Compositional data analysis for migration studies

Authors: Javier Elío[1], Marina Georgati[2], Henning S. Hansen[3], Carsten Keßler[4].

Corresponding author: Carsten Keßler

1 Prerequisites

The data have been analysed with **R** (version 4.1.0) and **Rstudio** (version 1.4.1717). The required packages are automatically checked and installed if needed from CRAN.

Furthermore, we have created our own package for downloading kortforsyningen data to a local repository directly from R (i.e. dangeo). The package can be downloaded from GitHub:

Although the *kortforsyningen* data are free, we would need to create a *username* and a *password* for getting access to them (you can make it here: "Opret ny bruger"). By default **dangeo** looks for credentials on `.Renviron` as: `kortforsyningen_id = "username"` and `kortforsyningen_pwd = "password"`. You would need to save them with `usethis::edit_r_environ()`:

You would also need to define with `dangeo_set_param()` the local directory where the data are downloaded (`loc_dir`). It is defined as `loc_dir = rappdirs::user_cache_dir()`, although it can be changed `loc_dir = "./your/local/path"`. The first time a file is downloaded with `dangeo_get_data()`, the process can be time consuming (there are some very big files). However, it will not be downloaded in subsequent calls if the files is already in the local directory (the dataset can be overwritten be setting `overwrite = TRUE` on `dangeo_get_data()`). Once we have our *username* and *password*, and we have define the local repository for the data, we can set them on our R-session:

Individual house prices data (from BBR) are not free, and we have save here the summary statistics by parish (i.e. median value).

[1]Department of Planning, Aalborg University, Copenhagen, Denmark; ORCID: 0000-0003-0624-2345; javierdem@plan.aau.dk - javiereliomedina@gmail.com

[2]Department of Planning, Aalborg University, Copenhagen, Denmark; ORCID: 0000-0001-7794-8308; marinag@plan.aau.dk

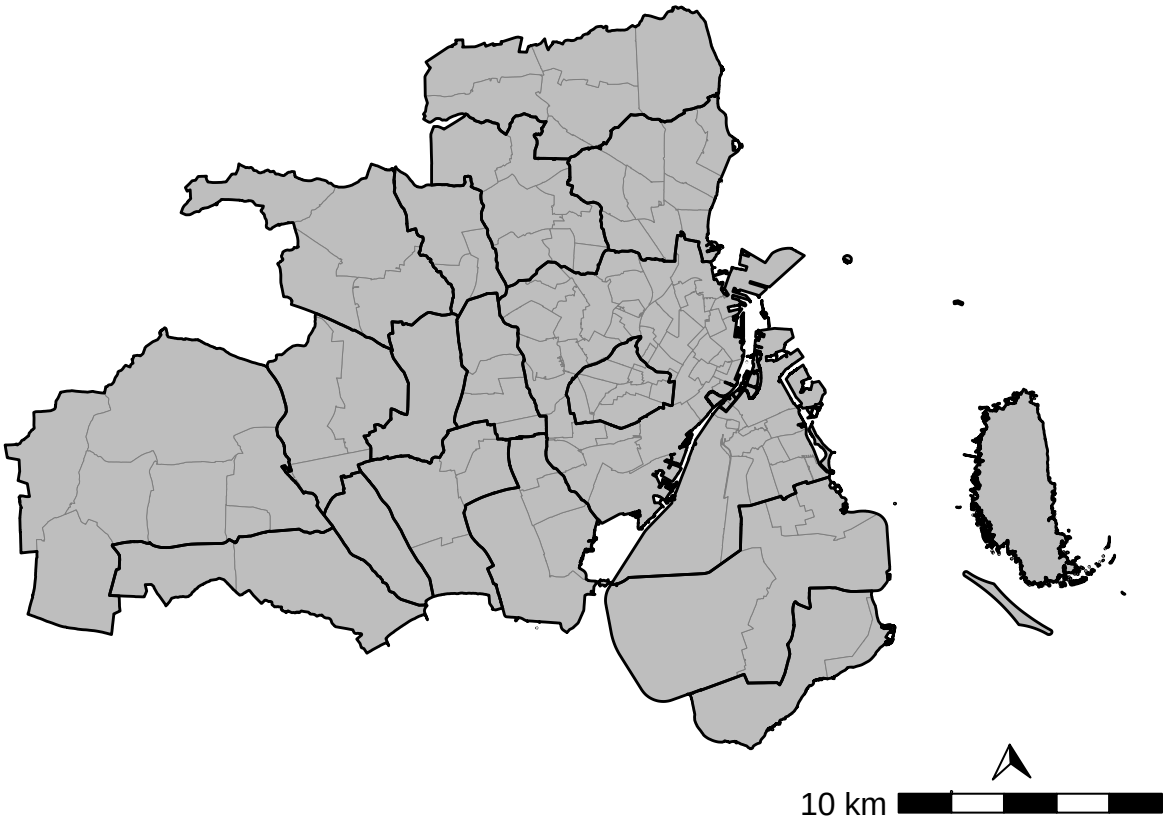
[3]Department of Planning, Aalborg University, Copenhagen, Denmark; ORCID: 0000-0001-7004-0698; hsh@plan.aau.dk

[4]Department of Planning, Aalborg University, Copenhagen, Denmark; ORCID: 0000-0002-9724-820X; kessler@plan.aau.dk

Set ggplot theme.

2 Data

2.1 Spatial data



2.2 Population

Population data in 2020 were uploaded from Denmark Statistics:

- KMSTA001: Population 1. January by parish, ancestry and National Church.

2.3 House prices

House prices in 2020 from BBR.

