

Building an Itinerary:

First, we need an itinerary in the form of a detailed data sheet. There are three methods we have used so far for creating the datasheet: digitizing a pre-existing itineraries (the easiest), building an itinerary from published sources or collections (moderately difficult), or creating our own from primary sources (the most labor intensive). Pavel Oleinikov at Wesleyan (one of the members of the Travelers Lab) has done some work trying to extract the data from books using text recognition, which is promising, but we still do a lot of the data construction by hand.

(Here's Pavel's write-up of the first shot:

<http://travelerslab.research.wesleyan.edu/2016/09/13/notes-on-the-margins/>) The first itinerary we worked with at Marlboro College came from a very well laid out, already published itinerary of Jaume II of Aragon you can see here:

https://ifc.dpz.es/recursos/publicaciones/28/86/_ebook.pdf Chances are what you use will look different from either of these - there is a lot of variation.

The Datasheet:

Despite the varieties of itinerary data, it is useful for us to have relatively uniform data formats to make our work more comparable. The data also needs to be flexible enough to capture the real variety of the historical sources themselves, which we do not have the power to change. After some experimentation, there are a few specifics that may seem odd, but prove to be important down the line. Every record gets its own unique ID - this should be fairly clear. Despite the redundancy of it, we've found it works best to give every record different columns for the day, month, and year. Some of the columns will be obvious - the location as stored in the record or publication itself, as well as the normalized location we have decided on in a modern spelling. Notes holds any other important historical data or even temporary notes as work progresses on the project. Finally, we record a set of codes that keep tabs on certain qualities about the locations themselves including uncertainty or probable errors from the sources themselves. There is a rubric and guidelines for that column that we will return to below. Lastly, since this is a group project, everyone should initial each record they produce, making it easier to ask about odd entries or get answers to questions as necessary (and, honestly, it makes it easier to grade...) Our final data headings tend to look like this:

Weekday | day | month | year | source_location | modern_name | geo_id | loc_codes | notes | initials

Use the Sources:

If you don't have a pre-built itinerary, you get to build one yourself using sources! This is the usual work of history. You should still build a datasheet along the lines described above, but filling it in will feel substantially different. Instead of filling in the information by date (starting at the beginning of the itinerary, going to the end), you will be entering lines as you find them in the data. It will be slower going in part because each line will need a reference for the source of the

information (similar to the references found in the itinerary of Jaume II above. Also, you will be entering information as you find it, which will likely be chronologically out of order. As the information comes out of reference works full of letters or edicts that indicate the location and date for the person you are interested in, the timeline will slowly fill in, but we will need to check back periodically to find out what portions of time are short on data. When you fill in a line, you should include as many details as are available from the source. Any detail that the source does not provide, you should leave blank and explain in the notes.

Going from a Datasheet to a Map:

First, we had to take the list of names and remove duplicates. The easiest way to do this was:

Step 1: Check the Itinerary against an existing gazetteer if available. The best way to do this is to use the `functions_list` available in the "Itinerary-Project-Code" folder. The easiest way to do this is to download all five components from the code library (everything except the "README") into a single folder (there are many ways to do this - cloning the repository to github desktop or using `downgit`: <https://minhaskamal.github.io/DownGit/#/home> are probably the easiest - copy the folder url into `downgit` and you're done!) You will need to run the modules in an environment that supports `pandas` - if you are unsure how to do that, see the supplementary instructions in the Tool Box. Then place your itinerary (as a `.csv`) as well as the existing gazetteer (also as `.csv`) in the folder with everything else. Open the `functions_list.txt` file and enter "yes" to process an itinerary, the appropriate filename, "yes" to lookup names in gazetteer, and the gazetteer filename. You can also write "yes" in the "output itinerary as gazetteer" section along with the filename for export. Finally, run the "interactive_gaz_itin_commands.py" program.

Step 2: If an existing gazetteer is not available where you think some entries might be found, follow the same steps as the above, except enter "yes" under "output itinerary as gazetteer" without the gazetteer lookup function.

