

ANÁHUAC



Programación de la Información

1. Introducción

Mtro. Roberto Arturo Morales Cruz



Roberto Arturo Morales Cruz

- Licenciatura en Actuaría, UDLAP
- MS in Analytics, Georgia Tech
- Data Scientist por 6 años.
- Docente por 2 años.



Calendario Tentativo

Lunes	Martes	Miércoles	Jueves	Viernes	Temas
09-ago	10-ago	11-ago	12-ago	13-ago	Introducción
16-ago	17-ago	18-ago	19-ago	20-ago	
23-ago	24-ago	25-ago	26-ago	27-ago	
30-ago	31-ago	01-sep	02-sep	03-sep	Componentes del lenguaje de Programación Python
06-sep	1er Parcial	08-sep	1er Parcial	10-sep	Construcción de Funciones en Python / Estructuras de Datos en Python
13-sep	14-sep	15-sep	Independencia	17-sep	
20-sep	21-sep	22-sep	23-sep	24-sep	
27-sep	28-sep	29-sep	30-sep	01-oct	
04-oct	Medio Término	06-oct	Medio Término	08-oct	Repaso / Examen / Revisión
11-oct	12-oct	13-oct	14-oct	15-oct	Manejo y Visualización de Datos en Python
18-oct	19-oct	20-oct	21-oct	22-oct	
25-oct	26-oct	27-oct	28-oct	29-oct	
01-nov	Día de Muertos	03-nov	3er Parcial	05-nov	Repaso / Examen / Revisión
08-nov	Acciona Negocios	10-nov	3er Parcial	12-nov	
15-nov	16-nov	17-nov	18-nov	19-nov	
22-nov	23-nov	24-nov	25-nov	26-nov	
Exámenes Finales					

Objetivos de esta sesión

- Entender cómo funciona y qué actores participan en el ecosistema actual de programación de la información.
- Comprender la importancia de saber programar.
- Conocer los lenguajes de programación más utilizados en la actualidad para el análisis de datos.

El Choro

- Hay más datos que nunca: Se estima que el 90% de los datos en el mundo se creados en los últimos dos años.
- Ciencia de Datos: “Profesión más sexy del siglo XXI”.
- Para 2022 habrán 700,000+ de vacantes para personas con un rol en el análisis de datos.

La Realidad

- Varias profesiones “tradicionales” están reemplazadas parcialmente por Computadoras.
- ¿Quiénes están a salvo?
 1. Creativos
 2. Profesiones que involucren Trato Humano
 3. Quienes trabajen en conjunto con IA
(Humano + Computadora > Computadora)
 4. Quienes le enseñen a las máquinas.

¿Qué es la “Ciencia de Datos”?

El uso de técnicas avanzadas de Matemáticas, Estadísticas y de cómputo para obtener **respuestas** a problemas actuales.

Parecido a lo que hace un Economista, ¿no?

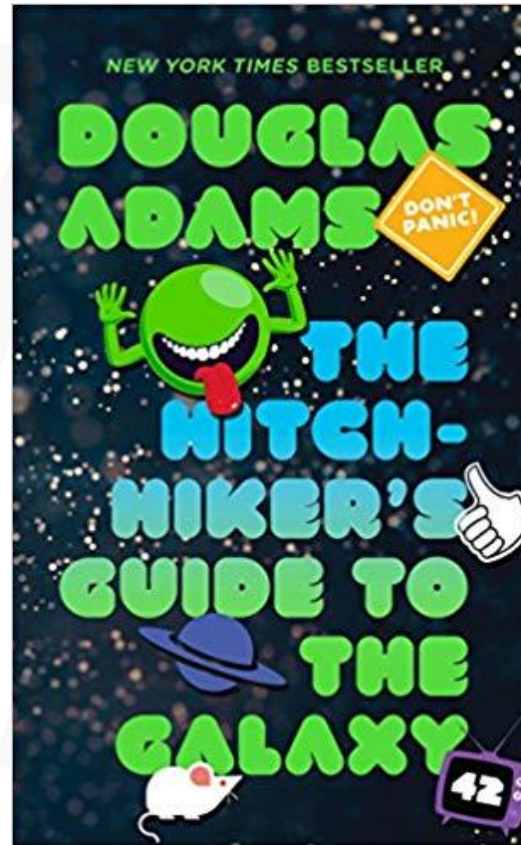
¿Qué es la “Ciencia de Datos”?

- X** No sustituye a la Experiencia.
- ✓ La complementa, para poder tomar decisiones basadas en datos.
- X** No da soluciones mágicas; necesita tener armonía entre el problema a resolver y la parte técnica.

¿Qué es la “Ciencia de Datos”?

- Se analizan datos que provienen de fuentes variadas.
- Se trata de encontrar patrones en los datos.
- Se monitorean Key Performance Indexes (KPI) y se analizan tendencias.
- Se comunican los resultados encontrados a gente con perfil no técnico.

¿Por dónde empiezo?



SELECT LEVEL



THE LORD OF THE RINGS



¿Cuál es la Respuesta a la Gran Pregunta del Universo, la Vida y todo lo demás?

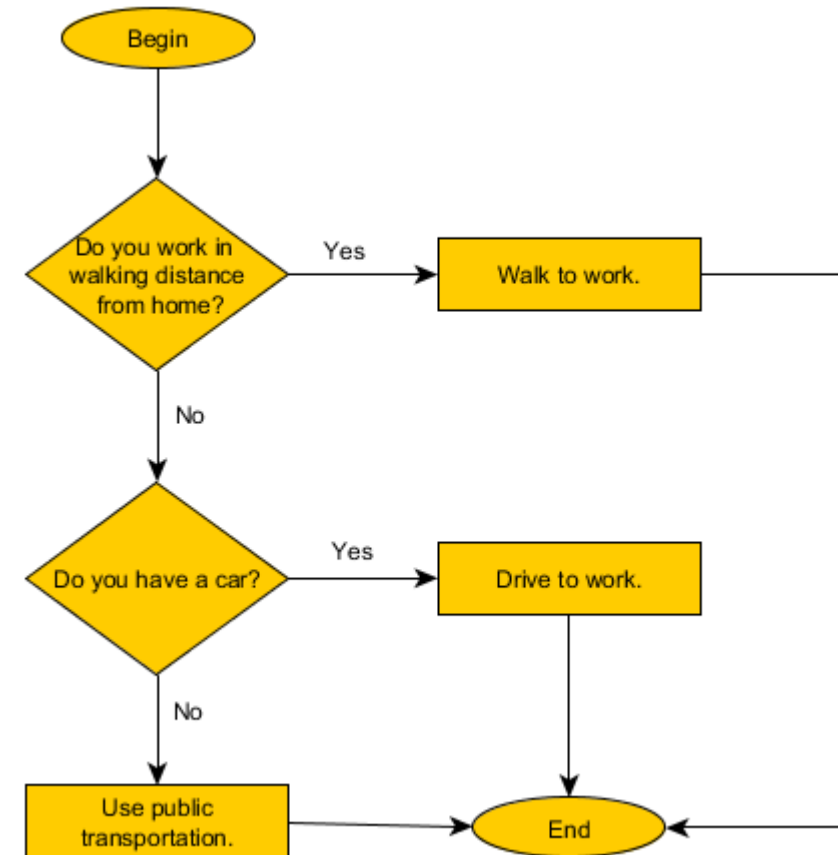


Humano + Computadora > Humano
Humano + Computadora > Computadora

- La computadora, por sí sola, no sabe qué es lo que debe hacer.
- ¡Hay que “traducirle” a la computadora, nuestra problemática.
- Afortunadamente, los Economistas ya tienen experiencia ligando problemas sociales, con planteamientos matemáticos.

¿Qué es un algoritmo?

- Una serie de pasos que se llevan a cabo para llevar a cabo una tarea.
- Tienen un principio y un fin claros.
- El ejemplo “típico”: una receta de cocina:
- ¡Pero cualquier serie de pasos, con un propósito, principio y fin, puede ser un algoritmo!



¿Qué es un algoritmo?

Ejercicio (10 minutos):

- Escriban, en el formato y complejidad que escojan, un algoritmo para:
 1. Viajar de su casa, a la Universidad.
 2. Tener un entrenamiento de su deporte favorito.
 3. Sacar dinero del cajero.

