Project Report Team 03
Integrated Map of Risk Information in Colombia

Johnathan Salamanca, Mario Cerón,
Carol Martinez, Javier Cocunubo, Jairo Niño, Alvaro Muñoz

December 10, 2019

1 Introduction

In 2012 the National System of Risk Management (NSRM) was created in Colombia. The system includes public, private, and community entities that will work closely with the government to coordinate the different risk management procedures. The NSRM is comprised of 6 instances:

- The National Risk Management Council (Consejo Nacional para la Gestión de Riesgo): coordinates the national system. At the head is the President and his govement.
- The National Risk of Disaster Management Unit UNGRD (Unidad Nacional para la Gestión del Riesgo de Desastres): it coordinates the nacional system and manage the risk management system.
- National Comittee for Risk Awareness (Comité Nacional para el Conocimiento del Riesgo): advises and plans the constant implementation process of risk awareness
- National Comittee for Risk Reduction (Comité Nacional para la Reduccin del Riesgo): it advises and plans the implementation of the process to reduce the risk of disasters.
- National Comittee for Risk Management (Comité Nacional para el Manejo de Desastres): it advises and plans the implementation of the process of disaster management
- City and Departmental Risk Management Council (Consejos departamentales distritales y municipales para la Gestión del Riesgo): they coordinate, advise, plan and control the processes of risk management in each territorial subdivision.

All six instances are responsible of preventing and managing possible disasters that occur in the country.

In April 2018, the National Planning Department (DNP) presented a report [1] that shows the national situation of the Risk Management in Colombia. The report presents a general overview of Disaster Risk in the world and the situation of Colombia in that matter.

Some of the information from that report is summarized as follows:

International Situation

- From 1980's the **disasters have triplicate worldwide**. 90% of disasters are hydrometeorological and generate 74% financial losses (e.g. Japan Tsunami, Katrina Hurricane, Japan Earthquake).

- The **number of deaths** due to disasters is **higher in developing countries** that in developed countries.

- Countries with high incomes are the ones that have more policy frameworks on risk management.

National Situation

- **88% of the disasters in Colombia are hydro-meteorological** (Inundaciones, movimientos de masa, flujo torrenciales, sequías e incendios, geológicos, otros).

- Infrastructure looses increase by Nina and Nino natural phenomena.

- Colombian **departments with less incomes** are the ones that have **more people affected** during the disasters.

Additionally, the report introduces the **Risk Management Index of Colombia** adjusted on the basis of capacities. The index measures **the risk of a territorial subdivision under hydrometeorological events** and the **capacity of that subdivision to manage the risk**. The index takes into account two indexes: the risk index and the capacity index. The risk component analyzes the thread, exposition, and vulnerability to a risk. Additionally, the capacity to manage the risk is analyzed based on the economic point of view, socio-economic, and risk management.

The index was created based on the following information:

- 15% of deaths are due to slow flooding (generated by constant and heavy rain that increases the rivers levels) and 85% of the homes affected during a disaster are due to this phenomena.

- Landslide: it causes 19% of death and 1% of affected homes.

- Torrential flow: it causes 66% of death and 14% of affected homes.

- 29% of the national territory has conditions of critical thread of hydrometereological phenomena.

- 13% of the population are socially vulnerable and are highly exposed to the most critical hydrometereological threads.

- Colombia territorial subdivisions have heterogeneous capacity of risk management.

Figure 1 describes the country situation on the basis of the 3 indexes: the capacity index (image on the left), the disaster risk index (image in the center), and the risk management index that combines both (image on the left).

Based on the previously mentioned facts **in this project** we present the development process of the tool called: **Integrated Map of Risk Information in Colombia** which we consider a tool that is a means to an end. It presents key information of the general situation of the country in terms of risk and can be used as a tool to guide future decisions in the different regions.

The following sections describe the steps follow to develop the tool.

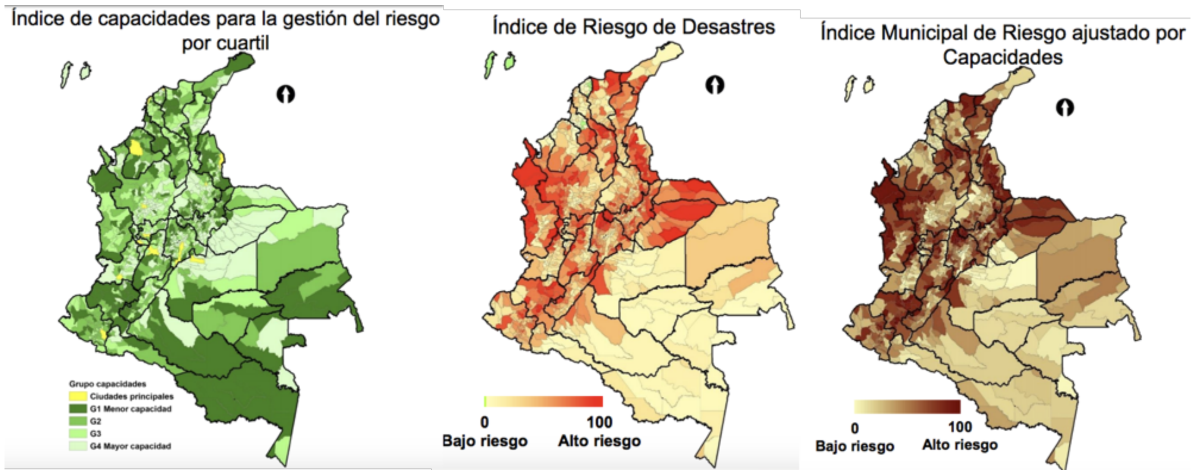


Figure 1: Risk Management Index of Colombia adjusted on the basis of capacities. The index which is illustrated on the right image combines the capacity of territorial subdivisions to manage the risk (image on the left), and their risk of a disaster (image in the center). Image taken from [1]

2 Problem Description

However, in this project, we challenged ourselves to present all the data together, in a way that any of us, as citizens, could make a decision, derive a conclusion or even form an opinion about what is happening in the region.

