World Happiness Report 2023

Mia Gmiza, Gabrijela Perković, Matija Roginić, Erika Tomakić

2023-12-06

Uvod

*dodati opis naseg zadatka

Deskriptivna analiza

Učitavanje podataka.

```
opis_var = read.csv("datasets/opis_varijabli.csv")
WHR_22 = read.csv("datasets/WHR_2022.csv")
WHR_22 = head(WHR_22, -1) # preskacem zadnji red jer je "xx"
WHR_23 = read.csv("datasets/WHR_2023.csv")
```

Podatci za 2022. godinu sastoje se od 146 država i dvije varijable. Podatci za 2023. godinu sastoje se od 137 država i 15 varijabli.

```
cat("Varijable za 2022. godinu:\n")
```

```
## Varijable za 2022. godinu:
```

```
names(WHR_22)
```

```
## [1] "Country"         "Happiness.score"
```

```
cat("Varijable za 2023. godinu:\n")
```

```
## Varijable za 2023. godinu:
```

```
names(WHR_23)
```

```
##  [1] "Country.name"
##  [2] "Regional.indicator"
##  [3] "Ladder.score"
##  [4] "GDP.per.capita"
##  [5] "Social.support"
##  [6] "Healthy.life.expectancy"
##  [7] "Freedom.to.make.life.choices"
##  [8] "Generosity"
##  [9] "Perceptions.of.corruption"
## [10] "Alcohol.consumption.Both.Sexes..L.year."
## [11] "Alcohol.consumption.Male..L.year."
## [12] "Alcohol.consumption.Female..L.year."
## [13] "Crime.rate.Crime.Index"
## [14] "Healthcare.Legatum.Prosperity.Index.Health.Score"
## [15] "Gini.Coefficient...World.Bank"
```

```r
any(is.na(WHR_22))
```

```
## [1] FALSE
```

```r
cat("U podatcima za 2022. godinu nema nedostajućih vrijednosti.\n")
```

```
## U podatcima za 2022. godinu nema nedostajućih vrijednosti.
```

```r
any(is.na(WHR_23))
```

```
## [1] TRUE
```

```r
cat("U podatcima za 2023. godinu ima nedostajućih vrijednosti.\n")
```

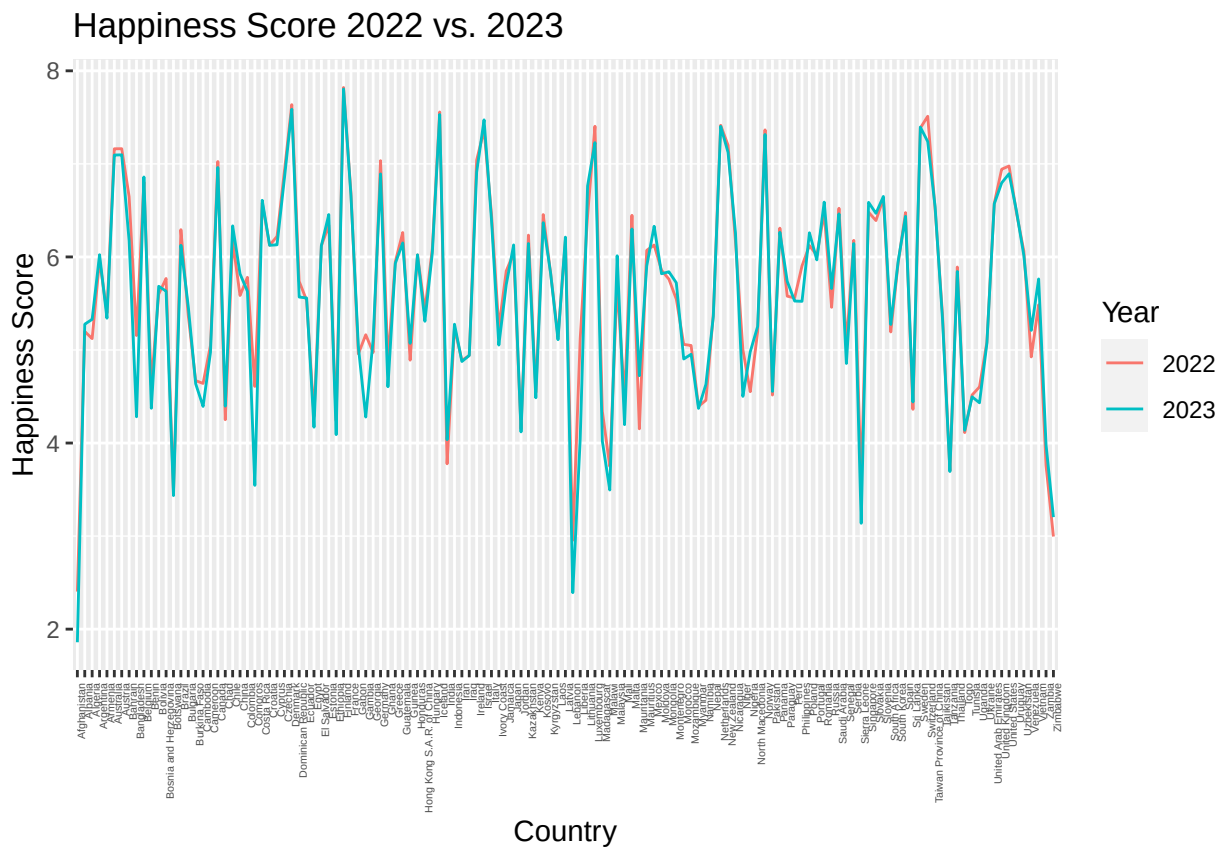
```
## U podatcima za 2023. godinu ima nedostajućih vrijednosti.
```

```r
for (col_name in names(WHR_23)) {
  if (sum(is.na(WHR_23[,col_name])) > 0){
    cat('Ukupno nedostajućih vrijednosti za varijablu ',col_name, ': ', sum(is.na(WHR_23[,col_name])),'
  }
}
```

```
## Ukupno nedostajućih vrijednosti za varijablu  Healthy.life.expectancy :  1
## Ukupno nedostajućih vrijednosti za varijablu  Alcohol.consumption.Both.Sexes..L.year. :  6
## Ukupno nedostajućih vrijednosti za varijablu  Alcohol.consumption.Male..L.year. :  6
## Ukupno nedostajućih vrijednosti za varijablu  Alcohol.consumption.Female..L.year. :  6
## Ukupno nedostajućih vrijednosti za varijablu  Crime.rate.Crime.Index :  24
## Ukupno nedostajućih vrijednosti za varijablu  Healthcare.Legatum.Prosperity.Index.Health.Score :  2
## Ukupno nedostajućih vrijednosti za varijablu  Gini.Coefficient...World.Bank :  10
```

