

R for Data Science

Fundamentos y Aplicaciones

Guillermo de Anda-Jáuregui y Cristóbal Fresno

Instituto Nacional de Medicina Genómica

2019

¿Qué es la Ciencia de Datos?

Ciencia de Datos

- Campo multidisciplinario
- Extraer DATOS ÚTILES de conjuntos de datos: estructurados y no estructurados
- Usando método científico, estadística, algoritmos y métodos computacionales
- Transformar los datos en CONOCIMIENTO a partir de las diferentes herramientas utilizadas.

]

¿Qué aplicaciones tiene?

¿Qué aplicaciones tiene?

-> Explorar

-> Clasificar

-> Predecir

-> Entender

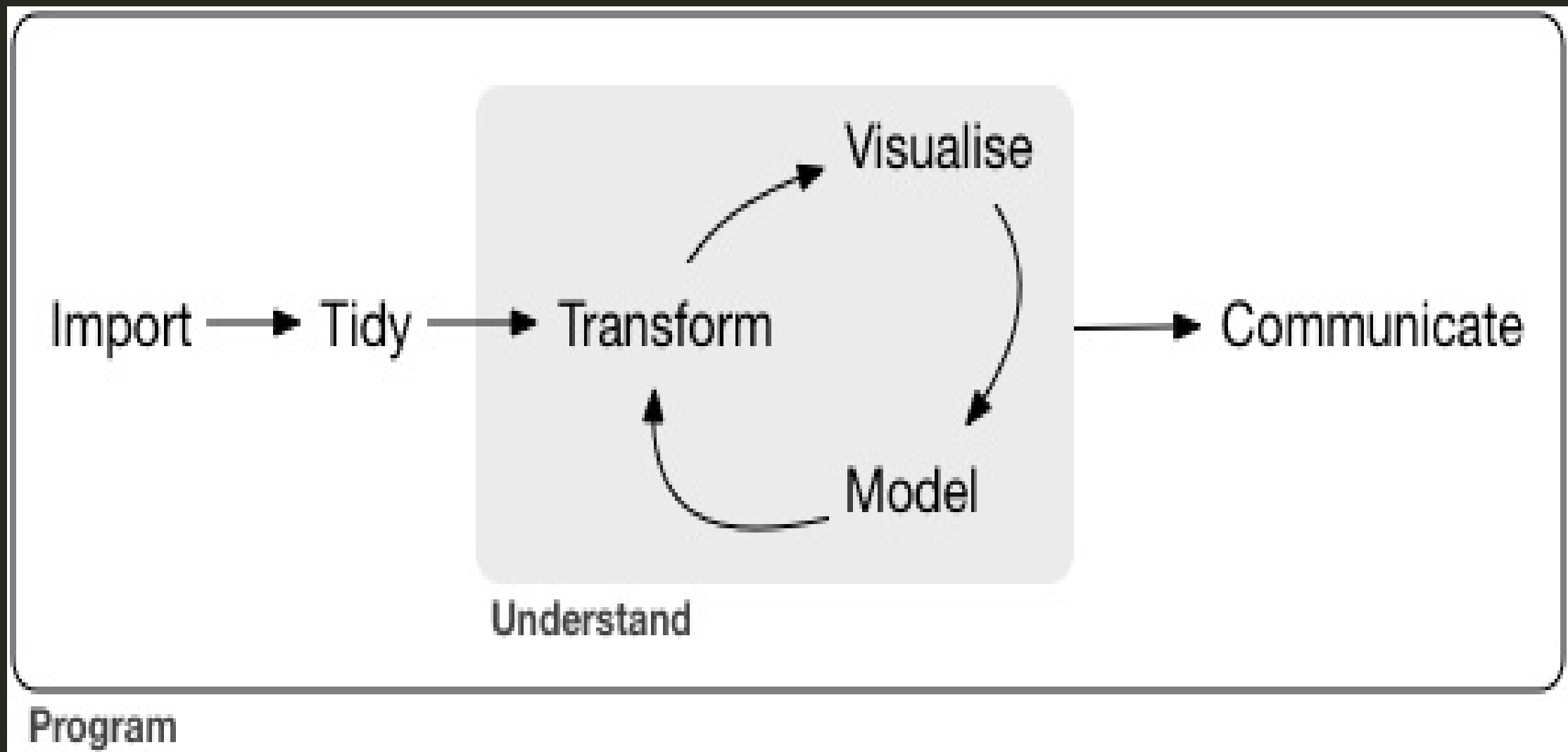
-> Tomar decisiones

¿Qué herramientas existen para hacer ciencia de datos?

