LECTURE 4: CUSTOMER SEGMENTATION

Clustering

Universal AI

Customer Segmentation: Airline

- Who's flying next Monday?
- Who's flying this weekend?
- Who's flying for Thanksgiving?
- Let's think about it from an airline's perspective:
  - What different types of customers might the airline have?
  - How many types are enough to capture customer behavior?
  - How might the airline market differently to different customer types?
- Airlines have access to massive data from frequent flyer program and global distribution systems to answer such questions

Customer Segmentation: Automobile

- Consider an automobile manufacturer that collects information about customers' preferences for their automobile purchases:
  - What different types of customers might the auto manufacturer have?
  - How many types are enough to capture customer behavior?
  - How might the manufacturer target different customer types?
- Auto manufacturer can leverage data from past sales and customer surveys to answer such questions

Customer Segmentation

- Segmentation: subdivision of customers into a relatively small number of groups that share similar characteristics

→ Analytics to **divide the market into meaningful and measurable segments** according to customers' needs, past behaviors, and demographic profiles

→ **Targeted marketing**: design of tailored products/promotions/services to meet customer needs within each segment
  - E-commerce has enabled much more prevalent and personalized use of customer segmentation and targeted marketing

→ Moving from "one size fits all" offerings for competitive edge

Data and Visualization

Airline Customer Data

- Data on 3,999 customers obtained from the loyalty program of a former airline
- Six numerical values describing customers:
  - Number of miles eligible for award travel from historical activity
  - Number of non-flight bonus transactions in the past 12 months
  - Number of miles earned from non-flight bonus transactions in the past 12 months
  - Number of flight miles in the past 12 months
  - Number of flight transactions in the past 12 months
  - Tenure in the loyalty program (days)

Airline Customer Data

n: # of observations ($n = 3,999$)

$n = 3,999$

	Balance	BonusMiles	BonusTrans	FlightMiles	FlightTrans	DaysSinceEnroll
1	48296	31329	9	500	1	3061
2	10021	0	0	0	0	7879
3	49280	22370	16	0	0	3312
4	213539	2750	15	0	0	4751
5	125465	14750	9	0	0	7206
6	7698	0	0	0	0	1734
7	201259	40755	34	0	0	3398
8	350608	50988	26	2643	5	3630
9	146232	83783	19	375	1	3566
10	2080	0	0	0	0	4635
11	93971	62023	22	450	3	4580
12	20999	15914	13	0	0	6206
13	15832	8130	20	500	1	1698
14	207021	3600	4	100	1	5412
15	31504	7358	2	0	0	1022
16	27619	83726	68	14050	46	1325
...	...	...	...	...	...	...
3995	3016	0	0	0	0	1398
3996	28577	48564	14	0	0	3586
3997	276571	42044	23	0	0	7872
3998	28848	0	0	0	0	3069
3999	96522	61105	19	0	0	6924

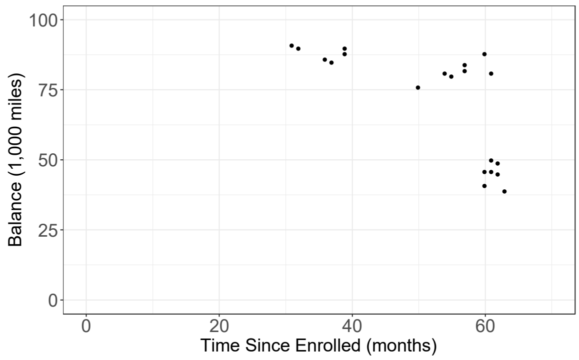
What is Different from Before?

- So far, we have seen dataset with a *dependent variable* that we are trying to predict as a function of *independent variables*
  - These are called **supervised learning** problems
- In segmentation, there is no dependent variable to predict; instead, we aim to create groups of "similar" data points
  - These are called **unsupervised learning** problems
- No training/testing split for unsupervised learning
  - Without a dependent variable, out-of-sample performance is irrelevant
  - Unsupervised learning is used for summary / exploratory analysis
  - Unsupervised models are harder to validate and more subjective

Distances, Clusters, Normalization

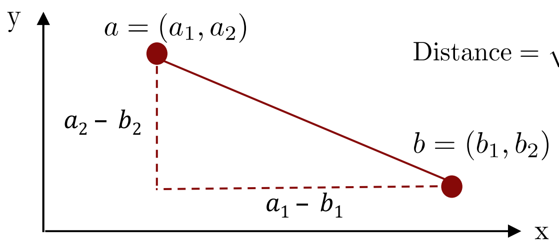
Mini-dataset

- For visualization purposes, let us consider a simple dataset with 20 customers and 2 characteristics



Measuring Distance between 2 Points

- Clustering relies on the *distance* between pairs of observations
  - Typically, Euclidean distance (recall the Pythagorean theorem!)



$$\text{Distance} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

- Can be directly generalized to instances with more variables

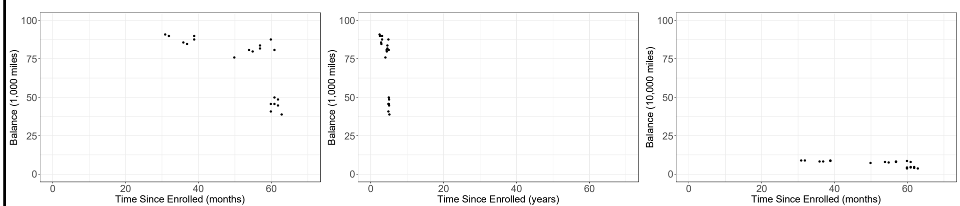
$$\boxed{\text{Distance} \;=\; \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_6 - b_6)^2}}$$

Scaling Issues

- **Issue 1:** Hard to compare different units (e.g., days vs. miles)
- **Issue 2:** Distance primarily driven by the variables of larger scale

Variable	Balance	BonusMiles	BonusTrans	FlightMiles	FlightTrans	Tenure
Mean	73,601	17,145	12	460	1	4,118
Std. dev.	100,776	24,151	10	1,400	4	2,65

- **Issue 3:** The distance metric is sensitive to the choice of units



Normalization

- Normalized variables are expressed in terms of "number of standard deviations from mean"

$$\text{Normalized data} = \frac{\text{Data - Sample mean}}{\text{Sample standard deviation}}$$

Original Data

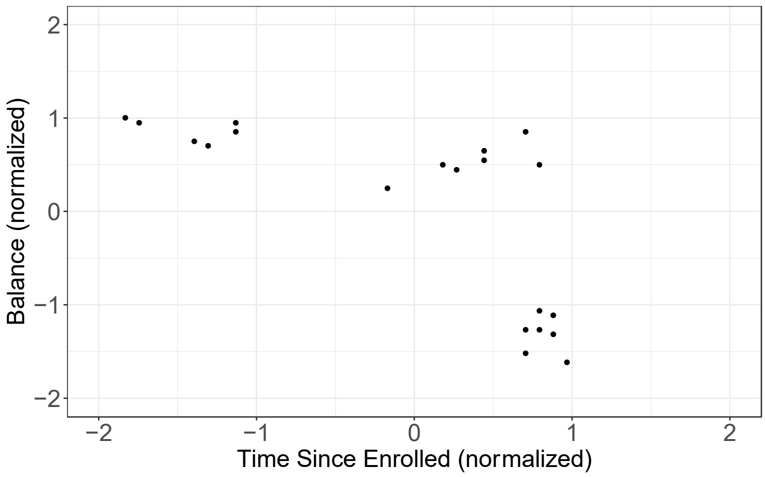
Member Number	Tenure (Months)	Balance (1,000 miles)
1	60	44
2	61	46
3	62	48
4	63	44
5	62	46
…	…	…
19	35	86
20	37	88
Sample mean	**51.85**	**70.90**
Sample S.D.	**11.43**	**19.86**

Normalized Data

Member Number	Tenure (normalized)	Balance (normalized)
1	0.71	-1.51
2	0.80	-1.25
3	0.89	-1.10
4	0.98	-1.61
5	0.89	-1.30
…	…	…
19	-1.30	0.71
20	-1.12	0.86

[For example, 0.71 = (60 – 51.85)/11.43]

