

Global House Market Analysis

Julius

2025-03-30

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
library(grid)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4          v readr      2.1.4
## v forcats    1.0.0          v stringr   1.5.1
## v ggplot2    3.5.0.9000     v tibble    3.2.1
## v lubridate  1.9.4          v tidyr     1.3.1
## v purrr      1.0.2
## -- 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
```

```
library(sf)
```

```
## Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE
```

```
library(rnaturalearth)
```

```
## The legacy packages mapproj, rgdal, and rgeos, underpinning the sp package,
## which was just loaded, were retired in October 2023.
## Please refer to R-spatial evolution reports for details, especially
## https://r-spatial.org/r/2023/05/15/evolution4.html.
## It may be desirable to make the sf package available;
## package maintainers should consider adding sf to Suggests:.
## Support for Spatial objects ('sp') will be deprecated in {rnaturalearth} and will be removed in a future
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##   discard
##
## The following object is masked from 'package:readr':
##
##   col_factor
```

```
library(naniar)
library(ggplot2)
library(tinytex)
library(RColorBrewer)
library(rvest)
```

```
##
## Attaching package: 'rvest'
##
## The following object is masked from 'package:readr':
##
##   guess_encoding
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(leaflet)
```

```
#tinytex::install_tinytex()
```

```
house_market <- read_csv("/Users/julius/Personal/Personal_Project/House_market_analysis/Global-Housing-I
```

```
## Rows: 200 Columns: 11
## -- Column specification -----
## Delimiter: ","
## chr  (1): Country
## dbl (10): Year, House Price Index, Rent Index, Affordability Ratio, Mortgage...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
Countries_coordinate <- read_csv("/Users/julius/Personal/Personal_Project/House_market_analysis/Global-I
```

```
## Rows: 245 Columns: 8
## -- Column specification -----
```

```
## Delimiter: ","
## chr (4): country_code, country, usa_state_code, usa_state
## dbl (4): latitude, longitude, usa_state_latitude, usa_state_longitude
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
Countries_coordinate <- select(Countries_coordinate,1:4)
```

```
#Run this once only
```

```
#Filtering the list of countries that are inside the house market dataset and. joining the two dataset
country_filter <- c("United States","Canada","United Kingdom","Germany","France","Italy","Spain","Austria")
```

```
Countries_coordinate <- Countries_coordinate %>%
  filter(country %in% country_filter) %>%
  mutate(country = ifelse(country == "United Arab Emirates", "UAE", country)) %>%
  mutate(country = ifelse(country == "United States", "USA", country)) %>%
  mutate(country = ifelse(country == "United Kingdom", "UK", country))
```

```
colnames(Countries_coordinate)[4] <- "Country"
```

```
#joining the two files
```

```
House_market_coordinate <- (left_join(house_market, Countries_coordinate, by='Country'))
```

```
#initial analysis
```

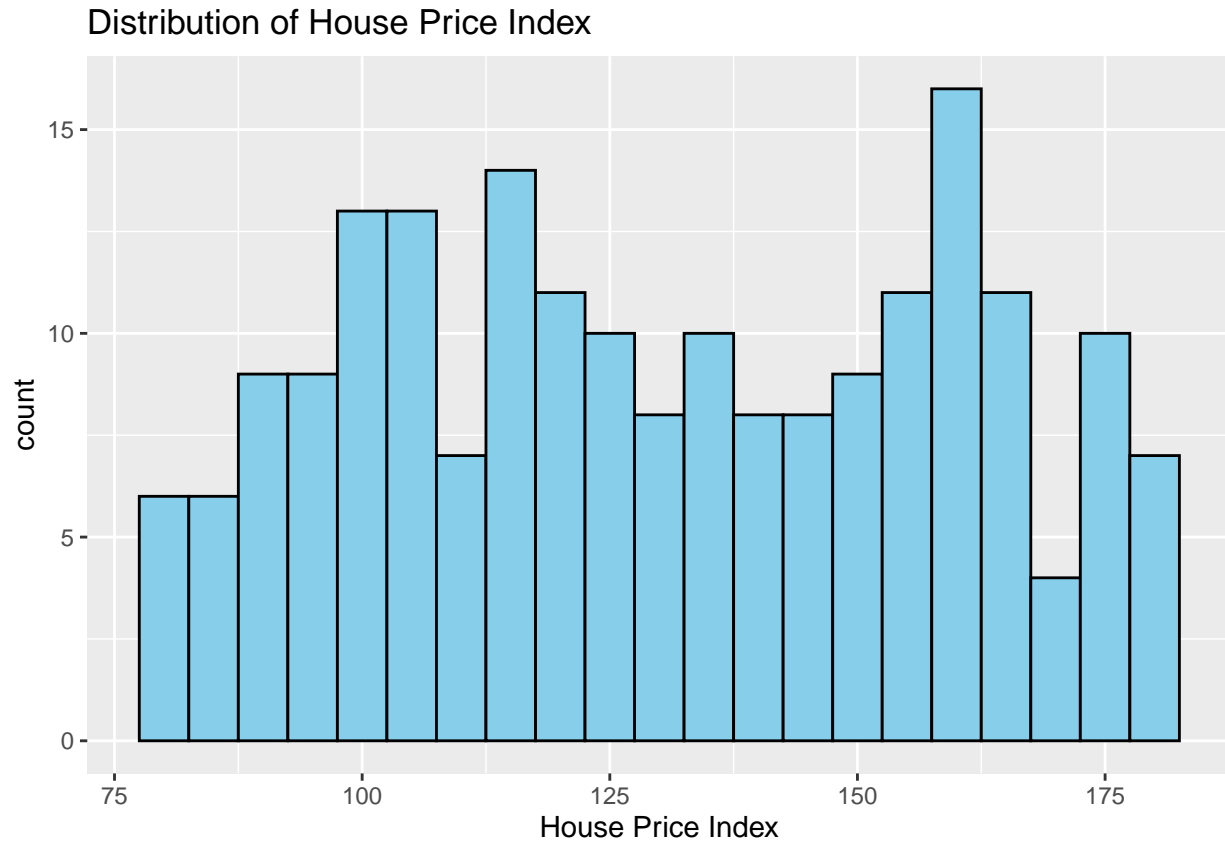
```
summary(House_market_coordinate)
```

