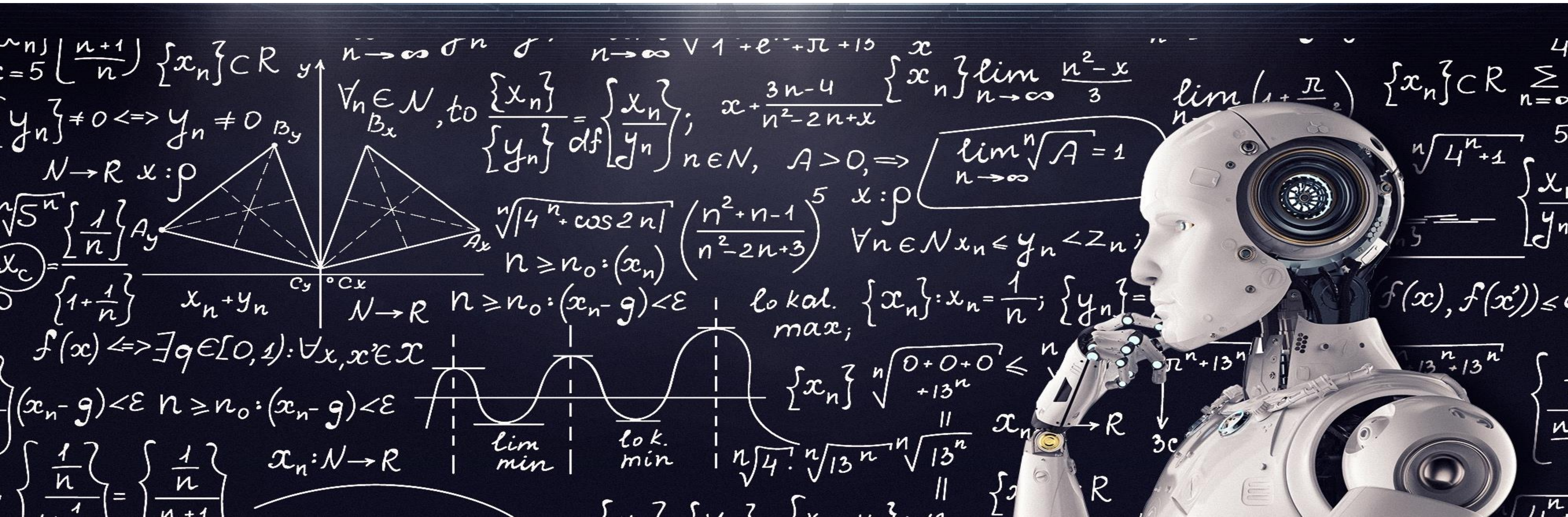


# Análisis multivariado





# Análisis multivariado







# Objetivo

## Análisis multivariado

- Comprender la estadística multivariante, su aplicación y sus diferentes métodos y análisis.
- Resolver un problema de análisis multivariado utilizando python y un conjunto de datos de UCI Machine Learning Repository.







# Bases

- Reglas de asociación ->  $P(Y|X)$
- Métodos de clasificación -> IF income >  $\theta_1$  AND savings >  $\theta_2$  THEN low-risk ELSE high-risk
- Predicción ->  $P(\tilde{Y} = 1|X = x)$
- Detección de outliers
- Regresión ->  $y = wx + w_0$
- Teorema de Bayes ->  $P(C|x) = \frac{P(C)p(x|C)}{p(x)}$
- Funciones discriminantes ->  $g_i(x) = p(x|C_i)P(C_i)$
- Métodos paramétricos
- Maximum Likelihood Estimation ->  $l(\theta|X) \equiv p(X|\theta) = \prod_{t=1}^N p(x^t|\theta)$
- Bernoulli Density , Multinomial Density, Gaussian Density







# Planteamiento del problema

- Tenemos los datos resultado de un análisis químico de vinos cultivados en la misma región de Italia pero derivados de tres cultivos diferentes. El análisis determinó las cantidades de 13 elementos que se encuentran en cada uno de los tres tipos de vinos.
- El objetivo es establecer por medio de técnicas de inteligencia artificial **las variables más influyentes** en la calidad sensorial del vino. [Galeano-Arias 2020]





# Planteamiento del problema

- Los atributos son
- 1) alcohol
- 2) ácido málico
- 3) Ceniza
- 4) Alcalinidad de la ceniza
- 5) magnesio
- 6) fenoles totales
- 7) Flavonoides
- 8) Fenoles no flavonoides
- 9) proantocianinas
- 10) Intensidad de color
- 11) tonalidad
- 12)  $OD_{280}/OD_{315}$  de vinos diluidos
- 13) prolina

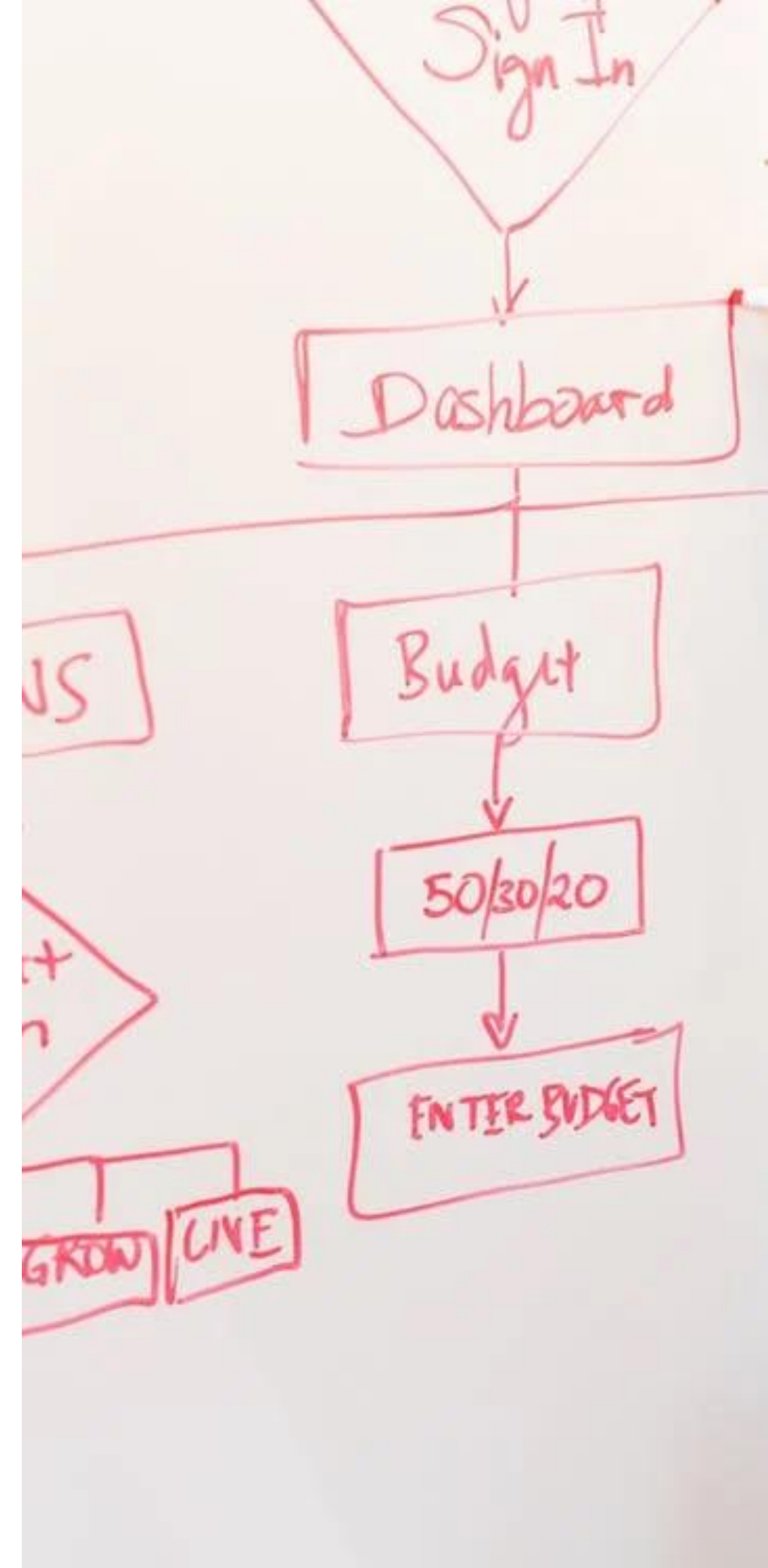






# Desarrollo de ideas

- Análisis químico
- Método robóticos
- Espectroscopia infrarroja
- técnicas de espectroscopia
- PCA
- K-vecinos
- Maquinas de soporte vectorial
- Árboles de clasificación
- Regresión

