Template

Mark Barkalov (2782988)   Joël Dijkstra (2854215)   Shabana Esmati (2868438)
Jasper Hoogendoorn (2856970)   Giovanni Koek (2857652)
Robin Kolmus (2862402)   Elijah Kruize (2868400)   Tim van der Laan (2854528)

2025-06-24

Set-up your environment

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr)
library(readr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v forcats   1.0.0      v stringr   1.5.1
## v lubridate 1.9.4      v tibble    3.2.1
## v purrr     1.0.4

## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(cbsodataR)
library(sf)
```

```
## Linking to GEOS 3.13.1, GDAL 3.10.2, PROJ 9.5.1; sf_use_s2() is TRUE
```

Title Page

Include your names:

Mark Barkalov (2782988)

Joël Dijkstra (2854215)

Shabana Esmati (2868438)

Jasper Hoogendoorn (2856970)

Giovanni Koek (2857652)

Robin Kolmus (2862402)

Elijah Kruize (2868400)

Tim van der Laan (2854528)

Include the tutorial group number:

Group 4

Include your tutorial lecturer's name:

Miss. C. Schouwenaar

Part 1 - Identify a Social Problem

Use APA referencing throughout your document.

1.1 Describe the Social Problem

In recent years it has become obvious that the housing crisis has risen across the developed countries and become a major issue for their residents. Unfortunately, the Netherlands did not become an exception in the current trend. One of the clear indicators of a serious housing problem could serve the WOZ-value, which reflects the official estimated market value of homes, and which has been increasing much faster than the average income. This growing gap makes it more difficult for people, especially first-time buyers and young adults, to afford accommodation. According to recent data (Mouissie & Kraniotis, 2023), there is now a severe shortage of housing across the country — so severe that no single province in the Netherlands had a housing surplus at the moment of 2023. In many places, demand for housing is still far greater than supply. This has made home ownership a distant goal for many Dutch citizens, pushing more people into the rental market, which, as the consequence of increased demand, is also becoming less affordable.

This issue must be taken seriously because it affects both individuals and society as a whole. A lack of affordable housing can reduce people's quality of life and create long-term financial stress. It can also make it harder for people to move for work or start a family, which affects the economy as a whole. In addition, the growing number of single-person households, due to aging and changing lifestyles, has further increased the pressure on the housing market, while at the same time, building new homes has become more difficult. As a result, as noted in the recent ABN AMRO article (Langen, 2025), the country is expected to be short by over 400,000 homes in 2025. Given the scale of the problem and its social impact, it is crucial that housing policy is rethought and action is taken to improve affordability. Without intervention, the Dutch housing market will continue to exclude more people from owning a home.

Part 2 - Data Sourcing

2.1 Load in the data

```
data2 <- read.csv("wozdata.csv")
data <- read.csv("inkdata.csv")
zeefle <- read.csv2("zeefle.csv")
provincie_2020 <- cbs_get_sf("provincie", 2020)
df_2023 <- read_csv2("bedrijvenNL2023.csv", show_col_types=FALSE)
```

```
## i Using "','" as decimal and "'.'" as grouping mark. Use `read_delim()` for more control.
```

