

# MANEJO E INTERPRETACIÓN DE DATOS EN QUÍMICA

---

Universidad Industrial de Santander

Escuela de Química

Código: 27661

# Introducción & background

## Enrique



- **1994: Químico Universidad Industrial de Santander (UIS)**
- 1997: Magister Química UIS
- 2001 – 2004: Doctorado UNAM
- 2004 – 2010: Profesor Asociado UIS
- 2010 – presente: Profesor Titular UIS

**Consultas:** Lunes 3:00 a 5:00 virtual, Martes 3:00 a 5:00  
oficina (CT 302)

# Grupo

## 1 Minutos Biografía:

- Su nombre completo
- Semestre que cursa y programa / área
- Background Técnico
  - Lenguaje de programación
  - Cursos de estadística

## Preguntas:

- Qué espera de este **curso**?
- Tiene algún problema que podríamos solucionar en el curso?

# Recursos: cursos en la web



<https://www.datacamp.com/>

# Recursos: tutoriales, mini-cursos, etc.



**Medium**  
Daily Digest

<https://medium.com/>

# Recursos: tutoriales, mini-cursos, etc.



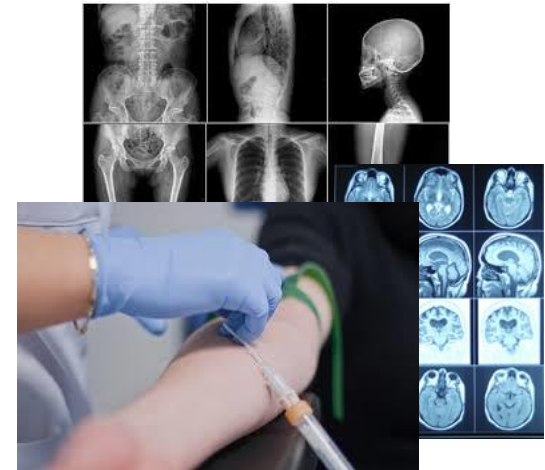
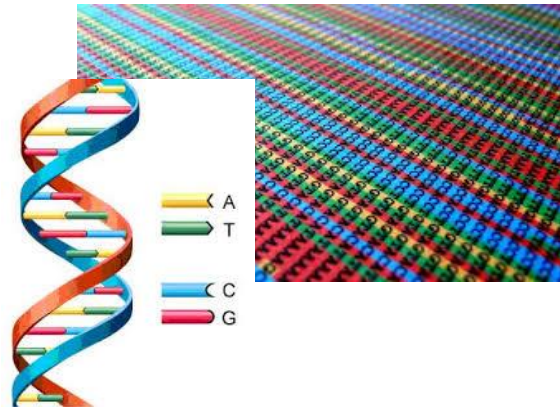
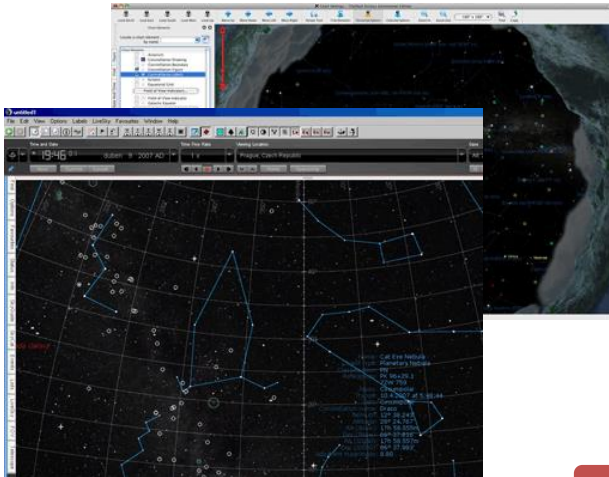
kaggle



