

# Veri Setlerinin İlişkilerinin ve Zaman Serilerinin Görselleştirilmesi

Sahranur İnce

30 Kasım 2022

## Özet

Bu raporda, Kaggle'dan alınan 2 farklı veri setinin ilişkilerinin görselleştirme çalışmaları ve R'dan elde edilen zaman serilerinin görselleştirme çalışmaları bulunmaktadır. Kaggle'dan alınan ilk veri seti 2009-2019 yılları arasında Amazon'da en çok satan 50 kitap ve ikinci veri seti Türk TV dizileri hakkındadır. R'dan elde edilen zaman serileri ise döviz kurları hakkındadır. İlk önce son on yılda Türk lirasının, Euro(EUR), Amerikan doları(USD) ve Pound(GBP) karşılıklarının değişimi veri görselleştirme yöntemleriyle incelenmiştir. Daha sonra, belirtilen yıllar arasında 3 farklı döviz kurunun (Avustralya Doları(AUD), Japonya Yeni(JPY), Kanada Doları(CAD)) Amerikan doları karşılıkları veri görselleştirme yöntemleriyle incelenmiştir.

## Gerekli Paketlerin Yüklenmesi

```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("priceR")
install.packages("tidyverse")
library(ggplot2)
library(dplyr)
library(priceR)
library(tidyverse)
```

# Uygulama 1: En Çok Satan Kitaplar

## Veri Setinin İncelenmesi

Bu veri seti 2009-2019 yılları arasında Amazon'da en çok satan 50 kitabı içermektedir. Bu veri setinde 550 gözlem ve 7 değişken vardır. Bu değişkenler şunlardır: Kitap Adı, Yazar, Okuyucu Puanı, Yorumlar, Fiyat, Tür.