The Municipal Capacities-Adjusted Disaster Risk Index is an innovative indicator for policy-makers to make informed decisions about how to better preserve citizensẃell-being in the presence of real and potential threats. However, to be actionable, information needs not only to be available but efficiently delivered to communities, as a mean of protecting citizenś rights, foster economic growth, and make government officials accountable.

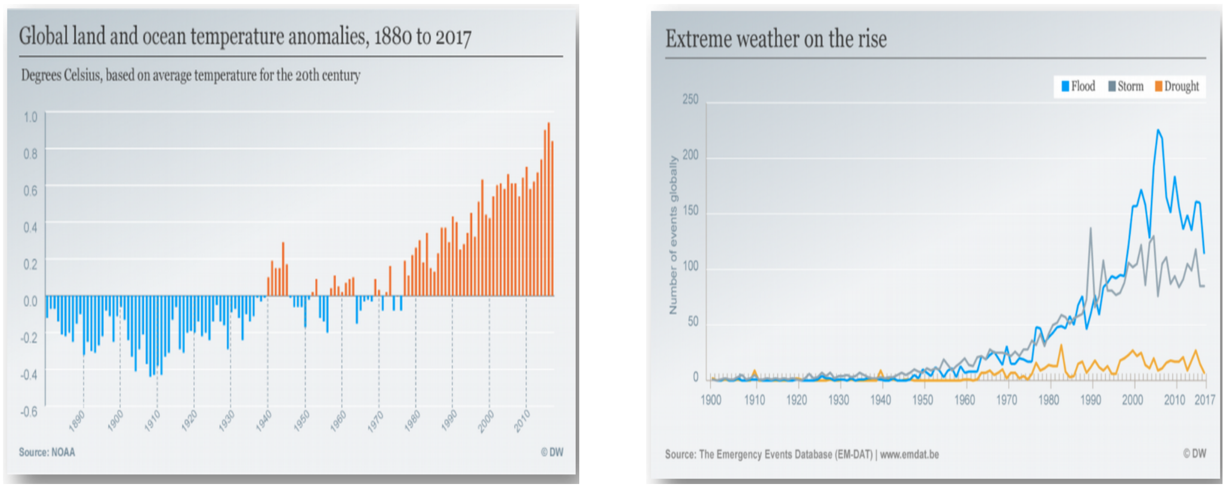
As mentioned previously, **Colombia** already **has** an index, **the Risk Management Index of Colombia adjusted on the basis of capacities**. However, official risk management **information lacks of a understandable delivery system**, that enables local communities **to improve their risk awareness** and disaster coping capabilities, in scenarios such as extreme temperatures and changing weather patterns.

For instance, in the official information, **it is not apparent how similar events have impacted communities with different risk and vulnerability profiles**, and there is no relevant information to assess the performance of risk management activities.

The decision to choose a topic in the field of **Risk and Vulnerability to natural disasters** was also based on two main reasons:

1. Besides any discrepancies people may have in terms of the cause of **climate change**, most people and the scientific community agree that it **is happening**. The world is experiencing changes in temperature patterns worldwide, and that has a direct correlation with the frequency of extreme hydrometeorological events. As it is shown in Figure 2, global temperatures are increasing, and extreme weather events have increased in the last 60 years. In Colombia, the hydrometereological events are responsible of 88% of the disasters according to [1]

2. The exposition to the effects of such hydrometereological events is not equally distributed. **Risk from a natural disaster** is somewhat **random** but **vulnerability is anthropogenic**.



(a) Global temperature anomalies

(b) Number of extreme weather events per year

Figure 2: Frequency of extreme hydro-meteorologica events.

In other words, the cost of two events similar in magnitude, in two different places, can vary from some physical damage, to the loss of vital infrastructure and even human lives.

3 Proposal

With the proposed **Dashboard Web Application**, the team expects **to enable any person** to find, read and **understand the basic risk profile of their region of interest** and also understand about climate associated risks. All this based on open data philosophy.

We set out to do 3 things:

- Find historic evidence of those claims.
- Create an index to capture and monitor that imbalance in the allocation of coping capabilities and resources.
- Offer actionable information to both the general public and administration officials to improve their understanding of future impact of climatic events.

Open data is not just about putting information in the cloud and making it available for people to download it. It is **a philosophy and practice that guarantees availability for all** in order to use data to foster economic growth, protect citizen's and business' rights, and delimit administration's accountability.

Put in another way: it is not open data if there is a knowledge barrier that keep it out of most people's reach.

3.1 Possible use

The developed tool could be of value for the people in general, outside our regular interests, activities and industries. Imagine you are a person casting a vote:

- The worst equipped your region is to deal with emergencies, the more important is for you to know which candidate is prioritizing those capabilities.

- If you are an elected official, getting a clear picture of the risk profile of your district at a glance is a core asset when allocating resources, and designing preemptive plans. For instance, knowing that now precipitations level is close to the historical maximum can trigger an early alarm.

