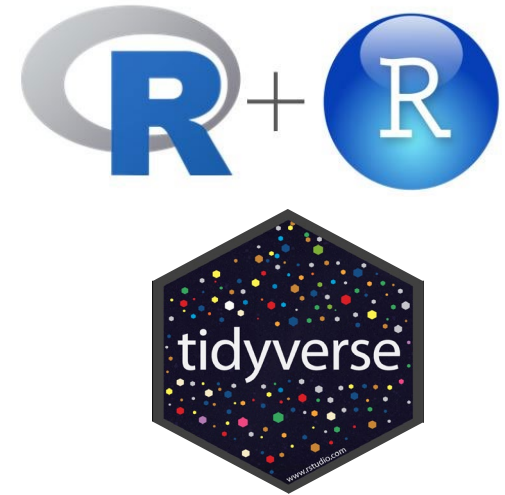


Manipulación y visualización de datos con R y Tidyverse



TALLER DE INTRODUCCIÓN A R (PARTE 1)

JUNIO – 2020

AMADEO GUZMÁN C.

¿Qué es R?

¿Por qué deberíamos aprender a usarlo?

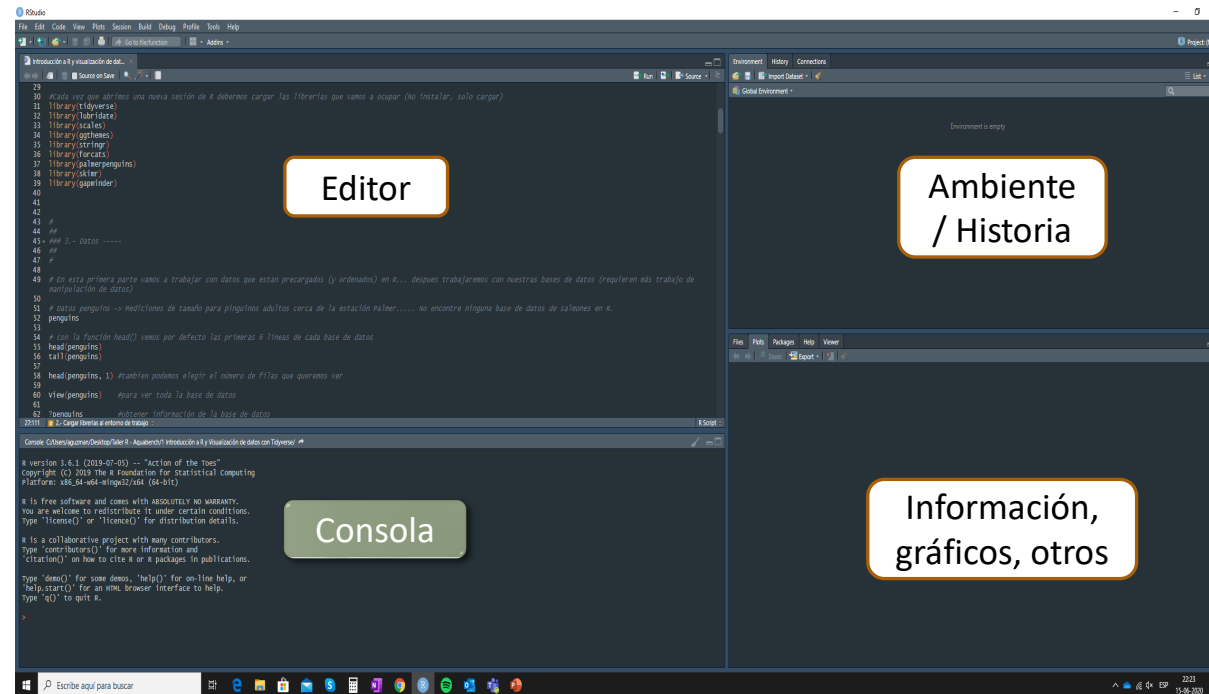
- **R** es un entorno y lenguaje de programación con enfoque en análisis estadístico y una gran capacidad para crear visualizaciones de alto nivel.
- Herramienta flexible y que puede ampliarse fácilmente mediante el uso de paquetes y funciones...incluso podemos crear nuestras propias funciones
- Programación orientada a objetos (datos, funciones, variables, resultados, etc... todo se pueden guardar en la memoria activa del computador con un nombre específico y las podemos usar en cualquier momento)
- Todo el flujo de trabajo queda escrito en el código – Trabajo/Investigación Reproducible.

