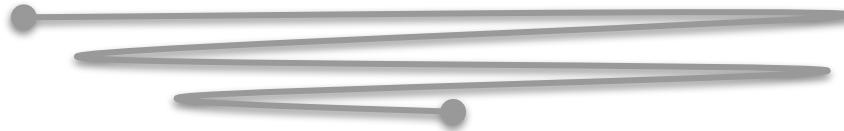


# *Laboratorio de Datos*

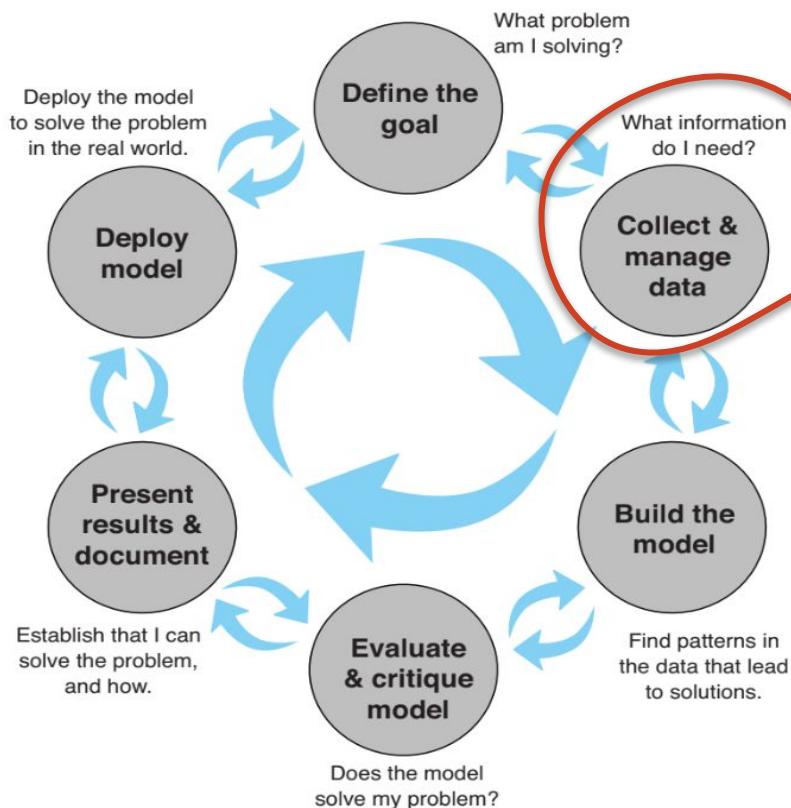


## *Introducción a Modelado*

*Pablo Turjanski*



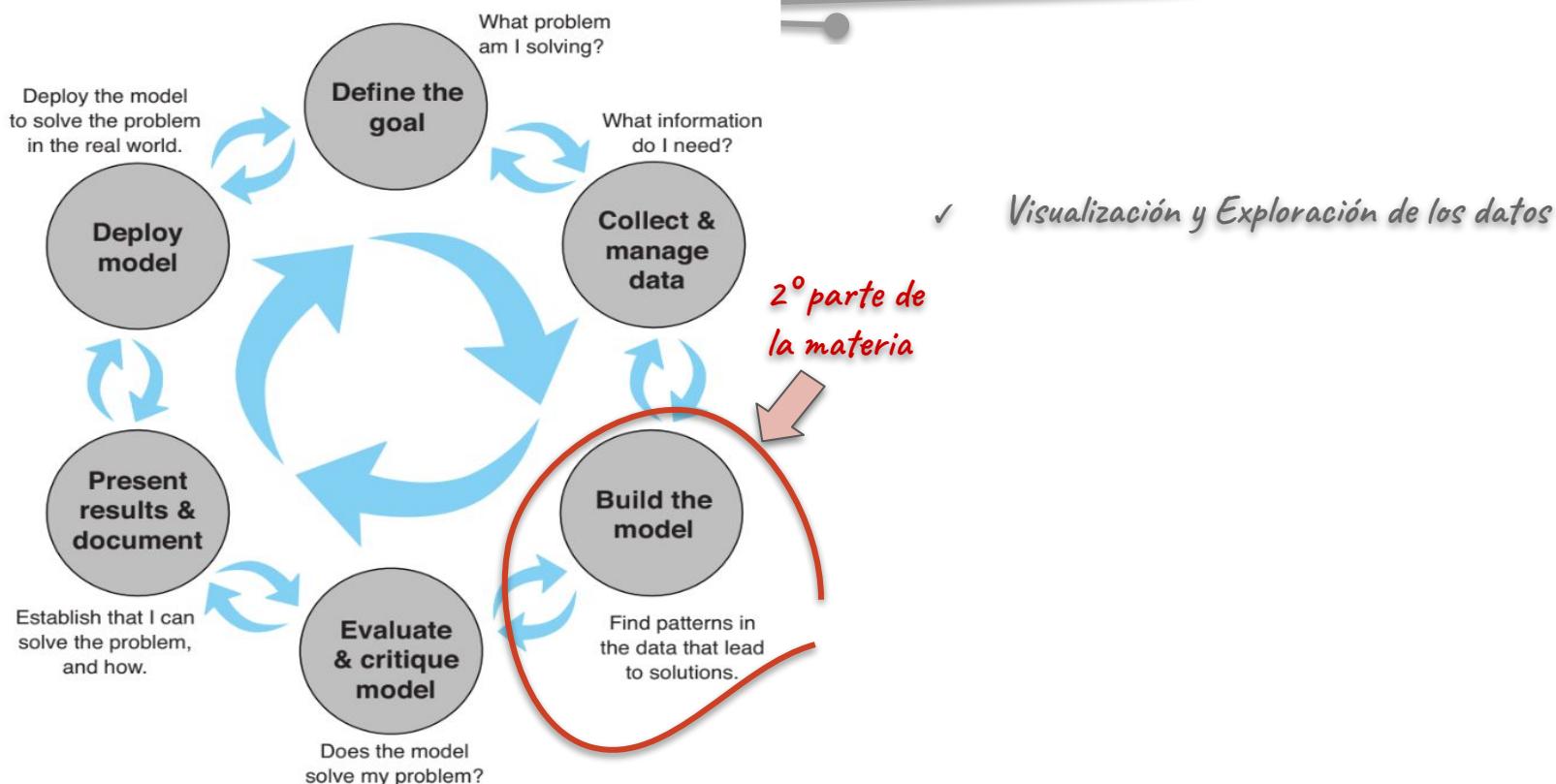
# Recorrido de la materia (hasta ahora)



1º parte de  
la materia

- ✓ Lenguaje de programación (Python)
- ✓ Modelado conceptual de los datos (DER)
- ✓ Representación de los datos (modelo relacional)
- ✓ Formas de consultar los datos (AR/SQL)
- ✓ Recomendaciones para el diseño (Normalización)
- ✓ Calidad de datos
- ✓ Leyes acerca de la Protección de Datos

# Recorrido de la materia (hasta ahora)

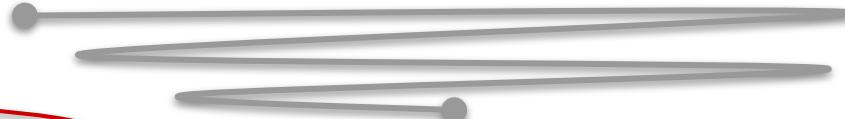


*Preguntas, experimentos & datos, análisis, conclusiones ...*



*Empecemos con un problema en particular ...*

Preguntas, experimentos & datos, análisis, conclusiones ...



Queremos saber qué tan rápido reaccionamos a un estímulo

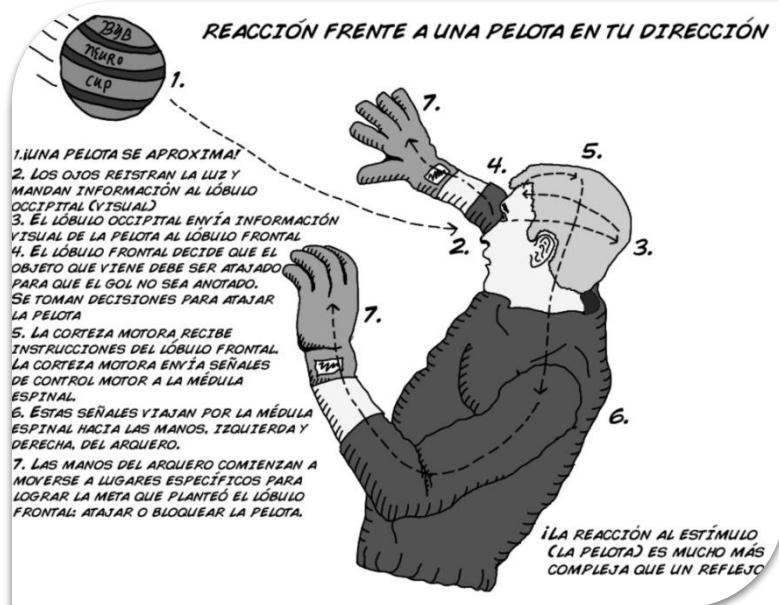
El tiempo de reacción es una medida de la rapidez con la que un organismo responde a algún tipo de estímulo.

Estímulo -----> Respuesta  
Tiempo de  
reacción

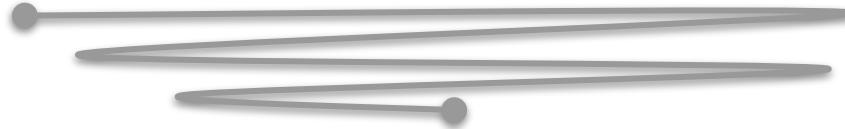
# Preguntas, experimentos & datos, análisis, conclusiones ...



