

Lic. Patricia Andrea Loto

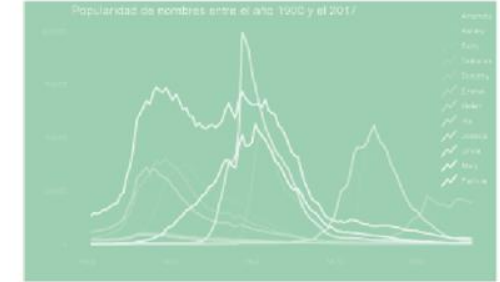
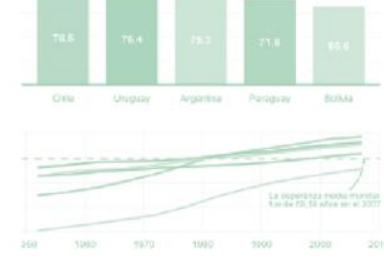
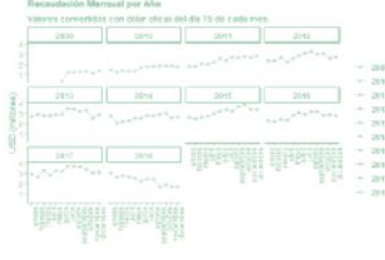
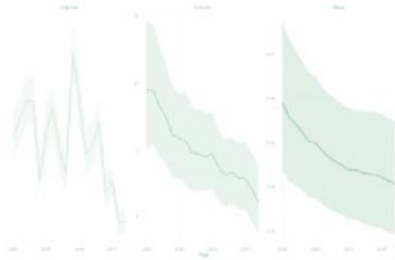
Desarrollo de Software
Gestión de Proyectos Tecnológicos
Análisis de datos

 Patricia Loto

 @patriloto

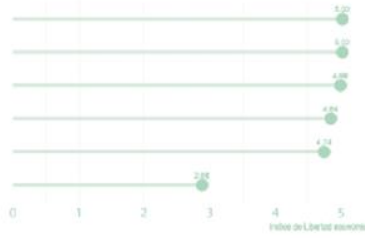
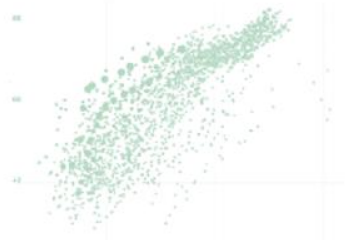


WOMEN IN DATA SCIENCE



VISUALIZACIÓN DE DATOS

en R con ggplot2



Hoy hablamos sobre...

PARTE 1:

- **¿De qué hablamos cuando hablamos de visualización de datos?**
 - ¿Qué es la visualización de datos? ¿Para qué visualizamos?
 - Importancia de la visualización dentro del proceso de la ciencia de datos
 - Cualidades de una buena visualización
- **Representando mis datos**
 - Componentes visuales: ¿Cuáles son los ingredientes de una visualización?
 - Visualizando con claridad
- **Un recorrido por los gráficos más típicamente utilizados**
 - Gráficos para representar: **cantidad**, **distribución**, **relación** y **dispersión**.
 - Gráficos elementales: Bar chart- Line Plot - Scatter Plot- Density Plot.



Hoy hablamos sobre...

PARTE 2:

- **El paquete ggplot2**

El paquete ggplot2 y la gramática de gráficos en capas

Capas de un gráfico

Sintaxis - Mi primer gráfico



- **¿Por dónde empiezo?**

Comunidades de aprendizaje: #Rladies #R4DSEs #DatosdemierRcoles

¿Cómo Participo?

- **Hands-on con ggplot2**

Practicamos con los datasets de:

- Países (versión al español de gapminder) - Legos- Babynames





Parte 1

¿De qué hablamos
cuando hablamos de
visualización de datos?

Podemos entender a la visualización como un medio que puede ser usado como una **herramienta** y a la vez como una **forma de expresar datos**.

*Source: Nathan Yau, escritor del libro **Data Points**.*



Visualización



El mundo
real

Datos

Formas y
colores



Interpretación

Source: Nathan Yau, escritor del libro **Data Points**.



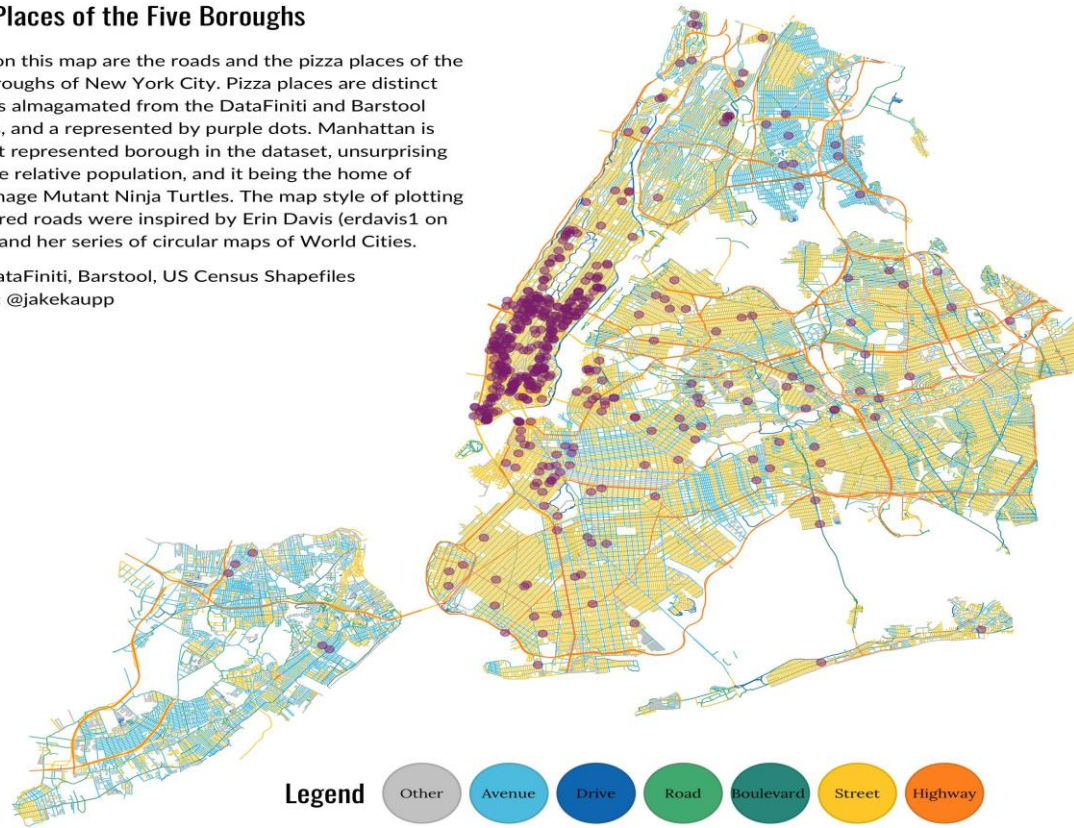
Ejemplos

Pizza Places of the Five Boroughs

Shown on this map are the roads and the pizza places of the Five Boroughs of New York City. Pizza places are distinct locations amalgamated from the DataFiniti and Barstool datasets, and are represented by purple dots. Manhattan is the most represented borough in the dataset, unsurprising given the relative population, and it being the home of the Teenage Mutant Ninja Turtles. The map style of plotting the colored roads were inspired by Erin Davis (erdavis1 on github), and her series of circular maps of World Cities.

Data: DataFiniti, Barstool, US Census Shapefiles

Graphic: @jakekaupp

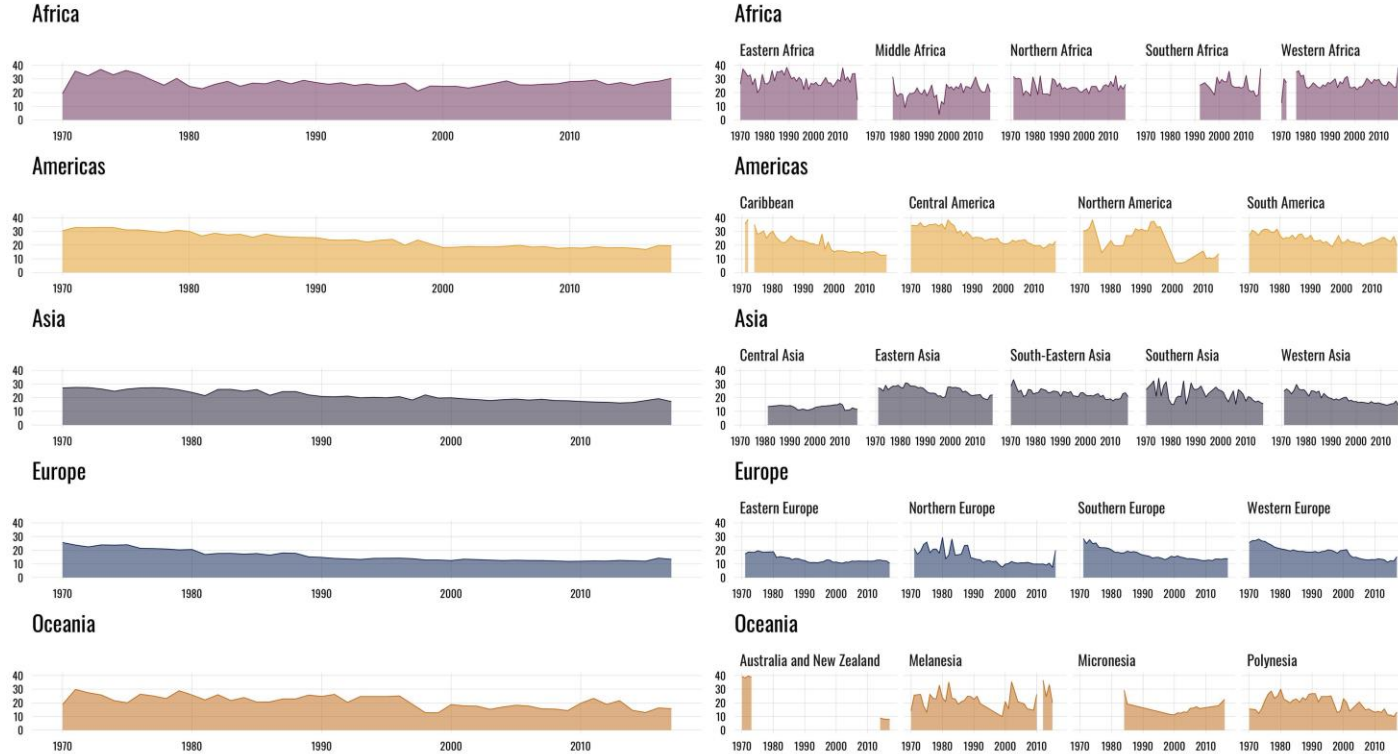


Source: @jakekaupp



Working to Two Sigma: Student Teacher Ratios Improving Since the 1970s

Illustrated below is the average student to teacher ratio across each continent (left column) and region (right column). Continent and region assigned from iso3c coding of country name and are consistent with the World Bank Development Indicators.



