

Data Mining aplicado a Ciencia y Tecnología

Pablo Riera (pablo.riera@gmail.com)

Juan E. Kamienkowski (juank@dc.uba.ar)

Nosotros

Pablo Riera (pablo.riera@gmail.com). Lic. y Dr. en Física - UBA

Procesamiento de sonido, música y habla. Acústica y psicoacústica. Deep learning.

Juan E. Kamienkowski (juank@dc.uba.ar). Lic. y Dr. en Física - UBA

Neurociencia. Visión. Movimientos oculares. Neurofisiología. Modelos computacionales.



Ustedes?

background? ...

python? ...

R? ...

git? ...

latex? ...

Un poco de filosofía

Data Mining aplicado a Ciencia y Tecnología

→ Clustering

→ Grafos

→ Música

→ Neurociencia

Papers!!

Un poco de burocracia

TP1. Clustering + Música

Seminario paper 1

TP2. Grafos + Neurociencia

Seminario paper 2

Examen

Asistencia

Un poco de burocracia

CRONOGRAMA - DATA MINING EN CIENCIA Y TECNOLOGÍA - 2019

	Martes	TEMA	Laboratorio	Viernes	TEMA	Laboratorio
0	13/08	Data mining y Ciencia – Ciencia de Datos en Música + Tema de TP		16/08	Data mining y Ciencia – Ciencia de Datos en Música + Tema de TP	
1	20/08	Estándares en Ciencia de Datos	git + latex	23/08	Estándares en Ciencia de Datos	git + ...
2	27/08	Introducción a aplicaciones en el tema del TP		30/08	Introducción a aplicaciones en el tema del TP	
3	03/09	Temas especiales de clustering (distancias)		06/09	Temas especiales de clustering (distancias)	
3	10/09	Temas especiales de clustering, parte 2 (validación interna y externa de clusters)	Clustering	13/09	Temas especiales de clustering, parte 2 (validación interna y externa de clusters)	
3	17/09	Análisis crítico y discusión sobre paper de medidas de validación externa. Temas especiales de clustering, parte 3 (métodos: PAM, difuso, por densidad). Presentación del TP1.		20/09	Análisis crítico y discusión sobre paper de medidas de validación externa. Tema clustering, parte 3 (métodos: PAM, difuso, por densidad). Presentación del TP2.	
4	24/09	Herramientas de estadística para análisis de datos	Clustering	27/09	Herramientas de estadística para análisis de datos	
Pre-Entrega TP1						
4	01/10	Visualización de datos y reducción de dimensionalidad. Mapas autorganizativos.	Consultas	04/10	Visualización de datos y reducción de dimensionalidad. Mapas autorganizativos.	
6	08/10	Visualización de datos y reducción de dimensionalidad. Mapas autorganizativos.		11/10	Visualización de datos y reducción de dimensionalidad. Mapas autorganizativos.	
Entrega TP1						


5	15/10	Análisis de grafos (introducción)		18/10	Análisis de grafos (introducción)	
5	22/10	Análisis de grafos, parte 2 (medidas de centralidad, caracterización global). Análisis de grafos, parte 3 (grafos al azar, búsqueda de comunidades)		25/10	Análisis de grafos, parte 2 (medidas de centralidad, caracterización global). Análisis de grafos, parte 3 (grafos al azar, búsqueda de comunidades)	
5	29/10	Laboratorio de análisis de grafos con R. Devolución TP1. Presentación TP2	Grafos	01/11	Laboratorio de análisis de grafos con R. Devolución TP1. Presentación TP2	Grafos
5	05/11	Tutorial sobre manejo de datos de grafos con R (si falta dar algo). Análisis crítico y discusión sobre paper de redes.	Grafos	08/11	Tutorial sobre manejo de datos de grafos con R (si falta dar algo). Análisis crítico y discusión sobre paper de redes.	Grafos
6	12/11	Herramientas de estadística para análisis de datos.		15/11	Herramientas de estadística para análisis de datos.	
7	19/11	Mining de datos geográficos. Parte 1		22/11	Mining de datos geográficos. Parte 1	
7	26/11	Mining de datos geográficos. Parte 2	Consultas	29/11	Mining de datos geográficos. Parte 2	Consultas
03/12 Examen final						
	10/12	Cierre general o consulta	Consultas	13/12	Cierre general o consulta	Consultas
	15/12	Entrega TP2				


Data Mining aplicado a Ciencia y Tecnología

Pablo Riera (pablo.riera@gmail.com)



Juan E. Kamienkowski (juank@dc.uba.ar)

Aplicaciones

- Minería de Texto
 - Clasificar o categorizar documentos
 - Análisis de encuestas
 - Procesamiento automático de mensajes
 - Construcción de Ontologías
 - Buscadores
- Patrones de evolución de Enfermedades
- **Análisis de sonido** 
- Astronomía
- Imágenes satelitales
- Economía

- Patrones de uso / Perfiles
- Redes sociales
- Neurociencia
 - Neuroimágenes
 - **Conectividad** 
- Biología
 - Patrones en estructuras moleculares
 - Alineamiento de secuencias
 - Expresiones de genes

Aplicaciones

- Minería de Texto
 - Clasificar o categorizar documentos
 - Análisis de encuestas
 - Procesamiento automático de mensajes
 - Construcción de Ontologías
 - Buscadores
- Patrones de evolución de Enfermedades
- **Análisis de sonido** 
- Astronomía
- Imágenes satelitales
- Economía
- Patrones de uso / Perfiles
- Redes sociales
- Neurociencia
 - Neuroimágenes
 - **Conectividad** 
- Biología
 - Patrones en estructuras moleculares
 - Alineamiento de secuencias
 - Expresiones de genes

Diferentes problemas pueden tratarse con técnicas emparentadas, con algunas especificaciones propias de cada dominio.

