

Identifying family clusters within the US Holstein population to manage genetic diversity

Yvette Steyn, T. Lawlor, Y. Masuda, S. Tsuruta, D.A.L. Lourenco, I. Misztal

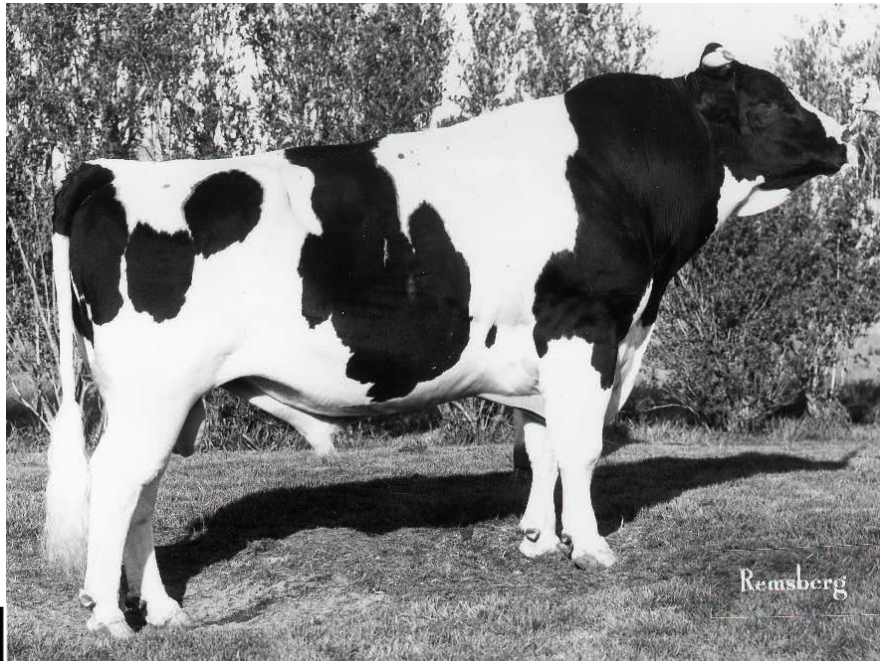


UNIVERSITY OF
GEORGIA



AI reach

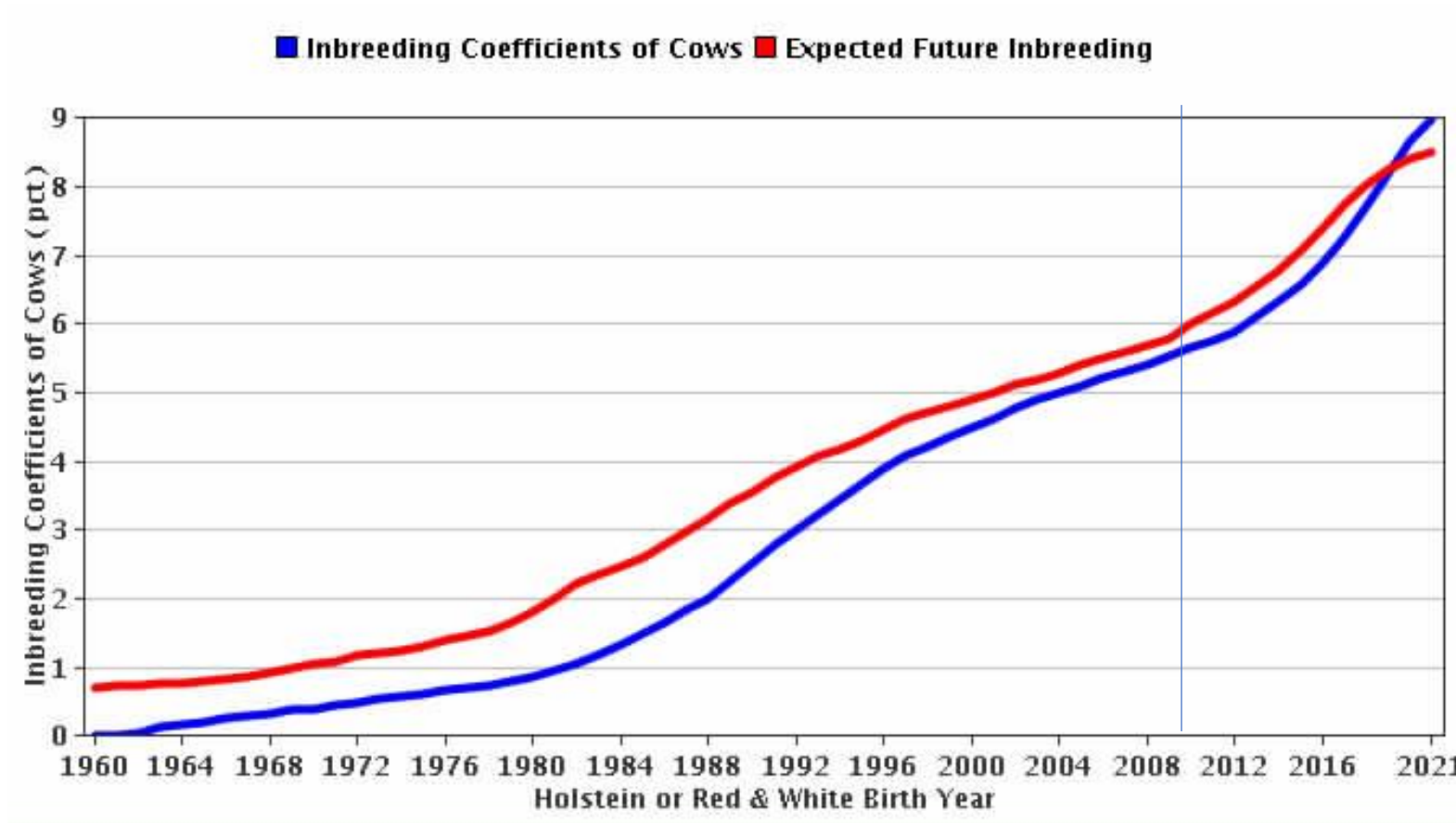
Round Oak Rag Apple **Elevation**



Pawnee Farm Arlinda **Chief**



Introduction



Objectives



Identify potential family clusters in the US Holstein population and their key founders

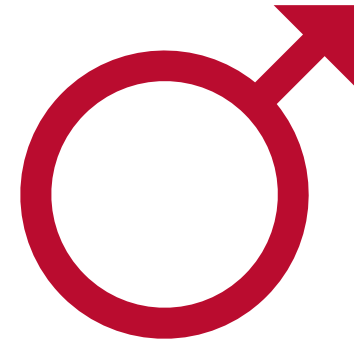


Evaluate the expected inbreeding that would result from within- or across-cluster mating



Data

- Sires of animals born after 1985
- Genotyped sires with most progeny
- 1,145 Sires
- Nr progeny: 312 to 49,146
- 7 Countries



Relationships

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2 \sum p_i (1 - p_i)}$$

G: Genomic relationship matrix

Z : SNP marker matrix (0,1,2) minus 2 times the allele frequency

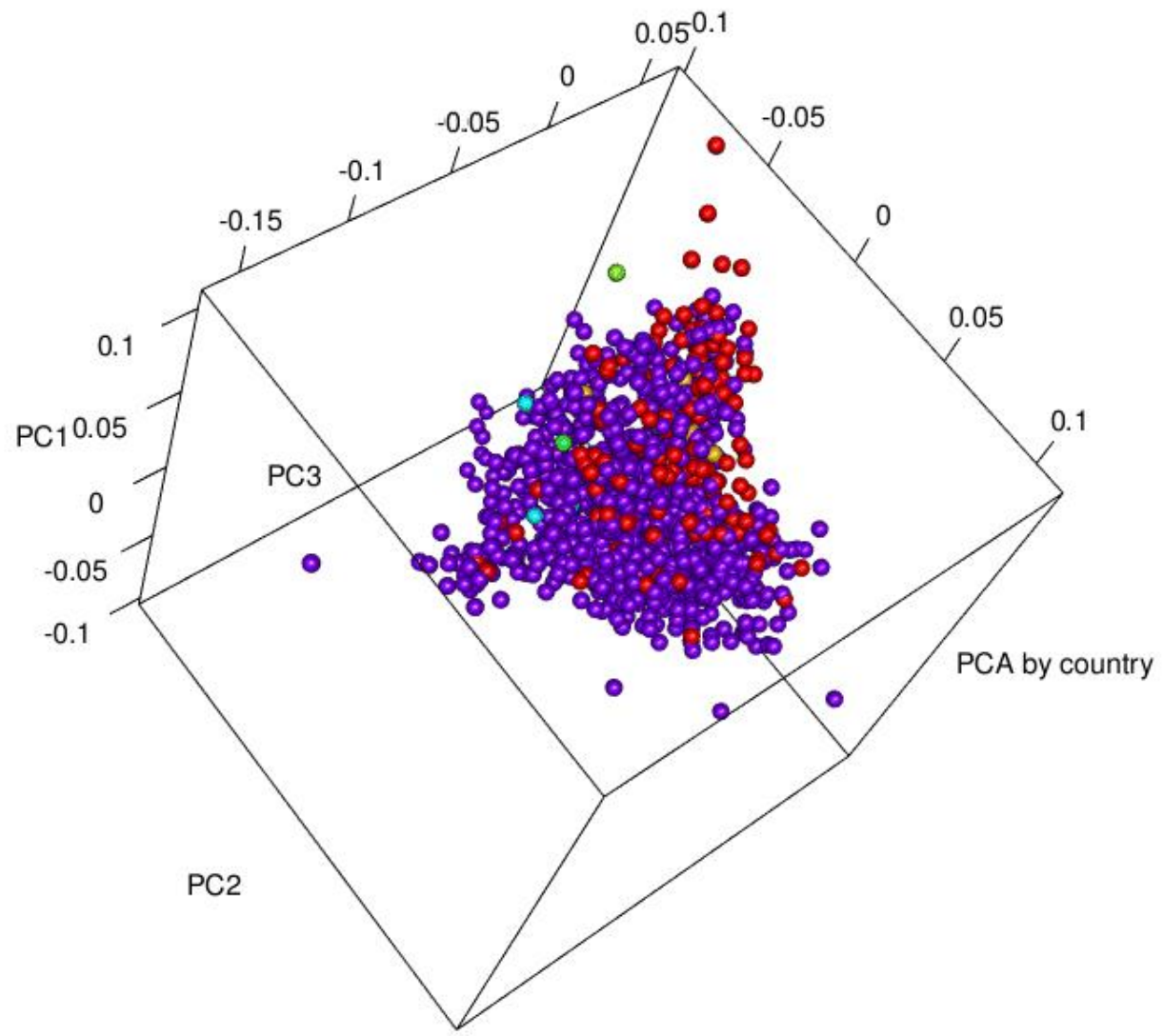
p_i : allele frequency of SNP i



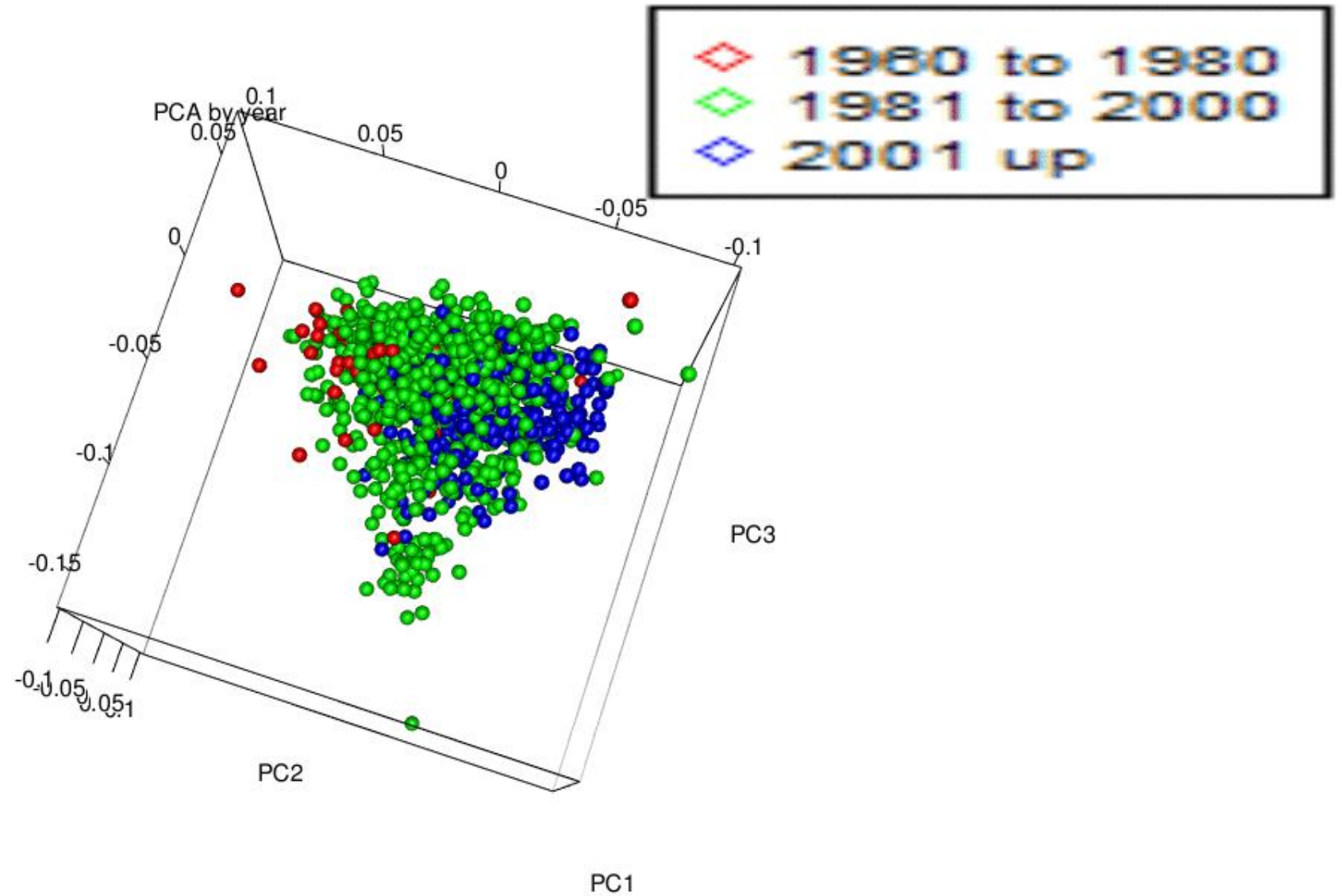
Principal Component Analysis

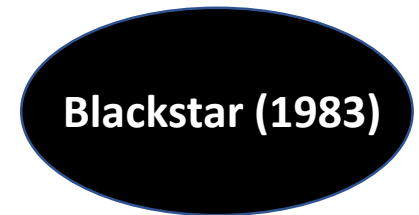
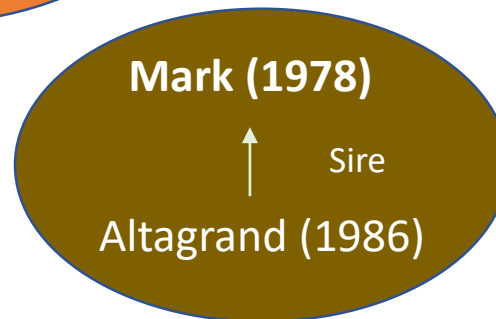
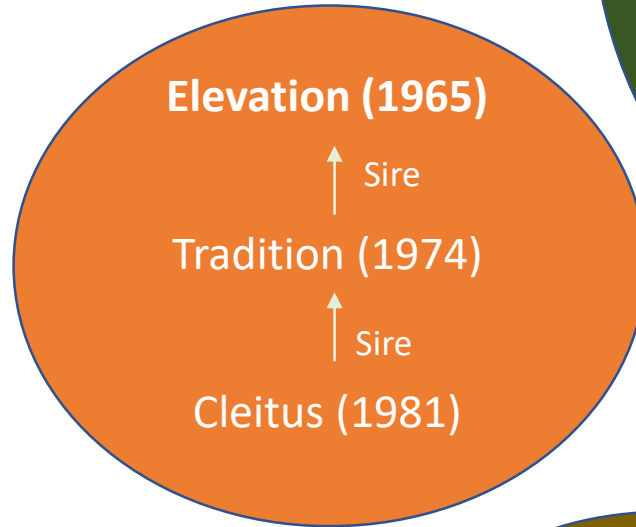
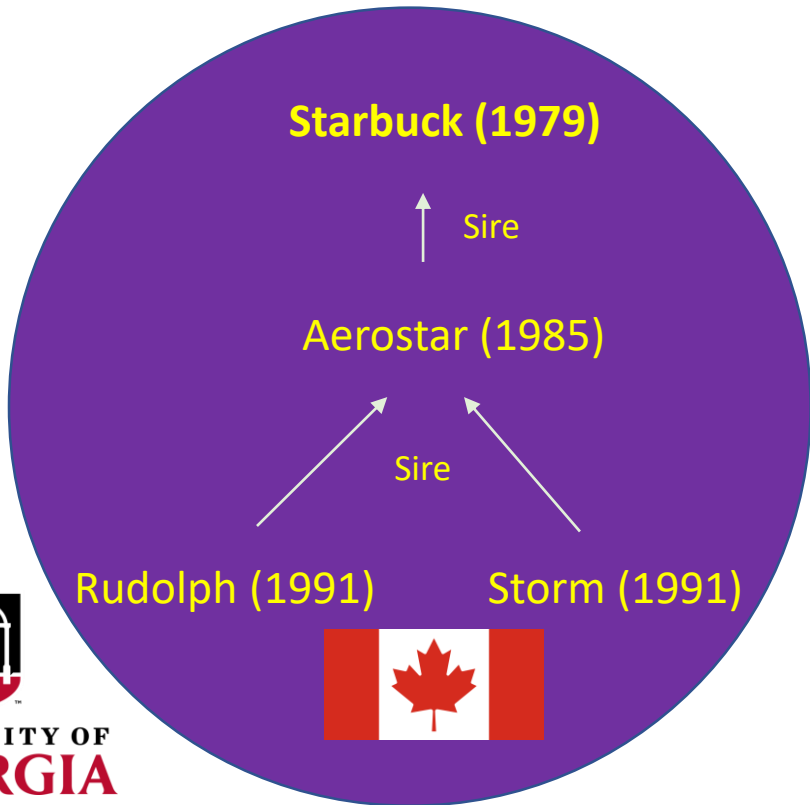
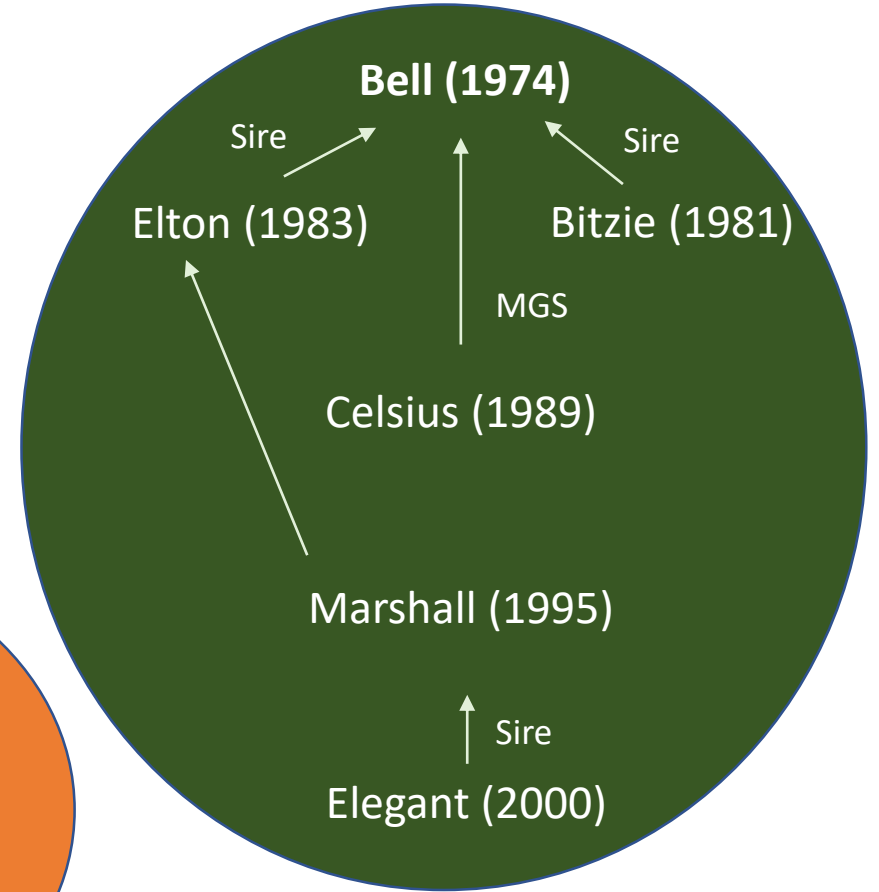
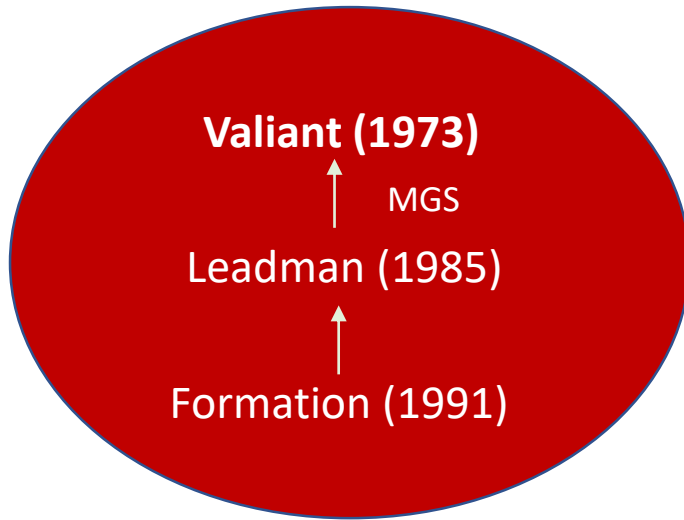
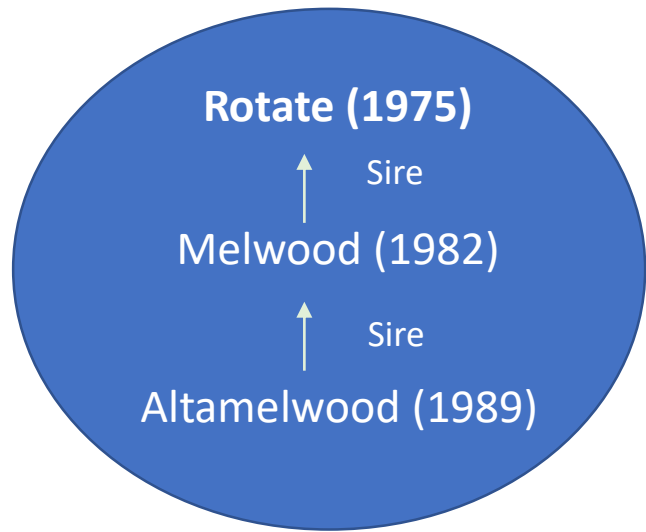
- Eigenvalue decomposition
- 1st PC:
 - Captures most variance
- 2nd PC:
 - Uncorrelated to PC1
 - Captures 2nd most variance





