



**T.C**

**HACETTEPE ÜNİVERSİTESİ FEN FAKÜLTESİ**

**İSTATİSTİK PROGRAMI**

**İST-432 GRAFİKSEL VERİ ANALİZİ FİNAL  
ÖDEVİ**

**MELİH AKBABA-2200329033**

**DERSİN SORUMLUSU**

**Dr. Öğr. Üyesi Onur TOKA**

**24/05/2024**

## İçindekiler

YATAY ÇUBUK GRAFİĞİNİN TANITILMASI.....	3
YATAY ÇUBUK GRAFİĞİ ÖRNEKLERİ .....	3
VERİ KÜMESİNİN AÇIKLANMASI .....	5
GRAFİK KODLARININ AÇIKLANMASI.....	6
GRAFİĞİN YORUMLANMASI .....	10
KAYNAKÇA.....	11

# YATAY ÇUBUK GRAFİĞİNİN TANITILMASI

Yatay çubuk grafiği, kategorik değişkenleri numerik değerlere göre karşılaştırmak için kullanılmaktadır. Dikey eksen karşılaştırılan kategorileri temsil ederken, yatay eksen bir değeri temsil etmektedir. Yatay eksenindeki çubukların uzunluğu temsil ettikleri değerlerle orantılıdır. Temsil ettikleri değerler genellikle, kategorik verinin verideki sayısı ve oranıyla doğrudan ilişkilendirilebileceği gibi, aynı zamanda o kategorik değişken için farklı bir değişkenin ortalaması veya toplamı da olabilir.

Dikey çubuk grafiğinde ekstra olarak kategorik verilerin dağılımı da görselleştirilebilirken yatay çubuk grafiğinde daha çok kategorik veriler arasındaki farka ve değişime odaklanılmaktadır. Yatay çubuk grafiğinin bir avantajı ise etiketlerin görüntülenmesinin daha kolay olmasıdır.

## YATAY ÇUBUK GRAFİĞİ ÖRNEKLERİ

1-Şekil 1'deki grafikte PGA Tour'un düzenlemiş olduğu golf turnuvalarını kazanarak en yüksek kazancı elde eden ilk 10 golf oyuncusunu göstermektedir. Grafiğin y-ekseninde golf oyuncuları, resimleri ve isimleri ile verilmiştir. x-ekseni ise turnuvalardan kazanılan toplam kazancı içermektedir. Her bir çubukta, kazanılan toplam para dolar cinsinden gösterilmiştir.

