

# Reportes automatizados con R Markdown

Pablo Aguirre Hörmann

Departamento de Gestión de la Información

[pablo.aguirre@sma.gob.cl](mailto:pablo.aguirre@sma.gob.cl)

26 de agosto de 2019



# Qué veremos

1. ¿Qué es R, RStudio, y *R Markdown*?
2. Introducción al uso de R Markdown
3. Demostración con ejemplo real de la Superintendencia del Medio Ambiente (SMA)

**Este no es un taller para aprender a programar en R**

Se busca mostrar el potencial de estas herramientas para la implementación en sus instituciones

# A lo que queremos llegar

## Minuta Región Metropolitana

Departamento de Gestión de la Información  
22/08/2019

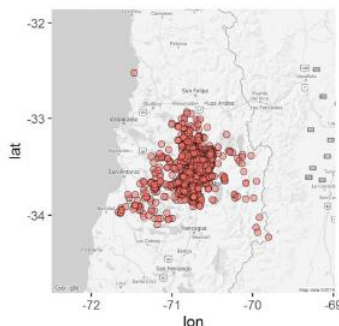
### 1. Sobre la información presentada <sup>1</sup>

A continuación, se presenta una minuta con diferentes indicadores asociados a la **Región Metropolitana**. Los datos son recopilados a partir de los diferentes sistemas de información disponibles de la Superintendencia del Medio Ambiente y consideran la información actualizada al 22/08/2019.

### 2. Unidades Fiscalizables

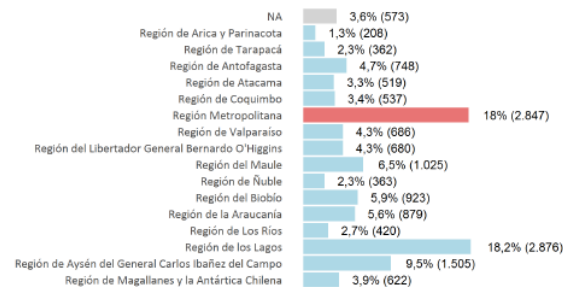
- 2.847 de las 15.773 Unidades Fiscalizables (UF) a nivel nacional corresponden a la **Región Metropolitana** (18%).
- Estas se concentran mayoritariamente en la categoría **Vivienda e Inmobiliarios** (727 UFs), seguida por la **Equipamiento** (676 UFs)

#### 2.1 Localización UFs en la Región Metropolitana

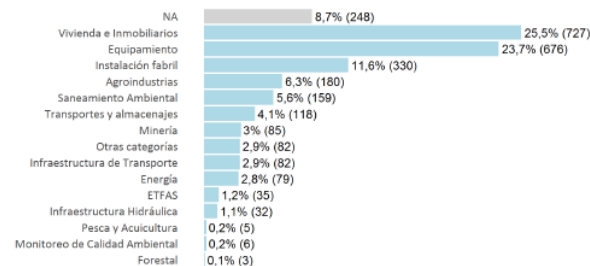


<sup>1</sup> UF: Unidad Fiscalizable / RCA: Resolución de Calificación Ambiental / IA: Inspección Ambiental / PPDA: Plan de Prevención y Descontaminación Ambiental / RSMA: Resolución de normas e instrucciones de carácter de la Superintendencia de Medio Ambiente / PC/PDC: Programa de Cumplimiento / NC: Norma de Contaminación / LEY: Ley Ambiental / REG: Reglamento / NE: Norma de Emisión / MP: Medida Provisional / FDC: Formulación de Cargos / MM: Millones

### 2.2 Distribución de UFs a nivel nacional



### 2.3 UFs de la Región Metropolitana por sector



### 2.4 Las 5 UFs de Región Metropolitana con más RCA

Unidad Fiscalizable	RCA
PRC VITACURA	34
PRC SANTIAGO	25
METRO S.A. LINEA 5	18
RELLENO SANITARIO SANTA MARTA	13
MINERA LA FLORIDA LTDA. - ALHUE	12
METRO S.A. LINEA 2	12
PRC PEÑALOEN	12

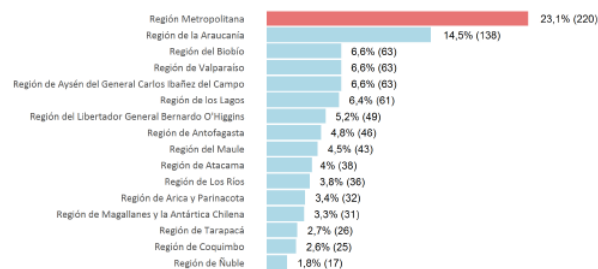
# A lo que queremos llegar

## 3. Formulaciones de cargo

### 6.1 Formulaciones de cargo (FDC)

- La Región Metropolitana ocupa el lugar 1 respecto al resto de las regiones en número de FdC.
- De 951 FdC, 220 (23,1%) han sido en la Región Metropolitana.
- Dentro de la Región Metropolitana, 74,1% de las FdC fueron iniciadas a partir de denuncias.
- Equipamiento es el sector que concentra el mayor número de FdC en la Región Metropolitana (82) seguida por Agroindustrias (34).

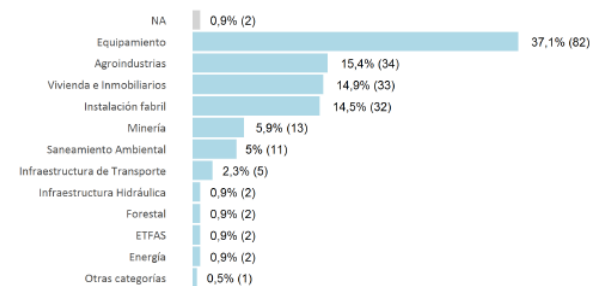
#### FDC por región



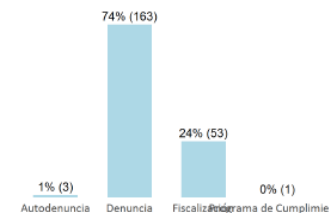
#### FDC asociadas a la Región Metropolitana por año



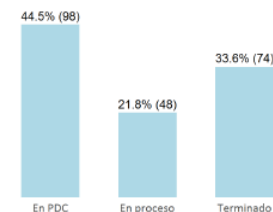
#### Distribución de FDC asociadas a la Región Metropolitana por sector



