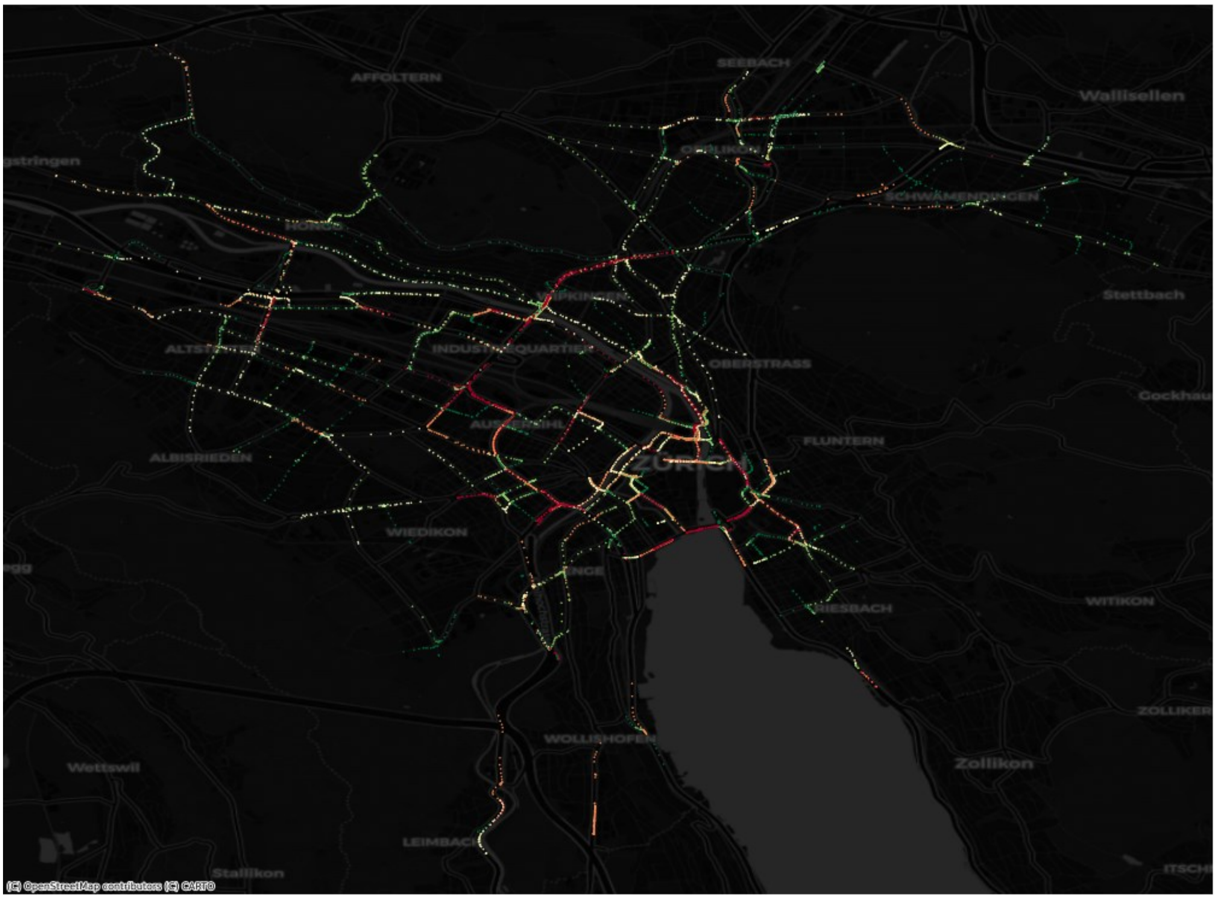
Understanding the relationship between urban mobility and cycle network

EPA1316 – Introduction to Urban Data Science – Group 15

Ziyi Liu – 5006430
Joppe van Veghel – 5199115
Anniek van Veldhuizen – 4454200
Chamon Wieles – 4448200



Abstract

In order to mitigate the effect of climate change, policy makers need to get insights on traffic flows, congestion and the ways citizens travel. By finding out the correlation between traffic flows in the cities and the accessibility bicyclist experience in the urban environment, policy makers can make well-grounded decisions. In this research, we aim to help this challenge by answering the research question: 'What is the effect of urban cycle network on the congestion and sustainability in and of a city?'. In addition, we aim to find out how the use of lime e-scooters relate to the traffic flows. Traffic flow data (UTD19 from ETH Zurich), OpenStreetMap data, Urban Typologies data and lime scooter data is used in answering the research question. After performing several analyses, it can be stated that increasing the bicycle network will have a positive effect on congestion and thus on sustainable transition in a city.

Introduction

Urban passage mobility accounts for over 2 Gt CO_2e greenhouse gas (GHG) emission in 2016, equivalent to over 4% of global annual emission (Oke et al., 2019). To mitigate the mounting climate change, increasing the level of bicycle mode share is a desirable goal, but that it will not happen if people believe that there are too few safe places to cycle (Transportation Research Board, 1992). A lot of research has been done on cycling in the cities and the effect this has on the environment and health. Due to policy making, an increase cycling can be created by improving both actual and perceived environmental conditions (Moudon et al., 2005). However, little study has been done on the effect of vehicle sharing systems. It has been found that bicycle-sharing systems have a positive effect on reducing rush-hour congestion (Wang & Zhou, 2017). However, the effect of e-scooters sharing systems has not yet been researched. For instance, what is the effect of e-scooters sharing on urban mobility in the first place, and what is the effect of cycle network on e-scooters use?

Congestion is a critical issue in the field of urban mobility. Loder, et al. (2019) investigates the relationship between urban vehicle road and bus network typology and traffic congestion using UTD19 dataset[1]. However, the effect of urban cycle network including bicycle parking amenity, cycle lanes and mobility sharing systems on the congestion remains to be explored. Therefore, this research aims to fill the knowledge gap on the relationship between urban congestion and cycle network. Besides UTD19 dataset, Urban Typologies dataset[2], geographic information from OpenStreetMap[3] and Lime e-scooters location dataset[4] are collected and applied to the research. Both inter-city-level and city-level analysis will be conducted to answer the research question:

What is the effect of urban cycle network on the congestion and sustainability in and of a city?

Figure 1 illustrates the research flow diagram with six major steps, namely literature review, dataset collection, data cleaning, exploratory data analysis, data analysis and conclusion. And the data analysis part future segments into three sub-steps including inter-city benchmarking, city-level analysis on congestion and cycle network, and on the effect of e-scooters.

---

[1] UTD19 is a publicly available dataset including the vehicle speed, flow and density data by 23541 stationary detectors on urban roads in 40 cities worldwide. Retrieved from: https://utd19.ethz.ch/

[2] Urban Typologies is a publicly available dataset including 64 indicators to categorize 12 urban typologies for 331 cities worldwide. Retrieved from: http://web.mit.edu/afs/athena.mit.edu/org/i/its-lab/www/dashboard/new%20dashboard/index.html

[3] OpenStreetMap is a collaborative project including free-to-use geodata worldwide under an open license. Retrieved from: https://www.openstreetmap.org/

[4] Lime is one of the largest dockless electric scooters operators around the world. The e-scooters location dataset is collected by the research team.

The findings of this research showed that on an inter-city level, no relevant correlations were found between the CO2 emissions, cycle network shares, congestion and bicycle. When zooming in on Zurich, Birmingham, Bordeaux, Frankfurt, Manchester & Toulouse, a city level analysis showed that also no significant correlation between traffic congestion and bicycle parking capacity was found. When looking at traffic congestion versus cycle lane lengths, for Zurich, Manchester, and Toulouse it could be inferred that less congestion is the causality of high cycle lanes accessibility. Regarding the Lime e-scooters, no clear relations where found between the location of the scooters and traffic flow, bicycle parking & cycling paths. However, it can be stated that people prefer to use Lime scooters near bicycle paths, indicating that the total length of bicycle paths do have an effect on the use of Lime scooters.

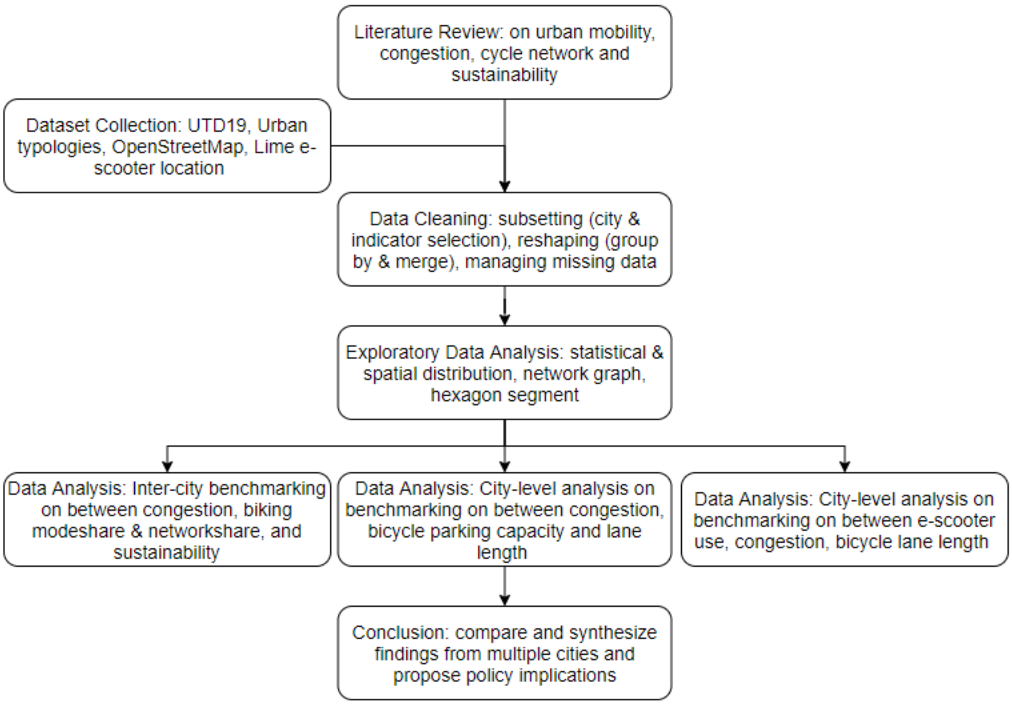


Figure 1: Research flow diagram

In this research possible relations have been investigated in order to be able to write a policy advise for improving the sustainability of a city. The gap that is tackled is that little is known yet on the effect of upcoming shared e-scooters in cities. The improvement of sustainability can be done by reducing congestion in the city, which leads to less CO2 emissions. Possible solutions for congestion reducing travel modes are cycling and Lime e-scooters. It has been shown that an improved bicycle network can have a desired effect on the use of both travel means, leading to less congestion and thus a more sustainable city. The policy advise that flows from this research will thus be: Improvement of the bicycle network.

Theoretical fundamentals

It is already known what the effect of congestion is on air quality of the city and how this relates to sustainability. Congestion leads to 'stop-and-go traffic conditions', leading to an increased number of acceleration and deceleration events, which lead to increased CO_2 emissions. Therefore, congestion is major factor when estimating road traffic emissions (Grote, Williams, Preston, & Kemp, 2016). To guarantee a more sustainable environment, these CO_2 emissions should be reduced (Akerboom, Botzen, Buijze, Michels, & Rijswick, 2020).

(Loder, et al., 2019) argued that vehicle road and bus network characteristic, such as network density, betweenness centrality and bus production, could interpret most of congestion. However, it also indicates a decreasing benefit of network expansion. However, the role of cycle network is ignored, which could potentially improve the biking behaviours and thereby reduce the usage of vehicles as well as congestion. Such hypothesis implicates a policy shift in terms of infrastructure investment from vehicle road network into cycle lane network.

Besides, as an emerging mobility mode, the effect of e-scooter sharing is less addressed in literature. More and more cities are being taken over by companies that put shared scooters in the cities. Despite that not all municipalities welcome the e-scooters they have the potential to fill an important role in urban mobility when solutions to congestion and pollution are urgently needed. (Schellong, Sadek, Schaetzberger, & Barrack, 2019).

(Oke et al., 2019) synthesise 64 urban indicators into 9 factors and categorizes 331 cities into 12 urban typologies through an exploratory factor analysis. To construct a comparable data pipeline, the cities in the MassTransit (either Heavyweight or Moderate) typologies are selected to conduct the inter-city benchmarking. Those cities are mostly located in developed countries and have a relatively advanced infrastructural network, which could mitigate the influences of socio-economic gap on the analysis (Oke et al., 2019). In a future step, six common cities of UTD19 and Urban Typologies datasets are shortlisted for the city-level analysis, namely Zurich, Toulouse, Bordeaux, Frankfurt, Manchester & Birmingham. These cities are all middle-size European cities with the similar congestion level, therefore, it is interesting to zoom into the cities and explore the spatial distribute of the congestion & cycle network.

In this study, the focus is to investigate the effect of urban cycle network on the congestion and sustainability. More specifically, four research sub-questions will be answered:

- SQ1: What is the relationship between congestion, biking and sustainability based on city-level benchmarking?
- SQ2: What is the effect of cycle network including parking capacity and lane length on the congestion level within a city?
- SQ3: What the relationship between e-scooter use, congestion and cycle lanes network within a city?
- SQ4: What conclusions can be derived when comparing the synthesising the finding from multiple cities?

SQ1 corresponds to the inter-city benchmarking, which will give a holistic view on the cities of MassTransit Typologies and set up a basis for city-level analysis. To answer SQ2, we will connect the congestion with the urban cycle-related network in 6 cities, including bicycle parking amenity and cycle lanes. And in SQ3, we further the discussion to explore the impact of the e-scooter sharing as a novel part of cycle network. In the end, the finding in the last three SQs will be concluded to fill the knowledge gap on the relationship of urban congestion and cycle network as well as answer SQ4. Those finding could contribute to the urban policymaking towards sustainable mobility future.

Data pre-processing

Data collection

Four major datasets (shown in Table 1) are utilized in this research. The UTD19 dataset is used for the traffic flow indicator and the Urban Typologies is used for data on congestion, bicycle modeshare and CO2 emission. The empirical data from UTD19 is recorded from stationary traffic sensors including vehicle speed, flow or density (Loder, Ambühl, Menendez, & Axhausen, 2019). The cycle network-related data is obtained from OpenStreetMap with the help of OSMnx python package (Boeing, 2017). The data retrieved from OpenStreetMap covers city administrative boundary, vehicle road and cycle lanes network and bicycle amenities (e.g. parking and rental). The e-scooter location data is retrieved from a constructed Lime API to indicate the usage pattern at the street-level.

Table 1 Data sources and main indicators

Dataset	Source/access	Indicator	Coverage & granularity
UTD19	EHT Zurich, publicly accessible	Traffic flow	Street-level, 40 cities
Urban Typologies	MIT, publicly accessible	Congestion Bicycle modeshare CO2 Emissions	City-level, 331 cities
Cycle network	OpenStreetMap, access by OSMnx (Boeing, 2017)	Bicycle parking capacity Cycle lanes length Cycle network share	Street-level, worldwide
E-scooter locations	Lime, accessible through application	E-scooter locations	Street-level, 100+ cities

Lime is a shared electric scooter company that offers their users to use any scooter that is available in real time. The lime dataset is built by requesting the location of all available lime scooters in a city at several moments per day. By requesting this data at the server of the company, real-live data is obtained. This is done at multiple moments on a day to see the changes in available scooters. Only the scooters that have changed location will be taken into the dataset as a 'new' scooter. The scooters that did not change location, thus, will only be included once. Preventing assigning extra weight to scooters that did not move overtime by taking them into account more than once. At each data-request, for every scooter, the longitude, latitude and moment of last activity is selected. In total the dataset consists out of 4000+ scooter locations. (Further elaboration can be found in Appendix A).

