Using R to Introduce Students to Principal Component Analysis, Cluster Analysis, and Multiple Linear Regression

DAVID HARVEY

PITTCON 2018

http://bit.ly/2j10Hai

Analytical Chemistry 2.0

A Collection of Free Digital Resources for Teaching Analytical Chemistry

Analytical Chemistry 2.0 is a reenvisoning of the print textbook *Modern Analytical Chemistry* (first published in 1999 by McGraw-Hill) following the return to the author of the copyright. The textbook's title is inspired by a system of software version numbering in which the sequence ***X.Y.Z*** identifies a major change in functionality (***X***), a minor change in functionality (***Y***), and small changes and/or corrections of errors that do not affect functionality (***Z***). In this scheme, *Modern Analytical Chemistry* is equivalent to *Analytical Chemistry 1.0*, with the transition from a print to a digital format representing a fundamental change in how a user experiences the textbook.

A new edition of an existing textbook is more than a correction of errors and less than a fundamental change in how a user interacts with the textbook; thus, the initial edition of *Analytical Chemistry 2.0*, which was released in fall 2009, becomes *Analytical Chemistry 2.1*, which was released in summer 2016. Files for all editions and a consideration of what might make up Analytical Chemistry 3.0 are available using the dropdown menu in the navigation bar or the links here:

Version 2.0 Version 2.1 What is Analytical Chemistry 3.0?

Additional resources developed to support this project are available here:

case studies shiny apps R scripts and packages

The following resources provide additional perspectives on this project:

- "Analytical Chemistry 2.0-An Open-Access Digital Textbook," Harvey, D. T. *J. Anal. Bioanal. Chem.* **2011**, *399*, 149-152 (DOI).
- "Analytical Chemistry 2.1: An Open-Access Digital Resource for Undergraduate Education in Analytical Chemistry," poster presented at the 2016 Pittcon Conference (link).

If you do make use of these materials, please let me know by sending an email to harvey@depauw.edu.

http://bit.ly/2j10Hai

Context

Prelude

Main Feature

(in 3 Parts w/2 Themes)

Chem 351: Chemometrics

This course, **Chem 351: Chemometrics**, provides an introduction to how chemists and biochemists can extract useful information from the data they collect in lab, including, among other topics, how to summarize data, how to visualize data, how to test data, how to build quantitative models to explain data, how to design experiments, and how to separate a useful signal from noise.

Two 60 minute class periods per week for 14 weeks. Course is divided into 12 units.

- Unit 1: Introduction
- Unit 2: Basic Statistics
- Unit 3: Distribution of Data
- Unit 4: Confidence Intervals
- Unit 5: Analysis of Variance
- Unit 6: Linear Regression
- Unit 7: 3D Visualizations
- Unit 8: Experimental Design
- Unit 9: Signal Processing
- Unit 10: Principal Component Analysis
- Unit 11: Cluster Analysis
- Unit 12: Multiple Linear Regression

http://bit.ly/2j10Hai

ℝ as a Tool for Teaching Chemometrics

- data-centric programming language and environment for statistical computing
- large number of users ensures longevity of software
- base installation provides access to a wide variety of computational methods for processing data and tools for visualizing data
- highly extensible through user-written scripts and packages of functions
- available via Free Software Foundation's GNU General Public License
- versions for UNIX, Linux, Windows, and MacOS platforms
- easy to interweave text, tables, and figures
- base packages are very stable so code is resilient

for further details regarding ℝ, see https://www.r-project.org/

Make ℝ a tool, not a barrier. Emphasis is on adapting coding examples, using coding templates, using available packages, and using functions provided by instructor

The Analytical System: Beer's Law

- stock standards
  - 0.0500 M Cu^{2+}
  - 0.1000 M Co^{2+}
  - 0.0375 M Cr^{3+}
  - 0.1300 M Ni^{2+}
  - all in 0.10 M HNO_3
- samples prepared from stocks
  - single metal ions
  - binary mixtures of metal ions
  - ternary mixtures of metal ions
  - quaternary mixtures of metal ions

- spectra collected with a Vernier SpectroVis Plus spectrometer; exported as .csv files
- individual .csv files combined into a single file, cleaned up, and saved as a .csv file.
- file has 80 rows (one per sample) and 642 columns (seven with information on composition of samples and absorbance values at 635 wavelengths).
- data set is brought into ℝ using the readr package's read_csv( ) function and then further subsetted within R to create individual data files.

Context

Prelude

Main Feature

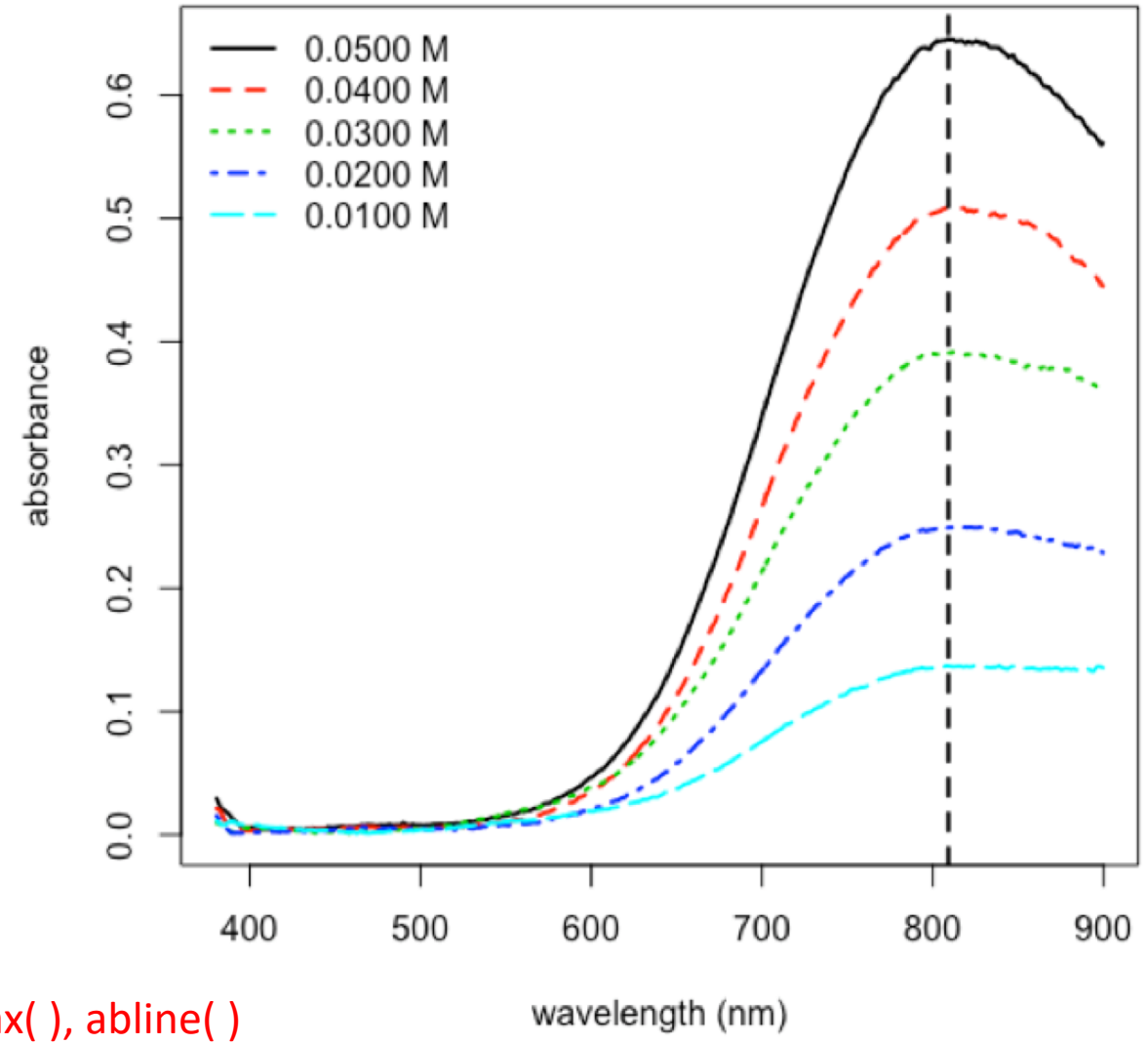
(in 3 Parts w/2 Themes)

External Standardization for Copper

$$A_{\lambda,Cu} = \varepsilon_{\lambda,Cu} b C_{Cu}$$

1. **plot spectra for set of standards and identify the wavelength of maximum absorbance**

$$\lambda = 809.1 \text{ nm}$$



R functions: readr::read_csv( ), matplot( ), legend( ), which.max( ), abline( )

http://bit.ly/2j10Hai

External Standardization for Copper

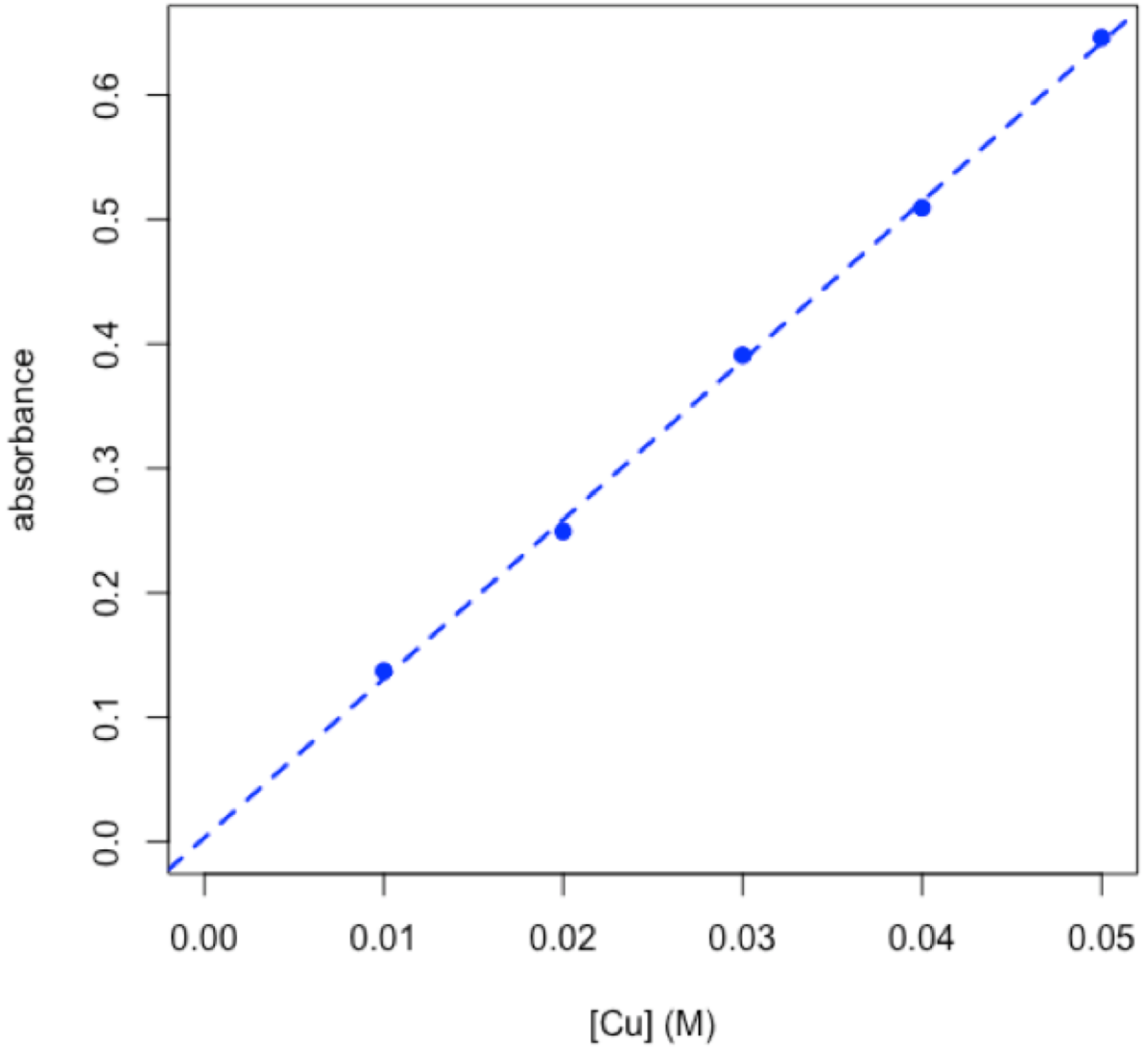
$$A_{\lambda,Cu} = \varepsilon_{\lambda,Cu} b C_{Cu}$$

1. plot spectra for set of standards and identify the wavelength of maximum absorbance
2. **plot calibration data and determine equation for calibration curve**

```
Coefficients:     Estimate     Std. Error    t value      Pr(>|t|)
(Intercept)       0.003202     0.008222      0.389         0.723
cuStd_conc       12.778883   0.247900       51.548       1.61e-05 ***
---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.007839 on 3 degrees of freedom
Multiple R-squared:  0.9989, Adjusted R-squared:  0.9985
F-statistic:  2657 on 1 and 3 DF,  p-value: 1.608e-05
```

$$A = 12.78\ M^{-1} \times C + 0.0032$$

R functions: plot( ), lm( ), abline( ), summary( )



