All Countries Environmental Data

## Method: PLS-C

PLS is used to find the fundamental relations between two matrices (X and Y), i.e. a latent variable approach to modeling the covariance structures in these two spaces. A PLS model will try to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. PLS regression is particularly suited when the matrix of predictors has more variables than observations, and when there is multicollinearity among X values. PLS bears some relation to principal components regression; instead of finding hyperplanes of maximum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the X and Y data are projected to new spaces, the PLS family of methods are known as bilinear factor models.

## Dataset

* Data: Measurements of environment conditions in Countries
* Rows: There are 137 observations, 1 for each country.
* Columns: Columns: Total 2 Tables – All Rain and All Temp
* Structure of Data

str(country\_env\_df)

## 'data.frame': 137 obs. of 29 variables:  
## $ Country : Factor w/ 137 levels "Afghanistan",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Happiness\_Rank : Ord.factor w/ 3 levels "VH"<"H"<"U": 3 2 2 3 1 3 1 1 2 2 ...  
## $ accessibility\_to\_cities: num 317.7 73.8 1212.8 378.2 209.2 ...  
## $ elevation : num 1832 652 557 1061 683 ...  
## $ aspect : num 201 192 185 174 145 ...  
## $ slope : num 1.516 1.89 0.171 0.193 0.624 ...  
## $ cropland\_cover : num 9.51 23.35 3.69 2.79 21.96 ...  
## $ tree\_canopy\_cover : num 0.375 12.805 0.177 19.87 8.834 ...  
## $ isothermality : num 35.9 33.2 40.3 64.3 49.9 ...  
## $ rain\_coldestQuart : num 128.72 392.51 25.29 8.05 79.09 ...  
## $ rain\_driestMonth : num 1.722 40.088 0.935 0.26 17.183 ...  
## $ rain\_driestQuart : num 8.3 138.15 6.09 4.43 60.49 ...  
## $ rain\_mean\_annual : num 311.3 1151.1 79.5 1023.4 539.9 ...  
## $ rain\_seasonailty : num 91.6 38.5 67.1 91.5 48.3 ...  
## $ rain\_warmestQuart : num 12.69 138.33 9.51 318.54 183.14 ...  
## $ rain\_wettestMonth : num 67.8 159 13.4 202.2 79.2 ...  
## $ rain\_wettestQuart : num 175.8 435.9 33.3 524.3 211.7 ...  
## $ temp\_annual\_range : num 40.3 27.1 36.5 21.5 26.8 ...  
## $ temp\_coldestQuart : num -0.261 3.58 13.152 18.794 8.024 ...  
## $ temp\_diurnal\_range : num 14.72 9.11 14.87 13.85 13.46 ...  
## $ temp\_driestQuart : num 21.1 19.6 26.9 18.9 11.1 ...  
## $ temp\_max\_warmestMonth : num 32 26.3 41.5 31 28.2 ...  
## $ temp\_mean\_annual : num 11.5 11.5 23 21.6 14.2 ...  
## $ temp\_min\_coldestMonth : num -8.312 -0.806 5.058 9.549 1.443 ...  
## $ temp\_seasonality : num 88.2 62.7 75.1 18.5 47.6 ...  
## $ temp\_warmestQuart : num 22.7 19.6 32.5 23.3 20.2 ...  
## $ temp\_wettestQuart : num 3.95 5.27 20.81 22.76 16.48 ...  
## $ wind : num 3.43 2.47 4.03 2.16 4.27 ...  
## $ cloudiness : num 114.2 181.1 90.7 187.5 159 ...

* Research Question

Which variables in Rain and Temperature contribute most towards happiness

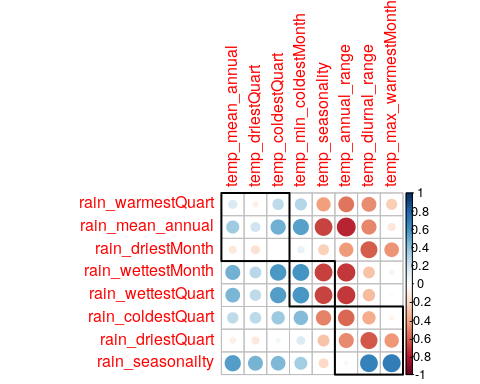
## Analysis

### Correlation Plot

Visually analyze multicollinearity between all varaibles in Rain and Temperature tables.

rain = country\_env\_df\_for\_pca[,grep('rain',colnames(country\_env\_df\_for\_pca))]  
temp = country\_env\_df\_for\_pca[,grep('temp',colnames(country\_env\_df\_for\_pca))]  
  
corr\_result = cor(rain, temp)  
corrplot(corr\_result,order = 'hclust', addrect = 3)

## Warning in as.dist.default(1 - corr): non-square matrix



### Identify Latent Components

#### PLS-C

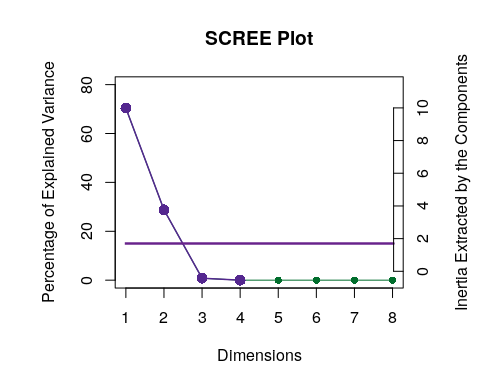
Now we have Latent Variables and Saliences. \* Latent Variables are the new Data points w.r.t. correlation between both the tables. Latent Variables exists for each table. \* Saliences represent correlation between variables of each table.