**Kitap fiyatları ve okuyucu yorum sayıları arasındaki ilişkiyi veri görselleştirme yöntemleriyle araştırınız. Kitap fiyatlarındaki artışın okuyucu yorum sayıları üzerinde nasıl bir etkisi vardır?**

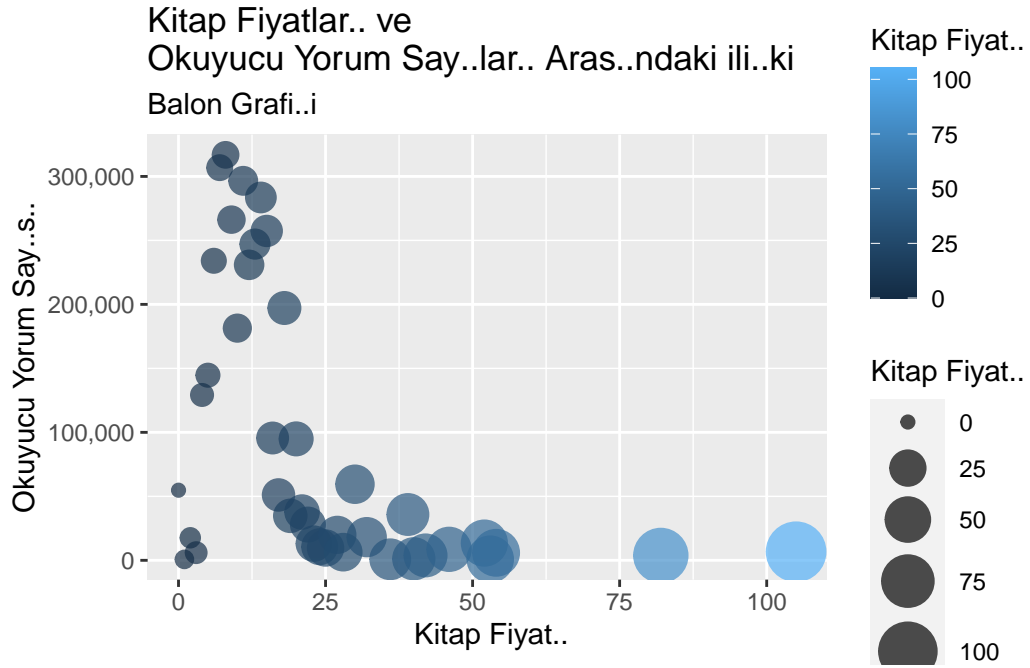
```
bestsellers_with_categories <- read_csv("bestsellers with categories.csv")
```

```
okuyucuyorum <- bestsellers_with_categories %>%  
  group_by(Price, Reviews) %>%  
  summarise(sayı1 = n(), sayı2 = n())
```

```
yorumsayıları <- okuyucuyorum %>%  
  group_by(Price) %>%  
  summarize(Toplamyorumsayısı = sum(Reviews))  
yorumsayıları
```

```
# A tibble: 40 x 2  
  Price Toplamyorumsayısı  
  <dbl>          <dbl>  
1     0           54787  
2     1            720  
3     2          17609  
4     3           5983  
5     4         129219  
6     5         144662  
7     6         234065  
8     7         306840  
9     8         316865  
10    9         266221  
# ... with 30 more rows
```

```
ggplot(yorumsayıları, aes(x=Price, y=Toplamyorumsayısı, size = Price, color=Price)) +
  scale_y_continuous(labels = scales::comma)+
  geom_point(alpha=0.7) +
  scale_size(range = c(2, 10), name="Kitap Fiyatı")+
  labs(x=" Kitap Fiyatı",
       y="Okuyucu Yorum Sayısı",
       size= "Kitap Fiyatı",
       color = "Kitap Fiyatı",
       title ="Kitap Fiyatları ve
Okuyucu Yorum Sayıları Arasındaki ilişki",
       subtitle = "Balon Grafiği")
```



