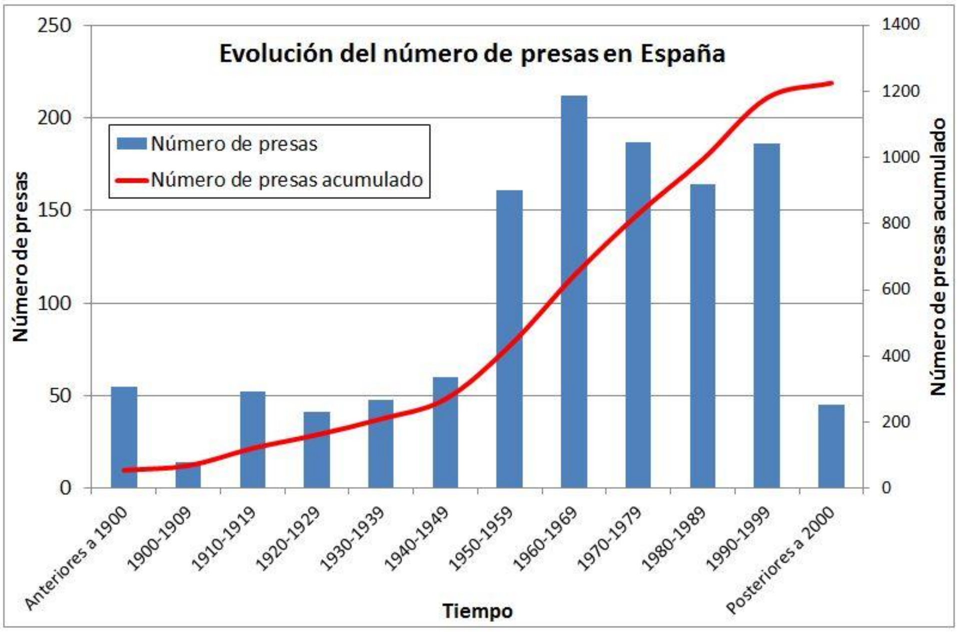
Developing a Visualization Tool to Monitor Reservoir Reserves in Spain

Problem Description and Background

"I have survived a civil war, a world war and a pertinacious drought, so this is not going to kill me" - My grandma, anytime something bad happens. Own translation.

The drought which my grandma mentions so often happened right after the Spanish civil war and is in fact a succession of multiple droughts between years 1944 and 1955. These severe droughts put the already ravaged Spanish economy at the time under huge stress as these droughts caused famine and food rationing.

These famines have left a terrible memory in the minds of those who experienced it. They also explain the urge the Spanish government felt to build many reservoirs and water management infrastructures, as these dry cycles have repeated throughout Spanish history. This was translated in just under 1000 reservoirs being built only during Franco's dictatorship 1939-1975 (see image below).



This makes of Spain the 5th country in the world and the 1st in the EU with the most reservoirs, 1225, whilst being the 29th country by population as of 2019, with just under 47 million inhabitants.

Even though the great number of reservoirs has helped manage droughts in recent time, extreme drought phenomena have remained relatively common, the most recent in 2017. Furthermore, with climate change extreme droughts and floods are expected to increase making water management more important than ever.

Spain is an agricultural powerhouse at a European level, so even if nowadays with all these reservoirs, we would not expect Spain to experience a famine, a very severe drought could be very damaging for Spain's economy, especially in its rural areas.

This also highlights who might be interested in a tool to monitor reservoir fill: the Spanish farmers and other inhabitants of rural Spain, as well as any stakeholder in Spain's agriculture. As we will discuss in the following section, the data which is currently available is presented in a dull and obsolete format, making it hard to use, so developing a tool which allows for better monitoring of the country's water reserves will be interesting for many people and businesses in Spain.

Data Description and Use for the Solution

The data currently available on reservoirs can be found on the site https://www.embalses.net/ . Data in the site is displayed in charts such as the following:

Embalses en Asturias

Pantano	Capacidad	Embalsada	Variacion
ALFILORIOS	8	8	0
ARBON	41	37	-1
DOIRAS	97	87	-1
LA BARCA	34	15	-2
RIOSECO	4	4	1
SALIME	266	225	2
TANES	34	29	-1

Embalses en Asturias (Sin datos Semanales)

Pantano	Capacidad
EL FURACON	1
LA FLORIDA	1
LA GRANDA	3
LA JOCICA	1
LA MORTERA	0
PRIAÑES	2
SALIENCIA	0
SAN ANDRES TACONES	4
SOMIEDO	6
TRASONA	4
VALDEMURIO	1
VALDUNO	0
VALLE	0

We have a chart like this for every province in Spain, autonomous region and fluvial basin. We could use any set of tables to perform our study, but we have chosen to use the tables from the autonomous regions (provinces are contained inside them), for practical reasons.

