

R Programlama

Muhammed Fatih Tüzen

Table of contents

Preface	3
I R Programlama Hakkında	4
R Programı ile Neler Yapılabilir	5
R Programlama ile ilgili Faydalı Kaynaklar	6
R ve RStudio'nun Bilgisayara Kurulması	7
R Studio Kişiselleştirme	10
II R Programlamaya Giriş	12
1 Summary	15
2 Introduction	16
III Veri Manipulasyonu	17
3 select	21

Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 + 1

[1] 2

Part I

R Programlama Hakkında

R programlama, veri analizi, istatistiksel ve ekonometrik hesaplamalar, veri görselleştirme ve veri madenciliği gibi istatistiksel ve veri analitiği işlemleri için kullanılan bir programlama dilidir. İlk olarak 1990 yılında **Ross Ihaka** ve **Robert Gentleman** tarafından geliştirilmeye başlanmıştır ve o zamandan bu yana istatistiksel analiz alanında çok popüler bir araç haline gelmiştir. Yazılım ismini yazarların isimlerinin baş harflerinden almaktadır.

R Programı ile Neler Yapılabilir

R, açık kaynaklı bir programlama dili ve yazılım ortamıdır, bu da onu geniş bir kullanıcı topluluğu tarafından desteklenen ve geliştirilen bir platform yapar. R ile yapılabilecek başlıca işler şunlardır:

1. **Veri Analizi:** R, veri çerçeveleri ve veri setleri üzerinde işlem yapmak için bir dizi fonksiyon ve araç sunar. Veri temizleme, dönüştürme, özeti alma ve analiz etme işlemleri R ile kolayca gerçekleştirilebilir.
2. **Veri Görselleştirme:** R, ggplot2 gibi grafik paketleri ile verilerinizi görselleştirmenize olanak tanır. Çeşitli grafik türleri (çizgi grafikleri, sütun grafikleri, dağılım grafikleri vb.) oluşturabilirsiniz.
3. **İstatistiksel Analiz:** R, istatistiksel modelleri oluşturmak, hipotez testleri yapmak ve regresyon analizi gibi istatistiksel analizler gerçekleştirmek için zengin bir araç seti sunar. Ayrıca zaman serisi analizi ve kümeleme gibi konularda da kullanılır.
4. **Veri Madenciliği:** R, veri madenciliği uygulamaları için kullanılabilir. Makine öğrenimi algoritmaları uygulamak ve veri madenciliği projeleri geliştirmek için paketler içerir.
5. **Raporlama:** R Markdown kullanarak veri analizi ve sonuçlarını raporlama için kullanılır. Bu, anlamlı ve formatlı raporlar oluşturmanıza yardımcı olur.
6. **Paketler ve Genişletilebilirlik:** R, kullanıcıların işlevselliği genişletmek için paketler ekleyebileceği bir sistem sunar. CRAN (Comprehensive R Archive Network) gibi kaynaklar, binlerce paketi içeren bir depo sağlar.

Not

R programlama özellikle istatistik, veri bilimi ve akademik araştırmalar alanlarında çok kullanılır, ancak endüstriyel uygulamalarda da giderek daha fazla kullanılmaktadır. R'nin açık kaynaklı olması ve geniş bir kullanıcı topluluğuna sahip olması, bu dilin popülerliğini artırmıştır. R ile çalışmak için temel programlama bilgisine sahip olmak yararlı olacaktır, ancak öğrenmesi oldukça erişilebilir bir dildir ve çevrimiçi kaynaklar ve kurslar mevcuttur.

R Programlama ile ilgili Faydalı Kaynaklar

R programlamayı öğrenmek ve geliştirmek için bir dizi faydalı kaynak bulunmaktadır. R programlamaya başlamak veya ilerlemek için kullanabileceğiniz bazı kaynaklar:

