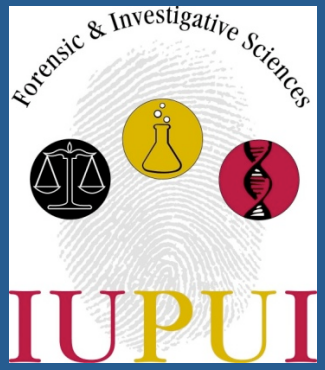
Analysis of Trace Evidence Using Microspectrophotometry and Multi-Variate Statistics



John V. Goodpaster, Ph.D.
Department of Chemistry and Chemical Biology
Forensic and Investigative Sciences Program
Indiana University Purdue University Indianapolis

Outline

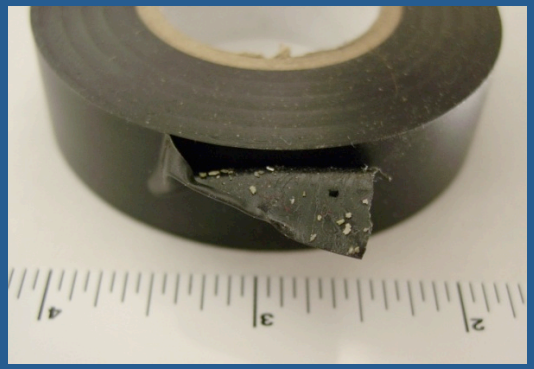
- The Nature of Class Evidence
- The Role of Chemometrics
- Research at IUPUI
  - Dyed Hair (Julie Barrett and Jay Siegel)
  - Clearcoats (Elisa Liszewski, Simon Lewis and Jay Siegel)
  - Dyed Cotton Fibers (Elisa Liszewski)

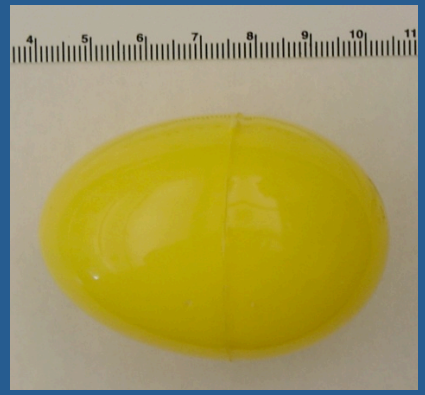
The Realities of Trace Evidence Comparisons

- You are ask to compare two items:
  - A sample, collected by someone else, from a questioned source that consists of a small amount of poor quality material deposited under unknown conditions
  - A exemplar from a known source that consists of a larger amount of good quality material which can be sampled by you under known conditions
- How can we reliably approach this analytical problem?

Some “Q”s

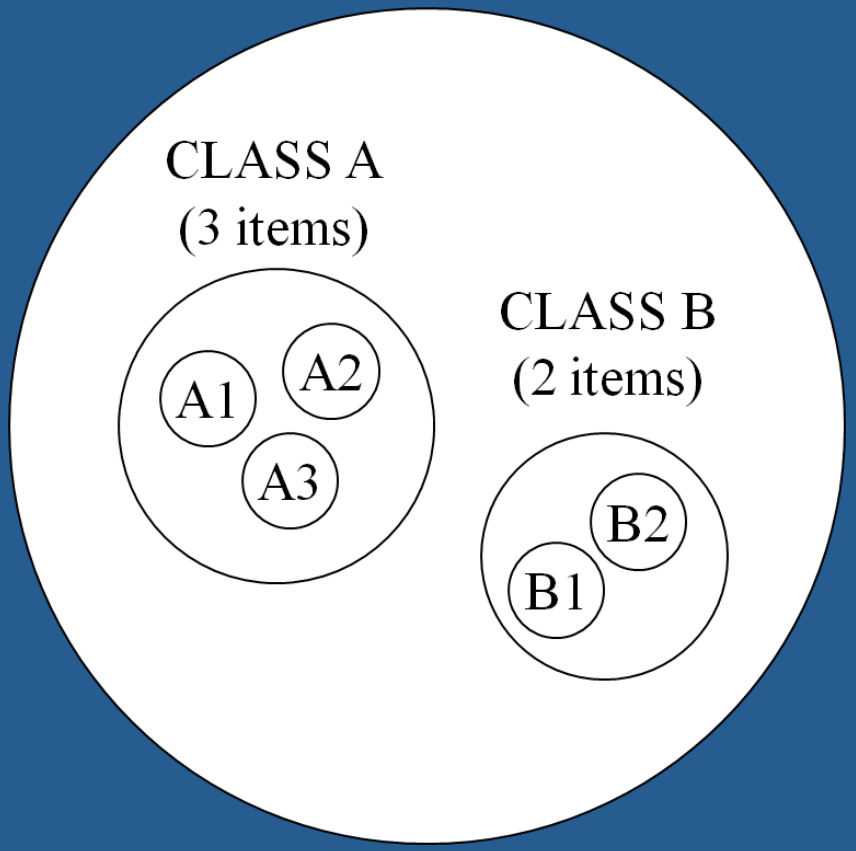
Some “K”s





Step 1: Understand the Nature of Class Evidence

- In order for class evidence to be useful, it must be:
  - Widespread
  - Diverse
  - Reliably differentiable
- We ACCEPT and EXPECT that some items will be in the same class and will have indistinguishable physical and/or chemical characteristics.



Step 2: Apply "Forensic Logic"

- Given:
  - An exemplar (K) from a known source
  - A questioned sample (Q) from an unknown source
- We MUST Assume that the source is homogenous and/or K and Q are representative
- This leads us to two statements
  - Statement A: Q and K originated from the same source
  - Statement B: Q and K have indistinguishable class characteristics

Premise: If A then B

Consequence: If not B then not A (Exclusion)

HOWEVER – "If B Then A" does not automatically follow!!!

Step 3: "Don't oversell the evidence. . . but give it its due"

- Exclusion
  - "Q is excluded as having the same source as K"
- Association
  - "Q is consistent with K"
  - "Q could not be excluded as having the same source as K"
  - "Q could have originated from K or another source with the same physical and chemical characteristics"
- How can we articulate the significance of an association?
- Ideally, we know something about the number of classes and/or their relative size

Assessing Significance of Trace Evidence Comparisons

- Class Discrimination Studies
  - A collection of samples is analyzed using multiple techniques and the extent to which samples can be placed into multiple differentiable classes is assessed
- Pair-Wise Studies
  - A collection of samples is analyzed using multiple techniques and the extent to which pairs of samples remain indistinguishable is assessed
- Environmental Studies
  - A specific type of sample is selected and the frequency with which that sample is found by chance is assessed

A Comprehensive Approach

1. Understand the sample population
2. Acquire a large and representative collection
3. Use multiple, orthogonal analytical methods that are not subject to microheterogeneity
4. Acquire replicate data and quantitatively assess reproducibility and the extent of differentiation (Chemometrics)
5. Monitor changes in the sample population over time

The Role of Chemometrics

Unsupervised

- "Here's my data, how is it organized?"
- No prior knowledge of groupings is input into the analysis
- Examples
  - Clustering Analysis
  - PCA

Supervised

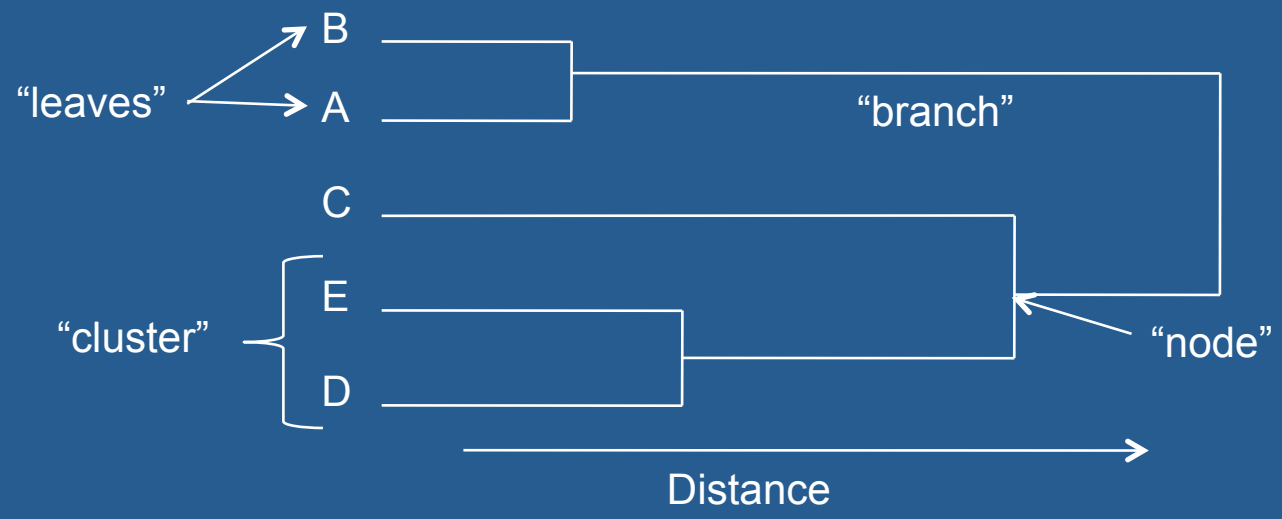
- "Here's my data, how does this new data fit in?"
- Data is grouped by the user prior to analysis
- Examples
  - Discriminant Analysis

Typical Chemometric Procedure

1. Pre-treatment
    1. Baseline correction
    2. Normalization (sum of squares)
    3. Mean centering
2. Agglomerative Hierarchical Clustering (AHC)
3. Principal Components Analysis (PCA)
4. Discriminant Analysis (DA)

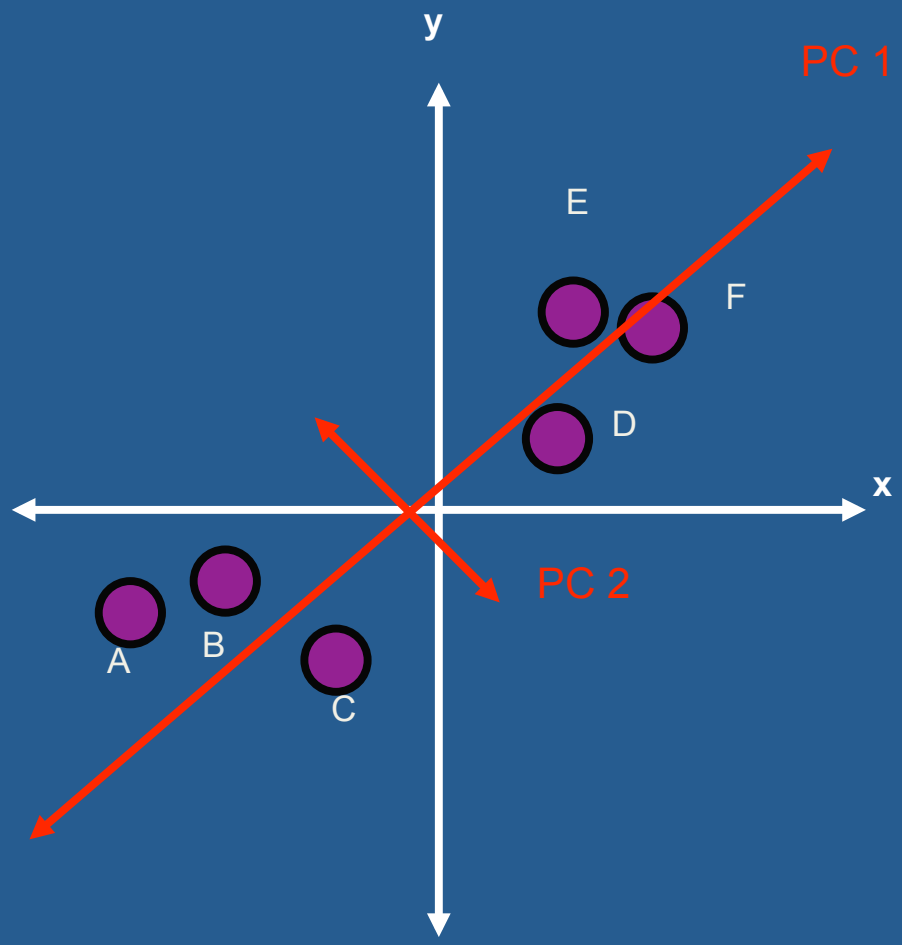
Agglomerative Hierarchical Clustering (AHC)

- Makes up homogeneous groups of objects (classes) on the basis of their description by a set of variables
- Produces a dendrogram, whose root is the class that contains all the observations.
- One can choose to truncate the tree at a given level, depending upon either user-defined constraints or more objective criteria.