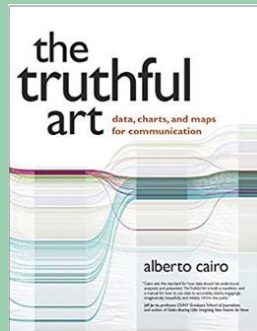
Data: UNESCO Institute of Statistics | Graphic: @jakekaupp

Source: @jakekaupp



¿Y Para qué visualizamos?

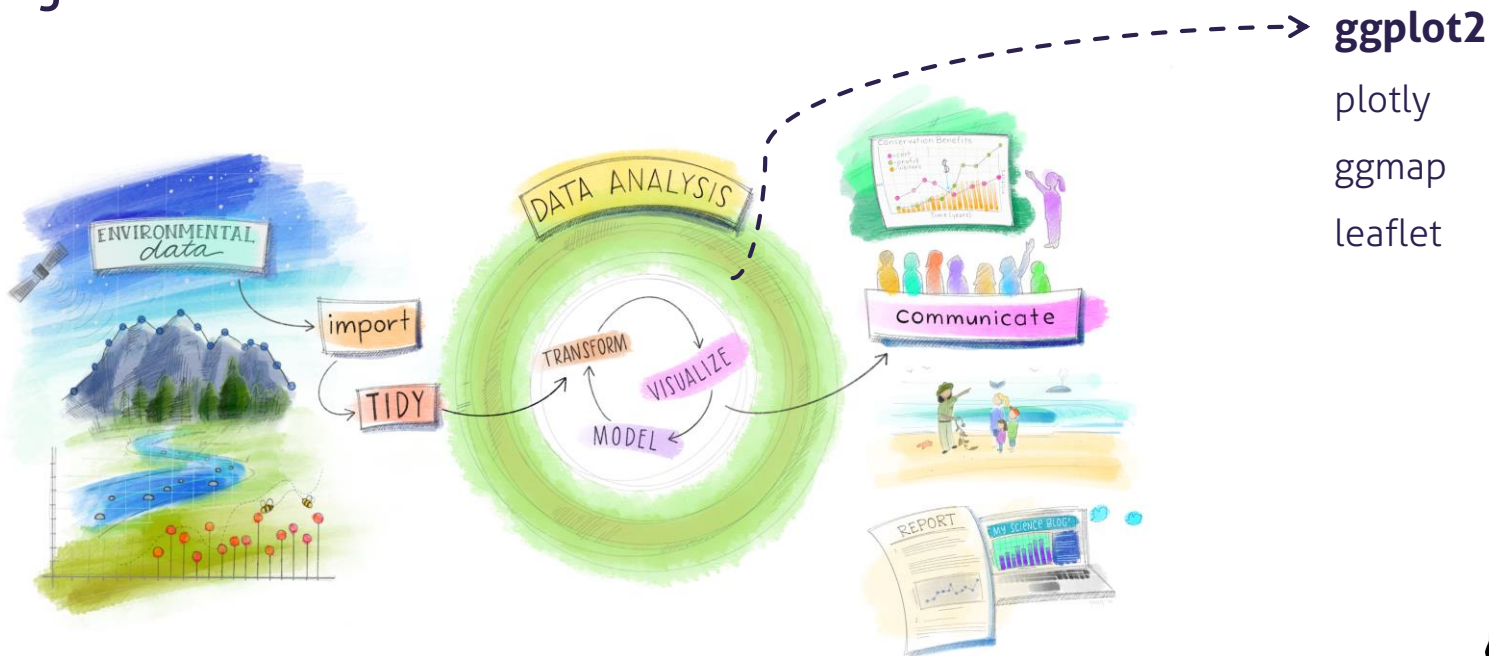
Para expresar y transmitir el significado de los datos de manera precisa, clara, atractiva, imaginativa, bella y confiable con el objetivo de informar al público, a nuestro público.



Source: Alberto Cairo, escritor del libro **The Truthful Art**.



La visualización dentro del esquema de trabajo en ciencia de datos



Art: Allison Horst

A hand is visible on the right side of the frame, holding a piece of white chalk and writing on a green chalkboard. The chalkboard has some faint, partially visible writing in white chalk, including the word 'Konm' and some numbers. A dark purple rectangular box is overlaid on the left and center of the image, containing white text.

Una imagen vale más que mil
palabras pero...

Cualidades de una buena visualización

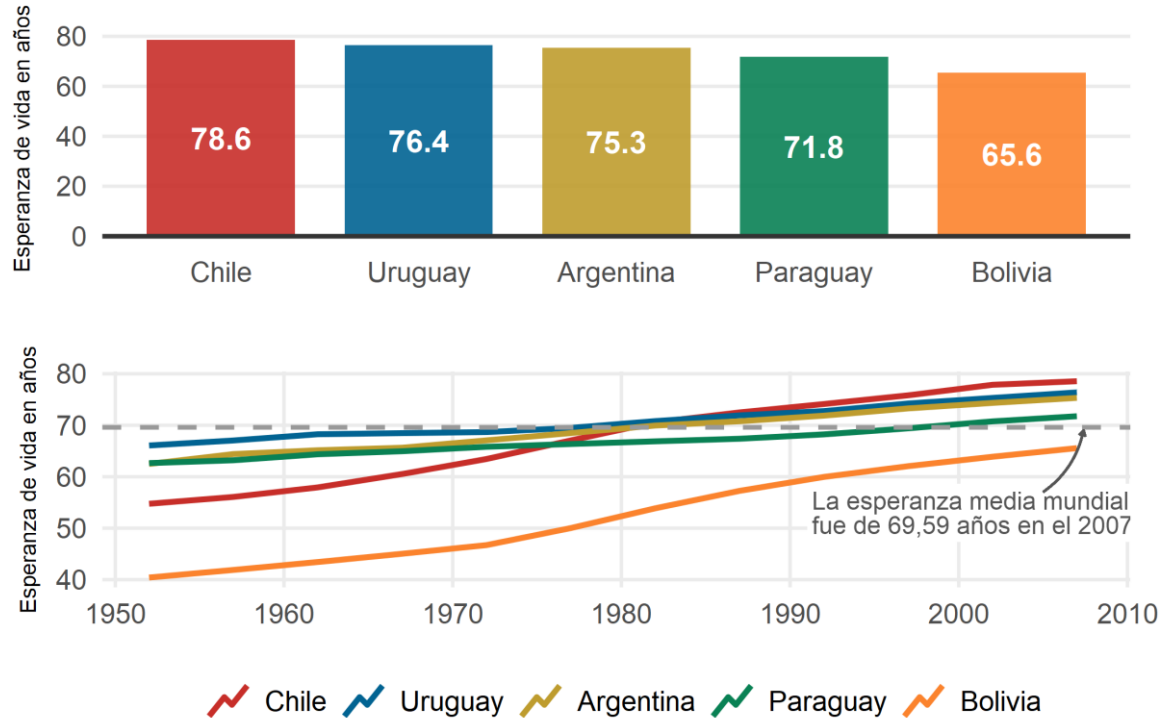
Las cinco cualidades de una visualización memorable:

- Que sea agradable a la vista
- Que sea funcional
- Que muestre hallazgos
- Que esté basada en datos confiables, es decir, que transmita la verdad



Expectativa de vida de Argentina y países limítrofes

Período 1950-2017



Data: países del paquete datos | Por Patricia Loto



WOMEN IN DATA SCIENCE

Representando mis datos



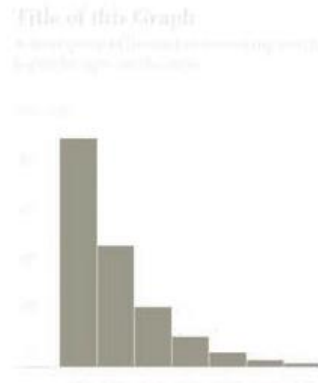
¿Cuáles son los ingredientes de una **visualización**?

- Cada visualización se construye sobre **datos** y **cuatro componentes**:



¿Cuáles son los ingredientes de una visualización?

1. Señales visuales +



Visual Cues

Visualization involves encoding data with shapes, colors, and sizes. Which cues you choose depends on your data and your goals.



¿Cuáles son los ingredientes de una visualización?

1. Señales visuales +
2. Sistemas de coordenadas +

Title of Your Graph

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Coordinate System

You map data differently with a scatterplot than you do with a pie chart. It's x- and y-coordinates in one and angles with the other; it's cartesian versus polar.



WOMEN IN DATA SCIENCE



¿Cuáles son los ingredientes de una visualización?

