



Visualizing and comparing population projection rasters

Author:
Andreas Gram Riisgaard

Supervisor:
Carsten Keßler

Titel:	Abstract:
Visualizing and comparing population projection rasters	In this project an interactive tool for visual comparison of raster datasets have been developed using population projections as a case.
Projekt:	To develop a tool able to enable such comparisons it is important to know how population projections should be visualized and which functionalities are important for the tool. There is also a technical challenge in visualizing large raster datasets, while still maintaining a responsive user experience. The conventions for visualization of population projections was explored through a literature review. A quantitative sequential dataset like a population projection should be colored, so that the areas with least population is colored in a lighter color, than the denser areas.
Project Period:	The functionalities for the tool was determined by comparing with another interactive map. It was decided to have two maps showing different population projections. The map can be navigated either by panning and zooming or using a search bar.
Author:	To ensure a responsive user experience the raster was not loaded into the tool in its entirety. Instead it was divided into smaller tiles, which got loaded based on the extent of the map. These tiles were then colored on the client.
Supervisor:	The tool was created as an Openlayers map displaying tiles, which was created with a modified version of the python program gdal2tiles. While the user experience is responsive while using the map the creation of tiles is time consuming. The tool could therefore be improved in the feature by using cloud optimized geotiffs instead of tiles.
Number of pages: 78	
Number of annexes: 9	
Afsluttet: 4-6-2020	

Preface

This is edited in the file Formalia/Preface

Contents

Preface	iii
1 Introduction	1
1.1 Problem statement	1
1.2 Limitations	2
1.3 Target audience	2
1.4 Report structure	2
I Design	5
2 Case data	7
2.1 The data behind the simulation	7
2.2 The simulation	7
2.3 Shared Socioeconomic Pathways	8
2.4 Technical details about the case data	9
3 Raster formatting	11
3.1 Bit depth	11
3.2 Tiled raster	11
4 Current methods of comparing rasters	15
4.1 Desktop Gis program	15
4.2 Interactive visualization tool for population simulations	16
4.3 MakeCityWebsite	17
5 Visual conventions	19
5.1 Color	19
5.2 Map projection	21
6 Related work	23
6.1 Tilebased raster visualisation	23
7 Components of an interactive maps	27
7.1 Relevant functionalities	28
7.2 Comparing layers	29
8 Final design	31
8.1 Technical concepts	31
8.2 Visualisation	35
8.3 Functionalities	35
8.4 Implemented design	35

8.5 Number of data classes	36
II Development	37
9 Coding languages and plugins	41
9.1 Languages	41
9.2 Libraries	42
9.3 Testing server	44
10 Developing the tool	45
10.1 Creating the tiles	45
10.2 Processing time	48
10.3 Visualizing tiles	49
10.4 Loading data at a wrong resolution	49
10.5 Calculate max value in current extent	51
10.6 Recolor when the max value change	54
10.7 Polish	54
11 Result	57
III Evaluation	59
12 Evaluation Methods	61
12.1 Setup	61
12.2 Overview	61
12.3 Metrics for performance	62
12.4 Calculating score	63
12.5 Additional measurements	64
13 Evaluation	65
13.1 Postload performance	69
14 Discussion	71
14.1 Loading tiles from the wrong zoom level	71
14.2 Further development	71
14.3 Optimizing the tile generation	72
14.4 Cloud Optimized Geotiff	72
14.5 Performance enhancements	73
14.6 Not calculating minimum values in tiles	74
15 Conclusion	77
Bibliography	79
A This is at test	85

Introduction

1

In 2019 the population in the world reached 7.7 billion people, which is an increase of one billion over the past twelve years. According to The United Nations Department of Economic and Social Affairs' (UN DESA) median scenario the growth is expected to continue reaching 9.7 billion in 2050. [United Nations Department of Economic and Social Affairs, 2019b]

To be able to adapt infrastructures to this population growth it is necessary to predict where these people will settle. While UN DESA provides this information on a national level [United Nations Department of Economic and Social Affairs, 2019a], it is more ideal with a more nuanced picture, since most planning are based on local or regional scale spatial projections. [Jones and O'Neill, 2016]

Other researchers (SEDAC, CISC) have used simulations to distribute the population within each country as raster layers. However due to the high resolution and/or small scales, visually comparing these raster datasets is a time-consuming task. The purpose of this project is to create a tool allowing fast and easy comparison of such raster datasets, focusing on the use case of population projections.

1.1 Problem statement

To explore the possibilities for creating such a comparison tool the following research question have been defined:

How can population rasters be visualized and compared efficiently and effectively?

This broad main question will be answered by answering the following three subquestions:

Which conventions exist for visualization of population projections?

Which functionalities are relevant for comparing different rasters?

How can a responsive user experience be ensured, when loading and visualizing large raster dataset?

1.2 Limitations

Determining relevant functionalities and the responsiveness would ideally have been done with user testing. However it was determined that both the creation of the tool and a scientific approach to user testing would require too much time. Therefore the tool creation got prioritised and other evaluation methods were chosen. This is expanded upon in section 13

1.3 Target audience

The target audience for this project is academical researchers. It is meant as a tool for these researcher to be able to quickly compare different population projections.

This prototype of the tool have only been created for internal and individual use. It will therefore not be created with multiple users in mind and there will be no considerations for security.

The tool is also created with only computers in mind, so it will not be optimized for mobile.

Alternative target audiences are being discussed in section x.

1.4 Report structure

1

The report has been divided into three parts; design, development and evaluation.

In the design part the thought process behind the design of the solution is explained. It starts with some background information about the case data in chapter x, raster formatting in chapter x and chapter x about how the raster data currently is being compared.²

¹FiXme Note: ADD: quick overview of what the solution is going to be, what is population projections, SSP

²FiXme Error: Add these three to figure -case,raster,current

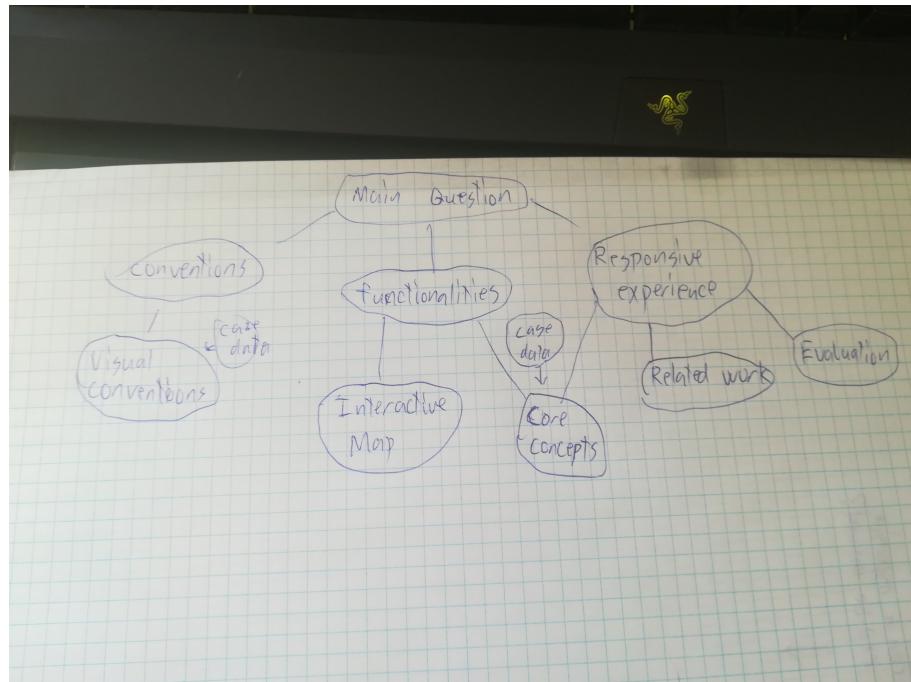


Figure 1.1: Overview of how the research question gets answered in the first part of the report

The rest of the first part is focused on answering the subquestions as illustrated in figure 1.1. The first subquestion is answered through a literature review in chapter x. After the visual conventions have been explained an example of an interactive map gets analysed in chapter x. This is done to understand which functionalities are important for such a map. This is followed by a review of related work in chapter x. Based on the related work and an initial exploration of the data five core concepts for the tool are created. These are described in chapter x. All of the considerations presented in this part then gets collected into the design in chapter x.

The second part is the development of the tool. In chapter x the building blocks, which the tool is build from, are presented. How these building blocks are put together is then explained in chapter x. Chapter x is an overview of the final product.

The last part starts with an evaluation of the developed tool in chapter x. This is followed by chapter x with a discussion of the tool and how it could be developed further. The last chapter is the conclusion in chapter x

³Fixme Note: Update structure and create figure

Part I

Design

Case data 2

The population projection visualized in this project have been created by a geosimulation, which geographically distribute the predicted population. This geosimulation have been run every tenth year from 2010 to 2100. It has also been run for different scenarios for the feature.

2.1 The data behind the simulation

¹ This projection is a geographical distribution of the global prospects created by The United Nations Department of Economic and Social Affairs. These prospects estimate the population numbers in each country living in rural and urban areas. The spatial distribution has been based on raster data from the Global Rural Urban Mapping Project (GRUMP). In this dataset the world has been divided into 1x1 km cell, each of which contains the number of estimated people. GRUMP also have a dataset for the extent of urban areas. This urban extent has been estimated based on the light emitted after dark, which as a result means that it overestimates the extent of urban areas. To compensate for this the dataset was combined with the Global Land Cover Map from the European Space Agency (GlobCover). This dataset more accurately defines the extent of urban areas. Only the areas which qualified as urban in GRUMP dataset and were classified as “artificial surfaces and associated urban areas” in the GlobCover data were treated as urban areas.

2.2 The simulation

Using this dataset to define the extent of urban, suburban, and rural areas, the simulation was run as illustrated in figure 2.1

¹Fixme Note: Rewrite this - data is not a mix of GRUMP and GlobCover, but seperated.

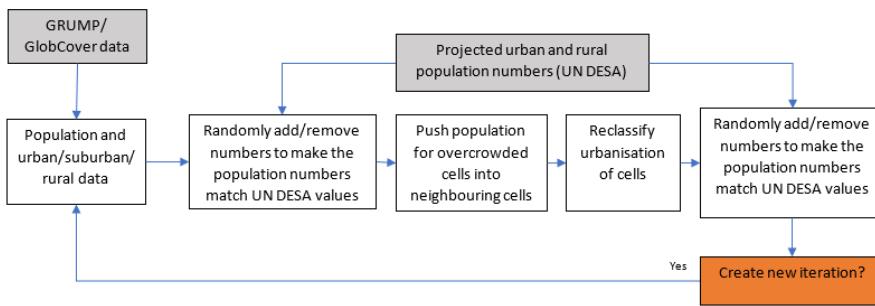


Figure 2.1: A flowdiagram of simulation creating the population projections

The population in rural and urban areas were adjusted to match the projected values from UN DESA. This was done by randomly adding or removing people from the area types.

The cells would have a country-specific limit on how many people could live in them. Any cells with a population above this limit would push the excess population to neighboring cells.

After this, the cells would be reclassified as more urbanized if their population had increased above a country-determined threshold. This way a rural cell might become a suburban one or a suburban become urban if its population increased. The data is then adjusted to match projected values again with the same approach as before. This is required because the reclassification would mean that the number for the different urbanization classes would no longer match the numbers from UN DESA.

These steps are then repeated for every iteration.

[Keßler and Marcotullio, 2017]

2.3 Shared Socioeconomic Pathways

2

The geosimulation have been run for five different future scenarios called the Shared Socioeconomic Pathways (SSPs). The different scenarios are based on different ways, climate change could affect society. As illustrated in figure x the five scenarios are based on two axes: socio-economic challenges for mitigation and for adaptation of climate change impacts. “Socio-economic” is referring to a large array of societal and socioecological system. Covering aspects of a political, social, demographic, cultural, lifestyle, institutional, economic, and technological nature and condition of ecosystems. The horizontal axis is environmental or societal conditions, which would make adapting to climate change more difficult. This include climate hazards (temperature and precipitation changes, sea level rise and extreme weather phenoms); what and whom these hazards will affect. [O’Neill et al., 2013] The vertical axis is factors, which in absence of climate policy would increase that amount of emissions of greenhouse gasses and factors reducing the society’s capacity

²FiXme Note: Create own figure

to lessen those emissions. Examples of factors reducing society's capacity is inadequate technologies and insufficient resources to support mitigation policies. The increase in greenhouse gasses could for example be a result of population and economic growth. [Ebi et al., 2014]

	SSP1 Sustainability	SSP2 Middle of the road	SSP3 Regional rivalry	SSP4 Inequality	SSP5 Fossil-fueled Development
Population Growth					
High fertility	Low	Medium	High	High	Low
Other low fertility	Low	Medium	High	Medium low	Low
Rich low fertility	Medium	Medium	Low	Medium low	High
Urbanization level					
High income	Fast	Central	Slow	Central	Fast
Medium income	Fast	Central	Slow	Fast	Fast
Low income	Fast	Central	Slow	Fast	Fast

Table 2.1: Population growth and urbanization for each of the SSPs

How the population growth and the urbanization levels will be in the different scenarios can be seen in table 2.1. This table have been created by Jones and O'Neill [2016]. The population growth values are based on current income and fertility condition. The urbanization predictions are based on the current income.

3 4

2.4 Technical details about the case data

Table 2.2 is highlighting some technical specifications of the dataset, which will be important for the development. The technical details have been acquired using GDAL's raster information program, gdalinfo. GDAL is further described in section x.

Technical information		
Attribute	Value	Further explanation
NoData value	-2147483648	
Bit depth	32	Section ??
Projection	EPSG: 4326	Section ??
Projection unit	Degree	Section ??

Table 2.2: Technical attributes of the case data

The majority of these concepts will be explained in upcoming sections. Only the NoData value will be explained here. A raster file has values for every pixel. If no data is available for

³FiXme Error: Need more context to the table - what does it mean? Do I need it? - use it to select what to show in results

⁴FiXme Error: Write about domain sharding being unneeded

a pixel, then the pixel get assigned the NoData value. <https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/performing-analysis/nodata-and-how-it-affects-analysis.htm>

If the value was set to 0 it would be impossible to determine if an area had nobody living in it or if no data was available. That is the reason for the NoData value is set to a negative value, since the population density never can be negative.

5

⁵Fixme Note: what is the noData in this case - lakes?

Raster formatting 3

Visualizing raster data requires an understanding of how it works from a technical perspective.

3.1 Bit depth

Common raster formats include jpeg, png and tiff. One of the key differences between these formats are the bit depth. This value determines how many different colors that can be assigned each pixel in the image. Each bit can only be assigned one of two values, 0 or 1. This number of available colors for a pixel can therefore be calculated as $2^{\text{number of bits}}$. As an example, a bit depth of 8 bits would then result in $2^8 = 256$ different potential colors, since is the number of combinations of ones and zeroes, that are possible. For most maps this amount of colors is enough. The human eye can comprehend around 10 million different colors, but often that many colors are not needed. When displaying aerial photographs, the depth of 24 bits is often used, since it can display 16.7 million different colors and therefore can appear in true-colors to humans. The limitations of the human eye does not mean that having more than 10 million bit combination is pointless. The bit combination can also be used for other thing than colors. It can be used to create transparency values or to store metadata about the image file. For instance, the geotiff format can use the additional bit to store georeferencing information. These advantages of higher bit count come at a cost of a larger file size. Having a larger color depth will result in a slower loading and larger requirements for storage space. [Dent et al., 2009]

3.2 Tiled raster

This section describes standards for dividing large raster files into smaller tiles. A way of only visualising the necessary parts of a raster is to divide it into smaller raster tiles and then only load the relevant tiles. When loaded into the map these tiles then gets places next to eachother, so they appear as a single large map image. These tiles can also be created with different resolutions, so that zooming in on the map with return tiles more details. As illustrated in figure 3.1 each of these zoom layers have more tiles. [Liedman, No Date] When zoomed all the way out the entire world is rendered as a single tile. Whenever the zoom level gets increased by one each tile in the previous layer get replaced with four smaller tiles. The number of tiles on a zoom level z can therefore be calculated as 2^{2z} . [OpenStreetMap Wiki, 2020]

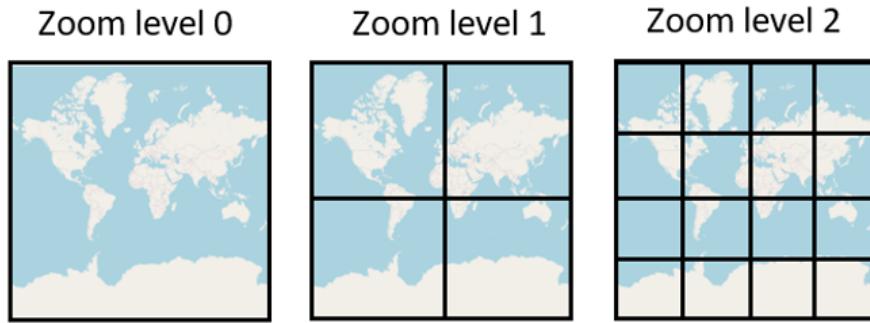


Figure 3.1: Illustration of the increase in tiles for each zoom level

Openstreetmaps is an example of a service, which use tiled rasters for their map. When this service is used, requests for tiles are send to <http://tile.openstreetmap.org/zoom/x/y.png>. In the request the values zoom, x and y are replaced with the current zoom level, tile column and tile row. [OpenStreetMap Wiki, 2020]

