

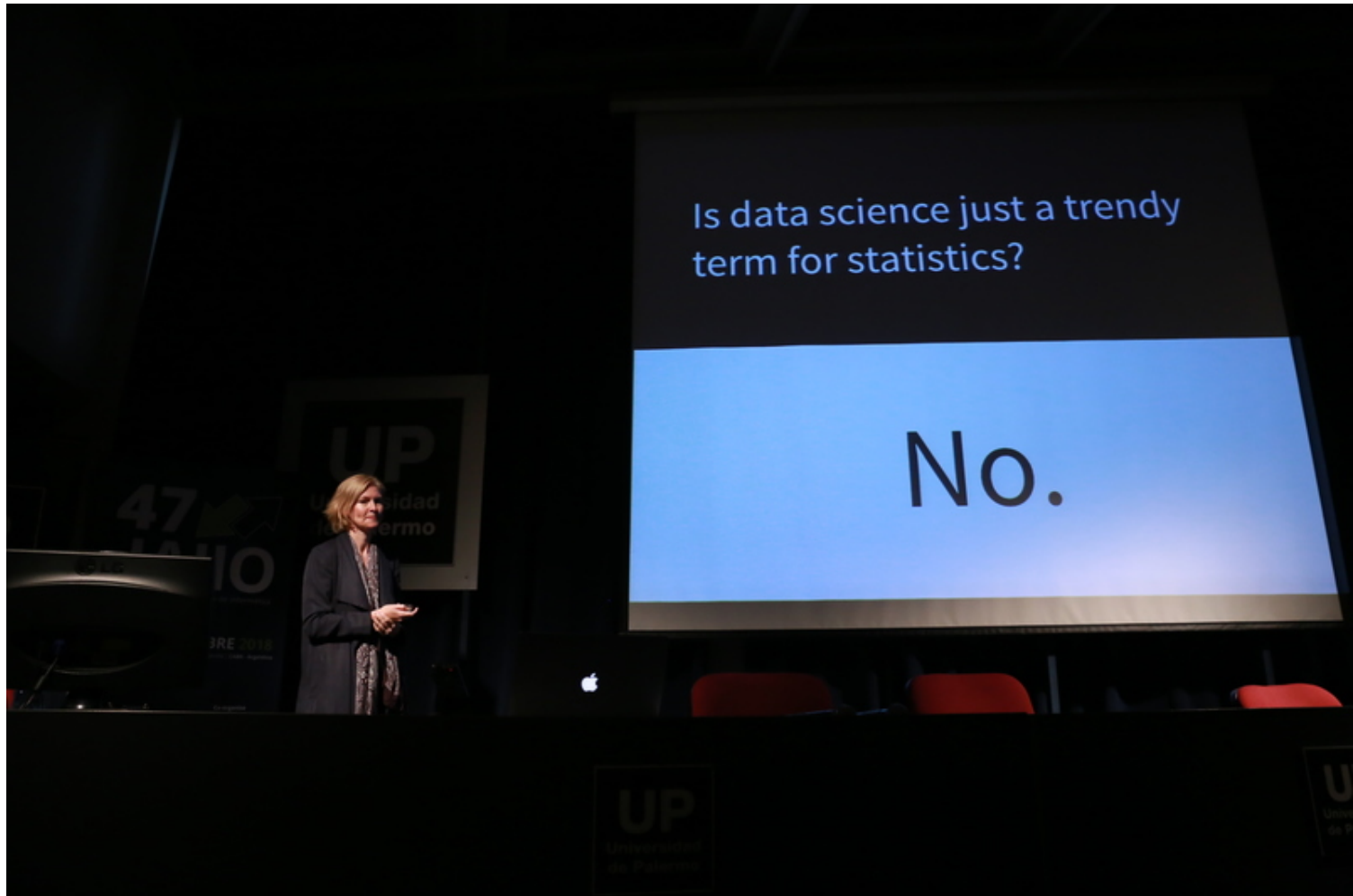
# Grupo de Usuarios de R en Uruguay

GURU::mvd

Primer Encuentro - Noviembre 2018

Muchas gracias por participar!!

# La ciencia de los datos



# La ciencia de los datos

- Es difícil pensar en algún área de la ciencia, industria o gobierno donde no se ha utilizado el término: *Data Science*
- The American Statistical Association (ASA, 2015), identifica tres áreas fundamentales:
  - Manejo de bases de datos: transformación, organización y agrupamiento de datos
  - Estadística: transformar los datos en conocimiento
  - Sistemas: infraestructura computacional para el análisis de datos
- ¿Cómo interactúan estadística, data science, análisis de datos y R ?