```
##      Country          Year      House Price Index      Rent Index
## Length:200      Min.   :2015      Min.   : 80.55      Min.   : 50.35
## Class :character 1st Qu.:2017      1st Qu.:104.14      1st Qu.: 60.47
## Mode  :character Median :2020      Median :129.19      Median : 83.72
##              Mean   :2020      Mean   :130.38      Mean   : 83.05
##              3rd Qu.:2022      3rd Qu.:157.13      3rd Qu.:100.60
##              Max.   :2024      Max.   :179.97      Max.   :119.86
## Affordability Ratio Mortgage Rate (%) Inflation Rate (%) GDP Growth (%)
## Min.   : 3.042      Min.   :1.538      Min.   :0.5321      Min.   : -1.92183
## 1st Qu.: 5.034      1st Qu.:3.045      1st Qu.:1.9392      1st Qu.: -0.09563
## Median : 7.376      Median :4.330      Median :3.6646      Median : 2.30755
## Mean   : 7.238      Mean   :4.151      Mean   :3.6498      Mean   : 2.13372
## 3rd Qu.: 9.276      3rd Qu.:5.218      3rd Qu.:5.2617      3rd Qu.: 4.27278
## Max.   :11.880      Max.   :6.486      Max.   :6.9123      Max.   : 5.95893
## Population Growth (%) Urbanization Rate (%) Construction Index
## Min.   : -0.9614      Min.   :60.17      Min.   : 70.97
## 1st Qu.: -0.1833      1st Qu.:66.92      1st Qu.: 90.18
## Median : 0.7224      Median :75.10      Median :110.59
## Mean   : 0.7228      Mean   :74.77      Mean   :111.20
## 3rd Qu.: 1.6213      3rd Qu.:82.68      3rd Qu.:133.78
## Max.   : 2.4979      Max.   :89.79      Max.   :149.74
## country_code      latitude      longitude
## Length:200      Min.   : -30.56      Min.   : -106.347
## Class :character 1st Qu.: 23.58      1st Qu.: -3.514
## Mode  :character Median : 38.78      Median : 11.509
##              Mean   : 32.72      Mean   : 22.936
```

