Lab1: RMarkdown

Lara Katz

2024-01-19

library(terra)  
library(ggplot2)  
library(corrplot)

# Read in multivariate data set

bds <- read.csv("./Data/Katz\_BDS\_data.csv")  
bds <- bds[,-1] # remove first column

# Use the head function to show the first few rows of the dataset

head(bds)

## pab ID catchment clay coast elev lith marine ph X01\_water  
## 1 1 1 1 63.32500 23.88172 46.83685 11 1 47.40862 0.2698153  
## 2 1 2 3 88.93841 14.55324 46.41621 13 1 46.98804 4.8307953  
## 3 1 3 3 88.93841 14.12124 46.41621 13 1 46.98804 4.8307953  
## 4 1 4 3 167.59267 12.29321 28.18165 14 1 48.33503 1.5440967  
## 5 1 5 3 167.59267 16.95244 28.18165 14 1 48.33503 1.5440967  
## 6 1 6 3 74.02357 23.47776 46.53899 11 1 46.68098 1.7269258  
## X02\_Barren X03\_LAc\_NHardwd X05\_NAtl\_CoastPlain\_Hardwd X07\_LAc\_NPine.Oak  
## 1 0 0.00000000 0.00000000 0.26889712  
## 2 0 0.00000000 0.01054837 0.27736539  
## 3 0 0.00000000 0.01054837 0.27736539  
## 4 0 0.01619243 0.42797562 0.06868713  
## 5 0 0.01619243 0.42797562 0.06868713  
## 6 0 0.00000000 0.25053707 0.03955692  
## X08\_LAc\_Pine.Hemlock.Hardwd X09\_CAp\_Dry\_Oak.Pine X10\_Ap\_Hemlock.N\_Hardwood  
## 1 0.0000000000 76.39704 4.531043  
## 2 0.0005120558 62.91043 6.073166  
## 3 0.0005120558 62.91043 6.073166  
## 4 0.0099962549 52.63204 12.638550  
## 5 0.0099962549 52.63204 12.638550  
## 6 0.0000000000 68.32129 6.403162  
## X11\_Ac\_Low.Elev\_Spruce.Fir.Hardwood X12\_AcAp\_Montane\_Spruce.Fir  
## 1 0 0  
## 2 0 0  
## 3 0 0  
## 4 0 0  
## 5 0 0  
## 6 0 0  
## X13\_CentAp\_Pine.Oak\_Rocky\_Wd X14\_NAtl\_Coast\_Plain\_Maritime  
## 1 0.09100428 0  
## 2 0.18215849 0  
## 3 0.18215849 0  
## 4 0.66283411 0  
## 5 0.66283411 0  
## 6 0.01318510 0  
## X15\_NAtl\_Coast\_Plain\_Dune X16\_CentIntAp\_FloodplainSys  
## 1 0 5.861996  
## 2 0 8.068892  
## 3 0 8.068892  
## 4 0 1.664302  
## 5 0 1.664302  
## 6 0 3.784316  
## X17\_CentIntAp\_RiparianSys X18\_LAc\_FloodplainSys X19\_Bor\_Acidic\_PeatSys  
## 1 0.0000000 0 0.00000000  
## 2 0.0158223 0 0.11075780  
## 3 0.0158223 0 0.11075780  
## 4 0.0000000 0 0.03238561  
## 5 0.0000000 0 0.03238561  
## 6 0.0000000 0 0.00000000  
## X20\_CentIntAp\_SwampSys X21\_GulfAtl\_CoastPlain\_SwampSys  
## 1 7.366721 0.01075615  
## 2 13.477438 0.01582230  
## 3 13.477438 0.01582230  
## 4 18.614346 0.00000000  
## 5 18.614346 0.00000000  
## 6 13.890319 0.00000000  
## X22\_GulfAtl\_CoastPlain\_TMarshSys X23\_LAc\_Shrub.Herb\_WetlSys  
## 1 0.00000000 2.380809  
## 2 0.01582256 2.096747  
## 3 0.01582256 2.096747  
## 4 0.00000000 3.753188  
## 5 0.00000000 3.753188  
## 6 0.00000000 3.715234  
## X24\_NCentInt\_Wet\_Flatwd X25\_LAc\_SwampSys X26\_NeInt\_PineBarrens  
## 1 2.821912 0 0  
## 2 1.913718 0 0  
## 3 1.913718 0 0  
## 4 7.935406 0 0  
## 5 7.935406 0 0  
## 6 1.855471 0 0  
## X27\_AcAp\_WdHeath.Krummholz X28\_Bor\_JackPine.BlackSpruce X29\_AcAp\_AlpineTundra  
## 1 0 0 0  
## 2 0 0 0  
## 3 0 0 0  
## 4 0 0 0  
## 5 0 0 0  
## 6 0 0 0  
## sand silt slope  
## 1 606.5103 330.1371 3.106801  
## 2 561.0895 349.9612 2.958915  
## 3 561.0895 349.9612 2.958915  
## 4 417.9165 414.4939 2.663899  
## 5 417.9165 414.4939 2.663899  
## 6 598.0278 327.9486 2.491327

