Household Shortage

Joris Heeremans (2787294), Tom Nethe (2784796), Mohamed Hannun (2824390),
Jelle Menninga (2864828), Daan Notenboom (2853496), Suleyman Yigitsoy (2861891),
Soufyan Attokoki (2829868)

2025-06-24

Contents

1 Set-up your environment

2 Household Shortage

Joris Heeremans, Tom Nethe, Mohamed Hannun, Jelle Menninga, Daan Notenboom, Suleyman Yigitsoy, Soufyan Attokoki

Tutorial group number: 5

Tutorial team number: 1

Tutorial lecturer's name: Jack Fitzgerald

3 Part 1 - Identify a Social Problem

3.1 1.1 Describe the Social Problem

The Netherlands is facing a severe housing shortage, estimated to be around 400,000 homes.(CBS, 2025) This imbalance between supply and demand was created by population growth (including immigration), an increase in single-person households, and a lack of new construction (Langen, January 2025). This imbalance has led to skyrocketing prices and fierce competition for available houses. The house prices have risen significantly, making it difficult for many, especially solo buyers and those with low to middle incomes, to find suitable and affordable housing. The government needs to come up with new ideas; otherwise, this problem will only grow bigger.

4 Part 2 - Data Sourcing

4.1 2.1 Load in the data

```
#load Housing Stock data
Voorraad_woningen <- read.csv("Vooraad_woningen_Google - Voorraad_woningen.csv")
KW1_voorraad <- Voorraad_woningen[grepl("1e kwartaal", Voorraad_woningen$Perioden), ]

Huishoudens <- read.csv("Aantal_huishoudens - Blad1 (2).csv")

Migratie <- read.csv("Migratie - Blad1 (2).csv")

Bevolking <- read.csv("Bevolking - Blad1.csv")
```

4.2 2.2 Provide a short summary of the dataset(s)

```
head(KW1_voorraad)
```

```
##           Regio.s         Perioden  aantal
## 2       Nederland 2019 1e kwartaal 7830489
## 6       Nederland 2020 1e kwartaal 7909246
## 10      Nederland 2021 1e kwartaal 7987921
## 12 Groningen (PV) 2019 1e kwartaal  279996
## 16 Groningen (PV) 2020 1e kwartaal  283748
## 20 Groningen (PV) 2021 1e kwartaal  285538
```

```
head(Huishoudens)
```

```
##          Regio.s Perioden  aantal
## 1      Nederland     2018 7857914
## 2      Nederland     2019 7924691
## 3      Nederland     2020 7997800
## 4      Nederland     2021 8043443
## 5 Groningen (PV)     2018  292255
## 6 Groningen (PV)     2019  293740
```

Both data sets include information about the stock houses and the number of households in the Netherlands per province and period.

5 Part 3 - Quantifying

5.1 3.1 Data cleaning

We already filtered the data through the filter on the site to make the filtering process easier. We did this because there were a lot of unnecessary data we do not need.

```
# Add a year column by extracting the first 4 characters of the period column
Voorraad_woningen$Year <- substr(Voorraad_woningen$Perioden, 1, 4)
Huishoudens$Year <- substr(Huishoudens$Periode, 1, 4)


# Set which years we want to use
years_to_keep <- c("2019", "2020", "2021")

# Set which provinces
provincies <- c("Groningen (PV)", "Fryslân (PV)", "Drenthe (PV)", "Overijssel (PV)", "Flevolan
                "Gelderland (PV)", "Utrecht (PV)", "Noord-Holland (PV)", "Zuid-Holland (PV)",
                "Zeeland (PV)", "Noord-Brabant (PV)", "Limburg (PV)")

# Select only data for Netherlands or provinces, and only the needed years, only first quarter
housing_stock_selected <- subset(Voorraad_woningen,
                                 Regio.s %in% provincies &
                                 Year %in% years_to_keep &
```

```
                                grepl("1e kwartaal", Perioden))


households_selected <- subset(Huishoudens,
                              Regio.s %in% provincies &
                              Year %in% years_to_keep)

# Reshape migration data from wide to long
migratie_long <- Migratie %>%
  filter(Onderwerp %in% c("Immigratie", "Emigratie")) %>%
  select(-X) %>%
# remove the column causing an error
  pivot_longer(
    cols = starts_with("X"),
    names_to = "Year",
    names_prefix = "X",
    values_to = "Aantal"
  ) %>%
  select(Regio.s, Onderwerp, Year, Aantal) %>%
  pivot_wider(
    names_from = Onderwerp,
    values_from = Aantal
  ) %>%
  mutate(
    Immigratie = as.numeric(Immigratie),
    Emigratie = as.numeric(Emigratie)
  )
# Reshape population data from wide to long
bevolking_long <- Bevolking %>%
  filter(Onderwerp == "Bevolking") %>%
  select(-X) %>%
# Remove the column that causes an error
  pivot_longer(
    cols = starts_with("X"),
    names_to = "Year",
    names_prefix = "X",
    values_to = "Bevolking"
  ) %>%
  select(Regio.s, Year, Bevolking) %>%
  mutate(Bevolking = as.numeric(Bevolking))

# Merge and calculate net migration as % of population
migratie_bevolking <- migratie_long %>%
  left_join(bevolking_long, by = c("Regio.s", "Year")) %>%
  mutate(
    Netto_Migratie = Immigratie - Emigratie,
    Netto_Migratie_Perc = 100 * Netto_Migratie / Bevolking
```

```
  )

head(migratie_bevolking)
```