3 Analysis

3.1 Population

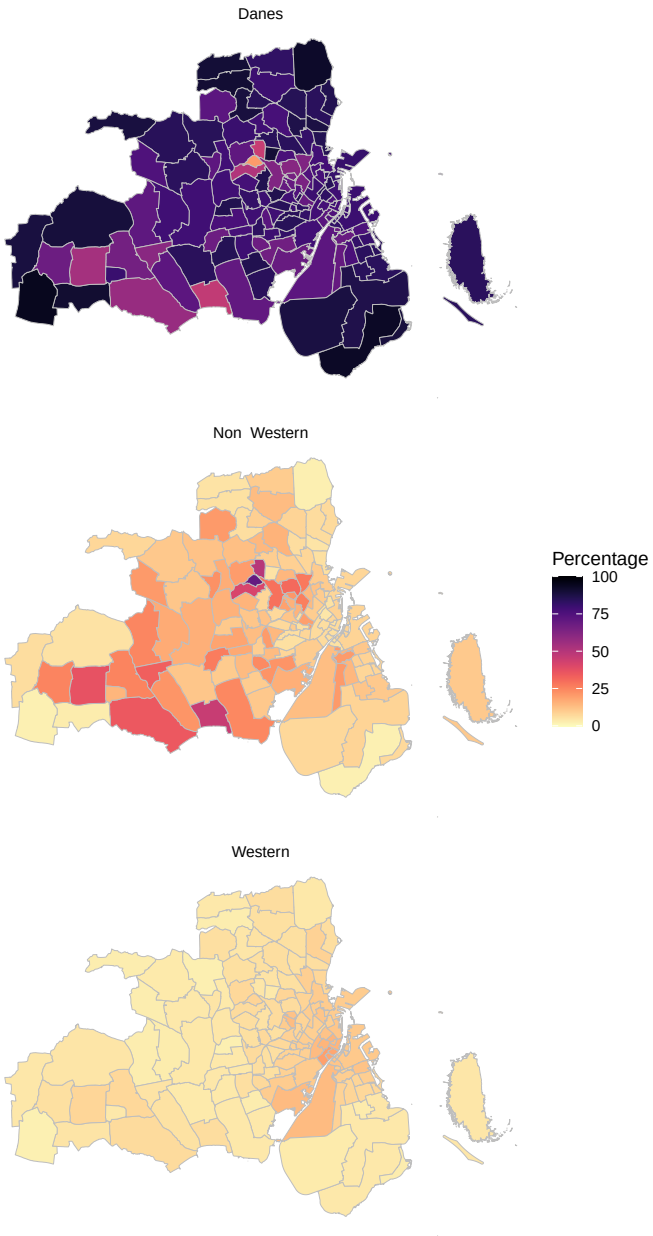


Figure 1: Population distribution

3.2 House prices

The total number of residential units used for the analysis is therefore 18764 (Table **??**). The summary descriptive statistics of the housing prices are:

Characteristic	**Overall**, N = 18,764	**Multi-storey**, N = 12,628
___House prices (kDKK)___		
Mean	8,875	10,762
Median	3,700	3,400
IQR	2,425 - 5,595	2,100 - 5,600
Range	0 - 460,000	0 - 460,000
___Dwelling size (m2)___		
Mean	100	83
Median	91	77
IQR	68 - 122	59 - 98
Range	11 - 857	11 - 857
___Prices per square meter (kDKK/m2)___		
Mean	109	143
Median	39	42
IQR	30 - 52	33 - 55
Range	0 - 5,679	0 - 5,679

House prices by type (removing very low prices; i.e. <1 kDKK/m2, n = 242).

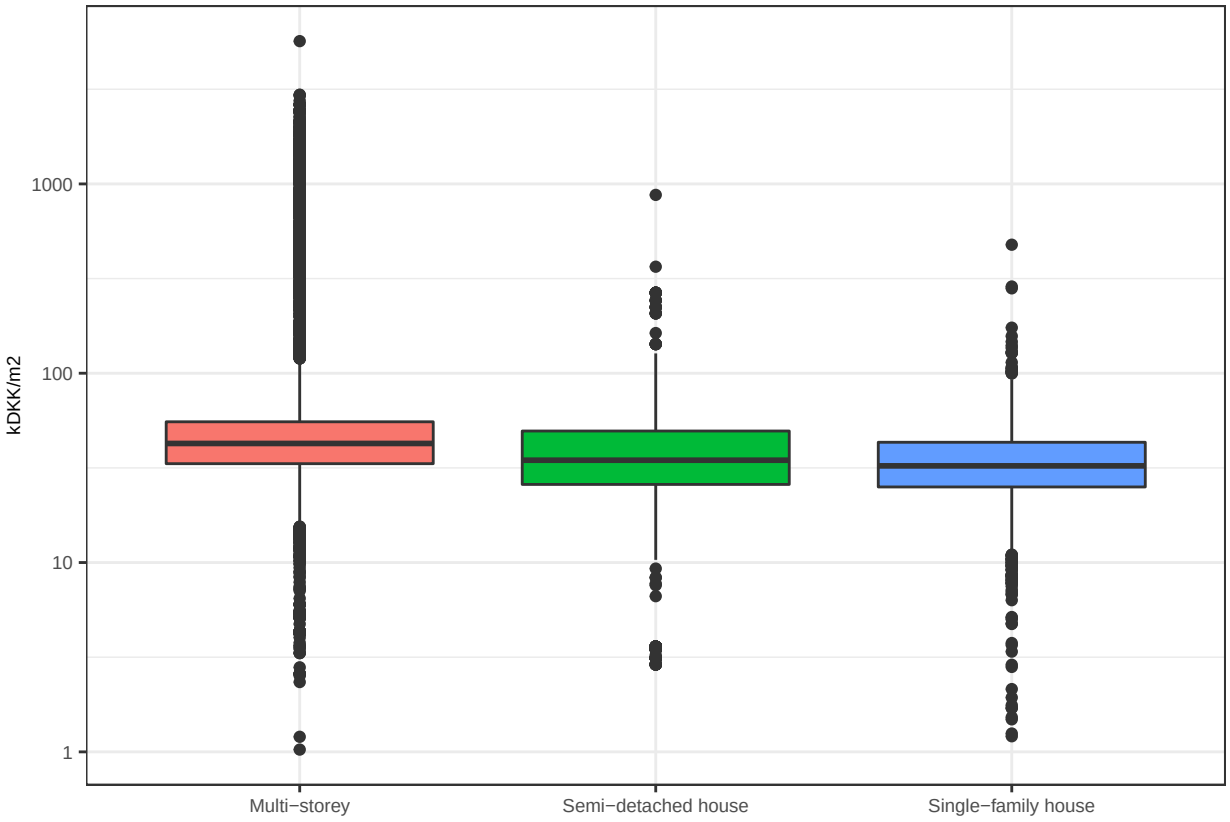


Figure 2: Boxplot of residential units in the open free trade by house typr (values have been truncated with house prices >1 kDkk/m2)