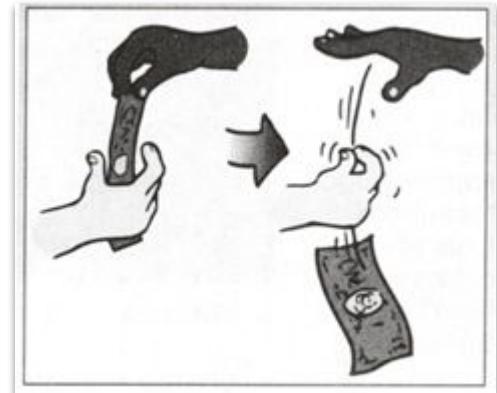
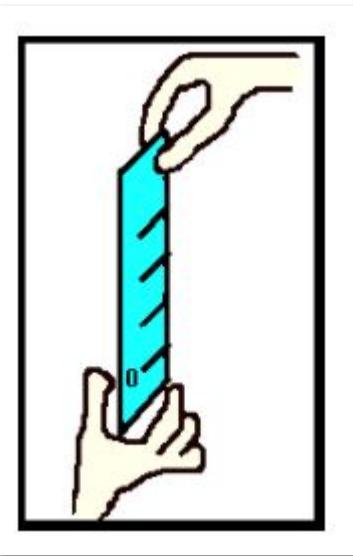
Quizás conozcan algunos experimentos ...



*Preguntas, experimentos & datos, análisis, conclusiones ...*

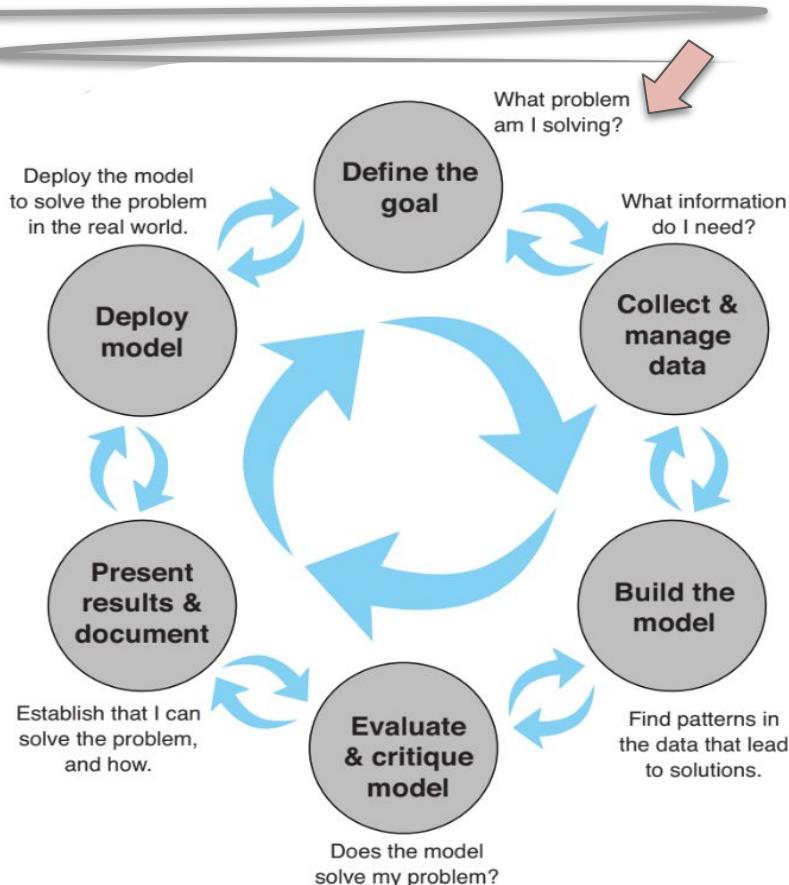


*Otros ...*

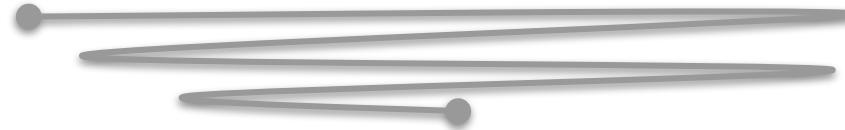


# Preguntas, experimentos & datos, análisis, conclusiones ...

Pregunta muy amplia ...

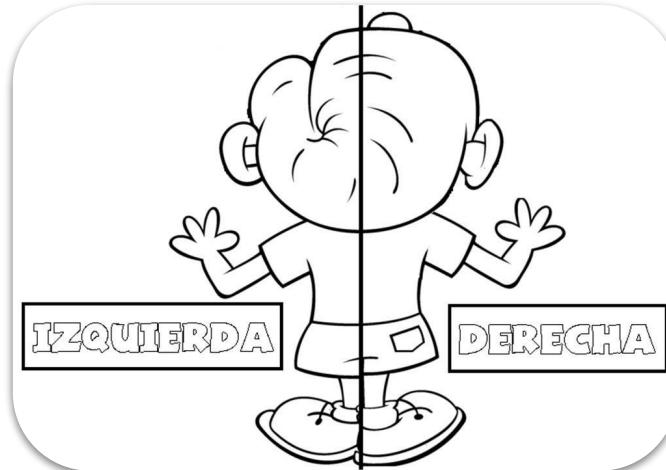


## Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

$$t_{\text{reacción}}^{\text{(mano izquierda)}} = t_{\text{reacción}}^{\text{(mano derecha)}} ?$$



# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea: "Hacer clic sobre número asignado"

Tiempo 0  
Mano Izquierda  
Número al azar 7

1	2	3
4	5	6
7	8	9

Comenzar

# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea
2. Cada estudiante ejecuta la tarea con la mano que le es asignada (derecha o izquierda)

Tiempo 0

Mano Izquierda

Número al azar 7

1	2	3
4	5	6
7	8	9

Comenzar

# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

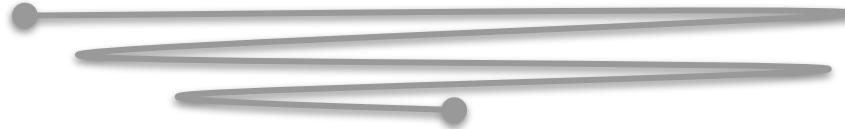
Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea
2. Cada estudiante ejecuta la tarea con la mano que le es asignada (derecha o izquierda)
3. Elegimos un/unos estudiantes para que realicen la experiencia

Tiempo 0 Mano Izquierda Número al azar 7

1	2	3
4	5	6
7	8	9
Comenzar		

# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea
2. Cada estudiante ejecuta la tarea con la mano que le es asignada (derecha o izquierda)
3. Elegimos un/unos estudiantes para que realicen la experiencia
4. Se registra el **tiempo de reacción** (tiempo que tardó desde que se le asignó la tarea hasta que la ejecutó)

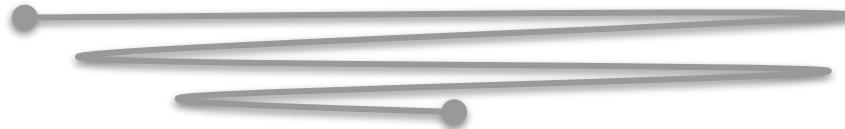
Tiempo 0

Mano Izquierda

Número al azar 7

1	2	3
4	5	6
7	8	9
<span style="background-color: green; color: white; padding: 5px;">Comenzar</span>		

# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea
2. Cada estudiante ejecuta la tarea con la mano que le es asignada (derecha o izquierda)
3. Elegimos un/unos estudiantes para que realicen la experiencia
4. Se registra el **tiempo de reacción** (tiempo que tardó desde que se le asignó la tarea hasta que la ejecutó)

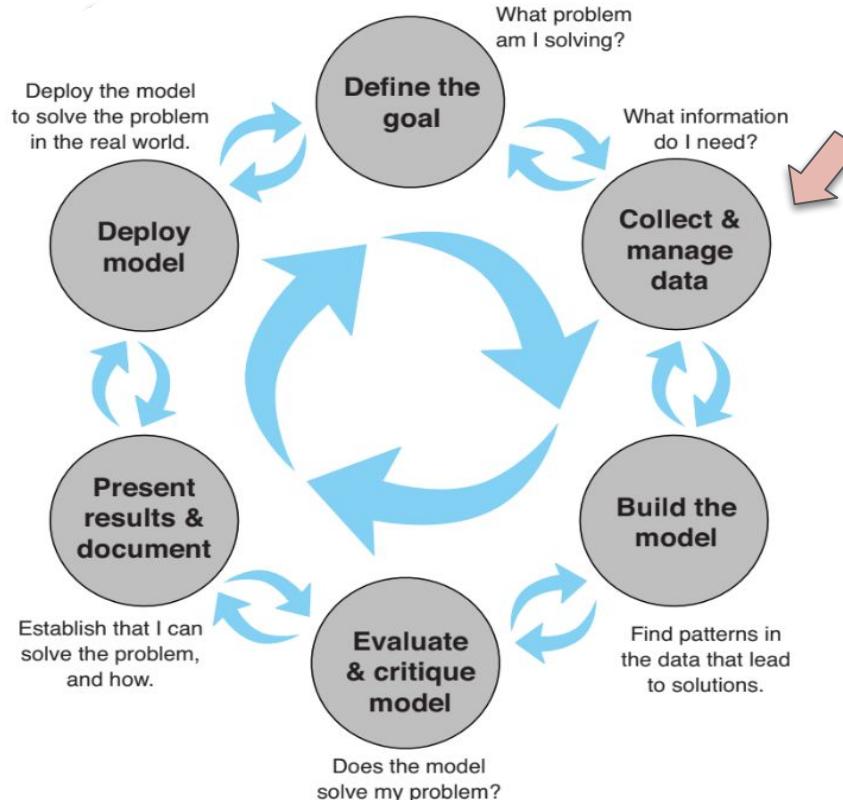
The interface includes the following elements:

- Tiempo: 0
- Mano: Izquierda
- Número al azar: 7
- Table:

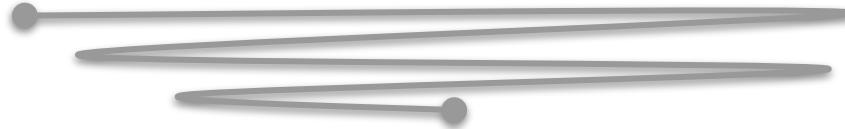
A	B	C
Estudiante N°	Mano	tiempo
1		
2		
3		
4		
7		

# Preguntas, experimentos & datos, análisis, conclusiones ...

Recolección de datos ...



# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

Tiempo 0

Mano Izquierda

Número al azar 7

1	2	3
4	5	6
7	8	9

Comenzar

[Link a dispositivo](#)

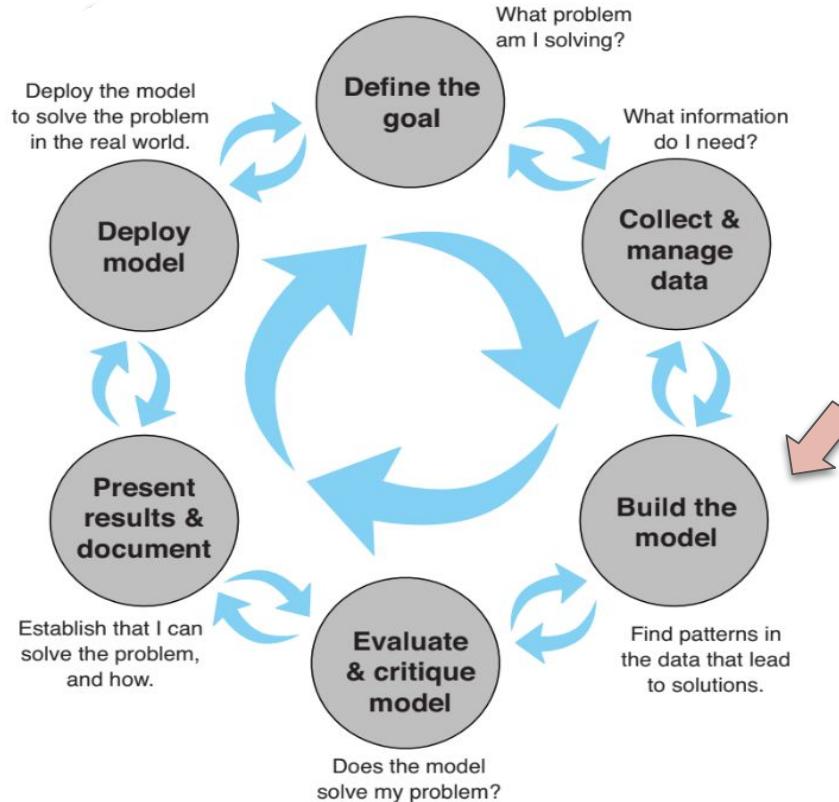


A	B	C
Estudiante N°	Mano	tiempo

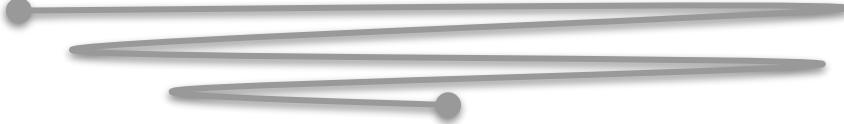
[Link a planilla](#)

# Preguntas, experimentos & datos, análisis, conclusiones ...

Busquemos evidencia en  
los datos ...



## Preguntas, experimentos & datos, análisis, conclusiones ...

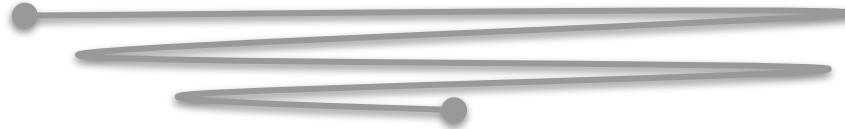


Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

¿Qué podemos hacer con los datos  
para intentar responder esta pregunta?

- Promedio como medida de tendencia central
- Gráficos. Ejemplo: Histograma de frecuencias relativas

*Preguntas, experimentos & datos, análisis, conclusiones ...*

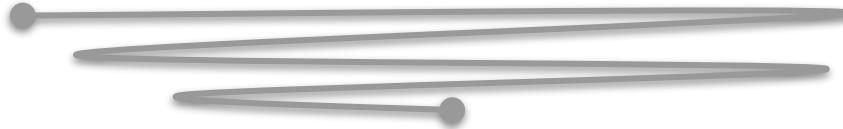


*Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?*



*(Promedio, Gráficos)*

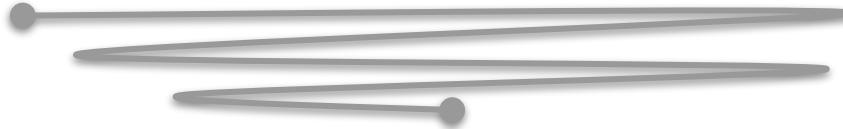
## *Preguntas, experimentos & datos, análisis, conclusiones ...*



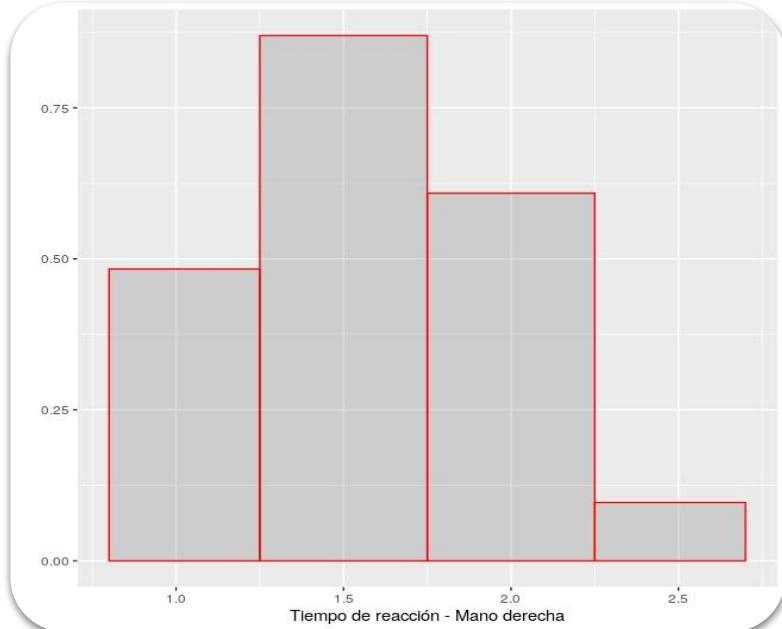
*Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?*

<i>Promedio Total</i>	1.85
<i>Promedio Derecha</i>	1.58
<i>Promedio Izquierda</i>	2.06

# Preguntas, experimentos & datos, análisis, conclusiones ...



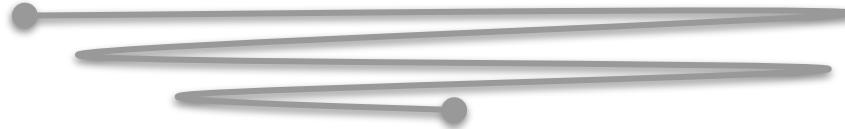
Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?



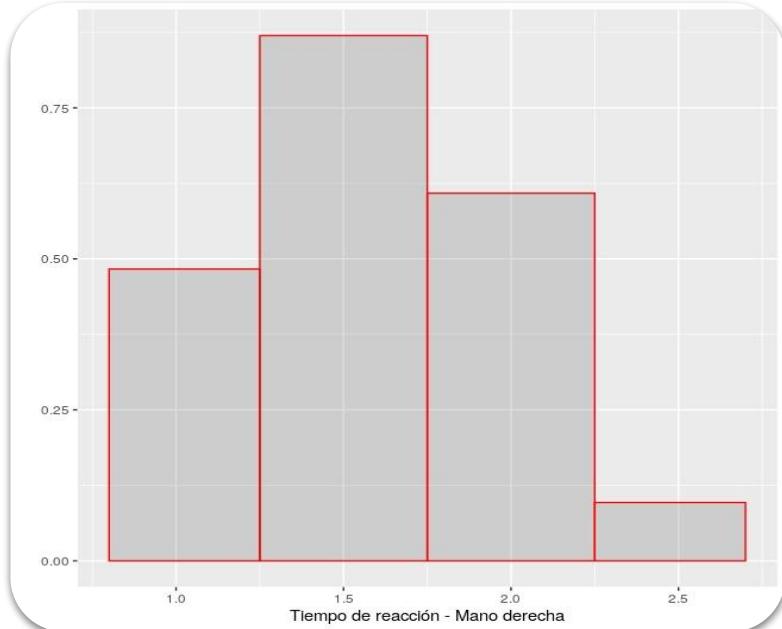
Histograma

- Manera visual de representar los datos
- Se agrupan los datos en diferentes intervalos
- El área sobre cada intervalo es la proporción de observaciones de ese intervalo.

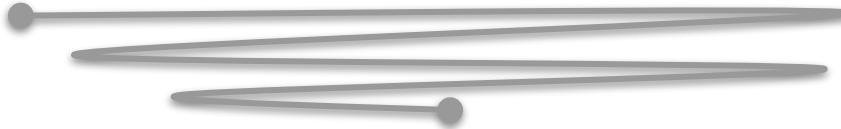
# Preguntas, experimentos & datos, análisis, conclusiones ...