#### FDC asociadas a la Región Metropolitana por origen



#### FDC asociadas a la Región Metropolitana por estado



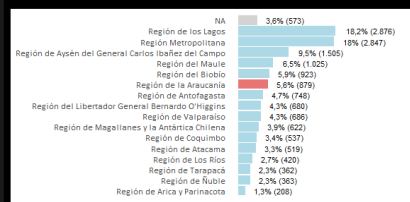
# Por lo que tenemos que pasar

```
Reporte_Región_Julio.Rmd
1:
2: author: Departamento de Gestión de la Información
3: output:
4:   word_document:
5:     reference_docx: FormatoreporteReg.docx
6: params:
7:   reg1: "Araucanía" #reemplazar con las opciones de abajo
8:   #noconsiderar: 2019 #poner 0 para considerar todo
9:   ---
10:
11:   {r, include=FALSE, include=FALSE}
12:   reg1 <- c("Región de Tarapacá"
13:     "Región de Antofagasta"
14:     "Región de Atacama"
15:     "Región de Coquimbo"
16:     "Región del Libertador General Bernardo O'Higgins"
17:     "Región de Valparaíso"
18:     "Región del Maule"
19:     "Región del Biobío"
20:     "Región de la Araucanía"
21:     "Región de Los Lagos"
22:     "Región de Aysén del General Carlos Ibañez del Campo"
23:     "Región de Magallanes y la Antártica Chilena"
24:     "Región Metropolitana"
25:     "Región de Los Ríos"
26:     "Región de Arica y Parinacota"
27:     "Región de Ñuble")
28:
29:   reg2 <- c("Tarapacá"
30:     "Antofagasta"
31:     "Atacama"
32:     "Coquimbo"
33:     "O'Higgins"
34:     "Valparaíso"
35:     "Maule"
36:     "Biobío"
37:     "Araucanía"
38:     "Lagos"
39:     "Aysén"
40:     "Magallanes"
41:     "Metropolitana"
42:     "Ríos"
43:     "Arica"
44:     "Ñuble")
45:
46:   reg3 <- as.data.frame(cbind(reg1,reg2))
47:
48:   reg <- as.character(reg1[reg3$reg2 == params$reg, 1]) #esto facilita la impresión desde "ImprimirReportes.Región.R"
49:
50:   ---
```

```
Reporte_Región_Julio.Rmd
507:
508: UP_1 <- resumen_reg %>% filter("Región" == reg) %>% summarise(
509:   UF_2 <- resumen_reg %>% summarise(
510:     RCA_1 <- detalle_unidadfiscalizableinstrumento %>%
511:       filter(SiglaInstrumento == "RCA", #Instrumento de Inter?
512:         EstadoRegistro == 1, #siempre considerar con tablas "detalle"
513:         NombreRegion == reg #filtro region
514:         #existen RCAs "interregionales" que en "NombreRegionProyecto" especifican región (acá se consideran
515:         | (NombreRegion == "Interregional" & NombreRegionProyecto == reg)) %>% summarise(RCA = n())
516:     RCA_2 <- detalle_unidadfiscalizableinstrumento %>%
517:       filter(SiglaInstrumento == "RCA", #Instrumento de Inter?
518:         EstadoRegistro == 1) %>% summarise(RCA = n())
519:     Denuncias_1 <- denuncias_año %>% filter(RegionSelect == reg) %>% summarise(Denuncias_2 = denuncias_año %>% summarise(
520:     Fisc_1 <- datos_fiscalizaciones_reg %>% filter("Región" == reg) %>% summarise(
521:     Fisc_2 <- datos_fiscalizaciones_reg %>% filter(Region == reg) %>% summarise(
522:     FDC_1 <- FDC_reg_porcentaje %>% filter(Region == reg) %>% summarise(
523:     FDC_2 <- FDC_reg_porcentaje %>% filter(Region == reg) %>% summarise(
524:     PDC_aprobado_1 <- PDC_estado %>% filter(Estado == "Aprobados") %>% summarise(
525:     PDC_rechazado_1 <- PDC_estado %>% filter(Estado == "Rechazados") %>% summarise(
526:     PDC_aprobado_2 <- datos %>% distinct(Expediente, AñoEstado, Estado) %>%
527:     group_by(Año = AñoEstado, Estado) %>% summarise(PDC = n()) %>%
528:     filter(!is.na(Año)) %>% filter(Estado == "Aprobados") %>% summarise(
529:     PDC_rechazado_2 <- datos %>% distinct(Expediente, AñoEstado, Estado) %>%
530:     group_by(Año = AñoEstado, Estado) %>% summarise(PDC = n()) %>%
531:     filter(!is.na(Año)) %>% filter(Estado == "Rechazados") %>% summarise(
532:     Monto_PDC_1 <- PDCMontosMontoPDC(MM CLP) %>% summarise(
533:     Monto_PDC_2 <- datos %>% select(Nombre, MontoPDC_CLP, AñoPDCAprobación) %>%
534:     arrange(desc(MontoPDC_CLP)) %>% group_by(AñoPDCAprobación) %>%
535:     summarise(MontoPDC(MM CLP) = round(sum(MontoPDC_CLP, na.rm = TRUE)/1000000) %>%
536:     filter(!is.na(AñoPDCAprobación), MontoPDC(MM CLP) != 0) %>% summarise(
537:     multas_firme_1 <- multas_total %>% filter(TipoMulta == "firme") %>% summarise(
538:     multas_nofirme_1 <- multas_total %>% filter(TipoMulta == "No firme") %>% summarise(
539:     ---
```

```
Reporte_Región_Julio.Rmd
971:
972: group_by(TipoMulta) %>% summarise(UTA = sum(MultasTotalUTA, na.rm = TRUE)) %>% filter(TipoMulta == "No firme") %>% summarise(
973:   multas_cursadas_1 <- multas_firme_1 + multas_nofirme_1
974:   multas_cursadas_2 <- multas_firme_2 + multas_nofirme_2
975:
976:   variable <- c("Unidades Fiscalizables (UF)", "RCAs", "Denuncias", "Expedientes de Fiscalización", "Formulaciones de Cargo",
977:     "Multas no firmes (UTA)")
978:   reg1 <- c(UF_1, RCA_1[1], Denuncias_1, Fisc_1, FDC_1, PDC_aprobado_1, PDC_rechazado_1, Monto_PDC_1, multas_cursadas_1,
979:     multas_firme_1, multas_nofirme_1)
980:   todos <- c(UF_2, RCA_2[1], Denuncias_2, Fisc_2, FDC_2, PDC_aprobado_2, PDC_rechazado_2, Monto_PDC_2, multas_cursadas_2,
981:     multas_firme_2, multas_nofirme_2)
982:   tab_resumen <- as.data.frame(cbind(variable, reg1, todos)) %>%
983:     rename("Indicador" = variable,
984:       "Región" = reg1,
985:       "Todas las regiones" = todos) %>%
986:     mutate("Región" = round(as.numeric(as.character("Región"))),
987:       "Todas las regiones" = round(as.numeric(as.character("Todas las regiones"))),
988:       "% región seleccionada" = round("Región"/"Todas las regiones"*100,1))
989:   colnames(tab_resumen)[c(2,4)] <- c(reg, paste0("% del total"))
990:
991:   tab_resumen %>%
992:     kable(format = "markdown", format.args = list(decimal.mark = ".", big.mark = ","))
993:
994:   ---
```

```
Reporte_Región_Julio.Rmd
1031:
1032: ## 3.2 Distribución de UPS a nivel nacional
1033:
1034: {r, fig.width=6, fig.height=2.9, fig.align="center"}
1035: resumen_reg %>%
1036:   ggplot(aes(reorder("Región", Porcentaje), Porcentaje, label = paste0(prettyNum(Porcentaje, big.mark = ".", decimal.mark =
1037:   ), "%", " (", prettyNum(UPS, big.mark = ".", decimal.mark = ","), ")"), fill = Destacar)) +
1038:   geom_bar() +
1039:   ylim(0, max(resumen_catesPorcentaje)*1.22 +
1040:   coord_flip() +
1041:   geom_text(hjust = -0.22, size = 3) +
1042:   scale_fill_manual(values = c("51" = "#E87474", "No" = "light blue", "512" = "light gray")) +
1043:   theme(panel.background = element_rect(fill = NA),
1044:     legend.position = "none",
1045:     plot.title = element_text(hjust = 0.5),
1046:     axis.title.x = element_blank(),
1047:     axis.title.y = element_blank(),
1048:     axis.ticks.x = element_blank(),
1049:     axis.ticks.y = element_blank(),
1050:     axis.text.x = element_blank(),
1051:     text = element_text(family = "callibri"))
1052:
1053:   ---
1054:
1055: ## 3.3 UPS de la 'r' reg' por sector
1056:
1057: {r, fig.width=6, fig.height=3.9, fig.align="center"}
1058: resumen_cate %>%
1059:   ggplot(aes(reorder("Categoría Económica", Porcentaje), Porcentaje, label = paste0(prettyNum(Porcentaje, big.mark = ".",
1060:   ---
```



