

Introducción a Python

Sebastian Arpón, PhD
Físico y Data Scientist
MetricArts



METRICARTS



github.com/sarpon

¿Qué hacen los científicos?

Astronomer



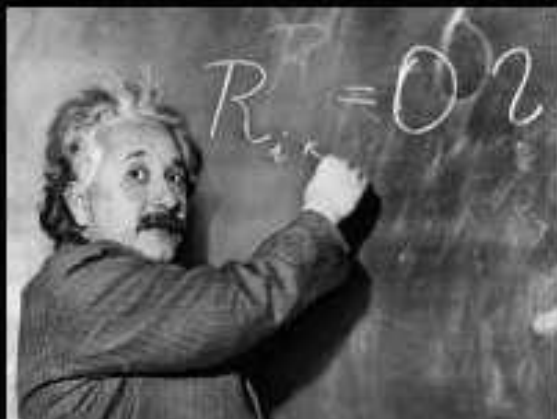
what my friends think
I do



what my parents think
I do



what undergrads think
I do



what I think I do

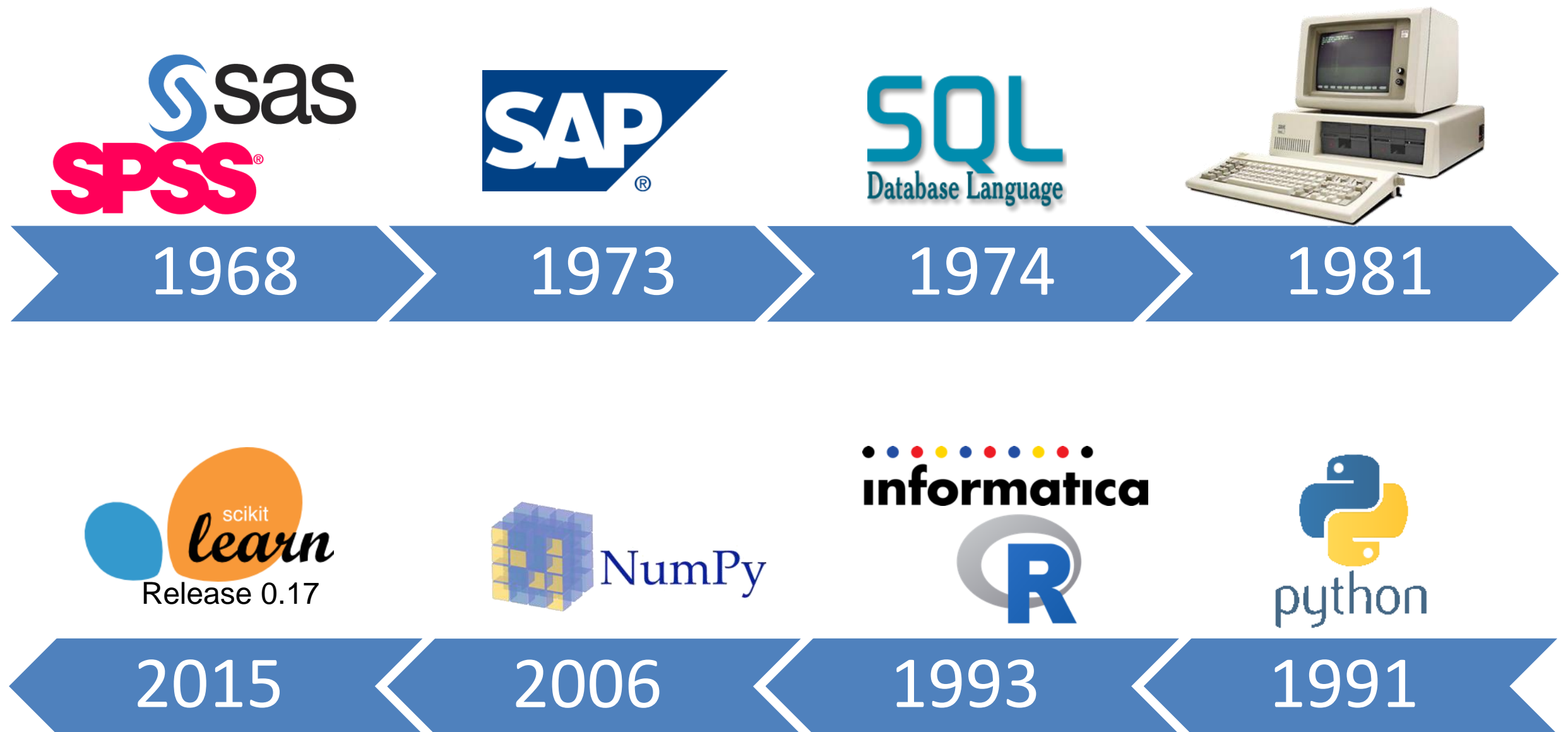


what I wish I did



what I really do

Evolución del Analytics



Lenguajes de programación

- Los lenguajes más usados por la comunidad de Data Science son Python y R. Se estima que Python tiene más de 30M de usuarios y R más de 16M.



- R es más funcional y los módulos de análisis estadístico vienen incorporados.



- Python es más orientado a objetos y deben cargarse módulos para hacer análisis.

R vs Python

R	Python
<ol style="list-style-type: none">1. Big community who creates libraries2. Free3. Early adopter in explanatory and predictive modeling.4. Easy to connect to data sources, including NoSQL and webscraping.	<ol style="list-style-type: none">1. Scalability2. General purpose language3. Easy to learn4. Good in machine learning5. Big community6. Free
<ol style="list-style-type: none">1. Can be slow with big datasets2. Steep learning curve3. No official support4. No user interface	<ol style="list-style-type: none">1. Not as strong in explanatory modeling2. Choice of version: 2.7 or 3.5?3. No user interface4. No official support

¿Qué es Python?

- Python es un lenguaje de programación creado por Guido van Rossum a principios de los años 90 y cuyo nombre está inspirado en el grupo de cómicos ingleses "Monty Python"
- Es un lenguaje similar a Perl, pero con una sintaxis muy limpia y que favorece un código legible.



Características de Python

- Multi propósito
- Multi plataforma
- Lenguaje interpretado
- Orientado a objetos
- Tipado dinámico y tipado fuerte
- Foco en legibilidad y producción