¿Qué es un programa?

- Una serie de instrucciones que sigue la computadora.
- Los ejemplos “típicos” son las distintas aplicaciones de nuestros equipos electrónicos: procesadores de texto, como Microsoft Word; las hojas de cálculo, como Excel; mensajería instantánea, como WhatsApp.
- Para que funcionen, antes alguien tuvo que “enseñarle” a la computadora qué imágenes mostrar, qué hacer cuando el usuario diera clic, etc.
- ¡Pero esos son sólo casos particulares!

¿Qué es un programa?

- Todos los programas son algoritmos, pero no todos los algoritmos son programas.
- ¿Qué marca la diferencia?

El programa ya puede ser entendido y ejecutado por la computadora.

¿Qué es un programa?

- Word, WhatsApp, Chrome, son programas, puesto que la computadora sigue instrucciones complejas para abrirlos, y luego para modificarse.
- A ese modelo se le llama *Programación Orientada a Objetos*, puesto que los programadores se enfocan en productos en específico.

¿Qué es un programa?

- En esta clase, **NO** buscamos aprender eso, sino potencializar su trabajo como Economistas.
- Aquí aprenderemos la *Programación Funcional*; es decir, usaremos computadoras para facilitarnos la vida al realizar cálculos matemáticos de todo tipo.

¿Qué es un Lenguaje de Programación?

- Para poder traducir de un algoritmo a un programa, necesito ponerlo en términos que la computadora me pueda entender. Es decir, necesito hablar en su *lenguaje*.



Varios de los Distintos Lenguajes de Programación.

¿Qué es un Lenguaje de Programación?

- Así como en el ser Humano tenemos lenguajes más *Románticos*, y otros más *Serios*, así pasa con los lenguajes de programación.
- Todos son útiles, pero dependiendo de lo que quiero hacer, algunos me sirven mejor.

¿Qué es un Lenguaje de Programación?

- La computadora, originalmente, sólo piensa en 0s y 1s.
- Todas las instrucciones que la máquina tiene, se pueden traducir de esta manera.
- Si le hablo a la computadora en esos términos, estaré usando el **Lenguaje de Máquina**.
- Sus programas suelen correr más rápido.

¿Qué es un Lenguaje de Programación?

- También le puedo hablar a través de un lenguaje un poco más complicado, con algunas frases cortas.
- Un programa aparte, llamado *ensamblador*, traduce esas frases a 0s y 1s.
- Un lenguaje de este tipo es un **Lenguaje Ensamblador**.

Assembly Language

```
mov ecx, ebx  
mov esp, edx  
mov edx, r9d  
mov rax, rdx
```

Programmer

Assembler + Linker

Machine Language

```
100101011001  
010011111011  
111010101101  
01010101010
```

Processor

¿Qué es un Lenguaje de Programación?

- Los lenguajes más prácticos, ocupan palabras y frases conocidas (generalmente en inglés). Así, puedo ser más directo con las instrucciones que le doy a la máquina.
- Estos se conocen como **Lenguaje de Alto Nivel**.

index.java

```
1. public class Holamundo {  
2.  
3.     public static void main(String[] args) {  
4.         System.out.println("Hola mundo");  
5.     }  
6. }
```

index.html

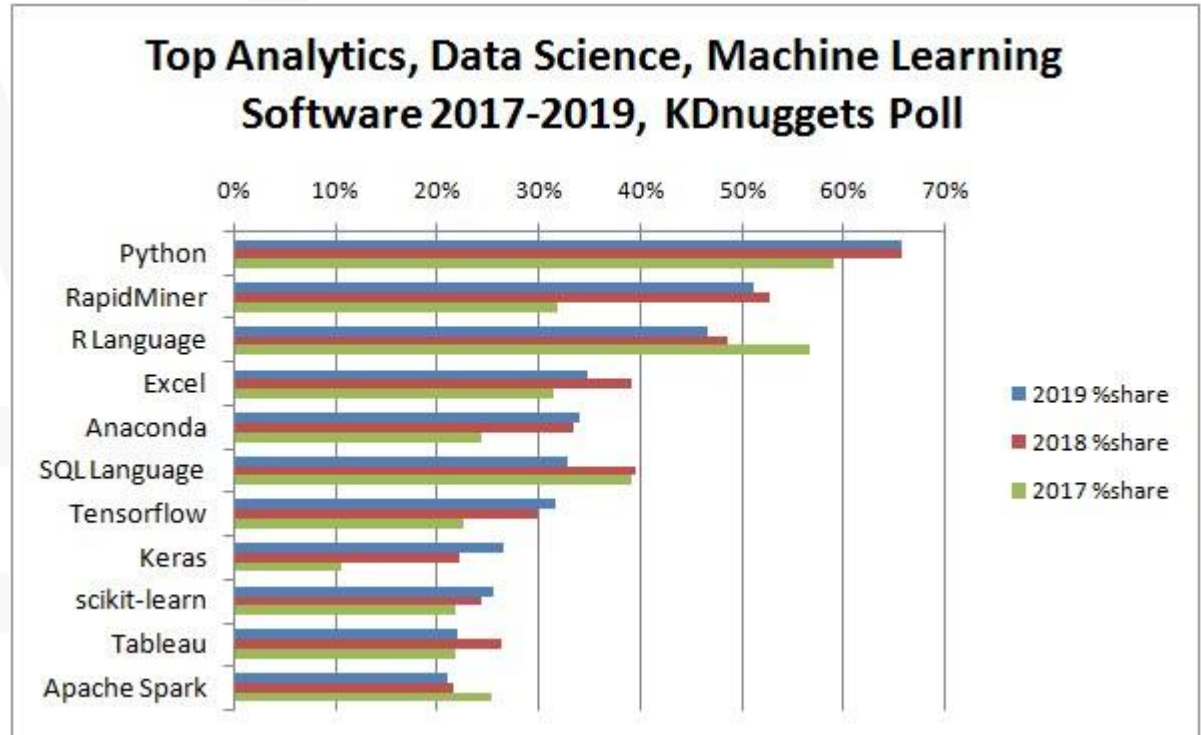
```
1. <!doctype html>  
2. <html>  
3. <head>  
4. <title>Hola mundo</title>  
5. </head>  
6. <body>  
7.     <h1>Hola mundo</h1>  
8. </body>  
9. </html>
```


¿Qué es un Lenguaje de Programación?

- Algunos lenguajes son **compilados**; necesito terminar todo el programa y *compilar*, para poder ver los resultados.
- Otros, son **interpretados**; puedo ir observando mis resultados en cada paso.
- Por el tipo de tareas que realizaremos, ¿qué clases de lenguajes creen que nos sirvan más?

¿Qué Lenguaje debo Ocupar?

- Con tantas herramientas en el mercado, es bastante común confundirse sobre cuál es la que necesito utilizar.



¿Qué Lenguaje debo Ocupar?

- Un proceso típico es conseguir primero el Software, y después ver si es útil llevar a cabo mi tarea.
- ¡Error!
- Primero se deben identificar los problemas que se pretende resolver, y sobre ello, ver qué programas pueden servirme.
- El Programa se adapta al problema. No al revés.

¿Qué Lenguaje debo Ocupar?

Aspectos a considerar:

- **Curva de Aprendizaje:** ¿Qué tan difícil es de aprender?
- **Escalabilidad:** ¿Qué tan fácil será de crecer?
- **Apoyo:**
 - Si es comprado, ¿me darán soporte?
 - Si es libre, ¿hay una comunidad en línea que ayude con dudas?
- **Hardware requerido:** ¿Necesito algún equipo en especial?
- **Velocidad**

¿Qué Lenguaje debo Ocupar?

- Para llevar a cabo las tareas de un Economista, necesito un lenguaje fácil de entender. Por lo tanto, alguno de **alto nivel**.
- Como estaré utilizando operaciones matemáticas, necesito un lenguaje donde pueda estar viendo los resultados en cada paso. Para ello, un lenguaje **interpretado**.
- La velocidad no es tan importante.
- De preferencia, que ya tenga mucho uso, para que me sea más fácil conseguir ayuda y aplicar lo que sé.