1. **Resmi R Web Sitesi:** R'nin resmi web sitesi (<https://www.r-project.org/>) R programlamaya başlamak için temel kaynaktır. Burada R'nin indirilmesi, kurulumu ve temel belgelendirme bilgilerine erişebilirsiniz.
2. **RStudio:** R programlama için yaygın olarak kullanılan RStudio IDE'si (Entegre Geliştirme Ortamı), R kodlarını yazmak, çalıştırmak ve yönetmek için güçlü bir araçtır. RStudio'nun resmi web sitesi (<https://www.rstudio.com/>) RStudio'nun indirilmesi ve kullanımı hakkında bilgi sunar.
3. **R Dersleri ve Kurslar:** İnternette birçok ücretsiz R dersi ve kursu bulabilirsiniz. Coursera, edX, Udemy ve DataCamp gibi platformlar, R programlamayı öğrenmek için çeşitli kurslar sunmaktadır.
4. **R Belgeleri:** R'nin resmi belgeleme (<https://cran.r-project.org/manuals.html>) kaynakları, R dilinin temellerini ve paketlerini öğrenmek için çok faydalıdır. R'deki komutlar ve fonksiyonlar hakkında ayrıntılı bilgi içerirler.
5. **Kitaplar:** R programlamayı öğrenmek için yazılmış birçok kitap bulunmaktadır. Örnek olarak, "R Graphics Cookbook" (Hadley Wickham), "R for Data Science" (Hadley Wickham ve Garrett Grolemund), "Advanced R" (Hadley Wickham) gibi kitaplar önerilebilir.
6. **Stack Overflow:** Programlama sorunları ve hatalarıyla karşılaştığınızda, Stack Overflow gibi forumlarda R ile ilgili sorular sormak ve cevaplamak için topluluktan yardım alabilirsiniz.
7. **GitHub:** R ile ilgili açık kaynaklı projeleri incelemek ve kendi projelerinizi paylaşmak için GitHub gibi platformları kullanabilirsiniz. GitHub'da R kodlarını içeren birçok depo bulunmaktadır.
8. **Bloglar ve Videolar:** R ile ilgili bloglar ve YouTube kanalları, öğrenmek ve güncel kalmak için harika kaynaklardır. RStudio Blog (<https://blog.rstudio.com/>) ve YouTube'da R ile ilgili videoları bulabileceğiniz RStudio'nun resmi kanalı bunlara örnektir.

Tavsiye

R programlamayı öğrenmek ve geliştirmek için sürekli olarak yeni kaynaklar ve materyaller üretilmektedir. İhtiyacınıza ve seviyenize uygun kaynakları seçmek için zaman ayırın ve kendi hızınıza göre öğrenmeye devam edin.

R ve RStudio'nun Bilgisayara Kurulması

R'ın internet sitesinden işletim sisteminize uygun programı indirip kurabilirsiniz. Linux, Mac OS ve Windows işletim sistemleri için sürümleri mevcuttur.

Windows İşletim Sistemi İçin R Kurulumu

1. R programını indirmek için R resmi web sitesini ziyaret edin: <https://cran.r-project.org/>
2. Sayfanın üst kısmında “Download R for Windows” başlığını bulun ve tıklayın.

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2023-06-16, Beagle Scouts) [R-4.3.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

3. İndirilen sayfada “base” sekmesine tıklayın.

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to install R for the first time .
contrib	Binaries of contributed CRAN packages (for R >= 3.4.x).
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 3.4.x).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

4. Açılan sayfada “Download R 4.3.1 for Windows” linkine tıklayın ve dosyayı indirin.

[Download R-4.3.1 for Windows](#) (79 megabytes, 64 bit)

[README on the Windows binary distribution](#)
[New features in this version](#)

This build requires UCRT, which is part of Windows since Windows 10 and Windows Server 2016. On older systems, UCRT has to be installed manually from [here](#).

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is
CRAN.MIRROR>bin/windows/base/release.html.

Last change: 2023-06-16

⚠ Dikkat

Sayfayı ziyaret ettiğiniz tarihlerde farklı sürümlerin olabileceğine dikkat edin. Örneğin ileri bir tarihte bu sayfayı ziyaret ettiğinizde R programının yeni sürümü ile karşılaşabilirsiniz. O yüzden sürüm bilgisi değişiklik gösterebilir.

5. İndirilen dosyayı çift tıklayarak çalıştırın ve yükleyiciyi başlatın.
6. Yükleyici, R'nin temel sürümünü yüklemek için sizi yönlendirecektir. Varsayılan ayarları genellikle kabul edebilirsiniz.
7. Kurulum tamamlandığında, R'yi çalıştırmak için masaüstünüzde veya Başlat menüsünde “R” simgesini bulabilirsiniz.

Windows İşletim Sistemi İçin R Studio Kurulumu

R editörü grafiksel bir arayüz olmayıp eski tip bir yazılım konsoludur. **R Studio**, R programlama dili için geliştirilmiş entegre bir geliştirme ortamı (IDE) ve arayüzüdür. R Studio, R kodlarını daha verimli bir şekilde yazmanıza, çalıştırmanıza ve yönetmenize olanak tanıyan daha modern ve kullanışlı bir arayüz sunmaktadır. Ayrıca veri analizi, görselleştirme ve raporlama işlemleri için güçlü bir platform sunar. R Studio, açık kaynak bir projedir ve ücretsiz olarak kullanılabilir.

R Studio'nun kurulumu aşağıdaki adımlarla gerçekleştirilebilir:

1. R Studio'nun en son sürümünü indirmek için aşağıdaki bağlantıyı kullanın:
<https://www.rstudio.com/products/rstudio/download/>
2. Sayfada “Download RStudio Desktop for Windows” kısmına tıklayın ve indirmeyi başlatın.

