



Translating datasets using  
{datalang}: the development of  
{datos} package for the R4DS  
Spanish translation 

Riva Quiroga

 @rivaquiroga



# #useR2019: the French keyboard experience



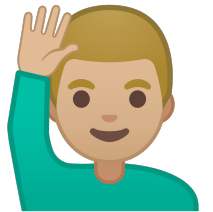
## #useR2019: the French keyboard experience

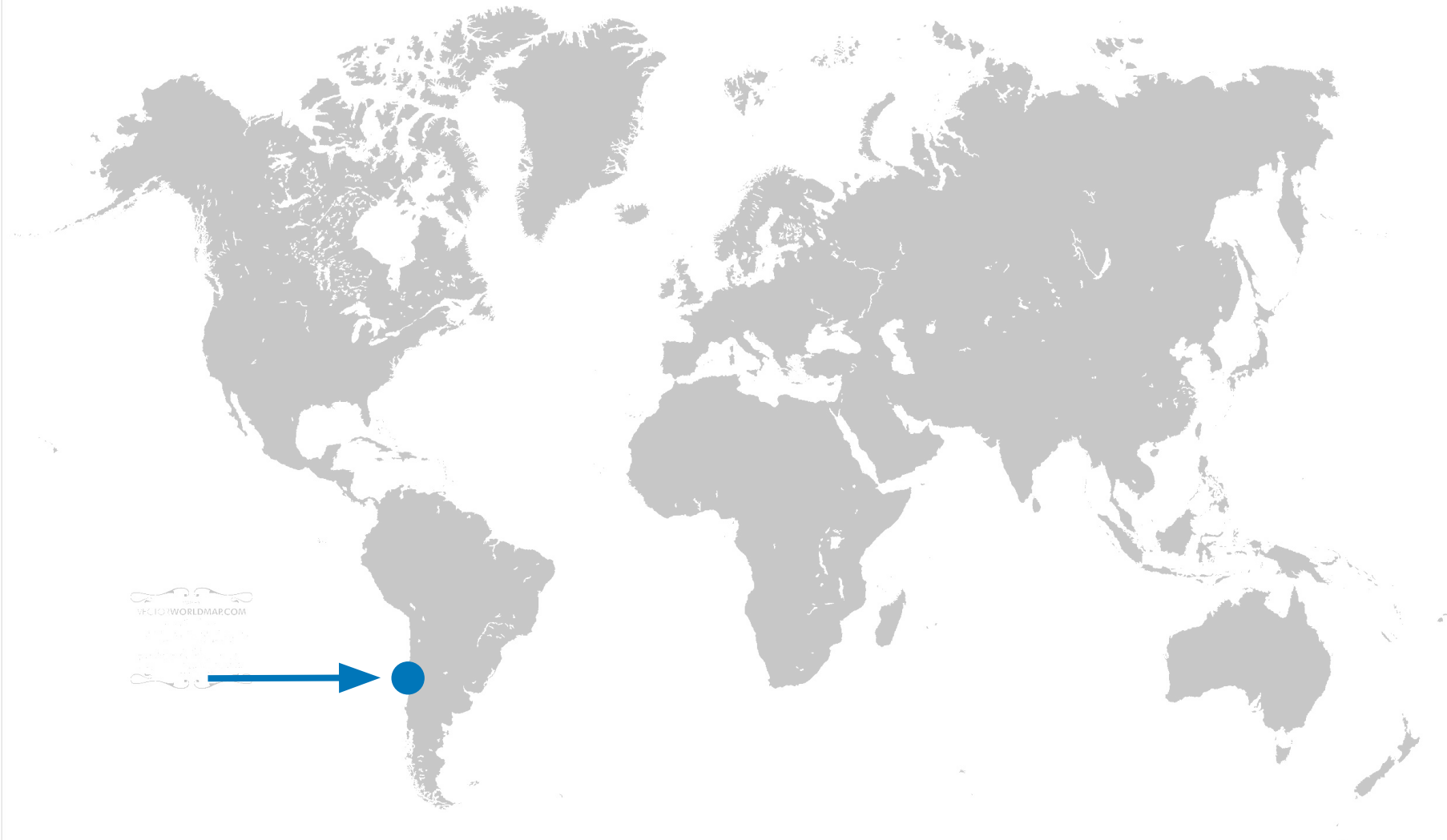


**> when technology  
is not built around your  
own language**

**language gap**

# language gap





©KJOWORLDMAR.COM

©KJOWORLDMAR.COM



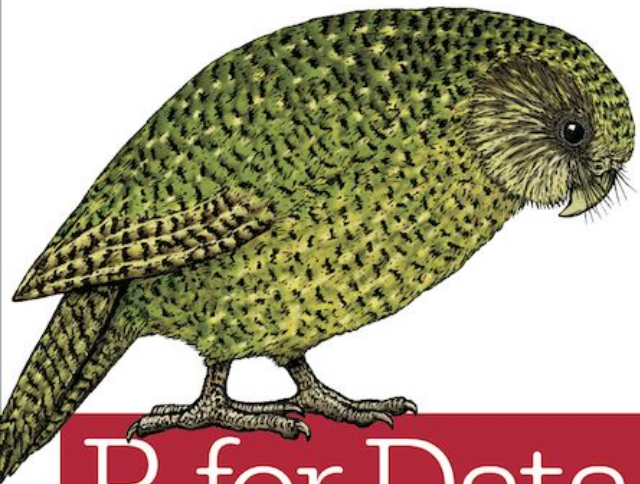
**500 dollars**  
**english course**

450 dollars  
**minimum wage**

**language gap** 

**happy R users**

O'REILLY®

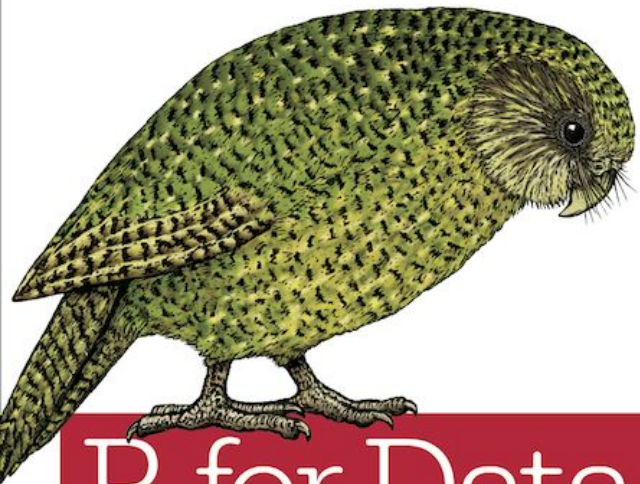


# R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &  
Garrett Golemund

O'REILLY®



# R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &  
Garrett Golemund

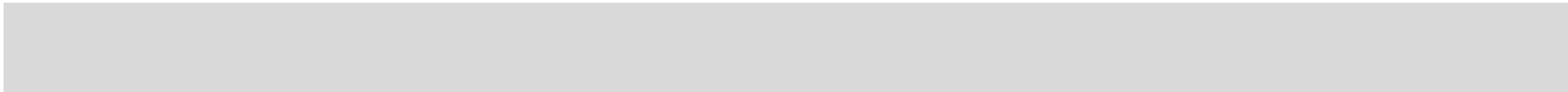
O'REILLY®



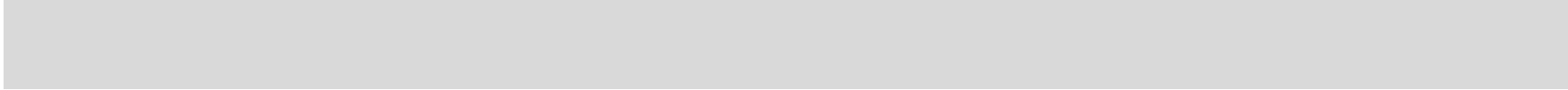
R PARA CIENCIA  
DE DATOS

IMPORT, TIDY, TRANSFORM, VISUALIZE, AND MODEL DATA

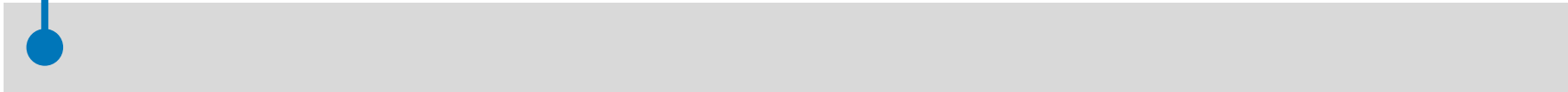
Hadley Wickham &  
Garrett Golemund



**2017**







2017



**Hadley Wickham**  @hadleywickham · 20 dic. 2017

Just learned that R for data science is being translated in to Chinese (simplified & complex), French, Korean, Portuguese, Russian and Serbian next year!