```
df_2019 <- read_csv2("bedrijvenNL2019.csv", show_col_types=FALSE)
```

```
## i Using "','" as decimal and "'.'" as grouping mark. Use `read_delim()` for more control.
```

2.2 Provide a short summary of the dataset(s)

Data2: Data2 (CBS, 2025b) is a dataset from CBS that mentions the average house prices of houses over the years 2019-2023. This dataset gives us data for the Netherlands, per city and per province. The dataset gives average house prices of houses you can buy and the value of houses that people rent out. We used this dataset to find out what the changes in house prices were over the years.

Data and zeefle: Data (CBS, 2025a) is a dataset from CBS that mentions the average income of groups with different hour contracts over the years 2019-2023 in the Netherlands. It also gives the total average of personal income which we used. We also have the dataset zeefle (CBS, 2024) which is the same as dataset Data but it mentions the average income of Zeeland and Flevoland which we used for the temporal variation.

provincie_2020: Dataset provincie_2020 is a dataset from the package cbsodataR that we used to make a map for the spatial variation. It provides the coordinates of every province.

df_2023 and df_2019: These two datasets (CBS, 2025c) are from the same dataset from CBS and they give the same information. They provide us with data about the amount of businesses per province but one in the year 2023 and the other in 2019. We used this for our subgroup analysis.

2.3 Describe the type of variables included

The datasets makes use of multiple variables, including:

- WOZ Value in the Netherlands in each province from 2019 - 2023

This data was gathered from CBS and is an administrative source about the house prices in the years 2019 to 2023.

- Average income of working population in the Netherlands, Zeeland and Flevoland from 2019-2023

This data was gathered from CBS and is an administrative source about the average income in the Netherlands and the provinces in which we specifically picked out Zeeland and Flevoland.

- Number of businesses per province in 2019 and 2023

This data was gathered from CBS and is an administrative source about the number of businesses per province in the years 2019 and 2023.

Part 3 - Quantifying

3.1 Data cleaning

If we wanted to keep certain rows in a dataset and delete everything else then we used this code:

```r
data2 <- data2[-c(1:5, 19:379), -c(1, 2, 5:7, 9:11, 13:15, 17:19, 21:23)] #verwijdert onnodige columns
```

If we wanted to make a certain column in the dataset the name of the rows we used this code:

```r
rownames(data2) <- data2[, 1]  # Set row names using the first column
```

Because it is a Dutch dataset the decimals are commas, so to make the commas into decimal points we used this code:

```r
data <- apply(data, 2, function(x) gsub(",", ".", x)) # replaces every comma into a decimal
```

Because we wanted to put two datasets under each other we used this code:

```r
data <- rbind(data, data2) # merges the two data sets
```

If we wanted to make two data sets join we needed to change names so it could join and the codes we used looked like this:

```r
data2$Regio.s <- gsub(" \\(PV\\)", "", data2$Regio.s)
data2$Regio.s <- gsub("Fryslân", "Friesland", data2$Regio.s)
```

3.2 Generate necessary variables

Variable 1: House prices to average personal income ratio. We used this to see how much yearly income is needed to fully purchase the average house in each province and the country as a whole. We did this for the Netherlands and for the provinces Zeeland and Flevoland. For the provinces we used the average personal income in their respective province and for the Netherlands we used the average personal income of the Netherlands as a whole.

```r
new_row <- data[5, ] / data[3,]  # Element-wise division of average Netherlands house prices divided by
```

```r
new_row_fle <- data[11, ] / data[19,]  # Element-wise division of Flevoland's houseprices divided by th
new_row_zee <- data[16, ] / data[20,]  # Element-wise division of Zeeland's houseprices divided by the
```

Variable 2(dif): This variable is the house prices of 2023 divided by the house prices in 2019. This way we can see which city had the most increase during the years. We could find the city with the most increase and the city with the least amount of increase and find out if both cities increased more than the income. We can also use this in the subgroup analysis to find if an increase in businesses has a direct effect on the house prices.

```r
dif <- c() # give variable dif the value of a vector

for(i in 1:nrow(data2))
{
  dif[i] = data2$"2023"[i] / data2$"2019"[i] # divides the house prices in 2023 by the houseprices in 2
}
```

Variable 3(busi): This variable is the amount of businesses in each province in 2023 divided by the businesses in 2019. This way we can compare the change in house prices with the change in businesses, to see if an increase in businesses has a direct impact on the house prices in the Netherlands.