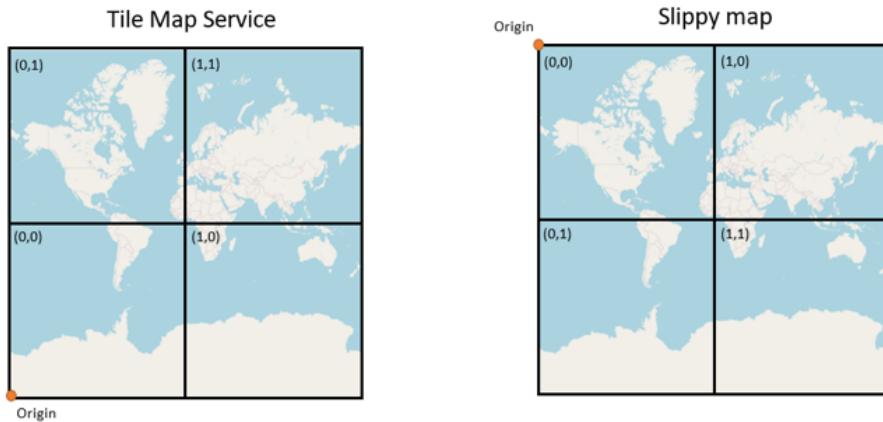


Figure 3.2: Illustration of the difference between the TMS and XYZ formats. The individual maps are inspired similar maps from Planet [No Date] and OSGeo Wiki [2012]

The naming of these tiles are done differently for different standards. In this section only the Tile Map Service (TMS) and Slippy map (XYZ) standard will be addressed. Both of these have the same approach to naming zoom level and columns, but different approach to naming the rows. This difference has been illustrated in figure 3.2. The reason for the difference is that TMS tiles number their rows from south northwards, whereas XYZ are numbering rows the reverse way. [OpenStreetMap Wiki, 2019]

Current methods of comparing rasters 4

When trying to solve an issue it is important to understand what the current methods for achieving the same is. This chapter will present three current methods for comparing rasters

4.1 Desktop Gis program

One way of visualising the population data would be to load the data into a desktop GIS programs like QGIS or ArcGIS. Both of these can visualise TIFF files and it is possible to visually compare different raster layers by changing back and fourth between the different layers. However one would have to manually style the layer to make spatial patterns become visible. Without styling the layer the tiff file will be colorized based on the most and least populated areas in the world. When coloring based on these extremes most of the world will appear in the same color as illustrated in figure 4.1. Only the most populated areas in India and China will appear in a different color. This is further expanded upon in section ¹

²

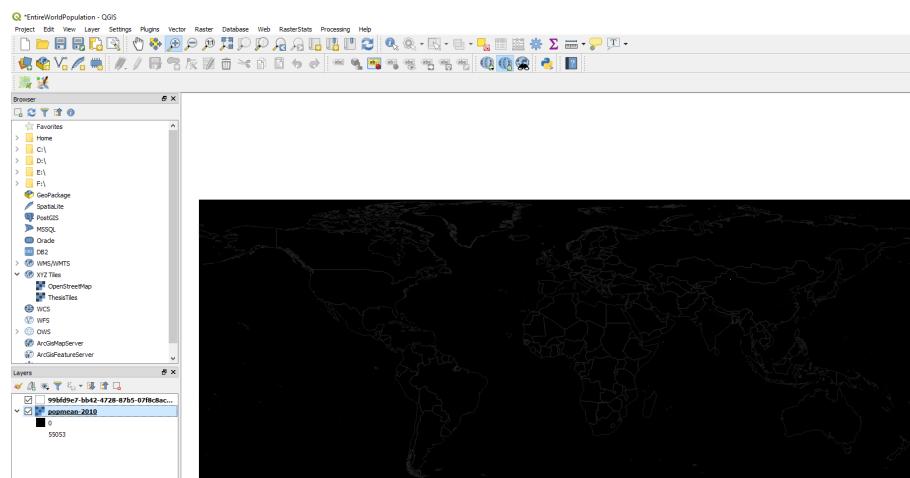


Figure 4.1: The population distribution visualised in QGIS and colored based on data from the entire world

¹FiXme Note: Ref to core concepts

²FiXme Note: Remember Carsten mail about locally, on-the-fly and subdomains

4.2 Interactive visualization tool for population simulations

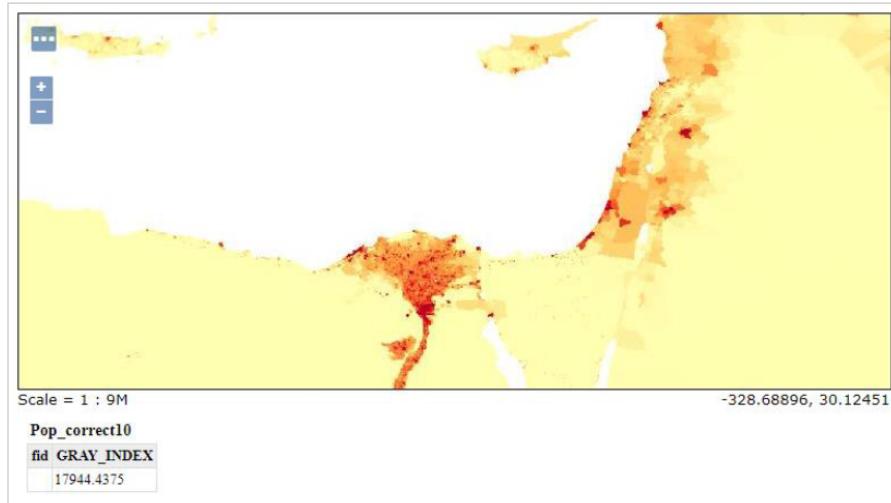


Figure 4.2: An interactive visualization tool made in GeoServer. Source: ?

? wrote a her master thesis about interactive visualisation of the same population projections, which is the case data for this project. In the thesis she tests different methods for interactive visualisation. The first tests are the python libraries Bokeh and Folium. Both of these are creating interactive webbased maps, but were unable to visualize the population projects due to the file size of the rasters. Therefore other solutions were created using servers for geographic imagery. The servers GeoServer and MapServer were used in the thesis, which both uses Openlayers for their interactive interface. Figure 4.2 shows the Geoserver visualisation of the raster data. Figure 4.3 shows the Mapserver solution. Both tools required a manual styling.



Figure 4.3: An interactive visualization tool made in MapServer. Source: ?

The conclusion in the thesis was that MapServer was the faster option, whereas GeoServer had a less steep learning curve. Both servers were slower than viewing in QGIS. Both had limited customizability. As shown in the two figures the tool were only showing a single map and not having an interface for changing the raster layers. The tools were therefore more of an visualization tool, than a comparison tool.

4.3 MakeCityWebsite

One of the creators of the case data created a python script, called MakeCityWebsite, for comparing the different population projections for a chosen city. The final product of the tool can be seen in figure x and x.

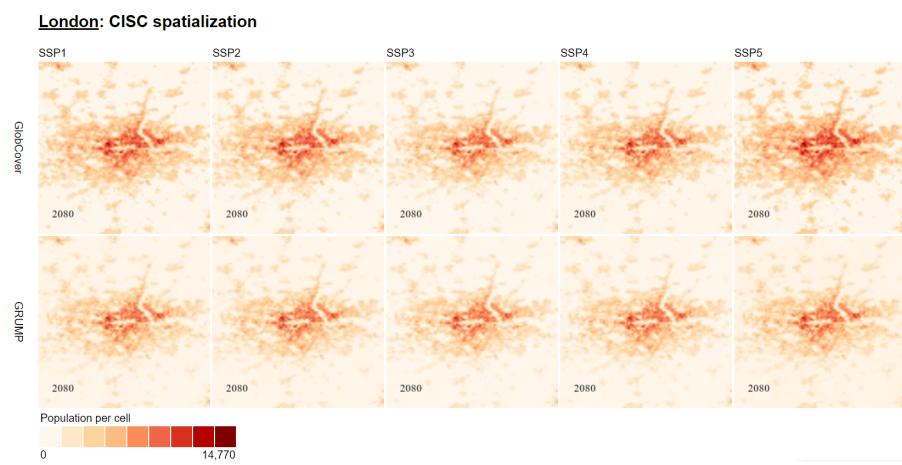


Figure 4.4: The top half of a product made by MakeCityWebsite script

The input for the tool is the rasters from each of the two urbanization models, the five SSPs and every tenth year from 2010-2100, so a total of hundred different rasters. The tool first creates new raster files, which have been cut to the extent of the chosen city. Then the minimum and maximum value in those files are calculated. The maximum and minimum values among all the rasters are then used to determine a color scale. This color scale is then used to create a new set of colored rasters.

So at this point the tool has generated one hundred rasters, which all have been colored using the same color scale. For the user to be able to comprehend this amount of data two additional steps are taken. First the individual combinations of urbanization models and SSPs are collected in gifs. These gifs are displayed how the population evolves in their scenario by chronology changing between different years. These different gifs are then placed next to each other in a website as illustrated in figure x.

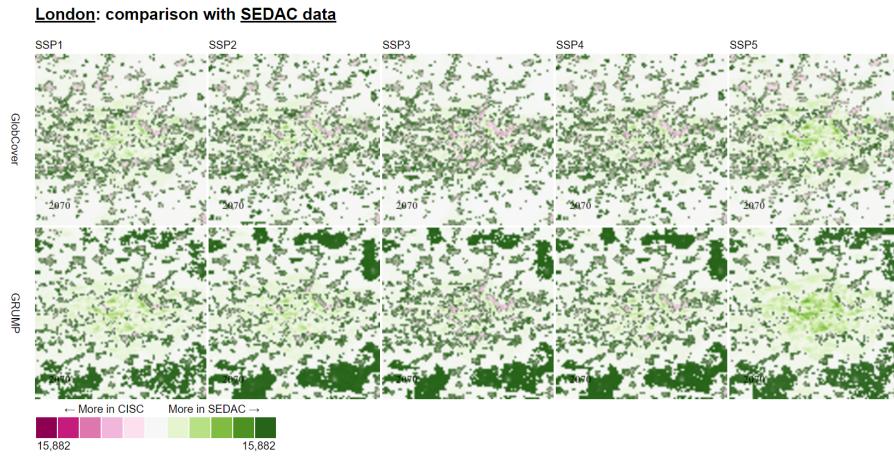


Figure 4.5: The bottom half of a product made by MakeCityWebsite script

The tool also creates a comparison with another population projection called SEDAC. This is done by going through the same steps above, but with a custom raster. This raster visualises the difference between the two projections. This comparison part of the website can be seen in figure z <https://github.com/crstn/CISC/blob/master/MakeCityWebsite.py>

The creation of these comparison rasters take 20 minutes for a single SSP <https://github.com/crstn/CISC/blob/master/MakeCityWebsite.py>

3

³Fixme Note: How long time does it take to run MakeCityWebsite

Visual conventions

5

¹ This chapter will address the visual convention connected to population projection.

Within the field of spatial convention there are multiple convention, however not all of these are relevant in this particular context. To understand which conventions are relevant for visualizing data one must understand the properties of the data to be visualized. Spatial data can be presented as either discrete-objects or continuous-fields also known as vector and raster data. [de Smith et al., 2020] These two types of data have different ways of being visualised. As mentioned in chapter ?? the data is in the raster format and is sequential. Visual conventions connected to discrete-objects will therefore not be covered.

2

5.1 Color

A color can be defined by three parameters: hue, saturation and lightness. These different concept have all been illustrated in figure 5.1.

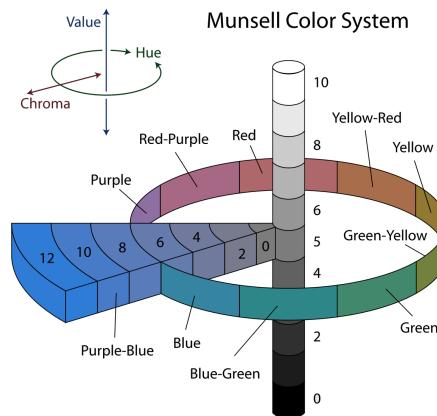


Figure 5.1: Illustration of the concepts: hue, saturation and lightness. Source: Rus [2007]

Hue

The hue is what would traditionally be referred to as colors (red, green, yellow). In the figure the hue is illustrated as the circle around the column.

¹Fixme Note: Add text about what a convention is and scale and projection

²Fixme Note: data is sequential

Saturation

The saturation is also referred to by some authors as chroma, intensity or purity. It is a measurement for how vivid the color is. A color with a low saturation would be close to the color grey. If the saturation increases more of the color pigment is added to the color, until there is no trace of grey left. The saturation has a value from 100% (fully colored) to 0% (grey). It is illustrated in the figure as the distance from the center.

Lightness

A measurement of the color's lightness or darkness. In the literature this is often referred to as the colors value, but Brewer remarks that this use of terminology is not ideal in data science since "value" also could refer to the data values. It is illustrated as the vertical axis in the figure. [Dent et al., 2009]

5.1.1 Selecting a color scheme

There are multiple elements, that should be considered, when choosing a color scheme. Naturally the different colors should be easily distinguishable, but one should also consider the user of the map and the medium used for presenting the map. The right choice of color pattern allows the user to see patterns in complex data, which otherwise would be obscured.

Cindy Brewer divides color schemes into the four categories; binary, qualitative, diverging and sequential. These categories and combinations between them have been visualized in figure 5.2. [Brewer, 1994]

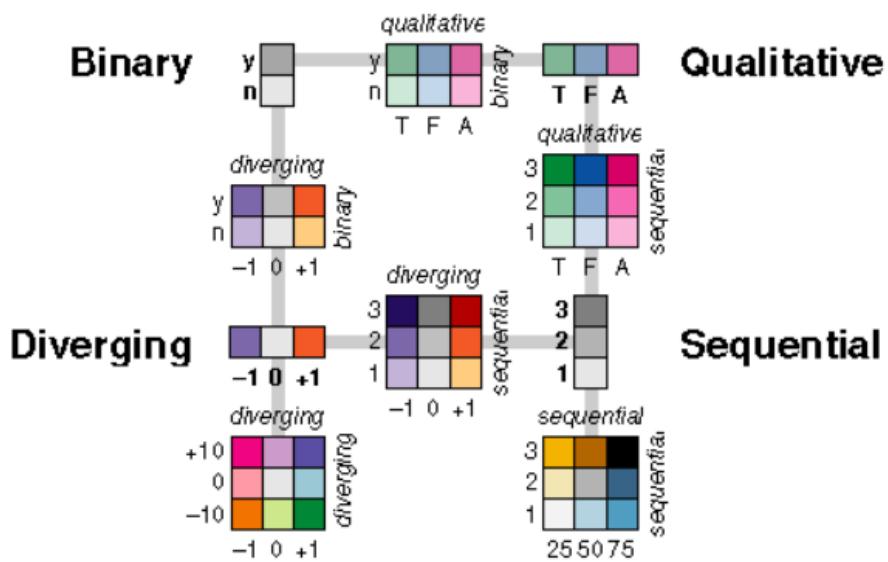


Figure 5.2: Illustration of the different categories of color schemes. Source: Brewer [No data]

Each of these categories are good for visualizing different kinds of data. The qualitative

is well suited for illustrating nominal data. This color scheme is for instance commonly used for land use maps. The binary scheme is a qualitative scheme, but with only two categories. A example of a use case for this scheme could be visualizing whether countries are members of the European Union or not. The sequential scheme is useful for illustrating the ordered data going from low values to higher ones. An example of this could be height of vegetation. The differentiation between the values are illustrated with differences in lightness. The diverging scheme is similar to having two sequential color schemes. This enables highlighting a critical value in the middle of the data. It is for example being used to illustrate temperatures, where the neutral middle value would be 0°C. [Brewer, 1994]

Based on these categories of color schemes Brewer have developed an online tool, colorbrewer2.org, which can aid mapmakers in picking a color scheme. This tool also takes into colorblindness into consideration and inform the user if a colorscheme is suitable for printing or viewing on small screens. [Brewer et al., 2013]

5.1.2 Visual conventions for colors

The conventions can be divided based on whether the visualized data is qualitative or quantitative. The qualitative datasets have many historical convention – like using green colors for vegetation or using blue for water. It is not the same case for quantitative data, where “No conventions exist for color choice on quantitative maps. (for example, population density maps are always blue, income maps are always green, and so on)” [Dent et al., 2009]. There are however convention for the choice of lightness, where "light is less - dark is more"

[Garlandini and Fabrikant, 2009]

5.2 Map projection

When creating a map, the three-dimensional earth gets transformed into a two-dimensional map, which leads to distortion. The severity of these distortions depends on the scale. For large areas, the distortion has bigger consequences compared to the smaller scale. When mapping at large scale, it is impossible to prevent distortion to some of the projection’s properties. This means that area, angles, distance, or direction on the map will be distorted. Which properties get affected depends on the choice of projection. When visualizing quantitative data an equivalent projection is often used.

Equivalent projections

In Equivalent projections area is not distorted, but distances and shapes often are. This projection type is important for quantitative thematic maps, since maintaining area properties usually is desirable. This is especially important, when then area is part of the mapped data as is the case in density maps. The case data projection is an equivalent projection.

Related work 6

1 2

This is not the first project attempting to visualising large raster datasets in an interactive way. It is also not the first attempt at visualizing this particular dataset.

