

# Final Report

## “The Impact of Refugees on House Prices in European Countries”

### Group Members

Ahmed Albarqi (aa5550)

Muhammad Usman Anwar Goraya (mag2385)

Sarthak Gulati (sg4450)

2025-05-09

## 1 Introduction

In recent decades, global conflicts, political instability, and humanitarian crises have led to a significant rise in refugee movements across the world. Prolonged wars in Syria, Afghanistan, and Ukraine; ethnic and religious persecution in countries like Myanmar and Eritrea; and widespread violence & the unlawful occupation in regions such as Palestine have forced millions of individuals to seek safety beyond their national borders. The European Union (EU) presents an interesting case with regards to being an area of protection for individuals who are seeking protection and refuge. EU states recognize asylum as a fundamental human right under the 1951 Geneva Convention (European Commission, Migration & Home Affairs 2021). In order to share responsibility to welcome asylum seekers in a dignified manner, EU member states have established the Common European Asylum System (CEAS) in 1999. However the year 2015 saw a record number of refugees and asylum seekers emigrating to the EU seeking asylum and refuge, which made the EU to consider revamping its migration policies as a whole.

The arrival of refugees in European countries has generated a debate on its economic and social impacts. One key concern is the effect of refugee inflows on housing markets, particularly house prices. While some argue that increased demand from refugees raises house prices, others go on to suggest that concentration in low-income areas may depress local markets. This research aims to examine the relationship between refugee intake and house price dynamics in different European countries. In this context, countries which have restrictive refugee policies can act as a control group while countries which have more open integration policies for refugees will be a treatment group. The policy shock that will be exploring will be increased acceptance of refugees post the year 2015-16 by the European countries.

In view of the above, and in order to explore the relationship between the variable under scrutiny, our research questions will be as follows:

- **How does the variation in refugee intake amongst European countries impact house prices?**

## 2 Policy Background

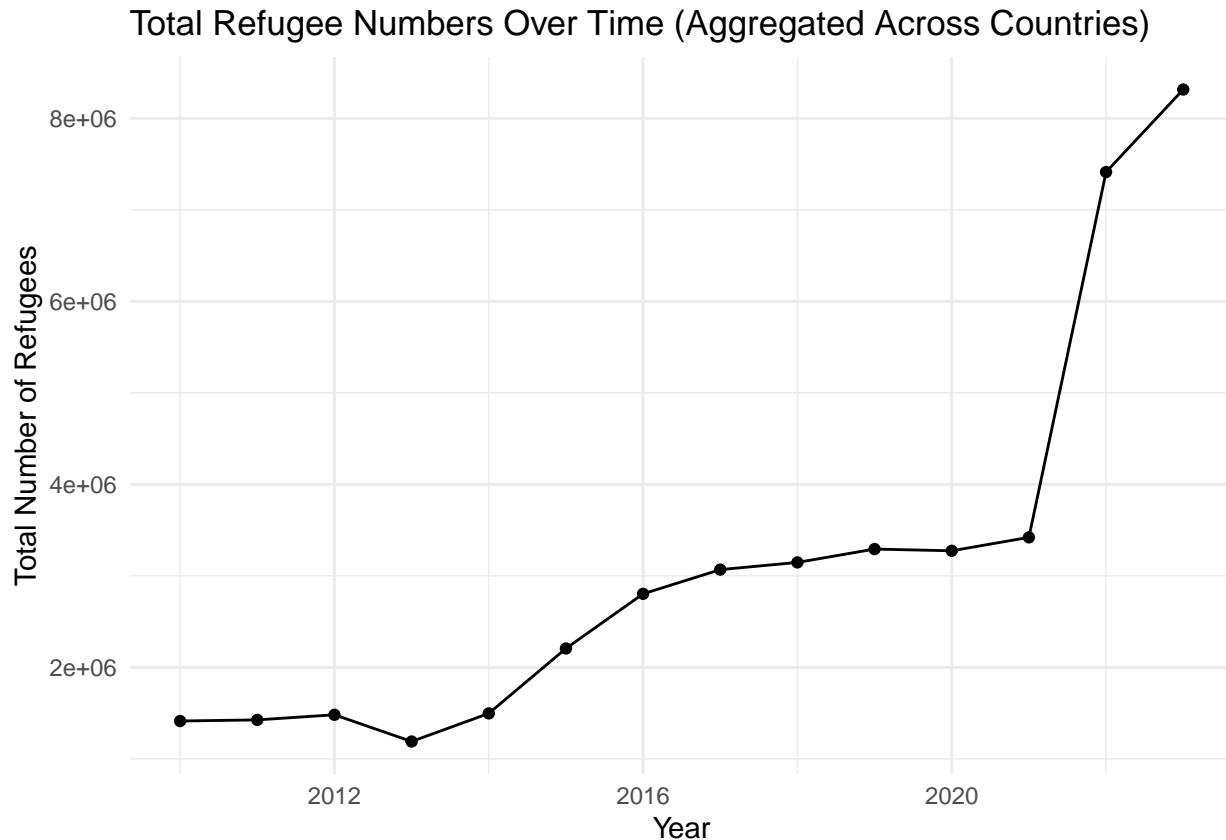
The onset of civil war in Syria in 2011 and other global crises forced millions of people to flee to neighboring states and Europe. Migrants fled from the conflict struck states such as Syria, Afghanistan, Iraq, and some even closer countries like Albania and Kosovo. As discussed prior, in the year 2015, European countries witnessed an almost unprecedented record influx of refugees - with some estimates quoting 1.3 million individuals arriving in Europe (Pew Research 2016) . In this backdrop the European governments forced

through a deal to impose refugee quotas, sharing 120,000 people between them in a watershed decision that several states bitterly opposed.

The relocation would be done according to a mandatory distribution key using objective and quantifiable criteria (40% of the size of the population, 40% of the GDP, 10% of the average number of past asylum applications, 10% of the unemployment rate). It applies to nationalities of applicants with an EU-wide average recognition rate of 75% or higher.

## 2.1 Refugees Landscape

Referring to the graph below, we can observe the asylum applicants across EU countries overtime. At the year 2015 we can see a huge shock in the asylum applications



## 2.2 Differential Country Policies

It was observed that Germany, Hungary and Sweden were the leading destinations for the refugees while the United Kingdom (UK) and France saw fewer applications of asylum seekers and refugees (Pew Research 2016). This makes for a good case of comparison in research analysis which will be discussed in detail further in an ensuing section. While access to the labor market may be restricted by administrative (e.g. work permits, authorizations or visas) and practical barriers (e.g. language, social skills, etc.), refugees have the legal right to work in all OECD countries. For asylum seekers, the rights and obligations differ significantly from one country to another. Most countries, however, grant access to the labour market to some groups of applicants on certain conditions, including a prior waiting period, ranging between 2 months in Italy to 12 months in the United Kingdom.

### 3 Data description

The UN Refugee Agency, namely United Nations High Commissioner for Refugees (UNHCR), collates population data relating to persons who are forcibly displaced or stateless. The data is sourced primarily from governments and also from UNHCR operations. Based on UNHCR definitions (included in Appendix), our data set will be focused on the number of refugees and asylum seekers referenced by their host countries. To have a fair comparison across the EU countries, we will use the number of refugees per 1000 residents.

We use data for housing prices from the Eurostat database. The House Price Index (HPI) measures inflation in the residential property market. The HPI captures price changes of all types of dwellings purchased by households (flats, detached houses, terraced houses, etc.). Only transacted dwellings are considered, self-build dwellings are excluded. The land component of the dwelling is included. The HPI is available for all European Union Member States (except Greece), the United Kingdom (only until the third quarter of 2020), Iceland, Norway, Switzerland and Türkiye.

### 4 Descriptive Findings

As shown before in the graph, the number of refugees saw substantial increases during the period 2014-16 and 2021-23, and that can be due to:

#### 2014-16

- This surge likely corresponds to the heightened migration flows related to global crises and wars in the middle east and other countries.
- This year also EU countries decided to open their borders and accept more refugees because of the unprecedented number of refugees.

#### 2021-23

- Probably increased global tensions and Covid-19 pandemic played a role in this but this needs further investigation.

