Active Fires Monitoring Project

Sapienza University of Rome

Visual Analytics

Alexandru Sarpe, Leonardo Sarra

February 17, 2020

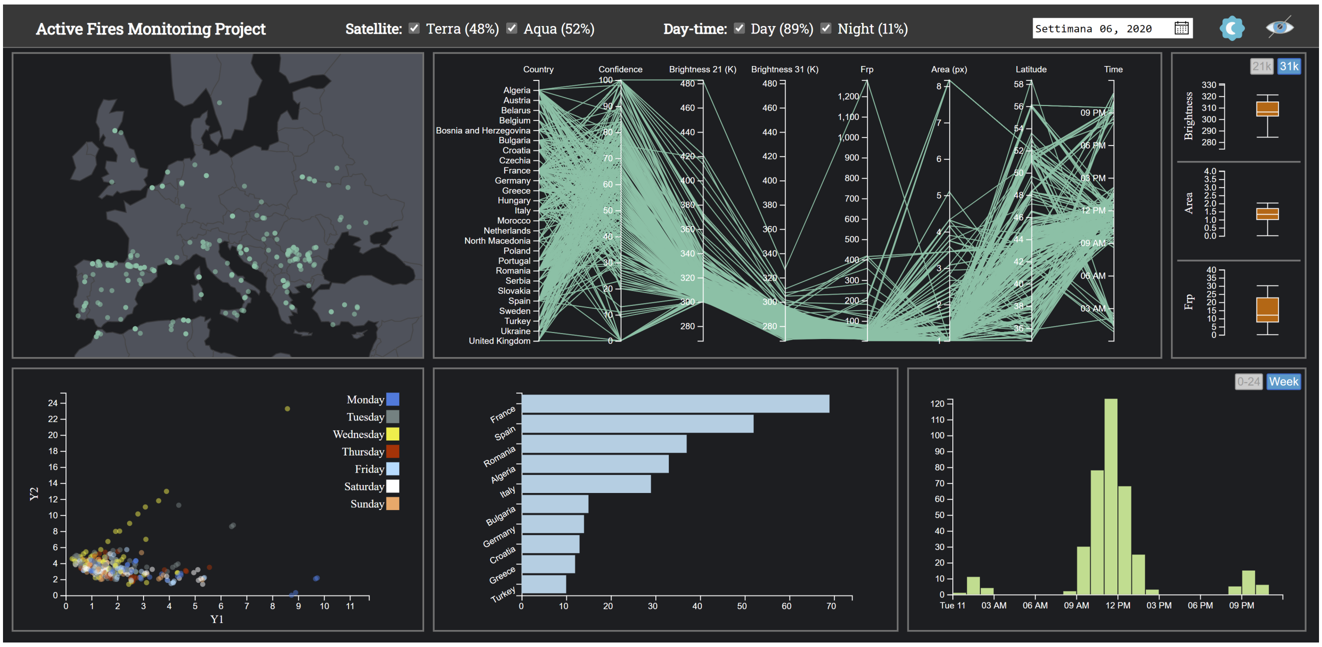

Figure 1: *Visualization with dark mode and colorblind-friendly mode activated*

Abstract

Fire events are a serious issue in Europe, with this problem in mind we developed a visual analytics tool to help users in analyzing fire events from the MODIS instruments on NASA's satellites.
The multiple interactive views of the project provide both low-level and high-level information and help the user in understanding fire events patterns and identifying outliers.

I. Introduction

Forest fires constitute a serious problem throughout Europe, data show that there are more than 65000 fires every year, burning approximately half a million of wildland and forest areas [1].
Usually, the fires are seen as a problem of just France, Italy, Greece, Spain and Portugal but in reality, the problem is widespread to every country.
Statistics about those fires are very important because they are fundamental in creating fire danger forecasts, monitoring of fire activity, assessment of damages due to fires and post-fire recovery are essential tools to support the services in charge of forest fire prevention and fighting. Because of the importance of the topic, the Europe Union started developing in 1998 the EFFIS (European Forest Fire Information System) [1] to keep updated and reliable information on wildland fires in Europe.
EFFIS uses NASA's FIRMS, an active fire detection system that works by analyzing data coming from multiple satellites at different distances of earth and with different instruments.