# 50 Years of Data Science

Trabajo en JCGS (2017) de David Donoho, donde revisa 4 trabajos clave:

1. John Tukey: “The Future of Data Analysis” (1962)
2. John Chambers: “Greater or Lesser Statistics: A Choice for Future Research” (1993)
3. Leo Breiman: “Statistical Modeling: The Two Cultures” (2001)
4. William S. Cleveland: “Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics” (2001)

# 50 Years of Data Science

- *... mi interés central es en el análisis de datos ..* (Tukey, 1962)
- *... el análisis de datos es una disciplina compleja, debe adaptarse a lo que la gente puede y necesita hacer con datos.* (Tukey, 1962)
- *...estadística enfrenta una elección entre la visión de estadística matemática y una visión más amplia de aprender de los datos..* (Chambers, 1993)
- *esfuerzo: ... modelos y métodos para datos (20%), cómputo con datos (15%), teoría (20%), ...* (Cleveland, 2001)
- *... modelado predictivo puede ser usado para analizar conjuntos de datos grandes y complejos ...* (Brieman, 2001)

# De todas partes ...

Ciencia de datos:

1. Toda la estadística
2. Otras ciencias (todas?) donde el manejo de datos es clave

# Lenguaje R



# 25 años ... un montonaso!

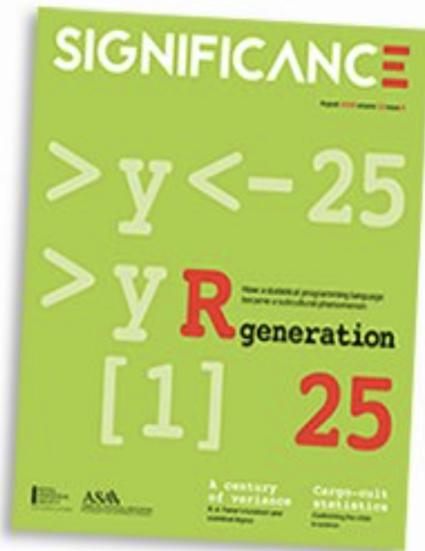
📌 Tweet fijado



**Royal Statistical Society** @RoyalStatSoc · 16 ago.

Our @signmagazine piece on the history of #R is our most shared article! Read it now while it's still #openaccess [rss.onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2018.01169.x](https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2018.01169.x)

🌐 Traducir Tweet



<https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2018.01169.x>

# Lenguaje R

Es uno de los lenguajes más comunmente utilizado para ciencia de datos. Es el estandar de la comunidad estadística.

