Project SETI

Classification of the Unknown "Squiggle" Time Series

Frank Fan, Kenny Smith, Jason Wang
Advised by Professor Jeffrey Ullman

Goals

Supervised

1) Build a real-time classifier to distinguish between squiggle and non-squiggle
2) Build a multi-class classifier to stratify new squiggles into subgroups

Unsupervised

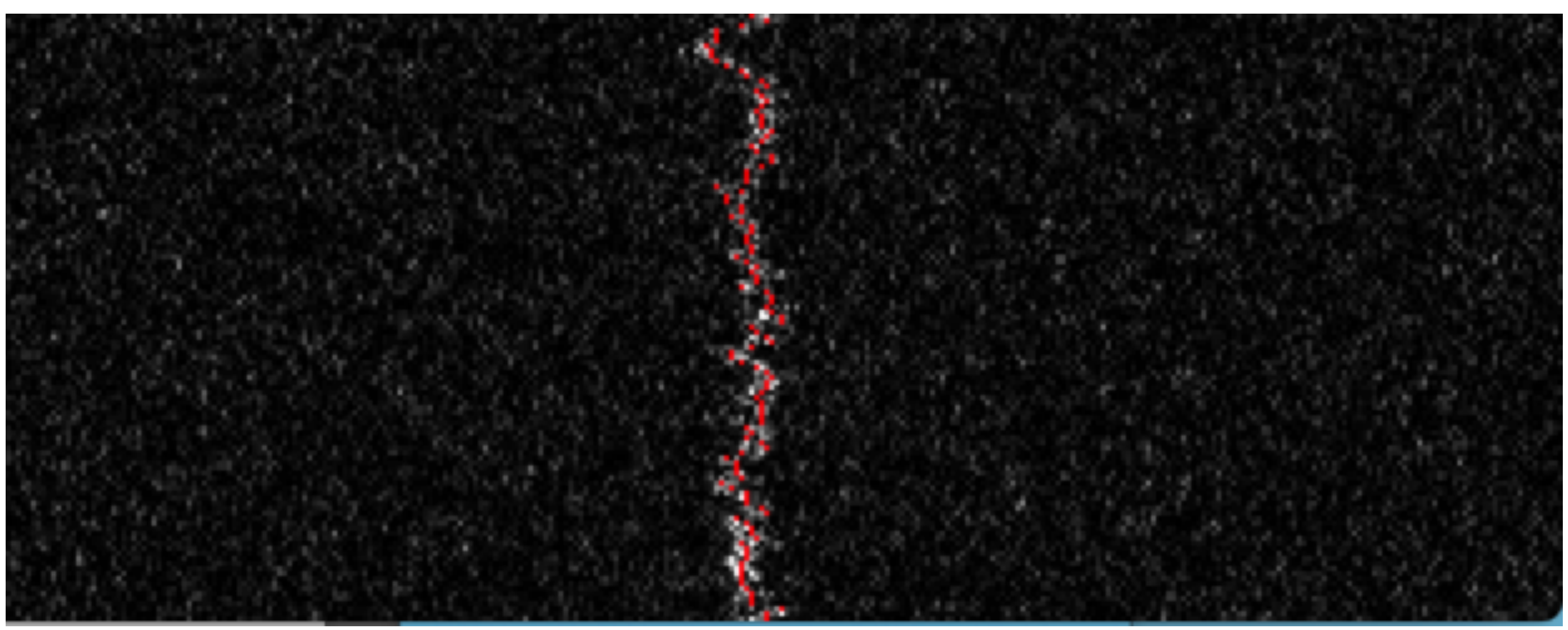
1) Identify squiggle subgroups
2) Pinpoint key characteristics of each subgroup

Approach

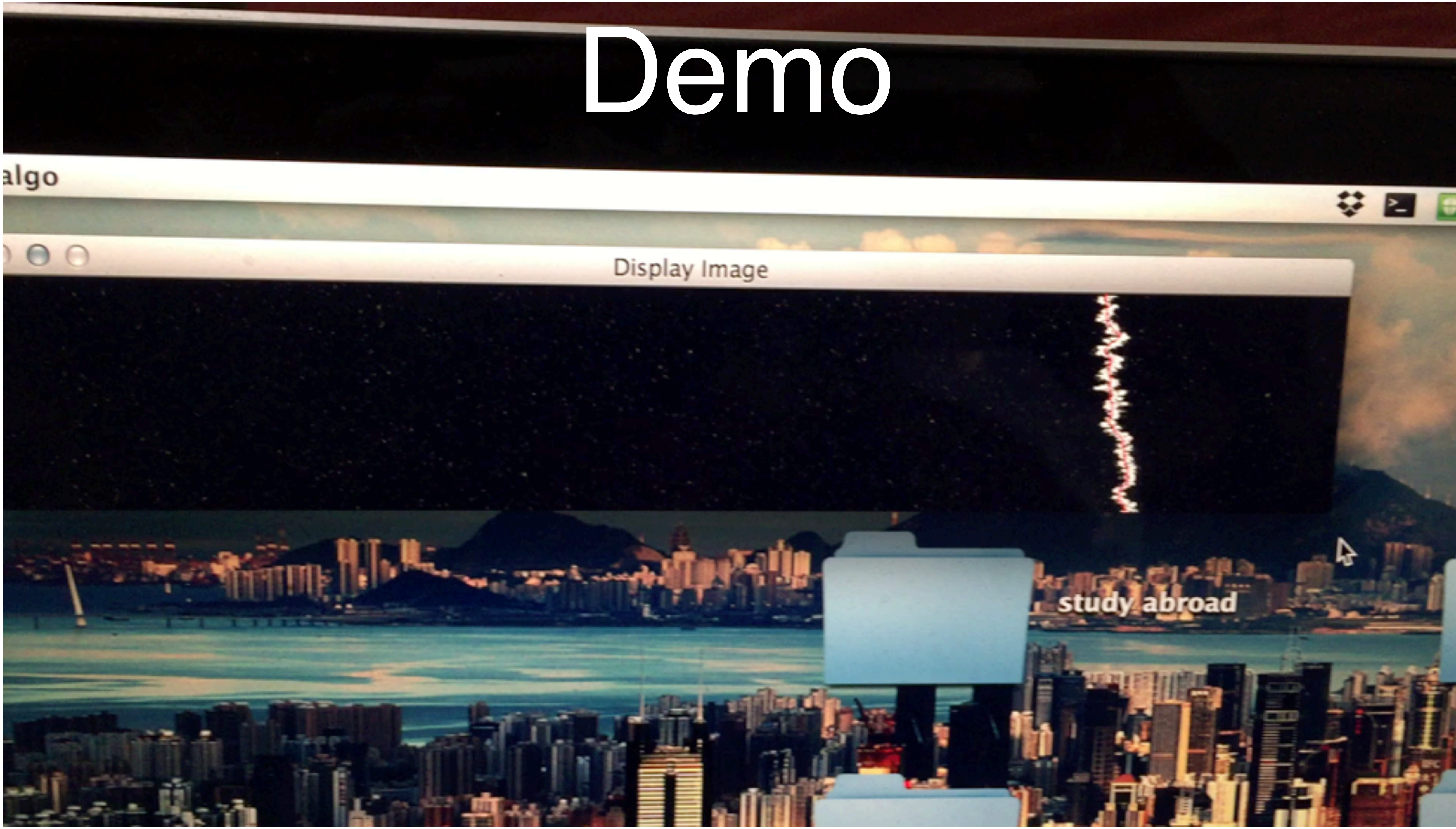
Discretization

Waterfall Plots → Discrete Time Series

$$L(\alpha,\beta) = \alpha*(\text{Intensity})+\beta*(\text{Neighboring Intensities})+(1-\alpha-\beta)(\text{Deviation})^2$$



Demo



Feature Extraction

- 63 Discrete Fourier Transform samples
- Variance of raw time series
- Loss from DP algorithm
- AR & MA parameters from ARIMA(1,1,1) fitted model
- MSE from fitted linear regression*