```
##          3rd Qu.: 51.41    3rd Qu.: 85.271
##          Max.      : 61.52    Max.      : 138.253
```

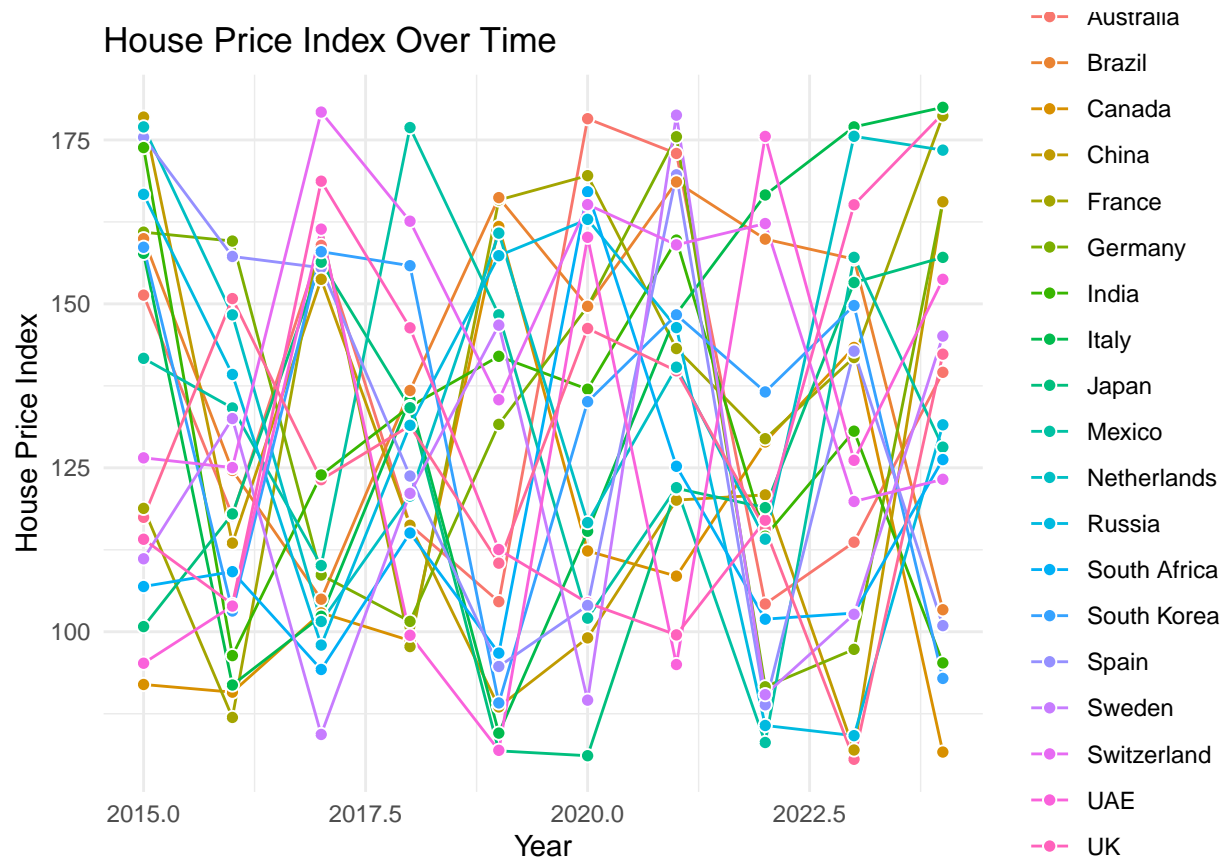
#Distribution of key metrics

```
ggplot(House_market_coordinate, aes(x = `House Price Index`)) +
  geom_histogram(binwidth = 5, fill = "skyblue", color = "black") +
  labs(title = "Distribution of House Price Index")
```



#Time Trends

```
ggplot(House_market_coordinate, aes(x = Year, y = `House Price Index`, fill = Country)) +
  geom_line(aes(color = Country)) +
  geom_point(aes(color = Country), shape = 21, size = 2, color = "white") +
  labs(title = "House Price Index Over Time") +
  theme_minimal()
```



```
# Create the correlation matrix
correlation_matrix <- cor(House_market_coordinate[, c("House Price Index", "Rent Index", "Affordability
Mortgage Rate (%)", "Inflation Rate (%)", "GDP Gr
Population Growth (%)", "Urbanization Rate (%)",
Construction Index"])]

# Enhanced correlation plot
corrplot(correlation_matrix,
  method = "circle",
  type = "upper",
  col = colorRampPalette(c("blue", "white", "red"))(200),
  tl.col = "black",
  tl.srt = 90,
  addCoef.col = "black",
  number.cex = 0.7,
  diag = FALSE,
  cl.lim = c(-1, 1),
  cl.pos = "r",
  title = "Correlation Heatmap of House Market Variables",
  mar = c(0, 0, 1, 0)
)
```

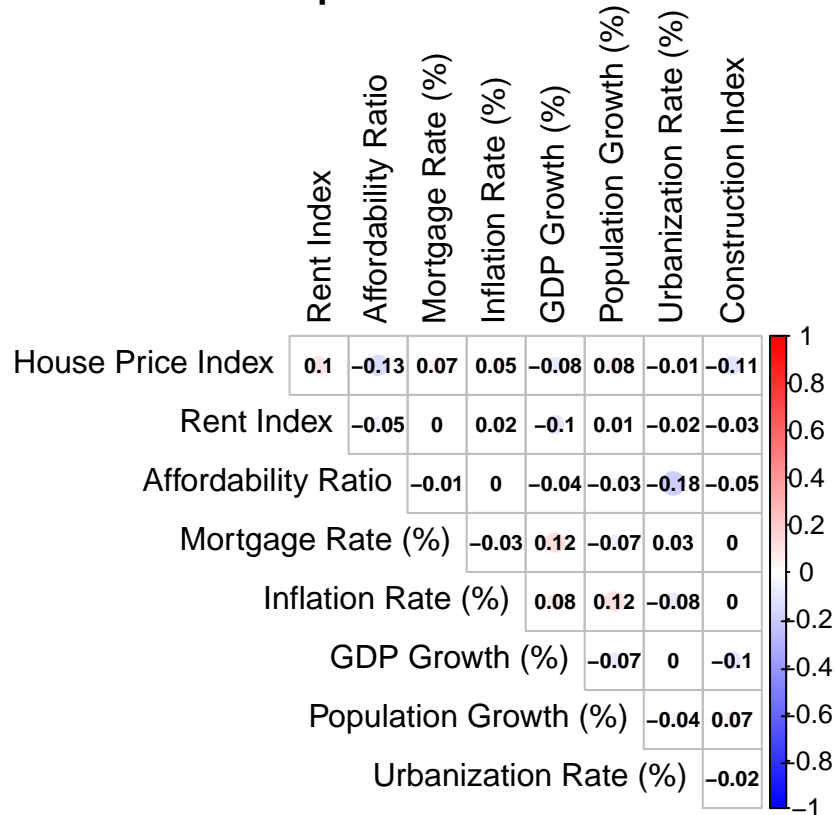
```
## Warning in text.default(pos.xlabel[, 1], pos.xlabel[, 2], newcolnames, srt =
## tl.srt, : "cl.lim" is not a graphical parameter
```