First of all, we find that there are two different charts for most regions, one of them with "live" data, updated weekly, and the other just displaying the capacity of the reservoirs.

For the weekly updated chart, we find the reservoir capacity, the water reserves and the variation since the previous measurement. In both cases, the first column displays the reservoir's name.

The objective of my project will be to develop a visualization tool that allows us to visualize and monitor this data in a more intuitive and visually attractive manner.

The data will be used as following:

- The reservoir name will be used to find the reservoir's location and precise name using the Nominatim API to search OpenStreetMaps.

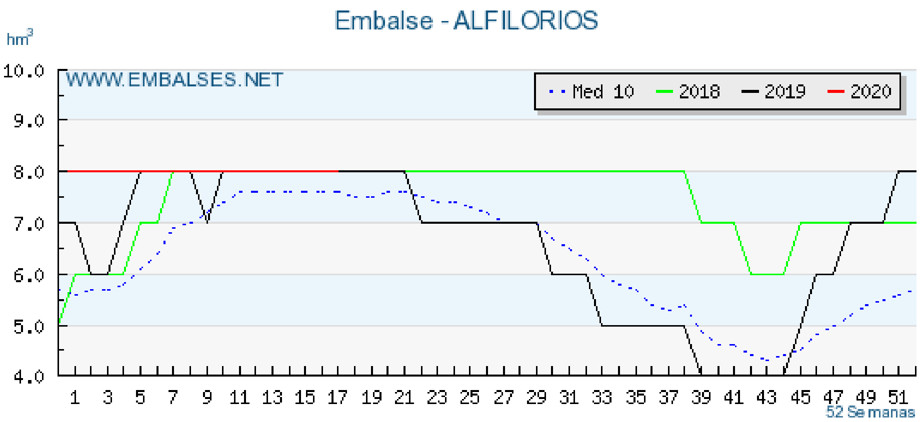
- We will then use the reservoir's precise name to retrieve the reservoir's shape from OpenStreetMaps using the OverPass API, in order to use it for data visualization.
- We will use the reservoir's capacity and reserves (in absolute terms) to calculate the reserves as a percentage of capacity. We will use this metric to give a colour to each reservoir's shape in the map as a function of their relative reserves.

We will also use this data to find other insights on the distribution of reservoirs:

- How many of these are large, medium or small?
- Is there some correlation between the relative reservoir's reserves and their geographical position?

Data Limitations:

The website includes graphs such as the one below, showing the evolution of the reservoir's reserves in time, for several years, however, we cannot access the data from the time series.



Having the data from the time series, would allow for a much deeper analysis of the Spanish reservoirs, unfortunately, I have not been able to come across a source with this data.

Methodology:

Exploratory Data Analysis:

First of all, we want to have a quick look at the data in order to extract some early insights on how to deal with the data, and maybe also interesting hypothesis to be verified throughout the project.

In addition to some basic metrics from the dataset, we will look at the size distribution of reservoirs in terms of capacity, the level of relative reserves for each province in Spain and the amount of capacity built in each province in Spain.