PC1 to PC3

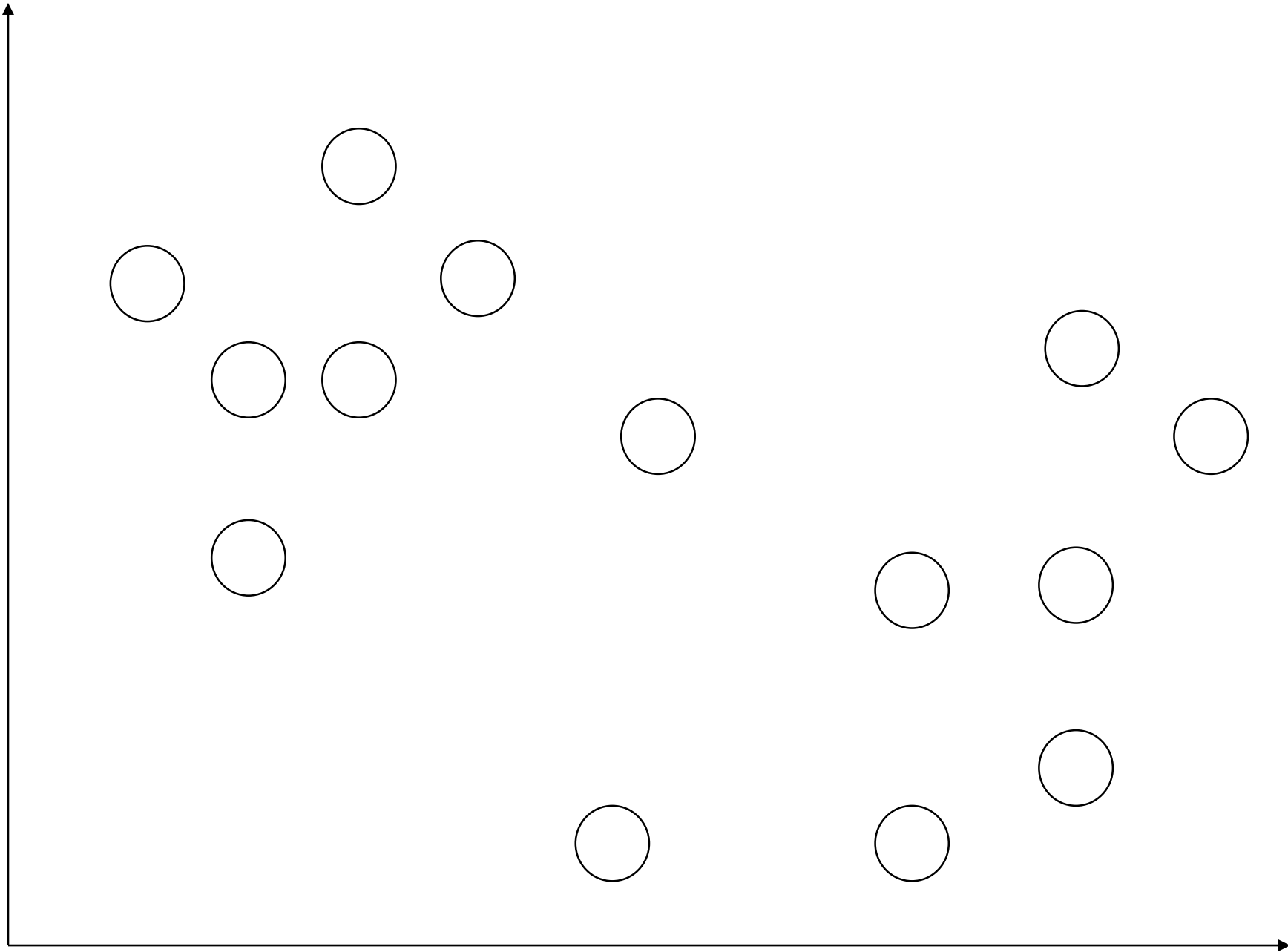


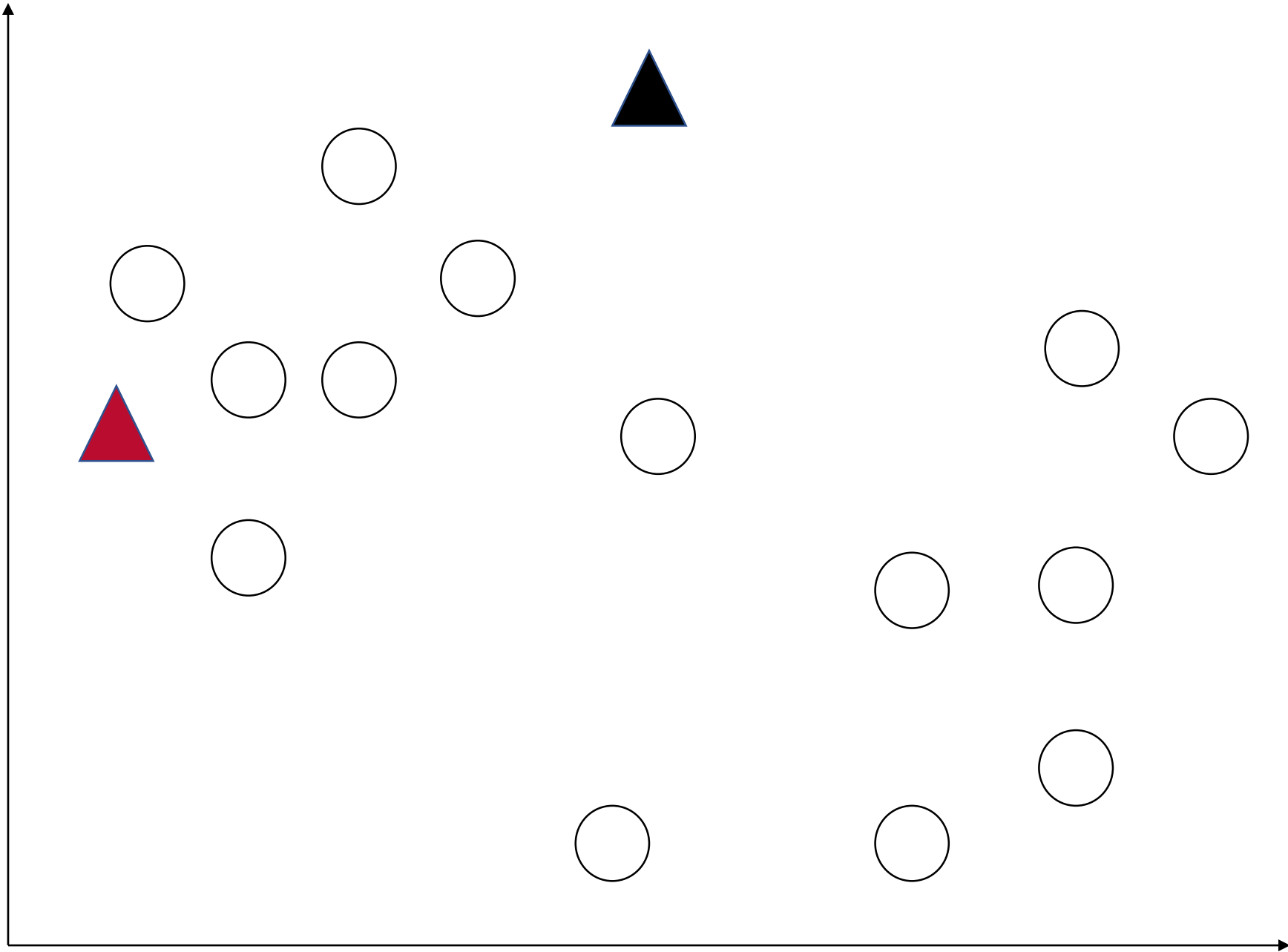


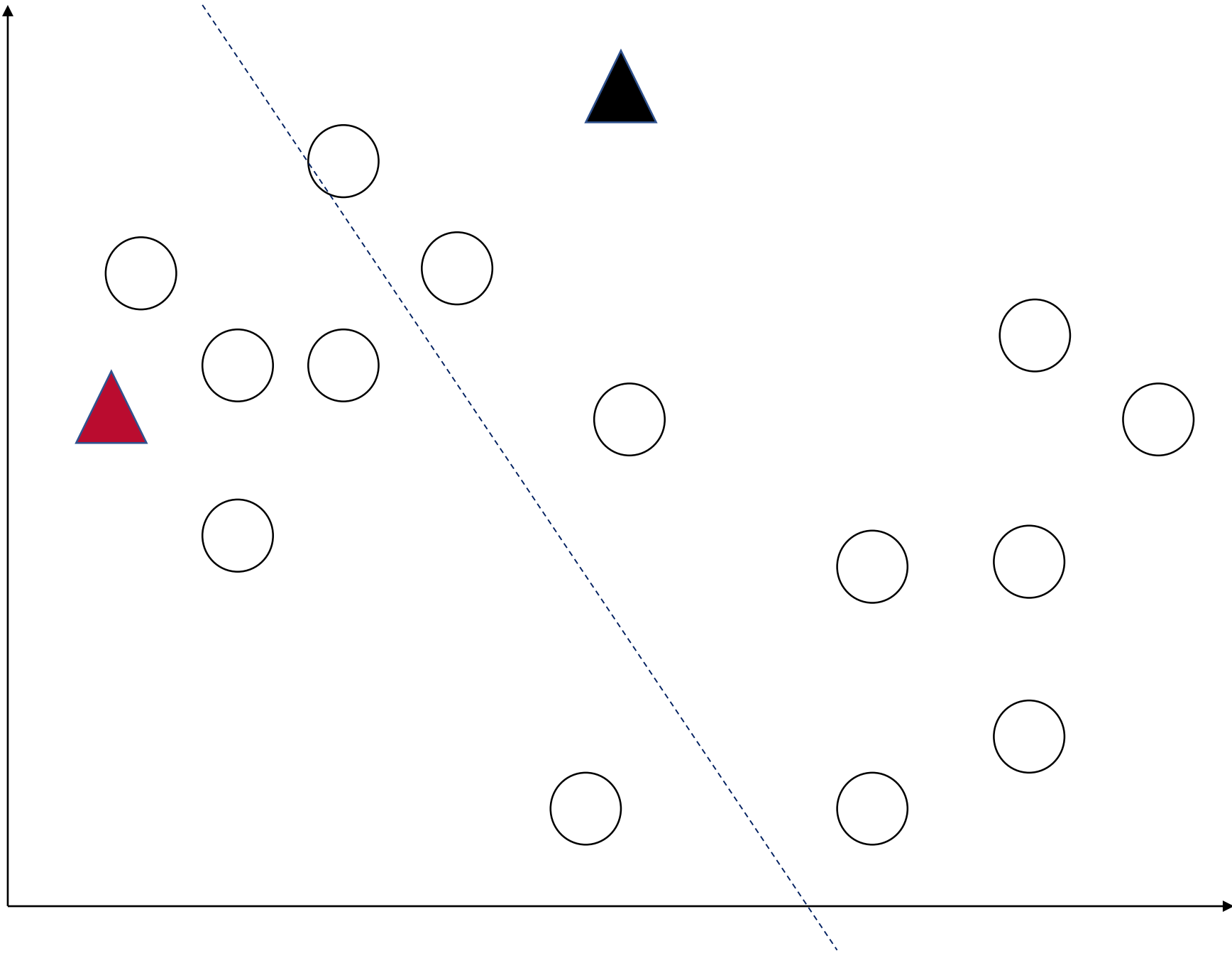
K-means clustering

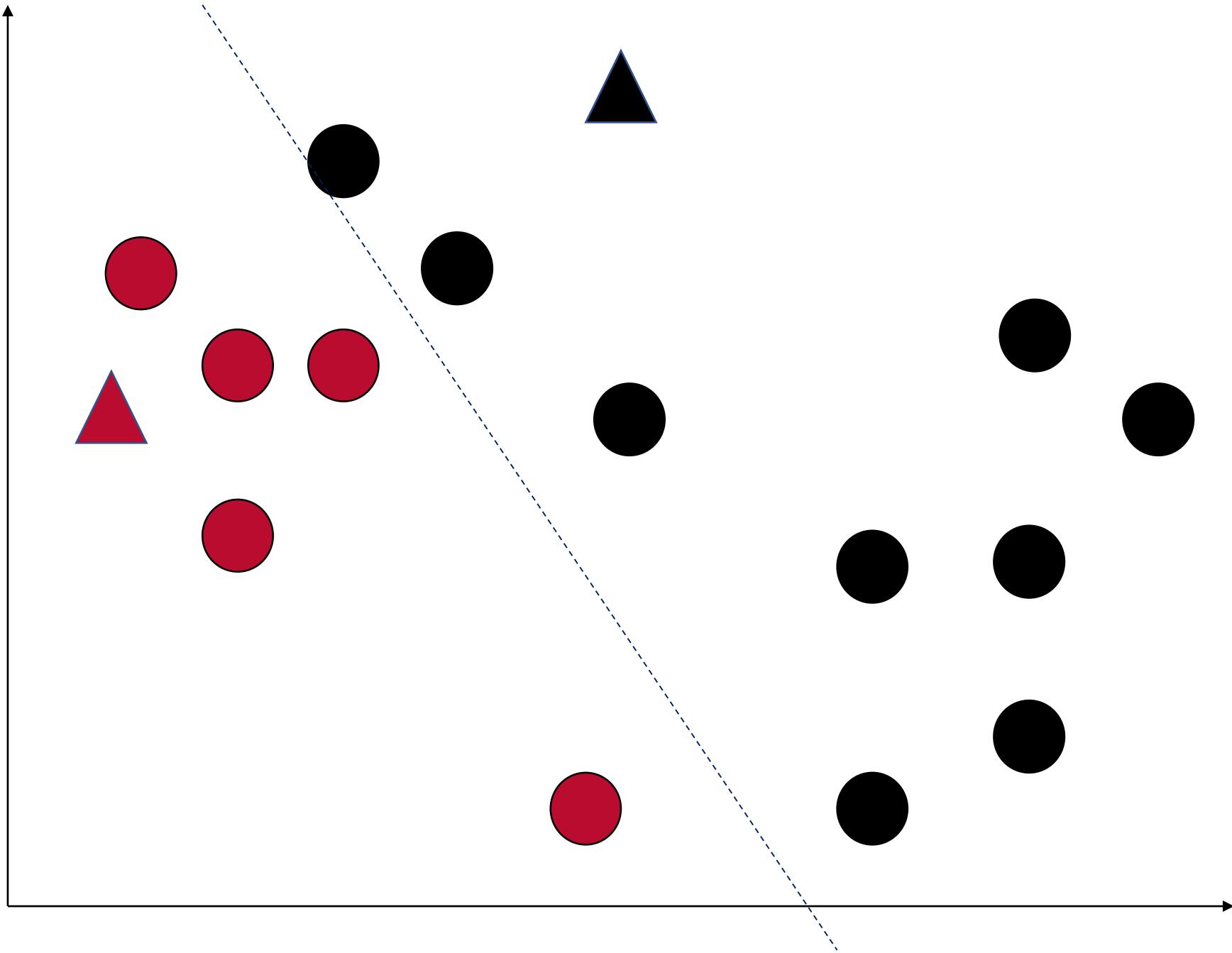
- Hard clustering
- Reduces within-cluster variation
- Increases between-cluster variation
- User choice of clusters
- Simple
- Fast
- Large data

