IE360 - HW1

Derya Özşeker

13/04/2022

In this homework, data that are provided by EVDS are visually analyzed separately and with relation to each other. Search data provided by Google Trends are observed and are compared with the EVDS data. Data sets that are observed and analyzed are consumer price index of total products, employment rate and house sales. The question that is aimed to answer is whether monthly house sales are related with the other two aspects. All data are monthly and from January 2014 to January 2022.

Consumer price index (CPI) is a measure of the cost of the required goods and services for a typical consumer in order to live in an average urban city. It's used to analyze the economical situation with respect to time and other locations, reflecting the inflation. In this study, it's analyzed in order to understand the economical condition of an average person living in Turkey during different time periods so that the effect of the economical situation on the house sales and the relation between them can be further understood.

Employment Rate is the percentage of employed population over the working population, which is the population over 15 years old and under the retirement age, able to work. Seasonal workers are also counted in the working population. In this study, it's used to understand whether the employment situation at a given time is effective over the house sales at that time.

Gathering Required Packages and Data Sets

Install packages:

```
install.packages("readxl", repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpAHU8kE/downloaded
_packages

install.packages("ggplot2", repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpAHU8kE/downloaded
_packages
```

```r
install.packages("dplyr", repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpAHU8kE/downloaded
_packages

install.packages("tidyr", repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpAHU8kE/downloaded
_packages

install.packages("lubridate", repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpAHU8kE/downloaded
_packages

install.packages("RColorBrewer", repos = "http://cran.us.r-
project.org")

##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpAHU8kE/downloaded
_packages

install.packages("zoo", repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpAHU8kE/downloaded
_packages

library(readxl)
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

library(tidyr)
library(lubridate)

##
## Attaching package: 'lubridate'

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

library(RColorBrewer)
library(zoo)

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Import data sets:

```
EVDS_HouseSales <- read_xlsx("~/Desktop/HouseSales.xlsx")

## New names:
## • `` -> `...3`

Trends_WordSearch_SatılıkDaire = read.csv("~/Desktop/SatılıkDaire.csv",
dec=",")
Trends_WordSearch_Emlak = read.csv("~/Desktop/Emlak.csv", dec=",")
Trends_WordSearch_Konut = read.csv("~/Desktop/Konut.csv", dec=",")
EVDS_CPI <- read_xlsx("~/Desktop/CPI.xlsx")

## New names:
## • `` -> `...3`

Trends_WordSearch_Alışveriş = read.csv("~/Desktop/Alışveriş.csv",
dec=",")
Trends_WordSearch_Market = read.csv("~/Desktop/Market.csv", dec=",")
Trends_WordSearch_Fiyat = read.csv("~/Desktop/Fiyat.csv", dec=",")
EVDS_InterestRate <- read_xlsx("~/Desktop/InterestRate.xlsx")

## New names:
## • `` -> `...3`

Trends_WordSearch_KonutKredisi = read.csv("~/Desktop/KonutKredisi.csv",
dec=",")
EVDS_EmploymentRate <- read_xlsx("~/Desktop/EmploymentRate.xlsx")
```

```
## New names:
## • `` -> `...3`

Trends_WordSearch_İşİlanı = read.csv("~/Desktop/İşİlanı.csv", dec=",")
Trends_WordSearch_İşBaşvurusu = read.csv("~/Desktop/İşBaşvurusu.csv",
dec=",")
```

Preparation of Data

Manipulate data sets:

```
Date=EVDS_HouseSales$'Tarih'[13:109]
HouseSales=EVDS_HouseSales$'TP AKONUTSAT1 TOPLAM'
EmploymentRate=EVDS_EmploymentRate$`TP YISGUCU2 G7`
CPI=EVDS_CPI$`TP FG J0`
InterestRate=EVDS_InterestRate$`TP KTF12`

WordSearch_SatılıkDaire =
as.numeric(Trends_WordSearch_SatılıkDaire$'Kategori..Gayrimenkul'[123:2
19])
WordSearch_Emlak
=as.numeric(Trends_WordSearch_Emlak$'Kategori..Gayrimenkul'[123:219])
WordSearch_Konut =
as.numeric(Trends_WordSearch_Konut$'Kategori..Gayrimenkul'[123:219])
WordSearch_Alışveriş =
as.numeric(Trends_WordSearch_Alışveriş$'Kategori..Alışveriş'[123:219])
WordSearch_Market
=as.numeric(Trends_WordSearch_Market$'Kategori..Alışveriş'[123:219])
WordSearch_Fiyat =
as.numeric(Trends_WordSearch_Fiyat$'Kategori..Alışveriş'[123:219])
WordSearch_KonutKredisi
=as.numeric(Trends_WordSearch_KonutKredisi$'Kategori..Tüm.kategoriler'[
123:219])
WordSearch_İşBaşvurusu
=as.numeric(Trends_WordSearch_İşBaşvurusu$'Kategori..Tüm.kategoriler'[1
23:219])
WordSearch_İşİlanı =as.numeric(Trends_WordSearch_İşİlanı
$'Kategori..Tüm.kategoriler'[123:219])

years <-  c('2014', '2015', '2016', '2017', '2018', '2019',
'2020','2021','2022')
Date=EVDS_HouseSales$Tarih[13:109]
Year=0
```

Form data frames:

```
Dataframe_Trends_HouseSales  =
data.frame(Date=Date,TrendsHouseSales=WordSearch_SatılıkDaire+WordSearc
h_Emlak+WordSearch_Konut)
Dataframe_Trends_EmploymentRate=
data.frame(Date=Date,TrendsEmploymentRate=WordSearch_İşBaşvurusu+WordSe
```

```
arch_İşİlanı)
Dataframe_Trends_CPI=
data.frame(Date=Date,TrendsCPI=WordSearch_Alışveriş+WordSearch_Market+W
ordSearch_Fiyat)
Dataframe_Trends_InterestRate=
data.frame(Date=Date,TrendsInterestRate=WordSearch_KonutKredisi)

Dataframe_HouseSales =
data.frame(Year=Year,Date=Date,HouseSales=as.integer(HouseSales[13:109]
))
Dataframe_EmploymentRate=
data.frame(Year=Year,Date=Date,EmploymentRate=as.numeric(EmploymentRate
[1:97]))
Dataframe_CPI=
data.frame(Year=Year,Date=Date,CPI=as.numeric(CPI[54:150]))
Dataframe_InterestRate=
data.frame(Year=Year,Date=Date,InterestRate=as.numeric(InterestRate[53:
149]))