# ¿Qué es ?

## Lenguaje y plataforma

- Lenguaje de programación estadística
- Herramienta de visualización de datos
- Gratuito

## Ecosistema

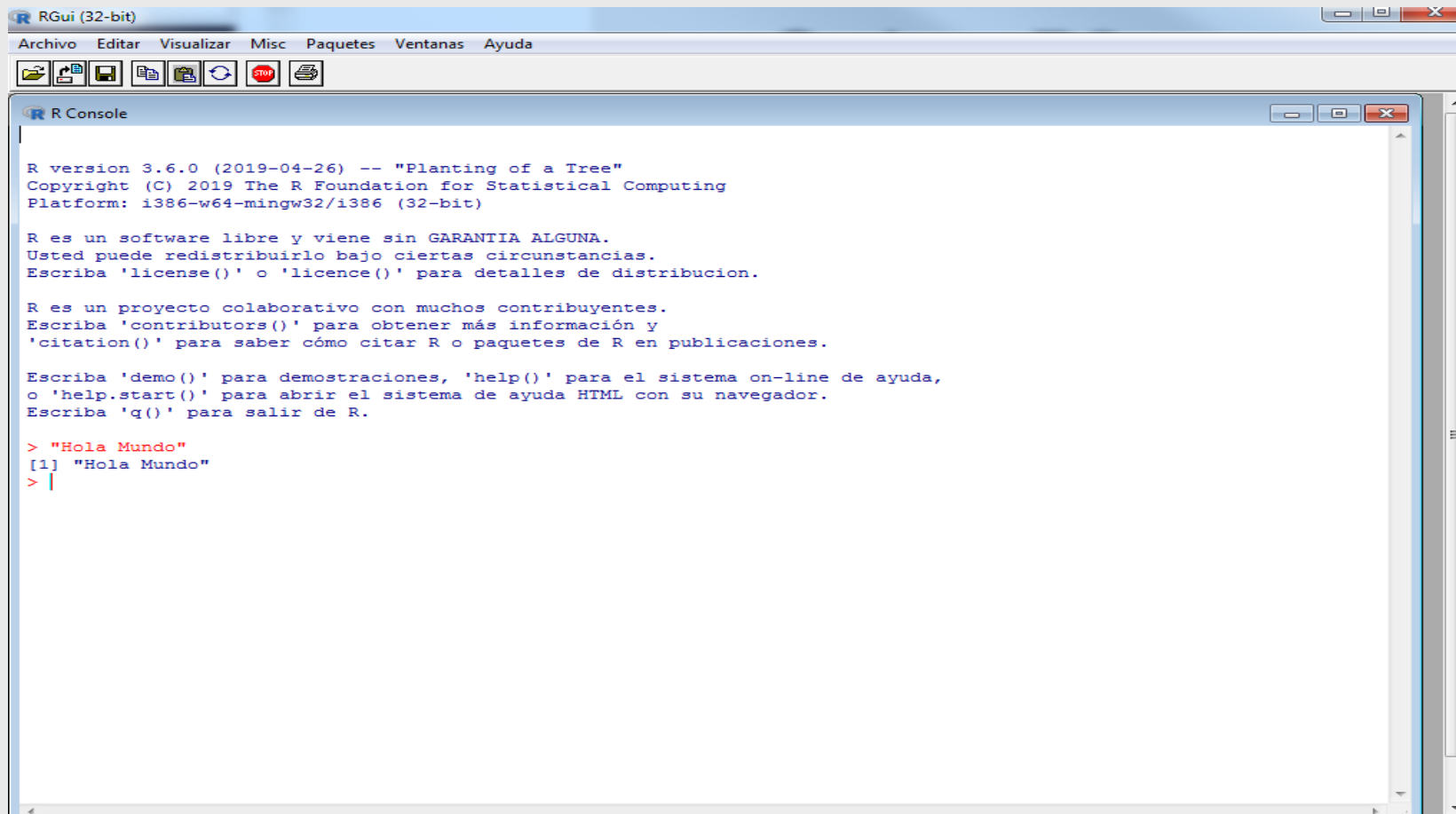
- Muchas aplicaciones e integraciones con otras plataformas
- 12.000+ librerías gratuitas disponibles

## Comunidad

- 2.5+ millones de usuarios
- Muchos y diversos grupos de usuarios a nivel mundial

<https://www.r-project.org/>

# ¿Qué es R?



```
RGui (32-bit)
Archivo  Editar  Visualizar  Misc  Paquetes  Ventanas  Ayuda

R Console

R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> "Hola Mundo"
[1] "Hola Mundo"
> |
```



