Applying Hierarchical Clustering to Broad Absorption Line Profiles for Quasar Classification

Nathalie Thibert

Collaborators:
Dr. Sarah Gallagher
Dr. Mark Daley
Dr. Aycha Tammour

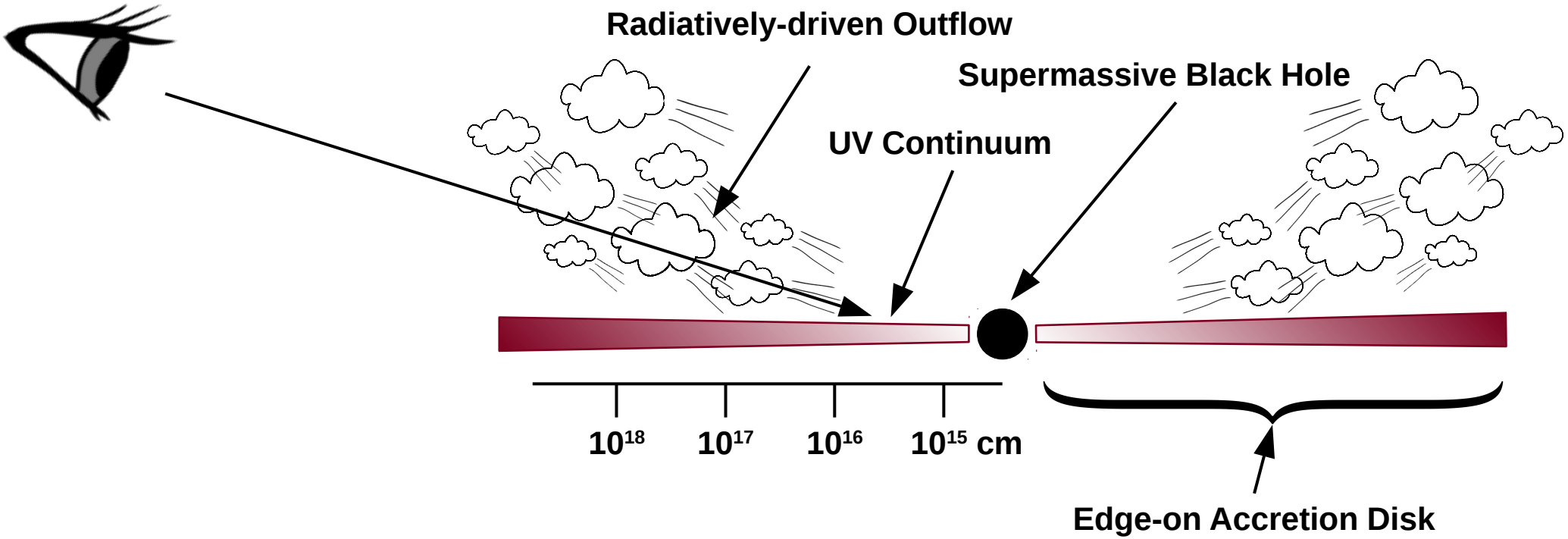
Great Lakes 2016 Quasar Symposium

Western Science

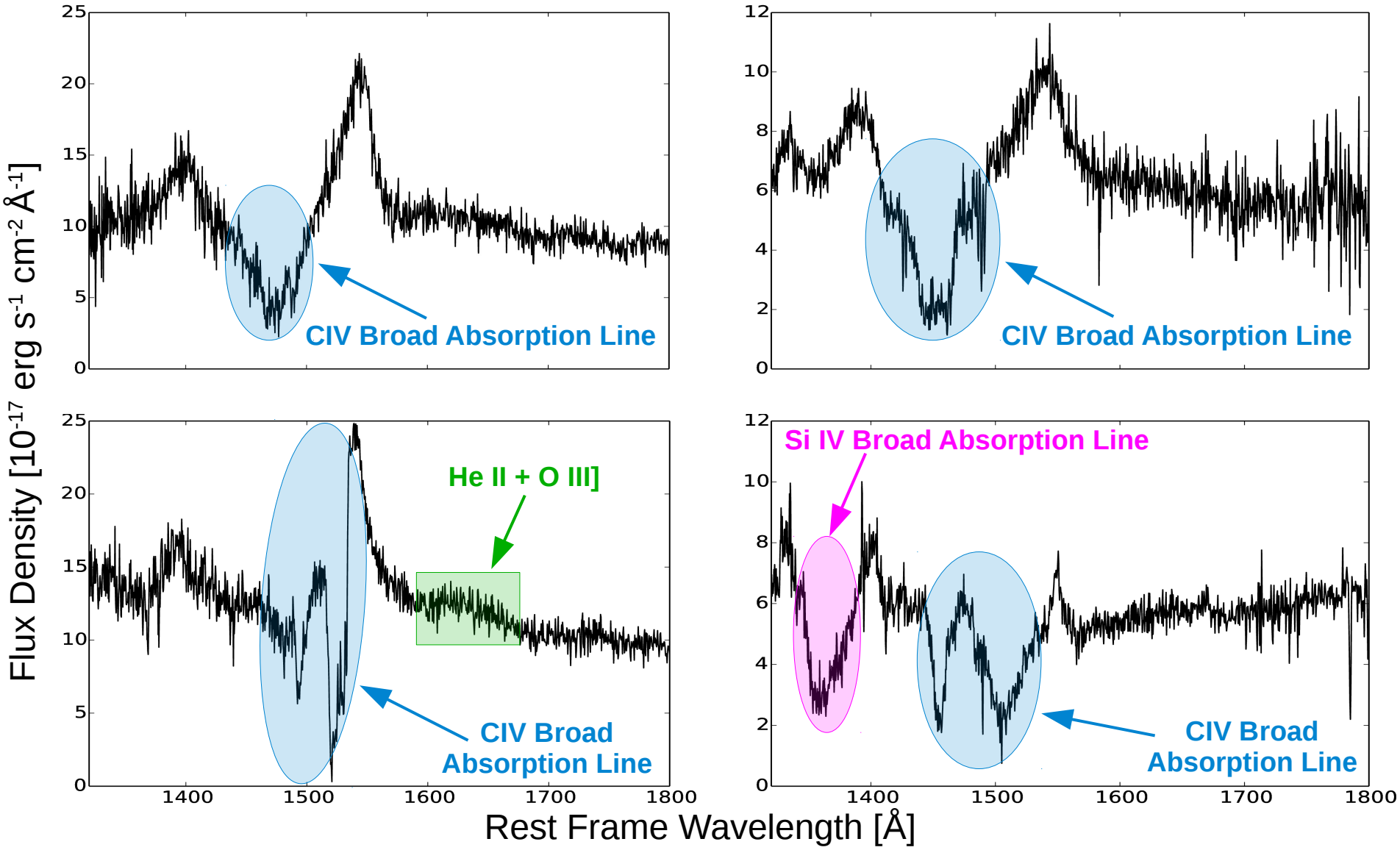
NSERC CRSNG

Image credit: NASA

Broad Absorption Line (BAL) Quasars



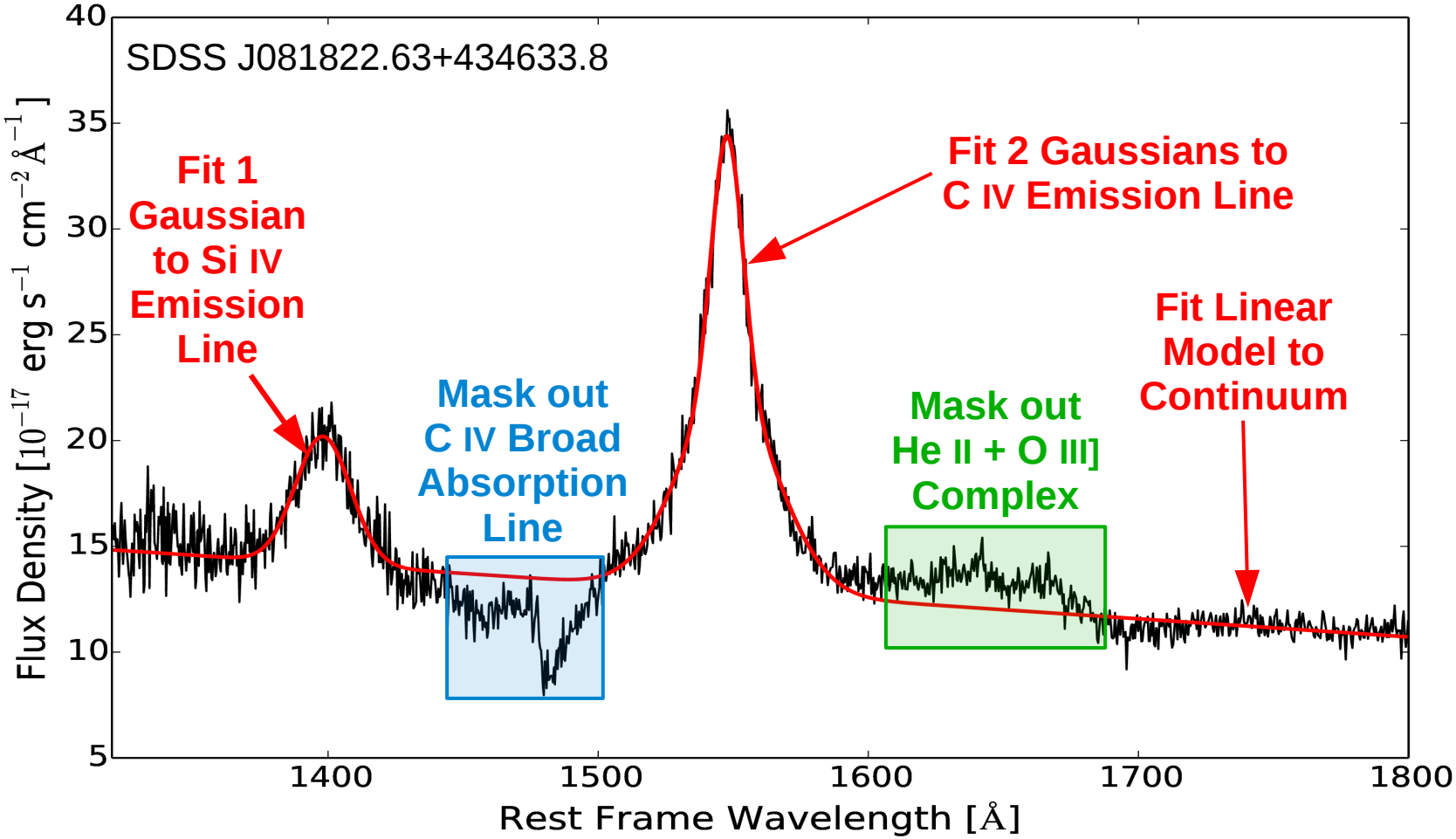
Diversity of C IV BALs



Why are these profiles different?

