



Estadística Descriptiva e Inferencia utilizando R (*Updated*)

Programa Estadístico de “Código Abierto”

versión 2.10.1

Dr. Orville M. Disdier

Objetivos

1. Exponer las posibles ventajas y desventajas del programa R.
2. Definir y demostrar algunos comandos y operaciones básicas.
3. Demostrar el uso de R para realizar estadísticas
 - ✓ Descriptivas
 - ✓ Inferenciales
 - ✓ Medidas de correlación y regresión
4. Practicar y aplicar lo aprendido.



Diseño del Taller

Parte 1

Introducción



Parte 2

Definiciones, comandos y operaciones



Parte 3

Estadísticas

Práctica y
Aplicación

Parte I

Introducción

¿Qué es R?

- R es un programa estadístico creado por Ross Ihaka y Robert Gentleman de la Universidad de Auckland, Nueva Zelanda
 - ✓ Para el manejo de datos, gráficas y cálculos.
- Desde 1995 se distribuye gratis a través de Internet
 - ✓ Bajo los términos de la GNU General Public Licence
 - ▶ <http://www.gnu.org/>
- Desde el 1997 un “core group” lo maneja
 - ✓ Asistido de una gran cantidad de colaboradores internacionales.

¿Qué es R? (continuación)

- Está basado en los programas S y S-Plus, es gratis y completamente programable, lo que brinda flexibilidad en el análisis.
- Un gran número de paquetes suplementan a R, también accesibles a través de Internet, lo que convierte a R en un sistema integrado de herramientas para el análisis de datos.

Gráficas en R

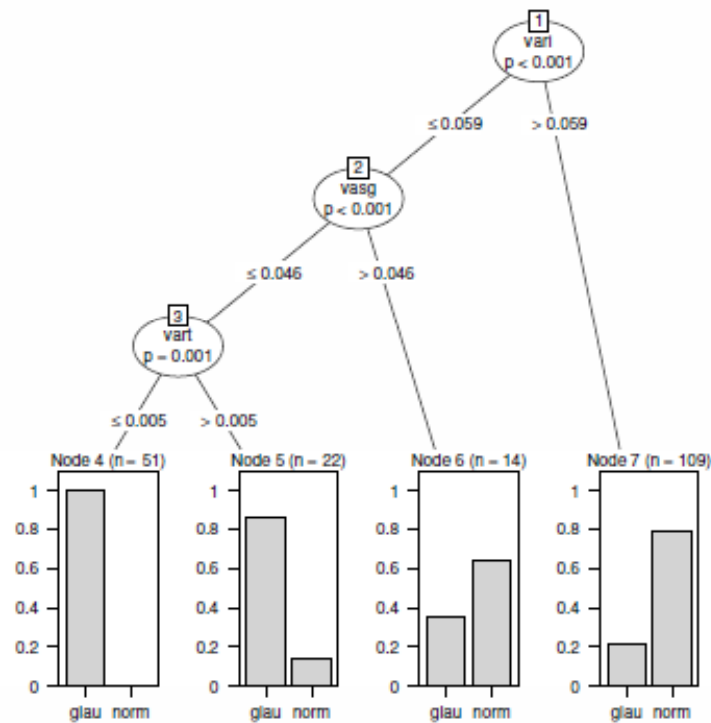


Figure 1.7

A novel decision tree plot, visualizing the distribution of the dependent variable in each terminal node. Produced using the `party` package by Torsten Hothorn, Kurt Hornik, and Achim Zeileis.

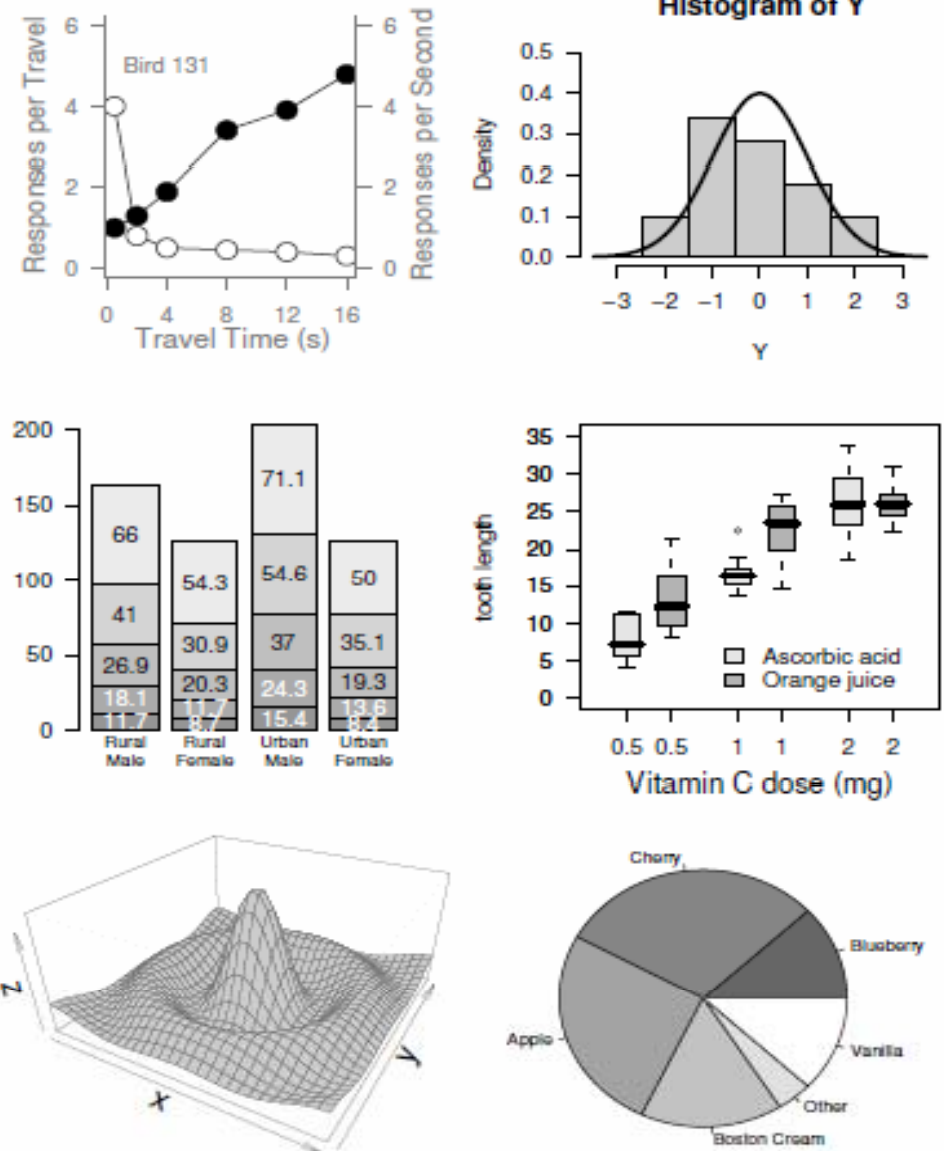
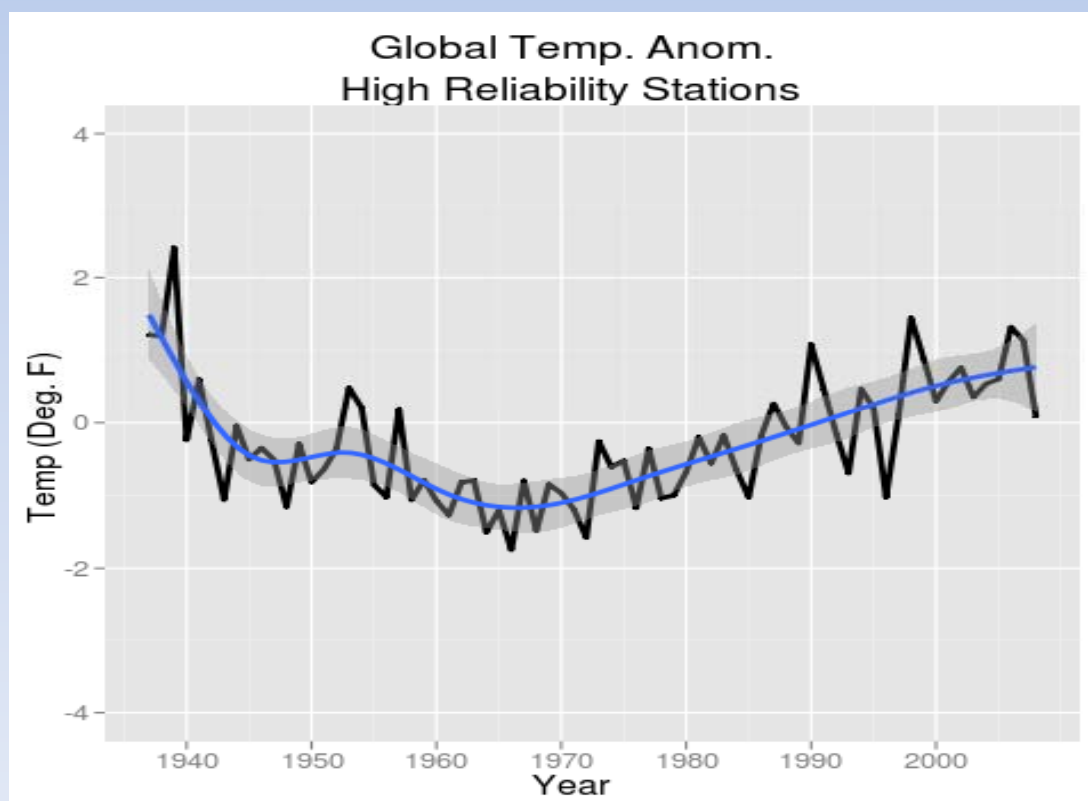


Figure 1.2

Some standard plots produced using R: (from left-to-right and top-to-bottom) a scatterplot, a histogram, a barplot, a boxplot, a 3D surface, and a piechart. In the first four cases, the basic plot type has been augmented by adding additional labels, lines, and axes. (The boxplot is adapted from an idea by Roger Bivand.)