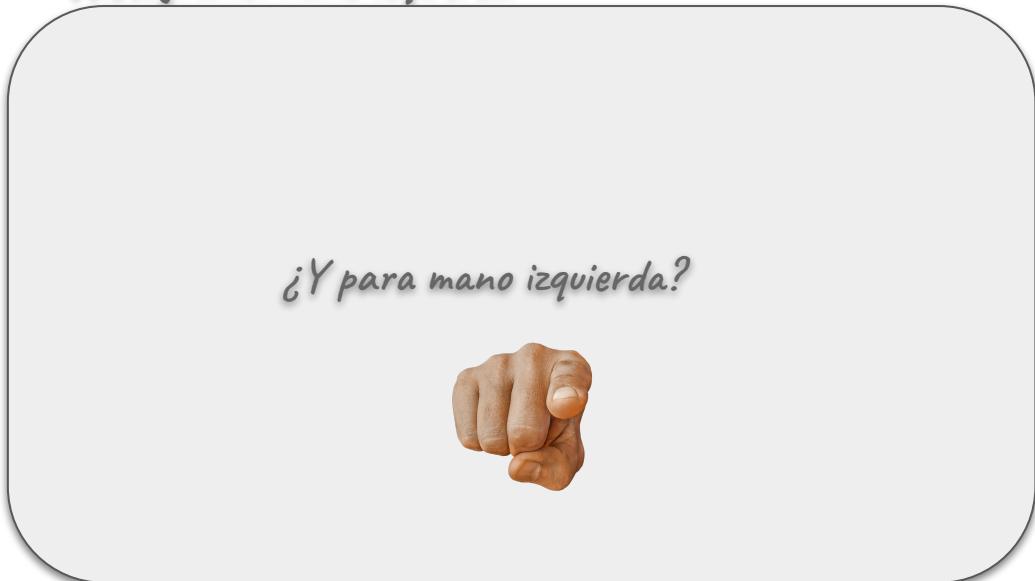
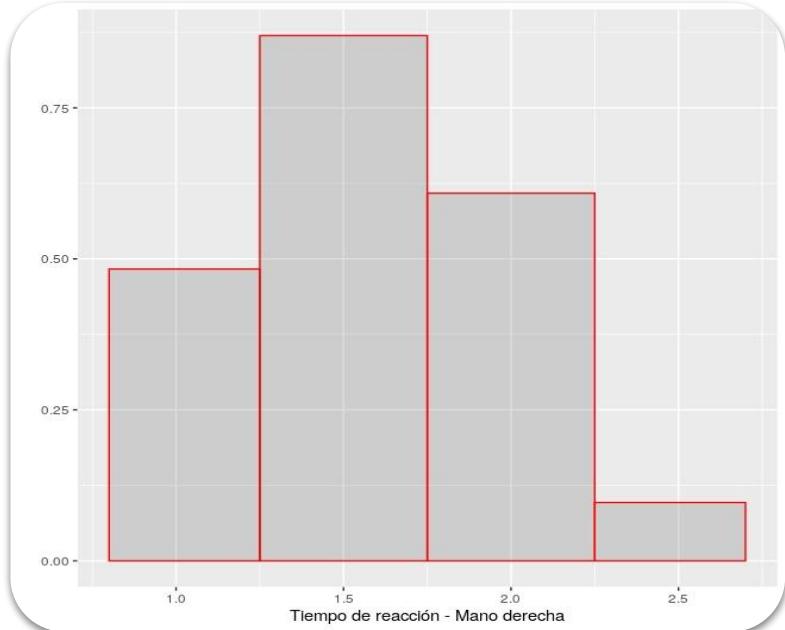
Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?



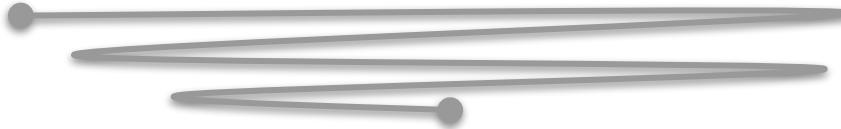
# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?



# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?

Simulador de observaciones ...

[Link a Simulador](#)  
(Desarrollado por Mariela Sued)



Nos permite generar varios casos  
¿Qué obtenemos ...?

## Tiempo de Reacción

### Diseño experimental

Mano Tiempo

Mano	Tiempo
Derecha	2.27
Derecha	1.93
Izquierda	5.64
Izquierda	1.05
Izquierda	4.50
Derecha	5.11
Izquierda	4.16
Izquierda	6.73
Derecha	1.98
Derecha	3.21
Derecha	0.78
Izquierda	0.81
Izquierda	2.36
Izquierda	3.74
Izquierda	4.08
Izquierda	4.52
Izquierda	2.33
Izquierda	3.88
Izquierda	4.30
Derecha	2.69

Cantidad de datos (entre 5 y 50):  
20

Últimos tres números de tu celular:  
264

### Manos a la obra!

#### Filtrar Mano

- Derecha
- Izquierda

¿Cuántos individuos utilizaron la mano derecha?

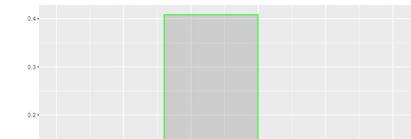
Esperando respuesta

¿Qué proporción de ellos tardó entre 1.6 y 3 segundos?

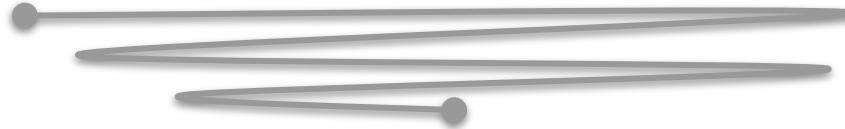
Esperando respuesta

### Histograma de tiempos de reacción de manos derechas

#### Frecuencias relativas - Mano Derecha



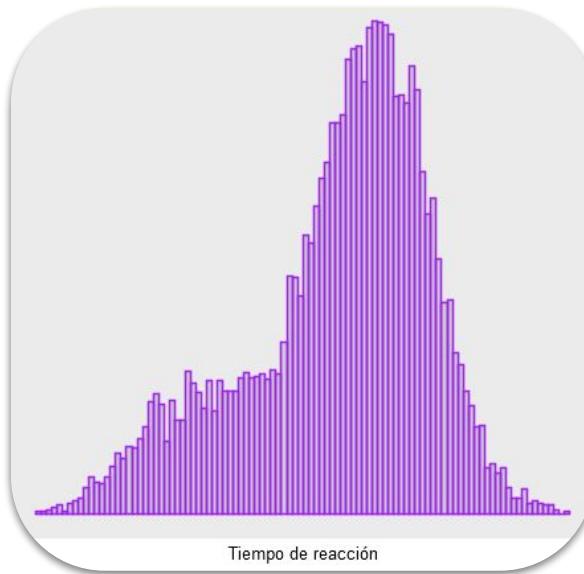
# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?



Mano derecha

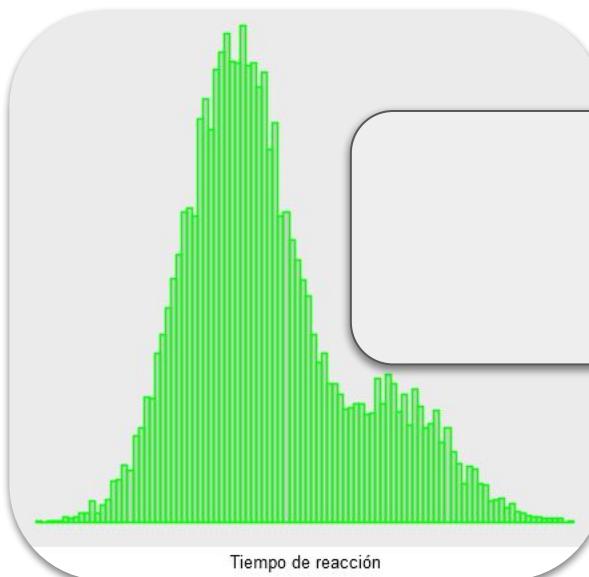


Mano Izquierda

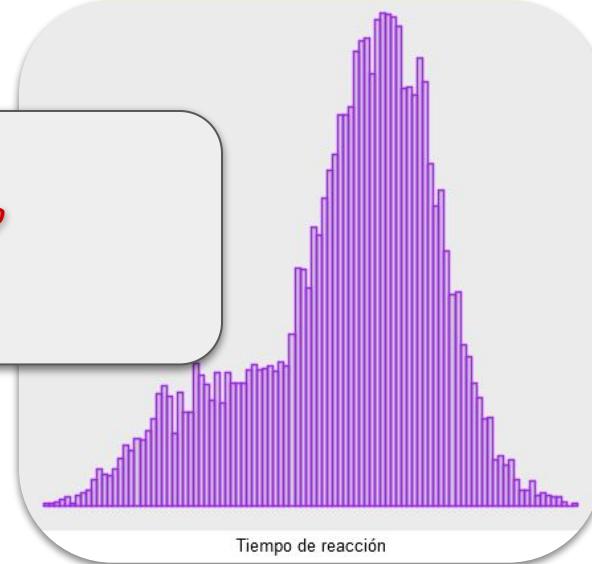
# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?



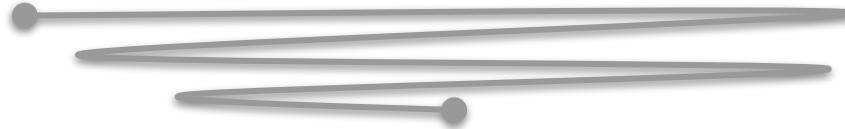
Mano derecha



Mano Izquierda

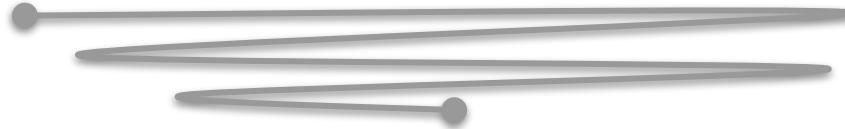
¿Alguna observación?

*Preguntas, experimentos & datos, análisis, conclusiones ...*



*Inicio de espacio para críticas*

*Preguntas, experimentos & datos, análisis, conclusiones ...*

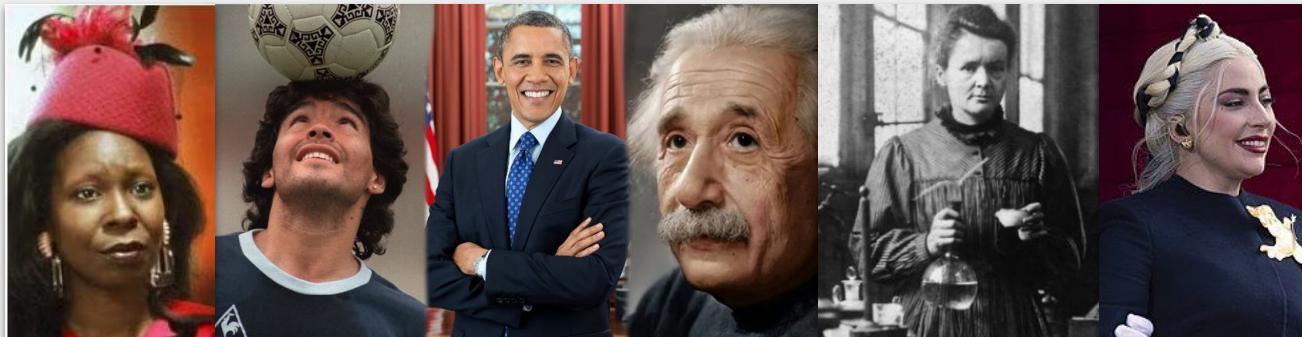


*Espacio para críticas*

# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?