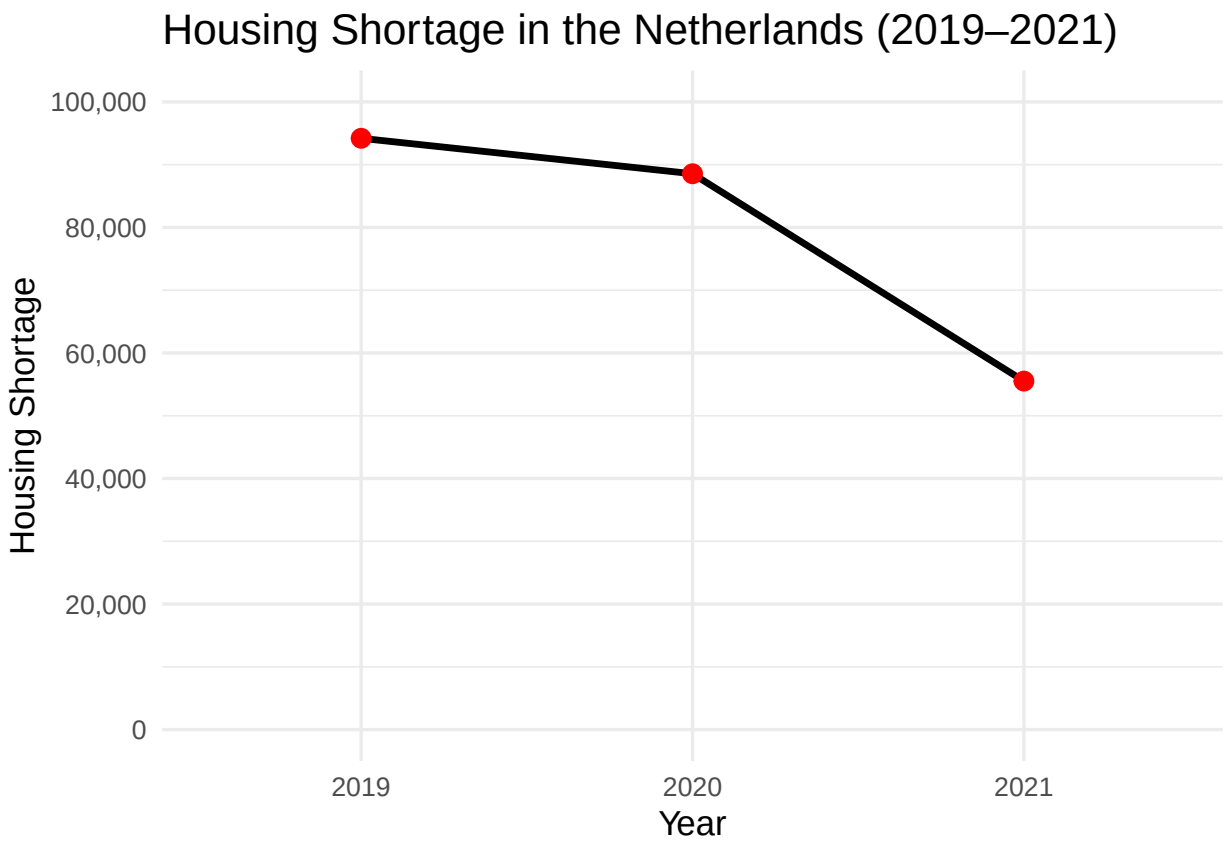
```
## # A tibble: 6 x 7
##   Regio.s        Year  Immigratie Emigratie Bevolking Netto_Migratie
##   <chr>          <chr>      <dbl>     <dbl>     <dbl>          <dbl>
## 1 Groningen (PV) 2019       16664      6239    583990          10425
## 2 Groningen (PV) 2020       14043      6406    585866           7637
## 3 Groningen (PV) 2021       22575      5665    586937          16910
## 4 Fryslân (PV)   2019        4497      2676    647672           1821
## 5 Fryslân (PV)   2020        3513      2532    649957            981
## 6 Fryslân (PV)   2021        3349      2520    651435            829
## # i 1 more variable: Netto_Migratie_Perc <dbl>
```