Plot of Normalized Mini-dataset



Normalized Airline Customer Data

n: # of observations ($n = 3,999$)

$n = 3,999$

	Balance	BonusMiles	BonusTrans	FlightMiles	FlightTrans	DaysSinceEnroll
1	-0.25	0.59	-0.27	0.03	-0.10	-0.51
2	-0.63	-0.71	-1.21	-0.33	-0.36	1.82
3	-0.24	0.22	0.46	-0.33	-0.36	-0.39
4	1.39	-0.60	0.35	-0.33	-0.36	0.31
5	0.51	-0.10	-0.27	-0.33	-0.36	1.50
6	-0.65	-0.71	-1.21	-0.33	-0.36	-1.15
7	1.27	0.98	2.33	-0.33	-0.36	-0.35
8	2.75	1.40	1.50	1.56	0.96	-0.24
9	0.72	2.76	0.77	-0.06	-0.10	-0.27
10	-0.71	-0.71	-1.21	-0.33	-0.36	0.25
11	0.20	1.86	1.08	-0.01	0.43	0.22
12	-0.52	-0.05	0.15	-0.33	-0.36	1.01
13	-0.57	-0.37	0.87	0.03	-0.10	-1.17
14	1.32	-0.56	-0.79	-0.26	-0.10	0.63
15	-0.42	-0.41	-1.00	-0.33	-0.36	-1.50
16	-0.46	2.76	5.87	9.71	11.76	-1.35
...	...	...	...	...	...	...
3995	-0.70	-0.71	-1.21	-0.33	-0.36	-1.32
3996	-0.45	1.30	0.25	-0.33	-0.36	-0.26
3997	2.01	1.03	1.19	-0.33	-0.36	1.82
3998	-0.44	-0.71	-1.21	-0.33	-0.36	-0.51
3999	0.23	1.82	0.77	-0.33	-0.36	1.36

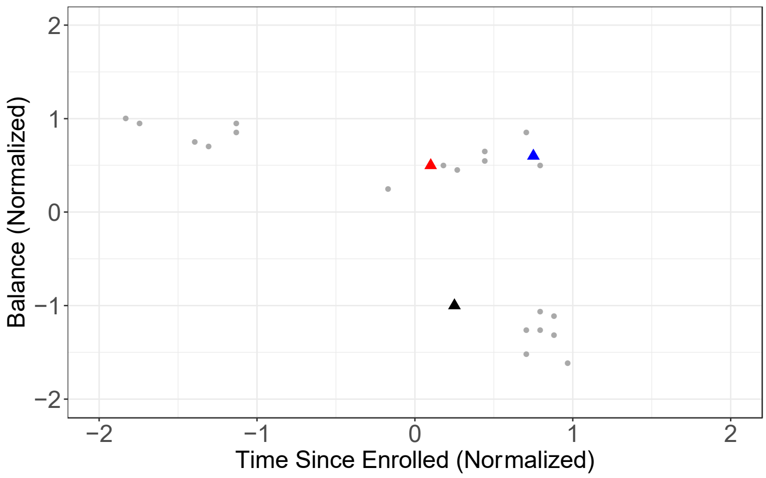
Method 1. k-Means Clustering

Principles

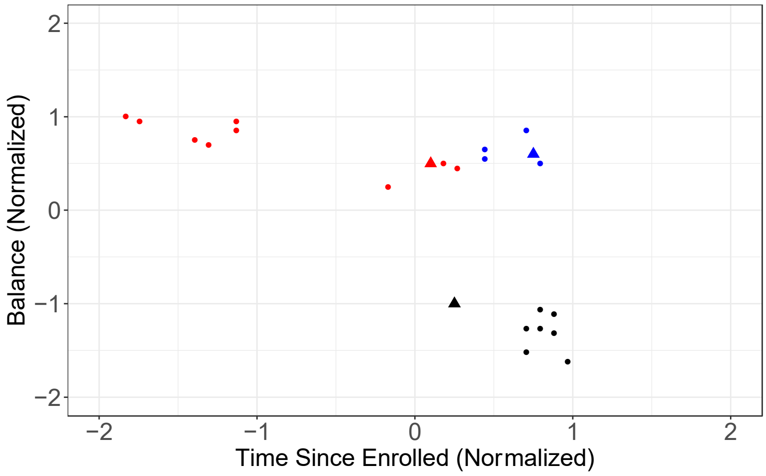
k-Means Clustering

Set number of clusters	• Select a number of clusters, denoted by *k*
Initialize	• Randomly select *k* centroid locations • A "centroid" can be thought of as a point that is "representative" of each cluster • A centroid is not necessarily a data point
Iterate	• Repeat the following two steps, until convergence • Assign each observation to the nearest centroid • Recalculate centroids as average of assigned observations
Terminate	• Terminate when no observations get reassigned