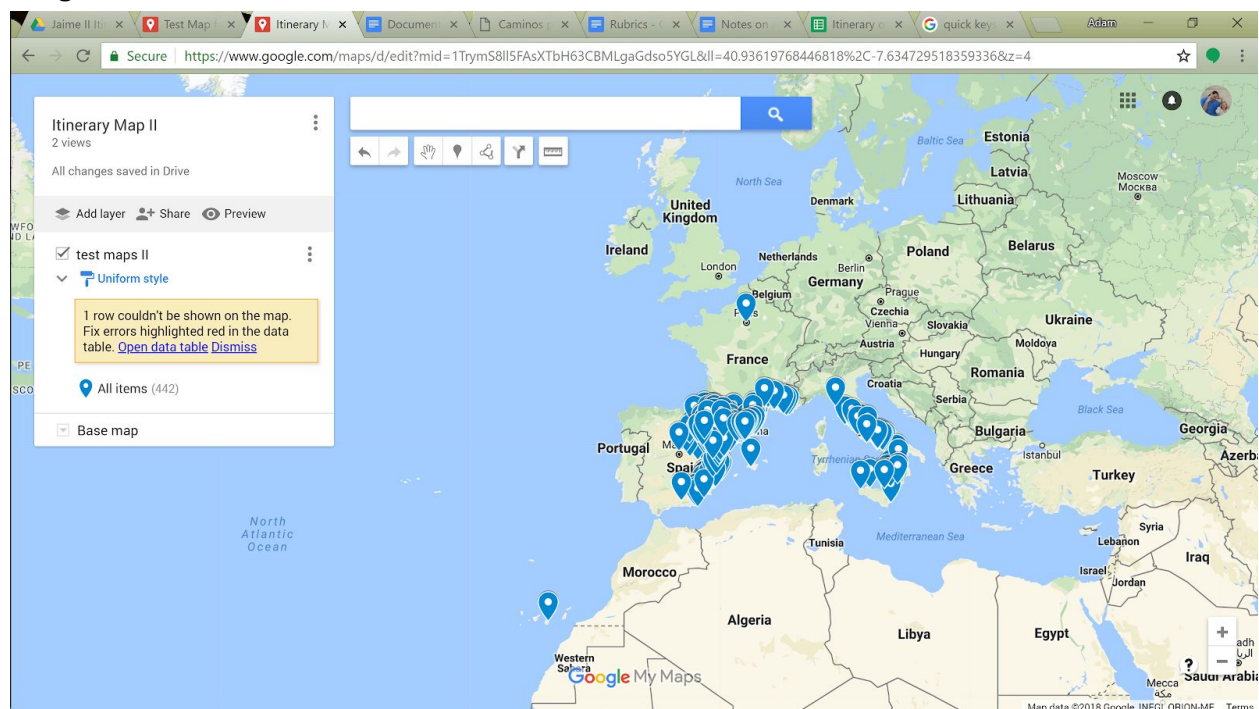
This should produce an alphabetized gazetteer to work through with those entries already found marked appropriately. If you checked the itinerary against an existing Gazetteer, you should work through the list of names in the new "gaz_match" column. Any name with an exact match will simply say "exact match" in the column, while those with similar names will have guesses in them for review. Those without any similar name will be blank. To continue working with only those unknown locations, you can filter out anything with an exact matching using this column.

At this point, reading through that list was very important since it reveals a lot of misspellings and other small errors - for instance, the alphabetized list of our original itinerary included "Zargaoza," "Zaragosa," and "Zaragoza" with a carriage return in the box, along with the actually correct "Zaragoza." Consolidating these errors before moving on will save you a lot of work down the road. After checking the spelling and other errors of the list, repeat steps 1 and 2 above.

Using this new gazetteer of (hopefully) unique places, upload the new .csv file into a google sheet and run the Google Sheets add-on (search under add-ons) called “Geocoder by Awesome Table” to automatically find and assign coordinates to each place name. The plug-in works better if you can create a country column and restrict the location of each city by its country. However, this can be difficult when there are lots of towns you do not recognize and can not place in their modern nation state. When running the plug-in, it will work better to check the box labelled, “Try wider results.” This will return more hits for more items (but also more errors...but that will be dealt with later.)

