

File types in R

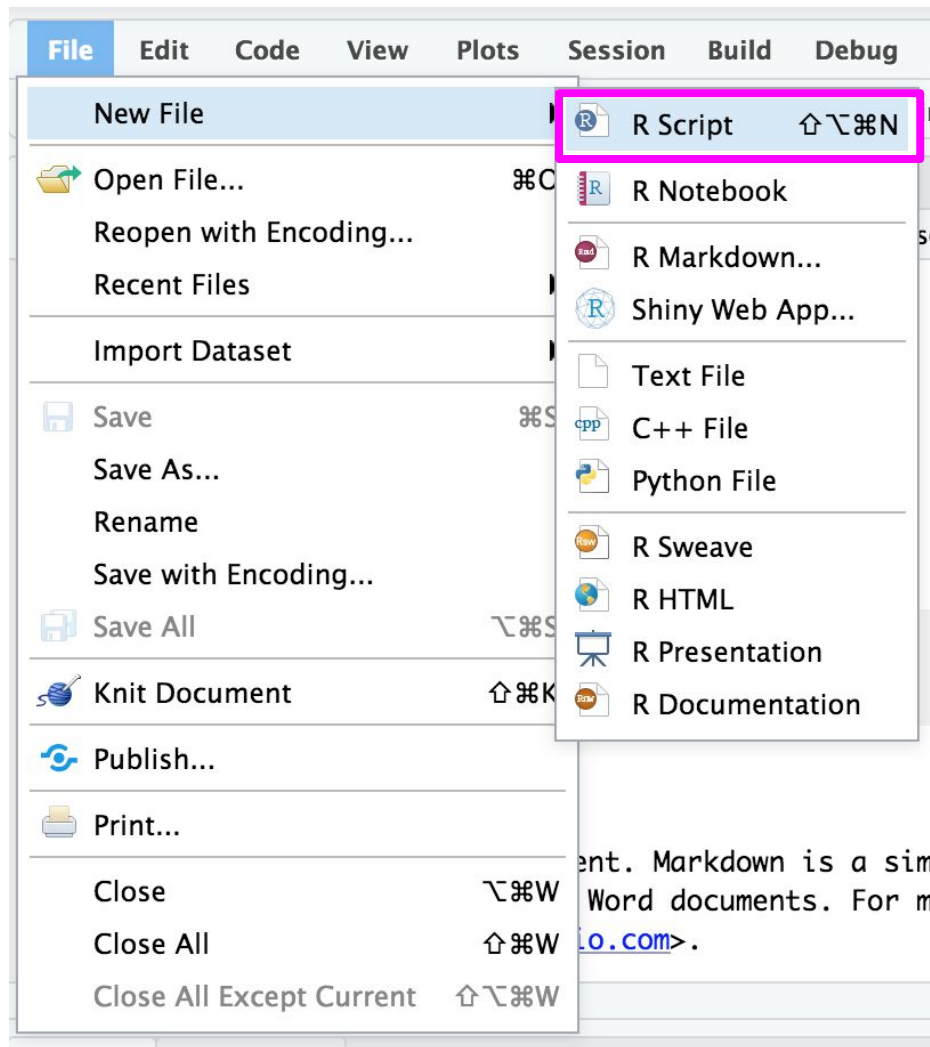


Getting and Cleaning Data

Data in R



- Recommended capitalization
 - Scripting files: .R
 - R Markdown: .Rmd
 - Saved R Objects: .rda
 - Serialized R Objects: .rds
 - R Project files .Rproj