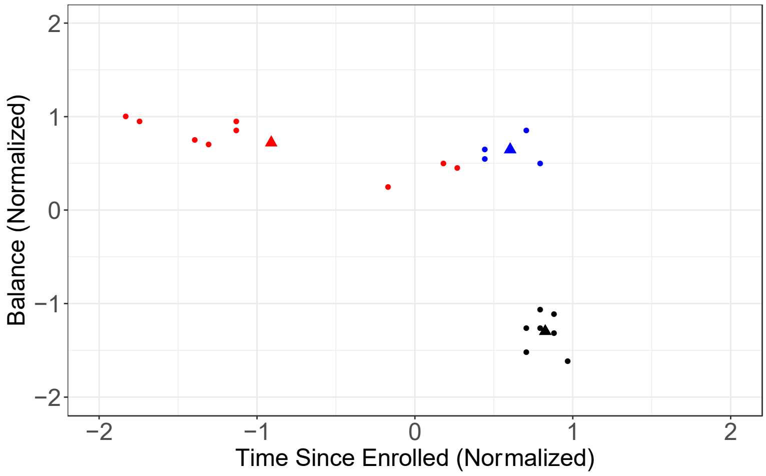
1. Select Random Centroid Locations



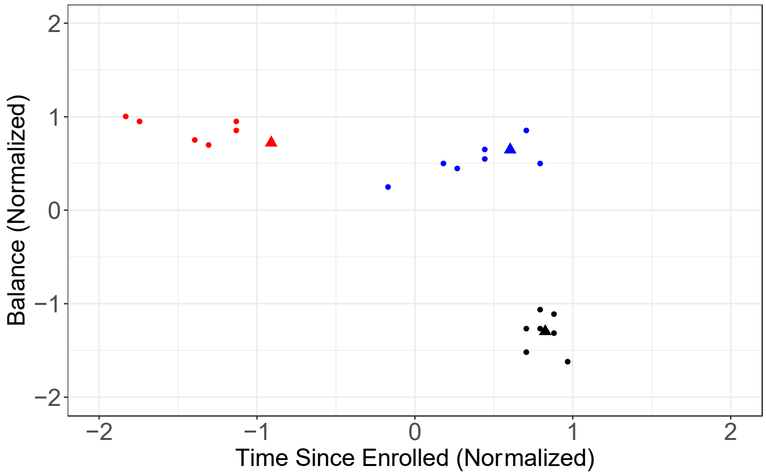
2a. Assign Observations to Centroids



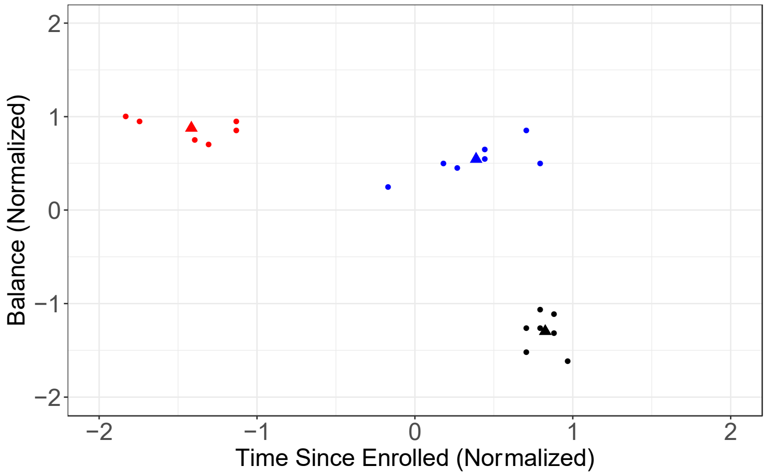
2b. Recalculate Centroids



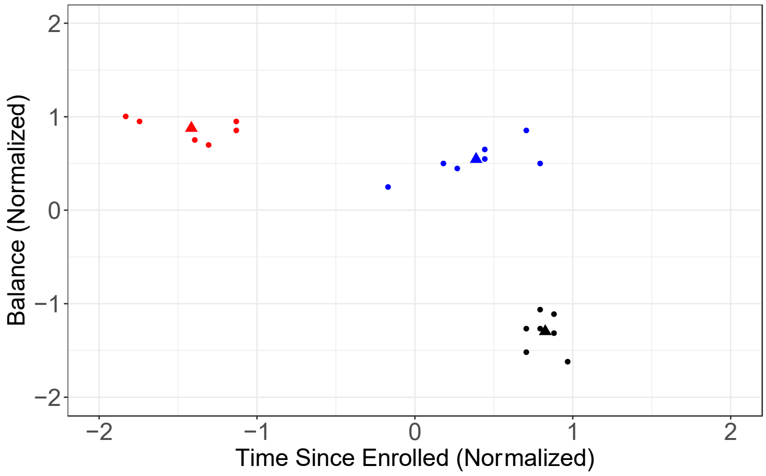
2a. Assign Observations to Centroids



2b. Recalculate Centroids



2a. Assign Observations to Centroids



Method 1. *k*-Means Clustering

Application to Airline Customers Data

Normalized Cluster Centroids

Normalized Variable	Cluster							
	1	2	3	4	5	6	7	8
Balance	-0.12	-0.16	0.54	0.18	-0.42	0.95	4.89	0.95
BonusMiles	0.08	-0.40	1.70	-0.03	-0.61	1.11	1.47	1.21
BonusTrans	0.56	-0.36	0.99	0.55	-0.87	2.20	0.79	3.31
FlightMiles	-0.24	-0.22	-0.09	1.64	-0.25	3.85	0.48	9.84
FlightTrans	-0.27	-0.23	-0.08	1.69	-0.26	4.37	0.72	8.21
DaysSinceEnroll	-0.58	0.95	0.66	-0.08	-0.88	0.50	1.06	-0.33
Cluster Size	**893**	**1,124**	**504**	**212**	**1,107**	**69**	**76**	**14**

- How would you describe these groups of customers?
- How might we market to these clusters?

Clusters 2 and 5

Normalized Variable	Cluster	
	2	5
Balance	-0.16	-0.42
BonusMiles	-0.40	-0.61
BonusTrans	-0.36	-0.87
FlightMiles	-0.22	-0.25
FlightTrans	-0.23	-0.26
DaysSinceEnroll	0.95	-0.88
Cluster Size	**1,124**	**1,107**

Original Variable	Cluster	
	2	5
Balance	57,207	31,165
BonusMiles	7,565	2,308
BonusTrans	8	3
FlightMiles	147	114
FlightTrans	0	0
DaysSinceEnroll	6,074	2,300
Cluster Size	**1,124**	**1,107**

- Dormant customers: low-activity customers across the board

→ Promotional one-time events to incentivize new purchases?

Clusters 1 and 3

Normalized Variable	Cluster 1	Cluster 3
Balance	-0.12	0.54
BonusMiles	0.08	1.70
BonusTrans	0.56	0.99
FlightMiles	-0.24	-0.09
FlightTrans	-0.27	-0.08
DaysSinceEnroll	-0.58	0.66
Cluster Size	**893**	**504**

Original Variable	Cluster 1	Cluster 3
Balance	61,201	127,761
BonusMiles	19,073	58,156
BonusTrans	17	21
FlightMiles	118	333
FlightTrans	0	1
DaysSinceEnroll	2,923	5,484
Cluster Size	**893**	**504**

- The point addicts: focus on bonus transactions

→ Target bonuses for flying? Special offers for bonus transactions?

Clusters 6 and 7

Normalized Variable	Cluster 6	Cluster 7
Balance	-0.35	0.26
BonusMiles	-0.58	0.18
BonusTrans	-0.76	0.46
FlightMiles	-0.22	-0.18
FlightTrans	-0.23	-0.19
DaysSinceEnroll	0.69	1.04
Cluster Size	**875**	**664**

Original Variable	Cluster 6	Cluster 7
Balance	61,201	127,761
BonusMiles	19,073	58,156
BonusTrans	17	21
FlightMiles	118	333
FlightTrans	0	1
DaysSinceEnroll	2,923	5,484
Cluster Size	**893**	**504**

- The old guard: Long-lasting customers with moderate spending

→ Thank for loyalty? Special offers? Refer a friend?

Clusters 4 and 8

Normalized Variable	Cluster	
	4	8
Balance	0.29	0.95
BonusMiles	0.16	1.21
BonusTrans	0.84	3.31
FlightMiles	1.89	9.84
FlightTrans	1.94	8.21
DaysSinceEnroll	-0.01	-0.33
Cluster Size	**211**	**14**

Original Variable	Cluster	
	4	8
Balance	91,719	168,897
BonusMiles	16,360	46,301
BonusTrans	17	43
FlightMiles	2,763	14,244
FlightTrans	8	32
DaysSinceEnroll	3,964	3,446
Cluster Size	**211**	**14**

- The new oil: Recent customers with very high spend

→ Retain, retain, retain: bonus miles, flying challenges, perks, etc.

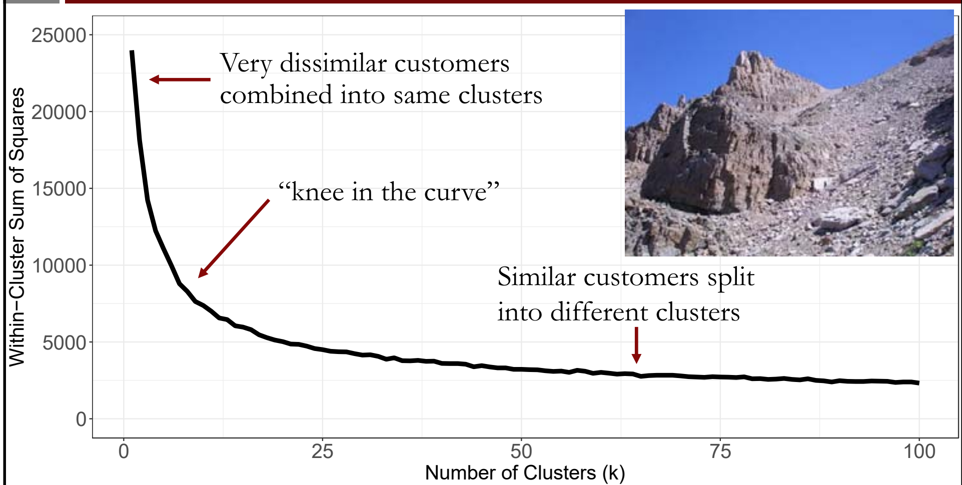
Method 1. *k*-Means Clustering

Selecting the Number of Clusters k

Selecting the Number of Clusters *k*

- Ideally, the clusters should be homogeneous
- Measure of cluster dissimilarity: sum of squared distances of each observation from its cluster centroid
- Trade-off in selecting the number of clusters *k*
  - Low *k*: dissimilar customers will be combined into the same cluster, leading to heterogeneous clusters
  - High *k*: clusters will be too specific, leading to unactionable clusters—and we won't be able to efficiently target clusters
- A *scree plot* displays the tradeoff between number of clusters *k* and cluster dissimilarity—in order to find a "sweet spot"

Scree Plot for Airline Clusters



Method 2. Hierarchical Clustering

Principles

Hierarchical Clustering

Initialize

- Start with each observation in its own cluster
  → Initially, as many clusters as data points (3,999 here)

Iterate

- Iteratively combine the pair of clusters that are have the smallest cluster dissimilarity
  - Prefer combining clusters that are "close" to each other
  - Prefer combining small clusters instead of large clusters