After identifying our period of interest, now we will examine the independent variable, which is Refugees inflows per 1000 residents.

starting with simple summary statistics of the variable from the year 2010 - 2021, the summary statistics are as follows:

Table 1: Summary Statistics for Refugees per 1000 Residents by Year

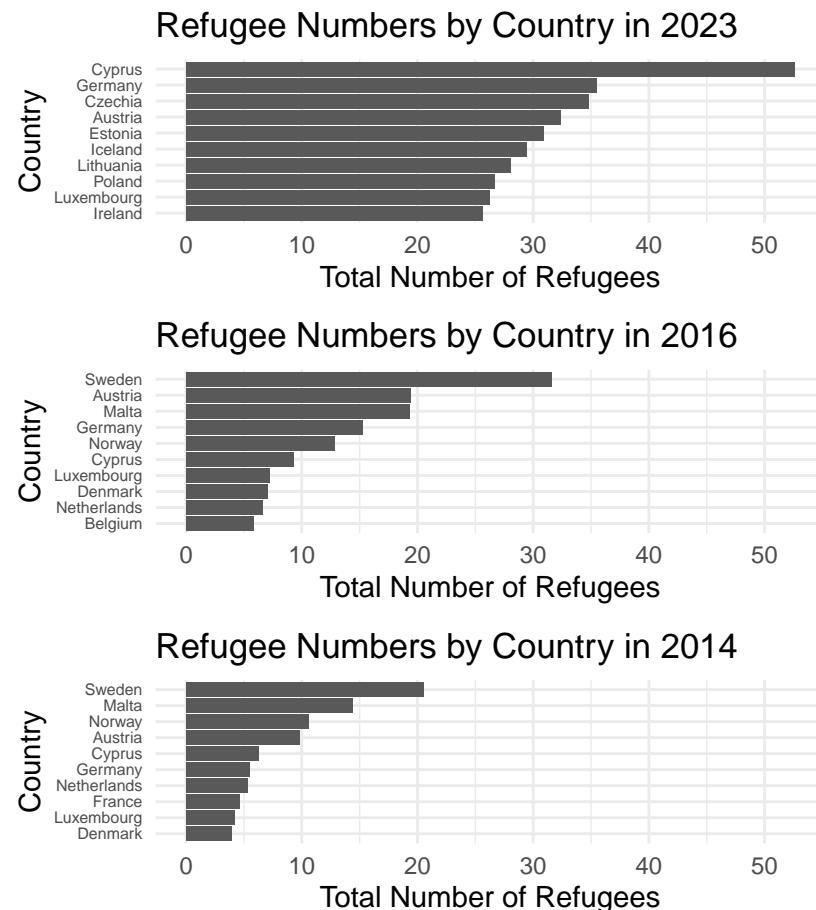
Year	Count	Mean	SD	Min	Q1	Median	Q3	Max
2010	28	3.47	4.51	0.03	0.23	0.98	5.90	17.91
2011	28	3.51	4.78	0.03	0.22	1.08	5.24	20.17
2012	28	3.52	5.01	0.04	0.22	0.91	5.02	21.42
2013	28	3.52	5.57	0.05	0.20	1.35	4.08	25.35
2014	28	3.69	4.89	0.10	0.27	2.26	4.81	20.52
2015	28	5.46	7.47	0.13	0.33	3.28	6.79	33.32
2016	28	5.76	7.65	0.15	0.39	3.08	7.11	31.56
2017	28	5.82	7.48	0.16	0.41	2.83	6.53	29.09
2018	28	6.05	7.60	0.18	0.40	2.93	6.79	28.12

Year	Count	Mean	SD	Min	Q1	Median	Q3	Max
2019	28	6.61	8.21	0.18	0.44	3.77	7.13	27.41
2020	28	6.81	8.34	0.17	0.39	3.85	8.51	26.48
2021	28	7.20	8.80	0.20	0.48	4.60	8.78	31.59
2022	28	19.05	11.53	3.67	9.93	17.14	26.81	49.41
2023	28	20.60	11.38	4.65	11.71	21.18	27.02	52.57

The summary statistics reveal a clear upward trend in the refugee influx across European countries from 2010 to 2023. The mean number of refugees per 1000 residents has steadily increased from 3.47 in 2010 to 20.60 in 2023, with significant fluctuations in standard deviation (SD), reflecting varying levels of influx across countries. The years 2022 and 2023 show a sharp rise in the median values, reaching 17.14 and 21.18, respectively, accompanied by higher Q1, median, and Q3 values, indicating a broader shift towards higher refugee arrivals. Notably, the maximum values for these years (49.41 in 2022 and 52.57 in 2023) suggest a few countries experienced exceptional levels of refugee influx.

#### 4.1 Refugees Landscape across the EU Countries

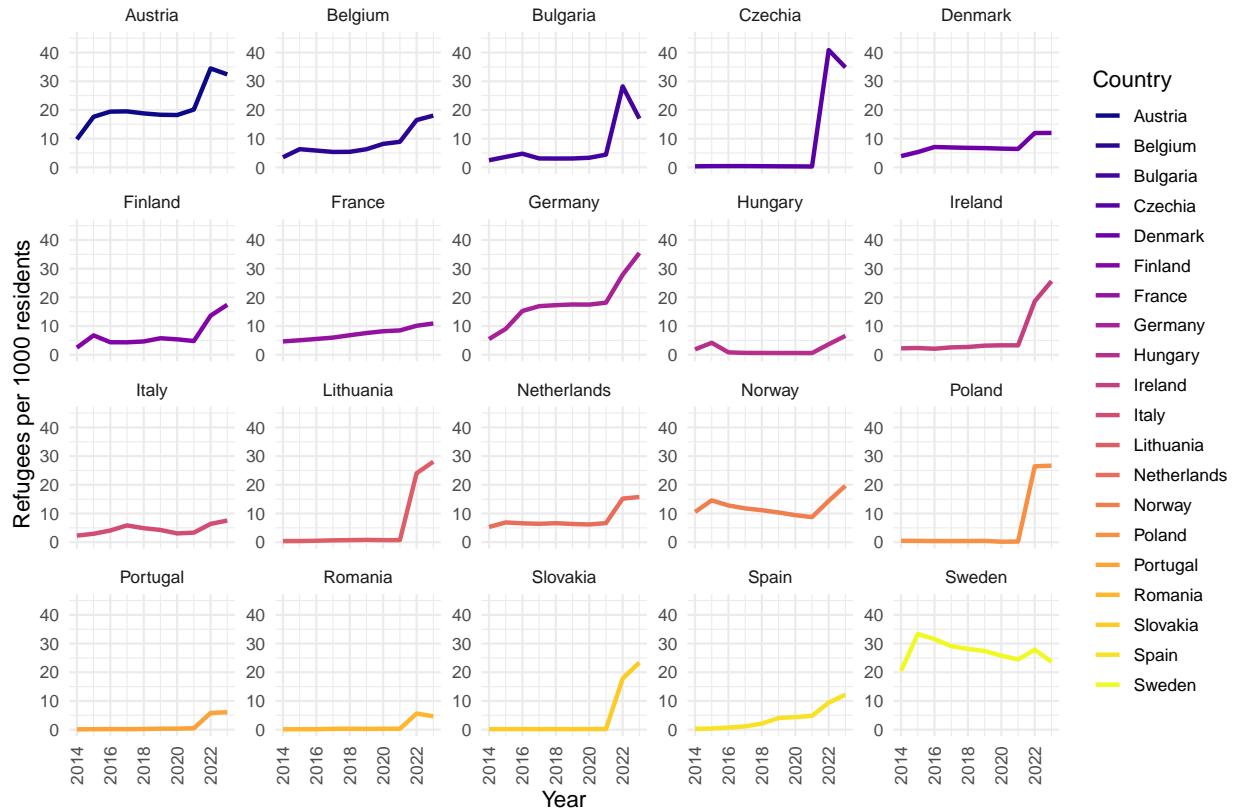
Now we will examine the trends of the refugees influx for EU countries for our period of interest 2014-2016-2023



We can observe that countries like Sweden, Malta, Norway, and Austria admitted the highest number of the refugees per 1000 residents over the years 2014-16.

However, looking the recent data in 2023, Cyprus, Germany, Czechia and Estonia emerge as the leading receptors of refugees. Also, there is some evidence of convergence among number of refugees accepted by countries in the range of 25-35 refugees per 1000 residents.

### Refugee per 1000 residents trends in Top Countries (2014–2023)



Looking the faceted plot, we see while some countries have seen gradual increasing trends in refugees per 1000 residents, others have seen sharp rise in the last 3-4 years.

