Visual Analytics
VICrime

Francesco Artibani, Flavia Ferranti and Valerio Longo

2021, February

Contents

1 Introduction

In this paper we describe the various visualization techniques realized to assist crime analysis.
We decided to analyze crimes because they affect the quality of life, economic growth and reputation of a nation. Furthermore, they impact over important life's decision, like even moving to a new place, avoiding dangerous areas. Due to all these reasons it's important to find a way to reduce every kind of infractions.

The aim of *VICrime*, our visualization tool, is exactly this one: assist users into studying the behaviours of crimes in specific Italian areas, identifying hotspots, crime patterns and their evolution over time.
In order to do that we needed to find a way to represent information such that they could be easily visualized, analyzed and understood, supporting Italian authorities into monitoring the situation and, if necessary, conceiving plan of actions.

Therefore, we realized three types of visualization:

- Crime mapping is used to understand the geographic crime distribution, identifying hotspots, and highlight how much the demographic incidence can affect the rate of felonies. Once a particular area has been detected, a more precise study can be carried on by analyzing subareas.
- Parallel coordinates are specifically addressed to find patterns, analyzing every crime composing a territory, reporting also their recurrence rate. In addition, this view is very useful for detecting evolution over time.
- Multidimensional scaling projection was inserted for having a tangible situation of the differences existing between areas without having to compare each location crime by crime.

The possibility of having interaction between the user and the visualizations allows to better inspect some data, while the coordination among all views is fundamental to see each possible projection of the information, for the purpose of studying them according different point of view.

The work has been realized following the steps of Visual Analysis, that we will see in the following chapter, starting from data gathering.

2 Related work

Throughout the development of the project, to have an idea of what we could do, we analyzed the literature, studying papers used both for the criminal domain and for analyzing statistical data.
For what concern the views used we took inspiration from **CrimeVis: An Interactive Visualization System for Analyzing Crime Data in the State of Rio de Janeiro (2017)**. They investigated data in a period of time going from 2003 to 2015 stating that understanding this type of problem is difficult because violence varies very much both geographically and over time, in addition, data may be inconsistent or unrelated. *CrimeVis* is mainly based on map view and parallel coordinates to visualize the crimes of the Police Department that exposed them. Moreover, criminal data can be contrasted with socioeconomic data related to the region associated to each department. Their idea is that views can be asked on demand, only if needed, to not weight down the vision. The tool offers three clustering techinques, containing MDS, with the main scope of finding similarities and patterns. Then, is also possible to see a 3D visualization of the parallel coordinates, to highlight something that could be missed in the 2D case, making the user choose for the attribute of the Z axis.
This paper was definitely an inspiration for us, so we used it as a guidelines, with the proper changes. First of all, we decided to focus on the whole country taking the available data, without having the possibility to access the denounces from each police station. The idea of map and parallel views were considered optimal for our case, to represent patterns in crimes, but the use of a 3D visualization, instead, was perceived as difficult to interact with and redundant. In fact, since we concentrate purely on the crimes situation, rather than on the reason of happening, then the use of a third axis will loose his purpose. The three kind of clustering techniques could be too much as well, so we opted for only one algorithm but with three different outcome based on the additional information required by the user.

Another paper that we considered was **CrimAnalyzer: Understanding Crime Patterns in Sao Paulo (2015)**. The goal is to understand the dynamics of crimes over the city by analyzing the variation of crime patterns over space and time. The analysis is done only over the city focusing on some particular crimes rather than on all possible ones. Here the idea is to use some visualization techniques that are considered not trivial for crime analysis.
As before, user visually query the data set by interacting with a map but we also have an additional map to represents hotspots giving information about number of crimes, temporal rate and giving more relevance to hotspots having a larger number of delinquencies. Then among these non usual choice there is a ranking type view (to understand crime type evolution, crime type ranking and the number of crimes in each time slice) and a radial type view where the 12 bar depict the months rate of each year.
This work was useful especially for the purpose to define what an hotspot is, with the intuition of weights a specific crime with the penalty gave from the italian government; this parameter permits the MDS to be more clear in some case studies. Following the same idea, the fact to permit the user to balance the crimes all over the population came also from here. For what concern the visualizations, even if they were considered useful, we thought that they could be a little less intuitive for users that are not experts, while we wanted to create an environment adaptable to more situations.