# ¿Qué es Studio?

IDE para R: *Entorno de desarrollo integrado*

## Consola

## Editor de código

- Auto-completación
- Ayuda de sintaxis
- Ejecución directa

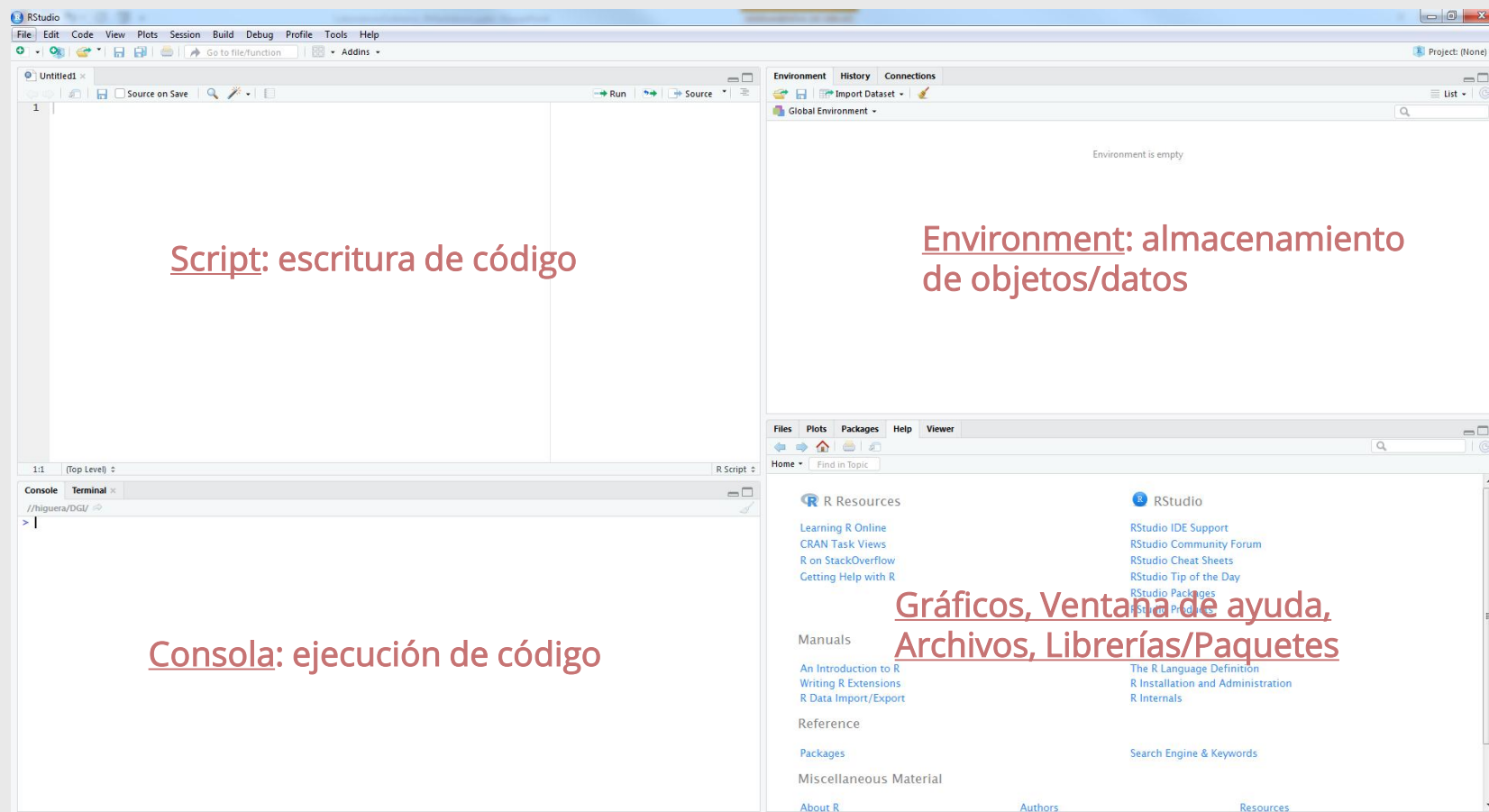
## Herramientas para distintas tareas

- Visualización
- Conexión con otras plataformas
- Depuración de código
- Manejo del ambiente de trabajo

