

Enseñanza basada en reproducibilidad

Juan C. Correa, Ph.D.

Fundación Universitaria Konrad Lorenz
juanc.correan@konradlorenz.edu.co

Curso en: **Tecnologías Reproducibles en la Enseñanza de la Metodología y la Estadística**

Objetivo del Curso

Comprender, a través de ejemplos concretos, los beneficios de usar repositorios de datos públicos para la enseñanza de contenidos metodológicos y estadísticos.

Agenda

- 1 El por qué de los repositorios de datos
- 2 Repositorios de datos
- 3 Algunos Ejemplos
 - Psicología de la corrupción y la riqueza
 - Consumismo Ideológico en Elecciones Colombianas
- 4 Un ejemplo guiado en Rstudio Cloud
- 5 Recomendaciones



Repositorios de datos ¿Por qué?

https://www.researchgate.net/publication/6763307_The_poor_availability_of_psychological_research_data_for_reanalysis

DOI: 10.1037/0003-066X.61.7.726

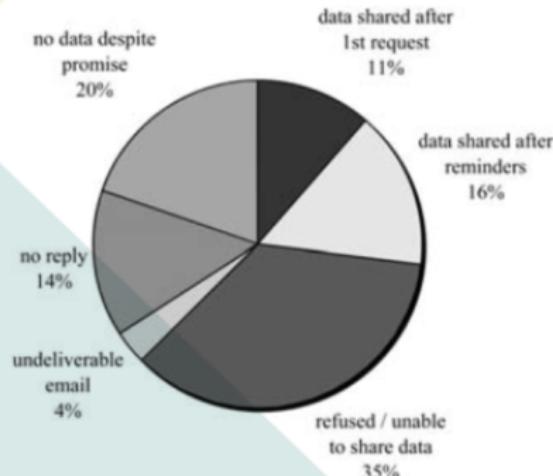
The Poor Availability of Psychological Research Data for Reanalysis

Jelte M. Wicherts, Denny Borsboom,
Judith Kats, and Dylan Molenaar
University of Amsterdam

The Problem

As the 141 articles included a total of 249 studies, we considered acquiring 90 to 100 data sets a realistic aim. We reasoned that adding a follow-up request after the original e-mail would take us a long way in that direction. Unfortunately, 6 months later, after writing more than 400 e-mails—and sending some corresponding authors detailed descriptions of our study aims, approvals of our ethical committee, signed assurances not to share data with others, and even our full resumes—we ended up with a meager 38 positive reactions and the actual data sets from 64 studies (25.7% of the total number of 249 data sets). This means that 73% of the authors did not share their data.

Figure 1.
Percentages of Empirical Articles' Corresponding Authors in Different Response Categories



Fuente original: Wicherts, M., Borsboom, D., Kats, J. & Molenaar, D. (2006). The poor availability Of Psychological research data for reanalysis. *The American Psychologist*, 61(7), 726 – 728.



Repositorios de datos ¿Por qué?



Iniciativas como GitHub o Stackoverflow (fundadas en 2008) visibilizaron los beneficios pedagógicos de compartir conocimientos aplicados para el desarrollo de software en ramas como la estadística, la genética, la manufactura, el control de procesos, etc.

Repositorios de datos ¿Por qué?

<https://www.redalyc.org/pdf/727/72711895025.pdf>

Psicothema 2009, Vol. 21, nº 4, pp. 652-655
www.psicothema.com

ISSN 0214 - 9915 CODEN PSOTEG
Copyright © 2009 Psicothema

¿Existe vida más allá del SPSS? Descubre R

Paula Elosua Oliden
Universidad del País Vasco

R es un entorno de programación, análisis estadístico y generación de gráficos distribuido bajo licencia GNU. Es un poderoso aliado para la investigación y una excepcional herramienta de trabajo para la docencia. Está constituido por más de 1.400 paquetes integrados con los que es posible ejecutar simples análisis descriptivos o aplicar los más complejos y novedosos modelos formales. Además, la incorporación a R de interfaces gráficas como Rcommander que crean entornos de trabajo amigables muy similares al entorno del SPSS permiten saltar la barrera de la accesibilidad, y utilizarlo sin ningún tipo de reparo en la docencia. ¿Existe algo mejor? Libre, gratuito, asequible, accesible y siempre a la vanguardia.

Iniciativas como GitHub o Stackoverflow facilitaron la popularización de software y datos abiertos.



Repositorios de datos

Un repositorio de datos abarca a un conjunto de recursos de hardware y/o software donde se almacenan conjuntos de datos estructurados (numéricos, geoespaciales, textuales, audiovisuales, o de cualquier otro formato) para su uso secundario con fines de investigación o docencia.