4 Datasets sourced

The main dataset used in the project is from the Colombia Risk of Disaster Management Unit (Unidad de Gestión de Riesgos y Desastres) UNGRD [2]. The dataset contains information about the risk management associated with natural phenomena, socio-natural, technologic, and human-based non-intentional incidents reported in Colombia in the last 10 years (38626 records). Some of the fields found in the dataset are: Date, Department, Municipality, Event Name, Code, Dead, Wounded, Disappeared, Affected People, Affected Families, Affected Houses, among others.

The team will also use a dataset from the National Administrative Department for Statistics DANE. It is a time series between 1985 to 2020 and contains demographic information, per department code [3] .

Both datasets contain "DIVIPOLA" codes, which is the codification of the Politica-Administrative Division of Colombia (Codification of the departments,). Figure 3 describes the meaning of the code. The first two numbers correspond to the department, followed by the Municipality Code and the Populated Center [4].

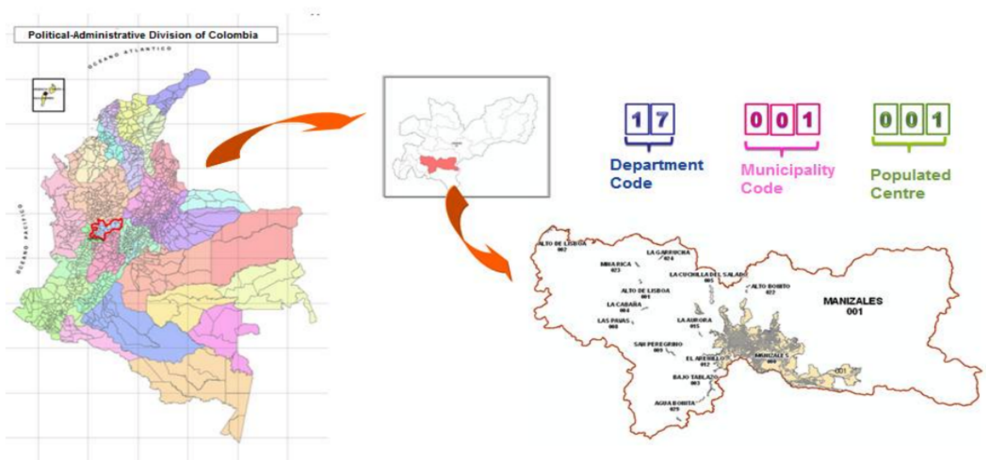


Figure 3: Explanation of "DIVIPOLA" code. The codes provide information of the Politica-Administrative Division of Colombia. Image taken from [4]

5 Application Overview

5.1 Users

Any person, Government officials, or someone in the private sector who wants to read and understand the basic risk profile of their region and make decisions with that information.

5.2 Architecture

Figure 4 shows the architecture of the proposed solution including the elements of the application at component level and its connections at high level (see deployment diagram). Additionally, it shows the application elements used for the Front and Back End. The figure also shows the names of the technologies used hosted on AWS cloud, i.e.: (Python, dash and libraries).

The following is the list AWS components used in the project:

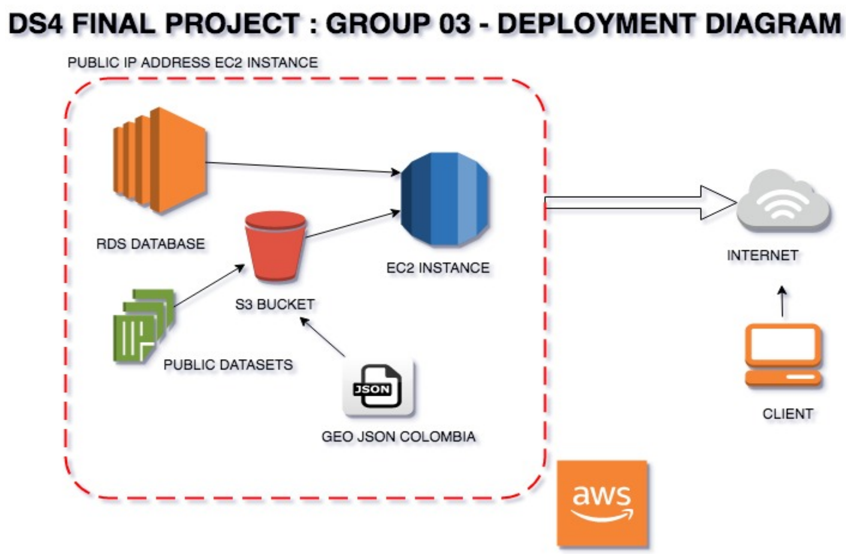
1. The machine who host the solution (Elastic Compute Cloud - EC2).
2. The Database (Relational Database Service -RDS).
3. The storage for the datasets and GeoJson files for Colombia on the service (Simple Storage Service - S3) to save these files.
4. The Security group for these services talks with each other and have access from the internet as well.
5. The remote DNS (Domain Name Service) to have a friendly URL for the application on the Apache Web server.
6. A remote code repository (hosted by github). It is used for hosting the source code and documentation.