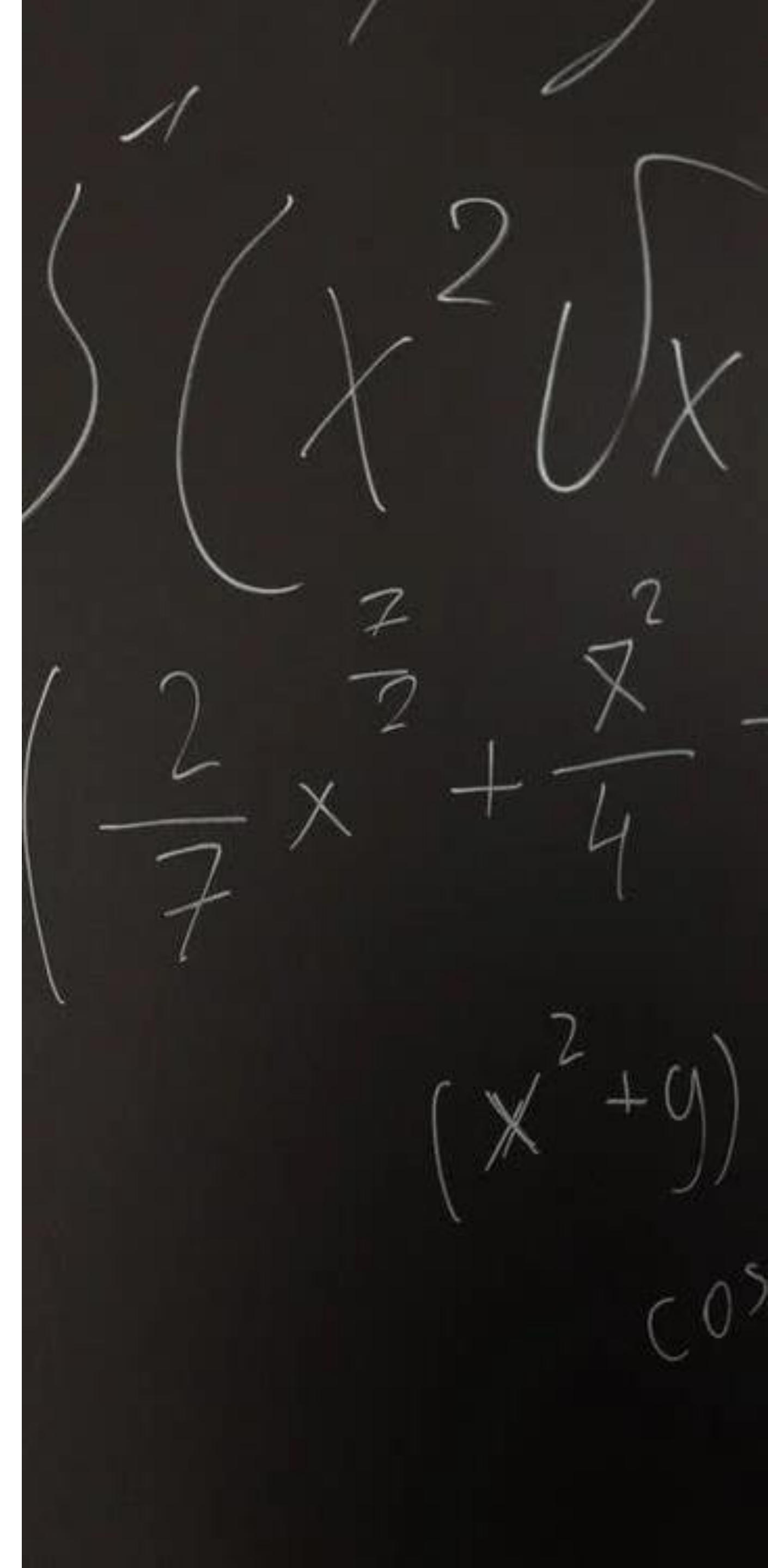
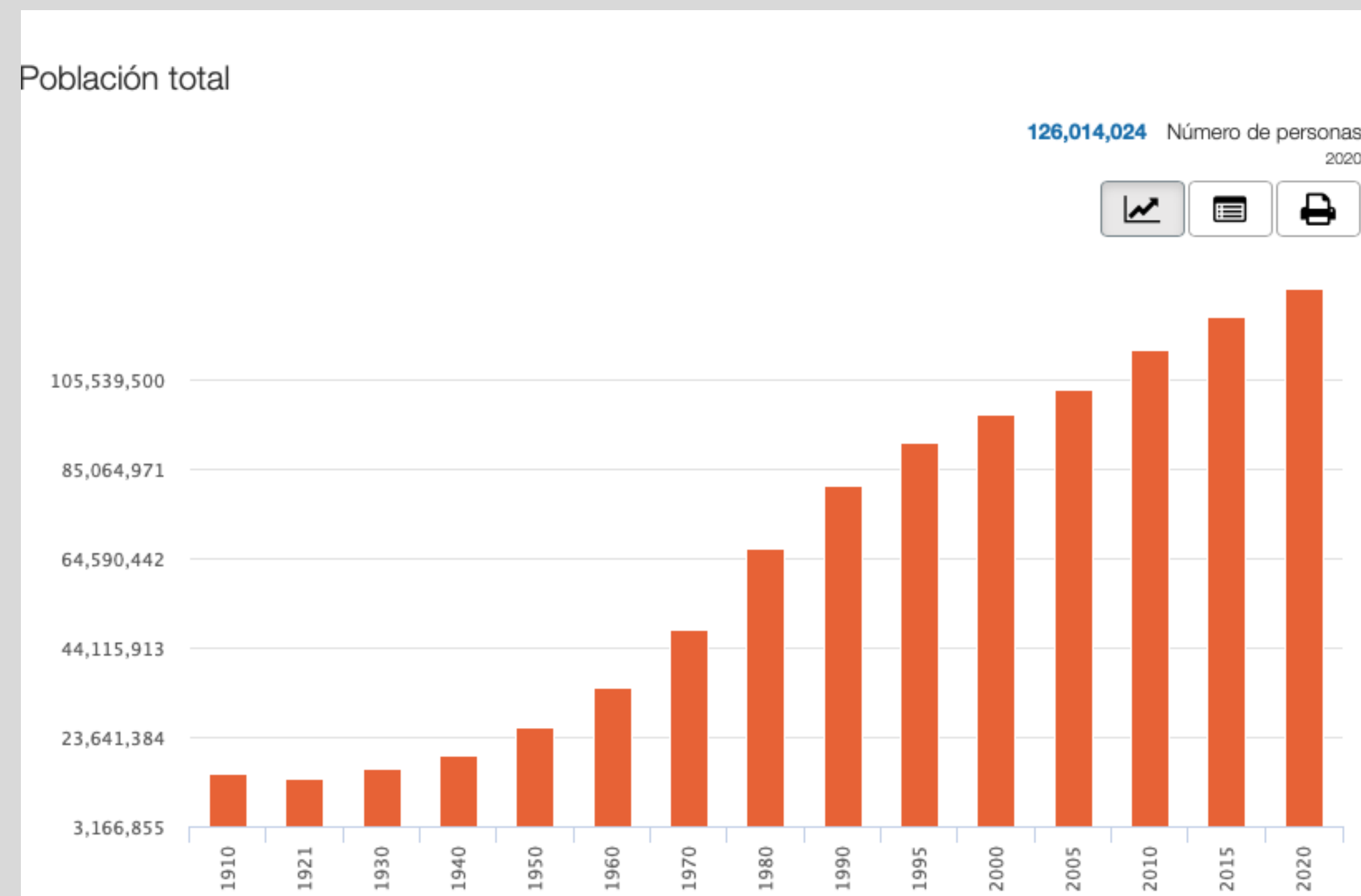




# Desarrollo de conceptos

## Análisis univariado

Análisis estadístico descriptivo que nos indica el comportamiento de una variable.