```
busi <- c() # give variable busi the value of a vector

for(i in 1:nrow(df_2019))
{
  busi[i] = df_2019$"Vestigingen (Aantal).x"[i] / df_2019$"Vestigingen (Aantal).y"[i] # Element division
}
```

3.3 Visualize temporal variation

```
ggplot(graph_data_long, aes(x = Year, y = Value, color = Region, group = Region)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  labs(
    title = "House Prices Relative to Average Personal Income (2019-2023)",
    x = "Year",
    y = "Relative Value"
  ) +
  theme_bw()
```
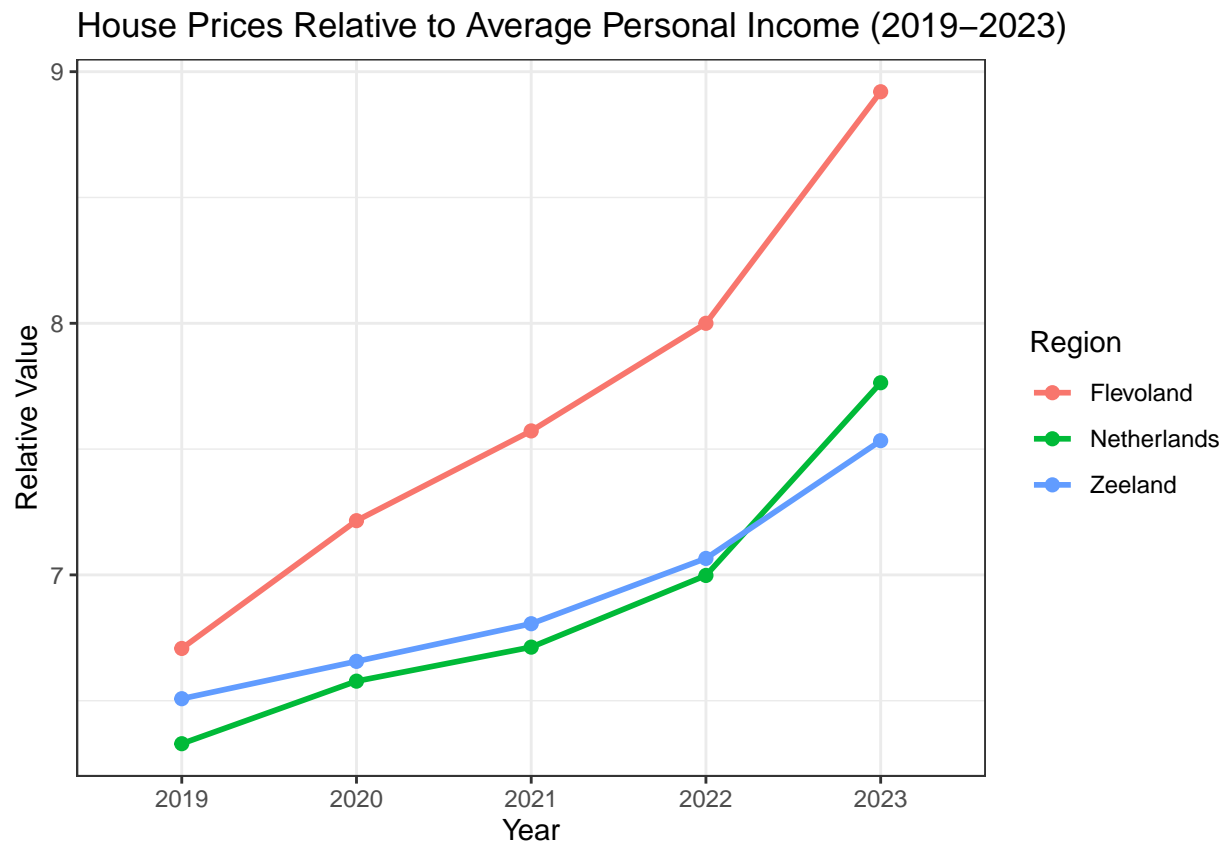

With this temporal variation we wanted to see if over the years the WOZ value relative to the average income actually increased, with which we can conclude if the increase in house prices actually had an impact on

the affordability of houses. For this we decided to measure this in the Netherlands, Flevoland, who has the highest change in WOZ value of the provinces, and Zeeland, who has the lowest change in WOZ value of the provinces. For each of these places we divided the house prices with the average income in that area. When we looked at the graph we were able to confirm that the house prices relatively increased compared to the average income and in particular did Flevoland have the highest relative increase. However there was a limitation in our findings: that this graph did not take into account that there are other expenses that have changed over time. So we cannot decide based solely on this graph how big of an impact the increase in house prices has on the percentage of income that people spend on their houses.

3.4 Visualize spatial variation