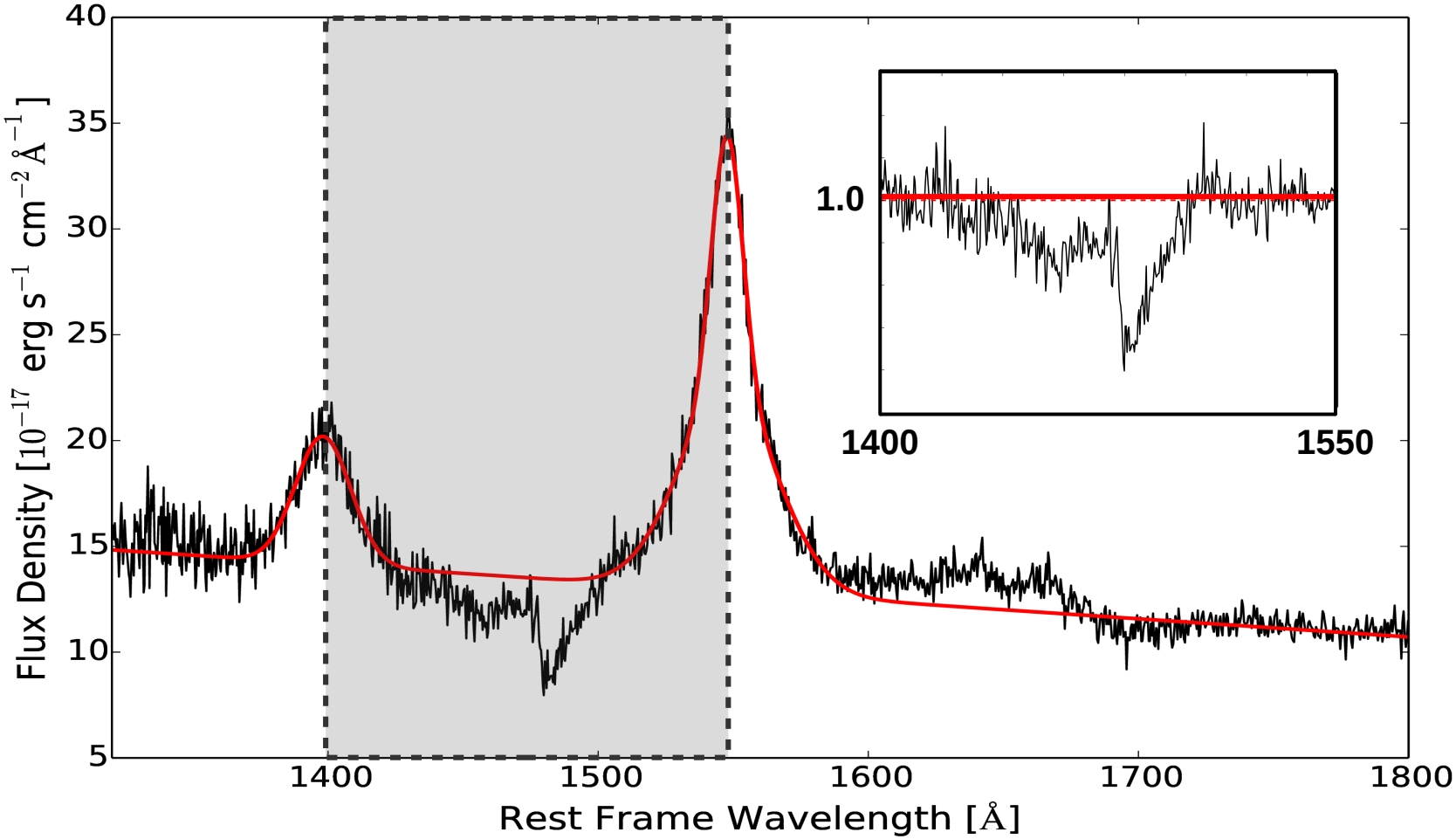
BAL Quasar Catalog & Data Reduction

- **Sample: 1,084 BAL quasars (1.68 ≤ *z* ≤ 4.93) from SDSS DR5 (Gibson et al., 2009)**
- **Method: (1) fit local continuum and prominent, broad emission lines (1320-1800 Å)**
  (2) normalize and "compare" spectra (1400-1550 Å)



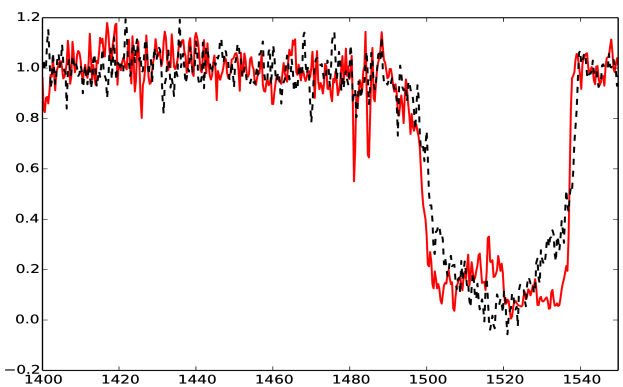
BAL Quasar Catalog & Data Reduction

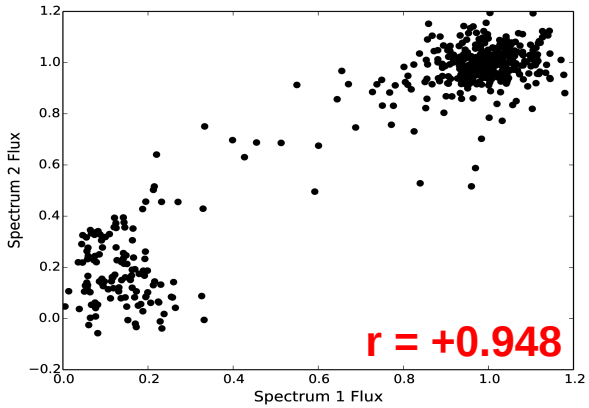
- **Sample: 1,084 BAL quasars (1.68 ≤ *z* ≤ 4.93) from SDSS DR5 (Gibson et al., 2009)**
- **Method: (1) fit local continuum and prominent, broad emission lines (1320-1800 Å)**
  (2) normalize and "compare" spectra (1400-1550 Å)



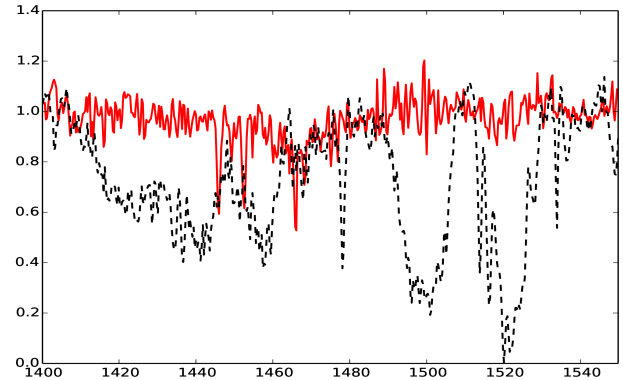
Pearson Product-moment Coefficients

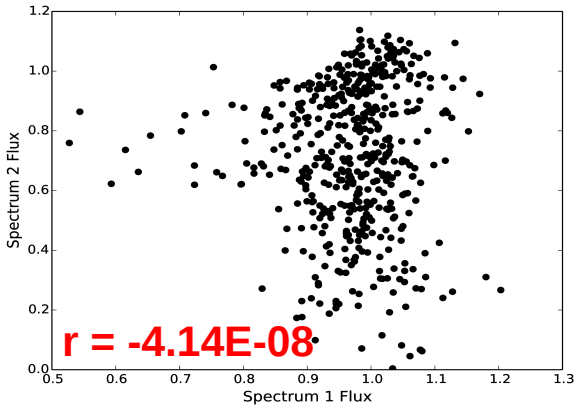
Strong Positive Correlation:



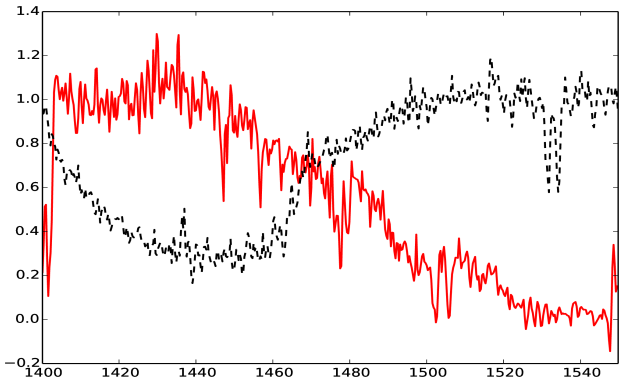


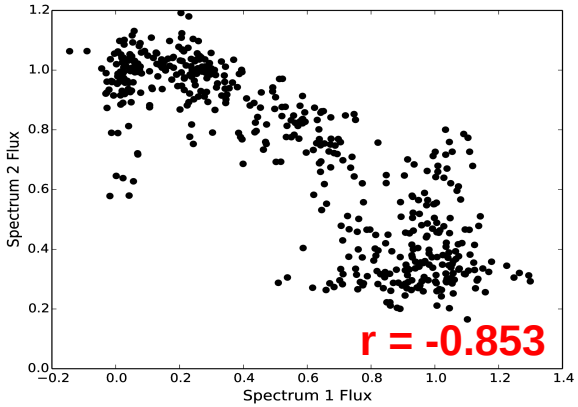
Weak/No Correlation:





Strong Negative Correlation:



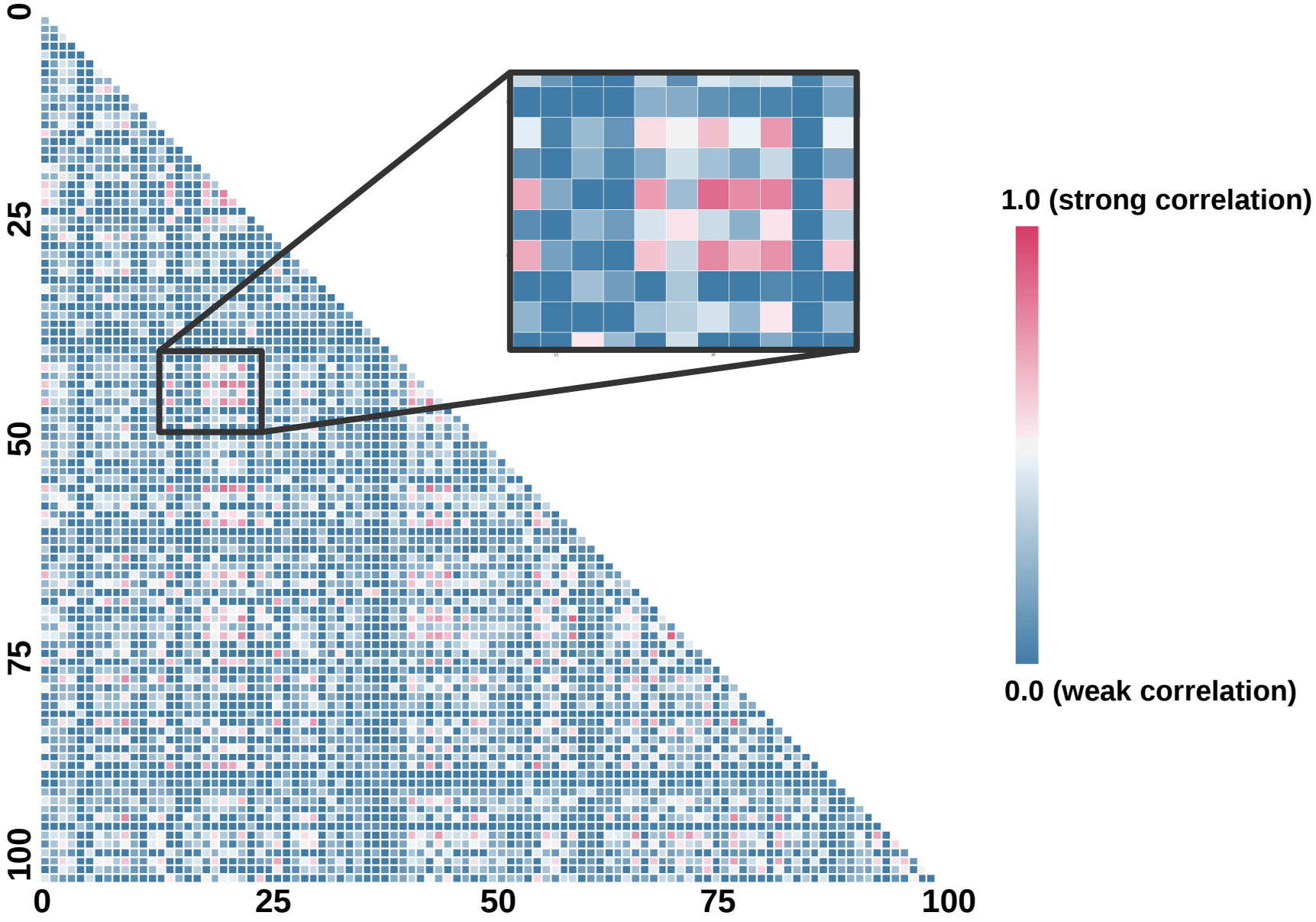


+1.0

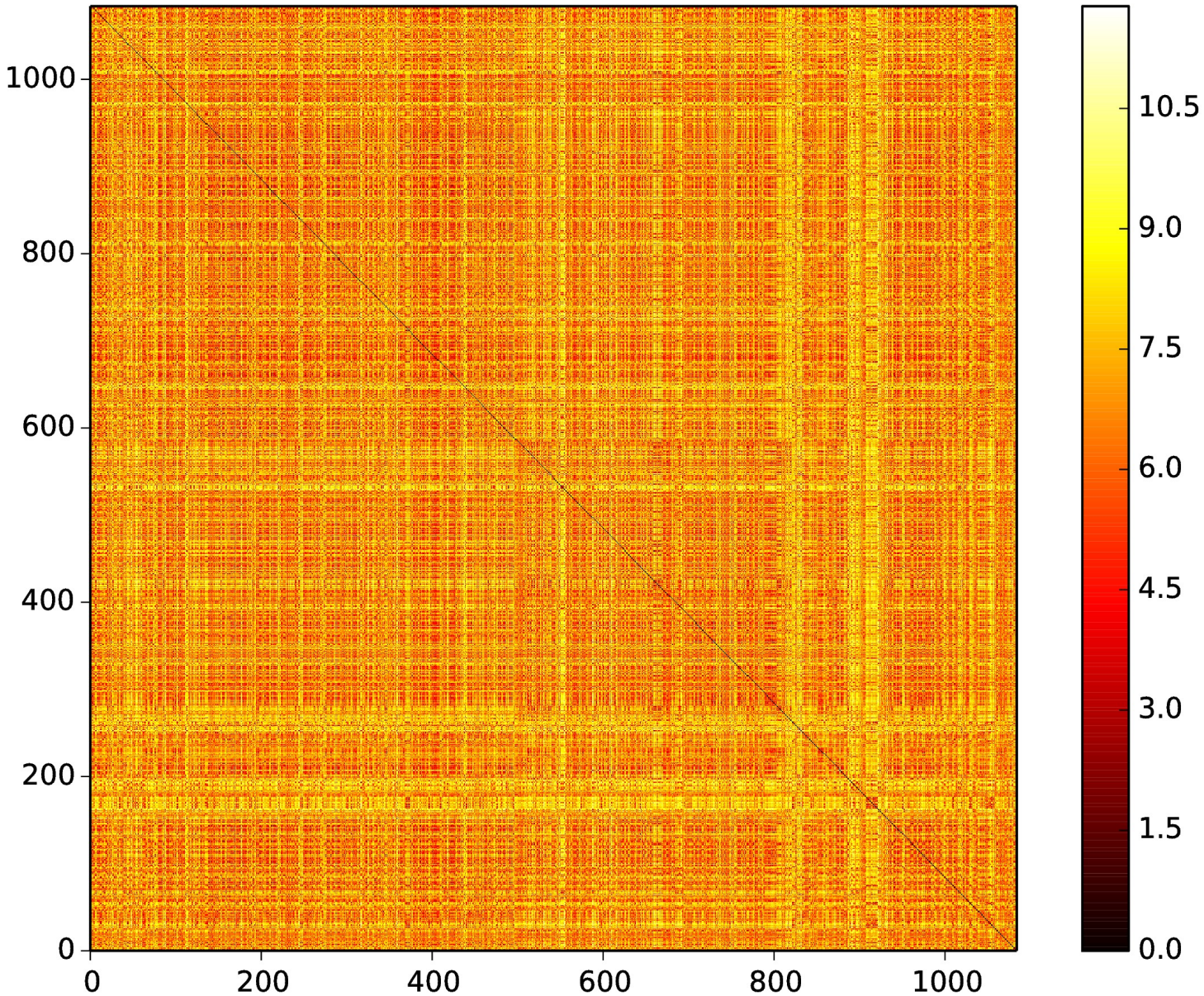
0.0

-1.0

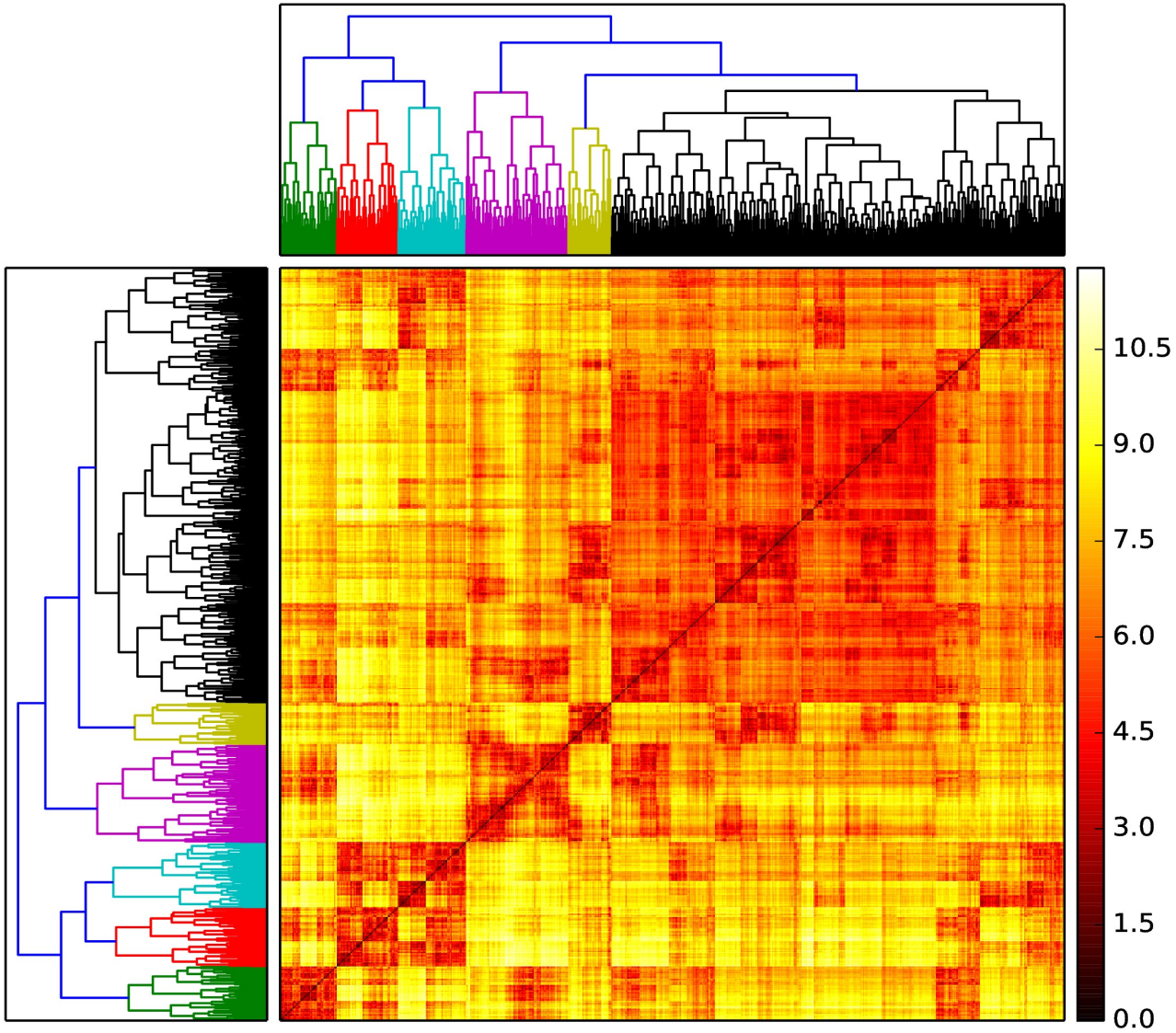
Matrix of Correlation Coefficients



Unclustered Distance Matrix