5.3 Front End Design

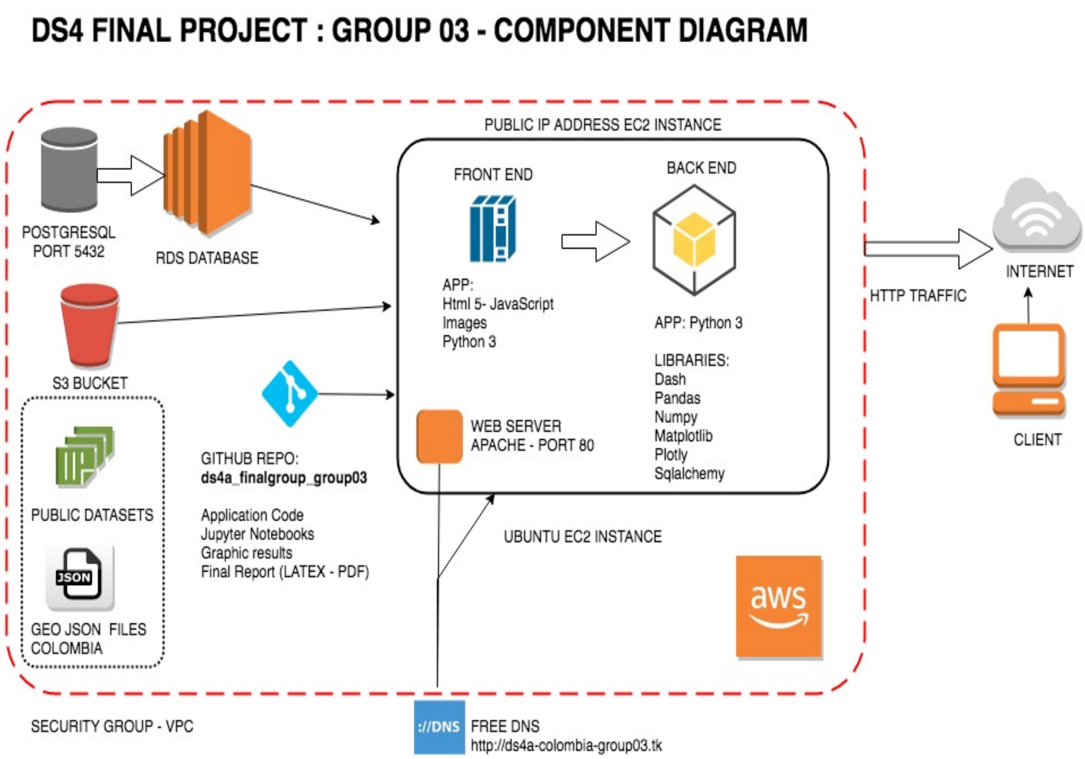
The following is the link of the web-page created for the project.
http://ds4a-colombia-group03.tk/

In general terms, the proposed information system will provide a dynamic map of Colombia with:

1. Impact variables such as Deaths, People and Houses affected.
2. A map (Main View) that will show the index adjusted by capacity (Índice Ajustado por Capacidades). By default, the map starts with the country view with political division (departamento).
3. A timeline slider will facilitate the visualization of the index using time intervals.
4. Considering the established association between extreme temperatures and the frequency of hydro-meteorological events, a projected extreme-temperature indicator for the 100 most vulnerable municipalities with 3 data points: Indicator value at Time 0 (1998), Time 1 (2018) and Time 3 (projected 2040). The indicator corresponds to the extreme temperature projection made by Climate Impact Lab for the number of days a year that register temperatures above 32 degrees Celsius.
5. The option of showing political divisions views (vista de departamento). The user can select the small political subdivision (municipio) and the graphics on the right side of the dashboard screen are updated.



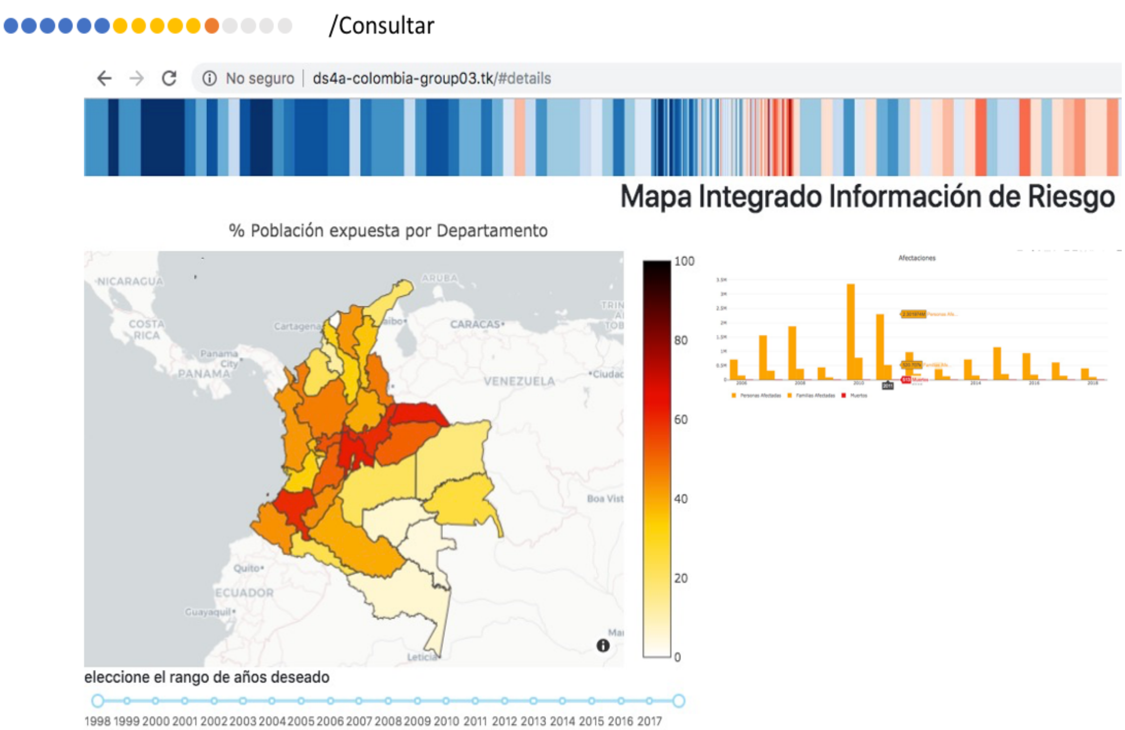
(a) Deployment diagram



(b) Component diagram

Figure 4: Project architecture.

