Country Environment Factors correlated with Happiness

Ritesh Malaiya

2018-11-22

Contents

Chapter 1

Introduction

1.1 Context

Assessing country-level social and economic statistics are often limited to socio-economic data. Not any more! This dataset will be maintained and updated with miscellaneous environmental data for countries across the globe.

1.2 Content

This data is all acquired through Google Earth Engine (https://earthengine.google.com/) where publicly available remote sensing datasets have been uploaded to the cloud to be manipulated by the average Joe like you and I. Most of the data is derived by calculating the mean for each country at a reduction scale of about 10km.

1.3 Inspiration

Can you use environmental statistics to predict social and economic data? Are people more happy in sunny countries? How do economies in forested countries compare with those dominated by grassland/desert?

1.4 Research Question (scope of this book)

How do the 137 countries differ on these variables?

Chapter 2

Dataset

- Data: Measurements of environment conditions in Countries
- Rows: There are 137 observations, 1 for each country.
- Columns: Total 29 variables
- Qualitative: Country (nominal), Happiness (Ordinal).
- Quantitative: Aspect, Slope Crop Land, Tree Canopy Wind Cloud & Multiple variables for Temp & Rain

2.1 Structure of Data

2.2 Datatable

Table 2.1: Here is a nice table!

Country	Happiness_Rank	accessibility_to_cities	elevation	aspect	slope	cropland_cov
Afghanistan	U	317.71575	1831.74440	201.4298	1.5156001	9.5118
Albania	H	73.83086	651.81554	192.1303	1.8900753	23.3460
Algeria	H	1212.79982	556.75832	184.9747	0.1708615	3.6908
Angola	U	378.20239	1061.47899	174.2569	0.1926286	2.7944
Argentina	VH	209.21958	682.79925	145.0314	0.6238553	21.9625
Armenia	U	97.29452	1850.48297	183.5375	2.3188956	21.3382
Australia	VH	845.86802	278.03171	183.0250	0.0892681	7.9385
Austria	VH	51.98273	950.53000	166.5693	1.4500095	33.0035
Azerbaijan	H	86.63017	640.06762	127.1189	1.4408533	36.1450
Bangladesh	H	33.64179	29.93988	168.6095	0.0325392	68.1313
Belarus	H	79.96755	161.15274	173.2924	0.0014257	51.4663
Belgium	VH	14.38773	143.45749	218.8504	0.0953660	60.7762
Benin	U	115.20319	267.04215	178.5135	0.0380440	25.2792
Bhutan	H	501.28316	2860.92535	181.5962	3.6983940	2.4223
Bolivia	H	445.12942	1290.02061	151.5460	0.6737767	4.0574
Bosnia and Herzegovina	H	73.43443	710.18800	160.8869	1.3528183	37.5556
Botswana	U	386.90666	1037.20212	162.7331	0.0062352	1.0776
Brazil	VH	736.72980	326.64308	176.1375	0.1606525	17.2823
Bulgaria	U	89.12241	467.44587	160.3145	0.9749420	52.3124
Burkina Faso	U	115.86369	302.30818	169.5040	0.0053619	14.9930

Chapter 3

Principal Component Analysis

3.1 Description

Principal component analysis (PCA), part of descriptive analytics, is used to analyze one table of quantitative data, specifically useful for *high dimensional data* and comparitively lesser data rows. PCA mixes the input variables to give new variables, called principal components. The first principal component is the line of best fit. It is the line that maximizes the inertia (similar to variance) of the cloud of data points. Subsequent components are defined as orthogonal to previous components, and maximize the remaining inertia.

PCA gives one map for the rows (called factor scores), and one map for the columns (called loadings). These 2 maps are related, because they both are described by the same components. However, these 2 maps project different kinds of information onto the components, and so they are *interpreted differently.* Factor scores are the coordinates of the row observations and Loadings describe the column variables. Both can be interpreted through their distance from origin. However, Factor scores are also interpreted by the distances between them and Loadings interpreted by the angle between them.

The distance from the origin is important in both maps, because squared distance from the mean is inertia (variance, information; see sum of squares as in ANOVA/regression). Because of the Pythagorean Theorem, the total information contributed by a data point (its squared distance to the origin) is also equal to the sum of its squared factor scores.

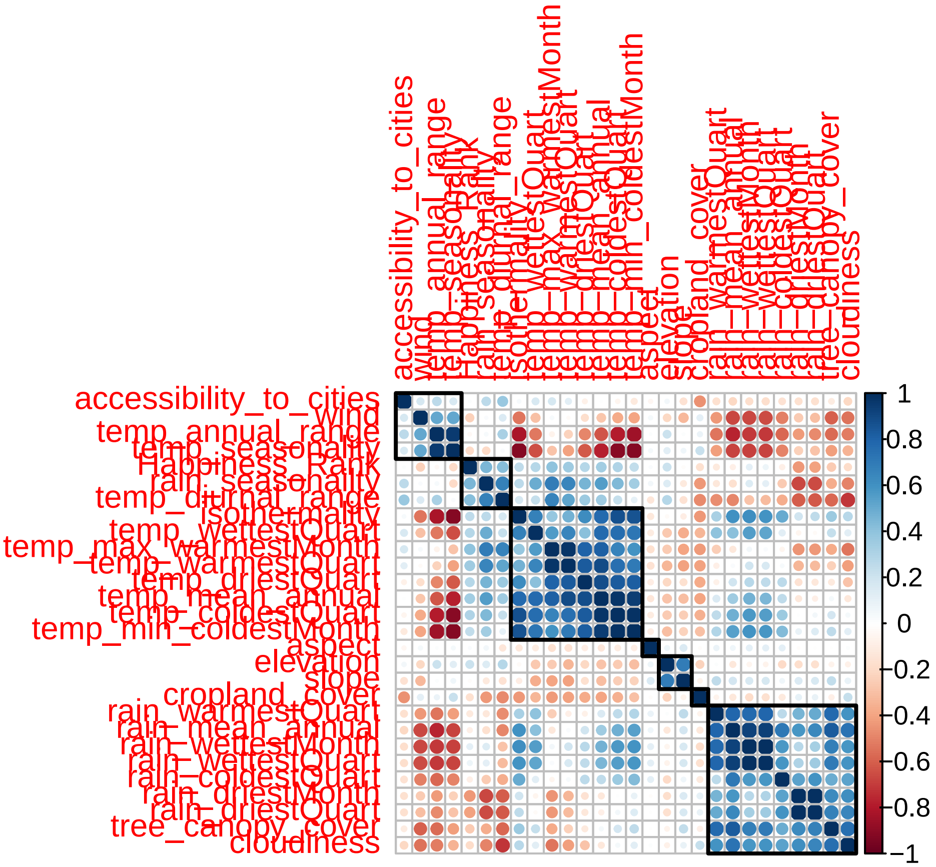
With both Factor and Loadings maps combined we can interpret which grouping criteria of rows of data is most impacted by which columns. This can interpreted visually by observing which a factors and loadings on a particular component and the distance on this component.

PCA also helps in *dimensionality reduction.* Using SVD, we get eigen values arranged in descending order in the diagonal matrix. We can simply ignore the lower eigen values to reduce dimensions. We can also take help of SCREE plot to visually analyze importance of eigen values.

There are multiple variables representing rain and Temp. Hence, for analysis purposes, lets choose annual mean for Rain and Temp.

3.2 Correlation Plot

Visually analyze multicollinearity in the system.

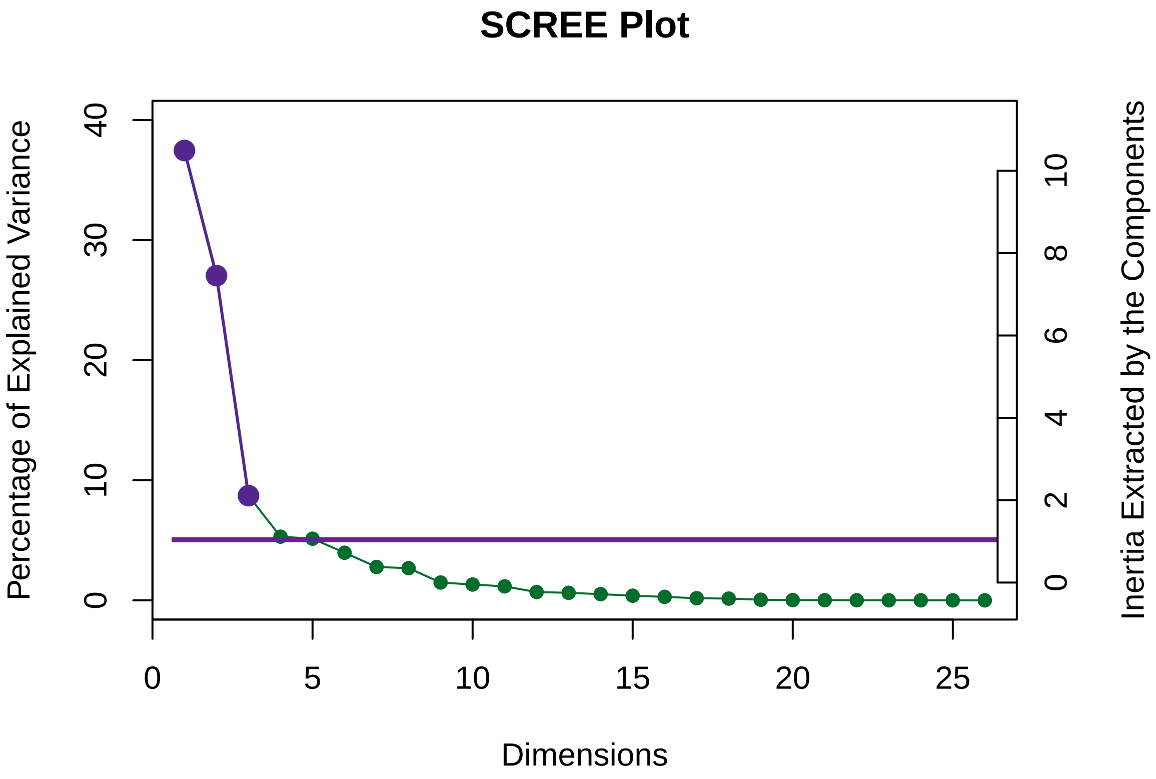


Now we have Factor scores and Loadings. * Factor Scores are the new Data points w.r.t. new Components achieved with help of SVD. * Loadings represent correlation between variables w.r.t the choosen Components. Can be interpreted in 3 ways + As slices of inertia of the contribution data table w.r.t. the choosen Components + As correlation between columns (features) of Original Data and Factor scores of each Components (latent features). + As coefficients of optimal linear combination i.e. Right Sigular Vectors (Q matrix of SVD)

3.3 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.



3.4 Factor Scores

Lets visualize happiness categories for components 1-10, to make a decision (visually) on the most important components.

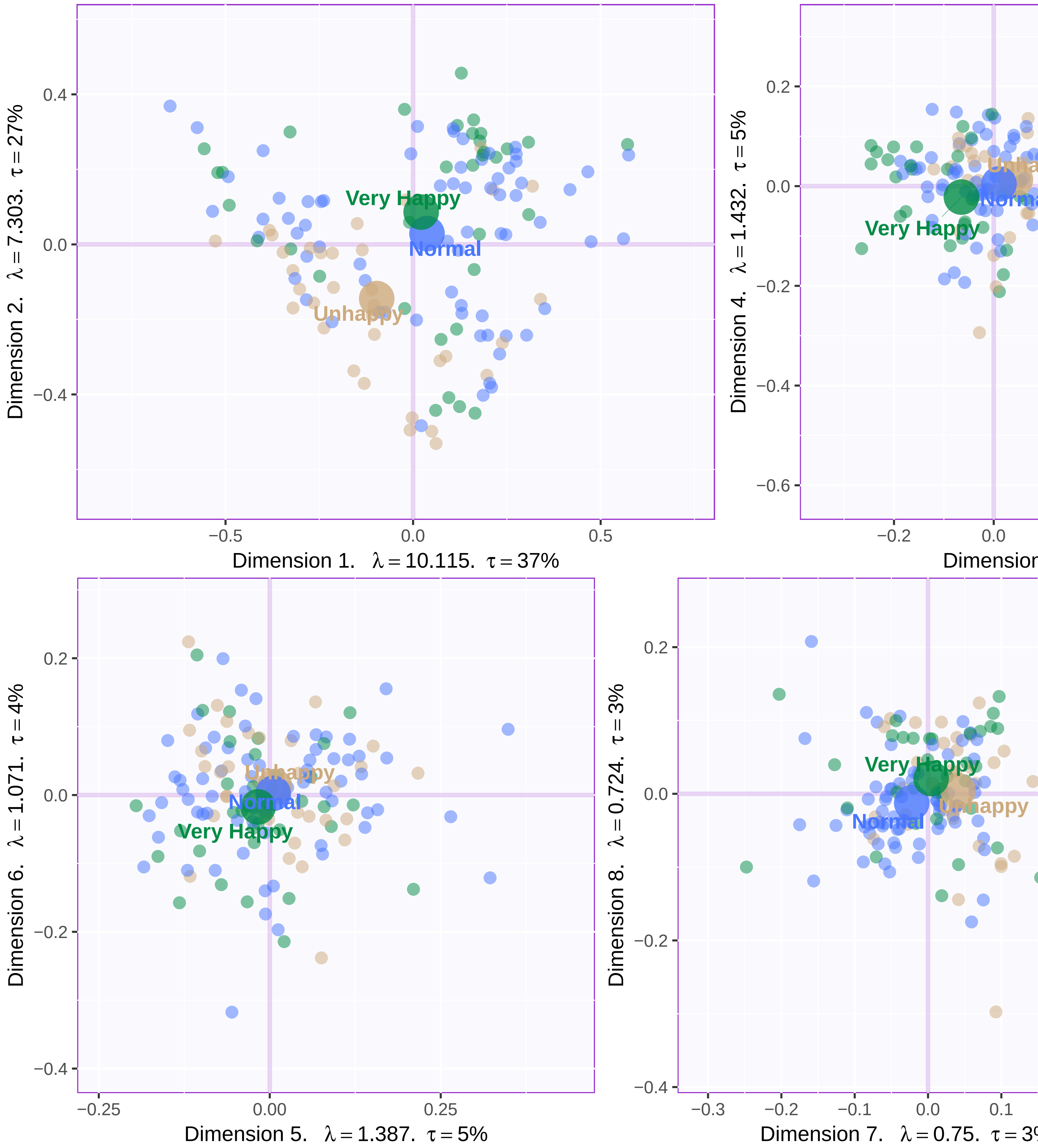
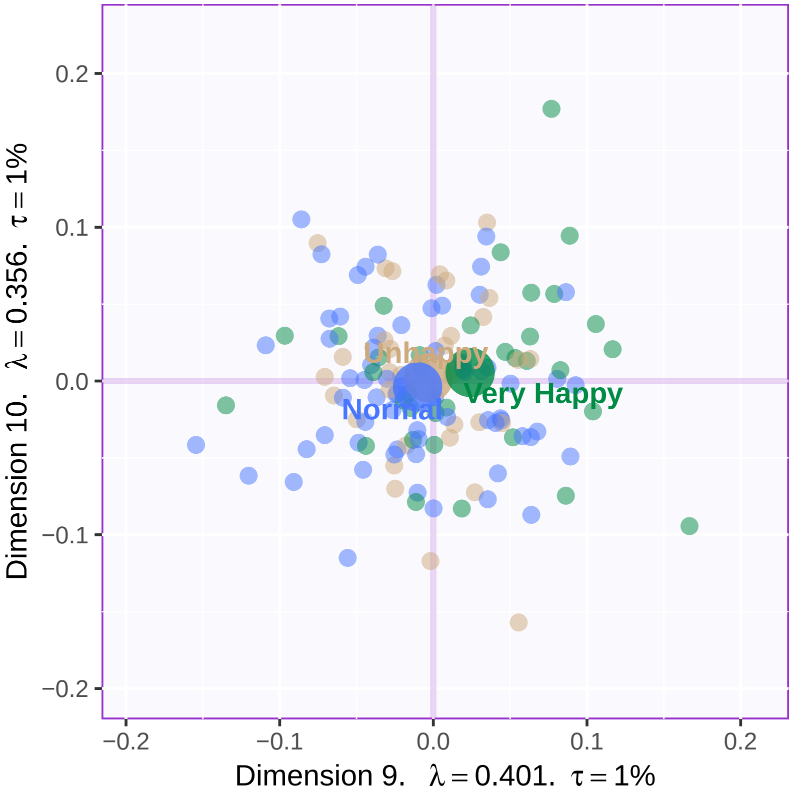

Very Happy
Normal
Unhappy
0.4
0.0
−0.4
−0.5
0.0
0.5
Dimension 1. λ = 10.115. τ = 37%
Dimension 2. λ = 7.303. τ = 27%
Unha
Norma
Very Happy
0.2
0.0
−0.2
−0.4
−0.6
−0.2
0.0
Dimension
Dimension 4. λ = 1.432. τ = 5%
Unhappy
Normal
Very Happy
0.2
0.0
−0.2
−0.4
−0.25
0.00
0.25
Dimension 5. λ = 1.387. τ = 5%
Dimension 6. λ = 1.071. τ = 4%
Very Happy
Normal
Unhappy
0.2
0.0
−0.2
−0.4
−0.3
−0.2
−0.1
0.0
0.1
Dimension 7. λ = 0.75. τ = 3
Dimension 8. λ = 0.724. τ = 3%

Table 3.1: Identify Components best describing happiness levels

	happy	happier	happiest
Component 1	-1	1	0
Component 2	-1	0	1
Component 3	1	0	-1
Component 4	0	0	0
Component 5	0	0	0
Component 6	0	0	0
Component 7	1	-1	0
Component 8	0	0	0
Component 9	0	-1	1
Component 10	0	0	0

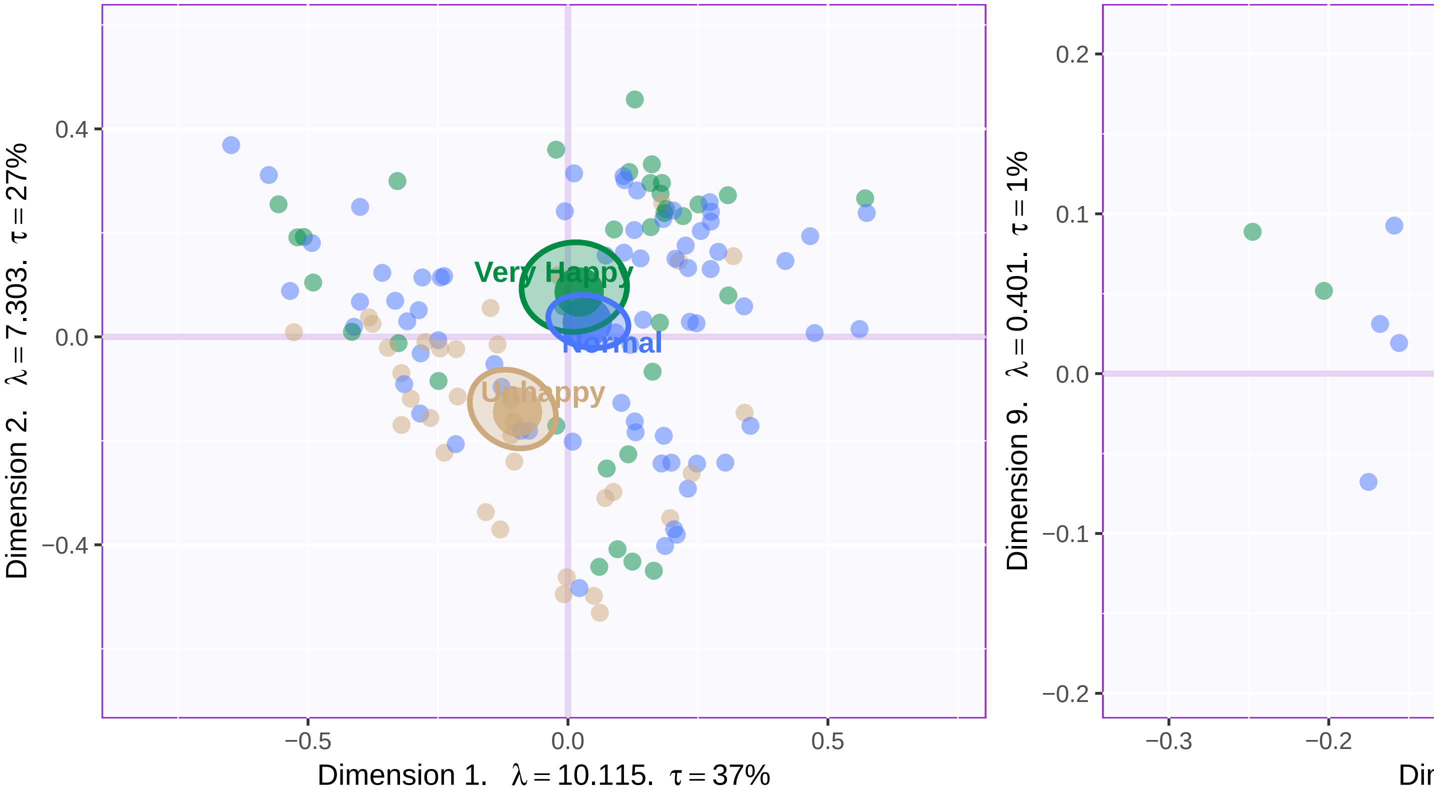


Since, it's not very straightforward to decide which components may be best suited for the research question at hand, let's represent, in a tabular format, which component helps to differentiate between which design variable values (Unhappy, Normal, Very Happy)

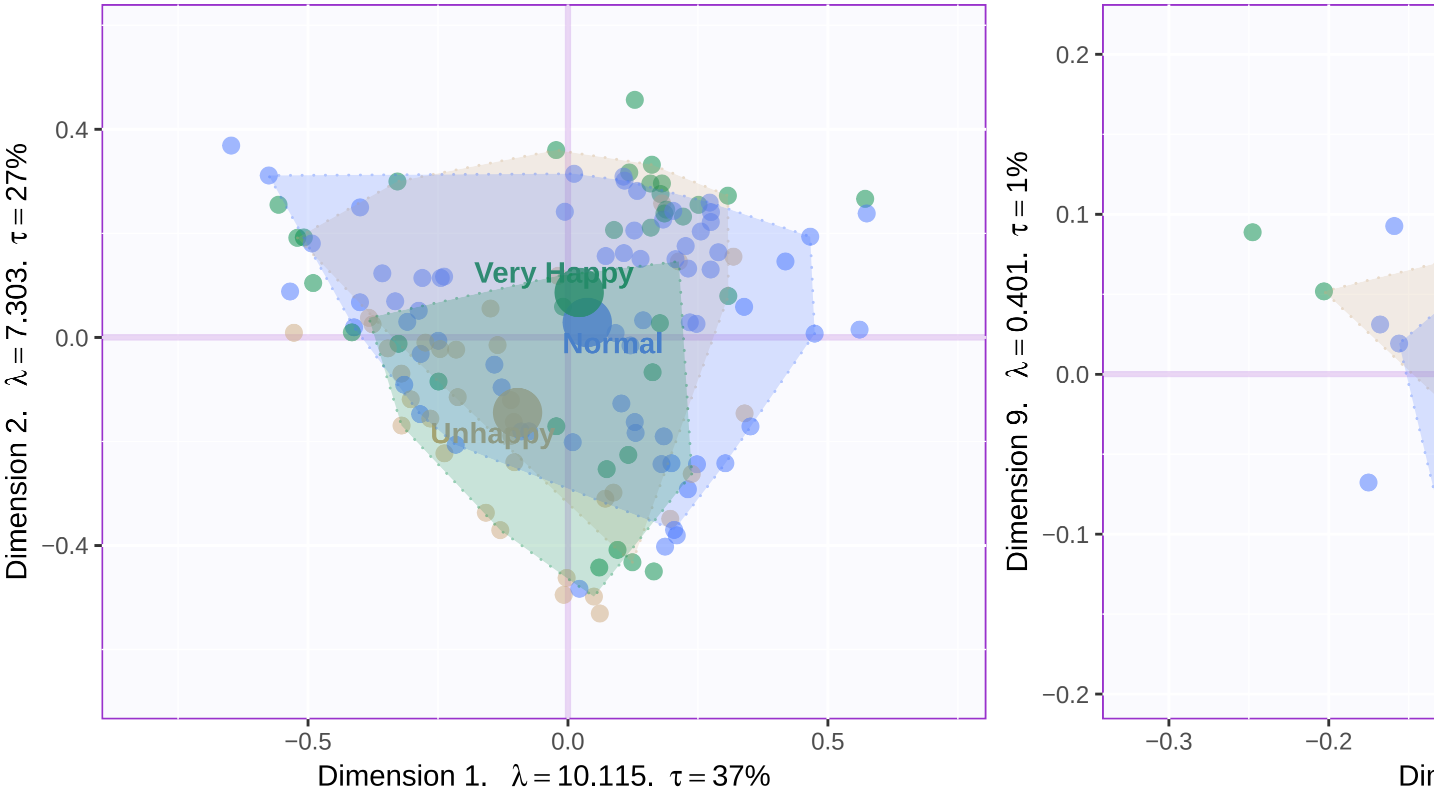
P.S. here -1 represents -ve quadrant of the component and +1 represent +ve quadrant. 0 represents that component was not decisive enough to clearly seperate happiness levels.

