Master's Thesis

Influence of state subsidies and local conditions on the development of public EV charging infrastructure within Europe.

Case Study Research on the example of France, Germany and Italy.

submitted to

Dr. Konstantin Hopf

in Information Systems and Energy Efficient Systems at the Faculty of Information Systems and Applied Computer Sciences
University of Bamberg

by

Cedric Dean Easton

on 16 October 2020

Matriculation number: 1919422
E-mail: cedric-dean.easton@stud.uni-bamberg.de
Degree programme: M.Sc. Information Systems

Influence of state subsidies and local conditions on the development of public EV charging infrastructure within Europe.

Case Study Research on the example of France, Germany and Italy.

by

Cedric Dean Easton

on 16 October 2020

Abstract

To turn the EU into a resource-efficient economy, national legislators are ramping up efforts to reduce total greenhouse gas emissions from the transport sector. For this, electric vehicles (EV) prove advantageous as they produce no on-road emissions, which is manifested in an increasing expansion of public charging infrastructure (CI) across Europe. While fast chargers are generally considered not commercially viable by default, their benefits to EV adoption rate were found to be positive. As is the case with a generally more developed CI, which is why continued subsidies are suggested in the literature. The latter also includes different allocation methods for optimal CI locations indicating prevailing discourse.

To help expand literature in this regard, an R framework was built to categorise locations using a Latent Dirichlet Allocation topic modelling approach. This framework relies on CI from Open Charge Map and comprehensive local data provided by Geofabrik extracts from OpenStreetMap. The framework matches each CI location with all localities found within a specific walking distance and uses the derived local conditions to categorise accordingly. This, in turn, enables the evaluation of category development over time, facilitating conclusions regarding the influence of state subsidies and local conditions. While the latter's influence was found to be dominated by residential localities, state subsidies were found to yield a significantly positive impact. Especially fast chargers along motorways benefited from focused subsidies, while CI in residential areas was found to evolve even without.

Contents

List of Abbreviations

BEV Battery Electric Vehicle

CI Charging Infrastructure

CRS Coordinate Reference System

CS Charging Station

DTM Document Term Matrix

EAFO European Alternative Fuels Observatory

EC European Commission

ETRS89 European Terrestrial Reference System from 1989

EU European Union

EV Electric Vehicle

fclass Locality Feature Class

GIS Geographic Information System

LDA Latent Dirichlet Allocation

MPCA Multinomial Principal Component Analysis

OCM Open Charge Map

OSM OpenStreetMap

PHEV Plugin Hybrid Electric Vehicle

SF Simple Features for R

VGI Volunteered Geographic Information

WGS84 World Geodetic System from 1984

List of Tables

List of Figures

1 Introduction

In 2014, the European Union (EU) initiated their Seventh Environment Action Programme period (2014-2020) with a priority objective being to “turn the Union into a resource-efficient, green, and competitive low-carbon economy” (European Union, 2018). A central aspect to this are the general emissions from the road transport sector, which “accounts for 72 % of total greenhouse gas emissions” (EEA, 2019). For this, Electric Vehicles (EVs), “which do not require petroleum fuel, can provide many benefits over internal combustion engine-based vehicles” (Mersky et al., 2016), as “they produce no on-road greenhouse gas emissions” (Mersky et al., 2016). According to the 2018 environmental indicator report, the goal to “reduce the overall environmental impact of production and consumption in the mobility sector” (European Union, 2018) compared to 1990 levels is unlikely to be met by the end of 2020 (European Union, 2018). Nonetheless, national legislators across the European Union have ramped up the expansion of Charging Infrastructure (CI) for EVs in recent years (EAFO, 2020a), following an EU legislation from 2014 urging its member states to “ensure, by means of their national policy frameworks, that an appropriate number of recharging points accessible to the public are put in place by 31 December 2020” (European Union, 2014). This has ignited significant research efforts into, among others, the general necessity, economic feasibility, location optimisation and required numbers of public Charging Stations (CSs).

Especially for fast chargers, the consensus is that a public CI network is “unlikely to be profitable” (Schroeder and Traber, 2012) by default, which is why “continuing financial incentives” (Serradilla et al., 2017) are proposed. For example due to “the risks associated with continuing technological development” (Serradilla et al., 2017), one of which being the increase of EV battery range which can “make a dense network of fast chargers redundant” (Wenig et al., 2019). Nonetheless, “fast chargers could provide assurance and comfort to reduce range anxiety” (Neaimeh et al., 2017), because “drivers want the comfort of knowing they can recharge if and when required” (Serradilla et al., 2017). Further agreement can be found in the importance of “home charging [...] for all segments” (Wenig et al., 2019), as the “vast majority of the charges (approx. 88%) are carried out at the place of residence” (Baresch and Moser, 2019). Discourse, however, prevails in the search of optimal CI locations, as indicated by the presence of various optimisation and allocation methods as well as in the search for the appropriate number of required CSs in relation to registered EVs. Another question to discuss is the maximum distance between a CS and a driver’s destination before they stop considering it as a viable charging location. For this reason, three different walking distances are defined (200m, 400m and 600m) to analyse the impact a change in distance has on the categorisation results. To help mitigate the prevailing discourse regarding CI locations, this thesis aims at categorising the existing CSs by their surroundings. This location-based categorisation shows how the public CI is distributed and balanced across locations and subsequently enables a characteristics-based analysis. As "an increased number of public charging points has a significant and positive impact on the Battery Electric Vehicle (BEV) adoption" (Egnér and Trosvik, 2018), a well-developed CI is considered advantageous in this thesis. The characteristics-based analysis uses the date at which each CS was created to model the development of CI over time and enable conclusions regarding the influence of local conditions. This development over time is then compared with the time frame of state subsidies to derive their impact on the overall CI development as well as on individual categories.

As a consequence of this, the research question that this thesis is focused on is: How does the public EV charging infrastructure develop in Europe *(on the example of France, Germany and Italy)* and what influence do state subsidies and local conditions have in that regard? To answer this, a framework was established in R that uses Geofabrik extracts from the OpenStreetMap (OSM) database to conduct a categorisation of locations based on surrounding local conditions using a Latent Dirichlet Allocation (LDA) topic modelling approach. The robustness of the results is increased by applying this categorisation for three different walking distances. Aside from the location-based categorisation, a characteristics-based analysis is conducted using individual attributes from the Open Charge Map (OCM) database to, for example, investigate the CI development over time.

2 Theoretical Background

2.1 Building Theory

The structural framework of this thesis is based on Building Theories from Case Study Research (Eisenhardt, 1989). Although their focus is “the larger context of social science research” (Eisenhardt, 1989), the presented roadmap “focuses on understanding the dynamics present within single settings” (Eisenhardt, 1989) and “can involve either single or multiple cases, and numerous levels of analysis” (Eisenhardt, 1989). The single research setting in this thesis is the development of CI within Europe. Its analysis, in turn, is based on multiple cases in the form of Germany, France and Italy and includes *within-country* and *cross-country* analyses as well as a sensitivity analysis for increased robustness of the results. For these reasons, the compiled general roadmap (table 2.1 as based on table 1 in Eisenhardt (1989)) is suitable for this thesis and only requires minor adjustments for the *Crafting Instruments and Protocols* and *Entering the Field* steps. These two are mostly linked to the context of social science research as they rely on data collection, which is not conducted in this thesis. Therefore, these two steps are replaced with the selection of appropriate, publicly available databases on the one hand and the preparation of the available data on the other.

Table 2.1: Roadmap for building theory. Based on table 1 in Eisenhardt (1989)

Step	Activity
Getting Started	Definition of research question.
	Possibly a priori constructs.
	Neither theory nor hypotheses.
Selecting Cases	Specified population.
	Theoretical, not random, sampling.
Crafting Instruments and Protocols	Multiple data collection methods.
	Qualitative and quantitative data combined.
	Multiple investigators.
Entering the Field	Overlap data collection and analysis, including field notes.
	Flexible and opportunistic data collection methods.
Analysing Data	Within-case analysis.
	Cross-case pattern search using divergent techniques.
Shaping Hypotheses	Iterative tabulation of evidence for each construct.
	Replication, not sampling, logic across cases.
	Search evidence for “why” behind relationships.
Enfolding Literature	Comparison with conflicting literature.
	Comparison with similar literature.
Reaching Closure	Theoretical saturation when possible.

2.2 Differentiating Research

Prior to the definition of the research question in the first step, a literature assessment was conducted by retrieving all papers from the journals *Energy Policy*, *Applied Energy* and *Energy* that matched against either of the keywords `'charging infrastructure'` or `'charging station'`. The abstracts of the retrieved papers were subsequently scanned to evaluate their relevance for this thesis. A further, but broader assessment was conducted for those papers that matched against `'electric vehicle'` by evaluating only their titles for relevance. From the relevant papers and their references, the following seven major research strands were identified and grouped to help specify the focus of this thesis.

1. Charging infrastructure and battery capacity as substitutes
2. Standard and fast charging infrastructure
3. Charging infrastructure to sooth range anxiety
4. Economic feasibility of large-scale charging infrastructure
5. Optimal placement of charging stations and rollout strategies
6. Appropriate number of required charging stations
7. Acceptable walking distances from charging stations

2.2.1 Charging infrastructure and battery capacity as substitutes

One strand of research considers CI and the battery capacity of EVs to be substitute investments to a certain degree. That assumption is related to this thesis, because "batteries, after reaching a critical size, make a dense network of fast chargers redundant for most [driver] segments" (Wenig et al., 2019) as concluded after an analysis of driving data by Wenig et al. (2019). With an analysis of commercially owned conventional passenger cars in Germany, Funke et al. (2019) investigated the economics of this interaction and suggest that "specific investments (€/BEV) in the expansion of [fast charging infrastructure] are three to seven times lower than those for a significant increase in battery capacity assuming the same market share of BEV." (Funke et al., 2019) Anderson et al. (2016) found diverging results for everyday and long-distance travel, respectively, with the defining threshold being a range of 100 km (Anderson et al., 2016). For everyday travel, Anderson et al. (2016) found that increased battery range only results in a small decline in CS need. For long-distance travel, however, increased battery range results in a significantly lower requirement for public CI. This conclusion implies a difference of use between standard and fast CI which manifests the second research strand.

2.2.2 Standard and fast charging infrastructure

The biggest distinguishing feature of CSs is their power supply: Usually, a distinction is drawn between normal power and high power; or standard charging and fast charging, respectively. According to EU legislation, normal power refers to power levels of "less than or equal to 22 kW" (European Union, 2014), while a high power CS "allows for a transfer of electricity to an electric vehicle with a power of more than 22 kW" (European Union, 2014). This threshold of 22 kW is also used by Gnann et al. (2018) for their analysis of charging behaviour and Anderson et al. (2016) for their concept of a demand-driven CI for Germany. After analysing EV charging behaviour in Ireland, Morrissey et al. (2016) concluded that energy "consumption values were higher for the fast charging infrastructure" (Morrissey et al., 2016) while the "charge duration values were higher at the public standard charging infrastructure" (Morrissey et al., 2016), implying that "fast chargers are used more regularly than the standard network." (Morrissey et al., 2016) Modelling the Austrian charging behaviour, Baresch and Moser (2019) found that the "vast majority of the charges (approx. 88%) are carried out at the place of residence" (Baresch and Moser, 2019). "It is clear then that the majority of daily driving can be met with current BEV models and standard slow chargers at private locations (i.e. home or work)" (Neaimeh et al., 2017), which suggests "that the BEV range is primarily a psychological barrier" (Neaimeh et al., 2017) introducing the third research strand.

2.2.3 Charging infrastructure to sooth range anxiety

While examining EV adoption in Sweden, Egnér and Trosvik (2018) found that "the main barriers affecting the consumers' decision of purchasing an EV are found to be battery range limitations resulting in range anxiety, high purchasing cost, and limited charging infrastructure" (Egnér and Trosvik, 2018). "Drivers want the comfort of knowing they can recharge if and when required, even if they subsequently don't often use the public [CI] provided to meet those perceived needs." (Serradilla et al., 2017) Therefore, "fast chargers are increasingly being considered as a practical and efficient method to ease the concerns of long charging durations." (Morrissey et al., 2016) "While this raises the question on whether a fast charge infrastructure is required" (Neaimeh et al., 2017), it would be false to interpret this "as evidence against supporting the roll-out of a fast charge infrastructure. Without fast chargers, the transition from liquid-fuel vehicles to BEVs will be affected" (Neaimeh et al., 2017), because drivers want "to be able to make all their journeys, not just the majority" (Neaimeh et al., 2017). Furthermore, "both drivers and public bodies continue to state that public [CI] is a requirement for private EV uptake, particularly stressing the need for rapid chargers because of their speed and therefore convenience" (Serradilla et al., 2017). "Fast chargers could provide assurance and comfort to reduce range anxiety and the perceived unsuitability of BEVs beyond short city driving." (Neaimeh et al., 2017) However, "the fast charge infrastructure provision is expensive and its utilisation levels are going to be low in the coming few years [...] which is not appealing to private investors." (Neaimeh et al., 2017)

2.2.4 Economic feasibility of large-scale charging infrastructure

For commercial viability, in a "best case scenario, a fast charger requires a usage frequency of 5.04–5.91 charges/day and a standard public charger requires a usage frequency of 2.24 charges/day" (Morrissey et al., 2016). While the "cost of fast charger installation is greatly in excess of the cost of installing standard charging infrastructure" (Morrissey et al., 2016), the former's "deployment may be the most desirable option in terms of commercial viability" (Morrissey et al., 2016). An evaluation of the 2017 CI in the United Kingdom resulted in the conclusion that a "credible financial business case does exist [...], if drivers are willing to pay 3.3 mark-up on electricity prices." (Serradilla et al., 2017) They continue, however, that due to "the risks associated with continuing technological development, consumer acceptance and drivers' willingness to pay the mark-up required, alternative solutions focussing on wider non-financial value should also be investigated." (Serradilla et al., 2017) These concerns are reflected in the already assessed literature as the consideration of CI and battery capacity as substitutes presents a technological development potentially impacting the use statistics of CSs. Neaimeh et al. (2017) concluded that "at current BEV market share, fast charge networks might not be profitable in the near-term" (Neaimeh et al., 2017). It is, therefore, not surprising that the study by Serradilla et al. (2017) "proposes continuing financial incentives to protect investors from uncertainties in the marketplace." (Serradilla et al., 2017) An analysis of the German CI from 2012 yielding the conclusion that a "market-driven roll-out [...] is unlikely to be profitable in Germany at 2011 EV penetration rates" (Schroeder and Traber, 2012) leads to the fifth identified research strand.

2.2.5 Optimal placement of charging stations and rollout strategies

The aforementioned conclusion that "the vast majority of the charges (approx. 88%) are carried out at the place of residence" (Baresch and Moser, 2019), coupled with the finding that another "approx. 8.8%, are carried out at the workplace" (Baresch and Moser, 2019) shows the general consensus that "home charging was essential for all [driver] segments" (Wenig et al., 2019). It "appears sufficient for 'frequent local drivers' [...], 'steady commuters' [...], 'short-distance delivery vehicles' [...], and 'long-distance delivery vehicles' [...] (if only a small number of fast chargers are available)." (Wenig et al., 2019) Morrissey et al. (2016) also found that "EV users prefer to carry out the majority of their charging at home in the evening" (Morrissey et al., 2016). Therefore, "in large cities where people to a higher extent live in flats rather than houses, charging infrastructure may be even more important since the convenience of charging at home may be limited" (Egnér and Trosvik, 2018). Additionally, "the effect of charging infrastructure is higher in urban municipalities than in other municipality types." (Egnér and Trosvik, 2018) This agrees with the 2014 EU legislation specifying that "recharging points accessible to the public should be installed, in particular at public transport stations" (European Union, 2014) and "in collective parking lots, such as in apartment blocks and office and business locations." (European Union, 2014)

Aside from this, the placement of CSs depends on the market saturation and rollout strategy. After an analysis of charging data from the Netherlands, Helmus et al. (2018) argue that "in a less mature market, demand-driven [CSs] seem the most likely candidate as they show significant higher performance on energy transfer" (Helmus et al. (2018)). *'Demand-driven'* rollout refers to CSs built in residential areas close to EV drivers based on their request Helmus et al. (2018). "In a more mature market [...] a more balanced portfolio [between demand-driven and strategic CSs] may be more applicable." (Helmus et al., 2018) They conclude that "demand-driven [CSs] outperform strategic [CSs] on weekly energy transfer and connection duration, while strategic [CSs] outperform their demand-driven counterparts on charging time ratio." (Helmus et al., 2018) Furthermore, various methods of calculating the optimal locations for CI are available, often with a multi-faceted approach for example considering “the location, the capacity, and the coverage area of any kind of facility” (Frade et al., 2011) or “considering operators, drivers, vehicles, traffic flow and power grid” (Kong et al., 2019). For the European Commission (EC), Gkatzoflias et al. (2016) presented a methodology to provide optimal “placement in a city network [...] and placement in a regional or national network [...]” (Gkatzoflias et al., 2016).

2.2.6 Appropriate number of required charging stations

The optimisation model by Frade et al. (2011) specified the estimation of CI demand as the “main difficulties encountered in the development of the study” (Frade et al., 2011). It, therefore, includes formulas for the estimation of refuelling demand and also differentiates between *night-time demand* “related to the population (households) living in the neighborhood” (Frade et al., 2011) and *daytime demand* “related to the employment (jobs) offered there.” (Frade et al., 2011) Legislation by the EU indicates that “the appropriate average number of recharging points should be equivalent to at least one recharging point per 10 cars” (European Union, 2014), “taking into account the number of electric vehicles estimated to be registered by the end of 2020 in each Member State.” (European Union, 2014) Agreeing with this, Anderson et al. (2016) found that around 33.000 CS are required for everyday traffic and a further 2.600 for long-distance travel for all of Germany. Putting it in relation, one public CS for 10 EVs may be required (Anderson et al., 2016). A differing conclusion was reached by Funke et al. (2019), where “a ratio of less than two charging points per 1000 BEV can satisfy demand.” (Funke et al., 2019) A charging behaviour analysis conducted by Gnann et al. (2018) prompted results even further away from the indication in the EU legislation. Their analysis found that “the ratio of refueling stations per 1000 vehicles, for fast charging points and battery electric vehicles can be close to conventional cars if charging power and battery sizes keep increasing” (Gnann et al., 2018). In 2018, this ratio was at “0.3 for Germany and 1.8 for Sweden” (Gnann et al., 2018).

2.2.7 Acceptable walking distances from charging stations

The maximum allowed distance localities may have to be considered relevant for the local condition of CSs is the last strand of research identified by the literature assessment. In their search of an optimal allocation of CI, Gkatzoflias et al. (2016) considered 300 metres as the maximum “acceptable walking distance between a charging station and POI” (Gkatzoflias et al., 2016). Frade et al. (2011), however, assumed “400 m as the maximum desirable walking distance [...] and 600 m as the maximum acceptable walking distance [...].” (Frade et al., 2011) In a related thesis conducting a parking location classification for the ‘Information Systems and Energy Efficient Systems’ faculty of the University of Bamberg, Germany, a “bounding box of 100 meters in each cardinal direction is used as query area, resulting in a 200 by 200 meters square with the parking spot in its center.” (Wölfl, 2020) For each square, a list of all available features within the bounding box is retrieved from OSM (Wölfl, 2020). Another related thesis for the same faculty divided Germany into a 6 km by 6 km grid mesh analysing the CI development by investigating the localities within each grid cell (Schepp, 2020). For their planning of CI, Cavadas et al. (2014) assumed "that 20 min is the maximum time-distance that the drivers are willing to walk from the parking space where they leave the vehicle charging and their destination" (Cavadas et al., 2014). For a walking speed of around 10 km/h that results in a distance of over 3 km to be considered for the location-based categorisation of CSs.

2.2.8 Implications

Agreement exists in the literature regarding fast chargers not being commercially viable by default and of home charging being important for most drivers. Although their practical necessity remains questionable due to technological development, the absence of fast chargers was found to affect BEV adoption rate. The only differentiation between standard and fast charging in this thesis is the location category of CSs on parking and service areas along motorways, which are considered to be predominantly fast chargers. A more in-depth investigation into the distribution of chargers with different power levels and connection types is not in the scope of this thesis as it focuses on the influence of local conditions on CI development to help expand literature in search for optimal locations. The questionable economic feasibility of public CSs is related to this thesis as continued financial incentives were suggested. Therefore, this thesis investigates the influence of state subsidies on the overall development. As a well-developed CI is considered advantageous in this thesis, the number of required CSs is largely disregarded as an increased number was found to positively impact BEV adoption, which is highly beneficial for the aforementioned EU goal of reducing total greenhouse gas emissions in the transport sector. The discourse regarding the maximum walking distance around CSs is mitigated by conducting the categorisation for three different measures to infer the framework dependence on this parameter.

2.3 Describing Subsidies

In the identified research strands, the "expansion of charging infrastructure is [...] indicated to be an effective measure to promote BEVs" (Egnér and Trosvik, 2018) and "continuing financial incentives to protect investors from uncertainties in the marketplace" (Serradilla et al., 2017) are proposed. A core initiating factor in this development is an EU legislation from 2014, in which the EU urged its member states to "ensure, by means of their national policy frameworks, that an appropriate number of recharging points accessible to the public are put in place by 31 December 2020" (European Union, 2014). "The number of such recharging points shall be established taking into consideration [...] the number of electric vehicles estimated to be registered by the end of 2020" (European Union, 2014). Therefore, the following section will describe the individual subsidy programmes Germany, France and Italy put in place to meet the mandated goals. Incentives for the purchase of EVs themselves, however, are not considered by this thesis.

2.3.1 Germany

In September 2019, the federal government of Germany initiated their 'Climate Action Programme 2030' (Federal Government of Germany, 2019). This plan specifies the goal that "a total of one million charging stations are to be available by 2030." (Federal Government of Germany, 2019) One step in achieving this is the order to all petrol stations in the country for the installation of CSs on their parking areas (Federal Government of Germany, 2019). This mandate "will promote the development of a network of public charging stations" (Federal Government of Germany, 2019). In line with the general consensus of 2.2.5, they understand that "most charging will [...] take place at home or at the workplace." (Federal Government of Germany, 2019) For that reason, "a buyer's premium will thus also be made available for private and commercial charging infrastructure." (Federal Government of Germany, 2019) The efforts and subsidies for a CI expansion exist since early 2017 with an amount of 300 million € specified for the time frame of 2017 to 2020 (BMVI, 2020b). According to the sixth call for subsidy application, CSs with normal power recharging are subsidised with up to 2,500 €, high power CSs below 100 kW with up to 12,000 € and high power CSs with or above 100 kW are subsidised with 30,000 € (BMVI, 2020b). Eligible for those applications are private investors, cities and municipalities and subsidies include the connection to the grid as well as assembly (BMVI, 2020a). In 2020, for the first time, charging equipment for private individuals is subsidised with a total of 50 million € (BMVI, 2020a). From around 22,000 CSs already subsidised, nearly one quarter are fast chargers (BMVI, 2020b).

2.3.2 France

In a recent push to promote France as a leading producer of EVs, President Emmanuel Macron announced an 8 billion € plan (Rose, 2020). Even before that, however, citizens could apply for a 300 € tax credit for the purchase of an EV charger at their primary residence (AVERE, 2020). Already in 2016, France started their ADVENIR programme to boost the French CI for which nearly 20 million € were accumulated in the time frame from 2018 to 2020 alone (ADVENIR, 2020). The ADVENIR programme currently offers 40% and 50% of the purchase and installation costs of CS for companies (ADVENIR, a) and neighbourhoods (ADVENIR, 2019) respectively. It also includes an offer of up to 2,160 € for *'demand-driven'* CS to public entities (ADVENIR, b).

2.3.3 Italy

Italy only recently started to subsidise the electrification of transport with the Eco-Bonus programme launched in mid 2019 (Ministero dello Sviluppo Economico, 2019). The only active subsidy for CI offered by this programme is a "tax deduction of 50% on a total amount of maximum 3,000 € [...] for the purchase and installation costs of EV chargers" (Wallbox, 2020). An annual total of 60 million € is available for 2019 and 70 million € for 2020 and 2021, respectively.

2.4 Describing Databases

The framework presented in this thesis will rely entirely on existing, publicly available databases and collect no data of its own, as mentioned in 2.1. The necessary data for the analyses is provided by the OSM and OCM databases, which can be considered as Volunteered Geographic Information (VGI). While traditional authoritative sources of geographic information follow procedures, content specifications and periodic quality assessment (Elwood et al., 2012), "VGI initiatives typically use none of these practices" (Elwood et al., 2012). Nevertheless, it is "proving valuable in addressing research questions that involve [...] concepts of place" (Elwood et al., 2012), as is the case in this thesis with the consideration of local conditions around CSs to estimate their influence on the CI development. After investigating VGI for predictive energy data analytics, Hopf (2018) found that "VGI data can help to improve the quality of predictive algorithms" (Hopf, 2018) providing further justification for the use of VGI in analytical research.
OSM, the first database, is VGI in the traditional sense. It is an open, free and community-driven database (OSM) containing comprehensive local data which is used to investigate the individual surroundings of each CS. The reason why OSM "has been able to collect so much data in such a short time frame" (Topf and Ramm, 2019), however, is that "rules on how to map certain features are often not well defined and there is no mandatory quality control." (Topf and Ramm, 2019) This fact makes the export and use of non-standardised data a critical problem for a framework aiming for reproducibility and broader application. For this reason, the OSM data is not retrieved from OSM directly, but rather accessed through shape files offered by Geofabrik. Those shape files, "which are normally updated every day" (Geofabrik, 2018), are pre-processed and "have specified a general-use mapping of the basic features like roads, waterways, different land use types, and points of interest" (Topf and Ramm, 2019) to facilitate use in Geographic Information Systems (GISs). This thesis falls back on the free shape files offered by Geofabrik, which "have the same structure as [commercial] shape files [...], but the free files contain fewer layers, and are only available for smaller areas" (Geofabrik, 2018), as specified in the technical details linked on Geofabrik (2018). Table A.6 provides a comprehensive overview of available GIS layers and their respective Locality Feature Classes (fclasses), which follow the same structure across all areas. According to the Geofabrik documentation, there are two reasons why some layers may have more than one shape file. One reason is that "the availability of high-resolution aerial imagery has led to many POI features being recorded as areas [...], not points [...]. Features drawn as area in OpenStreetMap will be written to the layer with an `_a` suffix. Features drawn as line/-point in OpenStreetMap will be written to the layer without an `_a` suffix." (Topf and Ramm, 2019) An inspection of the downloads makes it clear that this is the case for six layers: `'natural'`, `'places'`, `'pofw'`, `'pois'`, `'traffic'` and `'transport'`. Another specified reason for more than one shape file for a single layer are spillover shape files. "When a certain layer becomes too large for one shape file (shape files are limited to 2 GB in size), it will automatically spill over into additional shape files. A shape file named "`osm_pois_v07_1.shp`" will have spillover shape files names "`osm_pois_v07_2.shp`", "`osm_pois_v07_3.shp`" and so on." (Topf and Ramm, 2019)

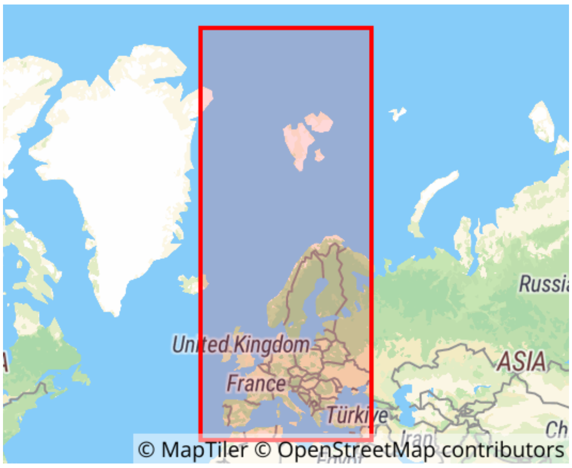


Figure 2.1: European Terrestrial Reference System. As provided by EPSG (2012)

The second database, OCM, takes VGI a step further. It is a "public, free, open database of charging equipment locations globally" (OCM, a) with geographic information provided in the form of longitude, latitude and address data. By default "all coordinates are unprojected WGS84" (Topf and Ramm, 2019). This is valid not only for the OCM, but also the OSM database. The World Geodetic System from 1984 (WGS84) is a *geographic* Coordinate Reference System (CRS) and therefore uses `'degrees'` as base units to specify where the geometry is located (Smith, 2020). It is for example "used by the GPS satellite navigation system and for NATO military geodetic surveying." (EPSG, 2007) On the grounds that the earth is not flat, the second step is converting the geometric data to a *projected* CRS for R to also know how to draw the geometries on a flat map. For consistency across all three analysed countries, the choice fell on the European Terrestrial Reference System from 1989 (ETRS89) (EPSG, 2012), which can be seen in figure 2.1. This CRS specifies `'metres'` as the default unit and has an accuracy of up to 1.0 m for onshore Europe (EPSG, 2012). This is crucial for the geometric calculations and manipulations conducted later-on in the script.