(a) Presentation page



(b) Main page

Figure 5: Front End: Integrated Map of Risk Information in Colombia. The main page (Figure 5(a)) will show the map of Colombia with the Risk Index, and will offer interactive options to get detail information of selected Municipalities, risk index and a projected temperature indicator.

6 Data Wrangling and Data Cleaning

The main dataset used in the project contains information about Disasters in Colombia from 1998 to 2017. We found 8 files to integrate from the following links:

- Datos Abiertos Colombia:
  https://www.datos.gov.co/
- Unidad Nacional para la Gestin del Riesgo de Desastres:
  http://portal.gestiondelriesgo.gov.co/http://portal.gestiondelriesgo.gov.co/.

The files are Microsoft Excel Files and contain the same information in the first 40 columns. However, the column names are not equal.

All the information has been analyzed in notebooks, the notebooks can be found in the following links:

- Disasters dataset integration: link
- Disaster exploratory analysis: link.

For Data Wrangling and cleaning the following steps were conducted:

- Load the 8 Excel Files as Pandas Data Frames. df_desastres_1998_2012, df_desastres_2013, df_desastres_2014, df_desastres_2015, df_desastres_2016, df_desastres_2017, df_desastres_2018
- Parsing the date columns as pandas.datetime.
  - Two corrections were conducted with some missing values:
    31/11/2018 to 30/11/2018, 27/11/201 to 27/11/2018
- Columns names standarization:
  - Change column names in df_desastres_2018.
  - Normalize type of disaster: we found the same kind of disaster written in different ways
- Concatenate all dataSets.
- Transform Data type of the following columns to numerical values:
  - muertos (deaths)
  - personas (Population)
  - heridos (Injured)
  - familias (Families)
  - viviendasdestru (Affected houses)
  - ptepeat
  - hectareas (hectares)
  - otros (others)

7 Technical Information

This section will provide a general idea of the different systems used to create the Web Application. The following items summarize important technical information of the project.

- What technologies did you use to set up the front-end? Python, Pandas, Dash libraries, GeoJSON files (Colombian map), html,css
- What types of visualizations do you display and why? A General Colombian Map with information associated to the risks, number of death, number of people and houses affected. A Detailed map but with information that correspond to the political division and subdivision.
- How does the Dashboard pass and receive information for use from the AWS-hosted components? views and database queries were created. This to create pandas DataFrames that are used to generate the information presented in the front end.
- What computation tool did you use? an EC2 Linux Ubuntu instance
- Code / program design paradigms used: Standard Python Dash application with callbacks with CSS, also to embed the page on HTML page with an Iframe.

7.1 AWS-hosted database

Figure 6 shows the entity relationship model of the project. The information contained in the model is:

1. Disasters : information of disasters of Colombia from 1998 to 2017, with the following fields: date, Colombian political division, number of disasters, death, injuries, missing people, families, houses, public and education services infrastructures and some economical information. (table disasters)
2. Events: name and category of the event (table eventos)
3. Political division of Colombia (divipola): ID of division, subdivision and name of both (divipola table)
4. Population estimates: relates population by period, age groups, political division, id, and gender
5. Weather estimates: weather information from 1990 to 2018 (minimal temperatures, maximal temperatures, precipitations) (tables: historico_cond_metereologicas, load_mintemp, load_maxtemp, load_precipitaciones)
6. Load tables: temporal tables for loading disasters, political division, population (tables: load_disasters, load_divipola, load_populations)
7. Views: wv_disasters (summary table for disasters).

The SQL script for the creation of the database on AWS can be download from this link:
https://marioceron-case-51.s3.amazonaws.com/final_project/Script_Desastres_DB.sql

Full backup Database:
https://marioceron-case-51.s3.amazonaws.com/final_project/Script_Desastres_DB_full.sql

