HORUS
URBAN HEALTH

D2.3. REPORT ON ESTIMATES

Submission Date	30 September 2024
Responsible Partner	UVEG

Project Documentation	
Project Acronym	HORUS
Project Full Title	Health Outcomes from Raised Urban Settings
Grant Agreement	101136516
Call Identifier	HORIZON-HLTH-2023-DISEASE-03
Topic	HORIZON-HLTH-2023-DISEASE-03-03
Funding Scheme	Research and Innovation action (RIA)
Project Duration	36 months
Coordinator	Universitat de Valencia
Website	www.horus-urbanhealth.eu

Deliverable Documentation	
Number	D2.3
Title	Report on estimates
Related WP	WP2. Identifying and measuring the association between urban environment features and NCDs
Lead Beneficiary	UVEG
Nature of Deliverable	Report
Dissemination Level	Public
Submission Due Date	30/09/2024
Actual Submission Date	27/09/2024

Authors	
Author(s)	
Alfonso Gallego Valadés	UVEG
Tamara Alhambra Borrás	UVEG
Mirian Fernández Salido	UVEG
Jorge Garcés Ferrer	UVEG
Contributor(s)	
All partners	

TABLE OF CONTENT

LIST OF ACRONYMS AND ABBREVIATIONS

BMI: Body Mass Index

CBS: Statistics Netherlands

CONNECTIVITY: Count of accessible services

CVD: Cardiovascular Disease

DBT: Diabetes

DRK: Alcohol consumption

DRU: Drug use

GREEN SURFACE: Sqm of green surface per inhabitant

IND.ACCES: Accessibility

IND.CY.INF: Cycling infrastructure

IND.ECO: Economic activity

IND.FOOD: Food environment

IND.INTERS: Street network intersections

IND.OPEN: Open public areas

IND.PED.FAC: Pedestrian facilities

IND.PED.INF: Pedestrian infrastructure

IND.PT: Public transport

IND.TOP: Topology

LSL: Low sleep

NAC: Non-active commuting

NCD: Non-Communicable Disease

NDVI: Normalized Difference Vegetation Index

OR: Odds Ratio

OWT: Overweight

PCA: Principal Component Analysis

PHI: Physical inactivity

POP_DENS: Population density

PRF: Processed food

SMK: Smoking

T2DM: Type 2 Diabetes Mellitus

TRAFFIC_POP: % of population exposed to traffic and noise

WP: Work Package

EXECUTIVE SUMMARY: OBJECTIVE OF THE DELIVERABLE

Deliverable 2.3 – *Report on estimates* is a public document summarising the process of analysis for obtaining estimates of the association between the multidimensional physical-functional characteristics of the urban environment in the three project cities (Valencia, Rotterdam and Rijeka) and the prevalence of risk behaviours and factors for Non-Communicable Diseases (NCDs) and, in particular, Cardiovascular Diseases (CVDs) and Type 2 Diabetes Mellitus (T2DM). In addition to presenting these estimates, the main purpose of this Deliverable is to constitute a necessary precedent for the formation of criteria and decision-making in subsequent phases of the project. In particular, this document aims to provide insights on what types of neighbourhoods can be found in each city and what behavioural and health outcomes they are systematically associated with, in order to define the urban areas in which the tasks planned under the framework of Work Package (WP) 3 and 4 will be implemented.

This Deliverable is closely related to Deliverable 2.2 - *Thematic Maps*. In particular, this Deliverable builds on the same working line as its preceding one, while broadening the focus of the analysis. Data on urban characteristics used in the preparation of that document are in turn used in the present Deliverable, tracing a clear sequence of analysis on which the project as a whole is built. In addition, the mapping of the spatial distribution of risk factors and risk behaviours is extended to the case of Rijeka (Annex 2), which could not be included in the previous Deliverable due to timing issues.

The Deliverable is structured as follows: firstly, the methodological background behind the analysis approach is presented. This section summarises the sourcing and processing of the data, as well as the structure and methods of the analysis for each of the three project cities. Next, the main results are presented in the form of highlights, involving both the results from the neighbourhood classification analysis and the estimation of association measures. Annex 1 consists of the output of the .Rmd document in which the R code was written in order to allow traceability of the data analysis process performed for each of the three cities separately. As access to some of the survey microdata sets (in particular those from the GGD Health Monitor - Rotterdam) is restricted, the document as a whole is not fully reproducible, although the sections for Valencia and Rijeka can be. Finally, Annex 2 presents natural break maps (Jenks) plotting the spatial distribution of the behavioural and health outcomes of interest, based on the survey data from the municipality of Rijeka.

1. Methodological background

1.1 Data

1.1.1 Classification of neighbourhoods

The procedure for obtaining and processing data on the physical-functional characteristics of the urban environment is detailed in Deliverable 2.2. From this data, the neighbourhood-level variables of interest for this analysis are defined as follows:

- Urban determinant indices are specified in a 500 m wide cell hexagonal grid layer. For neighbourhoods, the indices are obtained by interpolating the cell-level indices to the residential area of the neighbourhood (only on land classified as residential by Corine Land Cover) through applying the weighted mean of the cell values (or their proportional share) corresponding to each neighbourhood:
  - **IND.CY.INF**: Cycling infrastructure. Relation of cycle lanes and general network. (Values from 0 to 1, being 1 equivalent to 100%, meaning all network contained in the cell is completely cyclable) (Valencia and Rotterdam).
  - **IND.FOOD**: Food environment. Counting of food services such as shops, markets and production spaces such as vegetable gardens.
  - **IND.OPEN**: Open public areas. Counting open public spaces such as squares.
  - **IND.PED.INF**: Pedestrian infrastructure. Relation of pedestrian paths and the general road network (values from 0 to 1, being 1 equivalent to 100%, meaning all network contained in the cell is completely pedestrian).
  - **IND.PT**: Public transport. Counting public transport stops.
  - **IND.ACCES**: Street network: Accessibility. Relation of the accessible area based on street network (service area) and the "theoretical" area (perfect circle of 250m radius) from the centroid of each cell (values from 0 to 1, being 1 equivalent to 100%).
  - **IND.TOP**: weighted mean of the cumulative altitude difference (Rijeka).
  - **IND.PED.FAC**: Pedestrian facilities. Counting pedestrian facilities and furniture, such as benches or pergola-type shade shelters.
  - **IND.INTERS**: Street network: intersections. Intersection counts (crossroads).
  - **IND.ECO**: Economic activity. Count of economic activities, shops or services (non-public).
- The rest of the urban determinants indices are directly specified by neighbourhood:

– **POP_DENS**: Population density.
– **CONNECTIVITY**: Count of accessible services within a given radius.
– **TRAFFIC_POP**: % of population exposed to traffic and noise.
– **GREEN SURFACE**: Sqm of green surface per inhabitant (based on NDVI).

Deliverable 2.2 also allows visualising the spatial distribution of these variables, both in the subset of indicators specified at grid level and those aggregated at neighbourhood level.

1.1.2 Individual-level survey data

From the individual-level survey data, a set of binary variables regarding different NCD risk behaviours and factors were constructed, as well as specific variables on the prevalence of CVD and T2DM diagnosis. Variables were defined as follows:

- **NAC** (Non-active commuting): 1 if respondent does not regularly spend at least 30 minutes for 5 days or more per week on daily commuting, walking or cycling; or at least 60 minutes for 3 or 4 days per week. 0 otherwise.
- **PHI** (Physical inactivity): 1 if respondent does not normally spend at least 30 minutes for 3 days or more per week exercising; or at least 60 minutes for 2 days per week. 0 otherwise
- **LSL** (Low sleep): 1 if respondent usually sleep less than 6 hours a day. 0 otherwise (Valencia and Rijeka).
- **PRF** (Processed food): 1 if respondent usually eats ultra-processed food at least 3 times a week. 0 otherwise.
- **SMK** (Smoking): 1 if respondent reports currently smoking. 0 otherwise.
- **DRK** (Alcohol consumption): 1 if respondent reports weekly average alcohol consumption above the low-risk threshold limit: more than 20 g/day for men (2 standard drinks) or more than 10 g/day for women (1 standard drink). 0 otherwise.
- **DRU** (Drug use): 1 if respondent reports having used drugs (including cannabis and nitrous oxide) in the last 4 weeks. 0 otherwise (Rotterdam).
- **OWT** (Overweight): 1 if respondent states Body Mass Index (BMI) > 25. 0 otherwise.
- **CVD** (Cardiovascular disease): 1 if respondent has diagnosed CVD. 0 otherwise.
- **DBT** (Diabetes): 1 if respondent has diagnosed diabetes. 0 otherwise.

For each city, variables were constructed from the sources listed below.

1.1.2.1 Valencia

The individual-level survey data for the city of **Valencia** comes from the City Council's *Health, Food and Sport Barometer*, which aimed to find out the opinion of the citizens of Valencia regarding various aspects of health and sporting practice. The study covered topics such as the assessment of the sports facilities available in the neighbourhoods, the sporting and healthy habits of the population, and commuting on foot, by bicycle or other non-motorised means of transport. It also inquired into the frequency and reasons for practising sport in leisure time, the most common sports and the places where they are practised, as well as the reasons why some citizens do not practise sport. In addition, the barometer assessed the population's perception of health, the main fears related to their well-being, the prevalence of diseases and their treatment, the prevention of health problems, sleeping habits, nutrition and addictive behaviours, such as tobacco and alcohol consumption.

The study was conducted through 2,298 personal interviews with people registered in Valencia, aged 18 and over. The sample was selected by quota sampling, controlled by sex and age at district level. The interviews were carried out in 287 sampling points distributed throughout the city, in the morning and afternoon, from Monday to Saturday, during the months of June and July 2019.

1.1.2.2 Rotterdam

The individual-level survey data for the city of **Rotterdam** comes from the *Dutch Health Monitor*, the *Gezondheidsmonitor*. This is a national questionnaire aimed at collecting information on the health status and wellbeing of Dutch citizens. This study includes topics such as perceived health, health behaviors (drinking, smoking, poor diet, exercise, among others), chronic conditions, anxiety and depression, stress, loneliness and noise pollution.

National and regional measurements of the Health Monitor take place every 4 years. Every four years, Statistics Netherlands (CBS) conducts an initial sampling of the general Dutch population aged 18 and above on a national level. Those selected receive an invitation letter to complete the *Gezondheidsmonitor* either digitally or on paper. In this study, participants have been contacted a maximum of four times; the fourth contact moment only took place in regions with a low response rate. At local level, the municipal health service is responsible for data collection.

The data used in this study was collected in 2020 among residents of the Rotterdam-Rijnmond region by the municipal health service of Rotterdam, GGD Rotterdam-Rijnmond. Data collection started on the 11th of September 2020 and ended on the 18th of December (a period of 15 weeks). Within the region of Rotterdam-Rijnmond, a total of 95,179 people have been contacted for participation within this study. Data from 13,194 participants from Rotterdam who provided informed consent and filled in the questionnaire are available.

1.1.2.3 Rijeka

A cross-sectional survey was conducted at the municipality of Rijeka between March and July 2024, during which 2,448 participants completed the *City of Rijeka Health Barometer* questionnaire. The study included individuals of both sexes, aged 18 and older, residing in the city of Rijeka. The questionnaires were distributed through various channels, including email (with a link to access the questionnaire via the Microsoft Forms platform), utility bills accompanied by a letter containing a QR code for online access, and social media platforms such as Facebook. Additionally, the QR code for the questionnaire was made available on the RCC portal of the City of Rijeka, where residents manage utility payments online. Paper versions of the questionnaire were distributed at primary healthcare centers under the jurisdiction of the Primorje-Gorski Kotar County Community Health Centre, as well as to local boards throughout the city and the Kantrida Retirement Home.

1.2 Methods and structure of the analysis

Annex 1 provides a comprehensive and traceable overview of the data analysis process conducted separately for the three selected cities: Valencia, Rotterdam and Rijeka. The analysis focuses on estimating the association between living in different types of neighbourhoods and individual behavioural and health outcomes. This is achieved by first classifying neighbourhoods according to a set of physical-functional urban characteristics using the k-means clustering algorithm. Such data are drawn from the work performed in the framework of the preparation of Deliverable 2.2. The *k* clusters for Valencia and Rotterdam are set to k = 4, assuming a quadrant typology bounding the location of observations in the two-dimensional space defined by the two main axis from Principal Component Analysis (PCA). For Rijeka, on the other hand, it is assumed that the major source of variability in the data comes from the difference between the inner-city and peri-urban environment, so k = 2 main clusters are set.

After the classification, a spatial distribution map of the clusters is presented, visually representing the classification and allowing to observe the degree of spatial autocorrelation of the urban physical-functional characteristics. In general, the degree of spatial autocorrelation in the assignment of neighbourhoods to clusters is remarkably high in each city. Considering the absence of spatial weighting in the implementation of the algorithm, this fact constitutes a test of the internal validity of the classification, which adequately reflects the relationships between variables within the dataset.

Estimation is then conducted using individual-level survey data, linking outcomes to the classification of respondents' residential neighborhoods. Binary logistic regressions are applied for each outcome of interest, enabling the analysis of deviance between model pairs — one with sociodemographic covariates and the other a full model incorporating the neighborhood cluster factor —. Adjusted Odds Ratios (ORs) are derived from these models, allowing for the assessment of whether individuals living in certain neighborhood clusters have higher or lower odds of exhibiting a behavior or outcome compared to those in a reference cluster, which in all cities is represented by peri-urban areas.

Finally, further specification analyses are presented to assess the robustness and consistency of the results. To do so, neighbourhoods are classified into three levels (low, medium, high) for each variable characterising the urban environment, establishing cut-off points along the distribution through natural breaks (Jenks). Logistic regressions are then estimated for the behavioural variables only, using plausible causal predictors of the outcome of interest.

2. Main results

2.1 Valencia

2.1.1 Classification of neighbourhoods

4 clusters are obtained, with the following sizes: 24, 15, 38, 8. The percentage of the between-cluster sum of squares with respect to the total is 53%. No overlapping between clusters is observed, although there are some differences in size and relative internal variability (Fig. 1). Clusters are defined as follow:

- **Cluster 1**: Shows moderate-low availability of public open spaces, public transport and commercial services. It has a moderate proportion of traffic-calmed streets, a moderate-low average density of road intersections and high exposure to traffic noise. Population density is moderate-low but variable, as is exposure to green space, which is moderate-low but also variable. It corresponds to secondary peripheral areas, some of them on the urban border.
- **Cluster 2** (reference Cluster for comparisons): Characterised by lower availability of food services, public open spaces, proportion of footpaths and traffic-calmed streets,

pedestrian facilities, public transport stops, economic activity and accesibility to services. It has the highest exposure to green areas. It corresponds mainly to the peripheral areas and secondary urban settlements in the municipality.

- **Cluster 3**: Has a moderate-high availability of food services, public open spaces and economic activity. It has the highest average presence of public transport stops, high average functional and service accesibility, as well as a moderate presence of pedestrian infrastructure. It is characterised by the highest population density and lowest exposure to green areas.
- **Cluster 4**: Includes areas with greater commercial and food service presence, high accessibility to services and connectivity, predominantly pedestrianised roads and moderate-high population density. Exposure to green areas has the lowest values. It corresponds mainly to neighbourhoods in the historic centre and *extra-muros*, as well as a neighbourhood located contiguous to the university area.

For further details, see Figs. 1-2 and Tab.1.

2.1.2 Association estimates

Compared to living in a **Cluster 2** neighbourhood, living in a **Cluster 1** neighbourhood is associated with:

- 47% higher odds of NAC (adjusted OR = 1.47, 95% CI [1.07, 2.02]).
- 32% lower odds of PHI (adjusted OR = 0.68, 95% CI [0.5, 0.92]).
- 46% lower odds of PRF (adjusted OR = 0.54, 95% CI [0.4, 0.74]).
- 40% lower odds of OWT (adjusted OR = 0.6, 95% CI [0.43, 0.82]).

Compared to living in a Cluster 2 neighbourhood, living in a **Cluster 3** neighbourhood is associated with:

- 36% lower odds of PHI (adjusted OR = 0.64, 95% CI [0.49, 0.83]).
- 47% lower odds of PRF (adjusted OR = 0.53, 95% CI [0.41, 0.69]).
- 50% lower odds of OWT (adjusted OR = 0.5, 95% CI [0.37, 0.66]).

Compared to living in a Cluster 2 neighbourhood, living in a **Cluster 4** neighbourhood is associated with:

- 6-fold higher odds of NAC (adjusted OR = 6.64, 95% CI [4.35, 10.25]).
- 63% lower odds of PRF (adjusted OR = 0.37, 95% CI [0.24, 0.56]).
- 54% lower odds of SMK (adjusted OR = 0.46, 95% CI [0.28, 0.75]).

- 44% lower odds of DRK (adjusted OR = 0.56, 95% CI [0.32, 0.93]).
- 2-fold higher odds of DBT (adjusted OR = 2.18, 95% CI [1.03, 4.59]).

For more details, see Table 3.

At conventional statistical significance level ($p < 0.05$), the full model specification, including the Cluster factor, improves the deviance in NAC, PHI, PRF, SMK and OWT compared to the socio-demographic variables model.

Additionally, it is observed that, at conventional statistical significance levels, compared to living in a neighbourhood with level = low, living in a neighbourhood with:

- level = medium and level = high of cycling infrastructure decrease 43% and 54%, respectively, the odds of NAC.
- level = high of cycling infrastructure decreases the odds of PHI by 31%.
- level = medium and level = high of public transport increase 72% and 91%, respectively, the odds of NAC.
- level = high of pedestrian facilities decreases the odds of PHI by 25%.
- level = medium of exposure to traffic increases by 2-fold the odds of LSL.
- level = medium and level = high of food environment decrease by 26% and 51%, respectively, the odds of PRF.

