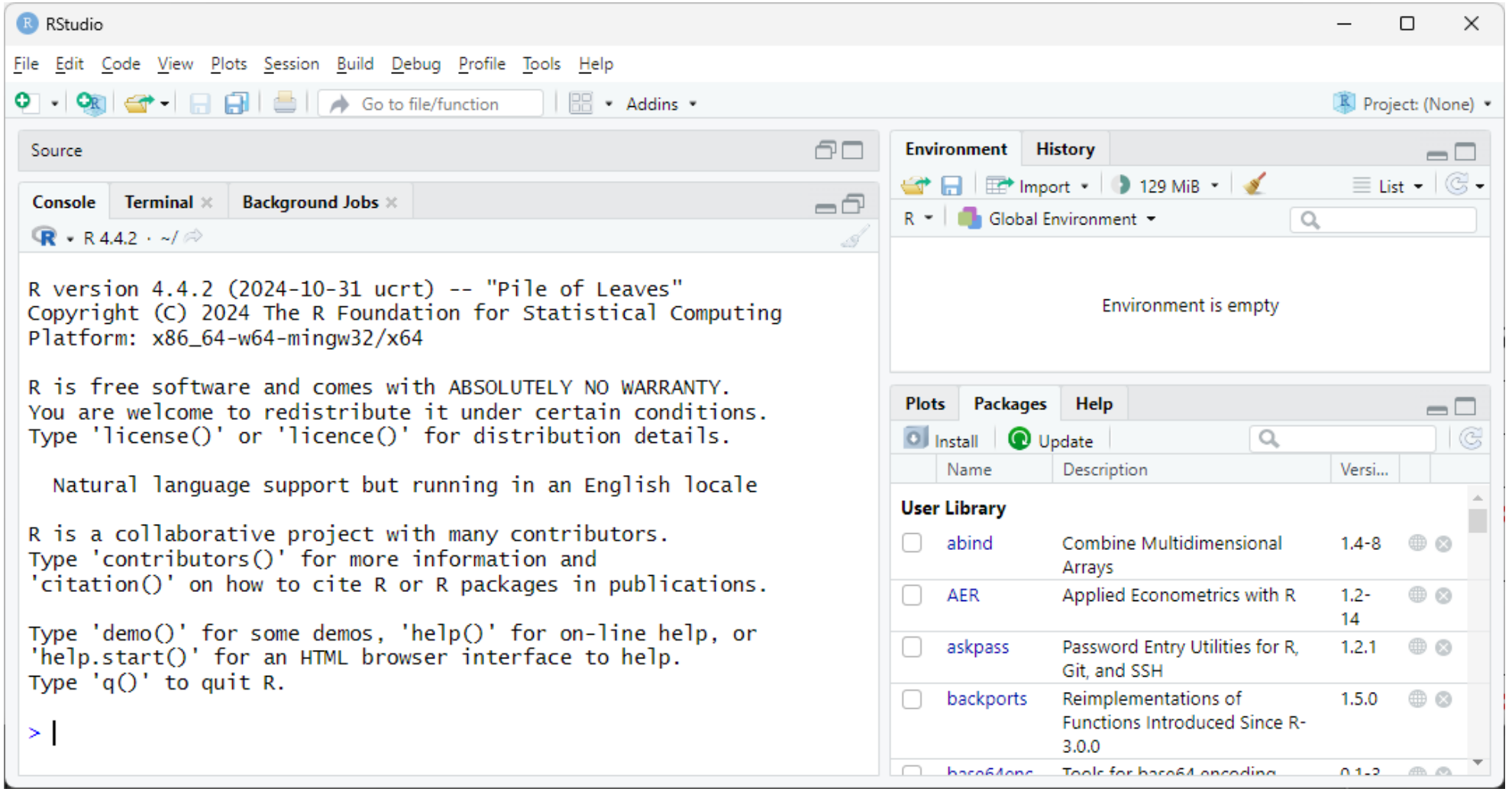
THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

ECON2300 - Introductory Econometrics

Tutorial 1: R and Basic Operations

Tutor: Francisco Tavares Garcia

RStudio IDE


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Source
Console
Terminal
Background Jobs
R 4.4.2 · ~/
R version 4.4.2 (2024-10-31 ucrt) -- "Pile of Leaves"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Import
129 MiB
List
R
Global Environment
Environment is empty
Plots
Packages
Help
Install
Update
Name
Description
Versi...
User Library
abind
Combine Multidimensional Arrays
1.4-8
AER
Applied Econometrics with R
1.2-14
askpass
Password Entry Utilities for R, Git, and SSH
1.2.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.5.0

ECON2300 – Tutorial 01

Install R – 4.4.2

https://cran.r-project.org/

Install RStudio – 2024.12.1+563

https://posit.co/download/rstudio-desktop/

Update all packages –

In RStudio >>

Tools >>

Check for Package Updates >>

Select All >>

Install Updates

Who's your Tutor?

ECON2300 – Tutorial 01

Install R – 4.4.2

https://cran.r-project.org/

Install RStudio – 2024.12.1+563

https://posit.co/download/rstudio-desktop/

Update all packages –

In RStudio >>

Tools >>

Check for Package Updates >>

Select All >>

Install Updates

Who's your Tutor?

Born in 1986 in Ourinhos,
São Paulo state, Brazil

2004 – 2008
Bachelor of Computer Science
2008 – 2012
Supervisor at Procter & Gamble
2009 – 2011
MBA - FGV
2012 – 2018
Built and ran a Hostel
2021 – 2023
Bachelor of Economics – UQ
2024 – 2025
Bachelor of Mathematics - UQ

ECON2300 – Tutorial 01

Install R – 4.4.2

https://cran.r-project.org/

Install RStudio – 2024.12.1+563

https://posit.co/download/rstudio-desktop/

Update all packages –

In RStudio >>

Tools >>

Check for Package Updates >>

Select All >>

Install Updates

Econometrics/Statistics

ECON1310 - Introductory Statistics for Social Sciences

ECON2300 - Introductory Econometrics

ECON2105 - Statistical Theory for Economists

ECON3330 - Econometric Analysis

ECON3350 - Applied Econometrics for Macroeconomics and Finance

ECON3360 - Causal Inference for Microeconometrics

ECON6300 - Advanced Microeconometrics

STAT2003 - Mathematical Probability

STAT2004 - Statistical Modelling & Analysis

STAT3001 - Mathematical Statistics

STAT3004 - Probability Models & Stochastic Processes

ECON3350 – Tutorial 01

Install R – 4.4.2

https://cran.r-project.org/

Install RStudio – 2024.12.1+563

https://posit.co/download/rstudio-desktop/

Update all packages –

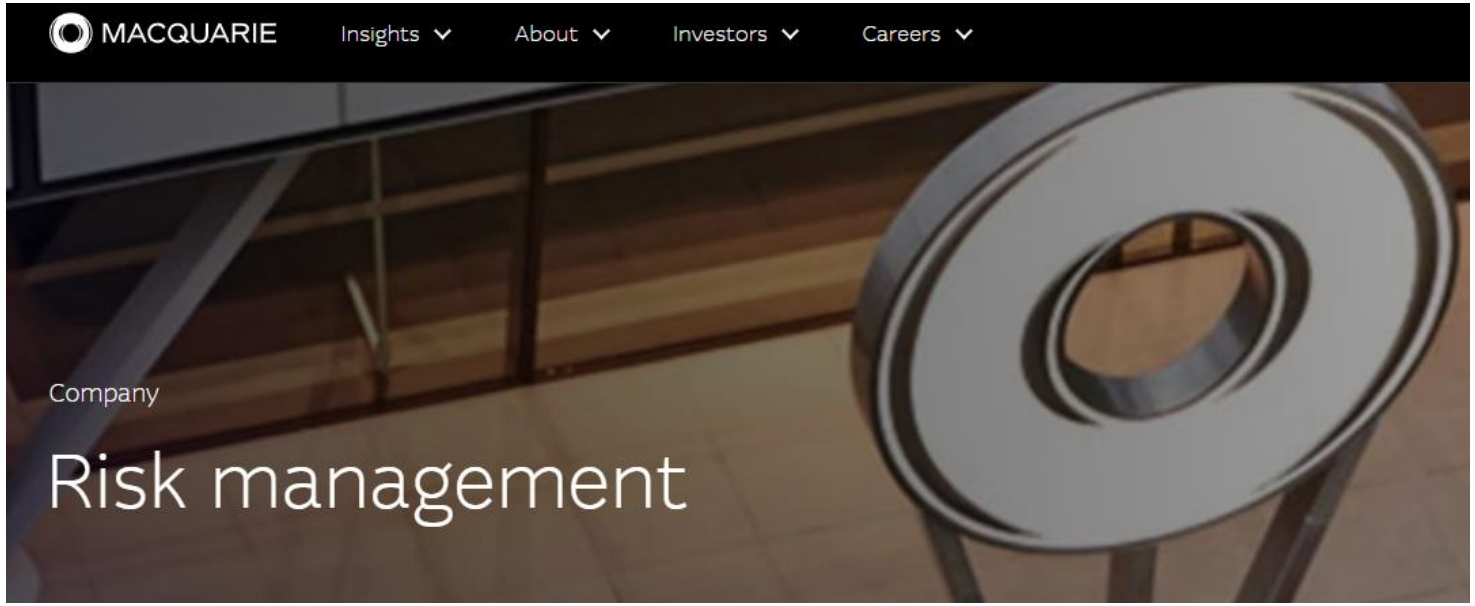
In RStudio >>

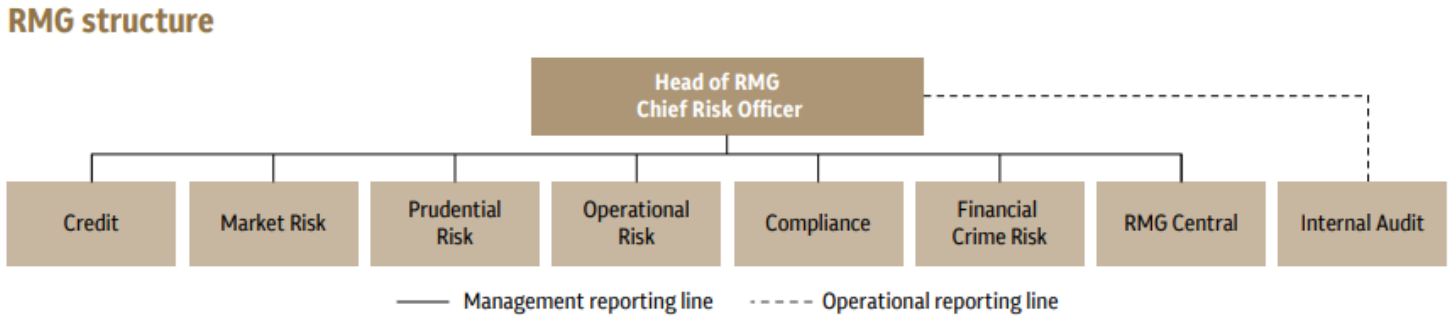
Tools >>

Check for Package Updates >>

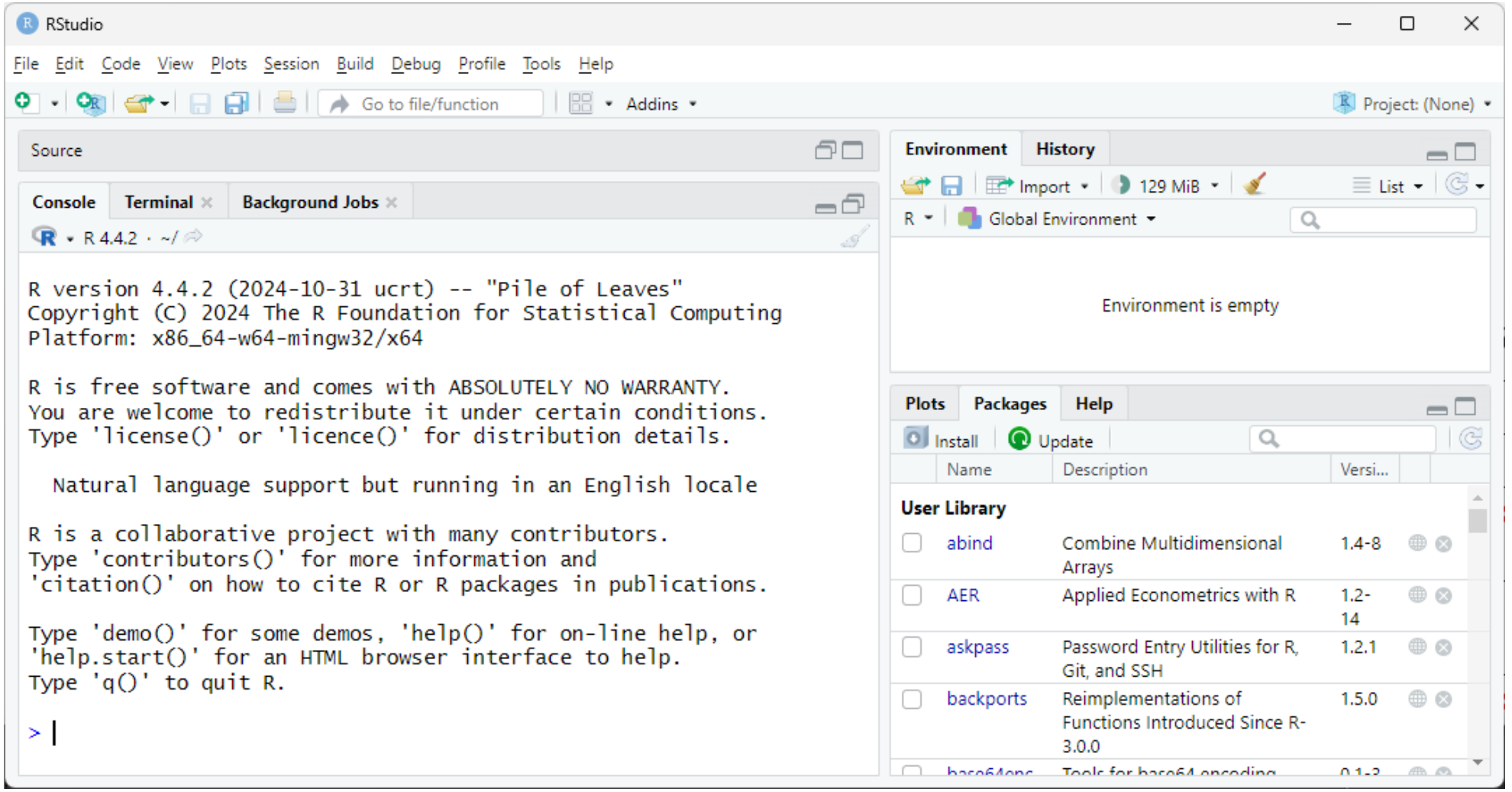
Select All >>

Install Updates





RStudio IDE


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Source
Console
Terminal
Background Jobs
R 4.4.2 · ~/
R version 4.4.2 (2024-10-31 ucrt) -- "Pile of Leaves"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Import
129 MiB
List
R
Global Environment
Environment is empty
Plots
Packages
Help
Install
Update
Name
Description
Versi...
User Library
abind
Combine Multidimensional Arrays
1.4-8
AER
Applied Econometrics with R
1.2-14
askpass
Password Entry Utilities for R, Git, and SSH
1.2.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.5.0

Assessments

Assessment summary

Category	Assessment task	Weight	Due date
Quiz	Problem Solving, Data Analysis and Short Report Online	25% 7 best out of 10	Weeks 3,4,5,6,7,8,9,10,11,12 Online Periodic Assessments Throughout the Semester
Project	Project: Assignment and Brief Research Report Online	25%	17/04/2025 4:00 pm The project can be submitted at anytime before the due date.
Examination	Final Exam Hurdle Identity Verified In-person	50%	End of Semester Exam Period 7/06/2025 - 21/06/2025

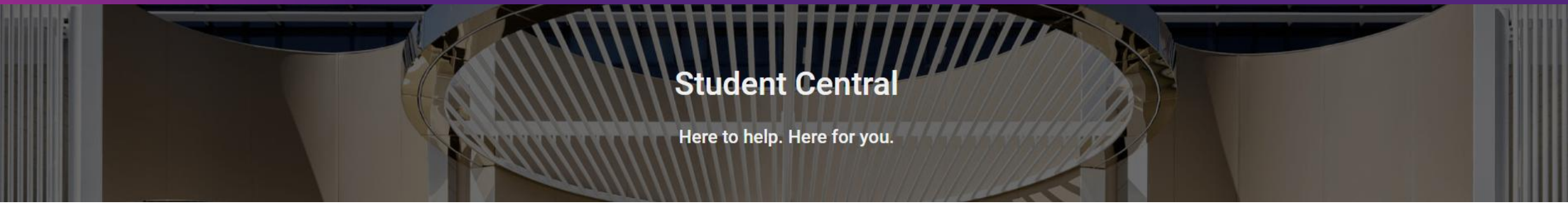
A hurdle is an assessment requirement that must be satisfied in order to receive a specific grade for the course. Check the assessment details for more information about hurdle requirements.