Gráficas en R

```
data2 = read.table(file="/home/scott/resultset.csv",sep="," ,header=TRUE)
install.packages("ggplot2")
library(ggplot2)
png(file="temp_anom.png")
qplot(year,temp_anom,data=data2,geom=c("point","smooth","jitter"),alpha=l(1/10),y
lim=c(-4,4),main="Global Temp. Anom.\nHigh Reliability
Stations",xlab="Year",ylab="Temp (Deg. F)")
dev.off()
```



Gráficas en R

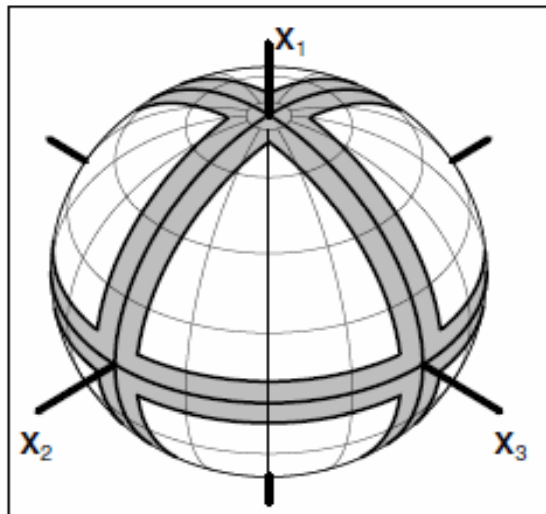
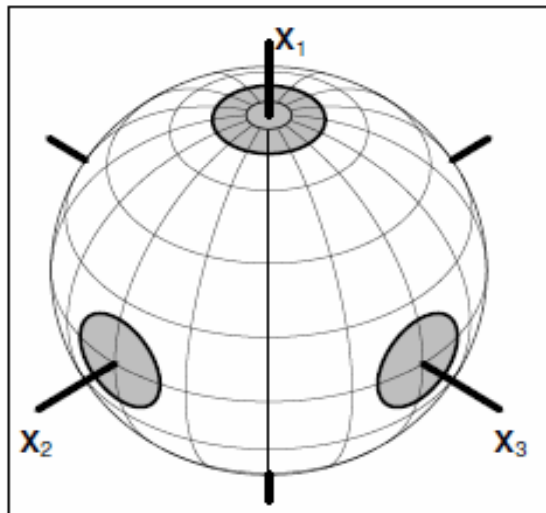


Figure 1.9
Didactic diagrams produced using R and functions provided by Arden Miller. The figures show a geometric representation of extensions to F-tests.

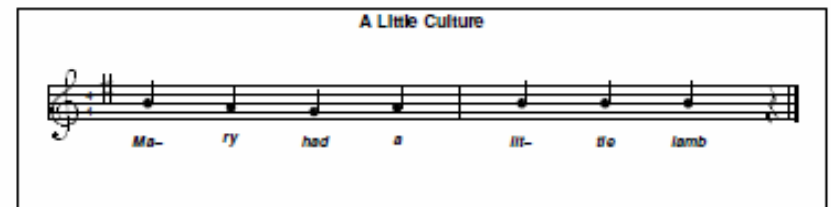


Figure 1.10
A music score produced using R (code by Steven Miller).



Figure 1.11
A piece of clip art produced using R.

Instalación

- Se accede a través del “Comprehensive R Archive Network” (CRAN) en la página: <http://cran.r-project.org>.
- Disponible en tres plataformas: Windows, MacOS X y Linux.
- La versión actual para Windows es R 2.10.1.
 - ✓ Es actualizada constantemente

The Comprehensive R Archive Network - Windows Internet Explorer

http://cran.r-project.org/ Live Search

File Edit View Favorites Tools Help Convert Select

The Comprehensive R Archive Network



The Comprehensive R Archive Network

Frequently used pages

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)
- [The R Journal](#)

Software

- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

Documentation

- [Manuals](#)
- [FAQs](#)
- [Contributed](#)

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:

- [Linux](#)
- [MacOS X](#)
- [Windows](#)

Source Code for all Platforms

Windows and Mac users most likely want 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** (2009-12-14): [R-2.10.1.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.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN

Internet 100%

The Comprehensive R Archive Network - Windows Internet Explorer

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

File Edit View Favorites Tools Help Convert Select

The Comprehensive R Archive Network

R for Windows

This directory contains binaries for a base distribution and packages to run on i386/x64 Windows.

Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal precautions with downloaded executables.

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch)
contrib	Binaries of contributed packages (managed by Uwe Ligges)

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)
- [The R Journal](#)

Software

- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

Documentation

- [Manuals](#)
- [FAQs](#)
- [Contributed](#)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or 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](#).

Last modified: April 4, 2004, by Friedrich Leisch

Done Internet 100%

The Comprehensive R Archive Network - Windows Internet Explorer

http://cran.r-project.org/ Live Search

File Edit View Favorites Tools Help Convert Select

The Comprehensive R Archive Network

R-2.10.1 for Windows



[Download R 2.10.1 for Windows](#) (32 megabytes)

[Installation and other instructions](#)

New features in this version: [Windows specific](#), [all platforms](#).

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [How do I install R when using Windows Vista?](#)
- [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.htm.

Last change: 2009-12-14, by Duncan Murdoch

Internet 100%

The Comprehensive R Archive Network - Windows Internet Explorer

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

File Edit View Favorites Tools Help

Convert Select

Instituto de Esta... National Center f... The U.S. Depart... Secretariado del ... CDC Centers for Disea... Bienvenidos The Compreh...

18b113bc9c20816da239f7fb3a30bf63 *R-2.9.1-win32.exe



<http://cran.r-project.org/bin/windows/base/md5sum.txt>

MD5sums:

http://www.openoffice.org/dev_docs/using_md5sums.html

CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R](#)

[Homepage](#)

[The R](#)

[Journal](#)

MD5sums: Generating...

File	MD5 Hash
R-2.9.1-win32.exe	18b113bc9c20816da239f7fb3a30bf63

Key

- ☐ Unprocessed
- ☒ OK / Done
- ☐ Processing
- ☐ Error (0 so far)

Batch (1 of 1):

