David Cabornero

Carlo Magnani

Fernando Madrazo

A visualization tool for analyzing bicycles accidents in the “Comunidad de Madrid”

1. Introduction

Data visualization is an extremely powerful tool for knowledge discovery among other things. Over the years more resources have been tasked for the development of visualization best practices and standard. This type of methods looks very well suited for tackling problems policy makers face every day in order to be able to structure more effective policies and optimal decisions for the overall wellbeing.

For this reason, we decided to make use of such techniques to share new knowledge around some specific problems afflicting the *Comunidad de Madrid* which we believe are becoming more crucial nowadays; we decided to investigate patterns related to accidents in the city of Madrid to help distil new information by the use of some visualization tools to uncover patterns, trends, correlations and outliers among other things.

This paper will be structured as follows: we will first present the problem at hand and what we’re trying to achieve, we will then discuss the type of task at hand and questions we’re trying to answer, along with the data used for the task; as a third part we will discuss the visual encoding we found most useful to answer the question we posed and finally we will briefly present the application we developed for the problem.

2. Problem Characterization

A big issue in metropolis such as the city of Madrid has always been viability and safety on roads; accidents play a major part in public perception of a municipal administration’s effectiveness. As we all know, global warming and worldwide problems such as COVID-19 accelerated the transition from classic means of transportation to more environmentally friendly/personal ways of commuting. Bicycles therefore are becoming an ever-increasing way for citizens to move around, although poses a problem for administrations in terms of the effective tackling of eventual disruption and for accidents management. By studying data related to bicycle accidents through time, we believe we can help focus problem solving initiative on more focused topics therefore making them more effective. As we don’t know what these topics are beforehand, we need to leverage interactivity to investigate the dataset to get the answers we need; to do that we developed an interactive visualization tool leveraging R and the Shiny package.

General questions we’re trying to answer are on the likes of: Is there any seasonal pattern in terms of number of accidents for some specific district of Madrid? How is the density of accidents distributed through the districts of the city? Who are the common victims of accidents in each district? How has the pandemic affected to the number of accidents in each district these previous years? Is there a mayor concentration of accidents in any sector of the city? Is weather a conditioning factor for accidents through time?

3. Data and Task abstraction

The dataset we chose relates to the “Open Data Project” from the *Ayuntamiento de Madrid*; purpose of this initiative is to allow citizens to be an active part of the of the decision-making process by exploring, studying or developing applications in order to help municipalities increase welfare for their citizens. We felt this dataset was the ideal candidate for our effort as it is quite detailed and well structured.

The dataset (a Table) comprises of items and attributes: 2291 items and 16 attributes of which several are keys. The dataset structure is highlighted in table 1.

Attribute Name	Attribute Type	eaxmple
num_expediente	double (idenfitier)	2019S002373
fecha	date	08/01/2019
hora	time	20:55:00
localizacion	categorical	CALL. PUERTO DE LA CRUZ VERDE, 49
numero	ordinal	49
distrito	categorical	ARGANZUELA
tipo_accidente	categorical	Atropello a persona
estado_meteorolÃ³gico	categorical	Despejado
tipo_vehiculo	categorical	Bicicleta
tipo_persona	categorical	PeatÃ³n
rango_edad	categorical	De 35 a 39 aÃ±os
sexo	boolean	Hombre
lesividad	ordinal	7.0
coordenada_x_utm	quantitative (spatial position)	441775,74
coordenada_y_utm	quantitative (spatial position)	4470816,41
positiva_alcohol	boolean	N
positiva_droga	boolean	N

table 1

We will present (highlight) information already known by the general public and administration officials (i.e. the number of accidents) while trying to discover new as well by looking for specific patterns or unusual things in the data (this is why Visualization will be used).

As discussed already we will be searching for elements within the visualization framework as we

And finally, we will be looking for trends, outliers and specific features in our data while also searching for correlation or dependencies.

4. Visual Encoding and Interaction

- Choropleth

Include 2 different sets of options to filter the map shown:

  - Type of victims: all, drivers and pedestrians.
  - Number of accidents: total amount and per capita.

Idiom	Choropleth Map
What: Data	Geographic geometry data. Table with one quantitative attribute for the number of accidents per district.
How: Encode	Space: use given geometry for area mark boundaries of the districts. Color: sequential segmented color map representing the density of accidents in that region.