Desde el punto científico ¿Cuál debería ser el objetivo del modelado?

¿Describir la realidad?

¿Ordenar conocimientos en un cuerpo coherente?

¿Hacer predicciones?

Desde el punto científico ¿Cuál debería ser el objetivo del modelado?

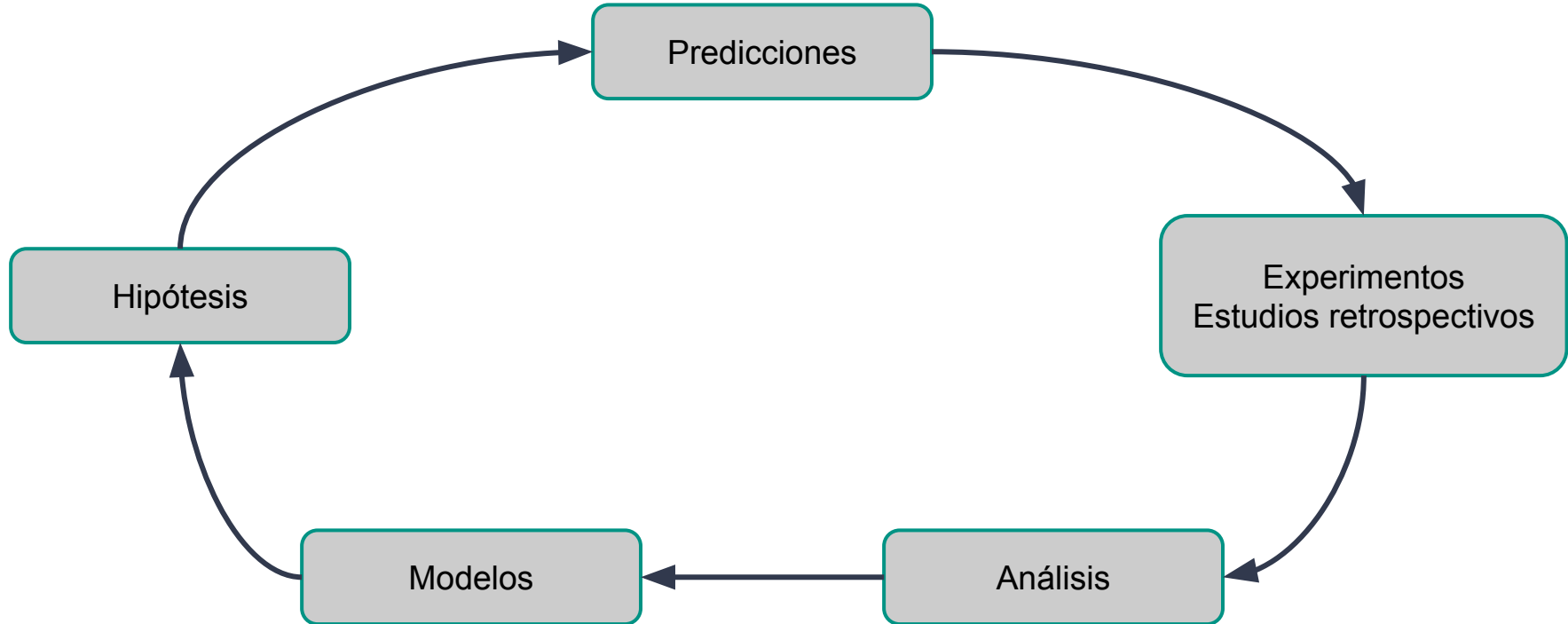
¿Describir la realidad?

¿Ordenar conocimientos en un cuerpo coherente?

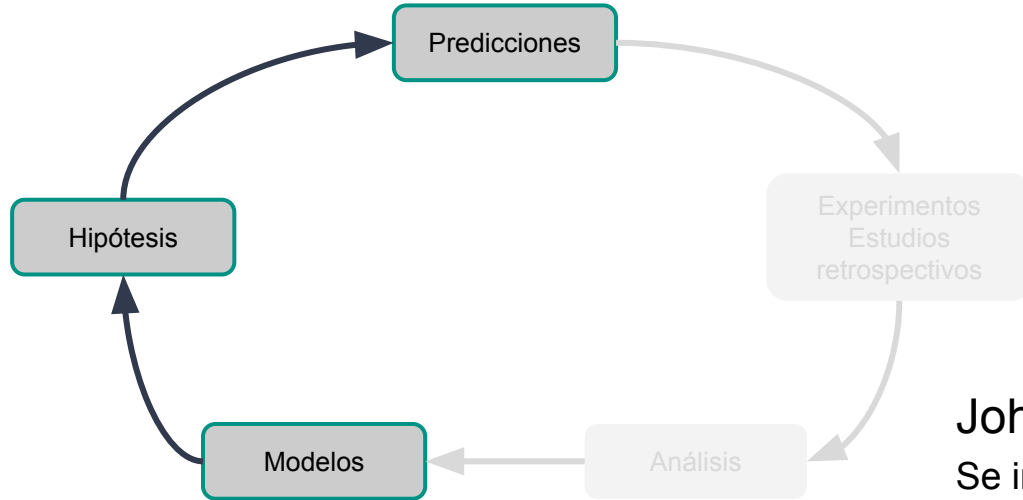


¿Hacer predicciones?

En realidad, el proceso es continuo....