I need HELP!!!

- Ed Discussion (Blackboard/Learn.UQ)
- Consultation Tuesday to Friday!!
- https://www.econometrics-with-r.org/

(good source for R codes – same book)

- ECON2300@uq.edu.au – for general questions
- cml.2300@uq.edu.au – for CML/quizzes



I really need HELP…

https://my.uq.edu.au/contact/student-central

Contact us

Monday-Friday, 8am-5pm:

Email
1300 275 870 (8.30am-5pm)
Building 42, St Lucia

Chat – unavailable

Emergency help

For immediate risk:

UQ Campus Security
07 3365 3333 (24/7)

Off-campus emergency
000 (24/7)

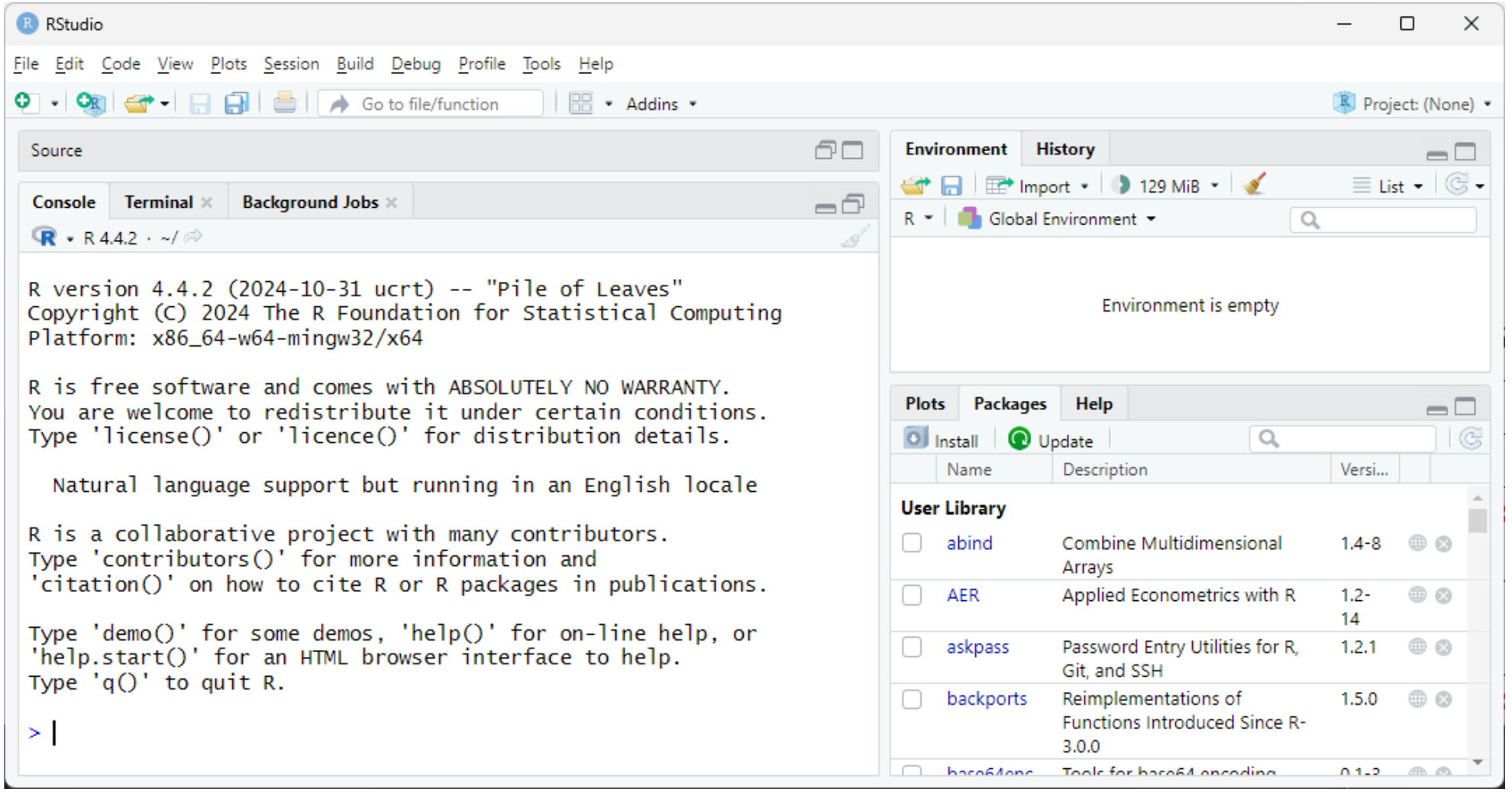
Crisis support

For urgent mental health support:

UQ Counselling and Crisis Line
1300 851 998 (24/7)

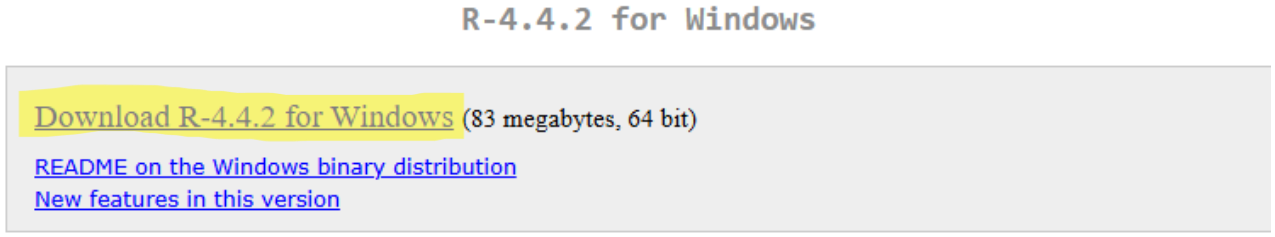
Text a Crisis Counsellor
0488 884 115 (4.30pm-8am)

RStudio IDE


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Source
Console
Terminal
Background Jobs
R 4.4.2 · ~/
R version 4.4.2 (2024-10-31 ucrt) -- "Pile of Leaves"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Import
129 MiB
List
R
Global Environment
Environment is empty
Plots
Packages
Help
Install
Update
Name
Description
Versi...
User Library
abind
Combine Multidimensional Arrays
1.4-8
AER
Applied Econometrics with R
1.2-14
askpass
Password Entry Utilities for R, Git, and SSH
1.2.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.5.0

Installing R (not RStudio yet)

R base distribution – 4.4.2
https://cran.r-project.org/



R-4.4.2 for Windows

Download R-4.4.2 for Windows (83 megabytes, 64 bit)

README on the Windows binary distribution
New features in this version

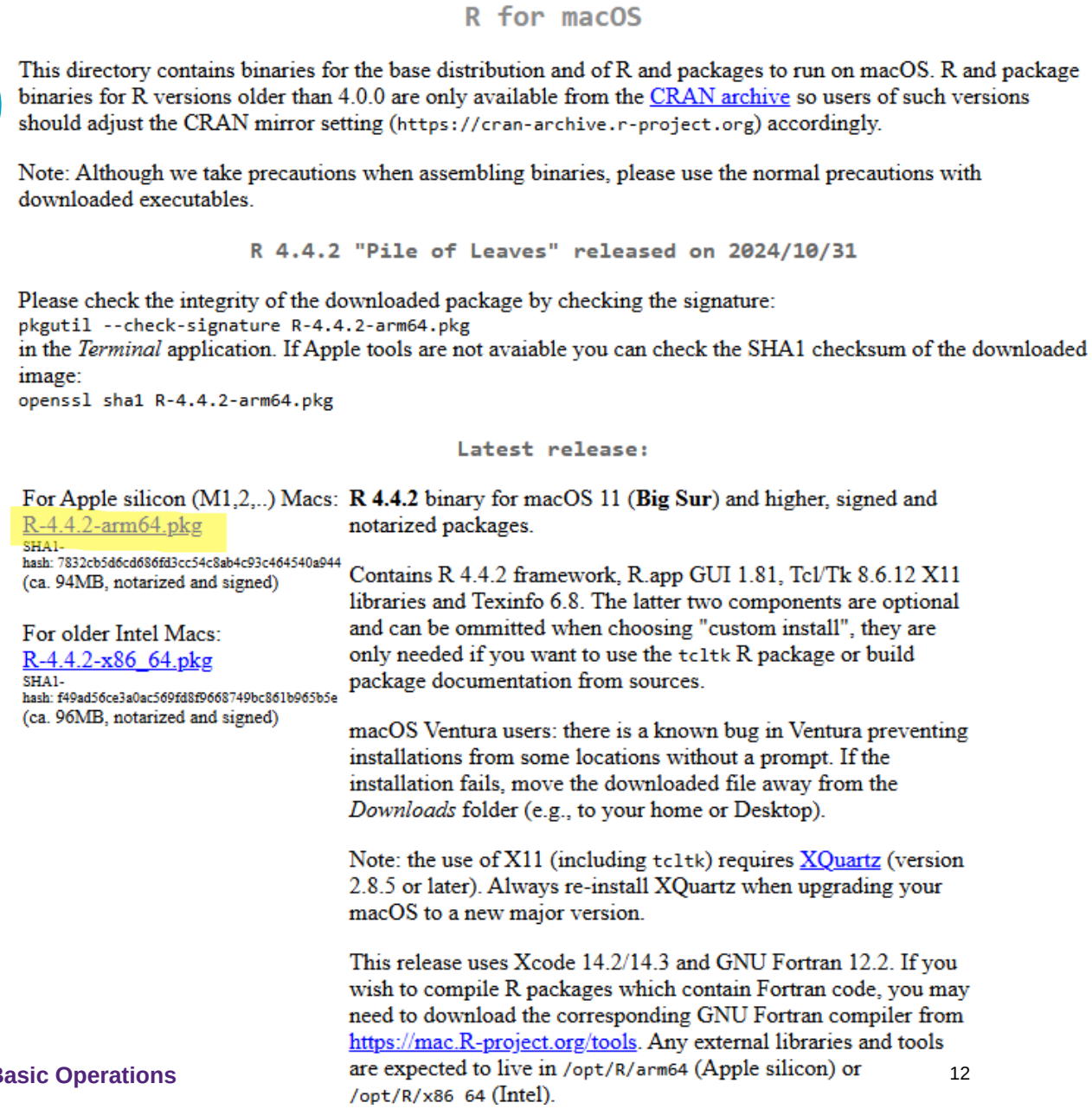
This build requires UCRT, which is part of Windows since Windows 10 and Windows Server 2016. On older systems, UCRT has to be installed manually from here.

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.



R for macOS

This directory contains binaries for the base distribution and of R and packages to run on macOS. R and package binaries for R versions older than 4.0.0 are only available from the CRAN archive so users of such versions should adjust the CRAN mirror setting (`https://cran-archive.r-project.org`) accordingly.

Note: Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

R 4.4.2 "Pile of Leaves" released on 2024/10/31

Please check the integrity of the downloaded package by checking the signature:
`pkgutil --check-signature R-4.4.2-arm64.pkg`
in the *Terminal* application. If Apple tools are not avaiable you can check the SHA1 checksum of the downloaded image:
`openssl sha1 R-4.4.2-arm64.pkg`

Latest release:

For Apple silicon (M1,2,..) Macs:
R-4.4.2-arm64.pkg
SHA1-hash: 7832cb5d6cd686fd3cc54c8ab4c93c464540a944
(ca. 94MB, notarized and signed)

For older Intel Macs:
R-4.4.2-x86_64.pkg
SHA1-hash: f49ad56ce3a0ac569fd8f9668749bc861b965b5e
(ca. 96MB, notarized and signed)

R 4.4.2 binary for macOS 11 (**Big Sur**) and higher, signed and notarized packages.

Contains R 4.4.2 framework, R.app GUI 1.81, Tcl/Tk 8.6.12 X11 libraries and Texinfo 6.8. The latter two components are optional and can be ommitted when choosing "custom install", they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

macOS Ventura users: there is a known bug in Ventura preventing installations from some locations without a prompt. If the installation fails, move the downloaded file away from the *Downloads* folder (e.g., to your home or Desktop).

Note: the use of X11 (including `tcltk`) requires XQuartz (version 2.8.5 or later). Always re-install XQuartz when upgrading your macOS to a new major version.

This release uses Xcode 14.2/14.3 and GNU Fortran 12.2. If you wish to compile R packages which contain Fortran code, you may need to download the corresponding GNU Fortran compiler from https://mac.R-project.org/tools. Any external libraries and tools are expected to live in `/opt/R/arm64` (Apple silicon) or `/opt/R/x86_64` (Intel).

Installing RStudio

Rstudio IDE – 2024.12.1+563
https://posit.co/download/rstudio-desktop/

DOWNLOAD

RStudio IDE

The most popular coding environment for R, built with love by Posit.

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. It includes a console, syntax-highlighting editor that supports direct code execution. It also features tools for plotting, viewing history, debugging and managing your workspace.