To be able to perform city-level and inter-city-level analyses the following variables were chosen: congestion (%), cycle network share (%), bicycle modeshare (%) and CO_2 emissions (metric tonnes). The congestion, bicycle modeshare and the CO_2 emissions (metric tonnes). These variables are retrieved from the Urban Typologies dataset. The cycle network share is introduced to reflect the development and importance of cycle network in the urban network. Cycle network share is defined as the ratio of total street length for the bike network type and the total street length of all network types, which are calculated based on the network graph retrieved separately from OpenStreetMap for each city. The CO_2 emissions in the Urban Typology dataset were given as emissions in metric tons per capita per year. Considering that larger cities produce a lot of CO_2 emissions but may have less emissions per capita due to the high population, the total CO_2 emissions in metric tonnes were calculated by multiplying the CO_2 emissions in metric tonnes with the population living in the urban agglomeration.

Data cleaning

As for Urban Typologies, the dataset is filtered into 49 cities within MassTransit Typologies, of which 40 cities could be used for the further analysis after the data cleaning process. When comparing the UTD19 dataset to the Urban Typologies dataset six common cities were found on which city level analysis could be performed. Table 2 gives an overview of these cities. As it is shown, the six cities all belong to the category of MassTransit Moderate and exhibit similar characteristics on congestion level, urbanization rate, economic development, which will mitigate the impact of irrelevant socio-economic factors for the analysis. On the contrary, those cities have huge difference on the mobility modeshare, cycle network share as well as GHG emission. Therefore, we conduct a solid data pipeline to investigate the relationship between mobility and street network and thereby to conduct the comparison and synthesis.

For the lime dataset, after retrieving all the data from the application, the relevant data is selected. This consisted of the scooter-id, the location (longitude and latitude) and the last time the scooter was used. The longitude and latitude have been combined into a point geometry location. If the same scooter is observed multiple times (same scooter-id, location and time used), this observation is removed.

When retrieved, the UTD19 data was divided into three datasets; traffic flow per detector, location of each detector and the lane on which the detector is and with consists traffic dataset is merged with link and detector files using the link id and detector id respectively to connect the traffic flow with geographical locations. The traffic flow is used as an indicator to show locations with the highest possibility for congestion.

By defining the geometry information for traffic and cycle network, the geographical relationship among them can be established. In order to get a fine granularity, a hexagon grid with the radius of around 1 kilometre is constructed for city-level analysis to group all the geometries into different segmentations (e.g. 3762 hexagons for Zurich). The average value of traffic flow, the sum value of bicycle parking capacity, the sum value of bicycle lanes length, the total number of e-scooters in each segmentation is calculated and related to the hexagon geometry. This is the basis of the analysis. What's more, in each future analysis, only hexagons with no NAN values for all the concerned indicators will be kept to obtain the results.

Exploratory Data Analysis

The data for the inter-city is scaled between 0 and 1. By doing this, the different variables can be compared to each other to find how the variables relate. A pair-plot, shown in Figure 22 Appendix C, is made to visualize potential relations. To find out whether the indicators are correlated, the correlation coefficient had to be found. The table that shows correlations between each indicator can be seen in Appendix C. In Table 2, the basic characteristics from the Urban Typologies dataset for the six selected cities can be seen. Most cities have comparable values for congestion, urbanization rate and GDP per capita. Zurich has the highest public transit modeshare (46.3% compared to the second highest (20% for Frankfurt)). Even though Frankfurt has the highest walking modeshare (31% compared to 5.7% for Zurich), their CO_2 emission per capita per year is the largest with 13.1 tonnes compared 6.7 tonnes for Zurich. We have to keep in mind the possibility that the data is biased and that the way these data is observed can differ between cities.

Table 2 Basic characteristic of six cities for city-level analysis

City	**Toulouse**	**Bordeaux**	**Frankfurt**	**Zurich**	**Birmingham**	**Manchester**
Typology	MassTransit Moderate					
Country	France	France	Germany	Switzerland	UK	UK
Car Modeshare (%)	66.0	67.0	38.0	42.9	64.8	70.6
Public Transit Modeshare (%)	9.0	9.0	20.0	46.3	17.0	15.0
Bicycle Modeshare (%)	3.0	3.0	11.0	4.1	1.5	2.2
Walking Modeshare (%)	21.0	21.0	31.0	5.7	8.6	11.0
Congestion (%)	27.0	31.0	28.0	31.0	26.0	38.0
Urbanization Rate (%)	79.5	79.5	75.3	73.9	82.6	82.6
GDP per Capita (USD)	50061.0	41631.0	69523.0	56666.0	31572.0	38151.0
CO_2 Emissions per Capita per Year (tonnes)	6.4	5.8	13.1	6.7	4.7	5.0
Cycle network share (%, OpenStreetMap)	79.6%	79.6%	71.9%	54.6%	78.9%	84.9%

Zooming into the city-level, the UTD19 dataset is used to generate the traffic flow heatmap across the street network in each city. As shown in Figure 2 for Zurich, the most congested streets, indicated with the red colour, are in the city centre and on certain streets in the suburbs. The visualisations for the five other cities are presented in Appendix C. Figure 3 visualizes the data from OpenStreetMap. The left figure showcase the spatial distribution of bicycle parking amenities and its relations with the locations of traffic detectors, and the middle figure original cycle network graph retrieved and the right two figures illustrated the gpkg and shp files for network nodes and edges respectively.

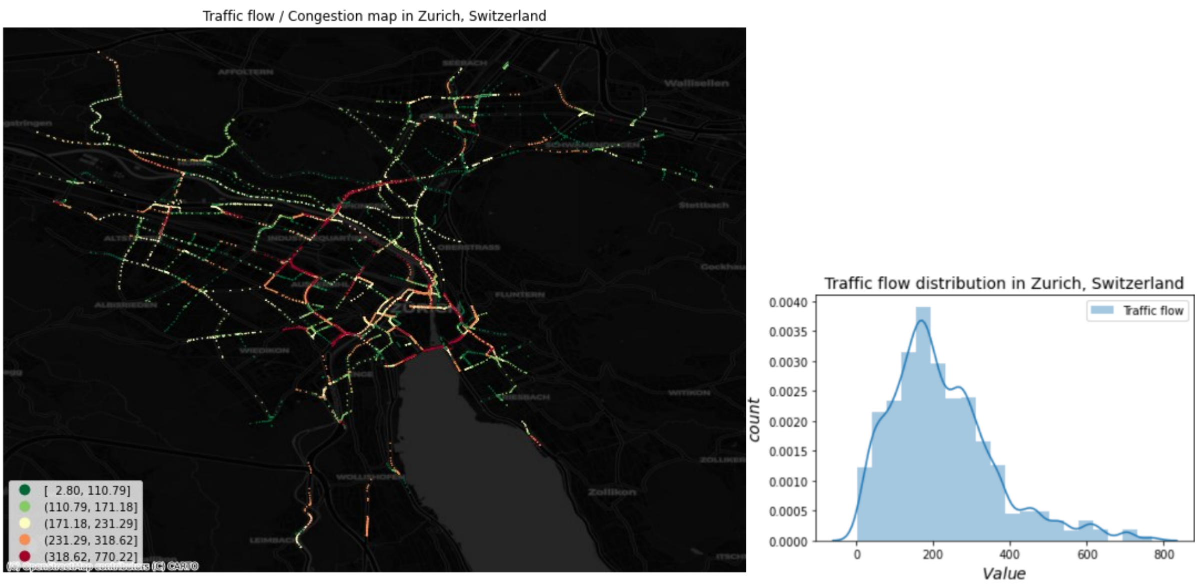


Figure 2 Traffic flow (congestion) heatmap (left) & statistical distribution (right) in Zurich, Switzerland

Figure 3 Distribution of traffic detectors & bicycle parking amenities (left), Cycle lanes network (middle), network nodes(gpkg, up right) and network edges (shp, bottom right) in Zurich, Switzerland

When assigning values for the indicators to the hexagon grid, Figure 4 and Figure 5 are obtained. These figures represent choropleth maps for traffic flow, parking, cycle lanes length and e-scooters usage. The hexagon grid covers the whole region, but hexagons with NAN value are set as close to transparent. In general, high concentrations of vehicle flow, amenities and network density can be found in centres, therefore, we introduce a new indicator x per vehicle flow in the city-level analysis to represent the degree of abundance of x. Specifically, bicycle parking capacity/traffic flow will reflect the abundant/scarce degree of such bicycle amenity; and cycle lanes length/traffic flow will reflect the cycle lanes accessibility. In this way, we could mitigate the bias from the absolute value due to the huge differences of infrastructure density between city centre and suburbs. As for the in-depth analysis with e-scooter data, a higher granular hexagon grid is constructed to explore the potential differences on the results. The definition of the hexagon grid can affect the results and this should be kept in mind when doing the analysis.

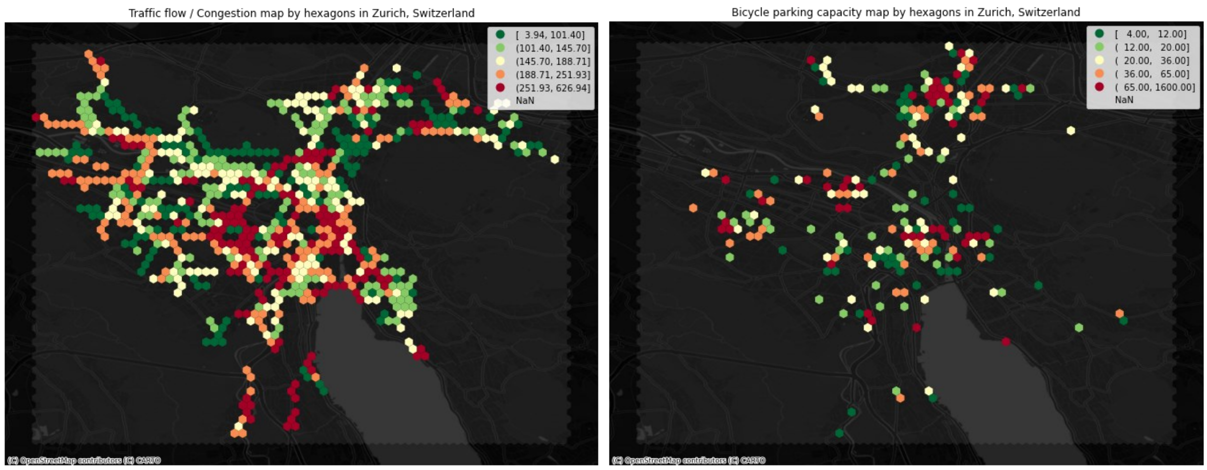


Figure 4 The choropleth map for traffic flow (left) and bicycle capacity (right) by hexagon in Zurich, Switzerland

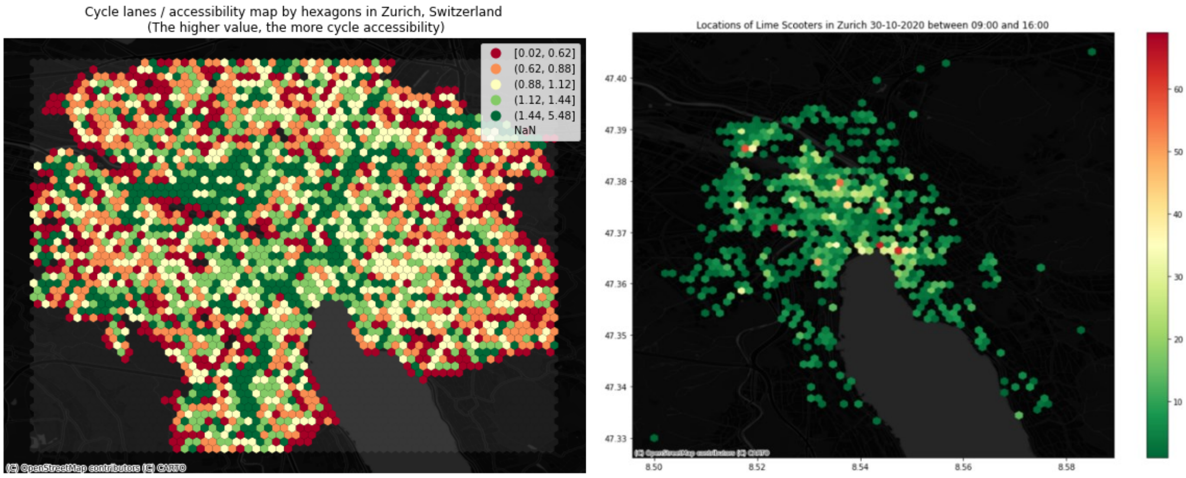


Figure 5 The choropleth map for cycle lanes length (left) and e-scooter usage (right) by hexagon in Zurich, Switzerland

Data Analysis & Results

Method

To investigate the effect of urban cycle network on the congestion and sustainability in and of a city, it must be researched whether a correlation exists between certain urban characteristics. This correlation is searched for at two different levels: inter-city level and city level. Several analyses are conducted in answering the research (sub)-questions. We have to keep in mind that correlation does not mean causation.

First, an analysis on correlation was performed on an inter-city level. At this level, a simple correlation is tried to be found between characteristics of 40 Mass-Transit cities around the world. To be able to find correlation, the data had to be scaled to values between 0 and 1. For city level research, there has been more attention for the spatial autocorrelation, which is he correlation among characteristics due to their relative location proximity. In other words, whether the observed value at a location is influenced (in a positive or negative way) by its immediate neighbors rather than a randomly assigned value. Such analysis will be conducted in six cities of interest. The rationale behind the method selection is that the global autocorrelation could provide the overview for the spatial distribution within the city-level and local autocorrelation will show the best/worst/outlier clusters, which could mitigate the randomization effect of the hexagon grid (e.g. a hexagon might accidently excludes the cycle lanes along with the same vehicle road so that this hexagon will be identified as less accessible by biking than it should be.)

Moran's I statistic is used as an indicator for the global spatial autocorrelation. This statistic takes into account the spatial lag and the deviation of the variable from its mean. Based on the significance of the Moran's I, one might reject the null-hypothesis of spatial randomness; if the significance level is below a certain predefined threshold, it can be assumed that the values are not spatially random. The data can be visualized in the Moran scatter plot including a line with the slope of the Moran's I. Based on this scatterplot, four quadrants can be made (HH, LL, HL, LH (H: high, L: low). The HH and LL quadrant show us a positive spatial autocorrelation, HL and LH negative spatial autocorrelation. In a further step, Local Indicators of Spatial Association (LISA) LISA provides a statistic for each location with its significance. It shows which areas are in what way showing some type of clustering.

In a bivariate Moran's statistic, two variables are analyzed. This is done in such way that the value of one and the spatial lag of the other variable are combined. It does not take into account the correlation between two variables at one location, but, rather, it compares one variable in one location to the other variable in another location. In this situation it will look at whether the number of scooters in one location is affected by the traffic flow in a neighboring location.

The in-depth analysis on Zurich using the retrieved Lime free-floating e-scooter data consists of three parts. First, the location of the Lime scooters is analysed for a relation with the traffic congestion. This is done in three hexagon grid sizes (hexagon grid radius of 5 m, 10 m and 50 m). These values are based upon the likely width of a street and are aimed to show us what the effect is in the direct surroundings of busy streets. Secondly, the location of the Lime scooters is analysed on a relation with the bicycle parking capacity. Thirdly, and finally, the location of the Lime scooters is analysed on a relation with the presence of cycle paths.

Results

Inter-city analysis – MassTransit urban typology

A pair-plot, see is made to visualize potential relations between the different variables. This pair-plot is shown in Appendix C. To find out whether the indicators are correlated, the correlation coefficient had to be found. These correlation coefficients are shown in Table 7, which can be found in Appendix C. No significant correlations were found between the any of the indicators. These results necessitate the analysis to zoom into the city-level.

City-level analysis – Zurich, Birmingham, Bordeaux, Frankfurt, Manchester & Toulouse

Traffic congestion versus Bicycle parking capacity

Table 3 demonstrates the correlation between flow and capacity for Zurich. The flow and capacity show a slight positive correlation. On the contrary, the introduced indicator could better reflect the abundancy of such bicycle amenity and exhibit a weak negative correlation with traffic flow. According to the four quantiles identified in the choropleth map for traffic flow (1-4 from low to high traffic flow, Figure 4), a similar correlation analysis is conducted for different congestion scenarios. It could be observed that the correlation improves in the middle range. Overall, we cannot conclude significant correlation between traffic congestion and bicycle parking capacity in Zurich, while the similar results are also obtained in the other 5 cities. One potential reason could be observed from the Figure 6 that the most of the bicycle capacity concentrate on a lower range without enough variance, regardless of some extreme outliers with a rather high value.