#### Scree Plot

Gives amount of information explained by corresponding component. Gives an intuition to decide which components best represent data in order to answer the research question.

P.S. The most contribution component may not always be most useful for a given research question.

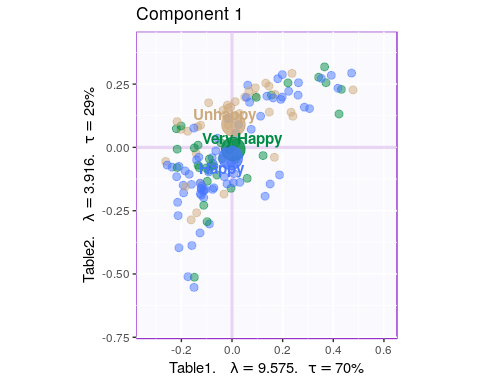
PTCA4CATA::PlotScree(ev = resPLSC$TExPosition.Data$eigs,  
 title = 'SCREE Plot',  
 p.ev = resPerm4PLSC$pEigenvalues,  
 plotKaiser = TRUE  
)



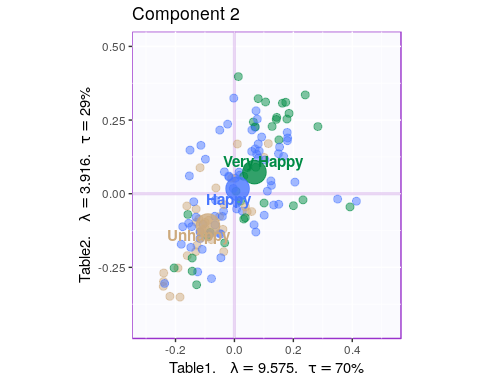
#### Latent Variables

Lets visualize happiness categories for Components 1 of each table

##### Component 1 for both Tables: Rain and Temperature

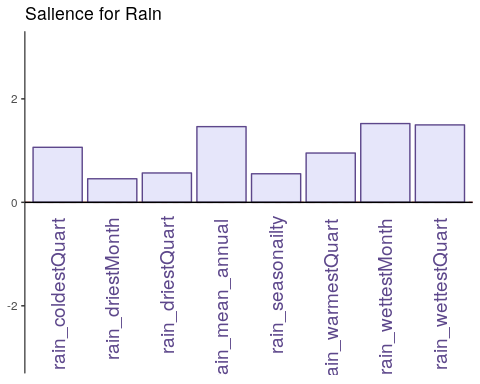


##### Component 2 for both Tables: Rain and Temperature

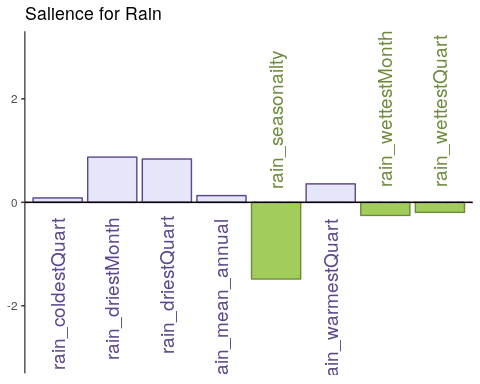


#### Salience for Rain

##### Components 1

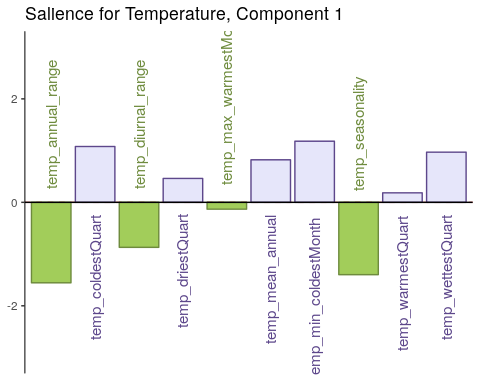


##### Component 2

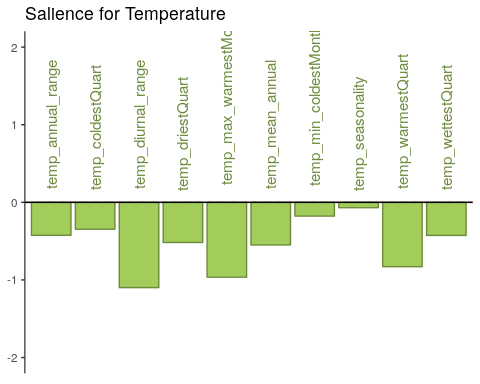


#### Salience for Temperature

##### Component 1



##### Component 2

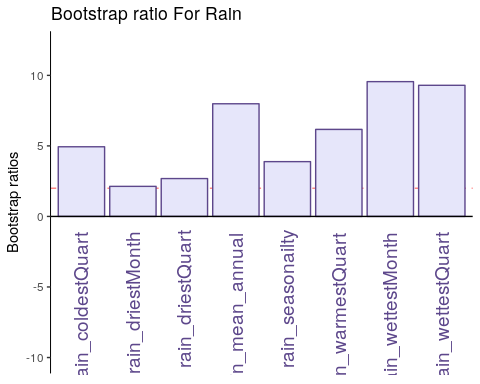


#### Most Contributing Variables - PLS-C (with Inference)

* Bootstrap Test

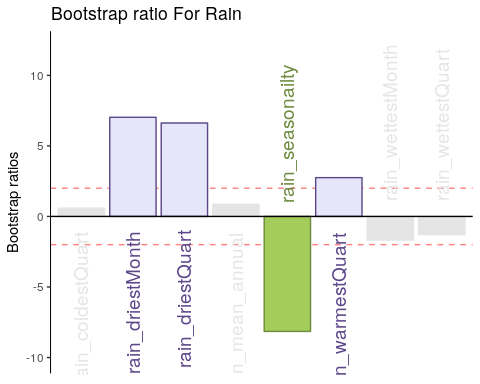
##### Rain - Component 1

BR = resBoot4PLSC$bootRatios.i  
  
PrettyBarPlot2(BR[,1],  
 threshold = 2,  
 font.size = 5,  
 main = 'Bootstrap ratio For Rain ',  
 ylab = 'Bootstrap ratios',  
 horizontal = TRUE,  
 ylim = c(-10,12)  
 )