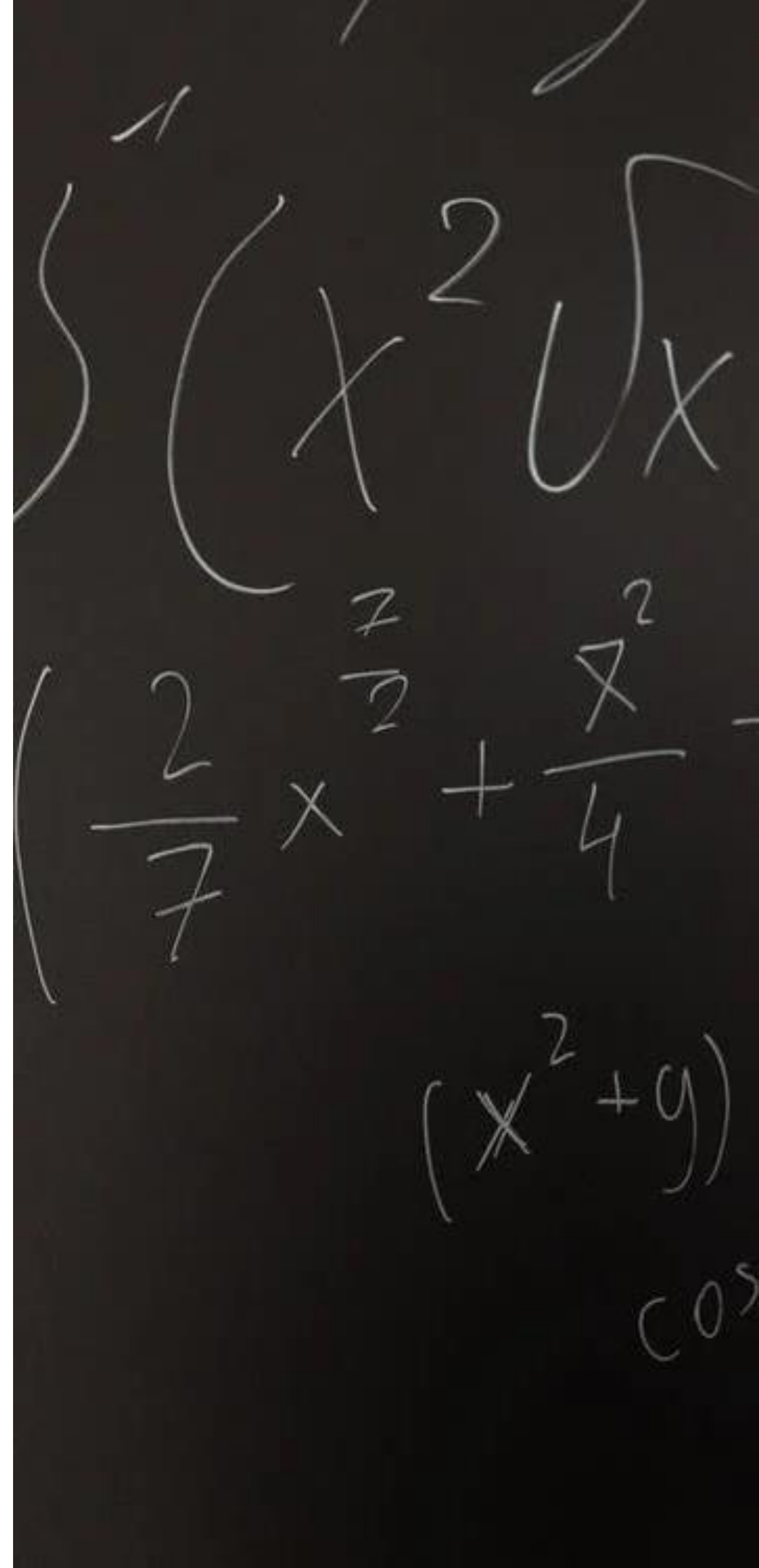
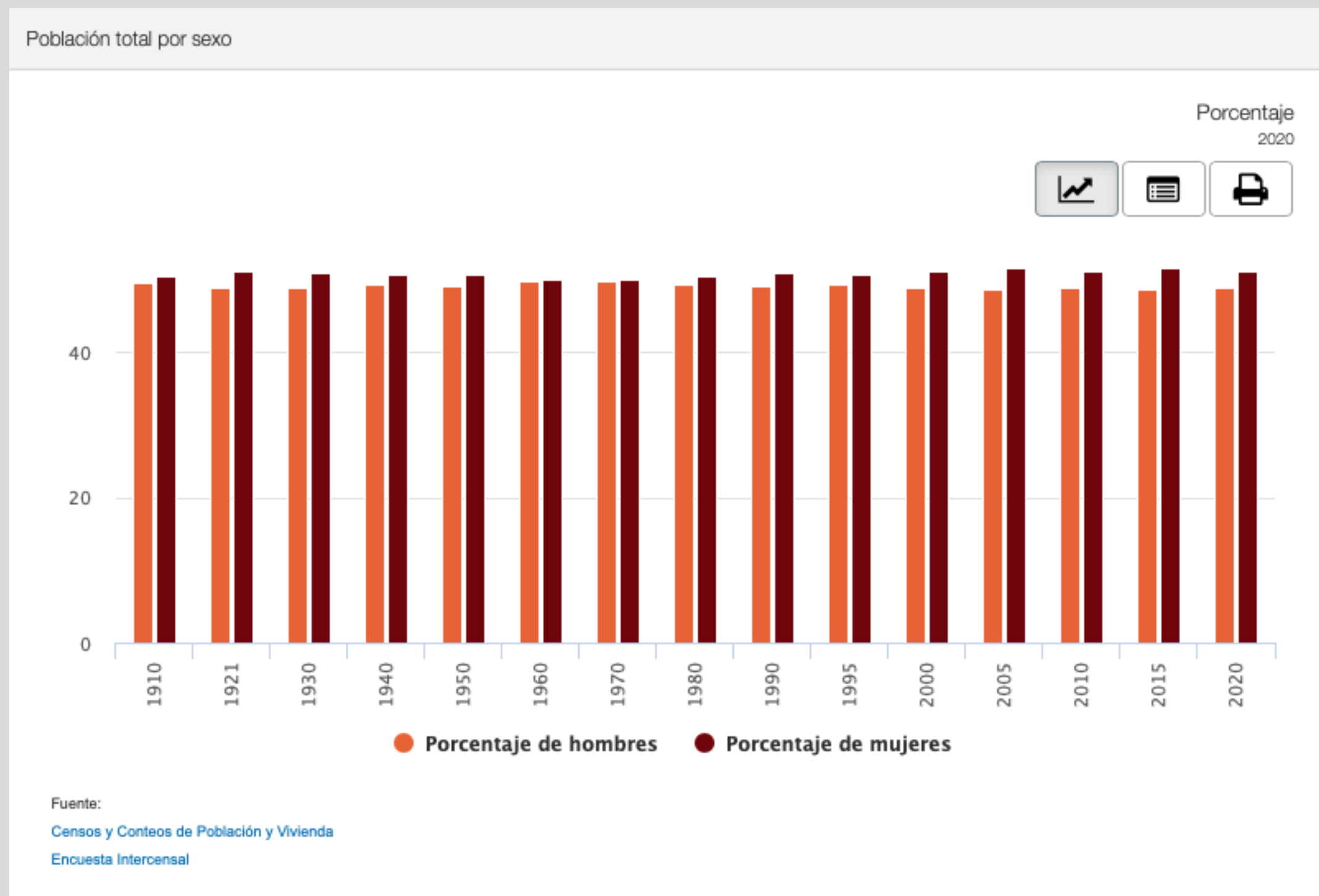