- Paquetes cerrados
- Python

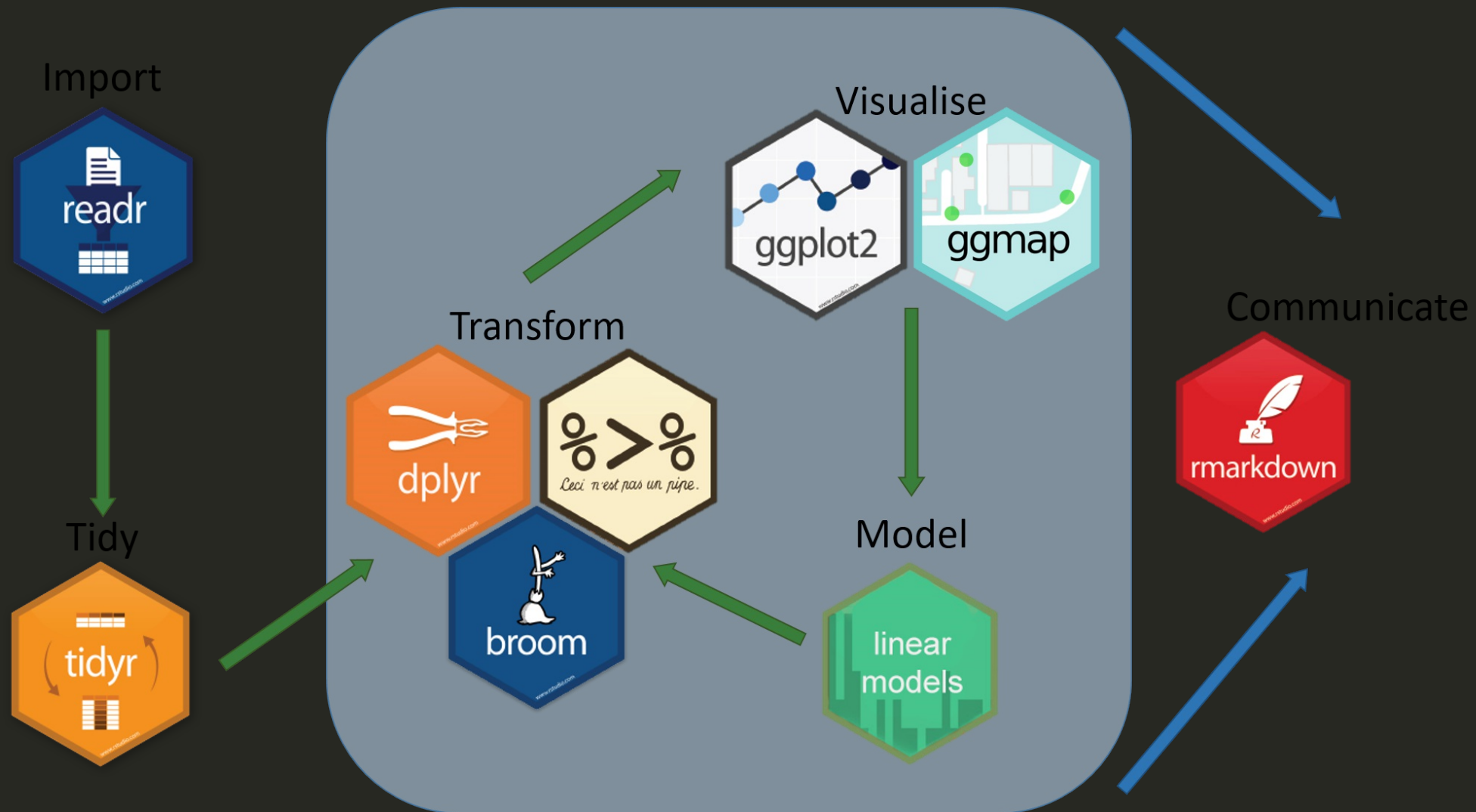


El paradigma Tidy



tomado de <https://r4ds.had.co.nz/introduction.html>

El paradigma Tidy



tomado de <https://ourcodingclub.github.io>

Alcances del curso

- Datos tabulares

-Censos, encuestas, registros, experimentos masivos...

```
iris %>% head
```

| ## | | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|----|---|--------------|-------------|--------------|-------------|---------|
| ## | 1 | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| ## | 2 | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| ## | 3 | 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| ## | 4 | 4.6 | 3.1 | 1.5 | 0.2 | setosa |
| ## | 5 | 5.0 | 3.6 | 1.4 | 0.2 | setosa |
| ## | 6 | 5.4 | 3.9 | 1.7 | 0.4 | setosa |

...

- No datos ...

- Masivos
- Audio, imagen, texto natural ...
- Modelado