5.2 3.2 Generate necessary variables

```
# Merge datasets
data_combined <- housing_stock_selected %>%
  select(Regio.s, Year, Voorraad = aantal) %>%
# Replace value_column if needed
  left_join(
    households_selected %>%
      select(Regio.s, Year, Huishoudens = aantal),
# Replace value_column if needed
    by = c("Regio.s", "Year")
  )

# Create shortage variable
data_combined <- data_combined %>%
  mutate(Tekort = Huishoudens - Voorraad)
```

```
# SUM the shortage across all provinces for each year
nederland_shortage <- data_combined %>%
  group_by(Year) %>%
  summarize(Total_Tekort = sum(Tekort, na.rm = TRUE))

# Make year a factor
nederland_shortage$Year <- as.factor(nederland_shortage$Year)

# Create the line plot with fixed y-axis (0 to 100,000)
ggplot(nederland_shortage, aes(x = Year, y = Total_Tekort, group = 1)) +
  geom_line(color = "black", size = 1.2) +
  geom_point(size = 3, color = "red") +
  scale_y_continuous(
```

```
    limits = c(0, 100000),
    breaks = seq(0, 100000, by = 20000),
    labels = scales::comma
  ) +
  labs(
    title = "Housing Shortage in the Netherlands (2019-2021)",
    x = "Year",
    y = "Housing Shortage"
  ) +
  theme_minimal(base_size = 13)
```

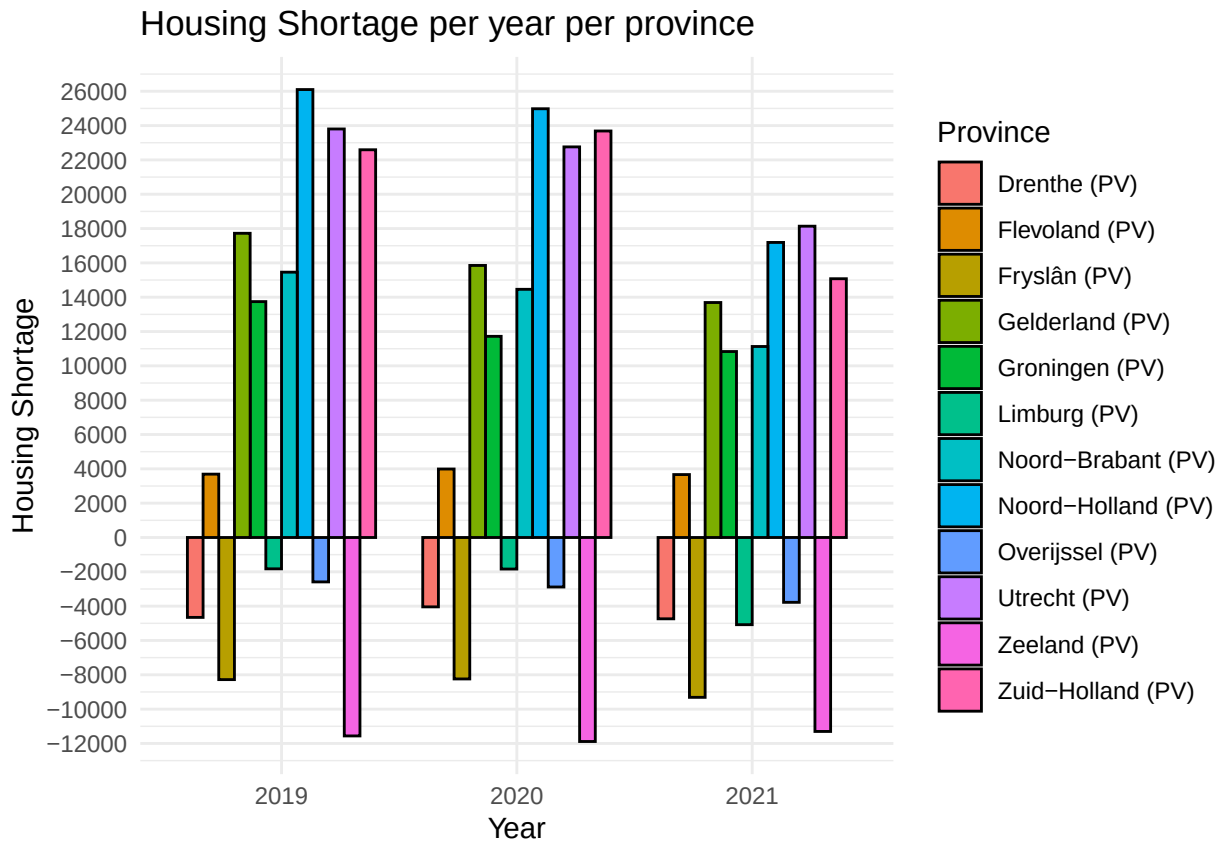


The line plot illustrates the trend of housing shortage in the Netherlands between 2019 and 2021. The red dots and the line represent the estimated magnitude of the shortage in each year. This visual is important because it provides a clear overview of the national trend in housing scarcity over this period. A clear decrease in the housing shortage can be observed between 2019 and 2021, which is a positive development in addressing housing market issues.