<https://www.kaggle.com/>



# Nuestro mundo gira en torno a los datos



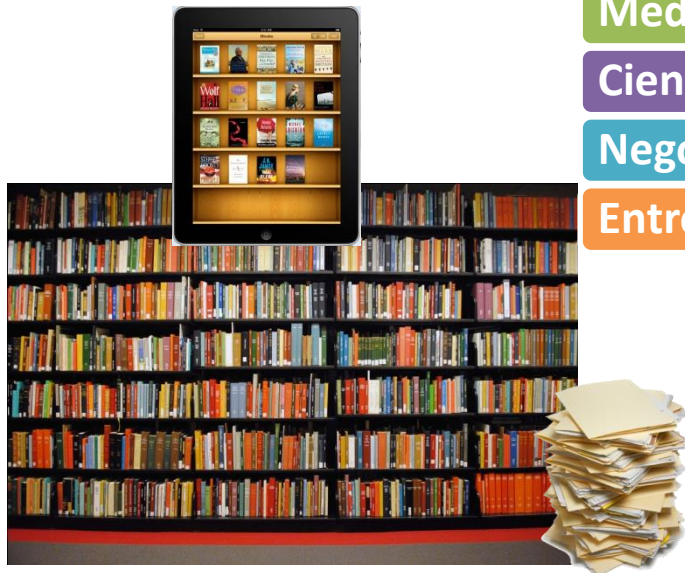
Ciencia

Medicina

Ciencias Sociales y Humanidades

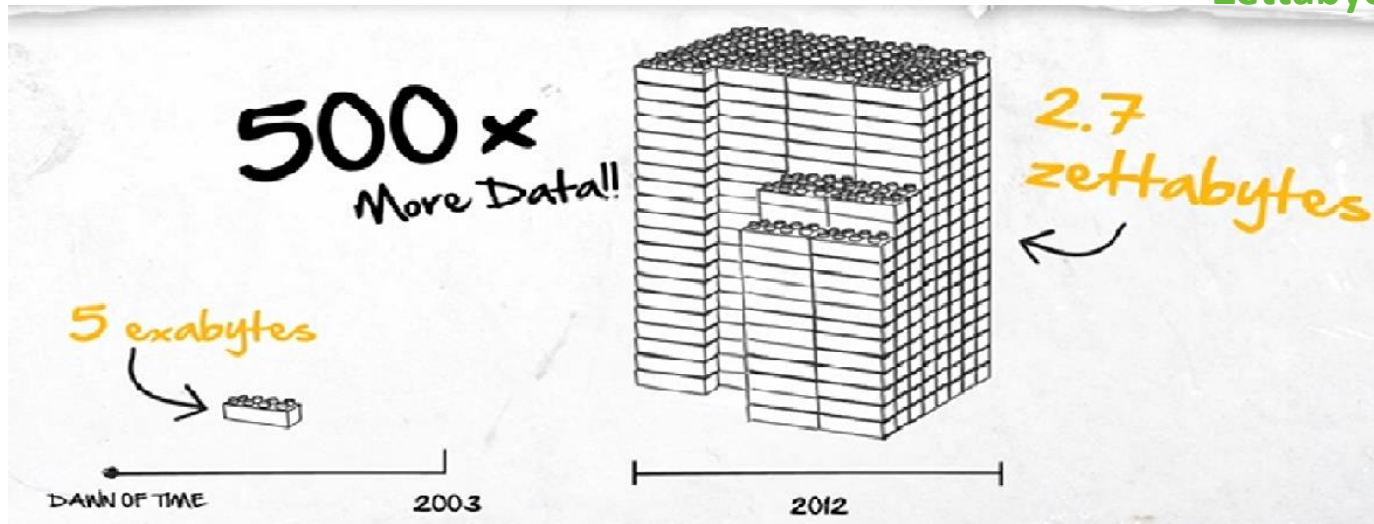
Negocio, Comercio e Industria

Entretenimiento y Ocio



# La explosión de los datos

Zettabytes:  $10^{21}$  bytes.



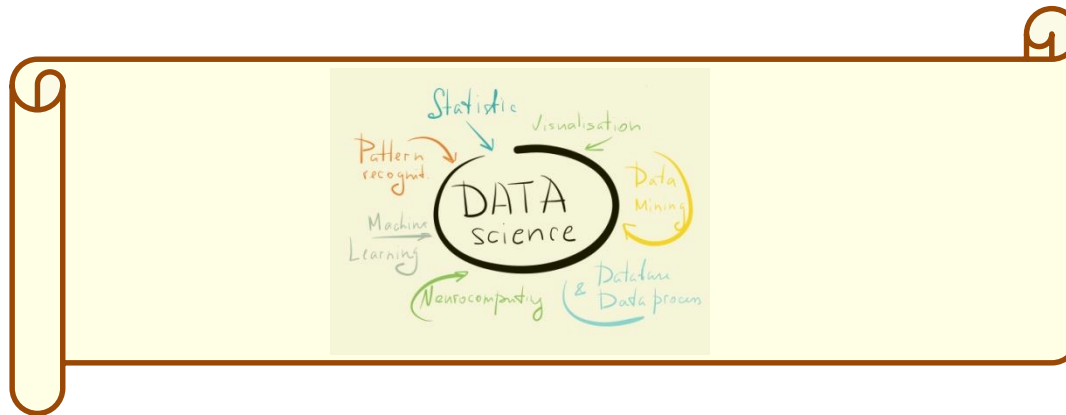


# ¿Qué hacemos con estos datos?

El progreso y la innovación ya no se ven obstaculizados por la capacidad de recopilar datos, sino por la capacidad de

