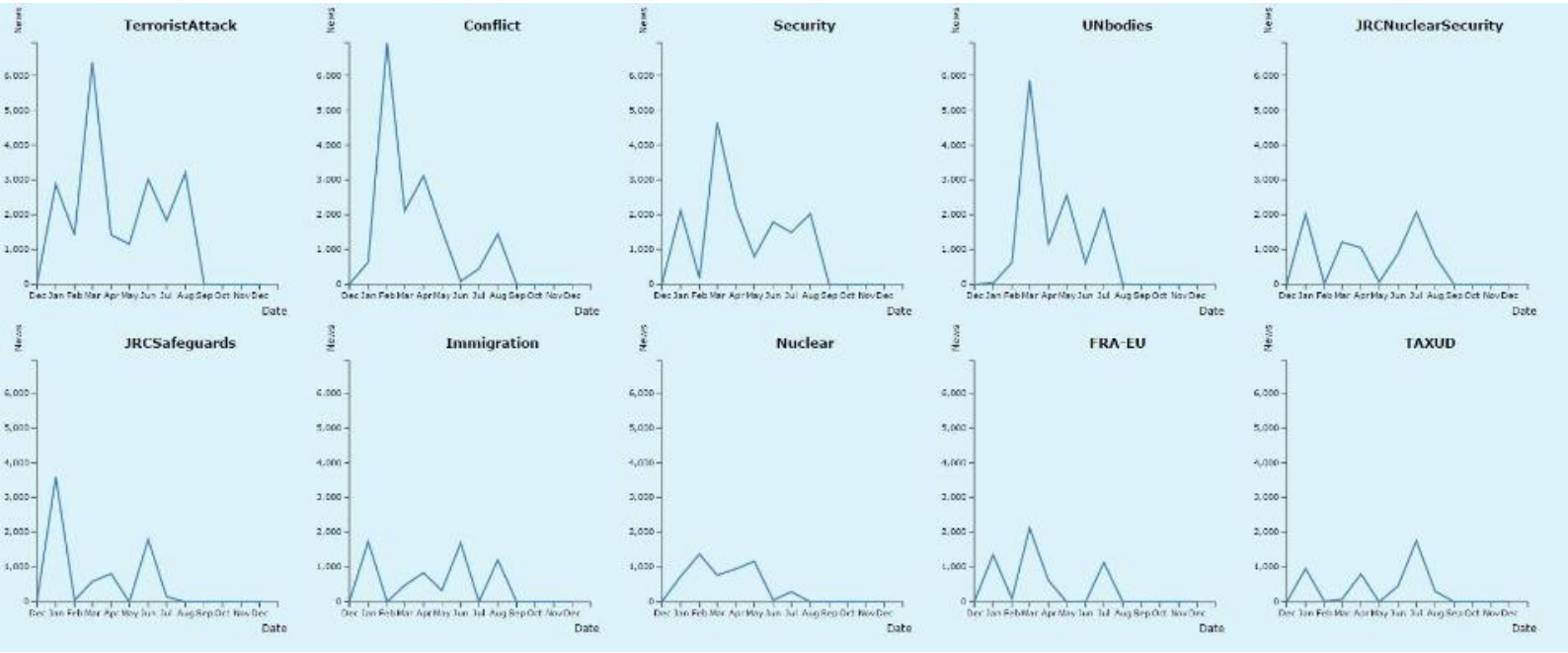
News Visualization with small multiples



Elena Akrytova: 336087895

Katerina Portman: 326812880

Ilana Veitsblit: 305235814

Data

Background

Today, in the age of mass communication, news are published every minute in the telecommunications networks. The news are published on the television, news sites, and even on our phone apps. This fact, causes all of us to be constantly connected with everything happening in the country and the world. News reports published in various fields: political, diplomatic, security, economic, sport etc. Looking at these categories can be particularly interesting. If we filter the data in a different ways, we can learn a lot about different countries, the events that happened at certain times and more.