R vs Python



Escuela de
Negocios



ANÁHUAC

R y Python

- Programación Funcional
- Software Libre
- Compatibilidad
- Fáciles de usar



R y Python

Amplio soporte web

Capacidades para:

Leer

Manipular

Describir

Modelar

Visualizar

...



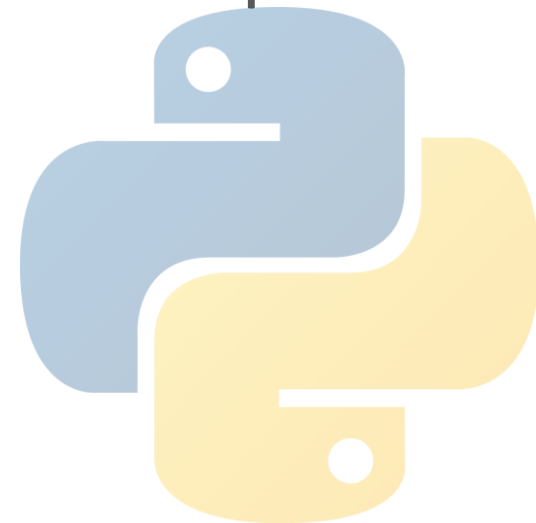
Python

- Creado en 1981.
- Su nombre se debe al programa Monty Python (¡hace siglos...!)
- Lenguaje Interpretado, de alto nivel, y multiparadigma: se puede ocupar orientado a objetos, en programación imperativa y en programación funcional.



Python

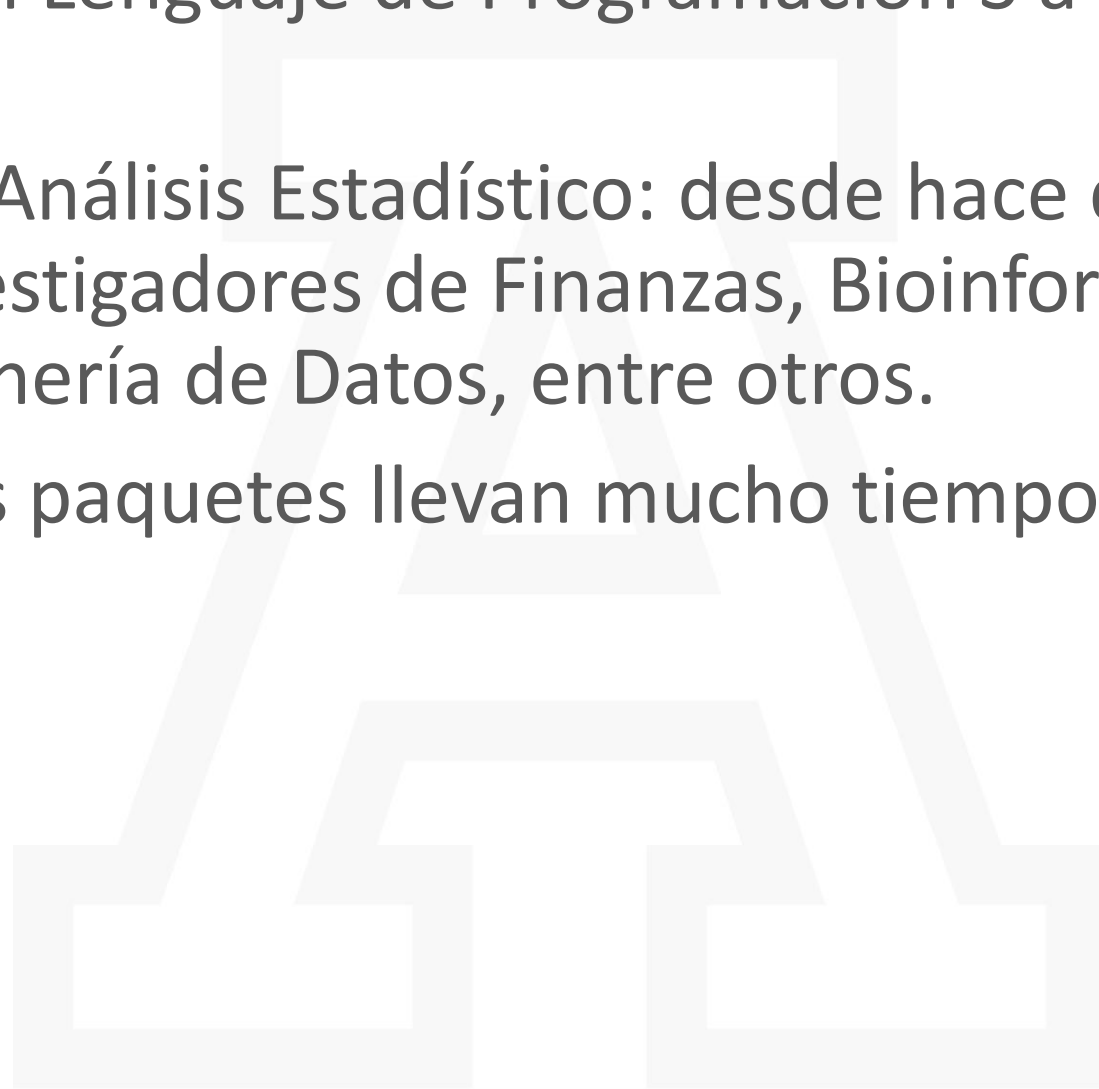
- Sirve para muchas más cosas aparte del análisis de datos. De hecho, esa es de las últimas cosas que pudo hacer...
- Por lo mismo, es de los lenguajes más populares en la actualidad.
- Su estructura es más parecida a un lenguaje de programación “tradicional”.
- Actualmente va en la versión 3.9. Ha cambiado mucho con respecto a versiones anteriores.



