A Visual tool for Data Quality Checking of Geomagnetic Data

Juan Carlos Méndez
jc.mendez (at) uniandes.edu.co

https://beta.observablehq.com/@dersteppenwolf/a-visual-tool-for-data-quality-checking-of-geomagnetic-data

Description

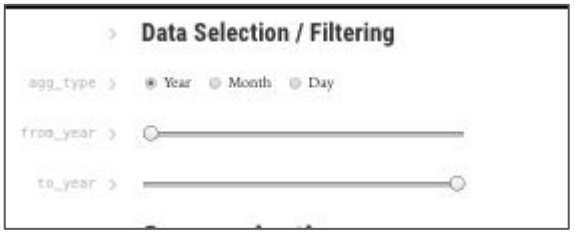
- **Dataset Title:** Geomagnetic Components of the Fúquene Observatory between 1955 and 2014
- **Number of Rows:** 525.960
- **Data Types :** Quantitative, temporal
- **The Problem:** The dataset has a huge number of rows (525.960). The owner of data doesn't have a visual tool that let him easily check data quality
- **The Task:** Allow the user to explore the data set in a visual and interactive way in order to detect possible quality problems in the data.

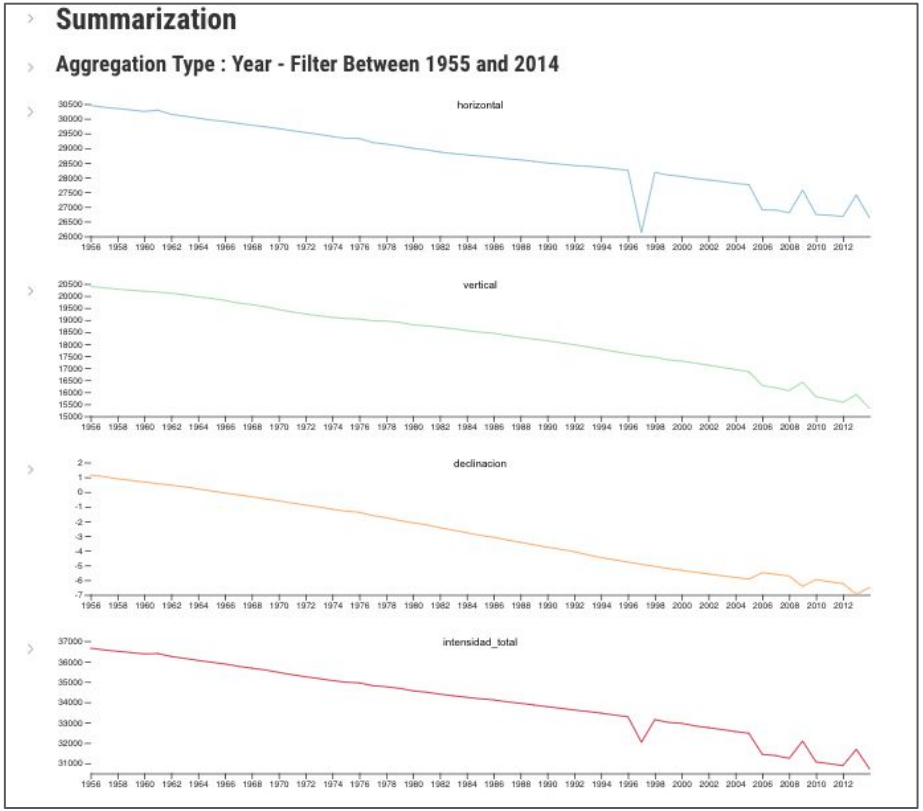
The Solution

- An interactive visualization was created using **d3** and **observablehq** that allows the user to explore the horizontal, vertical, declination and total intensity geomagnetic components in a summarized way by aggregation per year, month or day, as well as the exploration of the detail of the time series for a specific day.
- **Technologies:**
  - **Visualization:**
    - d3, observablehq, html5, css
  - **Data Preprocessing:** *(Summarization and Exploration)*
    - Python, Pandas, Jupyter, JupyterLab