3

6.1 Tilebased raster visualisation

This is not the first project to follow many of these core concepts. In Bernhard Baumrocks master thesis, he created a webmap, where raster tiles were being colored locally. [Baumrock, 2018a]

4

¹Fixme Note: Write about Cloud optimized Geotiff somewhere

²Fixme Note: Write about gdal2tiles parallelly

³Fixme Note: Write about Sarahs stuff here

⁴Fixme Note: Rewrite into after sarah

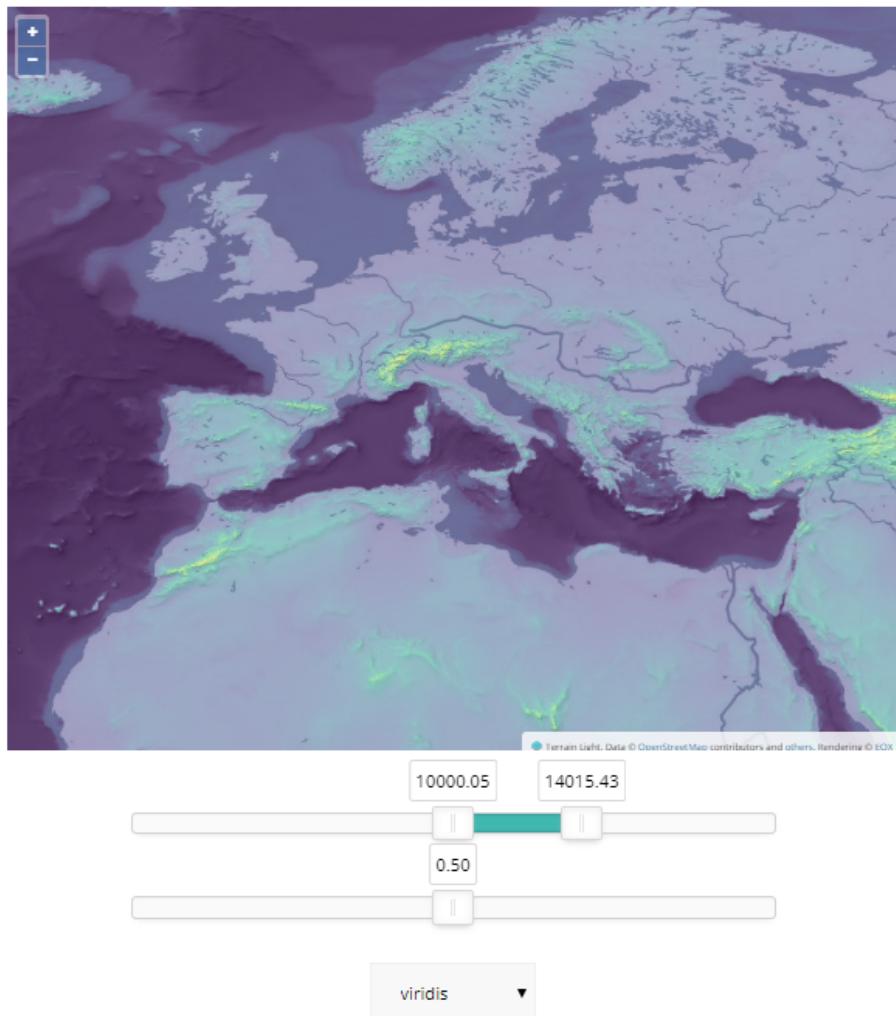


Figure 6.1: The map created by [Baumrock, 2018b]

A picture of one of his maps can be seen in figure 6.1. This particular map is not included in his thesis, but it is the most similar to this project. The map shows an elevation data . Below the map are two sliders and a dropdown list with the label “viridis”. Changing the value in this dropdown list allows the user to switch to another color scheme. The bottom slider is controlling the opacity for the raster layer. The upper slider controls which maximum and minimum values the coloring should be based on. How the map change, when changing the values in the upper slider can be seen in figure 6.2. When zooming in another more detailed layer gets loaded and rendered.

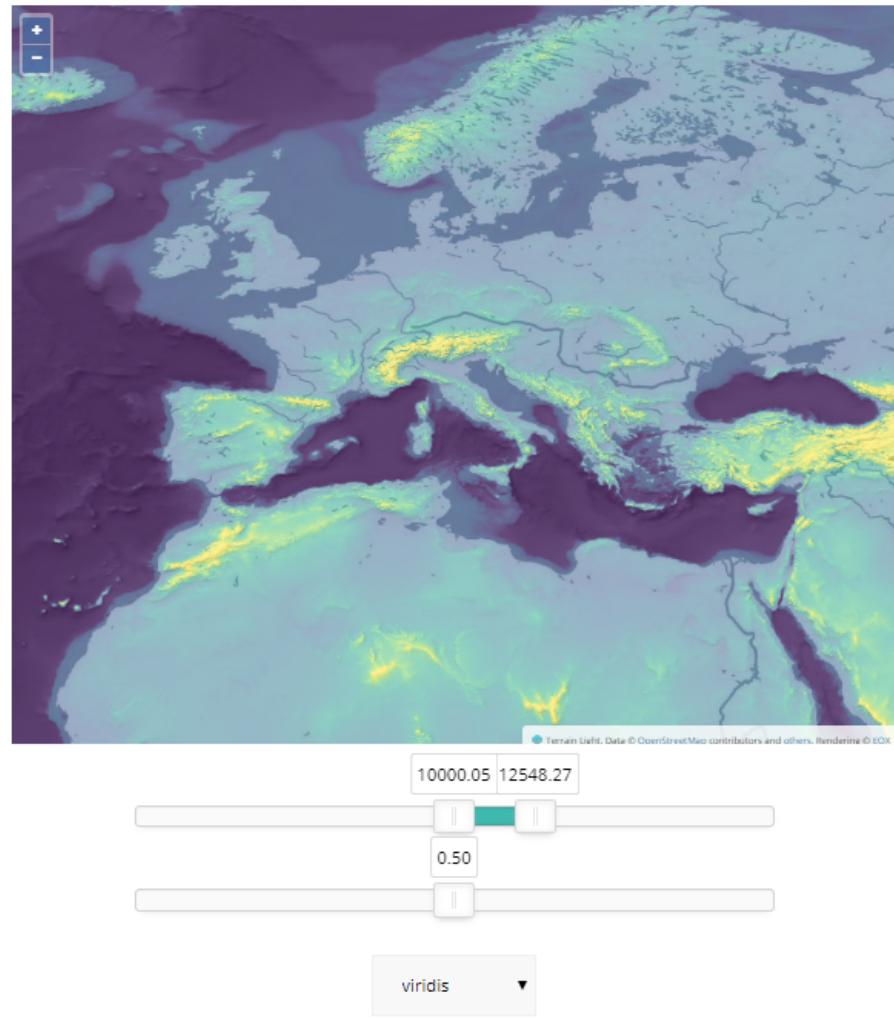


Figure 6.2: How the map change, when using the slider

When taking a closer look at the source code it can be seen that the raster is being loaded as tiff tiles with a bit depth of 16 bit. [Baumrock, 2018b] Only tiles the currently are visible are being requested unless the tile already have been requested. [?]

How the tool functions from a technical perspective is further detailed in section x.

⁵FiXme Note: Add interactive somewhere in the research question

⁶FiXme Note: Who is the target audience?

Components of an interactive maps

7

When creating an interactive map one must consider which features to add to enhance the user experience. To give an overview of some of the basic features used for navigating and get information from a map an example of an interactive map has been used. www.openstreetmap.org has been used as an example of an interactive map. [OpenStreetMap, 2020] A picture of this map can be seen in figure 7.1. Its User Interface (UI) have been numbered with boxes, which have been colored green if the feature would make sense for exploring and comparing large population dataset. The reasoning behind the coloring is explained in section 7.1

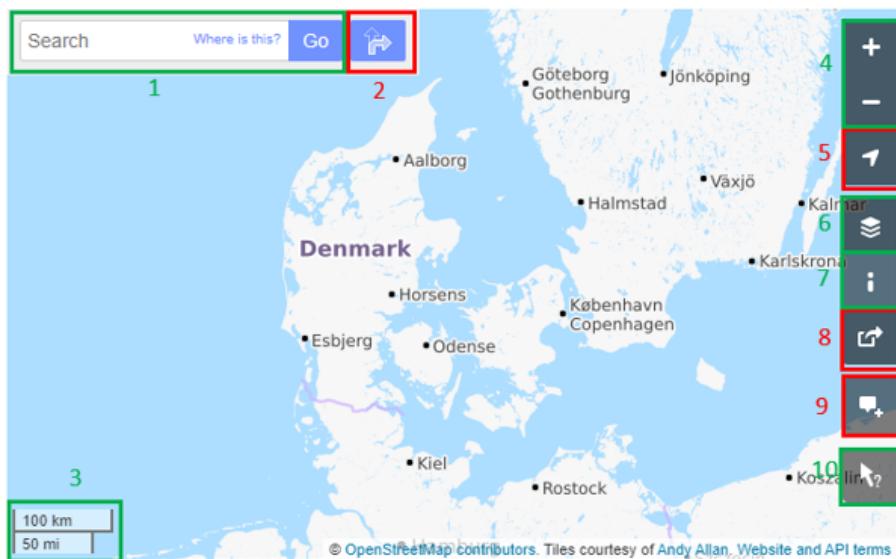


Figure 7.1: An example of an interactive map

What the different parts of the UI does can be seen in table 7.1. The functions of the UI can be classified in five different categories. Using the map navigation UI the user can control which part of the map is being displayed. The real-life navigation tool is used for navigating in the real world. The map explanation category gives the user information about the map. This can be general information about the scale or symbols on the map or specific information about a clicked point. The tool for controlling the displayed data is used for controlling what is being visualized on the map. The last category, communication with others, is for sharing information with other users of the map.¹ ²

¹Fixme Note: Write that my own search bar can be improved

²Fixme Note: Did I write about toggling dual legends?

Number	Name	Category	Function
1	Search bar	Map navigation	Pans to the searched location
2	Route calculation	Real-life navigation	Calculate a route between two given points
3	Scale	Map explanation	Shows the scale of the map
4	Zoom	Map navigation	Allow the user to zoom in or out
5	Show my location	Map navigation	Show the location of the user and navigates to it
6	Layers	Control data display	Control which layer is being displayed
7	Map key	Map explanation	Displays the map legend explaining the meaning of the map keys
8	Share	Communication with others	Generate a link for what currently is displayed on the map or download an image for the map
9	Add note to the map	Communication with others	Add notes about mistakes, so that the map can be corrected
10	Query features	Map explanation	Give more information about a clicked map object

Table 7.1: Overview of the UI elements highlighted in figure 7.1

7.1 Relevant functionalities

Not all of these categories are relevant for this tool and this target audience. Most of the tools in the Map navigation category are relevant. Having a search bar allow the user to quickly pan to an area of interest and the zoom function allows the user to control the extent. The "Show my location" function is less relevant for this target audience. It would be relevant if the target audience was citizens, who just wanted to explore a population projection out of curiosity. These would potentially be interested in knowing what the expected population would be in their neighbourhood. If the researcher are working with the same area, where they happen to live they can pan to this area using the search bar.

The Real-life navigation tool is only useful for route planning, which is not the purpose of this map. The map explanation tools are all useful for the application.

While a scale can give the user a sense of scale, this is only useful if the scale is presented with a unit, which the user easily can comprehend. As mentioned in section x the unit of the projection is in degrees, which would result in the scale also being in degrees. It was therefore decided not to include a scale.³

Having a legend is also central to explain the user which values the colors of the raster would correspond to. Getting more information about a clicked point could also be relevant for the tool. This information could be the exact value in the clicked point. The layer control is important, since this would allow the map to be able to display multiple different population projection.

The last category, Communication with others, is outside of the scope of this project, since these rely on multiple users.

4

³FiXme Note: write about scale - update figure to exclude scale

⁴FiXme Note: Update the figure to have the share box being relevant

5

7.2 Comparing layers

The functions presented so far does not enable an easy visual comparison of raster data. It is possible currently by changing back and forth between different layers with different population projections. This however is a bit inconvenient. An alternative way would be to use comparison layers in a similar fashion to how it currently is being done with the makeCityWebsite tool mentioned in section x. This would require the creation of comparison layers for each combination of SSPs and years, which should be compared. A third option would be to display multiple maps at the same time as illustrated in figure x. These two maps are linked together by a shared view. The means the panning or zooming in one map would do the same operation on the other map.

6

<https://openlayers.org/en/latest/examples/side-by-side.html>

This third option was chosen.

7

⁵Fixme Note: Write here if you come up with more relevant information, which you could throw in here

⁶Fixme Note: Make figure

⁷Fixme Note: Make frontpage

Final design 8

In this chapter the content of the previous chapter is summarised into design guidelines, which the tool will be build upon. These fall into three categories: Technical concepts, visualisation, and functionalities.

8.1 Technical concepts

Five technical concepts have been developed based on the current solutions, related work and initial tests with the case data:

- Load only the needed data
- Coloring should be based on the current extent
- Tiles should not be precolored
- Coloring should be done locally
- File format should have same bit depth as input format

The purpose of these concepts is to ensure that a large dataset can be loaded and visualized correctly.

8.1.1 Load only the needed data

The case data presented in chapter x holds information for the entire world. However, loading data for the entire world would be unnecessarily time consuming. Initial experiments of rendering this quantity of data also revealed that the browser did not have enough memory to load that amount of data. Instead only the relevant data should be downloaded. The relevant data in this case would refer to the data, which the user is able to see. Getting a subset of a raster dataset can be accomplished by dividing it into smaller tiles. Section x will present two different ways this can be achieved.

8.1.2 Coloring should be based on the current extent

The initial exploration of the data showed that the coloring of the tiles should be based of the current extent. This is due to how much the values are varying across the world. Figure 8.1 shows an example of the difference between a map where the coloring is based on all values or if it is limited to the current extent. Both maps illustrate the population in Denmark. The left map has its coloring based on values from the entire world, while

the right has its colors based on the current extent. The left map appears empty since the population density in Denmark is negligible compared with the densest areas in the world. In the right map it is possible to see the location of the most populated cities. Since the left map provide the user with no relevant information and the right one does, the coloring should be based on the current extent.

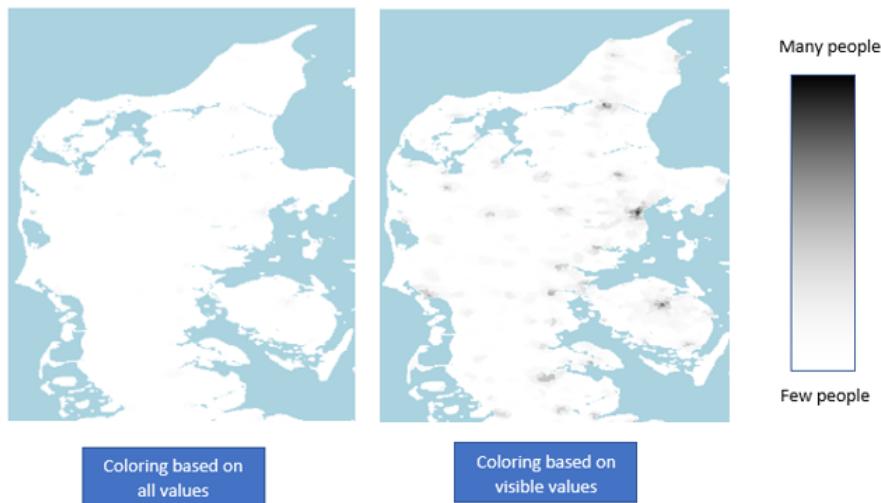


Figure 8.1: Population density in Jutland

8.1.3 Tiles should not be precolored

If the coloring should be based on the current extent and the map is interactive the coloring of the tiles should be done on the fly instead of once and for all. The reason for this is best explained with the example in figure 8.2. This figure is illustrating the population in a small subset of India. The three boxes are illustrating examples of possible map extents, which the user could get by zooming in on different parts of the map. Each of these three have a different max value in their cell with the highest population. The blue square is covering the city center of Indore and therefore have a high value than the green, which is covering the outskirts of the city. The last square does not include any part of a bigger city and therefore have a lower value. Since their max values are different, they would each need their own coloring. If the tiles were to be precolored it would be necessary, create three coloring of the tiles – one for each the extents. However, since the map is interactive, the extent is not limited to those three. If the user instead were to zoom in on an area between the red and blue square a new max value would maybe be found. When scaling this example up to the entire world and to multiple zoom extents there would be thousands of combinations. If all of these needed to be prerender it would both require lots of initial processing time and storage space for an immense amount of data.

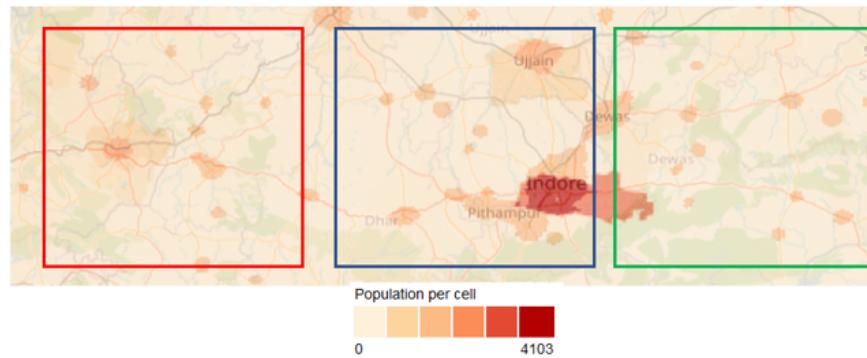


Figure 8.2: Population density. The three boxes are examples of possible extents an interacting user could get by zooming.

The alternative is to color the tiles on the fly. When the user would zoom to a new area the map, a script could register the highest currently visible value. This information could then be sent to recoloring script, which would color the tiles based on this. This way only tiles with the needed colors would be rendered.

