

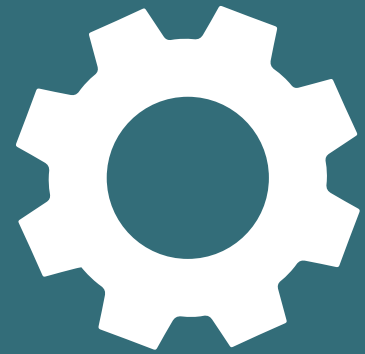
# Entornos de trabajo

JAVIER CÓZAR

# Objetivos



Identificar el rol y las **necesidades tecnológicas** de los Data Scientists



Familiarizarnos con los **entornos de trabajo**



Conocer las principales características de un **lenguaje de programación**

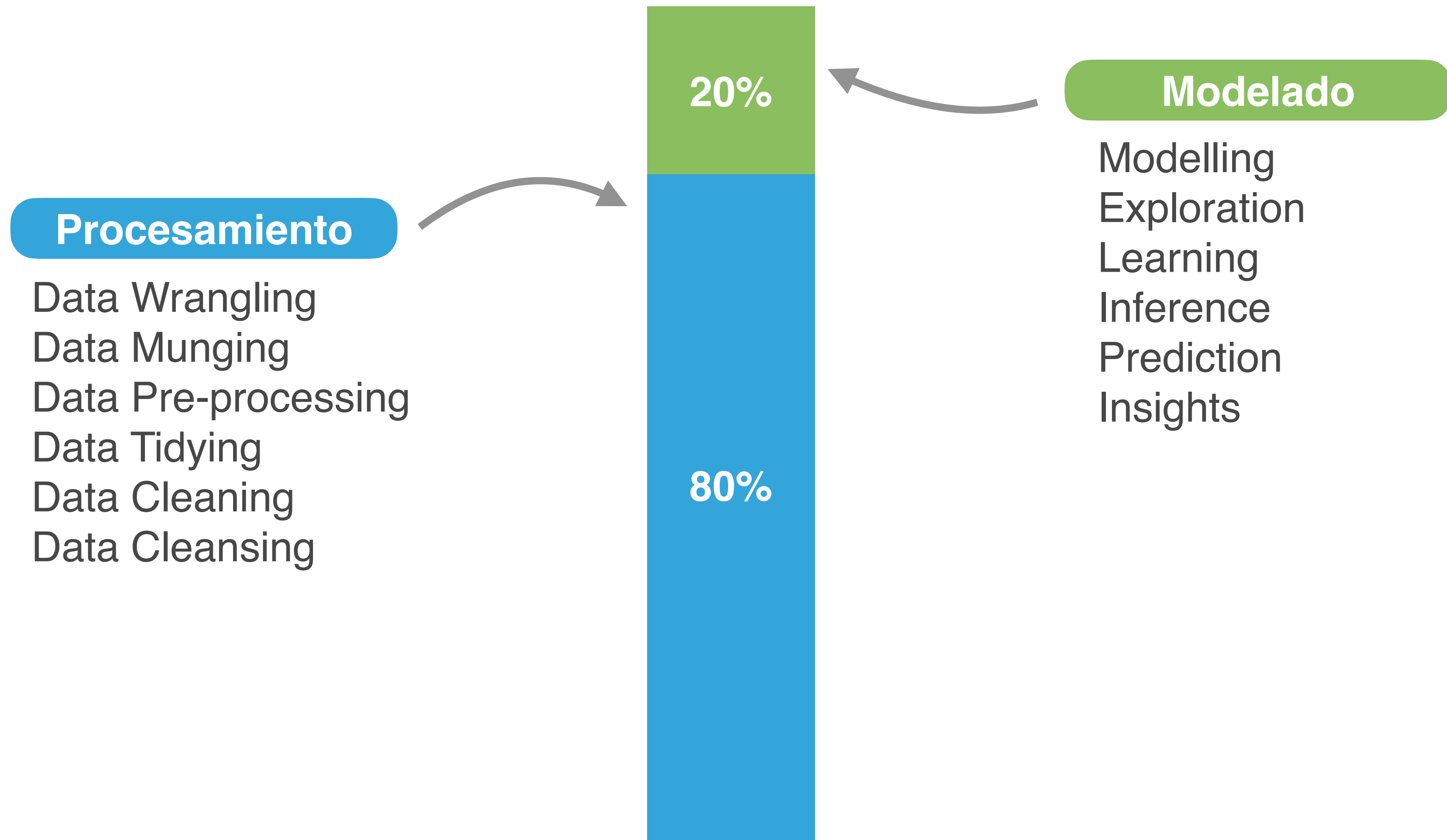
# Índice

1. Introducción
2. Ecosistema Data Science
3. Entornos de desarrollo
4. Introducción a programación

# Índice

## 1. Introducción

# ¿A qué se dedica un Data Scientist?



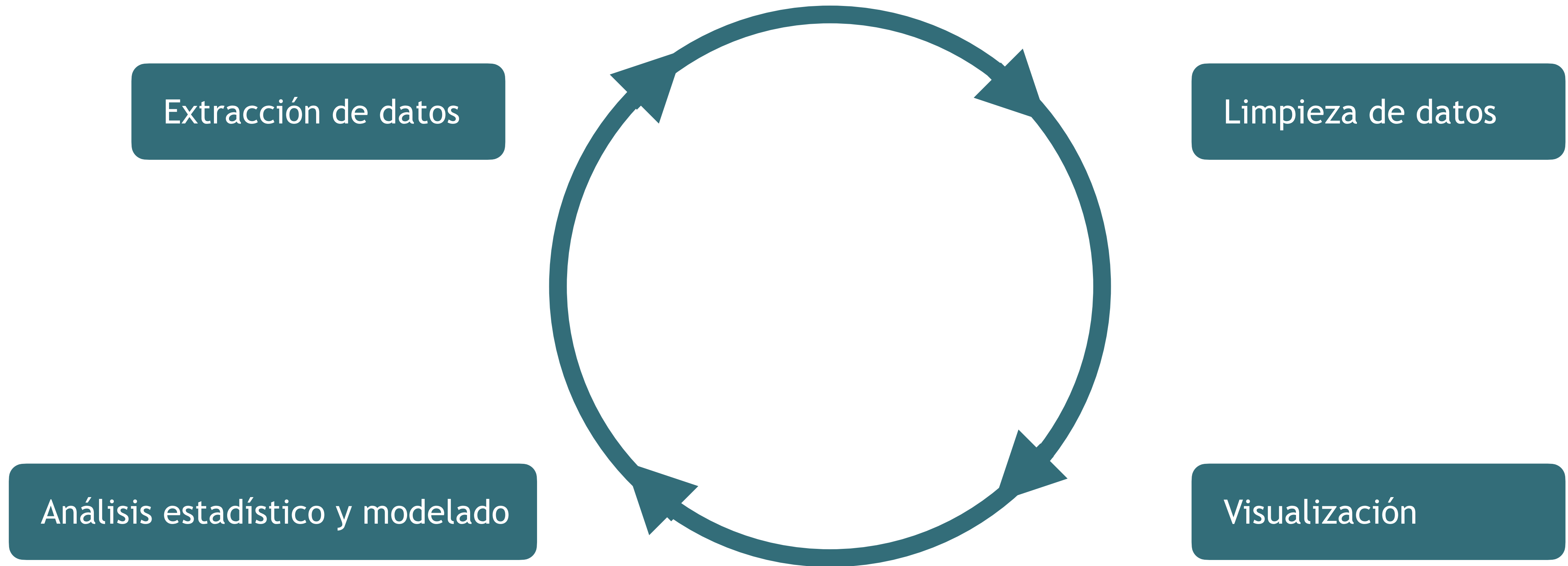
