Clustering the Car Market
An Exploratory Analysis of the EU Car-Market

Nigar Salmanzade

Term Paper of the Course: Advanced Econometrics

Supervising Professor: Prof. Dr. Cristian Huse

October 15, 2024

Keywords: EU Car Market, Clustering Methods, Exploratory Data Analysis

1 Introduction

The EU car-market plays a large role in the global economy and creates a revenue of 157 billion euros annually (ACEA 2024a). With 10.5 million new cars being registered to operate on EU-roads and over 12 million cars being produced in the EU in 2023 importance of this market is immense (ACEA 2024b). It comes as a surprise that this market has not been studied as intensively as its American counterpart. Over last decades, only few studies analysed the market segmentation within the market with the focus of most studies lying on electric vehicles (EV). This paper aims at providing a first look into the segmentation of the European car-market using clustering methods. Namely, hierarchical clustering using Ward's Method and K-Means clustering will be employed to study the development of market segmentation over last decade. The basis for this analysis will be the EEA 2023 dataset which covers car registrations in all EU member states from 2012 until 2022 .

This paper aims at paving the way for future analysis on the EU car-market by providing first looks at market segmentation and their respective changes across the EU member states and across a time period from 2010 to 2022.

To do this, we begin in Section 2 by laying down a solid foundation on why this research not only matters but is a necessary first step to take in order to think about the EU car-market in a structured manner which has not yet been done for large parts of the market.

Section 3 follows up with a brief review of existing literature, which at this point in time has not been sufficient to cover the market extensively.

In Section 4 an investigation of the dataset used is performed. The EEA 2023 dataset created by the European Environment Agency which covers car registrations and characteristics in a time period from 2010 until 2022 is used in this analysis.

Next, Section 5 describes the approach taken to conduct the clustering analysis. We also apply multiple clustering methods since due to large amounts of data, hierarchical clustering is not feasible on its own. This will be covered in 5.

Afterward, the results of the analysis will be presented with a specific focus on changes in clustering over time and differences in clustering cross-country - see Section 6.

Lastly, Section 7 will provide an outlook and possibilities for future research resulting from our analysis.

2 Motivation

The car market is amidst one of largest transformations ever following the popularization of EV and new policies aimed at reducing or outright preventing CO_2 stemming from passenger cars in order to limit Climate Change and reach the agreed upon goals of the Paris Agreement. These changes in policy have increased substantially over the last decade with EV now being more accessible and affordable than ever before. Currently, EV and hybrid vehicles make up a share of 21.5% of all new cars sold in the EU (ACEA 2024a) and there has been some analysis on how the newly forming EV market is segmented. However, barely any research has been conducted into how the overall car market segmentation has evolved in this critical time frame.

How does this change to more sustainable transport affect the segmentation of the car market? How does this change differ from country to country? And what kind of market segments emerge during this time period?

All these questions have not been studied well, as most studies only focus on single countries' markets and on more limited time frames as well as excluding combustion engine vehicles from the analysis, even though they remain a popular choice for new cars purchased. Also, with the popularization of SUV-type vehicles over last one to two decades, does this segment remain as strongly supported by consumers as it currently is?

In our analysis we employ clustering methods to look at the market segmentation of the EU car market in the years 2010 to 2022 by comparing not only time variance but also cross-country differences. Our results will then provide a solid foundation for researchers and policy makers alike to base further research upon and extend the investigation into the drivers of this segmentation.

The EU car market lends itself well for this analysis as it includes a multitude of different countries with vastly different demographics. Furthermore, European car makers have in recent years struggled to match their Asian competition in sales, especially in terms of EV and Hybrid vehicles (EPRS - European Parliament Research Service 2024). Having a deeper understanding of the market segmentation can prove helpful for the industry to potentially identify new opportunities for new models of cars which will in turn be beneficial to the EU economy. With the automotive industry still being one of the most important industries in all of Europe, its importance to the overall economy is substantial: Trips by car make up about 85% of all inland journeys and in 2022 over 5.5 million cars have been exported from the EU alone generating a revenue of 157 billion euros (ACEA 2024a). On top of that, the automotive industry provides a large share of the overall employment with 6.1% (13.8 million) of all jobs being directly or indirectly connected to this industry (EPRS - European Parliament Research Service 2024). Going forward, it will be crucial for both policy makers and car makers to build a framework in which this industry can continue to generate economic opportunities for many EU citizens. In order to achieve this endeavor, it is necessary to understand how the car market developed over the years and what trends can be derived from this historical perspective. Our analysis will provide a first step into this direction by painting a clear picture of the development of market segmentation in the EU, which will show the large shift that the EU car market is undergoing.

3 Surrounding Literature

As hinted to in the previous section, the existing literature on market segmentation in the EU car-market is sparse. This section will provide a brief look into the scientific literature that does exist. However, it should be noted that this will not be an exhaustive literature review. Rather, in this section the aim is to position our analysis amongst existing literature and identify potential literature gaps to be filled by future research.

In a recent publication, Moering-Martinez, Senzeybek, and Jochem 2024 analyse the market-segmentation of the European market for electric vehicles (EV). Similar to our analysis they use the "Monitoring of CO2 emissions from passenger cars, 2022"-dataset (EEA 2023) dataset to conduct their analysis. Nonetheless, they limit their scope to only a select few European countries and more importantly just EV. This is something we extend upon by broadening the scope with more countries and vehicle types covered. Moering-Martinez, Senzeybek, and Jochem 2024 rely on both clustering methods as well as Principal Component Analysis to investigate the market segmentation in an exploratory matter. Another noticeable difference to our analysis is their clustering of countries and not vehicles with the goal being to identify front-runners (and late-adopters) in EV-adoption across the EU. Our goal is to determine how the car market changed over last decade and to uncover structural characteristics.

Similarly Kubiczek and Hadasik 2021 analyse market segmentation on the Polish EV market using hierarchical clustering methods, namely Ward's Method. This leads them to find two to four clusters of EV. Interestingly they find that market segmentation remains robust even when prices are not included in the clustering procedure. The latter proves particularly interesting for our analysis since the "Monitoring of CO2 emissions from passenger cars, 2022"-dataset (EEA 2023) dataset does in fact not include prices.
Eisenmann and Buehler 2018 cluster cars based on their owners' usage pattern. Using survey data on the usage of a car they applied hierarchical clustering methods whilst taking into account variability and intensity of the car use.
More commonly, cluster analysis is applied when investigating the behaviour of individuals. One example for this is the study by Morton, Anable, and Nelson 2017. They analyse the segmentation of car owners based on their personal attitudes toward EV and cars in general using a two-stage clustering approach which is a very similar setup top the one employed here but for cars registered. This analysis slots in as one of only a few covering market segmentation on the EU car-market and is unique in its holistic coverage of the total market without major limitations in countries and vehicles covered.

4 Data

4.1 Car Registration Data

The basis of our analysis is the "Monitoring of CO2 emissions from passenger cars, 2022"-dataset (EEA 2023). It covers all new car registrations in the EU from 2010 to 2022. Each country provides its own subset of registrations which are then combined by the European Environment Agency (EEA). This results in a large dataset with millions of registered cars all over Europe but also comes with considerable heterogeneity in data quality which complicates applying a universal methodology.
The complete dataset is split into a panel of subsets both by country and by year. This enables us to not only compare cross-country clustering differences but also time variance over the course of the 23 periods. In total, there are 367 country-year subsets. These also include Norway, Iceland, the UK and Croatia, countries that are not part of the EU (anymore) or are not covered over all periods in case of the former two, the latter does not have coverage for all periods since Croatia joined the EU in 2013 which is why they are excluded them from our analysis.
Furthermore, the subsets for Ireland in 2010 and 2011 have too many missing values which makes our Clustering procedure impossible for these two subsets. Nonetheless, Ireland, as it has been a constant member state of the EU and has data for all periods is part of the analysis. Overall, we take into consideration 336 subsets, although only a select few countries are looked at in this paper. Typically, at least one country of all European regions will be considered and presented. In the Results Section (6) we present a select few countries in the periods 2010, 2016 and 2022, which enables us to see the development clearly. Additional material will also be provided without being shown in this paper as to not inflate the content too much.

4.2 Variables

The dataset covers a multitude of different variables ranging from name and make of the car to its respective emissions or engine power. All variables included in the clustering procedure are stated in Table 1. Some of these have to altered to allow for a uniform picture. Mainly, the Fuel Type variable includes many different variations of describing the same fuel which are unified in spelling. Then, three binary variables are created - Combustion, Hybrid and Electric - such that each registered car has exactly one of these variables set to equal one whilst the others are zero. In some period-country combinations there are no new registrations of EV and hybrid vehicles such that for those periods, we have to exclude these variables from the clustering. However, it should be stated that this happens only for few subsets, mainly in 2010 or 2011 in smaller countries.
In addition, to not exclude combustion or electric cars by matter of missing values in some columns, for all combustion vehicles the Watt Hours per Kilometer variable are set to equal zero and for all EV the engine capacity is set to zero.

Table 1: Variables and Description

Variable	Description
r	Number of Car Registrations
m (kg)	Mass in Running Order in kg
AT1 (mm)	Axle Width of the Steering Axle in mm
AT2 (mm)	Secondary Axle Width in mm
ec (cm3)	Engine Capacity in cubic cm (Combustion Only)
ep (KW)	Engine Power in Kilowatts
z (Wh/km)	Watt Hours per Kilometer (Electric Only)
W (mm)	Wheelbase in mm
Ft	Fuel Type

5 Clustering Method

5.1 Hierarchical Clustering

The clustering method of our choice is hierarchical clustering using Ward's Method. This method is an agglomerative clustering method (Murtagh and Contreras 2012), meaning each observation starts as its own cluster. Merging then happens pair-wise until every observation is contained within one cluster. The procedure can visualized nicely using dendrograms which offer a first way for graphical inspection and give direction to a potential choice of cluster count. We augment this by using silhouette scores and scree plots to find an appropriate number of clusters. Furthermore, Ward's Method minimizes within-group variance resulting in compact, easy to interpret clusters. (Murtagh and Legendre 2014).

Other hierarchical clustering methods like Single-, Average- and Complete-Linkage Clustering did not yield promising results. With clustering being an exploratory method, trial and error is part of the analysis. Whereas the Ward's Method produced well distinguished groupings, the Single Linkage only produced a uniform "mass" which did not hold much interpretability. The main drawback of this method is its high computational complexity which for us facilitates the need to augment it by also applying K-Means clustering in a two-stage setup. In this, a limit on observations is set to use before drawing a random sample and performing hierarchical clustering on this sample. In our case the threshold is 10,000 observations. We regularly ran into memory issues when exceeding this number of observations. It has to be stated, that hierarchical clustering in our analysis serves mainly as a means to determine a suitable number of clusters.