http://bit.ly/2j10Hai

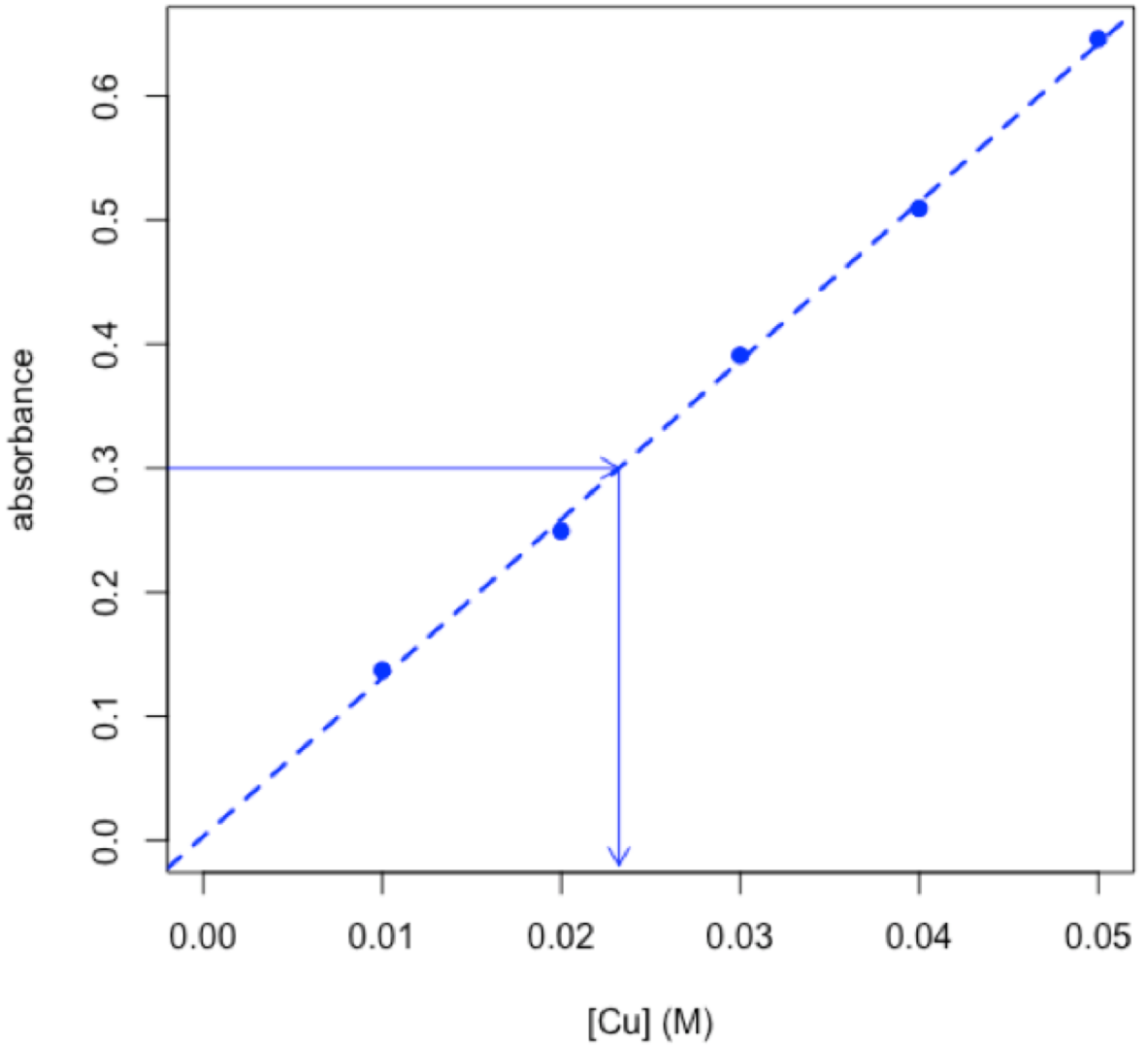
External Standardization for Copper

$$A_{\lambda,Cu} = \varepsilon_{\lambda,Cu} b C_{Cu}$$

1. plot spectra for set of standards and identify the wavelength of maximum absorbance
2. plot calibration data and determine equation for calibration curve
3. **determine concentration of copper in an unknown using the chemCal package**

```
$Prediction[1] 0.02322567
$`Standard Error`[1] 0.0006847376
$Confidence[1] 0.002179141
$`Confidence Limits`[1] 0.02104653  0.02540481
```

$$C_{cu} = 0.0232 \text{ M} \pm 0.0022 \text{ M}$$



R functions: chemCal::inverse.predict( ), arrows( )

Analysis of Binary Mixture: Cu and Ni

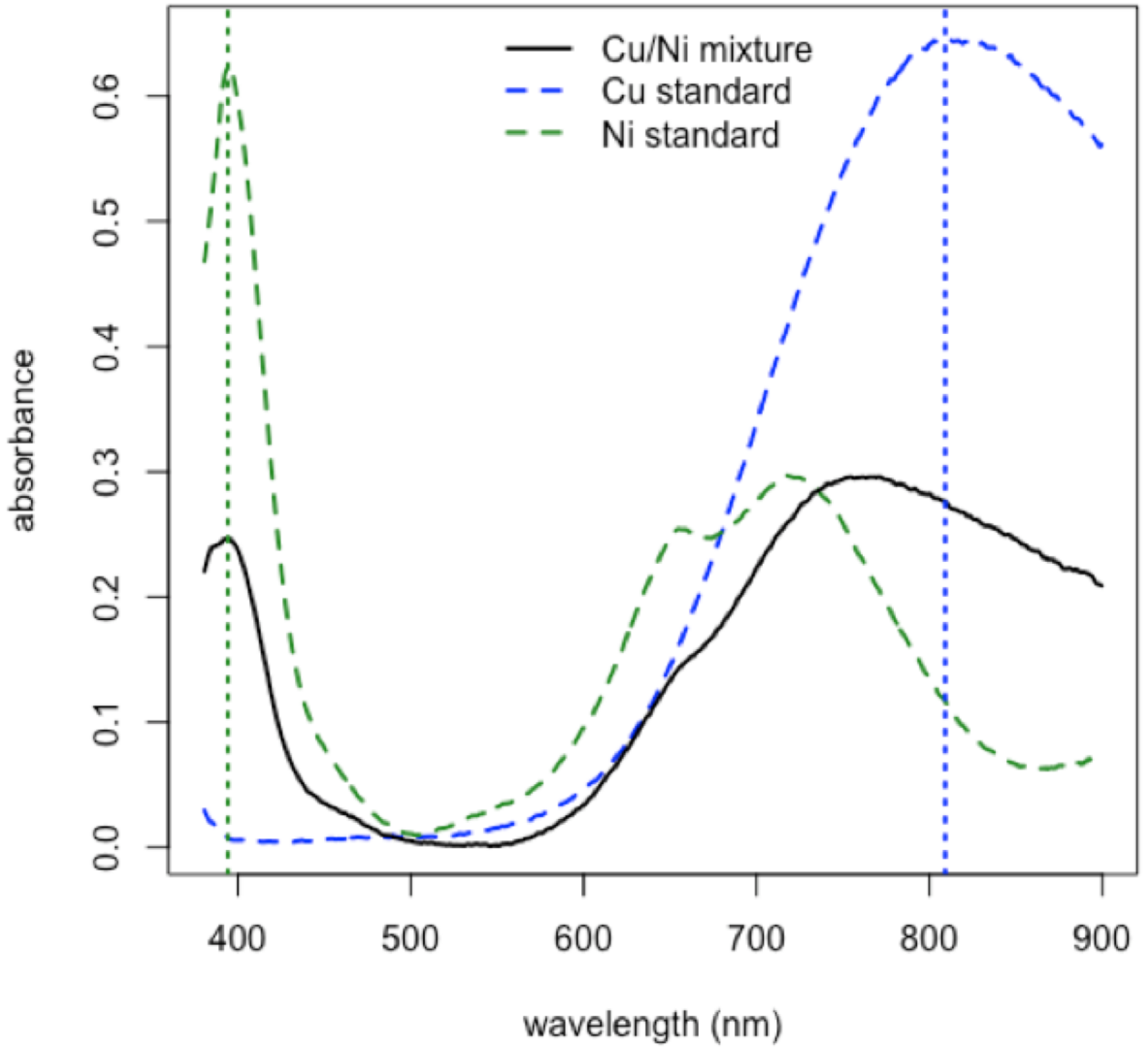
$$A_{\lambda 1,mix} = \varepsilon_{\lambda 1,Cu} b C_{Cu} + \varepsilon_{\lambda 1,Ni} b C_{Ni}$$

$$A_{\lambda 2,mix} = \varepsilon_{\lambda 2,Cu} b C_{Cu} + \varepsilon_{\lambda 2,Ni} b C_{Ni}$$

1. **plot spectra for a Cu standard, for a Ni standard, and for a mixture, and identify the wavelengths to use for the analysis**

$\lambda 1 = 809.1$ nm

$\lambda 2 = 394.2$ nm



R functions: plot( ), lines( ), legend( ), which.max( ), abline( )

http://bit.ly/2j10Hai

Analysis of Binary Mixture: Cu and Ni

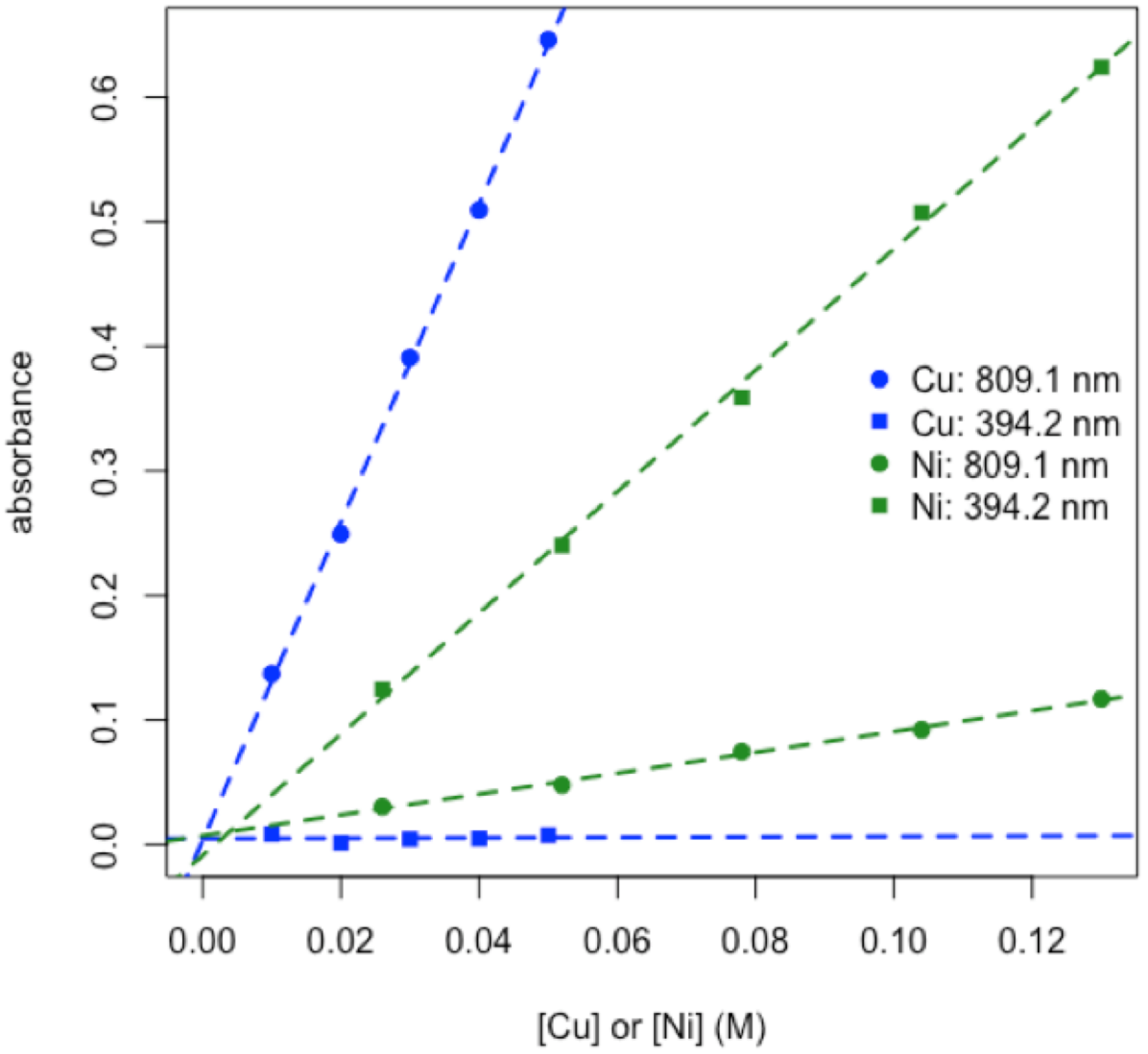
$$A_{\lambda 1,mix} = \varepsilon_{\lambda 1,Cu} b C_{Cu} + \varepsilon_{\lambda 1,Ni} b C_{Ni}$$

$$A_{\lambda 2,mix} = \varepsilon_{\lambda 2,Cu} b C_{Cu} + \varepsilon_{\lambda 2,Ni} b C_{Ni}$$

1. plot spectra for a Cu standard, for a Ni standard, and for a mixture, and identify the wavelengths to use for the analysis
2. **plot calibration data and determine values for εb for each metal at each wavelength**

$\varepsilon_{809.1,Ni} b = 0.837\ M^{-1}$ and $\varepsilon_{394.2,Ni} b = 4.870\ M^{-1}$

$\varepsilon_{809.1,Cu} b = 12.78\ M^{-1}$ and $\varepsilon_{394.2,Cu} b = 0.017\ M^{-1}$



R functions: plot( ), legend( ), lm( ), abline( )

http://bit.ly/2j10Hai

Analysis of Binary Mixture: Cu and Ni

$$A_{\lambda 1,mix} = \varepsilon_{\lambda 1,Cu} b C_{Cu} + \varepsilon_{\lambda 1,Ni} b C_{Ni}$$

$$A_{\lambda 2,mix} = \varepsilon_{\lambda 2,Cu} b C_{Cu} + \varepsilon_{\lambda 2,Ni} b C_{Ni}$$

1. plot spectra for a Cu standard, for a Ni standard, and for a mixture, and identify the wavelengths to use for the analysis
2. plot calibration data and determine values for εb for each metal at each wavelength
3. **use R's solve function to calculate concentrations of copper and nickel in mixture**

C_{Cu} = 0.0182 M and C_{Ni} = 0.0506 M

ℝ functions: matrix( ), colnames( ), rownames( ), solve( )

2×2 matrix of εb values

	Cu	Ni
809.1 nm	12.77888326	0.8369356
394.2 nm	0.01747073	4.8700205

2×1 matrix of absorbance values

	mixture
809.1 nm	0.2789825
394.2 nm	0.2415476

calculate 2×1 matrix of concentrations

[eb] × [conc] = [abs]

	mixture
Cu	0.01858748
Ni	0.04953220

http://bit.ly/2j10Hai

Matrix Notation for Beer's Law ($A = \varepsilon bC$)

two analytes: two samples with absorbance measured at two wavelengths

$$[A]_{2 \text{ samples} \times 2 \text{ wavelengths}} = [C]_{2 \text{ samples} \times 2 \text{ analytes}} \times [\varepsilon b]_{2 \text{ analytes} \times 2 \text{ wavelengths}}$$

one analyte: one sample with absorbance measured at one wavelength

$$[A]_{1 \text{ sample} \times 1 \text{ wavelength}} = [C]_{1 \text{ sample} \times 1 \text{ analyte}} \times [\varepsilon b]_{1 \text{ analyte} \times 1 \text{ wavelength}}$$

overdetermined system: more samples and wavelengths than analytes

$$[A]_{8 \text{ samples} \times 5 \text{ wavelengths}} = [C]_{8 \text{ samples} \times 2 \text{ analytes}} \times [\varepsilon b]_{2 \text{ analytes} \times 5 \text{ wavelengths}} + [RE]_{8 \text{ samples} \times 5 \text{ wavelengths}}$$

generalize: n analytes, s samples, and w wavelengths where $n \leq$ smaller of s or w