Repositorios de datos

Algunos repositorios de datos disponibles:

- En Ciencias Naturales

- <https://rda.ucar.edu/>
- <https://datadryad.org/stash>
- <http://archive.eso.org/cms.html>
- <https://www.dataone.org/>
- <https://figshare.com/>

- En Ciencias Sociales

- <https://www.icpsr.umich.edu/web/pages/>
- <https://www.re3data.org/>
- <https://ciser.cornell.edu/data/data-archive/>
- <https://dataverse.harvard.edu/>
- <https://www.europeansocialsurvey.org/>

Psicología de la corrupción y la riqueza



Cornell University
Library

arXiv.org > q-fin > arXiv:1604.00283

Quantitative Finance > General Finance

We gratefully acknowledge support from
the Simons Foundation
and member institutions

Search or Article-id

(Help | Advanced search)

All papers ▾

Go!

Corruption and Wealth: Unveiling a national prosperity syndrome in Europe

Juan C. Correa, Klaus Jaffe

(Submitted on 5 Nov 2015)

Data mining revealed a cluster of economic, psychological, social and cultural indicators that in combination predicted corruption and wealth of European nations. This prosperity syndrome of self-reliant citizens, efficient division of labor, a sophisticated scientific community, and respect for the law, was clearly distinct from that of poor countries that had a diffuse relationship between high corruption perception, low GDP/capita, high social inequality, low scientific development, reliance on family and friends, and languages with many words for guilt. This suggests that there are many ways for a nation to be poor, but few ones to become rich, supporting the existence of synergistic interactions between the components in the prosperity syndrome favoring economic growth. No single feature was responsible for national prosperity. Focusing on synergies rather than on single features should improve our understanding of the transition from poverty and corruption to prosperity in European nations and elsewhere.

Comments: 22 pages, 4 figures

Subjects: General Finance (q-fin.GN)

Cite as: arXiv:1604.00283 [q-fin.GN]

(or arXiv:1604.00283v1 [q-fin.GN] for this version)

Submission history

From: Juan C. Correa [view email]

[v1] Thu, 5 Nov 2015 10:58:44 GMT (973kb)

