

Project plan

Humanities research question

I am planning on doing my project about the domestic travelling of Finnish politicians during the latter half of 1900s or maybe a shorter period within that period. I would like to see if the following questions can be answered.

1. Which places did the politicians visit and does it change over time?
2. Does the party of the politician affect where they travel?

Data

I am trying to answer these questions by using data from Finna API. I am planning on using all the photos that come up by searching a politician's name and the metadata that accompanies it. Based on a quick glance to the data, photos are quite likely to include the creation year or even date and place with the actual picture.

In addition to Finna's data I will use data from Wikipedia for getting lists of both Finnish politicians and towns with coordinates to allow creating the queries and plotting the results on map.

Processing, cleaning, transforming

For the list of names, I will probably have to use python to change them to form that can be used in API queries. In other words, remove all Scandic characters. And from the list of towns I will probably have to remove all the duplicates, if there is any.

For the actual data I get from Finna, I will try to match with Python the metadata with the list of towns and with regex try to find strings that are in format of dates. Then I will have to create a data frame based of all necessary data extracted, so I will in the end have data frame that includes every photo that has a date and a town included in its metadata.

Analysing the data

I will use geopandas and matplotlib to visualize the data. I would like to draw map from every year of the wanted timeframe to show where there have been visits and how many – maybe show which party has been the most active. In additions, I think it will be useful to make a line or a bar graph to show how the visits evolve over the years in different towns, since it might be hard to see from the maps alone.

I think this will help to answer the humanities question since it can quite clearly show where the visits have occurred and how much of them has been. Also, it can be made so that it shows where the different parties visited. In the end, I think this visualization is quite good, but there are other problems, which I will discuss next.

Critical analysis

First problem I see is with matching the town names and metadata as well as matching the dates and metadata. Since Finna's metadata structure is quite flaky, I think it will be best to query for quite a bit of metadata and not just try to search just for the date of creation or place of creation, since it seems that the useful information might not be there and can be included only in the summary or the title of the photo. But because of this there can be included also false or additional information that can match with the town name or date, causing a bad hit. For example, if there is photo that is taken in Espoo and there is a description "A politician arrived from Tampere to Espoo", the program will match with Tampere instead of the correct Espoo. Other very big challenge is the problem of querying for an exact politician. Even when searching with whole name, you can't be sure if you are getting pictures about the politician or a same named person.

There are also problems that could be dealt bit more easily but might be too time consuming for this course or don't have clear answer. For example, I will probably match the town names in lemma form, which causes that some searches might not find a town name even though it is described. For example, if there is only a description of "Helsingissä" the lemma of Helsinki will not match that. Other similar problem is the question about duplicates. If there is a picture about a politician's visit to some event, there can be multiple pictures about the same event. If there is an exact date, it will be easy to remove duplicates, but if there is only a year, it can be hard to know if the pictures are from the same or different visit.

The biggest problem I see with this project is the bias of the data and uncertainty if there is enough of it. Finna's collections don't seem exhaustive when it comes to photos, so it can be that the data doesn't really tell who has been visiting or where, but more about where did they take photos of visiting politicians at the time and which archives have put their data into Finna. It is also probable that many politicians visited different towns, but that alone was not necessary enough to cause someone to take pictures. In other words, the photos may also tell more about where important events were hosted, not exhaustively where the politicians visited.