

DATA VISUALIZATION WITH PLOTLY

PLOTLY NEDİR?

Plotly, açık kaynak JavaScript grafik kütüphanesi. plotly.js aracılığıyla etkileşimli web tabanlı grafikler oluşturmak için bir R paketidir. Plotly görselleştirme kütüphanesi D3.js, HTML ve CSS üzerine inşa edilmiş en iyi veri görselleştirme araçlarından biridir. Çevrimiçi etkileşimli veri görselleştirmeleri oluşturmayı seçebilir veya bu görselleştirmeleri tercih edilen dilde/ araçta oluşturmak için teklif edilen kitaplıkları kullanabilirsiniz. R, Python, MATLAB, Perl, Julia, Arduino gibi birçok dil/araçla uyumludur. Sürüm 2.0'dan (17 Kasım 2015) itibaren, R paketi ile oluşturulan grafikler plotly htmlwidgets çerçevesi aracılığıyla yerel olarak işlenir. Varsayılan olarak, R paketi web tarayıcınızda veya görüntüleyicide yerel olarak çalışır.

Plotly kütüphanesi 2012 yılında Montreal, Kanada'da kurulmuştur. Genel merkezi Kanada'dır. Alex Johnson, Jack Parmer Chris Parmer ve Matthew Sundquist tarafından kurulmuştur. Kurucuların geçmişi bilim, enerji-veri analizi ve görselleştirme. İşletme türü olarak da özel şirkettir.

R paketinin kullanılabilmesi için öncelikle **installed.packages("plotly")** yazarak yüklenmelidir ve bu kütüphane **library(plotly)** ile aktifleştirilmelidir.

PLOTLY AVANTAJLARI VE DEZAVANTAJLARI

Plotly'nin bazı avantajlarına ve dezavantajlarına bir göz atalım.

AVANTAJLARI:

1. D3.js bilmek zorunda kalmadan .js D3 kullanılarak oluşturulmuş etkileşimli görselleştirmeler oluşturmanıza olanak tanır.
2. R, Python, MATLAB, Perl, Julia, Arduino gibi farklı dillerin/ araçların sayısı ile uyumluluk sağlar.
3. Çizim kullanarak, etkileşimli çizimler kolayca birden fazla kişi ile çevrimiçi paylaşılabilir.
4. Plotly da veri yükleyerek ve plotly GUI kullanarak etkileşimli çizimler oluşturmak için hiçbir teknik arka plan olan kişiler tarafından kullanılabilir.
5. Plotly R ve Python ggplots ile uyumludur.
6. Bu iframes veya html kullanarak projeler veya web sitelerine etkileşimli çizimler gömmek için izin verir.
7. Plotly kullanarak etkileşimli çizimler oluşturmak için sözdizimi de çok basittir.

DEZAVANTAJLARI:

1. Plotly topluluk sürümü kullanılarak yapılan arşivler her zaman herkese açıktır ve herkes tarafından görüntülenebilir.
2. Çizim topluluk sürümü için, api aramaları günde bir üst sınırı vardır.
3. Renklendirme seçenekleri bir üst sınır olarak hareket eden topluluk sürümünde mevcut renk Paletleri sınırlı sayıda da vardır.

PLOTLY İLE GRAFİK OLUŞTURMA ADIMLARI

Veri görselleştirme, katı ve hızlı kuralları olmayan bir sanattir. Plotly kullanarak etkileşimli grafikler oluşturmak için bir dizi tipik adımlar şöyle sıralanabilir:

1. Görselleştirme oluşturmak için kullanılacak verileri elde etmek ve bunları istenen formata dönüştürmek için önceden işlemek
2. Seçtiğiniz dilde / araçta Plotly API'sini çağırmak
3. Grafiğin her ekseninde temsil edilecek veriler, en uygun çizim türü (histogram, kutu grafikleri, 3B yüzeyler gibi), grafikteki veri noktalarının veya çizimin rengi ve diğer özellikler gibi hedefler belirleyerek grafik oluşturma.
4. Çizim başlığı, eksen başlığı / etiketleri, eksen başlığı / etiket yazı tipleri, vb. gibi düzen alanlarının eklenmesi

Grafikleri oluşturmadan önce, görselleştirmeleri oluşturmak için söz dizimine bir göz atalım.

=> plot_ly(x, y, type, mode, color, size)

1. x = x eksenini için değerler
2. y = y eksenini için değerler
3. type = oluşturulması gereken çizim türünü belirtir. Örneğin 'scatter', 'box' ya da 'histogram'
4. mode = grafiğin temsil edilmesi gereken mod
5. color = veri noktalarının rengini temsil eder
6. size = grafikteki veri noktalarının veya çizgilerin boyutunu temsil eden x, y ve z ile aynı uzunluktaki değerler.

NOT : Plotly, grafikleri başka biriyle çeşitli formatlarda paylaşmanıza da olanak tanır. Bunun için bir Plotly hesabında oturum açmanız gerekiyor. Grafiklerinizi paylaşmak için şu kimlik bilgilerine ihtiyacınız olacak: kullanıcı adınız ve benzersiz API anahtarınız.

NOT : Ben görselleştirmeyi sadece grafiklerin nasıl oluşturulduğunu göstermekten ziyade veri setinin üzerinde incelenerek ve yorumlarıyla anlatılarak öğrenilmesinden yanayım. O yüzden plotly ile bütün grafiklerin tek tek nasıl oluşturulduğunu göstermiyorum. Sadece belirlediğim bir veriyi plotly ile en iyi nasıl analiz edebiliriz mantığıyla hareket ederek veri setim için ihtiyacım olan grafikleri çizdireceğim. Bu analizimde de elimden geldiğince her grafik tipini göstermeye çalıştım.

GGPLOT2 İLE PLOTLY KULLANMA

Ggplot2, piyasadaki en iyi görselleştirme kitaplıklarından biridir. Plotly ile ilgili en iyi kısmı ggplotly() ile etkileşim eklenebilmesidir, bu da grafikleri daha da geliştirecektir. Böylelikle ggplot2 ile oluşturduğumuz grafikleri ggplotly() ile çok rahat interaktif grafiklere dönüştürebilirsiniz.

PLOTLY İLE VERİ GÖRSELLEŞTİRİLMESİ

KULLANILAN PAKETLER

```
library(plotly)
library(ggplot2)
library(GGally)
library(dplyr)
library(kableExtra)
library(reshape2)
library(maps)
library(plot3D)
library(magrittr)
```

VERİ HAZIRLAMA

Veri Hakkında

Dünya Mutluluk Raporu küresel mutluluk durumunun önemli bir anketidir.Raporlar bugün dünyadaki mutluluğun durumunu gözden geçiriyor ve yeni mutluluk biliminin mutluluktaki kişisel ve ulusal varyasyonları nasıl açıkladığını gösteriyor. Mutluluk puanları ve sıralamaları Gallup Dünya Anketi verilerini kullanmaktadır.Puanlar, ankette sorulan temel yaşam değerlendirme sorusuna verdiği yanıtlara dayanmaktadır.Cantril merdiveni olarak bilinen bu soru, katılımcılardan 10 ve mümkün olan en kötü yaşamın 0 olmasını düşünerek kendi mevcut yaşamlarını bu ölçekte derecelendirmelerini ister. Puanlar 2013-2016 yılları için ulusal temsili örneklerden ve tahminleri temsil etmek için Gallup ağırlıklarını kullanır. Mutluluk puanını izleyen sütunlar, altı faktörün (ekonomik üretim, sosyal destek, ortalama yaşam süresi, özgürlük, yolsuzluğun olmaması ve cömertlik) her bir ülkede yaşam değerlendirmelerinin Distopya'da olduğundan daha yüksek olmasına ne ölçüde katkıda bulunduğunu tahmin etmektedir. (Altı faktörün her biri için dünyanın en düşük ulusal ortalamalarına eşit değerlere sahip varsayımsal ülkeye Distopya denmektedir.)