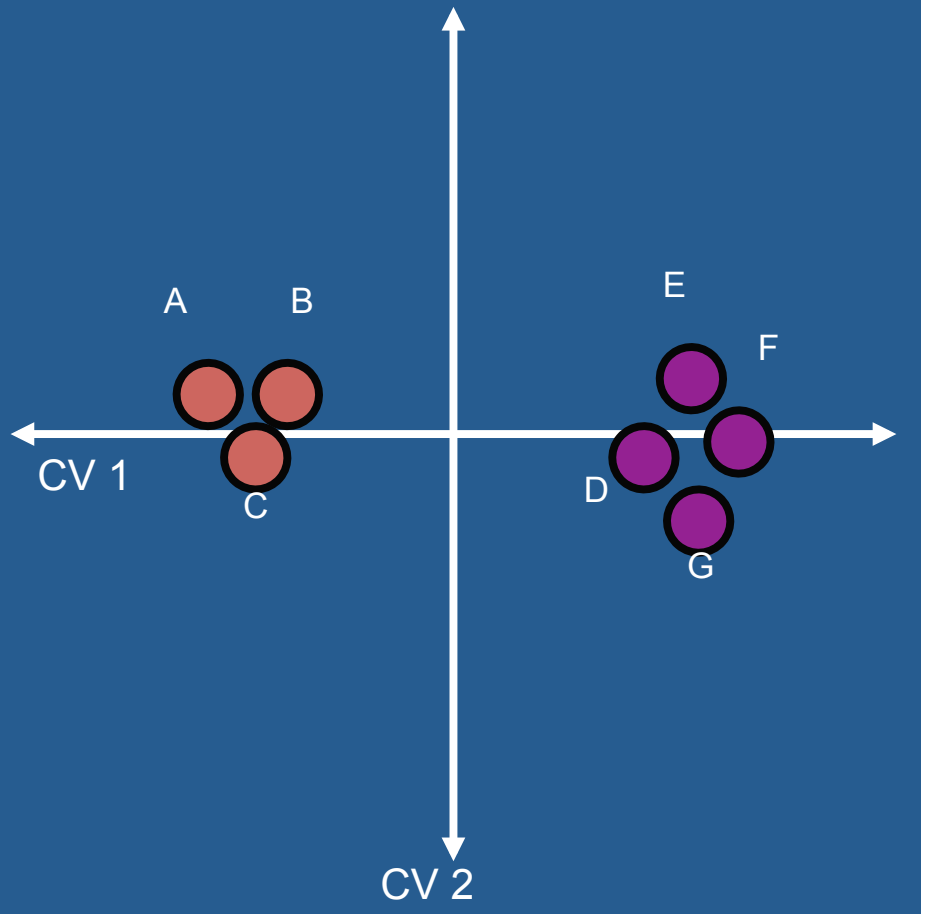
Principal Components Analysis (PCA)

- Visualizes correlations between variables
- Obtains non-correlated factors which are linear combinations of the initial variables
- Visualizes observations in a 2- or 3-dimensional space



Discriminant Analysis (DA)

- Given observations with known groups (a "learning sample"), DA optimizes these groups using canonical variates (CVs)
- New observations are then assigned to one of the existing groups



“So, what’s in it for us?”

Chemometrics CAN:

- Help determine how many reliably differentiable groups are present in a sample population
- Find characteristics that best discriminate different samples
- Indicate classes that are easily confused
- Assess the extent to which a known and unknown belong to the same class

Chemometrics CANNOT:

- Determine an “error rate” for a forensic comparison that incorporates multiple microscopic and instrumental steps
- Decide for you whether or not a Q and K came from the same source

The Problem with Chemometrics

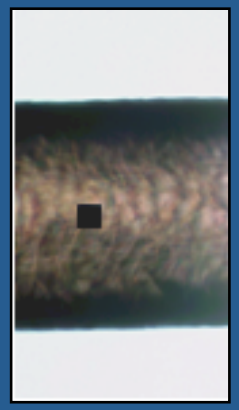
- It consists of advanced mathematical and statistical methods that can be difficult to understand
- How can we be comfortable with chemometric studies if we do not understand them?
- Are statisticians trying to usurp the role of our training and experience in interpreting data?

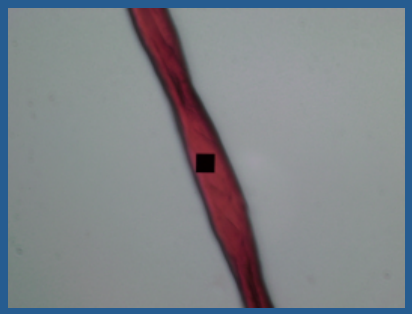
MSP and Chemometrics

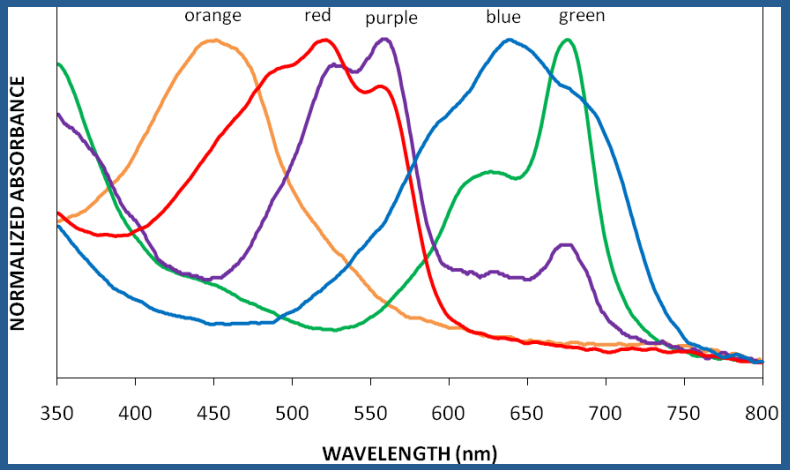
- Chemometrics on MSP data can:
  - Extract information from a large data set
  - Reduce complexity
  - Discern subtle differences in spectra
  - Provide quantitative predictions of the classification of unknown samples

UV-visible Microspectrophotometry

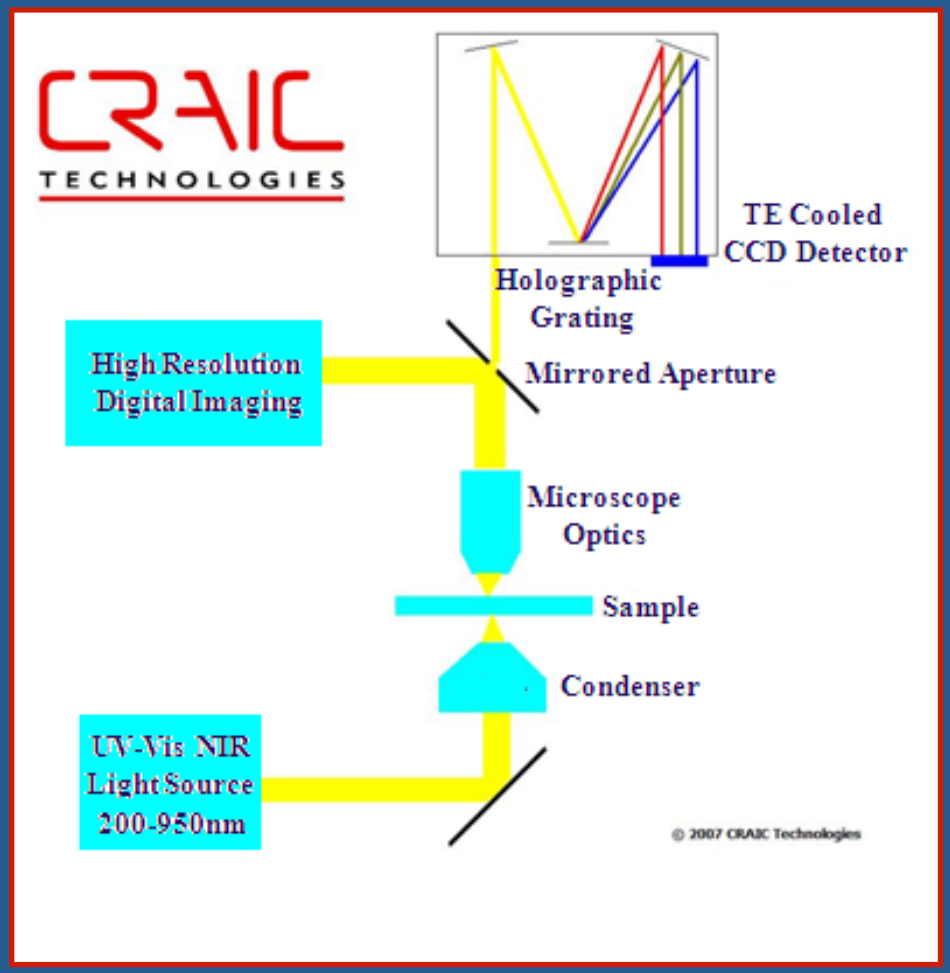
- Combines microscopy and spectrophotometry
- Small sample sizes (fibers, ink, paint)
- Non-destructive
- Minimal sample preparation
- Provides quantitative color determination





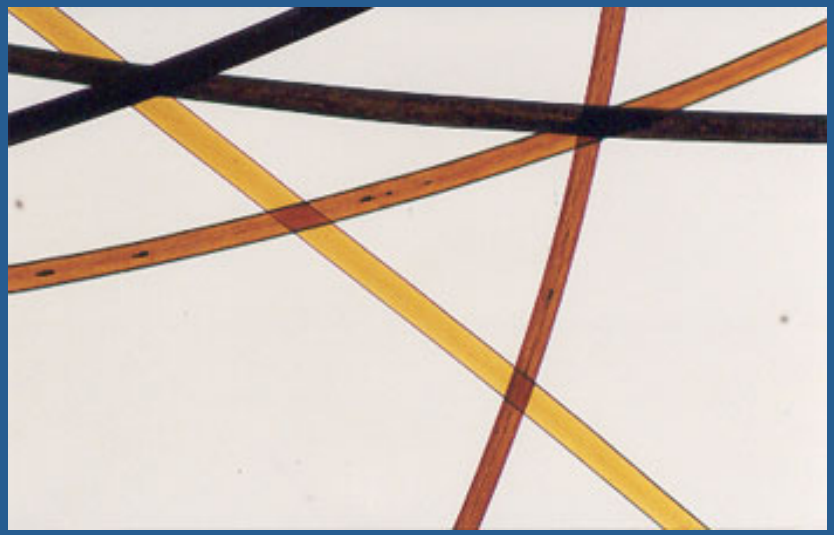


UV/visible/NIR MSP (transmission/reflectance/fluorescence)