Looking at the table, it seems component 1, 2, 7, 9 may be able to best represent all 3 happiness levels. Although, SCREE Plot suggests that 3^{rd} and 4^{th} components might be useful, from our above analysis we know otherwise. Also, SCREE plot suggests that component 6^{th} and onwards might not be useful which is contradicting our findings above. Hence, let's plot components 1 vs 2 and 7 vs 9. Similarily, we will also plot Loading plots for these componenets.

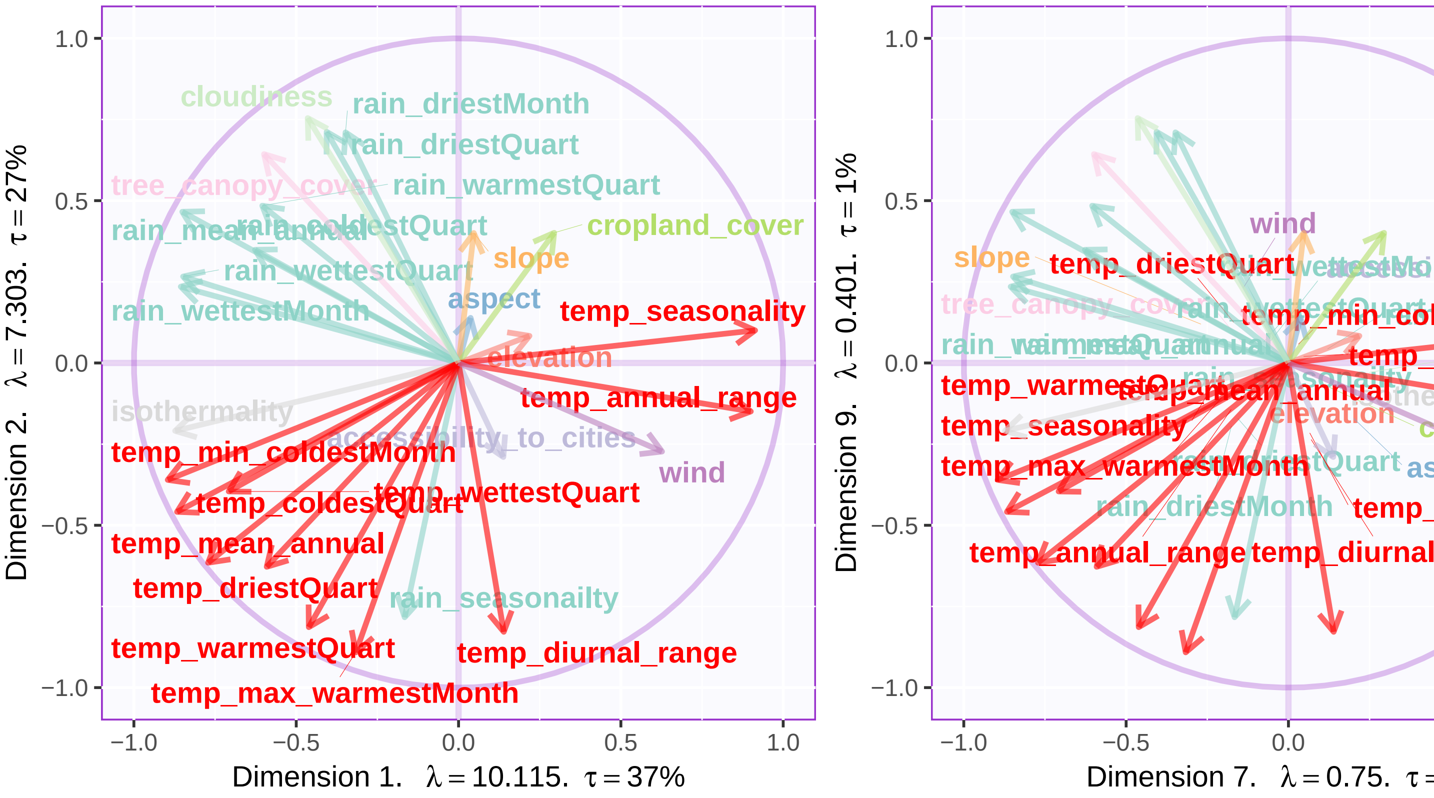
- With Confidence Interval



- With Tolerance Interval



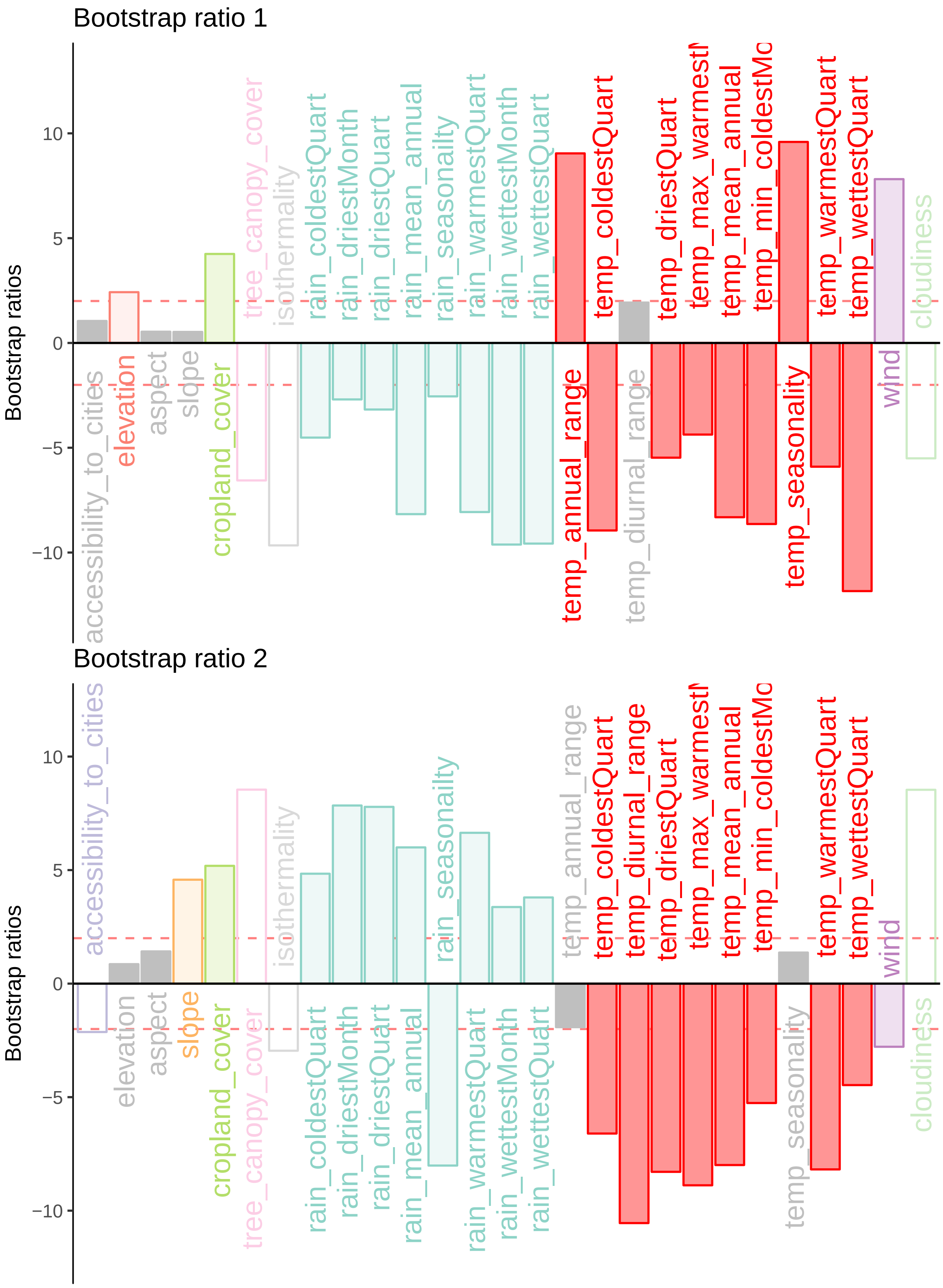
3.5 Loadings

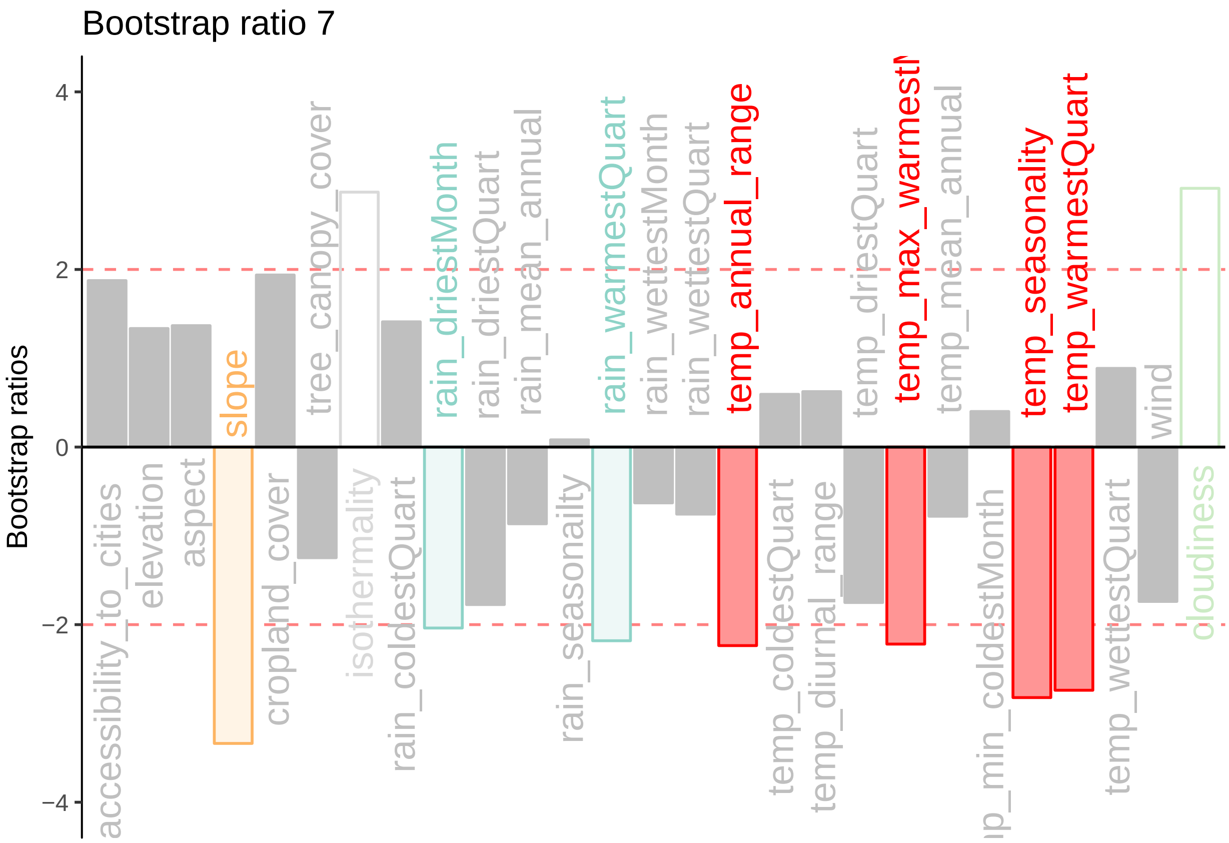


3.6 Most Contributing Variables

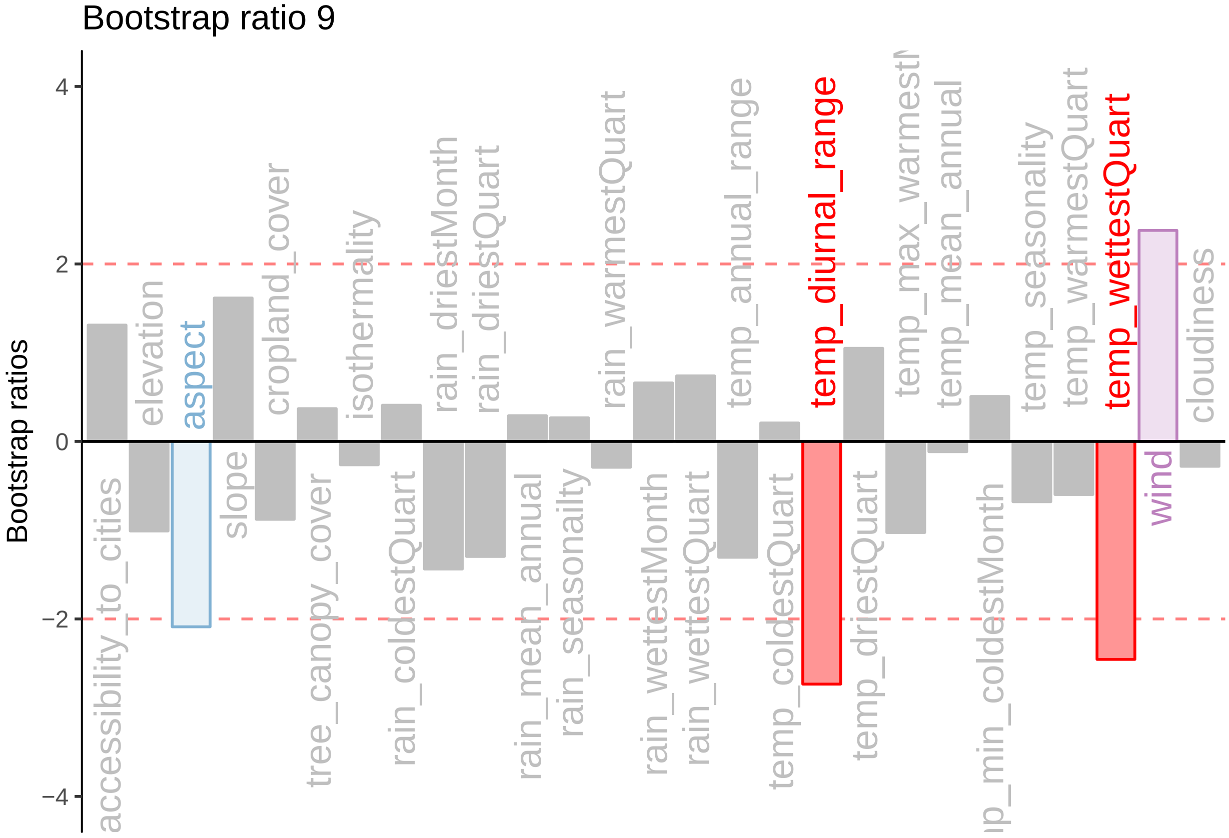
Let's plot variable contributions against each chosen components i.e. 1, 7, 9.

- With Bootstrap Ratio


Bootstrap ratio 1
Bootstrap ratios
10
5
0
−5
−10
accessibility_to_cities
elevation
aspect
slope
cropland_cover
tree_canopy_cover
isothermality
rain_coldestQuart
rain_driestMonth
rain_driestQuart
rain_mean_annual
rain_seasonailty
rain_warmestQuart
rain_wettestMonth
rain_wettestQuart
temp_annual_range
temp_coldestQuart
temp_diurnal_range
temp_driestQuart
temp_max_warmestM
temp_mean_annual
temp_min_coldestMo
temp_seasonality
temp_warmestQuart
temp_wettestQuart
wind
cloudiness
Bootstrap ratio 2
Bootstrap ratios
10
5
0
−5
−10
accessibility_to_cities
elevation
aspect
slope
cropland_cover
tree_canopy_cover
isothermality
rain_coldestQuart
rain_driestMonth
rain_driestQuart
rain_mean_annual
rain_seasonailty
rain_warmestQuart
rain_wettestMonth
rain_wettestQuart
temp_annual_range
temp_coldestQuart
temp_diurnal_range
temp_driestQuart
temp_max_warmestM
temp_mean_annual
temp_min_coldestMo
temp_seasonality
temp_warmestQuart
temp_wettestQuart
wind
cloudiness