Vizualizacija podataka

Za usporedbu razine sreće u publikaciji iz 2022. i 2023. godine možemo uzeti presjek zajedničkih država. To nas ostavlja s podatcima za 133 države.




```
library(dplyr)
```

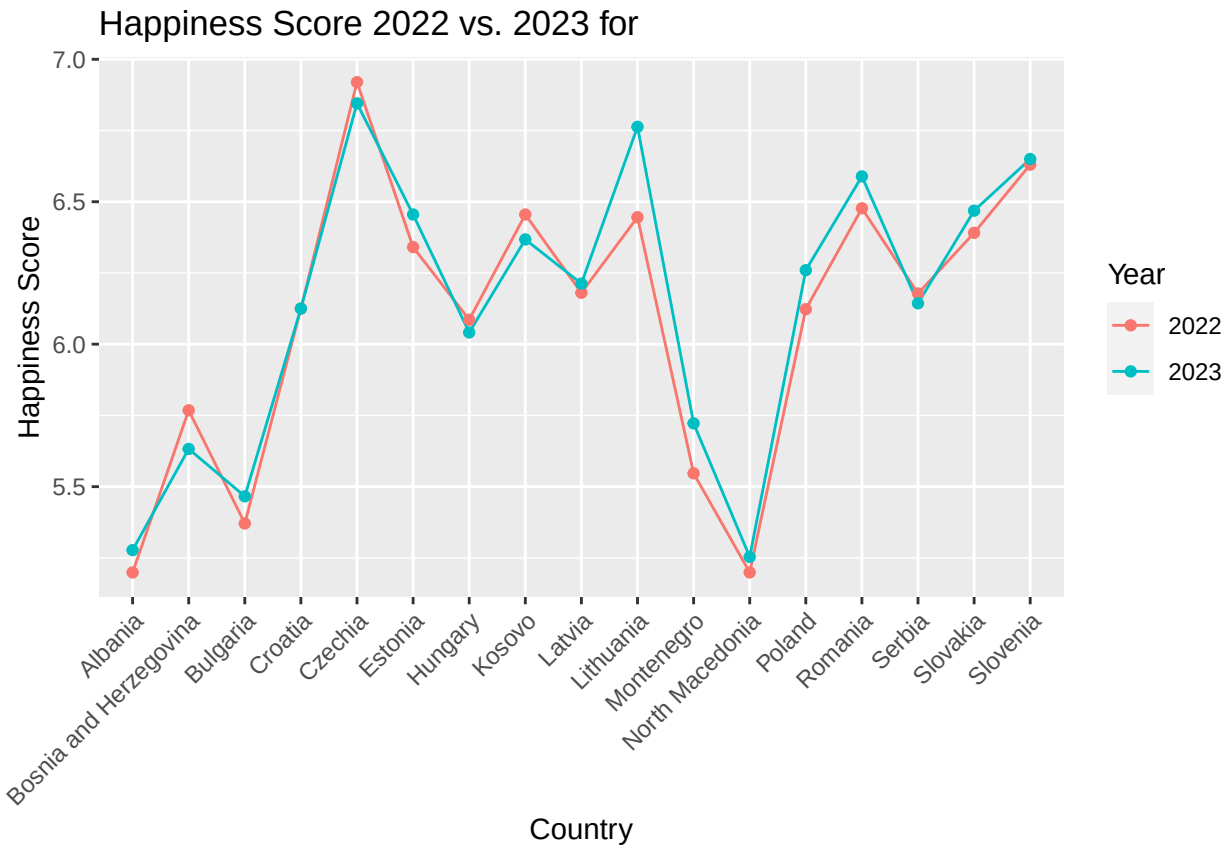
```
##
## Attaching package: 'dplyr'
```