Ben 2016 yılına ait Dünya Mutluluk Raporu verilerini almak istiyorum.. Şimdi Mutluluk durumunu hangi değişkenlerin ne ölçüde etkilediğini, birbirleriyle olan ilişkilerini ve en mutlu olan kıtayı Plotly kullanarak inceleyelim.

Veriyi Düzenleme

Öncelikle Kaggle'dan aldığım 2016 yılına ait World Happiness Report (Dünya Mutluluk Raporu) verisini okutup yapısını gözden geçirelim.

```
Happiness <- read.csv("C:/Users/Dilan/Desktop/2016.csv")

str(Happiness)
```

```
## 'data.frame':   157 obs. of  13 variables:
## $ Country      : chr  "Denmark" "Switzerland" "Iceland" "Norway" ...
## $ Region       : chr  "Western Europe" "Western Europe" "Western Europe" ...
## $ Happiness.Rank : int   1 2 3 4 5 6 7 8 9 10 ...
## $ Happiness.Score : num  7.53 7.51 7.5 7.5 7.41 ...
## $ Lower.Confidence.Interval : num  7.46 7.43 7.33 7.42 7.35 ...
## $ Upper.Confidence.Interval : num  7.59 7.59 7.67 7.58 7.47 ...
## $ Economy..GDP.per.Capita. : num  1.44 1.53 1.43 1.58 1.41 ...
## $ Family       : num  1.16 1.15 1.18 1.13 1.13 ...
## $ Health..Life.Expectancy. : num  0.795 0.863 0.867 0.796 0.811 ...
## $ Freedom      : num  0.579 0.586 0.566 0.596 0.571 ...
## $ Trust..Government.Corruption.: num  0.445 0.412 0.15 0.358 0.41 ...
## $ Generosity   : num  0.362 0.281 0.477 0.379 0.255 ...
## $ Dystopia.Residual : num  2.74 2.69 2.83 2.66 2.83 ...
```

Happiness veri setimde 157 gözlem ve 13 değişken bulunmaktadır. Country ve Region değişkenleri karakter (Kesikli/kategorik), geriye kalan değişkenler numerik (Süreklili) tipindedir. Peki bu veri setimdeki değişkenler neleri ifade ediyor? Söyle bir göz gezdirip tanıtalım verimizi.

1. Country : Ülke
2. Region : Bölge yada kıta
3. Happiness_Rank : Mutluluk sıralaması
4. Happiness_Score : Mutluluk skoru/puanı
5. Lower.Confidence.Interval : Alt Güven Aralığı
6. Upper.Confidence.Interval : Üst Güven Aralığı
7. Economy..GDP.per.Capita. : Ekonomi, kişi başına GSYİH
8. Family : Aile
9. Health..Life.Expectancy. : Sağlık, Yaşam Beklentisi
10. Freedom : Özgürlük
11. Trust..Government.Corrupction. : Güven, Hükümet Yolsuzluk
12. Generosity : Cömertlik
13. Dystopia.Residual : Distopi Kalıntı

Eksik veriler var mı? Kontrol edelim.

```
complete.cases(Happiness)
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [31] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [46] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [61] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [76] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [91] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [106] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [121] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [136] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [151] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Eksik veri yok.

Veri setimin daha anlaşılır olması için değişken isimlerini tekrar adlandırmak istiyorum.

```
colnames(Happiness) <- c("Country", "Region", "Happiness_Rank", "Happiness_Score", "lower", "Upper", "Economy", "Family",  
                          "Life_Expectancy", "Freedom", "Trust", "Generosity", "Dystopia_Residual")
```

Ayrıca Region değişkenini silip bunun yerine bölgeleri altı farklı kıtada Continent adında yeni bir değişkene atamak istiyorum.

Lower.Confidence.Interval ve Upper.Confidence.Interval değişkenlerinde bu veri setinde çok kullanmayacağım. O yüzden bu iki değişkeni de veri setimden çıkarıyorum.

```
Happiness <- Happiness[, -c(2,5,6)]  
  
Happiness$Continent <- NA  
  
Happiness$Continent[which(Happiness$Country %in% c("Israel","United Arab Emirates","Singapore","Thailand",  
                                                    "Taiwan Province of China","Qatar","Saudi Arabia","Kuwait",  
                                                    "Bahrain","Malaysia","Uzbekistan", "Japan","South Korea",  
                                                    "Turkmenistan", "Kazakhstan", "Turkey","Hong Kong S.A.R., China",  
                                                    "Philippines","Jordan","China","Pakistan","Indonesia","Azerbaijan",  
                                                    "Lebanon","Vietnam","Tajikistan","Bhutan","Kyrgyzstan","Nepal",  
                                                    "Mongolia","Palestinian Territories","Iran","Bangladesh","Myanmar",  
                                                    "Iraq", "Sri Lanka", "Armenia", "India", "Georgia", "Cambodia",  
                                                    "Afghanistan", "Yemen", "Syria"))] <- "Asia"  
  
Happiness$Continent[which(Happiness$Country %in% c("Norway", "Denmark", "Iceland", "Switzerland", "Finland",  
                                                    "Netherlands", "Sweden", "Austria", "Ireland", "Germany",  
                                                    "Belgium", "Luxembourg", "United Kingdom", "Czech Republic",  
                                                    "Malta", "France", "Spain", "Slovakia", "Poland", "Italy",  
                                                    "Russia", "Lithuania", "Latvia", "Moldova", "Romania",  
                                                    "Slovenia", "North Cyprus", "Cyprus", "Estonia", "Belarus",  
                                                    "Serbia", "Hungary", "Croatia", "Kosovo", "Montenegro",  
                                                    "Greece", "Portugal", "Bosnia and Herzegovina", "Macedonia",  
                                                    "Bulgaria", "Albania", "Ukraine"))] <- "Europe"  
  
Happiness$Continent[which(Happiness$Country %in% c("Canada", "Costa Rica", "United States", "Mexico",  
                                                    "Panama","Trinidad and Tobago", "El Salvador", "Belize", "Guatemala",  
                                                    "Jamaica", "Nicaragua", "Dominican Republic", "Honduras",  
                                                    "Haiti"))] <- "North America"  
  
Happiness$Continent[which(Happiness$Country %in% c("Chile", "Brazil", "Argentina", "Uruguay",  
                                                    "Colombia", "Ecuador", "Bolivia", "Peru",  
                                                    "Paraguay", "Venezuela"))] <- "South America"  
  
Happiness$Continent[which(is.na(Happiness$Continent)))] <- "Africa"  
  
Happiness <- Happiness %>% select(Country,Continent, everything())  
  
# Kıta sütununu faktör olarak değiştiriyorum  
  
Happiness$Continent <- as.factor(Happiness$Continent)
```

Şimdi düzenlediğim Happiness veri setimin yapısına bakıp görselleştirmeye başlayabiliriz.

