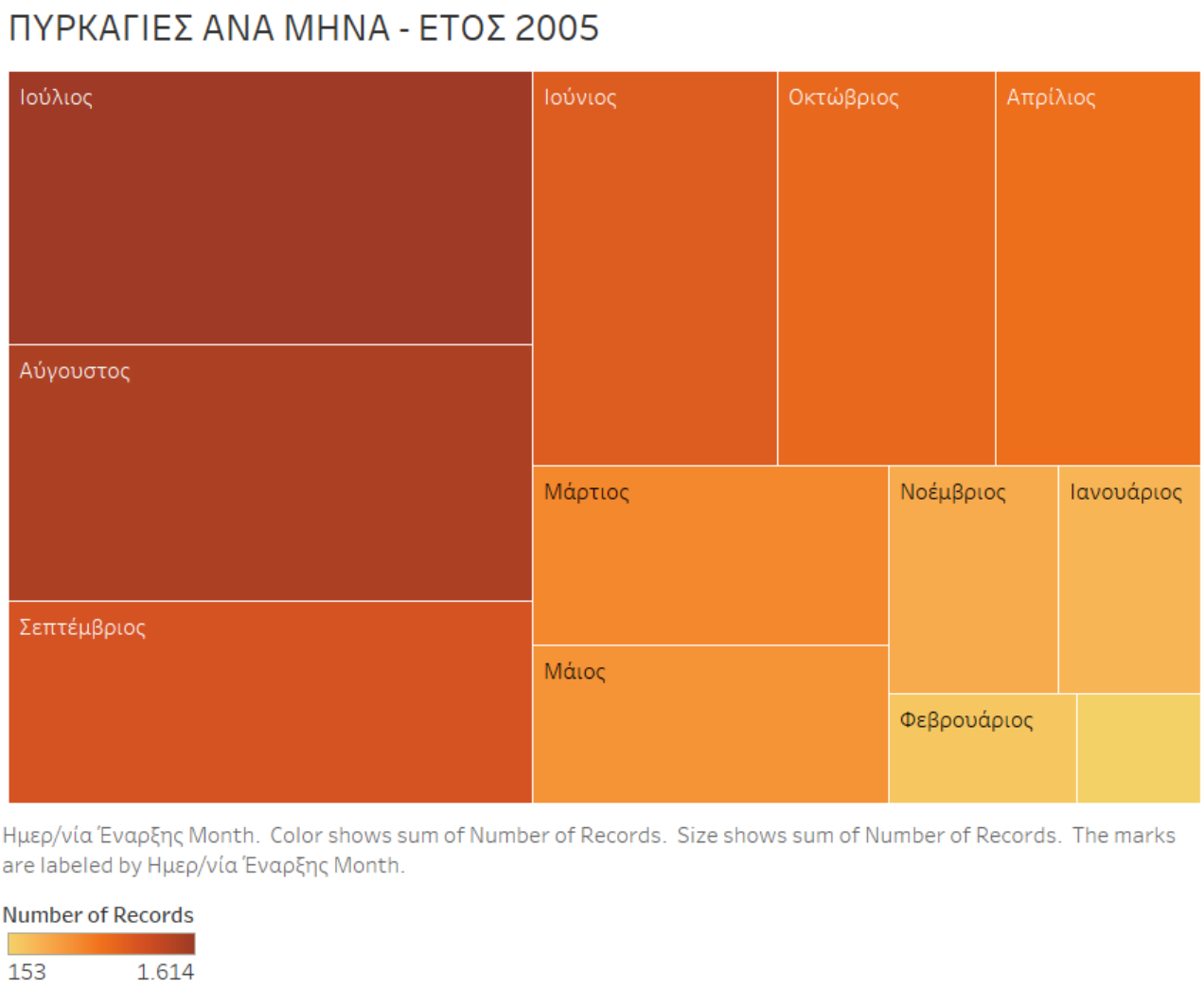
The aim of this assignment is to create some useful data visualizations using a given dataset, in order to present it with the most comprehensible way. For a visualization to be successful, it should have a clear purpose, include only the relevant context, use appropriate structure and has useful formatting.

The dataset contains information about fires in Greece. The subset of the dataset, which I used, contains the instances that represent the fires that took place in Greece in the year 2005 (for one of the visualizations, the whole dataset was used).