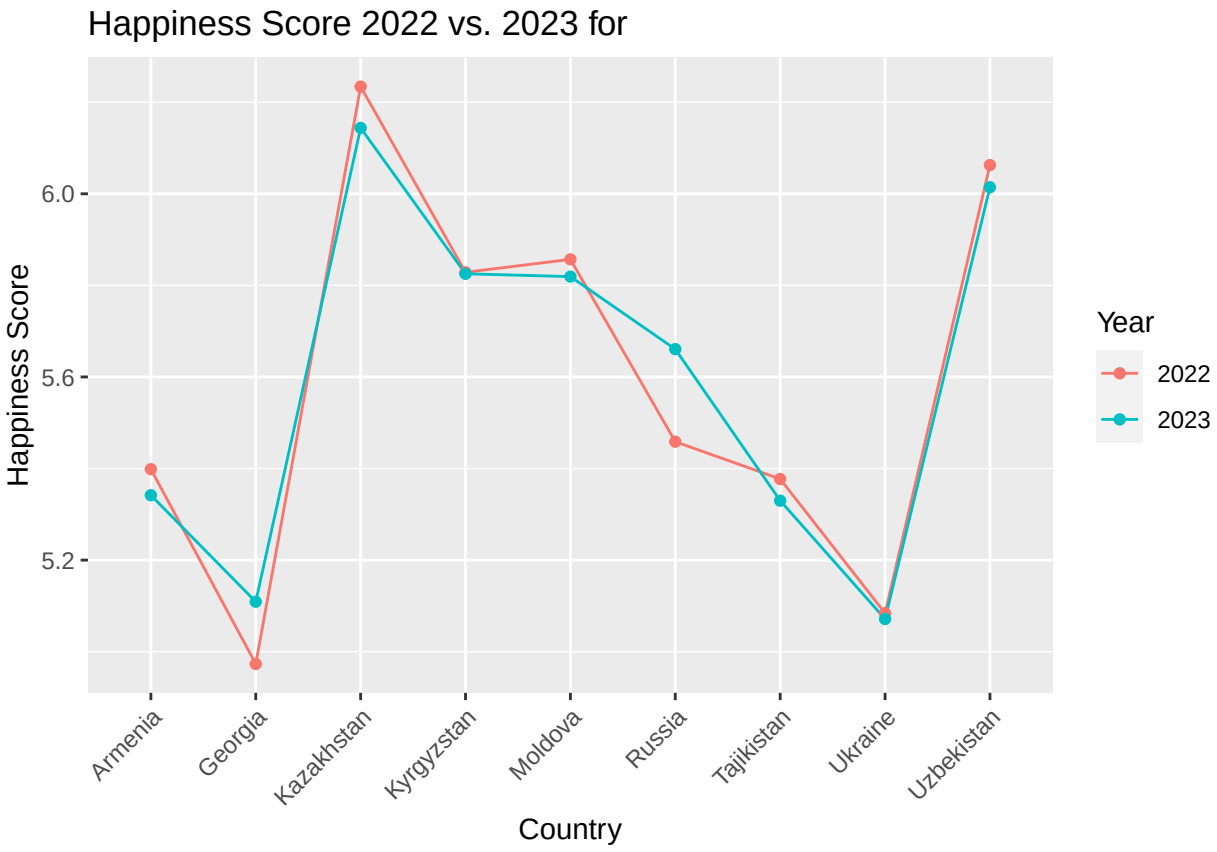
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

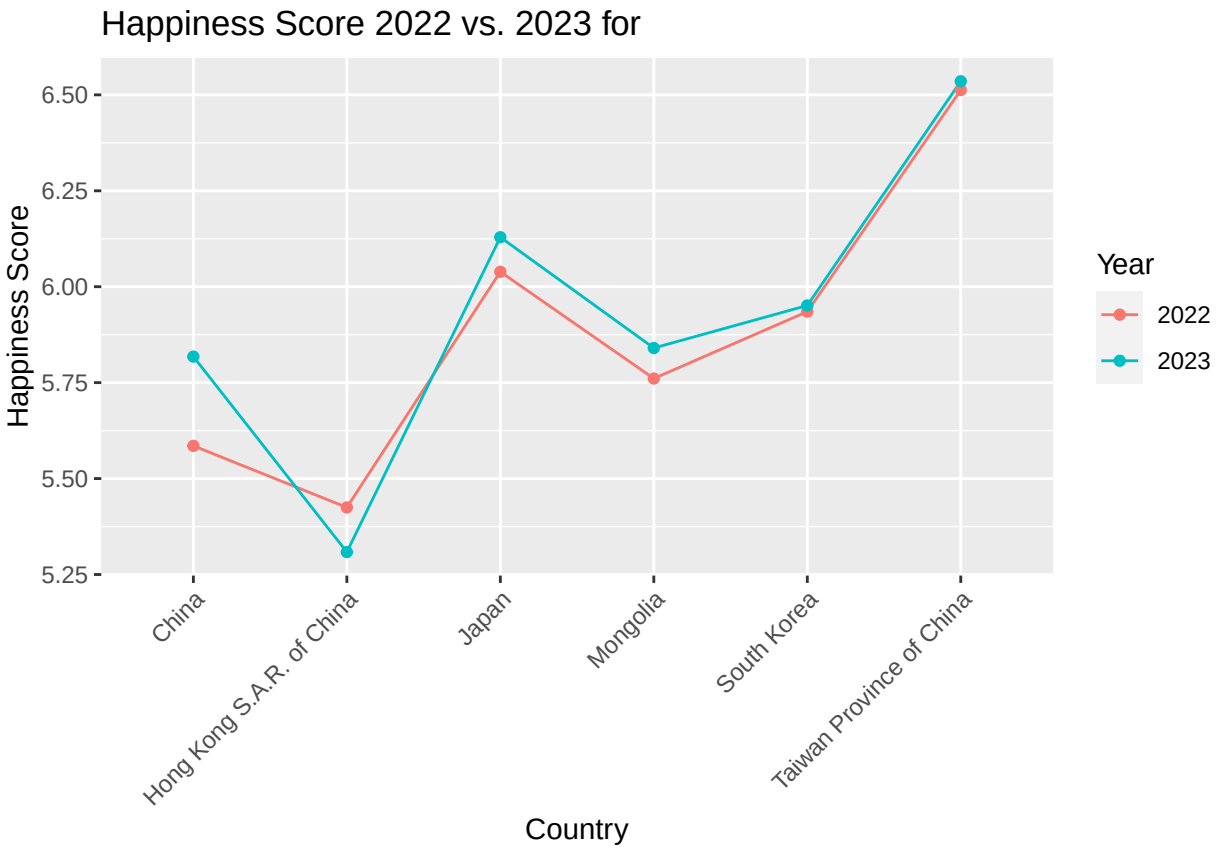
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

