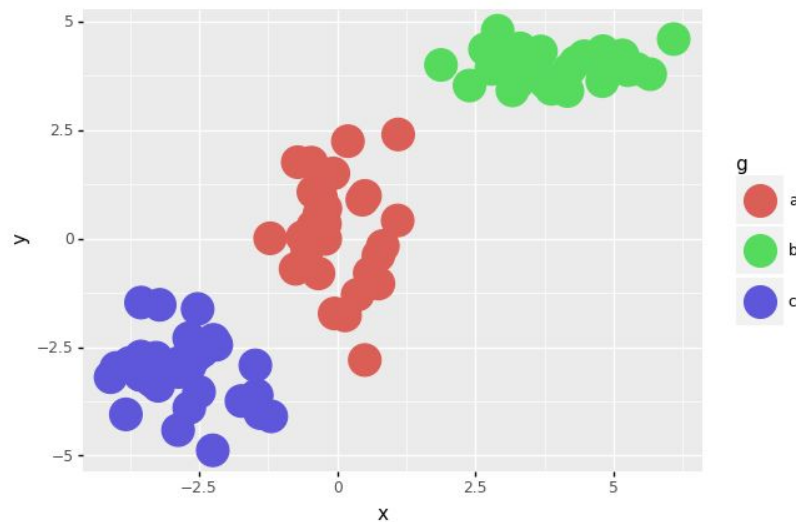
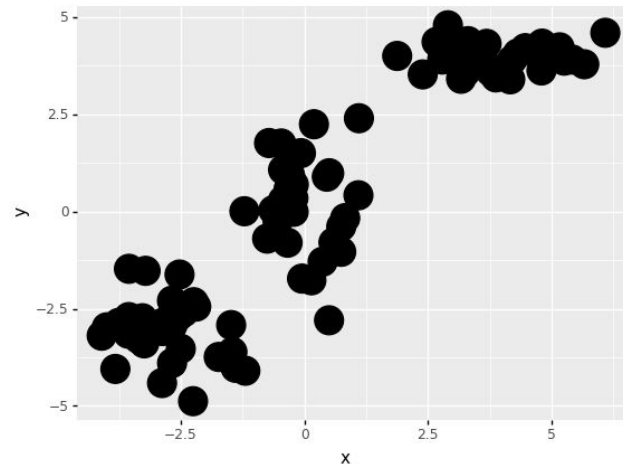


Clustering Overview

Zach Gulde

What is clustering?

- Unsupervised ML -- works with unlabeled data
 - I.e. data without a target
- Grouping observations together so that groups are more similar than not
- Often Exploratory



WHEN IS CLUSTERING USEFUL?

Fresh only



♥ our OWN Brand



Frozen + partying



Marketing, Customer Segmentation

Anomaly Detection

file Access



Document Classification

Support
Uptime
Support
Price

Survey response.. 12:00
Survey response.. 12:01
Survey response.. 1:05
Survey Response.. 1:11
...

THEMES of responses

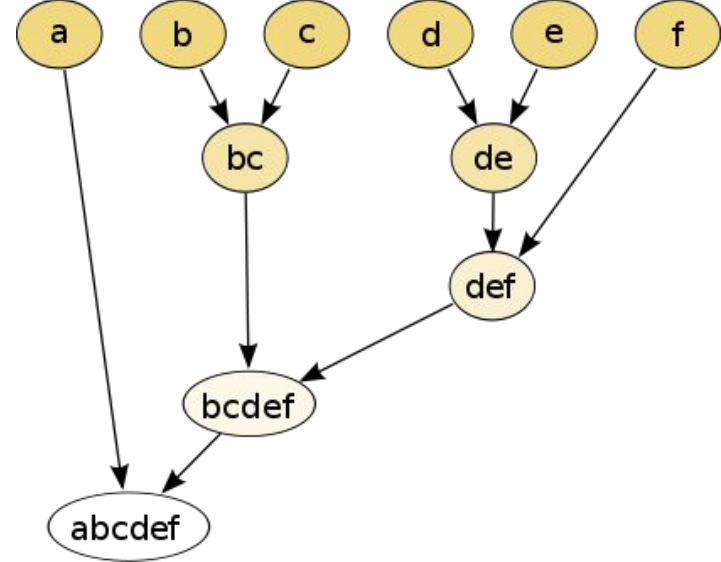
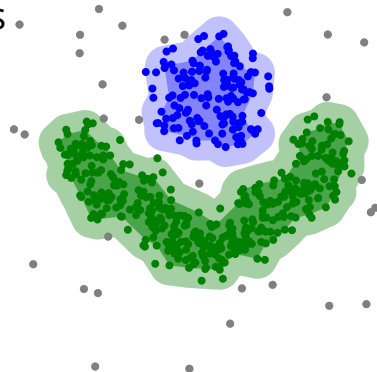
Image Processing



What are the different
Subjects of these photos?

What are common clustering techniques?

- Hierarchical Clustering
 - bottom up or top down
- K Means Video
 - Finds k "centroids"
 - have to choose a k
 - most popular
- DBSCAN Video
 - density-based approach
 - can find arbitrary shapes
 - expensive



Hierarchical Clustering Dendrogram

What do we do with clusters?

For Exploration

- Give names to groups/labels
- Treat as separate subgroups
- How do clusters interact with other features / targets?

For Machine Learning

- Use them as labels
- Use them as features
- Model clusters separately

What else is coming up?

- Wrangling: outlier examination
- Exploration: more question asking, visualization, hypothesis testing
- Modeling: How K-Means works and how to use it
- Clustering Applications
- Project: zillow again, but different goal + clustering