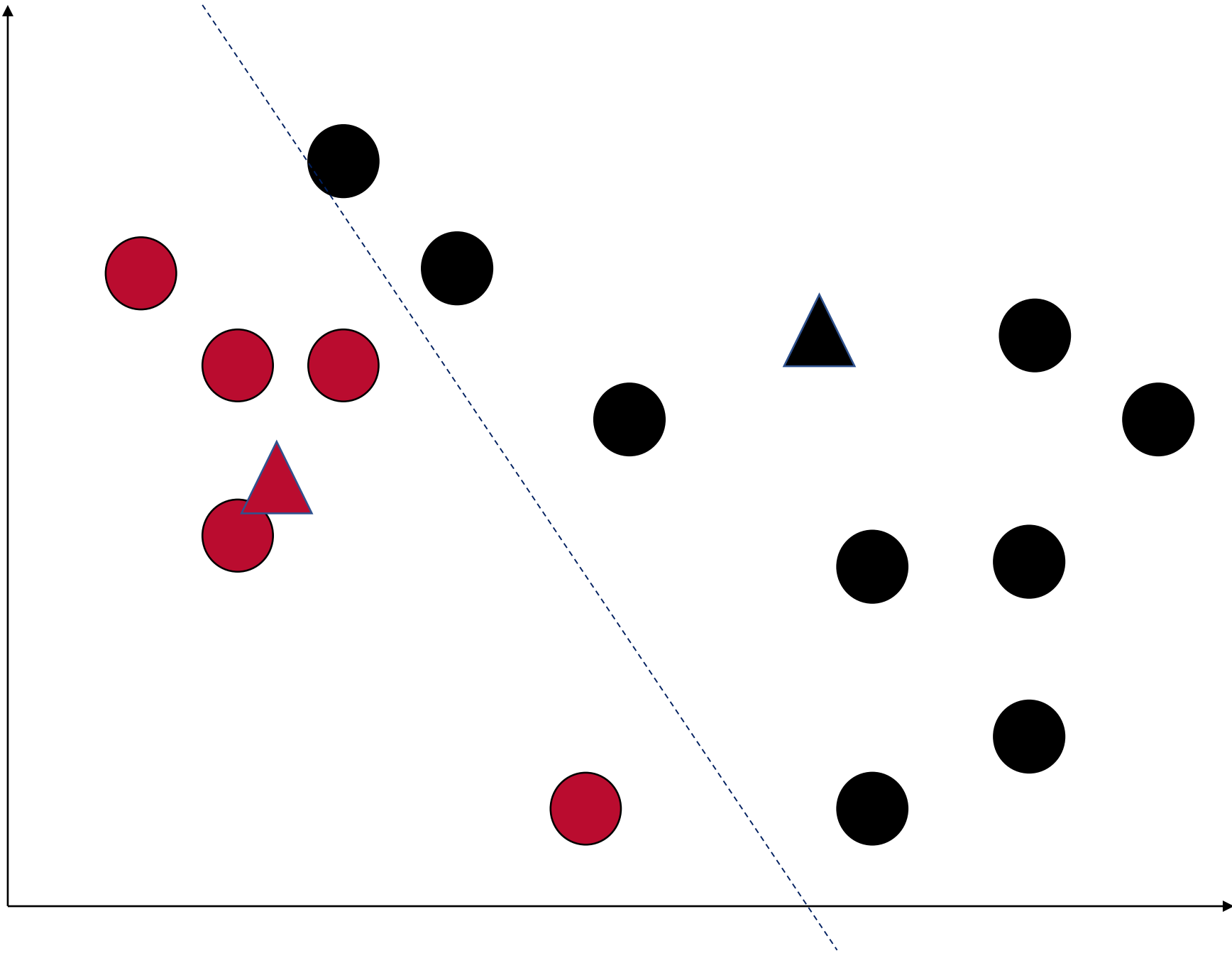


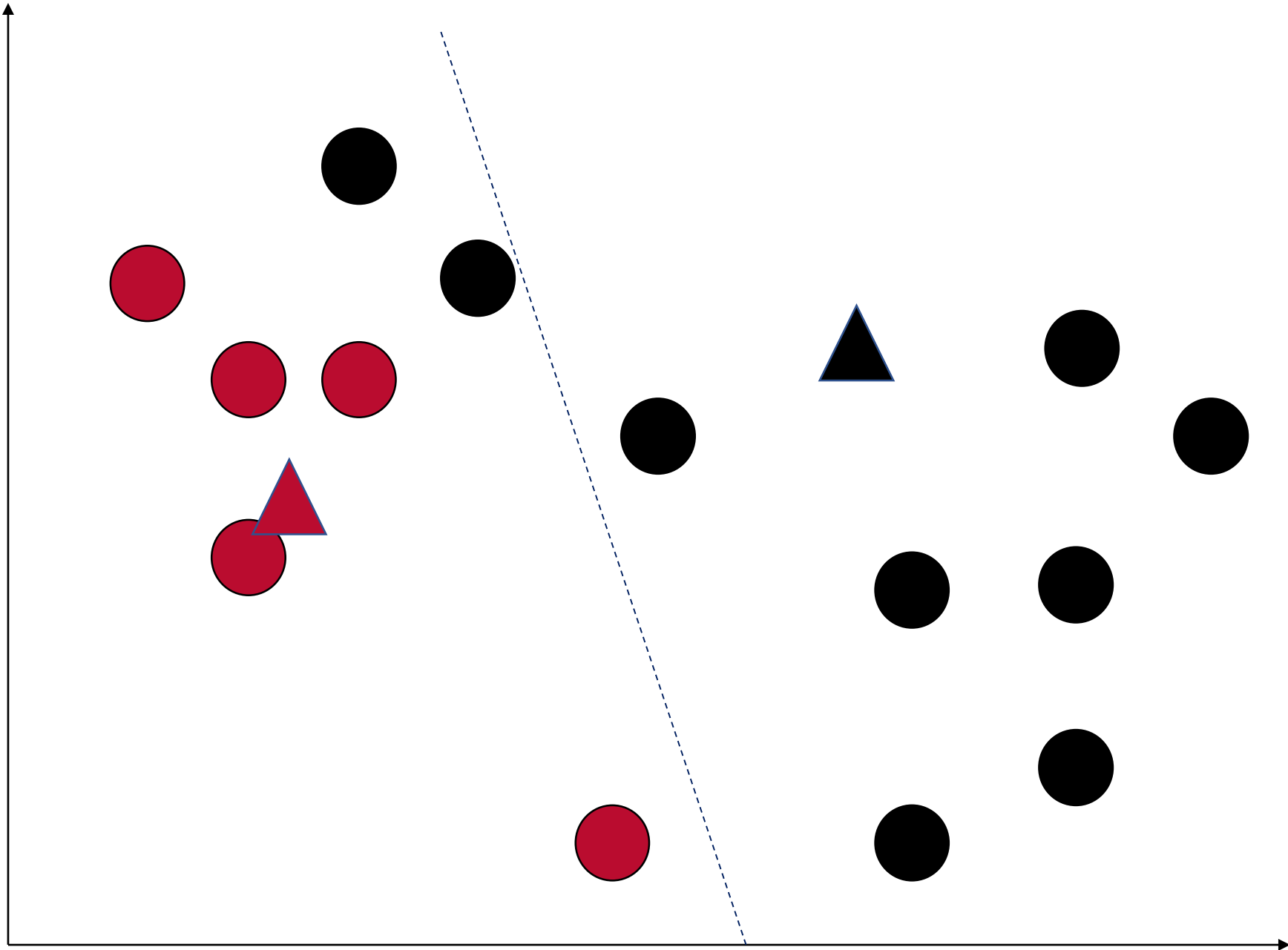


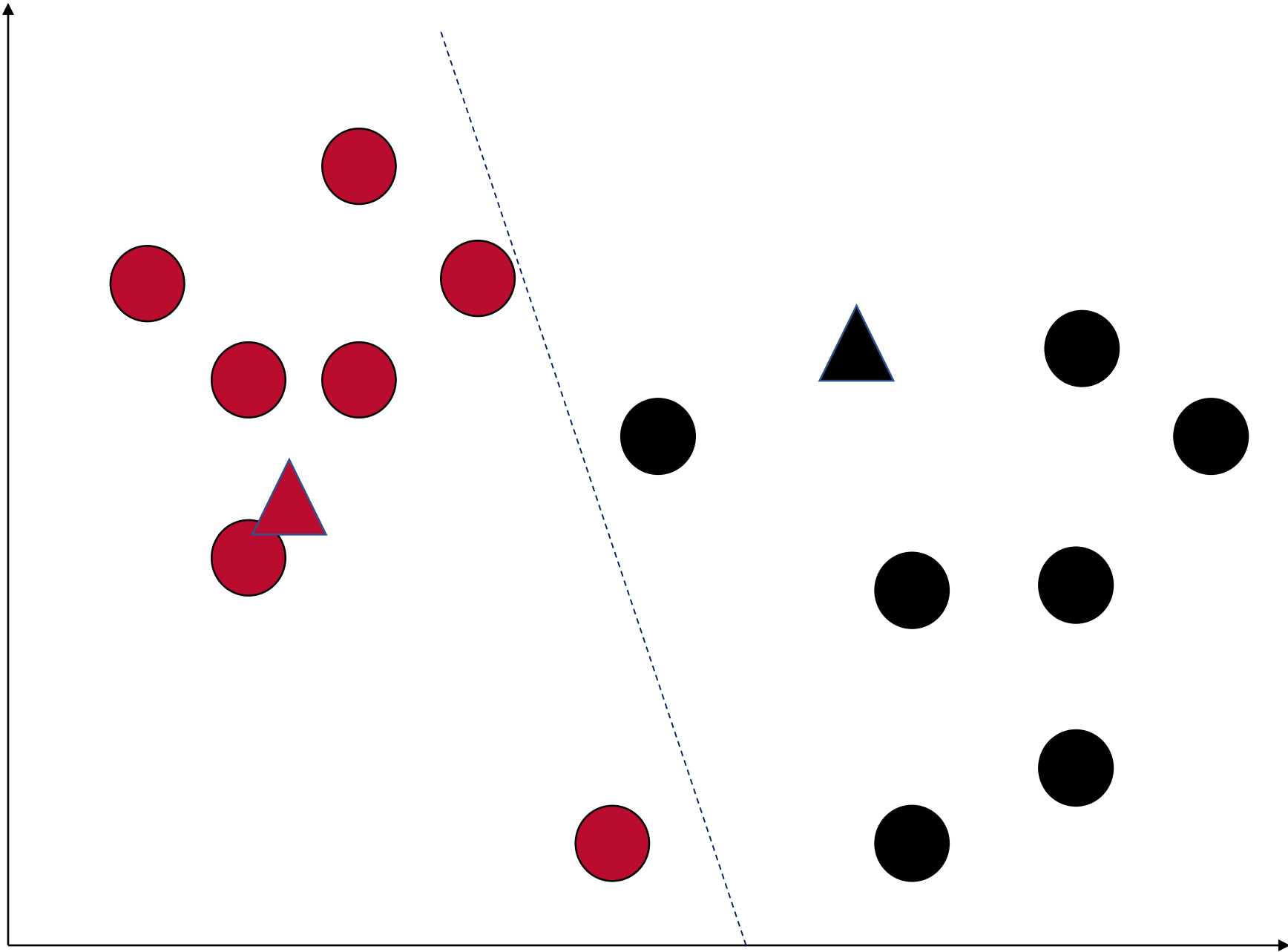


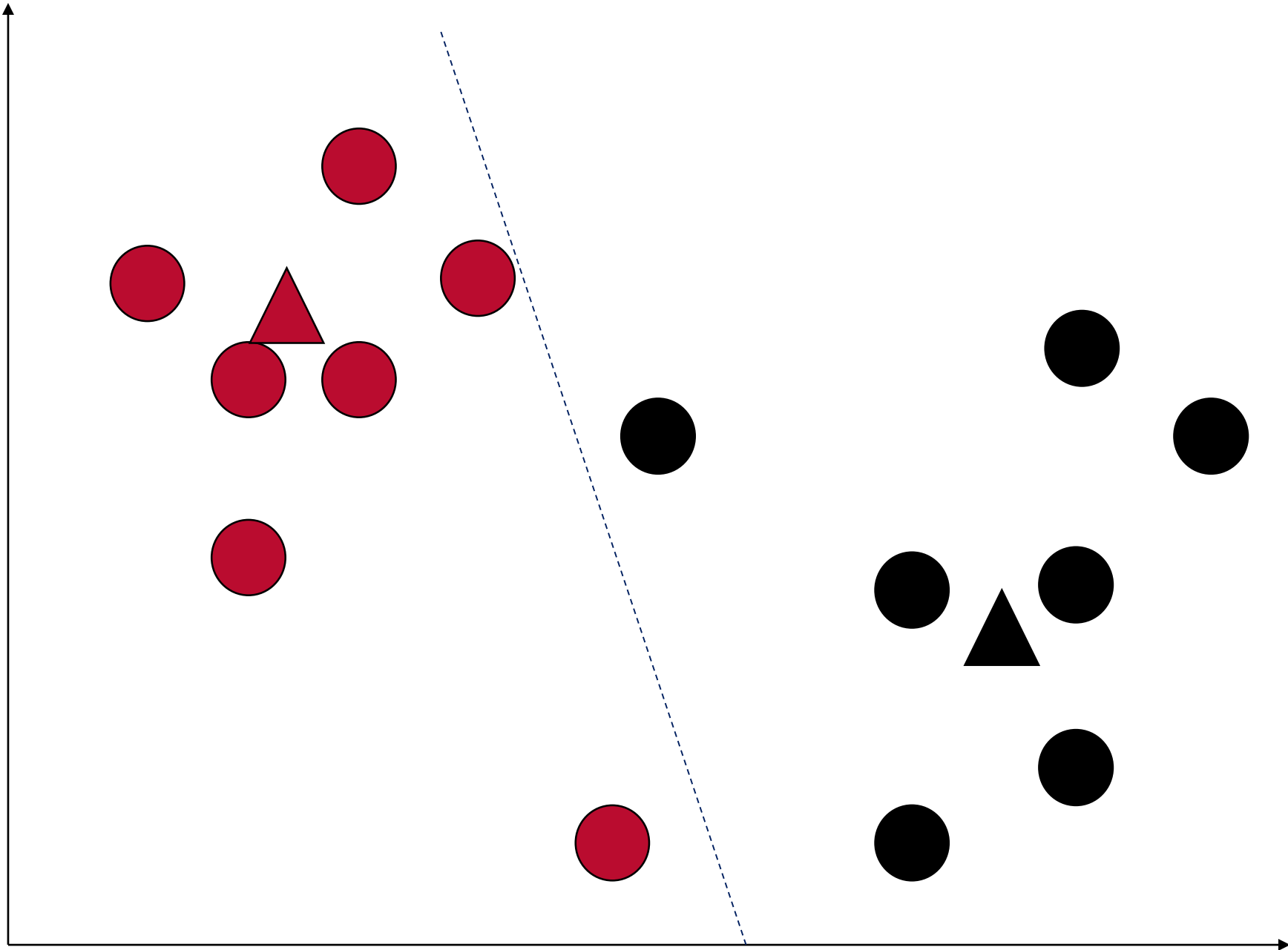


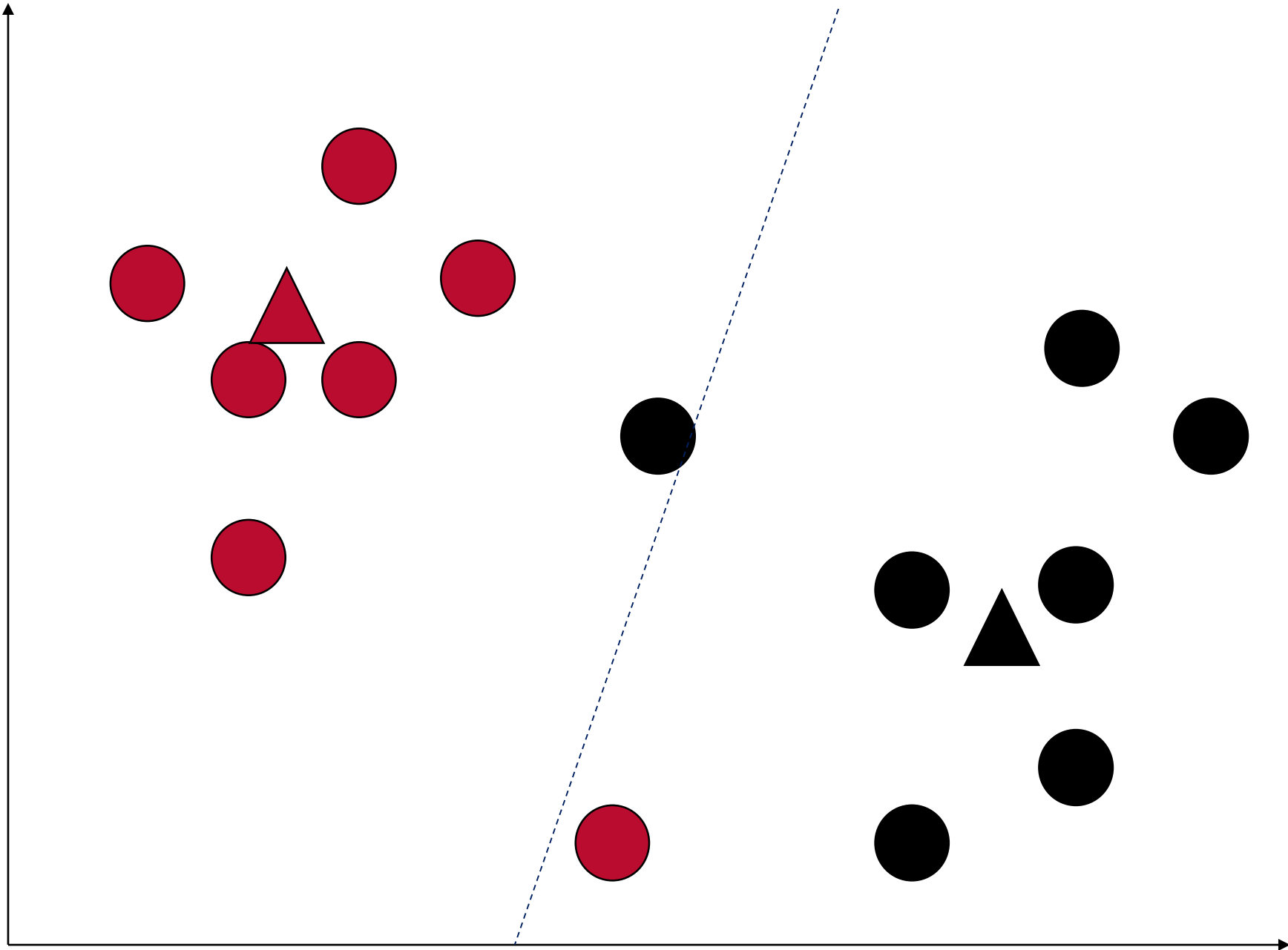


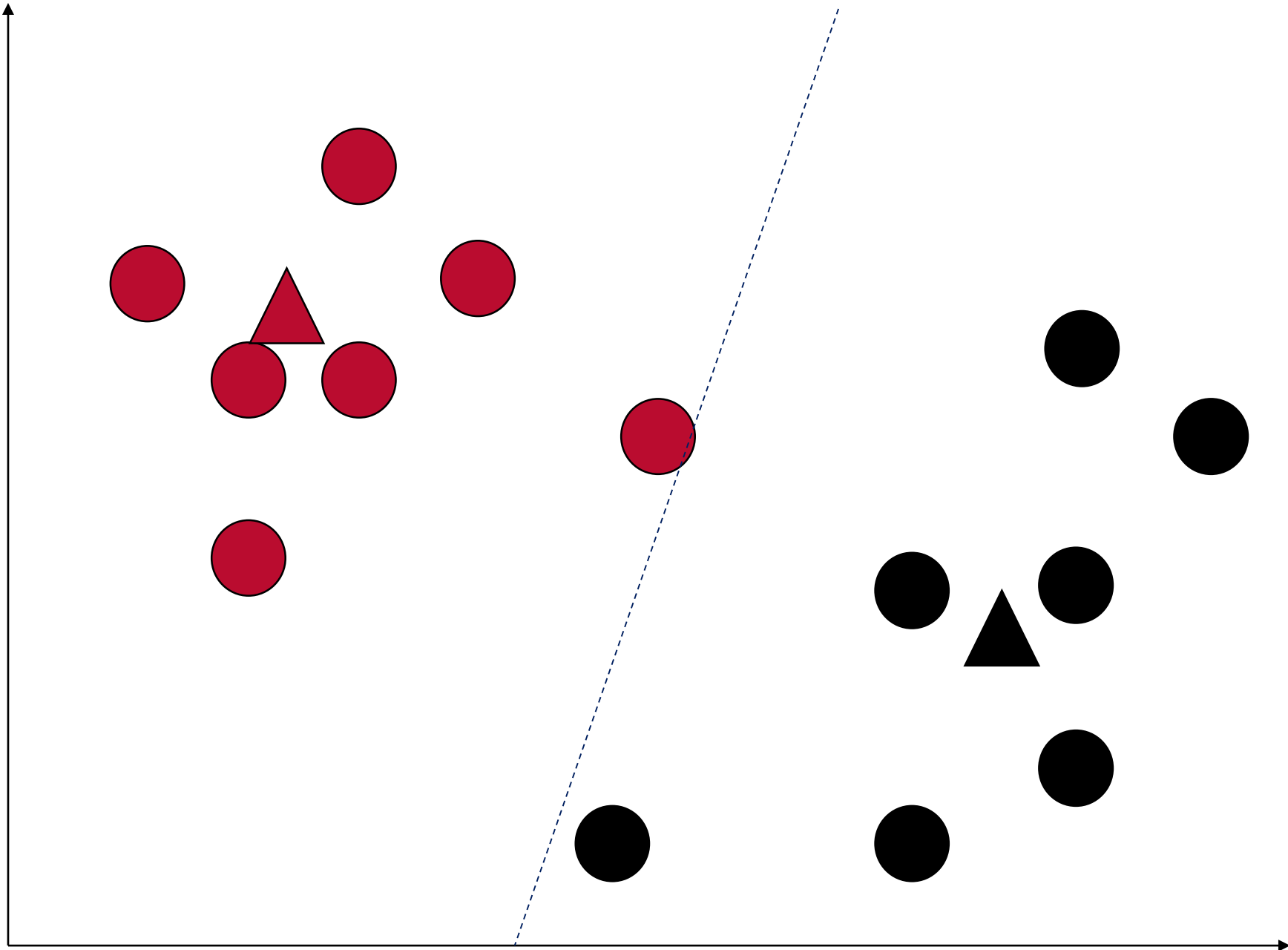