DataBase

In our project we decided to work with database of news information in the United States. Each item in the original data contain the following attributes: id, images, keywords,category, author, fullText, title, date and more. Of course, all these features are essential and needed a person to understand whether it is relevant knowledge. Following what we wrote in the previous paragraph, the amount of news being published at a certain time is enormous. Therefore, we filter the database and left only the features that are relevant to us. For our needs, we left the **date** and **category** only. This reduced significantly the database and allowed us to get a large amount of data that are relevant to us. After the changes, our data look like this:


```
{
  "date": "2016-03-21T01:33:00Z",
  "category": { "term": "UNbodies" }
},
{
  "date": "2016-03-21T01:33:00Z",
  "category": [
    { "term": "Erasmus" },
    { "term": "EU-Pacific" }
  ]
},
```

As you can see there are reports which are associated with a number of categories, these reports will be counted with each category. For example: the item that published in 21/03/2016, will be counted ones in "Erasmus" category and ones in "EU-Pacific" category. News Uncategorized, did not enter to our visualization.

Data Distribution

In the original data, there are 16,186 news items that spread in a one month (21/03/2016 – 21/04/2016). And because we want to compare between categories for a longer period, one month is not enough. The amount of data is very big, and we cannot have more from it, therefore we created "fake" data for demonstrate our visualization and maximizing our goal. So, we created approximately 141,000 new items that spread in another 6 month. In the end we have news items from 01/01/2016- 28/08/2016.

How we created our data?

We did not create the data randomly, because we wanted to see that there are categories that have more news items form other categories. If we produce the data randomly, is distributed evenly and that's not our goal. Therefore, when we produced the data, we gave a priority to certain categories in certain dates in order to make our point and see the differences between the categories.

Task

Our visualization establishes the relation between the number of news items created during a set period of time. In our visualization you can see this relation over number of graphs. Each graph represents a category. This demonstration can be very interesting because we can understand which categories were more popular in certain dates.

The main goal is to compare between categories and to understand by that which subjects are more popular in a certain country, in our case- United States. By knowing which news items are more popular in a certain period of time, we can tell at a single glance what is happening generally in the country and in which topics we have breaking news. Because the news, should reflect reality.

To compare between the categories we used the small multiples view.

Why is this interesting?

At first glance, by looking at the visualization, we can understand if there was an unusual event in a particular time period. For example: if there are many news from the security category, we may conclude it was a security incident at that time. And if it lasted a long period of time, we may thought that there are deterioration in the security situation, as has happened in our country following the stabbing incident that happened in the center of the country.

In the opposite case, the user can detect concealment of certain information. If he knew it was an unusual event in a certain area and there is not much news in its category at the time the incident happened, he could conclude that someone trying to hide the event because of certain reasons. For example, in our country there are emergency situations created following the Haifa Bay refineries (בתי זיקוק). But for a some reason, we do not hear about it in the news. Is there a reason for this? We expect that near these dates, there will be a lot of news items in the environment category.

Who could be interested in that?

Analysts who analyze news and want to see general statistics about the number of news items in his country. This is a good way for the analist to see and analyze things, because if the news were ordered in a list, he could not get that value. So, for his needs, this visualization provides added value and the ability to compare the categories quickly by human eye.

Critics who criticize the news system. By looking at this visualization, they can easily tell whether the news system has faithfully served the purpose. They can see whether residents are getting an accurate picture of reality, or if they get any deception from the system and the state.

Journalists that want to write about "hot" topics. This visualization can be useful for them to know what happened and what issues the public may interested in.

Any user who wants to see the character of the state: political? security? environmental?. Actually ,anyone can be interested in that visualization and learn from it about the news system and events occurring in his country.

Visualization - Small multiple

Visual mapping

Each category is mapped to a single graph.