```
str(Happiness)
```

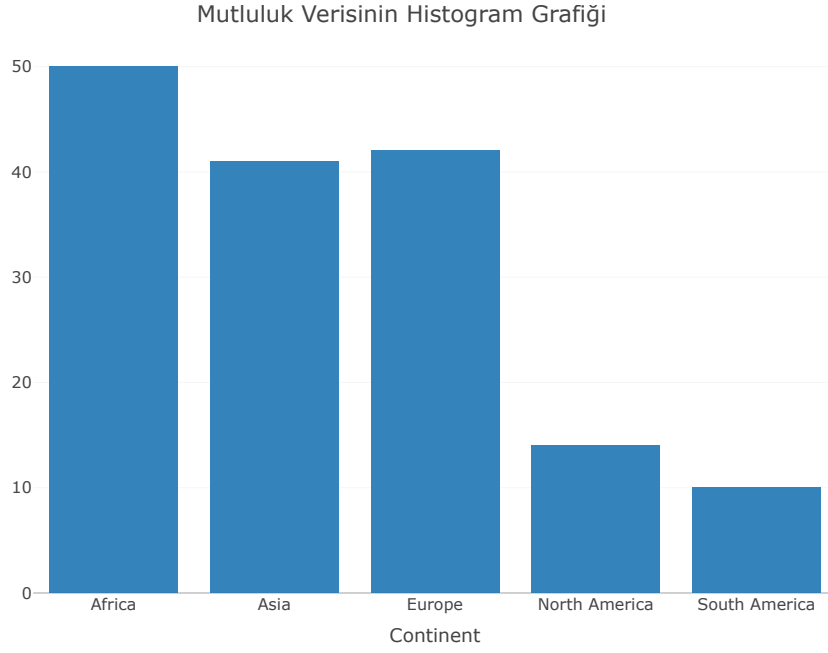
```
## 'data.frame':   157 obs. of  11 variables:  
## $ Country      : chr  "Denmark" "Switzerland" "Iceland" "Norway" ...  
## $ Continent    : Factor w/ 5 levels "Africa","Asia",...: 3 3 3 3 3 4 3 1 1 3 ...  
## $ Happiness_Rank : int  1 2 3 4 5 6 7 8 9 10 ...  
## $ Happiness_Score : num  7.53 7.51 7.5 7.5 7.41 ...  
## $ Economy       : num  1.44 1.53 1.43 1.58 1.41 ...  
## $ Family        : num  1.16 1.15 1.18 1.13 1.13 ...  
## $ Life_Expectancy : num  0.795 0.863 0.867 0.796 0.811 ...  
## $ Freedom       : num  0.579 0.586 0.566 0.596 0.571 ...  
## $ Trust         : num  0.445 0.412 0.15 0.358 0.41 ...  
## $ Generosity    : num  0.362 0.281 0.477 0.379 0.255 ...  
## $ Dystopia_Residual: num  2.74 2.69 2.83 2.66 2.83 ...
```

Yeni veri setim 157 gözlem ve 11 değişkenden oluşmaktadır.

GÖRSELLEŞTİRME

Analiz için ilk başvurulacak grafiklerden sıklık dağılımları yani histogram grafiği değişken hakkında önemli bilgiler vermektedir. Bu yüzden ilk önce histogram grafiğine bakmak istiyorum.

```
plot_ly(Happiness[,2:11], type = "histogram", mode = "markers", x = ~Continent, alpha = 0.9) %>%
  layout(title = "Mutluluk Verisinin Histogram Grafiği",
    xaxis = list(title = "Continent"),
    yaxis = list(title = ""))
```

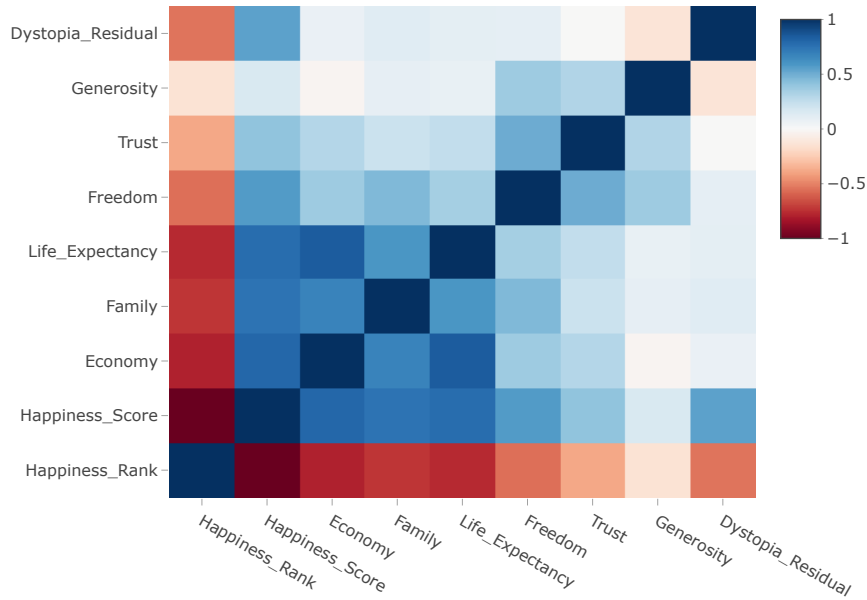


Histogram grafiğine baktığım zaman verimin sağa çarpık olduğu görülmektedir.

Veri Setindeki Değişkenlerin Birbirleriyle Olan İlişkilerinin İncelenmesi

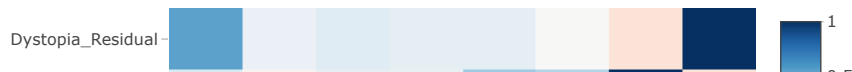
Veri setimizdeki sayısal değişkenler arasındaki ilişkiyi anlamak için korelasyon matrisini görüntüleyelim.

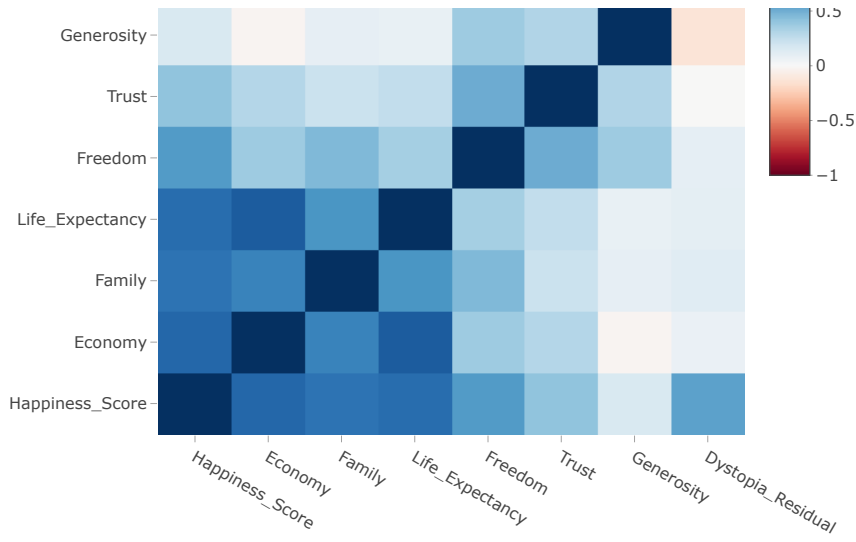
```
corr <- cor(dplyr::select_if(Happiness, is.numeric))
plot_ly(colors = "RdBu") %>%
  add_heatmap(x = rownames(corr), y = colnames(corr), z = corr) %>%
  colorbar(limits = c(-1, 1))
```



Happiness_Rank (Mutluluk sıralaması) ne kadar düşüğe Happiness_Score (Mutluluk skoru/puanı) ve diğer değişkenlerin o kadar yüksek olduğu görülmekte. Sonuç olarak Happiness_Rank ve diğer değişkenler arasında ters bir ilişki olduğunu söyleyebiliriz. O halde Happiness_Rank değişkenini çıkarıp sonuçları tekrar gözden geçirmek istiyorum.