House prices parish (spatial distribution)

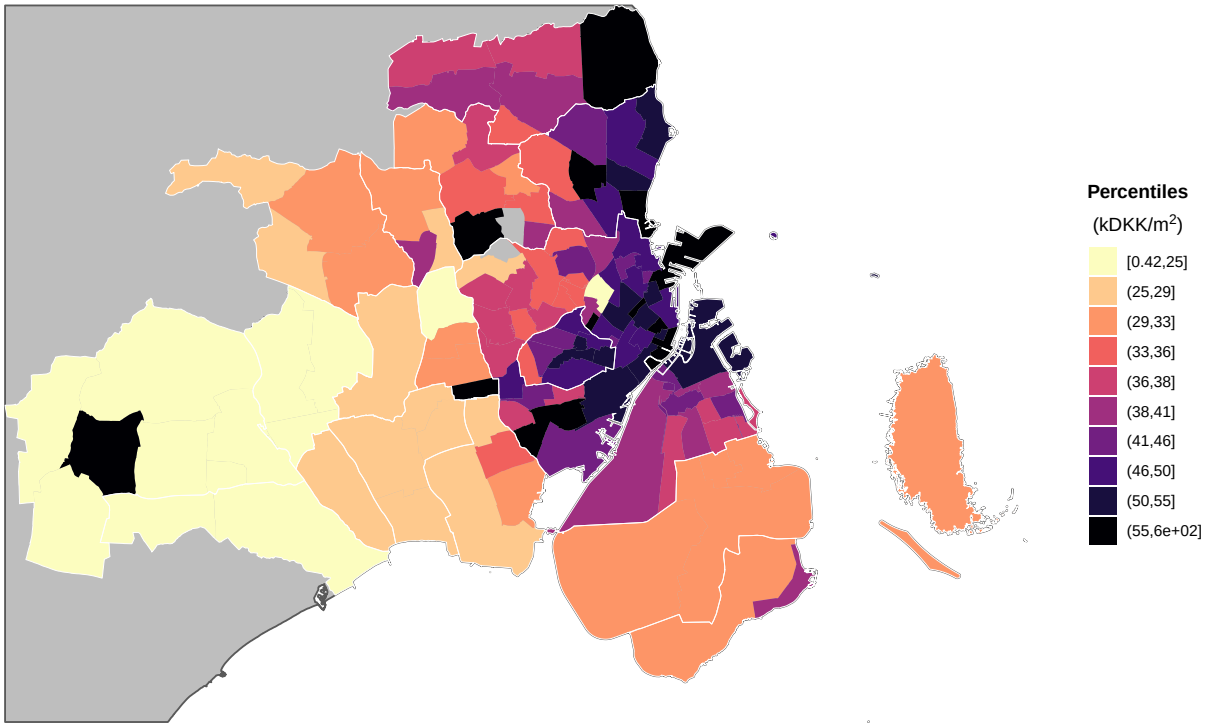


Figure 3: Median house prices in the ordinary free trade

3.3 Compositional data analysis

Ternary plot (there are conflicts between *ggtern* and *ggplot* and we make the ternary plots in a separate project; https://javiereliomedina.github.io/ternary_maps_DK/)