[#rstats](#)

 Traducir Tweet

 14

 41

 429



2017



**Hadley Wickham**  @hadleywickham · 20 dic. 2017

Just learned that R for data science is being translated in to Chinese (simplified & complex), French, Korean, Portuguese, Russian and Serbian next year!

[#rstats](#)

 Traducir Tweet

 14

 41

 429



**Marce** @Fichulina · 20 dic. 2017

What about Spanish? :)

 Traducir Tweet

 2



 2



2017



**Hadley Wickham**  @hadleywickham · 20 dic. 2017

Just learned that R for data science is being translated in to Chinese (simplified & complex), French, Korean, Portuguese, Russian and Serbian next year!

[#rstats](#)

 Traducir Tweet


 14  41  429 



**Marce** @Fichulina · 20 dic. 2017

What about Spanish? :)

 Traducir Tweet

 2   2 



**Hadley Wickham** 

@hadleywickham

Siguiendo

En respuesta a [@Fichulina](#)

Strangely, no. But let me ask about it

 Traducir Tweet



2017



**Hadley Wickham**  @hadleywickham · 20 dic. 2017

Just learned that R for data science is being translated in to Chinese (simplified & complex), French, Korean, Portuguese, Russian and Serbian next year!

[#rstats](#)

 Traducir Tweet


 14  41  429 



**Marce** @Fichulina · 20 dic. 2017

What about Spanish? :)

 Traducir Tweet

 2   2 



**Hadley Wickham** 

@hadleywickham

Siguiendo

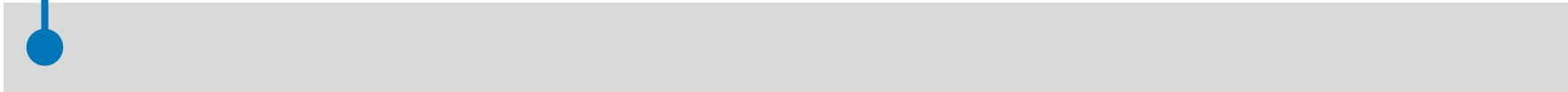
En respuesta a [@Fichulina](#)