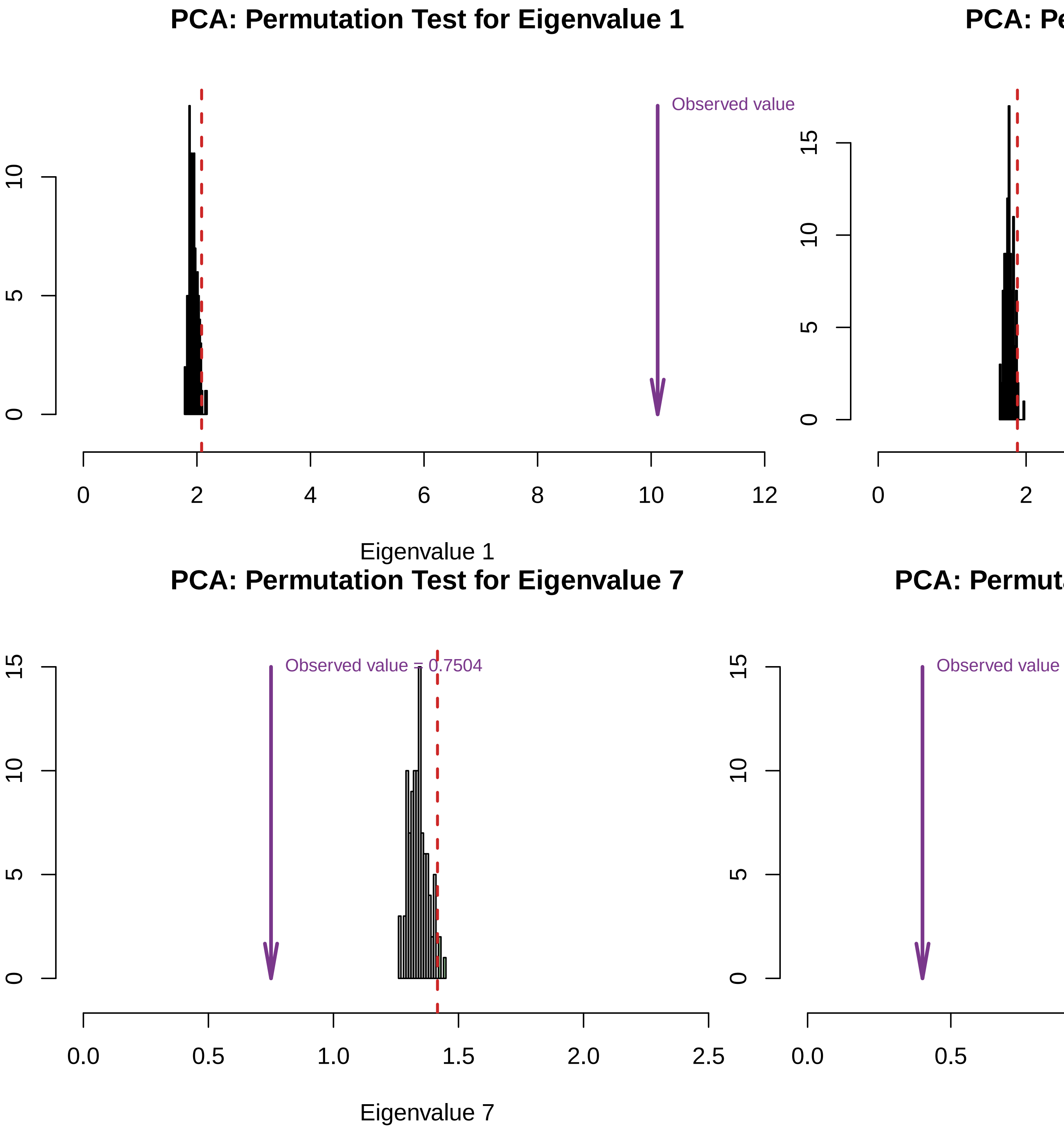


Bootstrap ratio 7

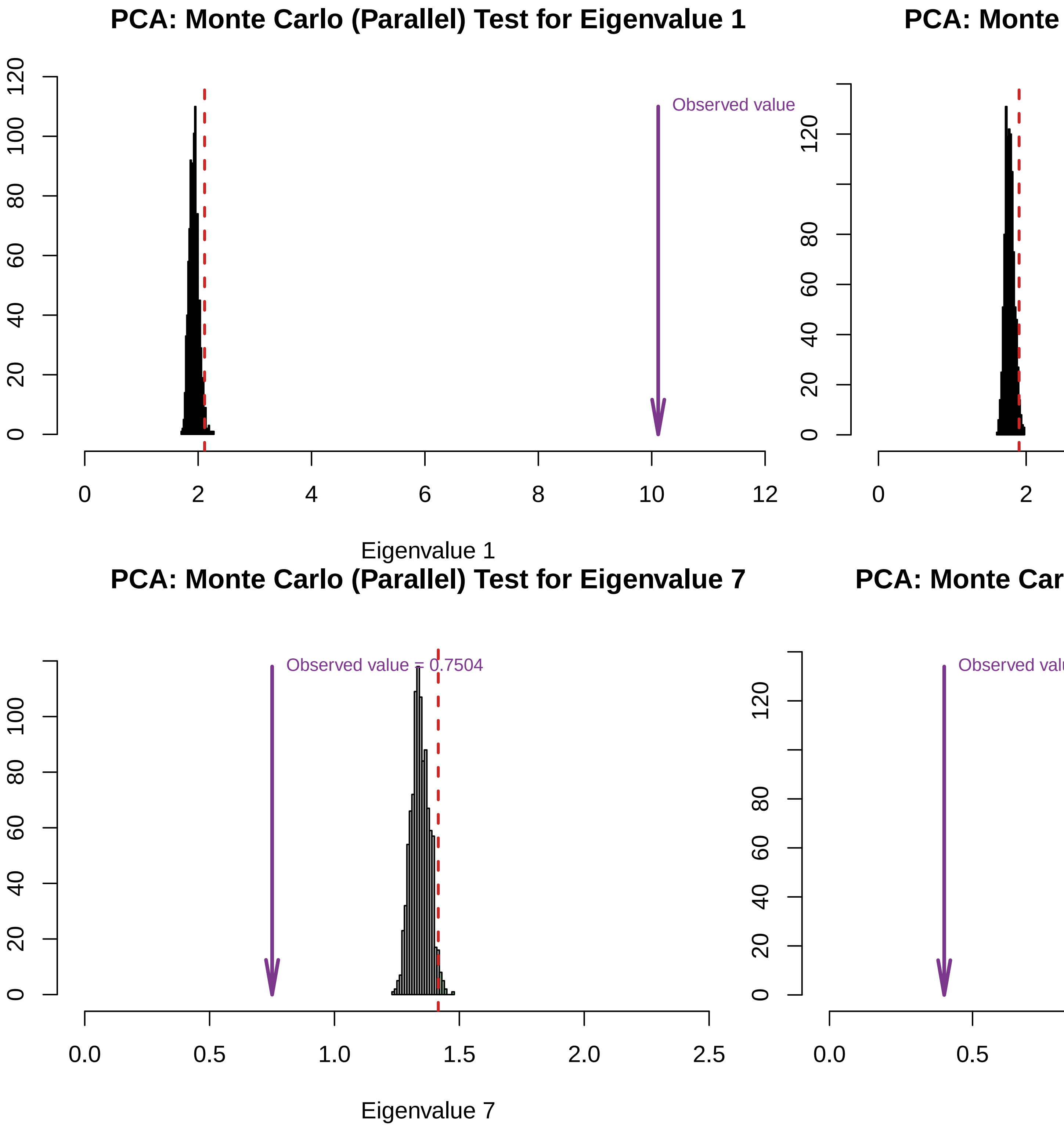


Bootstrap ratio 9

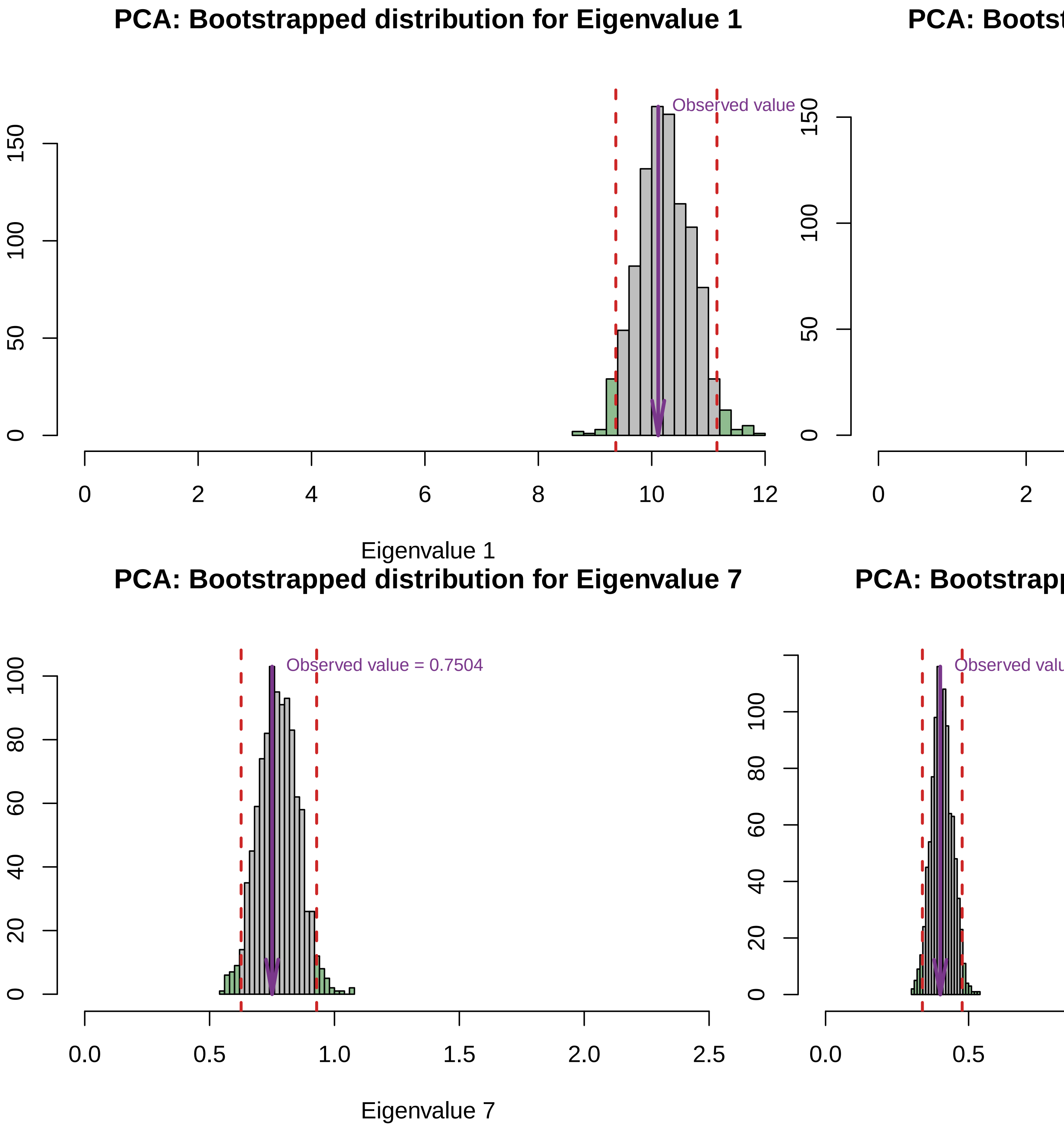
3.7 Permutation Test



3.8 Parallet Test



3.9 Bootstrap Test



3.10 Conclusion

- Component 1:
  - Rows: Normal & Happy
  - Columns: Cloudiness & Rain vs Cropland, Aspect, Elevation
  - Interpret: People in countries with more Cloudiness, Trees and Rain tends to be happier.
- Component 7:
  - Rows: Happy & Unhappy
  - Columns: Temp and Rain vs Accessibility and Cropland
  - Interpret: Rain and Temp seems to be main reason for unhappiness and Cropland is important for Happiness.
- Component 9:
  - Rows: Happy & Very Happy
  - Columns: Temp vs Rain
  - Interpret: Rain and Temp seems to be main reason for Happiness. *This contradicts with Component 7 and 1.*

Chapter 4

Multiple Component Analysis

4.1 Description

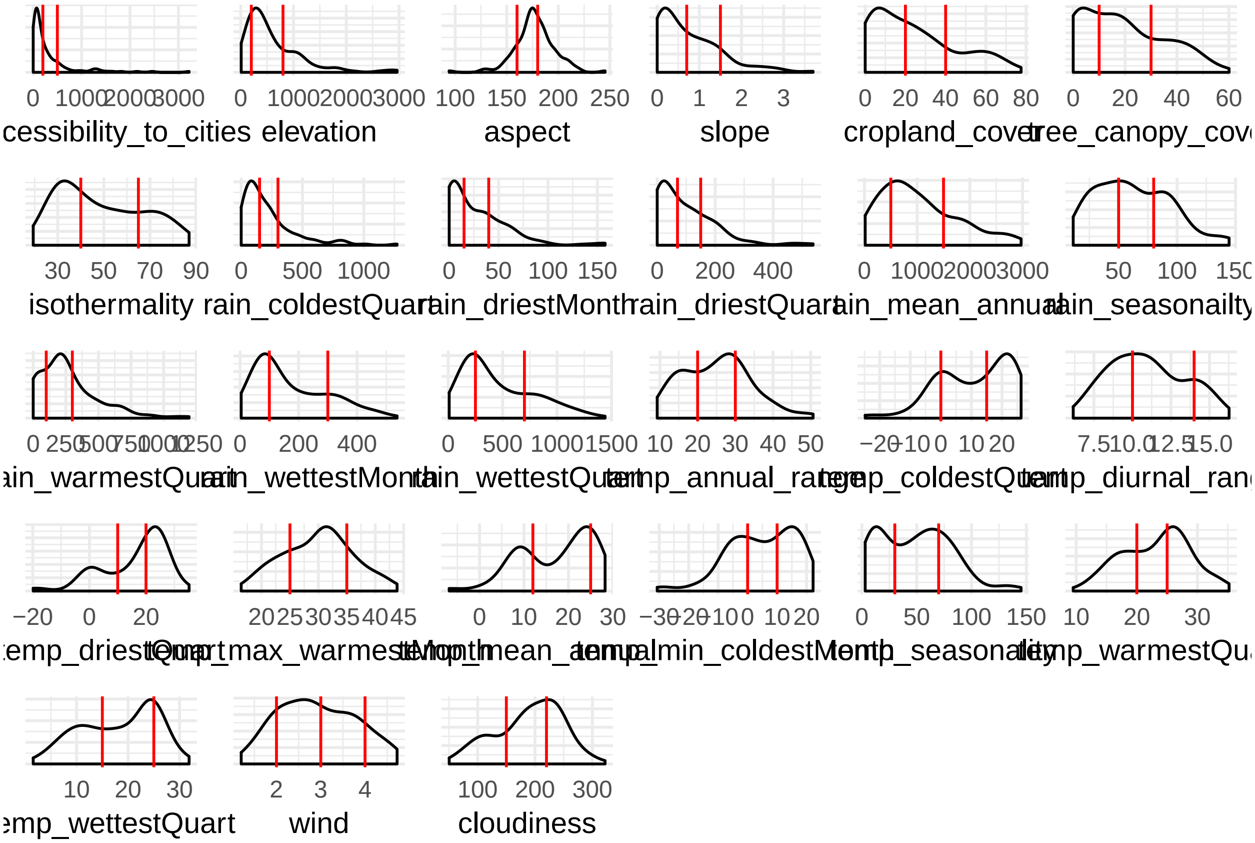
Multiple correspondence analysis (MCA) is an extension of correspondence analysis(CA) which allows one to analyze the pattern of relationships of several categorical dependent variables. As such, it can also be seen as a generalization of principal component analysis when the variables to be analyzed are categorical instead of quantitative. Because MCA has been (re)discovered many times, equivalent methods are known under several different names such as optimal scaling, optimal or appropriate scoring, dual scaling, homogeneity analysis,scalogram analysis, and quantification method.

Interpreting MCA Multiple correspondence analysis locates all the categories in a Euclidean space.

- The first two dimensions of this space are plotted to examine the associations among the categories.
- The top-right quadrant of the plot shows the categories.
- The bottom-left quadrant shows the association.
- This interpretation is based on points found in approximately the same direction from the origin and in approximately the same region of the space. Distances between points do not have a straightforward interpretation.

4.2 Density Plot

Let's observe the distribution of each variables to get an intuition of how we can bin these variables. It's important to have nearly equal number of observations in the each bin and to try to cut the variables in a way to so that each new binned distribution is nearly Gaussian. We can also verify that our binning is appropiate by calculating Spearman Correlation for each of original variable and binned variable, the correlation coefficient should be close to 1.



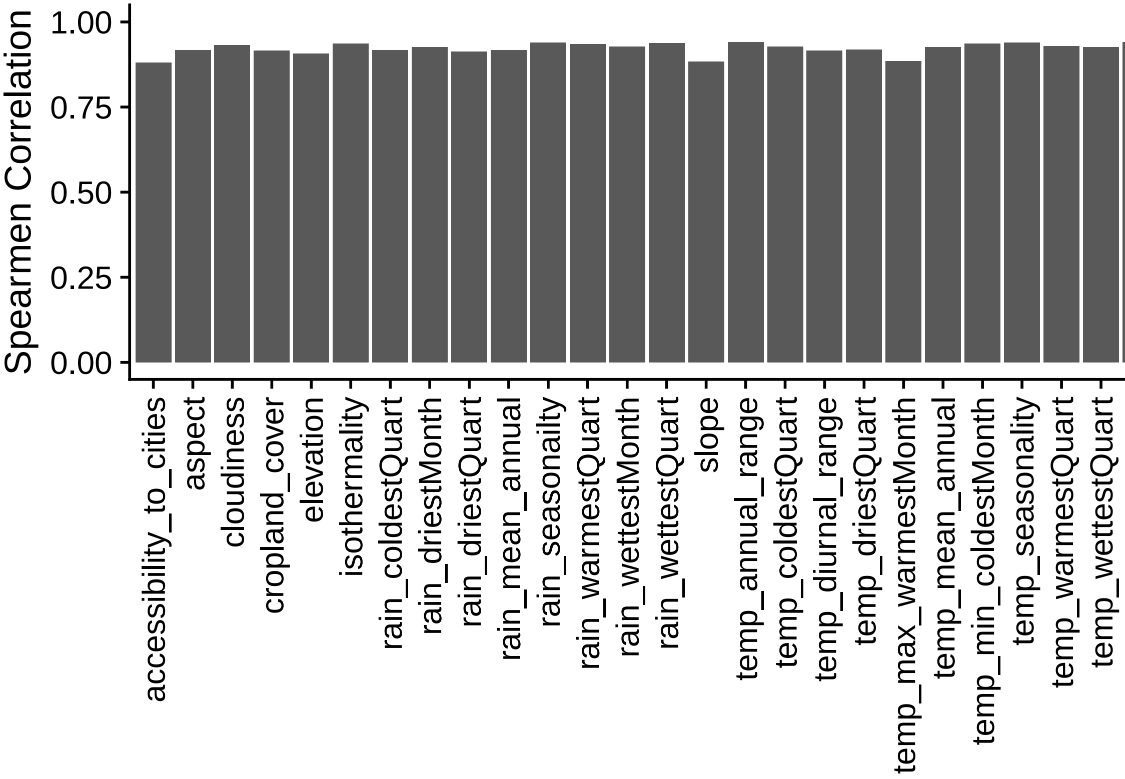
4.3 Binning

Structure of Data after binning based on above observation.

```
## 'data.frame':    137 obs. of  27 variables:
##  $ accessibility_to_cities: Factor w/ 3 levels "1","2","3": 2 1 3 2 2 1 3 1 1 1 ...
##  $ elevation              : Factor w/ 3 levels "1","2","3": 3 2 2 3 2 3 2 3 2 1 ...
##  $ aspect                 : Factor w/ 3 levels "1","2","3": 3 3 3 2 1 3 3 2 1 2 ...
##  $ slope                  : Factor w/ 3 levels "1","2","3": 3 3 1 1 1 3 1 2 2 1 ...
##  $ cropland_cover         : Factor w/ 3 levels "1","2","3": 1 2 1 1 2 2 1 2 2 3 ...
##  $ tree_canopy_cover      : Factor w/ 3 levels "1","2","3": 1 2 1 2 1 1 1 3 1 2 ...
##  $ isothermality          : Factor w/ 3 levels "1","2","3": 1 1 2 2 2 1 2 1 1 2 ...
##  $ rain_coldestQuart      : Factor w/ 3 levels "1","2","3": 1 3 1 1 1 1 1 2 1 1 ...
##  $ rain_driestMonth       : Factor w/ 3 levels "1","2","3": 1 3 1 1 2 2 1 3 2 1 ...
##  $ rain_driestQuart       : Factor w/ 3 levels "1","2","3": 1 2 1 1 1 1 1 3 1 1 ...
##  $ rain_mean_annual       : Factor w/ 3 levels "1","2","3": 1 2 1 2 2 2 1 2 1 3 ...
##  $ rain_seasonailty       : Factor w/ 3 levels "1","2","3": 3 1 2 3 1 1 2 1 1 3 ...
##  $ rain_warmestQuart      : Factor w/ 3 levels "1","2","3": 1 2 1 3 2 2 2 3 1 3 ...
##  $ rain_wettestMonth      : Factor w/ 3 levels "1","2","3": 1 2 1 2 1 1 1 2 1 3 ...
##  $ rain_wettestQuart      : Factor w/ 3 levels "1","2","3": 1 2 1 2 1 1 1 2 1 3 ...
##  $ temp_annual_range      : Factor w/ 3 levels "1","2","3": 3 2 3 2 2 3 2 2 3 2 ...
##  $ temp_coldestQuart      : Factor w/ 3 levels "1","2","3": 1 2 2 3 2 1 2 1 2 3 ...
##  $ temp_diurnal_range     : Factor w/ 3 levels "1","2","3": 3 1 3 2 2 2 2 1 1 1 ...
##  $ temp_driestQuart       : Factor w/ 3 levels "1","2","3": 3 2 3 2 2 1 2 1 2 2 ...
##  $ temp_max_warmestMonth  : Factor w/ 3 levels "1","2","3": 2 2 3 2 2 1 3 1 2 2 ...
##  $ temp_mean_annual       : Factor w/ 3 levels "1","2","3": 1 1 2 2 2 1 2 1 1 3 ...
##  $ temp_min_coldestMonth  : Factor w/ 3 levels "1","2","3": 1 1 2 2 2 1 2 1 1 3 ...
##  $ temp_seasonality       : Factor w/ 3 levels "1","2","3": 3 2 3 1 2 3 2 2 3 2 ...
##  $ temp_warmestQuart      : Factor w/ 3 levels "1","2","3": 2 1 3 2 2 1 3 1 2 3 ...
##  $ temp_wettestQuart      : Factor w/ 3 levels "1","2","3": 1 1 2 2 2 1 2 1 1 3 ...
##  $ wind                   : Factor w/ 4 levels "1","2","3","4": 3 2 4 2 4 1 4 2 2 2 ...
##  $ cloudiness             : Factor w/ 3 levels "1","2","3": 1 2 1 2 2 2 1 3 2 2 ...
```

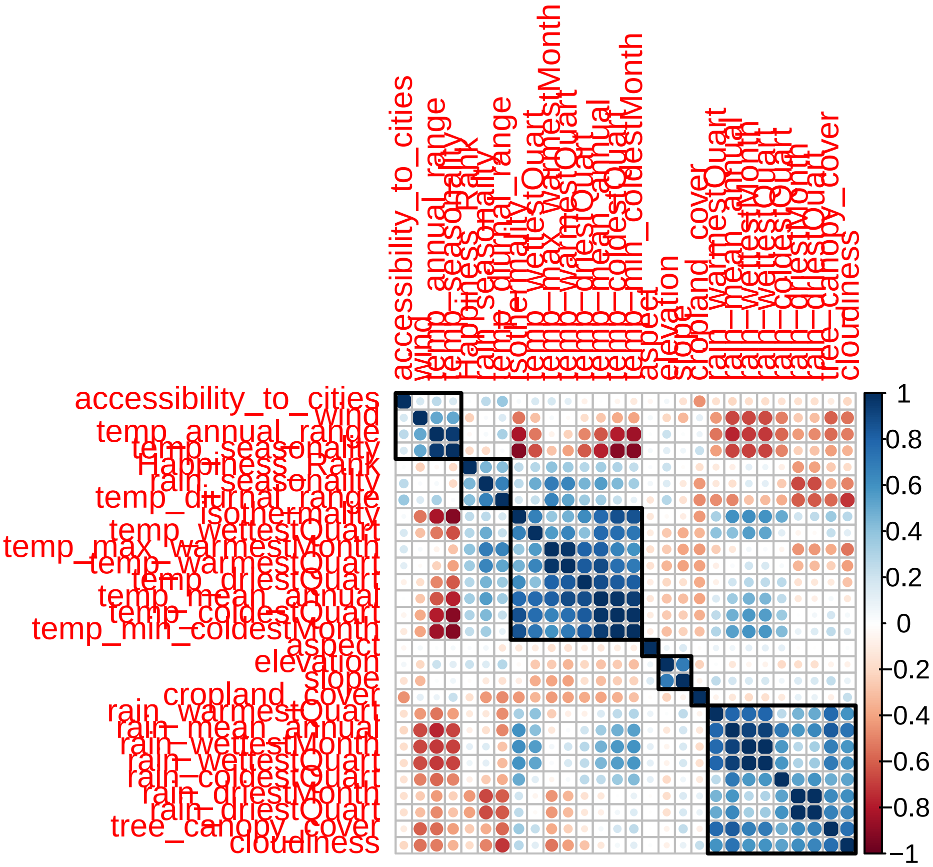
4.4 Spearman Correlation

Let's observe correlation between original data and binned data to make sure that neither the correlation ceoficient is too low or perfect.



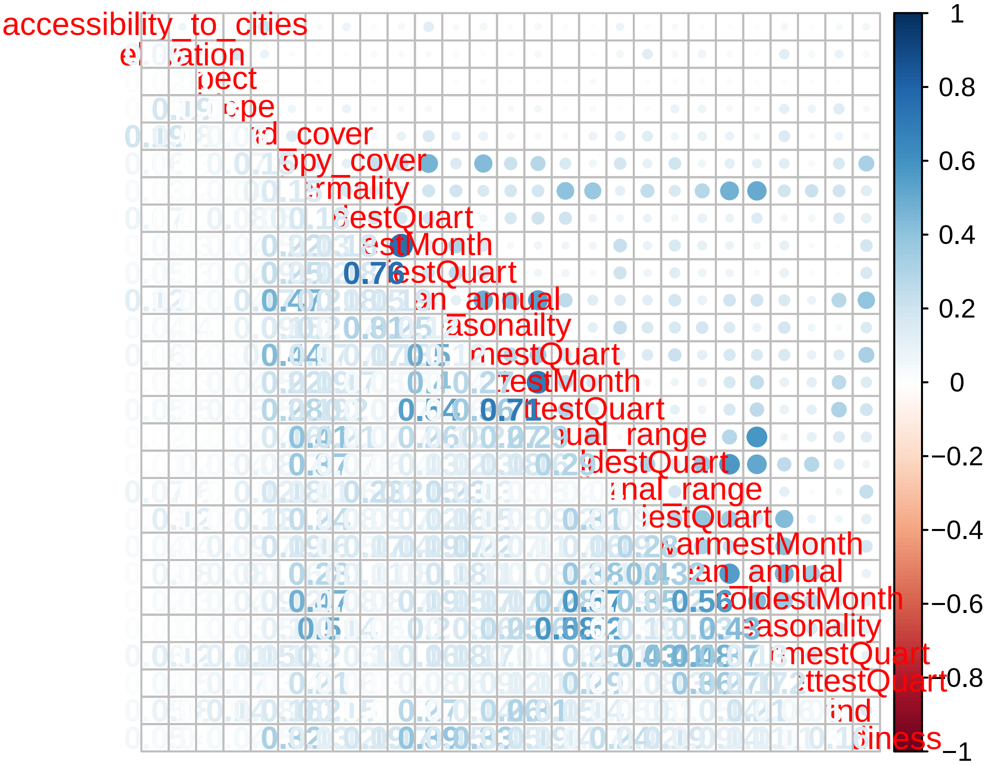
4.5 Correlation Plot

Visually analyze multicollinearity in the system of the original data



4.6 Heatmap

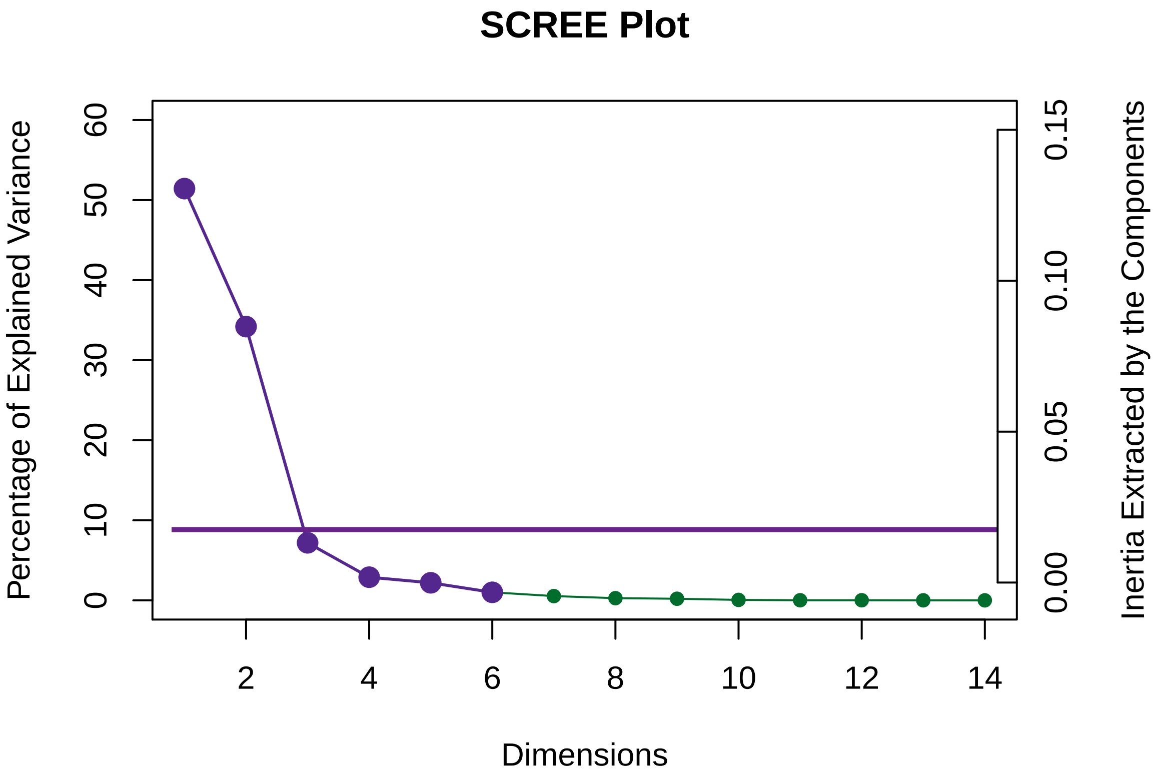
For binned data.