2020

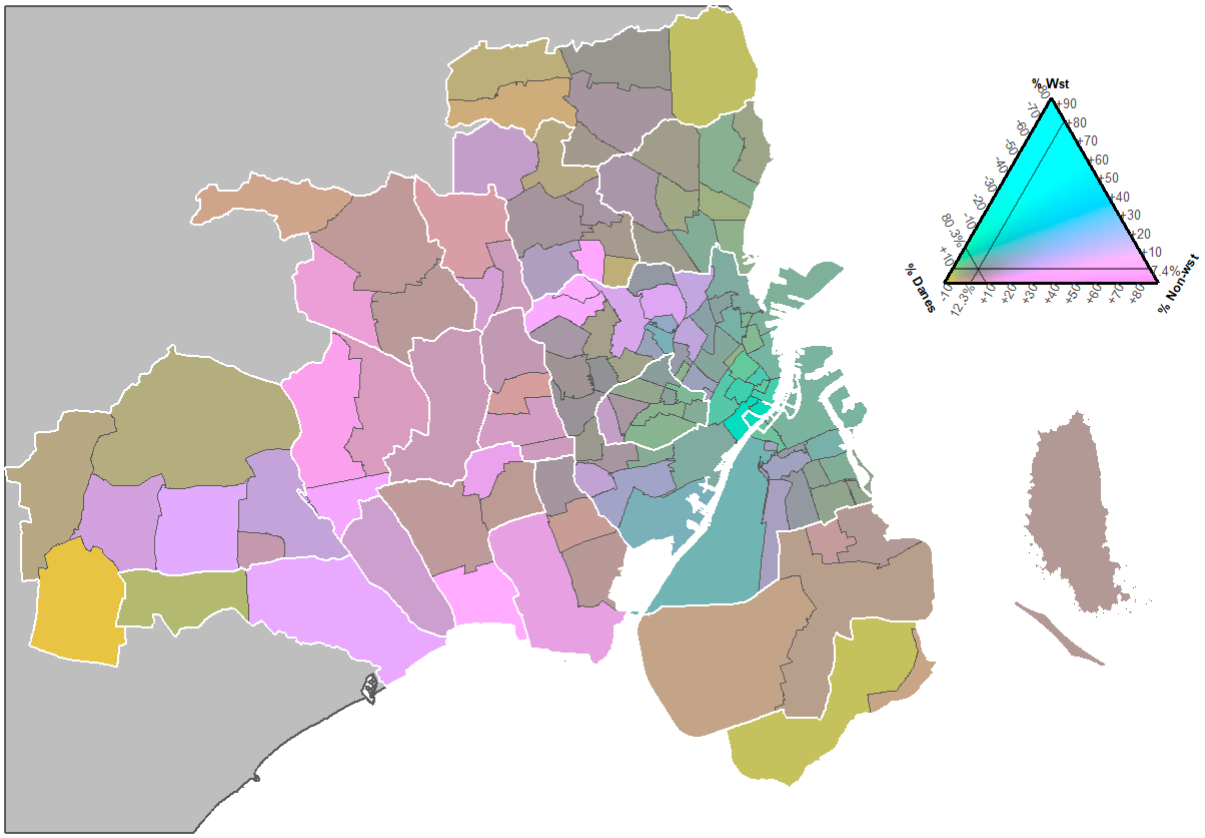


Figure 4: Population distribution in 2020

PCA (with clr transformations). Helps to identify the variables that account the most for the variability of the results and chose the balance. ….

Table 1: Spatial autocorrelation of balances

balance	moran_I	expectation	variance	statistic	p.value	method
b1	0.470	-0.008	0.003	8.563	0	Moran I test under randomisati
b2	0.542	-0.008	0.003	9.837	0	Moran I test under randomisati

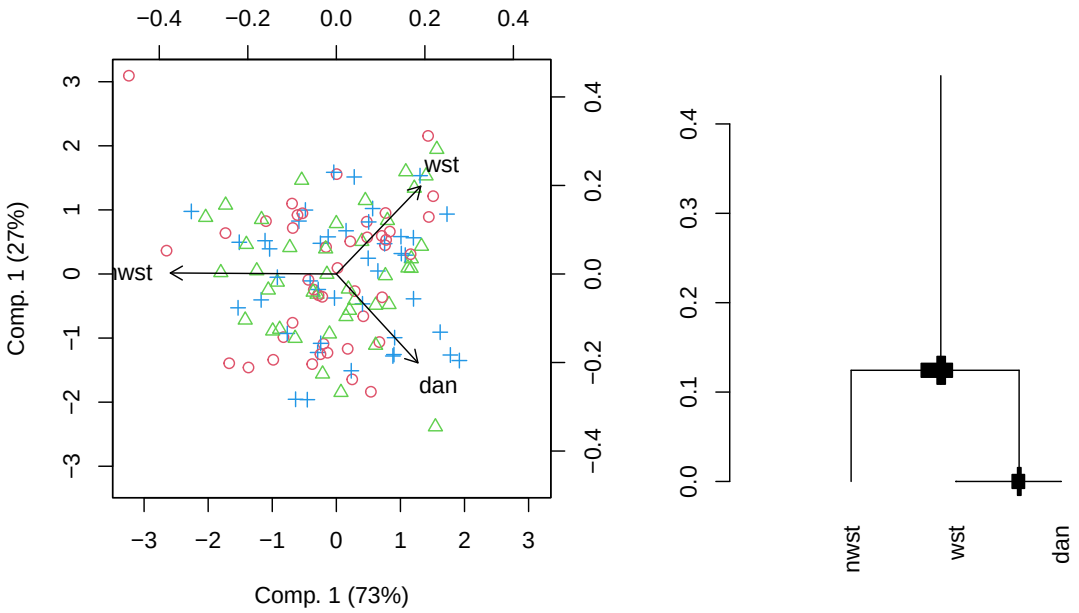


Figure 5: Biplot of clr transfomation and balance dendrogram

In our cases, with only three variables, our balances are:

$$b_1 = \sqrt{\frac{2}{3}} * ln(\frac{x_1 x_2}{x_3^2})$$

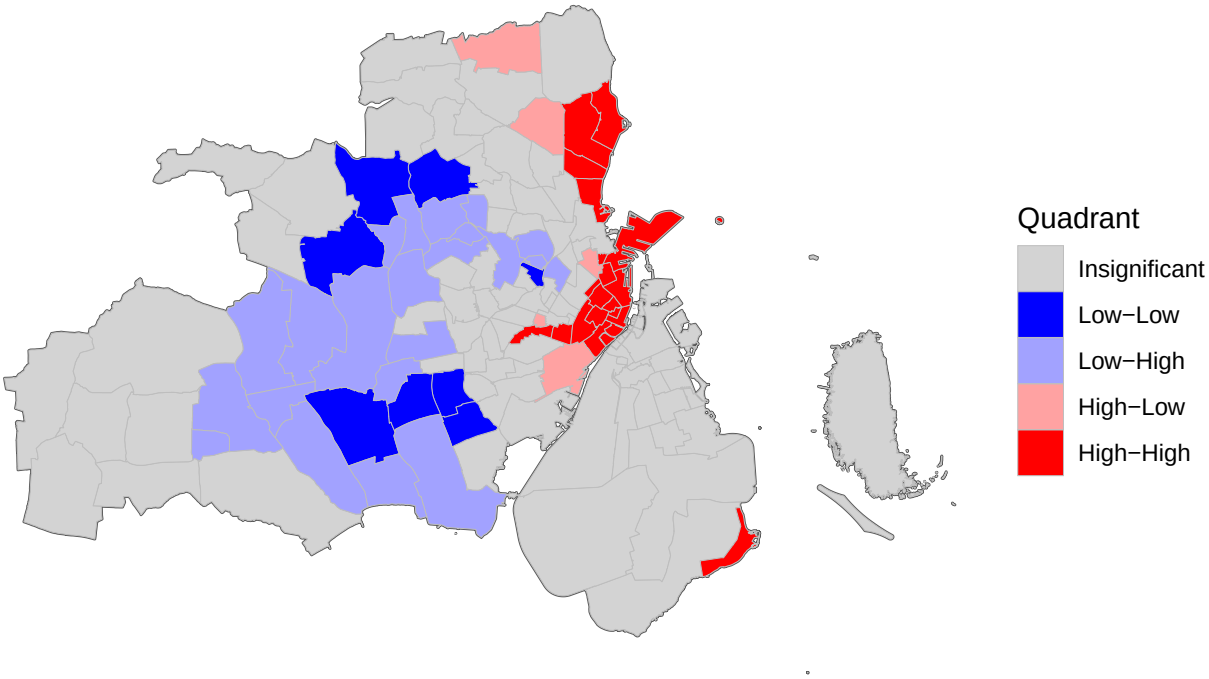
$$b_2 = \sqrt{\frac{1}{2}} * ln(\frac{x_1}{x_2})$$

Where x_1, x_2, x_3 are the Danes, Western, and Non-wester population in the parish.