For more details, see Table 4.

2.2 Rotterdam

2.2.1 Classification of neighbourhoods

4 clusters are obtained, with the following sizes: 20, 16, 4, 21. The percentage of the between-cluster sum of squares with respect to the total is 41.3%. Slight overlap between clusters is observed (Fig. 3). Clusters are defined as follow:

- **Cluster 1** (reference Cluster for comparisons): Predominantly composed of peri-urban areas, characterised by low population density and low connectivity to services. It has the lowest levels of traffic exposure index and economic activity. The provision of pedestrian and public transport infrastructure is also limited. However, it stands out for having the highest green area per inhabitant, suggesting environments with high availability of natural spaces, but reduced access to key urban services.
- **Cluster 2**: This cluster includes areas with medium and low population densities, and is characterised by an intermediate supply of services and connections. The food

environment presents moderate average values, while the availability of public open spaces is the highest of all clusters, although the difference with respect to the average values of the rest of the clusters is possibly not very relevant. Public transport infrastructure and economic activity are at average levels. Despite having the lowest green area per capita, connectivity to services is medium-high, suggesting relatively easy access to basic services and transport.

- **Cluster 3**: This is the most densely populated cluster, mainly representing the historic centre of Rotterdam. Public infrastructure indices, such as transport and connectivity to services, are the highest in the municipality. The supply of food services and economic activity are also the highest, reflecting a high concentration of services and commerce in the area. Overall, the cluster has geographic accessibility to urban green infrastructure, although the median green area per capita is low.
- **Cluster 4**: Includes neighbourhoods with medium population densities, and peri-urban and urban sprawl areas. It is characterised by low connectivity to services and limited provision of pedestrian infrastructure. The economic activity and food environment also has low average values. In terms of exposure to traffic and noise, it has moderate to low values. The green area per inhabitant is at medium levels, characteristic of the discontinuous urban fabric typical of cities with medium-low densities.

For further details, see Figs. 3-4 and Tab.5.

2.2.2 Association estimates

Compared to living in a **Cluster 1** neighbourhood, living in a **Cluster 2** neighbourhood is associated with:

- 21% lower odds of NAC (adjusted OR = 0.79, 95% CI [0.63, 1]).
- 33% higher odds of PRF (adjusted OR = 1.33, 95% CI [1.07, 1.64]).
- 37% higher odds of SMK (adjusted OR = 1.37, 95% CI [1.21, 1.54]).
- 48% higher odds of DRU (adjusted OR = 1.48, 95% CI [1.19, 1.84]).
- 19% higher odds of DBT (adjusted OR = 1.19, 95% CI [1.01, 1.39]).

Compared to living in a Cluster 1 neighbourhood, living in a **Cluster 3** neighbourhood is associated with:

- 15% lower odds of PHI (adjusted OR = 0.85, 95% CI [0.73, 1]).
- 30% higher odds of SMK (adjusted OR = 1.3, 95% CI [1.08, 1.56]).
- 51% higher odds of DRK (adjusted OR = 1.51, 95% CI [1.27, 1.79]).

- 2-fold higher odds of DRU (adjusted OR = 2.05, 95% CI [1.53, 2.72]).
- 24% lower odds of OWT (adjusted OR = 0.76, 95% CI [0.65, 0.88]).

Compared to living in a Cluster 1 neighbourhood, living in a **Cluster 4** neighbourhood is associated with:

- 14% higher odds of DRK (adjusted OR = 1.14, 95% CI [1.02, 1.27]).
- 32% higher odds of DRU (adjusted OR = 1.32, 95% CI [1.07, 1.62]).
- 15% lower odds of OWT (adjusted OR = 0.85, 95% CI [0.78, 0.92]).

For more details, see Table 7.

At conventional statistical significance level, the full model specification, including the Cluster factor, improves the deviance in PHI, PRF, SMK, DRK, DRU, OWT and DBT compared to the socio-demographic variables model.

Additionally, it is observed that, at conventional statistical significance levels, compared to living in a neighbourhood with level = low, living in a neighbourhood with:

- level = medium of pedestrian facilities decreases 17% de odds of PHI.
- level = medium of economic activity increases 28% de odds of SMK.
- level = medium and level = high of economic activity increase 12% and 37%, respectively, the odds of DRK.
- level = medium and level = high of economic activity increase 54% and 79%, respectively, the odds of DRU.

For more details, see Table 8.

2.3 Rijeka

2.3.1 Classification of neighbourhoods

2 clusters are obtained, with the following sizes: 15, 19. The percentage of the between-cluster sum of squares with respect to the total is 34.2%. A stronger ratio between the variability attributable to the first component with respect to the second can be observed than in the previous cases, supporting the idea of the urban-peri-urban division (Fig. 5). Clusters are defined as follow:

- **Cluster 1** (reference Cluster for comparisons): This cluster is composed of predominantly peri-urban and pre-rural neighbourhoods characterised by low population density and connectivity to services. Areas in this cluster have limited food and pedestrian infrastructure. Access to public transport is relatively low, with modest average access to the street network. On the other hand, these areas are distinguished by a higher green area per capita and lower exposure to traffic and noise. Other indicators, such as pedestrian facilities and economic activity, are also low compared to Cluster 2.

- **Cluster 2**: Includes neighbourhoods located in the inner-city area and along the coast, which stand out for higher population density and significantly higher connectivity. In this cluster, food infrastructure is much more prevalent and the provision of public open space is considerably higher compared to Cluster 1. In addition, access to public transport is higher, as is the proportion of pedestrian infrastructure and pedestrian facilities. This cluster has a denser and more connected urban structure and high economic activity. Despite the higher exposure to traffic and noise, these areas have considerably improved accessibility compared to Cluster 1. In contrast, the green area per capita is very low, reflecting a more limited availability of green space.

For further details, see Figs. 5-6 and Tab.9.

2.3.2 Association estimates

Compared to living in a **Cluster 1** neighbourhood, living in a **Cluster 2** neighbourhood is associated with:

- 22% lower odds of NAC (adjusted OR = 0.78, CI 95% [0.66, 0.92]).
- 12% lower odds of PHI (although not reaching conventional statistically significant levels, adjusted OR = 0.88, CI 95% [0.74, 1.04]).
- 30% lower odds of LSL (adjusted OR = 0.7, CI 95% [0.5, 0.97]).
- 23% lower odds of PRF (adjusted OR = 0.77, CI 95% [0.65, 0.92]).
- 18% lower odds of OWT (adjusted OR = 0.82, CI 95% [0.69, 0.98]).

At conventional statistical significance level, the full model specification, including the Cluster factor, improves the deviance in NAC, LSL, PRF and OWT compared to the socio-demographic variables model.

3. Conclusions

This study explored the relationship between living in different types of neighbourhoods and the prevalence of NCD risk behaviours in three European cities: Valencia, Rotterdam and Rijeka. The results revealed significant associations between urban environment characteristics and several risk factors, including both expected and some unexpected findings. The main insights, as well as the limitations, strengths, weaknesses and added value of these results, are presented below.

Despite providing a comprehensive analysis of the influence of neighbourhood type on various risk behaviours, the study has certain limitations. First, its cross-sectional design prevents establishing direct causal relationships between the urban environment and the observed behaviours, as these associations could be affected by self-selection, generating a reverse causality problem characteristic of observational studies of this type. In addition, data availability varied between cities, limiting the generalisability of some findings, such as drug use, which was only measured in Rotterdam. Factors not considered, such as ethno-cultural, socio-economic or psychological influences, could be conditioning these associations. Also, the surveys did not include sufficient comparable information on socio-economic status, a key factor in lifestyle studies, although education level was used as a proxy. The five-year difference in data collection between cities should also be taken into account; although this time gap is considered reasonable due to the relative stability in the socio-demographic composition of neighbourhoods in the short term. Finally, heterogeneity in the urban characteristics of cities, such as their morphology and historical development, affect the overall interpretation of the results.

This study revealed some unexpected results. For example, in Valencia, living in neighbourhoods with greater accessibility to services and economic and commercial dynamism (Cluster 4) was associated with a reduction in the odds of alcohol consumption (DRK) and smoking (SMK), while in Rotterdam, the association was inverse. This could be due to socio-cultural characteristics not captured in this study. In Rotterdam, in addition, a positive association was observed between living in areas with high economic activity and drug use (DRU), which could be related to greater accessibility compared to other neighbourhoods. These findings suggest that the dynamics of risk behaviour may be influenced by a complex interaction of socio-economic factors and characteristics of the urban environment, which requires further research.

This study has several key strengths. The use of models adjusted for socio-demographic variables allowed controlling for the effect of individual factors on risk behaviour, which strengthens the findings on the relation with the environment. In addition, the classification of neighbourhoods based on specific urban indicators provided a contextualised characterisation of each city. However, a weakness is the heterogeneity in the size of the clusters and the partial overlap between some of them, which might make direct comparisons somewhat more difficult. The percentage of variability explained by the clusters ranged from 34.2% (Rijeka) to 53% (Valencia), indicating a moderate ability to distinguish urban areas according to the factors studied. Furthermore, the number of clusters was established a priori, based on theoretical assumptions; other classification methods might have generated different results.

This study provides certain conclusions for urban planning and public health. The urban environment was found to be significantly associated with the prevalence of risk behaviours, particularly in relation to physical inactivity (PHI) and consumption of ultra-processed foods (PRF). In terms of active mobility (NAC), there may be an interaction not captured in this study between public transport and the need to commute. In Valencia, living in neighbourhoods with better pedestrian and cycling infrastructure is associated with a reduction in physical inactivity and non-active commuting, suggesting that policies that encourage active mobility could have a positive impact on health. In Rotterdam, a similar pattern was observed in relation to pedestrian facilities and physical inactivity. Furthermore, the consistent relationship between higher connectivity and urban density with lower odds of being overweight and consuming processed foods reinforces the importance of creating environments that promote healthy habits. However, the discrepancy in findings on smoking and alcohol consumption in areas of high economic activity between cities indicates that there may be unexplored socio-cultural factors that could influence these associations and should be addressed in future research.

ANNEX 1. R CODE OF THE ANALYSIS FOR REPRODUCIBILITY

This chunk first loads all required packages, including those for reading Excel files (*readxl*), data manipulation (*dplyr*), spatial analysis (*sf*), and more. It then read the Excel file containing the neighbourhood-level data from the multi-city dataset into the data frame *original_dataset*, followed by the import of shapefiles for neighborhoods in Valencia, Rotterdam, and Rijeka with the *st_read* function from the *sf* package, enabling spatial analysis.

```
# Load all required packages
lapply(c("readxl","dplyr","factoextra","vtable",
         "sf", "sjPlot","sjmisc","sjlabelled","classInt"),
       require, character.only = TRUE)

# Import urban characteristics dataset
original_dataset <- read_excel("local/path/urbandata.xlsx")

# Import shapefiles
shape_VLC <- st_read("local/path/Valencia-neighborhoods.shp")
shape_RTM <- st_read("local/path/Rotterdam-neighborhoods.shp")
shape_RJK <- st_read("local/path/Rijeka-neighborhoods.shp")
```

1. Valencia

1.1. Classification of neighbourhoods

This chunk begins by filtering the dataset to include only neighbourhoods from Valencia with valid population values (*POP*). Next, a set of relevant variables related to urban characteristics (e.g., infrastructure, green surface, traffic, and connectivity, among others) is selected for further analysis. These variables are then scaled to standardize their ranges.

To classify the neighborhoods into groups, the k-means clustering algorithm is applied with *centers = 4* predefined clusters, ensuring reproducibility by setting a random seed. The result of the clustering is printed, showing the size of each cluster, the means of the variables within each cluster, and the clustering assignments for each neighborhood. Finally, the distribution of the clusters in the space of principal components as defined by *prcomp* is visualized using the *fviz_cluster* function, providing a clear representation of the neighborhood types in Valencia.

```
# Filter observations with valid values in column POP (total population) an
d City = Valencia
data_VLC <- original_dataset %>%
  filter(!is.na(POP) & City == "Valencia")

# Select variables of interest
variables_VLC <- data_VLC %>%
  select(IND.CY.INF, IND.FOOD, IND.OPEN, IND.PED.INF, IND.PT, IND.ACCESS,
         IND.PED.FAC, IND.INTERS, IND.ECO, POP_DENS, CONNECTIVITY,
         TRAFFIC_POP, GREEN.SURFACE)
```

```
# Scale variables
scaled_variables_VLC <- as.data.frame(scale(variables_VLC))

# Set seed for reproducibility
set.seed(123)

# Run k-means clustering
kmeans_result_VLC <- kmeans(scaled_variables_VLC, centers = 4, nstart = 25)
kmeans_result_VLC

## K-means clustering with 4 clusters of sizes 24, 15, 38, 8
##
## Cluster means:
##   IND.CY.INF   IND.FOOD    IND.OPEN  IND.PED.INF     IND.PT  IND.ACCESS
## 1 -0.2198560 -0.6887831 -0.20761035 -0.007117505 -0.3564954 -0.1519370
## 2 -1.1977178 -0.5754911 -0.93864411 -1.115991789 -1.3529448 -1.4966926
## 3  0.7241250  0.2781964  0.03611472  0.051375028  0.7356180  0.4969898
## 4 -0.5343047  1.8239621  2.21124382  1.869805736  0.1120722  0.9014079
##   IND.PED.FAC IND.INTERS    IND.ECO    POP_DENS CONNECTIVITY TRAFFIC_POP
## 1  0.24061756 -0.2330045 -0.5531986 -0.52573127   -0.3850627   0.4459734
## 2 -1.02390476 -1.3614619 -0.7310571 -1.17099754   -1.4149297  -1.1101781
## 3  0.04383288  0.4107931  0.1976022  0.77459448    0.5169296   0.3305341
## 4  0.98976256  1.3004875  2.0917173  0.09349043    1.3527661  -0.8263732
##   GREEN.SURFACE
## 1    -0.1932600
## 2     1.0428301
## 3    -0.2396507
## 4    -0.2371855
##
## Clustering vector:
##  [1] 4 4 4 4 4 4 3 3 3 4 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 1 1 1 1 3 1 1
 1 2 3 3
## [39] 1 1 1 3 3 1 1 3 1 1 1 1 1 3 2 3 3 3 3 3 3 4 3 3 1 3 1 3 3 1 3 1 1 2
 2 2 2 2
## [77] 1 2 2 2 2 2 2 2 2
##
## Within cluster sum of squares by cluster:
## [1] 118.73333 116.56421 208.74370  68.68681
##  (between_SS / total_SS =  53.0 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.wit
hinss"
## [6] "betweenss"    "size"         "iter"         "ifault"

# Visualize clusters
fviz_cluster(kmeans_result_VLC, data = scaled_variables_VLC,
             geom = "point", stand = FALSE) +
  scale_colour_manual(values = c("#f1c40f", "#2ecc71",
```

```
                                "#3498db", "#e74c3c")) +
  scale_fill_manual(values = c("#f1c40f", "#2ecc71", "#3498db", "#e74c3c"))
```

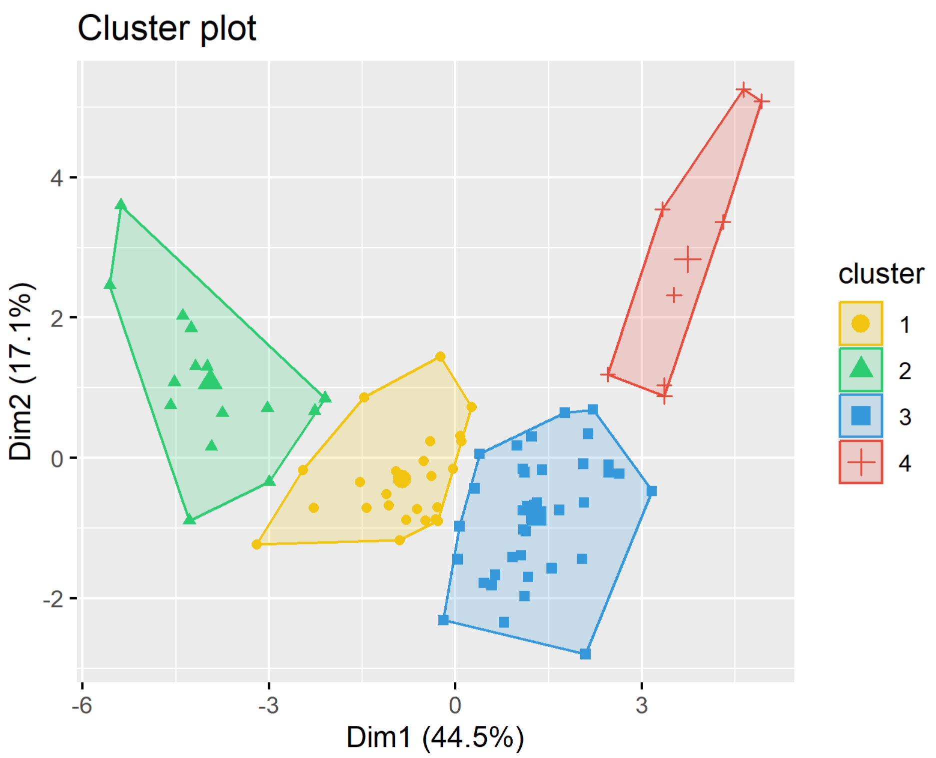

Fig. 1. K-means clusters in a biplot. Valencia data

Cluster labels are then assigned to the data frame and summary statistics are generated for each cluster, including medians, using the *st* function.

```
# Assign Clusters to a new data frame
variables_cluster_VLC <- variables_VLC %>%
  mutate(Cluster = kmeans_result_VLC$cluster)

# Summary statistics
st(variables_cluster_VLC,
   add.median = TRUE,
   group = 'Cluster',
   title = "**Tab. 1**. Summary statistics for clusters in Valencia")
```