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$

Context

Prelude

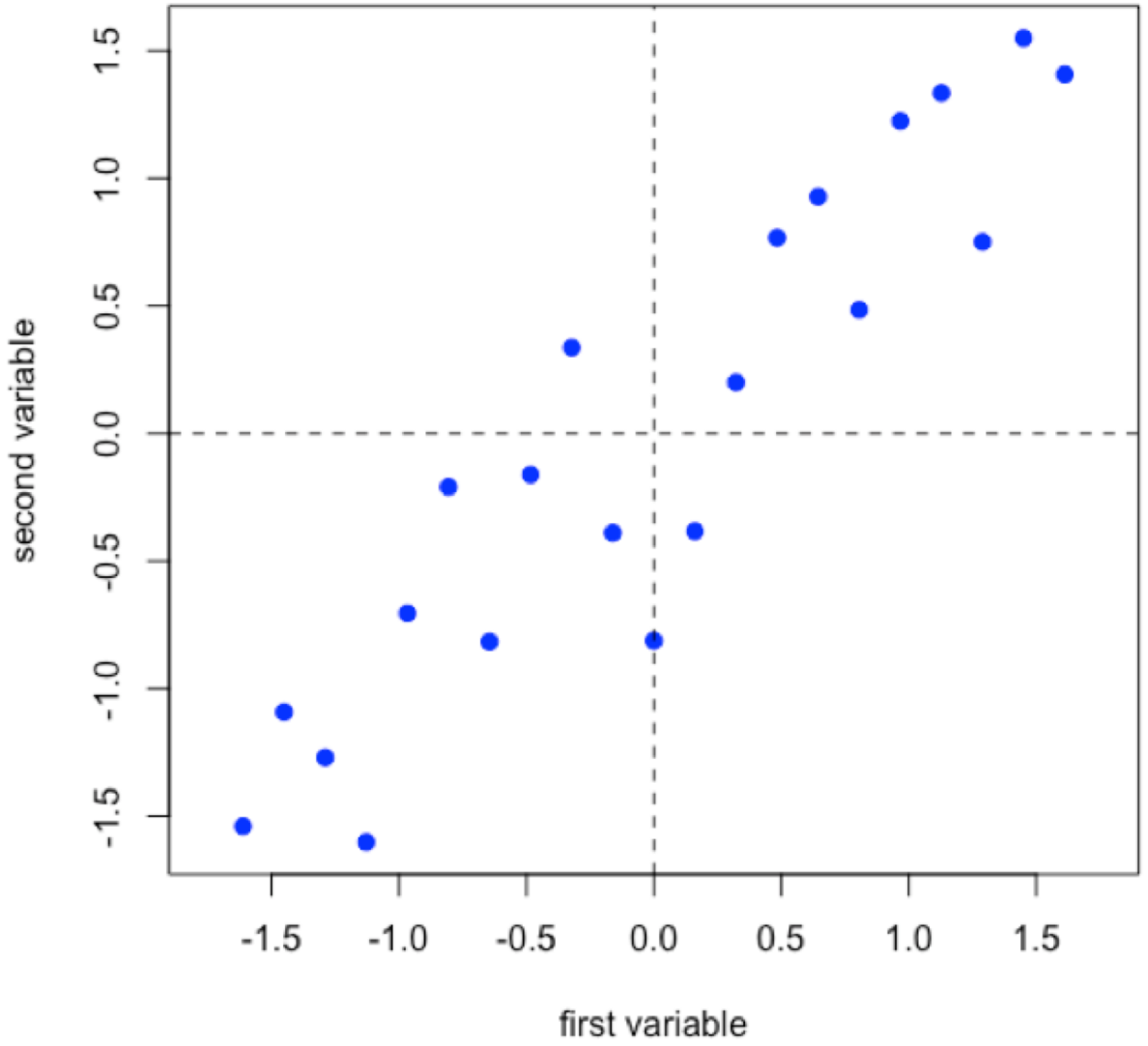
Main Feature

(in 3 Parts w/2 Themes)

How Does PCA Work?

Suppose we have 21 samples and that we measure two properties—first variable and second variable—for each sample giving a matrix of data, [*D*], that has 21 rows and 2 columns.

$$[D]_{21 \times 2}$$

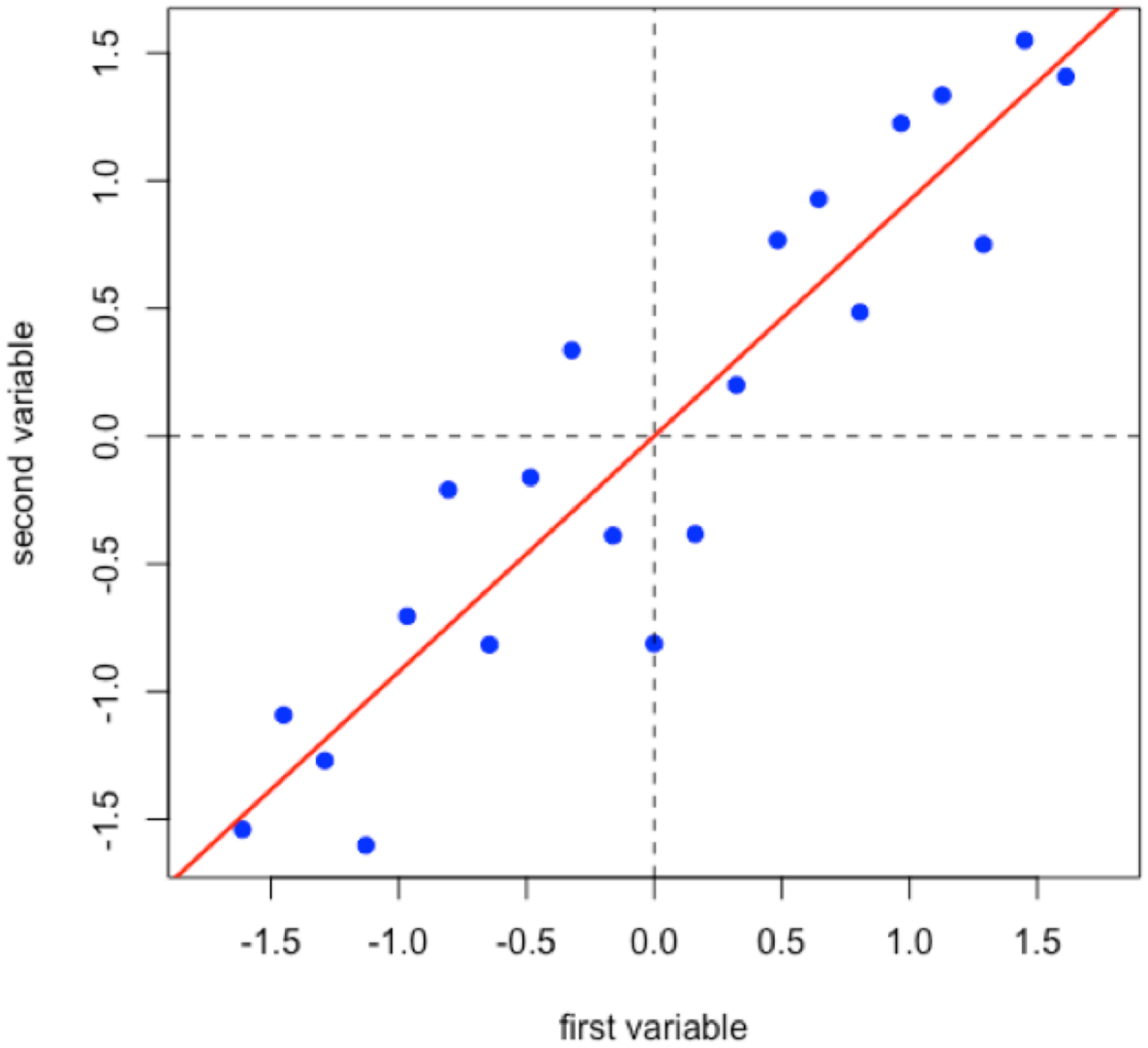


R functions: seq( ), rnorm( ), scale( ), plot( ), abline( )

http://bit.ly/2j10Hai

How Does PCA Work?

Linear regression provides the line of best fit to the data and explains more of the data's overall variance than either of the two individual variables; we call this the first principal component.



ℝ functions: lm( ), abline( )

http://bit.ly/2j10Hai

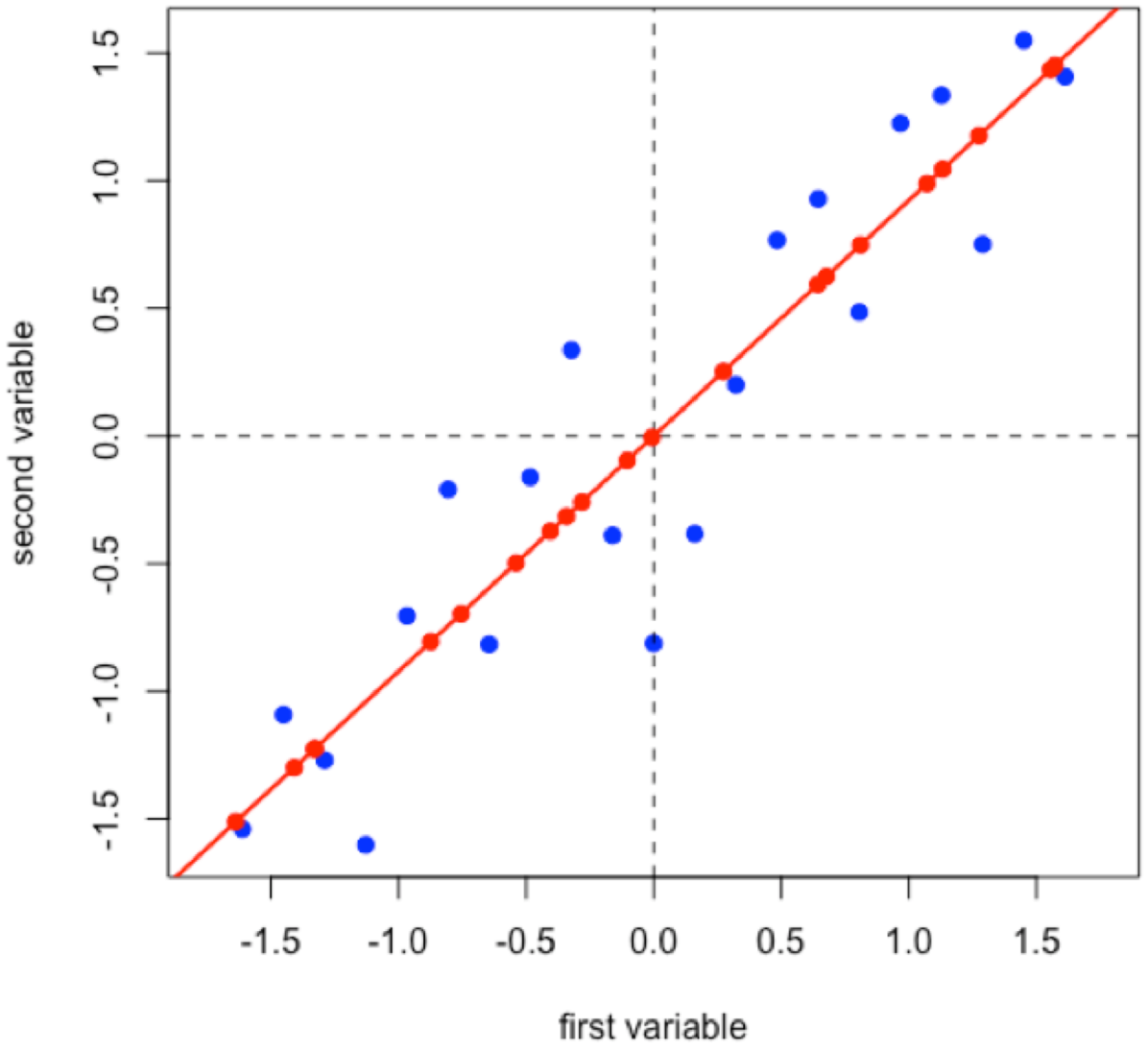
How Does PCA Work?

Projecting the data onto the regression line gives the location of the data on the first principal component; these are called scores, (*S*). The cosines of the angles between the first principal component and each of the original axes are called loadings, (*L*).

$$[D]_{21 \times 2} = [S]_{21 \times 1} \times [L]_{1 \times 2}$$

each sample has a score

each variable has a loading



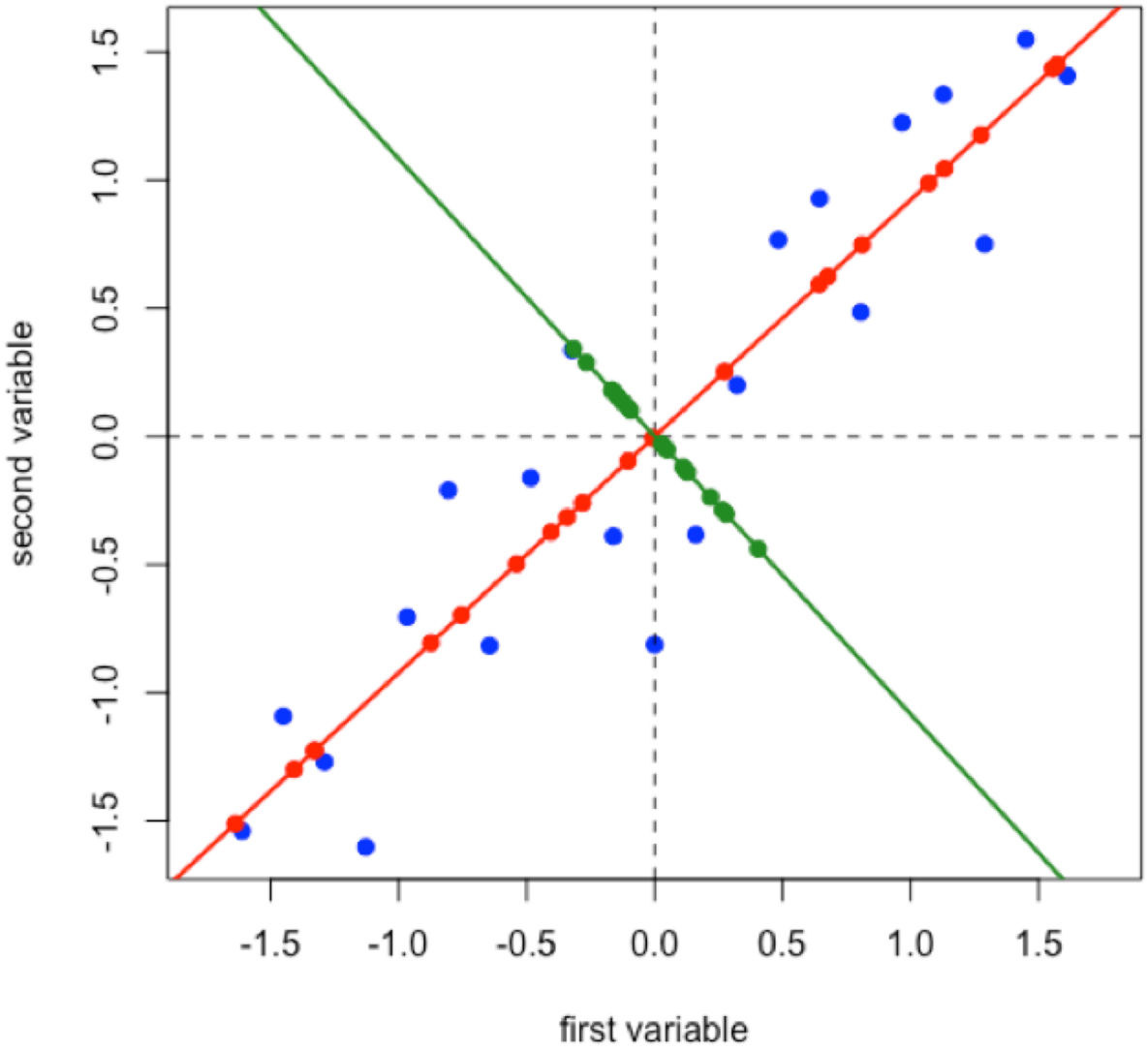
R functions: points( )

http://bit.ly/2j10Hai

How Does PCA Work?