We can therefore analyse the spatial autocorrelation of the balances:

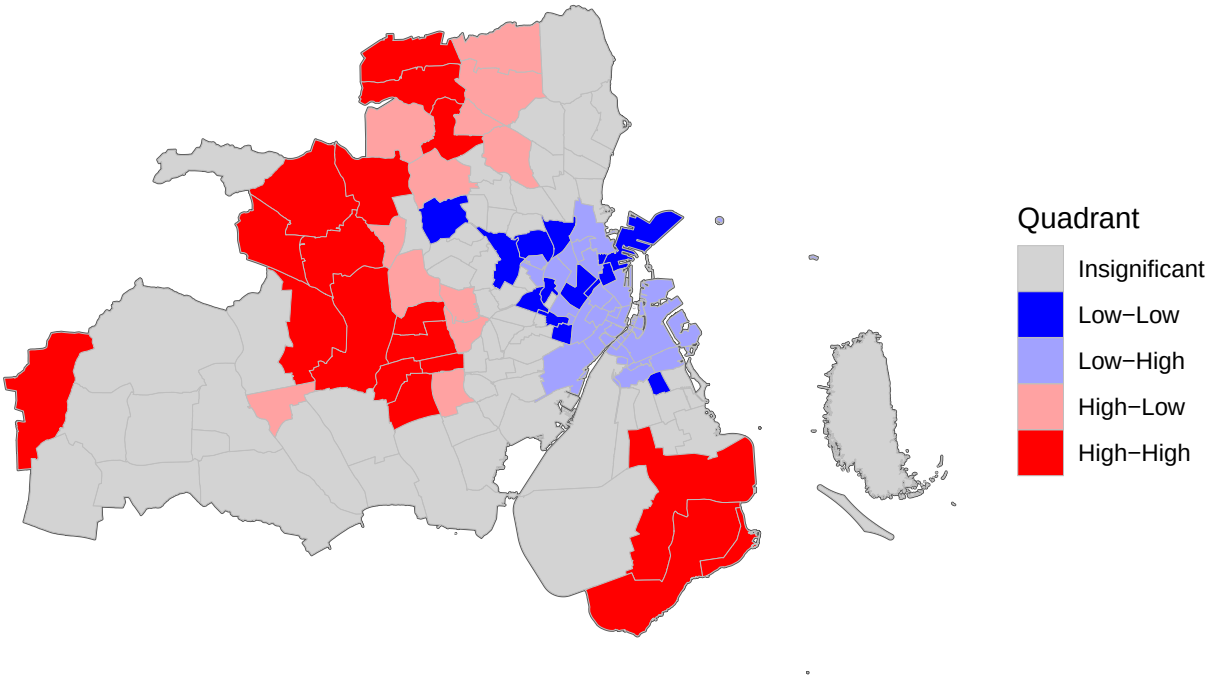
2020

b1



2020

b2



Spatial clusters (k-means cluster) with balances: separate Non-western from Danes and Western citizens.