Los orígenes en Astronomía

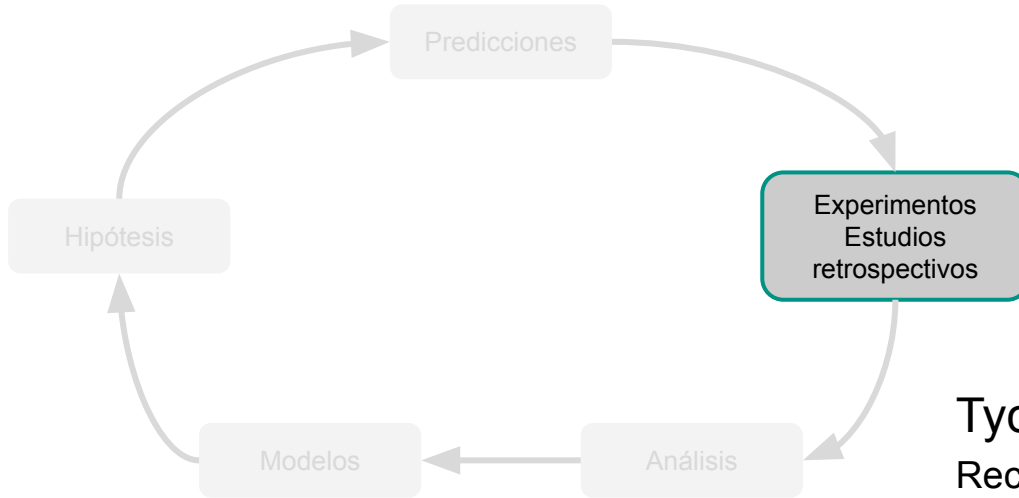


Johannes Kepler (1571-1630)

Se inicia con el sistema de Ptolomeo, luego aprendí el sistema Copernicano.

Propone una modificación al sistema Copernicano, en la cual las órbitas de los planetas estaban sobre 5 esferas (modelo).

Los orígenes en Astronomía

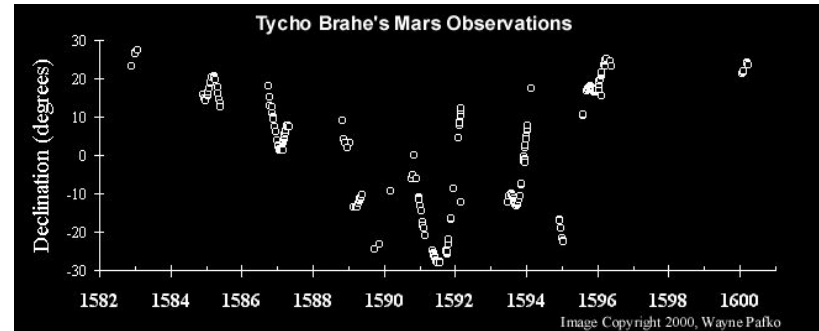
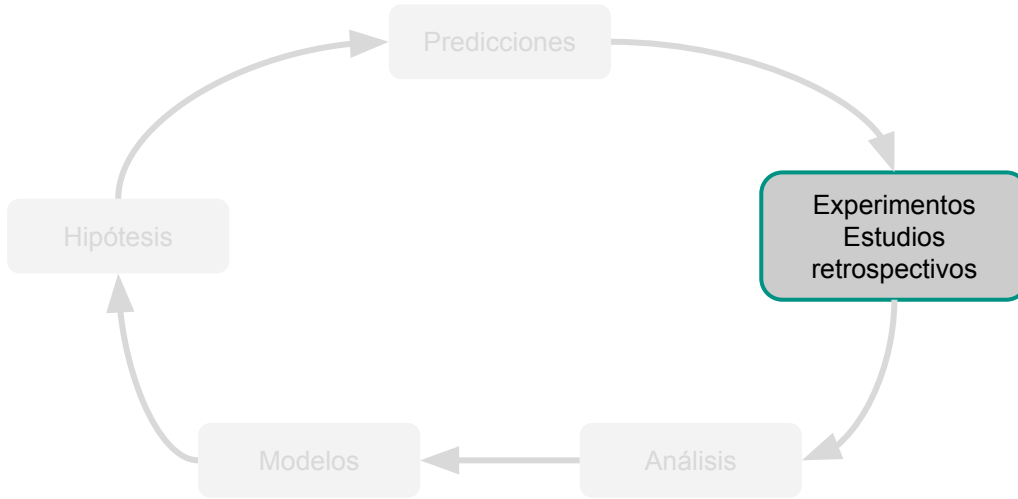


Tycho Brahe (1546 -1601)

Reconoce que los datos astronómicos del momento eran de muy mala calidad, decide recolectar datos de buena calidad (sin telescopio).