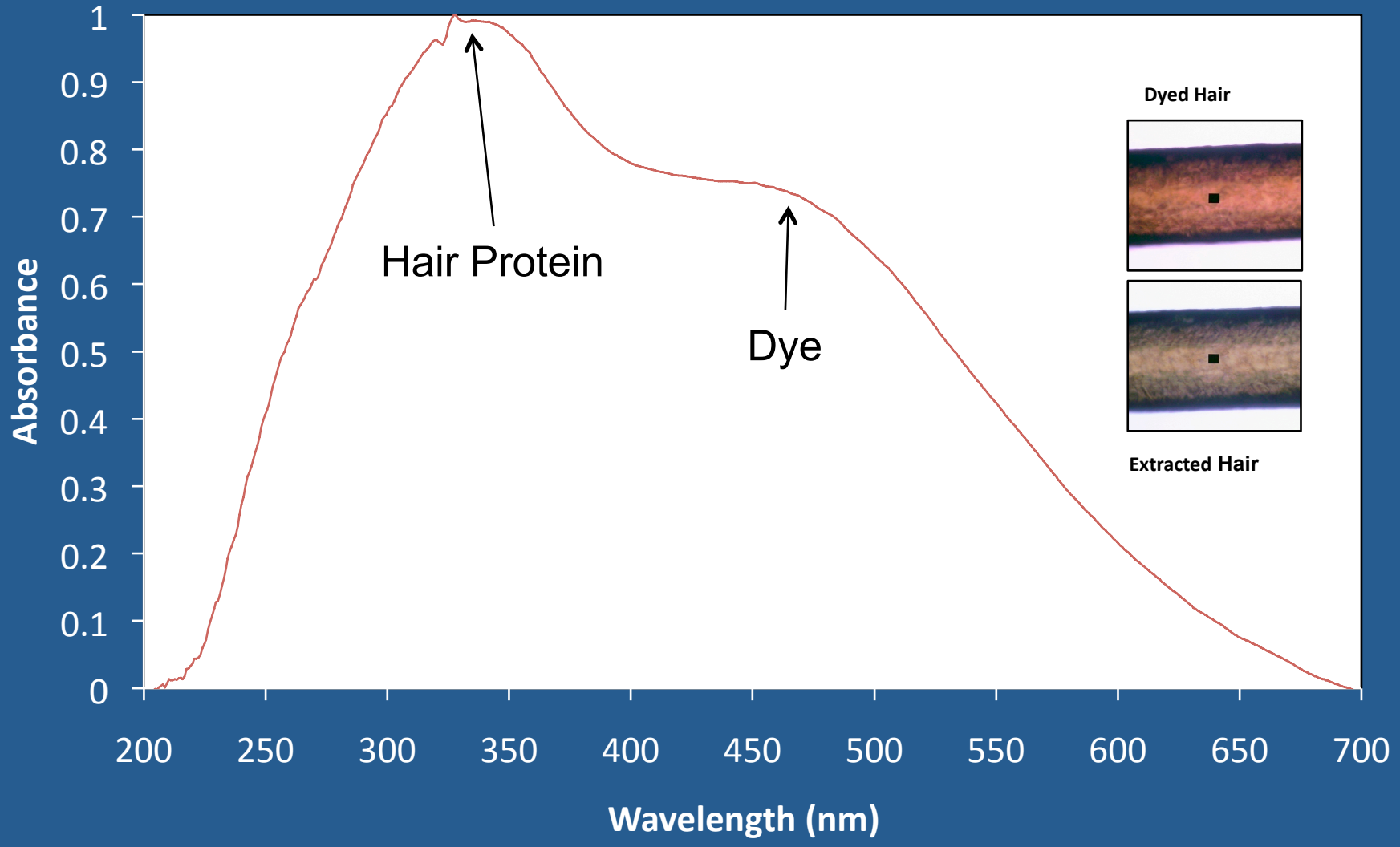


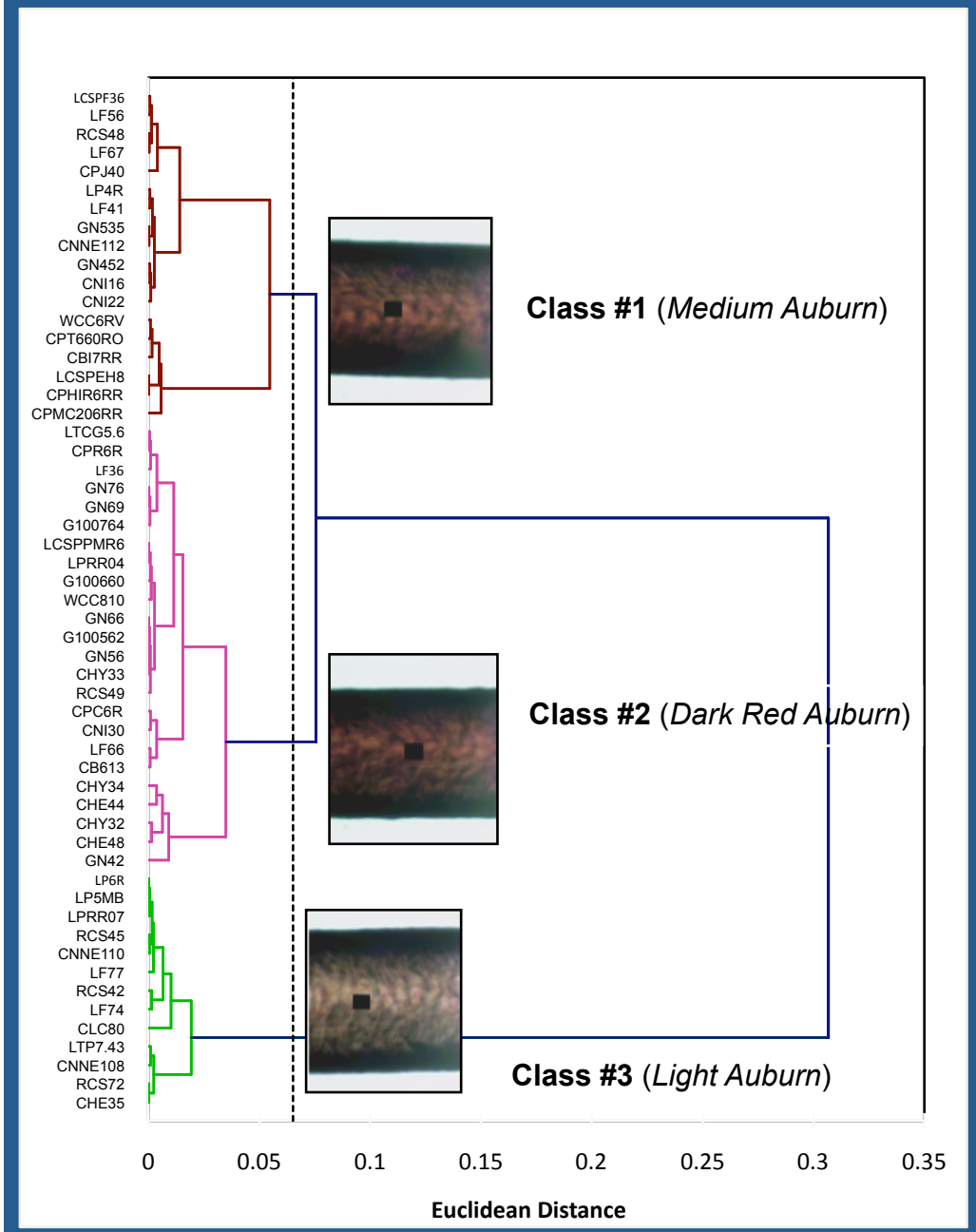
Part I: Analysis of Dyed Hair

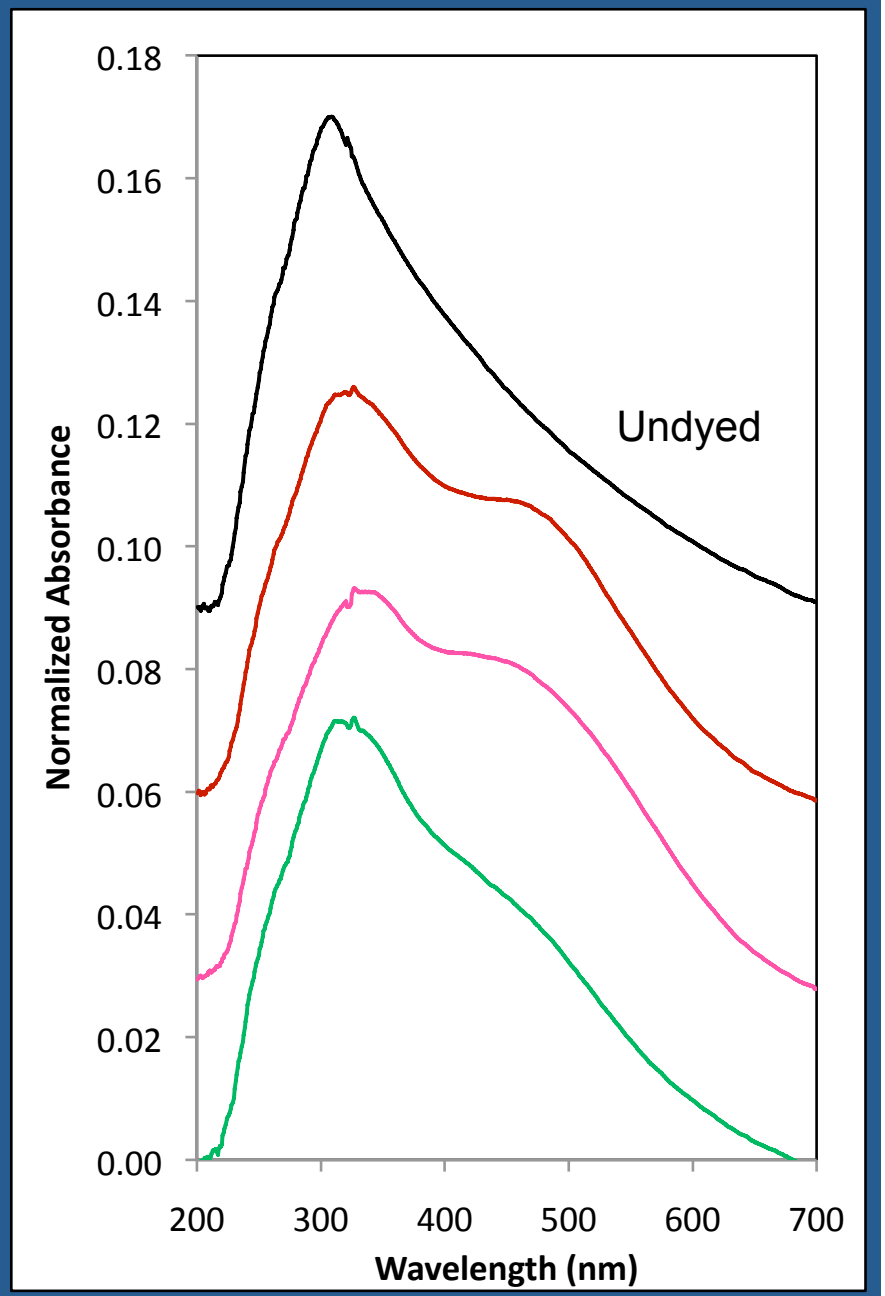
- Can spectroscopic analysis yield forensically useful information about hair dyes?
- Can you identify the manufacturer of a hair dye based on its absorbance characteristics?