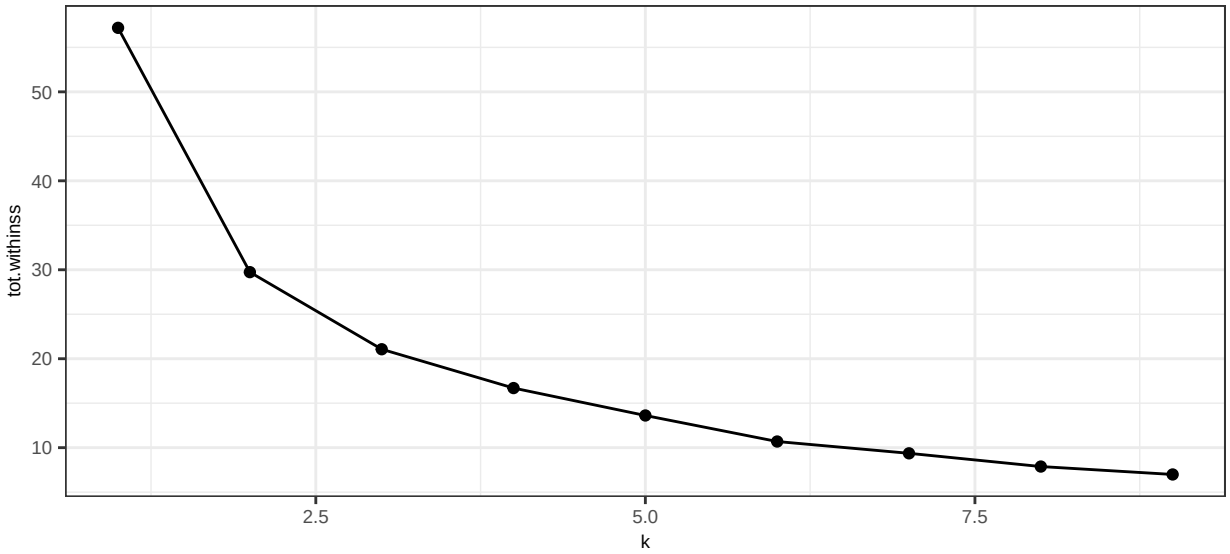
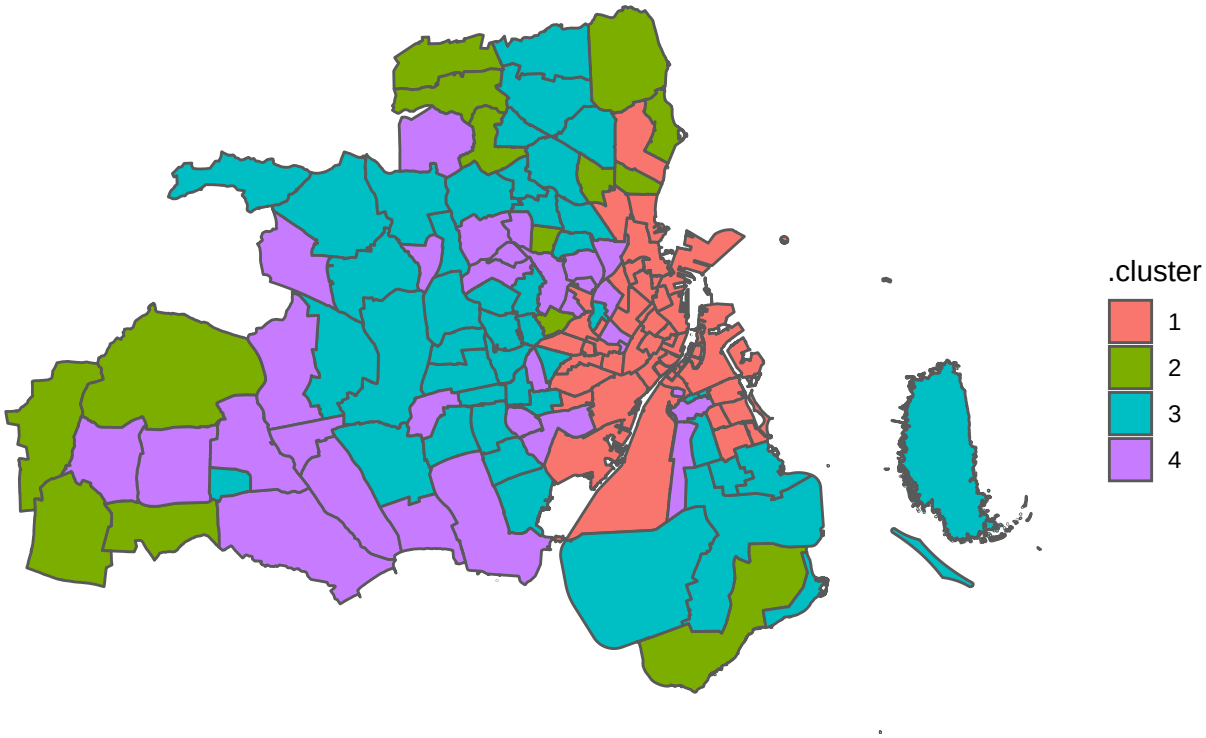


Table 2: Median population percentage by cluster

.cluster	pop_dan_pct	pop_frgn_wst_pct	pop_frgn_nwst_pct	pop_total_pct
1	80.15	10.26	8.38	100
2	89.66	4.76	5.61	100
3	81.89	5.58	12.54	100
4	68.06	7.22	25.04	100



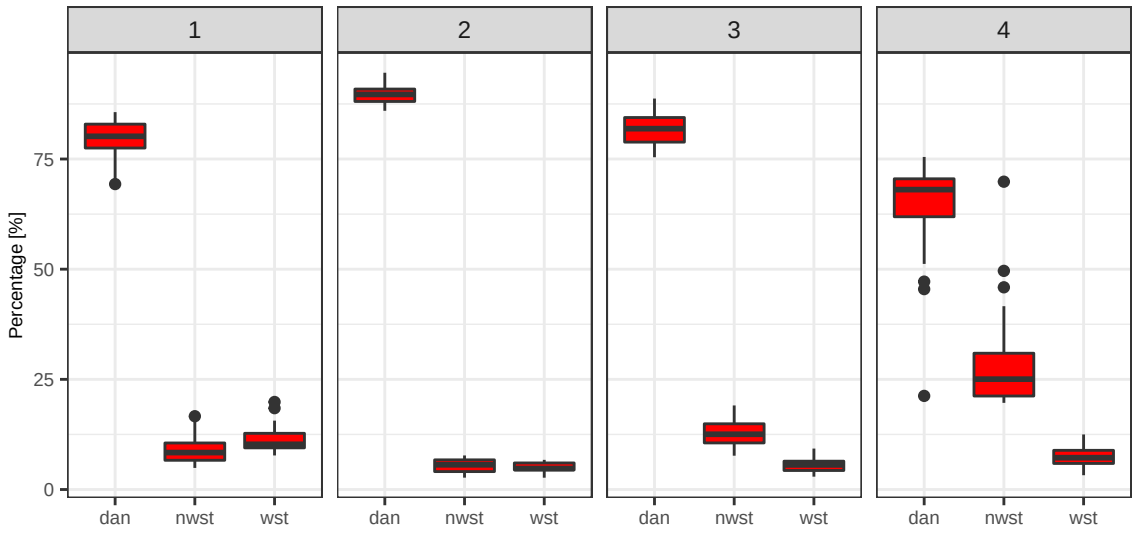


Figure 6: Cluster characteristics

3.3.1 Linear model

Links with house prices (zoom the figure to the parishes with median values).

Table 3: Regression coefficients (ß) of the log-linear model

	Estimate	CI (lower)	CI (upper)	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.691	1.451	1.931	0.121	13.971	<0.001	***
b1	0.163	0.077	0.249	0.043	3.757	<0.001	***
b2	-0.124	-0.256	0.008	0.067	-1.857	0.066	.

