

R 2019

18/03/2019

PRESENTACIONES

Profe Luis G. Moyano

lgmoyano@gmail.com

@luisgmoyano

Profe Carolina Gandini

r-lectures.github.io



QUÉ ES R?

- Software estadístico para GNU
- Manipulación de datos, cálculo y gráficos
- Implementación de S
- Abierto, extensible y colaborativo

QUÉ VAMOS A APRENDER EN ESTA MATERIA

- R cómo herramienta científica
- Foco en **datos**
- Horizontal para todas las carreras
- Autonomía

QUÉ NO VAMOS A APRENDER EN ESTA MATERIA

- No vamos a hacer énfasis en la lógica de programación
- No vamos a ver algoritmos en profundidad
- No vamos a cubrir R en su totalidad

R EN LA PROGRAMACIÓN Y EL ANÁLISIS DE DATOS

- Adoptado ampliamente en estadística, biología y ciencias de la computación
- R vs. Python y otros
- *Scripting vs. Interactive*
 - Workflow científico

VENTAJAS DE R

- Exploración
- Flexibilidad
- Comunidad

DESVENTAJAS DE R

- Memoria
- Velocidad (pero están **Rcpp** y **Rfortran**)
- Nivel de adopción heterogéneo

PREPÁRENSE PARA FRUSTRARSE

- R es idiosincrático
- Frustrarse es típico y temporal
- Persistencia!



ORGANIZACIÓN DE LA MATERIA

HORARIOS

- Lunes de 15 a 18, miércoles de 16 a 19
- ~ 1+ hora de teoría
- resto de práctica/consulta

GUÍAS DE EJERCICIOS

- Parte principal del curso
- Aproximadamente uno por clase
- Haremos hincapié en desarrollar "reproducible research"
 - informe + código + datos

PARCIALES

- Dos parciales, en la mitad y al final, aprox.
- Tendrán toda la información disponible durante el parcial
- Estarán fuertemente basados en las guías prácticas

PROYECTO FINAL

- Informe escrito
- Cada uno buscará tema y datos propios, acordando conmigo

CALENDARIO

- Parcial 1: ~ 8 de Abril
- Parcial 2: ~ 29 de Abril
- Proyecto final

HERRAMIENTAS

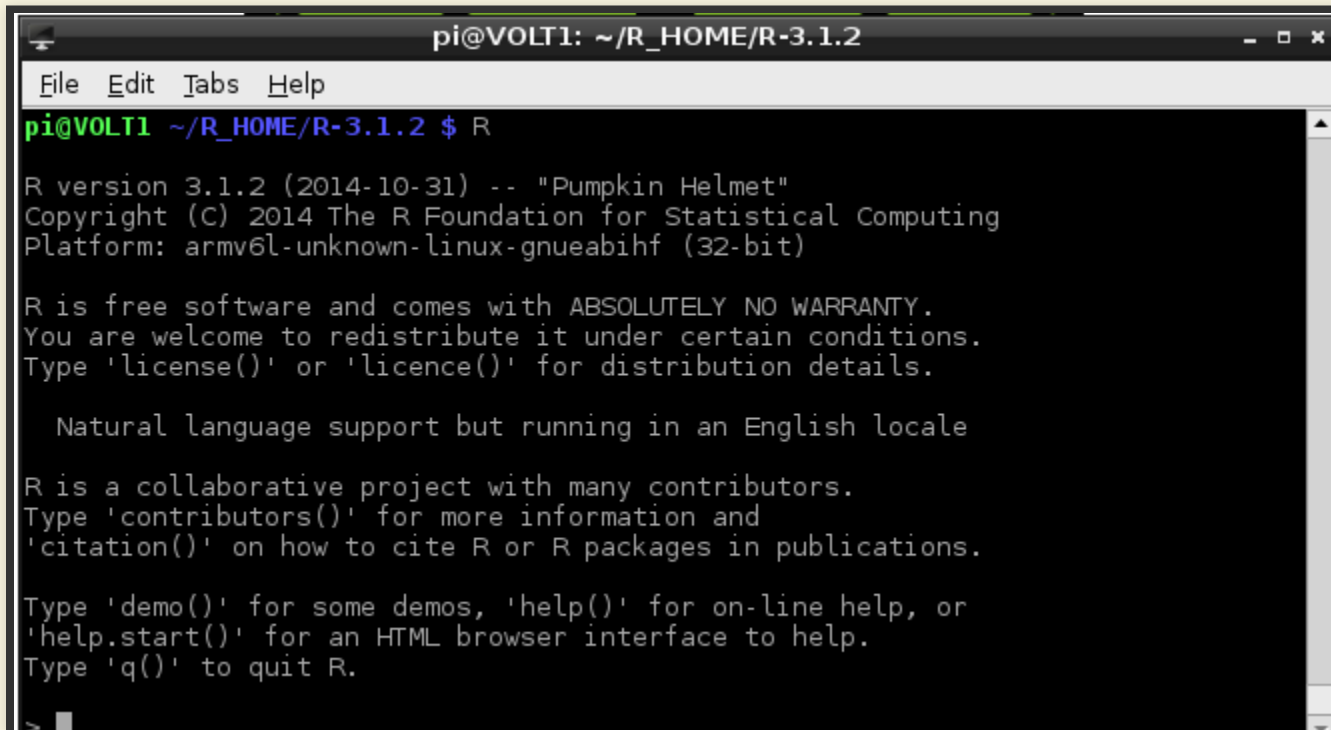
- Web del curso
- Google Classroom
- R en una terminal
- R con RStudio