5.3 3.3 Visualize temporal variation

```
# Make sure your factor level for year is logically arranged (for the x-axis)
data_combined$Year <- as.factor(data_combined$Year)
```

```r
# Black border, narrower bars optional
ggplot(data_combined, aes(x = factor(Year), y = Tekort, fill = Regio.s)) +
  geom_col(position = "dodge", color = "black", width = 0.8) +
  scale_y_continuous(
    breaks = seq(
      floor(min(data_combined$Tekort, na.rm = TRUE) / 2000) * 2000,
      ceiling(max(data_combined$Tekort, na.rm = TRUE) / 2000) * 2000,
      by = 2000)
  ) +
  labs(
    title = "Housing Shortage per year per province",
    x = "Year",
    y = "Housing Shortage",
    fill = "Province"
  ) +
  theme_minimal()
```



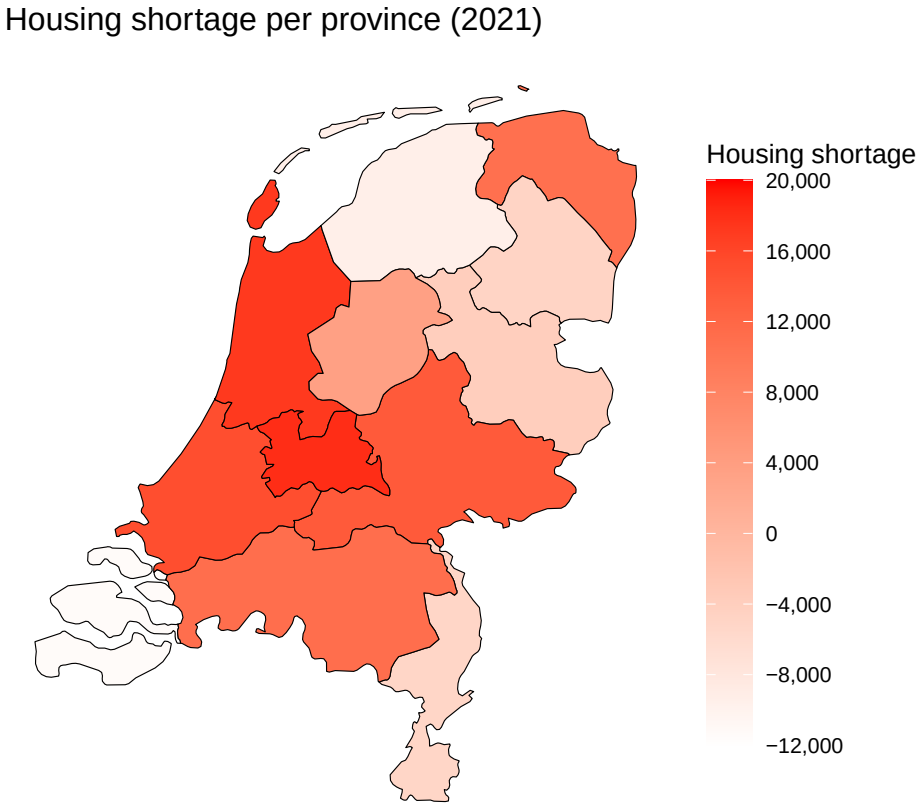
The bar chart visualizing the housing shortage per year per province in the Netherlands from 2019 to 2021. Each group of bars represents a specific year, and within each year, individual bars show the housing shortage (or surplus, if negative) for different provinces, color-coded according to the legend on the right. This visual is crucial for understanding regional disparities and temporal changes in housing supply and demand across the Netherlands. For instance, it allows for direct comparison of housing shortages in provinces like Noord-Holland and Zuid-Holland over the years,

and also highlights provinces that experienced surpluses.

5.4 3.4 Visualize spatial variation

```
# Load shapefile with Dutch provinces
nl_prov <- ne_states(country = "Netherlands", returnclass = "sf")

# Select only the 12 real provinces of the Netherlands
provincies_nederland <- c(
  "Drenthe", "Flevoland", "Friesland", "Gelderland", "Groningen",
  "Limburg", "Noord-Brabant", "Noord-Holland", "Overijssel",
  "Utrecht", "Zeeland", "Zuid-Holland"
)
nl_prov_alleen_prov <- nl_prov %>% filter(name %in% provincies_nederland)

# Create a data frame for 2021 with province column without (PV)
data_2021 <- data_combined %>%
  filter(Year == "2021") %>%
  mutate(provincie_kort = gsub(" \\(PV\\)", "", Regio.s),
         provincie_kort = ifelse(provincie_kort == "Fryslân", "Friesland", provincie_kort),
         )

# Join the shortages to the map
kaart_met_tekort <- nl_prov_alleen_prov %>%
  left_join(data_2021, by = c("name" = "provincie_kort"))

# Plot the map with color gradient white (low) to red (high)
ggplot(kaart_met_tekort) +
  geom_sf(aes(fill = Tekort), color = "black", size = 0.5) +
  scale_fill_gradient(
    low = "white",
    high = "red",
    name = "Housing shortage",
    limits = c(-12000, 20000),
    breaks = seq(-12000, 20000, by = 4000),
    labels = scales::comma
  ) +
  labs(
    title = "Housing shortage per province (2021)",
  ) +
  theme_minimal(base_size = 10) +
  theme(
    legend.key.height = unit(1.5, "cm"),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank()
```

```
)
```