5.2 K-Means Clustering

As laid out in the previous section, the main drawback of using hierarchical clustering is its reliance on a fully calculated distance matrix. This, due to its high computational complexity of $O(n^3)$, leads to rapid growth in calculation effort with rising number of observations (Murtagh and Contreras 2012). So much so, that it is infeasible for us to calculate some distance matrices. Examples would be the German datasets which can consist of more than two houndred thousand observations.

Since we deem these large markets as highly important to our analysis, we employ a second clustering approach, better suited to handle larger data sets. K-means clustering is one such method which helps us to analyze also the large countries' datasets. Other than hierarchical clustering, it does not require a full distance matrix, which reduces complexity to $O(nkl)$ (with n observations, k clusters and l iterations) but it also has some drawbacks:

First and foremost, it requires a pre-specified number of clusters in which the data is supposed to be split into. The choice of this number is not straightforward. Choosing a number of clusters too low will lead to poor differentiation between clusters and this will fail to capture key differences between market segments. Choosing to many clusters will lead clusters to capture features too specific to derive meaning from (similar to overfitting in Regression Analysis) (Ikotun et al. 2023).

Multiple tools are employed to define a suitable number of clusters *a priori*. First, we use our hierarchical clustering approach on a random subsample (n = 10,000) and calculate silhouette scores and inspect resulting scree plots to define a suitable number of clusters by applying the "Elbow-Method". Second, we inspect the dendrograms of the hierarchical clustering procedure to

provide a better foundation for the optimal cluster count. In some cases the scree plot does not display a clear "elbow-like" shape or display two or more possible "elbows". It is then, when we emphasize the dendrogram more. In some fringe cases it is also necessary to put the choice of cluster numbers into context with previous and later periods to make the correct choice. It should be mentioned that this is one of the more ambigous aspects of cluster analysis and has room for interpretation. This is tackled by providing examples of our reasoning and choice of cluster counts in Section 6, acknowledging the subjectivity of this matter.
A second problem is its reliance on random initialization of cluster centroids. Tackling this is done by running multiple iterations of the code in order to prove the validity of the results. Furthermore, it prevents accepting results that converged on local minima (Ikotun et al. 2023).
Graphical interpretability also suffers when using K-Means clustering since it is not an agglomerative procedure and thus no dendrogram can be presented.
Overall, our approach is akin to the two-stage clustering approach used by Morton, Anable, and Nelson 2017 with a some noticeable differences in choosing the number of clusters and initialization of cluster centroids in the K-Means Clustering.

5.3 Example Case - Sweden

To provide a better overview of our method, we select one case for which we will elaborate the procedure step by step. The choice of the Swedish subsets is made deliberately since the subsets have a mix of well-defined, clear to identify cases as well as some periods in which the there is some ambiguity. Especially in the choice of cluster numbers, it will become apparent that the "Elbow-Method" is not suitable on its own.

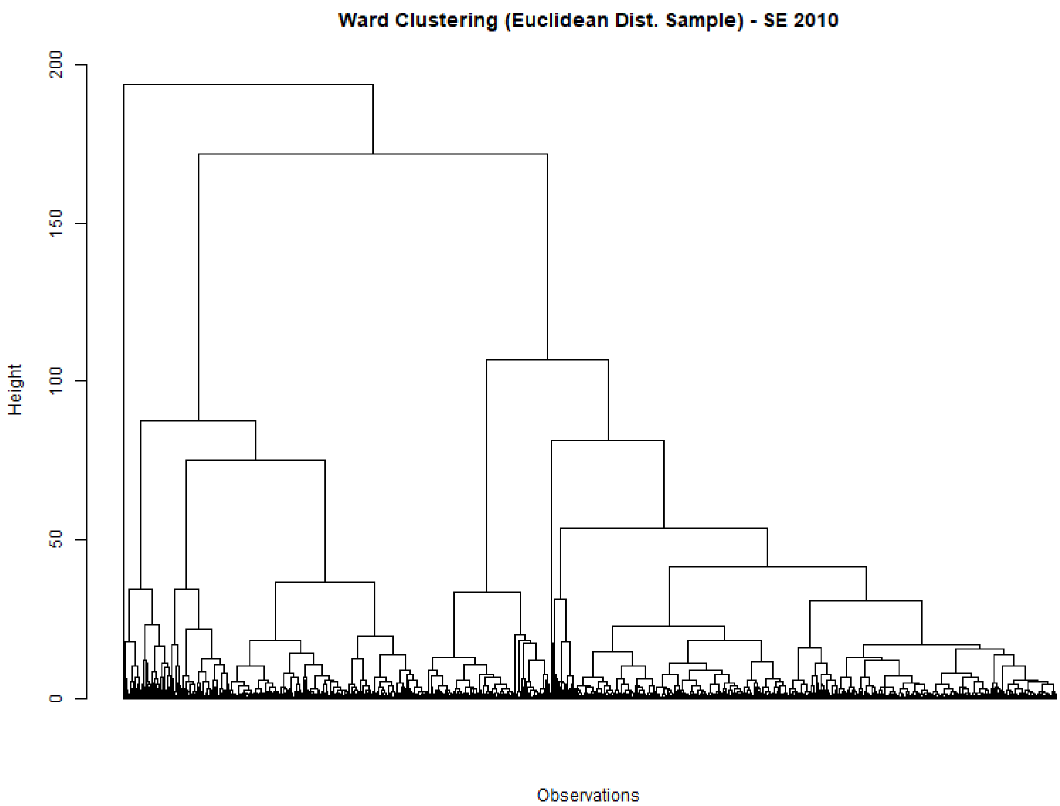

Figure 1: Dendrogram of the Swedish Subset 2010

Figure 1 shows the dendrogram of the Swedish data subset for the year 2010. Based solely on this, one would be able to deduct a proper cluster count of three: One cluster to the far left of the graph which is comparatively small in observations included and two broader clusters in the middle and the right of the graph which both include more observations. As we do not want to rely solely on this method alone, the Scree Plot is also investigated, which displays the silhouette width or scores on the ordinate and the number of clusters on the abcissa. By applying the "Elbow-Method" to Figure 2 it can clearly be seen that a cluster count of three seems to be appropriate for this subset which strengthens our initial thought based on the dendrogram.

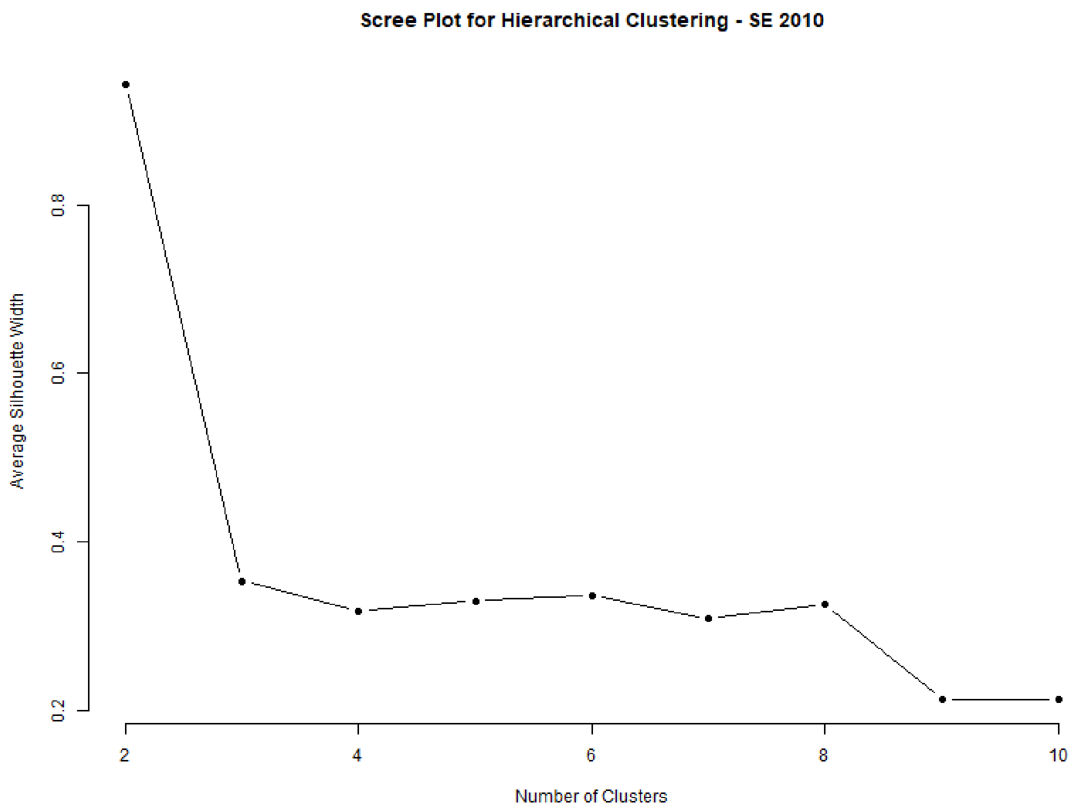

Figure 2: Scree Plot of the Swedish Subset 2010

On the contrary, whilst the subset for 2022 still yields an interpretable dendrogram (see Figure 3), the "Elbow-Method" cannot be used on the associated dendrogram. Whilst Figure 4 might display an "Elbow-like" structure at a cluster count of nine, this does not match the dendrogram, in which a cluster number of four seems reasonable and it also does not match with the overall structure. Almost all cluster counts are in a range from three to six clusters, thus choosing nine would create an obvious outlier which is not backed by the associated dendrogram.
Overall, we see that not only do the dendrogram and the scree plot matter but especially if the latter fails to produce a reasonable "Elbow", the context of previous and following periods as well as structure of comparable countries cluster counts matter.

After choosing an appropriate number of clusters, we continue with K-Means clustering using said number. Initialization of cluster centroids is done randomly, which proved to be stable over multiple iterations of the code. Another way of initialization could be to calculate cluster centroids based on the hierarchical clustering, similar to what Morton, Anable, and Nelson 2017 do. Given that our results are stable, the need for this does not seem to be particularly high, although analysing the difference this makes, might be a future extension for this application. With the K-Means clustering having been performed, we then move to creating summaries of each cluster which will provide crucial information about the market segments and the characteristics of the cars inside each segment.