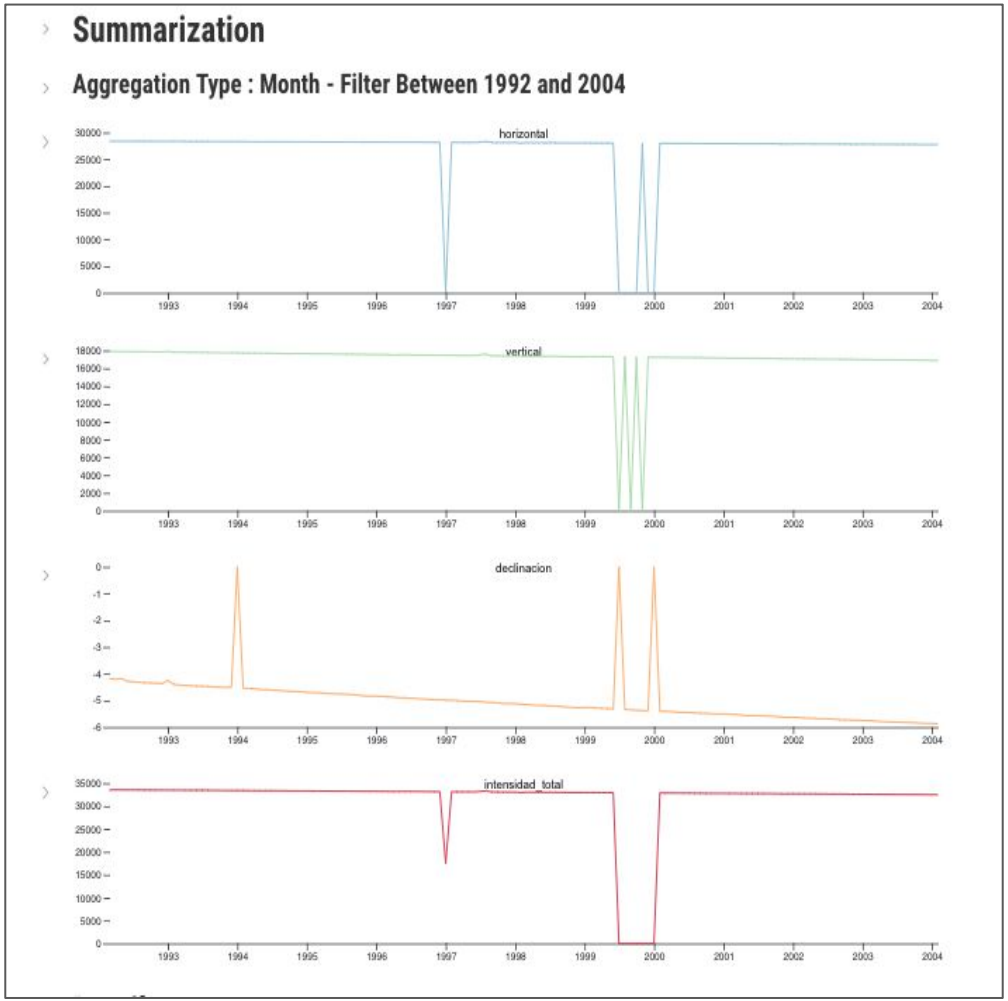
Demo

[A Visual tool for Data Quality Checking of Geomagnetic Data]