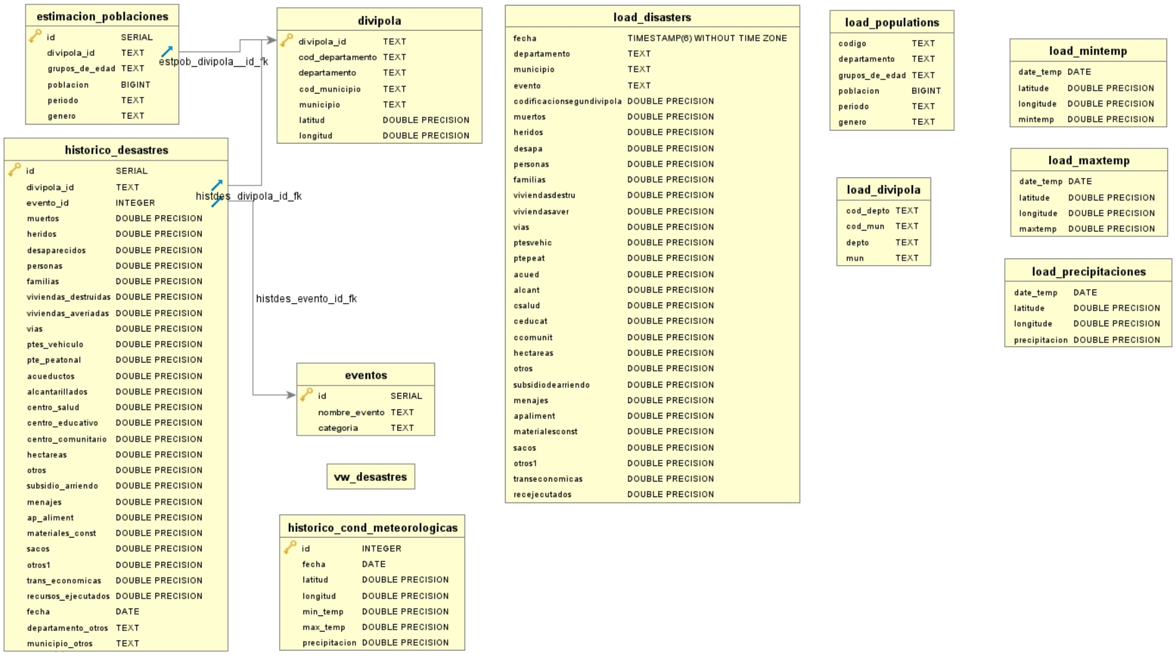


Figure 6: Entity relationship model.

Figure 7 shows the database uploaded to AWS, and Figure 8 shows the database loaded to the AWS hosted database. The link between the front-end and the AWS-hosted databased was also stablished (see Figure 9).

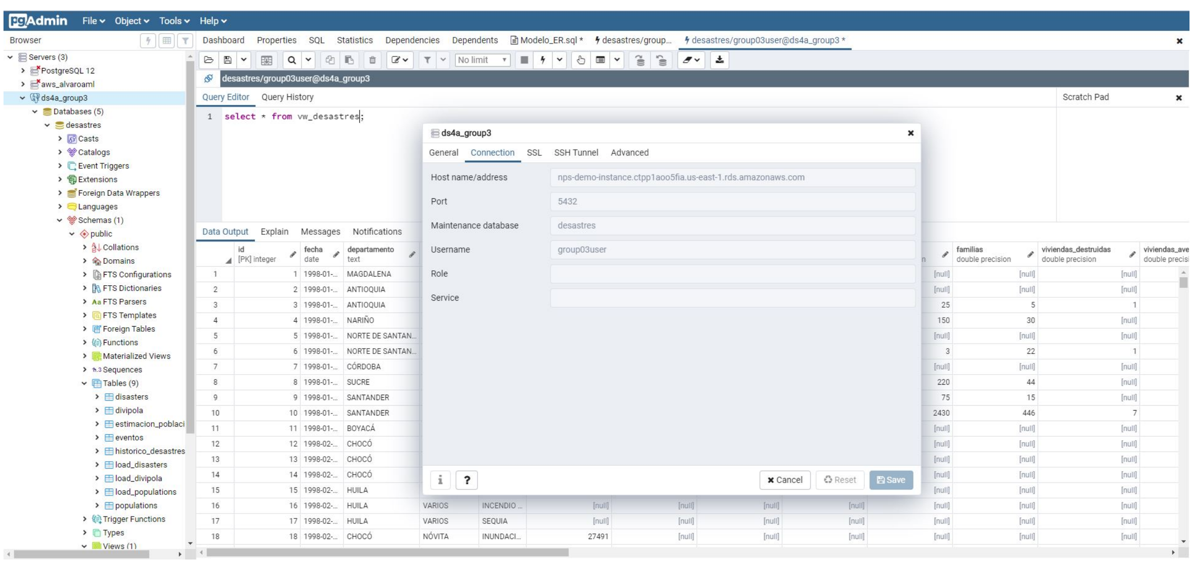


Figure 7: AWS connection.

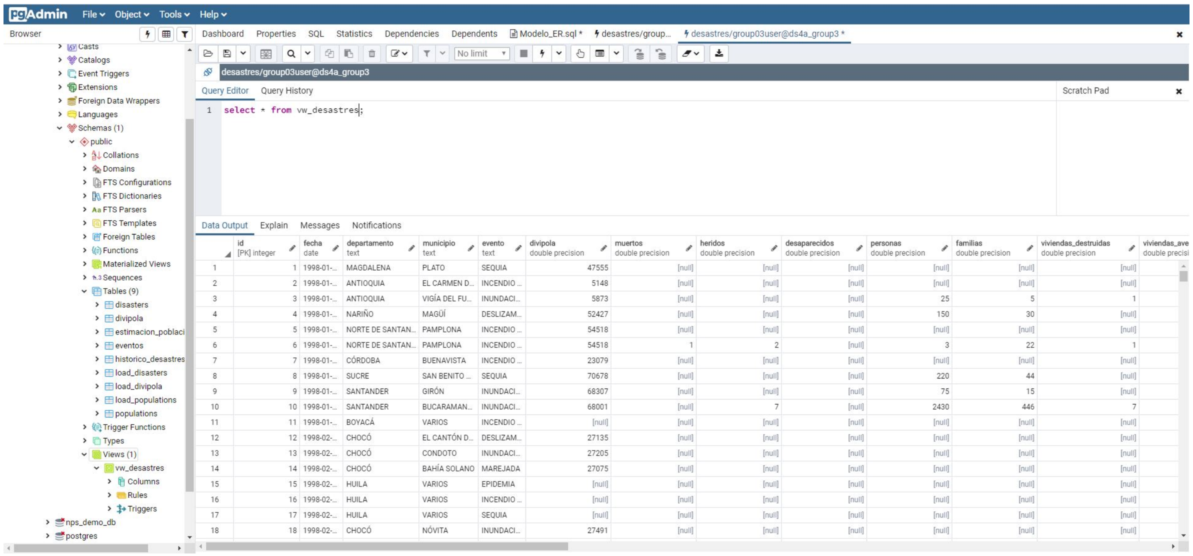


Figure 8: Database loaded to AWS.