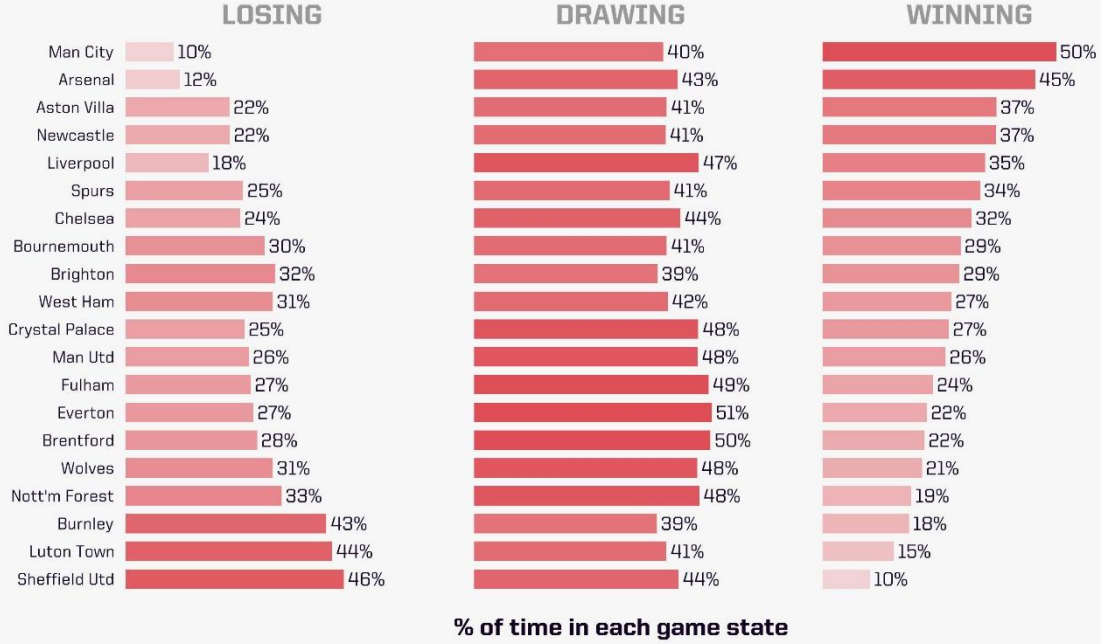


Şekil 1

## How much time has each team spent in each game state?

English Premier League 2023-24

Opta Analyst



Şekil 2

2- Şekil 2'deki grafikte İngiltere Premier Lig'inin 2023-2024 sezonunda ligde mücadele eden 20 takımın oynadıkları her bir maçın, maçın farklı durumlarında ne kadar süre geçirdiğini toplam süreye oranlayarak göstermektedir. Her bir durum, 3 farklı yatay çubuk grafik olarak sunulmuş ve çubuklar, oranın değişimine göre renklendirilmiştir. Grafiğin, y-ekseninde ligdeki takımlar, x-ekseninde ise durumların görülme sürelerinin oranları verilmiştir.

Örnek yorum olarak Manchester City ve Sheffield United'in olduğu grafik yorumlanacak olursa:

Manchester City, 2023-2024 sezonunda oynadığı maçlarda, toplam sürenin %10'unda skor olarak geride, %40'ında berabere ve %50'sinde ise önde geçirmiştir.

Sheffield United, 2023-2024 sezonunda oynadığı maçlarda, toplam sürenin %46'sında skor olarak geride, %44'ünde berabere ve %10'unda ise önde geçirmiştir. Bu performans, lig tablosunda son sırada yer almasının bir nedeni olarak gösterilebilir.

# VERİ KÜMESİNİN AÇIKLANMASI

Grafik çiziminde kullanılacak olan veri kümesi, 2020-2024 yılları arasında farklı pozisyonlarda çalışan kişilerin bilgilerini içermektedir. Değişkenler:

DEĞİŞKENLER	AÇIKLAMASI
<b>work_year</b>	2020 ile 2024 arası yılları içermektedir.
<b>experience_level</b>	["MI","SE","EN","EX"] olmak üzere 4 farklı deneyim seviyesi içermektedir.
<b>employment_type</b>	["FT","CT","PT","FL"] olmak üzere 4 farklı istihdam türü içermektedir.
<b>job_title</b>	155 tane unvan içermektedir.
<b>salary</b>	Yerel para birimi cinsinden maaşları içermektedir.
<b>salary_currency</b>	Maaşların yerel para birimlerini içermektedir
<b>salary_in_usd</b>	Maaşların dolar cinsini içermektedir.
<b>employee_residence</b>	Çalışanların ikamet ettiği ülkeleri içermektedir.
<b>remote_ratio</b>	Uzaktan çalışma oranlarını içermektedir.
<b>company_location</b>	İstihdam edilen şirketin konumunu içermektedir.
<b>company_size</b>	["S","M","L"] olarak şirketin büyüklüğünü içermektedir.

