

Veri Görselleştirme - Ödev 4

Sezai Ufuk Oral - 36664666542

Kullanılacak Kütüphanelerin Yüklenmesi

```
library("ggplot2")
library("dplyr")
library("priceR")
library("readr")
library("MetBrewer")
library("ggmosaic")
```

1) En Çok Satan Kitaplar

1.1 Veri Seti Hakkında

Bu veri seti 2009-2019 yılları arasında Amazon'un "En Çok Satan 50 Kitap" sıralamasına girebilmiş kitapları bulundurmaktadır. Bu veri setinde 550 adet kitap bulunmaktadır.

<https://www.kaggle.com/datasets/sootersaalu/amazon-top-50-bestselling-books-2009-2019>

```
bestsellers <- read.csv("bestsellers.csv")
```

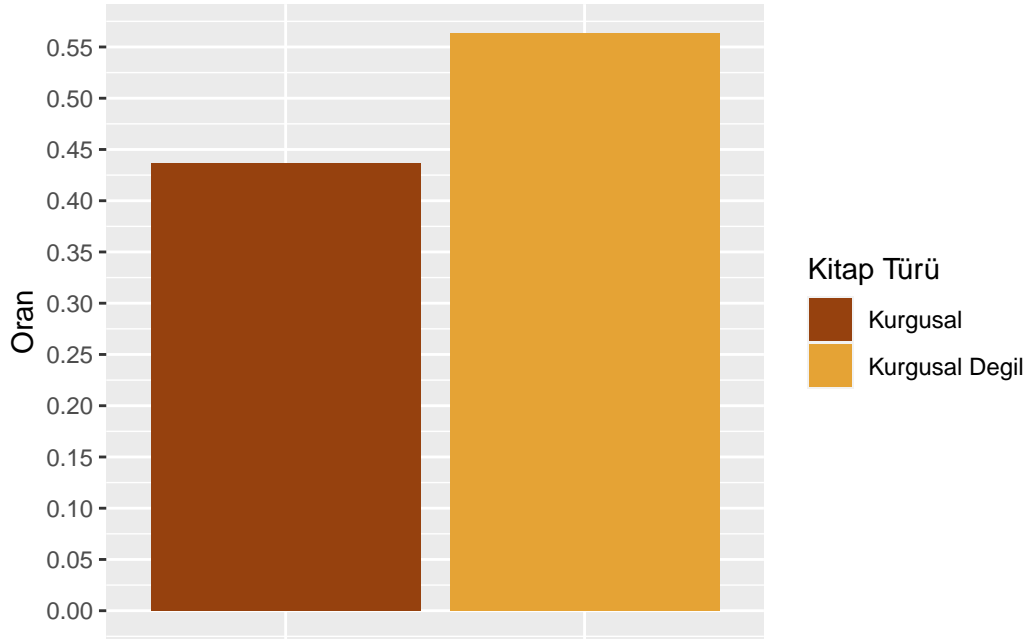
1.2 Veri Seti Hakkındaki Çıkarımlar

(1.2.1) En çok satan kitapların türlerine göre oranlarını veri görselleştirme yöntemleriyle araştırınız.

```
bestsellerGroups <- bestsellers %>% group_by(Genre) %>%
  summarise(Sum=n())
totalBooks <- as.numeric(bestsellerGroups[1,2]
  + bestsellerGroups[2,2])
bestsellerRatios <- bestsellerGroups %>%
  summarise(Ratio = Sum/totalBooks)

bestsellerGroups[3] <- bestsellerRatios
```

```
ggplot(bestsellerGroups,aes(fill=Genre,x=Genre,y=Ratio))+
  geom_col()+scale_y_continuous(breaks = seq(0,8,0.05))+
  scale_fill_manual(values=met.brewer("Degas",2),
    labels=c("Kurgusal","Kurgusal Değil"))+
  labs(x="Kitap Türü",y="Oran",fill="Kitap Türü") +
  theme(axis.text.x=element_blank(),axis.ticks.x=element_blank(),
    axis.title.x = element_blank())
```



Yukarıdaki grafikten yapabileceğimiz çıkarıma göre, kurgusal olmayan kitaplar, kurgusal oranlara göre daha fazla oran ile “Amazon En Çok Satanlar” listesine girmiştir.