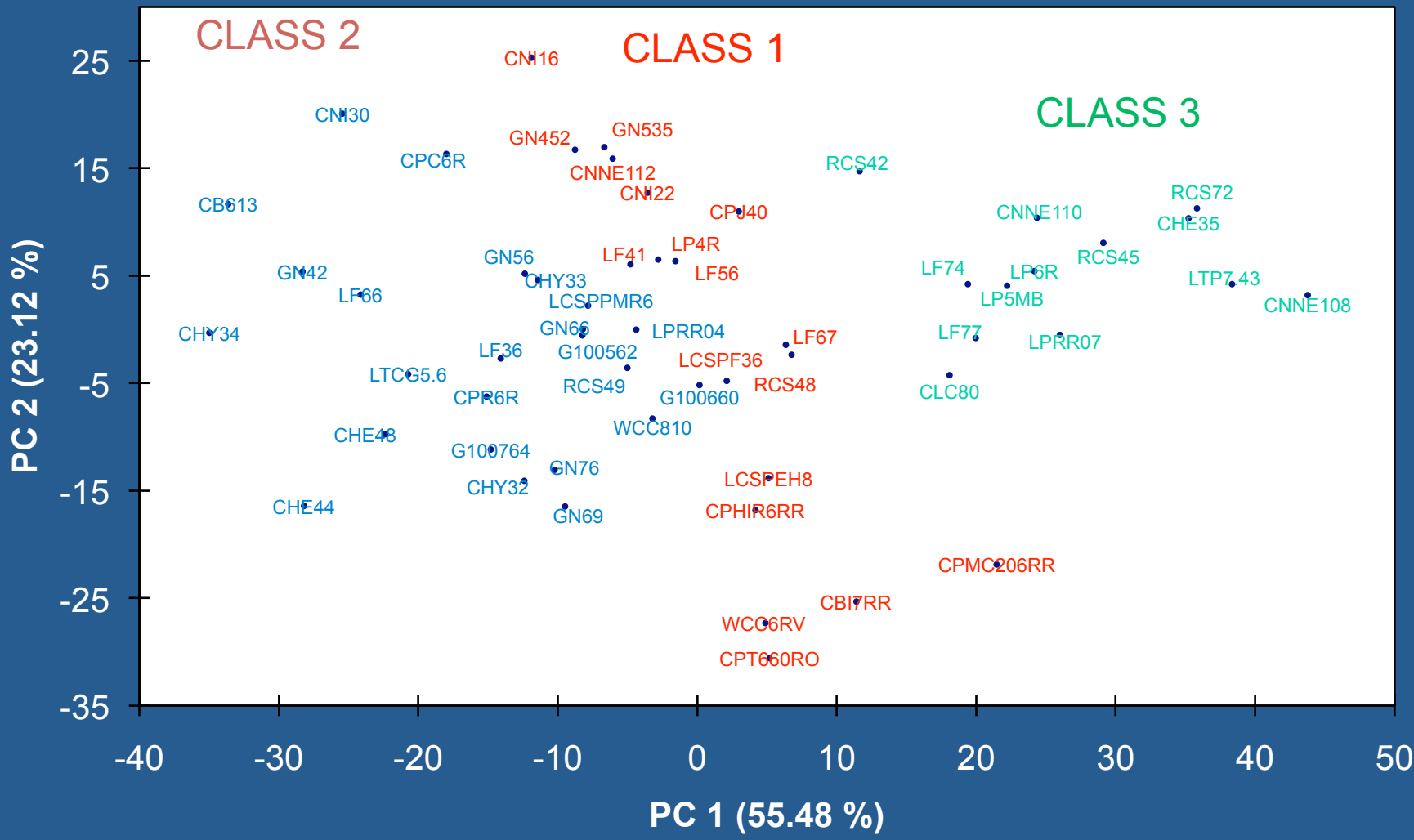
UV-visible Spectrum of Dyed Hair

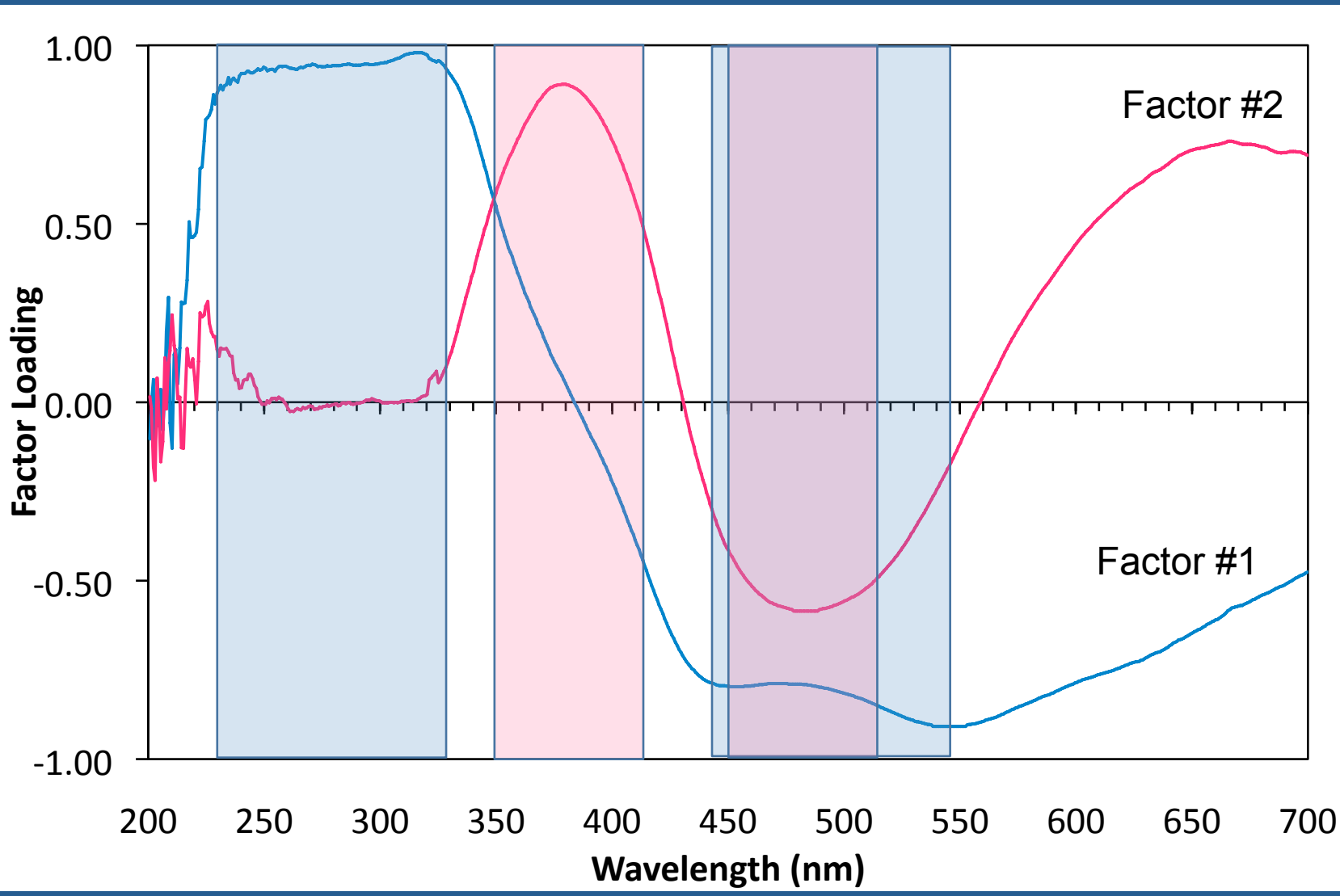



LCSPF36
LF56
RCS48
LF67
CPJ40
LP4R
LF41
GN535
CNNE112
GN452
CNI16
CNI22
WCC6RV
CPT660RO
CBI7RR
LCSPEH8
CPHIR6RR
CPMC206RR
LTCG5.6
CPR6R
LF36
GN76
GN69
G100764
LCSPPMR6
LPRR04
G100660
WCC810
GN66
G100562
GN56
CHY33
RCS49
CPC6R
CNI30
LF66
CB613
CHY34
CHE44
CHY32
CHE48
GN42
LP6R
LP5MB
LPRR07
RCS45
CNNE110
LF77
RCS42
LF74
CLC80
LTP7.43
CNNE108
RCS72
CHE35
Class #1 (Medium Auburn)
Class #2 (Dark Red Auburn)
Class #3 (Light Auburn)
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Euclidean Distance

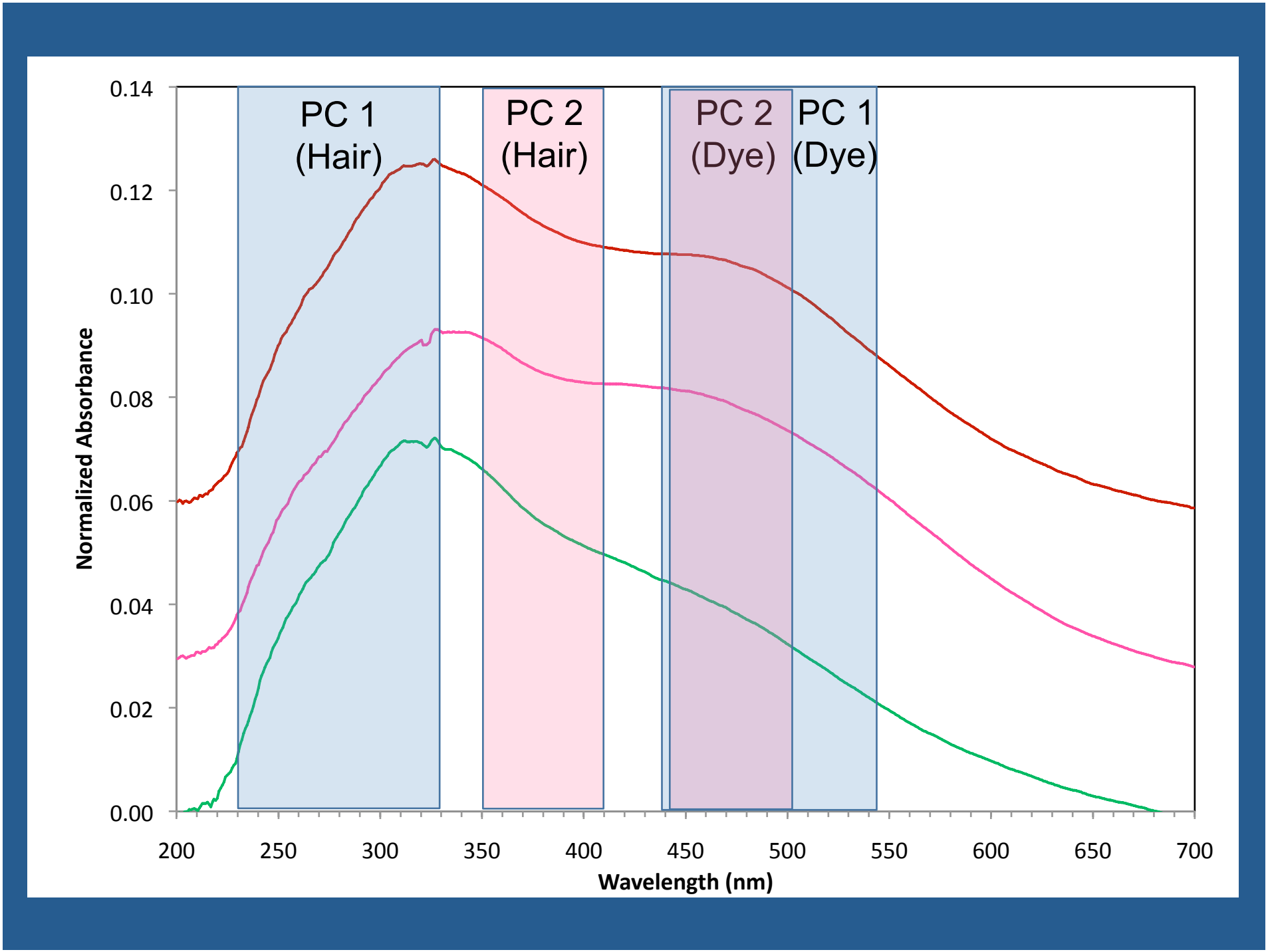

0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Normalized Absorbance
Undyed
200
300
400
500
600
700
Wavelength (nm)