Los orígenes en Astronomía

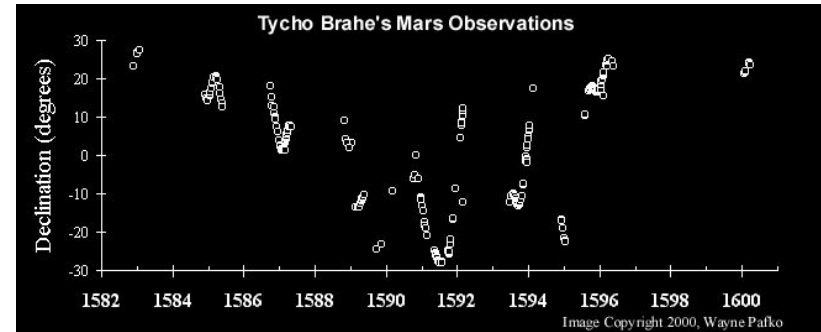
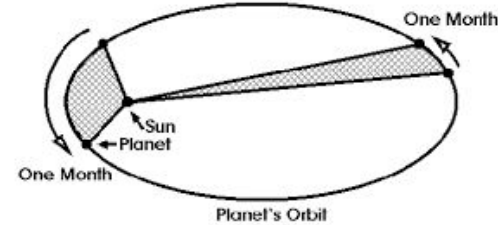
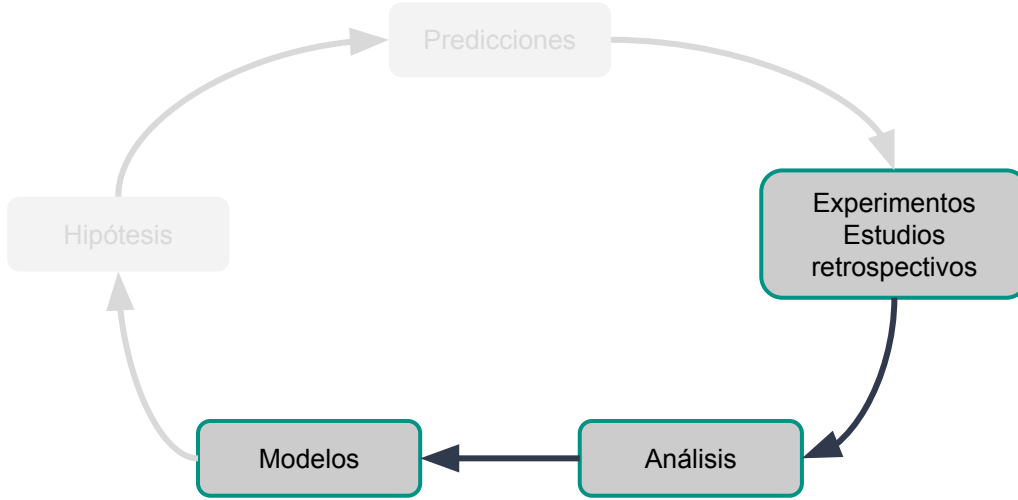
Tycho Brahe (1546 -1601)



A lo largo de 25 años recolectó las mediciones más precisas de su época de las posiciones de los planetas conocidos y muchas estrellas (sin telescopio).

Los orígenes en Astronomía

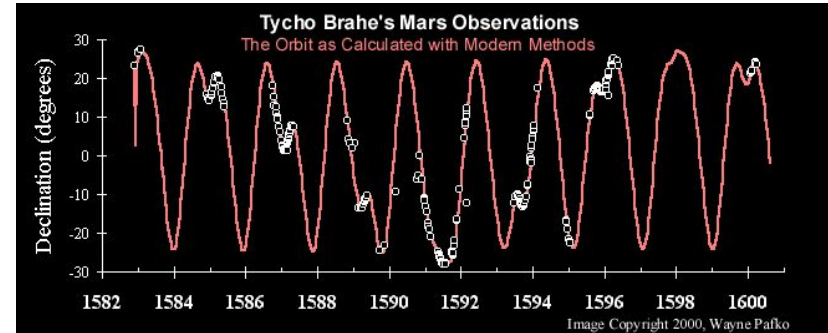
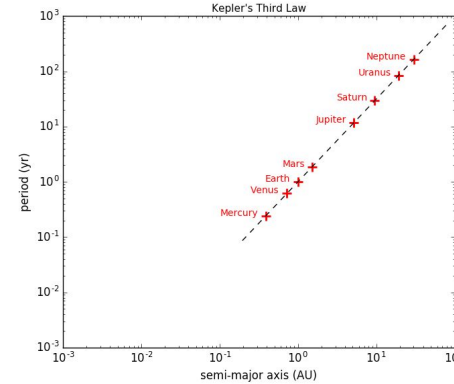
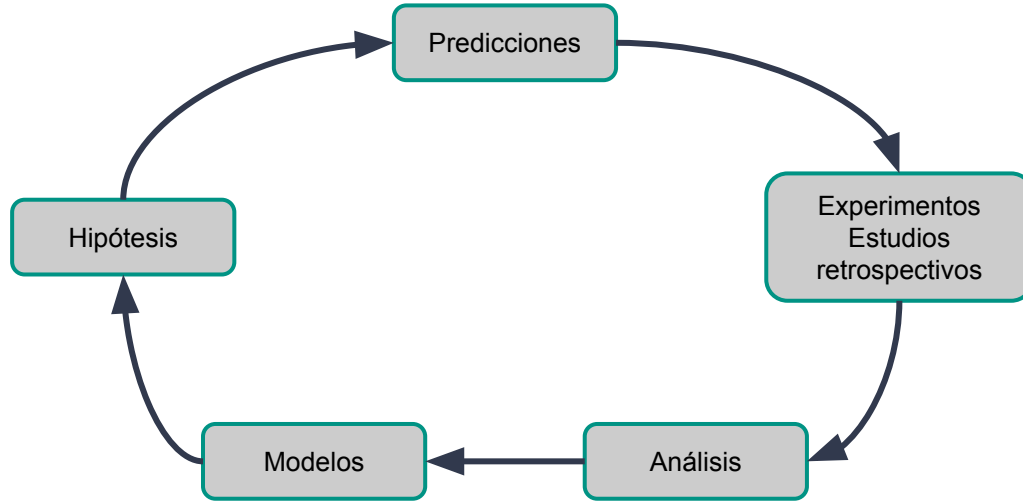
Johannes Kepler (1571-1630)



Utilizó los datos de la órbita de Marte que Brahe había recopilado para crear un modelo sobre el movimiento de los planetas

Los orígenes en Astronomía

Johannes Kepler (1571-1630)



- Los planetas siguen órbitas elípticas, con el sol en uno de sus focos.
- Los planetas barren áreas iguales en tiempos iguales.
- Los cuadrados de los periodos son proporcionales al cubo de las distancias medias al sol.

Astronomía y Data Mining hoy

Objetivos:

- Explotar los datos / Generar nuevos descubrimientos.
- Establecer vínculos entre la teoría astrofísica y los datos / evaluar predicciones.