2.5 Identifying Categories

To bring these two databases together, a matching will be conducted which retrieves all localities from the OSM dataset that are within walking distance to each CS from the OCM dataset. Therefore, the result of this matching is of type discrete count data in the form of number of occurrences of each locality feature per CS. Prior to the analysis, the different categories of CSs that exist in the analysed countries are unknown. Previous categorisation methods involve Morrissey et al. (2016), who categorised CI by its use case and differentiated between *'Car Park'*, *'On-Street'*, *'Petrol Station'*, *'Household'* and *'Multi-Modal"*. While the first four are self-explanatory, *'Multi-Modal'* refers to CSs "located in areas where individuals leave their vehicle and use an alternative form of transport [...] (park and ride)" (Morrissey et al., 2016). The aforementioned related thesis by Wölfl (2020) conducted a categorisation of parking locations based on the surrounding environment and manually specified the following categories: *'Commercial'*, *'Food'*, *'Industrial'*, *'Leisure'*, *'Public'*, *'Residential'*, *'Nature'* and *'Parking'* (Wölfl, 2020). As this thesis is concerned with analysing the influence of local conditions specifically, a manual definition of categories is avoided. Therefore, the predominant option for the analysis in this thesis is to rely on unsupervised learning, which is "attractive in applications where data is cheap to obtain, but labels are either expensive or not available." (Wittek, 2014) "A learning algorithm, having no guidance, must identify structures on its own, relying solely on the data instances." (Wittek, 2014) This applies to this thesis,

because the data is free to obtain, but labels in the form of categories are not available and need to be derived from the data instances in an automated fashion.
Just this year, Jouvin et al. (2020) published a new algorithm for greedy clustering of count data based on a mixture of Multinomial Principal Component Analysis (MPCA). This combines the "dimension reduction to clustering [to] drastically improve performance and stability" (Jouvin et al., 2020) with the "flexibility of topic modeling while being able to assign each observation to a unique cluster." (Jouvin et al., 2020) The analysis in this thesis will be entirely conducted within R, "a language and environment for statistical computing and graphics." (R Foundation) Due to the recency of the mentioned publication, its implementation within languages for statistical computing is not yet established and not in the scope of this thesis. For this reason, the choice fell on their used foundation, the "building blocks of the so-called *topic models*" (Jouvin et al., 2020). These are the LDA and the MPCA, whose "key advantage [...], compared to other models for count data, is their flexibility. In particular, they allow observations to have mixed memberships towards the various topics." (Jouvin et al., 2020) Blei and Lafferty (2007) define topic models as "latent variable models of documents that exploit the correlations among the words and latent semantic themes" (Blei and Lafferty, 2007), which can "extract [...] interpretable and useful structure without any explicit "understanding" of the language" (Blei and Lafferty, 2007). "The topics discovered by the algorithm are found in a completely unsupervised fashion, using no information except the distribution of the words themselves, implying that the minor categories capture real differences" (Griffiths and Steyvers, 2004). "Although MPCA allows dimension reduction on discrete data, it is not designed for clustering *per se*" (Jouvin et al., 2020) and will, therefore, not find application in this thesis. This combination makes LDA a suitable candidate for the analysis in this thesis, because the categories are not yet known and it is entirely possible for CSs to be located in the vicinity of different categories. As a result of this, the advantage of mixed memberships of CSs towards the various categories is a great fit.

2.6 Introducing Latent-Dirichlet-Allocation

LDA was first described as a technique to reduce dimensions "for collections of discrete data such as text corpora." (Blei et al., 2003) "The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words." (Blei et al., 2003) The model can be defined as follows:

- "A *word* is the basic unit of discrete data, defined to be an item from a vocabulary indexed by $\{1,\ldots,V\}$." (Blei et al., 2003) In the context of this thesis, the vocabulary contains all possible fclasses, or *words*, that may be in the vicinity of CSs.
- "A *document* is a sequence of N words denoted by $\mathbf{w} = (w_1,w_2,\ldots,w_N)$, where w_n is the nth word in the sequence." (Blei et al., 2003) In the context of this thesis, a *document* corresponds to a single CS characterised by a sequence of all fclasses from its vicinity.

- “A *corpus* is a collection of M documents denoted by **D** $= \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M\}$.” (Blei et al., 2003) In the context of this thesis, a *corpus* represents a single country.

According to Blei et al. (2003), LDA relies on the *exchangeability*, or *bag-of-words*, assumption. This implies that the order within the *documents* as well as within the *corpus* can be neglected (Blei et al., 2003), which is given in the case of this thesis. The representation of a corpus is a so-called Document Term Matrix (DTM), “in which the frequencies of words in documents are captured.” (Ponweiser, 2012) Table 2.2 shows a small exemplary *corpus* as DTM with **D** $= \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ and vocabulary $V = \{\texttt{Locality\_1}, \texttt{Locality\_2}, \texttt{Locality\_3}, \texttt{Locality\_4}, \texttt{Locality\_5}\}$ as adapted from figure 3.1 in Ponweiser (2012).

Table 2.2: Example DTM with **D** $= 3$ *documents* and $V = 5$ *words*.

	Locality_1	Locality_2	Locality_3	Locality_4	Locality_5
Charging Station 1	2	1	0	0	1
Charging Station 2	1	0	2	3	1
Charging Station 3	0	2	4	0	0

Although the analysis in this thesis is not directly linked to text mining, LDA is considered to be an appropriate technique, because it “has applications to other problems involving collections of data, including data from domains such as collaborative filtering, content-based image retrieval and bioinformatics.” (Blei et al., 2003) The “recent success of “bag-of-words” representation for object recognition problems in computer vision” (Wang et al., 2007) even inspired the analysis of video sequences on the basis of LDA, “where each frame corresponds to a “word”.” (Wang et al., 2007) The implementation of the LDA in R is based on Silge and Robinson (2020) and Ponweiser (2012) and is founded upon the *topicmodels* package for R by Gruen (2020).

3 Methodology

This thesis aims to provide an answer to the research question: How does the public EV charging infrastructure develop in Europe *(on the example of France, Germany and Italy)* and what influence do state subsidies and local conditions have in that regard? To answer this, a framework was established that uses Geofabrik extracts from the OSM database to conduct a categorisation of desired locations based on surrounding OSM locality feature classes using an LDA topic modelling approach.
For the methodology, the approach of Building Theories from Case Study Research (Eisenhardt, 1989) (table 2.1) is used, starting with the research question above. Afterwards, the cases and databases are inspected and prepared for the subsequent analyses. These, in turn, are split into *within-country* and *cross-country* analyses. The former is important to become “familiar with each case as a stand-alone entity [...] before investigators push to generalize patterns across cases” (Eisenhardt, 1989) in the latter. Prior to this, however, one central issue needs to be addressed: The number of latent variables, or categories, needs to be specified before the execution of the LDA. While Ponweiser (2012) presents multiple possible learning algorithms to infer the number of categories, the maximum likelihood stands out as a method for estimating how well a given statistical model fits the set of observations as also conducted by Griffiths and Steyvers (2004). Therefore, in this thesis multiple iterations of the LDA are conducted with different numbers of categories and subsequently evaluated by their log-likelihood. For an increased robustness of the results, the prevailing discourse from 2.2.7 is balanced by a sensitivity analysis, which features three different walking distances around the CSs in which locality features are considered for the categorisation.

3.1 Selecting Countries

The aim of the thesis is to evaluate the development of CI within Europe. A geographically constraining factor was restricting the country selection to the EU rather than all of Europe. This was done due to considerable standardisation and research effort by the European Union (2014) and European Commission (Gkatzoflias et al., 2016). One criteria for the country selection was the investigation of road transport-related figures. In terms of total number of passenger cars registered per EU member state as well as millions of passenger-kilometres on national territory, Germany, Italy and France are leading in descending order as provided by the statistical office of the EU (Database Eurostat, 2020). Additionally, Germany, France and Italy are only eclipsed by Spain regarding the total length of motorways (Database Eurostat, 2020), while remaining at or above the EU average amount of public CSs per 100 km motorway, according to EAFO (2020a). This implies the existence of an already decent CI, which makes these three countries a good base for this thesis. Aside from transport-related figures, the

EU countries are also investigated in terms of number of residents, or population. In a post-Brexit EU, Germany, France and Italy have the largest population in descending order (European Union, 2020) and make up around 47% of the entire EU population (European Union, 2020).
Due to those statistics, Germany, France and Italy were selected to be used as research cases in this thesis.

3.2 Selecting Databases

A guiding principle of this thesis is to exclusively use open, free and publicly available databases. This is done to support the intention of the presented framework being applicable in a broad, universal range of research concerned with the categorisation of desired locations. For retrieving geographic data, OSM is the best decision as it is not only free of charge with an open licence, but also allows access to the raw vector data of localities (OSM, 2018). OSM further includes CSs represented by the `'charging_station'` amenity tag (OSM, 2020), providing basic information. These, however, are not included in the Geofabrik extracts, which is why an alternative is required: It exists in the form of OCM, a "public, free, open database of charging equipment locations globally." (OCM, a) This database offers comprehensive information regarding the characteristics of each CS as listed in their documentation (OCM, b).
To enable the estimation of the comprehensiveness of OCM, an initial investigation into the database is conducted for the case of Germany. In OCM, 12,581 CSs are listed, which stand in gross contradiction to the 41,461 charging points (EAFO, 2020b) listed on the aforementioned European Alternative Fuels Observatory (EAFO). Here, it is important to note that a single CS "can have more than one supply point to allow several vehicles to be charged simultaneously." (Frade et al., 2011) According to the EAFO knowledge centre, their means of providing CI statistics "can be translated to counting each socket and each plug" (EAFO, 2020e), which corresponds to the `'sumOfQuantities'` variable in the OCM dataset as shown in table 3.1. Disregarding `'NA'` values, the total number of connections for Germany is 33,146, which substantially narrows the gap to the EAFO data. According to the German government, a total of 15,426 public CSs is available in Germany as of September 9, 2020 (Bundesnetzagentur, 2020), resulting in a difference of 2,845 CSs between OCM and the official data as provided by the German government.
This thesis uses the number of distinct locations (CSs) as base, rather than the number of supply points or even number of sockets and plugs. This is done, because for drivers it is considered more important to know at how many locations they can charge their EV. Using the number of available supply points or even the number of connections as base gives a false sense regarding the coverage of CI. For conventional refuelling infrastructure also the number of petrol stations is considered instead of the number of fuel pumps or even number of 'taps', as done by Gnann et al. (2018) for the vehicle-to-refuelling station index mentioned in 2.2.6. Furthermore, this would falsify the location-based analyses as one CS park with 100 supply points and even more connections, as is in fact present twice in the German OCM dataset, would strongly bias the categorisation.

3.3 Retrieving Data

The foundation for this thesis is, therefore, provided by two databases. OCM on the one hand, for detailed information regarding the CI characteristics of each country and OSM on the other, for comprehensive local data which is used to categorise CSs according to their surroundings. In an attempt to build the R script as a flexible, stand-alone procedure that is largely independent from manual tasks as well as copyright restrictions, both databases are retrieved and prepared directly within R rather than being shipped alongside this thesis. The date of the last data retrieval for both OSM and OCM is 11 October 2020.

3.3.1 Open Charge Map

In the case of OCM, a download call to the official OCM GitHub (OCM, 2020) is executed in the code to retrieve the complete database as a JSON export. Upon import into R, this database is a structure of nested data frames which makes inspection and, subsequently, use highly unintuitive. For that reason, a selection of desired variables is performed and stored in a separate standard data frame for easier handling. This dataset encompasses a total of 16 variables as visible in table 3.1 where each observation represents a CS.

Table 3.1: List of selected variables for the OCM dataset.

Variable	Description
`id`	Used for querying individual CS.
`title`	Description of the location. *(Highly arbitrary)*
`address`	Usually street and nearest house number.
`town`	Name of nearest town.
`ZIP`	Postal/ZIP code of nearest town.
`country`	Country codes according to `‘ISO 3166-1 alpha-2’` (ISO, 2020)
`latitude`	Latitude data in WGS84 format.
`longitude`	Longitude data in WGS84 format.
`usage`	Specifies level of accessibility to the public. *(Full list in table A.4)*
`numberOfPoints`	Number of supply points for simultaneous charging.
`connections`	Types of available connection standards. *(Full list in table A.3)*
`chargers`	Power level of available chargers. *(Full list in table A.1)*
`quantities`	Number of each available connection type.
`sumOfQuantities`	Total number of available connections at each CS.
`status`	Status of CS operability. *(Full list in table A.2)*
`DateLastUpdate`	Date at which the CS was last updated in the database.
`DateCreated`	Date at which the CS was added to the database.

3.3.2 OpenStreetMap

As previously mentioned, the free shape files from Geofabrik “have the same structure as [commercial] shape files [...], but the free files contain fewer layers, and are only available for smaller areas” (Geofabrik, 2018). Because of that restriction, the desired OSM data needs to be retrieved per federal state rather than for the whole country. This results in multiple download calls to the Geofabrik download server executed from within R. In 2.4 it was specified that “shape files are limited to 2 GB in size” (Topf and Ramm,

2019) and if one should exceed that limit it "will automatically spill over into additional shape files." (Topf and Ramm, 2019) An inspection of the downloads shows that this limitation is not present in the datasets used by this thesis, which, therefore, does not account for this issue. All shape files include `'osm_id'` and `'code'` data plus specific data which can be considered irrelevant for this thesis. The `'places'` shape files for example include `'population'` and `'name'` information for each locality if available. Because of that, the shape file import is restricted to only the respective `'fclass'` and its `'geometry'` for increased consistency. The only exception to that is the `'building'` layer where the `'type'` variable is imported instead of `'fclass'` due to more detailed information. For consistency reasons necessary for the data preparation, it is immediately renamed to `'fclass'`. All resulting OSM datasets follow the structure depicted in table 3.2 below, with two variables and each observation representing one locality feature.

Table 3.2: List of selected variables for the OSM datasets.

Variable	Description
`fclass`	Locality feature class depicting the type of each locality.
`geometry`	Shape and position data of each locality.

3.4 Preparing Data

3.4.1 Feature Catalogue

Owing to its role as theoretical foundation, the catalogue with each possible fclass is initiated first and contains four variables. `'layer'`, `'fclass'` and `'description'`, which are retrieved manually from Topf and Ramm (2019) and OSM (2019a), plus `'filter'` which serves as a filter for undesired fclasses as well as a grouping method for semantically similar fclasses.

There are four possible reasons why certain fclasses were marked as undesired. One possibility is that the locality is unlikely to be the target of a car ride or that its duration of use or stay does not facilitate a charging event. Park benches and vending machines were filtered due to this as an example. A second possibility is that the locality does not provide any useful implication for characteristics of the surrounding environment. For example, general service roads or unspecified bodies of water will be taken out by this filter. Another filter is whether or not the locality is accessible to the public, as for example military compounds or all sorts of railways are not. Railways have been filtered here, because they are considered to be closed circuit. Meaning, public access to those localities can only be gained through designated points of entry (e.g. railway stations) which are filed separately. So, while a CS could be located next to train tracks, their mere existence is unimportant for its categorisation, as long as there is not also a railway station nearby, which would allow to actually use them. The last possible filter is for temporary localities like construction or excavation sites.

The second use of the `'filter'` variable is the semantic grouping. Due to the aforementioned absence of mandatory quality control, some localities can potentially be filed under multiple fclasses. A town hall for example can be a filed as `'building_public'`, `'building_civic'`, `'building_governmental'` or `'pois_public_building'`. Therefore, semantically similar fclasses have been assigned to semantic groups to improve

the analysis by removing semantically redundant geometries. A complete list of the semantic grouping as conducted in this thesis can be found in the combination of tables A.5 and A.6.
This feature catalogue is also the only aspect in which the R script is affected by copyright restrictions. The Geofabrik documentation, which the feature catalogue is based on, contains copyright restrictions in the form of the 'Creative Commons Attribution-Share Alike 2.0' licence. In an attempt to minimise restrictions on the remaining work, the function initiating the feature catalogue as an R data frame was outsourced into a separate script.
After the initialisation, the first step in the preparation of this catalogue is to add an incremental catalogue index, so each fclass has its own unique integer. This will be used to handle all future referencing, as comparing integers is generally more efficient than comparing strings. The second step is to generate unique long names for all fclasses. This is necessary because some shape files share fclasses, which could cause issues in the analysis. `'residential'`, for example, is an fclass present in shape files `'landuse'`, `'roads'` and `'building'`. Those long names are generated by connecting the shape file title and the fclass with an underscore, resulting in `'landuse_residential'`, `'roads_residential'` and `'building_residential'` for the previous example. After that, an automated generalisation is conducted on those long names to merge subcategories into their parent category. This is done to reduce the number to fewer, but more distinct fclasses. For example, the long names `'traffic_parking'`, `'traffic_parking_site'`, `'traffic_parking_multistorey'` and `'traffic_parking_underground'` are all present in the dataset and the distinction between them is not relevant in this thesis. The long name `'traffic_parking'` can be considered to be the parent category as the others are specifications of the same thing. Therefore, the automated generalisation iterates through all observations and checks whether the current long name is present within another. Here it is crucial to not check for the long name itself, but rather a concatenation of long name and underscore `'_'` to avoid false positives like `'pois_public_building'` being considered as a subcategory for and subsequently replaced by `'pois_pub'`. While the latter is present in the former, `'pois_pub_'` is not and therefore avoids a false positive. For true positives, the subcategory is replaced by the parent category. The result of this generalisation is stored in a separate variable in the feature catalogue, because the compiled long name is still required for the indexing of OSM localities with the corresponding feature catalogue index. As a result of this, localities of fclasses `'traffic_parking'`, `'traffic_parking_site'`, `'traffic_parking_multistorey'` and `'traffic_parking_underground'` all have the same catalogue index. Another part of this generalisation is also the processing of the semantic grouping. This is a simple check whether the individual fclass has been assigned to a semantic group through the `'filter'` variable. If it has, its catalogue index and generalised long name are replaced by the ones specified in the combination of tables A.5 and A.6.

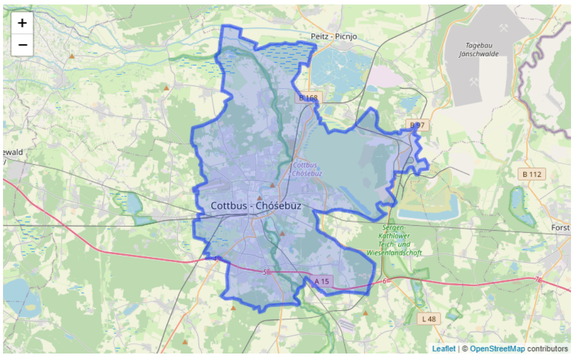
3.4.2 OpenStreetMap

After preparing the theoretical foundation, the OSM datasets are the first to be prepared. The first step is to match each locality from the OSM shape files with the corresponding catalogue index of the fclass. This is achieved by generating long names for all OSM localities as done previously and matching them to their respective counterpart in the feature catalogue. In some cases, it is possible that the dataset contains an fclass which is not specified in the Geofabrik documentation and, therefore, not present in the feature catalogue. Those occasions are treated as faulty and are marked accordingly, so they can be removed later-on. Throughout the R script, all data that is treated as faulty, redundant or is otherwise marked for removal, has its catalogue index replaced with `'NA'`. This way, the geometric data is still present, but can easily be dropped if need be. All geometric data is first converted to Simple Features for R (SF) geometries, with the WGS84 specified as their default CRS. To allow correct calculations and manipulations on that data, it is subsequently transformed to the ETRS89 (figure 2.1). Most processes involving geometric data rely on functions from the SF package by Pebesma (2020).

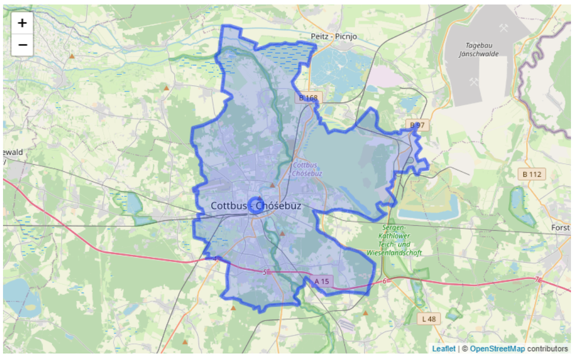
Converting the OSM geometries opens up the option for multiple cleaning stages. Prior to these, however, all geometries are processed by the `'st_make_valid'` function. This step is necessary, because on one occasion, the German *landuse* shape file contained an invalid geometry, which caused the following error in the preparation function: `'Error in CPL_geos_binop(st_geometry(x), st_geometry(y), op, par, pattern, : Evaluation error: TopologyException: side location conflict at [Longitude] [Latitude]'`. As it is possible for the problem to arise again in the future, the validity of all geometries is enforced to avoid this error.

The dataset is now ready for the various cleaning stages, the first of which is the identification and removal of redundant geometries within each shape file. This step relies on the `'st_equals'` function and returns a list for each observation containing the reference indices of all geometries equal in shape and position. The resulting list is always in ascending order and, since a geometry is always equal to itself, also always at least one element long. This is important because it allows only keeping the first unique localities from the list and removing all its subsequent occurrences. Whether localities are redundant or not is checked by using the `'duplicated'` function to process their catalogue indices. Every first unique catalogue index is kept and all subsequent occurrences of those are removed. Here is where the semantic grouping is required, because different OSM contributors could have filed the same locality under different specifications. Due to the semantic grouping, all fclasses that express the same meaning have the same catalogue index and can, therefore, be identified easily. For example, `'building_house'` and `'building_residential'` have different fclasses, but the same catalogue index as they are considered to be semantically similar. Nevertheless, it is still possible that equal geometries can have multiple unique fclasses. It is not good practice, however, to simply remove all but one of the equal geometries, because it is possible that a locality has been correctly filed under multiple specifications. For example, a student canteen can be both a `'pois_restaurant'` as well as a `'building_university'`. These two are semantically different, but both valid in the presented scenario. Another example of frequent co-occurrences of semantically different fclasses are `'landuse_forest'` and `'landuse_nature_reserve'`. Not every forest is a nature reserve and not every nature reserve is a forest. Therefore, it is not feasible in this thesis to decide in an automated fashion whether the semantically different fclasses are faulty or whether they are valid in that specific scenario.

After the data inspection in 3.3, it is clear that some layers include both polygon and point shape files. For a multitude of localities, OSM stores not only the outlines of the geometric shape, but also a point at which a descriptive label is displayed. Removing those redundant labels from the dataset is important for the analysis conducted later-on in the thesis. Every OSM layer containing a polygon and a point shape file, is therefore being iterated through to identify all intersecting points for each polygon. This step relies on the 'st_intersects' function. To avoid removing false positives from the dataset, it is crucial to check whether the intersecting points and polygon have identical catalogue indices. It is ensured only in those cases that the point is truly a redundancy of the polygon. For example, the polygon of a shopping mall may contain not only its own label point, but also the points of smaller shops or other localities within the shopping mall. Removing those would present a significant loss of information and should be avoided. Figure 3.1 visualises this using the example of the city of Cottbus, Germany. Figure 3.1a is a plot of only the polygon and figure 3.1b shows both the polygon and the point with the city outlines of Cottbus highlighted and the label printed at the point. The point does not hold any additional information, because if it is within range of a hypothetical CS then the polygon enclosing it definitely is as well. For that reason, it is already known that a city is in the vicinity. By keeping both the polygon and the point in the dataset, the algorithm finds two instances of the same fclass leading to incorrect conclusions.



(a) Only city polygon without label



(b) City polygon with label point

Figure 3.1: Cottbus, Germany as example of point labels

There is one additional step in the OSM preparation process that may be conducted after the specification of the correct CRS, but before the geometry cleaning stages. It is a remnant from the past that is disabled by default and not applied in this thesis, but remains fully functional in the code. The matching between OCM and OSM data was previously done by calculating the distance between each CS and all OSM localities to find those that are within walking distance. This proved to be time-consuming, so the OSM geometries were simplified to improve efficiency. The method remained unacceptably inefficient all the same and has since been replaced. The simplification concept shall, nonetheless, be described here briefly in case an increasing amount of data justifies its use in the future. It relies on the 'st_simplifies' function and simplifies geometries with a tolerance of 'X' metres. All distances below that tolerance are smoothed out, while the general shape of the geometry is preserved. Figure 3.2 shows that functionality on the example of the *Zoologischer Stadtgarten* in Karlsruhe, Germany with figure 3.2a displaying the polygon in its original state as used in this thesis and figure 3.2b displaying the same polygon simplified with a tolerance of 75 metres.

(a) Original polygon without simplification **(b)** Simplified polygon with 75m tolerance

Figure 3.2: *Zoologischer Stadtgarten* in Karlsruhe, Germany as example of geometry simplification

3.4.3 Open Charge Map

The last data to be prepared is the OCM dataset. Corresponding to the first step in the OSM preparation, this process starts by generating SF point geometries based on the longitude and latitude data and assigning the default WGS84. The next step is again reducing duplicate and redundant data. A quick first measure in that regard is to simply filter out all duplicate rows. After that, the next step is again analogous to finding redundant geometries in the OSM dataset. For each observation, all geometries equal to the current one are calculated. Prior to that, however, the dataset needs to be sorted differently. First, by the date of the last update, then by the total number of available connections and lastly by the number of supply points at each CS. This sorting is important because for each CS the 'st_equals' function checks for similar geometries and it does that by iterating through the dataset top to bottom. Due to the sorting, the first equal geometry found is always the observation with the most recent data. If there are equal geometries that also have been updated simultaneously, the observation with the highest number of connections and subsequently charging points is selected. All other subsequent observations can safely be removed.
Next, the countries desired for the analysis can directly be filtered by their ISO code and are stored in separate per-country data frames. Analogue to the OSM geometries, the CSs are converted to the ETRS89 to allow for proper manipulation and projection. The conversion was not done prior to this step, because the original dataset contains countries around the globe for which this CRS would not be suitable. Additionally, for calculating equal geometries, the CRS is only secondary. It specifies only the projection on a flat map and does not alter the position and shape of the geometry itself.

3.5 Matching Localities

After preparing the datasets, it is time to match the OSM to the OCM data. In other words, for each charging station from OCM, it is time to retrieve all localities from OSM that are within walking distance. With the 'st_buffer' function all CSs are converted from point geometries to circular polygons, or buffer zones, with a specified radius and the CS in the centre. To make sense of the sensitivity analysis, buffer zones with three different radii are created and stored in separate variables to avoid overwriting information. In an attempt to appropriately reflect the findings in 2.2.7, these have been specified as 200 m for the *maximum convenient walking distance*, 400 m for the *maximum desirable walking distance* and 600 m for the *maximum acceptable walking distance*. The matching itself is now conducted by checking whether or not a buffer zone intersects with a locality. This method relies on the 'st_intersects' function which, contrary to the previously used 'st_distance' function, contains the option for constructing a spatial index (Pebesma and Bivand, 2017). Figure 3.3 displays the efficiency gain of using a spatial index, which reduces "the time to get results from quadratic in the number of geometries to linear in the number of geometries." (Pebesma and Bivand, 2017) The 'st_distance' function does not offer a spatial index, because it needs to cycle through all possible combinations to return a dense matrix (Pebesma and Bivand, 2017). Using 'st_intersects' does not directly facilitate a sensitivity analysis with different walking distances. However, due to the immense gain in computational time, multiple computations can be conducted with the three walking distances specified above. This matching has to be done between OCM and every OSM shape file individually, each yielding a list as the result. The resulting lists of the OSM shape files are then merged together into one per-country list and in turn used to compile a joint list representing the EU for the *cross-country* analysis later-on. This is conducted for each of the three walking distances, each resulting in four lists: One each for Germany, France, Italy and the EU reference. Each observation in those lists represents one OCM charging station and contains another list itself. These sublists contain the catalogue indices of all localities that intersect the respective CS's walking distance.

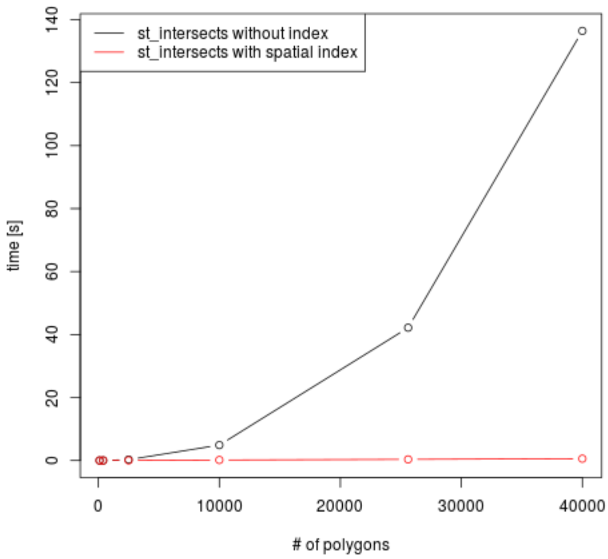


Figure 3.3: Benchmark with and without spatial index on 'st_intersects'

3.6 Generating Document-Term-Matrices

As specified in 2.6, for the LDA to work, the data needs to be present in a DTM, in which the number of occurrences of each fclass are captured per CS. Therefore, following the procedure established in Silge and Robinson (2020), the twelve lists are converted into individual DTMs with each row representing one CS, each column one fclass and each element the number of occurrences of the respective fclass for the respective CS.