8.1.4 Coloring should be done locally

The coloring can be done in two ways as illustrated in figure 8.3. It can be done by sending the information to a tileserver, which then would color the tiles and send them to the client. Alternatively, the coloring could be done on the client.

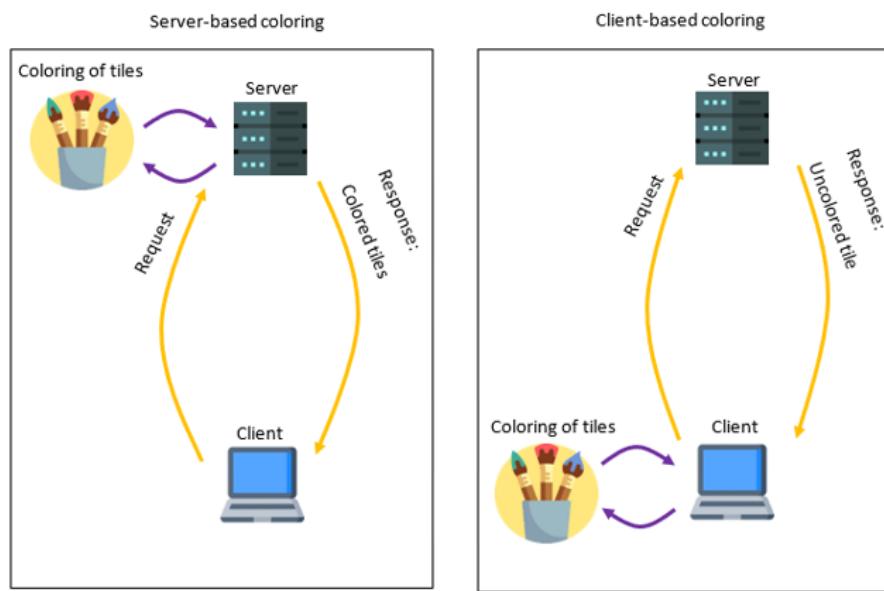


Figure 8.3: Difference between coloring the tile on the server and locally

Comparing the two options is best done with an example, which has been illustrated in figure 8.5. Here the blue and green box are two different possible views both containing 16 raster tiles. The two boxes have different max values, since blue one contains the city

center. In the example a user will pan from the blue view to the green one. If the coloring is done on the server side the initial 16 tiles in the blue box will be requested on load. Then when the user pan to the green view the 16 tiles in that box will be requested and colored. The four tiles that are shared between the blue and the green boxes must be requested again, since they need a new color due to the change in max value. This is not the case if the coloring is done locally. In this case only 12 tiles would have to be requested, when changing from the blue to the green view. Since the coloring is happening locally the coloring script can recolor the four tiles it already has downloaded from the initial load. In the right part of figure 8.3 the yellow loop would be run 12 times, whereas the purple one would be run 16 times. This should result in a faster experience for the user.

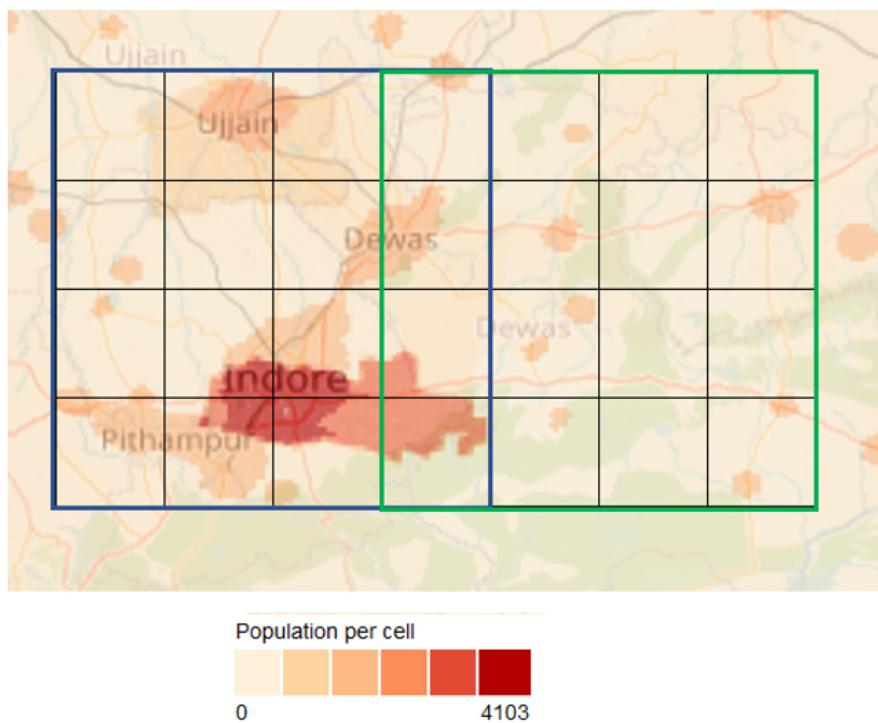


Figure 8.4: Example of why local coloring is needed. The squares are different extents used for the example

8.1.5 File format should have same bit depth as input format

As mentioned in section x the data from the case has a bit depth of 32 bit. To visualize the data the file format must not be changed to a format with less than 32 bits. Figure 8.5 is an example of how a subset of the case data look originally and when converted to an 8-bit format. As mentioned in section x the pixels in an 8-bit raster can have values between 0-255. This means that this format is unable to correctly display the original data range, where the data range is covering thousands of values. The original data get clamped into the 8-bit format producing wrong result <https://gdal.org/programs/gdal2tiles.html>.

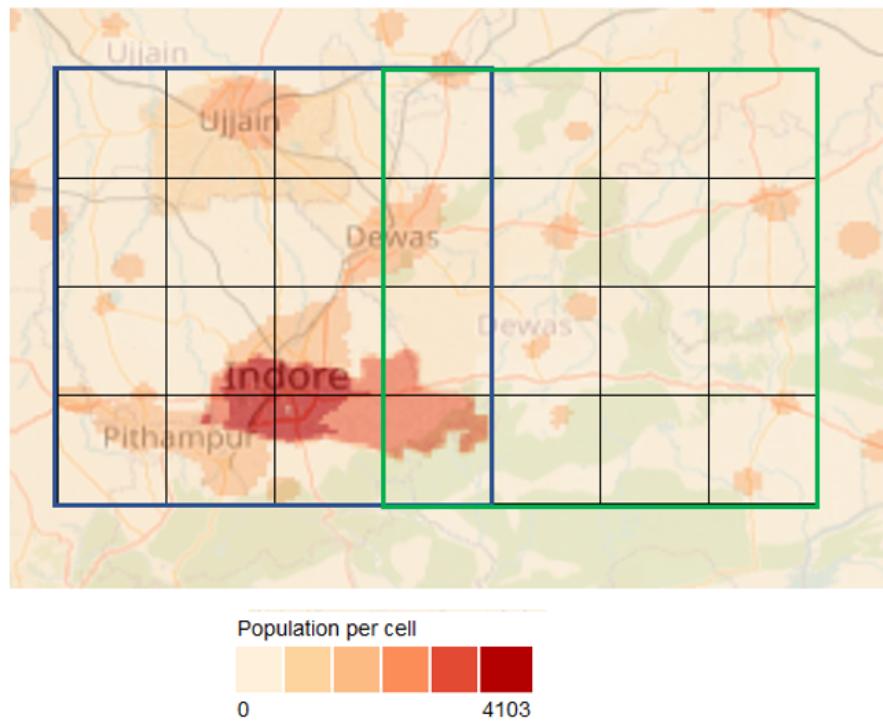


Figure 8.5: 32 bits crammed into 8 bits

Normally this issue could be worked around by rescaling the input data to the new format [GDAL, 2020]. This rescaling would be a lossy compression resulting in loss of information since the original data cannot be expressed with values between 0-255. [?] For the normal use of tiles this would not be a problem since tiles would only be used for visualization. If the tiles are being recolored locally the tiles are being used for more than just visualization. It is necessary to access the data within the tile. This means that solutions resulting in a lossy compression are not an option. To ensure that the correct data reach the client the file format used for tiles must have a bit depth of 32.

8.2 Visualisation

8.3 Functionalities

This section lists the functionalities, which¹

8.4 Implemented design

Most of these technical concepts are fulfilled by the map created by Baumrock [2018a] as mentioned in section ??.

¹Fixme Note: Rewrite the beginning, so that it does not mention Visualisation and Functionalities

It is locally visualizing only the necessary data. The technical concept related to bit depth is relevant in the creation of the tiles, so it does not really apply to this. It is also to some extent allowing the user to color the layer based on the current extent. The user can adjust the maximum and minimum values but does not know what the maximum values are in the current extent.

It has therefore been decided to build the projection upon the foundation created by Baumrock. One major change will be the automatization of the coloring. Instead of the user defining maximum and minimum values to color the map based on, it is instead done automatically. The map detects the maximum and minimum values, which are within the current map extent and adjust the color scale dynamically to fit these.²

The concept for a solution has been illustrated in figure x. The tool will use two maps with a shared view, so that both maps show the same area. Each map will show a different scenario, so that different population projections can be compared. Not all of the functionalities presented in chapter x have been included in the design. Some of these functionalities will be mentioned in the discussion in chapter x.

8.4.1 Visualisation

The sequential color scale chosen for the visualized rasters have been created by colorbrewer.org. It was decided to use six different colors on the maps. This is despite the fact that ? advise that more colors safely can be used, when similar colors are neighboring each other. The reason for this is that even though similar colors are neighboring each other within the individual map, the user still has to compare none-neighboring colors, since comparison is done between the two maps. The breakpoints between the different classes are determined as percentage steps of 20 % of the current maximum value.

The projection used EPSG:4326, since this was both used in Baumrocks original map and was the projection for the case data. This projection turned out to have consequences for the implementation of a scale. This will be expanded upon in section x.

3

8.5 Number of data classes

After selecting a color scale, it is important to select how many different colors should be displayed on the map. Having a high number of data classes produces a detailed map, since there with more classes is less data generalization. There can also be too many different data classes to the point where it becomes difficult to tell them apart. How many classes can be used depends on the spatial distribution of the data. Maps with irregular spatial patterns should have five to seven different classes, whereas regularly patterned maps can have more. The reason for this is that when similar colors are placed side by side it is easier to distinguish between them.[Brewer et al., 2013]

²FiXme Note: change code so that min values also get calculated

³FiXme Note: write about classes in visualisation

Part II

Development

In this chapter it is explained how the tool has been developed. The first section is describing the coding languages used and the plugins used to create the tool. The following sections are describing the necessary steps to build the tool. These steps can be seen in figure 8.6.

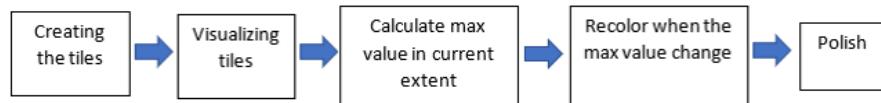


Figure 8.6: An overview of the code

First it is necessary to divide the raster into tiles to be able to limit the loaded data to the current extent. How this is done is explained in section x. Section x describes how the tiles have been visualized. To be able to color the tiles based on the current values it is required to know what the current values are. In section x it is explained how this information is calculated. The raster layer is then rerendered with these values as described in section x. Lastly some measurements were taken to ensure a more user-friendly experience. These additions are described in section x.

Coding languages and plugins 9

For this project, the coding language Python was used to create the tiles, while visualization of the tiles was done in a webgis build with the languages HTML, CSS and Javascript. This webgis have been tested on a local caddy server for reasons explained in section x

9.1 Languages

Python

Python is programming language with a simple syntax, which functions across multiple different platforms. This simple syntax means that Python can achieve the same as some other coding languages in fewer lines.[W3Schools, No dateb] Python was used in this project because of the gdal library expanded further upon in section x This project has been using version 3.6 of Python.

HTML

HTML is short for Hyper Text Markup Language. It is the language used for defining and structuring a web page's content.

CSS

CSS is an abbreviation for Cascading Style Sheets. This language defines how the HTML will be displayed.

Javascript

How a webpage behaves is defined by the language Javascript. This is what makes the web page interactive. [W3Schools, No datea]

9.2 Libraries

To transform the raster data into smaller tiles the python library Gdal is used.

Gdal

GDAL is library for translating between multiple different geospatial data formats. [GDAL, 2020a]

Included in this library is the gdal2tiles program, which can divide raster files into smaller tiles. At the time of using this program it was only able to generate tiles structured after the TMS standard. The function to follow the XYZ structure was added the 3th of May 2020 [GDAL, 2020] [GDAL, 2020b] The script rendering the tiles in the map was based on the XYZ structure. This meant that the rows of tiles were ordered incorrectly when the generated tiles were imported. Therefore, the official version of gdal2tiles was replaced by a version made by a github user named commenthol. This version is modified to allow the creation of tiles following the XYZ structure. [Commenthol, 2020] Both the official version and the modified version have their output format as mbtiles with a bit depth of 8 bit. To be usable for this project a bit depth of at least 32 bit is needed, as mentioned in section x. The workaround for this issue will be explained in section x.

Openlayers

The map in which the tiles are being showed are created in Openlayers, which is an open source JavaScript library for creating dynamic maps for web pages. [Openlayers, No Date] Openlayers was chosen because the tool presented in related work was built in Openlayers. Therefore, using Openlayers would enable expanding upon this existing tool instead of starting from scratch.

The existing tool was created with the debug version of Openlayers, which this tool also have build on. The debug version is a version of the Openlayers, which have not been minified. Because of this it is a larger file than the regular one. <https://gis.stackexchange.com/questions/155529/whats-difference-between-ol-js-and-ol-debug-js-files-in-openlayers-3>¹ Attempts at replacing it with the minified version resulted in errors from olGeoTiff, which have been build based on the debug version. Optimizing other parts of the code was given higher priority, so the debug version never got replaced.

olGeoTiff

olGeoTiff is a Javascript class for visualising geotiff tiles in Openlayers, utilising the libraries geotiff.js and Plotty. The visualized tiles are being processing in the client instead of on a server. It was used in the map presented in section x. The class is a modified version of Openlayers WMTS layer, where the internal tile loading function has been changed.

¹Fixme Note: link

The regular function would request precolored tiles and then add them to the map. The tiles requested by the modified version are not precolored and need to be processed before getting added to the map. A simplified illustration of this processing is illustrated in figure x. This simplified figure is enough to explain the mechanics of the class but does not detail the callback function structure. Aside from being used for error handling callbacks are also necessary to ensure that Openlayers do not try to add the tiles to the map, before they have been processed. More detailed figures can be found in [Baumrock, 2018a] thesis.

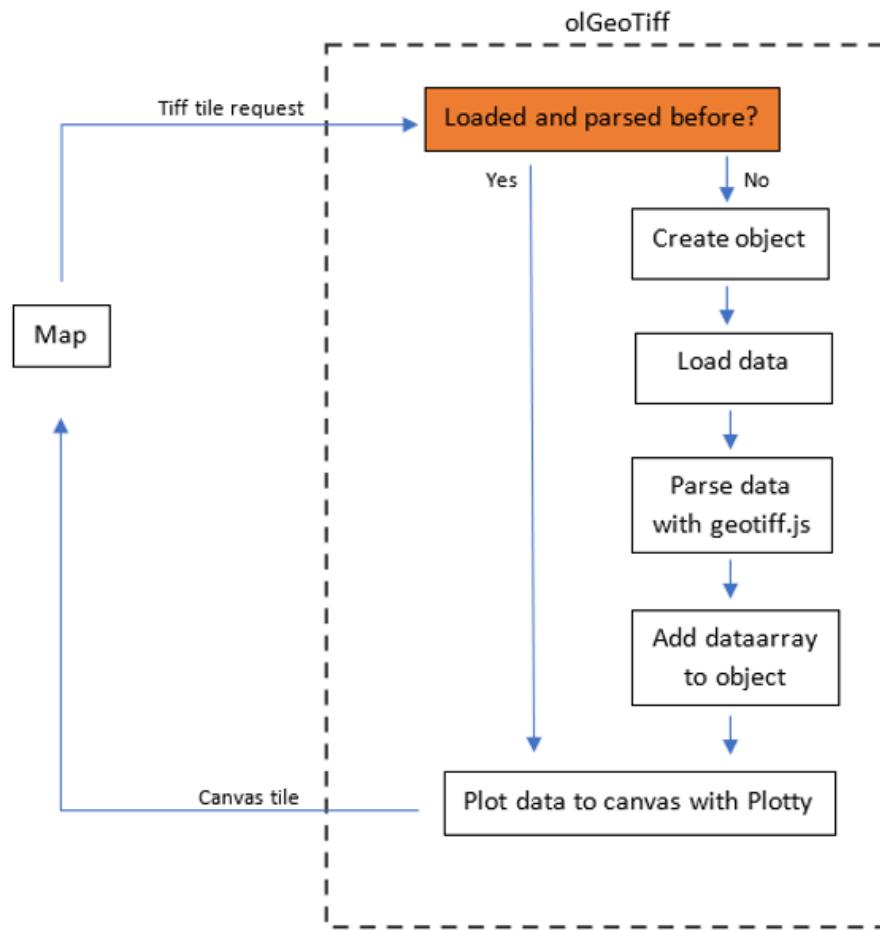


Figure 9.1: A simplified illustration of how olGEoTiff functions

To ensure that tiles are only being downloaded once an object keeps track of all the downloaded data. The object is organized by the tiles' url. Whenever tiles are being requested the object will always be checked to see if it already contains the url. If it does not the object will be updated to include the requested tile. The tile will then be loaded before being processed. [Baumrock, 2018a] The processing is done with the TIFF parser geotiff.js [Schindler, 2018] and Plotty, which is a library for creating images from data arrays [Santillan, 2020].