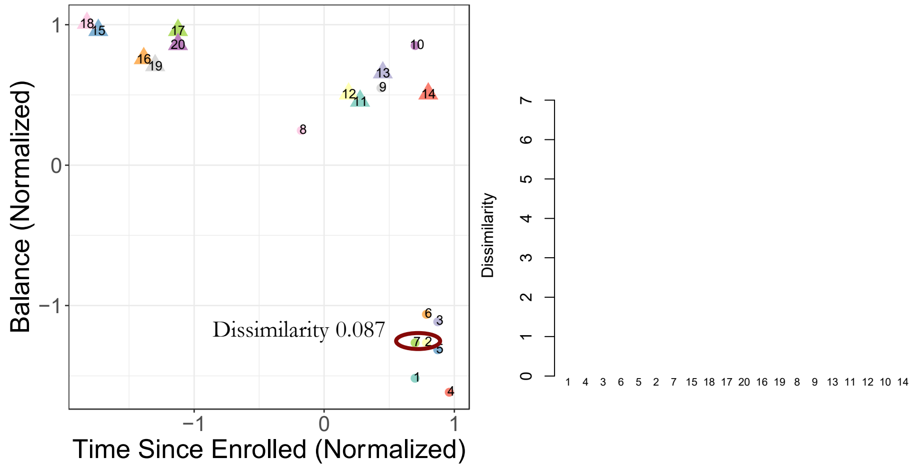
→ Number of clusters goes down by 1 at each iteration

Terminate

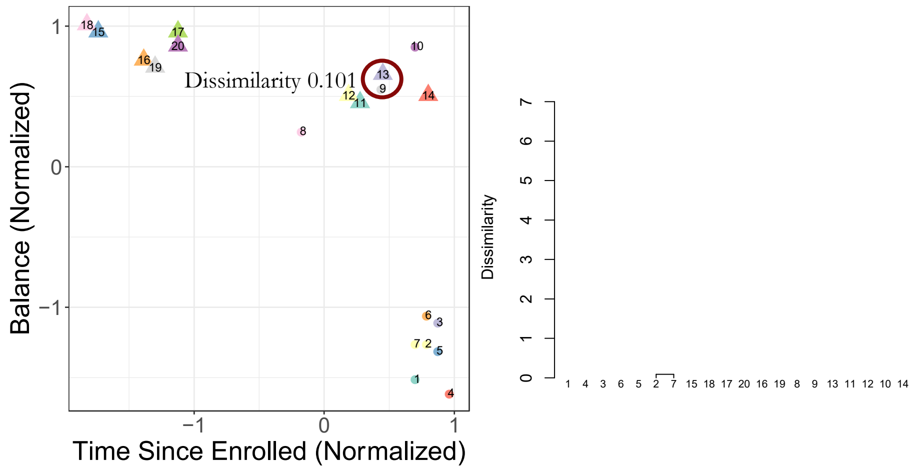
- Ultimately, 1 cluster with all data points