Tab. 1. Summary statistics for clusters in Valencia

Variable	N	M	SD	Md	N	M	SD	Md	N	M	SD	Md	N	M	SD	Md
Cluster		1				2				3				4		
IND.CY.INF	24	0.062	0.027	0.066	15	0.024	0.022	0.024	38	0.099	0.029	0.095	8	0.05	0.03	0.055
IND.FOOD	24	1.8	1.9	1	15	2.2	2.5	1.7	38	5.1	2.4	4.9	8	10	3.1	11
IND.OPEN	24	3.8	2.1	3.4	15	0.82	0.82	0.42	38	4.7	2.1	4.7	8	13	5.5	11
IND.PED.INF	24	0.43	0.08	0.44	15	0.25	0.11	0.25	38	0.44	0.12	0.47	8	0.74	0.14	0.74
IND.PT	24	4.3	1.5	3.9	15	1.8	1.4	1.3	38	7	1.6	6.9	8	5.5	2.2	5.7

Variable	N	M	SD	Md	N	M	SD	Md	N	M	SD	Md	N	M	SD	Md
Cluster		1				2				3				4		
IND.ACCESS	24	0.37	0.087	0.4	15	0.2	0.11	0.19	38	0.46	0.069	0.46	8	0.51	0.09	0.53
IND.PED.FAC	24	7.6	5.6	5.9	15	1.7	1.8	1.6	38	6.7	3.4	6.4	8	11	3.8	9.8
IND.INTERS	24	946	364	919	15	199	172	135	38	1373	540	1435	8	1962	394	1955
IND.ECO	24	11	7.6	9.7	15	4.1	6.4	1.8	38	40	26	34	8	113	54	112
POP_DENS	24	116	87	94	15	30	43	5.2	38	291	100	304	8	199	75	178
CONNECTIVITY	24	15	6.1	17	15	3.5	4.3	1.9	38	26	7.1	25	8	35	6.6	35
TRAFFIC_POP	24	34	9.4	33	15	15	12	9	38	32	8.5	32	8	18	12	16
GREEN.SURFACE	24	19	48	3	15	461	763	139	38	2.4	4.2	1	8	3.2	3.9	1.5

1.2. Mapping clusters

In order to map the spatial distribution of the clusters, this chunk assigns cluster labels to the Valencia neighbourhood dataset, merges it with the *sf* object *shape_VLC* containing the bounding polygons through a left join, and visualises the spatial distribution of the clusters using *ggplot2* with custom color palette.

```
# Assign Cluster to data
data_VLC$Cluster <- as.factor(kmeans_result_VLC$cluster)

# Do the left join
VLC_geo_sep <- left_join(shape_VLC, data_VLC, by = c("City" = "City", "codd
istbar" = "number"))

# Define colors for clusters
cluster_colors_VLC <- c("1" = "#f1c40f", "2" = "#2ecc71",
                        "3" = "#3498db", "4" = "#e74c3c", "lightgray")

# Plot the results: Valencia
ggplot(data = VLC_geo_sep) +
  geom_sf(aes(fill = Cluster), color = "white") +
  scale_fill_manual(values = cluster_colors_VLC, na.value = "lightgray",
                    name = "Cluster") +
  theme_minimal() +
  labs(fill = "Cluster")
```

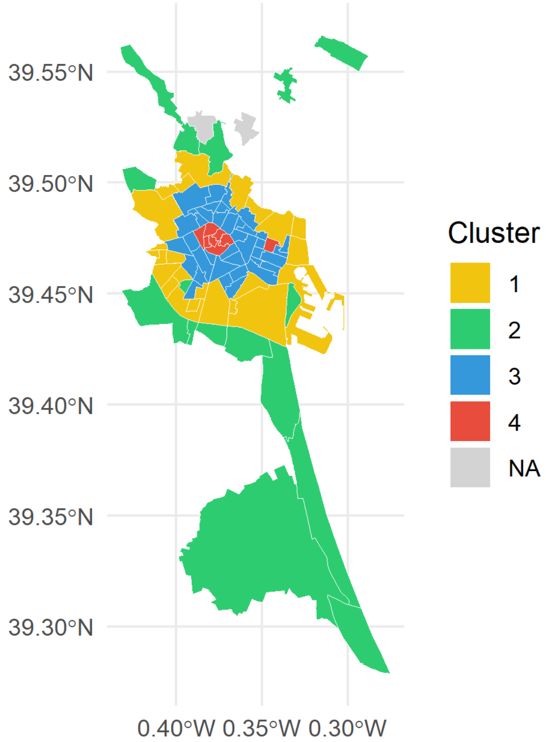

Fig. 2. K-means clusters map from Valencia data analysis

1.3. Estimation of associations

Once the clusters have been defined, the estimation of the effects proceeds, starting with the import of the data.

```
# Import data
survey_data_VLC <- read_excel("local/path/survey_data_VLC.xlsx")
```

This chunk displays the first few rows of the survey dataset for Valencia and prepares the data for analysis by selecting the relevant variables. A left join is then performed to merge the survey data with the neighborhood clusters based on the neighborhood codes.

```
# Print first rows
head(survey_data_VLC)

## # A tibble: 6 x 15
##      ID city     neighbourhood Sex     Age Education act.commuting
##   <dbl> <chr>    <chr>         <chr> <dbl> <chr>             <dbl>
## 1     4 Valencia 153           1        18 3                     1
## 2     6 Valencia 094           0        25 5                     1
## 3     7 Valencia 091           1        44 2                     0
## 4     8 Valencia 091           0        37 4                     0
## 5     9 Valencia 092           0        62 2                     0
## 6    10 Valencia 092           1        35 3                     0
## # i 8 more variables: physical.activity <dbl>, sleep <dbl>, diet <dbl>,
```

```
## #   smoking <dbl>, drinking <dbl>, overweight <dbl>, CVD <dbl>, Diabetes
 <dbl>
```

```
# Prepare the data for doing the left join
data_to_join_VLC <- VLC_geo_sep %>%
  select(coddistbar, Cluster)

# Left join
VLC_data_model <- left_join(survey_data_VLC, data_to_join_VLC, by = c("neig
hbourhood" = "coddistbar"))
```

The following chunk fits multiple logistic regression models to examine the relationship between various socio-demographic factors (*Sex*, *Age*, *Education*) and the health-related behavioral outcomes of interest. Rows with complete data for the specified columns are first identified to ensure valid model comparisons. Separate logistic regression models are then fitted using the `glm` function with the binomial family and logit link for each outcome. The subset parameter ensures that only complete cases are included in each model, and the variable *Age* is standardized for better interpretability of the intercept.

```
# Set the subset for further model comparison
valid_rows_VLC <- complete.cases(VLC_data_model[, c("Sex", "Age", "Educatio
n", "Cluster")])

# Models for socio-demographics
m.com.vlc.0 <- glm(act.commuting ~ Sex + scale(Age) + Education, data = VLC
_data_model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.pa.vlc.0 <- glm(physical.activity ~ Sex + scale(Age) + Education, data =
VLC_data_model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.sl.vlc.0 <- glm(sleep ~ Sex + scale(Age) + Education, data = VLC_data_mod
el, family = binomial(link='logit'), subset = valid_rows_VLC)
m.di.vlc.0 <- glm(diet ~ Sex + scale(Age) + Education, data = VLC_data_mode
l, family = binomial(link='logit'), subset = valid_rows_VLC)
m.sm.vlc.0 <- glm(smoking ~ Sex + scale(Age) + Education, data = VLC_data_m
odel, family = binomial(link='logit'), subset = valid_rows_VLC)
m.dr.vlc.0 <- glm(drinking ~ Sex + scale(Age) + Education, data = VLC_data_
model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.bmi.vlc.0 <- glm(overweight ~ Sex + scale(Age) + Education, data = VLC_da
ta_model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.cvd.vlc.0 <- glm(CVD ~ Sex + scale(Age) + Education, data = VLC_data_mode
l, family = binomial(link='logit'), subset = valid_rows_VLC)
m.db.vlc.0 <- glm(Diabetes ~ Sex + scale(Age) + Education, data = VLC_data_
model, family = binomial(link='logit'), subset = valid_rows_VLC)
```

Once implemented the regressions, this chunk allows to display the outputs with `tab_model`.

```
# Visualization of results
tab_model(m.com.vlc.0, m.pa.vlc.0, m.sl.vlc.0, m.di.vlc.0,
          m.sm.vlc.0, m.dr.vlc.0, m.bmi.vlc.0, m.cvd.vlc.0, m.db.vlc.0,
          collapse.ci = TRUE,
          p.style = "stars",
```

```
           show.r2 = FALSE,
           pred.labels = c("Intercept", "Sex (Man)","Age (standarized)",
                           "Education: Primary school",
                           "Education: Secondary school",
                           "Education: Vocational studies",
                           "Education: University"),
           dv.labels = c("NAC", "PHI", "LSL", "PRF",
                         "SMK", "DRK", "OWT", "CVD", "DBT"),
           title = "**Tab 2**. ORs for socio-demographics. Valencia")
```

Tab 2. ORs for socio-demographics. Valencia

Predictors	NAC	PHI	LSL	PRF	SMK	DRK	OWT	CVD	DBT
Intercept	0.95 (0.64 – 1.39)	1.92 ** (1.30 – 2.86)	0.19 *** (0.11 – 0.31)	0.82 (0.55 – 1.22)	0.30 *** (0.18 – 0.47)	0.09 *** (0.04 – 0.16)	0.82 (0.53 – 1.27)	0.02 *** (0.01 – 0.05)	0.06 *** (0.03 – 0.11)
Sex (Man)	1.07 (0.90 – 1.28)	0.64 *** (0.54 – 0.76)	0.64 ** (0.48 – 0.83)	1.50 *** (1.26 – 1.78)	1.42 *** (1.18 – 1.72)	1.54 *** (1.25 – 1.91)	1.55 *** (1.30 – 1.86)	2.03 ** (1.28 – 3.26)	0.95 (0.67 – 1.35)
Age (standarized)	0.99 (0.90 – 1.09)	0.99 (0.91 – 1.09)	1.50 *** (1.30 – 1.75)	0.69 *** (0.63 – 0.76)	0.72 *** (0.65 – 0.80)	0.98 (0.87 – 1.09)	1.62 *** (1.47 – 1.78)	3.19 *** (2.34 – 4.46)	3.31 *** (2.61 – 4.25)
Education: Primary school	0.52 ** (0.35 – 0.78)	0.55 ** (0.36 – 0.82)	1.01 (0.61 – 1.73)	0.75 (0.49 – 1.14)	1.34 (0.83 – 2.23)	2.33 *(1.26 – 4.74)	1.39 (0.87 – 2.19)	0.82 (0.41 – 1.67)	0.93 (0.55 – 1.60)
Education: Secondary school	0.45 *** (0.29 – 0.68)	0.40 *** (0.26 – 0.60)	0.70 (0.40 – 1.24)	0.61 * (0.40 – 0.94)	1.00 (0.61 – 1.70)	2.76 ** (1.47 – 5.67)	0.97 (0.61 – 1.55)	0.89 (0.41 – 1.97)	0.35 ** (0.17 – 0.69)
Education: Vocational studies	0.49 *** (0.32 – 0.75)	0.44 *** (0.28 – 0.67)	0.65 (0.36 – 1.18)	0.66 (0.43 – 1.03)	1.07 (0.65 – 1.81)	2.81 ** (1.49 – 5.79)	0.89 (0.55 – 1.42)	0.41 (0.14 – 1.11)	0.80 (0.41 – 1.56)
Education: University	0.43 *** (0.28 – 0.65)	0.23 *** (0.14 – 0.35)	0.54 * (0.30 – 0.99)	0.44 *** (0.28 – 0.68)	0.59 (0.36 – 1.02)	1.66 (0.87 – 3.45)	0.57 * (0.36 – 0.92)	0.34 * (0.12 – 0.89)	0.40 * (0.20 – 0.82)
Observations	2271	2275	2276	2274	2276	2274	2148	2276	2276

* p<0.05   ** p<0.01   *** p<0.001

The ecological variable `Cluster` is then incorporated into the models for each outcome of interest. This factor allows to obtain the OR associated with living in each type of neighbourhood, versus living in Cluster 2 (reference), adjusting for individual socio-demographic characteristics.

```
# Relevel factor
VLC_data_model <- within(VLC_data_model, Cluster <- relevel(Cluster, ref =
2))

# Adjusted models
m.com.vlc.1 <- glm(act.commuting ~ Cluster + Sex + scale(Age) + Education,
data = VLC_data_model, family = binomial(link='logit'), subset = valid_rows
```

```
_VLC)
m.pa.vlc.1 <- glm(physical.activity ~ Cluster + Sex + scale(Age) + Educatio
n, data = VLC_data_model, family = binomial(link='logit'), subset = valid_r
ows_VLC)
m.sl.vlc.1 <- glm(sleep ~ Cluster + Sex + scale(Age) + Education, data = VL
C_data_model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.di.vlc.1 <- glm(diet ~ Cluster + Sex + scale(Age) + Education, data = VLC
_data_model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.sm.vlc.1 <- glm(smoking ~ Cluster + Sex + scale(Age) + Education, data =
VLC_data_model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.dr.vlc.1 <- glm(drinking ~ Cluster + Sex + scale(Age) + Education, data =
 VLC_data_model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.bmi.vlc.1 <- glm(overweight ~ Cluster + Sex + scale(Age) + Education, dat
a = VLC_data_model, family = binomial(link='logit'), subset = valid_rows_VL
C)
m.cvd.vlc.1 <- glm(CVD ~ Cluster + Sex + scale(Age) + Education, data = VLC
_data_model, family = binomial(link='logit'), subset = valid_rows_VLC)
m.db.vlc.1 <- glm(Diabetes ~ Cluster + Sex + scale(Age) + Education, data =
 VLC_data_model, family = binomial(link='logit'), subset = valid_rows_VLC)

# Visualization of results
tab_model(m.com.vlc.1, m.pa.vlc.1, m.sl.vlc.1, m.di.vlc.1,
          m.sm.vlc.1, m.dr.vlc.1, m.bmi.vlc.1, m.cvd.vlc.1, m.db.vlc.1,
          collapse.ci = TRUE,
          p.style = "stars",
          show.r2 = FALSE,
          pred.labels = c("Intercept", "Cluster: 1", "Cluster: 3",
                          "Cluster: 4", "Sex (Man)", "Age (standarized)",
                          "Education: Primary school",
                          "Education: Secondary school",
                          "Education: Vocational studies",
                          "Education: University"),
          dv.labels = c("NAC", "PHI", "LSL", "PRF",
                        "SMK", "DRK", "OWT", "CVD", "DBT"),
          title = "**Tab 3**. Adjusted ORs in Valencia")
```

Tab 3. Adjusted ORs in Valencia

Predictors	NAC	PHI	LSL	PRF	SMK	DRK	OWT	CVD	DBT
Intercept	0.70 (0.44 – 1.11)	2.57 *** (1.64 – 4.07)	0.14 *** (0.07 – 0.25)	1.45 (0.92 – 2.29)	0.36 *** (0.21 – 0.61)	0.11 *** (0.05 – 0.21)	1.31 (0.80 – 2.18)	0.02 *** (0.01 – 0.06)	0.05 *** (0.02 – 0.10)
Cluster: 1	1.47 * (1.07 – 2.02)	0.68 * (0.50 – 0.92)	1.50 (0.91 – 2.52)	0.54 *** (0.40 – 0.74)	0.88 (0.64 – 1.22)	0.70 (0.48 – 1.00)	0.60 ** (0.43 – 0.82)	1.01 (0.48 – 2.18)	0.91 (0.47 – 1.80)
Cluster: 3	0.95 (0.71 – 1.26)	0.64 *** (0.49 – 0.83)	1.53 (0.99 – 2.45)	0.53 *** (0.41 – 0.69)	0.83 (0.63 – 1.11)	0.87 (0.65 – 1.19)	0.50 *** (0.37 – 0.66)	0.76 (0.40 – 1.53)	1.36 (0.80 – 2.45)
Cluster: 4	6.64 *** (4.35 – 10.25)	1.18 (0.80 – 1.76)	1.09 (0.54 – 2.14)	0.37 *** (0.24 – 0.56)	0.46 ** (0.28 – 0.75)	0.56 * (0.32 – 0.93)	0.81 (0.53 – 1.24)	0.81 (0.27 – 2.17)	2.18 * (1.03 – 4.59)

