Exercise 1

Problem 1

Task 1.1

Image of **population density using postal code areas** with five color-coded ranges is shown in Fig 1.

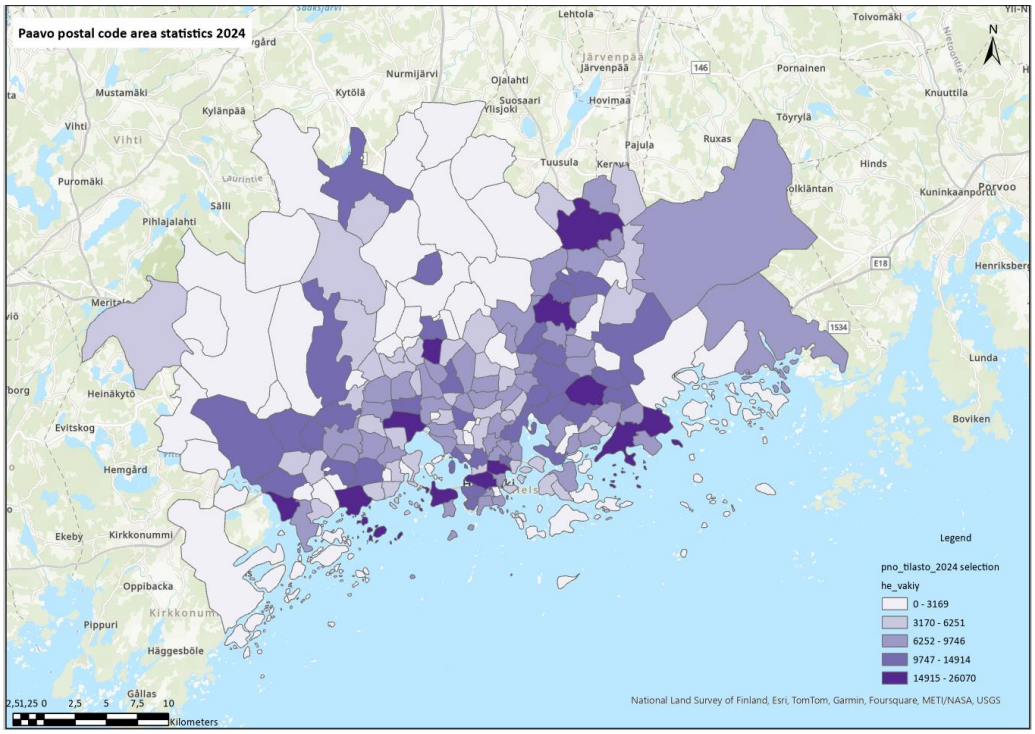

Fig 1. Population density using the postal code areas

Image of **population density using the population grid** with five color-coded ranges is shown in Fig 2:

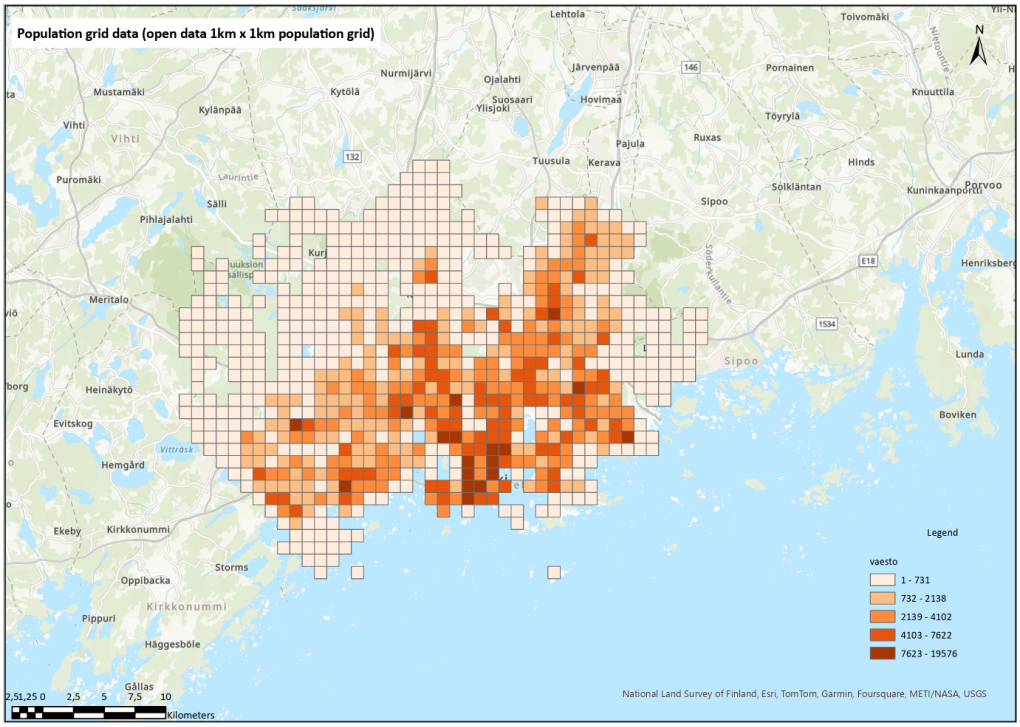

Fig 2. Population density using the population grid

Image of the map legend for both the postal code and population grid representations of the population density were as follow, the method for symbolizing the legends in both maps are the same, which is `**Natural Break**`:

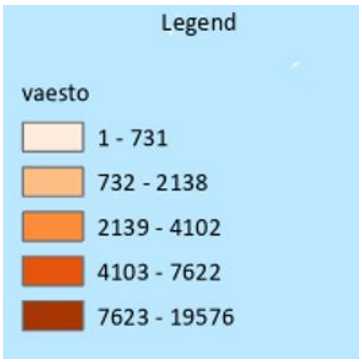

Fig 3. Legend for population density using postal code areas (left)

Fig 4. Legend for population density using the population grid (right)

Task 1.2

Question 1.1: How have you ensured that the representations of population density in the two maps are comparable?

Answer: To make the two maps comparable, both maps need to use the method for symbolizing the legends to represent population density. Additionally, the population density is categorized into five groups in both maps, ensuring the visual representation aligns in scale and range. While the postal code map represents population density based on administrative boundaries, the population grid uses fixed spatial units, allowing a direct comparison of how population is distributed in both types of representations.

Question 1.2: What are the most important similarities and differences between the representation of population density on the two maps?

Answer:

Similarities:

- Both maps show higher population density concentrated in the southern and central regions, particularly around Helsinki and nearby urban areas. This indicates that both maps are aligned in reflecting that these areas have a higher population density.
- Both maps use different shades to represent population density, with darker colors indicating higher population concentrations and lighter colors representing lower densities.

Differences:

- The Paavo postal code area statistics map (Fig 1) uses postal code areas to depict population distribution, which results in irregularly shaped polygons, as the boundaries follow postal code divisions. These areas vary in size and shape depending on the postal code.
- The population grid data map (Fig 2) uses a 1km x 1km grid, meaning population density is displayed uniformly in square blocks, providing a more even, grid-based distribution of data across the region.
- The postal code map aggregates data over larger, often uneven areas, potentially smoothing out variations within a postal code region. The population grid map provides a finer resolution by breaking the area into smaller grid cells, potentially offering more detailed insights into population distribution within smaller geographic spaces.

Question 1.3: Is this a good example of the Modifiable Areal Unit Problem (MAUP)? If so, why?

Answer: Yes, this is a clear example of the Modifiable Areal Unit Problem (MAUP). The way population density is represented shifts noticeably depending on the spatial units used—whether they're postal code areas or grid cells. MAUP arises because different aggregation units (such as postal codes versus uniform grid cells) can result in different interpretations of population patterns. This emphasizes how the choice of spatial units can impact both statistical outcomes and visual representations.

Problem 2

Task 2.1

Image of **Comparison of HSY data with HKI data** is shown in Fig 5.