- Kurgusal kitaplar: ~0.44%
- Kurgusal olmayan kitaplar: ~0.56%

(1.2.2) En çok satan kitapların türlerine ve kullanıcı reytingine (iki gruba ayırınız: 4 puan altı ve üstü) göre oranlarını veri görselleştirme yöntemleriyle araştırınız.

```
options(dplyr.summarise.inform = FALSE)

bestsellers_lower4Rating <- bestsellers %>% filter(User.Rating < 4.0)
bestsellers_higher4Rating <- bestsellers %>% filter(User.Rating >= 4.0)

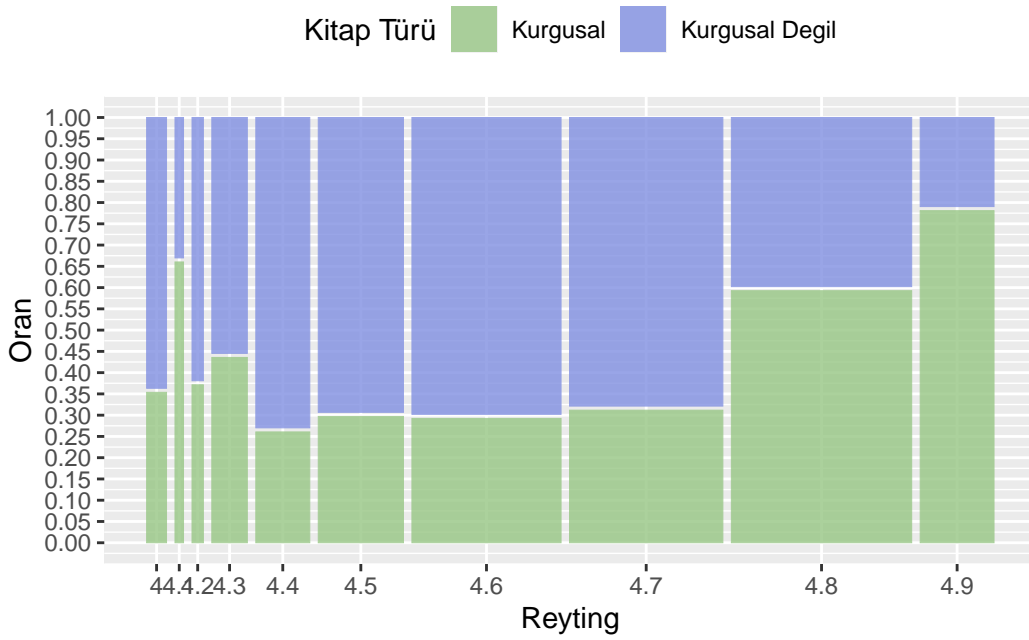
by_rating <- bestsellers_higher4Rating %>%
```

```

group_by(Genre, User.Rating) %>%
summarise(Ratio = n()/totalBooks)

ggplot(by_rating) +
  geom_mosaic(aes(x = product(User.Rating),
                        weight = Ratio,
                        fill = Genre))+
  scale_y_continuous(breaks = seq(0,1,0.05))+
  scale_fill_manual(values=met.brewer("Derain",2),
                    labels=c("Kurgusal", "Kurgusal Değil"))+
  labs(x="Reyting", y="Oran", fill="Kitap Türü") +
  theme(legend.position="top")

```



Yukarıdaki grafikten yapabileceğimiz çıkarımlara göre, 4.4 ve 4.7 aralığındaki reytinge sahip kitaplarda yüzdesel üstünlük kurgusal olmayan kitaplarda olsa da 4.7 üzeri reytinglerde kurgusal olan kitaplar daha çoğunluktadır. 4.0 ve 4.3 aralığındaki reytinglerde yüzdesel olarak kurgusal olmayan kitaplar üstün iken 4.1 reytingine sahip kitaplar arasında, farklı olarak, kurgusal kitaplar öndedir. Grafik incelendiğinde yüzdesel çoğunluğun genel olarak kurgusal olmayan kitaplarda bulunduğu söylenebilir.

