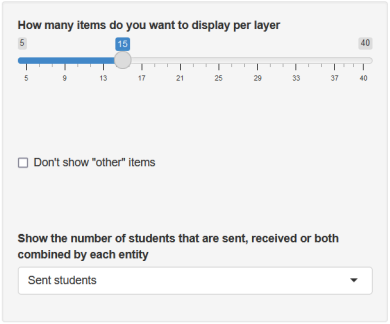
Report group 24:

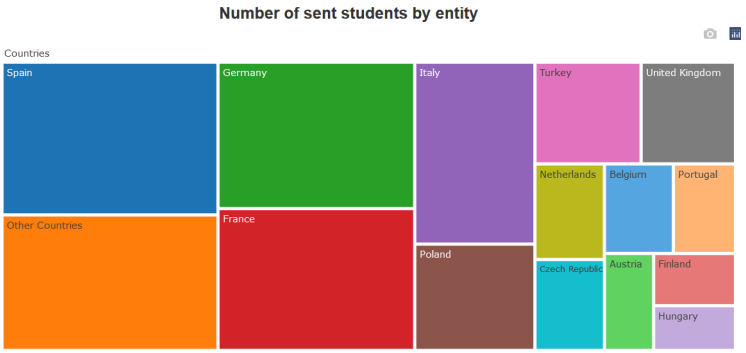
Treemap: (https://j4ohnm-tobias-engelbrecht.shinyapps.io/Treemap/)

(Tobias Engelbrecht)
Question:
How much is the Erasmus programme utilised by different geographic regions?
Meaning, how is the distribution among the countries included in the Erasmus program, as well as the distribution among the cities within each country, and finally the distribution among the universities within each city.

How to run the tool:
The tool can be run by opening the link to the application and inspecting the treemap to see the distribution of students for the displayed items. The user is able to get more information about an item (e.g. a country) by clicking on it, which displays a deeper layer of the treemap. This means that by clicking on a country the user will get information about the cities within that country and by clicking on the city the user is able to see the information about the universities within that city. Hovering over an item also shows the concrete number of students for that item.

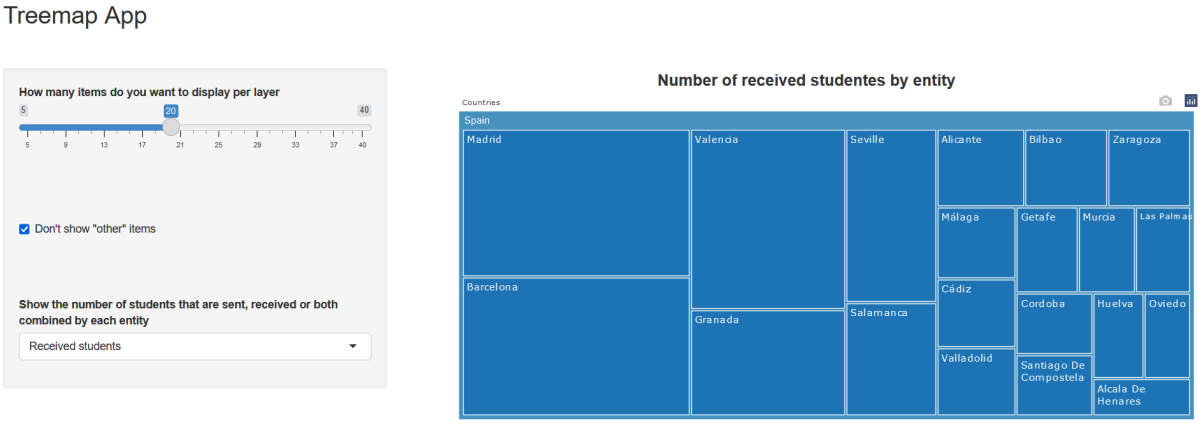
Treemap App





The left side panel offers a number of options for the user to make the visualisation more suited for the users needs.
The slider input at the top allows the user to change the number of displayed items per layer. This makes it possible to get a more complete overview by including more items or an overview with less items to make it easier to process all the information.
The check box in the middle is an option to display “other items” or to exclude them from the rendered treemap. The visualisation displays x items, where x is the number that is chosen in the aforementioned slider input. Every other item that is smaller than the x-th item is aggregated to a new “other item” and can be found by clicking on the item. E.g. the user has put up the slider to 10 and is looking up his own university “Malmö Högskolan” for fun. He knows that this university is located in Malmö in Sweden. As Sweden can not be found in the first layer of countries he clicks on “Other Countries” where he can find it. In the panel for Sweden he can not find Malmö yet so he is pressing on “Other cities”, where he can find it. After pressing on Malmö he can see that all students from Malmö are from his university “Malmö Högskolan”. While this is the wanted behaviour the “other items” take up quite a big part in the first country layer and would take up an even bigger part if the user would want to