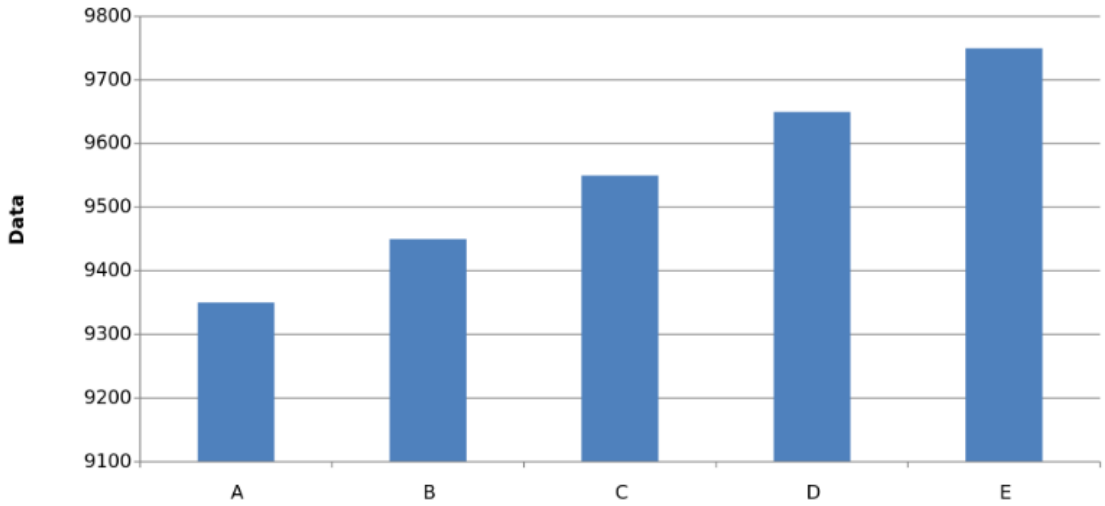
Data	Visualization
Date	**X Axis**
# Items in a category	**Y Axis**

Why we choose small multiple view?

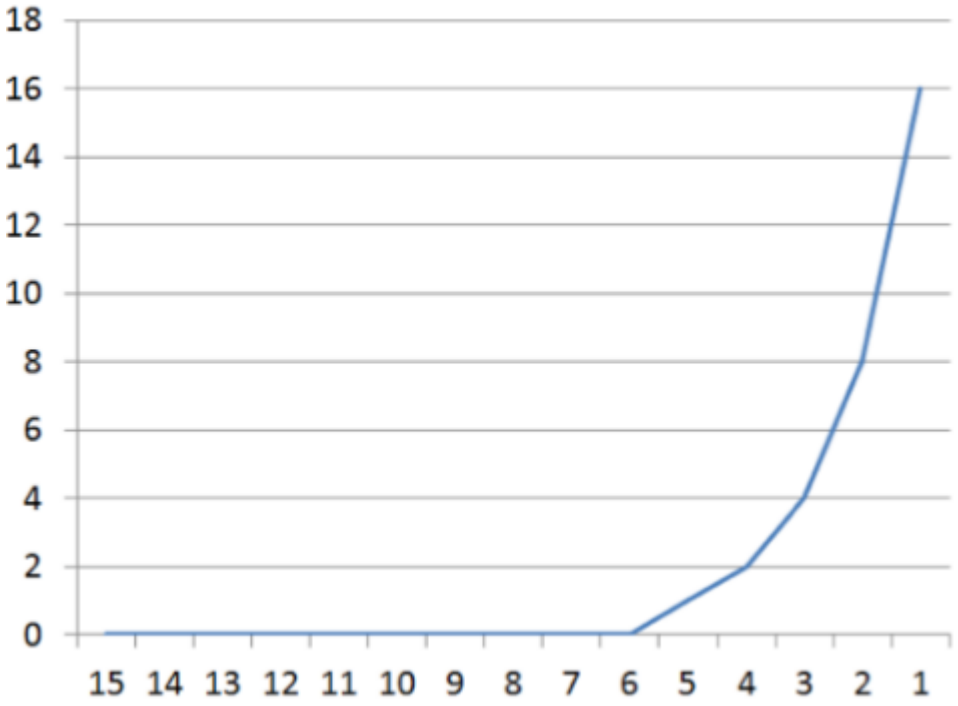
A **small multiple** (sometimes called trellis chart, lattice chart, grid chart, or panel chart) is a series of similar graphs or charts using the same scale + axes, allowing them to be **easily compared**. It uses multiple views to show different partitions of a dataset. And this is our main goal. By this view we can easily compare between categories and see the differences between each category in a specific period of time.

Type of the graph

In order to show the changes in various periods of time, we debated what type of graph to use. The two main suggestions were: **Bar graph** that looks like that:



The secong option was: **Line graph** that looks like that:



We decided that the graph that more suitable for us is the line graph, because we want to see the changes every moment to recognize ups and maximum points.

Interaction

Design

In our visualization we have the following elements:

Title- the title shows in real time which graphs we compare. This is important because we have different filters for our graphs, so the user can know in a real time which comparisons he sees. We chose the title's **background** to be related to the subject- news. By this background a user can at first glance understand the subject of our visualization.