Bu grafikte kitap fiyatları ve okuyucu yorum sayıları arasındaki ilişki görselleştirilmiştir. Grafikte balonların büyüklüğü ve renkleri kitapların fiyatına göre değişiklik gösterir. Kitapların fiyatı arttıkça baloncuklar büyür ve rengi açılır, azaldıkça küçülür ve rengi koyulaşır. Grafikte elde edilen sonuçlara göre kitap 1 arttıkça okuyucu yorum sayısının azaldığını, kitap fiyatı azaldıkça okuyucu yorum sayısının çoğaldığını görüyoruz. Bu durumda kitap fiyatı ile okuyucu yorum sayısı arasında negatif yönlü bir ilişki olduğunu söyleyebiliriz.

**Kitap fiyatları ve okuyucu yorum sayıları arasındaki ilişkiyi, kitap türüne göre veri görselleştirme yöntemleriyle araştırınız. Kitap fiyatlarındaki artışın okuyucu yorum sayıları üzerinde etkisi kitap türüne göre bir değişiklik göstermekte midir?**

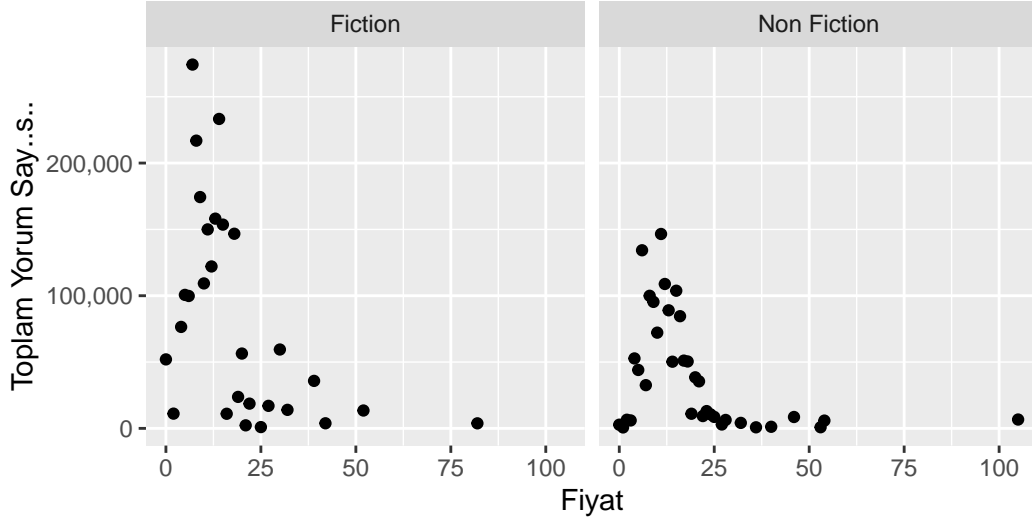
```
kitap <- bestsellers_with_categories %>%  
  group_by(Price,Reviews,Genre) %>%  
  summarise(sayı =n())
```

```
kitaptörü <- kitap %>%  
  group_by(Price, Genre) %>%  
  summarize(Toplamyorumsayısı = sum(Reviews))
```

```
ggplot(kitaptörü, aes(x = Price, y = Toplamyorumsayısı)) +  
  geom_point() +  
  scale_y_continuous(labels = scales::comma)+  
  labs(title = "mpg ve hp arasındaki ilişkinin vites türüne göre görselleştirilmesi") +  
  facet_grid(~Genre) +  
  labs(x = "Fiyat",  
       y= "Toplam Yorum Sayısı",  
       title = "Kitap Fiyatları ve Okuyucu Yorum Sayıları Arasındaki İlişkinin  
Kitap Türüne Göre Görselleştirilmesi",  
       subtitle= "Saçılım Grafiği")
```

## Kitap Fiyatlar.. ve Okuyucu Yorum Say..lar.. Aras..ndaki ..li.. Kitap Türüne Göre Görselle..tirilmesi

### Saç..l..m Grafi..i



Bu grafikte kitap fiyatları ve okuyucu yorum sayıları arasındaki ilişki kitap türüne göre incelenmiştir. Kitap türlerine göre iki farklı saçılım grafiği elde edilmiştir. Grafikte elde edilen sonuçlara göre; kurgusal kitap türünde fiyat arttıkça okuyucu yorum sayısının azaldığını, fiyat azaldıkça okuyucu yorum sayısının çoğaldığını görüyoruz. Kurgusal olmayan kitap türünde de aynı durum geçerlidir. Bu durumda her iki kitap türü içinde, fiyat ile okuyucu yorum sayısı arasında negatif yönlü bir ilişki olduğunu söyleyebiliriz. Bunlara ek olarak, kitap fiyatlarındaki artışın okuyucu yorum sayıları üzerinde etkisi kitap türüne göre bir değişiklik göstermemektedir.

## Uygulama 2: Türk TV dizileri

### Veri Setinin İncelenmesi

Bu veri seti türk tv dizileri hakkında veriler içermektedir. Bu veri setinde 2128 gözlem ve 103 değişken vardır. Değişkenlerde dizilerin adları, bölüm süreleri, sezon sayıları, reytingleri vb. değişkenler yer almaktadır.