File: 100%

Save Close

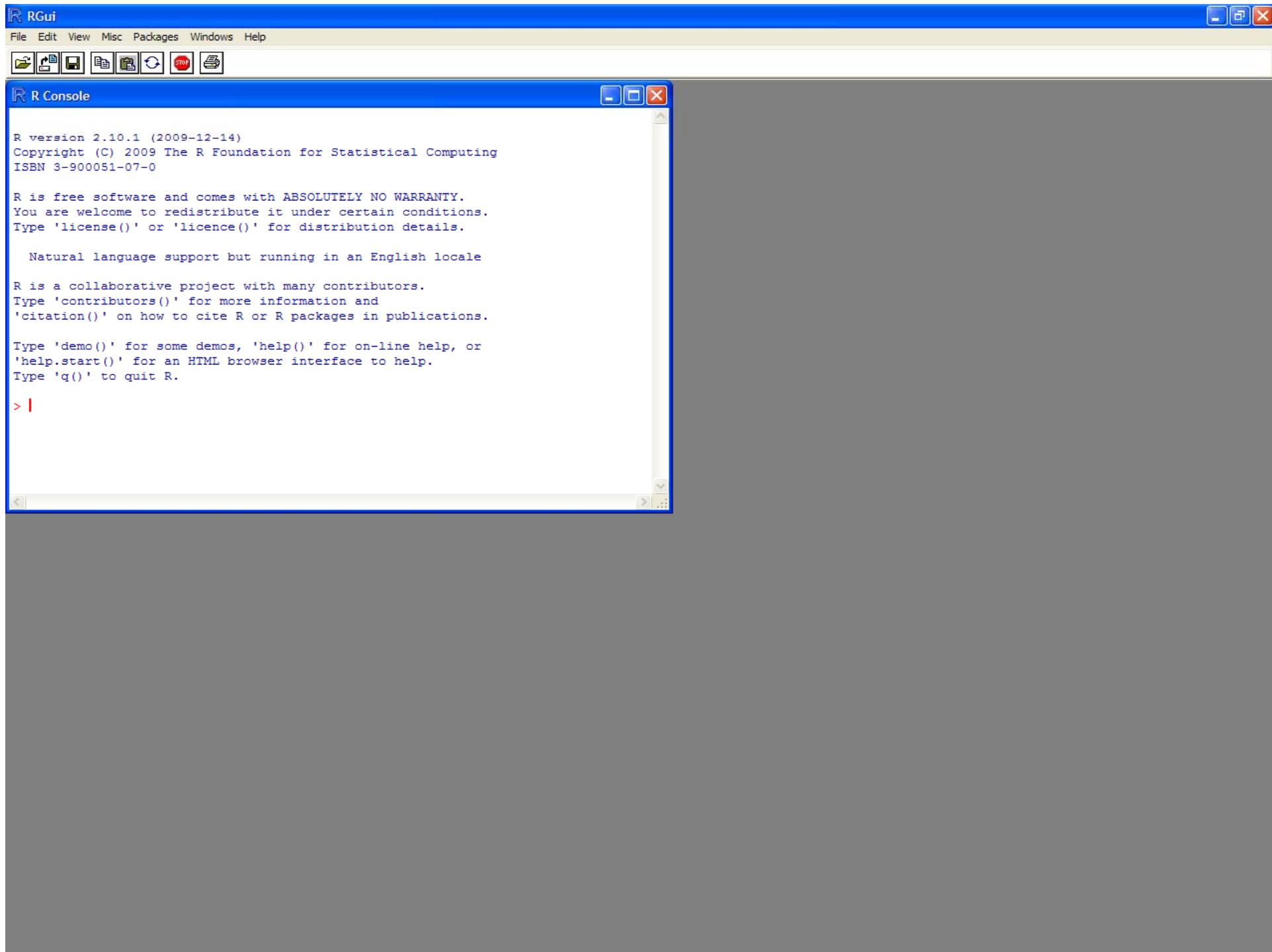
1 Sec

File Information

Path: C:\Downloads\
Name: R-2.9.1-win32.exe
Size: 36,568.88 Kb

14

Internet 240%



General Rweb interface - Windows Internet Explorer

http://bayes.math.montana.edu/Rweb/Rweb.general.html

File Edit View Favorites Tools Help

Convert Select

The Comprehensive R Archiv... General Rweb interface

Statistical Analysis On The Web

Rweb

To run **Rweb** just type the **R** (or Splus) code you want to execute into the text window and then click on the submit button. You will get a new html page with the text output of your code followed by the graphical output (if any) from your code.

Below the submit button is a text area where you can enter the URL for a Web accessible dataset and a browse button for selecting a dataset on your computer. Either way, the dataset will be read in using `read.table` with `header=T` and stored in a dataframe called `X`. The dataframe, `X`, will then be attached so you can use the variable names. Eventually I hope to add several other options for data entry ... let me know if you have any suggestions.

If you use the back button on your browser to come back to this page you can modify your old code and then resubmit it, or you can clear the text area and type in all new code. The computer time for all of this is donated by the Department of Mathematical Sciences, Montana State Univeristy.

Submit Erase Everything ...

External Data Entry

Enter a dataset URL :

Or

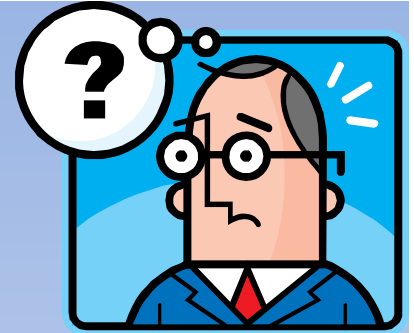
Select a local file to submit:

<http://bayes.math.montana.edu/Rweb/Rweb.general.html>

Parte II

Definiciones, comandos y operaciones

Ayuda (Help)

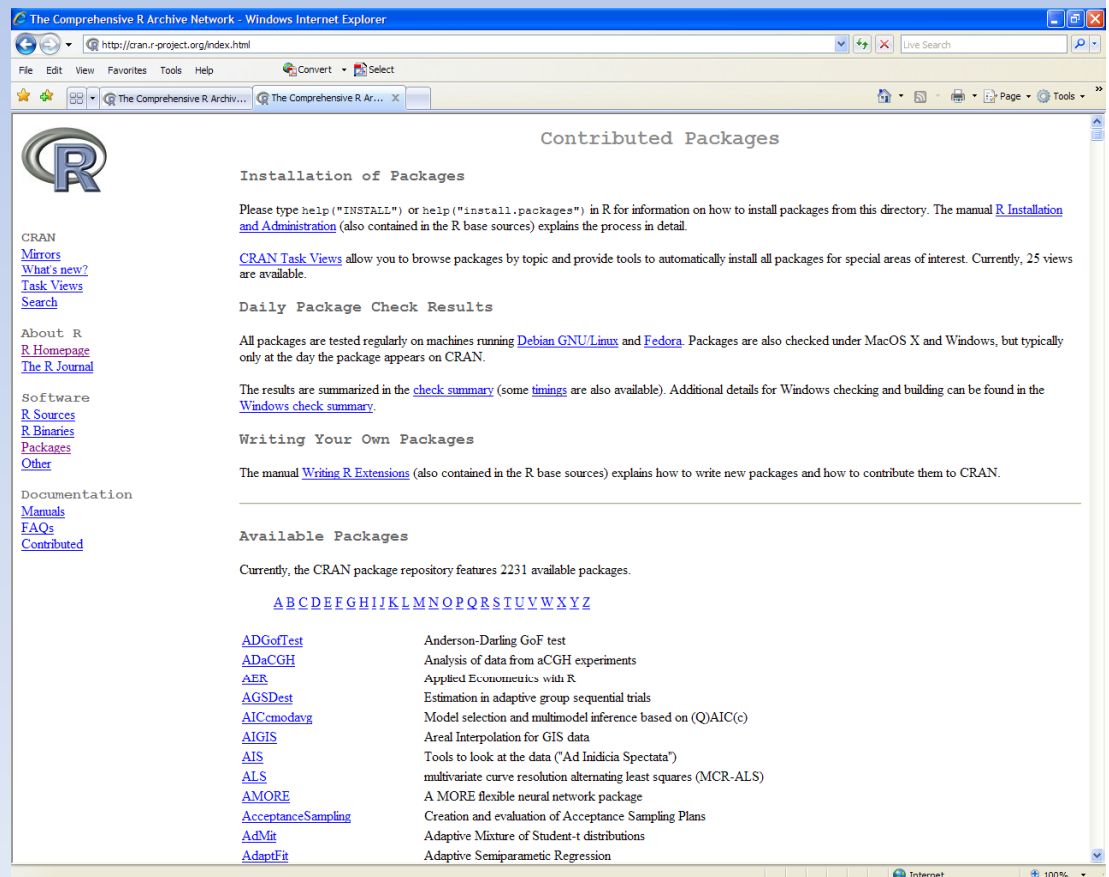


- Utilizar el “Help” en el menu del programa.
- Utilizar el comando:
 - > help.start()*
 - > help.search(search string)*
 - > help(function)*, ej. *> help(plot)*
 - > ?plot*

Paquetes de R

- Programas que se le pueden añadir a R para que realice funciones especializadas.
- Al instalar R se instalan algunos paquetes básicos

<http://cran.r-project.org/index.html>



Paquetes de R

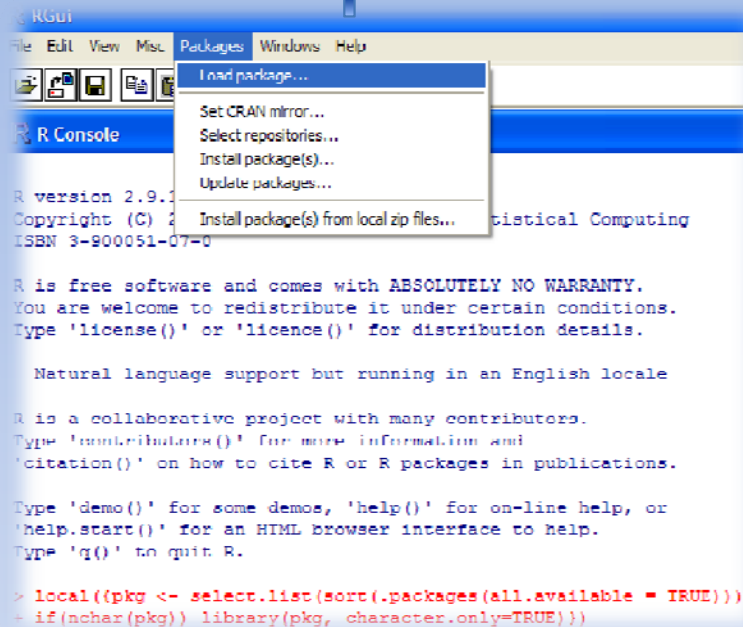
- Paquetes activos
>search()
- Paquetes disponibles para activar
>library()

