

# Exploring potential causes, consequences and visualizing evolution of major air pollutant emissions over EU28

Jules Sauvinet, Marine Ruiz

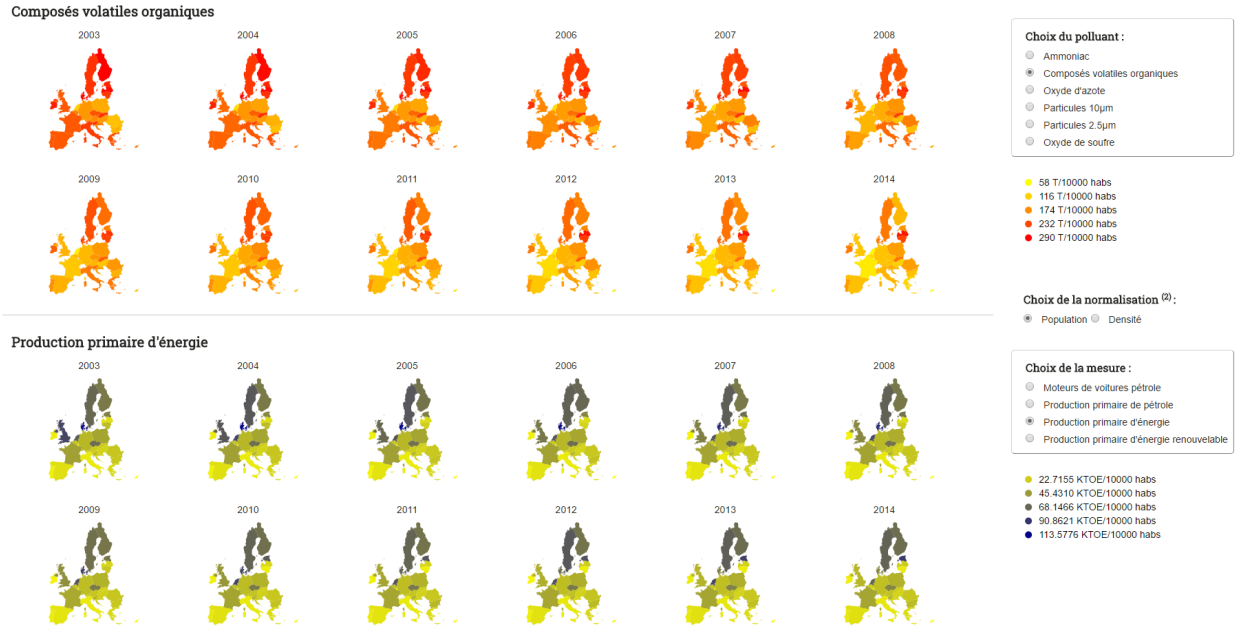


Fig. 1. Main view of the visualization: 12 small maps of EU28 colored according to Non-methane volatile organic compounds emissions on the top and 12 small multiple maps of EU28 showing primary production of energy from 2003 to 2014.

**Abstract**— From now on, the media talks about environmental issues on a daily basis. Pollution peaks have become commonplace in large metropolises and air pollution is becoming a subject that is increasingly affecting citizens. As pollution is not only a visual or olfactory discomfort, it turns now the main environmental health risk in the world. On average, it causes the premature death of 7 million people worldwide, including 600 000 in Europe and more than 50 000 in France, according to the World Health Organization, the Ministry of the Environment and the Environment European Environment Agency. We propose here a model which allows to visualize the evolution of the emissions of the main harmful pollutants during 12 years and to explore the potential causes, consequences by analyzing the correlations. For instance, does it seems to be an obvious correlation between the fines particles emissions and pulmonary diseases?

**Index Terms**—Air pollution, Pollutant, Ammonia, Sulfur oxides, Non-methane volatile organic compounds, Nitrogen oxides, Particles fines, EU28, Ecology, Small multiple, Small multiple maps, Maps, Data visualization, Geographical data visualization, Data visualisation.

## 1 INTRODUCTION

The emissions of most harmful air pollutants has globally decreased over the past 25 years in European Union (see the Fig. 2 below). Nevertheless, emissions remain very high, particularly in some countries whose economy depends on sectors responsible for the discharge of certain pollutants. While some countries have taken stock of the issue and taken steps to reduce their emissions, others are struggling to keep pace with economic and political conflicts. As a result, there are regularly many pollution peaks in major European cities, posing serious problems for the environment and health.