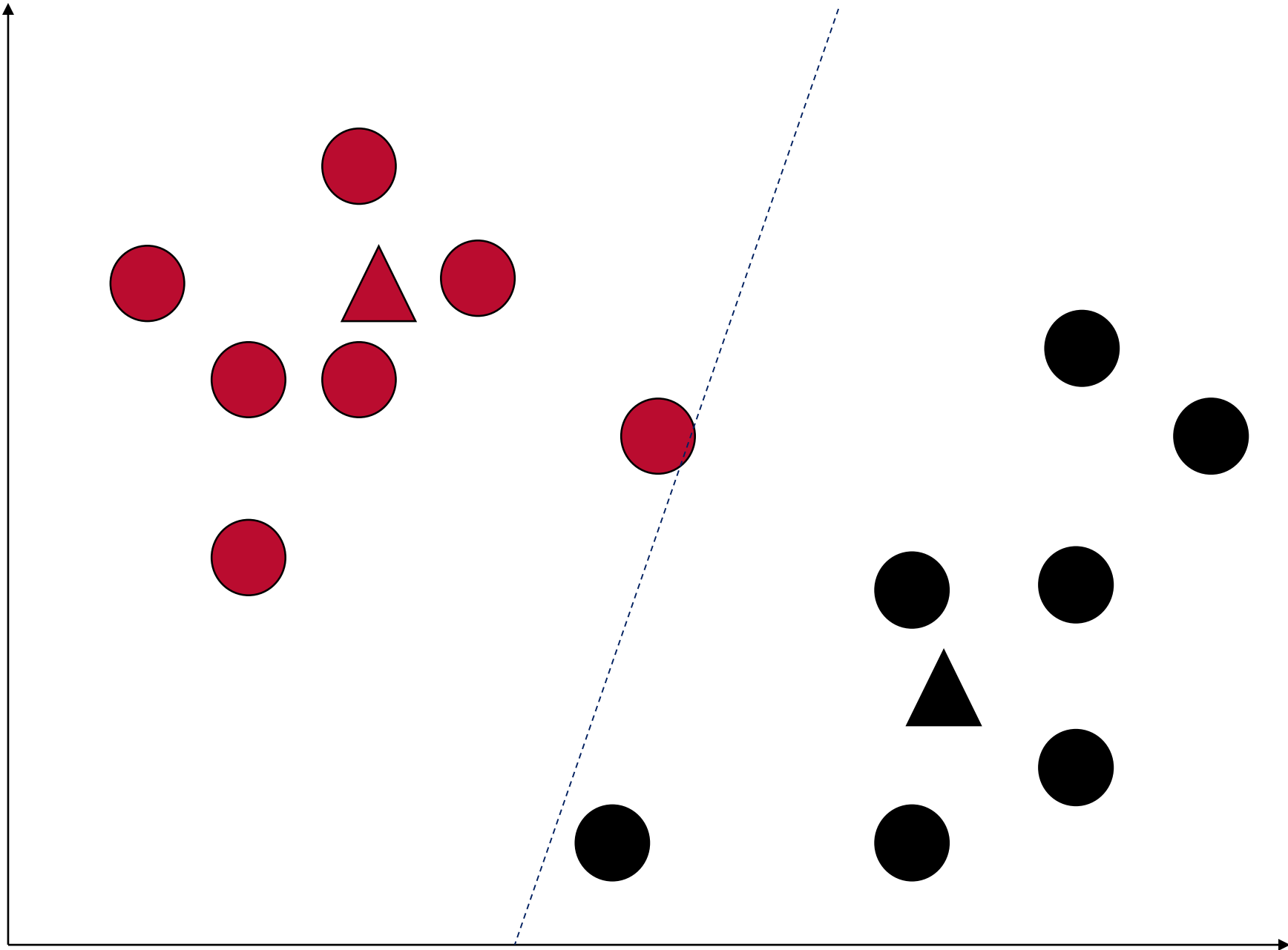


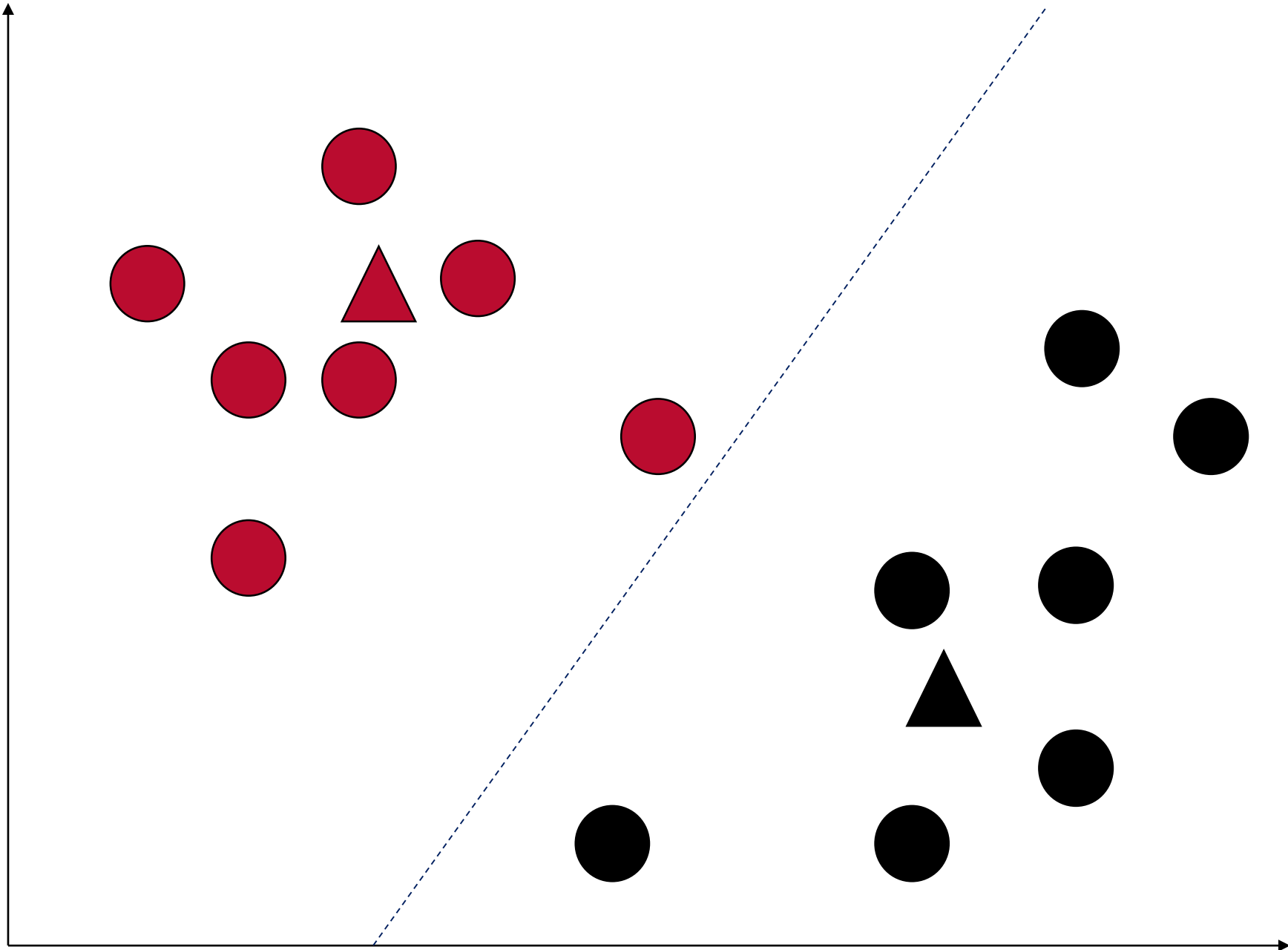


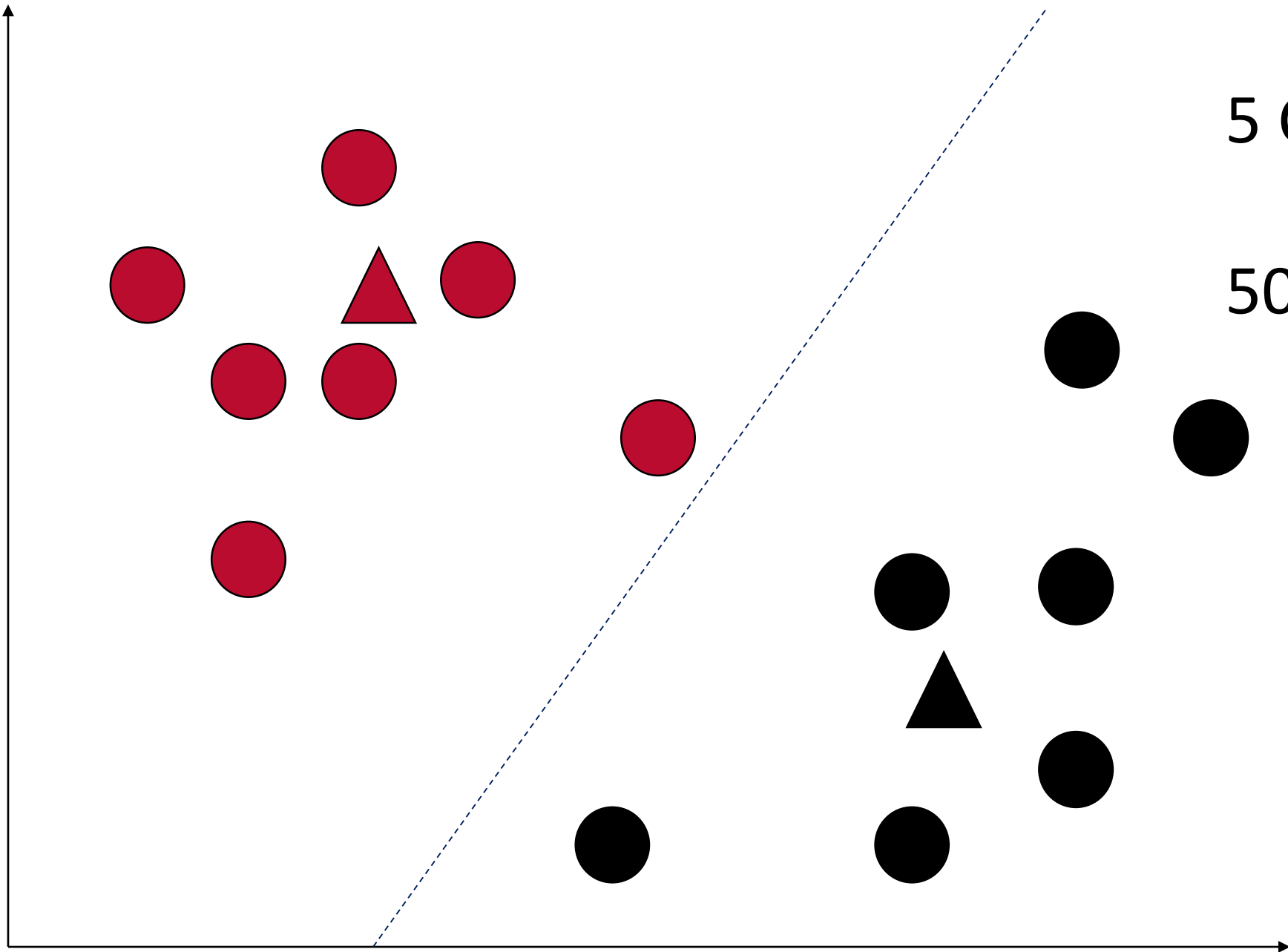






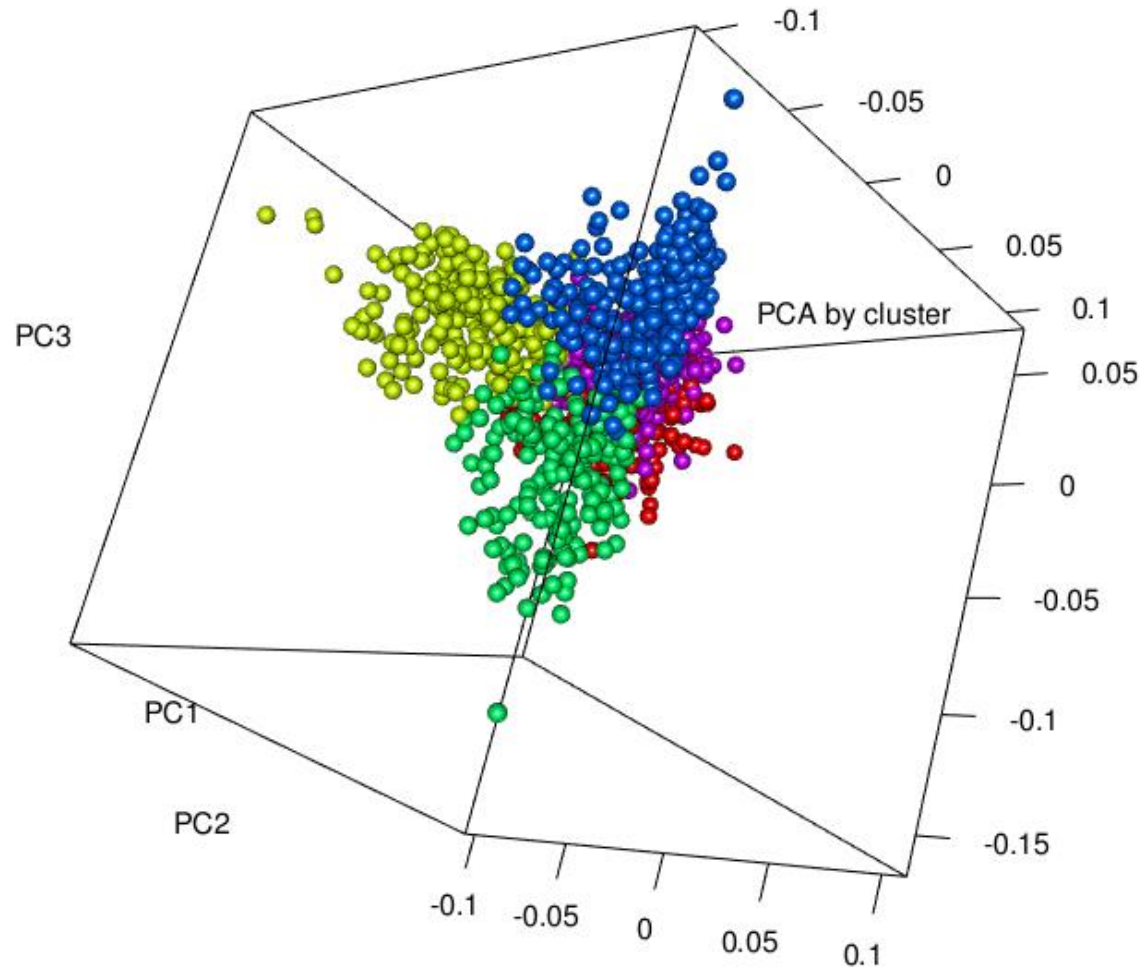






5 Clusters

50 replicates



1

Valiant (1973)