INSTALACIÓN

- Instalar R
- Instalar RStudio

R EN EL TERMINAL

- Abran un terminal
- Pongan "R"
- Voilá

A screenshot of a terminal window titled 'pi@VOLT1: ~/R_HOME/R-3.1.2'. The terminal shows the command 'R' being executed, which displays the R version 3.1.2 startup information, including the date (2014-10-31), the version name 'Pumpkin Helmet', copyright information for The R Foundation, and the platform 'armv6l-unknown-linux-gnueabi (32-bit)'. It also includes a disclaimer about warranty and a list of useful commands like 'license()', 'demo()', and 'q()'.

```
pi@VOLT1: ~/R_HOME/R-3.1.2
File Edit Tabs Help
pi@VOLT1 ~/R_HOME/R-3.1.2 $ R
R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: armv6l-unknown-linux-gnueabi (32-bit)

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.

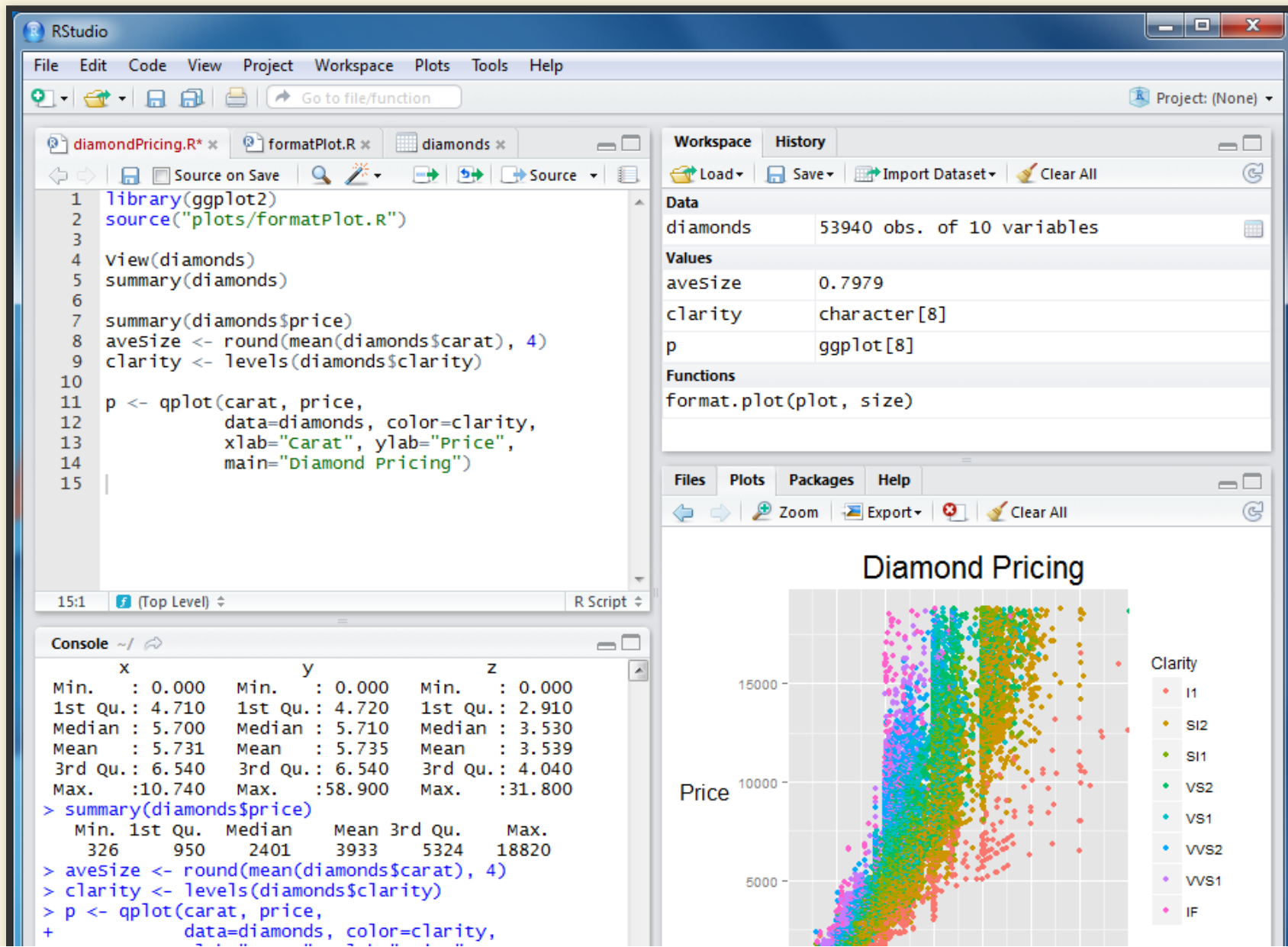
Natural language support but running in an English locale

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.
```