1. Señales visuales +
2. Sistemas de coordenadas +
3. Escalas +



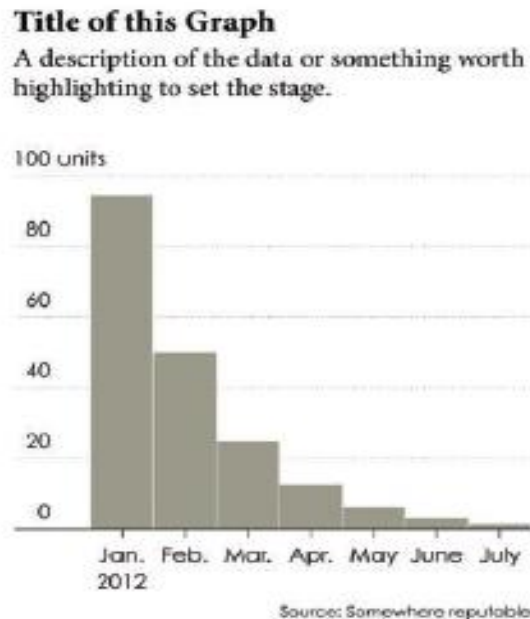
Scale

Increments that make sense can increase readability, as well as shift focus.

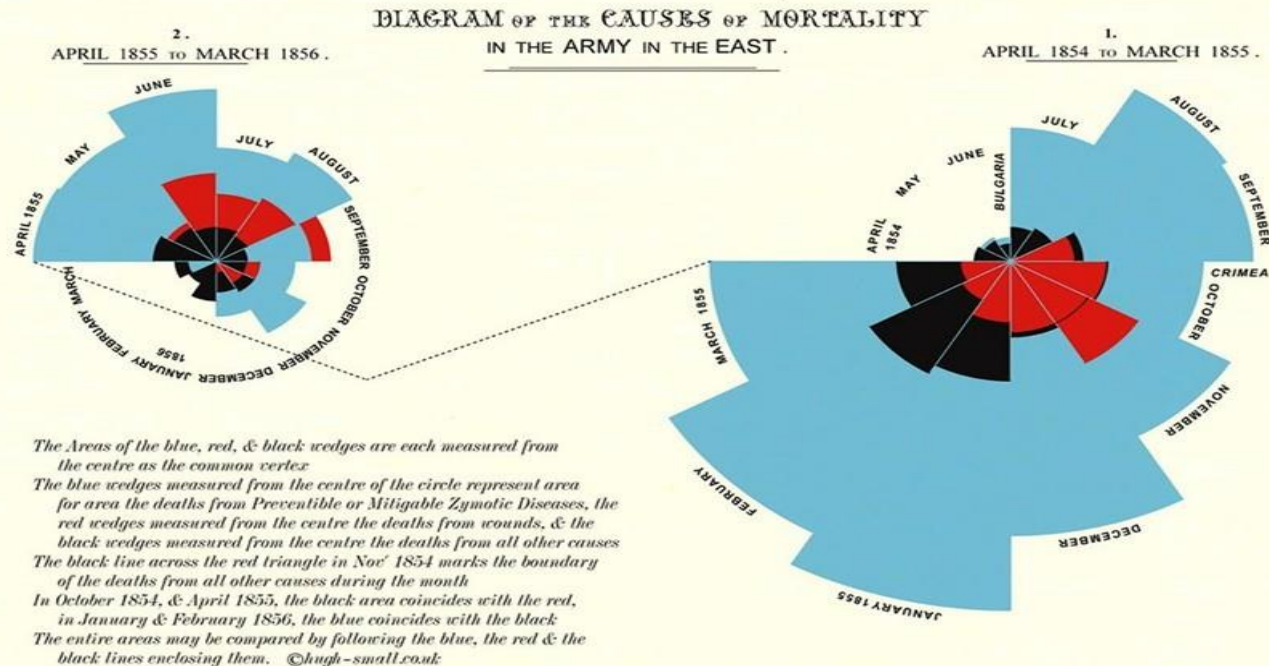


¿Cuáles son los ingredientes de una visualización?

1. Señales visuales +
2. Sistemas de coordenadas +
3. Escalas +
3. Contexto

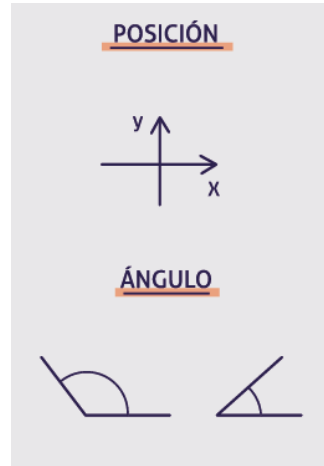


¿Qué sucede si unimos todos los ingredientes?



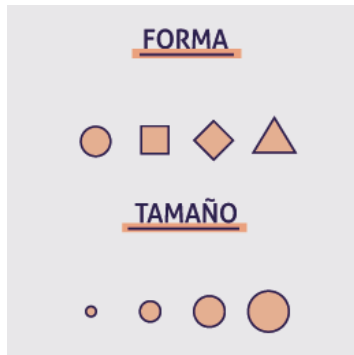
Componentes visuales

1. Señales visuales



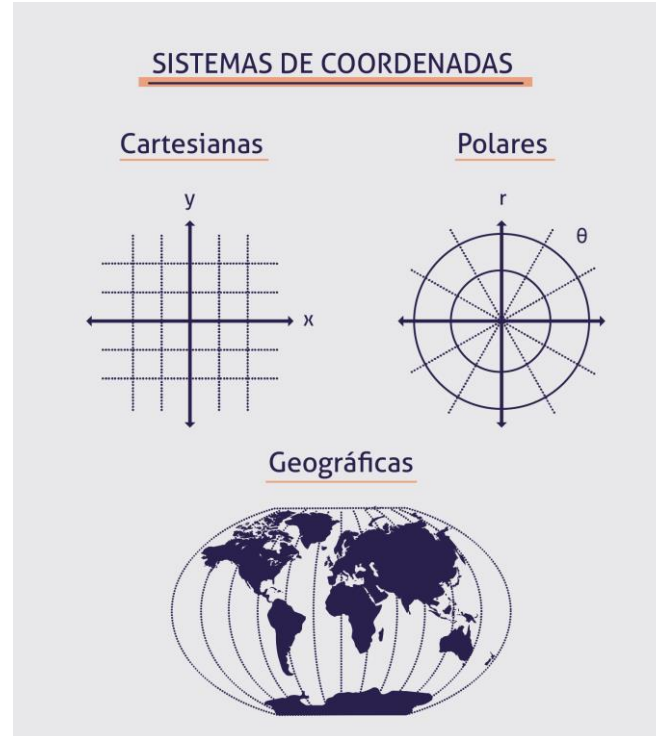
Componentes visuales

1. Señales visuales



Componentes visuales

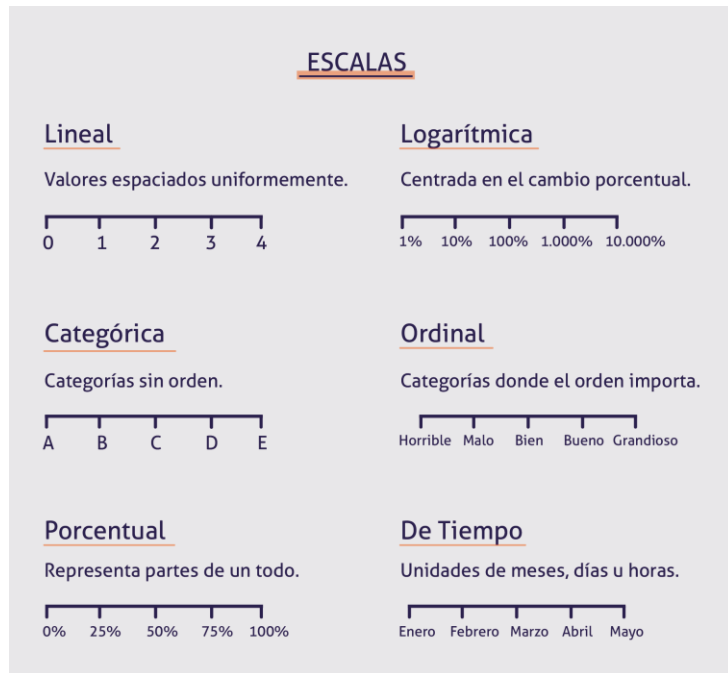
2. Sistemas de coordenadas



Componentes visuales

3. Escalas

4. Contexto



Source: Nathan Yau, escritor del libro *Data Points*.