4.7 Scree Plot

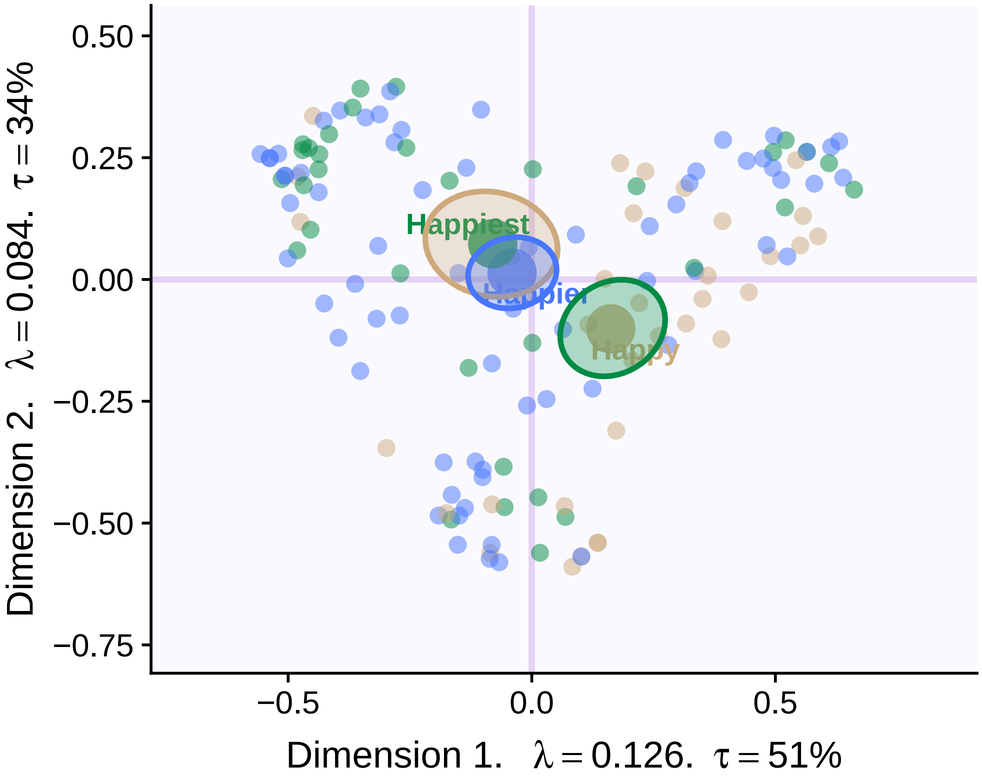
Gives amount of information explained by corresponding component. Gives an intuition to decide which components best represent data in order to answer the research question.

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

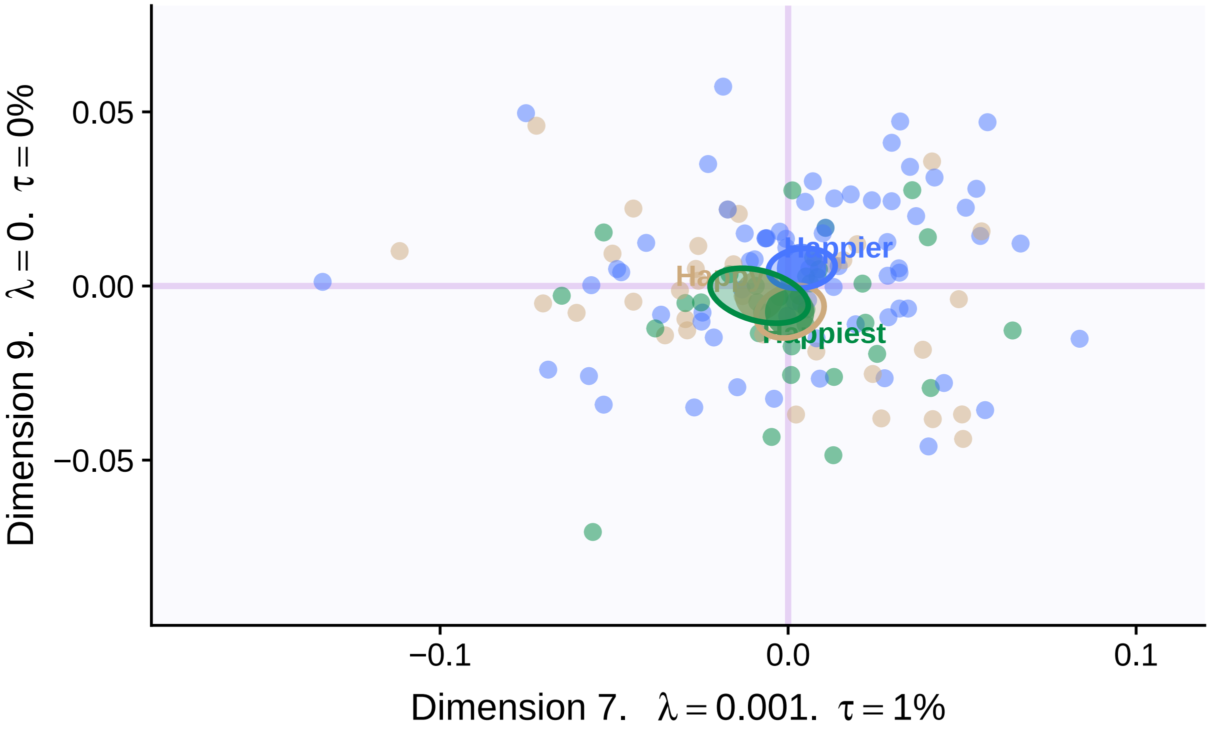


4.8 Factor Scores

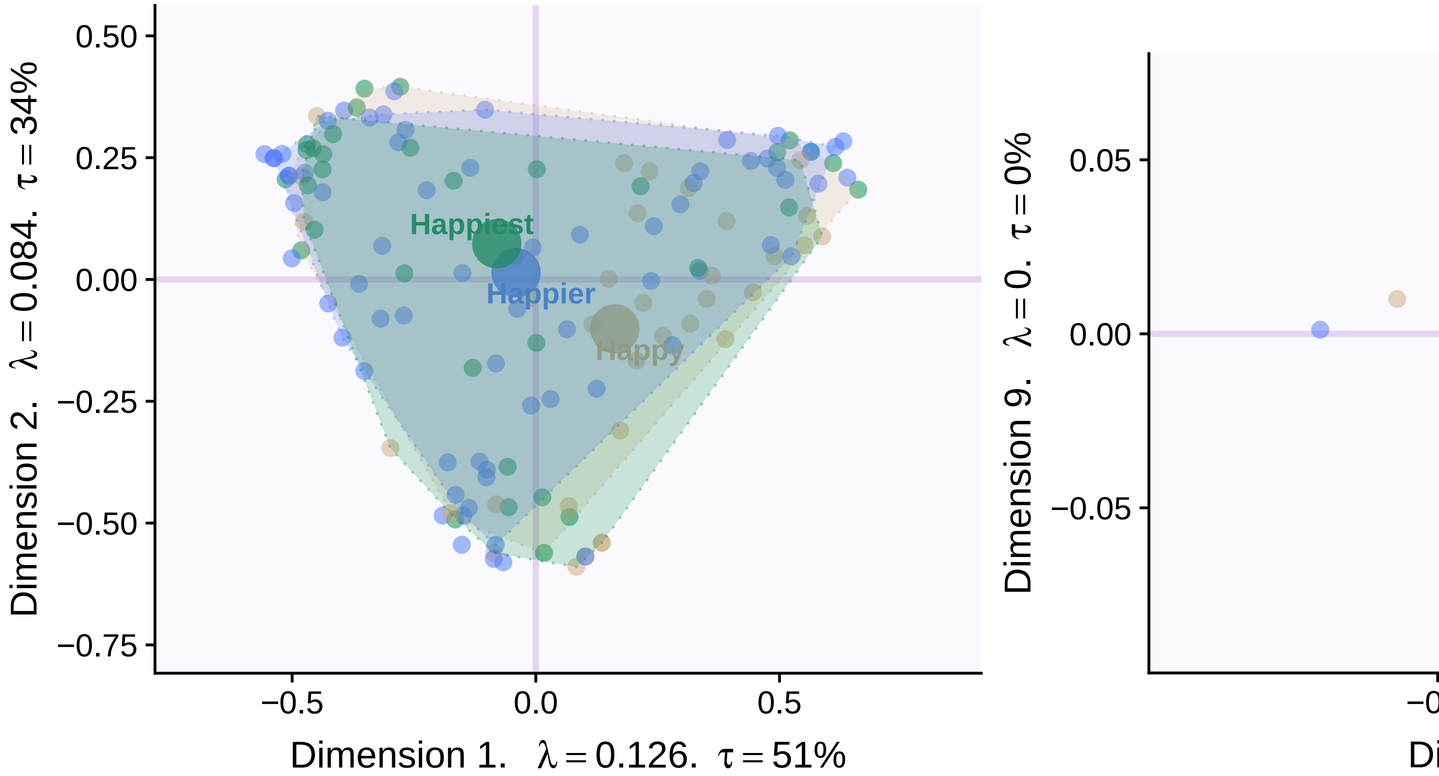
Lets visualize happiness categories for components 1-10, to make a decision (visually) on the most important components.



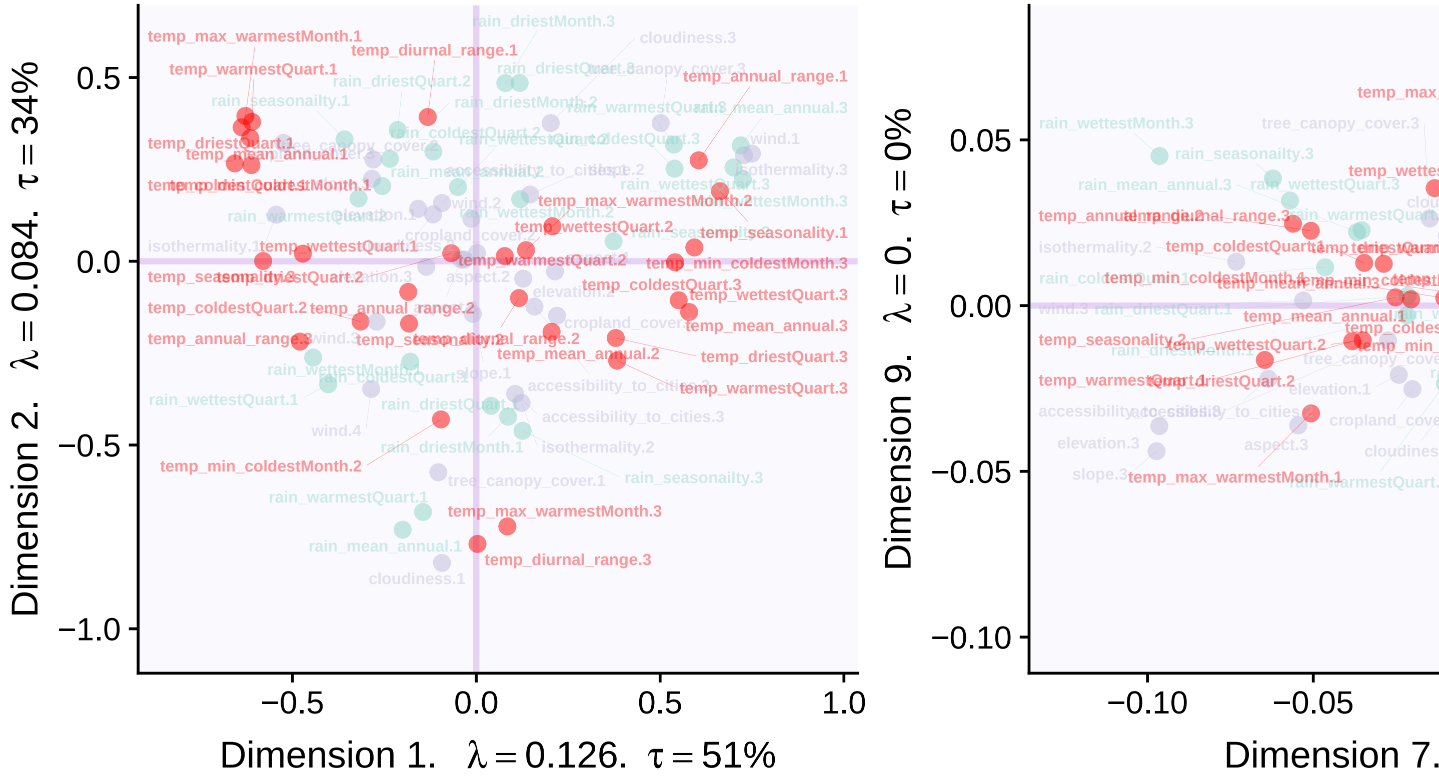
With Confidence Interval



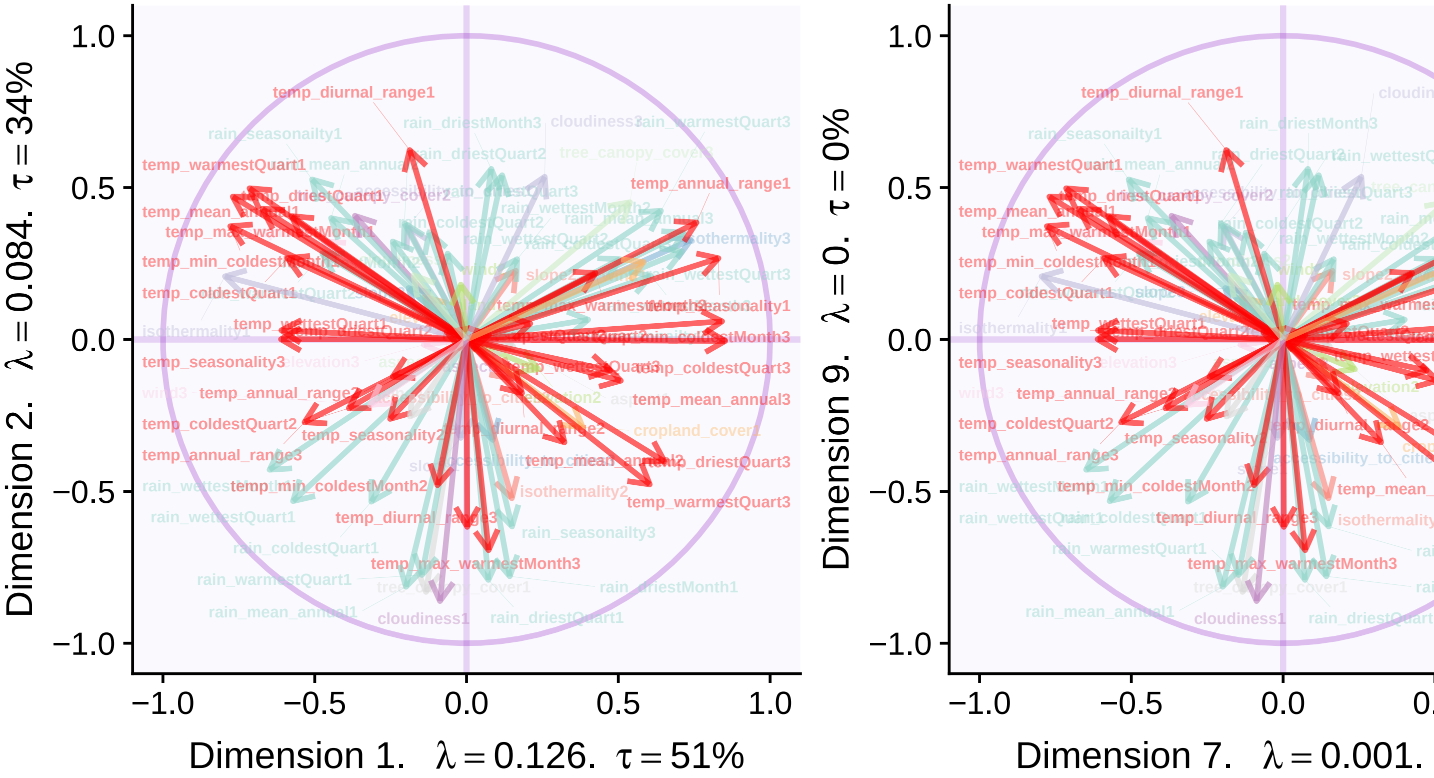
With Tolerance Interval


0.50
0.25
0.00
−0.25
−0.50
−0.75
Dimension 2. λ = 0.084. τ = 34%
Happiest
Happier
Happy
−0.5
0.0
0.5
Dimension 1. λ = 0.126. τ = 51%
0.05
0.00
−0.05
Dimension 9. λ = 0. τ = 0%
−0
Di

4.9 Loadings


0.5
0.0
−0.5
−1.0
Dimension 2. λ = 0.084. τ = 34%
temp_max_warmestMonth.1
temp_diurnal_range.1
temp_warmestQuart.1
rain_driestMonth.3
cloudiness.3
rain_driestQuart.2
tree_canopy_cover.3
temp_annual_range.1
rain_seasonailty.1
rain_driestMonth.2
rain_warmestQuart.3
rain_mean_annual.3
temp_driestQuart.1
temp_mean_annual.1
temp_min_coldestMonth.1
temp_coldestQuart.1
rain_coldestQuart.2
rain_wettestQuart.2
rain_coldestQuart.3
wind.1
isothermality.3
rain_mean_annual.2
accessibility_to_cities.2
slope.2
rain_wettestQuart.3
temp_max_warmestMonth.2
rain_wettestMonth.3
rain_warmestQuart.2
elevation.1
wind.2
rain_wettestMonth.2
temp_wettestQuart.2
temp_seasonality.1
isothermality.1
temp_wettestQuart.1
cropland_cover.2
temp_warmestQuart.2
temp_min_coldestMonth.3
temp_seasonality.3
temp_driestQuart.2
aspect.2
temp_coldestQuart.3
elevation.2
temp_wettestQuart.3
temp_coldestQuart.2
temp_annual_range.2
cropland_cover.1
temp_mean_annual.3
temp_annual_range.3
wind.3
temp_seasonality.2
temp_diurnal_range.2
temp_mean_annual.2
temp_driestQuart.3
rain_wettestMonth.1
rain_coldestQuart.1
slope.1
accessibility_to_cities.1
temp_warmestQuart.3
rain_wettestQuart.1
rain_driestQuart.1
accessibility_to_cities.3
wind.4
rain_driestMonth.1
isothermality.2
temp_min_coldestMonth.2
tree_canopy_cover.1
rain_seasonailty.3
rain_warmestQuart.1
temp_max_warmestMonth.3
rain_mean_annual.1
temp_diurnal_range.3
cloudiness.1
−0.5
0.0
0.5
1.0
Dimension 1. λ = 0.126. τ = 51%
0.05
0.00
−0.05
−0.10
Dimension 9. λ = 0. τ = 0%
rain_wettestMonth.3
tree_canopy_cover.3
rain_seasonailty.3
rain_mean_annual.3
rain_wettestQuart.3
temp_annual_range.3
temp_diurnal_range.2
isothermality.2
temp_coldestQuart.1
rain_coldestQuart.1
temp_min_coldestMonth.1
wind.3
rain_driestQuart.1
temp_mean_annual.1
temp_seasonality.2
rain_driestMonth.1
temp_wettestQuart.2
tree_canopy_cover
temp_warmestQuart.1
temp_driestQuart.2
elevation.1
accessibility_to_cities.3
accessibility_to_cities
cropland_cover
elevation.3
aspect.3
slope.3
temp_max_warmestMonth.1
rain_warmestQuart.
−0.10
−0.05
Dimension 7.

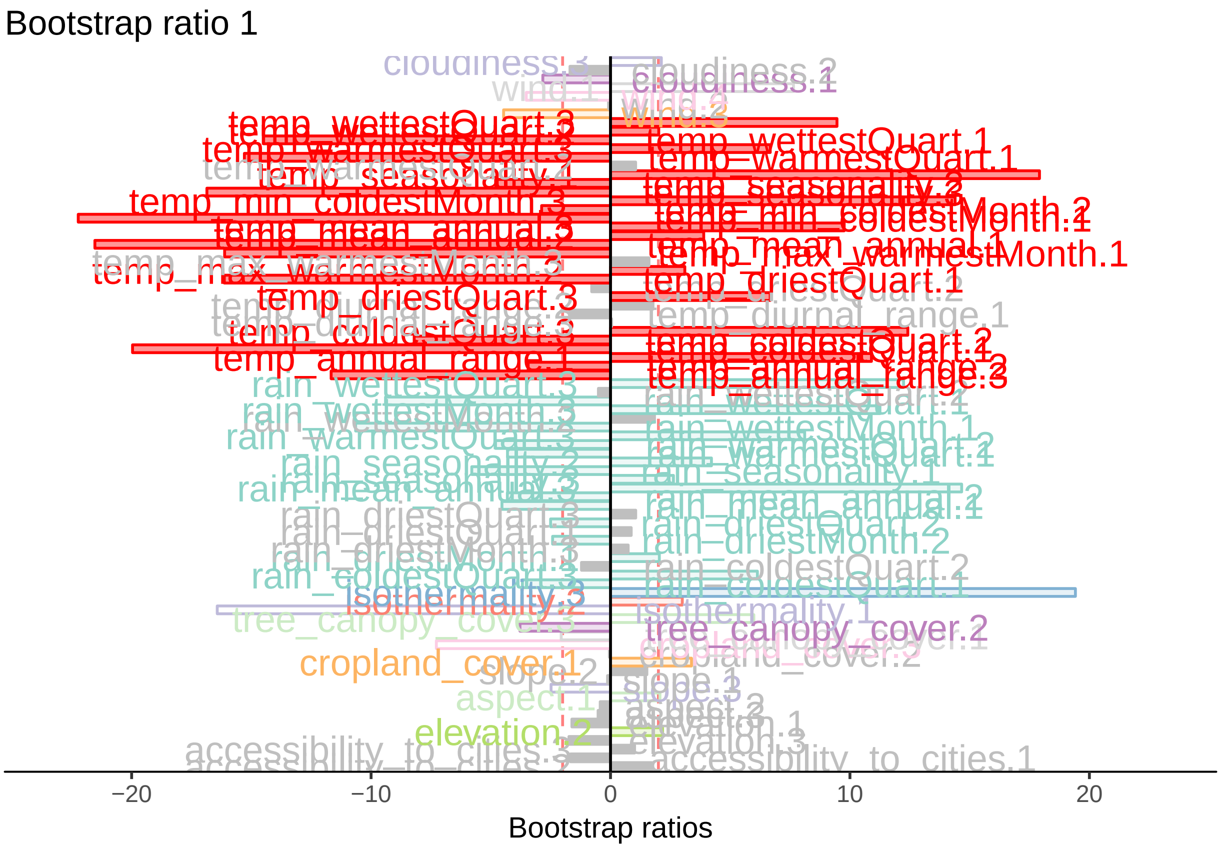
4.10 Loadings (correlation plot)

