Estonian drinking water analysis

Mihkel Vaino, Kätlin Protsin

Idea

The idea behind this project was born from this year's events happening in Estonia when in May 2023 there were news that the drinking water in Kuressaare had been contaminated.

Although the quality of drinking water in Estonia has generally been quite high thanks to regulations and frameworks, the described event shows that keeping an eye on the quality of drinking water is important to ensure safe and healthy water to all Estonians.

Objectives

The goals in this project were:

- Overall analysis of drinking water in general and within counties
- Which are the main indicators which cause the drinking water quality to turn bad
- Trends over time
- How is the quality of water in correlation with other environmental factors

Data

The main datasets for drinking water analysis are taken from Terviseamet's open data source. The data was in xml format which needed parsing and a lot of cleaning. This part was done mostly using Power BI Power Query as it included automatic xml parser and the tool is very visual which helped to see what else needs cleaning in the data.

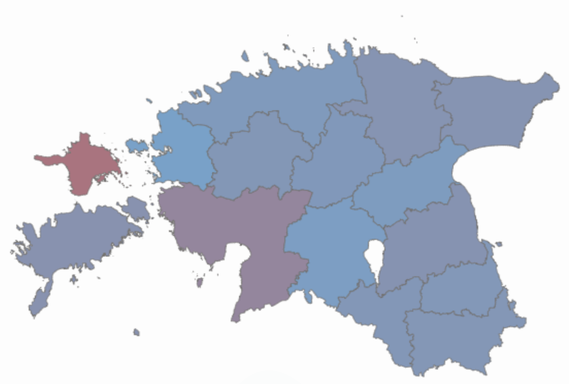
In addition, a dataset (in csv format) from Statistikaamet was used which contains the information about pollution load to surface water bodies with discharged wastewater.

Analysis and results

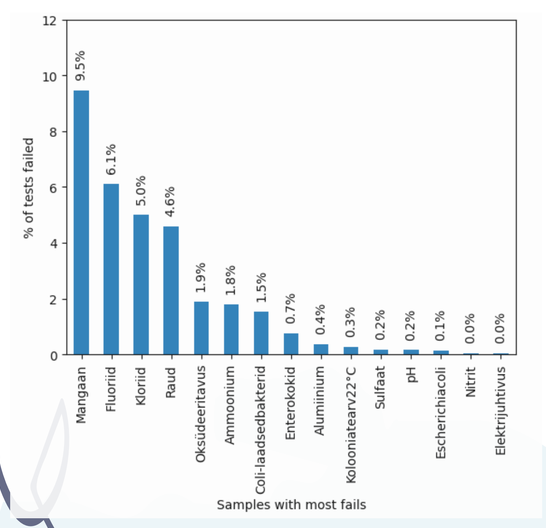
Overall, it can be seen that the quality of drinking water in Estonia is quite high, less than 1% of all the tests made for different indicators over time showed failure with requirements. Looking at the percentage of tests where least one of the samples failed the average over Estonia is 4.6% within the past 2 years.

Lowest percentage can be seen in Hiiumaa where there has been issues with high iron concentration before. This is most likely caused by iron-rich soil there. Although there was a distinguishable issue with drinking water in Saaremaa, the percentage of test is not actually much above average.

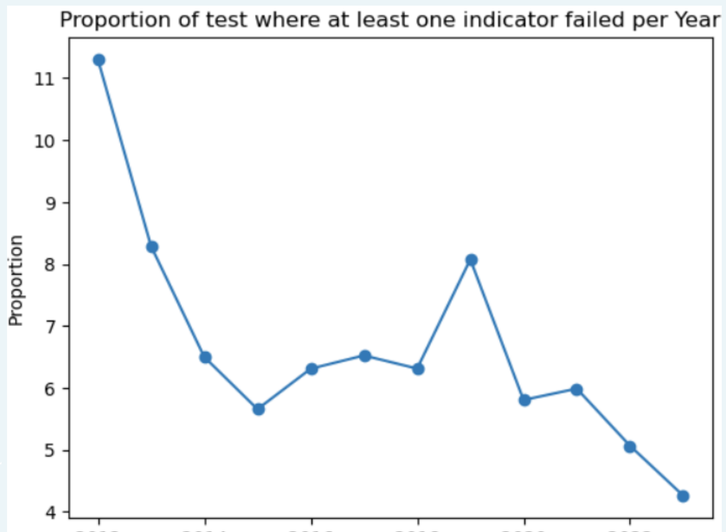
% of tests which didn't correspond to requirements by county from 2022