Projecting the original data onto a line that is perpendicular to the first principal component gives the second principal component and adds in a second set of scores and loadings.

$$[D]_{21 \times 2} = [S]_{21 \times 2} \times [L]_{2 \times 2}$$



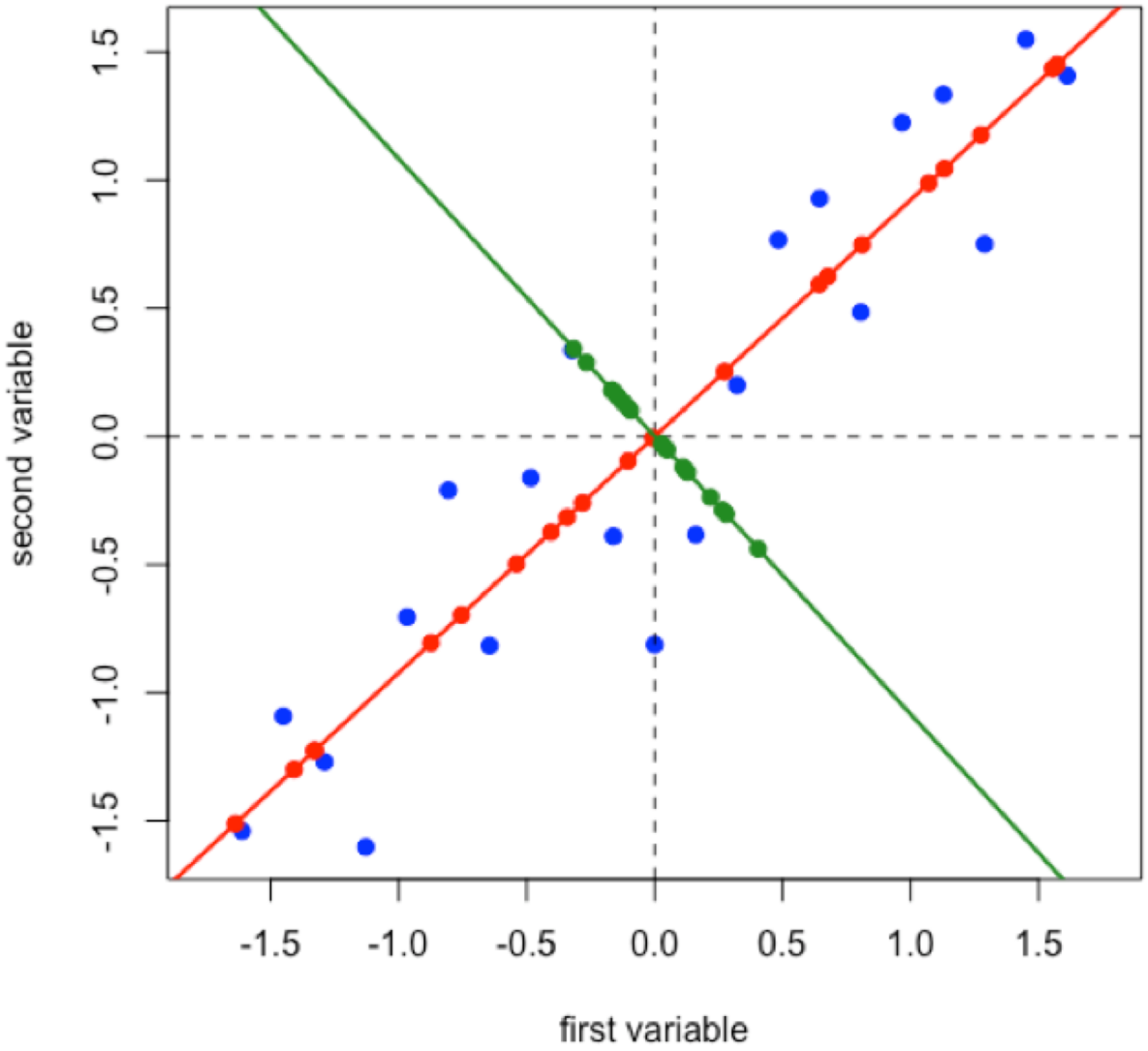
R functions: abline( ), points( )

How Does PCA Work?

Projecting the original data onto a line that is perpendicular to the first principal component gives the second principal component and adds in a second set of scores and loadings.

$$[D]_{21 \times 2} = [S]_{21 \times 2} \times [L]_{2 \times 2}$$

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$



R functions: abline( ), points( )

http://bit.ly/2j10Hai

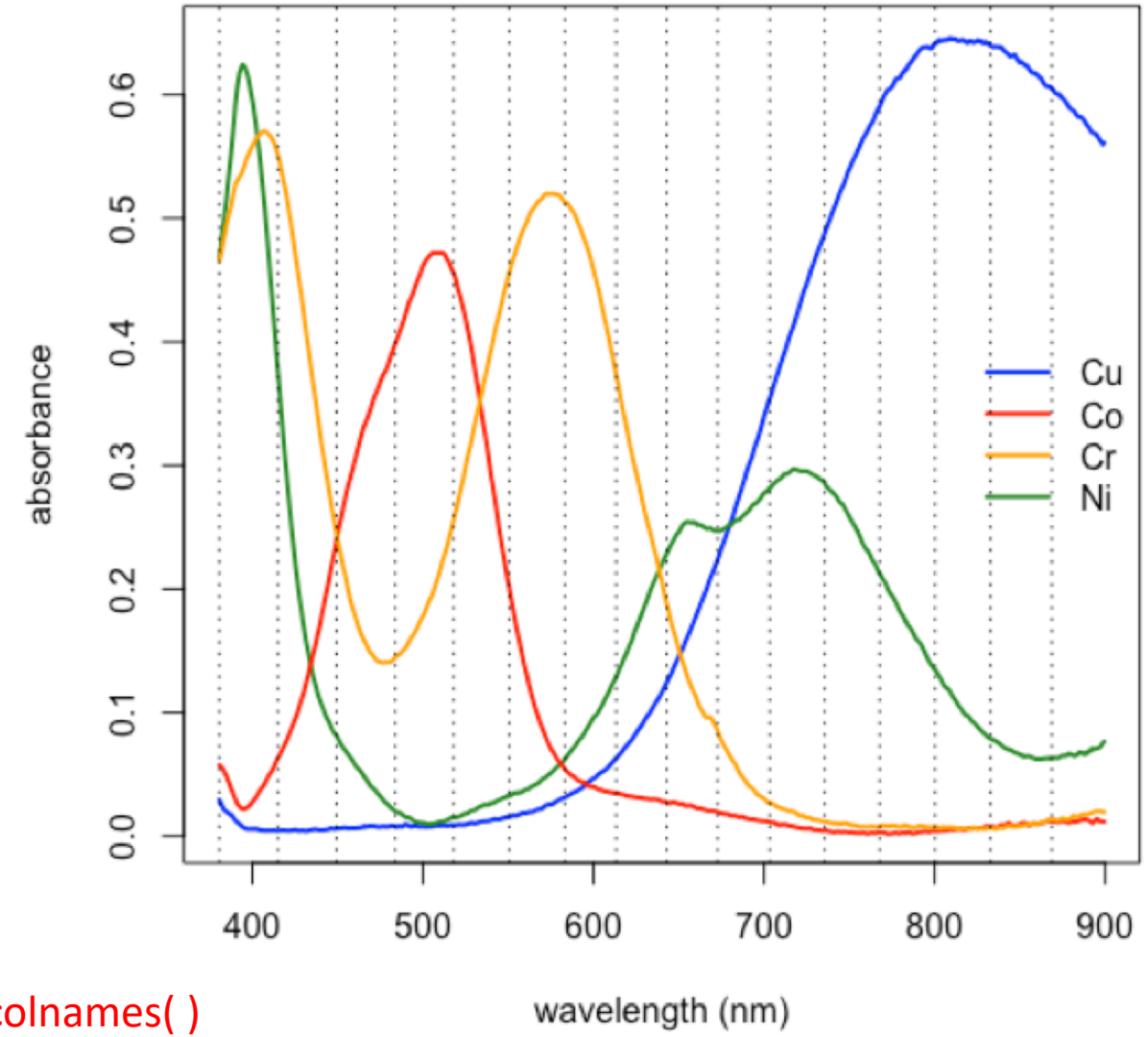
PCA: Worked Example

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. choose a subset of the original 635 wavelengths

```
[1] 380.5  414.9  449.3  483.7  517.9  550.6  583.2  613.3
[9] 642.9  672.7  703.3  735.5  767.8  800.2  832.6  868.7
```



R functions: plot( ), lines( ), abline( ), legend( ), as.numeric( ), colnames( )

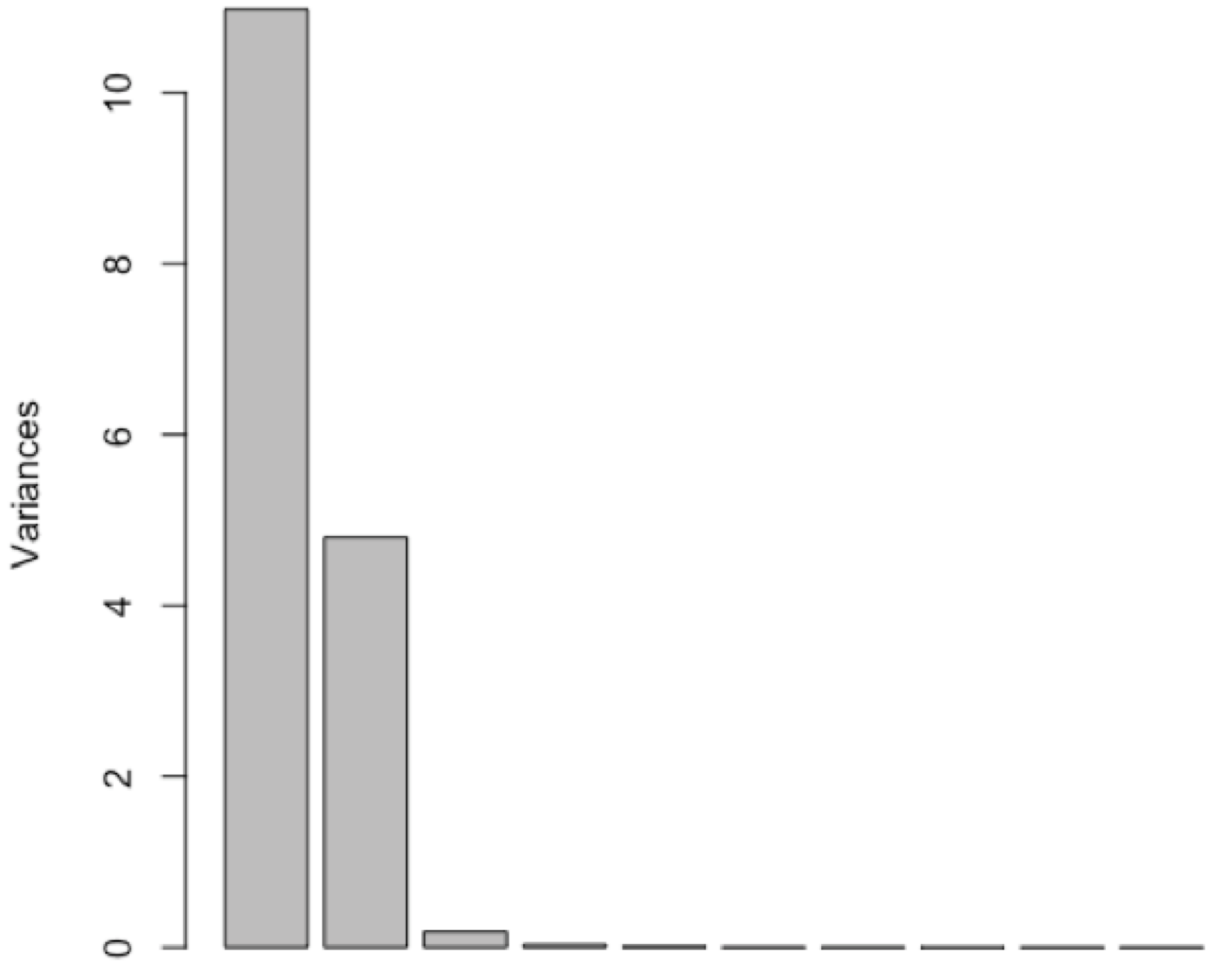
PCA: Worked Example

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. choose a subset of the original 635 wavelengths
2. **perform PCA and determine relative importance of the 16 principal components**

	PC1	PC2	PC3	PC4
Standard deviation	3.3134	2.1901	0.42561	0.17585
Proportion of Variance	0.6862	0.2998	0.01132	0.00193
Cumulative Proportion	0.6862	0.9859	0.99725	0.99919



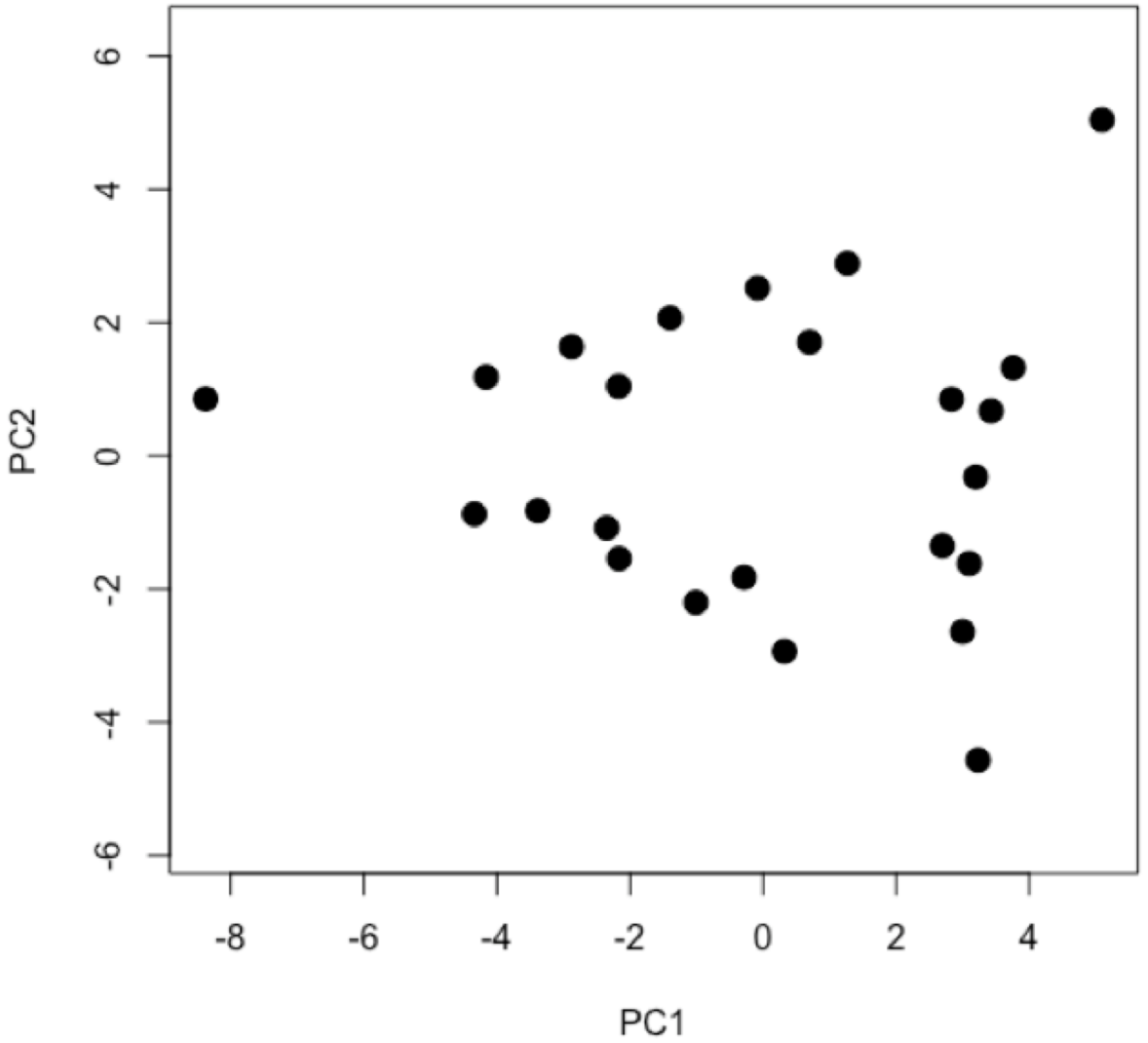
R functions: prcomp( ), plot( ), summary( )