script_test.R File uses .R extension

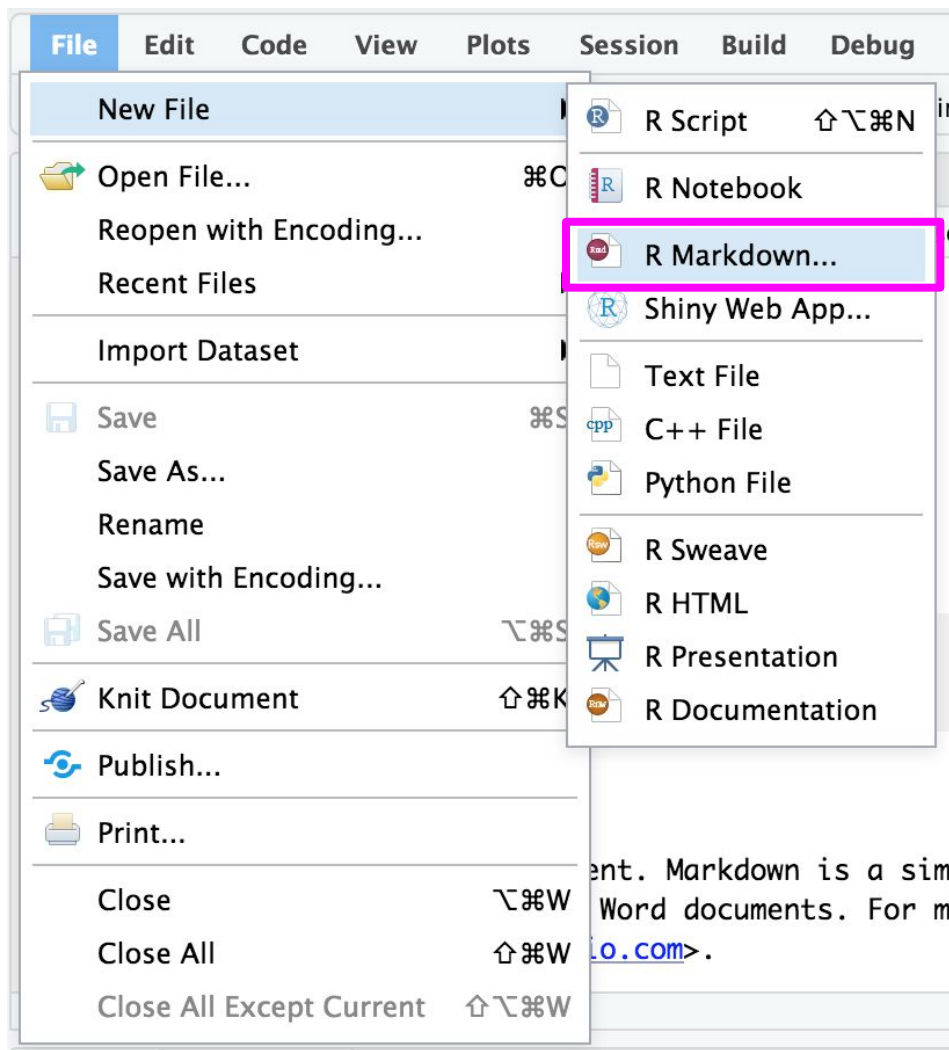
Source on Save Run Source

```
1 ### Scripting in R happens in .R files
2
3 ### They contain R code
4
5 mpg <- mtcars$mpg
6 summary(mpg)
7 hist(mpg)
```

File contains
R code

7:10 (Top Level)

R Script



File uses .Rmd extension

```
1 ---
2 title: "My First R Markdown Document!"
3 author: "Jane Doe"
4 date: "5/31/2018"
5 output:
6   html_document: default
7   pdf_document: default
8 ---
9
10 ```{r setup, include=FALSE}
11 knitr::opts_chunk$set(echo = TRUE)
12 ```
13
14 ## R Markdown
15
16 This is an R Markdown document. Markdown is a simple formatting syntax for
17 authoring HTML, PDF, and MS Word documents. For more details on using R Markdown
18 see <http://rmarkdown.rstudio.com>.
19
20
21
22:1 [C] Chunk 2: cars [R Markdown]
```

File uses
Markdown
formatting and
includes code in
code chunks

script_test.R x

Source on Save Run Source

```
1  ### Scripting in R happens in .R files
2
3  ### Let's create a few objects|
4  mpg <- mtcars$mpg
5  cyl <- mtcars$cyl
6
7
```

mpg and cyl
are objects

3:31 (Top Level) R Script

script_test.R x

Source on Save

Run

Source

```
1 ### Scripting in R happens in .R files
2
3 ### Let's create a few objects
4 mpg <- mtcars$mpg
5 cyl <- mtcars$cyl
6
7
```

3:31 (Top Level) ↕ R Script ↕

Console Terminal x

/cloud/project/ ↕

```
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> |
```

Environment History Connections

Import Dataset

List

Global Environment

Values

cyl	num [1:32]	6 6 4 6 8 6 8 4 4 6 ...
mpg	num [1:32]	21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8...