ANÁHUAC

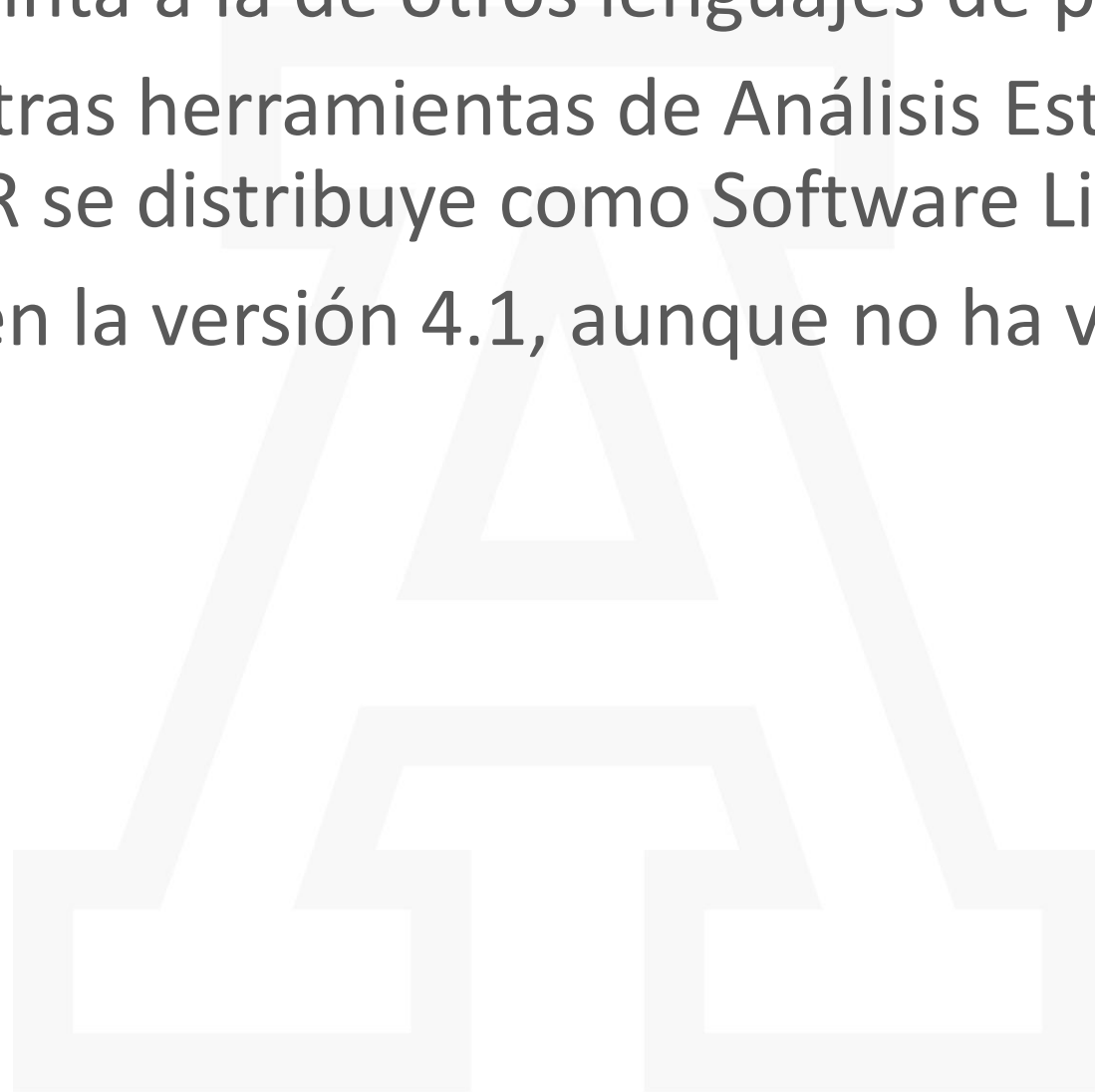
R

- Surgió a partir del Lenguaje de Programación S a finales de los años 70.
- Está enfocado al Análisis Estadístico: desde hace décadas es ocupado por Investigadores de Finanzas, Bioinformática, Biomédica, Matemáticas, Minería de Datos, entre otros.
- Por lo mismo, sus paquetes llevan mucho tiempo siendo probados.



R

- Su sintaxis es distinta a la de otros lenguajes de programación.
- A diferencia de otras herramientas de Análisis Estadístico (SAS, SPSS, Stata, MATLAB), R se distribuye como Software Libre.
- Actualmente va en la versión 4.1, aunque no ha variado tanto con los años.



R vs Python

Ejercicio (10 minutos):

- ¿Cuál es mejor para el análisis estadístico? ¿R o Python?
 - Investigar y resumir algunas ideas.
 - Cada quién debe compartir un *bullet point*.
 - *Ejemplo*: “R es mejor porque tiene un nombre más corto...”

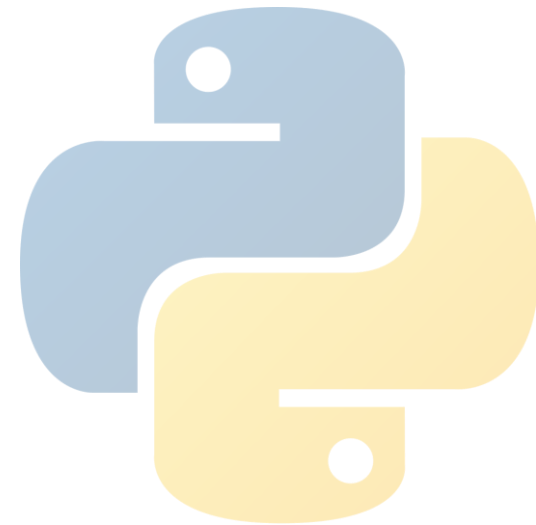
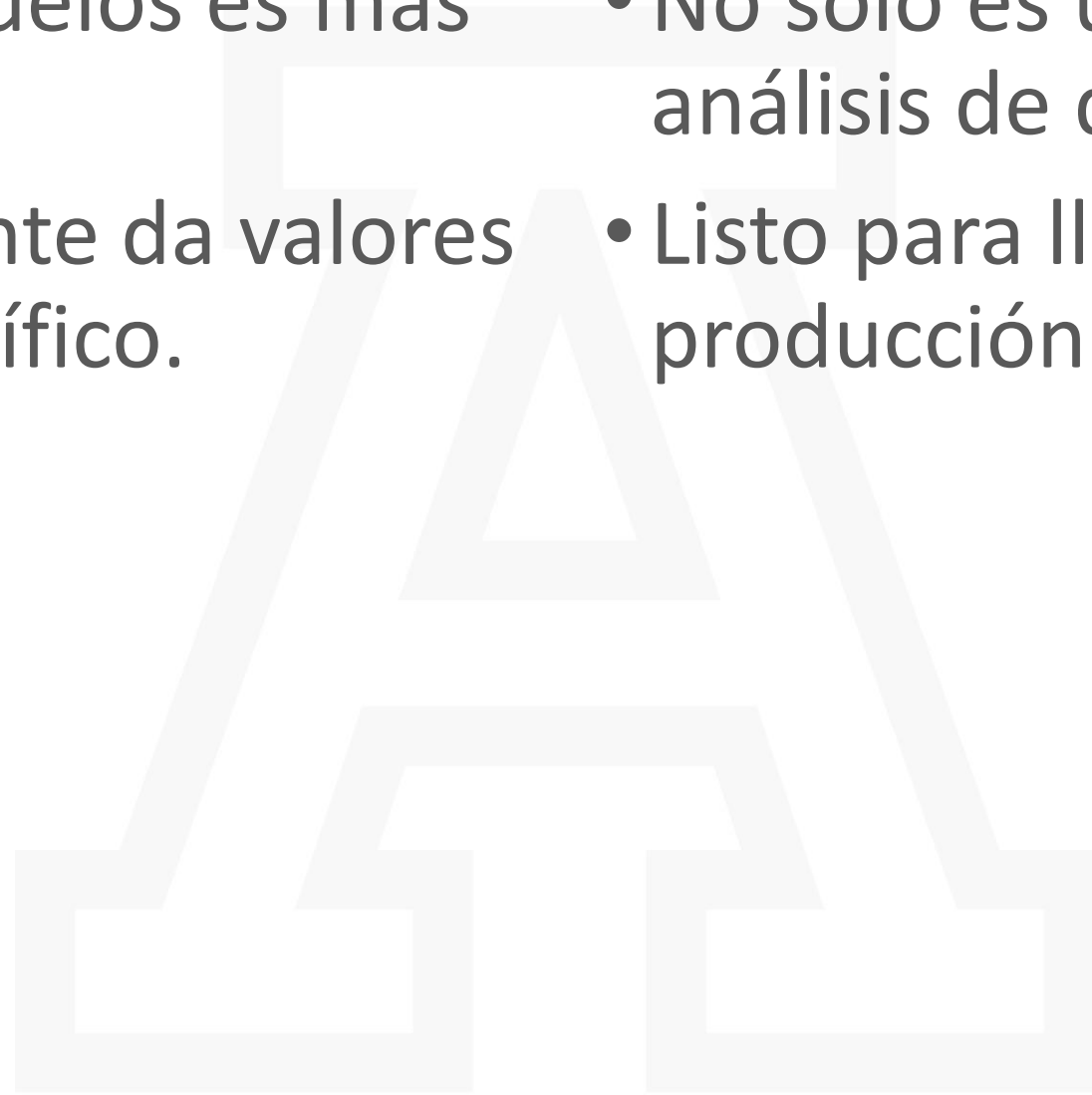
R vs Python

- Desarrollado en la Academia.
- Preferido por estadísticos, y otros científicos.
- Desarrollado pensado en la Industria
- Preferido por quienes tienen antecedentes de computación.



R vs Python

- Desarrollar modelos es más directo.
- Automáticamente da valores de interés científico.
- No sólo es un lenguaje de análisis de datos.
- Listo para llevarse a producción.