Strangely, no. But let me ask about it

 Traducir Tweet



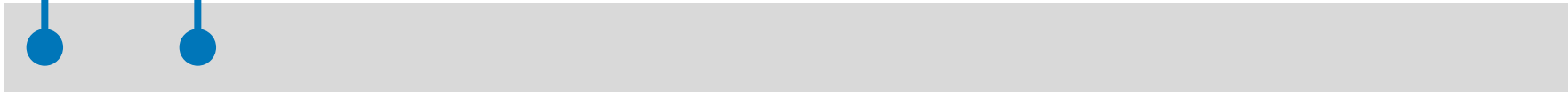
2017



2017



2018



## r4ds-es

Forked from hadley/r4ds

R for data science - Spanish translation

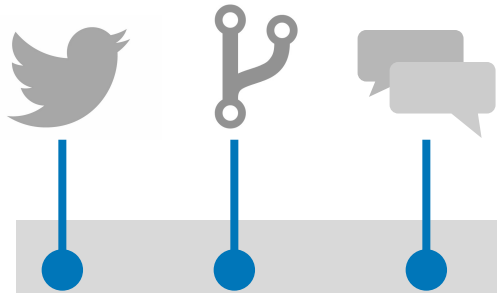
★ Star

● R ★ 5 🍴 2,439 📄 Other Updated on 25 Mar 2018

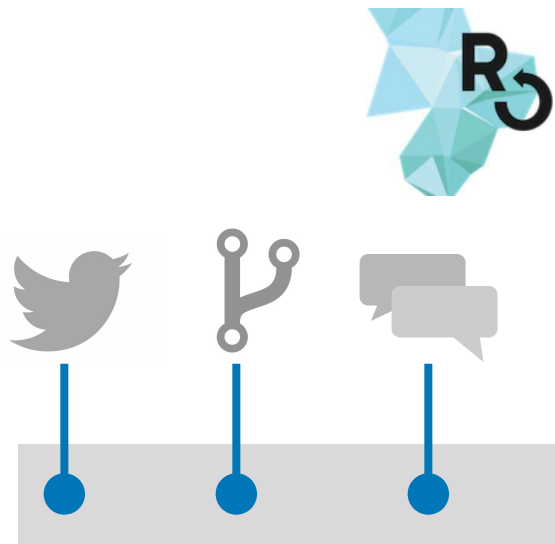


2018





2018



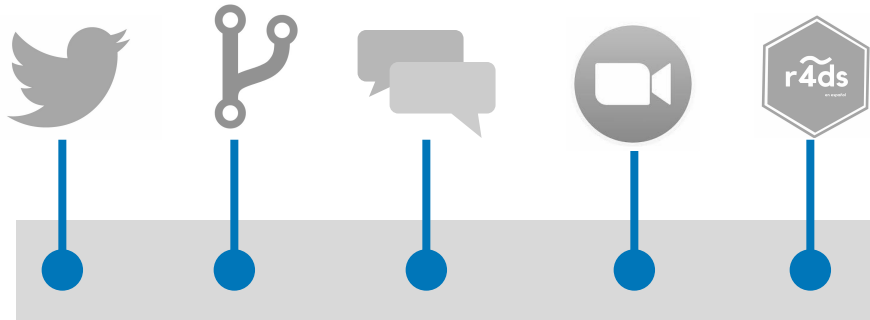
2018



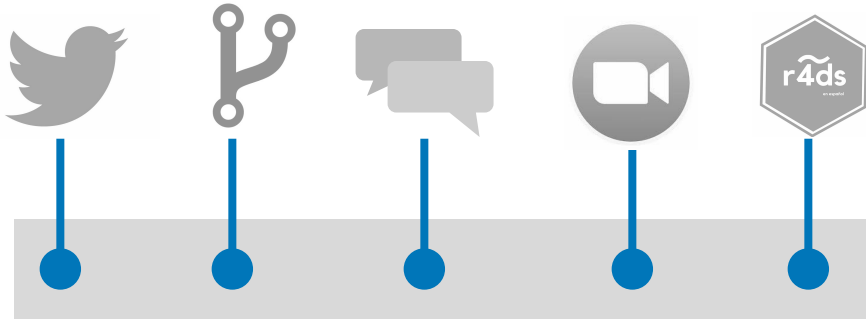
2018



2018



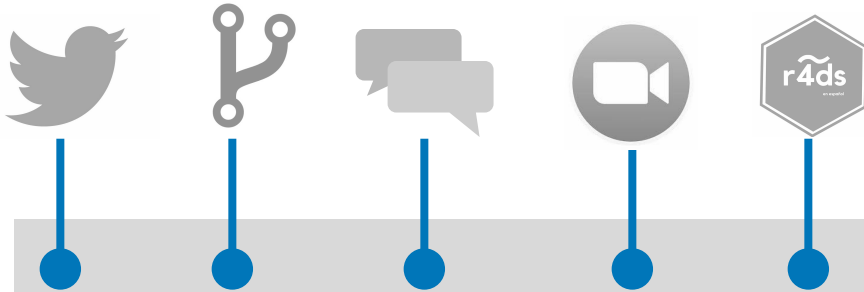
2018



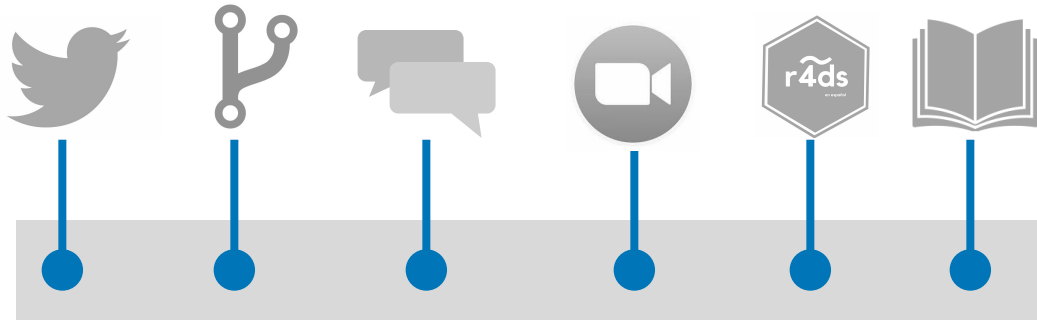
2018



**sticker-driven  
project**

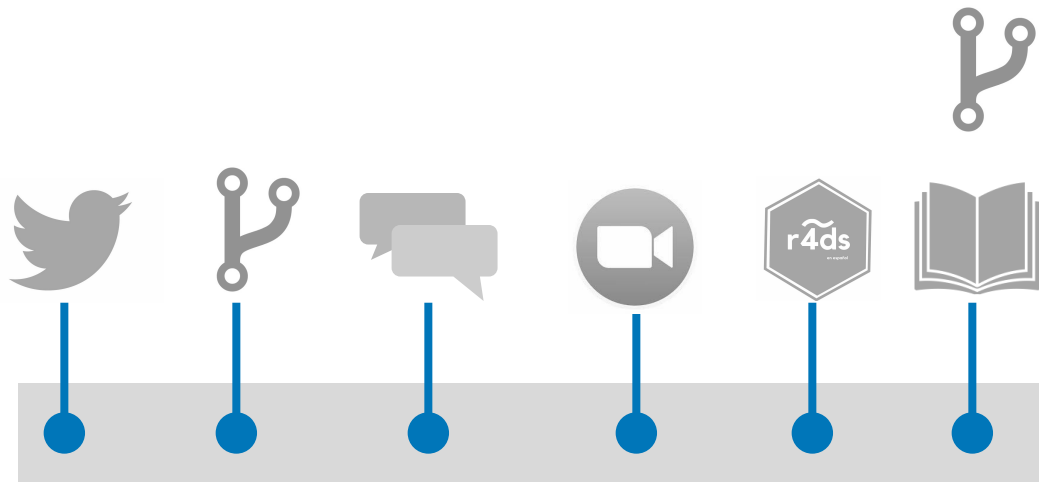


**2018**



2018





2018



Marcela  
Alfaro



Mónica  
Alonso



Fernando  
Álvarez



Zulemma  
Bazurto



Yanina  
Bellini



Juliana  
Benítez



Elio  
Campitelli



María Paula  
Caldas



Florencia  
DAndrea



Rocío  
Espada



Joshua  
Kunst



Patricia  
Loto



Pamela  
Matías



Lina  
Moreno



Paola  
Prieto



Riva  
Quiroga



Lucía  
Rodríguez



Pacha



Daniela  
Vázquez



Melina  
Vidoni



Roxana  
Villafañe

## the translators

**translators + reviewers**



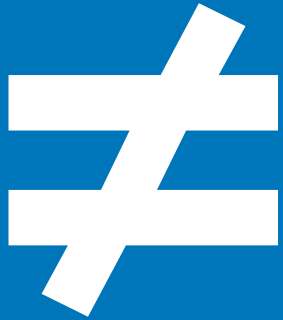
**varieties of Spanish**



**backgrounds**



**R experience**

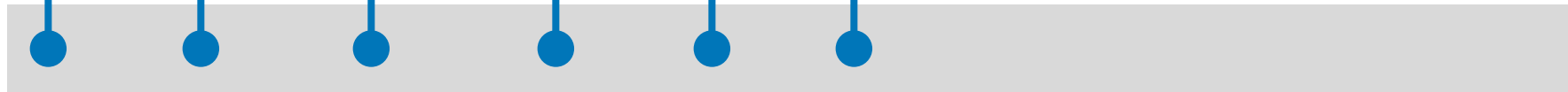


**github experience**



**github experience**

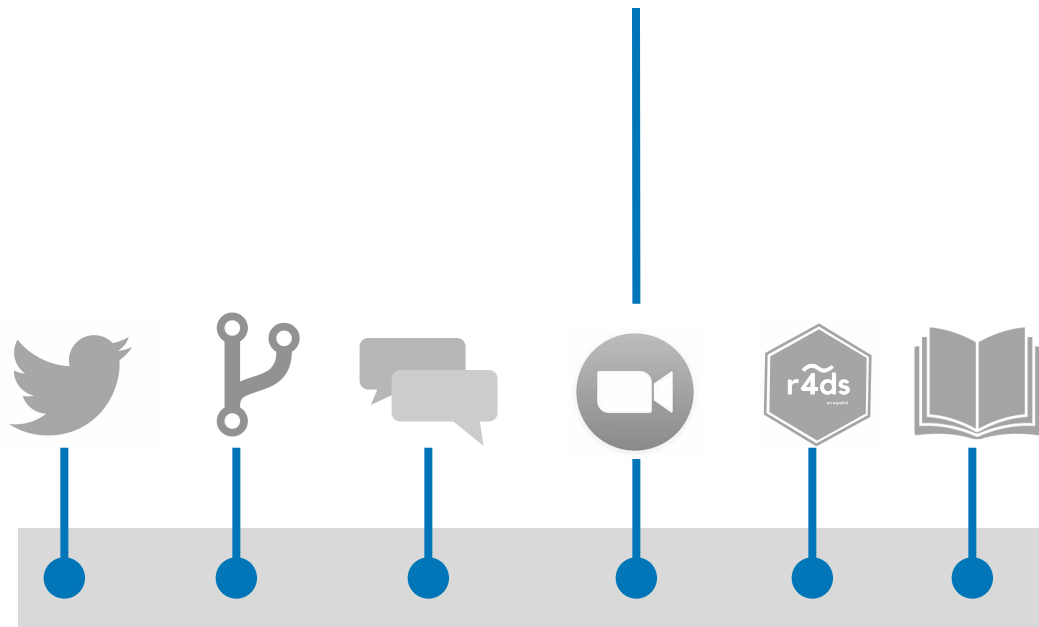




2017

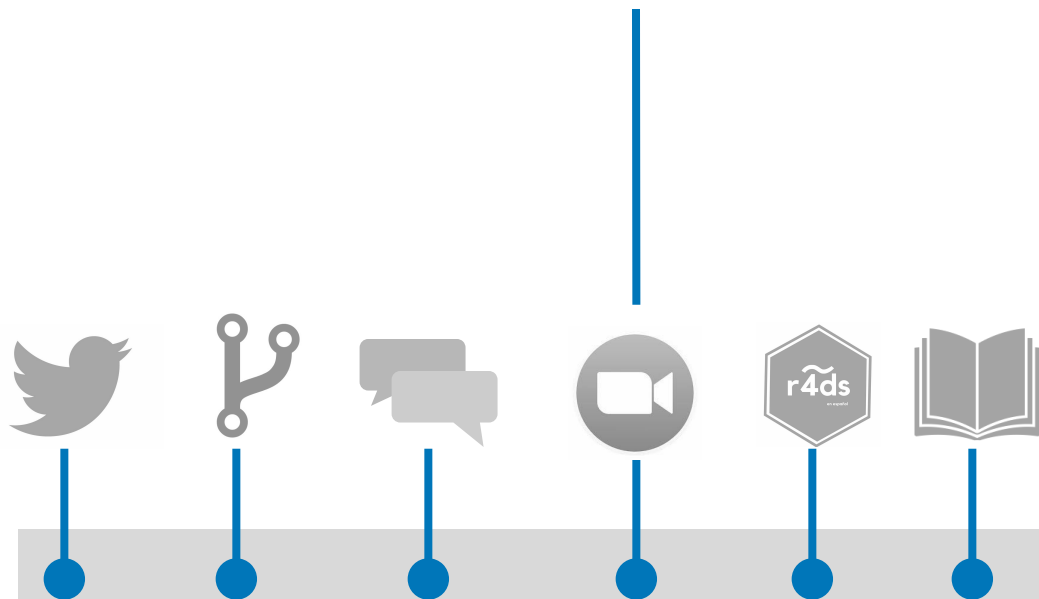
2018

# translation: text



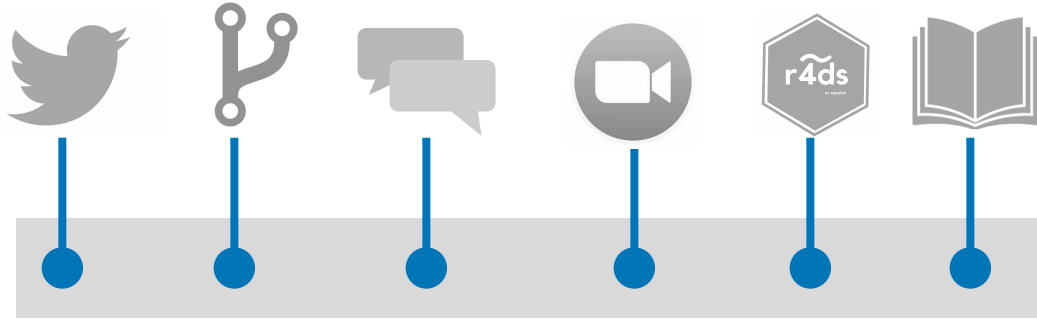
2017 2018

# translation: text + datasets

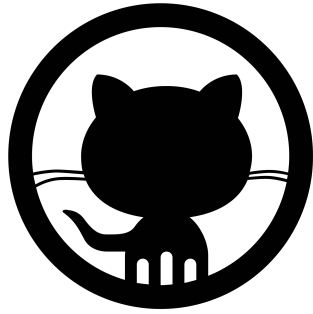


2017 2018

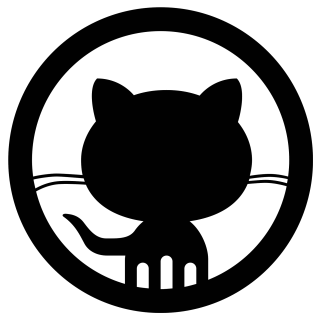
# infrastructure



2017 2018



# github.com/cienciadedatos

The screenshot displays the GitHub profile page for the user 'cienciadedatos'. At the top left is the 'r4ds' logo, a blue hexagon with a white bird-like shape and the text 'r4ds' and 'by Hadley'. To its right is the username 'cienciadedatos'. Below this is a navigation bar with tabs for 'Repositories 6', 'Packages', 'People 16', 'Teams', and 'Projects'. A search bar with the placeholder 'Find a repository...' and a 'Type: All' dropdown is positioned below the navigation. The main content area lists six repositories, each with a title, description, language, and statistics (forks, stars, issues, pull requests). Each repository entry also features a green activity line graph on the right side.