Using the latitude and longitude generated by the plug-in, we created a map using the Google maps interface, with each point marked using pins. [See Image 1] A number of points were obviously put in the wrong location, Spanish cities ending up in Florida, North Africa, the Philippines, etc. Multiple cities that the google API cannot recognize also get dumped at the location of the country; hence there are a small handful of locations near Paris standing in as "France," and numerous unknown locations at the official coordinates for "Spain" (40.4636 N Latitude and -3.7496 Longitude). [See Image 2] In the case of Jaime II there turned out to be nearly twenty such points. Simply disaggregating the points by clicking and dragging them to a nearby location helped us to clarify to students how many points there are to work on in that space. [See Image 3]

Image 1:



Note: This is the map with all attempted coordinates looked up from the plug-in and imported from a google sheet into a google map. Each pin can be clicked on and interacted or edited.

Image 2:

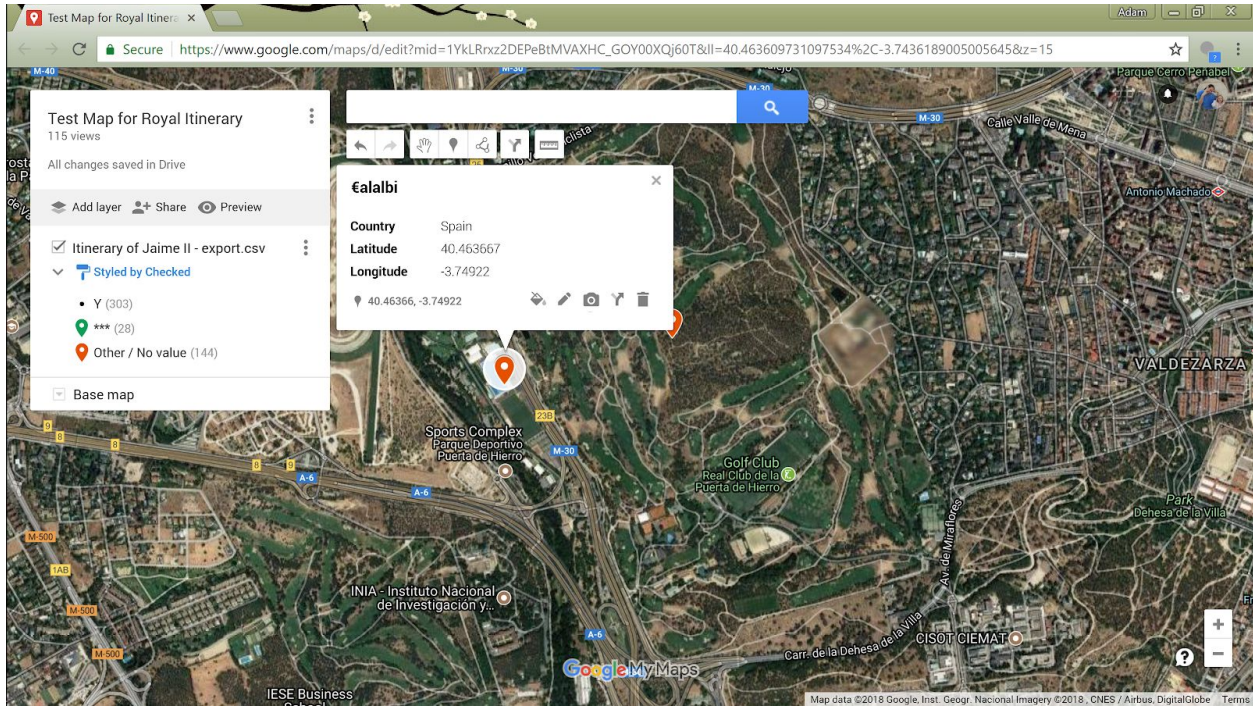
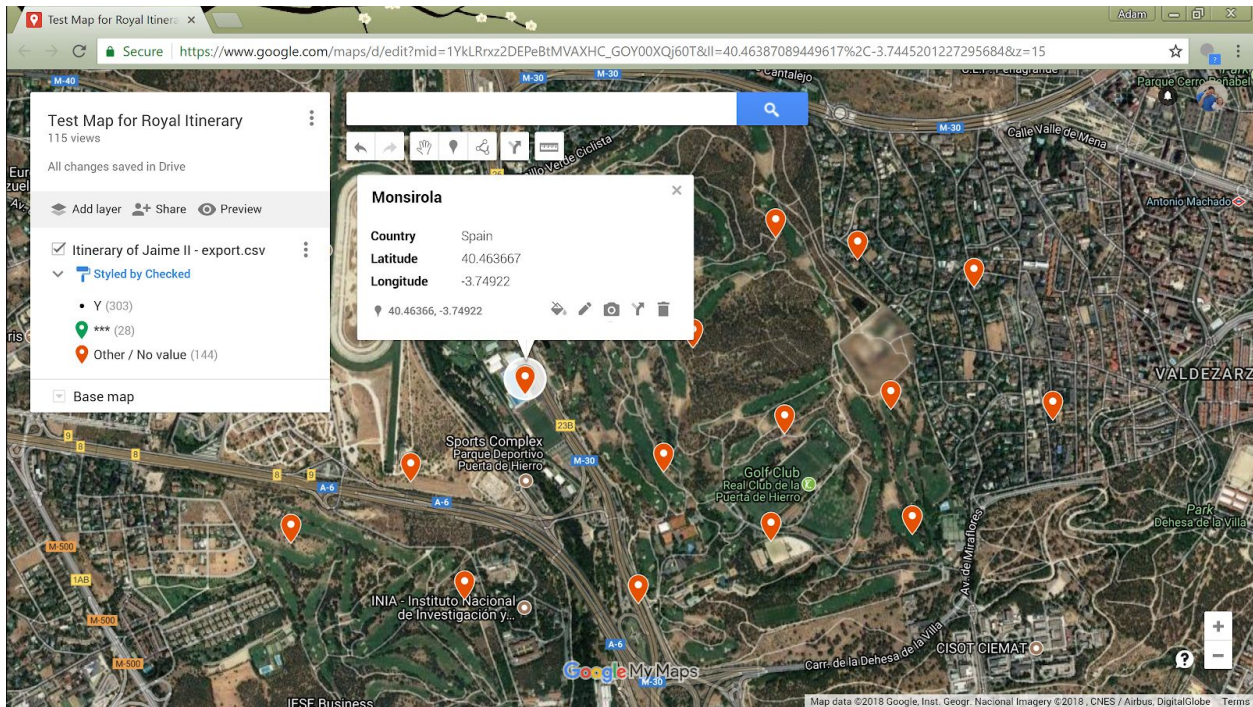