6 Results

6.1 Results Sweden

Table 2 shows the cluster summary for Sweden in 2010. The three clusters show some notable difference from each other: Cluster 1's average car can be described as a rather large vehicle with

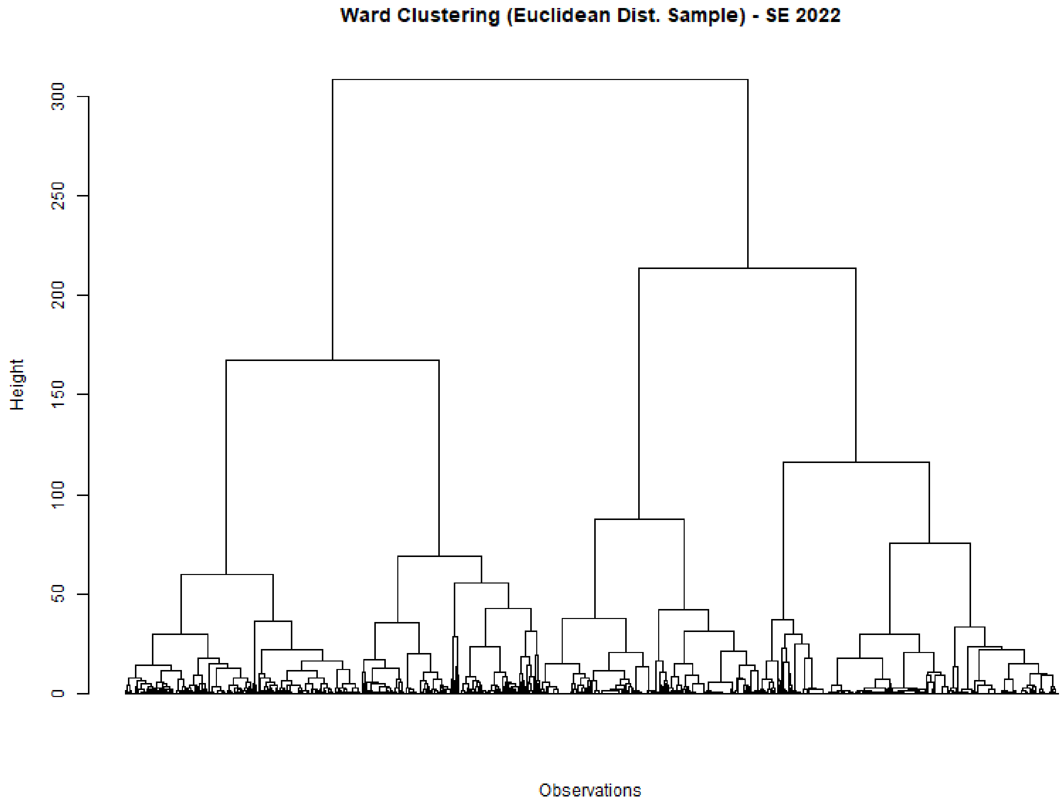

Figure 3: Scree Plot of the Swedish Subset 2022

a large engine. Overall, this cluster consists of 1030 observations (n) with an average number of registrations per observation of r = 14.09, making it a smaller cluster in terms of cars covered. Cluster 2 includes all EV that are present in this period, however these amount to only a small share. Primarily, this cluster can be seen as a small-car cluster. Cluster 3 on the other hand consist of only combustion engine cars with larger measurements but medium sized engines on average. Assigning labels to these clusters could look the following: Cluster 1 - High Powered Large Cars, Cluster 2 - Hatchbacks and Small Cars, Cluster 3 - Station Wagons/Family Cars.

Period 2016's cluster summary can be seen in Table 3. The associated dendrogram (Figure 5) and scree plot (Figure 6) can be found in the Appendix (Section 8)

Compared to that, Table 4 shows the cluster summary for 2022. Not only can an additional cluster be observed as compared to 2010, we also see vastly different characteristics within each cluster.

Cluster 1 now displays a cluster of only EV, with typically high mass and rather high engine power. This is a large difference to previous periods. Additionally, this cluster is the largest of the four clusters with a total number of 86,260 (n = 86,260; note that for the periods after 2016 the variable r ceases to have meaning. This is due to a change in record-keeping from the EEA's side. The total number of observations can still be gauged properly). Cluster 2 can be seen as the new hybrid cluster with large cars with comparatively powerful larger engines. Cluster 3 consists of mostly combustion engine cars. On average these are comparatively small in size and in engine capacity and power. A proper label for this cluster might be City Car/Small Car. In terms of vehicle size cluster 4 is similar to cluster 2. This last cluster consists of only combustion engine vehicles with moderate engine capacity and power but with large wheelbases. This likely reflects typical family sedans or station wagons. Cluster 4 makes up the largest share of vehicles with n = 86,260 registrations covered. The other three clusters show comparable cluster sizes with 64,833 for cluster 2, 67,794 for cluster 3 and 52,779 for cluster 4. The contrast from this clustering to the displayed for the 2010 period is large and especially the emergence of EV and hybrid vehicles is something that is clearly noticeable.

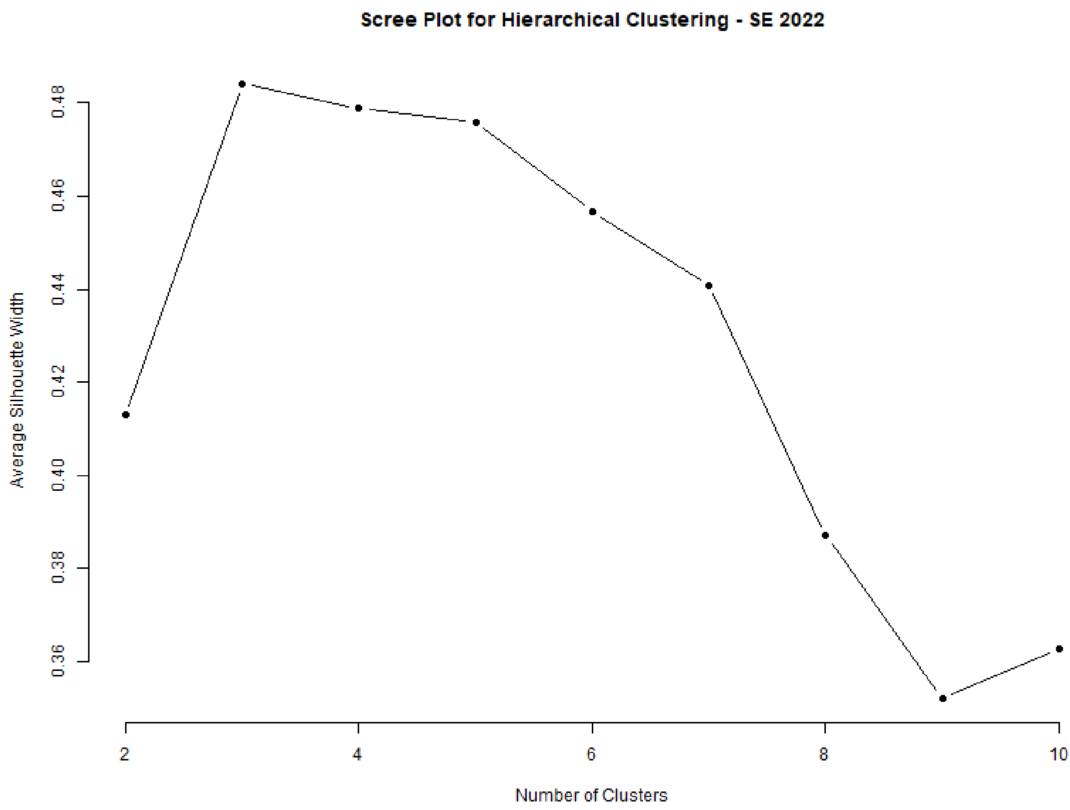

Figure 4: Scree Plot of the Swedish Subset 2022

6.2 Results Germany

The German subsets and their respective clustering paint a similar picture: For 2010 the German clustering (Table 5) looks remarkably similar to the Swedish counterpart. A cluster made up of small cars and two clusters with larger vehicles characterise this procedure. Furthermore, for both counties these large vehicle clusters are split into one with modestly high engine capacity (Cluster 2) and one with higher engine capacity values (Cluster 3). Also, in 2010 EV are not very prevalent in either country.

In 2016, the clustering results are still remarkably similar (Table 6): The Swedish subset still displays three mainly combustion-engine clusters, the German subsets mirrors this but with lower shares in EV and hybrid vehicles inside the clusters. Both countries show three clusters that could be labelled as small, medium and large vehicle clusters. Furthermore, the small and medium clusters make up a larger share of registrations as compared to its respective large car cluster.

2022 is an outlier in our analysis with a cluster count assumed to be six (see Figures 11 and 12). The summary, found in Table 7, shows that each cluster is mostly made up of one type of engine. Cluster 1 includes smaller EV with lower engine power and smaller wheelbase and axle width. Likely, these vehicles are EV for urban traffic which are not used for larger distances. On the other hand, cluster 2 contains only larger EV. With an average mass of around 2,000kg and a larger wheelbase this cluster likely contains most Tesla vehicles, since they tend to be larger in terms (Tesla 2024). Cluster 3 consists of small combustion engine only cars with small, low horsepower engine. These could be classified as the typical "city car". The largest cluster is cluster 4 with more than 800,000 registrations and it contains only combustion engine vehicles with modest characteristics. With modest engine power and smaller wheelbase and axle width this likely contains most sedans and station wagons and more compact SUV. Thus, this cluster could be seen as the family car cluster. Cluster 5 can be seen as a large vehicle cluster (SUV) with mostly combustion engines and high engine power with an average of 222.68KW. Lastly, cluster 6 solely consists of hybrid vehicles which are a bit smaller than cluster 2's cars.