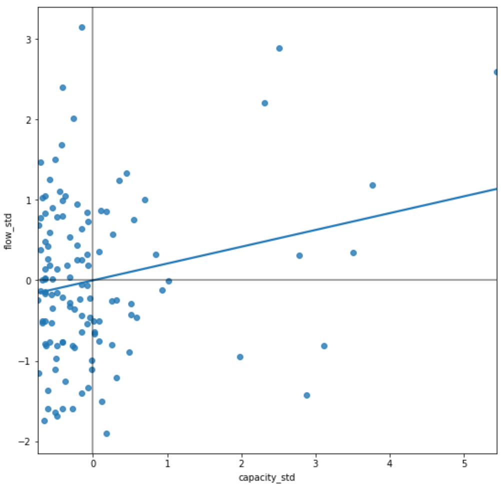


Figure 6 Scatterplot of standardized traffic flow versus standardized bicycle parking capacity

Table 3 Correlation coefficient between traffic flow and bicycle parking capacity under different congestion scenarios

Correlation coefficient		flow	capacity	capacity/flow
≤ quantile 1	flow	1	0.034935	-0.32111
≤ quantile 2	flow	1	0.016573	-0.36839
≤ quantile 3	flow	1	-0.06261	-0.43673
≤ quantile 4	flow	1	0.007692	-0.40809
All	flow	1	0.208973	-0.3198

However, it is still interesting for policymakers to know which regions lack of the bicycle amenity but with rather high travel demand. The red regions in Figure 7 showcases the most scare regions, especially for the LL region from LISA (right figure), which cluster requires the urgent attention to expand the bicycle parking capacity.

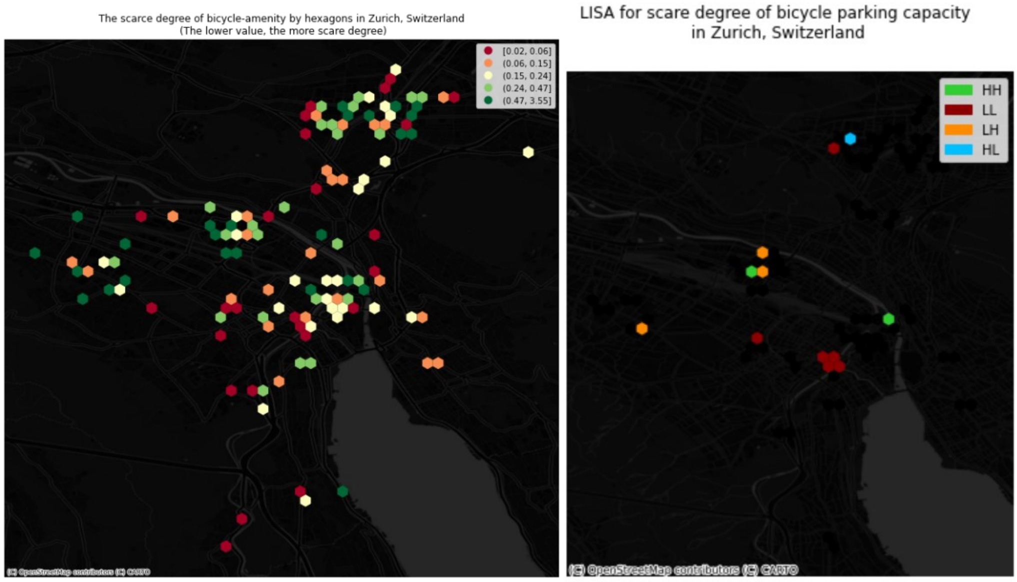


Figure 7 The choropleth map for the scare degree of bicycle parking amenity and its LISA result in Zurich, Switzerland

Traffic congestion versus Cycle lanes length

Then we switch to the analysis about cycle lanes length, on which a much larger dataset is retrieved compared to the bicycle parking capacity. Still taking Zurich as an example, Table 4 demonstrates the results from the statistical correlation analysis where a slightly better correlation is observed between traffic flow and bicycle lane length. Especially in the least congested areas (≤ 1 Quantiles), the corelation coefficient is -0.56 for Zurich, 0.57 for Manchester, and 0.64 for Toulouse (Seen in the Table 4). It could be inferred that an advanced cycle lane network has been established in those areas, so that less congestion is the causality of high cycle lanes accessibility. One way to test the inference is to investigate the traffic flow and cycle network changes over the last decade combined with the investigation on biking culture and urban policies, which, however, is out of the scope of this research. As for Frankfurt, a consistent acceptable correlation is observed which indicates an evenly development of cycle lane and vehicle road network across the city.

Table 4 Correlation coefficient between traffic flow and cycle lanes length under different congestion scenarios and regions

		All region		City-centre region	
Correlation coefficient		length	length/flow	length	length/flow
≤ quantile 1	flow	-0.06885	-0.56656	0.123361	-0.61745
≤ quantile 2	flow	0.071468	-0.48631	0.223301	-0.51572
≤ quantile 3	flow	0.109878	-0.46057	0.218991	-0.48219
≤ quantile 4	flow	0.161188	-0.42444	0.106514	-0.46742
All	flow	-0.00285	-0.33959	0.105092	-0.38243

Table 5 Correlation coefficient between traffic flow and length/flow under different congestion scenarios for 5 other cities

	Toulouse	Manchester	Frankfurt		Birmingham	Bordeaux
	Correlation coefficient between traffic flow and cycle lane accessibility					
≤ quantile 1	-0.645149	-0.56656	-0.578352	All	-0.583440	-0.338359
≤ quantile 2	-0.634494	-0.48631	-0.619638	≥ quantile 1	-0.496004	-0.525803
≤ quantile 3	-0.566558	-0.46057	-0.600981	≥ quantile 2	-0.464077	-0.429668
≤ quantile 4	-0.525010	-0.42444	-0.620804	≥ quantile 3	-0.583440	-0.338359
All	-0.369100	-0.33959	-0.571450	≥ quantile 4	-0.784477	-0.19360

On the contrary, in Birmingham, the negative correlation between congestion and cycle lanes accessibility increase along with the increase of traffic flow. Especially in the most congested areas (top 1 Quantiles), the correlation coefficient is -0.78 (Table 5). Therefore, it is rather safe to claim that higher cycle lanes accessibility could mitigate the congestion, especially in the most congested area. Therefore, in order to mitigate the vehicle congestion level, the City Council of Birmingham should consider to expand the cycle lanes network rather than vehicle lanes network. As for Bordeaux, no acceptable correlation is observed from each congestion scenario, although the improvement during the middle range might implicate the influence of extreme values on the result.

To learn more on the city-centre effect, similar analysis is implemented on the city-centre region only for Zurich. A better correlation than the original results in Table 4 consolidate our initial assumption for the difference of city centre and suburb, and thereby support the selection for x per vehicle flow indicators.

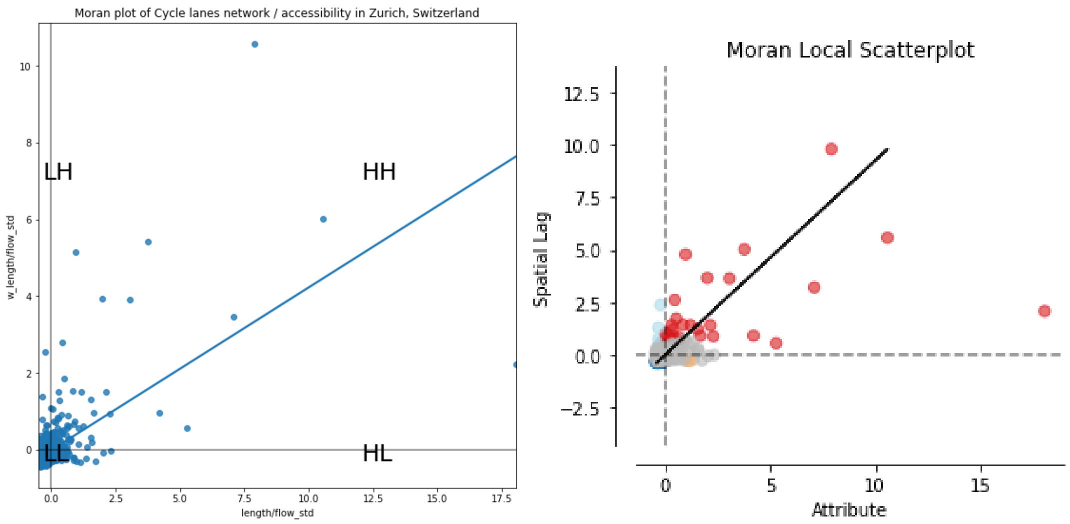


Figure 8 Global Moran plot (I=0.42, p_sim=0.001, left) and Local Moran plot(right)

In a future step, we will explore the global and local autocorrelation of the cycle lanes accessibility (length/flow). Figure 8 provides the scatterplot for to show the spatial autocorrelation with a significant level of 0.001 for global Moran and 0.05 for LISA. Figure 9 showcases the results from LISA analysis for Zurich and Manchester. The red regions of LL indicate a low cycle accessibility cluster, and on the contrary, the green regions of HH represents a high cycle accessibility cluster. Both yellow region of LH and blue region of HL shows a unordinary low accessibility compared to their neighbours, but this could also be due to the randomization effect of hexagon grid. It is suggested that the policymakers should attach more attention on the LL regions and benchmark those with the HH regions to find out the key success factors to improve the related cycle infrastructure.

Figure 9 The choropleth map for LISA result of cycle lanes accessibility in Zurich, Switzerland (left) and Manchester, UK (right)

Besides a single variable spatial autocorrelation, Bivariate Moran Statistics is implemented to stimulate the discussion and reach out a more robust result. Similarly, Figure 10 show the statistical distribution between the traffic flow and average value of cycle lanes length of one's neighbours. Both red regions of HL and yellow regions of HH requires more attention to construct a new or expand the existing cycle network. Related to the finding from the single variable Moran, some regions are categorized as red in both analysis, which implicate a more reliable conclusion that those regions require the most urgent policy attention on relevant infrastructure investment, as well as for the yellow regions. Some red regions that flip up to blue implicates that those regions have far less travel demand so that it is not necessary to expand the existing cycle network.

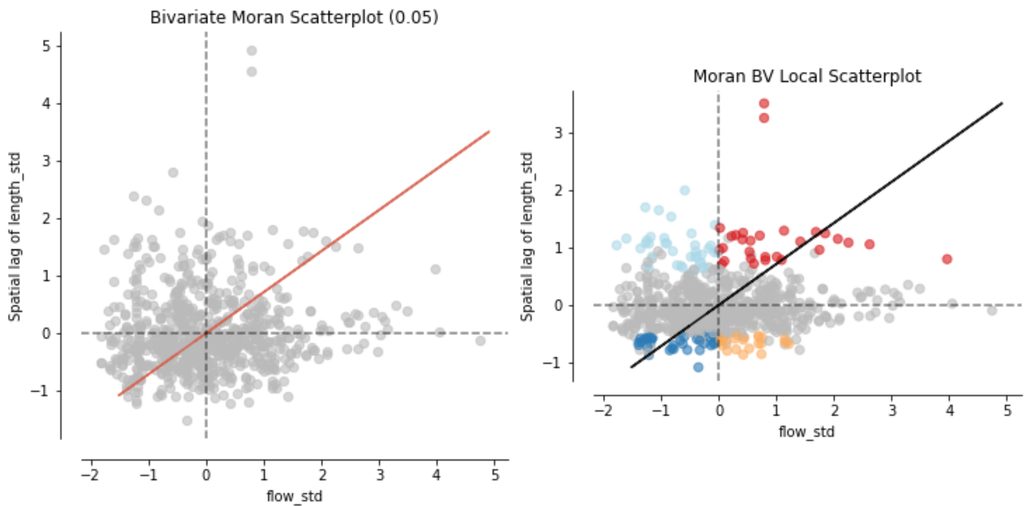


Figure 10 Global Moran plot (I=0.50, p_sim=0.001, left) and Local Moran plot(right)

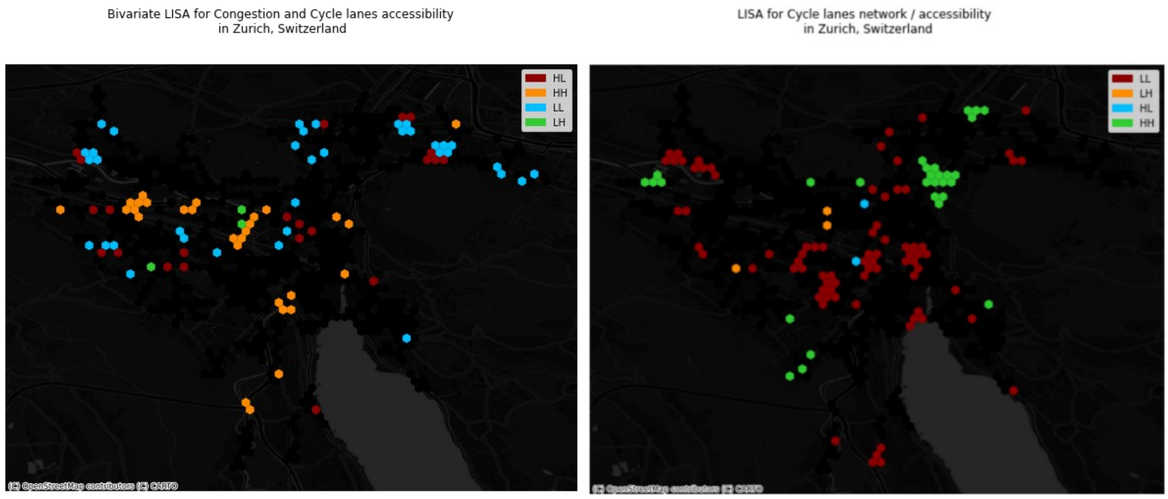


Figure 11 The choropleth map for bivariate LISA result of traffic flow and cycle lanes length (left) and single variable LISA result for comparison in Zurich, Switzerland

In-depth analysis with Lime e-scooter data in Zurich

In general, no clear relations where found between the location of the scooters and the other indicators (traffic, bicycle parking & cycling paths). However, some analyses show useful results. Below we will explain per relation that has been analysed what the results.

Lime scooters versus traffic congestion

The analysis conducted on the relation between the location of the Lime scooters and the traffic congestion has been conducted on different sizes of hexagons (radius of approx. 5 m, 10 m and 50 m). Table 6 shows the correlation between the lime scooters and traffic congestion for each of the hexagon. The two hexagon grids with the larger hexagons, show a very small positive relation between the number of cars and the number of Lime scooters. The finest hexagon grid shows a negative relation between the number of cars and the number of lime scooters. A scatterplot is included for the finest hexagon grid between the number of scooters and traffic congestion in Figure 12.

Table 6 Correlation between location of lime scooters and traffic congestion for different hexagon grid values

Hexagon grid radius [m]	50	10	5
Correlation	0.00894	0.05529	-0.085619

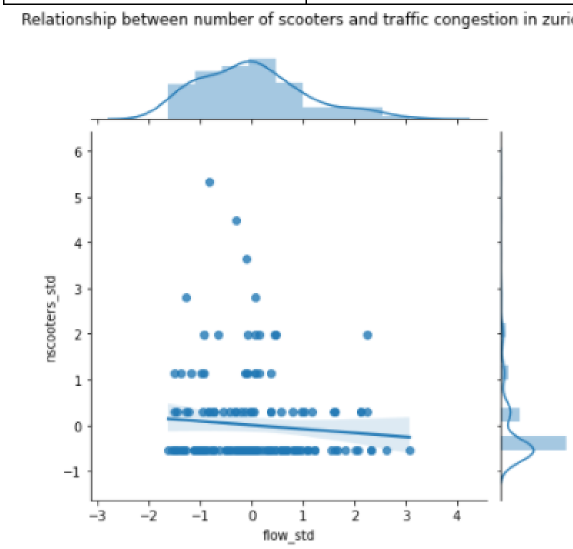


Figure 12 Pairplot between number of scooters and traffic congestion in Zurich hexagon grid with 5 m radius

Lime scooters versus Bicycle parking

When taking the hexagon grid value of approx. 10 m radius, we see a small positive correlation (0.23404) between the lime scooter locations and the bicycle parking. Figure 13 shows the pair-plot between the number of scooters and the bicycle parking in Zurich for a hexagon grid of 10 m radius. Large outliers can be seen for the bicycle parking at a value of 2.