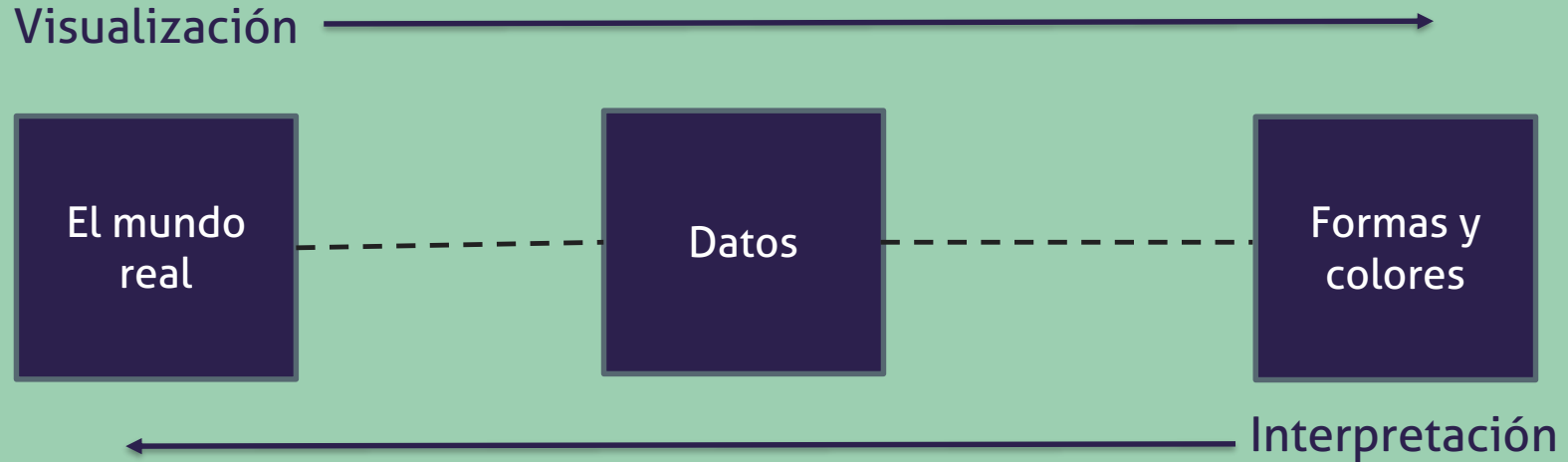
Visualizando con claridad

“When you use graphics to present results to other people, you must make your graphics readable to those who don’t know your data as well as you do.”

*Source: Nathan Yau, escritor del libro **Data Points**.*



Visualizando con claridad

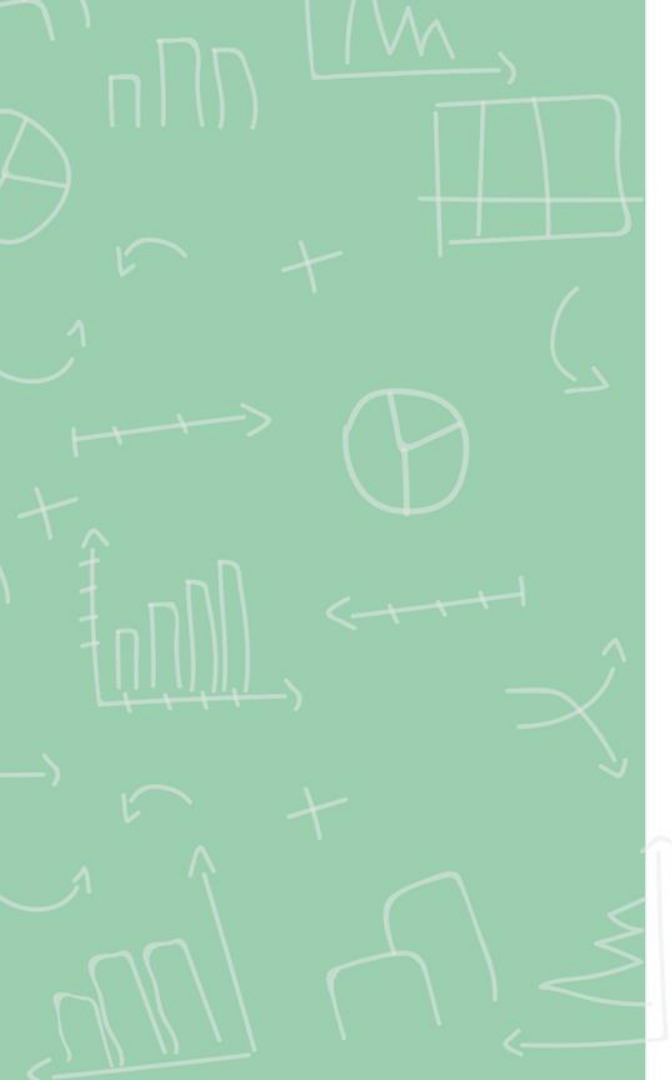


Source: Nathan Yau, escritor del libro *Data Points*.



Recursos que favorecen la interpretación de los gráficos

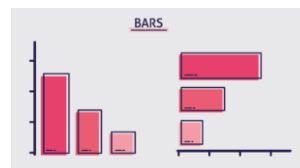
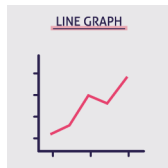
- Jerarquía en la visualización
- Resaltado de la información más importante
- Anotaciones: brindando contexto a los datos
- Medidas estadísticas: media, quartiles, etc.
- Color



Recorrido por los gráficos más típicamente utilizados

Tipos de gráficos

- Cantidad
- Distribución
- Proporción
- Relación entre 'x' e 'y'
- Dispersión
- Datos geoespaciales



Source: Clauss Wilke, escritor del libro [Fundamentals of Data Visualization](#).



Cantidad (una variable)

Gráfico de barras

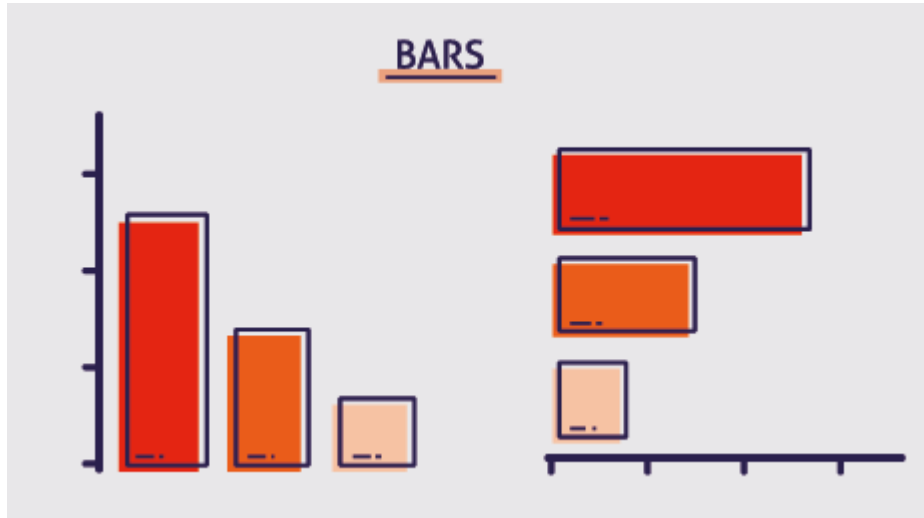
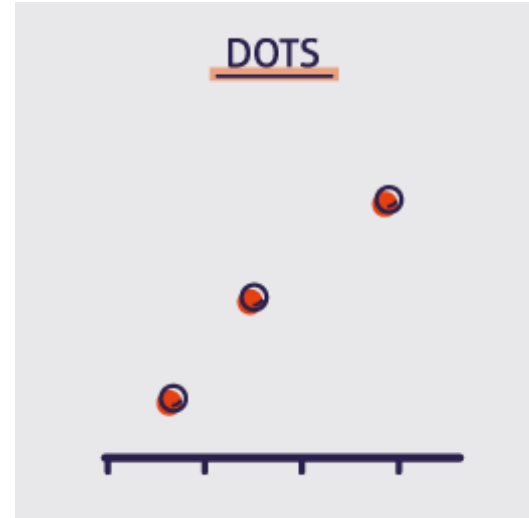


Gráfico de puntos



Cantidad (múltiples variables)

Gráfico de barras agrupadas

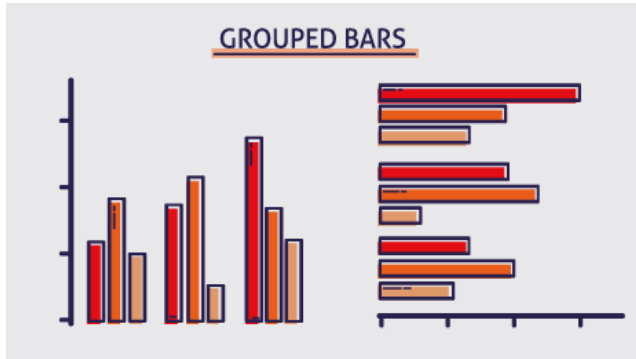
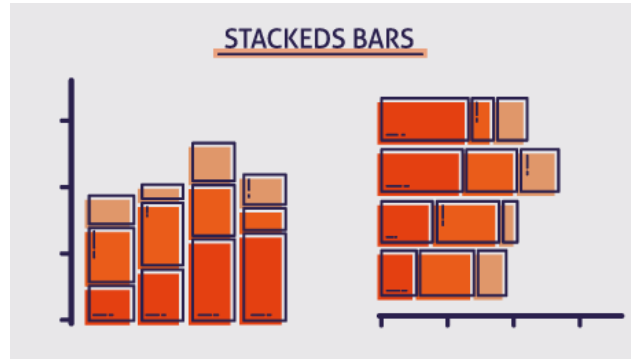
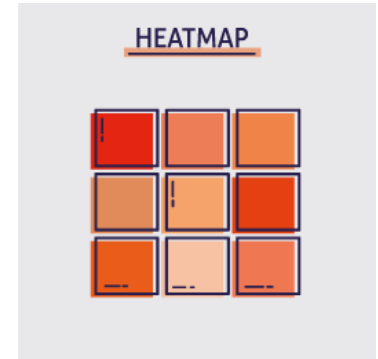