```
corr <- cor(dplyr::select_if(Happiness, is.numeric))
newcor <- cor(Happiness[c(4:11)])
plot_ly(colors = "RdBu") %>%
  add_heatmap(x = rownames(newcor), y = colnames(newcor), z = newcor) %>%
  colorbar(limits = c(-1, 1))
```

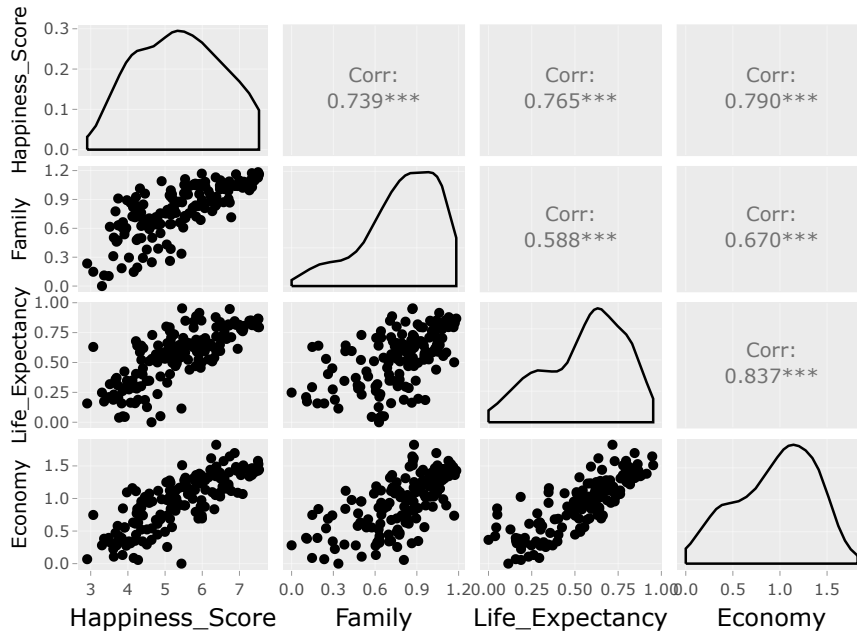




Economy, Life_Expectancy ve Family mutluluğa katkıda bulunurken Trust ve Generosity mutluluk puanının üzerinde en az etkiye sahip olduğu söylenebilir.

Şimdide Happiness veri setinde Mutluluk skoru (Happiness_Score) ile ilişkili olan değişkenlerinin birbirleriyle olan ilişkilerini ggpairs fonksiyonu ile inceleyelim.

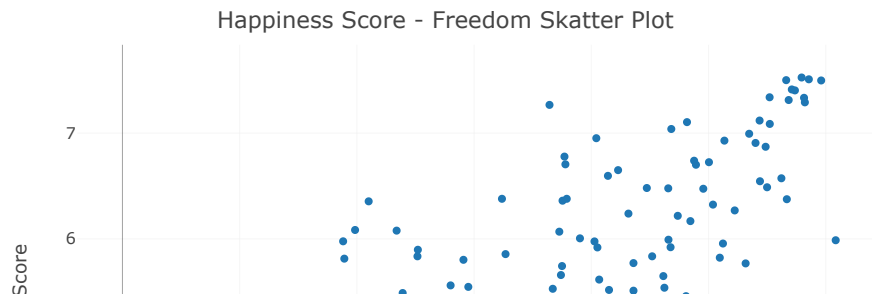
```
fig <- Happiness %>%
  select(Happiness_Score, Family, Life_Expectancy, Economy) %>%
  ggpairs(.)
ggplotly(fig)
```

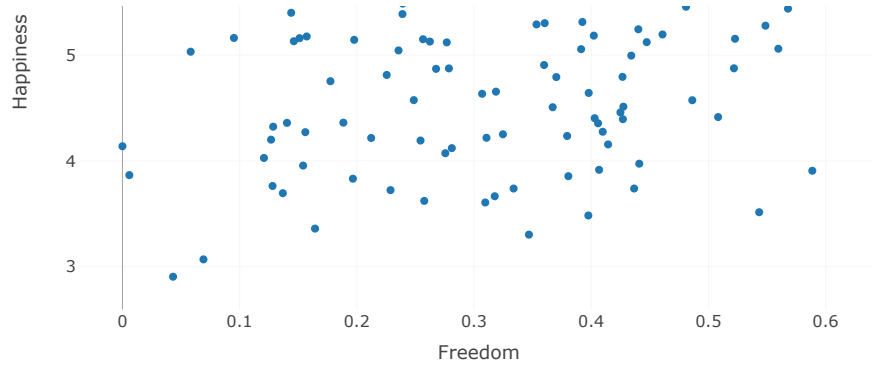


Yaşam beklentisi (Life_Expectancy) ile Ekonomi (Economy) arasında pozitif yönlü güçlü bir ilişki vardır. Aynı Zamanda bu iki değişkenin Happiness_Score değişkeniyle oldukça güçlü pozitif ilişkisi olduğunu gördük. Bundan sonraki analizlerimizi bu doğrultuda yapabiliriz.

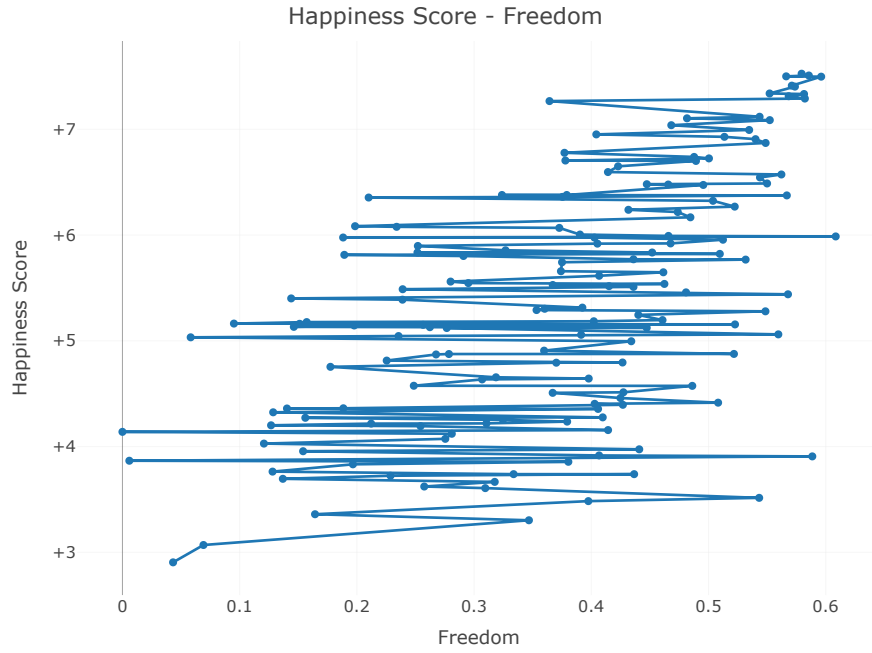
Peki ya Freedom değişkeni? Bu değişkenin Happiness_Score ile nasıl bir ilişkisi var? Bunun için Scatter plot grafiğine bir göz atalım. (Sürekli iki değişken arasındaki ilişkiyi görmek için önemli bir grafik olduğundan scatter plot grafiğini çizdirmek istiyorum.)