→ Enables to select the appropriate number of clusters

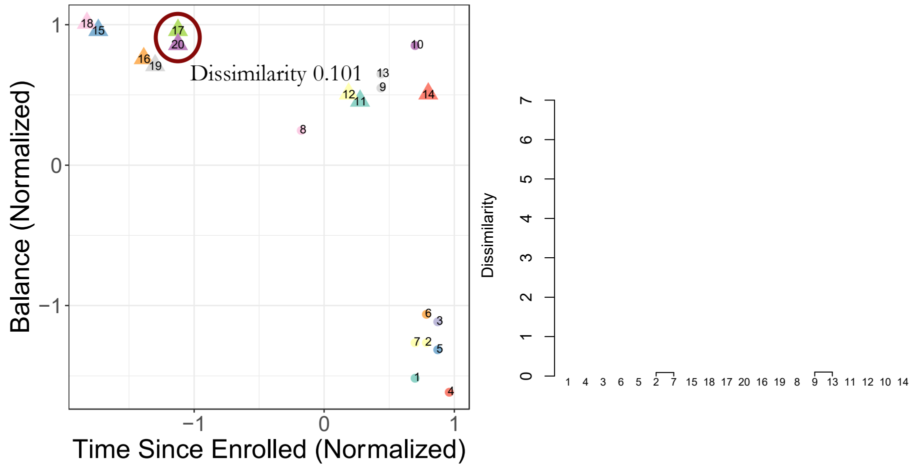
Hierarchical Clustering on Mini-dataset



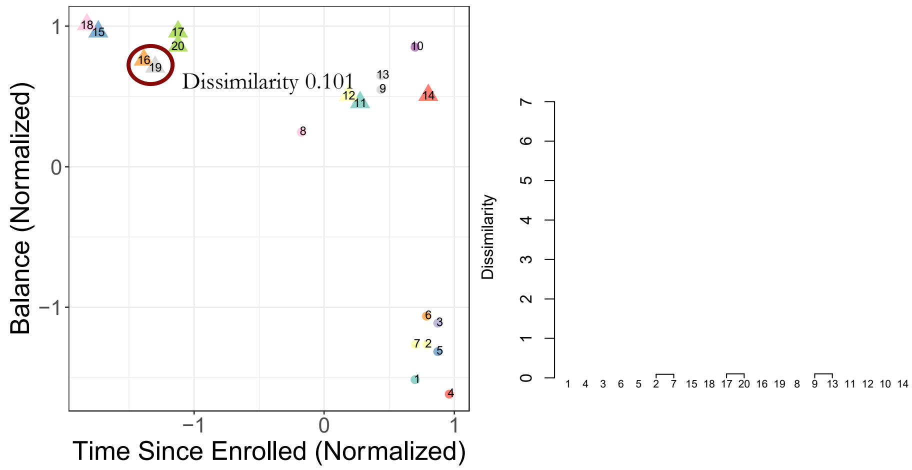
Hierarchical Clustering on Mini-dataset



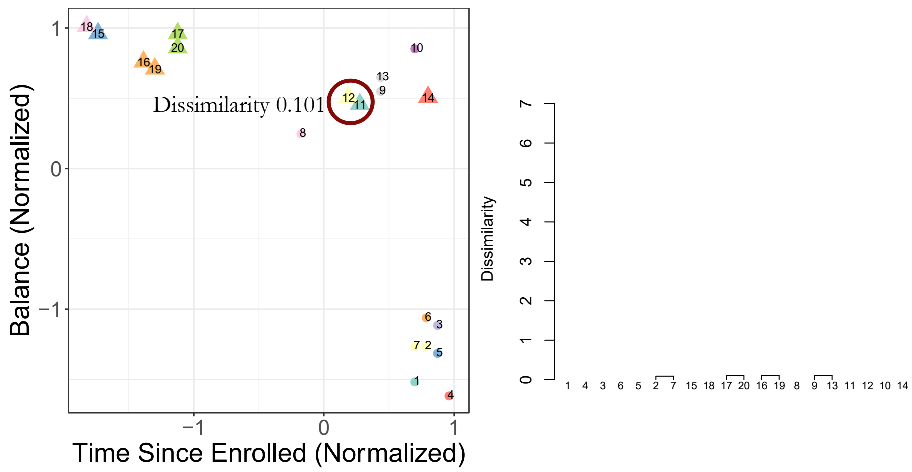
Hierarchical Clustering on Mini-dataset



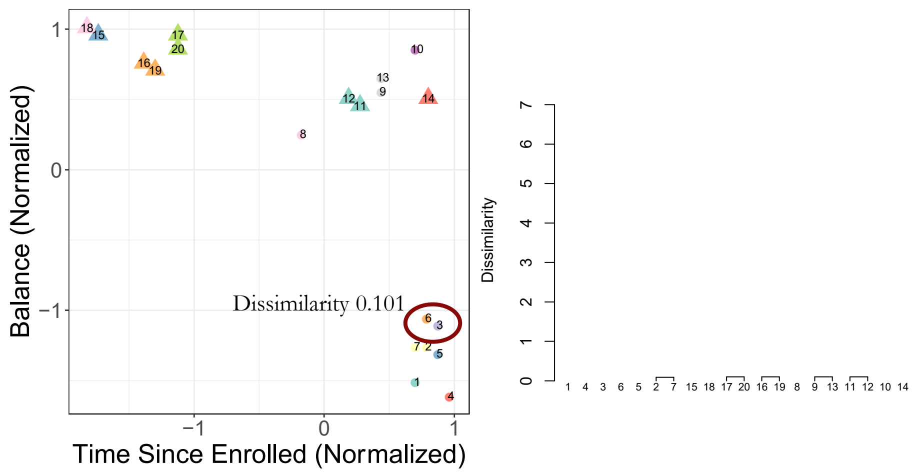
Hierarchical Clustering on Mini-dataset



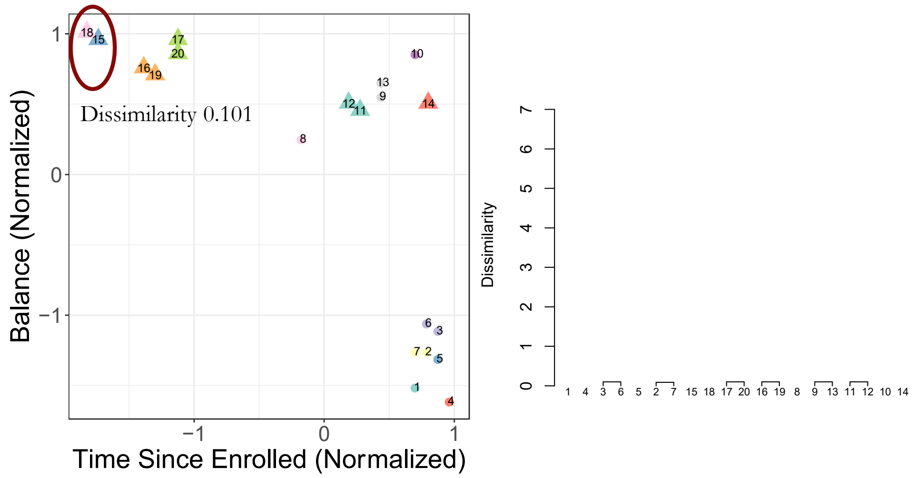
Hierarchical Clustering on Mini-dataset



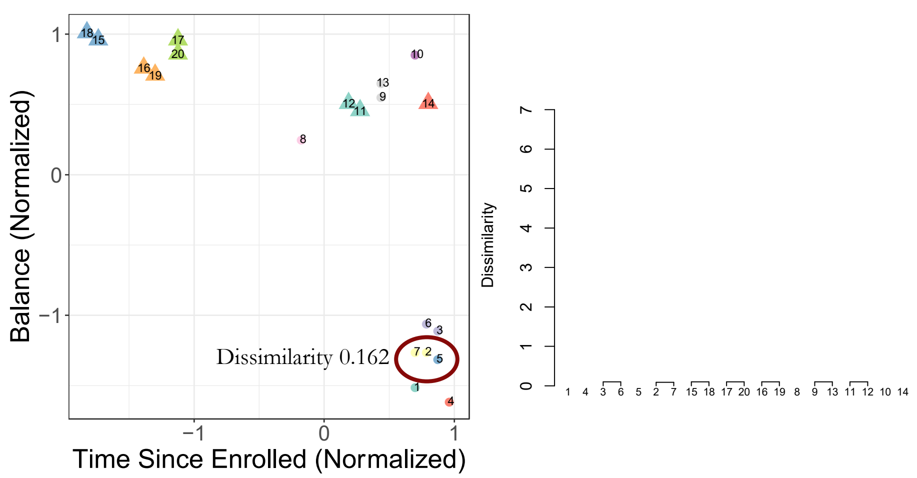
Hierarchical Clustering on Mini-dataset



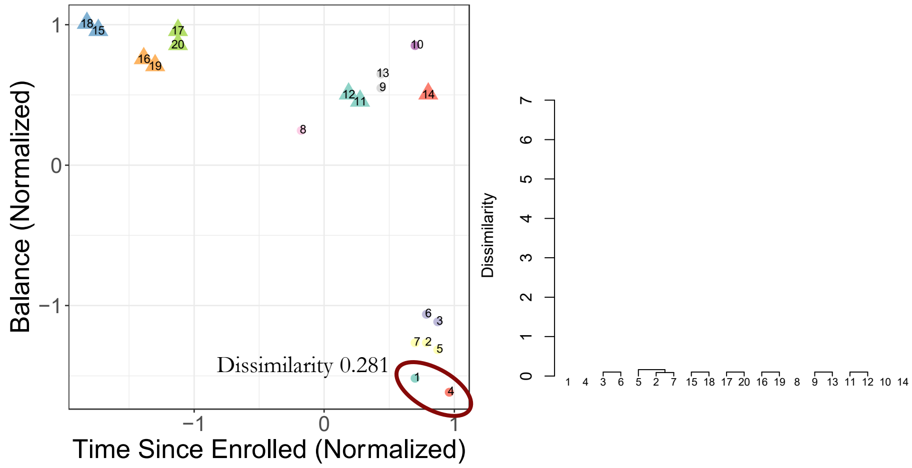
Hierarchical Clustering on Mini-dataset



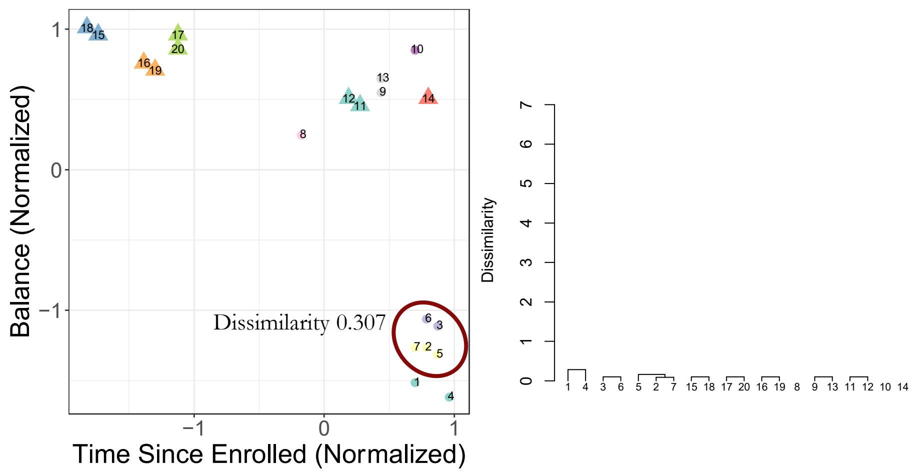
Hierarchical Clustering on Mini-dataset



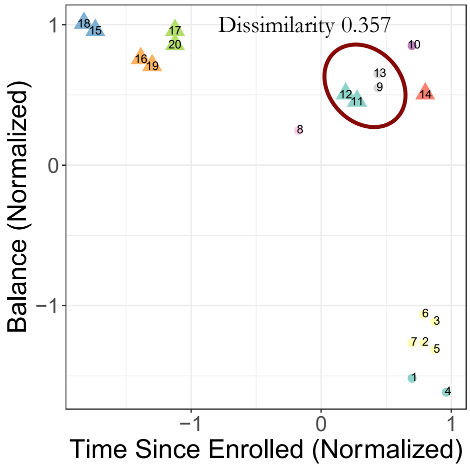
Hierarchical Clustering on Mini-dataset

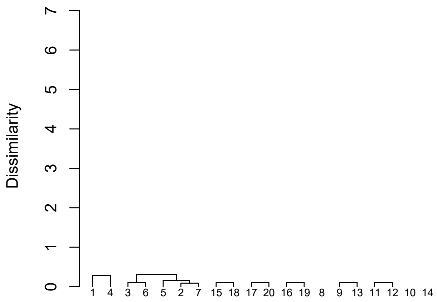


Hierarchical Clustering on Mini-dataset

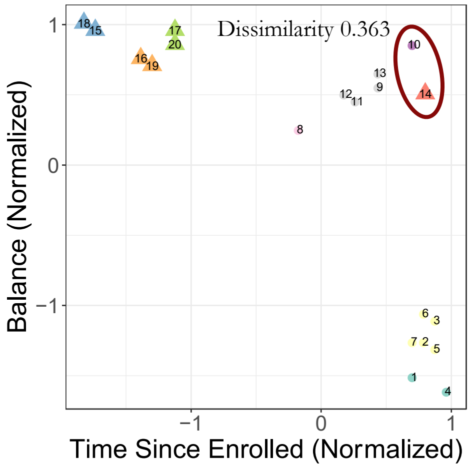


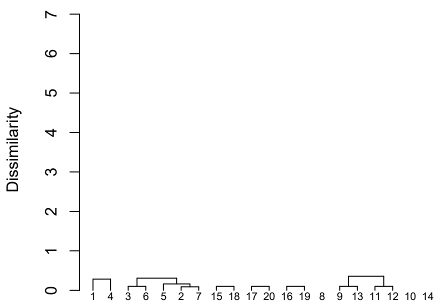
Hierarchical Clustering on Mini-dataset



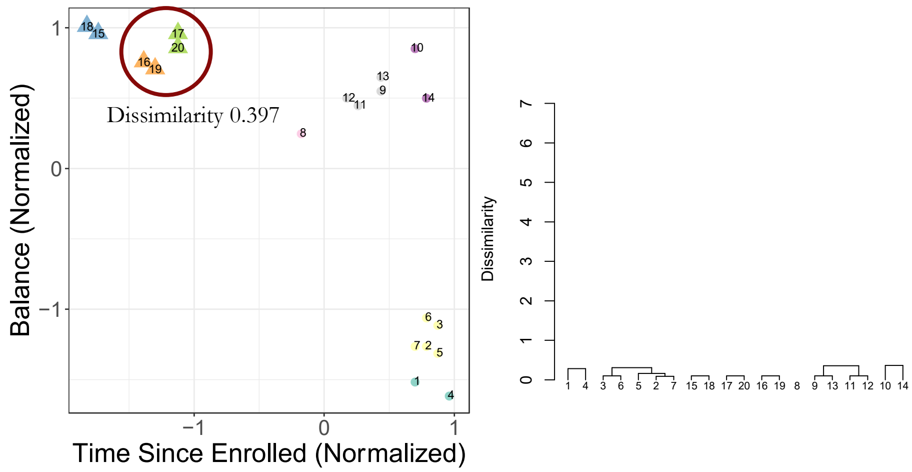