DOWNLOAD

RStudio Desktop

Used by millions of people weekly, the RStudio integrated development environment (IDE) is a set of tools built to help you be more productive with R and Python.

Don't want to download or install anything? Get started with RStudio on [Posit Cloud for free](#). If you're a professional data scientist looking to download RStudio and also need common enterprise features, don't hesitate to [book a call with us](#).

1: Install R

RStudio requires R 3.3.0+. Choose a version of R that matches your computer's operating system.

DOWNLOAD AND INSTALL R

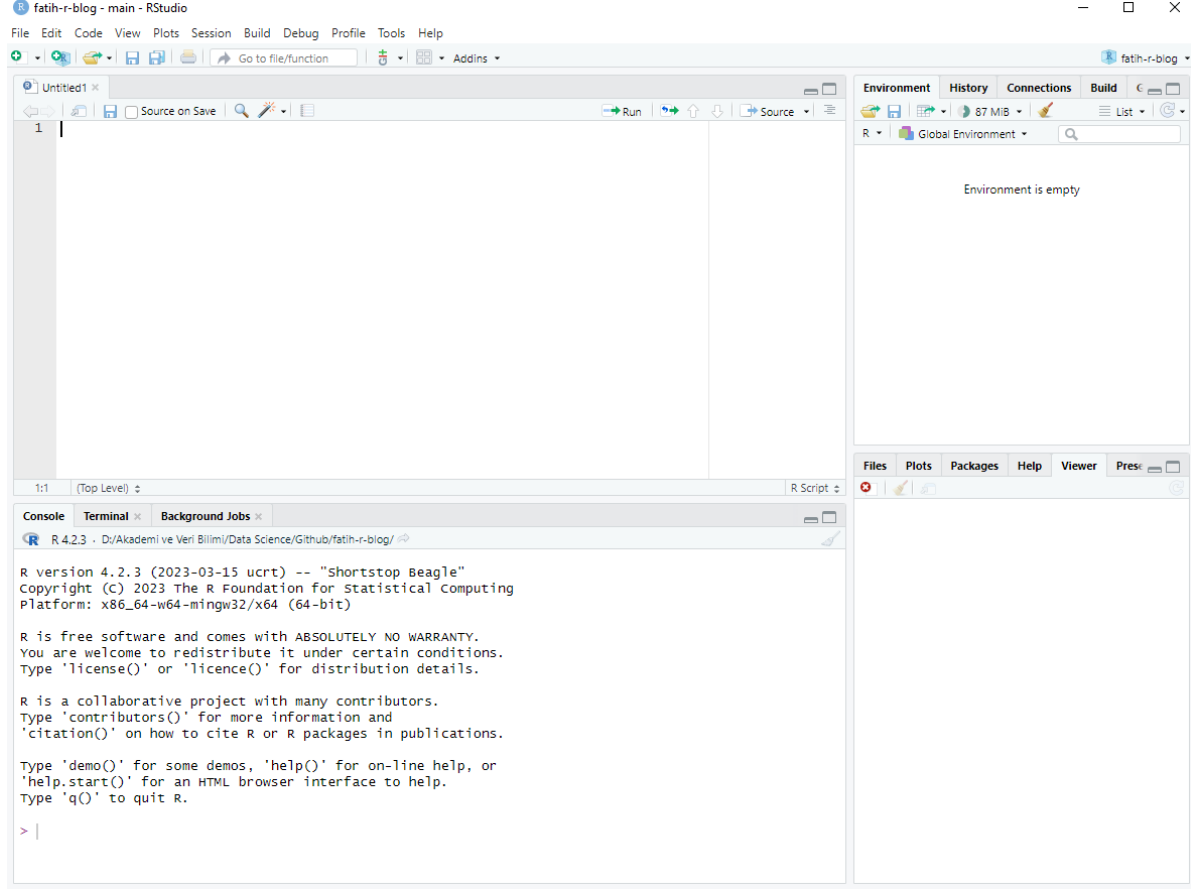
2: Install RStudio

DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 212.78 MB | [SHA-256: BCF6B866](#) | Version: 2023.06.2+561 |
Released: 2023-08-30

- İndirilen dosyayı çift tıklayarak çalıştırın ve kurulumu başlatın. Kurulum sırasında varsayılan ayarları genellikle kabul edebilirsiniz.
- Kurulum tamamlandığında, R Studio'yu başlatmak için masaüstünüzde veya Başlat menüsünde **"RStudio"** simgesini bulabilirsiniz.

