

# Notas escépticas sobre el Machine Learning

Lecture 4: Unsupervised data

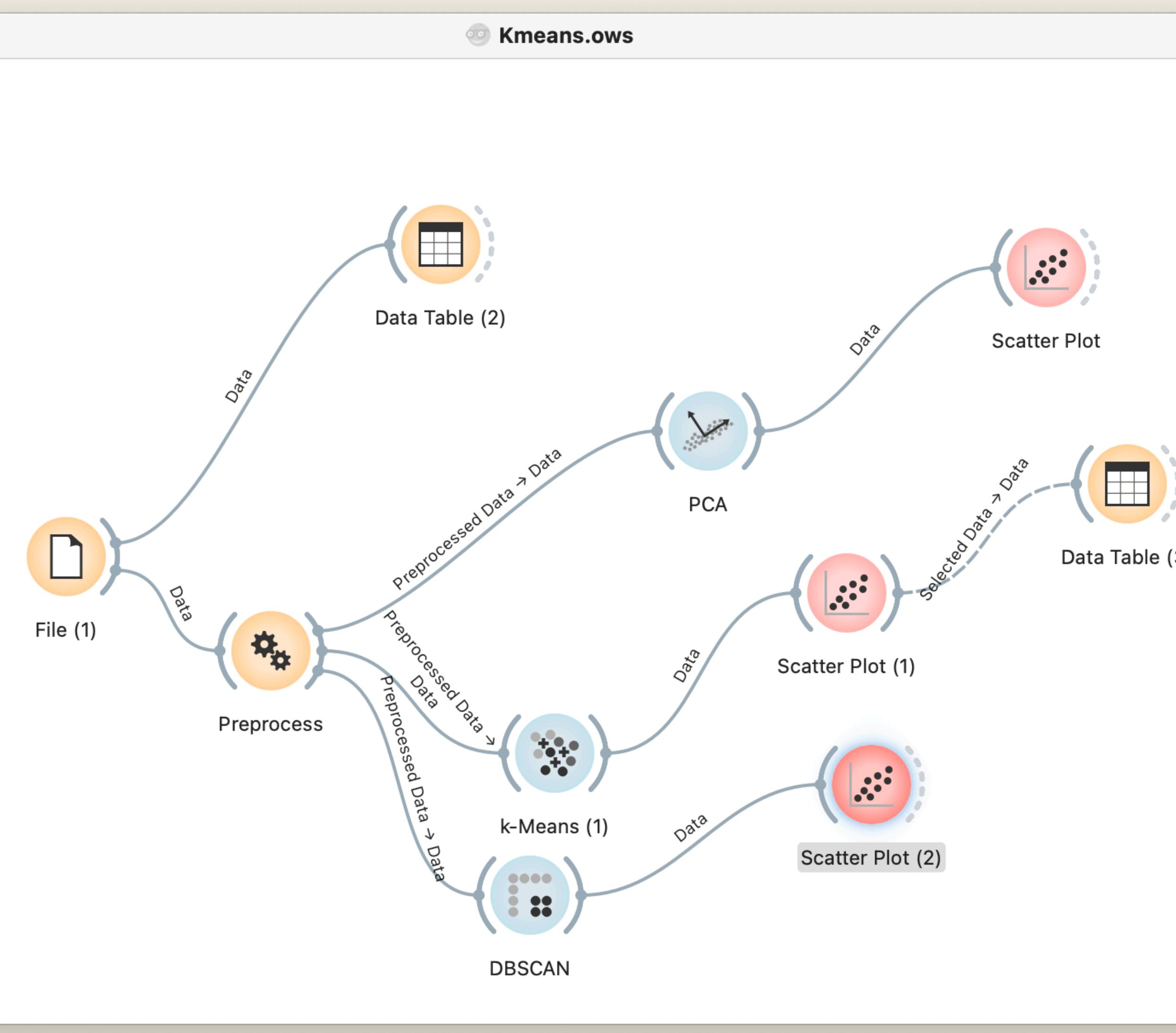
# Type of unsupervised learning

- **Unsupervised transformations** of a dataset are algorithms that create a new representation of the data which might be easier for humans or other machine learning algorithms to understand compared to the original representation of the data (dimensionality reduction)
- **Clustering**, on the other hand, partition data into distinct groups of similar items.