Hierarchical Clustering on Mini-dataset

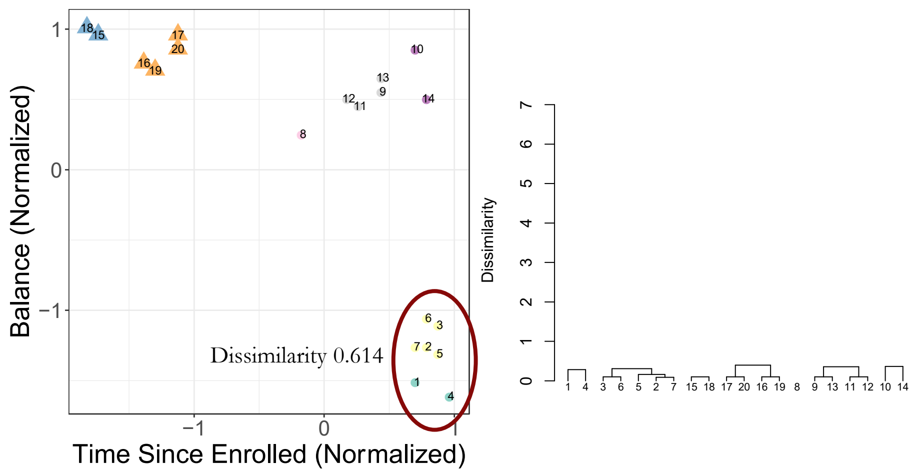




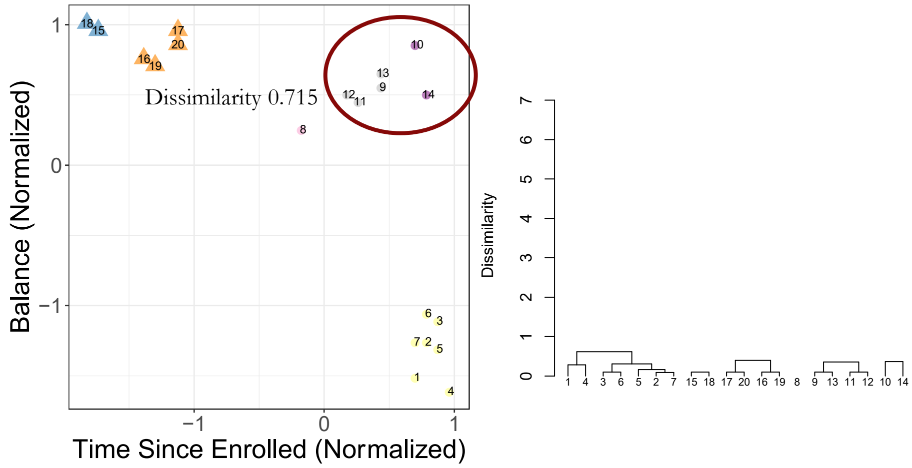
Hierarchical Clustering on Mini-dataset



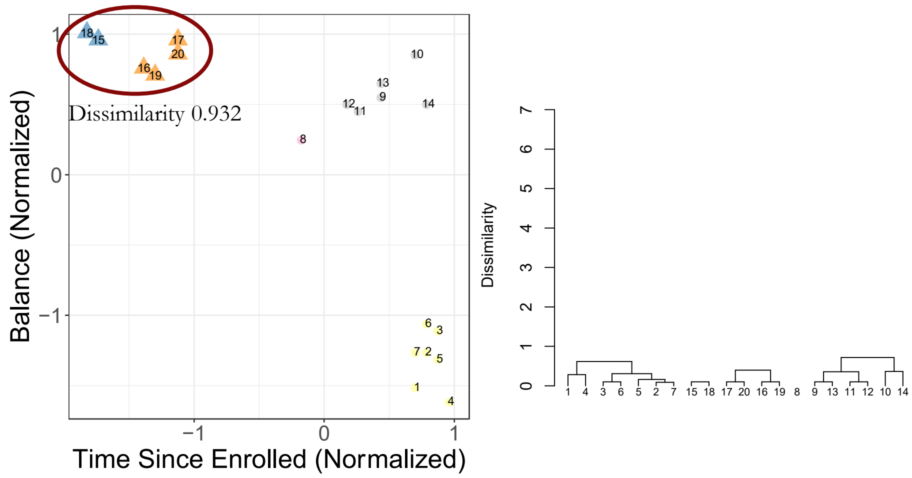
Hierarchical Clustering on Mini-dataset



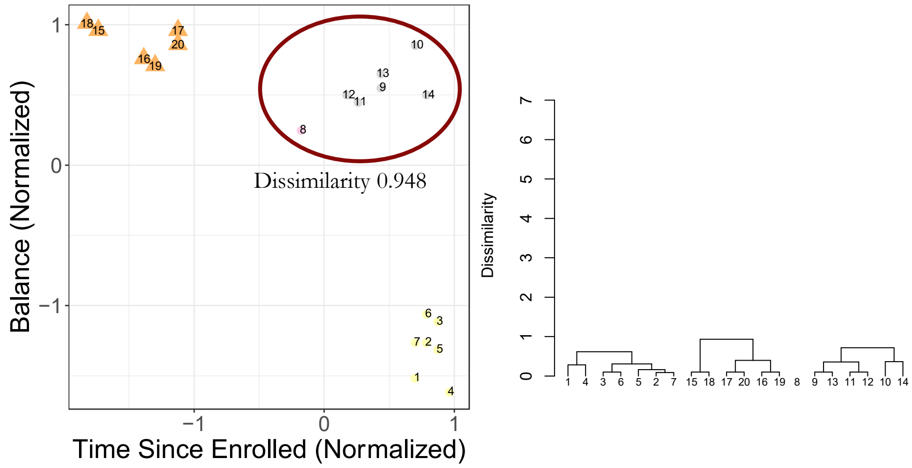
Hierarchical Clustering on Mini-dataset



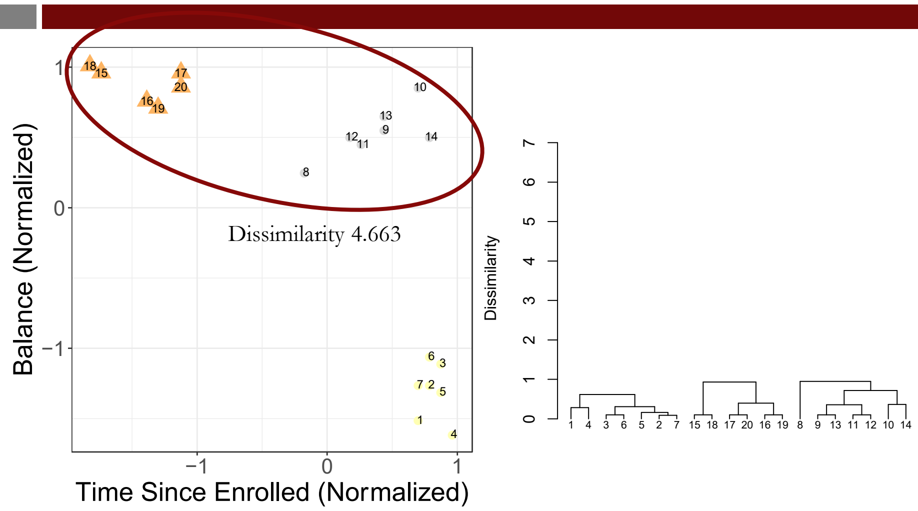
Hierarchical Clustering on Mini-dataset



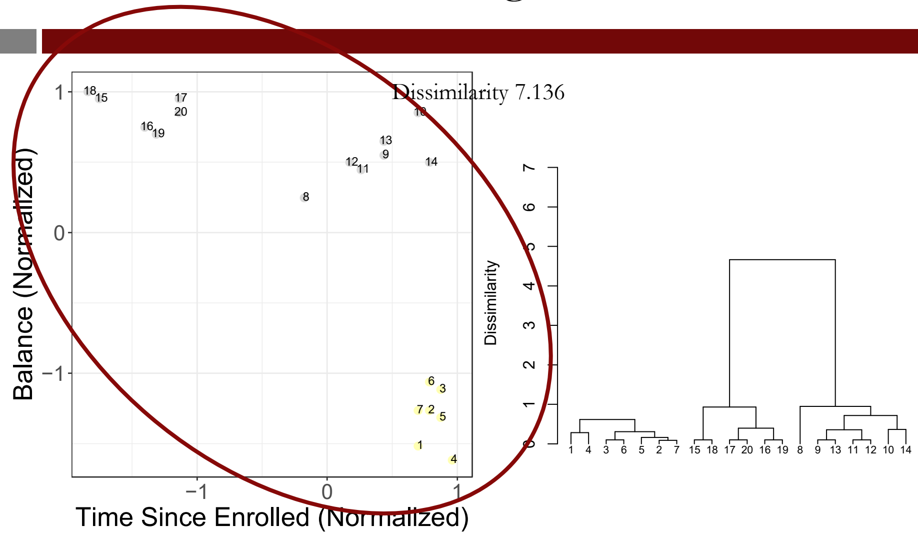
Hierarchical Clustering on Mini-dataset



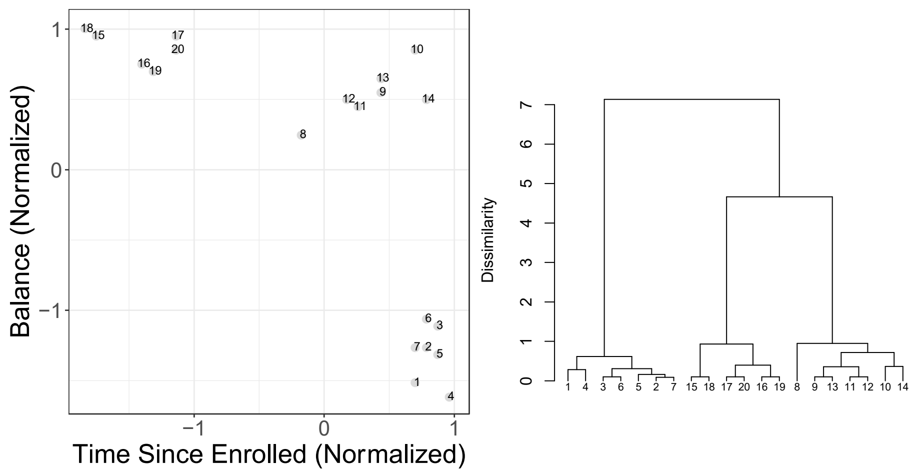
Hierarchical Clustering on Mini-dataset



Hierarchical Clustering on Mini-dataset

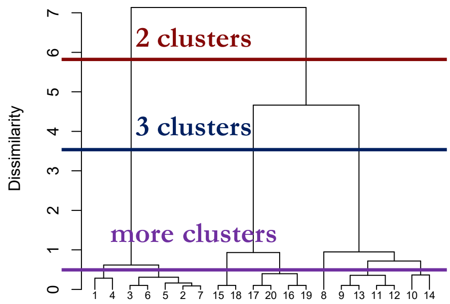


Hierarchical Clustering on Mini-dataset



Use of Dendogram

- Full hierarchy of clusters, all the way to single observations
- Height indicates dissimilarity
  - E.g., 3 & 6 are close to each other, as are 2 & 7
- Common mistake: the x axis does not indicate similarity
  - (1 & 4) are far from (3 & 6)
- One can "reconstruct" a cluster set for each value of *k*