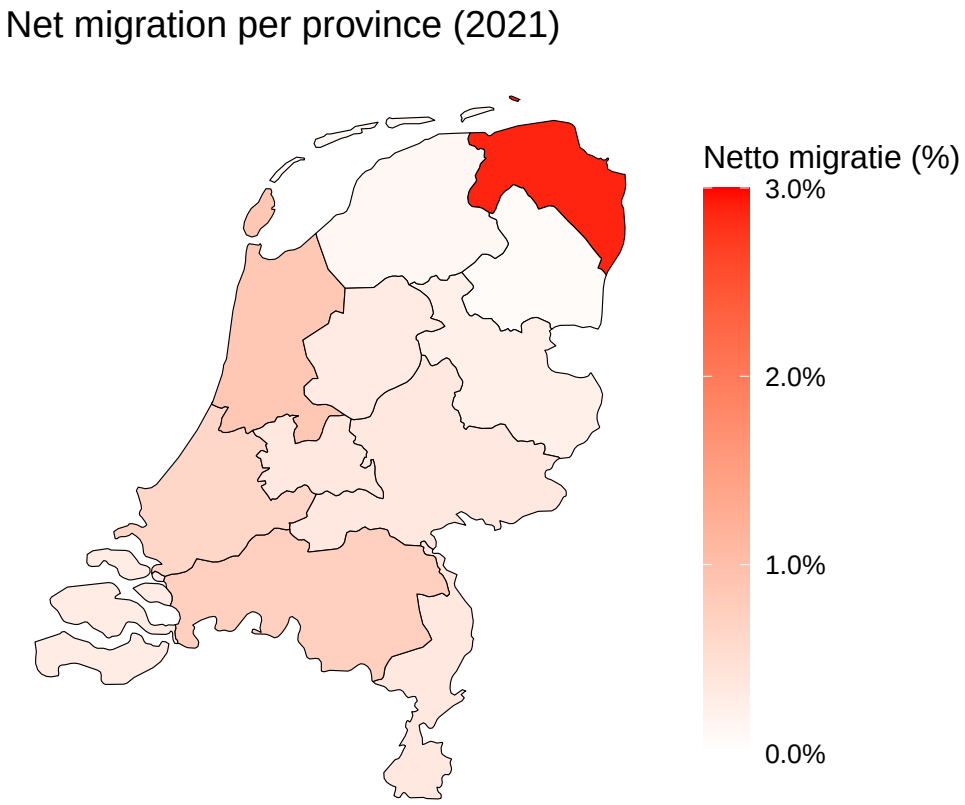


This map visualizes the housing shortage per province in the Netherlands in 2021. Provinces with a higher shortage are marked in darker red, while those with smaller shortages (or even surpluses, indicated by lighter shades tending towards white/light pink and potentially beyond zero if the scale allowed) are lighter. This visual is important because it highlights regional disparities in housing supply and demand. For example, urbanized provinces like Zuid-Holland and Noord-Holland (as well as Utrecht) appear to experience greater shortages, suggesting targeted policy interventions may be needed in these areas. Provinces with lighter shades might indicate a more balanced housing market or even a surplus in 2021.

5.5 3.5 Visualize sub-population variation

```
# 2. Filter to the 12 main provinces
provincies_nederland <- c(
  "Drenthe", "Flevoland", "Friesland", "Gelderland", "Groningen",
  "Limburg", "Noord-Brabant", "Noord-Holland", "Overijssel",
  "Utrecht", "Zeeland", "Zuid-Holland"
)
nl_prov_12 <- nl_prov %>% filter(name %in% provincies_nederland)
```

```r
# 3. Prepare CBS migration data for 2021
data_migratie <- migratie_bevolking %>%
  filter(Year == "2021") %>%
  mutate(
    provincie_kort = gsub(" \\(PV\\)", "", Regio.s),
    provincie_kort = ifelse(provincie_kort == "Fryslân", "Friesland", provincie_kort),
    Netto_Migratie_Perc = Netto_Migratie / Bevolking * 100
  )

# 4. Join shapefile with migration data
kaart_cbs_migratie <- nl_prov_12 %>%
  left_join(data_migratie, by = c("name" = "provincie_kort"))

ggplot(kaart_cbs_migratie) +
  geom_sf(aes(fill = Netto_Migratie_Perc), color = "black", size = 0.4) +
  scale_fill_gradient(
    low = "white", high = "red",
    name = "Netto migratie (%)",
    limits = c(0, 3),
    breaks = seq(0, 3, by = 1),
    labels = scales::percent_format(accuracy = 0.1, scale = 1)
  ) +
  labs(
    title = "Net migration per province (2021)",
    fill = "Migration %"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    legend.key.height = unit(1.5, "cm"),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank()
  )
```