<https://arxiv.org/abs/1604.00283>



Psicología de la corrupción y la riqueza



THE WORLD BANK



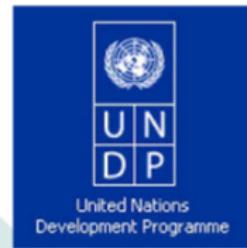
TRANSPARENCY
INTERNATIONAL



SCImago
Journal & Country
Rank

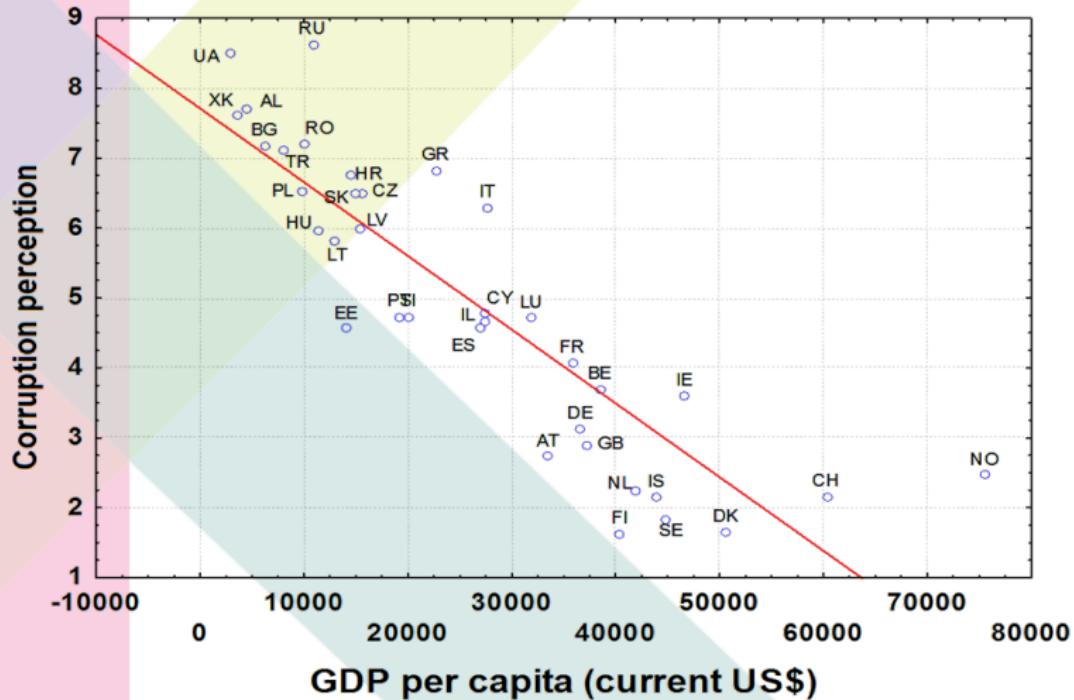


European
Social
Survey

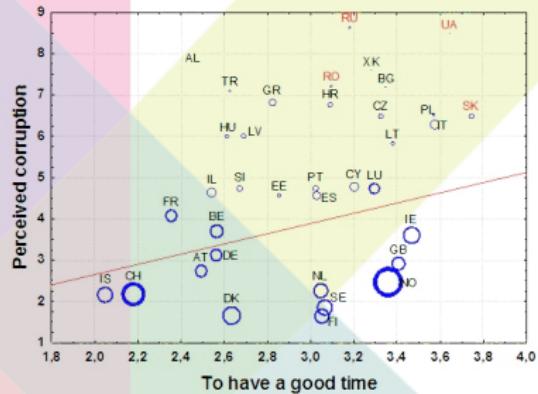


United Nations
Development Programme

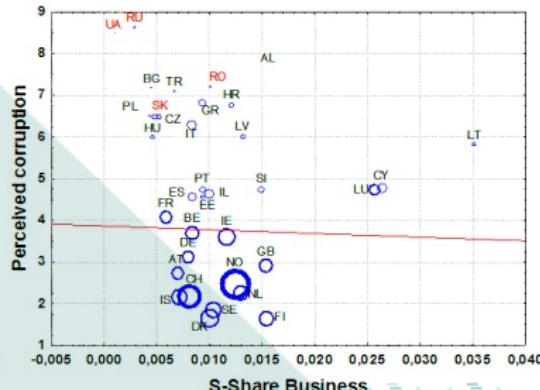
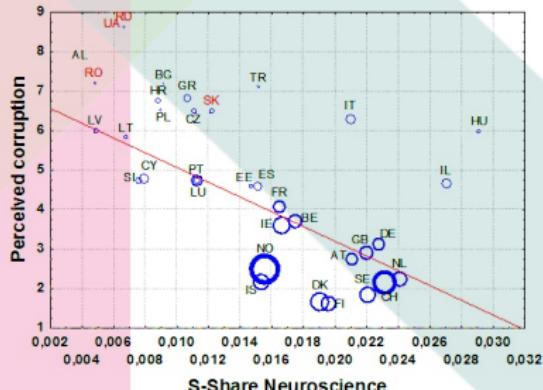
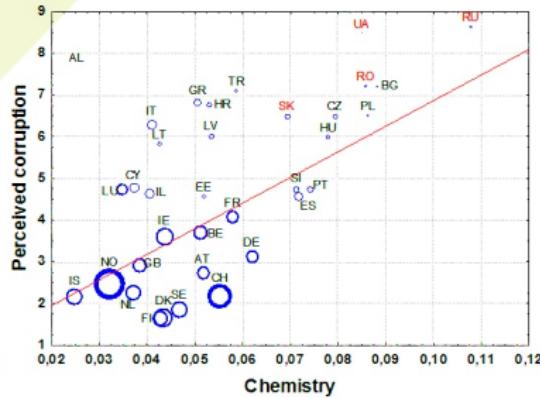
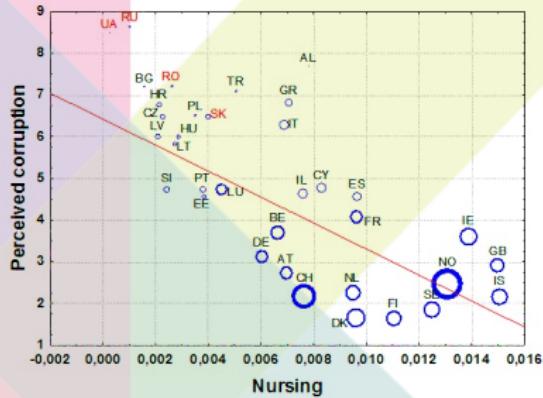
Psicología de la corrupción y la riqueza



Psicología de la corrupción y la riqueza



Psicología de la corrupción y la riqueza



Psicología de la corrupción y la riqueza



From: "Ladislav Kristoufek" <kristouf@utia.cas.cz>
Date: Apr 6, 2016 7:36 AM
Subject: Corruption paper
To: <juancorrea@usb.ve>
Cc:

Dear Prof. Correa,

I have come across your "Corruption and Wealth" on Econophysics Forum and I believe you might be interested in a related work of mine. Please find attached our paper published in Physica A.