Then we searched for work explicitly done on our dataset finding a report done by a Padova University student: **Un'analisi sul rapporto tra criminalità e benessere in Italia**. This study was conducted for the statistic faculty aiming to find a relationship between crime rates and environmental wellness. The data was taken as well from ISTAT and each region crimes rate was displayed in a tabular form, using some color to encode the frequency. Red was used for low value and green for high. The problem with this table was that the reader find hard to compare all these values and, although the color may help, the missing legend makes it harder to understand. Although, the writer tried to reduce the values of numbers by considering crimes rate over 10000 habitants, thing that we did as well. Then there was present some scatter plot to try in finding correlation among numbers of crimes and factors like income and unemployment. The writer had to explain and comment every result, while some interactive view could have made everything easier. In the end, the analyst itself stated that understanding the reason behind diffusion of crimes are very hard to find because even family fights in childhood, friendship or work dissatisfaction could led a person to commit a crime, and this is why we did not treat this aspect.

Finally, to understand what kind of analysis could be made with our type of visualizations we decided to search for other tools exploiting them. We thought that map and clustering techniques could be integrated well especially when considering geographically distant place but with common attributes. The clustering techniques is also well linked with the parallel coordinates that gives the reasons why some places are similar or not. From this we concluded that, any sort of anlysis concerning various aspects characterizing territories could benefit from this views. For example we could study the efficiency of public transport, the economic situation, the level of instruction and more in general the quality of life. Following this intuition we searched for a tool able to quantify this property. The result was the following paper **Design Study: Using Multiple Coordinated Views to Analyze Geo-referenced High-dimensional Datasets** which aimed to encode the attractiveness of Swiss cities considering attributes going from economic field, to social and environmental one. *City'O'Scope* was designed for help the user understand common city ranking given by a country newspaper. The simplicity and effectiveness of map combined to parallel coordinates makes this tool extremely appreciated. Additional interesting work came from **CrimeVis: Visualizing Vancouver Crime** This work shares the same objectives: Represent the crime changes over the years and how a crime rates changes in different locations. Besides, they use very different views as an Isotype plot coordinated to a neighborhood selection map and a gridmap where in each grid it's plotted the amount of changes of crimes (in that territory) got from two different ranges of time. In the end we must cite some works as

- **NYC crime map**, a governmental visualization tool composed of 3 switchable kind of maps (Precinct/Crime Location/Heat)

- **CrimeMapping**, TriTech Software System to help US community. Provides the exact address of the crime reported by most of all US Police departments using a google map.
- **Indice della Criminalità**, Elaboration of similar data from "Il sole 24 ore" journal. It provides histograms, density map and a radial view of italian Provinces.

3 Dataset

In order to try to be consistent, it has been decided to use as dataset all the *denounced* crimes reported by the ISTAT in the years between 2012 and 2019, the latest years available. Moreover, the dataset has been preprocessed due to the amount of too specific data. In fact, some crimes, as rapes, homicides, damage and so on, were too fine grained, because they were split in "subcrimes" based on the purpose (rapes against adult, rapes against minor and so on). So it has been decided to sum up all these subcrimes in order to have a more clear and general dataset. Finally, to weight the crimes over the population, we retrieve the number of inhabitants of each location, in the interested years, and we added this information to our dataset.
The output of this simple edit of the source dataset is a csv file containing 1073 rows (representing all italian regions and provinces in the year interval of 2012-2019) and 35 columns (representing the crimes, the name of the territory and the sum), with the total of a AS index of 37.555. Here a snippet:

```
territorio;anno;popolazione;strage;tentati_omicidi;omicidi;[.....];altri_delitti;totale
//.....
Piemonte;2012;4357663;1;84;142;[.....];31621;243077
Piemonte;2013;4374052;1;69;112;[.....];34851;248366
Piemonte;2014;4436798;0;75;123;[.....];32094;240892
Piemonte;2015;4424467;0;66;134;[.....];30359;227047
Piemonte;2016;4404246;3;57;124;[.....];29684;207885
Piemonte;2017;4392526;1;73;122;[.....];30207;193783
Piemonte;2018;4375865;0;67;158;[.....];29369;184594
Piemonte;2019;4328565;1;73;128;[.....];30873;180478
//......
```

Later on, to have a better interface, we decided to translate all crimes name in English, so that the tool could be as international as possible. In addition, the MDS visualization, in order to retrieve the danger coefficients, exploits the convicted criminals data (still reported by ISTAT) containing for each crime the detention time that Italy's government had adopted in the last 10 years.