**Extraer el valor de los datos**

**Convertir datos en conocimiento**





# Datos: una definición

Un dato es la representación de una variable que puede ser cuantitativa o cualitativa que indica un valor que se le asigna a las cosas y se representa a través de una secuencia de símbolos, números o letras.



*Número de matrículas e inscripciones en universidades pública frente a privadas, etc.*

# Datos: una definición

Cada variable puede ser independiente o dependiente:

- Una variable independiente (iv) no está controlada o afectada por otra variable (por ejemplo, el tiempo en un conjunto de datos de series temporales)
- Una variable dependiente (vd) se ve afectada por la variación de una o varias variables independientes asociadas (por ejemplo, la temperatura de una región)

Un set de datos también contiene un conjunto de observaciones (también llamadas registros) sobre estas variables.



# Datos: una definición

El concepto de dato está estrechamente ligado al de información, pero existe una diferencia fundamental entre ambos términos. Mientras los datos refieren a eventos o hechos registrados, la información está constituida por aquellos datos brutos que son procesados de manera tal que generen contenido que puede ser conocido e interpretado por los usuarios.

Los datos no tienen sentido por sí mismos, pero al ser procesados y contextualizados se convierten en información certera y disponible para conocer un fenómeno, tomar decisiones o ejecutar acciones.

# Tipos de datos

- **Nominales**
- Ordinales
- Escalas / Cuantitativos
  - Relaciones
  - Intervalos

No ordenados {...}

Valores no-numéricos

Por ejemplo:

- Datos Categóricos (finitos)
  - {manzanas, naranjas, peras}
  - {roja, verde, azul}
- Datos Arbitrarios (infinitos)
  - {"12 Main St. Boston MA", "45 Wall St. New York NY", "Cra 27 #9"...}
  - {"Juan Castro", "Diana Lane", ...}

# Tipos de Datos

- Nominales
- **Ordinales**
- Escalas/Cuantitativa
  - Relaciones
  - Intervalos

## Ordenados

<...>

(conocidos como tuples)

Ejemplos:

- Numéricos: <2, 4, 6, 8>
- Binarios: <0, 1>
- Non-numéricos:  
<G, PG, PG-13, R>

# Basic data types

- Nominales
- Ordinales
- **Escala/Cuantitativos**
  - Relaciones
  - Intervalos

Un **rango numérico** [...]

## Relaciones

- Distancia del “cero absoluto”
- Pueden ser comparados matemáticamente
- Por ejemplo: altura, peso, etc.

## Intervalos

- Elementos numéricos  
Ordenados que pueden ser matemáticamente manipulados, pero no pueden ser comparados
- Ej.: fechas, eventos temporales