Gráfico de barras apiladas



Mapa de calor



Distribución Simple

Histograma

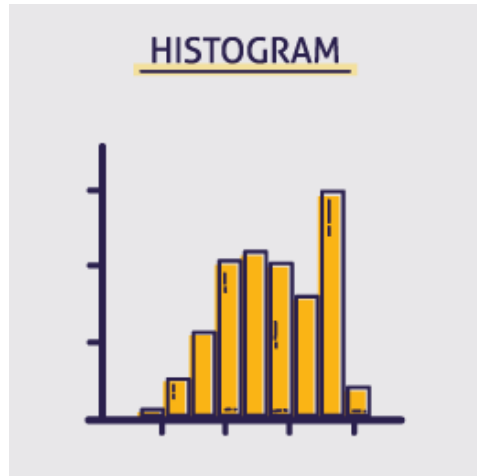
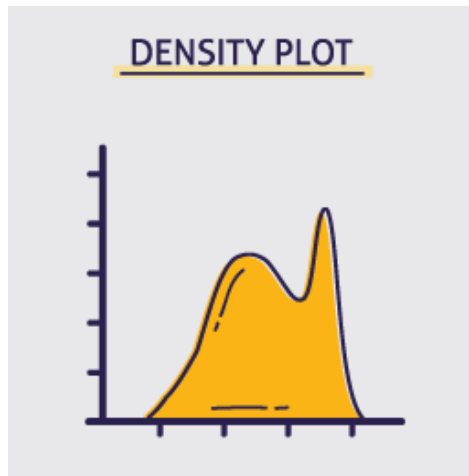
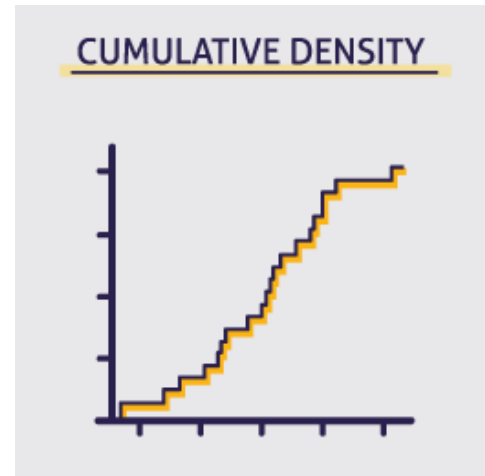


Gráfico de Densidad

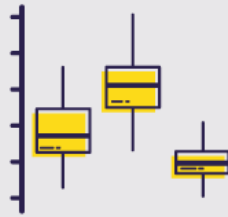


Densidad Acumulada

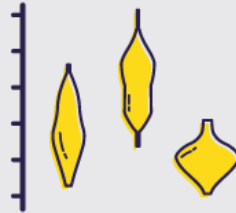


Distribución Múltiple

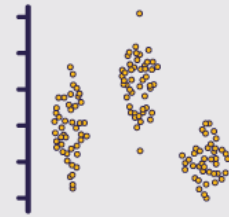
BOXPLOTS



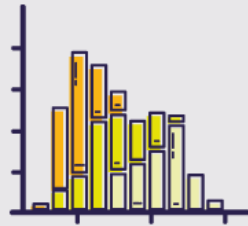
VIOLINS



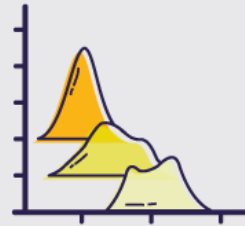
SINA PLOTS



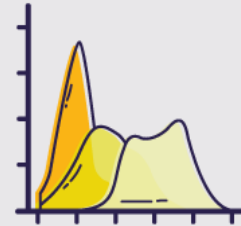
STACKED HISTOGRAMS



RIDGELINE PLOT



OVERLAPPING DENSITIES



Proporción (una variable)

Gráfico de torta

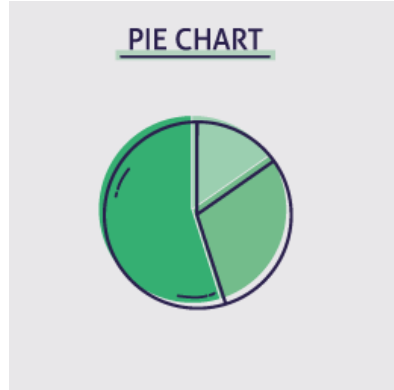
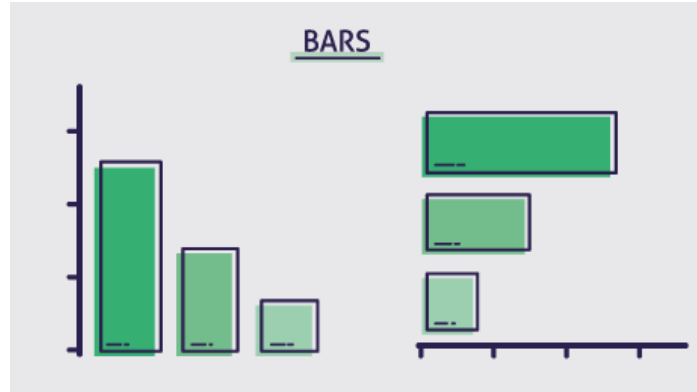


Gráfico de barras



Proporción (múltiples variables)

Múltiples gráficos de torta

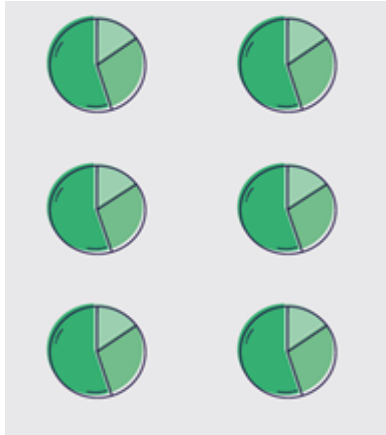
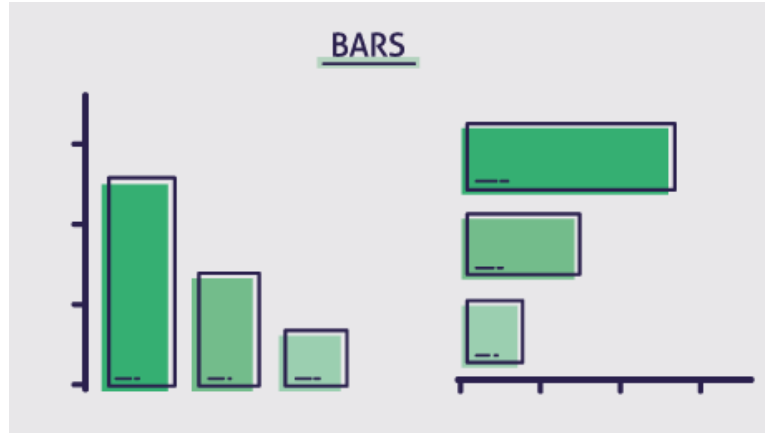
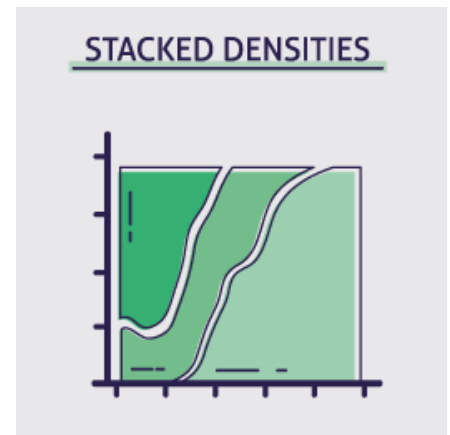


Gráfico de barras

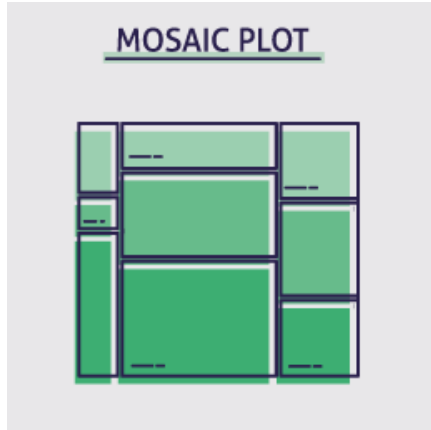


Densidades apiladas



Proporción (múltiples variables)

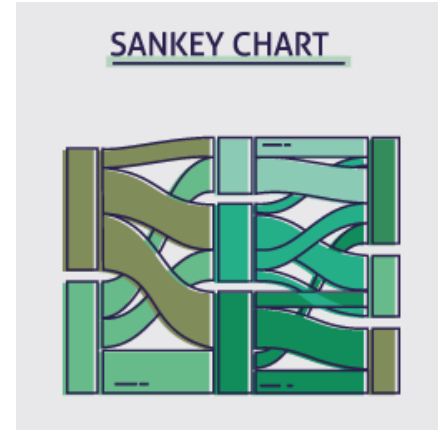
Gráfico de Mosaico



Mapa de árboles

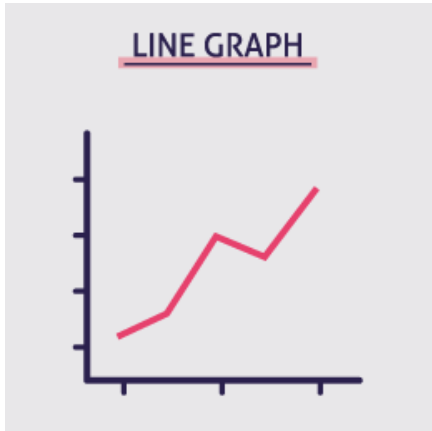


Diagrama de sankey



Relación entre 'x' e 'y' (una variable)