Next, we classified countries into three groups based on their average number of refugees per 1,000 residents over the years 2010–2023. The thresholds were derived from the summary statistics (specifically, the median and 3rd quartile “because the data is right skewed”) of the variable “refugees per 1,000 residents”:

Table 2: Average Refugees per 1000 Residents (2010–2023) and Influx Classification

Country	avg_refugees_per_1000	refugee_influx_group
Sweden	22.90	High
Malta	21.46	High
Cyprus	18.61	High
Austria	17.38	High
Germany	14.91	High
Norway	11.82	High
Luxembourg	9.67	High
Netherlands	7.26	Medium
Belgium	7.00	Medium
France	6.39	Medium
Denmark	6.08	Medium
Czechia	5.71	Medium
Finland	5.58	Medium
Ireland	5.51	Medium
Bulgaria	5.47	Low
Iceland	5.00	Low
Estonia	4.55	Low
Lithuania	4.16	Low
Poland	4.15	Low
Italy	3.55	Low
Latvia	3.48	Low
Slovakia	3.09	Low
Spain	2.86	Low
Hungary	1.59	Low
Croatia	1.12	Low
Portugal	1.01	Low
Slovenia	0.98	Low
Romania	0.89	Low

High Refugee Influx: Average above 7.8631702 (above the 3rd quartile, Q3)

Medium Refugee Influx: Average between 5.4905071 and 7.8631702 (between the 1st and 3rd quartiles, Q1 and Q3)

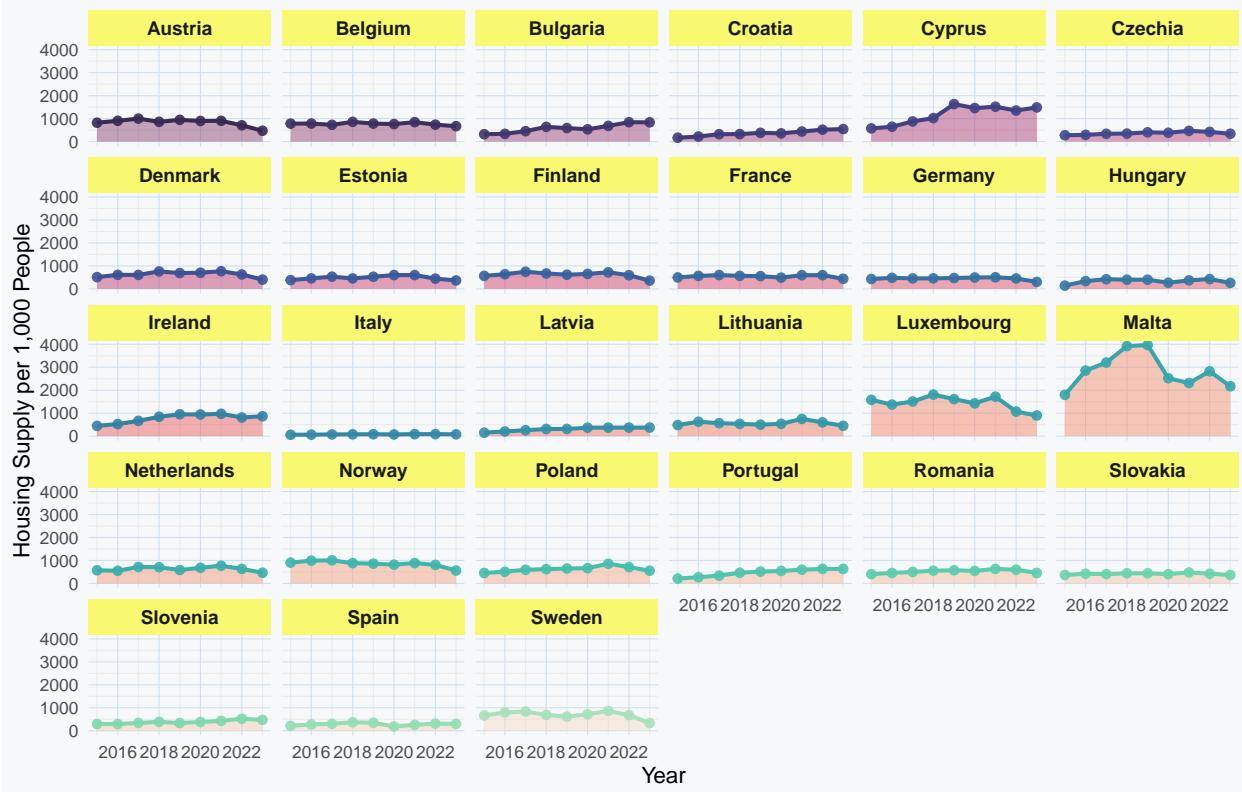
Low Refugee Influx: Average below 5.4905071 (below the 1st quartile, Q1)

This classification method ensures a balanced and data-informed grouping of countries, highlighting those with unusually high or low refugee influxes relative to the sample distribution.

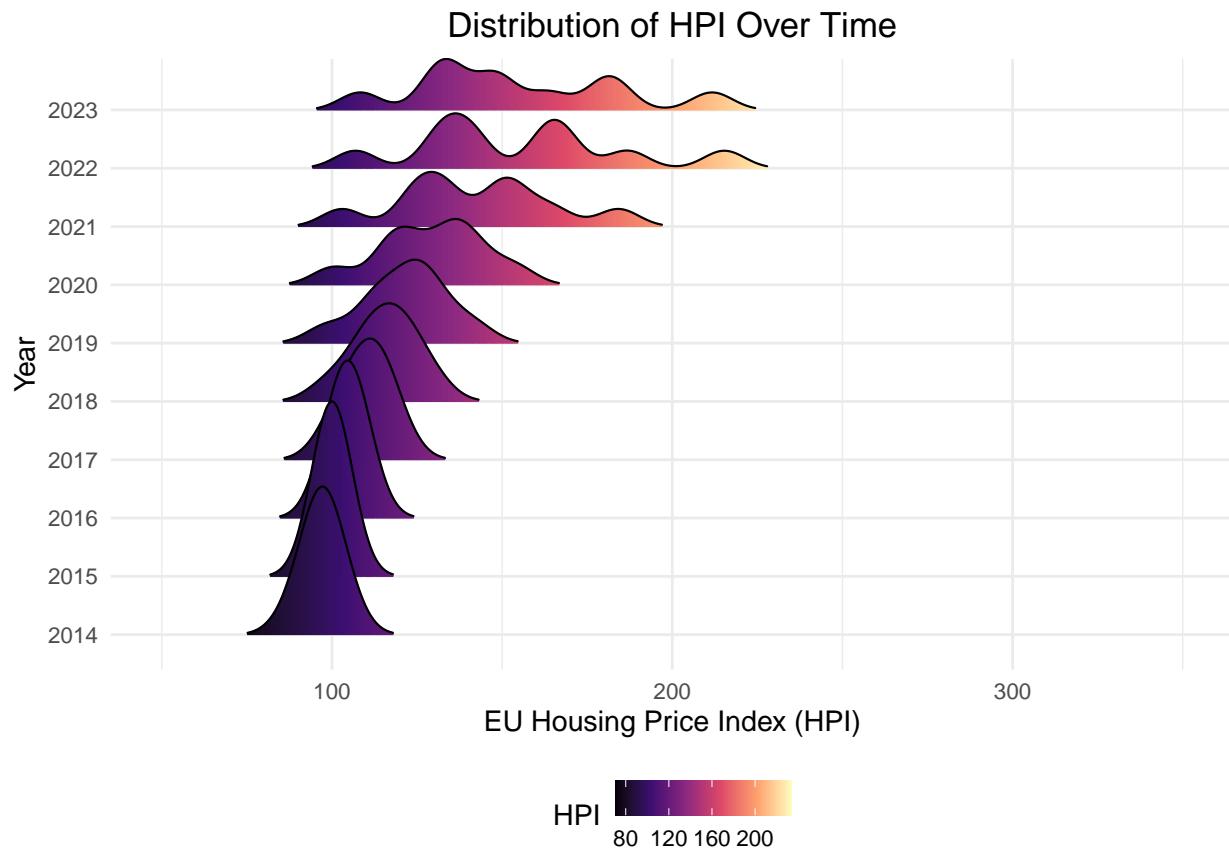
Let’s now turn our focus to the Housing Prices and the Housing Price Index (HPI). The HPI is used as proxy for the housing prices for the purpose of this report. The following data visualization presents how different countries had their housing supply change over time.