check cities in Germany for example. Therefore the checkbox provides an option to only focus on the biggest items of each layer and completely hide smaller items. In this view it is hard to get information about these items. It is still possible though to hover over items to see the correct number of students for that item. So the number of students that are displayed in a sublayer will not add up to the number of the parent if the sublayer would have more items. To achieve a treemap representation that accurately fills the whole square extra calculations had to be made.
The last option lets the user decide whether he wants to see students that were sent to the locations (i.e., countries, cities, unis) or students that were received by these locations. It is also possible to see the combined number of sent and received students. These options might help the user to solve different questions that he might have.



What:
The dataset consists of multiple tables that are linked with different identifying attributes. The main table contains single student mobilities as one row. The original dataset is very detailed with lots of details about the single student mobilities as well as more information in other tables about the locations and more. The parts that interested us were in different tables that we had to link. The data set is very clean and does not contain empty columns or seemingly false information anywhere in it. The only data that had to be changed for us were the names of a few universities in Italy that had the exact same name (now with the city in braces afterwards) and the city Maastricht that existed twice in the data, once in the Netherlands and once in Belgium. For this visualisation we aggregate the number of students that are either sent, received or both. So the main attribute is a quantitative ordered attribute with a sequential ordering. There is a categorical attribute whether the students are sent or received, but this difference is not displayed in the same visualisation and instead there are separate visualisations for each of them.

Why: The Erasmus program aims to provide an opportunity for students from all over Europe to study abroad for a limited time. It can be important for different users with different reasons to get more knowledge about the distribution of where people are going with the Erasmus program.
A student might be interested if there are a lot of other international students at a specific place. He might simply get an overview about the country distribution to know where a lot of students are going. Or he might have already chosen one or a few countries and looks up more information about the cities within these countries. Even if he has already chosen a country and city he might look for a university that already has experience with many

international students, or contrary, for a university that has few international students to connect more easily with locals.
Otherwise a journalist might want to get an overview about the distribution as well. The journalist might want to highlight geographic regions that are not utilising the Erasmus program as much as others. And this can be on the country level or within a country and even the university level might be useful for specific stories.
In general the tool aims to help the user discover more about the given question, or simply for the enjoyment of the user as it provides a pleasing visual.

Chord Diagram:(https://j4ohnm-tobias-engelbrecht.shinyapps.io/ChordDiagram/)

(Elise Maistre)

Question:

Which country do students choose as their destination for an Erasmus course, depending on their country of origin? What proportion of incoming and outgoing students are from each country? Which countries have the highest flows?

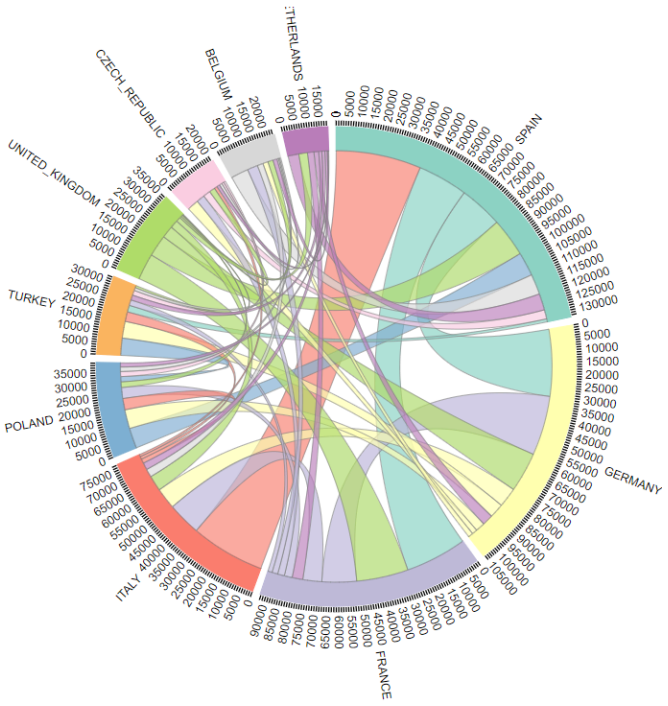
From this erasmus data and by looking at the first two columns of our dataset, which correspond to the countries of entry and exit, it is possible to deduce some information. This information is difficult to use without visualisation tools.