Table 2: Cluster Summary Sweden 2010

Cluster	1	2	3
r	14.09	70.47	49.89
m (kg)	1984.61	1164.00	1567.70
W (mm)	2985.90	2457.12	2703.36
At1 (mm)	1610.82	1454.59	1542.78
At2 (mm)	1612.41	1452.55	1542.77
ec (cm3)	3251.19	1385.06	1991.22
z (Wh/km)	0	1.30	0
Combustion Share	1	99.30%	1
Electric Share	0	0.70%	0
Emission Standard	233.41	129.61	165.07
n	1030	857	3973

Table 3: Cluster Summary Sweden 2016

Cluster	1	2	3
r	49.69	11.05	24.75
m (kg)	1283.08	2032.68	1581.39
W (mm)	2580.15	3004.63	2755.90
At1 (mm)	1510.16	1634.03	1568.04
At2 (mm)	1499.03	1638.86	1566.73
ec (cm3)	1396.16	2842.40	1912.88
ep (KW)	87.43	226.21	130.06
z (Wh/km)	5.50	11.88	0
Combustion Share	95.70%	93.58%	1
Electric Share	1.31%	4.04%	0
Hybrid Share	2.99%	2.38%	0
Emission Standard	110.16	174.62	131.98
n	3209	2229	7140

6.3 Results Bulgaria

So far, only northern and central Europe subsets have been analysed, which leads us to our last region - Bulgaria. Bulgaria being in the South-East of Europe enables us to get a broader view and the differences of clustering within the EU.
Table 8 shows the cluster summary for 2010 in Bulgaria. Here, the clusters are vastly different to the two previous countries. The cars registered in Bulgaria are, on average, lighter by a sizeable amount. Further, the clusters themselves are not as heterogeneous as those seen for Sweden and Germany in 2010. It appears as if there are a small, medium and large car cluster. However, this large car cluster is still much lighter in terms of mass and less powerful than its counterpart in the German or Swedish subset. One explanation for this might be a higher popularity of smaller cars which drives up registration numbers. Whereas in Germany and Sweden you will typically see SUV-type vehicles, Bulgaria seems to have a higher concentration of small sedans.

Interestingly, in 2016 (see Table 9) there are still no registrations of EV in Bulgaria and just a small number of hybrid vehicle registrations. But, the cluster summary changed in some notable ways: First, the cluster summary now is more akin to that of Sweden and Germany in 2010 with higher vehicle mass on average and similar shares of fuel types within each cluster. Second, the difference between clusters in 2016 is much more pronounced than it was in 2010, making the classification more robust. Third and not unimportantly, the number of observations within this subset of data is considerably larger and provides a better basis for the clustering procedure. We cannot exclude the possibility of 2010's results being driven by a small sample size.

In 2022 (see Table 10), the assumption of three clusters still seems optimal. The dendrogram resembles strongly the hierarchical clustering of Sweden in 2016 (compare Figure 5 and 17). One

Table 4: Cluster Summary Sweden 2022

Cluster	1	2	3	4
m (kg)	1969.54	1962.60	1339.05	1808.97
W (mm)	2764.28	2791.26	2609.23	2825.62
At1 (mm)	1586.54	1608.57	1534.40	1612.58
At2 (mm)	1584.57	1612.41	1529.58	1616.74
ec (cm3)	0	1872.14	1384.01	2080.48
ep (KW)	193.02	142.24	88.22	153.34
z (Wh/km)	175.92	177.67	1.24	0
Combustion Share	0	0	99.13%	1
Electric Share	1	0	0.87%	0
Hybrid Share	0	1	0	0
Emission Standard	0	30.20	124.89	166.95
n	86260	64833	67794	52779

Table 5: Cluster Summary Germany 2010

Cluster	1	2	3
r	97.10	396.11	35.93
m (kg)	1589.43	1211.02	2184.23
W (mm)	2704.83	2501.18	3125.74
At1 (mm)	1550.91	1463.55	1645.54
At2 (mm)	1550.23	1461.68	1645.99
ec (cm3)	2012.28	1409.40	3011.18
z (Wh/km)	0	0.11	0
Combustion Share	1	99.92%	1
Electric Share	0	0.08%	0
Emission Standard	167.59	134.65	241.40
n	12045	3794	5438

explanation for this could be that Bulgaria is lacking behind in the adoption of EV and hybrid vehicles compared to Sweden and Germany. Bulgaria being less developed than both of these countries might give an explanation as to why this phenomenon occurs. Still, this would need to be analysed further in order to gain solid evidence for our hypothesis. It cannot go unmentioned however that there now is a cluster made up of only EV and hybrid vehicles which is smaller in number of registrations as compared to the other two clusters (n = 1,631 as compared to 22,810 for the more modest specification car cluster 1 and 3,450 for cluster 3 which covers more powerful larger combustion engine cars).

These are just some of the subsets of data on which we performed the analysis. Overall the picture this analysis paints is clear. Countries in Central, Northern and Western Europe tend to display an earlier adoption of EV and hybrid technologies when compared to their Southern or Eastern European counterparts. The emergence of EV and hybrid vehicles typically leads to an increased number of clusters. Usually, this shows itself in the form of one to two EV and/or hybrid clusters.

7 Conclusion and Outlook

As hinted to in the previous section, the amount of research that can build upon our exploratory approach is immense. Taking a deeper look into what the drivers of the segmentation are and how economic development plays a role in this, is only one possible step. Additionally, countries that have yet to undergo the transformation to more sustainable modes of transport might be able to derive proper strategies for the implementation of such technologies when looking at our project and connecting it to infrastructure availability. Especially for EV this is bound to matter significantly.
What we showed is, that the market segmentation inside the EU member states' car-markets underwent and is still undergoing massive changes. Especially with the emergence of EV and

Table 6: Cluster Summary Germany 2016

Cluster	1	2	3
r	172.27	48.17	26.29
m (kg)	1191.67	1522.22	2036.06
W (mm)	2506.36	2730.52	3077.06
At1 (mm)	1486.80	1564.74	1642.67
At2 (mm)	1480.86	1560.82	1644.91
ec (cm3)	1295.05	1787.85	2659.25
ep (KW)	76.80	118.15	182.90
z (Wh/km)	1.88	0	2.59
Combustion Share	98.69%	1	98.60%
Electric Share	0.97%	0	0.99%
Hybrid Share	0.34%	0	0.41%
Emission Standard	111.31	128.46	177.29
n	8162	31441	13697

Table 7: Cluster Summary Germany 2022

Cluster	1	2	3	4	5	6
m (kg)	1370.66	2006.80	1231.29	1585.47	2057.36	1906.03
W (mm)	2421.67	2822.51	2556.48	2729.04	2966.54	2756.91
At1 (mm)	1470.83	1587.35	1499.52	1569.24	1637.72	1585.56
At2 (mm)	1467.47	1596.80	1509.69	1579.92	1645.81	1595.77
ec (cm3)	0	0	1240.31	1773.20	2668.69	1798.28
ep (KW)	77.33	183.65	81.30	126.34	222.68	126.57
z (Wh/km)	155.76	170.60	0	0	0.41	175.89
Combustion Share	0	0	1	1	99.79%	0
Electric Share	1	1	0	0	0	0
Hybrid Share	0	0	0	0	0.21%	1
Emission Standard	0	0	125.69	149.48	211.47	33.06
n	135906	287421	643567	826189	255893	351782

hybrid vehicles the segmentation had a lasting change. Whether or not the new segmentation will soon resemble the previous market segmentation when mostly combustion engine cars were present is unclear. It seems that, overall cars have tended to get larger and heavier the first of which is not a trend that is in line with more sustainability. The latter is likely unavoidable in the near future due to the battery weight. It is clearly visible that there are large differences in clustering and development of clustering across European countries making it difficult to paint a complete picture for the EU. Our recommendation is to divide Europe into subsections or investigate member countries individually in order to grasp potential differences properly.

8 Appendix

Table 8: Cluster Summary Bulgaria 2010

Cluster	1	2	3
r	10.12	3.39	5.56
m (kg)	1166.92	1661.41	1355.21
W (mm)	2566.17	2736.11	2704.68
At1 (mm)	1462.23	1568.61	1523.89
At2 (mm)	1450.87	1571.46	1528.96
ec (cm3)	1415.32	2144.41	1619.24
Emission Standard	138.82	183.95	146.57
n	60	44	72

Table 9: Cluster Summary Bulgaria 2016

Cluster	1	2	3
r	9.00	2.58	3.79
m (kg)	1197.81	2079.49	1530.27
W (mm)	2560.15	3007.86	2717.93
At1 (mm)	1492.65	1637.17	1564.93
At2 (mm)	1485.28	1647.05	1565.10
ec (cm3)	1348.45	3047.10	1788.72
ep (KW)	73.96	223.99	114.66
z (Wh/km)	0	0	0.58
Combustion Share	1	1	99.55%
Hybrid Share	0	0	0.45%
Emission Standard	113.60	185.77	129.69
n	1171	644	2014

Table 10: Cluster Summary Bulgaria 2022

Cluster	1	2	3
m (kg)	1381.03	1880.93	2152.25
W (mm)	2648.58	2732.46	2981.69
At1 (mm)	1547.34	1571.21	1653.14
At2 (mm)	1547.10	1570.76	1663.96
ec (cm3)	1460.35	867.96	2794.22
ep (KW)	94.25	156.03	230.45
z (Wh/km)	0	176.92	0
Combustion Share	1	0	1
Electric Share	0	55.67%	0
Hybrid Share	0	44.33%	0
Emission Standard	131.77	17.07	216.85
n	22810	1631	3450