With kind wishes,

Ladislav Kristoufek

====

Ladislav Kristoufek

Assistant Professor
Charles University in Prague, Faculty of Social Sciences
Institute of Economic Studies
Department of Finance and Capital Markets

Research Fellow

[Physica A 428 \(2015\) 351–358](#)



Worldwide clustering of the corruption perception

Michal Paulus, Ladislav Kristoufek*

Institute of Economic Studies, Faculty of Social Sciences, Charles University, Opletalova 26, 110 00, Prague 1, Czech Republic



HIGHLIGHTS

- Clusters of corruption perception are analyzed.
- Well-defined clusters are identified.
- Corruption is found to be related with economic development.



Consumismo Ideológico en Elecciones Colombianas

<https://www.liebertpub.com/doi/pdfplus/10.1089/cyber.2016.0402>

CYBERPSYCHOLOGY, BEHAVIOR, AND SOCIAL NETWORKING
Volume 20, Number 1, 2017
Mary Ann Liebert, Inc.
DOI: 10.1089/cyber.2016.0402

Ideological Consumerism in Colombian Elections, 2015: Links Between Political Ideology, Twitter Activity, and Electoral Results

Juan C. Correa, PhD¹ and Jorge E. Camargo, PhD²

Abstract

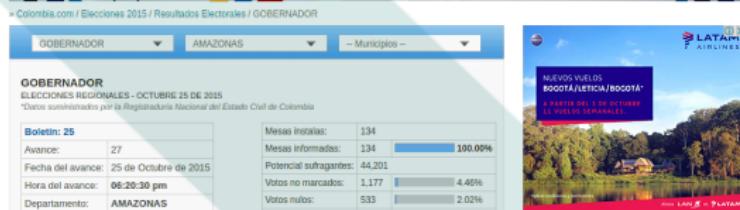
Propagation of political ideologies in social networks has shown a substantial impact on voting behavior. Both the contents of the messages (the ideology) and the politicians' influence on their online audiences (their followers) have been associated with such an impact. In this study we evaluate which of these factors exerted a major role in deciding electoral results of the 2015 Colombian regional elections by evaluating the linguistic similarity of political ideologies and their influence on the Twitter sphere. The electoral results proved to be strongly associated with tweets and retweets and not with the linguistic content of their ideologies or politicians' followers in Twitter. Finally, suggestions for new ways to analyze electoral processes are discussed.