## Housing Supply per 1,000 People Across Countries (2014 – 2023)

Standardized by population for a more meaningful comparison

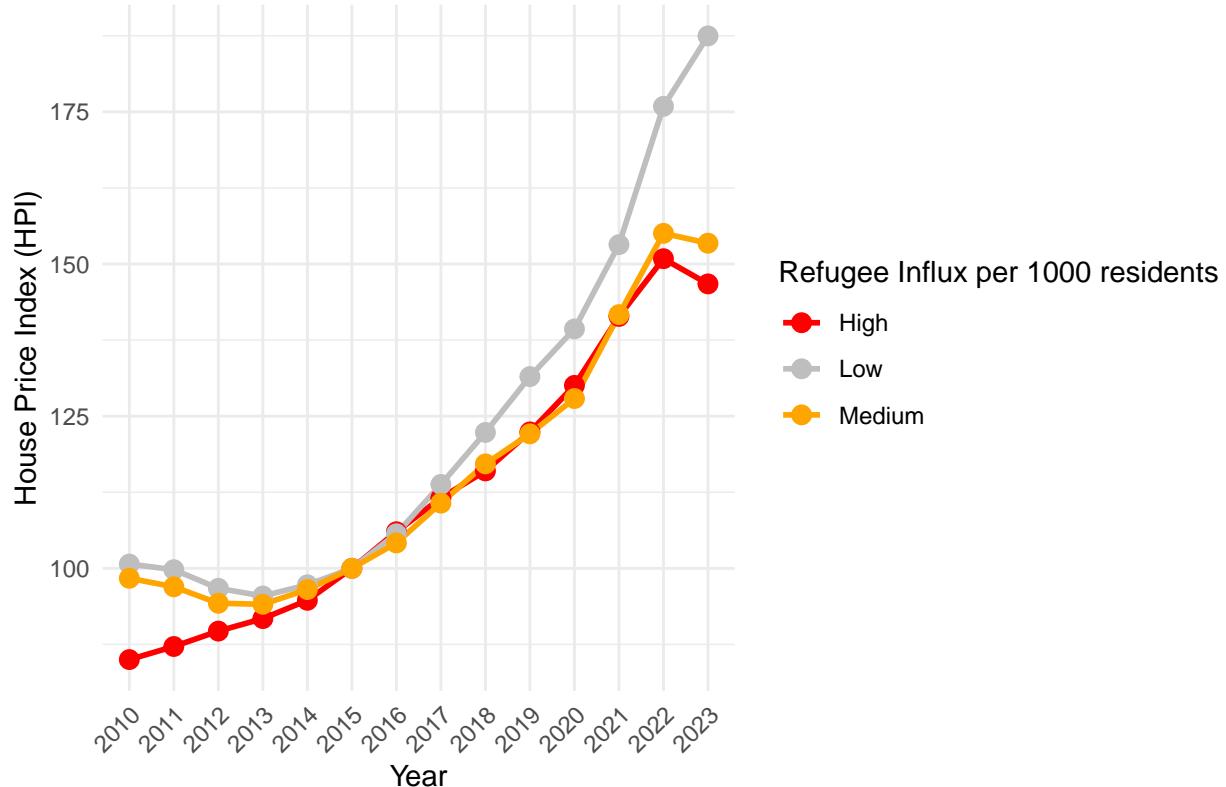


After looking at the housing supply in different EU countries, in this section, we will have a look at what the Housing Price Index (HPI) with the help of a Ridge Plot. This visualizes how the EU HPI is distributed across the years 2014 to 2023 for countries with high refugee inflows. Each ridge represents the distribution of HPI values in a given year, with the height of the curve indicating the density (concentration) of countries around particular HPI values. The color gradient—ranging from dark purple to orange—corresponds to the HPI scale, enhancing visual separation of low and high housing price levels. Most interestingly, this plot helps reveal shifts in housing price trends over time: for example, the ridges shifting rightward (toward higher HPI values), indicates a general increase in housing prices over the decade without the impact or effect of refugee intake by EU countries.



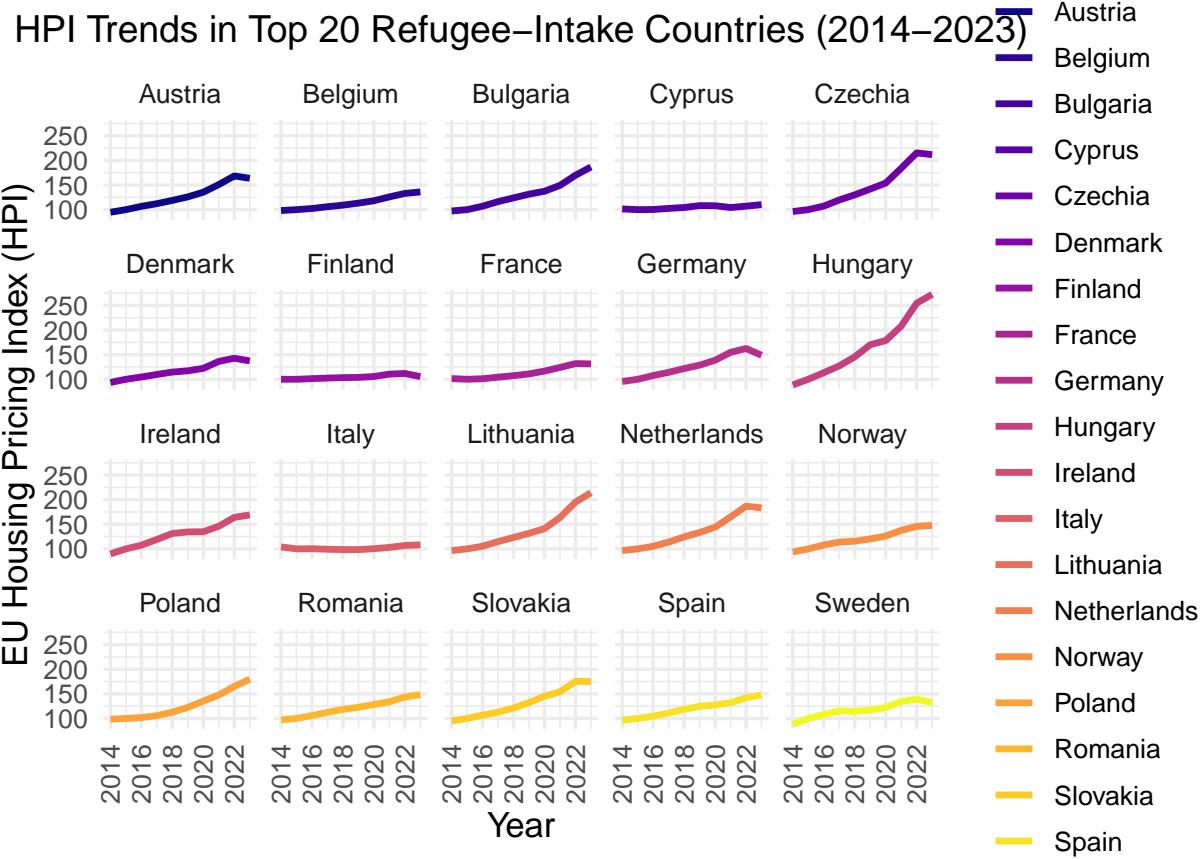
Now we will examine the average House Price Index (HPI) over the years 2010 to 2021 to identify trends across the three refugee influx groups: High, Medium, and Low. This will help visualize how housing markets evolved over time within each group.

## HPI Over Time (2010–2023) by Refugee Influx Groups



Countries with medium and low refugee influx saw faster growth in housing prices after 2016, while those with high refugee influx experienced more moderate growth. This may imply that a higher influx of refugees could be associated with housing market stabilization effects or greater public housing intervention, but further analysis would be needed to isolate causal mechanisms. Additionally, we can observe that the trend before the 2015 shock wasn't consistent across these groups.