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

/ > cloud > project

	Name	Size	Modified
	..		
	.Rhistory	0 B	May 30, 2018, 12:34 PM
	project.Rproj	205 B	Jun 12, 2018, 11:18 AM
	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM
	script_test.R	108 B	Jun 12, 2018, 11:47 AM




```
script_test.R x
Source on Save
Run
Source

1 ### Scripting in R happens in .R files
2
3 ### Let's create a few objects
4 mpg <- mtcars$mpg
5 cyl <- mtcars$cyl
6
7 save(mpg, cyl, file="mtcars_objects.rda")
8 |
```

8:1 (Top Level) ↕ R Script ↕

Console Terminal x

```
/cloud/project/ ↗
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> save(mpg, cyl, file="mtcars_objects.rda")
> |
```

save()
generates an
R Data file

Environment History Connections

Import Dataset ↕

Global Environment ▾

Values

cyl	num [1:32]	6 6 4 6 8 6 8 4 4 6 ...
mpg	num [1:32]	21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8...

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More ▾

/ > cloud > project

	▲ Name	Size	Modified
	..		
<input type="checkbox"/>	.Rhistory	0 B	May 30, 2018, 12:34 PM
<input type="checkbox"/>	project.Rproj	205 B	Jun 12, 2018, 11:18 AM
<input type="checkbox"/>	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
<input type="checkbox"/>	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	script_test.R	150 B	Jun 12, 2018, 11:51 AM
<input type="checkbox"/>	mtcars_objects.rda	589 B	Jun 12, 2018, 11:51 AM

script_test.R x

Source on Save

Run

Source

```
1 ### Scripting in R happens in .R files
2
3 ### Let's create a few objects
4 mpg <- mtcars$mpg
5 cyl <- mtcars$cyl
6
7 save(mpg, cyl, file="mtcars_objects.rda")
8 |
```

8:1 (Top Level) ↕ R Script ↕

Console Terminal x

/cloud/project/ ↗

```
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> save(mpg, cyl, file="mtcars_objects.rda")
> |
```

Environment History Connections

Import Dataset ↗

Global Environment ▾

Values

cyl	num [1:32]	6 6 4 6 8 6 8 4 4 6 ...
mpg	num [1:32]	21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8...

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More ▾

/ > / > cloud > project

	▲ Name	Size	Modified
📁	..		
<input type="checkbox"/>	.Rhistory	0 B	May 30, 2018, 12:34 PM
<input type="checkbox"/>	project.Rproj	205 B	Jun 12, 2018, 11:18 AM
<input type="checkbox"/>	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
<input type="checkbox"/>	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	script_test.R	150 B	Jun 12, 2018, 11:51 AM
<input type="checkbox"/>	mtcars_objects.rda	589 B	Jun 12, 2018, 11:51 AM

Confirm Remove Objects



Are you sure you want to remove all objects from the environment? This operation cannot be undone.

☐ Include hidden objects

Yes

No

script_test.R x

Source on Save Run Source

```
1 ### Scripting in R happens in .R files
2
3 ### Let's create a few objects
4 mpg <- mtcars$mpg
5 cyl <- mtcars$cyl
6
7 save(mpg, cyl, file="mtcars_objects.rda")
8 |
```

8:1 (Top Level) R Script

Console Terminal

/cloud/project/ ↗

```
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> save(mpg, cyl, file="mtcars_objects.rda")
> |
```

Environment History Connections

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

/ > / > cloud > project

	▲ Name	Size	Modified
	..		
<input type="checkbox"/>	.Rhistory	0 B	May 30, 2018, 12:34 PM
<input type="checkbox"/>	project.Rproj	205 B	Jun 12, 2018, 11:18 AM
<input type="checkbox"/>	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
<input type="checkbox"/>	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	script_test.R	150 B	Jun 12, 2018, 11:51 AM
<input type="checkbox"/>	mtcars_objects.rda	589 B	Jun 12, 2018, 11:51 AM

```
script_test.R x
Source on Save
Run
Source

1  ### Scripting in R happens in .R files
2
3  ### Let's create a few objects
4  mpg <- mtcars$mpg
5  cyl <- mtcars$cyl
6
7  save(mpg, cyl, file="mtcars_objects.rda")
8
9  load("mtcars_objects.rda")
10 |
```

10:1 (Top Level) R Script

Console Terminal x

```
/cloud/project/
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> save(mpg, cyl, file="mtcars_objects.rda")
> load("mtcars_objects.rda")
> |
```

Environment History Connections

Global Environment

Values

cyl	num [1:32]	6 6 4 6 8 6 8 4 4 6 ...
mpg	num [1:32]	21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8...