1

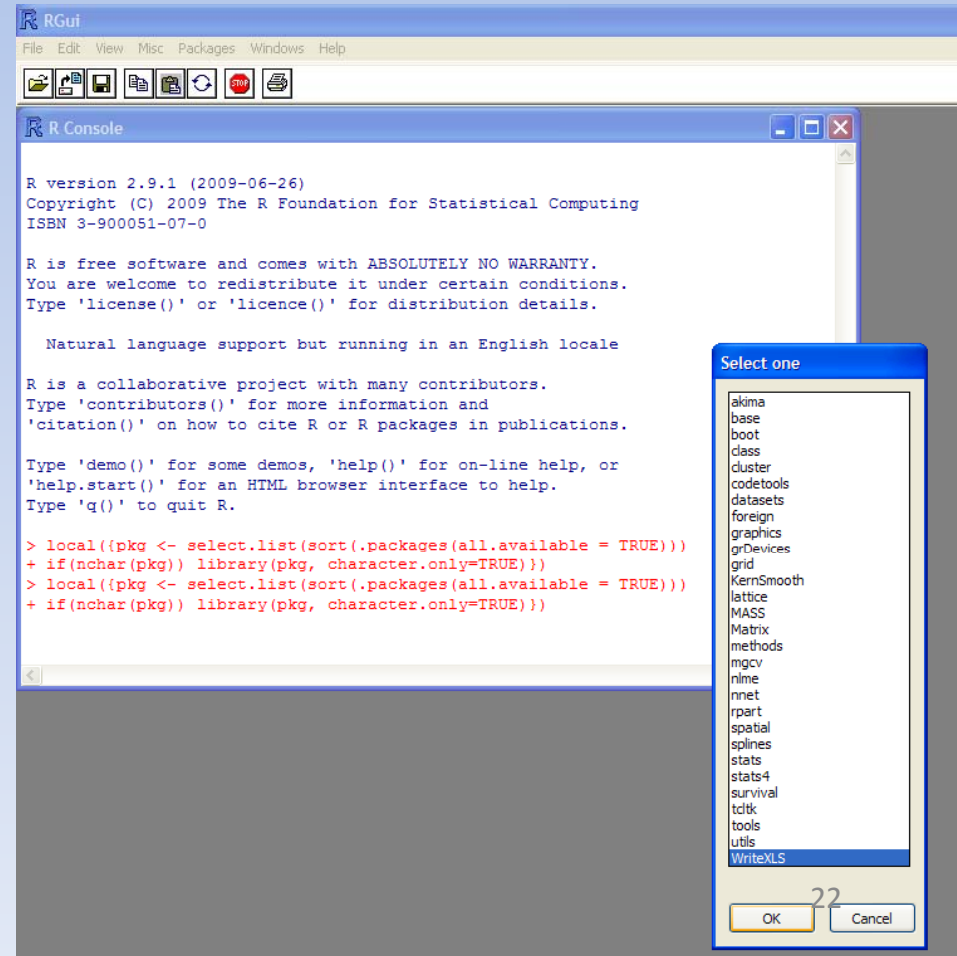


Activar los paquetes R

1



2



Operaciones matemáticas básicas

<i>Operación</i>	<i>Símbolo en R</i>
Suma	+
Resta	-
Multiplicación	*
División	/
Potencia	^

Orden de evaluación matemática



$$10+5^3*2/10$$

$$10+5^3*2/10 = 10+125*2/10$$

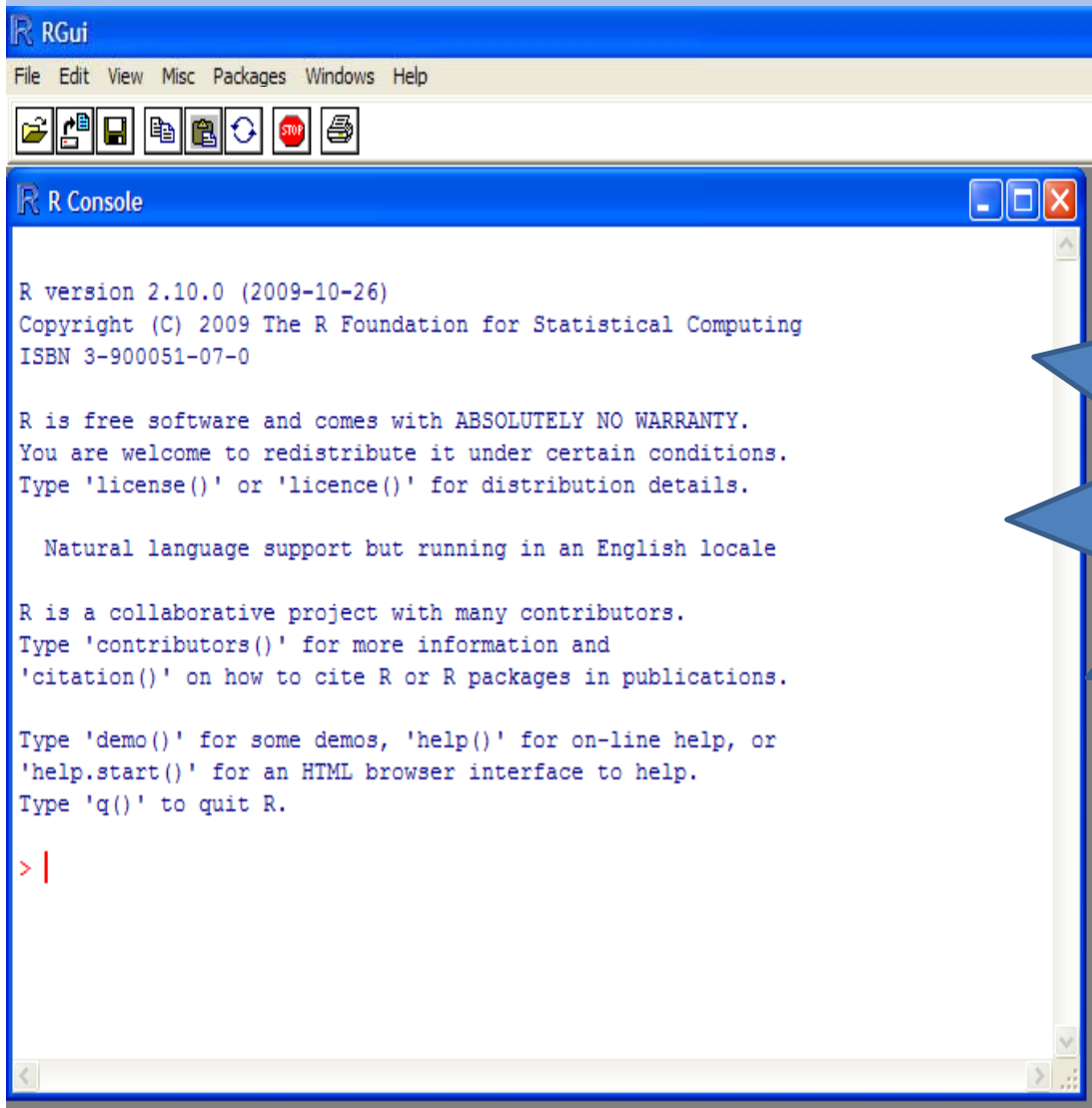
$$10+125*2/10 = 10+25$$

$$10+25 = 35$$

Funciones matemáticas

Raíz cuadrada	<code>sqrt()</code>
Exponenciación, e^x , $e=2.7182$	<code>exp()</code>
Logaritmo natural	<code>log()</code>
Logaritmo base 10	<code>log10()</code>
Seno	<code>sin()</code>
Coseno	<code>cos()</code>
Número aleatorios	<code>runif(<i>n,min,max</i>)</code> <code>round(runif(<i>n,min,max</i>),0)</code>
Valor absoluto	<code>abs()</code>

Utilizando R



The screenshot shows the RGui application window. The title bar reads 'RGui'. The menu bar includes 'File', 'Edit', 'View', 'Misc', 'Packages', 'Windows', and 'Help'. Below the menu bar is a toolbar with icons for file operations and execution. The 'R Console' window is open, displaying the following text:

```
R version 2.10.0 (2009-10-26)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

> |
```

**Números
Aleatorios**

Variables

- Se puede crear o asignar utilizando “=” ó “<-”.
- Mantiene el valor hasta que sea borrada o se le asigne otro valor.
- Sus nombres no pueden empezar con un número.
- R es “case sensitive”.
- Las variables pueden ser numéricas o de caracteres.
- Evitar los siguientes nombres: c, q, t, C, D, F, I, T.