```
# Create the map using ggplot2
ggplot(data2) +
  geom_sf(aes(fill = `2020`)) +  # Mapping values from the 2020 column
  scale_fill_gradient(low = "blue", high = "red", limits = c(150, 500)) +
  theme_minimal() +
  labs(title = "Map of House prices per region", fill = "House prices in 2020\nx1000 in euros")
```
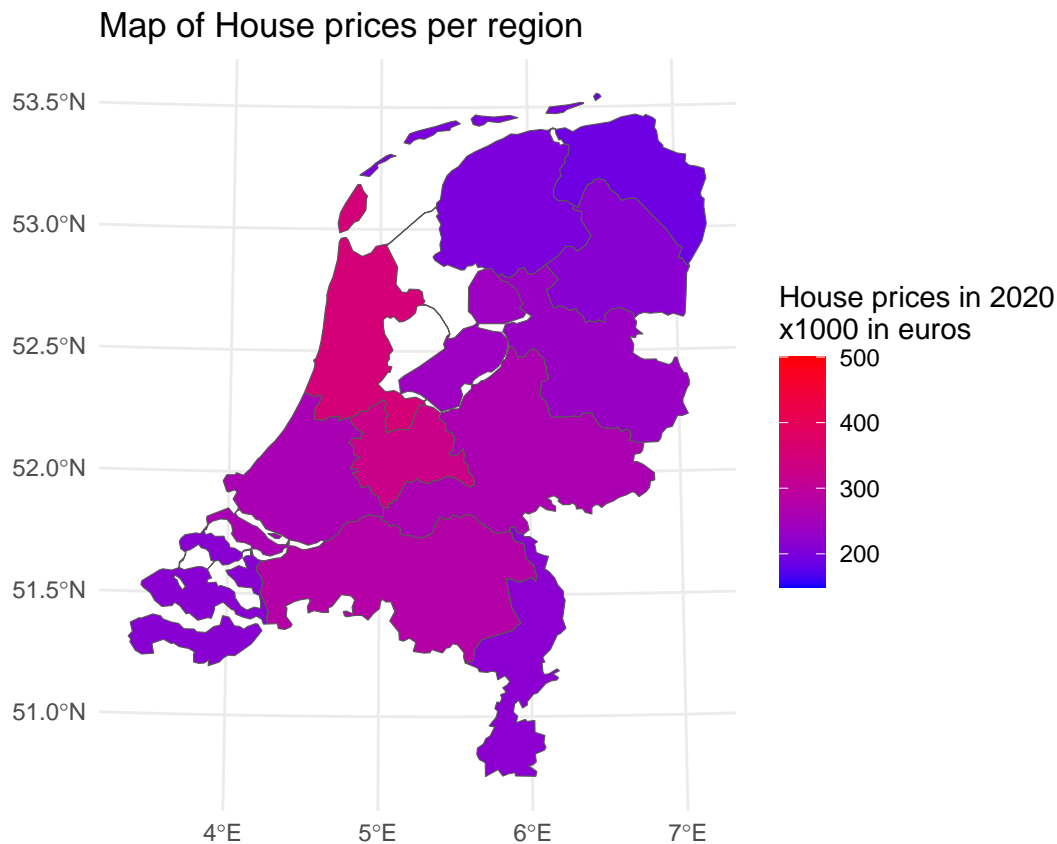


