Global Warming & Natural Disaster: Is there a correlation?

Marta Carlos, 47592 *
FCT NOVA

Vicente Almeida, 47803 †
FCT NOVA

Marcin Konieczny, 56551 ‡
FCT NOVA

Figure 1: *Hurricane Florence, Sept. 13*

Abstract

Global warming has been a reason of concern over the last decades. The need to change the way we live, and how we can adapt to a world with a considerable reduction of greenhouse gas emissions, is even more clear: we have barely over a decade to change our ways before it becomes irreversible [United Nations 2019]. In the light of these recent news, we decided to study if there's a correlation between this global phenomena and natural disasters. With the help of Tableau, we work on 7 datasets and propose interactive ways to visualize the available information through the use of Tableau Dashboards, allowing whoever uses these dashboards can get their own conclusions.

1 Introduction

Earth - our dear home. It should be in the interest of all people to take care of it and since one of the most recent dangers which concerns our planet is a global warning problem, everyone should be striving to understand what this implies and what we should aim to do.

This atmospheric phenomenon is caused by the increased amount of carbon dioxide and greenhouse gases emission in the atmosphere and it is proved that global temperature has risen since 1880 [NASA]. Furthermore, we have started to see the consequences of global warming is having on our the planet and a big change needs to happen soon, given the time restraint and how the human life style affects the rest of the ecosystem [United Nations 2019].

It is in the concern of many to truly understand what possible interrelationships could there be between climate change and natural disasters. We believe this understanding could help change the minds of many and how they take action.

In our work we aim to investigate whether global warming implies that disasters will tend to become stronger and more frequent. We also compare some observations from Nature (like CO_2 emissions, storms' and floods') to look for observable patterns.

Unfortunately worldwide research on the subject is difficult as long as we don't have access to the global data. Because of this setback, we decided to get an overall feel of the world with temperatures and CO_2 emissions but focus on the United States when it comes to analysing possible patterns in natural disasters, given how this information is documented through out many datasets.

In order to study global warming and natural disasters we start by listing questions regarding the subject, followed by a short description of what datasets we've worked with.

Before preparing any visualizations, we studied of what has been done since there is a need to actually have some sort of an under-

*e-mail:m.carlos@campus.fct.unl.pt
†e-mail:vo.almeida@campus.fct.unl.pt
‡e-mail: marcin.l.konieczny.student.put.poznan.pl

standing of the science behind global warming, natural disasters and what visualizations have already been proposed by experts in the field.

In the end, we propose some visualizations for the data and interactive dashboards that let a user get their own conclusions.

2 Research Questions

In order to understand if there is any correlation between global warming and natural disasters, an initial step is understanding how exactly is Global Warming affecting Earth. With this in mind, we present the following questions:

1. How have global temperatures changed in the last years?
2. How are drastic temperature changes related to CO_2 emissions? Is there a pattern?
3. How are high temperatures affecting the glaciers?

Due to the easiness in finding USA related data, we next focused on understanding the behaviour of natural disasters in the USA. We propose the following questions:

1. Has the frequency of natural disasters increased in the last decades?
    - Has the storms' frequency increased?
    - Has the floods' frequency increased?
    - Has the fires' frequency increased?
2. Has the intensity of natural disasters increased?
    - Has the intensity of storms increased?
    - Has the intensity of floods increased?
    - Has the intensity of fires increased?

3 Datasets

3.1 Cumulative CO_2 Emissions

A simple dataset with the cumulative CO_2 emissions, from 1751 to 2016. It's information can be summed up as:

- **Entity:** the name of the country or region
- **Code:** the FIPS identifier for the country (missing when the *Entity* is a region)
- **Year:** the year of the CO_2 measurement
- **Cumulative CO_2 Emissions:** The cumulative amount of CO_2 emission, in tones

Comprised of 62 thousand rows, this dataset contains some null values, especially until the second half of the XX century, which is to be expected from it's time range. The dataset is available at [1] along with some other analysis related with CO_2 and greenhouse gas emissions.

3.2 Temperature change

Contains data about annual temperature anomalies, with the baseline period of 1961 to 2017. It's a dataset with 12K rows and 14 fields, such as:

- **Area:** Area corresponding to the country (and area code)
- **Value:** the value corresponding to the temperature anomaly

Further details on the dataset can be found in the appendix B

3.3 Global Glacier Recession

The International Association of Cryospheric Sciences provided this dataset that contains over a million observations. The most important values of this dataset will be:

- **Glacier name** The name of the glacier
- **Geographic latitude and longitude:** coordinates of the glacier
- **Year of measurement**
- **Thickness:** the size of the thickness in meters
- **Elevation:** the height of the glacier in meters
- **Glacier country:** country which the glacier belongs to

The dataset can be found in International Association of Cryospheric Sciences. More information is avaiable on C

3.4 Floods

G.R.Brakenridge, at the Dartmouth Flood Observatory, aggregated data from "news, governmental, instrumental, and remote sensing sources"[2], from 1985 till 2017, to make this small (4.5 thousand rows) but rich dataset about floods on a global scale. It includes information such as:

- **Country:** the name of the country affected
- **Detailed Location:** the city or region of the country affected
- **Start Date:** the date in which the flood started
- **Duration:** duration of the flood, in days
- **Main Cause:** the reason why the flood happened (ex: heavy rain)
- **Magnitude:** the magnitude of the flood, calculated with duration, severity and affected Area

The dataset contains even more relevant details which are available in appendix E. There is also a shapefile included alongside the data.

3.5 Storms

A dataset that contains information on hurricanes and typhoons on the North Pacific and Atlantic area, from the National Hurricane Center. Each occurrence is divided in six-hourly recordings and contain pertinent information such as:

- **Location:** the location where the storm occurred, in coordinates
- **Status:** the status of the hurricane (for exemple *TD* for a tropical cyclone of tropical depression intensity)
- **Maximum Wind:** the strongest wind registered, in knots
- **Minimum Pressure:** the lowest pressure registered, in millibars

[1] https://ourworldindata.org/CO2-and-other-greenhouse-gas-emissions

[2] http://floodobservatory.colorado.edu/Archives/index.html

The time frame of this dataset is between 1851 and 2015 and together they are comprised of 75 thousand records. More information, like metrics regarding the wind, can be viewed in the appendix F.

3.6 USA Wildfires

This dataset contains around 180K records of USA wildfires, ranging from 1992 to 2015. Holds information about day, time of day of when it started (when it was discovered) as well as when it was considered to be contained/controlled. With data about its' latitude and longitude, as wells as fire size, size class (from A to G) and even the cause of said fires, we can truly get a feel of the magnitude of these wildfires.

In total, this dataset has 39 fields and regarding specific fire information we have:

- **Fire IDs:** fields regarding different a fire's identifiers
- **Fire location:** latitude, longitude, state and county
- **Date:** date, time, day of year and year regarding the discover of the fire and also day and day of the year of when the fire was controlled/contained
- **Cause:** code and description for the (statistical) cause of the fire
- **Other**

A more detailed information about the dataset's fields can be found in appendix G and it is available on Kaggle.

3.7 USA State Information

This dataset was obtained with R and its' dataset packages. The main purpose of this dataset will be to normalize the 3.6's dataset.

- **Area :** information regarding a state's ara, in square miles
- **Region:** a state name
- **State abb:** a state's abbreviation
- **Other:** information about life expectancy, illiteracy in the 1970's. income, population, etc

4 State of the art

One key aspect in presenting a good visualization is knowing what has already been done in terms of visualizations related to our work. This step is important since it does help in understanding what works and what doesn't. But before analysing any visualization, it's very important to understand what's the science behind those visualizations and also what we're aiming to answer with our own visualizations. Due to this, we divide the state of art in two subsections:

1. Understanding the science behind climate change and natural disasters
2. A study of what visualizations have been done in the field of climate change and natural disasters

4.1 The Scientific Background

It's thanks to CO_2 that planet Earth has surface temperatures that allow life as we know it. CO_2 is also known as the greenhouse gas since it emits and absorbs thermal radiation, creating what we call the greenhouse effect [Ritchie and Roser 2019]. Unfortunately, the raise of CO_2 emissions caused by the modern human lifestyle has now disturbed the cycle and led to what we call global warming.