Keywords: Twitter, Colombian elections, ideological consumerism

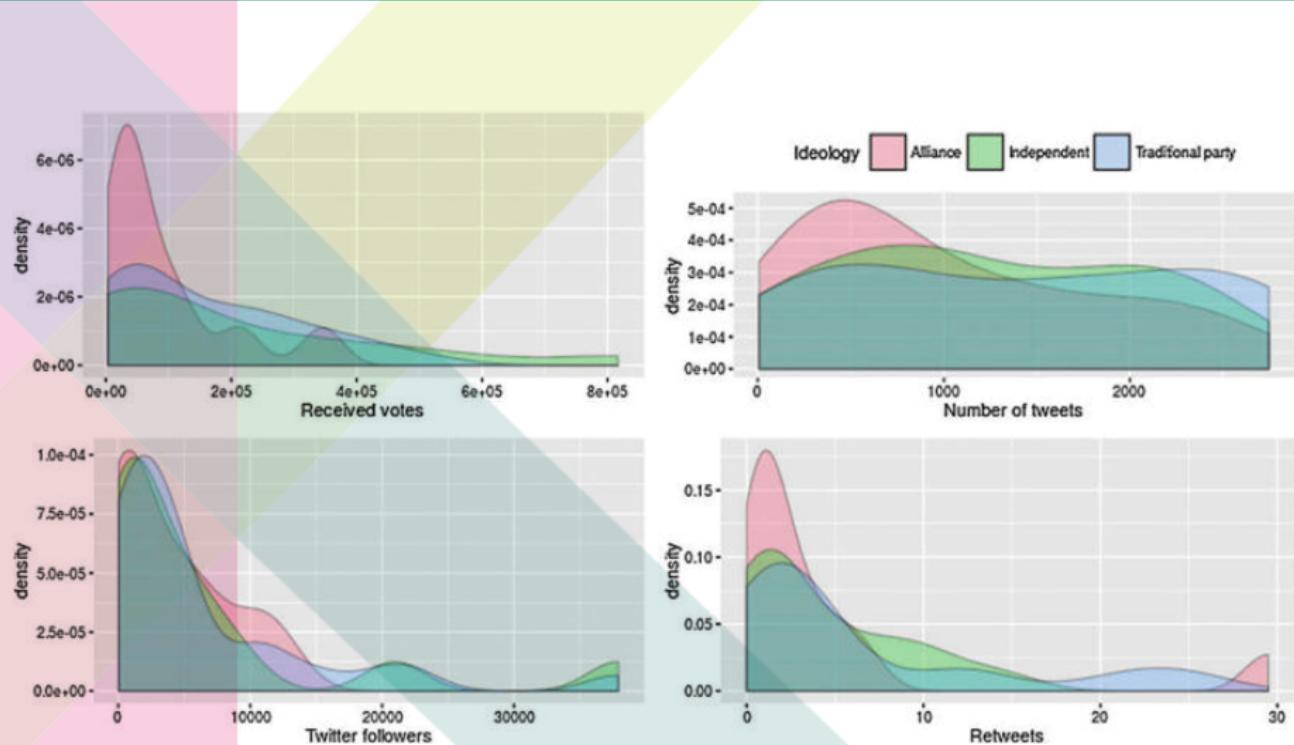


Consumismo Ideológico en Elecciones Colombianas

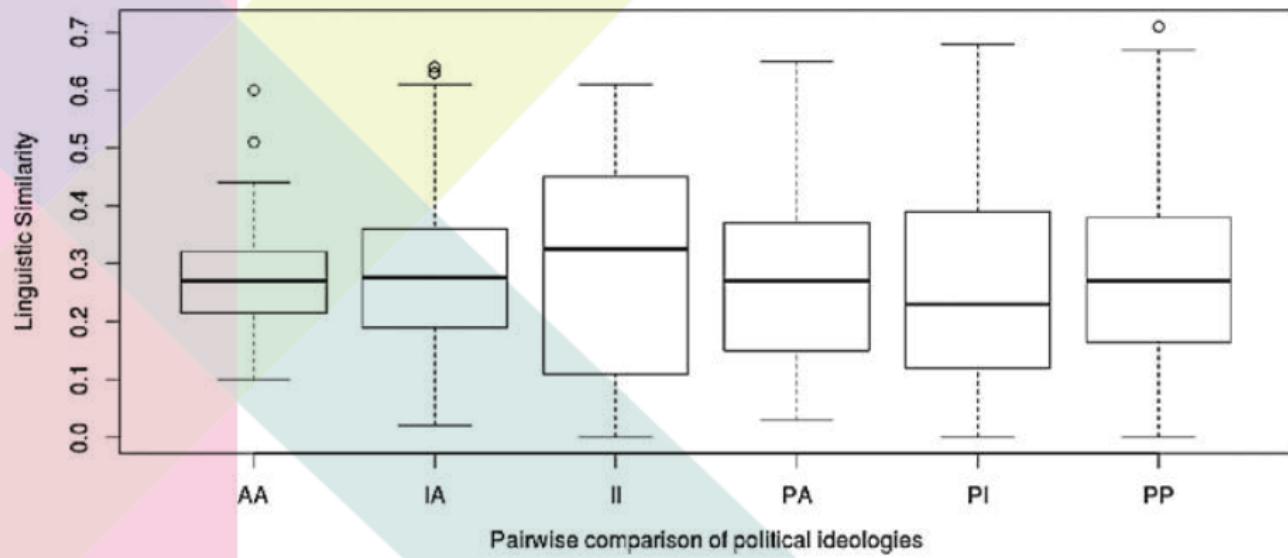
52 candidatos a gobernadores; 52 colecciones de tweets; 69202 tweets; número promedio de tweets por candidato; número promedio de retweets por tweet por candidato; ideología política; similaridad lingüística entre ideologías; resultados electorales (número de votos recibidos).