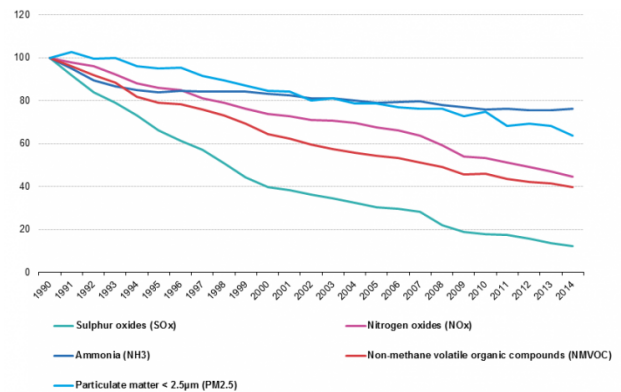


Fig. 2. A visualization of pollution over EU28 for the past 25 years. The image is from [1] and is in the public domain.

- Jules Sauvinet is studying at Université Claude Bernard Lyon I. E-mail: [contact@julessauvinet.fr](mailto:contact@julessauvinet.fr).
- Marine Ruiz is studying at Université Claude Bernard Lyon I. E-mail: [marine.ruiz@etu.univ-lyon1.fr](mailto:marine.ruiz@etu.univ-lyon1.fr).

Taking note of this issue, we created a model which would have as ambition to inform people on the evolution of the population over the last years. Our model will also have to be able to geolocate (by country) pollution according to the pollutants. We wanted to highlight the countries that pollute the most, and force us to question why? Is it a function of the sector of activity? Do these countries comply with European ecological standards? Does the European policy on environmental policy therefore have a real interest or should it be reformed?

We thought our model, created our model and developed our visualization in 1 month. So, of course, we did not have time to answer all these questions. Nevertheless, we tried to introduce them by giving visual keys to ask questions such as : does the rejection values of ammonia of the countries of the European Union responds to the same trend as the use of pesticides? In other words, do the values of our data indicate that the more pesticides a country uses, the more ammonia it releases?

The comparative power of the model we built resides in the use of small multiple maps. Essentially, a small multiple is a series of displays with the same design structure repeated for all the images, arranged in a grid. When a visualization is made of several dimensions including time and space, it is a good way to display them all in a single view.

The first part is to be able to visualize the evolution of pollution in the European Union countries for 6 pollutants identified by Eurostats as among the most dangerous for the environment and health: Ammonia (NH<sub>3</sub>), Sulfur Oxides (SO<sub>x</sub>) Nitrogen Oxides (NO<sub>x</sub>), Non-methane volatile organic compounds (NMVOC), Fine Particles less than 2.5  $\mu$ m (PM<sub>2.5</sub>), and Fines Particles less than 10  $\mu$ m (PM<sub>10</sub>).

In order to be able to visualize the evolution of the emission of these pollutants, we have developed an interface composed of 12 small maps from the European Union countries to 28 concerning 12 recent years (the data we have of the pollutants range from 1990 to 2014, and these are more complete the longer the time). The user chooses the pollutant from which he wishes to consult changes in the list of 6 specified earlier. Then, each small map (see the Fig. 3 below) is colored according to a scale of color more or less intensely depending on the amount of pollution rejected per inhabitant. The countries whose data we do not have are colored gray. The color scale is constructed from two minimum and maximum values which are the minimum and maximum rejected pollution values by a country during the 12 years considered.

Thus, the user can observe the evolution of the pollution to compare the countries between two countries or for the same country for two different years. The small maps give him an overview showing the general trend of the evolution of pollution for the 12 years and the pollutant considered. Moreover, the color scale makes it possible to isolate clearly the countries which have a much higher emission than the others (e.g. Ireland in Fig. 3).



Fig. 3. A visualization of a small map of Non-methane volatile organic compounds emissions. The image is from our visualization.

The second part of our model is the exploration of the potential causes and consequences of pollution among those that we propose. Once the pollutant has been chosen, the user can choose from a list what we will call a comparison measure. For example, for ammonia, the user can choose from the "Nitrogen Fertilizer" measure. A second set of 12 small maps is then made showing this time the evolution of the measure considered for the maximum 12 years common with those of the pollutants whose data are available.

Thus, by comparing the intensity of the coloration of the two series of small maps, the user can identify if there is a potential link between the pollutant and the measure considered. First, it is possible to observe if over a year the color nuances correspond and potentially indicates a correlation for the target year. Secondly, it is possible to observe whether the evolution trends (growth, decay, stagnation) are connected (similar, or opposite). The goal is to invite the user to wonder but obviously does not affirm the observed link.