---

# Challenges in Unsupervised Learning

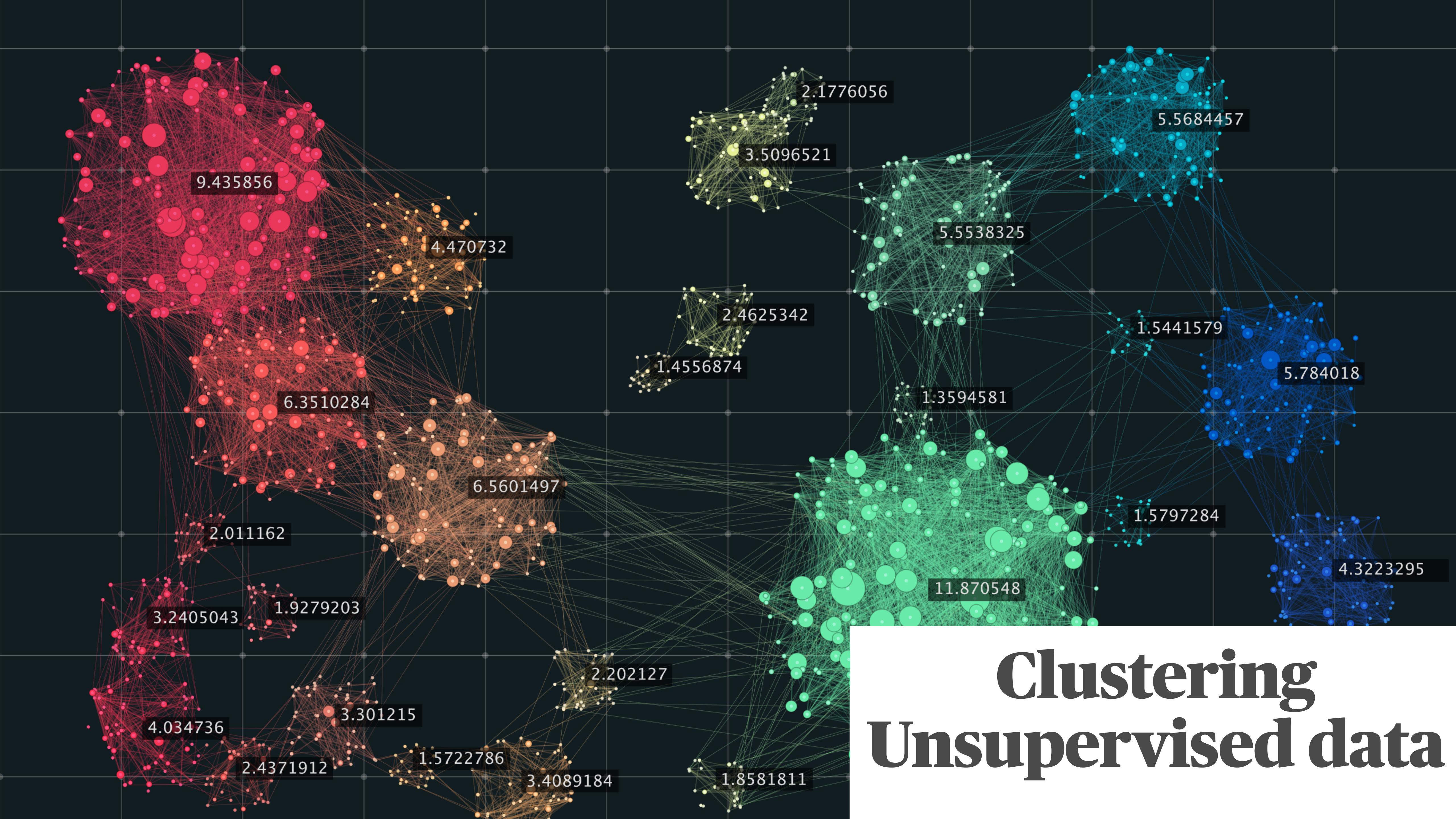
A major challenge in unsupervised learning is evaluating whether the algorithm learned something useful. Unsupervised learning algorithms are usually applied to data that does not contain any label information, so we don't know what the right output should be. Therefore, **it is very hard to say whether a model "did well."**