RStudio Desktop | RStudio Server

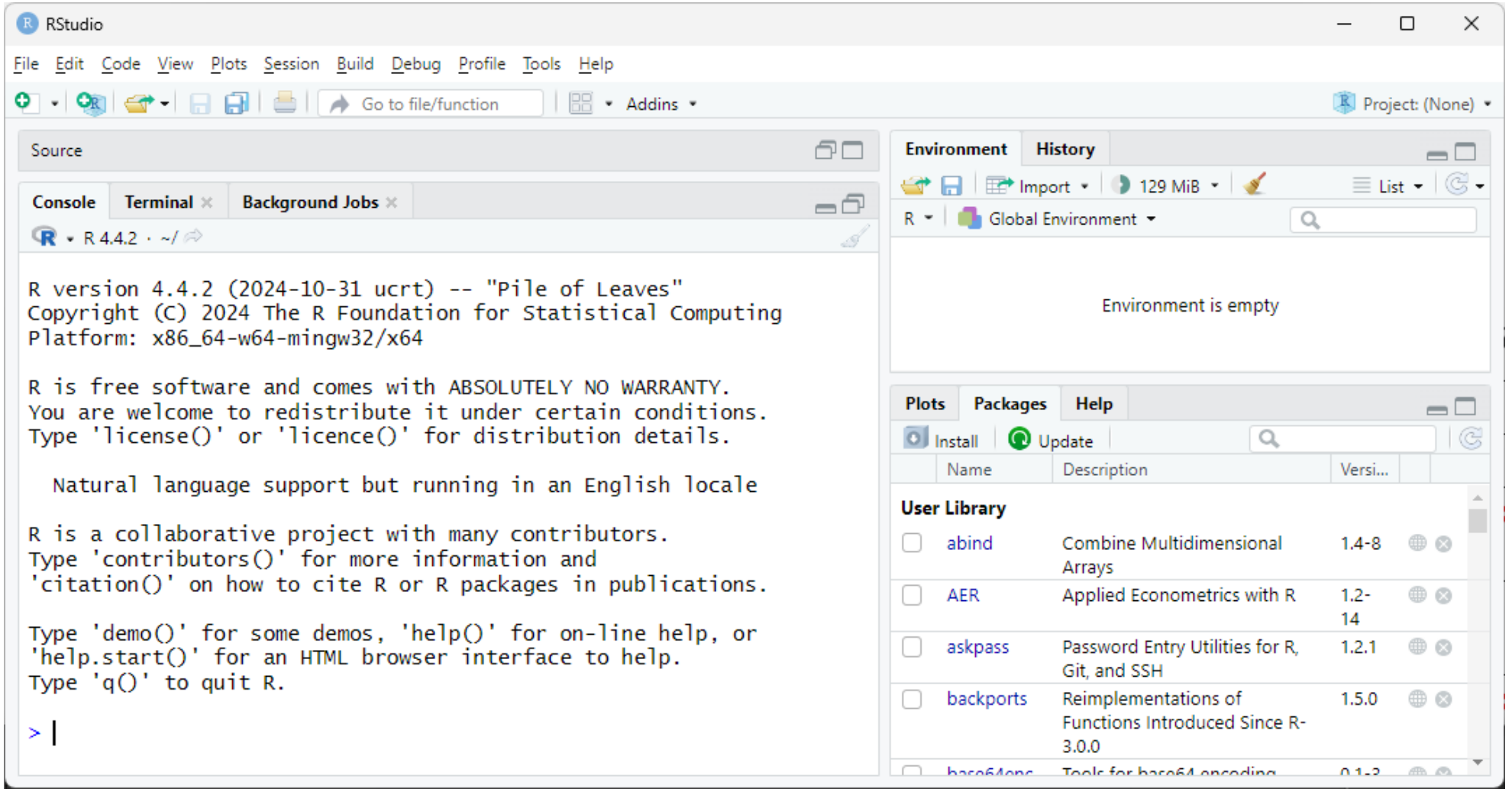
RStudio Desktop

Find out more about RStudio Desktop and RStudio Desktop Pro below.

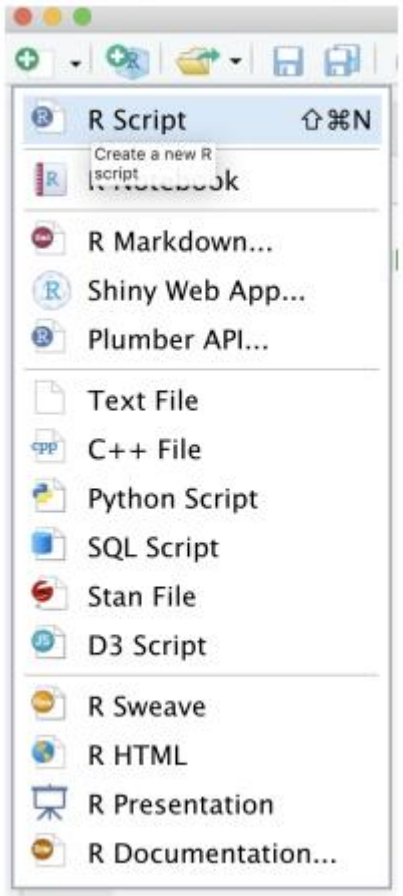
DOWNLOAD RSTUDIO

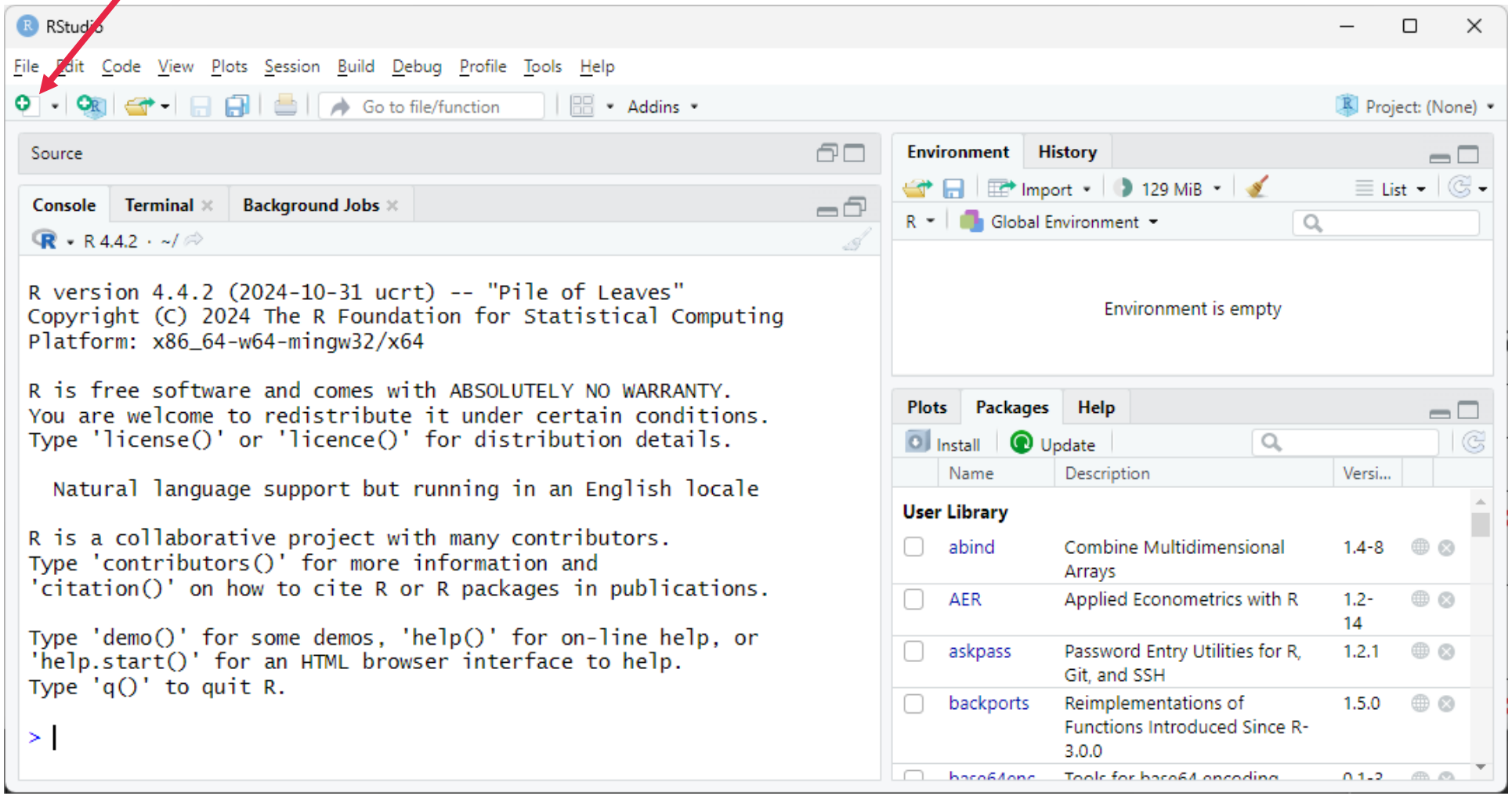
OS	Download	Size	SHA-256
Windows 10/11	RSTUDIO-2024.12.1-563.EXE	265.28 MB	BB369743
macOS 13+	RSTUDIO-2024.12.1-563.DMG	557.15 MB	BE73D3A9
Ubuntu 20/Debian 11	RSTUDIO-2024.12.1-563-AMD64.DEB	203.14 MB	EE259A88
Ubuntu 22/Debian 12	RSTUDIO-2024.12.1-563-AMD64.DEB	203.17 MB	710931EC
Ubuntu 24	RSTUDIO-2024.12.1-563-AMD64.DEB	203.17 MB	710931EC
OpenSUSE 15	RSTUDIO-2024.12.1-563-X86_64.RPM	205.07 MB	9C7E7109

RStudio IDE


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Source
Console
Terminal
Background Jobs
R 4.4.2 · ~/
R version 4.4.2 (2024-10-31 ucrt) -- "Pile of Leaves"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Import
129 MiB
List
R
Global Environment
Environment is empty
Plots
Packages
Help
Install
Update
Name
Description
Versi...
User Library
abind
Combine Multidimensional Arrays
1.4-8
AER
Applied Econometrics with R
1.2-14
askpass
Password Entry Utilities for R, Git, and SSH
1.2.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.5.0

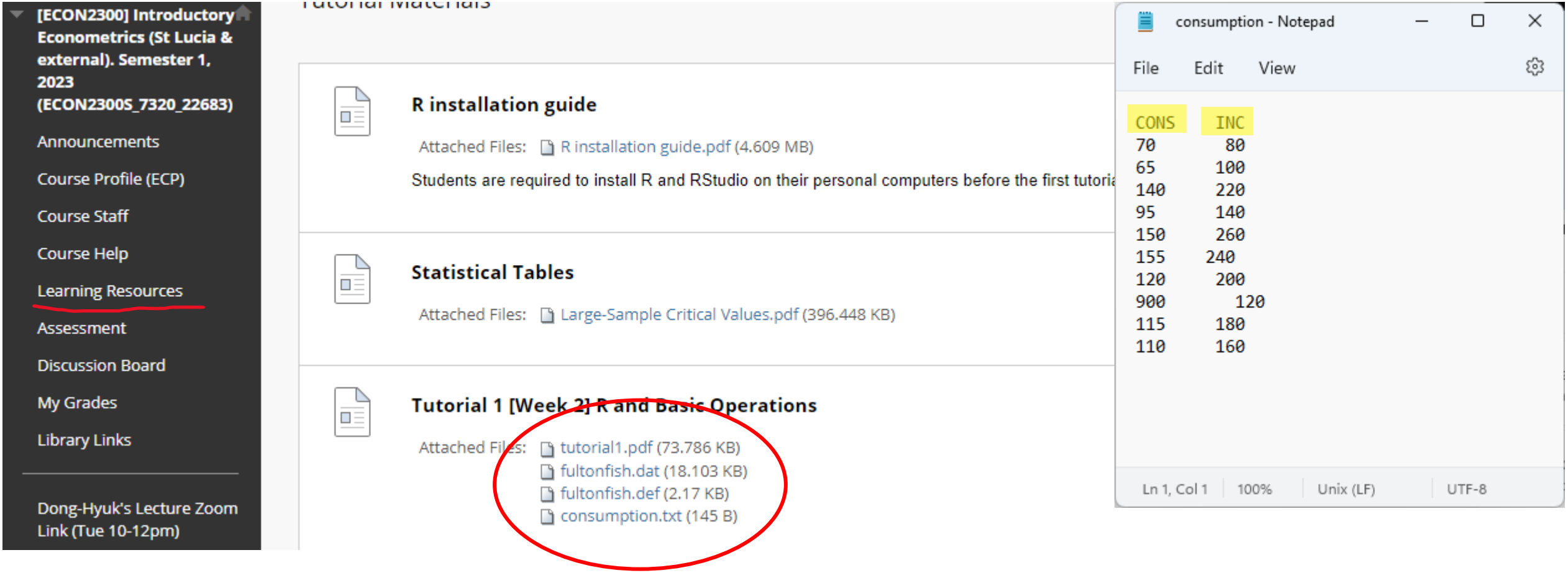
New script


R Script ⇧⌘N
Create a new R script
R Markdown...
Shiny Web App...
Plumber API...
Text File
C++ File
Python Script
SQL Script
Stan File
D3 Script
R Sweave
R HTML
R Presentation
R Documentation...


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Source
Console
Terminal
Background Jobs
R 4.4.2 · ~/
R version 4.4.2 (2024-10-31 ucrt) -- "Pile of Leaves"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Import
129 MiB
List
R
Global Environment
Environment is empty
Plots
Packages
Help
Install
Update
Name
Description
Versi...
User Library
abind
Combine Multidimensional Arrays
1.4-8
AER
Applied Econometrics with R
1.2-14
askpass
Password Entry Utilities for R, Git, and SSH
1.2.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.5.0

1. The text file `consumption.txt` contains observations on the weekly family consumption expenditure (`CONS`) and income (`INC`) for a sample of 10 families.

consumption.txt



Let's download the script for this tutorial.