Reati	fino a un r	1 mese - 3	3 mesi - 6	6 mesi - 12	1 anno - 2	2 anni - 3	3 anni - 5	5 anni - 10	10 anni e	Totale cor	Rapporto	Coefficien	Coefficien	c. totale x
strage	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tentati omicidi	0	0	0	1	2	16	34	25	4	82	0,004927	0,37	0,00182	0,182
omicidi	0	0	3	11	5	0	3	29	44	95	0,005708	0,64	0,00364	0,364
percosse	4	11	2	1	0	0	0	0	0	18	0,001082	0,01	0,00001	0,001
lesioni dolose	43	253	315	202	65	9	2	1	0	890	0,053476	0,03	0,0016	0,16
sequestri di persona	0	0	0	0	0	0	0	2	0	2	0,00012	0,5	0,00006	0,006
violenze sessuali	0	2	7	19	94	144	49	10	0	325	0,019528	0,18	0,00351	0,351
atti sessuali con minorenne	0	0	1	2	11	13	8	0	0	35	0,002103	0,18	0,00038	0,038
corruzione di minorenne	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sfruttamento e favoreggiamento della prostit	0	0	0	1	1	1	0	0	0	3	0,00018	0,12	0,00002	0,002
pornografia minorile e detenzione di materiali	1	3	0	2	11	13	8	0	0	38	0,002283	0,16	0,00037	0,037
furti	158	1331	3045	2143	394	23	6	3	0	7103	0,426786	0,03	0,0128	1,28
rapine	3	14	84	626	1543	884	242	17	0	3413	0,205071	0,13	0,02666	2,666
estorsioni	0	0	8	62	201	142	34	3	0	450	0,027038	0,14	0,00379	0,379
truffe e frodi informatiche	1	7	11	10	4	0	0	0	0	33	0,001983	0,04	0,00008	0,008
delitti informatici	0	0	0	0	0	0	0	0	0	0	0	0	0	0
contraffazione di marchi e prodotti industriali	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ricettazione	32	199	418	550	402	48	5	0	0	1654	0,099381	0,05	0,00497	0,497
riciclaggio e impiego di denaro, beni o utilità	0	3	16	31	28	6	1	0	0	85	0,005107	0,07	0,00036	0,036
usura	0	0	0	0	4	0	0	0	0	4	0,00024	0,1	0,00002	0,002
incendi	0	0	4	10	14	0	2	0	0	30	0,001803	0,09	0,00016	0,016
normativa sugli stupefacenti	19	50	484	762	614	307	102	26	1	2365	0,142102	0,09	0,01279	1,279
associazione per delinquere	0	0	0	0	1	2	1	1	0	5	0,0003	0,26	0,00008	0,008
associazione di tipo mafioso	0	0	0	3	4	2	2	0	2	13	0,000781	0,27	0,00021	0,021
contrabbando	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Totale condanne per fascia di pena	261	1873	4398	4436	3398	1610	499	117	51	16643	1	4,15	0,07331	7,333

Finally, the Map view use two others ISTAT datasets for draw the Italian maps divided for regions and for provinces. The downloaded file is a shapefile so for the use, it needs a pre-processing for the transformation in a GeoJson file.

4 Visualization Techniques

After this data manipulation phase we started the visualization one through the following three views.