**cienciadedatos**

Repositories 6 Packages People 16 Teams Projects

Find a repository... Type: All

**datos-de-miercoles**  
Versión en español del proyecto #tidytuesday  
HTML 21 forks 36 stars 7 issues 3 pull requests Updated yesterday

**datos**  
Sets de datos para la traducción del libro R4DS  
R 6 forks 11 stars 5 issues 3 pull requests Updated on 5 Jun

**recursos-en-espanol**  
Directorio con recursos en español para aprender R  
0 forks 0 stars 0 issues 0 pull requests Updated on 4 Jun

**r4ds**  
Forked from hadley/r4ds  
Traducción del libro "R for Data Science"  
R 2,442 forks 15 stars 5 issues 5 pull requests Updated on 1 Jun

**documentacion-traducccion-r4ds**  
Documentación del proyecto de traducción colaborativa de R4DS y orientaciones para la traducción y revisión de capítulos  
12 forks 19 stars 1 issue 0 pull requests Updated on 1 Jun

# github.com/cienciadedatos



cienciadedatos

Repositories 6

Packages

People 16

Teams

Projects

Find a repository...

Type: All

## datos-de-miercoles

Versión en español del proyecto #tidytuesday

HTML 21 36 7 3 Updated yesterday



## datos

Sets de datos para la traducción del libro R4DS

R 6 11 5 3 Updated on 5 Jun



## recursos-en-espanol

Directorio con recursos en español para aprender R

0 0 0 0 Updated on 4 Jun



## r4ds

Forked from hadley/r4ds

Traducción del libro "R for Data Science"

R 2,442 15 5 5 Updated on 1 Jun



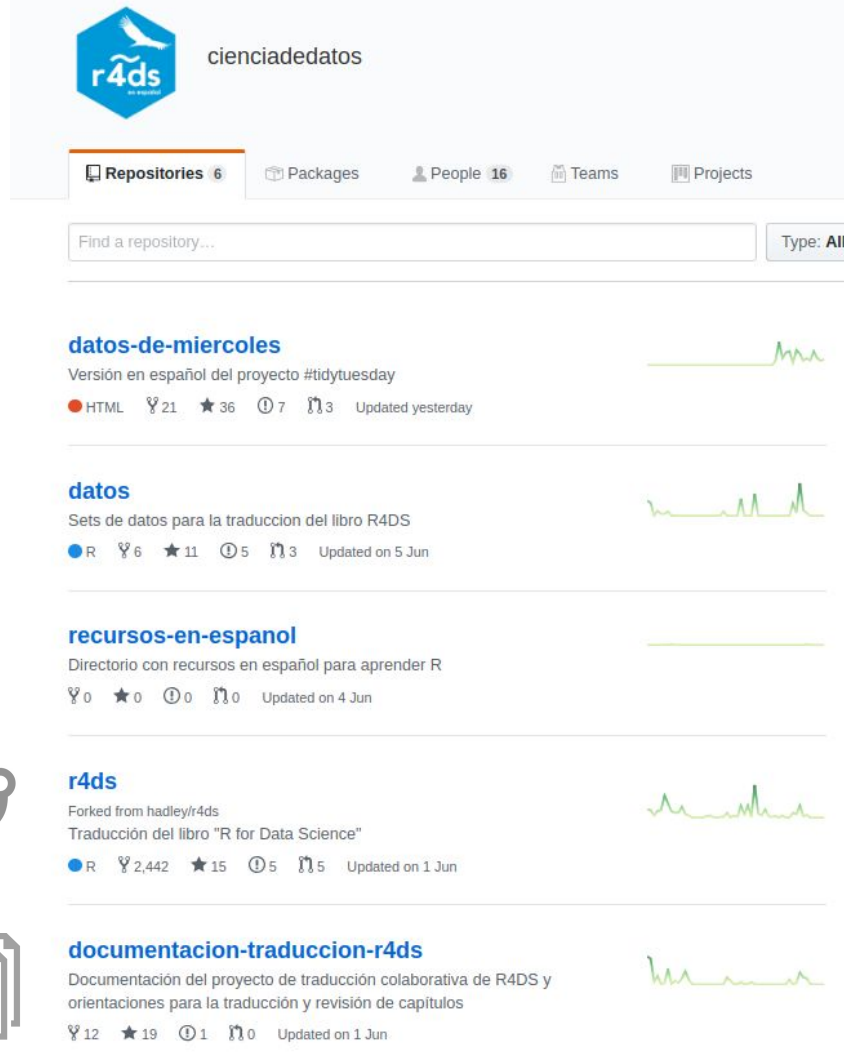
## documentacion-traducccion-r4ds

Documentación del proyecto de traducción colaborativa de R4DS y orientaciones para la traducción y revisión de capítulos

12 19 1 0 Updated on 1 Jun



# github.com/cienciadedatos



The screenshot shows the GitHub profile page for the user 'cienciadedatos'. At the top left is the 'r4ds' logo, a blue hexagon with a white bird-like shape and the text 'r4ds' and 'by Hadley'. To the right of the logo is the username 'cienciadedatos'. Below this are navigation tabs: 'Repositories 6', 'Packages', 'People 16', 'Teams', and 'Projects'. A search bar with the placeholder 'Find a repository...' and a 'Type: All' dropdown is located below the tabs. The main content area displays a list of repositories, each with a title, description, and statistics (language, forks, stars, issues, pull requests, and last update). The repositories listed are 'datos-de-miercoles', 'datos', 'recursos-en-espanol', 'r4ds', and 'documentacion-traduccion-r4ds'. Each repository entry includes a green line graph on the right side, representing the repository's activity over time.

**cienciadedatos**

Repositories 6 Packages People 16 Teams Projects

Find a repository... Type: All

**datos-de-miercoles**  
Versión en español del proyecto #tidytuesday  
● HTML 🍴 21 ★ 36 ⓘ 7 🏗️ 3 Updated yesterday

**datos**  
Sets de datos para la traducción del libro R4DS  
● R 🍴 6 ★ 11 ⓘ 5 🏗️ 3 Updated on 5 Jun

**recursos-en-espanol**  
Directorio con recursos en español para aprender R  
🍴 0 ★ 0 ⓘ 0 🏗️ 0 Updated on 4 Jun

**r4ds**  
Forked from hadley/r4ds  
Traducción del libro "R for Data Science"  
● R 🍴 2,442 ★ 15 ⓘ 5 🏗️ 5 Updated on 1 Jun

**documentacion-traduccion-r4ds**  
Documentación del proyecto de traducción colaborativa de R4DS y orientaciones para la traducción y revisión de capítulos  
🍴 12 ★ 19 ⓘ 1 🏗️ 0 Updated on 1 Jun





# github.com/cienciadedatos



cienciadedatos

Repositories 6

Packages

People 16

Teams

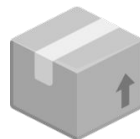
Projects

Type: All

## datos-de-miercoles

Versión en español del proyecto #tidytuesday

HTML 21 36 7 3 Updated yesterday



## datos

Sets de datos para la traducción del libro R4DS

R 6 11 5 3 Updated on 5 Jun



## recursos-en-espanol

Directorio con recursos en español para aprender R

0 0 0 0 Updated on 4 Jun



## r4ds

Forked from hadley/r4ds

Traducción del libro "R for Data Science"

R 2,442 15 5 5 Updated on 1 Jun



## documentacion-traducccion-r4ds

Documentación del proyecto de traducción colaborativa de R4DS y orientaciones para la traducción y revisión de capítulos

12 19 1 0 Updated on 1 Jun



[github.com/cienciadedatos](https://github.com/cienciadedatos)