# Skills necesarias



# Índice

1. Introducción
2. Ecosistema Data Science

# ¿Qué necesita un Data Scientist?





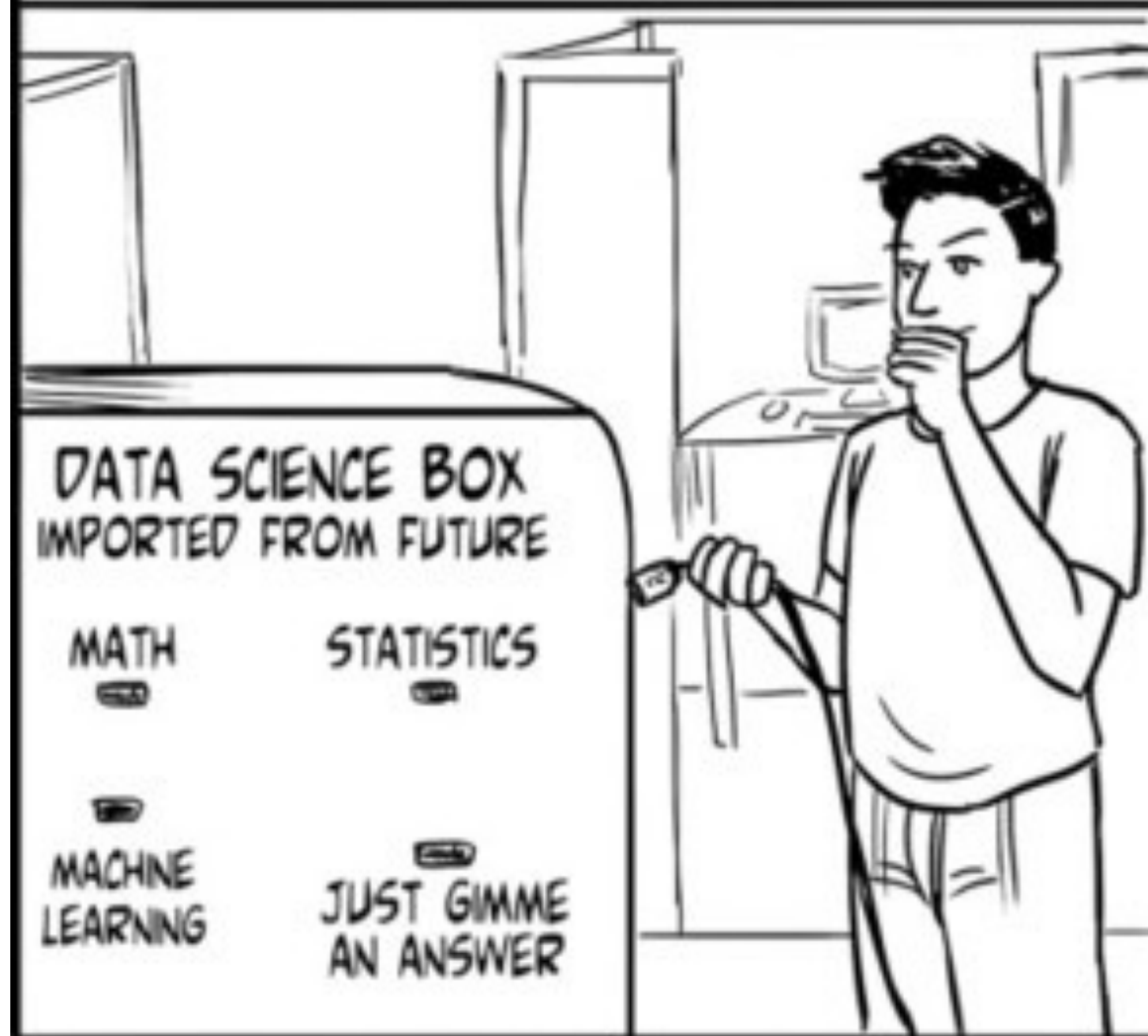
WHAT MY BOSS THINKS DATA SCIENCE IS



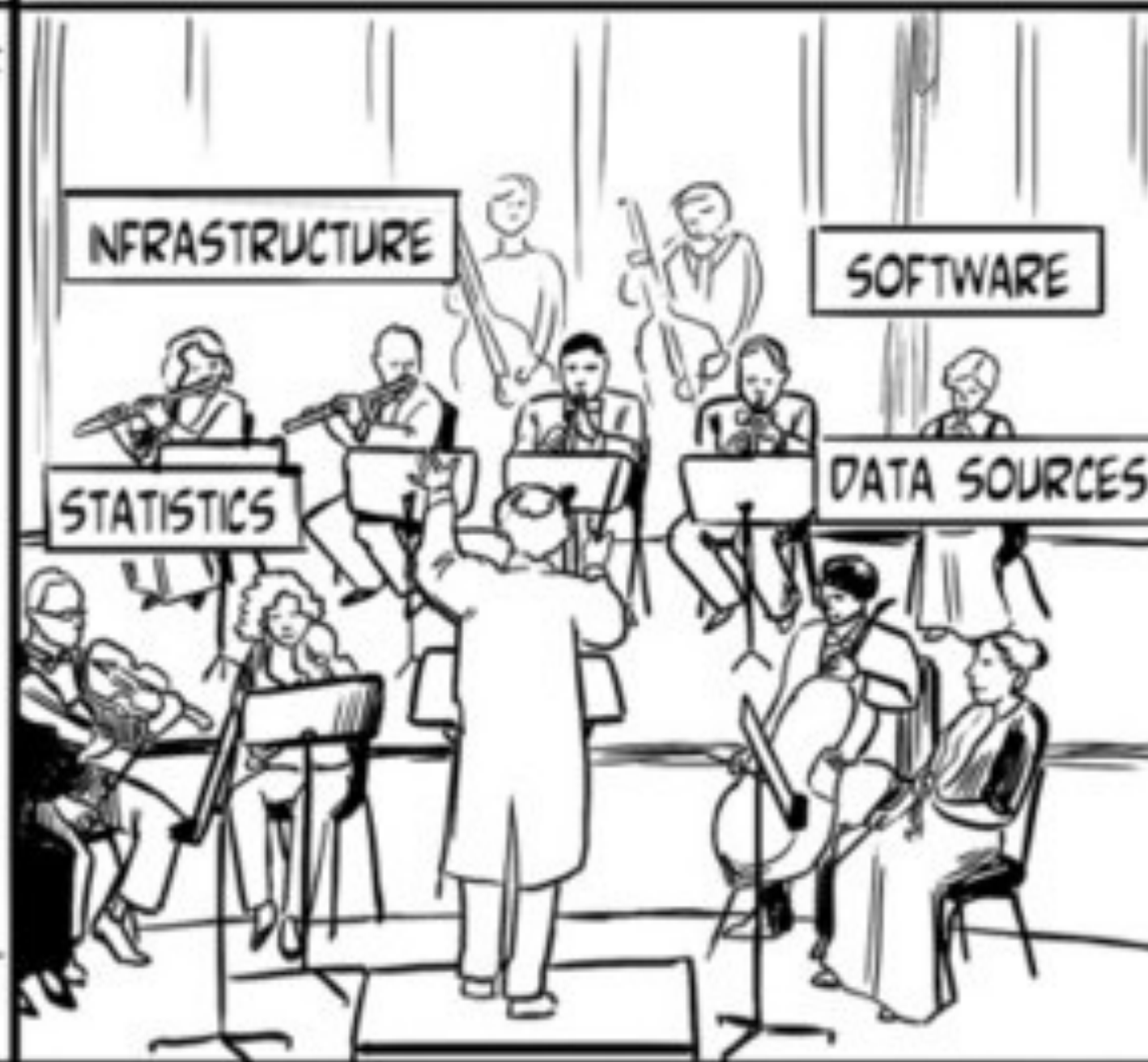
WHAT MY CUSTOMERS THINK DATA SCIENCE IS



WHAT SOFTWARE ENGINEERS THINK DATA SCIENCE IS



WHAT I THINK DATA SCIENCE IS



# ¿Qué necesita un Data Scientist?

- Un **lenguaje de programación** de propósito general
- Flexibilidad y programación **interactiva**
- Herramientas de **reporting**
- **Potencia** de cálculo
- Amplia **comunidad** software