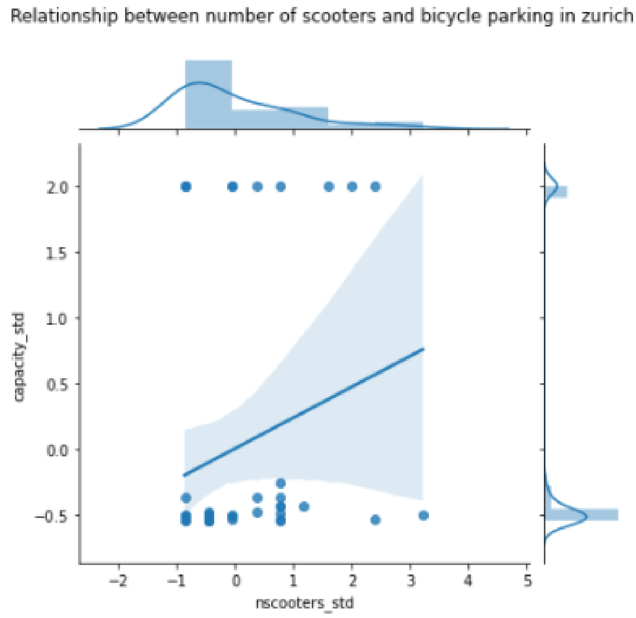


Figure 13 Pair-plot between number of scooters and bicycle parking in Zurich

Lime scooters vs Cycle lanes length

For the hexagon grid with a radius of approx. 5 m, a correlation of 0.15864 is seen between the lime scooters and cycle lanes length. This is a slight positive relation which can possibly be caused by the likeliness that people prefer using a scooter (and cycling) close to or on cycling paths. Figure 14 shows the pair-plot between the number of scooters and cycle lanes in Zurich. The cycle length is skewed to the left.

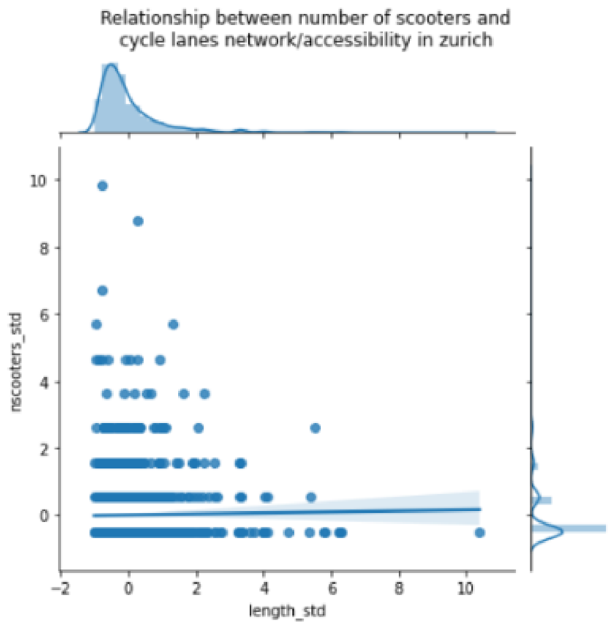


Figure 14 Pair-plot between number of scooters and cycle lanes in Zurich

Discussion & Conclusion

The findings of this research showed that on an inter-city level, no relevant correlations were found between the CO2 emissions, cycle network shares, congestion and bicycle. When zooming in on Zurich, Birmingham, Bordeaux, Frankfurt, Manchester & Toulouse, a city level analysis showed that also no significant correlation between traffic congestion and bicycle parking capacity was found. When looking at traffic congestion versus cycle lane lengths, for Zurich, Manchester, and Toulouse it could be inferred that less congestion is the causality of high cycle lanes accessibility. And in Birmingham, it concludes that higher cycle lanes accessibility could mitigate the congestion, especially in the most congested area. Therefore, a rather counter-intuitive suggestion -- expand the cycle lanes network rather than vehicle road network—should be considered in order to reduce the congestion level. Combining with the single-variable and bivariate Moran statistics, the least cycle-accessible regions are identified to suggest further policy attention on the related infrastructure investment. Regarding the Lime e-scooters, no clear relations were found between the location of the scooters and traffic flow, bicycle parking & cycling paths. However, it can be stated that people prefer to use Lime scooters near bicycle paths, indicating that the total length of bicycle paths do have an effect on the use of Lime scooters.

In this research possible relations have been investigated in order to be able to write a policy advise for improving the sustainability of a city. The gap that is tackled is that little is known yet on the effect of upcoming shared e-scooters in cities. The improvement of sustainability can be done by reducing congestion in the city, which leads to less CO2 emissions. Possible solutions for congestion reducing travel modes are cycling and Lime e-scooters. It has been shown that an improved bicycle network can have a desired effect on the use of mobility modes, leading to less congestion and thus a more sustainable city. The policy advice that flows from this research will thus be: Improvement of the bicycle network.

When conducting this research, multiple limitations were faced. These limitations are, amongst others, related to data availability, data bias, time consuming data retrieval, assumptions made and the predefined scope of the research. Based on these limitations, suggestions for further research are provided.

One of the issues we faced was the data availability and the data bias of the datasets. The UTD19 dataset shows the traffic flow for only a selected number of roads and cities. Further research might look at more cities and include more roads, if data is available. The data we have used is the yearly data for different years within a time frame of only several years. It is assumed that within this time frame these characteristics are not changed significantly.

The urban typologies dataset has a lot of missing values, which limited the number of indicators that are used. It is not known how all the data is retrieved and therefore there might be a bias in the data. For example the way the mode shares data is retrieved in each city might differ, which leads to inaccurate comparisons.

The retrieval of the lime data is very time consuming. Because of our limited time frame, it was only possible to retrieve this data for Zurich. The use of lime data can be debatable as the company can set areas were scooters can be placed and where not. It will also put more scooters in busy areas and not in areas where they are hardly used; it is debatable whether this also shows the availability for bicycle use.

In the congestion, we see a 'city-centre effect': in the city centre more people and more general amenities are presented which leads to busier roads and higher number of scooters, hence more congestion and more flow in the city centre compared to other areas. Further research could implement a more sophisticated method that relates the vehicle accumulation and network capacity (e.g. Macroscopic Fundamental Diagram) to represent the congestion level, which will give a more holistic view compared to the traffic flow only (Loder, Ambühl, Menendez, & Axhausen, 2019b).

References

Akerboom, S., Botzen, W., Buijze, A., Michels, A., & Rijswick, M. v. (2020). Meeting goals of sustainability policy: CO2 emission reduction, cost-effectiveness and societal acceptance. An analysis of the proposal to phase-out coal in the Netherlands. *Energy Policy, Volume 138*.

Bettencourt, L., Lobo, J., Strumsky, D., & West, G. (2010). Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. *PLOS ONE 5(11): e13541*.

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks.

Grote, M., Williams, I., Preston, J., & Kemp, S. (2016). Including congestion effects in urban road traffic CO2 emissions modelling: Do Local Government Authorities have the right options? *Transportation Research Part D: Transport and Environment*, 95-106.

Loder, A., Ambühl, L., Menendez, M., & Axhausen, K. W. (2019). Understanding traffic capacity of urban networks. *Scientific reports, 9(1), 1-10.*

Moudon, A. V., Lee, C., Cheadle, A. D., Collier, C. W., Johnson, D., Schmid, T. L., & Weather, R. D. (2005). Cycling and the built environment, a US perspective. *Transportation Research Part D: Transport and Environment, Volume 10, Issue 3*, 245-261.

Oke, J. B., Aboutaleb, Y. M., Akkinepally, A. A., C. L., H., Y., Z., & P. C., .. &.-A. (2019). A novel global urban typology framework for sustainable mobility futures. *Environmental Research Letters, 14(9), 095006.*

Schellong, D., Sadek, P., Schaetzberger, C., & Barrack, T. (2019). *The Promise and Pitfalls of E-Scooter Sharing.* Boston Consulting Group.

Transportation Research Board. (1992). *Transportation research record No. 1372 Operations and Safety: Non motorized Transport.* Washington, D.C.: National Academy Press.

Wang, M., & Zhou, X. (2017). Bike-sharing systems and congestion: Evidence from US cities. *Journal of Transport Geography,*, 147-154.

Appendix A Lime scooter data retrieval process

Briefly formulated, in the written code we pretend to be an application user. The map in the application shows the location of the scooters in the surrounding. Only if you zoom in far enough, it is possible to see these locations rather than a cluster of scooters. Also, the application only provides you with the location of 50 scooters. If the area you want to get the scooter location has more than 50, it will provide you with 50 scooters that are located the closest towards the centre of the area on which you want the scooter data. This is what makes the code rather complex, but this is not what causes the high effort required to retrieve the data after creating the pipeline. This is caused by the way the data is retrieved.

As mentioned, the code is written in a way that it looks as if data is being requested by the application. This requires there to be a user with an account. Using a personal lime account and the belonging phone number, we first have to login and verify the account. To do this, we send my phone number to a server which returns a verification code to the phone number. Then we have to send this verification code back. In addition, we send the locations back on which we want to get a json back with the scooter locations. This includes a 'personal' location, which is not our real location of course but the location that we want the lime server to think that we are at.

The complex part is that the server only provides us with the location of 50 scooters. Which is not a fair distribution over the requested area. This means that for one city, different smaller areas have to be requested on the location of the scooters in order to retrieve the location of all scooters in the city. In order to automize this, a for-loop is created that creates a grid of squared polygons from the assigned area. Then, using a for-loop and a function, the scooter data is retrieved individually for every created squared polygon in the grid. The scooter data is than combined in one data frame and the scooter locations that have been taken into account more than once are removed.

The coding and the structure behind this process results in a solid pipeline for scooter data. You only need to login once and you can repeat the process limitless times. Resulting in a large dataset. However, there is one problem that is caused by the 'backdoor' aspect of this method. Since a grid is required also a lot of data is requested in a lot of separate requests. In order to gain a realistic distribution over the city, a fine grid is required due to the limitation of only retrieving 50 scooters per retrieval request.

In order to retrieve better, more and more detailed data we created a grid for the centre and for the overall city which have different patterns. One having a pattern of 5 by 6 and one having a pattern of 7 by 6. Making sure that the centres of the grid squares are not located at the same location. Maximizing the number of different scooter locations.

Together with the maximal level of detail for the grids, which actually is 5 by 6 and 7 by 6 for respectively the city centre and the whole city, it is possible to retrieve proper and well represented data for in this case Zurich at one point in time. However, for every new data retrieval we need to login again. Also doing this a lot results in Lime blocking the used phone number. Eventually after repeating the process for a day we managed to retrieve over 4000 lime scooter locations in Zurich.

In conclusion, even though the pipeline is solid, it is very time consuming to retrieve lime data using this 'backdoor' method. Therefore, the Lime data has only been retrieved for Zurich. With this method, we had to use many different data science aspects which we learned during the course. In addition, we sought for an alternative method to retrieve data. We think we managed to do so. In order to show you the process and to better understand what we did, since it is a non-regular method, we have added the code in this appendix. Only for the retrieval and not for the preparation for the analysis.

Appendix B: Exploratory data analysis city level

Exploratory data analysis Birmingham, United Kingdom

Figure 15 Traffic flow in Birmingham, United Kingdom

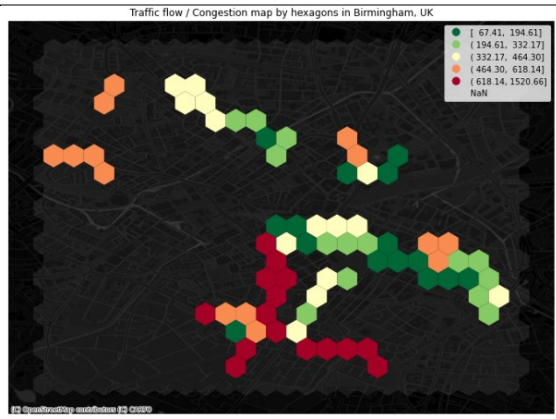


Figure 16 Hexagon map of traffic flow in Birmingham, United Kingdom

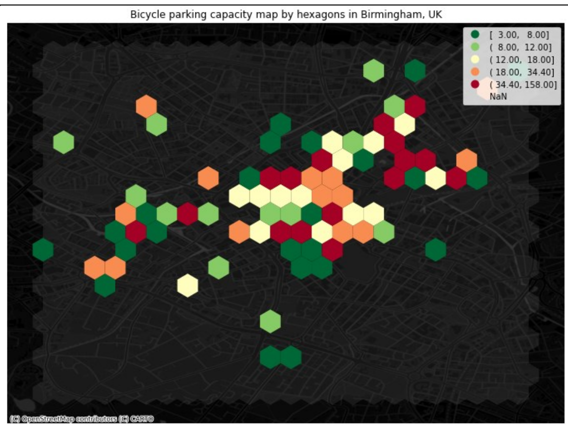
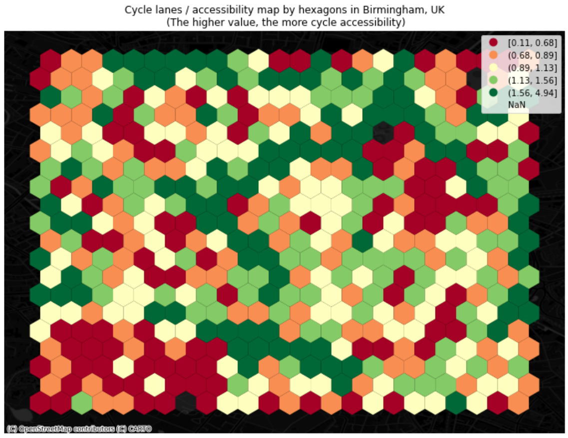
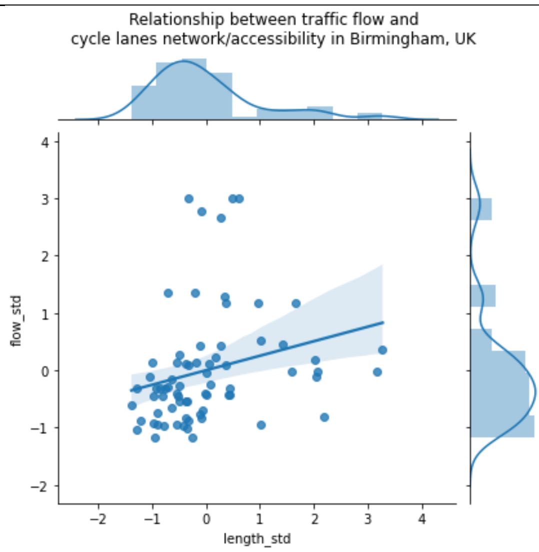
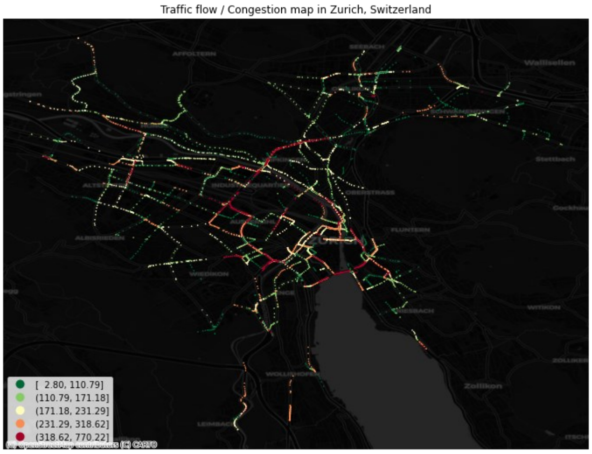