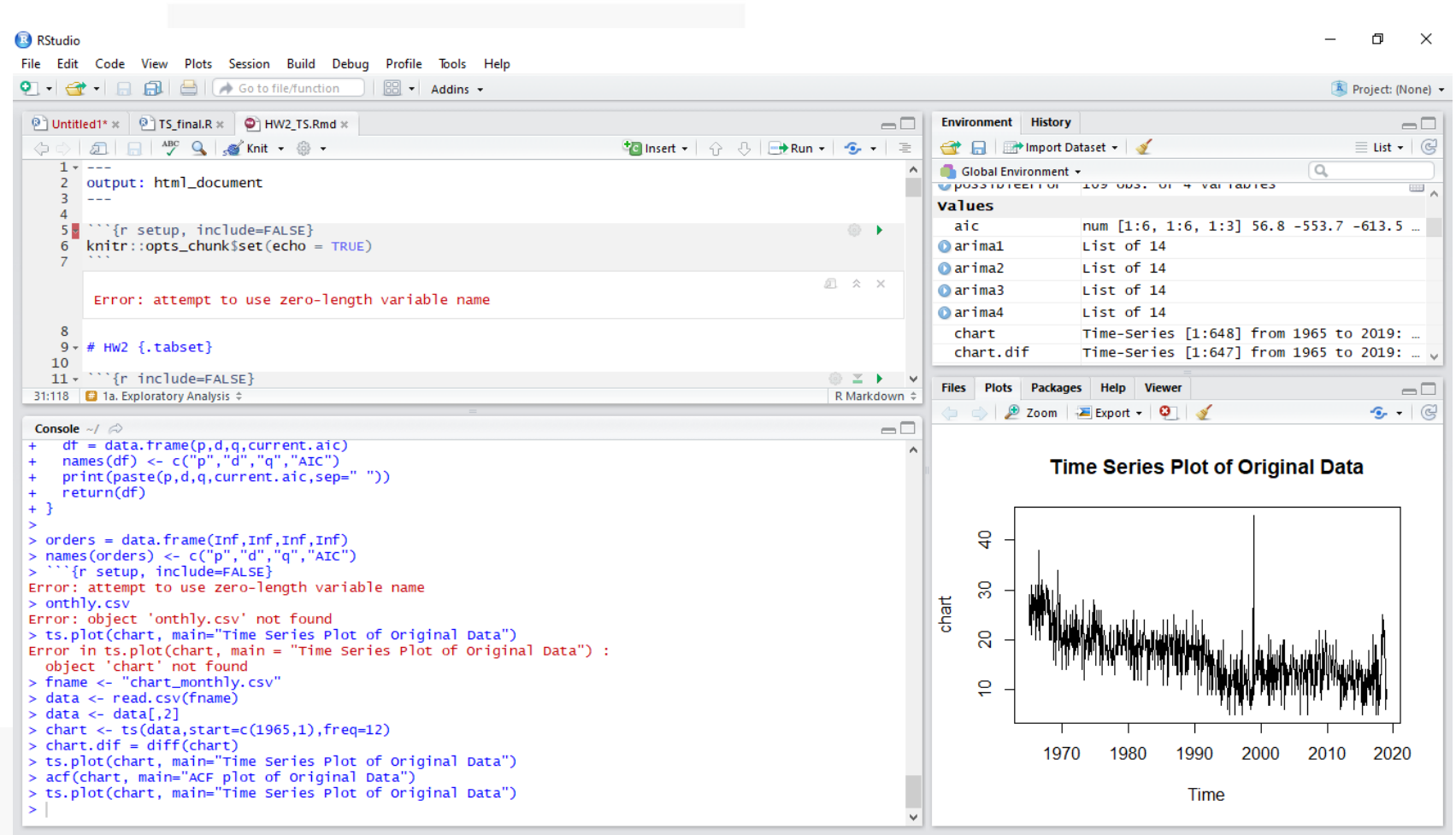


Trabajos

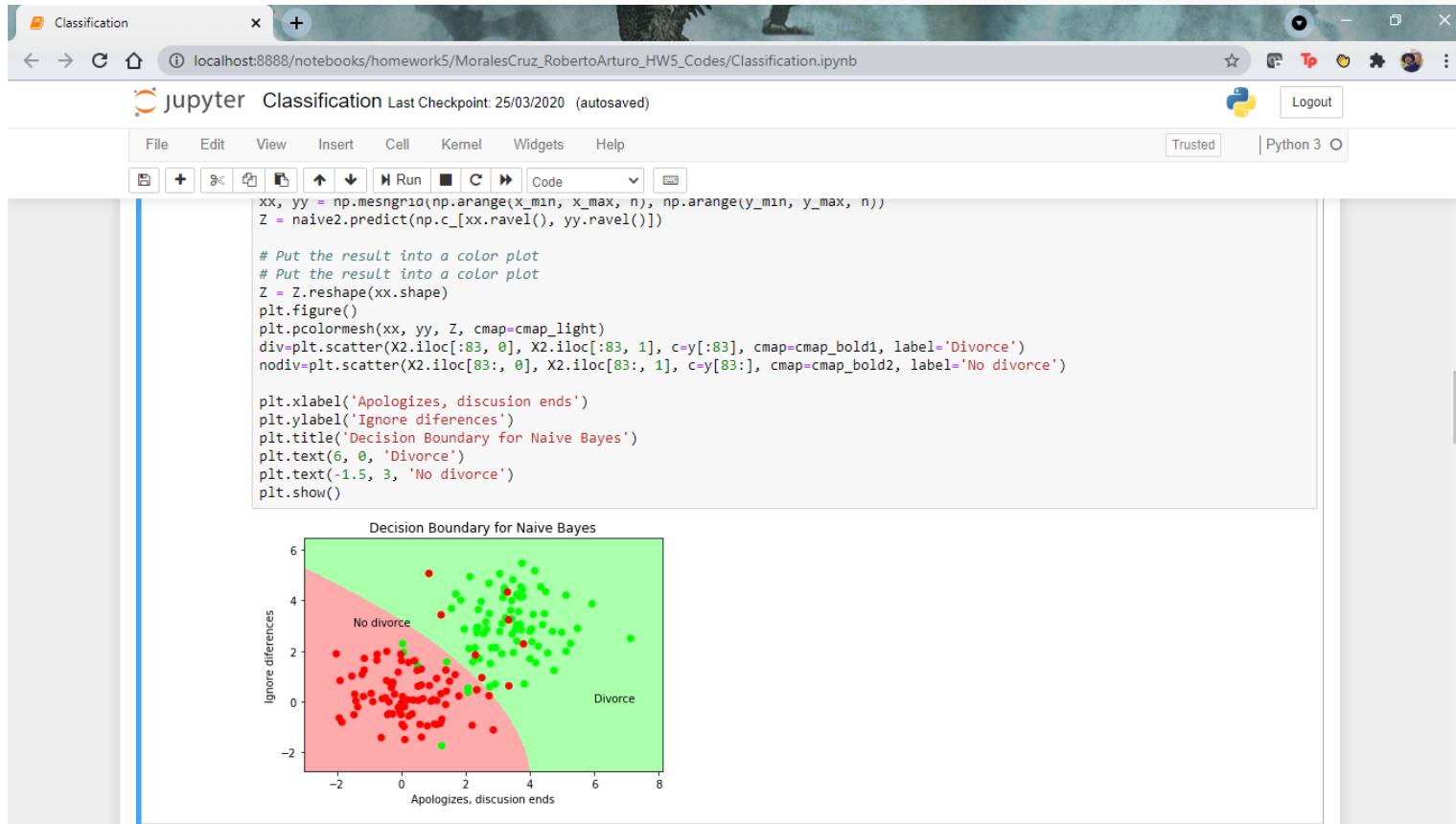


Integrated Development Environment

- RStudio



Integrated Development Environment

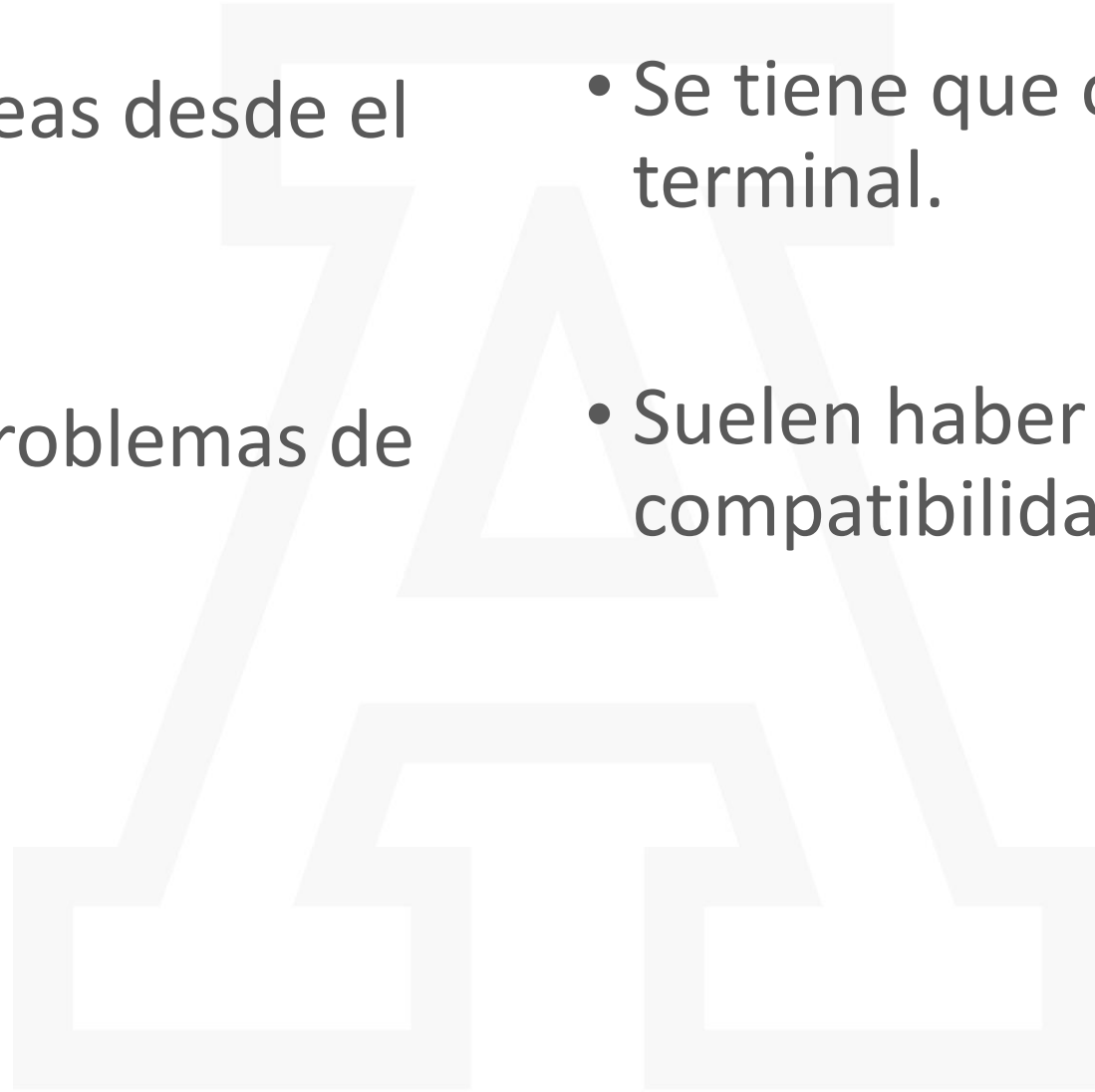


- Jupyter Notebooks



Paquetes/Librerías para ampliar funcionalidades

- Con un par de líneas desde el programa.
- Se tiene que ocupar una terminal.
- Difícil que haya problemas de compatibilidad.
- Suelen haber problemas de compatibilidad.



Asignación de Variables

- Se puede ocupar una flecha (\leftarrow) o un signo e igual (=)