Múltiples instrumentos distintos, basados en tierra y en el espacio, recolectan datos constantemente.



Se obtienen datos multidimensionales (de distinto tipo y calidad).



Los datos se ordenan de acuerdo a sus coordenadas, al momento de su captura y, si es posible, se registra el objeto involucrado (etiquetas).

Astronomía y Data Mining hoy

- **Caracterizar los objetos conocidos:** métodos de aprendizaje no supervisados, clustering.
- **Asignar los nuevos objetos a clases ya existentes:** métodos de aprendizaje supervisados, clasificación.
- **Descubrir objetos de clases desconocida:** aprendizaje semi-supervisado, detección de outliers.
- **Desarrollar nuevos algoritmos, adaptar los existentes.**

Múltiples instrumentos distintos, basados en tierra y en el espacio, recolectan datos constantemente.

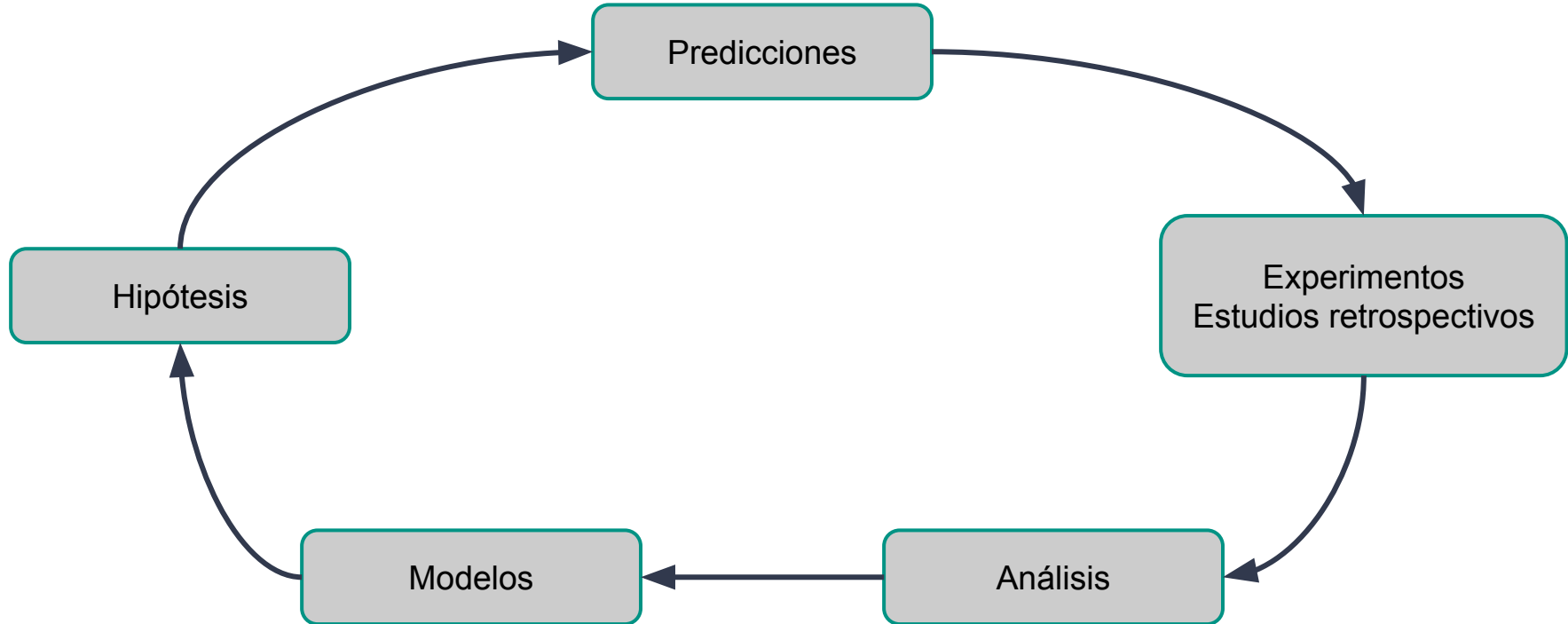


Se obtienen datos multidimensionales (de distinto tipo y calidad).

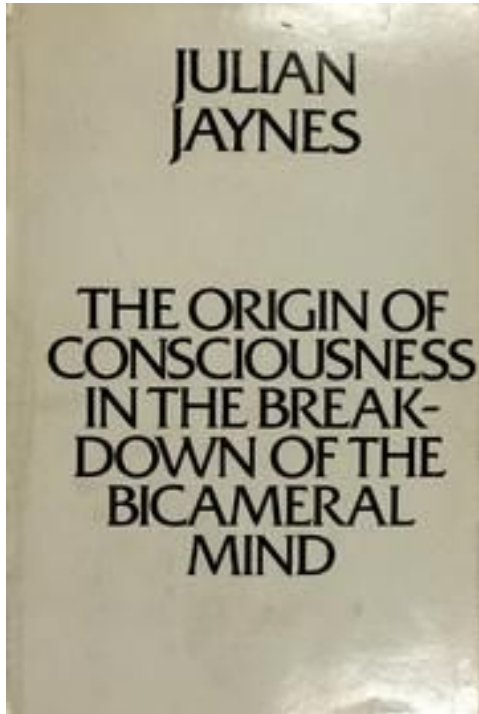


Los datos se ordenan de acuerdo a sus coordenadas, al momento de su captura y, si es posible, se registra el objeto involucrado (etiquetas).

En realidad, el proceso es continuo....