*Key feature for squiggle v.s. non-squiggle

68 features total normalized to unit variance, mean 0

Progress

Supervised

Dataset: discretized time series

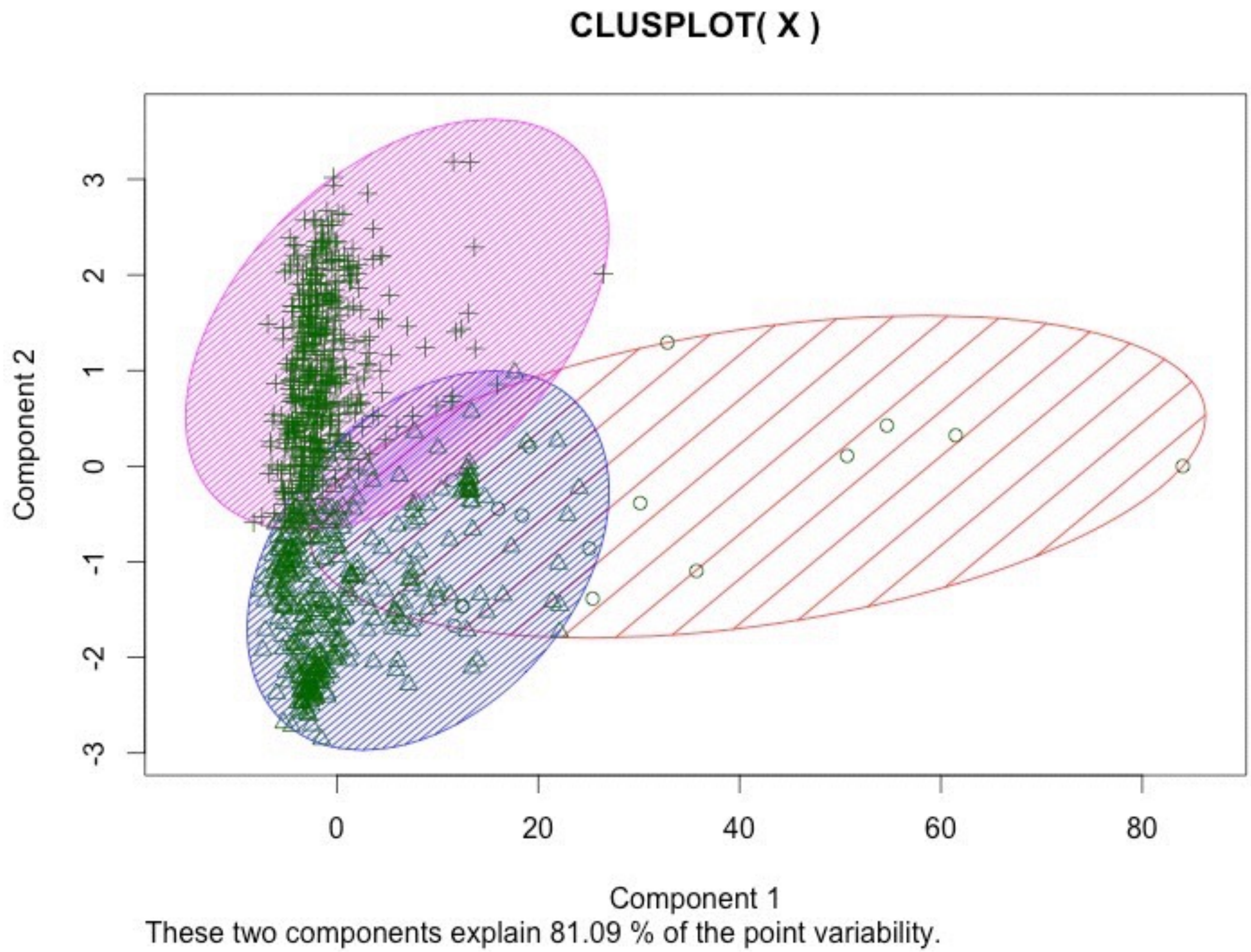
- 833 Squiggle Examples
- 843 Non-Squiggle Examples

Model	AUC	ACC
Lasso Regularization on Validation Set	0.996	0.994
Lasso Regularization using CV	0.995	0.984
Logistic Regression	0.999	0.988

Unsupervised

Iterative K-Means: Euclidean

Round 1

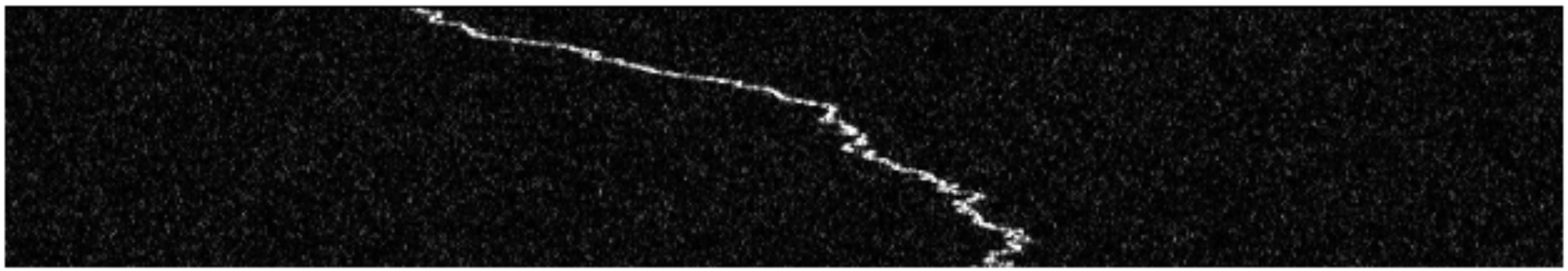


Unsupervised

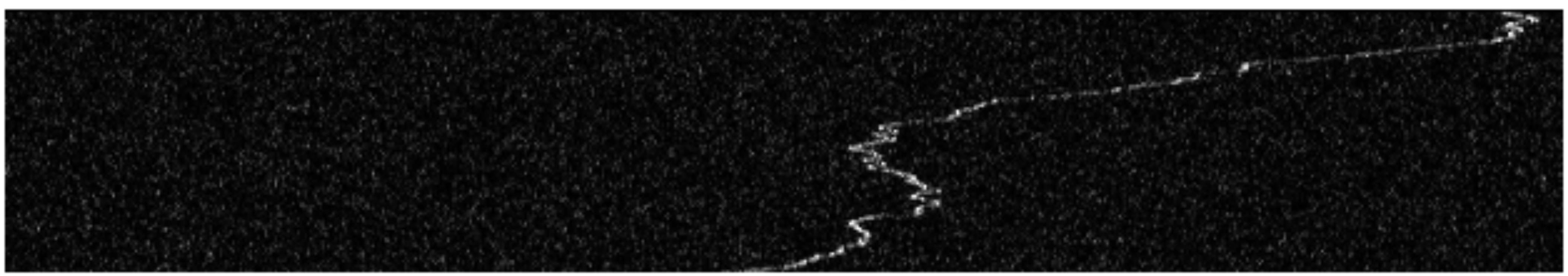
Iterative K-Means: Euclidean

Cluster 3: Greatest variance in frequency, independent of intensity

2014-09-12_03-22-09_UTC.act32064.dx1006.id-5.L.png



2014-09-19_02-52-15_UTC.act33784.dx1008.id-3.L.png



2014-09-06_05-55-05_UTC.act30577.dx1016.id-0.L.png

Unsupervised

Iterative K-Means: Euclidean

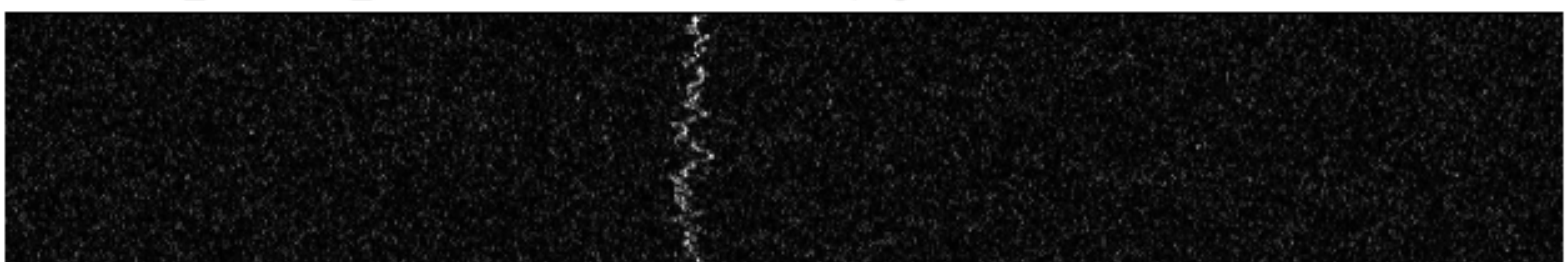
Cluster 1: Least variance in frequency, low intensity