4.1 Italian Map

Italian map allow to visualize data through a color scale. For the color scale decision we have used the tool "colorBrewer", which permit to choose the best one to be used. Each colour represent the range of number of crimes given the number of infractions on each territory. Every selected area has a fill color weighted over the number of crimes in that place. In addition to filling, the color of the borders of each territory is meaningful. This characteristic permit to understand if the fill of territory is changed after the last modification on the menu, that brings to a re-computation. The border will be black if the fill of territory has not changed, while it will be green if the fill has changed. The legend of the Italian Map permit to understand the number of crimes given a color, indeed it relate a color with respect to the amount of felony happening. A territory can be selected clicking on it and is colored with the color correspondent to the range in which falls its number of crimes. Hovering the mouse on the selected territory can be obtained its name, the number of crimes, the number of crimes each 10000 inhabitants, the population and the area (km^2). The map provides zoom for a better selection and analysis of the small territories. In this visualization is possible to filter all territories of a certain color, this function is implemented on the legend, in fact, by clicking a color on it only territories of that class are selected. Having data of the population and territory area, we have decided to add a further tool for visualize the population density on the map. The density is encoded in a categorical scale of 5 classes. The value that distinguish a class from the others is the size of a circle (lower density means smaller circle). This information can be shown or not using a checkbox over the map. To explain the mean of the circles there is a legend that is shown if the checkbox is checked. Every density values is shown going over the circle of a specific territory and to see only territories belonging to a certain class we can simply click on it on the legend. The entire map is interactive with all others visualizations.

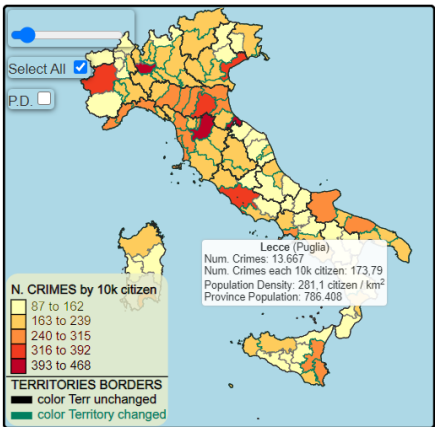


Figure 1: Italian Map with mouseover on a province

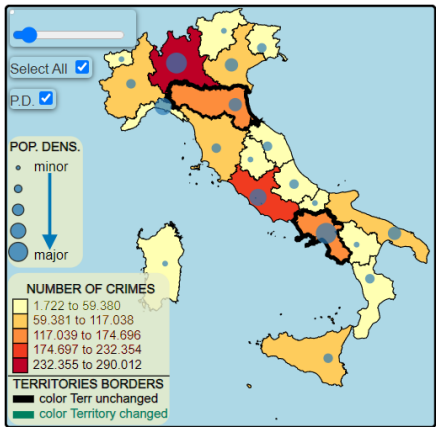
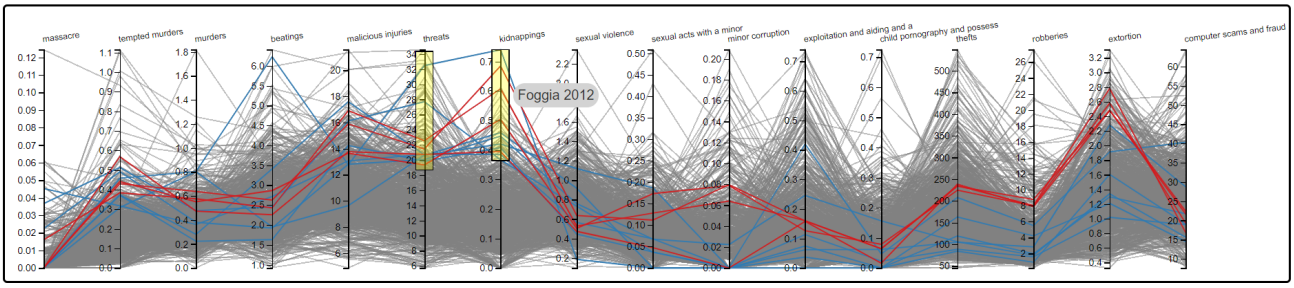


Figure 2: Italian Map with mouseover on legend

4.2 Parallel coordinates

Parallel coordinates is a technique for visualizing multidimensional data through parallel axes in a 2D chart. In our case, each axis represents a crime, while a polyline represents a territory (region or province) that intersects each axis at its corresponding crime value. However, parallel coordinate charts have some limitations. Even with an average-sized data set, it may suffer from overplotting, making it difficult to identify characteristics, trends or patterns. One way to alleviate this limitation is the brushing feature that can filter only certain territories. In our implementation, the parallel coordinates offers a brush function for each axis, useful to filter only the territories that for each crime have a certain range, a tooltip that provides the name and the year

(when selected) of the territory selected, the possibility to drag and re-position axis to see correlation among attributes and finally the very useful highlights of the same territory in different year that allows the user to check the crime trend that a certain territory had during that years.