Predictors	NAC	PHI	LSL	PRF	SMK	DRK	OWT	CVD	DBT
Sex (Man)	1.07 (0.90 – 1.29)	0.63 *** (0.53 – 0.75)	0.64 ** (0.48 – 0.84)	1.50 *** (1.26 – 1.78)	1.43 *** (1.18 – 1.73)	1.55 *** (1.25 – 1.91)	1.56 *** (1.30 – 1.86)	2.04 ** (1.28 – 3.27)	0.96 (0.67 – 1.36)
Age (standarized)	0.99 (0.90 – 1.10)	0.99 (0.91 – 1.09)	1.50 *** (1.29 – 1.74)	0.69 *** (0.63 – 0.76)	0.72 *** (0.65 – 0.80)	0.98 (0.87 – 1.09)	1.63 *** (1.48 – 1.80)	3.22 *** (2.35 – 4.51)	3.32 *** (2.61 – 4.27)
Education: Primary school	0.60 * (0.39 – 0.91)	0.57 ** (0.37 – 0.86)	0.98 (0.59 – 1.67)	0.73 (0.48 – 1.12)	1.32 (0.81 – 2.20)	2.22 * (1.20 – 4.53)	1.46 (0.91 – 2.33)	0.84 (0.43 – 1.74)	0.95 (0.55 – 1.65)
Education: Secondary school	0.50 ** (0.32 – 0.77)	0.41 *** (0.27 – 0.63)	0.67 (0.38 – 1.18)	0.61 * (0.39 – 0.94)	0.99 (0.60 – 1.68)	2.67 ** (1.42 – 5.49)	1.04 (0.64 – 1.67)	0.92 (0.42 – 2.06)	0.34 ** (0.17 – 0.69)
Education: Vocational studies	0.58 * (0.37 – 0.90)	0.47 *** (0.30 – 0.72)	0.61 (0.34 – 1.11)	0.65 (0.42 – 1.02)	1.04 (0.63 – 1.77)	2.70 ** (1.43 – 5.56)	0.96 (0.59 – 1.54)	0.43 (0.15 – 1.16)	0.82 (0.42 – 1.62)
Education: University	0.41 *** (0.26 – 0.64)	0.23 *** (0.15 – 0.36)	0.52 * (0.29 – 0.95)	0.46 *** (0.30 – 0.72)	0.62 (0.37 – 1.06)	1.64 (0.86 – 3.42)	0.61 * (0.38 – 0.99)	0.36 * (0.13 – 0.95)	0.39 * (0.19 – 0.80)
Observations	2271	2275	2276	2274	2276	2274	2148	2276	2276

* p<0.05   ** p<0.01   *** p<0.001

To estimate how much the deviance is improved by the inclusion of the ecological variable `Cluster` in the individual socio-demographic variable models, the *anova* function is applied specifying `test = "Chisq"`, which allows to obtain the probability for the Chi-squared-distributed likelihood ratio test statistic, given the null is true.

```
# Model comparison
anova(m.com.vlc.0, m.com.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: act.commuting ~ Sex + scale(Age) + Education
## Model 2: act.commuting ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      2264     2845.7
## 2      2261     2713.1  3   132.63 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.pa.vlc.0, m.pa.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: physical.activity ~ Sex + scale(Age) + Education
## Model 2: physical.activity ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      2268     2952.5
## 2      2265     2931.0  3   21.465 8.427e-05 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.sl.vlc.0, m.sl.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: sleep ~ Sex + scale(Age) + Education
## Model 2: sleep ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2269     1520.0
## 2      2266     1515.2  3   4.7699   0.1894

anova(m.di.vlc.0, m.di.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: diet ~ Sex + scale(Age) + Education
## Model 2: diet ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      2267     2938.6
## 2      2264     2909.5  3    29.09 2.144e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.sm.vlc.0, m.sm.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: smoking ~ Sex + scale(Age) + Education
## Model 2: smoking ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2269     2556.5
## 2      2266     2546.2  3   10.377  0.01562 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.dr.vlc.0, m.dr.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: drinking ~ Sex + scale(Age) + Education
## Model 2: drinking ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2267     2224.2
## 2      2264     2216.4  3     7.72  0.05217 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.bmi.vlc.0, m.bmi.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: overweight ~ Sex + scale(Age) + Education
```

```
## Model 2: overweight ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      2141     2778.3
## 2      2138     2749.4  3   28.856 2.401e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.cvd.vlc.0, m.cvd.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: CVD ~ Sex + scale(Age) + Education
## Model 2: CVD ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2269     595.66
## 2      2266     594.37  3   1.2872   0.7322

anova(m.db.vlc.0, m.db.vlc.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: Diabetes ~ Sex + scale(Age) + Education
## Model 2: Diabetes ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2269     945.41
## 2      2266     938.10  3   7.3079  0.06271 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1.4. Specification tests

The adjusted association between living in each type of neighborhood and NCD-related health outcomes has already been estimated. However, while the clusters are composed of observations that are similar on average in terms of the observed characteristics, they also exhibit internal heterogeneity. Therefore, it is considered relevant to implement additional specification tests to obtain more information on the main sources of variability among the variables included in the cluster classification. The specified variables are categorized into three groups ("Low," "Medium," "High") using Jenks natural breaks, with new columns (suffix '.pb') storing the classified values. Specific columns are printed to verify the classification results, and the newly classified columns are selected and merged with *data_VLC* based on the common identifier, resulting in *VLC_sp_data_model*.

```
# Variables to classify
variables_to_classify_VLC <- c("IND.CY.INF", "IND.FOOD", "IND.OPEN", "IND.P
ED.INF", "IND.PT", "IND.ACCESS", "IND.PED.FAC", "IND.INTERS", "IND.ECO", "P
OP_DENS", "CONNECTIVITY", "TRAFFIC_POP", "GREEN.SURFACE")

# Classify the variables and create new columns with suffix 'pb'.
for (var in variables_to_classify_VLC) {
  breaks <- classIntervals(data_VLC[[var]], n = 3, style = "jenks")$brks
  data_VLC[[paste0(var, ".pb")]] <- cut(data_VLC[[var]], breaks = breaks, l
```

```
abels = c("Low", "Medium", "High"), include.lowest = TRUE)
}

# Print only specific columns for checking the result
print(data_VLC[, c("IND.CY.INF", "IND.CY.INF.pb", "IND.FOOD", "IND.FOOD.pb"
)])

## # A tibble: 85 x 4
##    IND.CY.INF IND.CY.INF.pb IND.FOOD IND.FOOD.pb
##         <dbl> <fct>            <dbl> <fct>
##  1   0.0378   Low              13.5  High
##  2   0.0961   Medium            6.86 Medium
##  3   0.0614   Medium            7.54 Medium
##  4   0.0182   Low              13.4  High
##  5   0.000565 Low              13.9  High
##  6   0.0495   Low              11.6  High
##  7   0.104    High              7.49 Medium
##  8   0.121    High              8.33 Medium
##  9   0.131    High              6.85 Medium
## 10   0.0683   Medium            9.65 High
## # i 75 more rows

# Select only the new columns generated for the join
columns_for_join_VLC <- paste0(variables_to_classify_VLC, ".pb")

# Perform left join with 'VLC_data_model' based on the variable 'neighbourh
ood'.
VLC_sp_data_model <- VLC_data_model %>%
  left_join(data_VLC[, c("number", columns_for_join_VLC)], by = c("neighbou
rhood" = "number"))
```

Alternative specifications are proposed below, focusing only on the models with behavioral outcomes, as the health endpoints involve a more complex combination of causal factors. These models include, as predictors, the original variables hypothesized to represent the largest source of conditional variability for each outcome.

```
# Specification models
sp.m.com.vlc <- glm(act.commuting ~ Sex + scale(Age) + Education + IND.CY.I
NF.pb + IND.PED.INF.pb + IND.PT.pb, data = VLC_sp_data_model, family = bino
mial(link='logit'))
sp.m.pa.vlc <- glm(physical.activity ~ Sex + scale(Age) + Education + IND.C
Y.INF.pb + IND.OPEN.pb + IND.PED.INF.pb + IND.PED.FAC.pb, data = VLC_sp_dat
a_model, family = binomial(link='logit'))
sp.m.sl.vlc <- glm(sleep ~ Sex + scale(Age) + Education + TRAFFIC_POP.pb +
IND.ECO.pb + POP_DENS.pb, data = VLC_sp_data_model, family = binomial(link=
'logit'))
sp.m.di.vlc <- glm(diet ~ Sex + scale(Age) + Education + IND.FOOD.pb, data
= VLC_sp_data_model, family = binomial(link='logit'))
sp.m.sm.vlc <- glm(smoking ~ Sex + scale(Age) + Education + IND.ECO.pb, dat
a = VLC_sp_data_model, family = binomial(link='logit'))
sp.m.dr.vlc <- glm(drinking ~ Sex + scale(Age) + Education + IND.ECO.pb, da
ta = VLC_sp_data_model, family = binomial(link='logit'))
```

```r
# Visualization of results
tab_model(sp.m.com.vlc, sp.m.pa.vlc, sp.m.sl.vlc,
          sp.m.di.vlc, sp.m.sm.vlc, sp.m.dr.vlc,
          collapse.ci = TRUE,
          p.style = "stars",
          auto.label = TRUE,
          show.r2 = FALSE,
          pred.labels = c("Intercept", "Sex (Man)", "Age (standarized)",
                          "Education: Primary school",
                          "Education: Secondary school",
                          "Education: Vocational studies",
                          "Education: University",
                          "Cycle infrastructure: Medium",
                          "Cycle infrastructure: High",
                          "Pedestrian infrastructure: Medium",
                          "Pedestrian infrastructure: High",
                          "Public transport: Medium",
                          "Public transport: High",
                          "Open public areas: Medium",
                          "Open public areas: High",
                          "Pedestrian facilities: Medium",
                          "Pedestrian facilities: High",
                          "Traffic exposure: Medium",
                          "Traffic exposure: High",
                          "Economic activity: Medium",
                          "Economic activity: High",
                          "Population density: Medium",
                          "Population density: High",
                          "Food environment: Medium",
                          "Food environment: High"),
          dv.labels = c("NAC", "PHI", "LSL", "PRF", "SMK", "DRK"),
          title = "**Tab 4**. Adjusted ORs for specification variables in V
alencia")
```

Tab 4. Adjusted ORs for specification variables in Valencia

Predictors	NAC	PHI	LSL	PRF	SMK	DRK
Intercept	0.81 (0.51 – 1.29)	2.55 *** (1.66 – 3.95)	0.13 *** (0.07 – 0.24)	0.94 (0.62 – 1.40)	0.32 *** (0.19 – 0.50)	0.08 *** (0.04 – 0.15)
Sex (Man)	1.10 (0.91 – 1.32)	0.63 *** (0.53 – 0.75)	0.64 ** (0.48 – 0.84)	1.50 *** (1.26 – 1.78)	1.43 *** (1.18 – 1.73)	1.57 *** (1.27 – 1.93)
Age (standarized)	1.01 (0.92 – 1.11)	1.01 (0.92 – 1.10)	1.49 *** (1.29 – 1.73)	0.69 *** (0.63 – 0.76)	0.73 *** (0.66 – 0.80)	0.97 (0.87 – 1.08)
Education: Primary school	0.59 * (0.39 – 0.91)	0.58 ** (0.38 – 0.87)	1.00 (0.60 – 1.71)	0.75 (0.50 – 1.15)	1.33 (0.83 – 2.23)	2.13 * (1.14 – 4.34)
Education: Secondary school	0.53 ** (0.34 – 0.82)	0.44 *** (0.28 – 0.67)	0.67 (0.38 – 1.21)	0.63 * (0.41 – 0.97)	1.01 (0.61 – 1.70)	2.43 ** (1.29 – 5.01)

Predictors	NAC	PHI	LSL	PRF	SMK	DRK
Education: Vocational studies	0.62 * (0.39 – 0.96)	0.49 ** (0.31 – 0.75)	0.62 (0.35 – 1.14)	0.67 (0.43 – 1.04)	1.06 (0.64 – 1.80)	2.62 ** (1.39 – 5.40)
Education: University	0.46 *** (0.29 – 0.72)	0.25 *** (0.16 – 0.39)	0.55 (0.30 – 1.02)	0.47 *** (0.30 – 0.73)	0.61 (0.36 – 1.05)	1.42 (0.74 – 2.96)
Cycle infrastructure: Medium	0.57 *** (0.46 – 0.72)	0.82 (0.65 – 1.02)				
Cycle infrastructure: High	0.46 *** (0.35 – 0.61)	0.69 ** (0.53 – 0.90)				
Pedestrian infrastructure: Medium	0.77 ** (0.63 – 0.94)	0.82 (0.67 – 1.01)				
Pedestrian infrastructure: High	14.63 *** (7.92 – 29.35)	1.26 (0.60 – 2.63)				
Public transport: Medium	1.72 *** (1.31 – 2.29)					
Public transport: High	1.91 *** (1.40 – 2.62)					
Open public areas: Medium		1.04 (0.84 – 1.28)				
Open public areas: High		1.58 (0.64 – 3.95)				
Pedestrian facilities: Medium		0.85 (0.69 – 1.05)				
Pedestrian facilities: High		0.75 * (0.58 – 0.98)				
Traffic exposure: Medium			2.14 ** (1.37 – 3.40)			
Traffic exposure: High			1.28 (0.79 – 2.10)			
Economic activity: Medium			1.03 (0.74 – 1.41)		0.91 (0.74 – 1.11)	1.57 *** (1.26 – 1.95)
Economic activity: High			1.05 (0.56 – 1.87)		0.63 (0.39 – 0.98)	0.82 (0.48 – 1.34)

Predictors	NAC	PHI	LSL	PRF	SMK	DRK
Population density: Medium			0.90 (0.58 – 1.39)			
Population density: High			0.79 (0.51 – 1.23)			
Food environment: Medium				0.74 ** (0.61 – 0.89)		
Food environment: High				0.49 *** (0.35 – 0.69)		
Observations	2271	2275	2276	2274	2276	2274

* p<0.05 ** p<0.01 *** p<0.001

2. Rotterdam

The code is presented as in the previous section.

2.1. Classification of neighbourhoods

```
# Filter observations with valid values in column POP (total population) an
d City = Rotterdam
data_RTM <- original_dataset %>%
  filter(!is.na(POP) & City == "Rotterdam")

# Select variables of interest
variables_RTM <- data_RTM %>%
  select(IND.CY.INF, IND.FOOD, IND.OPEN, IND.PED.INF, IND.PT, IND.ACCESS,
         IND.PED.FAC, IND.INTERS, IND.ECO, POP_DENS, CONNECTIVITY,
         TRAFFIC_POP, GREEN.SURFACE)

# Scale variables
scaled_variables_RTM <- as.data.frame(scale(variables_RTM))


# Set seed for reproducibility
set.seed(123)

# Run k-means clustering
kmeans_result_RTM <- kmeans(scaled_variables_RTM, centers = 4, nstart = 25)
kmeans_result_RTM
## K-means clustering with 4 clusters of sizes 20, 16, 4, 21
##
## Cluster means:
##    IND.CY.INF   IND.FOOD   IND.OPEN IND.PED.INF     IND.PT IND.ACCESS
## 1 -0.31187789 -0.5028623 -0.3889990   0.2701522 -0.2855686  0.1691311
## 2 -0.23022629  0.4913126  0.9777476   0.8847092  0.1861971  0.8911831
## 3 -0.08773978  2.4286357  0.4934487   0.1202957  2.3466059  0.6393457
## 4  0.48914940 -0.3580142 -0.4684655  -0.9542654 -0.3168669 -0.9618540
```

```
##   IND.PED.FAC IND.INTERS    IND.ECO   POP_DENS CONNECTIVITY TRAFFIC_POP
## 1  -0.3764938  0.5731712 -0.5006649 -0.6351538   -0.5926657  -0.2592798
## 2   0.4286199  0.2829843  0.3675768  0.1190108    0.9503848  -0.3918465
## 3   1.6790558  0.4478353  2.9594289  1.2261081    1.7977952   0.9921251
## 4  -0.2878222 -0.8467864 -0.3669356  0.2806891   -0.5020964   0.3565066
##   GREEN.SURFACE
## 1   0.492762201
## 2  -0.295990252
## 3  -0.002746622
## 4  -0.243257786
##
## Clustering vector:
##  [1] 3 4 4 4 4 2 3 4 4 4 2 3 1 4 1 4 2 1 2 1 1 1 3 4 1 2 4 2 2 2 2 4 1 2
 1 1 1 4
## [39] 1 1 1 1 1 4 2 4 2 2 4 2 1 1 1 4 1 2 4 4 2 4 4
##
## Within cluster sum of squares by cluster:
## [1] 127.31575 132.76911  58.58001 139.44102
##  (between_SS / total_SS =  41.3 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.wit
hinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```r
# Visualize clusters
fviz_cluster(kmeans_result_RTM, data = scaled_variables_RTM,
             geom = "point", stand = FALSE) +
  scale_colour_manual(values = c("#2ecc71", "#3498db",
                                 "#e74c3c", "#f1c40f")) +
  scale_fill_manual(values = c("#2ecc71", "#3498db", "#e74c3c", "#f1c40f"))
```

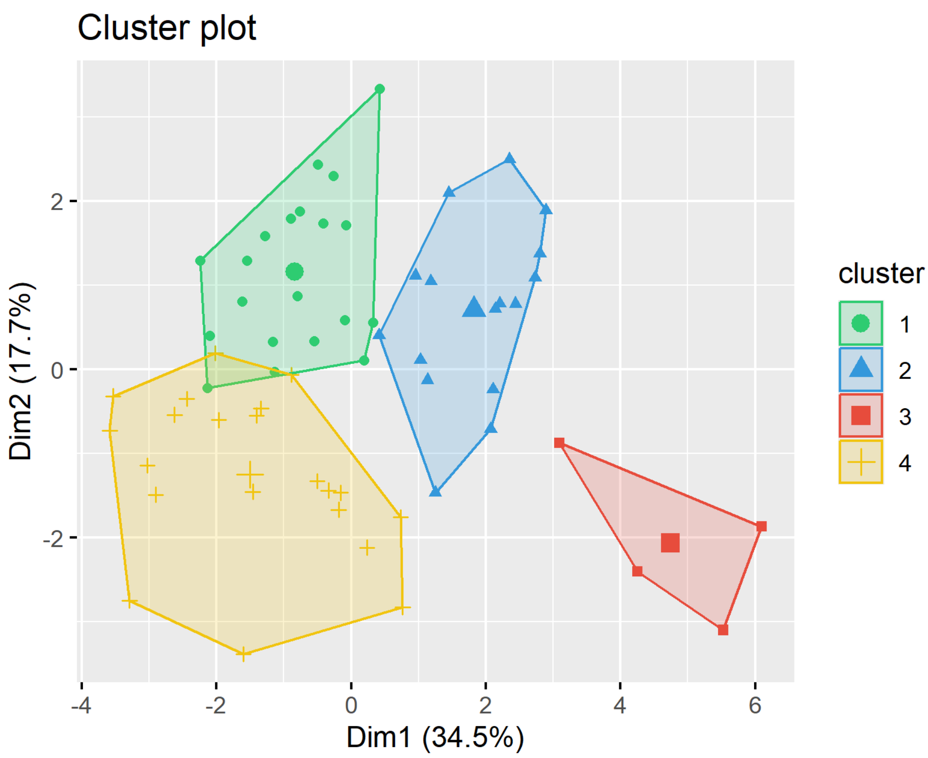

Fig. 3. K-means clusters in a biplot. Rotterdam data