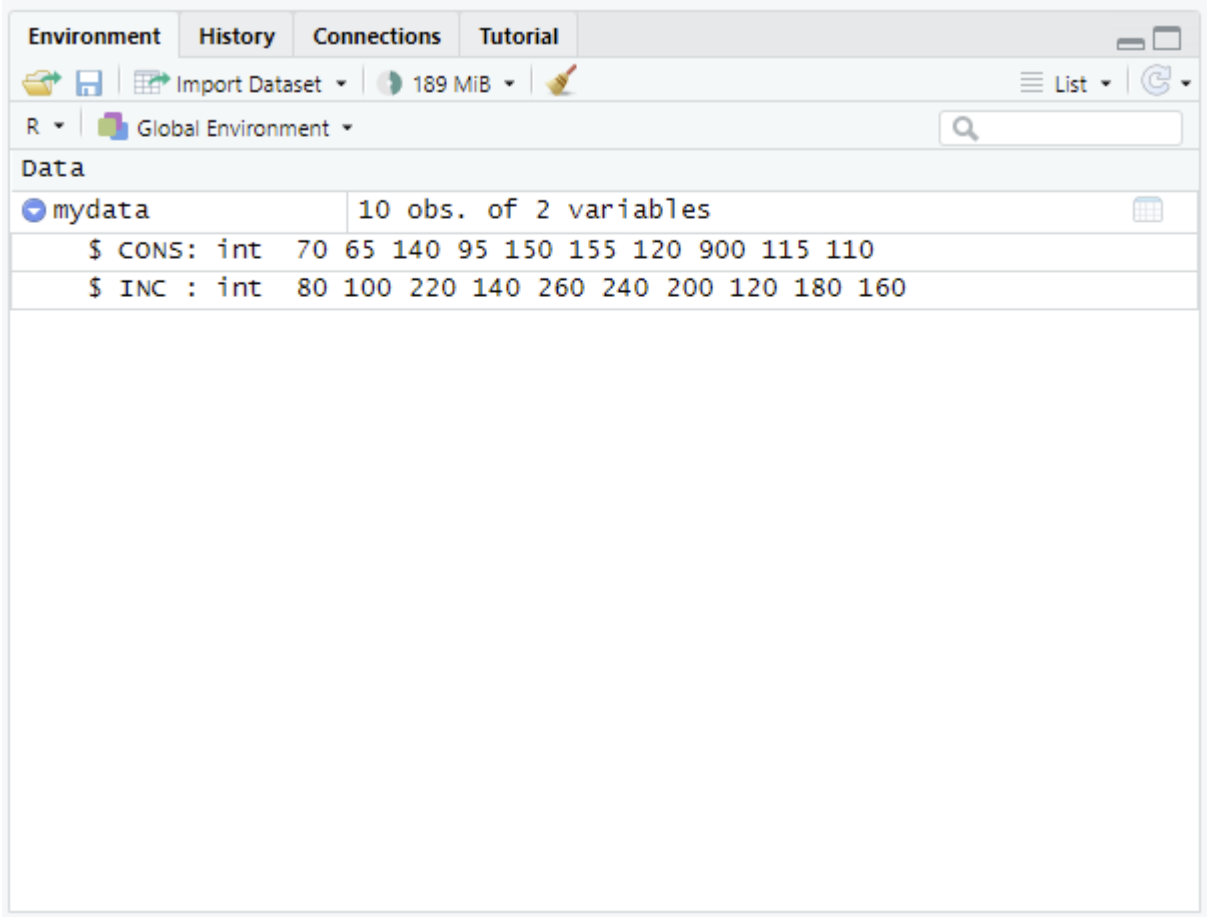
- Copy the code from Codeshare,
  - https://github.com/tavaresgarcia/teaching
- Paste the code in a new script in RStudio,
- Save the script in the same folder as the data.

1. The text file **consumption.txt** contains observations on the weekly family consumption expenditure (CONS) and income (INC) for a sample of 10 families.

   (a) Read the data into R.

Set Work Directory

Session > Set Working Directory > To Source File Location



Solution The data is loaded using the R command **read.delim**.

```
mydata <- read.delim("consumption.txt", header = TRUE,  sep = "")
```

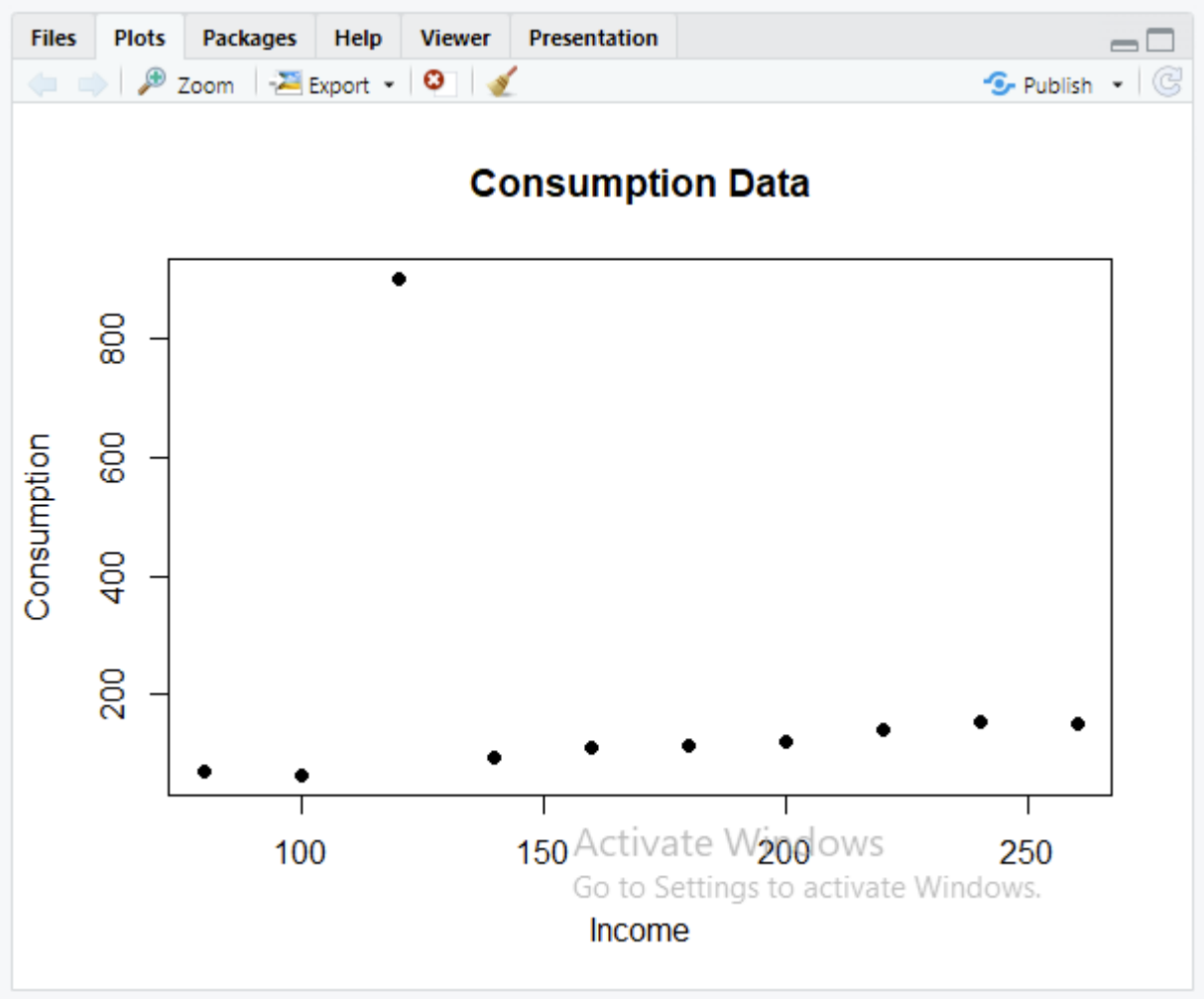
We use the option **header = TRUE** to inform R that the first line contains variable names, and the option **sep = ""** to indicate that the variables are separated by a space. At the same, we create an R variable **mydata** to store the data.

(b) Draw a scatter diagram of CONS against INC.

Solution The simplest way to draw a scatter gram is to `attach` the data and use the `plot` command.

```
attach(mydata)
plot(INC, CONS, main="Consumption Data",
     xlab="Income", ylab="Consumption", pch=19)
```

The command `plot` has several arguments. The first two are the X and Y variables. In addition, it has options to choose a title (`main`) and labels (`xlab` and `ylab`), as well as the point style (`pch`).



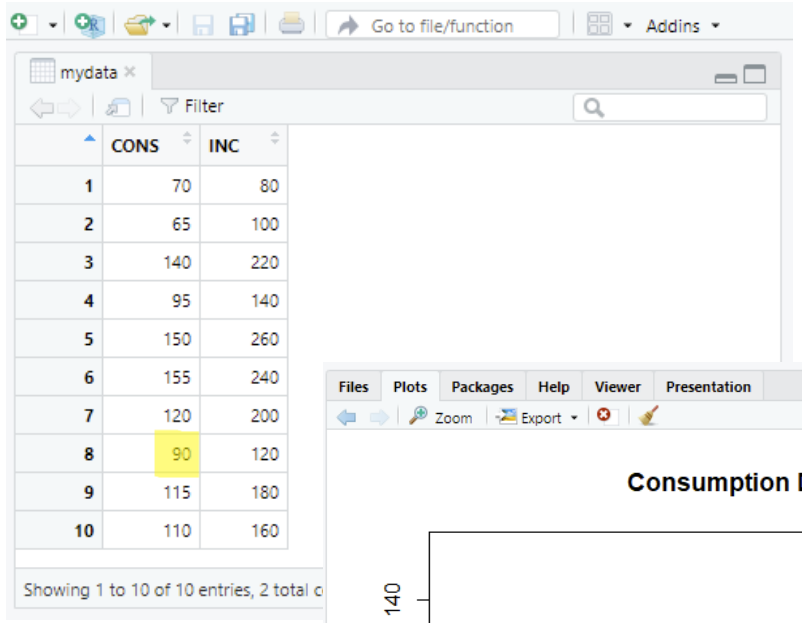
(c) On checking the data, you find that your assistant has recorded the weekly consumption expenditure for Family 8 as \$900 instead of \$90. Correct this error and redraw the scatter diagram.

Solution The data are in the form of a matrix whose (8,1) element has the error, so we assign the correct value to it. Next, we need to "refresh" the data in memory by "detaching" and "attaching" **mydata** again. Once done, redraw the scatter diagram by repeating the command in part (b).

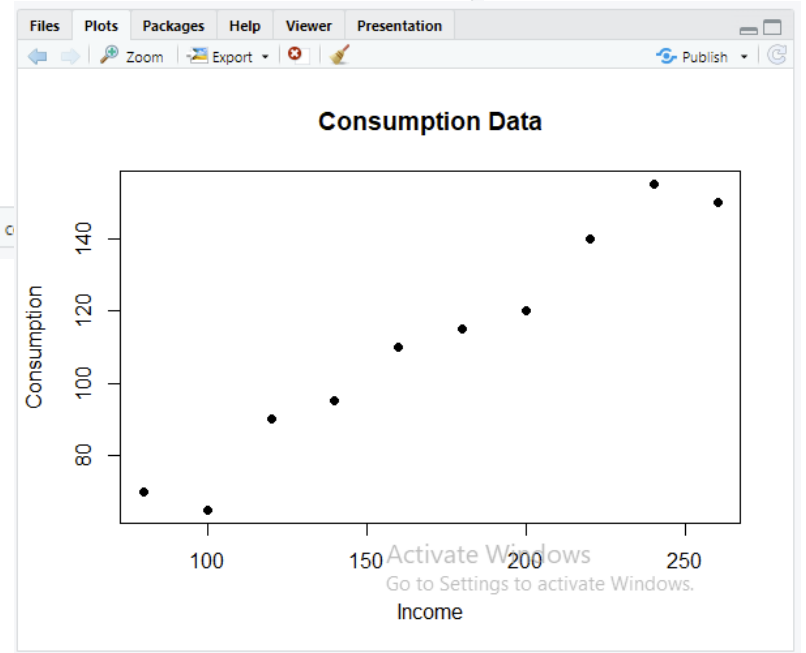
```
mydata[8,1] <- 90

detach(mydata)
attach(mydata)

plot(INC, CONS, main="Consumption Data",
     xlab="Income", ylab="Consumption", pch=19)
```



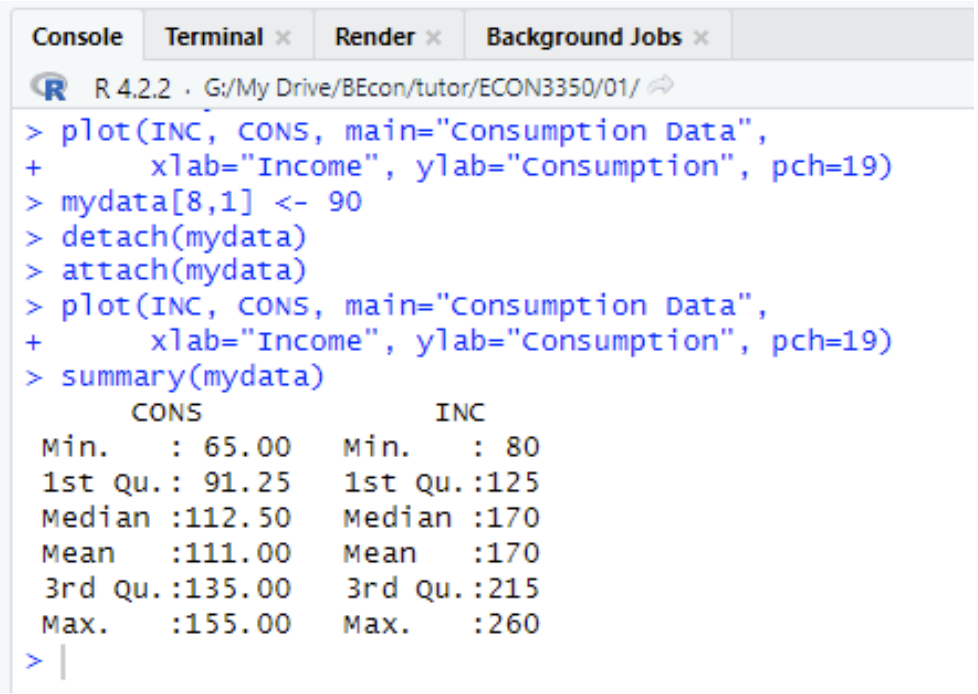
	CONS	INC
1	70	80
2	65	100
3	140	220
4	95	140
5	150	260
6	155	240
7	120	200
8	90	120
9	115	180
10	110	160



(d) Compute the mean, median, maximum and minimum values of `INC` and `CONS`.

Solution All these statistics are neatly summarised by the `summary` command.

```
summary(mydata)
```



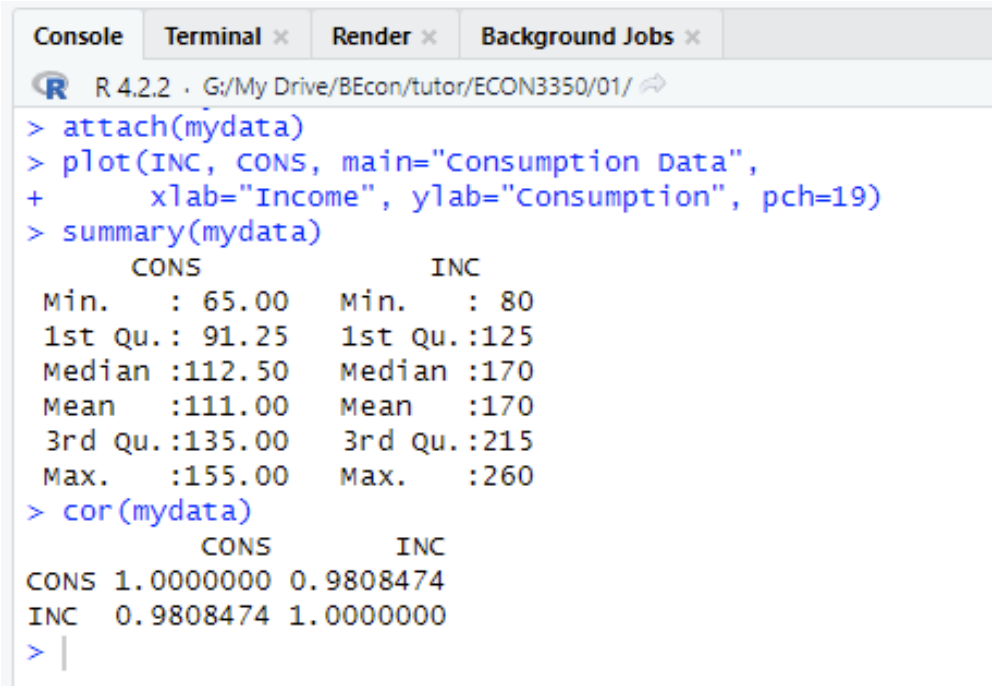
(e) Compute the correlation coefficient between CONS and INC. Comment on the result.