Y.....

¿Qué es RStudio?

¿Por qué deberíamos usarlo?

- **RStudio** es una interfaz de usuario (Entorno de Desarrollo Integrado IDE)
- Nos permite trabajar de manera más “amigable” con R.



Al principio R puede ser un poco...

“Complicado”R utiliza una consola de comandos y no una interfaz amigable para los usuarios, lo cual puede parecer un gran desafío cuando recién empiezas a utilizar este programa.

“Frustrante”Tiene una curva de aprendizaje inicial lenta, lo cual muchas veces nos va a hacer querer volver a nuestra zona de confort, en base a clicks del mouse sobre menús desplegables o plantillas Excel intuitivas (pero con limitada capacidad).

Para que esta primera etapa de aprendizaje sea menos “complicada” y “frustrante” vamos a conocer el funcionamiento de R a través de **Tidyverse** y la **visualización de datos** con fines exploratorios.

¿Qué es Tidyverse?

¿Por qué deberíamos aprender a usarlo?



Tidyverse es una colección de paquetes de R diseñado para **ciencia de datos**.

Todos los paquetes comparten una filosofía de diseño, gramática y estructura.

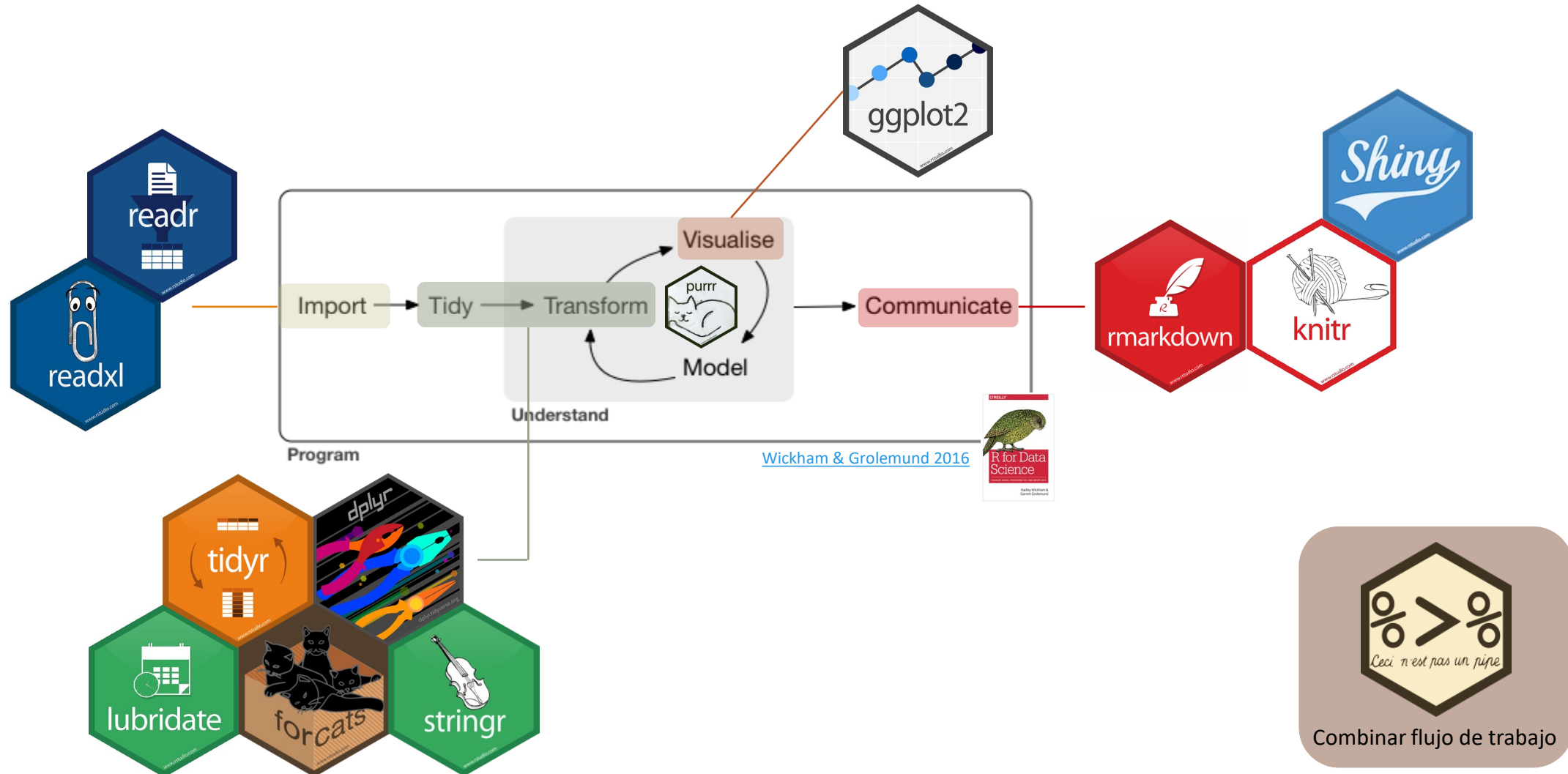
Se pueden combinar todas las funciones de los distintos paquetes en un mismo flujo de trabajo.