<https://www.rstudio.com>



# ¿Qué es R Studio?

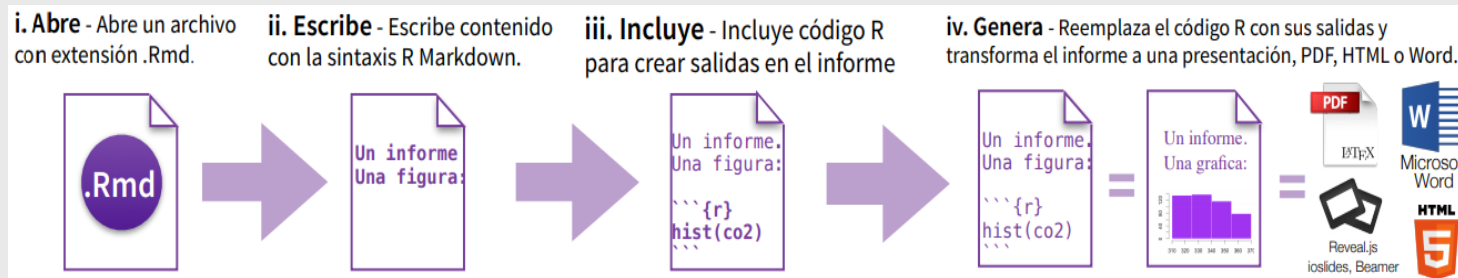


# ¿Qué es R Markdown?



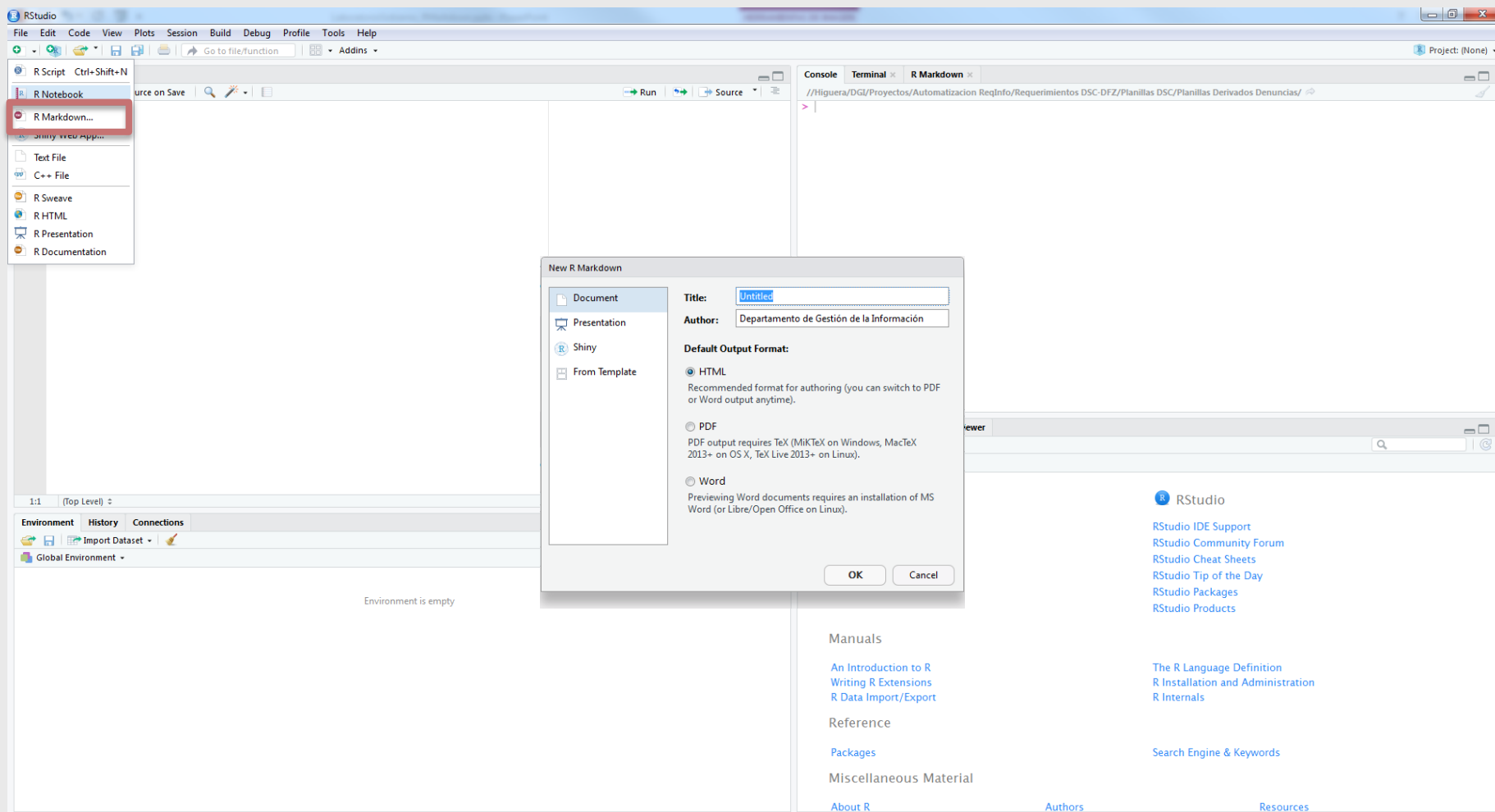
## Entorno para la creación de reportes/documentos

- Microsoft Word
- HTML
- PDF
- Power Point
- Y más...