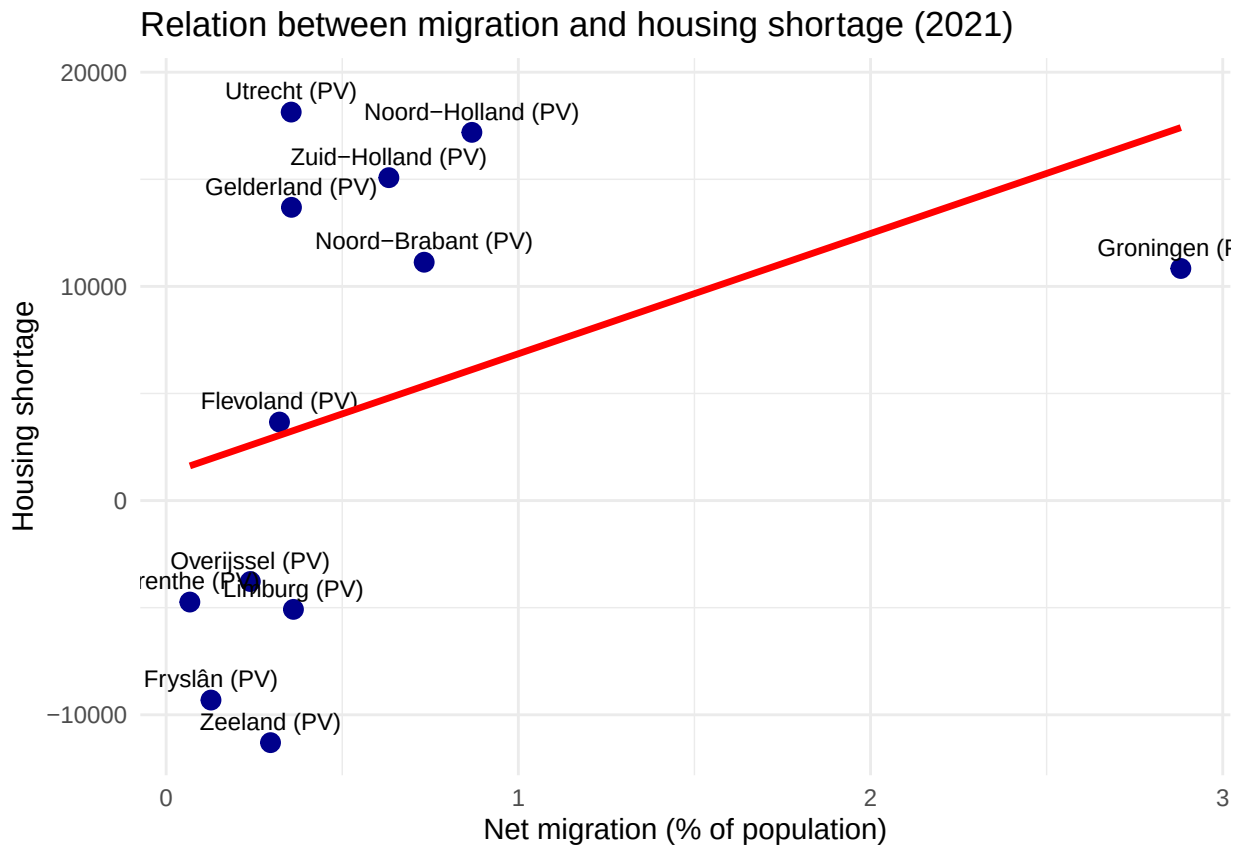
This map displays the net migration per province in the Netherlands for the year 2021. Provinces with higher net migration percentages are depicted in darker red, while those with lower percentages are lighter. This visualization is important as it illustrates the geographical distribution of population movement within the Netherlands, highlighting areas that are experiencing significant population growth or decline due to migration. For example, the province of Groningen appears to have a notably higher net migration percentage, suggesting it attracted more residents than it lost in 2021, which could have implications for local housing demand and infrastructure.