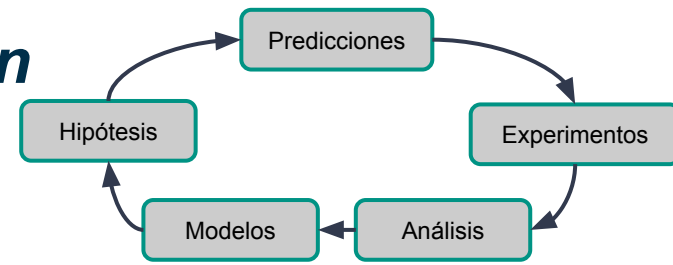


A quantitative philology of introspection



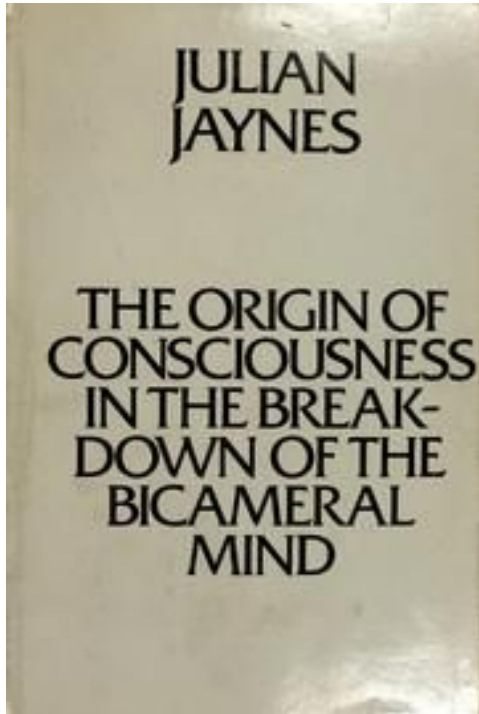
J. Jaynes, *The origins of consciousness in the breakdown of the bicameral mind*, Mariner Books, 2000.

Hipótesis: “*Un pasado donde la mente contaba con funciones cognitivas separadas en una zona parlante (alucinaciones acústicas) y otra que escuchaba y obedecía*”



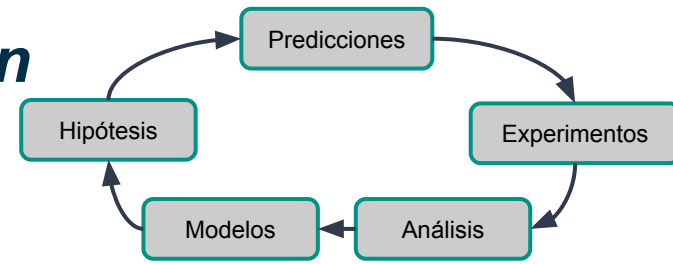
Diuk CG, et al “A quantitative philology of introspection” **Front. Int. Neurosci.** 2012.

A quantitative philology of introspection



J. Jaynes, *The origins of consciousness in the breakdown of the bicameral mind*, Mariner Books, 2000.

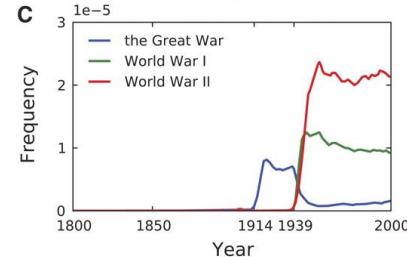
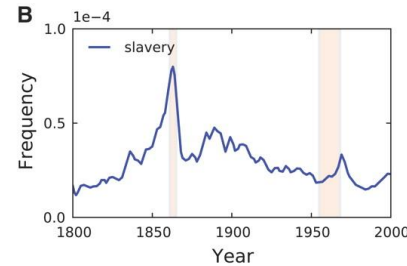
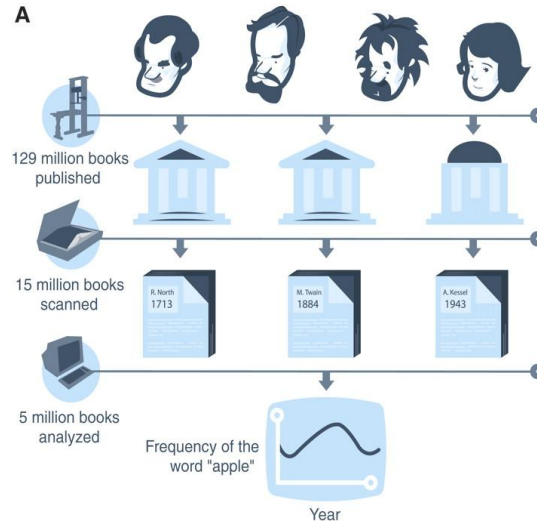
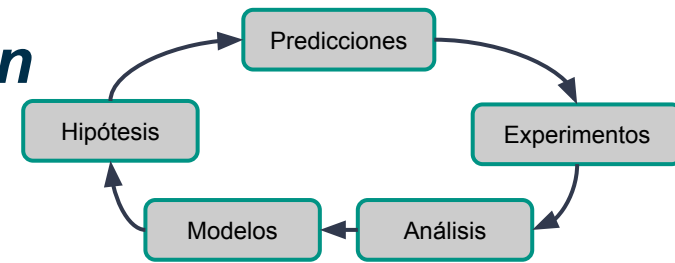
Predicción: “*En la literatura se debería observar un aumento monotónico del concepto de introspección*”



Diuk CG, et al “A quantitative philology of introspection” **Front. Int. Neurosci.** 2012.

A quantitative philology of introspection

Experimento: ¿Cómo aparece la palabra introspección en los textos a través de la historia?