- **Extensiones:**

- C++ , python
- R notebooks, Shiny apps
- Reportes dinámicos y reproducibles

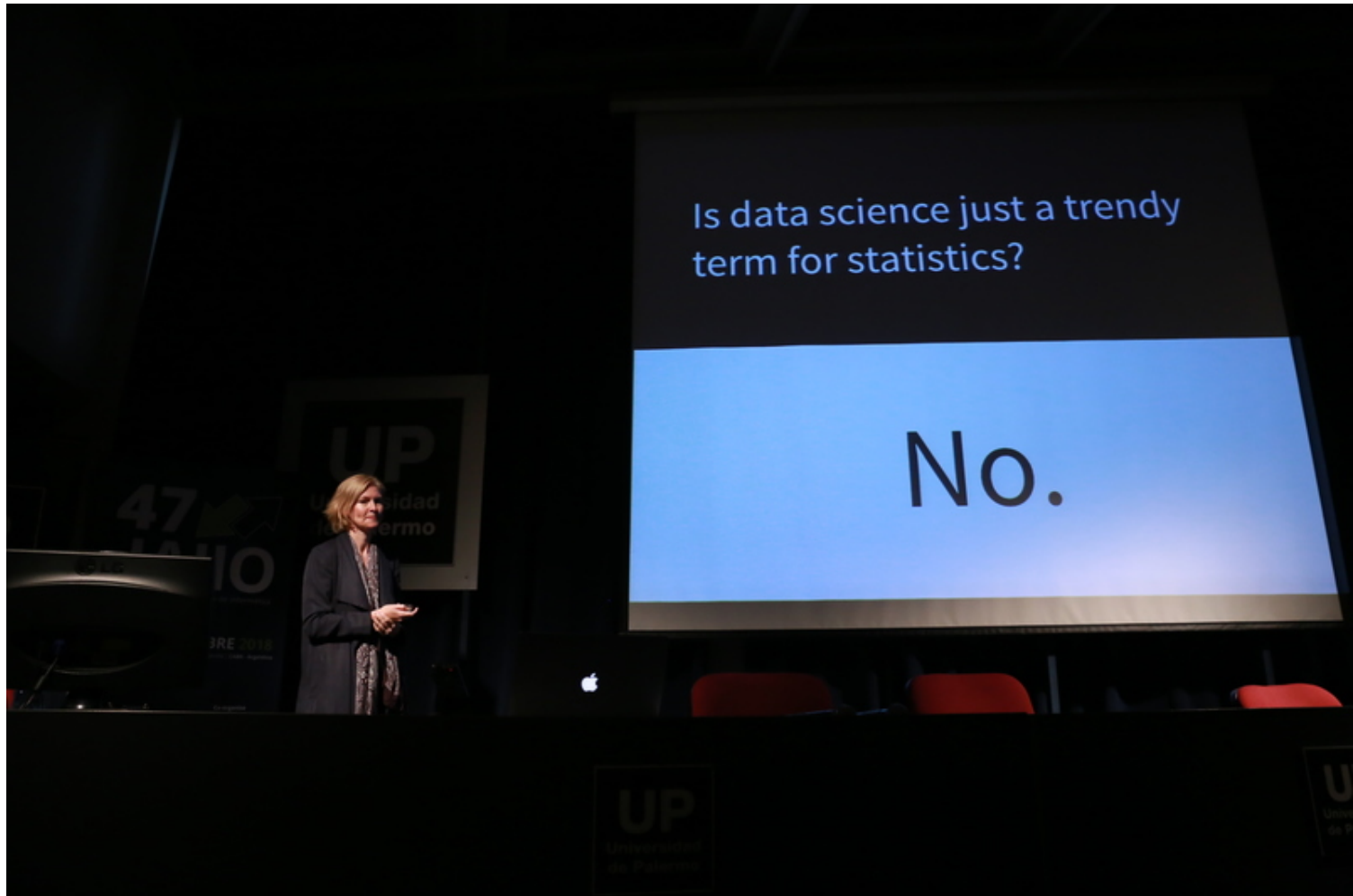
- **Ecosistema de bibliotecas:**

- CRAN (15000 bibliotecas)
- Bioconductor (2000 bibliotecas)
- Github (versiones de desarrollo)
- Ropensci

- **Comunidades:**

- Grupos de Usuarios
- Rladies
- Rforwards

¿Te acordás de esta foto ?



# LatinR 2018

Primer conferencia de uso y desarrollo de R en América Latina.



# ¿Porqué un Grupo de Usuarios ?

- Mucha gente interesada
  - Cursos de R en FCS, FCIEN, FCEA, CURE
  - Lic. Estadística de FCEA, FCIEN y FING
  - <https://espejito.fder.edu.uy/>
- Aportar a la comunidad R en la región

# R-Ladies Montevideo

- Primer iniciativa de comunidad R en Montevideo

R-Ladies: organización internacional que promueve la diversidad de género en la comunidad de R en un espacio amigable y seguro.

- Primer encuentro Agosto 2017
- 169 miembros

Contacto:

meetup [rladies.uy](https://www.meetup.com/rladies-uy/)

twitter: [@RLadiesMVD](https://twitter.com/RLadiesMVD)

Facebook: [RLadiesMVD](https://www.facebook.com/RLadiesMVD)

email: [montevideo@rladies.org](mailto:montevideo@rladies.org)

# R-consortium

- Brinda soporte a GURU::mvd (y todos los grupos de usuarios a nivel global)
- Código de conducta: Una comunidad *abierta, considerada y respetuosa*  
[https://wiki.r-consortium.org/view/R\\_Consortium\\_and\\_the\\_R\\_Community\\_Code\\_of\\_Conduct](https://wiki.r-consortium.org/view/R_Consortium_and_the_R_Community_Code_of_Conduct)

# Tryolabs

 **Tryolabs**  @tryolabs · 1 nov.

Proud to be hosting the inauguration of the [@gurumvd](#), the first group of [#R](#) users in Uruguay, dedicated to sharing experience and strengthen the R community. 🙌 [@\\_R\\_Foundation](#)



**Grupo de Usuarios de R:Montevideo** [@gurumvd](#)

📌 ¡Reservá la fecha!

El 21 de noviembre a las 19 horas en [@tryolabs](#) inauguramos el Grupo de Usuarios de R de Montevideo. [meetup.com/GURU-mvd...](#)

 Traducir Tweet

  3  17 



Se viene el Popurri!!!

