

# Recap

Adrian Jackson, EPCC, The University of Edinburgh

[a.jackson@epcc.ed.ac.uk](mailto:a.jackson@epcc.ed.ac.uk)

August 2025

[www.archer2.ac.uk](http://www.archer2.ac.uk)



# Reusing this material



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

This means you are free to copy and redistribute the material and adapt and build on the material under the following terms: You must give appropriate credit, provide a link to the license and indicate if changes were made. If you adapt or build on the material, you must distribute your work under the same license as the original.

Note that this presentation contains images owned by others. Please seek their permission before reusing these images.

# Partners



Engineering and  
Physical Sciences  
Research Council

Natural  
Environment  
Research Council



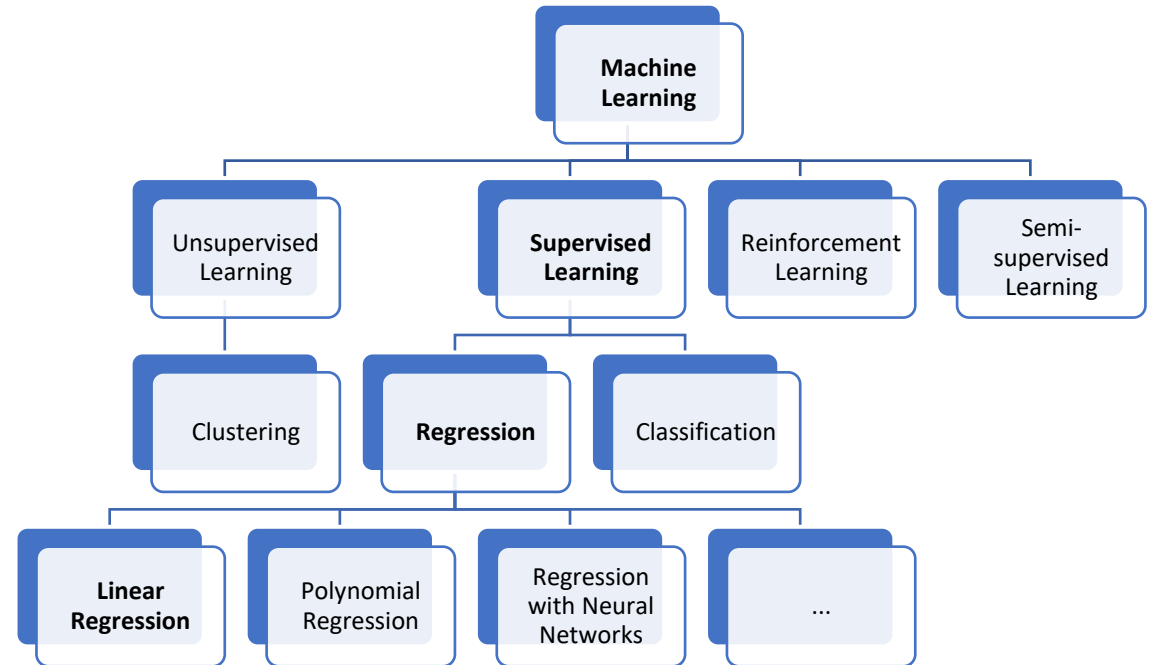
THE UNIVERSITY  
*of* EDINBURGH

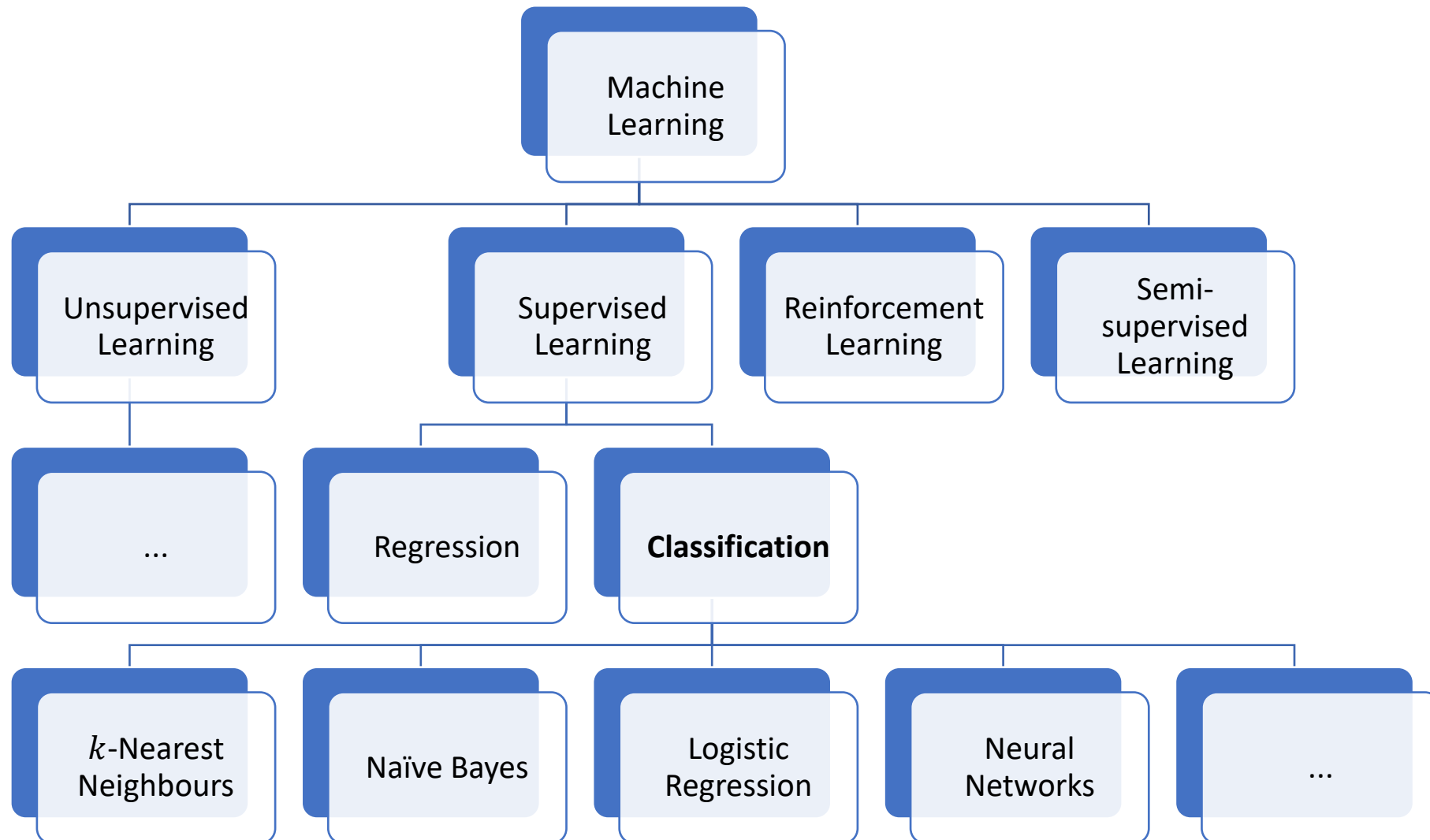


**Hewlett Packard  
Enterprise**

# Linear Regression

- Linear Regression predicts a continuous variable from one or more variables
- One of the simplest predictive models
- Existing data is used to create a **linear model**



