Data Visualization - Assignment 4
FRENCH TEAM
Identifying Fake News with BuzzFeed News

Julien BALOCHE-ROUSSEAU, Laure DASSY, Lea MOUKARZEL, Alban VEAUTE

1. Project description

Over the past four years, BuzzFeed News has maintained lists of sites that publish completely fabricated stories. As they encounter new ones and debunk their content, they add them to the list. To produce this story, they used this list as well as some sites brought to their attention by fact-checking websites.

Currently in the middle of an election period, fake news has been a sociological phenomenon for almost 10 years now, and its impact has been growing steadily. This period is particularly conducive to the propagation of such news, as evidenced by the last American presidential campaigns marked by numerous fake news.

Thanks to this dataset, we wanted to study different aspects of this sociological phenomenon, whether it be their origin, their chronology or their typology. In a way, thanks to our project, we are trying to draw up a draft identity card of fake news. For our study, we focus on data available for 2018: the year of the elections in the United States, a time period that we believe will guide our research.

For the realization of this project, we mainly used three tools: OpenRefine for data cleaning and preparation, python and the Altair library for the realization of the various graphics and finally Streamlit for the creation of our web interface.

2. Feature description

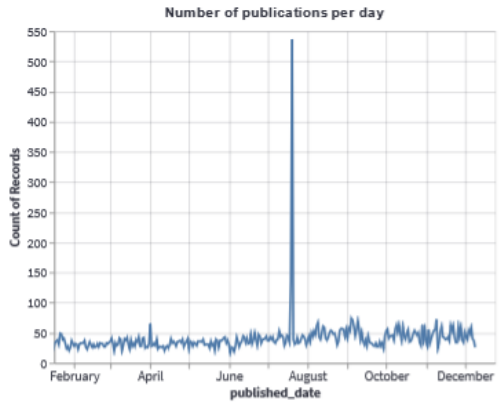
Our tool consists of 10 visualizations, specially designed and arranged to answer all the questions we have asked ourselves during previous assignments.
At the opening, our tool first offers the user an overview of the fake news situation in 2018.
Aiming to be as interactive as possible, all the graphs are connected either to each other or by selectors on the left side of the screen.
The idea behind this is to make the tool easy to use, while allowing the user to go into detail and inspect the impact of different categories or origins of fake news.
Thus, the user starts with an overview of the number of fake news and Facebook engagement generated over the year, with the possibility to zoom in on certain periods.
Then, the next 5 views are connected to the selectors on the left and allow the user to focus on the origins and/or categories of fake news that interest them.
Finally, the last 3 graphs offer a monthly view over the year 2018 of the most important fake news sources. These 3 elements are connected to each other, allowing to explain in detail the relationship between categories and origins while keeping an eye on the temporality and the reactions on social networks.

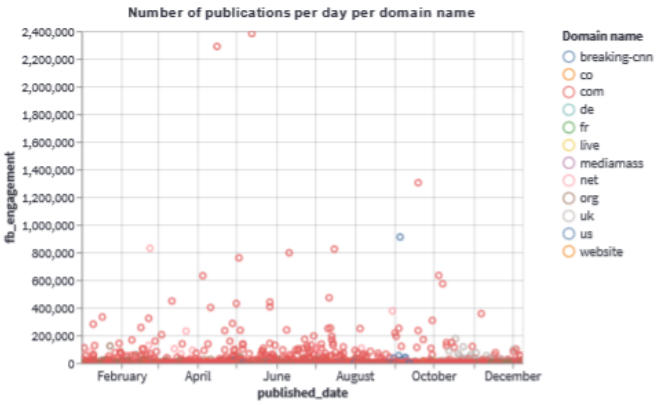
3. Interesting findings

We can divide our findings from this dataset into three main parts.

First, we look at the activity of fake news published throughout 2018. Thus, we can see whether a certain period of the year was particularly prone to this phenomenon.

We can see that engagement is relatively stable with a general growth over the whole year. Over the course of 2018, no major phenomenon has significantly changed the pace of the fake news phenomenon. However, it is advisable to remain cautious about this analysis as it concerns only the fake news identified by BuzzFeed. Finally, there was a peak on 20 July, which we will come back to in more detail in the rest of this project.





Then, we can observe in detail the origin and/or the category we are interested in. We will then have some focused insights from five graphs. The first one will allow us to understand which categories can be found in one origin. Thanks to this graph, we will see whether a category dominates an origin or if it is more balanced. Moreover, we can compare several origins by selecting them in the left sidebar.

We can also see how origins and categories are impacted by the days of the week, or the length of titles. These views are interesting when we want to dig into one or several specific categories/origins and get some insights.

For this example, we selected the Category “politics” and the origin “yournewswire.com”

Finally, the last part of our tool allows us to quickly highlight, on a monthly basis, the main categories and origins that generate the most reactions on Facebook, as well as the number of publications per day.
It also allows you to see, within a category, how the reactions on social networks are distributed according to the different origins and vice versa.
Furthermore, it also allows you to see which origin/category duos create the most engagement on the internet according to the number of publications.

4. Member contribution

We have tried to divide the work equally for this project. First we decided on the ideas, concepts and designs together. We have collaborated for all of the tasks, the responsibilities were divided as follows:

a. Julien Baloche-Rousseau & Alban Veaute: Altair graphs + report
b. Laure Dassy & Lea Moukarzel: Streamlit + video