**Edgar Ruiz**

[github.com/cienciadedatos](https://github.com/cienciadedatos)



**Edgar Ruiz**



**Mauricio Vargas**

[github.com/cienciadedatos](https://github.com/cienciadedatos)



**Edgar Ruiz**



**Mauricio Vargas**



**Riva Quiroga**

# traducción capítulo EDA - Pam #44

Edit

Open pamelinni wants to merge 4 commits into `cienciadatos:traduccion` from `pamelinni:chapter-explore-data`

Conversation 13 Commits 4 Checks 0 Files changed 1

+294 -293

Changes from all commits File filter... Jump to... Settings

0 / 1 files viewed

Review changes

587 EDA.Rmd

Viewed

@@ -1,595 +1,596 @@

1 -# Exploratory Data Analysis

1 +# Análisis exploratorio de datos (`_EDA_`)

2

2

3 -## Introduction

3 +## Introducción

4

4

5 -This chapter will show you how to use visualisation and transformation to explore your data in a systematic way, a task that statisticians call exploratory data analysis, or EDA for short. EDA is an iterative cycle. You:

5 +Este capítulo te mostrará cómo usar herramientas de visualización y transformación para explorar tus datos sistemáticamente, una tarea que las personas expertas en estadística suelen llamar análisis exploratorio de datos, o `_EDA_` (del inglés `_exploratory data analysis_`). El `_EDA_` es un ciclo iterativo, pues tú:



306 ■ ■ ■ model-building.Rmd

Viewed



32 library(lubridate)

33 ```

34

35 **-## Why are low quality diamonds more expensive? {#diamond-prices}**

36

37 **-In previous chapters we've seen a surprising relationship between the quality of diamonds and their price: low quality diamonds (poor cuts, bad colours, and inferior clarity) have higher prices.**

32 library(lubridate)

33 ```

34

35 **+++ ¿Por qué los diamantes de baja calidad son más caros? {#diamond-prices}**

36

37 **+En el capítulo anterior vimos una sorprendente relación entre la calidad de los diamantes y su precio: baja calidad de diamantes (cortes pobres, colores malos, y claridad inferior) tienen más altos precios.**



**pmnatural** on 14 Nov 2018 Member



baja calidad de diamantes > los diamantes de baja calidad



Reply...

Resolve conversation

38

39 ```{r dev = "png"}

40 **-ggplot(diamonds, aes(cut, price)) + geom\_boxplot()**

41 **-ggplot(diamonds, aes(color, price)) + geom\_boxplot()**

42 **-ggplot(diamonds, aes(clarity, price)) + geom\_boxplot()**

38

39 ```{r dev = "png"}

40 **+ggplot(diamantes, aes(corte, precio)) + geom\_boxplot()**

41 **+ggplot(diamantes, aes(color, precio)) + geom\_boxplot()**


42 **+ggplot(diamantes, aes(claridad, precio)) + geom\_boxplot()**



r4ds-es ▾



 Riva

 Todos los hilos

Canales



# aleatorio

# archivos-útiles

# de-donde-eres

# dudas-github

# dudas-proceso-trabajo

# dudas-revisión

# dudas-traducción

# general

# latin-r


# paquete-datos



r4ds-es ▾



 Riva

 Todos los hilos

Canales 

# aleatorio

# archivos-útiles

# de-donde-eres

# dudas-github

# dudas-proceso-trabajo

# dudas-revisión

# dudas-traducción

# general

# latin-r


# paquete-datos

# polls!

r4ds-es ▾



 Riva

 Todos los hilos

Canales 

# aleatorio

# archivos-útiles

# de-donde-eres

# dudas-github

# dudas-proceso-trabajo

# dudas-revisión

# dudas-traducción

# general

# latin-r

# paquete-datos

mtcars 



# aleatorio

# archivos-útiles

# de-donde-eres

# dudas-github

# dudas-proceso-trabajo

# dudas-revisión

# dudas-traducción

# general

# latin-r

# paquete-datos

# mtcars



**Simple Poll** APP 10:27

¿Cómo traducir 'cars'?

1 autos 12

@elio, @Joshua, @yabellini, @Patri Loto, @Florencia D'Andrea, @Laura Ación, @Monica Alonso, @Maria Paula, @Roxana, @pmoracho, @Priscilla Minotti, @Ana Diedrichs

2 coches 5

@yabellini, @Melina, @Laura Ación, @Monica Alonso, @Maria Paula

3 carros 1

@Zulemma



# aleatorio

# archivos-útiles

# de-donde-eres

# dudas-github

# dudas-proceso-trabajo

# dudas-revisión

# dudas-traducción

# general

# latin-r

# paquete-datos

# mtcars 🚗 -> mtautos



**Simple Poll** APP 10:27

¿Cómo traducir 'cars'?

1 autos 12

@elio, @Joshua, @yabellini, @Patri Loto, @Florencia D'Andrea, @Laura Ación, @Monica Alonso, @Maria Paula, @Roxana, @pmoracho, @Priscilla Minotti, @Ana Diedrichs

2 coches 5

@yabellini, @Melina, @Laura Ación, @Monica Alonso, @Maria Paula


3 carros 1

@Zulemma

r4ds-es ▾



 Riva

 Todos los hilos

Canales 

# aleatorio

# archivos-útiles

# de-donde-eres

# dudas-github

# dudas-proceso-trabajo

# dudas-revisión

# dudas-traducción

# general

# latin-r

# paquete-datos

%>%



Todos los hilos

Canales

# aleatorio

# archivos-útiles

# de-donde-eres

# dudas-github

# dudas-proceso-trabajo

# dudas-revisión

# dudas-traducción

# general

# latin-r

# paquete-datos

%>%



Riva 09:24

/poll "Ustedes dicen?" "un pipe" "una pipe" "lo traduzco"



Simple Poll APP 09:24

Ustedes dicen?

1 un pipe 11

@Joshua, @Pao, @Edgar, @pacha, @Marce, @Monica Alonso, @yabellini, @Patri Loto, @Daniela Vázquez, @Pam, @Melina

2 una pipe 1

@Riva

3 lo traduzco



# aleatorio

# archivos-útiles

# de-donde-eres

# dudas-github

# dudas-proceso-trabajo

# dudas-revisión

# dudas-traducción

# general

# latin-r

# paquete-datos

%>%



Riva 09:24

/poll "Ustedes dicen?" "un pipe" "una pipe" "lo traduzco"



Simple Poll APP 09:24

Ustedes dicen?

1 un pipe 11

@Joshua, @Pao, @Edgar, @pacha, @Marce, @Monica Alonso, @yabellini, @Patri Loto, @Daniela Vázquez, @Pam, @Melina

2 una pipe 1

@Riva

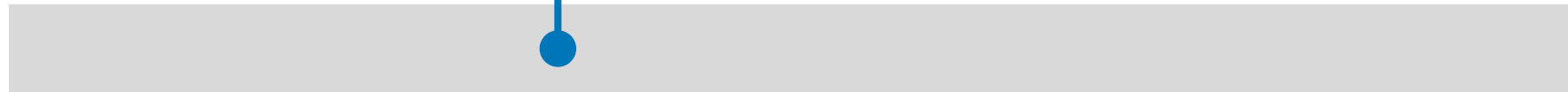


3 lo traduzco



Delete Poll

# translation: text + datasets



2017 2018



translation: **text** + datasets



2017 2018



# THE CARPENTRIES

Search or jump to... 7 Pull requests Issues Marketplace Explore

Carpentries-ES / board Watch 21 Star 4 Fork 7

Code Issues 3 Pull requests 2 Projects 0 Wiki Insights

Branch: master board / Convenciones\_Traduccion.md Find file Copy path

raynamharris Update Convenciones\_Traduccion.md 788df1f on 5 Apr

12 contributors

132 lines (118 sloc) | 4.92 KB Raw Blame History

## Lineamientos de traducción

Este será un documento compartido entre todos los traductores, donde podemos acceder a los lineamientos de traducción. Por favor vayan agregando algunos lineamientos que consideren apropiados. Si necesitan discutir algún alineamiento contactarse con Hely o Paula.