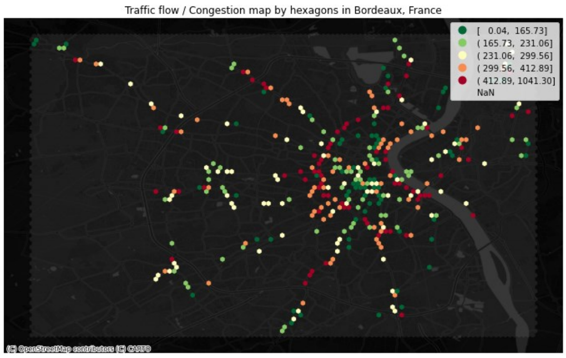
Figure 17 Hexagon map of cycle lanes / accessibility in Birmingham, United Kingdom

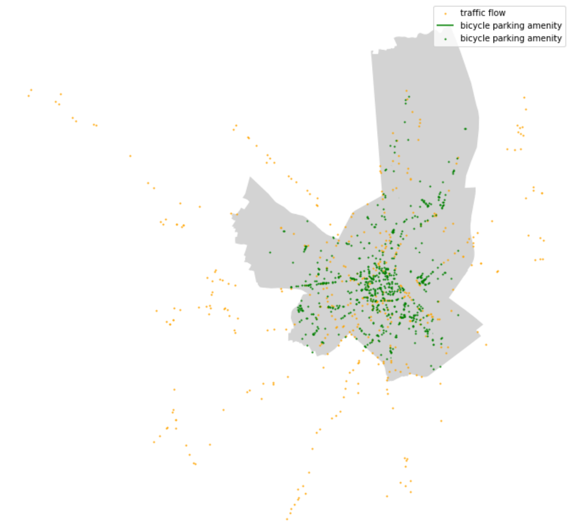


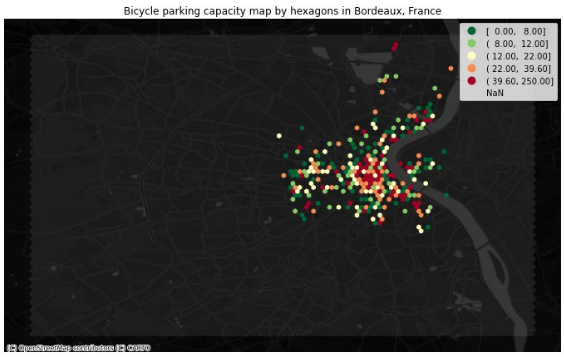


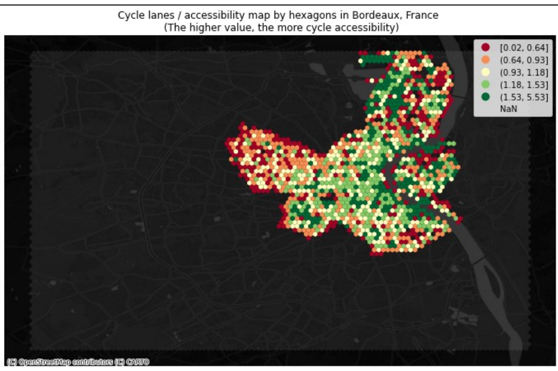
Exploratory data analysis Bordeaux, France

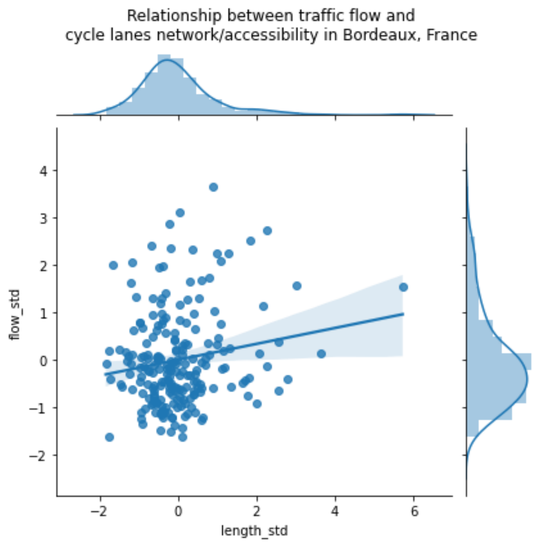




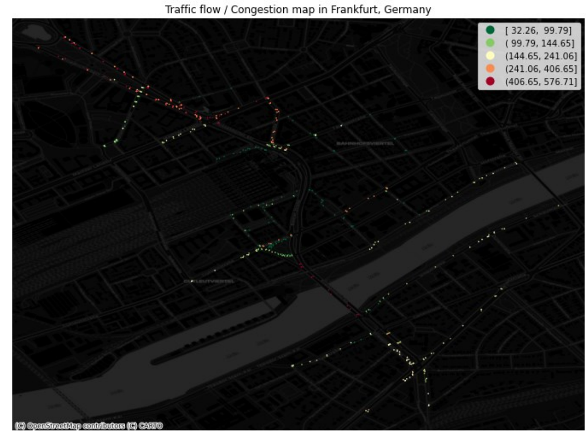


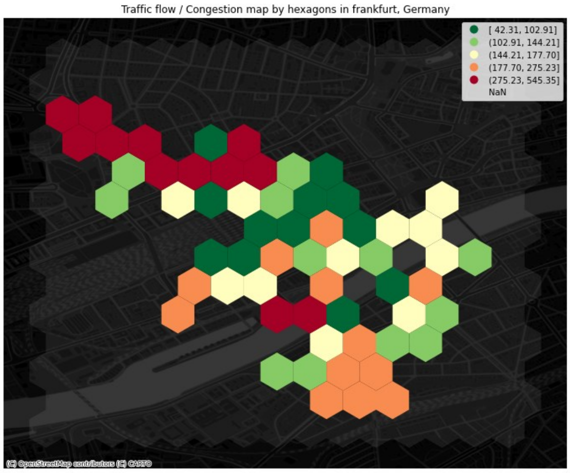


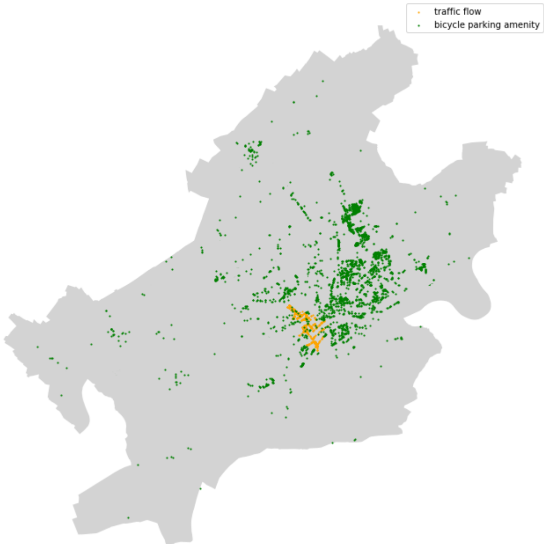


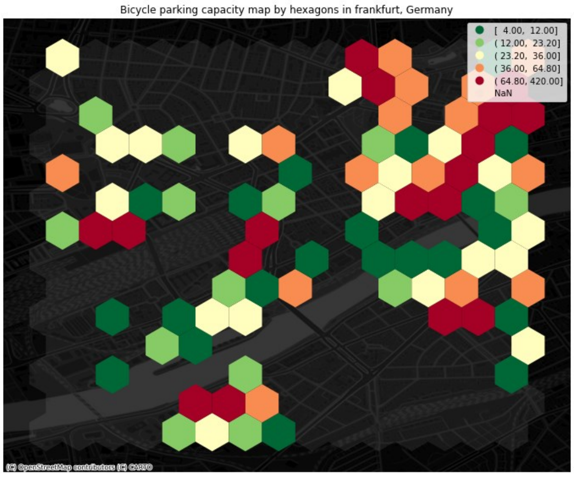


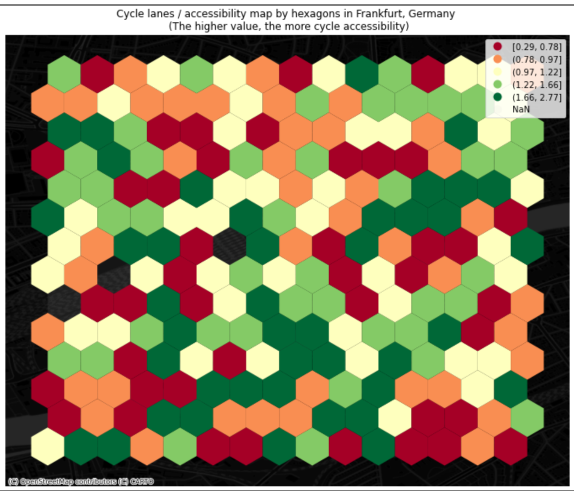
Exploratory data analysis Frankfurt, Germany

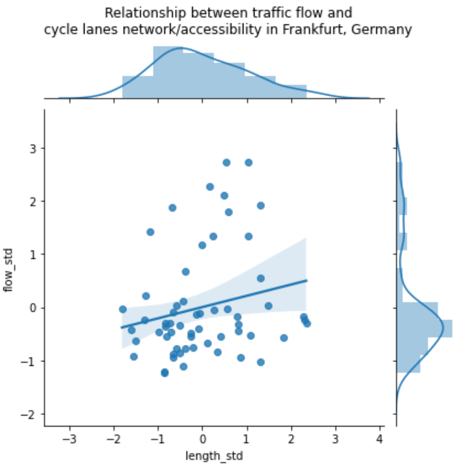








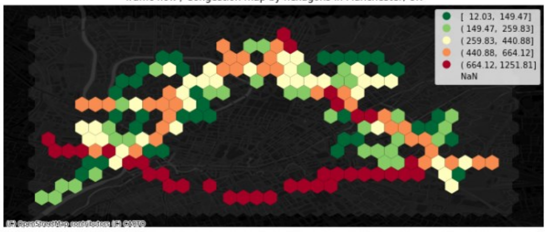




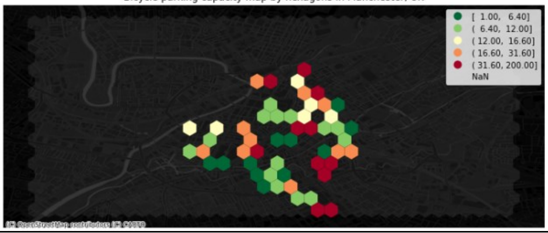
Exploratory data analysis for Manchester, United Kingdom

Traffic flow / Congestion map in Manchester, United Kingdom

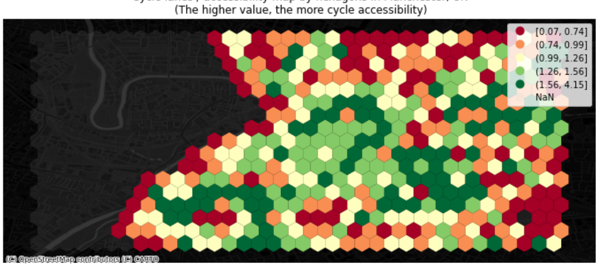
Traffic flow / Congestion map by hexagons in Manchester, UK



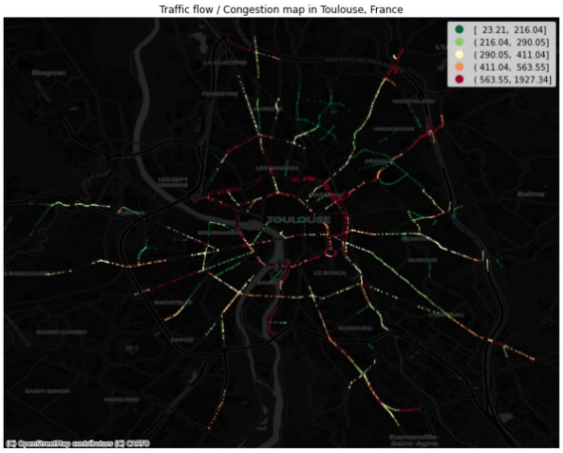
Bicycle parking capacity map by hexagons in Manchester, UK



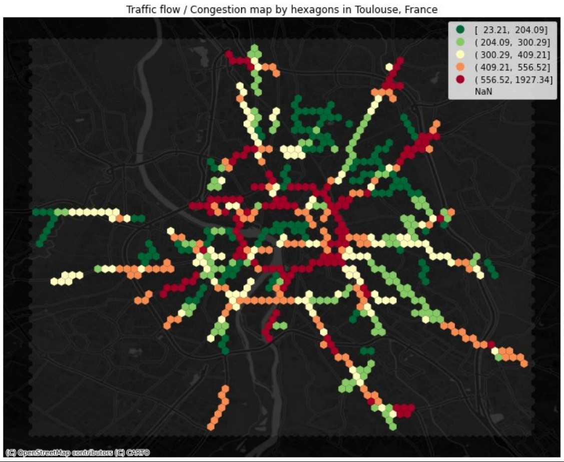
Cycle lanes / accessibility map by hexagons in Manchester, UK
(The higher value, the more cycle accessibility)



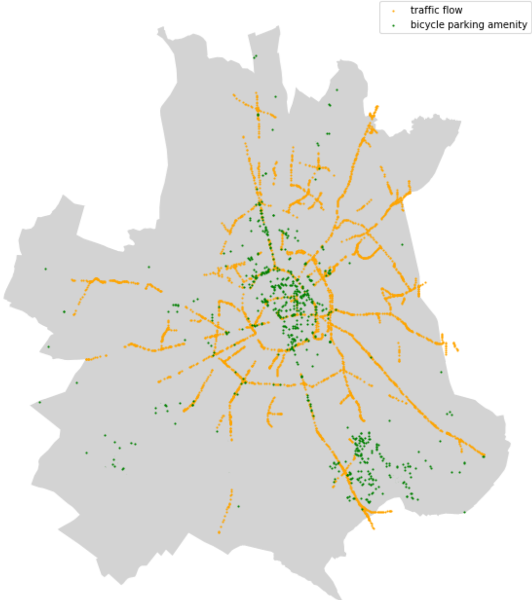
Exploratory data analysis for Toulouse, France

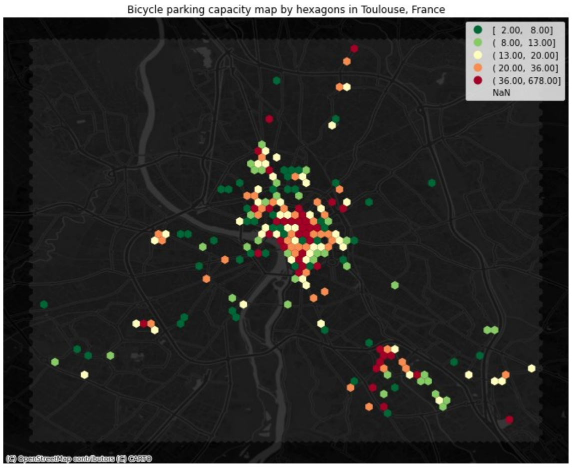


Traffic flow / Congestion map in Toulouse, France

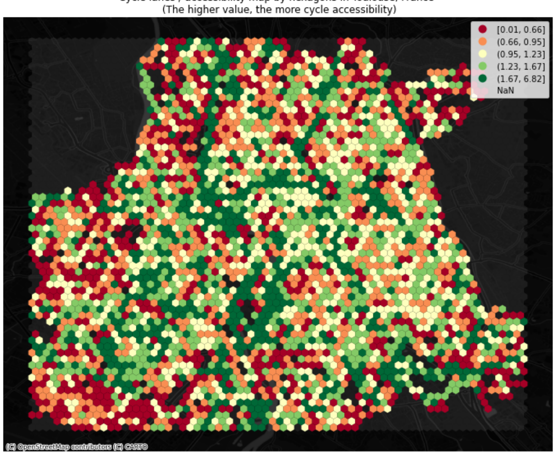


Traffic flow / Congestion map by hexagons in Toulouse, France





Bicycle parking capacity map by hexagons in Toulouse, France



Cycle lanes / accessibility map by hexagons in Toulouse, France
(The higher value, the more cycle accessibility)