Tenemos que reformular la pregunta

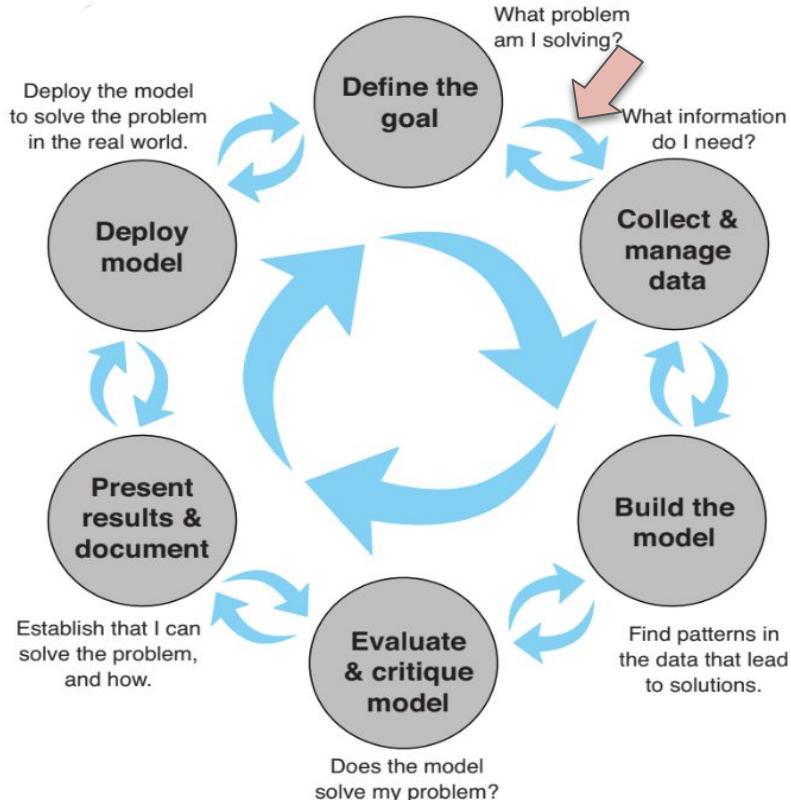
¿Qué tienen en común estas personalidades?

¡¡¡Su mano hábil es la zurda!!!

¿Qué consecuencias tiene?

# Preguntas, experimentos & datos, análisis, conclusiones ...

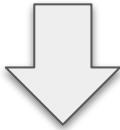
Reformular la pregunta ...



Preguntas, experimentos & datos, análisis, conclusiones ...

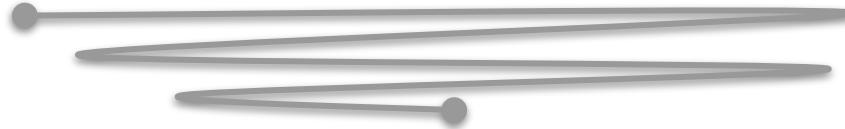


Pregunta: ¿El tiempo de reacción, es el mismo con la mano derecha que con la mano izquierda?



¿El tiempo de reacción, es el mismo con la mano **hábil** que con la mano **no hábil**?

*Preguntas, experimentos & datos, análisis, conclusiones ...*

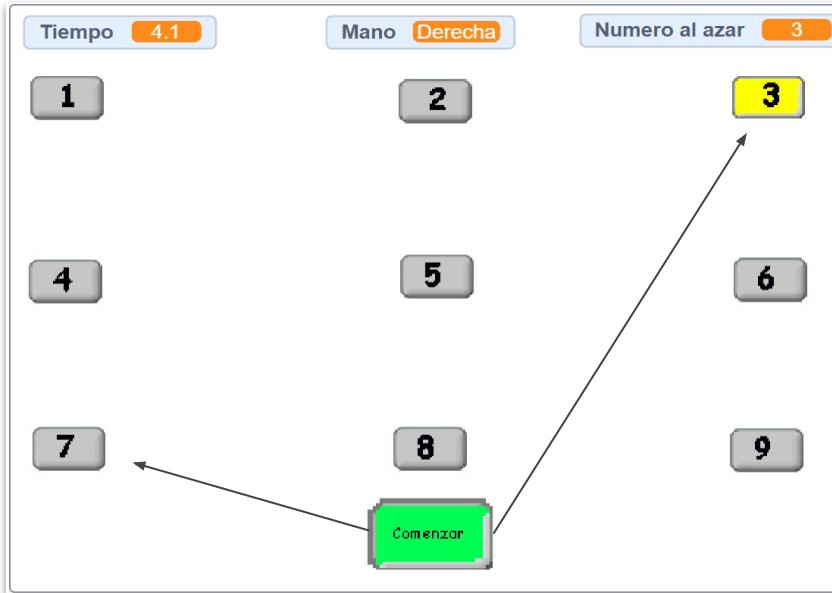


*Espacio para críticas*

# Preguntas, experimentos & datos, análisis, conclusiones ...



## Dispositivo experimental: Dial Cuadrado



¿Qué consecuencias tiene?

¿Cómo se podría solucionar?

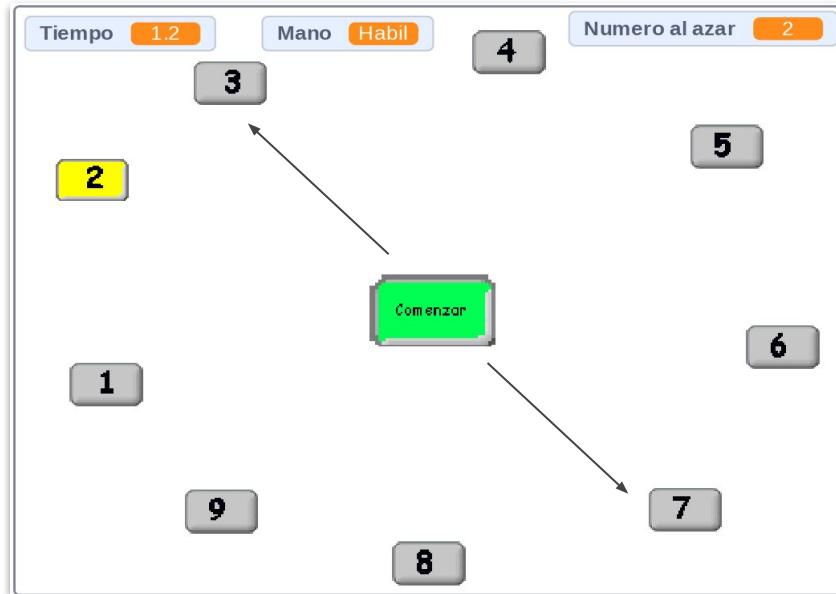
¡A mí me tocó el 7!

¡y a mí el 3!

Preguntas, experimentos & datos, análisis, conclusiones ...



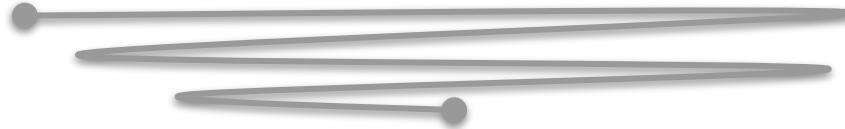
Dispositivo experimental: Dial Circular



¡A mí me tocó el 7!

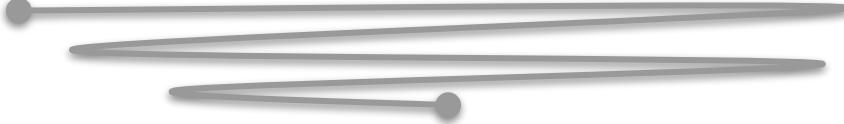
¡y a mí el 3!

*Preguntas, experimentos & datos, análisis, conclusiones ...*



*Espacio para críticas*

## Preguntas, experimentos & datos, análisis, conclusiones ...

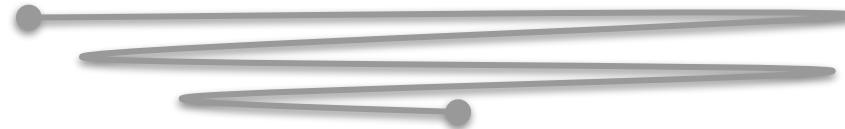


Pregunta: ¿El tiempo de reacción, es el mismo con la mano hábil que con la mano no hábil?

Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea
2. Cada estudiante ejecuta la tarea con la mano que le es asignada (derecha o izquierda)
3. Elegimos un/unos estudiantes para que realicen la experiencia
4. Se registra el **tiempo de reacción** (tiempo que tardó desde que se le asignó la tarea hasta que la ejecutó)

# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano hábil que con la mano no hábil?

Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea
2. Cada estudiante ejecuta la tarea con la mano que le es asignada (derecha o izquierda)
3. Elegimos un/unos estudiantes para que realicen la experiencia
4. Se registra el tiempo de reacción (tiempo que tardó desde que se le asignó la tarea hasta que la ejecutó)

¡Qué consecuencias tiene?

Estudiante 1



Estudiante 2



Mano hábil	Mano no hábil
0.8	1.1

Mano hábil	Mano no hábil
1.5	2.3

¡Cómo se podría solucionar?

# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano hábil que con la mano no hábil?

Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea
2. Cada estudiante ejecuta la tarea con la mano que le es asignada (derecha o izquierda)
3. Elegimos un/unos estudiantes para que realicen la experiencia
4. Se registra el tiempo de reacción (tiempo que tardó desde que se le asignó la tarea hasta que la ejecutó)

Estudiante N°	Mano	Tiempo
38	Hábil	1.3
39	No hábil	2.1
40	Hábil	1.1
41	Hábil	1.5
42	Hábil	1.3
43	No hábil	2.4
44	No hábil	2.1
45	Hábil	1.4

A cada estudiante le tomamos el tiempo de reacción de ambas manos



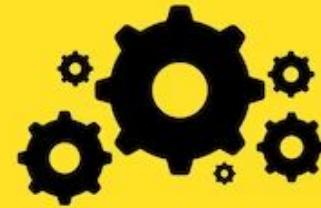
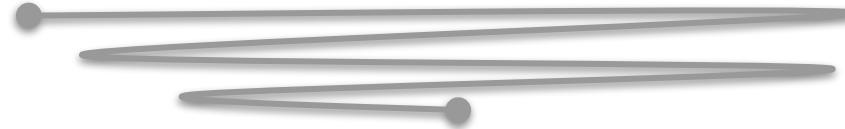
Estudiante N°	Tiempo	
	Mano hábil	Mano no hábil
38	1.3	2.2
39	1.5	2.1
40	1.1	2.9
41	1.5	2.4
42	1.3	2.0
43	1.2	2.4
44	1.1	2.1

*Preguntas, experimentos & datos, análisis, conclusiones ...*



*Fin de espacio para críticas*

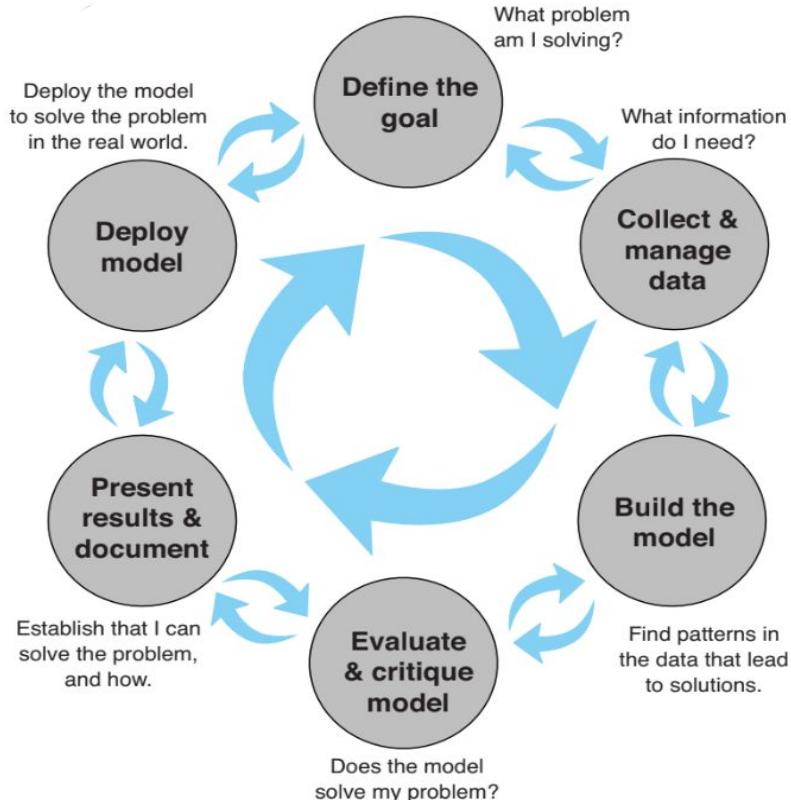
*Preguntas, experimentos & datos, análisis, conclusiones ...*



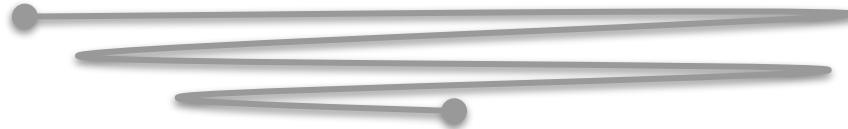
**Update...**

# Preguntas, experimentos & datos, análisis, conclusiones ...

Tenemos que ir y volver ...



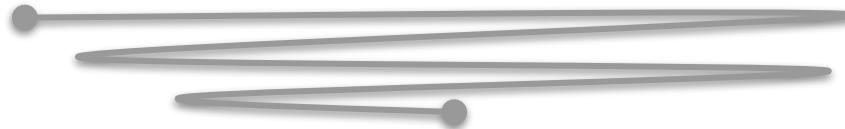
*Preguntas, experimentos & datos, análisis, conclusiones ...*



*Pregunta: ¿El tiempo de reacción, es el mismo con la mano hábil que con la mano no hábil?*



# Preguntas, experimentos & datos, análisis, conclusiones ...



Pregunta: ¿El tiempo de reacción, es el mismo con la mano hábil que con la mano no hábil?

Diseñamos una experiencia para responder esta pregunta

1. Definimos una tarea
2. Cada estudiante ejecuta la tarea con la mano que le es asignada (derecha o izquierda)
3. Elegimos un/unos estudiantes para que realicen la experiencia
4. Se registra el **tiempo de reacción** (tiempo que tardó desde que se le asignó la tarea hasta que la ejecutó)



[Link a dispositivo](#)

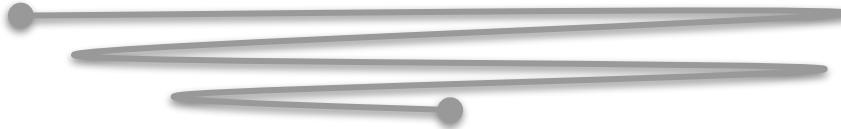
Estudiante N°	Tiempo	
	Mano hábil	Mano no hábil
38	1.3	2.2
39	1.5	2.1
40	1.1	2.9
41	1.5	2.4
42	1.3	2.0
43	1.2	2.4
44	1.1	2.1

[Link a planilla](#)

Promedio Total	1.85
Promedio Habil	1.58
Promedio No Habil	2.06



## Cierre de actividad

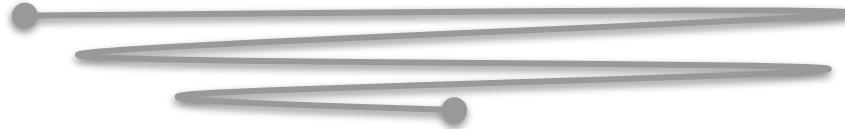


1. Los datos tienen variabilidad.

Son 'representantes' de esa población (verdad) que no conozco.

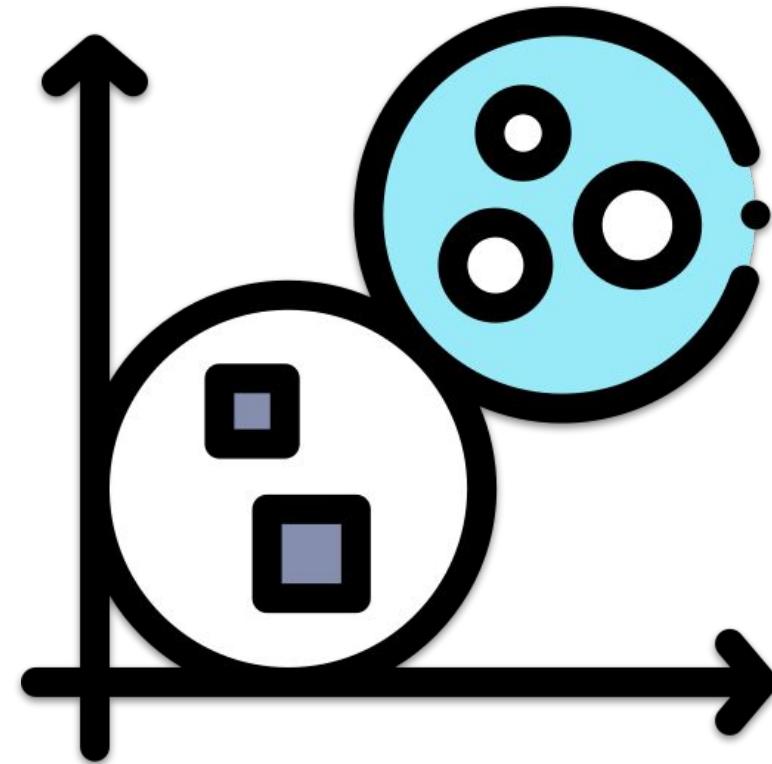
2. La Ciencia de Datos (incluyendo a la estadística) nos ayuda a tomar decisiones basados en los datos (con conocimientos previos) sobre cuestiones de la población, que son desconocidas.

## *Cierre de actividad*



3. *El diseño de una experiencia/estudio para responder una pregunta de investigación no es una tarea menor*
  - a. *Analizar los datos que surgen de una experiencia/estudio*
  - b. *Participan diversas herramientas matemáticas y computacionales*
  - c. *Involucra trabajo en equipo*
  - d. *Necesita de espíritu crítico*
  - e. *Distintas disciplinas involucradas. Conocimiento del problema, diálogo con individuos de otras disciplinas.*

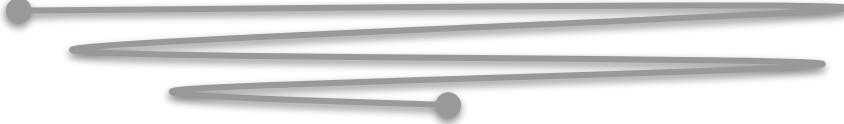
*Agrupados*



*Agrupados*



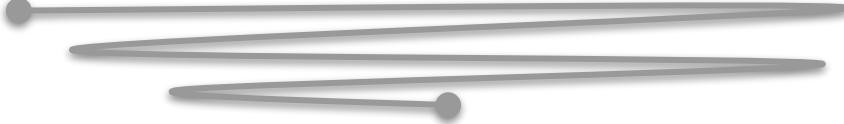
## Agrupados - Consigna - Parte 1



### Consigna.

- Armar grupos de 6 alumnxs
- Los docentes les van a facilitar un conjuntos de elementos
- Agruparlos según el criterio que les parezca más razonable
- No se pueden hacer preguntas
- Tienen 5 minutos para agruparlos
- Sacar una foto y subirla a la presentación compartida: AGRUPADOS

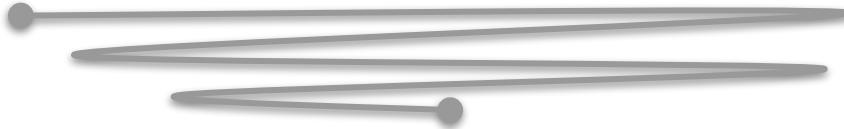
## Agrupados - Consigna - Parte 2



Cada grupo ahora escriba:

1. ¿Qué característica pueden extraer de cada objeto? Forma, alto, ancho, etc.
2. ¿Si las tuvieran que plasmar en un modelo relacional de qué forma lo harían?
3. ¿Cuáles de esas características tuvieron en cuenta para realizar el agrupamiento?
4. ¿Podrían asignarle un nombre representativo a cada grupo formado?
5. Escribir "las reglas" que tuvieron en cuenta para agrupar los objetos

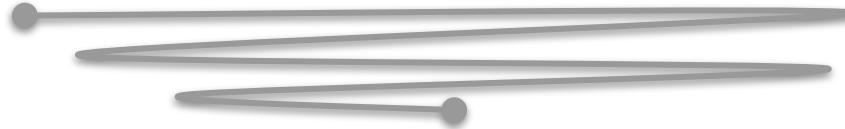
## Agrupados - Consigna - Parte 3



Cada grupo ahora escriba:

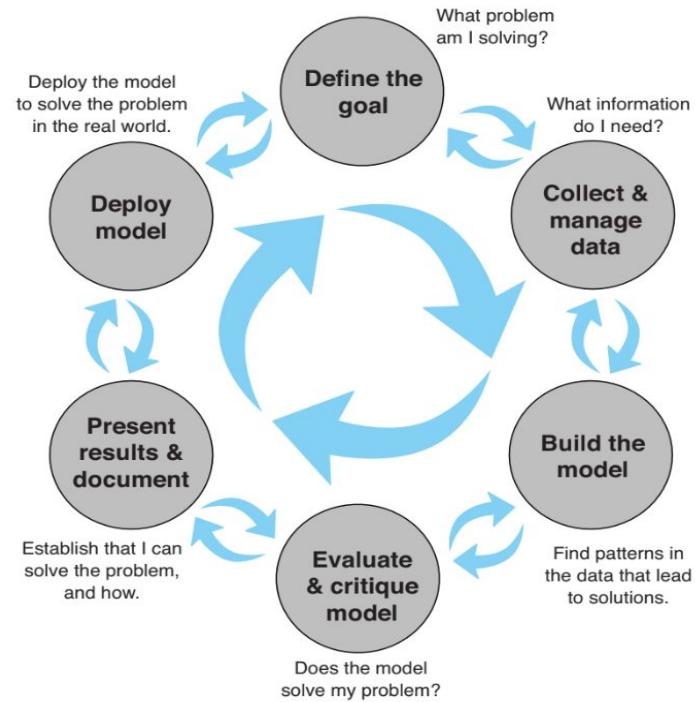
1. Considerar un objeto nuevo (que se los acercará alguno de los docentes), utilizando las reglas escritas ¿podrian asignarlo a un grupo?
2. Comentar con el resto de la clase cuál fue el criterio para asignarlo

## *Agrupados - Reflexiones*

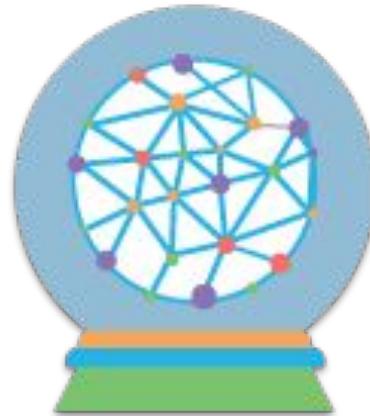
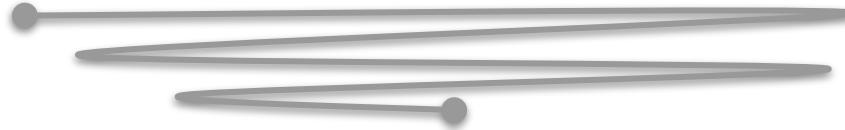


- *Los objetos tienen características*
- *En nuestro problema sólo representamos algunas (minimundo)*
- *Tenemos un objetivo: generar un agrupamiento de objetos*
- *Tratamos de agrupar los que tienen características similares*
- *Generamos reglas (un modelo)*
- *Utilizamos el modelo para agrupar un nuevo elemento*

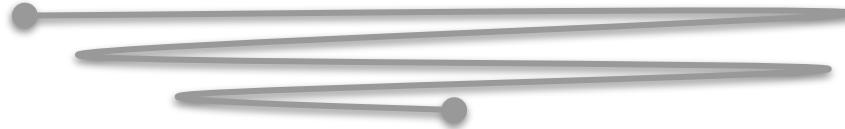
# Agrupados - Reflexiones



# *Modelado*



## Modelado



¿Quién de ustedes alguna vez se hizo una de estas preguntas?

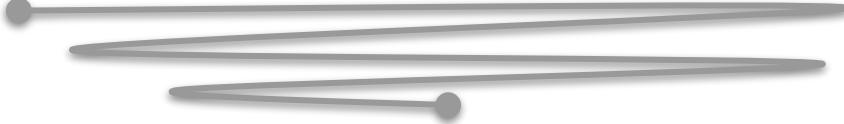
- ¿Qué camino debería tomar para volver a casa?
- ¿Debería cambiar de empresa proveedora de servicio de celular?
- ¿Es posible que en algún futuro tenga cierta enfermedad?

Este tipo de preguntas evidencian nuestra necesidad de conocer eventos futuros, y por lo general queremos tomar las mejores decisiones hacia ese futuro.

¿Qué suelen hacer para ...?



## Modelado



¿Quién de ustedes alguna vez se hizo una de estas preguntas?

- ¿Qué camino debería tomar para volver a casa?
- ¿Debería cambiar de empresa proveedora de servicio de celular?
- ¿Es posible que en algún futuro tenga cierta enfermedad?

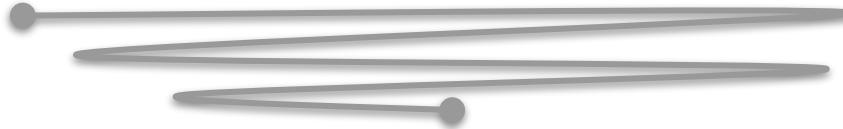
Este tipo de preguntas evidencian nuestra necesidad de conocer eventos futuros, y por lo general queremos tomar las mejores decisiones hacia ese futuro.

---

¿Qué suelen hacer para ...?



## Modelado

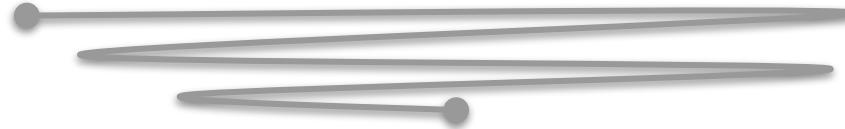


¿Quién de ustedes alguna vez se hizo una de estas preguntas?

- ¿Qué camino debería tomar para volver a casa?

**Respuesta  
basada en** { Datos → Reporte del tráfico, Reporte meteorológico, etc.  
Experiencia → Debería evitar tal calle los días que llueve

## *Modelado*

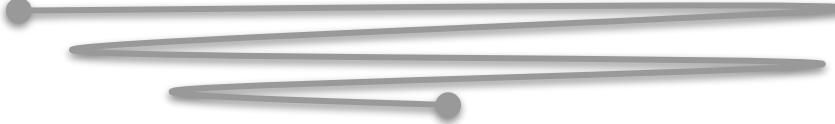


*Cada vez hay más datos a disposición y aplicaciones que:*

- 1. Toman información actual*
- 2. Filtran los datos en busca de patrones que sean relevantes para nuestro problema*
- 3. Devuelven respuestas*

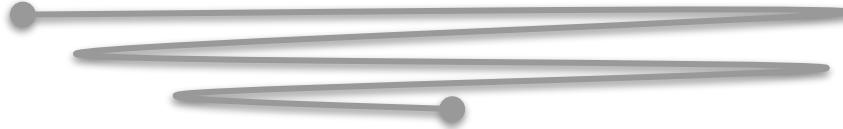
*Estos procesos involucran varias áreas del conocimiento*

## Modelado



- Evolución a lo largo de varios campos (química, informática, física y estadística)
- Distintas denominaciones:
  - "aprendizaje automático"
  - "inteligencia artificial"
  - "reconocimiento de patrones"
  - "minería de datos"
  - "análisis predictivo"
  - "descubrimiento de conocimiento"
- Si bien cada campo aborda el problema utilizando diferentes perspectivas y conjuntos de herramientas, el objetivo final es el mismo:  
Hacer una predicción precisa

## *Modelado*



Modelado predictivo: Es el proceso de desarrollar una herramienta o modelo matemático que genera una predicción precisa.\*

\* "Applied Predictive Modeling", Kuhn & Johnson 2016

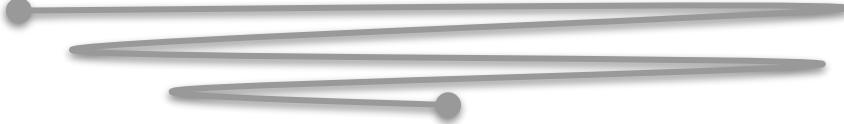
## *Modelado*



*¿Se les ocurren ejemplos de modelados predictivos?*



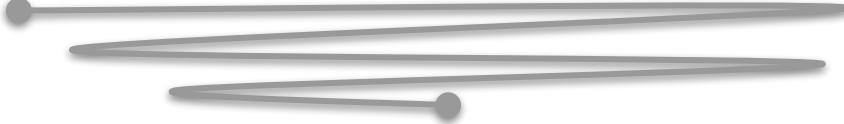
## Modelado



¿Se les ocurren ejemplos de modelados predictivos?

- ¿Tiene un paciente una enfermedad específica?
- Según sus elecciones anteriores, ¿qué películas le interesarán a este espectador?
- ¿Con qué personas debería emparejar la aplicación de citas?
- Dado un correo electrónico ¿Es SPAM?
- ¿Responderá este paciente a esta terapia?

## Modelado

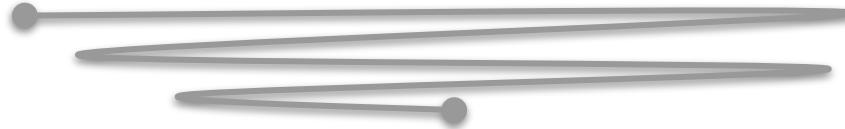


*Si bien los modelos predictivos tienden a generarnos productos más satisfactorios, mejores tratamientos médicos e inversiones más rentables, regularmente generan predicciones inexactas y brindan respuestas incorrectas.*

### ¿Ejemplos?

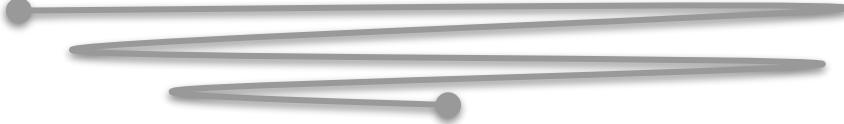
- *Alguna vez hemos perdido algún correo electrónico importante debido a un modelo predictivo (lo identificó como SPAM)*
- *Modelos predictivos de diagnóstico médico han diagnosticado erróneamente enfermedades*

## Modelado



“Nuestra habilidad para predecir o tomar decisiones está restringida por nuestro conocimiento presente y pasado, y se ve afectada por factores que quizás no hemos considerado. Estas realidades son los límites de cualquier modelo, pero estas realidades no deben impedirnos buscar mejorar nuestro proceso y construir mejores modelos.”