### Atención

- Idioma oficial: Español de latinoamerica (en si, por que la mayoría de los traductores son de latinoamerica)
- Esta es una traducción en tono de voz informal
- Vamos a tutear en las traducciones, decir "Tú" en vez de "vos" o "usted" y todas las conjugaciones de la persona "tú"
- Usar "ustedes" en vez de "vosotros" y todas las respectivas conjugaciones
- Ser cuidadosos con los acentos o tildes
- Abrir y cerrar símbolos de pregunta y exclamación
- Ortografía
- Por defecto las traducciones llevarán género femenino (la-las-investigadoras-programadoras-etc) a fin de equilibrar el sesgo existente en la mayoría de los documentos, que usan el género masculino
- Todos los números del cero al diez se escriben literalmente

### NO se traduce

- Los títulos en la cabecera de cada capitulo no se traducen, son símbolos de estilo (**title**, **objectives**, **keypoints**, etc.)
- `{:.callout}` y `{:.challenge}` Son símbolos de estilo
- Nombres de paquetes y funciones en R como `ggplot2`, `tidyr`, `dplyr`
- Nombres de variables en ejemplos y en general el código encerrado entre `~~~` `~~~` `{:.bash}`, `{:.r}`. Nota: Los comentarios dentro del código pueden traducirse si ven que el traducirlos dejaría mas claro las instrucciones.
- Nombres de libros (referencias o citas)
- Nombres de programas
- Nombres de directorios o **paths**
- Nombres de la organización, por ejemplo **The Carpentries**, **Software Carpentry** and **Data Carpentry**

## Enter *The Programming Historian*

(The initial English version)

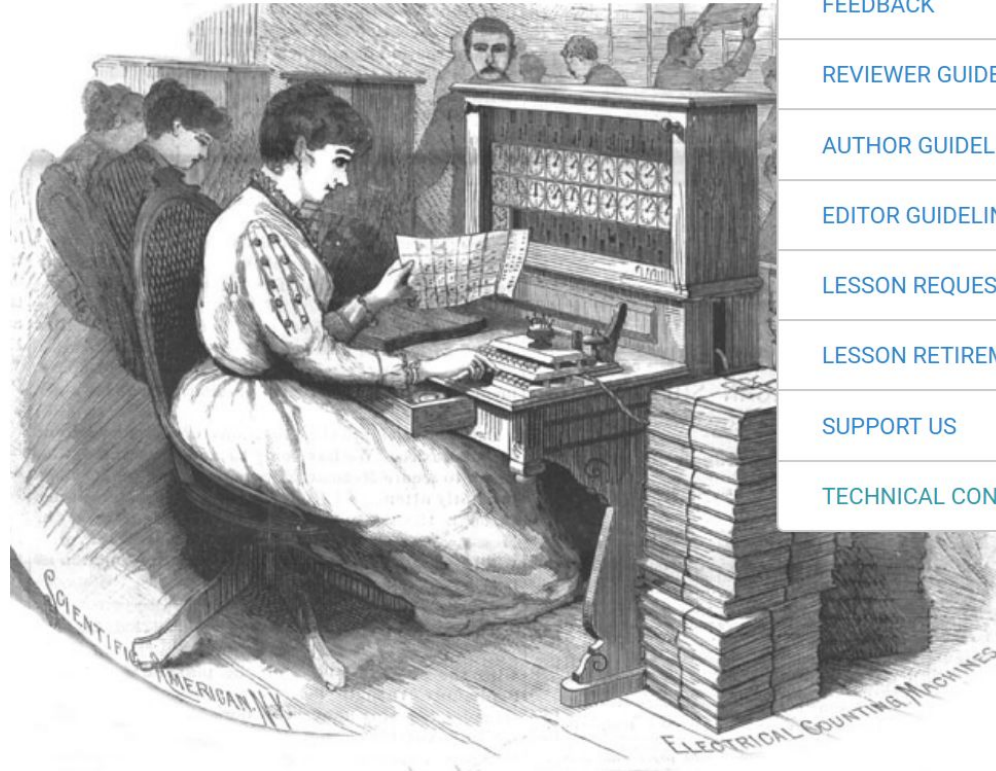
[PH]

80 lessons

ISSN: 2397-2068

We publish novice-friendly, peer-reviewed tutorials that help humanists learn a wide range of digital tools, techniques, and workflows to facilitate research and teaching.

[programminghistorian.org](http://programminghistorian.org) / [@ProgHist](https://twitter.com/ProgHist)

[OVERVIEW](#)[FEEDBACK](#)[REVIEWER GUIDELINES](#)[AUTHOR GUIDELINES](#)[EDITOR GUIDELINES](#)[LESSON REQUESTS](#)[LESSON RETIREMENT POLICY](#)[SUPPORT US](#)[TECHNICAL CONTRIBUTIONS](#)

# translation: text + **datasets**



2017 2018

```
> head(diamonds)
```

```
# A tibble: 6 x 10
```

	price	carat	cut	color	clarity	depth	table	x	y	z
	<int>	<dbl>	<ord>	<ord>	<ord>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	326	0.23	Ideal	E	SI2	61.5	55	3.95	3.98	2.43
2	326	0.21	Premium	E	SI1	59.8	61	3.89	3.84	2.31
3	327	0.23	Good	E	VS1	56.9	65	4.05	4.07	2.31
4	334	0.29	Premium	I	VS2	62.4	58	4.2	4.23	2.63
5	335	0.31	Good	J	SI2	63.3	58	4.34	4.35	2.75
6	336	0.24	Very good	J	VVS2	62.8	57	3.94	3.96	2.48

```
> head(diamantes)
```

```
# A tibble: 6 x 10
```

	precio	quilate	corte	color	claridad	profundidad	tabla	x	y	z
	<int>	<dbl>	<ord>	<ord>	<ord>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	326	0.23	Ideal	E	SI2	61.5	55	3.95	3.98	2.43
2	326	0.21	Premium	E	SI1	59.8	61	3.89	3.84	2.31
3	327	0.23	Bueno	E	VS1	56.9	65	4.05	4.07	2.31
4	334	0.29	Premium	I	VS2	62.4	58	4.2	4.23	2.63
5	335	0.31	Bueno	J	SI2	63.3	58	4.34	4.35	2.75
6	336	0.24	Muy bueno	J	VVS2	62.8	57	3.94	3.96	2.48

# one possible approach

```
> diamantes <- rename(diamonds, quilate = carat, corte = cut, ...
```



# one possible approach

```
> diamantes <- rename(diamonds, quilate = carat, corte = cut, ... %>%
```

# one possible approach

```
> diamantes <- rename(diamonds, quilate = carat, corte = cut, ... %>%  
  mutate(corte = case_when(cut == "Good" ~ "Bueno" ...
```

# another approach



**{datalang}**

[github.com/edgararuiz/datalang](https://github.com/edgararuiz/datalang)

```
> library(datalang)
> my_spec <- system.file("specs/diamonds.yml", package = "datos")
> translate_data(my_spec)
```

```
> library(datalang)
> my_spec <- system.file("specs/diamonds.yml", package = "datos")
> translate_data(my_spec)
```

	precio	quilate	corte	color	claridad	profundidad	tabla	x	y	z
1	326	0.23	Ideal	E	SI2	61.5	55	3.95	3.98	2.43
2	326	0.21	Premium	E	SI1	59.8	61	3.89	3.84	2.31
3	327	0.23	Bueno	E	VS1	56.9	65	4.05	4.07	2.31
4	334	0.29	Premium	I	VS2	62.4	58	4.2	4.23	2.63
5	335	0.31	Bueno	J	SI2	63.3	58	4.34	4.35	2.75
6	336	0.24	Muy bueno	J	VVS2	62.8	57	3.94	3.96	2.48

```
1- df:
2-   source: ggplot2::diamonds
3-   name: diamantes
4- variables:
5- price:
6-   trans: precio
7-   desc: Precio en dólares americanos (\$326-\$18,823)
8- carat:
9-   trans: quilate
10-  desc: Peso del diamante (0.2-5.01)
11- cut:
12-   trans: corte
13-   desc: Calidad del corte (Regular, Bueno, Muy bueno, Premium, Ideal)
14-   values:
15-     Good: Bueno
16-     Very Good: Muy bueno
17-     Fair: Regular
18- color:
19-   trans: color
20-   desc: Color del diamante, de J (peor) a D (mejor)
21- clarity:
22-   trans: claridad
23-   desc: Medida de que tan claro es el diamante (I1 (peor), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (mejor))
24- depth:
25-   trans: profundidad
26-   desc: Porcentaje de la profundidad total =  $z / \text{mean}(x, y) = 2 * z / (x + y)$  (43-79)
27- table:
28-   trans: tabla
29-   desc: Ancho de la parte superior del diamante con relación a su punto más ancho (43-95)
30- x:
31-   trans: x
32-   desc: Largo en milímetros
33- "y":
34-   trans: "y"
35-   desc: Ancho en milímetros
36- z:
37-   trans: z
38-   desc: Profundidad en milímetros
39- help:
40-   name: diamantes
41-   alias: diamantes
42-   title: Precio de 50.000 diamantes
43-   description: Un conjunto de datos que contiene los precios de casi 54.000 diamantes.
44-   usage: diamantes
45-   format: Un data.frame con 53.940 filas y 10 columnas|
46
```

```

1- df:
2-   source: ggplot2::diamonds
3-   name: diamantes
4- variables:
5-   price:
6-     trans: precio
7-     desc: Precio en dólares americanos (\$326-\$18,823)
8-   carat:
9-     trans: quilate
10-    desc: Peso del diamante (0.2-5.01)
11-  cut:
12-    trans: corte
13-    desc: Calidad del corte (Regular, Bueno, Muy bueno, Premium, Ideal)
14-    values:
15-      Good: Bueno
16-      Very Good: Muy bueno
17-      Fair: Regular
18-  color:
19-    trans: color
20-    desc: Color del diamante, de J (peor) a D (mejor)
21-  clarity:
22-    trans: claridad
23-    desc: Medida de que tan claro es el diamante (I1 (peor), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (mejor))
24-  depth:
25-    trans: profundidad
26-    desc: Porcentaje de la profundidad total = z / mean(x, y) = 2 * z / (x + y) (43-79)
27-  table:
28-    trans: tabla
29-    desc: Ancho de la parte superior del diamante con relación a su punto más ancho (43-95)
30-  x:
31-    trans: x
32-    desc: Largo en milímetros
33-  "y":
34-    trans: "y"
35-    desc: Ancho en milímetros
36-  z:
37-    trans: z
38-    desc: Profundidad en milímetros
39- help:
40-   name: diamantes
41-   alias: diamantes
42-   title: Precio de 50.000 diamantes
43-   description: Un conjunto de datos que contiene los precios de casi 54.000 diamantes.
44-   usage: diamantes
45-   format: Un data.frame con 53.940 filas y 10 columnas|
46-