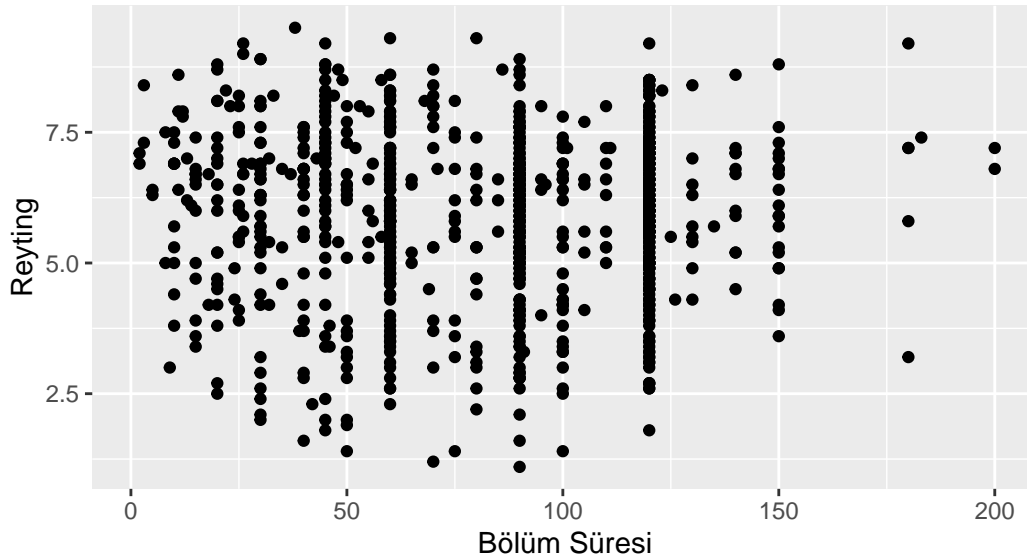
Bölüm süresi 200 dakika altında olan dizilerin bölüm süreleri ile reytingleri arasındaki ilişkiyi veri görselleştirme yöntemleriyle araştırınız. Bölüm süresinin artmasının reyting üzerinde nasıl bir etki göstermiştir?

```
turkish_tvseries <- read_csv("turkish_tvseries.csv")

newdata <- turkish_tvseries[ which(turkish_tvseries$runtimes <= 200),]

ggplot(newdata, aes(x = runtimes, y = rating)) +
  geom_point() +
  labs(x = "Bölüm Süresi",
       y = "Reyting",
       title = "Bölüm Süreleri ve Reytingler Arasındaki İlişkinin Görselleştirilmesi",
       subtitle = "Saçılım Grafiği")
```

Bölüm Süreleri ve Reytingler Arasındaki İlişkinin Görselleştirilmesi  
Saçılım Grafiği



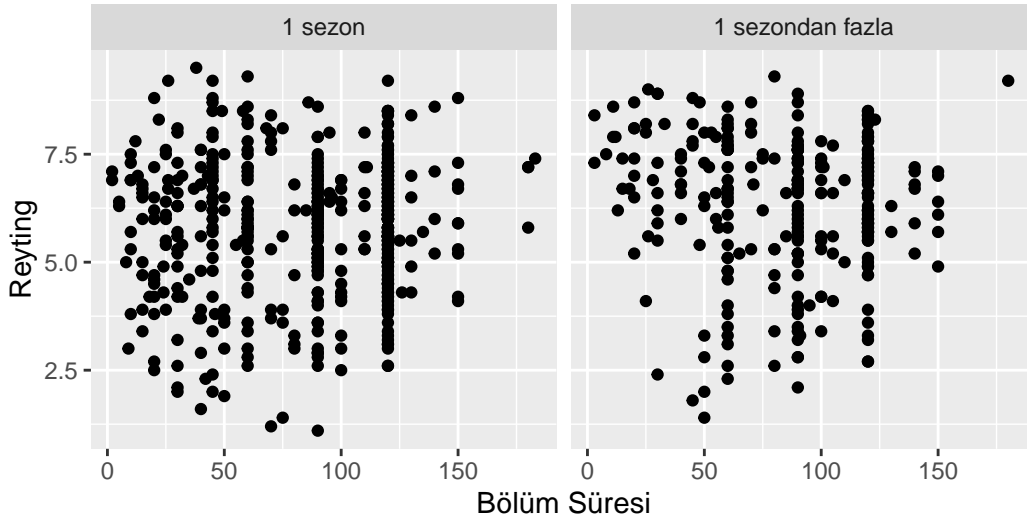
Bu grafikte, bölüm süreleri ve reytingler arasındaki ilişki görselleştirilmiştir. Saçılım grafiğinde elde edilen sonuçlara göre, bölüm süresi ile reyting arasında bir ilişki olmadığını söyleyebiliriz. Diğer bir deyişle, bölüm süresinin artmasının reyting üzerinde hiçbir etkisi yoktur.