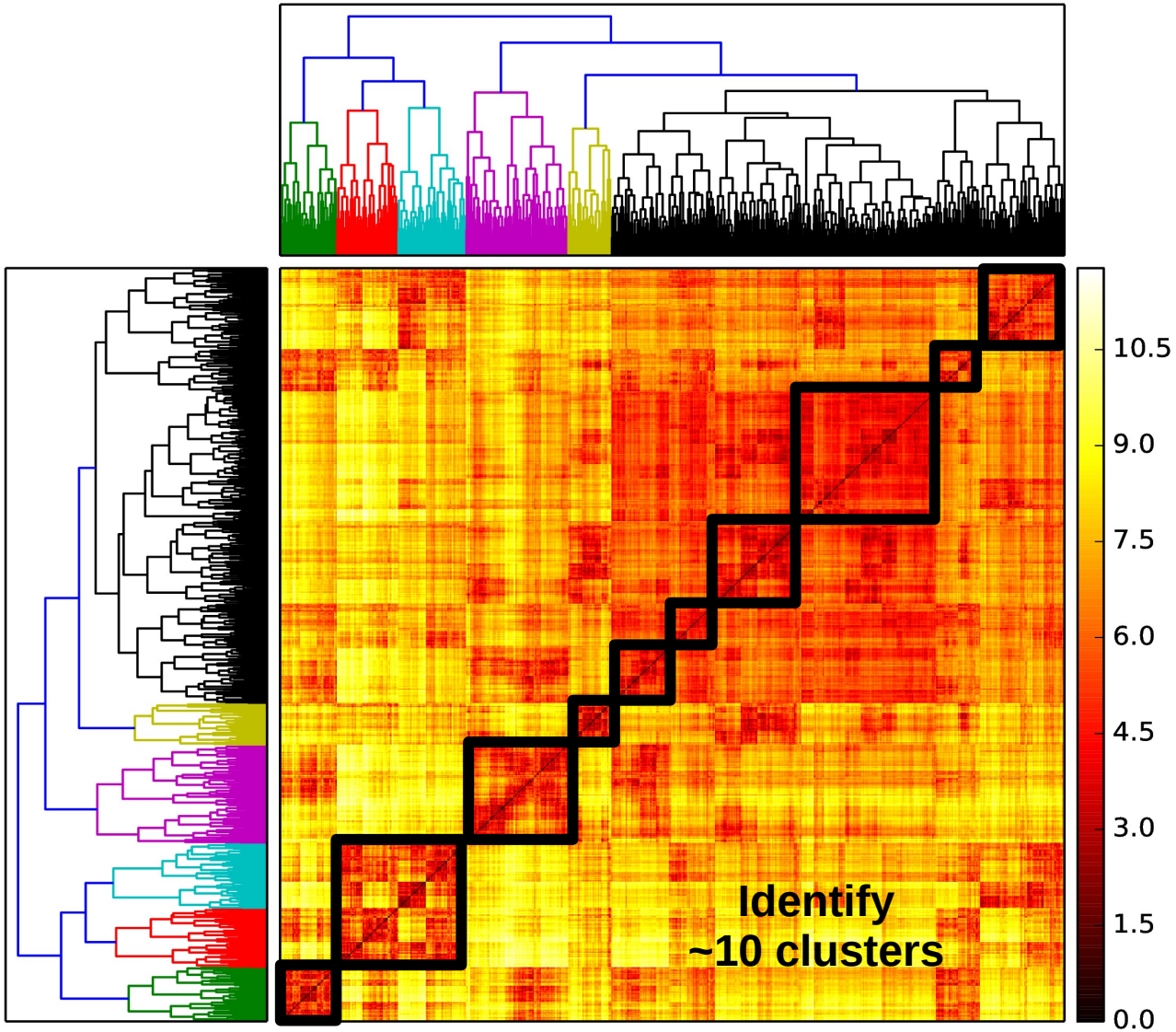


Results from Hierarchical Clustering



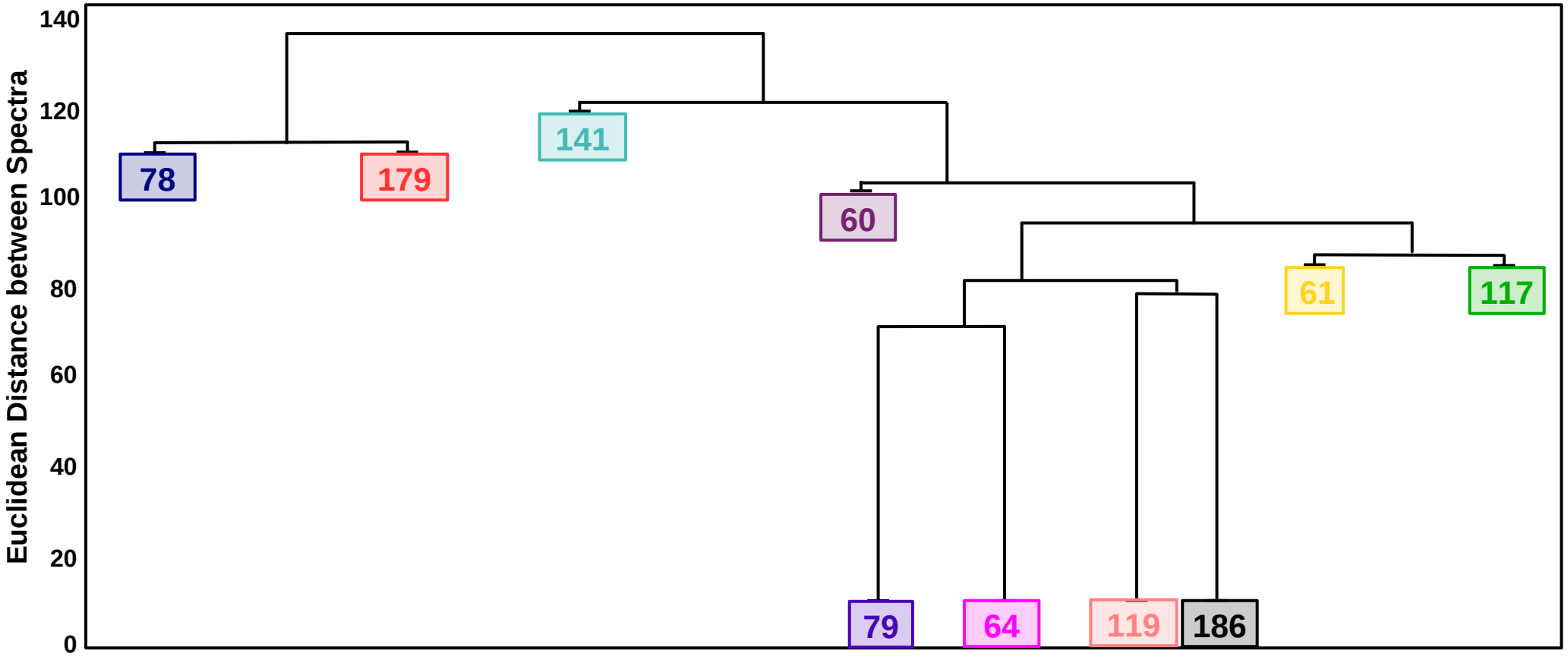
scipy.cluster.hierarchy.linkage
scipy.cluster.hierarchy.dendrogram