In the figure we can see that if the user want to check the trend of a territory in the years, it's highlighted. Moreover, using the brush function it's possible to have a more clean visualization of that territory among all others and the range selected of some particular crimes (rapes and theft).

4.3 Multidimensional scaling

Multidimensional scaling (MDS) is a technique used for dimensionality reduction, allowing to summarize the original p-dimensional data space in the form of lower k-dimensional components that are easier to plot and visualize.

The data used for multidimensional scaling are dissimilarities between pairs of objects. We want to represent these dissimilarities as distances between points in a low dimensional space (here 2d space) such that the distances correspond as closely as possible to the dissimilarities.

In our case, we used MDS to output a spatial representation, consisting of a geometric configuration of points representing Italy's locations (both regions and provinces). The input dataset for the MDS consisted of these territories having 31 variables each, representing different crimes. To understand the relationship among objects we had to generate a proximity matrix. This is a symmetric matrix which encode this dissimilarity: the more the value of each cell *c[i][j]* is bigger the more are different the objects *[i]* and *[j]*. To calculate this proximity values we used the Euclidean distance, but first we decided to calculate some coefficient in order to weight every crimes according to their danger.

In order to produce these coefficients we considered the convicted criminals dataset.

The first coefficient is used to derive a sort of scale of danger, where longer sentences correspond to a more dangerous felony. We summed up the number of sentences for each infraction weighting them by the period of detention (going from 1 month to 120 months). The result was then divided by the totality of judgment for that specific crime weighted on the maximum period of imprisonment possible (120 months).

The second coefficient is similar as before but, in this case, it was taken into account also the frequency with which a certain felony happens, in order to highlight overly committed crimes (like for example robberies).

At this point the user has the ability to choose the coefficient that he prefers or he can even compute a dissimilarity matrix without these additional information.

To make a complete study according to user needs is also possible to change the dataset considering the demographic incidence and to select a per regions study or a per provinces one.

Once the matrix is calculated, it can be given as input to the MDS algorithm which results can be projected into a scatter plot.

We have to highlight that the resulting axis has no semantic meaning, their only purpose is to quantify the distance between locations to understand how delinquencies impact differently on each part of the country.

For this reason particular attention can be given to outliers detected by this approach. However, even if MDS can give us a rapid idea of the situation, it cannot give alone the reason why points are placed in certain position. To understand that, we can combine this result with the parallel coordinates visualization. In fact, by selecting

specific points with a brush feature is possible to see them both on the map and on the parallel coordinates. Finally, to have a clear view of what happens, especially with cluster, there is the possibility to zoom over the chart and inspect locations and to add or remove labels when needed (to not occlude points distribution).