```
## Warning in text.default(pos.ylabel[, 1], pos.ylabel[, 2], newrownames, col =
```

```
## tl.col, : "cl.lim" is not a graphical parameter
```

```
## Warning in title(title, ...): "cl.lim" is not a graphical parameter
```

Correlation Heatmap of House Market Variables



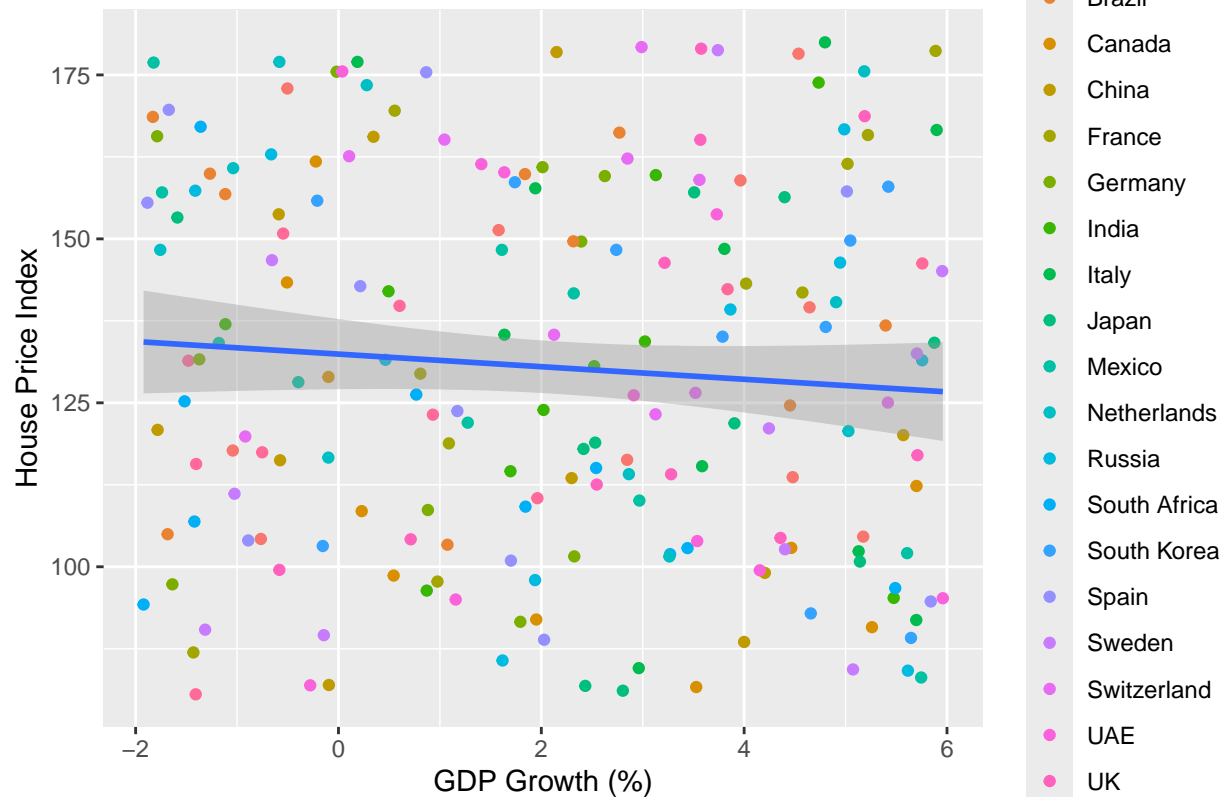
```
summary(House_market_coordinate[, c("House Price Index", "Rent Index", "Affordability Ratio", "Mortgage Rate (%)")])
```