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

/ > / > cloud > project

	Name	Size	Modified
↑	..		
<input type="checkbox"/>	.Rhistory	0 B	May 30, 2018, 12:34 PM
<input type="checkbox"/>	project.Rproj	205 B	Jun 12, 2018, 11:18 AM
<input type="checkbox"/>	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
<input type="checkbox"/>	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	script_test.R	178 B	Jun 12, 2018, 11:59 AM
<input type="checkbox"/>	mtcars_objects.rda	589 B	Jun 12, 2018, 11:51 AM

```
script_test.R x
Source on Save Run Source
1 ### Scripting in R happens in .R files
2
3 ### Let's create a few objects
4 mpg <- mtcars$mpg
5 cyl <- mtcars$cyl
6
7 saveRDS(mpg, file="mpg.rds")
8
9
```

8:1 (Top Level) R Script

Console Terminal x

```
/cloud/project/
> ### Let's create a few objects
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> saveRDS(mpg, file="mpg.rds")
>
```

saveRDS ()
generates a
serialized R
object file

Environment History Connections

Global Environment

Values

cyl	num [1:32]	6 6 4 6 8 6 8 4 4 6 ...
mpg	num [1:32]	21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8...

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

/ > / > cloud > project

	Name	Size	Modified
	..		
	.Rhistory	0 B	May 30, 2018, 12:34 PM
	project.Rproj	205 B	Jun 12, 2018, 11:18 AM
	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM
	script_test.R	138 B	Jun 12, 2018, 12:06 PM
	mtcars_objects.rda	589 B	Jun 12, 2018, 11:51 AM
	mpg.rds	142 B	Jun 12, 2018, 12:06 PM

script_test.R x

Source on Save

Run

Source

```
1 ### Scripting in R happens in .R files
2
3 ### Let's create a few objects
4 mpg <- mtcars$mpg
5 cyl <- mtcars$cyl
6
7 saveRDS(mpg, file="mpg.rds")
8
9
```

7:29 (Top Level) R Script

Console

Terminal x

/cloud/project/

```
> ### Let's create a few objects
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> saveRDS(mpg, file="mpg.rds")
>
```

Environment

History

Connections

Import Dataset

Global Environment

Environment is empty

Files

Plots

Packages

Help

Viewer

New Folder

Upload

Delete

Rename

More

/ > cloud > project

	Name	Size	Modified
	..		
<input type="checkbox"/>	.Rhistory	0 B	May 30, 2018, 12:34 PM
<input type="checkbox"/>	project.Rproj	205 B	Jun 12, 2018, 11:18 AM
<input type="checkbox"/>	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
<input type="checkbox"/>	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	script_test.R	138 B	Jun 12, 2018, 12:06 PM
<input type="checkbox"/>	mtcars_objects.rda	589 B	Jun 12, 2018, 11:51 AM
<input type="checkbox"/>	mpg.rds	142 B	Jun 12, 2018, 12:06 PM



```
script_test.R x
Source on Save
Run
Source

1 ### Scripting in R happens in .R files
2
3 ### Let's create a few objects
4 mpg <- mtcars$mpg
5 cyl <- mtcars$cyl
6
7 saveRDS(mpg, file="mpg.rds")
8
9 readRDS("mpg.rds")
10
11 mpg <- readRDS("mpg.rds")
12
13

12:1 (Top Level) R Script
```

Console Terminal

```
/cloud/project/
> ### Let's create a few objects
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> saveRDS(mpg, file="mpg.rds")
> readRDS("mpg.rds")
[1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4
[16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7
[31] 15.0 21.4
> mpg <- readRDS("mpg.rds")
>
```

Serialized R objects
prints contents of
object if not assigned

readRDS() reads
object back in

Environment History Connections

Global Environment

Values

cyl	num [1:32]	6 6 4 6 8 6 8 4 4 6 ...
mpg	num [1:32]	21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8...