Gráfico de líneas



Scatterplot conectado

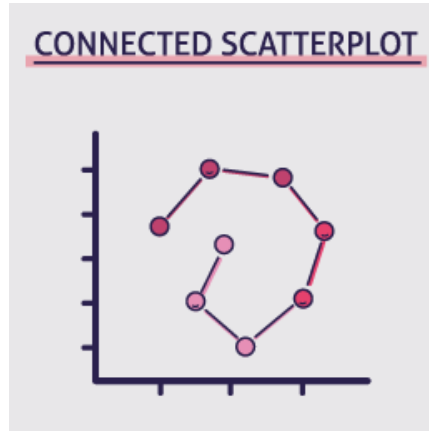
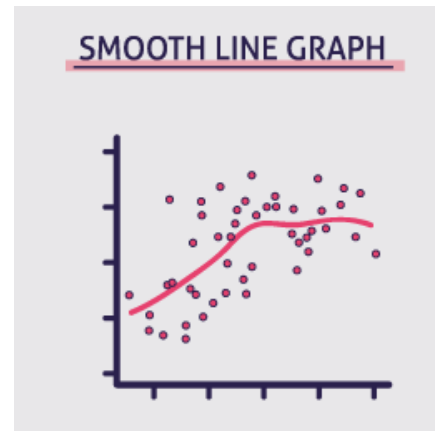
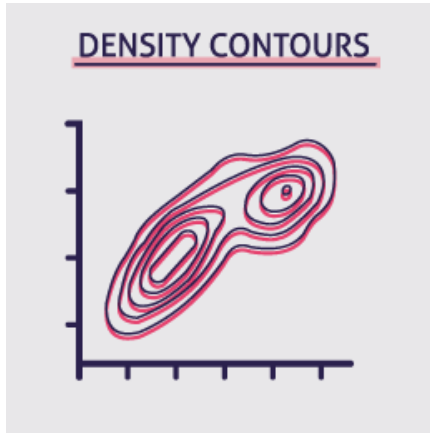


Gráfico de líneas
suavizada



Relación entre más de una variable

Gráfico de líneas



Hex Bins

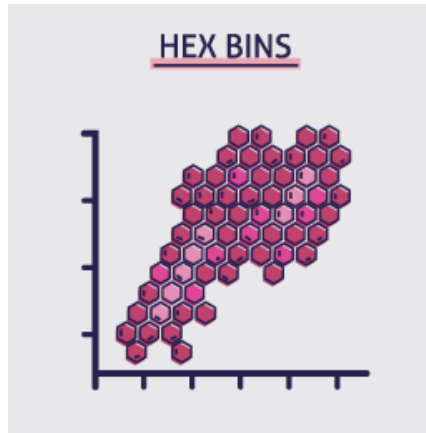
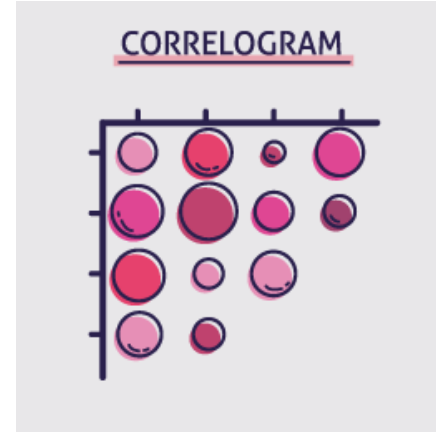


Gráfico de correlación



Dispersión

Gráfico de dispersión

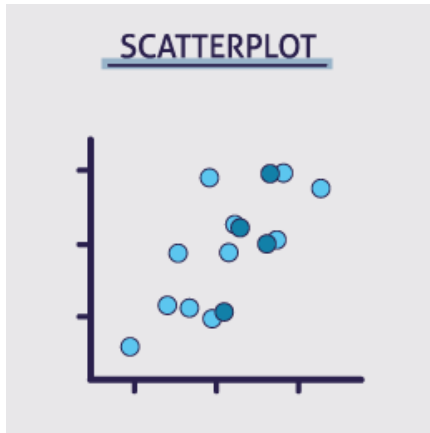


Gráfico de burbujas

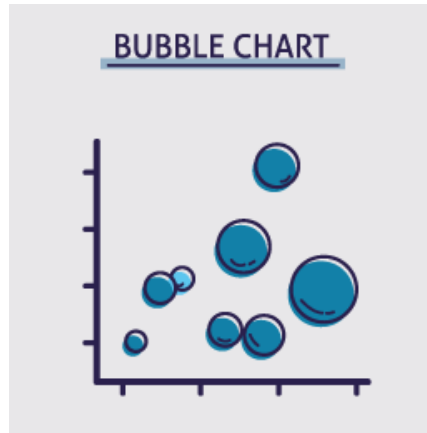
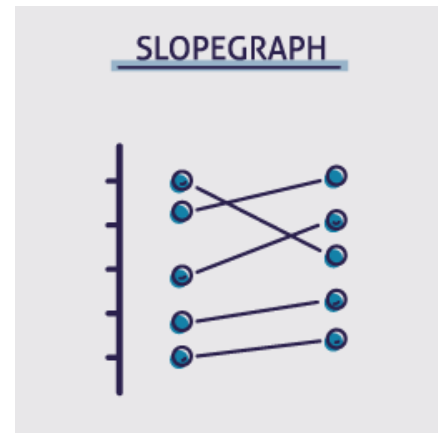
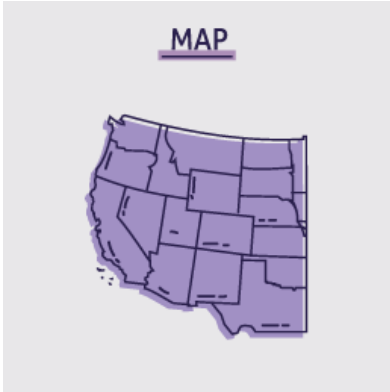


Gráfico de líneas suavizada

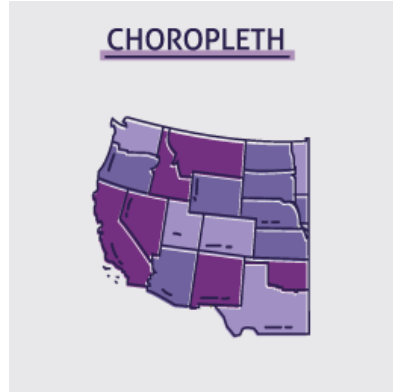


Datos geoespaciales

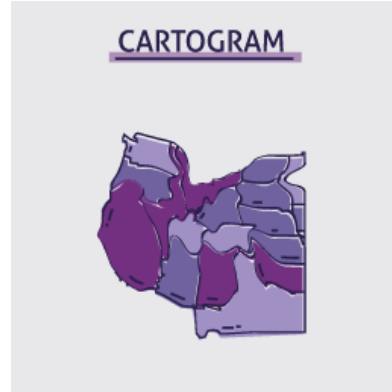
Mapa



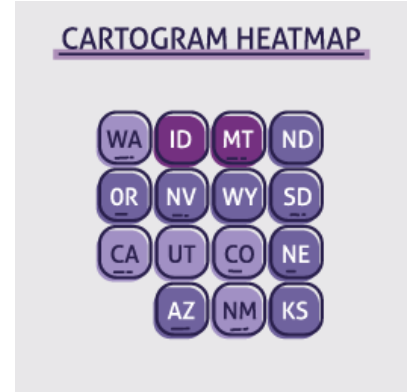
Mapa coroplético



Cartograma



Cartograma Heatmap



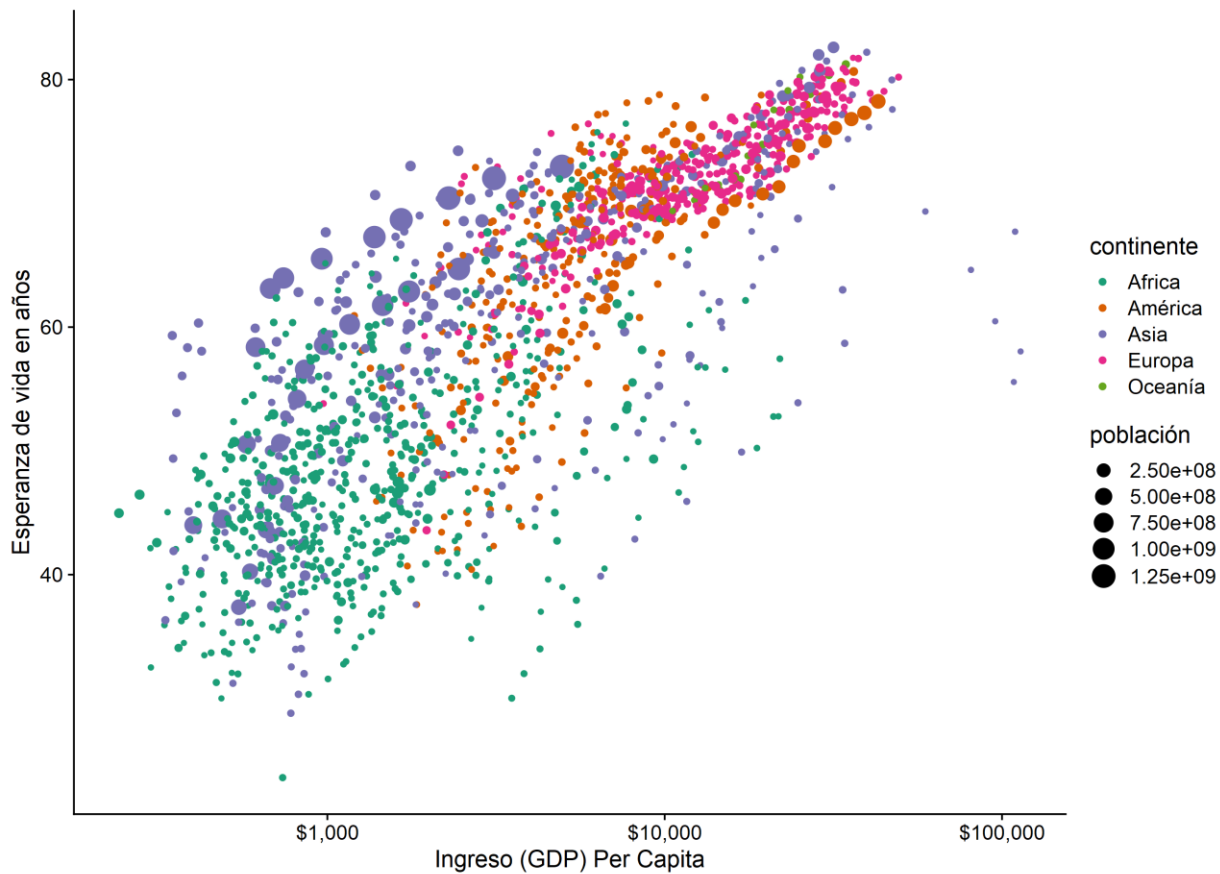
Parte 2



¿Qué hay atrás de cada gráfico?