2014-11-02_12-12-19_UTC.act40628.dx3014.id-4.R.png

2014-11-01_06-36-46_UTC.act40177.dx3003.id-0.L.png

Cluster 2: Low variance in frequency, higher intensity

2014-11-02_08-42-43_UTC.act40567.dx3041.id-4.L.png

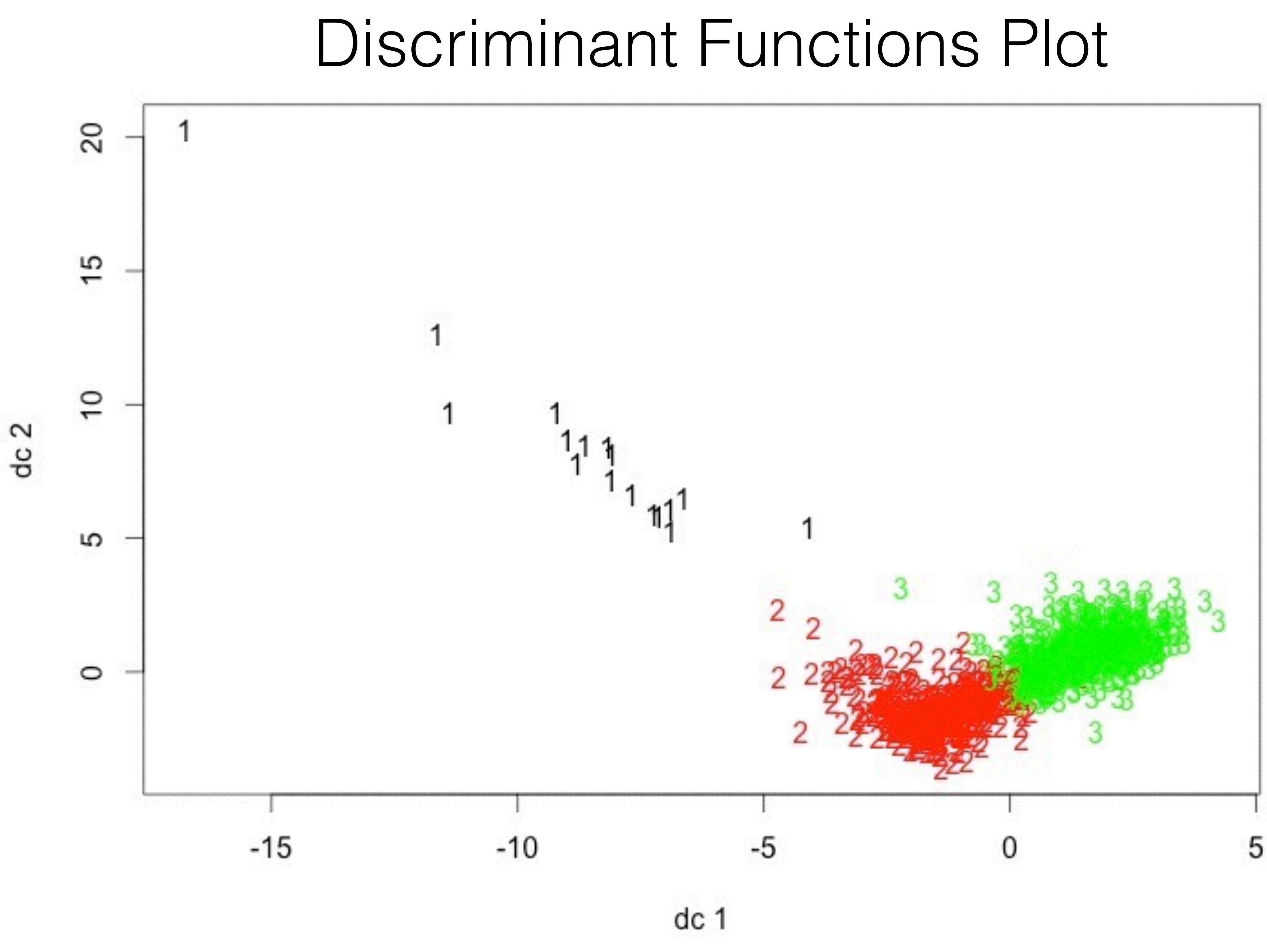


2014-11-01_06-28-14_UTC.act40174.dx3036.id-3.R.png

Unsupervised

Iterative K-Means: Euclidean