```
## House Price Index   Rent Index   Affordability Ratio Mortgage Rate (%)
## Min.      : 80.55    Min.      : 50.35    Min.      : 3.042    Min.      :1.538
## 1st Qu.:104.14    1st Qu.: 60.47    1st Qu.: 5.034    1st Qu.:3.045
## Median :129.19    Median : 83.72    Median : 7.376    Median :4.330
## Mean   :130.38    Mean   : 83.05    Mean   : 7.238    Mean   :4.151
## 3rd Qu.:157.13    3rd Qu.:100.60    3rd Qu.: 9.276    3rd Qu.:5.218
## Max.    :179.97    Max.    :119.86    Max.    :11.880    Max.    :6.486
```

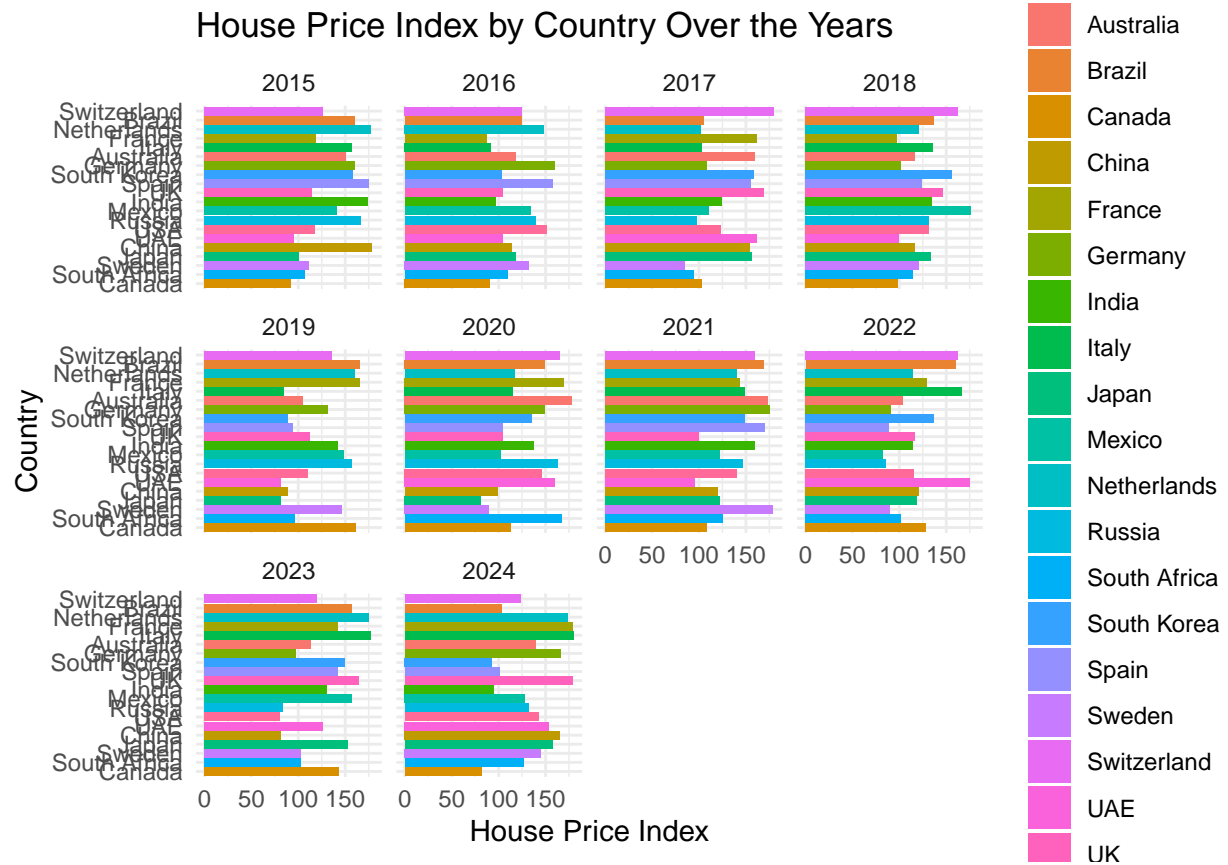
```
ggplot(House_market_coordinate, aes(x = `GDP Growth (%)`, y = `House Price Index`)) +
  geom_point(aes(color = Country)) +
  geom_smooth(method = "lm") +
  labs(title = "Correlation Between GDP Growth and House Price Index")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