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

/ > cloud > project

	Name	Size	Modified
	..		
<input type="checkbox"/>	.Rhistory	0 B	May 30, 2018, 12:34 PM
<input type="checkbox"/>	mpg.rds	142 B	Jun 12, 2018, 1:14 PM
<input type="checkbox"/>	mtcars_objects.rda	589 B	Jun 12, 2018, 11:51 AM
<input type="checkbox"/>	project.Rproj	205 B	Jun 12, 2018, 1:14 PM
<input type="checkbox"/>	script_test.R	185 B	Jun 12, 2018, 1:15 PM
<input type="checkbox"/>	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
<input type="checkbox"/>	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM


```
script_test.R x
Source on Save
Run
Source

1 ### Scripting in R happens in .R files
2
3 ### Let's create a few objects
4 mpg <- mtcars$mpg
5 cyl <- mtcars$cyl
6
7 saveRDS(mpg, file="mpg.rds")
8
9 readRDS("mpg.rds")
10
11 mpg <- readRDS("mpg.rds")
12
13

12:1 (Top Level) R Script
```

Console Terminal

```
/cloud/project/
> ### Let's create a few objects
> mpg <- mtcars$mpg
> cyl <- mtcars$cyl
> saveRDS(mpg, file="mpg.rds")
> readRDS("mpg.rds")
[1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4
[16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7
[31] 15.0 21.4
> mpg <- readRDS("mpg.rds")
>
```

.Rproj file is the
project file extension

Environment History Connections

Global Environment

Values

cyl	num [1:32]	6 6 4 6 8 6 8 4 4 6 ...
mpg	num [1:32]	21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8...

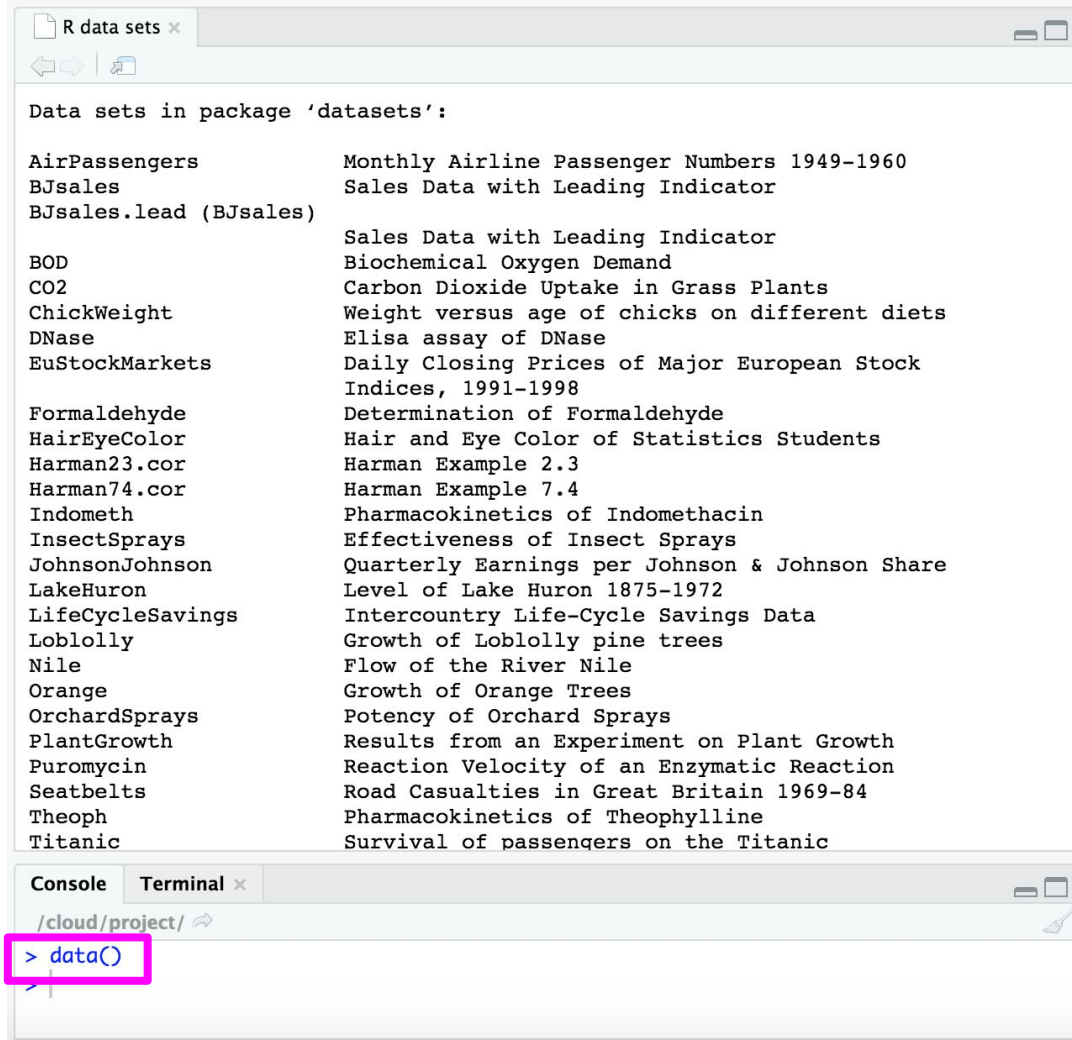
Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