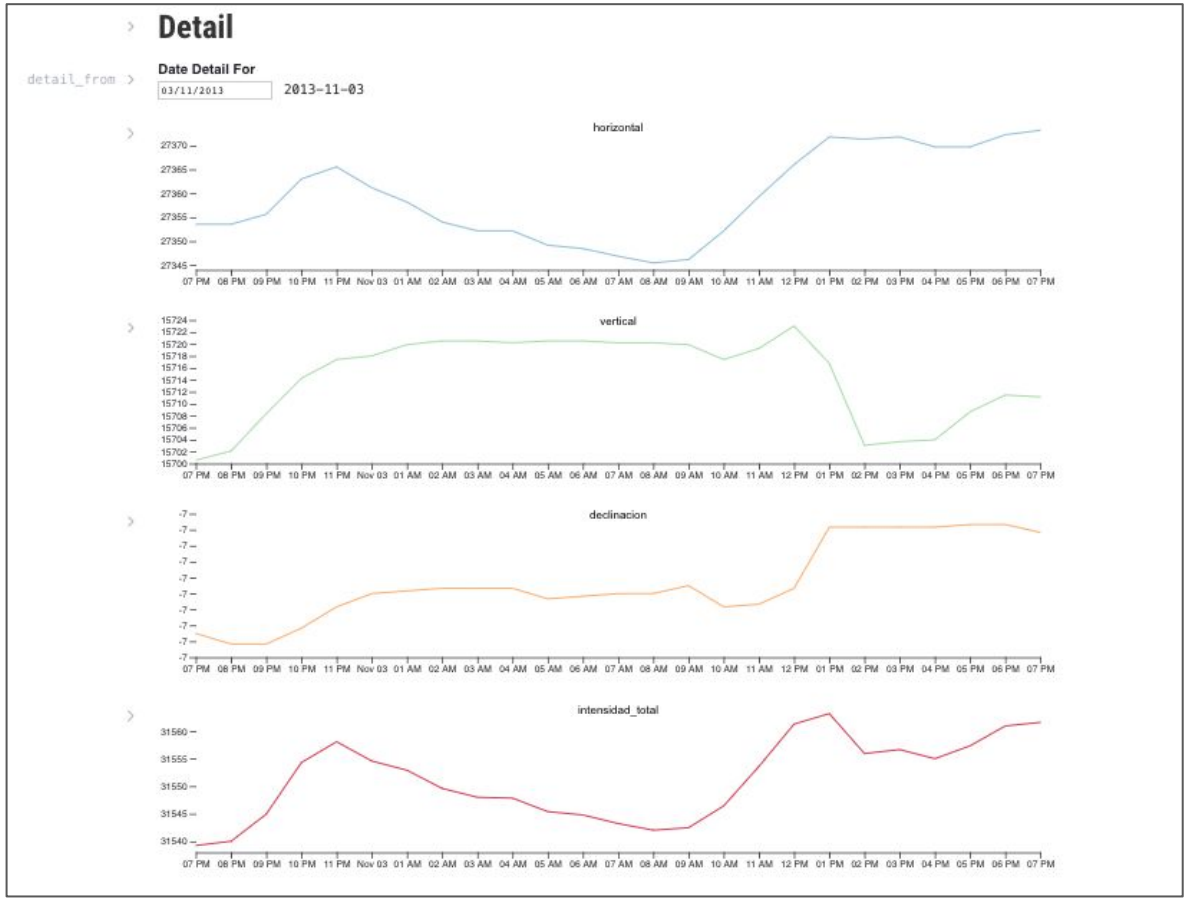




Demo


Summarization
Aggregation Type : Month - Filter Between 1992 and 2004
horizontal
vertical
declinacion
intensidad_total

Demo


Detail
detail_from
Date Detail For
03/11/2013
2013-11-03
horizontal
27370
27365
27360
27355
27350
27345
07 PM 08 PM 09 PM 10 PM 11 PM Nov 03 01 AM 02 AM 03 AM 04 AM 05 AM 06 AM 07 AM 08 AM 09 AM 10 AM 11 AM 12 PM 01 PM 02 PM 03 PM 04 PM 05 PM 06 PM 07 PM
vertical
15724
15722
15720
15718
15716
15714
15712
15710
15708
15706
15704
15702
15700
07 PM 08 PM 09 PM 10 PM 11 PM Nov 03 01 AM 02 AM 03 AM 04 AM 05 AM 06 AM 07 AM 08 AM 09 AM 10 AM 11 AM 12 PM 01 PM 02 PM 03 PM 04 PM 05 PM 06 PM 07 PM
declinacion
-7
07 PM 08 PM 09 PM 10 PM 11 PM Nov 03 01 AM 02 AM 03 AM 04 AM 05 AM 06 AM 07 AM 08 AM 09 AM 10 AM 11 AM 12 PM 01 PM 02 PM 03 PM 04 PM 05 PM 06 PM 07 PM
intensidad_total
31560
31555
31550
31545
31540
07 PM 08 PM 09 PM 10 PM 11 PM Nov 03 01 AM 02 AM 03 AM 04 AM 05 AM 06 AM 07 AM 08 AM 09 AM 10 AM 11 AM 12 PM 01 PM 02 PM 03 PM 04 PM 05 PM 06 PM 07 PM

Insights

1. The user discovered some anomalies in the data using the interactive visualization.

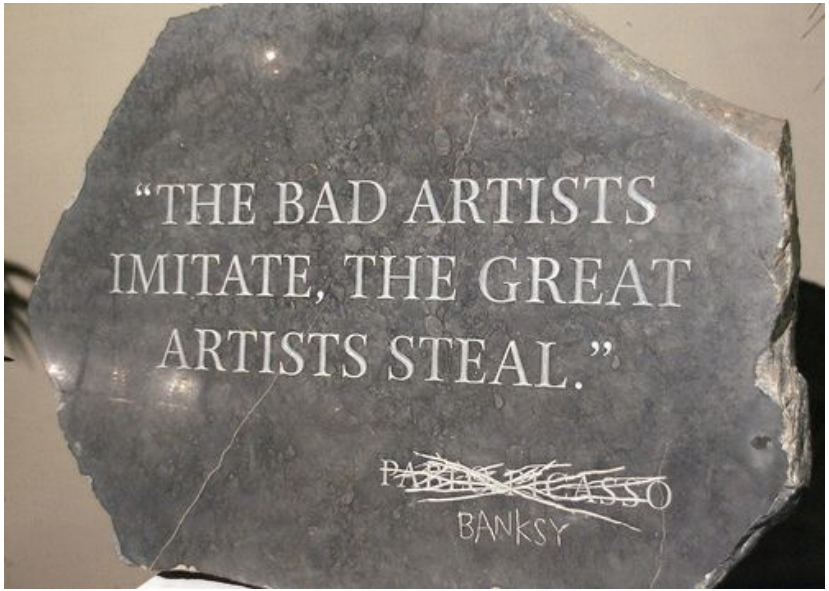
2. According to the owner of the data, those anomalies were caused by errors during the process of the extraction and transformation of the original data.

3. The owner of the data will publish in the following days a new corrected version of the dataset

4. The owner of data is considering to include visual tools like this in his daily work.

Difficulties

- D3's Learning Curve (i.e. data parsing ,)
- It is not easy to *"debug"* using observablehq. *console.log("...blah...");*
- Non-Reproducible examples (i.e. old versions of D3)



Achievements

- The user actually got insights!
- My first work using *observablehq* and it was possible to create a “*reusable*” chart.
- The user got interested about using interactive visualization tools