Results from Hierarchical Clustering

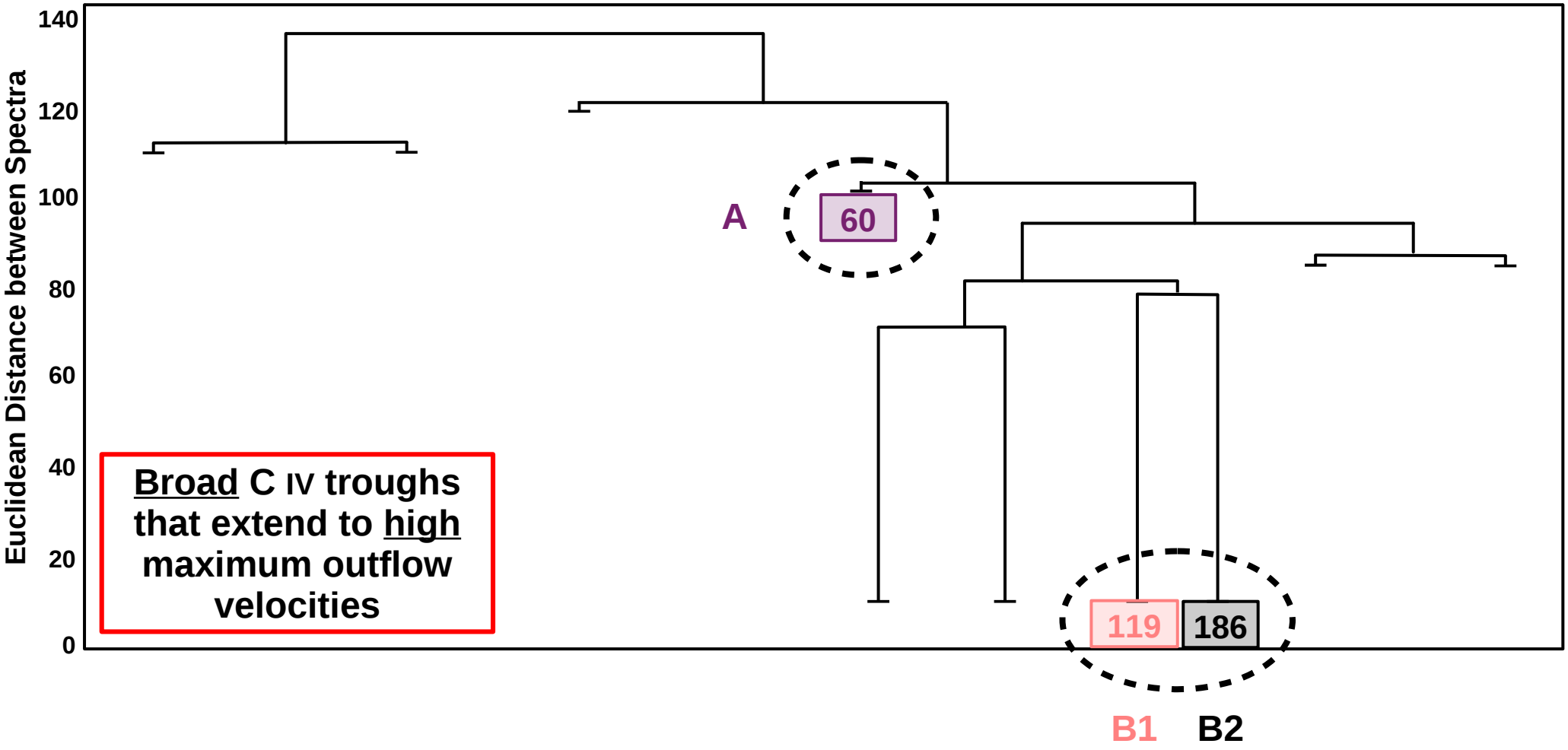


scipy.cluster.hierarchy.linkage
scipy.cluster.hierarchy.dendrogram

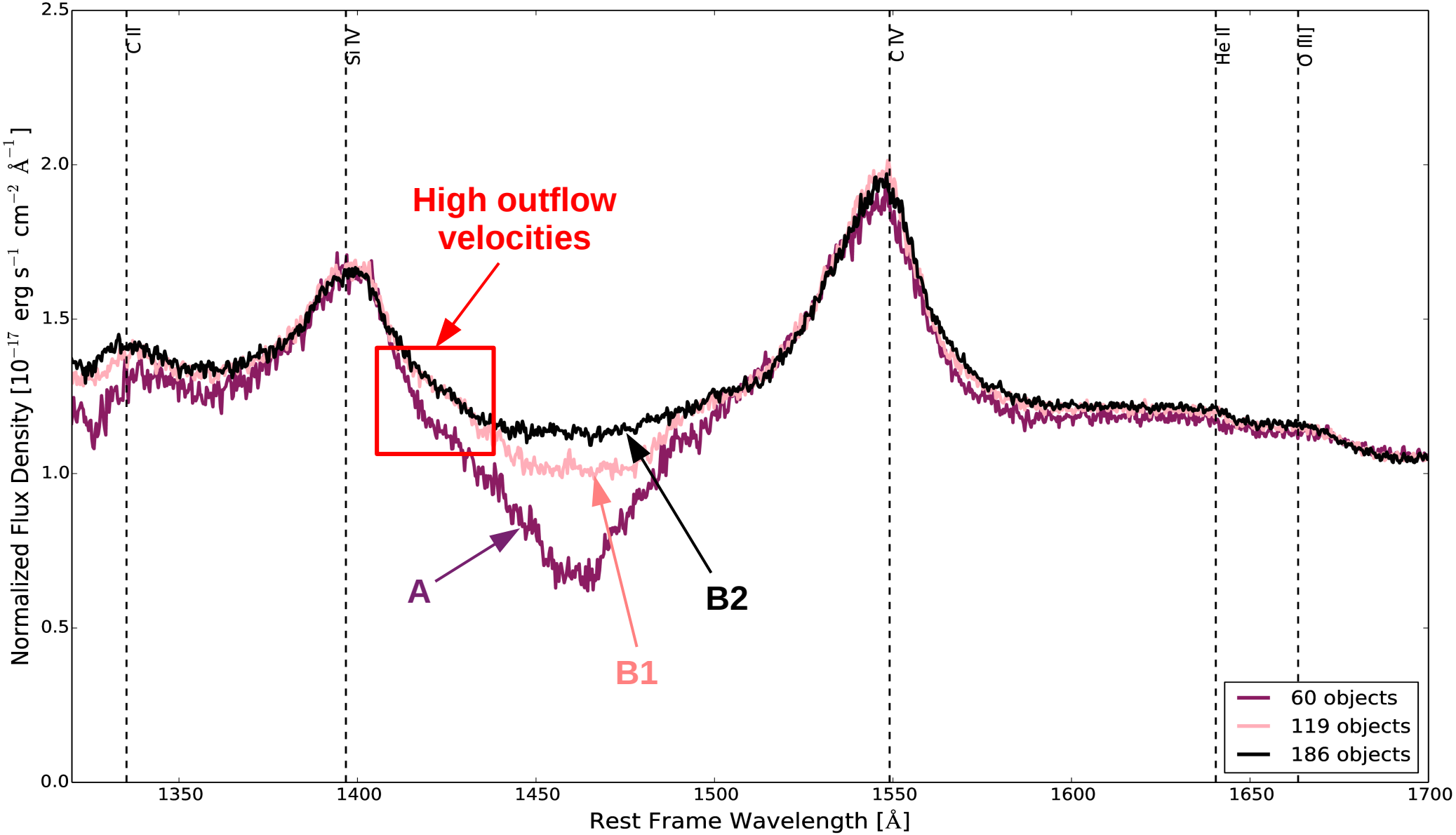
Dendrogram of the Merging Process



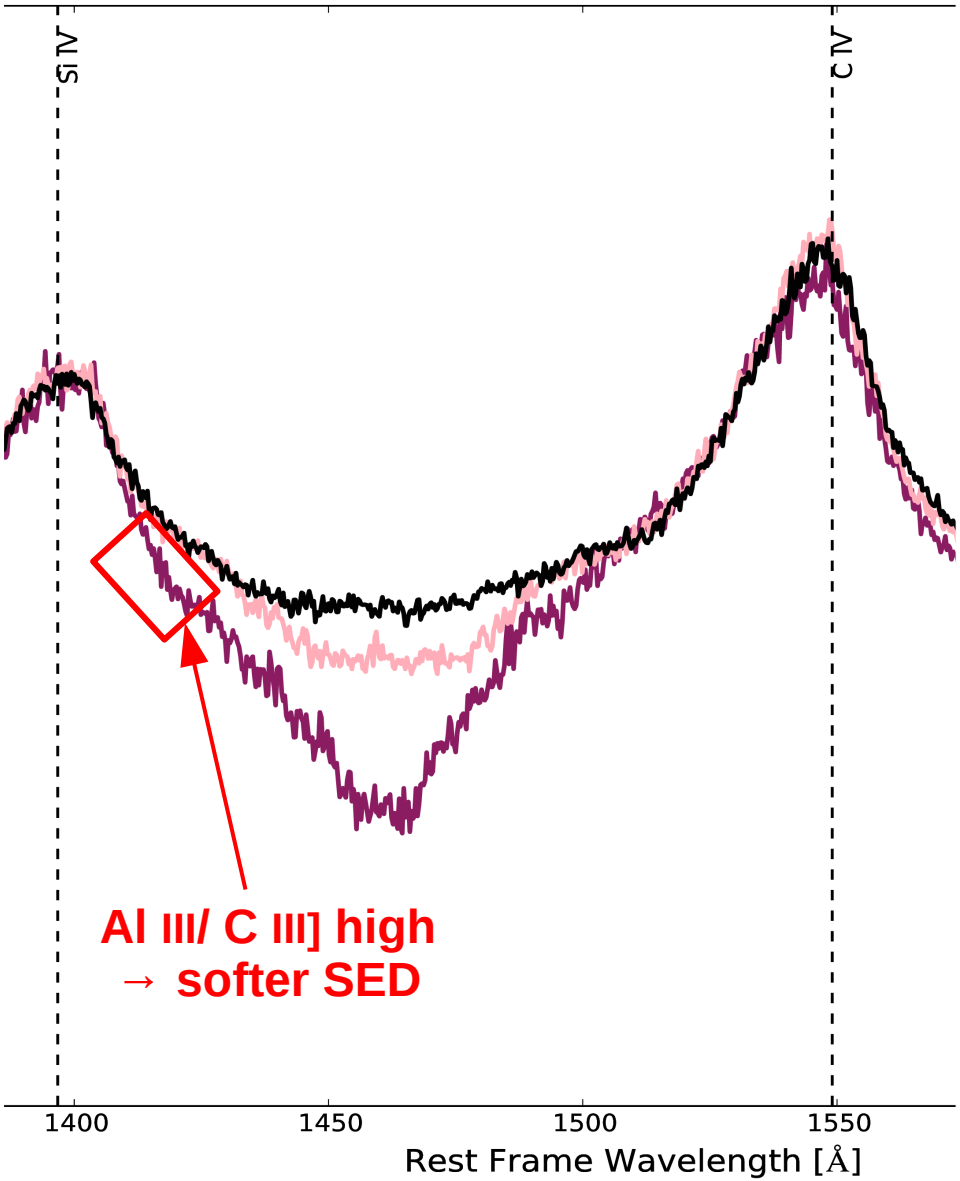
Example: Broad, High-velocity C IV Troughs



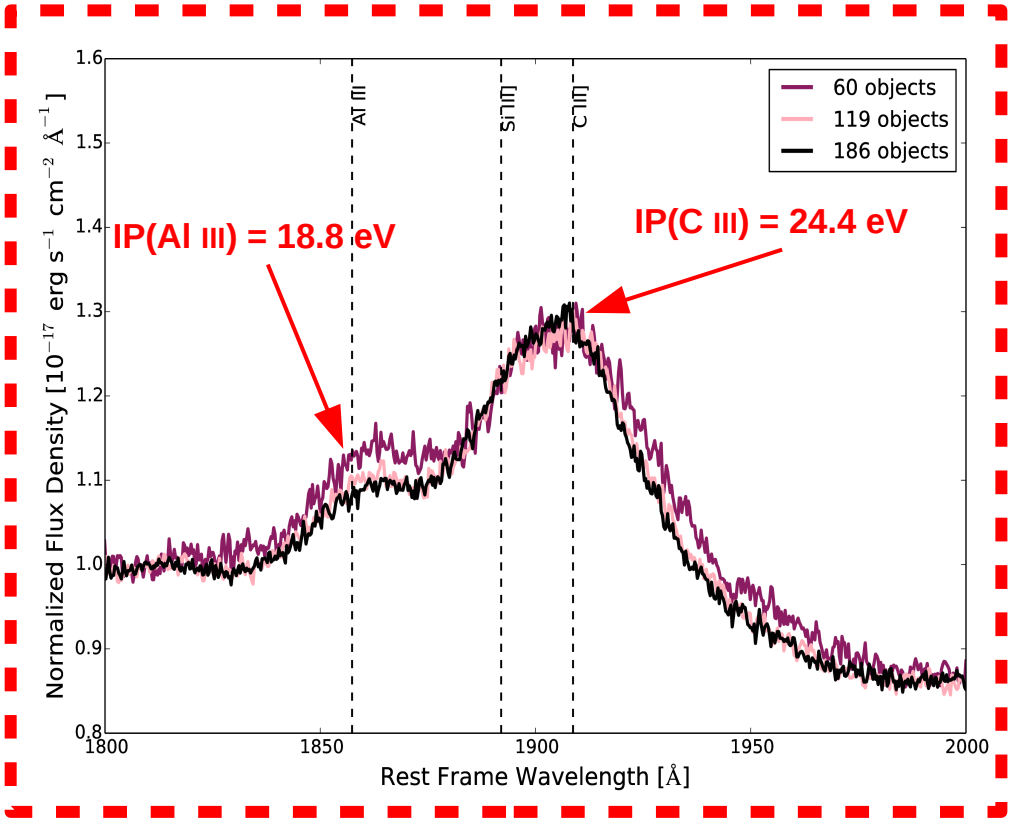
Broad, High-velocity C IV Troughs



Broad, High-velocity C IV Troughs

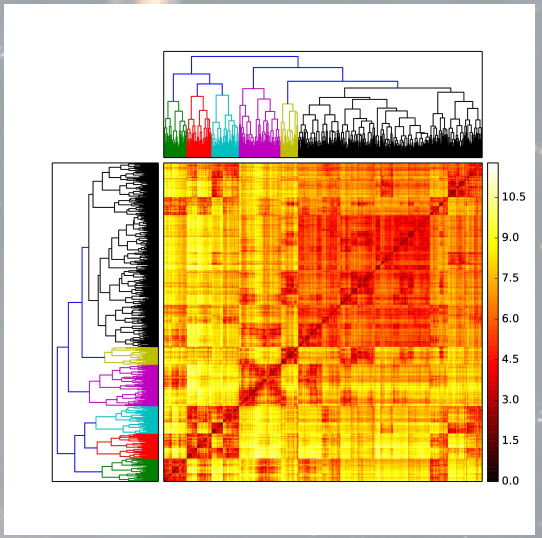


Emission line ratios as a proxy for the hardness of the SED.

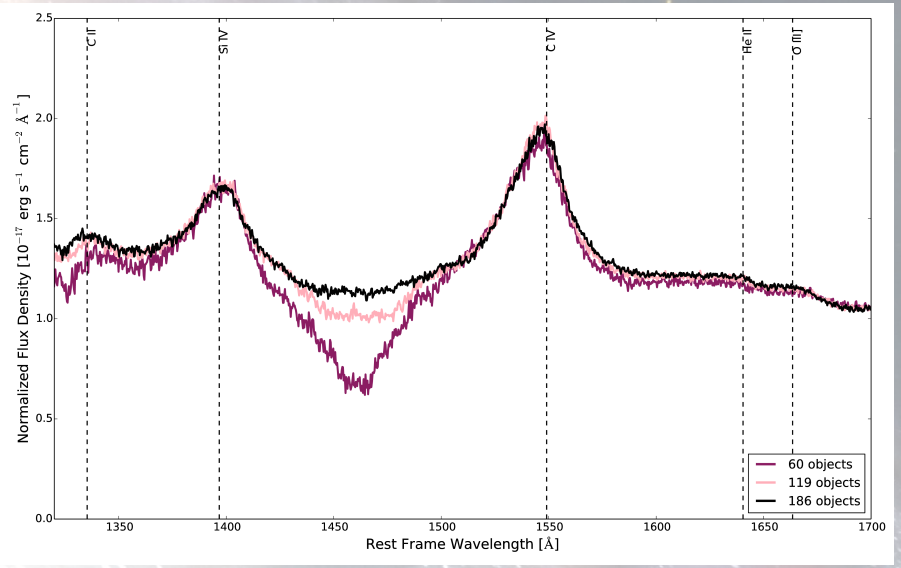


Conclusions/Next Steps

- We apply hierarchical clustering to the C IV troughs of 1,084 BAL quasar spectra. We see at least 10 distinct clusters.

