

Tipos de Graficos

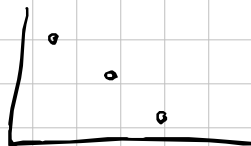
Qualitative data

PIE CHART



→ % total

Dot plot



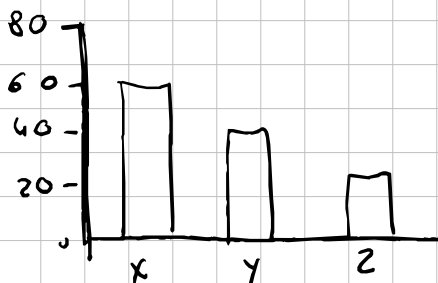
→ Categorias

Quantitative - Bar graph

→ Numbers

→ Distance

...



Histogram

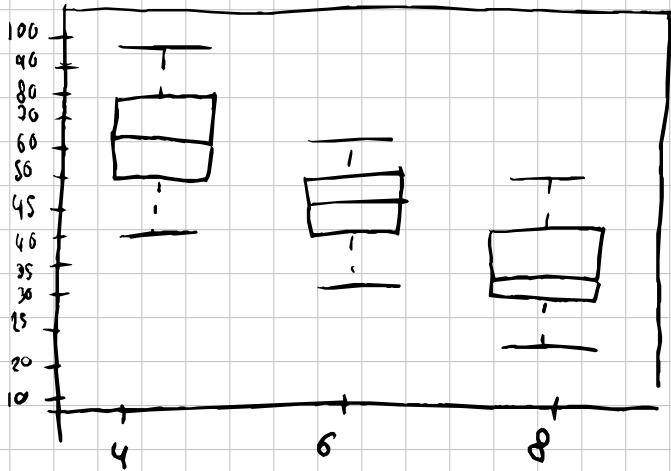
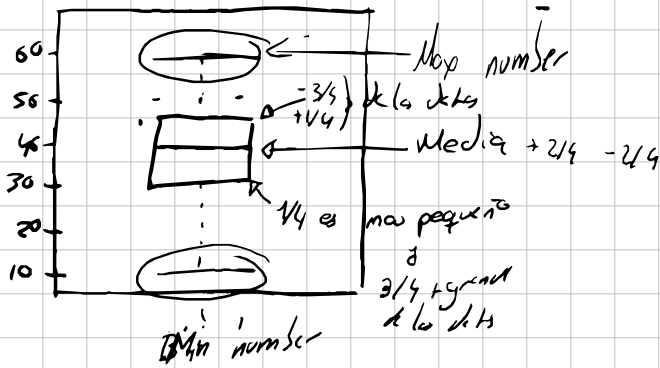
→ Permite adaptar el tamaño de los bloques a la frecuencia

→ Densidad: (altura de la barra)

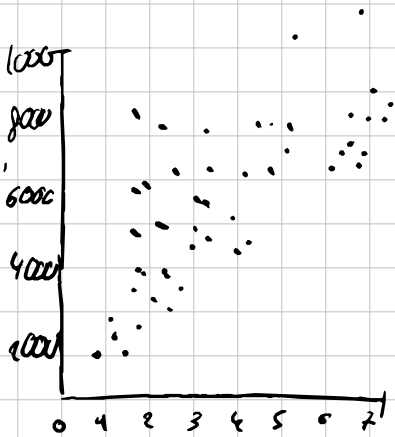
→ Porcentajes:



Box-Plot - Caja y Bigote



Scatterplot - Dispersión



Uso principalmente
para visualizar
la relación entre dos
valores

Como por ejemplo
años de estudio y
salarios

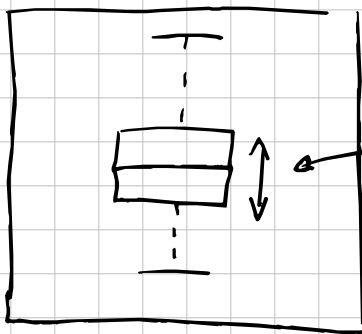
* Es importante dar todo el contexto necesario
, pero tampoco excesivo

Numerical Summary Measures

- ▷ Mean = average = media
- ▷ Median: punto medio, la mitad de los datos más Mediana alto y mitad más bajos.
- ▷ Mode: más repetido

- ▷ 1st Quartile o cuartil: 25% más bajo
- ▷ 2nd Quartile: media
- ▷ 3rd Quartile: 75% más bajo

Percentiles, the five Numbers Summary, and Standard Deviation



Interquartile range
3rd quartile - 1st quartile
it measures how spread the data is

Standard Deviation: Same with the values

\bar{x} average of the numbers $x_0 \dots x_n$

formula
$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Takes the square root of the average of each number minus the average

EXAMING OF HOW TO PRODUCE DATA

STATISTICAL INTERFERENCE

- Pick a random sample of the population to use it to calculate everything
- Population: whole dataset
- Parameter: quantity / value / data we are interested from the population
- Sample - muestra
- Statistic (estimate): parameter solo medido en el sample

Simple Random vs Stratified Random

SAMPLING

► Sample of convenience:

No suele ser una buena manera de sacar datos. Un ejemplo de este tipo sería coger solo clientes de un barrio en lugar de toda España.

Bias. Sample que favorece un resultado concreto.

Selection Bias: sample of convenience hace más probable, uno u otro sobre otro.

Non-responsive bias: la gente que responde puede ser diferente a la que no responde.

voluntary response: generalmente los reviews solo vienen de clientes con o muy buenas o muy malas experiencias.

► Simple random sample

• Select subject without replacement

► Stratified Random Sample

Dividir el population en grupos de topicos similares "strata".

Bias and Chance Error

$$\text{Estimate} = \text{parameter} + \underbrace{\text{Bias}}_{\substack{\text{bias / error} \\ \text{partly caused} \\ \text{by the} \\ \text{sample}}} + \text{chance error}$$

Observation Studies

Son los resultados de comparación de datos
atribuir de una variable aunque no tiene
porque ser directamente la que altera la otra
pero se predice que una variable este ligada
a otra determinada.

Ej. + red meat =? + cancer

Pero realmente no es la carne roja porx es que la
gente que come carne roja suele hacer menos
ejercicio y beber mas alcohol

A este alcohol + ejercicio se le llama confounding variable

Para poder asegurar un vinculo es necesario un
experimento para asegurar su efecto.

Similar a los estudios medicos y los
grupos medicados y de control con el placebo

Un experimento debe ser double-blind ni los pacientes ni los examinadores deben saber
en que grupo esta el paciente

Randomization

▷ \mathcal{R}