```
# Create the map using ggplot2
ggplot(data2) +
  geom_sf(aes(fill = `2023`)) +  # Mapping values from the 2020 column
  scale_fill_gradient(low = "blue", high = "red", limits = c(150, 500)) +
  theme_minimal() +
  labs(title = "Map of House prices per region", fill = "House prices in 2023\nx1000 in euros")
```

Map of House prices per region
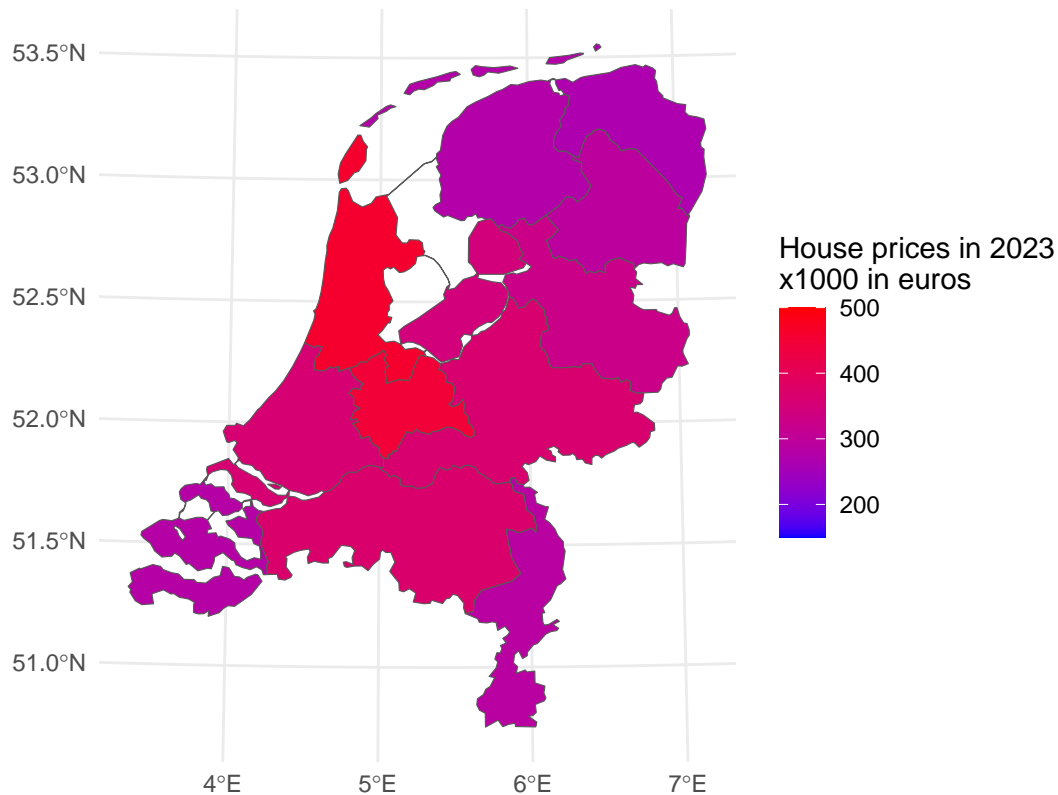
For this spatial variation we mapped the house prices in 2020 and 2023 in order to see if there was a province that exceptionally grew compared to the other provinces. When you look at the two maps you can clearly see that every province had an increase in house prices and you can also see that the provinces all had relatively the same growth. While creating the spatial variation, we encountered a limitation. We could not find any data on the average m2 of houses per province so that we could make this map based on the average house prices per m2 in each province. So this map isn't as accurate as it could have been if we found that data.

3.5 Visualize sub-population variation

Does the growth of new businesses have a direct impact on the change in WOZ value?