<u>Exploratory data analysis Zurich, Switzerland</u>

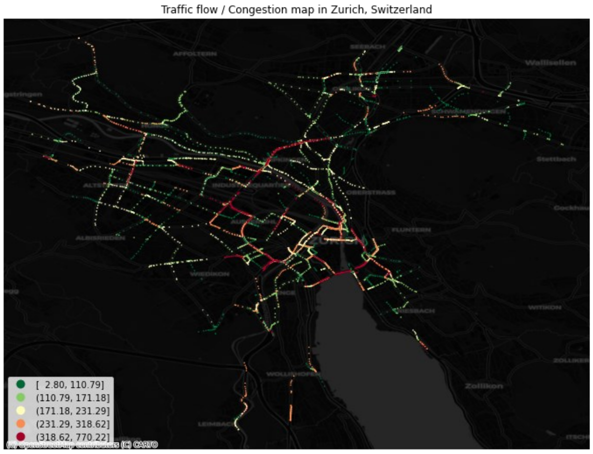


Figure 18 Traffic flow in Zurich, Switzerland

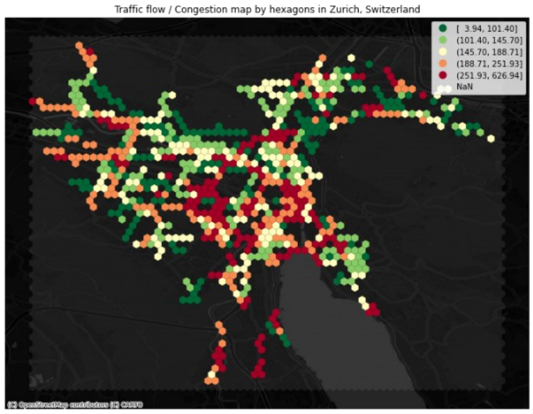


Figure 19 Hexagon map of traffic flow in Zurich, Switzerland

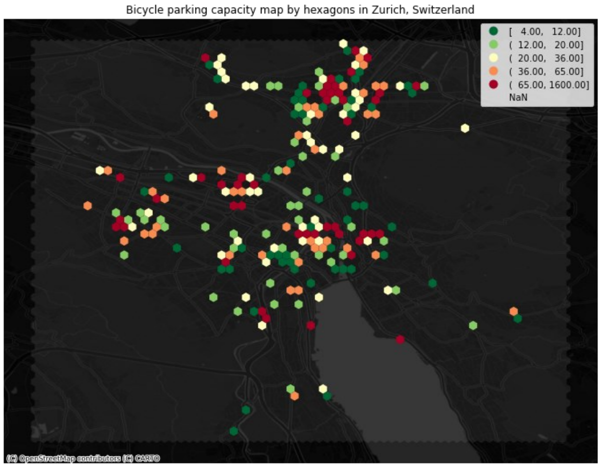


Figure 20 Hexagon map of bicycle parking capacity in Zurich, Switzerland

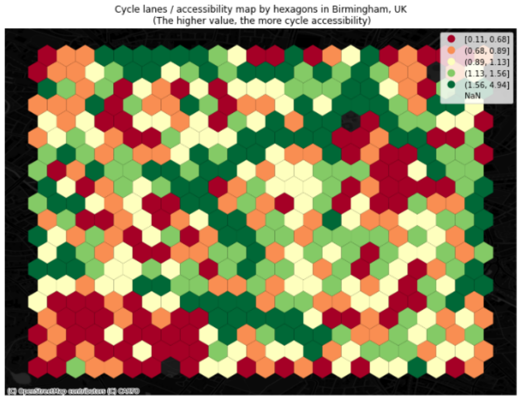


Figure 21 Hexagon map of cycle lanes/ accessibility in Birmingham, United Kingdom

Appendix C: Inter-city analysis

The relations between indicators are visualized in figure(??), which shows a pairplot for the relations between the chosen indicators as described in the section on the intercity analysis. As can be seen in table (??), in which the correlation coefficients between the different indicators is shown no significant correlation was found.

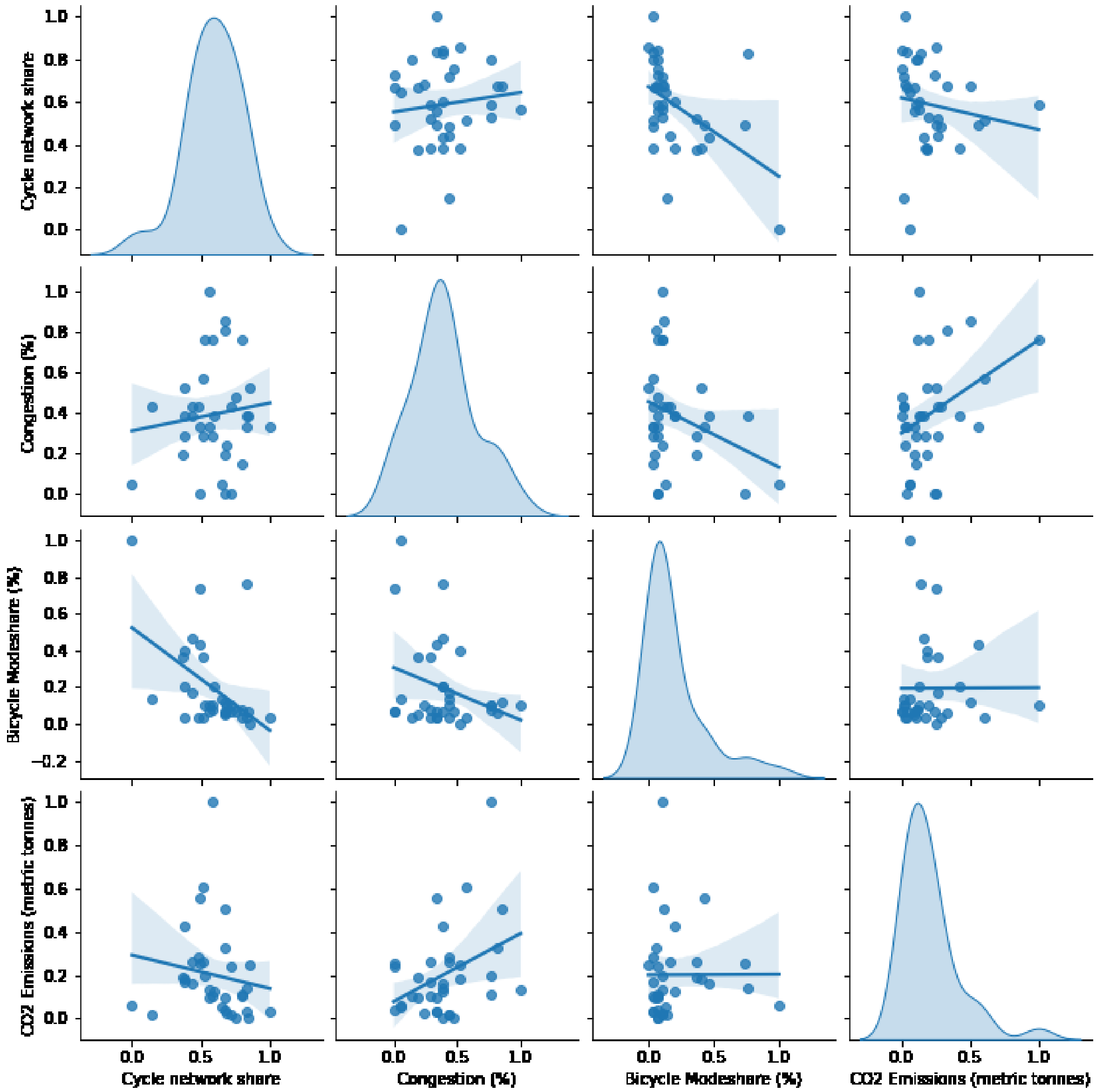


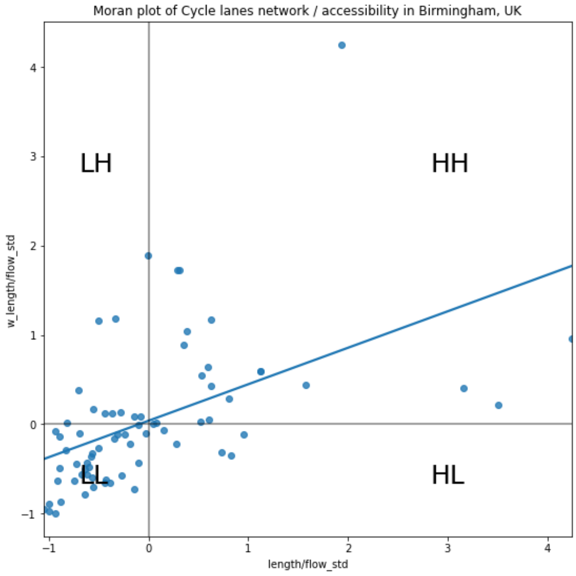
Figure 22 Pair-plot of inter-city characteristics

Table 7 Correlation coefficients for inter-city analysis

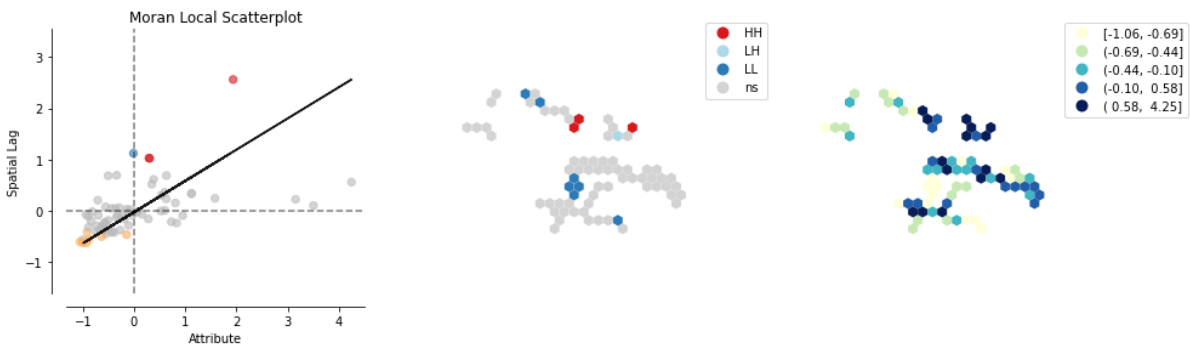
	Cycle network share	Congestion (%)	Bicycle Modeshare (%)	CO2 Emissions (metric tonnes)
Cycle network share	1.000000	0.112131	-0.484981	-0.151493
Congestion (%)	0.112131	1.000000	-0.304339	0.380610
Bicycle Modeshare (%)	-0.484981	-0.304339	1.000000	0.003290
CO2 Emissions (metric tonnes)	-0.151493	0.380610	0.003290	1.000000

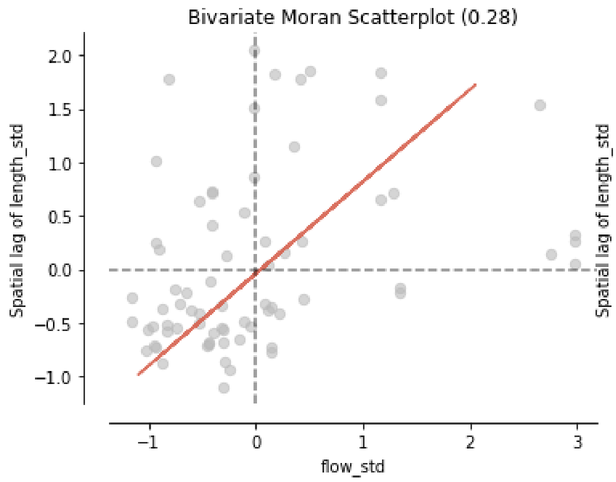
Appendix D: City level results for Birmingham, Bordeaux, Frankfurt, Manchester, Toulouse

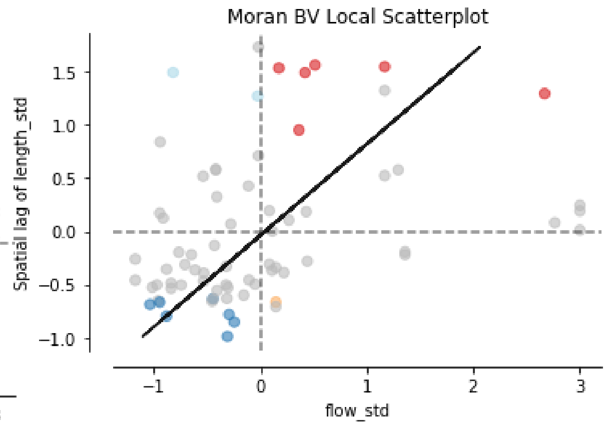
<u>Birmingham:</u>

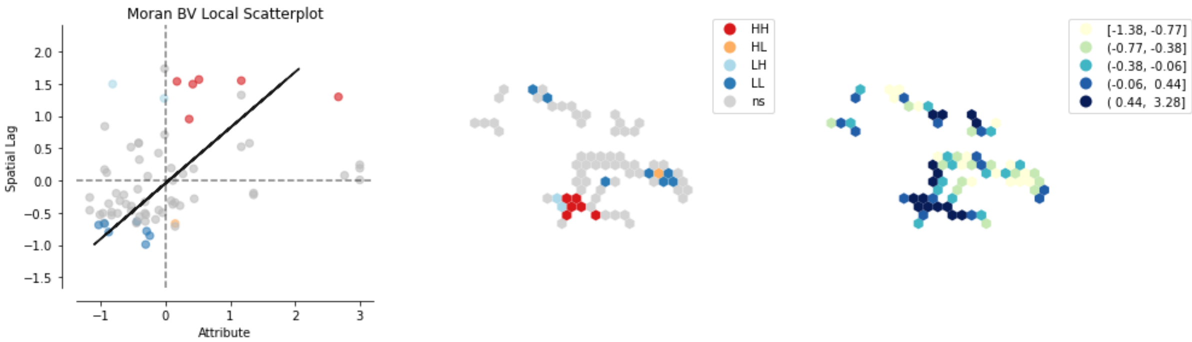


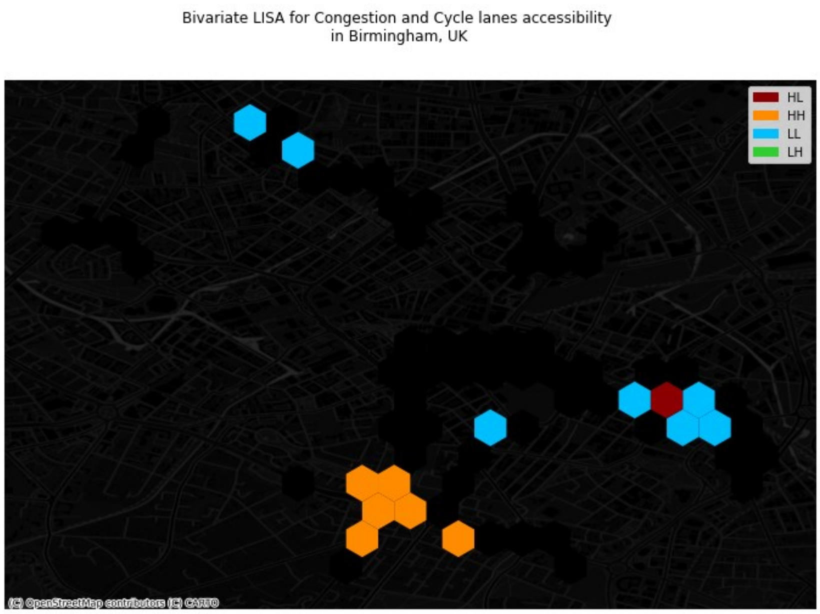
LISA for Cycle lanes network / accessibility
in Birmingham, UK