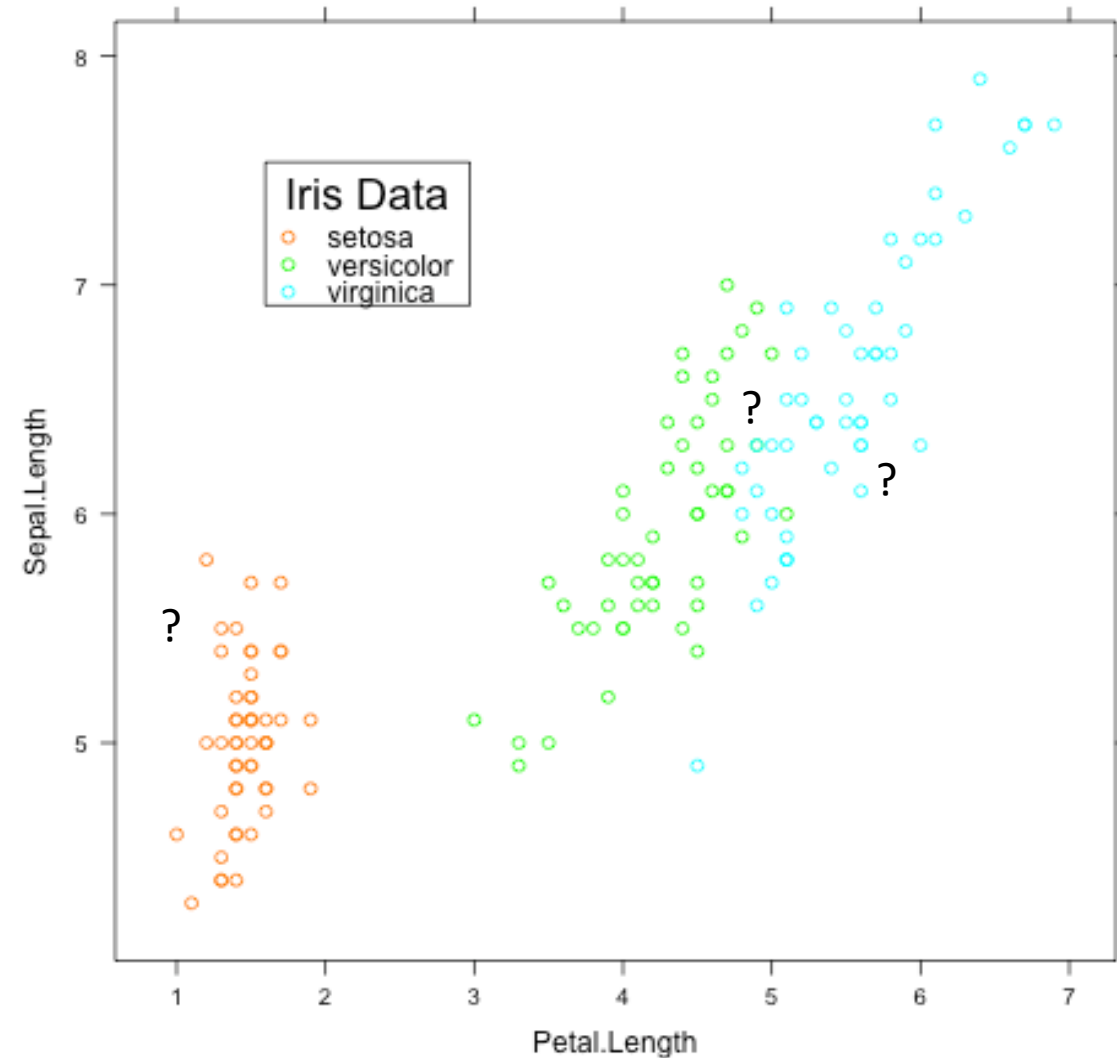


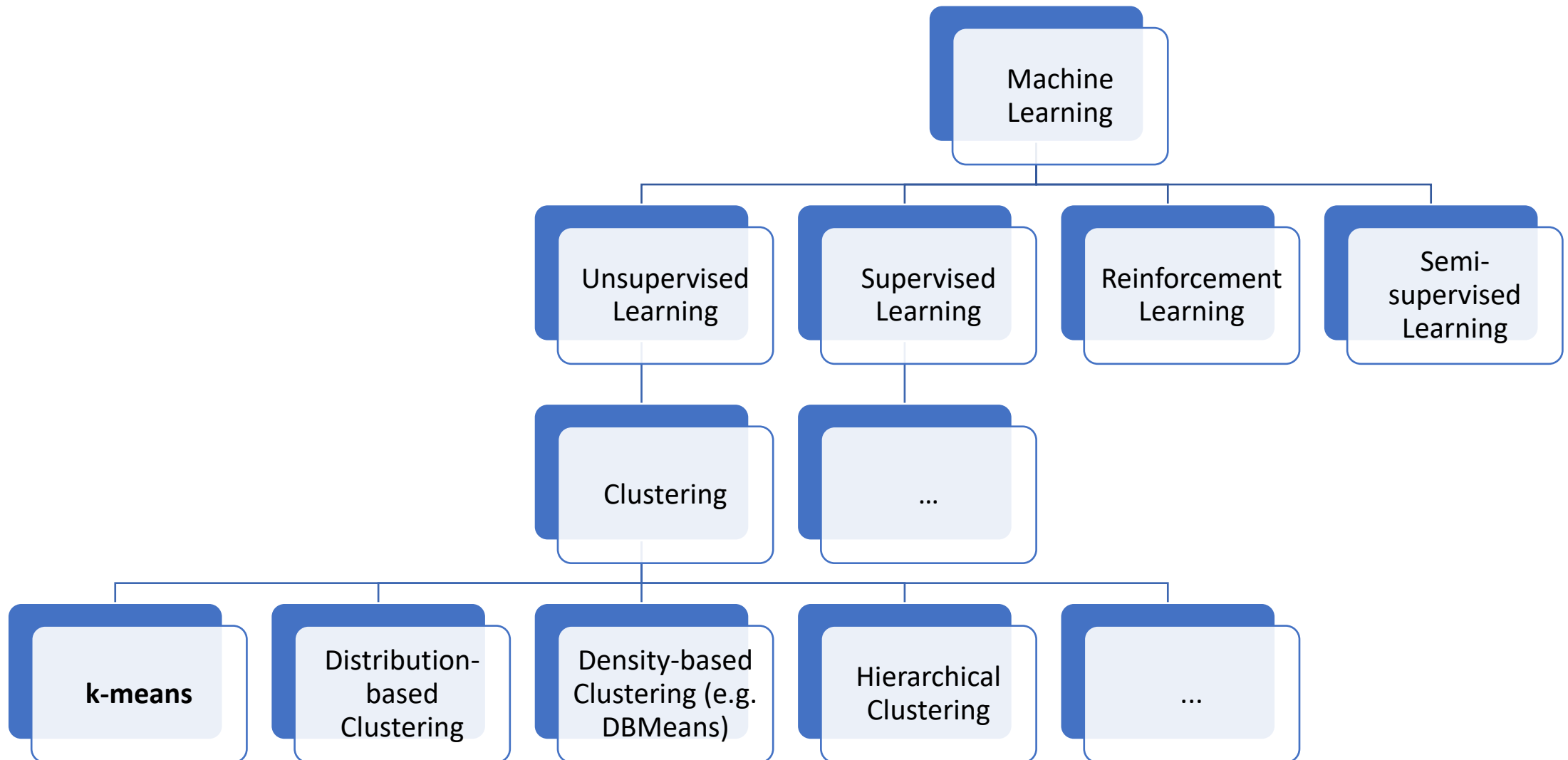
# Classification: problem overview

- Input is one or more features
  - numerical or categorical
- Output is a category (or class)