The Industrial Revolution changed the world in many ways, from the way of living of the general population to, unfortunately, the start of the rise of global temperatures. Any change in the average global temperature means a change in the received and radiated energy of the planet. When taken into account the amount of heat it takes to warm the oceans, the rise of 0.8^o Celsius (since 1880) is very worrisome [Earth Observatory].

Given the concerns of this rise in temperatures, one wonders if we can already observe its impact, more specifically, if we can observe a change in natural disasters, like fires, floods, storms, etc. and question their correlation to the global warming.

Fires are mostly influenced by fuels, climate-weather, ignition agents and people [Flannigan et al. 2006]. Thus the impact of climate change on the wildfires' increase in frequency (if there's an increase) and in intensity is put into question. Since changes in temperatures are easily observable (and they do play a big role on the activity and prevision of wildfires) [Yang et al. 2017] proposed methods to improve predictive services on wildfires using timely weather information.

It is projected that the increase of summer drying over most mid-latitude continental interiors (and associated risk of drought) will not only decrease water resources (in quantity and quality) but also increase the risk of forest fires[Aalst 2006]. Additionally, in the USA. the seasonal length[3] of wildfires has increased significantly by 64% (78 days) when comparing 1970-1986 with 1987-2003[Westerling et al. 2006].

Besides the decrease in winter precipitation, summer droughts are a given a to happen and these droughts affect the flammability of forests' fuels. It has also been discussed how early snow melting can cause longer dry seasons, resulting in bigger opportunities for large fires to occur [Westerling et al. 2006].

Similar to fires, glaciers are also influenced by the increase of the Earth's average temperature [M. Ruby, C. O'Neill 2019]. The long term measurement of glaciers' balance changes shows that world's ice resources are rapidly decreasing and in all cases, long records have shown a general recession [Krimmel 2002]. In the United States, Alaska's glaciers lose about 75 billion tons of ice each year and over the last 40 years thinning has been on the order of 50 to 100 meters at lower elevations of glacier occurrence and about 18 meters at higher elevations.[Dennis O. Nelson 2019] It was proven that glaciers in western North America, excluding Alaska, are melting four times faster than a decade ago.

As the observation frequency is increased, most of the records show evidence of the general retreat, however, glaciers placed side by side geographically sometimes behave differently. It may be caused by complex interactions between the effects of elevation, latitude, exposure, basin area-elevation distribution, the variables of climate and the individual glacier's dynamic-response characteristics [Krimmel 2002]. Only a few glaciers are actually advancing (in locations that were well below freezing, and where increased precipitation has outpaced melting) [PRESS 2009].

In 1998, [Doornkamp 1998] considered what could be the causes for the natural hazard that are floods. While there was already a talk about the effects of global warming on the frequency and magnitude of floods, these were not strong enough to justify the changes. [Doornkamp 1998] instead attributed much more responsibility to other human causes like man-induced subsidence, the supply of silt

[3] Seasonal length is the time between first wildfire discovery (reported) and the last wildfire

on the coastal areas and the occupancy of floodable land and even the creation of flood defences which produced unexpected negative effects.

In a little less than decade the situation seemed to change: the United States suffered an unusual number of hurricanes and the Intergovernmental Panel [on Climate Change 2007] indicates that in response to global warming, there will be a surge in hurricanes and sea the sea level rise. [Mousavi et al. 2011], in response to these projections, studied the impact of these changes. It is important to note that wind storms or hurricanes and coastal floods have a high chance of occurring at the same time, so it's valid to assume that a rise in hurricanes will lead to a rise in coastal flooding.

[Mousavi et al. 2011] concentrated their analysis on a hot-spot for high coastal surges in the United States and the Gulf of Mexico (more specifically the city of Corpus Christi). By choosing three of the major hurricanes of the area's history, they ran simulations to evaluate potential future floods due to hurricanes in the next eighty years. They concluded that if these global warming projections prove to be true, hurricanes and flooding will be substantially more intense and damaging to humans, both economically and socially.

4.2 The Visualization Background

Given the importance of climate-change and the greenhouse effect, visualizations help the general public truly understand the science behind it.

When it comes to visualizing temperature information, this is a theme that has been widely done in various ways. Probably one of the most well known temperature visualizations is the temperature mapping to color, all over the world. [NASA/GISS 2019] presents this in a timeline, starting in 1884, ending in 2018 (figure **??**).

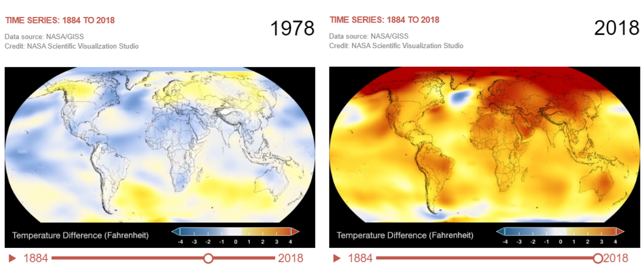


Figure 2: *Global Temperature: Time series 1884 - 2018 by NASA/GISS*

While this type of visualization does show change overtime, it gives a very "general" information since the user can't directly interact with the visualization besides browsing the years' timeline.

On the other hand, in a more interactive way, [InMeteo 2019] developed a very interesting and dynamic application about meteorological weather over the whole world. This application lets the user choose exactly what variable (temperature, wind, snow coverage, etc) they want to map with color and also lets the user zoom in a specific city, allowing the visualization of the specific temperature. The usage of color to map different variables, according to the user's will and the possibility of city's details are important aspects to take into mind when planning an interactive visualization.

Even though [InMeteo 2019] makes for a very interactive visualization, it does not allow for a easy or observable comparison of temperatures.

One thing are temperatures and another thing are temperature anomalies. When it comes to analysing temperature changes over the years, temperatures anomalies are more important to study than absolute temperatures [NOAA 2019]. This is because a temperature anomaly is the difference from an average temperature (average of 30 or more years) and with an anomaly we can clearly see when it was warmer or colder than the average.

The visualization of temperature anomalies proposed by [Lipponen 2019] shows the different anomalies values over the years, where bigger blue circles indicate colder than average temperatures and bigger red circles hotter than average. Even if we can quickly capture what the visualization is trying to convey (global warming is happening), the countries representation is not the best: it gives the user geographical information (countries' names) but it is difficult to visualize which countries are exactly colder/warmer since reading a country name is not the same as seeing it in a map.

With this in mind, in Chapter 5 we propose a mix of the typical temperature information provided by [NASA/GISS 2019] with temperature anomalies, where the user will control which year they want to visualize.

Regarding CO_2 emissions, [Ritchie and Roser 2019] present a map of the world, also with a timeline, which makes it possible to clearly see the start of the Industrial Revolution in Britain and its' progression over the world. The problem with this visualization is the choice of colors for the mapping of CO_2 quantities, since between 50 million tonnes and 5 billion tonnes are not easily distinguishable (Figure 5).

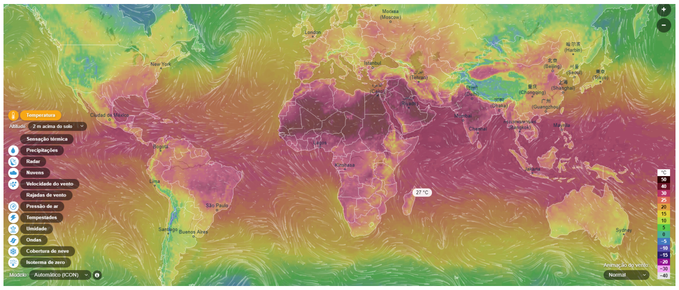
Another look at CO_2 emissions is its' division by source (Figure 7b) and annual emissions by world regions (Figure 7a). While these two visualizations present very interesting information by themselves, maybe it would be interesting to see how we could visualize the annual CO_2 emissions and their relation to temperature changes.

Concerning wildfires, [Gonçalves et al. 2019] (Figure 6 developed a special a new data visualization interface with the main goal of providing new, interesting and informative ways of displaying Portugal's fire data. These visualizations give a good understanding of what is happening all over the country, and when it comes to the district, one can specify if they want to see also details about fire causes. But when divided for each district, the graphic continues on showing the same graphic on the right (Figure 6). It's true that this graphic has information about every district but an alternative way would be to show the total burnt area (by cause or not) over the years, for that district. In the Proposal (5) we take this idea and apply it to the USA wildfires.