PCA Observations Plot

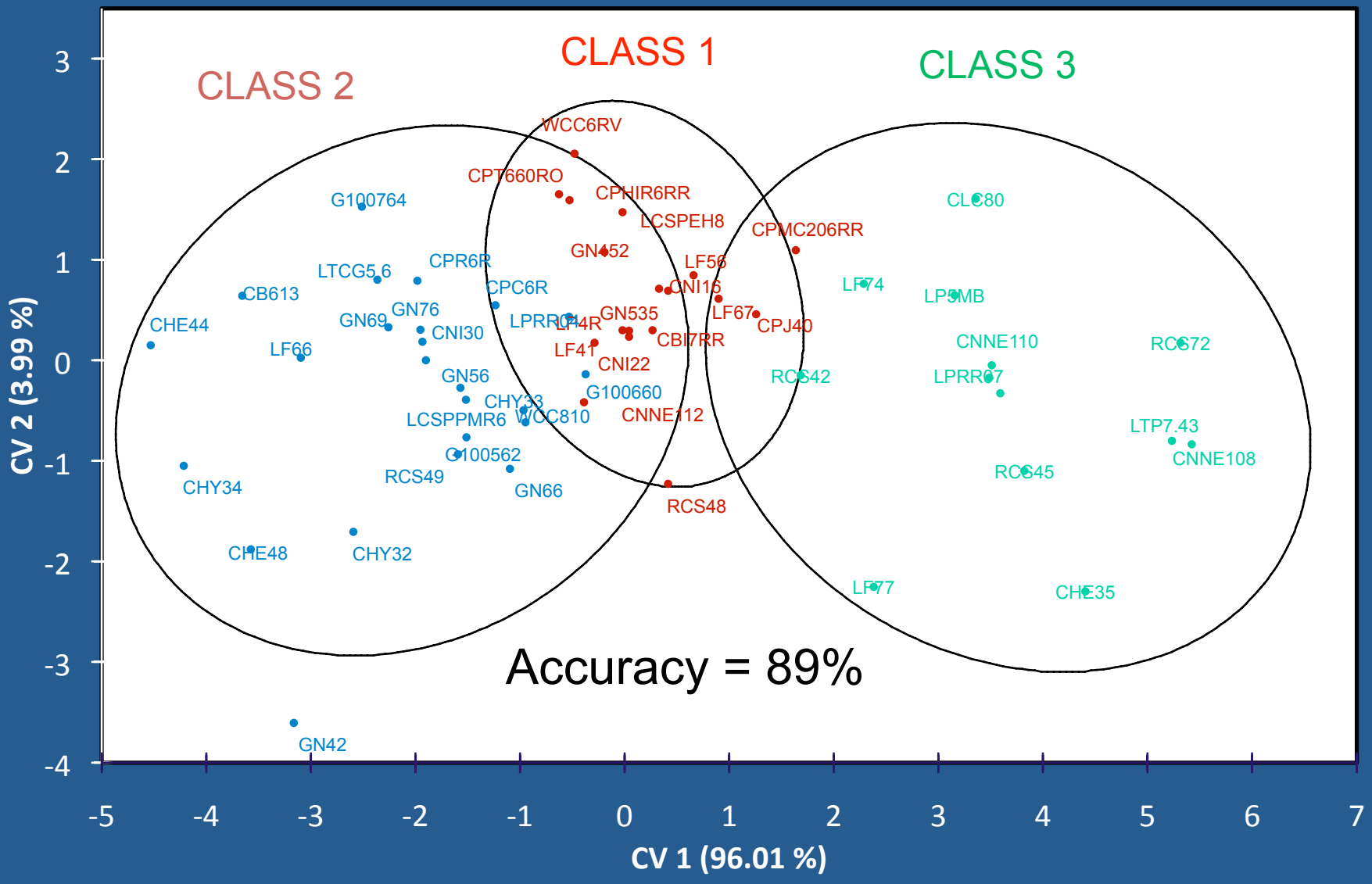




- Positive loadings indicate a positive correlation between the variable and the factor
- Negative loadings indicate a negative correlation between the variable and the factor
- Loadings around zero indicate no correlation


PC 1
(Hair)
PC 2
(Hair)
PC 2
(Dye)
PC 1
(Dye)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Normalized Absorbance
200
250
300
350
400
450
500
550
600
650
700
Wavelength (nm)

Discriminant Analysis

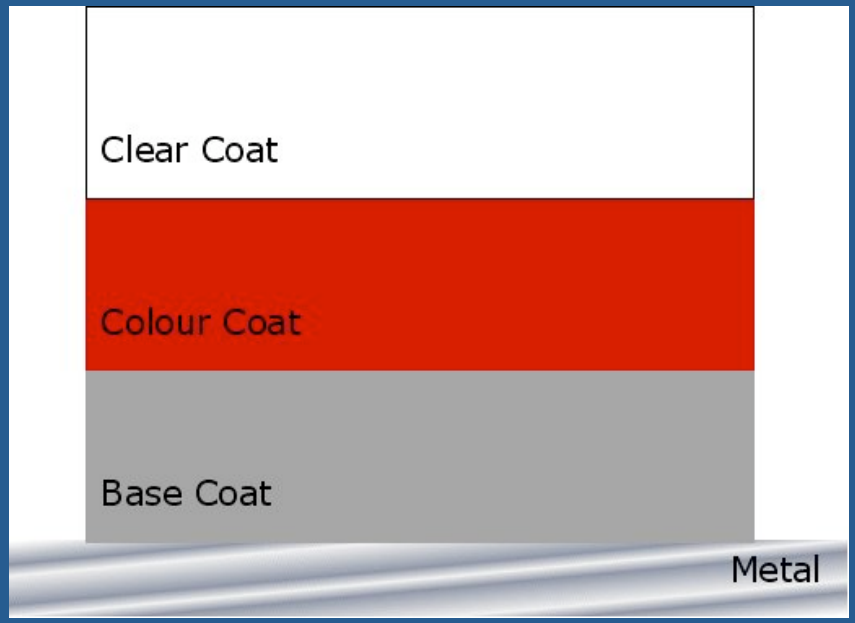


Part I Conclusions

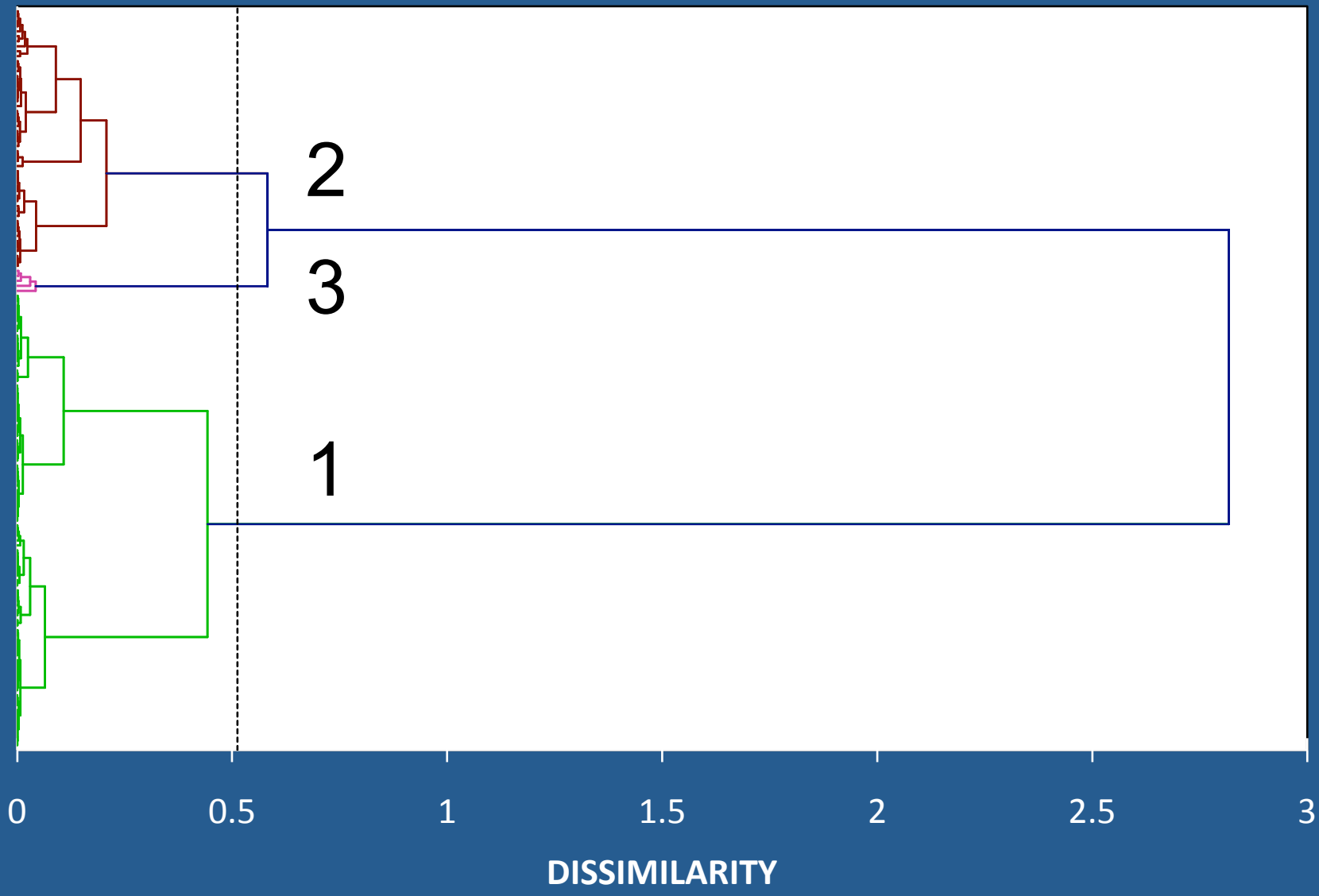
- Red hair dyes were successfully classified into three primary groups
- Important spectral regions include 230 – 330 nm, 350 – 410 nm, and 440 – 540 nm
- Classifying by dye brand or manufacturer was not successful (Accuracy = 56%)
- There was no change in results upon long-term storage, but there is significant fading of dyed hair samples with successive washing