```

df:

source: ggplot2::diamonds

name: diamantes

```

1- df:
2-   source: ggplot2::diamonds
3-   name: diamantes
4- variables:
5-   price:
6-     trans: precio
7-     desc: Precio en dólares americanos (\$326-\$18,823)
8-   carat:
9-     trans: quilate
10-    desc: Peso del diamante (0.2-5.01)
11-  cut:
12-    trans: corte
13-    desc: Calidad del corte (Regular, Bueno, Muy bueno, Premium, Ideal)
14-    values:
15-      Good: Bueno
16-      Very Good: Muy bueno
17-      Fair: Regular
18-  color:
19-    trans: color
20-    desc: Color del diamante, de J (peor) a D (mejor)
21-  clarity:
22-    trans: claridad
23-    desc: Medida de que tan claro es el diamante (I1 (peor), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (mejor))
24-  depth:
25-    trans: profundidad
26-    desc: Porcentaje de la profundidad total =  $z / \text{mean}(x, y) = 2 * z / (x + y)$  (43-79)
27-  table:
28-    trans: tabla
29-    desc: Ancho de la parte superior del diamante con relación a su punto más ancho (43-95)
30-  x:
31-    trans: x
32-    desc: Largo en milímetros
33-  "y":
34-    trans: "y"
35-    desc: Ancho en milímetros
36-  z:
37-    trans: z
38-    desc: Profundidad en milímetros
39- help:
40-   name: diamantes
41-   alias: diamantes
42-   title: Precio de 50.000 diamantes
43-   description: Un conjunto de datos que contiene los precios de casi 54.000 diamantes.
44-   usage: diamantes
45-   format: Un data.frame con 53.940 filas y 10 columnas|
46-

```

variables:

price:

trans: precio

desc: Precio en dólares americanos



```
1- df:
2-   source: ggplot2::diamonds
3-   name: diamantes
4- variables:
5- price:
6-   trans: precio
7-   desc: Precio en dólares americanos (\$326-\$18,823)
8- carat:
9-   trans: quilate
10-  desc: Peso del diamante (0.2-5.01)
11- cut:
12-   trans: corte
13-   desc: Calidad del corte (Regular, Bueno, Muy bueno, Premium, Ideal)
14-   values:
15-     Good: Bueno
16-     Very Good: Muy bueno
17-     Fair: Regular
18- color:
19-   trans: color
20-   desc: Color del diamante, de J (peor) a D (mejor)
21- clarity:
22-   trans: claridad
23-   desc: Medida de que tan claro es el diamante (I1 (peor), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (mejor))
24- depth:
25-   trans: profundidad
26-   desc: Porcentaje de la profundidad total = z / mean(x, y) = 2 * z / (x + y) (43-79)
27- table:
28-   trans: tabla
29-   desc: Ancho de la parte superior del diamante con relación a su punto más ancho (43-95)
30- x:
31-   trans: x
32-   desc: Largo en milímetros
33- "y":
34-   trans: "y"
35-   desc: Ancho en milímetros
36- z:
37-   trans: z
38-   desc: Profundidad en milímetros
39- help:
40-   name: diamantes
41-   alias: diamantes
42-   title: Precio de 50.000 diamantes
43-   description: Un conjunto de datos que contiene los precios de casi 54.000 diamantes.
44-   usage: diamantes
45-   format: Un data.frame con 53.940 filas y 10 columnas|
46
```

**values:**  
**Good: Bueno**  
**Very Good: Muy bueno**  
**Fair: Regular**



```
1- df:
2-   source: ggplot2::diamonds
3-   name: diamantes
4- variables:
5- price:
6-   trans: precio
7-   desc: Precio en dólares americanos (\$326-\$18,823)
8- carat:
9-   trans: quilate
10-  desc: Peso del diamante (0.2-5.01)
11- cut:
12-   trans: corte
13-   desc: Calidad del corte (Regular, Bueno, Muy bueno, Premium, Ideal)
14-   values:
15-     Good: Bueno
16-     Very Good: Muy bueno
17-     Fair: Regular
18- color:
19-   trans: color
20-   desc: Color del diamante, de J (peor) a D (mejor)
21- clarity:
22-   trans: claridad
23-   desc: Medida de que tan claro es el diamante (I1 (peor), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (mejor))
24- depth:
25-   trans: profundidad
26-   desc: Porcentaje de la profundidad total =  $z / \text{mean}(x, y) = 2 * z / (x + y)$  (43-79)
27- table:
28-   trans: tabla
29-   desc: Ancho de la parte superior del diamante con relación a su punto más ancho (43-95)
30- x:
31-   trans: x
32-   desc: Largo en milímetros
33- "y":
34-   trans: "y"
35-   desc: Ancho en milímetros
36- z:
37-   trans: z
38-   desc: Profundidad en milímetros
39- help:
40-   name: diamantes
41-   alias: diamantes
42-   title: Precio de 50.000 diamantes
43-   description: Un conjunto de datos que contiene los precios de casi 54.000 diamantes.
44-   usage: diamantes
45-   format: Un data.frame con 53.940 filas y 10 columnas|
46
```



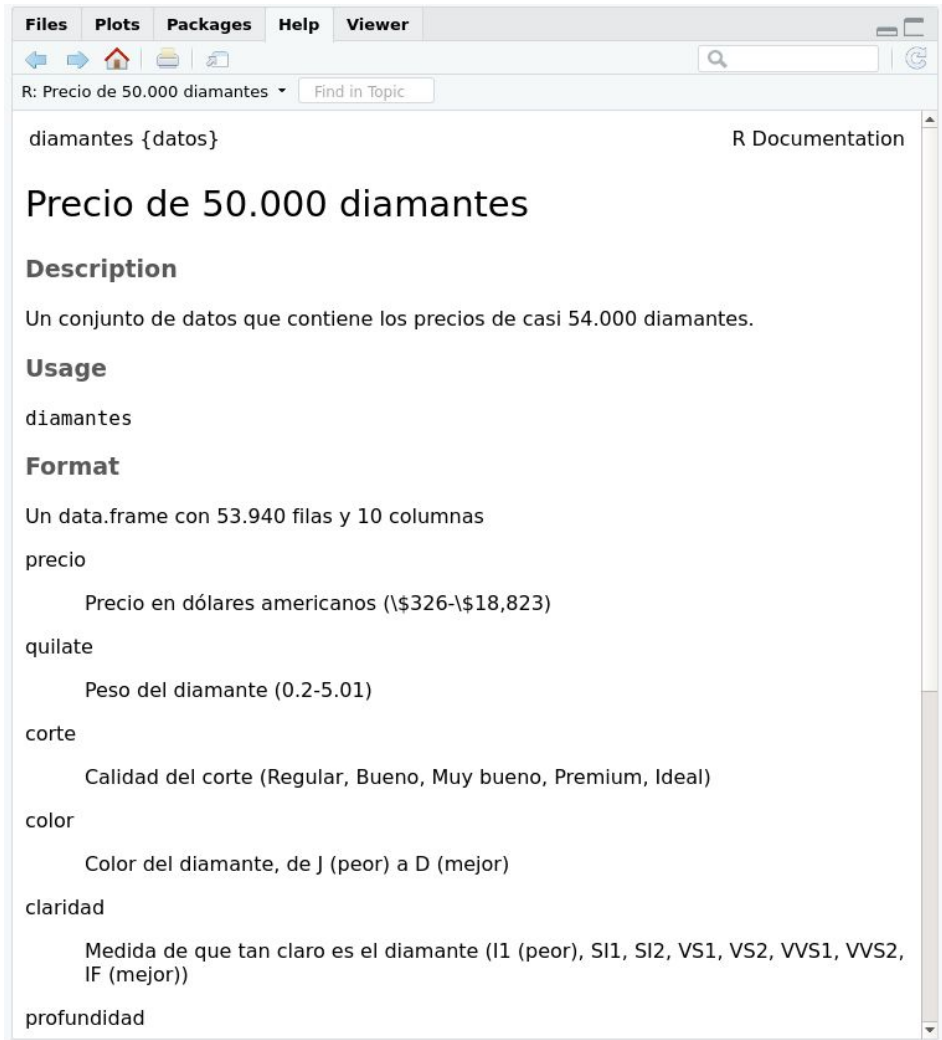
help:

```
> head(diamantes)
```

```
# A tibble: 6 x 10
```

