

Gapminder Data inceleme

ilke

2022-07-26

GAPMINDER VERİ SETİ İNCELEME

Tidyverse paketini yükleyip aktive edelim.

```
#install.packages("tidyverse")
library(tidyverse)
```

```
## — Attaching packages — tidyverse 1.3.1 —
```

```
## ✓ ggplot2 3.3.6      ✓ purrr 0.3.4
## ✓ tibble 3.1.7       ✓ dplyr 1.0.9
## ✓ tidyr 1.2.0        ✓ stringr 1.4.0
## ✓ readr 2.1.2        ✓ forcats 0.5.1
```

```
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag() masks stats::lag()
```

R'in kütüphanelerinde bulunan verileri görmek için data() fonksiyonunu kullanıyoruz.

R'in içerisinde mevcut olan veri setlerini inceleyelim. Bu veri setlerinden gapminder verisi ile çalışalım.

```
library(datasets)

library(gapminder)

data(gapminder)

View(gapminder)
```

Veri seti hakkında genel bilgilere erişelim.

```
summary(gapminder)
```

```
##      country      continent      year      lifeExp
## Afghanistan: 12 Africa :624 Min. :1952 Min. :23.60
## Albania : 12 Americas:300 1st Qu.:1966 1st Qu.:48.20
## Algeria : 12 Asia :396 Median :1980 Median :60.71
## Angola : 12 Europe :360 Mean :1980 Mean :59.47
## Argentina : 12 Oceania : 24 3rd Qu.:1993 3rd Qu.:70.85
## Australia : 12 Max. :2007 Max. :82.60
## (Other) :1632
##      pop      gdpPercap
## Min. :6.001e+04 Min. : 241.2
## 1st Qu.:2.794e+06 1st Qu.: 1202.1
## Median :7.024e+06 Median : 3531.8
## Mean :2.960e+07 Mean : 7215.3
## 3rd Qu.:1.959e+07 3rd Qu.: 9325.5
## Max. :1.319e+09 Max. :113523.1
##
```

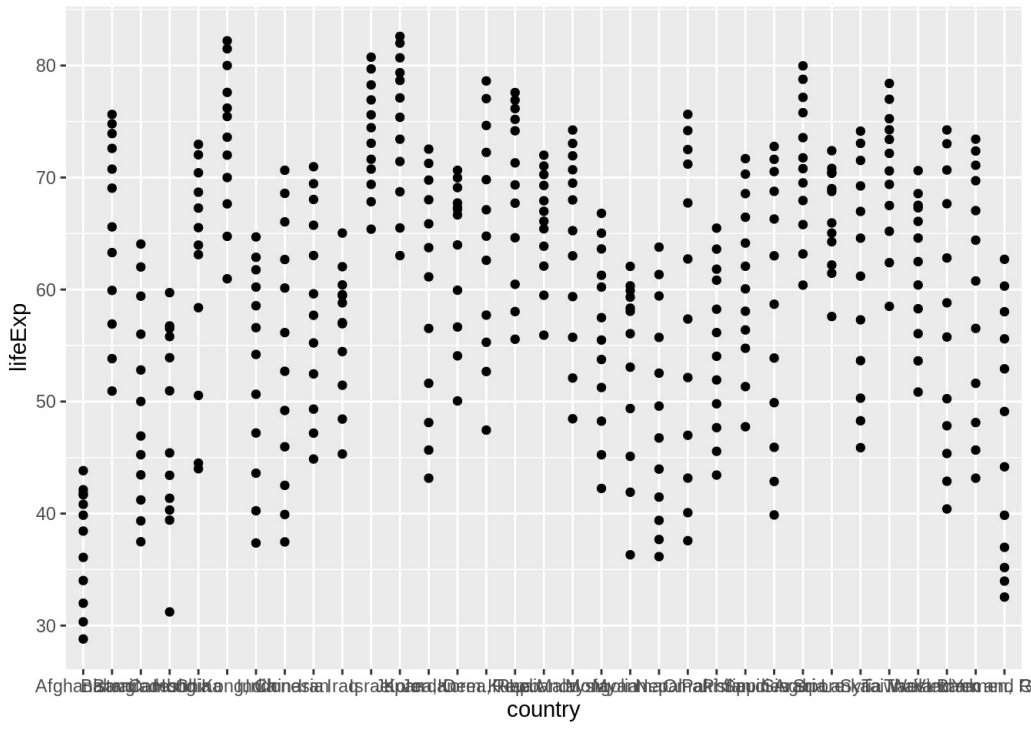
Country, continent, year,lifeExp değişkenlerinden oluşan alt veri setini çekelim. Sadece kıtası Asya olanları filtreliyoruz.

```
asya_yaş<- gapminder %>%
  select(country,continent,year,lifeExp) %>%
  filter(continent=='Asia')
```

Alt veri setinden, Country ve lifeExp arasındaki serpilme diyagramını çizdirelim.

```
asya_yaş_serpilme<-ggplot(data=asya_yaş, aes(x=country ,y=lifeExp))+
  geom_point()

asya_yaş_serpilme
```



Şimdi grafiğimizi biraz daha

detaylandıralım. Bir önceki grafikte tüm noktalar siyah olduğu için yıla göre country ve lifeExp değişkenlerinin nasıl değiştiğini gözlemleyemedik. Biraz daha bilgi katmak için year bilgilerini ayrı renklerde görselleştirmek istediğimizi düşünelim. Geliştirmek istersek; eksenlere değişken isimleri yerine başka isimler verebilir, grafiği isimlendirebiliriz.

```
asya_yaş_serpilme<- ggplot(data=asya_yaş, aes(x=country ,y=lifeExp,colour=year))+
  geom_point()+
  labs(x="Country",y="LifeExp",title="Asya Kıtası Ülke&Yaş")
```

asya_yaş_serpilme