Permite que las operaciones comunes en el proceso de análisis de datos / ciencia de datos sean más intuitivas.



¿Qué es la ciencia de datos?

¿Cómo la aplicamos en nuestro trabajo?

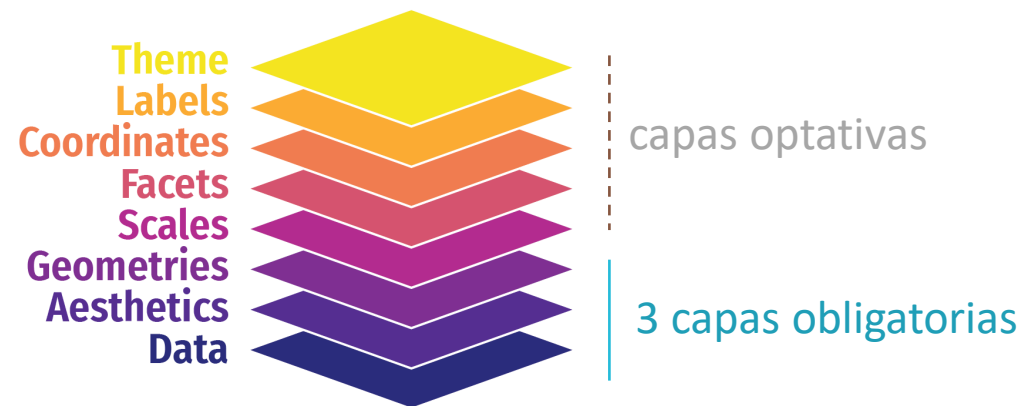


Hoy vamos a trabajar principalmente con....

ggplot2

ggplot2 es un paquete de visualización de datos para la programación estadística en lenguaje R. Fue creado en el año 2005 por Hadley Wickham y esta basado en The Grammar of Graphics.

Su funcionamiento y filosofía se basa en un esquema general que divide los gráficos en componentes semánticos como **escalas** y **capas**.



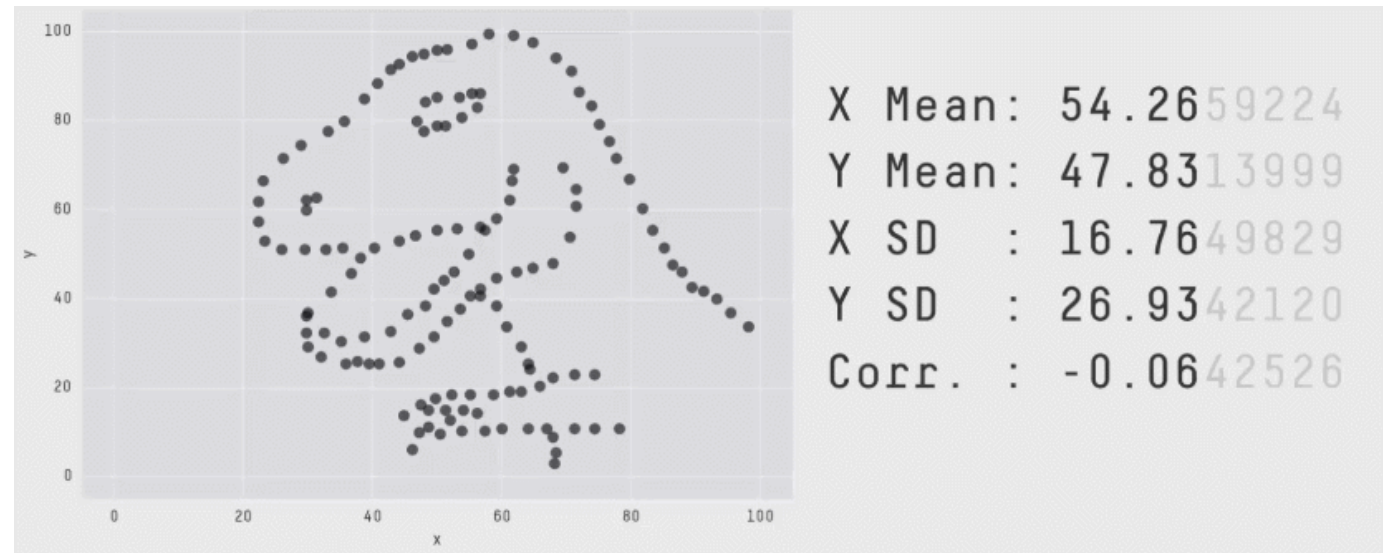
<https://ggplot2.tidyverse.org/reference/index.html#section-layer-geoms>