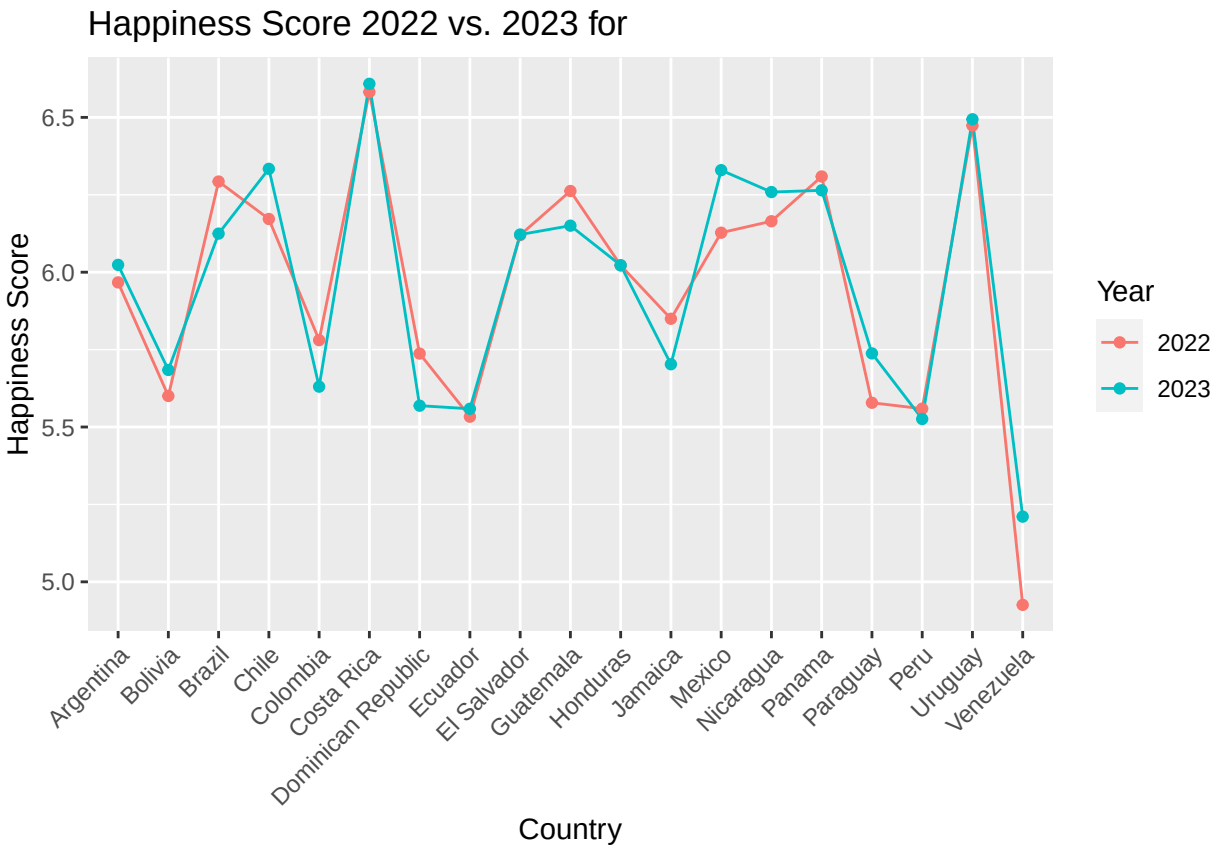
```
# grupiranje po regijama radi preglednije vizualizacije
presjek_drzava = merge(WHR_23, WHR_22, by.x = "Country.name", by.y = "Country")[c("Country.name", "Regi
colnames(presjek_drzava) = c("Country", "Region", "2022", "2023")
grouped_by_regions <- presjek_drzava %>%
  group_by(Region) %>%
  group_split()

num_of_regions = 10
for (i in 1:10) {
  region = levels(grouped_by_regions[[i]]$Region)[i]
  title = paste("Happiness Score 2022 vs. 2023 for", region)
  data = grouped_by_regions[[i]][c("Country", "2022", "2023")]
  df_long <- melt(data, id.var = "Country")

  line_plot = ggplot(df_long, aes(x = Country, y = value, color = variable)) +
```