2) Pokemon Çizgi Filmi

2.1 Veri Seti Hakkında

Bu veri seti Pokemon Çizgi Filmi karakterlerinin detaylı özelliklerini içermektedir.

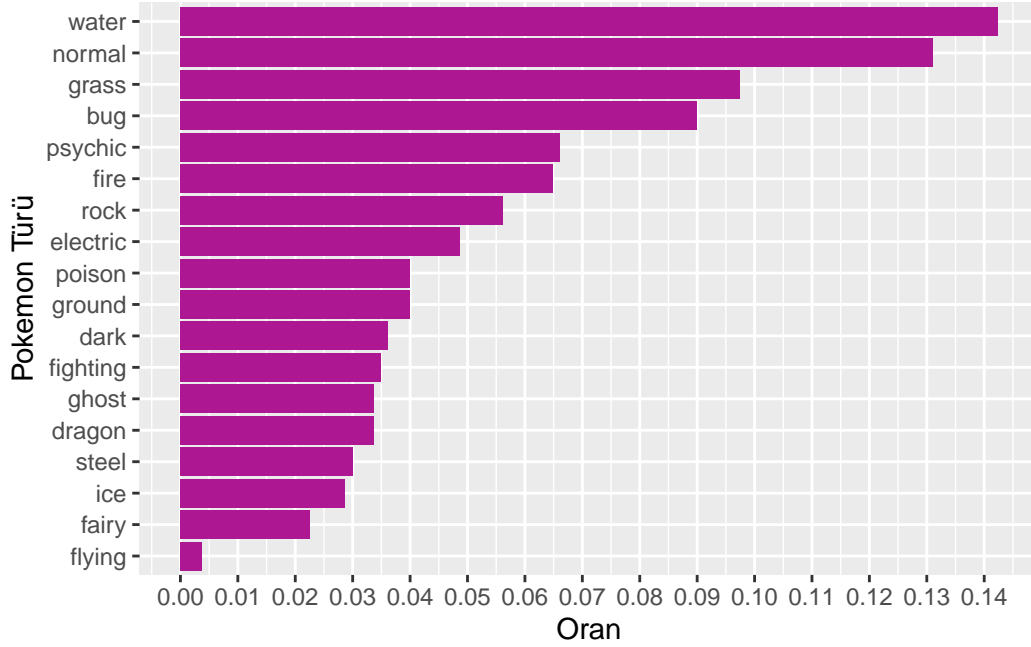
<https://www.kaggle.com/datasets/kingabzpro/pokmon-legendary-data>

```
pokedex <- read.csv("pokedex.csv")
```

2.2 Veri Seti Hakkındaki Çıkarımlar

(2.2.1) Tiplerine göre pokemonların oranlarını veri görselleştirme yöntemleriyle araştırınız.

```
pokemon_type_count <- pokedex %>% group_by(type) %>%  
  summarise(n())  
pokemon_total_count <- as.numeric(colSums(pokemon_type_count[,2]))  
colnames(pokemon_type_count)[2] <- "count"  
  
pokemon_type_count["ratio"] <- pokemon_type_count %>%  
  summarise(count/pokemon_total_count)  
  
ggplot(pokemon_type_count, aes(x=ratio, y=reorder(type, ratio)))+  
  geom_col(fill = "#ad1792") +  
  labs(x="Oran", y="Pokemon Türü")+  
  scale_x_continuous(breaks = seq(0,1,0.01))
```



Yukarıdaki grafik incelenerek pokemonlar arasında yüzdesel olarak en fazla “water” yani “su” ve “normal” tipinde pokemonlar bulunmaktadır, en az bulunan, en nadir pokemon tipi ise “flying” yani “uçan” pokemonlardır.