Bölüm süresi 200 dakika altında olan dizilerin bölüm süreleri ile reytingleri arasındaki ilişkiyi, dizinin bir sezon veya daha fazla sürmesi durumuna göre veri görselleştirme yöntemleriyle araştırınız. Tek sezonluk ve daha uzun süren dizilerin bölüm süresinin artmasının reyting üzerinde oluşturduğu etki nasıl bir değişim göstermiştir?

```
newdata <- newdata %>%
  drop_na(`number of seasons`) %>%
  drop_na(runtimes) %>%
  drop_na(rating) %>%
  mutate(sezonsayısı = ifelse(`number of seasons` <= 1, "1 sezon", "1 sezondan fazla"))

ggplot(newdata, aes(x = runtimes, y = rating)) +
  geom_point() +
  labs(title = "mpg ve hp arasındaki ilişkinin vites türüne göre görselleştirilmesi") +
  facet_grid(~sezonsayısı) +
  labs(x = "Bölüm Süresi",
       y = "Reyting",
       title = "Bölüm Süreleri ve Reytingler Arasındaki İlişkinin
Sezon Sayısına Göre Görselleştirilmesi",
       subtitle = "Saçılım Grafiği")
```

Bölüm Süreleri ve Reytingler Arasındaki İlişkinin  
Sezon Sayısına Göre Görselleştirilmesi  
Saçılım Grafiği



Bu grafikte bölüm süreleri ve reytingler arasındaki ilişki sezon sayısına göre incelenmiştir. Sezon sayısı “1 sezon” ve “1 sezondan fazla” olmak üzere iki gruba ayrılmış ve iki ayrı saçılım grafiği elde edilmiştir. Grafikte elde edilen sonuçlara göre, hem 1 sezon hem de 1 sezondan fazla süren diziler için, bölüm süresi ile reytingler arasında bir ilişki olmadığını söyleyebiliriz. Diğer bir deyişle bölüm sürelerindeki artışın reytingler üzerinde etkisi sezon sayılarına göre bir değişiklik göstermemektedir.

## Uygulama 3: Döviz Kurları

**Son 10 yılda, Türk lirasının, Euro, Amerikan doları ve Pound karşılıklarının değişimini veri görselleştirme yöntemleriyle araştırınız.**

### Veri Setinin İncelenmesi

Bu kısımda 3 farklı zaman serisi veri seti bulunmaktadır. Birinci veri seti 01.01.2012 ile 30.11.2022 tarihleri arasında Türk Lirasının Euro karşısındaki değişimini, ikinci veri seti 01.01.2012 ile 30.11.2022 tarihleri arasında Türk Lirasının Amerikan Doları karşısındaki değişimini ve üçüncü veri seti 01.01.2012 ile 30.11.2022 tarihleri arasında Türk Lirasının Pound karşısındaki değişimini gösteren veriler içermektedir.

```
x <- historical_exchange_rates(from = "TRY", to = "EUR",
start_date = "2012-01-01", end_date = "2022-11-30")

ggplot(x, aes(x = date, y = one_TRY_equivalent_to_x_EUR)) +
  geom_line(color = "dodgerblue3") +
  labs(x = "Yıl",
       y = "TRYEUR",
       title = "2012-2022 Yılları Arasında Türk Lirasının
Euro Karşısındaki Değişimi",
       subtitle = "Çizgi grafiği") +
  theme_classic()
```



## 2012–2022 Yılları Arasında Türk Lirasının Euro Karşısındaki Değişimi

Çizgi grafiği



Grafikte elde edilen sonuçlara göre, 2012-2022 yılları arasında Türk Lirasının Euro'ya karşı uzun dönemli azalış gösterdiğini söyleyebiliriz. Türk lirası bazı zamanlarda kısa dönemli artışlar gösterse de zaman geçtikçe Euro'ya karşı sürekli değer kaybetmiştir.

```
y <- historical_exchange_rates(from = "TRY", to = "USD",  
start_date = "2012-01-01", end_date = "2022-11-30")  
  
ggplot(y, aes(x = date, y = one_TRY_equivalent_to_x_USD)) +  
  geom_line(color = "dodgerblue3") +  
  labs(x = "Yıl",  
       y = "TRYUSD",  
       title = "2012-2022 Yılları Arasında Türk Lirasının  
Amerikan Doları Karşısındaki Değişimi",  
       subtitle = "Çizgi Grafiği") +  
  theme_classic()
```

