

Incidenti a Trento: analisi e statistiche dal 2003 ad oggi

Schematic Explanation

Course: Data Journalism
prof. Ivano Bison and prof. Maurizio Napolitano

Alessandra Corso
Department of Mathematics
University of Trento
Student number 205149
alessandra.corso@studenti.unitn.it

1. Asking a Question

The subject of the article [1] is the accident. The project focuses on road accidents in Trento. The first approach to the topic was in terms of three questions: how many clashes? Where do they happen? How do they spread over different areas? The aim was to study the development of accidents over time, explore the history and the latest news about the clashes in Trento. The city residents should be informed about this issue. It is important to raise the awareness about current events and, specifically, road conditions. Data scientists analyze big data from a statistical and a mathematical point of view with the purpose of communicating the results to people and increasing their knowledge on these current problems. The findings of this research should contribute to a deeper understanding of accidents in Trento and it could be published in a local online newspaper to raise residents awareness around the current issue.

2. Finding Data

The municipality of Trento makes data available at its specific web site [2]. These open data contain lots of information about different topics regarding the city. By looking at *Aree tematiche* section, each person can easily find data about environment, economy, education, sport and others. In this case,

specific data about accidents are in the *Cartografia* section. Another part of data was useful for the final paragraph of the article. These data can be found at the Trentino website [3] in the section concerning sport advice. One specific point gives detailed information about Monte Bondone and it is used for analyses [4].

3. Getting Data

From the *Cartografia* section, in the *Download* part, there is the connection to the chapter of *Incidenti (Open Data)* [5]. The dataset used for all the analyses is in the part of *Risorsa 2* and, from the related link, it is possible to download the *incidenti.zip* folder. At this point, the dataset, called *incidenti.dbf*, can be opened. It has more than 16000 rows, each referring to a single accident in Trento. 8 different variables in the columns describe the time and the place of each event.

4. Cleaning Data

R code is available in an html form where outputs are showed [6] or in the folder of github [7]. First cleaning operations were performed to delete NA values from the dataset. Specifically, more than 30 rows contained null values in the *numero* and in the *anno* columns (R code: rows 47 – 54). Some mistakes were made in the *numero* column. The

variable describes the numeric sequence of the occurred accident and each dataset row refers to an accident case. The major value of the numeric sequence would be equal to the length of the rows, but in some cases, two numbers were different. This inequality was corrected considering the length of dataset rows (R code: 56 – 70). Moreover, the downloaded dataset contains accident information from 2002 and it is updated until 2020. Since the starting month of 2002 was not clear, dataset rows referring to accidents of that year were deleted (R code: 73 – 91). The analysis was performed on data containing information from January 2003 to August 2020.

5. Analysing Data

All analyses of this project were made through the programming language R (html file [6] and github source [7]). The first important operation was dividing the big dataset in different smaller parts, each containing information about one single year (R code: 135 – 145). Another next important step was running *revgeo()* function for all datasets. In this way, by starting from latitude and longitude variable, it is possible to add, for each row, information related to street, house number and zip code of the place where each accident happened (R code: 149 – 167). Then, the writing of a little algorithm was useful to know the exact number of accidents in a specific geographic area, considering latitude and longitude information (R code: 356 – 379). Other functions, calculations and statistics were applied to make all analyses. In a Google Drive folder [8] some files were saved, in this way, R code can be run to see analyses and all results.

6. Presenting Data

For this article, some plots were built, through the software R, in order to summarize results for an easier and a more immediate understanding. Histograms were chosen to show a unique numeric variable in the y axis in relation with a temporal variable in the x axis. For the first histogram (R

code: 104 – 128), colors help to perceive the difference of bar heights. For the other two (R code: 389 – 534), colors remind the plot subject and help the viewers to make attention to one specific bar with a different characteristic. Connected scatter (R code: 298 – 352) was chosen to show more than one trends. Colors help the viewers to easily distinguish different cases. Lollipop plots (R code: 617 – 756) were built horizontally because the categorical variable is described with a text and, in this way, viewer can easily read it. At the end of the article, a customized plot (R code: 773 – 832) relates a first numeric variable with a categorical one. Different intensities of color add one other information: the brighter the color, the greater is the value of a second numeric variable. Finally, maps (R code: 205 – 225) help the viewers to have a real feedback. In this way, it is possible to see city streets and locate accidents in an exact geographic point.

Sources and references

- [1] Article *Incidenti a Trento: analisi e statistiche dal 2003 ad oggi*, available at <https://medium.com/@alessandra.corso/incidenti-a-trento-73abb90d23ac>
- [2] Comune di Trento website <https://www.comune.trento.it/>
- [3] Trentino website <https://www.visittrentino.info/it>
- [4] Monte Bondone section of Trentino website https://www.visittrentino.info/it/guida/sport-estate/salite-da-mito/monte-bondone_tour_18085714
- [5] Incidenti (Open Data) at Comune di Trento website <https://www.comune.trento.it/Aree-tematiche/Cartografia/Download/Incidenti-Open-Data2>
- [6] R code of the project in an html form, available at https://alessandracorso.github.io/project_datajournalism/R_code_project_data_journalism.html
- [7] R code of the project from github, available at https://github.com/alessandracorso/project_datajournalism/blob/master/R_code_project_Incidenti_Trento.Rmd
- [8] Folder in Google Drive available at <https://drive.google.com/drive/folders/1rurddYtPGcTJi3Jf2FJ-E2jUSCslQpjj?usp=sharing>