Dataset

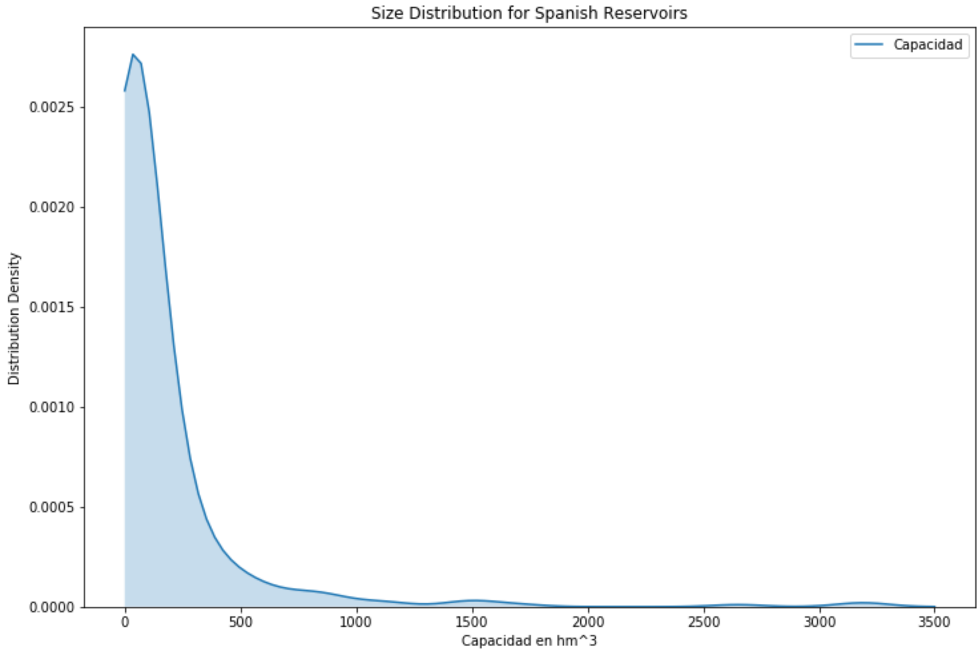
Our dataset contains 356 reservoirs, many less than the 1200 that exist in Spanish soil. This is due to the fact that most reservoirs in the country do not have weekly updates and have not been considered useful for this exercise.

[15]:

	Pantano	Capacidad	Embalsada en Porcentaje
0	AGUASCEBAS	6.0	100.000000
1	ALMODOVAR	6.0	33.333333
2	ANCHURICAS	6.0	50.000000
3	ANDEVALO	634.0	52.839117
4	ARACENA	129.0	51.937984
...	...	...	...
351	ARGOS	10.0	70.000000
352	LA CIERVA	7.0	71.428571
353	PUENTES	26.0	69.230769
354	SANTOMERA	26.0	23.076923
355	VALDEINFIERNO	12.0	16.666667

Size Distribution for Spanish Reservoirs:

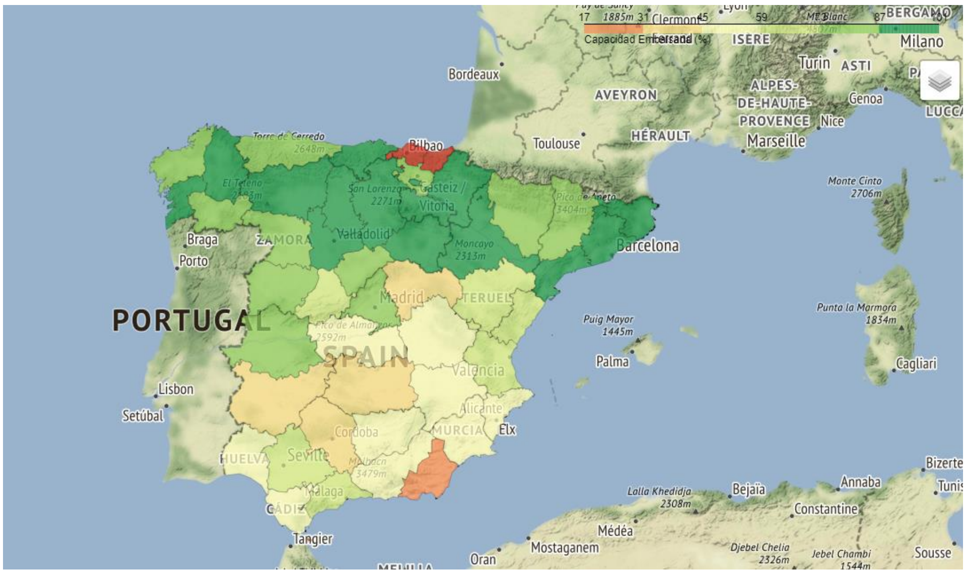
One of the first metrics we want to look at is the distribution in size for the reservoirs.



We can see that there are much more small reservoirs in comparison to large ones and in fact the amount decreases sharply, nearly exponentially for larger reservoirs.