```
scatter_plot <- Happiness %>%
  plot_ly(x = ~Freedom, y = ~Happiness_Score, type="scatter", mode="markers") %>%
  layout(title = "Happiness Score - Freedom Scatter Plot",
    xaxis = list(title = "Freedom"), yaxis = list(title = "Happiness Score"))
scatter_plot
```



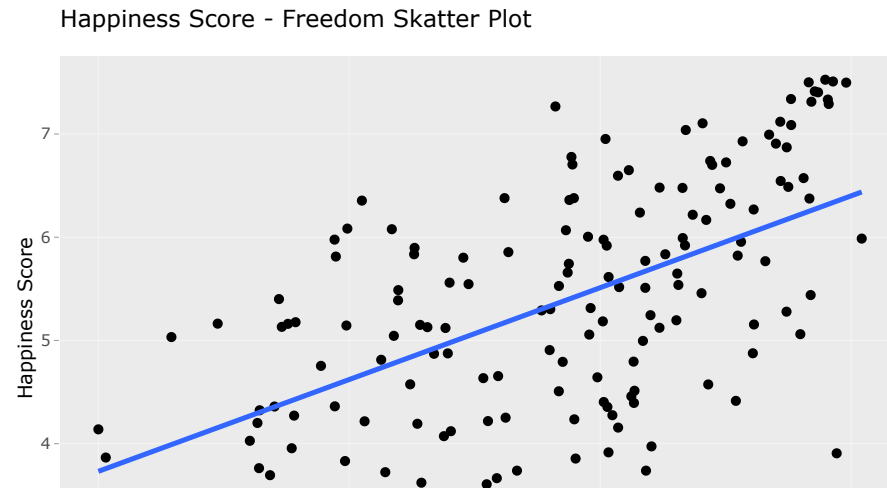


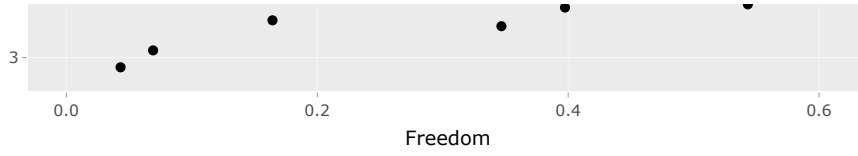
```
scatter_plot <- Happiness %>%
  plot_ly(x = ~Freedom, y = ~Happiness_Score, type="scatter", mode="markers+lines") %>%
  layout(title = "Happiness Score - Freedom ",
    xaxis = list(title = "Freedom"), yaxis = list(title = "Happiness Score"))
scatter_plot <- scatter_plot %>%
  layout(
    yaxis = list(
      tickformat = "+"
    )
  )
scatter_plot
```



Grafiğe bakıldığında zaman noktalarının dağınık durduğunu görmekteyiz. İlişkiyi biraz daha anlamlandırabilmek için ggplot ile scatter plot grafiğini biraz daha güzelleştirelim.

```
# GGLOT kullanarak scatter plot çizdirme
scatter <- ggplot(Happiness, aes(Freedom, Happiness_Score)) +
  geom_point() + geom_smooth(method=lm, se = FALSE) +
  xlab("Freedom") + ylab("Happiness Score") + ggtitle("Happiness Score - Freedom Skatter Plot")
ggplotly(scatter)
```





Grafikte Happiness Score ve Freedom arasında zayıf pozitif bir ilişkinin olduğunu söyleyebiliriz. O halde bu değişkeni çok kullanmayacağız.

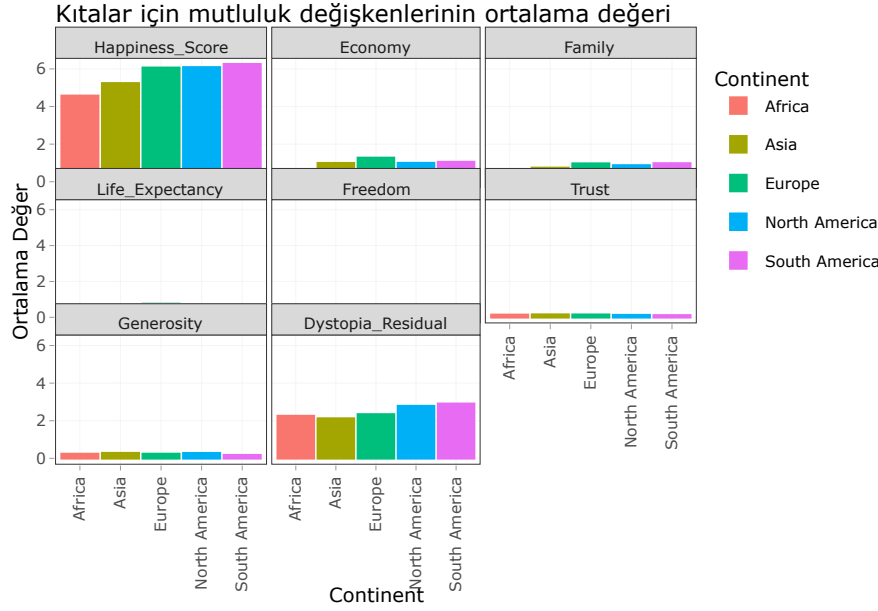
Happiness (Mutluluk) Değişkenlerinin Kıtalar Göre Analizi

Değişkenlerin birbiriyle olan ilişkilerini inceledik ; şimdide kıtalara göre nasıl değişkenlik gösterdiklerine ggplotly ile bakalım.