# Desarrollo de conceptos

## Análisis univariado

Medida de tendencia central, medidas de dispersión



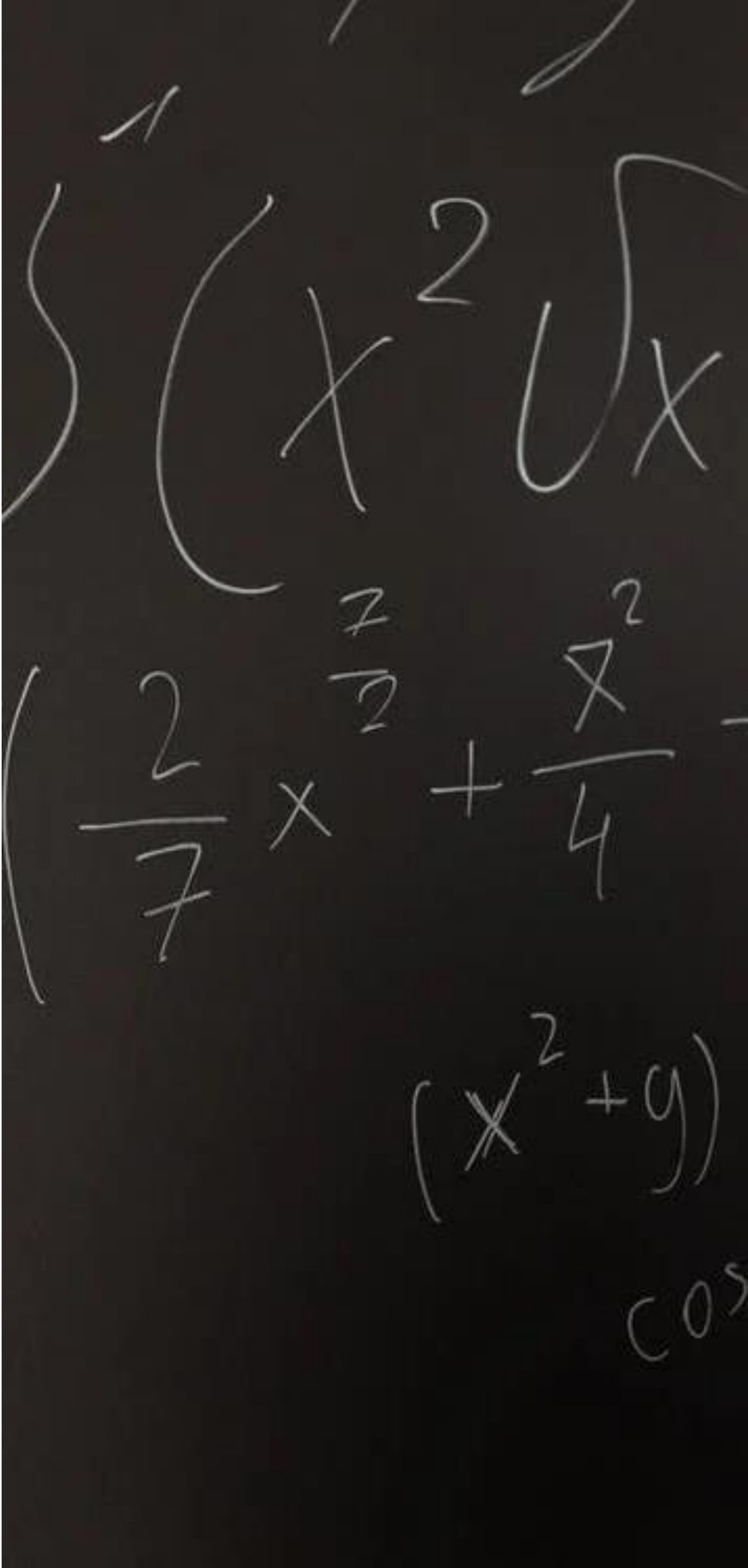
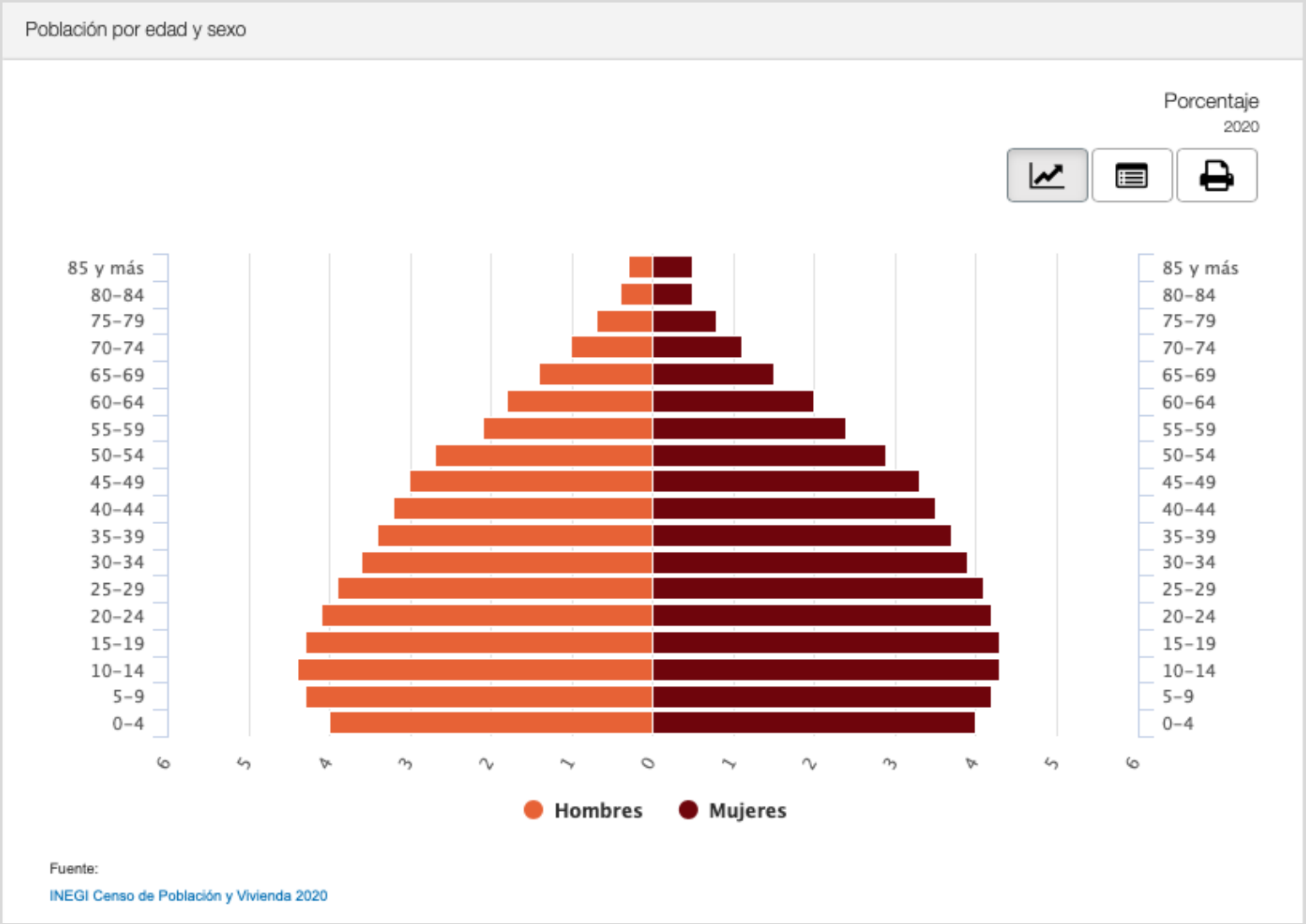
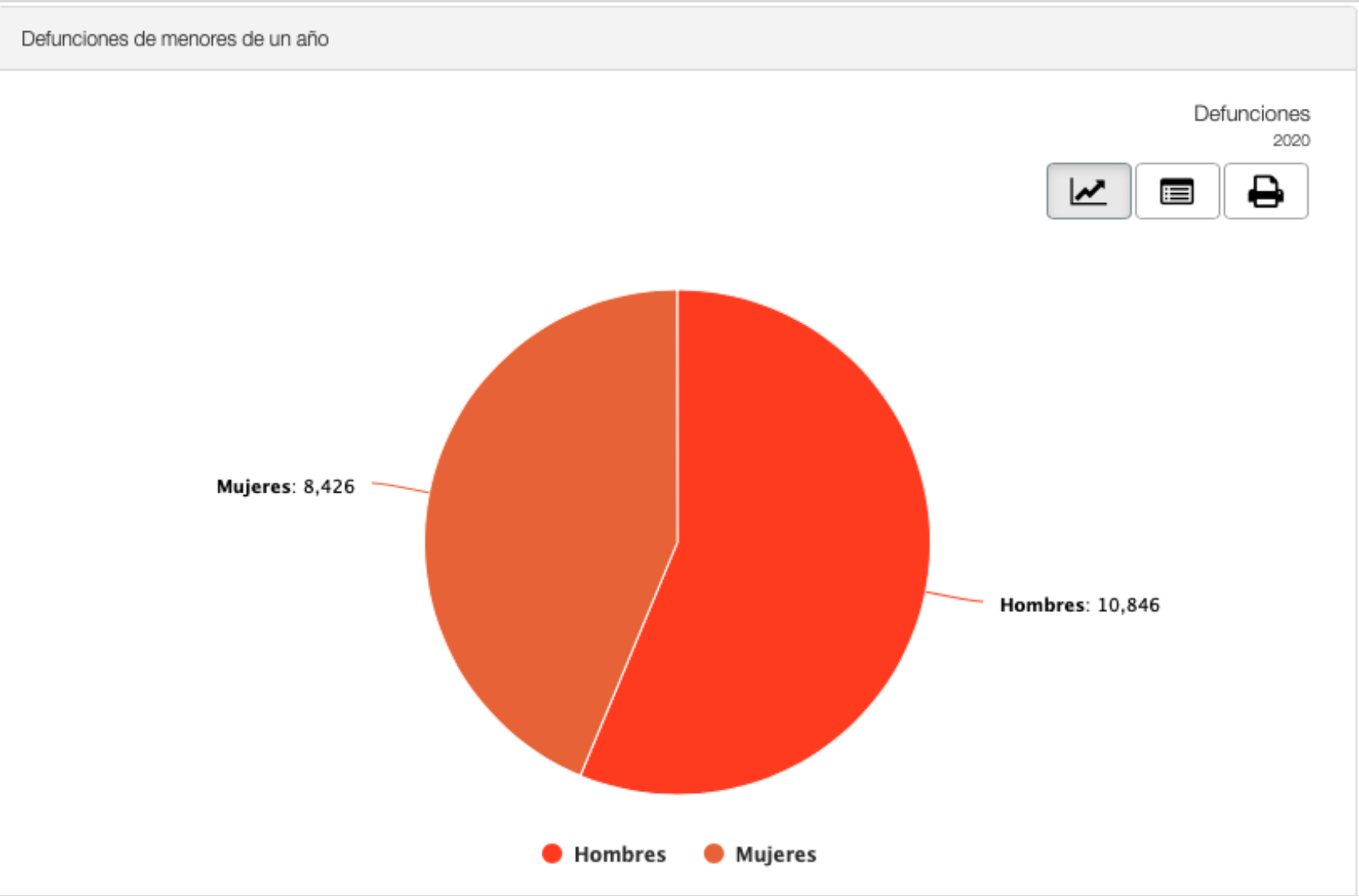




# Desarrollo de conceptos

## Análisis bivariado

Relación entre un par de atributos







# Desarrollo de conceptos

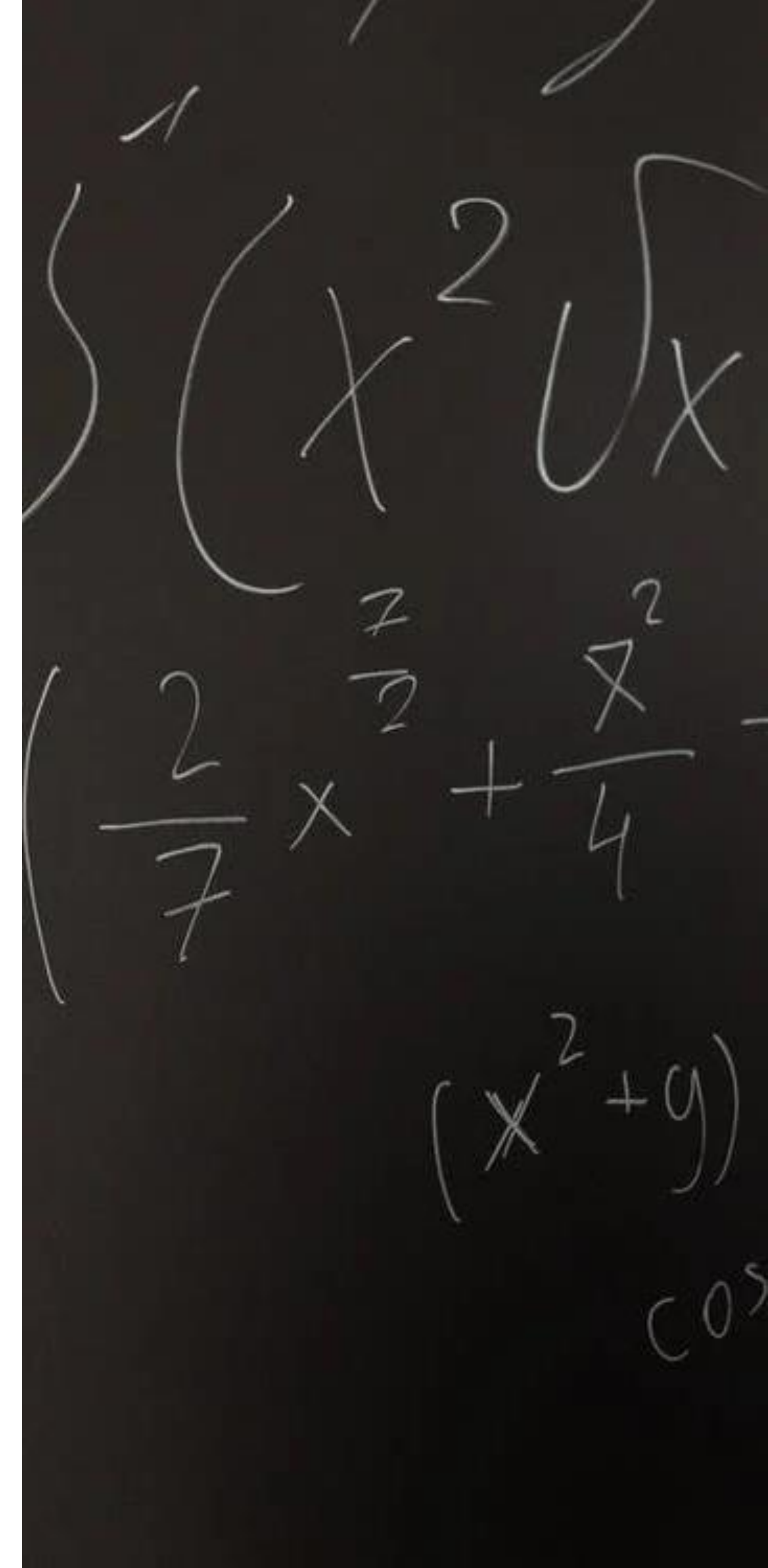
## Análisis multivariado

El **análisis estadístico multivariante** se ocupa de los datos que consisten en conjuntos de medidas sobre una serie de individuos u objetos. Los datos de la muestra pueden ser alturas y pesos de algunos individuos extraídos al azar de una población de niños en edad escolar en una ciudad dada.