- Con un signo de igual (=)



Vectores / Listas

- Con la letra **c** (*to combine*)
- Paréntesis o Corchetes
- `my_vector <- c("Tec", "ITAM")`
- `my_list = ["IPN", "UNAM"]`
- `my_tuple = ("UDLAP", "ULSA")`



Seleccionar un elemento

`my_list[1]`



Contar

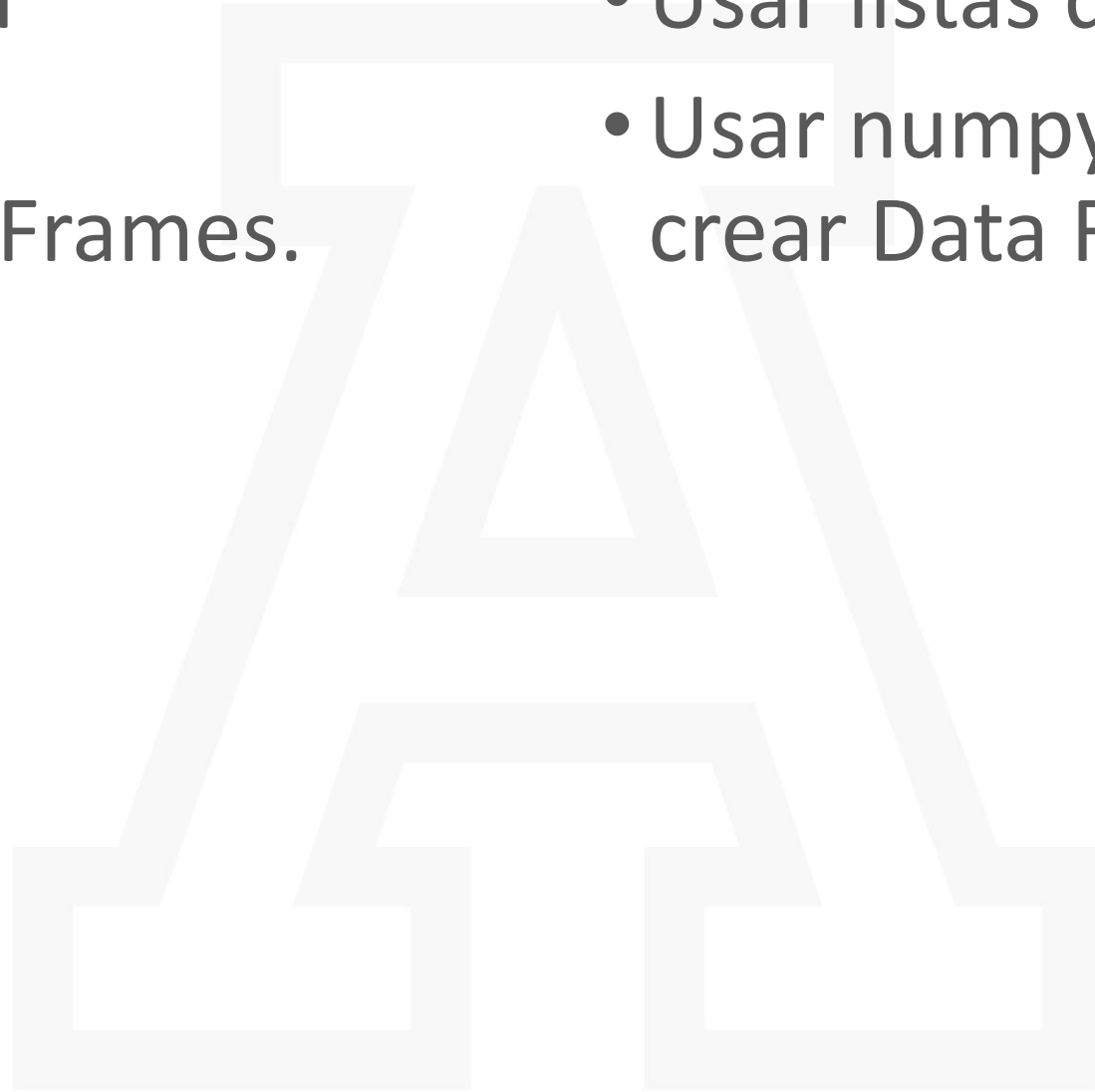
- 1, 2, 3...

- 0, 1, 2...



Trabajar con más de una dimensión (tablas)

- ¡Se pueden usar naturalmente!
- Se llaman Data Frames.
- Usar listas de listas.
- Usar numpy, pandas para crear Data Frames.