Objetos

- Variables, valores
- Ver los objetos disponibles:
 - >ls()
 - >objects()
- Ver el valor del objeto
 - >*nombre del objeto*
- Borrar o eliminar objetos
 - >*rm(nombre del objeto)*

Vectores

- Colección de objetos (números, caracteres, otros)
- Concatenación de objetos nuevos o existentes,.
- Comando:
 `>vector1= c(1,2,3,4,5)`
- Vector secuencial
 `>seq(min,max,increment)`
 `>(desde):(hasta)`
- “Sort”
 `>sort(x)`

Vectores (continuación)

- Extracción de valores de un vector

```
>x=c(1,10,3,20,14)
```

```
> vector[posición del valor a extraer]
```

```
>x[5]
```

```
[1] 14
```

```
> x[c(2,4)]    # Extrae elementos en la posición 2 y  
                4 solamente
```

- [1] 10 20

```
> x[-c(2,4)]    # Remueve los elementos 2 y 4
```

```
[1] 10 12 13 14 15 16 17 18 19 20
```

Funciones para vectores

<i>Función</i>	<i>Descripción</i>
<code>length(v)</code>	Largo del vector v
<code>min(v)</code>	Mínimo
<code>max(v)</code>	Máximo
<code>sum(v)</code>	Suma de todos los elementos de v
<code>prod(v)</code>	Producto de los elementos de v
<code>mean(v)</code>	Media
<code>median(v)</code>	Mediana
<code>var(v)</code>	Varianza
<code>sd(v)</code>	Desviación estándar
<code>cor(v, w)</code>	Correlación entre dos vectores
<code>sort(v)</code>	Organiza los elementos en forma ascendente
<code>Quantile(v)</code>	Cuartilas empíricas

Operaciones lógicas

<i>Operación</i>	<i>Símbolo en R</i>
Menor a	<
Menor o igual a	<=
Mayor a	>
Mayor o igual a	>=
Exactamente igual a	==
Diferente a	!=



Observaciones incompletas (*missing values*)

- Los missing se establecen con las letras **NA**.
- Si una variable incluye un **NA** entre sus valores, los cálculos no se podrán realizar y el resultado será: **NA**.
- Para que el NA no se considere en los cálculos hay que escribir la siguiente instrucción:
`>na.rm=TRUE`



Observaciones incompletas:

Ejemplo

- Asignar valores a una variable.
`>pagos=c(200,300,150,NA)`
- Calcular la media
`> mean(pagos)`
[1] NA
- Calcular la media sin considerar el missing
`> mean(pagos, na.rm=TRUE)`
[1] 216.6667

Matrices

- Son arreglos de filas y/o columnas
- Se crean utilizando:

```
> matrix(data,nrow,ncol,byrow=F)
```

Se crearán por columnas

- Crear una matriz 2x4 (row by col)

```
>z=c(2,4,5,7,2,9,14,15)
```

```
>matrix1=matrix(z,2)
```

```
      [,1] [,2] [,3] [,4]
```

```
[1,]  2    5    2   14
```

```
[2,]  4    7    9   15
```

Ejercicio

- Crear una matriz utilizando la siguiente tabla:

Año	Núm. matrimonios	Núm. divorcios	Tasa de natalidad (%)
2001	28,598	13,870	14.6
2002	25,645	14,578	13.7
2003	25,236	14,225	13.1
2004	23,650	15,197	13.2
2005	23,511	15,816	13

Ejercicio (respuesta)

```
>datospr=  
matrix(c(2001,2002,2003,2004,2005,28598,25645,2  
5236,23650, 23511,13870,  
14578,14225,15197,15816,14.6,13.7,13.1,13.2,13),  
nrow=5)
```

	[,1]	[,2]	[,3]	[,4]
[1,]	2001	28598	13870	14.6
[2,]	2002	25645	14578	13.7
[3,]	2003	25236	14225	13.1
[4,]	2004	23650	15197	13.2
[5,]	2005	23511	15816	13.0

Ejercicio

- Calcule el promedio del número de matrimonios:

Año	Núm. matrimonios	Núm. divorcios	Tasa de natalidad (%)
2001	28,598	13,870	14.6
2002	25,645	14,578	13.7
2003	25,236	14,225	13.1
2004	23,650	15,197	13.2
2005	23,511	15,816	13

Lectura de datos externos

- Directorio de trabajo (“working directory”)
 - > getwd()
Indica cual es el directorio de trabajo
 - > setwd("Z:/Proyecto 2 (Academias)/Taller R")
Establece el nuevo directorio de trabajo

Lectura de datos externos (continuación)

- Importar de Excel

1

- Convertir a CSV
- “Save as CSV comma delimited”

2

- Asignar nombre y leer
- `>Datos.Tabla1=read.csv(“TABLA1-INFORME DE PUNTUACIONES.csv”,header=T)`

3

- Ver y corroborar
- `>Datos.Tabla1`

4

- Activar los datos (“Attach”)
- `>attach(Datos.Tabla1)`

Editor de datos

- Activar la pantalla de edición
 `>fix(objeto o variable)`
- Editar el contenido de una variable
 `>fix(pagos)`
- Remover (objetos o variables)
 `>rm(pagos)`

Gráficas

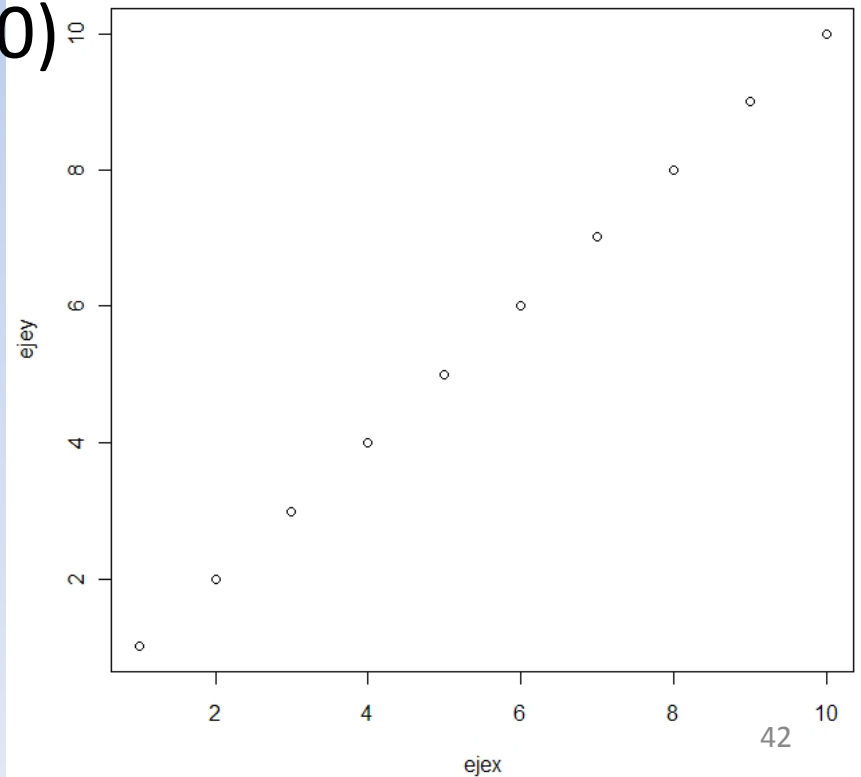
- Plot (x versus y)

```
>plot (x, y)
```

```
>ejex=c(1,2,3,4,5,6,7,8,9,10)
```

```
> ejey=c(1,2,3,4,5,6,7,8,9,10)
```

```
> plot(ejex,ejey)
```

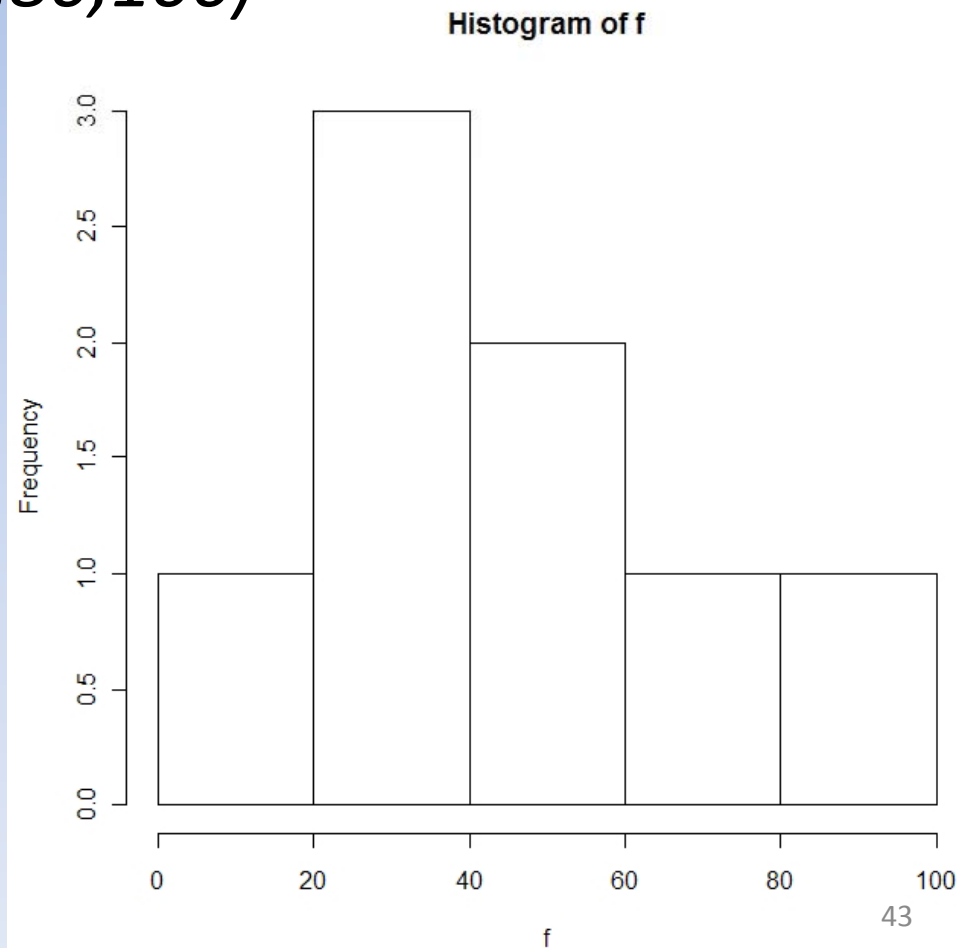


Gráficas (continuación)