PCA: Worked Example

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. choose a subset of the original 635 wavelengths
2. perform PCA and determine relative importance of the 16 principal components
3. **examine and interpret scores for first two principal components**



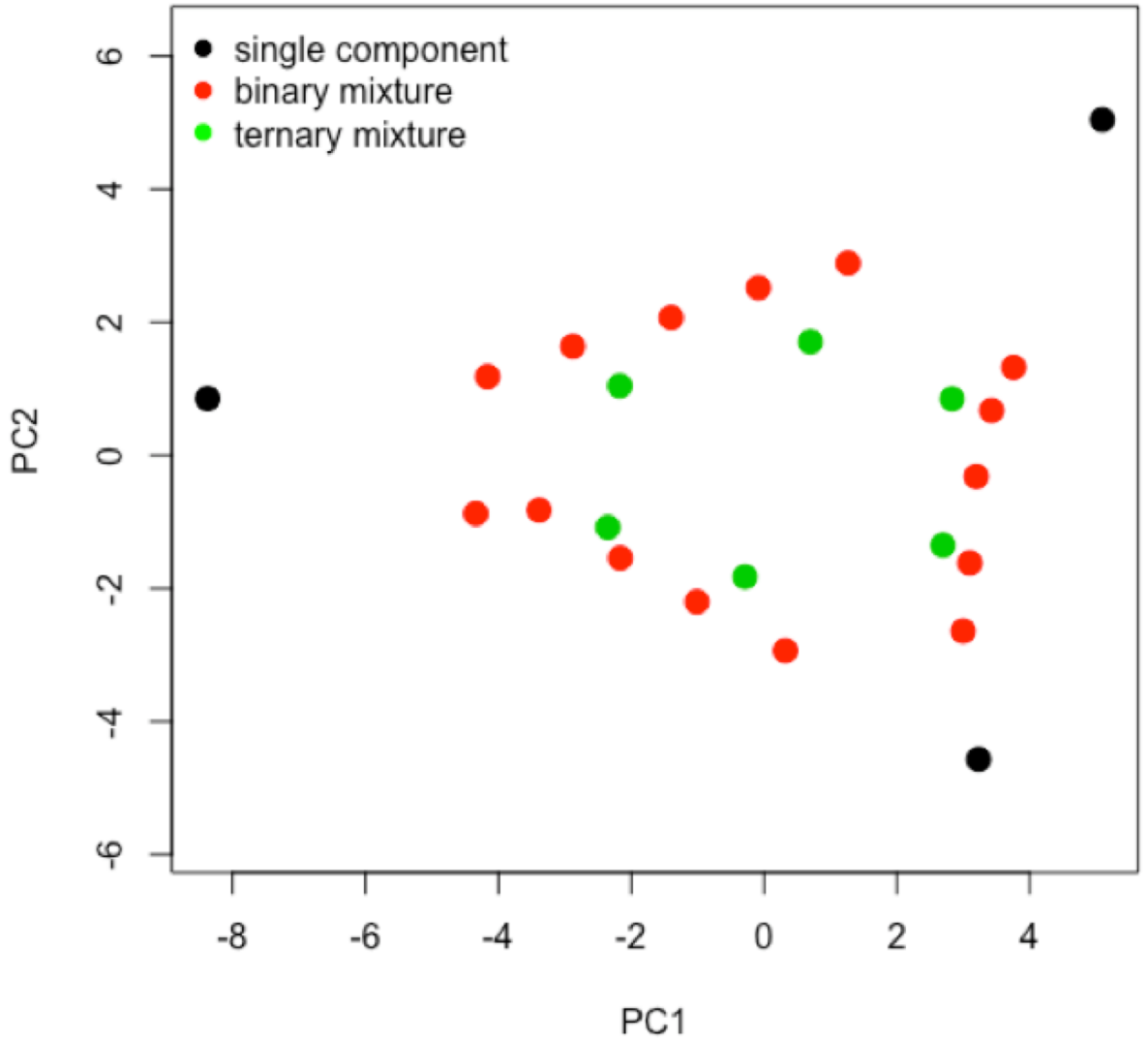
R functions: plot( )

PCA: Worked Example

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. choose a subset of the original 635 wavelengths
2. perform PCA and determine relative importance of the 16 principal components
3. **examine and interpret scores for first two principal components**



ℝ functions: plot( ), factor( ), legend( )

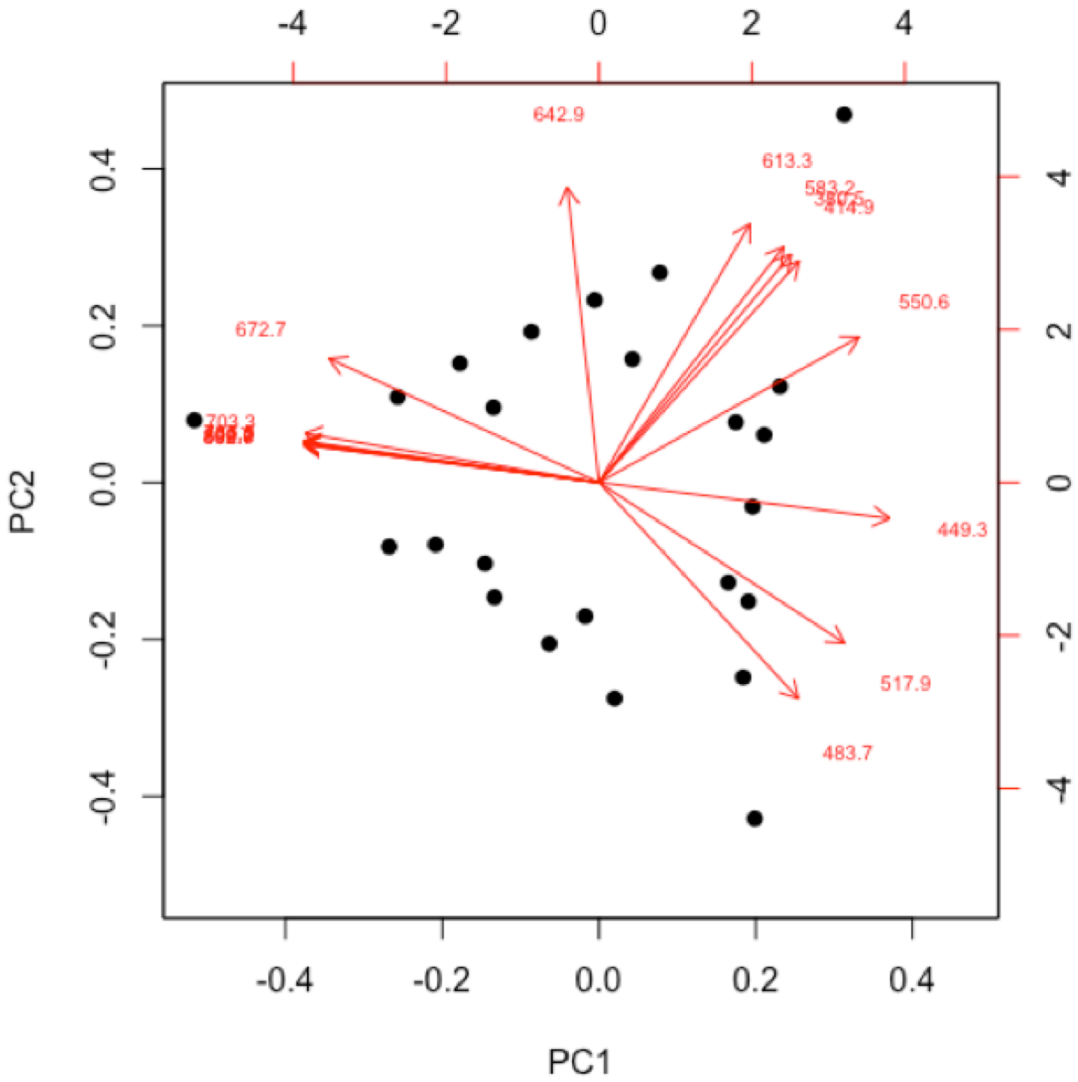
http://bit.ly/2j10Hai

PCA: Worked Example

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$

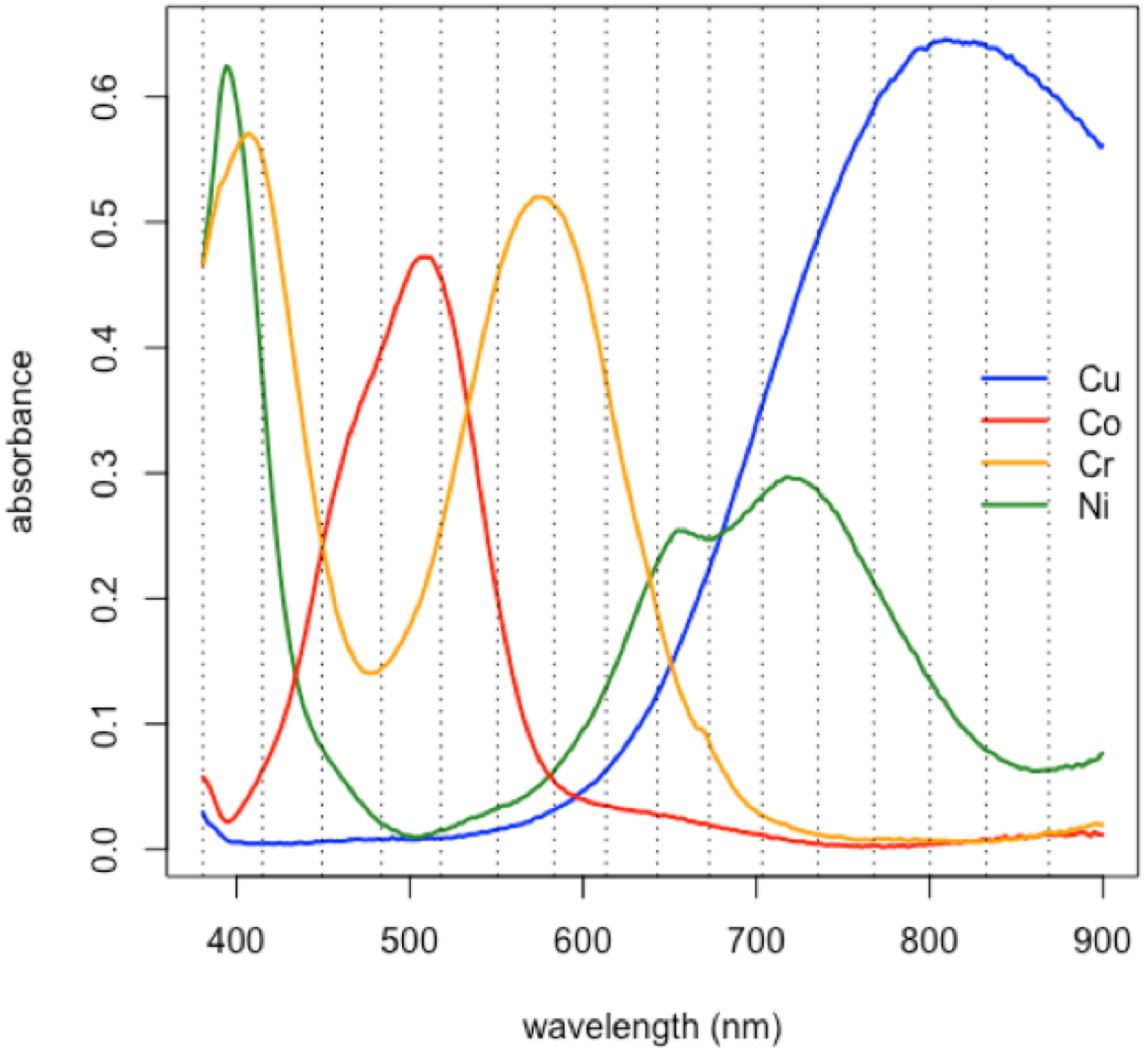
Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

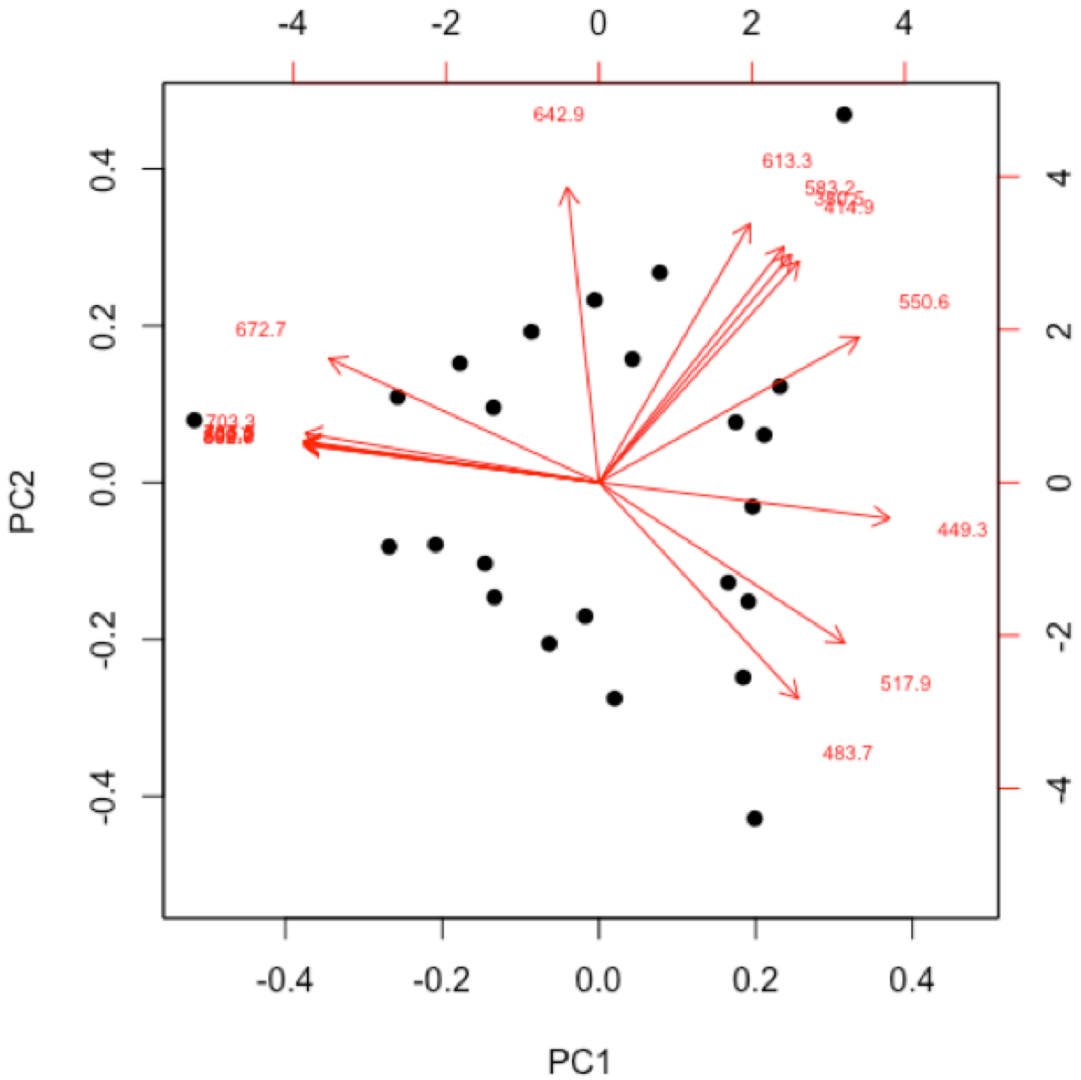
1. choose a subset of the original 635 wavelengths
2. perform PCA and determine relative importance of the 16 principal components
3. examine and interpret scores for first two principal components
4. **examine and interpret biplot of loadings and scores for the first two principal components**