## 2 RELATED WORK

- In his visualization [3], Boy and Fekete shows CO<sub>2</sub> emissions in countries around the world as clouds of visual smoke. The visualization is reused here [2] as a narrative visualization. The literature in InfoVis (Information Visualization) often states that a narrative visualization (storytelling) can engage the reader in data mining. The visualization allows an exploration in time and the comparison of emission between the different countries. But the reader must take the time to "read the visualization". In addition, it is not possible to directly compare the issue between two years.
- Here [4], Brewer talks about the adapted colors to use for mapping and visualization. Some of his work is used here Color brewer" for the design of color scales.
- The concept of displaying a series of small maps is also known as small multiple as popularized by Tufte [11] and Cleveland [5]. Tufte defines the small multiples as "Illustrations of postage-stamp size are indexed by category or a label, sequenced over time like the frames of a movie, or ordered by a quantitative variable not used in the single image itself". According to him, small multiples can show rich, multi-dimensional data without trying to cram all that information into a single, overly-complex chart. In his book [12], p.67, he says that "For a wide range of problems in data presentation, small multiples are the best design solution".
- In his website article [7], Zach Gemignani says about small multiple that :
  - "They allow for the display of many variables with less risk of confusing your audience."
  - "The reader can quickly learn to read an individual chart and apply this knowledge as they scan the rest of the charts. This shifts the readers effort from understanding how the chart works to what the data says."
  - "Small multiples enable comparison across variables and reveal the range of potential patterns in the charts."
- In his article [10], Smith show the interest of small multiples maps for geographical and time visualization. He says that multiple small maps may prove useful as a method of data visualization and/or seeking. Furthermore, according to him, small multiples maps have the potential to address the problem of highly dimensional data, and simplify searches for information over time.
- In this article [9], the authors show how small multiples maps might be useful to complete the task in of comparison of the variation of data values over other variables. They say that small multiples of a geographical map is relevant to obtain to compare different spatial configurations of values.
- In his article [6], Stephen Few talks about choropleth maps and the use of color intensity for displaying quantities.
- In this article [1], from the website Eurostat site where we picked our datas, a study of the evolution of the pollutants of our visualization

is presented with an explanation of the emitting sectors. The format of this kind of article does not have the catchy side that we want to have with our visualization.

- Here is this website's article [8], the author explain how to use the NUTS geographical data that we use, some Eurostats' data (roughly similar to the format of the measurement data we use) to display crimes in Europe according to density with D3.js. This article helped us to create our visualization.
- These two blocks using D3.js to display small multiple (block 1 and block 2) technically helped us to realize our small multiple maps.

### 3 PROJECT DESCRIPTION

#### 3.1 Project's choice

##### 3.1.1 Why we have chosen this project

First, we wanted to exhibit atmosphere and water pollutant data for France. But, we had too many pollutant indicators and not enough measures to notice a possibly correlation. In fact, we had 33 atmosphere pollutant indicators and 33 water pollutant indicators and no correlated measure.

So, we decided to change our course project deciding to exhibit only atmosphere pollutant data, and this time, for the European Union. We also decided to keep only 6 pollutants contrary to France and to search more correlated measures. We wanted to find countries that emit the most. We also wanted to put forward a significant number of correlations with our measure data.

##### 3.1.2 Where we found our data

We found our data in the Eurostat website. This website offers many databases for the European Union countries and it is possible to see a visualization of these databases in this same website. We can find databases with information about:

- Economy and finance
- Population and social conditions
- Industry, trade and services
- Agriculture, forestry and fisheries
- International trade
- Transport
- Environment and energy
- Science, technology, digital society

All of our data are taken from these databases. Eurostat offers data files that contain a subsection of databases. So, even if we make use of a piece of a database, it is not necessary to download an entire database, but just a piece of it.

#### 3.2 Web Page description

We organized our Web Page into 3 parts:

- First, a header which describes the project.
- Then an area with 12 small maps of the European countries, showing pollution-related data, thanks to the pollutant list, the legend and the colors. On the top of this area, there is a title which displays the name of the pollutant.
- Finally, an other area with the same 12 small maps, this time to show pollutant-related data such as causes, consequences or correlations to the emission of the pollutant selected.

##### 3.2.1 The pollutant area

To be more accurate, next to the 12 pollutant small maps, there is a button radio list of 6 pollutants (see the Fig. 4 below). By default, when the Web Page is loaded, the pollutant on top of the list is selected. When an other button radio of the list is selected, the 12 small maps are updated, as well as the legend and the title.