The loaded tiles first get parsed with geotiff.js and added to the object before Plotty get used to render tiles in the designated colors. Then the tiles get added to the map. [Baumrock, 2018a] The olGeoTiff also have a redraw function, which when triggered will

redraw the tile layer based on the current designated colors. This was for instance used in Bernnard Baumrocks map, whenever the color sliders were changed.

9.3 Testing server

To test the tool a local test server was set up. This was initially a Python-based http server, but was later replaced with a Go-based Caddy server to improve performance. The performance difference between the two is elaborated upon in section x.

Caddy was chosen because it supports the second version of Hyper Text Transfer Protocol (HTTP) <https://dassur.ma/things/h2setup/>

which is used for the communication between the client and server. <https://www.w3schools.com/whatis/whatishttp2.asp>

HTTP/2 is allowing a faster communication between them https://web.dev/uses-http2/?utm_source=lighthouse&utm_medium=devtools

Developing the tool

10

This chapter details how the visualization tool was built.

10.1 Creating the tiles

The tiles got created using a modified version of gdal2tiles, which was changed to create tiles following the XYZ format. However, since the default output bit depth is too low, it would have to be modified further. The file format would also have to be changed to tiff in order to be visualized with the geotiff.js library.

10.1.1 Changing the file type

Changing the file format can be done by changing two lines in the gdal2tiles script:

```
1 #Original code
2 #self.tiledriver = 'PNG'
3 #self.tileext = 'png'
4 #New code
5 self.tiledriver = 'GTiff'
6 self.tileext = 'tiff'
```

Listing 10.1: Changing the file format

This changes the raster driver from png to the geotiff format and the file extension from png to tiff. [GDAL, No date] The geospatial information, which is the difference between a geotiff and a regular tiff, gets lost in process. This information is not important for this project since the tiles are being loaded based on their name and folder placement, not based on the internal metadata.

Running gdal2tiles with these changes will produce a tiff file, which still would be limited to 8 bits.

10.1.2 Increasing the bit depth

The reason for the bit limit is that gdal2tiles uses the memory dataset driver, which have 8 bits as default. This default can overwritten to 32 bits by adding “gdal.GDT_Int32” to every instance where the driver is being used as demonstrated in code x. [Tosovsky and Strip, 2014]

```

1 self.mem_drv = gdal.GetDriverByName('MEM')
2 ...
3 #Old code
4 #dstile = mem_drv.Create('', tile_size, tile_size, tilebands)
5 #New code
6 dstile = self.mem_drv.Create(' ', self.tilesize, self.tilesize, tilebands, gdal.GDT_Int32)

```

Listing 10.2: Increasing the bit depth

The memory driver is being used four times, which all have been changed in a similar fashion.

The script is now generating tiff tiles correctly. However it also generates KML files for visualization in Google Earth, which is undesired since it would increase the processing time and required storage space.

10.1.3 Prevent generation of Google Earth files

gdal2tiles is automatically generating KML files if the projection is EPSG:4326. According to the documentation for the official gdal2tiles this can be disabled with the command "-no.kml". [GDAL, 2020]

This command does not prevent the creation of the files in the modified version. Therefore the automatic generation has been removed from the code. This was done by commenting out the lines shown in x.

```

1 if self.out_srs and srs4326.ExportToProj4() \
2     == self.out_srs.ExportToProj4():
3     self.kml = True
4     self.isepsg4326 = True
5     if self.options.verbose:
6         print('KML autotest OK!')

```

Listing 10.3: Increasing the bit depth

1

10.1.4 Command for tile creation

After all the modification are in place the script can be run by running the command illustrated below:

```
python <path to modified gdal2tiles script> -leaflet -zoom=<desired zoom levels> -profile=raster -webviewer=none <input file> <output directory>
```

Most of these inputs are from the documentation for the original gdal2tiles, though the leaflet command is from the modified version. This command ensures that tiles are being created following the XYZ formatting instead of the TMS. The zoom command defines which zoom levels should be generated. An example of an input could be 0-2, which would generate tiles for the zoom levels 0, 1 and 2.

¹Fixme Note: Write about the addition to time the whole thing

Setting the profile to raster was done because the two other options (mercator and geodetic) resulted in the error message "list index out of range", while the raster profile gave a useable product. The selection of this options and its consequences is expanded upon in section x.

The webviewer option prevent the generation of webviewers to visualize the tiles. These default webviewers is created to visualize tiles of the original format, so they would be unable to visualize the modified tiles. The input file is the name of the processed file and output directory is the folder, where the tiles get created.

10.1.5 Result

While creating the tiles the error message "ERROR 1: Buffer too small" gets displayed. The tiles are still being generated with the correct formatting and bit depth. This is an error message from numpy, which gdal2tiles are using. [Tosovsky and Strip, 2014] The generated tiles appears as they should, so this error message is potentially referring to the geospatial information, which gets lost.

The tiles at the rightmost and bottom edges have an edge as shown in figure x. This edge is not being shown, when the tile is being loaded into the map. A comparison between the input file and the tile shows that all data is being displayed. The bottom part of the input file is the last part before the edge. This edge seems to be generated because the input file was not perfectly dividable with the tile size as illustrated in figure x. <https://github.com/ARIisgaard/Thesis/issues/22> This edge could be another explaination for the error message.

2

3

The script also produces an xml file with metadata. An example of the content of said metadata file can be seen in code x.

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <TileMap tilemapservice="http://tms.osgeo.org/1.0.0" version="1.0.0"><Abstract/><SRS>GEOGCS["WGS
   84",DATUM["WGS_1984",SPHEROID["WGS 84",6378137,298.257223563,AUTHORITY["EPSG","7030"]]],
3  AUTHORITY["EPSG","6326"]],PRIMEM["Greenwich",0,AUTHORITY["EPSG","8901"]],UNIT["degree",0.0174532925199433,
4  AUTHORITY["EPSG","9122"]],AXIS["Latitude",NORTH],AXIS["Longitude",EAST],AUTHORITY["EPSG","4326"]]
5  </SRS><BoundingBox maxy="25.00000000000814" maxx="80.9999999999989" miny="19.00000000000815"
   minx="72.9999999999989"/><Origin y="19.00000000000815" x="72.9999999999989"/><TileFormat
   extension="tiff" mime-type="image/tiff" height="256" width="256"/><TileSets
   profile="raster"><TileSet order="2" units-per-pixel="0.0083333333333" href="2"/><TileSet
   order="3" units-per-pixel="0.0041666666667" href="3"/><TileSet order="4"
   units-per-pixel="0.0020833333333" href="4"/><TileSet order="5"
   units-per-pixel="0.00104166666667" href="5"/><TileSet order="6"
   units-per-pixel="0.0005208333333" href="6"/><TileSet order="7"
   units-per-pixel="0.00026041666667" href="7"/><TileSet order="8"
   units-per-pixel="0.0001302083333" href="8"/><TileSet order="9"
   units-per-pixel="0.00006510416667" href="9"/></TileSets></TileMap>
```

Listing 10.4: The metadata from the xml file generated by the modified gdal2tiles

²Fixme Note: Make figures - remove link

³Fixme Note: Test if normal gdal2tiles tiles have edges - if false, then point it out here

10.2 Processing time

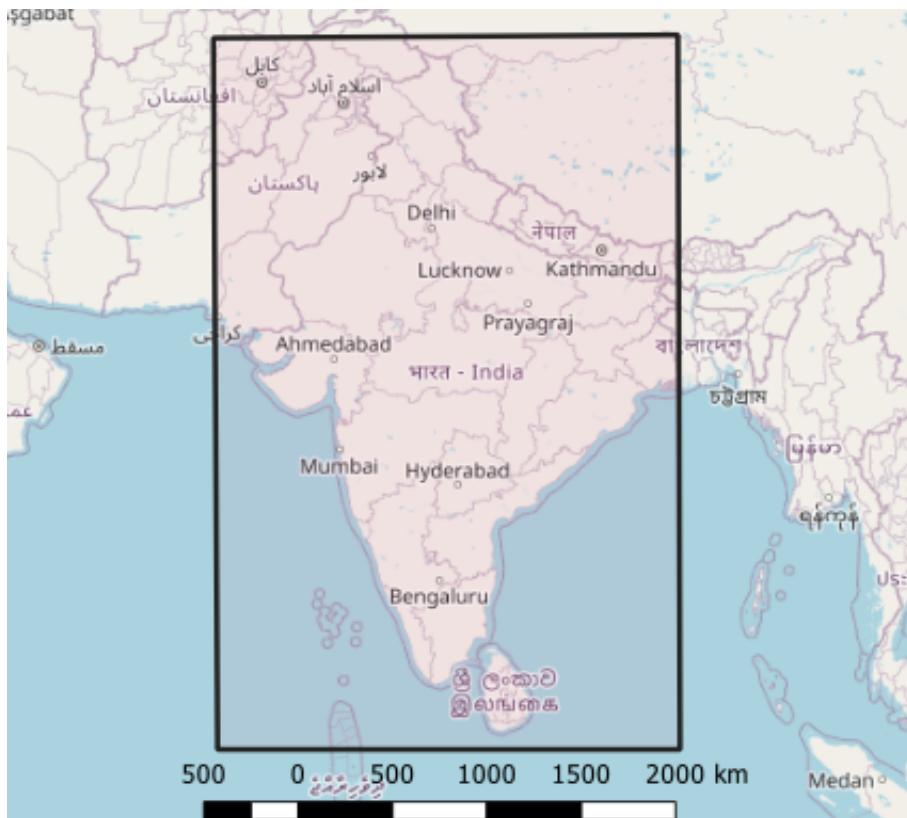


Figure 10.1: Illustration of the area used for timing the tile generation

The processing time is tested using the area of India shown in figure [?]. This polygon with a $9\ 291\ 842\ \text{km}^2$ area was processed into tiles for each zoom level between 0 to 9. This is the equivalent of 349525 tiles as calculated with the formula presented in section x.

$$2^{2n}2^{2*9} + 2^{2*8}...2^{2*0} = 349525 \text{ tiles} \quad (10.1)$$

The creation of these tiles took 6 hours and 4 minutes. The amount of time mainly depends on the highest zoom level, since the number of tiles in a layer get quadrupled whenever the zoom level increases. In comparison the same area at the zoom levels 0-7 only took 22 minutes.

This processing time is not ideal, but it can theoretically be shortened with multiprocessing.

10.2.1 Parallel processing of tiles

Multiprocessing is the usage of multiple of the computer's processors. Python is by default only able to use a single processor, even if multiple are available. This can be circumvented with the multiprocessing module. [Data@Urban, 2018] The modified version by Commenthol [2020] also has a multiprocessing version allowing each processor of the computer to work separately on generating tiles. However, the task is not properly

distributed between the processes. This means that not all of the tiles are being generated. [Herjar, 2019] Therefore, the slower single process version has been used for this project.

10.3 Visualizing tiles

After the tiles got created, they are stored in a folder, which gets uploaded to the testserver along the index file. The metadata from the xml file must be loaded into the map to be able to visualize the tiles. Normally this could be done using the WMTSCapabilities() function in Openlayers. [Openlayers, No dateb] However, the formatting of the xml file produced by gdal is different from the format, which this function can read. Therefore, a small script has been created to parse the xml file and store the information in a metadata object. This object, tileMetadata, stores the bounding box, origin, center coordinates as well as tilesize. The initial version of the object also stored the resolution data, which is called units-per-pixel in code x. This led to some inconsistency when loading tiles, which had been generated without some of the lower zoom level. This will be further expanded upon in section x. Therefore, the resolution was instead generated using the script x, where 0.0333 is the value for units-per-pixel at zoom level 0.

```

1 for (var z = 0; z < 14; ++z) {
2 // generate resolutions and matrixIds arrays for this WMTS
3 //The number in the resolution calculation is the units-per-pixel value at zoomlayer 0 in the xml
4 // file generated by gdal2tiles
5 resolutions[z] = 0.03333333333514 / Math.pow(2, z);
6 matrixIds[z] = z;
7 }
```

Listing 10.5: The JavaScript in the project

Using this metadata, the tiles could be visualized using the olGeoTiff class.

10.3.1 Custom colors scheme

Using colorbrewer a custom sequential colorscheme was generated. This scheme was added to Plotly and selected as a color palette. *This code might change a bit, so wait with writing

10.4 Loading data at a wrong resolution

Figure x shows how the map looked, when loaded with a manually defined max value.

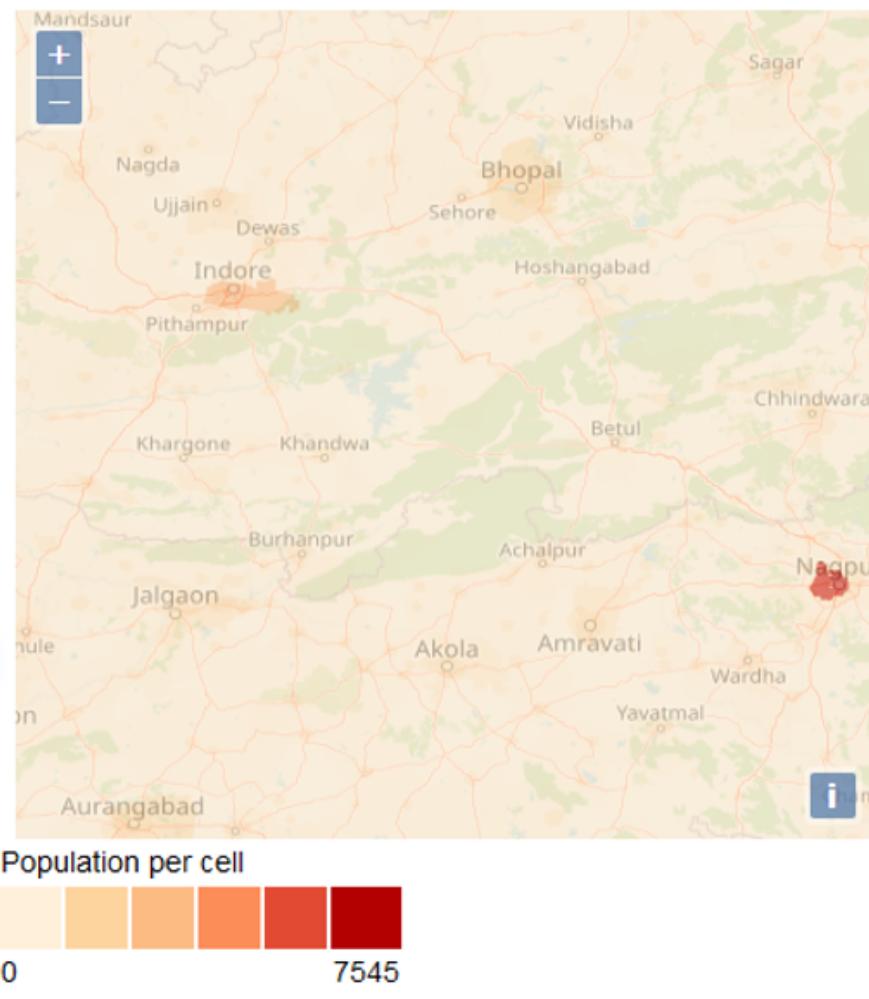


Figure 10.2: The map looking right, but with tiles from the wrong zoom level

While the map appeared to look alright, it was loading tiles from the wrong zoom level. The tiles always got loaded from a zoom level 3 lower than intended. So the map in figure x is visualizing the map at zoom level 7 but loading and displaying tiles from zoom level 4.

Some experiments with loading the tiles from the correct zoom level Setting up Openlayers to download tile from the current view and correct zoom level crashed the map with the error message “Insufficient Resources”. The amount of loaded tiles seems unnecessarily high and size of the tiles too small. This seems to indicate, that the tiles, which the modified version of gdal2tiles associates with zoom level 7 in fact belongs to a higher zoom level.

The reason for this bug was never discovered and the bug never got fixed. When the resolution data was gathered from the metadata file, the difference between the loaded and the actual zoom level would be between different sets of tiles. This variance would depend on the which zoom levels were not being generated. So, if only the zoom levels 2-7 were generated, the difference would increase by one for each missing layer. In this case the loaded tiles would instead be wrong by 5, calculated as the default wrongness of 3 plus 2 for missing zoom layer 0 and 1. This bug will be defining for the rest of the code.

10.5 Calculate max value in current extent

Calculating the highest value in the current extent can be divided into two smaller tasks. Figuring out which tiles currently are being displayed and processing these tiles. It was decided not to calculate the lowest value in each tiles, since the calculation noticeable increased the processing time without yielding any benefit. This is expanded upon in section 14.6

4

10.5.1 Current displayed tiles

The tiles, which currently is within the view, can be found using the Openlayers tileGrid method forEachTileCoord. This method can trigger a function for each tile within a given zoom level and extent. [Openlayers, No datec] The method is going through the tiles based on their coordinates, but this can be translated to the tile urls using the getTileUrlFunction.