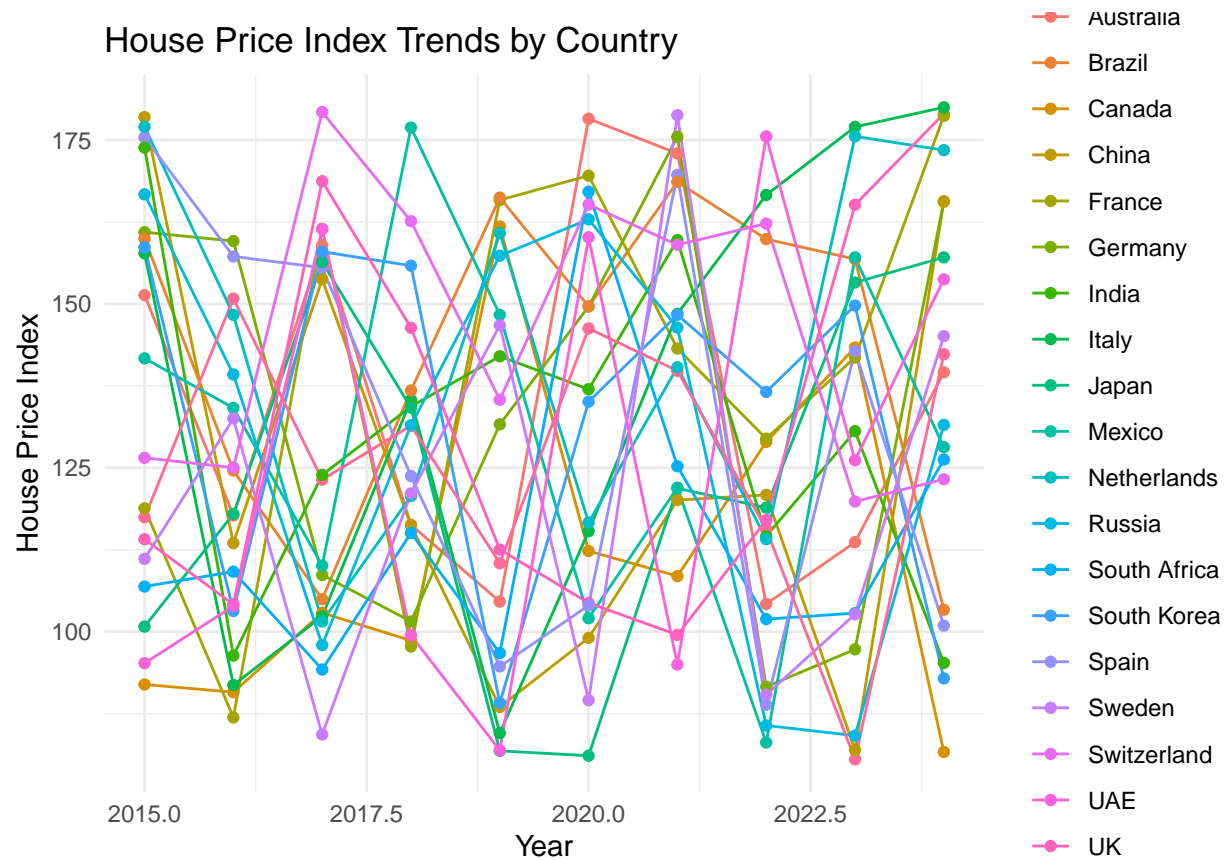
Correlation Between GDP Growth and House Price Index



```
#comparing other countries
ggplot(House_market_coordinate, aes(x = reorder(Country, `House Price Index`), y = `House Price Index`,
  geom_bar(stat = "identity", position = "dodge") +
  coord_flip() +
  facet_wrap(~ Year) + # Creates separate plots for each year
  labs(title = "House Price Index by Country Over the Years",
    x = "Country",
    y = "House Price Index") +
  theme_minimal()
```

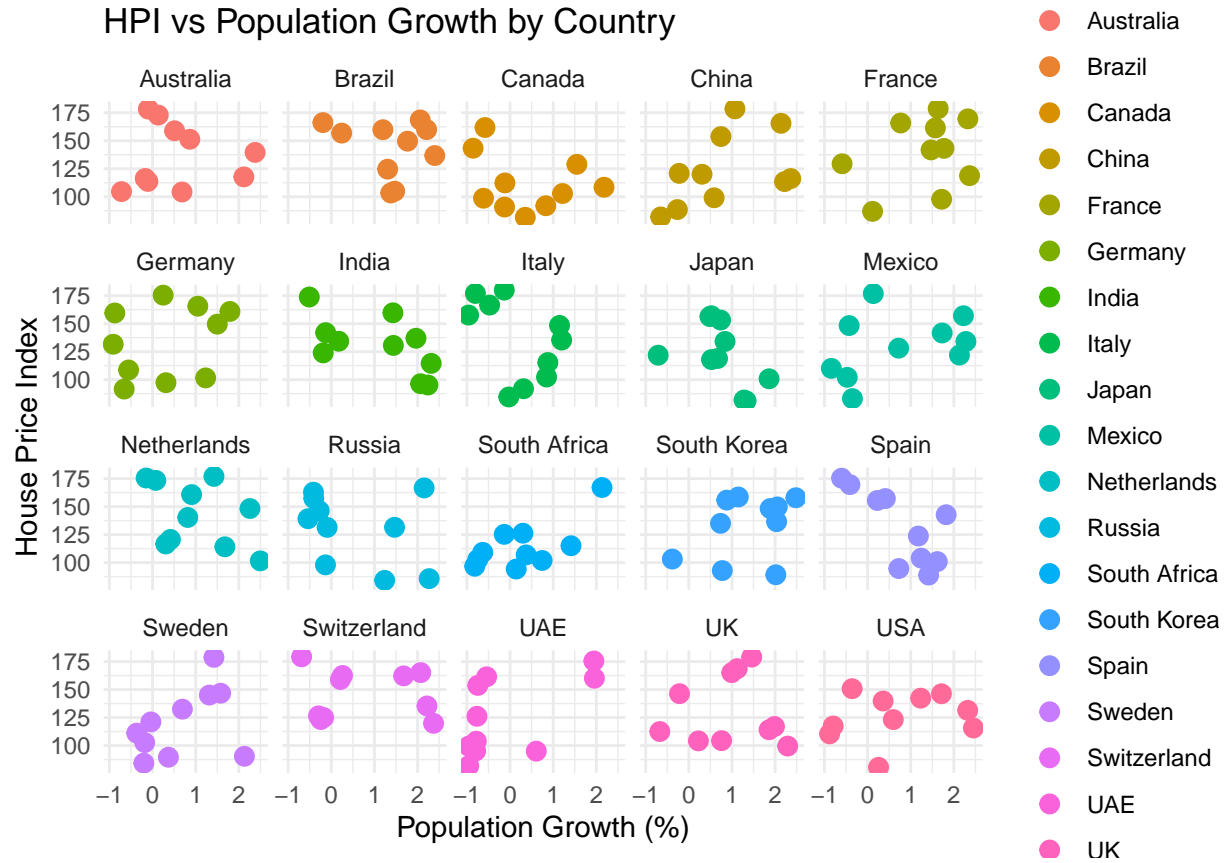


```
ggplot(House_market_coordinate, aes(x = Year, y = `House Price Index`, color = Country)) +
  geom_line() +
  geom_point() + # Adds dots at each year
  labs(title = "House Price Index Trends by Country",
        x = "Year",
        y = "House Price Index") +
  theme_minimal()
```