# GRAFİK KODLARININ AÇIKLANMASI

```
library(grid)
library(tidyverse)
library(shadowtext)
library(sysfonts)

data <- read.csv("C:/Users/MELİH/Desktop/GRAFİKSEL VERİ ANALİZİ DÖKÜMAN/final ödevi/salaries_2.csv")
```

İlk olarak yukarıdaki kodlarla gerekli kütüphaneler çağrıldı ve data değişkenine kullanılacak olan veri seti atandı.

```
str(data)

## 'data.frame':    16494 obs. of  11 variables:
##  $ work_year      : int  2024 2024 2024 2024 2024 2024 2024 2024 2024
## 2024 ...
##  $ experience_level : chr  "MI" "MI" "MI" "MI" ...
##  $ employment_type  : chr  "FT" "FT" "CT" "CT" ...
##  $ job_title        : chr  "Data Scientist" "Data Scientist" "Data Scientist" "Data Scientist" ...
##  $ salary           : int  120000 70000 130000 110000 240000 180000 202800 115000 80000 45000 ...
##  $ salary_currency   : chr  "USD" "USD" "USD" "USD" ...
##  $ salary_in_usd     : int  120000 70000 130000 110000 240000 180000 202800 115000 100000 56250 ...
##  $ employee_residence: chr  "AU" "AU" "US" "US" ...
##  $ remote_ratio      : int  0 0 0 0 0 0 0 0 0 ...
##  $ company_location  : chr  "AU" "AU" "US" "US" ...
##  $ company_size      : chr  "S" "S" "M" "M" ...
```

**str()** fonksiyonu ile veri setinin yapısı basit şekilde gösterildi.

```
salary_in_peryear <- as.data.frame(data[data$job_title == "Data Scientist", ]
%>%
  group_by(work_year) %>%
  summarise(avg_salary=mean(salary_in_usd)))

salary_in_peryear$work_year <- as.character(salary_in_peryear$work_year)
```

Yukarıdaki kodla, unvanı “Data Scientist” olan kişilerin yıllara göre ortalama maaşları bulunarak DataFrame yapısında kaydedildi. Daha sonra çalışma yılları karakter olarak tanımlandı.

```
font_add_google("Noto Serif")
```

**font\_add\_google()** fonksiyonuyla beraber grafikte kullanılacak font olan “Noto Serif” yazı ailesi google’dan çekildi.

```
plt <- ggplot(salary_in_peryear) +  
  geom_col(aes(x=avg_salary,y = work_year), fill = "#FFC464", width = 0.6)
```

salary\_in\_peryear dataframe’ini kullanarak , x-ekseni ortalama maaş ve y-ekseni çalışma yılları olan yatay çubuk grafiği çizdirildi. Çubuklar #FFC464 kodlu turuncu rengine boyandı. Ve plt adlı değişkene kaydedildi.

```
plt <- plt +  
  scale_x_continuous(  
    limits = c(0, 200000,10000),  
    breaks = seq(0, 180000, by = 60000),  
    expand = c(0, 0),  
    position = "top"  
  ) +  
  scale_y_discrete(expand = expansion(add = c(0, 0.5))) +  
  theme(  
    panel.background = element_rect(fill = "white"),  
    panel.grid.major.x = element_line(color = "#D38A00", size = 0.7),  
    axis.ticks.length = unit(0, "mm"),  
    axis.title = element_blank(),  
    axis.line.y.left = element_line(color = "black",size=0.7),  
    axis.text.y = element_blank(),  
    axis.text.x = element_text(family="Noto Serif",size = 14)  
  )
```

Yukarıdaki kod bloğunda:

- **scale\_x\_continuous()** fonksiyonu ile grafiğin x-ekseninin etiketlemesi yapıldı. position="top" ile x-ekseni grafiğin üst kısmına alındı.
- **scale\_y\_discrete()** fonksiyonu ile y-eksenindeki genişleme miktarı belirlendi. expand argümanı ile birlikte y-ekseninin alt sınırında bir genişleme olmadığını, üst sınırında bir miktar boşluk bırakılacağı belirlendi.
- **theme()** fonksiyonu ile grafiğin teması oluşturuldu.
  - a- **panel.background** argümanı ile grafiğin arka planı beyaz yapıldı.
  - b- **panel.grid.major.x** argümanı ile grafiğin x-eksenindeki ızgaraların rengi #D38A00 kodlu turuncu renkli yapıldı.
  - c- **axis.ticks.length** argümanı ile eksen işaretlerinin uzunluğu 0 mm olarak ayarlandı yani işaretler görünmez yapıldı.
  - d- **axis.title** argümanı ile eksen başlıkları kaldırıldı.