```

1 var tileUrlFunction = wmslayer.getSource().getTileUrlFunction()
2 var zoomlevelAdjustment = 3
3 wmslayer.getSource().getTileGrid().forEachTileCoord(loadExtent, mapZoom - zoomlevelAdjustment,
4   function(tileCoord) {
5     tileName = tileUrlFunction(tileCoord, ol.proj.get('EPSG:4326'))
6     %Find max value
7   }

```

Listing 10.6: The JavaScript in the project

forEachTileCoord is triggering for each tile in the given zoom extent. This means that it on zoom level 7 it would load the tiles, that should be rendered on zoom level 7. Due to the bug mentioned section 10.4 it is not the tiles from that zoom level, which are being displayed. Instead the tiles from zoom level 2 are being displayed. This will complicate some of the next steps. The adjustment to the zoom level is in order to load the tiles, which should have displayed.

10.5.2 Max value for each tile

If the tiles added to the map had been from the correct zoom level calculating max value of a tile could have been done as olGeoTiff were running. olGeoTiff holds the values for the tiles in a dataarray already, so finding the maximum value could be done with a single line adding a tiles max value to the dictionary for that tile. This line is shown in code x, where urlToTiff is the name of the object, which holds the data. maxValueTileData[url].maxValue = Math.max(...urlToTiff[url][0]) However, since olGeoTiff holds data from an incorrect zoom layer this would not function. A possible solution would be to trigger forEachTileCoord for the same wrong zoom layer as the displayed tiles are from. This solution was not implemented since it would result in a map with could potentially so misleading information. The explanation for this has been illustrated in figure x.

⁴Fixme Note: Mention the negative values in case data

Todo: Figure with showing how large the loaded tiles are compared with how large they should be.

The tiles from the wrong zoom level are so large, that they would show data outside the view of the map. This means that the coloring could be based on the information that was outside of the current view.

The alternative solution is to run another function going through the tiles, which should have been displayed and find the highest value among these. This would result in a map where the wrong tiles were being colored based on information from the smaller correct tiles. This solution is by no means ideal. It means requesting tiles from two layers, one for displaying on the map a one for calculating the relevant max values. Loading more data than necessary will make the script slower, but since no better solution was not found this have been implemented. The calculation of the maximum value for each tile was done in a very similar fashion to how olGeoTiff operates as illustrated in figure 10.3. An object holds information about all tiles, which have been processed and the max value is known. When the function is run for a tile it first checks if the tile already has been processed. If that is not the case, then it will create an object for the data, which it then will load and parse with geotiff.js. The data array with parsed data will then be run through as descripted in code x to calculate the maximum value. This maximum value will then be returned to the map.

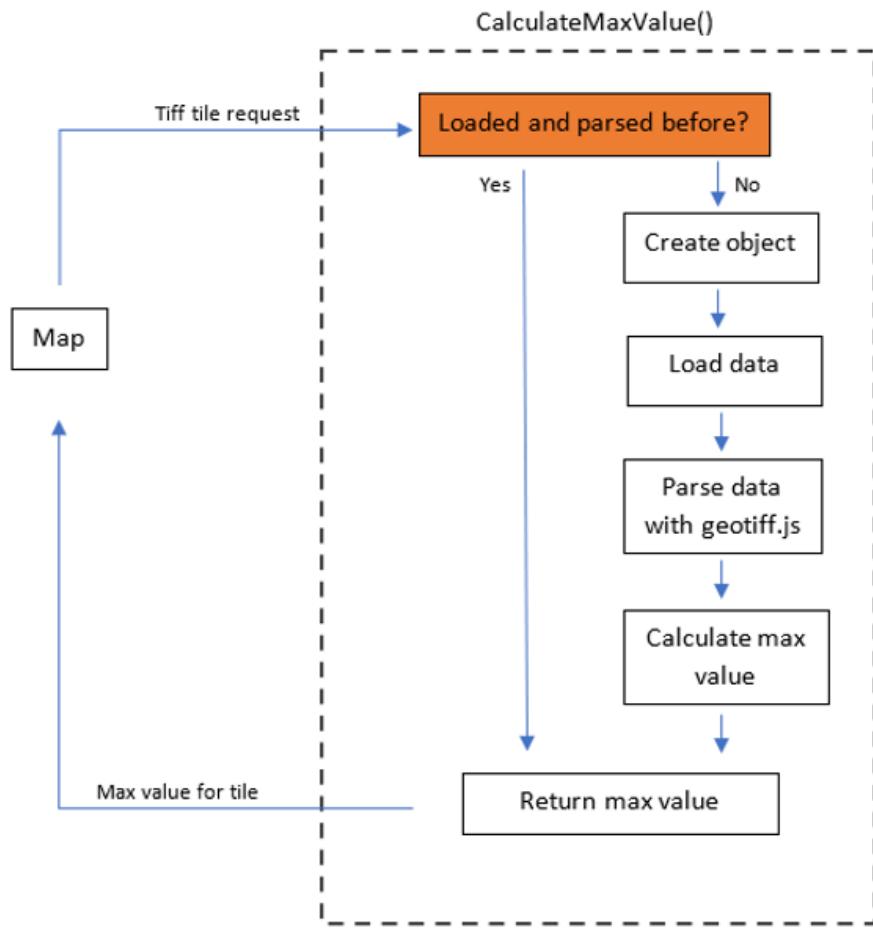


Figure 10.3: Flow diagram of the calculation over max value in a tile

10.5.3 Highest tile value the current displayed

To find the largest tile value among all the currently displayed tiles `forEachTileCoord` as mentioned earlier is used. `forEachTileCoord` does not have a trigger for when it has run through all the tiles. This functionality is necessary for ensuring that the produced maximum value is actually the maximum value. Without a precise end trigger the coloring applied to the map could be based on the biggest value found before the coloring script started running instead of the absolute highest value in the current display. Since this trigger is necessary it has been created by running `forEachTileCoord` another time to count the number of tiles. This part of the code can be seen in code x.

```

1 var tileNumber = 0;
2 wmslayerMap1.getSource().getTileGrid().forEachTileCoord(loadExtent, mapZoom - zoomlevelAdjustment,
3   function(tileCoord) {
4     tileNumber++;
5   })
  
```

Listing 10.7: The JavaScript in the project

The total number of tiles in the current extent can then be used as a trigger. The recoloring of the map can then be delayed until the function for calculating the maximum value have

been running for the same amount of times as there are tiles on the map. In practical terms this can be accomplished by adding a counter and an if statement to the maximum-value-function. The counter would check how many times the function has run. The if statement would check if the amount of time the function had been run was equal to the number of tiles. If this is the case and the registered max value is different from the previous one the recolor function would trigger. The calculation of the maximum value is the function presented in figure 10.4. It is running asynchronously because the array otherwise would be filled with “undefined” values instead of actual values. Running it asynchronously ensures that the script awaits the calculation of the value.

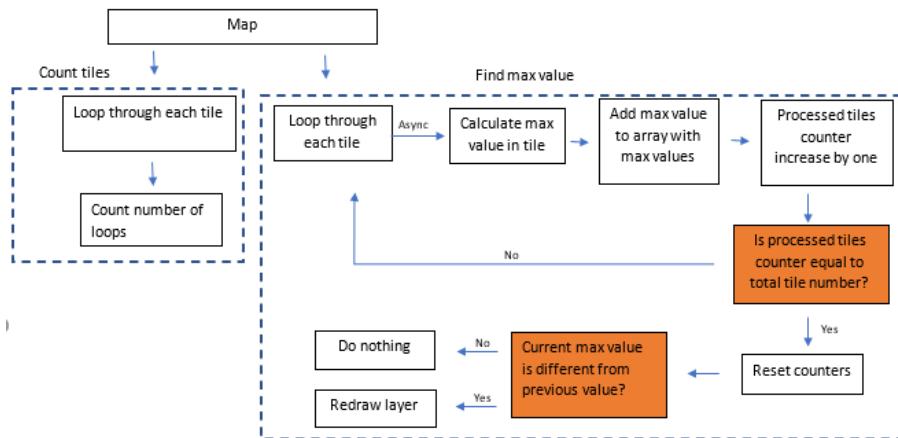


Figure 10.4: Finding the largest value in all currently displayed tiles

10.6 Recolor when the max value change

The recoloring function was already part of olGeoTiff. However, in Bernhard Baumrocks’ thesis this recoloring was triggered manually by the user when changing the color sliders. In this project the changing of colors will trigger automatically. The function for recoloring the map have therefore been set up to trigger on the user stopping this changing the view. By not triggering before the map movement is finished a smoother user experience is ensured, since new tiles does not have to processed before the user is finished with interacting with the map. The recolor function is running through all of the code mentioned in the previous sections.