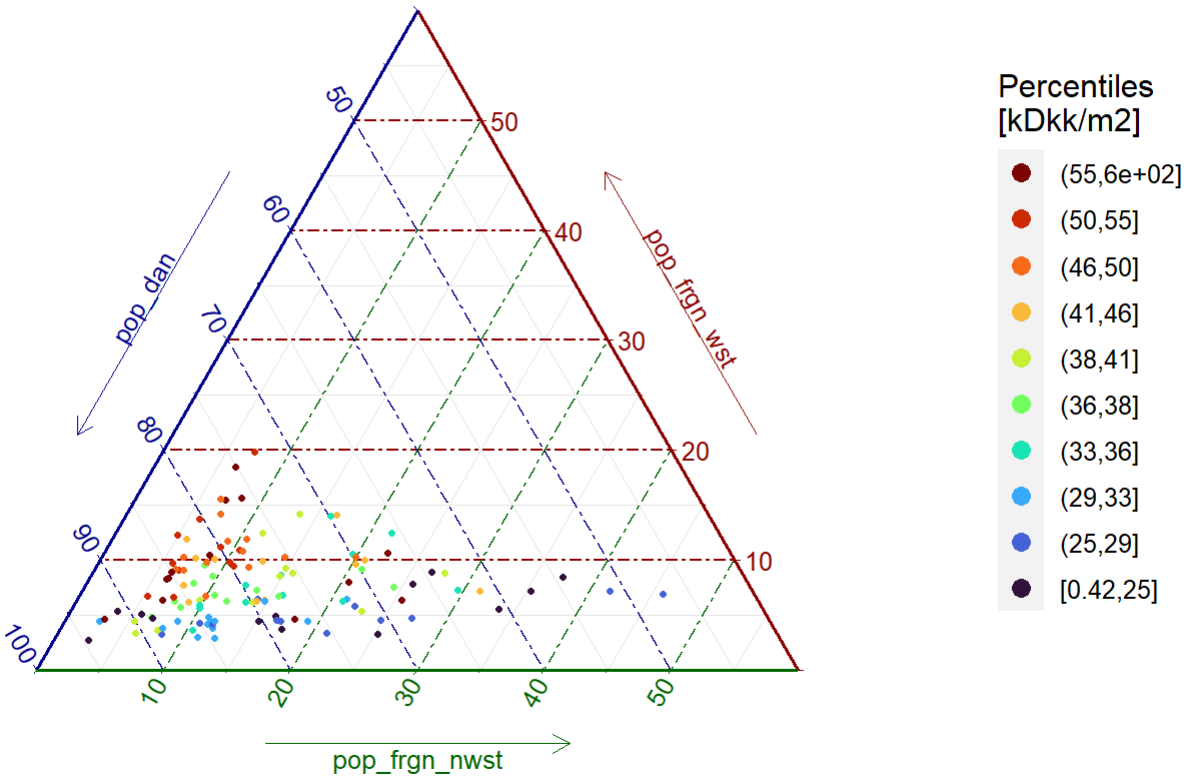


Figure 7: Population distribution in 2020

Linear models with N > 5.

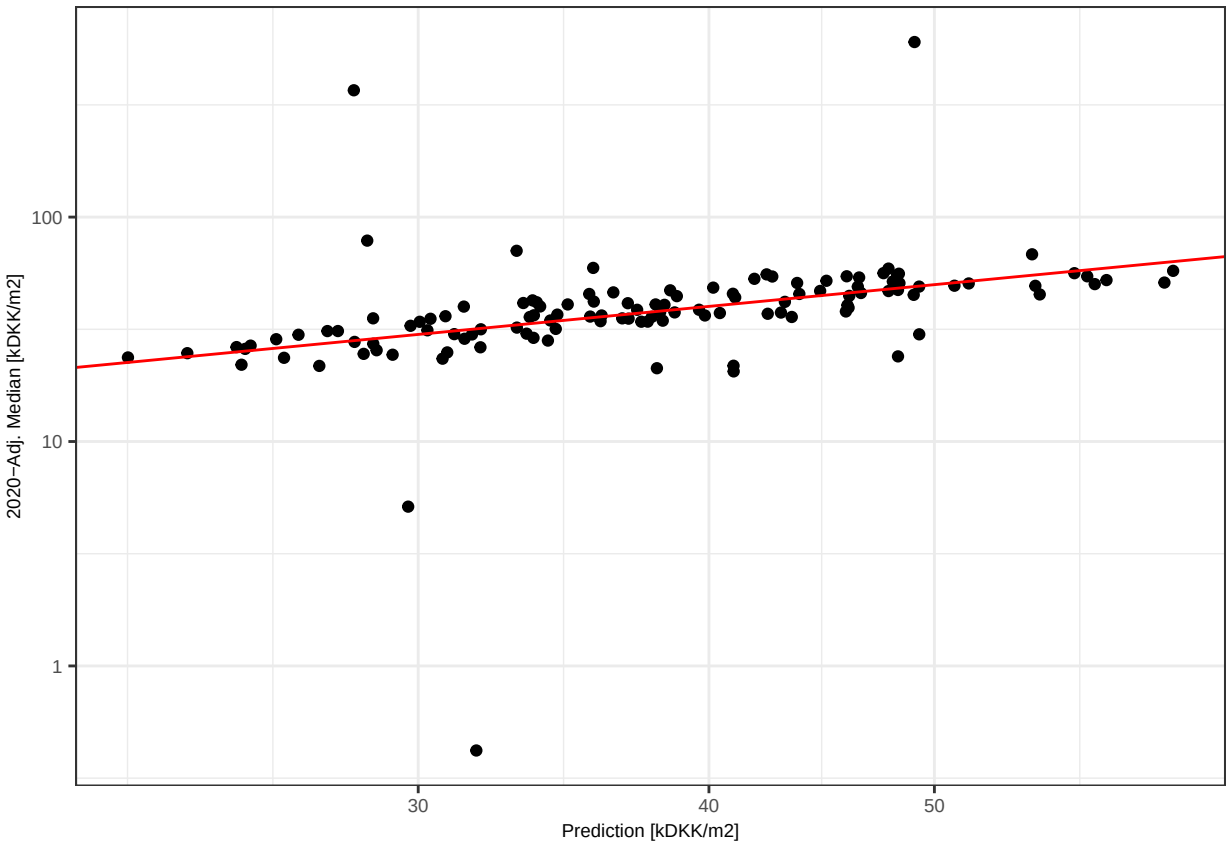


Figure 8: Linear model housing prices vs. balances (red line: x = y)

Acknowledgements

This work has been financed by Aalborg University - AAU (Project: Global flows of migrants and their impact on north European welfare states - FLOW). The sole responsibility of this publication lies with the authors. AAU is not responsible for any use that may be made of the information contained therein.