Dataframe_HouseSales$Date=parse_date_time(Dataframe_HouseSales$Date,"Ym
")
Dataframe_HouseSales$Year=factor(format(Dataframe_HouseSales$Date,
"%Y"),levels = years)
Dataframe_EmploymentRate$Date=parse_date_time(Dataframe_EmploymentRate$
Date,"Ym")
Dataframe_EmploymentRate$Year=factor(format(Dataframe_EmploymentRate$Da
te, "%Y"),levels = years)
Dataframe_CPI$Date=parse_date_time(Dataframe_CPI$Date,"Ym")
Dataframe_CPI$Year=factor(format(Dataframe_CPI$Date, "%Y"),levels =
years)
Dataframe_InterestRate$Date=parse_date_time(Dataframe_InterestRate$Date
,"Ym")
Dataframe_InterestRate$Year=factor(format(Dataframe_InterestRate$Date,
"%Y"),levels = years)
```

Data Visualization and Analysis

1. Monthly House Sales Data in Turkey Between 2014 and 2022

EVDS data set includes monthly observations of total number of houses sold in
Turkey between 01.2014 and 01.2022; Google Trends data set includes the total
number of searches in the category of real estate between 01.2014 and 01.2022 of
the words "Satılık Daire", "Konut" and "Emlak", assuming that sum of the number of
searches of these categories will provide a reasonable lead about the number of
houses sold. The summation of these search numbers are used in order to explain
the correlation between these two sets of data.

Line plot of EVDS data:

```
ggplot(Dataframe_HouseSales, aes(x=Date, y=HouseSales)) +
  geom_line(color = "black", size = 1, aes(group=1),linetype="solid") +
  labs(x = "Date",
       y = "House Sales",
       title = "Monthly House Sales in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black")) +
  expand_limits(y=0)
```
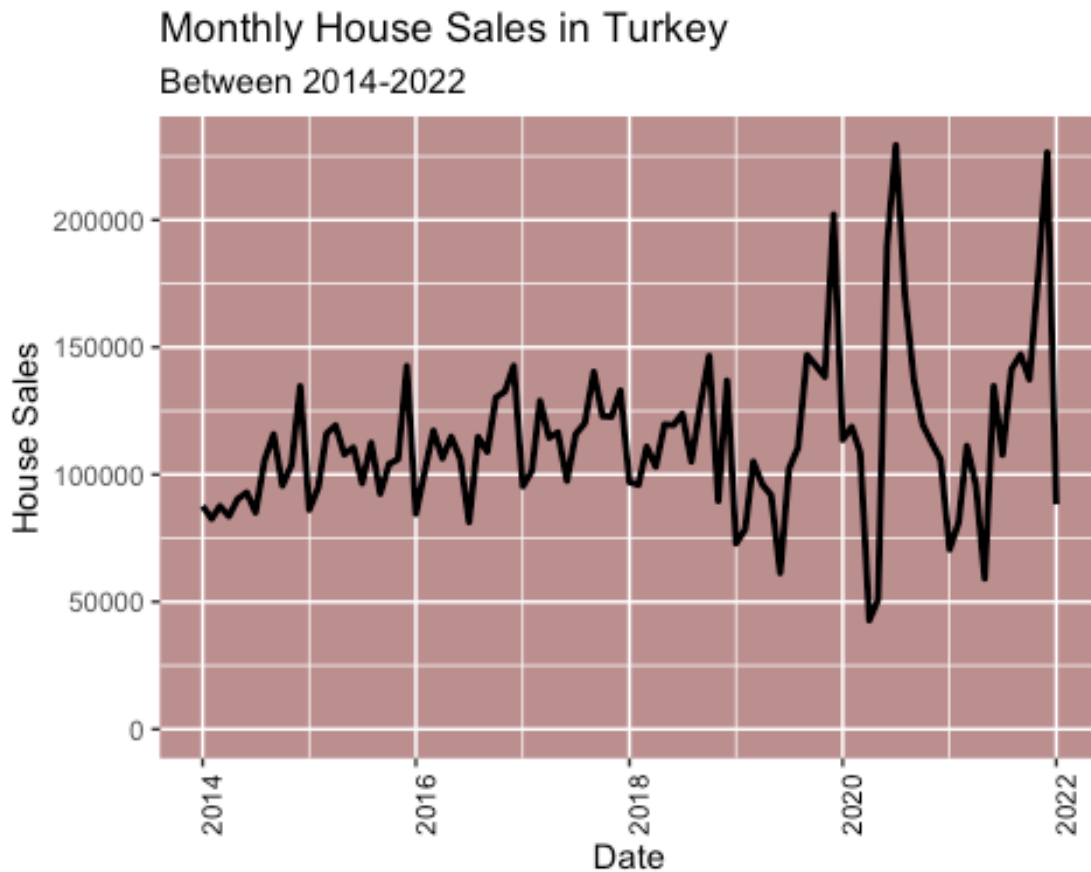

General trend of house sales between 2014 and 2022 is increasing. This can be explained by the growing population and increase in the number of houses constructed. There are visible peaks in September and December nearly every year, then sudden decreases in January is observed. This seasonality effect might be explained by the tendency to do major investments at the end of the year rather than at the beginning of the year because of possible new regulations at the beginning of the year. This phenomenon might also be affective on the additional sales in the summer and autumn months that is in general slightly more than the sales in winter and spring. Other explanation might be the increasing tendency to buy houses in the summer that are not used usually in the winter. This seasonality effect doesn't explain some of the peaks and crashes. The sudden crash in April 2020

can be explained by the pandemic and the global-local market freeze that it has caused. And, sudden peak in July 2020 can be seen as the comeback of this crash where many of the restrictions were lifted. It can also be seen that in the last years, the fluctuations has increased, which might be related to the fluctuating market conditions.

Line plot of Google Trends data:

```
ggplot(Dataframe_Trends_HouseSales, aes(x=Date, y=TrendsHouseSales)) +
  geom_line(color = "black", size = 1, aes(group=1),linetype="solid") +
  labs(x = "Date",
       y = "Search Count",
       title = "Google Search Count Related to House Sales in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black")) +
  expand_limits(y=0)
```
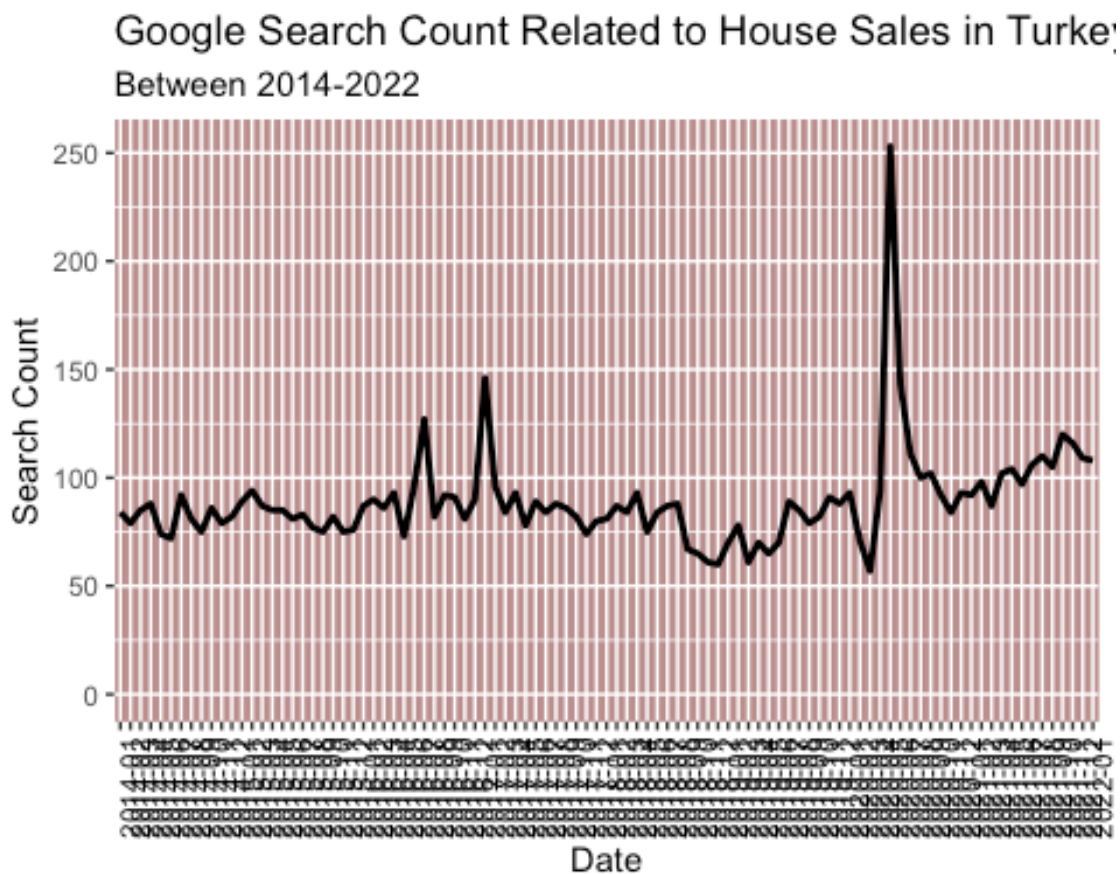

The correlation between the number of searches in Google for sum of the chosen words, which are "Satılık Daire", "Emlak" and "Konut", and the number of houses sold seems to be very strong. This shows that number of searches for houses online increases correlated with the tendency to buy houses. The Google Trends data also

show the same seasonality effect that is explained previously. Also, the sudden market crash and peak in 2020 are visible in this data as well. However, some peaks and crashes are seen in the months that are just before the months that the EVDS data show the peaks or crashes. This can be explained by the tendency to search for the houses prior to the actual sales.

Yearly histograms:

```
ggplot(Dataframe_HouseSales%>%group_by(Year), aes(HouseSales)) +
  geom_histogram(bins=24,color = "black", size = 0.25,
aes(fill=Year,color=Year),linetype="solid") +
  facet_wrap(Year~.,scales = "free_x",ncol = 4) +
  scale_color_brewer(palette = "Paired") +
  scale_fill_brewer(palette = "Paired") +
  labs(x = "House Sales",
       title = "Yearly Histograms of House Sales in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black"))
```
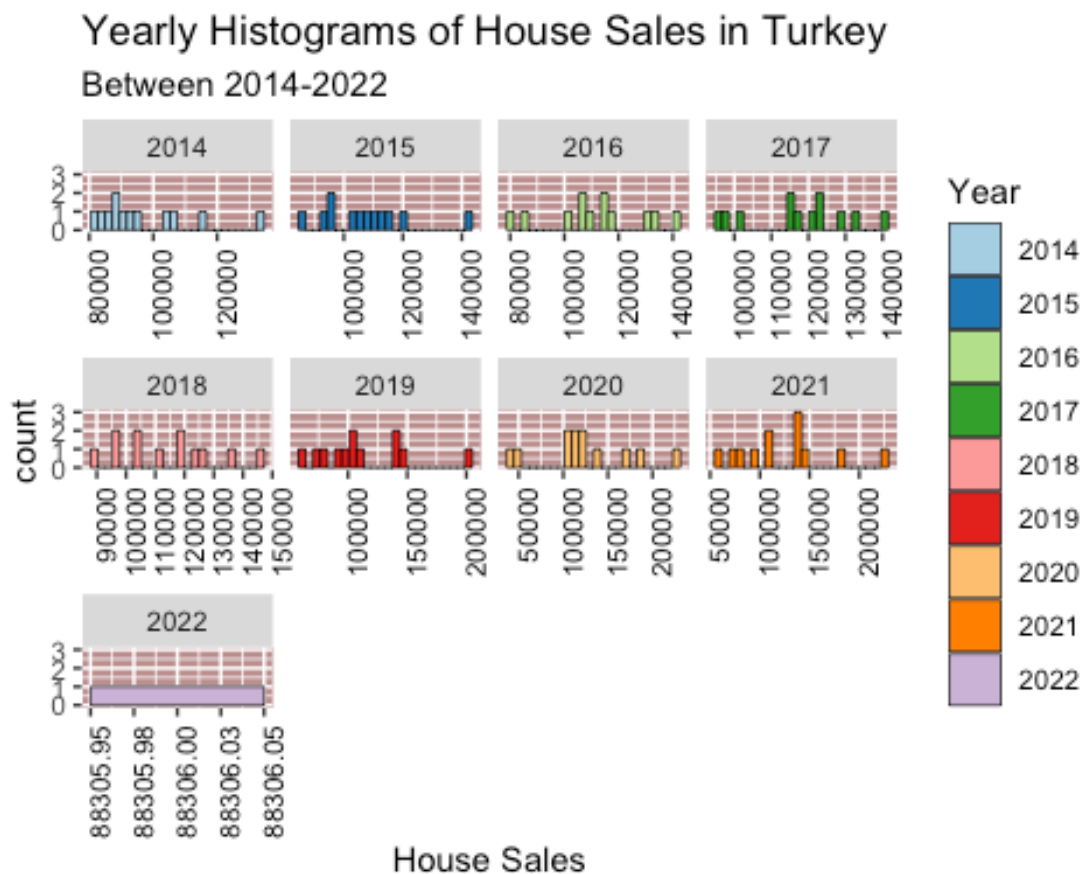

From the yearly histograms, it can be seen that the mean of the house sales increase slightly from 2014 to 2021, which was also visible in the line graphs. However, the

variance also increase significantly, which can be seen from the interval range of data. This variance increase explains the increasing fluctuations in the later years that was visible in the line graphs. Other than these changes, the general shape of the histograms are similar, which shows the general tendency between the years are similar.

Yearly box plots:

```
ggplot(Dataframe_HouseSales%>%group_by(Year), aes(x=Date,
y=HouseSales)) +
  geom_boxplot(color = "black", size = 0.25,
aes(fill=Year,color=Year),linetype="solid") +
  scale_fill_brewer(palette="Paired") +
  labs(x = "Date",
       y = "House Sales",
       title = "Yearly Box Plots of House Sales in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black"))
```
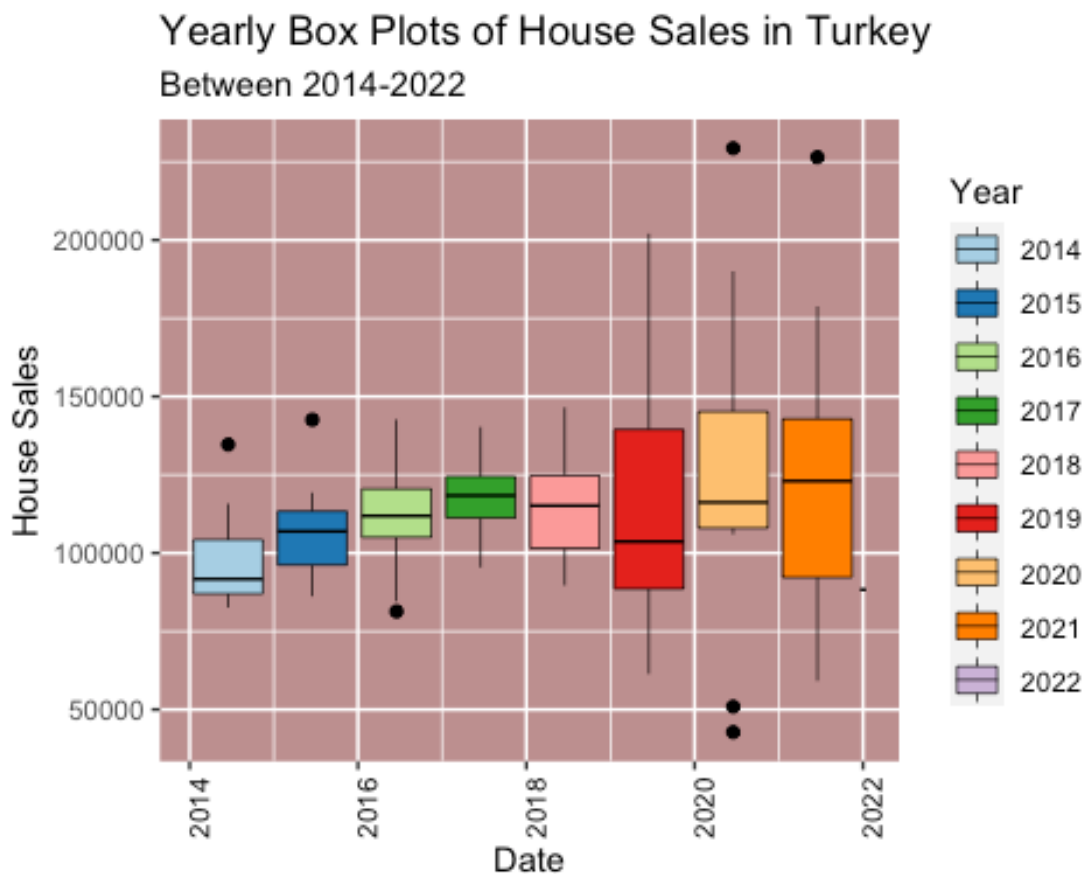

The yearly box plots show the increasing variance of the house sales through the years. Especially the house sales in the last three years show major increase in the

inter quartile range, which explains the increased fluctuations in the sales within the years. The mean house sales show a steady increase between 2014 and 2017, an irregular decrease between 2018 and 2020 possibly due to the economical environment, and increasing trend from 2020.

2. Monthly Employment Rate Data in Turkey Between 2014 and 2022

EVDS data set includes monthly observations of employment rate in Turkey between 01.2014 and 01.2022; Google Trends data set includes the total number of searches in the general category between 01.2014 and 01.2022 of the words "İş başvurusu" and "İş ilanı", assuming that sum of the number of searches of these categories will provide a reasonable lead about the employment rate. The summation of these search numbers are used in order to explain the correlation between these two sets of data.

Line plot of EVDS data:

```
ggplot(Dataframe_EmploymentRate, aes(x=Date, y=EmploymentRate)) +
  geom_line(color = "black", size = 1, aes(group=1),linetype="solid") +
  labs(x = "Date",
       y = "Employment Rate (%)",
       title = "Monthly Employment Rate in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black")) +
  expand_limits(y=0)
```

## Monthly Employment Rate in Turkey
Between 2014-2022
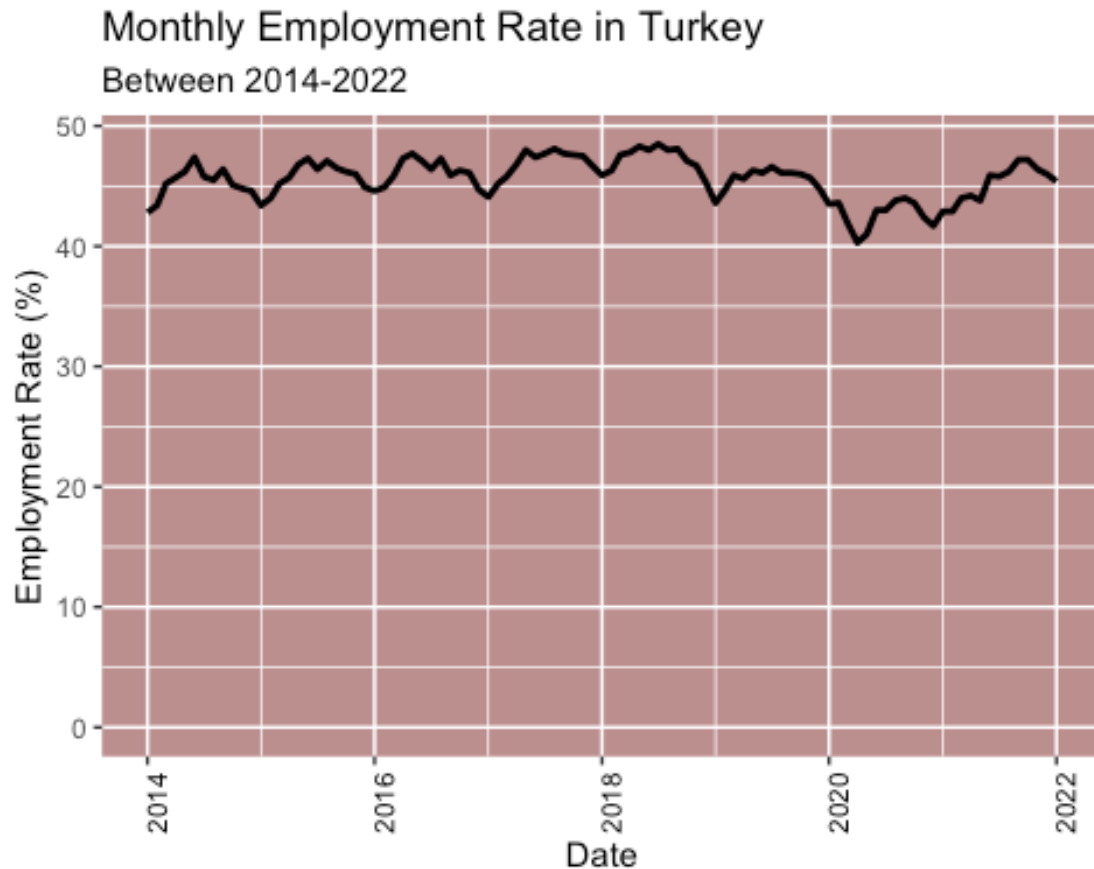
Employment rate in Turkey seems to be at approximately 45 percent of the working population. Between 2014 and 2018, the employment rate seems to be at a slow increase, however it starts to decrease after 2018. There is visible yearly seasonality in the employment rate data. The peaks in the employment rate in the summer and early autumn and decreasing tendency in the winter months might be explained by the seasonal workers that only work in the summer. However, in 2020, the employment rate decreases significantly, which shows a major similarity with the house sales, proving that the market freeze caused by the pandemic in 2020 decreased the employment rate, related with the sunken businesses and laid-off workers. The employment rate seems to be coming to its previous level since 2020.

Line plot of Google Trends data:

```
ggplot(Dataframe_Trends_EmploymentRate, aes(x=Date,
y=TrendsEmploymentRate)) +
  geom_line(color = "black", size = 1, aes(group=1),linetype="solid") +
  labs(x = "Date",
       y = "Search Count",
       title = "Google Search Count Related to Employment Rate in
Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
```

```
90,vjust=1,hjust=1,color="black")) +
  expand_limits(y=0)
```

Google Search Count Related to Employment Rate in T
### Between 2014-2022
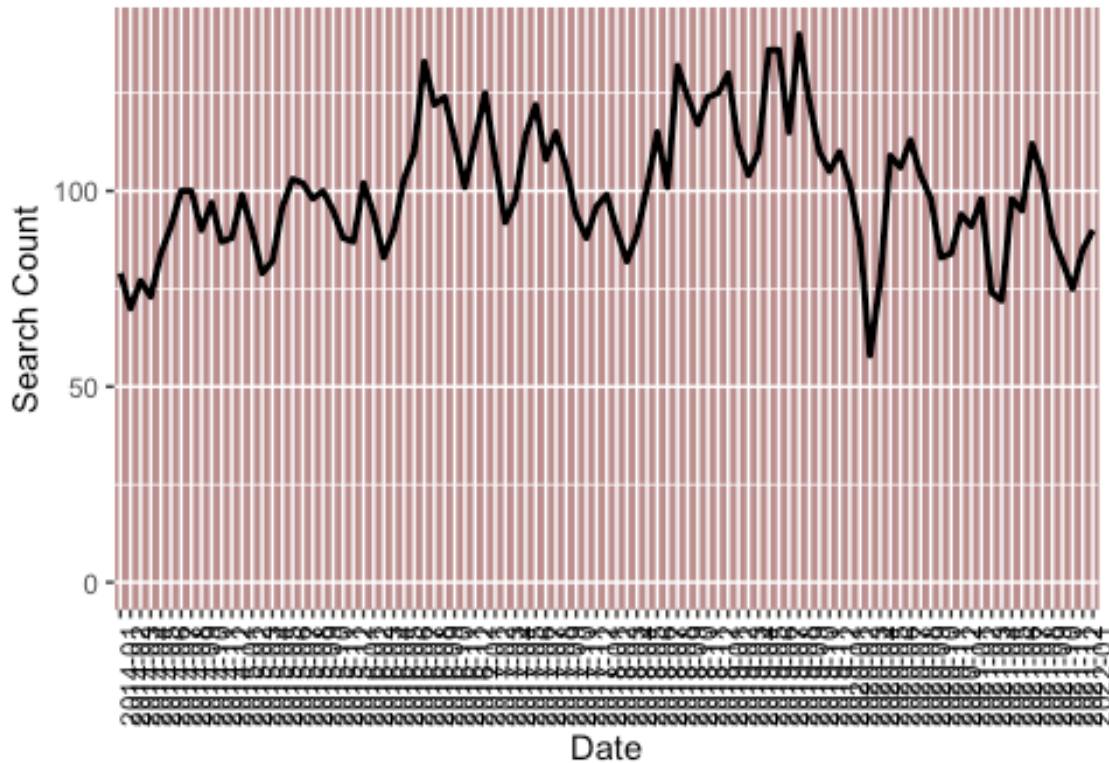
EVDS data and Google Trends data show a clear correlation between them, however the number of searches is much more fluctuating than the employment rate, which is understandable since the employment rate is not very volatile under normal circumstances. Google Trends data show the same seasonality effect with the EVDS data. The increasing number of searches for jobs in the summer and autumn months and the decreasing number of searches for jobs in the winter months show the similar seasonality condition. The volume of searches seems to be at a much higher level after 2016 till 2020. At March 2020, very similar to the house sales, the search for jobs suddenly decreases significantly, which may be the result of closed market place, interrupted working environment, no new hires in most businesses because of the pandemic.

Yearly histograms:

```
ggplot(Dataframe_EmploymentRate%>%group_by(Year), aes(EmploymentRate))
+
  geom_histogram(bins=24,color = "black", size = 0.25,
aes(fill=Year,color=Year),linetype="solid") +
  facet_wrap(Year~.,scales = "free_x",ncol = 4) +
  scale_color_brewer(palette = "Paired") +
```