Mark (1978)

Mandingo (1979)

Tesk (1981)

3

Bell (1974)

Belltone (1982)

Durham (1994)

4

Tradition (1974)

Cleitus (1981)

Leadman (1985)

2

Chairman (1976)

Blackstar (1983)

Emory (1989)

Integrity (1990)

4

Starbuck (1979)

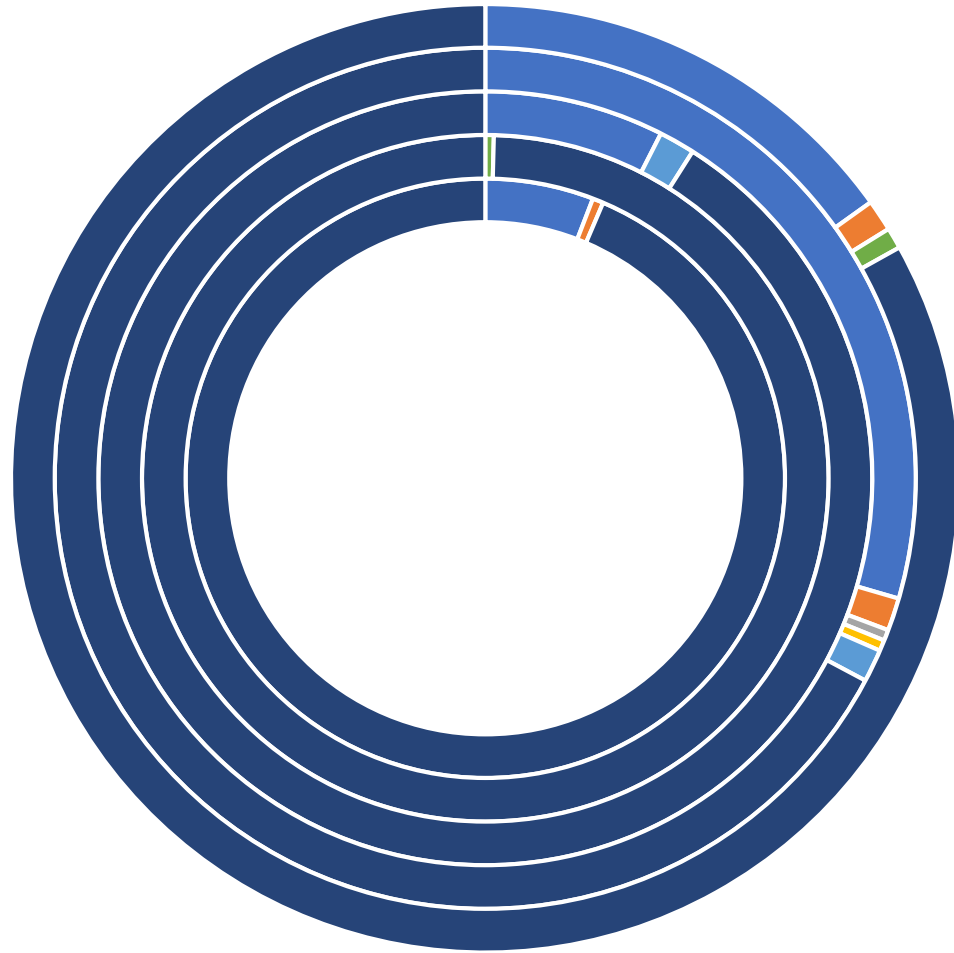
Aerostar (1985)

Outside (1994)

Morty (1997)



Country composition of clusters



Clusters are reflecting different populations

Blackstar (2) contains almost all U.S. bulls.

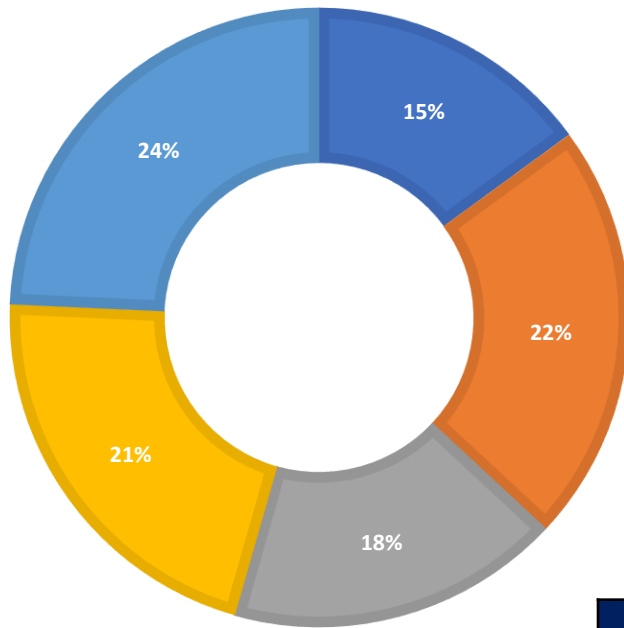
Starbuck (4) is the most “international” group

■ CAN ■ DEU ■ GBR ■ HUN ■ ITA ■ NLD ■ USA



ANIMALS IN CLUSTERS

■ Chief ■ Blackstar ■ Bell ■ Starbuck ■ Tradition



The smallest group includes Valiant (Cluster 1), older U.S. genotyped bulls
Bulls in this group have a strong relationship to a single progenitor

PAWNEE FARM ARLINDA CHIEF

The largest group is from Tradition (Cluster 5)

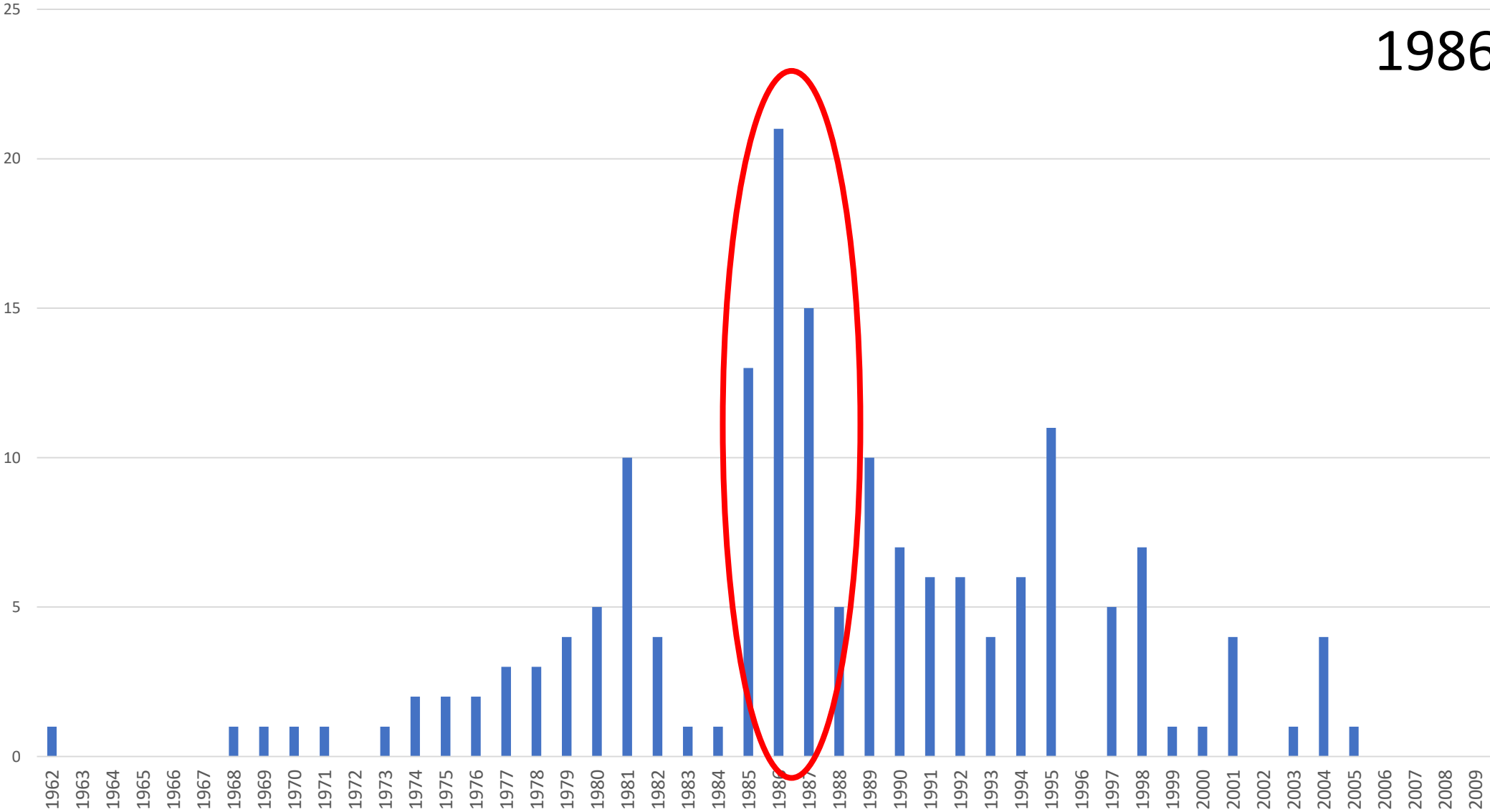
Multiple U.S. sire families are represented

Tradition is the son of **ROUND OAK RAG APPLE ELEVATION**

	Sire Family	
Cluster	Progenitors	Key Sires
1	Chief	Mark and Valiant
2	Ivanhoe & Fond Matt	Chairman & Blackstar
3	Ivanhoe Star	Bell
4	Canadian breeding	Starbuck, Outside, etc.
5	US breeding Multiple families	Tradition, Leadman, etc.