Filters- we allow the user to select combination of categories that he wants to compare. The categories are divided into 2 groups: **main categories**: Culture, Economics, Education, Environment, Health, Politics, Security, Sport and Transportation. And **sub categories**: Unbudies, JRCNuclearSecurity etc. (each sub category, related to main category. For example: TerroristAttack is a sub category of Security). This is important for comparing between big issues like Security and Politics. We chose the filters we thought would be most relevant for users to compare. After choosing each filter the user need to press "Go" button for seeing the visualization.

The filters are:

Popular- the user can choose the most popular sub categories for comparison. In addition the user can decide how many category to show. We allowed to see from 1 till 12 different graphs. The minimum graph is 1 because may be the user want to learn about the most popular category and to investigate the behavior of specific category in a particular time. The maximum graphs that user can compare is 12, because we don’t want to overload user's eye.

Evaluation- Value

This filter is important because by looking at the graphs of this filter, the user can know immediately which specific categories were the most popular and to distinguish if there are unordinary event in this category.

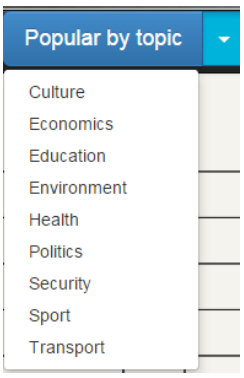
Self reflection

Advantages:

- The user can choose how many graphs he want to compare for his needs.

Disadvantages:

- If the user choose only 2 categories to compare by their popularity, one of them can be popular because there is one peak in a specific time. On the other hand, the other graph is monotonic without momentary peaks. Both of them considered popular graphs but in a different way. The visualization is not comparing the graphs with momentary peaks.
- There are categories that their names are not intuitive for the user and the user cannot understand about what the category are about and he should investigating this name to understand.



Popular by topic- in this filter the user can see the most 9 popular sub categories in each main category. We choose to show 9 graphs because we think that this number achieves the goal without overload user's eye. For example- the user want to see which sub categories were most popular in the Security section. So he will see the most 9 popular categories that related to the Security section.

Evaluation- Value

This filter is important because we allow the user to focus its filtering according to his individual needs. For example if I am a journalist focusing on Political issues, I can know which topics are more relevant for the public.

Self reflection

Advantages:

- Anyone can customize the visualization in his personal needs.

Disadvantages:

- We decided for the user how many categories to show. May be he wants more than 9 sub categories.
- We decided to show the most popular sub categories, but may be the user wants it in a different order. For example he can be intrested in categories that were less popular for his needs.

Compare topics- this is the **default** filter. In this filter the user can see a comparison between the main categories: Culture, Economics, Education, Environment, Health, Politics, Security, Sport and Transportation in all months together. The categories ordered by popularity in a descend order because we think that this is the most intuitive order for a user. we choose this view to be the main view because we think that this is the most interesting view that we can learn a lot from it.

Evaluation- Value

This is one of the most important filters for maximizing our goal. By this filter each user can understand which main category was more popular and to distinguish abnormal behaviors at general issues.

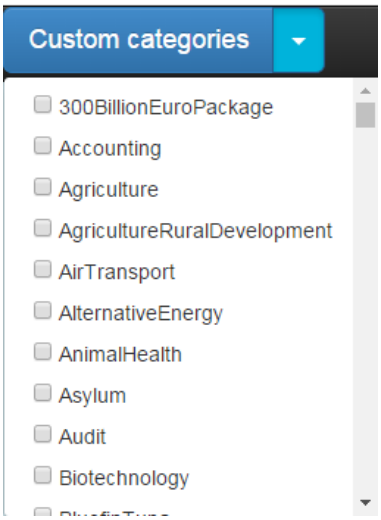
Advantages:

- It can contribute for user experience because he able to see which was the most popular main category and then dive into it and see which sub categories were popular in the subject itself by using "popular by topic" filter.

Disadvantages:

- Very general, do not know what's going on inside.
- We choose the topics that considered "main topics", may be there is a different partition for main topics.
- The user can be interested in topics that don’t appear there.