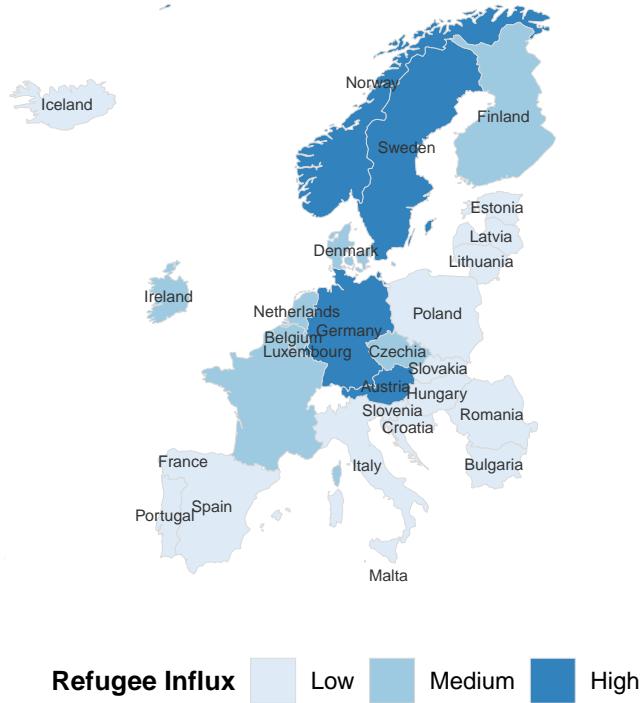
While remaining on the data visualization front of HPI, we can also decipher the country wise trends in those nations which accepted the highest number of refugees from the year 2014 to 2023 in the form of a Faceted line plot, which is as follows:



In this section we will examine the refugees influx spatially across European countries from 2010 to 2023. This spatial analysis will help identify potential spatial patterns, clusters, and trends in refugee movement, facilitating deeper insights into refugee patterns across Europe.

# Refugee Influx Groups in Europe (2010–2023)

Average refugee influx per 1,000 residents



Interestingly, while one might expect that countries in closer geographical proximity would experience similar refugee influxes, the map suggests otherwise. It challenges the assumption that neighboring countries would necessarily share similar refugee patterns, as countries farther apart may exhibit comparable levels of refugee arrivals.

Scatter Plot of HPI Growth vs. Refugees\_per\_1000 residents Change



Next, we try to do some exploratory analysis to understand the relationship between growth of HPI and change in refugees per 1000 residents. Scatterplot indicates weak correlation between the two variables.

## 5 Empirical Strategy

The empirical analysis which our group intends to employ, seeks to explore the causal relationship between the influx of refugees and asylum seekers in EU countries and subsequent changes in housing prices (proxied by the EU Housing Price Index - HPI). In order to explore the said relationship, we have constructed an entity-year panel dataset (country-year observations), enabling the exploitation of policy variation both over time and across EU member states.

Further, to address the potential endogeneity of refugee inflows and the resulting omitted variable bias, we plan to implement a shift-share instrumental variable approach. Consequently we have constructed an instrument that interacts past refugee settlement shares across EU countries (the “share”) with total EU-wide refugee inflows by country of origin in a given year (the “shift”). This allows us to isolate plausibly exogenous variation in refugee allocation driven by historical settlement patterns rather than contemporary economic or policy conditions on the refugee intake front. This approach will strengthen the identification of the causal impact (if any) of refugee inflows on housing prices.

## 5.1 Regression Approach

The primary specification is a two-way fixed effects model, accounting for country and year fixed effects, structured as follows:

$$\text{HPIGrowth}_{ct} = \alpha_c + \gamma_t + \beta_1 \cdot \text{RefugeeGrowth}_{ct} + \epsilon_{ct}$$

Where:

- $\text{HPI}_{ct}$ : Housing Price Growth in country  $c$  at year  $t$
- $\alpha_c$ : Country fixed effects (controls for time-invariant country-specific factors)
- $\gamma_t$ : Year fixed effects (controls for time-specific shocks across countries)
- $\text{RefugeeGrowth}_{ct}$ : Growth in refugees and asylum seekers
- $\epsilon_{ct}$ : Error term

	Fixed Effects Only	Full Model
(Intercept)	-1339.977*** (132.534)	-1350.308*** (147.205)
Refugee_Growth		0.001*** (0.000)
Refugee_Growth_Lag1		-0.001** (0.000)
Num.Obs.	364	336
R2	0.386	0.438
R2 Adj.	0.335	0.383
AIC	2187.9	1979.8
BIC	2304.8	2101.9
Log.Lik.	-1063.958	-957.890
F	7.536	7.937
RMSE	4.50	4.19

+ p <0.1, \* p <0.05, \*\* p <0.01, \*\*\* p <0.001

## 6 Findings

### 6.1 Regression Results

Here we present the results of the two regression models to explore the relationship between refugee growth and housing price index (HPI) growth.

We estimate two models: one with only fixed effects (for country and year) and another including both refugee growth and lagged refugee growth.

The “Full Model” shows that higher current refugee growth is linked to a significant increase in the outcome variable, while refugee growth in the previous year is linked to a significant decrease. The “Full Model,” which includes more factors than just fixed effects, explains the data better than the “Fixed Effects Only” model.

### 6.2 Multicollinearity

Examining the Multicollinearity of our variables with VIF

Table 3: Variance Inflation Factor (VIF) Table

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
Refugee_Growth	1.13	1	1.06
Refugee_Growth_Lag1	1.15	1	1.07
Country	1.17	27	1.00
Year	1.10	1	1.05

The GVIF results indicate no serious multicollinearity concerns in the model. Both Refugee\_Growth and its lag have GVIF<sup>1/(2\*Df)</sup> values around 1.06–1.07, well below the commonly used thresholds of 2 or

	Fixed Effects Only	Full Model (df_fin)	Full Model + Housing
(Intercept)	-1339.977*** (132.534)	-1350.308*** (147.205)	-1097.375*** (326.683)
Refugees_per_1000_Growth		0.066*** (0.019)	0.046*** (0.012)
Refugees_per_1000_Growth_Lag1		-0.065** (0.020)	-0.023+ (0.012)
Housing_per_1000_Growth			10.172*** (1.730)
Housing_per_1000_Growth_lag1			5.401*** (1.568)
Num.Obs.	364	336	189
R2	0.386	0.438	0.581
R2 Adj.	0.335	0.383	0.498
AIC	2187.9	1979.8	1047.3
BIC	2304.8	2101.9	1154.3
Log.Lik.	-1063.958	-957.890	-490.657
F	7.536	7.937	7.028
RMSE	4.50	4.19	3.25

+ p <0.1, \* p <0.05, \*\* p <0.01, \*\*\* p <0.001

5. The fixed effects for Country and Year also show very low adjusted GVIFs ( 1.00 and 1.05 respectively), suggesting that the inclusion of these categorical variables does not inflate variance significantly. Overall, multicollinearity is not a concern in this specification.

### 6.3 Regression with the addition of another controlling variable

We added another variable, which is the housing supply across the countries, we got the below results:

The results suggest that refugee inflows have a significant impact on housing price growth in the EU, with both current and past refugee growth influencing housing dynamics. Specifically, current refugee inflows lead to higher housing price growth, as seen in both the Full Model and the Full Model + Housing, where a positive relationship is established. This likely reflects the demand for housing that arises when refugee populations increase, either through direct housing needs or indirectly due to the economic integration of refugees in the host country. On the other hand, lagged refugee inflows show a negative effect on housing growth, suggesting that initial boosts in housing demand may be followed by a lagged response, possibly due to market adjustments, overbuilding in earlier periods, or delayed policy interventions.

The inclusion of housing supply growth control suggests positive effect of refugees dominate, however, the positive coefficient on housing supply is contrary to expectations.

This suggests possible endogeneity concerns in the data that we have been working with. Therefore, we have come forth with the Shift Share approach which can be viewed in the next section.