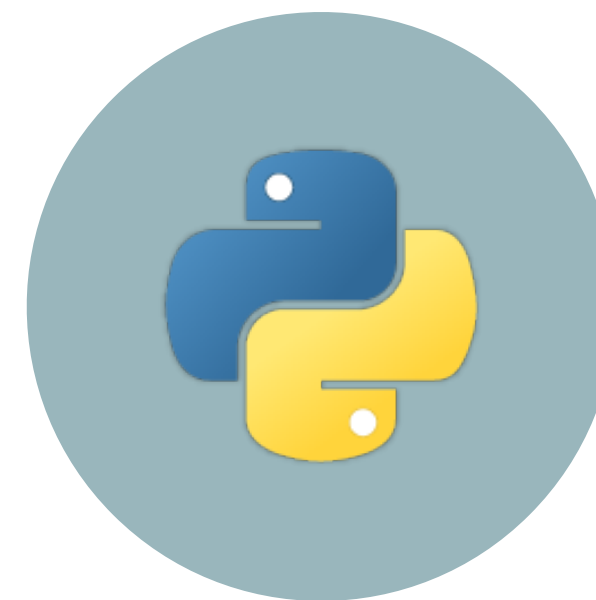
# Lenguajes de programación

- Para un **desarrollador**, la elección de un lenguaje de programación es similar a la elección de una herramienta para realizar un trabajo físico
- Es habitual que sea **políglota** y pueda elegir el lenguaje más adecuado para cada proyecto






















# Lenguajes de programación

- En Data Science existen requisitos comunes para un lenguaje de programación
  - Habilidad de representar y manipular **estructuras de datos**
  - Fácil introducción al lenguaje (profesionales **multidisciplinaes**)
  - **Interactivos** y que faciliten el **trabajo en equipo**





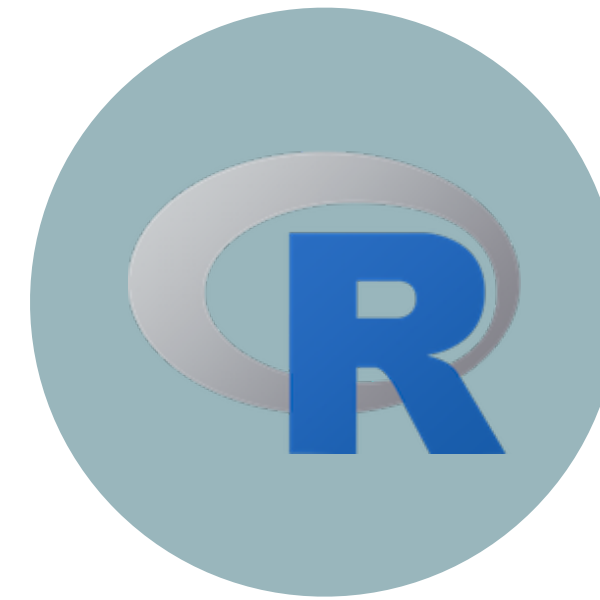
Rank	Language	Type	Score
1	Python ▼	  	100.0
2	Java ▼	  	95.3
3	C ▼	  	94.6
4	C++ ▼	  	87.0
5	JavaScript ▼		79.5
6	R ▼		78.6
7	Arduino ▼		73.2
8	Go ▼	 	73.1
9	Swift ▼	 	70.5
10	Matlab ▼		68.4

[Fuente: IEEE Spectrum](#)

# ¿Qué lenguaje elegir?



- Lenguaje de propósito general
- Ampliamente usado en producción
- Stack potente en DataScience y Machine Learning