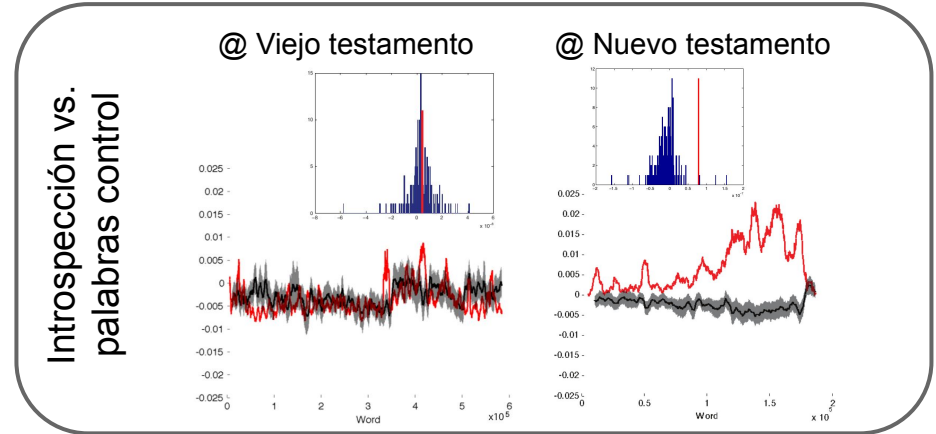
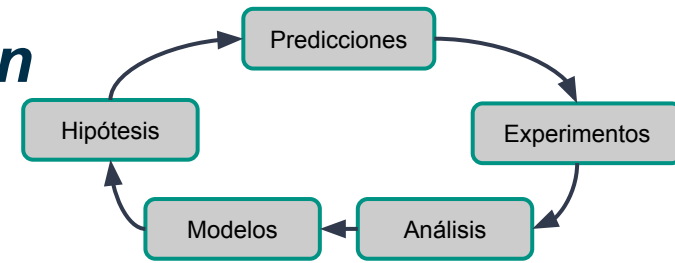
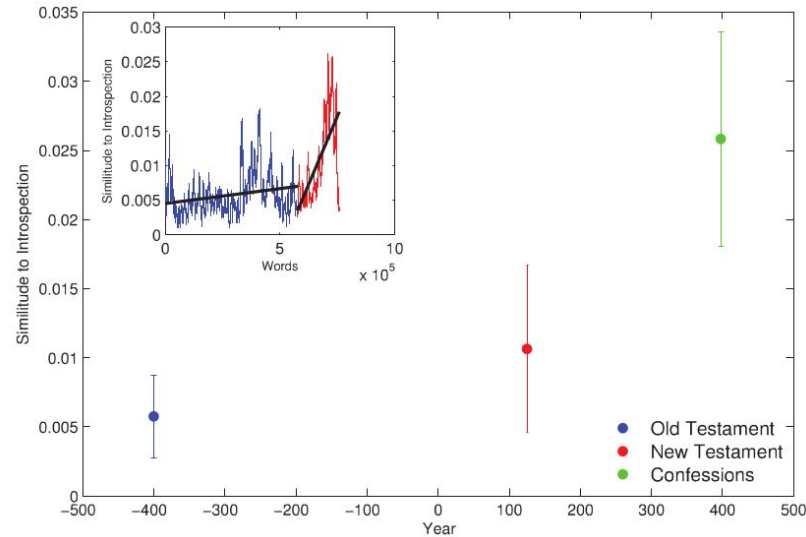


Michel, J.B. et. "AI, Quantitative analysis of culture using millions of digitized books", **Science**, 2011.

Diuk CG, et al "A quantitative philology of introspection" **Front. Int. Neurosci.** 2012.

A quantitative philology of introspection

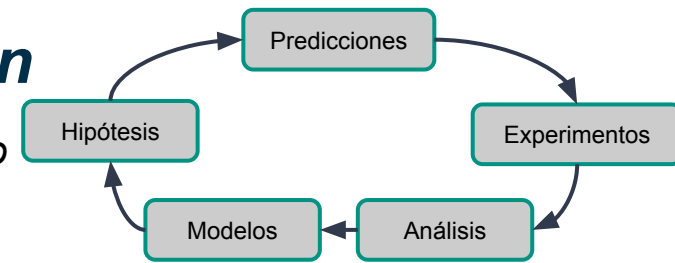
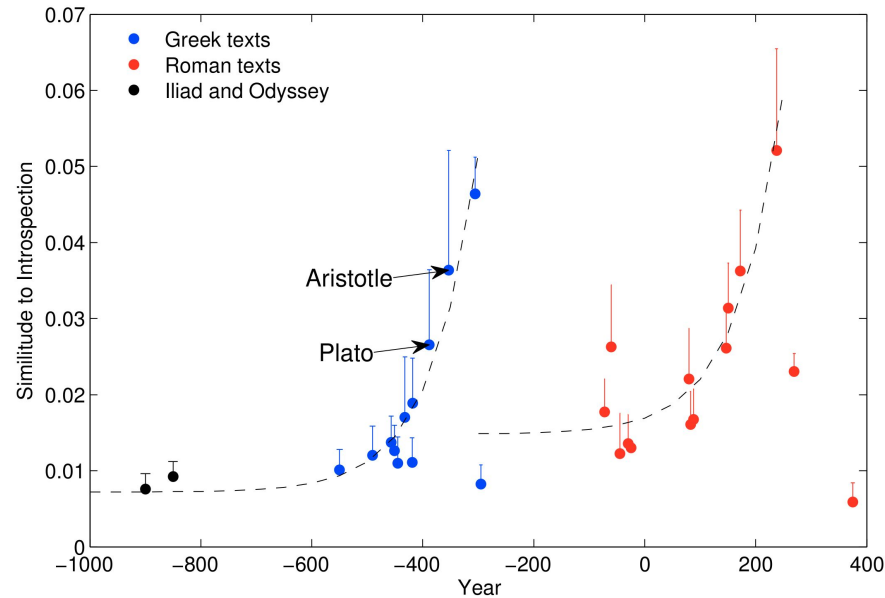
Análisis: ¿Cómo aparecen *palabras relacionadas a introspección* en los textos a través de la historia? (respecto a un conjunto de palabras control)



Diuk CG, et al "A quantitative philology of introspection" **Front. Int. Neurosci.** 2012.