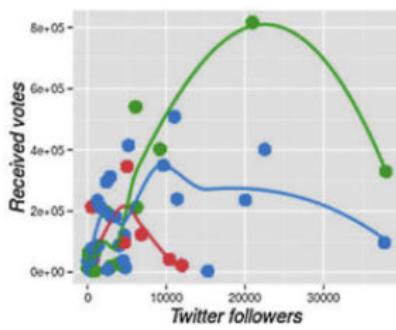
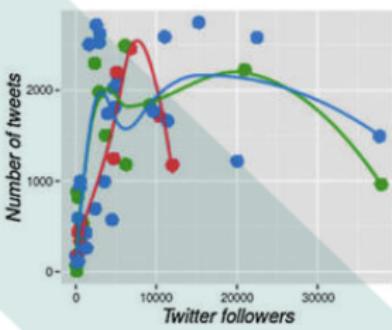
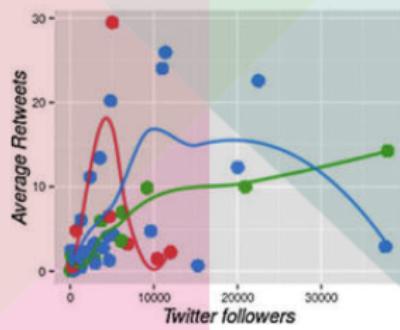
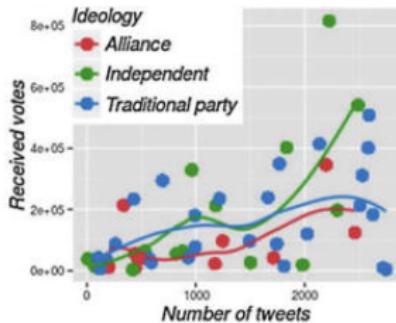
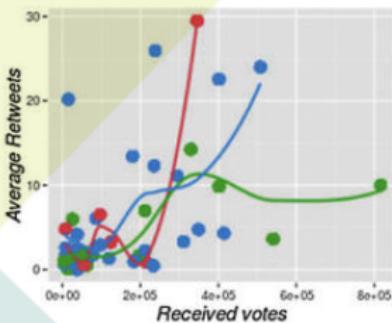
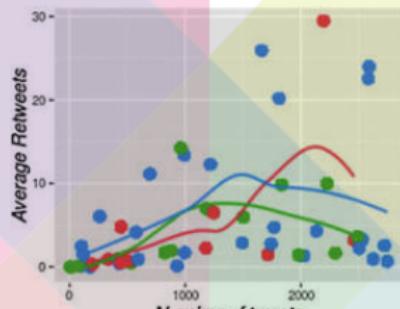
Consumismo Ideológico en Elecciones Colombianas



Consumismo Ideológico en Elecciones Colombianas



Consumismo Ideológico en Elecciones Colombianas



Un ejemplo guiado en Rstudio Cloud

Descargar los datos y el script para reproducir los gráficos incluidos en el artículo de consumo ideológico en elecciones colombianas.

jcurrean / Visualizaci-on-de-Datos-en-R

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 4 branches 0 tags

Go to file Add file Code

Aqui se muestran algunos de los usos que puede darse al paquete ggplot2 de R.

Readme MIT License

No releases published Create a new release

No packages published Publish your first package

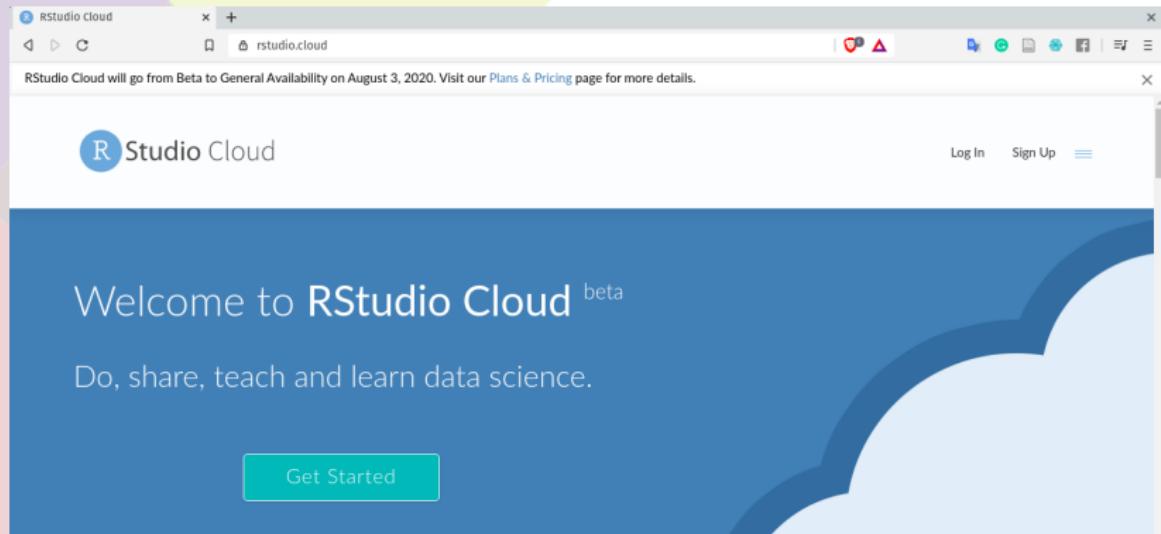
R 100.0%

<https://github.com/jcorrean/Visualizaci-on-de-Datos-en-R>



Un ejemplo guiado en Rstudio Cloud

Crear una cuenta en Rstudio Cloud



<https://rstudio.cloud/>



Un ejemplo guiado en Rstudio Cloud

Step 1: "Sign Up"