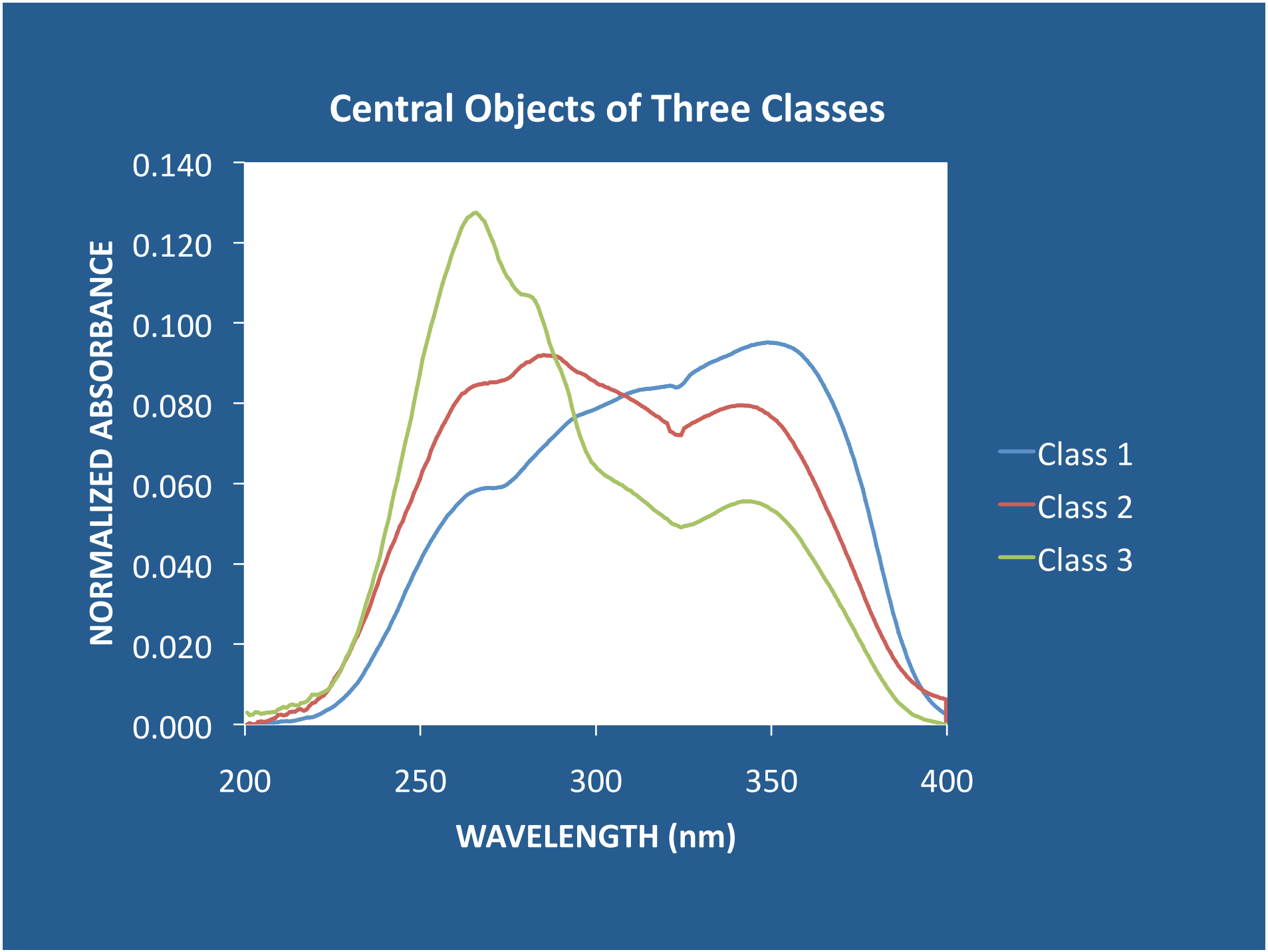
Part II: Automotive Clear Coats

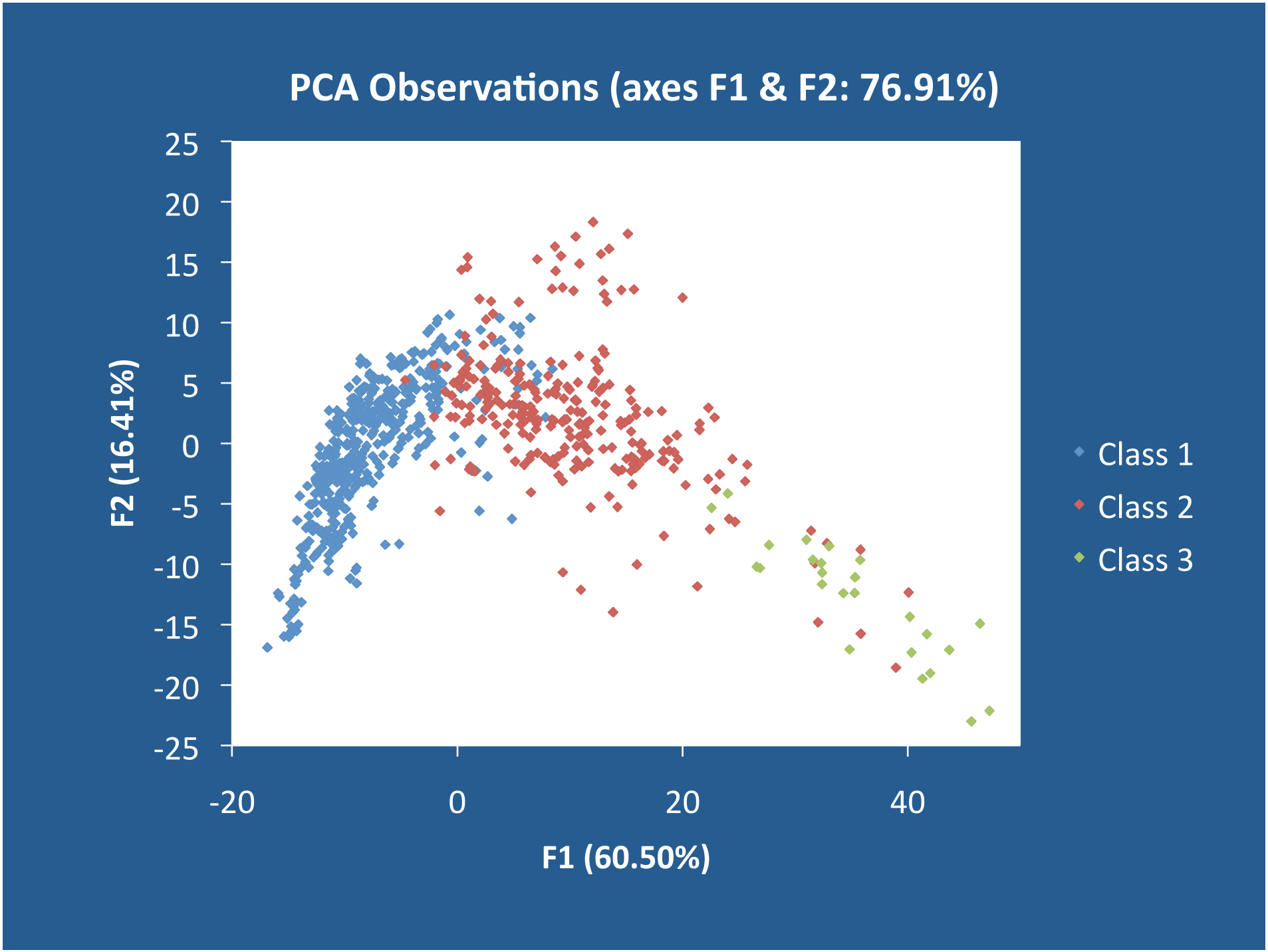
- Top coat of paint that contains no pigmentation or color
- Sprayed over top of certain color coats to serve as a protective film
- Contains 1-2% UV absorbers (absorb light between 290-350 nm)