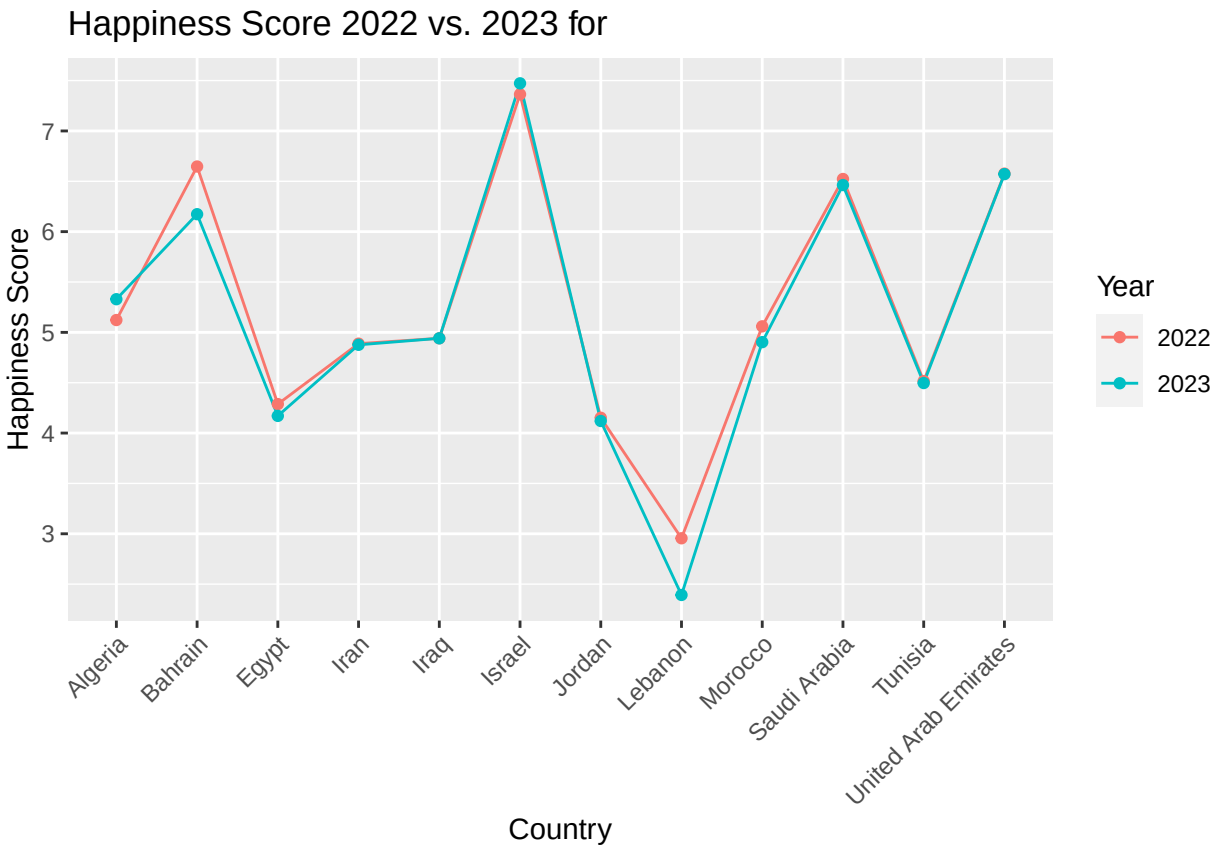
```
  geom_line(aes(group = variable)) +
  geom_point() +
  labs(title = title,
       y = "Happiness Score",
       color = "Year") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

  print(line_plot)
}
```