- Muy extendido en el ámbito académico
- Lenguaje específico para análisis de datos y estadística
- RStudio es muy sencillo de utilizar



# Índice

1. Introducción
2. Ecosistema Data Science
3. Entornos de desarrollo

# Herramienta de desarrollo

- Potenciar la **experimentación** y la iteración
- Funcionalidad para generar **reportes**
- Que permita la **reproducibilidad** de los experimentos





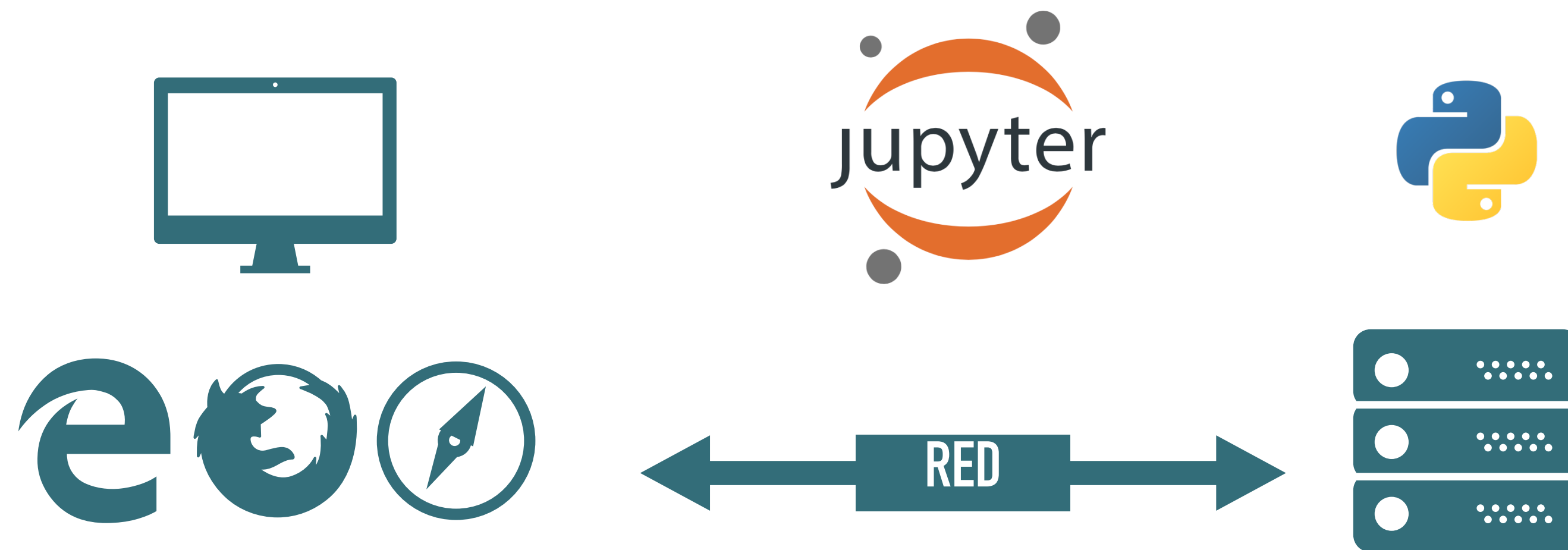
# Herramienta de desarrollo

- [Software gratuito](#), licencia GPL. Versión server de pago.
- Centrado en el trabajo con los datos en R
  - Vistas específicas para tablas y gráficos
- Dispone de una versión online ([RStudio cloud](#))
  - Límite de 15 horas de uso al mes



# Herramienta de desarrollo

- Entorno de ejecución interactivo y visual
- Permite mezclar código, documentación y gráficas
- Herramienta por excelencia en **Data Science**
- Se instala como un paquete de python (pero se puede configurar para usar R)