Solution The command `cor` gives a correlation matrix. The off-diagonal elements are correlation coefficients between the variables indicated in the rows and columns.

```
cor(mydata)
```

```
##             CONS       INC
## CONS 1.0000000 0.9808474
## INC  0.9808474 1.0000000
```

In this example, we have only two variables, which gives only one correlation coefficient (0.981). Since the correlation coefficient is close to (positive) one, consumption and income are moving in the same direction and they are closely related.

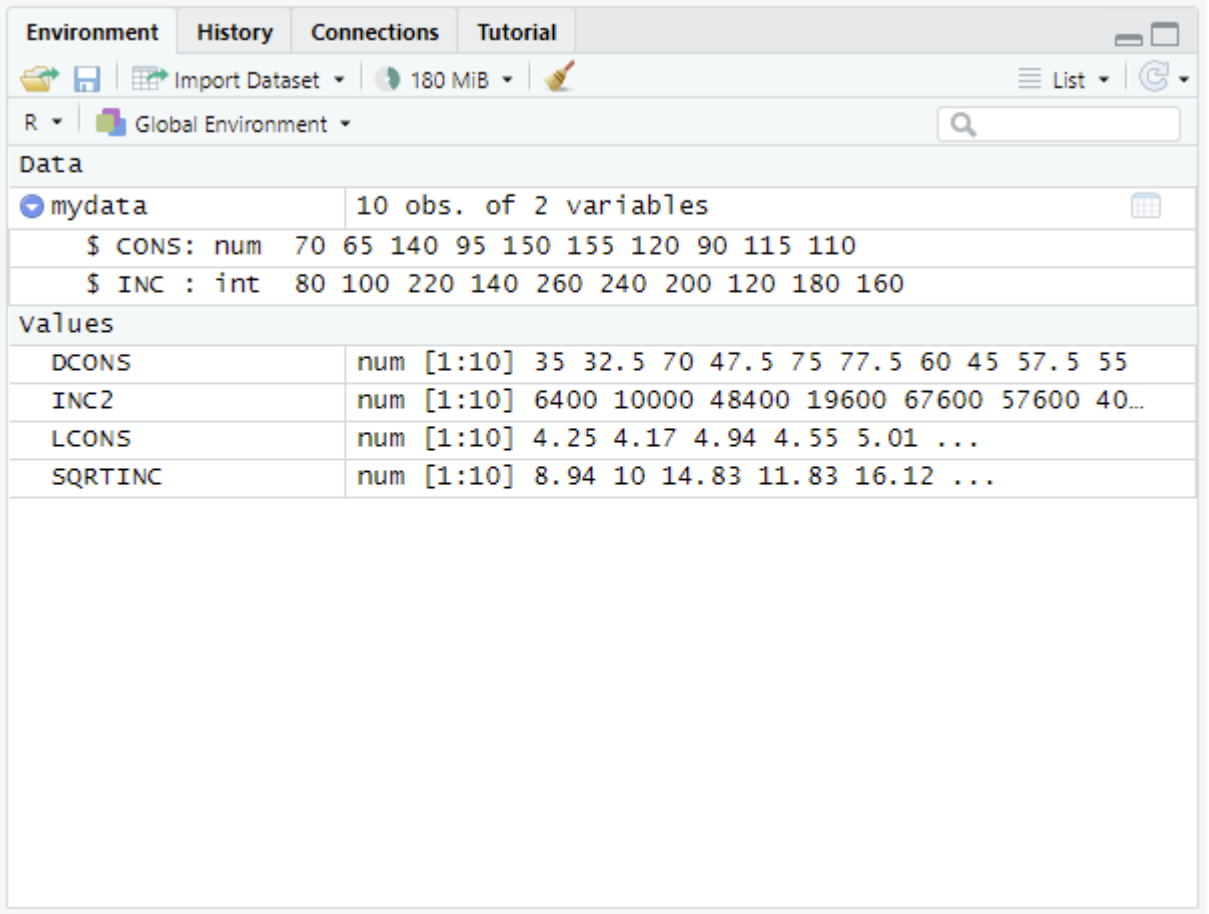


(f) Create the following new variables

$$\texttt{DCONS} = 0.5\texttt{CONS}$$
$$\texttt{LCONS} = \log(\texttt{CONS})$$
$$\texttt{INC2} = \texttt{INC}^2$$
$$\texttt{SQRTINC} = \sqrt{\texttt{INC}}$$

Solution Variables are created using either <- or =. The function log applied the "natural logarithm'' transformation.

```
DCONS <- 0.5 * CONS
LCONS <- log(CONS)
INC2 = INC^2
SQRTINC = sqrt(INC)
```

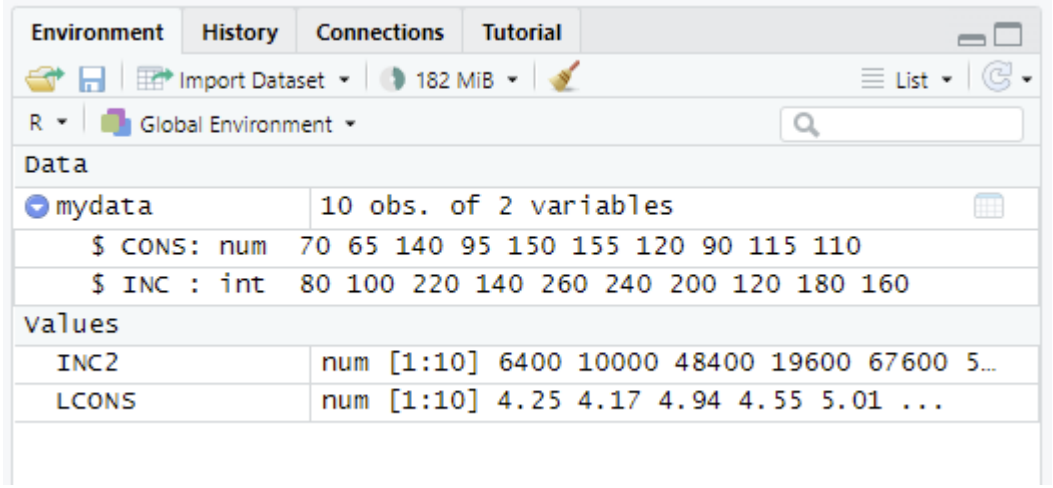


(g) Delete the variable DCONS and SQRTINC.

(h) Delete everything.

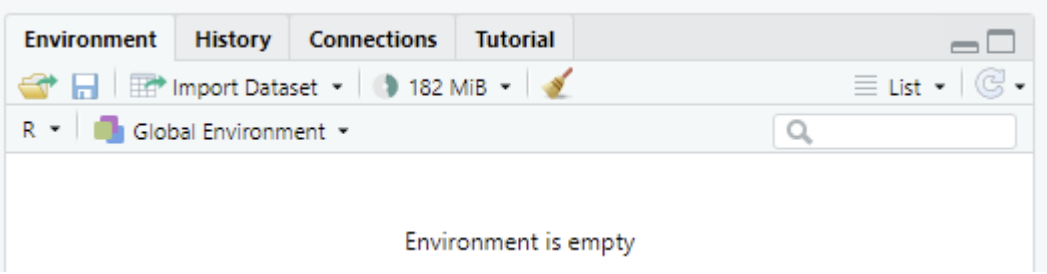
Solution Use the rm command to delete variables.

```
rm(DCONS, SQRTINC)
```



Solution Delete all the variables by passing the output of the ls command to rm.

```
rm(list = ls())
```



2. At the Famous Fulton Fish Market in New York city, sales of whiting (a type of fish) vary from day to day. Over a period of several months, daily quantities sold (in pounds) were observed. These data are in the file `fultonfish.dat`. Description of the data is in the file `fultonfish.def`. Describe the first four columns.

fultonfish.dat

Whiting

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

2. At the Famous Fulton Fish Market in New York city, sales of whiting (a type of fish) vary from day to day. Over a period of several months, daily quantities sold (in pounds) were observed. These data are in the file `fultonfish.dat`. Description of the data is in the file `fultonfish.def`. Describe the first four columns.

fultonfish.dat

fultonfish - Notepad

File Edit View

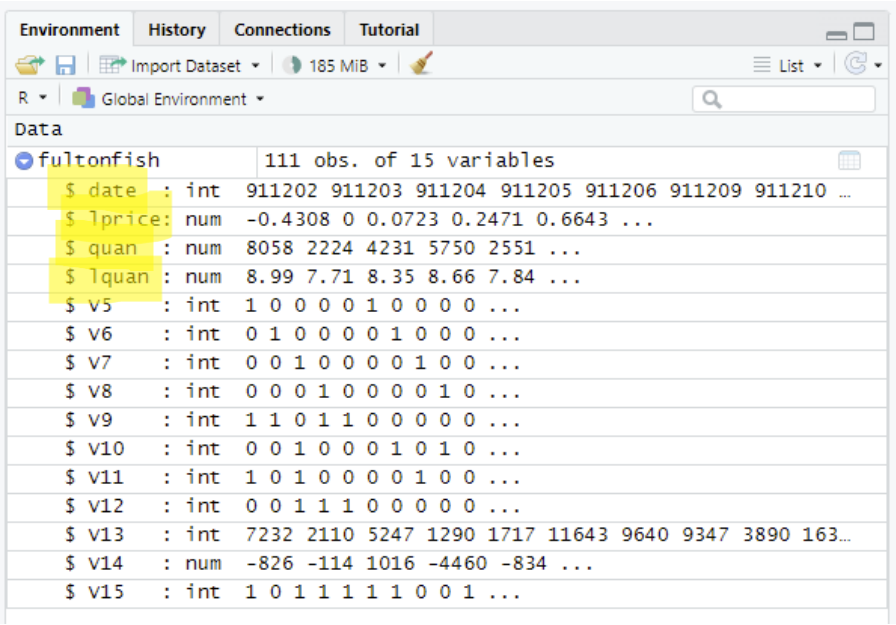
911202	-.4307829	8058.003	8.994421	1	0	0	0	1	0	1	0	7232	-826.0029	1
911203	0	2224.001	7.707063	0	1	0	0	1	0	0	0	2110	-114.0012	0
911204	.0723207	4231.001	8.350194	0	0	1	0	0	1	1	1	5247	1015.999	1
911205	.247139	5749.998	8.656955	0	0	0	1	1	0	0	1	1290	-4459.998	1
911206	.6643268	2551.001	7.844241	0	0	0	0	1	0	0	1	1717	-834.001	1
911209	-.2065143	10952	9.301277	1	0	0	0	0	0	0	0	11643	691.002	1
911210	-.1158318	7485	8.920656	0	1	0	0	0	1	0	0	9640	2155	1
911211	-.2598674	9008.996	9.105979	0	0	1	0	0	0	1	0	9347	338.0039	0
911212	-.1171254	4055	8.307706	0	0	0	1	0	1	0	0	3890	-164.9998	0
911213	-.3420761	9992.003	9.20954	0	0	0	0	0	0	0	0	16318	6325.997	1
911216	-.1255632	5180.002	8.552561	1	0	0	0	1	0	0	1	8725	3544.998	1
911217	.027399	5030	8.523175	0	1	0	0	1	0	0	1	2780	-2250	1
911218	-.0712275	7083	8.865453	0	0	1	0	1	0	0	1	9078	1995	1
911219	.1230601	9762.996	9.186355	0	0	0	1	1	0	0	1	5066	-4696.996	1
911220	.2130932	5999.002	8.699348	0	0	0	0	1	0	0	1	4796	-1203.002	1
911223	-.3172045	12196	9.408863	1	0	0	0	0	1	0	1	13647	1451.003	1
911224	-.1088388	3463.999	8.150179	0	1	0	0	0	1	0	1	1255	-2208.999	1
911226	.2231435	814.9999	6.703188	0	0	0	1	0	1	0	1	1115	300.0001	0
911227	.2464593	6626.999	8.798907	0	0	0	0	0	0	0	1	6887	260.0015	0
911230	-.075431	14260.01	9.565214	1	0	0	0	0	0	1	1	15894	1633.993	1
911231	.2055992	4014.999	8.297792	0	1	0	0	0	0	1	1	5850	1835.001	1
920102	.2188098	4109.001	8.320935	0	0	0	1	0	0	0	0	409	-3700.001	1
920103	.307025	7221.997	8.884887	0	0	0	0	0	0	0	0	7222	.003418	0
920106	.399592	11344	9.336444	1	0	0	0	1	0	0	0	13036	1692.004	1

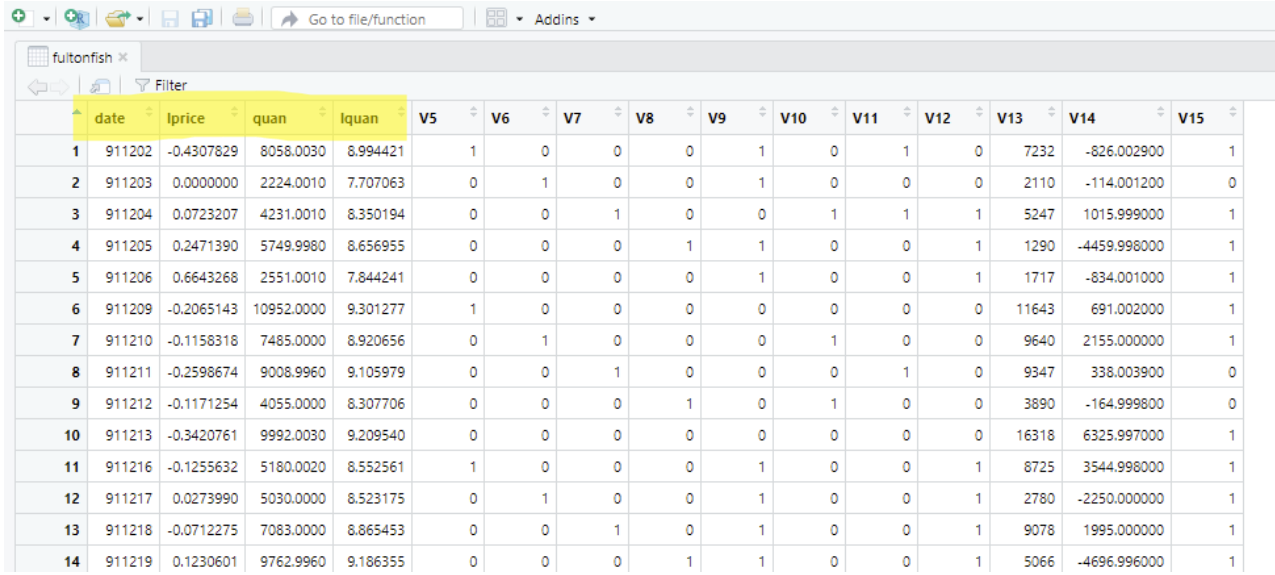
(a) Use R to open the data file and name the series in the first four columns as `date`, `lprice`, `quan` and `lquan`

Solution R assigns variable names `V1`, `V2`, ... when the variables do not have a name. Assign proper names to the first four variables using the command `colnames`.