```
Happiness.Continent <- Happiness %>%
  select(-3) %>%
  group_by(Continent) %>%
  summarise_at(vars(-Country), funs(mean(., na.rm=TRUE)))

Happiness.Continent.melt <- melt(Happiness.Continent)

bar <- ggplot(Happiness.Continent.melt, aes(y=value, x=Continent, color=Continent, fill=Continent)) +
  geom_bar( stat="identity") +
  facet_wrap(~variable) + theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Kıtalar için mutluluk değişkenlerinin ortalama değeri",
        y = "Ortalama Değer")
ggplotly(bar)
```



Oceania'nın distopya kalıntıları dışında tüm alanlarda yaklaşık olarak en yüksek ortalamaya sahip olduğunu görebiliriz. Bundan sonra Avrupa, Kuzey Amerika ve Güney Amerika, mutluluk puanı ve diğer yedi faktör açısından kabaca aynıdır. Son olarak, Asya ve Afrika tüm değişkenler için en düşük puana sahiptir.

Öncelikle kıtalar ve ülkeler hakkında bilgi edinmek için ilk 10 mutlu ülkeye bakalım.

```
kable(head(Happiness, 10)) %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
```

Country	Continent	Happiness_Rank	Happiness_Score	Economy	Family	Life_Expectancy	Freedom	Trust	Generosity	Dysto
Denmark	Europe	1	7.526	1.44178	1.16374	0.79504	0.57941	0.44453	0.36171	
Switzerland	Europe	2	7.509	1.52733	1.14524	0.86303	0.58557	0.41203	0.28083	
Iceland	Europe	3	7.501	1.42666	1.18326	0.86733	0.56624	0.14975	0.47678	
Norway	Europe	4	7.498	1.57744	1.12690	0.79579	0.59609	0.35776	0.37895	
Finland	Europe	5	7.413	1.40598	1.13464	0.81091	0.57104	0.41004	0.25492	
Canada	North America	6	7.404	1.44015	1.09610	0.82760	0.57370	0.31329	0.44834	
Netherlands	Europe	7	7.339	1.46468	1.02912	0.81231	0.55211	0.29927	0.47416	
New Zealand	Africa	8	7.334	1.36066	1.17278	0.83096	0.58147	0.41904	0.49401	

Country	Continent	Happiness_Rank	Happiness_Score	Economy	Family	Life_Expectancy	Freedom	Trust	Generosity	Dysto
Australia	Africa	9	7.313	1.44443	1.10476	0.85120	0.56837	0.32331	0.47407	
Sweden	Europe	10	7.291	1.45181	1.08764	0.83121	0.58218	0.40867	0.38254	

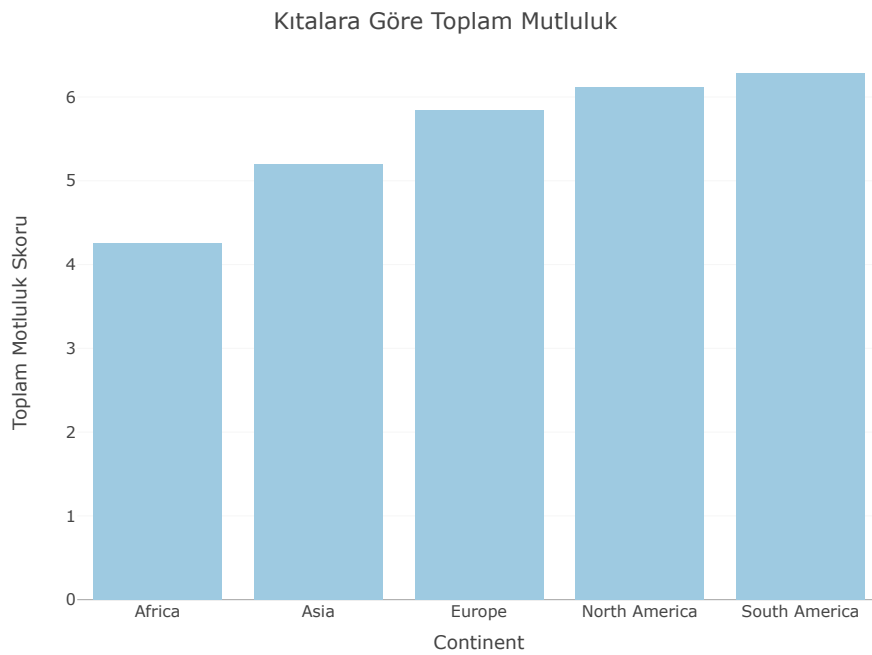
Genel olarak en mutlu kıtanın avrupa olduğu söylenebilir.Peki gerçekten en mutlu kıta avrupa mı? Görselleştirmelerimize devam edelim. Bunun için Happiness_Score değişkenim benim için oldukça önemli, kıtalara göre nasıl değiştiğine bakalım.

Kıtalara Göre Happiness_Score Karşılaştırılması

=> Kıtalara göre Happiness_Score karşılaştırılması için Bar Plot

```
Happiness_continent <- Happiness %>%
  group_by(Continent) %>%
  summarise(Ort.Mutluluk = mean(Happiness_Score, round(1)))

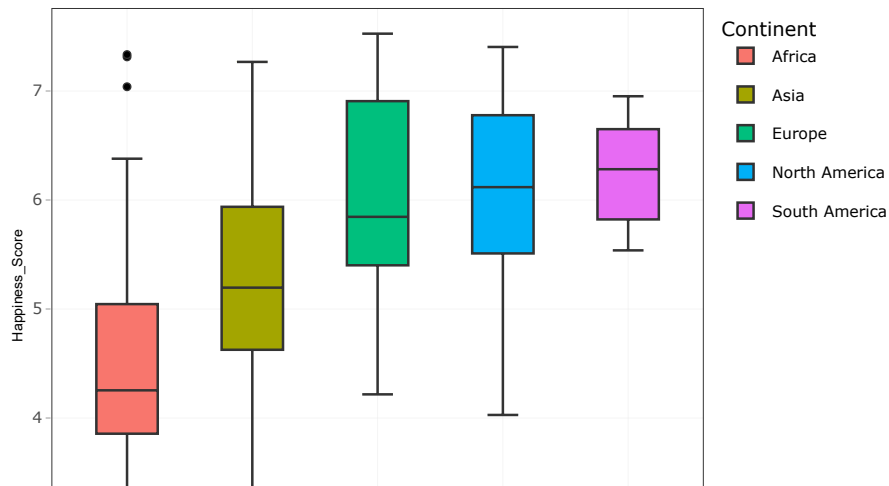
Total.Mutluluk <- plot_ly(Happiness_continent, x = ~Continent,
  y = ~Ort.Mutluluk,
  type = 'bar',
  name = 'Ortalama Mutluluk',
  marker = list(color = 'rgb(158,202,225)')) %>%
  layout(title="Kıtalara Göre Toplam Mutluluk", yaxis = list(title = "Toplam Mutluluk Skoru"))
Total.Mutluluk
```

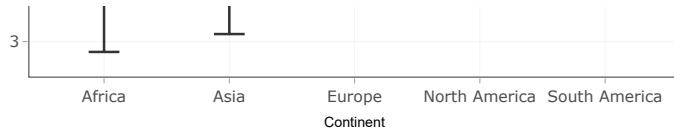


Toplam mutluluk skoruna baktığım zaman Oceania'nın diğer kıtalara göre puanının daha yüksek olduğu görülmektedir. Ayrıca en düşük puanın Afrika kıtası olduğu söylenebilir.

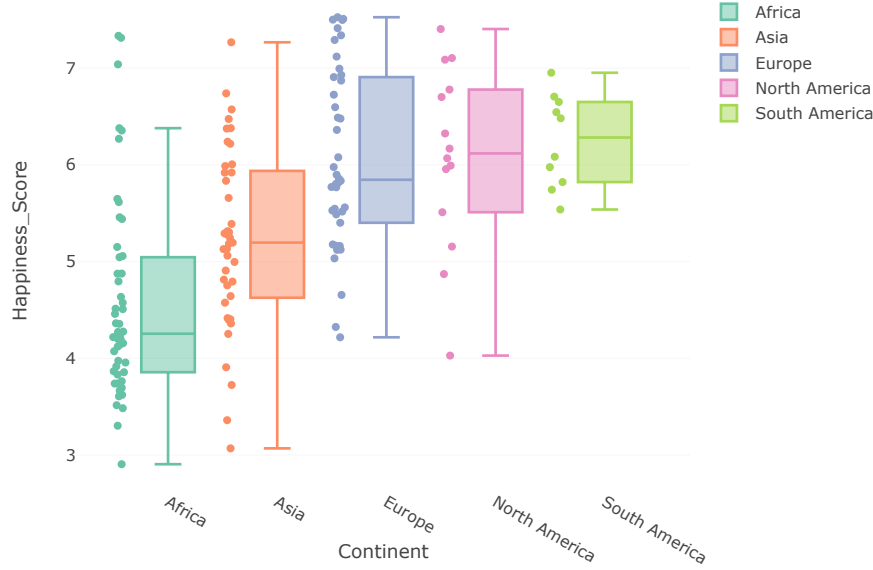
=> Kıtalara göre Happiness_Score karşılaştırılması için Box Plot