ℝ functions: biplot( )

PCA: Worked Example





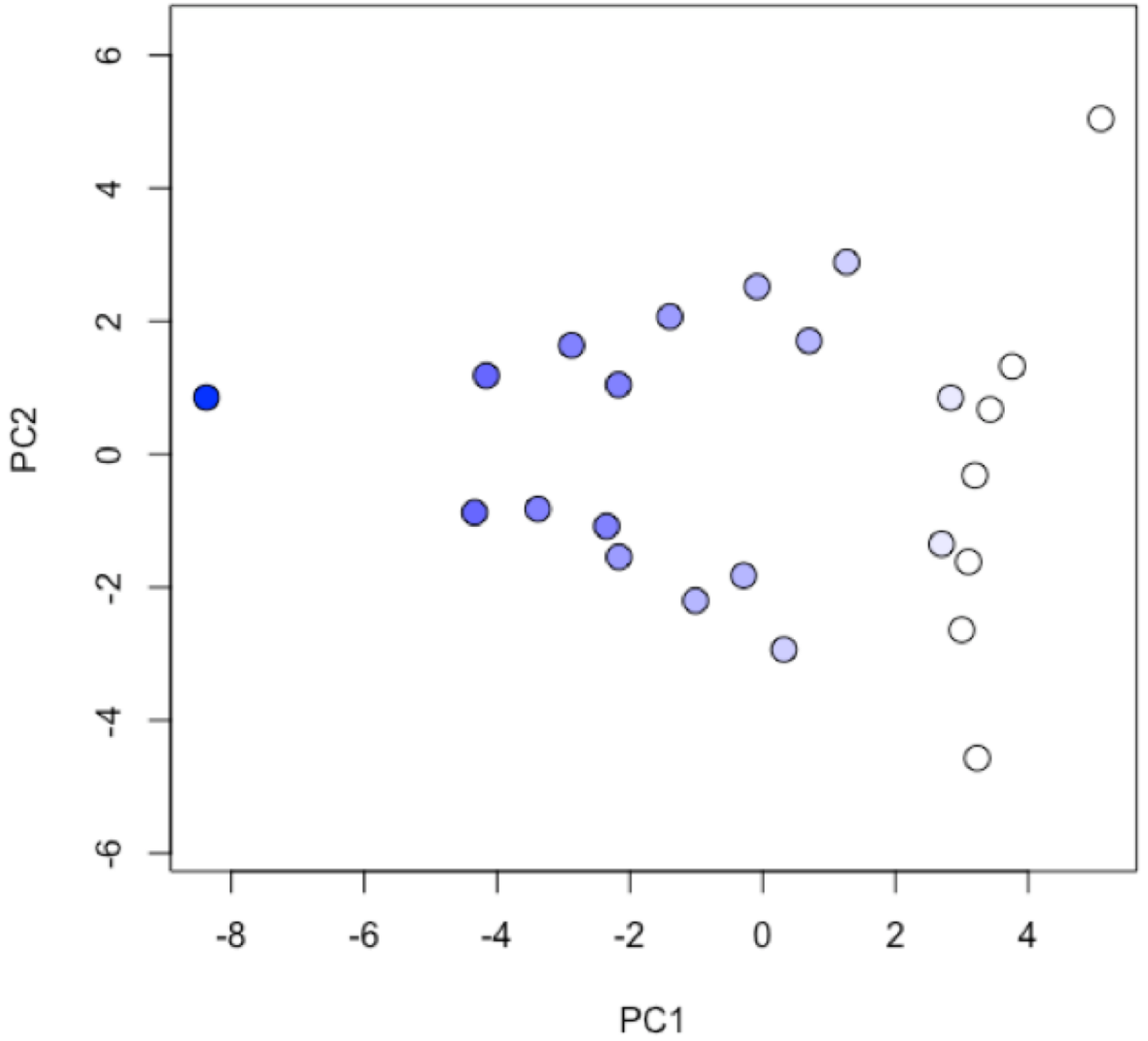
http://bit.ly/2j10Hai

PCA: Worked Example

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. choose a subset of the original 635 wavelengths
2. perform PCA and determine relative importance of the 16 principal components
3. examine and interpret scores for first two principal components
4. examine and interpret biplot of loadings and scores for the first two principal components



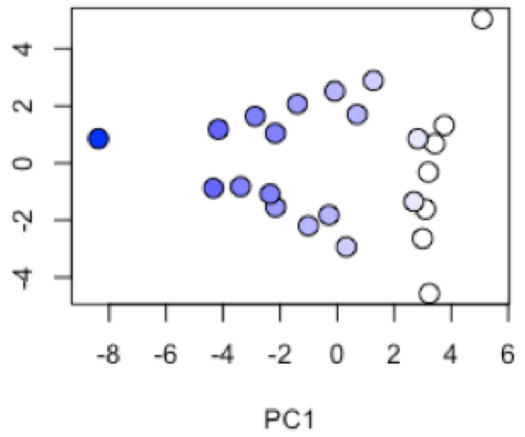
R functions: plot( ), colorRampPalatte( ), as.numeric( ), cut( )

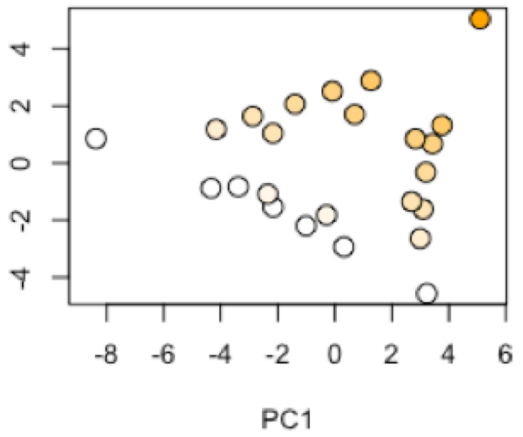
PCA: Worked Example

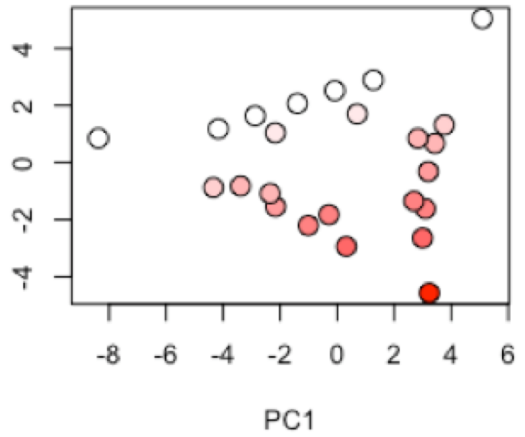
$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} + [RE]_{s \times w}$$

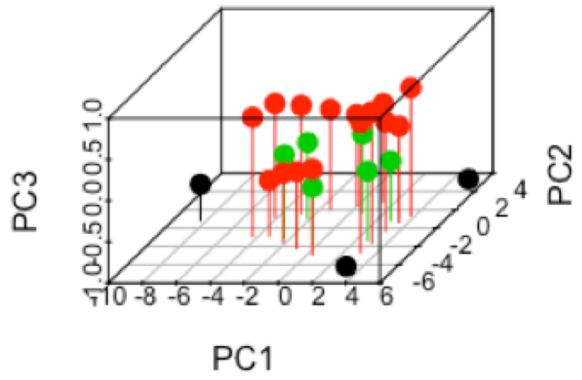
Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. choose a subset of the original 635 wavelengths
2. perform PCA and determine relative importance of the 16 principal components
3. examine and interpret scores for first two principal components
4. examine and interpret biplot of loadings and scores for the first two principal components





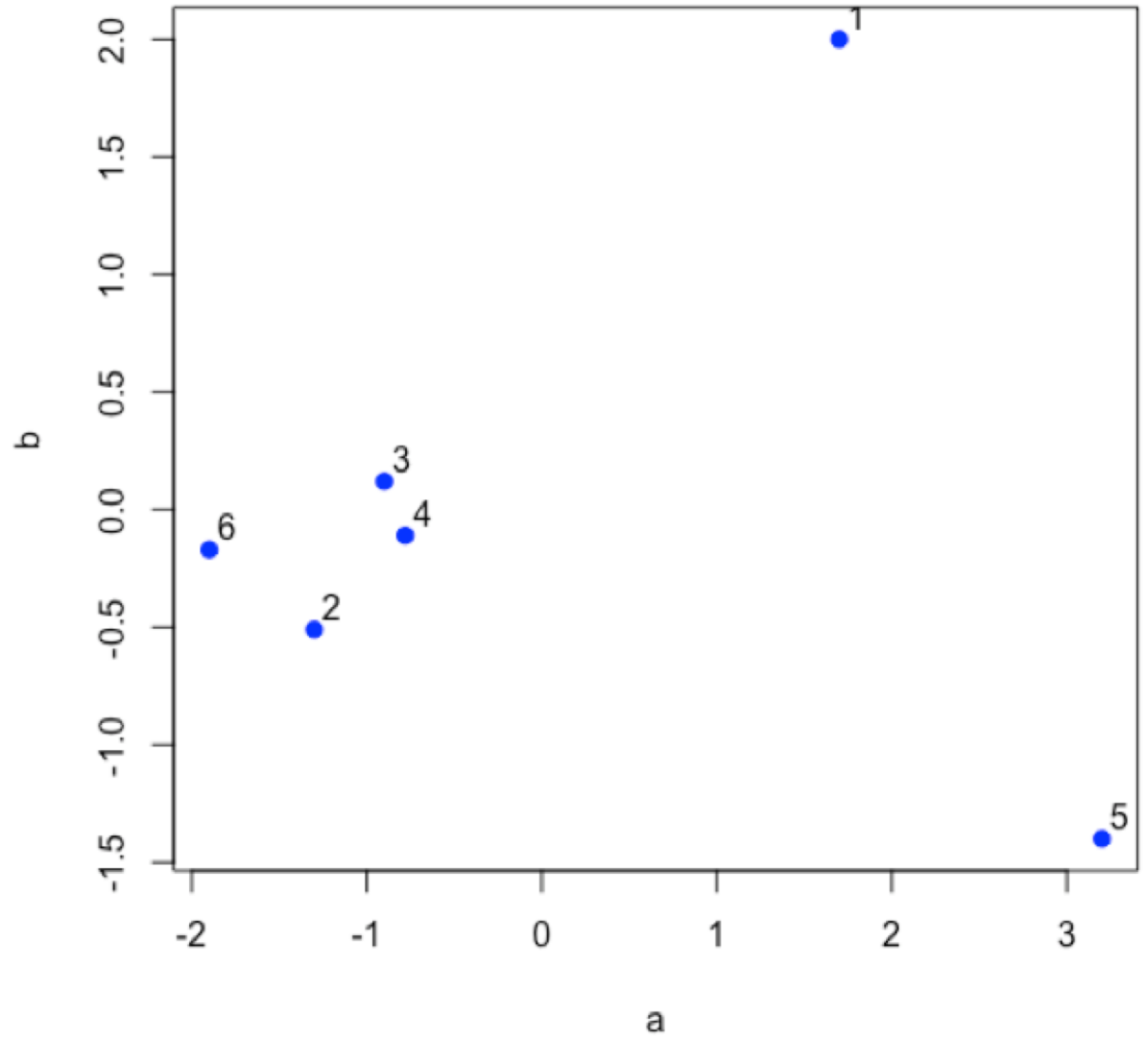




ℝ functions: par( ), plot( ), scatterplot3d::scatterplot3d

How Does Cluster Analysis Work?

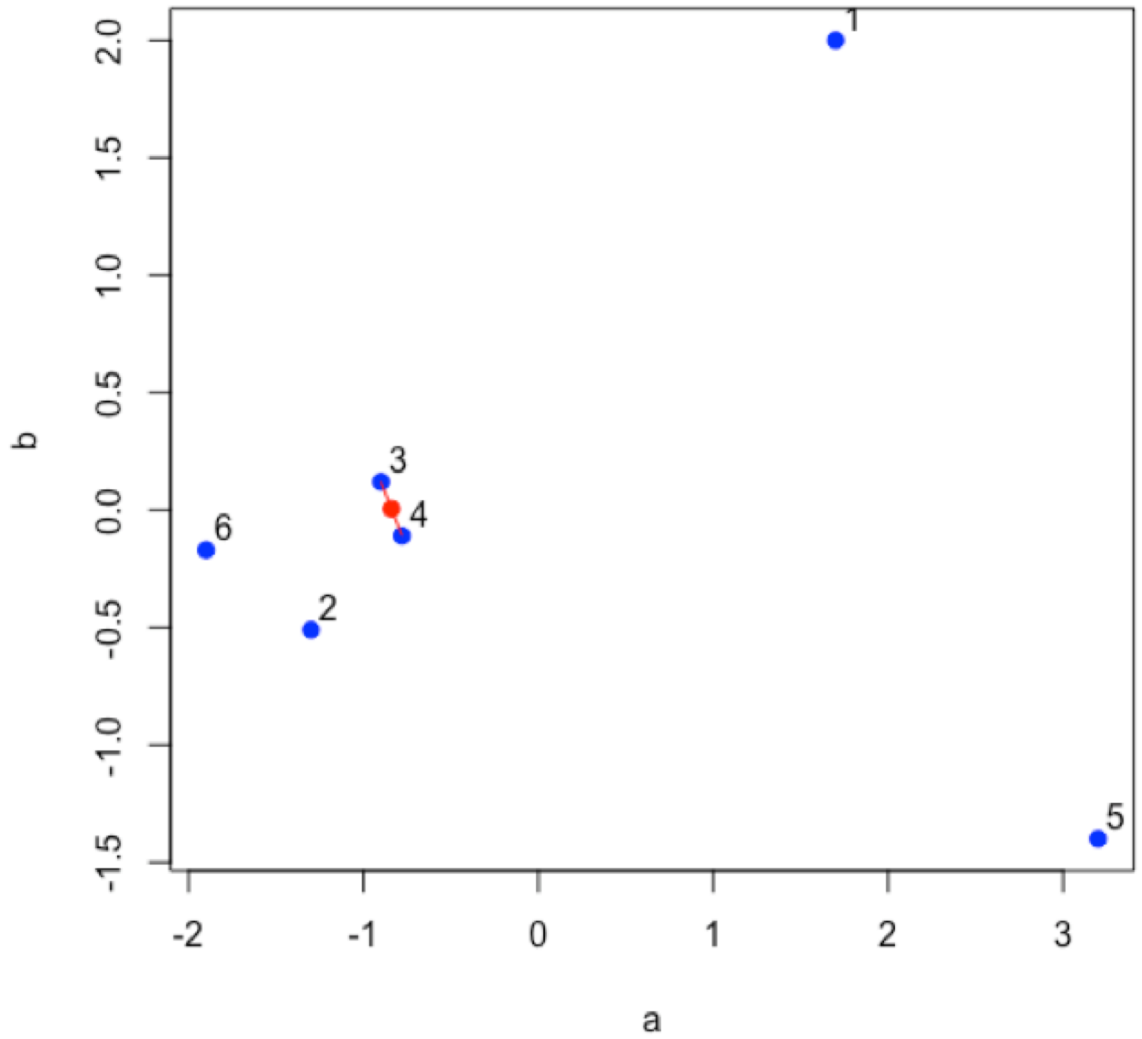
Suppose we have six samples and that we measure two properties—a and b—for each sample and create a scatterplot of the data.