### 6.4 Shift Share Approach

##

	(1)
(Intercept)	13 863.860
	(10 923.869)
Bartik_Z	12.226***
	(0.488)
Num.Obs.	234
R2	0.730
R2 Adj.	0.729
AIC	6257.6
BIC	6268.0
Log.Lik.	-3125.796
F	627.448
RMSE	153 148.31

+ p <0.1, \* p <0.05, \*\* p <0.01, \*\*\* p <0.001

```
## 2SLS Regression Results: Effect of Refugees on Housing Prices
## -----
## Dependent variable:
## -----
## HPI
## -----
## Total_Refugees          0.00000
##                           (0.00001)
##
## Constant                129.097*** 
##                           (2.181)
##
## -----
## Observations             234
## R2                      0.008
## Adjusted R2              0.004
## Residual Std. Error     30.074 (df = 232)
## -----
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Table 4: IV Regression Diagnostics (Weak Instruments, Endogeneity, etc.)

Test	df1	df2	statistic	p-value
Weak instruments	1	232	627.448	0.000
Wu-Hausman	1	231	2.963	0.087
Sargan	0	NA	NA	NA

The first-stage regression confirms that the Bartik instrument is a strong predictor of actual refugee inflows. The coefficient on Bartik\_Z is 12.23 (SE = 0.49), highly significant at the 1% level, indicating a strong positive relationship. The model explains 73% of the variation in refugee inflows ( $R^2 = 0.73$ ), and the

weak instruments test statistic of 627.45 far exceeds the conventional threshold, ruling out weak instrument concerns. These results validate the strength and relevance of the instrument that we will be employing in the ensuing the second-stage analysis.

Here, our group also wants to highlight before presenting the Bartik approach results that, we constructed the instrument by interacting global, origin-specific refugee supply shocks with each country's historical settlement shares, averaged over the 2010–2014 pre-crisis period. This design assumes that historical refugee settlement patterns are exogenous to future housing market shocks and that global refugee flows are driven primarily by geopolitical events outside the control of host countries. We deliberately limit our baseline specification to exclude additional controls such as GDP or inflation to avoid overfitting, multicollinearity, and possible post-treatment bias, focusing instead on the exogenous variation introduced by the instrument.

This table presents the second-stage results of a two-stage least squares (2SLS) regression estimating the causal impact of refugee inflows on housing prices (HPI), using a Bartik-style instrument to address endogeneity concerns. The coefficient on Total\_Refugees is effectively zero and statistically insignificant, suggesting that instrumented refugee inflows do not have a detectable impact on country level HPI throughout the countries and years studied. The constant term is highly significant, reflecting the baseline level of HPI when refugee inflows are *zero*. With an  $R^2$  of 0.008, the model explains less than 1% of the variation in HPI, reinforcing the conclusion that, at the aggregate level, refugee arrivals—when instrumented by exogenous global shocks—do not significantly influence housing market prices (which our group has proxied by way of HPI).

## 7 Conclusion

This report set out to investigate the impact of refugee inflows on housing prices across European countries, leveraging the policy shock of increased refugee acceptance post-2015/16. The study used a panel dataset of country-year observations and employed a shift-share instrumental variable approach to address potential endogeneity concerns. Initial empirical models suggested a nuanced relationship, with current refugee growth appearing to exert upward pressure on housing prices, while lagged refugee growth indicated a potential dampening effect. However, these preliminary findings were subject to concerns about endogeneity, which motivated the application of the shift-share instrumental variable approach.

The implementation of the Bartik-style instrument, utilizing historical refugee settlement patterns to isolate exogenous variation in refugee allocation, yielded a critical shift in our understanding. The strength and validity of the instrument were confirmed through robust first-stage regression results, effectively addressing the endogeneity issue.

Crucially, the second-stage analysis revealed that, once endogeneity is accounted for, refugee inflows do not demonstrate a statistically significant impact on housing prices at the aggregate country level. This finding implies that the observed correlations in initial models were likely confounded by other factors, and that exogenous variations in refugee inflows, driven by broader geopolitical events and historical settlement patterns, do not exert a measurable influence on housing market dynamics within the studied European context.

These results contribute to the ongoing debate on the socio-economic consequences of refugee migration, suggesting that concerns about inflationary pressures on housing markets due to refugee influx may be overstated at the macroeconomic level. It is important to acknowledge the limitations of this study, primarily its focus on aggregate country-level effects. Future research could explore more granular regional or local impacts, or investigate the role of specific housing policies and market structures in mediating the relationship between refugee inflows and housing prices. Further research can also examine the long term effects.

## 8 Appendix

```
df_fin <- read_csv("Final_df_pop.csv")
```

```

df_fin <- df_fin %>% filter(Country != "Turkiye")
df_fin <- df_fin %>% mutate(Total_Refugees2= Total_Refugees+Total_Asylum)
df_fin <- df_fin %>% mutate(Refugees_per_1000= Total_Refugees2*1000/Population)
refugees_over_time <- df_fin %>% group_by(Year) %>% summarise(Total_Refugees_Aggregated = sum(Total_Refugees2, na.rm = TRUE))
plot_refugees_time <- ggplot(refugees_over_time, aes(x = Year, y = Total_Refugees_Aggregated)) +
  geom_line() + geom_point() + labs( title = "Total Refugee Numbers Over Time (Aggregated Across Countries)", x = "Year", y = "Total Number of Refugees" ) + theme_minimal()
print(plot_refugees_time)

```

## 9 Calculate summary statistics per year

```

year_summary_stats <- df_fin %>% group_by(Year) %>% summarise( Count = n(), Mean = mean(Refugees_per_1000, na.rm = TRUE), SD = sd(Refugees_per_1000, na.rm = TRUE), Min = min(Refugees_per_1000, na.rm = TRUE), Q1 = quantile(Refugees_per_1000, 0.25, na.rm = TRUE), Median = median(Refugees_per_1000, na.rm = TRUE), Q3 = quantile(Refugees_per_1000, 0.75, na.rm = TRUE), Max = max(Refugees_per_1000, na.rm = TRUE) )

```

## 10 Print summary statistics per year

```

kable(year_summary_stats, digits = 2, caption = "Summary Statistics for Refugees per 1000 Residents by Year")

```

## 11 Data prep

```

latest_year <- max(df_fin$Year) refugees_latest_year <- df_fin %>% filter(Year == latest_year) %>% arrange(desc(Refugees_per_1000))

refugees_2014 <- df_fin %>% filter(Year == "2014") %>% arrange(desc(Refugees_per_1000))

refugees_2016 <- df_fin %>% filter(Year == "2016") %>% arrange(desc(Refugees_per_1000))

top_n <- 10

refugees_latest_year_top <- refugees_latest_year %>% slice_max(order_by = Refugees_per_1000, n = top_n)

refugees_2016_top <- refugees_2016 %>% slice_max(order_by = Refugees_per_1000, n = top_n)

refugees_2014_top <- refugees_2014 %>% slice_max(order_by = Refugees_per_1000, n = top_n)

max_refugees <- max(refugees_latest_year$Refugees_per_1000)

plot_refugees_latest_year <- ggplot(refugees_latest_year_top, aes(x = reorder(Country, Refugees_per_1000), y = Refugees_per_1000)) + geom_col() + coord_flip() + labs( title = paste("Refugee Numbers by Country in", latest_year), x = "Country", y = "Total Number of Refugees" ) + scale_y_continuous(limits = c(0, max_refugees), labels = scales::comma) + theme_minimal() + theme(axis.text.y = element_text(size = 6))

plot_refugees_2016 <- ggplot(refugees_2016_top, aes(x = reorder(Country, Refugees_per_1000), y = Refugees_per_1000)) + geom_col() + coord_flip() + labs( title = "Refugee Numbers by Country in 2016",
```

```
x = "Country", y = "Total Number of Refugees" ) + scale_y_continuous(limits = c(0, max_refugees),  
labels = scales::comma) + theme_minimal() + theme(axis.text.y = element_text(size = 6))
```

```
plot_refugees_2014 <- ggplot(refugees_2014_top, aes(x = reorder(Country, Refugees_per_1000), y =  
Refugees_per_1000)) + geom_col() + coord_flip() + labs( title = "Refugee Numbers by Country in 2014",  
x = "Country", y = "Total Number of Refugees" ) + scale_y_continuous(limits = c(0, max_refugees),  
labels = scales::comma) + theme_minimal() + theme(axis.text.y = element_text(size = 6))
```

## 12 Combine plots

```
combined_plot <- plot_refugees_latest_year / plot_refugees_2016 / plot_refugees_2014 print(combined_plot)  
data_filtered <- df_fin %>% filter(Year >= 2014 & Year <= 2023)
```

## 13 ===== 3. Facet Line Plot =====

## 14 Get top 20 countries by refugee intake (2014–2023)

```
top_20_countries <- data_filtered %>% filter(Year == 2023) %>% group_by(Country) %>% summarise(total_refugees = sum(Total_Refugees, na.rm = TRUE)) %>% arrange(desc(total_refugees)) %>% slice(1:20) %>% pull(Country)
```

## 15 Filter for top countries

```
top_data <- data_filtered %>% filter(Country %in% top_20_countries)
```

## 16 Faceted Line Plot with 2015 vertical reference line