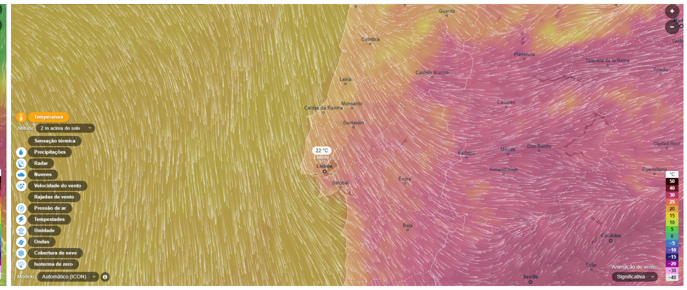
There has been already some work done with the dataset we're using for wildfires on Kaggle platform, more specifically, [Walters 2017] compares the number of wildfires by state and afterwards normalizes the data so it's possible to directly compare wildfires across states. The need for normalization comes because states vary a lot in terms of areas and normalization is used to help fully understand if one state has a truly significant larger amount of fires when compared with another. This information by state will be interesting to correlate with the cause behind the fire.

When it comes to floods we found an interesting visualization from the Federal Emergency Management Agency [FEMA], about the historical risks and costs of floods . It's composed by a map graph of the selected state, a 2D line graph with about the flood events over the years and a bar chart regarding the economic cost of floods in that state.

We find that the relation between the map graph and the line graph works really well, and the latter being split into counties easily allows us too find patterns in the data. Its' weak point could be the



(a) *Ventusky zoomed out*



(b) *Ventusky application zoomed in Portugal*

Figure 3: *Two possible temperature visualizations with Ventusky*

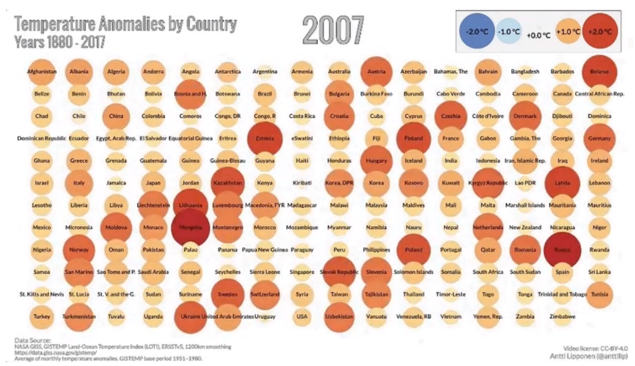


Figure 4: *Temperature anomalies 1880-2017 by country.No matter how you visualize it, it looks scary! by Antti Lipponen*

overlapping of lines, but the interaction with the map allows us to intuitively select the desired county and look at the details on the line graph. It would be interesting if there was a way to look at the entire US map at one time, in order to quickly find the states which are hot spots for this kind of natural hazard, but we understand that it would probably not translate well in the line chart.

When looking through hurricane data visualizations, one will come across many related with the trajectory or the wind direction in a map graph. However, we found this peculiar, at least the for the subject at hand, visualization regarding an historical overview titled "Thirty years of Atlantic Hurricanes" by [Canipe]. This yearly summary contains small polygons which represent the storms, color coded by category. Those "noteworthy" enough also have a small label with their name and maximum wind speed.

The graph requires an initial processing of the visual concepts but once that is achieved its' biggest strength is evident: it serves as a great tool to find patterns throughout the years. For example, one can clearly notice that there is a concentration of occurrences in the middle of the year, which is most likely due to the higher temperatures. It suffers, however, a grave problem of overlapping when one wants to look into the details of a specific hurricane, especially due to the fact that there isn't a click function to hold the information. We would like to try to reproduce this graph on Tableau but attempt to solve the problems we just referred.

Like we said in the Datasets section (section 3), our plan is to combine the sea level dataset with other like the one about the glacier, so we would like to discuss a visualization that has a combination of elements that we are studying. However we came across an interesting self contained animated graph by [Pluck], which illus-

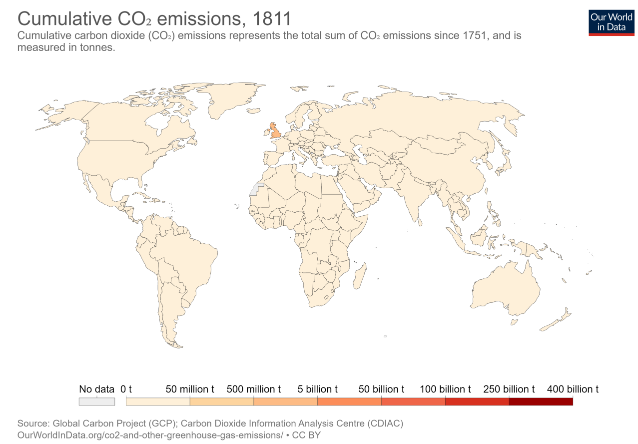
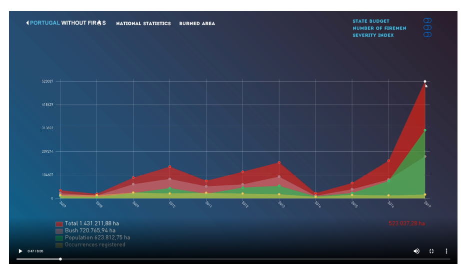
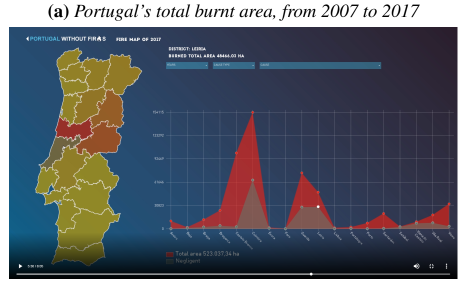


Figure 5: *Cumulative carbon dioxide (CO_2) emissions represents the total sum of CO_2, emissions since 1751, and is measured in tonnes.*

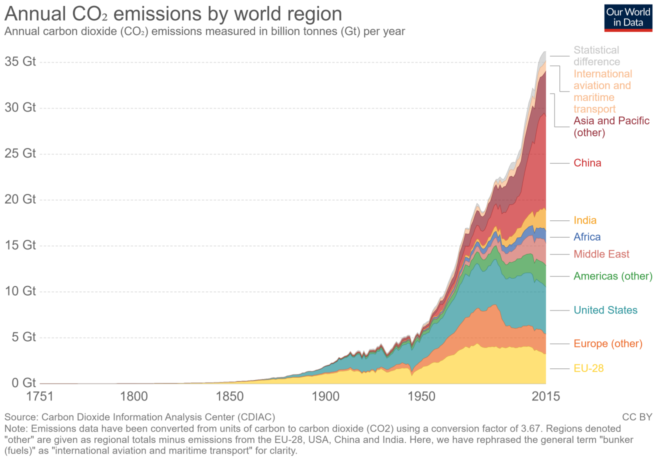


(a) *Portugal's total burnt area, from 2007 to 2017*

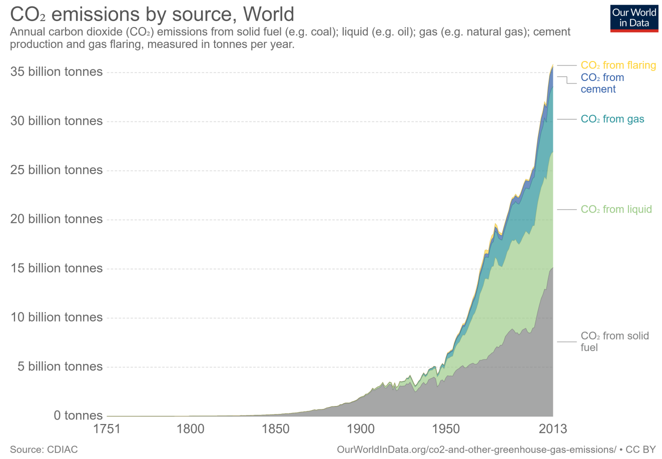


(b) *Burnt area by district, 2017*

Figure 6: *'Portugal without fires', a data visualization system to help analyse forest fire data in Portugal*