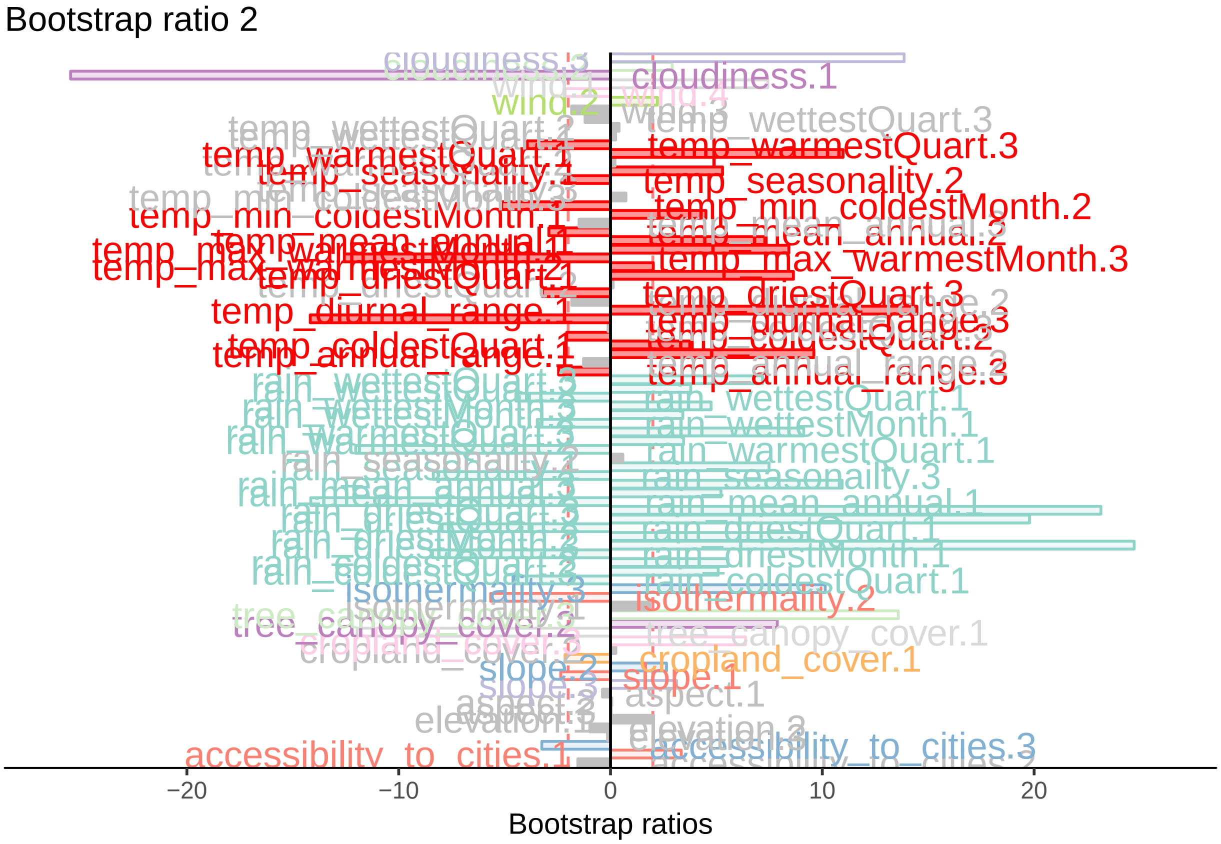


4.11 Most Contributing Variables (Inference)

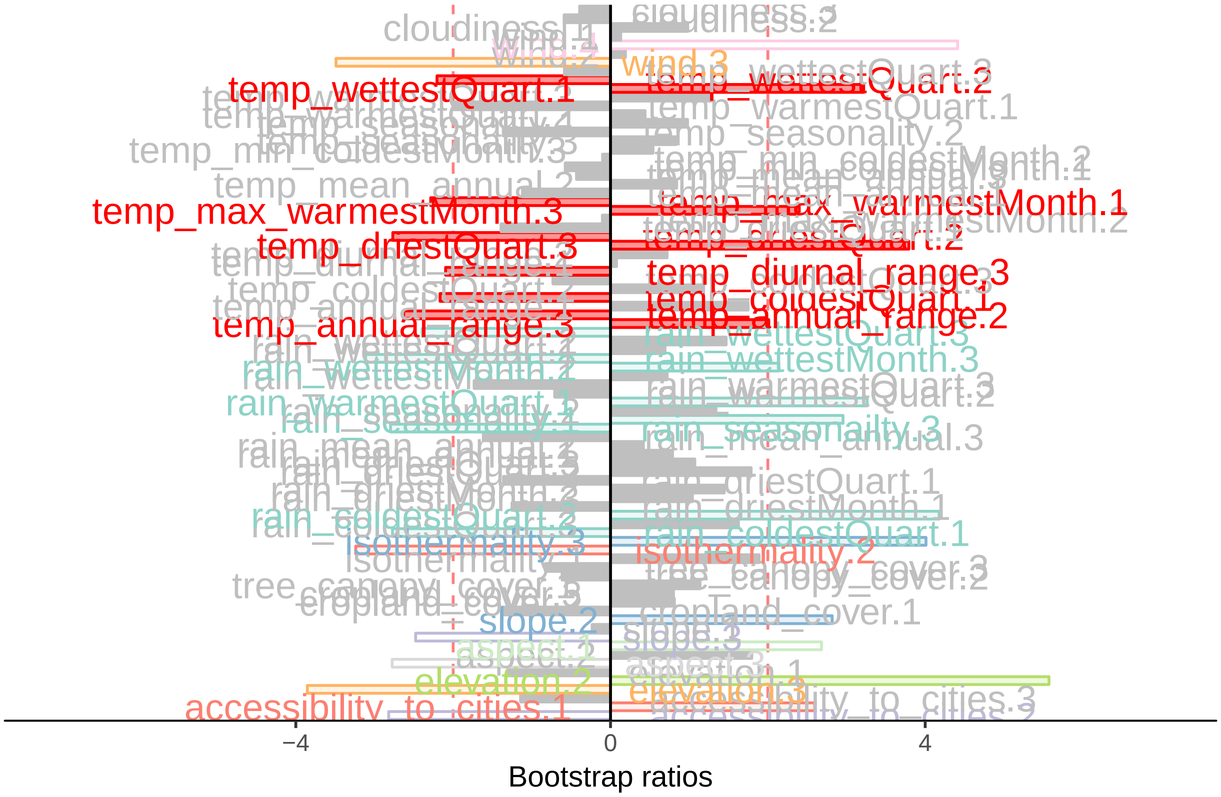
Let's plot variable contributions against each chosen components i.e. 1, 2, 7, 9.

- With Bootstrap Ratio

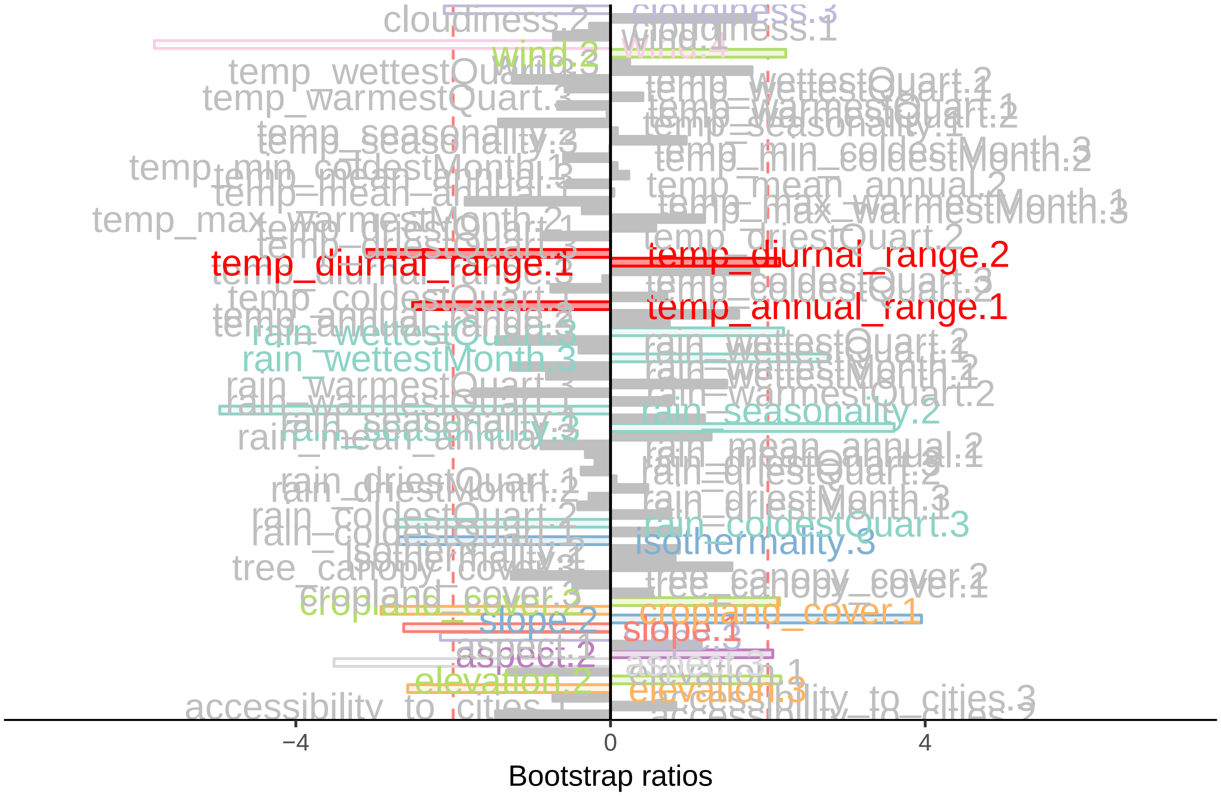

Bootstrap ratio 1
cloudiness.3
cloudiness.2
cloudiness.1
wind.1
wind.3
temp_wettestQuart.3
temp_wettestQuart.1
temp_warmestQuart.3
temp_warmestQuart.1
temp_seasonality.4
temp_seasonality.2
temp_min_coldestMonth.3
temp_min_coldestMonth.2
temp_min_coldestMonth.1
temp_mean_annual.3
temp_mean_annual.1
temp_max_warmestMonth.3
temp_max_warmestMonth.1
temp_driestQuart.3
temp_driestQuart.1
temp_driestQuart.2
temp_diurnal_range.3
temp_diurnal_range.1
temp_coldestQuart.3
temp_coldestQuart.1
temp_coldestQuart.2
temp_annual_range.1
temp_annual_range.2
temp_annual_range.3
rain_wettestQuart.3
rain_wettestQuart.1
rain_wettestMonth.3
rain_wettestMonth.1
rain_warmestQuart.3
rain_warmestQuart.2
rain_warmestQuart.1
rain_seasonality.2
rain_seasonality.1
rain_mean_annual.3
rain_mean_annual.2
rain_mean_annual.1
rain_driestQuart.3
rain_driestQuart.2
rain_driestMonth.3
rain_driestMonth.2
rain_coldestQuart.3
rain_coldestQuart.2
rain_coldestQuart.1
isothermality.3
isothermality.1
tree_canopy_cover.3
tree_canopy_cover.2
tree_canopy_cover.1
cropland_cover.1
cropland_cover.2
slope.2
slope.1
slope.3
aspect.1
aspect.3
aspect.2
elevation.2
elevation.1
elevation.3
accessibility_to_cities.3
accessibility_to_cities.1
−20
−10
0
10
20
Bootstrap ratios


Bootstrap ratio 2
cloudiness.3
cloudiness.2
cloudiness.1
wind.1
wind.2
wind.4
wind.3
temp_wettestQuart.2
temp_wettestQuart.1
temp_wettestQuart.3
temp_warmestQuart.1
temp_warmestQuart.2
temp_warmestQuart.3
temp_seasonality.1
temp_seasonality.3
temp_seasonality.2
temp_min_coldestMonth.3
temp_min_coldestMonth.1
temp_min_coldestMonth.2
temp_mean_annual.1
temp_mean_annual.2
temp_max_warmestMonth.1
temp_max_warmestMonth.2
temp_max_warmestMonth.3
temp_driestQuart.1
temp_driestQuart.2
temp_driestQuart.3
temp_diurnal_range.1
temp_diurnal_range.2
temp_diurnal_range.3
temp_coldestQuart.1
temp_coldestQuart.2
temp_annual_range.1
temp_annual_range.2
temp_annual_range.3
rain_wettestQuart.3
rain_wettestQuart.1
rain_wettestMonth.3
rain_wettestMonth.1
rain_warmestQuart.3
rain_warmestQuart.1
rain_seasonality.2
rain_seasonality.3
rain_mean_annual.3
rain_mean_annual.1
rain_driestQuart.3
rain_driestQuart.1
rain_driestMonth.3
rain_driestMonth.1
rain_coldestQuart.3
rain_coldestQuart.1
isothermality.3
isothermality.2
tree_canopy_cover.3
tree_canopy_cover.2
tree_canopy_cover.1
cropland_cover.3
cropland_cover.1
slope.2
slope.3
slope.1
aspect.2
aspect.1
elevation.1
elevation.3
accessibility_to_cities.1
accessibility_to_cities.3
accessibility_to_cities.2
−20
−10
0
10
20
Bootstrap ratios

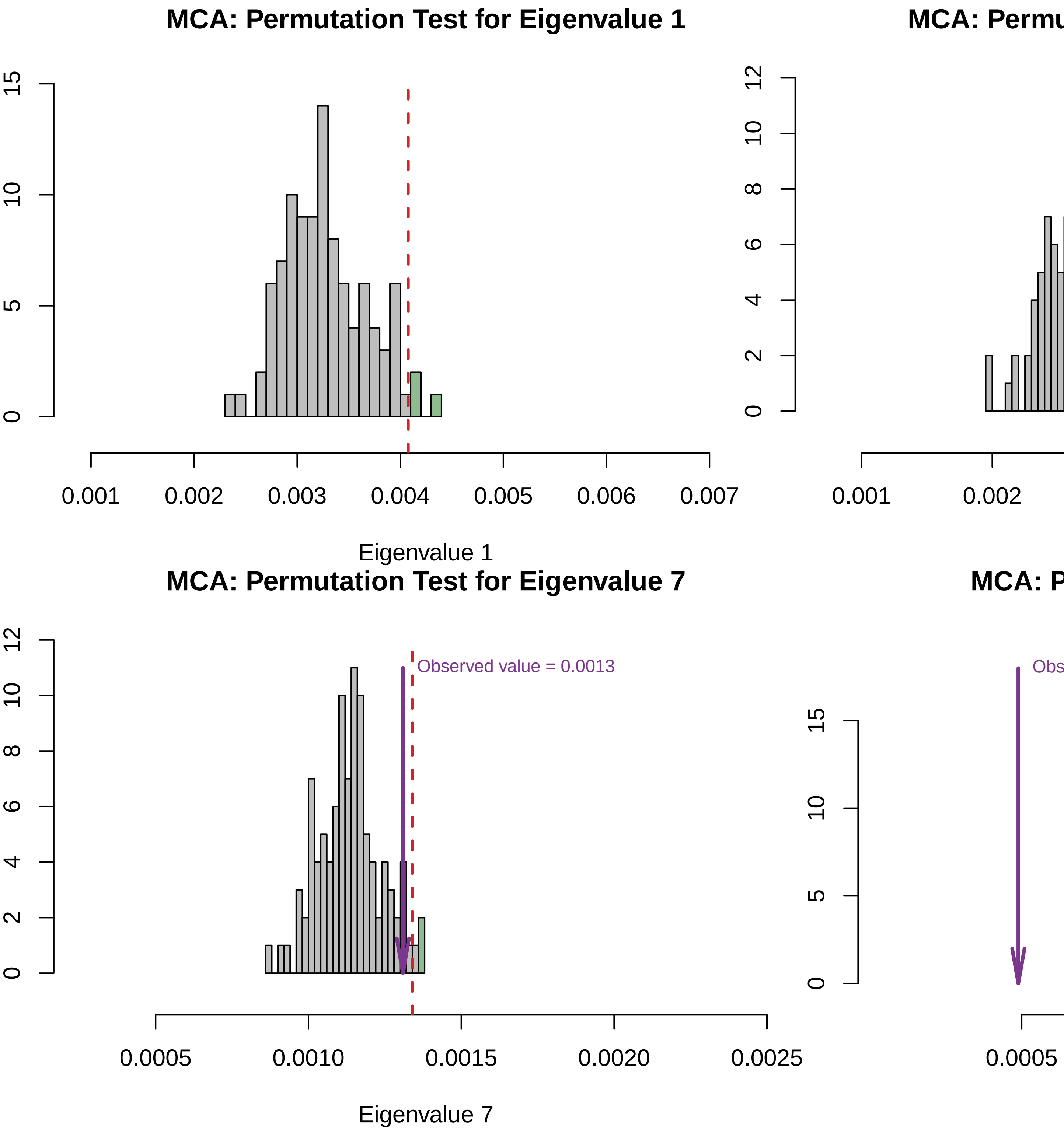
Bootstrap ratio 7



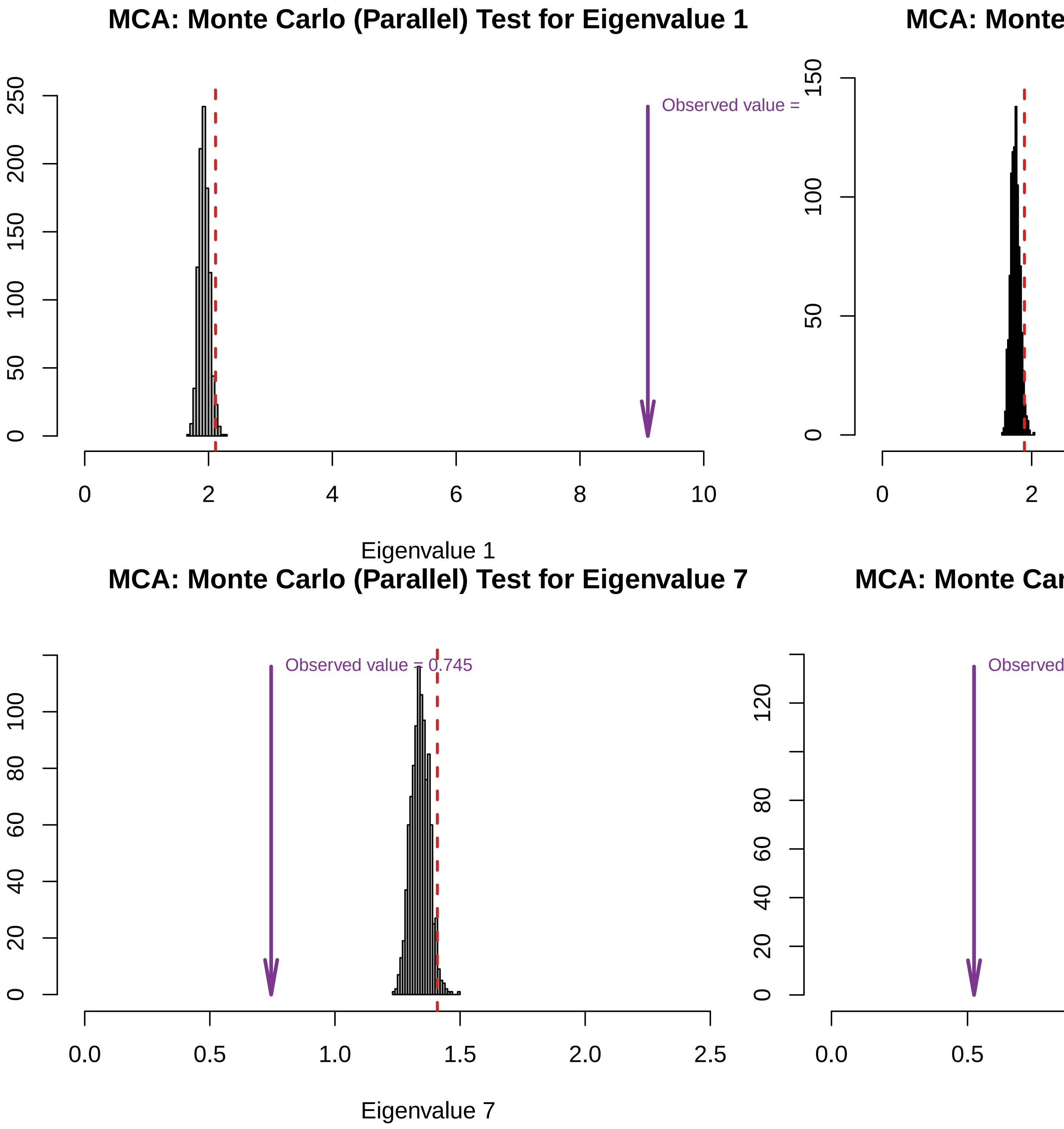
Bootstrap ratio 9



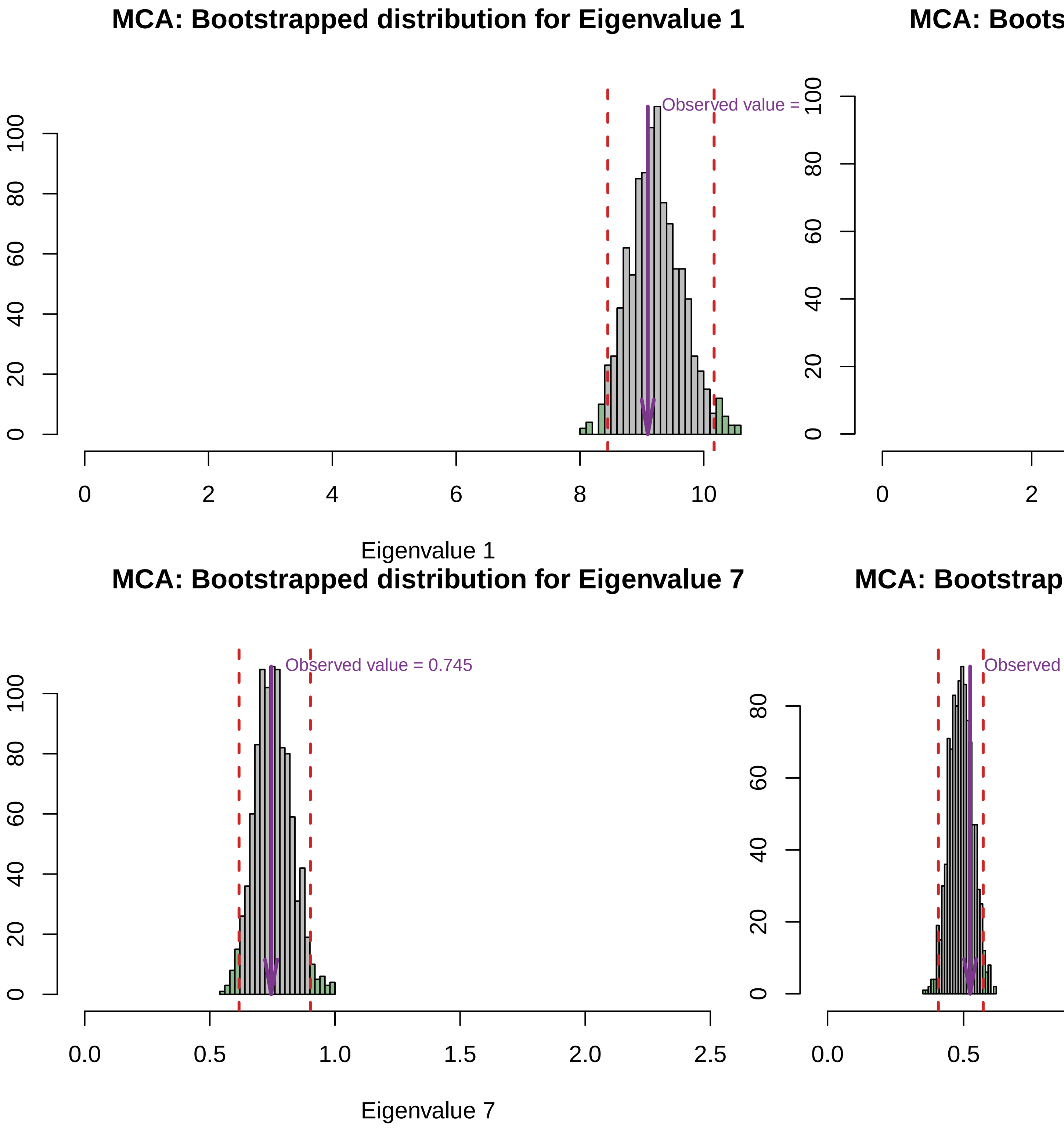
4.12 Permutation Test



4.13 Parallel Test



4.14 Bootstrap Test



4.15 Conclusion

- Component 1:
  - Rows: Normal & Happy
  - Columns: Cloudiness & Rain vs Cropland, Aspect, Elevation
  - Interpret: People in countries with more Cloudiness, Trees and Rain tends to be happier.
- Component 7:
  - Rows: Happy & Unhappy
  - Columns: Temp and Rain vs Accessibility and Cropland
  - Interpret: Rain and Temp seems to be main reason for unhappiness and Cropland is important for Happiness.
- Component 9:
  - Rows: Happy & Very Happy
  - Columns: Temp vs Rain
  - Interpret: Rain and Temp seems to be main reason for Happiness. *This contradicts with Component 7 and 1.*

Chapter 5

Partial Least Squares - Correlation

5.1 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.