However, the LDA cannot handle *documents* without any *words*. In the context of this thesis, this refers to CSs for which no locality features have been found in the vicinity. For this reason, the DTMs need to be cleaned before the LDA-based analyses can be conducted.
It is not ideal, however, to simply drop these unmatched CSs from the DTMs, because the indices of all subsequent CSs would differ from their actual indices in the OCM dataset. It is crucial, though, that the possibility of this referencing is preserved, because all CS characteristics like connection and use types as well as the creation dates remain in the OCM dataset. While removing those CSs from both DTM and OCM would solve the referencing problem, it is entirely possible for a CS to not have been matched with any localities in the *maximum convenient walking distance* of 200 m, but still have matched localities within its *maximum acceptable walking distance* of 600 m. To circumvent this problem, the OCM dataset is enlarged by the addition of three variables each representing the CS category for one of the walking distances. These categories are initiated with a default value of `'0'`. As each cell in the DTM represents the occurrences of the respective fclass, it follows that a CS which was not matched with any localities has its sum of all occurrences equal zero. Therefore, indices with a row sum of zero have their category value of `'0'` replaced by `'NA'` in the OCM dataset. Now, the unmatched CSs can be removed from the DTMs to ensure functionality of the LDA, while also leaving the index structure intact and preventing a loss of information, because for each walking distance, all CSs with a category of `'NA'` can simply be filtered out,resulting in the same index structure as the DTM. The final result are twelve DTMs, which can now be processed by the LDA for the analyses.

3.7 Within-Country Analyses

For reasons of consistency, the different *within-country* analyses all follow the same procedure as described in the following. This section is only concerned with the *within-country* analyses, which means the number of categories may differ across the researched cases.

3.7.1 Number of categories

As specified in 2.6, multiple iterations of the LDA are conducted with different numbers of categories and evaluated regarding their log-likelihood to find the appropriate number of categories. For this purpose, four log-likelihood graphs are provided: One for each walking distance, plus a fourth combining the three individual curves, which allows to draw a conclusion regarding which walking distance is best suited to describe the given data. The visualisation of this appropriate number is a (local) maximum or, alternatively, a visible *"bend"* in the course of the log-likelihood curve after which more categories would not significantly increase the log-likelihood value.
After knowing the appropriate number of categories to represent all CSs, the predominant category of each CS is extracted and assigned to the corresponding category variable created in 3.6. The predominant category of a CS is the one best suited to describe the majority of the matched localities. Although this step negates the advantage

of mixed memberships, it is necessary for the characteristics-based analysis for each category.

3.7.2 Location-based analysis

The mixed-memberships, however, are still used for the location-based analysis and visualised in so-called *scatterpie* charts as described in Duplan (2019). These rely on *t-Distributed Stochastic Neighbour Embedding* to reduce the mixed-memberships towards categories into an "easily-visualizable 2-dimensional space." (Duplan, 2019) At this point, the terms *category* and *topic* need to be differentiated for better understanding. *Category* refers to the semantic group of CSs, for example service areas along motorways. A *topic* refers to the number this semantic group has in the context of the respective walking distance. In other words, one *category* may for example be represented by *topic1* for 200 m, *topic2* for 400 m and *topic9* for the 600 m walking distance. Due to the entirely independent processing of the walking distances, which possibly contain a differing number of categories, it is furthermore possible for a *category* to not have a corresponding *topic* in each walking distance.
In the scatterpie chart, each CS is represented as a pie whose sections represent the degree of membership towards the corresponding topics, while its size is proportional to the degree of membership towards its predominant topic (Duplan, 2019). Their position in the scatterpie chart provides an impression of how CSs are dispersed between the topics, as individual CSs "are pulled in the direction of the "center of mass" of another topic they may also belong to." (Duplan, 2019) This initial impression of relationships between topics is backed by the computation of correlation plots to have this in clear and precise metrics. In line with Duplan (2019), all CSs whose most prominent membership is less than one third, are filtered out to reduce visual noise.
As an example, a CS that belongs entirely to topic 1 is uniformly coloured and of full size. On the other hand, a CS that has mixed memberships towards two topics, with its most predominant membership being 60% towards topic 1, is displayed as a pie chart with a size equal to 60% of full size and two pie sections, 60% and 40%. While being smaller than the other example, it is also pulled towards the area of the topic to which it belongs with 40%. Finally, a CS that belongs to five topics with 20% each, is filtered from the chart entirely.
Aside from this, one comprehensive table is provided for each country. These tables contain all three walking distances and display the top-5 fclasses as well as the absolute number of CSs belonging predominantly to the respective topic to facilitate drawing conclusions regarding the overall topic importance. Additionally to this, each fclass is directly followed by an importance value to its right. These values represent how many of all retrieved localities around CSs belonging predominantly to this topic were of the respective fclass, providing an indication on the balancing within the topic. The sum of CSs belonging predominantly to each topic is also followed by a percentage value representing their share in relation to the total number of existing CSs.
The degree of membership of CSs towards their predominant topics can give an indication of how capable those are of representing all fclasses found in the vicinity of the CSs. For this reason, the mean degree of membership is calculated for each walking distance

to once again enable the drawing of conclusion regarding which walking distance is best suited to describe the given data.

3.7.3 Characteristics-based analysis

As this thesis is concerned with the development of CI over time, three different time series are computed. While the first shows the annual expansion separated by topic, the second shows this expansion in an aggregated version. The third displays the expansion of each topic relative to its largest annual expansion in a violin chart. For the time aspect, the date at which each CS was added to the OCM database, `'DateCreated'`, is used. To simplify readability and enable faster drawing of conclusions, the colours of the various topics are consistent across the scatterpie and the three time series.

3.7.4 Germany

After the preparation and cleaning stages, Germany has 12,585 CSs offering 33,161 connections for simultaneous charging across a total of 30,903 supply points.

3.7.4.1 Number of categories

Figure 3.4a displays the log-likelihood curve for 200 m and holds one particular point of interest at 13 categories. Up to this, the overall increase is significantly stronger than for higher numbers. The 400 m curve in figure 3.4b follows a smoother course, which complicates identifying a point of interest. However, at a number of 12 categories, the general incline starts to flatten again making it a suitable candidate for this distance. For the last walking distance, the graph in figure 3.4c includes a local maximum at 13 categories, which simultaneously is another changing point for the overall curve incline. The joint graph (figure 3.4d) indicates that the categorisation generally works best for the 200 m walking distance. For Germany, this lead to the decision of 13, 12 and 13 categories for the three respective walking distances.

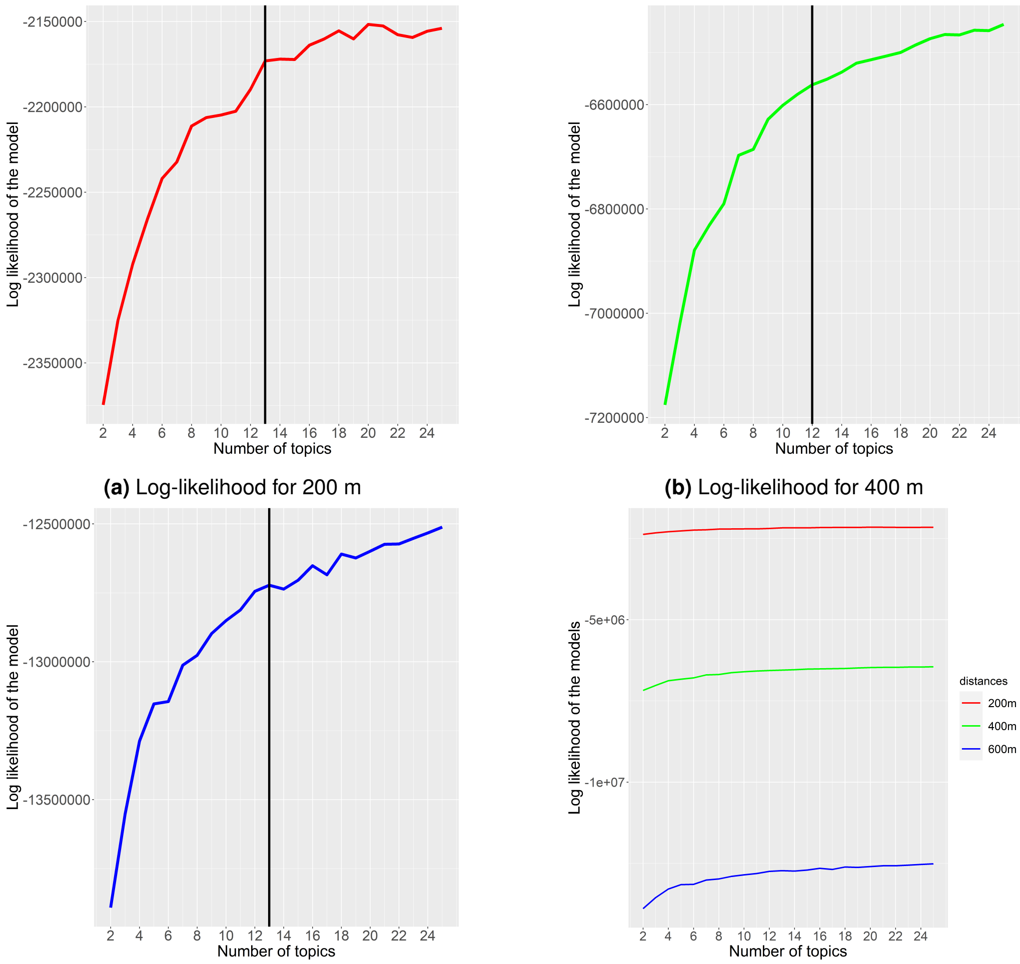
3.7.4.2 Location-based analysis

The mean degree of membership towards the predominant topic is relatively stable across the walking distances with 46%, 47% and 47%, respectively.
The comprehensive overview for Germany is provided in table 3.3 and shows two topics that dominate each walking distance by their number of assigned CSs. This twofold dominance is also visible in the scatterpie charts in figures 3.5e, 3.6e and 3.7e. *Topic9* for 200 m, *Topic11* for 400 m and *Topic10* for 600 m is one of those two. From the scatterpie charts, it is clear that these three topics follow similar structures with a relatively loose core, a far reach towards other categories and a broad, fanned out position of their CSs. Having the `'roads_residential'` fclass owning the majority of localities (69%, 57% and 53%, respectively) indicates that the CSs are located on roads in residential areas. This is further specified by the `'group_parking'` fclass with 4%, 12% and 14%, respectively. The fact that their structures follow the same patterns and their

fclasses match, implies that these three topics belong to one category. These are located at public parking areas at roads in residential areas.
The second dominant observation is *Topic5* for 200 m, *Topic10* for 400 m and *Topic13* for 600 m. Although they have a similarly far reach, they differ from the previous by having a denser core and tighter fan pattern, yet are again similar to each other across the walking distances. This fact again makes it likely that the three topics belong to the same category. With at least 89%, all three topics are absolutely dominated by `'group_residential_building'`, which represents any form of private housing. The fact that public parking areas are only present with 1% in the 400 m and 600 m distances, indicates that these CSs possibly facilitate road-side charging, as individual parking spaces along roads are not listed as parking features in the Geofabrik export (Topf and Ramm (2019) and OSM (2019a)). The overwhelming presence of private households further indicates densely populated urban areas.
One set of topics is striking due to being consistently small in structure and also having their clusters in the scatterpie charts largely independent from neighbouring topics. This set is represented by *Topic10* for 200 m, *Topic4* for 400 m and *Topic4* for 600 m and is located near private garages (around 90%) in residential areas. The low presence of residential buildings indicates that these are possibly conglomerates of garages belonging to separated apartment blocks in larger cities. The presence of turning circles, which are circular ends of residential roads (OSM, 2019a) further support that theory.
Two other categories with a medium reach have a visible relationship in the scatterpie charts and appear to be merged for the 400 m distance. These are parking areas along motorways (*Topic2* for 200 m, *Topic2* for 400 m and *Topic6* for 600 m) and rural tracks in a natural surrounding (*Topic3* for 200 m, *Topic2* for 400 m and *Topic12* for 600 m). Traffic-calmed areas at pedestrian roads (*Topic1* for 200 m, *Topic7* for 400 m and *Topic1* for 600 m) and at living streets (*Topic4*, *Topic9* and *Topic11*, respectively), which are both occasionally in vicinity of memorials or tourist infos are further sets with matching fclasses and similar structure across the walking distances.
Aside from these, there is also a number of smaller categories often lodged in-between others in the centre. One of them, *Topic8* for 200 m and *Topic8* for 400 m, is located on parking lots near commercial buildings (e.g. supermarkets). Another is consistently placed at storage units often in the vicinity of parking and agricultural localities and represented by *Topic7*, *Topic5* and *Topic5*, respectively. Furthermore, there are three categories primarily characterised by the type of road they are at. These are *Topic13*, *Topic3* and *Topic7*, respectively for public parking near primary *(national)* roads, *Topic6*, *Topic12* and *Topic9*, respectively for public parking near secondary *(regional)* roads and *Topic1* for 400 m and *Topic2* for 600 m for public parking near local roads.
Although the number of categories differs across the three walking distances, the scatterpie charts follow roughly the same structure and there appear to be similar relationships between the categories. The correlation plots (figures 3.5b, 3.6b and 3.7b) suggest mostly negative correlations between categories with only a few exceptions. These slightly positive correlations are for example between categories featuring `'roads_pedestrian'` and `'roads_living_street'` which seems logical because both are referring to traffic-calmed areas. Nonetheless, they were not grouped in the semantic grouping conducted earlier in this thesis as they do serve different purposes.

3.7.4.3 Characteristics-based analysis

Figures 3.5c, 3.6c and 3.7c display the annual expansion of CI for their respective walking distance. The most striking observation across all three figures is the sharp uptake in CSs for all categories in 2015. Prior to this date, there appears to be hardly any activity in CI development. Equally noteworthy is the decline in the construction of new CSs starting 2019. Interestingly, most categories follow a mostly similar course, however, there is one category with a more atypical development, as seen in figures 3.5a, 3.6a and 3.7a, respectively. While nearly all other categories have their peak in 2015, this category is way less pronounced in the initial CI boost and mostly follows a relatively steady increase only to amplify in 2017. Additionally, it is one of the few that achieves further growth after 2019. This category are the parking areas along motorways represented by *Topic2*, *Topic2* and *Topic6*, respectively. One further category with an atypical course is only present in the 600 m distance, *Topic8*. These CSs are located at public places around memorials in residential conglomerates. For 2015 and 2016, this category follows the general flow, but in 2017 has a sharp increase in its expansion. This expansion boost in 2017 is shared by most other categories as well, although less pronounced.



(a) Log-likelihood for 200 m

(b) Log-likelihood for 400 m

(c) Log-likelihood for 600 m

(d) Log-likelihood for all distances

Figure 3.4: Log-likelihoods for Germany

Table 3.3: Overview of German location categories

		Locality Feature 1		Locality Feature 2		Locality Feature 3		Locality Feature 4		Locality Feature 5		**Sum of CS/topic**	
200 m walking distance	Topic1	roads_pedestrian	33%	pois_tourist_info	10%	pois_hairdresser	6%	roads_residential	6%	group_parking	5%	**677**	**5%**
	Topic2	roads_tertiary	53%	group_parking	19%	roads_motorway	9%	transport_bus_stop	4%	landuse_scrub	3%	**856**	**7%**
	Topic3	landuse_forest	19%	roads_track	16%	landuse_meadow	16%	landuse_scrub	14%	group_parking	9%	**1293**	**10%**
	Topic4	roads_living_street	75%	group_parking	5%	group_religious_building	5%	building_greenhouse	4%	pois_tourist_info	1%	**104**	**1%**
	Topic5	group_residential_building	90%	roads_residential	5%	landuse_residential	2%	group_kindergarten	0%	pois_hairdresser	0%	**2651**	**21%**
	Topic6	roads_secondary	67%	transport_bus_stop	6%	group_parking	5%	roads_trunk	4%	landuse_residential	4%	**861**	**7%**
	Topic7	building_garages	86%	building_barn	6%	transport_bus_stop	1%	landuse_residential	1%	group_parking	1%	**44**	**0%**
	Topic8	group_parking	25%	building_commercial	20%	building_industrial	9%	landuse_commercial	8%	building_retail	7%	**1188**	**9%**
	Topic9	roads_residential	69%	landuse_residential	5%	transport_bus_stop	5%	group_religious_building	4%	group_parking	4%	**2869**	**23%**
	Topic10	group_garage	92%	roads_residential	4%	traffic_turning_circle	1%	transport_bus_stop	1%	landuse_residential	1%	**322**	**3%**
	Topic11	group_parking	27%	group_park	20%	roads_residential	17%	group_school	10%	landuse_residential	5%	**1111**	**9%**
	Topic12	pois_memorial	46%	building_office	13%	group_university	10%	pois_vending_parking	5%	group_hospital	5%	**156**	**1%**
	Topic13	roads_primary	71%	roads_residential	5%	landuse_residential	4%	transport_bus_stop	4%	group_parking	2%	**445**	**4%**
400 m walking distance	Topic1	roads_tertiary	59%	group_parking	11%	roads_trunk	7%	transport_bus_stop	7%	roads_residential	4%	**319**	**3%**
	Topic2	roads_track	19%	landuse_forest	16%	landuse_meadow	15%	roads_motorway	10%	landuse_scrub	10%	**1770**	**14%**
	Topic3	roads_primary	64%	roads_residential	10%	group_parking	7%	transport_bus_stop	4%	group_hotel	1%	**304**	**2%**
	Topic4	group_garage	90%	group_residential_building	3%	roads_residential	3%	traffic_turning_circle	1%	transport_bus_stop	1%	**495**	**4%**
	Topic5	building_garages	63%	group_parking	10%	roads_residential	6%	group_garage	5%	building_barn	2%	**75**	**1%**
	Topic6	landuse_residential	18%	group_park	13%	group_parking	12%	landuse_scrub	11%	group_university	5%	**260**	**2%**
	Topic7	roads_pedestrian	24%	pois_memorial	17%	roads_residential	8%	pois_hairdresser	6%	pois_vending_parking	5%	**771**	**6%**
	Topic8	building_commercial	24%	group_parking	21%	building_industrial	15%	building_retail	7%	landuse_commercial	5%	**569**	**5%**
	Topic9	roads_living_street	28%	pois_tourist_info	21%	group_parking	10%	roads_residential	8%	group_hotel	5%	**290**	**2%**
	Topic10	group_residential_building	89%	roads_residential	5%	group_parking	1%	group_school	1%	transport_bus_stop	0%	**3499**	**28%**
	Topic11	roads_residential	57%	group_parking	12%	transport_bus_stop	5%	group_religious_building	4%	group_school	3%	**3808**	**30%**
	Topic12	roads_secondary	66%	group_parking	8%	transport_bus_stop	6%	roads_residential	2%	landuse_residential	2%	**417**	**3%**
600 m walking distance	Topic1	roads_pedestrian	21%	roads_residential	10%	pois_tourist_info	8%	group_parking	6%	pois_vending_parking	5%	**805**	**6%**
	Topic2	roads_tertiary	55%	roads_residential	14%	transport_bus_stop	7%	group_parking	7%	landuse_residential	2%	**136**	**1%**
	Topic3	group_parking	20%	group_university	19%	group_hospital	9%	group_residential_building	6%	group_park	5%	**97**	**1%**
	Topic4	group_garage	89%	roads_residential	3%	group_residential_building	2%	traffic_turning_circle	1%	group_parking	1%	**633**	**5%**
	Topic5	building_garages	54%	group_parking	12%	roads_residential	9%	transport_bus_stop	3%	landuse_allotments	3%	**102**	**1%**
	Topic6	group_parking	20%	building_commercial	14%	roads_motorway	14%	building_industrial	14%	landuse_scrub	5%	**1012**	**8%**
	Topic7	roads_primary	64%	roads_residential	10%	group_parking	4%	transport_bus_stop	4%	landuse_residential	2%	**172**	**1%**
	Topic8	pois_memorial	26%	group_residential_building	17%	landuse_residential	14%	group_park	5%	pois_hairdresser	4%	**197**	**2%**
	Topic9	roads_secondary	68%	transport_bus_stop	6%	roads_residential	5%	group_parking	3%	landuse_residential	2%	**213**	**2%**
	Topic10	roads_residential	53%	group_parking	14%	transport_bus_stop	5%	group_school	3%	group_religious_building	3%	**3547**	**28%**
	Topic11	roads_living_street	38%	group_parking	14%	landuse_scrub	14%	roads_trunk	13%	group_park	4%	**132**	**1%**
	Topic12	roads_track	23%	landuse_meadow	16%	landuse_forest	15%	roads_residential	10%	landuse_scrub	7%	**1655**	**13%**
	Topic13	group_residential_building	89%	roads_residential	5%	group_parking	1%	building_commercial	1%	transport_bus_stop	1%	**3878**	**31%**

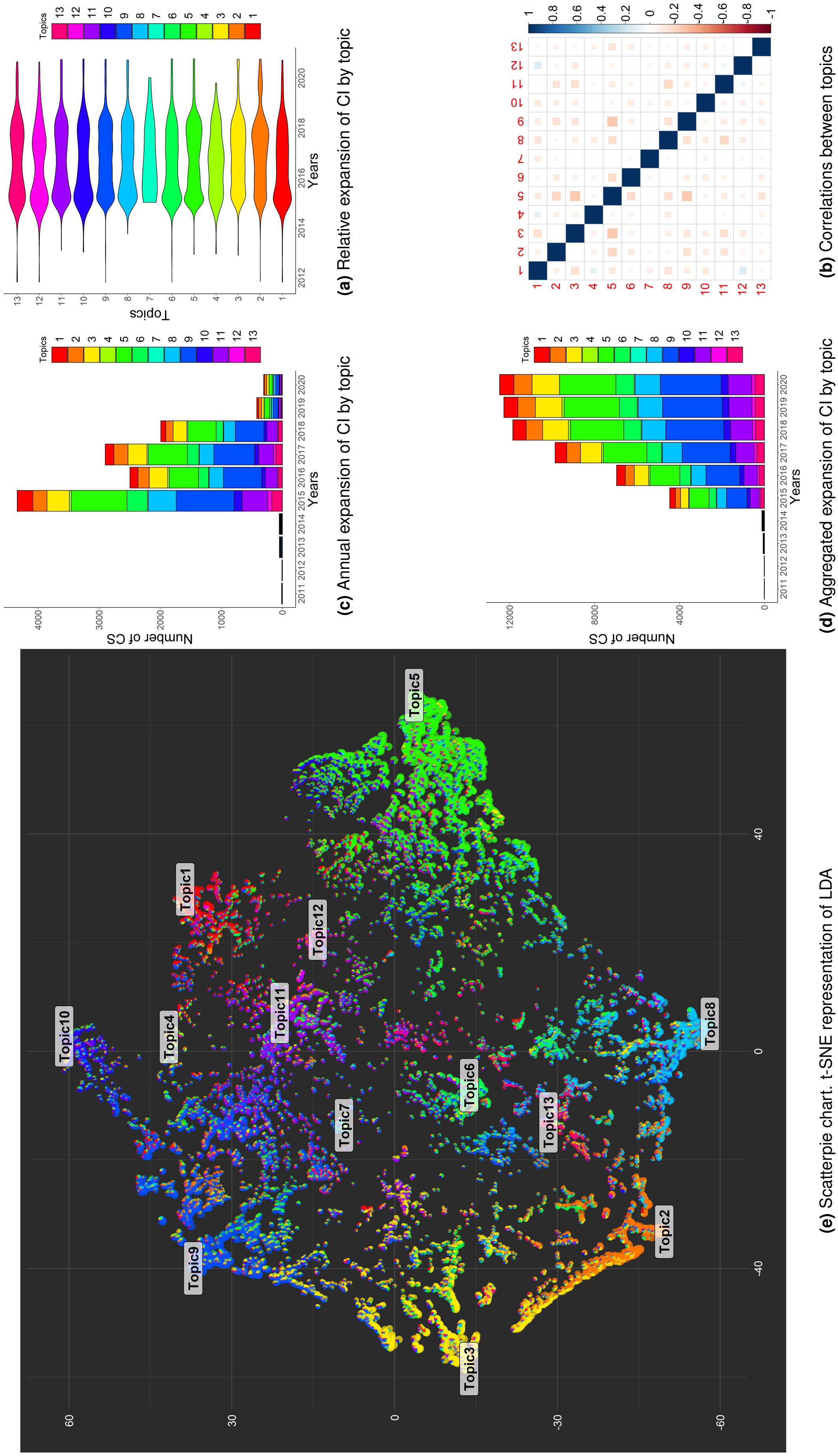


Figure 3.5: Maximum convenient walking distance (DE 200 m)

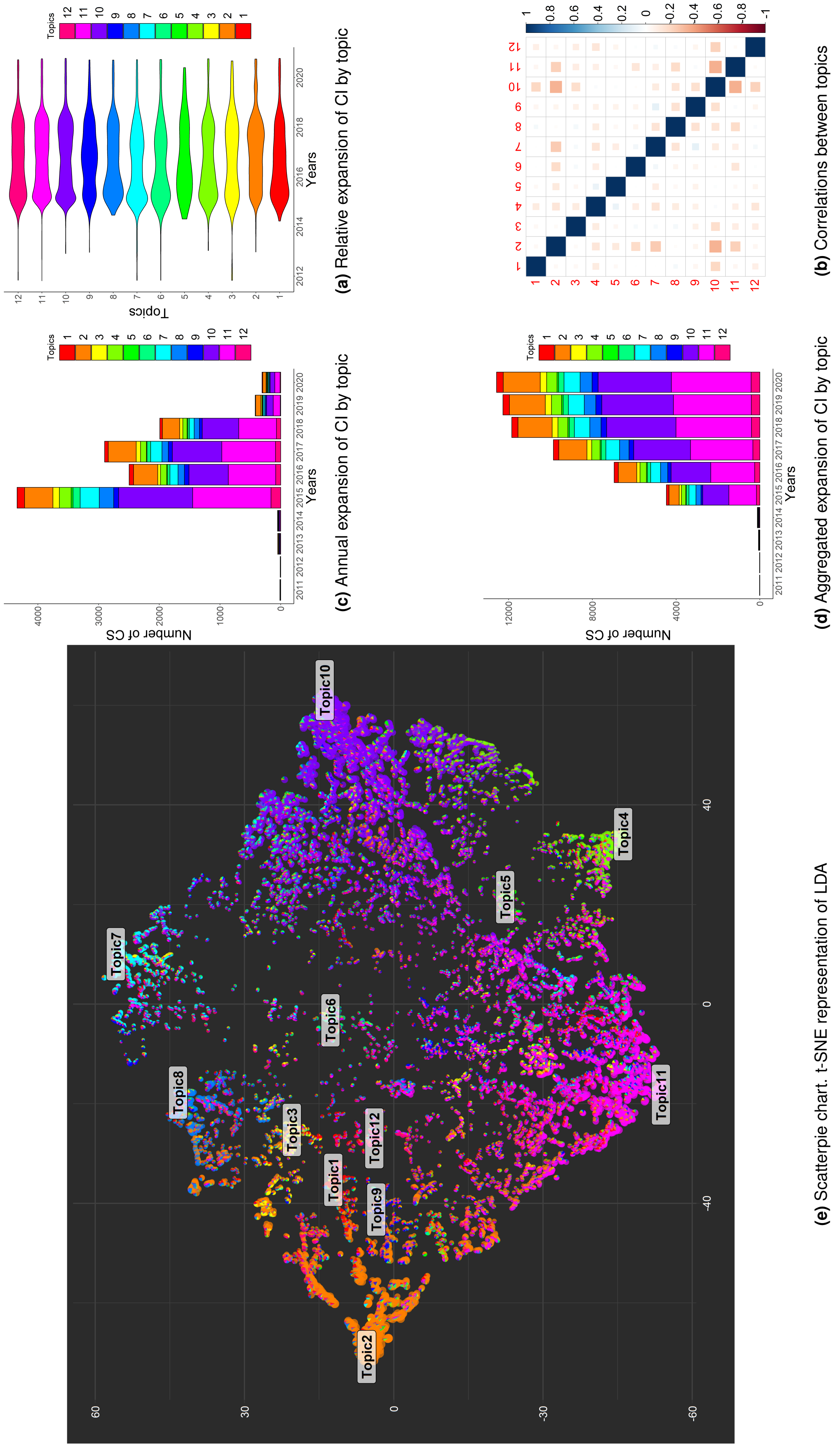


Figure 3.6: Maximum desirable walking distance (DE 400 m)

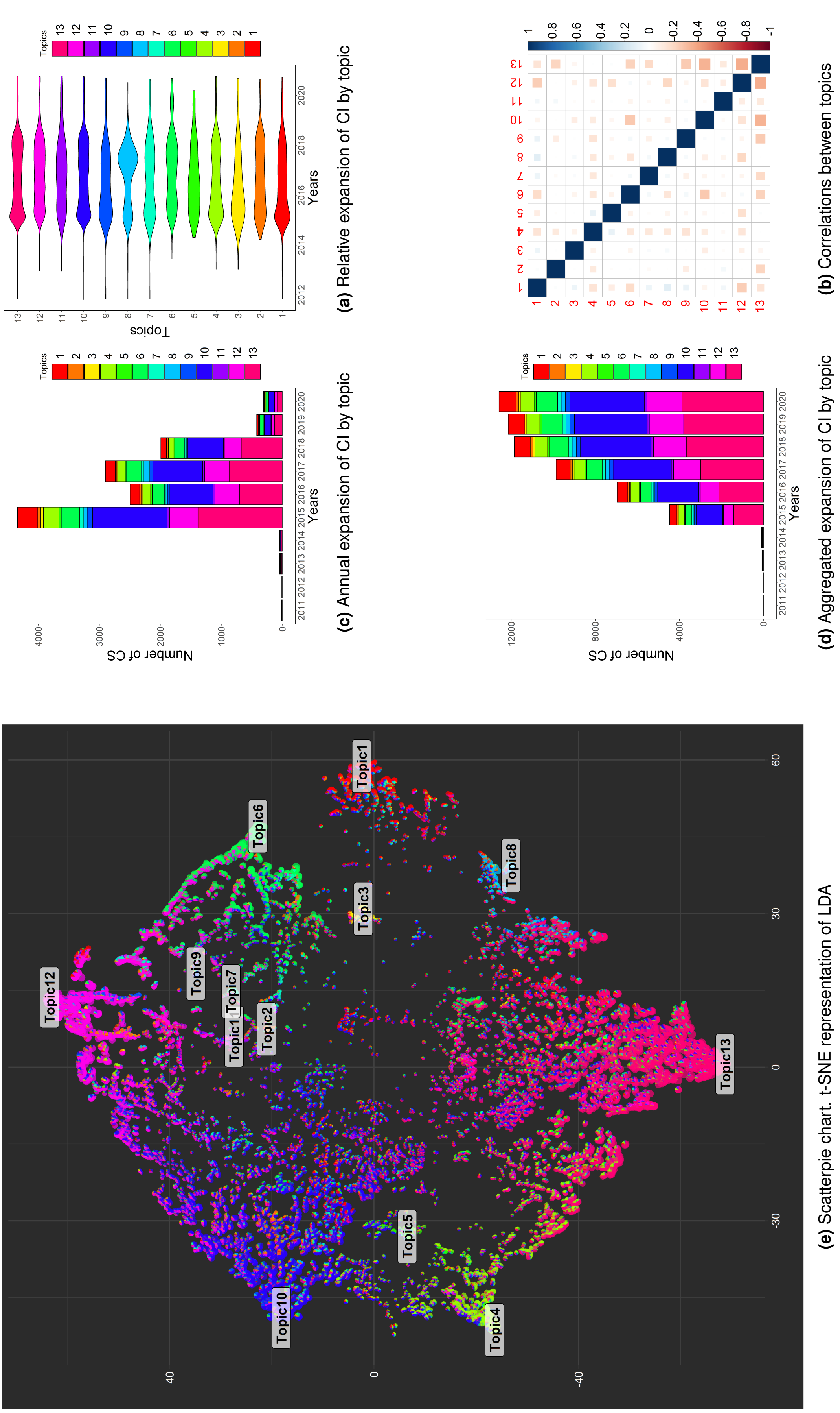


Figure 3.7: Maximum acceptable walking distance (DE 600 m)

3.7.5 France

After the preparation and cleaning stages, France has 8,400 CSs offering 4,588 connections for simultaneous charging across a total of 3,644 supply points. The discrepancy in the data is due to the majority of French CSs not having their `'numberOfPoints'` or connection `'quantities'` specified.