# Instalación

- [Anaconda](#) es un producto que permite configurar y acceder directamente a un stack preconfigurado que contiene las herramientas más comunes
- Permite el acceso de manera segura a paquetes y librerías
- Optimizado para diferentes arquitecturas
- Ofrecen productos de pago que lo extienden para gestionar proyectos de mayor envergadura



# Instalación

- [RStudio Desktop](#) es un software de escritorio que permite programar en R con facilidad
- Ofrece una consola, una ventana de edición, y una visión de salidas y resultados
- Permite gestionar las dependencias y las actualizaciones mediante un menú
- Instala R y otras dependencias junto con el IDE



# Instalación

- Es necesario instalar el **IDE** junto con el **lenguaje** de programación y otras **dependencias**
  - Versiones
  - Dependencias
  - Actualizaciones
- Existen opciones **online** tanto para trabajar en Jupyter Notebook como en RStudio
- Nos despreocupamos de la gestión, nos centramos en el **análisis de datos**

# Google Colaboratory

- [Google Colab](#) es un servicio online basado en jupyter para **python**
- Utiliza **Google Drive** para almacenar las libretas donde trabajamos
- Totalmente administrado: ¡no hay que configurar ni instalar nada!
- Es **gratuito**, pero no permite su uso para computación intensiva
  - Hasta 12h seguidas, y cierres de sesión tras inactividad
  - Existe versión de pago con menos restricciones y más recursos



# RStudio Cloud

- [RStudio Cloud](#) es un servicio que ofrece **RStudio** por navegador web
- Se divide en proyectos
- En la versión gratuita hay un límite de horas de uso por mes
  - 15 horas usando 1GB de RAM y 1CPU
- Existe opción de pago para un uso profesional



# Índice

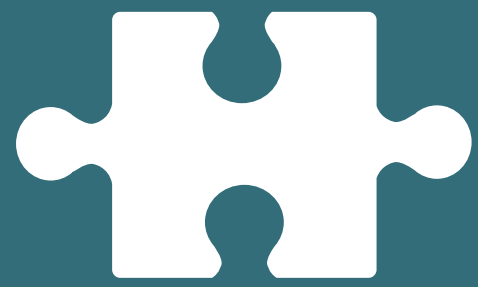
1. Introducción
2. Ecosistema Data Science
3. Entornos de desarrollo
4. Introducción a programación



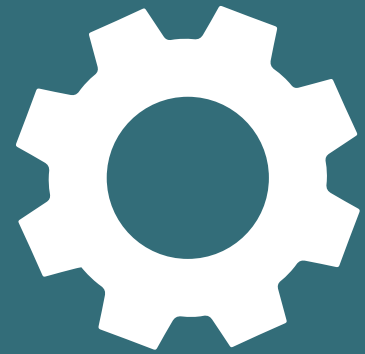
# Introducción a la programación

- Permite expresar una tarea mediante un lenguaje que entiende un computador
  - Leer un fichero, enviar un correo electrónico, analizar un historial de compras, etc.
- Todos los lenguajes de programación poseen
  - **Variables:** hace referencia a un valor, similar a una variable matemática
  - **Expresiones:** operaciones aritméticas (entre números) y con secuencias de caracteres
  - **Estructuras de control:** condiciones y bucles
  - **Funciones:** bloques de código con una funcionalidad determinada

# Conclusiones



Hemos identificado el papel que juega un Data Scientist y las **necesidades tecnológicas** que demandan



Hemos trabajado con el **entorno de trabajo** Google Colab



Hemos aprendido los conceptos básicos de un lenguaje de programación en **Python**

# ¡Aprender a programar es una carrera de fondo!

- Al igual que cuando aprendemos otro idioma, es importante practicar
- Tener motivación y curiosidad hace que aprendamos mejor y más rápido
- Preguntar y aprender de los demás
  - [stackoverflow.com](https://stackoverflow.com), [reddit.com](https://reddit.com), [kaggle.com](https://kaggle.com)
- ¡Estar al día!
  - Documentación oficial, blogs, cursos y tutoriales, ...

# Entornos de trabajo

JAVIER CÓZAR