- Research Question

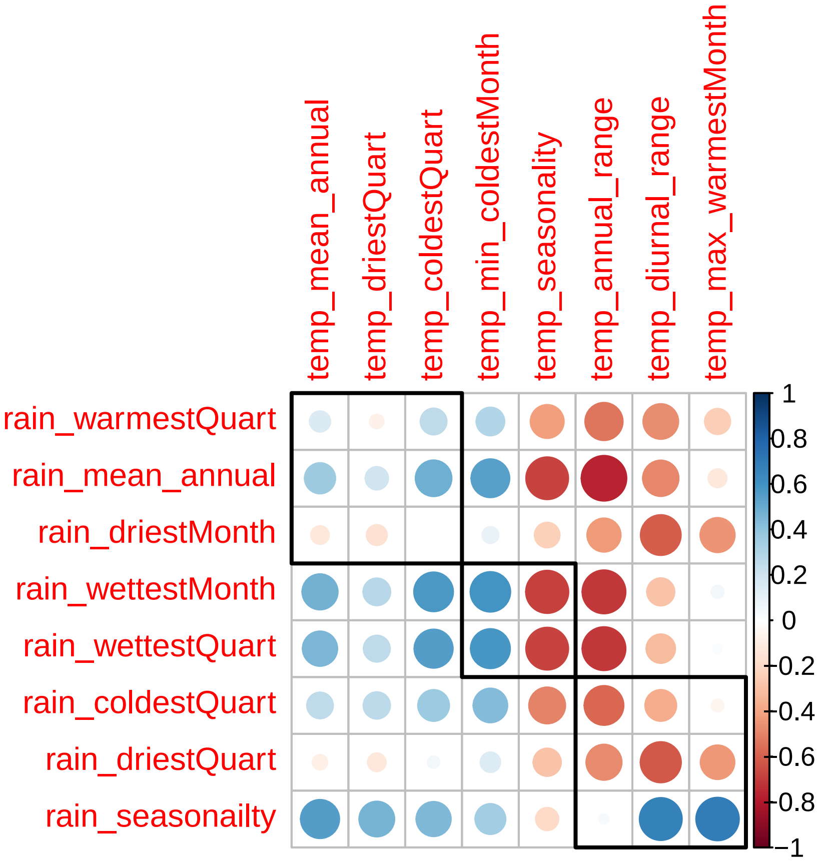
Which variables in Rain and Temperature contribute most towards happiness

5.2 Analysis

5.3 Correlation Plot

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

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



5.4 PLS-C

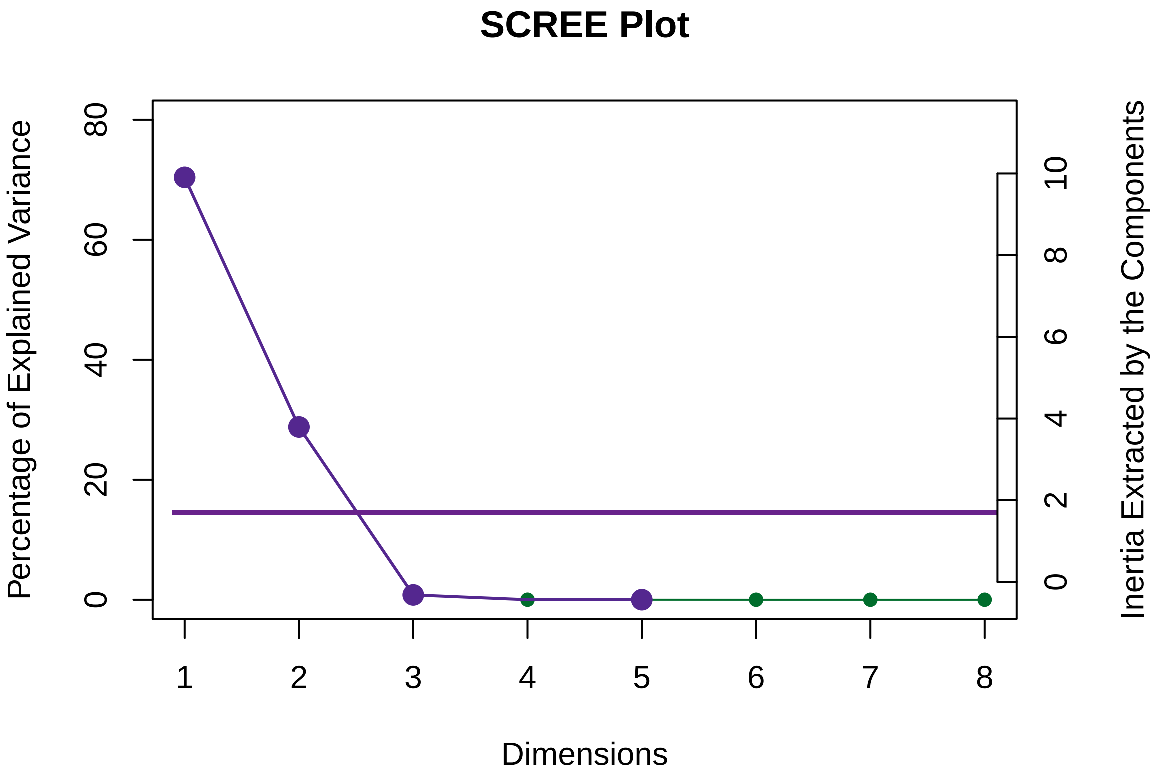
```
## ---------------------------------------------------------------------------
##  Results of Permutation Test for PLSC of X'*Y = R
##  for Omnibus Inertia and Eigenvalues
## ---------------------------------------------------------------------------
## $ fixedInertia      the Inertia of Matrix X
## $ fixedEigenvalues  an L*1 vector of the eigenvalues of X
## $ pOmnibus          the probablity associated to the Inertia
## $ pEigenvalues      an L* 1 matrix of p for the eigenvalues of X
## $ permInertia       vector of the permuted Inertia of X
## $ permEigenvalues   matrix of the permuted eigenvalues of X
## ---------------------------------------------------------------------------
```

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.

5.5 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.

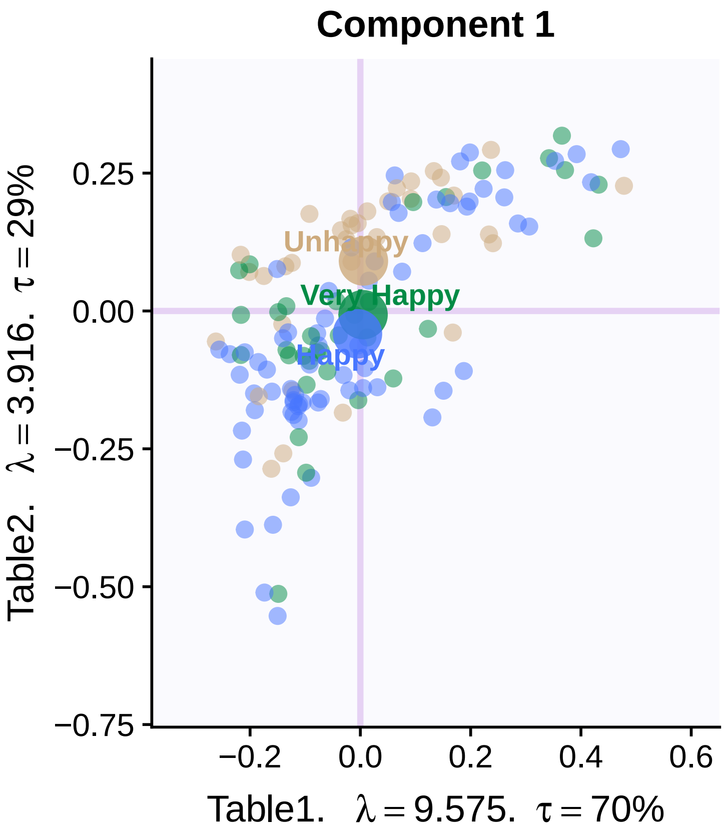
P.S. The most contribution component may not always be most useful for a given research question.



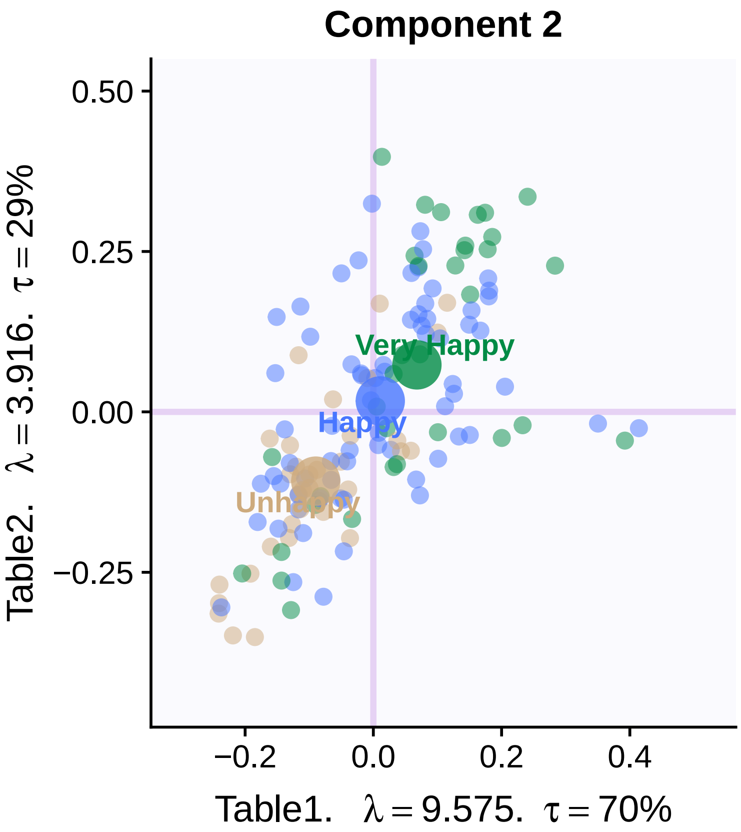
5.6 Latent Variables

Lets visualize happiness categories for Components 1 of each table

5.6.1 Component 1 for both Tables: Rain and Temperature

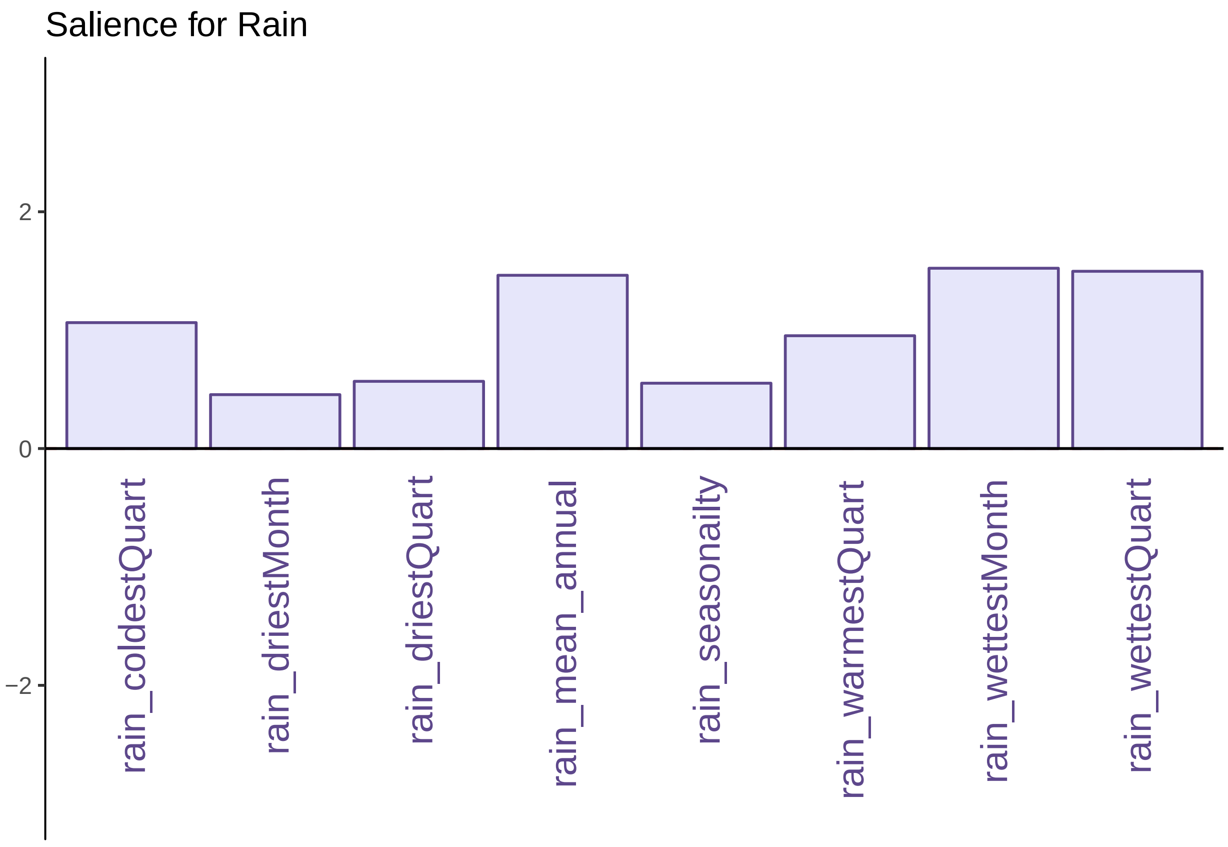


5.6.2 Component 2 for both Tables: Rain and Temperature

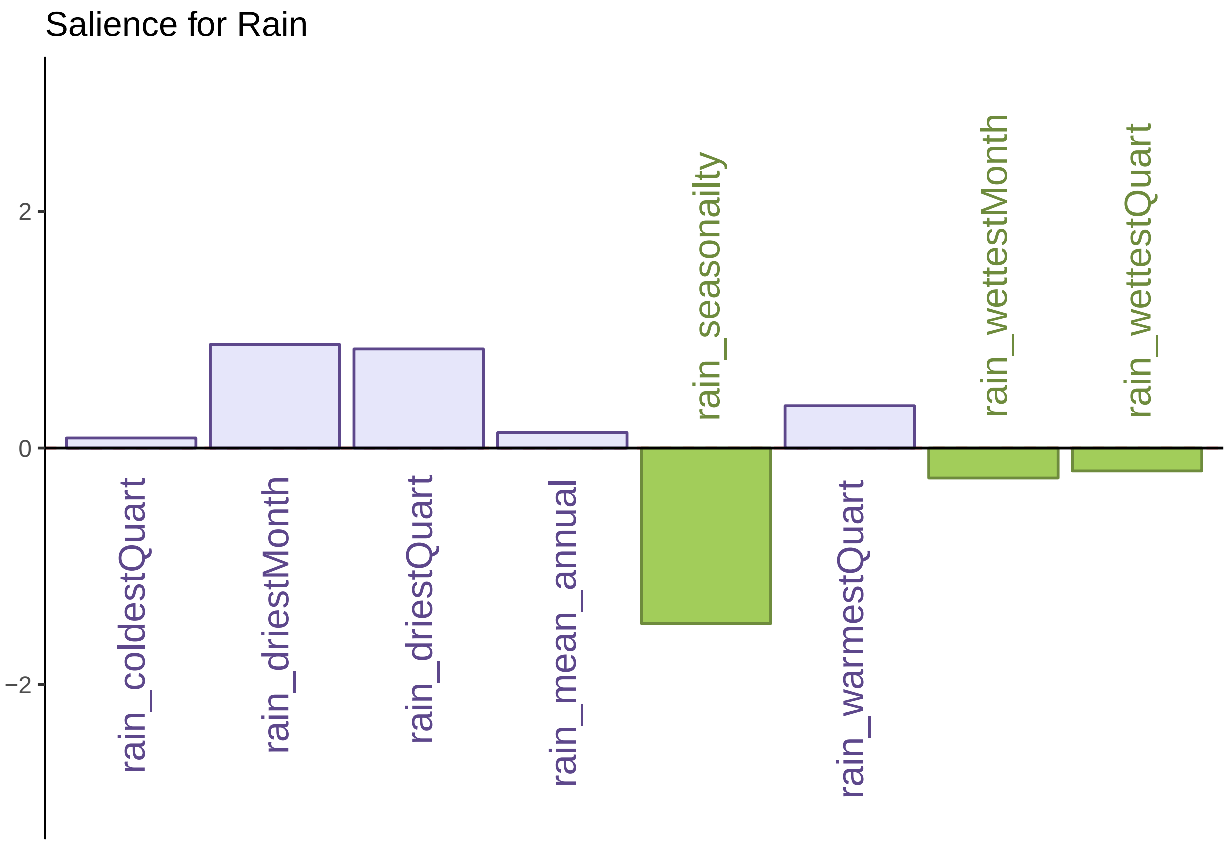


5.7 Salience for Rain

5.7.1 Components 1

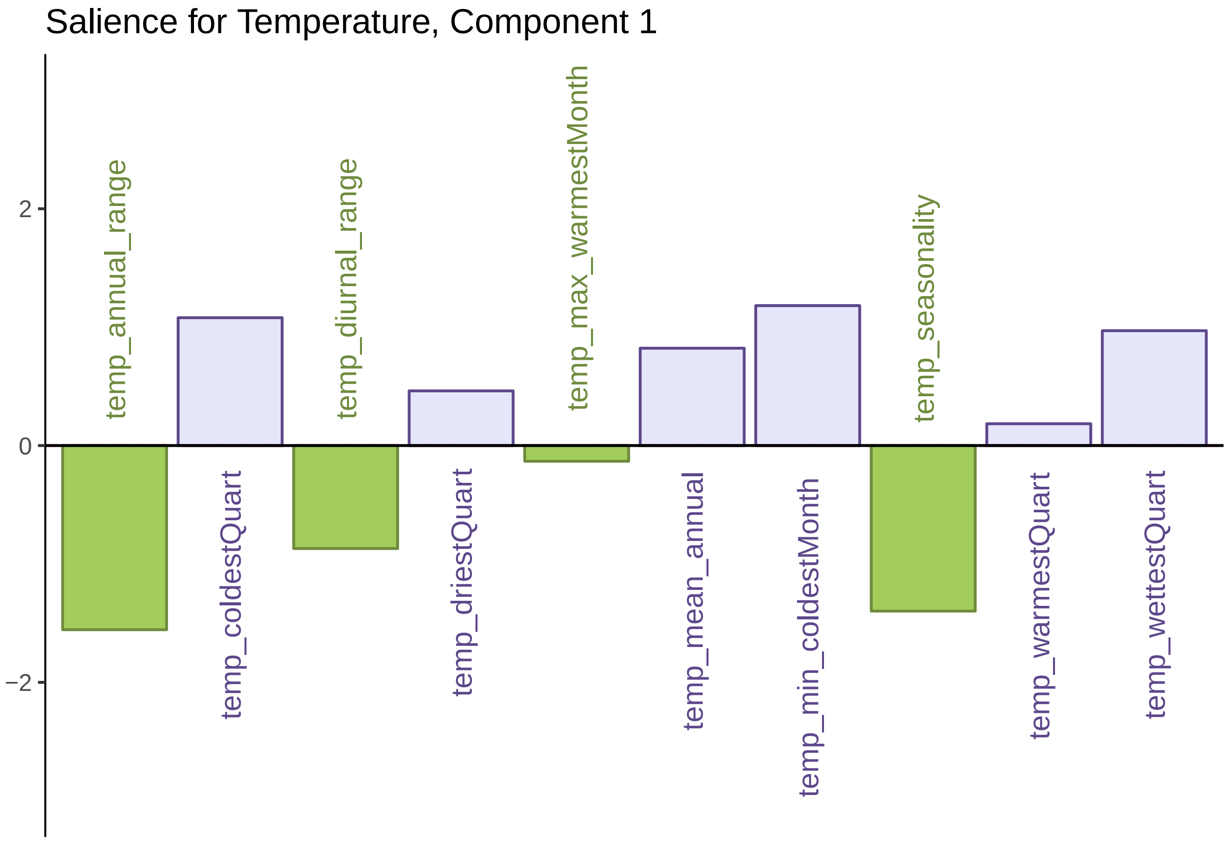


5.7.2 Component 2

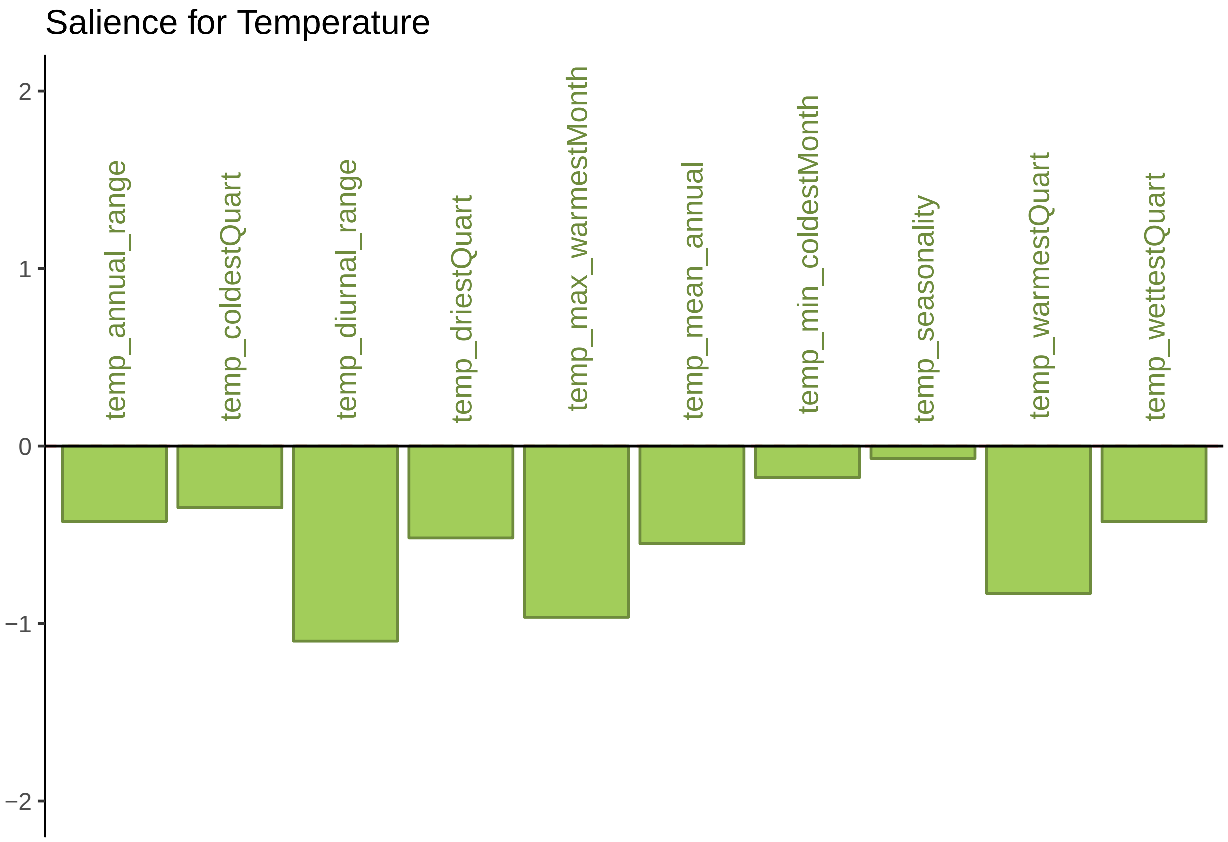


5.8 Salience for Temperature

5.8.1 Component 1



5.8.2 Component 2



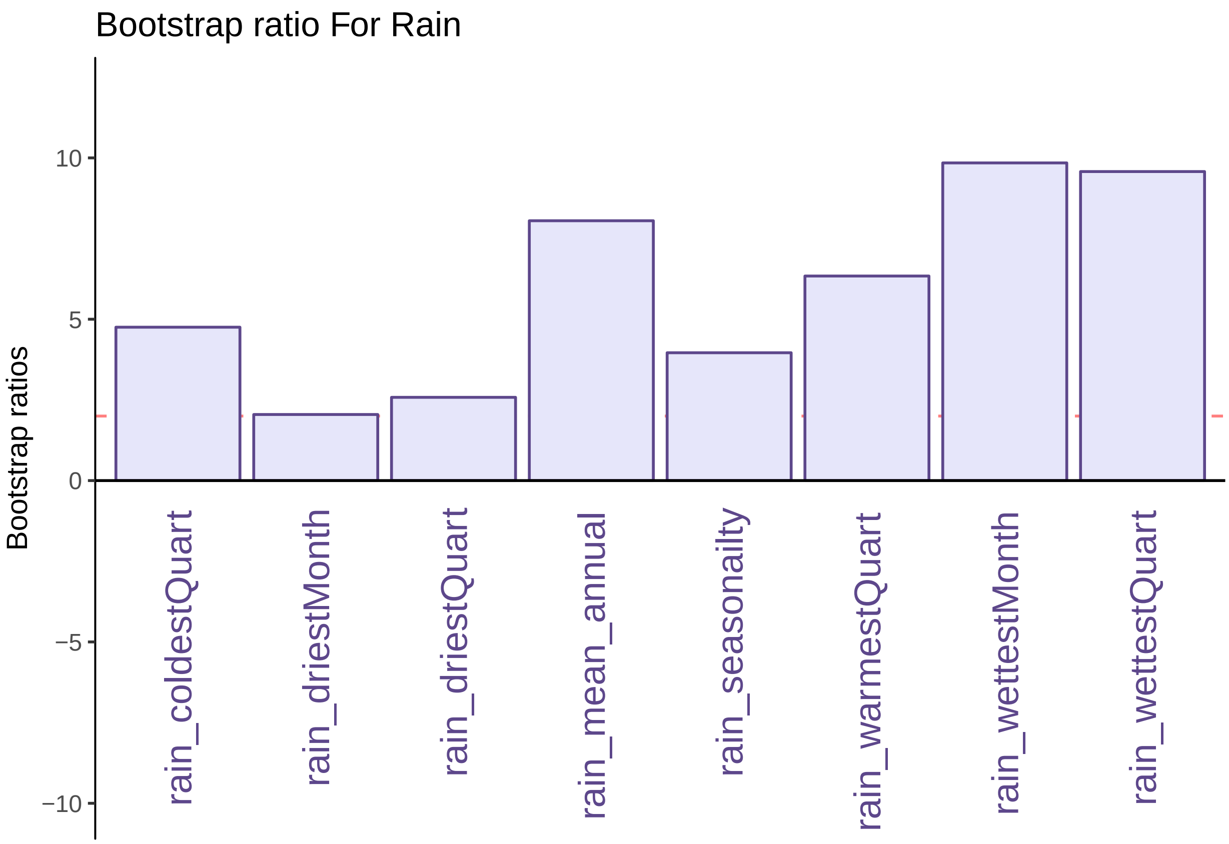
5.9 Most Contributing Variables - PLS-C (with Inference)

```
## ------------------------------------------------------------------------------
##  Bootstraped Factor Scores (BFS) and Bootstrap Ratios  (BR)
##  for the I and J-sets of a PLSC (obtained from multinomial resampling of X & Y)
## ------------------------------------------------------------------------------
## $ bootstrapBrick.i          an I*L*nIter Brick of BFSs  for the I-Set
## $ bootRatios.i              an I*L matrix of BRs for the I-Set
## $ bootRatiosSignificant.i   an I*L logical matrix for significance of the I-Set
## $ bootstrapBrick.j          a  J*L*nIter Brick of BFSs  for the J-Set
## $ bootRatios.j              a  J*L matrix of BRs for the J-Set
## $ bootRatiosSignificant.j   a  J*L logical matrix for significance of the J-Set
## ------------------------------------------------------------------------------
```