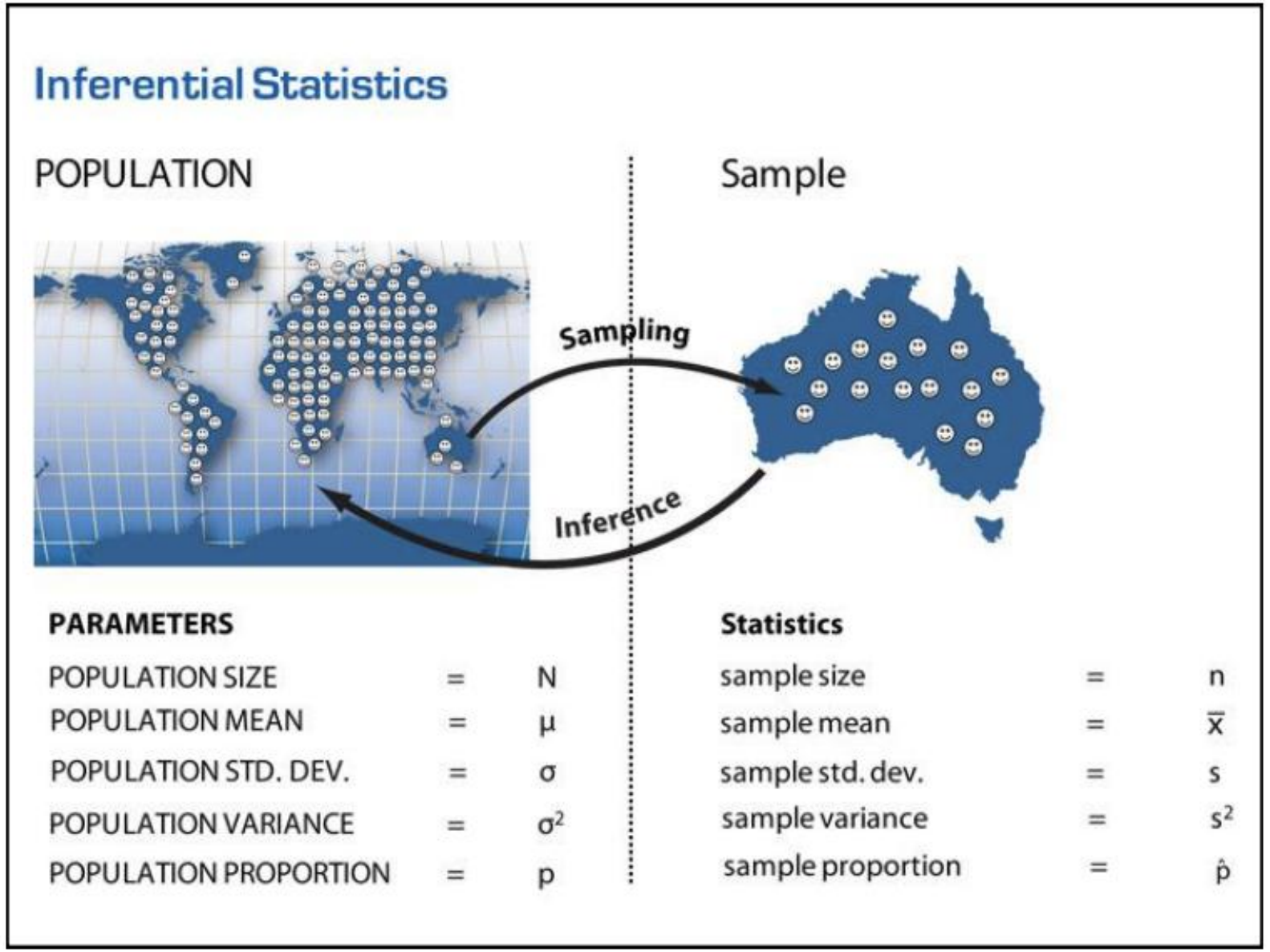
```
fultonfish <- read.delim("fultonfish.dat", header = FALSE,  sep = "")
colnames(fultonfish)[1:4] <- c("date", "lprice", "quan", "lquan")
```

The command `colnames` takes an R object as an argument—in this case `fultonfish`. The range in brackets, [1:4], chooses the columns (from the first to the fourth). The command `c` "concatenates" a list of variables.

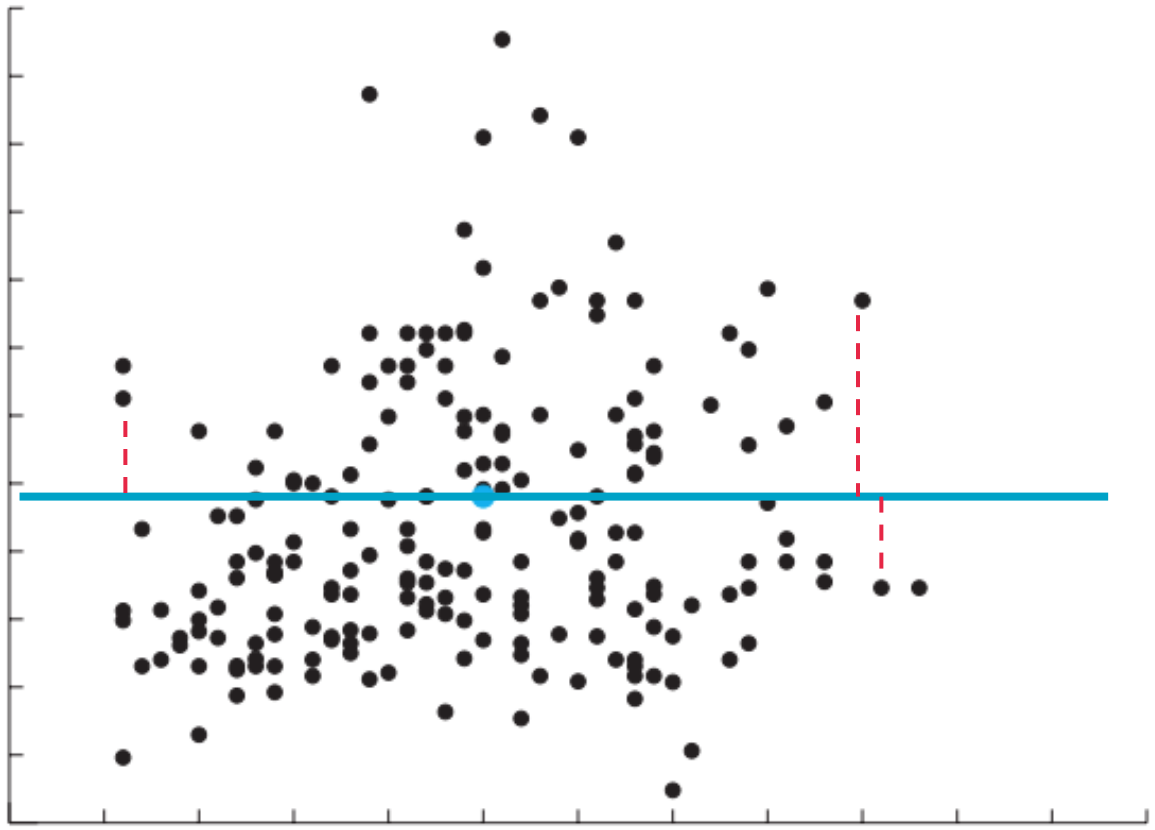




	date	lprice	quan	lquan	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
1	911202	-0.4307829	8058.0030	8.994421	1	0	0	0	1	0	1	0	7232	-826.002900	1
2	911203	0.0000000	2224.0010	7.707063	0	1	0	0	1	0	0	0	2110	-114.001200	0
3	911204	0.0723207	4231.0010	8.350194	0	0	1	0	0	1	1	1	5247	1015.999000	1
4	911205	0.2471390	5749.9980	8.656955	0	0	0	1	1	0	0	1	1290	-4459.998000	1
5	911206	0.6643268	2551.0010	7.844241	0	0	0	0	1	0	0	1	1717	-834.001000	1
6	911209	-0.2065143	10952.0000	9.301277	1	0	0	0	0	0	0	0	11643	691.002000	1
7	911210	-0.1158318	7485.0000	8.920656	0	1	0	0	0	1	0	0	9640	2155.000000	1
8	911211	-0.2598674	9008.9960	9.105979	0	0	1	0	0	0	1	0	9347	338.003900	0
9	911212	-0.1171254	4055.0000	8.307706	0	0	0	1	0	1	0	0	3890	-164.999800	0
10	911213	-0.3420761	9992.0030	9.209540	0	0	0	0	0	0	0	0	16318	6325.997000	1
11	911216	-0.1255632	5180.0020	8.552561	1	0	0	0	1	0	0	1	8725	3544.998000	1
12	911217	0.0273990	5030.0000	8.523175	0	1	0	0	1	0	0	1	2780	-2250.000000	1
13	911218	-0.0712275	7083.0000	8.865453	0	0	1	0	1	0	0	1	9078	1995.000000	1
14	911219	0.1230601	9762.9960	9.186355	0	0	0	1	1	0	0	1	5066	-4696.996000	1


Inferential Statistics
POPULATION
Sample
Sampling
Inference
PARAMETERS
POPULATION SIZE = N
POPULATION MEAN = μ
POPULATION STD. DEV. = σ
POPULATION VARIANCE = σ²
POPULATION PROPORTION = p
Statistics
sample size = n
sample mean = x̄
sample std. dev. = s
sample variance = s²
sample proportion = p̂

(b) Compute the sample mean and standard deviation of the quantity sold (**quan**).



What is the sample mean?
Should I add every observation and divide by the number of observations?

YES!

The bar represents mean →

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)$$

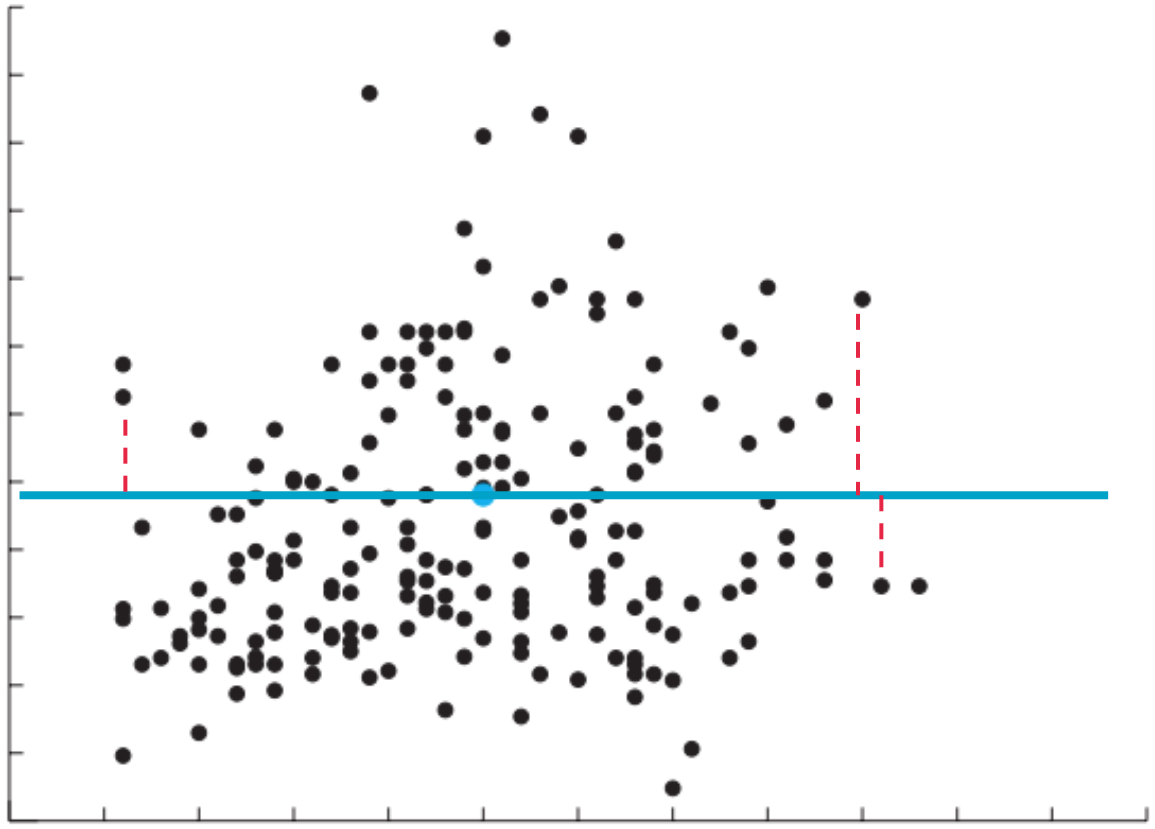
$$\overline{Y} \xrightarrow{p} \mu_Y.$$

What is the standard deviation?
From ECON1310, you might remember that standard deviation = σ and σ^2 = variance. So $\sigma = \sqrt{variance}$.

$$\mathrm{Var}(Y) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu_Y)^2$$

$$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

(b) Compute the sample mean and standard deviation of the quantity sold (**quan**).



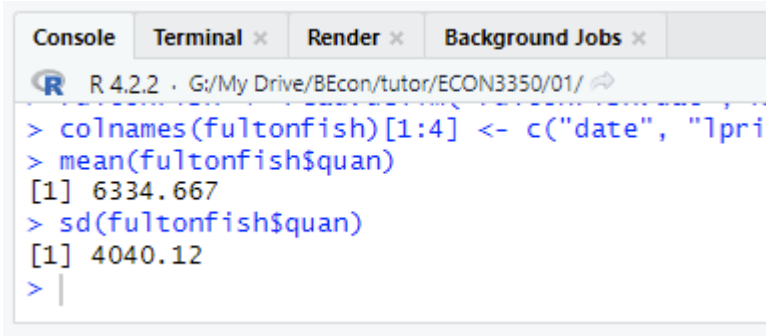
Solution This is straightforward using commands **mean** and **sd**.

```
mean(fultonfish$quan)
```

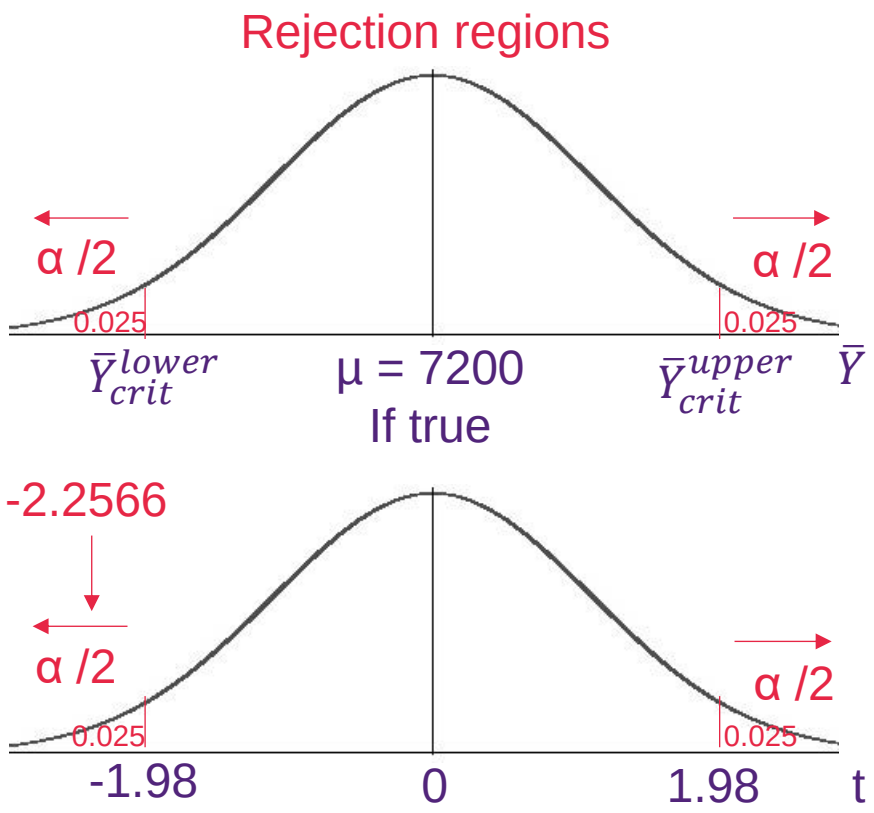
```
## [1] 6334.667
```

```
sd(fultonfish$quan)
```

```
## [1] 4040.12
```



(c) Test the null hypothesis that the mean quantity sold is equal to 7,200 pounds a day at the 5% level of significance.