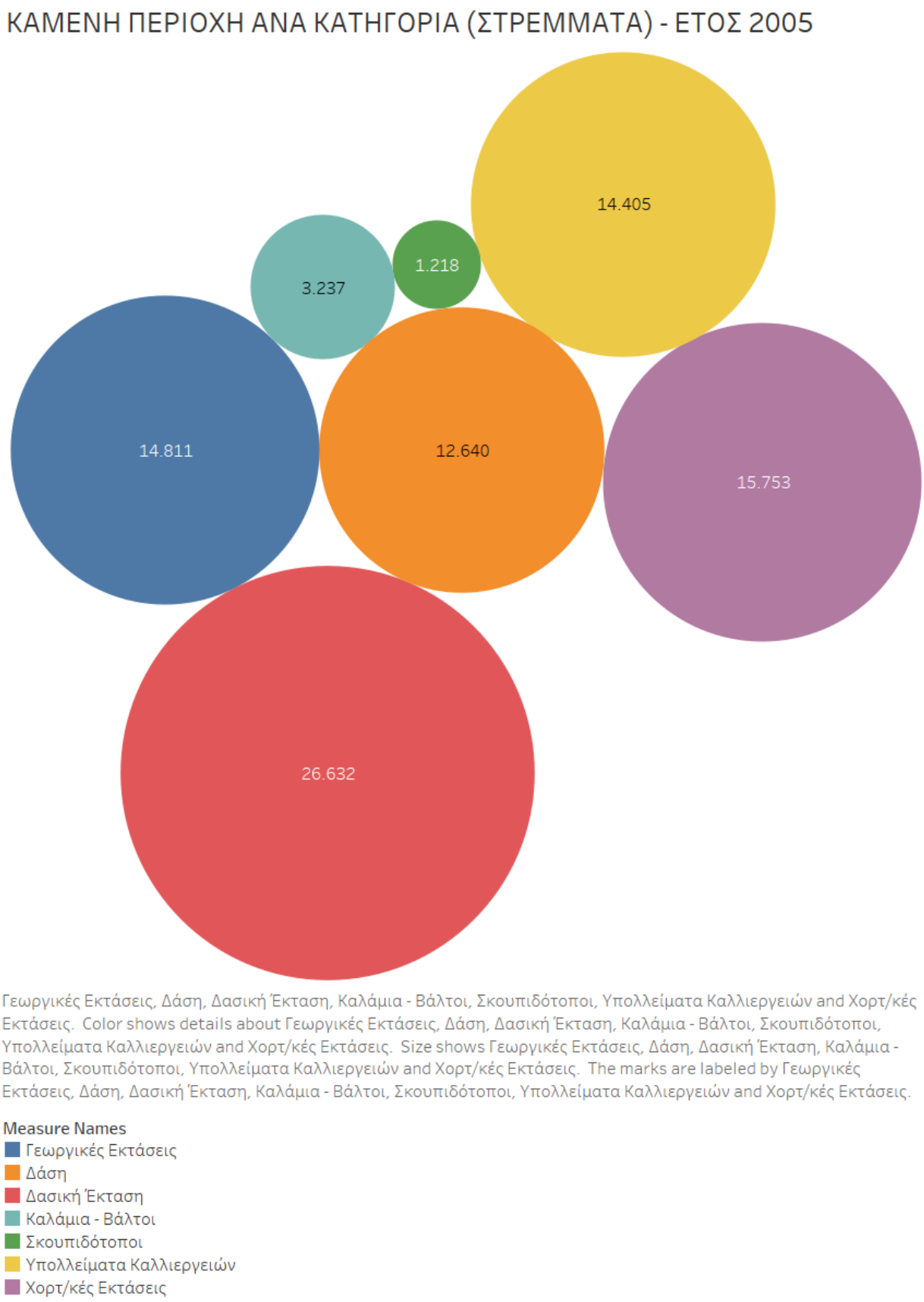
For the purpose of this assignment I used the Tableau 10.5 software and have made 9 visualizations.

Visualizations

I. The first visualization presents the number of fires that occurred in each month of the year 2005. As we can see, fires were more in the summer months, which makes sense as in the summer there are more factors that enhance the likelihood of fires, like exceed heat.



II. The second visualization presents the summation of the area (in acres) that was burnt, in relation to the type of the area. The size of the circles represents the size of the areas proportionally and the exact number of acres that were burnt is written in each circle. As we can conclude the type of area that was mostly burnt in 2005 was forest.