- `hist ()` #histograma

```
>f=c(10,30,30,40,50,60,80,100)
```

```
>hist (f)
```



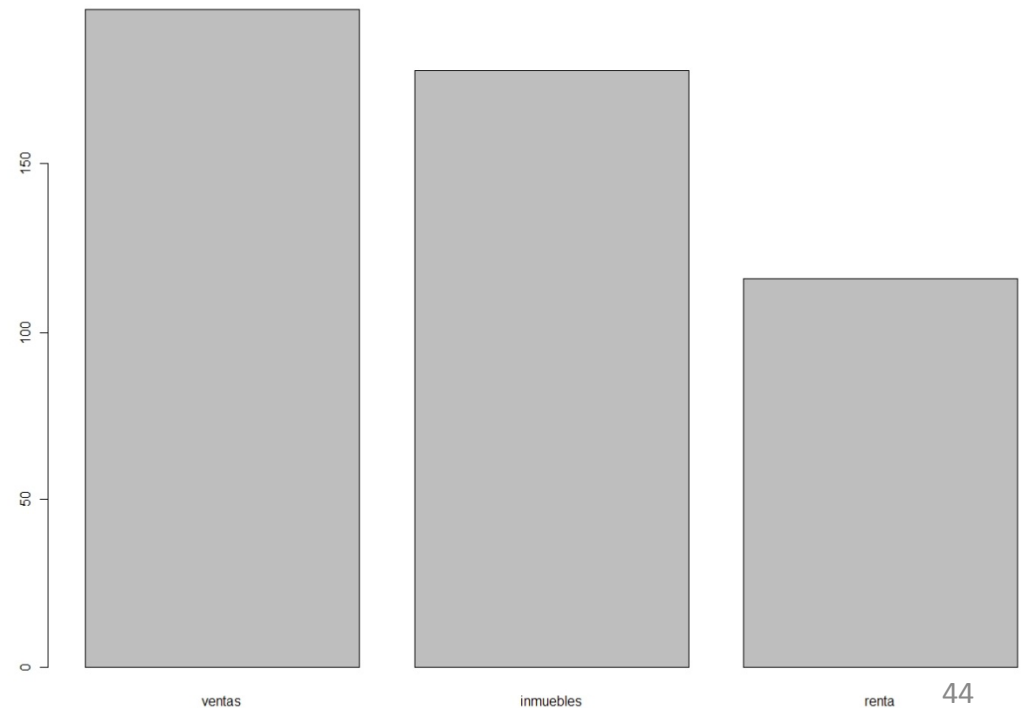
Gráficas de Barra (Plot)

- `barplot(x)`

>impuestos=c(196,178,116) #en miles de millones

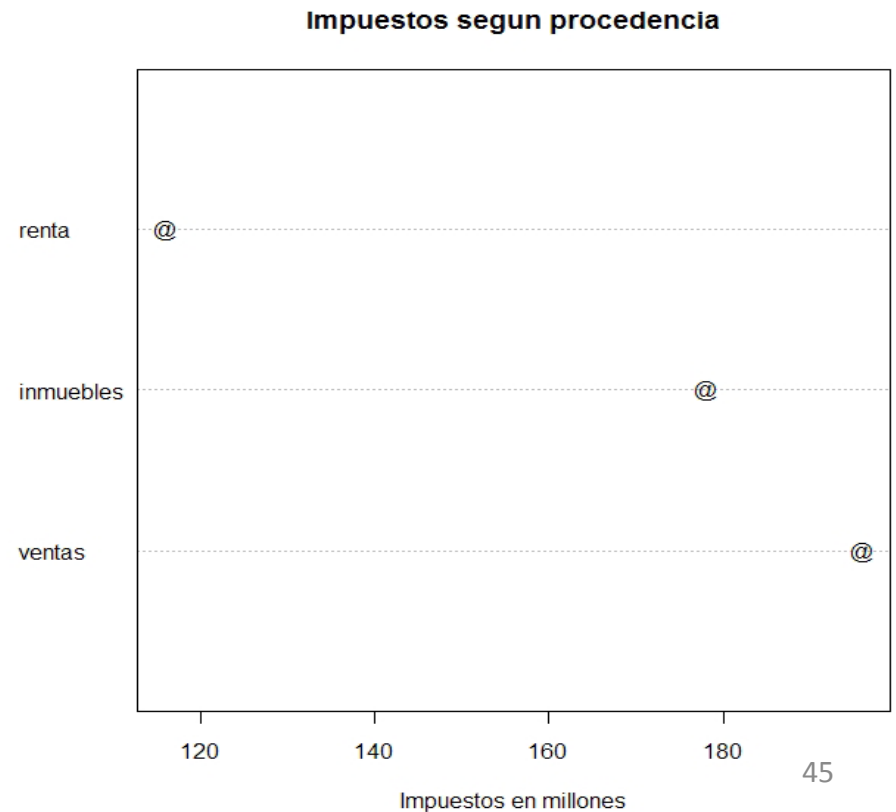
>names(impuestos)=c("ventas","inmuebles","renta")

>barplot(impuestos)

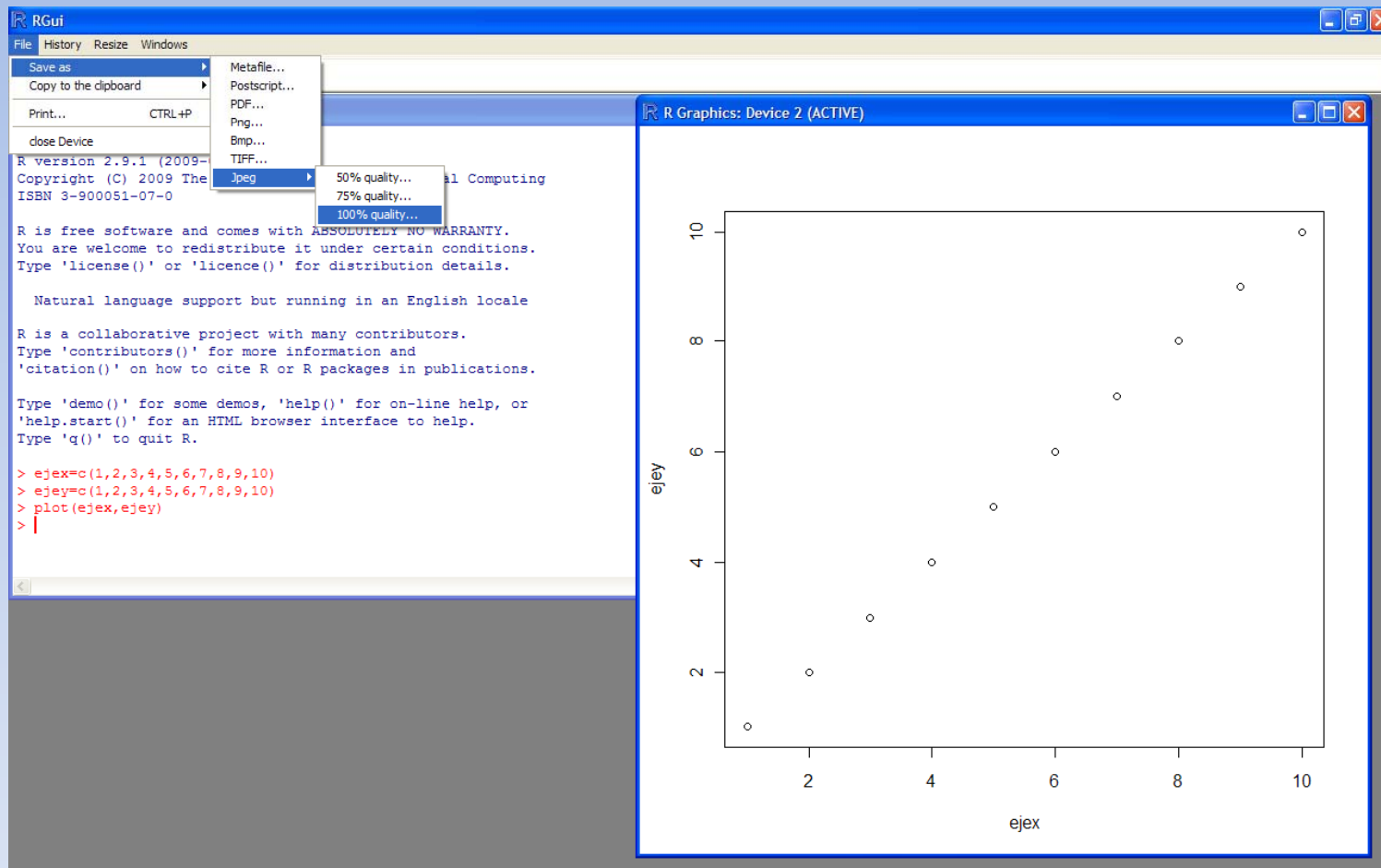


Gráficas de Barra (Dotchart)