Trabajar con más de una dimensión (tablas)

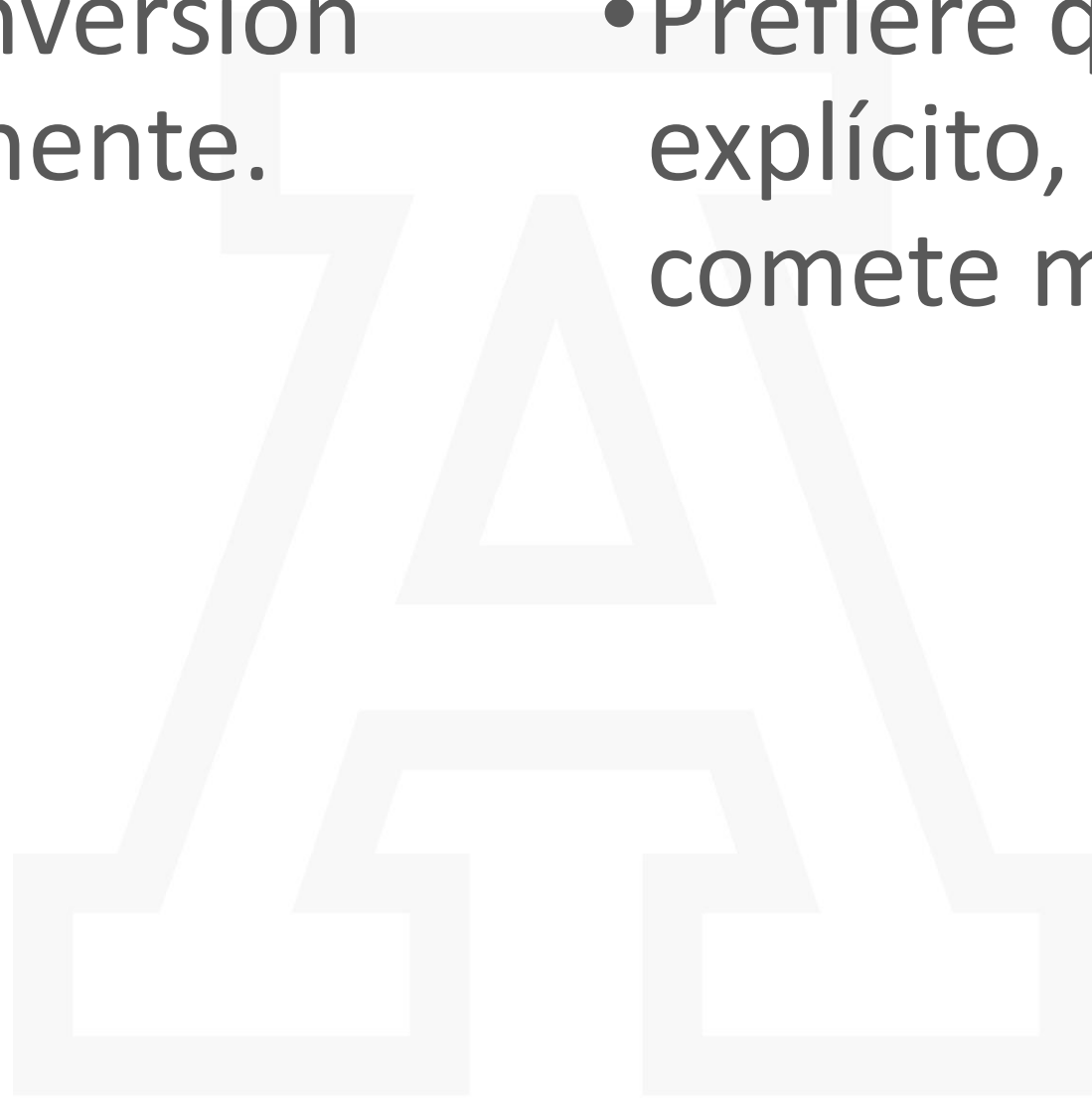
```
print('Los cinco delitos con mayor decremento y mayor incremento en Enero a Junio 2020, a comparación del año pasado, son:')  
  
delitos_cum['Cambio'] = (delitos_cum[2020]/delitos_cum[2019])-1  
delitos_cum[(delitos_cum[2020]>50) & (delitos_cum[2019]>50)].sort_values('Cambio')
```

Los cinco delitos con mayor decremento y mayor incremento en Enero a Junio 2020, a comparación del año pasado, son:

ao_hechos	2019.0	2020.0	All	Cambio
delito				
ROBO A PASAJERO A BORDO DE METRO SIN VIOLENCIA	1669	335	2004	-0.799281
ROBO A PASAJERO A BORDO DE METRO CON VIOLENCIA	223	55	278	-0.753363
ROBO A TRANSEUNTE A BORDO DE TAXI PÚBLICO Y PRIVADO CON VIOLENCIA	205	52	257	-0.746341
FALSEDAD ANTE AUTORIDADES	258	71	329	-0.724806
ROBO A NEGOCIO CON VIOLENCIA	2717	901	3618	-0.668384
...
DDH FRVT	52	85	137	0.634615
DESOBEDIENCIA Y RESISTENCIA DE PARTICULARES	190	353	543	0.857895
DDH OTRAS MATERIAS	97	190	287	0.958763
DDH RELACIONADAS	56	145	201	1.589286
ROBO A TRANSEUNTE EN PARQUES Y MERCADOS CON VIOLENCIA	85	252	337	1.964706

Múltiples tipos de datos

- R hace la conversión automáticamente.
- Prefiere que seas explícito, por lo que comete menos errores.



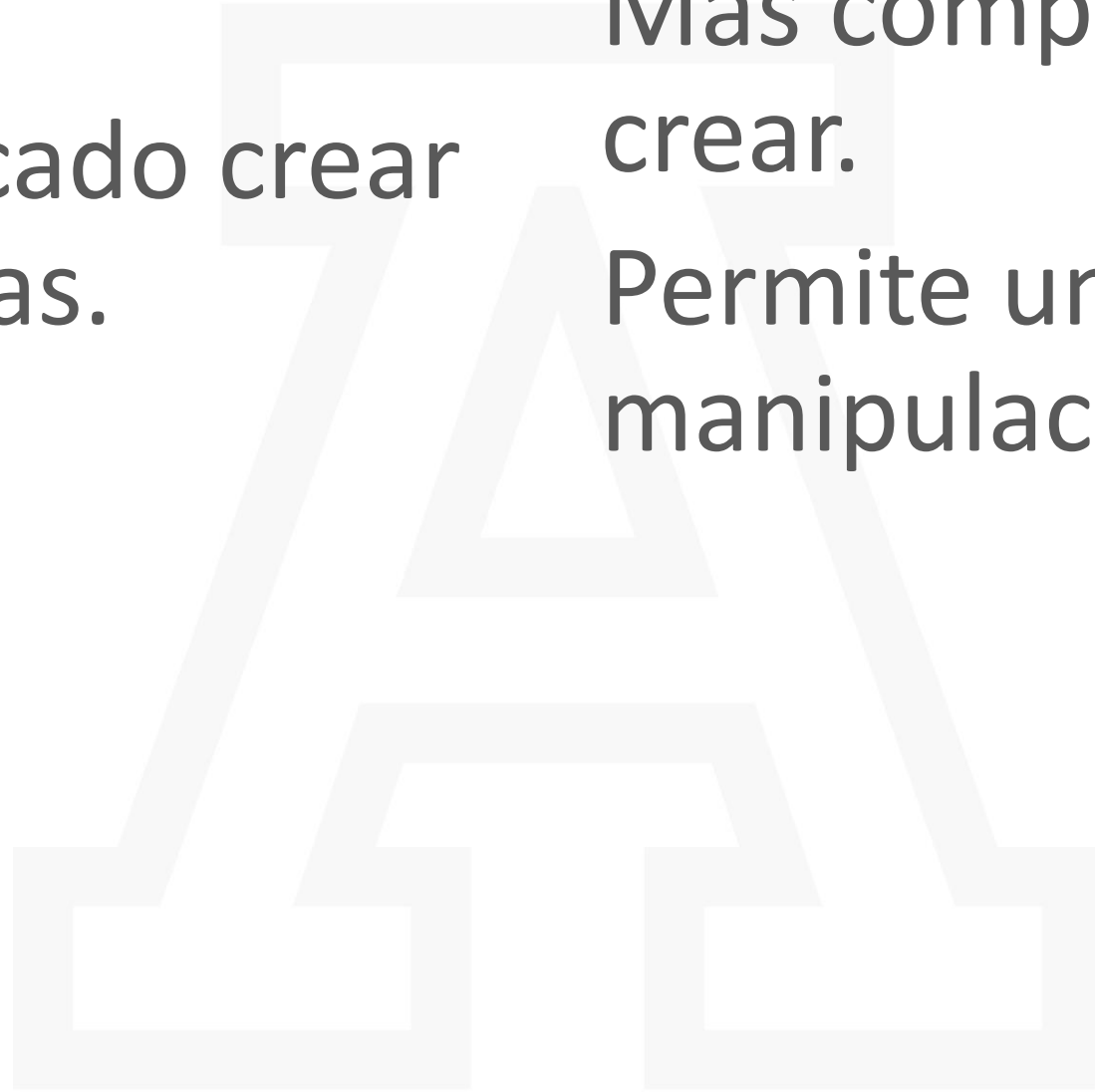
Visualizaciones (Gráficas)

Fácil de crear.