R Studio Kişiselleştirme



RStudio, kullanıcıların ihtiyaçlarına göre kişiselleştirilebilen bir entegre geliştirme ortamı (IDE) sunar. RStudio’yu kişiselleştirmek için aşağıdaki yolları kullanabilirsiniz:

1. **R Studio Arayüzündeki Alanları Değiştirme:** Resimde görüldüğü gibi yeni bir R Script açıldığında arayüzde 4 farklı alan görülmektedir. Bu alanlar isteğe göre yer değiştirilebilmektedir. Bunun için **“Tools”** (Araçlar) menüsünden **“Global Options”** (Genel Ayarlar) sekmesi açılır. Buradan **“Pane Layout”** kısmından istenilen ayarlar yapılabilir.
2. **Temayı ve Editör Stilini Değiştirme:** RStudio’nun görünümünü değiştirmek için birçok tema ve editör stilini seçebilirsiniz. Bu, yazılım geliştirme ortamınızın daha hoş veya kullanışlı olmasını sağlar. **“Tools”** (Araçlar) menüsünden **“Global Options”** (Genel Ayarlar) sekmesini seçerek bu ayarları değiştirebilirsiniz.
3. **Klavye Kısayollarını Kişiselleştirme:** RStudio’da kullanılan klavye kısayollarını özelleştirebilirsiniz. **“Tools”** (Araçlar) menüsünden **“Modify Keyboard Shortcuts”**

(Klavye Kısayollarını Düzenle) seçeneğini kullanarak klavye kısayollarını tanımlayabilir veya değiştirebilirsiniz.

4. **Eklentileri ve Paketleri Kullanma:** RStudio, kullanıcıların işlevselliği genişletmek için eklentileri ve R paketlerini kullanmalarını sağlar. Bu paketler, kod otomatik tamamlama, kod görselleştirme, proje yönetimi gibi birçok işlemi kolaylaştırabilir. R Studio’nun sol üst köşesindeki “**Tools**” (Araçlar) menüsünden “**Install Packages**” (Paketleri Yükle) seçeneği ile yeni paketleri yükleyebilirsiniz.
5. **R Markdown Belgelerini Özelleştirme:** R Markdown belgeleri, raporlar ve belgeler oluşturmak için kullanılır. Bu belgeleri kişiselleştirebilirsiniz. R Markdown belgelerinin başlık, stil, tablo düzeni ve grafikler gibi birçok yönünü özelleştirebilirsiniz.
6. **Proje Ayarlarını Yapılandırma:** RStudio’da projeler kullanmak, projelerinizi daha düzenli ve etkili bir şekilde yönetmenize yardımcı olabilir. “File” (Dosya) menüsünden “New Project” (Yeni Proje) seçeneği ile yeni projeler oluşturabilir ve projelerinizi kişiselleştirebilirsiniz.
7. **Kod Tarayıcı ve Çalışma Ortamını Özelleştirme:** RStudio’nun sağ tarafında bulunan “**Environment**” (Çalışma Ortamı) ve “**Files**” (Dosyalar) sekmelerini özelleştirebilirsiniz. Bu sekmeleri dilediğiniz gibi düzenleyebilirsiniz.
8. **Addins Kullanma:** RStudio’nun “Addins” (Eklentiler) menüsü, kullanıcıların özel işlevleri ekleyebileceği bir bölümdür. Bu sayede belirli işlemleri hızlıca gerçekleştirebilirsiniz.

RStudio’nun bu kişiselleştirme seçenekleri, kullanıcıların kendi ihtiyaçlarına ve tercihlerine göre IDE’yi özelleştirmelerine olanak tanır. Bu şekilde, RStudio’yu daha verimli ve kişiselleştirilmiş bir şekilde kullanabilirsiniz. RStudio’nun ana bileşenleri ve temel özellikleri ise şunlardır:

1. **Script Editörü:** RStudio’nun sol üst kısmında yer alan bu bölüm, R kodlarını yazmak, düzenlemek ve çalıştırmak için kullanılır. Renk vurguları, otomatik tamamlama ve hata işaretleme gibi birçok yazılım geliştirme özelliği içerir.
2. **Environment (Çalışma Ortamı) :** Sağ üst köşede bulunan “Çalışma Ortamı” sekmesi, çalışan nesneleri ve değişkenleri görüntülemenizi sağlar. “Files” sekmesi ise projenizdeki dosyaları ve klasörleri görüntülemenize yardımcı olur.
3. **Console:** Alt sol köşede bulunan bu bölüm, R kodlarını anlık olarak çalıştırmanıza ve sonuçları görmesinize olanak tanır. R komutlarını doğrudan konsola yazabilir ve çalıştırabilirsiniz.
4. **Diğer Sekmeler :** RStudio, çeşitli grafikler ve görselleştirmeler oluşturmanıza olanak tanır. R koduyla çizilen grafikler, “**Plots**” sekmesinde görüntülenir. Bunu yanı sıra “**Help**” kısmında fonksiyonlar ile ilgili bilgi alınabilir, “**Packages**” kısmından ise paket yükleme vb. işler yapılabilir.