```
# Assign Clusters to a new data frame
variables_cluster_RTM <- variables_RTM %>%
  mutate(Cluster = kmeans_result_RTM$cluster)

# Summary statistics
st(variables_cluster_RTM,
   add.median = TRUE,
   group = 'Cluster',
   title = "**Tab. 5**. Summary statistics for clusters in Rotterdam")
```

Tab. 5. Summary statistics for clusters in Rotterdam

Variable	N	M	SD	Md	N	M	SD	Md	N	M	SD	Md	N	M	SD	Md
Cluster		1				2				3				4		
IND.CY.INF	20	0.12	0.029	0.12	16	0.12	0.043	0.11	4	0.13	0.022	0.12	21	0.15	0.034	0.15
IND.FOOD	20	0.83	0.36	0.78	16	2.6	1	2.6	4	6	3.8	6.2	21	1.1	0.91	0.8
IND.OPEN	20	1	0.56	0.86	16	2.4	1.1	2.3	4	1.9	1.2	1.7	21	0.94	0.65	0.71
IND.PED.INF	20	0.62	0.065	0.6	16	0.69	0.055	0.69	4	0.6	0.049	0.61	21	0.49	0.094	0.51
IND.PT	20	2.4	1	2.3	16	3.4	1.7	3.2	4	8.2	4.3	6.5	21	2.3	1.6	1.9
IND.ACCESS	20	0.34	0.057	0.33	16	0.41	0.063	0.43	4	0.39	0.037	0.4	21	0.24	0.066	0.23
IND.PED.FAC	20	2.5	1.9	2	16	9.3	12	5.1	4	20	4.9	18	21	3.2	5.2	1.6
IND.INTERS	20	69	25	70	16	62	18	60	4	66	14	68	21	36	11	33

Variable	N	M	SD	Md	N	M	SD	Md	N	M	SD	Md	N	M	SD	Md
Cluster		1				2				3				4		
IND.ECO	20	4.7	2.9	3.9	16	18	7	19	4	59	25	64	21	6.8	6.8	5.3
POP_DENS	20	24	22	24	16	90	86	63	4	187	178	148	21	104	78	66
CONNECTIVITY	20	4.6	2.4	4.3	16	12	3.4	13	4	17	1.7	17	21	5	3.3	4.6
TRAFFIC_POP	20	16	8.3	15	16	15	6.5	15	4	31	12	28	21	23	14	25
GREEN.SURFACE	20	1431	2957	160	16	9.4	12	5.5	4	538	1072	2.5	21	104	213	23

2.2. Mapping clusters

```r
# Assign Cluster to data
data_RTM$Cluster <- as.factor(kmeans_result_RTM$cluster)

# Do the left join
RTM_geo_sep <- left_join(shape_RTM, data_RTM, by = c("City" = "City", "CBS_buurtc" = "number"))

# Define colors for clusters
cluster_colors_RTM <- c("1" = "#2ecc71", "2" = "#3498db", "3" = "#e74c3c", "4" = "#f1c40f", "lightgray")

# Plot the results: Rotterdam
ggplot(data = RTM_geo_sep) +
  geom_sf(aes(fill = Cluster), color = "white") +
  scale_fill_manual(values = cluster_colors_RTM, na.value = "lightgray",

  name = "Cluster") +
  theme_minimal() +
  labs(fill = "Cluster")
```

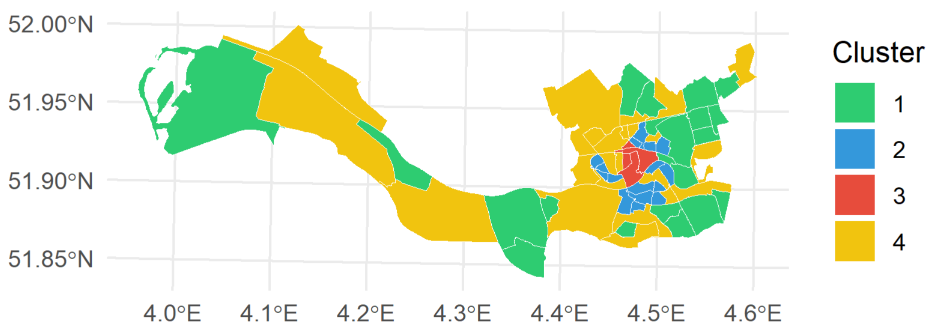

Fig. 4. K-means clusters map from Rotterdam data analysis

2.3. Estimation of associations

```
# Import data
survey_data_RTM <- read_excel("local/path/survey_data_RTM.xlsx")

# Print first rows
head(survey_data_RTM)

## # A tibble: 6 x 15
##           ID city      neighbourhood   Sex   Age Education act.commuting
##        <dbl> <chr>             <dbl> <dbl> <dbl>     <dbl>         <dbl>
## 1 1911010329 Rotterdam             8     1    NA        NA             1
## 2 1911010330 Rotterdam            30     0    NA        NA             1
## 3 1911010337 Rotterdam             5     0    NA        NA             1
## 4 1911010338 Rotterdam            43     1    NA        NA             1
## 5 1911010359 Rotterdam            55     0    NA        NA             1
## 6 1911010365 Rotterdam             6     0    NA        NA            NA
## # i 8 more variables: physical.activity <dbl>, diet <dbl>, smoking <dbl>
,
## #   drinking <dbl>, overweight <dbl>, CVD <dbl>, Diabetes <dbl>, drugs <
dbl>

# Prepare the data for doing the left join
data_to_join_RTM <- RTM_geo_sep %>%
  select(CBS_buurtc, Cluster)

# Required data type transformations
survey_data_RTM$neighbourhood <- as.character(survey_data_RTM$neighbourhood
)
survey_data_RTM$Education <- as.factor(survey_data_RTM$Education)

# Left join
RTM_data_model <- left_join(survey_data_RTM, data_to_join_RTM, by = c("neig
hbourhood" = "CBS_buurtc"))
```

```
# Set the subset for further model comparison
valid_rows_RTM <- complete.cases(RTM_data_model[, c("Sex", "Age", "Educatio
n", "Cluster")])

# Models for socio-demographics
m.com.rtm.0 <- glm(act.commuting ~ Sex + scale(Age) + Education, data = RTM
_data_model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.pa.rtm.0 <- glm(physical.activity ~ Sex + scale(Age) + Education, data =
RTM_data_model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.di.rtm.0 <- glm(diet ~ Sex + scale(Age) + Education, data = RTM_data_mode
l, family = binomial(link='logit'), subset = valid_rows_RTM)
m.sm.rtm.0 <- glm(smoking ~ Sex + scale(Age) + Education, data = RTM_data_m
odel, family = binomial(link='logit'), subset = valid_rows_RTM)
m.dr.rtm.0 <- glm(drinking ~ Sex + scale(Age) + Education, data = RTM_data_
model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.dg.rtm.0 <- glm(drugs ~ Sex + scale(Age) + Education, data = RTM_data_mod
```

```r
el, family = binomial(link='logit'), subset = valid_rows_RTM)
m.bmi.rtm.0 <- glm(overweight ~ Sex + scale(Age) + Education, data = RTM_da
ta_model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.cvd.rtm.0 <- glm(CVD ~ Sex + scale(Age) + Education, data = RTM_data_mode
l, family = binomial(link='logit'), subset = valid_rows_RTM)
m.db.rtm.0 <- glm(Diabetes ~ Sex + scale(Age) + Education, data = RTM_data_
model, family = binomial(link='logit'), subset = valid_rows_RTM)

# Visualization of results
tab_model(m.com.rtm.0, m.pa.rtm.0, m.di.rtm.0, m.sm.rtm.0,
          m.dr.rtm.0, m.dg.rtm.0, m.bmi.rtm.0, m.cvd.rtm.0, m.db.rtm.0,
          collapse.ci = TRUE,
          p.style = "stars",
          show.r2 = FALSE,
          pred.labels = c("Intercept", "Sex (Man)","Age (standarized)",
                          "Education: Secondary school",
                          "Education: Vocational studies",
                          "Education: University"),
          dv.labels = c("NAC", "PHI", "PRF", "SMK",
                        "DRK", "DRU", "OWT", "CVD", "DBT"),
          title = "**Tab 6**. ORs for socio-demographics. Rotterdam")
```

Tab 6. ORs for socio-demographics. Rotterdam

Predictors	NAC	PHI	PRF	SMK	DRK	DRU	OWT	CVD	DBT
Intercept	18.5 *** (13.31 – 26.68)	6.61 *** (5.56 – 7.92)	0.11 *** (0.08 – 0.14)	0.23 *** (0.20 – 0.27)	0.09 *** (0.07 – 0.11)	0.03 *** (0.02 – 0.04)	1.39 *** (1.23 – 1.58)	0.13 *** (0.11 – 0.15)	0.14 *** (0.12 – 0.17)
Sex (Man)	0.91 (0.77 – 1.08)	0.70 *** (0.65 – 0.76)	1.37 *** (1.16 – 1.62)	1.63 *** (1.48 – 1.78)	0.89 * (0.81 – 0.98)	2.09 *** (1.78 – 2.45)	1.23 *** (1.14 – 1.32)	1.23 *** (1.11 – 1.36)	1.52 *** (1.35 – 1.71)
Age (standarized)	1.91 *** (1.74 – 2.09)	1.70 *** (1.63 – 1.77)	0.49 *** (0.43 – 0.56)	0.72 *** (0.69 – 0.76)	1.15 *** (1.09 – 1.21)	0.36 *** (0.31 – 0.40)	1.56 *** (1.50 – 1.62)	2.72 *** (2.54 – 2.93)	2.69 *** (2.47 – 2.94)
Education: Secondary school	1.15 (0.78 – 1.66)	0.55 *** (0.45 – 0.66)	0.52 *** (0.37 – 0.73)	0.85 (0.73 – 1.00)	2.44 *** (1.97 – 3.04)	1.18 (0.78 – 1.87)	0.74 *** (0.65 – 0.85)	0.95 (0.81 – 1.12)	0.52 *** (0.44 – 0.61)
Education: Vocational studies	0.94 (0.63 – 1.38)	0.52 *** (0.43 – 0.64)	0.62 ** (0.44 – 0.89)	0.87 (0.72 – 1.04)	2.08 *** (1.64 – 2.65)	0.97 (0.63 – 1.57)	0.91 (0.78 – 1.07)	0.95 (0.78 – 1.15)	0.47 *** (0.38 – 0.57)
Education: University	1.26 (0.86 – 1.81)	0.28 *** (0.23 – 0.34)	0.25 *** (0.18 – 0.35)	0.48 *** (0.41 – 0.57)	3.23 *** (2.61 – 4.04)	1.00 (0.66 – 1.56)	0.44 *** (0.38 – 0.51)	0.70 *** (0.59 – 0.83)	0.24 *** (0.20 – 0.29)
Observations	10,541	13,194	7593	13,150	12,763	7458	12,919	12,882	13,004

* p<0.05 ** p<0.01 *** p<0.001

```r
# Adjusted models
m.com.rtm.1 <- glm(act.commuting ~ Cluster + Sex + scale(Age) + Education,
data = RTM_data_model, family = binomial(link='logit'), subset = valid_rows
```

```
_RTM)
m.pa.rtm.1 <- glm(physical.activity ~ Cluster + Sex + scale(Age) + Educatio
n, data = RTM_data_model, family = binomial(link='logit'), subset = valid_r
ows_RTM)
m.di.rtm.1 <- glm(diet ~ Cluster + Sex + scale(Age) + Education, data = RTM
_data_model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.sm.rtm.1 <- glm(smoking ~ Cluster + Sex + scale(Age) + Education, data =
RTM_data_model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.dr.rtm.1 <- glm(drinking ~ Cluster + Sex + scale(Age) + Education, data =
 RTM_data_model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.dg.rtm.1 <- glm(drugs ~ Cluster + Sex + scale(Age) + Education, data = RT
M_data_model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.bmi.rtm.1 <- glm(overweight ~ Cluster + Sex + scale(Age) + Education, dat
a = RTM_data_model, family = binomial(link='logit'), subset = valid_rows_RT
M)
m.cvd.rtm.1 <- glm(CVD ~ Cluster + Sex + scale(Age) + Education, data = RTM
_data_model, family = binomial(link='logit'), subset = valid_rows_RTM)
m.db.rtm.1 <- glm(Diabetes ~ Cluster + Sex + scale(Age) + Education, data =
 RTM_data_model, family = binomial(link='logit'), subset = valid_rows_RTM)

# Visualization of results
tab_model(m.com.rtm.1, m.pa.rtm.1, m.di.rtm.1, m.sm.rtm.1, m.dr.rtm.1,
          m.dg.rtm.1, m.bmi.rtm.1, m.cvd.rtm.1, m.db.rtm.1,
          collapse.ci = TRUE,
          p.style = "stars",
          auto.label = TRUE,
          show.r2 = FALSE,
          pred.labels = c("Intercept", "Cluster: 2", "Cluster: 3",
                          "Cluster: 4", "Sex (Man)", "Age (standarized)",
                          "Education: Secondary school",
                          "Education: Vocational studies",
                          "Education: University"),
          dv.labels = c("NAC", "PHI", "PRF", SMK",
                        "DRK", "DRU", "OWT", "CVD", "DBT"),
          title = "**Tab 7**. Adjusted ORs in Rotterdam")
```

Tab 7. Adjusted ORs in Rotterdam

Predictors	NAC	PHI	PRF	SMK	DRK	DRU	OWT	CVD	DBT
Intercept	20.9 *** (14.61 – 30.90)	6.60 *** (5.48 – 7.99)	0.10 *** (0.07 – 0.14)	0.20 *** (0.17 – 0.23)	0.09 *** (0.07 – 0.11)	0.02 *** (0.01 – 0.04)	1.46 *** (1.27 – 1.68)	0.14 *** (0.12 – 0.17)	0.14 *** (0.11 – 0.16)
Cluster: 2	0.79 * (0.63 – 1.00)	1.10 (0.98 – 1.22)	1.33 ** (1.07 – 1.64)	1.37 *** (1.21 – 1.54)	1.03 (0.90 – 1.17)	1.48 *** (1.19 – 1.84)	1.04 (0.94 – 1.15)	0.93 (0.81 – 1.07)	1.19 * (1.01 – 1.39)
Cluster: 3	0.94 (0.67 – 1.37)	0.85 * (0.73 – 1.00)	0.85 (0.58 – 1.22)	1.30 ** (1.08 – 1.56)	1.51 *** (1.27 – 1.79)	2.05 *** (1.53 – 2.72)	0.76 *** (0.65 – 0.88)	1.02 (0.82 – 1.26)	0.87 (0.66 – 1.14)
Cluster: 4	0.90 (0.73 – 1.12)	0.92 (0.84 – 1.01)	0.84 (0.68 – 1.04)	1.05 (0.94 – 1.18)	1.14 * (1.02 – 1.27)	1.32 ** (1.07 – 1.62)	0.85 *** (0.78 – 0.92)	0.89 (0.79 – 1.01)	0.94 (0.82 – 1.09)

Predictors	NAC	PHI	PRF	SMK	DRK	DRU	OWT	CVD	DBT
Sex (Man)	0.91 (0.77 – 1.08)	0.70 *** (0.65 – 0.76)	1.37 *** (1.16 – 1.63)	1.62 *** (1.48 – 1.78)	0.89 * (0.81 – 0.98)	2.08 *** (1.77 – 2.44)	1.23 *** (1.15 – 1.33)	1.23 *** (1.11 – 1.36)	1.52 *** (1.35 – 1.71)
Age (standarized)	1.90 *** (1.73 – 2.08)	1.70 *** (1.63 – 1.78)	0.50 *** (0.44 – 0.56)	0.73 *** (0.69 – 0.76)	1.15 *** (1.10 – 1.21)	0.36 *** (0.32 – 0.41)	1.57 *** (1.51 – 1.63)	2.71 *** (2.53 – 2.91)	2.71 *** (2.49 – 2.96)
Education: Secondary school	1.11 (0.75 – 1.60)	0.56 *** (0.46 – 0.67)	0.56 *** (0.40 – 0.79)	0.89 (0.76 – 1.05)	2.43 *** (1.96 – 3.03)	1.21 (0.79 – 1.92)	0.76 *** (0.66 – 0.87)	0.94 (0.80 – 1.11)	0.53 *** (0.46 – 0.63)
Education: Vocational studies	0.91 (0.60 – 1.34)	0.53 *** (0.43 – 0.65)	0.66 * (0.47 – 0.94)	0.91 (0.76 – 1.10)	2.09 *** (1.65 – 2.67)	1.02 (0.66 – 1.65)	0.92 (0.79 – 1.08)	0.94 (0.77 – 1.14)	0.48 *** (0.39 – 0.59)
Education: University	1.23 (0.84 – 1.76)	0.29 *** (0.24 – 0.35)	0.27 *** (0.19 – 0.38)	0.50 *** (0.42 – 0.59)	3.13 *** (2.52 – 3.92)	0.96 (0.64 – 1.51)	0.46 *** (0.40 – 0.53)	0.69 *** (0.58 – 0.83)	0.25 *** (0.21 – 0.31)
Observations	10,541	13,194	7593	13,150	12,763	7458	12919	12882	13004

* p<0.05 ** p<0.01 *** p<0.001