R functions: plot( )

How Does Cluster Analysis Work?

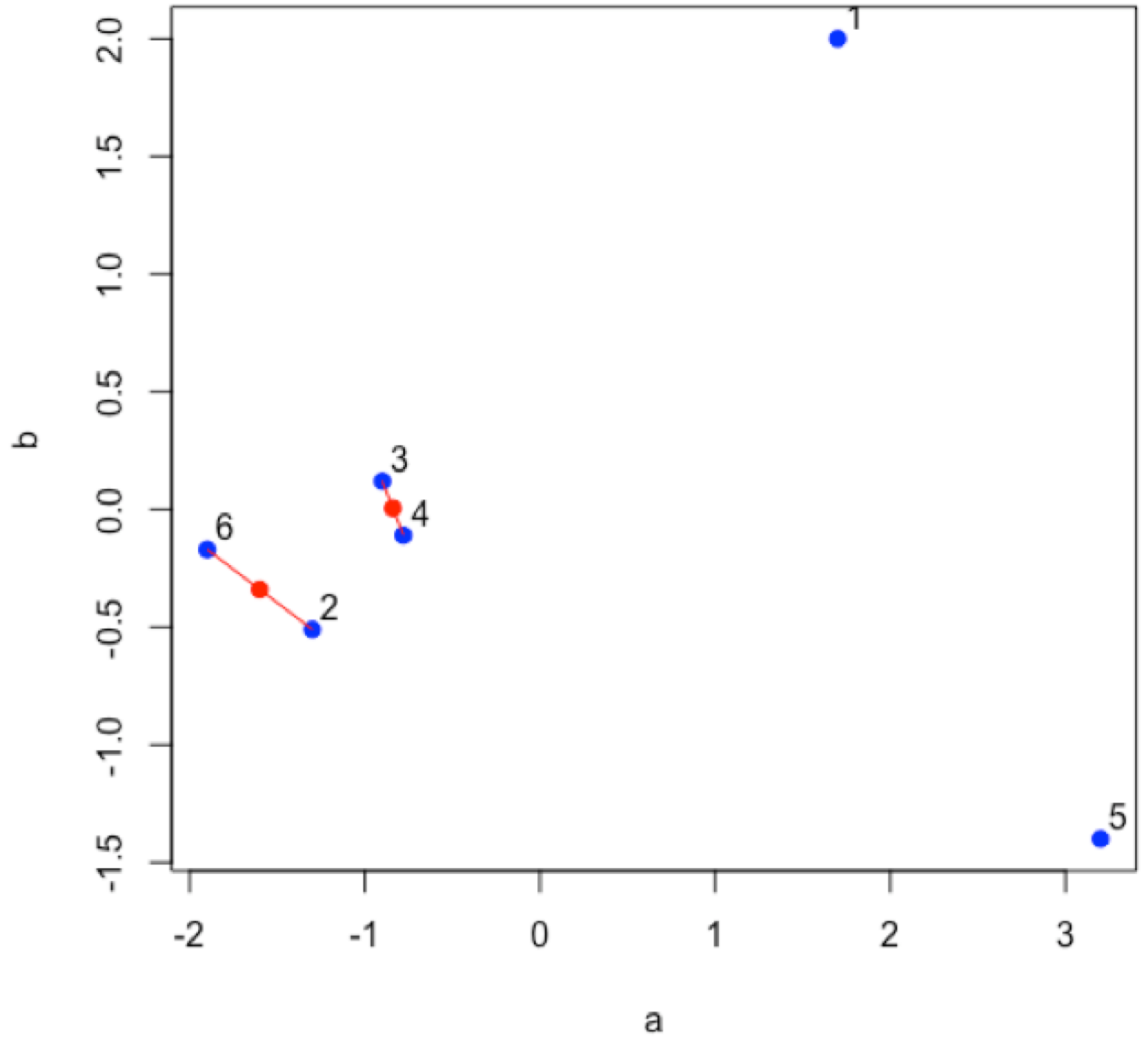
Find the two points that are closest to each other; this is the first cluster. Find the midpoint between the two points and define it as the position of the first cluster.



R functions: plot( ), segments( ), points( )

How Does Cluster Analysis Work?

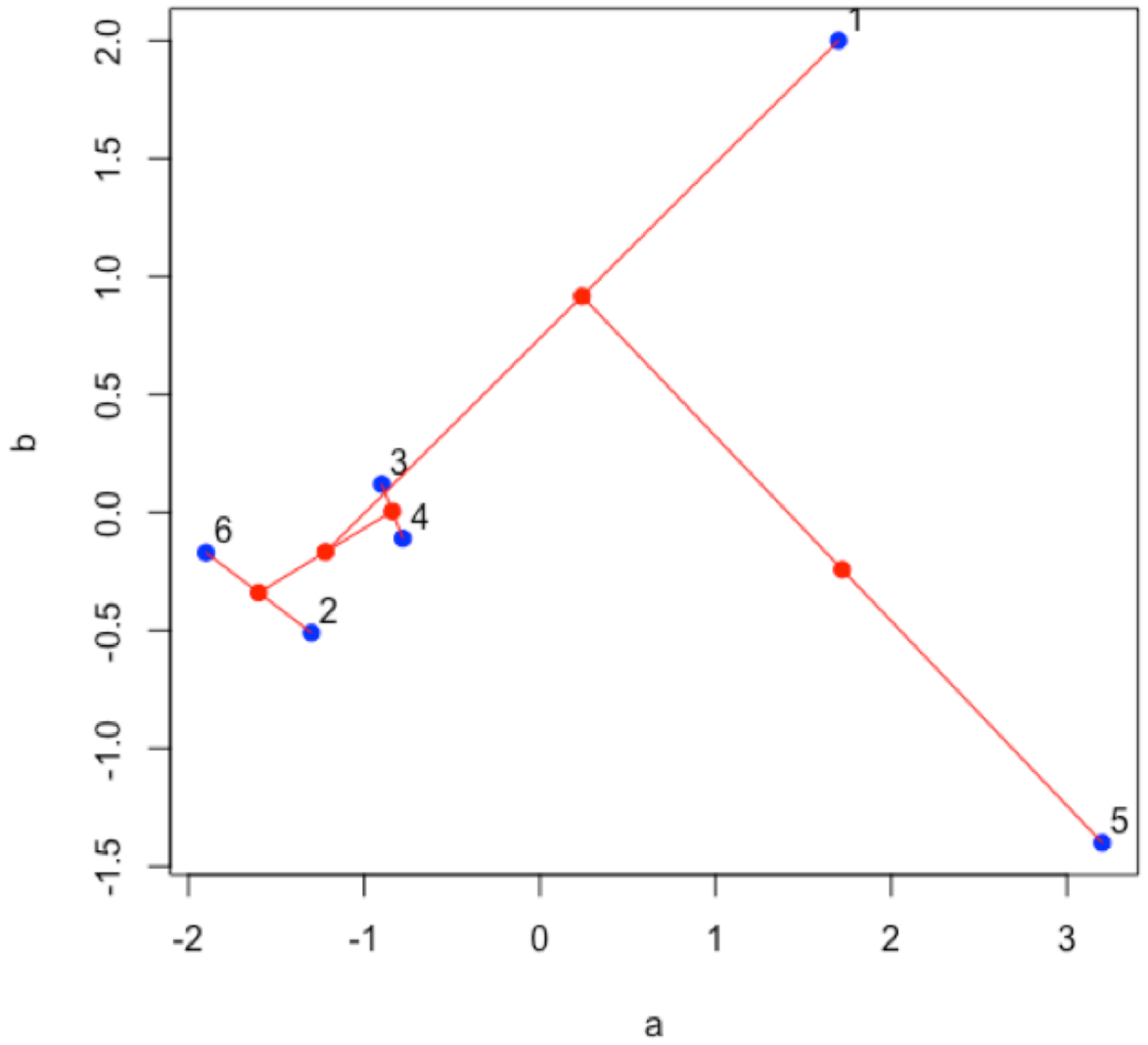
Find the next two points that are closest to each other; this is the second cluster. Find the midpoint between the two points and define it as the position of the second cluster.



R functions: plot( ), segments( ), points( )

How Does Cluster Analysis Work?

Continue until all of the original data points are included in a cluster.

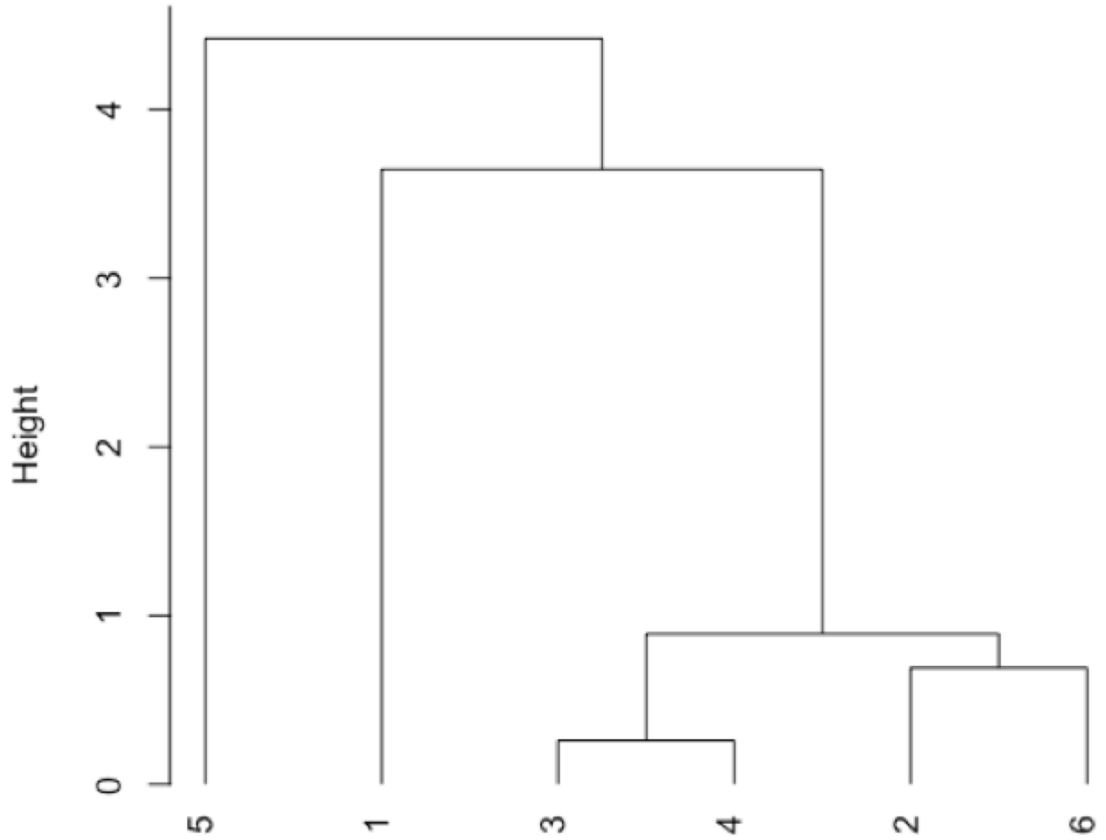


R functions: plot( ), segments( ), points( )

http://bit.ly/2j10Hai

How Does Cluster Analysis Work?

Plot a dendrogram, which shows the connectivity between points and clusters of points in terms of the distance (heights) separating them.



R functions: dist( ), hclust( ), plot( )

http://bit.ly/2j10Hai

Cluster Analysis: Worked Example

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. **calculate the distance between the individual data points using one of the available methods**

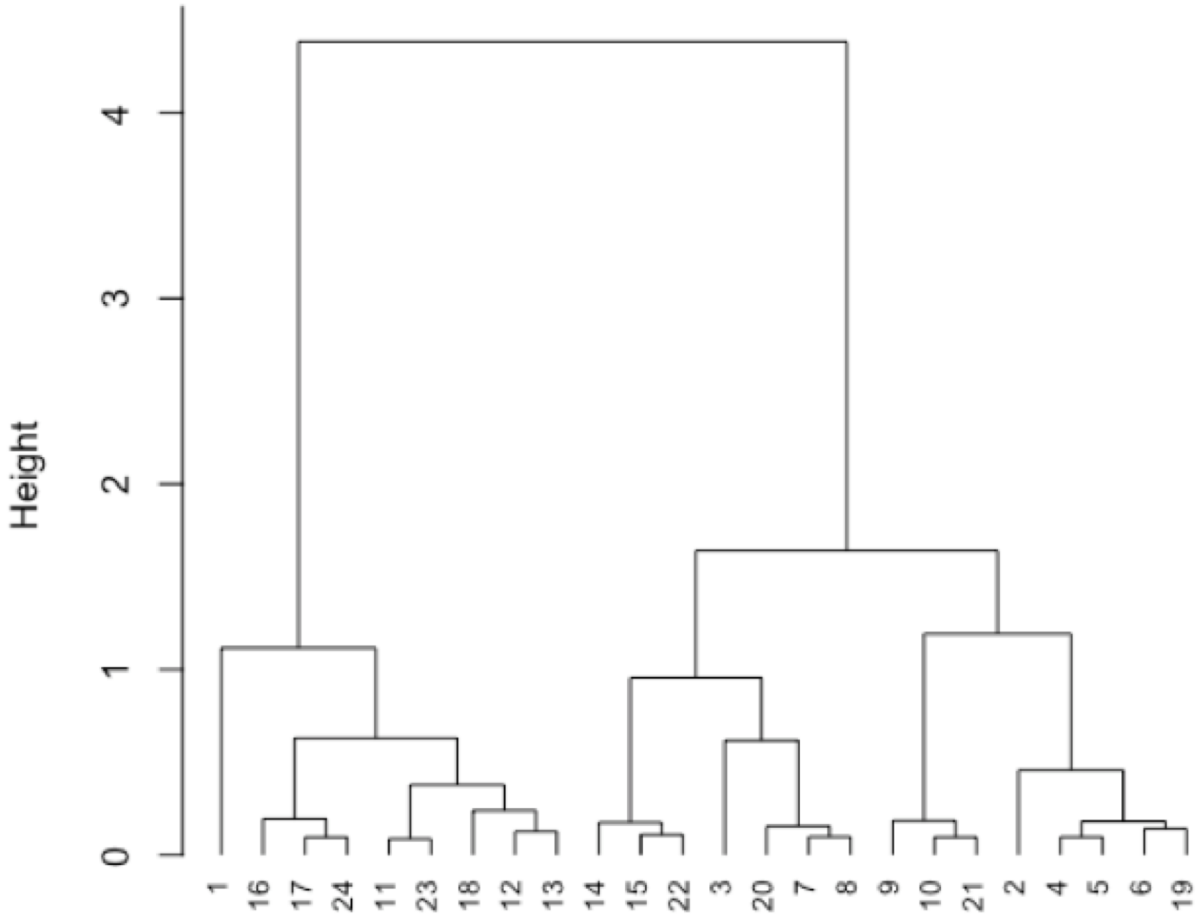
	1	2	3	4	5
2	1.53328104				
3	1.73128979	0.96493008			
4	1.48359716	0.24997370	0.77766228		
5	1.49208058	0.32863786	0.68852029	0.09664215	
6	1.49457333	0.42903074	0.57495499	0.21089686	0.11755129