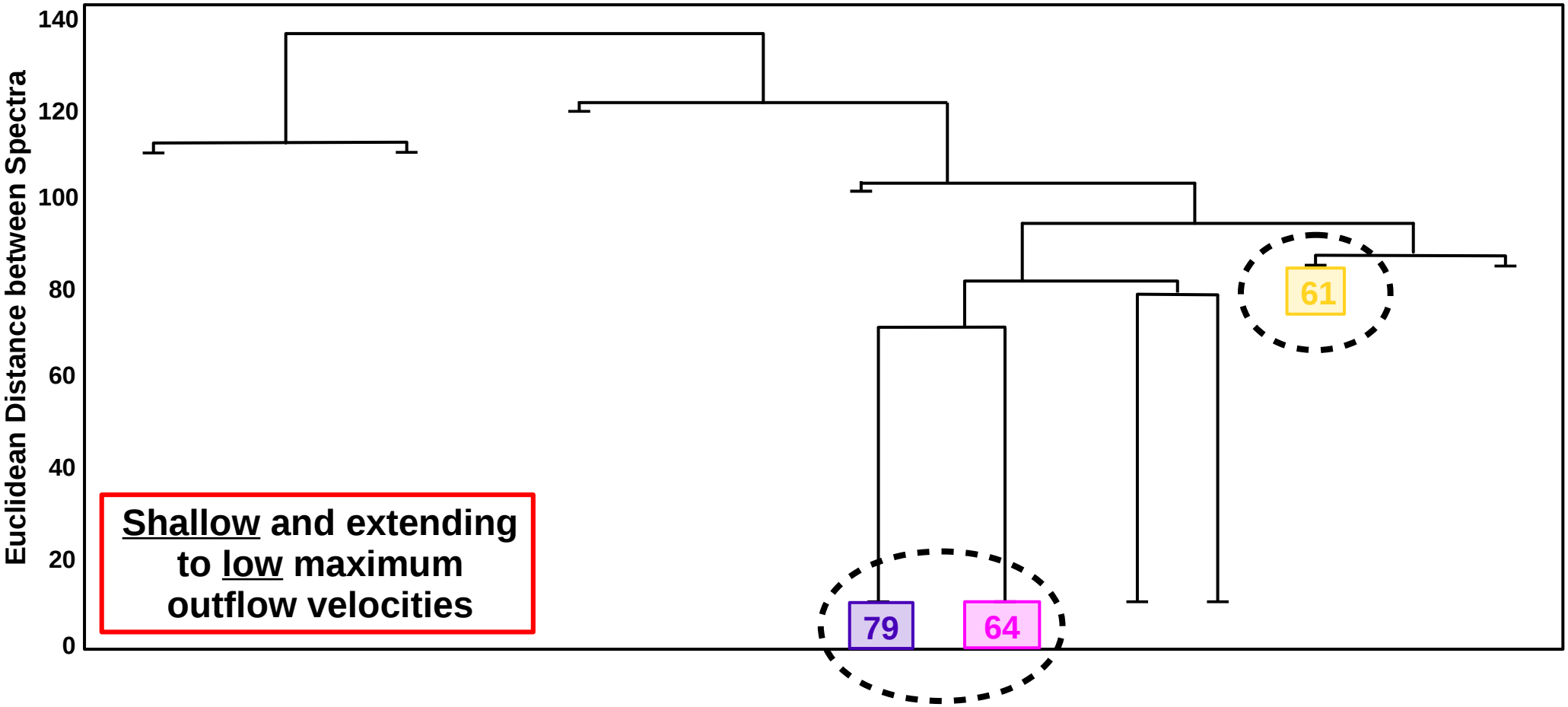


- Certain emission and aborption line properties are seen to group together in distinct categories, which could imply a common cause.

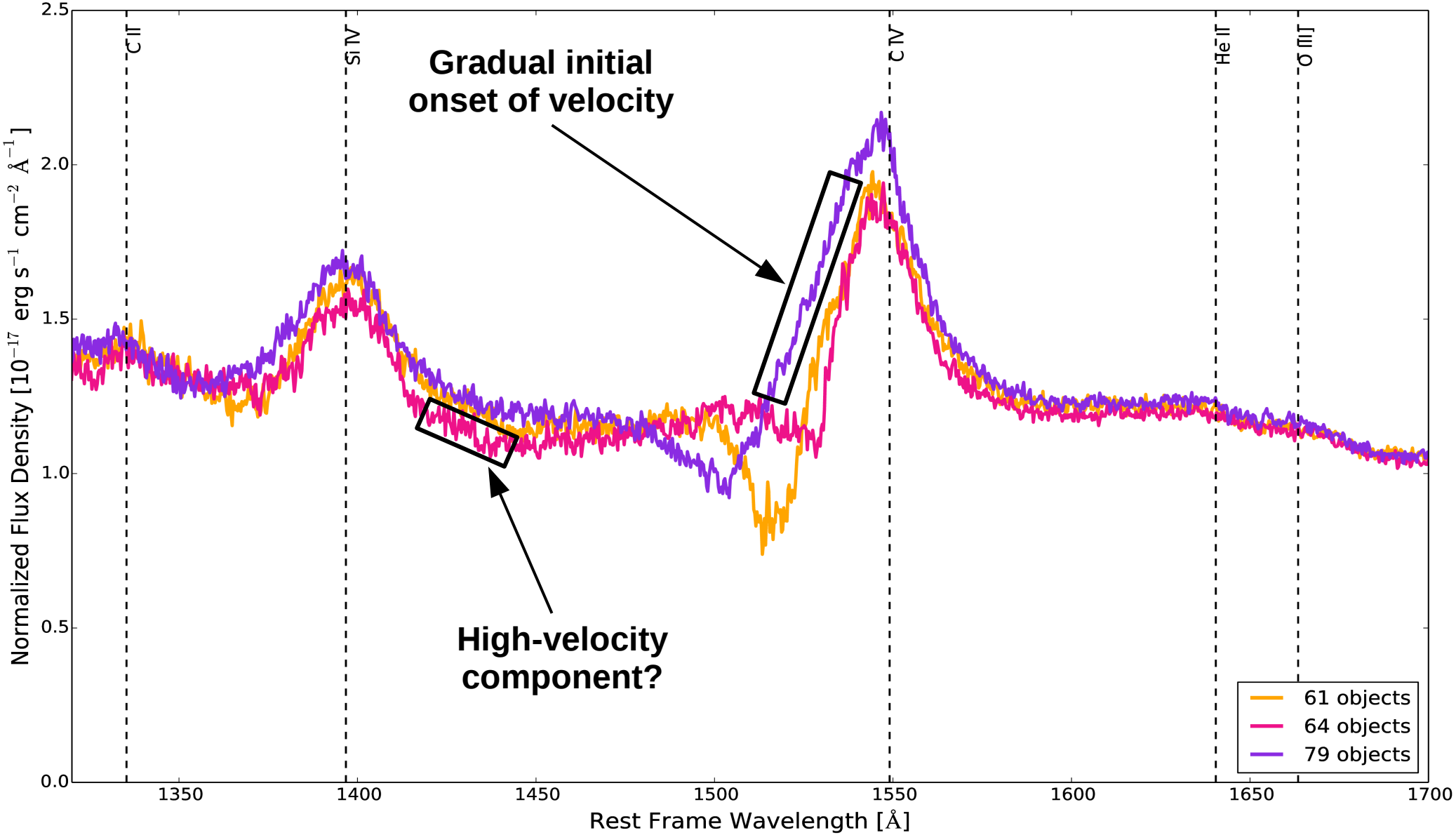


Additional Slides

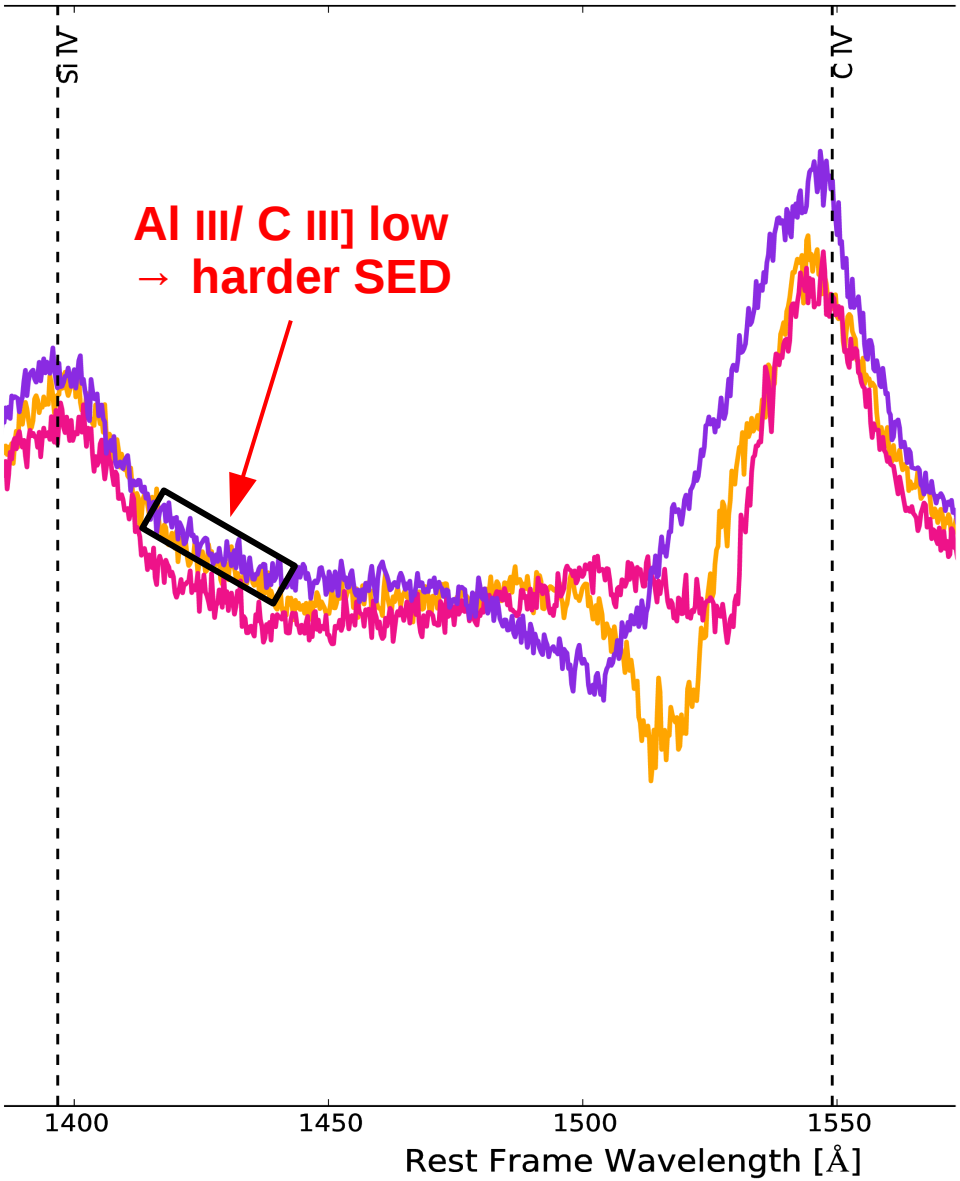
Example: Shallow, Low-velocity C IV Troughs



