

# Análisis y visualización de datos

...

Algoritmos y Estructura de Datos III

# Problemas comunes en los TPs

1. Que hacer cuando hay muchas variables
2. Desde y hasta donde experimentar
3. ¿Puedo comparar peras con manzanas ?
4. ¿Qué es mejor caso, peor caso y caso promedio?
5. Como NO mentir cuando mostramos los datos
  - Errores comunes
  - Problemas con las escalas
  - Gráficos poco informativos

# 1 - Que hacer cuando tenemos muchas variables

¿Podemos hacer un gráfico n-dimensional? ... ¿que pasa si tengo más de 4 variables?

En general no es una buena idea un gráfico n-dimensional, son costosos, y muchas veces poco intuitivos o directamente no se pueden graficar.

1. Fijar todas las dimensiones salvo una y plotear el resultado en base a la dimensión de la variable no fijada. Si, nos quedan varios gráficos, o uno con varias líneas.
2. trabajar sobre una función de las variables y graficar respecto de esa función, por ejemplo, de la suma. Dado un grafo con “n” nodos y “m” aristas, en algunos casos se puede considerar el tamaño del mismo como la suma de  $n+m$ . En tal caso, mientras que las muestras estén bien balanceadas, se puede plotear respecto a  $n+m$ . En general, la complejidad del algoritmo es quien indica respecto de qué consiste el tamaño del grafo.

## 2 - Desde y hasta donde experimentar

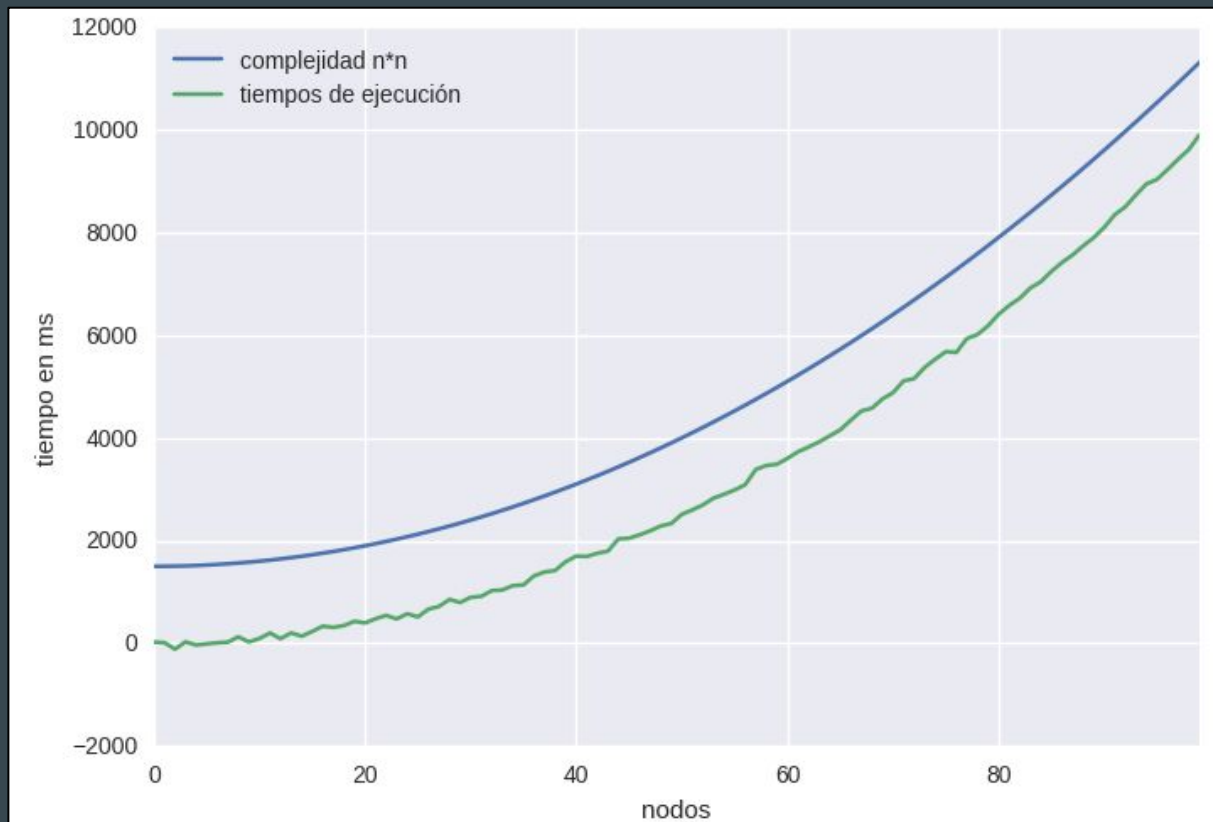
No hay una forma correcta de saber hasta dónde o desde donde experimentar. Un posible approach podría ser ir variando la granularidad con la cual se samplea de forma de tener un panorama amplio antes de lanzarse a medir todo.