So, we have a dynamic visualization of pollutant emission evolution in time.



Fig. 4. The pollutant selection box. The image is from our visualization.

##### 3.2.2 The measure area

Likewise, in the measure map area, there is a list of measures available, and when an other measure is selected, all the area is updated.

The origin of the emission is different for each pollutant, so to maximize the relevance of the comparison between pollutant and measure, we offer a different list of measures for each pollutant. That is why, when the pollutant changes the list also does.

Moreover, the number of maps for pollutants and measures depends on the data we have. In fact, when we did not have any data for a year, we preferred not to represent the map of this year at all instead of showing a total gray map. We only selected the years for which we had some data for both pollutant and measure. If we have some data available for more than 12 years since 2001, we take into account the data for the 12 more recent years (for example, from 2003 to 2014). Conversely, if since 2001 we have less than 12 years of available data, we take all the data we have since 2001. So, according to the pollutant and the measure selected, there are more or less maps.

All of these choices enable us to have a dynamic visualization. Indeed, we maximize information by deleting non-pertinent elements. For example, the user avoids losing time with comparisons between pollutant and measure which may lead to nothing and his eye will not be caught by some gray small maps that do not bring any piece of information.

##### 3.2.3 The maps coloration

The coloration allows a first visualization of the comparison between the pollutant and the measure. But, it is possible to be more specific, by running the mouse over a country on any pollutant- map or measure-map.

In fact, the name of the country appears and all the years are replaced by the pollutant emission value for this country in the pollutant area and by the measure value for this same country in the measure area.

Once again, in order to keep only the relevant elements, if we don't have any data for this country for a specific year, the year is replaced by a blank space.

Our visualization arrangement and the colored maps allow, at first sight, to have an idea of the correlation between the pollutant and the measure. Then, showing the values for a country in the 24 maps help us have an idea of the evolution of both of them. Visualizing pollutant and measure values together showcases correlation or anti-correlation in the evolution between the pollutant and the measure.

##### 3.2.4 The data normalization

Then, we have chosen to normalize our data because the most populated countries pollute more than the least populated countries. If we color the small maps with the untreated data, the legend will be slanted.

Then, to solve this problem, we offer two types of data normalization (see the Fig. 5 below). By default, when the Web Page is loaded, the population normalization is selected, but it is possible to change it with the two buttons radio. When density is selected, all of the 24 small maps are updated, as well as the two legends. Conversely, it is possible to return to population normalization selecting the population radio button.

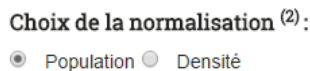


Fig. 5. The normalization selection box. The image is from our visualization.

### 3.2.5 The countries selection

However, even if we normalize the data, some countries have an outlier value for a pollutant or a measure. Consequently, the legend is slanted and the coloration does not bring an accurate piece of information because one country is red and all the others are yellow. We cannot see in the maps real shades.

So, to solve this problem, we offer on the top of the pollutant area, the list of the all countries for which we have some data. By default, when the Web Page is loaded, all the cases are checked. If a country has an outlier value for a pollutant-measure couple, it is possible to un-check this country.

That means that, in the pollutant and measure areas, this country turns gray on all the 24 maps and running the mouse over it does not generates any visual effect. The 24 maps and the two legends are updated, but this time, they do not take into account this un-checked country. It is possible to see now, a better shade with a legend more accurate.

To conclude this visualization presentation, we offer a small description and an image of the pollutant selected in top right corner. As the Ammonia is selected by default, we see when the Web Page is loaded, the Ammonia description.

### 3.3 Choice of data format

We decided to present 12 maps for one pollutant and 12 maps for one measure. One single map is very small to describe precisely 28 countries. We had the choice between NUTS1, NUTS2 and NUTS3 formats.

NUTS, The Nomenclature of Territorial Units for Statistics, is a geocode standard for referencing the administrative divisions of countries for statistical purposes. The standard was developed by the European Union. There are three levels of NUTS defined: NUTS1, NUTS2 and NUTS3. NUTS1 is the least precise and NUTS3 is the most precise.

In the case of France, NUTS1 stands for France, NUTS2 stands for a region (Bretagne, Normandie, etc ) and NUTS3 stands for a French department (Ain, Aisne, etc ). Not every country has all levels of division, depending on their size. One of the most extreme cases is Luxembourg, each of the three NUTS divisions corresponds to the entire country itself.