3.7.5.1 Number of categories

The 200 m distance, as displayed in figure 3.8a, is characterised by a declining log-likelihood with higher numbers of categories. Therefore, 9 categories are chosen, as they present a local maximum and tipping point in the overall course. Figure 3.8b represents the 400 m distance and follows a smoother development. For this curve, 10 categories stand out as the break from an overall steep to a more stale increase in log-likelihood. Lastly, the 600 m curve (figure 3.8c) contains a local maximum at 14 categories, which stands out in the overall course. Figure 3.8d shows the log-likelihoods of all distances in relation to each other and identifies 200 m as best suited for the given data. Therefore, the French CI is categorised into 9, 10 and 14 categories for the three walking distances.

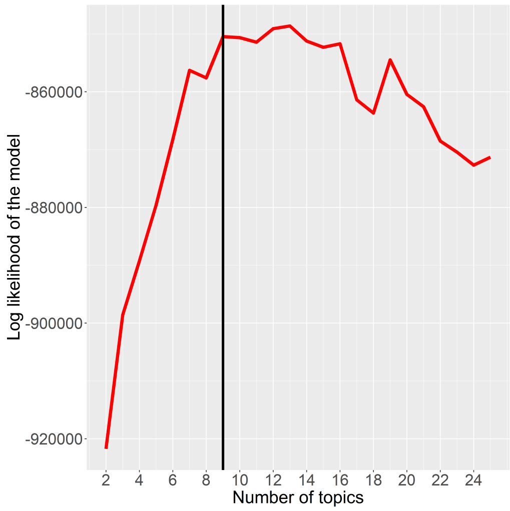
3.7.5.2 Location-based analysis

The mean degree of membership for the dominant topics is relatively stable across the distances with 49%, 46% and 42%, respectively.
Table 3.4 holds the comprehensive overview for the French results, which provides the first striking observation. Based on the number of CSs, one topic entirely overshadows the others for each walking distance. These topics hold 51%, 46% and 43% of all French CSs for the three respective walking distances and are mainly represented by `'roads_residential'`, which represents public parking areas at roads in residential areas. The scatterpie charts (figures 3.9e, 3.10e and 3.11e) clearly reflect that overshadowing, with one topic each expressing an immense gravitational pull. Compared to their far reach, the individual cores of *Topic2*, *Topic9* and *Topic6*, respectively are disproportionally small.
Three categories remain rather distinct in their structure across all three scatterpie charts. The first is located at parking areas along motorways, as represented by *Topic4*, *Topic2* and *Topic13*, respectively. The second is directly next to residential buildings and private garages as represented by *Topic9*, *Topic3* and *Topic7*, respectively. Lastly, the third of the group is located on traffic-calmed roads and in the vicinity of hotels or tourist information as represented by *Topic7*, *Topic6* as well as the pair of *Topic2* and *Topic3*, respectively.
Another striking observation from the scatterpie charts is the relationship of two categories. The first of the two, *Topic3*, *Topic8* and *Topic14*, respectively, is characterised by natural surroundings and religious buildings. The second is represented by *Topic5*, *Topic1* and *Topic5*, respectively. This category is characterised by tertiary *(local)* roads, but also by the presence of religious buildings.
A small, but rather distinct category is represented by *Topic6*, *Topic7* and *Topic1*, re-

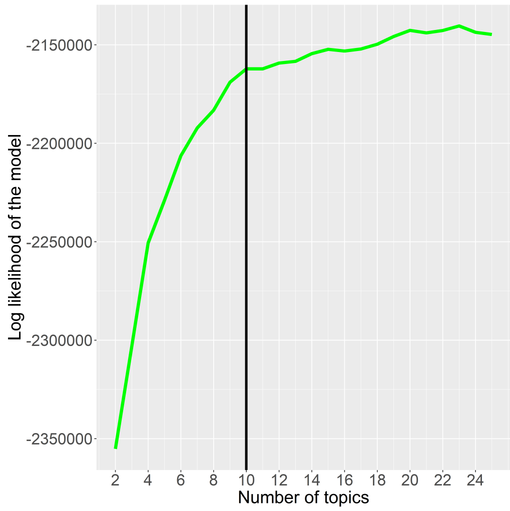
spectively and placed at bus stops near hotels or parks. Furthermore, two categories are again dominated by the type of road they are located at, which are primary *(national)* (*Topic1*, *Topic10* and *Topic4*, respectively) and secondary *(regional)* roads, as represented by *Topic8*, *Topic5* and *Topic8*, respectively.
The most prominent, slightly negative correlation is between the parking areas along motorways and public parking areas at roads in residential areas. The higher number of categories in the 600 m distance also prompted some slightly positive correlations. One pair are living and pedestrian streets (*Topic2* and *Topic3*), which for the other walking distances were merged in one category. Another slightly positive correlation exists from *Topic1*, *Topic2* and *Topic3* towards *Topic12*, which facilitates road-side charging as indicated by the presence of vending machines for parking tickets and the absence of public parking areas.

3.7.5.3 Characteristics-based analysis

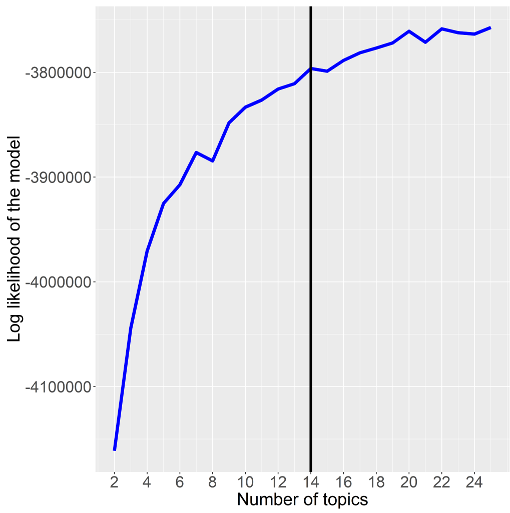
Figures 3.9c, 3.10c and 3.11c display the development of each topic over time. The most striking observation is a staggering increase in CI growth in 2020. Aside from that, there are two additional timeslots prompting overall growth, 2013 and 2017. According to a discussion between an OCM contributor and Christopher Cook, the creator of OCM, an entire French regional CI network was missing from the database (OCM and Cook, 2020). This was due to a faulty import from the official national database as provided by the French government and has been fixed on the 27th of August 2020. As mentioned previously, the time factor in these statistics relies on `'DateCreated'`, which is the date at which the CS was added to the database and not necessarily the date at which it was constructed. As a result of this, even though the CSs are now fully available, no conclusions can be drawn in retrospect as all CSs included in that bulk-import have the same creation date. Additionally, due to periodic downtimes of CI expansion in earlier time periods, it is uncertain whether or not that problem occurred in the past, which is why all further interpretation of the development of French CI over time has to be stopped as the validity can not be ensured. The way `'DateCreated'` works, further means that reliable time-based conclusions of French CI in this form can never be drawn based on OCM data.



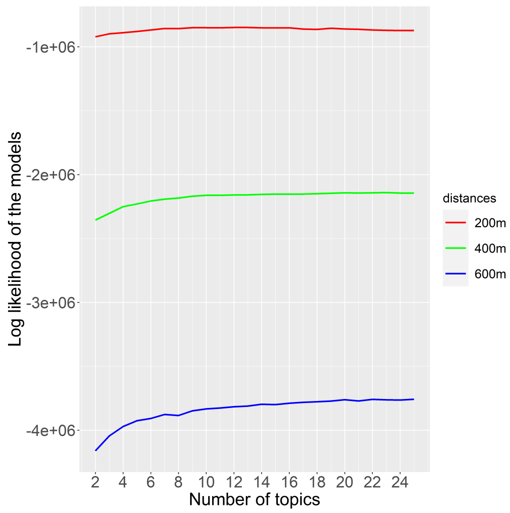
(a) Log-likelihood for 200 m



(b) Log-likelihood for 400 m



(c) Log-likelihood for 600 m

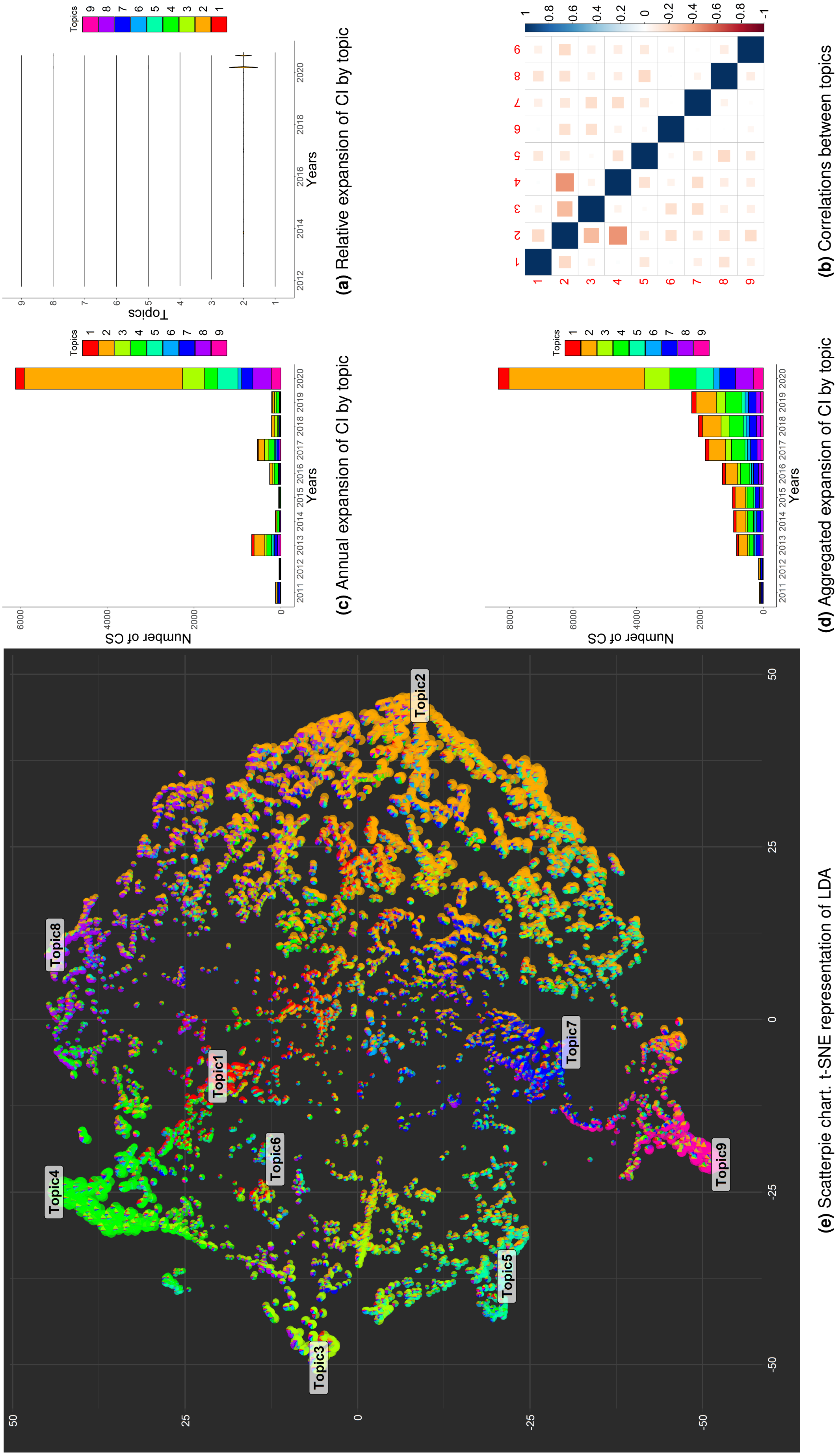


(d) Log-likelihood for all distances

Figure 3.8: Log-likelihoods for France

Table 3.4: Overview of French location categories

		Locality Feature 1		Locality Feature 2		Locality Feature 3		Locality Feature 4		Locality Feature 5		**Sum of CS/topic**	
200 m walking distance	Topic1	roads_primary	86%	landuse_residential	3%	group_parking	2%	group_supermarket	1%	landuse_industrial	1%	**341**	**4%**
	Topic2	roads_residential	73%	group_parking	9%	group_religious_building	4%	landuse_residential	4%	group_school	3%	**4278**	**51%**
	Topic3	landuse_forest	18%	roads_track	18%	group_religious_building	16%	landuse_meadow	10%	group_graveyard	10%	**799**	**10%**
	Topic4	group_parking	50%	roads_motorway	9%	building_retail	8%	roads_trunk	6%	traffic_fuel	4%	**823**	**10%**
	Topic5	roads_tertiary	75%	group_religious_building	7%	group_parking	5%	landuse_residential	4%	group_school	4%	**559**	**7%**
	Topic6	transport_bus_stop	46%	group_hotel	11%	group_park	11%	pois_hairdresser	3%	pois_bank	3%	**190**	**2%**
	Topic7	roads_pedestrian	30%	pois_tourist_info	9%	roads_living_street	9%	pois_bank	6%	pois_hairdresser	6%	**493**	**6%**
	Topic8	roads_secondary	79%	group_parking	8%	landuse_residential	4%	roads_residential	1%	group_school	1%	**566**	**7%**
	Topic9	group_residential_building	87%	group_garage	4%	building_garages	2%	roads_residential	2%	group_parking	1%	**313**	**4%**
200 m walking distance	Topic1	roads_tertiary	64%	roads_residential	14%	transport_bus_stop	7%	group_religious_building	3%	group_parking	2%	**556**	**7%**
	Topic2	group_parking	31%	roads_motorway	14%	roads_trunk	12%	building_retail	9%	building_industrial	3%	**663**	**8%**
	Topic3	group_residential_building	86%	group_garage	5%	roads_residential	3%	building_garages	1%	group_parking	1%	**434**	**5%**
	Topic4	group_parking	41%	roads_residential	15%	group_park	9%	group_school	8%	transport_bus_stop	3%	**591**	**7%**
	Topic5	roads_secondary	71%	roads_residential	10%	transport_bus_stop	7%	group_parking	3%	landuse_residential	2%	**493**	**6%**
	Topic6	roads_pedestrian	45%	roads_living_street	15%	roads_residential	5%	group_religious_building	4%	pois_bank	3%	**281**	**3%**
	Topic7	transport_bus_stop	12%	group_hotel	11%	pois_hairdresser	8%	pois_tourist_info	8%	pois_vending_parking	7%	**238**	**3%**
	Topic8	roads_track	25%	landuse_forest	17%	landuse_meadow	10%	group_religious_building	9%	landuse_residential	9%	**929**	**11%**
	Topic9	roads_residential	83%	group_religious_building	3%	traffic_turning_circle	2%	group_parking	2%	transport_bus_stop	2%	**3846**	**46%**
	Topic10	roads_primary	74%	transport_bus_stop	8%	roads_residential	5%	landuse_residential	2%	group_parking	2%	**337**	**4%**
200 m walking distance	Topic1	transport_bus_stop	46%	roads_tertiary	13%	roads_residential	11%	group_parking	9%	group_park	6%	**94**	**1%**
	Topic2	roads_living_street	15%	roads_residential	13%	group_hotel	10%	pois_hairdresser	7%	transport_bus_stop	6%	**303**	**4%**
	Topic3	roads_pedestrian	48%	pois_tourist_info	9%	group_park	3%	group_religious_building	3%	pois_attraction	3%	**180**	**2%**
	Topic4	roads_primary	79%	transport_bus_stop	4%	group_parking	2%	roads_residential	2%	group_supermarket	2%	**228**	**3%**
	Topic5	roads_tertiary	71%	roads_residential	11%	group_religious_building	4%	group_graveyard	2%	group_parking	2%	**359**	**4%**
	Topic6	roads_residential	87%	traffic_turning_circle	3%	group_religious_building	2%	group_school	1%	landuse_residential	1%	**3619**	**43%**
	Topic7	group_residential_building	93%	roads_residential	4%	group_parking	1%	group_school	0%	group_park	0%	**466**	**6%**
	Topic8	roads_secondary	67%	roads_residential	20%	transport_bus_stop	3%	group_school	2%	group_parking	1%	**521**	**6%**
	Topic9	group_garage	50%	building_garages	18%	building_office	5%	group_parking	4%	group_public_building	3%	**21**	**0%**
	Topic10	group_parking	24%	roads_trunk	22%	building_retail	13%	building_industrial	6%	building_commercial	4%	**273**	**3%**
	Topic11	group_parking	37%	roads_residential	29%	group_park	6%	group_school	6%	pois_bank	2%	**802**	**10%**
	Topic12	pois_vending_parking	26%	roads_residential	12%	pois_hairdresser	8%	group_residential_building	5%	transport_tram_stop	5%	**36**	**0%**
	Topic13	roads_motorway	28%	group_parking	23%	landuse_residential	20%	landuse_scrub	7%	landuse_forest	5%	**366**	**4%**
	Topic14	roads_track	31%	landuse_forest	18%	landuse_meadow	11%	group_religious_building	7%	landuse_residential	4%	**1101**	**13%**



(a) Relative expansion of CI by topic

(b) Correlations between topics

(c) Annual expansion of CI by topic

(d) Aggregated expansion of CI by topic

(e) Scatterpie chart. t-SNE representation of LDA

Figure 3.9: Maximum convenient walking distance (FR 200 m)

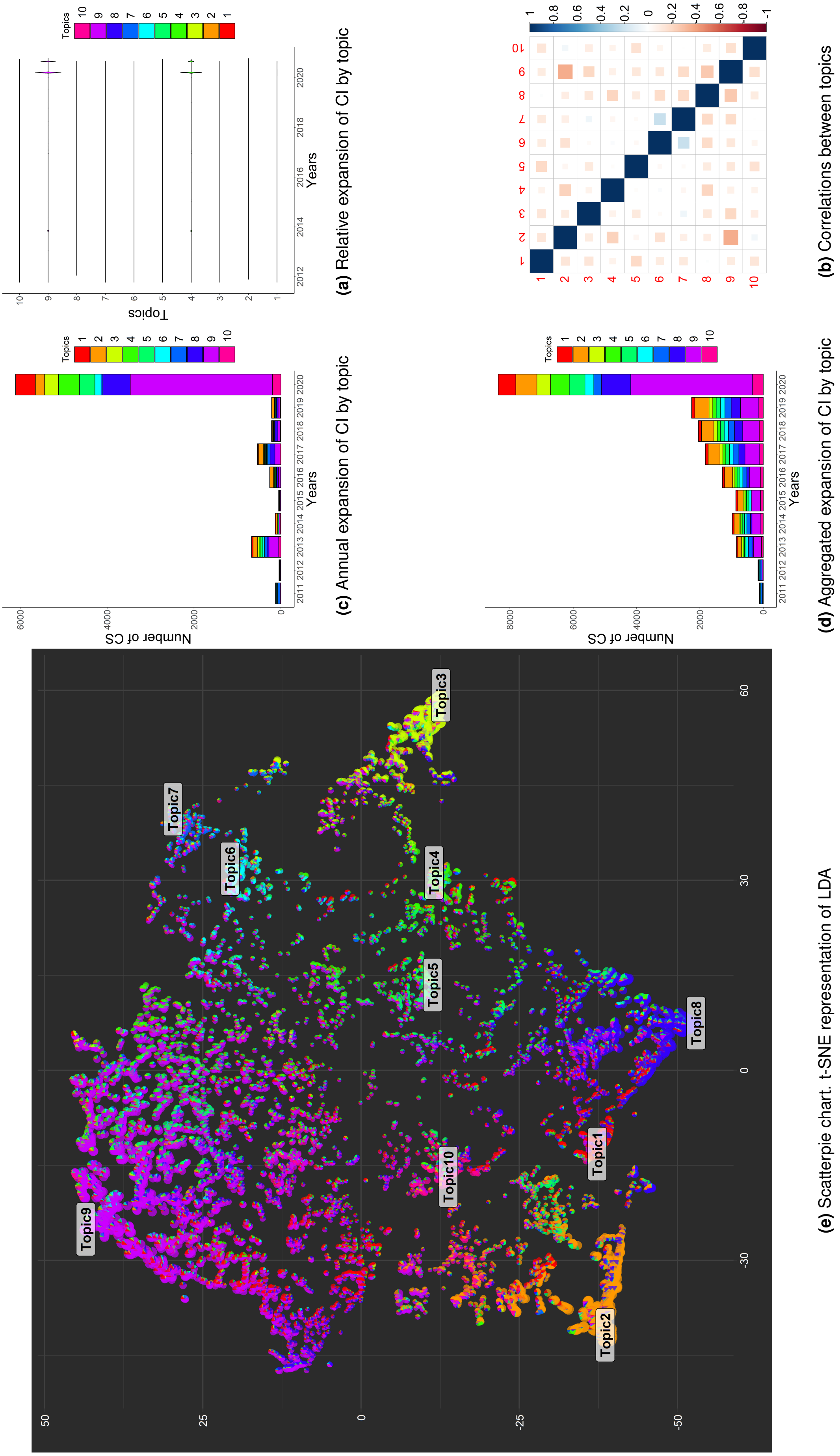
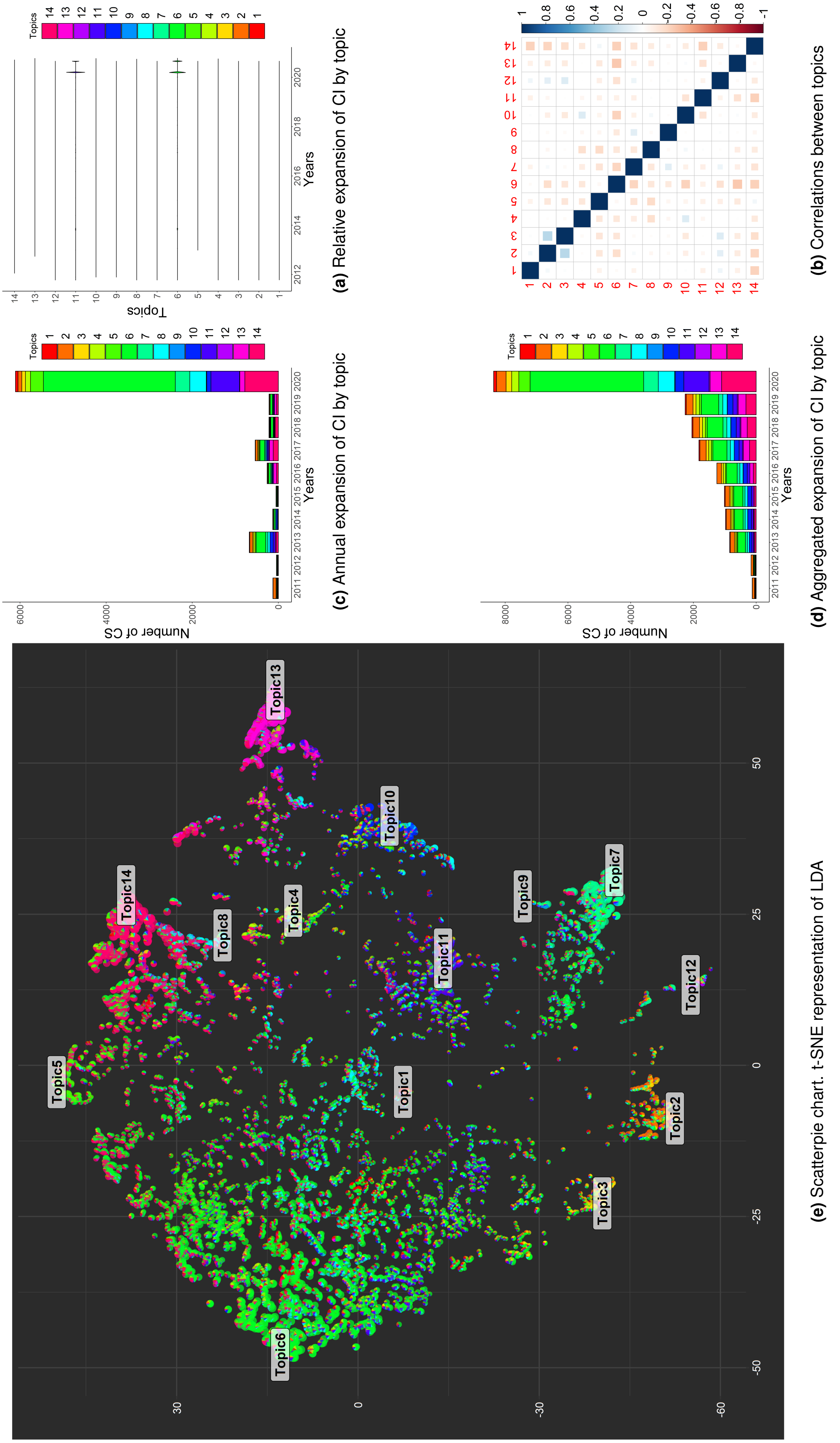


Figure 3.10: Maximum desirable walking distance (FR 400 m)



(a) Relative expansion of CI by topic

(b) Correlations between topics

(c) Annual expansion of CI by topic

(d) Aggregated expansion of CI by topic

(e) Scatterpie chart. t-SNE representation of LDA

Figure 3.11: Maximum acceptable walking distance (FR 600 m)

3.7.6 Italy

After the preparation and cleaning stages, Italy has 6,687 CSs offering 15,943 connections for simultaneous charging across a total of 14,203 supply points.