¿Por qué o cuando visualizar nuestros datos?

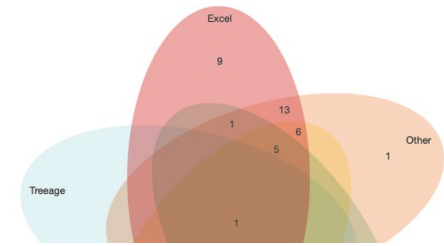
- Explorar y conocer nuestros datos
- Validar datos
- Transmitir información relevante / hallazgos importantes.
- Comunicar resultados con nuestro público objetivo (colegas, clientes, estudiantes, jefes, etc.)

Una imagen vale más que mil palabras.....

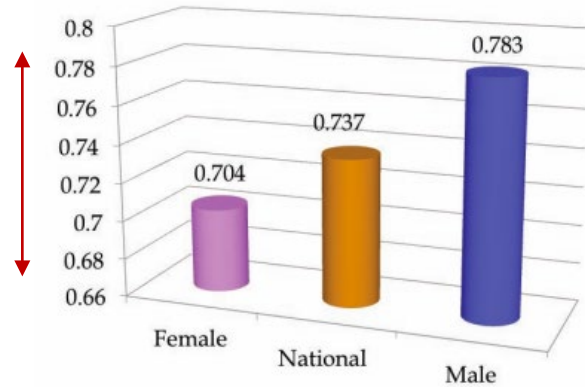
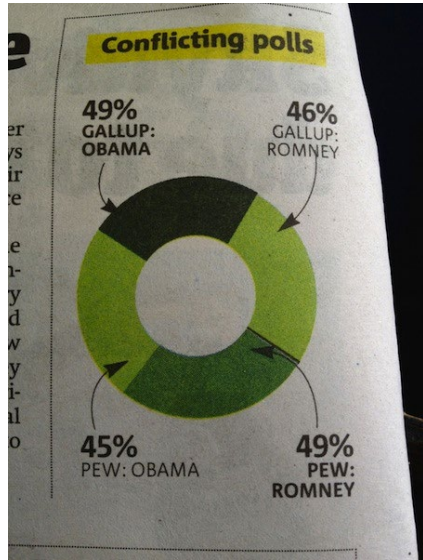
También aplica al análisis de datos