# Bulleted list describing the columns of the dataset

Columns of dataset:

* **pab** = Presence/absence/pseudoabsence data (binary)
* **catchment**: Catchment position (stream size or size of largest input stream into lake/pond) (categorical data)
* **clay**: Mean proportion clay composition of soil within point’s HUC12 subwatershed (continuous data)
* **coast**: Distance from the coastline in km (continuous data)
* **elev**: Mean elevation (m) within point’s HUC12 subwatershed (continuous data)
* **lith**: Lithology at point (categorical data)
* **marine**: Whether or not point is within the historic marine limit (binary data)
* **ph**: Mean pH of soil within point’s HUC12 subwatershed (continuous data)
* Columns **X01\_water** through **X29\_AcAp\_AlpineTundra**: Percent cover of the LANDFIRE landcover type within the point’s HUC12 subwatershed (continuous data)
  + **X01\_water**: Open Water
  + **X02\_Barren**: Barren-Rock/Sand/Clay
  + **X03\_LAc\_NHardwd**: Laurentian-Acadian Northern Hardwoods Forest
  + **X05\_NAtl\_CoastPlain\_Hardwd**: Northern Atlantic Coastal Plain Hardwood Forest
  + **X07\_LAc\_NPine.Oak**: Laurentian-Acadian Northern Pine(-Oak) Forest
  + **X08\_LAc\_Pine.Hemlock.Hardwd**: Laurentian-Acadian Pine-Hemlock-Hardwood Forest
  + **X09\_CAp\_Dry\_Oak.Pine**: Central Appalachian Dry Oak-Pine Forest
  + **X10\_Ap\_Hemlock.N\_Hardwood**: Appalachian (Hemlock-)Northern Hardwood Forest
  + **X11\_Ac\_Low.Elev\_Spruce.Fir.Hardwood**: Acadian Low-Elevation Spruce-Fir-Hardwood Forest
  + **X12\_AcAp\_Montane\_Spruce.Fir**: Acadian-Appalachian Montane Spruce-Fir Forest
  + **X13\_CentAp\_Pine.Oak\_Rocky\_Wd**: Central Appalachian Pine-Oak Rocky Woodland
  + **X14\_NAtl\_Coast\_Plain\_Maritime**: Northern Atlantic Coastal Plain Maritime Forest
  + **X15\_NAtl\_Coast\_Plain\_Dune**: Northern Atlantic Coastal Plain Dune and Swale
  + **X16\_CentIntAp\_FloodplainSys**: Central Interior and Appalachian Floodplain Systems
  + **X17\_CentIntAp\_RiparianSys**: Central Interior and Appalachian Riparian Systems
  + **X18\_LAc\_FloodplainSys**: Laurentian-Acadian Floodplain Systems
  + **X19\_Bor\_Acidic\_PeatSys**: Boreal Acidic Peatland Systems
  + **X20\_CentIntAp\_SwampSys**: Central Interior and Appalachian Swamp Systems
  + **X21\_GulfAtl\_CoastPlain\_SwampSys**: Gulf and Atlantic Coastal Plain Swamp Systems
  + **X22\_GulfAtl\_CoastPlain\_TMarshSys**: Gulf and Atlantic Coastal Plain Tidal Marsh Systems
  + **X23\_LAc\_Shrub.Herb\_WetlSys**: Laurentian-Acadian Shrub-Herbaceous Wetland Systems
  + **X24\_NCentInt\_Wet\_Flatwd**: North-Central Interior Wet Flatwoods
  + **X25\_LAc\_SwampSys**: Laurentian-Acadian Swamp Systems
  + **X26\_NeInt\_PineBarrens**: Northeastern Interior Pine Barrens
  + **X27\_AcAp\_WdHeath.Krummholz**: Acadian-Appalachian Subalpine Woodland and Heath-Krummholz
  + **X28\_Bor\_JackPine.BlackSpruce**: Boreal Jack Pine-Black Spruce Forest
  + **X29\_AcAp\_AlpineTundra**: Acadian-Appalachian Alpine Tundra
* **sand**: Mean proportion sand composition of soil within point’s HUC12 subwatershed (continuous data)
* **silt**: Mean proportion silt composition of soil within point’s HUC12 subwatershed (continuous data)
* **slope**: Mean slope (degrees) within point’s HUC12 subwatershed (continuous data)