# Natalia da Silva, @pacocuak

- Profesora Adjunta en el Instituto de Estadística (IESTA-UDELAR)
- Lic. en Estadística y Lic. Economía: UDELAR
- Doctorado y Maestría en Estadística: Iowa State University, USA
- Investigación: aprendizaje automático, estadística computacional, visualización estadística y meta-análisis
- Primeros pasos en R: como estudiante de Lic. Estadística (IESTA-UdelaR) ... cerca de 100 años atrás
- R-Ladies Ames, R-Ladies Montevideo y ahora GURU::mvd

# Ejemplos de lo que hago en R!

- **Paquetes:** <https://cran.r-project.org/web/packages/PPforest/index.html>
- **Reportes interactivos:** <https://arxiv.org/pdf/1807.07207.pdf>
- **Shiny apps:** <https://ndasilva.shinyapps.io/shiny-pov/>
- **Página personal:** <http://natydasilva.com>
- **Slides:** [https://natydasilva.github.io/taller\\_LatinR/#1](https://natydasilva.github.io/taller_LatinR/#1)
- **Curso en Lic. Estadística:** "Nuevas tecnologías para el análisis estadístico de datos"

# Nicolás Schmidt, @nicosch14

- Docente e investigador en el Departamento de Ciencia Política - FCS
- Primeros pasos en R: como ayudante de investigacion en FCEA. Necesidad de automatizar tareas que me dejaran tiempo para estudiar...

# Nicolás Schmidt, @nicosch14

- Docente e investigador en el Departamento de Ciencia Política - FCS
- Primeros pasos en R: como ayudante de investigacion en FCEA. Necesidad de automatizar tareas que me dejaran tiempo para estudiar...

Mi primer función en R!:

```
funcion_diaria <- function(paquete = "base"){  
  ruta <- search()  
  l <- ls(ruta[which(ruta==paste0("package:", paquete))])  
  sample(l, 1)  
}  
funcion_diaria()
```

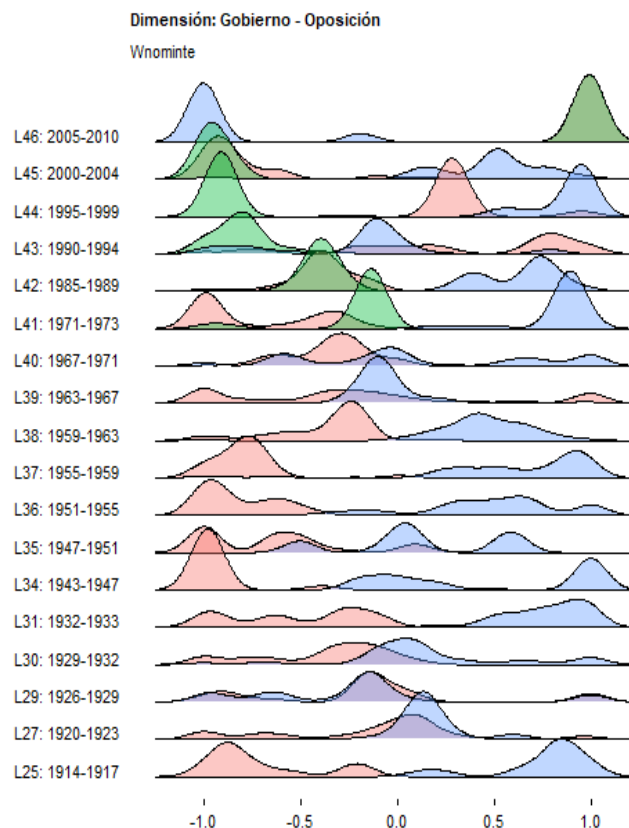
```
## [1] "as.data.frame.model.matrix"
```

# Qué hago en R...

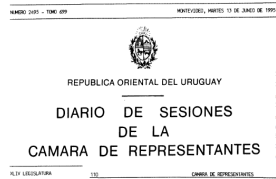
- Estudio sistemas electorales y sistemas de partidos y cómo éstos impactan en el regimen de gobierno.
- Producto de esto hice un paquete para simplificarme la vida:



- **esaps**: *Indicators of Electoral Systems and Party Systems*
- **Votaciones nominales en congresos**: Simulacion MCMC, multidimensional scaling...



# Actualmente... más de lo mismo pero con Text Analysis



onio Marchesano, Luis José Martínez, Miguel Mancía,  
Dr. León Morelli, Carlos E. Negro, Germán Oller, Juan  
Ramón Pereira Pabón, Juan Pintos Pereira, Carlos Pita  
Vivas, Baltasar Prieto, Alfonso Requena Vugt,  
Ricardo Rocha Inaz, Carlos Rodríguez Labruna,  
Jano, Carlos Rossi, Hebert Rossi Pasina, Walter  
Silveira Zavala, Carlos Norberto Soto, Guillermo  
sot Vallant, Tabaré Viera, Alfredo Zaffaroni Ortiz  
/ps