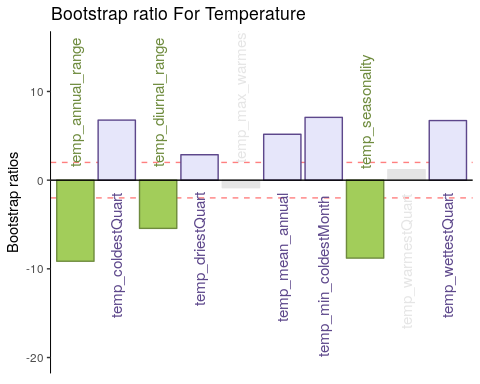
##### Rain - Component 2

BR = resBoot4PLSC$bootRatios.i  
  
PrettyBarPlot2(BR[,2],  
 threshold = 2,  
 font.size = 5,  
 #color4bar = gplots::col2hex(col4J), # we need hex code  
 main = 'Bootstrap ratio For Rain ',  
 ylab = 'Bootstrap ratios',  
 horizontal = TRUE,  
 ylim = c(-10,12)  
 )



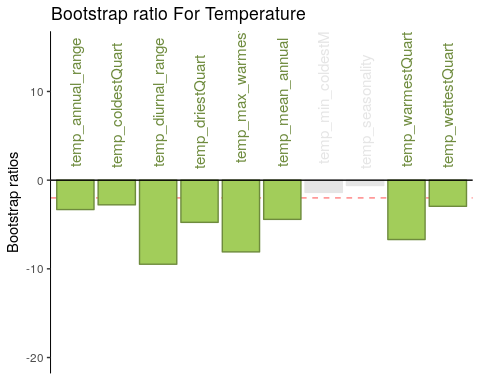
##### Temperature - Component 1

BR = resBoot4PLSC$bootRatios.j  
  
PrettyBarPlot2(BR[,1],  
 threshold = 2,  
 font.size = 4,  
 #color4bar = gplots::col2hex(col4J), # we need hex code  
 main = 'Bootstrap ratio For Temperature ',  
 ylab = 'Bootstrap ratios',  
 horizontal = TRUE,  
 ylim = c(-20,15)  
 )



##### Temperature - Component 2

BR = resBoot4PLSC$bootRatios.j  
  
PrettyBarPlot2(BR[,2],  
 threshold = 2,  
 font.size = 4,  
 #color4bar = gplots::col2hex(col4J), # we need hex code  
 main = 'Bootstrap ratio For Temperature ',  
 ylab = 'Bootstrap ratios',  
 horizontal = TRUE,  
 ylim = c(-20,15)  
 )



### Conclusion

Here Component 2 seems to best seperate Happiness levels. Let’s compare Component 2 for both tables.

* Table 1 & 2 Component 2
  + Latent Variables: Very Happy vs Unhappy (for Rain and Temperature both)
  + Salience:
    - Rain: It seems dryness and wetness at a montly scale have more effect than coldness or yearly patterns.
    - Temperature: All temperature variations at a monthly and yearly scale seems to impact happiness.