3.7.6.1 Number of categories

Figure 3.12a contains two potential points of interest: 10 and 15 categories. As the incline in log-likelihood starts to slow down after 10 categories, this number was ultimately decided as fewer, but more concise categories are considered advantageous. The overall course of figure 3.12b for 400 m, is a lot smoother, but also contains a bend at 10 categories in its curve, which is also the case for the third walking distance. Figure 3.12d shows the log-likelihoods of all distances in relation to each other and identifies 200 m as best suited for the given data. Ultimately, 10 categories were chosen for all three walking distances for the Italian CI.

3.7.6.2 Location-based analysis

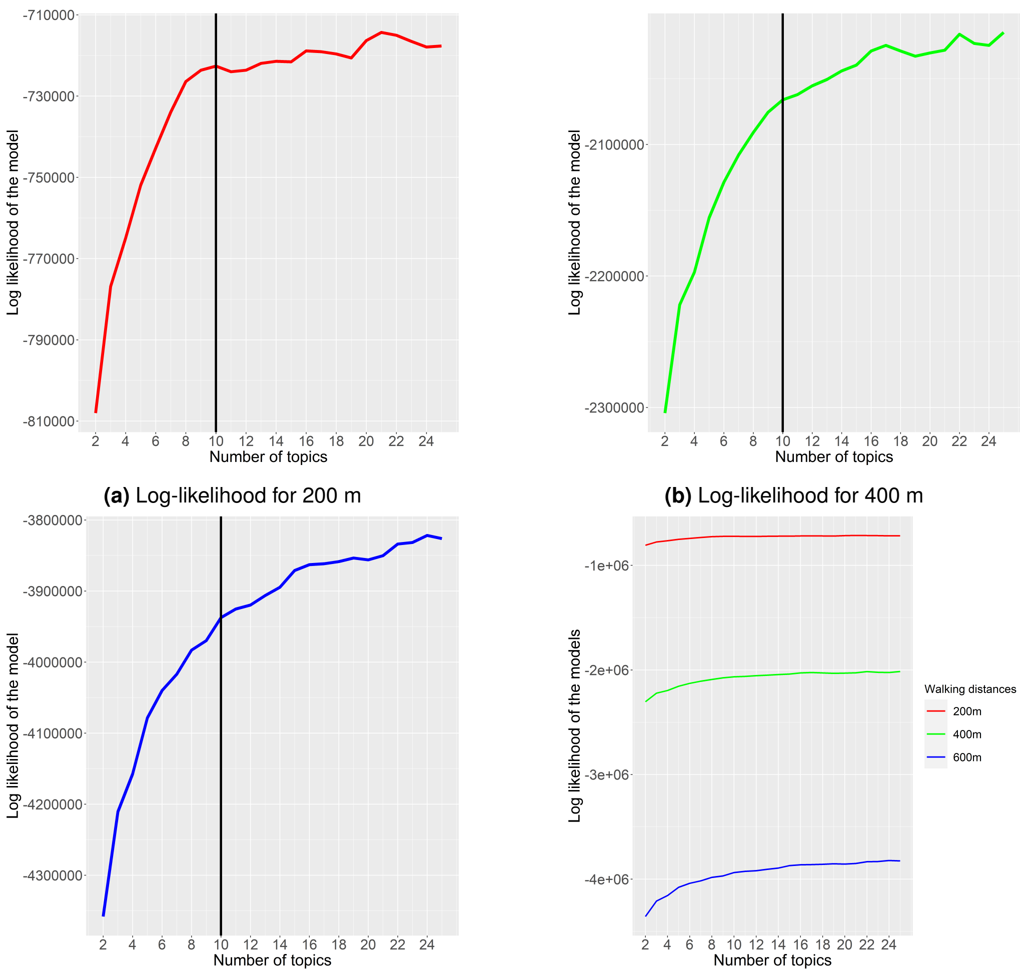
The mean degree of membership for the dominant topics is relatively stable across the distances with 49%, 50% and 48%, respectively.
The comprehensive overview for Italy is provided in table 3.5, which indicates one prominent topic for each walking distance, *Topic7*, *Topic3* and *Topic8*, respectively. As shown in the scatterpie charts (figures 3.13e, 3.14e and 3.15e), these topics have a relatively strong core, yet also a far and broad reach towards their neighbours. With an importance of 83%, 81% and 77%, respectively, `'roads_residential'` largely defines all three topics. A further specification is present with `'group_religious_building'` and `'group_park'`, ranging between an importance of 3% to 5% each. This category is also subject to the most prominent correlations, which are exclusively of negative nature, as visible in figures 3.13b, 3.14b and 3.15b. This indicates a very rather distinct category.
Another set of topics (*Topic5* *Topic9* and *Topic2*, respectively) follows a distinct pattern across all distances with a very dense core and tight gravitational pull. Table 3.5 shows that with above 93% each, all three topics are uniquely defined by `'group_residential_building'` indicating the same category of road-side charging in dense conglomerates of private households.
Additional categories exist with a very similar, medium-sized setup. *Topic1*, *Topic10* and *Topic3*, respectively represent the first and are located in traffic-calmed areas near religious buildings. Another, represented by *Topic6*, *Topic7* and *Topic5*, respectively, is characterised by industrial features and for 200 m and 400 m includes trunk roads and motorways, respectively. These two, however, are merged into their own *Topic6* for the 600 m distance. Natural surroundings manifest another category with medium-sized structure as represented by *Topic6* for 400 m and *Topic7* for 600 m.
The remaining topics are all rather small and largely lodged in the centre of the scatterpie charts in-between their neighbours. Three of them are characterised by the type of road they are at and, judging by the low presence of parking features, are likely to facilitate road-side charging. *Topic2*, *Topic2* and *Topic10*, respectively for national

roads, *Topic3*, *Topic8* and *Topic1*, respectively for regional roads and *Topic8*, *Topic4* and *Topic9*, respectively for local roads.

3.7.6.3 Characteristics-based analysis

The development over time of Italian CS categories can be seen in figures 3.13c, 3.14c and 3.15c. A general increase of CSs across all categories in 2019 is the first observation that catches the eye. Aside from that, other strong increases across the categories are visible in 2015 and 2017. Prior to 2015 only minor CI activity can be spotted. The overall development course of categories appears to be quite varied until 2019 where an expansion in the violin charts (figures 3.13a, 3.14a and 3.15a, respectively) is visible across most categories. CSs on residential roads appear to be among the biggest beneficiaries from this 2019 boost.

CSs on parking areas along motorways were not prominent enough to be categorised by themselves and were rather merged with industrial buildings or trunk roads. Even after this merger, however, the categories with motorways were found to not be overly affected by the 2019 boost.



(a) Log-likelihood for 200 m

(b) Log-likelihood for 400 m

(c) Log-likelihood for 600 m

(d) Log-likelihood for all distances

Figure 3.12: Log-likelihoods for Italy

Table 3.5: Overview of Italian location categories

		Locality Feature 1		Locality Feature 2		Locality Feature 3		Locality Feature 4		Locality Feature 5		**Sum of CS/topic**	
200 m walking distance	Topic1	roads_pedestrian	36%	group_religious_building	10%	pois_bank	5%	transport_bus_stop	5%	group_hotel	4%	**454**	**7%**
	Topic2	roads_primary	75%	transport_bus_stop	10%	traffic_fuel	3%	group_supermarket	2%	landuse_residential	2%	**320**	**5%**
	Topic3	roads_secondary	73%	transport_bus_stop	8%	roads_residential	5%	group_park	2%	landuse_residential	2%	**420**	**6%**
	Topic4	landuse_residential	24%	roads_track	22%	landuse_orchard	9%	group_garage	7%	landuse_vineyard	6%	**427**	**6%**
	Topic5	group_residential_building	95%	building_garages	1%	building_commercial	1%	roads_residential	1%	group_school	1%	**826**	**12%**
	Topic6	building_industrial	67%	landuse_industrial	15%	roads_trunk	11%	traffic_fuel	1%	landuse_retail	1%	**446**	**7%**
	Topic7	roads_residential	83%	group_park	5%	group_religious_building	3%	landuse_residential	3%	group_school	3%	**2292**	**34%**
	Topic8	roads_tertiary	63	transport_bus_stop	9%	group_park	6%	roads_residential	5%	group_parking	4%	**570**	**9%**
	Topic9	group_parking	61%	building_commercial	7%	group_park	6%	transport_bus_stop	4%	group_supermarket	3%	**491**	**7%**
	Topic10	group_hotel	27%	landuse_forest	18%	roads_motorway	11%	landuse_meadow	10%	landuse_scrub	7%	**419**	**6%**
200 m walking distance	Topic1	building_commercial	9%	transport_bus_stop	9%	pois_bank	8%	pois_vending_parking	7%	pois_hairdresser	7%	**88**	**1%**
	Topic2	roads_primary	63%	transport_bus_stop	10%	group_parking	7%	roads_residential	4%	traffic_fuel	2%	**296**	**4%**
	Topic3	roads_residential	81%	group_park	4%	group_religious_building	4%	group_parking	3%	group_school	3%	**2457**	**37%**
	Topic4	roads_tertiary	67%	transport_bus_stop	12%	roads_residential	8%	group_park	4%	group_parking	2%	**291**	**4%**
	Topic5	group_parking	40%	landuse_residential	16%	group_park	9%	group_garage	6%	roads_residential	5%	**369**	**6%**
	Topic6	roads_track	18%	landuse_forest	15%	group_parking	11%	group_hotel	11%	landuse_meadow	7%	**881**	**13%**
	Topic7	building_industrial	67%	roads_motorway	11%	landuse_industrial	7%	group_farm	3%	group_parking	2%	**538**	**8%**
	Topic8	roads_secondary	60%	group_parking	8%	roads_trunk	8%	transport_bus_stop	7%	group_park	3%	**412**	**6%**
	Topic9	group_residential_building	94%	roads_residential	3%	building_garages	1%	group_school	0%	group_park	0%	**1027**	**15%**
	Topic10	roads_pedestrian	46%	group_religious_building	13%	group_hotel	5%	roads_residential	5%	transport_bus_stop	4%	**322**	**5%**
200 m walking distance	Topic1	roads_secondary	55%	transport_bus_stop	10%	group_parking	7%	roads_residential	6%	group_park	5%	**286**	**4%**
	Topic2	group_residential_building	93%	roads_residential	2%	building_garages	1%	group_school	0%	building_commercial	0%	**1135**	**17%**
	Topic3	roads_pedestrian	46%	group_religious_building	14%	transport_bus_stop	5%	pois_bank	3%	group_university	3%	**269**	**4%**
	Topic4	roads_residential	22%	group_hotel	17%	pois_bank	6%	transport_bus_stop	5%	pois_hairdresser	5%	**177**	**3%**
	Topic5	building_industrial	79%	landuse_industrial	6%	group_farm	5%	landuse_residential	2%	group_parking	1%	**408**	**6%**
	Topic6	group_parking	43%	roads_motorway	10%	roads_trunk	10%	group_garage	9%	building_commercial	3%	**417**	**6%**
	Topic7	roads_track	19%	landuse_residential	15%	landuse_forest	15%	group_parking	10%	landuse_meadow	7%	**957**	**14%**
	Topic8	roads_residential	77%	group_parking	6%	group_park	4%	group_religious_building	4%	group_school	3%	**2436**	**36%**
	Topic9	roads_tertiary	50%	roads_residential	19%	transport_bus_stop	10%	group_park	6%	group_school	3%	**426**	**6%**
	Topic10	roads_primary	66%	transport_bus_stop	15%	roads_residential	4%	group_supermarket	2%	traffic_fuel	2%	**171**	**3%**

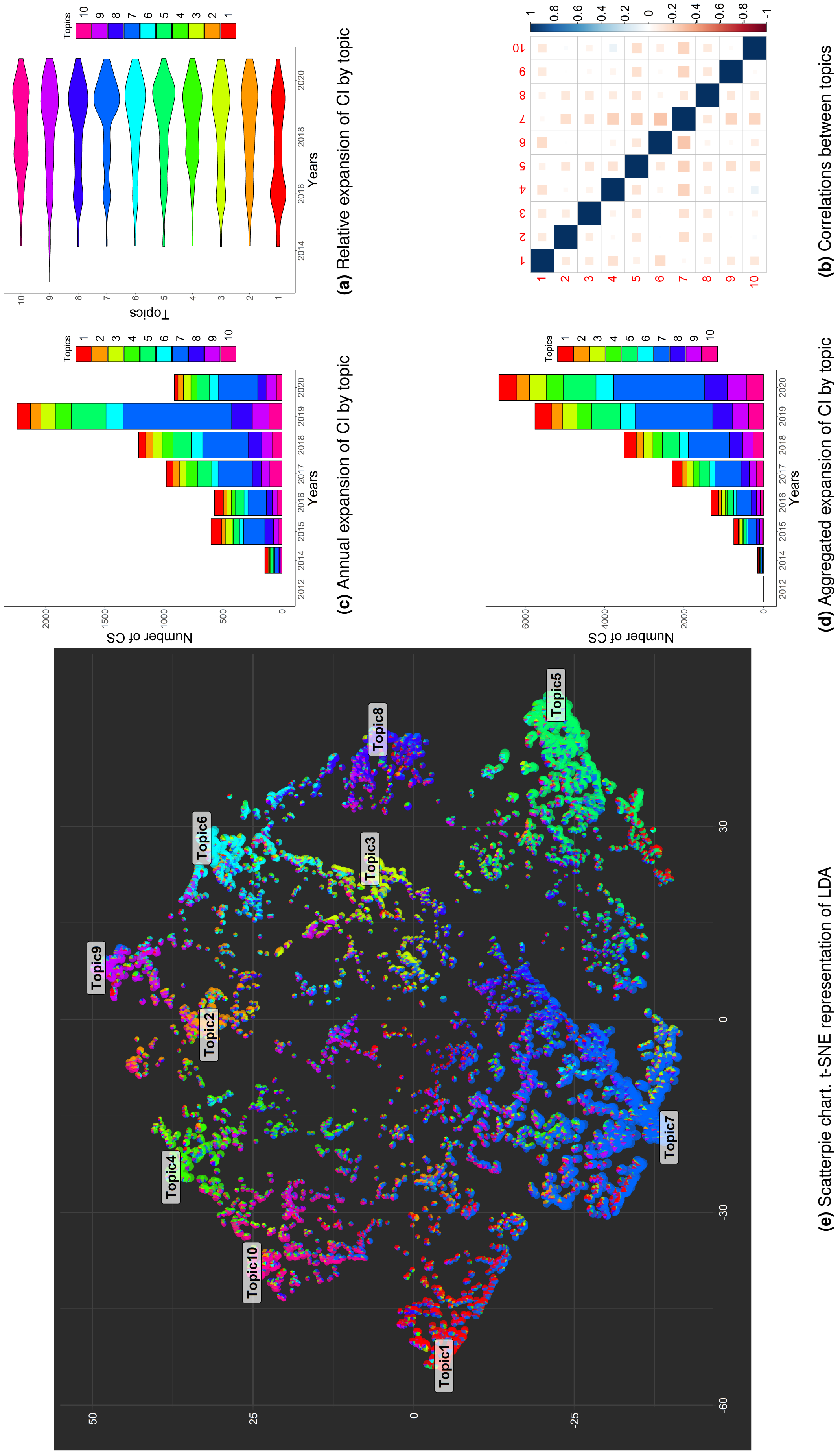
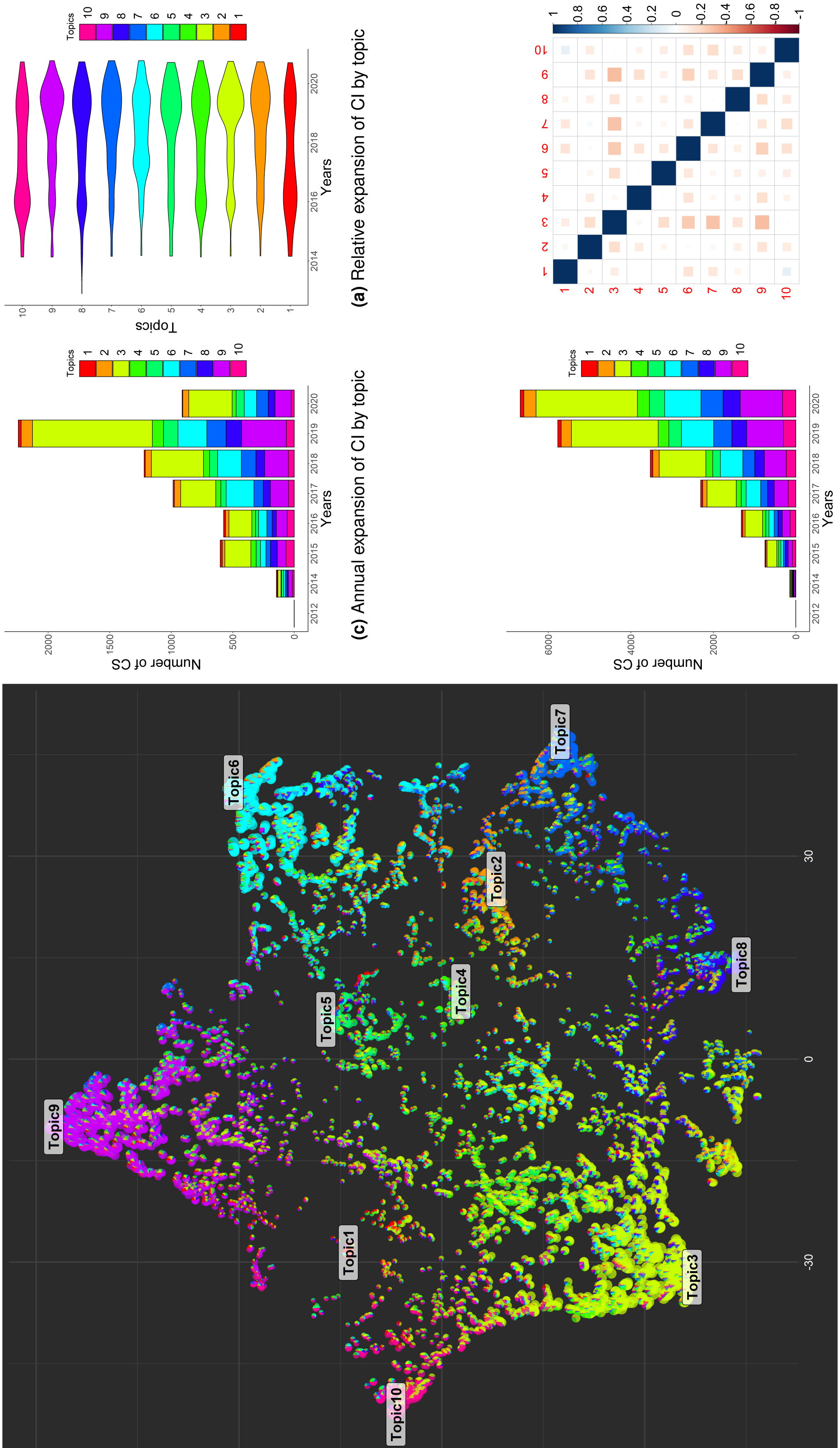


Figure 3.13: Maximum convenient walking distance (IT 200 m)



(a) Relative expansion of CI by topic

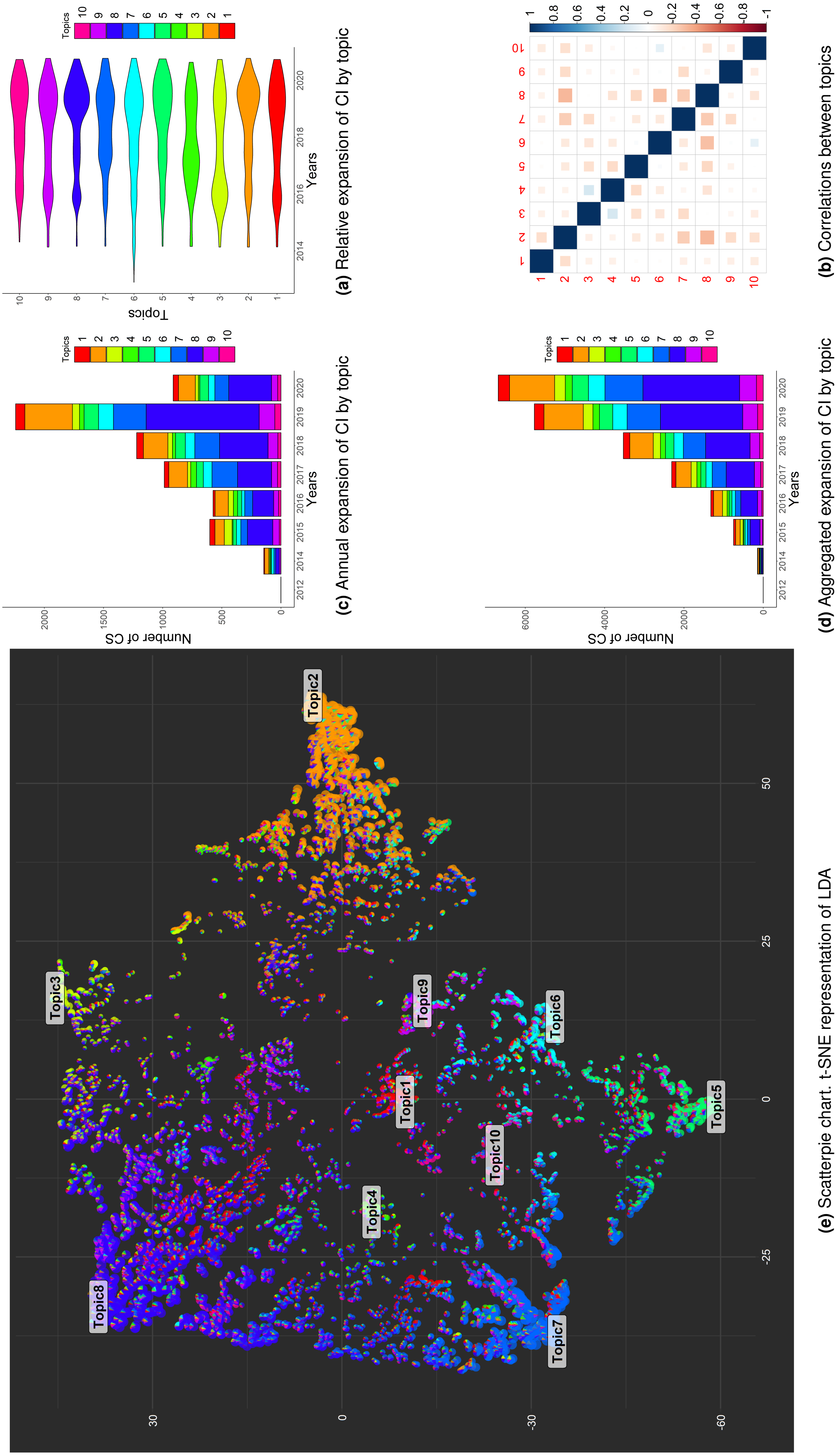
(b) Correlations between topics

(c) Annual expansion of CI by topic

(d) Aggregated expansion of CI by topic

(e) Scatterpie chart. t-SNE representation of LDA

Figure 3.14: Maximum desirable walking distance (IT 400 m)



(a) Relative expansion of CI by topic

(b) Correlations between topics

(c) Annual expansion of CI by topic

(d) Aggregated expansion of CI by topic

(e) Scatterpie chart. t-SNE representation of LDA

Figure 3.15: Maximum acceptable walking distance (IT 600 m)

3.8 Cross-Country Analysis

In this section, the three countries are merged and processed following the procedure of the *within-country* analyses and subsequently referred to as the EU. This is done to get familiar with the data as a whole and provide reference data against which the previous analyses can be compared in the search for patterns and differences. It is of utmost importance, however, to note that the inconsistency of the French OCM data is present within this merger as well. For this reason, the only fully reliable data is the location-based analysis. The characteristics-based analysis must be considered unreliable at least in parts. While after the fixed database import the number of French CSs appears to be representative again, the number of supply points and simultaneous connections, nonetheless, remain unreliable much like the evaluation against time.
After the preparation and cleaning stages, the EU reference has 27,672 CSs offering 53,692 connections for simultaneous charging across a total of 48,750 supply points.

3.8.1 Number of categories

Figure 3.16a contains one local maximum that is of significant interest: At 13 categories, the curve follows a transition from a steep incline to an overall stagnating course. The 400 m course in figure 3.16b is quite similar, but a lot smoother. This curve has its bend at 14 categories, after which higher numbers appear to not provide a significant log-likelihood increase. Figure 3.16c for the 600 m distance has the bend higher up for 17 categories, which is also a small local maximum. Figure 3.16d shows the log-likelihoods of all distances in relation to each other and identifies 200 m as best suited for the given data. Ultimately, 13, 14 and 17 categories were decided to represent the European CI for the respective walking distances.

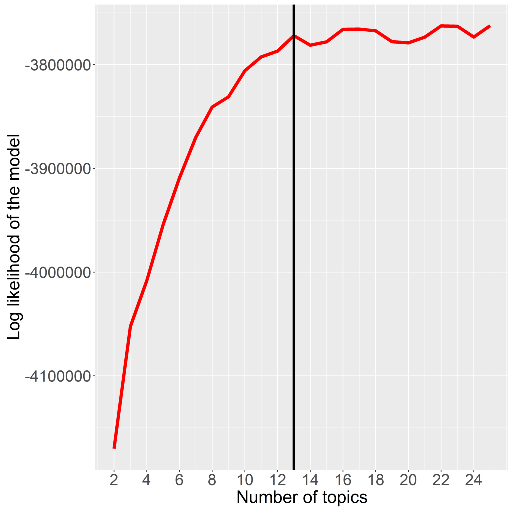
3.8.2 Location-based analysis

The mean degree of membership for the dominant topics is relatively stable across the distances with 46%, 47% and 43%, respectively.
Table 3.6 holds the comprehensive overview for the resulting EU topics for each walking distance, which are visualised in the scatterpie charts in figures 3.17e, 3.18e and 3.19e, respectively. The latter show the presence of two prominent topics for all walking distances. One set with a fanned-out structure whose CSs are dispersed broadly across the 2D space and the other with a similarly far reach, but a more concise pattern and a tighter pull. The former set, *Topic10*, *Topic2* and *Topic1*, respectively, represents a category of CSs along residential roads, occasionally on public parking near religious buildings or schools and has by far the largest amount of members with 32% to 35%.
The second striking category (*Topic5*, *Topic13* and *Topic8*, respectively) is overwhelmingly characterised by `'group_residential_building'` with 94% or above and is, therefore, most likely located in residential conglomerates of larger urban cities as indicated by the low presence (1%) of public parking and garages. A third category that catches the eye, resides along tracks in a natural surrounding and is represented by *Topic7*, *Topic10* and *Topic17*, respectively.
Topic13, *Topic12* and *Topic4*, respectively, follow a similar pattern, which is always in

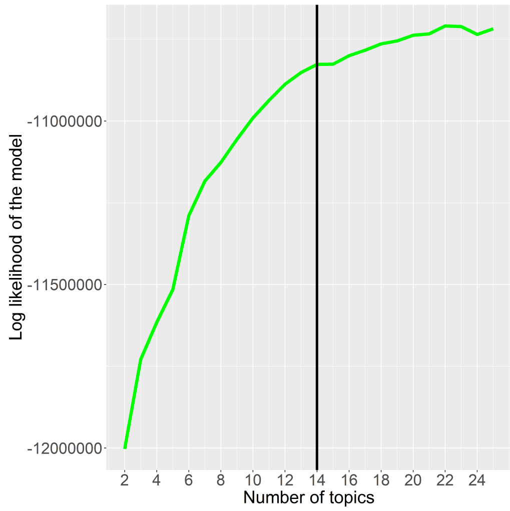
close proximity to the category of CSs in residential conglomerates. This indicates that the surrounding environment is similar in both categories, however, that the defining features, residential buildings and private garages, are spatially separate. This, in turn, implies urban areas where the residential buildings themselves do not offer enough space for garages, which subsequently are clustered in a separate location.
This distinctness in pattern is shared by CSs on parking or service areas along motorways, as represented by *Topic12*, *Topic4* and *Topic5*, respectively, as well as CSs in industrial areas, as represented by *Topic11*, *Topic1* and *Topic6*, respectively.
Three additional categories are again defined by the type of road at which they are located. These are *Topic2*, *Topic9* and *Topic2*, respectively, for national roads, *Topic4*, *Topic5* and *Topic10*, respectively, for regional roads as well as *Topic3*, *Topic8* and *Topic15*, respectively, for local roads. Aside from these, several smaller categories exist, CSs in traffic-calmed areas for example with *Topic1* and *Topic6* for 200 m, *Topic11* for 400 m as well as *Topic12* and *Topic16* for 600 m, respectively.
The correlations between categories for each walking distance are provided in figures 3.17b, 3.18b and 3.19b. The two dominant categories facilitating charging in residential areas hold the most prominent correlations, which are mostly slightly negative in nature, especially so, for the correlation between the two, indicating largely distinct categories.

