**DATEI: *metabolite\_arafield4\_overview.R***

* Laden der Original-Tabellen Metabolite\_AraField4.txt
* (Ranking des Anthocyan-Scores)
* Ranking des BBCH-Scores
* Normalisierung der BBCH-Scores
  + anhand der Check-Kultivare C24 und Col-0 über verschiedene Feldversuche
* Unterteilung des BBCH in 3 Kategorien, von 0 bis 49: vegetativ, von 50 bis 59: Blütenanlagen, ab 60: Blüte
* Negative Werte in sehr kleine postive Werte umwandeln

nonNeg4=as.data.frame(apply(orig.df4, 2, function(col)

{ min.val=min(col[col>0])

col[col<0]=(min.val/10)

col # Column index }))

* + Feld2:
  + Feld3: mehrere Werte für Fumarat und Malat
  + Feld4: eigentlich nur ein negativer Wert bei Fumarat
* Berechnung fehlender Messwerte
  + Feld3: für Fumarat (wahrscheinlich Problem bei Messung aufgetreten) mittels pca (pcaMethods)
  + Feld2: für Saccharose mittels pca (pcaMethods)
* Erstellen einer Kategorie für blühende und nicht-blühende Pflanzen
* Erstellen einer Kategorie für vegetative und nicht-vegetative Pflanzen
  + Feld4: keine vegetativen Pflanzen
* Associationplot 🡪 *Arafield4\_Assocplot.pdf*

assocplot(table(metabolite\_arafield4$antho\_value, metabolite\_arafield4$bbch\_categories))



* Paarweiser

pairwise.prop.test(table(metabolite\_arafield4$antho\_value, metabolite\_arafield4$bbch\_categories))

* Fisher-Test

fisher.test(table(metabolite\_arafield4$antho\_value, metabolite\_arafield4$bbch\_categories))

* Berechnung der Akklimatisierungskapazität als Differenz von LT50ACC und LT50NAC

**DATEI: *metabolite\_arafield4\_transformationen.R***

* Feld3: Extremwerte aus dem Datensatz herausnehmen: Glc>35, Frc>30, Suc>20, Mal>30, Fum>5 (sonst sind keine Normalverteilungen möglich)
* Histogramme und Shapiro-Test 🡪 *Metabolite\_AraField4\_Histogramme.pdf*
* Box-Cox-Transformationen und Box-Cox-Diagramme 🡪 resultierende Histogramme bzw. Q-Q-Plot
  + *Metabolite\_AraField4\_Transformationen.pdf*
  + *Metabolite\_AraField4\_boxcox.pdf*
* Hinzufügen der transformierten Metabolit-Daten zum Gesamtdatensatz
* write.table(metabolite\_arafield4, file="Masterarbeit/Ergebnisse/arafield4.txt", sep="\t")
* *arafield4.txt*

**DATEI: *metabolite\_arafield4\_plots.R***

* Boxplots
  + der Metabolite und LT50 für gesamte Proben (einzeln)
  + der Metabolite und LT50 aufgeteilt nach BBCH Kategorien
  + der Metabolite und LT50 aufgeteilt nach Flowering state
  + der Metabolite und LT50 aufgeteilt nach vegetative state
  + der Metabolite und LT50 aufgeteilt nach Anthocyan-Scores
  + der Metabolite und LT50 aufgeteilt nach BBCH Kategorien sowie nach Anthocyan-Scores

🡪 *Metabolite\_AraField4\_bwplots.pdf*

* Pairs Plot 🡪 *Metabolite\_AraField4\_Pairs.pdf*

**DATEI: *metabolite\_arafield4\_models***

* Auswahl aller Beobachtungen, die keine NAs in der Spalte mit LT50 und Anthoscore enthalten 🡪 idx\_LT50\_4
* Dataframe für Regressionsmodell für LT50 (ACC bzw. NAC) 🡪 transformierte Daten
  + write.table(arafield4\_regressiondata, file="Masterarbeit/Ergebnisse/arafield4\_regressiondata.txt", sep="\t")
  + arafield4\_regressiondata
* erstes Modell ohne weitere Veränderungen mit allen Parametern 🡪 **arafield4\_regressionACC**
  + glc, frc, suc, raf, aa, mal, fum, pro, antho, norm\_bbch, anthoscore
* Kreuzvalidierung 🡪 CVlm
* Plot der gemessenen LT50ACC vs. Vorhergesagte LT50ACC 🡪 *Metabolite\_AraField4\_ACCmodelplot.pdf*
* Datensatz für die Akzessionen zur Vorhersage definieren: arafield4\_predictiondata
* Vorhersage der LT50-Werte mit Modell zur Validierung
* Verwendung der kategorialen Variablen anthoscore und norm\_bbch als **Faktoren**!

**🡪 arafield4\_regressionACC\_factor**

**Modell-Selektion**

* Cp-Plot
* stepAIC für Modell mit numerischen Variablen und Modell mit Faktoren

**DATEI: *metabolite\_arafield4\_leaps***

* Verwendung von leaps zur Variablenselektion und Generierung von Abbildungen
* Verwendung der verschiedenen Kriterien: BIC, Radj², R², Cp