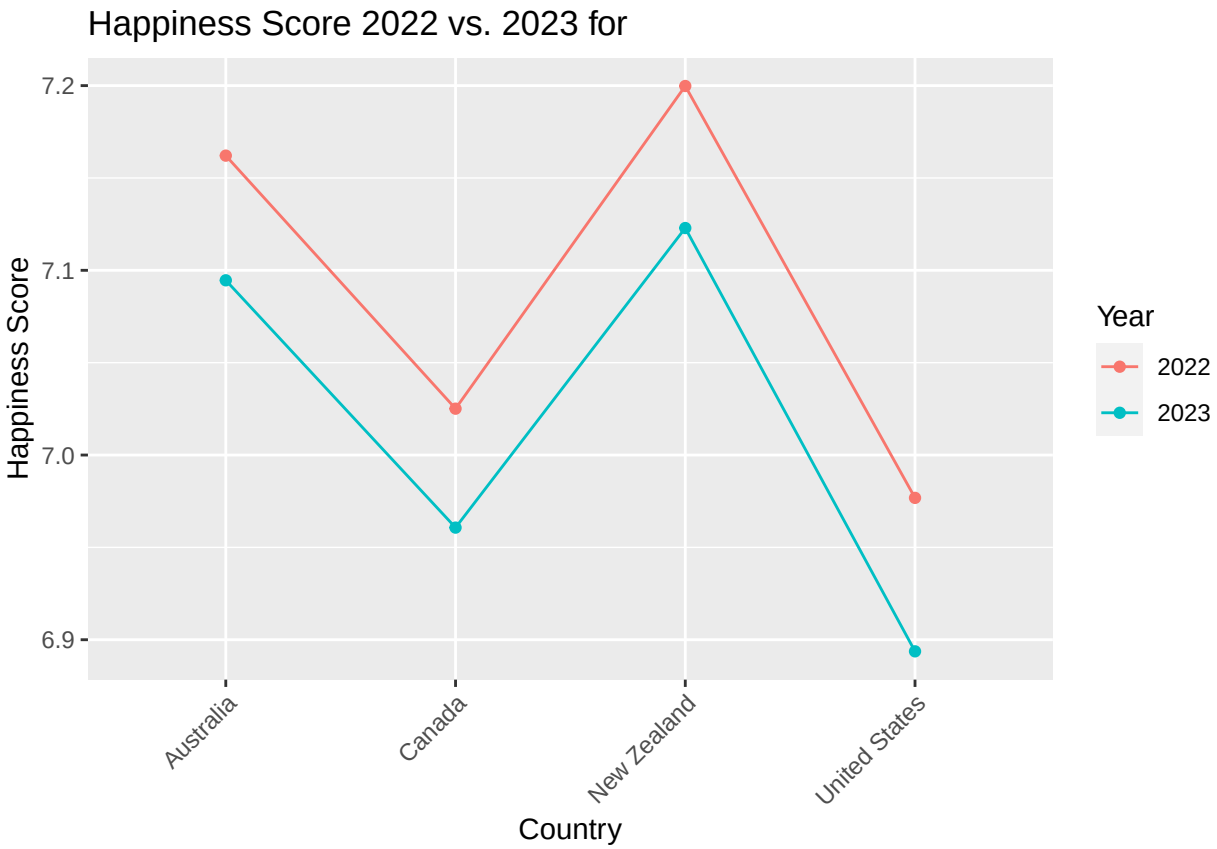


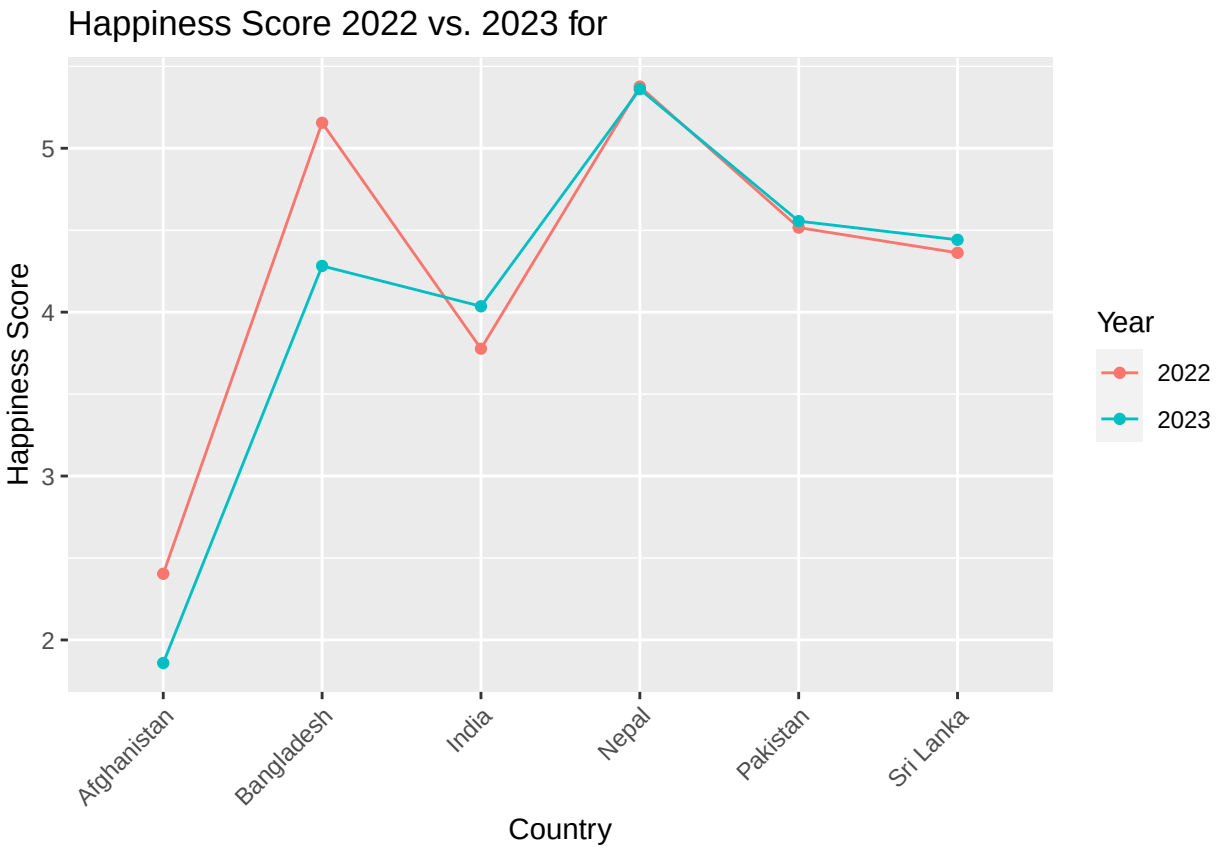

Happiness Score 2022 vs. 2023 for
Happiness Score
6.0
5.6
5.2
Armenia
Georgia
Kazakhstan
Kyrgyzstan
Moldova
Russia
Tajikistan
Ukraine
Uzbekistan
Country
Year
2022
2023