```
  scale_fill_brewer(palette = "Paired") +
  labs(x = "Employment Rate (%)",
       title = "Yearly Histograms of Employment Rate (%) in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black"))
```
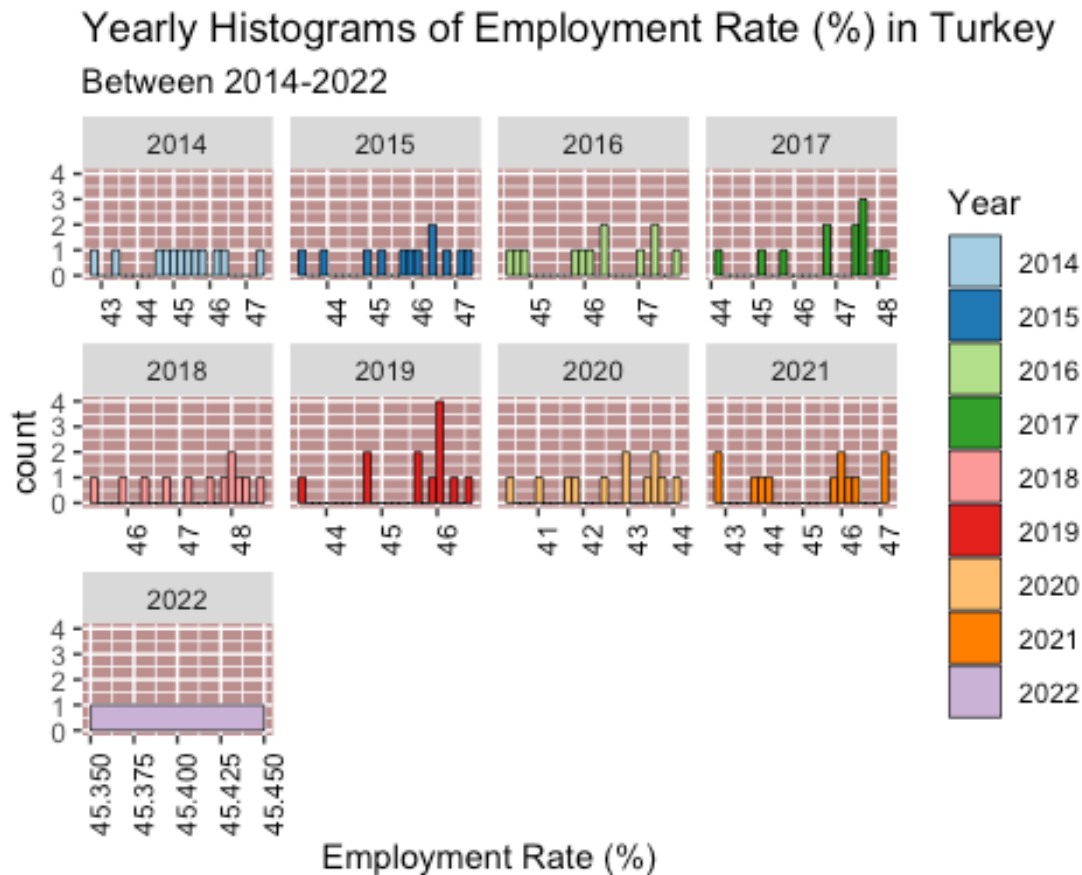

The yearly histograms show that till 2016, the data is evenly distributed with a mean of 45% as it was also shown in the line graphs. However, after 2017, the histograms get more and more skewed to the left, leaving most of the observations on the right side of the graph and forming a shape closer to the normal distribution. This causes a slight shift in the mean of the data towards right. And, in 2020, this left skewedness gets lesser and the mean of the data decreases significantly. In 2021, half of the data forms a cluster on the left whereas the other half forms a cluster on the right side. This shows the increasing fluctuations within the year, showing an unstable employment rate.
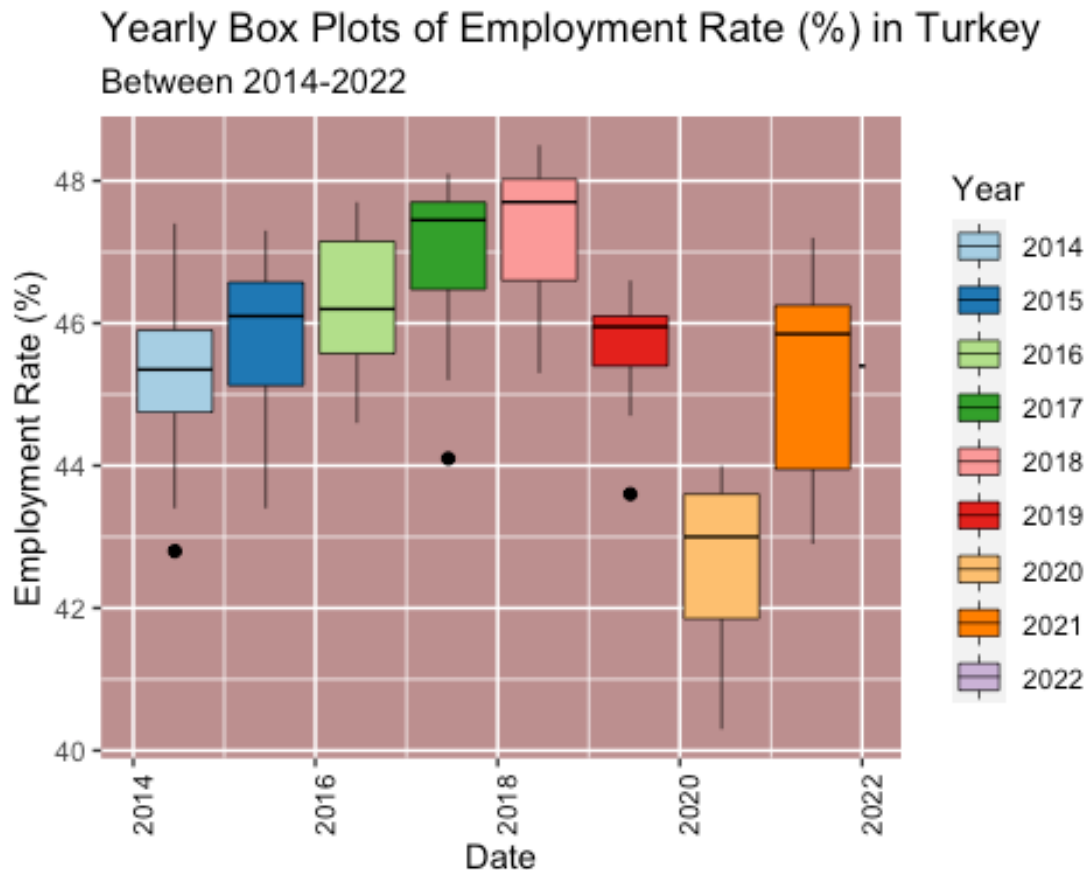
Yearly box plots:

```
ggplot(Dataframe_EmploymentRate%>%group_by(Year), aes(x=Date,
y=EmploymentRate)) +
  geom_boxplot(color = "black", size = 0.25,
```

```
aes(fill=Year,color=Year),linetype="solid") +
  scale_fill_brewer(palette="Paired") +
  labs(x = "Date",
       y = "Employment Rate (%)",
       title = "Yearly Box Plots of Employment Rate (%) in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black"))
```


As it can be seen from the yearly box plots, there is a slow increase in the employment rate between 2014 and 2018. However, it starts to decrease after 2018 and crashes suddenly at 2020. At 2021, the mean employment rate increases considering the 2020 box plot to a similar point with before 2018 rate, but with a much greater variance. Increasing trend till 2018, the halt after 2018 and decay in 2020 -after which a much volatile range is seen- show the similarity with the house sales box plots, showing that either employment rate is effective on the houses sold or the general economical condition affects both house sales and employment rate with the same manner.

3. Monthly Consumer Price Index Data in Turkey Between 2014 and 2022

EVDS data set includes monthly observations of consumer price index in Turkey between 01.2014 and 01.2022; Google Trends data set includes the total number of searches in the category of shopping between 01.2014 and 01.2022 of the words "Alışveriş", "Market" and "Fiyat", assuming that sum of the number of searches of these categories will provide a reasonable lead about the consumer price index (CPI). The summation of these search numbers are used in order to explain the correlation between these two sets of data.

Line plot of EVDS data:

```
ggplot(Dataframe_CPI, aes(x=Date, y=CPI)) +
  geom_line(color = "black", size = 1, aes(group=1),linetype="solid") +
  labs(x = "Date",
       y = "Consumer Price Index (%)",
       title = "Monthly Consumer Price Index in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black")) +
  expand_limits(y=0)
```

Monthly Consumer Price Index in Turkey
### Between 2014-2022
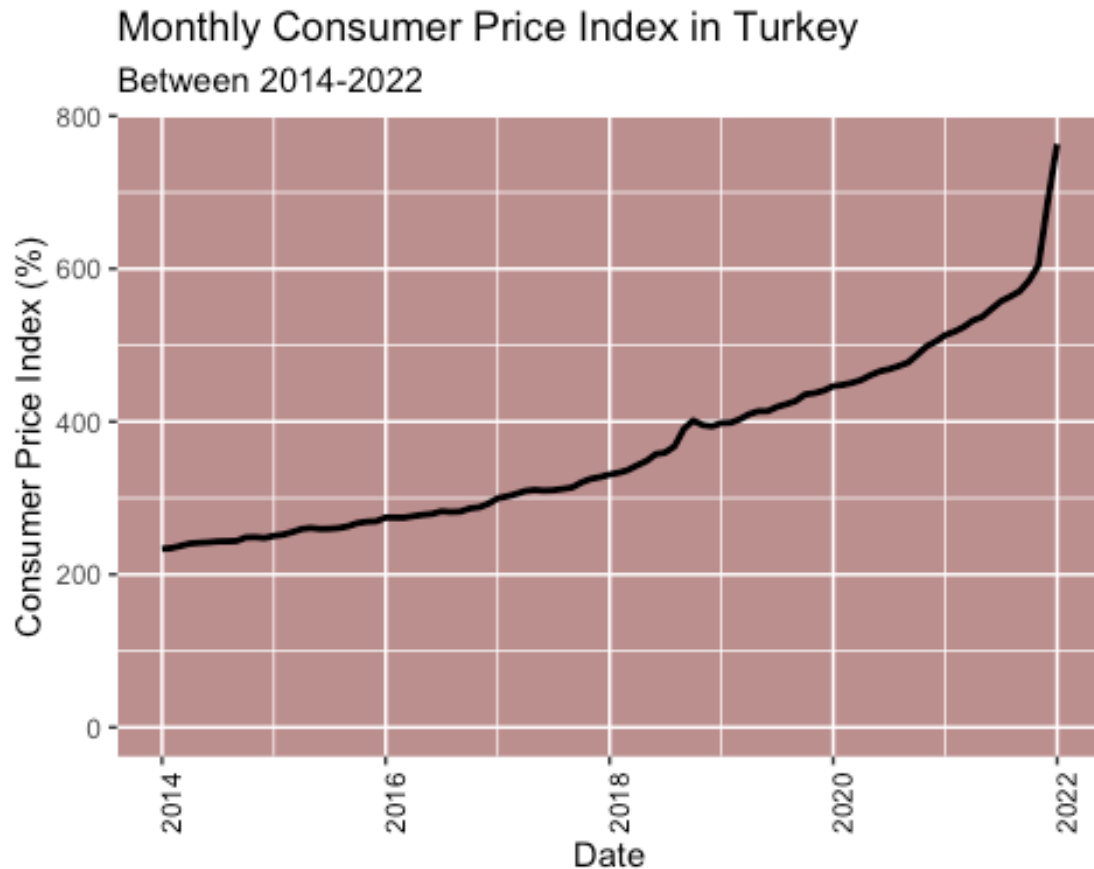
The consumer price index seem to be increasing exponentially from 2014 to 2022 with a slow increase till 2018, a break in 2018, which caused an increase at a higher rate and an almost vertical increase in the last year. The acceleration of rise is very large at the end of 2021 and at the beginning of 2022. As consumer price index is strongly correlated with inflation, the increasing inflation rate and the increasing gap between Turkish Liras and Euros or US Dollars might be the reason for the rapid consumer price increase in the last year. The general increasing tendency of this gap in the last decade might explain the increasing consumer price index.

Line plot of Google Trends data:

```
ggplot(Dataframe_Trends_CPI, aes(x=Date, y=TrendsCPI)) +
  geom_line(color = "black", size = 1, aes(group=1),linetype="solid") +
  labs(x = "Date",
       y = "Search Count",
       title = "Google Search Count Related to Consumer Price Index in
Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black")) +
  expand_limits(y=0)
```

Google Search Count Related to Consumer Price Index
Between 2014-2022
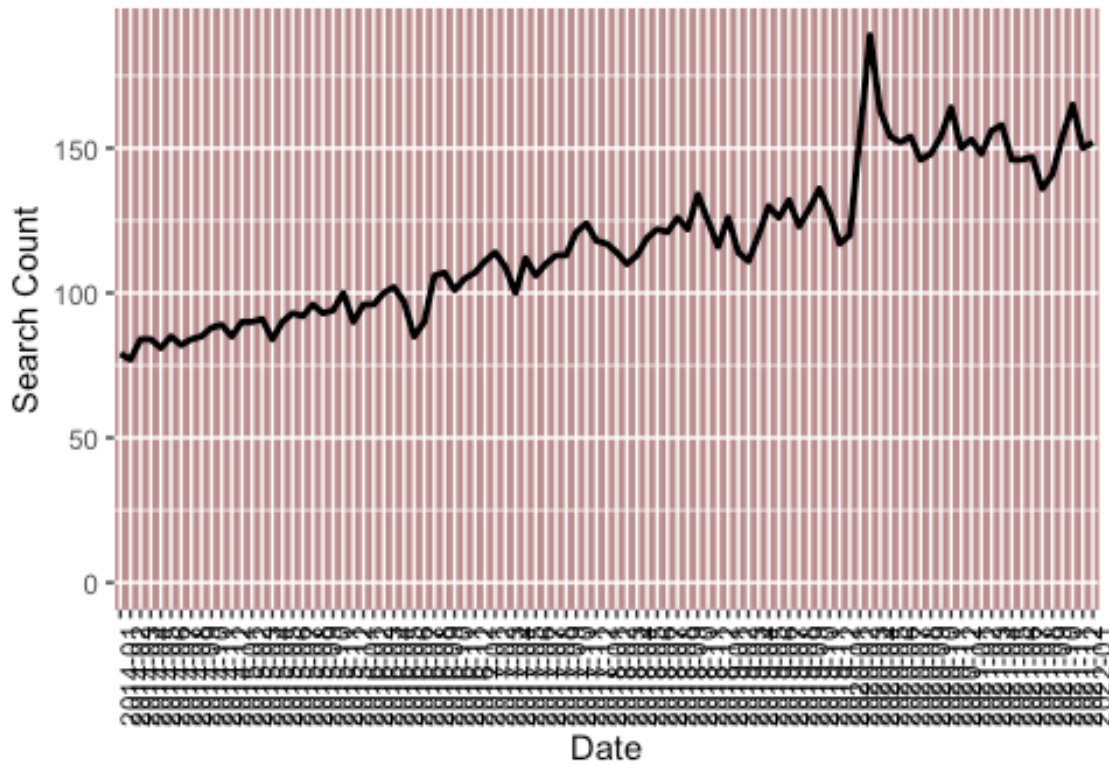
The total number of searches provided by the Google Trends data show a significant incresing trend that is very similar to the CPI data, proving the correlation between the two data sets. In March 2020, the number of searches has a sudden peak and the line gets shifted upwards after this peak and continues the increasing pattern at a new level. This coincides with the CPI graph, in which the rate of the increase shifts to a higher level in 2021.

Yearly histograms:

```
ggplot(Dataframe_CPI%>%group_by(Year), aes(CPI)) +
  geom_histogram(bins=24,color = "black", size = 0.25,
aes(fill=Year,color=Year),linetype="solid") +
  facet_wrap(Year~.,scales = "free_x",ncol = 4) +
  scale_color_brewer(palette = "Paired") +
  scale_fill_brewer(palette = "Paired") +
  labs(x =  "Consumer Price Index (%)",
       title = "Yearly Histograms of Consumer Price Index in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black"))
```

# Yearly Histograms of Consumer Price Index in Turkey
Between 2014-2022
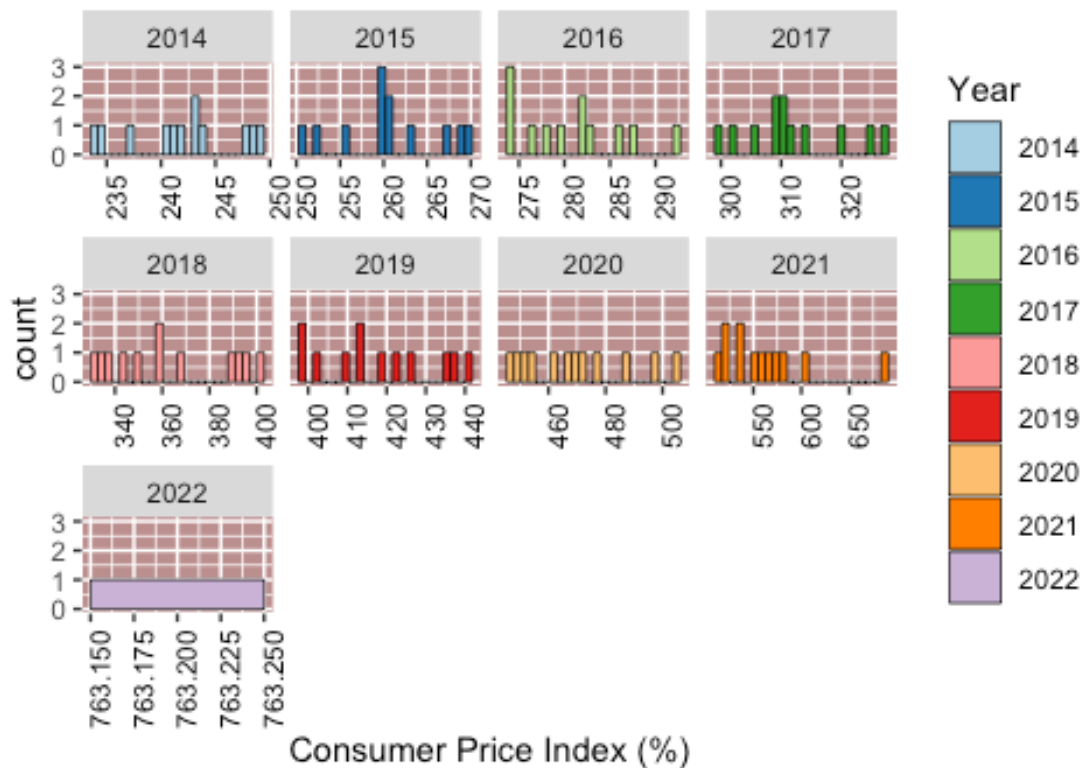
The yearly histograms show that the CPI value increases rapidly and significantly as the years pass. It also shows that the change within the year increases as well, as the data gets distributed between the values more evenly, the range of values gets larger, the gaps between the values get larger and the shape of the distribution gets farther away from the normal distribution. This shows that the rate of growth of the CPI value gets larger through time.

Yearly box plots:

```
ggplot(Dataframe_CPI%>%group_by(Year), aes(x=Date, y=CPI)) +
  geom_boxplot(color = "black", size = 0.25,
aes(fill=Year,color=Year),linetype="solid") +
  scale_fill_brewer(palette="Paired") +
  labs(x = "Date",
       y = "Consumer Price Index (%)",
       title = "Yearly Box Plots of Consumer Price Index in Turkey",
       subtitle = "Between 2014-2022") +
  theme(panel.background = element_rect(fill = "rosy brown"))+
  theme(axis.text.x = element_text(angle =
90,vjust=1,hjust=1,color="black"))
```

Yearly Box Plots of Consumer Price Index in Turkey
Between 2014-2022
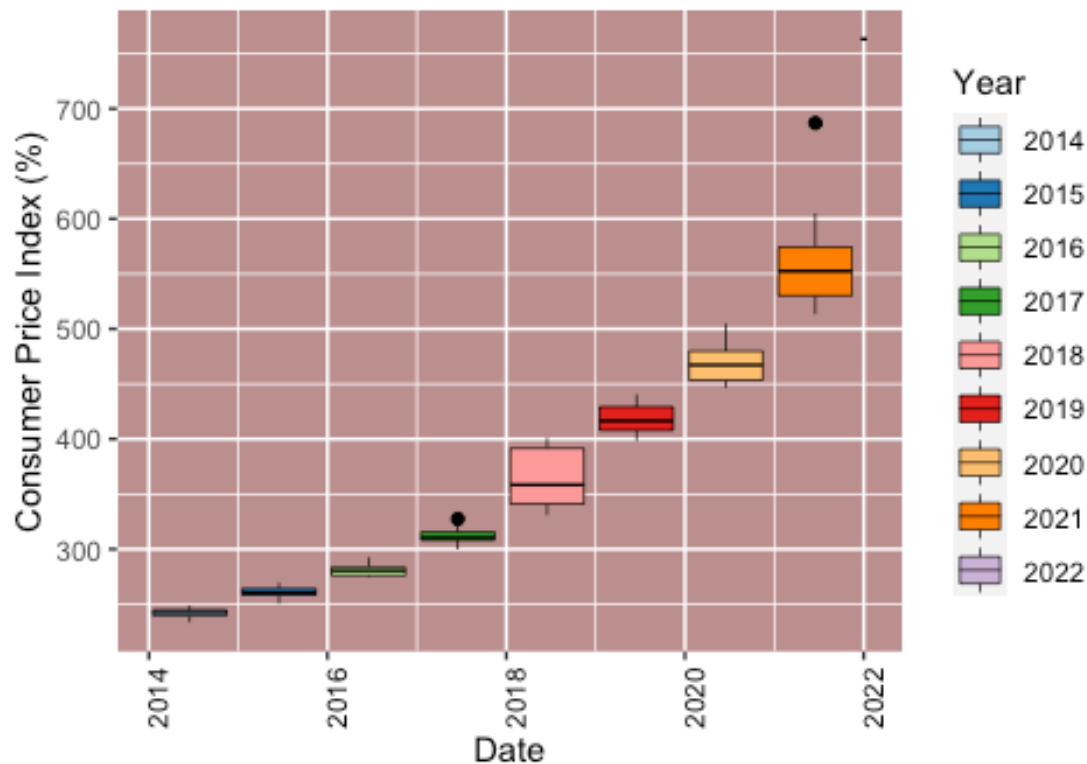
The yearly box plots prove both the rapid increase and the increase of the variance within the year, as it shows that the box plots till 2018 is much narrower than the others. In 2018 and in 2021, the variance within the year, the inter quartile range of data gets increased, which proves that at these years, there are breaking points in the CPI values that causes a faster increase. The increasing trend of the CPI value gets two breaks causing an upwards turn in the years that both house sales and employment rate get breaks causing a downwards turn. This shows that economical condition creates a correlation between all three measurements.

After analyzing every aspect of the employment rate, consumer price index and house sales from January 2014 to January 2022, it can be concluded that there is a definite correlation between these values that may be affected by each other. Consumer price index increase reflects the inflation, which does not create a major change in the economical standards of the public at a normal level but causes the economical standards to go down if it's at a very high rated increasing trend. This results in a lower tendency to buy houses, which might make the house sales go down. The employment rate decrease and too much fluctuating employment rate may cause an instability in the economical situation of the public. This also may decrease the house sales increase rate. The trend in the house sales may be affected by the other two factors, however the causality relation between these three variables cannot be concluded from this analysis.