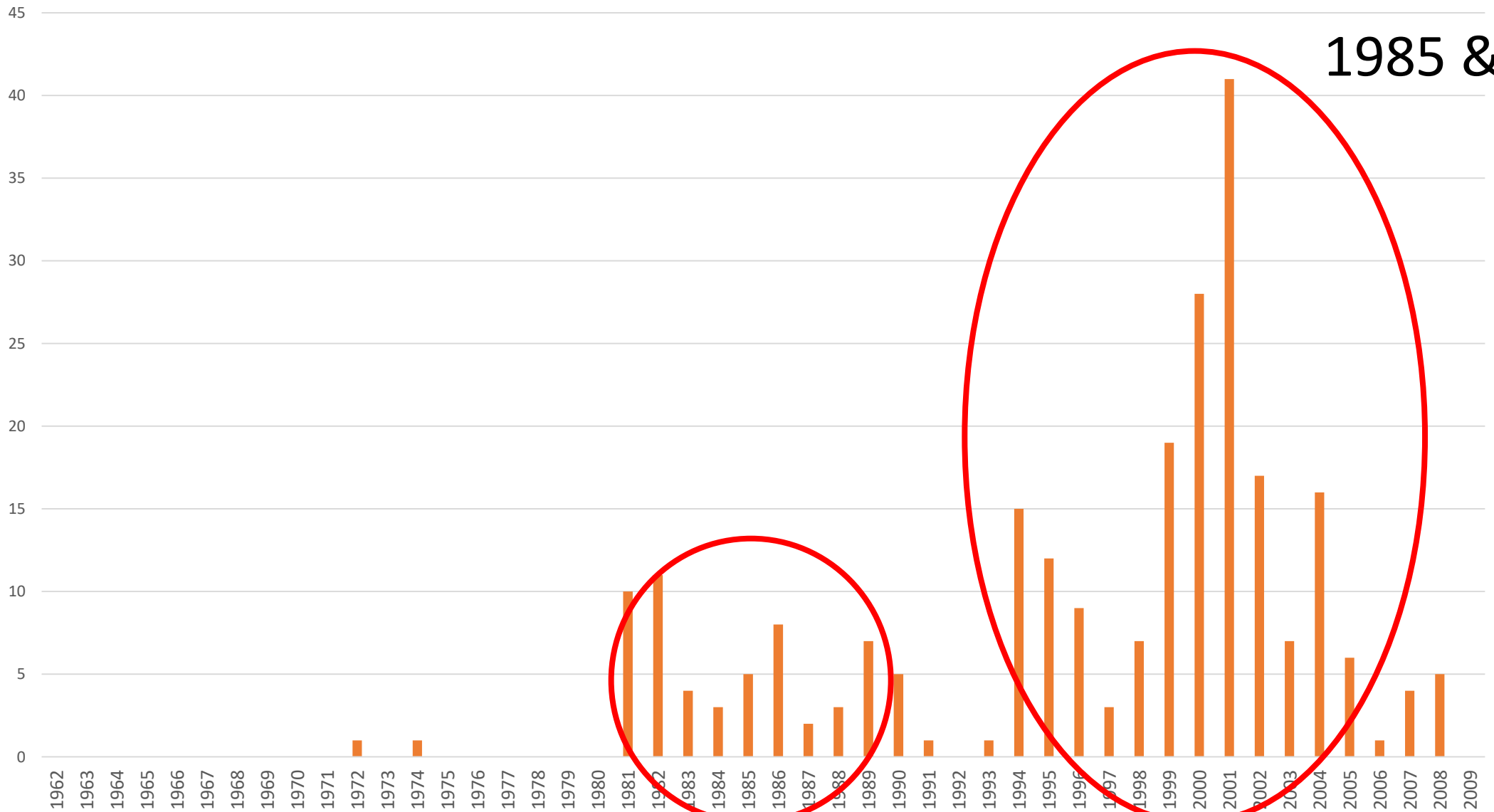
Cluster 1: Chief

1986

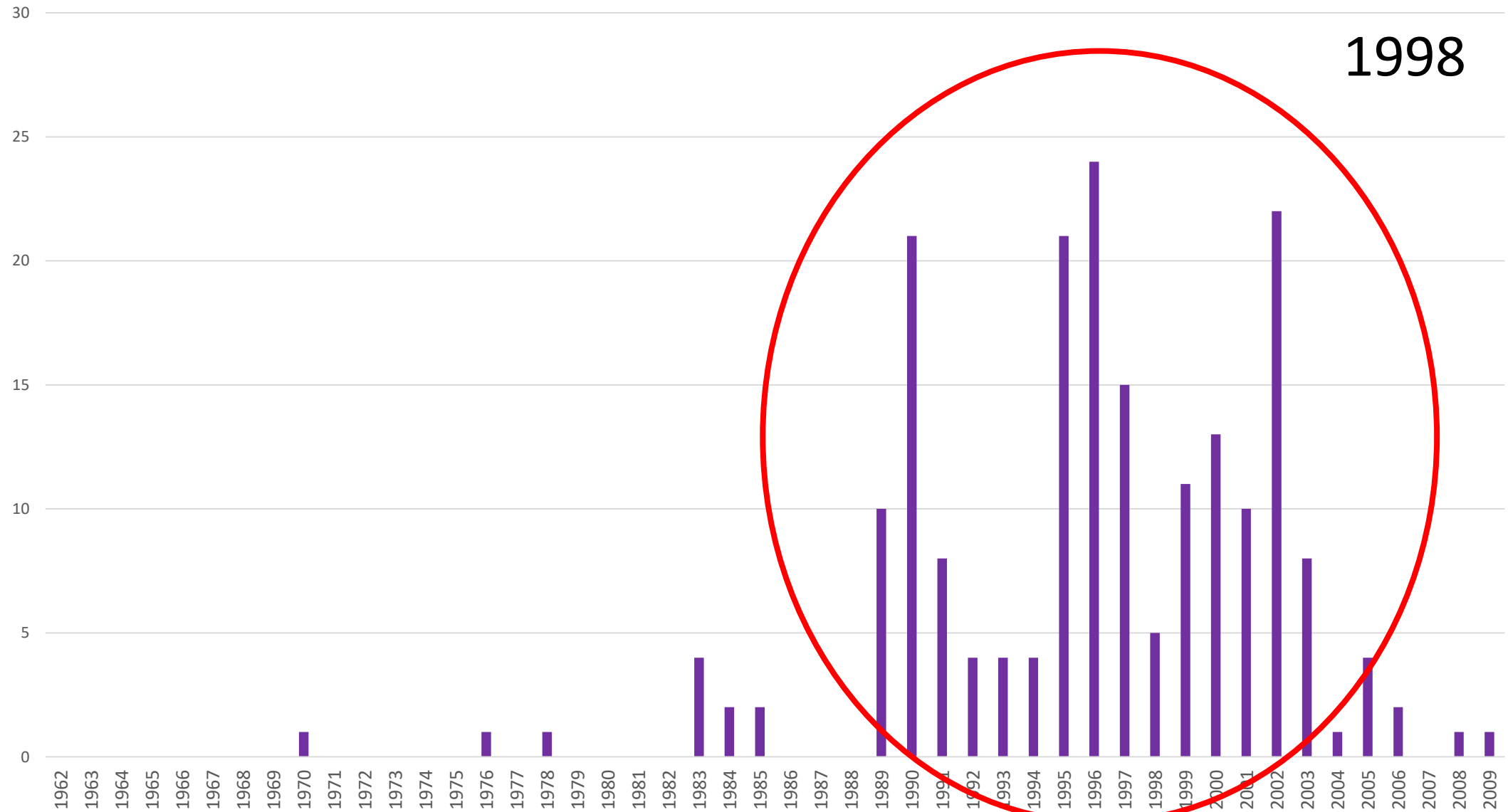


Cluster 2: Blackstar

1985 & 2001

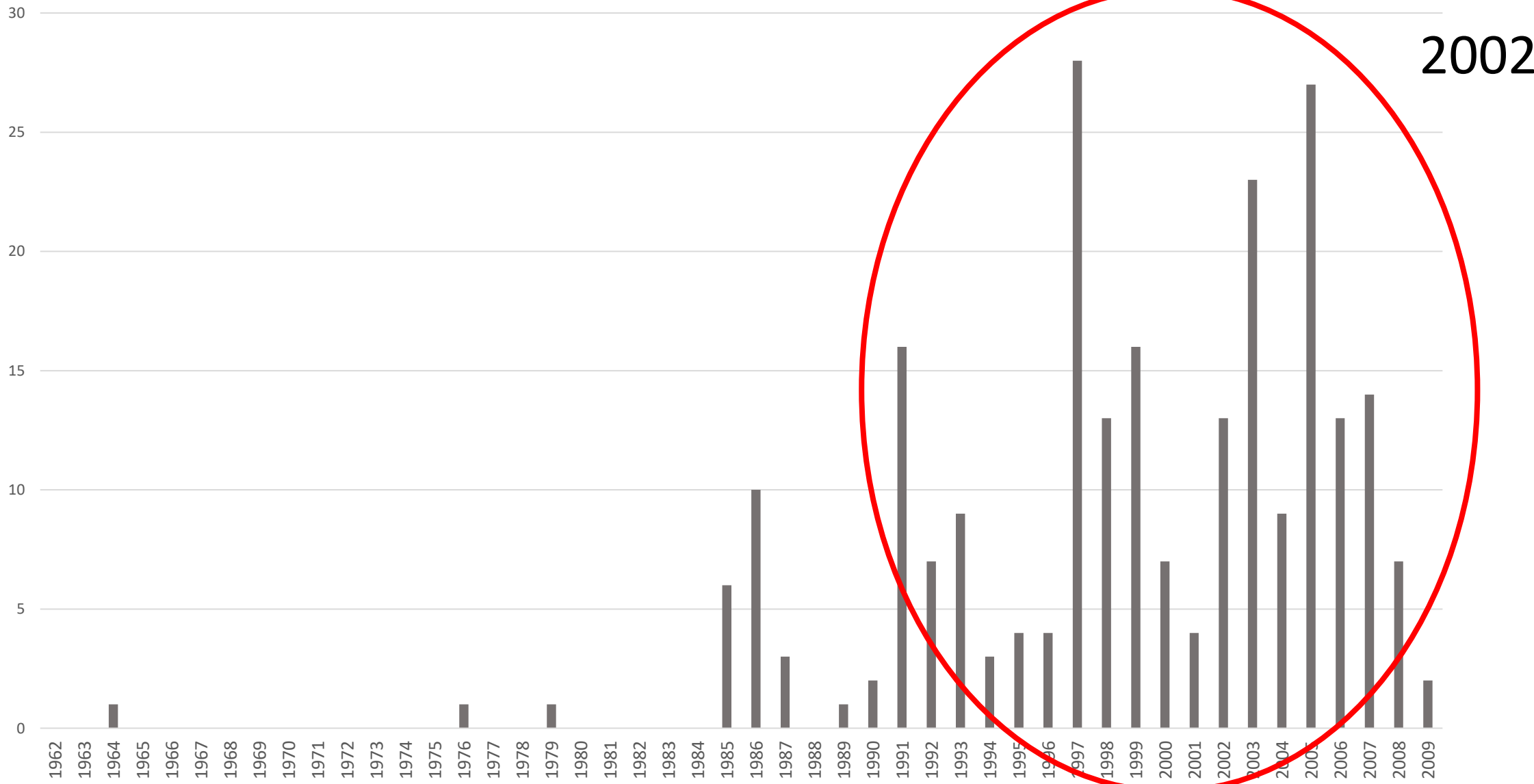


Cluster 3: Bell



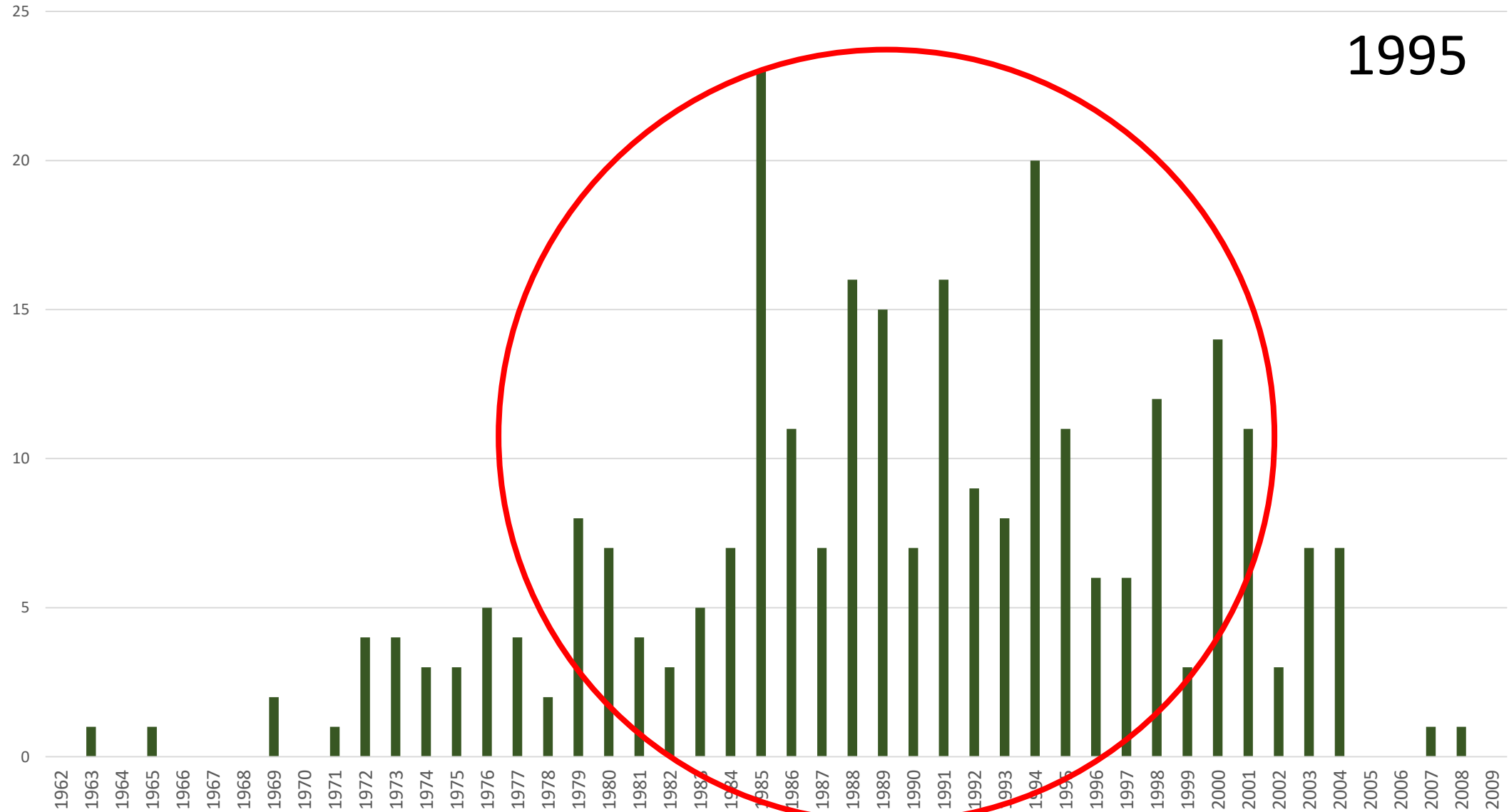
Canadian

Cluster 4: Starbuck



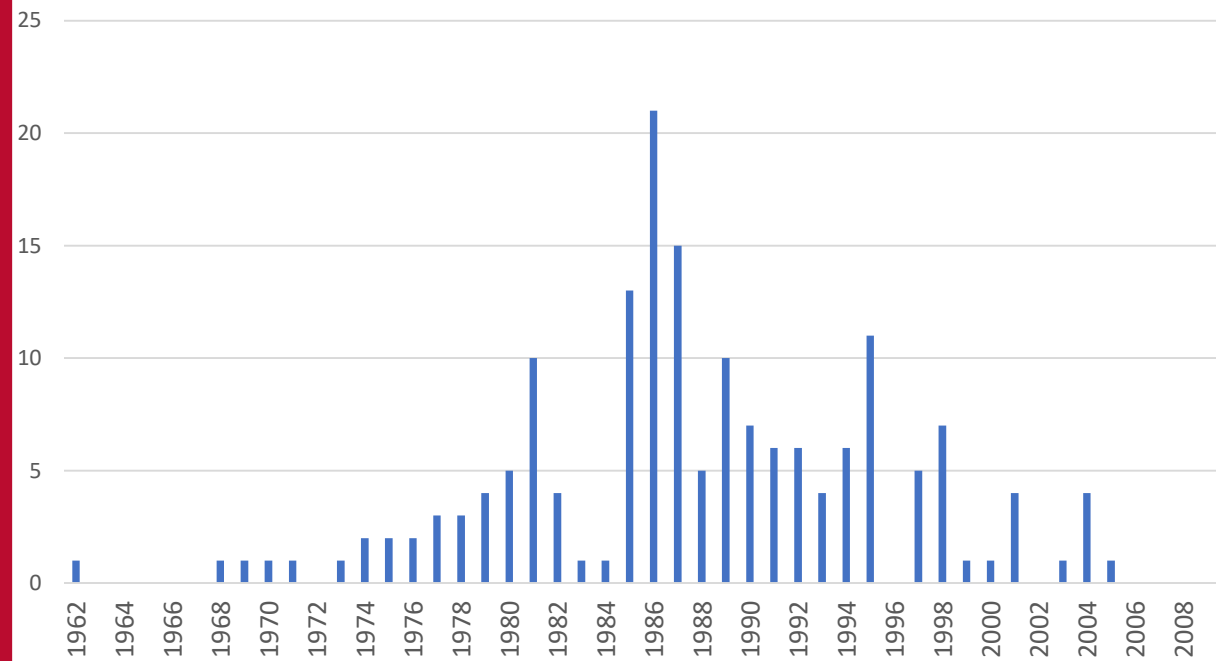
Multiple US

Cluster 5: Tradition



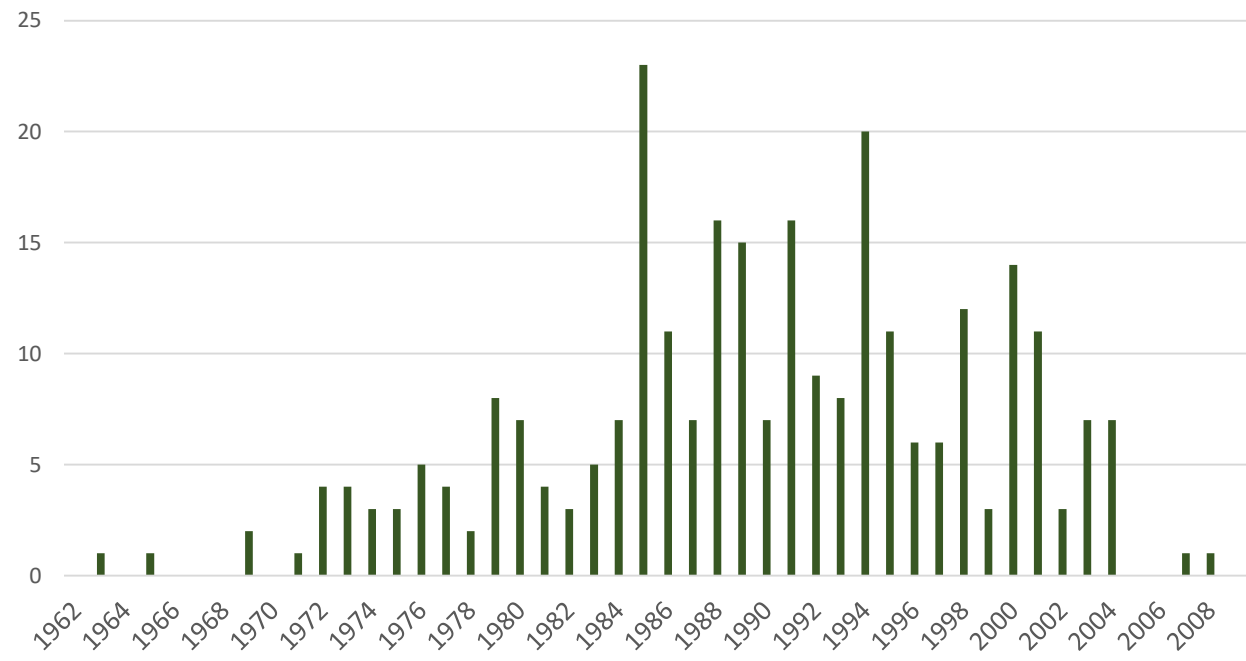
Oldest clusters

Chief (1)



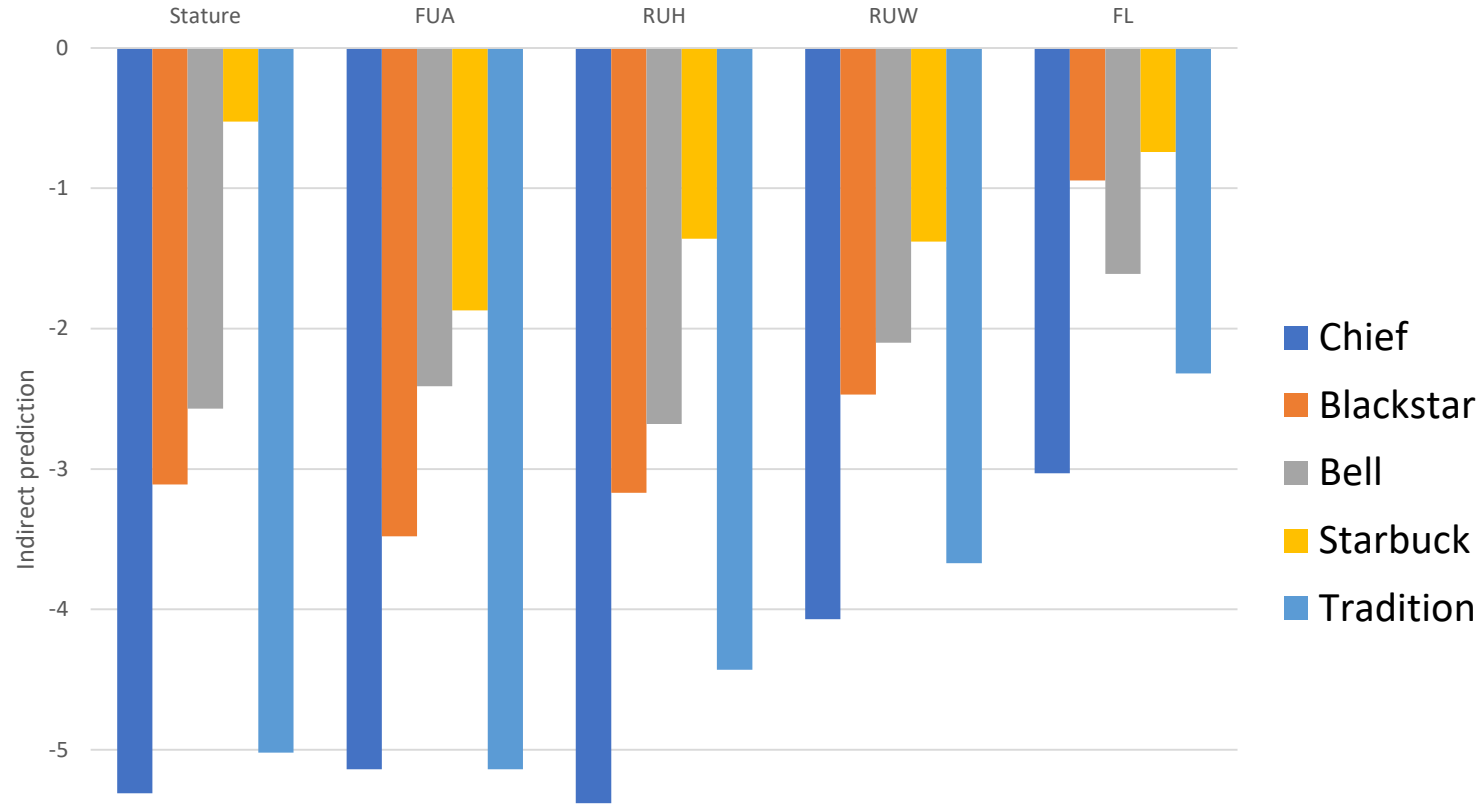
1986

Tradition (5)



1995

Traits means within cluster



Ranking	Stature	FUA	RUH	RUW	FL
First	Starbuck	Starbuck	Starbuck	Starbuck	Starbuck
Second	Bell	Bell	Bell	Bell	Blackstar
Third	Blackstar	Blackstar	Blackstar	Blackstar	Bell
Fourth	Tradition	Tradition	Tradition	Tradition	Tradition
Fifth	Chief	Chief	Chief	Chief	Chief

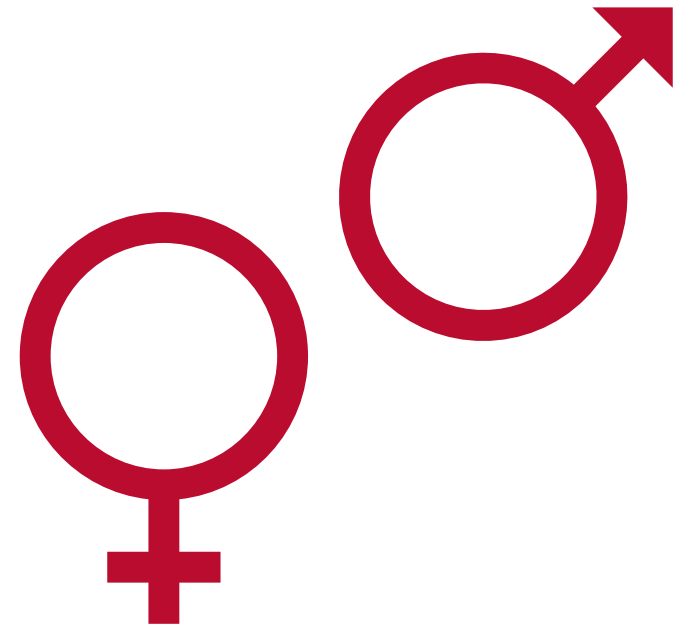
Practical application

- Mating decisions
- Across-cluster mating
- Selection candidates



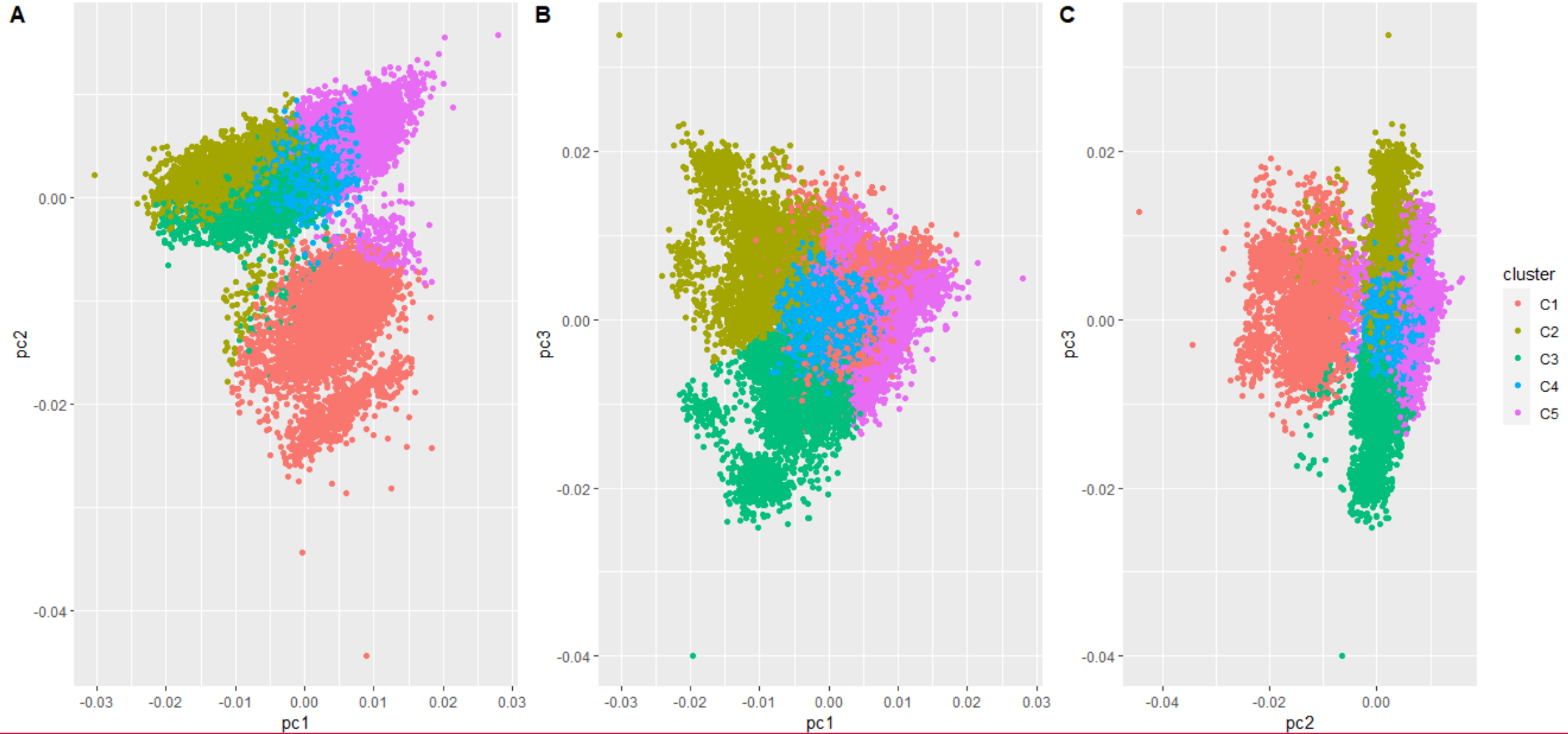
Data

- Sires of animals born after 2010 & at least 25 progeny
- Females measured after 2012 (only data to 2014)
- 20,099 selection candidates with genotypes
- Years: 1987 and up
- 14 countries
- 5 Clusters

