

Visualización de Datos en R

Una brevísima guía con ggplot2

Juan C. Correa, PhD

Fundación Universitaria Konrad Lorenz
Facultad de Psicología
Bogotá - Colombia

✉ juanc.correan@konradlorenz.edu.co



Agenda

- 1 Visualización de Datos: 101
 - 2 Motivación
 - 3 Ejemplos Concretos
 - 4 ...Un poquito más...
 - 5 Recomendaciones para otros gráficos
- References



Visualización de Datos: 101

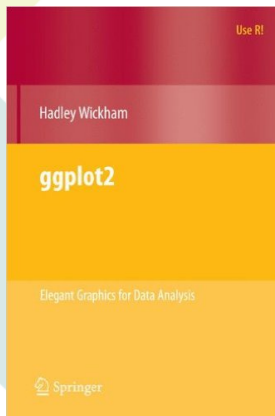
La visualización de datos se refiere a las técnicas utilizadas para comunicar datos o información codificándola como objetos visuales (e.g., puntos, líneas, barras, etc) contenidos en gráficos. El objetivo es comunicar la información clara y eficientemente a los usuarios. Es uno de los pasos en el análisis de datos o la ciencia de los datos.



WIKIPEDIA
The Free Encyclopedia

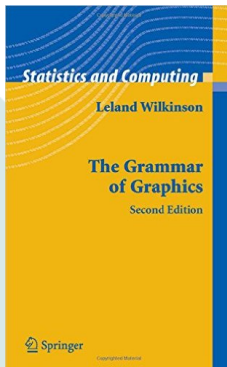
Visualización de Datos: 101

“**ggplot2** es un paquete de R para producir gráficos de datos estadísticos, pero es distinto a otros paquetes porque cuenta con una profunda gramática subyacente que los genera (Wickham, 2016, p. 1)”.



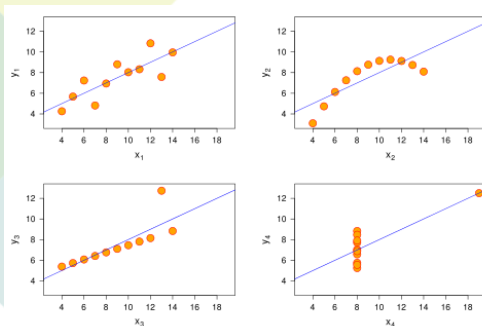
Visualización de Datos: 101

La gramática de **ggplot2** está basada en “La Gramática de los Gráficos” (Wilkinson, 2006) y le permite al usuario construir sus propios gráficos con base en un conjunto predefinido de instrucciones, que permite adaptar los gráficos a las necesidades del usuario.



Motivación: El Cuarteto de Anscombe

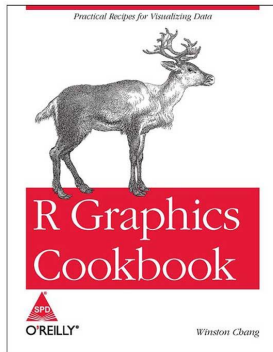
Cuatro conjuntos de datos con propiedades estadísticas iguales y que evidentemente son distintos cuando se examinan gráficamente.



https://es.wikipedia.org/wiki/Cuarteto_de_Anscombe

Motivación: Libro de Texto

Aunque hay un montón de libros, emplearemos este...



Disponible en Dropbox: “*Aprendiendo R/Libros de R*”

Ejemplo 1: Twitter y Elecciones

Reproduzcamos los gráficos de Correa and Camargo (2017)

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

<http://online.liebertpub.com/doi/pdfplus/10.1089/cyber.2016.0402>

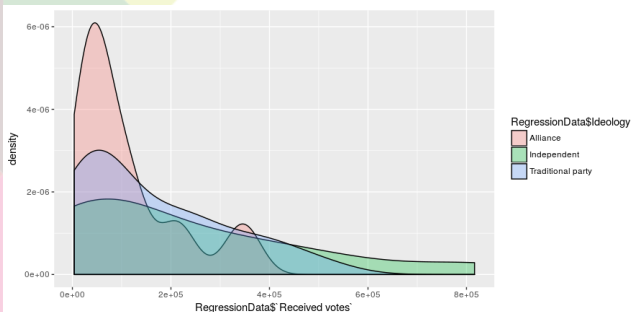


Ejemplo 1: Twitter y Elecciones

```
load('AprendiendoR/Datos usados en el grupo de estudio/IdeologyDifferences.RData')
```