(2.2.2) Jenerasyon ve efsane olup olmadıklarına göre pokemonların oranlarını veri görselleştirme yöntemleriyle araştırınız.

```
library(treemapify)

pokemon_by_gen <- pokedex %>% group_by(generation,is_legendary)%>%
  summarise(ratio= n()/pokemon_total_count)

pokemon_by_gen<-pokemon_by_gen %>%
  mutate(is_legendary=
    replace(is_legendary, is_legendary==0, "Normal"))

pokemon_by_gen<-pokemon_by_gen %>%
  mutate(is_legendary=
    replace(is_legendary, is_legendary==1, "Efsane"))

ggplot(pokemon_by_gen, aes(area = ratio, fill = ratio,
  label = generation, subgroup=is_legendary)) +
  geom_treemap() + geom_treemap_text(colour = "yellow",
  place = "centre",
```

```
size = 14) +  
geom_treemap_subgroup_text(place = "centre",  
grow = FALSE,  
alpha = 1,  
colour = "white",  
fontface = "italic") + labs(fill="Oran")
```



Yukarıdaki grafik incelendiğinde “Normal” ve “Efsane” olan pokemonların dağılım oranları hakkında bilgi edinebilmekteyiz. Grafikte kaplanılan alan büyüklüğüne bakılarak pokemonların genelinin “Normal” pokemonlardan oluştuğunu, “Efsane” pokemonların ise çok daha nadir olduğunu anlarız. Kutuların renkleri yüzdesel veri hakkında bizlere bilgi vermektedir, renkler açıldıkça yüzde artmaktadır, “Normal” pokemonları en fazla 5. ve 1. nesil pokemonlar oluşturmaktadır. “Efsane” pokemonlarda ise yüzdeler birbirine çok yakındır, her nesildeki nüfus birbirine çok yakındır.

3. Türk TV Dizileri

3.1 Veri Seti Hakkında

Bu veri seti Türk Televizyon Dizileri hakkında detaylı bilgi içermektedir. (Oyuncular, Yayınlandığı Diller, Ekran Oranları vs.)

<https://www.kaggle.com/datasets/emreokcular/turkish-tv-series>

```
turkishtv <- read.csv("turkishtv.csv")
```

3.2 Veri Seti Hakkındaki Çıkarımlar

3.2.1 Dizilerin türlerine göre oranlarına veri görselleştirme yöntemleriyle araştırınız. (Her bir türü grafikte ayrı ayrı kullanınız, oranları iki veya daha fazla tür üzerinden etiketlemeyiniz.)