Part II

R Programlamaya Giriş

R kodunun çalıştırılması oldukça basittir ve R Studio gibi entegre geliştirme ortamları (IDE'ler) kullanırken daha da kolaylaşır. R kodunu çalıştırmak için temel adımlar:

1. **R Studio'yu Açın:** İlk adım, R Studio veya başka bir R IDE'sini açmaktır.

2. **Yeni Bir script oluşturun veya mevcut bir script kullanın:**

- R Studio'da, sol üst köşede bulunan “File” (Dosya) menüsünden “New Script” seçeneği ile yeni bir R scripti oluşturabilirsiniz.
- Mevcut bir scripte gitmek istiyorsanız, “File” menüsünden “Open Script” seçeneğini kullanabilirsiniz.

3. **R Kodunu Scripte Yazın:** Oluşturduğunuz veya açtığınız R skriptinde, R kodlarını yazın veya yapıştırın. Örneğin, basit bir hesaplama yapmak için aşağıdaki kodu kullanabilirsiniz:

```
x <- 5
y <- 10
z <- x + y
z
```

```
[1] 15
```

4. **Kodu Çalıştırma:**

- Çalıştırmak istediğiniz kodu seçin veya imleci çalıştırmak istediğiniz satıra getirin.
- Çalıştırma işlemi için aşağıdaki yöntemlerden birini kullanabilirsiniz:
 - Klavyede varsayılan olarak “Ctrl+Enter” (Windows/Linux) veya “Command+Enter” (Mac) tuş kombinasyonunu kullanabilirsiniz.
 - R Studio'daki “Run” (Çalıştır) düğmesini veya “Run” (Çalıştır) menüsünü kullanabilirsiniz.
 - Çalıştırmak istediğiniz kodu seçtikten sonra sağ tıklarsanız, “Run” (Çalıştır) seçeneğini göreceksiniz.

5. **Sonuçları İnceleyin:** Çalıştırılan kodun sonuçları konsol penceresinde veya çıktı bölümünde görüntülenir. Örneğin, yukarıdaki örnekte “z” değişkeninin değeri olan “15” sonucunu göreceksiniz.

Dikkat

Bir script üzerinden çalıştırılan R kodunun sonuçlarını sol alt kısımda yer alan Console bölümünde görebilirsiniz. Aynı şekilde kodu Console bölümüne yazıp Enter tuşuna

bastığınızda yine sonuç alabilirsiniz. Ancak script içerisinde yazılan kodları bir **.R** uzantılı dosya olarak saklama ve daha sonradan bu dosyaya ulaşma şansınız varken, Console ile çalıştırılan kodları bir **.R** dosyası olarak saklama şansınız yoktur. Console tarafındaki sonuçlar geçici olarak ekranda kalır ve R Studio'yu kapatıp açtığınızda tekrar yazdığınız ve çalıştırdığınız kodlara ulaşamayabilirsiniz.

İpucu

Console tarafına yansıyan kodların ve sonuçların farklı formatlarda saklama şansımız vardır. Bunun için **sink** fonksiyonunu araştırmanızı önerebilirim.

1 Summary

In summary, this book has no content whatsoever.

$1 + 1$

[1] 2

2 Introduction

This is a book created from markdown and executable code.

See @knuth84 for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```


Part III

Veri Manipulasyonu



Veri manipölasyonu, veri çerçevesi üzerinde verileri dönüştürmek, filtrelemek, birleştirmek veya yeniden düzenlemek gibi işlemleri içeren önemli bir veri bilimi becerisidir. R programlama dili, veri manipölasyonu için oldukça güçlü ve esnek bir araç sunar. Bu yazıda, R kullanarak veri manipölasyonunu nasıl yapabileceğinizi öğreneceğiz.

Veri manipölasyonu için R’da yaygın olarak kullanılan iki ana kavram, “veri çerçevesi” ve “paketler”dir. Veri çerçevesi, verileri tablo şeklinde düzenleyen ve işleyen veri yapısıdır. R’da veri çerçevesi, **data.frame** türünden nesnelerdir. Veri manipölasyonu için kullanabileceğiniz birçok paket vardır, ancak en yaygın kullanılanlar arasında **dplyr** ve **tidyr** bulunur. Bu paketler, veri manipölasyonunu kolaylaştırmak için bir dizi işlev içerir.

dplyr, RStudio’den Hadley Wickham tarafından geliştirilmiş ve en yaygın veri işleme zorluklarını çözenize yardımcı olan bir veri işleme dil bilgisidir. **dplyr** paketi, **devtools** paketi ve **install_github()** fonksiyonu kullanılarak **CRAN**’dan veya **GitHub**’dan kurulabilir. GitHub deposu genellikle paketteki en son güncellemeleri ve geliştirme sürümünü içerir.

