

# R ile Veri Görselleştirme

**Dinçer GÖKSÜLÜK, Phd.**

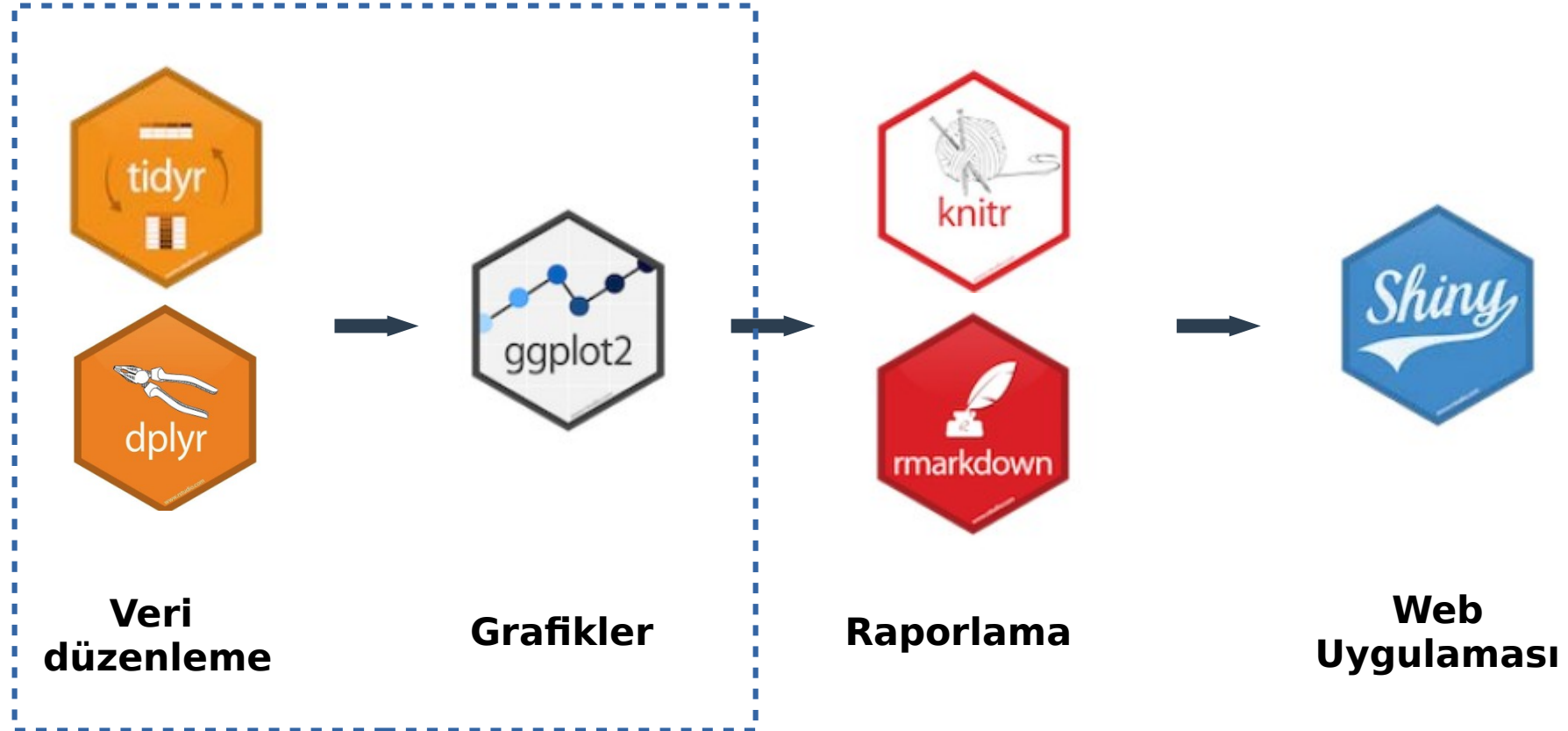
Erciyes Üniversitesi, Tıp Fakültesi  
Biyoistatistik Anabilim Dalı

# Grafiklerin kullanım alanları...

- Veri analizi
  - Tanımlayıcı istatistik
  - Modelleme / Tahmin
  - Varsayım kontrolleri
  - ...
- Raporlama
- Sunum

# Kullanılan araçlar...

- **R, R Studio**
- **R Kütüphaneleri**
  - graphics, lattice, **ggplot2**, ...
  - highcharts, rayrender, gganimate, ... (üç boytlı, interaktif grafikler)
- **Tekrar edilebilir raporlamalar**
  - TeX, pandoc, ...
  - knitr, Sweave, R Markdown, Quarto

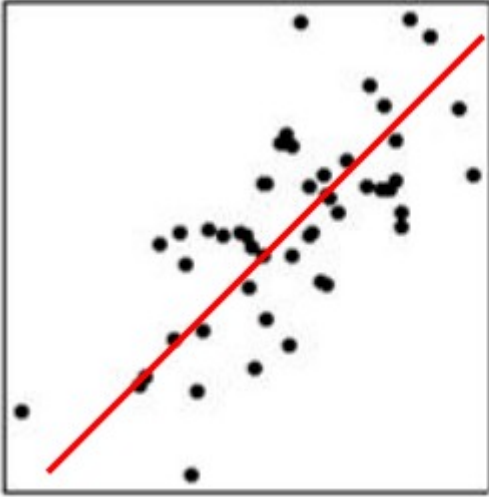


# Neden grafik?

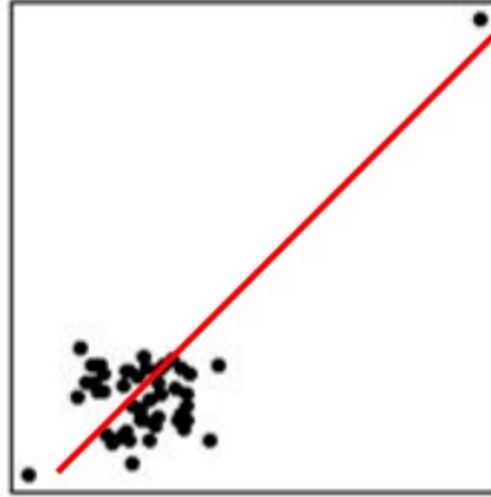
“A picture is worth a thousand words.”

“A **correctly selected** picture is worth a thousand words.”