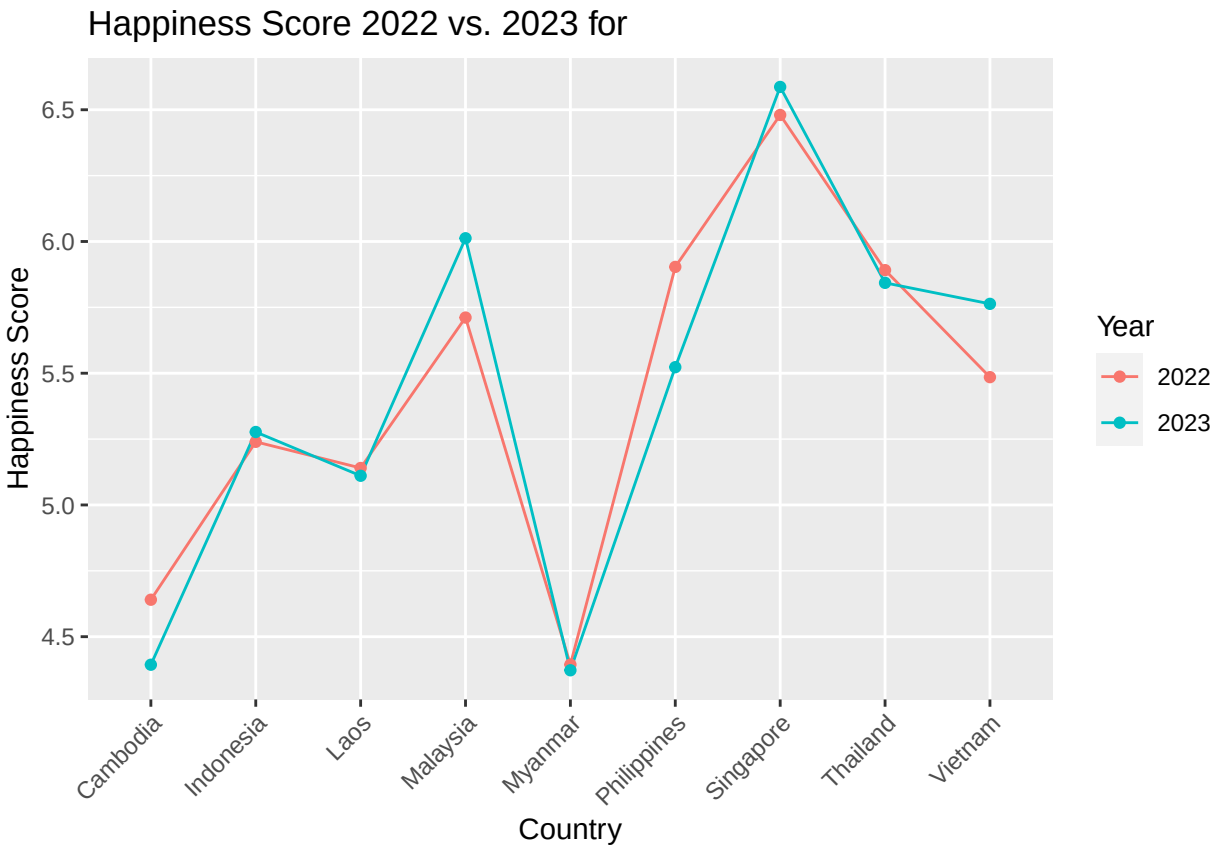

Happiness Score 2022 vs. 2023 for
Happiness Score
6.50
6.25
6.00
5.75
5.50
5.25
China
Hong Kong S.A.R. of China
Japan
Mongolia
South Korea
Taiwan Province of China
Country
Year
2022
2023

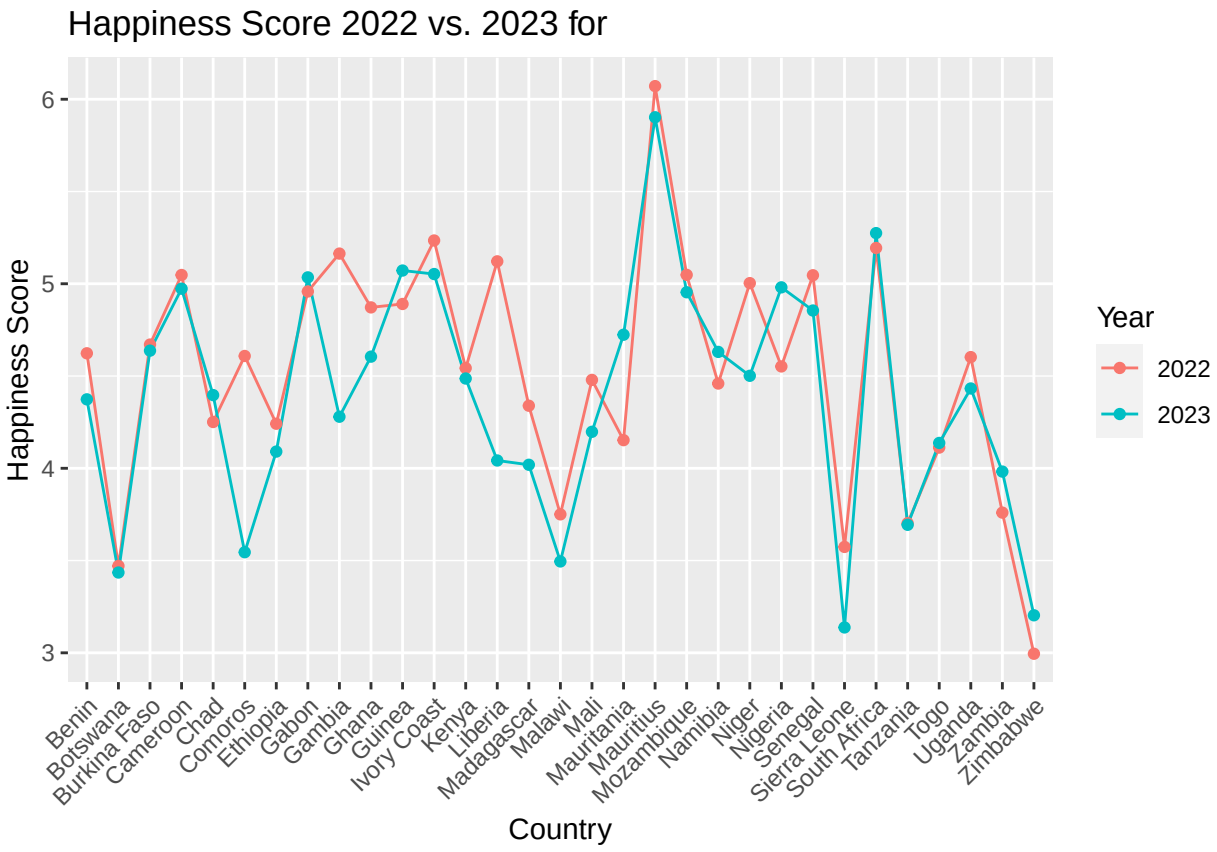

Happiness Score 2022 vs. 2023 for
Happiness Score
6.5
6.0
5.5
5.0
Argentina
Bolivia
Brazil
Chile
Colombia
Costa Rica
Dominican Republic
Ecuador
El Salvador
Guatemala
Honduras
Jamaica
Mexico
Nicaragua
Panama
Paraguay
Peru
Uruguay
Venezuela
Country
Year
2022
2023