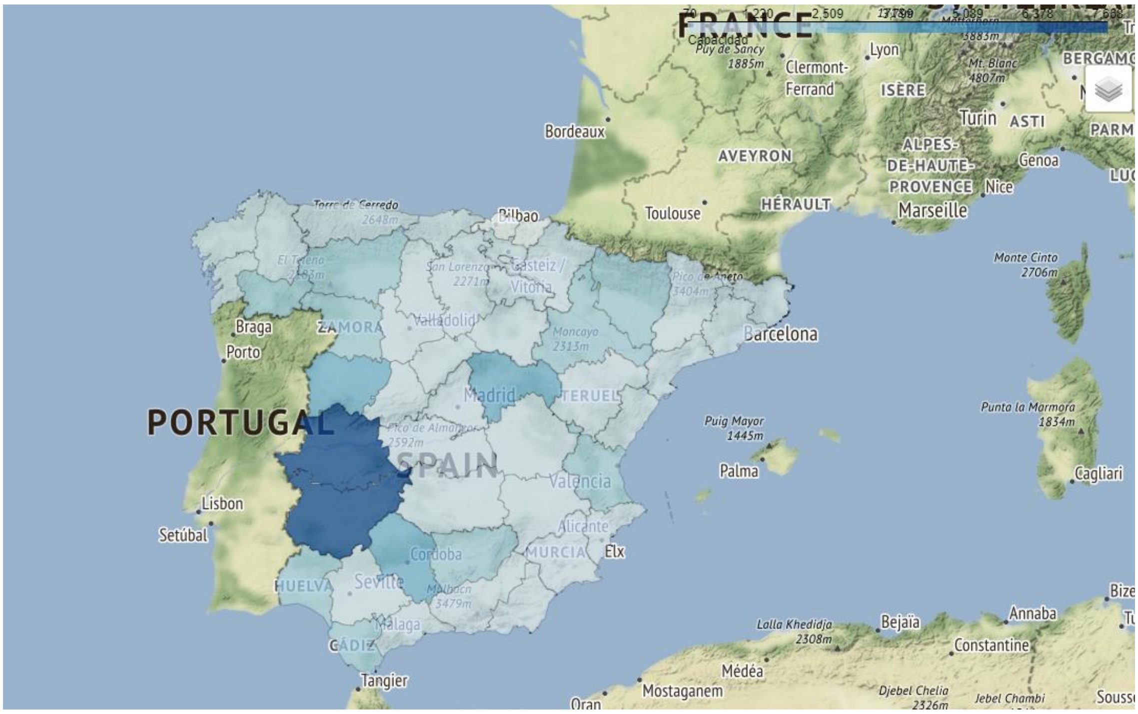
Geographical Distribution of Reserves (in Percentage of Capacity)

In this section we use the cumulated information in the site for the different provinces. This will allow us to see visually if there are regions which have lower levels in their reservoirs and if there is a geographical pattern. In our case, it seems like reservoirs further North tend to be fuller than those further South, however, it is soon to confirm this statement. We note that the two red provinces in the North do not have any reservoirs.