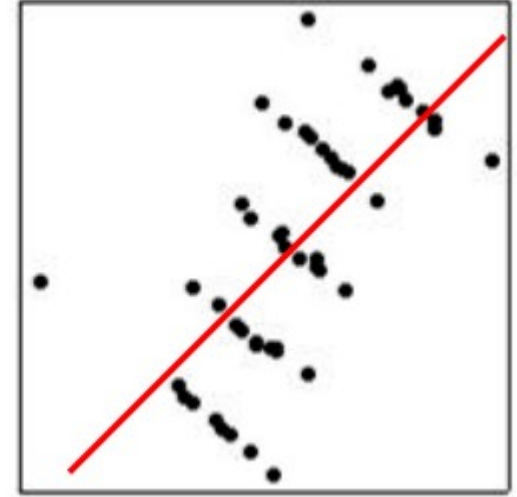
# Neden grafik?



**İstenilen durum**



**Korelasyon = 0**  
**Aykırı gözlem: 2**



**Doğrusal ilişki??**

İstatistiksel analizler sonucunda aynı korelasyon değerine sahip olduğu görülen üç farklı veri yapısı

# 80-20 Kuralı: Veri Analizi

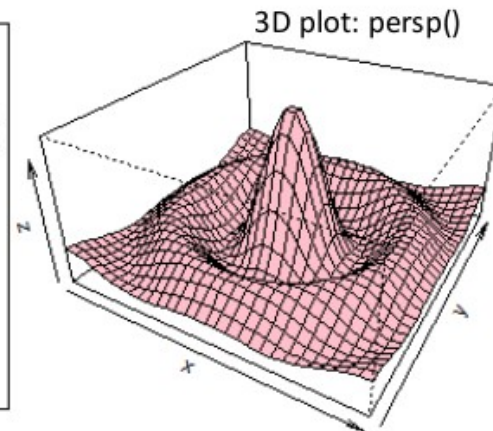
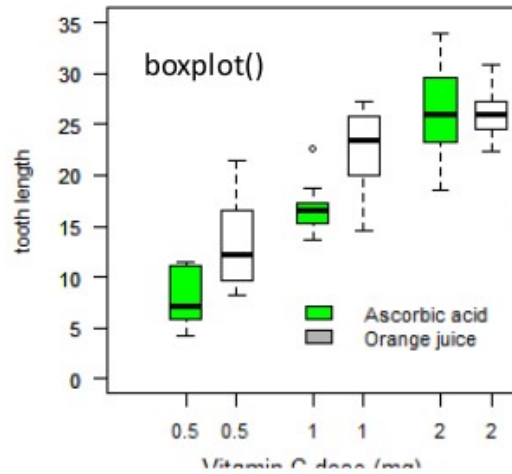
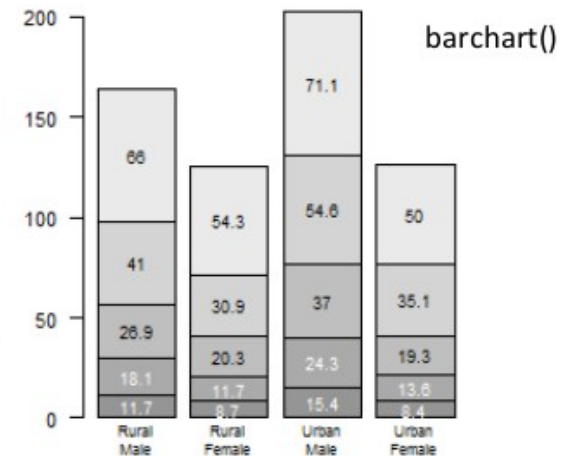
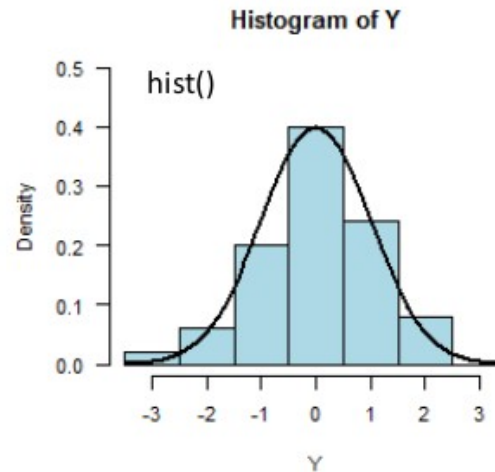
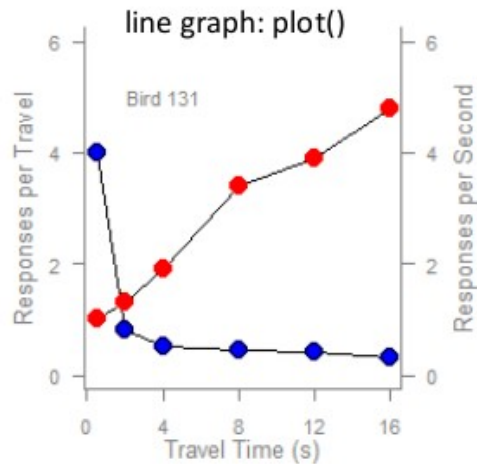
- Standart bir veri analizinde zamanın %80'i veri hazırlama ve düzenleme için harcanmaktadır.
  - Veri toplama, R'a aktarma, etiketleme (kategorik değişkenler)
  - Veri ön işleme (aykırı değerler, satır/sütun düzenlemeleri, ..)
- Analiz öncesi aşamalardaki süreyi azalttıkça;
  - Farklı modeller ile denemeler
  - Özgün ve etkin grafikler
  - Kapsamlı raporlama, değerlendirme ve tartışma

# 80-20 Kuralı: Grafikler

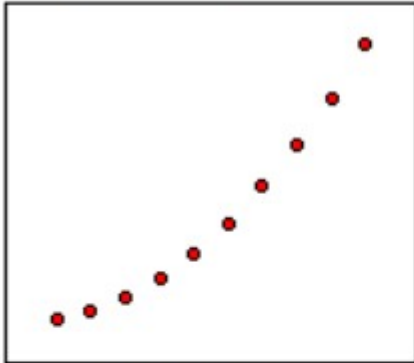
- **Veri Analizi:** Zamanın %20'si ile istenilen bir grafik %80 oranında hazırlanabilir.
  - Varsayılan grafik opsiyonlarını kullanmak
  - Hazır grafik kütüphaneleri ve fonksiyonlarını kullanmak
  - %10 fazladan zaman harcayarak grafikler daha etkin hale getirilebilir.
- **Raporlama ve Sunum:** Zamanın %80'i ile istenilen bir grafiğin %20'si hazırlanabilir.
  - Çözünürlük kalitesi (eksen isimlerinin ve grafik üzerindeki bilgilerin net olarak okunabilmesi)
  - Sayfa düzenine uygunluk