```
facet_plot <- ggplot(top_data, aes(x = Year, y = Refugees_per_1000, color = Country, group = Country))  
+ geom_line(size = 1.2) + facet_wrap(~Country, scales = "free_y") + scale_color_viridis_d(option =  
"plasma") + scale_y_continuous(limits = c(0, 45)) + theme_minimal(base_size = 13) + labs( title =  
"Refugee per 1000 residents trends in Top Countries (2014–2023)", x = "Year", y = "Refugees per 1000  
residents" ) + theme( plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angle = 90, vjust = 0.5) )  
print(facet_plot)  
df_avg <- df_fin %>% group_by(Country) %>% summarise( avg_refugees_per_1000 = mean(Refugees_per_1000,  
na.rm = TRUE) )  
medi <- median(df_avg$avg_refugees_per_1000, na.rm = TRUE)  
Q3 <- quantile(df_avg$avg_refugees_per_1000, 0.75, na.rm = TRUE)
```

## 17 Step 3: Classify countries based on Q1 and Q3

```
df_avg <- df_avg %>% mutate(refugee_influx_group = case_when( avg_refugees_per_1000 > Q3 ~  
"High", avg_refugees_per_1000 < medi ~ "Low", TRUE ~ "Medium" ))
```

## 18 Step 4: Sort and display as a table

```
df_avg_sorted <- df_avg %>% arrange(desc(avg_refugees_per_1000))
color_vector <- case_when( df_avg_sorted$refugee_influx_group == "High" ~ "red",
                           df_avg_sorted$refugee_influx_group == "Medium" ~ "orange",
                           df_avg_sorted$refugee_influx_group == "Low" ~ "gray" )
```

## 19 Step 4: Display table with colored influx group column

```
df_avg_sorted %>% kable( digits = 2, caption = "Average Refugees per 1000 Residents (2010–2023) and Influx Classification" ) %>% kable_styling(full_width = FALSE) %>% column_spec(3, color = "white", background = color_vector) %>% row_spec(0, bold = TRUE)
df_fin2 <- read_excel("Final_df_pop.xlsx", sheet = "Final_df_pop")
df_fin2 <- df_fin2 %>% filter(Country != "Turkiye")
df_fin2 <- df_fin2 %>% mutate(Total_Refugees2= Total_Refugees+Total_Asylum)
df_fin2 <- df_fin2 %>% mutate(Refugees_per_1000_2= Total_Refugees2*100/Population)
ggplot(df_fin2, aes(x = Year, y = Housing_per_1000*1000000)) + geom_area(aes(fill = Country), alpha = 0.4, position = 'identity') + geom_line(aes(color = Country), size = 1.1) + geom_point(aes(color = Country), size = 2.2, alpha = 0.9) + facet_wrap(~ Country, scales = "fixed") + scale_color_viridis_d(option = "mako", begin = 0.2, end = 0.9) + scale_fill_viridis_d(option = "rocket", begin = 0.3, end = 0.95) + labs( title = "Housing Supply per 1,000 People Across Countries (2014 - 2023)", subtitle = "Standardized by population for a more meaningful comparison", x = "Year", y = "Housing Supply per 1,000 People" ) + theme_minimal(base_size = 13) + theme( strip.background = element_rect(fill = "#F9F871", color = NA), strip.text = element_text(face = "bold", size = 11), plot.background = element_rect(fill = "#f6f8fa", color = NA), panel.grid.major = element_line(color = "#D0E0F0", size = 0.3), plot.title = element_text(face = "bold", size = 18, hjust = 0.5, color = "#333399"), plot.subtitle = element_text(size = 13, hjust = 0.5, color = "#555555"), legend.position = "none" )
data_filtered <- df_fin %>% filter(Year >= 2014 & Year <= 2023)
```

## 20 Top 10 refugee countries based on total refugee count

```
top_refugee_countries <- data_filtered %>% group_by(Country) %>% summarise(total_refugees = sum(Total_Refugees, na.rm = TRUE)) %>% arrange(desc(total_refugees)) %>% slice(1:10) %>% pull(Country)
```

## 21 Filter the dataset

```
top_data <- data_filtered %>% filter(Country %in% top_refugee_countries, HPI <= 500)
```

## 22 Ridge plot with reduced smoothing (bandwidth control)

```
ggplot(top_data, aes(x = HPI, y = factor(Year), fill = after_stat(x))) + geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01, bandwidth = 6 # Gentlemen, we can reduce this to 3 or 2 if you want it bumpier ) + scale_fill_viridis_c(option = "magma", name = "HPI") + labs( title = "Distribution of HPI Over Time", x = "EU Housing Price Index (HPI)", y = "Year" ) + coord_cartesian(xlim = c(50,
```

```
350)) + theme_minimal(base_size = 14) + theme( plot.title = element_text(hjust = 0.5), legend.position = "bottom" )
```

## 23 Filter data for 2010–2021 and merge influx group information

```
df_merged_2010_2021 <- df_fin %>% left_join(df_avg_sorted %>% select(Country, refugee_influx_group), by = "Country")
```

## 24 Calculate the average HPI and refugees per 1000 for each group over the period 2010–2021

```
df_avg_grouped <- df_merged_2010_2021 %>% group_by(refugee_influx_group, Year) %>% summarise( avg_HPI = mean(HPI, na.rm = TRUE), avg_refugees_per_1000 = mean(Refugees_per_1000, na.rm = TRUE) )
```

## 25 Calculate the overall average for each group across the years

```
df_avg_grouped_overall <- df_avg_grouped %>% group_by(refugee_influx_group) %>% summarise( avg_HPI = mean(avg_HPI, na.rm = TRUE), avg_refugees_per_1000 = mean(avg_refugees_per_1000, na.rm = TRUE) )
```

## 26 Graph 1: Average HPI Over Time (2010–2023) by Refugee Influx Group

```
ggplot(df_avg_grouped, aes(x = as.factor(Year), y = avg_HPI, color = refugee_influx_group)) + geom_line(aes(group = refugee_influx_group), size = 1) + geom_point(size = 3) + scale_color_manual(values = c( "Low" = "gray", "Medium" = "orange", "High" = "red" )) + labs( title = "HPI Over Time (2010–2023) by Refugee Influx Groups", x = "Year", y = "House Price Index (HPI)", color = "Refugee Influx per 1000 residents" ) + theme_minimal() + theme(axis.text.x = element_text(angle = 45, hjust = 1))  
data_filtered <- df_fin %>% filter(Year >= 2014 & Year <= 2023)
```

## 27 ===== 3. Facet Line Plot =====

## 28 Get top 20 countries by refugee intake (2014–2023)

```
top_20_countries <- data_filtered %>% group_by(Country) %>% summarise(total_refugees = sum(Total_Refugees, na.rm = TRUE)) %>% arrange(desc(total_refugees)) %>% slice(1:20) %>% pull(Country)
```

## 29 Filter for top countries

```
top_data <- data_filtered %>% filter(Country %in% top_20_countries)
```

```

ggplot(top_data, aes(x = Year, y = HPI, color = Country)) + geom_line(size = 1.2) + facet_wrap(~Country,
scales = "fixed") + scale_color_viridis_d(option = "plasma") + theme_minimal(base_size = 13) + labs(
title = "HPI Trends in Top 20 Refugee-Intake Countries (2014–2023)", x = "Year", y = "EU Housing
Pricing Index (HPI)" ) + theme( plot.title = element_text(hjust = 0.5, size = 14), # reduced title size
axis.text.x = element_text(angle = 90, vjust = 0.5) # vertical x-axis labels )

europe <- ne_countries(scale = "medium", returnclass = "sf", continent = "Europe")

```

## 30 Assuming 'df\_avg' is your dataframe