CRAN sayfasından yüklemek için;

```
> install.packages("dplyr")
```

GitHub sayfasından yüklemek için;

```
> install_github("hadley/dplyr")
```

dplyr paketinde sıklıkla kullanılan fonksiyonlar şunlardır:

- **select** : veri çerçevesinden istenilen sütunları seçer.

- **filter** : mantıksal koşullara dayalı olarak bir veri çerçevesinden satırları filtreler.
- **arrange** : satırları sıralar.
- **rename** : sütun isimlerini yeniden isimlendirir.
- **mutate** : yeni değişkenler/sütunlar ekler veya mevcut değişkenleri dönüştürür.
- **summarise/ summarize** : veri çerçevesindeki farklı değişkenlerin özet istatistiklerini oluşturur
- **%>%** (pipe) operatörü birden çok eylemi ardışık düzende zincirleme şekilde birbirine bağlamak için kullanılır.

Veri manipülasyonu ile örnekler bazen küçük veri setleri oluşturulacaktır bazen de 2015 yılı ABD nüfus sayımına ilişkin **counties** veri seti kullanılacaktır. Bu veri setinde eyalet ve şehir detayında nüfus, gelir, ırk, coğrafi yapı, işgücü gibi değişkenler yer almaktadır.

```
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
```

```
counties <- readRDS("datasets/counties.rds")
```

```
# veri setinin yapısı hakkında bilgi sağlar
glimpse(counties)
```

```
Rows: 3,138
```

```
Columns: 40
```

```
$ census_id      <chr> "1001", "1003", "1005", "1007", "1009", "1011", "10~
$ state          <chr> "Alabama", "Alabama", "Alabama", "Alabama", "Alabam~
$ county        <chr> "Autauga", "Baldwin", "Barbour", "Bibb", "Blount", ~
$ region        <chr> "South", "South", "South", "South", "South", "South~
$ metro         <chr> "Metro", "Metro", "Nonmetro", "Metro", "Metro", "No~
```

\$ population	<dbl> 55221, 195121, 26932, 22604, 57710, 10678, 20354, 1~
\$ men	<dbl> 26745, 95314, 14497, 12073, 28512, 5660, 9502, 5627~
\$ women	<dbl> 28476, 99807, 12435, 10531, 29198, 5018, 10852, 603~
\$ hispanic	<dbl> 2.6, 4.5, 4.6, 2.2, 8.6, 4.4, 1.2, 3.5, 0.4, 1.5, 7~
\$ white	<dbl> 75.8, 83.1, 46.2, 74.5, 87.9, 22.2, 53.3, 73.0, 57.~
\$ black	<dbl> 18.5, 9.5, 46.7, 21.4, 1.5, 70.7, 43.8, 20.3, 40.3,~
\$ native	<dbl> 0.4, 0.6, 0.2, 0.4, 0.3, 1.2, 0.1, 0.2, 0.2, 0.6, 0~
\$ asian	<dbl> 1.0, 0.7, 0.4, 0.1, 0.1, 0.2, 0.4, 0.9, 0.8, 0.3, 0~
\$ pacific	<dbl> 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0~
\$ citizens	<dbl> 40725, 147695, 20714, 17495, 42345, 8057, 15581, 88~
\$ income	<dbl> 51281, 50254, 32964, 38678, 45813, 31938, 32229, 41~
\$ income_err	<dbl> 2391, 1263, 2973, 3995, 3141, 5884, 1793, 925, 2949~
\$ income_per_cap	<dbl> 24974, 27317, 16824, 18431, 20532, 17580, 18390, 21~
\$ income_per_cap_err	<dbl> 1080, 711, 798, 1618, 708, 2055, 714, 489, 1366, 15~
\$ poverty	<dbl> 12.9, 13.4, 26.7, 16.8, 16.7, 24.6, 25.4, 20.5, 21.~
\$ child_poverty	<dbl> 18.6, 19.2, 45.3, 27.9, 27.2, 38.4, 39.2, 31.6, 37.~
\$ professional	<dbl> 33.2, 33.1, 26.8, 21.5, 28.5, 18.8, 27.5, 27.3, 23.~
\$ service	<dbl> 17.0, 17.7, 16.1, 17.9, 14.1, 15.0, 16.6, 17.7, 14.~
\$ office	<dbl> 24.2, 27.1, 23.1, 17.8, 23.9, 19.7, 21.9, 24.2, 26.~
\$ construction	<dbl> 8.6, 10.8, 10.8, 19.0, 13.5, 20.1, 10.3, 10.5, 11.5~
\$ production	<dbl> 17.1, 11.2, 23.1, 23.7, 19.9, 26.4, 23.7, 20.4, 24.~
\$ drive	<dbl> 87.5, 84.7, 83.8, 83.2, 84.9, 74.9, 84.5, 85.3, 85.~
\$ carpool	<dbl> 8.8, 8.8, 10.9, 13.5, 11.2, 14.9, 12.4, 9.4, 11.9, ~
\$ transit	<dbl> 0.1, 0.1, 0.4, 0.5, 0.4, 0.7, 0.0, 0.2, 0.2, 0.2, 0~
\$ walk	<dbl> 0.5, 1.0, 1.8, 0.6, 0.9, 5.0, 0.8, 1.2, 0.3, 0.6, 1~
\$ other_transp	<dbl> 1.3, 1.4, 1.5, 1.5, 0.4, 1.7, 0.6, 1.2, 0.4, 0.7, 1~
\$ work_at_home	<dbl> 1.8, 3.9, 1.6, 0.7, 2.3, 2.8, 1.7, 2.7, 2.1, 2.5, 1~
\$ mean_commute	<dbl> 26.5, 26.4, 24.1, 28.8, 34.9, 27.5, 24.6, 24.1, 25.~
\$ employed	<dbl> 23986, 85953, 8597, 8294, 22189, 3865, 7813, 47401,~
\$ private_work	<dbl> 73.6, 81.5, 71.8, 76.8, 82.0, 79.5, 77.4, 74.1, 85.~
\$ public_work	<dbl> 20.9, 12.3, 20.8, 16.1, 13.5, 15.1, 16.2, 20.8, 12.~
\$ self_employed	<dbl> 5.5, 5.8, 7.3, 6.7, 4.2, 5.4, 6.2, 5.0, 2.8, 7.9, 4~
\$ family_work	<dbl> 0.0, 0.4, 0.1, 0.4, 0.4, 0.0, 0.2, 0.1, 0.0, 0.5, 0~
\$ unemployment	<dbl> 7.6, 7.5, 17.6, 8.3, 7.7, 18.0, 10.9, 12.3, 8.9, 7.~
\$ land_area	<dbl> 594.44, 1589.78, 884.88, 622.58, 644.78, 622.81, 77~