R CON RSTUDIO

- Abran RStudio
- Voilà



AYUDA

1. El propio R:

```
> ?plot <- ayuda de la función *plot*  
> ??plot <- "apropos"
```

2. *Vignettes* (manuales de paquetes)

3. Google - StackOverflow - R-bloogers

4. Google Classroom (local), Stackoverflow (global)

5. Profes Carolina y Luis (último recurso! ;)

COMANDOS BÁSICOS I

- Ayuda: `? y ??`
- Cargar un paquete: `library(GenomicRanges)`
- Instalar un paquete:
`install.packages("GenomicRanges")`
- Explorar y cargar datos: `data()`, `data(CO2)`
- Qué objetos tenemos? `ls()`
- Cargar código de disco:
`source("./archivito.R")`

COMANDOS BÁSICOS II

- Imprimir algunas líneas de un objeto: `head (mpg)`
- Estructura de un objeto: `str (mpg)`
- Resumen de un objeto: `summary (mpg)`
- Gráficar dos variables: `plot (displ, hwy)`

THE TIDYVERSE (AKA HADLEYVERSE)

An ecosystem of packages designed with common APIs and a shared philosophy.

- Desarrollado por Hadley Wickham
- Principales paquetes: dplyr y ggplot2
- Orden y centralización a través de paquetes coherentes entre sí
- La gramática como orden conceptual: verbos y sustantivos

VISUALIZACIÓN

PACKAGE *BASE*: FUNCIÓN `PLOT ()`

```
# Define the cars vector with 5 values
cars <- c(1, 3, 6, 4, 9)

# Graph the cars vector with all defaults
plot(cars)

# Graph cars using blue points overlayed by a line
plot(cars, type="o", col="blue")

# Create a title with a red, bold/italic font
title(main="Autos", col.main="red", font.main=4)
```

PACKAGE *BASE*: FUNCIÓN `PLOT ()`

```
# Get a random log-normal distribution
r <- rlnorm(1000)

# Get the distribution without plotting it using tighter breaks
h <- hist(r, plot=F, breaks=c(seq(0,max(r)+1, .1)))

# Plot the distribution using log scale on both axes, and use
# blue points
plot(h$counts, log="xy", pch=20, col="blue",
      main="Log-normal distribution",
      xlab="Value", ylab="Frequency")
```

PACKAGE *GGPLOT2*: FUNCIÓN `QPLOT ()`

- similar a `plot()`
- para hacer figuras básicas

```
qplot(displ, hwy, data = mpg)
```

```
qplot(displ, hwy, colour = class, data = mpg)
```

PRÁCTICA

Descargar **práctica 1.**