III. The third visualization presents the average time (in minutes) that was consumed for the fires to be extinguished, at each region of Greece. To create this visualization, firstly, I created a calculated field to combine the date and time of the start of the fire and the extinguishing of it. The functions used are shown below:

For the date and time of the start of the fire:
DATEADD('hour', DATEPART('hour', [Ώρα Έναρξης]), DATEADD('minute', DATEPART('minute', [Ώρα Έναρξης]), DATEADD('second', DATEPART('second', [Ώρα Έναρξης]), [Ημερ/νία Έναρξης])))

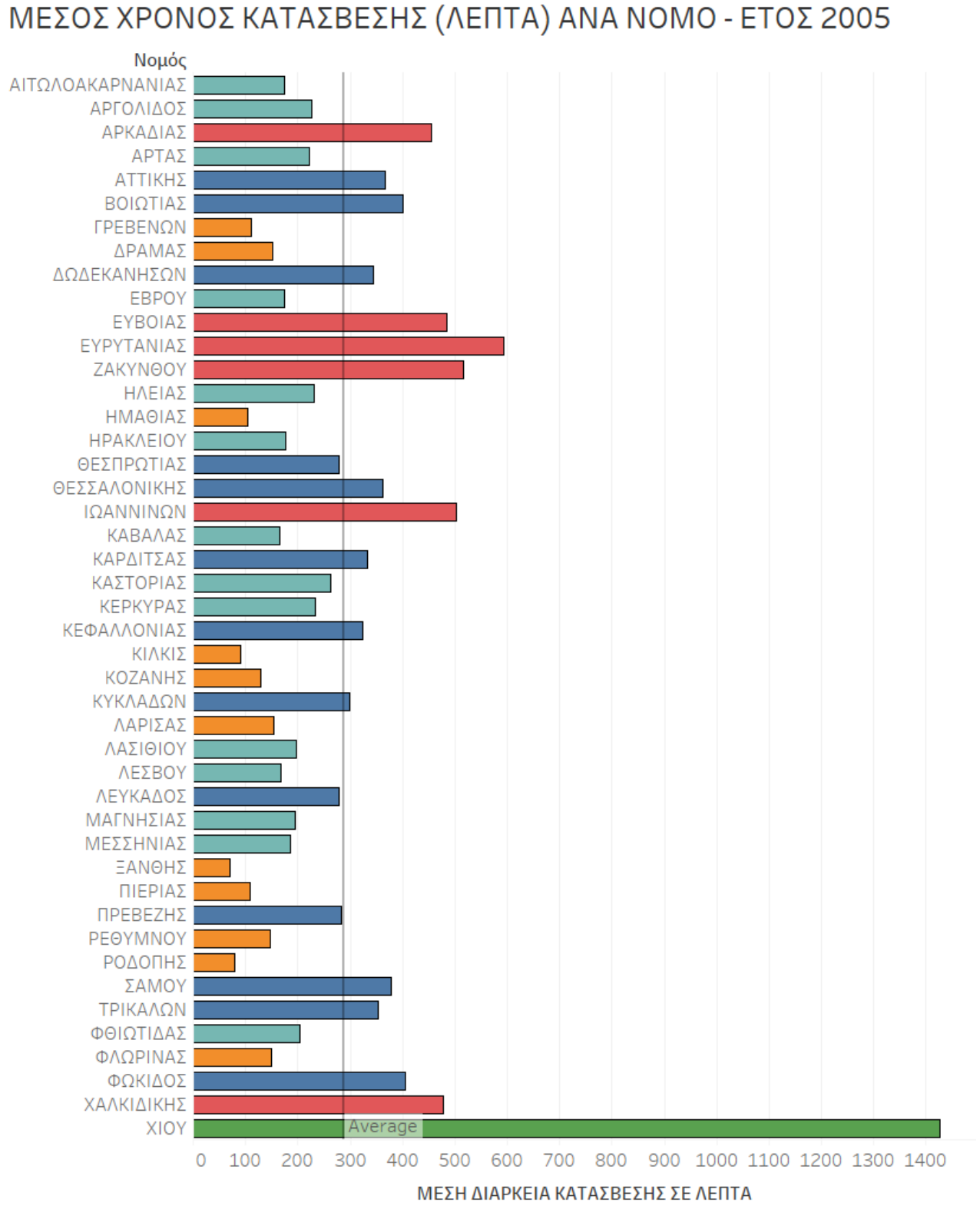
For the date and time of the extinguishing of the fire:
DATEADD('hour', DATEPART('hour', [Ώρα Κατάσβεσης]), DATEADD('minute', DATEPART('minute', [Ώρα Κατάσβεσης]), DATEADD('second', DATEPART('second', [Ώρα Κατάσβεσης]), [Ημερ/νία Κατασβεσης])))

I also created a calculated field that represents the difference of the two variables above to find the duration of the fire extinguish in minutes. For every visualization that this field has been used, the negative values were filtered out, because they are false values, because there cannot be a negative duration of time.