Crecimiento económico y esperanza de vida



DataSource: Gapminder.



WOMEN IN DATA SCIENCE


¿Qué hay atrás de cada gráfico?

- ✓ Cada **observación** está representada con un **punto**, cuya posición está dada por **dos variables** (posición horizontal y vertical).
- ✓ Cada punto tiene tamaño y color, estos atributos son denominados elementos estéticos (o aes por su denominación en inglés: aesthetics)
- ✓ Los aes son propiedades que pueden ser percibidas en el gráfico.

Cada **aes** puede ser mapeado a una variable o fijado en un valor constante.

El paquete ggplot2



ggplot2 es uno de los  más populares para visualización de datos dentro de la comunidad R.

Fue desarrollado por **Hadley Wickham** (2008) y está basado en la **gramática de gráficos en capas**.

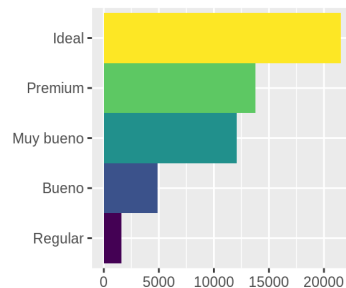
Es parte de un conjunto de paquetes que tiene foco en la ciencia de datos llamado **Tidyverse**.



Gramática de gráficos en capas o layered grammar of graphics

La Gramática de gráficos en capas (basada en "The Grammar of Graphics" by Wilkinson, Anand, and Grossman) nos permite conocer:

- ¿Qué es un gráfico?
- ¿Cuáles son los componentes de un gráfico?
- ¿Cómo describir y crear un gráfico?

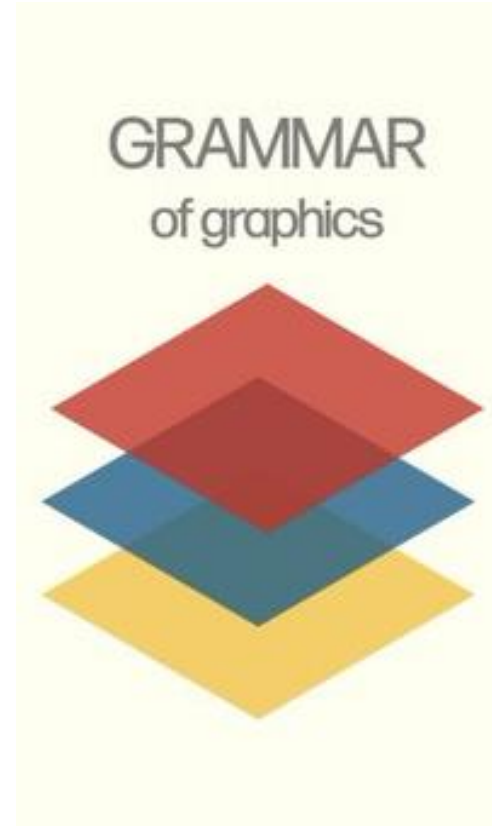


Capas de un gráfico

Función geom

Atributos estéticos

Datos



Tres componentes de todo gráfico

Datos (data):
nuestro set de datos

**2. Atributos o
elementos estéticos
(aes):**

un conjunto de mapeos
estéticos entre las
variables y las
propiedades visuales

3. Capas (layers):

al menos una capa que
describe cómo representar
cada observación. Cada **capa**
tiene un **objeto geométrico**,
una transformación estadística,
una posición y opcionalmente
un conjunto de mapeos
estéticos.

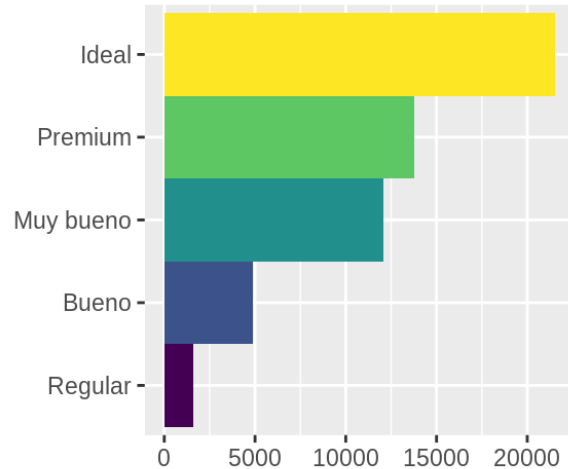
Otros Componentes

- **Escalas (Scales):** mapea valores en el espacio de los datos a valores en el espacio estético (ej. color, tamaño, forma o posición)
- **Coord (coordenadas):** normalmente Cartesianas..
- **Facets (facetas):** cómo se arregla el display cuando son muchos gráficos
- **Theme (temas),** items para mejorar el gráfico como fuente, tamaño, color, background.

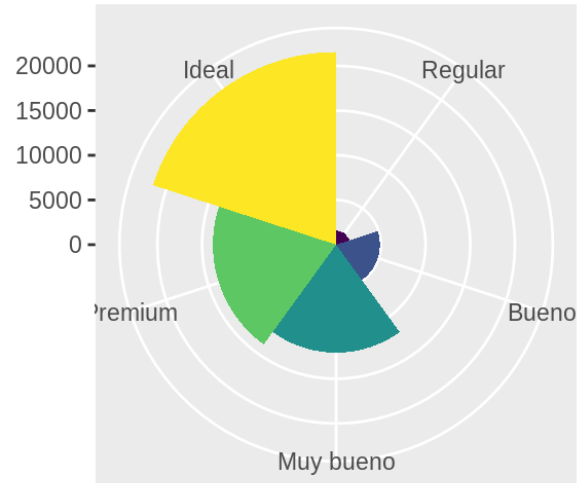


Sistema de coordenadas

■ Cartesianas

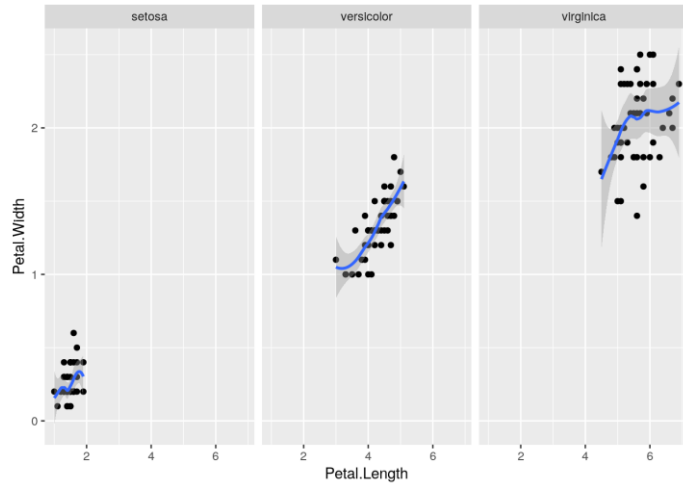


■ Polares

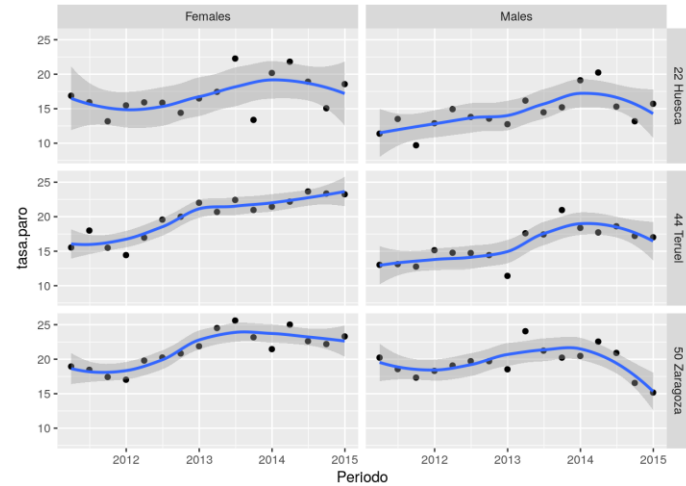


Facets

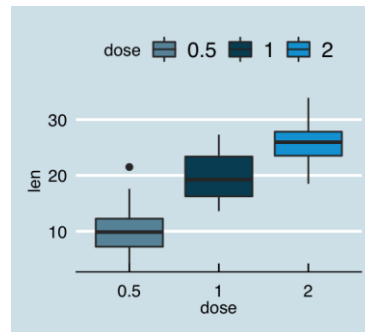
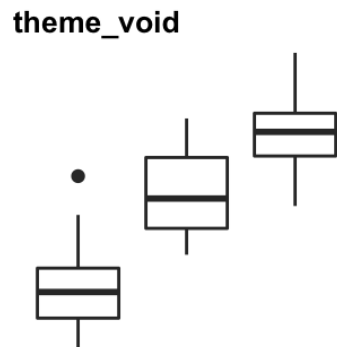
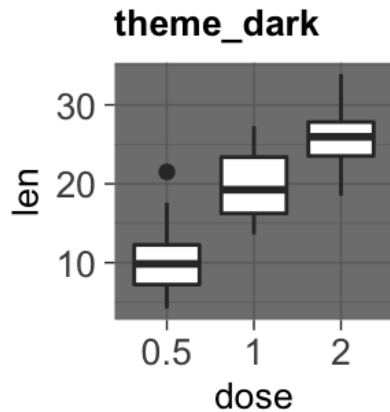
■ Facet_wrap