```
Box_plot <- ggplot(Happiness , aes(x = Continent, y = Happiness_Score)) +
  geom_boxplot(aes(fill=Continent)) + theme_bw() +
  theme(axis.title = element_text(family = "Helvetica", size = (8)))
ggplotly(Box_plot)
```





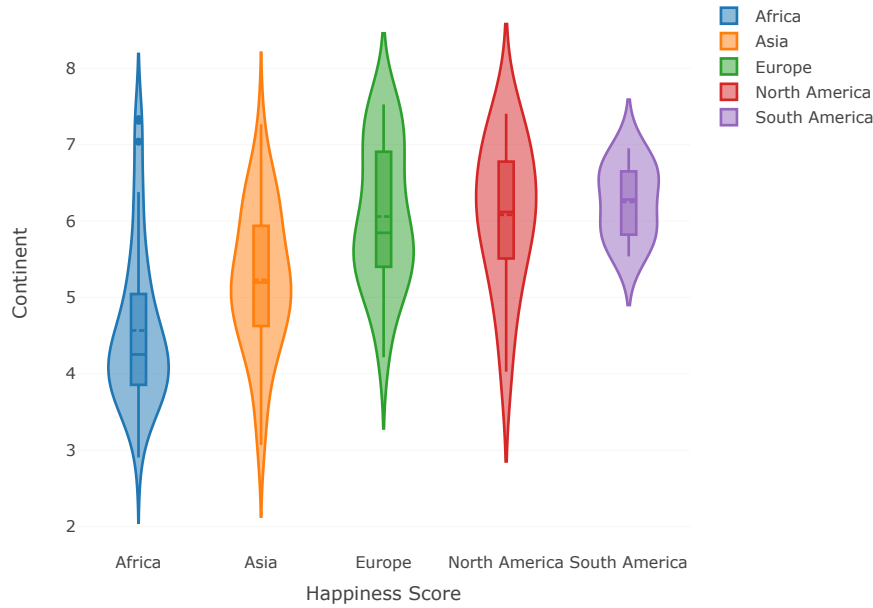
```
plot_ly(Happiness, x=~Continent, y = ~Happiness_Score, type = "box", boxpoints = "all",
        color=~Continent, jitter = 0.3, pointpos = -1.8)
```



Grafikten anlaşıldığı gibi Oceania'nın iki değişkeni olduğu için en yüksek mutluluk skoruna sahip olduğunu görülmektedir. Biraz daha netleştirmek için violin grafiğine bakalım.

=> Kıtalara göre Happiness_Score karşılaştırılması için violin grafiği

```
violin <- Happiness %>%
  plot_ly( x = ~Continent, y = ~Happiness_Score, split = ~Continent, type = 'violin', box = list(visible = T),
          meanline = list(visible = T), x0 = "")
violin <- violin %>%
  layout(xaxis = list( title = "Happiness Score"),
         yaxis = list(title = "Continent", zeroline = F))
violin
```



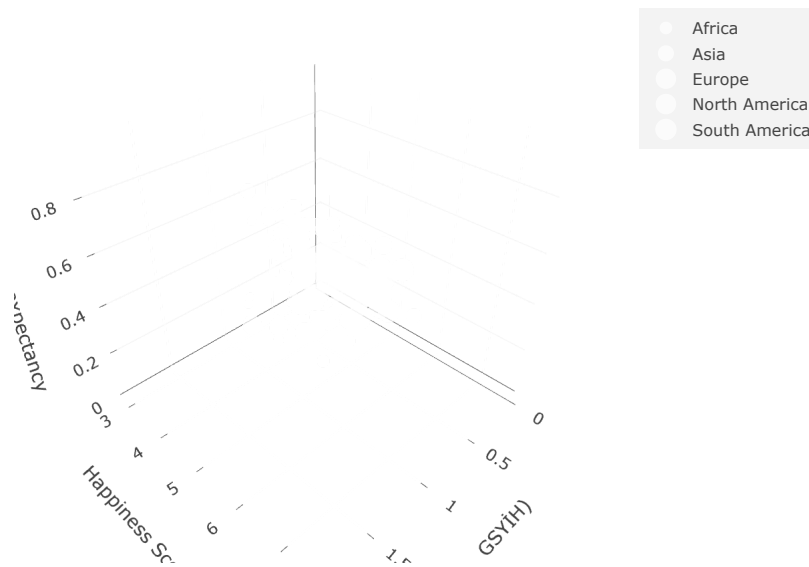
Sonuç olarak bu grafikler için Oceania'nın en yüksek medyan mutluluk puanına (Happines_Score) sahip olduğunu fakat iki değişkeni olduğu için en yüksek değeri gösterdiğini , ortalanca mutluluk puanı açısından ikinci sırada Avrupa, Güney Amerika ve Kuzey Amerikanın olduğunu, Asyanın ise Afrika'dan sonra en düşük medyana sahip olduğunu söyleyebiliriz.

Happiness_Score aynı zamanda Economy ve Life_Expectancy değişkenleriyle de güçlü pozitif bir ilişkisinin olduğunu daha önce görmüştük. Bu yüzden Happiness_Score, Life_Expectancy ve Economy değişkenlerinin birbiriyle olan ilişkilerini daha net olarak 3D grafiği kullanarak gösterelim. Ayrıca Ekonominin kıtalara göre (kişi başına GSYİH) mutluluk puanının nasıl değiştiğini animasyon grafiği şeklinde göstermek istiyorum.