The screenshot shows the RStudio Cloud sign-up page. At the top, there are tabs for 'Log In' and 'Sign Up'. Below that, a section titled 'Available Plans' displays four options: 'Cloud Free', 'Cloud Premium', 'Cloud Instructor', and 'Cloud Organization'. The 'Cloud Free' plan is highlighted with a dark grey background. It includes a brief description: 'If you make limited, occasional use of RStudio Cloud, or have your usage covered by your school/organization or an instructor, our free plan is all you need.' To the right of this description is a list of five features with checkmarks: 'Up to 10 projects total', '1 shared space (5 inviters and 10 projects max)', '15 project hours per month', 'Up to 1 GB RAM per project', and 'Up to 1 CPU per project'. A large blue arrow points from the bottom left towards the 'Sign Up' button at the bottom right of the page.

Step 2: Fill the form

The screenshot shows the RStudio Cloud sign-up form. It contains several input fields: 'Email' (juan.correa@vse.cz), 'Password' (represented by a series of asterisks), 'First Name' (Juan C.), and 'Last Name' (Correa). Below these is a large blue 'Sign up' button. To the right of the button is the text '— or —'. Further down are two additional sign-up options: 'Sign up with Google' (with a Google logo) and 'Sign up with GitHub' (with a GitHub logo).



Un ejemplo guiado en Rstudio Cloud

Renombrar el workspace como “VisualizacióndeDatosenR”

The screenshot shows the RStudio Cloud interface. On the left, there's a sidebar with sections like 'Spaces', 'Learn', and 'Info'. The main area is titled 'Your Projects' and shows a single project named 'Untitled Project' created on Sep 30, 2019, at 6:04 PM. There are buttons for 'New Project', 'Delete', and a search bar. On the right, there are options for sorting projects ('By name' is selected) and information about capacity. A navigation bar at the bottom includes icons for back, forward, and search.

RStudio Cloud will go from Beta to General Availability on August 3, 2020. Visit our [Plans & Pricing](#) page for more details.

Your Workspace Projects Info Juan C. Correa

Spaces Your Projects Untitled Project Created Sep 30, 2019 6:04 PM Delete Search Projects

New Project Options

Sort Projects

By name

By date created

Capacity

This is your personal workspace.

Learn more about [Your Workspace](#) in the [Guide](#).

Learn Guide What's New Primers Cheat Sheets Feedback and Questions

Plans & Pricing Terms and Conditions System Status

Juan C. Correa (<https://correajc.com>) Enseñanza basada en reproducibilidad Bogotá, Junio-2021 23 / 29

Un ejemplo guiado en Rstudio Cloud

The screenshot shows the RStudio Cloud interface. On the left, there's a sidebar with navigation links: Spaces, Your Workspace (containing 'Prueba' and '+ New Space'), Learn (containing 'Guide', 'What's New', 'Primer', 'Cheat Sheets'), Help, Current System Status, and RStudio Community.

The main workspace title is 'Your Workspace / Visualización de Datos en R'. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. A status bar at the bottom indicates RAM usage (4.0.3) and a connection to Juan C. Correa.

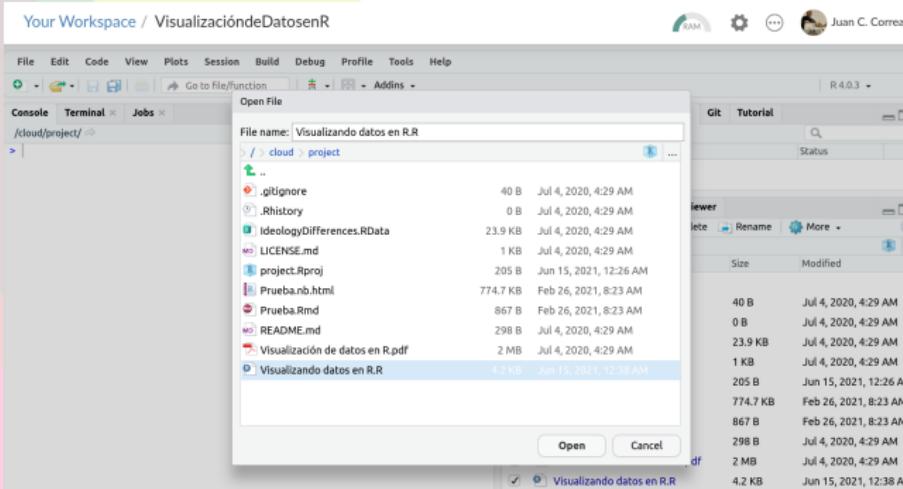
The workspace contains several tabs: Console, Terminal, and Jobs. The Console tab shows the path '/cloud/project/'. The Terminal tab has a command prompt (>). The Jobs tab is currently inactive.