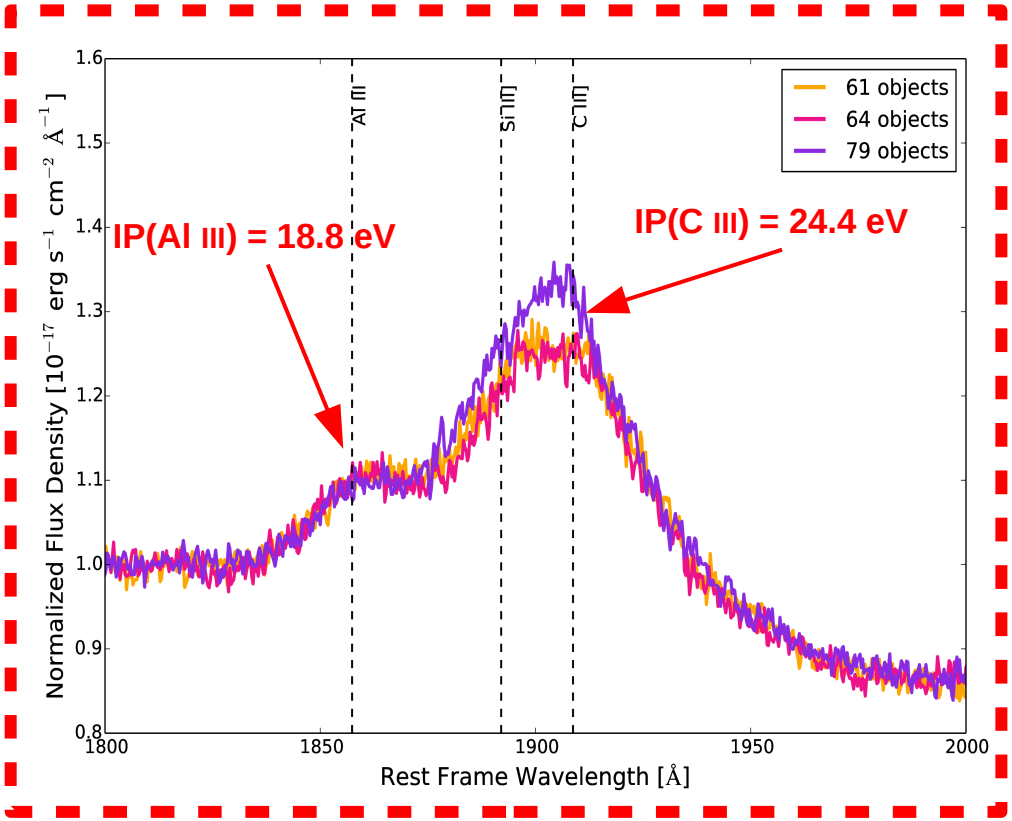
Shallow, Low-velocity C IV Troughs



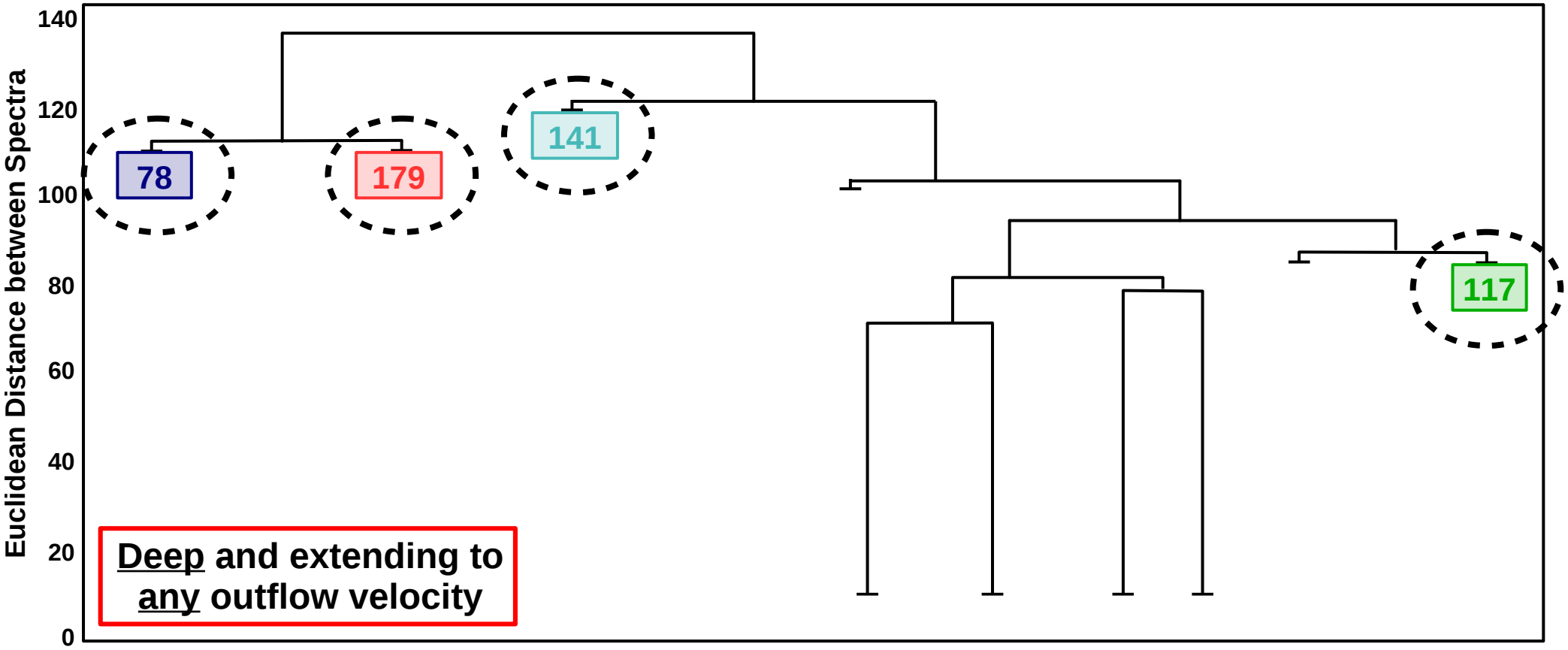
Shallow, Low-velocity C IV Troughs



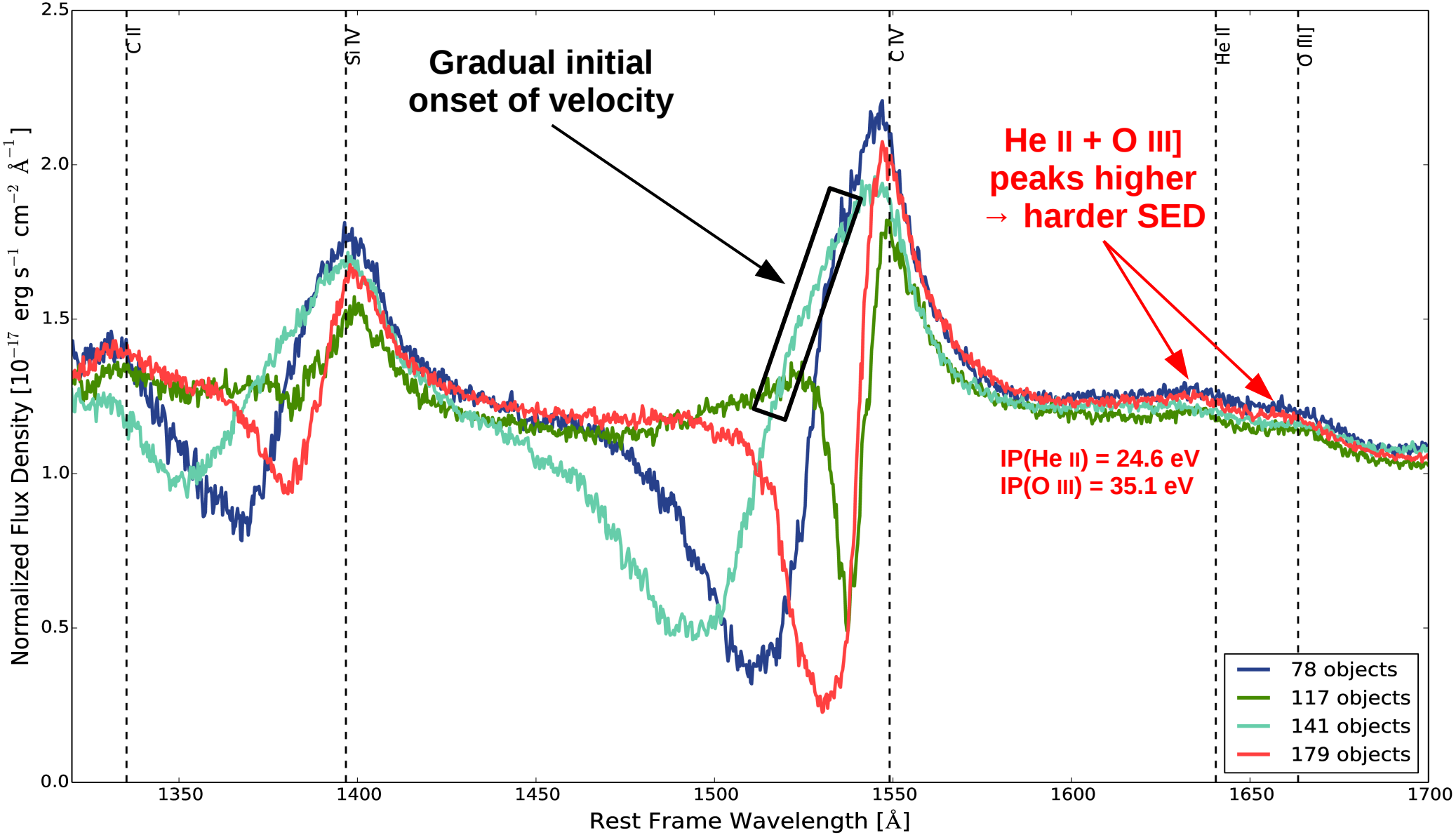
Emission line ratios as a proxy for the hardness of the SED.