Duration in minutes:
DATEDIFF('minute', [Ημέρα και Ώρα Έναρξης],[Ημέρα και Ώρα Κατάσβεσης])

Then I created the plot using the average of the duration in minutes and Greece's regions. I have also used the analytics tab of the Tableau software to show the total average of the duration of the extinguishing and to create a color clustering among the regions.

As we can see from this visualization the region of "Χιος" seem to have the most average duration of fire extinguishing which might indicate a lot of things, such as the event of fires too big to be quickly taken out or the lack of personnel and fire fighters on the island, but these are things that need to be further investigated.

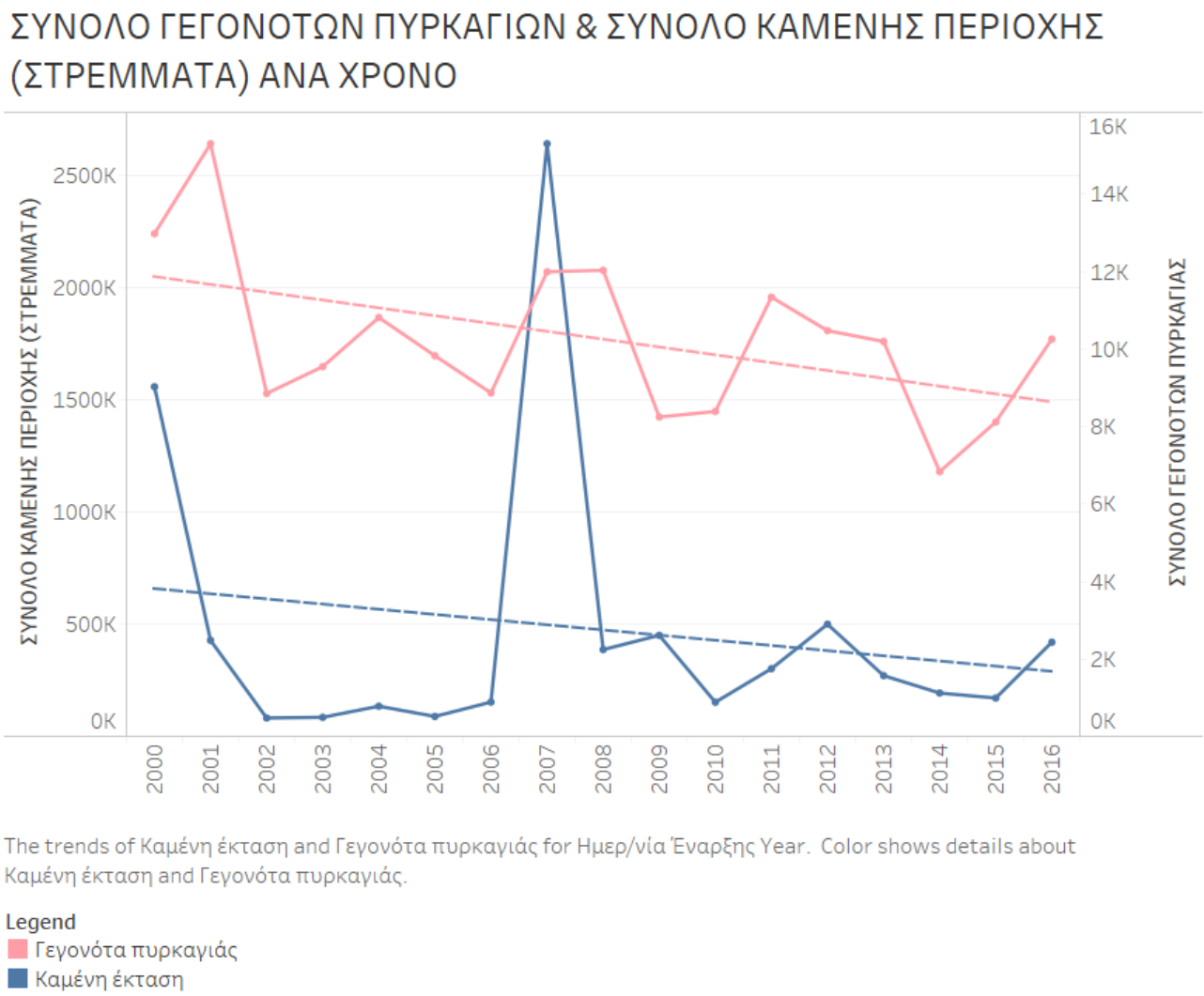


Average of Διάρκεια Κατάσβεσης for each Νομός. Color shows details about Clusters. The view is filtered on average of Διάρκεια Κατάσβεσης, which ranges from 0 to 1.425.