5.9.1 Bootstrap Test

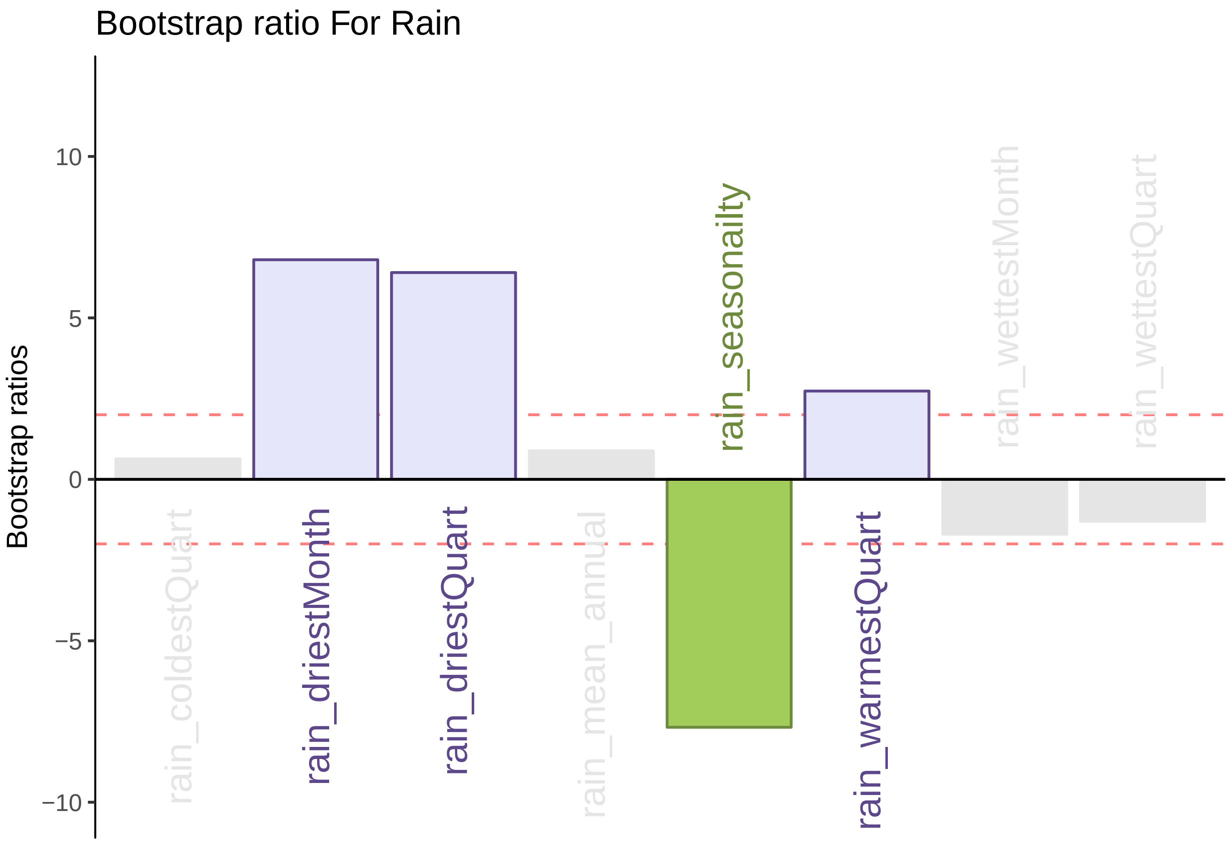
- Rain - Component 1

```
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)
              )
```



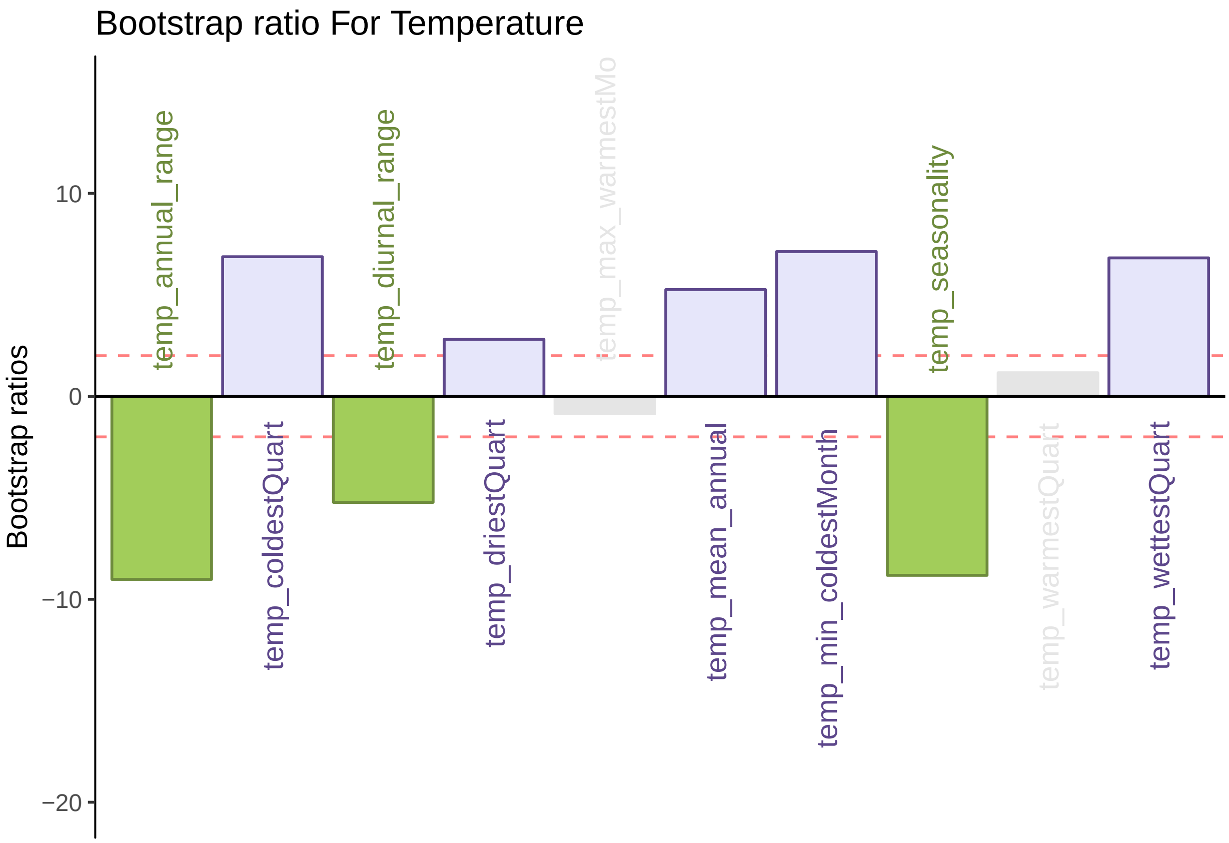
- Rain - Component 2

```
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)
              )
```



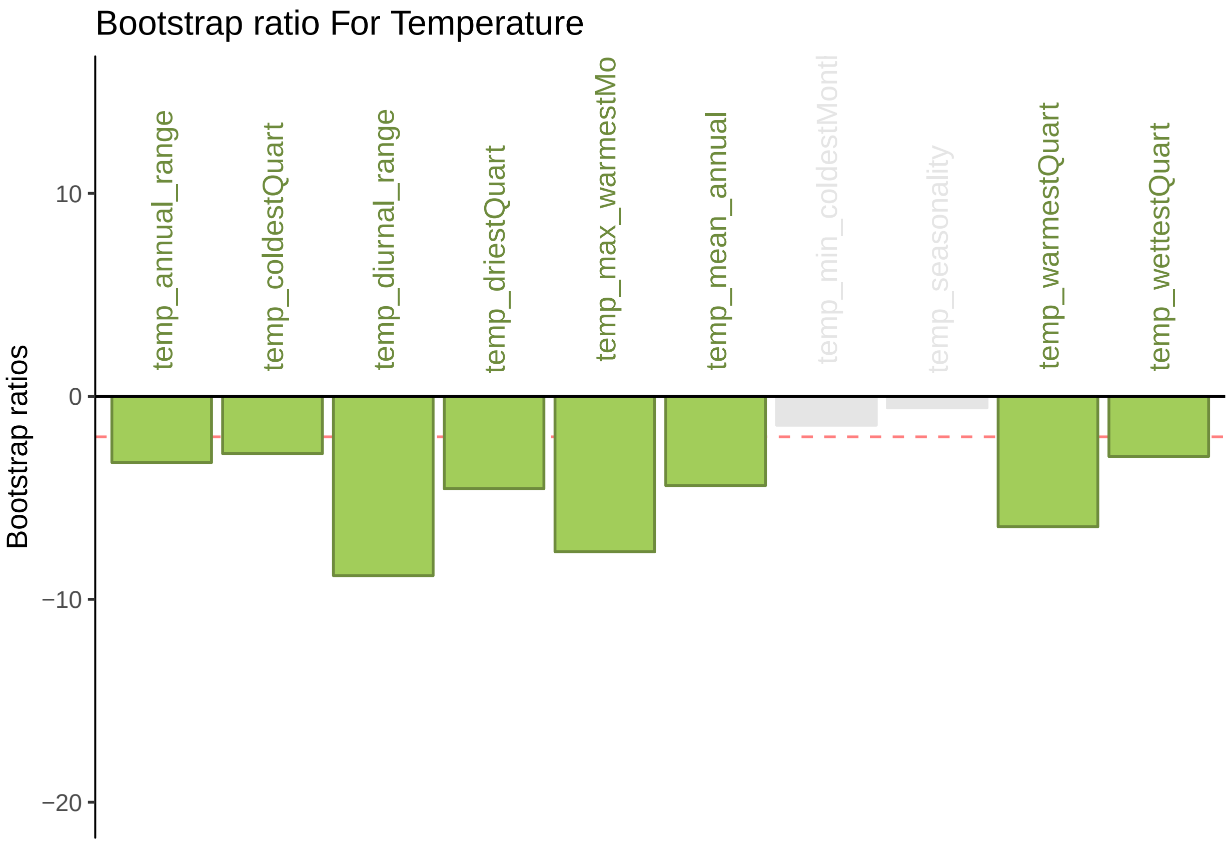
- Temperature - Component 1

```
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)
              )
```



5.10 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.

Chapter 6

Barycentric Discriminant Analysis

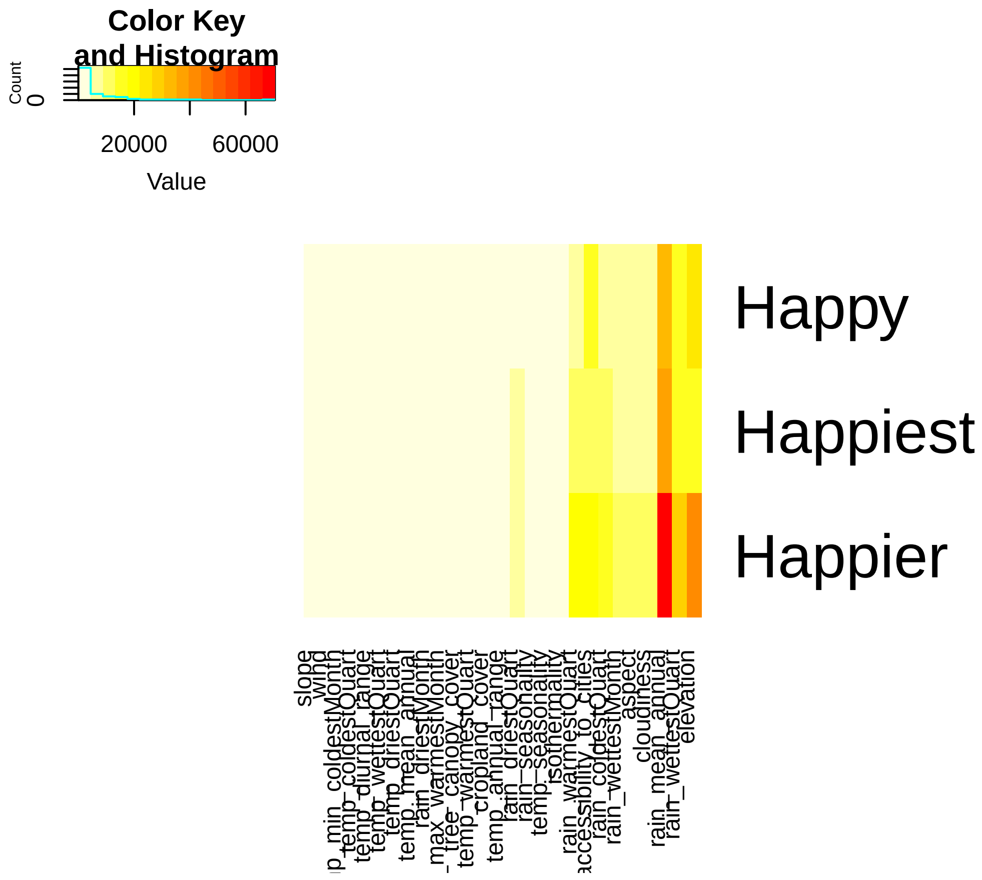
6.1 Method: BADA

6.2 Heatmap

Visually analyze multicollinearity in the system.

```
happiness_dummies = as.data.frame(dummy(country_env_df$Happiness_Rank))
colnames(happiness_dummies) <- c('Happiest', 'Happier', 'Happy' )

#heatmap(t(happiness_dummies) %*% as.matrix(country_env_df_for_pca))
```

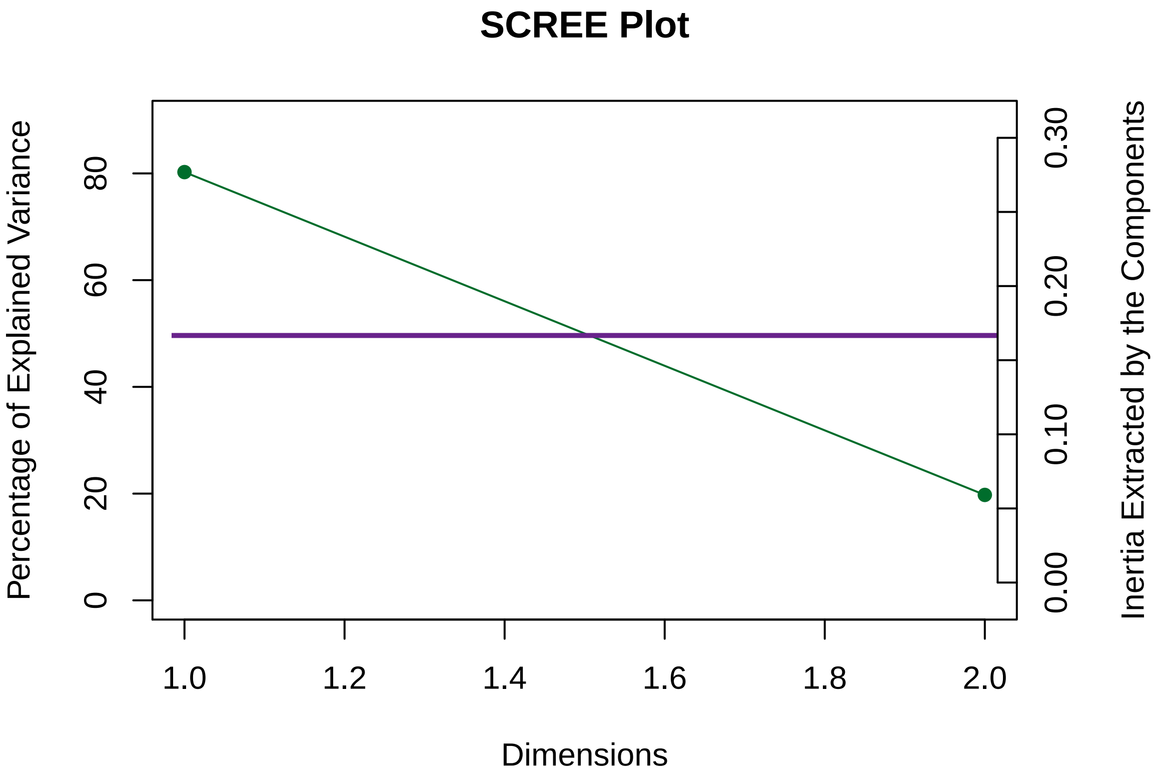
```
heatmap.2(t(happiness_dummies) %*% as.matrix(country_env_df_for_pca), col = rev(heat.colors(16)), dendr
```



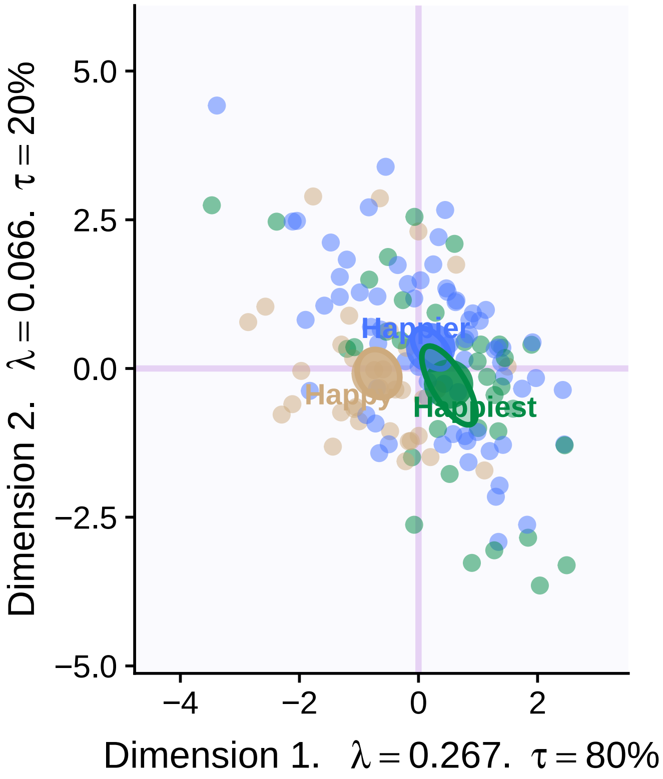
6.3 Scree Plot

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.

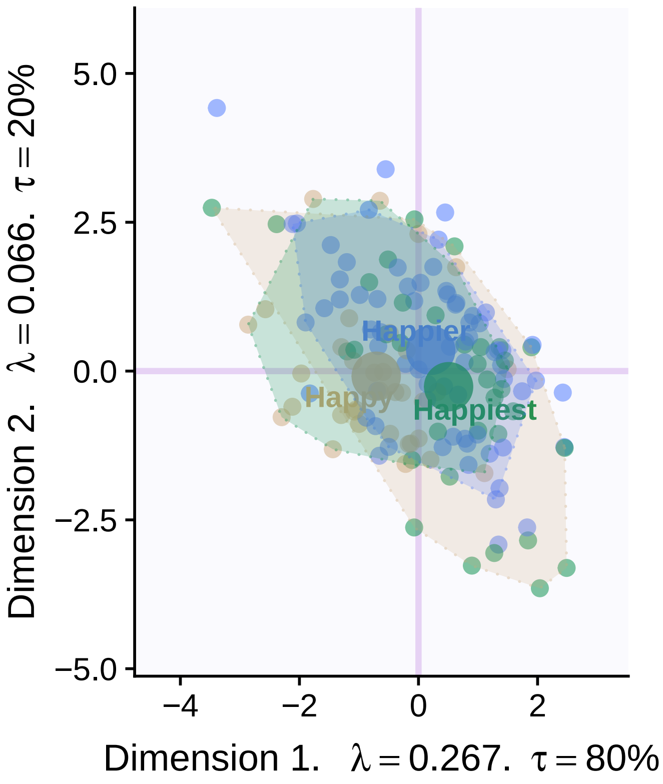
P.S. The most contribution component may not always be most useful for a given research question.