Starting from the same NASA dataset we developed our visual analytics application that also focuses on the European region by using the data from the MODIS instruments, all with the intent of helping the user in analyzing active fire events through different visualizations and interactions which expose both low-level and high-level information.

II. Dataset

The dataset used in this project is NASA's dataset of fire events from the MODIS instruments [2] (on Terra and Aqua satellites), only fire events from the European region are considered.
The NASA's dataset that is made publicly available covers only the last 7 days, in order to build a larger dataset we stored multiple weeks (from 10/12/19 to 17/02/20).
By doing so we obtained a total of 2428 tuples, moreover by having multiple weeks we can let the user select the week on which the visualization will be computed.
The dataset comes with different attributes:

- **Latitude** of the center of the fire event.
- **Longitude** of the center of the fire event.
- **Brightness T21**, Channel 21/22 brightness temperature of the fire pixel measured in Kelvin.
- **Brightness T31**, Channel 31 brightness temperature of the fire pixel measured in Kelvin.
- **Scan**, along-scan pixel size.
- **Track**, along-track pixel size.
- **Acquisition date** of the event by the satellite.
- **Acquisition time** (UTC) of the event by the satellite
- **Satellite**, Aqua or Terra (the two MODIS satellites).
- **Confidence** of the detection.
- **Version** that identifies the collection (e.g: MODIS collection 6) and the source of the data processing.
- **FRP**, Fire Radiative Power of the event.
- **DayNight**, express if it is daytime or nighttime at the time of the acquisition.

A. Dataset preprocessing

While the dataset comes with already many arguments out-of-the-box we opted to do some preprocessing to enhance it.
The preprocessing is executed once the dataset has been just downloaded and consists of adding two extra attributes. The first attribute that is added is "**Country**" which represents the country in which the event took place (computed through reverse-geocoding), the second, instead is "**Area**" which express the area of the fire event in the satellite image.
On top of that, we also perform PCA[3] on the various entries and store the principal components by appending them to the downloaded dataset.
The PCA is performed only after the features are standardized by removing the mean and scaling to unit variance, moreover, only a subset of the attributes ("Satellite" and "Version" are excluded) is used during the computation.
Those principal components will be later used in the scatterplot view to give a 2D representation of the weekly-events.
Note that PCA is executed only once and so the components will not change-over-time when the visualization is running.

After this phase the dataset to be loaded by the visualization tool which takes care of the representation.

III. Visualization

Different views are used in this project to build a complete tool for analyzing fire events.
Those views are built using D3.js and exposes a wide range of information.

A. Parallel Coordinates Graph

The main visualization element is the parallel coordinates graph. In this chart, the user can observe different attributes of the dataset in a single plot. The axes represent the following attributes:
Country, Confidence, Brightness 21k, Brightness 31k, Frp, Area, Time.
This visualization coordinates each other graph giving the possibility to the user to impose multiple filters on the dataset interacting with it.
Each axis of the graph constitutes a filter on the dataset for its particular dimension. The filtered data is shared among all the other graphs in the complete visualization.
Depending on the selection lines can take different colors to express to the user that those events got filtered out.
The domain of the axes is computed at run-time, most of the time is just a matter of taking the minimum and the maximum value of the attribute, but other times, like for brightness and brightness31k we opted to use a domain that is the most inclusive by taking into account the values from both attributes at the same time. We opted for this approach for these particular attributes as they are usually used in tandem because their difference expresses additional information that helps researchers in classifying fire events. If we didn't adopt this approach their difference would have been hard to compute visually because the scale would have been different.