```
ggplot(data, aes(x = Group, y = Value, fill = Category)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Growth of New Companies and WOZ change for the 3 groups",
       x = "Group", y = "Average Value", fill = "Category")
```

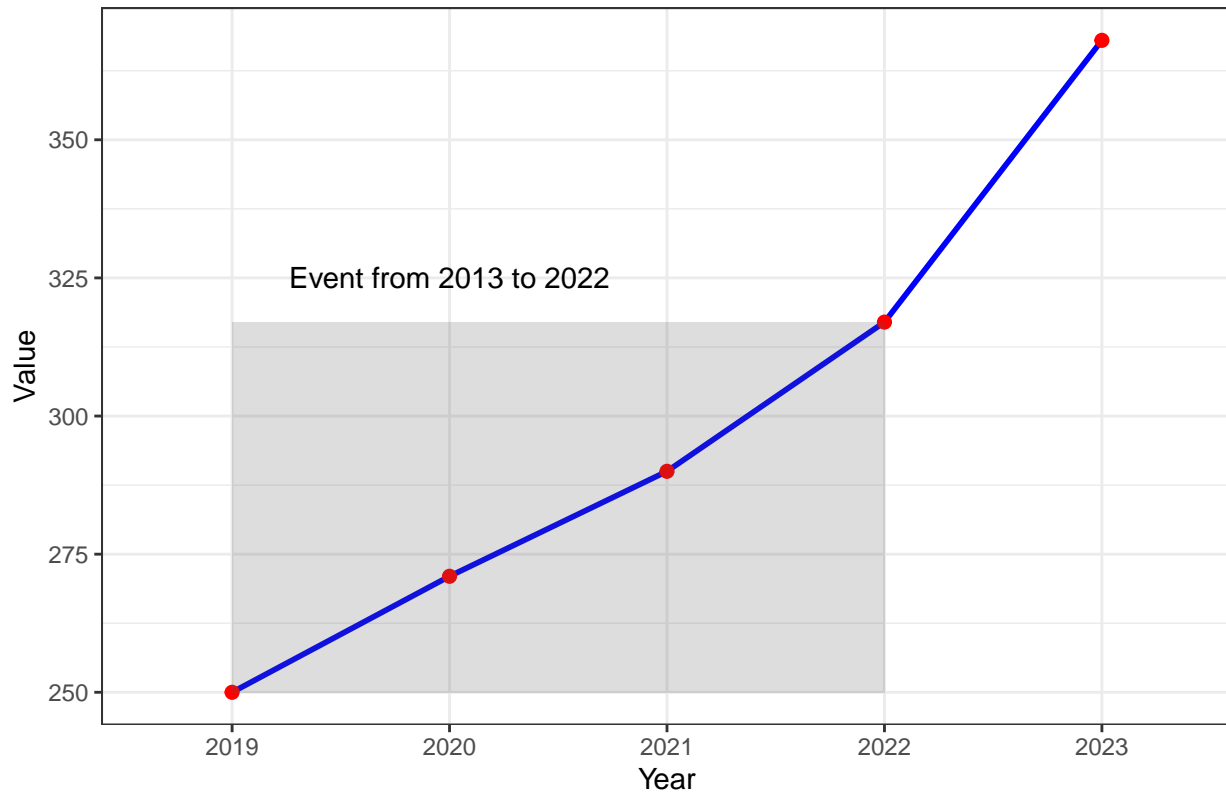Growth of New Companies and WOZ change for the 3 groups
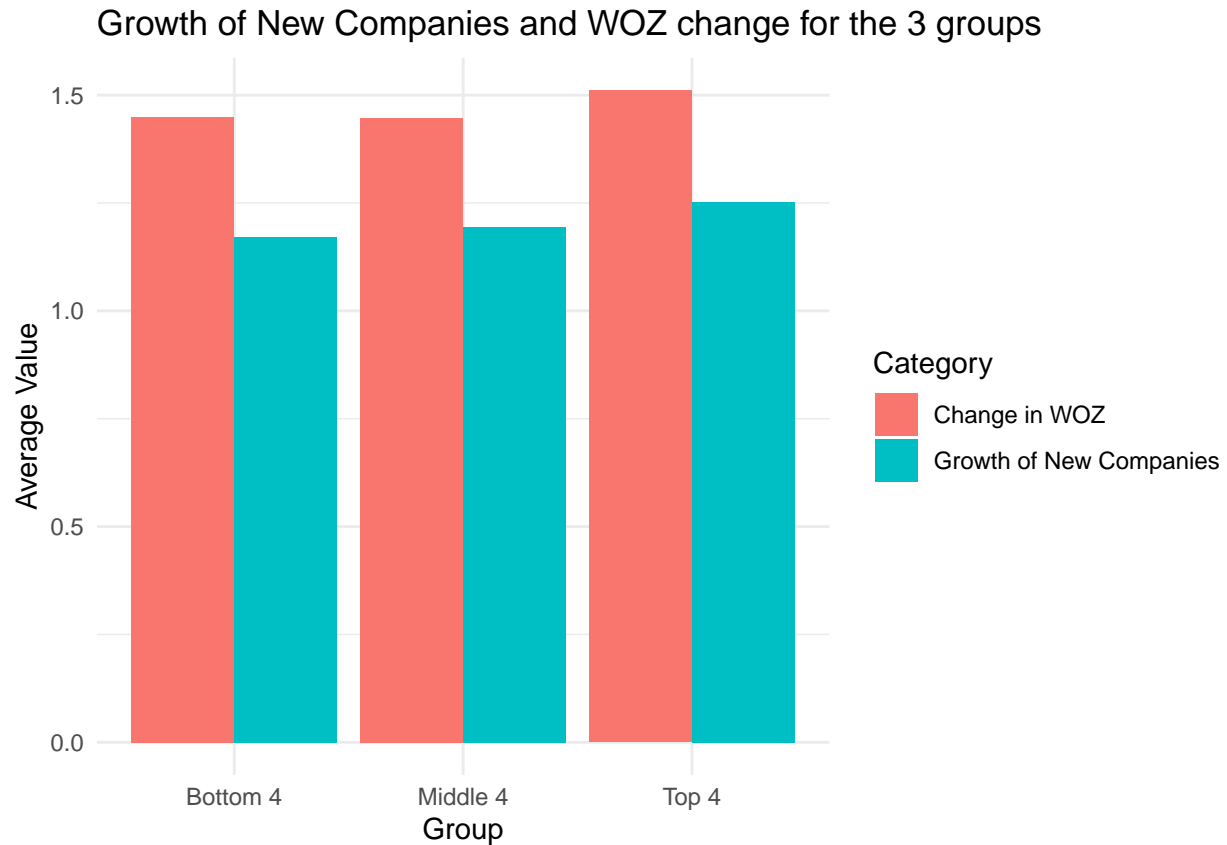
For this subgroup analysis we wanted to see if an increase in new companies had a direct impact on WOZ value. We analysed this by grouping the 12 provinces into 3 groups based on the growth of new companies. Highest 4 having the highest growth and lowest 4 having the least growth in new companies. Each group we took the average growth of new companies and the average change in WOZ value. Based on the bar plot we could conclude that growth of new companies and change in WOZ value have a positive correlation, but we could not conclude that the growth of new companies has a direct impact on the WOZ value. This is because of a limitation that the growth of new companies and the WOZ value is almost the same in the three groups.

3.6 Event analysis