To answer the question about the types of flows, it is useful to use a visualisation tool. The data therefore first had to be summed before it could be used. The data was summed line by line, according to the corresponding inputs and outputs, and then grouped into a matrix as shown below.

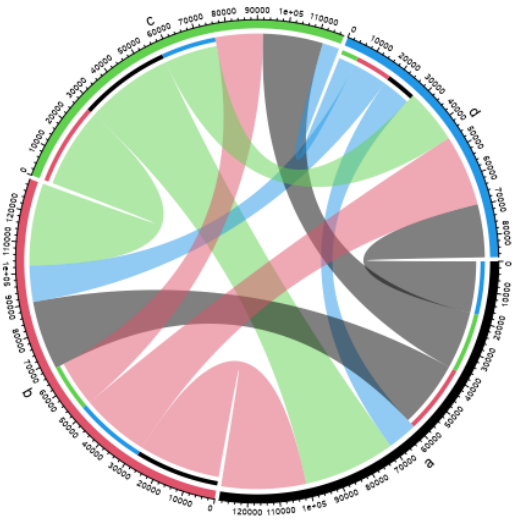
	SPAIN	GERMANY	FRANCE	ITALY	POLAND	TURKEY	UNITED_KINGDOM	CZECH_REPUBLIC	BELGIUM	NETHERLANDS	A
SPAIN	0	19883	22929	38960	10928	1410	17646	4172	9347	7487	
GERMANY	30523	0	27249	9058	4315	6570	18249	2365	2224	4959	
FRANCE	28070	15832	0	8640	3640	1973	23659	3124	2616	4858	
ITALY	33994	9729	15207	0	2032	1270	7098	944	3718	3189	
POLAND	9265	8663	5032	5168	0	2435	2534	2289	2047	1865	
TURKEY	3085	7668	2555	4368	7844	0	1015	2246	1304	2315	
UNITED_KINGDOM	9927	5561	12051	3688	346	275	0	733	788	2111	
CZECH_REPUBLIC	3360	5209	3770	1381	902	983	2589	0	1244	1370	
BELGIUM	7971	2405	4779	2631	580	646	2396	547	0	1871	
NETHERLANDS	4934	2153	2969	1400	328	1412	3941	349	1054	0	

The first line corresponds to data from Spain to all other countries receiving Spanish students. The number is the number of Spaniards in these countries. The first column corresponds to students from the countries concerned who come to Spain to study. The countries are sorted according to the number of students doing their erasmus in the country concerned.
The image above shows an extract from the data matrix, but the total number of countries involved is 33, giving a matrix of 33 x 33.
Reading such a matrix is very difficult and not easily comparable, so it is much easier to visualise it using visualisation tools. We could think of observing it with a correlation matrix, but that would still be difficult to analyse. A good way of looking at this data is to use a Chord Diagram.

A Chord Diagram is a circular object containing different data flows between different entities within it. An example of a chord diagram is shown below.



First attempt: circlize library



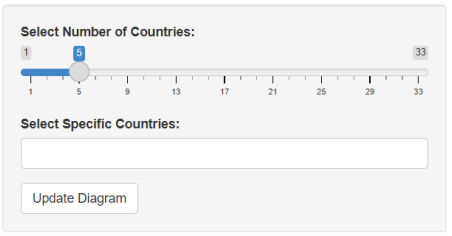
This first attempt works, but doesn't offer the desired features. With this library, it is very difficult to make the chord diagram interactive. It is also difficult to compare flows and display too many entities simultaneously. In fact, the input and output data are observed according to their proportion for each country, but their proportion is therefore doubled, once for each of the two countries concerned.