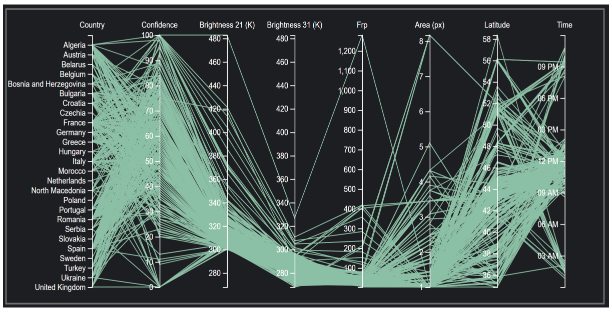

Figure 2: *Parallel graph*

B. Scatterplot

The scatterplot chart is used to represent the elements of the dataset after a dimensionality reduction on two dimensions.
In particular, this graph represents the first two components of the execution of PCA. Each point of the scatterplot represents a fire event, projected on the first two principal components.
By observing this chart the user can find similarities over the events looking at the distance between the points and searching for clusters.
This chart also allows to find outliers on the dataset.
In order to add more information on the graph, we added a chromatic encoding of the points, representing the days of the week in which the event happened.
For a deeper analysis, the user can select a zone of interest on the graph, this selection is highlighted on the parallel coordinates graph and the geographical map.

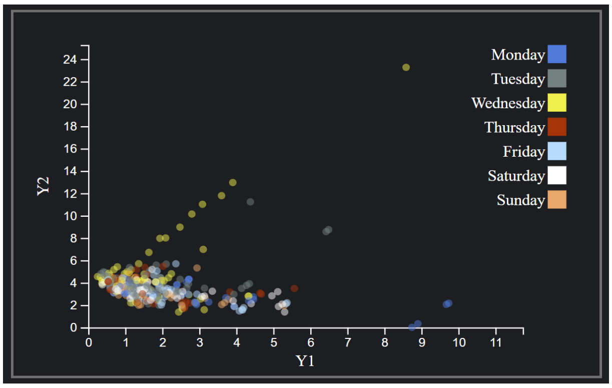

Figure 3: *Scatterplot*

C. Geographic Map

The geolocation of the events is represented by a geographic map in which events are placed by their latitude and longitude.
It is a very simple view but it helps the users by giving them the position of the events which otherwise is hard to get from the raw data.
In the geomap events are represented by semi-transparent circles which can potentially change color when the events are selected in the scatterplot.
The user also has the ability to pan and zoom which comes in handy when there is a cluster of events that otherwise are hard to distinguish.

Figure 4: *Geographic map*

D. Bar chart

In the project, we also used a bar chart to express the countries where most of the events occurred.
In particular, the bar chart represents a top-10 because depending on the week there could be 20+ countries with fire events, most of them with an insignificant number of events, therefore, representing all the countries represents a big problem.
Using a top-10 enabled us to make a smaller view which still contains all the information relevant for the user.
For this view, the top-10 is generated at run-time and is recomputed based on the interaction of the user with the other views and filtering options.

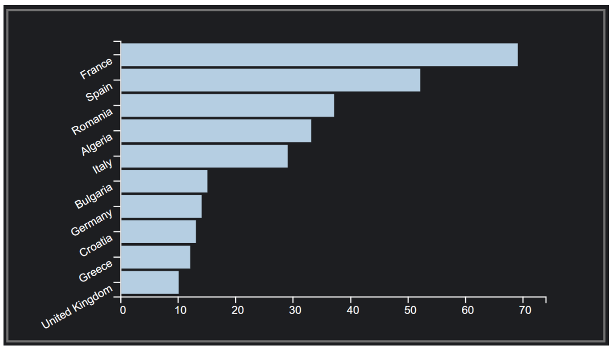

Figure 5: *Bar chart*

E. Histogram

The distribution of the fire events over time is represented using an histogram.
Through the histogram, the user is able to distinguish time-related patterns across the week.
For the histogram two different scales for the x-axis are available: hourly and weekly.
The user can change at will between the different scales by pressing a button.
The hourly scale is useful to find the hours in which most of the events took place while the weekly one comes in handy when finding the days with the most fire events.
This view, like the others, requires some level of computation done at run-time every time the user interacts with the other views and filtering options, in particular, we have to compute the number of events for each bin of the histogram.