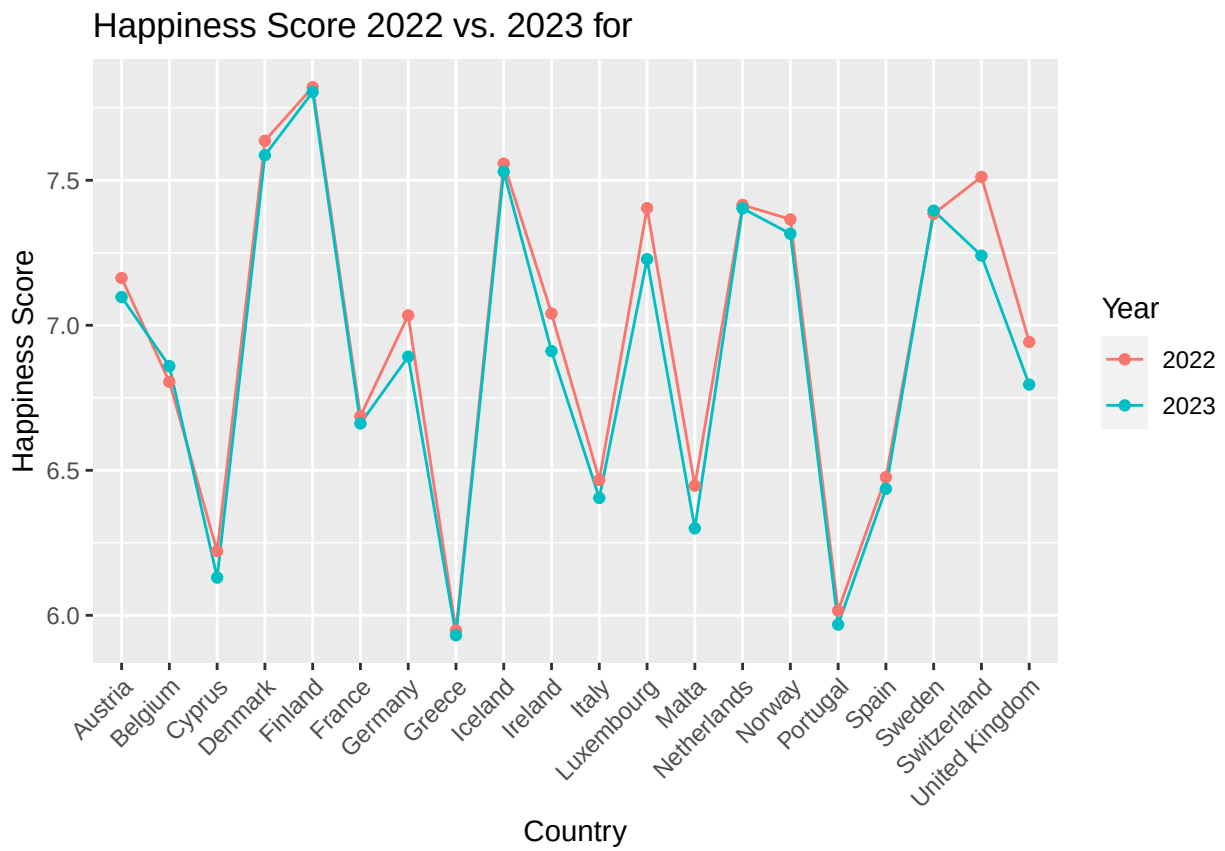

Happiness Score 2022 vs. 2023 for
Happiness Score
7
6
5
4
3
Algeria
Bahrain
Egypt
Iran
Iraq
Israel
Jordan
Lebanon
Morocco
Saudi Arabia
Tunisia
United Arab Emirates
Country
Year
2022
2023


Happiness Score 2022 vs. 2023 for
7.2
7.1
7.0
6.9
Happiness Score
Australia
Canada
New Zealand
United States
Country
Year
2022
2023


Happiness Score 2022 vs. 2023 for
Happiness Score
5
4
3
2
Afghanistan
Bangladesh
India
Nepal
Pakistan
Sri Lanka
Country
Year
2022
2023


Happiness Score 2022 vs. 2023 for
Happiness Score
6.5
6.0
5.5
5.0
4.5
Cambodia
Indonesia
Laos
Malaysia
Myanmar
Philippines
Singapore
Thailand
Vietnam
Country
Year
2022
2023


Happiness Score 2022 vs. 2023 for
Happiness Score
6
5
4
3
Benin
Botswana
Burkina Faso
Cameroon
Chad
Comoros
Ethiopia
Gabon
Gambia
Ghana
Guinea
Ivory Coast
Kenya
Liberia
Madagascar
Malawi
Mali
Mauritania
Mauritius
Mozambique
Namibia
Niger
Nigeria
Senegal
Sierra Leone
South Africa
Tanzania
Togo
Uganda
Zambia
Zimbabwe
Country
Year
2022
2023



Sljedeće što možemo napraviti je izračunati korelaciju između varijabli. Možemo izabrati i ispisati korelaciju između svakog para varijabli, ali takav ispis bi bio nepraktičan, a nije nam ni potreban. Stoga ćemo ispisati samo korelaciju svih varijabli s varijablom koja prikazuju indeks sreće u pojedinoj državi.

```
my_data <- WHR_23[, c(3,4,5,6,7,8,9,10,11,12,13,14,15)]
matrix = round(cor(my_data, use = "complete.obs"),2)
corrs <- matrix[, 1]
names <- colnames(matrix)
var = names[1]

df <- data.frame(Variable = colnames(matrix)[-1], Correlation = corrs[-1])
last12 <- tail(df,12)
cat(sprintf("%s %.2f\n", last12$Variable, last12$Correlation))
```

```
## GDP.per.capita 0.72
##  Social.support 0.80
##  Healthy.life.expectancy 0.71
##  Freedom.to.make.life.choices 0.60
##  Generosity 0.09
##  Perceptions.of.corruption -0.54
##  Alcohol.consumption.Both.Sexes..L.year. 0.54
##  Alcohol.consumption.Male..L.year. 0.51
##  Alcohol.consumption.Female..L.year. 0.60
##  Crime.rate.Crime.Index -0.38
##  Healthcare.Legatum.Prosperity.Index.Health.Score 0.74
##  Gini.Coefficient...World.Bank -0.32
```