# Operaciones Básicas

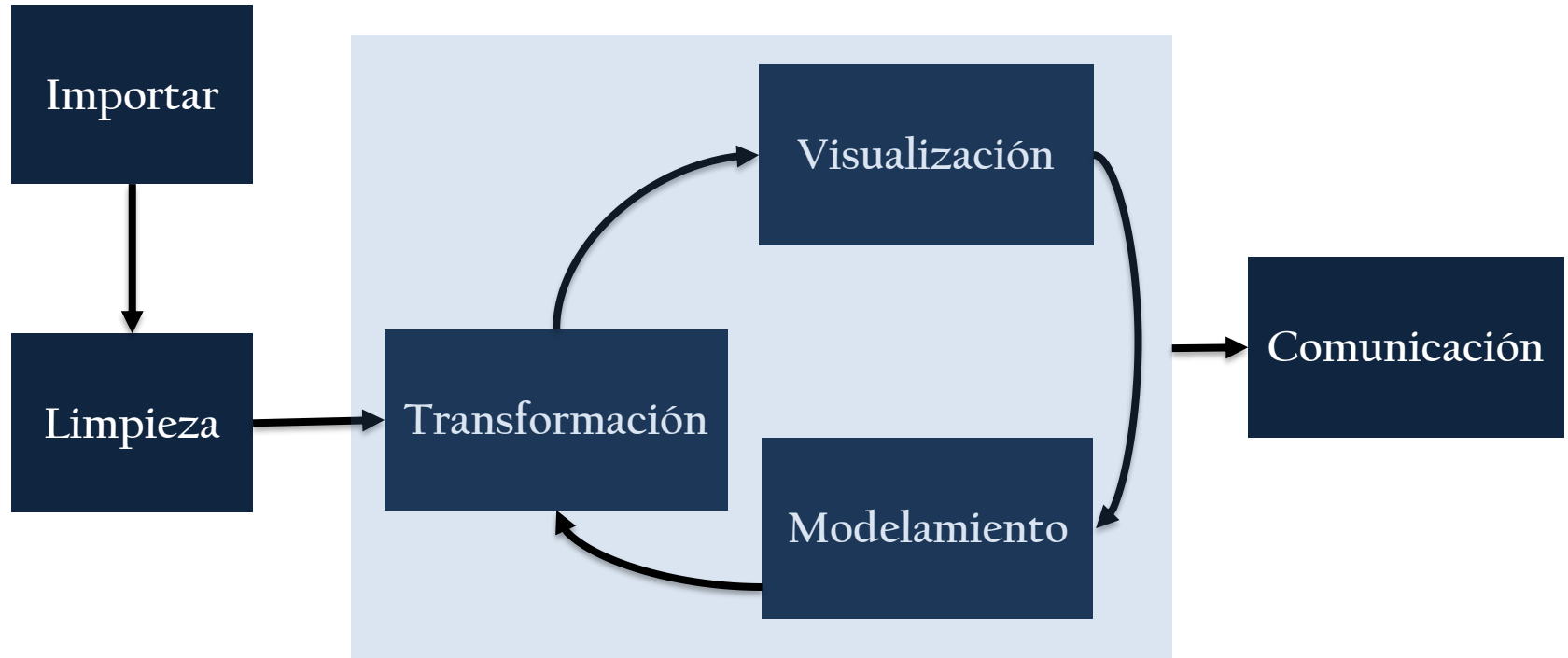
- Nominales (N)
  - Igualdad:  $=$  y  $\neq$
  - Frecuencia: qué tan a menudo aparece  $x$  ?
- Ordinales (O)
  - En Relación con otros:  $>$ ,  $<$ ,  $\geq$ ,  $\leq$
  - Distribución: inferencia sobre la frecuencia relativa
- Cuantitativa (Q)
  - Otras operaciones matemáticas:  $(+, -, *, /, \text{etc.})$
  - Estadística descriptiva: *promedios, desviación estándar, etc.*

# Etapas de la Ciencia de Datos

LEER

EXPLORAR

COMUNICAR



# Contenido

- 1. Instalación de plataformas (Anaconda, Jupyter, Python)
- 2. Introducción a Python (Operaciones básicas, librerías, etc.)
- 3. Lectura de archivo de datos (Importación de datos)
- 4. Lectura de datos; filtrado, selección, ordenado, agrupamiento, re-arreglo, etc.
- 5. Visualización de datos e Inspección

# Contenido

- 6. Pre-procesamiento de datos
- 7. Estadística descriptiva
- 8. Lectura de archivo de datos (Importación de datos)
- 9. Estadística Inferencial: Quimiometría
  - Clasificación
  - Regresión Lineal
  - Regresión Lineal multivariado
  - Selección y muestreo
  - Otros métodos avanzados



# Objetivos del curso



1. Entender  
qué es DS  
(y que no es)



2. Conocer  
fundamentos/  
herramientas



3. Ser capaz  
de escoger  
métodos

# Información General

- Website del curso:

<https://tic.uis.edu.co/ava/course/view.php?id=29946>

- Jupyter:

<https://jupyter.org/>



- Anaconda:

<https://www.anaconda.com/>



- Rstudio:

- <https://rstudio.com/>



# Salvando responsabilidades

## Este curso es un experimento en **construcción**

seguimos la idea de que las personas aprenden más eficazmente cuando construyen cosas personalmente significativas

- Mi trabajo:



# Notas

- Participación (10%):
- Labs (30%):
- 8 (cortos) asignaciones (40%):
- Proyecto de curso (20%)



# Preguntas?