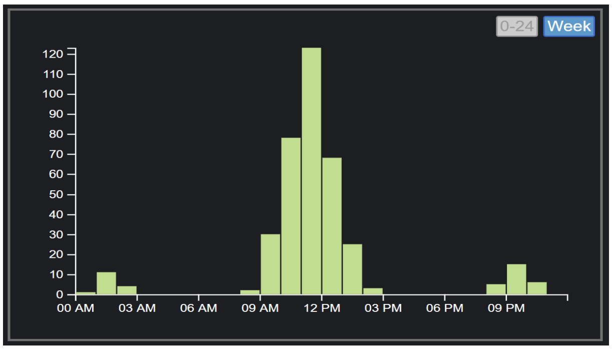

Figure 6: *Histogram*

F. Box plots

In other to aid the user in understanding the distribution of the attributes values shown in the parallel graph we also added multiple vertical box plots to the project.
Box plots are used to graphically depicting brightness, Frp, and area through their quantiles.
Of these three box plots, the one for the brightness also supports the possibility of changing the channel (21/31) through the interaction with a button.
The quantiles are computed at run-time and change over time depending on the user interactions on the other views.

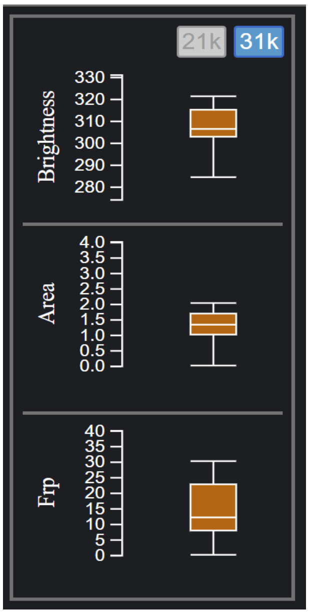

Figure 7: *Brightness, Area and Frp box plots*

G. Header

While not being a chart or a graph the header too plays an important role in the visualization.
The header comes with filtering and week selection capabilities, it also gives the possibility of enabling the dark mode and colorblind-friendly mode.
The check-boxes plays an important role because they let the user filter based on the satellite and the day-time while still providing a numerical proportion of the data that fall in their category.
Those proportions are computed at run-time and changes only when the user changes the week of analysis.

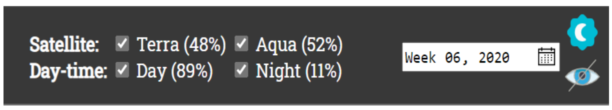

Figure 8: *Header*

H. Dark mode and colorblind-friendly mode

This project comes with two modes for the visualization: dark mode and colorblind-friendly mode. Both are activated through a button in the header of the page and the changes take place immediately without the need of refreshing the page.
The dark mode is useful to all the users it helps the user in focusing on the visualization because thanks to dark background the details on the views pop up more and are easier to read.
The dark mode is not just a matter of changing the background, all the views that we already presented

before will adapt to the new mode by choosing new colors for their visual element but also the text, the axes, the borders will change to a lighter color too.
The colorblind-friendly mode works similarly but instead of changing the visual elements to a darker tone it focuses more on changing the color of the visual elements to help people with red-green color blindness.
Thanks to this mode everyone can easily read the views without being misled by their colors.

Note that both the modes can be active at the same time, this means that we have a total of 4 color schemes to cover all the possible combinations.
In particular dark mode requires lighter colors while the color-friendly palette has a range of colors that is way more limited than a normal one, therefore, depending on the combination some adjustments were required to not create a negative effect on the readability of the visualization.

IV. Interactions and analytics

The project includes multiple possibilities of interactions.
These interactions allow the user to perform a deeper analysis of the dataset.
We already introduced them when talking about the views, here is a recap:

- **Parallel Coordinates**: allows brush selections on each axis, which performs multiple filters on the data and modifies the content of the other visualizations.
- **Scatterplot**: allows brush selection on the events, this selection is highlighted on the geographical map and on the parallel coordinates chart.
- **Geographic Map**: allows pan and zoom.
- **Histogram**: allows the selection of two different domains for the x-axis, giving the hourly and weekly distribution of the events.
- **Box plot**: allows the selection of the two different channels for the brightness value.
- **Header**: performs filter on the dataset on the two binary values, satellite and day/night. It allows also the user to filter the dataset over the weeks and to change the enable/disable the dark mode and the colorblindness-friendly mode.