### 2012–2022 Yılları Arasında Türk Lirasının Amerikan Doları Karşısındaki Değişimi Çizgi Grafiği



Grafikte elde edilen sonuçlara göre, 2012-2022 yılları arasında Türk Lirasının Amerikan Dolarına karşı uzun dönemli azalış gösterdiğini söyleyebiliriz. Türk lirası bazı zamanlarda kısa dönemli artışlar gösterse de zaman geçtikçe Amerikan Dolarına karşı sürekli değer kaybetmiştir.

```
z <- historical_exchange_rates(from = "TRY", to = "GBP",  
start_date = "2012-01-01", end_date = "2022-11-30")  
  
ggplot(z, aes(x = date, y = one_TRY_equivalent_to_x_GBP)) +  
  geom_line(color = "dodgerblue3") +  
  labs(x = "Yıl",  
       y = "TRYGBP",  
       title = "2012-2022 Yılları Arasında Türk Lirasının  
Pound Karşısındaki Değişimi",  
       subtitle = "Çizgi Grafiği") +  
  theme_classic()
```

### 2012–2022 Y..llar.. Aras..nda Türk Liras..n..n Pound Kar....s..ndaki De..i..imi Çizgi Grafi..i



Grafikte elde edilen sonuçlara göre, 2012-2022 yılları arasında Türk Lirasının Pound'a karşı uzun dönemli azalış gösterdiğini söyleyebiliriz. Türk lirası bazı zamanlarda kısa dönemli artışlar gösterse de zaman geçtikçe Pound'a karşı sürekli değer kaybetmiştir. En keskin düşüşün 2013-2014 yılları arasında gerçekleştiğini söyleyebiliriz.

### Seçtiğiniz 3 farklı döviz kurunun Amerikan doları karşılıklarını veri görselleştirme yöntemleriyle araştırınız.

#### Veri Setinin İncelenmesi

Bu kısımda 3 farklı zaman serisi veri seti bulunmaktadır. Birinci veri seti 01.01.2015 ile 01.01.2022 tarihleri arasında Avustralya Dolarının Amerikan Doları karşısındaki değişimini, ikinci veri seti 01.01.2012 ile 01.01.2020 tarihleri arasında Japon Yenisinin Amerikan Doları karşısındaki değişimini ve üçüncü veri seti 01.01.2014 ile 01.01.2021 tarihleri arasında Kanada Dolarının Amerikan Doları karşısındaki değişimini gösteren veriler içermektedir.

```
t <- historical_exchange_rates(from = "AUD", to = "USD",  
start_date = "2015-01-01", end_date = "2022-01-01")  
  
ggplot(t, aes(x = date, y = one_AUD_equivalent_to_x_USD)) +  
  geom_line(color = "dodgerblue3") +  
  labs(x = "Yıl",
```

```

y = "AUDUSD",
title = "2015-2022 Yılları Arasında Avustralya Dolarının
Amerikan Doları Karşısındaki Değişimi",
subtitle = "Çizgi Grafiği") +
theme_classic()

```



Grafikte elde edilen sonuçlara göre, 2015-2022 yılları arasında Avustralya Doları Amerikan Doları karşısında bir süre artış gösterip, bir süre azalış göstermiştir. Avustralya Dolarının Amerikan Doları karşısında 2018-2020 yılları arasında en büyük düşüşünü yaşadığını ve 2020-2021 yılları arasında en büyük artışını yaşadığını söyleyebiliriz.

```

v <- historical_exchange_rates(from = "JPY", to = "USD",
start_date = "2012-01-01", end_date = "2020-01-01")

ggplot(v, aes(x = date, y = one_JPY_equivalent_to_x_USD)) +
  geom_line(color = "dodgerblue3") +
  labs(x = "Yıl",
y = "JPYUSD",
title = "2012-2020 Yılları Arasında Japon Yenisinin
Amerikan Doları Karşısındaki Değişimi",
subtitle = "Çizgi Grafiği") +

