

GIS

Robert Xu 1100610

#This assignment follows FAIR (findable, accessible, interoperable, reusable) guiding principles.

#This assignment (include this file) has been uploaded on GitHub with a readme part.

Link:<https://github.com/Seoyangsam/QGIS>

- Status quo: The FAIRness level of the data in the use case is quite high since it's already on GitHub, other users can find it on Github and use all the data by their own. I didn't put any license since it's completely open for anyone for any use.
- Objective in FAIRness: The best way to improve in data FAIRness for the considered use case is to write more detailed readme part to make it more interoperable and reusable; in addition, make the files more accessible.
- Roadblocks: 1. For the files with suffix 'qgz', you need first download the file from GitHub and then use QGIS software to open it, so if you don't have the software, you need download the software at first then you can check the file. The solution could be to make the file become more accessible such as convert to another format such as json that you can directly open on GitHub.

2. The readme part is not detailed enough, other people may get confused somehow that how should they reuse this usecase. Thus, a more detailed readme part is needed.
- Recommended actions (list both technical and non-technical actions): What should be done to the data in the use case to reach the desirable objective in FAIRness is to check other people's readme part on GitHub to see how can I make my own part as well. What was missing, and how should I add it into mine. Also, figure out how to convert the format of some files, ask in community such as stackoverflow.
- Success criteria: How can it be ensured/checked that the recommended actions are successful? I would like to ask other users to give feedback if the usecase's FAIRness has been improved or not.

Question: open the attribute table and check what information is inside. How does it compare to the information on the data.overheid.nl site?

fid	148
gemeenteco	GM0489
jrstatcode	2019GM0489
gemeentena	Barendrecht
water	NEE
omgevingsa	1666
stedelijkh	2
bevolkings	2465
aantal_inw	48673
mannen	23999
vrouwen	24674
aantal_hui	19259
oppervlakt	2173
COVID-19_aantallen_gemeente_03042022_Total_reported	22722
COVID-19_aantallen_gemeente_03042022_Deceased	51
infected	
normalization	46.68296591539457

There are many attributes, compared to the information(csv file downloaded) on the data.overheid.nl site, the attributes are very different

Version	Date_of_report	Date_of_publication	Municipality_code	Municipality_name	Province	Total_reported	Hospital_admission	Deceased
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0003;	Appingedam;	Groningen;	0;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0010;	Delfzijl;	Groningen;	0;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0014;	Groningen;	Groningen;	3;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0024;	Loppersum;	Groningen;	0;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0034;	Almere;	Flevoland;	1;	1;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0037;	Stadskanaal;	Groningen;	0;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0047;	Veendam;	Groningen;	0;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0050;	Zeewolde;	Flevoland;	1;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0059;	Achtkarspelen;	Friesland;	0;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0060;	Ameland;	Friesland;	0;	0;	0
6;	2020-03-13 10:00:00;	2020-03-13 10:00:00;	GM0072;	Harlingen;	Friesland;	0;	0;	0

Question: What would you think is a proper way to normalize the data such that municipalities can be compared?

We can add a new column which the value is the number of infected people in a municipality divided by the population in that municipality.

Question: what does the function `to_int("layer name")` mean?

It converts the value into integer type.

Exercise 2

Question: Why is it important that the code is unique in the municipality dataset? What could happen if it is not?

It's important that we can identify each municipality with its own code, it's just like people's Id, it should be unique. If it's not unique, then it make things become harder to distinguish each municipalities.

Question: You see for some municipalities a value of *NULL* in the column containing the overweight percentages. What does this mean and how can this happen?

It's missing data, this could happen when you collecting data, or you lost your data when you save them in a place, or during converting between files.

Question: What is this command doing?

Takes an input vector layer and creates a new vector layer with additional fields in its attribute table. The additional attributes and their values are taken from a second vector layer. Features are joined by finding the closest features from each layer.

Question: what does Id and Id_2, and the distance refer to?

Id refers to the number of deaths, and Id_2 refers to the pump, while the distance refers to the distance the location of the registered cholera victims and the closest.

Questions: What does it mean what you see now? Why are not all pumps (Id_2) present?

I see the color of Id was changed, categorized into different group with different colors.

Because for some pumps, there's no victims associated with, therefore those pumps were not presented.

Categorized

Value: ⌵ ⌵


Symbol: ● ⌵

Color ramp: ⌵

Symbol	Value	Legend
<input checked="" type="checkbox"/> ●	1	1
<input checked="" type="checkbox"/> ●	2	2
<input checked="" type="checkbox"/> ●	3	3
<input checked="" type="checkbox"/> ●	5	5
<input checked="" type="checkbox"/> ●	6	6
<input checked="" type="checkbox"/> ●	8	8
<input checked="" type="checkbox"/> ●	all other...	

Advanced ⌵

Questions: what is the id of the pump near the locations with the most cholera outbreak? And in which street is it located?

 Statistics by category — Features Total: 6,

	fid_2	count	
1	1	171	
2	2	36	
3	8	9	
4	6	13	
5	5	19	
6	3	2	

No.1; Broad Street

Extra:

Question: do you see that some wells have NULL values? Which are they and why?

Joined layer — Features Total: 8, Filtered: 8, Selected: 0

	fid	ld	ld_2	count	unique	min	max	range	sum	mean
1	1	1	1	171	7	1	15	14	340	1.988304093567...
2	2	2	2	36	5	1	8	7	65	1.805555555555...
3	3	3	3	2	2	2	3	1	5	2.5
4	4	4	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
5	5	5	5	19	4	1	8	7	40	2.105263157894...
6	6	6	6	13	3	1	4	3	27	2.076923076923...
7	7	7	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
8	8	8	8	9	2	1	4	3	12	1.333333333333...

No.4&7

Because there's no victims nearby, so no values were associated with them