(Anderson 2003)

Análisis de datos de altas dimensiones

Múltiples atributos de un individuo (edad, compras, ingresos)







# Desarrollo de conceptos

## Tipos

### Dependencia

Variable dependiente – qué variables independientes explican el fenómeno. **Multivariable**

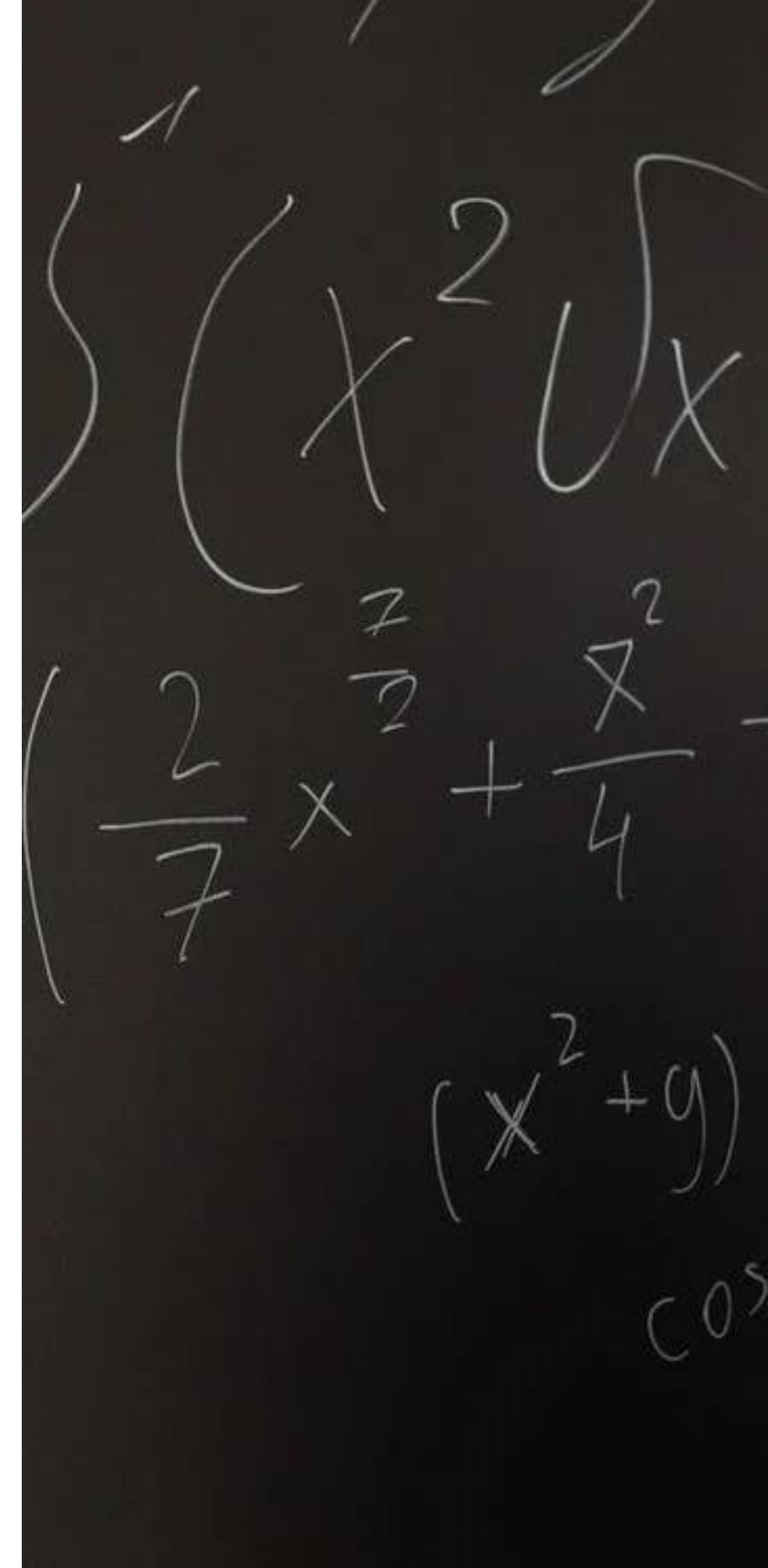
*Regresión lineal*

### Interdependencia

Análisis simultáneo de todas las variables – **Multivariado**

*Matriz de correlación Pearson Spearman*

*PCA*



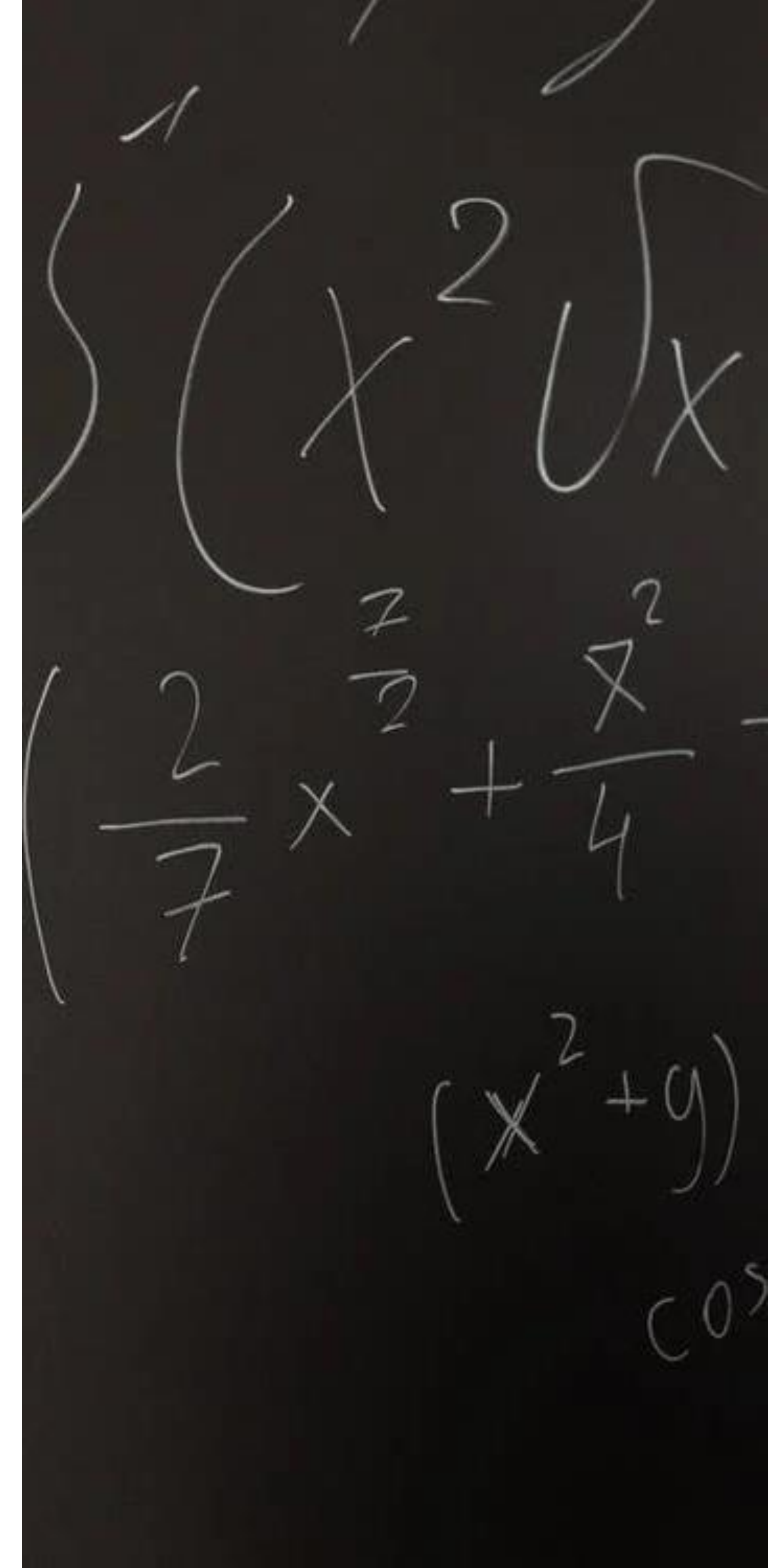




# Desarrollo de conceptos

## OBJETIVO DEL ANÁLISIS MULTIVARIADO

- Reducir el número de variables
  - Múltiples variables correlacionadas
  - Grupos más manejables
  - PCA
- Clasificación no supervisada
  - Identificar grupos (vinos, cervezas)
  - Conglomerados
- Clasificación supervisada
  - Con una variable clasificadora predecir el valor de una variable
  - Análisis discriminante lineal (LDA)