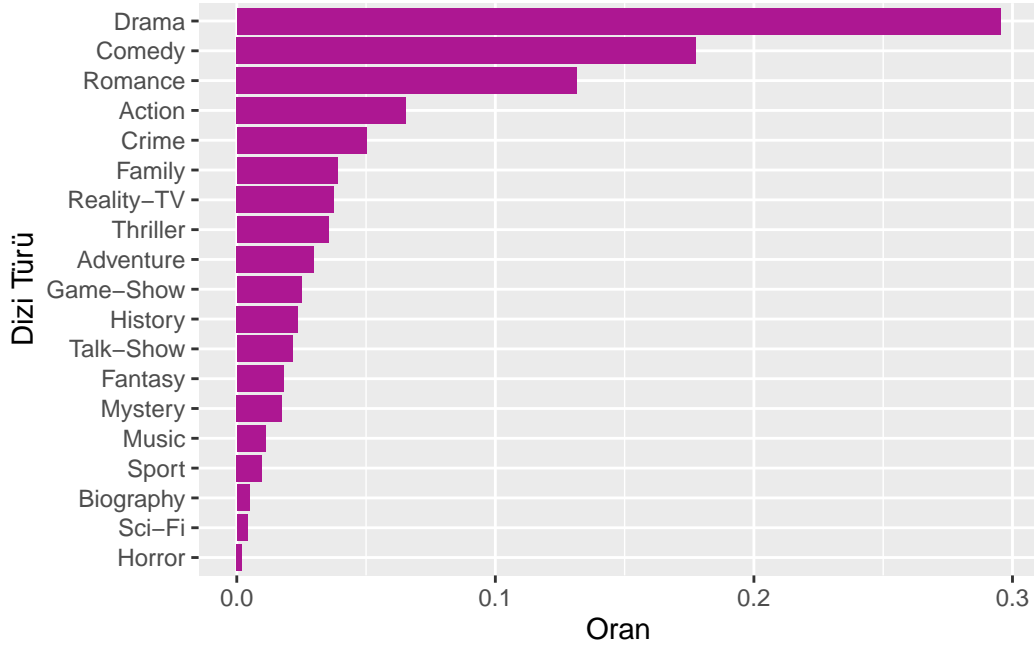
```
tv_type_count <- turkishtv %>% group_by(genres) %>% na.omit() %>%  
  summarise(count=n())  
tv_type_count<-tv_type_count[-1,]  
  
genres <- data.frame(matrix(ncol = 2, nrow = 0))  
names <- c("genre", "sum")  
colnames(genres) <- names  
  
newGenre <- genres[nrow(genres) + 1,]  
  
genreSolver <- function(toFilter) {  
  action <- tv_type_count %>% filter(grepl(toFilter, genres))  
  return(c(toFilter,as.numeric(colSums(action[,2]))))  
}  
  
genres[nrow(genres) + 1,] <- genreSolver("Action")  
genres[nrow(genres) + 1,] <- genreSolver("Adventure")  
genres[nrow(genres) + 1,] <- genreSolver("Comedy")  
genres[nrow(genres) + 1,] <- genreSolver("Crime")  
genres[nrow(genres) + 1,] <- genreSolver("Drama")  
genres[nrow(genres) + 1,] <- genreSolver("Romance")  
genres[nrow(genres) + 1,] <- genreSolver("Thriller")  
genres[nrow(genres) + 1,] <- genreSolver("Fantasy")  
genres[nrow(genres) + 1,] <- genreSolver("Horror")  
genres[nrow(genres) + 1,] <- genreSolver("Mystery")  
genres[nrow(genres) + 1,] <- genreSolver("Sci-Fi")  
genres[nrow(genres) + 1,] <- genreSolver("History")  
genres[nrow(genres) + 1,] <- genreSolver("Reality-TV")  
genres[nrow(genres) + 1,] <- genreSolver("Sport")  
genres[nrow(genres) + 1,] <- genreSolver("Game-Show")  
genres[nrow(genres) + 1,] <- genreSolver("Talk-Show")  
genres[nrow(genres) + 1,] <- genreSolver("Family")  
genres[nrow(genres) + 1,] <- genreSolver("Biography")  
genres[nrow(genres) + 1,] <- genreSolver("Music")  
  
genres[,2] <- sapply(genres[,2], as.numeric)  
genres_total_count <- as.numeric(colSums(genres[, 2, drop = FALSE]))
```

```

genres["ratio"] <- genres %>%
  summarise(ratio = sum/genres_total_count)

ggplot(genres,aes(x=ratio,y=reorder(genre,ratio)))+
  geom_col(fill = "#ad1792") +
  labs(x="Oran",y="Dizi Türü")+
  scale_x_continuous(breaks = seq(0,1,0.1))

```



Yukarıdaki grafikte görüldüğü üzere Türk Televizyon Dizileri arasında en fazla çekilen tür Drama sonrasında ise Komedi'dir. Türk Televizyonu olarak en az çekilen yapımlar ise Korku yapımlarıdır. Yapımcılar tarafından genel olarak Drama, Komedi ve Romantizm türleri seçildiği görülmektedir.