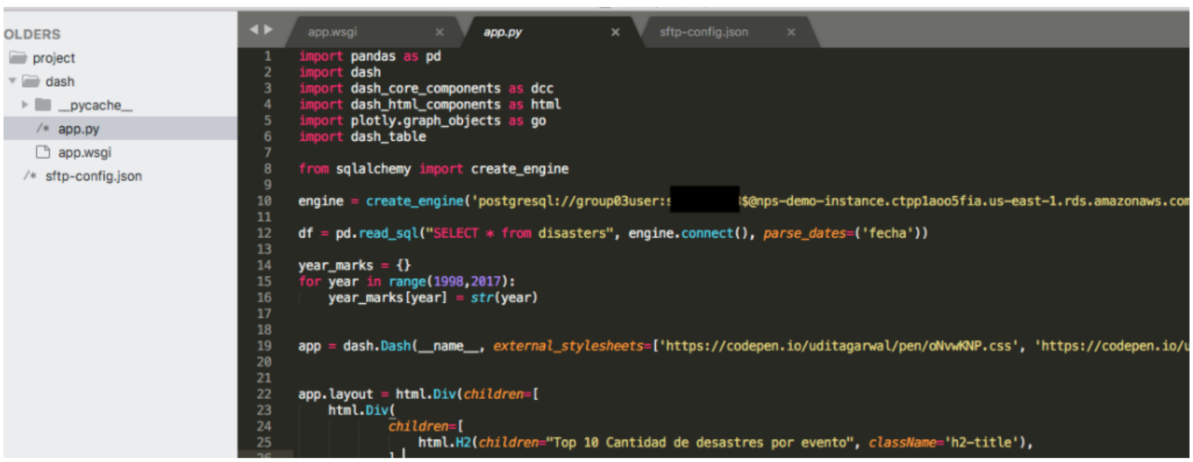


Figure 9: Link between DASH and AWS stablished

7.2 Interaction with the Dashboard

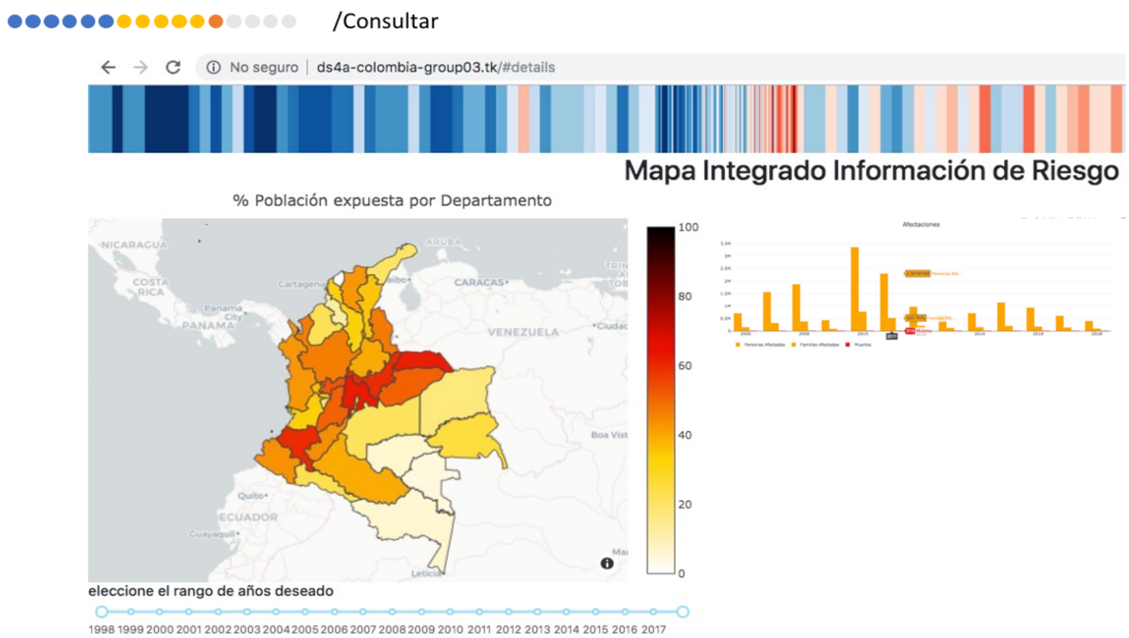
Figure 10 shows what it looks like when you go to the website and click explore. Figure 10(a) corresponds to the main view. On the left-hand side is dominated by a map where you see the municipalities with their Adjusted Risk Index, reflecting each city's built-up capabilities to cope with risk. That's by default but you can pick different variables for the heatmap.

In the upper right, the main view shows the possibility to drill down into impact variables: Affected people and families, affected and destroyed homes, and off course the deceased. Which accounts for the current vulnerability of the region of interest at any point in the last 30 years (which can be modified with the slider below the map).