```

1 map.on("moveend", function() {
2   recolorMap()
3 });
4 });
  
```

Listing 10.8: The JavaScript in the project

10.7 Polish

In addition to the rendering of the raster in the map, some features were also added to improve the user experience.

10.7.1 Two maps

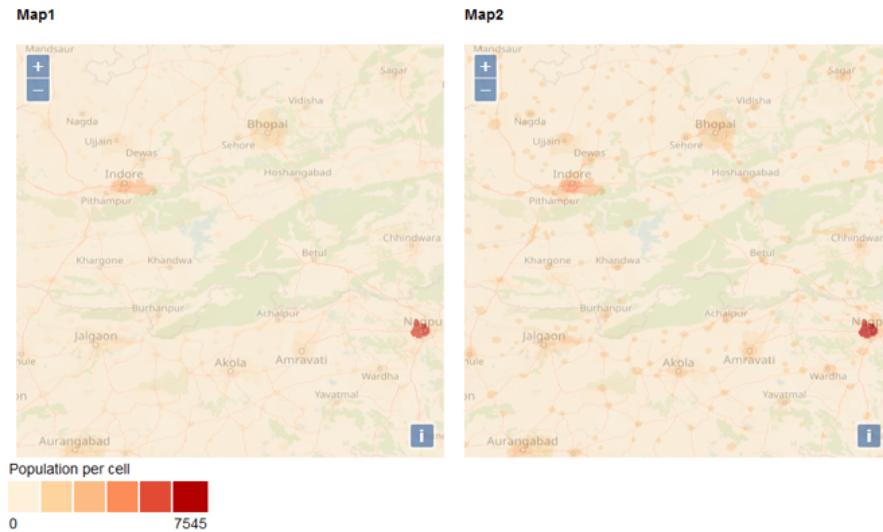


Figure 10.5: A second map

To be able to compare different rasters with each other a second map was added as illustrated in figure 10.5. This second map is having its own raster dataset but sharing the view with the first map. This means that the two maps would always show the same area. Panning or zooming in one map would do the same action in the other map. The two maps are sharing the same legend.

The two projection are also sharing the same color scale. This means that the coloring is done based on the maximum value in the current extent of either of the maps. To accomplish this changes were made to the function, which were finding the maximum value among the tiles as illustrated in figure x.

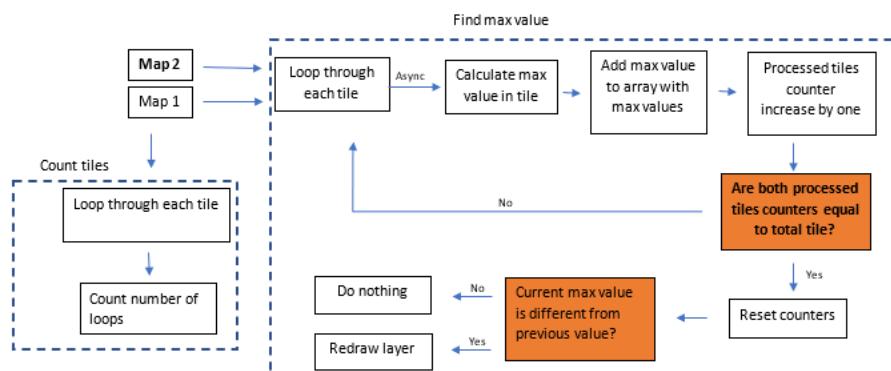


Figure 10.6: Calculating the maximum value for two maps. Changes from the original have been written in bold

This function is now being used twice, once for each layer. If the same function as presented in section x was used, then the redrawing would potentially trigger twice.

To avoid this each of the maps are given their own processed tiles counter. The if statement, which is resetting the counters and triggering the recoloring, has been changed, so that it now is checking if both of these processed tiles counters is equal to the total number of tiles. Thereby the recoloring can only happen, when the tile in both maps have been processed.

The two maps would always display the same amount of tiles, since they are sharing the same view. It is therefore only necessary to calculate the amount of tiles for one of the layers.

10.7.2 Search function

In order to be able to faster navigate the map a search function was created as shown in figure 10.7.

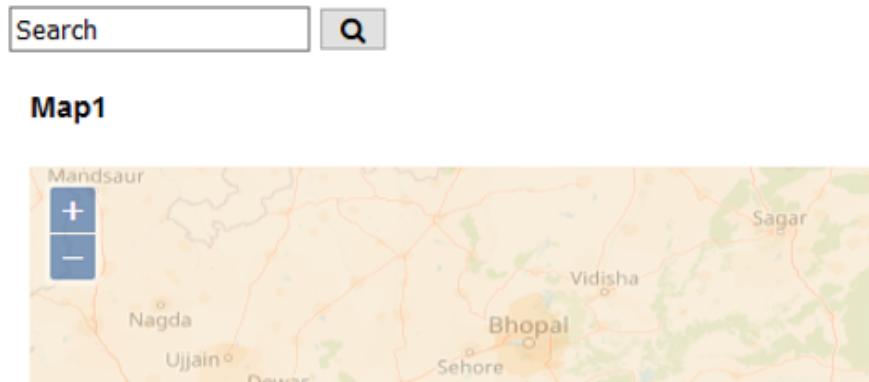


Figure 10.7: A searchbar

The user can use this to change the current view of the maps to a given location. This is accomplished through the help of Nominatim. This tool can search through Openstreetmap data by location names. It then returns data about the searched location.

Among this data is the latitude and longitude for the central point of the place. [Nominatim, No date] This coordinate is then used as the coordinate for the center of the map.

Result 11

1

[?]

¹Fixme Note: Write how the final product looks like and functions

Part III

Evaluation

Evaluation Methods

12

In this chapter the methods for evaluating the product will be presented.

Due to the limitations mentioned in section 1.2 user testing was not a possibility. This meant that it was necessary to find another way to evaluate the user experience. This was done using Google Lighthouse version 5.7. This chapter starts with an overview of the tool followed by an evaluation of which criteria are relevant for this project. The relevant criteria and their metrics are then expanded upon.

12.1 Setup

The tests have been performed in Google Chrome version 81. To ensure that tests were unaffected by outside influences all Chrome extensions were disabled during testing. The Cache

12.2 Overview

Google Lighthouse is an automated tool in Chrome DevTools, which is the developer tool included in Chrome. It can be used to check the quality of websites within the five categories: Best practice, performance, accessibility, search engine optimization and progressive web app. [Google Developers, 2020b]¹

Performance

The performance audit is measuring the site performance and load speed. This is done by measuring the time needed for different stages of loading and when the user is able to interact with the site. [Google Developers, No datec]

Accessibility

The accessibility check is for ensuring all users can effectively access and navigate the webpage. The check is mainly focused on Accessible Rich Internet Applications, which is used by assistive technologies such as screen readers.[MDN Web Docs, 2020] It is pointed out that accessibility is difficult to automatically test, so further manual testing is recommended.

[Google Developers, No datea]

¹FiXme Note: Inset Image here

Best practice

The best practices cover different types of website enhancements. There checks to ensure that the website is fast, secure, and not using deprecated technologies, among others. [?]

Search engine optimization

The checks within the Search Engine Optimization (SEO) category evaluate how well the page is optimized for ranking by search engine. This optimization is achieved by ensuring that the page is readable search engines and that search engines can access the page. It also includes some measurements for being mobile friendly. [Google Developers, No dated]

Progressive web app

The Progressive Web App (PWA) audits are target towards a mobile audience. These includes having a fast and reliable experience on mobile networks and to which degree the website act in a similar fashion to a native mobile application. [Google Developers, No dateb]

12.2.1 Relevant evaluation criteria

Not all the audit categories in Google Lighthouse are relevant for this particular project. As mentioned in section ?? the tool is developed for local use on a computer. This means that it is not relevant how optimized a site is for the phone. Since it is used locally it can not be found by a search engine, which makes the SEO category irrelevant. The best practices for enhancing the websites speed are relevant, while the security measures are not, since the tool is not accessible to others on the internet. While accessibility is important it has not been prioritised for the development of this prototype. The main argument is that the tool is being developed for data scientist and not for anybody. It is presumed that the majority of people, who daily work with computers have the sensory capacity to easily do so. A high performance is vital to ensuring a responsive user experience, so the performance audit will also be included.

2

12.3 Metrics for performance

3

4

First Contentful Paint

The First Contentful Paint (FCP) is the time in seconds it takes before the browser to render the first parts of the website. [Google Developers, 2019a]

Speed Index

²FiXme Note: Write target audience

³FiXme Note: Write about timing the python processing

⁴FiXme Note: Source: Return to this one later

The Speed Index (SI) is a time measurement of content appearing visually during load. To calculate it the visual progression between each frame of the loading process is measured. This is then being processed using the Speedline module. The unit is seconds. [Google Developers, 2019h]

The Speedline module calculates the SI by calculating an Interval Score between each frame and then summaries all the Interval Scores. The Interval Score can be calculated using the formula 12.3, where the Interval is the elapsed time since last frame and Completeness is the percentage of visual completeness. [WebpageTest.org, 2013]

$$\text{IntervalScore} = \text{Interval} * \left(1.0 - \frac{\text{Completeness}}{100} \right) \quad (12.1)$$

First Meaningful Paint

The First Meaningful Paint (FMP) is the seconds from the initial loading of the page to the primary content is visible. Where primary content is referring to the largest layout change, which is visible without scrolling.

It should be noted that FMP is being replaced with Largest Contentful Paint (LCP) in the next release of Chrome Lighthouse. The reason for this is that FMP did not give consistent result and was difficult to standardize in all web browsers. [Google Developers, 2019c] LCP is the seconds needed to render the single largest content element, which is visible without scrolling.[Google Developers, 2020a]

Time to Interactive

The Time to Interactive (TTI) is the seconds before a loaded website is completely interactive. This time is defined as after FCP when most of visible elements can be interacted with and respond within 50 milliseconds. [Google Developers, 2019k]

First CPU Idle

First CPU Idle is the seconds before a page becomes minimally interactive. Minimally interactive is defined as when the majority of the visible User Interface elements are interactive and giving responds within a reasonable time. The difference between this and TTI is the degree of interaction. When the user can begin interacting with a page the First CPU Idle can be measured. TTI is measured when all interactions can be performed. This feature is being replaced with Total Blocking Time (TBT) in the next Lighthouse release. The reason for this is that this and TTI are too similar to maintain both. [Google Developers, 2019b] TBT is the total amount of milliseconds where the website is not responding to user input. It is measured during the time between FCP and TTI. [Google Developers, 2019j]

12.4 Calculating score

Each of the metrics in the section above are then converted into a score between 0 and 100. This score is calculated by comparing each metric with performance data from real websites

in the HTTP Archive. <https://web.dev/performance-scoring/> The scores from each metrics are then weighted to reflect that each factor does not have the same importance for the perceived performance. The weight of each metric can be seen in table x.

5

12.5 Additional measurements

The measurement mentioned so far only take the initial load into consideration, not the postload performance. Therefore, two additional measurements were conducted. Both of these were measured using the Performance tool in the Google Chrome Developer tool. Since there are slight variation in the time needed to perform the requests, each operation was tested multiple time and the average was calculated.

12.5.1 Time to color new layer

This is the time in seconds required to load and color completely new sets of tiles for both maps. The scenario is created by zooming to a new zoom extent. This is done to ensure that none of the tiles on the map have been loaded before. The measured time is the time from the map is clicked, initiating the zooming, till the layer have been colored.

12.5.2 Time to recolor a loaded layer

This measurement is the time to recolor a layer with loading. This scenario is created by panning to a previously loaded part of the map, so that it is recolored. It is measured from the moment, the panning stops since the recoloring script is initiated on ended movement. The timer is stopped, when the layer is recolored.

⁵FiXme Note: Make table with score weights

Evaluation 13

This chapter is the evaluation of the script based on the Google Lighthouse audit. This chapter will highlight some elements of the audit. The full audit can be seen in appendix x. The given score in the relevant categories can be seen in figure x. There was a bug in the measurement of the performance score, which will be further expanded upon in section x. This means that the tool. It does however still provide some useful suggestions as to how the site could be improved.

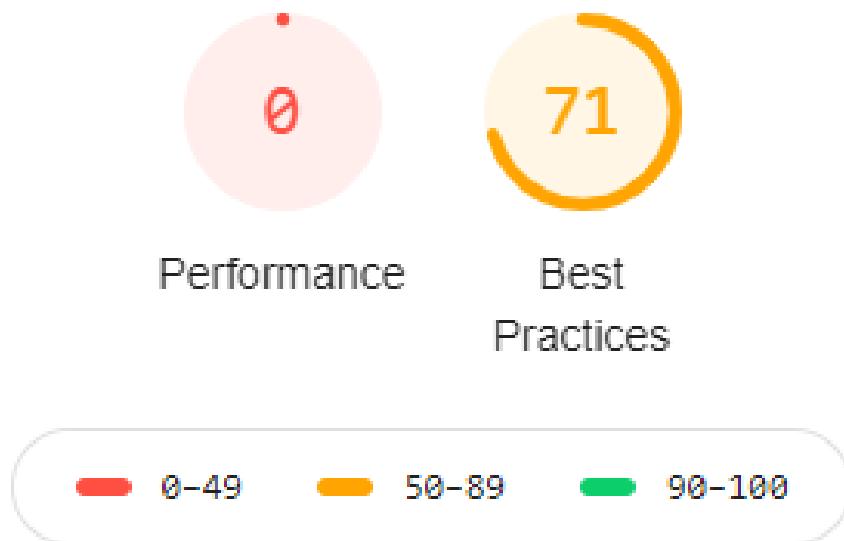


Figure 13.1: The score for performance and best practices from Google Lighthouse

13.0.1 Performance

As shown the performance scored nothing out of a hundred. The performance in each of the metrics can be seen in figure 13.2.

▲ First Contentful Paint	20.2 s	▲ First Meaningful Paint	20.2 s
▲ Speed Index	20.2 s	▲ First CPU Idle	20.2 s
▲ Time to Interactive	809.1 s	▲ Max Potential First Input Delay	330 ms

Figure 13.2: The result of the performance audit

1

It does seem like some bug in the tool must have occurred because all of the values are too high. This can be confirmed by comparing with the measurements from the developer tool's performance tool. These show that the page is fully loaded in 4.6 second, which does not fit with Lighthouse's 20.2 seconds for First Contentful and First Meaningful Paint. Another sign that something is amiss is the 13.5 minutes before the Time to Interactive, which the tool determined within 10 seconds of analyzing the site.

Even if the automatic calculation of the performance metrics failed the metrics are still relevant for the evaluation. Using the data from the performance tool these values can be manually estimated.



Figure 13.3: After 0.3 seconds the search bar and legend appears.

Figure x, x and x are all screenshots of the website taken by the performance tool, while the page was loading. Figure x is taken 0.3 seconds after the page started loading. This time is both the First Meaningful Paint and First CPU Idle since the user can interact with the searchbar.

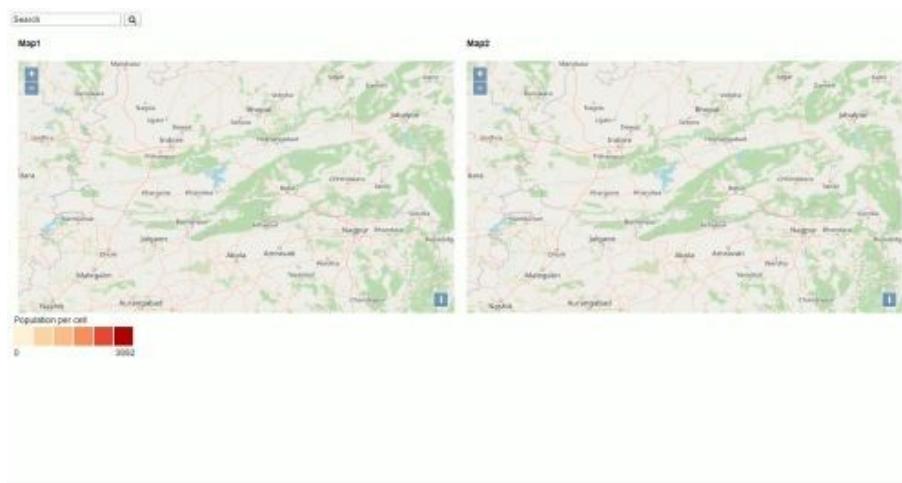


Figure 13.4: After 1 second the map appears without the raster layer

¹FiXme Note: Remove potential first input delay from figure?

After 1 second the maps appear on the website as shown in figure x. The user is able pan and zoom in the map, so this is the Time to Interactive.



Figure 13.5: After 4.6 seconds the map is fully loaded

The last screenshot at figure x shows the map fully loaded after 4.6 seconds. With the tiff tiles being loaded and visualized this is the time for the First Meaningful Paint. This can be used as a conservative estimate for the Speed Index. This would be the value for the Speed Index if everything got visualized at this time and nothing had been visualized on the website prior.

The real Speed Index would be lower, since the legend, map titles and search bar have been visualized at an earlier stage.

These metrics can then manually be calculated into a score using the Lighthouse Scoring Calculator as shown in figure x. The calculator is unable to take inputs below 1 second. Based on the given value the performance should have scored 88. <https://googlechrome.github.io/lighthouse/scorecalc/>



Figure 13.6: The website for manually calculating the scores, which Lighthouse grants

13.0.2 Opportunities

In addition to providing the metrics for the performance Lighthouse also comes with suggestion to how the loading time can be reduced.

Eliminate render-blocking resources

The opportunity for the largest estimated time saving is to eliminate render-blocking resources. These are the URLs, which must be loaded before the first paint can be applied to the page. [Google Developers, 2019i]

Table x shows the different libraries, which have to be loaded before the site is rendered. It can be seen that Openlayers is 82 % of the loaded data.³

Plugin name	Size	Unused bytes	
Openlayers	2784 KB	1188 KB	42.7 %
Geotiff.js	172 KB	100 KB	58.1 %
jQuery	87 KB	61 KB	70.7 %
Plotly	26 KB	1 KB	5.3 %
olGeoTiff	7 KB	0.1 KB	1.5 %

Table 13.1: Size of loaded modules and the much they are used

When analysed further with the Coverage tool it can be seen that a large part of the Openlayers library is not being used as show in table 13.1.⁴

5

Minification and compressing

Files sizes and script parsing time can be reduced by minifying and compressing the files.

A minified file is a file, where all the unnecessary parts have been cut away. Whitespaces and unused code are removed leaving a smaller file, which still functions perfectly.

Compression is using an algorithm to modify the data, so that it takes up less space. [Google Developers, 2019f]⁶

Avoid enormous network payloads

Long loading times are highly correlated with the amount of loaded data. [Google Developers, 2019e]

Loading the raster tiles for the map does require loading multiple files with a large bit depth, which requires a lengthy loading time.

³FiXme Note: Have I written about Jquery??

⁴FiXme Note: TODO: Make a table from the information in the figure below – tabtext – resource use of the 3 larges libraries

⁵FiXme Note: Write about Coverage earlier – maybe in evaluation tools

⁶FiXme Note: For exam: Try using Terser to minify and compress – Write about Terser in future work

Does not use HTTP/2 for all of its resources

This suggested improvement appears if some of the page's resources are being served with a version of HTTP/1. According to Google Audit all loaded resources are being delivered using HTTP 1.1.

7

[Google Developers, 2019d]

Uses deprecated APIs

When loading the metadata about the tiles a synchronous XMLHttpRequest is used. This is a deprecated API, which will be removed from Chrome in a feature edition. [Google Developers, 2019g]⁸

13.1 Postload performance

In addition to the measurements of the performance for loading the site the performance after loading was also measured. In table 13.7 is measurements of performance of using the tool. It can be seen that loading and coloring a new layer takes 3.89 seconds, whereas coloring a layer, which already have been loaded takes on average 0.56 seconds.⁹

Time to color new layer	Time to recolor loaded layer
3.92	0.6
3.85	0.4
3.98	0.5
3.82	0.8
3.88	0.5
Average	
3.89	0.56

Figure 13.7: Seconds to load a new layer and color it and to recolor a loaded layer

⁷Fixme Note: Write about HTTP/2 earlier?

⁸Fixme Note: borgerinddragelses værktøj

⁹Fixme Note: In section x is a further explanation for the larger variance in the recoloring measurements

Discussion 14

1 2

14.1 Loading tiles from the wrong zoom level

As mentioned in section x the tiles are being loaded from an incorrect zoom layer. A potential explanation for this could be the difference between what gdal2tiles and Openlayers define as zoom level 0. Openlayers defines zoom level 0 as the whole world, whereas gdal2tiles defines it as the whole raster.

It was discovered that the tile from zoom level 0 would be rendered in Openlayers, when the entirety of the raster was visible. This means that if tiles were generated for the entire world, then the zoom levels would be the same.

The tiles could possibly be ordered and named correctly by running gdal2tiles with a different profile. It was originally run with the raster profile, because the other profiles resulted in an error message as mentioned in section x.

14.2 Further development

Not all the intended functionality got added to the map. This section will list some of the features, which there was not time to implement.

14.2.1 Changing layers

Another functionality, which could have been added was the option to change the datasets of the maps. As mentioned in section x the case data contains population projection for ten different years. It does therefore make sense to have the option display more than the two datasets. As illustrated in figure x data from other years could be added as separate layers. This would allow the user to compare different datasets without having to reload the webgis with different datasets.

This way the map could also be used to visualize the different in the same projection from one year and another.

¹Fixme Note: Write about smaller maps for better performance

²Fixme Note: Update final design to not have scale

14.2.2 Option for multiple colorschemes

If there is a vast difference between the values in the two shown projection coloring based on the same maximum values does not necessarily make sense as was shown in the figure in core concept 2. It would therefore make sense to have the option to enable separate coloring as shown in figure x. TODO: make a figure showing the map with multiple legends and color schemes.

14.2.3 Additional information

Currently the user can not get exact information about the population density at a given point. They know the approximate value because of the legend, but not the exact value.

A function could be added to get the population counts from a clicked point for both maps. When the user clicks the map an infobox informs the user about the value at the clicked point for both maps.

The user could also be given addition information about the data in the current extent through histograms. In the empty space beside the maps there could be a histogram for each map, showing the distribution of the data.

14.3 Optimizing the tile generation

Currently the creation of the tiles is rather time consuming. This issue could be eliminated by either using multiple processors or a different technology.

14.3.1 Parallel generation of tiles

As mentioned in section x the generation of tiles were done without using multiple processors, which meant that it became a time-consuming process. With the official gdal2tiles being able to generate tiles following the XYZ standard it is possible that using this would allow for parallel generation of tiles.

14.4 Cloud Optimized Geotiff

An alternative to separating the rasterfile into smaller tiles would be to have the entire raster as one file, but then only send part of the file. This is possible using Cloud Optimized GeoTIFF (COG). These are GeoTiff files, which are organized in internal “tiles”. The files also contain multiple versions of the same image, where each is a downsampled version of the original. Each of these image versions can match a different zoom level.

This way of organizing the file is then combined with Range requests, which allows the client to request parts of a file instead of the whole file. <https://www.cogeo.org/in-depth.html>