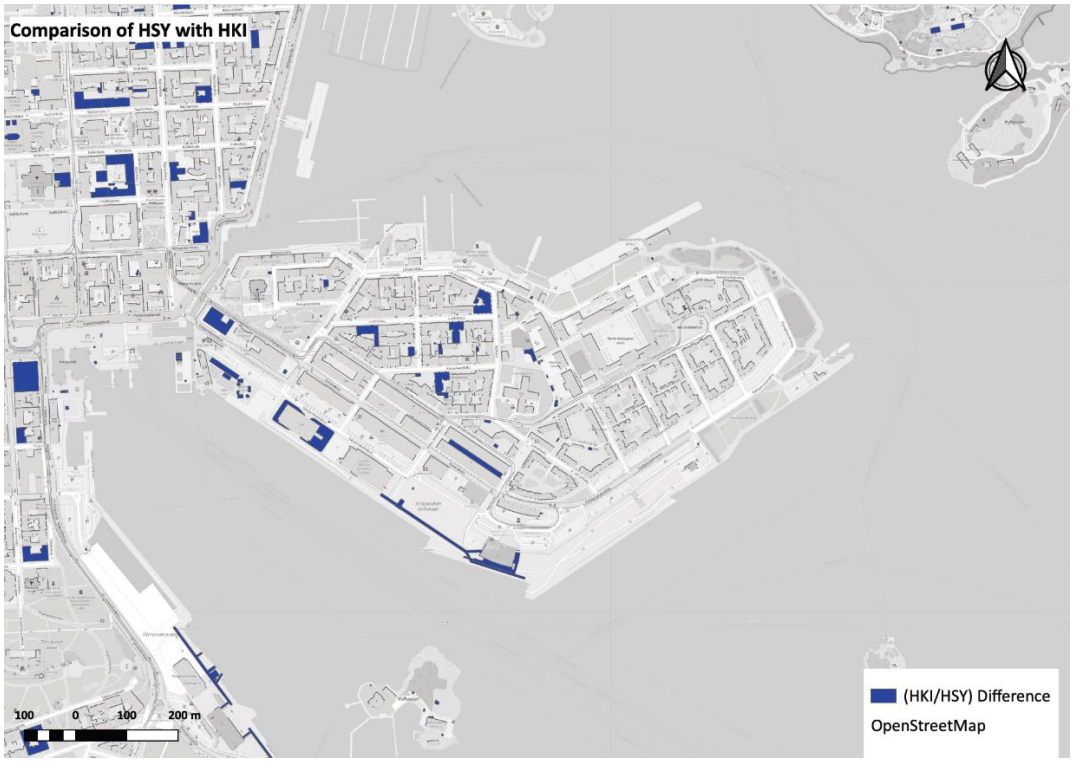

Fig 5. Comparison of HSY with HKI

Image of **Comparison of OSM data with HKI data** is shown in Fig 6.

Fig 6. Comparison of OSM with HKI

Task 2.2

Question 2.1: Assessment of HSY data when compared to HKI data

Answer: When comparing HSY data with HKI data, several discrepancies are noticeable, highlighted by the blue areas on the map (Fig 5), representing the differences between the two datasets.

Spatial Accuracy: Overall, HSY data aligns well with HKI data, with most buildings positioned correctly. However, there are some notable differences in spatial placement or the shapes of structures in certain areas. In some cases, buildings or features present in the HKI data are either missing or misaligned in the HSY dataset, affecting spatial accuracy.

Completeness: There are gaps in the HSY dataset, as some buildings present in the HKI data are either missing or incomplete in the HSY dataset. This lack of completeness is particularly evident in the areas marked in blue. The gaps suggest that some urban elements are either outdated in the HSY data or have not been fully recorded.

Major Issues with Dataset Quality:

1. Missing Buildings or Features: Some key structures present in the HKI dataset are not represented in HSY, leading to incompleteness.
2. Spatial Misalignment: Certain buildings are misaligned or shaped differently compared to HKI, indicating issues with spatial accuracy.

Question 2.2: Assessment of OSM data when compared to HKI data

Answer: When comparing OSM data to HKI data, the differences are highlighted in orange (Fig 6), reflecting variations in accuracy and completeness between the two datasets.

Spatial Accuracy: OSM generally exhibits high spatial accuracy, with most buildings and features correctly positioned relative to the HKI data. However, like HSY, some areas display minor discrepancies in building geometries or placement, suggesting slight spatial inaccuracies.

Completeness: While OSM appears more complete than HSY, it still has some missing or incomplete buildings. These discrepancies are shown in the orange-marked areas, where certain structures are either absent or inaccurately represented.

Major Issues with Dataset Quality:

1. Missing Features: Despite being more complete than HSY, OSM still lacks certain buildings or urban elements.
2. Outdated Information: Some data may be outdated, as OSM's open-source nature might lead to slower updates compared to the official HKI dataset.

Question 2.3: Comparison of OSM and HSY Data Quality

Answer: OSM data generally shows higher quality than HSY. The main reason is that OSM has fewer discrepancies when compared to HKI, indicating better spatial accuracy and more completeness. Although both datasets exhibit some missing or misaligned buildings, HSY shows more significant gaps and spatial inaccuracies. OSM's open-source nature may allow for more frequent updates, contributing to its better overall quality.

Question 2.4: HKI as a Baseline for Data Quality Assessment

Answer: It is reasonable to use HKI as a baseline for external data quality assessment. The HKI dataset likely represents official, authoritative geographic information for the region, which provides a reliable reference point. HKI data can be assumed to have undergone rigorous validation processes, ensuring its accuracy and completeness. Given this, it serves as a dependable standard for assessing external datasets like HSY and OSM.

Problem 3

For this exercise, I spent approximately 4 to 5 hours in total, including time spent familiarizing myself with ArcGIS and writing the report. Since my laptop failed to start the day before the submission, I had to redo Exercise 1 on a school computer. However, the time spent redoing the exercise is not included in my time estimate. By completing the exercise a second time, I found that my understanding of the data and my proficiency in using ArcGIS improved significantly, which is a positive outcome.

Additionally, while working on Problem 2, I noticed that after adding the HKI, HSY, and OSM datasets into ArcGIS, there were noticeable gaps in the building maps. Also, when I tried to fix these issues (during the FIXED process), the software frequently froze or encountered errors. As a result, I chose to complete this part using QGIS instead, and I did not face the same issues with QGIS. Therefore, I feel that QGIS may have certain advantages over ArcGIS for specific tasks.

Here are some of my suggestions: I believe it would be helpful to include more basic tutorials on software operations during the exercise sessions. This would make it easier for students who have no prior experience with these tools to get started.