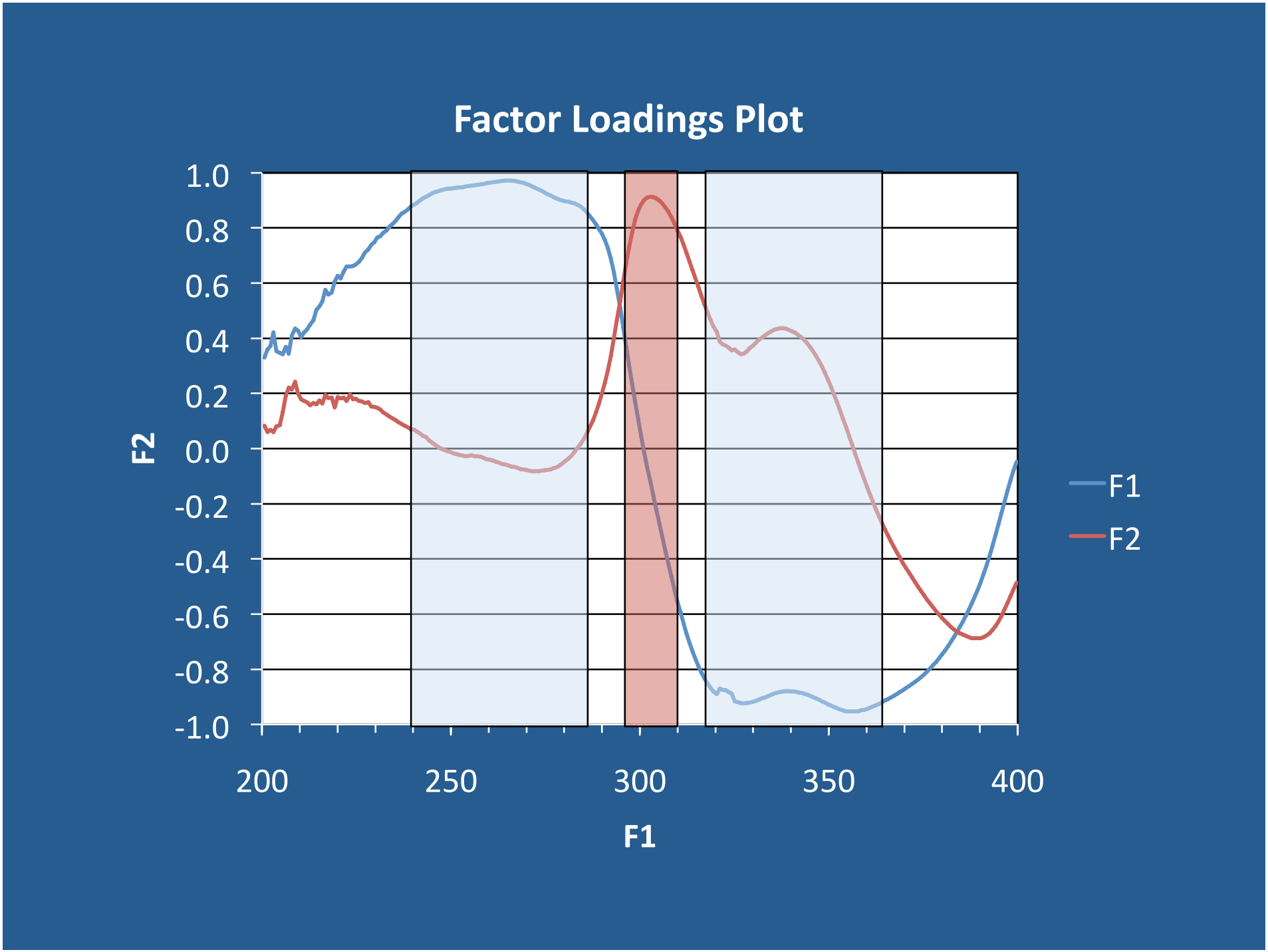


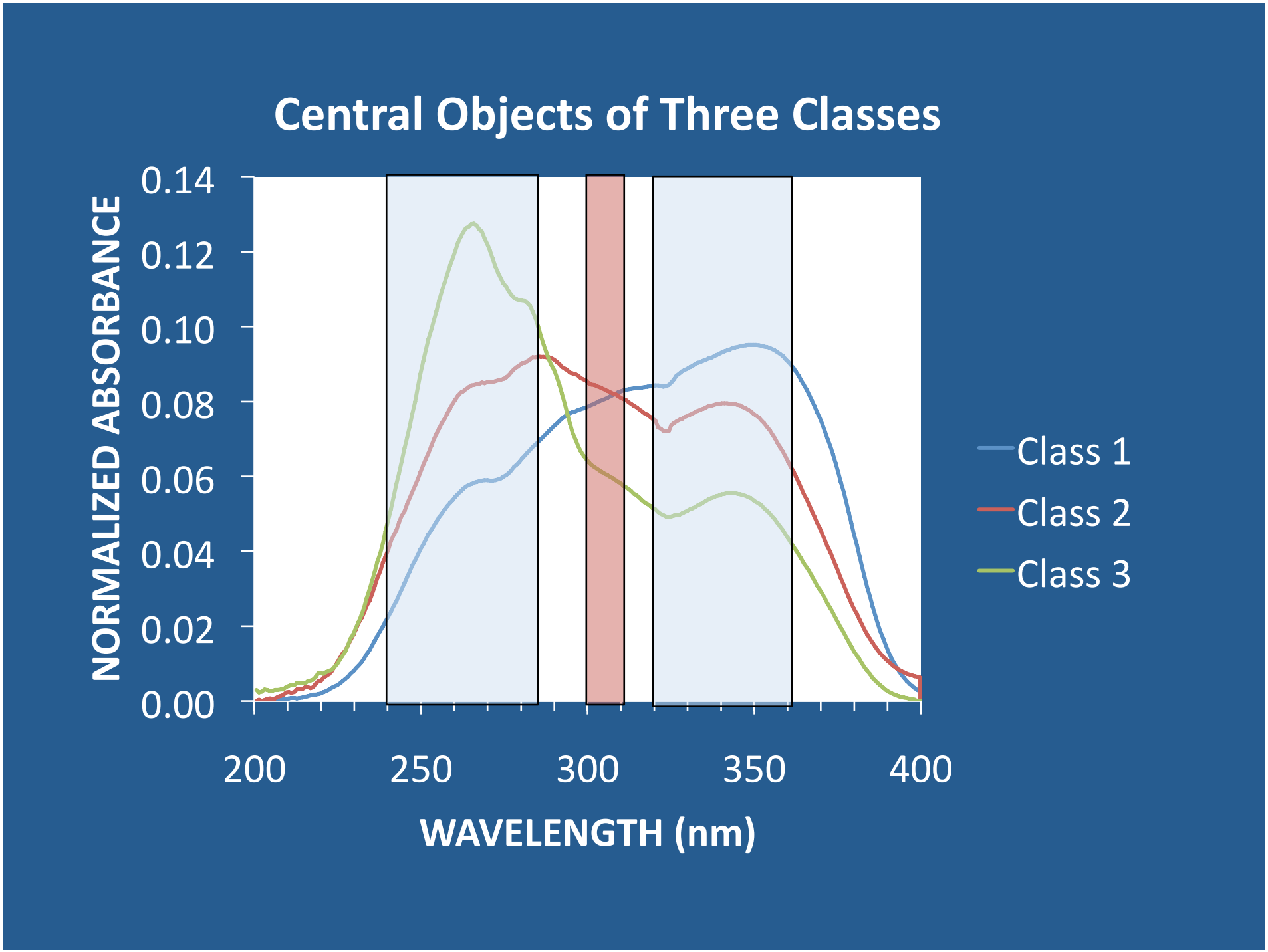
AHC Results

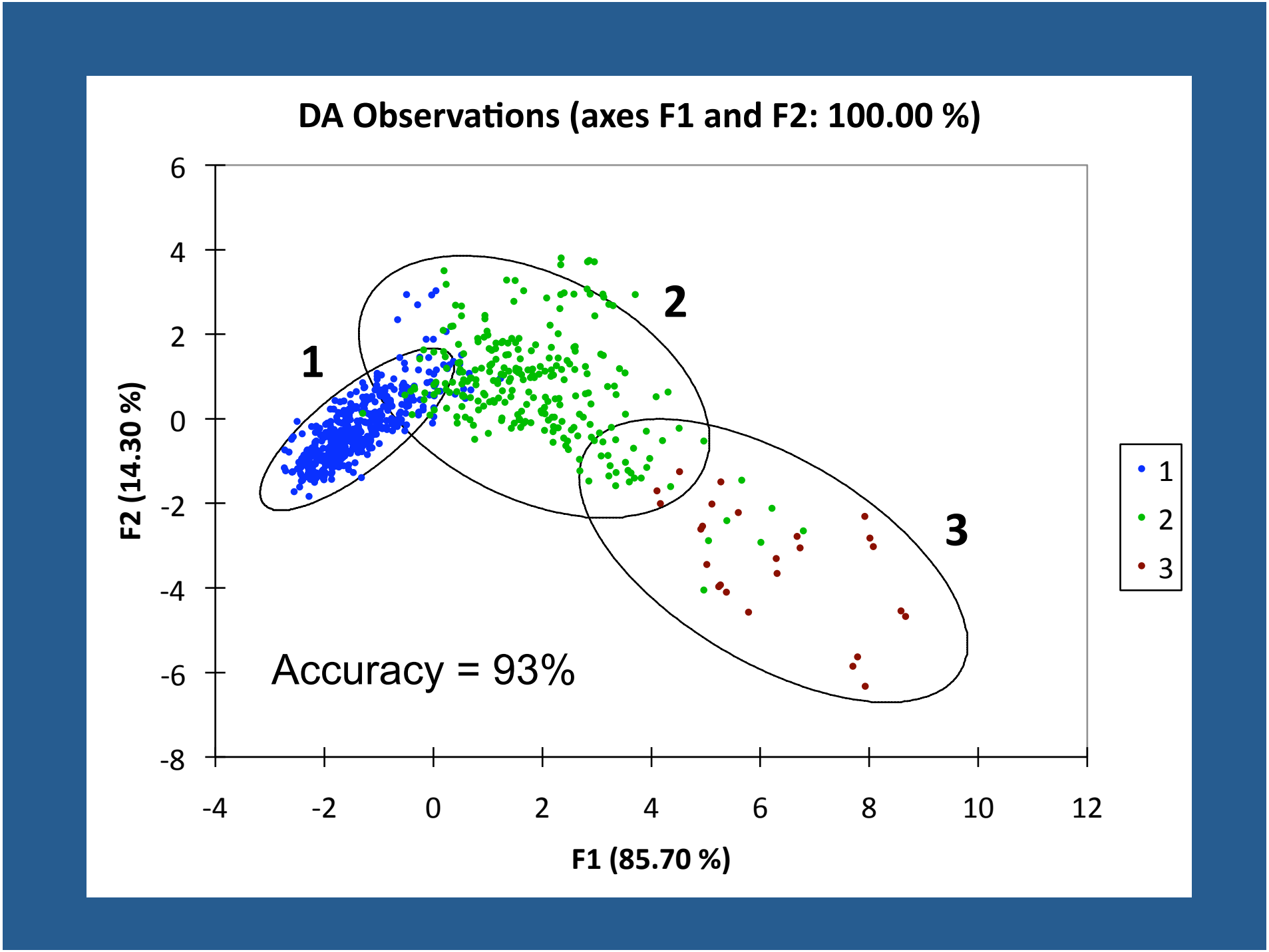



Central Objects of Three Classes
NORMALIZED ABSORBANCE
0.140
0.120
0.100
0.080
0.060
0.040
0.020
0.000
200
250
300
350
400
WAVELENGTH (nm)
Class 1
Class 2
Class 3


PCA Observations (axes F1 & F2: 76.91%)
F2 (16.41%)
25
20
15
10
5
0
-5
-10
-15
-20
-25
-20
0
20
40
F1 (60.50%)
Class 1
Class 2
Class 3


Factor Loadings Plot
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
F2
200
250
300
350
400
F1
F1
F2


Central Objects of Three Classes
NORMALIZED ABSORBANCE
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
200
250
300
350
400
WAVELENGTH (nm)
Class 1
Class 2
Class 3


DA Observations (axes F1 and F2: 100.00 %)
F2 (14.30 %)
F1 (85.70 %)
6
4
2
0
-2
-4
-6
-8
-4
-2
0
2
4
6
8
10
12
1
2
3
Accuracy = 93%
1
2
3

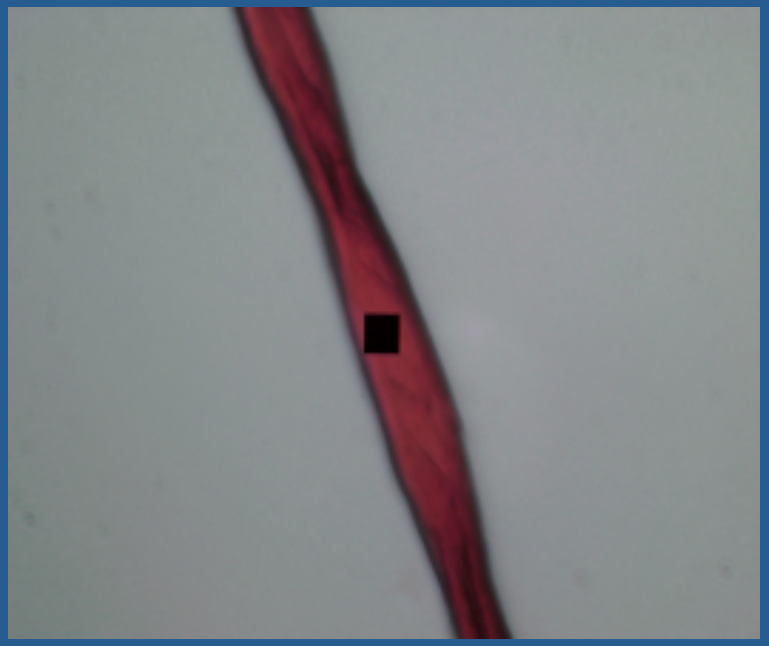
Confusion Matrix for the Cross-Validation Results:					
From/ To	***1***	***2***	***3***	***Total***	***% Correct***
1	**434**	**21**	0	455	95.38
2	**22**	**230**	**8**	260	88.46
3	0	**2**	**23**	25	92.00
Total	456	253	31	740	**92.84**