■ Facet_grid



Themes





Ggplot2: primeros pasos



ggplot2: Build a data MASTERpiece



Art: Allison Horst



WOMEN IN DATA SCIENCE

Sintaxis de ggplot2

```
p ← ggplot(data= <data>,  
           mapping= aes(<aesthetic> = <variable>,  
                        <aesthetic> = <variable>,  
                        <...> = <...>)
```

```
p + geom_<type>(<...>)+  
    scale_<mapping>_<type>(<...> )+  
    coord_<type>(<...> )+  
    labs(<...> )
```

1 er Paso: datos

1 DATOS ORDENADOS

p ← **ggplot** (data = paises , ...

pib_per_capita	esperanza_de_vida	poblacion	continente
340	65	31	Europa
227	51	200	América
909	81	80	Europa
126	40	20	Asia

2do paso: Mapeos estéticos

3er paso: Función geom

2 MAPEOS

```
p ← ggplot (data = paises ,  
            mapping = aes( x = gdp ,  
                           y = lifexp , size = poblacion ,  
                           color = continente) )
```

3 FUNCIÓN GEOM

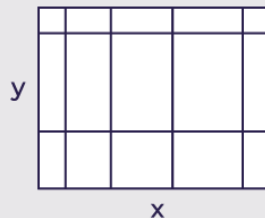


```
p + geom_point()
```

4to paso: Coordenadas y escalas

4 SISTEMAS DE COORDENADAS Y ESCALAS

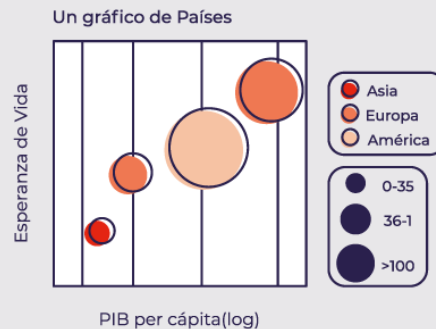
p + `coord_cartesian()` +
`scale_x_log10()`



5to paso: Etiquetas y guías

5 ETIQUETAS Y GUÍAS

p + `labs` (x = "PIB per cápita(log)",
Y = "Esperanza de vida",
title = "Un gráfico de Países")



¿Por dónde empiezo?

¿Cómo visualizar y no frustrarse en el intento?

- **Sé perseverante:** la única manera de aprender es practicando y experimentando.
- **Sé paciente contigo y con R.**
- Trabaja de manera **incremental**, comienza por un pequeño gráfico y luego en cada iteración mejóralo.
- No estás sólo, busca una **comunidad** abierta e inclusiva de la que puedas aprender, ejemplo **Rladies** .



¿Cómo visualizar y no frustrarse en el intento?

- Seguí en twitter a gente de la comunidad de R que se dedica a lo que vos querés aprender.

@CedScherer

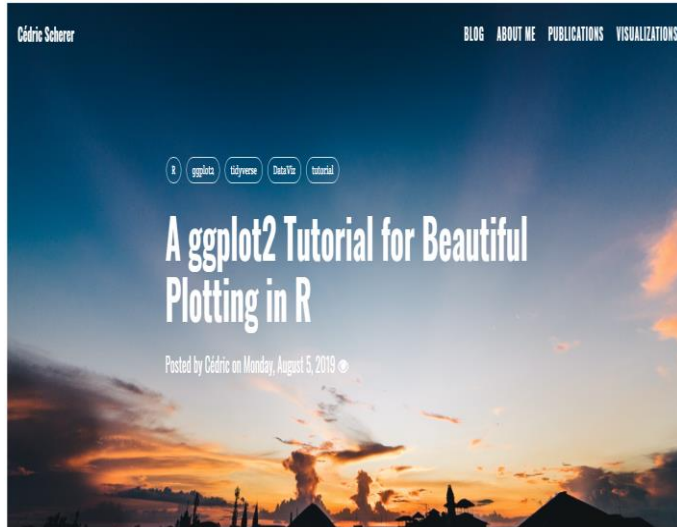
@jbkunst



@r0mymendez

@watzoever

¿Cómo visualizar y no frustrarse en el intento?



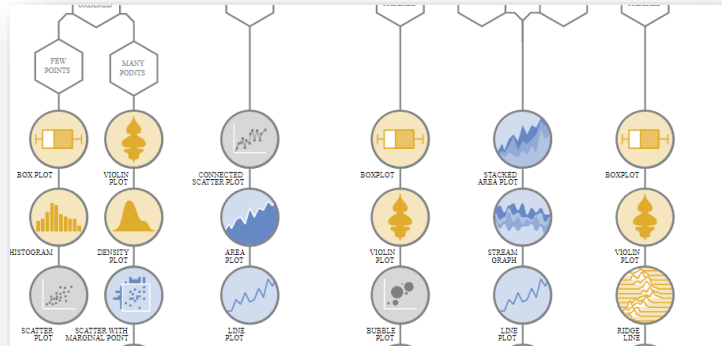
Blogs con mucho para aprender

<https://cedricscherer.netlify.com>

/



¿Cómo visualizar y no frustrarse en el intento?



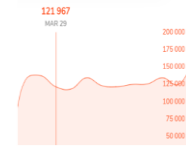
<https://www.data-to-viz.com>

Data, Code and Visualization

April 8, 2019

When charts are integrated in the web page

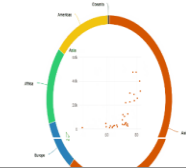
I saw a chart yesterday, a chart without frame, without title or captions just a very hidden chart at bottom of the web page. It will be a nice time to get the most interesting visual features of it.



Feb. 4, 2019

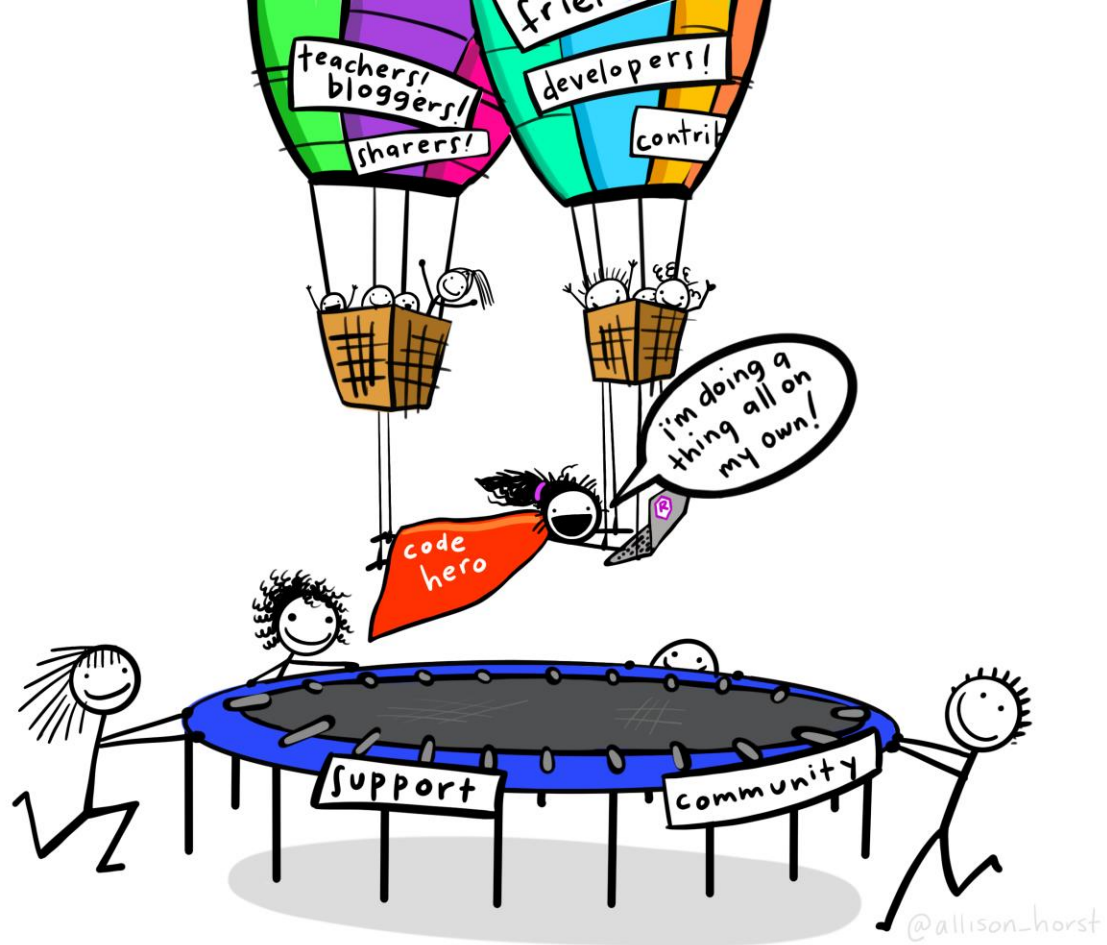
Using tooltips in unexpected ways

Highcharts tooltips are one of most underrated features. In the tooltips you can easily render html, that means we can put images, tables even charts. So the possibilities are almost limitless.



<https://jkunst.com/blog>

Comunidades



Source: Allison Horst

Comunidad de aprendizaje por proyecto



“R for Data Science”
en español

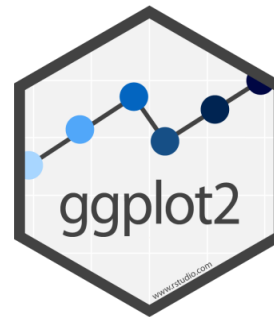


Comunidad RLadies



Organización global que promueve la diversidad de género en la comunidad de R mediante meetups y mentorías en un espacio amigable y seguro.

Tu turno



¡Gracias!

¿Alguna pregunta?

@patriloto

patricialoto@hotmail.com