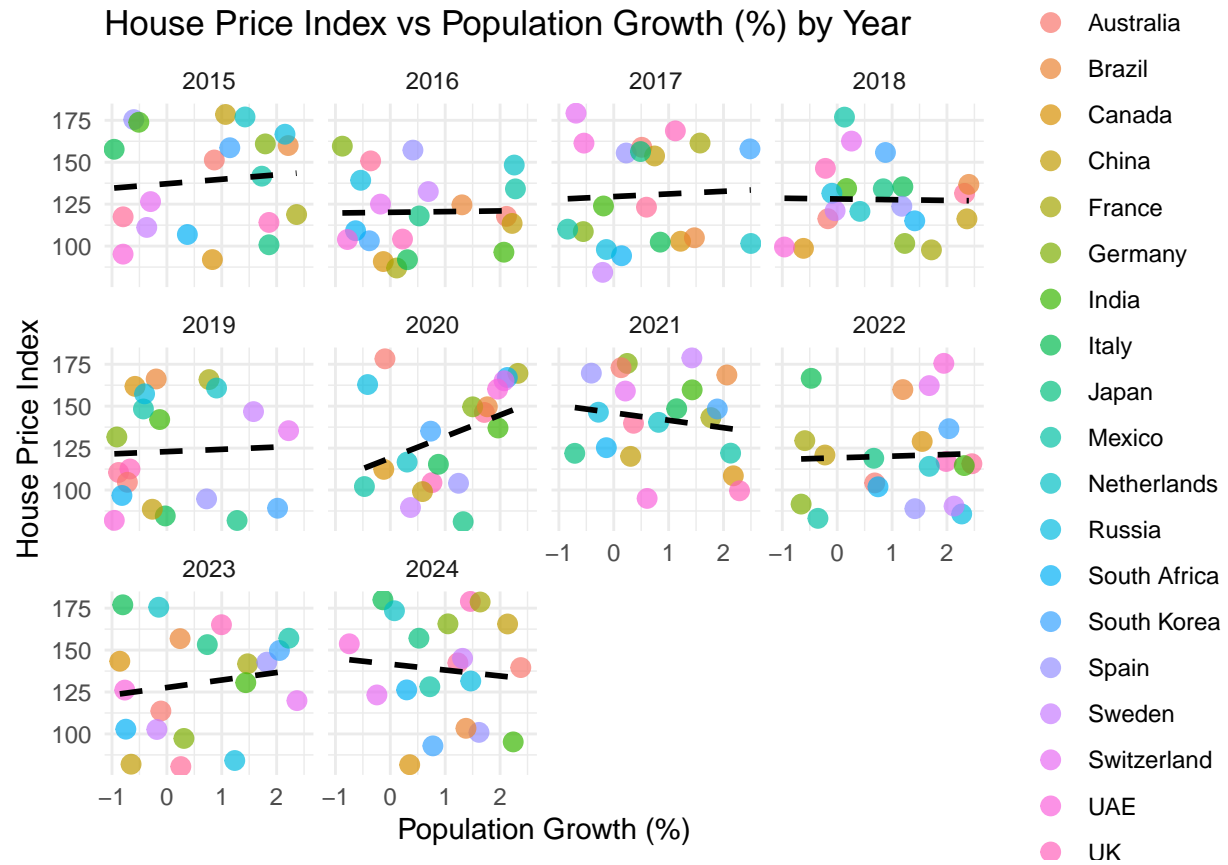
```
ggplot(House_market_coordinate, aes(x = `Population Growth (%)`, y = `House Price Index`, color = Country)) +
  geom_point(size = 3) +
  facet_wrap(~ Country) +
  labs(title = "HPI vs Population Growth by Country",
       x = "Population Growth (%)",
       y = "House Price Index") +
  theme_minimal()
```