e- **axis.line.y.left** argümanı ile y-eksen çizgisinin rengi siyah yapıldı ve kalınlığı 0.7 olarak belirlendi.

f- **axis.text.y** argümanı ile grafiğin y-eksenindeki etiketler kaldırıldı.

g- **axis.text.x** argümanı ile x-eksenindeki etiketlerin yazı ailesi “Noto Serif” olarak ayarlandı. Yazı boyutu 14 olarak belirlendi

```
plt <- plt +  
  geom_text(  
    data = subset(salary_in_peryear),  
    aes(0, y = work_year, label = work_year),  
    hjust = 0,  
    nudge_x = 0.3,  
    colour = "white",  
    family="Noto Serif",  
    size = 7  
  ) +  
  geom_text(  
    data = subset(salary_in_peryear),  
    aes(x = avg_salary, y = work_year, label = round(avg_salary)),  
    hjust = 1,  
    nudge_x = 0.3,  
    colour = "white",  
    family = "Noto Serif",  
    size = 7  
  )  
)
```

Yukarıdaki kod bloğundaki **geom\_text()** fonksiyonları ile çubukların üstüne Yıl ve Ortalama Maaş etiketlemeleri yapıldı. **geom\_text()** fonksiyonu ile etiketleme yapılacak metinlerin hizalanma yönü, rengi, boyutu, yazı tipi ve konumları ayarlandı.



```
plt <- plt +
  labs(
    title = "Veri Bilimci Maaşlarının Değişimi",
    subtitle = "Yıllara Göre Ortalama Veri Bilimci Maaşı (USD), 2020-2024"
  ) +
  theme(
    plot.title = element_text(
      family = "Noto Serif",
      face = "bold",
      size = 22
    ),
    plot.subtitle = element_text(
      family = "Noto Serif",
      size = 20
    )
  )
)
```

Yukarıdaki kod bloğundaki **labs()** fonksiyonuyla beraber grafiğin ana ve alt başlıkları oluşturuldu. Daha sonrasında **theme()** fonksiyonu yardımıyla başlıkların yazı tipi, kalınlığı ve boyutu belirlendi.

```
plt <- plt +
  theme(
    plot.margin = margin(0.05, 0, 0.1, 0.01, "npc")
  )
)
```

**theme()** fonksiyonunun **plot.margin** argümanı ile grafiğin sırasıyla sol, sağ, üst ve alt kenarları için boşluklar belirlendi. “npc” parametresi ile grafik boyutuna göre normalize edilmiş bir uzunluk biriminin kullanılacağı belirtildi.

```
grid.lines(x = unit(0, "npc"), y = c(0.85, 0.95), gp = gpar(lwd = 10, col =
"red", linejoin = "mitre", lineend = "butt"))
```