```
anova(m.com.rtm.0, m.com.rtm.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: act.commuting ~ Sex + scale(Age) + Education
## Model 2: act.commuting ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1     10535     4248.8
## 2     10532     4244.9  3   3.9094   0.2714

anova(m.pa.rtm.0, m.pa.rtm.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: physical.activity ~ Sex + scale(Age) + Education
## Model 2: physical.activity ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1     13188      14856
## 2     13185      14842  3   14.117  0.00275 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.di.rtm.0, m.di.rtm.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: diet ~ Sex + scale(Age) + Education
## Model 2: diet ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      7587     4055.1
```

```
## 2      7584     4036.9  3   18.137 0.0004122 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.sm.rtm.0, m.sm.rtm.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: smoking ~ Sex + scale(Age) + Education
## Model 2: smoking ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1     13144      11821
## 2     13141      11790  3   30.812 9.31e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.dr.rtm.0, m.dr.rtm.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: drinking ~ Sex + scale(Age) + Education
## Model 2: drinking ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1     12757      11874
## 2     12754      11850  3   23.723 2.854e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.dg.rtm.0, m.dg.rtm.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: drugs ~ Sex + scale(Age) + Education
## Model 2: drugs ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      7452     4362.6
## 2      7449     4336.2  3   26.445 7.694e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.bmi.rtm.0, m.bmi.rtm.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: overweight ~ Sex + scale(Age) + Education
## Model 2: overweight ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1     12913      16781
## 2     12910      16750  3   30.875 9.034e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.cvd.rtm.0, m.cvd.rtm.1, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: CVD ~ Sex + scale(Age) + Education
## Model 2: CVD ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1     12876     9516.8
## 2     12873     9512.6  3   4.1535   0.2454
```

```
anova(m.db.rtm.0, m.db.rtm.1, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Diabetes ~ Sex + scale(Age) + Education
## Model 2: Diabetes ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1     12998     7427.1
## 2     12995     7417.4  3   9.6998   0.0213 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.4. Specification tests

```
# Variables to classify
variables_to_classify_RTM <- c("IND.CY.INF", "IND.FOOD", "IND.OPEN", "IND.P
ED.INF", "IND.PT", "IND.ACCESS", "IND.PED.FAC", "IND.INTERS", "IND.ECO", "P
OP_DENS", "CONNECTIVITY", "TRAFFIC_POP", "GREEN.SURFACE")

# Classify the variables and create new columns with suffix 'pb'.
for (var in variables_to_classify_RTM) {
  breaks <- classIntervals(data_RTM[[var]], n = 3, style = "jenks")$brks
  data_RTM[[paste0(var, ".pb")]] <- cut(data_RTM[[var]], breaks = breaks, l
abels = c("Low", "Medium", "High"), include.lowest = TRUE)
}

# Print only specific columns for checking the result
print(data_RTM[, c("IND.CY.INF", "IND.CY.INF.pb", "IND.FOOD", "IND.FOOD.pb"
)])
```

```
## # A tibble: 61 x 4
##    IND.CY.INF IND.CY.INF.pb IND.FOOD IND.FOOD.pb
##         <dbl> <fct>            <dbl> <fct>
##  1     0.162  High             9.40  High
##  2     0.202  High             0.465 Low
##  3     0.164  High             0.854 Low
##  4     0.170  High             2.10  Medium
##  5     0.152  Medium           0.326 Low
##  6     0.0825 Low              2.63  Medium
##  7     0.113  Low              2.36  Medium
##  8     0.155  Medium           1.80  Medium
##  9     0.201  High             0.922 Low
## 10     0.151  Medium           2.24  Medium
## # i 51 more rows
```

```r
# Select only the new columns generated for the join
columns_for_join_RTM <- paste0(variables_to_classify_RTM, ".pb")

# Perform left join with 'RTM_data_model' based on the variable 'neighbourh
ood'.
RTM_sp_data_model <- RTM_data_model %>%
  left_join(data_RTM[, c("number", columns_for_join_RTM)], by = c("neighbou
rhood" = "number"))

# Specification models
sp.m.com.rtm <- glm(act.commuting ~ Sex + scale(Age) + Education + IND.CY.I
NF.pb + IND.PED.INF.pb + IND.PT.pb, data = RTM_sp_data_model, family = bino
mial(link='logit'))
sp.m.pa.rtm <- glm(physical.activity ~ Sex + scale(Age) + Education + IND.C
Y.INF.pb + IND.OPEN.pb + IND.PED.INF.pb + IND.PED.FAC.pb, data = RTM_sp_dat
a_model, family = binomial(link='logit'))
sp.m.di.rtm <- glm(diet ~ Sex + scale(Age) + Education + IND.FOOD.pb, data
= RTM_sp_data_model, family = binomial(link='logit'))
sp.m.sm.rtm <- glm(smoking ~ Sex + scale(Age) + Education + IND.ECO.pb, dat
a = RTM_sp_data_model, family = binomial(link='logit'))
sp.m.dr.rtm <- glm(drinking ~ Sex + scale(Age) + Education + IND.ECO.pb, da
ta = RTM_sp_data_model, family = binomial(link='logit'))
sp.m.dg.rtm <- glm(drugs ~ Sex + scale(Age) + Education + IND.ECO.pb, data
= RTM_sp_data_model, family = binomial(link='logit'))

# Visualization of results
tab_model(sp.m.com.rtm, sp.m.pa.rtm, sp.m.di.rtm,
          sp.m.sm.rtm, sp.m.dr.rtm, sp.m.dg.rtm,
          collapse.ci = TRUE,
          p.style = "stars",
          auto.label = TRUE,
          show.r2 = FALSE,
          pred.labels = c("Intercept", "Sex (Man)", "Age (standarized)",
                          "Education: Secondary school",
                          "Education: Vocational studies",
                          "Education: University",
                          "Cycle infrastructure: Medium",
                          "Cycle infrastructure: High",
                          "Pedestrian infrastructure: Medium",
                          "Pedestrian infrastructure: High",
                          "Public transport: Medium",
                          "Public transport: High",
                          "Open public areas: Medium",
                          "Open public areas: High",
                          "Pedestrian facilities: Medium",
                          "Pedestrian facilities: High",
                          "Food environment: Medium",
                          "Food environment: High",
                          "Economic activity: Medium",
                          "Economic activity: High"),
          dv.labels = c("NAC", "PHI", "PRF",
```

```
                        "SMK", "DRK", "DRU"),
          title = "**Tab 8**. Adjusted ORs for specification variables in R
otterdam")
```

Tab 8. Adjusted ORs for specification variables in Rotterdam

Predictors	NAC	PHI	PRF	SMK	DRK	DRU
Intercept	18.74 *** (12.45 – 28.92)	6.03 *** (4.91 – 7.45)	0.10 *** (0.07 – 0.14)	0.21 *** (0.18 – 0.24)	0.09 *** (0.07 – 0.11)	0.02 *** (0.02 – 0.04)
Sex (Man)	0.91 (0.77 – 1.08)	0.70 *** (0.65 – 0.76)	1.37 *** (1.16 – 1.63)	1.62 *** (1.48 – 1.78)	0.89 * (0.81 – 0.98)	2.08 *** (1.77 – 2.44)
Age (standarized)	1.91 *** (1.74 – 2.09)	1.70 *** (1.63 – 1.77)	0.49 *** (0.44 – 0.56)	0.73 *** (0.69 – 0.76)	1.16 *** (1.10 – 1.22)	0.37 *** (0.32 – 0.42)
Education: Secondary school	1.14 (0.77 – 1.63)	0.56 *** (0.46 – 0.67)	0.53 *** (0.38 – 0.75)	0.88 (0.75 – 1.04)	2.47 *** (2.00 – 3.09)	1.24 (0.81 – 1.95)
Education: Vocational studies	0.93 (0.62 – 1.37)	0.53 *** (0.43 – 0.64)	0.63 * (0.45 – 0.90)	0.91 (0.76 – 1.09)	2.13 *** (1.68 – 2.71)	1.04 (0.67 – 1.67)
Education: University	1.23 (0.84 – 1.77)	0.29 *** (0.24 – 0.35)	0.25 *** (0.18 – 0.35)	0.49 *** (0.42 – 0.58)	3.24 *** (2.61 – 4.05)	1.00 (0.66 – 1.56)
Cycle infrastructure: Medium	1.01 (0.82 – 1.25)	1.09 (0.98 – 1.20)				
Cycle infrastructure: High	0.92 (0.72 – 1.16)	1.07 (0.95 – 1.21)				
Pedestrian infrastructure: Medium	1.06 (0.83 – 1.35)	0.96 (0.85 – 1.09)				
Pedestrian infrastructure: High	0.97 (0.75 – 1.25)	1.12 (0.98 – 1.29)				
Public transport: Medium	1.00 (0.82 – 1.22)					
Public transport: High	1.70 (0.75 – 4.89)					
Open public areas: Medium		1.03 (0.94 – 1.14)				
Open public areas: High		1.09 (0.94 – 1.26)				
Pedestrian facilities: Medium		0.83 ** (0.74 – 0.94)				
Pedestrian facilities: High		0.89 (0.64 – 1.24)				
Food environment: Medium			1.09 (0.91 – 1.30)			
Food environment: High			0.88 (0.52 – 1.39)			

Predictors	NAC	PHI	PRF	SMK	DRK	DRU
Economic activity: Medium				1.28 *** (1.16 – 1.43)	1.12 * (1.01 – 1.25)	1.54 *** (1.29 – 1.83)
Economic activity: High				1.21 (0.99 – 1.47)	1.37 *** (1.14 – 1.64)	1.79 *** (1.34 – 2.38)
Observations	10,541	13,194	7593	13,150	12,763	7458

* p<0.05   ** p<0.01   *** p<0.001

3. Rijeka

3.1. Classification of neighbourhoods

```
# Filter observations with valid values in column POP (total population) an
d City = Rijeka
data_RJK <- original_dataset %>%
  filter(!is.na(POP) & City == "Rijeka")

# Select variables of interest
variables_RJK <- data_RJK %>%
  select(IND.FOOD, IND.OPEN, IND.PED.INF, IND.PT, IND.ACCESS,
         IND.TOP, IND.PED.FAC, IND.INTERS, IND.ECO, POP_DENS, CONNECTIVITY,
         TRAFFIC_POP, GREEN.SURFACE)

# Scale variables
scaled_variables_RJK <- as.data.frame(scale(variables_RJK))

# Set seed for reproducibility
set.seed(123)

# Run k-means clustering
kmeans_result_RJK <- kmeans(scaled_variables_RJK, centers = 2, nstart = 25)
kmeans_result_RJK

## K-means clustering with 2 clusters of sizes 15, 19
##
## Cluster means:
##     IND.FOOD   IND.OPEN IND.PED.INF     IND.PT IND.ACCESS    IND.TOP
## 1 -0.2913692 -0.7932857  -0.8471660 -0.7830117 -0.7381034  0.3010146
## 2  0.2300283  0.6262782   0.6688152  0.6181672  0.5827132 -0.2376431
##   IND.PED.FAC IND.INTERS    IND.ECO   POP_DENS CONNECTIVITY TRAFFIC_POP
## 1  -0.4373052 -0.9263417 -0.3405952 -0.7708383   -0.7525656  -0.3267217
## 2   0.3452409  0.7313224  0.2688910  0.6085566    0.5941307   0.2579382
##   GREEN.SURFACE
## 1     0.5976025
## 2    -0.4717914
##
## Clustering vector:
##  [1] 2 2 2 1 2 2 1 1 2 1 1 1 1 2 2 2 2 1 1 2 1 2 2 2 2 1 1 1 1 2 1 2 2 2
```

```
##
## Within cluster sum of squares by cluster:
## [1]  89.45347 192.71553
##  (between_SS / total_SS =  34.2 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.wit
hinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
# Visualize clusters
fviz_cluster(kmeans_result_RJK, data = scaled_variables_RJK,
             geom = "point", stand = FALSE) +
  scale_colour_manual(values = c("#2ecc71", "#e74c3c")) +
  scale_fill_manual(values = c("#2ecc71", "#e74c3c"))
```

Fig. 5. K-means clusters in a biplot. Rijeka data

```
# Assign Clusters to a new data frame
variables_cluster_RJK <- variables_RJK %>%
  mutate(Cluster = kmeans_result_RJK$cluster)

# Summary statistics
st(variables_cluster_RJK,
   add.median = TRUE,
```

```
  group = 'Cluster',
  title = "**Tab. 9**. Summary statistics for clusters in Rijeka")
```

Tab. 9. Summary statistics for clusters in Rijeka

Variable	N	M	SD	Md	N	M	SD	Md
Cluster		**1**				**2**		
IND.FOOD	15	0.91	1.5	0.45	19	2.4	3.4	1.2
IND.OPEN	15	0.62	0.59	0.42	19	2.9	1.4	2.6
IND.PED.INF	15	0.23	0.12	0.24	19	0.46	0.081	0.46
IND.PT	15	1.9	0.85	1.9	19	4.5	1.6	4.3
IND.ACCESS	15	0.19	0.082	0.21	19	0.33	0.08	0.32
IND.TOP	15	10	3.4	9.1	19	8.8	2.1	9.2
IND.PED.FAC	15	0.94	0.56	0.74	19	6.5	8.8	3.5
IND.INTERS	15	271	188	239	19	907	233	824
IND.ECO	15	2.4	2.4	1.6	19	19	35	6.4
POP_DENS	15	17	14	14	19	67	33	59
CONNECTIVITY	15	1.4	2.1	0.78	19	15	10	14
TRAFFIC_POP	15	31	19	21	19	42	19	42
GREEN.SURFACE	15	1017	1206	398	19	19	16	15

3.2. Mapping clusters

```
# Assign Cluster to data
data_RJK$Cluster <- as.factor(kmeans_result_RJK$cluster)

# Do the left join
RJK_geo_sep <- left_join(shape_RJK, data_RJK, by = c("City", "number"))

# Define colors for clusters
cluster_colors_RJK <- c("1" = "#2ecc71", "2" = "#e74c3c", "lightgray")

# Plot the results: Rijeka
ggplot(data = RJK_geo_sep) +
  geom_sf(aes(fill = Cluster), color = "white") +
  scale_fill_manual(values = cluster_colors_RJK, na.value = "lightgray", na
me = "Cluster") +
  theme_minimal() +
  labs(fill = "Cluster")
```

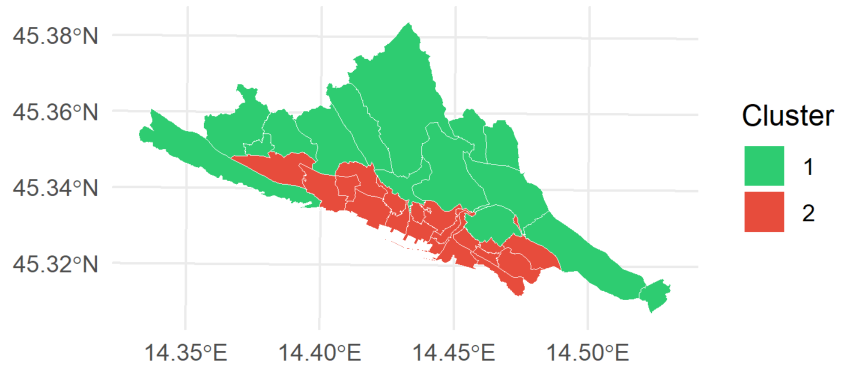

Fig. 6. K-means clusters map from Rijeka data analysis

3.3. Estimation of associations

```
# Import data
survey_data_RJK <- read_excel("local/path/survey_data_RJK.xlsx")
```

```
# Print first rows
head(survey_data_RJK)

## # A tibble: 6 x 15
##   ID    city   Neighbourhood Sex     Age Education act.commuting
##   <chr> <chr>  <chr>         <chr> <dbl> <chr>             <dbl>
## 1 20    Rijeka 2             1        35 4                     0
## 2 21    Rijeka 8             1        60 4                     0
## 3 22    Rijeka 1             1        45 4                     0
## 4 23    Rijeka 9             0        32 4                     1
## 5 25    Rijeka 1             1        55 4                     0
## 6 26    Rijeka 1             <NA>     45 <NA>                  0
## # i 8 more variables: physical.activity <dbl>, Sleep <dbl>, diet <dbl>,
## #   smoking <dbl>, drinking <dbl>, overweight <dbl>, CVD <dbl>, Diabetes
 <dbl>

# Prepare the data for doing the left join
data_to_join_RJK <- RJK_geo_sep %>%
  select(number, Cluster)

# Left join
RJK_data_model <- left_join(survey_data_RJK, data_to_join_RJK, by = c("Neig
hbourhood" = "number"))
```

```
# Set the subset for further model comparison
valid_rows_RJK <- complete.cases(RJK_data_model[, c("Sex", "Age", "Educatio
```

```r
n", "Cluster")])

# Models for socio-demographics
m.com.rjk.0 <- glm(act.commuting ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.pa.rjk.0 <- glm(physical.activity ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.sl.rjk.0 <- glm(Sleep ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.di.rjk.0 <- glm(diet ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.sm.rjk.0 <- glm(smoking ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.dr.rjk.0 <- glm(drinking ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.bmi.rjk.0 <- glm(overweight ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.cvd.rjk.0 <- glm(CVD ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.db.rjk.0 <- glm(Diabetes ~ Sex + scale(Age) + Education, data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)

# Visualization of results
tab_model(m.com.rjk.0, m.pa.rjk.0, m.sl.rjk.0, m.di.rjk.0,
          m.sm.rjk.0, m.dr.rjk.0, m.bmi.rjk.0, m.cvd.rjk.0, m.db.rjk.0,
          collapse.ci = TRUE,
          p.style = "stars",
          show.r2 = FALSE,
          pred.labels = c("Intercept", "Sex (Man)","Age (standarized)",
                          "Education: Secondary school",
                          "Education: Vocational studies",
                          "Education: University"),
          dv.labels = c("NAC", "PHI", "LSL", "PRF",
                        "SMK", "DRK", "OWT", "CVD", "DBT"),
          title = "**Tab 10**. ORs for socio-demographics. Rijeka")
```