```

df_avgCountry <- as.character(df_avg$Country) europecountryname <- as.character(europe$name)
europe_filtered <- europe %>% filter(country_name %in% df_avg$Country)
europe_merged <- left_join(europe_filtered, df_avg, by = c("country_name" = "Country"))
europe_merged$refugee_influx_group <- factor(europe_merged$refugee_influx_group, levels = c("Low",
"Medium", "High"))

europe_centroids <- st_centroid(europe_merged)

ggplot(data = europe_merged) + geom_sf(aes(fill = refugee_influx_group), color = "#d9d9d9", linewidth =
0.15) + geom_sf_text( data = europe_centroids, aes(label = country_name), size = 2, color = "#333333" )
+ scale_fill_manual( values = c("Low" = "#deeaf7", "Medium" = "#9ecae1", "High" = "#3182bd"),
name = "Refugee Influx", labels = c("Low", "Medium", "High") ) + coord_sf(xlim = c(-25, 45), ylim =
c(34, 72), expand = FALSE) + theme_minimal(base_family = "sans") + labs( title = "Refugee Influx
Groups in Europe (2010–2023)", subtitle = "Average refugee influx per 1,000 residents", x = NULL, # Remove x-axis label
y = NULL # Remove y-axis label ) + theme( plot.title = element_text(size = 16, face =
"bold", hjust = 0.5, margin = margin(b = 5)), plot.subtitle = element_text(size = 11, hjust = 0.5, color =
"#666666", margin = margin(b = 10)), legend.position = "bottom", legend.title = element_text(face =
"bold", size = 10), legend.text = element_text(size = 9), axis.text.x = element_blank(), # Remove x-axis
text axis.ticks.x = element_blank(), # Remove x-axis ticks axis.text.y = element_blank(), # Remove y-axis
text axis.ticks.y = element_blank(), # Remove y-axis ticks panel.grid = element_blank(), plot.background =
element_rect(fill = "white", color = NA), panel.background = element_rect(fill = "white", color = NA)
)

df_fin <- df_fin %>% arrange(Country, Year) %>% group_by(Country) %>% mutate( Refugee_Growth =
(Refugees_per_1000 - lag(Refugees_per_1000)) / lag(Refugees_per_1000) * 100, HPI_Growth = (HPI -
lag(HPI)) / lag(HPI) * 100 ) %>% ungroup()

correlation_hpi_refugees <- cor(df_fin$HPI_Growth, df_fin$Refugee_Growth, use = "pairwise.complete.obs")
scatter_plot_growth <- ggplot(df_fin, aes(x = Refugee_Growth, y = HPI_Growth)) + geom_point(alpha =
0.6, color = "#1f78b4") + labs( title = paste("Scatter Plot of HPI Growth vs. Refugees_per_1000
residents Change (Correlation = ", round(correlation_hpi_refugees, 3), ")"), x = "Refugees Change (%)", y =
"HPI Growth (%)") + scale_x_continuous(limits = c(-20, 20), labels = scales::percent_format(scale = 1))
+ scale_y_continuous(limits = c(-20, 20), labels = scales::percent_format(scale = 1)) + theme_minimal()

print(scatter_plot_growth)

df_fin <- df_fin %>% arrange(Country, Year) %>% group_by(Country) %>% mutate( Refugee_Growth_Lag1 =
lag(Refugee_Growth, 1), Refugee_Growth_Lag2 = lag(Refugee_Growth, 2) ) %>% ungroup()

```

## 31 Model with only fixed effects (Country and Year)

```

fixed_effects_only_model <- lm(HPI_Growth ~ Country + Year, data = df_fin) summary_fe_only <-
summary(fixed_effects_only_model)

```

```

full_model <- lm(HPI_Growth ~ Refugee_Growth + Refugee_Growth_Lag1 + Country + Year, data = df_fin)

modelsummary( list("Fixed Effects Only" = fixed_effects_only_model, "Full Model" = full_model),
coef omit = "Country|Year", stars = TRUE )

#Checking for multicollinearity concerns

vif_table <- vif(full_model) vif_df <- as.data.frame(vif_table)

kable(vif_df, format = "latex", booktabs = TRUE, digits = 2, caption = "Variance Inflation Factor (VIF) Table") %>% kable_styling(latex_options = c("striped", "hold_position"))

df_fin <- df_fin %>% arrange(Country, Year) %>% group_by(Country) %>% mutate(Refugees_per_1000_Growth = (Refugees_per_1000 - lag(Refugees_per_1000)) / lag(Refugees_per_1000), Refugees_per_1000_Growth_Lag1 = lag(Refugees_per_1000_Growth, 1), Refugees_per_1000_Growth_Lag2 = lag(Refugees_per_1000_Growth, 2)) %>% ungroup()

```

## 32 Step 2: Fixed effects only model (Country and Year)

```
fixed_effects_only_model <- lm(HPI_Growth ~ Country + Year, data = df_fin)
```

## 33 Step 3: Full model with Refugees\_per\_1000\_Growth and its lag

```
full_model <- lm( HPI_Growth ~ Refugees_per_1000_Growth + Refugees_per_1000_Growth_Lag1 + Country + Year, data = df_fin )
```

## 34 Step 3: Full model with Refugees\_per\_1000\_Growth and its lag, controling for housing supply growth

```
df_fin2 <- df_fin %>% arrange(Country, Year) %>% group_by(Country) %>% mutate(Housing_per_1000_Growth = Housing_per_1000 / lag(Housing_per_1000) - 1, Refugees_per_1000_Growth = (Refugees_per_1000 - lag(Refugees_per_1000)) / lag(Refugees_per_1000), Refugees_per_1000_Growth_Lag1 = lag(Refugees_per_1000_Growth, 1), Refugees_per_1000_Growth_Lag2 = lag(Refugees_per_1000_Growth, 2), Housing_per_1000_Growth_lag1=lag(Housing_per_1000_Growth,1) ) %>% ungroup()
```

## 35 Step 2: Run the full model including the new control variable

```
full_model2 <- lm( HPI_Growth ~ Refugees_per_1000_Growth + Refugees_per_1000_Growth_Lag1 + Housing_per_1000_Growth + Housing_per_1000_Growth_lag1+ Country + Year, data = df_fin2 )

modelsummary( list( "Fixed Effects Only" = fixed_effects_only_model, "Full Model (df_fin)" = full_model, "Full Model + Housing" = full_model2 ), coef omit = "Country|Year", # Omits fixed effect dummies from display stars = TRUE )
```

## 36 Step 1: Loading both sheets from Excel file, its an xlsx file now

```
main_df<- read_excel("Final_df_pop.xlsx", sheet = "Final_df_pop") refugee_df <- read_excel("Final_df_pop.xlsx", sheet = "persons_of_concern")
```

## **37 Step 2: Filtering base period (2010–2014) to compute average shares**

```
base_df <- refugee_df %>% filter(Year >= 2010 & Year <= 2014)
```

## **38 Step 3: Averaging and calculating the refugees by host and origin country**

```
avg_shares <- base_df %>% group_by(Country of Asylum, Country of Origin) %>% summarise(Avg_Refugees = mean(Refugees, na.rm = TRUE), .groups = "drop")
```

## **39 Step 4: Total refugees from each origin across all countries**

```
origin_totals <- avg_shares %>% group_by(Country of Origin) %>% summarise(Total-Origin = sum(Avg_Refugees, na.rm = TRUE), .groups = "drop")
```

## **40 Step 5: Compute 5-year base shares**

```
base_shares <- avg_shares %>% left_join(origin_totals, by = "Country of Origin") %>% mutate(Base_Share = Avg_Refugees / Total-Origin)
```

## **41 Step 6: Get national shocks by year and origin**

```
national_shocks <- refugee_df %>% group_by(Year, Country of Origin) %>% summarise(National_Shock = sum(Refugees, na.rm = TRUE), .groups = "drop")
```

## **42 Step 7: Merging base shares with shocks and calculate instrument components**

```
bartik_components <- base_shares %>% left_join(national_shocks, by = "Country of Origin") %>% mutate(Weighted_Shock = Base_Share * National_Shock)
```

## **43 Step 8: Aggregating to country-year level Bartik instrument**

```
bartik_df <- bartik_components %>% group_by(Year, Country of Asylum) %>% summarise(Bartik_Z = sum(Weighted_Shock, na.rm = TRUE), .groups = "drop")
```

```
#Step 9: merging final_merged <- main_df %>% left_join(bartik_df, by = c("Country" = "Country of Asylum", "Year" = "Year"))
```

```
#Step 10: # First-stage regression first_stage <- lm(Total_Refugees ~ Bartik_Z, data = final_merged)
```

## 44 View first-stage summary

```
modelsummary(first_stage, stars = TRUE)
```

## 45 Step 10:

## 46 Basic IV without controls

```
model_iv_basic <- ivreg( HPI ~ Total_Refugees | Bartik_Z, data = final_merged )
```

## 47 1. 2SLS Regression Output

```
suppressWarnings( stargazer(model_iv_basic, type = "text", title = "2SLS Regression Results: Effect of Refugees on Housing Prices") )
```

## 48 2. Diagnostics Table

```
diagnostics <- summary(model_iv_basic, diagnostics = TRUE)  
diagnostics$diag_df <- as.data.frame(diagnostics)$diag_df  
<- rownames(diagnostics) rownames(diag_df) <- NULL
```

## 49 Clean display with kable

```
kable(diag_df[, c("Test", "df1", "df2", "statistic", "p-value")], digits = 3, caption = "IV Regression Diagnostics (Weak Instruments, Endogeneity, etc.)")
```