4.4 Visualizations coordination and interaction

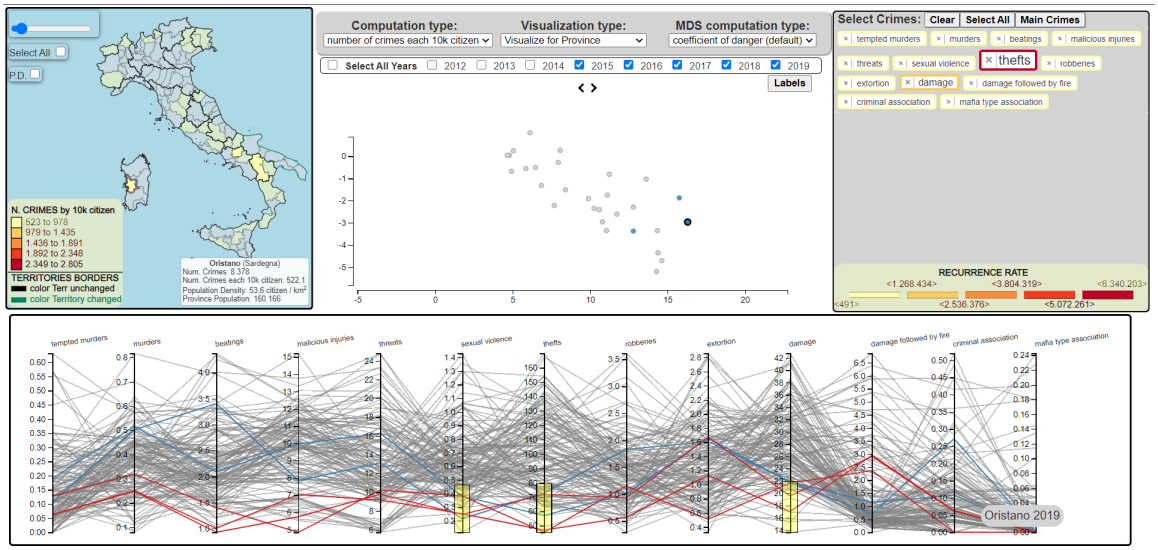
As previously said, the user has the ability to change various aspect of the data perceived.
First of all, through the specific menu, it is possible to change the year of study by selecting multiple periods, even not contiguous in time, and by clicking on specific button to go forward or backward the selected years. This result in a change of computation in all three visualizations triggering an analytic function that sum information of the dataset, when multiple time period are chosen. The same thing happens when deciding to take into account the weight that the population has on areas or not. Another important change of visualization is referred to the choose of analyzing regions or provinces.
Inside the menu we can find the feature concerning the possibility of making a selection over the crimes of interest. This choice trigger an analytic function used to re-organization of the coloration of the map and of the crime selection menu itself and it also leads to the change of the number of axis in the parallel coordinates. This selection menu provide a legend that makes sense to the colors showed around every label. The crimes are encoded by the same color scale present on the map legend representing their recurrence rate. In conclusion, by clicking on a range of the legend only the corresponding crimes will be selected.
For what concern the coordination between views, every selection, or even deselection, on the map leads to the adding or removing of the correspondent polyline on the parallel coordinates. While, going over a location on the map, a polyline on the parallel or a point in the MDS leads to an highlight of the same data on all others visualizations. Same as for the selection, through brushing, of multiple points on the scatter plot.
While making a brush on the parallel coordinates produce a filtering of the corresponding points in the other visualization, to have a clear view of what areas respect the chosen constraint.
All of this allows the user to conduct deepen studies, as we will now see.

5 Case of studies

Our project can have various goals and so be used by different type of users, from private user to public authorities, thus, just to give some examples and a semantic direction to our work, we built three different hypothetical user analysis.