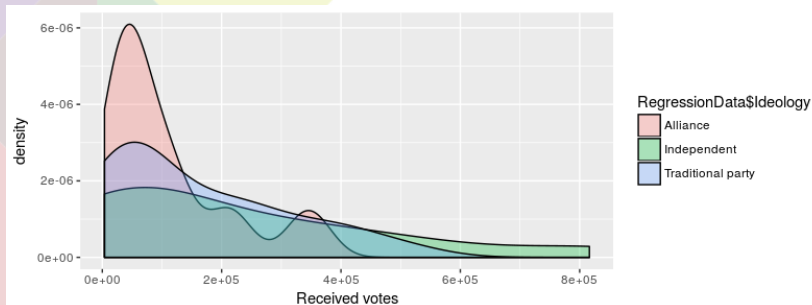
```
library(ggplot2)
```

```
ggplot(RegressionData, aes(x=RegressionData$`Received votes`, fill=RegressionData$`Ideology`)) +  
geom_density(alpha=0.3)
```



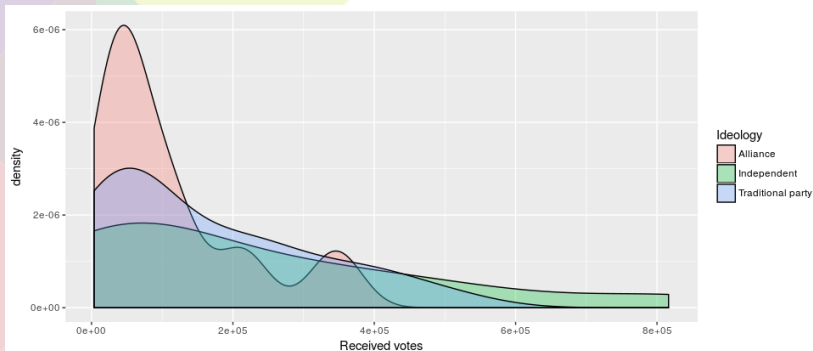
Ejemplo 1: Twitter y Elecciones

```
ggplot(RegressionData, aes(x=RegressionData$` Received votes` , fill=RegressionData$` Ideology` )) +  
geom_density(alpha=0.3) + xlab('Received votes')
```



Ejemplo 1: Twitter y Elecciones

```
ggplot(RegressionData, aes(x=RegressionData$` Received votes` , fill=RegressionData$` Ideology` )) +  
geom_density(alpha=0.3) + xlab('Received votes') + labs(fill='Ideology')
```



¿Cómo hacemos los otros gráficos de la Figura 1?



Ejemplo 1: Twitter y Elecciones

¿Cómo ponemos varios gráficos uno al lado del otro?

Una buena idea es ponerle un nombre a cada gráfico (digamos “f1” para la primera figura, “f2” para la segunda...) y entonces manipularlos como objetos dentro de una grilla...



Ejemplo 1: Twitter y Elecciones

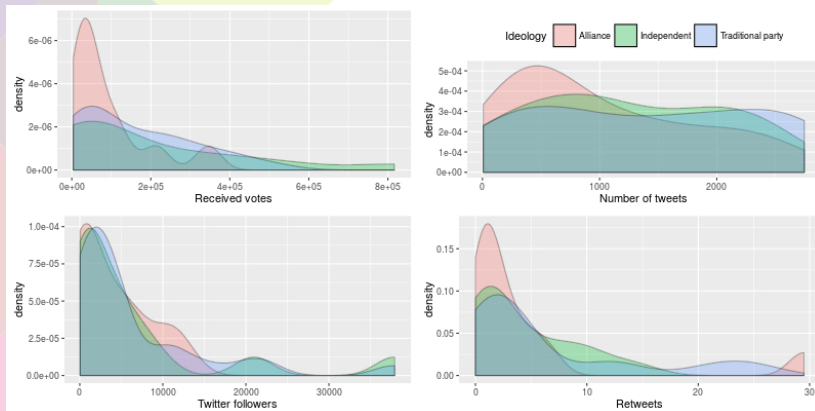
Vamos a crear una función llamada `multiplot` que nos permita colocar gráficos de `ggplot2` uno al lado del otro...

```
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL)
library(grid)
plots <- c(list(...), plotlist)
numPlots = length(plots)
if (is.null(layout)) {
  layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
    ncol = cols, nrow = ceiling(numPlots/cols))
}
if (numPlots==1) {
  print(plots[[1]])
} else {
  grid.newpage()
  pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))
  for (i in 1:numPlots) {
    matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))
    print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
      layout.pos.col = matchidx$col))
  }
}
multiplot(f1, f2, f3, f4, cols=2) #Pon el gráfico f1, al lado del f2, f3 y f4
```