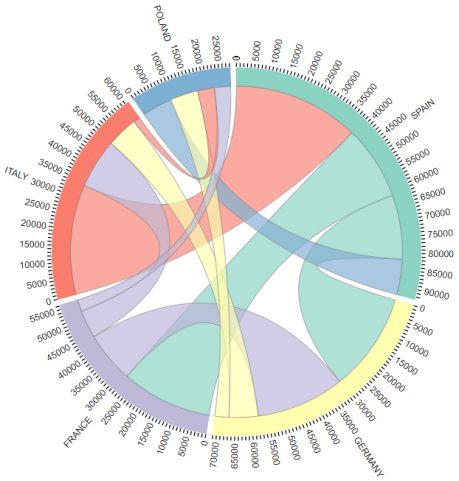
Second attempt: chorddiag library

This second attempt achieves the desired result. The proportions are not duplicated and the chord diagram is interactive.

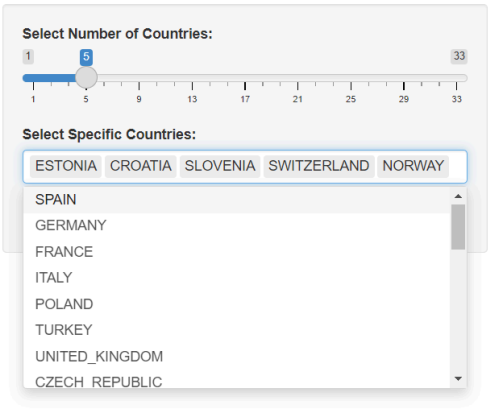
The chord diagram can also be modified by the user to display the countries of interest.

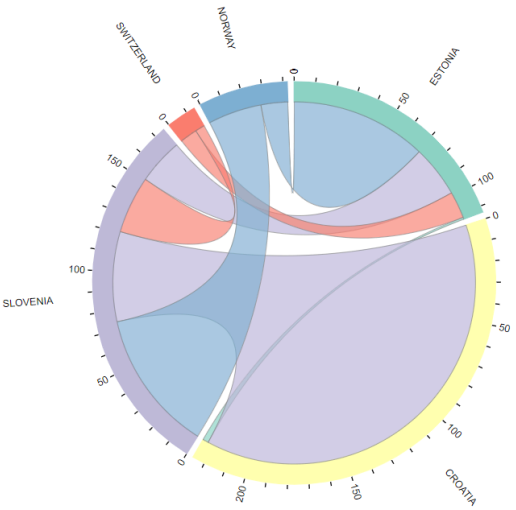
Modifications can be made in two ways. The first is by means of a slider input indicating the number of countries selected. The countries selected will be those with the most arrivals in descending order. Below is the chord diagram obtained when five of the thirty-three possible countries are selected.





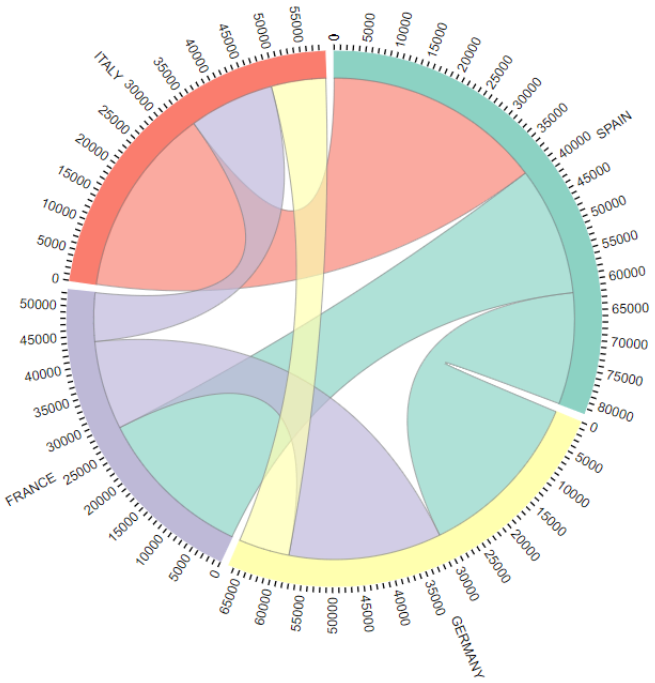
The second way to make changes is to select the countries you want to observe, using a drop-down selectable list. As shown below.