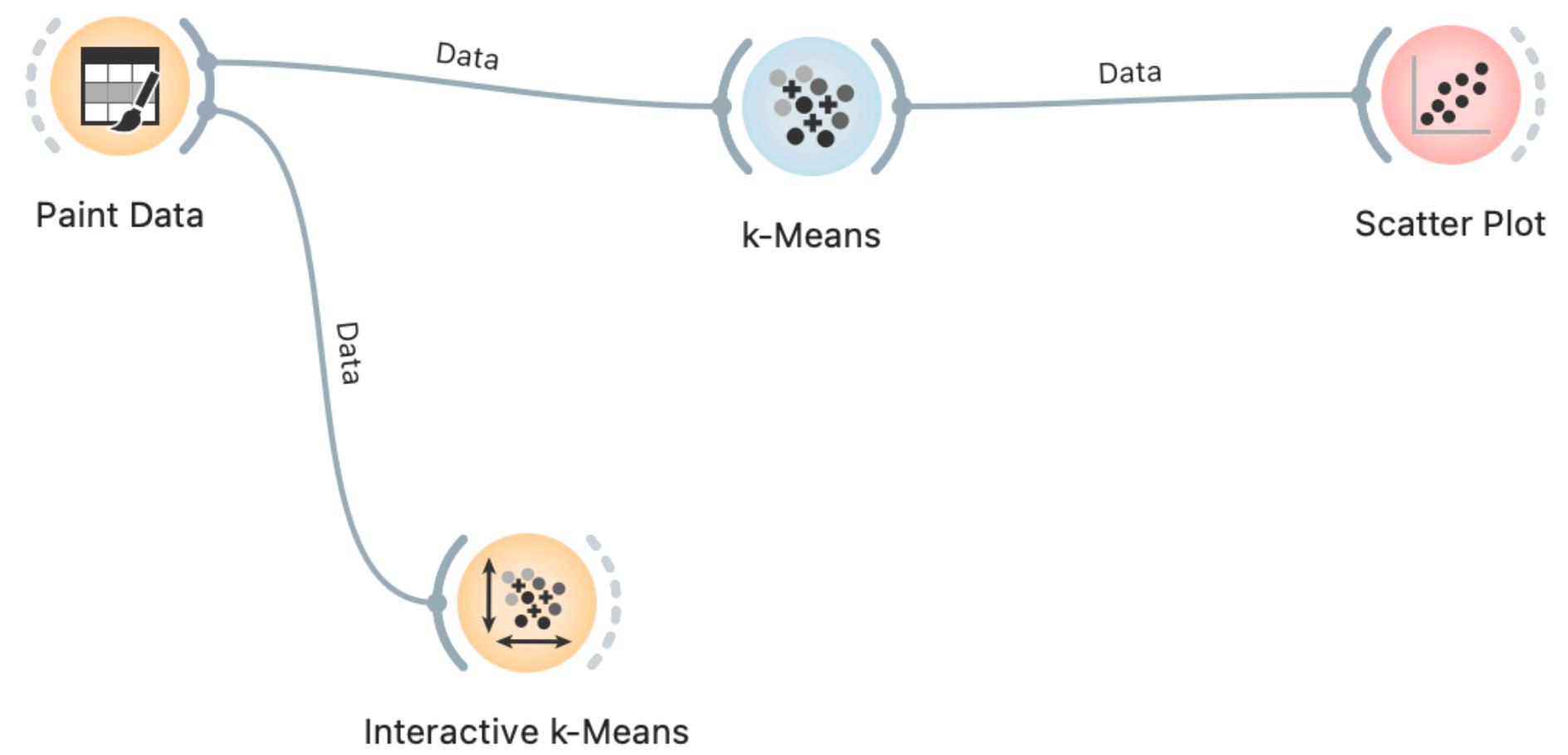
6/ABRIL/2022

# Preprocessing and Scaling

- StandardScaler
- RobustScaler
- MinMaxScaler



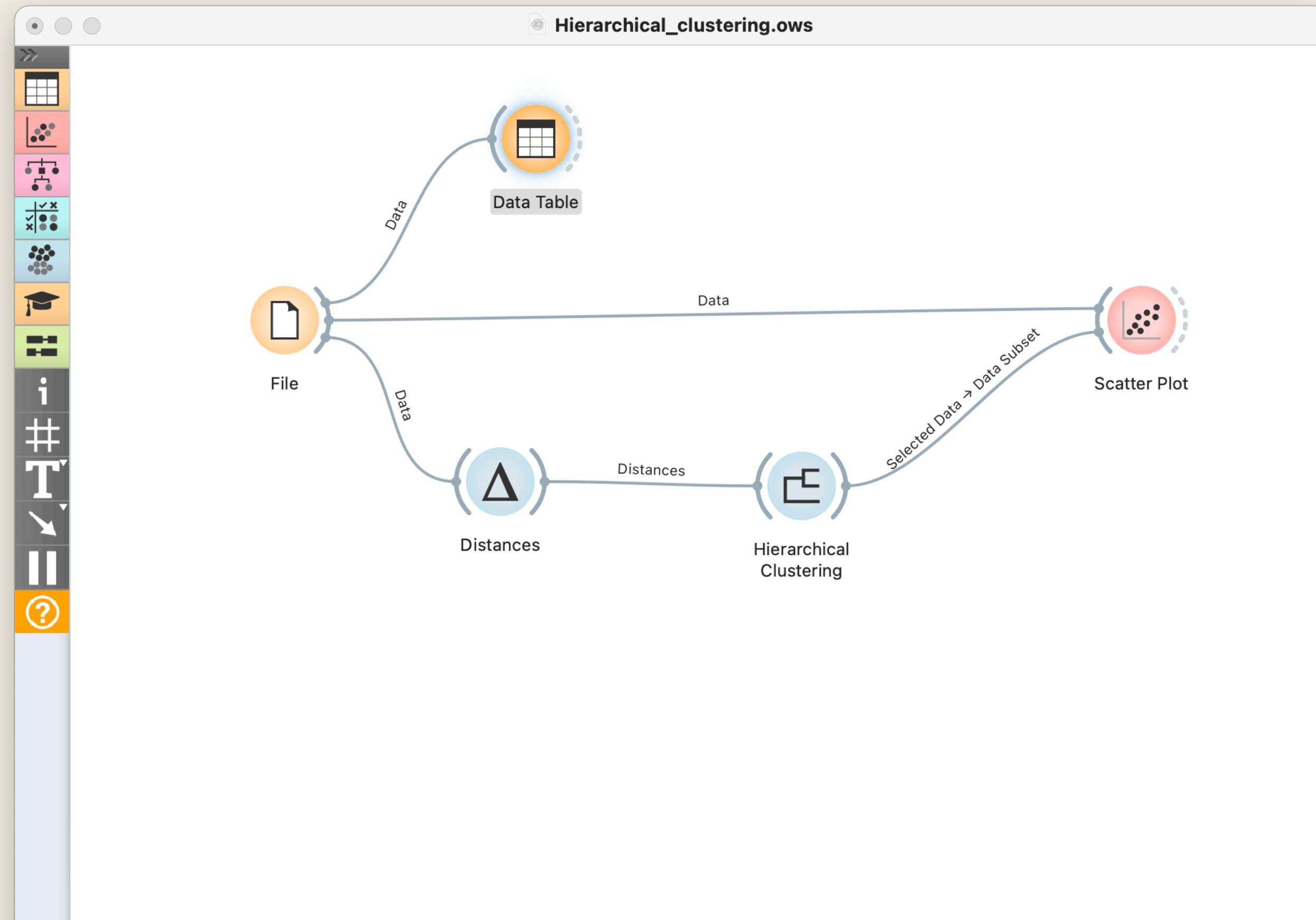
# Clustering Unsupervised data



11/MARZO/2021