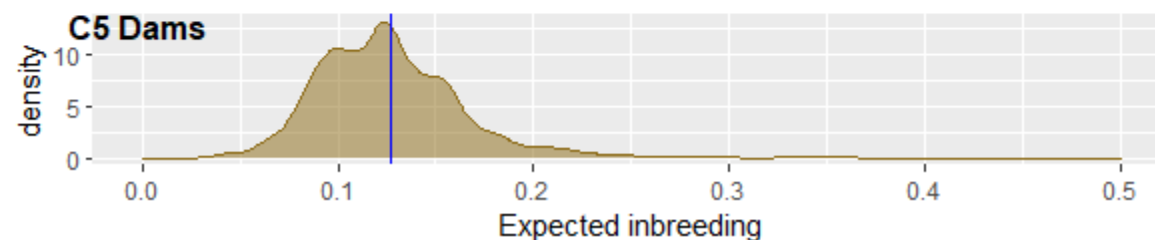
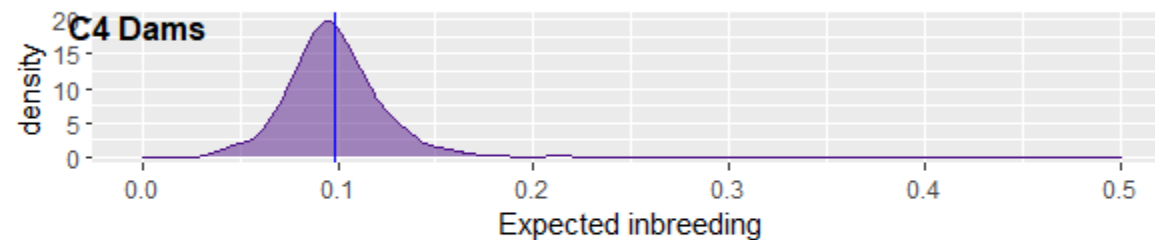
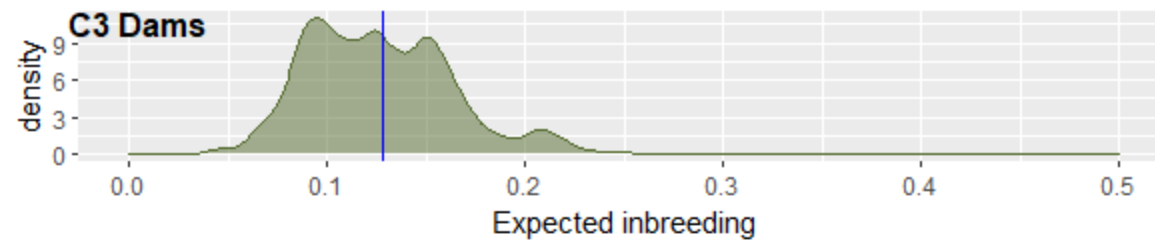
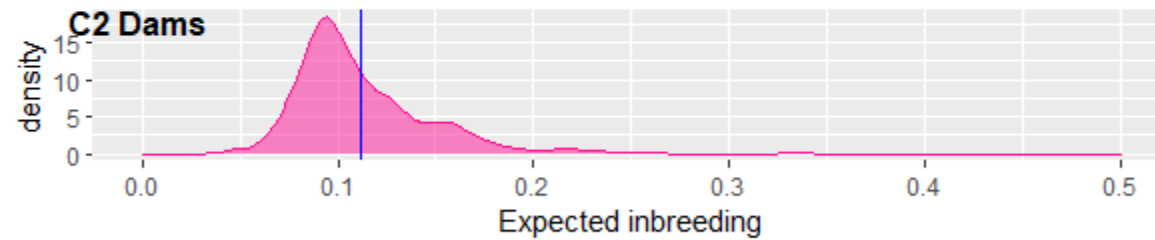
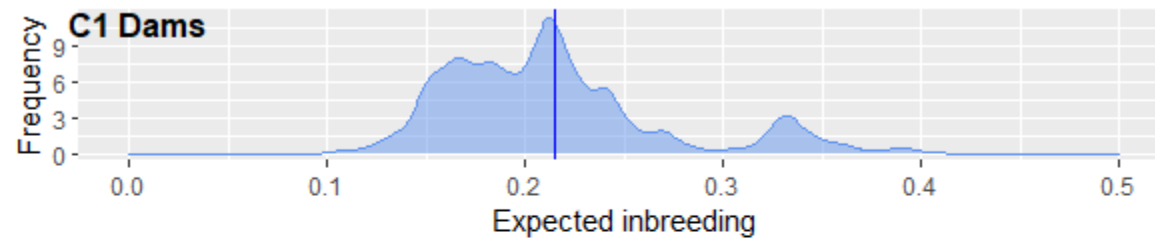


Larger data: Selection candidates

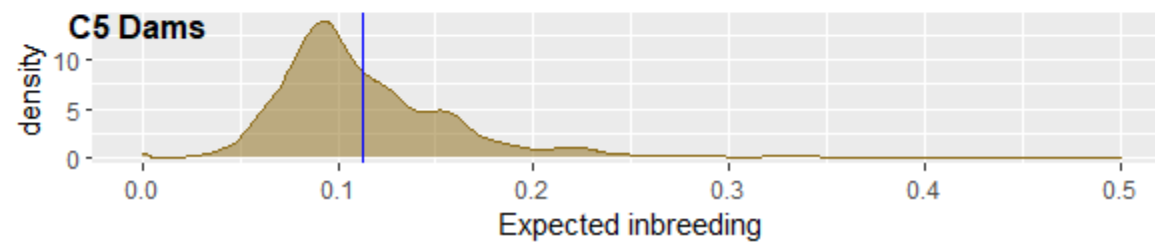
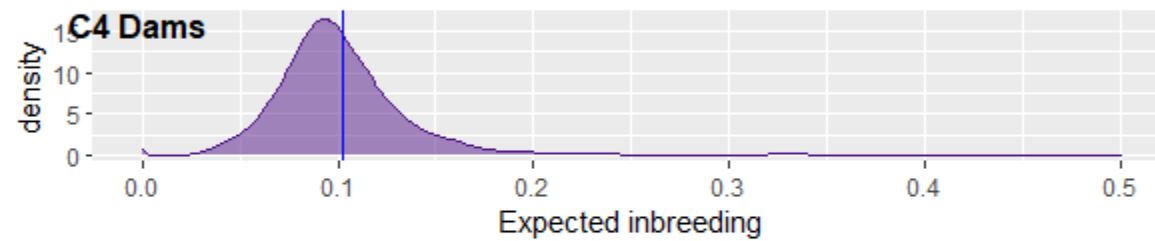
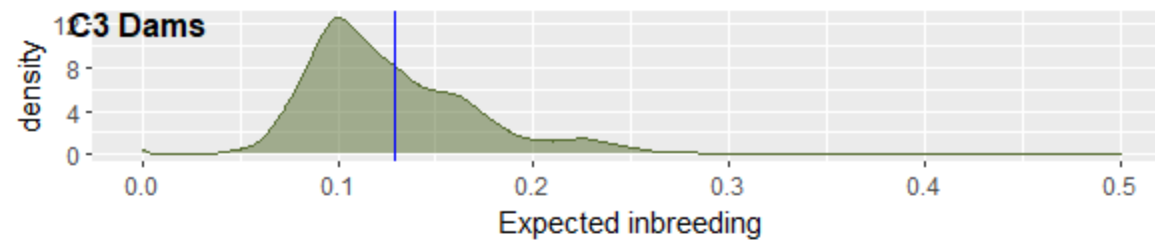
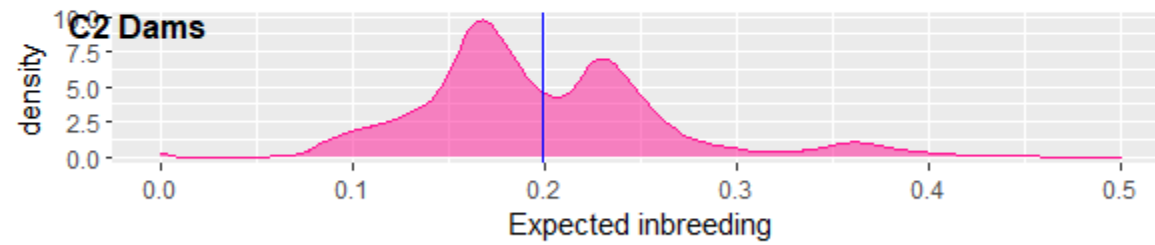
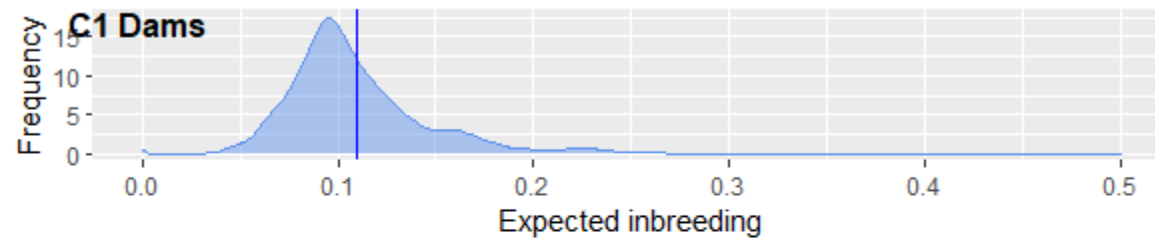
Principal Component Plots - 3 Dimensions



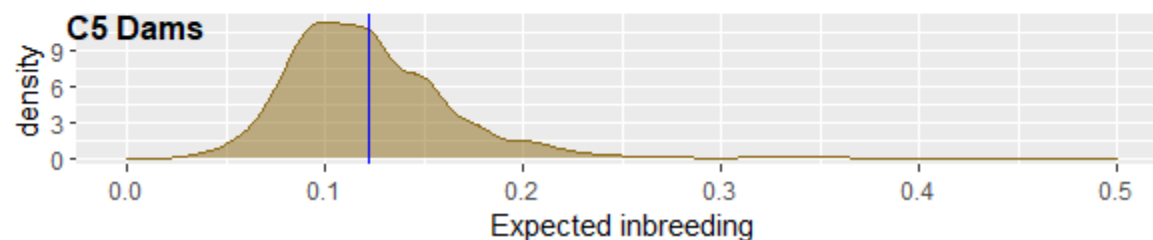
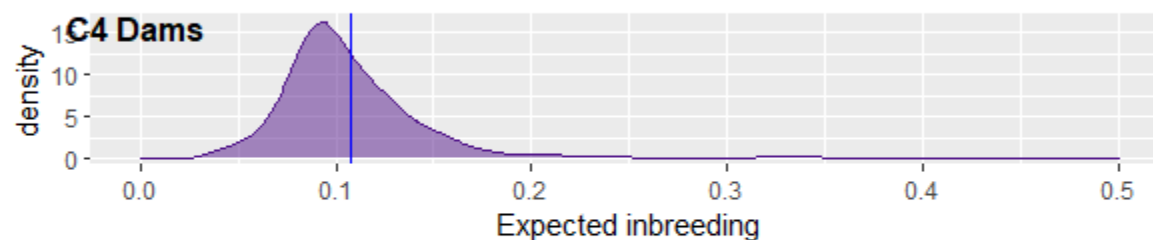
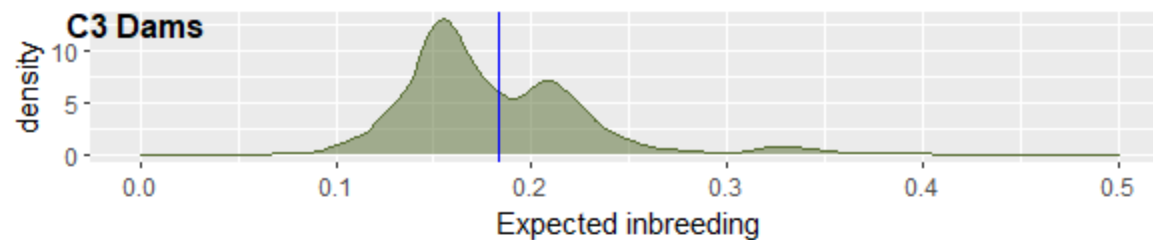
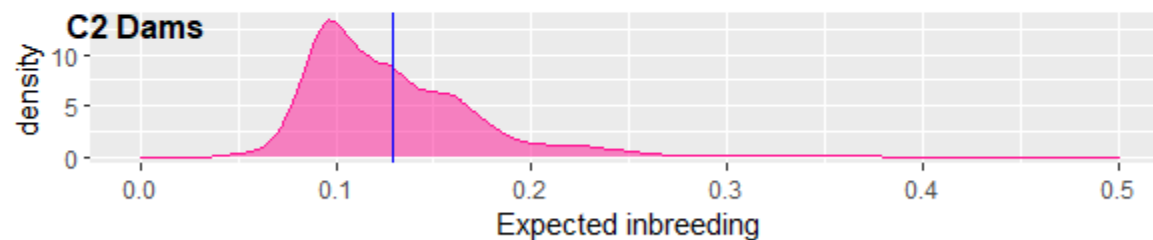
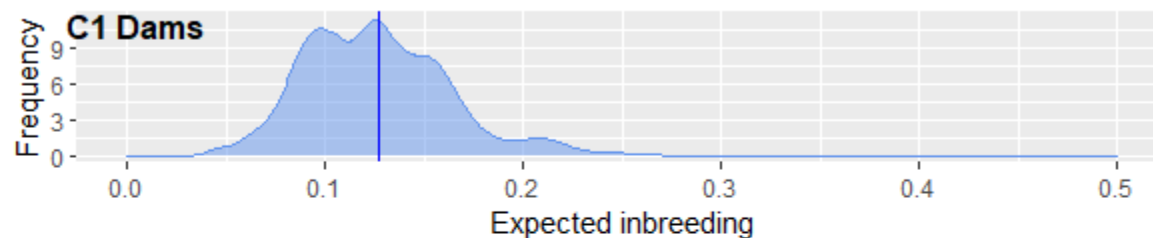
Expected inbreeding when using Cluster 1 sires



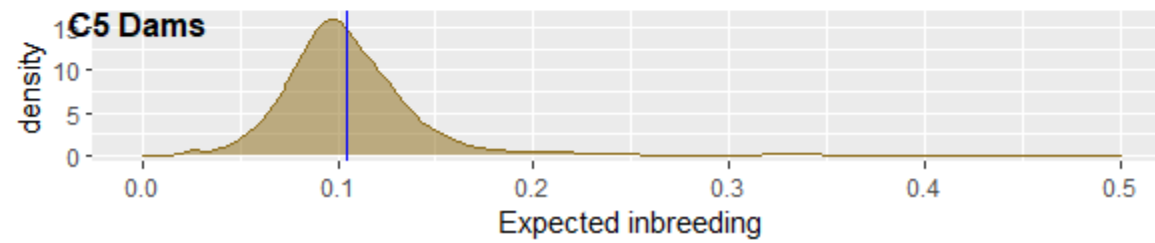
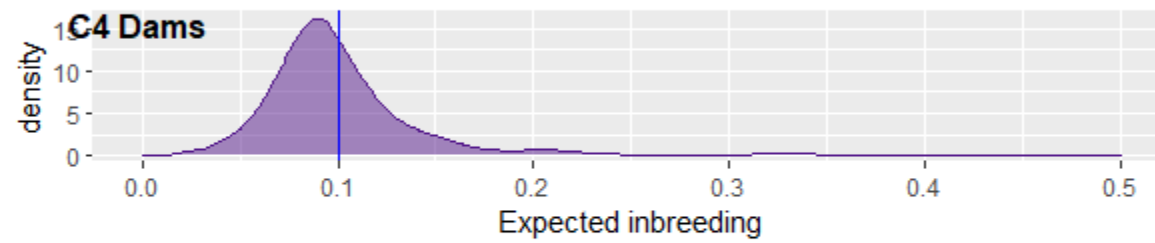
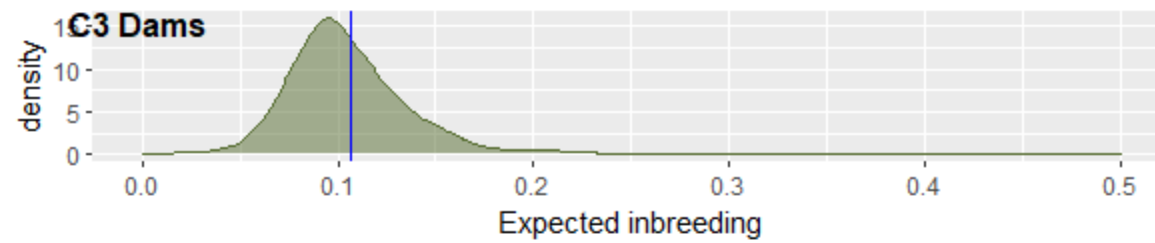
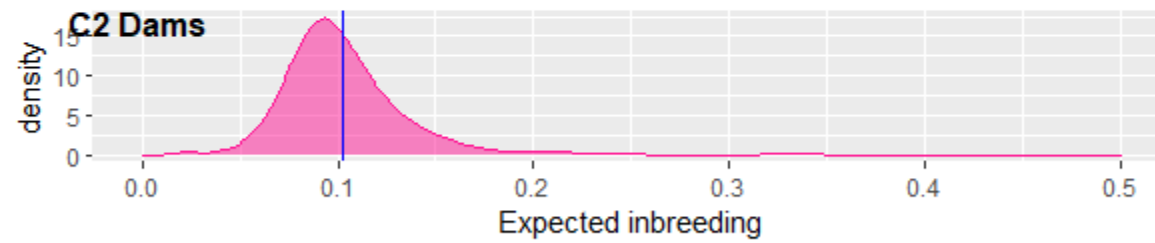
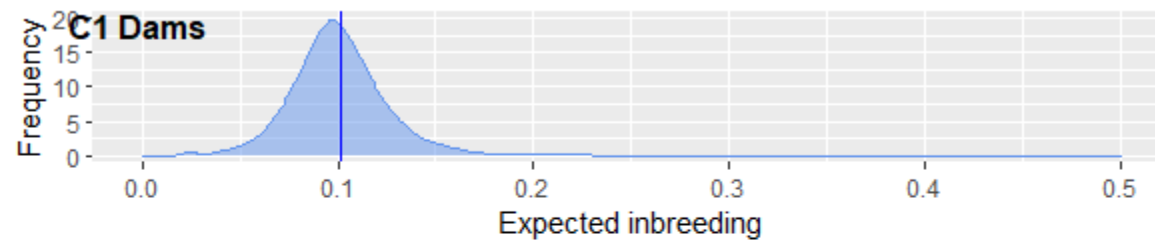
Expected inbreeding when using Cluster 2 sires