ont-size: 9pt">faltan:/span>/p>

ont-size: 9pt">Con licencia el señor  
sidentes <B>Ricardo Lombardo </B>by  
ntra qué se trata de la de dentro, porque  
para que la historia del mundo y de las cosas  
muy fácil, cuando un señor hablaba en la  
clase, la historia del mundo y de las cosas no  
especial, y meotro como una conversación  
parte de ella, pero después a medida que  
pero, que él, me agitas a para, con palabras  
y con palabras, y estar ya como lo de  
de dentro, incluso que voy a estar solo a como el  
entonces, después, entonces, pero, pero, pero,  
entonces, después, entonces, pero, pero, pero,  
sota a siete días que me que consecutivos a  
pero como

ont-size: 9pt">Sin aviso, los  
ber </B>by <B>Andrés Toriani.</span></p>

ont-size: 9pt">va name="numeral3">3</a>)</p>

ont-size: 9pt">SEÑOR PRESIDENTE (Dr.  
span>/p>

ont-size: 9pt">(Es la hora 19 y 04)</span></p>

ont-size: 9pt">Continúa la  
sunto que figura en primer lugar  
tad de miembros de la Suprema  
la República.</span></p>

ont-size: 9pt">Estaba en uso de la  
el señor legislador.</span></p>

ont-size: 9pt">SEÑOR TABOZO, - Me  
/p>



C:\Users\usuario\Google Drive\Text Analysis - RStudio									
File Edit Code View Plots Session Build Debug Profile Tools Help									
Go to file/function Addins									
editR® dat									
Filter									
* Intervencion	legislador	diario	FTV	preside	camara	camara2	legis	Nchar	talk
1 SEÑOR CARDOSO. - Señor Presidente de la Asamblea...	CARDOSO	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	12740	9
2 SEÑOR CIGLIUTI. - Pido la palabra. SEÑOR CIGLIUTI. D...	CIGLIUTI	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	2226	9
3 SEÑOR FERREIRA. - Pido la palabra. SEÑOR FERREIRA ...	FERREIRA	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	2388	9
4 SEÑOR PAZ AGUIRRE. - Pido la palabra. SEÑOR PAZ AG...	PAZ AGUIRRE	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	4016	9
5 SEÑOR PEREYRA. - Señor Presidente señores legisla...	PEREYRA	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	14906	9
6 SEÑOR RODRIGUEZ CAMUSO. - Pido la palabra. SEÑO...	RODRIGUEZ CAMUSO	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	5180	9
7 SEÑOR ROSSI PASINA. - Señor Presidente de la Asambl...	ROSSI PASINA	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	4060	9
8 SEÑOR TOURNE. - Pido la palabra. SEÑOR TOURNE. - S...	TOURNE	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	10411	9
9 SEÑOR VALLANT. - Pido la palabra. SEÑOR VALLANT. - S...	VALLANT	Diari...	NP 60 ...	PRESIDE EL SEÑOR SENA...	ASAMBLEA GENERAL	ASA...	42	1166	9
10 SEÑOR AGUIRRE. - Pido la palabra. SEÑOR AGUIRRE. - ...	AGUIRRE	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	6014	11
11 SEÑOR BATALLA. - Pido la palabra. SEÑOR BATALLA. - S...	BATALLA	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	2930	11
12 SEÑOR CIGLIUTI. - Pido la palabra. SEÑOR CIGLIUTI. - S...	CIGLIUTI	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	1341	11
13 SEÑOR FERREIRA. - Me permite una interrupción? SEÑ...	FERREIRA	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	1061	11
14 SEÑOR FORTEZA. - Pido la palabra. SEÑOR FORTEZA. - ...	FORTEZA	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	1736	11
15 SEÑOR LORENZO RIOS. - Pido la palabra. SEÑOR LORE...	LORENZO RIOS	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	6797	11
16 SEÑOR RITA. - Pido la palabra. SEÑOR RITA. - Señor Pre...	rita	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	523	11
17 SEÑOR RODRIGUEZ CAMUSO. - Pido la palabra. SEÑO...	RODRIGUEZ CAMUSO	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	703	11
18 SEÑOR SINGER. - Apoyado	SINGER	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	23	11
19 SEÑOR TOURNE. - Pido la palabra. SEÑOR TOURNE. - S...	TOURNE	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	2252	11
20 SEÑOR ZUMARAN. - Pido la palabra. SEÑOR ZUMARAN. ...	ZUMARAN	Diari...	NP 4-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	4078	11
21 SEÑOR CIGLIUTI. - Me permite una interrupción? SEÑO...	CIGLIUTI	Diari...	NP 5-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	16914	11
22 SEÑOR DAVEREDE. - Me permite una interrupción? SE...	DAVEREDE	Diari...	NP 5-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	1487	11
23 SEÑOR FAU. - Me permite una interrupción señor sena...	FAU	Diari...	NP 5-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	4003	11
24 SEÑOR FORTEZA. - Me permite una interrupción? SEÑ...	FORTEZA	Diari...	NP 5-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	14886	11
25 SEÑOR LESCANO. - Me permite una interrupción? SEÑO...	LESCANO	Diari...	NP 5-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	5925	11
26 SEÑOR MARTINEZ MORENO. - No sé si no cede un mil...	MARTINEZ MORENO	Diari...	NP 5-T...	PRESIDE EL DOCTOR EN...	ASAMBLEA GENERAL	ASA...	42	6623	11