/ > cloud > project

	Name	Size	Modified
	..		
<input type="checkbox"/>	.Rhistory	0 B	May 30, 2018, 12:34 PM
<input type="checkbox"/>	mpg.rds	142 B	Jun 12, 2018, 1:14 PM
<input type="checkbox"/>	mtcars_objects.rda	589 B	Jun 12, 2018, 11:51 AM
<input type="checkbox"/>	project.Rproj	205 B	Jun 12, 2018, 1:14 PM
<input type="checkbox"/>	script_test.R	185 B	Jun 12, 2018, 1:15 PM
<input type="checkbox"/>	test_document.html	741.2 KB	Jun 1, 2018, 2:01 PM
<input type="checkbox"/>	test_document.pdf	161.7 KB	Jun 1, 2018, 2:00 PM
<input type="checkbox"/>	test_document.Rmd	889 B	Jun 1, 2018, 2:01 PM

Opens up a list
of all datasets
available by
default in R



The screenshot shows an R console window with two tabs: 'R data sets x' and 'Terminal x'. The 'R data sets x' tab is active and displays a list of datasets in the 'datasets' package. The list is organized into two columns, with dataset names on the left and their descriptions on the right. The 'Terminal x' tab is also visible, showing the command `> data()` entered at the prompt, which is highlighted with a pink rectangle.

```
R data sets x
Data sets in package 'datasets':

AirPassengers      Monthly Airline Passenger Numbers 1949-1960
BJsales            Sales Data with Leading Indicator
BJsales.lead (BJsales) Sales Data with Leading Indicator
BOD                Biochemical Oxygen Demand
CO2                Carbon Dioxide Uptake in Grass Plants
ChickWeight        Weight versus age of chicks on different diets
DNase              Elisa assay of DNase
EuStockMarkets     Daily Closing Prices of Major European Stock
                  Indices, 1991-1998
Formaldehyde       Determination of Formaldehyde
HairEyeColor       Hair and Eye Color of Statistics Students
Harman23.cor       Harman Example 2.3
Harman74.cor       Harman Example 7.4
Indometh           Pharmacokinetics of Indomethacin
InsectSprays       Effectiveness of Insect Sprays
JohnsonJohnson    Quarterly Earnings per Johnson & Johnson Share
LakeHuron          Level of Lake Huron 1875-1972
LifeCycleSavings   Intercountry Life-Cycle Savings Data
Loblolly           Growth of Loblolly pine trees
Nile               Flow of the River Nile
Orange            Growth of Orange Trees
OrchardSprays      Potency of Orchard Sprays
PlantGrowth        Results from an Experiment on Plant Growth
Puromycin          Reaction Velocity of an Enzymatic Reaction
Seatbelts          Road Casualties in Great Britain 1969-84
Theoph             Pharmacokinetics of Theophylline
Titanic            Survival of passengers on the Titanic

Console   Terminal x
/cloud/project/
> data()
```

R data sets x

← → ↺

Data sets in package 'datasets':