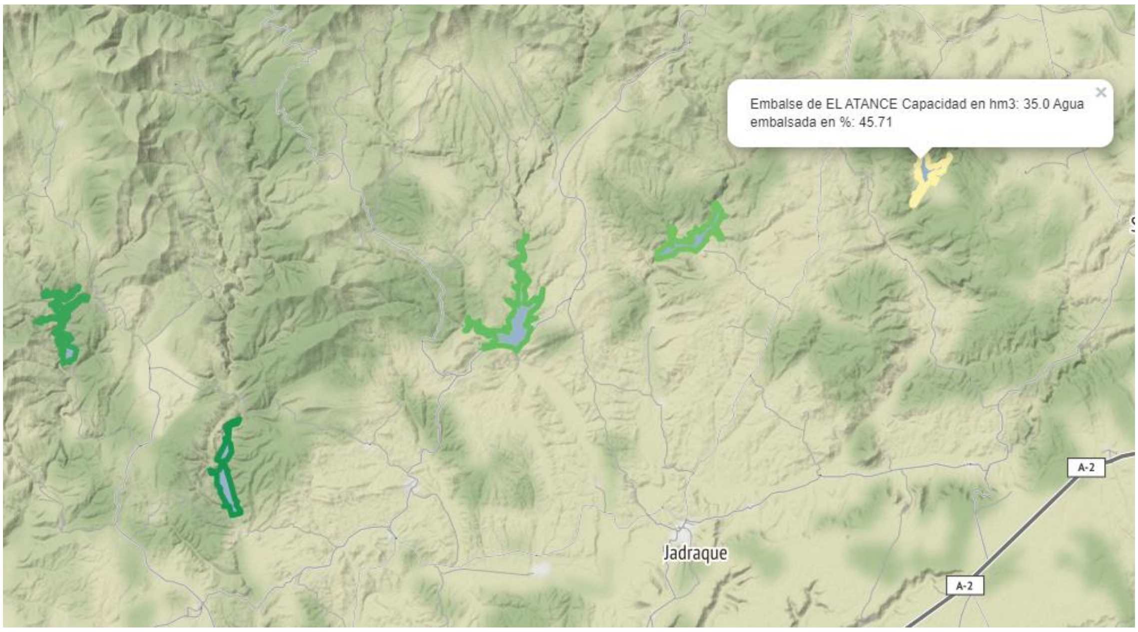
Geographical Distribution of Capacity

We will now quickly explore the amount of reservoir capacity built in each province of Spain. The objective is to see broadly if there is any pattern regarding this distribution. We can easily see that two provinces dominate in terms of absolute capacity, by a large margin, but other than that, we cannot see any pattern. We will verify this hypothesis later on.



Tool Development:

The objective is to create a tool which allow us to evaluate at a very local level the level of reserves in the surrounding reservoirs in an easy and intuitive way. The tool should allow us to see at a glance the state of the reservoirs in a certain area. See a screenshot of the tool below, for illustrative purposes:



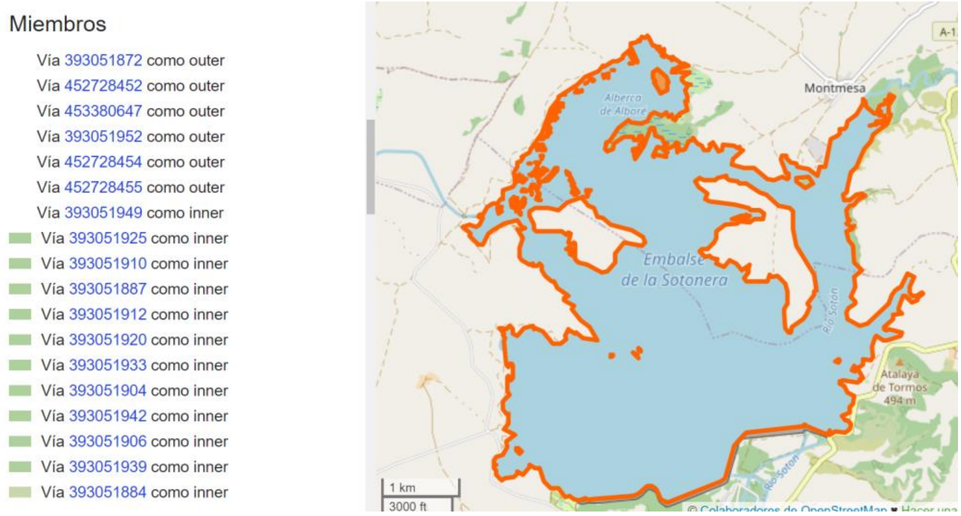
To produce such a map, we need to obtain a source of data with the shapes of the reservoirs, in order to highlight the silhouette. Then the colours will be assigned as a function of the relative reserves.

Reservoir shapes

Many of the online maps give us the highlighted silhouette of the element we look for, whether it is a city, a lake or a region. In our particular case, we also want a map provider whose API is free to use. In our case we found that the Foursquare API contained very few reservoirs.

On the other hand, we found that many of Spain's reservoirs appeared in OpenStreetMaps, so we decided to user their services. In order to retrieve their shapes, we used the OverPass API, used for bulk download from OSM. The problem with this API is that the quality of its search was very poor on its own, and I had to use the Nominatim API to retrieve a more precise version of the name than the one provided in *embalses.net*.

Once this has been done, we can recover the reservoir shapes, which are often composed of several 'members' which are parts of the contour. See image below:



Each of these members contains the nodes for their part of the contour, which we will search to retrieve. Once we have a list of the nodes, we will easily extract their coordinates and use them for the final tool.