Figure 1

- Cluster map

Include 2 buttons to zoom closer or farther with a "+" symbol and a "-" symbol respectively.

Idiom	Cluster map
What: Data	Geographic geometry data.
What: Derived	Clusters with one quantitative attribute: the number of accidents.
How: Encode	Distance: clusters are merged into one single cluster the farther you zoom (or splits again if you zoom in). Color: group of clusters color representing the number of accidents that happened in that cluster. Position: each accident is allocated in base of the coordinates assigned to it.

Figure 2

- Streamgraph

Include 2 different sets of options to filter the graph shown:

  - Current year: 2019, 2020 and 2021.
  - Weather selected: clear, hailing, light rain, heavy rain, cloudy and unknown.

Idiom	Streamgraph
What: Data	Multidimensional table: one quantitative value attribute (accidents), one ordered key (months), one categorical key attribute (weather) .
What: Derived	One quantitative attribute (for layer ordering).
How: Encode	Use derived geometry showing weather layers across the months, layer height encode accidents within the district.

Figure 3

- Radial chart

Include a set of options to filter the chart shown:

  - Select a district: include a list with all the different districts in the city.

Idiom	Radial Bar Charts
What: Data	Table: one quantitative attribute refering to the number of accidents that month, one categorical attribute refering to the month the accidents occured.
How: Encode	Length coding the line marks: radial layout. Color: sequencial points with unique color representing the year in which the accidents happened in that region.

Figure 4

5. Implementation

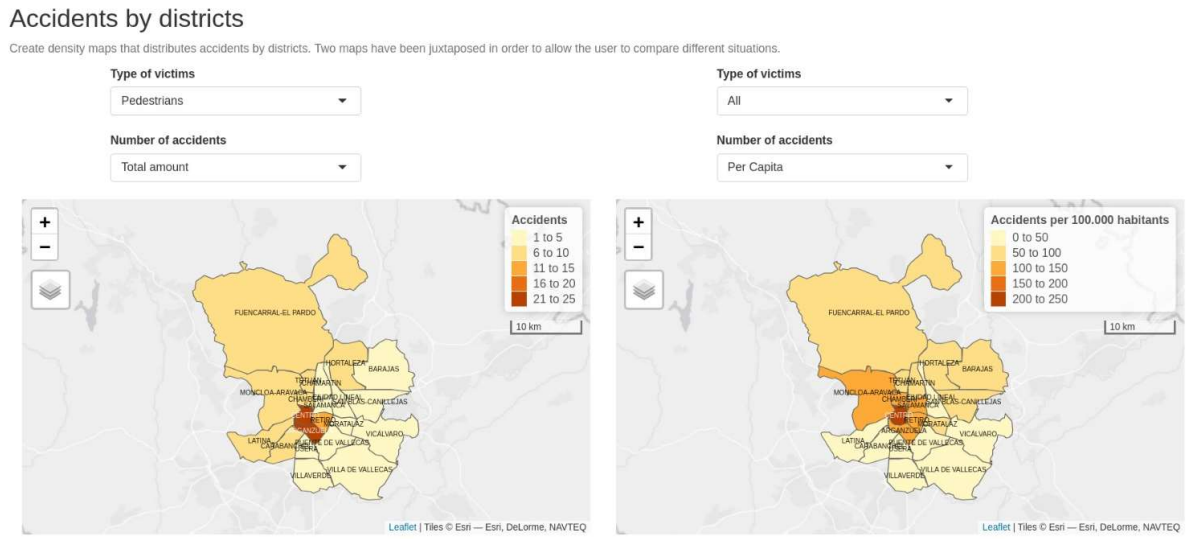
For this project we used the following list of libraries:

- *shiny*
- *tmap*
- *dplyr*
- *tidyverse*
- *rgdal*
- *ggstream*
- *leaflet*
- *fmsb*
- *lubridate*

To start the application it is only necessary to open the main.R with Rstudio and execute it.