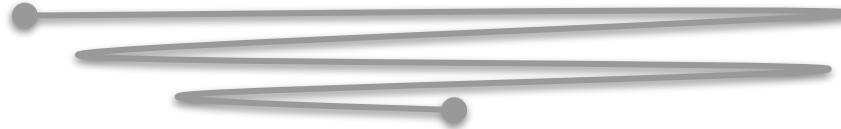
## Modelado



Razones más comunes por las que fallan los modelos predictivos:

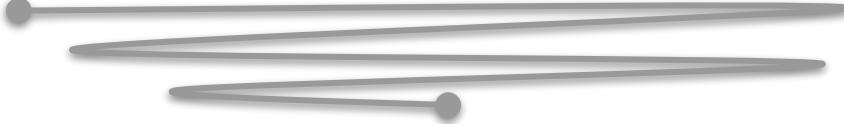
- Procesamiento previo inadecuado de los datos
- Validación inadecuada del modelo
- Extrapolación injustificada
- Ajuste excesivo del modelo a los datos existentes
- Se exploran relativamente pocos modelos cuando se buscan relaciones predictivas (en gral. debido a la preferencia, conocimiento o experiencia del modelador)

## *Predicción vs. Explicación*



*Retomando ... usamos datos históricos para crear una herramienta matemática para predecir casos futuros no vistos. Veamos algunos ejemplos ...*

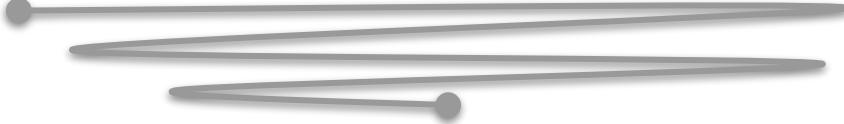
## Predicción vs. Explicación



### Ejemplo 1) Mail SPAM

- Interés en decidir con precisión si es SPAM (o no)
- Objetivo de este tipo de modelado es optimizar la precisión de la predicción  
(sin importar por qué el modelo marcó como SPAM (o no) un mail)

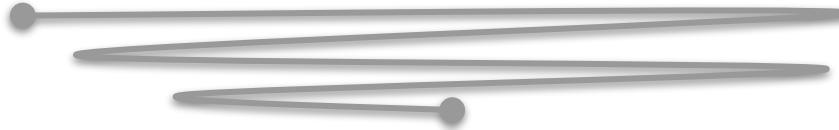
## Predicción vs. Explicación



### Ejemplo 2) Precio de venta de una propiedad

- Interés es que la tase correctamente (y conocer qué decisiones tomó)
  - Subvaluación producirá que hagan ofertas más bajas y por lo tanto un precio de venta más bajo;
  - Sobrevaluación puede ahuyentar a los potenciales compradores

## *Predicción vs. Explicación*



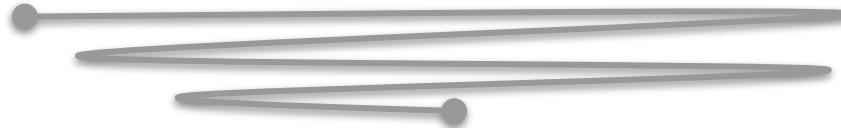
*Tensión entre ...*

*predicción (ej. Mail SPAM)*

*y*

*explicación (ej. Precio de venta de propiedad)*

## *Predicción vs. Explicación*

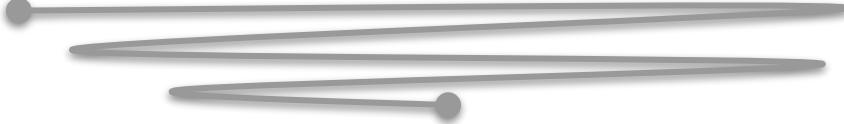


*Ejemplo 3) Proceso que enfrenta paciente con cáncer y médico cuando contemplan cambiar la terapia de tratamiento.*

*Algunos de los factores que el médico y el paciente deben considerar:*

- *Programa de dosificación*
- *Los posibles efectos secundarios*
- *Las tasas de supervivencia*

## Predicción vs. Explicación

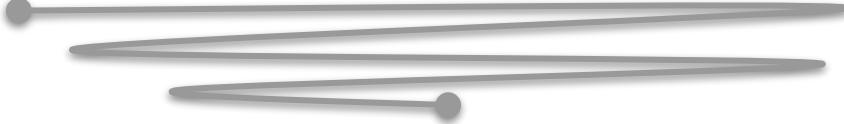


Sin embargo, si varios pacientes han seguido la terapia alternativa se podrían recopilar datos sobre estos pacientes relacionados con su enfermedad:

- Historial de tratamiento
- Datos demográficos
- Pruebas de laboratorio relacionadas con los antecedentes genéticos de los pacientes
- Otros datos biológicos (Ej. mediciones de proteínas)

Se podría crear un modelo predictivo para predecir la respuesta a la terapia alternativa en base a estos datos.

## Predicción vs. Explicación

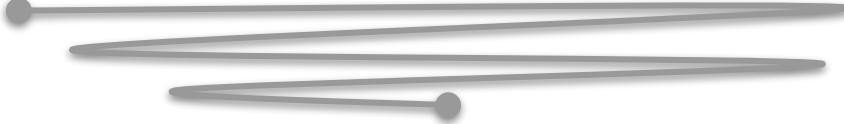


Pregunta crítica ...

*¿cómo va a reaccionar el paciente ante un cambio en la terapia?*

- *La predicción debe ser precisa, y si se crea un modelo para hacer esta predicción ... ¿Debería estar limitado por el requisito de interpretabilidad?*
- *Si no es así, se podría argumentar con fuerza que esto no sería ético*
- *¡O siempre que el modelo pueda validarse adecuadamente, no debería importar si se trata de una caja negra o de un modelo simple e interpretable?*

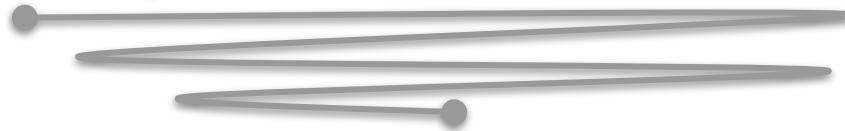
## Predicción vs. Explicación



### Modelado predictivo

- Interés principal. Generar **predicciones** precisas
- Interés secundario. Interpretar el modelo y comprender por qué funciona
- Realidad. A medida que avanzamos hacia una mayor precisión, los modelos se vuelven más complejos y su interpretabilidad se vuelve más compleja
- Aparece siempre el trade-off (“tire y afloje”) que tenemos cuando la precisión de la predicción es el objetivo principal

## *Construyendo modelos a partir de los datos*

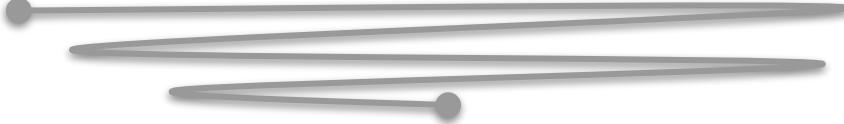


*¿Qué harían para responder las siguientes preguntas?*

- *¿Existe una relación entre la edad y los niveles de colesterol?*
- *¿Los pacientes en un ensayo clínico que toman un fármaco tienen mejores resultados en comparación con los pacientes que toman un placebo?*

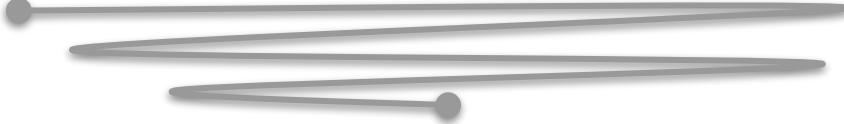
*Una posibilidad es tratar de cuantificar las relaciones entre las variables involucradas*

## *Construyendo modelos a partir de los datos*



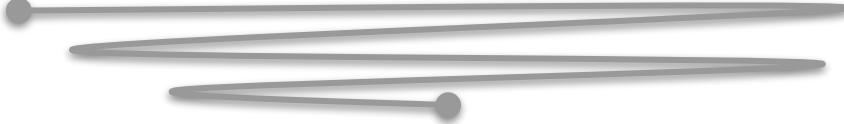
- La manera formal de describir, codificar y probar si una o más variables se relacionan con otras y cómo se relacionan es construir y evaluar modelos a partir de los datos
- Estos modelos describen relaciones importantes en los datos, incluida la fuerza y la dirección (positiva o negativa) de la relación
- Los modelos pueden codificar relaciones lineales y no lineales en los datos

## *Construyendo modelos a partir de los datos*



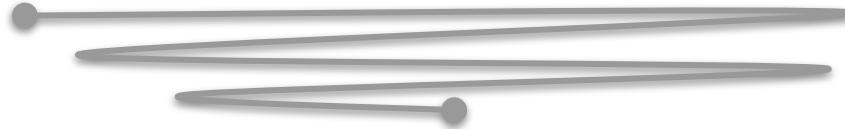
- Se pueden utilizar para confirmar una hipótesis sobre las relaciones  
(Ej. tiempo de reacción)
- Nos ayudan a resumir y comprender los datos
- Sin embargo, una de las aplicaciones más utilizadas de un modelo es la realización de predicciones (Ej. SPAM)

## Construyendo modelos a partir de los datos



- La variable a predecir a menudo se denomina "**variable a explicar**", "**variable y**" o "**variable de respuesta**"
- Las variables que se utilizarán para predecir esta respuesta se denominan "**variables explicativas**" o "**variables x**"
- Los modelos creados para predecir variables categóricas (como una variable binaria o una variable nominal) se denominan "**modelos de clasificación**"
- Los modelos que predicen variables continuas se denominan "**modelos de regresión**"
- Hay muchas formas de generar modelos de clasificación y regresión (vamos a ver algunos en la segunda parte de la materia)

## Cierre



1. Realizamos el diseño de una experiencia/estudio para responder una pregunta de investigación (*Tiempo de reacción*) y vimos que no es una tarea menor
2. Construimos un modelo de agrupamiento
3. Introdujimos el problema de predicción vs. explicación