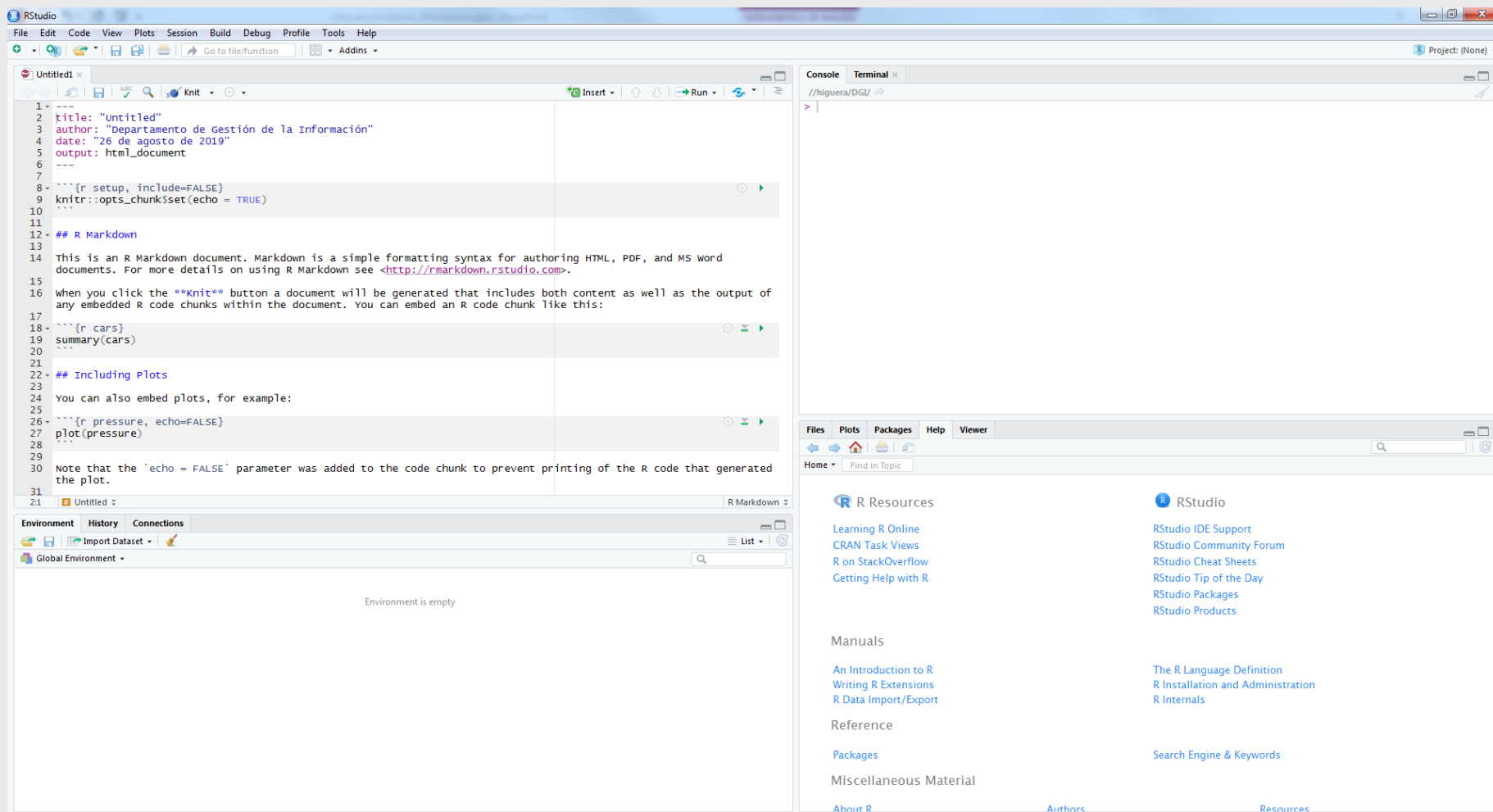
<https://rmarkdown.rstudio.com>

# ¿Qué es R Markdown?



The screenshot shows the RStudio IDE interface. In the top-left pane, the 'File' menu is open, and 'R Markdown...' is highlighted with a red rectangle. In the center, the 'New R Markdown' dialog box is displayed. The 'Document' option is selected in the left sidebar. The 'Title' field is 'Untitled', and the 'Author' field is 'Departamento de Gestión de la Información'. Under 'Default Output Format', the 'HTML' radio button is selected. The 'Word' option is also visible with a note: 'Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux)'. The 'OK' button is highlighted. In the bottom-right pane, the 'RStudio' sidebar is visible, showing links to 'RStudio IDE Support', 'RStudio Community Forum', 'RStudio Cheat Sheets', 'RStudio Tip of the Day', 'RStudio Packages', 'RStudio Products', 'Manuals', 'Reference', 'Miscellaneous Material', 'About R', 'Authors', and 'Resources'.

# ¿Qué es R Markdown?



The screenshot displays the RStudio IDE interface. The main editor window shows a new R Markdown document titled "Untitled1". The document content includes a YAML header with fields for title, author, date, and output format. Below the header, there is an R code chunk for setting knitr options and another for summarizing the 'cars' dataset. A third R code chunk is shown for plotting 'pressure' data, with a note explaining the 'echo = FALSE' parameter. The right sidebar shows the RStudio Resources page with links to learning resources, manuals, and reference materials.

**YAML Header:**

```
1 ---
2 title: "Untitled"
3 author: "Departamento de Gestión de la Información"
4 date: "26 de agosto de 2019"
5 output: html_document
6 ---
```

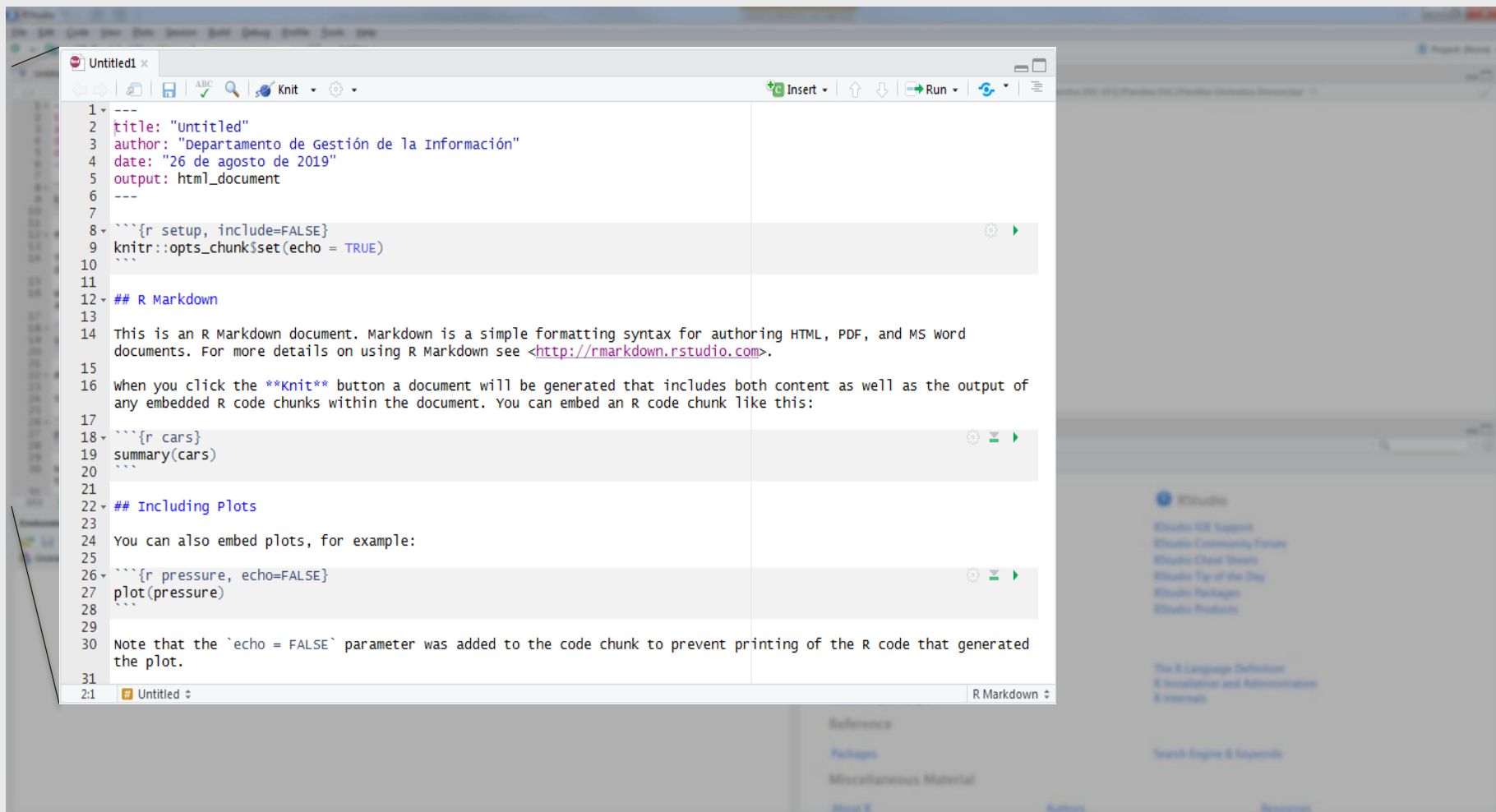
**R Code Chunks:**

```
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10
11 ## R Markdown
12
13 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word
14 documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
15
16 when you click the **knit** button a document will be generated that includes both content as well as the output of
17 any embedded R code chunks within the document. You can embed an R code chunk like this:
18
19 ```{r cars}
20 summary(cars)
21 ```
22
23 ## Including Plots
24
25 You can also embed plots, for example:
26
27 ```{r pressure, echo=FALSE}
28 plot(pressure)
29 ```
30
31 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated
32 the plot.
```