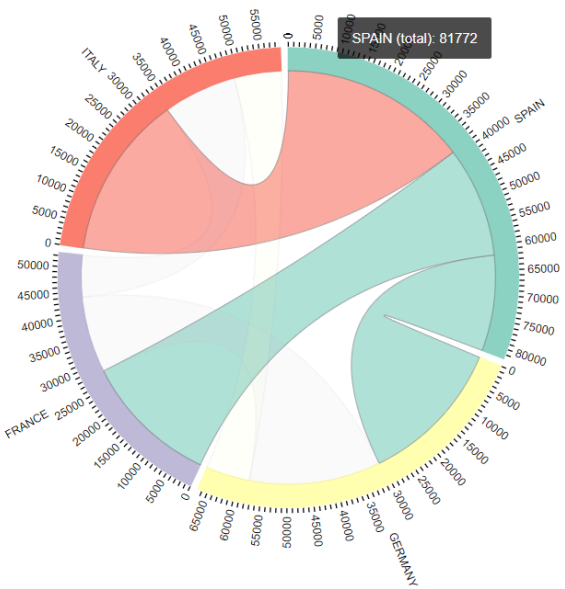


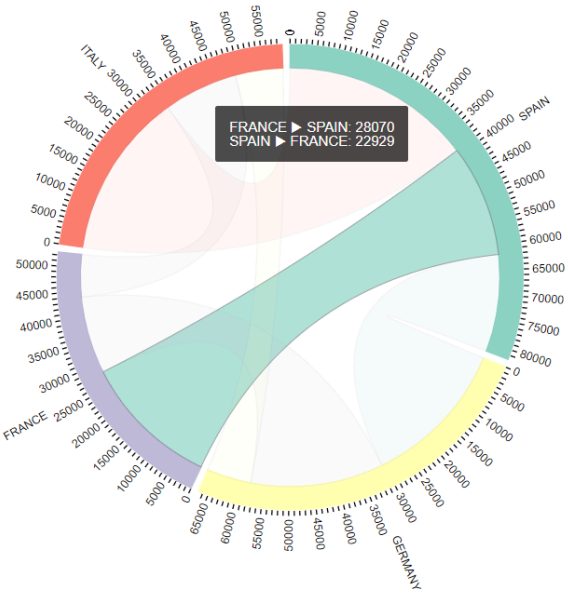
How these chord diagrams can be used to answer the questions posed.

We'll show you with the example below. This example corresponds to the first four countries receiving Erasmus students. These countries have been selected using the cursor.



First of all, we can see that the country receiving the most students is Spain, because its proportion in the diagram is greater than that of the other countries. As the diagram is interactive, it is possible to focus on just one receiving country. For example, as shown below, Spain receives a total of 81,772 students. There are also colours for flows, indicating the country sending the fewest students. It is therefore easy to see that France sends more students to Spain, as the colour of the flow between the two is the colour of Spain. What's more, to obtain the exact numbers in both directions, simply move the mouse over one of the flows and observe the result, as shown below.





This diagram answers the questions posed and makes it easier to observe the data.

Chloropleth map: (https://sachaitzko.shinyapps.io/choroplethmap/)

(Sacha Itzkowicht)

Question:

This choropleth map is designed to analyze student movement within the Erasmus exchange programme, specifically examining the directionality of student exchanges across Europe—whether students prefer to travel north, south, east, or west from their home countries.

Functionality of the Tool:

This interactive visualization is built using the Shiny framework, which allows for dynamic web applications in R. The tool is accessible via a direct link and employs Leaflet for map rendering, facilitating an intuitive exploration of student distribution data. The map utilizes color gradients, based on a quantile classification method, to represent the volume of students received by each country, providing an immediate visual distinction between different levels of student influx.

Data Presentation:

The interface is simple and intuitive, striving to present the data clearly without overwhelming the user. The colour gradient from yellow to red, using grey for European countries that do not receive students, is precisely applied, making it easy to interpret the data. The hover function is judiciously implemented to provide detailed information while maintaining the map's clean aesthetic.

Data Foundation:

The foundation of the map is a curated dataset, rigorously prepared and cleansed to ensure accurate representation. The R code demonstrates a comprehensive approach to data processing, including the use of dplyr for data summarization and sf for spatial data handling. The meticulous matching of country names between the dataset and Natural Earth data ensures geographical accuracy.