# Desarrollo de conceptos

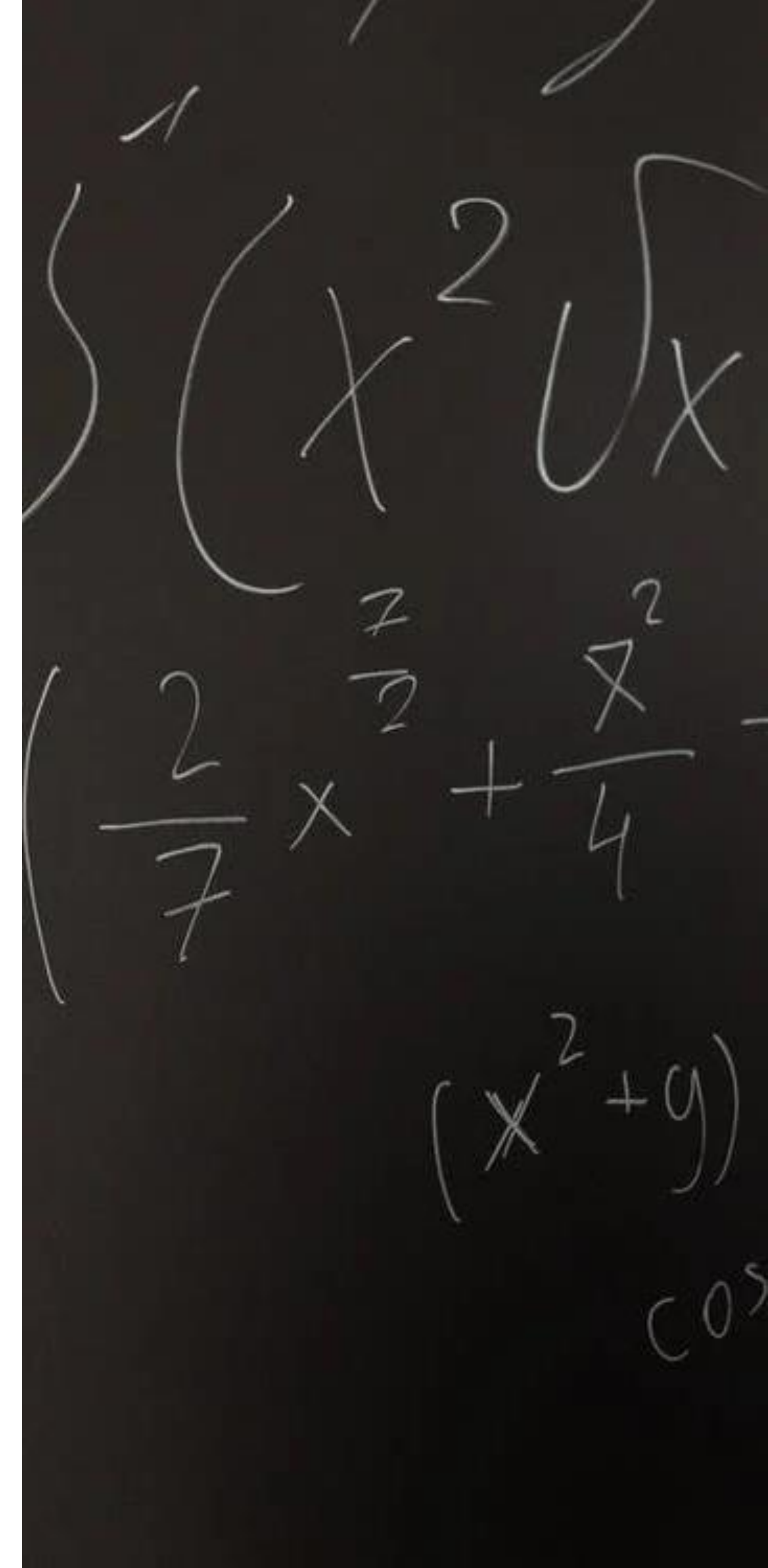
## Matemática

- Data Vector
- Data Matrix

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix};$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}.$$

- Fundamento del análisis multivariado es operaciones con matrices y vectores







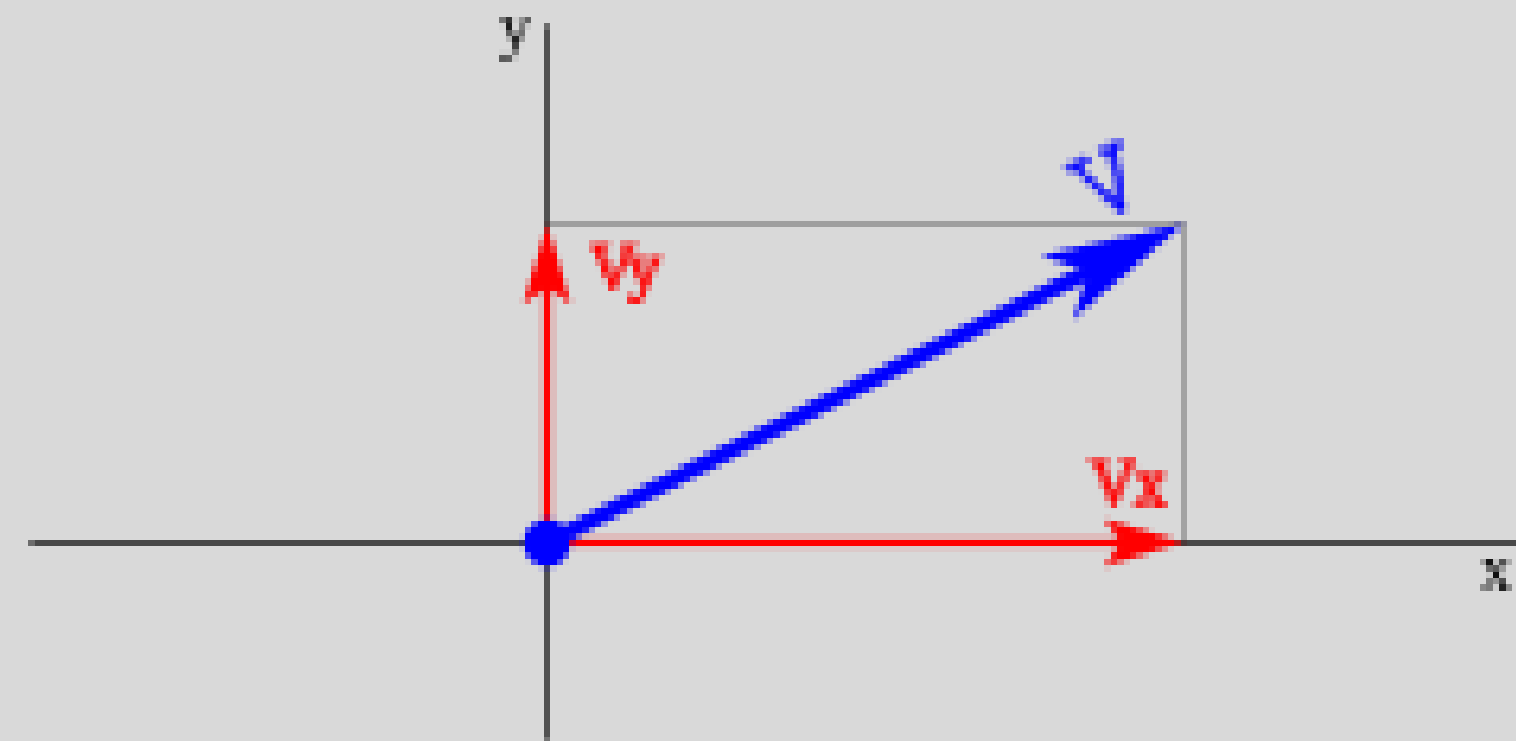
# Desarrollo de conceptos

# Métodos geométricos

- Álgebra lineal y álgebra matricial
- Los vectores tienen una representación espacial

# Vector

1  
2



$$\int (x^2 \sqrt{x})$$

$$\left( \frac{2}{7} x^{\frac{7}{2}} + \frac{x^2}{4} \right)$$

$$(x^2 + 9)$$

$$\cos$$

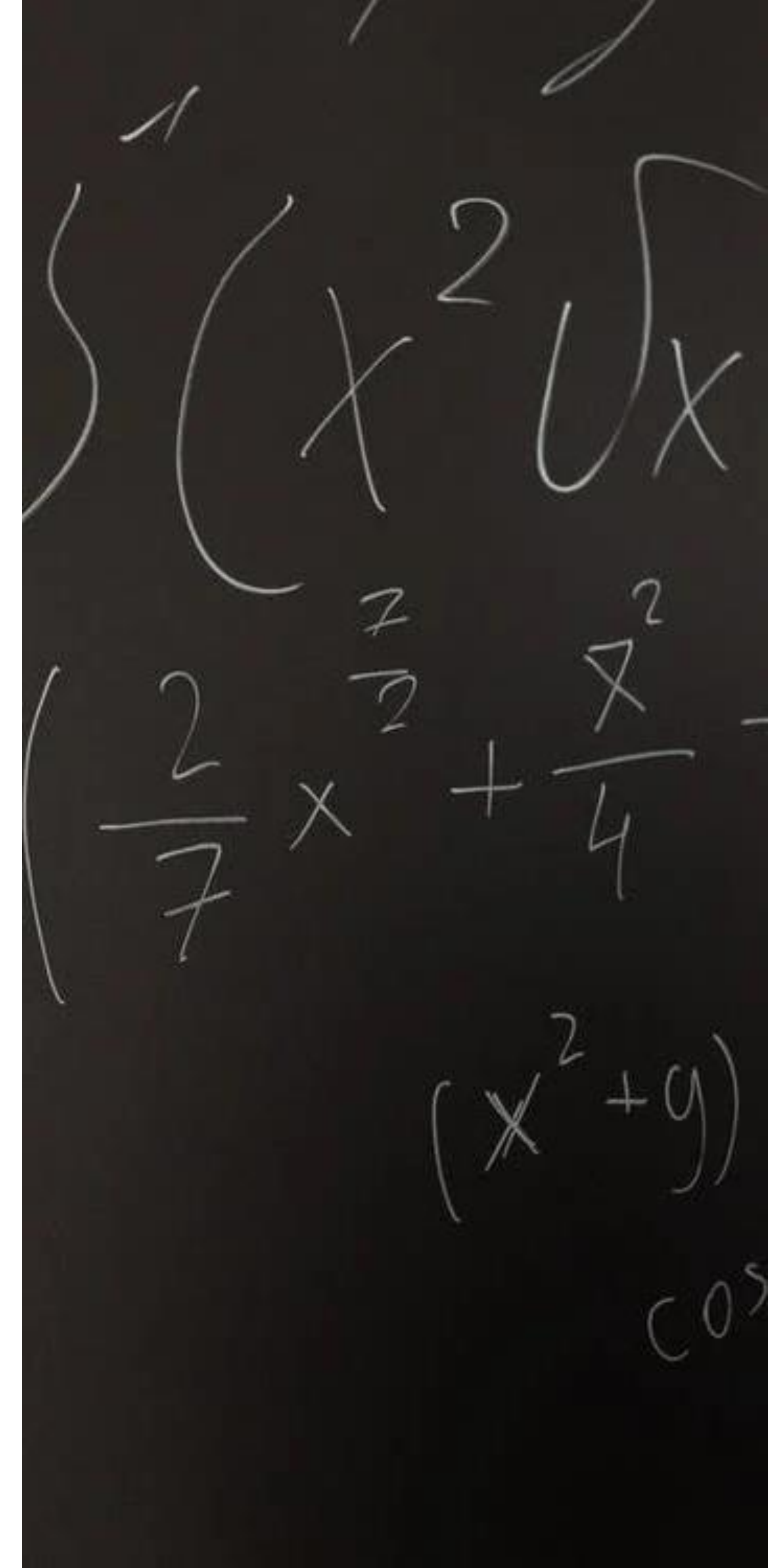




# Desarrollo de conceptos

## Proceso

1. Inspección visual de los datos
  - Distribución, máximos, mínimos
2. Análisis descriptivo multivariado y su visualización
  - Evaluar relaciones entre variables