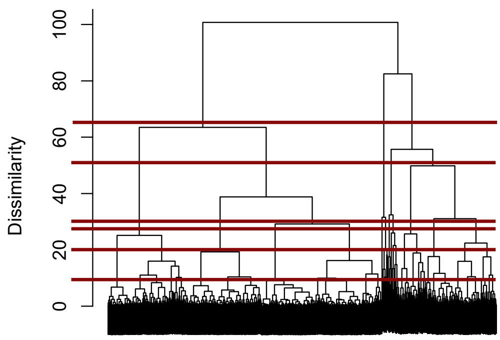
Method 2. Hierarchical Clustering

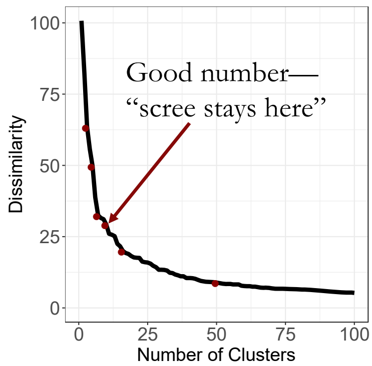
Application to Airline Customers Data

Hierarchical Clusters for Airline Data

- Output of our hierarchical clustering approach with the airline customer dataset:

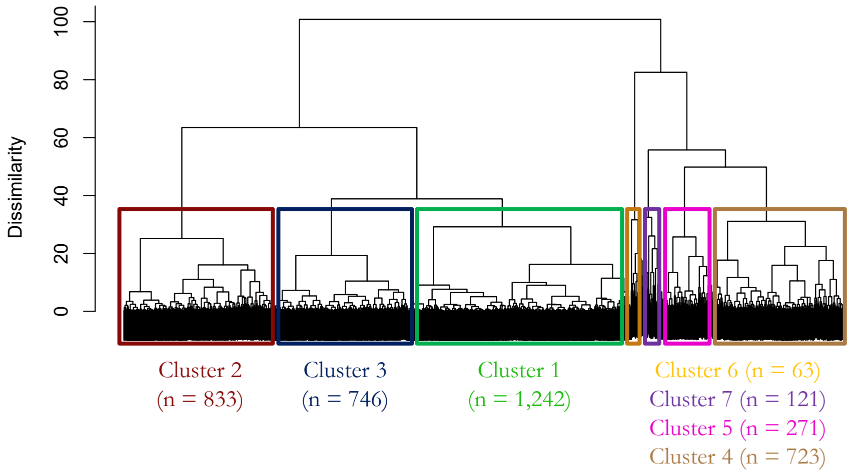
Scree Plot for Airline Clusters





- We will agglomerate any clusters that have dissimilarity 32 or smaller, yielding 7 final clusters

Dendrogram for Airline Clusters



Normalized Cluster Centroids

Normalized Variable	Cluster						
	1	2	3	4	5	6	7
Balance	-0.39	-0.09	-0.30	0.33	0.09	0.67	3.84
BonusMiles	-0.61	-0.29	-0.21	1.16	0.31	0.87	1.54
BonusTrans	-0.93	-0.20	0.27	0.92	0.79	2.35	0.77
FlightMiles	-0.23	-0.21	-0.27	-0.21	1.41	5.49	0.72
FlightTrans	-0.24	-0.23	-0.29	-0.21	1.49	5.59	0.90
DaysSinceEnroll	-0.61	1.18	-0.64	0.16	0.04	0.14	0.90
Cluster Size	**1,242**	**833**	**746**	**723**	**271**	**63**	**121**

- How are the clusters from hierarchical clustering similar to or different from those obtained with *k*-means clustering?