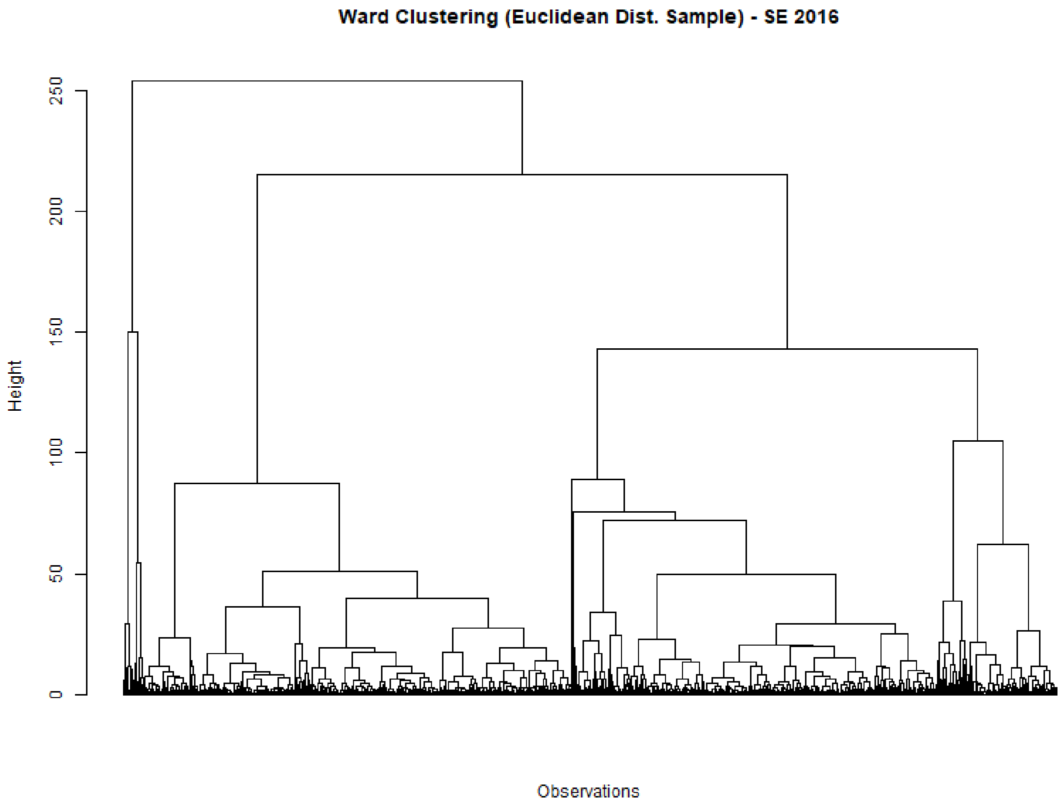

Figure 5: Dendrogram of the Swedish Subset 2016

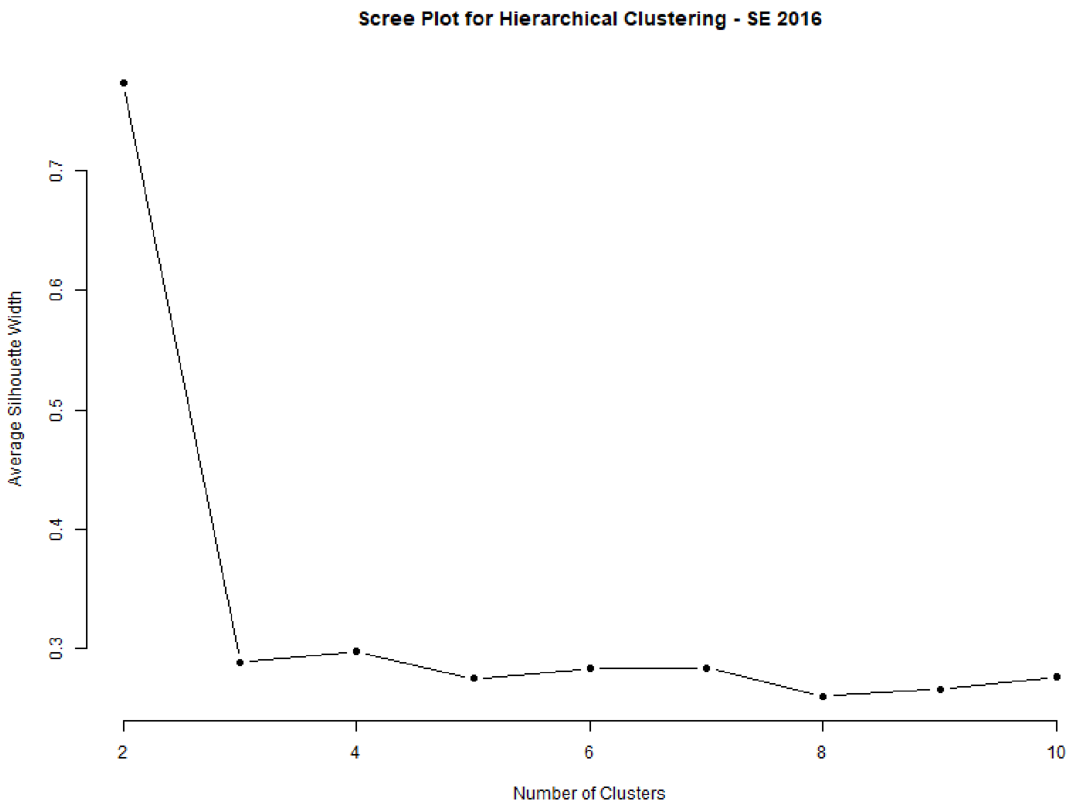

Figure 6: Scree Plot of the Swedish Subset 2016

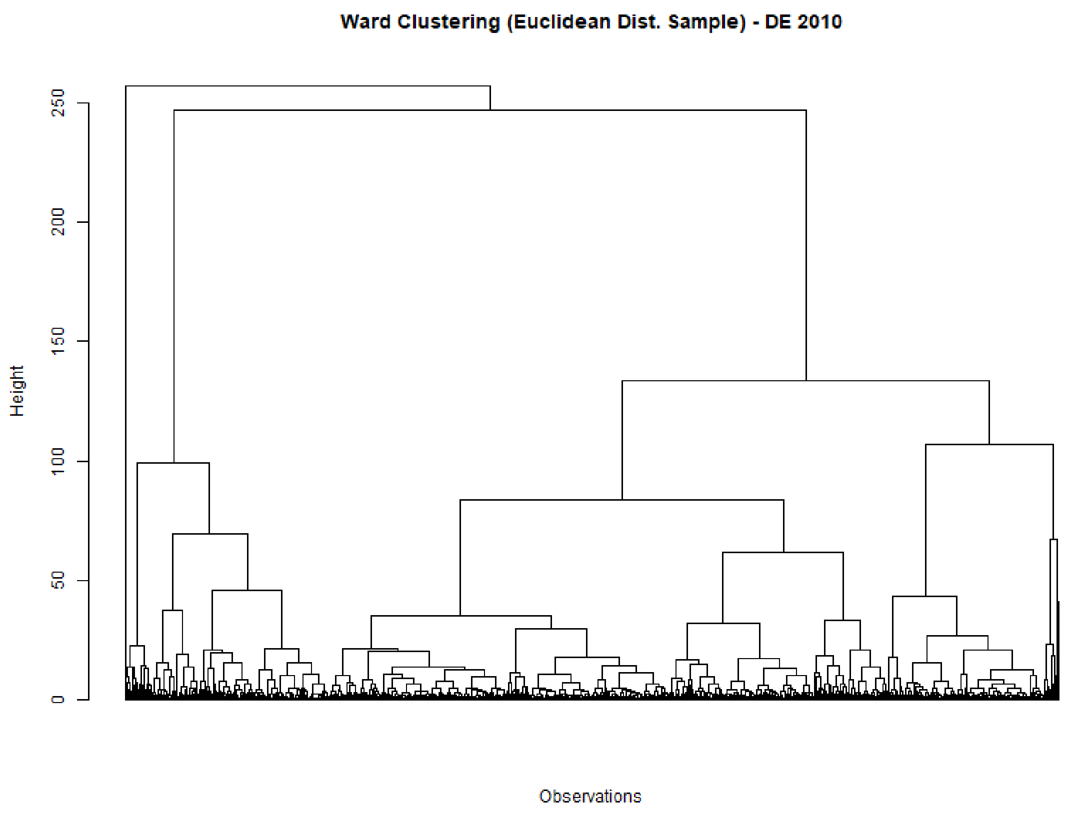

Figure 7: Dendrogram of the German Subset 2010

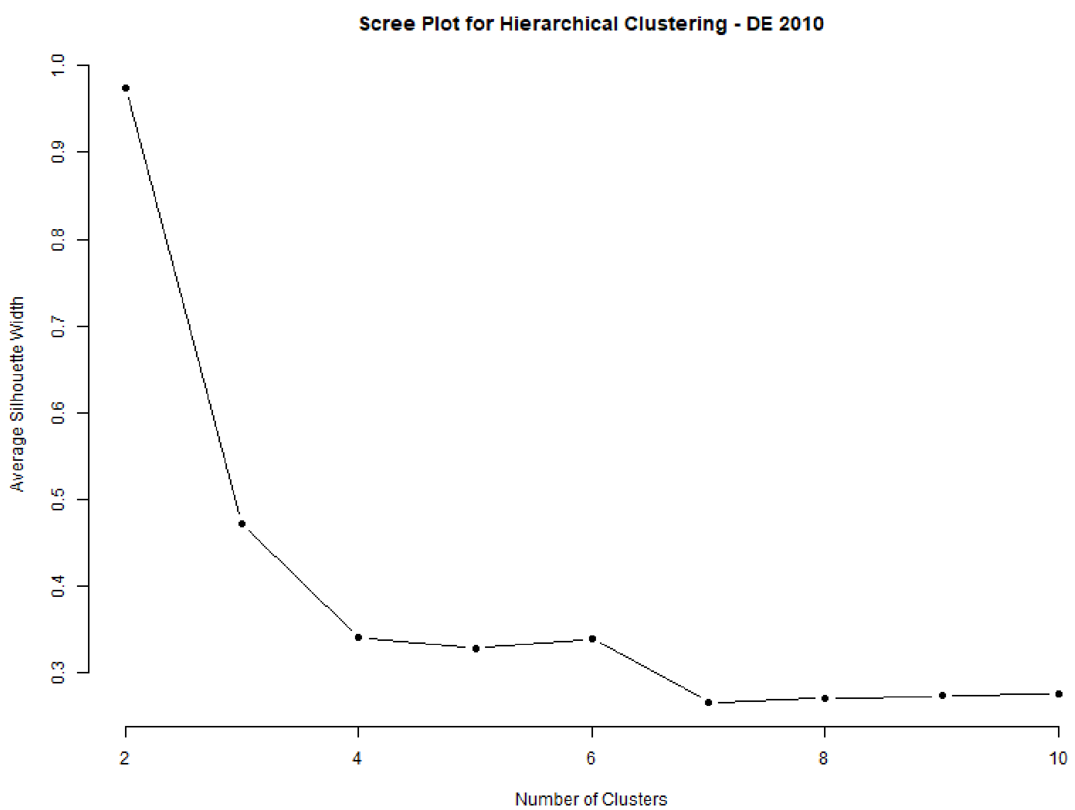

Figure 8: Scree Plot of the German Subset 2010

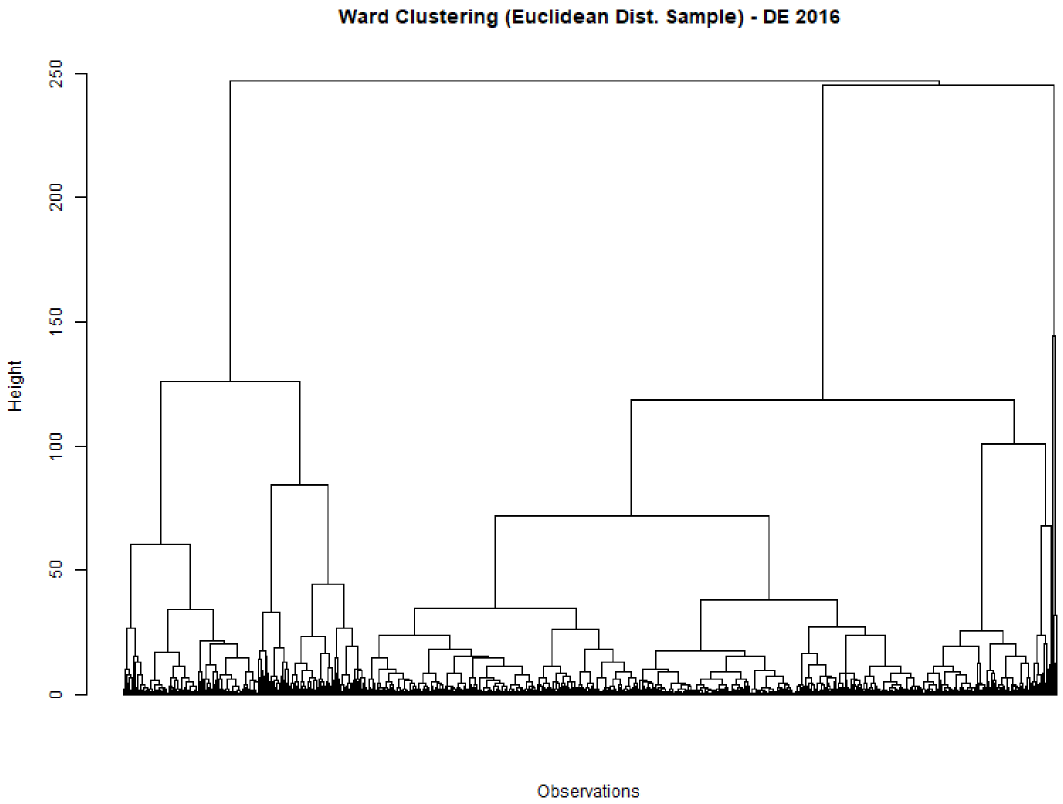

Figure 9: Dendrogram of the German Subset 2016

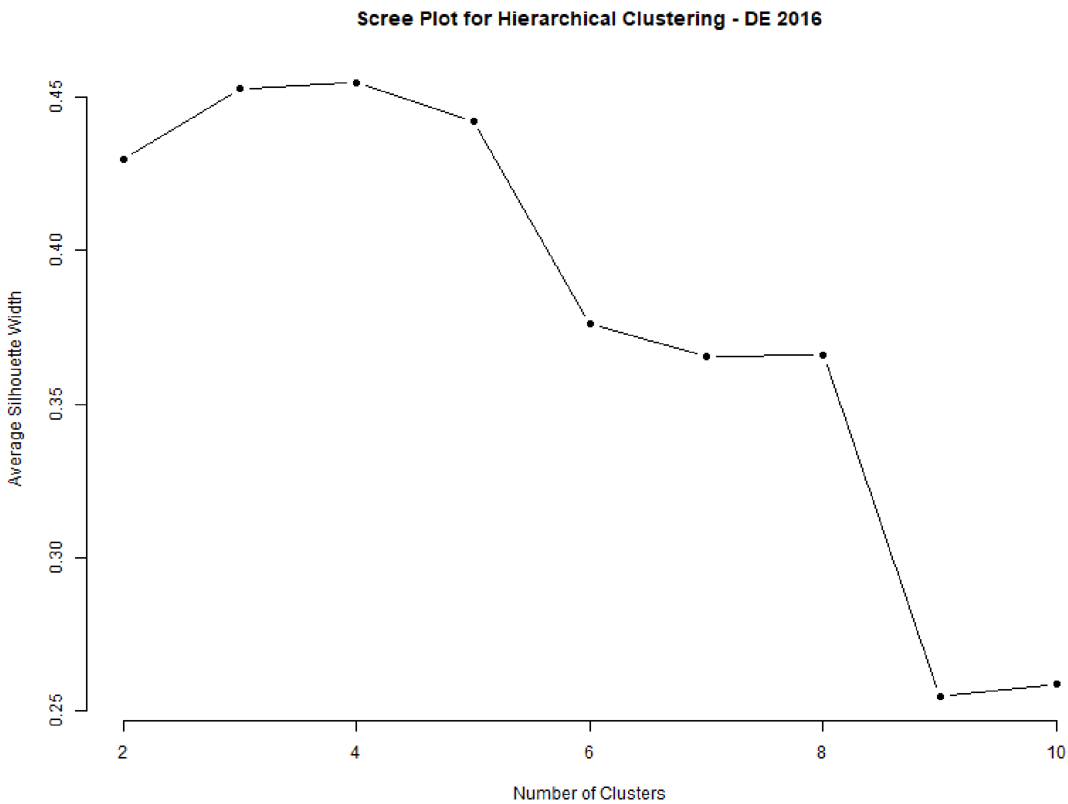

Figure 10: Scree Plot of the German Subset 2016

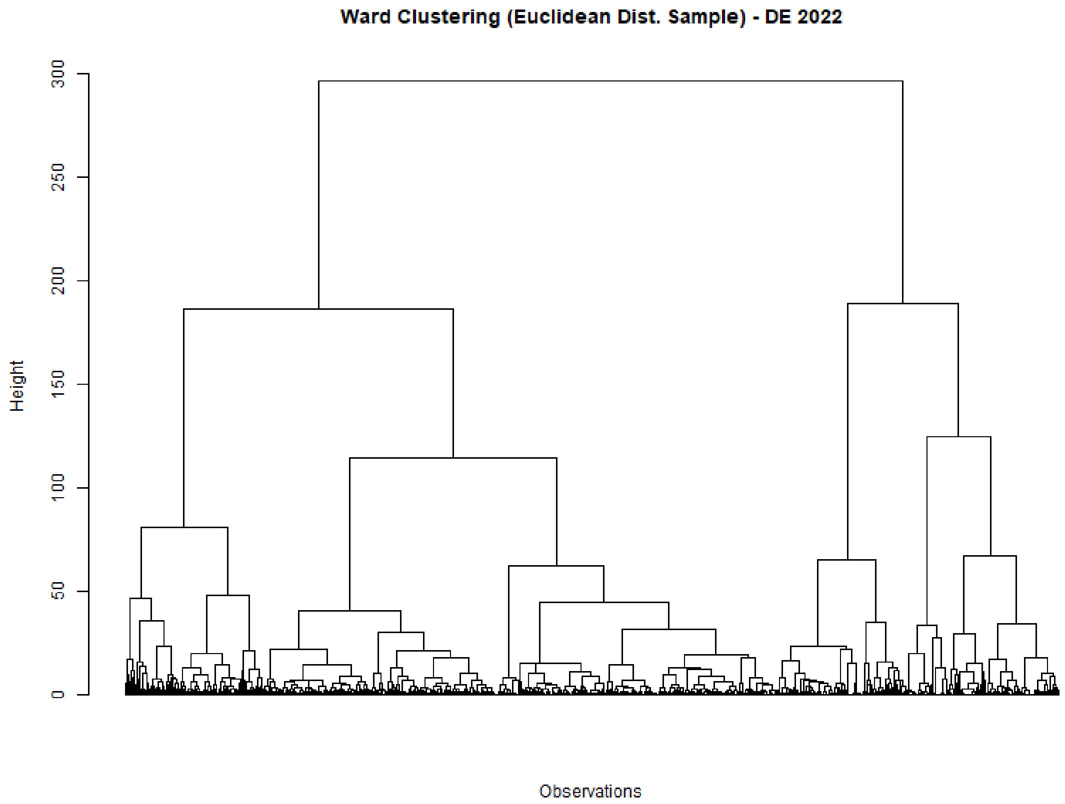

Figure 11: Dendrogram of the German Subset 2022

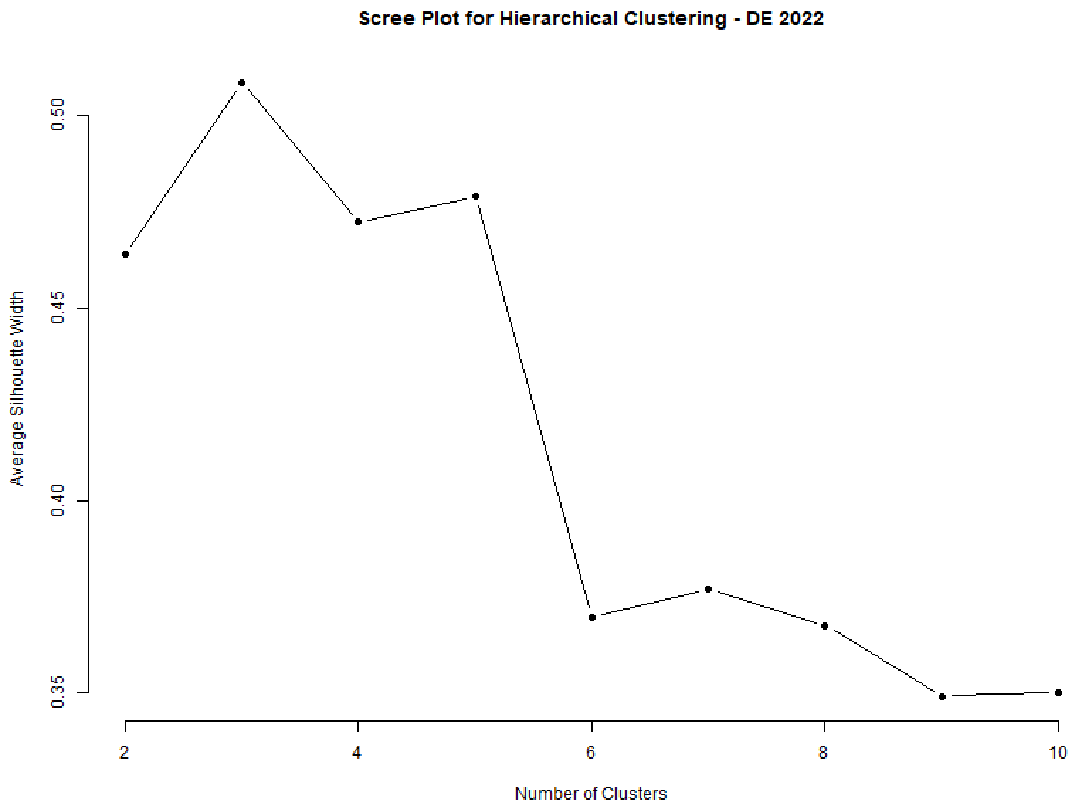

Figure 12: Scree Plot of the German Subset 2022

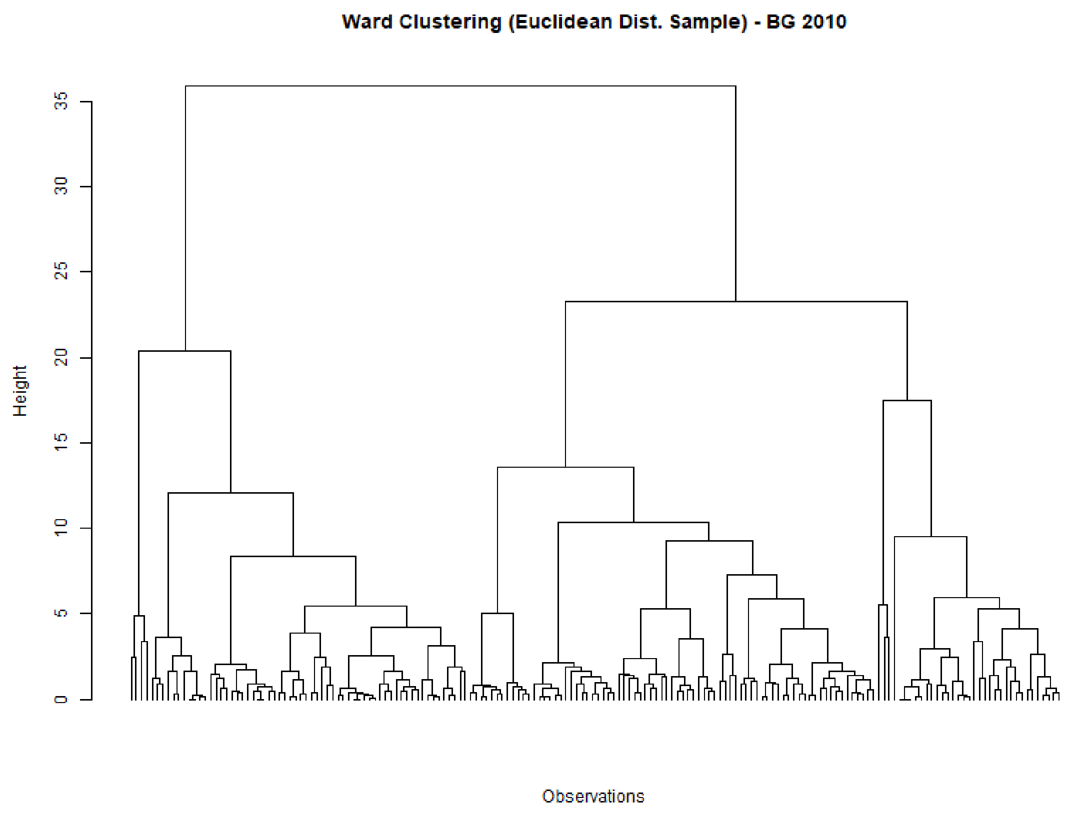

Figure 13: Dendrogram of the Bulgarian Subset 2010

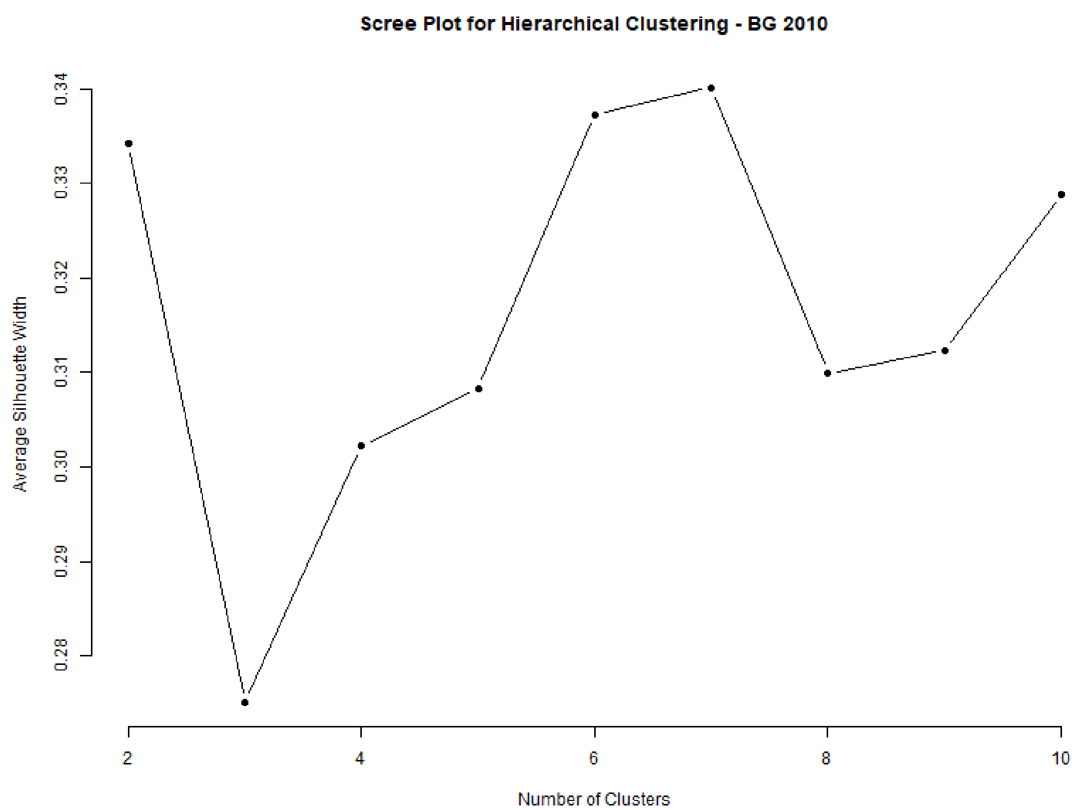

Figure 14: Scree Plot of the Bulgarian Subset 2010

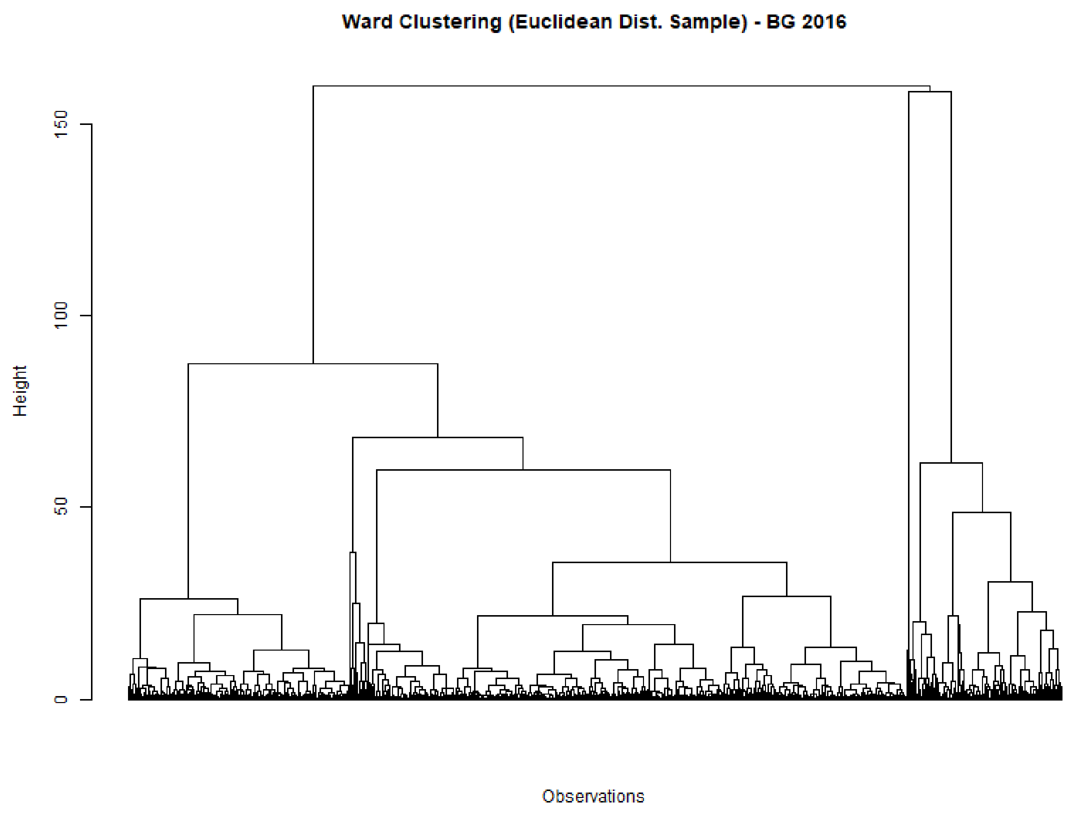

Figure 15: Dendrogram of the Bulgarian Subset 2016