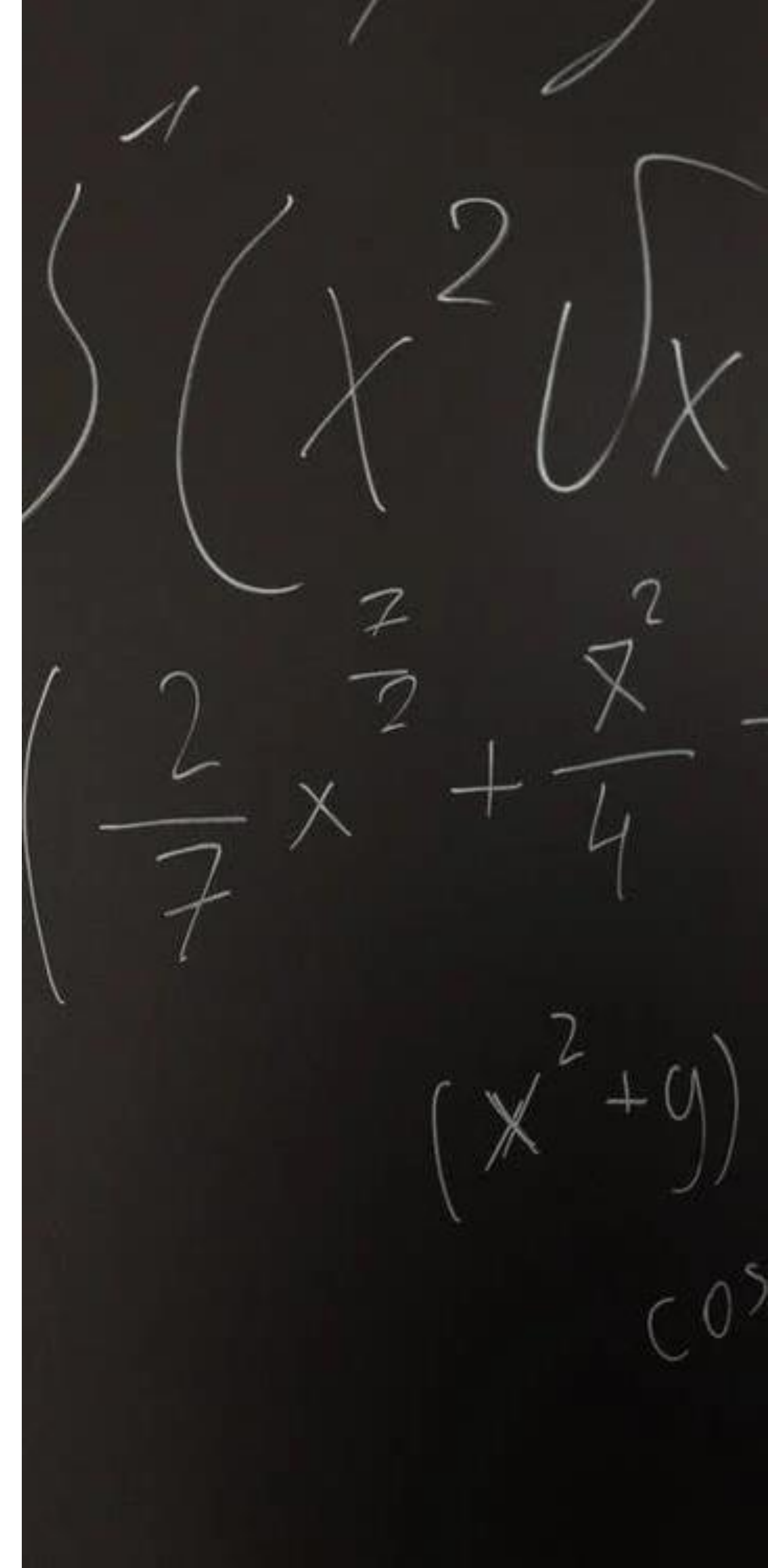
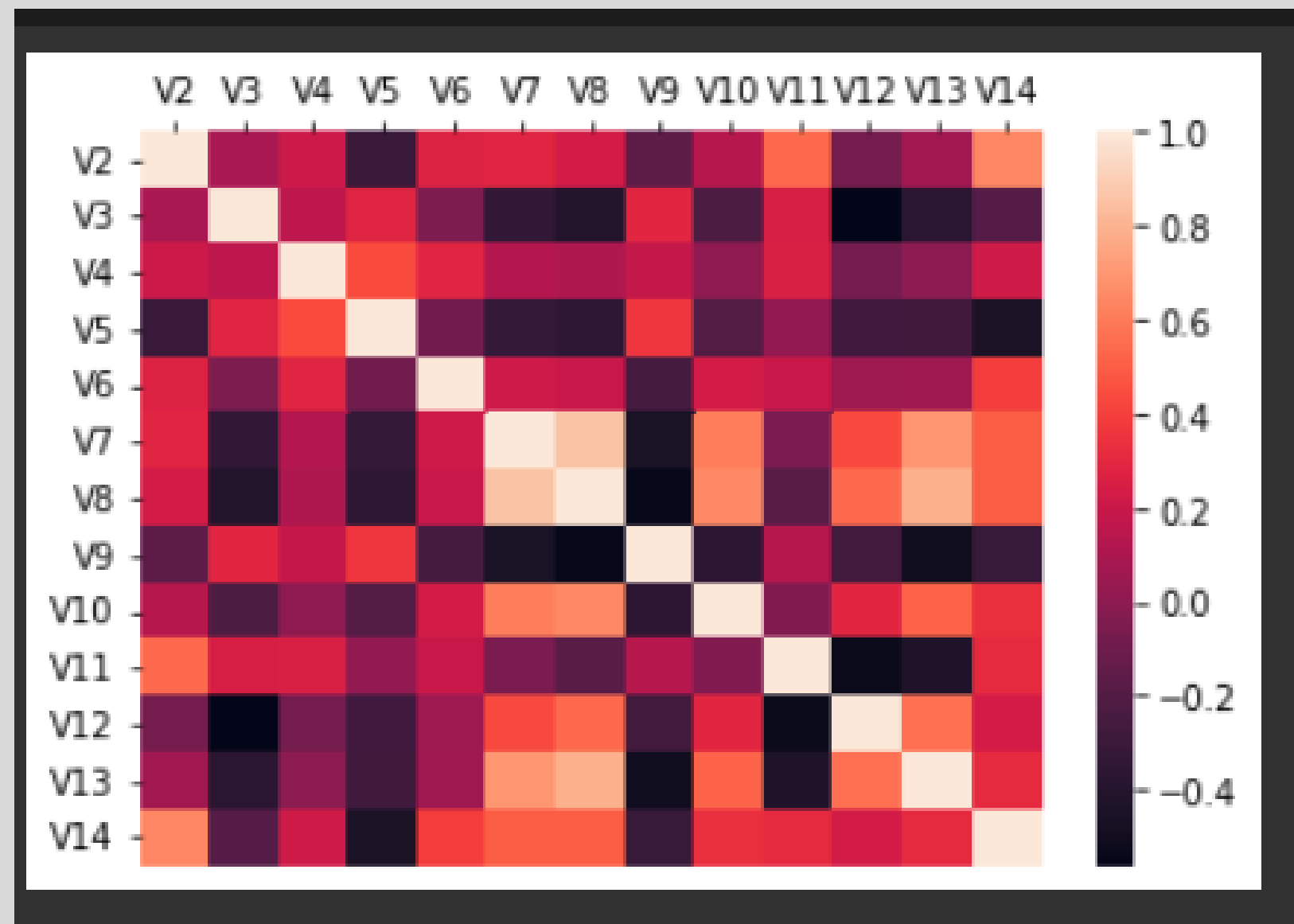
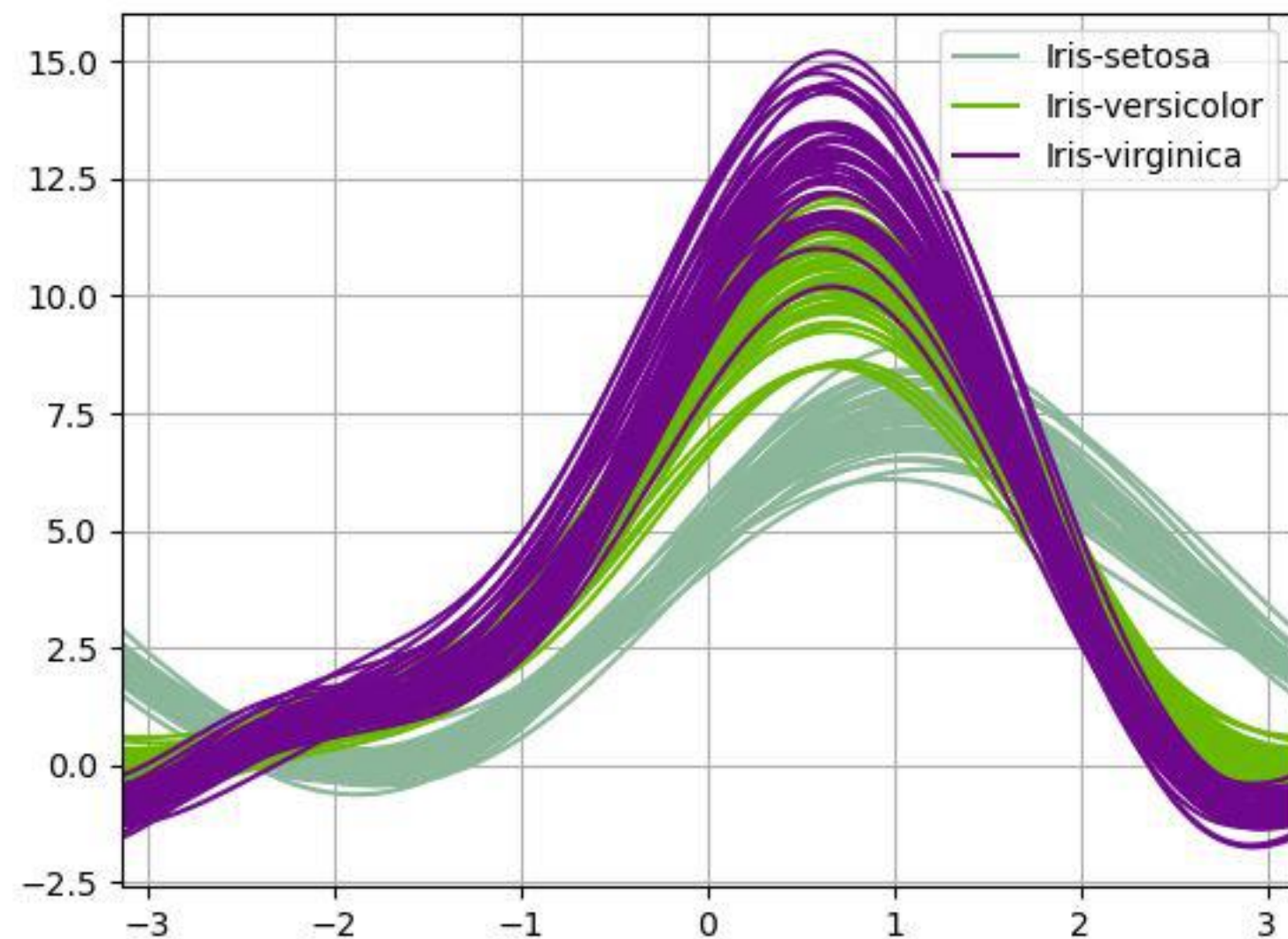