Showing 1 to 27 of 174 entries

# Guzmán López, @guzilop

- Desarrollador en el Instituto Uruguayo de Meteorología (INUMET)
- Lic. en Ciencias Biológicas (FCien-UdelaR)
- Estudié programación en la ORT y más de "autodidacta"
- Primeros pasos en R: como estudiante de FCien en 2009



# Guzmán López, @guzilop

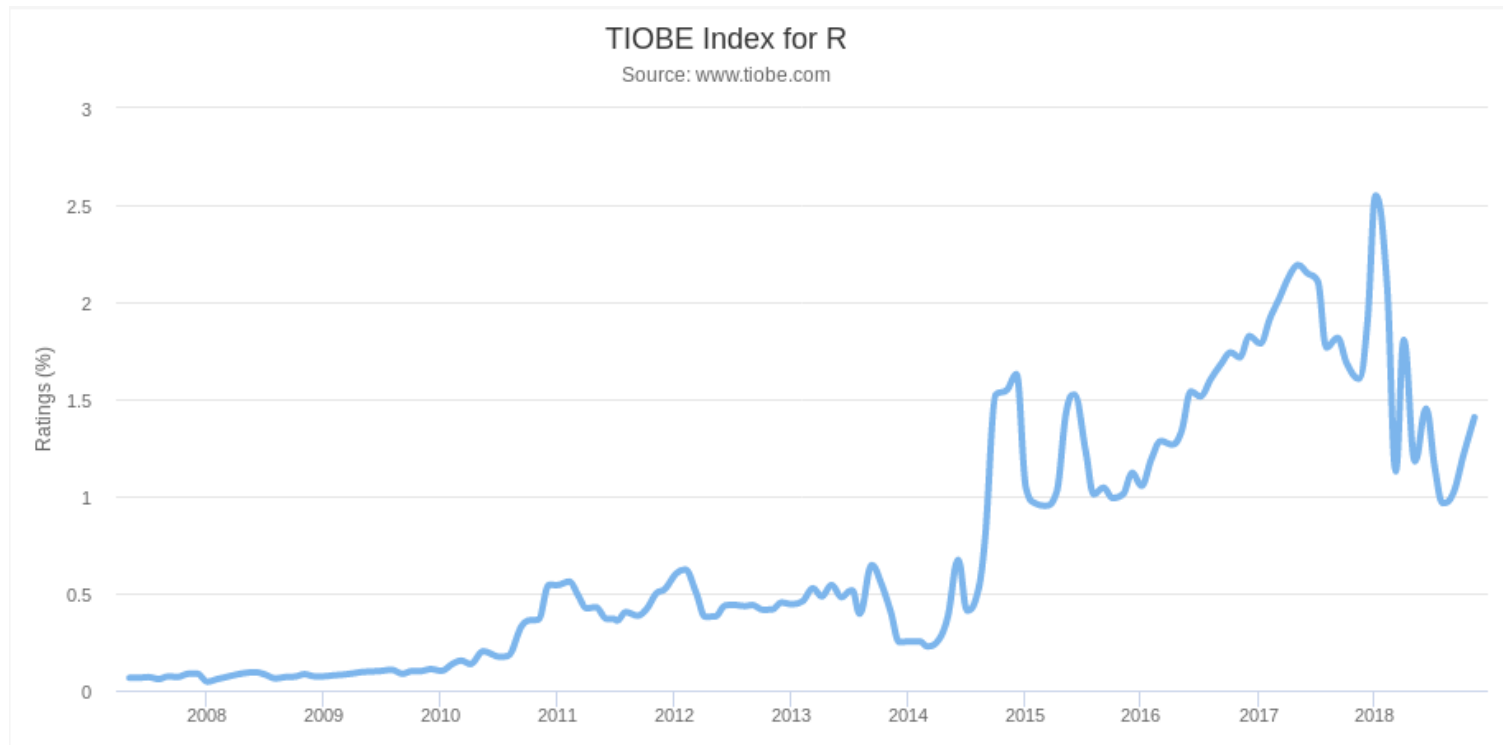
- Desarrollador en el Instituto Uruguayo de Meteorología (INUMET)
- Lic. en Ciencias Biológicas (FCien-UdelaR)
- Estudié programación en la ORT y más de "autodidacta"
- Primeros pasos en R: como estudiante de FCien en 2009