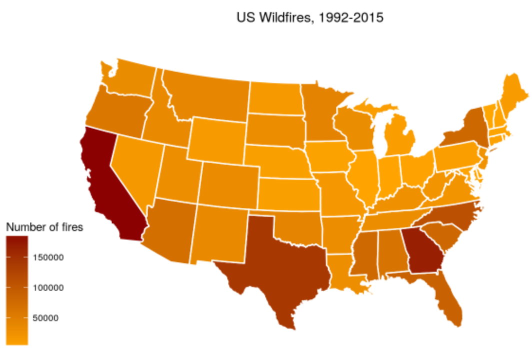


(a) *Annual carbon dioxide (CO_2) emissions measured in billion tonnes (Gt) per year*

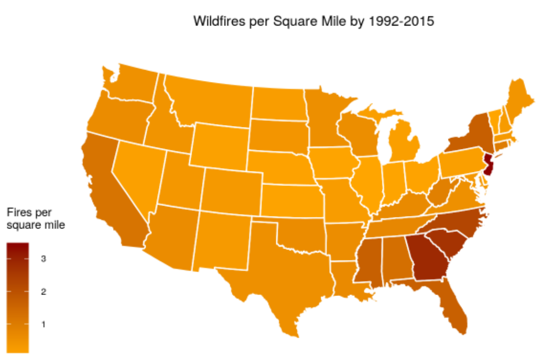


(b) *Annual carbon dioxide (CO_2) emissions from solid fuel; liquid; gas; cement production and gas flaring, measured in tonnes per year.*

Figure 7: *CO_2 emissions*



(a) *Wildfires by state*



(b) *Normalized data by state- Wildfires per square mile*

Figure 8: *Analysis on USA Wildfires, Kaggle dataset*

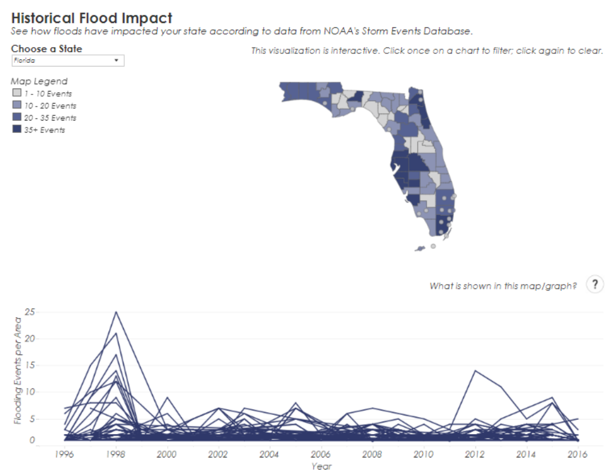


Figure 9: *The state of Florida with it's counties color coded by the amount of floods since 1996, along with a line graph of the occurrences withing the same time frame.*

trates the sea level rise over the years. Every decade another layer is added which enable us to process the data in a relative way to the layers of previous decades. However, this animation at first seemed really appealing because of it's resemblance with a wave, we later found out they are actually the monthly values being added. Another point, which is not necessarily a critical error, is that the color range doesn't make the distinction between decades clear enough.

5 Proposal

ur proposal can be clearly divided in two parts. On the first one we will try to build visualizations that answer our initial set of questions regarding CO_2, temperature and glaciers. On the second we will explore the natural disasters' datasets both in isolation and combined with others, in search of not only the questions of 2 but also new ones, some of which were introduced by the research made for the State of the Art.

Like it has been explained before (in 4), temperature anomalies are better than absolute temperatures, thus the analysis of temperature changes required a preparation of both the Cumulative CO_2 emis-

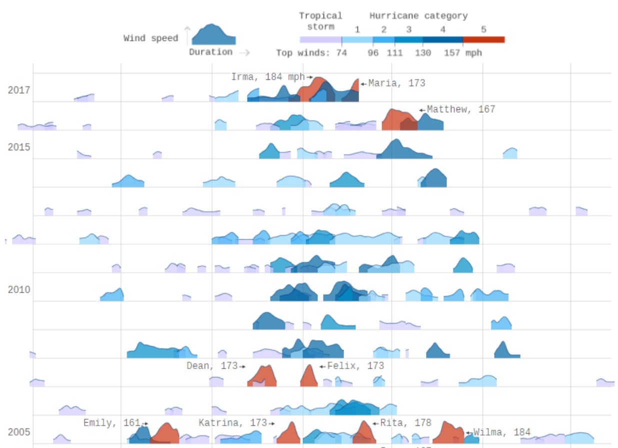


Figure 10: *A 12-year segment of the chart with its' caption. As mentioned in the text, there is a clear concentration of occurrences in the middle of the year.*

sions 3.1 and Temperature Changes 3.2. These datasets were joined by *Year* and *Country/Entity/Area* so it could be possible to analyse the correlation between these two phenomena over the years, for each country.

A problem we faced with this dataset was some inconsistent names for Russia, were it divided into USSR and Russian Federation. Here we decided to combine records with this problem with country name as "Russia". However, we acknowledge this is not the correct way to deal with these types of problems in the data, but for the purpose of this project, we opted for a simple solution in order to focus on other aspects.

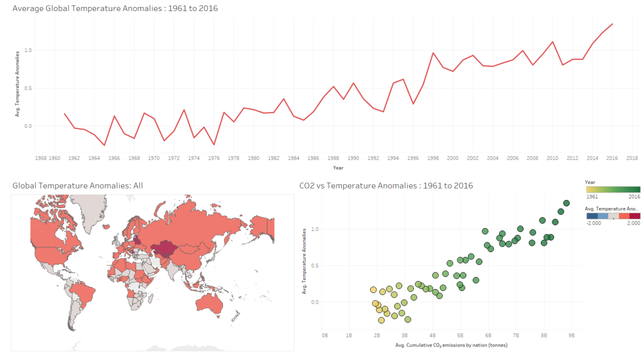
A similar problem occurred with records labeled as "Africa" and "Americas". For these records we decided to remove from the dataset, since we already had data for each country in these regions/continents.

Inspired by [Lipponen 2019]'s visualization, we chose to map temperature anomalies to color. But, contrary to [Lipponen 2019]'s, where we couldn't fully comprehend the temperature variation from country to country, we also used a choropleth map since visualizing a region is always better than mentally checking where it is (in the world).

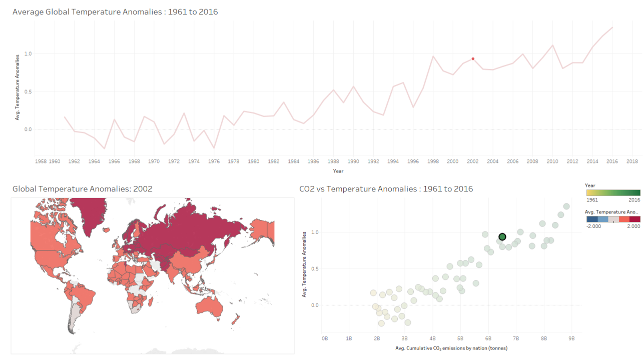
To enhance the information provided by the latter, we chose to add a complementary line chart with anomalies plotted from 1992 to 2017. With this addition, a user can chose a point in the graph (which represents a year) and it will change the information of the choropleth map by filtering the anomalies by the chosen year. The combination of these two different visualizations, the user will benefit from a better understanding of how temperatures have changed over the years, as well as how it has affected each country individually.

As for the correlation between CO_2 emissions and temperature anomalies, we opted to graph a scatter plot with these two variables. We mapped the average of anomalies to the y-axis, the average of cumulative CO_2 emissions to the x-axis and years to detail. The years were mapped to the visual variable color, in order to analyze the possible correlation of the two parameters across time.

These three visualizations, all combined together (Figure 11, form an interactive view of anomalies, CO_2 emissions and their possible correlation.



(a) *Dashboard without a filter.*



(b) *Dashboard filtered by year.*

Figure 11: *Dashboard with interactive views of temperature anomalies and CO_2 emissions.*

After temperatures and greenhouse emissions, we took a look at the geographical distribution of the glaciers data, which is shown in **??**. To achieve it, we created a point map, where each point represents a glacier from our dataset. Additionally, we mapped temperature anomalies to color so that we could show glaciers that are exposed to the biggest temperature anomalies.