Clusters
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5

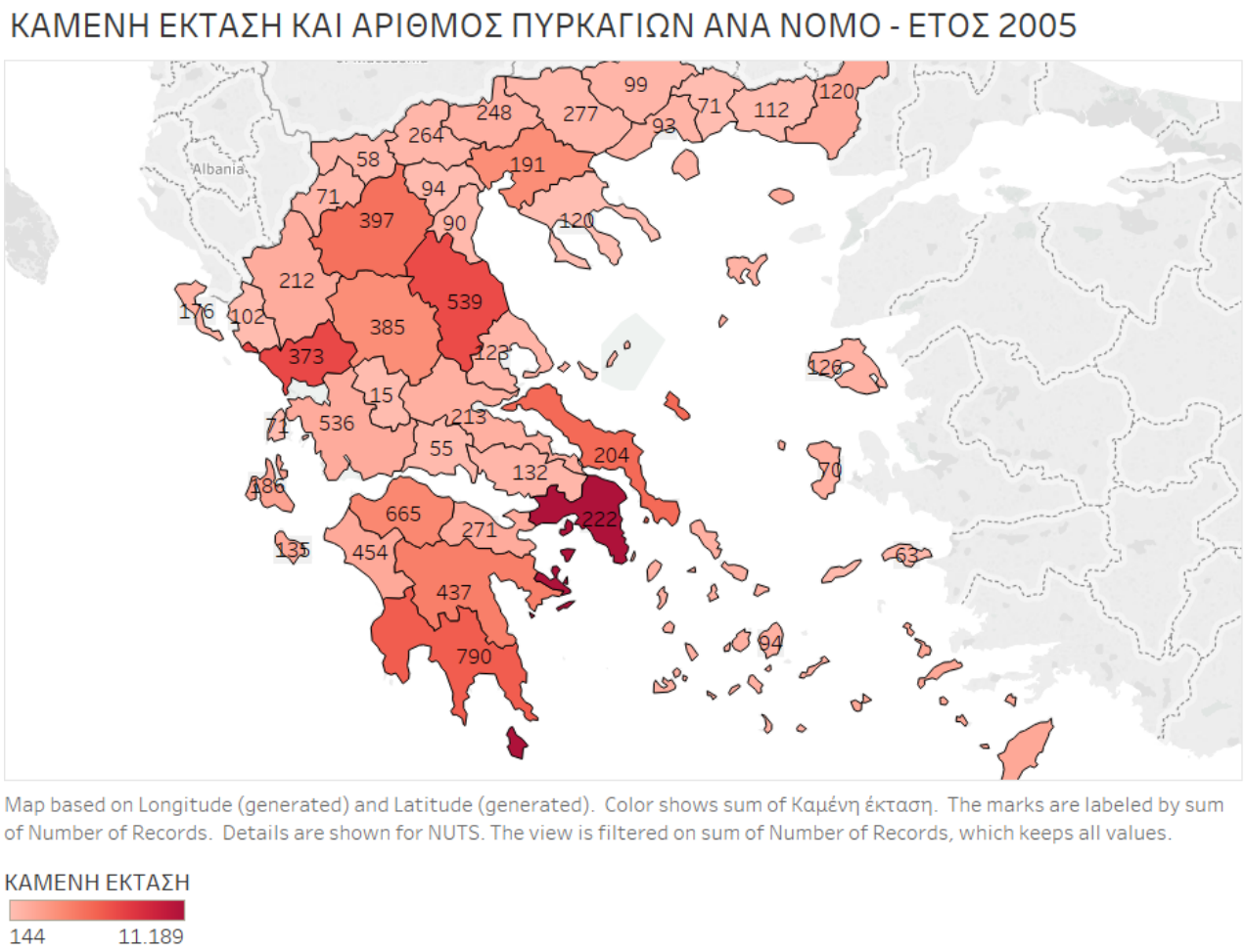
IV. The forth visualization is a dual axis plot that presents both the summation of the burnt area along with the number of fires that have occurred in each year (all years' data was used). I used another file that contained all the fire events throughout year 2000 till year 2016. I also used analytics to show the trend lines.