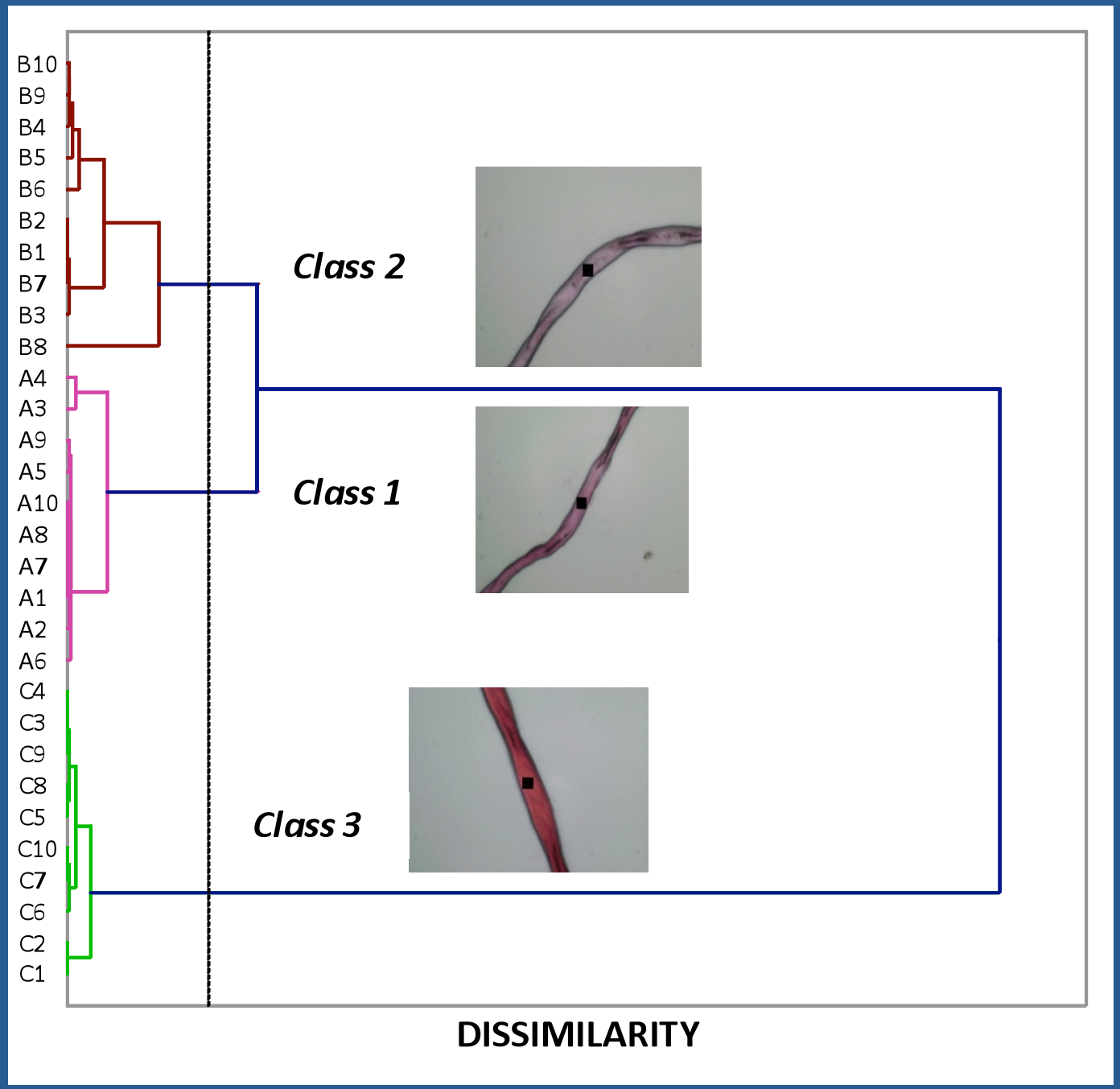
Part II Conclusions

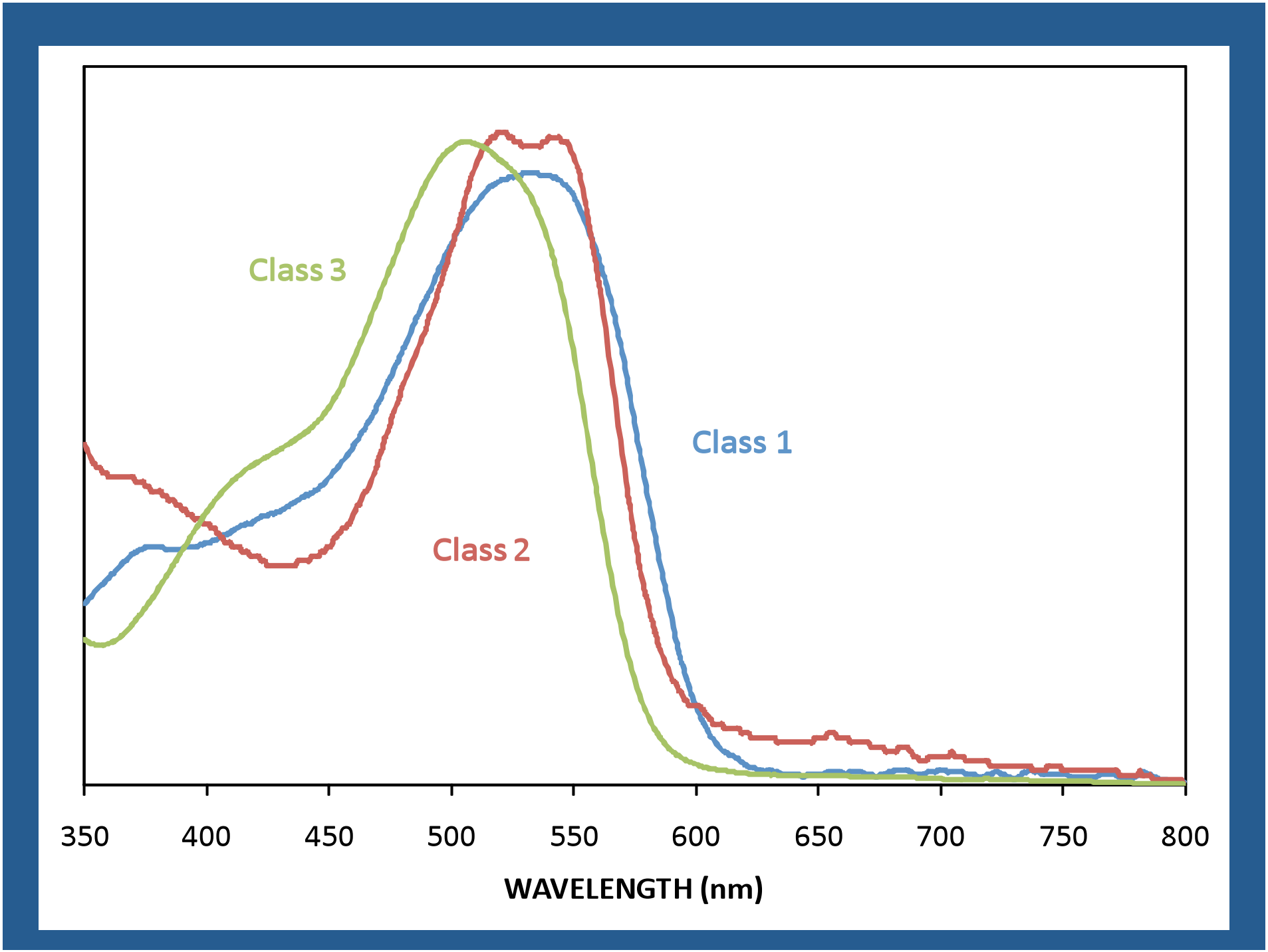
- Automotive clearcoats are reliably organized into three groups
- Important spectral regions include 240 – 280 nm, 300 – 310 nm and 320 – 360 nm
- Clear trends in the organization of the data by make, model and year were not observed

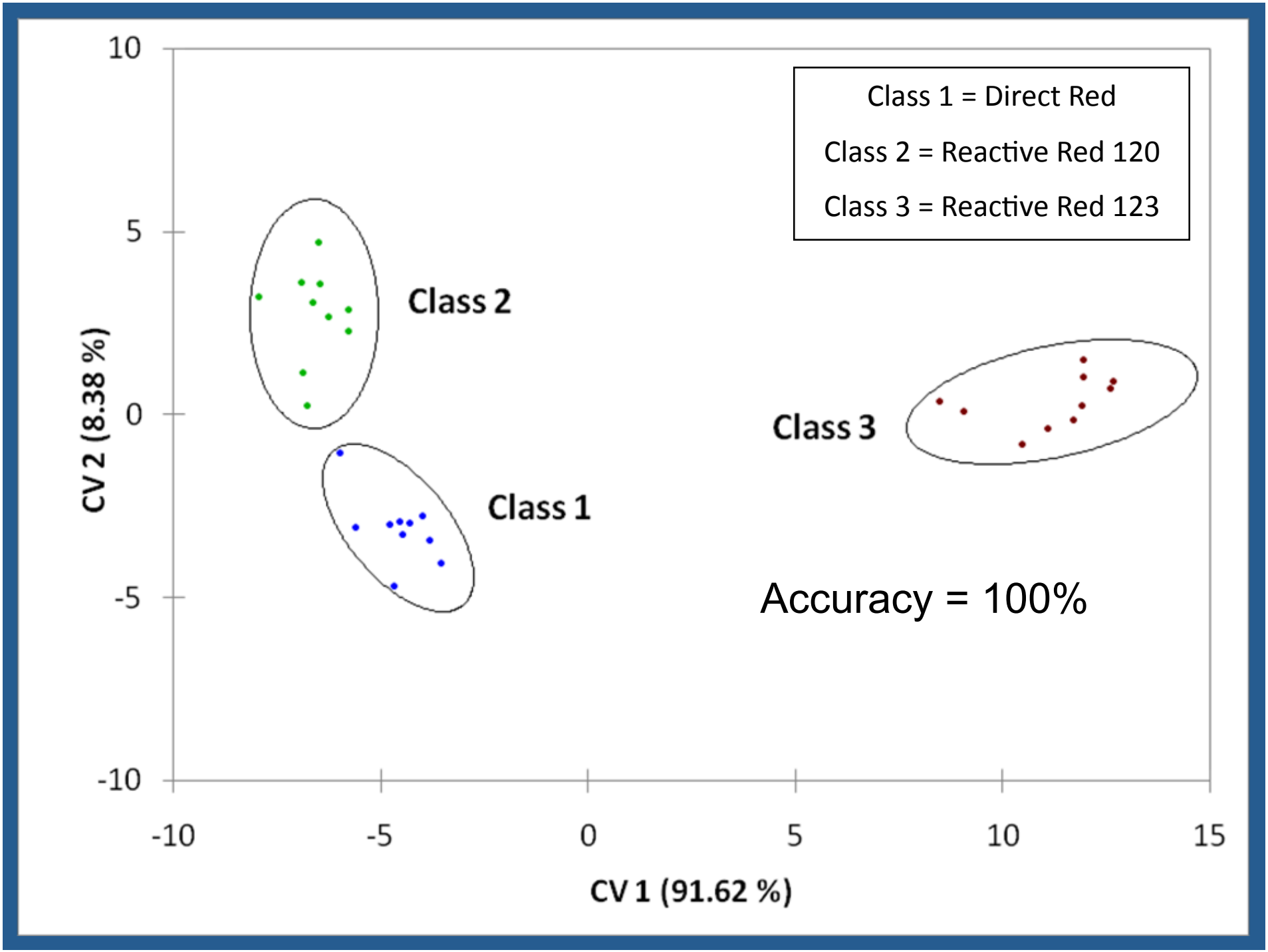
Part III: Analysis of Dyed Cotton Fibers

- The color of cotton fibers is considered to be the source of greatest variation.
- Association can be problematic due to sample heterogeneity and a lack of quantitative criteria for comparing spectra




B10
B9
B4
B5
B6
B2
B1
B7
B3
B8
A4
A3
A9
A5
A10
A8
A7
A1
A2
A6
C4
C3
C9
C8
C5
C10
C7
C6
C2
C1
Class 2
Class 1
Class 3
DISSIMILARITY


Class 3
Class 1
Class 2
350
400
450
500
550
600
650
700
750
800
WAVELENGTH (nm)


Class 1 = Direct Red
Class 2 = Reactive Red 120
Class 3 = Reactive Red 123
Class 2
Class 1
Class 3
Accuracy = 100%
CV 2 (8.38 %)
CV 1 (91.62 %)
10
5
0
-5
-10
-10
-5
0
5
10
15

Part III Conclusions

- MSP can reliably distinguish cotton fibers dyed with Direct Red, Reactive Red 120 and Reactive Red 123
- But – is this a representative sample of red cotton fibers?

Acknowledgements

- IUPUI University Fellowship (Elisa Liszewski)
- Indiana State Police Microanalysis Unit
- Aveda Fredric's Institute and Honors Beauty College
- Forensic Sciences Foundation Jan S. Bashinski Criminalistics Graduate Thesis Grant (Julie Barrett)