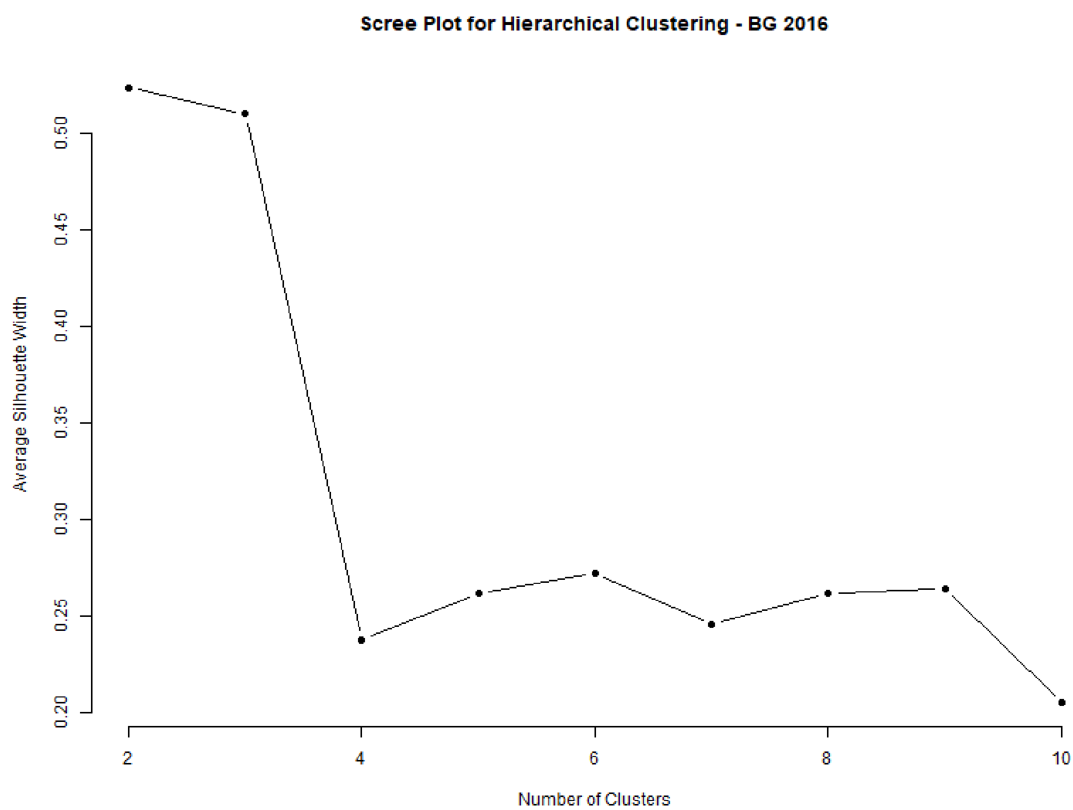

Figure 16: Scree Plot of the Bulgarian Subset 2016

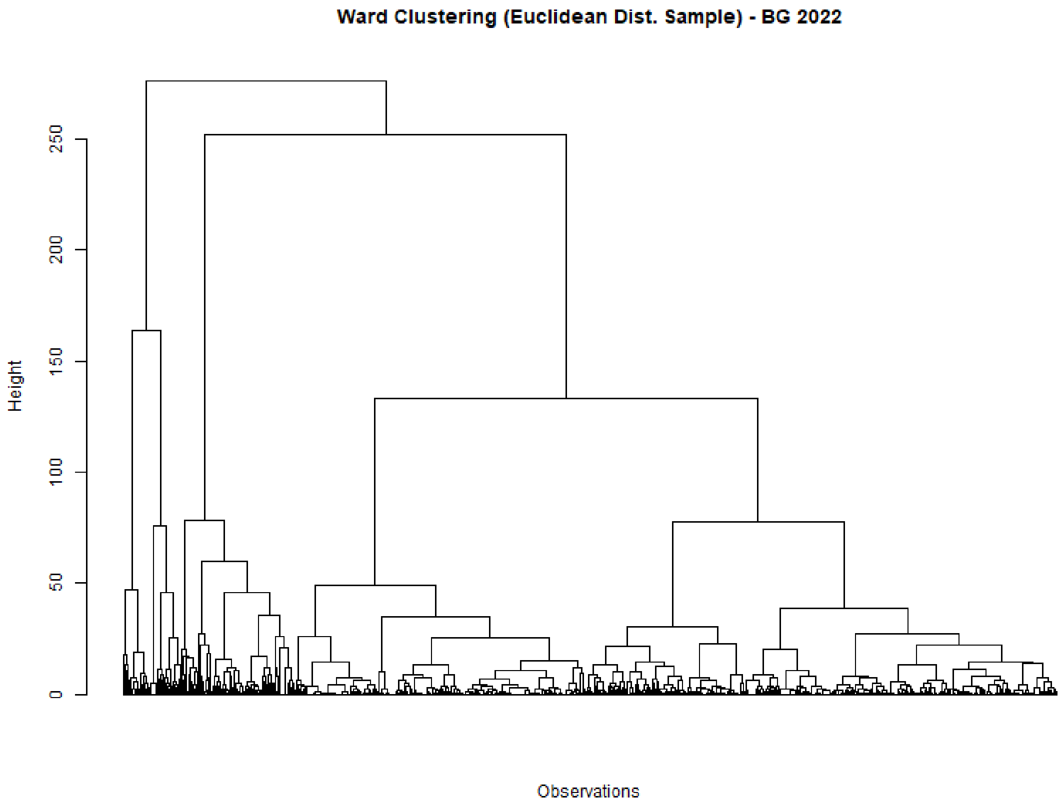

Figure 17: Dendrogram of the Bulgarian Subset 2022

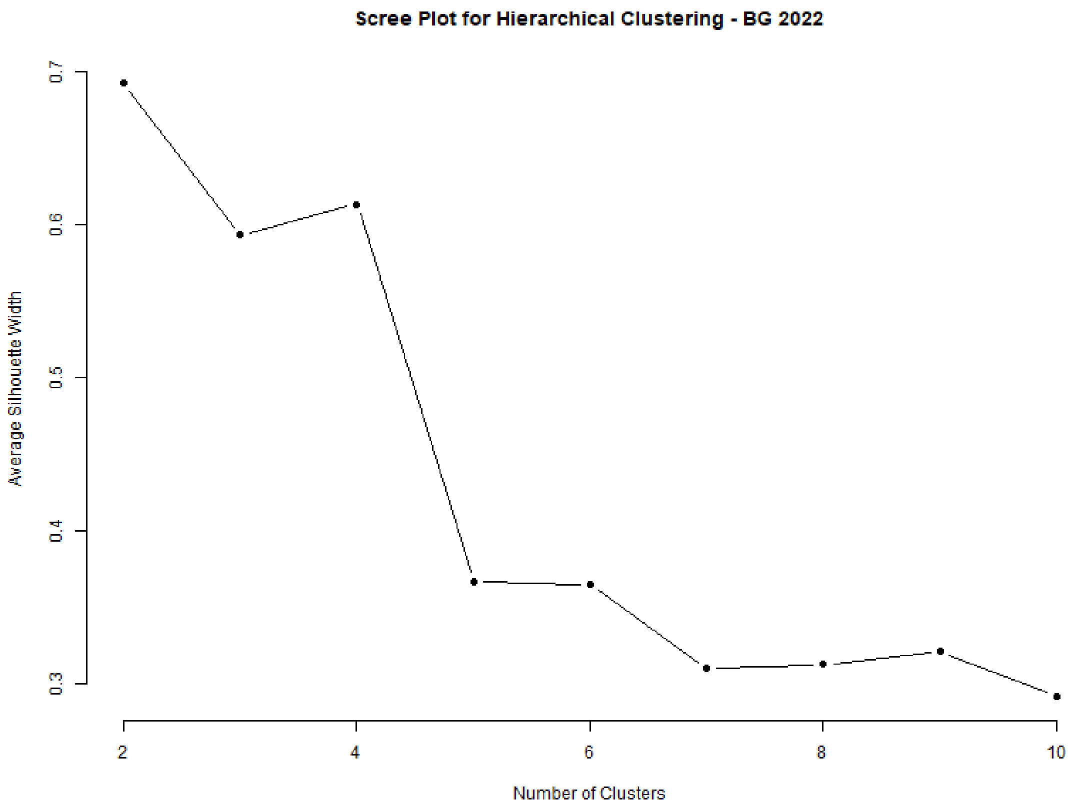

Figure 18: Scree Plot of the Bulgarian Subset 2022

References

ACEA (Jan. 2024a). *ACEA - Cars Fact Sheet*. Online via URL: https://www.acea.auto/fact/fact-sheet-cars/. URL: https://www.acea.auto/fact/fact-sheet-cars/.

— (Mar. 2024b). "Economic and Market Report Global and EU auto industry: Full year 2023". In: URL: https://www.acea.auto/files/Economic_and_Market_Report-Full_year_2023.pdf.

EEA (Aug. 2023). *Monitoring of CO2 emissions from passenger cars, 2022*. Online via EEA Datahub. DOI: 10.2909/992616f8-158f-4ecc-b978-814b81629db6.

Eisenmann, Christine and Ralph Buehler (May 2018). "Are cars used differently in Germany than in California? Findings from annual car-use profiles". In: *JOURNAL OF TRANSPORT GEOGRAPHY* 69, pp. 171–180. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2018.04.022.