**RStudio Resources:**

- R Resources**
  - Learning R Online
  - CRAN Task Views
  - R on StackOverflow
  - Getting Help with R
- Manuals**
  - An Introduction to R
  - Writing R Extensions
  - R Data Import/Export
- Reference**
  - Packages
- Miscellaneous Material**
  - About R
- RStudio**
  - RStudio IDE Support
  - RStudio Community Forum
  - RStudio Cheat Sheets
  - RStudio Tip of the Day
  - RStudio Packages
  - RStudio Products
- Search Engine & Keywords**

# ¿Qué es R Markdown?



# ¿Qué es R Markdown?

## Metadatos

```
1 ---
2 title: "untitled"
3 author: "Departamento de Gestión de la Información"
4 date: "26 de agosto de 2019"
5 output: html_document
6 ---
```

```
7
8 ## R Markdown
9
10 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.
11
12 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:
13
14 ## Including plots
15
16 you can also embed plots, for example:
17
18 ## Including plots
19
20 you can also embed plots, for example:
21
22 ## Including plots
23
24 you can also embed plots, for example:
25
26 ## Including plots
27
28 you can also embed plots, for example:
29
30 note that the 'echo = FALSE' parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```





# ¿Qué es R Markdown?

The image shows a screenshot of an R Markdown document in a code editor. The document contains several code chunks and text blocks. Three callouts are present:

- Código**: Points to a code chunk starting with ````{r setup, include=FALSE}` and ending with `````. The code inside is `knitr::opts_chunk$set(echo = TRUE)`.
- pedazos de código**: Explains that a code chunk starts with ````{r}` and ends with `````. It shows a callout box with the text "Aqui hay código" and a code chunk ````{r} dim(iris)````. To the right, the output of the code is shown: `## [1] 150 5`.
- Aqui hay código**: Points to a code chunk starting with ````{r cars}` and ending with `````. The code inside is `summary(cars)`.

The document also includes a section titled `## Including plots` with a code chunk starting with ````{r pressure, echo=FALSE}` and ending with `````. The code inside is `plot(pressure)`. The document footer includes links to "Reference", "Packages", "Miscellaneous Material", "About R", "Authors", and "Resources".

# Introducción a R Markdown

```

1 ---
2 title: "Introducción a R Markdown"
3 author: "Pablo Aguirre Hörmann"
4 output:
5   html_document:
6     df_print: paged
7 ---
8
9 # R Markdown
10
11 Este es un archivo de tipo R Markdown (.Rmd), en este tipo de archivos
12 podemos de manera simple intercalar texto con código.
13
14 El código se define en los "pedazos de código" (code chunks):
15
16 ```{r}
17 # Una suma por ejemplo
18 2 + 2
19 ```
20
21 ```{r, fig.align='center', fig.width=5, fig.height=4}
22 # o un gráfico simple
23 x <- 1:10
24 y <- rnorm(10)
25 plot(x, y)
26 ```
27
28 Y estos "pedazos de código" tienen distintas opciones que por ejemplos nos
29 permiten solo mostrar el resultado y no el código que lo genera (echo = FALSE):
30
31 ```{r, echo=FALSE}
32 # Manejo de datos
33 
```

## Introducción a R Markdown

Pablo Aguirre Hörmann

### R Markdown

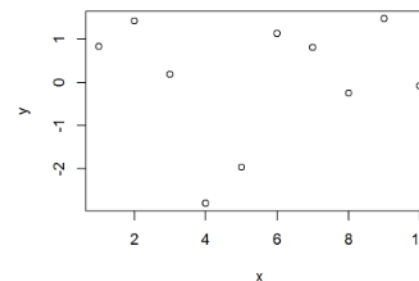
Este es un archivo de tipo *R Markdown* (`.Rmd`), en este tipo de archivos podemos de manera simple intercalar texto con código.

El código se define en los "pedazos de código" (*code chunks*):

```
# Una suma por ejemplo
2 + 2
```

```
## [1] 4
```

```
# o un gráfico simple
x <- 1:10
y <- rnorm(10)
plot(x, y)
```



Y estos "pedazos de código" tienen distintas opciones que por ejemplos nos permiten solo mostrar el resultado y no el código que lo genera (`echo = FALSE`):

```
50 ▾ # Título 1
51 ▾ ## Título 2
52 ▾ ### Título 3
53 ▾ #### Título 4
54 ▾ ##### Título 5
55 ▾ ##### Título 6
```



# Título 1

## Título 2

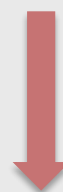
### Título 3

#### Título 4

##### Título 5


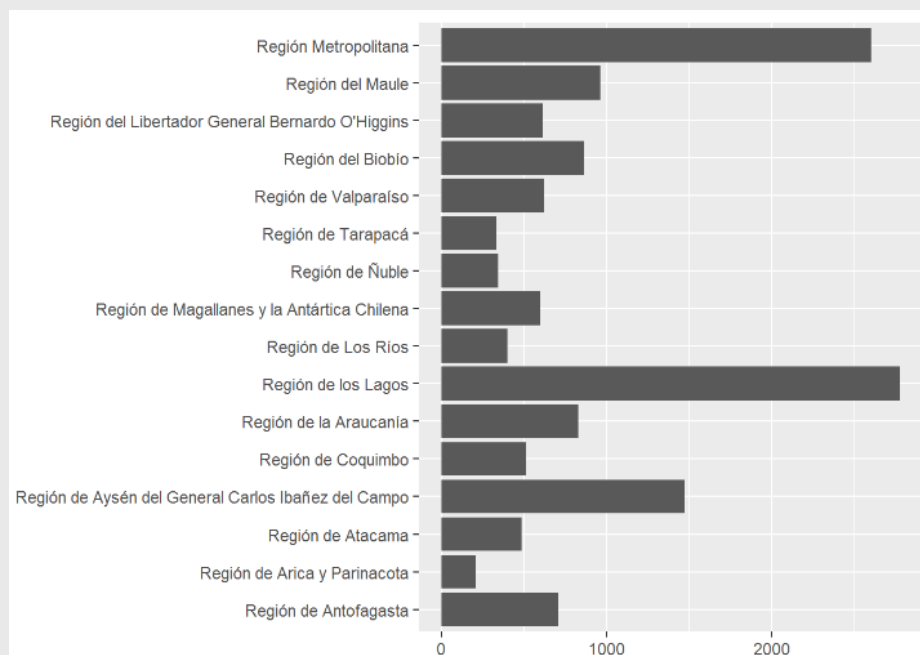
##### Título 6

```
40 - *cursiva* o _cursiva_  
41 - **negrita** o __negrita__  
42 - __*cursiva y negrita*__ o **_cursiva y negrita_**  
43 - ~tachado~  
44 - [link](https://portal.sma.gob.cl/)  
45 - `Objetos de código`  
46 - entre otras..
```



- *cursiva* o *cursiva*
- **negrita** o **negrita**
- ***cursiva y negrita*** o ***cursiva y negrita***
- tachado
- link
- `Objetos de código`
- entre otras..