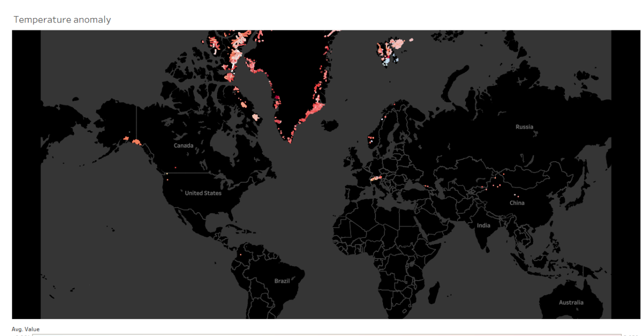


Figure 12: *Dashboard1*

The second dashboard we built aims to showcase the glaciers' properties, mainly their thickness and elevation. We made two visualizations based on point maps to represent those two features, both in color and size. The last visualization is an area graph meant to show how the elevation and thickness correlate.

Still in the topic of glaciers thickness, the last dashboard is constituted by two visualizations: the first one, a choropleth map repre-

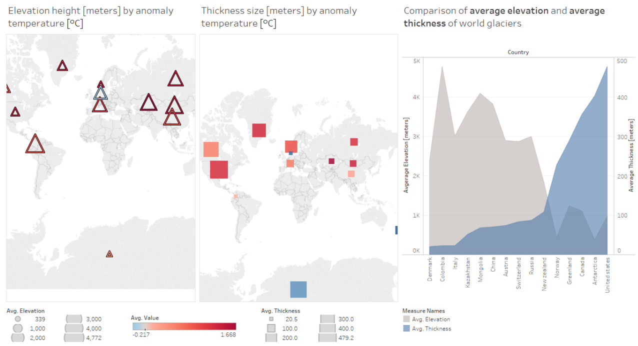


Figure 13: *Dashboard2*

senting the average thickness of each country glaciers; the second one, a historical overview of the average values. When selected a country on the first visualization, the second one is filtered to show the selected country's values. When a point is selected on the second one, it shows all the countries who had glaciers recorded on that year.

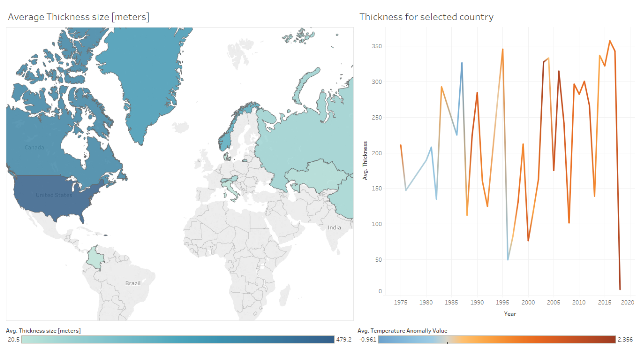


Figure 14: *Dashboard3*

Another dashboard touches on USA Wildfires information. Inspired by [Walters 2017]'s work, we decided to work with fires' duration, which could be related to a fire's intensity. To get this information, we used the fields *DISCOVERY_DATE* and *CONT_DATE* to calculate the total number of days a fire was burning . Unfortunately, both these fields were in a Julian data type, making it necessary to convert it to a *dd/mm/yy* date and only after this step we were able to calculate burn time by:

$$Burn_t = Contained - Discovered$$

This calculation ended with missing values because some fires have no registration of the date they were contained. We decided to exclude these cases and map the remaining results to a choropleth map, to be then filtered by year and fire cause. Furthermore, with the new calculated field, we wondered if there was any correlation between burn time and the overall fire size.

Sadly, after using a scatter plot with these variables, we couldn't visualize a perceptive pattern in the data. To overcome this, we used bar charts of Average *Burn Time* and Average *Fire Size* for fire causes and plotted them side by side.

To visualize a possible correlation with temperature anomalies and wildfire data, we started by calculating the global average temperature anomalies, so it wouldn't be separated by years. This step was necessary because otherwise the join with the 3.6 dataset would result in a lot of repeated information (every year for every country) which, when joined with the wildfires dataset, would have serious repercussions on performance.

After combining this data (by the *Year* fields), we created two different scatter plots: average global anomalies and average burn time, from 1992 to 2015; average global anomalies and average fire size, from 1992 to 2015.

In contemplation of possible seasonal patterns and the causes behind fires, two different barcharts were plotted. Both bartcharts are divided by fire cause but differ in the way they show information, namely, one shows a monthly distribution of fires and the other a comparative yearly overview of fire causes.

Motivated by the visualizations developed by [Gonçalves et al. 2019], we combined this dataset's visualizations into one dashboard (figure 15) , where by selecting a state (on the average burn time choropleth map) the other 3 visualizations are filtered by that state. Another filtering option, with analogous effect, is selecting a year in the scatter plots. The user may also choose to select a cause and, in the choropleth, it will be shown the average burn time of fires originated by the chosen cause.

Also in reference to [Walters 2017]'s work, we understood the need to normalize data. Every state in the United States has a different area so, to make comparisons between states normalization is needed. To achieve this visualization, we joined the 3.6 and 3.7 by the fields *STATE* and *state.abb*.

After normalizing the data, we were able to have information regarding the number of fires per square mile, from 1992 to 2015. This value is calculated at the state and year level, which makes it possible to form a choropleth map.

$$Fires_{persquaremile} = StateFires/StateArea$$

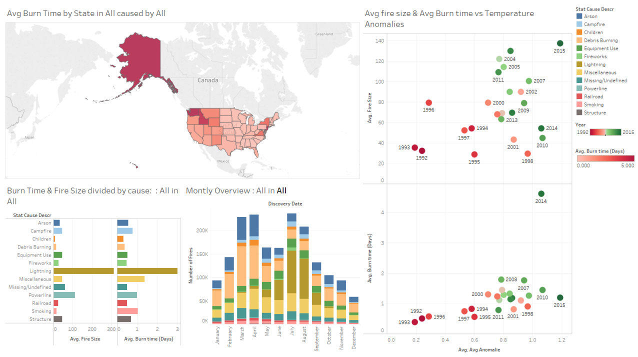
To boost this visualization, we thought it would be interesting to visualize big fires and their corrrelation to their cause, but instead of using barcharts like before, we built a density map for fires that had a size equal or bigger than 100 acres (with fire class D - G).

To complement these two maps, we graphed the number of fires over the years, separated by fire cause, with the purpose of having an overall view of not only the distribution of number fires over all years, but also an overview of what caused said fires.

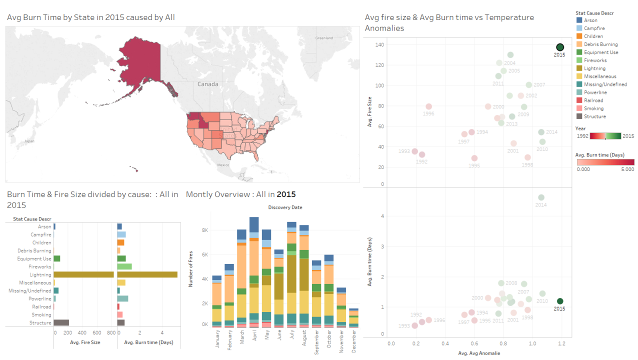
These three visualizations were put together in a dashboard (figure 16)and by clicking on the graph of fires over the years, it will filter the other two maps by the selected year.

Regarding the storms dataset, after we united the Atlantic and Pacific files into one and after we analyzed it's fields, we decided to approach the intensity aspect by relating the strongest winds with the lowest pressures. We also learned from the metadata that the hurricane status is classified by it's wind intensity, so we thought that it would be a nice addition to the visualization. We also took advantage that for the same storm there might be multiple rows describing it at different times, by enabling the user to click any storm in order to highlight all of it's stages.