R functions: dist( )

Cluster Analysis: Worked Example

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. calculate the distance between the individual data points using one of the available methods
2. **identify clusters and calculate and plot the heights between them using one of the available methods**

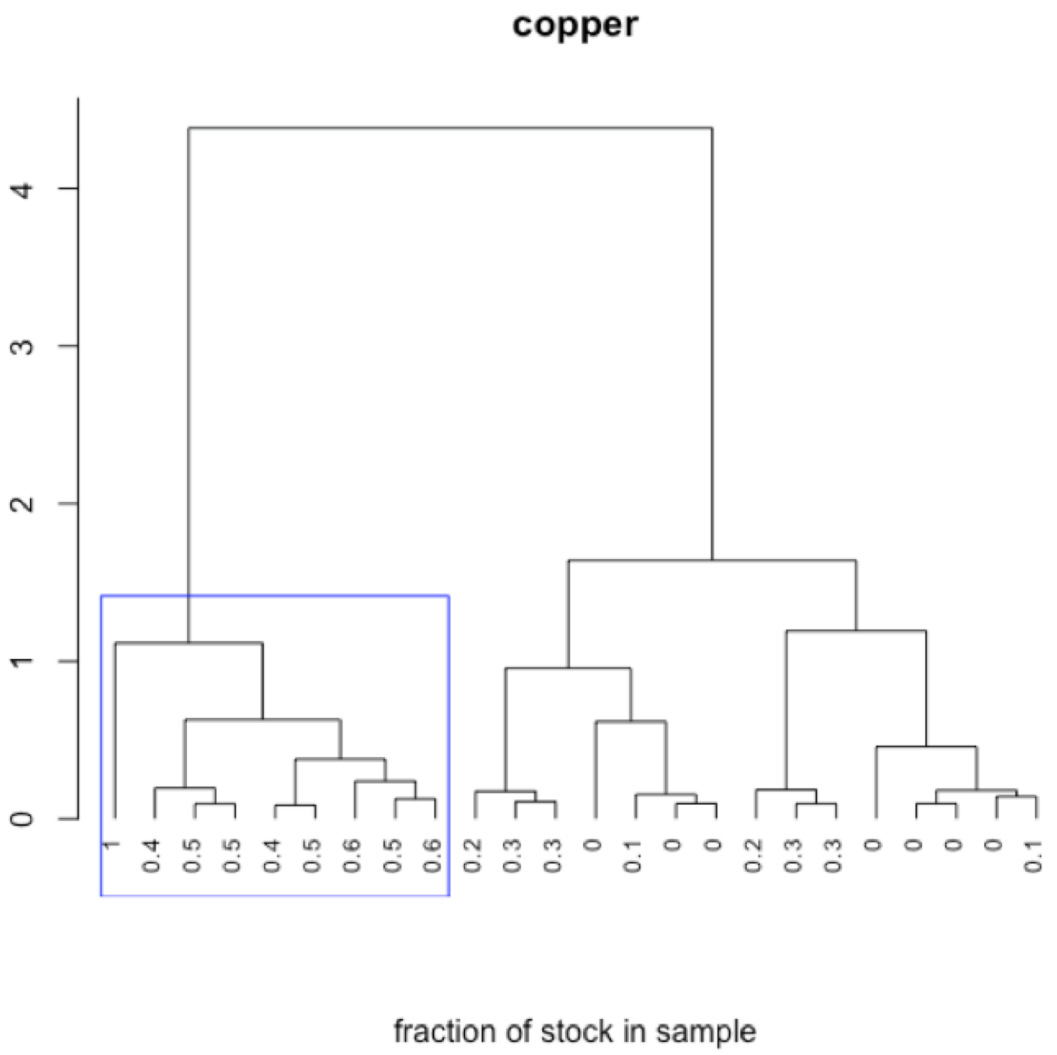


R functions: hclust( ), plot( )

Cluster Analysis: Worked Example

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. calculate the distance between the individual data points using one of the available methods
2. identify clusters and calculate and plot the heights between them using one of the available methods
3. **interpret dendrogram**

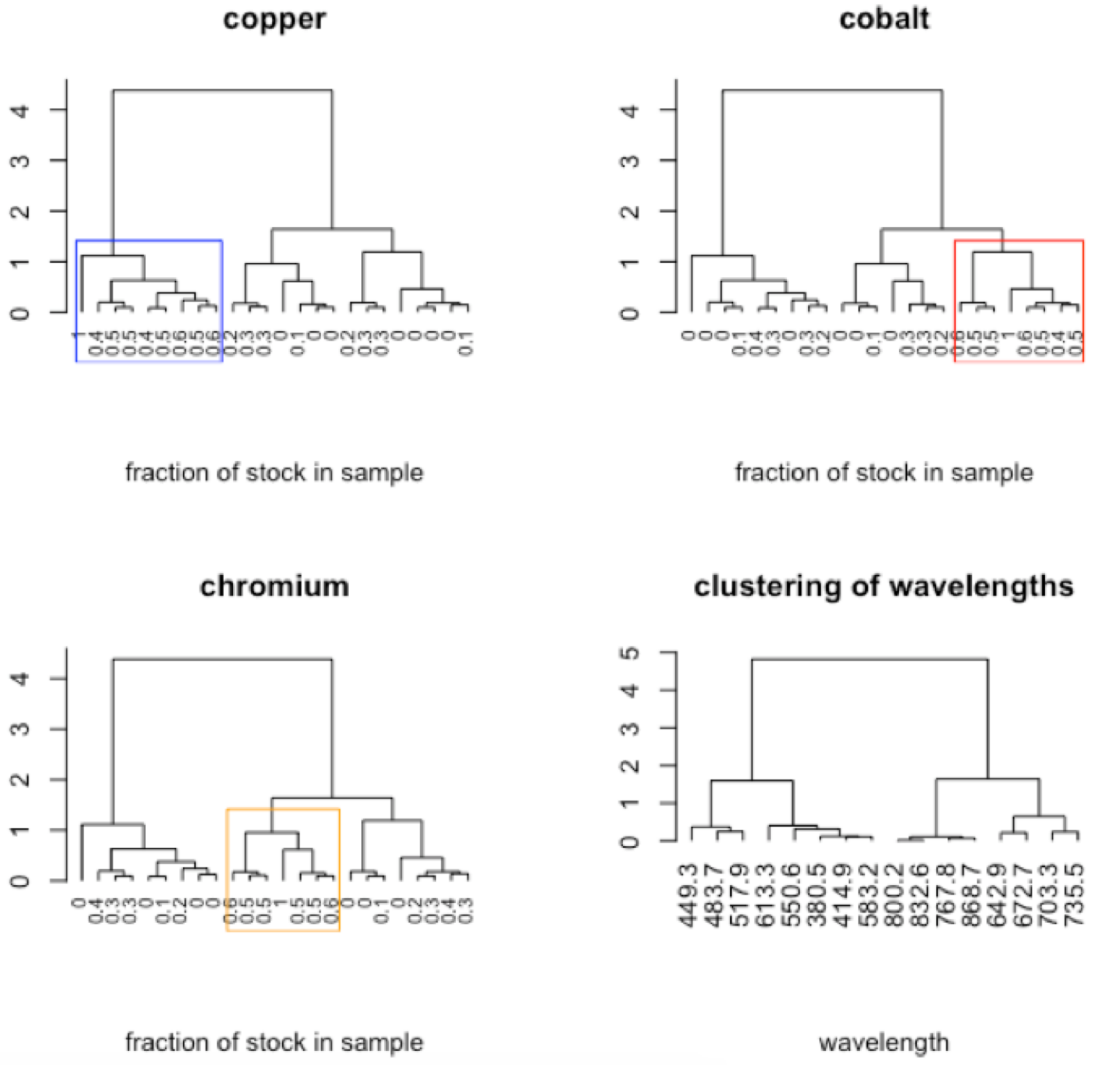


R functions: plot( ), rect.hclust( )

Cluster Analysis: Worked Example

Subset of data consisting of 24 of the 80 samples: stock Cu, stock Co, stock Cr, five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. calculate the distance between the individual data points using one of the available methods
2. identify clusters and calculate and plot the heights between them using one of the available methods
3. interpret dendrogram



ℝ functions: par( ), plot( ), rect.hclust( ), t( )

http://bit.ly/2j10Hai

How Does MLR Work?

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w}$$

Suppose we measure the absorbance at ***W*** wavelengths for ***S*** individual standard solutions with known concentrations for each of the ***N*** analytes. We can use these to determine a matrix of εb values for each analyte at each wavelength.

$$[C]^{T}_{n \times s} \times [A]_{s \times w} = [C]^{T}_{n \times s} \times [C]_{s \times n} \times [\varepsilon b]_{n \times w}$$

$$([C]^{T}_{n \times s} \times [C]_{s \times n})^{-1} \times [C]^{T}_{n \times c} \times [A]_{s \times w} = ([C]^{T}_{n \times n} \times [C]_{s \times n})^{-1} \times [C]^{T}_{n \times s} \times [C]_{s \times n} \times [\varepsilon b]_{n \times w}$$

$$([C]^{T}_{s \times n} \times [C]_{s \times n})^{-1} \times [C]^{T}_{s \times n} \times [A]_{s \times w} = [\varepsilon b]_{n \times w}$$

How Does MLR Work?

$$[A]_{s \times w} = [C]_{s \times n} \times [\varepsilon b]_{n \times w}$$

Having found the εb matrix, we can use it to calculate the concentrations for each of the ***N*** analytes in ***S*** samples given the absorbance for each sample at each wavelength.

$$[A]_{s \times w} \times [\varepsilon b]^{T}_{w \times n} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} \times [\varepsilon b]^{T}_{w \times n}$$

$$[A]_{s \times w} \times [\varepsilon b]^{T}_{w \times n} \times ([\varepsilon b]_{n \times w} \times [\varepsilon b]^{T}_{w \times n})^{-1} = [C]_{s \times n} \times [\varepsilon b]_{n \times w} \times [\varepsilon b]^{T}_{w \times n} \times ([\varepsilon b]_{n \times w} \times [\varepsilon b]^{T}_{w \times n})^{-1}$$

$$[A]_{s \times w} \times [\varepsilon b]^{T}_{w \times n} \times ([\varepsilon b]_{n \times w} \times [\varepsilon b]^{T}_{w \times n})^{-1} = [C]_{s \times n}$$

MLR: Worked Example

Standards are subset of data consisting of 15 of the 80 samples: five each prepared from stock Cu, stock Co, and stock Cr. Samples are subset of data consisting of 21 of the 80 samples: five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. **use absorbance values for a set of standards to calculate the εb values**

	380.5	414.9	449.3	483.7	517.9	550.6	583.2	613.3
concCu	0.5484511	0.1086153	0.1340763	0.1556545	0.1947192	0.3612272	0.6875421	1.3197158
concCo	0.7117778	0.7918523	2.5371205	4.0549583	4.5779242	2.0489508	0.5975168	0.3914665
concCr	13.2668054	15.1576056	6.9958232	4.0685312	6.7662738	12.0692592	13.6134665	9.8289364

R functions: findeb( )* * script written for this purpose

MLR: Worked Example

Standards are subset of data consisting of 15 of the 80 samples: five each prepared from stock Cu, stock Co, and stock Cr. Samples are subset of data consisting of 21 of the 80 samples: five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. use absorbance values for a set of standards to calculate the εb values
2. **use absorbance values for the samples and the calculated εb values to give the predicted concentrations of the analytes**

predicted concentrations of analytes

	concCu	concCo	concCr
[1,]	-0.00024	0.05991	0.00696
[2,]	-0.00037	0.04939	0.01050
[3,]	0.00036	0.03926	0.01488
[4,]	0.00075	0.03088	0.01879
[5,]	-0.00031	0.01947	0.02227
[6,]	0.01040	0.06076	0.00079

R functions: findconc( )* * script written for this purpose

MLR: Worked Example

Standards are subset of data consisting of 15 of the 80 samples: five each prepared from stock Cu, stock Co, and stock Cr. Samples are subset of data consisting of 21 of the 80 samples: five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. use absorbance values for a set of standards to calculate the εb values
2. use absorbance values for the samples and the calculated εb values to give the predicted concentrations of the analytes
3. **compare predicted and actual concentrations**

predicted concentrations of analytes

	concCu	concCo	concCr
[1,]	-0.00024	0.05991	0.00696
[2,]	-0.00037	0.04939	0.01050
[3,]	0.00036	0.03926	0.01488
[4,]	0.00075	0.03088	0.01879
[5,]	-0.00031	0.01947	0.02227
[6,]	0.01040	0.06076	0.00079

actual concentrations of analytes

	concCu	concCo	concCr
[1,]	0.000	0.06	0.00750
[2,]	0.000	0.05	0.01125
[3,]	0.000	0.04	0.01500
[4,]	0.000	0.03	0.01875
[5,]	0.000	0.02	0.02250
[6,]	0.010	0.06	0.00000

R functions: as.matrix( ), data.frame( )

MLR: Worked Example

Standards are subset of data consisting of 15 of the 80 samples: five each prepared from stock Cu, stock Co, and stock Cr. Samples are subset of data consisting of 21 of the 80 samples: five Cu/Co binary mixtures, five Cu/Cr binary mixtures, five Co/Cr binary mixtures, six Cu/Co/Cr ternary mixtures.

1. use absorbance values for a set of standards to calculate the εb values
2. use absorbance values for the samples and the calculated εb values to give the predicted concentrations of the analytes
3. compare predicted and actual concentrations
4. **report mean errors, standard deviations for errors, confidence intervals for errors, and identify maximum error for each analyte**

concCu	concCo	concCr
-0.000304	-0.000199	-0.000315

concCu	concCo	concCr
0.001102	0.000857	0.000662

concCu	concCo	concCr
±0.002298	±0.001787	±0.001381

Cu: 0.00219 (0.02719 vs. 0.0250)
Co: 0.00176 (0.05176 vs. 0.0500)
Cr: 0.00173 (0.00173 vs. 0)*

* **exceeds 95% confidence interval**

R functions: as.matrix( ), data.frame( ), apply( ), which.max( ), abs( ), round( )

Acknowledgments

- students in Chem 351
- Brian Saulnier (DePauw Chemistry major, class of 2018)
- DePauw University's Faculty Development Committee
- stackoverflow