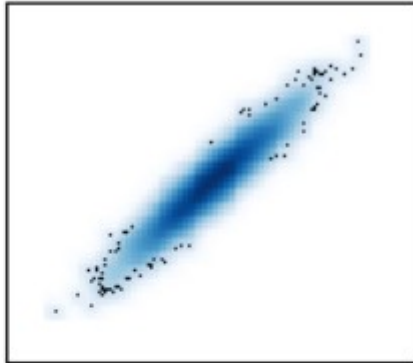
# Temel grafikler



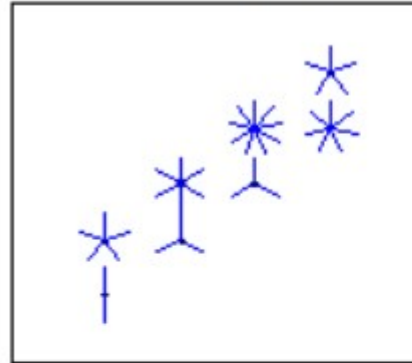
`plot(num, num)`



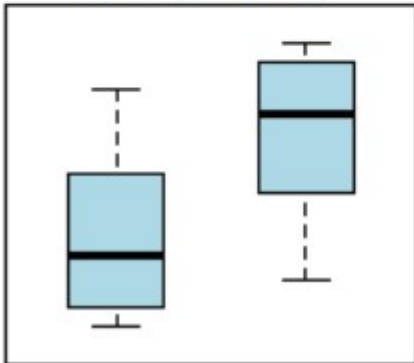
`smoothScatter()`



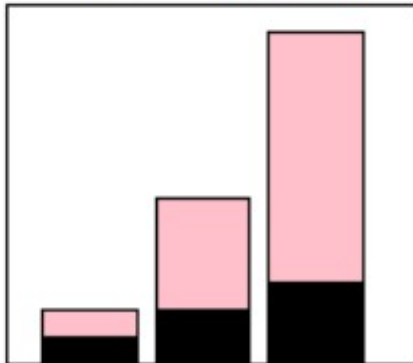
`sunflowerplot()`



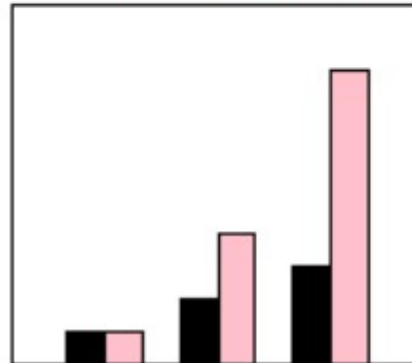
`boxplot(list)`  
`plot(fac, num)`



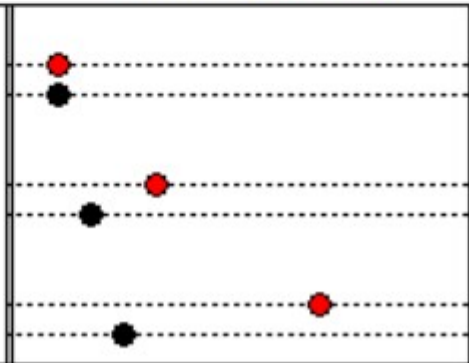
`barplot(matrix)`

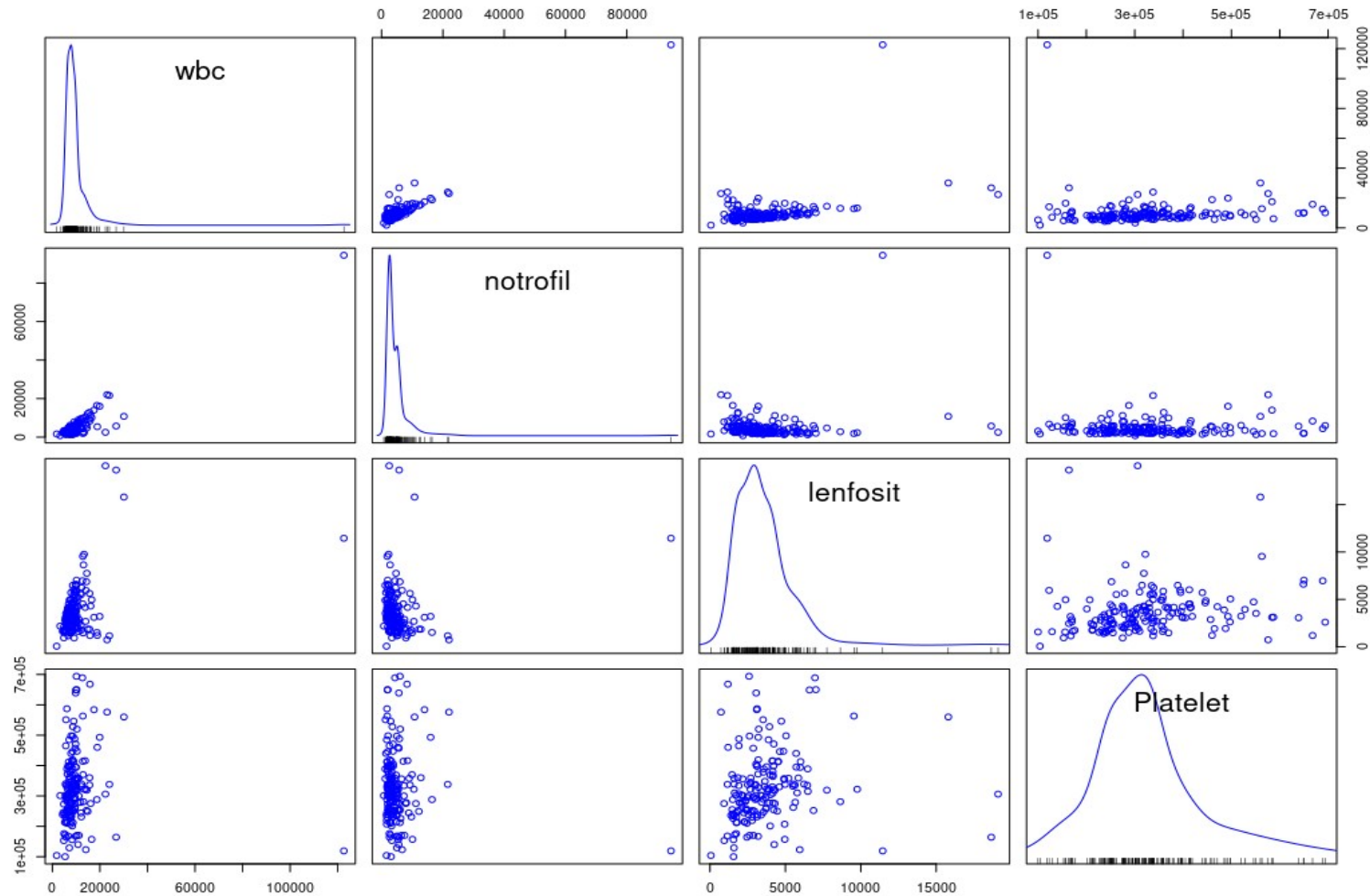


`barplot(matrix)`



`dotchart(matrix)`





```
library(car)
scatterplotMatrix(~ wbc + notrofil + lenfosit + Platelet, data = MPV_data,
  smooth = FALSE, regLine = FALSE
)
```