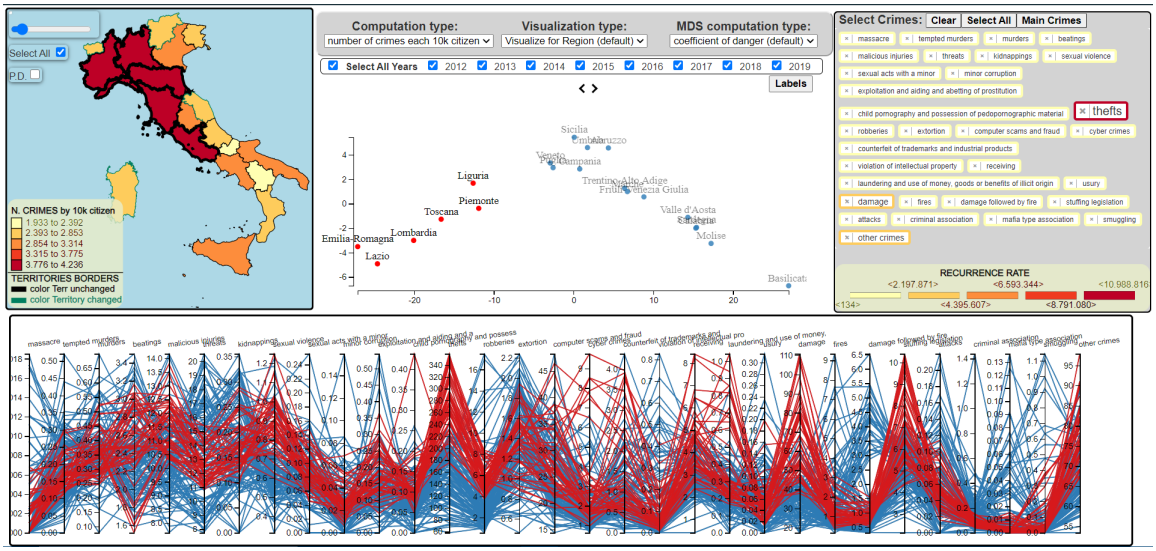
5.1 Moving house

We suppose that a private citizen wants to know which is the "safest" province where to live, in order to move there. So, he decides to check only the crimes related to its daily life, excluding felonies as kidnapping, fires, usury etc.. Moreover he must check the number of crimes related to the province population, not the absolute one. Then, going to the legend of the map we can select only territory that correspond to the class with the lowest number of crimes, choosing a particular year or all of them. Now it's possible to use the brush of the parallel coordinates, selecting the lowest values of sexual assault, theft and damages, the one considered as the most interesting crimes. As result, the lowest province is Oristano, while the most fraudulent are Rimini, Bologna, Firenze and Cremona.
After that in order to be sure that this result is not casual, due to a lucky year, he could also check that this low trend is maintained analyzing more years together. For example by taking the last 5 years we can still see that Oristano is a good choice.



5.2 Region trend over years

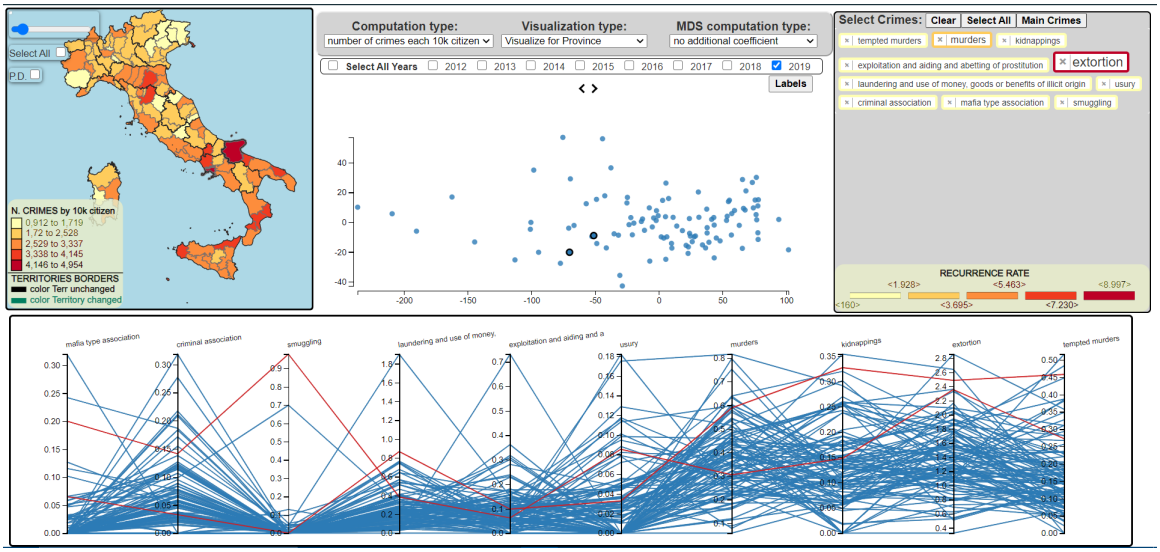
We suppose that the internal minister wants to visualize the trend over the years of the regions, in order to establish some kind of policies. So he decides at first to check all regions with all crimes, and for having a more balanced study, selecting crimes happening over 10k citizens. By starting from 2012 is possible to see that north center regions are red and highlighting them through the MDS allows to see their evolution year by year. In all of them they've always maintained a bad trend with usually medium high spike on thefts, damage, malicious injuries, kidnapping and sexual violence. However even though these regions are always the worst, the number of crimes is generally decreasing in all territories. At this points he wants to understand what can be done to improve the situation. By selecting all years we will see the total trend, from the parallel coordinates it's possible to notice some spikes in correspondence of thefts, narcotics and damages. However even homicides, kidnappings and sexual assault (that are pretty dangerous crimes) are in the medium-high slot. This means that these crimes are the ones that should be kept more under control to improve the situation.