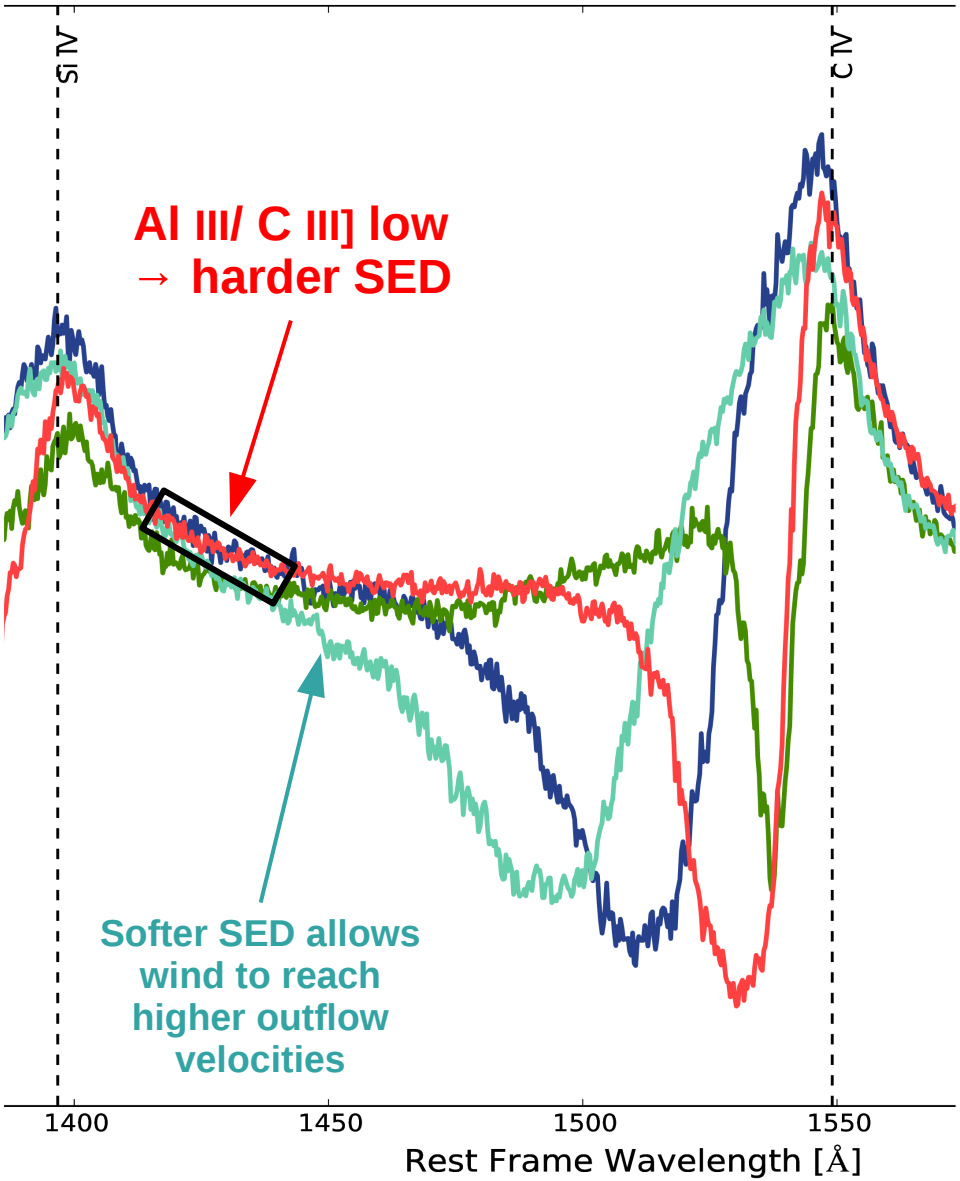
Example: Deep, Any-velocity C IV Troughs



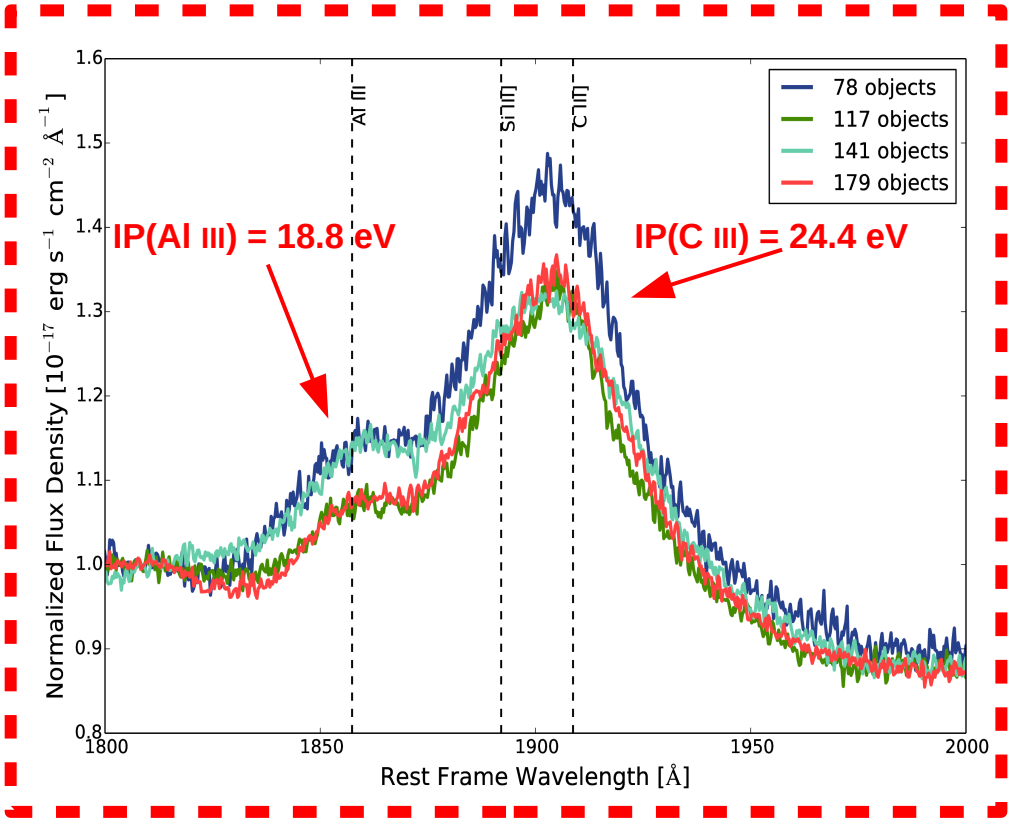
Deep, Any-velocity C IV Troughs



Deep, Any-velocity C IV Troughs



Emission line ratios as a proxy for the hardness of the SED.

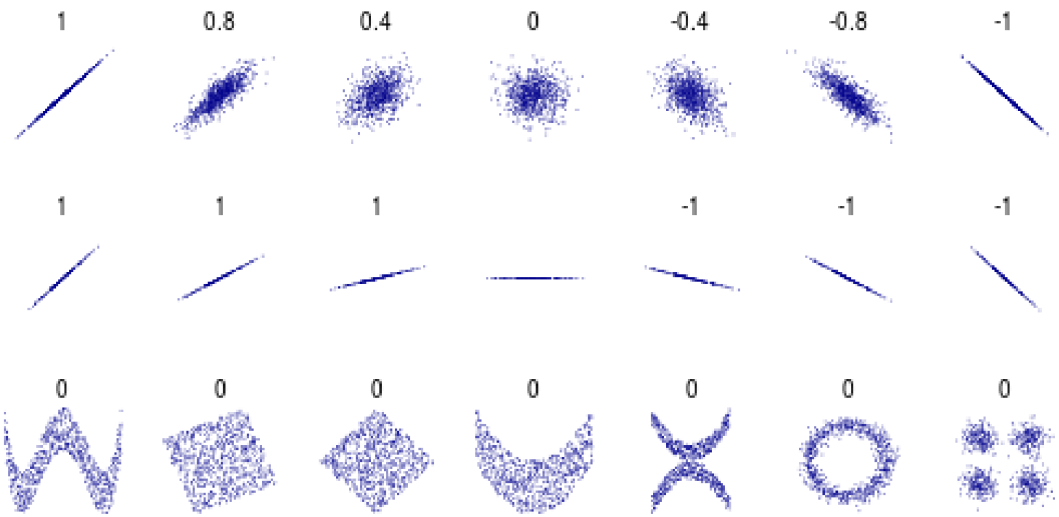


Pearson Correlation Coefficients

- Measures how two sets of data are correlated with one another.

$$r = \frac{\sum_{i=1}^{n}\left(\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

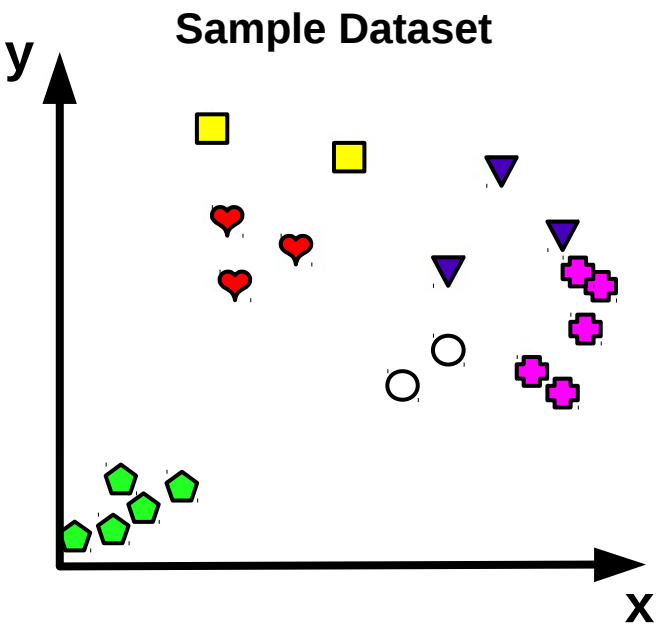


https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient

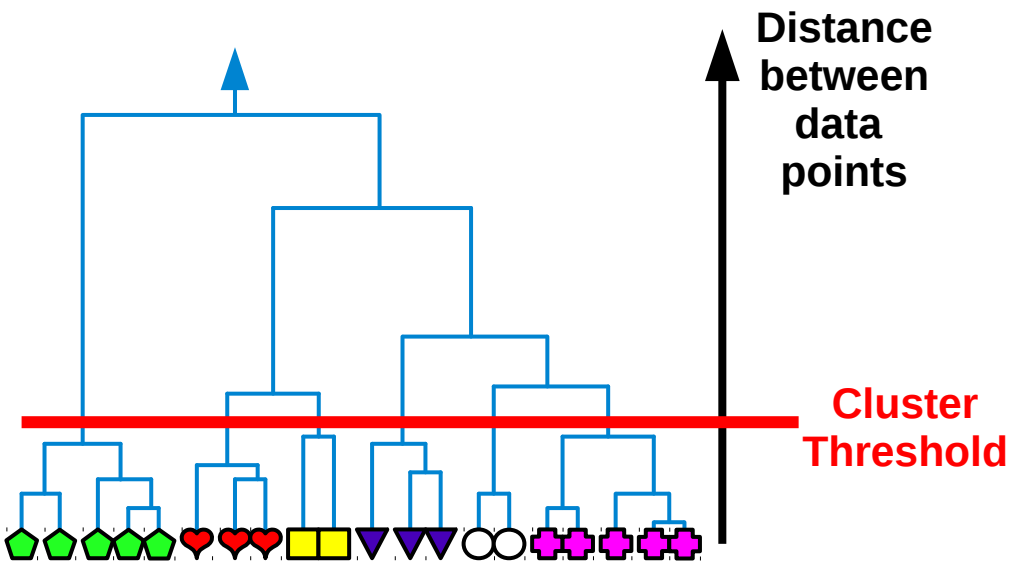
- Simple to compute in Python:

  >>> numpy.corrcoef(...)

Hierarchical Clustering Algorithm



Bovermann et al. (2008)



Dendrogram for Hierarchical Clustering of the Sample Dataset

Generalize to 1084-dimensional space...

Spectrum 1 feature vector: $(0.0, d_{1,2}, d_{1,3}, \ldots, d_{1,1084})$