	precio	quilate	corte	color	claridad	profundidad	tabla	x	y	z
	<int>	<dbl>	<ord>	<ord>	<ord>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	326	0.23	Ideal	E	SI2	61.5	55	3.95	3.98	2.43
2	326	0.21	Premium	E	SI1	59.8	61	3.89	3.84	2.31
3	327	0.23	Bueno	E	VS1	56.9	65	4.05	4.07	2.31
4	334	0.29	Premium	I	VS2	62.4	58	4.2	4.23	2.63
5	335	0.31	Bueno	J	SI2	63.3	58	4.34	4.35	2.75
6	336	0.24	Muy bueno	J	VVS2	62.8	57	3.94	3.96	2.48

## > ?diamantes



The image shows a screenshot of the R Documentation viewer for the 'diamantes' data set. The window title is 'R: Precio de 50.000 diamantes'. The content includes a description, usage instructions, and a list of variables with their units and ranges.

diamantes {datos} R Documentation

### Precio de 50.000 diamantes

#### Description

Un conjunto de datos que contiene los precios de casi 54.000 diamantes.

#### Usage

diamantes

#### Format

Un data.frame con 53.940 filas y 10 columnas

precio  
Precio en dólares americanos (\\$326-\\$18,823)

quilate  
Peso del diamante (0.2-5.01)

corte  
Calidad del corte (Regular, Bueno, Muy bueno, Premium, Ideal)

color  
Color del diamante, de J (peor) a D (mejor)

claridad  
Medida de que tan claro es el diamante (I1 (peor), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (mejor))

profundidad



{datos}



**translating book +  
datasets**



{datos}



**it's easy to contribute!**



2017

2018

2019



**conference-driven  
package**

2017

2018

2019





{datos}



**too bulky!**



# {datos}



```
create_promise <- function(path, package = "datos") {  
  spec_loc <- system.file(path, package = package)  
  spec <- read_yaml(spec_loc)  
  new_name <- spec$df$name  
  origin <- spec$df$source  
  
  envir <- as.environment(paste0("package:", package))  
  
  rlang::env_bind_lazy(  
    envir,  
    !! new_name := translate_data(spec_loc, eval(parse_expr(origin))  
  ))  
}
```



{datos}



```
> datos::diamantes
```

```
Error: diamantes is not an exported object  
from 'namespace:datos'
```



# {datos}



```
delayedAssign('diamantes',  
eval(parse(file.path(system.file('scripts','diamantes.txt', package = 'datos')))))
```



## datos

Este paquete provee la traducción al español de conjuntos de datos en inglés originalmente disponibles en otros paquetes de R. Los datos traducidos son los que se utilizan en los ejemplos del libro [R para Ciencia de Datos](#), la versión en español de [R for Data Science](#) de Hadley Wickham & Garrett Golemund. El paquete `datos` puede utilizarse junto con el libro o de manera independiente como fuente de datos de práctica en español.



## Instalación

La versión en CRAN está disponible por medio de:

```
# install.packages("datos")
```

La versión en GitHub puede ser instalada utilizando `remotes` :

```
# install.packages("remotes")
# remotes::install_github("cienciadedatos/datos")
```

## Requisitos

Este paquete traduce los datos **en el momento**. Esto implica que la versión en español de los datos no está contenida como un objeto dentro del paquete, sino que se genera al momento de utilizarlos. Por lo tanto, para poder usar `datos`, **el paquete que contiene los datos originales en inglés debe estar previamente instalado**.

Los paquetes necesarios son:

### Paquetes

nycflights13

## Links

Download from CRAN at  
<https://cloud.r-project.org/package=datos>

Browse source code at  
<https://github.com/cienciadedatos/datos>

Report a bug at  
<https://github.com/cienciadedatos/datos/issues>

## License

[Full license](#)

cc0

## Developers

Edgar Ruiz  
Author, maintainer

Riva Quiroga  
Author

Mauricio Vargas  
Author

Mauro Lepore  
Author

[All authors...](#)

## Dev status

CRAN **0.1.0**

build **passing**

codecov **94%**



**Edgar Ruiz**



## Links

Download from CRAN at  
<https://cloud.r-project.org/package=datos>

Browse source code at  
<https://github.com/cienciadedatos/datos>

Report a bug at  
<https://github.com/cienciadedatos/datos/issues>

## License

[Full license](#)

cc0

## Developers

Edgar Ruiz  
Author, maintainer

Riva Quiroga  
Author

Mauricio Vargas  
Author

Mauro Lepore  
Author

[All authors...](#)

## Dev status

CRAN 0.1.0

build passing

codecov 94%



**Edgar Ruiz**



**Riva Quiroga**



**Mauricio Vargas**



**Mauro Lepore**



## Links

Download from CRAN at  
<https://cloud.r-project.org/package=datos>

Browse source code at  
<https://github.com/cienciadedatos/datos>

Report a bug at  
<https://github.com/cienciadedatos/datos/issues>

## License

[Full license](#)

cco

## Developers

Edgar Ruiz  
Author, maintainer

Riva Quiroga  
Author

Mauricio Vargas  
Author

Mauro Lepore  
Author

[All authors...](#)

## Dev status

CRAN 0.1.0

build passing

codecov 94%



**Edgar Ruiz**



**Riva Quiroga**



**Mauricio Vargas**



**Mauro Lepore**



**Rayna Harris**



**Daniela Vázquez**



## Links

Download from CRAN at  
<https://cloud.r-project.org/package=datos>

Browse source code at  
<https://github.com/cienciadedatos/datos>

Report a bug at  
<https://github.com/cienciadedatos/datos/issues>

## License

[Full license](#)

cc0

## Developers

Edgar Ruiz  
Author, maintainer

Riva Quiroga  
Author

Mauricio Vargas  
Author

Mauro Lepore  
Author

[All authors...](#)

## Dev status

CRAN **0.1.0**

build **passing**

codecov **94%**





**datasets**



**code**



**translations**



**reviews**





**datasets**



**code**



**translations**



**reviews**





**datasets**



**code**



**translations**



**reviews**





**datasets**



**code**



**translations**



**reviews**





**datasets**



**code**



**translations**



**reviews**

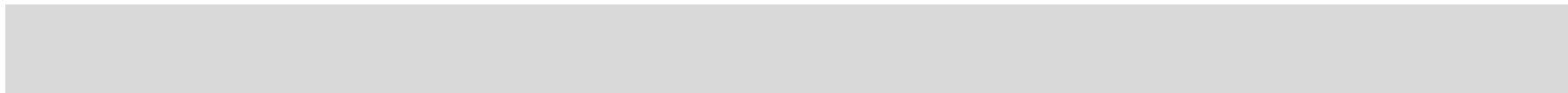


**but are we near the end of  
this project?**

**2018**



2018



∞



**what we have created is  
not just a book translation  
and a package**

**human and technical  
infrastructure**

**our biggest contribution  
developing {datos}**

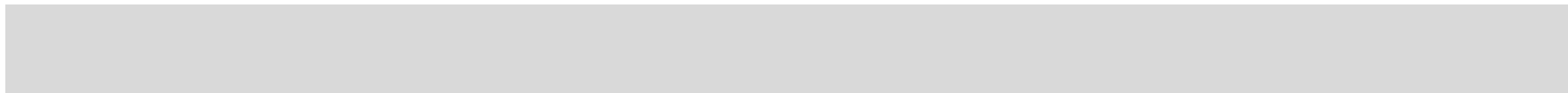
**> brand new errors**



 **Studio Community**



2018



∞

**MIND THE GAP**

**Riva Quiroga / @rivaquiroga**