<https://www.autodeskresearch.com/publications/samestats>

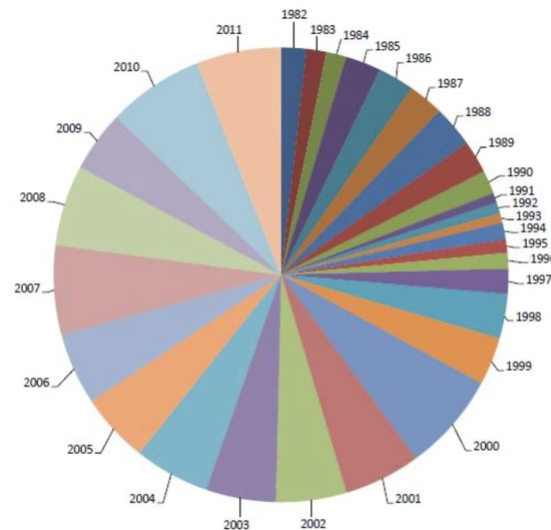
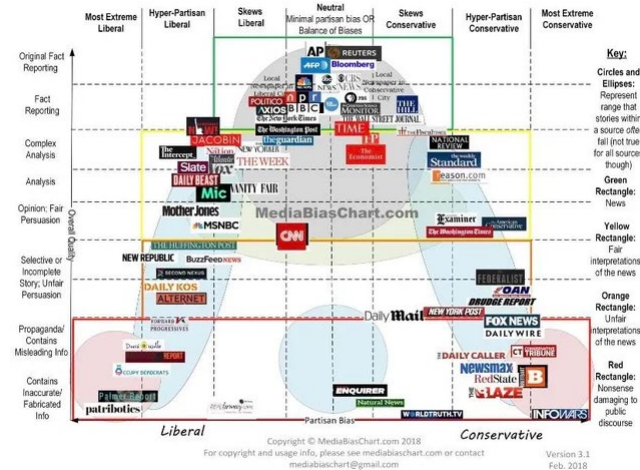
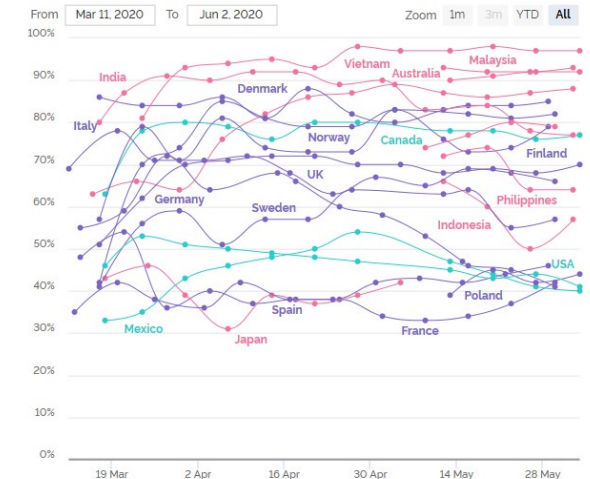


Algunas cosas que deberíamos evitar....



YouGov COVID-19 tracker: government handling

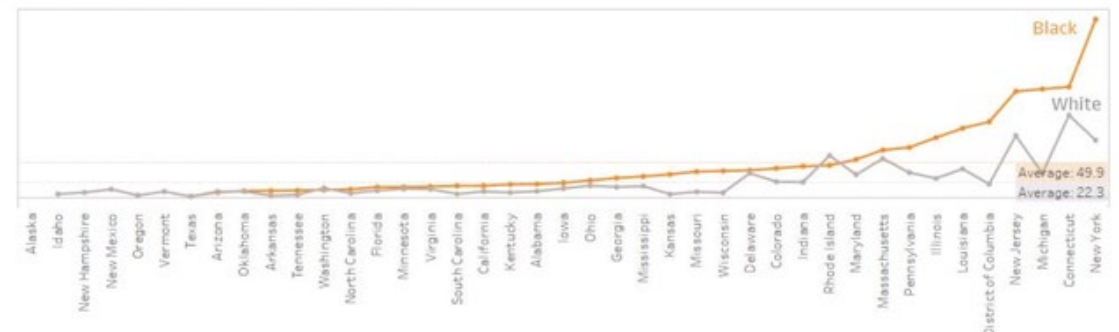
% of people in each country who think the government is handling the issue of coronavirus 'very' or 'somewhat' well



Lake Champlain, NY, water chestnut relative annual control costs, 1982 - 2011

Deaths per 100,000 people

For each 100,000 Americans (of their respective group), about 50 Blacks have died from COVID-19. Black's COVID-19 mortality rate is about 2.24 times of White's. New York has the highest Black COVID-19 mortality rate.



Ahora vamos a R + RStudio.....

- Descargar nuestros archivos de trabajo desde esta dirección de GitHub: <https://github.com/AmadeoGC/Taller-y-cursos-de-R>
- Descomprimir y guardar en el escritorio
- Abrir el script “Manipulación y visualización de datos con R y Tidyverse (Parte 1)”
 - Como probablemente es la primera vez que tienes que abrir este tipo de archivos la mejor opción es:
 - 1) Click con el botón derecho del mouse sobre el archivo .R
 - 2) Seleccionar Abrir con
 - 3) Seleccionar RStudio (y marcar el casillero de usar siempre esta aplicación para los archivos .R) y aceptar.