5.6 3.6 Event analysis

Analyze the relationship between two variables.

```
# Clean Regio.s before the join
event_analysis_data <- data_combined %>%
  filter(Year == "2021") %>%
  mutate(provincie_kort = gsub(" \\(PV\\)", "", Regio.s),
         provincie_kort = ifelse(provincie_kort == "Fryslân", "Friesland", provincie_kort)) %>%
  left_join(
    migratie_bevolking %>%
      filter(Year == "2021") %>%
      mutate(
        provincie_kort = gsub(" \\(PV\\)", "", Regio.s),
        provincie_kort = ifelse(provincie_kort == "Fryslân", "Friesland", provincie_kort),
```

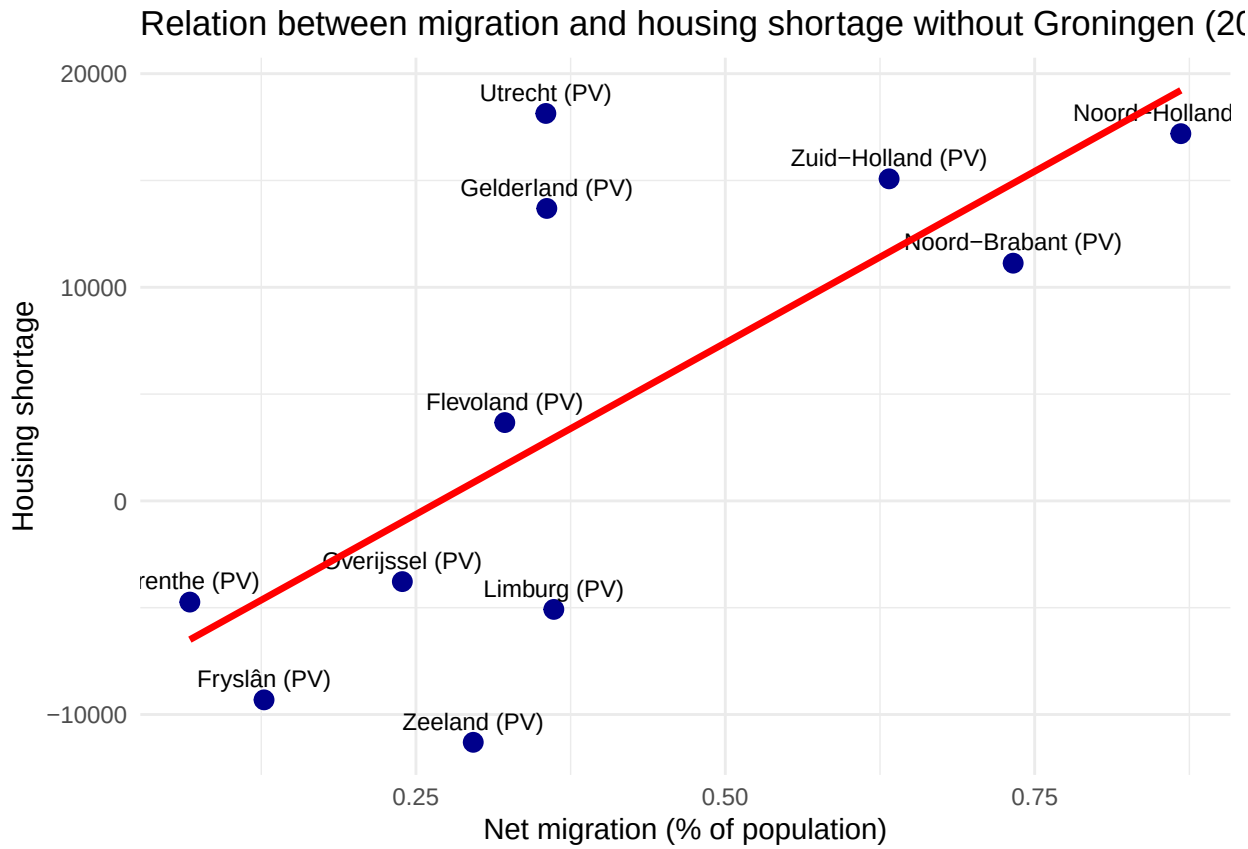
```
        Netto_Migratie_Perc = Netto_Migratie / Bevolking * 100
      ) %>%
      select(provincie_kort, Netto_Migratie_Perc),
    by = "provincie_kort"
  )

# Clean for plotting
event_analysis_data_clean <- event_analysis_data %>%
  filter(!is.na(Netto_Migratie_Perc), !is.na(Tekort))

# Plot with regression line and province labels
ggplot(event_analysis_data_clean, aes(x = Netto_Migratie_Perc, y = Tekort, label = Regio.s)) +
  geom_point(color = "darkblue", size = 3) +
  geom_text(nudge_y = 1000, size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "red", linewidth = 1.2) +
  labs(
    title = "Relation between migration and housing shortage (2021)",
    x = "Net migration (% of population)",
    y = "Housing shortage"
  ) +
  theme_minimal(base_size = 11)
```