36



69



92



301



331



348



374



407



460



484



506



546



585



618



652



768



856



861



883



888

## Chernoff Yüzleri Grafiği

# Coğrafik bilgi sistemleri...

This example, from Paul Murrell's *R Graphics* book shows a basic map of Brazil, with provinces and their capitals, shaded by region of the country.

Data-based maps can show spatial variation of some variable of interest



Murrell, Fig. 14.5

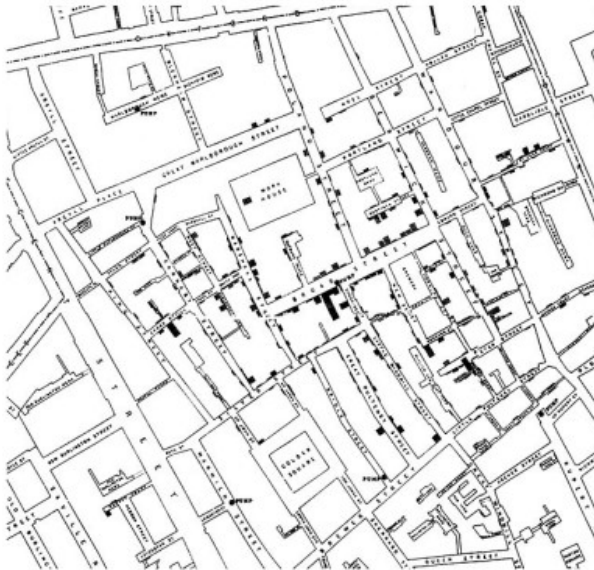


# Coğrafik bilgi sistemleri...

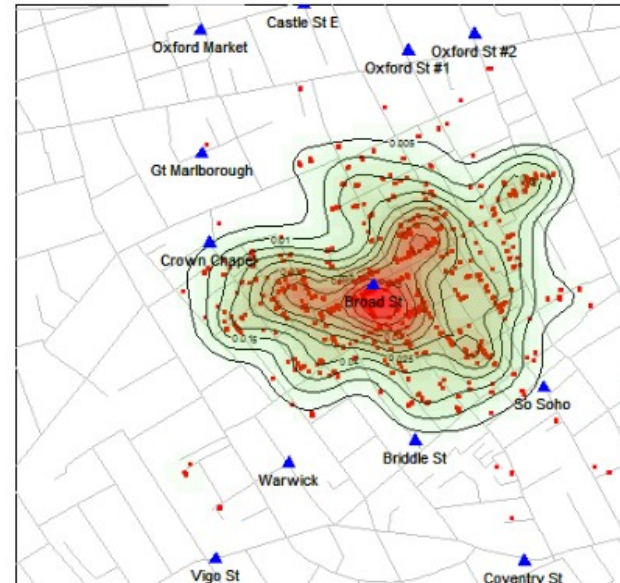
Dr. John Snow's map of cholera in London, 1854

Enhanced in R in the [HistData](#) package to make Snow's point

Portion of Snow's map:



Snow's Cholera Map, Death Intensity



```
library(HistData)
SnowMap(density=TRUE,
  main="Snow's Cholera Map, Death Intensity")
```

Contours of death densities are calculated using a 2d binned kernel density estimate, [bkde2D\(\)](#) from the [KernSmooth](#) package

# Coğrafik bilgi sistemleri...

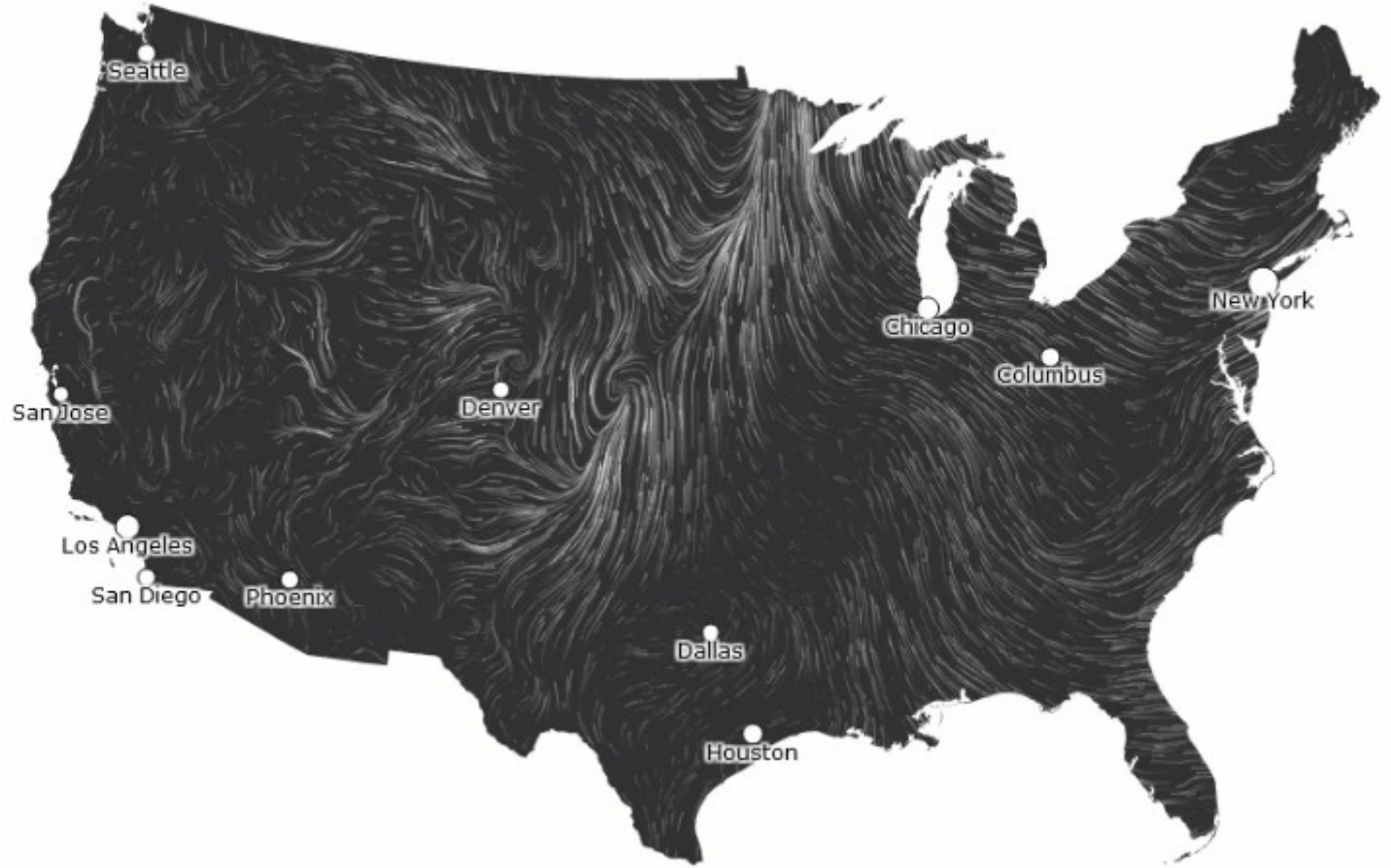
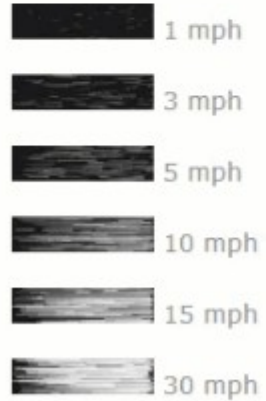
Sept. 24, 2018