Image 3:



Note: These are all the pins originally placed on “Spain” with no town name. They have been pulled apart from Image 2 so they are all easier to see - none of them are accurate or checked.

Checking and Verifying the Accuracy of All Points:

When checking and verifying points, the best way to understand this workflow will be to look at the flowchart that should help guide you through the process. The end goal of the process is a list of verified geographic references that includes all known locations in the itinerary. The references will make possible a range of geographic analyses once the data collection and verification is finished. The process also often catches a handful of extra errors or typos that entered in during the initial building of the itinerary.

There are two basic lines of attack for this part of the process. Some of us worked on the points going linearly through the dataset (taking the name in the first column, entering it in the map search bar and working through that point before going on to the next). The advantage of this method was that in order to check some of the points, we wanted to look at the locations the king (or pope or noble) visited directly before and after (so we would know if one location was impossibly far, etc.). Working through in a chronological order could make that simpler, especially if you aren't familiar with the area.

The second approach involves working geographically - zooming in on a part of the map with lots of pins and checking point by point within that region. This was, for some of us, a faster way of clearing data, especially if we already had any knowledge of the geography of the area. This can also be a quick way to go through "easy" pins that are already placed on the correct town and only involve checking, without further research.

Every single pin needed to be checked to ensure we had an appropriate latitude and longitude! While checking each point, we also assigned a "certainty value." This number described how confident we were that the coordinates assigned matched the actual location of the place. You will need a certainty column and put a certainty code for each location in the gazetteer following the rubric provided. **[See Location Certainty Rubric]** When you come across a difficult location or have a guess that you feel is reasonable but want more assurity, you should enter what you consider to be accurate and include more information in the notes, possibly flagging the entry for checking by a professor (we used a triple asterisk for the professor check tag: ***). **[See Image 4]**

Finally, as the pins get checked, remember to mark them on the map and the pins will change color, calling attention to the locations that still need checking or those that require professor attention. If you want to have a professor check on a specific pin, just place the usual tag (again, we used the asterisks "****") in the "check" box in the map - click on the pin, click on the "edit" button, change the box and save the pin and it will change to a third color. **[See Image 5]**

Image 4:

Itinerary of Jaime II - 1291-1327

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

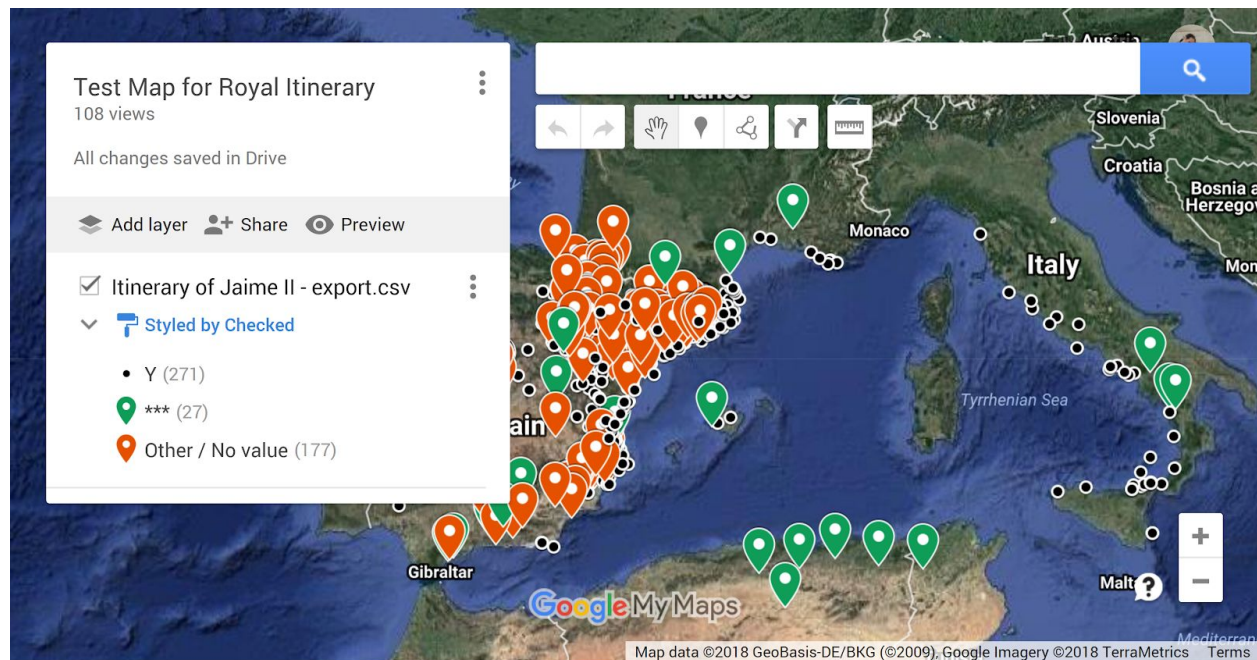
100% 123 Arial 10 B I S A

	A	B	C	D	E	F	G	H	I	J
53	Alpuente		Spain	Alpuente Spain	39.8757866	-1.0137828	Y	5		
54	Albalate de Zorita		Spain	Albalate de Zorita Spain	40.3084745	-2.8439975	Y	5		
55									The record has Jamie going to Guadalajara and from it to Tauste and back then to Gurrea de Gallego and back within the span of a couple of days, these distances are significant and Guerra de Gallego and Tauste are closer to one another than either is to Guadalajara so I am unsure if he could travel from Guadalajara to these other cities over the course of a single day and also why he would return to Guadalajara between trips rather than visiting both the two cities that were nearer each other before returning to Guadalajara. While this is true, it seems like Tauste and Gallego are the exception and Guadalajara is the more correct location	
	Guadalajara		Spain	Guadalajara Spain	40.632489	-3.16017	Y	5		
56	Tauste		Spain	Tauste Spain	41.9202247	-1.2578169	***			
57	Gurrea de Gállego		Spain	Gurrea de Gállego Spain	42.0161963	-0.7722974	***			
58	Atienza		Spain	Atienza Spain	41.1981179	-2.8708039	Y	5		
59	Ágreda		Spain	Ágreda Spain	41.8564653	-1.922138	Y	5		
60	Tarazona		Spain	Tarazona Spain	41.9042972	-1.7222535	Y	5		
61	Magalón		Spain	Magalón Spain	41.8321822	-1.4615955	Y	5		
62	Berbegal		Spain	Berbegal Spain	41.9611832	-0.0029143	Y	5		

Original Itinerary Data List of Place Names Calculations Lat-Long

Explore

Image 5:



Note: The black pins are complete, green need checking or are tagged for a professor, and orange have not yet been looked at.

Reference from flowchart:

*****Look back at the original itinerary. (Distance)**

You will want to check the original itinerary to find the known locations immediately before and after. Unless there is a large amount of unknown time (and maybe even then), you'll want to look for a matching location that seems reasonably on the way between the before and after. If not between, consider distance. For Jaime the II, it was highly unlikely that in the span of one day he would have been able to go a hundred miles out of the way. If there was a period of several weeks with unknown locations beforehand, though, that distance may have been possible.

*****Consider language changes (Naming)**

Many of the place names we were looking at had different names in Spanish, Latin, Catalan, and modern English names. Some variations were literal translations that didn't obviously resemble each other. If you have experience in languages, it will help to consider how pronunciation and usage might change the spelling of place names, and search other alternatives. It also comes into play for redundancies- there were plenty of repeated names that came down to translations of "mountain" "village" or other generic terms. A visual scan of the suspected area can help find modern places with similar names. (See 3. below)

*****Old Cities**

Some of our modern cities were built around or over the corresponding medieval cities. At times, these cases can be identified through the satellite view. Some of us didn't have any experience with medieval infrastructure and couldn't identify the older cities, but when we or our professor could find them, the pin would be placed inside the old city. (See 6. below)

Specific Problems We Encountered:

1. The location only exists in documents and is now unknown. We guessed at the location as best we could and included references to the documents in the notes. Often these got referred back to the professors.