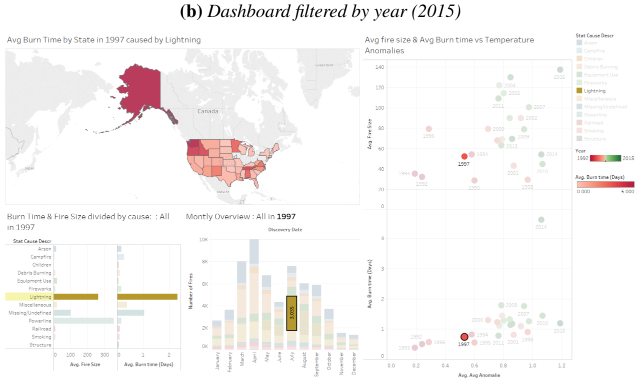
We then addressed the frequency question, starting by checking what the monthly distribution of storms looks like, and if there is a well defined "hurricane season" across the board. We achieved this by relating the amount of unique storm id's with the months, in a line chart. On top of this we decided to color code the line chart with the hurricane status, so that when the user clicks on a certain type of storm, the occurrences with that status will be highlighted



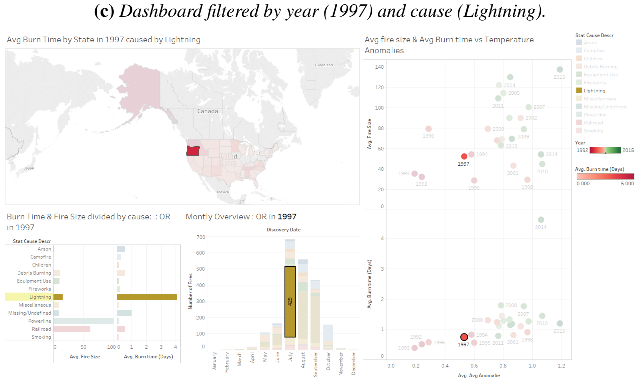
(a) *Dashboard with an interactive analysis on the Wildfires dataset*



(b) *Dashboard filtered by year (2015)*

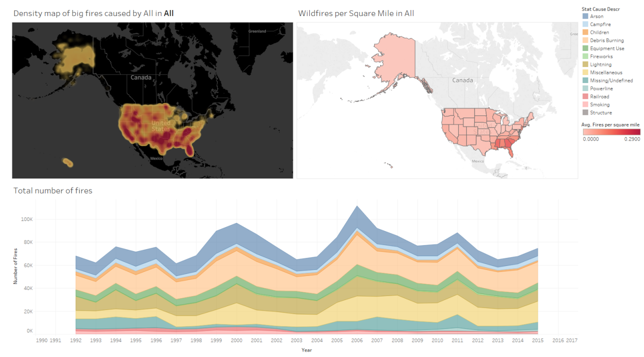


(c) *Dashboard filtered by year (1997) and cause (Lightning).*

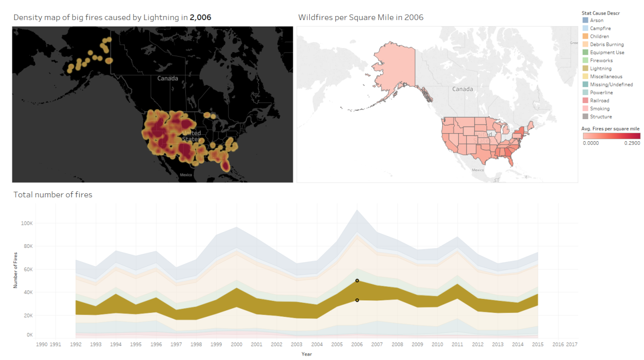


(d) *Dashboard filtered by year (1997), cause (Lightning) and state (Oregon).*

Figure 15: *Dashboard regarding USA Wildfire data and Average Global Temperature Anomalies.*



(a) *Dashboard with an interactive analysis on the Wildfires dataset*



(b) *Dashboard filtered by year (2006)*

Figure 16: *Dashboard regarding USA Wildfires and causes over the years.*

on the scatter plot. A user can also select a specific point on the scatter plot and a monthly distribution of the selected storm's year will be presented.

We also noticed our dataset contains information regarding the storm's coordinates. Since we already established that it's possible to establish a storm's "history" from it's records, we decided to build a map visualization with the paths of the natural hazards. We inserted the coordinates into the map and united them with the day of the observation. To represent the intensity of the hurricane we initially planned on using the status again, but encoding the maximum wind with both size and color provided better looking outputs. When the user clicks a point of the hurricane it produces the same feedback as when a point on the scatter plot is selected and, when a point is selected on the scatterplot, the year displayed on this visualization will be the year of the selected occurrence.

Lastly, we provide an overview of the dataset, over it's time range. We decided to simply stack the evolution of the number of occurrences, average strongest winds and average lowest pressure, through the last 17 decades. This visualization does not support any interactions with the others so it can serve as a constant reference point.

In the state of art we mentioned that when it comes to floods, hurricanes or storms in general are likely to have happened somewhere in the same time frame. Since we planned to start by looking into the datasets isolated, before comparing their values together, we took advantage of the *Main cause* field to try to find the yearly distribution of flood causes and overall what are the most common ones.

To achieve this we built a stacked bar chart, where each bar corre-

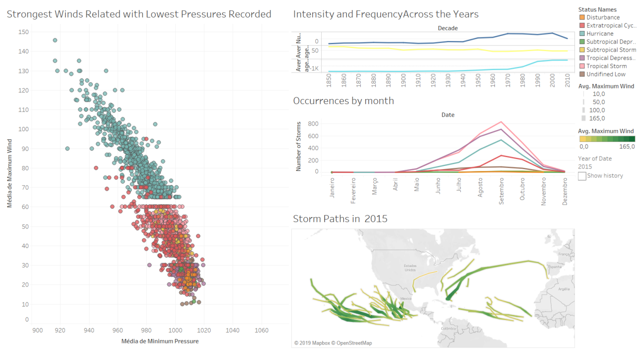


Figure 17: *Dashboard with interactive view of the relation between hurricane properties and their monthly and yearly distribution, along with a map that shows hurricane path in a certain year.*

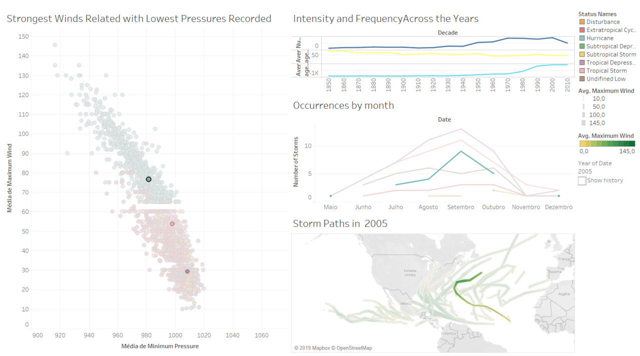


Figure 18: *When a specific storm is chosen it's status and path can be seen on the other graphs.*

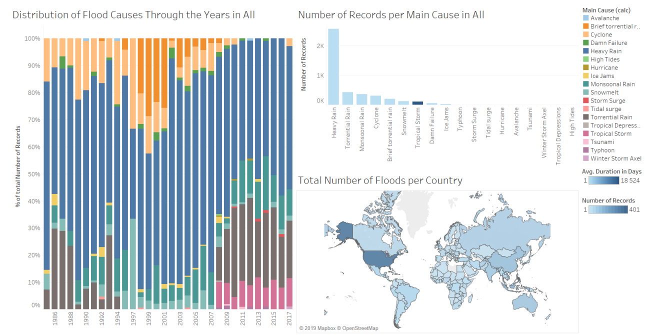


Figure 19: *Dashboard with interactive view of the relation between hurricane properties and their monthly distribution.*

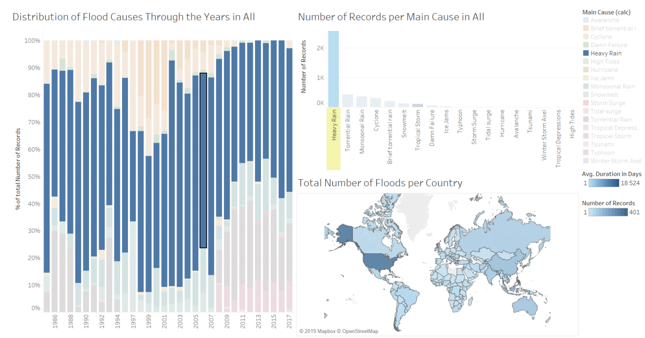


Figure 20: *When a flood cause is selected its highlighted in both graphs.*

sponds to the total records of a specified year, and the length of a color represents the percentage of occurrences with a certain cause.

For the top causes floods we did a normal bar chart relating these to the number of records. We also decided to add the average duration of these hazards to the color in order to see if there is any correlation with their frequency. In terms of interactivity the user can select any cause to highlight it in both graphs.