- `dotchart(x), xlab="label ejex", main="titulo"`
 `>dotchart(impuestos, xlab="Impuestos en millones",
 main="Impuestos según procedencia", pch="@")`

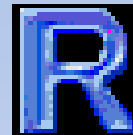
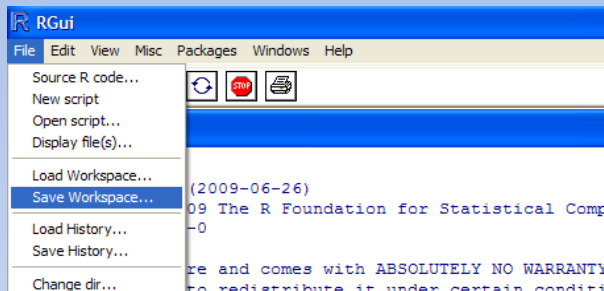


Guardar (save) las gráficas



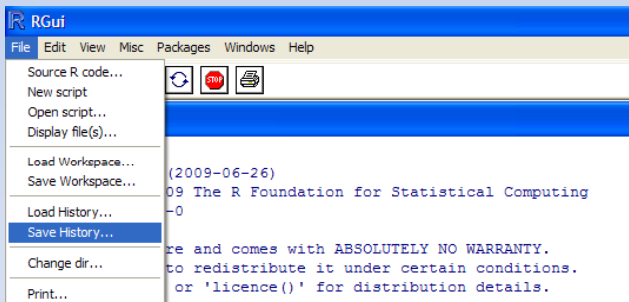
Resguardo (“Save”)

- Guardar (“save”) los datos y variables



Workspace.RData

- Guardar el historial (comandos y texto)



history.Rhistory

Práctica y Aplicación

TABLA 1
INFORME DE PUNTUACIONES
Escuela: *Del Monte Arriba*

ID	NOMBRE	SEXO	EDAD	Puntuaciones	
				<i>Examen #1</i>	<i>Examen #2</i>
1	José Hernández	M	8	80.5	90.2
2	Pietra Colón	F	9	70.6	76.9
3	Augusto Candela	M	10	54.9	50.1
4	Jorge Castro	M	9	100	100
5	Alexis Rodríguez	M	8	98.3	92.4
6	Julio Valle	M	8	32.1	45.2
7	María Valle	F	10	89.6	85.4
8	Pedro Caraballo	M	9	99.5	89.6
9	Ivelise Dessert	F	8	69.7	72.5
10	Félix Mirabal	M	9	84.5	88.7

Ejercicio #1

1. Sumar *Edad*
2. Promedio *Puntuaciones Examen #2*

ID	NOMBRE	SEXO	EDAD	Puntuaciones	
				Examen #1	Examen #2
1	José Hernández	M	8	82.5	92.2
2	Patricia Collins	F	9	70.6	76.9
3	Augusto Cardenas	M	10	54.9	50.1
4	Jorge Castro	M	9	100	100
5	Alexis Rodriguez	M	8	98.1	92.4
6	Julio Valdez	M	8	32.1	45.2
7	Maria Velez	F	10	89.6	85.4
8	Pedro Caraballo	M	9	99.5	89.6
9	Isabel Desautel	F	8	69.7	72.5
10	Felix Morales	M	9	84.5	88.7

EDAD
8
9
10
9
8
8
10
9
8
9

Examen #2
90.2
76.9
50.1
100
92.4
45.2
85.4
89.6
72.5
88.7

Resultados-Ejercicio #1

1. Sumar Edad = 88

```
> edad=8+9+10+9+8+8+10+9+8+9  
> edad  
[1] 88  
> edad2=c(8,9,10,9,8,8,10,9,8,9)  
> sum(edad2)  
[1] 88
```

Resultados-Ejercicio #1

2. Promedio *Puntuaciones Examen #2 = 79.1*

```
> examen2a=90.2+76.9+50.1+100+92.4+45.2+85.4+89.6+72.5+88.7
> examen2a
[1] 791
> examen2a/10
[1] 79.1
> examen2b=c(90.2,76.9,50.1,100,92.4,45.2,85.4,89.6,72.5,88.7)
> examen2b
[1] 90.2 76.9 50.1 100.0 92.4 45.2 85.4 89.6 72.5 88.7
> mean(examen2b)
[1] 79.1
```

TABLA 2
ESTATURA Y PESO DE LOS ESTUDIANTES
Escuela: *Del Monte Arriba*

ID	Estatura (pulg.)	Peso (lbs.)
1	66	140
2	67	180
3	58	130
4	73	200
5	69	175
6	67	180
7	71	179

Ejercicio #2

1. Media, Mediana, Var, Min, y Max *Estatura*
2. Media, Mediana, Var, Min, y Max *Peso*

ID	Estatura (pulg.)	Peso (lbs.)
1	66	140
2	67	180
3	58	130
4	73	200
5	69	175
6	67	180
7	71	179

Estatura (pulg.)
66
67
58
73
69
67
71

Peso (lbs.)
140
180
130
200
175
180
179

Resultados-Ejercicio #2

1. Media, Mediana, Var, MIN, y Max *Estatura:*

Media = **67.3**

Mediana = **67**

Var = **22.9**

Min = **58**

Max = **73**

```
> estatura=c(66,67,58,73,69,67,71)
> mean(estatura)
[1] 67.28571
> median(estatura)
[1] 67
> var(estatura)
[1] 22.90476
> min(estatura)
[1] 58
> max(estatura)
[1] 73
```

Resultados-Ejercicio #2

2. Media, Mediana, Var, MIN, y Max *Peso*:

Media = **169.9**

Mediana = **179**

Var = **616.8**

Min = **130**

Max = **200**

```
> peso=c(140,180,130,200,175,180,179)
> mean(peso)
[1] 169.1429
> median(peso)
[1] 179
> var(peso)
[1] 616.8095
> min(peso)
[1] 130
> max(peso)
[1] 200
```


TABLA 3

TABLA DE DEMANDA

Unidades vendidas de la novela “La Cosquilla” según su precio

Precio de venta (\$)	Unidades vendidas
10	5
9	9
8	13
7	17
6	21
5	25

Ejercicio #3

1. Correlación *Precio versus Unidades*
2. Plot *Precio versus Unidades* (curva de demanda)

Precio de venta (\$)	Unidades vendidas
10	5
9	9
8	13
7	17
6	21
5	25

Resultados-Ejercicio #3

1. Correlación *Precio* versus *Unidades*:

$$r = -1$$

```
> precio=c(10,9,8,7,6,5)
> unidades=c(5,9,13,17,21,25)
> cor(precio,unidades)
[1] -1
```

Resultados-Ejercicio #3

2. PLOT *Precio* versus *Unidades*:

```
> plot(unidades,precio)
```

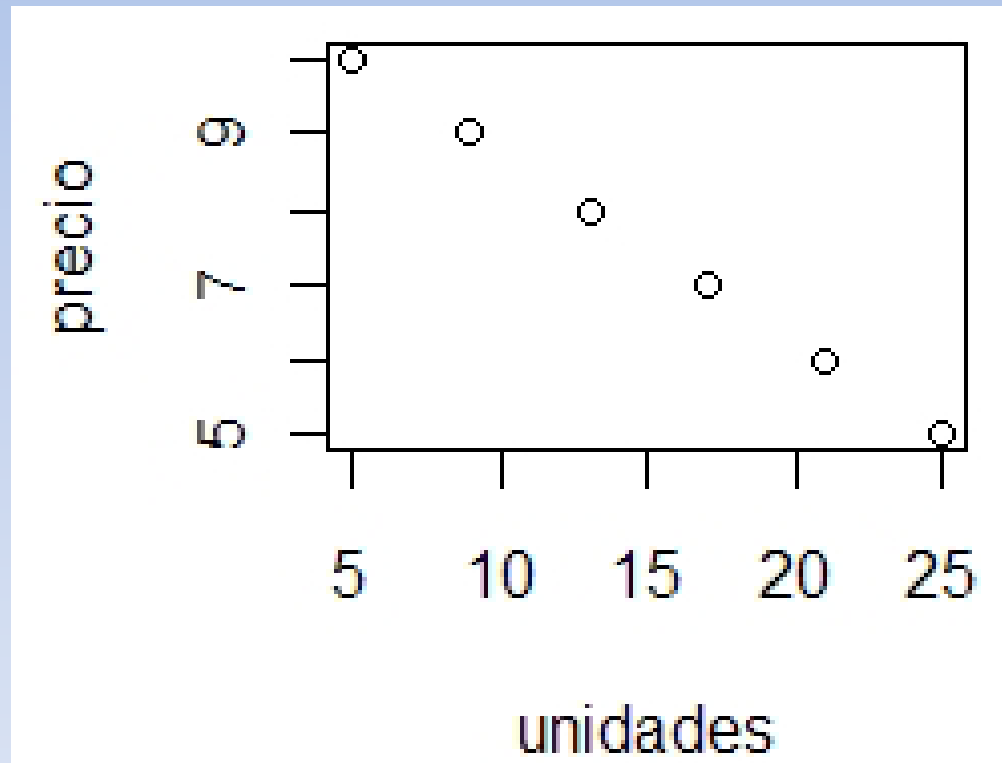


TABLA 4
TABLA DE EXPORTACION
Exportación de Frutas y Vegetales

Frutas y vegetales	Unidades exportadas (quintales)
Yautía	100
Guineos	680
Platanos maduros	789
Platanos verdes	150