Ejemplo 1: Twitter y Elecciones

El resultado luego de haberle puesto un nombre a cada gráfico y haberlos manipulado con la función “multiplot” debería ser semejante a esto...



Ejemplo 1: Twitter y Elecciones

Ok... No es exactamente igual... pero puede serlo si aprendemos a manipular algunos aspectos estéticos de ggplot2

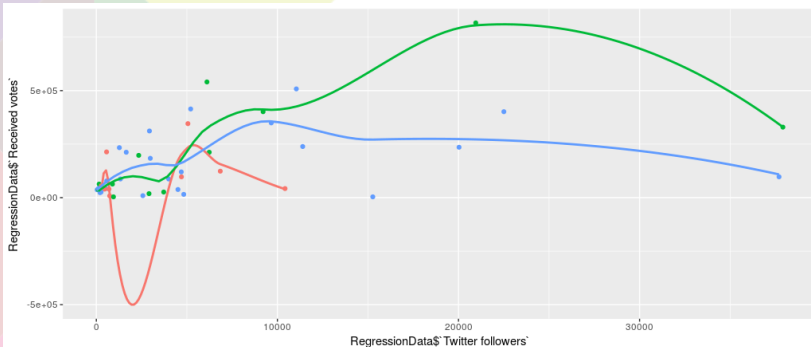
- 1 Quitar leyenda de un ggplot (Sección 10.1, página 225-226)
- 2 Cambiar la posición de la leyenda (Sección 10.2)
- ⋮
- 10 Usar etiquetas con múltiples líneas de texto (Sección 10.10)

La idea no es memorizar estos códigos, sino saber qué cosas se pueden hacer mientras se aprende ggplot2



Ejemplo 1: Twitter y Elecciones

```
ggplot(RegressionData, aes(x=RegressionData$`Twitter followers`, y=RegressionData$`Received votes`,  
  colour=RegressionData$Ideology)) + geom_point() + geom_smooth(method = loess, se=FALSE) +  
  theme(legend.position = 'none')
```



Ejemplo 1: Twitter y Elecciones

No siempre las funciones de ggplot2 son necesarias para visualizar datos. De hecho, la figura 2 fue obtenida con la siguiente función

```
boxplot(LinguisticSimilarity$LinguisticSimilarity ~  
LinguisticSimilarity$Comparison, xlab = "Pairwise  
Comparison")
```



Ejemplo 1: Elecciones y Twitter

¿Qué debemos hacer para “reproducir” la Figura 3?



Ejemplo 1: Elecciones y Twitter

¿Qué debemos hacer para “reproducir” la Figura 3?

- 1 Genere cada gráfico por separado.



Ejemplo 1: Elecciones y Twitter

¿Qué debemos hacer para “reproducir” la Figura 3?

- 1 Genere cada gráfico por separado.
- 2 Asígnele un nombre a cada gráfico, siguiendo la buena práctica de “F1” para la primera figura, “F2” para la segunda...



Ejemplo 1: Elecciones y Twitter

¿Qué debemos hacer para “reproducir” la Figura 3?

- 1 Genere cada gráfico por separado.
- 2 Asígnele un nombre a cada gráfico, siguiendo la buena práctica de “F1” para la primera figura, “F2” para la segunda...
- 3 Adapte la función “Multiplot” para 3 columnas



Ejemplo 1: Elecciones y Twitter

¿Qué debemos hacer para “reproducir” la Figura 3?