# K-means

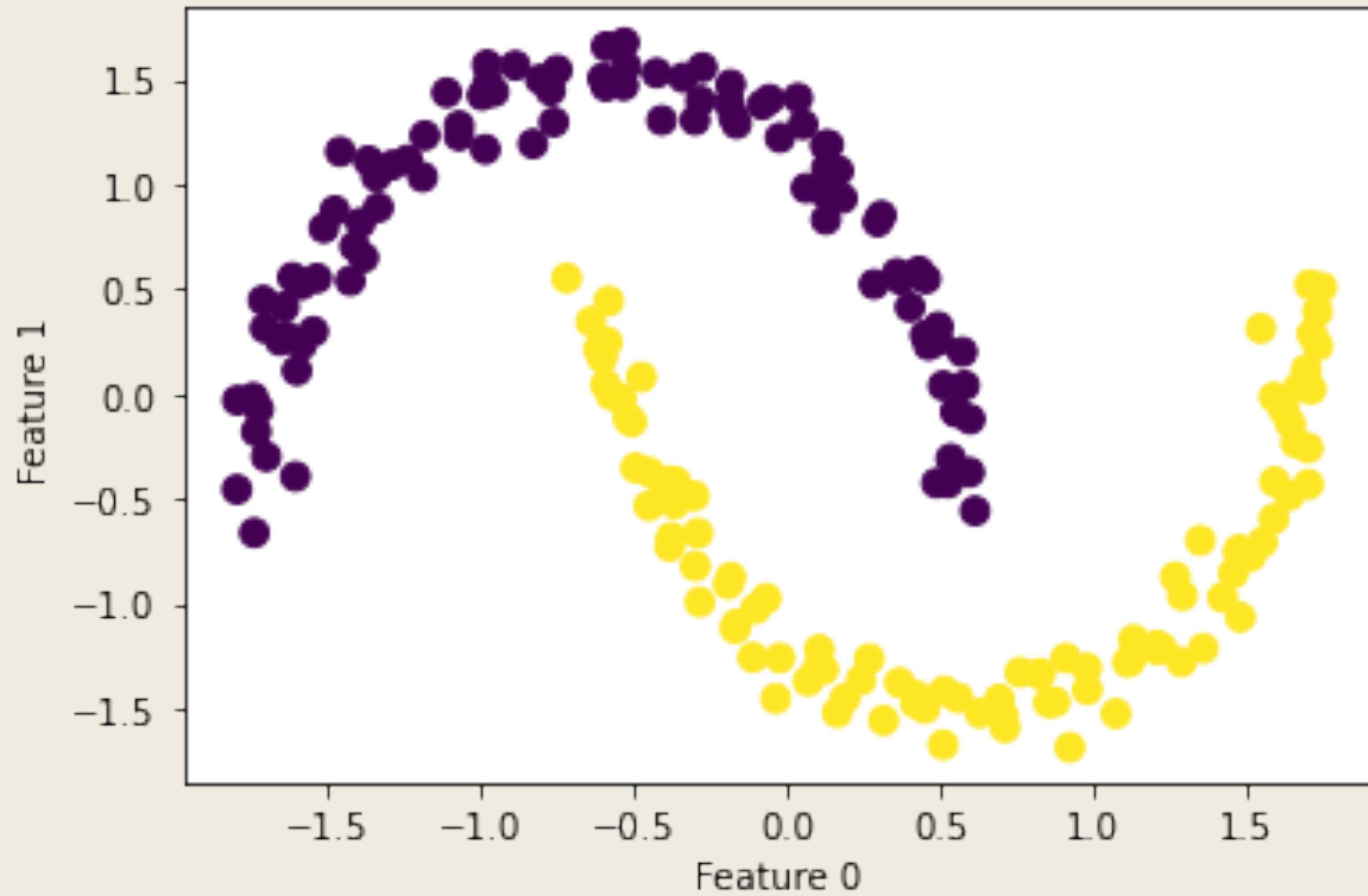
- Paint Data
- K-means → Scatter plot
- Interactive K-Means
- 2 example:
- Iris Data



11/MARZO/2021

# Hierarchical clustering

- Iris Data → Data Table
- Distances → hierarchical clustering
- Scatter plot



22/MARZO/2022

## DBSCAN

- DBSCAN works by identifying points that are in “crowded” regions of the feature space, where many data points are close together.
- `min_samples` and `eps`.
- Three kind of points: core, border, and outlier (noise)