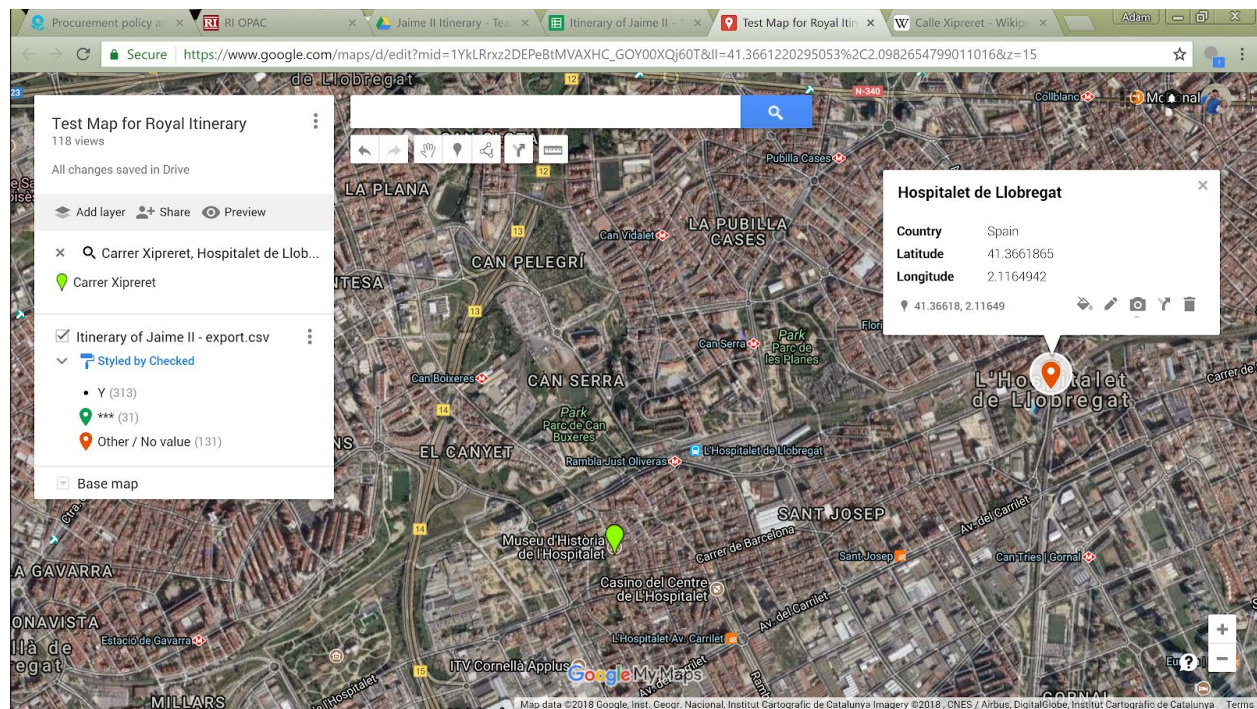
2. Distances that didn't make sense, but had no other explanation. We did a lot of digging on these ones, and some of them we ended up marking using the Itinerary code rubric rather than anything about the gazetteer of names itself.

3. Names changing over time was a common problem. This usually took one of three forms. One, small name changes, Vañón to Bañón for example. This is easier to fix if you know the language and know what letter changes are common. For students who were unfamiliar with Spanish or Catalan, this was something that we shuffled off to Adam because he would have a few good guesses in a couple minutes instead of a few hours and could generally resolve these quite quickly. The second, issue was cities just changing name. In some cases this was due to letter swaps like mentioned above but often it was a result of kingdoms conquering or getting conquered and therefore the name being changed at some point between when Jaime II was there and current day. An example of this was Fraino becoming Fresno (and located in a later itinerary with multiple names). For some cities, if we suspected the name had been changed, a google search and a scan of the Wikipedia page would sometimes give this history and list past names. Often this name change was combined with a letter swap and so students would get a short list of possible place names it could be under now, and Adam's knowledge of Spanish and Catalan would help narrow it down further. The third name issue we had was solid multi-name problems like Vanon and Banon, but also the modern French: Le Boulou which in Catalan is Voló.

4. At times locations are vague: near the castle, for example. When this was the case, the pin itself was dropped "nearby" and somewhat randomly and given a lower certainty accordingly. Sometimes, multiple location possibilities couldn't be ruled out by distance. In these instances, we made our best guess and marked a low certainty.

6. Relocating pins away from modern centers...Hospitalet is a good example. Many towns are quite small, both then and now, but those that have expanded vastly (Hospitalet which was a small parroquia outside of Barcelona - perhaps 2,000 people - is now part of the Barcelona megalopolis and the modern incorporated city has over half a million residents. Hence, the modern center of the city is nowhere near the actual village of the 14th century and should be moved - google places their pins over the name, which does not necessarily work. If you find an example, move the pin to the old center and change the coordinates accordingly. **[See Image 6]**

Image 6:



Note: The pin is currently on the modern name in the modern center and the green extra is the known center of the old city.