- 1 Genere cada gráfico por separado.
- 2 Asígnele un nombre a cada gráfico, siguiendo la buena práctica de “F1” para la primera figura, “F2” para la segunda...
- 3 Adapte la función “Multiplot” para 3 columnas
- 4 Corra la nueva función Multiplot



Ejemplo 1: Elecciones y Twitter

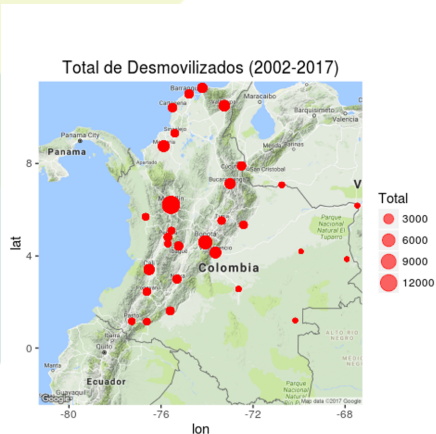
¿Qué debemos hacer para “reproducir” la Figura 3?

- 1 Genere cada gráfico por separado.
- 2 Asígnele un nombre a cada gráfico, siguiendo la buena práctica de “F1” para la primera figura, “F2” para la segunda...
- 3 Adapte la función “Mutiplot” para 3 columnas
- 4 Corra la nueva función Mutiplot
- 5 Exporte su gráfica al formato deseado (png, tiff, pdf)



...Un poquito más...

```
library(ggmap)
ColombiaMap <- get_map('Colombia', zoom = 6, scale = "auto", maptype = "terrain", source = "google")
d <- ggmap(ColombiaMap)
d + geom_point(colour = "red", alpha= .6, data = Cuadro1, aes(y = Latitud, x = Longitud, size = Total)) +
ggtitle("Total de Desmovilizados (2002-2017)")
```

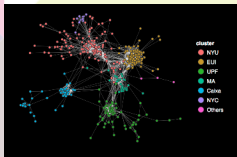


Recomendaciones para otros gráficos

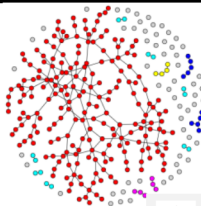
- 1 Inspírese mirando <http://www.r-graph-gallery.com/>
- 2 Seguir los ejemplos mostrados en **R Graphics Cookbook**
- 3 ¿Qué gráfico muestra más en menor cantidad de espacio?
- 4 Apóyese en los scripts generados por software comerciales. SPSS, por ejemplo, tiene la opción de generar los scripts para R.
- 5 Use otros paquetes como “vioplot”, “wordcloud”, “igraph”, “psych”, etc.



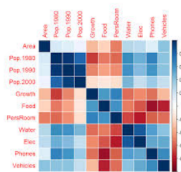
Recomendaciones para otros gráficos



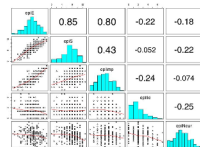
Rfacebook



igraph



corrplot

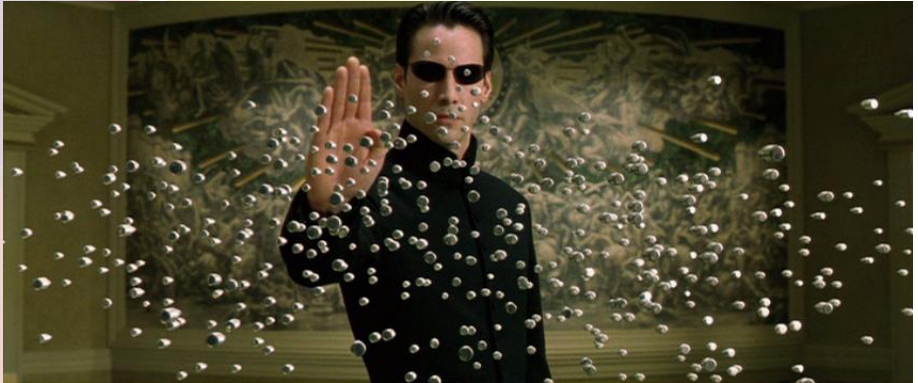


psych



aplpack

¿Preguntas?



¡Dispara!



¡Gracias!



- Correa, J. C., & Camargo, J. E. (2017). Ideological Consumerism in Colombian Elections, 2015: Links Between Political Ideology, Twitter Activity, and Electoral Results. *Cyberpsychology, Behavior, and Social Networking*, 20(1), 37–43. doi: 10.1089/cyber.2016.0402
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Wilkinson, L. (2006). *The grammar of graphics*. New York: Springer Science & Business Media.