3.8.3 Characteristics-based analysis

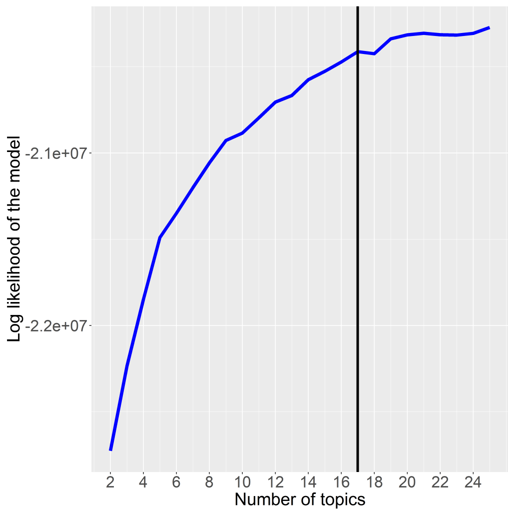
This section is highly affected by the faulty processing of large parts of the French CI as visible in figures 3.17c, 3.17c and 3.17c and will, therefore, only be touched briefly. There are three topics that share a very similar development course: *Topic11*, *Topic1* and *Topic6*, respectively. These represent CI in industrial areas and have a strong initial expansion in 2015, however, no significant development in the following years. Across all categories a general uptake in 2015 is again noticeable, which unfortunately is overshadowed by the belated bulk-import of French data in 2020. Therefore, no further observations can be made based on that data.



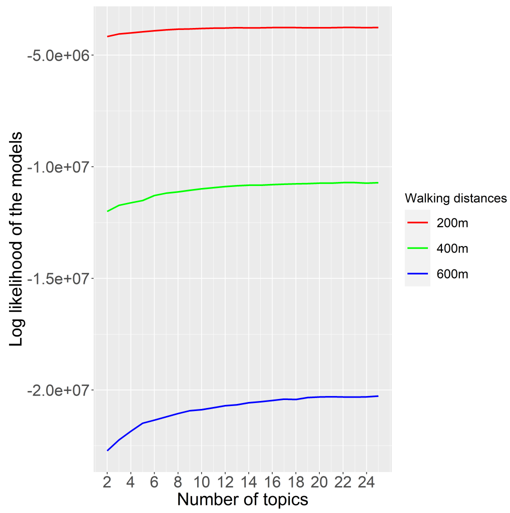
(a) Log-likelihood for 200 m



(b) Log-likelihood for 400 m



(c) Log-likelihood for 600 m

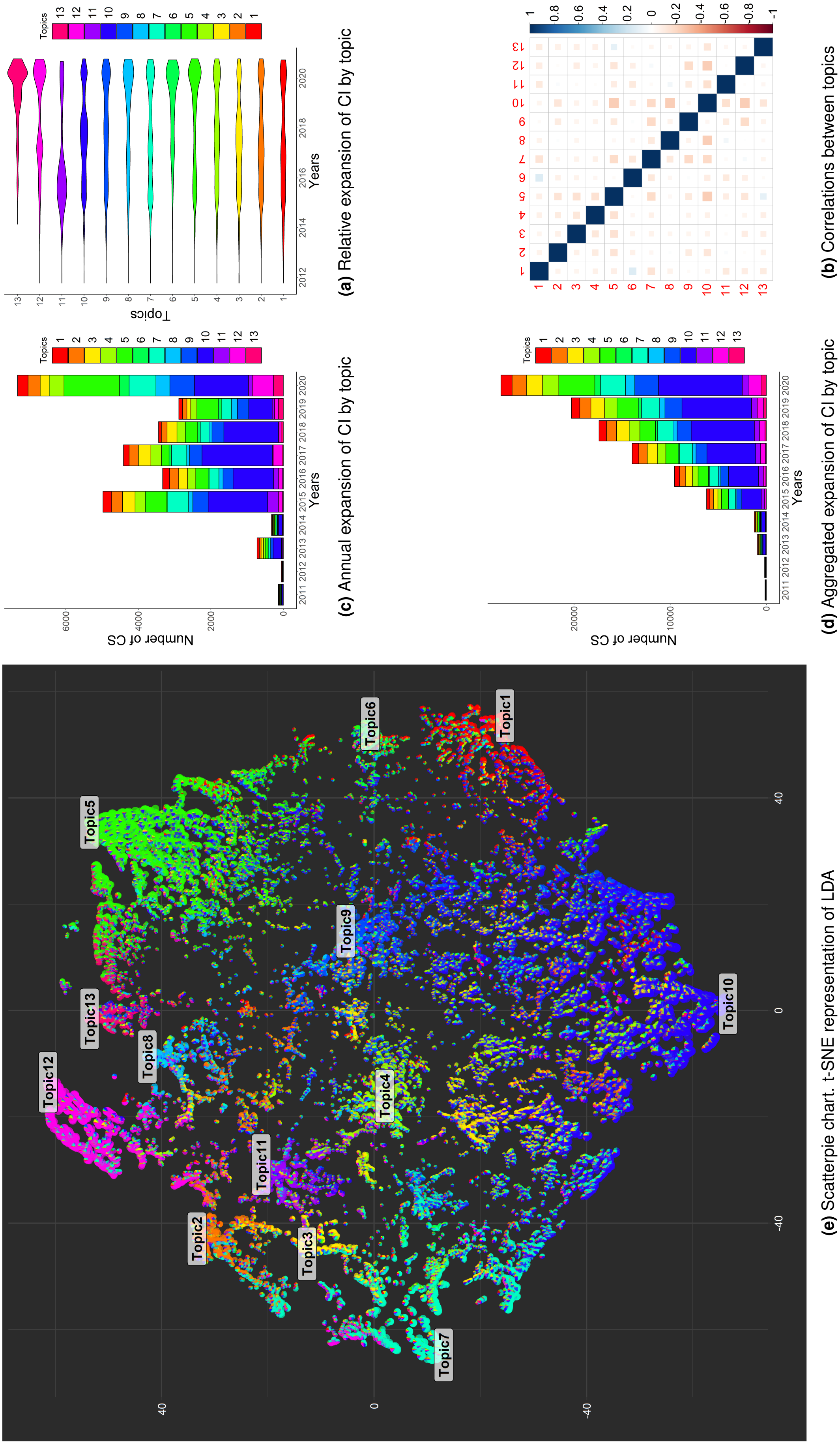


(d) Log-likelihood for all distances

Figure 3.16: Log-likelihoods for Europe

Table 3.6: Overview of European location categories

		Locality Feature 1		Locality Feature 2		Locality Feature 3		Locality Feature 4		Locality Feature 5		**Sum of CS/topic**	
200 m walking distance	Topic1	roads_pedestrian	46%	group_hotel	8%	roads_residential	6%	pois_bank	5%	pois_hairdresser	5%	**1163**	**4%**
	Topic2	roads_primary	69%	roads_trunk	8%	group_parking	6%	transport_bus_stop	5%	landuse_residential	2%	**1491**	**5%**
	Topic3	roads_tertiary	69%	roads_residential	10%	transport_bus_stop	8%	landuse_residential	4%	group_parking	3%	**1678**	**6%**
	Topic4	roads_secondary	78%	transport_bus_stop	8%	group_parking	4%	landuse_residential	2%	group_supermarket	1%	**1689**	**6%**
	Topic5	group_residential_building	95%	roads_residential	2%	landuse_residential	1%	building_garages	1%	group_kindergarten	0%	**3751**	**14%**
	Topic6	roads_living_street	23%	pois_memorial	21%	pois_tourist_info	15%	pois_vending_parking	8%	group_university	6%	**584**	**2%**
	Topic7	landuse_forest	20%	roads_track	20%	landuse_meadow	15%	group_parking	9%	landuse_residential	7%	**2603**	**9%**
	Topic8	building_commercial	20%	building_retail	13%	building_office	9%	landuse_commercial	9%	landuse_residential	8%	**970**	**4%**
	Topic9	group_parking	40%	group_park	18%	group_school	9%	transport_bus_stop	6%	roads_residential	6%	**2452**	**9%**
	Topic10	roads_residential	79%	group_religious_building	6%	landuse_residential	4%	transport_bus_stop	2%	group_parking	2%	**8717**	**32%**
	Topic11	building_industrial	51%	landuse_industrial	25%	group_parking	6%	building_warehouse	4%	group_farm	3%	**703**	**3%**
	Topic12	group_parking	35%	landuse_scrub	25%	roads_motorway	18%	landuse_forest	5%	traffic_fuel	4%	**1256**	**5%**
	Topic13	group_garage	68%	building_garages	14%	roads_residential	8%	landuse_residential	2%	transport_bus_stop	2%	**547**	**2%**
200 m walking distance	Topic1	building_industrial	56%	roads_trunk	20%	landuse_industrial	8%	group_farm	3%	group_parking	2%	**738**	**3%**
	Topic2	roads_residential	74%	group_parking	6%	group_religious_building	4%	transport_bus_stop	3%	group_school	3%	**9779**	**35%**
	Topic3	pois_memorial	70%	landuse_scrub	4%	building_cabin	3%	pois_hairdresser	3%	roads_residential	3%	**96**	**0%**
	Topic4	group_parking	54%	roads_motorway	14%	landuse_scrub	11%	landuse_forest	3%	traffic_fuel	2%	**1721**	**6%**
	Topic5	roads_secondary	70%	group_parking	9%	transport_bus_stop	7%	landuse_residential	2%	roads_residential	2%	**1223**	**4%**
	Topic6	building_commercial	25%	pois_tourist_info	15%	building_office	12%	building_retail	12%	landuse_commercial	5%	**348**	**1%**
	Topic7	group_park	16%	roads_residential	12%	transport_bus_stop	8%	pois_vending_parking	8%	pois_hairdresser	6%	**671**	**2%**
	Topic8	roads_tertiary	53%	roads_residential	16%	group_parking	10%	transport_bus_stop	6%	group_school	3%	**1702**	**6%**
	Topic9	roads_primary	68%	roads_residential	10%	transport_bus_stop	6%	group_parking	4%	landuse_residential	2%	**1001**	**4%**
	Topic10	roads_track	25%	landuse_forest	18%	landuse_meadow	15%	landuse_residential	5%	pois_tourist_info	4%	**2772**	**10%**
	Topic11	roads_pedestrian	36%	roads_living_street	11%	roads_residential	6%	group_hotel	6%	group_religious_building	5%	**1444**	**5%**
	Topic12	group_garage	81%	roads_residential	5%	group_parking	3%	roads_living_street	2%	traffic_turning_circle	1%	**661**	**2%**
	Topic13	group_residential_building	94%	roads_residential	3%	group_parking	1%	group_school	0%	roads_living_street	0%	**4969**	**18%**
	Topic14	landuse_residential	24%	building_garages	21%	group_parking	8%	roads_residential	6%	group_school	4%	**501**	**2%**
200 m walking distance	Topic1	roads_residential	80%	group_religious_building	4%	group_school	3%	group_park	2%	group_parking	2%	**9698**	**35%**
	Topic2	roads_primary	69%	group_parking	9%	transport_bus_stop	6%	roads_residential	4%	group_supermarket	1%	**747**	**3%**
	Topic3	group_parking	45%	roads_residential	16%	transport_bus_stop	10%	group_park	7%	traffic_turning_circle	4%	**1393**	**5%**
	Topic4	group_garage	89%	roads_residential	4%	traffic_turning_circle	1%	transport_bus_stop	1%	group_parking	1%	**791**	**3%**
	Topic5	roads_motorway	31%	landuse_scrub	30%	group_parking	15%	landuse_forest	6%	roads_track	3%	**1063**	**4%**
	Topic6	building_industrial	73%	landuse_industrial	8%	group_farm	5%	building_warehouse	3%	group_parking	2%	**548**	**2%**
	Topic7	pois_memorial	27%	building_office	9%	pois_vending_parking	9%	group_residential_building	8%	group_university	6%	**412**	**1%**
	Topic8	group_residential_building	94%	roads_residential	3%	group_school	0%	building_commercial	0%	group_parking	0%	**5989**	**22%**
	Topic9	roads_trunk	73%	group_parking	3%	transport_bus_stop	2%	landuse_industrial	2%	building_retail	2%	**175**	**1%**
	Topic10	roads_secondary	74%	group_parking	7%	transport_bus_stop	5%	roads_residential	4%	group_supermarket	1%	**779**	**3%**
	Topic11	pois_tourist_info	28%	building_commercial	23%	group_hotel	17%	roads_residential	4%	group_parking	4%	**238**	**1%**
	Topic12	roads_living_street	46%	group_parking	15%	roads_residential	9%	group_school	3%	group_park	3%	**232**	**1%**
	Topic13	landuse_residential	38%	building_retail	7%	roads_residential	7%	landuse_commercial	6%	group_park	5%	**257**	**1%**
	Topic14	building_garages	62%	group_parking	6%	group_residential_building	5%	landuse_allotments	4%	roads_residential	3%	**108**	**0%**
	Topic15	roads_tertiary	68%	transport_bus_stop	9%	group_parking	8%	roads_residential	6%	group_park	2%	**660**	**2%**
	Topic16	roads_pedestrian	36%	roads_residential	7%	group_religious_building	6%	transport_bus_stop	6%	pois_bank	5%	**1260**	**5%**
	Topic17	roads_track	30%	landuse_forest	18%	landuse_meadow	15%	landuse_residential	4%	roads_residential	4%	**3280**	**12%**



(a) Relative expansion of CI by topic

(b) Correlations between topics

(c) Annual expansion of CI by topic

(d) Aggregated expansion of CI by topic

(e) Scatterpie chart. t-SNE representation of LDA

Figure 3.17: Maximum convenient walking distance (EU 200 m)

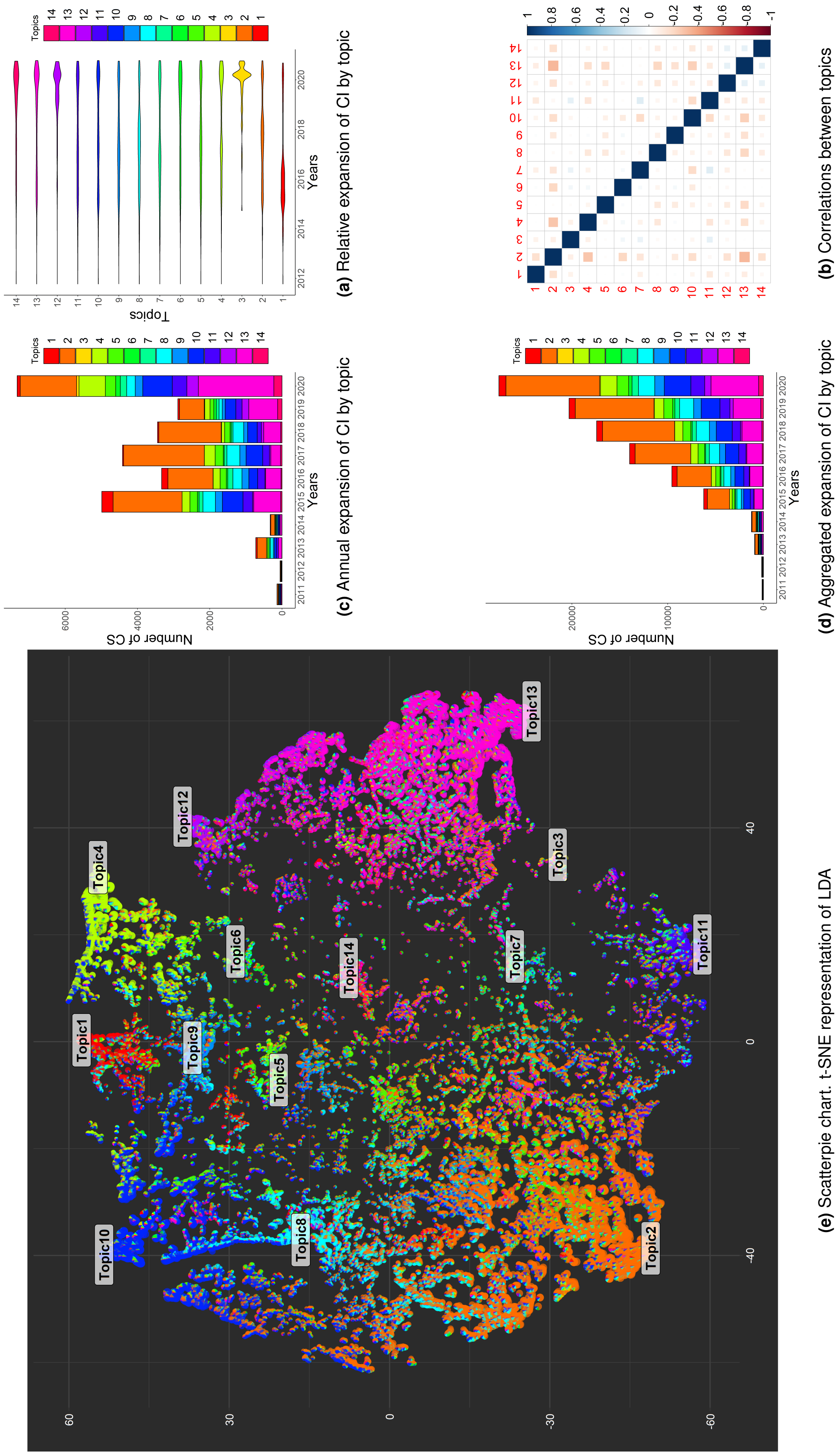


Figure 3.18: Maximum desirable walking distance (EU 400 m)

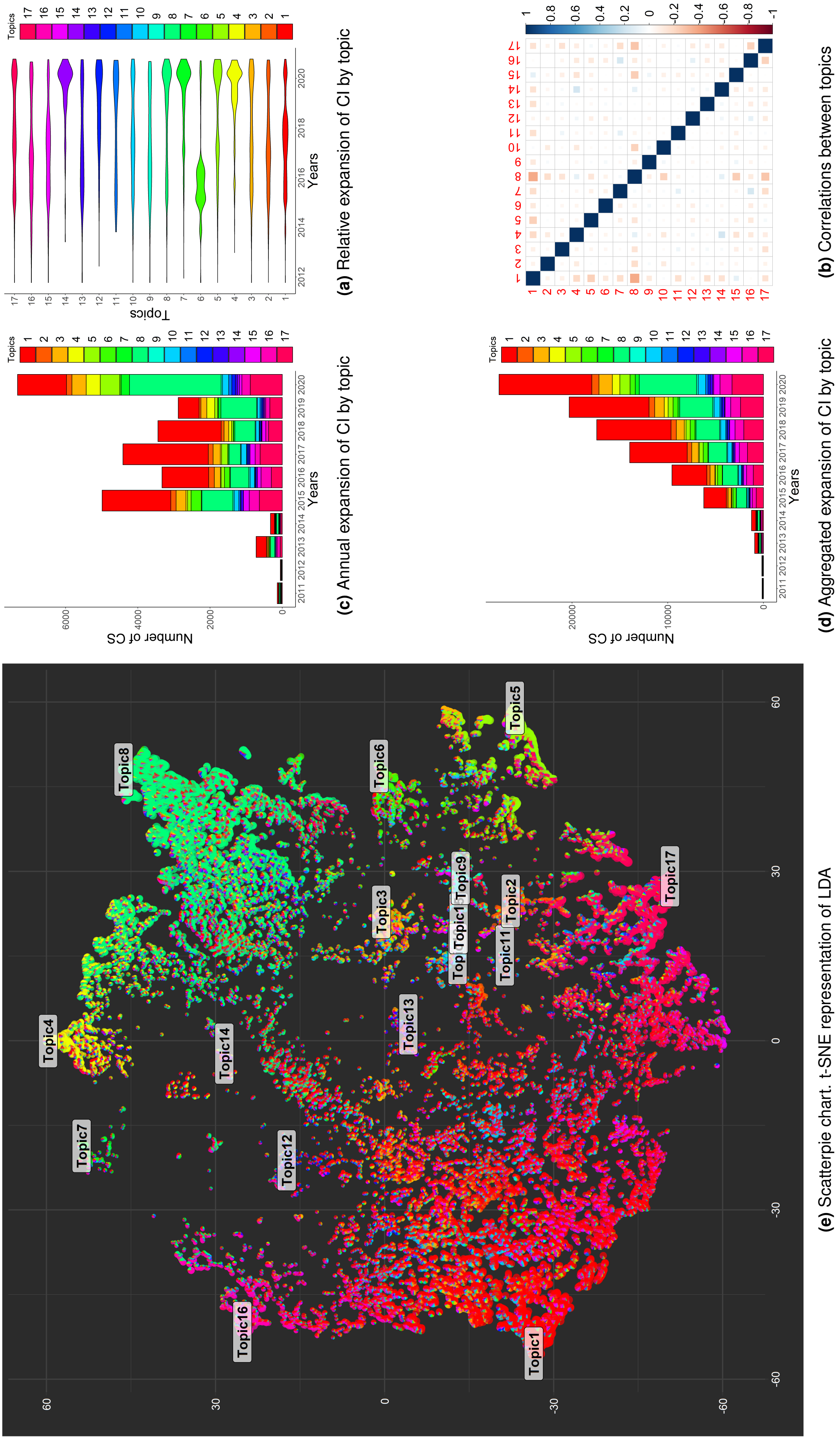


Figure 3.19: Maximum acceptable walking distance (EU 600 m)

4 Results and Discussion

The results of the descriptive analyses in the previous chapter are used to infer hypotheses regarding the overall development of CI within Europe. The hypotheses are shaped according to findings from the derived *cross-country* patterns and the relationships of the findings are investigated for their likely causes. Subsequently, the hypotheses and findings are enfolded by literature to help position this thesis in the assessed research.

4.1 Shaping Hypotheses

The most prominent result from the location-based analyses is the finding that CI along residential roads overshadows all other location types. With the exception of France, this is closely followed by road-side CSs next to conglomerates of private households, facilitating home charging. For all three individual countries, the combined percentage of these two ranges between 40% and 60%. This similarity based on the *within-country* analyses is supported by the merged CI data for the *cross-country* analysis. Although the national subsidy programmes differ strongly in terms of duration and financial size, the share of these two categories is stable as of 11 October 2020. This finding indicates the first hypothesis: Not only do CSs in residential areas dominate the public CI, they also evolve independent of subsidy programmes relative to the overall CI.

Another category was derived which proved to be of significant interest: CSs at parking or service areas along motorways. Italy was found to be the only country where these are not dominant enough to facilitate an own category and were rather merged with parking areas along national roads. The German subsidy programme specifically targets fast chargers with around one quarter of all approvals, which results in parking and service areas along motorways to be a strong beneficiary as indicated by the atypical development in the violin charts. For the French and Italian subsidy programme on the other hand, no detailed specifications of location requirements for CSs were found. The French subsidy programme has been running since 2016 and this category makes up between 4% and including 10% of the total CI. Italy, on the other hand, only initiated their subsidy programme at the end of 2019 and CI along motorways were found to not be dominant enough for an own category. This indicates the second hypothesis: CSs at parking or service areas along motorways do not significantly evolve on their own and highly depend on focused state subsidies.

Three smaller, but consistently present categories are primarily related to the type of road they are located at. These represent CSs at public parking areas near local, regional and national roads, with individual CI shares of consistently below 10%. This is in line with the share of CSs along traffic-calmed areas, which, however, are not consistently in the vicinity of public parking areas. This implies road-side charging as individual parking spaces along roads are not listed as parking features in the Geofabrik export (Topf and Ramm (2019) and OSM (2019a)) and large-scale parking facilities are unlikely to be located on roads not primarily intended for vehicles. This leads to the third hypothesis: Public parking areas near roads and road-side charging are consistently considered for CS placement unrelated to the respective road type. The only exception to this are motorways.

Two prominent results emerge from the characteristics-based analyses. First, for both the German and Italian CI, a general uptake in 2015 was noticeable across most categories. This may be explained by directives set by the EU. In its 2014 legislation regarding CI development, the EU urged that CSs “are built up with adequate coverage, in order to enable electric vehicles to circulate at least in urban/suburban agglomerations and other densely populated areas” (European Union, 2014). Second, the start of the respective subsidy programmes is clearly visible in the annual expansion and violin charts with noticeable changes in 2017 for Germany and 2019 for Italy. Therefore, both state-subsidies and mandated legislation appear to have a significantly positive effect on the expansion rate of public CI. This is based entirely on Germany and Italy as the characteristics-based analysis of France was overshadowed by faulty OCM database imports. While for the German CI development the effect of the EU legislation slightly surpasses the effect of the state subsidy programme, it is the other way around for Italy, where the effect of the state subsidy programme was found to be higher. Due to the corruption of the control case, France, no further specifications can be drawn in favour of either, leading to the more general fourth hypothesis: Legislation and subsidy programmes both have a significantly positive effect on CI development.

If correlations between categories are present in the correlation plots, they are largely of slightly negative nature, indicating mostly distinct categories. This is generally supported by the overall style of the scatterpie charts and the fact that with only few exceptions, adequate topics for all categories could be found across the walking distances. Additionally, the categorisation was consistently stable across the walking distances as indicated by the mean degree of membership towards the dominant topic. This mean degree of membership, however, was found to be consistently at or slightly below 50% indicating that on average the dominant category of each CS is not able to represent the absolute majority of all present fclasses. All four figures of joint log-likelihood graphs indicate that the *maximum convenient walking distance* (200 m) had the best goodness of fit for the given data. However, log-likelihood values were found to be very low overall. This combination indicates the fifth hypothesis: While the categorisation proves stable across all cases and walking distances, the derived categories are on average not capable of representing the majority of fclasses.

4.2 Enfolding Literature

Residential localities have been found to have the largest share of public CI. This is in line with the consensus in the literature that “home charging was essential for all [driver] segments” (Wenig et al., 2019) because “the vast majority of the charges (approx. 88%)

are carried out at the place of residence" (Baresch and Moser, 2019) as pointed out in 2.2.5. Additionally, "the charging points should also, regardless of the municipality type, be installed where they are visible to the public in order to increase the effectiveness of the policy instrument" (Egnér and Trosvik, 2018) and the reassurance of potential EV drivers. This guideline was implemented with great focus by the German government as CSs located at public places around memorials in residential conglomerates are one of the biggest beneficiaries from their subsidy programme. Another research finding that was considered by all three countries is that "in a less mature market, demand-driven [CSs] seem the most likely candidate as they show significant higher performance on energy transfer" (Helmus et al., 2018) compared to their strategic counterparts. The fact that CSs in residential areas evolve independent of subsidy programmes indicate that this category was developed with great focus from the start.

The German programme further specifically focused fast chargers with around a quarter of its subsidies (BMVI, 2020b) resulting in a more robust network along motorways than Italy. "Without fast chargers, the transition from liquid-fuel vehicles to BEVs will be affected" (Neaimeh et al., 2017) as they "could provide assurance and comfort to reduce range anxiety and the perceived unsuitability of BEVs beyond short city driving." (Neaimeh et al., 2017) The truth of this conclusion is visible in the numbers of BEVs and Plugin Hybrid Electric Vehicles (PHEVs) in those two countries. The EAFO provides total number of alternative fuel passenger cars grouped by their drive train. For Germany this number is nearly 300,000 (EAFO, 2020c) for BEVs and PHEVs combined, while it is around 55,000 (EAFO, 2020d) for BEVs and PHEVs combined in Italy. The conclusions that "fast charge networks might not be profitable" (Neaimeh et al., 2017) and the proposal for "continuing financial incentives to protect investors from uncertainties in the marketplace" (Serradilla et al., 2017) are reasonable indicators for why a network of fast chargers along motorways is unlikely to evolve by itself as seen in the Italian CI development. However, this category in Germany shows little activity prior to the subsidy programme, therefore, the findings of this thesis suggest that there is a significantly positive effect of focused state subsidies on CI development. This is in line with, for example, Egnér and Trosvik (2018) who found that "the local policy instrument of public charging infrastructure has a significant and positive impact on the BEV adoption rate" (Egnér and Trosvik, 2018). Furthermore, "by developing public charging infrastructure, the barriers of range anxiety and limited charging possibilities decrease which, in turn, increase the utility of owning a BEV." (Egnér and Trosvik, 2018) This development, on the other hand, requires "continuing financial incentives to protect investors from uncertainties in the marketplace." (Serradilla et al., 2017)

Although not financial in nature, mandatory CI development as enforced by EU legislation was found to be another incentive with significantly positive effect. This is supported by the Norwegian approach regarding CI, which is a mixture between financial and regulatory incentives. The financial aspect is covered by the budget of 2.1 million € "to housing associations for installing chargers" (Wallbox, 2020). The other aspect is regulation for the construction of "fast charging stations every 50 km" (Lorentzen et al., 2017) along main roads and that "for parking lots and parking areas of new buildings, a minimum amount of 6% has to be allocated to electric cars" (EAFO, 2020f). Policies like these enabled Norway to achieve an EV share of around 70% in August 2020 (Holland, 2020). As stated earlier, the road transport sector "accounts for 72 % of total greenhouse gas emissions" (EEA, 2019), which is why the achievement by Norway is a milestone towards the goal of European Union (2018) to "reduce the overall environmental impact of production and consumption in the mobility sector" (European Union, 2018). Therefore, the findings of this thesis are in line with the presented literature in the recommendation for legislators to provide state subsidies to support the development of public CI and facilitate growth of the EV share. Aside from state subsidies for the purchase of CI by non-governmental entities, regulation and legislation were found to also present a large impact on the CI expansion.