Step 1: State H_0 and H_1
H_0: μ = 7,200
H_1: μ ≠ 7,200

Step 2: Decision rule
Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2,n-1} = t_{0.025,110} = 1.98$

Step 3: Calculate t_{calc}

$$t_{calc} = \frac{\bar{Y} - \mu}{s_{\bar{Y}}} = \frac{\bar{Y} - \mu}{\frac{s}{\sqrt{n}}} = \frac{6334.67 - 7200}{\frac{4040.12}{\sqrt{111}}} = -2.2566$$

Step 4: Make a decision
$|t_{calc}| > t_{crit} \rightarrow |-2.26| > 1.98 \rightarrow$ Reject H_0.

Step 5: Conclusion
There is sufficient evidence to suggest that the mean quantity sold is not equal to 7,200 pounds a day at the 5% level of significance.

Five Steps for Hypothesis Testing.

1. State H_0 and H_1
2. State the decision rule for the appropriate test statistic and sampling distribution
3. Calculate the test statistic
4. Make a decision (reject H_0 or do not reject H_0)
5. State a conclusion

Note:
steps 1 and 2 are prior to any sample information.

26

(c) Test the null hypothesis that the mean quantity sold is equal to 7,200 pounds a day at the 5% level of significance.

Solution This is straightforward using the command `t.test`.

```
t.test(fultonfish$quan, mu = 7200)

##
##  One Sample t-test
##
## data:  fultonfish$quan
## t = -2.2566, df = 110, p-value = 0.02601
## alternative hypothesis: true mean is not equal to 7200
## 95 percent confidence interval:
##  5574.717 7094.617
## sample estimates:
## mean of x
##  6334.667
```

Five Steps for Hypothesis Testing.

1. State H_0 and H_1
2. State the decision rule for the appropriate test statistic and sampling distribution
3. Calculate the test statistic
4. Make a decision (reject H_0 or do not reject H_0)
5. State a conclusion

Note:
steps 1 and 2 are prior to any sample information.

26

Step 1: State H_0 and H_1
H_0: μ = 7,200
H_1: μ ≠ 7,200

Step 2: Decision rule
Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2,n-1} = t_{0.025,110} = 1.98$

Step 3: Calculate t_{calc}

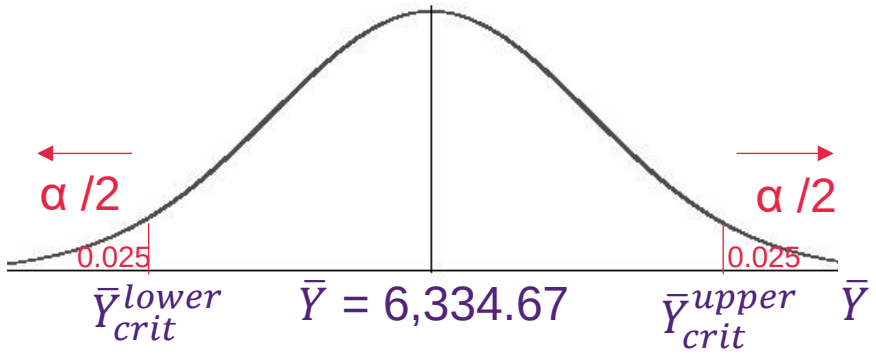
$$t_{calc} = \frac{\bar{Y} - \mu}{s_{\bar{Y}}} = \frac{\bar{Y} - \mu}{\frac{s}{\sqrt{n}}} = \frac{6334.67 - 7200}{\frac{4040.12}{\sqrt{111}}} = -2.2566$$

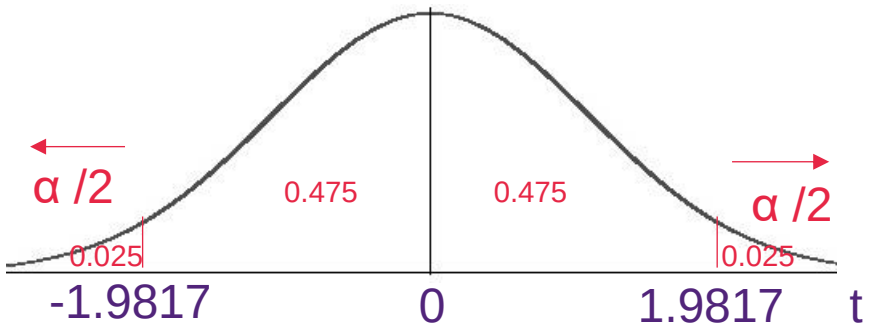
Step 4: Make a decision
$|t_{calc}| > t_{crit}$ → |-2.26| > 1.98 → Reject H_0.

Step 5: Conclusion
There is sufficient evidence to suggest that the mean quantity sold is not equal to 7,200 pounds a day at the 5% level of significance.

(d) Construct the 95% confidence interval for part (c)

Confidence interval



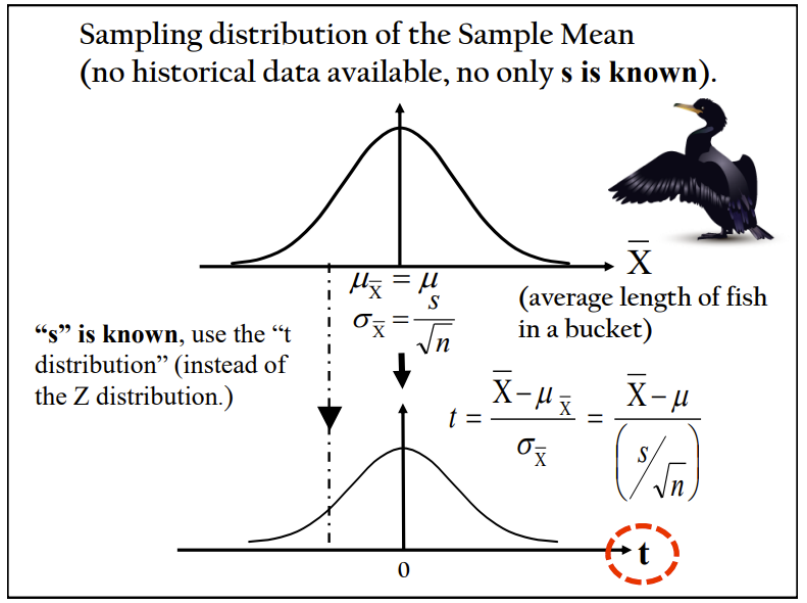


$$\bar{Y} \pm t_{\alpha/2,\,df} * s_{\bar{Y}} =$$

$$\bar{Y} \pm t_{\alpha/2,\,df} * \frac{s}{\sqrt{n}} =$$

$$6{,}334.67 \pm 1.9817 * \frac{4040.12}{\sqrt{111}} =$$

$$5{,}574.72 < \mu < 7{,}094.62$$



Confidence Interval Estimate for μ, (σ unknown, and only have s).

Lower limit: $\bar{X} - t_{\alpha/2,\,n-1}\frac{s}{\sqrt{n}}$

Upper limit: $\bar{X} + t_{\alpha/2,\,n-1}\frac{s}{\sqrt{n}}$

where $t_{\alpha/2,\,n-1}$ is the critical value t_{crit} of the t distribution with:

- n -1 degrees of freedom
- an area of α/2 in **each** tail
- t distribution assumptions must be satisfied

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

(d) Construct the 95% confidence interval for part (c)

```
Console  Terminal  Render  Background Jobs
R 4.2.2 · G:/My Drive/BEcon/tutor/ECON3350/01/
> t.test(fultonfish$quan, mu = 7200)

        One Sample t-test

data:  fultonfish$quan
t = -2.2566, df = 110, p-value = 0.02601
alternative hypothesis: true mean is not equal to 7200
95 percent confidence interval:
 5574.717 7094.617
sample estimates:
mean of x
 6334.667

>
```

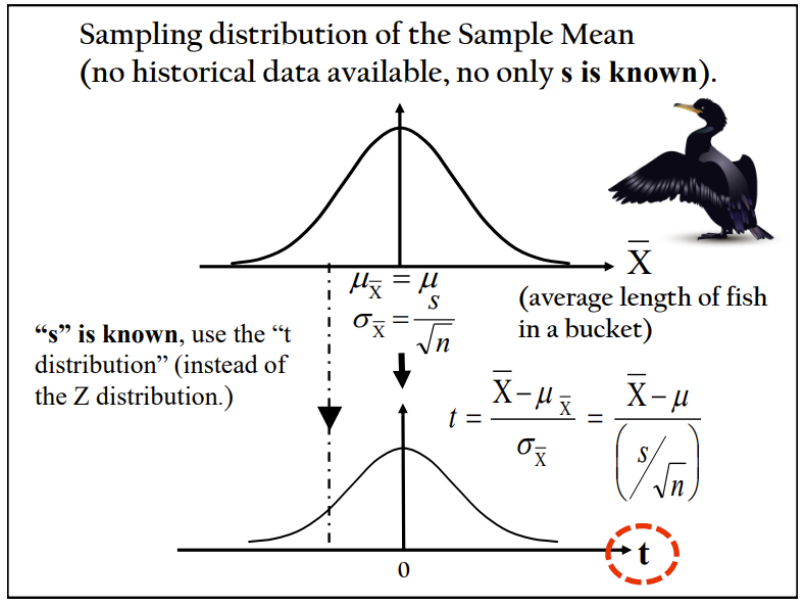
$$\bar{Y} \pm t_{\alpha/2,\,df} * s_{\bar{Y}} =$$

$$\bar{Y} \pm t_{\alpha/2,\,df} * \frac{s}{\sqrt{n}} =$$

$$6{,}334.67 \pm 1.9817 * \frac{4040.12}{\sqrt{111}} =$$

$$5{,}574.72 < \mu < 7{,}094.62$$

All the necessary information is available form the output of the `t.test` command. Indeed, the confidence interval itself is included in the output!



Confidence Interval Estimate for μ, (σ unknown, and only have s).

Lower limit: $\bar{X} - t_{\alpha/2,\,n-1} \frac{s}{\sqrt{n}}$

Upper limit: $\bar{X} + t_{\alpha/2,\,n-1} \frac{s}{\sqrt{n}}$

where $t_{\alpha/2,\,n-1}$ is the critical value t_{crit} of the t distribution with:

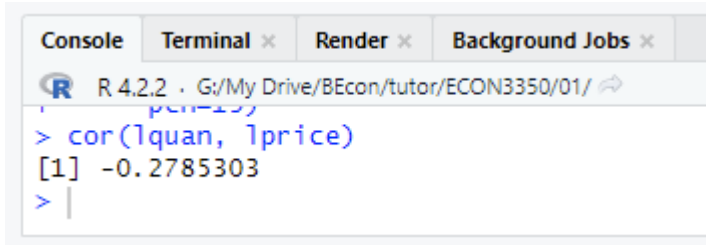
- n -1 degrees of freedom
- an area of α/2 in **each** tail
- t distribution assumptions must be satisfied

(e) Plot `lprice` against `lquan` and label the variable `lprice` as "log(Price) of whiting per pound" and `lquan` as "log(Quantity)". Then, comment on the nature of the relationship between these two variables.

Solution Generate the plot the same way as in Question 1, part (b).

```
attach(fultonfish)
plot(lquan, lprice,
     main = "Log Price and Log Quantity",
     xlab="log(Quantity)",
     ylab="log(Price) of whiting per pound",
     pch=19)
```

(e) Plot `lprice` against `lquan` and label the variable `lprice` as "log(Price) of whiting per pound" and `lquan` as "log(Quantity)". Then, comment on the nature of the relationship between these two variables.



Conceptually, we expect price and quantity to be negatively related, but there does not to appear to be a clear relationship between price and quantity in this data. We can investigate it further by computing the sample correlation.

```
cor(lquan, lprice)
```

```
## [1] -0.2785303
```

The correlation coefficient is slightly negative but not particularly strong. Does this mean demand for whiting is not very affected by prices?

(f) Save this workfile to any folder on any drive.