We preferred to use NUTS1 because as each map is very small, it is not easy to run the mouse over an area if we use NUTS2 or NUTS3 format (a single area would be too tiny).

### 3.4 Description of pollutants

We offer 6 very important pollutants:

- Ammonia
- Volatile Organic Compounds
- Nitrogen Oxide
- 10  $\mu\text{m}$  particles
- 2.5  $\mu\text{m}$  particles
- Sulfur Oxide

Each pollutant is told in ton in our data files. As we normalize the values by population or by density, the resultant value is very small. So, it is necessary to multiply this resultant value by 10000.

In fact, in the legend, each color represents a value of pollutant in ton per 10000 inhabitants.

### 3.5 Description of measures

We chose 18 measures that we considered relevant regarding the pollutants we selected.

There are measures about energy production:

- Primary production of electricity
- Primary production of renewable electricity
- Primary production of oil
- Primary production of gas
- Primary production of coal
- Nuclear heat

There are measures dealing with health:

- Cancer death
- Heart diseases deaths
- Length of hospitals stays for lung diseases

Some measures about agriculture:

- Pesticides
- Nitrogen fertilizers
- Potassium fertilizers
- Phosphor fertilizers

We have some measures on tax prices:

- Environmental tax
- Transport tax

Finally, there are some measures dealing with transport:

- Diesel motor cars
- Petrol motor cars

Likewise, for pollutant, it is necessary to multiply the normalize value by 10000. So, in the legend, each color represents a value of measure in the measure unit per 10000 inhabitants.

### 3.6 The measure suggestion

To optimize the comparison between these pollutants and these measures, we offer a list of measures for each pollutant. To do so, we first searched for the possible causes and consequences of the emissions of each pollutant and any data which be correlated with it. Then we tried to find adapted measures in our data files, or in the Eurostats database.

Thus, the computing of the all of the information gives this:

- For the Ammonia pollutant, we identified that the major sector that produces rejects was the agriculture. So we created a list with Pesticides, Nitrogen, Potassium and Phosphor fertilizers.
- For the Volatile Organic Compounds pollutant, transports and industrial activity are responsible for the emission of this pollutant. So we created a list with our data on Primary production of electricity, renewable electricity and coal and Petrol motor cars.
- For the Nitrogen Oxide pollutant, every energy transformation produces Nitrogen Oxide. So we created a list with all of our data concerning energy production, that is to say Primary production of electricity, renewable electricity, oil, gas, coal and Nuclear heat.
- For the 10  $\mu\text{m}$  particles and 2.5  $\mu\text{m}$  particles pollutants, we identified that the major sector that produces rejects was car diesel combustion. So, we created a list with Diesel motor cars as a first choice. To see if at least one correlation exists, we add in this list all of our measures dealing with health, this means Cancer death, Heart diseases deaths and Length of hospitals stays for lung diseases. To finish, as particles are a real issue for the government, we add to this same list data about environment and transport taxes.



- For the Sulfur Oxide pollutant, we observed that it was usually created during burning and refining processes, from the sulfur contained in raw materials such as coal, petrol and other ores that contain some metal. That is why we offer, for this pollutant, the same list as for the Nitrogen Oxide pollutant.

## 4 EVALUATION

We evaluated our interaction model by looking at whether known correlations were visually highlighted. If one takes Ammonia, which is emitted at 94% by agriculture, it would seem logical to establish a clear link between the colorings of colors for pollutant emissions and for the total rejection of pesticides. We asked people around us (10 persons in total) to make assumptions about looking at the visualization set on the pollutant 'Ammonia' and measure 'Pesticides'. All responded that visualization should, in their opinion, indicate that the more a country releases ammonia (per capita), the greater the release of pesticides. We have thus deduced that for some public, our visually was potentially relevant for the linkage between pollutant and measurement.

## 5 DISCUSSION

We will discuss here the potential improvements we have imagined for our visualization and that we had not had time to develop.

### 5.1 Technical points

First of all, from a technical point of view, it is possible that the visualization is not a satisfactory rendering on all the supports due to the dimensions of the screens. As the various graphics components do not all have flexible dimensions, their size and positioning may be different from those we have defined from our screens, hindering the readability and efficiency of the visualization.

### 5.2 Getting more visual clue

In addition, we thought about being able to display graphs of changes in pollution values and the chosen measure of a country for all the years of which we have data available. When the user clicks on a country on a small map, a graphic would appear in a container at the bottom of the visualization, giving an extra visual key to observe the correlation and evolution of the data.