Finally we added a simple choropleth map with the total number of records from each country to provide an overview of the most affected places on earth. To increase compatibility with the first view we filtered out any value who's date was missing. This way, we can let the user select any country to filter out the information on the two other views accordingly.

For both of these visualizations we had to modify the majority of values on the main causes field. This was because they were too specific and didn't have a count big enough to justify increasing the visual complexity, in order to be grouped up by just their most primitive type like "Heavy Rain" or "Tropical Storms". We also tried to take advantage of a shapefile included with this dataset, which covered the are of the floods registered, but unfortunately we haven't found a way to take advantage of this file without sacrificing substantial visual clarity.

In the State of the Art we learned about the frequency of floods occurring along with storms. To dive deeper on this subject we decided to cross our datasets to see if there were any visible correlation between these two natural hazards. We elaborated two questions to analyze any possible links between the two: are there a visible correlation between the number of hurricanes and floods and is there any similarity between the intensity of the flood and the strength of the storm winds along the months of the year?

To answer the first question we built three visualizations of which two correlate the number of occurrences with the years. The first one does this in an indirect matter, it shows the distribution of the number of occurrences, from 1984 to 2015, with a simple line chart. The second one diver more deeply into the correlation analysis, with a scatter plot of the same values as the last visualization, we also added a trend line easily convey the presence, or lack of, a link between hurricanes and floods. The user can interact with these visualizations by selecting a point of interest, which will then highlight it's correspondent value on the opposing graph and highlight the year on the line chart.

The last visualization which approaches the relation between the number of occurrences of these two hazards is a monthly distribution analogous covers to the one made for the storms dataset, we speculate that if there is a "hurricane season", according with the pre-established supposed relationship between floods and hurricanes, there should also be a "flood season", or at least an interval coincident with the storm's where floods occur at an higher rate.

The second question was inspired by a visualization from the storms dashboard which showed the monthly distribution of storms along the months of the year, color coded by their status. In this case, we wanted to see if the strongest winds coincide in the same months as the most intense floods, so we did another line chart with their monthly distribution.

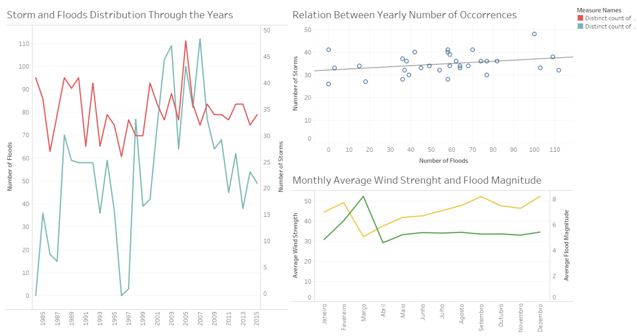


Figure 21: *Dashboard with interactive view of the relation between storms and floods, across multiple metrics.*

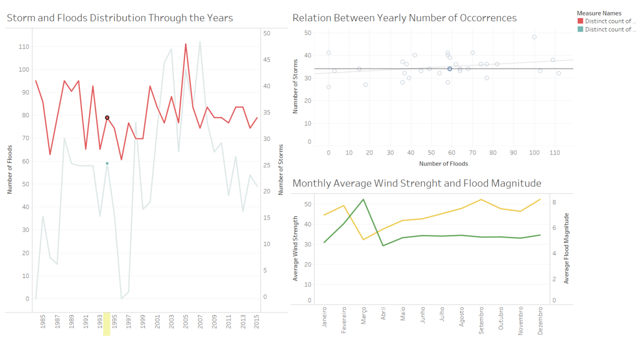


Figure 22: *When a specific year is selected on one of the views, it is also highlighted on the opposite view.*

6 Conclusions and Final Remarks

The visualizations about CO_2 emissions and temperature anomalies confirm what we were already expecting: CO_2 and changes in temperatures are correlated. In the scatter plot from figure 11a we can clearly observe how, as the years passed by, CO_2 emissions and temperatures increased. Another clear observation is how temperatures are getting warmer over the years.

Even with the strong correlation between CO_2 emissions and increase in temperatures, we can't as strongly visualize a correlation between temperature anomalies and fires' duration, as well as fire's size. Nonetheless it's possible to see how recent years have bigger and longer fires.

With the wildfires' dashboards it is also possible to draw some basic conclusions about fires, namely, how Lightning seems to be the main cause behind the biggest and longest fires. With the monthly overview, we recognize lightning's seasonality in summer time (figure 15). Also, when selecting the different years in the dashboard in figure 15 it's possible to see the average burn time increasing in all states

While we hoped to see strong correlations between summer and bigger numbers of fires, this isn't something we can conclude from the visualizations we presented.

Addressing the fires' intensity question, we can't give any solid answers, because of its' relation to factors we don't have present in our data, such as flame length and flame heat.

By analysing the graph about the number of fires over the years, unfortunately, we can't also draw any conclusions about the increase of fires' frequency.

From the hurricanes dashboard we can conclude that there is in fact, as expected, a strong correlation between the the wind intensity and storm's status. Two distinct categories can be identified in the first visualization of 10, clearly separated by a wind strength threshold. Above it there are almost exclusively Hurricane class storms and, below it the remaining values. In the lower cluster one can also identify strict thresholds separating the different classes, with the only exception being extratropical cyclone storms, which are evenly distributed along this section. Regarding the correlation between wind and pressure, although it's not as strong as the one between wind and status, one can still identify that these seem to be, in general, inversely proportionate, especially for the Hurricane storm status.[4]

We also find a clear answer to the question of "is there such thing as hurricane season?" and it's a resounding yes, as it can be seen on the "Occurrences by Month", there are close to none storms recorded from December to April when compared with the period between May and November, with a clear peak in September.

But have hurricanes become more frequent in the last years? Our data seems to indicate so, from 1940 to 1970 there has been large increase in hurricane numbers. This increase stabilized in the following decades, which leads us to believe that, even though storms have become more frequent, that high increase could be linked with better tracking and recording of these natural hazards in those three decades. Regarding the intensity of storms, which we already concluded has a clear link between minimum pressure and wind intensity, there seems to be a slight decrease in the maximum winds recorded and a very big increase in minimum pressure, which could again be attributed to better equipment being developed to record this type of parameter.

We set out to find the main causes of floods throughout the years in this natural disaster's dashboard. We can clearly identify severe rain (heavy rain, torrential rain, monsoonal rain) as the leading cause of floods around the world. These events could be linked with storms but unfortunately we don't have the information to conclude the cause of the extreme rainfall events, so there is no way we can confidently link the causes of floods with climate change. We can however affirm that monsoonal rains cause the longest floods, with an average twenty-one day duration.

In the dasboard that crossed the floods and hurricanes dataset we were aiming to find a correlation between these two occurrences, which would be corroborated by what we learned on the State of the Art. However, the results were mixed.

On the monthly distribution view we can very clearly see that "hurricane season" and "flood season" occur almost simultaneously, since they have very similar distributions. On the two other views, which correlate the number of records with along the years, we were not able to draw any strong conclusions. Both had small similar patterns but these were not prominent enough for us to declare any correlation along those parameters. This divergence of values could be caused due to the lack of values on the floods dataset compared to the hurricane's, which would justify why their monthly distribution is so close but yearly no patterns are noticeable.

Lastly, we were also unable to conclude anything about the correlation between the strength of storms' winds and floods' magnitude,

[4]In the last iteration of this report we presented a version of this visualization that was using the sum of maximum winds and the sum of minimum pressures, as axis. This was not our intention, we wanted the absolute values of each point and not a cumulative version of these. This change was extremely impactful to our interpretation of the data, since it provided two very different visualizations.

if anything, our visualization seems to show that these have opposite peaks. Since we already learned from the storms' dashboard that there are very few storms in January and February, we can confidently say that the high values for average wind strength, in these two months, are probably due to some strong outlier storms.

A subject of study was to check the influence of the temperature on the glacier area. Our goal was to check if higher moving average values by year have more powerful and visible impact on glacier melting. What was mentioned in the State of the Art (4) is that the glaciers are sensitive to the increase in global temperatures. However, with the presented data, we can't clearly confirm this thesis.