Results:

The visualization tool:

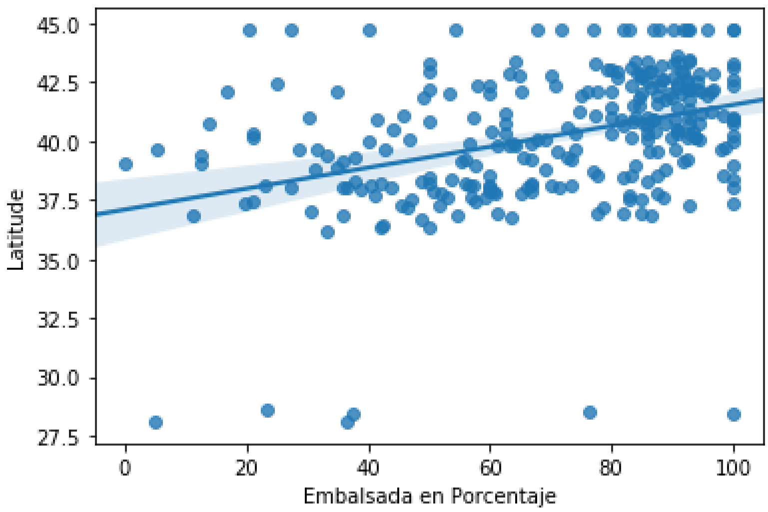
The final version of the tool allows us to easily see the main reservoirs in Spain as well as their level of reserves:

Furthermore, it allows us to see at a glance if certain areas might be subject to hydric stress, even at a very local level. I think this tool represents a massive improvement in terms of ease and aesthetics when monitoring the level of water reserves, satisfying our initial objective.

Other results:

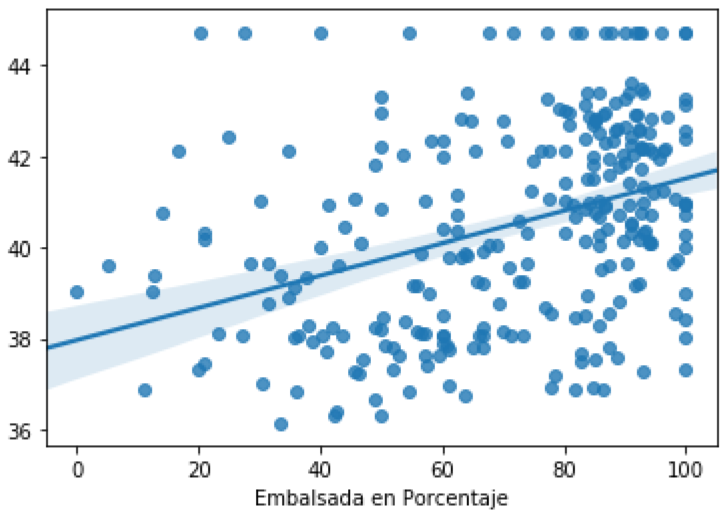
To verify our initial hypothesis that there might be some geographical correlation between the latitude (North-South) location and the level of reserves, we can plot a regression plot using the relative reserves of each reservoir and their latitude:

```
R-squared= 0.14040099740295647
```

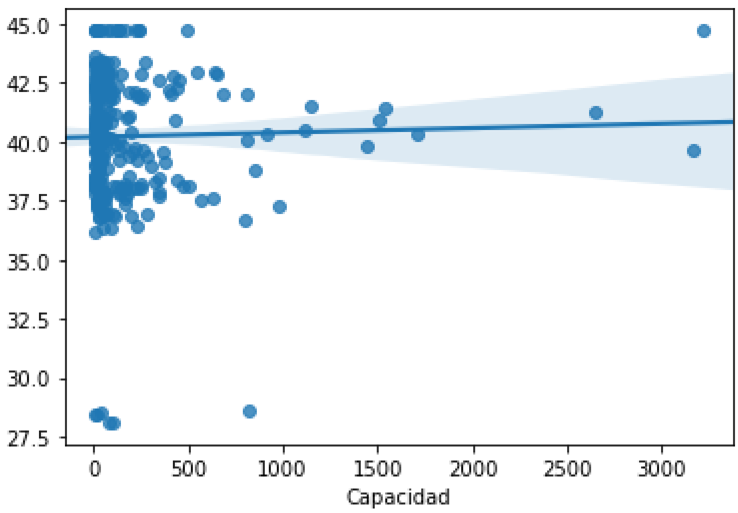


Due to the extremely low R-squared, it seems that there is nearly no correlation between both, however, given that there are a group of reservoirs to the South, in the Canary Islands (far away from continental Europe) could be skewing the results. However, when we eliminate these values, the R-squared is even lower, so we cannot conclude that there is a strong correlation between latitude and reserve levels. This was also verified for different degrees polynomial fits, with similar result.

```
R-squared= 0.13658572415560635
```