The limited capability of the derived categories to appropriately represent the majority of fclasses was the last identified hypothesis. This issue only surfaced, because of the used LDA *topic modelling* approach, which allows for mixed-memberships towards categories. This, however, is a hugely beneficial advantage and the suitability of this methodology has been proven through different use-cases mentioned in 2.6. Connected to this is the fact that since its initial publication in 2003, attempts have been made to improve upon LDA. Namely, the correlated topic model approach that is also provided in the *topicmodels* package for R by Gruen (2020). This approach was first published by Blei and Lafferty (2007) and "gives better predictive performance and uncovers interesting descriptive statistics" (Blei and Lafferty, 2007). In the same paper, however, Blei and Lafferty (2007) acknowledge "one issue that we did not thoroughly explore is model selection, that is, choosing the number of topics for a collection" (Blei and Lafferty, 2007). Therefore, "tackling the model selection issue in this setting is an important area of future research." (Blei and Lafferty, 2007) It was not applied in this thesis for the same reason, as due to disproportionately longer computation times, substantial testing was not feasible in the given time frame. In an attempt to mitigate this, the correlations between topics were computed after the LDA processing, but not taken into consideration during the actual categorisation. While LDA is still considered to be a highly suitable approach for the presented framework, overall low log-likelihood values of the categories suggest that a different method of deriving the appropriate number of categories may be beneficial. For example using the coherence score of topics as presented by Tang (2019) and Kapadia (2019).

Development of CI on public parking areas along roads was identified as the third hypothesis, which agrees with Morrissey et al. (2016). They found that "standard chargers may become commercially viable at car park locations in the shortest period of time due to the higher usage frequencies" (Morrissey et al., 2016). The overall dominance of roads in their respective categories, however, may also be attributed to the fact that longer roads are often denominated into smaller line strings, where each line string is its own geometry (OSM, 2019b). This is one of the limitations discussed in the following chapter.

5 Reaching Closure

5.1 Limitations

The biggest constraining factor emerging from the databases used in this thesis, is the unreliability of all time-related data for the French and therefore EU OCM dataset, which presented itself in September. This limitation comes back to using the `‘DateCreated’` variable declaring when each CS was added to OCM as the indicator of when it was constructed. Generally, this would be an acceptable estimator as the discrepancy between actual construction date and the filing by an OCM contributor is considered to be balanced across the cases. Due to a bulk import of French data on 27 August 2020, however, this estimator became unreliable for this and subsequently the EU time factor. A possible further limitation founded within the used data is the previously mentioned fact that longer roads are often denominated into smaller line strings. If, for example, a CS were to be in the vicinity of a single large road which, due to curves, is split up into four smaller line strings in OSM, the framework interprets this similarly to a CS that is in fact located in the vicinity of four separate roads at an intersection. This factor can potentially be mitigated by also importing the `‘ref’` attribute from the Geofabrik road shape file, which, according to Topf and Ramm (2019), provides a reference number for the roads if present. Because an initial inspection of this attribute was deemed unsuitable due to a large presence of `‘NA’` values, an investigation into alternatives is recommended, as it may produce a possibility to reliably merge road sections into one geometry. Additionally, the framework is currently only capable of identifying semantic duplicates within shape files and does not consider semantic duplicates across various shape files. An initial investigation into the German dataset prompted the result that there are only around 117,000 semantic duplicates across shape files, which stands in comparison to a total of over 17 million localities.

5.2 Future Work

The R framework built in this thesis is intended to be applicable in any desired categorisation of given longitude and latitude combinations based on surrounding OSM locality feature classes. The framework is planned to be made publicly available on Github so the following constraints, possibly preventing universal application, can be addressed. There are two constraints regarding the import of the OSM datasets. First, the framework does not account for possible spillover shape files as mentioned in 2.4, because no shape file of the researched cases exceeds 2 GB. Second, the framework requires 32 GB of RAM due to the OSM data import being rather inefficient, for which an investigation of alternatives is suggested. To reduce the amount of manual input from researchers and subsequently reducing the chance of errors, the implementation of an algorithm is recommended to automatically detect which topics from the three walking distances belong to the same category. This could potentially be based on a mixture between number of CSs, similarities of the top-5 fclasses per topic as well as the correlation between topics in each walking distance. It would, furthermore, improve the clarity and intuitiveness of the results.

In the past, a random forest implementation was planned to derive those local conditions which have a dominant influence on whether or not a CS is present at a given location. This was intended to be used as a comparison to the results of the LDA topic modelling approach, to estimate the robustness of the framework. For this to work, a functionality has been implemented that generates sample point geometries for each country that are not in vicinity of actual CSs. The result is a dataset with `'X'` actual CSs and a dataset with `'X'` dummy CSs. It was planned to generate DTMs for both of these datasets and then use a random forest implementation to derive those locality feature classes that are predominantly responsible for the presence of an actual CS. The already fully functional generation of sample point geometries remains in the code in case future research deems it useful. Last but not least, the framework also already facilitates broader characteristics-based analyses, for example, considering the annual expansion of connection types (table A.3) as visible in figure 5.1. Further options include the consideration of using the `'quantities'` of connections as base for the categorisation to be more in line with the data provided by EAFO or to investigate the power levels of the chargers per category to derive knowledge on how standard and fast chargers are dispersed across the categories. Combined with electricity use data of individual CSs, conclusions could be drawn regarding the overall energy consumption of categories, which, in turn, would enable deriving implications towards the search of the optimal CS location.

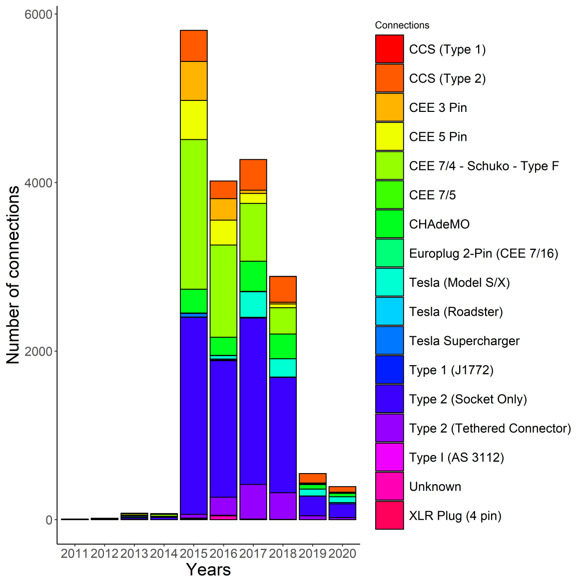


Figure 5.1: Annual expansion of connection types for Germany

5.3 Conclusion

In general, this thesis helps with the expansion of literature concerned with the evaluation of different CS locations. The presented framework is capable of categorising the entire national CI of any country that is supported by the OCM database and OSM extracts by Geofabrik as well as in general any given longitude and latitude dataset. This is done by matching the longitude and latitude information of the CSs as provided by the OCM database with the surrounding locality features as provided by the OSM database extracts from Geofabrik. The resulting categories grant insights into the distribution of the respective national CI and provide possibilities for future research if, for example, coupled with electricity consumption data per CS. The findings of this thesis suggest that state subsidies for the purchase of public CI by non-governmental entities have a substantially positive effect on the overall development, which is in line with the identified research strands. Through their categorisation using an LDA topic modelling approach, local conditions were found to have a strong effect as especially CSs in residential areas consistently dominate the overall CI across all cases. Furthermore, aside from local conditions and state subsidies, regulations by legislators were identified as a third influencing factor.
To "reduce the overall environmental impact of production and consumption in the mobility sector" (European Union, 2018), which currently "accounts for 72 % of total greenhouse gas emissions" (EEA, 2019), a mixture of financial and regulatory incentives in the form of state subsidies and legislation is recommended for decision makers. The research question of this thesis, as initially specified, is: How does the public EV charging infrastructure develop in Europe *(on the example of France, Germany and Italy)* and what influence do state subsidies and local conditions have in that regard?

Based on the findings, it can be concluded that there is a strong effort from EU and national legislators for developing charging infrastructure within Europe, with the goal of reducing overall greenhouse gas emissions from the transport sector. The location-based analysis of the resulting CS categories for Germany, France and Italy using OSM data yielded the conclusion that the influence of local conditions is decisively lead by residential locality feature classes and followed by public parking along local, regional and national roads. The characteristics-based analysis of CSs using OCM data yielded the conclusion that the influence of state subsidies was significant, positive and especially for the development of fast chargers along motorways highly beneficial as their development proved stale without specific incentives. CI in residential areas, on the other hand, was found to evolve even outside of time frames with state subsidies, suggesting that financial resources dedicated for this category may have a more effective use for other categories.

Bibliography

ADVENIR. Point de recharge sur parking privé à destination d'une flotte, a. URL https://advenir.mobi/flottes/. Accessed on 2020-08-21.

ADVENIR. Point de recharge sur la voirie, b. URL https://advenir.mobi/voirie-2/. Accessed on 2020-08-21.

ADVENIR. Equipement d'infrastructures collectives pour les 3000 premières copropriétés, October 2019. URL https://advenir.mobi/3-000-coproprietes/. Accessed on 2020-08-21.

ADVENIR. Le programme CEE ADVENIR s'ouvre à de nouveaux financeurs, January 2020. URL https://advenir.mobi/2018/01/15/le-programme-cee-advenir-souvre-a-de-nouveaux-financeurs/. Accessed on 2020-08-21.

John Erik Anderson, Moritz Bergfeld, Bastian Chlond, Horst E Friedrich, Niklas Hoffmann, Lenz, Florian Kleiner, Matthias Klötzke, Tobias Kuhnimhof, Stephan Schmid, Tamer Soylu, Felix Steck, and Christine Weiss. LADEN 2020: Konzept zum Aufbau einer bedarfsgerechten Ladeinfrastruktur in Deutschland von heute bis 2020. Technical report, Bundesministerium für Wirtschaft und Energie, December 2016. URL https://publikationen.bibliothek.kit.edu/1000093882.

AVERE. Comment bénéficier du crédit d'impôt transition énergétique de 300 euros pour bornes de recharge ?, January 2020. URL http://www.avere-france.org/Site/Article/?article_id=6091. Accessed on 2020-08-21.

Martin Baresch and Simon Moser. Allocation of e-car charging: Assessing the utilization of charging infrastructures by location. *Transportation Research Part A: Policy and Practice*, 124:388–395, June 2019. ISSN 09658564. doi: 10.1016/j.tra.2019.04.009. URL https://linkinghub.elsevier.com/retrieve/pii/S0965856418311042. Accessed on 2020-08-03.

David M. Blei and John D. Lafferty. A correlated topic model of SCIENCE. *The Annals of Applied Statistics*, 1(1):17–35, June 2007. ISSN 1932-6157. doi: 10.1214/07-AOAS114. URL http://www.cs.columbia.edu/~blei/papers/BleiLafferty2007.pdf. Accessed on 2020-09-12.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, page 30, January 2003. URL http://www.cs.columbia.edu/~blei/papers/BleiNgJordan2003.pdf.

BMVI. 500 Millionen Euro zusätzlich für Ladeinfrastruktur - 6. Förderaufruf abgeschlossen, August 2020a. URL https://www.bmvi.de/SharedDocs/DE/Artikel/G/infopapier-sechster-foerderaufruf-ladeinfrastruktur.html. Accessed on 2020-09-18.

BMVI. Förderrichtlinie Ladeinfrastruktur für Elektrofahrzeuge, August 2020b. URL https://www.bmvi.de/DE/Themen/Mobilitaet/Elektromobilitaet/Ladeinfrastruktur/Ladeinfrastruktur.html. Accessed on 2020-08-21.

Bundesnetzagentur. Ladesäulenkarte im Rahmen der Ladesäulenverordnung, September 2020. URL https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/HandelundVertrieb/Ladesaeulenkarte/Ladesaeulenkarte_node.html. Accessed on 2020-10-04.

Joana Cavadas, Gonçalo Correia, and João Gouveia. Electric Vehicles Charging Network Planning. In Jorge Freire de Sousa and Riccardo Rossi, editors, *Computer-Based Modelling and Optimization in Transportation*, volume 262, pages 85–100. Springer International Publishing, Cham, 2014. ISBN 978-3-319-04629-7 978-3-319-04630-3. doi: 10.1007/978-3-319-04630-3_7. URL http://link.springer.com/10.1007/978-3-319-04630-3_7. Accessed on 2020-08-20.

Database Eurostat. Database of Eurostat, the statistical office of the European Union., 2020. URL https://ec.europa.eu/eurostat/data/database. Accessed on 2020-08-17.

Siena Duplan. Visualizing Topic Models with Scatterpies and t-SNE, November 2019. URL https://towardsdatascience.com/visualizing-topic-models-with-scatterpies-and-t-sne-f21f228f7b02?gi=d6ec68701fa6. Accessed on 2020-09-24.

EAFO. European Union - Infrastructure Electricity, September 2020a. URL https://eafo.eu/countries/european-union/23640/infrastructure/electricity. Accessed on 2020-08-18.

EAFO. Germany - Infrastructure Electricity, September 2020b. URL https://eafo.eu/countries/germany/1734/infrastructure/electricity. Accessed on 2020-09-20.

EAFO. Germany - Vehicles and Fleet, September 2020c. URL https://eafo.eu/countries/germany/1734/vehicles-and-fleet. Accessed on 2020-09-30.

EAFO. Italy - Vehicles and Fleet, September 2020d. URL https://eafo.eu/countries/italy/1739/vehicles-and-fleet. Accessed on 2020-09-30.

EAFO. Knowledge Center, 2020e. URL https://eafo.eu/knowledge-center. Accessed on 2020-10-04.

EAFO. Norway - Incentives and Legislation, July 2020f. URL https://eafo.eu/countries/norway/1747/incentives. Accessed on 2020-10-09.

EEA. Transport greenhouse gas emissions. Technical report, European Environment Agency, November 2019. URL https://www.eea.europa.eu/airs/2018/resource-efficiency-and-low-carbon-economy/transport-ghg-emissions. Accessed on 2020-10-02.

Filippa Egnér and Lina Trosvik. Electric vehicle adoption in Sweden and the impact of local policy instruments. *Energy Policy*, 121:584–596, October 2018. ISSN 03014215. doi: 10.1016/j.enpol.2018.06.040. URL https://linkinghub.elsevier.com/retrieve/pii/S0301421518304282. Accessed on 2020-08-03.

Kathleen M Eisenhardt. Building Theories from Case Study Research. *The Academy of Management Review*, 14(4):20, October 1989. doi: 10.2307/258557. URL https://www.jstor.org/stable/258557.

Sarah Elwood, Michael F. Goodchild, and Daniel Z. Sui. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, 102(3):571–590, May 2012. ISSN 0004-5608, 1467-8306. doi: 10.1080/00045608.2011.595657. URL http://www.tandfonline.com/doi/abs/10.1080/00045608.2011.595657. Accessed on 2020-08-19.

EPSG. World Geodetic System 1984 (EPSG:4326), August 2007. URL https://epsg.io/4326. Accessed on 2020-08-07.

EPSG. European Terrestrial Reference System 1989 (EPSG:4839), November 2012. URL https://epsg.io/4839. Accessed on 2020-08-07.

European Union. Directive 2014/94/EU of the European Parliament and of the Council of 22 October 2014 on the deployment of alternative fuels infrastructure, October 2014. URL https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32014L0094&from=EN. Accessed on 2020-08-19.

European Union. *Environmental Indicator Report 2018: In Support to the Monitoring of the Seventh Environment Action Programme.* Publications Office, LU, 2018. URL https://data.europa.eu/doi/10.2800/180334. Accessed on 2020-10-02.

European Union. Living in the EU, October 2020. URL https://europa.eu/european-union/about-eu/figures/living_en. Accessed on 2020-08-14.

Federal Government of Germany. Climate Action Programme 2030, September 2019. URL https://www.bundesregierung.de/breg-en/issues/climate-action/klimaschutzprogramm-2030-1674080. Accessed on 2020-08-21.

Inês Frade, Anabela Ribeiro, Gonçalo Gonçalves, and António Pais Antunes. Optimal Location of Charging Stations for Electric Vehicles in a Neighborhood in Lisbon, Portugal. *Transportation Research Record: Journal of the Transportation Research Board*, 2252(1):91–98, January 2011. ISSN 0361-1981, 2169-4052. doi: 10.3141/2252-12. URL http://journals.sagepub.com/doi/10.3141/2252-12. Accessed on 2020-08-20.

Simon Árpád Funke, Patrick Plötz, and Martin Wietschel. Invest in fast-charging infrastructure or in longer battery ranges? A cost-efficiency comparison for Germany. *Applied Energy*, 235:888–899, February 2019. ISSN 03062619. doi: 10.1016/j.apenergy.2018.10.134. URL https://linkinghub.elsevier.com/retrieve/pii/S0306261918317021. Accessed on 2020-08-03.

Geofabrik. Geofabrik Download Server, 2018. URL https://download.geofabrik.de/. Accessed on 2020-08-05.

Dimitrios Gkatzoflias, Yannis Drossinos, Alyona Zubaryeva, Pietro Zambelli, Panagiota Dilara, and Christian Thiel. Optimal allocation of electric vehicle charging infrastructure in cities and regions. Technical Report JRC 101040, Joint Research Centre, 2016. URL https://publications.jrc.ec.europa.eu/repository/bitstream/JRC101040/allocatechargingpoints_sciencepolreport_eurreport_online.pdf.

Till Gnann, Simon Funke, Niklas Jakobsson, Patrick Plötz, Frances Sprei, and Anders Bennehag. Fast charging infrastructure for electric vehicles: Today's situation and future needs. *Transportation Research Part D: Transport and Environment*, 62:314–329, July 2018. ISSN 13619209. doi: 10.1016/j.trd.2018.03.004. URL https://linkinghub.elsevier.com/retrieve/pii/S1361920917305643. Accessed on 2020-08-03.

T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Supplement 1):5228–5235, April 2004. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0307752101. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0307752101. Accessed on 2020-09-12.

Bettina Gruen. Topicmodels package, April 2020. URL https://www.rdocumentation.org/packages/topicmodels/versions/0.2-11. Accessed on 2020-09-12.

J.R. Helmus, J.C. Spoelstra, N. Refa, M. Lees, and R. van den Hoed. Assessment of public charging infrastructure push and pull rollout strategies: The case of the Netherlands. *Energy Policy*, 121:35–47, October 2018. ISSN 03014215. doi: 10.1016/j.enpol.2018.06.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0301421518303999. Accessed on 2020-08-03.

Maximilian Holland. Norway In August — Over 70% EV Market Share And Heading Higher, September 2020. URL https://cleantechnica.com/2020/09/02/norway-in-august-over-70-ev-market-share-and-heading-higher/. Accessed on 2020-10-09.

Konstantin Hopf. Mining volunteered geographic information for predictive energy data analytics. *Energy Informatics*, 1(1):4, December 2018. ISSN 2520-8942. doi: 10.1186/s42162-018-0009-3. URL https://energyinformatics.springeropen.com/articles/10.1186/s42162-018-0009-3. Accessed on 2020-08-19.

ISO. ISO 3166 — Country Codes, March 2020. URL https://www.iso.org/iso-3166-country-codes.html. Accessed on 2020-09-20.

Nicolas Jouvin, Pierre Latouche, Charles Bouveyron, Guillaume Bataillon, and Alain Livartowski. Greedy clustering of count data through a mixture of multinomial PCA. *Computational Statistics*, July 2020. ISSN 0943-4062, 1613-9658. doi: 10.1007/s00180-020-01008-9. URL http://link.springer.com/10.1007/s00180-020-01008-9. Accessed on 2020-09-12.

Shashank Kapadia. Evaluate Topic Models: Latent Dirichlet Allocation (LDA) — by Shashank Kapadia — Towards Data Science, August 2019. URL https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0. Accessed on 2020-10-06.

Weiwei Kong, Yugong Luo, Guixuan Feng, Keqiang Li, and Huei Peng. Optimal location planning method of fast charging station for electric vehicles considering operators, drivers, vehicles, traffic flow and power grid. *Energy*, 186:115826, November 2019. ISSN 03605442. doi: 10.1016/j.energy.2019.07.156. URL https://linkinghub.elsevier.com/retrieve/pii/S0360544219314987. Accessed on 2020-08-03.

Erik Lorentzen, Petter Haugneland, Christina Bu, and Espen Hauge. Charging infrastructure experiences in Norway - the worlds most advanced EV market. page 11, October 2017.

Avi Chaim Mersky, Frances Sprei, Constantine Samaras, and Zhen (Sean) Qian. Effectiveness of incentives on electric vehicle adoption in Norway. *Transportation Research Part D: Transport and Environment*, 46:56–68, July 2016. ISSN 13619209. doi: 10.1016/j.trd.2016.03.011. URL https://linkinghub.elsevier.com/retrieve/pii/S1361920916000407. Accessed on 2020-09-17.

Ministero dello Sviluppo Economico. Ecobonus del Ministero dello Sviluppo Economico, 2019. URL https://ecobonus.mise.gov.it/. Accessed on 2020-08-21.

Patrick Morrissey, Peter Weldon, and Margaret O'Mahony. Future standard and fast charging infrastructure planning: An analysis of electric vehicle charging behaviour. *Energy Policy*, 89:257–270, February 2016. ISSN 03014215. doi: 10.1016/j.enpol.2015.12.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0301421515302159. Accessed on 2020-08-03.

Myriam Neaimeh, Shawn D. Salisbury, Graeme A. Hill, Philip T. Blythe, Don R. Scoffield, and James E. Francfort. Analysing the usage and evidencing the importance of fast chargers for the adoption of battery electric vehicles. *Energy Policy*, 108:474–486, September 2017. ISSN 03014215. doi: 10.1016/j.enpol.2017.06.033. URL https://linkinghub.elsevier.com/retrieve/pii/S0301421517303877. Accessed on 2020-08-03.

OCM. About, a. URL https://openchargemap.org/site/about. Accessed on 2020-08-05.

OCM. The Open Charge Map API, b. URL https://openchargemap.org/site/develop/api. Accessed on 2020-10-04.

OCM. Snapshots of current Open Charge Map data, October 2020. URL https://github.com/openchargemap/ocm-data. Accessed on 2020-09-08.

OCM and Christopher Cook. Add a whole French regional network, August 2020. URL https://community.openchargemap.org/t/add-a-whole-french-regional-network/122. Accessed on 2020-09-27.

OSM. About. URL https://www.openstreetmap.org/about. Accessed on 2020-08-05.

OSM. Comparison Google services – OSM, February 2018. URL https://wiki.openstreetmap.org/wiki/Comparison_Google_services_%E2%80%93_OSM. Accessed on 2020-10-04.

OSM. Map Features, July 2019a. URL https://wiki.openstreetmap.org/wiki/Map_Features. Accessed on 2020-08-05.

OSM. Proposed features: Multilinestring, August 2019b. URL https://wiki.openstreetmap.org/wiki/Proposed_features/Multilinestring. Accessed on 2020-10-06.

OSM. Amenity tag: Charging_station, September 2020. URL https://wiki.openstreetmap.org/wiki/Tag:amenity%3Dcharging_station. Accessed on 2020-10-04.

Edzer Pebesma. Simple Features for R, September 2020. URL https://r-spatial.github.io/sf/reference/index.html. Accessed on 2020-09-21.

Edzer Pebesma and Roger Bivand. Spatial indexes coming to sf, June 2017. URL https://www.r-spatial.org/r/2017/06/22/spatial-index.html. Accessed on 2020-08-12.

Martin Ponweiser. *Latent Dirichlet Allocation in R*. Diploma Thesis, Vienna University of Economics and Business, May 2012. URL https://epub.wu.ac.at/3558/1/main.pdf.

R Foundation. Introduction to R. URL https://www.r-project.org/about.html. Accessed on 2020-08-21.

Michel Rose. Macron wants France to be Europe's top clean car producer. *Thomson Reuters*, May 2020. URL https://www.reuters.com/article/us-health-coronavirus-france-autos-idUSKBN2322D6. Accessed on 2020-08-21.

Dagmar Schepp. *Analysis of the E-Mobility Charging Infrastructure Expansion on the Basis of OpenStreetMap Data Using Machine Learning Methods*. M.Sc. Thesis, Otto-Friedrich-Universität Bamberg, March 2020. Unpublished.

Andreas Schroeder and Thure Traber. The economics of fast charging infrastructure for electric vehicles. *Energy Policy*, 43:136–144, April 2012. ISSN 03014215. doi: 10.1016/j.enpol.2011.12.041. URL https://linkinghub.elsevier.com/retrieve/pii/S0301421511010470. Accessed on 2020-08-03.

Javier Serradilla, Josey Wardle, Phil Blythe, and Jane Gibbon. An evidence-based approach for investment in rapid-charging infrastructure. *Energy Policy*, 106:514–524, July 2017. ISSN 03014215. doi: 10.1016/j.enpol.2017.04.007. URL https://linkinghub.elsevier.com/retrieve/pii/S030142151730232X. Accessed on 2020-08-03.

Julia Silge and David Robinson. *Text Mining with R - A Tidy Approach*. March 2020. URL https://www.tidytextmining.com/. Accessed on 2020-09-22.

Heather Smith. Geographic vs Projected Coordinate Systems, February 2020. URL https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/gcs_vs_pcs/. Accessed on 2020-08-08.

Farren Tang. Beginner's Guide to LDA Topic Modelling with R - Identifying topics within unstructured text, July 2019. URL https://towardsdatascience.com/beginners-guide-to-lda-topic-modelling-with-r-e57a5a8e7a25. Accessed on 2020-10-06.

Jochen Topf and Frederik Ramm. OpenStreetMap Data in Layered GIS-Format, November 2019. URL http://download.geofabrik.de/osm-data-in-gis-formats-free.pdf.

Wallbox. EV and EV Charger Incentives in Europe: A Complete Guide for Businesses and Individuals, 2020. URL https://blog.wallbox.com/en/ev-and-ev-charger-incentives-in-europe-a-complete-guide-for-businesses-and-individuals/. Accessed on 2020-08-18.

Yang Wang, Payam Sabzmeydani, and Greg Mori. Semi-Latent Dirichlet Allocation: A Hierarchical Model for Human Action Recognition. In Ahmed Elgammal, Bodo Rosenhahn, and Reinhard Klette, editors, *Human Motion – Understanding, Modeling,* *Capture and Animation*, volume 4814, pages 240–254. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-75702-3. doi: 10.1007/978-3-540-75703-0_17. URL http://link.springer.com/10.1007/978-3-540-75703-0_17. Accessed on 2020-09-14.

Jürgen Wenig, Mariya Sodenkamp, and Thorsten Staake. Battery versus infrastructure: Tradeoffs between battery capacity and charging infrastructure for plug-in hybrid electric vehicles. *Applied Energy*, 255:113787, December 2019. ISSN 03062619. doi: 10.1016/j.apenergy.2019.113787. URL https://linkinghub.elsevier.com/retrieve/pii/S0306261919314746. Accessed on 2020-08-03.

Peter Wittek. Unsupervised Learning. In *Quantum Machine Learning*, pages 57–62. Elsevier, 2014. ISBN 978-0-12-800953-6. doi: 10.1016/B978-0-12-800953-6.00005-0. URL https://linkinghub.elsevier.com/retrieve/pii/B9780128009536000050. Accessed on 2020-09-12.

Laura Ulrike Wölfl. *Parking Location Classification for the Efficient Deployment of Electric Vehicle Charging Infrastructure*. M.Sc. Thesis, Otto-Friedrich-Universität Bamberg, June 2020. Unpublished.

A Appendix

A.1 OCM characteristics

The table below provides an overview of the power levels CSs may have in the OCM dataset. It is important to note here, that the differentiating threshold between normal and high power charging is given at a power level of 40kW. This contradicts the threshold defined by EU legislation (see 2.2.2), where normal power refers to power levels of "less than or equal to 22 kW" (European Union, 2014), while a high power CS "allows for a transfer of electricity to an electric vehicle with a power of more than 22 kW" (European Union, 2014).