As we can see the two variables seem to have similar trend apart from the points of year 2007, in which the burnt area seems to be a lot more, in proportion to the number of events and to the burnt area of the other years. This might mean that there has been one or several big fires in Greece in 2007 that caused the burnt area increment, but this has to be further investigated.

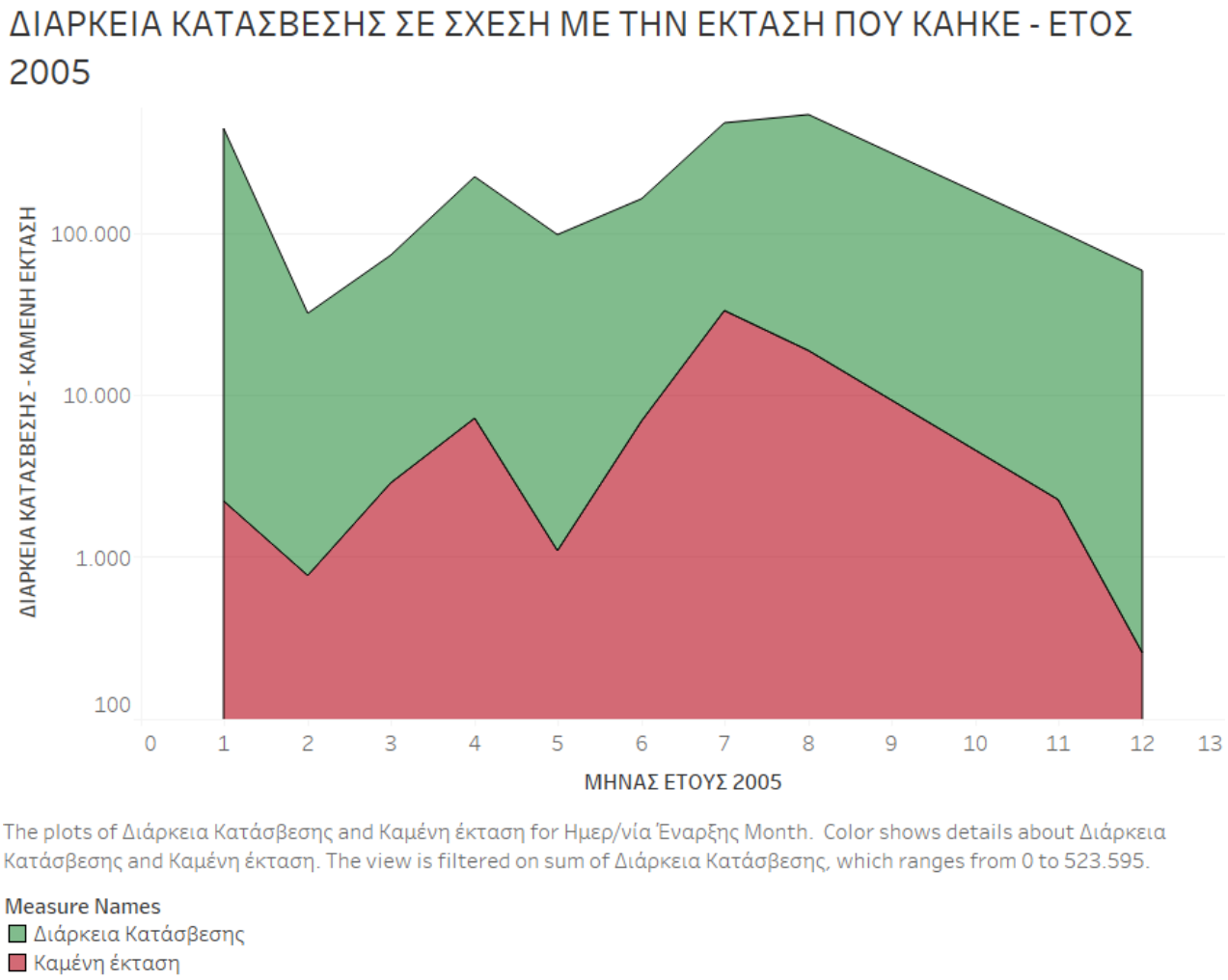


V. The fifth visualization is a colored map that presents the summation of the burnt area along with the number of events of fire for the year 2005. The color of the areas represents the size of the area in acres that was burnt and the number on each region represents the number of the fires that have occurred.

For this visualization, I have added one more attribute in the file of the fires of 2005, which contained the NUTS (Nomenclature of Territorial Units for Statistics) which is a code for the classification of territorial units. As long as this code was induced to the file, Tableau could recognize which region each instance refers to and generate the appropriate latitudes and longitudes that were used to load and create the map.