County	Percentage
Harju maakond	3.3%
Hiiu maakond	18.0%
Ida-Viru maakond	6.0%
Järva maakond	4.0%
Jõgeva maakond	2.9%
Lääne maakond	0.7%
Lääne-Viru maakond	6.0%
Pärnu maakond	10.9%
Põlva maakond	4.4%
Rapla maakond	4.8%
Saare maakond	6.7%
Tartu maakond	6.1%
Valga maakond	4.7%
Viljandi maakond	2.0%
Võru maakond	4.9%
Total	4.6%



One of the hypothesis made before starting the analysis was that the main indicator which causes failed samples is the coli bacteria. After analysing the percentage of failed samples it can be seen that the main indicators which fail are manganese (9.5%), fluoride (6.1%), chloride (5%) and iron (4.6%). Only approx. 1.5% of samples failed due to coli bacteria.



Overall yearly trend in Estonia is showing that the quality of drinking water has been improving over time. This is based on how many tests had at least one metric which didn't correspond to requirements compared with all tests made during that year. Using Mann-Kendall Trend test it was concluded that the trend is statistically significant and decreasing.

Similar test was conducted also within all counties and out of 15 Estonian counties 7 showed statistical significance that the number of tests where at least one metric has failed has improved over the period of 12 years (those were Ida-Viru, Järva, Jõgeva, Lääne-Viru, Pärnu, Põlva and Valga)

Correlation matrix between Phosporous in wastewater and top 4 quality detractors

	Fluoriid	Kloriid	Mangaan	Raud
Harju maakond	0.66	0.34	0.83	0.56
Ida-Viru maakond	nan	0.42	0.34	0.47
Järva maakond	-0.1	nan	0.29	0.78
Jõgeva maakond	0.44	nan	0.65	0.78
Lääne-Viru maakond	nan	-0.1	-0.2	0.14
Pärnu maakond	0.78	-0.09	0.78	0.65
Rapla maakond	0.64	nan	0.72	0.88
Saare maakond	-0.04	-0.04	0.45	0.01
Tartu maakond	0.04	nan	0.21	0.46
Viljandi maakond	0.5	nan	0.34	0.25
Võru maakond	nan	nan	-0.01	-0.05

Correlation matrix between Nitrogen in wastewater and top 4 quality detractors

	Fluoriid	Kloriid	Mangaan	Raud
Harju maakond	0.47	-0.33	0.25	0.3
Hiiu maakond	0.53	0.17	-0.43	-0.28
Ida-Viru maakond	nan	0.43	0.31	0.35
Järva maakond	-0.41	nan	0.37	-0.59
Jõgeva maakond	0.69	nan	0.4	0.36
Lääne maakond	0.07	-0.12	0.1	-0.01
Lääne-Viru maakond	nan	-0.08	-0.56	-0.07
Pärnu maakond	-0.54	0.61	-0.51	-0.78
Põlva maakond	nan	nan	0.38	0.28
Rapla maakond	-0.44	nan	-0.57	-0.71
Saare maakond	-0.12	-0.13	0.2	-0.2
Tartu maakond	0.16	nan	0.09	0.11
Valga maakond	nan	0.12	0.47	0.41
Viljandi maakond	-0.24	nan	-0.05	-0.5
Võru maakond	nan	nan	0.38	0.24

Using the data about pollution load to surface water bodies with discharged wastewater it was investigated if there is a statistically significant correlation between the phosphorous and nitrogen levels with the top 4 indicators which caused failed test results – manganese, fluoride, chloride and iron.

Using Pearson correlation coefficient in can be seen that there are some counties where the correlation between the values is statistically significant. Phosphorous from households and economic sectors discharged into aquatic ecosystems with waste water has more statistically significant counties than nitrogen.