```

plot_ly(Happiness, x = ~Economy, y = ~Happiness_Score, z = ~Life_Expectancy,
        color = ~Continent, size = ~Happiness_Score, colors = colors,
        marker = list(symbol = 'circle', sizemode = 'diameter'), sizes = c(2, 20),
        text = ~paste('Country:', Country,
                      '<br>Life Expectancy:', Life_Expectancy,
                      '<br>Happiness Score:', Happiness_Score,
                      '<br>GDP:.', Economy)) %>%
layout(title = '',
       scene = list(xaxis = list(title = 'GSYİH',
                                gridcolor = 'rgb(277, 277, 277)',
                                zerolinewidth = 1,
                                ticklen = 5,
                                gridwidth = 2),
                    yaxis = list(title = 'Happiness Score',
                                gridcolor = 'rgb(261, 261, 261)',
                                zerolinewidth = 1,
                                ticklen = 5,
                                gridwidth = 2),
                    zaxis = list(title = 'Life expectancy',
                                gridcolor = 'rgb(243, 243, 243)',
                                zerolinewidth = 1,
                                ticklen = 5,
                                gridwidth = 2)),
       paper_bgcolor = 'rgb(243, 243, 243)',
       plot_bgcolor = 'rgb(263, 263, 263)')

```

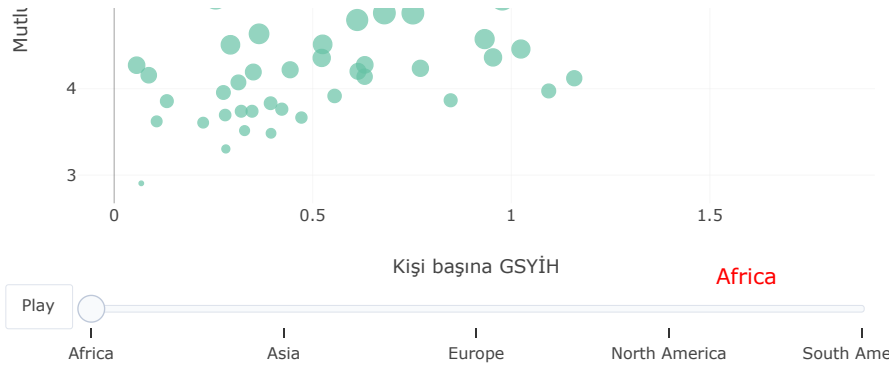


```

plot_ly(Happiness, x = ~Economy, y = ~Happiness_Score, hoverinfo = "text", text = ~Country) %>%
add_markers(frame = ~Continent, size = ~Happiness_Score, color = ~Continent,
           marker = list(sizemode = "diameter", sizeref = 3)) %>%
animation_opts(
  frame = 1000,
  transition = 300,
  easing = "elastic"
) %>%
animation_slider(
  currentvalue = list(
    prefix = NULL,
    font = list(color = "red")
  )
) %>%
layout(
  xaxis = list(title = "Kişi başına GSYİH"),
  yaxis = list(title = "Mutluluk Puanı"),
  title = "Happiness - Economy")

```





Grafikte Avrupadaki ülkelerinin mutluluk puanlarının daha yüksek olduğu söylenebilir. Aynı zamanda Happiness_Score, Economy ve Life_Expectancy değişkenleri arasındaki güçlü ilişkilerinin olduğu görülmektedir.

Happiness Score için Dünya Haritası

Son olarak Happiness Score için Dünya Haritası çizdirmek istiyorum. Böylelikle fareyi gezdirerek dünyadaki yerlerin mutluluk skorunu rahatlıkla görebilirim.

```
# Locations için code değişkenini bu veri seti ile kendi veri setimi birleştirerek elde ediyorum.

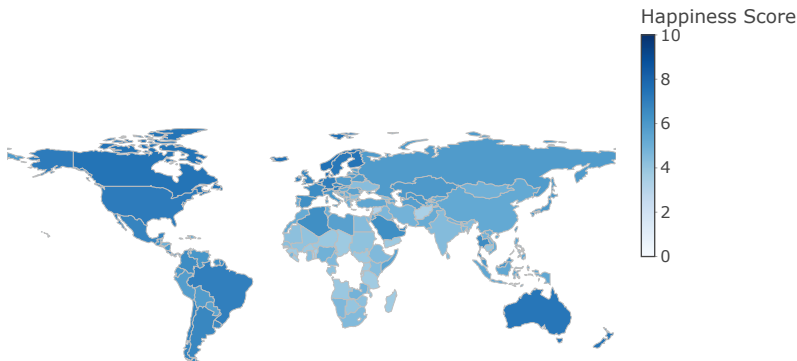
df <- read.csv('https://raw.githubusercontent.com/plotly/datasets/master/2014_world_gdp_with_codes.csv') #açılabilmesi için
interneteye bağlı olmalısınız.
names(df) <- c("Country", "gdp_billions", "code")
Happiness <- Happiness[,c("Country", "Happiness_Rank", "Happiness_Score")]
names(Happiness) <- c("Country", "Happiness_Rank", "Happiness_Score")
New_Happiness <- Happiness %>% left_join(df, by="Country")
write.csv(New_Happiness, "New_Happiness.csv")
New_Happiness <- read.csv('New_Happiness.csv')

l <- list(color = toRGB("grey"), width = 0.2)

g <- list(
  showframe = FALSE,
  showcoastlines = FALSE,
  projection = list(type = 'Mercator')
)

world_map <- plot_geo(New_Happiness) %>%
  add_trace(
    z = ~Happiness_Score, color = ~Happiness_Score, colors = 'Blues',
    text = ~Country, locations = ~code, marker = list(line = 1)
  ) %>%
  colorbar(title = 'Happiness Score', limits=c(0,10), thickness=10) %>%
  layout(
    title = 'World Happiness Score',
    geo = g)
world_map
```

World Happiness Score



ÖZET

Plotly ile Happiness veri setinin analizlerini göz önüne alarak en mutlu yerlerin Avrupa ülkeleri olduğu söylenebilir. Mutluluk durumu üzerinde Life_Expectancy, Family ve Economy değişkenlerinin olumlu etkilerinin olduğunu, Freedom değişkeninin çok etkisinin olmadığını, Trust ve Generosity değişkenlerinin ise etkisinin olmadığı söylenebilir.

KAYNAKÇA

1. Analytics Vidhya, "Interactive data visualizations using Plotly in R and Python?",Erişim: 26 Nisan 2022,
<https://www.analyticsvidhya.com/blog/2017/01/beginners-guide-to-create-beautiful-interactive-data-visualizations-using-plotly-in-r-and-python/>
2. Plotly, "Interactive web-based data visualization with R",Erişim: 26 Nisan 2022,<https://plotly-r.com/> (<https://plotly-r.com/>)
3. Plotly, "Plotly R Open Source Graphing Library",Erişim: 26 Nisan 2022,<https://plotly.com/r/> (<https://plotly.com/r/>)
4. Plotly, "R Figure Reference",Erişim: 26 Nisan 2022,<https://plotly.com/r/reference/index/> (<https://plotly.com/r/reference/index/>)
5. Kaggle, "World Happiness Report",Erişim: 26 Nisan 2022,<https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2016.csv>
(<https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2016.csv>)