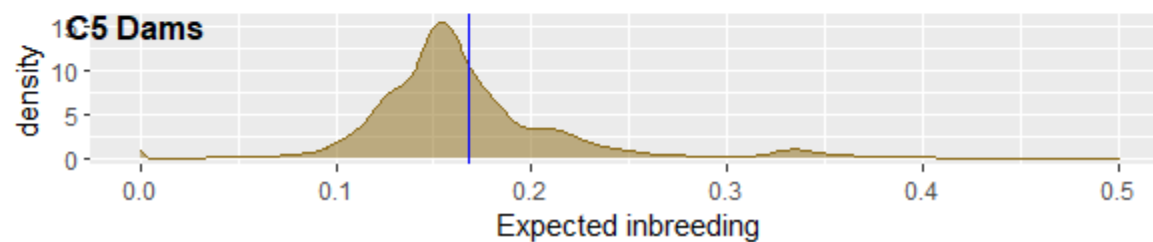
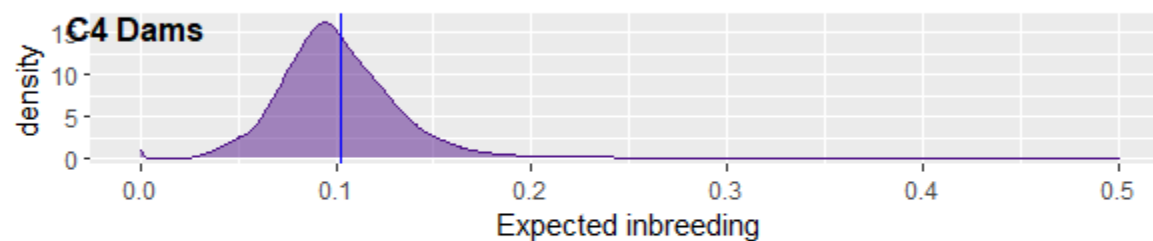
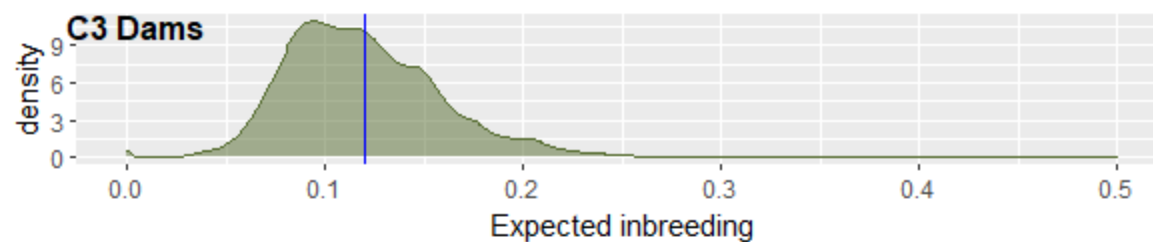
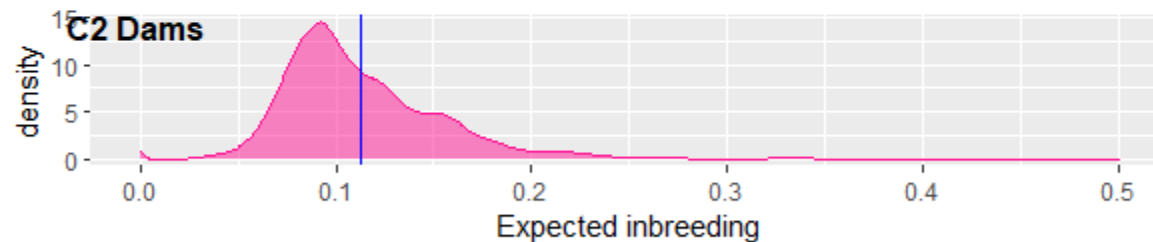
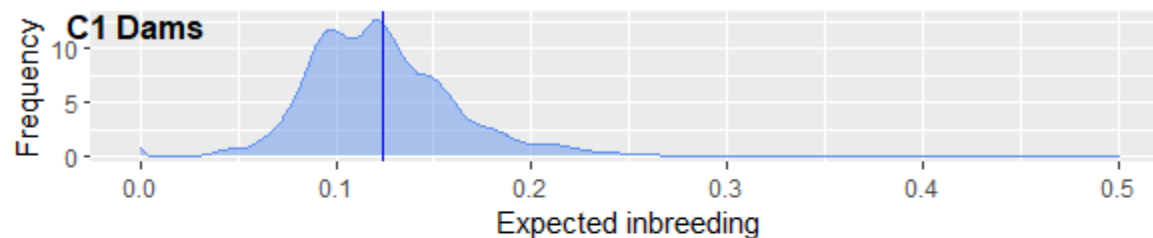
Expected inbreeding when using Cluster 3 sires



Expected inbreeding when using Cluster 4 sires



Expected inbreeding when using Cluster 5 sires



Expected inbreeding: 8 Clusters



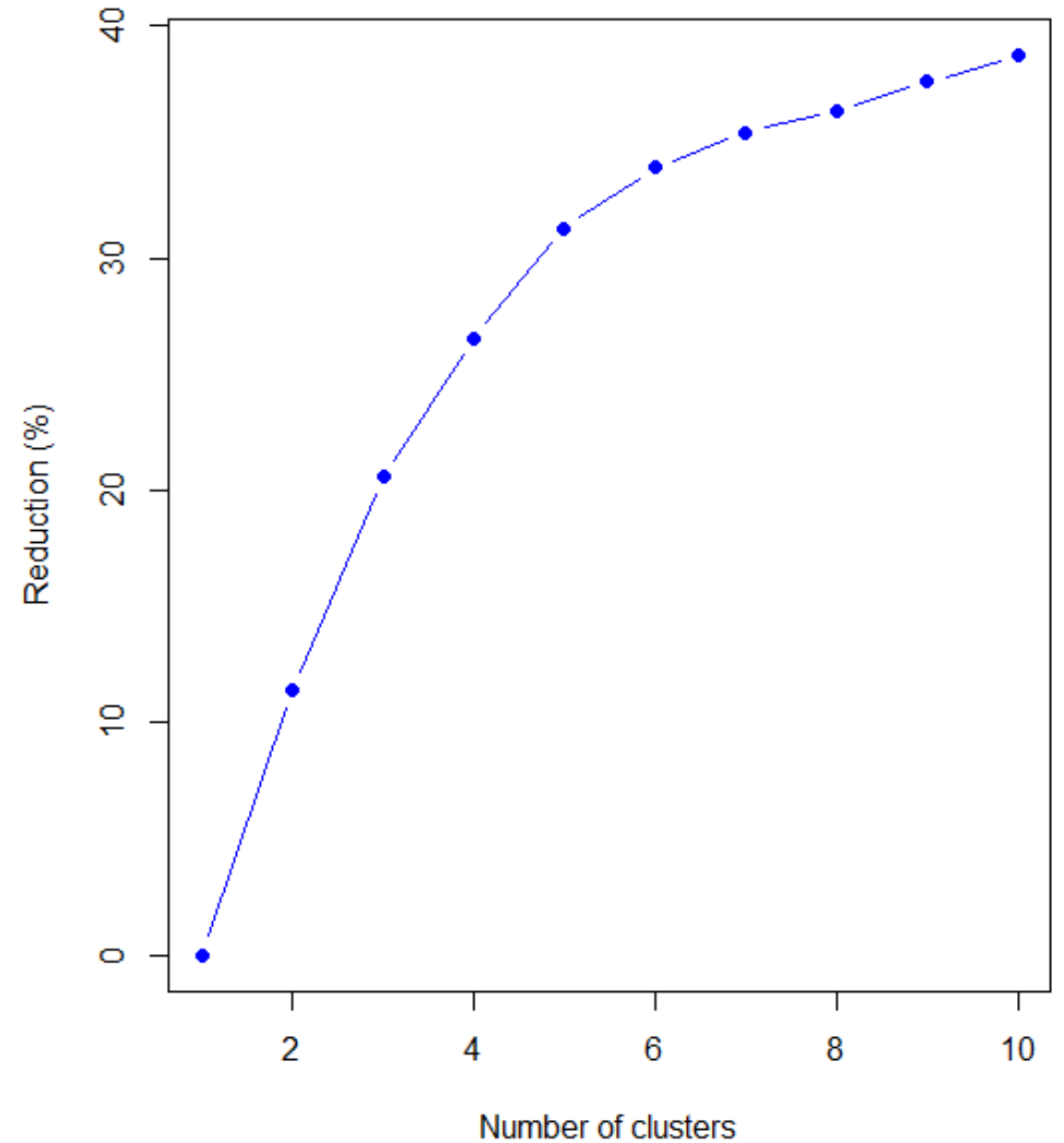
	Dams							
Sires	C1	C2	C3	C4	C5	C6	C7	C8
C1	0.26	0.20	0.19				0.13	
C2	0.20	0.20						
C3	0.18		0.18				0.13	
C4				0.14				
C5	0.12		0.12		0.26			0.13
C6								
C7	0.13		0.13				0.22	0.13
C8					0.13		0.12	0.17

Expected inbreeding of all candidate sires and dams: 0.121



$$\frac{\textit{between cluster SS}}{\textit{total SS}} \times 100$$

The reduction in sum of squares

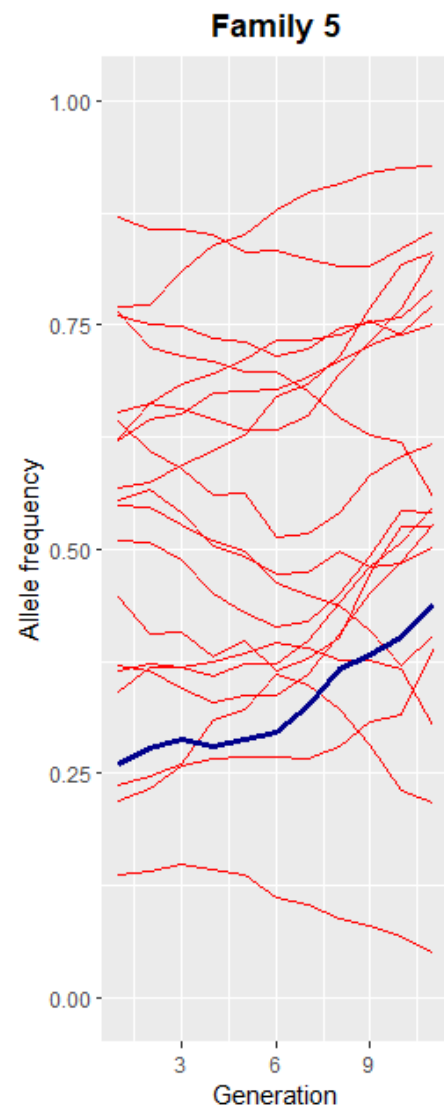
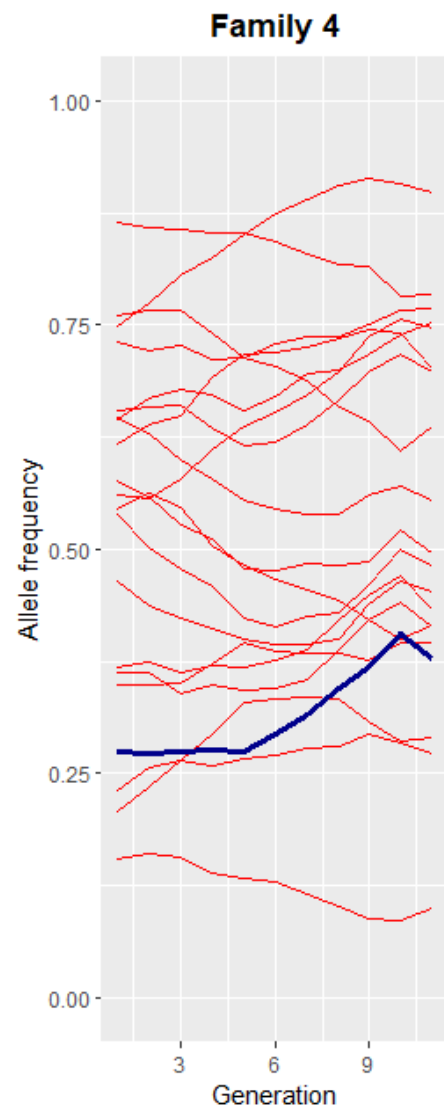
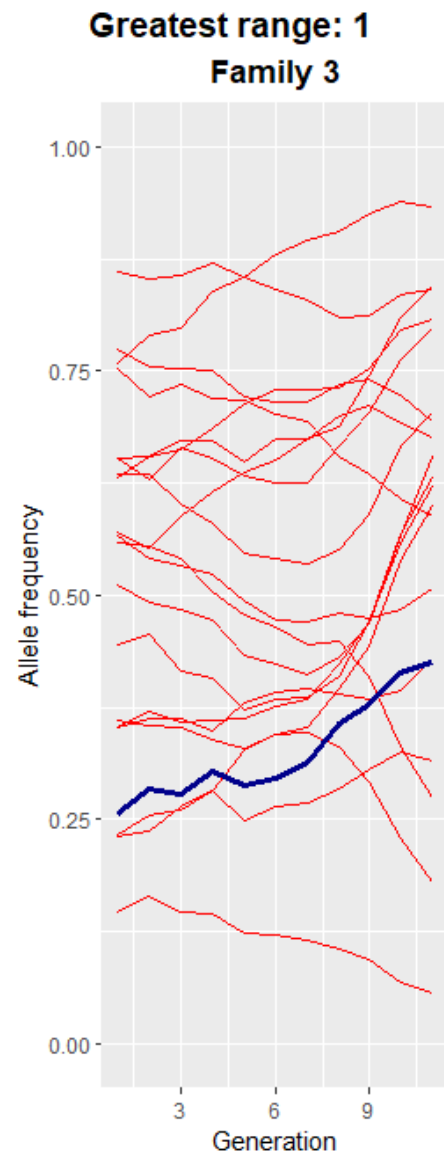
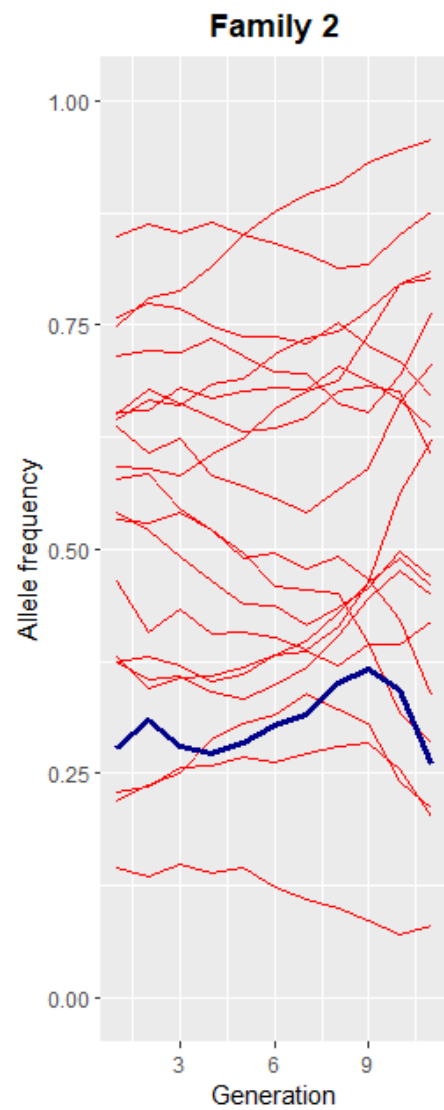
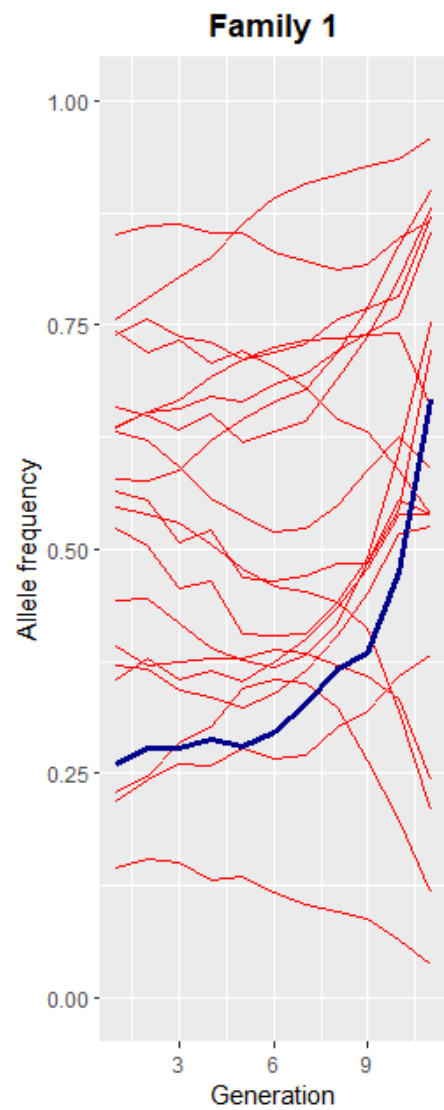


Conclusion

- PCA show important bulls
- Corresponds to k-means clustering

- Can reduce expected inbreeding
- Does not take genetic merit into account
- Selection intensity?
- Genetic redundancy





Thank you

yvette.steyn@uga.edu



**UNIVERSITY OF
GEORGIA**