Lab1: RMarkdown

Lara Katz

2024-01-19

```
library(terra)
```

```
## terra 1.7.29
```

```
library(ggplot2)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

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)
- Columns **X01_water** through **X29_AcAp_AlpineTundra**: Percent cover of the LANDFIRE land-cover 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

- **catch**: 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)
- **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

Read in shapefile of model extent:

```
# Check variables for multicollinearity
data.num <- as.data.frame(bds[,-c(1,2,30,34,35)]) # remove non-continuous data
data.num.names <- c("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",
                    "Clay",
                    "Coast",
                    "Elevation",
                    "pH",
                    "Sand",
                    "Silt",
                    "Slope") # make land cover class names shorter
colnames(data.num) <- data.num.names
str(data.num)
```

```
## 'data.frame':    10000 obs. of  34 variables:
##  $ Water                          : int  1 3 3 3 3 3 2 2 3 3 ...
##  $ Barren                         : num  63.3 88.9 88.9 167.6 167.6 ...
##  $ LAc Northern Hardwoods         : num  23.9 14.6 14.1 12.3 17 ...
##  $ NAtl Coast Plain Hardwood      : num  46.8 46.4 46.4 28.2 28.2 ...
##  $ LAc Northern Pine(-Oak)        : int  11 13 13 14 14 11 13 11 11 11 ...
##  $ LAc Pine-Hemlock-Hardwood      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ CAp Dry Oak-Pine               : num  47.4 47 47 48.3 48.3 ...
##  $ Ap (Hemlock-)N Hardwood        : num  0.27 4.83 4.83 1.54 1.54 ...
##  $ Ac Low-Elev Spruce-Fir-Hardwood: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ AcAp Montane Spruce-Fir        : num  0 0 0 0.0162 0.0162 ...
##  $ CAp Pine-Oak Rocky Wd          : num  0 0.0105 0.0105 0.428 0.428 ...
##  $ NAtl Coast Plain Maritime      : num  0.2689 0.2774 0.2774 0.0687 0.0687 ...
##  $ NAtl Coast Plain Dune          : num  0 0.000512 0.000512 0.009996 0.009996 ...
##  $ CentIntAp Floodplain Sys       : num  76.4 62.9 62.9 52.6 52.6 ...
##  $ CentIntAp Riparian Sys         : num  4.53 6.07 6.07 12.64 12.64 ...
##  $ LAc Floodplain Sys             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Bor Acidic Peat Sys            : num  0 0 0 0 0 ...
##  $ CInt and Ap Swamp Sys          : num  0.091 0.182 0.182 0.663 0.663 ...
##  $ Gulf/Atl Coast Plain Swamp Sys : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Gulf/Atl Coast Pl,in TMarsh Sys: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ LAc Shrub-Herb Wetl Sys        : num  5.86 8.07 8.07 1.66 1.66 ...
##  $ NCentInt Wet Flatwoods         : num  0 0.0158 0.0158 0 0 ...
##  $ LAc Swamp Sys                  : num  0 0 0 0 0 ...
##  $ NInt Pine Barrens              : num  0 0.1108 0.1108 0.0324 0.0324 ...
##  $ AcAp Wd/Heath-Krummholz        : num  7.37 13.48 13.48 18.61 18.61 ...
##  $ Bor Jack Pine-Black Spruce     : num  0.0108 0.0158 0.0158 0 0 ...
##  $ AcAp Alpine Tundra             : num  0 0.0158 0.0158 0 0 ...
##  $ Clay                           : num  2.82 1.91 1.91 7.94 7.94 ...
##  $ Coast                          : num  0 0 0 0 0 ...
##  $ Elevation                      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pH                             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Sand                           : num  607 561 561 418 418 ...
##  $ Silt                           : num  330 350 350 414 414 ...
##  $ Slope                          : num  3.11 2.96 2.96 2.66 2.66 ...
```

```
max(data.num$X01_water)
```

```
## Warning in max(data.num$X01_water): no non-missing arguments to max; returning
## -Inf
```

```
## [1] -Inf
```

```
corr <- cor(data.num, method = "pearson")
# head(corr)
```

Plot correlogram:

```
corrplot <- corrplot(corr,
                     type="lower", #shape of the plot itself: full, upper, lower
                     method="color", #shape of the data: circle, square, ellipse, number, shade, color,
                     diag = FALSE,
                     order="FPC", #how to cluster samples: AOE, hclust, FPC, alphabet, or leave blank
                     #p.mat = corrtest$p, #which correlations to use
                     #sig.level=0.05, #sets significance cutoff
                     #insig="label_sig",#leaves p > 0.05 blank
                     # addCoef.col = "black",
                     # number.cex = 0.5,
                     pch.col = "black",
                     pch.cex = 0.9,
                     tl.col="black", # text color
                     tl.cex=.7,#text size
                     tl.srt = 45,
                     col = COL2(diverging = "RdYlBu"))
```