```
ggplot(graph_data, aes(x = Year, y = Value, group = 1)) +  # Explicitly set `group = 1`
  geom_line(color = "blue", linewidth = 1) +
  geom_point(color = "red", size = 2) +
  annotate("text", label = "Event from 2013 to 2022", size = 4, x = 2, y = 325) +
  annotate("rect", xmin = 1, xmax = 4, ymin = 250, ymax = 317, alpha = .2) +
  labs(
    title = "Average house prices in the Netherlands",
    x = "Year",
    y = "Value"
  ) +
  theme_bw()
```

Average house prices in the Netherlands

Event from 2013 to 2022

(Chart: Average house prices in the Netherlands, with Value on the y-axis ranging from 250 to over 350, and Year on the x-axis from 2019 to 2023. The blue line rises steadily from about 250 in 2019 to about 367 in 2023. A shaded rectangle spans the region labeled "Event from 2013 to 2022.")

Over the past decade, the Netherlands has experienced a steep rise in house prices, driven largely by a growing housing shortage. As the population expanded and more people began living alone, the demand for homes increased significantly. However, the supply of new housing failed to keep pace. This imbalance became one of the most pressing challenges in the Dutch housing market, with an estimated shortage of nearly 390,000 homes by the end of 2023.

At the same time, historically low interest rates in the period from 2008-2022 following the financial crisis made mortgages more affordable, allowing buyers to borrow more and pushing property prices higher. Real estate increasingly became an investment tool, particularly in major cities like Amsterdam, Rotterdam and Utrecht. Homes were often bought not to live in, but to rent out, reducing availability for first-time buyers. Government policies, such as the mortgage interest tax deduction and borrowing rules that allowed couples to combine their incomes, further fueled demand and purchasing power.

In 2023, house prices surged more sharply than in previous years. After a short dip in 2022, interest rates stabilized and wages began to rise, giving buyers more financial room. Meanwhile, the housing shortage had worsened, leading to fierce competition and widespread overbidding. These conditions caused the strongest price increases in years, with some areas seeing double-digit growth within months.

Part 4 - Discussion

4.1 Discuss your findings

Our analysis confirms that from 2019 to 2023, housing prices in the Netherlands, measured through WOZ values, have increased at a much faster rate than personal incomes. This widening gap has made it increasingly difficult for average residents, especially first-time buyers, to purchase a home.

The national house price to income ratio has risen steadily, reflecting a decline in overall affordability. Provinces like Flevoland showed the sharpest increases in WOZ values relative to income, while Zeeland followed a more moderate trend. Still, in both cases, income growth failed to keep up with rising housing prices.

Our spatial analysis demonstrated that affordability issues are not limited to major cities. The problem is widespread and affects nearly all provinces, both urban and rural.

We also examined the relationship between business growth and house price development. Although some provinces with strong economic growth saw larger increases in housing prices, the connection was not consistent. This suggests that other structural factors, such as limited housing supply and long permit procedures, play a more significant role.

In conclusion, the data clearly shows a national housing affordability problem. Without targeted policy action, this trend will likely continue, putting further pressure on both the housing and rental markets.

Part 5 - Reproducibility

5.1 Github repository link

Provide the link to your PUBLIC repository here: https://github.com/jasperdaniel06/Team-Daap

5.2 Reference list

- ABN AMRO. (2025, January 16). Woningtekort in 2025: hoe groot is het probleem?. https://www.abnamro.nl/nl/prive/hypotheken/wonen/woningtekort-in-2025.html

- Langen, M. (2025, January 16). Woningmarkt: Bouwen naar behoefte. ABN AMRO. https://assets.abnamro.com/api/public/content/woningtekort-in-nederland.pdf

- Mouissie, S., & Kraniotis, L. (2023, November 11). Schreeuwend tekort Aan Woningen en Hoge Huizenprijzen: Hoe is Het Zo Gekomen?. NOS.nl - Nieuws. https://nos.nl/collectie/13960/artikel/2497415-schreeuwend-tekort-aan-woningen-en-hoge-huizenprijzen-hoe-is-het-zo-gekomen

- Centraal Bureau voor de Statistiek (CBS). (2025a, February 26). Regionale inkomensverschillen; gestandaardiseerd besteedbaar inkomen. https://opendata.cbs.nl/#/CBS/nl/dataset/85277NED/table?searchKeywords=gemiddeld%20inkomen

- Centraal Bureau voor de Statistiek (CBS). (2025b, June 19). Gemiddelde WOZ-waarde van woningen. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85036NED/table

- Centraal Bureau voor de Statistiek (CBS). (2025c, April 11). Vestigingen van bedrijven. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81578NED/table?fromstatweb

- Centraal Bureau voor de Statistiek (CBS). (2024, 1 November). Inkomen van personen; persoonskenmerken, regio (indeling 2024). https://opendata.cbs.nl/statline/#/CBS/nl/dataset/86005NED/table?dl=C2433