Custom categories- in this filter, the user can choose which sub categories he want to compare. The user can choose from 1 till 12 categories. The list of the sub categories the user can choose, ordered by alphabetical order.



Evaluation- Value

This filter can be interesting for professionals who work with news items. If they know the categories, they can filter in an appropriate way their exact needs.

Self reflection

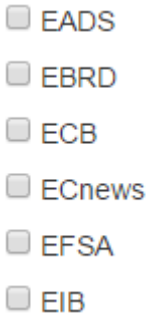
Advantages:

- The user can focus on the categories he want to compare. This filter allows him to compare categories in smaller resolutions and accurate the comparisons.

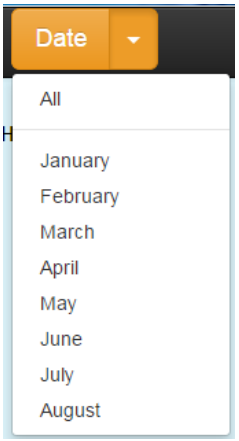
- The categories are arranged in alphabetical order, so if you know which category you are looking for, it will be easy to find.

Disadvantages:

- It is very difficult to understand the sub category names and to know which main categories they belong. For example:



- We limit the user to select up to 12 categories. This is a restriction which is not a good thing to prevent a user to compare more than 12 categories.



Date - This filter allows the user to choose which dates to view.The user can choose a specific month or all months together in 2016. We chose in those options because we think that it is the most relevant options for users to know if there are breaking news.

Evaluation- Value

This filter can help the user to focus on specific dates. This is very important because the main gaol is to compare categories in specific period of time. This filter gives the user free choice which dates are most relevant for his needs. Because of the importance of this filter, we allow the user to choose date in any moment.

Self reflection

Advantages:

- We give to the user the freedom to decide which dates relevant to him.
- All the above filters integrates with this filter- generic.

Disadvantages:

- Unable to see comparisons for a single day.
- The user cannot select a dates in dynamic way for example: from 03/01/2016 till 03/03/2016 .

User experience

When the user sees the visualization for the first time, he sees the Compare topics filter in front of him. If the user wants to choose a certain filter, he need to select filter and then press 

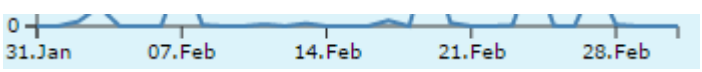
.

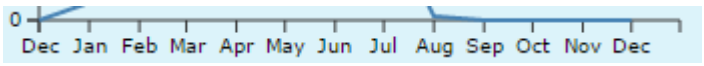
In every choice of filter, the user knew which visualization he sees in the header:

Self reflection in general

Advantages:

- The background is related to the subject so the user knows at first glance what the topic of visualization.
- The choice in small multiples serves our purpose in a good way.
- The filters are generic and can conform to any type of user and any kind of purpose.
- X axis- that shows the date ,written in a very informative and useful way:

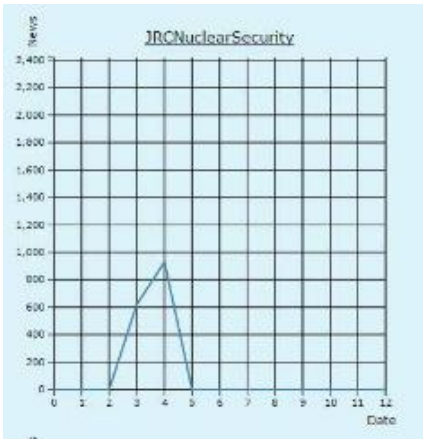




- The Y-axis is always constant in all graphs that are presented at a given moment and it gives a reliably comparison between the categories. It also changes dynamically depending on the number of news in each comparison.
- In regular home monitors we can see all the graphs in one picture and this maximize our gaol.

Disadvantages:

- In small monitors- the graphs are not placed in a one picture. For comparing all graphs that are chosen, we have to scroll down the page, and then we do not see the graphs next to each other and because of that the idea goes wrong.
- There are **no** grid in the graphs:
  This prevents the user from knowing exactly the news items number. In addition, the popularity of graphs can be misleading in this way. The following view could be more informative:



- You can not see changes in a specific category for some periods of time. For example: shows a number of graphs of politics spread over January-February and March.

References

http://metanit.com/web/d3js/4.3.php