3 select

Tabloyu (veri çerçevesi) seçmek ve dönüştürmek için R'da **dplyr** paketinde bulunan **select()** fonksiyonu oldukça kullanışlıdır. Bu fonksiyon, belirli sütunları seçmek veya sütun adlarını değiştirmek için kullanılır. **select()** fonksiyonunu kullanarak veri çerçevesinde sütunları seçme ve dönüştürme işlemlerinin nasıl yapıldığına dair aşağıda örnekler mevcuttur.

i Not

select() fonksiyonu ayrıca sütunları seçerken veya döndürürken bazı özel işlevler de kullanmanıza olanak tanır. Örneğin, **starts_with()**, **ends_with()**, **contains()** gibi işlevleri kullanarak sütun adlarının belirli bir örüntüyü karşılayanları seçebilirsiniz. Bu fonksiyon, veri manipülasyonu işlemlerinde oldukça kullanışlıdır ve veri çerçevelerini istediğiniz şekilde özelleştirmenize yardımcı olur.

```
# library(dplyr)
# counties <- readRDS("datasets/counties.rds")

# belirli sütunları seçmek
counties %>% select(state, county, population, unemployment)
```

```
# A tibble: 3,138 x 4
  state    county    population unemployment
  <chr>   <chr>         <dbl>         <dbl>
1 Alabama Autauga         55221           7.6
2 Alabama Baldwin       195121          7.5
3 Alabama Barbour        26932          17.6
4 Alabama Bibb           22604           8.3
5 Alabama Blount         57710           7.7
6 Alabama Bullock        10678           18
7 Alabama Butler         20354          10.9
8 Alabama Calhoun       116648          12.3
9 Alabama Chambers       34079           8.9
10 Alabama Cherokee     26008           7.9
# i 3,128 more rows
```

```
# belli aralıkta bütün sütunların seçilmesi
counties %>% select(state, county, drive:work_at_home)
```

```
# A tibble: 3,138 x 8
```

	state	county	drive	carpool	transit	walk	other_transp	work_at_home
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Alabama	Autauga	87.5	8.8	0.1	0.5	1.3	1.8
2	Alabama	Baldwin	84.7	8.8	0.1	1	1.4	3.9
3	Alabama	Barbour	83.8	10.9	0.4	1.8	1.5	1.6
4	Alabama	Bibb	83.2	13.5	0.5	0.6	1.5	0.7
5	Alabama	Blount	84.9	11.2	0.4	0.9	0.4	2.3
6	Alabama	Bullock	74.9	14.9	0.7	5	1.7	2.8
7	Alabama	Butler	84.5	12.4	0	0.8	0.6	1.7
8	Alabama	Calhoun	85.3	9.4	0.2	1.2	1.2	2.7
9	Alabama	Chambers	85.1	11.9	0.2	0.3	0.4	2.1
10	Alabama	Cherokee	83.9	12.1	0.2	0.6	0.7	2.5

```
# i 3,128 more rows
```

```
# belirli bir ifadeyi içeren sütunları seçmek
counties %>% select(state, county, contains("employed"))
```

```
# A tibble: 3,138 x 4
```