## Qué cosas hago/hice en R...

- Desarrollando herramientas de monitoreo y visualización para meteorología:
  - {Elastic Stack, Grafana, Docker} + R y Python
- Antes: acústica submarina, oceanografía y tráfico marítimo
  - R como SIG (sp, sf, leaflet, mapview, raster)
  - Geoestadística (gstat, geoR)
  - Aplicaciones web (shiny)

# Qué intereses tengo actualmente...

- Integración de R con otras tecnologías
- R en producción
- R + Python = <3
- Machine Learning para la maestría
- Que R siga creciendo...



# Elina Gómez, @elinagomezb

- Socióloga - UdelaR
- Maestría en Estudios Contemporáneos de América Latina (DS - FCS - UdelaR) (*tesis pendiente*)
- Asistente técnica Dirección Nacional de Evaluación y Monitoreo (MIDES) y DS - FCS
- Primeros pasos en R + Latex: compañerxs de trabajo me convencieron de la utilidad de R para el trabajo

## **Dictado:**

- Cursos internos de iniciación en R en DINEM - MIDES
- Curso *R: informes automatizados y análisis de datos en un único software* (Educación Permanente - FCS - UdelaR)

# Qué hago en R...

- Manipulación y procesamiento de bases de datos
- Automatización de informes sistemáticos con indicadores de monitoreo de Programas sociales (MIDES) y otros proyectos específicos
- **knitr**, **R Markdown**, **tidyverse** y otros...

Reporte de campo  
Encuesta Habitar Urbano Montevideo (EHUM)  
Facultad de Ciencias Sociales - Udelar  
2018-11-20

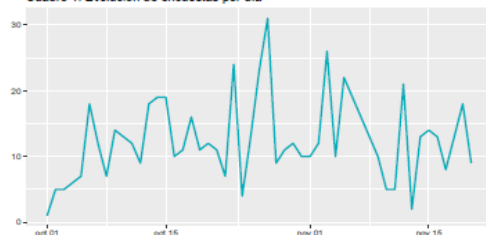
## I. Reporte de campo

A la fecha 2018-11-20 se registran 629 encuestas completas. En lo que va de trabajo de campo, los 70 encuestadores han realizado un promedio de 9.1 encuestas. La duración promedio de cada encuesta es de 56.2 minutos.

Table 1: Estadísticos descriptivos - Cantidad de encuestas

Min.	1
1st Qu.	1
Median	2
Mean	9
3rd Qu.	10
Max.	69

Cuadro 1. Evolución de encuestas por día



## Edad

En cuanto a la edad de los encuestados, se presentan estadísticos descriptivos y frecuencias según tramos de edad:

Table 3: Estadísticos descriptivos - Edad

Min.	6
1st Qu.	34
Median	52
Mean	50
3rd Qu.	64
Max.	90

Table 4: Tramos de edad

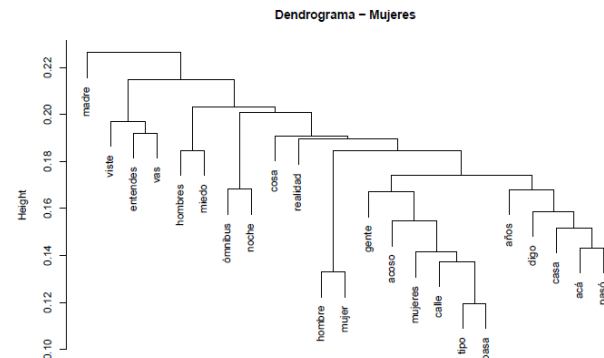
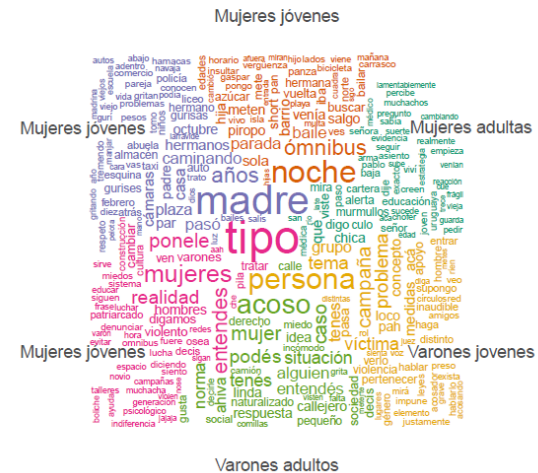
	N	Porcentaje
18 a 24	51	11.8
25 a 29	31	7.2
30 a 39	61	14.1
40 a 49	58	13.4
50 a 59	70	16.2
60 a 69	38	8.6
70 o más	69	15.3
Total	438	100.0