Lectura y escritura



Limpieza



Transformar



Visualizar



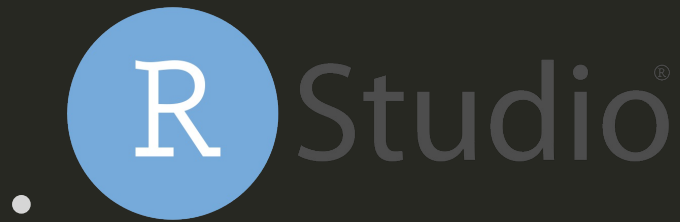
Reportar



Ecosistema de paquetes de R

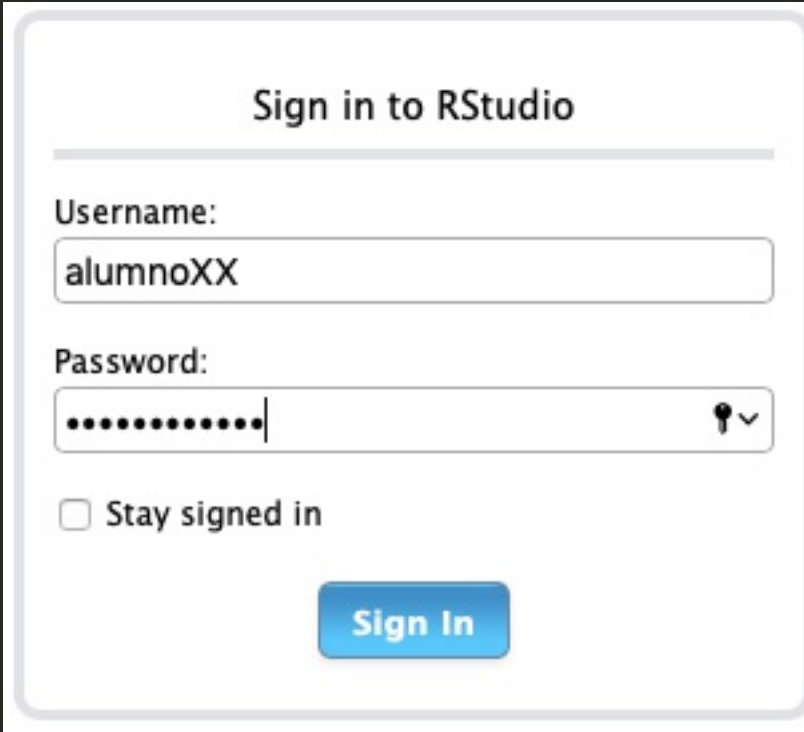


Herramientas



Recursos

- RStudio Server: <http://cluster.inmegen.gob.mx:8787>
- Github: <https://github.com/inmegen>
- Sistema de administración escolar (SAE) INMEGEN



Sign in to RStudio

Username:
alumnoXX

Password:
.....

☐ Stay signed in

Sign In

Estructura del curso

Estructura *espiral*

- 1 Conceptos y demostración de análisis
- 2 Ejercicio de análisis I (horario de clase)
- 3 Ejercicio de análisis II (horario de clase)
- 4 Intervalo de coffe breaks
- 5 Ejercicio de análisis III (horario extra clase)

Contacto

- gdeanda@inmegen.edu.mx
- cfresno@inmegen.gob.mx
- @gdeandajauregui
- @FresnoCristobal

Github

- github.com/guillermodeandajauregui
- github.com/kachelo