Tab 10. ORs for socio-demographics. Rijeka

Predictors	NAC	PHI	LSL	PRF	SMK	DRK	OWT	CVD	DBT
Intercept	0.83 (0.42 – 1.60)	1.17 (0.60 – 2.30)	0.13 *** (0.05 – 0.30)	1.12 (0.57 – 2.26)	0.60 (0.28 – 1.21)	0.08 *** (0.02 – 0.24)	0.86 (0.43 – 1.79)	0.01 *** (0.00 – 0.03)	0.04 *** (0.01 – 0.10)
Sex (Man)	1.20 * (1.01 – 1.42)	0.79 ** (0.66 – 0.94)	1.04 (0.73 – 1.46)	2.48 *** (2.08 – 2.97)	0.89 (0.72 – 1.09)	1.75 *** (1.40 – 2.18)	2.57 *** (2.14 – 3.09)	3.69 *** (2.35 – 5.87)	1.69 * (1.10 – 2.58)
Age (standarized)	0.90 * (0.83 – 0.98)	1.03 (0.95 – 1.12)	1.52 *** (1.30 – 1.79)	0.82 *** (0.76 – 0.90)	0.72 *** (0.65 – 0.79)	0.81 *** (0.72 – 0.91)	1.66 *** (1.52 – 1.81)	5.08 *** (3.84 – 6.86)	2.95 *** (2.35 – 3.75)
Education: Secondary school	0.81 (0.41 – 1.62)	0.83 (0.42 – 1.64)	0.62 (0.26 – 1.75)	0.44 * (0.21 – 0.88)	0.68 (0.33 – 1.48)	1.50 (0.51 – 6.41)	0.99 (0.47 – 2.03)	1.02 (0.34 – 3.80)	0.65 (0.24 – 2.08)

Predictors	NAC	PHI	LSL	PRF	SMK	DRK	OWT	CVD	DBT
Education: Vocational studies	1.04 (0.53 – 2.09)	0.81 (0.41 – 1.59)	0.55 (0.23 – 1.54)	0.42 * (0.21 – 0.85)	0.70 (0.34 – 1.52)	1.89 (0.65 – 8.04)	0.86 (0.40 – 1.75)	0.79 (0.27 – 2.92)	0.59 (0.22 – 1.87)
Education: University	1.17 (0.60 – 2.32)	0.56 (0.29 – 1.10)	0.41 (0.17 – 1.16)	0.38 ** (0.19 – 0.76)	0.45 * (0.22 – 0.98)	2.19 (0.77 – 9.24)	0.62 (0.29 – 1.25)	0.63 (0.21 – 2.34)	0.48 (0.18 – 1.54)
Observations	2447	2447	2447	2446	2447	2446	2447	2447	2447

* p<0.05 ** p<0.01 *** p<0.001

```
# Adjusted models
m.com.rjk.1 <- glm(act.commuting ~ Cluster + Sex + scale(Age) + Education, 
data = RJK_data_model, family = binomial(link='logit'), subset = valid_rows
_RJK)
m.pa.rjk.1 <- glm(physical.activity ~ Cluster + Sex + scale(Age) + Educatio
n, data = RJK_data_model, family = binomial(link='logit'), subset = valid_r
ows_RJK)
m.sl.rjk.1 <- glm(Sleep ~ Cluster + Sex + scale(Age) + Education, data = RJ
K_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.di.rjk.1 <- glm(diet ~ Cluster + Sex + scale(Age) + Education, data = RJK
_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.sm.rjk.1 <- glm(smoking ~ Cluster + Sex + scale(Age) + Education, data = 
RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.dr.rjk.1 <- glm(drinking ~ Cluster + Sex + scale(Age) + Education, data =
 RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.bmi.rjk.1 <- glm(overweight ~ Cluster + Sex + scale(Age) + Education, dat
a = RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJ
K)
m.cvd.rjk.1 <- glm(CVD ~ Cluster + Sex + scale(Age) + Education, data = RJK
_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)
m.db.rjk.1 <- glm(Diabetes ~ Cluster + Sex + scale(Age) + Education, data =
 RJK_data_model, family = binomial(link='logit'), subset = valid_rows_RJK)

# Visualization of results
tab_model(m.com.rjk.1, m.pa.rjk.1, m.sl.rjk.1, m.di.rjk.1,
          m.sm.rjk.1, m.dr.rjk.1, m.bmi.rjk.1, m.cvd.rjk.1, m.db.rjk.1,
          collapse.ci = TRUE,
          p.style = "stars",
          auto.label = TRUE,
          show.r2 = FALSE,
          pred.labels = c("Intercept", "Cluster: 2", "Sex (Man)",
                          "Age (standarized)",
                          "Education: Secondary school",
                          "Education: Vocational studies",
                          "Education: University"),
          dv.labels = c("NAC", "PHI", "LSL", "PRF",
                        "SMK", "DRK", "OWT", "CVD", "DBT"),
          title = "**Tab 11**. Adjusted ORs in Rijeka")
```

Tab 11. Adjusted ORs in Rijeka

Predictors	**NAC**	**PHI**	**LSL**	**PRF**	**SMK**	**DRK**	**OWT**	**CVD**	**DBT**
Intercept	0.93 (0.47 – 1.81)	1.25 (0.64 – 2.45)	0.15 *** (0.05 – 0.35)	1.26 (0.63 – 2.55)	0.57 (0.27 – 1.16)	0.09 *** (0.02 – 0.24)	0.93 (0.46 – 1.96)	0.01 *** (0.00 – 0.04)	0.04 *** (0.01 – 0.11)
Cluster: 2	0.78 ** (0.66 – 0.92)	0.88 (0.74 – 1.04)	0.70 * (0.50 – 0.97)	0.77 ** (0.65 – 0.92)	1.11 (0.91 – 1.35)	0.98 (0.79 – 1.23)	0.82 * (0.69 – 0.98)	0.76 (0.48 – 1.19)	0.89 (0.58 – 1.37)
Sex (Man)	1.19 * (1.00 – 1.42)	0.79 ** (0.66 – 0.94)	1.03 (0.72 – 1.45)	2.49 *** (2.08 – 2.97)	0.89 (0.72 – 1.09)	1.75 *** (1.40 – 2.18)	2.57 *** (2.14 – 3.09)	3.61 *** (2.29 – 5.74)	1.68 * (1.09 – 2.56)
Age (standarized)	0.90 ** (0.83 – 0.97)	1.03 (0.95 – 1.12)	1.51 *** (1.29 – 1.78)	0.82 *** (0.75 – 0.89)	0.72 *** (0.65 – 0.79)	0.81 *** (0.72 – 0.91)	1.66 *** (1.52 – 1.81)	4.96 *** (3.75 – 6.72)	2.93 *** (2.34 – 3.72)
Education: Secondary school	0.84 (0.42 – 1.69)	0.84 (0.42 – 1.67)	0.65 (0.27 – 1.84)	0.45 * (0.22 – 0.92)	0.66 (0.32 – 1.46)	1.51 (0.52 – 6.43)	1.03 (0.48 – 2.11)	1.04 (0.35 – 3.87)	0.66 (0.24 – 2.11)
Education: Vocational studies	1.09 (0.55 – 2.19)	0.82 (0.41 – 1.63)	0.58 (0.24 – 1.64)	0.44 * (0.21 – 0.89)	0.68 (0.33 – 1.50)	1.90 (0.65 – 8.06)	0.89 (0.42 – 1.82)	0.82 (0.28 – 3.05)	0.60 (0.22 – 1.90)
Education: University	1.23 (0.63 – 2.45)	0.58 (0.29 – 1.13)	0.44 (0.19 – 1.24)	0.40 * (0.20 – 0.80)	0.44 * (0.21 – 0.95)	2.20 (0.77 – 9.28)	0.65 (0.31 – 1.31)	0.65 (0.22 – 2.41)	0.49 (0.18 – 1.57)
Observations	2447	2447	2447	2446	2447	2446	2447	2447	2447

* p<0.05   ** p<0.01   *** p<0.001

```r
anova(m.com.rjk.0, m.com.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: act.commuting ~ Sex + scale(Age) + Education
## Model 2: act.commuting ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2441     3363.9
## 2      2440     3355.2  1   8.6966 0.003188 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.pa.rjk.0, m.pa.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: physical.activity ~ Sex + scale(Age) + Education
## Model 2: physical.activity ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2441     3310.6
## 2      2440     3308.4  1   2.2172   0.1365
```

```r
anova(m.sl.rjk.0, m.sl.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: Sleep ~ Sex + scale(Age) + Education
## Model 2: Sleep ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2441     1155.0
## 2      2440     1150.4  1   4.6187  0.03162 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.di.rjk.0, m.di.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: diet ~ Sex + scale(Age) + Education
## Model 2: diet ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2440     3123.9
## 2      2439     3115.6  1   8.3111  0.00394 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.sm.rjk.0, m.sm.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: smoking ~ Sex + scale(Age) + Education
## Model 2: smoking ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2441     2686.6
## 2      2440     2685.5  1   1.1085   0.2924

anova(m.dr.rjk.0, m.dr.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: drinking ~ Sex + scale(Age) + Education
## Model 2: drinking ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2440     2158.4
## 2      2439     2158.3  1 0.023897   0.8771

anova(m.bmi.rjk.0, m.bmi.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: overweight ~ Sex + scale(Age) + Education
## Model 2: overweight ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2441     3107.9
## 2      2440     3102.9  1   5.0127  0.02516 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(m.cvd.rjk.0, m.cvd.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: CVD ~ Sex + scale(Age) + Education
## Model 2: CVD ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2441     588.13
## 2      2440     586.68  1   1.4498   0.2286

anova(m.db.rjk.0, m.db.rjk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: Diabetes ~ Sex + scale(Age) + Education
## Model 2: Diabetes ~ Cluster + Sex + scale(Age) + Education
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2441     723.06
## 2      2440     722.76  1  0.29354    0.588
```

3.4. Specification tests

```
# Variables to classify
variables_to_classify_RJK <- c("IND.FOOD", "IND.OPEN", "IND.PED.INF", "IND.
PT", "IND.ACCESS", "IND.TOP", "IND.PED.FAC", "IND.INTERS", "IND.ECO", "POP_
DENS", "CONNECTIVITY", "TRAFFIC_POP", "GREEN.SURFACE")

# Classify the variables and create new columns with suffix 'pb'.
for (var in variables_to_classify_RJK) {
  breaks <- classIntervals(data_RJK[[var]], n = 3, style = "jenks")$brks
  data_RJK[[paste0(var, ".pb")]] <- cut(data_RJK[[var]], breaks = breaks, l
abels = c("Low", "Medium", "High"), include.lowest = TRUE)
}

# Select only the new columns generated for the join
columns_for_join_RJK <- paste0(variables_to_classify_RJK, ".pb")

# Perform left join with 'RJK_data_model' based on the variable 'Neighbourh
ood'.
RJK_sp_data_model <- RJK_data_model %>%
  left_join(data_RJK[, c("number", columns_for_join_RJK)], by = c("Neighbou
rhood" = "number"))

# Specification models
sp.m.com.rjk <- glm(act.commuting ~ Sex + scale(Age) + Education + IND.PED.
INF.pb + IND.PT.pb, data = RJK_sp_data_model, family = binomial(link='logit
'))
sp.m.pa.rjk <- glm(physical.activity ~ Sex + scale(Age) + Education + IND.O
PEN.pb + IND.PED.INF.pb + IND.PED.FAC.pb, data = RJK_sp_data_model, family
= binomial(link='logit'))
```

```r
sp.m.sl.rjk <- glm(Sleep ~ Sex + scale(Age) + Education + TRAFFIC_POP.pb +
IND.ECO.pb + POP_DENS.pb, data = RJK_sp_data_model, family = binomial(link=
'logit'))
sp.m.di.rjk <- glm(diet ~ Sex + scale(Age) + Education + IND.FOOD.pb, data
= RJK_sp_data_model, family = binomial(link='logit'))
sp.m.sm.rjk <- glm(smoking ~ Sex + scale(Age) + Education + IND.ECO.pb, dat
a = RJK_sp_data_model, family = binomial(link='logit'))
sp.m.dr.rjk <- glm(drinking ~ Sex + scale(Age) + Education + IND.ECO.pb, da
ta = RJK_sp_data_model, family = binomial(link='logit'))

# Visualization of results
tab_model(sp.m.com.rjk, sp.m.pa.rjk, sp.m.sl.rjk,
          sp.m.di.rjk, sp.m.sm.rjk, sp.m.dr.rjk,
          collapse.ci = TRUE,
          p.style = "stars",
          auto.label = TRUE,
          show.r2 = FALSE,
          pred.labels = c("Intercept", "Sex (Man)", "Age (standarized)",
                          "Education: Secondary school",
                          "Education: Vocational studies",
                          "Education: University",
                          "Pedestrian infrastructure: Medium",
                          "Pedestrian infrastructure: High",
                          "Public transport: Medium",
                          "Public transport: High",
                          "Open public areas: Medium",
                          "Open public areas: High",
                          "Pedestrian facilities: Medium",
                          "Pedestrian facilities: High",
                          "Traffic exposure: Medium",
                          "Traffic exposure: High",
                          "Economic activity: Medium",
                          "Economic activity: High",
                          "Population density: Medium",
                          "Population density: High",
                          "Food environment: Medium",
                          "Food environment: High"),
          dv.labels = c("NAC", "PHI", "LSL",
                        "PRF", "SMK", "DRK"),
          title = "**Tab 12**. Adjusted ORs for specification variables in
Rijeka")
```

Tab 12. Adjusted ORs for specification variables in Rijeka

Predictors	NAC	PHI	LSL	PRF	SMK	DRK
Intercept	1.22 (0.57 – 2.60)	1.29 (0.61 – 2.74)	0.17 *** (0.06 – 0.41)	1.12 (0.56 – 2.25)	0.60 (0.28 – 1.21)	0.08 *** (0.02 – 0.24)
Sex (Man)	1.21 * (1.01 – 1.43)	0.79 * (0.66 – 0.95)	1.01 (0.71 – 1.42)	2.49 *** (2.08 – 2.98)	0.89 (0.72 – 1.09)	1.75 *** (1.40 – 2.18)

Predictors	NAC	PHI	LSL	PRF	SMK	DRK
Age (standarized)	0.90 ** (0.83 – 0.97)	1.04 (0.96 – 1.13)	1.51 *** (1.29 – 1.77)	0.82 *** (0.76 – 0.90)	0.72 *** (0.65 – 0.79)	0.81 *** (0.72 – 0.91)
Education: Secondary school	0.86 (0.43 – 1.74)	0.86 (0.43 – 1.70)	0.64 (0.26 – 1.82)	0.44 * (0.22 – 0.90)	0.68 (0.33 – 1.48)	1.51 (0.52 – 6.44)
Education: Vocational studies	1.12 (0.56 – 2.25)	0.83 (0.42 – 1.64)	0.58 (0.24 – 1.63)	0.43 * (0.21 – 0.87)	0.69 (0.34 – 1.51)	1.90 (0.65 – 8.06)
Education: University	1.27 (0.65 – 2.54)	0.59 (0.30 – 1.15)	0.44 (0.18 – 1.24)	0.39 ** (0.19 – 0.78)	0.45 * (0.22 – 0.97)	2.21 (0.77 – 9.32)
Pedestrian infrastructure: Medium	0.71 (0.47 – 1.06)	0.95 (0.64 – 1.42)				
Pedestrian infrastructure: High	0.69 (0.45 – 1.05)	0.86 (0.57 – 1.30)				
Public transport: Medium	0.84 (0.69 – 1.03)					
Public transport: High	0.77 (0.53 – 1.11)					
Open public areas: Medium		0.97 (0.81 – 1.16)				
Open public areas: High		0.88 (0.61 – 1.27)				
Pedestrian facilities: Medium		1.02 (0.71 – 1.47)				
Pedestrian facilities: High		0.36 (0.05 – 1.41)				
Traffic exposure: Medium			0.90 (0.59 – 1.39)			
Traffic exposure: High			0.64 (0.25 – 1.45)			
Economic activity: Medium			1.13 (0.50 – 2.29)		1.15 (0.77 – 1.68)	0.95 (0.58 – 1.48)
Economic activity: High			2.35 (0.12 – 15.81)		0.73 (0.11 – 2.84)	0.41 (0.02 – 2.15)
Population density: Medium			0.65 * (0.43 – 0.97)			
Population density: High			0.57 (0.31 – 1.01)			
Food environment: Medium				0.67 (0.42 – 1.04)		
Food environment: High				0.50 (0.11 – 1.73)		

Predictors	NAC	PHI	LSL	PRF	SMK	DRK
Observations	2447	2447	2447	2446	2447	2446

* p<0.05 ** p<0.01 *** p<0.001

ANNEX 2. SPATIAL DISTRIBUTION OF NCD RISK BEHAVIOURS AND OUTCOMES

Natural breaks maps (Jenks) plotting the spatial distribution of the behavioural and health outcomes of interest, based on the survey data in Rijeka, are presented below:

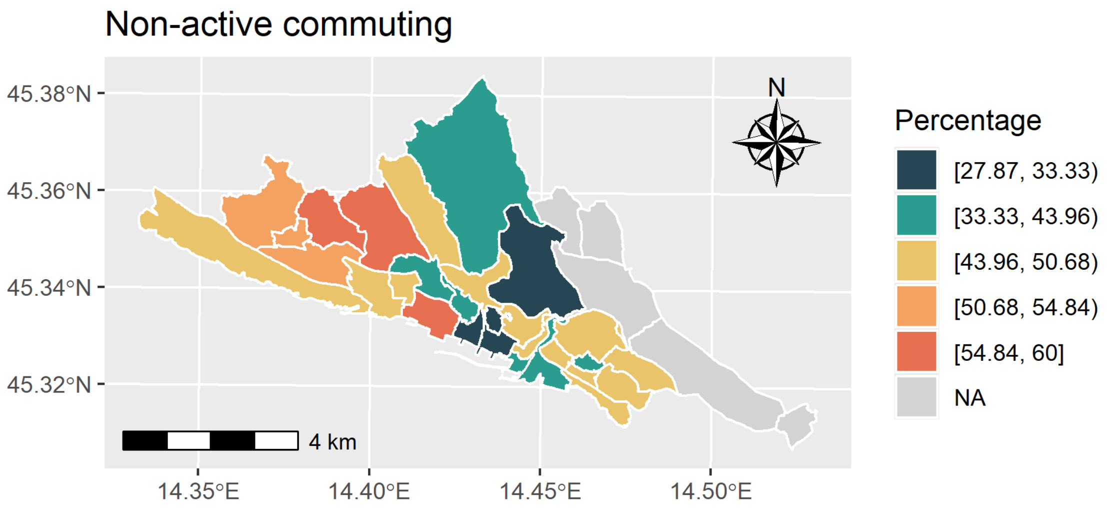

Figure 7. Percentage of respondents who do not regularly spend at least 30 minutes for 5 days or more per week on daily commuting, walking or cycling; or at least 60 minutes for 3 or 4 days per week (**NAC**).