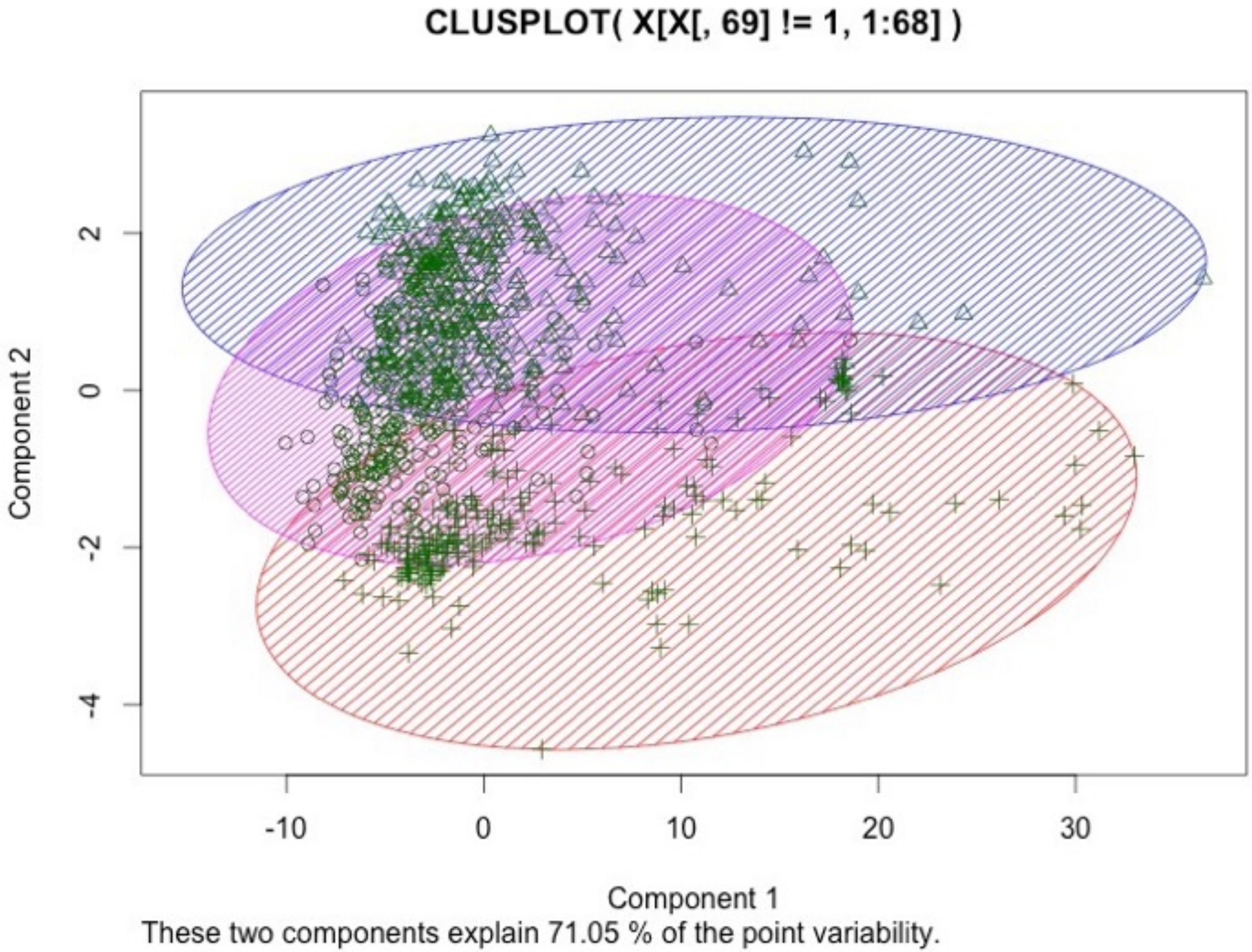
Round 1



Unsupervised

Iterative K-Means: Euclidean

Round 2

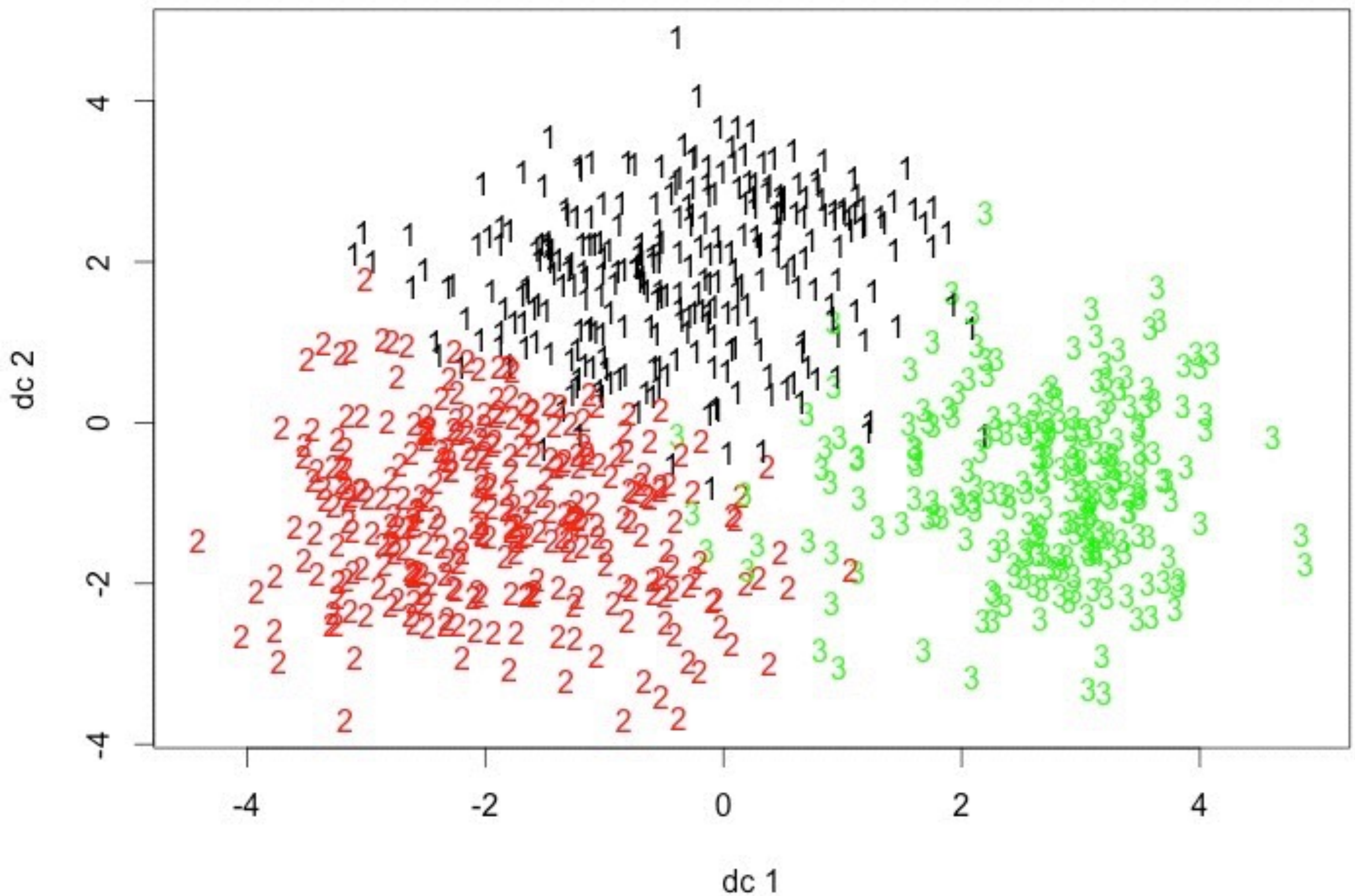


Unsupervised

Iterative K-Means: Euclidean

Round 2

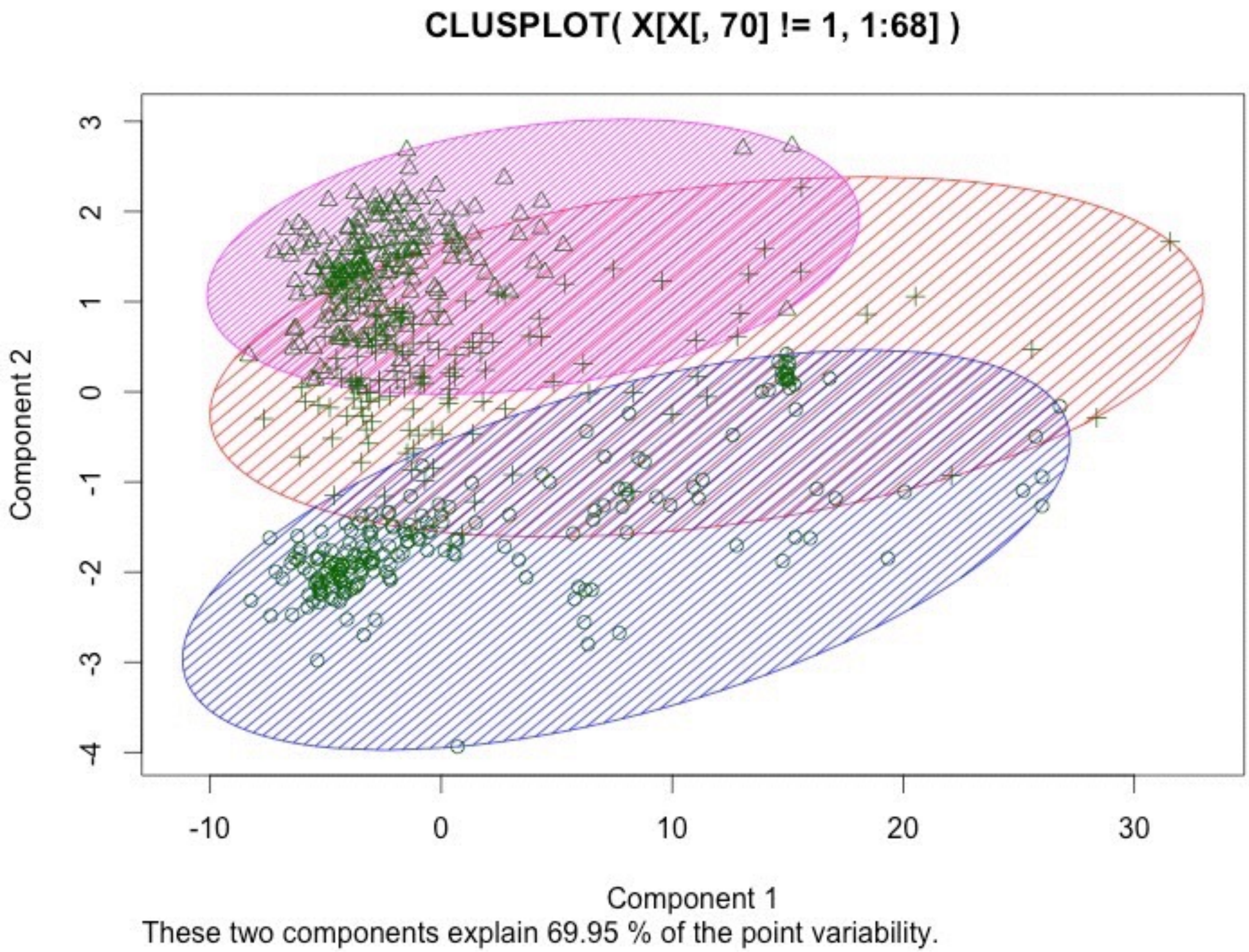
Discriminant Functions Plot



Unsupervised

Iterative K-Means: Euclidean

Round 3

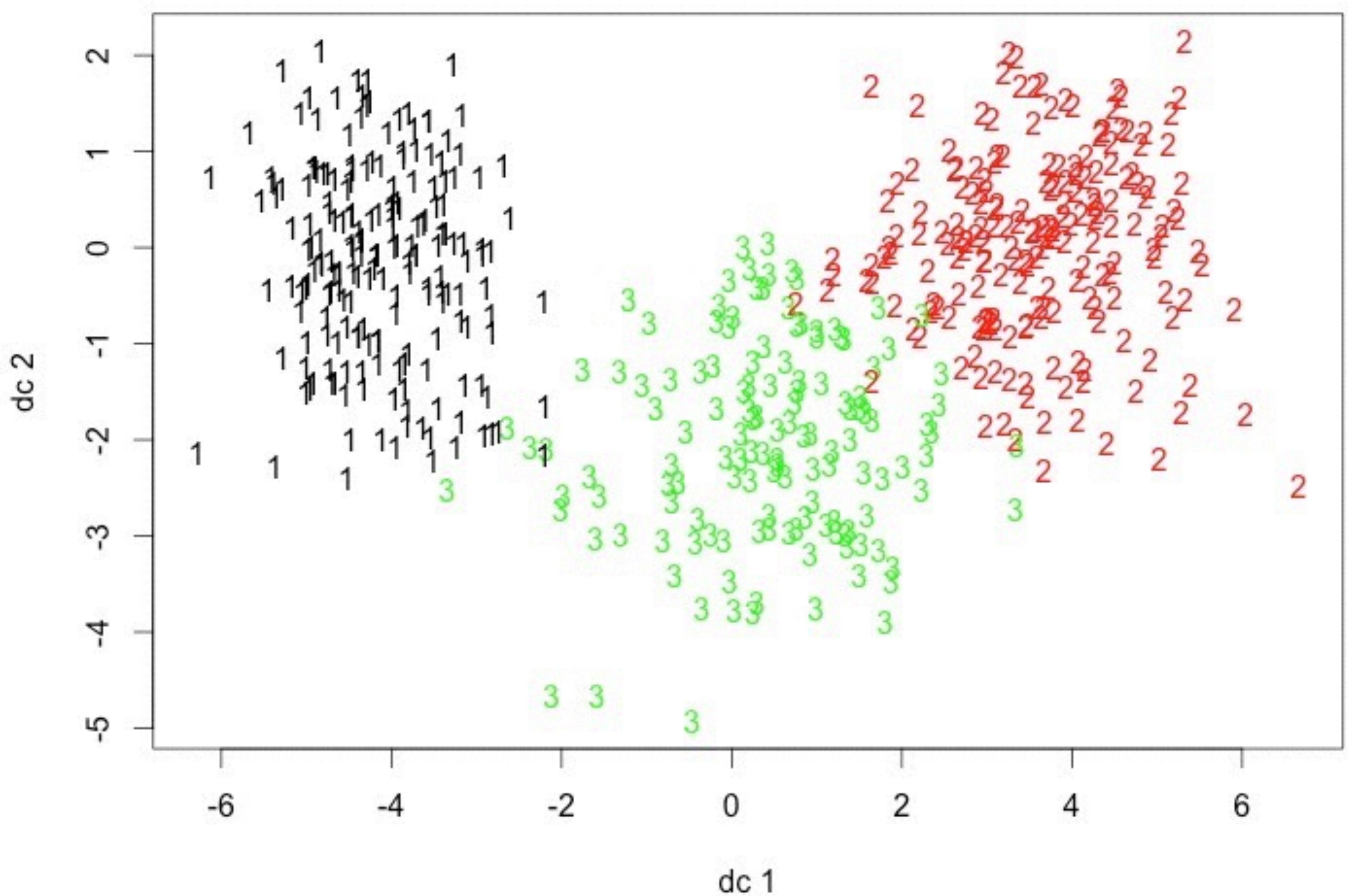


Unsupervised