AirPassengers

Monthly Airline Passenger Numbers 1949-1960

BJsales

Sales Data with Leading Indicator

BJsales.lead (BJsales)

Sales Data with Leading Indicator

BOD

Biochemical Oxygen Demand

CO2

Carbon Dioxide Uptake in Grass Plants

ChickWeight

Weight versus age of chicks on different diets

DNase

Elisa assay of DNase

EuStockMarkets

Daily Closing Prices of Major European Stock Indices, 1991-1998

Formaldehyde

Determination of Formaldehyde

HairEyeColor

Hair and Eye Color of Statistics Students

Harman23.cor

Harman Example 2.3

Harman74.cor

Harman Example 7.4

Indometh

Pharmacokinetics of Indomethacin

InsectSprays

Effectiveness of Insect Sprays

JohnsonJohnson

Quarterly Earnings per Johnson & Johnson Share

LakeHuron

Level of Lake Huron 1875-1972

LifeCycleSavings

Intercountry Life-Cycle Savings Data

Loblolly

Growth of Loblolly pine trees

Nile

Flow of the River Nile

Orange

Growth of Orange Trees

Console

Terminal x

/cloud/project/ ↗

> data()

> ?Orange

>

|

Opens up the Help page for this dataset

Environment

History

Connections

Import Dataset ↗

Global Environment ▾

Search

Files

Plots

Packages

Help

Viewer

R: Growth of Orange Trees ▾

Find in Topic

Orange {datasets}

R Documentation

Growth of Orange Trees

Description

The Orange data frame has 35 rows and 3 columns of records of the growth of orange trees.

Usage

Orange

Format

An object of class `c("nfnGroupedData", "nfGroupedData", "groupedData", "data.frame")` containing the following columns:


Tree

an ordered factor indicating the tree on which the measurement is made. The ordering is according to increasing maximum diameter.

age

a numeric vector giving the age of the tree (days since 1968/12/31)


R data sets x




Data sets in package 'ggplot2':

diamonds	Prices of 50,000 round cut diamonds
economics	US economic time series
economics_long	US economic time series
faithfuld	2d density estimate of Old Faithful data
luv_colours	'colors()' in Luv space
midwest	Midwest demographics
mpg	Fuel economy data from 1999 and 2008 for 38 popular models of car
msleep	An updated and expanded version of the mammals sleep dataset
presidential	Terms of 11 presidents from Eisenhower to Obama
seals	Vector field of seal movements
txhousing	Housing sales in TX

ConsoleTerminal x



/cloud/project/ 

```
> data(package="ggplot2")  
> |
```



R data sets x

Data sets in package 'ggplot2':

diamonds	Prices of 50,000 round cut diamonds
economics	US economic time series
economics_long	US economic time series
faithfuld	2d density estimate of Old Faithful data
luv_colours	'colors()' in Luv space
midwest	Midwest demographics
mpg	Fuel economy data from 1999 and 2008 for 38 popular models of car
msleep	An updated and expanded version of the mammals sleep dataset
presidential	Terms of 11 presidents from Eisenhower to Obama
seals	Vector field of seal movements
txhousing	Housing sales in TX

Console Terminal x

```
/cloud/project/
> data(package="ggplot2")
> ?msleep
No documentation for 'msleep' in specified packages and libraries:
you could try '?msleep'

> library(ggplot2)
> ?msleep
```

Package has to be
loaded before dataset
can be used

Once the package is
loaded ?msleep will
open up Help page for
that dataset

Environment History Connections

Import Dataset

Global Environment

Files Plots Packages Help Viewer

R: An updated and expanded version of the mammals sleep dataset Find in Topic

msleep {ggplot2} R Documentation

An updated and expanded version of the mammals sleep dataset

Description

This is an updated and expanded version of the mammals sleep dataset. Updated sleep times and weights were taken from V. M. Savage and G. B. West. A quantitative, theoretical framework for understanding mammalian sleep. Proceedings of the National Academy of Sciences, 104 (3):1051-1056, 2007.

Usage

```
msleep
```

Format

A data frame with 83 rows and 11 variables

name
common name
genus
year

Data in R



- Multiple file types
 - .R
 - .Rmd
 - .rda
 - .rds
 - .Rproj

- `data()`
 - Default datasets available
 - Datasets from packages

Summarizing: File types in R



Getting and Cleaning Data