Multi propósito

- Desarrollo web
- Desarrollo de aplicaciones
- Procesamiento y análisis de datos
- Análisis descriptivo
- Construcción de modelos predictivos
- Machine learning y Deep learning

Multi plataforma

- El intérprete de Python está disponible en multitud de plataformas
 - UNIX
 - Linux
 - DOS
 - Windows
 - Mac OS
- Nuestro programa podrá correr en todos estos sistemas sin grandes cambios.

Lenguaje interpretado

- Un lenguaje interpretado o de script es aquel que se ejecuta utilizando un programa intermedio llamado intérprete, en lugar de compilar el código a lenguaje máquina que pueda comprender y ejecutar directamente una computadora.
- Ventaja de los lenguajes compilados es que su ejecución es más rápida. Sin embargo los lenguajes interpretados son más flexibles y portables.

Orientado a objetos

- La orientación a objetos es un paradigma de programación en el que los conceptos del mundo real relevantes para nuestro problema se trasladan a clases y objetos en nuestro programa.
- La ejecución del programa consiste en una serie de interacciones entre los objetos.
- Programación imperativa, funcional y orientada a aspectos.

Tipado dinámico

- La característica de tipado dinámico se refiere a que no es necesario declarar el tipo de dato que va a contener una determinada variable.
- Su tipo se determinará en tiempo de ejecución según el tipo del valor al que se asigne, y el tipo de esta variable puede cambiar si se le asigna un valor de otro tipo.

Tipado fuerte

- Python impone fuertemente las reglas de tipificación, es decir, no se permite tratar a una variable como si fuera de un tipo distinto al que tiene.
- Es necesario convertir de forma explícita dicha variable al nuevo tipo previamente. Por ejemplo, si tenemos una variable que contiene un texto no podremos tratarla como un número (sumar la cadena “9” y el número 8).