```
185 Resumen_UnidadFiscalizable_reg %>%  
186   ggplot(aes(RegionNombre, `Número de UFs`)) +  
187   geom_col() +  
188   coord_flip() +  
189   xlab("") +  
190   ylab("")
```

 Knit

```
75 {r}
76 resultado <- 2 + 2
77 rel <- round(cor(x,y), 2)
78
79
80 La suma del cálculo hecho arriba es igual a resultado y la correlación entre x e y es rel.
```

 Knit

La suma del cálculo hecho arriba es igual a **4** y la correlación entre **x** e **y** es **0.45**.

# Ejemplo SMA

## Reportes de información por región

- Definir parámetros

- Cargar librerías

- Importar datos

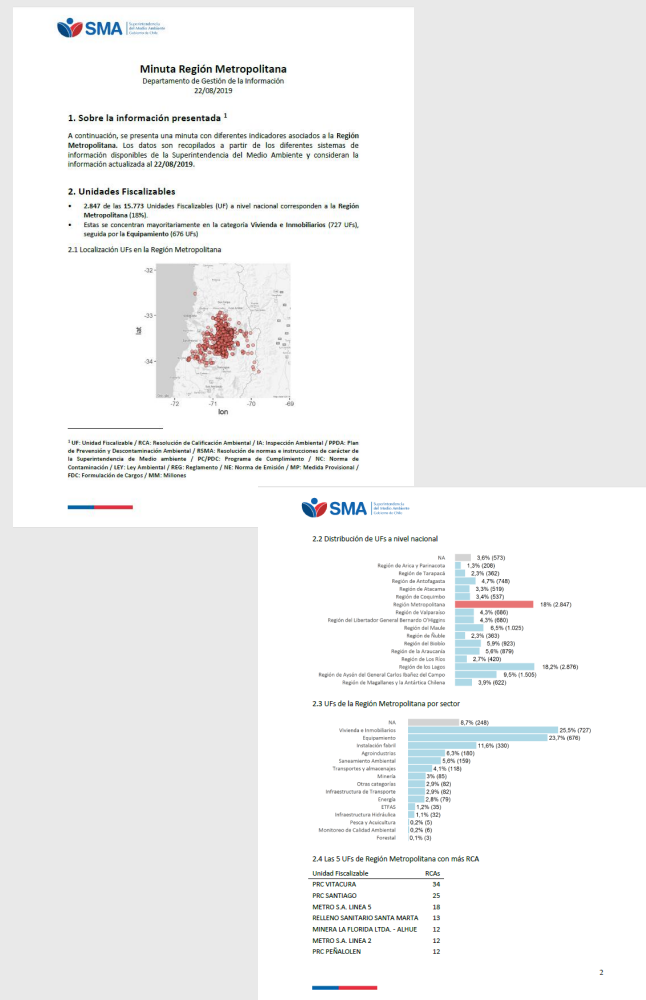
- Realizar modificaciones a los datos

- Generar cuerpo del reporte

Texto

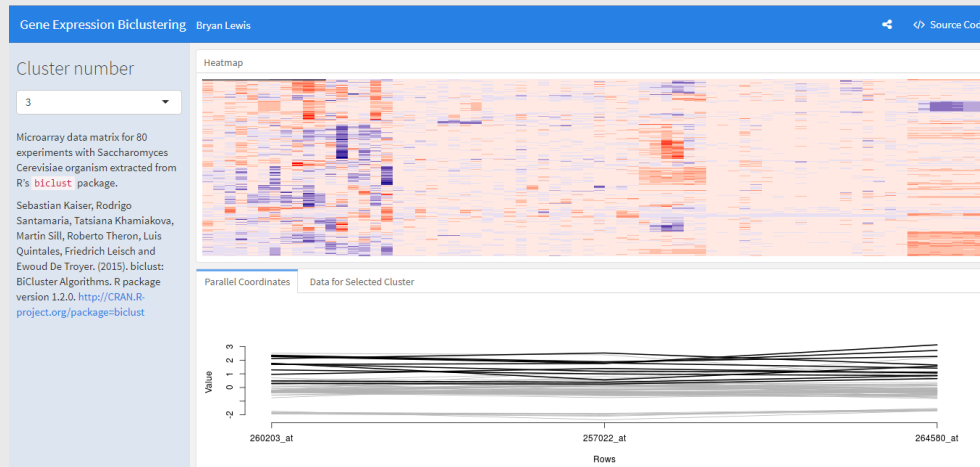
Tablas y gráficos

- Definir formato del documento final





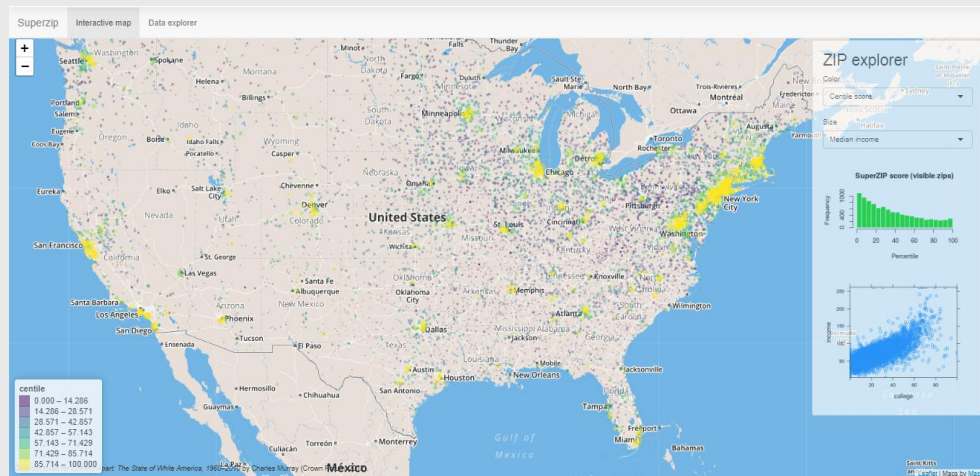
# Mucho más por delante...



flexdashboard for R

<https://ijallaire.shinyapps.io/shiny-biclust/>

Shiny  
by RStudio



<https://shiny.rstudio.com/gallery/superzip-example.html>

# Reportes automatizados con R Markdown

Pablo Aguirre Hörmann

Departamento de Gestión de la Información

[sebastian.elgueta@sma.Gob.cl](mailto:sebastian.elgueta@sma.Gob.cl) - [pablo.aguirre@sma.gob.cl](mailto:pablo.aguirre@sma.gob.cl)

26 de agosto de 2019