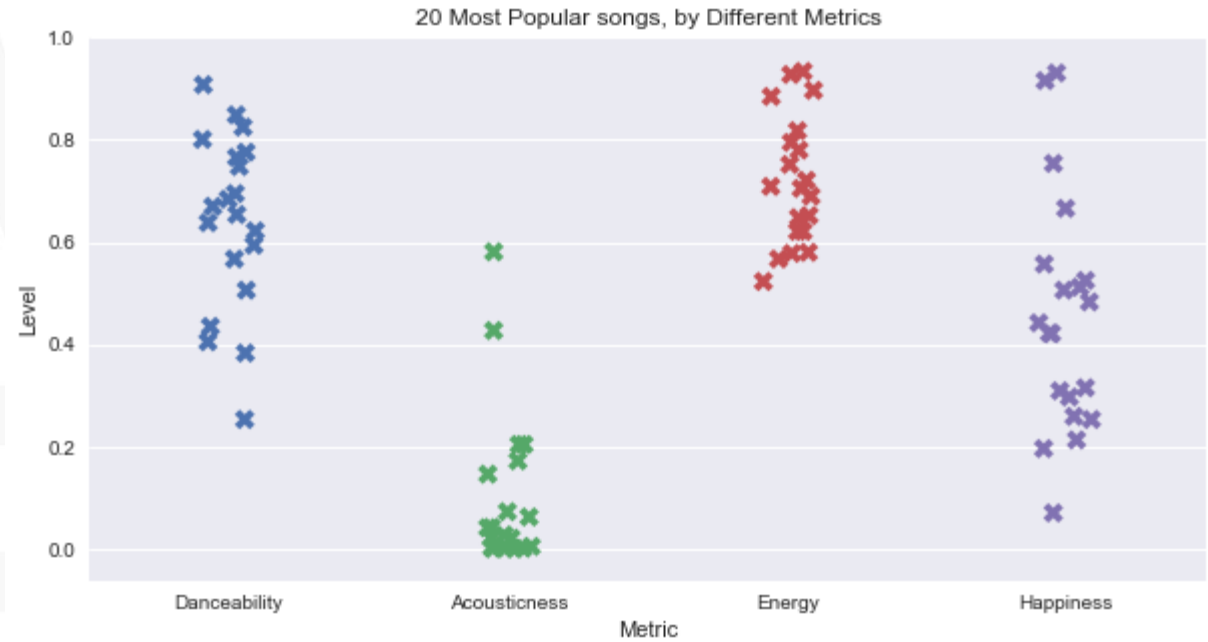
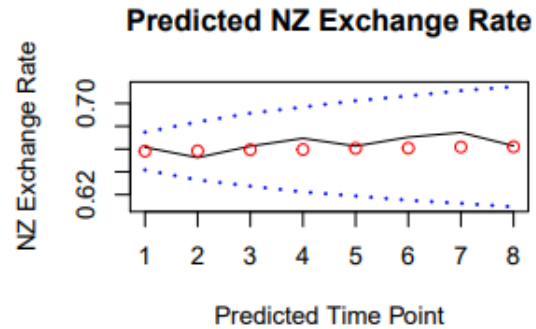
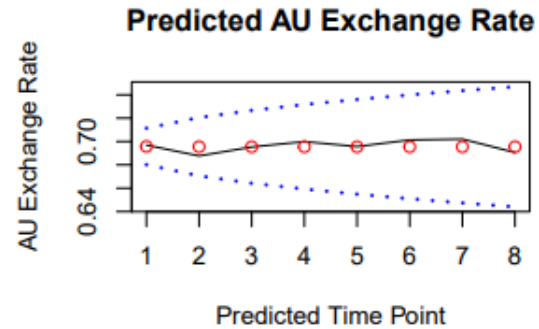
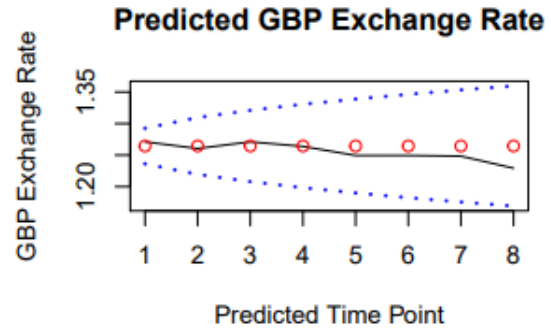
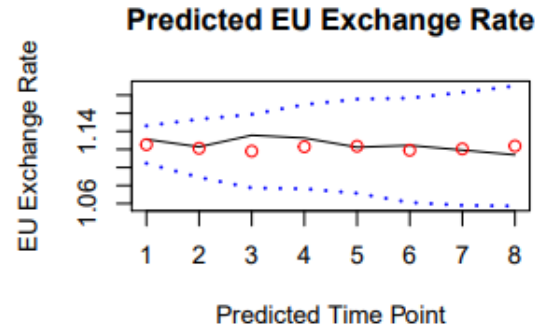
No es complicado crear buenas gráficas.

Más complicadas de crear.

Permite una mayor manipulación.



Visualizaciones (Gráficas)



Modelado Estadístico

```
> model <- lm(mass~height)
> summary(model)
```

```
Call:
lm(formula = mass ~ height)
```

Residuals:

Min	1Q	Median	3Q	Max
-61.43	-30.03	-21.13	-17.73	1260.06

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-13.8103	111.1545	-0.124	0.902
height	0.6386	0.6261	1.020	0.312

Residual standard error: 169.4 on 57 degrees of freedom
(28 observations deleted due to missingness)

Multiple R-squared: 0.01792, Adjusted R-squared: 0.0006956

F-statistic: 1.04 on 1 and 57 DF, p-value: 0.312



Modelado Estadístico

```
In [3]: from sklearn.linear_model import LinearRegression  
model = LinearRegression()  
model.fit(X, y)
```

```
Out[3]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,  
normalize=False)
```

```
In [4]: model.coef_
```

```
Out[4]: array([12.44002424])
```

```
In [5]: model.intercept_
```

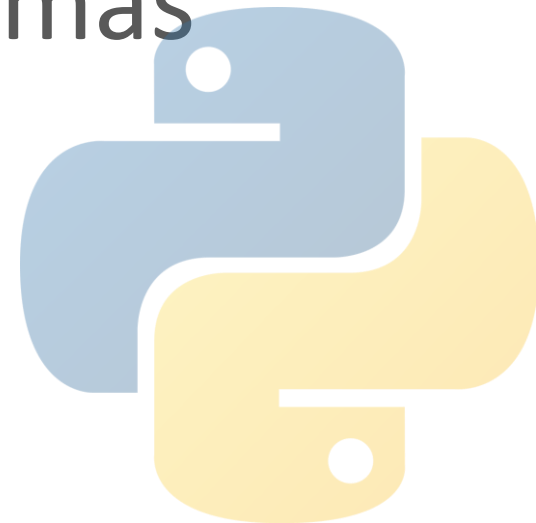
```
Out[5]: 101.89622505659258
```



ANÁHUAC

Puesta en Producción

- Python es un lenguaje que se utiliza para muchas otras aplicaciones, además de para el análisis.
- Por lo tanto, es mucho más fácil de compartir resultados en aplicaciones, páginas web, etc.
- Por ello, últimamente se utiliza un poco más seguido que R.



ANÁHUAC

El Futuro

- ¿Julia reemplazará a R y Python?
- ¿Llegará otro nuevo lenguaje?
- Sin importar qué pase, el desarrollar la lógica de programación es más importante que sólo aprender código de memoria.
- De esa forma, ¡nos podremos adaptar!