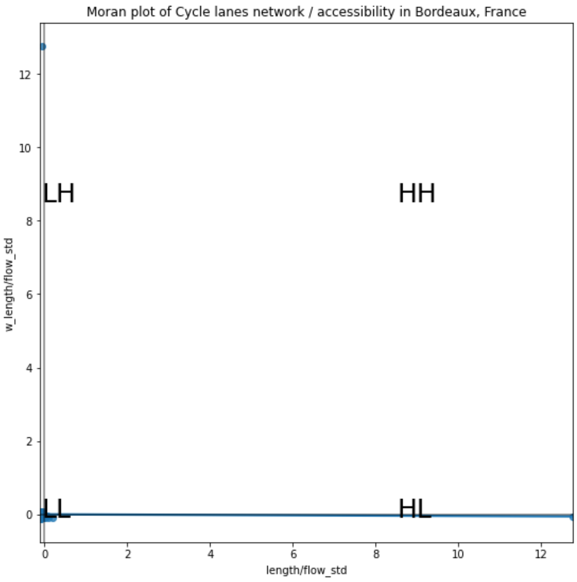

Bivariate Moran Scatterplot (0.28)
Spatial lag of length_std
flow_std


Moran BV Local Scatterplot
Spatial lag of length_std
flow_std

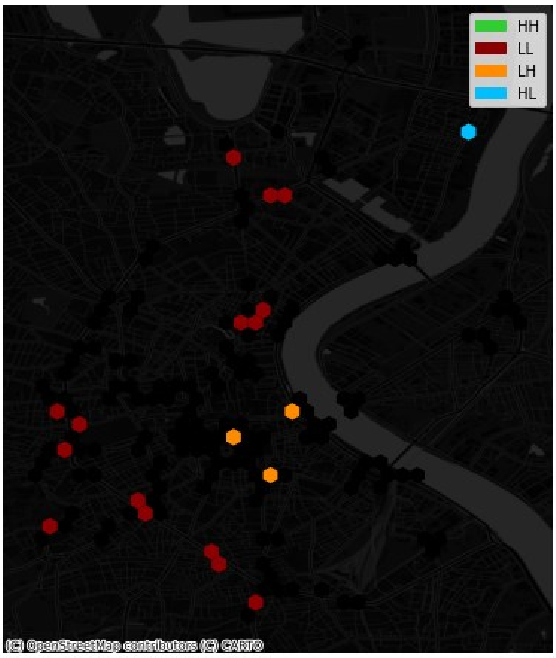

Moran BV Local Scatterplot
Spatial Lag
Attribute
HH
HL
LH
LL
ns
[-1.38, -0.77]
(-0.77, -0.38]
(-0.38, -0.06]
(-0.06, 0.44]
(0.44, 3.28]

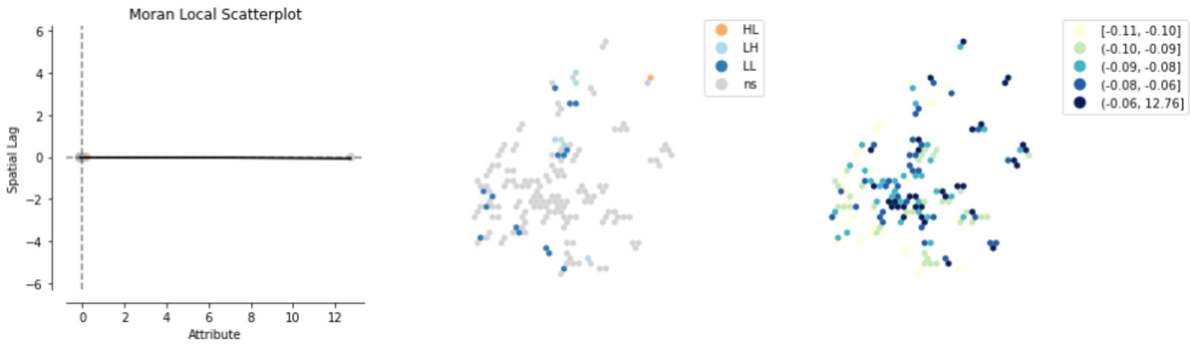

Bivariate LISA for Congestion and Cycle lanes accessibility
in Birmingham, UK
HL
HH
LL
LH
(C) OpenStreetMap contributors (C) CARTO

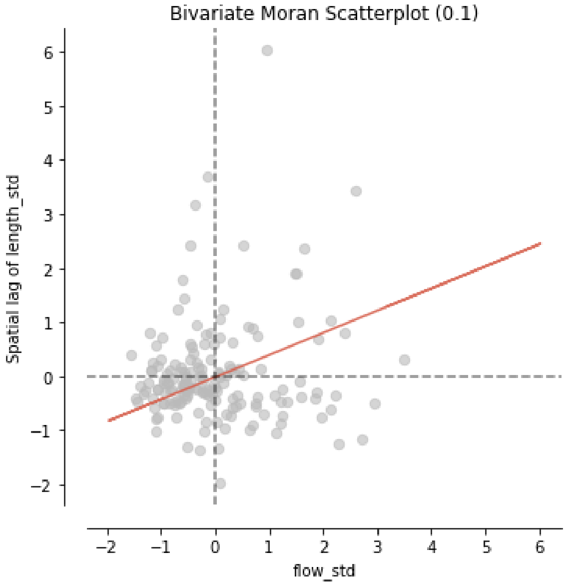
Bordeaux

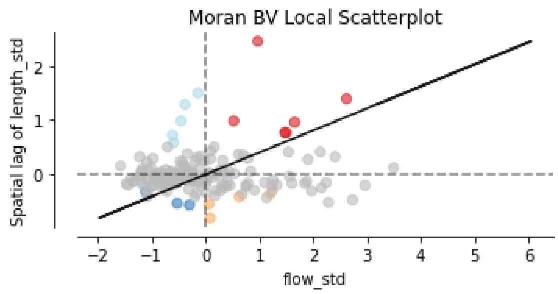


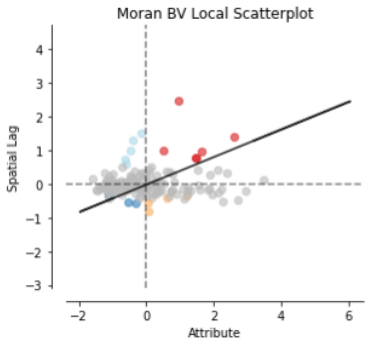
LISA for Cycle lanes network / accessibility
in Bordeaux, France

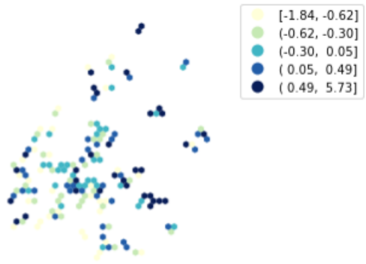




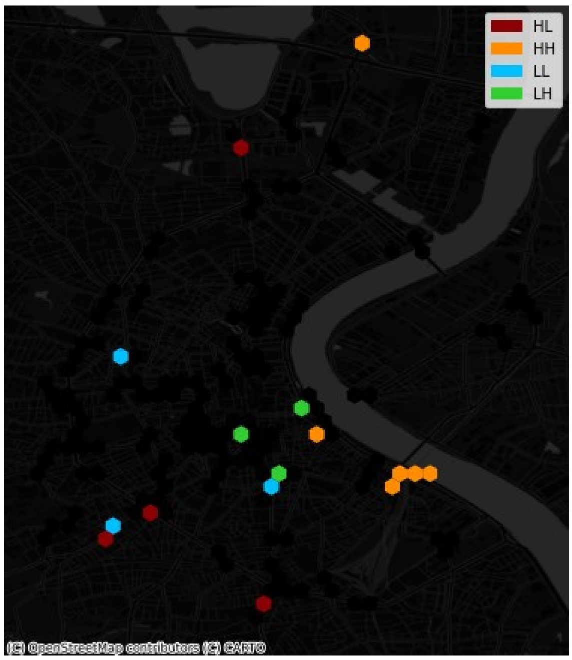




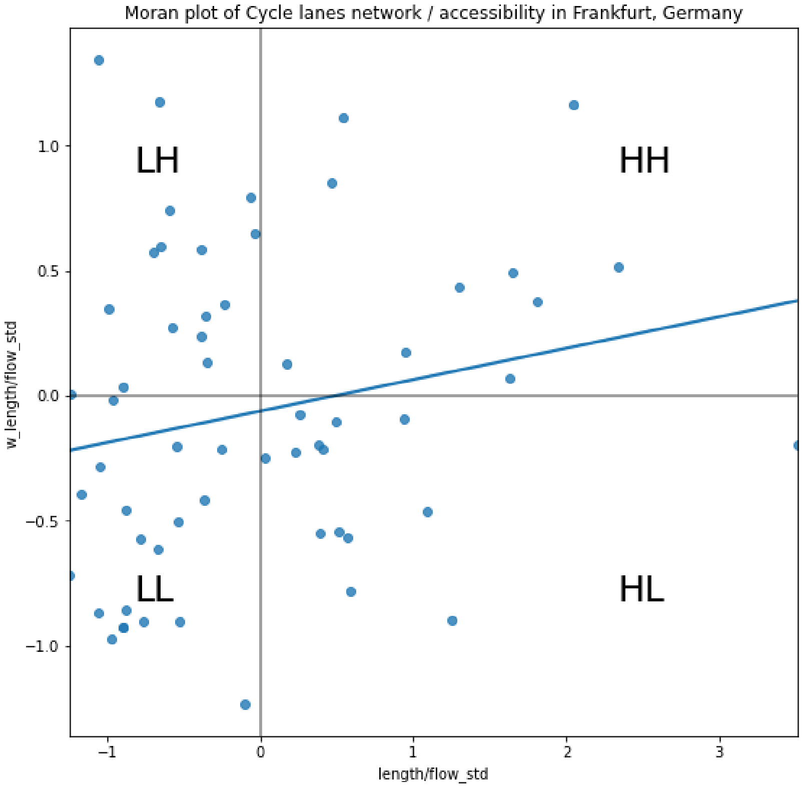


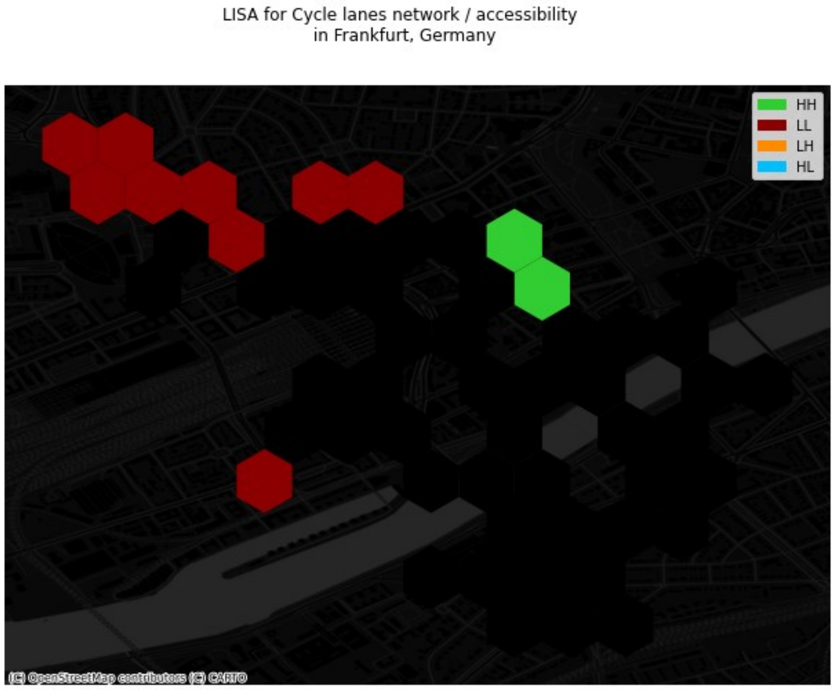


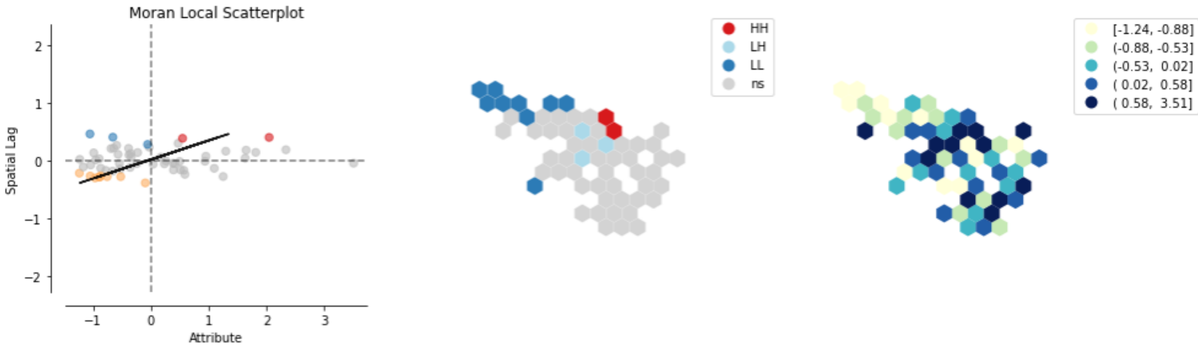
Bivariate LISA for Congestion and Cycle lanes accessibility
in Bordeaux, France

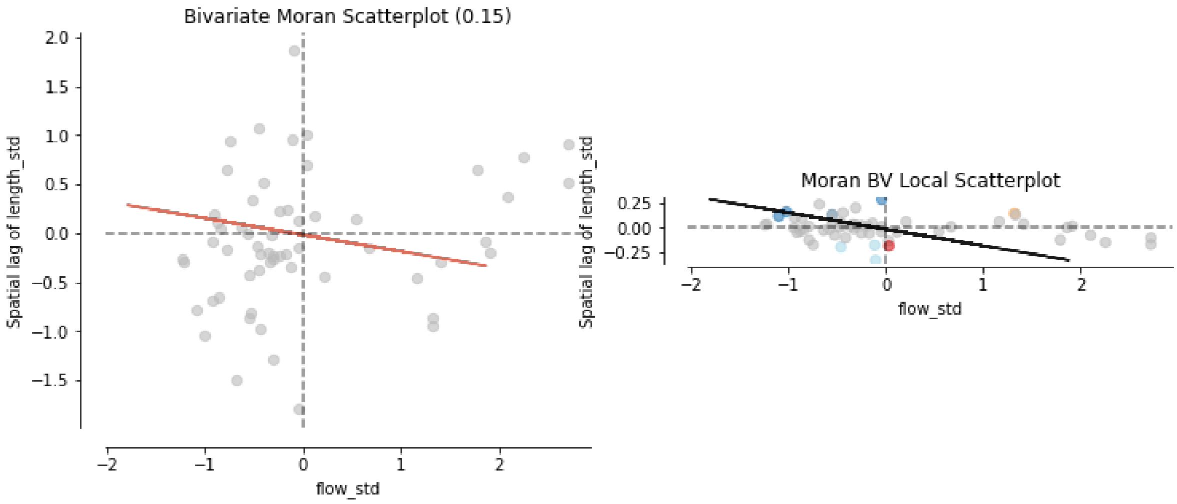


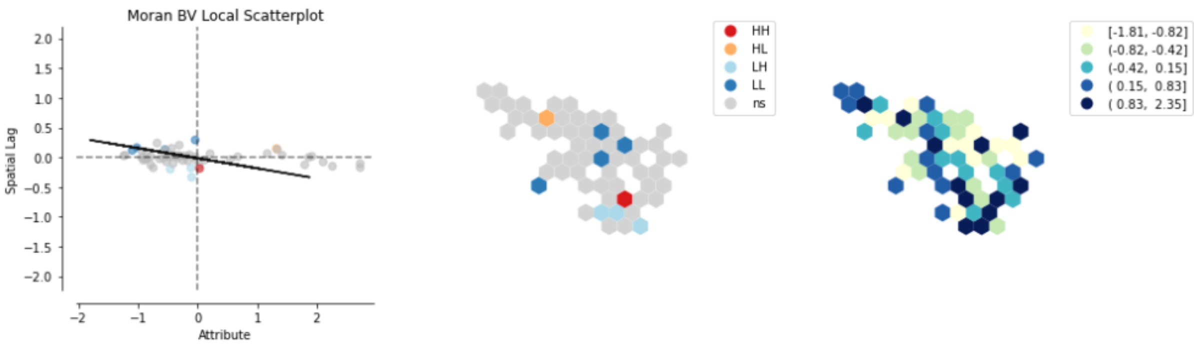
Frankfurt

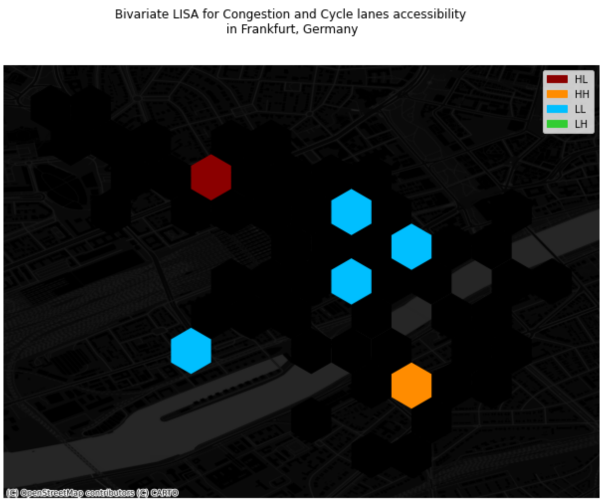

Moran plot of Cycle lanes network / accessibility in Frankfurt, Germany
LH
HH
LL
HL
w_length/flow_std
length/flow_std
1.0
0.5
0.0
−0.5
−1.0
−1
0
1
2
3