HPI vs Population Growth by Country

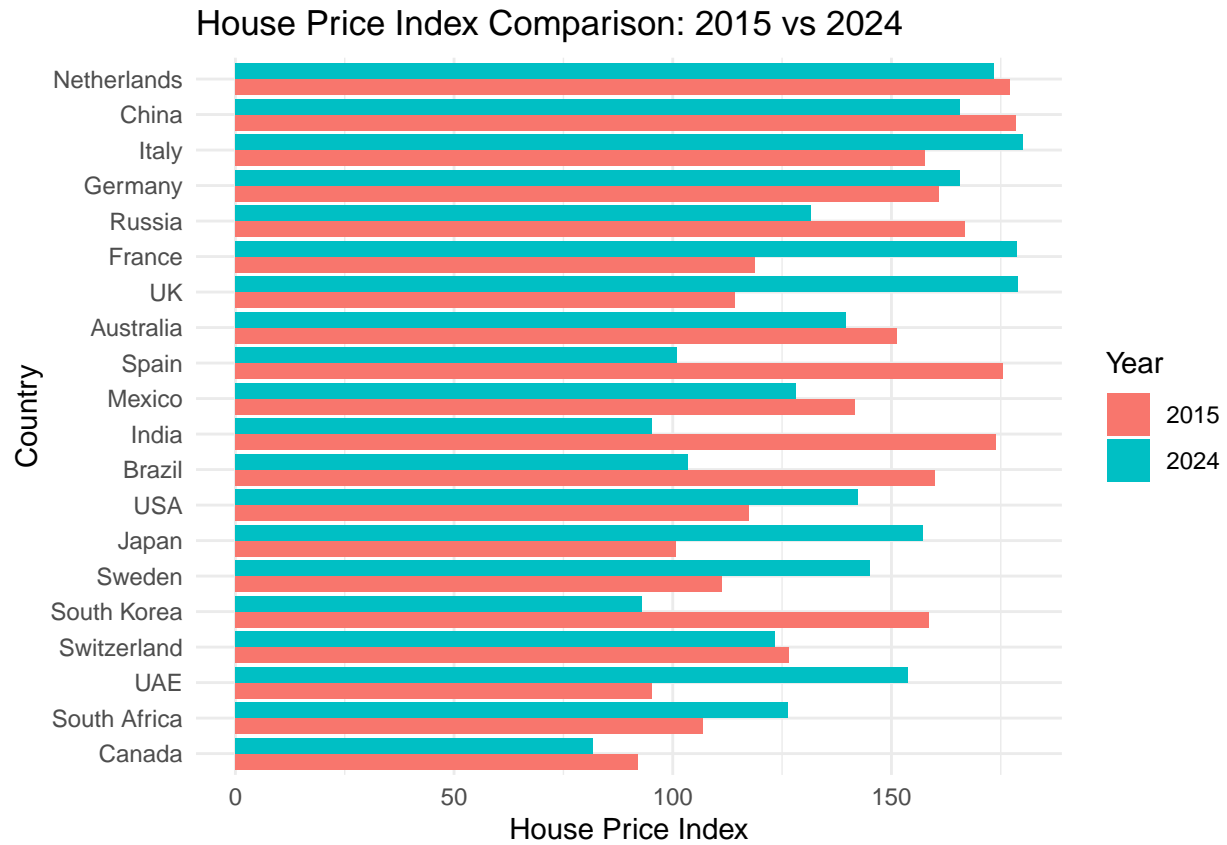


```
ggplot(House_market_coordinate, aes(x = `Population Growth (%)`, y = `House Price Index`, color = Count)) +
  geom_point(size = 3, alpha = 0.7) +
  geom_smooth(method = "lm", se = FALSE, linetype = "dashed", color = "black") +
  labs(title = "House Price Index vs Population Growth (%) by Year",
       x = "Population Growth (%)",
       y = "House Price Index") +
  facet_wrap(~Year) + # This creates a separate plot for each year
  theme_minimal()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



```
ggplot(House_market_coordinate %>% filter(Year %in% c(2015, 2024)),
  aes(x = reorder(Country, `House Price Index`),
    y = `House Price Index`,
    fill = factor(Year))) +
geom_bar(stat = "identity", position = "dodge") +
coord_flip() +
labs(title = "House Price Index Comparison: 2015 vs 2024",
  x = "Country",
  y = "House Price Index",
  fill = "Year") +
theme_minimal()
```



Main Heading

Subheading 1

Subheading 2

Subheading 3