R session

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19042)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.1252
## [2] LC_CTYPE=English_United Kingdom.1252
```

```
## [3] LC_MONETARY=English_United Kingdom.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.1252
##
## attached base packages:
## [1] tools     stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] dangeo_0.0.0.9000     devtools_2.4.2        usethis_2.0.1
##  [4] tm_0.7-8              NLP_0.2-1             tidytable_0.6.3
##  [7] tidytext_0.3.1        dplyr_1.0.7           purrr_0.3.4
## [10] readr_1.4.0           tidyr_1.1.3           tibble_3.1.2
## [13] tidyverse_1.3.0       table1_1.4.2          viridis_0.6.1
## [16] viridisLite_0.4.0     units_0.7-2           spdep_1.1-8
## [19] spData_0.3.10         sp_1.4-5              stars_0.5-3
## [22] abind_1.4-5           SnowballC_0.7.0       stringr_1.4.0
## [25] sf_1.0-1              rappdirs_0.3.3        RColorBrewer_1.1-2
## [28] remotes_2.4.0         rmarkdown_2.9         papeR_1.0-5
## [31] xtable_1.8-4          car_3.0-11            carData_3.0-4
## [34] potential_0.1.0       patchwork_1.1.1       opentripplanner_0.3.1
## [37] osrm_3.4.1            osmextract_0.3.0      mapview_2.10.0
## [40] latex2exp_0.5.0       knitr_1.33            kableExtra_1.3.4
## [43] janitor_2.1.0         ggforce_0.3.3         gifski_1.4.3-1
## [46] gganimate_1.0.7       ggplot2_3.3.5         gtsummary_1.4.2
## [49] giscoR_0.2.4          ggspatial_1.1.5       furrr_0.2.3
## [52] future_1.21.0         forcats_0.5.1         dint_2.1.3
## [55] danstat_0.1.0         data.table_1.14.0     compositions_2.0-2
## [58] bookdown_0.22         bit64_4.0.5           bit_4.0.4
## [61] biscale_0.2.0         animation_2.6
##
## loaded via a namespace (and not attached):
##   [1] utf8_1.2.1          tidyselect_1.1.1    htmlwidgets_1.5.3
##   [4] grid_4.1.0          munsell_0.5.0       codetools_0.2-18
##   [7] withr_2.4.2         colorspace_2.0-2    rstudioapi_0.13
##  [10] stats4_4.1.0        robustbase_0.93-8   bayesm_3.1-4
##  [13] listenv_0.8.0       labeling_0.4.2      slam_0.1-48
##  [16] polyclip_1.10-0     farver_2.1.0        rprojroot_2.0.2
##  [19] coda_0.19-4         parallelly_1.27.0   LearnBayes_2.15.1
##  [22] vctrs_0.3.8         generics_0.1.0      xfun_0.24
##  [25] R6_2.5.0            isoband_0.2.5       cachem_1.0.5
##  [28] assertthat_0.2.1    scales_1.1.1        gtable_0.3.0
##  [31] globals_0.14.0      lwgeom_0.2-6        processx_3.5.2
##  [34] rlang_0.4.11        systemfonts_1.0.2   splines_4.1.0
##  [37] broom_0.7.8         yaml_2.2.1          modelr_0.1.8
```

```
##  [40] crosstalk_1.1.1     backports_1.2.1     tokenizers_0.2.1
##  [43] tensorA_0.36.2      ellipsis_0.3.2      raster_3.4-13
##  [46] proxy_0.4-26        sessioninfo_1.1.1   Rcpp_1.0.7
##  [49] base64enc_0.1-3     progress_1.2.2      classInt_0.4-3
##  [52] ps_1.6.0            prettyunits_1.1.1   deldir_0.2-10
##  [55] haven_2.4.1         fs_1.5.0            leafem_0.1.6
##  [58] magrittr_2.0.1      openxlsx_4.2.4      gmodels_2.18.1
##  [61] reprex_2.0.0        pkgload_1.2.1       hms_1.1.0
##  [64] evaluate_0.14       leaflet_2.0.4.1     rio_0.5.27
##  [67] readxl_1.3.1        gridExtra_2.3       testthat_3.0.4
##  [70] compiler_4.1.0      KernSmooth_2.23-20  gt_0.3.0
##  [73] crayon_1.4.1        htmltools_0.5.1.1   Formula_1.2-4
##  [76] expm_0.999-6        lubridate_1.7.10    DBI_1.1.1
##  [79] tweenr_1.0.2        dbplyr_2.1.1        MASS_7.3-54
##  [82] broom.helpers_1.3.0 boot_1.3-28         Matrix_1.3-4
##  [85] cli_3.0.1           gdata_2.18.0        parallel_4.1.0
##  [88] pkgconfig_2.0.3     foreign_0.8-81      xml2_1.3.2
##  [91] svglite_2.0.0       webshot_0.5.2       rvest_1.0.0
##  [94] snakecase_0.11.0    callr_3.7.0         janeaustenr_0.1.5
##  [97] digest_0.6.27       cellranger_1.1.0    curl_4.3.2
## [100] gtools_3.9.2        satellite_1.0.2     lifecycle_1.0.0
## [103] nlme_3.1-152        jsonlite_1.7.2      desc_1.3.0
## [106] fansi_0.5.0         pillar_1.6.1        lattice_0.20-44
## [109] pkgbuild_1.2.0      fastmap_1.1.0       httr_1.4.2
## [112] DEoptimR_1.0-9      survival_3.2-11     glue_1.4.2
## [115] zip_2.2.0           png_0.1-7           class_7.3-19
## [118] stringi_1.6.2       memoise_2.0.0       e1071_1.7-7
```