By using these two technologies tiles of different spatial resolution can be sent to the client, which is similar to what the current solution is doing. The main difference is that with COG there would be one file, while the current solution has thousands of files. The creation of this file takes significantly less time compared to creating the individual tiles. India was as mentioned in section x processed into tiles in 6 hours and four minutes. Converting the same file into a COG takes 3 seconds. <https://www.cogeo.org/developers-guide.html> This way parts of only the parts of the raster, which the user can see on the map get requested.

There is currently no support for COG in Openlayers <https://github.com/openlayers/openlayers/issues/1073> and no extensions enabling it <https://github.com/cogeo/tiff/www.cogeo.org/issues/36>, but there is support for visualizing the files in Leaflet <https://github.com/GeoTIFF/georaster-layer-for-leaflet>

14.5 Performance enhancements

Based on the performance and best practice audits by Google Lighthouse the performance could be enhanced in multiple ways.

Eliminate render-blocking resources

As mentioned in section x the largest render-blocking resource was Openlayers. The same section also highlighted that 43.2 % of the library never get used. The performance could therefore be improved by only loading the necessary parts of the library.³ [Openlayers, No datea]

Avoid enormous network payloads

The size of data loads will be reduced by only having to load one set of tiles with the solution mentioned in section x.

⁴ - Tenser

⁵

Does not use HTTP/2 for all of its resources

Caddy was used to serve the resources with HTTP/2. However even after setting up Caddy the error message was still present. It is uncertain if this means that the connection still is HTTP/1 based or if it is HTTP/2 but labelled incorrectly. Figure x is two screenshots from the network tab of chrome's developer tool when the page is served with a python server (top) and caddy server (bottom). The grey bars are when a loaded file is being stalled. The blue bars are when they are loaded. This figure shows that the serving is being optimized, where the stalling largely is gone. This could indicate that the server no longer has the HTTP/1 limit of only having 6 TCP connections.⁶

³Fixme Note: Write about debug-ol

⁴Fixme Note: Write about asynchronous XMLHttpRequest and minifying/compressing

⁵Fixme Note: Add user testing to future work

⁶Fixme Note: Source!!!

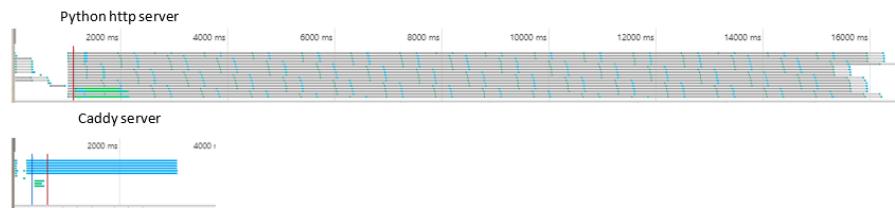


Figure 14.1: Screenshot of the network connection for a python and a caddy server

7

14.6 Not calculating minimum values in tiles

In the original version of the map both the minimum and maximum values were being calculated. However the program consistently defined the lowest value in currently visible tiles as 0. This does not mean that all the tiles had minimum values of 0, just that there always were a tile in the map extent with a value of 0. In the vast majority of case the value would be 0. Few would be below 100. In the densest of cities the minimum value for some tiles was higher than 1000. It is possible that by zoom close enough to a major city it none of the tiles within view would have a value of 0 in which case the minimum value should be adjusted. This is the main argument for calculating a minimum value. The downside of calculating it is that it has an significant effect on the processing time of the raster layer. Table 14.1 shows the time in seconds it takes for the layer to be rendered with and without calculating the minimum value. On average the rendering time is a second faster, when not calculating the minimum value.

Calculating minimum value	Not calculating minimum value
4.77	3.66
4.78	3.56
4.7	3.77
4.79	3.79
4.61	3.62
Average	
4.73	3.68

Table 14.1: Seconds to fully render the website with and without calculating the minimum value

The reason that the calculation of minimum value has such an effect on the processing time is presumably that the minimum value cannot be calculated in the same way as the maximum value. The maximum value is calculated by finding the maximum value in the array with all values.⁸ If the same operation is done to find the minimum value, the found minimum value often is -2147483648. As mentioned in section x that value means

⁷Fixme Note: Mention caddy – 6 connection issue earlier

⁸Fixme Note: Write about nodata

that no data is available for that pixel. It is therefore necessary to find the lowest positive value since no data is not the same as 0. This can be done by first filtering the raster, so only the positive values are left. The smallest of these values can then be calculated in the same way as the maximum value get calculated. It was decided that calculating a minimum value, which mostly would be 0 was not worth an additional second of loading.

Conclusion 15

In this chapter the research questions listed below will be answered. How can population rasters be visualized and compared efficiently and effectively? Which conventions exist for visualization of population projections? Which functionalities are relevant for comparing different rasters? How can a responsive user experience be ensured, when loading and visualizing large raster dataset?

The intent with this project was to create a tool for easy visual comparison of raster datasets using population projections as a case. To create this, it was necessary to understand how raster data should be visualized and which functionalities a comparison tool should have. How population raster should be visualized have been determined through a literature review of visualization conventions. Population projections are based on quantitative data, which is ordered sequentially. While many visual conventions exist for the qualitative data, the same is not the case for quantitative data. A notable convention, which is relevant for quantitative data is that the “light is less dark is more”. The functionalities needed for an interactive tool got selected by investigating a state-of-the-art interactive map. Based on this it was decided to add a search bar to enable easy navigation. To be able to compare different population projection two maps was displayed side by side. These were set up to always show the same area. The color of the projections inside was set up to automatically adjust to the population within the visible extent. A lot of other features also present in the state-of-the-art map did not get implemented due to time constraints. To ensure a responsive user experience when visualizing large raster dataset technical requirements were set up for the solution. Some of which were important for the responsive user experience. The data loaded into the map should be limited to a minimum, since loading unnecessary data would slowdown or crash the tool. To limit the loaded data the population projection was divided into raster tiles. Only the tiles visible within the map’s extent would be loaded. These tiles were then colored based on the maximum and minimum values in the current extent. This meant that the same tile would be colored in different ways dependent on what was in the current extent. This coloring was done on the client instead of on the server. This limited the number of requests to the server, since already loaded tiles could be reused. To be able to color the tiles at the client the tiles would contain information about population within each of their pixels instead of the pixels just having a color. The responsiveness of the tool also got evaluated using the Lighthouse tool. Using this tool, it was discovered that the main factor for the responsiveness was the type of test server. Changing to Caddy as a test server reduced the tile loading time from 16 to below 4 seconds.¹

¹FiXme Note: Maybe change technical concepts to technical requirements

²FiXme Note: Write how it was accomplished from a technical perspective

Bibliography

- Baumrock, 2018a.** Bernhard Baumrock. *Client-side Visualisation of Scientific RasterData Using WebGL and Open-Source WebMappingTechnologies*. <https://pdfs.semanticscholar.org/16e2/ee741969d14ba6effb9eb9686415edbb32c9.pdf>, 2018. Last Accessed: 24-05-2020.
- Baumrock, 2018b.** Bernhard Baumrock. *Visualizing GeoTIFF Tiles with OpenLayers*. <https://web.archive.org/web/20191031034339/https://eox.at/2018/01/visualizing-geotiff-tiles-with-openlayers/>, 2018. Last Accessed: 24-05-2020.
- Brewer, 1994.** Cynthia Brewer. *Guidelines for use of the perceptual*. <https://www.spiedigitallibrary.org/conference-proceedings-of-spie>, 1994. Last Accessed: 05-05-2020.
- Brewer, No data.** Cynthia Brewer. *Color Use Guidelines for Mapping and Visualization*. <http://www.personal.psu.edu/cab38/ColorSch/Schemes.html>, No data. Last Accessed: 05-05-2020.
- Brewer et al., 2013.** Cynthia Brewer, Mark Harrower, Ben Sheesley, Andy Woodruff og David Heyman. *ColorBrewer 2.0*. <https://colorbrewer2.org/#>, 2013. Last Accessed: 24-05-2020.
- Commenthol, 2020.** Commenthol. *gdal2tiles-leaflet*. <https://github.com/commenthol/gdal2tiles-leaflet>, 2020. Last Accessed: 24-05-2020.
- Data@Urban, 2018.** Data@Urban. *Using Multiprocessing to Make Python Code Faster*. https://medium.com/@urban_institute/using-multiprocessing-to-make-python-code-faster-23ea5ef996ba, 2018. Last Accessed: 27-05-2020.
- de Smith et al., 2020.** Michael J. de Smith, Michael F. Goodchild og Paul A. Longley. *Geospatial Analysis 6th Edition 2.1.6 - Fields*. <https://spatialanalysisonline.com/HTML/index.html?fields.htm>, 2020. Last Accessed: 05-05-2020.
- Dent et al., 2009.** Borden D. Dent, Jeffrey S. Torguson og Thomas W. Hodler. *Cartography - Thematic Map Design*. ISBN 978-0072943825, 2009. Last Accessed: 05-05-2020.
- Ebi et al., 2014.** Kristie L. Ebi, Tom Kram, Detlef P. van Vuuren, Brian C. O'Neill og Elmar Kriegler. *A New Toolkit for Developing Scenarios for Climate*. <https://doi.org/10.1080/00139157.2014.881692>, 2014. Last Accessed: 24-05-2020.

- Garlandini og Fabrikant, 2009.** Simone Garlandini og Sara Irina Fabrikant. *K. Stewart Hornsby et al. (Eds.): COSIT 2009, LNCS 5756, pp. 195–211, 2009. © Springer-Verlag Berlin Heidelberg 2009 Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization.* http://www.geo.uzh.ch/~sara/pubs/garlandini_fabs09.pdf, 2009. Last Accessed: 24-05-2020.
- GDAL, 2020a.** GDAL. *GDAL*. <https://gdal.org/>, 2020. Last Accessed: 24-05-2020.
- GDAL, 2020b.** GDAL. *Download*. <https://gdal.org/download.html#current-releases>, 2020. Last Accessed: 24-05-2020.
- GDAL, No date.** GDAL. *Raster drivers*. <https://gdal.org/drivers/raster/index.html>, No date. Last Accessed: 24-05-2020.
- GDAL, 2020.** GDAL. *gdal2tiles*. <https://gdal.org/programs/gdal2tiles.html>, 2020. Last Accessed: 24-05-2020.
- Google Developers, 2019a.** Google Developers. *First Contentful Paint*. <https://web.dev/first-contentful-paint/>, 2019. Last Accessed: 27-05-2020.
- Google Developers, 2019b.** Google Developers. *First CPU Idle*. <https://web.dev/first-cpu-idle/>, 2019. Last Accessed: 27-05-2020.
- Google Developers, 2019c.** Google Developers. *First Meaningful Paint*. <https://web.dev/first-meaningful-paint/>, 2019. Last Accessed: 27-05-2020.
- Google Developers, 2019d.** Google Developers. *Does not use HTTP/2 for all of its resources*. https://web.dev/uses-http2/?utm_source=lighthouse&utm_medium=devtools, 2019. Last Accessed: 25-05-2020.
- Google Developers, 2020a.** Google Developers. *Largest Contentful Paint*. <https://web.dev/lcp/>, 2020. Last Accessed: 27-05-2020.
- Google Developers, No datea.** Google Developers. *Accessibility audits*. <https://web.dev/lighthouse-accessibility/>, No date. Last Accessed: 24-05-2020.
- Google Developers, No dateb.** Google Developers. *PWA audits*. <https://web.dev/lighthouse-pwa/>, No date. Last Accessed: 24-05-2020.
- Google Developers, No datec.** Google Developers. *Performance audits*. <https://web.dev/lighthouse-performance/>, No date. Last Accessed: 24-05-2020.
- Google Developers, No dated.** Google Developers. *SEO audits*. <https://web.dev/lighthouse-seo/>, No date. Last Accessed: 24-05-2020.
- Google Developers, 2020b.** Google Developers. *Lighthouse*. <https://developers.google.com/web/tools/lighthouse/>, 2020. Last Accessed: 24-05-2020.

- Google Developers, 2019e.** Google Developers. *Avoid enormous network payloads.*
https://web.dev/total-byte-weight/?utm_source=lighthouse&utm_medium=devtools, 2019. Last Accessed: 25-05-2020.
- Google Developers, 2019f.** Google Developers. *Minify JavaScript.* https://web.dev/unminified-javascript/?utm_source=lighthouse&utm_medium=devtools, 2019. Last Accessed: 25-05-2020.
- Google Developers, 2019g.** Google Developers. *Uses deprecated APIs.*
https://web.dev/deprecations/?utm_source=lighthouse&utm_medium=devtools, 2019. Last Accessed: 25-05-2020.
- Google Developers, 2019h.** Google Developers. *Speed Index.*
https://web.dev/speed-index/?utm_source=lighthouse&utm_medium=devtools, 2019. Last Accessed: 27-05-2020.
- Google Developers, 2019i.** Google Developers. *Eliminate render-blocking resources.*
https://web.dev/render-blocking-resources/?utm_source=lighthouse&utm_medium=devtools, 2019. Last Accessed: 25-05-2020.
- Google Developers, 2019j.** Google Developers. *Total Blocking Time.*
<https://web.dev/lighthouse-total-blocking-time/>, 2019. Last Accessed: 27-05-2020.
- Google Developers, 2019k.** Google Developers. *Time to Interactive.*
<https://web.dev/interactive/>, 2019. Last Accessed: 27-05-2020.
- Herjar, 2019.** Herjar. *gdal2tiles.py multiprocessor option not working for overview tiles.*
<https://github.com/OSGeo/gdal/issues/1188>, 2019. Last Accessed: 27-05-2020.
- Jones og O'Neill, 2016.** Bryan Jones og Brian C. O'Neill. *Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways.*
<https://iopscience.iop.org/article/10.1088/1748-9326/11/8/084003/meta#references>, 2016. Last Accessed: 24-05-2020.
- Keßler og Marcotullio, 2017.** Carsten Keßler og Peter J. Marcotullio. *A Geosimulation for the Future Spatial Distribution.*
<http://carsten.io/agile-2017-short-Kessler-Marcotullio.pdf>, 2017. Last Accessed: 24-05-2020.
- Liedman, No Date.** Per Liedman. *The Hitchhacker's Guide To Tiled Maps.*
<http://www.liedman.net/tiled-maps/>, No Date. Last Accessed: 24-05-2020.
- MDN Web Docs, 2020.** MDN Web Docs. *ARIA.*
<https://developer.mozilla.org/en-US/docs/Web/Accessibility/ARIA>, 2020. Last Accessed: 24-05-2020.
- Nominatim, No date.** Nominatim. *Search queries.*
<https://nominatim.org/release-docs/develop/api/Search/>, No date. Last Accessed: 25-05-2020.

Openlayers, No Date. Openlayers. *A high-performance, feature-packed library for all your mapping needs.* <https://openlayers.org/>, No Date. Last Accessed: 24-05-2020.

Openlayers, No datea. Openlayers. *Introduction.*

<https://openlayers.org/en/latest/doc/tutorials/bundle.html>, No date. Last Accessed: 25-05-2020.

Openlayers, No dateb. Openlayers. *WMTS Layer from Capabilities.* <https://openlayers.org/en/latest/examples/wmts-layer-from-capabilities.html>, No date. Last Accessed: 25-05-2020.

Openlayers, No datec. Openlayers. *ol/tilegrid/TileGrid TileGrid.*

https://openlayers.org/en/latest/apidoc/module-ol_tilegrid_TileGrid-TileGrid.html#forEachTileCoord, No date. Last Accessed: 25-05-2020.

OpenStreetMap, 2020. OpenStreetMap. *OpenStreetMap.* www.openstreetmap.org, 2020. Last Accessed: 24-05-2020.

OpenStreetMap Wiki, 2020. OpenStreetMap Wiki. *Slippy map tilenames.*

https://wiki.openstreetmap.org/wiki/Slippy_map_tilenames, 2020. Last Accessed: 24-05-2020.

OpenStreetMap Wiki, 2019. OpenStreetMap Wiki. *TMS.*

<https://wiki.openstreetmap.org/wiki/TMS>, 2019. Last Accessed: 24-05-2020.

OSGeo Wiki, 2012. OSGeo Wiki. *Tile Map Service Specification - TileMap Diagram.*

https://wiki.osgeo.org/wiki/Tile_Map_Service_Specification#TileMap_Diagram, 2012. Last Accessed: 24-05-2020.

O'Neill et al., 2013. Brian C. O'Neill, Elmar Kriegler, Keywan Riahi, Kristie L. Ebi, Stephane Hallegatte, Timothy R. Carter, Ritu Mathur og Detlef P. van Vuuren. *A new scenario framework for climate change research:*

<https://link.springer.com/article/10.1007/s10584-013-0905-2>, 2013. Last Accessed: 24-05-2020.

Planet, No Date. Planet. *Slippy maps 101.*

<https://developers.planet.com/tutorials/slippy-maps-101/>, No Date. Last Accessed: 24-05-2020.

Rus, 2007. Jacob Rus. *File:Munsell-system.svg.*

<https://commons.wikimedia.org/wiki/File:Munsell-system.svg>, 2007. Last Accessed: 24-05-2020.

Santillan, 2020. Daniel Santillan. *plotly.* <https://github.com/santilland/plotly>, 2020. Last Accessed: 24-05-2020.

Schindler, 2018. Fabian Schindler. *geotiff.js.* <https://geotiffjs.github.io/>, 2018. Last Accessed: 24-05-2020.

Tosovsky og Strip, 2014. Jan Tosovsky og David Strip. *gdal2tiles for 16bit data.*

<http://osgeo-org.1560.x6.nabble.com/gdal-dev-gdal2tiles-for-16bit-data-td5163094.html#a5163098>, 2014. Last Accessed: 24-05-2020.

United Nations Department of Economic and Social Affairs, 2019a. United Nations Department of Economic and Social Affairs. *Total Population - Both Sexes.*

<https://population.un.org/wpp/Download/Standard/Population/>, 2019. Last Accessed: 24-05-2020.

United Nations Department of Economic and Social Affairs, 2019b. United Nations Department of Economic and Social Affairs. *World Population Prospects 2019 Highlights.*

https://population.un.org/wpp/Publications/Files/WPP2019_HIGHLIGHTS.pdf, 2019. Last Accessed: 24-05-2020.

W3Schools, No datea. W3Schools. *JavaScript Tutorial.*

<https://www.w3schools.com/js/default.asp>, No date. Last Accessed: 24-05-2020.

W3Schools, No dateb. W3Schools. *Python Introduction.*

https://www.w3schools.com/python/python_intro.asp, No date. Last Accessed: 24-05-2020.

WebpageTest.org, 2013. WebpageTest.org. *Speed Index.* <https://sites.google.com/a/webpagetest.org/docs/using-webpagetest/metrics/speed-index>, 2013.

Last Accessed: 27-05-2020.

This is at test A

A.1 Or is it?

List of Corrections

Note: ADD: quick overview of what the solution is going to be, what is population projections, SSP	2
Error: Add these three to figure -case,raster,current	2
Note: Update structure and create figure	3
Note: Rewrite this - data is not a mix of GRUMP and GlobCover, but seperated. . .	7
Note: Create own figure	8
Error: Need more context to the table - what does it mean? Do I need it? - use it to select what to show in results	9
Error: Write about domain sharding being unneeded	9
Note: what is the noData in this case - lakes?	10
Note: Ref to core concepts	15
Note: Remember Carsten mail about locally, on-the-fly and subdomains	15
Note: How long time does it take to run MakeCityWebsite	18
Note: Add text about what a convention is and scale and projection	19
Note: data is sequential	19
Note: Write about Cloud optimized Geotiff somewhere	23
Note: Write about gdal2tiles parallelly	23
Note: Write about Sarahs stuff here	23
Note: Rewrite into after sarah	23
Note: Add interactive somewhere in the research question	26
Note: Who is the target audience?	26
Note: Write that my own search bar can be improved	27
Note: Did I write about toogling dual legends?	27
Note: write about scale - update figure to exclude scale	28
Note: Update the figure to have the share box being relevant	28
Note: Write here if you come up with more relevant information, which you could throw in here	29
Note: Make figure	29
Note: Make frontpage	29

Note: Rewrite the beginning, so that it does not mention Visualisation and Functionalities	35
Note: change code so that min values also get calculated	36
Note: write about classes in visualisation	36
Note: link	42
Note: Write about the addition to time the whole thing	46
Note: Make figures - remove link	47
Note: Test if normal gdal2tiles tiles have edges - if false, then point it out here	47
Note: Mention the negative values in case data	51
Note: Write how the final product looks like and functions	57
Note: Inset Image here	61
Note: Write target audience	62
Note: Write about timing the python processing	62
Note: Source: Return to this one later	62
Note: Make table with score weights	64
Note: Remove potential first input delay from figure?	66
Note: Update to new scoring and time numbers	67
Note: Have I written about Jquery??	68
Note: TODO: Make a table from the information in the figure below – tabtext – resource use of the 3 larges libraries	68
Note: Write about Coverage earlier – maybe in evaluation tools	68
Note: For exam: Try using Terser to minify and compress – Write about Terser in future work	68
Note: Write about HTTP/2 earlier?	69
Note: borgerinddragelses værktøj	69
Note: In section x is a further explaination for the larger variance in the recoloring measurements	69
Note: Write about smaller maps for better performance	71
Note: Update final design to not have scale	71
Note: Write about debug-ol	73
Note: Write about asynchronous XMLHttpRequest and minifying/compressing	73

Note: Add user testing to future work	73
Note: Source!!!	73
Note: Mention caddy – 6 connection issue earlier	74
Note: Write about nodata	74
Note: Maybe change technical concepts to technical requirements	77
Note: Write how it was accomplished from a technical perspective	78