11:22 am EST

(time of forecast download)

top speed: 31.3 mph

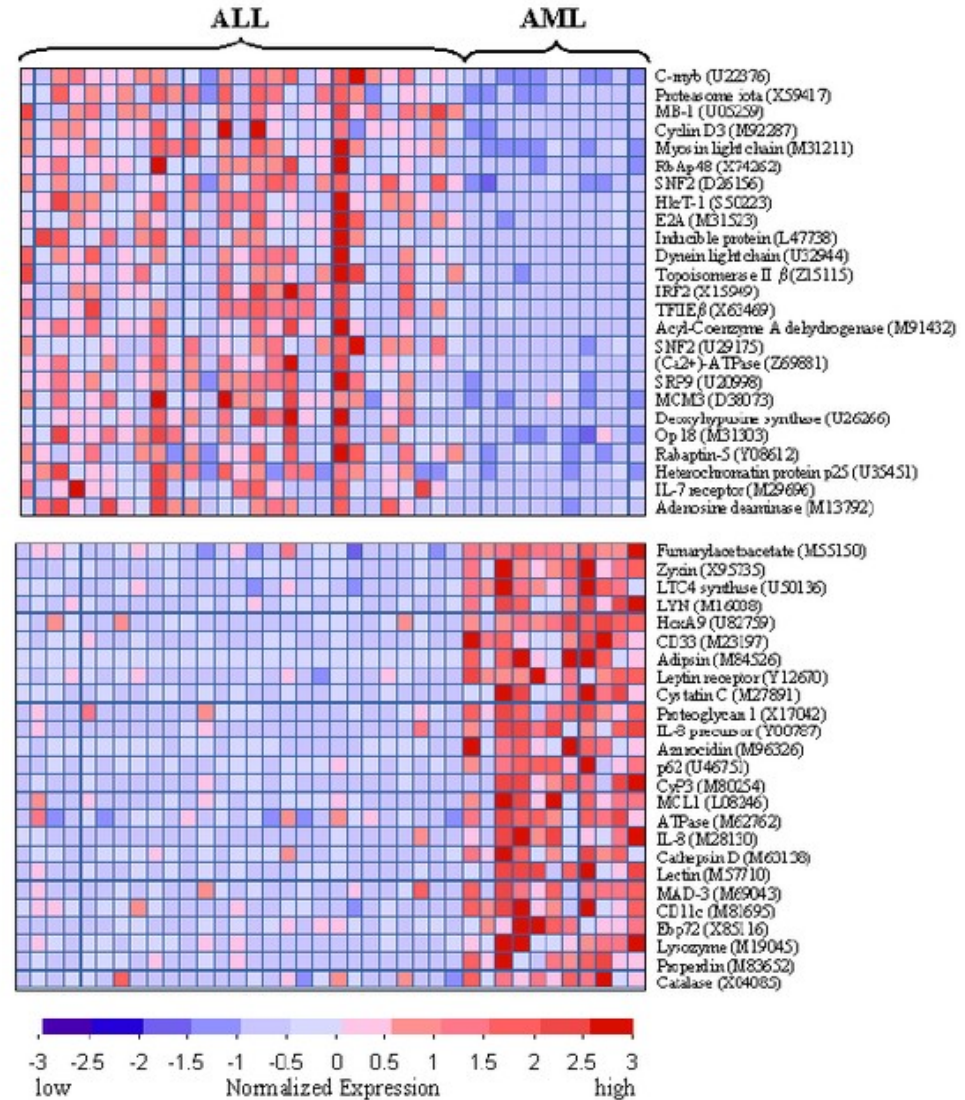
average: 7.9 mph



# Biyoinformatik...

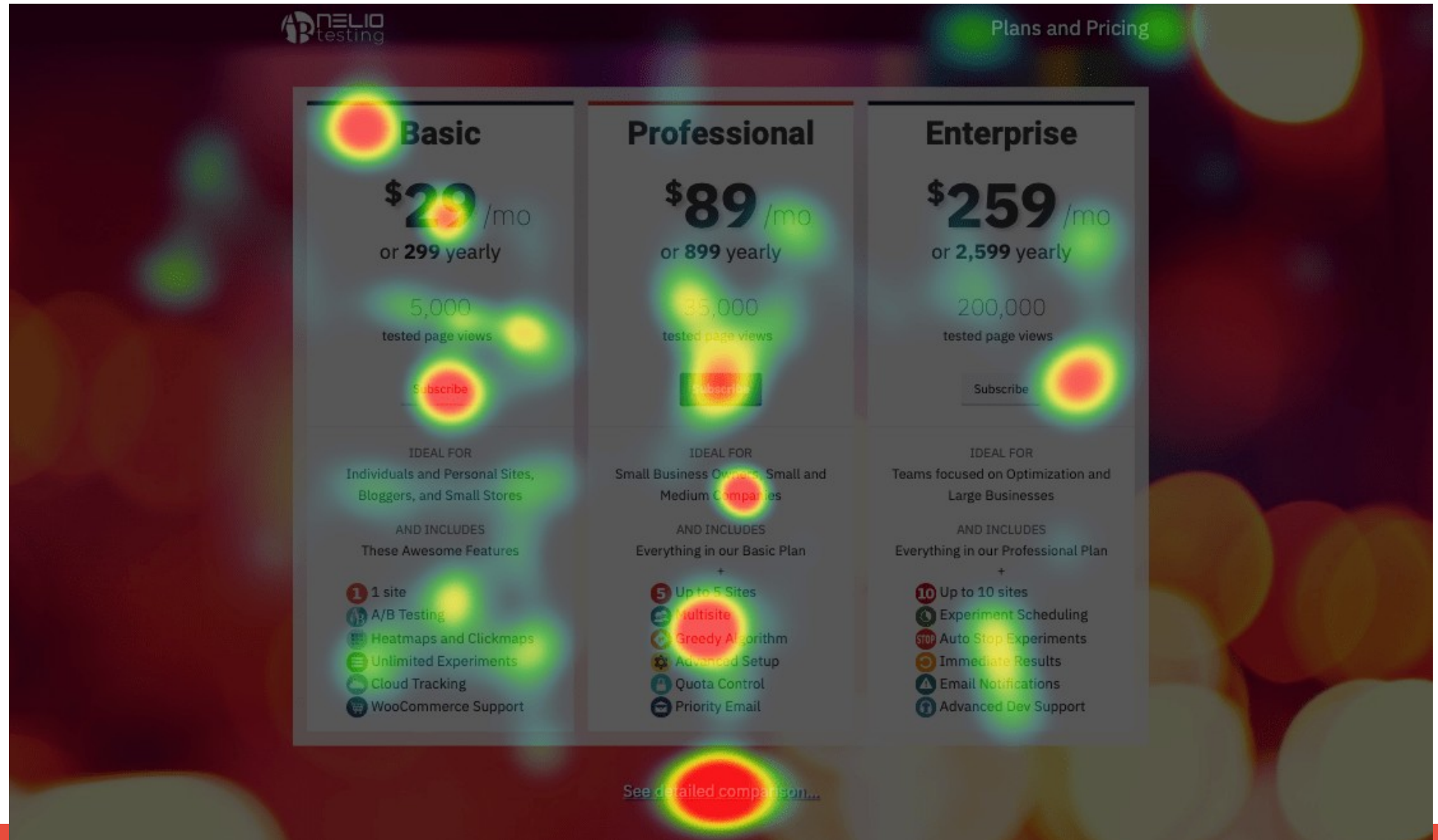
Gen ifade verileri üzerinden hastalık ile ilişkili genetik mekanizmaların belirlenmesi

**Kaynak:** Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., & Lander, E. S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (New York, N.Y.), 286(5439), 531-537.



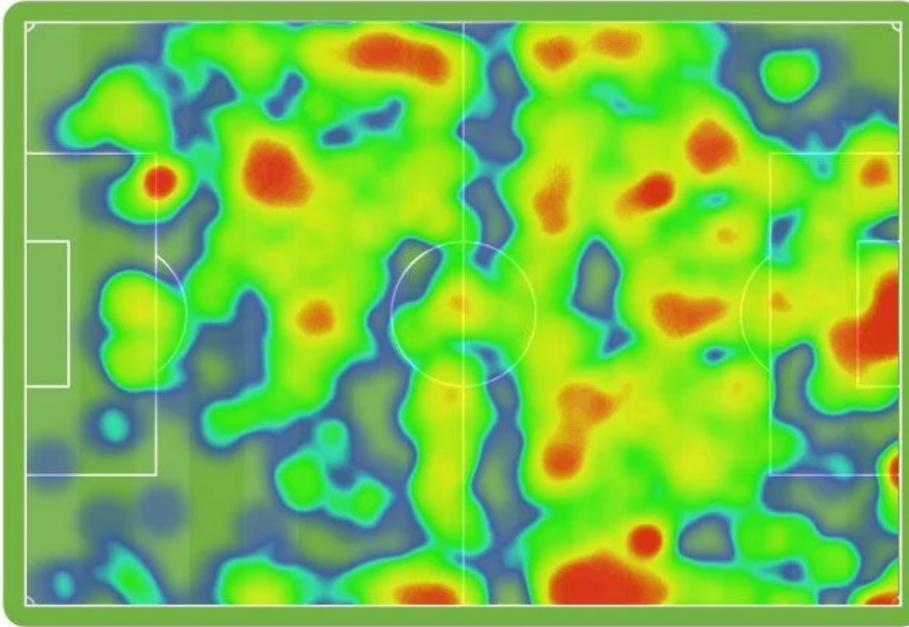


# Nereye odaklanıyoruz?



# Futbol istatistikleri

## Heat Map



Germany

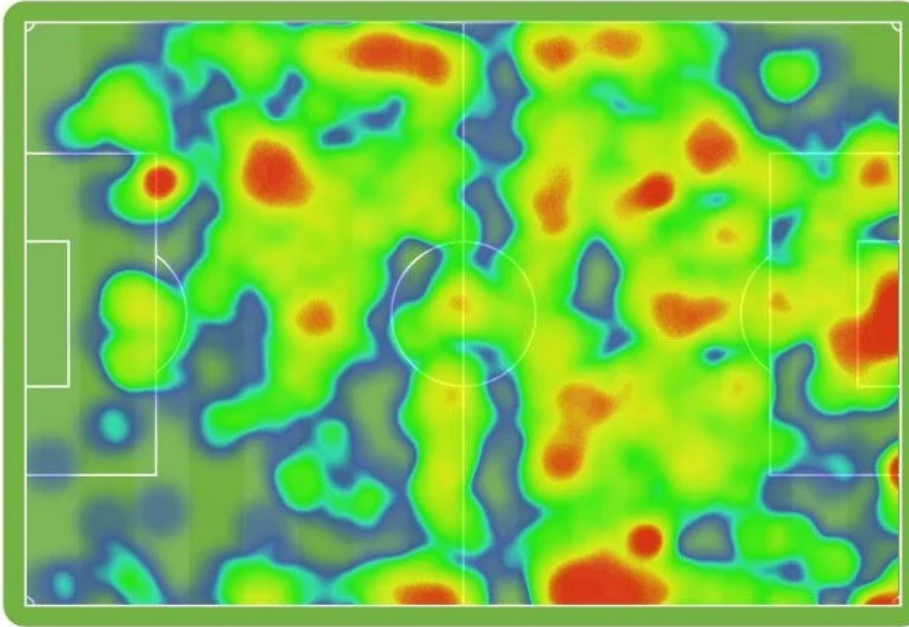


Brazil

- 2014 Dünya Kupası Yarı Final maçı.
- Her iki takımda yarı finale hiç yenilmeden çıkma başarısı göstermiştir.