5.3 Mob crimes

We suppose that the Anti-mafia Police Department (DIA) want to study all felonies strictly related to the mob, mafia and racket traffic. So, after selected only the related crimes, from the map is evident that the worst situation concern south Italy, involving especially Campania followed by Puglia and Calabria. Doing a deeper analysis we can see that the provinces of Foggia and Campania has respectively the higher spikes for what concern kidnapping and contraband. However we can also see that in the medium high range appears

provinces like Firenze or Bologna that were not spotted through the regions view because of the low values of the surrounding provinces. While Calabria is almost entirely composed by pretty risky provinces.
This stress the fact the major redevelopment must be done in the south area without forgetting of the other dangerous territories.



5.4 Correlation between crime and population density

We suppose that an Italian government department want to understand the reasons of a certain concentration of crimes in some places. They want to verify if exist a correlation among population density and crimes committed (their number or type), so activate the visualization of the density circles and set the computation of crimes weighing them on 10000 citizens. On the map it can be seen that selecting smaller circles (regions with the lower population density) it's more likely to have a correspondence with regions with a lower number of crimes. so can be now concluded that a less closeness among peoples means lower number of crimes.

5.5 Most frequent crimes

Thanks to the legend is possible to understand the most frequent crimes happening in Italy throughout the years. We can notice that the most frequents is always thefts, category that include lots of subcategories, as well as other crimes. By eliminating them we give more space to damages, computers scam and threats. However is fascinating to notice how computers scam slowly increases during the years becoming more and more frequent. So we decided to study the evolution of these crimes. We can see that thefts and damage generally decrease through years while computer scam and cyber crimes increase. Furthermore we can see how the prevalence of these last crimes concern the North Italy, in provinces like Milano, Imperia, Savona, Novara, Gorizia, Belluno and surprisingly Cagliari.

6 Conclusion

We introduced *VICrime*, a visual analytics tool to support the analysis of crimes in Italian territories. This tool has the dual property to offer to the user both a general and global view of all crimes denounced to the authorities, and a fine grained visualization that can inform the user about what's happening in a particular place, giving him a geographic localization, the dissimilarity of that place with respect to others and the trend of that place all over the crimes selected.
All of this conceived in a manageable and fitting-screen environment, that can help users to start solving crime problems.

6.1 Future work

When the new and updated data will be published by ISTAT, this visualization will easily integrate them giving an up-to-date vision of all crimes in Italy. With similar data, the work could be expanded including more geographical territories (Municipalities, Other nations and so on)

6.2 References

`https://github.com/FrancescoArtibani97/VA-project`

http://dati.istat.it/Index.aspx?DataSetCode=DCCV_DELITTIPS

http://dati.istat.it/Index.aspx?QueryId=25157#

https://colorbrewer2.org

Luiz Jose Schirmer Silva , Sonia Fiol Gonzales, Cassio F. P. Almeida, Simone D. J. Barbosaand Helio Lopes : CrimeVis: An Interactive Visualization System for Analyzing Crime Data in the State of Rio de Janeiro

https://www.scitepress.org/papers/2017/62587/62587.pdf

Germain Garcia, Jaqueline Silveira, Jorge Poco Member, IEEE, Afonso Paiva, Marcelo Batista Nery, Claudio T. Silva - CrimAnalyzer: Understanding Crime Patterns in Sao Paulo

https://www.researchgate.net/publication/336571559_CrimAnalyzer_Understanding_Crime_Patterns_in_Sao_Paulo

Lucia Andrighetti: Un'analisi sul rapporto tra criminalità e benessere in Italia

https://core.ac.uk/download/pdf/11658542.pdf?repositoryId=377

Dominique Brodbeck, Luc Girardin - Design Study: Using Multiple Coordinated Views to Analyze Geo-referenced High-dimensional Datasets

https://www.researchgate.net/publication/4024205_Design_study_Using_multiple_coordinated_views_to_analyze_geo-referenced_high-dimensional_datasets

https://www.crimemapping.com/map/agency/426

https://maps.nyc.gov/crime/

https://lab24.ilsole24ore.com/indice-della-criminalita

https://alexandra.kim/projects/crimevis