	state	county	employed	self_employed
	<chr>	<chr>	<dbl>	<dbl>
1	Alabama	Autauga	23986	5.5
2	Alabama	Baldwin	85953	5.8
3	Alabama	Barbour	8597	7.3
4	Alabama	Bibb	8294	6.7
5	Alabama	Blount	22189	4.2
6	Alabama	Bullock	3865	5.4
7	Alabama	Butler	7813	6.2
8	Alabama	Calhoun	47401	5
9	Alabama	Chambers	13689	2.8
10	Alabama	Cherokee	10155	7.9

```
# i 3,128 more rows
```

```
# belirli bir ifade ile başyalan sütunları seçmek
counties %>% select(state, county, starts_with("income"))
```

```
# A tibble: 3,138 x 6
```

	state	county	income	income_err	income_per_cap	income_per_cap_err
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Alabama	Autauga	51281	2391	24974	1080
2	Alabama	Baldwin	50254	1263	27317	711
3	Alabama	Barbour	32964	2973	16824	798
4	Alabama	Bibb	38678	3995	18431	1618
5	Alabama	Blount	45813	3141	20532	708
6	Alabama	Bullock	31938	5884	17580	2055
7	Alabama	Butler	32229	1793	18390	714
8	Alabama	Calhoun	41703	925	21374	489
9	Alabama	Chambers	34177	2949	21071	1366
10	Alabama	Cherokee	36296	1710	21811	1556

```
# i 3,128 more rows
```

```
# belirli bir ifade ile biten sütunları seçmek
counties %>% select(state, county, ends_with("work"))
```

```
# A tibble: 3,138 x 5
```

	state	county	private_work	public_work	family_work
	<chr>	<chr>	<dbl>	<dbl>	<dbl>
1	Alabama	Autauga	73.6	20.9	0
2	Alabama	Baldwin	81.5	12.3	0.4
3	Alabama	Barbour	71.8	20.8	0.1
4	Alabama	Bibb	76.8	16.1	0.4
5	Alabama	Blount	82	13.5	0.4
6	Alabama	Bullock	79.5	15.1	0
7	Alabama	Butler	77.4	16.2	0.2
8	Alabama	Calhoun	74.1	20.8	0.1
9	Alabama	Chambers	85.1	12.1	0
10	Alabama	Cherokee	73.1	18.5	0.5

```
# i 3,128 more rows
```

```
# belirli sütunları hariç tutarak seçmek
counties %>% select(census_id:population,-c(men:land_area))
```

```
# A tibble: 3,138 x 6
```

	census_id	state	county	region	metro	population
	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>
1	1001	Alabama	Autauga	South	Metro	55221

```

2 1003      Alabama Baldwin  South  Metro      195121
3 1005      Alabama Barbour  South  Nonmetro    26932
4 1007      Alabama Bibb     South  Metro      22604
5 1009      Alabama Blount   South  Metro      57710
6 1011      Alabama Bullock   South  Nonmetro    10678
7 1013      Alabama Butler    South  Nonmetro    20354
8 1015      Alabama Calhoun    South  Metro      116648
9 1017      Alabama Chambers  South  Nonmetro    34079
10 1019     Alabama Cherokee  South  Nonmetro    26008
# i 3,128 more rows

```

```

# belirli veri tipindeki sütunları seçmek
counties %>% select(where(is.character))

```

```

# A tibble: 3,138 x 5
  census_id state  county  region metro
  <chr>      <chr>   <chr>   <chr> <chr>
1 1001      Alabama Autauga  South  Metro
2 1003      Alabama Baldwin South  Metro
3 1005      Alabama Barbour South  Nonmetro
4 1007      Alabama Bibb   South  Metro
5 1009      Alabama Blount   South  Metro
6 1011      Alabama Bullock   South  Nonmetro
7 1013      Alabama Butler    South  Nonmetro
8 1015      Alabama Calhoun    South  Metro
9 1017      Alabama Chambers  South  Nonmetro
10 1019     Alabama Cherokee South  Nonmetro
# i 3,128 more rows

```

```

# select ile kolon adı değiştirmek
counties %>% select(census_id, pop = population)

```

```

# A tibble: 3,138 x 2
  census_id  pop
  <chr>      <dbl>
1 1001      55221
2 1003     195121
3 1005     26932
4 1007     22604
5 1009     57710

```


6	1011	10678
7	1013	20354
8	1015	116648
9	1017	34079
10	1019	26008

i 3,128 more rows