Some of those interactions, like the one on the geomap, the histogram or the box plot doesn't influence the other views, they are intended to make the information easier to read or change drastically the information that is provided by the view the user interact with.
Then we have interactions that end up updating other views, those interactions are useful because they give the possibility for a fine-grained analysis and potentially expose additional knowledge.
The project comes with three interactions that fall in this category:

- **Header → ***: The usage of the week selector, the checkboxes will trigger an update of the views in all the other views in order to change the data elements that are shown, instead, in case the color-friendly mode or the dark mode buttons are triggered the other views will update only the colors of their visual elements.
- **Parallel graph → * (except header)**: Brushing on one of the multiple axes of the parallel graph triggers an update in all the views that are going to show only a subset of the data.
  The parallel graph itself changes the color of the elements that got filtered out.
- **Scatterplot → Scatterplot, Geomap, Parallel graph**: Brushing on the scatterplot will highlight with a different color the elements on the geomap and the parallel graph, moreover, the elements in the selection rectangle are also highlighted (opacity increased, black border added).

Of these interactions we have the ones between the parallel and the scatterplot that creates a coordination in both ways between the views.

The analytics part can be subdivided in two categories. The first one is the most evident, and includes three elements:

- **Top 10 countries on bar chart**: this first analytics is computed on the bar chart graph, at run-time the 10 countries with most fire events are computed and this result is ordered in a data structure. This data is then displayed on the graph.
- **Percentage of events for Terra/Aqua, day/night**: at run-time we compute for each one of the two binary values the percentage of events present. This values are then displayed in the header.
- **Quantiles of box plots**: The quantiles of the box plots are computed at runtime, these values are the principal component of the box plot chart.

The second category includes less evident computations, for example when the week is changed the axis of all the views, or the number of the bins in the histogram have to be calculated again.

V. Reasons behind our approach

Our main goal was to create a compact visualization which helps the user to perform analysis on fire events, therefore, we chose the views with their utility in mind.
For example, by using the parallel chart the user can isolate different events in different ranges applying fil-

ters on the axis, combining the possible ranges of each value and obtaining a fine-grained selection on the events in the dataset.
The analysis of the events can be subdivided also in macro ranges using the checkboxes in the header or selecting different weeks to observe the differences over the weeks in the month.
The scatterplot chart, together with box plots and parallel coordinates chart have two roles, help the user to identify cluster of fire events while making the outliers easily identifiable.
The geographical map provides an easy location of the events, encoding them on the chart with their geographical position. The histogram chart, instead, offers high-level information about the week events, allowing the analysis over two different domains and offering additional information as the distribution of the fire events over the day's hours.
Finally, the bar chart provides additional information over the dataset as a rank of the countries with most fire events over the week. The coordination among the graphs helps in focusing the analysis on a specific group of data. Moreover the different animations of the views help in giving a visual feedback of the changes when filtering/selection is used.

VI. Conclusion

We believe that the end result represents a good tool for the analysis and monitoring of the different fire events in Europe.
The tool comes in handy because it offers the possibility of making fine-grained selections of data, getting instant visual information of the selection and additional insight information obtained with different analytics.

In the future, different visualization approaches could be developed and compared with the intent of finding the approach that is the most effective for representing the data.
But in order to make a comparison possible, there is the need for feedback from the researches that deals with those type of data with a daily-basis.
An interesting alternative approach would be to remove the parallel graph and merge its filtering capabilities with the geographic map.
Otherwise, it would be interesting reworking the existing views using WebGL, reducing the load on the DOM, and analyzing the performance improvements, this could virtually open up the possibility of adding the coverage for other continents.

References

[1] EFFIS - https://effis.jrc.ec.europa.eu/

[2] NASA FIRMS Dataset - https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms

[3] PCA implementation - https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

[4] The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios
Steve Wexler, Jeffrey Shaffer, Andy Cotgreave

[5] Color palettes for color blindness,
Martin Krzywinski - http://mkweb.bcgsc.ca/colorblind/

[6] ColorBrewer2 - http://colorbrewer2.org/