# Desarrollo de conceptos

## Proceso

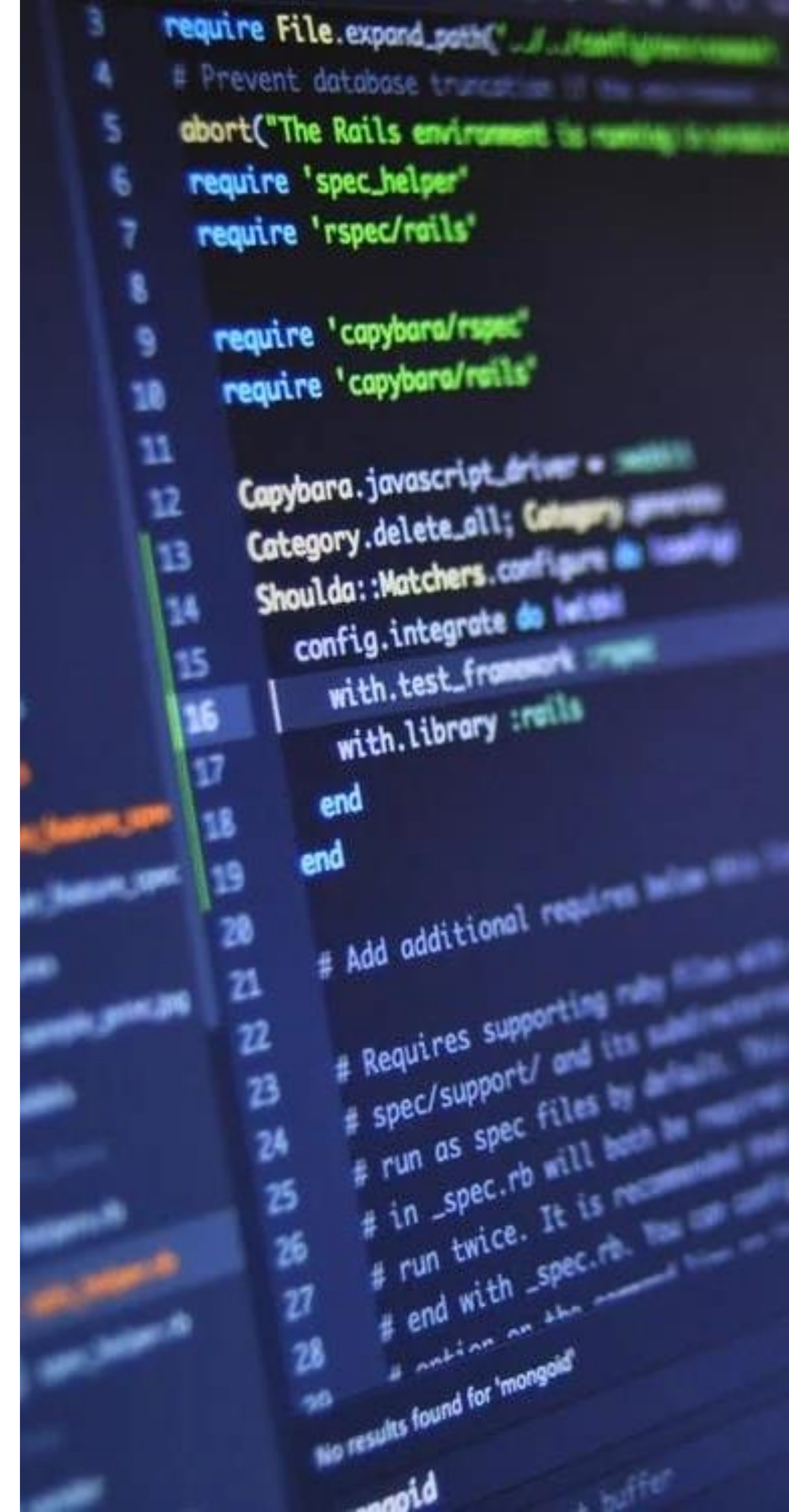
- Análisis descriptivo multivariado y su visualización





# Solución

- Crear Cuenta en Colab
  - <https://colab.research.google.com>
- Descargar Orange 3
  - <https://orangedatamining.com>







## Bibliográficos

- Introduction to machine learning - Ethem Alpaydın
- [Galeano-Arias 2020] - <https://scielo.conicyt.cl/pdf/infotec/v32n1/0718-0764-infotec-32-01-17.pdf>
- An Introduction to Multivariate Statistical Analysis (Wiley Series in Probability and Statistics) (T. W. Anderson)
- [Alaminos 2015] Análisis multivariante para las Ciencias Sociales, Antonio Alaminos Chica
- D. Garcia-Alvarez, Estudio comparativo de técnicas de detección de fallos basadas en el Análisis de Componentes Principales (PCA), <https://www.sciencedirect.com/science/article/pii/S1697791211000070>
- <https://www.python.org/doc/>
- <https://realpython.com/world-class-companies-using-python/>
- <https://www.anaconda.com>
- <https://jupyter.org>
- <https://numpy.org>
- <https://pandas.pydata.org>

# Recursos



## Dataset

<https://archive.ics.uci.edu/ml/datasets/wine>

[https://jdvelasq.github.io/courses/notebooks/sklearn\\_dataset\\_utilities/2-11\\_load\\_wine.html](https://jdvelasq.github.io/courses/notebooks/sklearn_dataset_utilities/2-11_load_wine.html)

## Notas Python

<https://github.com/erlinares/mlengineer/tree/main/cheatsheets>

## Software

- Python
- Colab
- Orange





# Solución

- Dentro de Colab
- <https://colab.research.google.com>

```
3 require File.expand_path("../..", __FILE__)
4 # Prevent database truncation if the environment is production
5 abort("The Rails environment is running in production")
6 require 'spec_helper'
7 require 'rspec/rails'
8
9 require 'capybara/rspec'
10 require 'capybara/rails'
11
12 Capybara.javascript_driver = :webkit
13 Category.delete_all; Category.create
14 Shoulda::Matchers.configure do |config|
15   config.integrate do |integrate|
16     with.test_framework :rspec
17     with.library :rails
18   end
19 end
20
21 # Add additional requires below this to modify RSpec's default
22 # behavior. The first require is for the rails gem, which is
23 # required by default. The second is for the rails_spec_helper
24 # gem, which is required by default. The third is for the
25 # rails_spec_helper gem, which is required by default. The
26 # fourth is for the rails_spec_helper gem, which is required
27 # by default. The fifth is for the rails_spec_helper gem,
28 # which is required by default. The sixth is for the
29 # rails_spec_helper gem, which is required by default.
```



## Bibliográficos

- <https://www.python.org/doc/>
- <https://realpython.com/world-class-companies-using-python/>
- <https://www.anaconda.com>
- <https://jupyter.org>
- <https://numpy.org>
- <https://pandas.pydata.org>