This scatter plot illustrates the relationship between net migration (as a percentage of the

population) and housing shortage for each province in the Netherlands in 2021. Each blue point represents a province, labeled with its name, showing its respective net migration on the x-axis and housing shortage on the y-axis. The red line represents a trend line, suggesting a positive correlation between net migration and housing shortage. This visual is important because it indicates that provinces experiencing higher net migration tend to also face greater housing shortages, implying that population influx is a significant factor contributing to housing demand and scarcity in certain regions. For instance, Groningen, with a high net migration, also shows a substantial housing shortage, while provinces like Zeeland and Fryslân, with lower net migration, appear to have a housing surplus.

```
# Create dataset and exclude Groningen
event_analysis_data <- data_combined %>%
  filter(Year == "2021") %>%
  mutate(
    provincie_kort = gsub(" \\(PV\\)", "", Regio.s),
    provincie_kort = ifelse(provincie_kort == "Fryslân", "Friesland", provincie_kort)
  ) %>%
  filter(provincie_kort != "Groningen") %>%
# Removing Groningen
  left_join(
    migratie_bevolking %>%
      filter(Year == "2021") %>%
      mutate(
        provincie_kort = gsub(" \\(PV\\)", "", Regio.s),
        provincie_kort = ifelse(provincie_kort == "Fryslân", "Friesland", provincie_kort),
        Netto_Migratie_Perc = Netto_Migratie / Bevolking * 100
      ) %>%
      select(provincie_kort, Netto_Migratie_Perc),
    by = "provincie_kort"
  )

# Filter on valid rows
event_analysis_data_clean <- event_analysis_data %>%
  filter(!is.na(Netto_Migratie_Perc), !is.na(Tekort))

# Create scatterplot with labels
ggplot(event_analysis_data_clean, aes(x = Netto_Migratie_Perc, y = Tekort, label = Regio.s)) +
  geom_point(color = "darkblue", size = 3) +
  geom_text(nudge_y = 1000, size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "red", linewidth = 1.2) +
  labs(
    title = "Relation between migration and housing shortage without Groningen (2021)",
    x = "Net migration (% of population)",
    y = "Housing shortage"
  ) +
  theme_minimal(base_size = 11)
```



This scatter plot displays the relationship between net migration (as a percentage of the population) and housing shortage for Dutch provinces in 2021, with the province of Groningen explicitly excluded. Each blue point represents a province, showing its net migration on the x-axis and housing shortage on the y-axis. The red line indicates a positive linear trend, suggesting that provinces with higher net migration percentages generally experience larger housing shortages. This visualization is important because by excluding Groningen (likely due to the presence of a large asylum seekers' center, which might skew migration data and its impact on the housing market), it provides a potentially clearer view of the underlying relationship between general migration patterns and housing shortages across the other provinces. It still shows that provinces like Utrecht, Noord-Holland, and Zuid-Holland, with relatively higher net migration, also face significant housing shortages.

6 Part 4 - Discussion

6.1 4.1 Discuss your findings

Analysis of Housing Shortage and Migration in the Netherlands (2019-2021)

The analysis of the graphs confirms a persistent and locally housing shortage in the Netherlands between 2019 and 2021. While there's a slight national decrease, the data clearly shows that urban and western provinces, such as Noord-Holland, Zuid-Holland, and Utrecht, are the most severely affected.

Migration patterns significantly exacerbate this issue. There is a clear correlation: regions experiencing higher net migration also report larger housing shortages. This indicates a demand-driven pressure on the housing market, caused by demographic growth and the attractiveness of these regions. The outlier of Groningen (possibly due to an asylum seekers' center) shows that specific migration factors can have a large local impact, but the trend remains clear even without this province.

This suggests that policy measures should prioritize housing development in high-demand areas, and perhaps also strategies to distribute population growth more evenly across the country.

7 Part 5 - Reproducibility

7.1 5.1 Github repository link

https://github.com/ProgrammingforEcon-Team-5-Groep-1/Programming-Group5-Team1

7.2 5.2 Reference list

Centraal Bureau voor de Statistiek. (2025, May 23). Stock of dwellings (and non-dwellings). CBS StatLine. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81955NED/table?fromstatweb

Centraal Bureau voor de Statistiek. (2025, June 3). Households. CBS StatLine. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/71486ned/table?fromstatweb

Centraal Bureau voor de Statistiek. (2025, May 28). Population development. CBS StatLine. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37230ned/table?ts=1750160957830

Centraal Bureau voor de Statistiek. (2025, May 28). Population development. CBS StatLine. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37230ned/table?ts=1750159602778

Centraal Bureau voor de Statistiek. (2025). Housing. CBS. https://www.cbs.nl/en-gb/visualisations/monitor-broad-prosperity-and-the-sustainable-development-goals/here-and-now/housing#:~:text=Door%20ABF%20Research%20is%20voor,een%20kleine%20aanvullende%20aardgaslevering%20

Langen, M. (2025, January 16). Housing market – building according to need. ABN AMRO Bank. https://www.abnamro.com/research/en/our-research/housing-market-building-according-to-need