Hay que tener cuidado con sacar conclusiones cuando los datos no son suficientemente descriptivos sobre el fenómeno que se está estudiando.

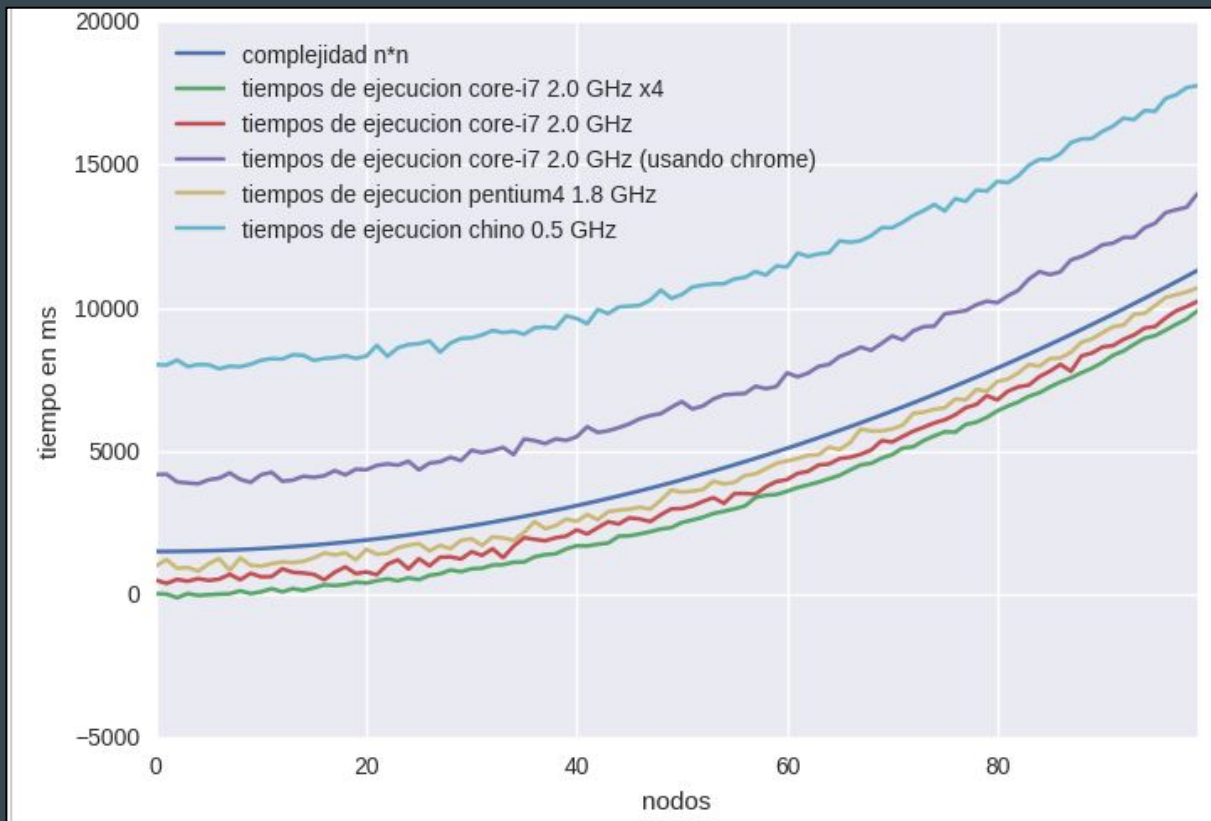
Respecto a los TPs de esta materia hay dos momentos en los que tenemos que tener cuidado de esto:

- Cuántas iteraciones son necesarias para poder sacar conclusiones los tiempos.
- Cuántos datos son necesarios para poder inferir la curva correcta que mejor explica los datos.

### 3 - ¿puedo comparar peras con manzanas?



### 3 - ¿puedo comparar peras con manzanas?



# 4 - ¿Qué es mejor caso, peor caso y caso promedio?

Se llaman mejor, peor y caso promedio a los valores de entrada que cumplen con ciertas propiedades y/o características que garantizan/indican que la cantidad de operaciones realizadas con esos valores son las mínimas, máximas o intermedias respectivamente.

**IMPORTANTE:** Los mejores, peores y casos promedios SIEMPRE se calculan con valores de entrada del MISMO tamaño.

Dado el valor de entrada  $v$  de tamaño fijo ' $n$ ' y el algoritmo  $m$ , entonces:

- $v$  es un mejor caso si cumple que para cualquier otro valor  $v_1$  con el MISMO tamaño ' $n$ ' que  $v$ , el algoritmo  $m$  sobre  $v$  realiza menos o igual cantidad de operaciones que  $v_1$ .
- $v$  es un peor caso si cumple que para cualquier otro valor  $v_1$  con el MISMO tamaño ' $n$ ' que  $v$ , el algoritmo  $m$  sobre  $v$  realiza más o igual cantidad de operaciones que  $v_1$ .
- $v$  es un caso promedio si cumple que para cualquier otro valor  $v_1$  con el MISMO tamaño ' $n$ ' que  $v$ , el algoritmo  $m$  sobre  $v$  tiene probabilidad 0.5 de realizar más o menos cantidad de operaciones que  $v_1$ .

# 5 - Como NO mentir cuando mostramos los datos

Los gráficos son una forma de mostrar los datos que facilitan la comprensión y el análisis de los mismos, suele ser más fácil expresar un fenómeno en gráficos que sobre los datos crudos. Pero también es fácil cometer errores de forma que el análisis de los mismos se vuelva complicado y/o falso.

## Errores comunes

- No mostrar los datos completos
- Comparar datos con escalas muy distintas en un mismo gráfico
- Mostrar distintas escalas



# 5 - Como NO mentir cuando mostramos los datos

## Gráficos poco informativos

Hay muchos tipos de gráficos y cada uno de ellos sirve para mostrar diferentes características de los datos, la pregunta entonces es:

¿que quiero mostrar de estos datos?



# 5 - Como NO mentir cuando mostramos los datos

## Problemas con las escalas

Cuando se está mostrando varios gráficos donde muestran distintas variables de un mismo fenómeno, hay que tener cuidado con estar mostrando siempre las mismas escalas o en caso contrario especificar para que quede bien claro que las escalas cambiaron. Si esto no se realiza correctamente suceden varios fenómenos en la lectura de los datos:

- El lector ingenuo que cree en ustedes solo mira las escalas del primer gráfico y luego mira el resto de los gráficos respecto a esa escala.
- Peor aún, el lector ingenuo compara las variables utilizando como referencia la misma escala para todos
- El lector perspicaz para evitar esto, tiene que volver a graficar en su cabeza los datos para poder analizarlos, en definitiva los gráficos perdieron gran parte de su poder expresivo.