If we apply the same reasoning to our second hypothesis, we can easily corroborate that there is no pattern between geographical position and storage capacity built:



45.0
42.5
40.0
37.5
35.0
32.5
30.0
27.5
0
500
1000
1500
2000
2500
3000
Capacidad

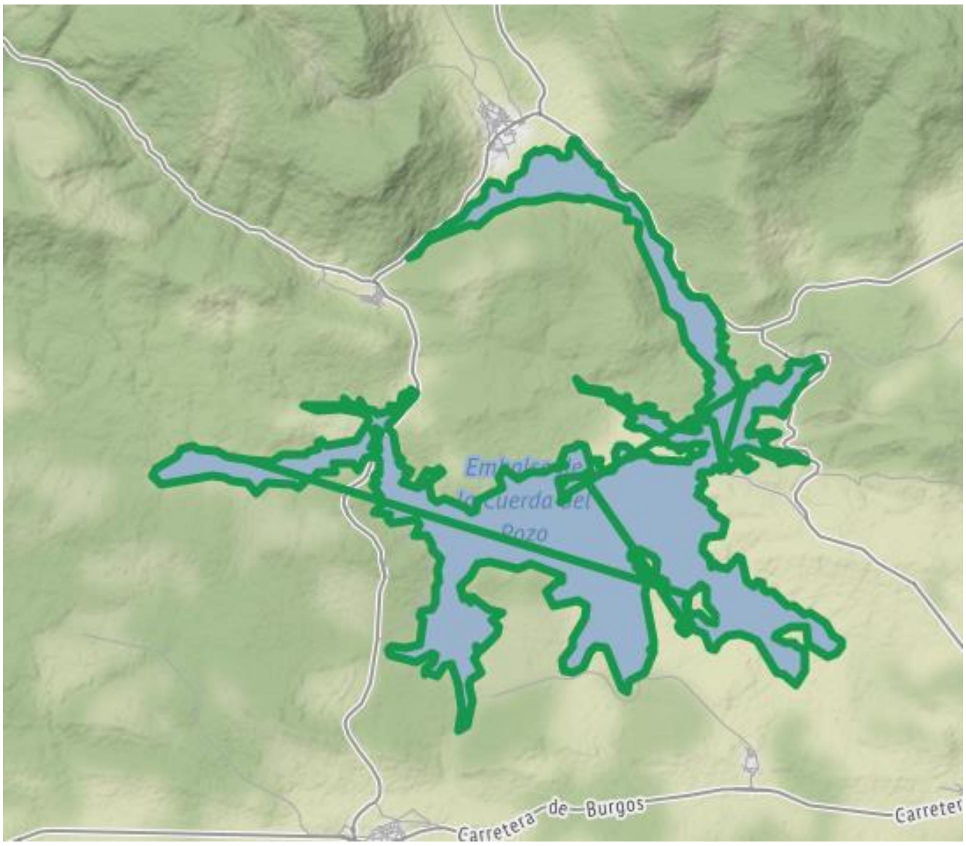
Discussion:

In general terms I am quite happy with the final result of the tool. However, there are certain things that should be improved in order to deliver a top-quality tool.

First, I would like to mention two issues with the data sources. On the one hand, the database in embalses.net is not exhaustive and only has current data for around 360 reservoirs out of more than 1200. Furthermore, out of these 360, I was only able to retrieve data from around 290. This is due to the fact that not all reservoirs appear in OpenStreetMaps, and some that do are not appropriately tagged, meaning that they cannot be picked up easily in an automatic way. It would be very time consuming to amend this database and it is not in the scope of this project.

I also faced an issue when retrieving the shape data when using the OverPass API for the largest and most complex reservoirs, because of the number of requests needed. It was not possible to include them in the map.

Finally, and also for reservoirs with complex shapes, built with multiple lines, the final shape shown has unexpected lines crisscrossing it (see image below). This is because the different line segments do not follow a specific order when retrieved from OSM. It would be possible to solve this issue, but again, it would be extremely time consuming and not of much interest for this project.



Conclusion:

Even though the tool has room for improvement, overall, it fulfils its purpose very well, as it will allow to better monitor water reserves even at very local levels.. I am very happy with the result; however, it was quite unfortunate that there was no data to complete a time series. This would have been very interesting to include in my project.