# Plot data

Remove categorical data and rename variables:

# Check variables for multicollinearity  
data.num <- as.data.frame(bds[,-c(1,2,30,34,35)]) # remove non-continuous data  
data.num.names <- c("Clay",  
 "Coast",  
 "Elevation",  
 "pH",  
 "Water",   
 "Barren",   
 "LAc Northern Hardwoods",   
 "NAtl Coast Plain Hardwood",  
 "LAc Northern Pine(-Oak)",   
 "LAc Pine-Hemlock-Hardwood",  
 "CAp Dry Oak-Pine",   
 "Ap (Hemlock-)N Hardwood",  
 "Ac Low-Elev Spruce-Fir-Hardwood",  
 "AcAp Montane Spruce-Fir",  
 "CAp Pine-Oak Rocky Wd",   
 "NAtl Coast Plain Maritime",  
 "NAtl Coast Plain Dune",  
 "CentIntAp Floodplain Sys",  
 "CentIntAp Riparian Sys",  
 "LAc Floodplain Sys",   
 "Bor Acidic Peat Sys",   
 "CInt and Ap Swamp Sys",   
 "Gulf/Atl Coast Plain Swamp Sys",  
 "Gulf/Atl Coast Pl,in TMarsh Sys",  
 "LAc Shrub-Herb Wetl Sys",   
 "NCentInt Wet Flatwoods",  
 "LAc Swamp Sys",  
 "NInt Pine Barrens",   
 "AcAp Wd/Heath-Krummholz",  
 "Bor Jack Pine-Black Spruce",  
 "AcAp Alpine Tundra",   
 "Sand",  
 "Silt",  
 "Slope") # make land cover class names shorter  
colnames(data.num) <- data.num.names  
corr <- cor(data.num, method = "pearson")  
# head(corr)

Plot correlogram:

corrplot <- corrplot(corr,   
 type="lower",  
 method="color",  
 diag = FALSE,  
 order="FPC",  
 pch.col = "black",  
 pch.cex = 1,  
 tl.col="black",  
 tl.cex=1,  
 tl.srt = 45,  
 col = COL2(diverging = "RdYlBu"))