Objective and professional application:

The primary objective of the choropleth map is to serve as a reliable and informative resource for stakeholders involved or interested in the Erasmus programme. It is designed as a professional-quality tool that offers educational perspectives on student exchange models,

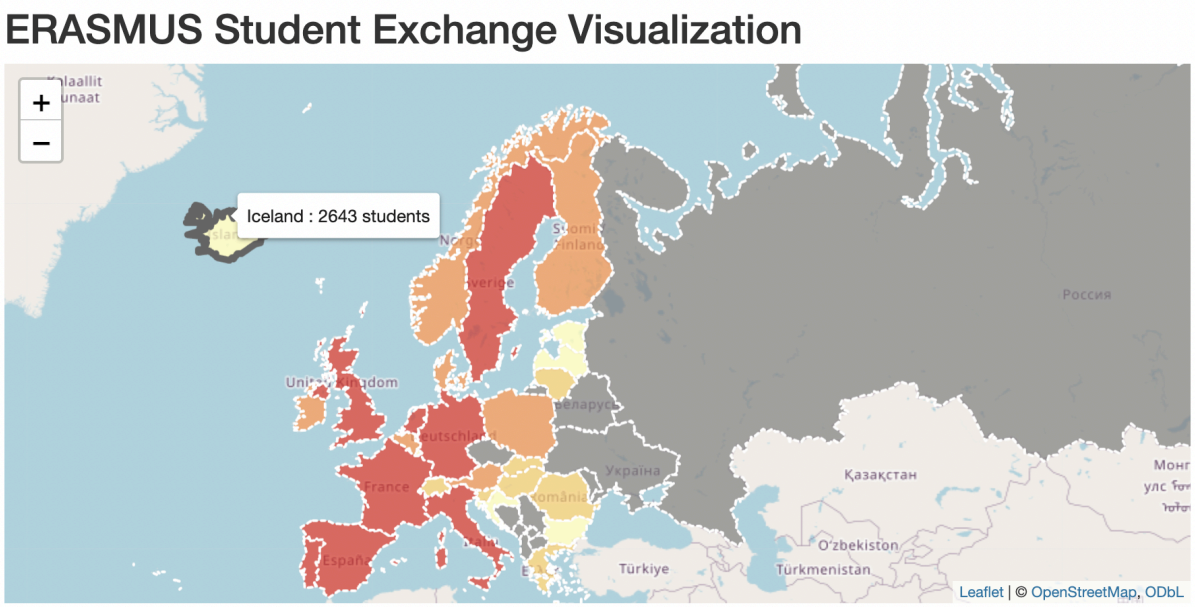
useful for both students and universities wishing to take stock of the latest trends in student exchanges in Europe.

Code Integration in Explanation:

The code underpinning this map reveals a deep integration of data science and geographic information system (GIS) techniques. It starts with the loading of necessary libraries, showcasing a multidisciplinary approach combining web development, data analysis, and spatial visualization. The preprocessing of data is handled with precision, grouping the data by receiving countries and summarizing the total number of students—a testament to the analytical rigour applied.

Normalization of country names reflects an understanding of the importance of data consistency, critical for merging datasets from different sources. The interactive elements of the map, such as the colorQuantile function for dynamic coloring and highlightOptions for enhanced user engagement, are indicative of sophisticated data visualization practices.

The application's UI is noted for its clean design, enabled by Shiny and shinyjs for JavaScript interactions, enhancing the usability and responsiveness of the map. The server function of the application brings the map to life, rendering it reactively as users interact with it, signifying a seamless integration of front-end and back-end development.



Detailed explanation of this usage example:

This capture of our choropleth map shows a colour-coded map of Europe with a particular focus on Iceland, showing that it received 2,643 students and being characterised by the colour yellow, synonymous with countries receiving a low proportion of European students.

The dark overlay indicates the selection of Iceland, and the pop-up box contains the precise figure, demonstrating the interactive functionality of the map. The overall colour scheme differentiates countries according to the number of students hosted, with a clear legend and visible zoom controls, ensuring user-friendly navigation. The base map, provided by OpenStreetMap and ODbL, provides a geographical context that contributes to the educational value of the map.

The map serves not only as a geographical index of student mobility, but also as a means of contextualising the scope of the Erasmus programme within the European educational landscape. User-centred design elements, such as the pop-up window and intuitive zoom controls, reflect a commitment to making data accessible and usable.

To conclude, the choropleth map is not just a static representation; it is a dynamic and interactive application that stands as a product of extensive data manipulation, careful design, and thoughtful implementation. It serves as a professional-grade visualization tool, enabling a comprehensive understanding of the geographical distribution of Erasmus students.