Iterative K-Means: Euclidean

Round 3

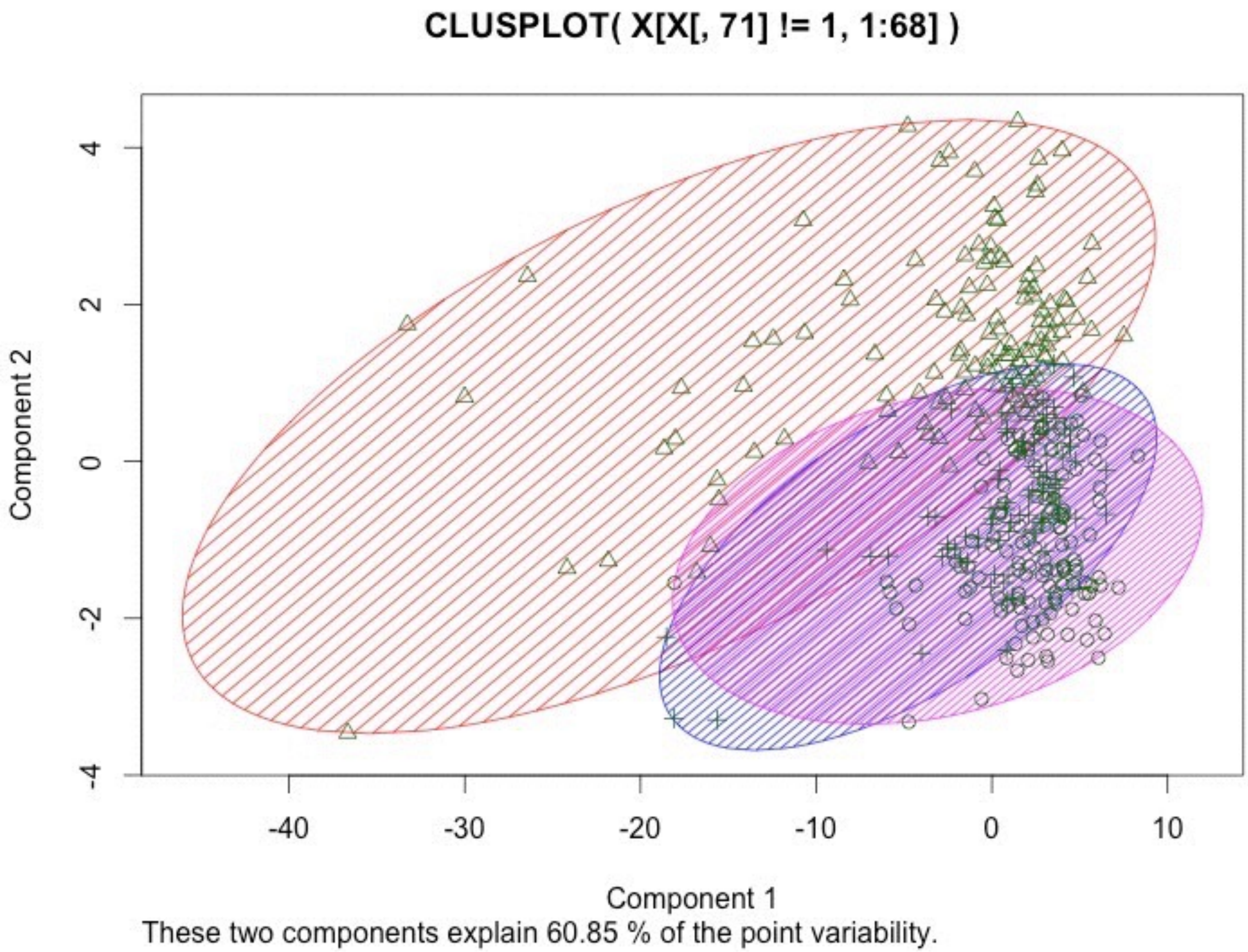
Discriminant Functions Plot



Unsupervised

Iterative K-Means: Euclidean

Round 4

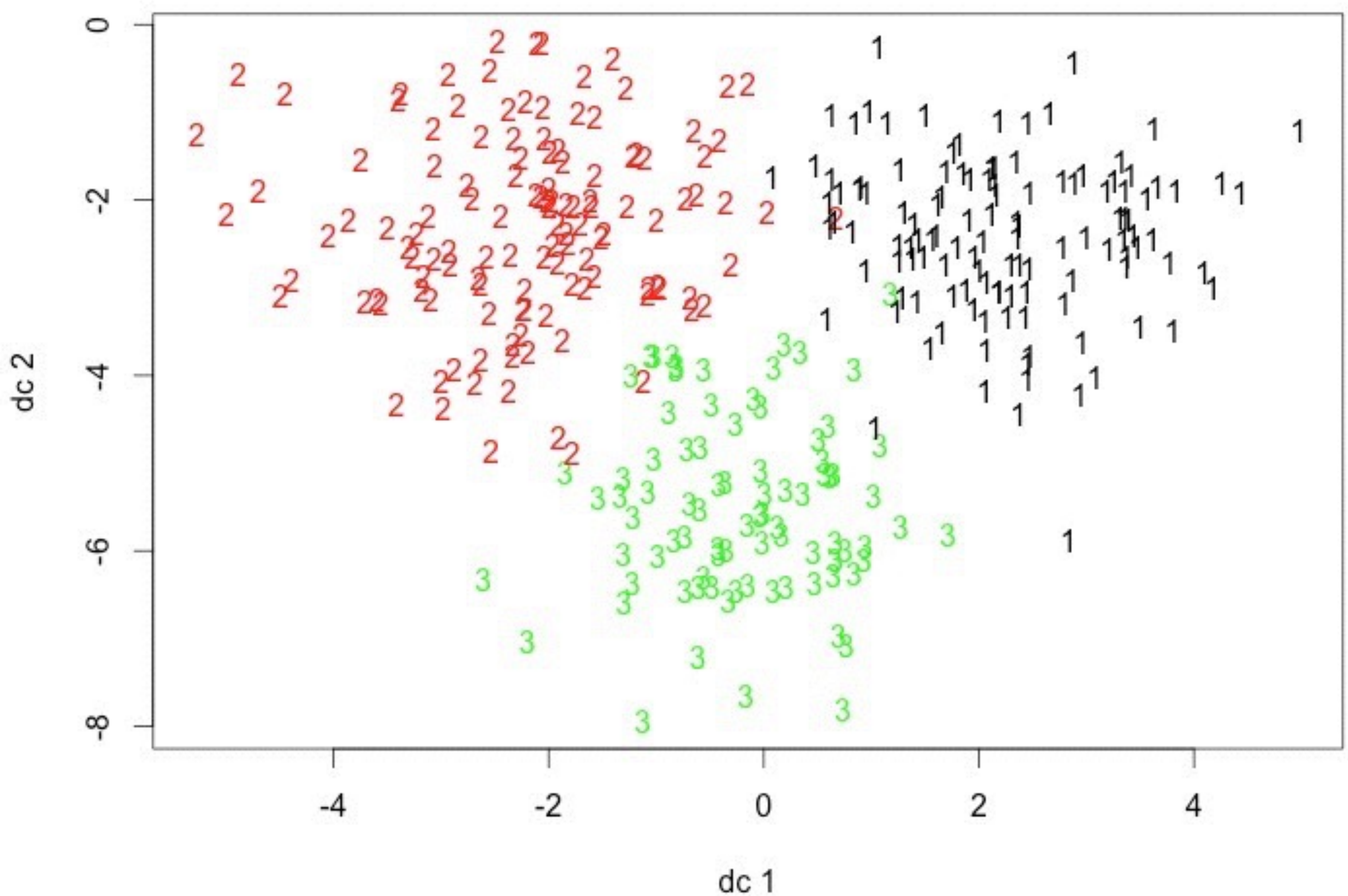


Unsupervised

Iterative K-Means: Euclidean

Round 4

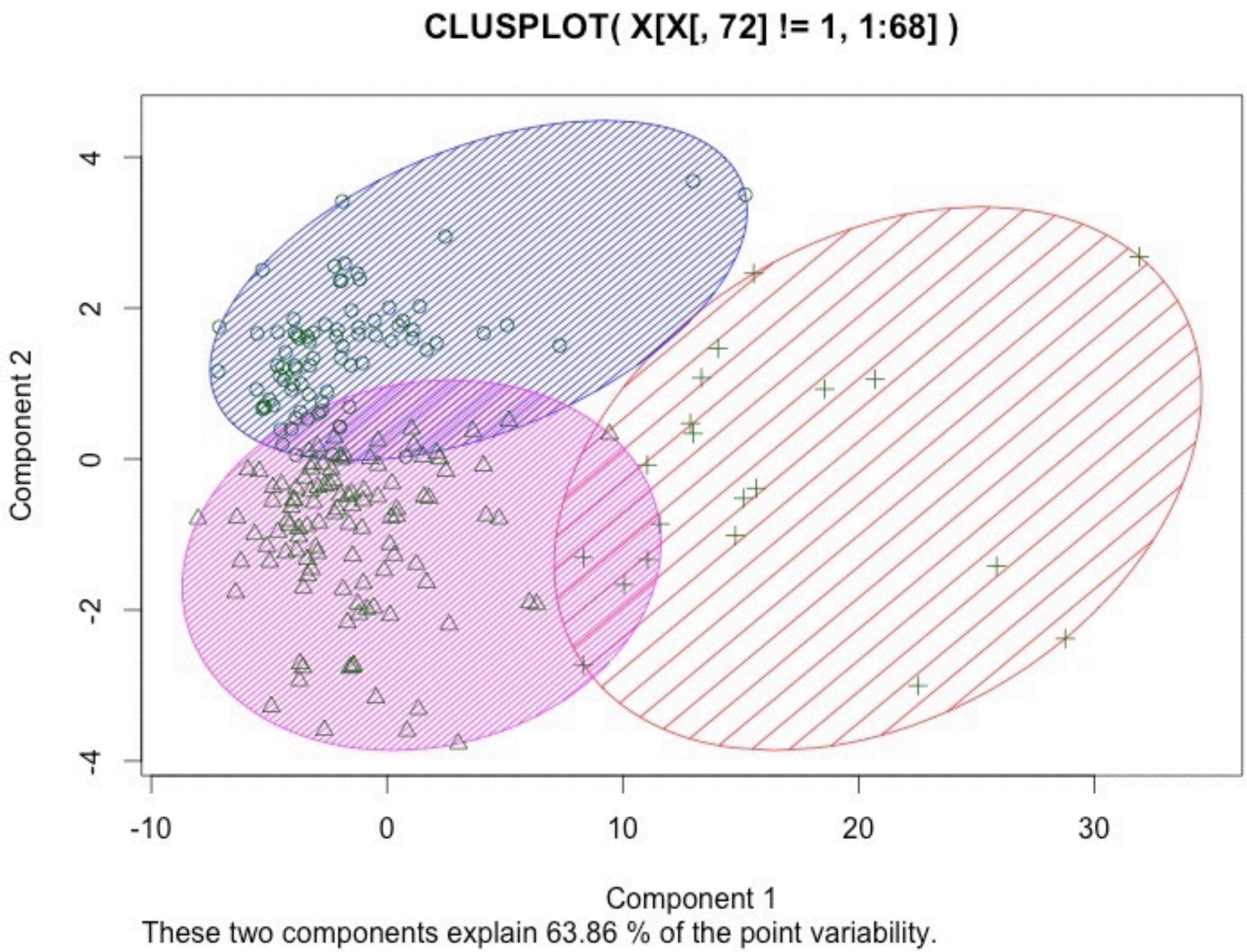
Discriminant Functions Plot



Unsupervised

Iterative K-Means: Euclidean