Solution Save the entire workspace in `RData` format using the `save` command in combination with the `ls` command.

```
save(list = ls(all = TRUE), file = "tuturial01.RData")
```

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Thank you

Francisco Tavares Garcia

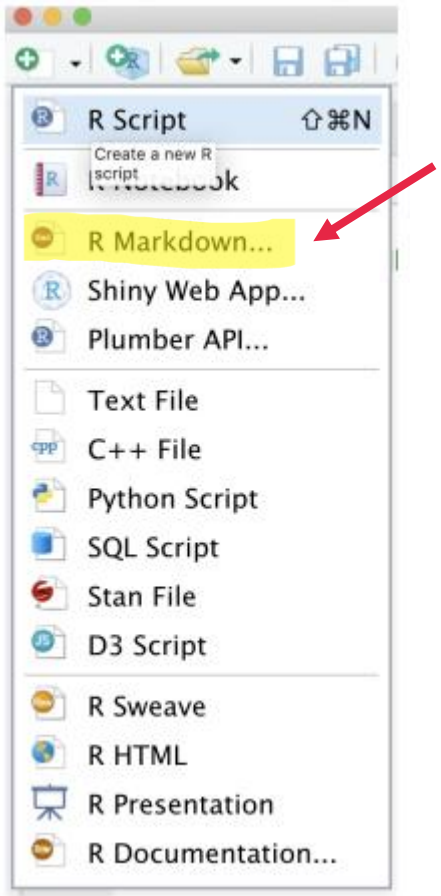
Academic Tutor | School of Economics

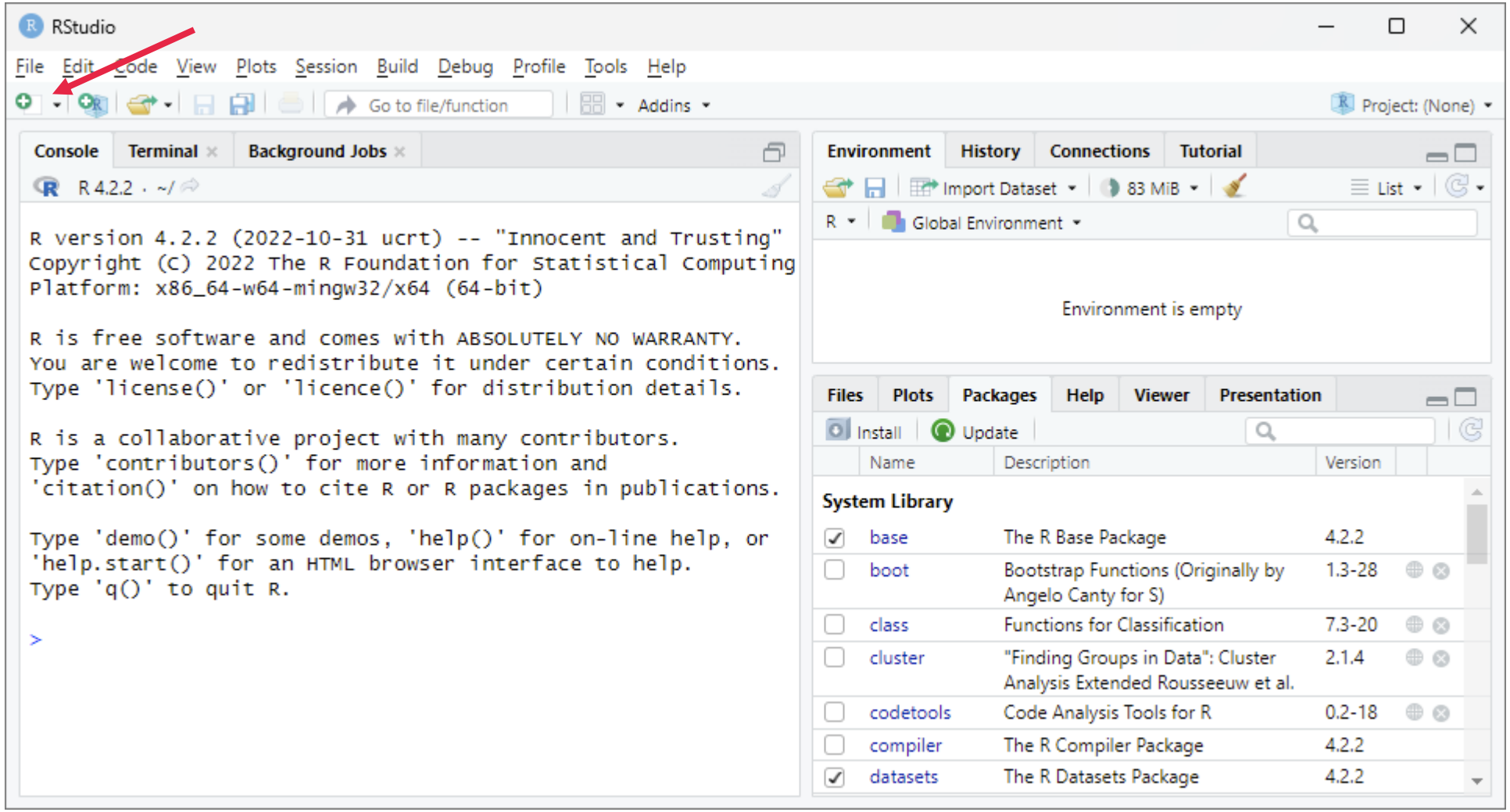
tavaresgarcia.github.io

Reference

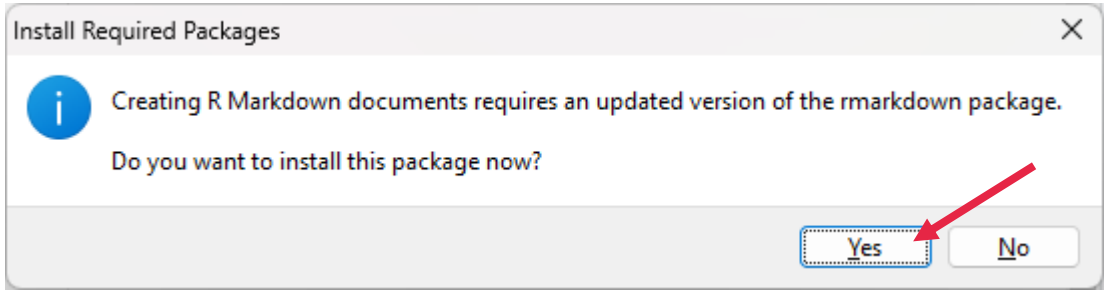
Stock, J. H., & Watson, M. W. (2019). Introduction to Econometrics, Global Edition, 4th edition. Pearson Education Limited.

Bonus – R Markdown



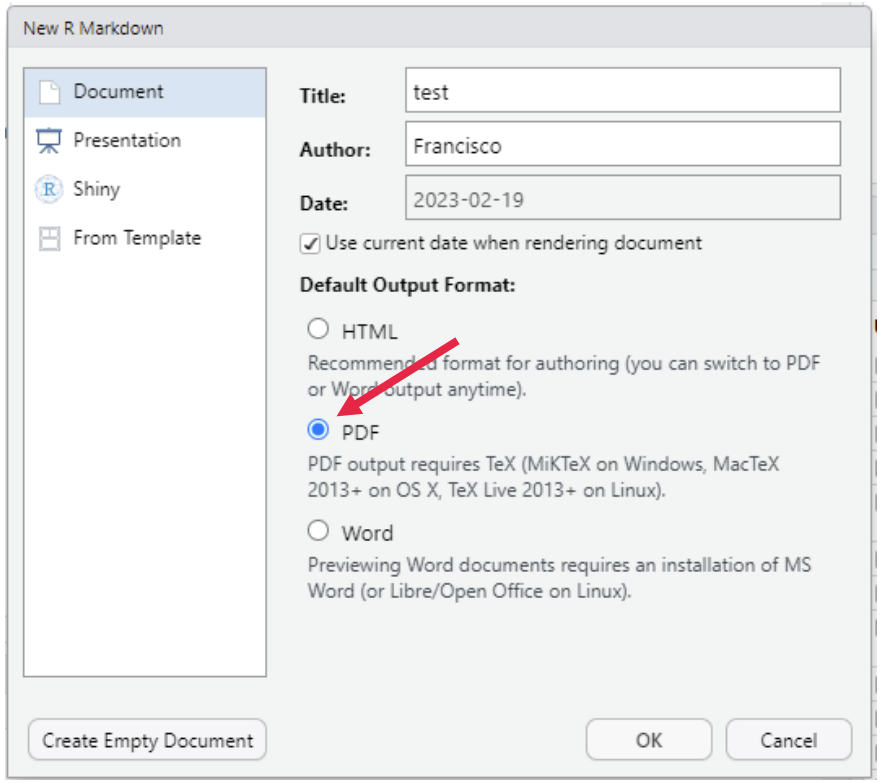


R Markdown - installation



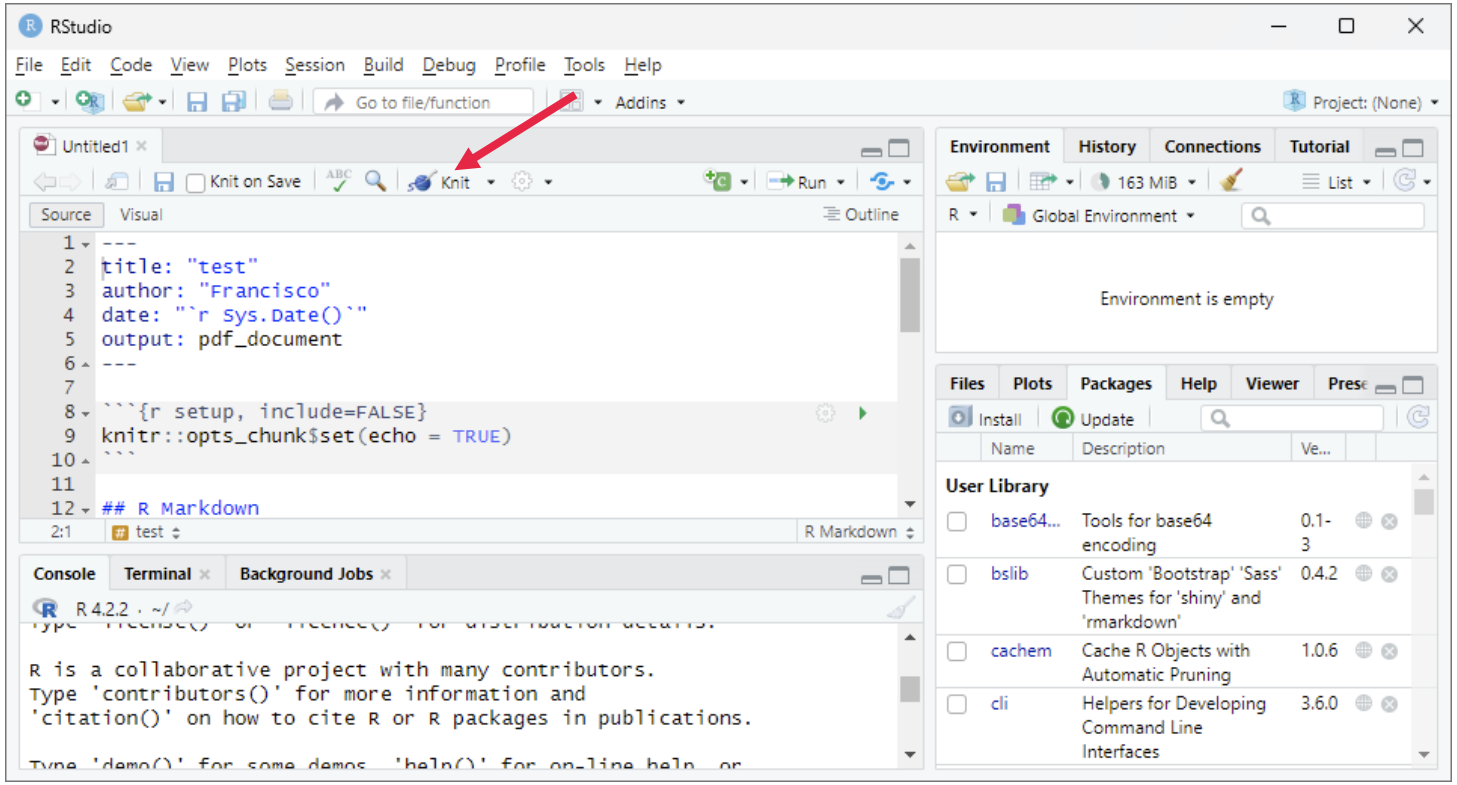
It will install 20+ packages to run R Markdown.

R Markdown - new document and Knit



Choose pdf to create documents using LaTeX.

Save your file, then Knit to PDF.

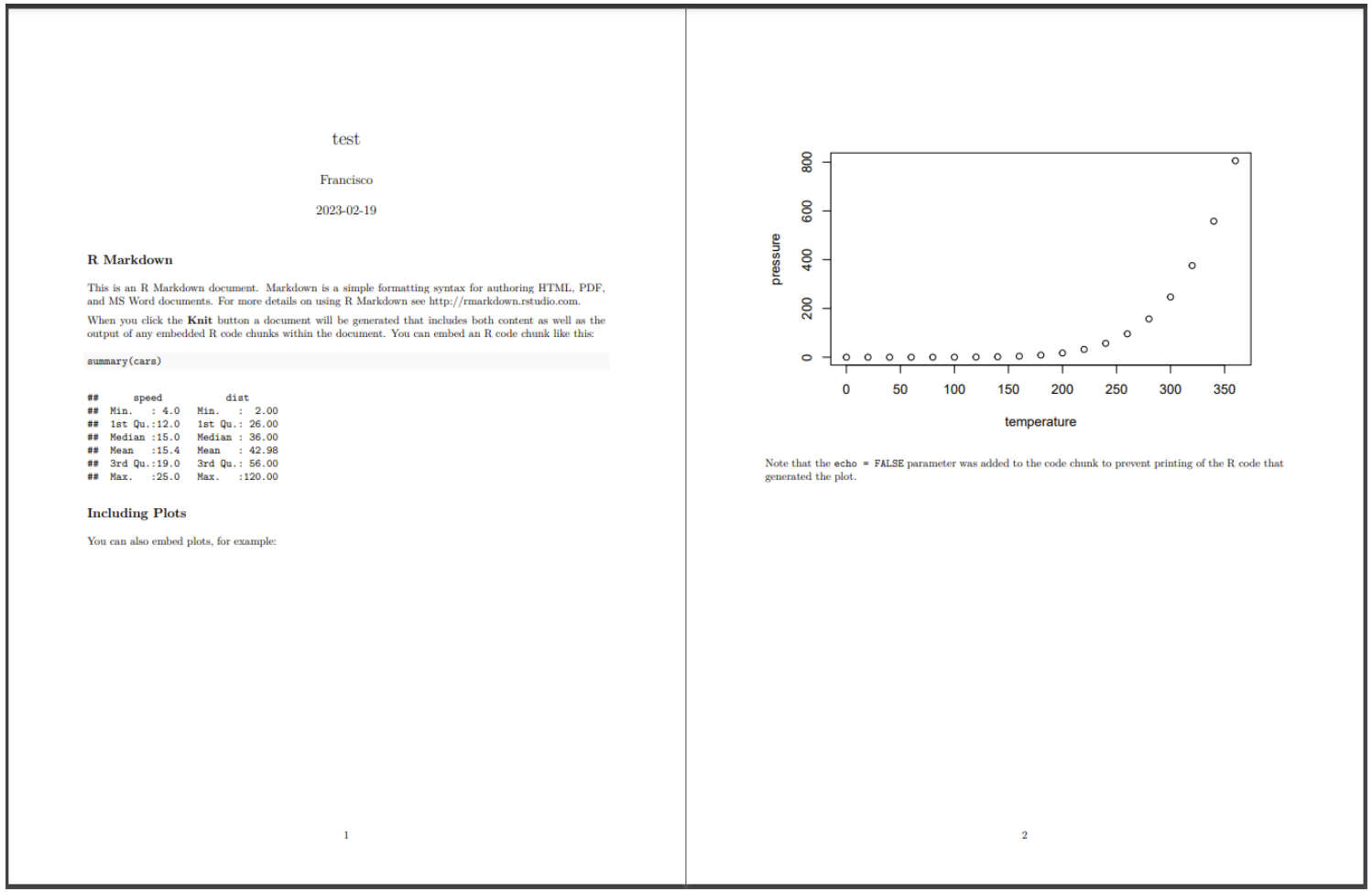


R Markdown - PDF

Go to the same folder you saved your .rmd file.
There you will find the PDF generated

You need to install the package tinytex.
To do so, run the following code:

```
tinytex::install_tinytex()
# to uninstall TinyTeX, run
# tinytex::uninstall_tinytex()
```



test

Francisco

2023-02-19

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:

1

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

2