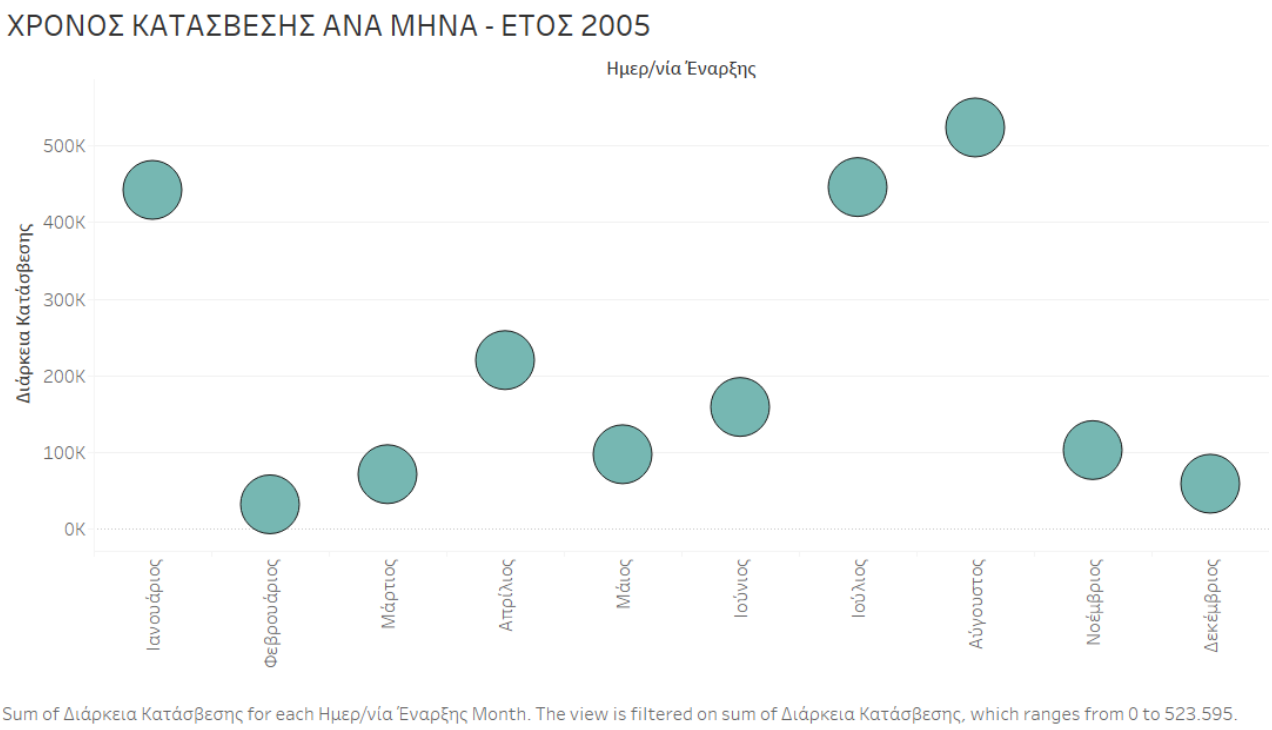
VI. The sixth visualization represents the duration of the fire extinguish along with the burnt area throughout the months of the year 2005. As we can see, the two areas of the plot follow the same pattern which means that there hasn't been a month in which there was an anomaly between those two variables.



The plots of Διάρκεια Κατάσβεσης and Καμένη έκταση for Ημερ/νία Έναρξης Month. Color shows details about Διάρκεια Κατάσβεσης and Καμένη έκταση. The view is filtered on sum of Διάρκεια Κατάσβεσης, which ranges from 0 to 523.595.

Measure Names
- Διάρκεια Κατάσβεσης
- Καμένη έκταση

VII. The seventh visualization represents the duration of the fires' extinguishing for each month of the year 2005. As we can see, in July and August the duration was a more than the other months and this is because the events of fires were more. In January, the duration is also high, even though the number of the events were a lot less than in July and August. This means we have to further search if the data for January is false or it there was a reason for this.