LISA for Cycle lanes network / accessibility
in Frankfurt, Germany
HH
LL
LH
HL
(C) OpenStreetMap contributors (C) CARTO

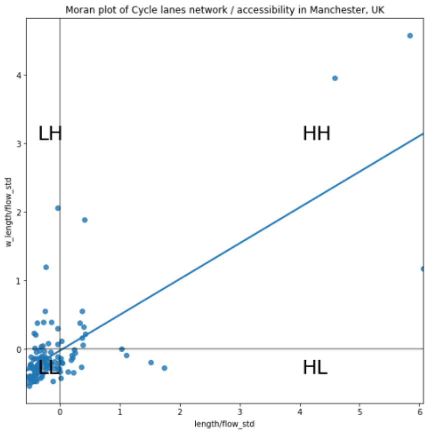

Moran Local Scatterplot
Spatial Lag
Attribute
HH
LH
LL
ns
[-1.24, -0.88]
(-0.88, -0.53]
(-0.53, 0.02]
(0.02, 0.58]
(0.58, 3.51]


Bivariate Moran Scatterplot (0.15)
Spatial lag of length_std
flow_std
Moran BV Local Scatterplot
Spatial lag of length_std
flow_std

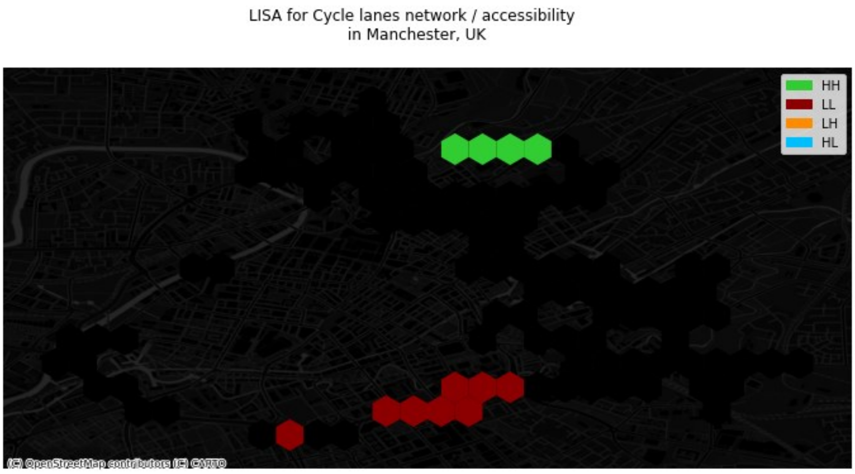

Moran BV Local Scatterplot
Spatial Lag
Attribute
HH
HL
LH
LL
ns
[-1.81, -0.82]
(-0.82, -0.42]
(-0.42, 0.15]
(0.15, 0.83]
(0.83, 2.35]


Bivariate LISA for Congestion and Cycle lanes accessibility
in Frankfurt, Germany
HL
HH
LL
LH
(C) OpenStreetMap contributors (C) CARTO

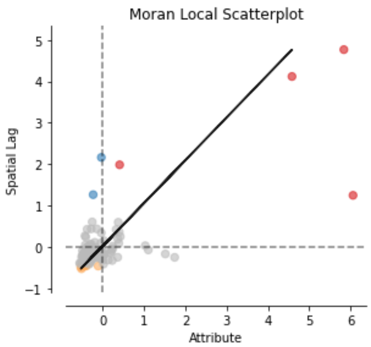
Manchester



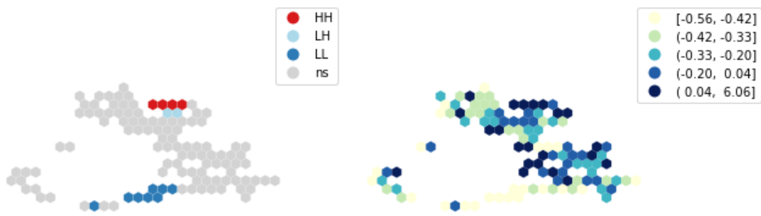
Moran plot of Cycle lanes network / accessibility in Manchester, UK
LH
HH
LL
HL
w_length/flow_std
length/flow_std



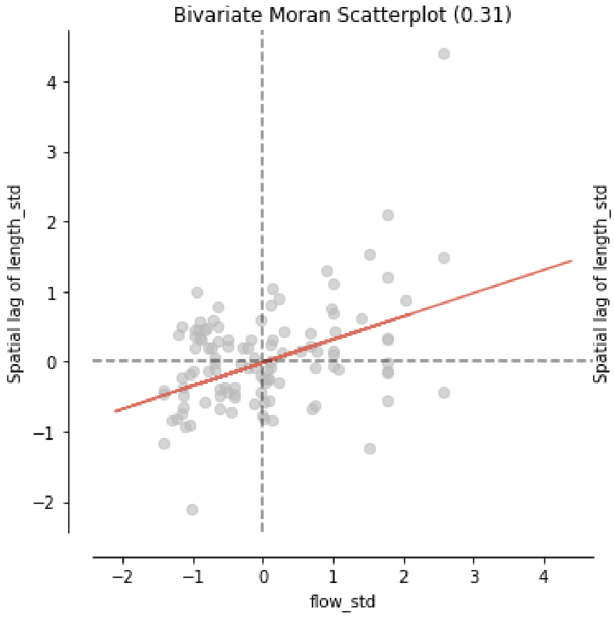
LISA for Cycle lanes network / accessibility in Manchester, UK
HH
LL
LH
HL
(C) OpenStreetMap contributors (C) CARTO

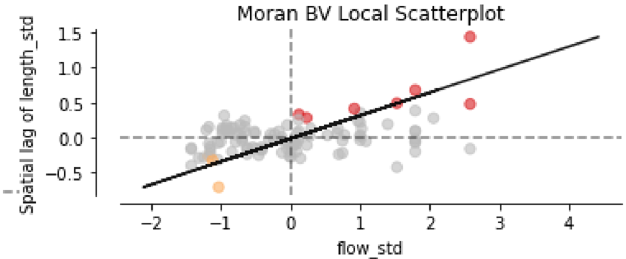


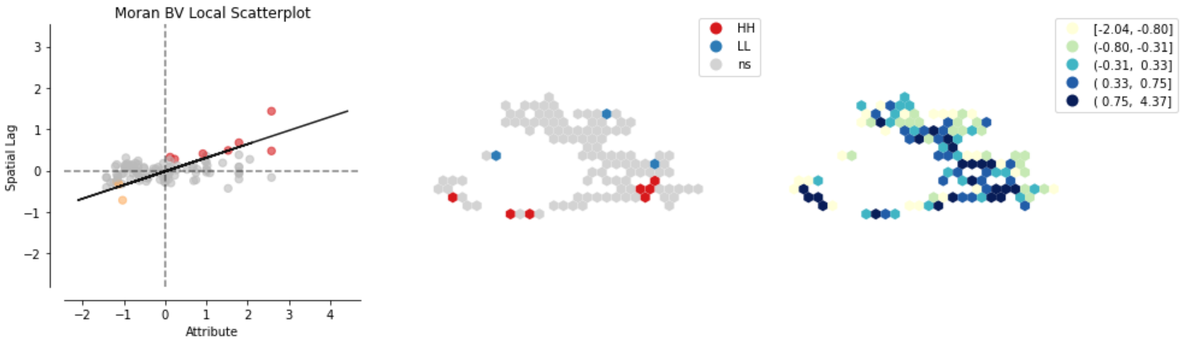
Moran Local Scatterplot
Spatial Lag
Attribute



HH
LH
LL
ns
[-0.56, -0.42]
(-0.42, -0.33]
(-0.33, -0.20]
(-0.20, 0.04]
(0.04, 6.06]

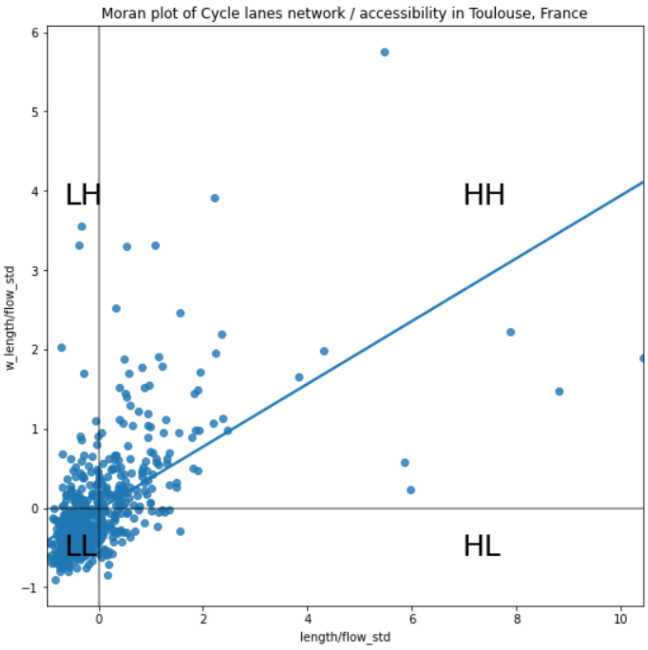


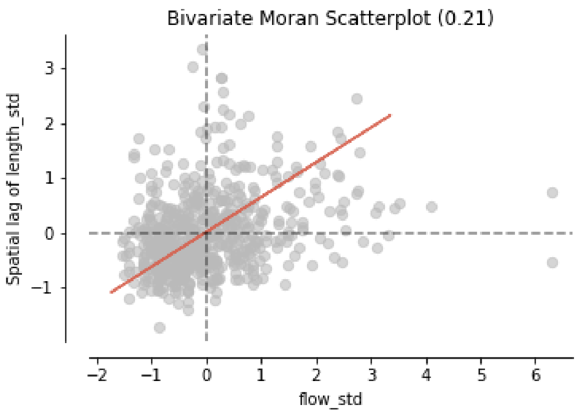


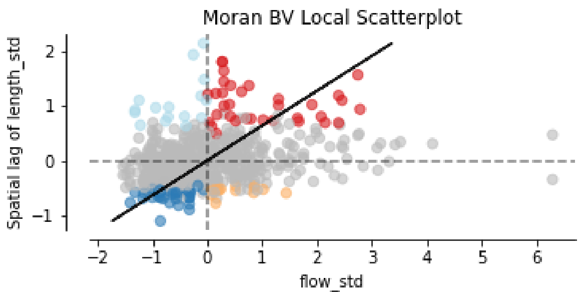


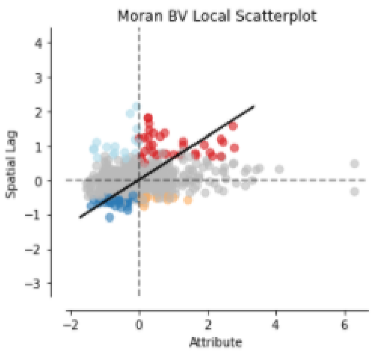
Bivariate LISA for Congestion and Cycle lanes accessibility
in Manchester, UK

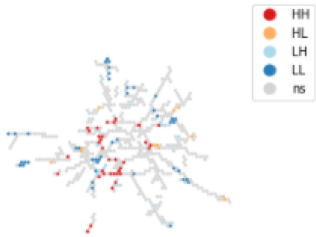
Toulouse

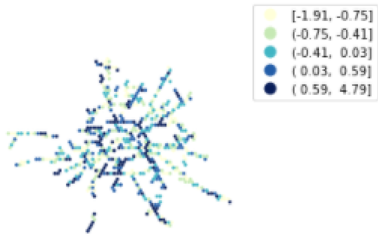

Moran plot of Cycle lanes network / accessibility in Toulouse, France
w_length/flow_std
length/flow_std
LH
HH
LL
HL

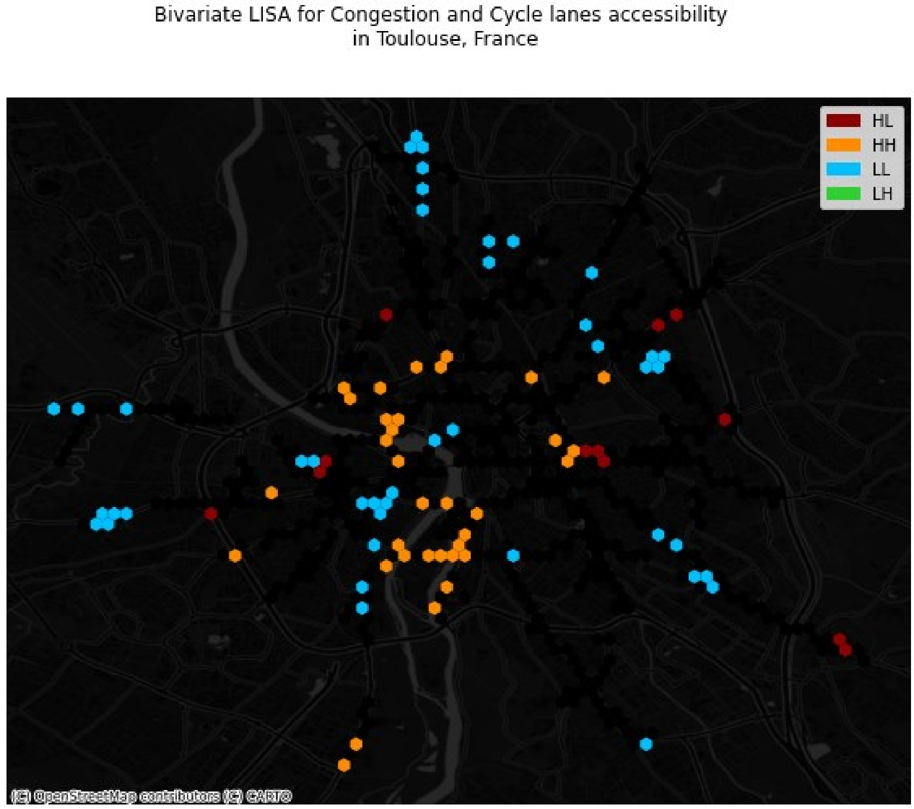

Bivariate Moran Scatterplot (0.21)
Spatial lag of length_std
flow_std


Moran BV Local Scatterplot
Spatial lag of length_std
flow_std


Moran BV Local Scatterplot
Spatial Lag
Attribute


HH
HL
LH
LL
ns


[-1.91, -0.75]
(-0.75, -0.41]
(-0.41, 0.03]
(0.03, 0.59]
(0.59, 4.79]


Bivariate LISA for Congestion and Cycle lanes accessibility
in Toulouse, France
HL
HH
LL
LH
(C) OpenStreetMap contributors (C) CARTO