```

```
theme_classic()
```

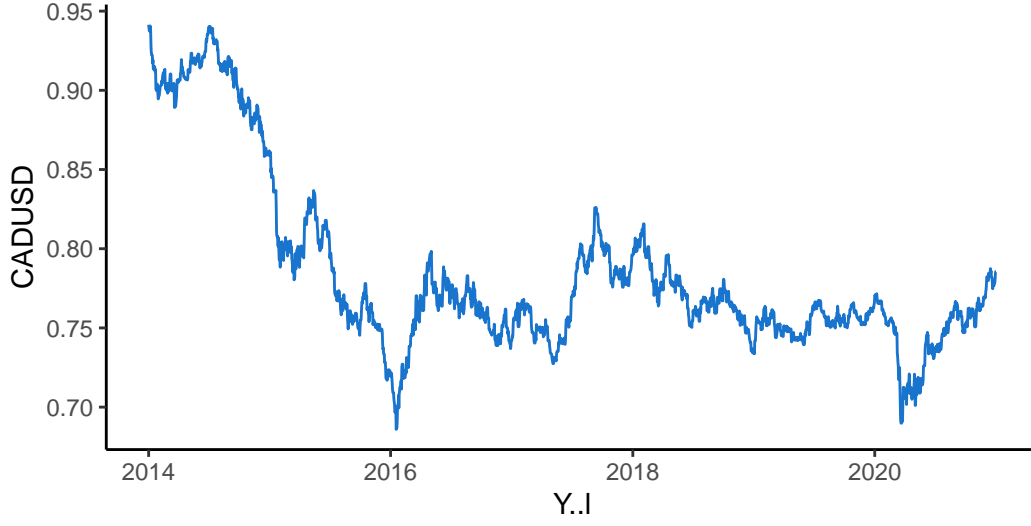
### 2012–2020 Y..llar.. Aras..nda Japon Yenisinin Amerikan Dolar.. Kar....s..ndaki De...imi Çizgi Grafi..i



Grafikte elde edilen sonuçlara göre, 2012-2016 yılları arasında Japon Yenisinin Amerikan Doları karşısında bazı zamanlarda artış gösterse de sürekli değer kaybettiğini görüyoruz. 2016-2017 yılları arasında tekrardan artış gösterip, daha sonra tekrardan düşüş göstermiştir. Japon Yenisinin 2018-2020 yılları arasında Amerikan Dolarına karşı büyük bir artış veya azalış göstermediğini söyleyebiliriz. Japon Yenisinin Amerikan Dolarına karşı en keskin düşüş gösterdiği yılın 2013, en keskin artış gösterdiği yılın ise 2016 olduğunu söyleyebiliriz.

```
c <- historical_exchange_rates(from = "CAD", to = "USD",  
start_date = "2014-01-01", end_date = "2021-01-01")  
  
ggplot(c, aes(x = date, y = one_CAD_equivalent_to_x_USD)) +  
  geom_line(color = "dodgerblue3") +  
  labs(x = "Yıl",  
       y = "CADUSD",  
       title = "2014-2021 Yılları Arasında Kanada Dolarının  
Amerikan Doları Karşısındaki Değişimi",  
       subtitle = "Çizgi Grafiği") +  
  theme_classic()
```

2014–2021 Y..llar.. Aras..nda Kanada Dolar..n..n  
Amerikan Dolar.. Kar....s..ndaki De..i..imi  
Çizgi Grafi..i



Grafikte elde edilen sonuçlara göre, 2014-2016 yılları arasında Kanada Doları Amerikan Doları karşısında büyük bir düşüş göstermiştir. 2016 yılından sonra tekrar artışa geçmiş ancak 2020 yılında yine büyük bir düşüş göstermiştir. 2020-2021 yılları arasında Kanada Dolarının Amerikan Dolarına karşı tekrar yükselişe geçtiğini söyleyebiliriz.