Moreover, we thought about giving an additional visual key to compare values across all countries, bidding on the information provided by the color scales. When the user clicks on a country, two value tables for all other countries for the current year are displayed for the pollution values and the current measure. Not all users have clear interpretations of the differences between colors, which would have allowed some users to more accurately visualize the targeted country according to the others and thus reach a wider audience. Nevertheless, since the visualization is already rich in the number of visual elements, the added information could also overload the information and obscure the view of the user.

### 5.3 Getting more information

Moreover, we have imagined that we can integrate the notion of respect for ecological norms to highlight countries that do not respect it and thus highlight the relevance of European environmental policy. Non-compliant countries could have an additional visual code such as being covered by hatching or dotted lines.

We also thought we could add more pollutants. To do this, we would have had to extend our knowledge of the field of air pollution in order to choose those with relevance for comparison with our measurement data. Similarly, we wanted to add further measurement data (particularly in the field of health), including data moving away from commonly established correlations (eg diesel cars and fine particle rejection), thus showing the consequences of probably less well-known pollution. However, the available data are often not available and not always easily interpretable for non-specialists in the field of pollution.

Another idea was to be able to navigate through all the common years available for the polluting / measuring combination with a 12-year sliding beach slider. That would have enabled us to make full use of our data. In addition to this idea of enriching the temporal dimension, we would have liked to be able to propose a monthly time scale in addition to the annual scale to observe with more refinement the peaks of pollution. Unfortunately, no asser data is available today with such a fine time dimension.

Finally, one possibility would have been to use a finer geographical breakdown, such as that proposed by the NUTS2 or the NUTS3 format. However, the scale of small maps is no longer suitable for reading a visualization with depictions of such fine geographical areas. It would have been necessary to reduce the number of small maps by removing years. It is a compromise to find, providing maximum information without drowning the user, while maintaining legibility and complexity allowing the visualization to remain fast and dynamic for a comfort of use.

## 6 CONCLUSION

We constructed an interaction model that allows a user to quickly observe the evolution of pollution within the EU28 for 12 years. User of small maps and color scale is an unusual approach still in evolution representations for serious subjects like pollution and brings visual comfort. This approach should reach a wider audience than current statistical graphs of statistics, being a less frightening visual approach for uninitiated people.

Our interaction model also explores the potential causes and consequences of the emission of certain pollutants. It allowed us to confirm certain intuitions (e.g. use of chemical fertilizers and rejection of ammonia), highlighting with potential visual correlations. The use of color scale and small maps also makes it possible to develop a playful pedagogy, aimed at reaching a wide target audience, including people who are not familiar with pollution issues or children.

Finally, our interaction model allows the user to compare which countries are the biggest polluters, depending on whether you choose per capital emissions or take into account population density. This may raise the question of the effectiveness of European environmental policies. This can lead to questions about countries that potentially meet certain standards or are lagging behind in the development of their reduction policies.

## ACKNOWLEDGMENTS

The authors wish to thank Aurélien Tabard, Nicolas Bonneel and Romain Vuillermot for their piece of advice and their support during our work.

## REFERENCES

- [1] Air pollution statistics, November 2016.
- [2] Co2pollutionmap, March 2014.
- [3] J. Boy and J.-D. Fekete. The CO2 Pollution Map: Lessons Learned from Designing a Visualization that Bridges the Gap between Visual Communication and Information Visualization. *IEEE Conference on Information Visualization [Poster paper]*, November 2014.
- [4] Brewer. *Color use guidelines for mapping and visualization*. MacEachren and D. R. F. Taylor ed., Oxford, UK, 1994.
- [5] W. S. Cleveland. *The elements of graphing data*. Wadsworth Advanced Books and Software, Belmont, CA, USA, 1985.
- [6] S. Few. Introduction to geographical data visualization. *Visual Business Intelligence Newsletter*, March 2009.
- [7] Z. Gemignani. Better know a visualization: Small multiples, April 2010.
- [8] R. V. i Rovira. Using eurostat's data with d3.js, October 2013.
- [9] A. Reimer, A. Unger, W. Meulemans, and D. Dransch. Schematized small multiples for the visual comparison of geospatial data, 2011.
- [10] B. K. Smith. Multiple small maps as an information seeking tool, 2005.
- [11] E. Tufte. *The visual display of quantitative information*. Graphics Press, Cheshire, CT, USA, 1983.
- [12] E. Tufte. *Envisioning information*. Graphics Press, Cheshire, CT, USA, 1990.