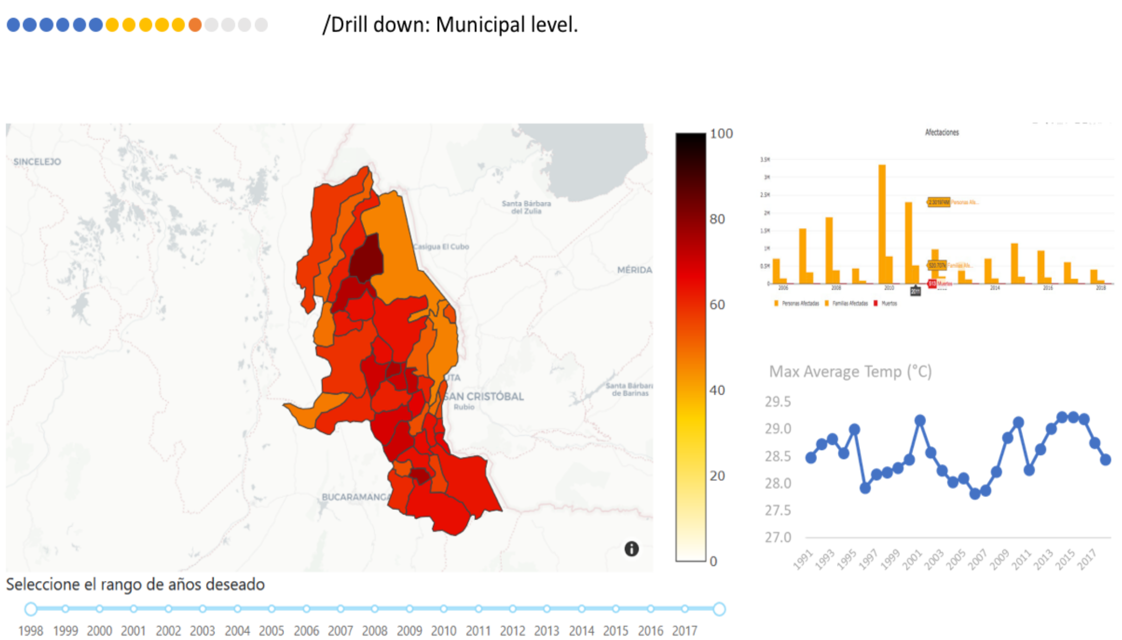
Once a person has clicked on the region of interest, the Map changes and shows the selected division and the different subdivisions, as shown in Figure 10(b). The figure shows El Tarra Norte de Santander, a municipality with a population of around 11K located at Colombia's north east. Heavily impacted by and - ill prepared to deal with- hydro-meteorological events, which you can see from their vulnerability index and the trend in affected people and homes.

If you check the evolution of average max temperatures you can see the size of the problem: Between 1990 and 2000 this indicator reached 29 degrees only once, in 1995. Then in 2001 it reached 29.17 and then again it stayed below 29 until 2010. Roughly a 10 year's cycle. However, it reached 29 again only 3 years later and hit a historical maximum at 29.23 in 2014. Stayed there for the next two years until 2016. You can also see how the minimum value for the last 10 years cycle is considerably higher than in the past.

Tipping points are crucial in environmental data. If earth's temperature is to rise 1.5 degrees, global population will be highly expose to severe heat waves, to Water Scarcity, and to flooding from sea level rise.



(a) Main page



(b) Information of political division and subdivision

Figure 10: Interacting with the Integrated Map of Risk Information in Colombia. Figure 10(a) shows the main page with the map of Colombia with the Risk Index. It provides interactive options to get detail information of selected Municipalities, risk index and a projected temperature indicator. Figure 10(b) shows the municipality of El Tarra Norte de Santander, heavily impacted by hydrometeorological events with a hight vulnerability index and high number of people and home affected by those events.

8 Final Thoughts

Something we learned when building data visualizations for climate data is that you are always trying to balance simplicity of the approach with closeness to the phenomena. Ed Hawkins, climate scientist from Reading University created the warming stripes visualization which you can see Colombia's version atop of the web application.

They show a clear picture of global warming since the 1900 in such an elegant way that they went viral and did a great service in terms of raising awareness. We took inspiration from that but decided to go into more detail in order to show how global warming affects you directly. The town you were born, the city you live.

Hopefully we accomplished some of that: Surfacing information from the past, removing friction to access current data and giving some clues about the future. This was the **challenge accepted** by 6 college educated professionals, experienced in the fields of data analytics, information management and research to put all the data together, in a way that any of us, as citizens, could make a decision, derive a conclusion or even form an opinion upon it.

Finally, to accomplish what was shown in this document, the work was divided as follows:

- Database AWS Computation / SQL queries: Alvaro Muñoz and Mario Cerón
- Front End Dash App: Javier Coconubo, Jhonatan Salamanca and Jairo Niño
- Data Wrangling: Javier Coconubo, Carol Martinez, Jhonatan Salamanca and Alvaro Muñoz.
- Dataset Research: Jairo Niño and Carol Martinez
- Project Management: Mario Cerón and Javier Coconubo
- Documentation and reporting: Carol Martinez and Mario Cerón

All of the 6 members of the group collaborate in equal efforts as well. $(16,6\%)$

References

[1] "Disaster risk index of colombia." https://colaboracion.dnp.gov.co/CDT/Prensa/Presentaci%C3%B3n%20%C3%8D%C3%8Dndice%20Municipal%20de%20Riesgo%20de%20Desastres.pdf. Date retrieved 16-10-2019.

[2] "Colombia risk management unit." http://portal.gestiondelriesgo.gov.co/Paginas/Consolidado-Atencion-de-Emergencias.aspx. Date retrieved 16-10-2019.

[3] "National administrative department for statistics." https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/series-de-poblacion. Date retrieved 10-2019.

[4] "Methodology for the codification of the politicaladministrative division of colombia -divipola-." https://www.dane.gov.co/files/nomenclaturas/ingles/Methodology%20for%20the%20Codification%20of%20the%20Political-Administrative%20Division%20of%20Colombia.pdf. Date retrieved 16-10-2019.