Table A.1: Power levels of charger types as provided by OCM

Type	Description
Level 1 : Low	Under 2 kW, usually domestic socket types
Level 2 : Medium	Over 2 kW, usually non-domestic socket type
Level 3: High	40kW and higher

The table below provides an overview of the status types that are supported by the OCM dataset. These do not refer to the CS itself, but rather to the individual connections. As this thesis uses the number of distinct locations (CSs) as base for the analysis this filter has no effect on the outcome. If this filter is applied in future research cases, note that *Partly Operational (Mixed)* is used for cases where some connections are not operational, while others of the same type are. Because the numbers of connections are grouped by type, it can not be derived exactly how many of the corresponding connection type are non-operational.

Table A.2: Status types as provided by OCM

Title	`IsOperational`
Unknown	
Currently Available (Automated Status)	`TRUE`
Currently In Use (Automated Status)	`TRUE`
Temporarily Unavailable	`TRUE`
Operational	`TRUE`
Partly Operational (Mixed)	`TRUE`
Not Operational	`FALSE`
Planned For Future Date	`FALSE`
Removed (Decommissioned)	`FALSE`
Removed (Duplicate Listing)	`FALSE`

The table below provides an overview of the different connection types that are supported by the OCM dataset.

Table A.3: Standards of connection types as provided by OCM

Title	IsDiscontinued	IsObsolete
Avcon Connector	TRUE	FALSE
Blue Commando (2P+E)		
BS1363 3 Pin 13 Amp		
CCS (Type 1)	FALSE	FALSE
CCS (Type 2)	FALSE	FALSE
CEE 3 Pin	FALSE	FALSE
CEE 5 Pin	FALSE	FALSE
CEE 7/4 - Schuko - Type F	FALSE	FALSE
CEE 7/5	FALSE	FALSE
CEE+ 7 Pin	FALSE	FALSE
CHAdeMO		
Europlug 2-Pin (CEE 7/16)	FALSE	FALSE
GB-T AC - GB/T 20234.2 (Socket)	FALSE	FALSE
GB-T AC - GB/T 20234.2 (Tethered Cable)	FALSE	FALSE
GB-T DC - GB/T 20234.3	FALSE	FALSE
IEC 60309 3-pin	FALSE	FALSE
IEC 60309 5-pin	FALSE	FALSE
LP Inductive	TRUE	TRUE
NEMA 14-30	FALSE	FALSE
NEMA 14-50	FALSE	FALSE
NEMA 5-15R	FALSE	FALSE
NEMA 5-20R	FALSE	FALSE
NEMA 6-15	FALSE	FALSE
NEMA 6-20	FALSE	FALSE
NEMA TT-30R	FALSE	FALSE
SCAME Type 3A (Low Power)	FALSE	FALSE
SCAME Type 3C (Schneider-Legrand)	FALSE	FALSE
SP Inductive	TRUE	TRUE
T13 - SEC1011 (Swiss domestic 3-pin) - Type J	FALSE	FALSE
Tesla (Model S/X)	FALSE	FALSE
Tesla (Roadster)	TRUE	FALSE
Tesla Battery Swap	FALSE	FALSE
Tesla Supercharger	FALSE	FALSE
Three Phase 5-Pin (AS/NZ 3123)	FALSE	FALSE
Type 1 (J1772)		
Type 2 (Socket Only)	FALSE	FALSE
Type 2 (Tethered Connector)	FALSE	FALSE
Type I (AS 3112)	FALSE	FALSE
Unknown		
Wireless Charging	FALSE	FALSE
XLR Plug (4 pin)	FALSE	FALSE

The table below provides an overview of the usage types that are supported by the OCM dataset. This filter was not applied in this thesis, because it is concerned with the overall CI and all of the usage types below are accessible to a certain degree.

Table A.4: Usage types as provided by OCM

Title	`IsMembershipRequired`	`IsAccessKeyRequired`
(Unknown)		
Private - For Staff or Customers	FALSE	FALSE
Private - Restricted Access	TRUE	
Privately Owned - Notice Required		
Public		
Public - Membership Required	TRUE	TRUE
Public - Notice Required	FALSE	FALSE
Public - Pay At Location	FALSE	FALSE

A.2 Feature catalogue and groups

The tables below represent the semantic groups and the feature catalogue respectively. The `‘filter’` variable in the feature catalogue shows whether the feature is unique (`‘0’`), whether it is undesired (`‘NA’`) or part of a semantic group. The filter number corresponds to the `‘group’` variable in the semantic group table directly below.

Table A.5: Semantic feature groups

group	title	index
1	group_farm	361
2	group_emergency_services	362
3	group_university	363
4	group_kindergarten	364
5	group_public_building	365
6	group_hospital	366
7	group_stadium	367
8	group_hotel	368
9	group_supermarket	369
10	group_kiosk	370
11	group_ruins	371
12	group_religious_building	372
13	group_parking	373
14	group_glacier	374
15	group_residential_building	375
16	group_school	376
17	group_railway_station	377
18	group_garage	378
19	group_water_tower	379
20	group_graveyard	380
21	group_park	381

Table A.6: Feature catalogue

layer	fclass	description	filter	index
places	city	Often over 100,000 people	0	1
places	town	Generally smaller than a city, between 10,000 and 100,000 people	0	2
places	village	Generally smaller than a town, below 10,000 people	0	3
places	hamlet	Generally smaller than a village, just a few houses	0	4
places	national_capital	A national capital	0	5
places	suburb	Named area of town or city	0	6
places	island	Identifies an island	NA	NA
places	farm	Named farm	1	361
places	dwelling	Isolated dwelling (1 or 2 houses, smaller than hamlet)	0	9
places	region	A region label (used in some areas only)	NA	NA
places	county	A county label (used in some areas only)	NA	NA
places	locality	Other kind of named place	NA	NA
public	police	A police post or station	2	362
public	fire_station	A fire station	2	362
public	post_box	A post box (for letters)	NA	NA
public	post_office	A post office	0	16
public	telephone	A public telephone booth	0	17
public	library	A library	0	18
public	town_hall	A town hall	0	19
public	courthouse	A court house	0	20
public	prison	A prison	0	21
public	embassy	An embassy	0	22
public	community_centre	A public facility which is mostly used by local associations for events and festivities	0	23
public	nursing_home	A home for disabled or elderly persons who need permanent care	0	24
public	arts_centre	A venue at which a variety of arts are performed or conducted, and may well be involved with the creation of those works, and run occasional courses	0	25
public	graveyard	A graveyard	20	380
public	market_place	A place where markets are held	0	27
public	recycling	A place (usually a container) where you can drop waste for recycling	NA	NA
public	recycling_glass	A place for recycling glass	NA	NA
public	recycling_paper	A place for recycling paper	NA	NA
public	recycling_clothes	A place for recycling clothes	NA	NA
public	recycling_metal	A place for recycling metal	NA	NA
public	university	A university	3	363
public	school	A school	16	376
public	kindergarten	A kindergarten (nursery)	4	364
public	college	A college	3	363
public	public_building	An unspecified public building	5	365
health	pharmacy	A pharmacy	0	38
health	hospital	A hospital	6	366
health	doctors	A medical practice	0	40
health	dentist	A dentist's practice	0	41
health	veterinary	A veterinary (animal doctor)	0	42
leisure	theatre	A theatre	0	43
leisure	nightclub	A nightclub, or disco	0	44
leisure	cinema	A cinema	0	45
leisure	park	A park	21	381
leisure	playground	A playground for children	0	47
leisure	dog_park	An area where dogs are allowed torun free without a leash	0	48
leisure	sports_centre	A facility where a range of sports activities can be pursued	0	49
leisure	pitch	An area set aside for a specific sport	0	50
leisure	swimming_pool	A swimming pool or water park	0	51
leisure	tennis_court	A tennis court	0	52
leisure	golf_course	A golf course	0	53
leisure	stadium	A stadium. The area of the stadium may contain one or several pitches	7	367
leisure	ice_rink	An ice rink	0	55
catering	restaurant	A normal restaurant	0	56
catering	fast_food	A fast-food restaurant	0	57
catering	cafe	A cafe	0	58
catering	pub	A pub	0	59

Continued on next page

Table A.6 – continued from previous page

layer	fclass	description	filter	index
catering	bar	A bar. The difference between a pub and a bar is not clear but pubstend to offer food while bars do not	0	60
catering	food_court	A common seating area with various fast-food vendors	0	61
catering	biergarten	An open-air area where food and drinks are served	0	62
accommodation	hotel	A hotel	8	368
accommodation	motel	A motel	0	64
accommodation	bed_and_breakfast	A facility offering bed and breakfast	0	65
accommodation	guesthouse	A guesthouse. The difference between hotel, bed and breakfast, and guest houses is not a strict one and OSM tends to use whatever the facility calls itself	0	66
accommodation	hostel	A hostel (offering cheap accomodation, often bunk beds in dormitories)	0	67
accommodation	chalet	A detached cottage, usually self-catering	0	68
accommodation	shelter	All sorts of small shelters to protect against bad weather conditions	NA	NA
accommodation	camp_site	A camp site or camping ground	0	70
accommodation	alpine_hut	An alpine hut is a building typically situated in mountains providing shelter and often food and beverages to visitors	0	71
accommodation	caravan_site	A place where people with caravans or motorhomes can stay overnight or longer	0	72
shopping	supermarket	A supermarket	9	369
shopping	bakery	A bakery	0	74
shopping	kiosk	A very small shop usually selling cigarettes, newspapers, sweets, snacks and beverages	10	370
shopping	mall	A shopping mall	0	76
shopping	department_store	A department store	0	77
shopping	convenience	A convenience store is a small shop selling a subset of items you might find at a supermarket	0	78
shopping	clothes	A clothes or fashion store	0	79
shopping	florist	A store stelling flowers	0	80
shopping	chemist	A shop selling articles of personal hygiene, cosmetics, and householdcleaning products	0	81
shopping	bookshop	A book shop	0	82
shopping	butcher	A butcher	0	83
shopping	shoe_shop	A shoe shop	0	84
shopping	beverages	A place where you can buy alcoholic and non-alcoholic beverages	0	85
shopping	optician	A place where you can buy glasses	0	86
shopping	jeweller	A jewelry shop	0	87
shopping	gift_shop	A gift shop	NA	NA
shopping	sports_shop	A shop selling sports equipment	0	89
shopping	stationery	A shop selling stationery for private and office use	0	90
shopping	outdoor_shop	A shop selling outdoor equiment	0	91
shopping	mobile_phone_shop	A shop for mobile phones	0	92
shopping	toy_shop	A toy store	0	93
shopping	newsagent	A shop selling mainly newspapersand magazines	NA	NA
shopping	greengrocer	A shop selling fruit and vegetables	0	95
shopping	beauty_shop	A shop that provides personal beauty services like a nail salon ortanning salon	0	96
shopping	video_shop	A place where you can buy films	0	97
shopping	car_dealership	A car dealership	0	98
shopping	bicycle_shop	A bicycle shop	0	99
shopping	doityourself	A do-it-yourself shop where you can buy tools and building materials	0	100
shopping	furniture_shop	A furniture store	0	101
shopping	computer_shop	A computer shop	0	102
shopping	garden_centre	A place selling plants and gardening goods	0	103
shopping	hairdresser	A hair salon	0	104
shopping	car_repair	A car garage	0	105
shopping	car_rental	A place where you can rent a car	0	106
shopping	car_wash	A car wash	0	107
shopping	car_sharing	A car sharing station	0	108
shopping	bicycle_rental	A place where you can rent bicycles	0	109
shopping	travel_agent	A travel agency	0	110
shopping	laundry	A place where you can wash clothes or have them cleaned	0	111
shopping	vending_machine	An unspecified vending machine	NA	NA
shopping	vending_cigarette	A cigarette vending machine	NA	NA

Continued on next page

Table A.6 – continued from previous page

layer	fclass	description	filter	index
shopping	vending_parking	A vending machine for parking tickets	0	114
money	bank	A bank	0	115
money	atm	A machine that lets you withdraw cash from your bank account	NA	NA
tourism	tourist_info	Something that provides information to tourists; may or many not be manned	0	117
tourism	tourist_map	A map displayed to inform tourists	NA	NA
tourism	tourist_board	A board with explanations aimed at tourists	NA	NA
tourism	tourist_guidepost	A guide post	0	120
tourism	attraction	A tourist attraction	0	121
tourism	museum	A museum	0	122
tourism	monument	A monument	0	123
tourism	memorial	A memorial	0	124
tourism	art	A permanent work of art	0	125
tourism	castle	A castle	0	126
tourism	ruins	Ruins of historic significance	11	371
tourism	archaeological	An excavation site	NA	NA
tourism	wayside_cross	A wayside cross, not necessarily old	NA	NA
tourism	wayside_shrine	A wayside shrine	NA	NA
tourism	battlefield	A historic battlefield	0	131
tourism	fort	A fort	0	132
tourism	picnic_site	A picnic site	0	133
tourism	viewpoint	A viewpoint	0	134
tourism	zoo	A zoo	0	135
tourism	theme_park	A theme park	0	136
misc	toilet	Public toilets	NA	NA
misc	bench	A public bench	NA	NA
misc	drinking_water	A tap or other source of drinking water	NA	NA
misc	fountain	A fountain for cultural, decorative,or recreational purposes	NA	NA
misc	hunting_stand	A hunting stand	NA	NA
misc	waste_basket	A waste basket	NA	NA
misc	camera_surveillance	A surveillance camera	NA	NA
misc	emergency_phone	An emergency telephone	NA	NA
misc	fire_hydrant	A firy hydrant	NA	NA
misc	emergency_access	An emergency access point (signposted place in e.g. woods thelocation of which is known to emergency services)	NA	NA
misc	tower	A tower of some kind	NA	NA
misc	tower_comms	A communications tower	NA	NA
misc	water_tower	A water tower	19	379
misc	tower_observation	An observation tower	0	NA
misc	windmill	A windmill	NA	NA
misc	lighthouse	A lighthouse	NA	NA
misc	wastewater_plant	A wastewater treatment plant	0	153
misc	water_well	A facility to access underground aquifers	NA	NA
misc	water_mill	A mill driven by water. Often historic	NA	NA
misc	water_works	A place where drinking water is processed	0	156
pofw	christian	A christian place of worship (usually a church) without one of the denominations below.	12	372
pofw	christian_anglican	A christian place of worship where the denomination is known	12	372
pofw	christian_catholic	A christian place of worship where the denomination is known	12	372
pofw	christian_evangelical	A christian place of worship where the denomination is known	12	372
pofw	christian_lutheran	A christian place of worship where the denomination is known	12	372
pofw	christian_methodist	A christian place of worship where the denomination is known	12	372
pofw	christian_orthodox	A christian place of worship where the denomination is known	12	372
pofw	christian_protestant	A christian place of worship where the denomination is known	12	372
pofw	christian_baptist	A christian place of worship where the denomination is known	12	372
pofw	christian_mormon	A christian place of worship where the denomination is known	12	372
pofw	jewish	A jewish place of worship (usually a synagogue)	12	372

Continued on next page

Table A.6 – continued from previous page

layer	fclass	description	filter	index
pofw	muslim	A muslim place of worhsip, (usually a mosque) without one ofthe denominations below	12	372
pofw	muslim_sunni	A Sunni muslim place of worship	12	372
pofw	muslim_shia	A Shia muslim place or worship	12	372
pofw	buddhist	A Buddhist place of worship	12	372
pofw	hindu	A Hindu place of worship	12	372
pofw	taoist	A Taoist place of worship	12	372
pofw	shintoist	A Shintoist place of worship	12	372
pofw	sikh	A Sikh place of worship	12	372
natural	spring	A spring, possibly source of a stream	NA	NA
natural	glacier	A glacier	14	374
natural	peak	A mountain peak	0	178
natural	cliff	A cliff	0	179
natural	volcano	A volcano	0	180
natural	tree	A tree	NA	NA
natural	mine	A mine	0	182
natural	cave_entrance	A cave entrance	0	183
natural	beach	A beach. (Note that beaches areonly rarely mapped as point features.)	0	184
traffic	traffic_signals	Traffic lights	NA	NA
traffic	mini_roundabout	A small roundabout without physical strucutre, usually just painted onto the road surface	NA	NA
traffic	stop	A stop sign	NA	NA
traffic	crossing	A place where the street is crossed by pedestrians or a railway	NA	NA
traffic	ford	A place where the road runs through a river or stream	NA	NA
traffic	motorway_junction	The place where a slipway enters or leaves a motorway	NA	NA
traffic	turning_circle	An area at the end of a street where vehicles can turn	0	191
traffic	speed_camera	A camera that photographs speeding vehicles	NA	NA
traffic	street_lamp	A lamp illuminating the road	NA	NA
traffic	fuel	A gas station	0	194
traffic	service	A service area, usually along motorways	0	195
traffic	parking	A car park of unknown type	13	373
traffic	parking_site	A surface car park	13	373
traffic	parking_multistorey	A multi storey car park	13	373
traffic	parking_underground	An underground car park	13	373
traffic	parking_bicycle	A place to park your bicycle	NA	NA
traffic	slipway	A slipway	0	201
traffic	marina	A marina	0	202
traffic	pier	A pier	0	203
traffic	dam	A dam	0	204
traffic	waterfall	A waterfall	0	205
traffic	lock_gate	A lock gate	0	206
traffic	weir	A barrier built across a river or stream	0	207
transport	railway_station	A larger railway station of mainline rail services	17	377
transport	railway_halt	A smaller, local railway station, or subway station	0	209
transport	tram_stop	A tram stop	0	210
transport	bus_stop	A bus stop	0	211
transport	bus_station	A large bus station with multiple platforms	0	212
transport	taxi_rank	A taxi rank	0	213
transport	airport	A large airport	0	214
transport	airfield	A small airport or airfield	0	215
transport	helipad	A place for landing helicopters	NA	NA
transport	apron	A apron (area where aircraft are parked)	0	217
transport	ferry_terminal	A ferry terminal	0	218
transport	aerialway_station	A station where cable cars or lifts alight	0	219
roads	motorway	Motorway/freeway	0	220
roads	trunk	Important roads, typically divided	0	221
roads	primary	Primary roads, typically national	0	222
roads	secondary	Secondary roads, typically regional	0	223
roads	tertiary	Tertiary roads, typically local	0	224
roads	unclassified	Smaller local roads	NA	NA
roads	residential	Roads in residential areas	0	226
roads	living_street	Streets where pedestrians have priority	0	227
roads	pedestrian	Pedestrian only streets	0	228
roads	motorway_link	Roads that connect from one road to another of the same or lower category	0	220

Continued on next page

Table A.6 – continued from previous page

layer	fclass	description	filter	index
roads	trunk_link	Roads that connect from one road to another of the same or lower category	0	221
roads	primary_link	Roads that connect from one road to another of the same or lower category	0	222
roads	secondary_link	Roads that connect from one road to another of the same or lower category	0	223
roads	service	Service roads for access to buildings, parking lots, etc	NA	NA
roads	track	For agricultural use, in forests, etc. Often gravel roads	0	234
roads	track_grade1	Tracks can be assigned a 'tracktype' from 1 (asphalt or heavily compacted) to 5 (hardly visible). A detailed description is here: http://wiki.openstreetmap.org/wiki/Tracktype	0	234
roads	track_grade2	Tracks can be assigned a 'tracktype' from 1 (asphalt or heavily compacted) to 5 (hardly visible). A detailed description is here: http://wiki.openstreetmap.org/wiki/Tracktype	0	234
roads	track_grade3	Tracks can be assigned a 'tracktype' from 1 (asphalt or heavily compacted) to 5 (hardly visible). A detailed description is here: http://wiki.openstreetmap.org/wiki/Tracktype	NA	NA
roads	track_grade4	Tracks can be assigned a 'tracktype' from 1 (asphalt or heavily compacted) to 5 (hardly visible). A detailed description is here: http://wiki.openstreetmap.org/wiki/Tracktype	NA	NA
roads	track_grade5	Tracks can be assigned a 'tracktype' from 1 (asphalt or heavily compacted) to 5 (hardly visible). A detailed description is here: http://wiki.openstreetmap.org/wiki/Tracktype	NA	NA
roads	bridleway	Paths for horse riding	NA	NA
roads	cycleway	Paths for cycling	NA	NA
roads	footway	Footpaths	NA	NA
roads	path	Unspecified paths	NA	NA
roads	steps	Flights of steps on footpaths	NA	NA
roads	unknown	Unknown type of road or path	NA	NA
railways	rail	Regular railway tracks	NA	NA
railways	light_rail	Light railway tracks, often commuter railways	NA	NA
railways	subway	Underground railway tracks	NA	NA
railways	tram	Tram tracks (may be incident with roads)	NA	NA
railways	monorail	A monorail track	NA	NA
railways	narrow_gauge	A narrow gauge railway track	NA	NA
railways	miniature	A miniature railway track	NA	NA
railways	funicular	A funicular, or cable railway usually on a steep inlcine	NA	NA
railways	rack	A rack railway	NA	NA
railways	drag_lift	An overhead tow-line for skiers	NA	NA
railways	chair_lift	An open chairlift run	NA	NA
railways	cable_car	A cabin cable car run	NA	NA
railways	gondola	An aerialway where the cabins go around ina circle	NA	NA
railways	goods	An aerialway for the transport of goods	NA	NA
railways	other_lift	Another type of lift	NA	NA
waterways	river	A large river	NA	NA
waterways	stream	A smaller river or stream	NA	NA
waterways	canal	An artificial waterway	NA	NA
waterways	drain	A small drainage ditch or similar structure	NA	NA
landuse	forest	A forest or woodland	0	265
landuse	park	A park	21	381
landuse	residential	A residential area	0	267
landuse	industrial	An industrial area	0	268
landuse	cemetery	A cemetery or graveyard	20	380
landuse	allotments	An area with small private gardens	0	270
landuse	meadow	A meadow, possibly used for grazing cattle	0	271
landuse	commercial	A commercial area	0	272
landuse	nature_reserve	A nature reserve	0	273
landuse	recreation_ground	An open green space for general recreation	21	381
landuse	retail	An area mainly used by shops	0	275
landuse	military	Military landuse, usually no access for civilians	NA	NA
landuse	quarry	A quarry	0	277
landuse	orchard	An area used for growing fruit-bearingtrees	0	278
landuse	vineyard	An area used for growing grapes	0	279
landuse	scrub	An area where scrub grows	0	280
landuse	grass	An area where grass grows	NA	NA
landuse	heath	Heath areas	0	282
landuse	national_park	A national park	0	283

Continued on next page

Table A.6 – continued from previous page

layer	fclass	description	filter	index
landuse	farmland	Agricultural land (areas where crops are grown)	0	284
landuse	farmyard	Area of land with farm buildings and the shrubbery/trees around them	1	361
water	water	Unspecified bodies of water. Typically lakes, but can also be larger rivers, harbours, etc	NA	NA
water	reservoir	Artificial lakes, typically above a dam	0	287
water	river	Polygons for larger rivers	0	288
water	dock	Dock (to repair ships, don't confuse it with the American term "dock")	0	289
water	glacier	Glaciers	14	374
water	wetland	Swamp, bog, or marsh land	0	291
building	apartments	A building arranged into individual dwellings, often on separate floors. May also have retail outlets on the ground floor.	15	375
building	bungalow	A single-storey detached small house, Dacha.	15	375
building	cabin	A cabin is a small, roughly built house usually with a wood exterior and typically found in rural areas.	0	294
building	detached	A detached house, a free-standing residential building usually housing a single family.	15	375
building	dormitory	For a shared building, as used by college/university students (not a share room for multiple occupants as implied by the term in British English).	0	296
building	farm	A residential building on a farm (farmhouse).	1	361
building	ger	A permanent or seasonal round yurt or ger used as dwelling.	NA	NA
building	hotel	A building designed with separate rooms available for overnight accommodation.	8	368
building	house	A dwelling unit inhabited by a single household (a family or small group sharing facilities such as a kitchen).	15	375
building	houseboat	A boat used primarily as a home.	15	375
building	residential	A general tag for a building used primarily for residential purposes.	15	375
building	semidetached_house	A residential house that shares a common wall with another on one side.	15	375
building	static_caravan	A mobile home (semi)permanently left on a single site.	15	375
building	terrace	A single way used to define the outline of a linear row of residential dwellings, each of which normally has its own entrance, which form a terrace (row-house in North American English).	15	375
building	commercial	A building where non-specific commercial activities take place, not necessarily an office building.	0	306
building	industrial	A building where some industrial process takes place. Use warehouse if the purpose is known to be primarily for storage/distribution.	0	307
building	kiosk	A small one-room retail building.	10	370
building	office	An office building.	0	309
building	retail	A building primarily used for selling goods that are sold to the public.	0	310
building	supermarket	A building constructed to house a self-service large-area store.	9	369
building	warehouse	A building primarily used for the storage or goods or as part of a distribution system.	0	312
building	cathedral	A building that was built as a cathedral.	12	372
building	chapel	A building that was built as a chapel.	12	372
building	church	A building that was built as a church.	12	372
building	mosque	A mosque	12	372
building	religious	Unspecific religious building.	12	372
building	shrine	A building that was built as a shrine.	12	372
building	synagogue	A building that was built as a synagogue.	12	372
building	temple	A building that was built as a temple.	12	372
building	bakehouse	A building that was built as a bakehouse (i.e. for baking bread).	0	321
building	civic	For any civic amenity.	5	365
building	fire_station	A building which houses fire fighting equipment ready for use.	2	362
building	government	For government buildings in general, including municipal, provincial and divisional secretaries, government agencies and departments, town halls, (regional) parliaments and court houses.	5	365

Continued on next page

Table A.6 – continued from previous page

layer	fclass	description	filter	index
building	hospital	A building which forms part of a hospital.	6	366
building	kindergarten	For any generic kindergarten buildings.	4	364
building	public	A building constructed as accessible to the general public (a town hall, police station, court house, etc.).	5	365
building	school	For any generic school buildings.	16	376
building	toilets	A toilet block	NA	NA
building	train_station	A building constructed to be a train station building, including buildings that are abandoned and used nowadays for a different purpose.	17	377
building	transportation	A building related to public transport.	0	331
building	university	A university building.	3	363
building	barn	An agricultural building used for storage and as a covered workplace.	0	333
building	conservatory	A building or room having glass or tarpaulin roofing and walls used as an indoor garden or a sunroom (winter garden).	NA	NA
building	cowshed	A cowshed (cow barn, cow house) is a building for housing cows, usually found on farms.	0	335
building	farm_auxiliary	A building on a farm that is not a dwelling (use 'farm' or 'house' for the farm house).	1	361
building	greenhouse	A greenhouse is a glass or plastic covered building used to grow plants.	0	337
building	stable	A stable is a building where horses are kept.	0	338
building	sty	A sty (pigsty, pig ark, pig-shed) is a building for raising domestic pigs, usually found on farms.	0	339
building	grandstand	The main stand, usually roofed, commanding the best view for spectators at racecourses or sports grounds.	0	340
building	pavilion	A sports pavilion usually with changing rooms, storage areas and possibly an space for functions & events.	0	341
building	riding_hall	A building that was built as a riding hall.	0	342
building	sports_hall	A building that was built as a sports hall.	0	343
building	stadium	A building constructed to be a stadium building, including buildings that are abandoned and used nowadays for a different purpose.	7	367
building	hangar	A hangar is a building used for the storage of airplanes, helicopters or space-craft.	0	345
building	hut	A hut is a small and crude shelter.	NA	NA
building	shed	A shed is a simple, single-storey structure in a back garden or on an allotment that is used for storage, hobbies, or as a workshop.	NA	NA
building	carport	A carport is a covered structure used to offer limited protection to vehicles, primarily cars, from the elements.	18	378
building	garage	A garage is a building suitable for the storage of one or possibly more motor vehicle or similar.	18	378
building	garages	A building that consists of a number of discrete storage spaces for different owners/tenants.	0	350
building	parking	Structure purpose-built for parking cars.	13	373
building	digester	A digester is a bioreactor for the production of inflatable biogas from biomass.	0	352
building	service	Service building usually is a small unmanned building with certain machinery (like pumps or transformers).	NA	NA
building	transformer_tower	A transformer tower is a characteristic tall building comprising a distribution transformer and constructed to connect directly to a medium voltage overhead power line.	NA	NA
building	water_tower	A water tower	19	379
building	bunker	A hardened military building.	NA	NA
building	bridge	A building used as a bridge. Can also represent a gatehouse for drawbridges.	NA	NA
building	construction	Used for buildings under construction.	NA	NA
building	roof	A structure that consists of a roof with open sides, such as a rain shelter, and also gas stations.	NA	NA
building	ruins	Frequently used for a house or other building that is abandoned and in poor repair. However, some believe this usage is incorrect, and the tag should only be used for buildings constructed as fake ruins (for example sham ruins in an English landscape garden).	11	371

Declaration of Authorship / Eigenständigkeitserklärung

Ich erkläre hiermit gemäß § 9 Abs. 12 APO, dass ich die vorstehende Masterarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Cedric Dean Easton
Matrikelnummer: 1919422
E-Mail: cedric-dean.easton@stud.uni-bamberg.de
Studiengang: M.Sc. Wirtschaftsinformatik

Ort, Datum Unterschrift