The first step was the confrontation of the Global Glaciers Recession dataset with the Countries temperature anomaly dataset. As it is shown in the 12 and 14 the northern hemisphere has higher temperature anomalies than the south one, but both are under influence of positive temperature anomaly. Most of the glaciers exist in the areas where yearly average is significant. That leads us to the question how exactly temperature changes affect the glaciers.

Through the investigation of the datasets, we have discovered that despite undeniable increase in global temperatures and global ice thickness average has increased. That measure is not clearly the opposite to the mentioned state of the art thesis (which could be i.e comparison of global ice caps tonnage, what was unattainable because of the lack of the data), however it shed some light upon the matter of global warming.

13 shows the average ice thickness by average anomaly temperature through years. The data is sundry and unbalanced, from countries with just one entry (i.e Mongolia) to the countries with rich history of observations (Norway, Switzerland). The results, likewise the data, vary a lot. In the USA, Canada or Antarctica we can observe rising tendency of the average ice thickness, with the opposite i.e. in Switzerland,Italy or Greenland. Nevertheless, global level of glacier thickness is rising.

We pondered in some possible explanations for this phenomena in the data, but it's important to emphasize that it's only a speculation. First of all, higher yearly temperature don't have to signify warmer winters. Overall average is positive, but it may be caused by higher summer temperature increase than winter temperature decrease.

Another potential reason may be the fact that national temperature data, may not efficiently reflect the real temperature changes near the glaciers. I.e. in China where numerous glaciers are located on the western part of the country in high, cold mountains [Krimmel 2002], while most of the population and industry sector units occupy flat, eastern China. There is a concern that average temperature would be disrupted by the observations from warmer, east side. If so, the average of national temperature anomaly may be higher from the real glaciers temperature.

There are a few explanation of the results, but they are not confirm (because of lack of data) and are not the main purpose of the project.

13 shows comparison of ice elevation and ice thickness. Higher elevation usually means less ice. Mountain areas are not able to carry much ice because lack of space. The biggest ice clusters are placed relatively close to the see level. What may be interesting is influence of sea currents on the ice caps, but we couldn't fulfill this study with the data we have.

7 References

References

AALST, M. 2006. The impacts of climate change on the risk of natural disasters. *Disasters 30* (04), 5–18.

CANIPE, C. Thirty years of atlantic hurricanes. https://www.axios.com/thirty-years-of-atlantic-hurricanes-1513305322-8b3c056a-ff13-49dc-a95d-961481119907.html. Online; accessed 23 May 2019.

DENNIS O. NELSON, 2019. Global Warming and Glaciers. http://www.waterencyclopedia.com/Ge-Hy/Global-Warming-and-Glaciers.html. Online; accessed 17 May 2019.

DOORNKAMP, J. 1998. Coastal flooding, global warming and environmental management. *Journal of Environmental Management 52*, 4, 327 – 333.

EARTH OBSERVATORY, N. World of change: Global temperatures. https://earthobservatory.nasa.gov/world-of-change/DecadalTemp. Online; accessed 17 May 2019.

FEMA. Data visualization: Historical flood risk and costs. https://www.fema.gov/data-visualization-floods-data-visualization. Online; accessed 23 May 2019.

FLANNIGAN, M. D., AMIRO, B. D., LOGAN, K. A., STOCKS, B. J., AND WOTTON, B. M. 2006. Forest fires and climate change in the 21st century. *Mitigation and Adaptation Strategies for Global Change 11*, 4 (Jul), 847–859.

GONÇALVES, D., LIMA, B., MOURA, J. M., AND FERREIRA, L. 2019. *'Portugal Without Fires', A Data Visualization System to Help Analyze Forest Fire Data in Portugal*. 01, 520–525.

INMETEO, MAREK MOJZÍK, M. P., 2019. Ventusky. https://www.ventusky.com/?p=-12.3;-28.8;3&l=feel. Online; accessed 23 May 2019.

KRIMMEL. 2002. Glaciers of the conterminous united states. *Science* (01).

LIPPONEN, A., 2019. Temperature anomalies 1880-2017 by country. https://twitter.com/anttilip/status/1033342041474969601, 08. Online; accessed 23 May 2019.

M. RUBY, C. O'NEILL, 2019. Glaciers melting Alaska-Washington. https://www.nytimes.com/interactive/2019/04/16/climate/glaciers-melting-alaska-washington.html. Online; accessed 17 May 2019.

MOUSAVI, M. E., IRISH, J. L., FREY, A. E., OLIVERA, F., AND EDGE, B. L. 2011. Global warming and hurricanes: the potential impact of hurricane intensification and sea level rise on coastal flooding. *Climatic Change 104*, 3 (Feb), 575–597.

NASA. Global climate change, vital signs of the planet. https://climate.nasa.gov/. Online; accessed 13 May 2019.

NASA/GISS, 2019. Global temperature. https://climate.nasa.gov/vital-signs/global-temperature/. Online; accessed 23 May 2019.

NOAA, 2019. Anomalies vs. temperature. https://www.ncdc.noaa.gov/monitoring-references/dyk/anomalies-vs-temperature. Online; accessed 23 May 2019.

ON CLIMATE CHANGE, I. P. 2007. *Fourth Assessment Report: Climate Change 2007: The AR4 Synthesis Report*. Geneva: IPCC.

PLUCK, K. Sea level rise since 1880. https://old.reddit.com/r/dataisbeautiful/comments/9lypeb/sea_level_rise_since_1880_oc/. Online; accessed 23 May 2019.

PRESS, C. U. 2009. Global climate change impacts in the united states.

RITCHIE, H., AND ROSER, M. 2019. Co2 and other greenhouse gas emissions. *Our World in Data*. https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions.

UNITED NATIONS, 2019. Only 11 years left to prevent irreversible damage from climate change, speakers warn during general assembly high-level meeting. https://www.un.org/press/en/2019/ga12131.doc.htm, March. Online; accessed 17 May 2019.

VELARDE, O., 2019. The best data viz and infographics on climate change facts. https://visme.co/blog/climate-change-facts/. Online; accessed 23 May 2019.

WALTERS, T., 2017. Wildfire exploratory analysis. https://www.kaggle.com/captcalculator/wildfire-exploratory-analysis, 09. Online; accessed 23 May 2019.

WESTERLING, A. L., HIDALGO, H. G., CAYAN, D. R., AND SWETNAM, T. W. 2006. Warming and earlier spring increase western u.s. forest wildfire activity. *Science 313*, 5789, 940–943.

YANG, P., SPENCER, T., STRIPLING, C., SHOEMATE, D., AND MODALA, N. R. 2017. Reinforcing wildfire predictive services with timely weather information. In *2017 25th International Conference on Geoinformatics*, 1–4.

A Cumulative CO_2 emissions, 2016

Entity:	**String**
Code:	String
Year:	Integer
Cumulative CO2 Emissions:	Decimal
Number of Rows:	62000
Source:	Cumulative CO_2 emissions, 2016 - Our World in data

B Temperature change

Field	Description
Area	String
Value	Decimal number
Year	Integer
Source	Temperature Change - available

C Glaciers

Glacier id:	**String**
Latitude:	Geographic
Longitude:	Geographic
Glacier name:	String
Thickness:	Meters
Elevation:	Meters
Size:	1300000
Source:	International Association of Cryospheric Sciences

D Global Mean Sea Level Trend

Year:	**Integer - Date**
Fraction of the Year:	Decimal
Smoothed GMSL:	Decimal - Millimeters
Size:	960
Source:	GMSL - NASA

E Global Active Archive of Large Flood Events

Country:	**String**
Detailed Location:	String
Start Date:	DD/MM/YYYY
Duration	Integer
Main Cause:	String
Magnitude:	Decimal
Size:	4500
Source:	Dartmouth Flood Observatory

F Hurricanes and Typhoons, 1851-2014

Latitude:	**Geographical**
Longitude	Geographical
Status:	String
Maximum Wind:	Integer - Knots
Minimum Pressure:	Integer - Millibars
Size:	75000
Source:	National Hurricane Center

G 1.88 Million Wildfires

Fire IDs:	**String**
Longitude	Geographical
Latitude:	Geographical
State:	String
County:	String
Discovery Date:	DD/MM/YYYY
Contain Date:	DD/MM/YYYY
Fire Year	Integer
Cause:	String
Fire Size:	Acres
Cause:	String
Size:	180000
Source:	USA Wildfires - Kaggle