# Futbol istatistikleri

Heat Map



Germany



Brazil

**7**

-

**1**

# ggplot2: Grammars of Graphics

## ggplot2 kütüphanesi

- İleri seviye görselleştirme
- Zengin grafik seçenekleri
- ggplot() fonksiyonu etrafında kurgulanmış kod yapısı
- Ön tanımlı grafik türleri için qplot() fonksiyonu

Grammars of Graphics, **Leland Wilkinson**

**ggplot():** Grafik arayüzünü başlatan ana fonksiyon.

- Veri seti
- Grafik bileşenleri, `aes(...)` fonksiyonu ile tanımlanır.

**geom\_\*():** Grafik geometri bileşenleri. Grafikte ne görmek istiyoruz?

- Noktalar, çizgiler, geometrik şekiller, metin ifadeler, ...
- `geom_points()`, `geom_line()`, `geom_text()`, ...



**scale\_\*():** Veri setindeki değerlerin grafiğe nasıl aktarılacağını belirleyen çeşitli seçenekler sunar.

- Renkler, çizgi türleri, etiket isimleri
- Eksen değerlerinin dönüşümleri, vb.

**theme(), theme\_\*():** Grafikte kullanılan temalar.

- Yazı tipi, büyüklüğü, rengi, vb.
- theme\_bw(), theme\_classic(), ...

**facet\_\*():** Grafiklerin bölüntülenmesi ve alt gruplara göre grafiklerin çizilmesi

**stat\_\*:** Veri üzerinden özet istatistikler elde edip grafik çizerken bu bilgilerin kullanılması

```

ggplot (data = <DATA>) +
<GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),
stat = <STAT>, position = <POSITION>) +
<COORDINATE_FUNCTION> +
<FACET_FUNCTION> +
<SCALE_FUNCTION> +
<THEME_FUNCTION>

```

required

Not required, sensible defaults supplied

```

ggplot(data = MPV) +
geom_point(mapping = aes(x = notrofil, y = platelet)) +
theme(panel.grid = element_blank()) +
facet_wrap(~cinsiyet, ncol = 2)

```



# ggplot2: Grammars of Graphics

**Kitap:** <https://ggplot2-book.org/>

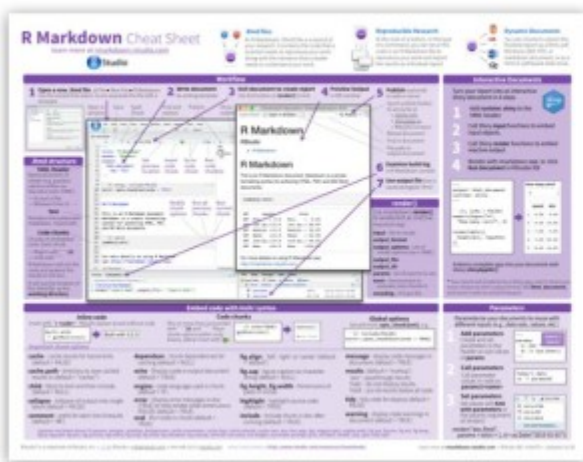
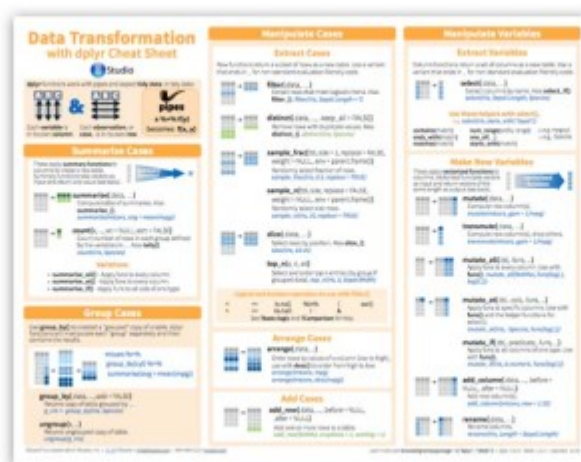
**Cookbook:** <https://r-graphics.org/index.html>