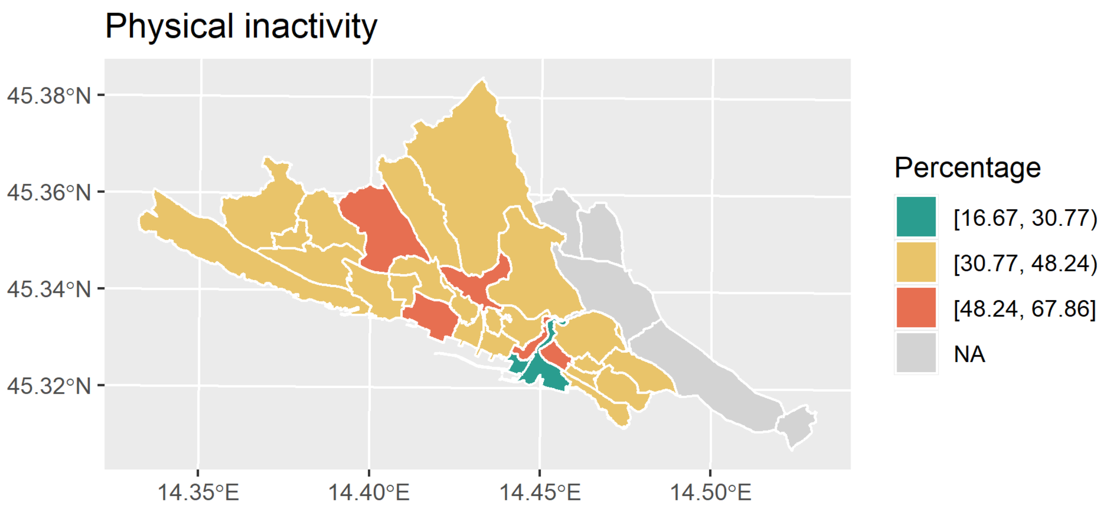

Figure 8. Percentage of respondents who do not normally spend at least 30 minutes for 3 days or more per week exercising; or at least 60 minutes for 2 days per week (**PHI**).

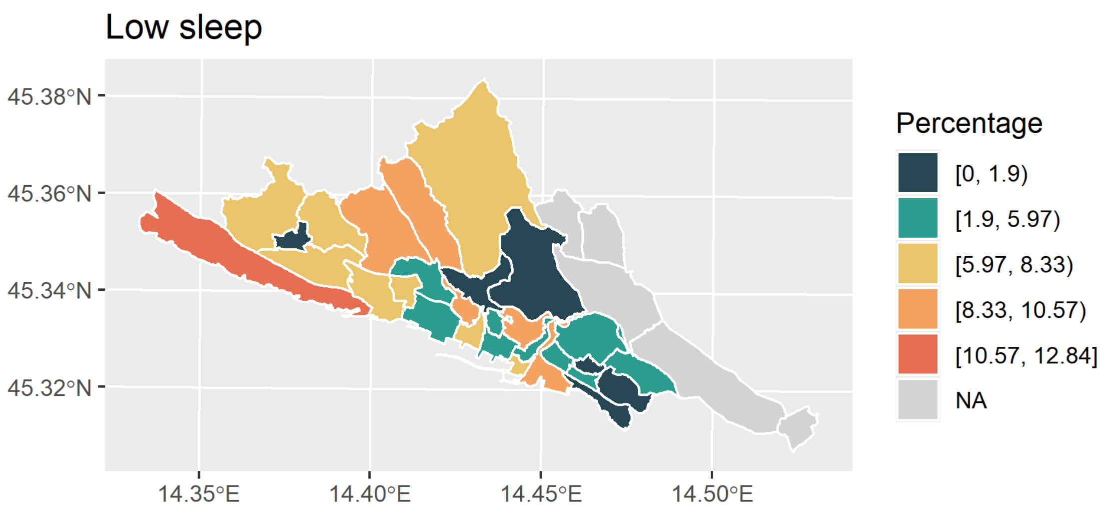

Figure 9. Percentage of respondents who usually sleep less than 6 hours a day (**LSL**).

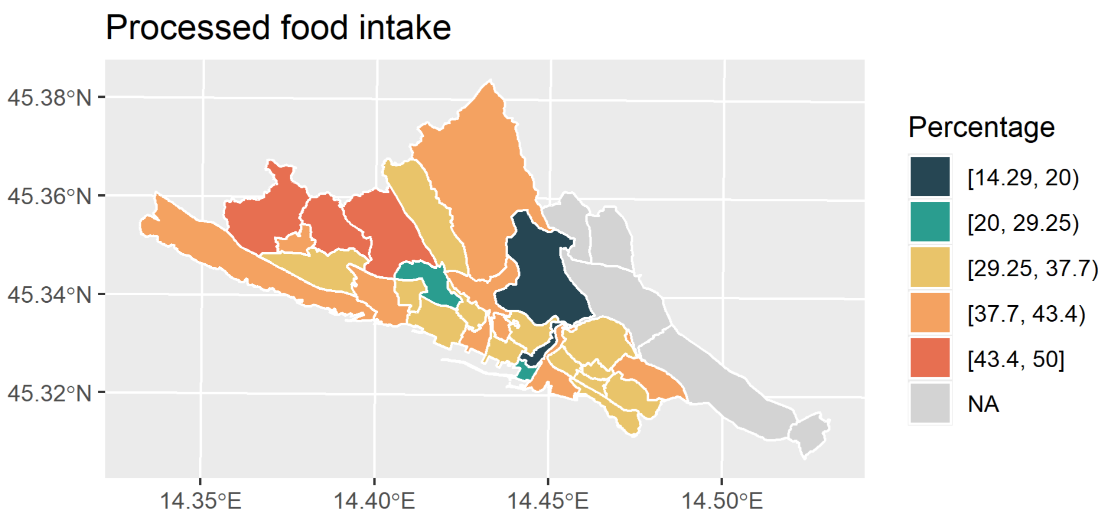

Figure 10. Percentage of respondents who usually eat ultra-processed food at least 3 times a week (**PRF**).

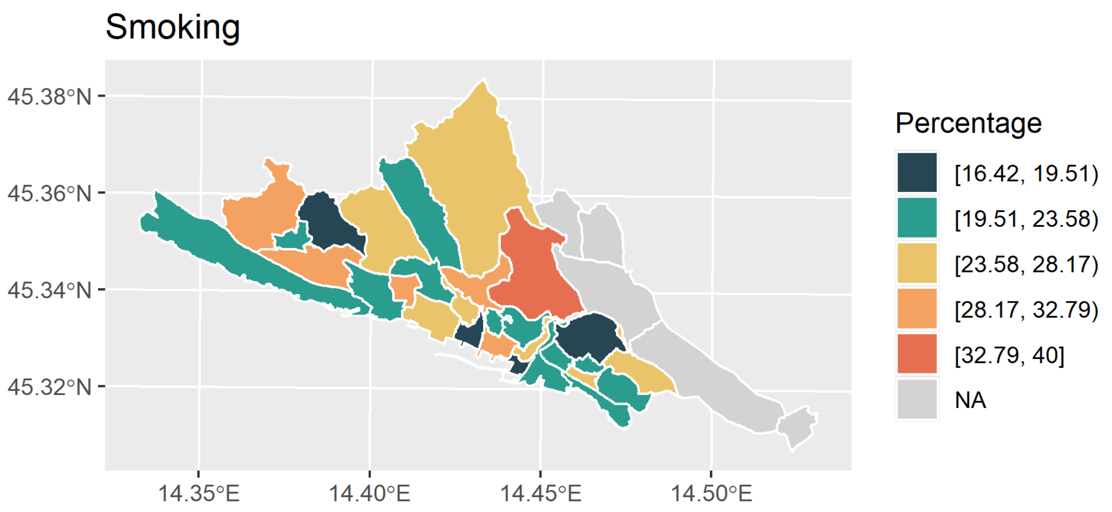

Figure 11. Percentage of respondents who report currently smoking (**SMK**).

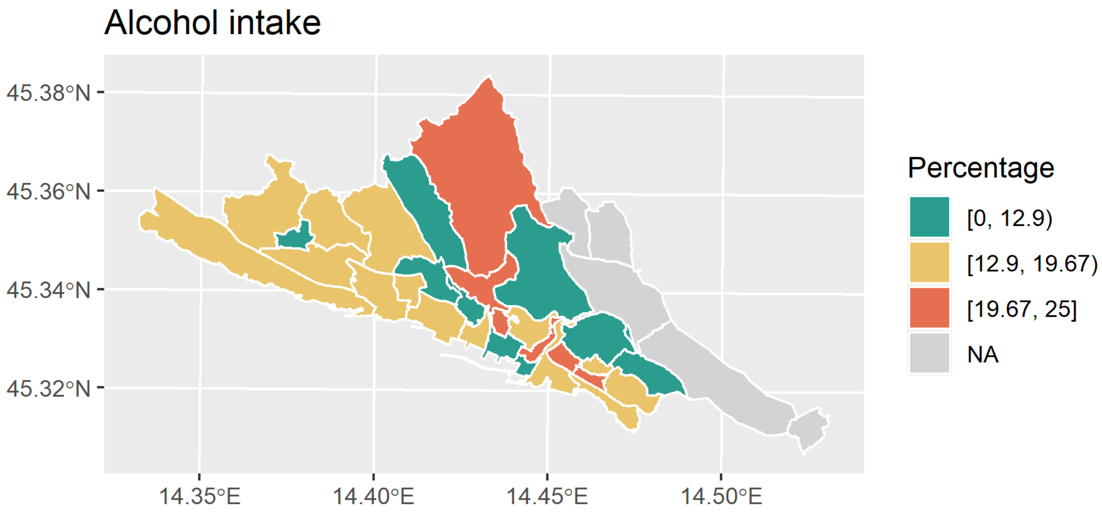

Figure 12. Respondents reporting alcohol consumption above the low-risk threshold limit: more than 20 g/day for men (2 standard drinks) or more than 10 g/day for women (1 standard drink) (**DRK**).

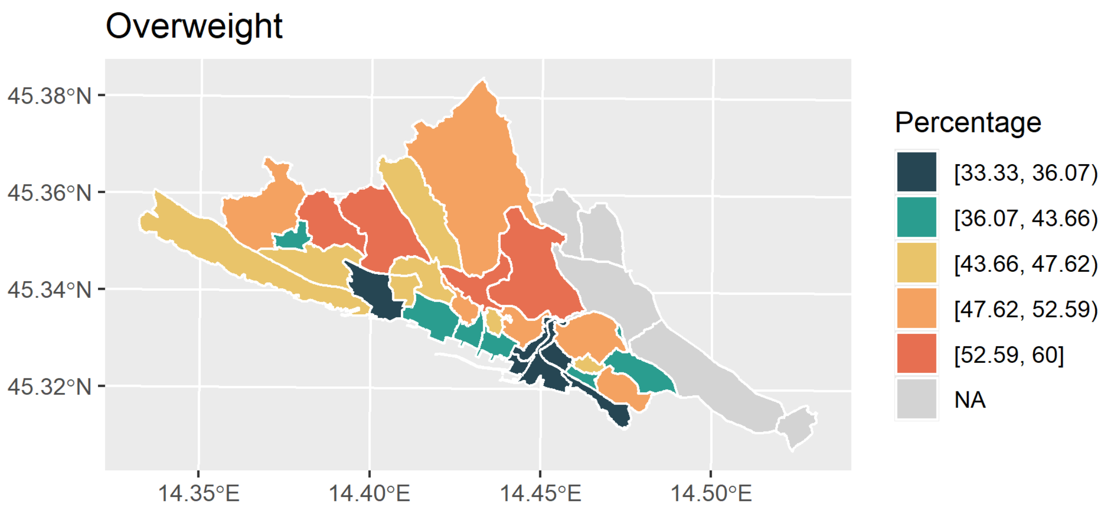

Figure 13. Percentage of respondents with BMI > 25 (**OWT**).

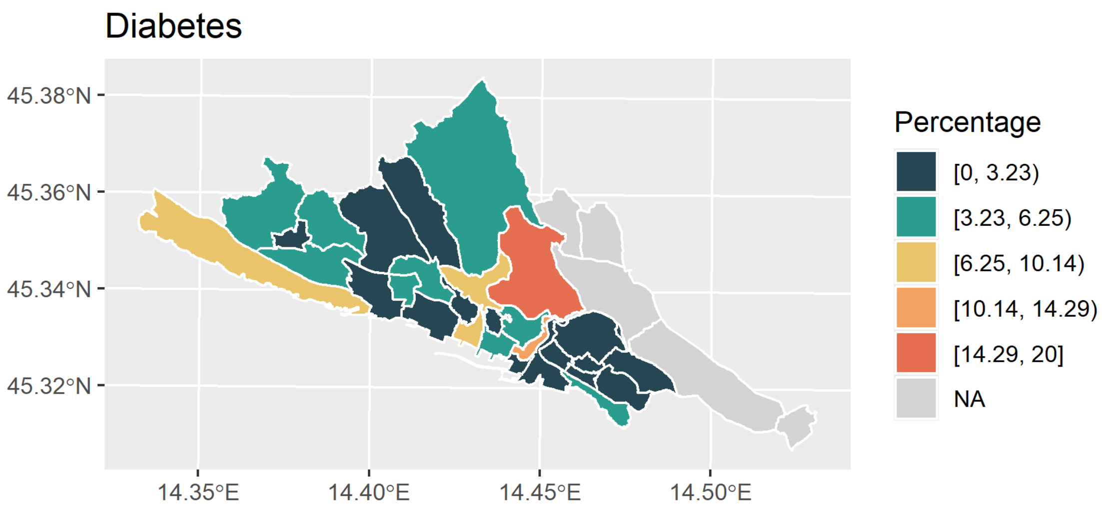

Figure 14. Percentage of respondents with diagnosed diabetes (**DBT**).

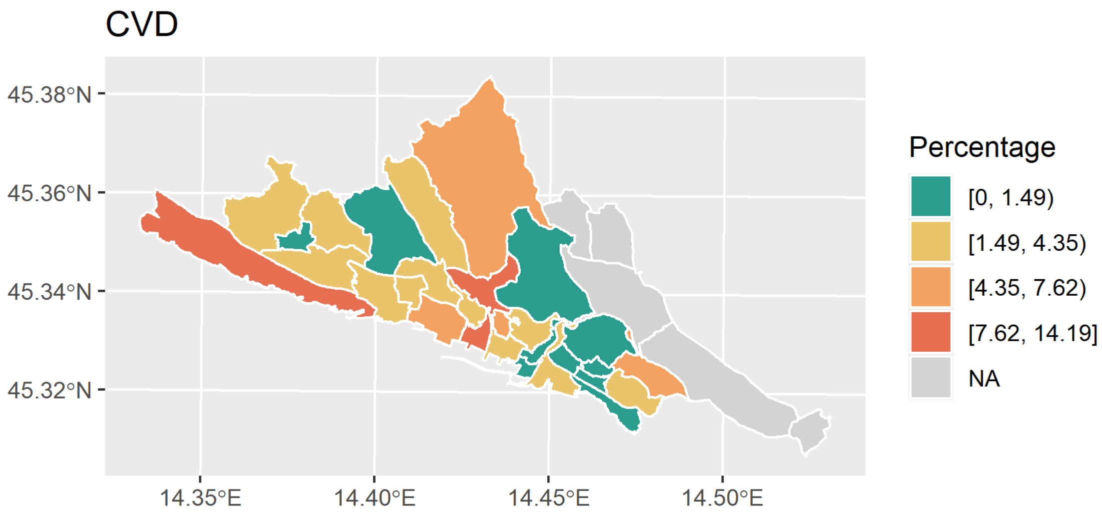

Figure 15. Percentage of respondents with diagnosed CVD (**CVD**).