Ejercicio #4

1. Generar una gráfica de barra *Tabla 4*

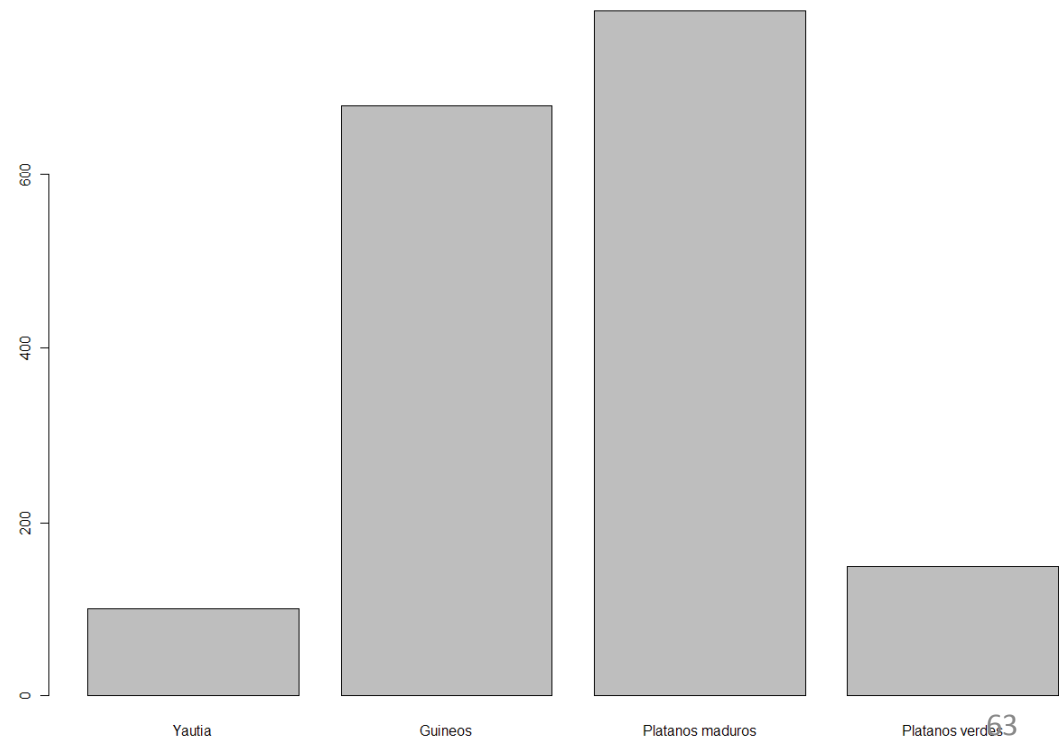
Frutas y vegetales	Unidades exportadas (quintales)
Yautía	100
Guineos	680
Platanos maduros	789
Platanos verdes	150

Resultados-Ejercicio #4

1. Generar una gráfica de barra *Tabla 4:*

```
> frutyveg=c(100,680,789,150)  
> names(frutyveg)=c("Yautia","Guineos","Platanos maduros","Platanos verdes")
```

```
> barplot(frutyveg)
```



Parte III

Análisis Estadístico

TABLA 2

Informe de Puntuaciones – Estudiantes de 5^{to} Grado Escuela: *Del Monte Arriba*

TABLA 1 - PUNTUACIONES 5to GRADO.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K
	ID	NOMBRE	SEXO	EDAD	Examen1	Examen2	Examen3				
2	1	José Hernández	M	8	81	90	100				
3	2	Pietra Colón	F	9	71	77	90				
4	3	Augusto Candela	M	10	55	50	68				
5	4	Jorge Castro	M	9	97	100	100				
6	5	Alexis Rodríguez	M	8	98	92	85				
7	6	Julio Valle	M	8	32	45	70				
8	7	María Valle	F	10	90	85	73				
9	8	Pedro Caraballo	M	9	90	90	71				
10	9	Ivelise Dessert	F	9	70	73	100				
11	10	Félix Mirabal	M	9	85	89	77				
12	11	Papo López	M	8	80	90	100				
13	12	Josefa Negrón	F	9	77	77	90				
14	13	Barbie Agosto	F	8	61	50	61				
15	14	Jorge Manso	M	9	99	100	100				
16	15	Roberto Diario	M	8	34	92	85				
17	16	Julieta Fragón	F	8	40	45	70				
18	17	Florinda Mesa	F	10	67	85	89				
19	18	Jaime Justo	M	9	90	90	81				
20	19	Claudia Natann	F	9	32	73	100				
21	20	Francisco Revedrón	M	9	44	89	77				
22	21	Jessica Fiel	F	8	66	90	55				
23	22	Juanita García	F	9	78	77	86				
24	23	Noel Zimbaue	M	10	44	50	32				
25	24	Heriberto Taniel	M	9	89	100	100				
26	25	Osvaldo Ortiz	M	8	66	92	85				
27	26	Darío Damian	M	8	32	45	70				
28	27	Harold Cruz	M	9	90	85	99				
29	28	Jason Killer	M	9	91	90	100				
30	29	Betty Correa	F	8	70	73	100				
31	30	Rafael Kintero	F	9	85	89	77				
32											
33											
34											

TABLA 1 - PUNTUACIONES 5to GRAD

Leer los datos externos

- Importar de Excel 2007

1

- Convertir a CSV
- “Save as CSV comma delimited”

2

- Asignar nombre y leer
- `>Datos.Tabla2=read.csv("TTABLA1-PUNTUACIONES 5to GRADO.csv",header=T)`

3

- Ver y corroborar
- `>Datos.Tabla2`

4

- Activar los datos (“Attach”)
- `>attach(Datos.Tabla2)`

Frecuencias

- Comandos
 - > Tablasexo= table(NOMBRE,SEXO) **Crear tabla**
 - > margin.table(Tablasexo,2)

• Ejemplo

> Tablasexo= table(NOMBRE,SEXO)

> Tablasexo

NOMBRE	SEXO
	F M
Alexis Rodríguez	0 1
Augusto Candela	0 1
Barbie Agosto	1 0
Betty Correa	1 0
Claudia Natann	1 0
Dario Damian	0 1
Félix Mirabal	0 1
Florinda Mesa	1 0
Francisco Revedrón	0 1
Harold Cruz	0 1
Heriberto Tariel	0 1
Ivelise Dessert	1 0
Jaime Justo	0 1
Jason Killer	0 1
Jessica Fiel	1 0
Jorge Castro	0 1
Jorge Manso	0 1
José Hernández	0 1
Josefa Negrón	1 0
Juanita García	1 0
Julieta Fragón	1 0
Julio Valle	0 1
María Valle	1 0
Noel Zimbaue	0 1
Oswaldo Ortiz	0 1
Papo López	0 1
Pedro Caraballo	0 1
Pietra Colón	1 0
Rafael Kintero	1 0
Roberto Diario	0 1

• Ejemplo

> margin.table(Tablasexo,2)

SEXO

F M

12 18

t-test (datos independientes)

- Comandos

- > var.test (grupo1, grupo2) *Prueba de varianzas*

- > t.test(grupo1, grupo2, var.equal=TRUE)

- Ejemplo

- > # Prueba de igualdad de varianzas

- > var.test(Examen1, Examen2)

F test to compare two variances

data: Examen1 and Examen2

F = 1.5171, num df = 29, denom df = 29, p-value = 0.2675

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval: 0.7220987 3.1874762

sample estimates: ratio of variances 1.517126

- Ejemplo

- > # Prueba t (dos lados)

- > t.test(Examen1, Examen2, var.equal=TRUE)

Two Sample t-test

data: Examen1 and Examen2

t = -1.7499, df = 58, p-value = 0.08542

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-19.223465 1.290131

sample estimates: mean of x mean of y
70.13333 79.10000

Nueva Pasta Dental

En un estudio se obtuvo una muestra de tres grupos de personas: se preguntó a 100 hombres, 130 mujeres y 90 niños, si les agradaba o no el sabor de una nueva pasta dental. Los resultados fueron los siguientes:

Las hipótesis son:

H0: La proporción de “gusto por la nueva pasta dental” es la misma en los tres grupos de personas

H1: Al menos en uno de los grupos la proporción es diferente.

Valores observados

	Hombres	Mujeres	Niños	
Les gustó el sabor	60	67	49	176
No les gustó el sabor	40	63	41	144
Total	100	130	90	320

Chi-cuadrada

- Comandos

- > a=matrix(c(*datos*),nc=*columnas*) Crear la matriz

- > chisq.test(*matriz*)

- Ejemplo

```
> # Crear la matriz
```

```
> a=matrix(c(60,40,67,63,49,41),nc=3)
```

```
> # Ver la matriz
```

```
> a
```

```
  [,1] [,2] [,3]
```

```
[1,]  60  67  49
```

```
[2,]  40  63  41
```

```
> #Prueba de Chi
```

```
> chisq.test(a)
```

```
      Pearson's Chi-squared test
```

```
data: a
```

```
X-squared = 1.6507, df = 2, p-value = 0.4381
```