Table 5: Sexo según tramos de edad

	Mujeres	Porcentaje	Varones	Porcentaje
18 a 24	28	11.4	23	12.4
25 a 29	15	6.1	16	8.6
30 a 39	38	15.4	22	11.8
40 a 49	33	13.4	25	13.4
50 a 59	42	17.1	28	15.1
60 a 69	47	19.1	46	24.7
70 o más	43	17.5	26	14.0
Total	246	100.0	186	100.0

# Qué hago en R...

- Análisis cualitativo y textual
- *Codificación de textos, nubes de palabras, análisis de asociación, agrupación jerárquica, análisis de sentimiento*
- RQDA, tm, quanteda



# Florencia Díaz Viraqué, @fdiazviraque

Estudiante de doctorado, Investigador Asistente en Laboratorio de Interacción Hospedero-Patógeno, Institut Pasteur Montevideo

- Lic. en Bioquímica
- Maestría en Biología Celular y Molecular
- Facultad de Ciencias, UdelaR

# Qué hago en R...

Investigación: ensamblado de genoma, análisis de transcriptoma, análisis y visualización de secuencias biológicas

Paquetes con los que suelo trabajar:

- `library("seqinr")`
- `library("Rsubread")`
- `library("DESeq2")`
- `library("GenomicRanges")`
- `library("ggplot2")`
- `library("pheatmap")`
- `library("geneplotter")`
- `library("karyoploteR")`

Muchas gracias!

# Ignacio Alvarez-Castro, @nachalca

- Profesor Adjunto en el Instituto de Estadística (IESTA-UDELAR)
- Primeros pasos en R: como estudiante de Lic. Estadística (IESTA-UdelaR) ... cerca de 2003



# Ignacio Alvarez-Castro, @nachalca

- Profesor Adjunto en el Instituto de Estadística (IESTA-UDELAR)
- Primeros pasos en R: como estudiante de Lic. Estadística (IESTA-UdelaR) ... cerca de 2003

## Qué cosas hago en R...

- Estadística Bayesiana (**STAN**, JAGS, INLA, NIMBLE)

*"The systematic application of probabilistic reasoning to all inferential problems, including inferring parameters of statistical models, is sometimes called a Bayesian approach"*

**Machine Learning - A probabilistic perspective**, Kevin Murphy.

# Redescubrir funciones de R-base

```
library(tidyverse)
library(stringr)

"http://www.impo.com.uy/convenios/1261-011.pdf" %>%
  as_data_frame() %>%
  mutate(aux=str_extract_all(value, pattern = '[:digit:]+') ) %>%
  mutate(doc=map(aux, paste0, collapse="-")) %>%
  unnest(doc) %>%
  mutate( destino  =  paste0(doc, '.pdf', sep = '') ) %>%
  pull( destino )
```

```
## [1] "1261-011.pdf"
```

# Redescubrir funciones de R-base

```
library(tidyverse)
library(stringr)

"http://www.impo.com.uy/convenios/1261-011.pdf" %>%
  as_data_frame() %>%
  mutate(aux=str_extract_all(value, pattern = '[:digit:]+') ) %>%
  mutate(doc=map(aux, paste0, collapse="-")) %>%
  unnest(doc) %>%
  mutate( destino  =  paste0(doc, '.pdf', sep = '' ) %>%
  pull( destino )
```

```
## [1] "1261-011.pdf"
```

```
"http://www.impo.com.uy/convenios/1261-011.pdf" %>%
  basename()
```

```
## [1] "1261-011.pdf"
```

# Daniela Vázquez, @d4tagirl

- Economista.
- Data Scientist Freelance (actualmente trabajando para Pyxis).
- Blog: dv.uy.
- R-Ladies co-fundadora.
- NASA Datanaut.

## Pasado:

- Data Scientist en Idatha (Análisis de datos con foco en Redes Sociales).
- Analytics Consultant en Equifax.

# Primeros pasos en R:

2010 en Análisis Multivariado 1 de la Licenciatura en Estadística.

Después autodidacta en el trabajo y con MOOCs varios.

# Qué cosas hago/hice en R..

- Text Analysis
- Machine Learning
- Visualización de datos
- Scrapping
- Blogging