Here is a user guide for the application:

- Choropleth map

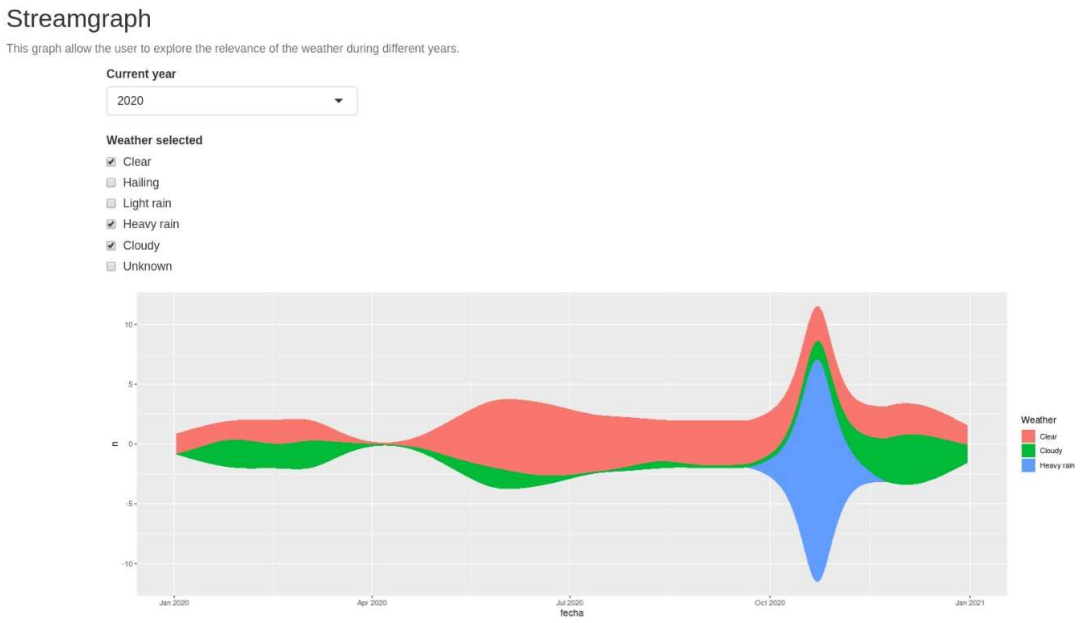


We juxtaposed 2 choropleth maps in order to compare different situations. In this case we can change the type of victim who suffered the accident shown for each of the choropleth map by clicking into the drowdown window assigned to the “Type of victims” and also we can change the number of accidents by clicking into the “Number of accidents” and choosing to see the total number of accidents or the ones occured per capita.

- Cluster map

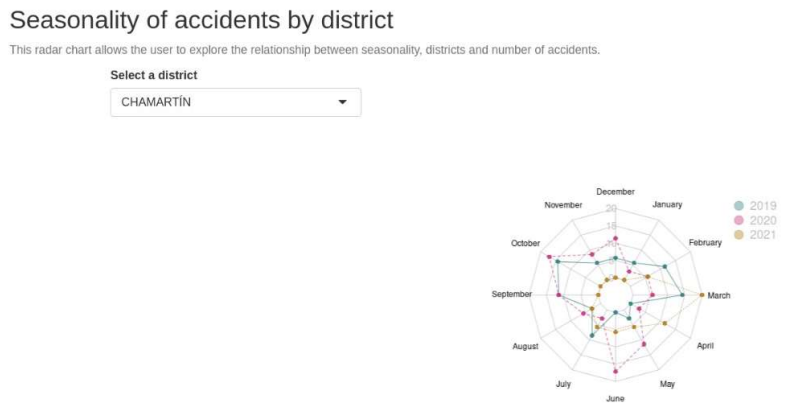
Using this map we can allocate each of the accidents geographically, but in case we are getting far with the zoom by clicking “-”, the accidents will merge into clusters composed with a number indentifying the number of accidents included into that cluster and also a diverging colormap between green and red that the higher the number of accidents included into a cluster the more close to red would be. The same way if we get closer by clicking “+”. A small animation is carried out when the user zooms in/out, preventing that they lose track of the clusters.

- Streamgraph



With this graph we can change the year we want to check by using the current year dropdown window, and by clicking into de different options of the weather selected we can choose which weathers we are interested in to compare how they affected each month into the number of accidents.

- Radial chart



In this graph we can check the number of accidents occured in a district for each of the months. In this radial chart we can observe all the accidents in that district also based on the year, and since all of them are shown at the same time we can compare them to see if some year had more or less incidents. To change the district we are checking, we can change it by clicking into the drowdown window fo select a district. This plot is specially good to find periodic patterns in our data.