A quantitative philology of introspection

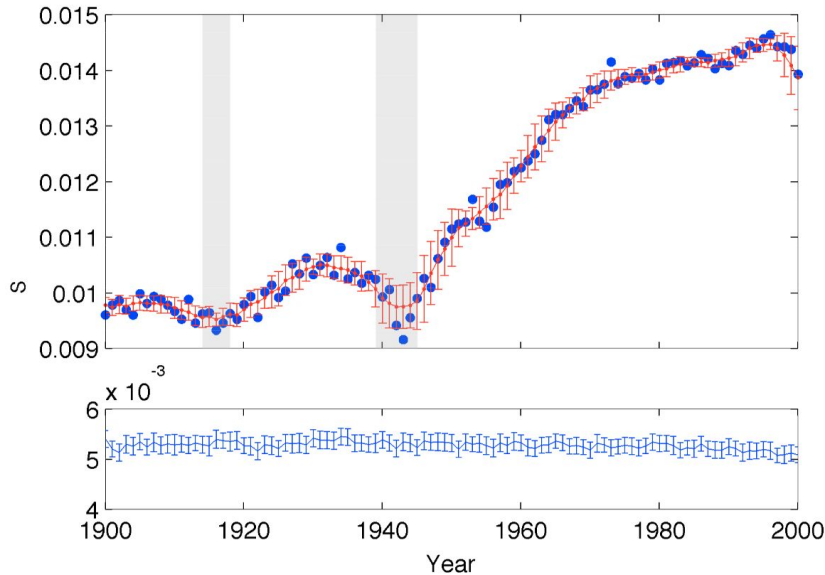
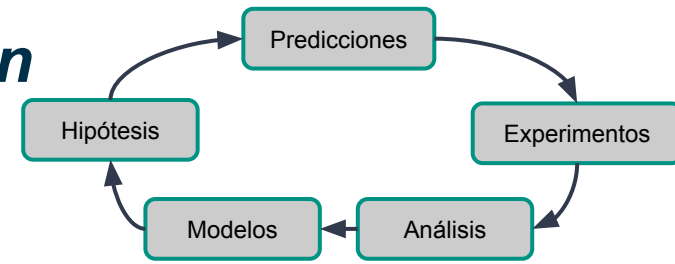
Modelo: “La introspección o conciencia es la evolución cultural (no de toda la humanidad)”



Diuk CG, et al “A quantitative philology of introspection” **Front. Int. Neurosci.** 2012.

A quantitative philology of introspection

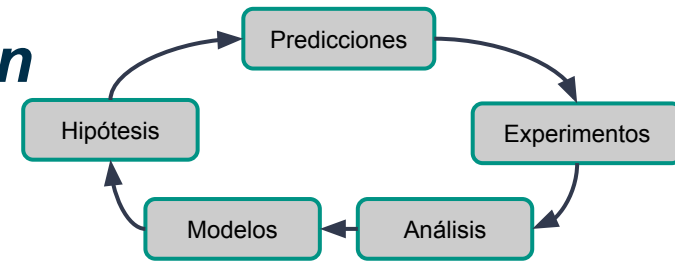
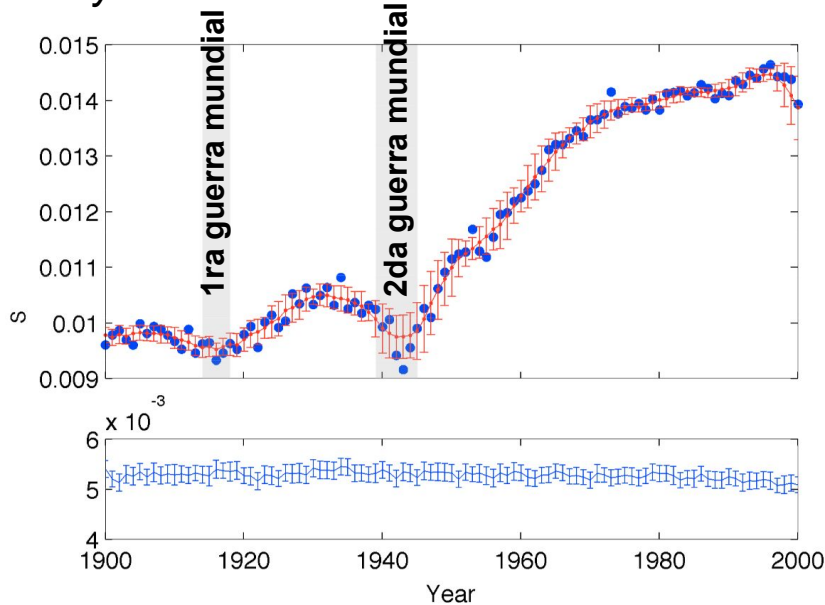
Predicciones: “En la cultura moderna se debería ver un aumento monótonico en el concepto de la introspección, con caídas en las mayores catástrofes culturales”



Diuk CG, et al “A quantitative philology of introspection” **Front. Int. Neurosci.** 2012.

A quantitative philology of introspection

Predicciones: “En la cultura moderna se debería ver un aumento monótonico en el concepto de la introspección, con caídas en las mayores catástrofes culturales”



Diuk CG, et al “A quantitative philology of introspection” **Front. Int. Neurosci.** 2012.

En realidad, el proceso es continuo....