The Environment tab in the top right shows a 'New Connection' dialog. The History tab shows a single entry: 'Connection'. The Connections tab lists 'Status' and 'Connection'. The Git tab shows a repository with files like '.gitignore', '.Rhistory', 'IdeologyDifferences.RData', 'LICENSE.md', 'project.Rproj', 'Prueba.nb.html', 'Prueba.Rmd', 'README.md', 'Visualización de datos en R.pdf', and 'Visualizando datos en R.R'. The Tutorial tab is also present.

At the bottom, there are standard RStudio navigation icons (back, forward, search, etc.). A large green 'K' logo is in the bottom right corner.

Un ejemplo guiado en Rstudio Cloud

Clic en el ícono de la carpeta amarilla (arriba a la izquierda, debajo del menú Edit).

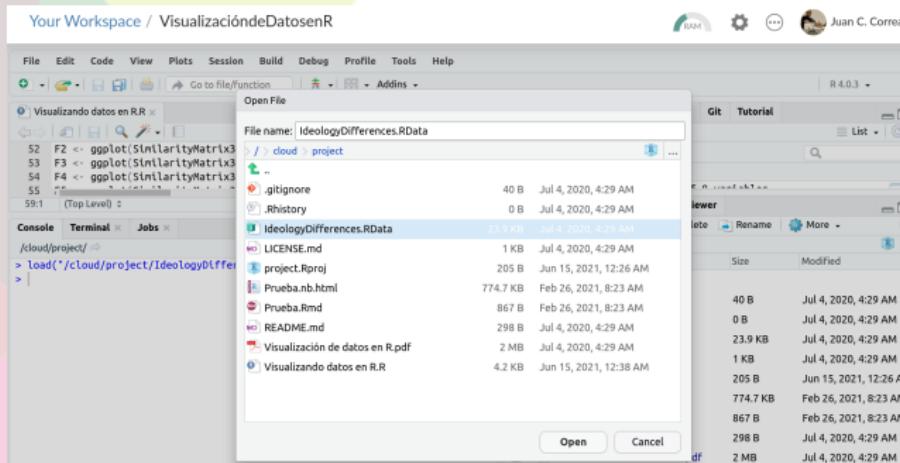


Clic en el archivo “Visualizando datos en R.R” y luego en el botón Open.



Un ejemplo guiado en Rstudio Cloud

Clic en el ícono de la carpeta amarilla (arriba a la izquierda, debajo del menú Edit).



Clic en el archivo “IdeologyDifferences.RData” y luego en el botón Open.



Un ejemplo guiado en Rstudio Cloud

Script Reproducible

The screenshot shows the RStudio Cloud interface. On the left, the 'R Script' pane contains a script with code for visualizing data. A red circle highlights the 'Run' button in the toolbar above the script. On the right, the 'Environment' pane displays various objects and their details. A blue arrow points from the 'Run' button to the 'Environment' pane, indicating that running the script will update these objects. Below the panes are the 'Console', 'Terminal', and 'Jobs' tabs.

R Script Content:

```
1 # El primer paso consiste en abrir los datos
2 # desde RStudio. Aquí doy por asumido que
3 # ya se sabe la manera de hacerlo.
4 library(ggplot2)
5 # Vamos a hacer nuestra primera Figura
6 ggplot(RegressionData, aes(x=RegressionData$`Received votes`, fill=Regr
7 # Ahora vamos a guardarla como un objeto
8 # que manipularemos con
9 # una función para hacer que
10 # aparezca tal como vemos en la Figura 1
11 # del paper de Correa & Camargo (2017)
12
```

Environment pane (Data section):

Object	Description
Cuadro1	544 obs. of 8 variables
Followers	List of 1
LinguisticSimil...	1326 obs. of 4 variables
ReceivedVotes	List of 1
RegressionData	45 obs. of 14 variables
Retweets	List of 1
SimilarityMatri...	52 obs. of 10 variables
Tweets	List of 1

Console pane:

```
/cloud/project/ >
> load("/cloud/project/IdeologyDifferences.RData")
>
```

Clic en el botón Run dentro del recuadro rojo para obtener los mismos gráficos.



Recomendaciones



.sav
.txt
.csv
.xlsx
.dat
.tab





PhD. Pablo Vallejo-Medina

R³

<https://youtu.be/Qdz7V94lhIO>