**Cheatsheets:**

<https://www.rstudio.com/resources/cheatsheets/>

**Paket Referans Sayfası:**

<https://ggplot2.tidyverse.org/index.html>

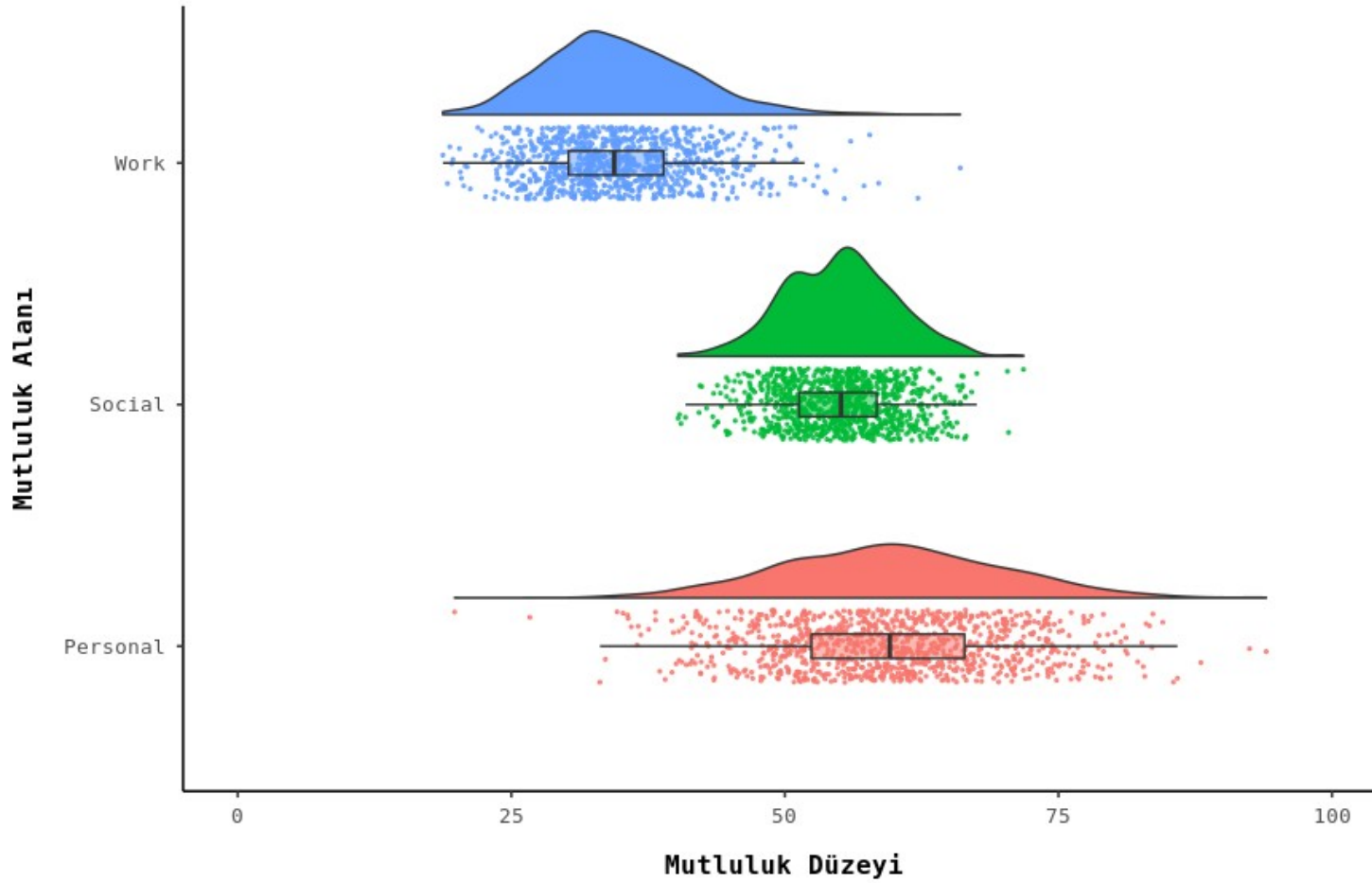


<https://www.rstudio.com/resources/cheatsheets/>

# GRAFİK UYGULAMASI

## Mutluluk Düzeyi Araştırması

Üç farklı alanda mutluluk düzeyleri [0 - 100] aralığında ölçülmüştür.

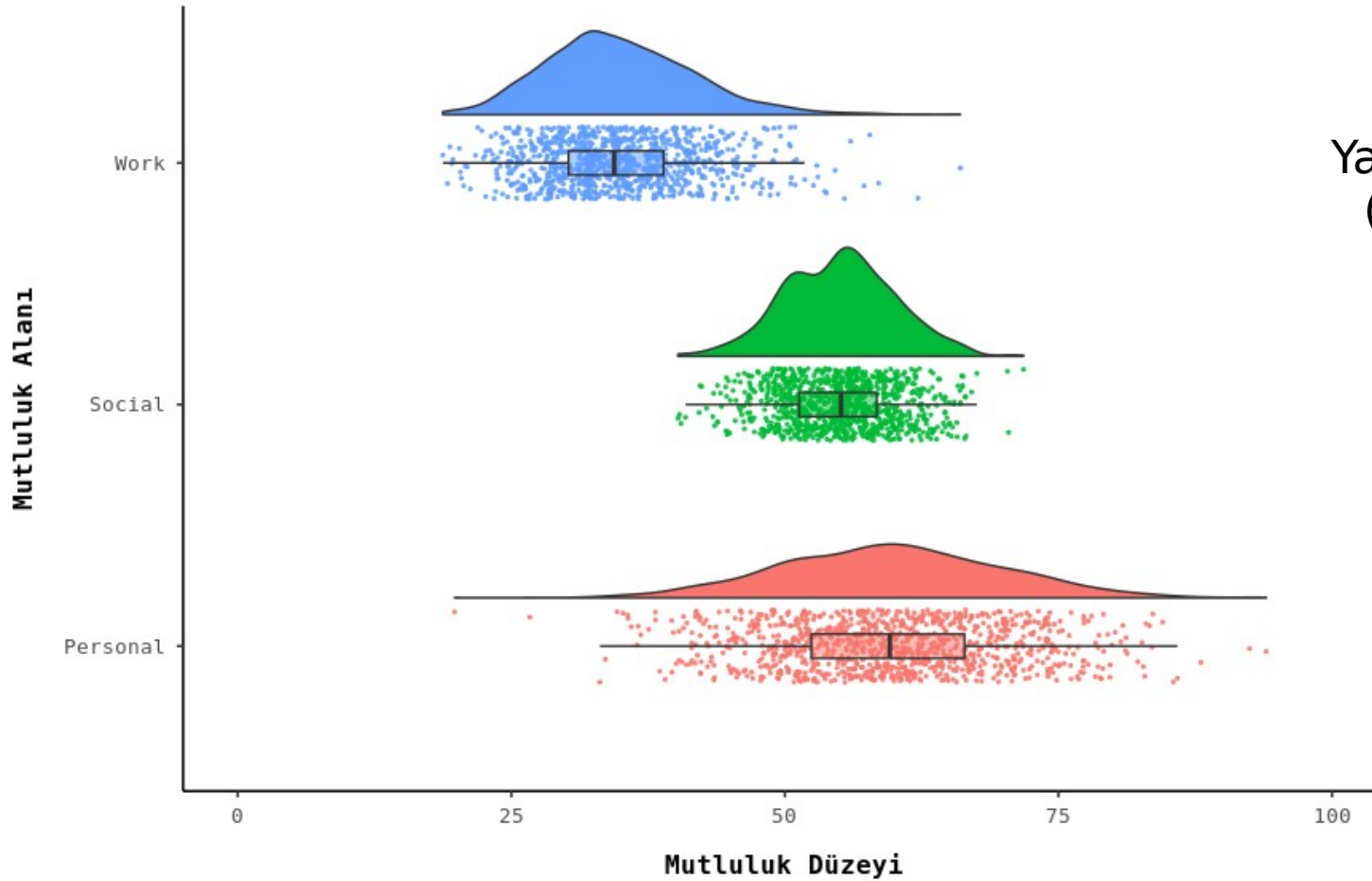


Grafiğin adı ??

# GRAFİK UYGULAMASI

## Mutluluk Düzeyi Araştırması

Üç farklı alanda mutluluk düzeyleri [0 - 100] aralığında ölçülmüştür.



Yağmur Bulutu Grafiği  
(Rain Cloud Graph)