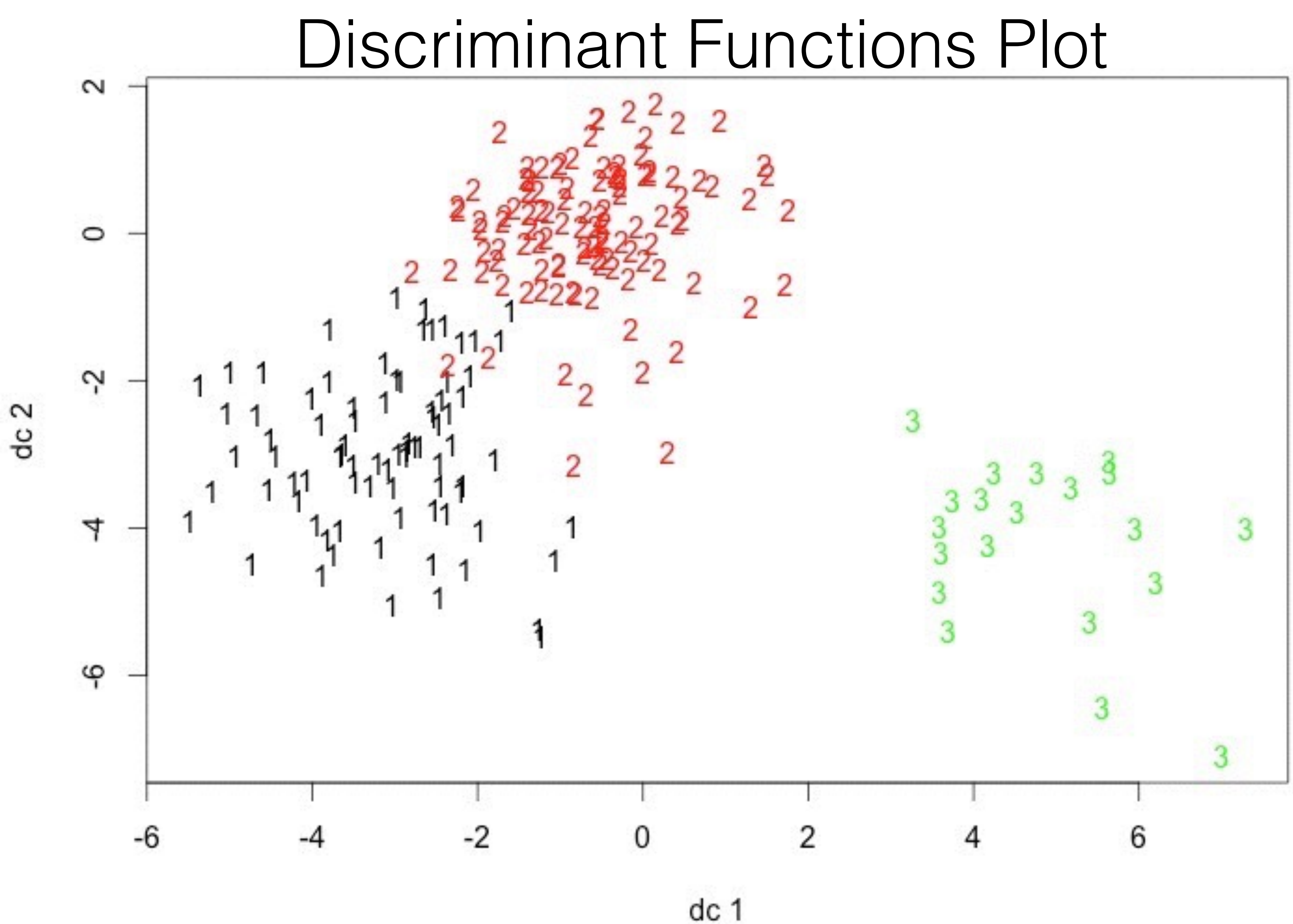
Round 5



Unsupervised

Iterative K-Means: Euclidean

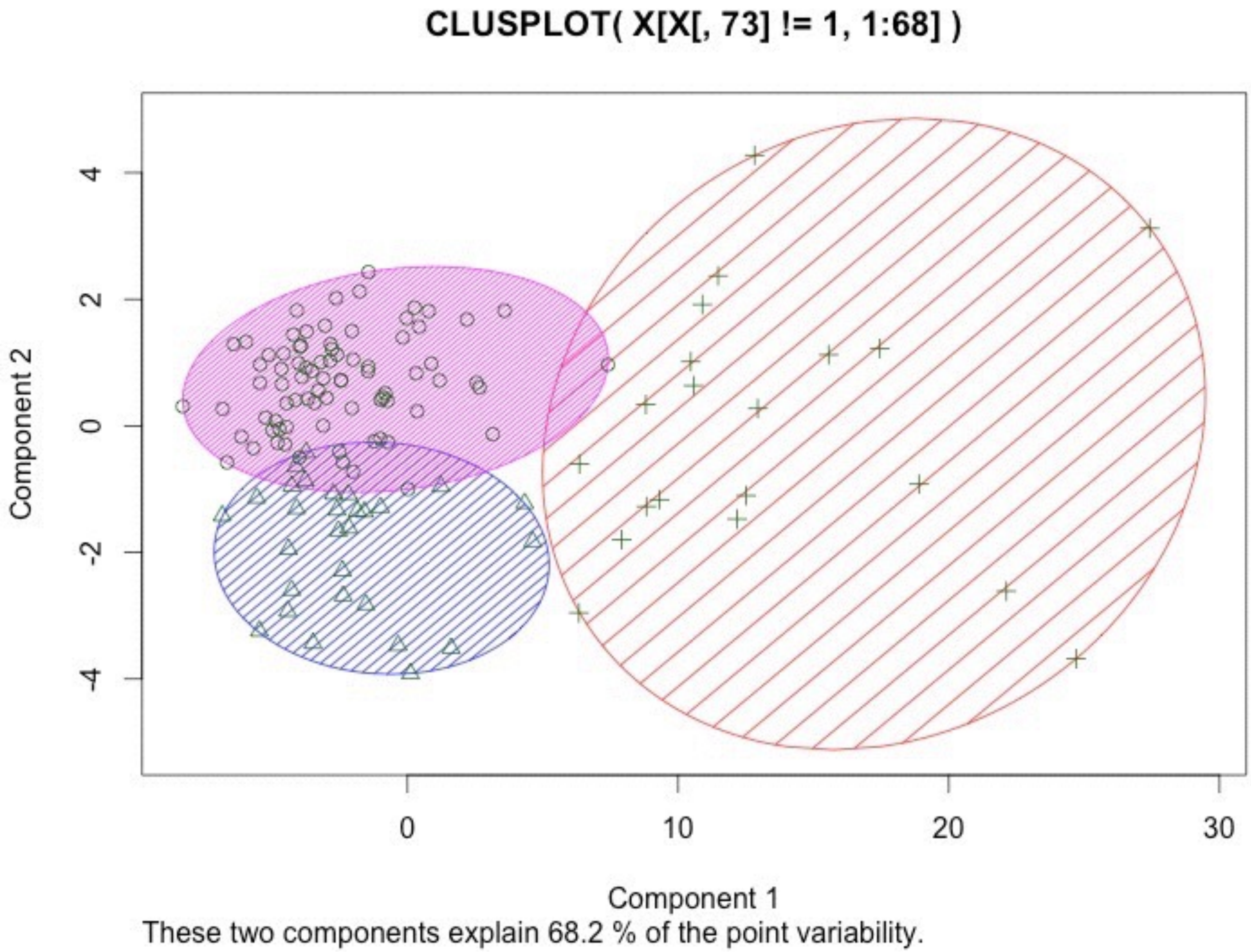
Round 5



Unsupervised

Iterative K-Means: Euclidean

Round 6

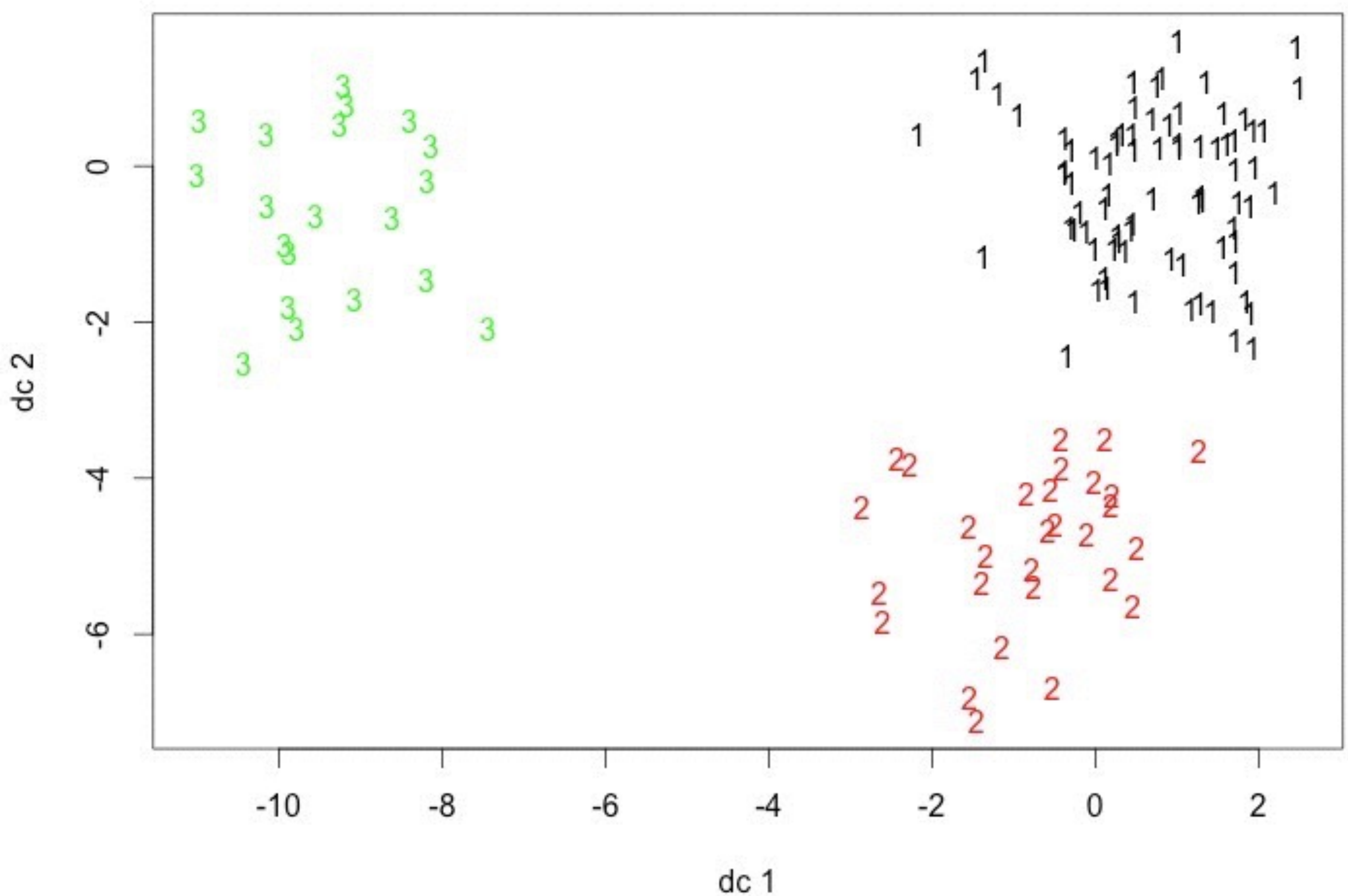


Unsupervised

Iterative K-Means: Euclidean

Round 6

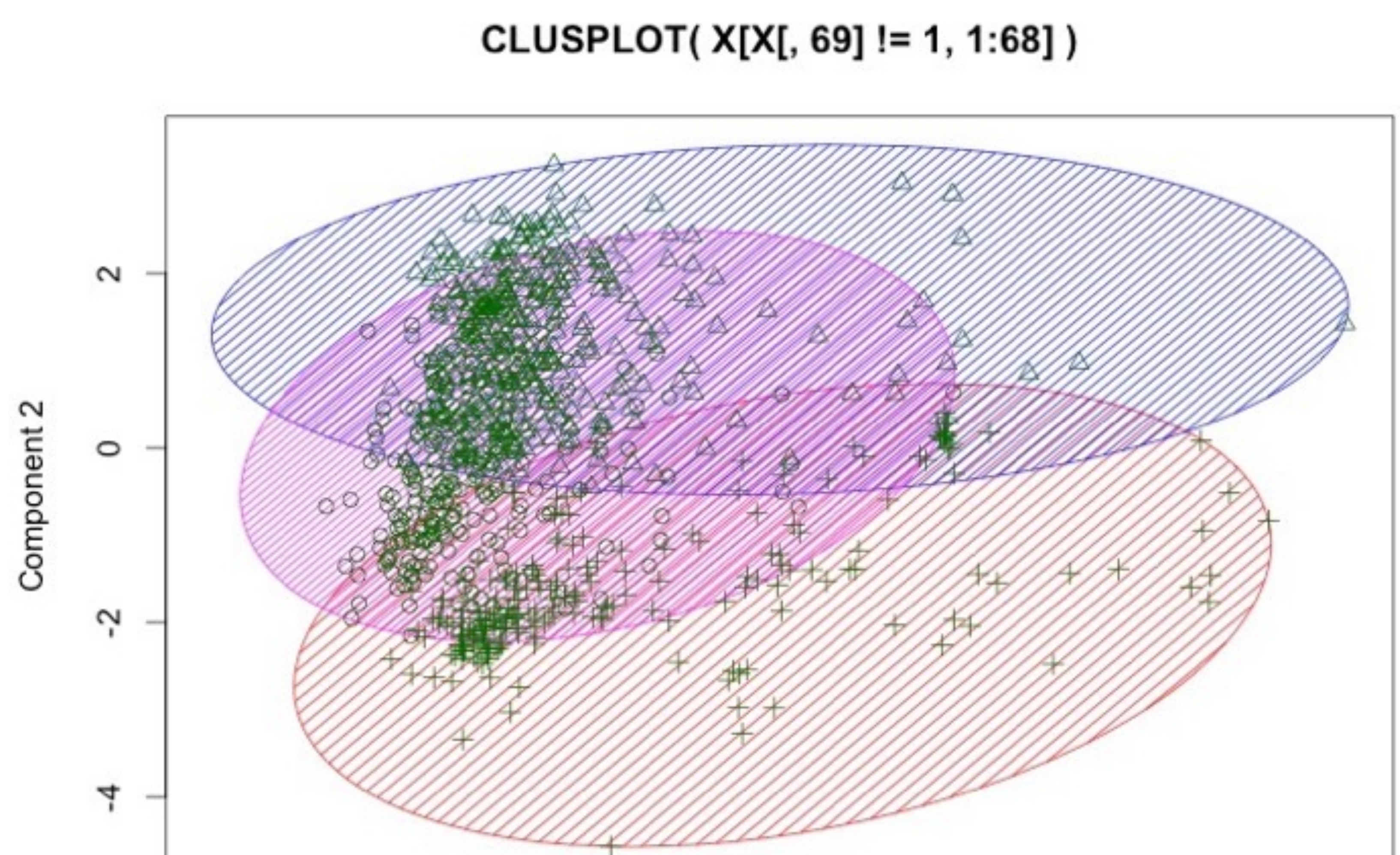
Discriminant Functions Plot



Conclusions

Mostly Hypotheses

- Modulation & variance in frequency = energy inputted by external source, extraterrestrial or not
- Majority squiggles have an archetypal energy, higher modulating squiggles are deviants



Conclusions

Mostly Hypotheses

- Transpositions across bandwidths is common
- Red shift as Earth is rotating towards source?

Live Demo?

Next Steps

Supervised

1) Identify which features are most significant in squiggle v.s. non-squiggle classifier
2) Build a multi-class classifier to stratify new squiggles into subgroups

Unsupervised

1) Iterative cluster in search of distinct archetypes
2) Exploratory data analysis to identify key attributes:

Periodicity (All at night? All on one day?), Modulation, etc.