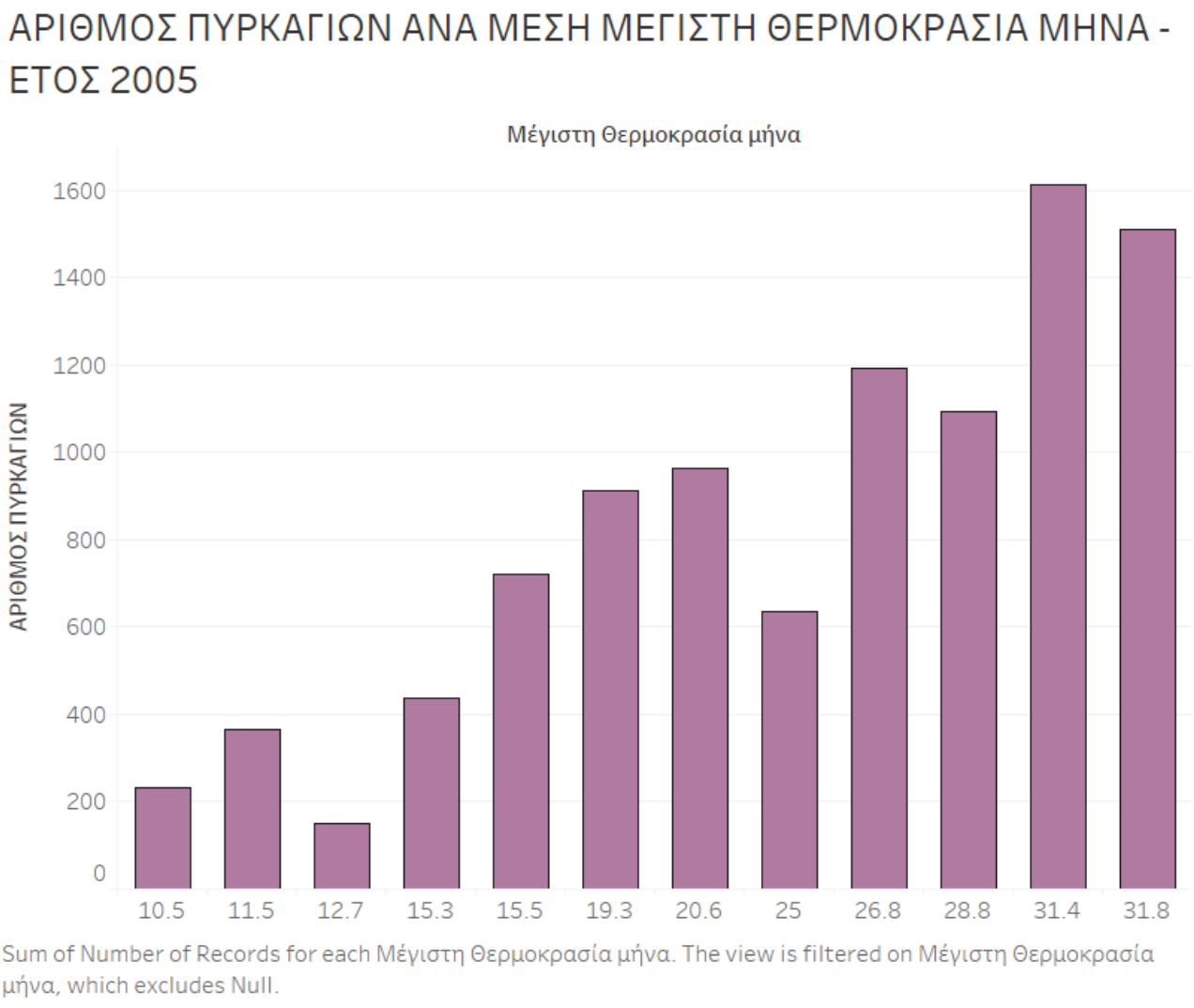


Sum of Διάρκεια Κατάσβεσης for each Ημερ/νία Έναρξης Month. The view is filtered on sum of Διάρκεια Κατάσβεσης, which ranges from 0 to 523.595.

VIII. The eighth visualization presents the number of events of fire in relation to the average maximum temperature in Greece.

There was not enough data regarding the temperature or other climate information easily obtained from the internet, so I used the following site: weatherdata.aws.gr/getmonthyears/result/10 to retrieve the average maximum temperature for each month which I added in the data.

As we can conclude from the bar chart below the number of fires are increased when the temperature gets higher, which makes sense as there is a lot more possibility for fire in the summer that the temperatures are higher than in the winter.



IX. The ninth visualization presents the number of the fires in relation to the moment of the day that they started (shown in minutes). As we can see, the events follow a Gaussian distribution with its mean around 4 p.m., which indicates that there is highest possibility for a fire to begin at around 4 p.m. than any other hour of the day.