**grid.lines()** fonksiyonu ile grafiğin başlıklarının sol kısmına kırmızı çizgi çekme işlemi yapıldı.

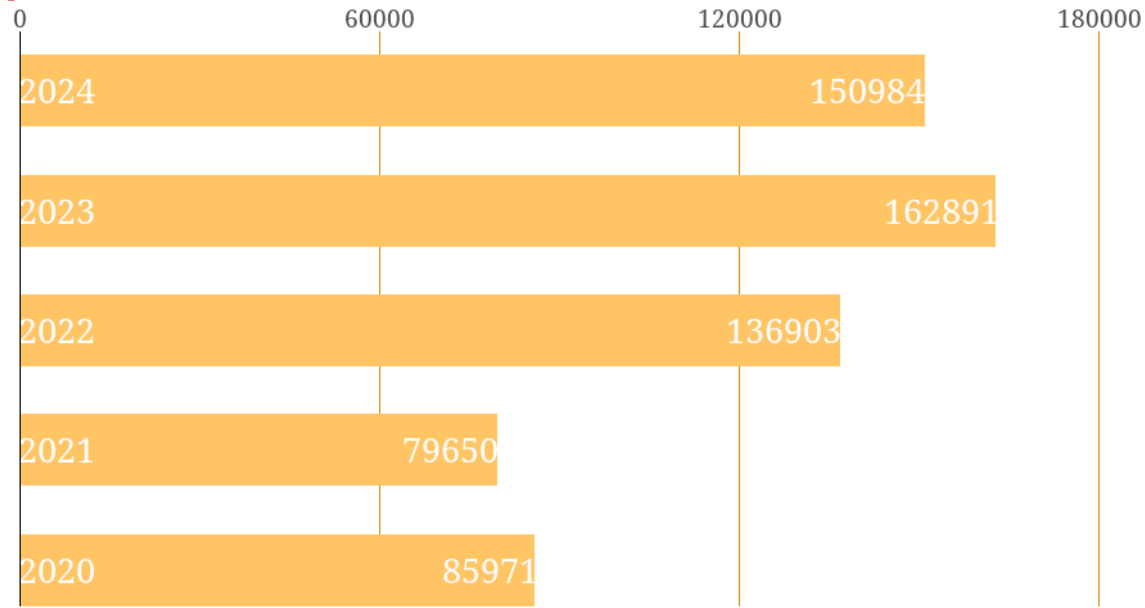
```
grid.text(  
  "Kaynak:Data Scientist Salary in 2024; Kaggle",  
  x = 0.005,  
  y = 0.06,  
  just = c("left", "bottom"),  
  gp = gpar(  
    col = "#8A8989",  
    fontsize = 16,  
    fontfamily = "Noto Serif"  
  )  
)
```

**grid.text()** fonksiyonu ile grafiğin sol alt kısmına alt bilgi (verinin adı ve kaynağı) yazdırıldı. x ve y parametreleri ile yazının başlayacağı koordinatlar belirlendi. just parametresi ile metnin hizalanma yönü belirlendi. gp parametresi ile yazının rengi,boyutu ve tipi belirlendi.

## GRAFIĞİN YORUMLANMASI

### Veri Bilimci Maaşlarının Değişimi

Yıllara Göre Ortalama Veri Bilimci Maaşı (USD), 2020-2024



Kaynak:Data Scientist Salary in 2024; Kaggle

Şekil 3

Şekil 3'teki grafik, 2020 ile 2024 yılları arasında "Data Scientist Salary in 2024" veri setindeki Veri Bilimcilerin yıllara göre dolar bazında ortalama maaşını göstermektedir. 2020'den 2024 yılına doğru ortalama maaşta artan bir trend olduğu görülmektedir. Bunun sebebinin son yıllarda veriye olan önemin artması ve bu nedenle Veri Bilimciye artan talep olduğu ileri sürülebilir. Bu grafik kariyerini Veri Bilimi alanında geliştirmek isteyenler için maaşlar konusunda bir fikir sahibi olmalarına da yardımcı olabilir.

# KAYNAKÇA

1- <https://www.kaggle.com/datasets/chopper53/data-scientist-salary-in-2024?resource=download>

2- <https://r-graph-gallery.com/web-horizontal-barplot-with-labels-the-economist.html>

3- <https://www.toucantoco.com/en/glossary/horizontal-barchart.html#:~:text=A%20horizontal%20bar%20chart%20is,indicates%20which%20data%20is%20represented.>

4- <https://www.jaspersoft.com/articles/what-is-a-bar-chart>

5- <https://twitter.com/OptaAnalyst/status/1787794271669538898>

6-

[https://www.reddit.com/r/dataisbeautiful/comments/1c029yb/oc\\_whos\\_won\\_the\\_most\\_pga\\_tournament\\_winnings/?utm\\_source=share&utm\\_medium=web3x&utm\\_name=web3xcss&utm\\_term=1&utm\\_content=share\\_button](https://www.reddit.com/r/dataisbeautiful/comments/1c029yb/oc_whos_won_the_most_pga_tournament_winnings/?utm_source=share&utm_medium=web3x&utm_name=web3xcss&utm_term=1&utm_content=share_button)