Clustering Automobile Customers

Opinions About Products

- We segmented airline customers based on **observed behavior** (bonus transactions, flights, …)
- We can also segment customers (or potential customers) based on their **stated preferences**—that is, what is important to them
  - → Design of targeted advertisements to users with particular preferences
  - → Design of products and services to satisfy customer needs
- Data for such segmentation often comes from surveys
  - Pros: they enable to test future and "what if" scenarios
  - Cons: they measure customers' intentions, not actual behaviors
- We use survey data from a German premium car manufacturer

Automobile Customer Data

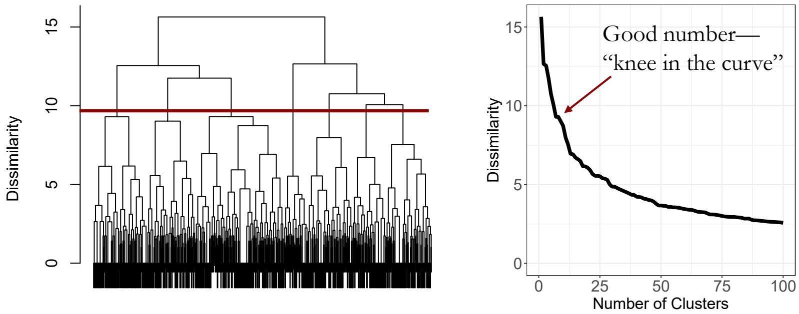
- Survey data on 793 customers based on their opinions about a variety of automobile features
  - All surveyed customers had purchased a car from the same German premium car manufacturer within 3 months
- Ten data fields indicating whether the surveyed customers found each of the following important:
  - Driving properties, interior, technology, comfort, reliability, handling, power, gas consumption, sportiness, safety
- Data are categorical (binary) here:
  - 1 if the factor was deemed important by the customer, 0 otherwise

Automobile Customer Data

n: # of observations ($n = 793$)

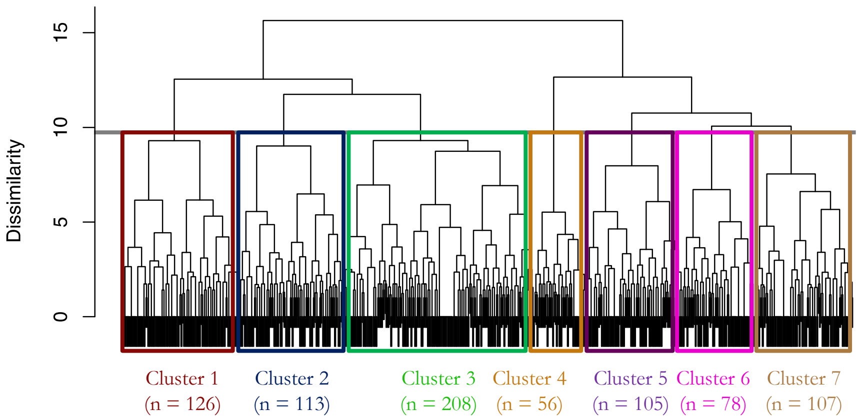
	driving_properties	interior	technology	comfort	reliability	handling	power	consumption	sporty	safety
1	0	0	1	0	1	0	0	0	1	1
2	1	0	0	0	0	1	0	0	0	1
3	1	0	0	1	0	0	1	0	1	1
4	0	1	0	1	1	0	0	0	1	1
5	0	1	1	1	0	0	0	0	0	0
6	1	1	1	1	0	0	0	0	1	0
7	0	0	0	0	0	1	1	0	0	1
8	1	0	1	0	1	0	0	0	0	1
9	0	1	1	1	1	0	0	0	0	0
10	1	0	0	0	1	0	0	0	0	1
11	1	0	0	0	0	1	0	0	1	1
12	1	0	0	0	0	0	1	1	1	1
13	1	0	0	0	0	1	1	0	1	0
14	1	0	0	0	1	0	1	0	0	1
15	0	1	0	0	0	0	1	0	0	0
16	1	0	0	0	1	1	0	0	0	1
...	...	...	...	...	...	...	...	...	...	...
789	1	0	1	0	1	0	1	0	0	1
790	0	0	0	0	0	0	0	0	0	0
791	1	0	0	0	0	0	1	1	0	1
792	1	1	0	1	1	1	1	0	1	0
793	1	0	0	0	0	1	1	0	1	0

Scree Plot for Automobile Clusters



- We will agglomerate any clusters that have dissimilarity 9.4 or smaller, yielding 7 final clusters

Dendrogram: Automobile Customers



Cluster Centroids (Not Normalized)

Variable	Cluster						
	1	2	3	4	5	6	7
driving_properties	0.20	**0.99**	**0.80**	0.61	**0.89**	**0.97**	**1.00**
interior	0.25	0.12	0.03	**0.71**	0.02	0.12	**0.61**
technology	0.34	0.52	0.47	0.35	0.36	0.44	**0.91**
comfort	0.11	0.30	0.51	**0.82**	0.32	0.26	**0.80**
reliability	0.49	0.46	0.53	0.27	0.10	0.40	**0.66**
handling	0.16	0.21	0.01	0.13	0.30	**0.96**	**0.86**
power	0.32	0.41	**0.81**	0.25	0.43	0.73	**0.93**
consumption	0.21	0.01	0.18	0.17	**0.75**	0.09	**0.62**
sporty	0.32	0.21	**0.79**	0.31	0.25	0.53	**0.77**
safety	0.38	**0.89**	0.14	0.35	0.21	0.06	**0.79**
Cluster Size	**208**	**126**	**107**	**113**	**105**	**78**	**56**

The AI Edge

Clustering & Customer Segmentation

- Clustering for customer segmentation is now pervasive
  - Better understanding of key customer types and their prevalence
  - Design of appropriate marketing strategies, tailored to each segment
- Key steps to cluster analysis:
  - Aim for data to be compatible in scale through normalization
  - Identify relevant clusters with *k*-means or hierarchical clustering
  - Analyze cluster centroids to interpret the type of customers in each cluster and determine appropriate managerial interventions
- Clustering cannot be subject to out-of-sample predictions
  → Reinforced need to engage decision-makers throughout model building

Comparison of Clustering Methods

- *k*-means & hierarchical clustering often yield consistent insights
- Still, the two methods have different strengths
- Hierarchical clustering
  - Flexibility through various distance and cluster dissimilarity metrics
  - Amenable to iterative decision making—at various levels of granularity
- *k*-means clustering
  - Is more computationally efficient—hence scalable to large datasets
- Many modeling questions can have significant impacts on results
  - whether to normalize data, how to quantify "distance", how many clusters to use, etc.

Main R Commands

```
library(flexclust)
airline <- read.csv("AirlinesCluster.csv")
pp <- preProcess(airline, method=c("center", "scale"))
airline.scaled <- predict(pp, airline)
mod <- kmeans(airline.scaled, iter.max=100, 8)
cluster.assignment.kmeans <- mod$cluster
d <- dist(airline.scaled)
mod.hclust <- hclust(d, method="ward.D2")
plot(mod.hclust,labels=F,xlab=NA,ylab="Dissimilarity")
dat.hc.airline <- data.frame(nclust = seq_along(mod.hclust$height),
                   dissimilarity = rev(mod.hclust$height))
ggplot(dat.hc.airline, aes(x=nclust, y=dissimilarity))+geom_line()
cluster.assignment.hierarchical <- cutree(mod.hclust, 7)
```

library

normalization

k-means

assignment

hierarchical clustering

dendogram

scree plot

assignment