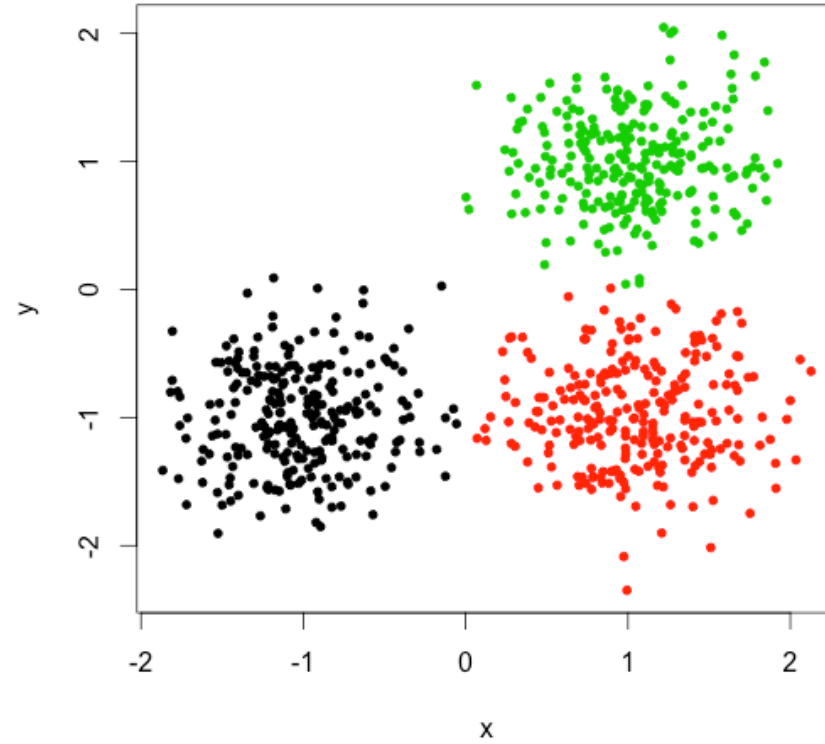
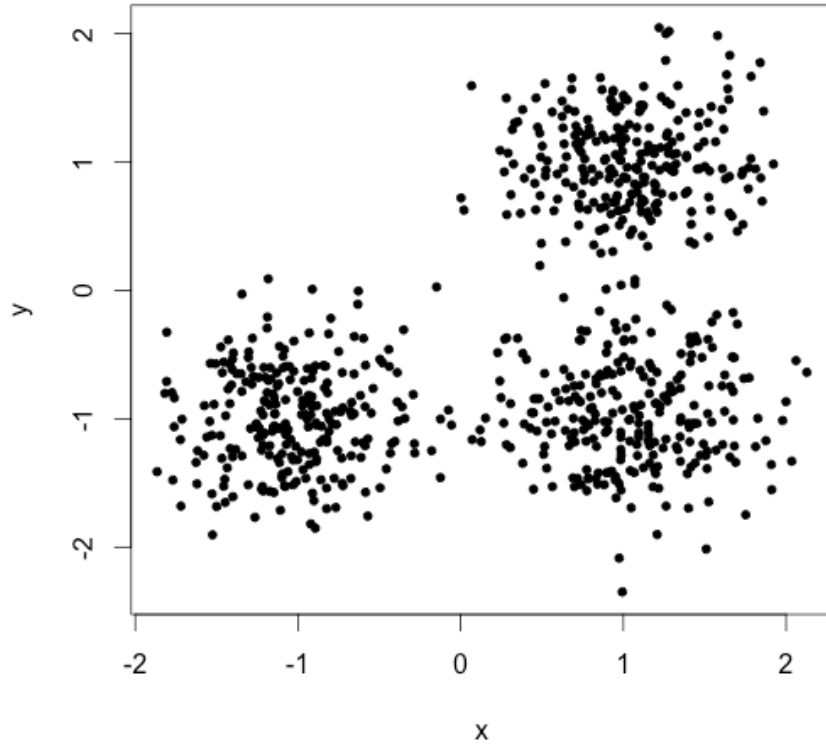
Iris image courtesy of David Berger under a CC-BY licence







# Clustering – the problem





# Data science and machine learning



- Lots more to data science than model building
  - Data collection, cleaning, analysis, visualisation, etc...
- Model choice and training/building requires understanding your problem
  - Might be simple approaches that work well
  - With complex problems or large datasets machine learning approaches can be useful