EPRS - European Parliament Research Service (Oct. 2024). *The crisis facing the EU's automotive industry*. Online via URL: https://www.europarl.europa.eu/thinktank/en/document/EPRS$_ATA$(2024)762419. URL: https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2024)762419.

Ikotun, Abiodun M. et al. (2023). "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data". In: *Information Sciences* 622, pp. 178–210. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2022.11.139. URL: https://www.sciencedirect.com/science/article/pii/S0020025522014633.

Kubiczek, Jakub and Bartlomiej Hadasik (Mar. 2021). "Segmentation of Passenger Electric Cars Market in Poland". In: *WORLD ELECTRIC VEHICLE JOURNAL* 12.1. ISSN: 2032-6653. DOI: 10.3390/wevj12010023.

Moering-Martinez, Gabriel, Murat Senzeybek, and Patrick Jochem (Oct. 2024). "Clustering the European Union electric vehicle markets: A scenario analysis until 2035". In: *TRANSPORTATION RESEARCH PART D-TRANSPORT AND ENVIRONMENT* 135. ISSN: 1361-9209. DOI: 10.1016/j.trd.2024.104372.

Morton, Craig, Jillian Anable, and John D. Nelson (2017). "Consumer structure in the emerging market for electric vehicles: Identifying market segments using cluster analysis". In: *INTERNATIONAL JOURNAL OF SUSTAINABLE TRANSPORTATION* 11.6, pp. 443–459. ISSN: 1556-8318. DOI: 10.1080/15568318.2016.1266533.

Murtagh, Fionn and Pedro Contreras (2012). "Algorithms for hierarchical clustering: an overview". In: *WIREs Data Mining and Knowledge Discovery* 2.1, pp. 86–97. DOI: https://doi.org/10.1002/widm.53. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.53. URL: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.53.

Murtagh, Fionn and Pierre Legendre (Oct. 2014). "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?" In: *JOURNAL OF CLASSIFICATION* 31.3, pp. 274–295. ISSN: 0176-4268. DOI: 10.1007/s00357-014-9161-z.

Tesla (Dec. 2024). *Model Y*. Online via URL: https://www.tesla.com/de$_de/modely$. URL: https://www.tesla.com/de_de/modely.