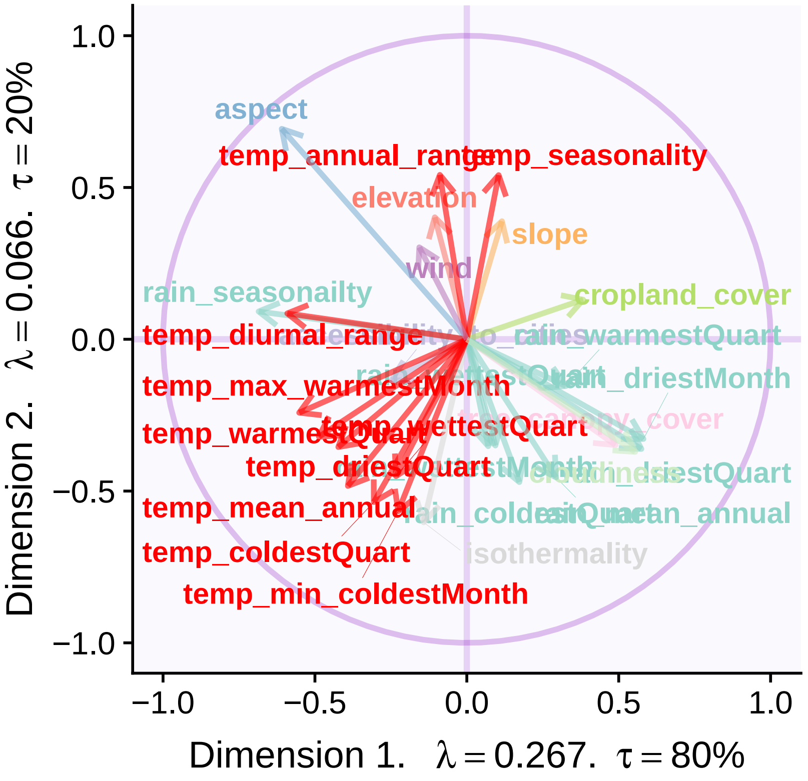
6.4 Factor Scores



- With Tolerance Interval

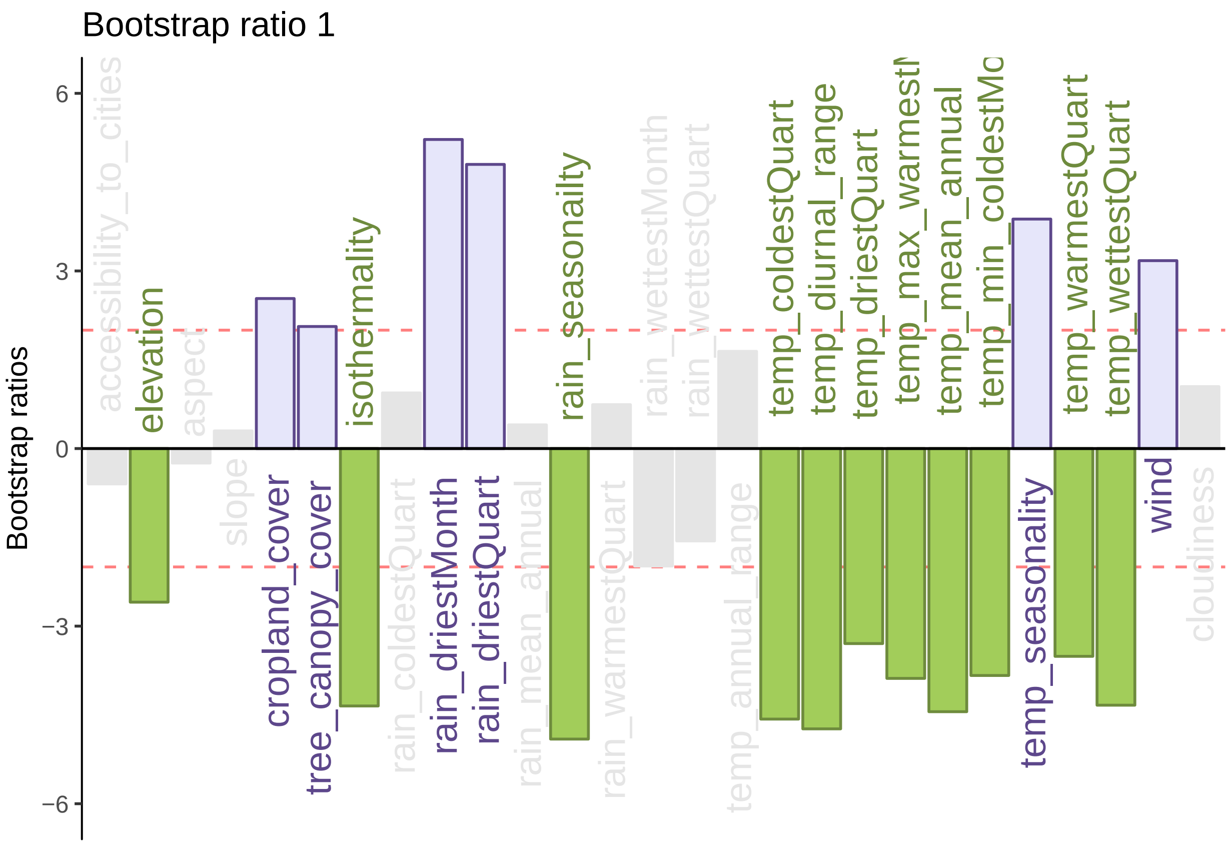

Happier
Happy
Happiest
5.0
2.5
0.0
−2.5
−5.0
−4
−2
0
2
Dimension 1. λ = 0.267. τ = 80%
Dimension 2. λ = 0.066. τ = 20%

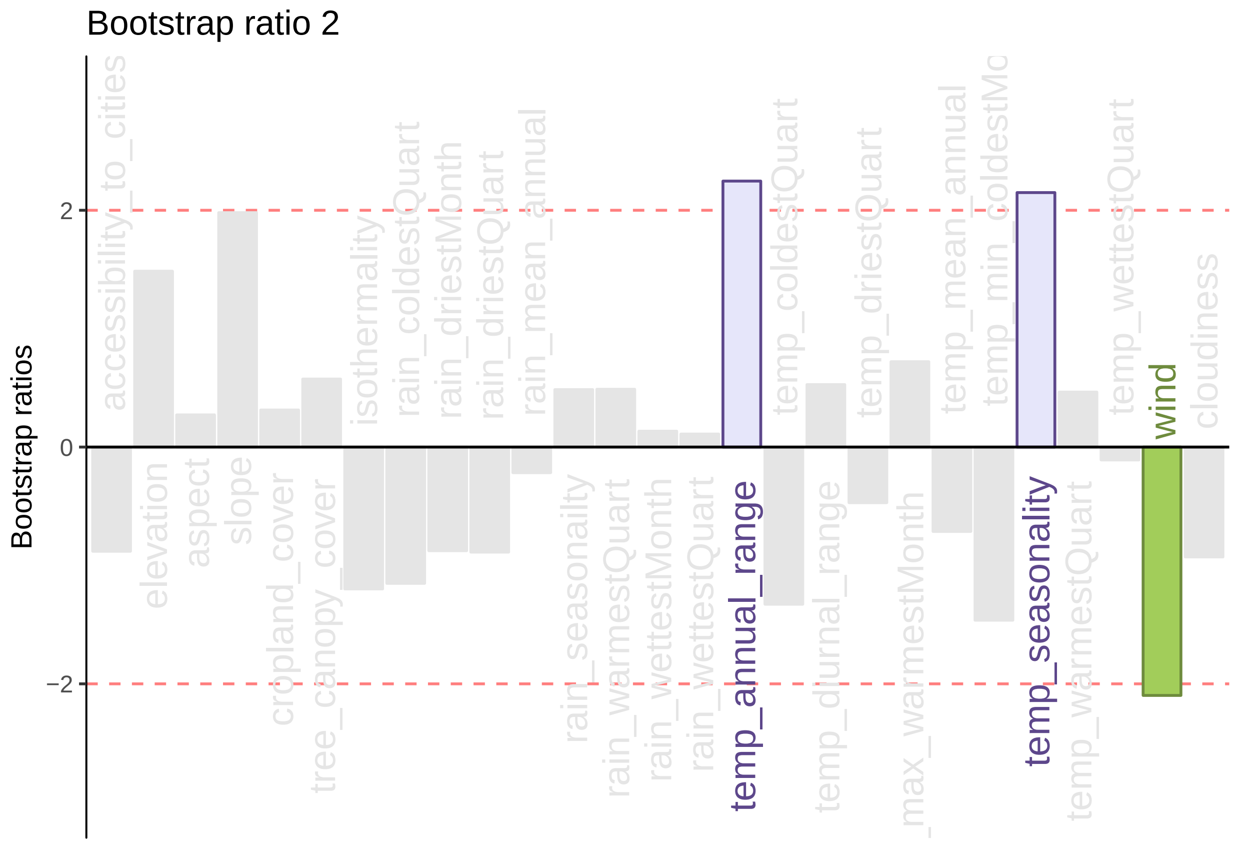
6.5 Loadings


aspect
temp_annual_range
temp_seasonality
elevation
slope
wind
rain_seasonailty
cropland_cover
temp_diurnal_range
accessibility_to_cities
rain_warmestQuart
rain_wettestQuart
rain_driestMonth
temp_max_warmestMonth
tree_canopy_cover
temp_warmestQuart
temp_wettestQuart
temp_driestQuart
rain_wettestMonth
cloudiness
rain_driestQuart
temp_mean_annual
rain_coldestQuart
rain_mean_annual
temp_coldestQuart
isothermality
temp_min_coldestMonth
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
Dimension 1. λ = 0.267. τ = 80%
Dimension 2. λ = 0.066. τ = 20%

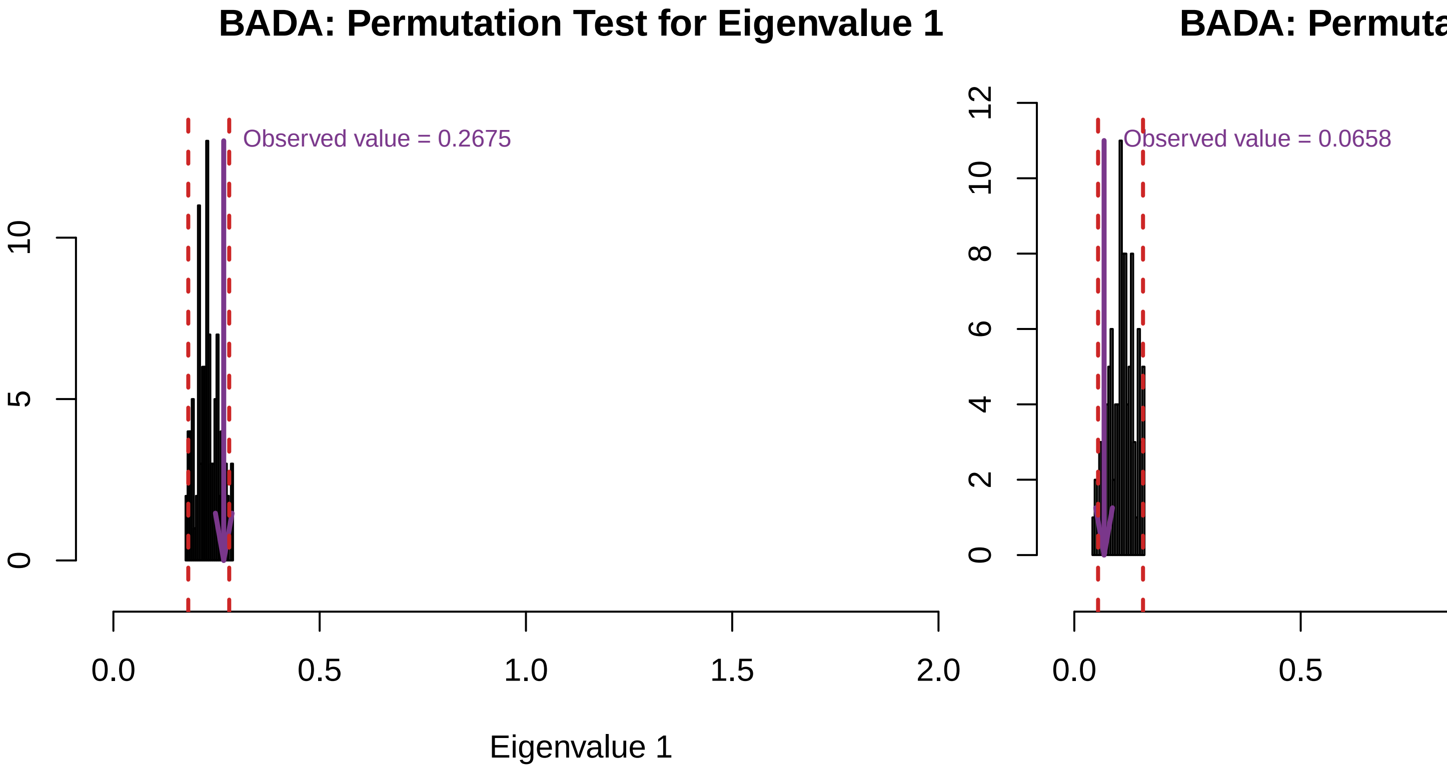
6.6 Most Contributing Variables

- With Bootstrap Ratio

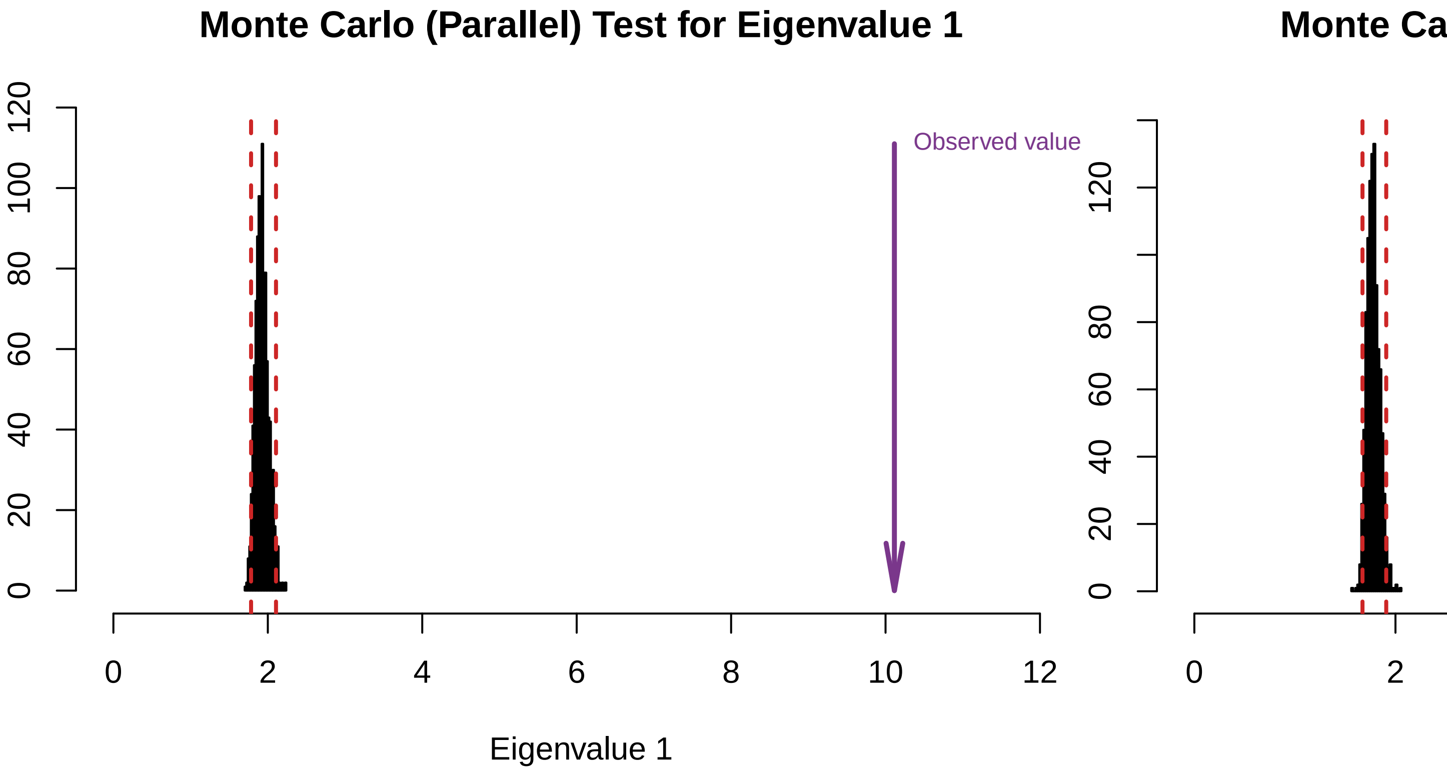



Bootstrap ratio 2
Bootstrap ratios
2
0
−2
accessibility_to_cities
elevation
aspect
slope
cropland_cover
tree_canopy_cover
isothermality
rain_coldestQuart
rain_driestMonth
rain_driestQuart
rain_mean_annual
rain_seasonality
rain_warmestQuart
rain_wettestMonth
rain_wettestQuart
temp_annual_range
temp_coldestQuart
temp_diurnal_range
temp_driestQuart
max_warmestMonth
temp_mean_annual
temp_min_coldestMo
temp_seasonality
temp_warmestQuart
temp_wettestQuart
wind
cloudiness

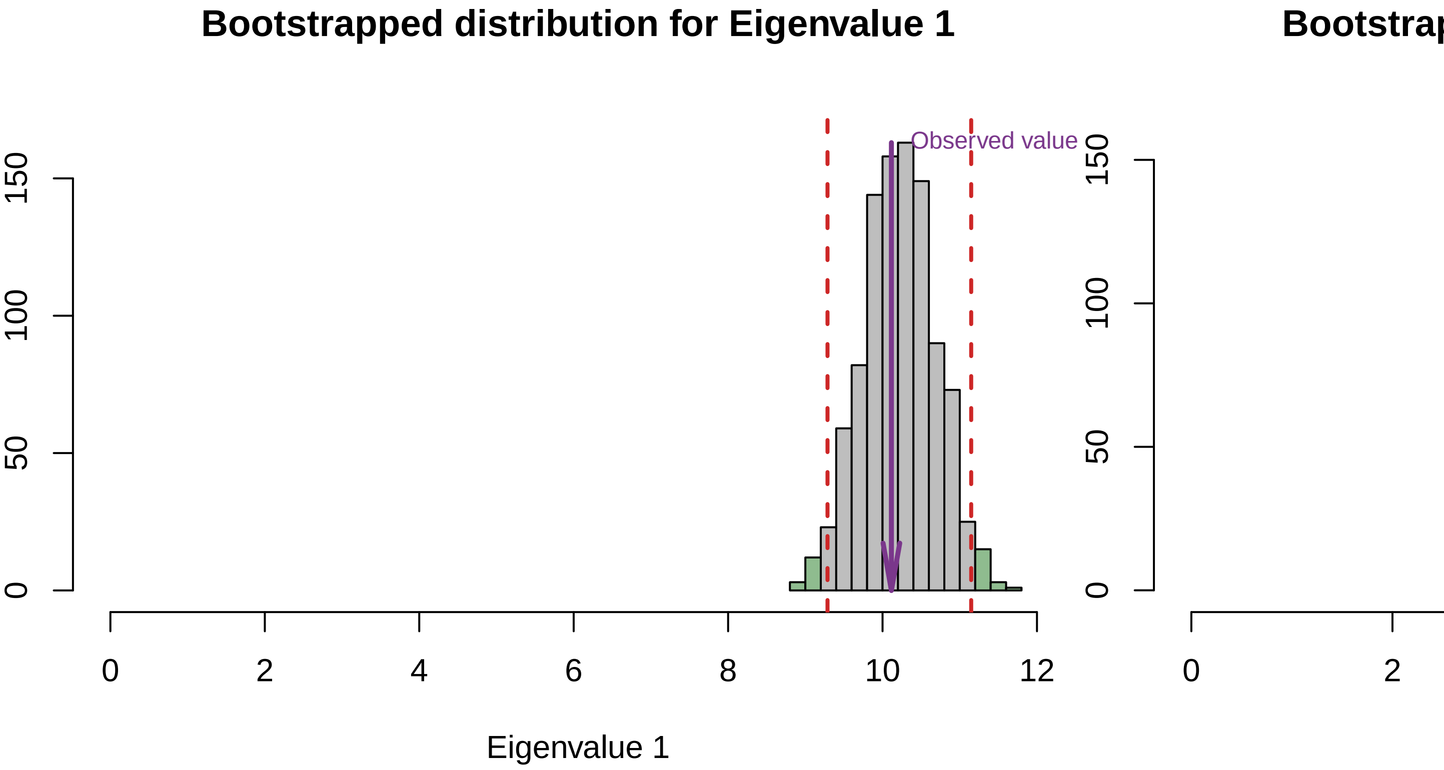
6.7 Permutation Test



6.8 Parallet Test



6.9 Bootstrap Test



6.10 Conclusion

- Component 1:
  - Rows: Normal & Happy
  - Columns: Cloudiness & Rain vs Cropland, Aspect, Elevation
  - Interpret: People in countries with more Cloudiness, Trees and Rain tends to be happier.
- Component 7:
  - Rows: Happy & Unhappy
  - Columns: Temp and Rain vs Accessibility and Cropland
  - Interpret: Rain and Temp seems to be main reason for unhappiness and Cropland is important for Happiness.
- Component 9:
  - Rows: Happy & Very Happy
  - Columns: Temp vs Rain
  - Interpret: Rain and Temp seems to be main reason for Happiness. *This contradicts with Component 7 and 1.*

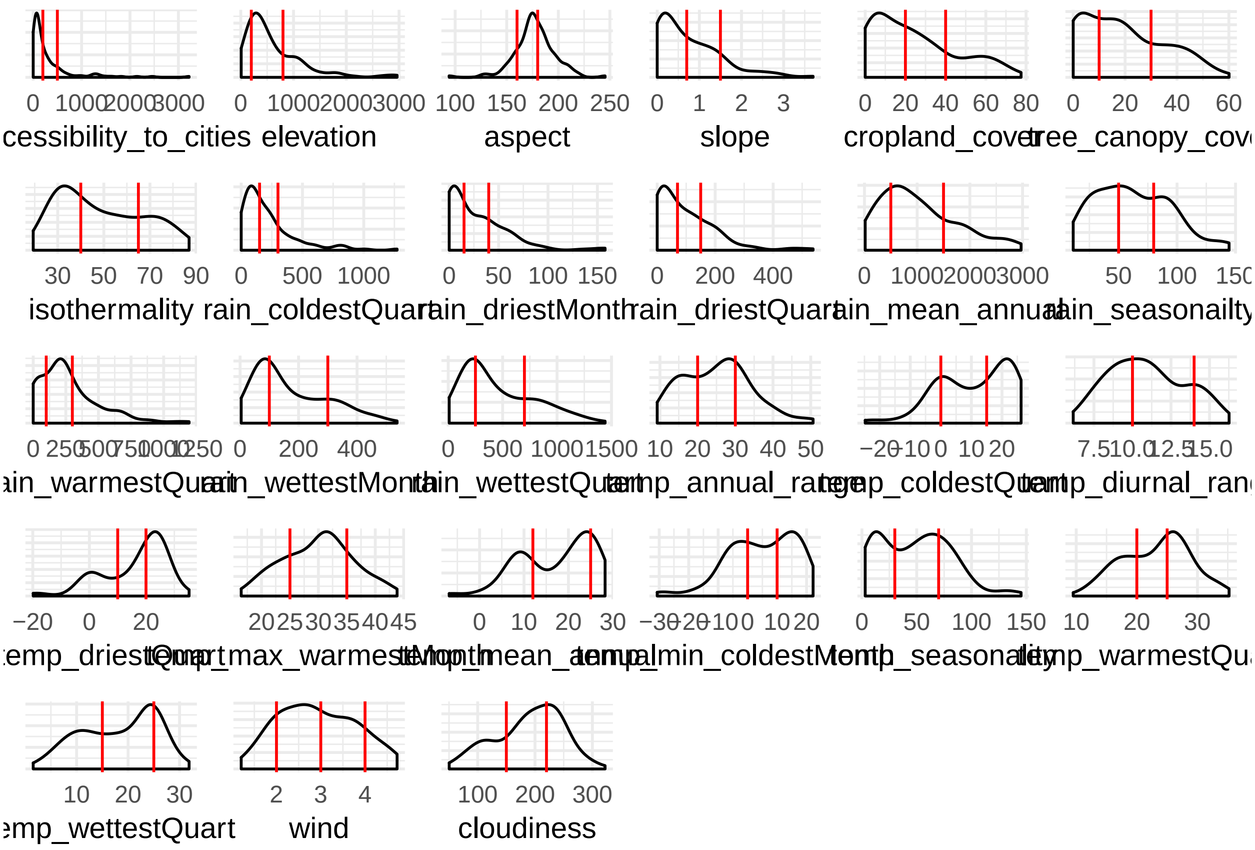
Chapter 7

Discriminant Correspondence Analysis

7.1 Method: DiCA

7.2 Density plot

Let's observe the distribution of each variables to get an intuition of how we can bin these variables. It's important to have nearly equal number of observations in the each bin and to try to cut the variables in a way to so that each new binned distribution is nearly Gaussian. We can also verify that our binning is appropiate by calculating Spearman Correlation for each of original variable and binned variable, the correlation coefficient should be close to 1.



7.3 Binning

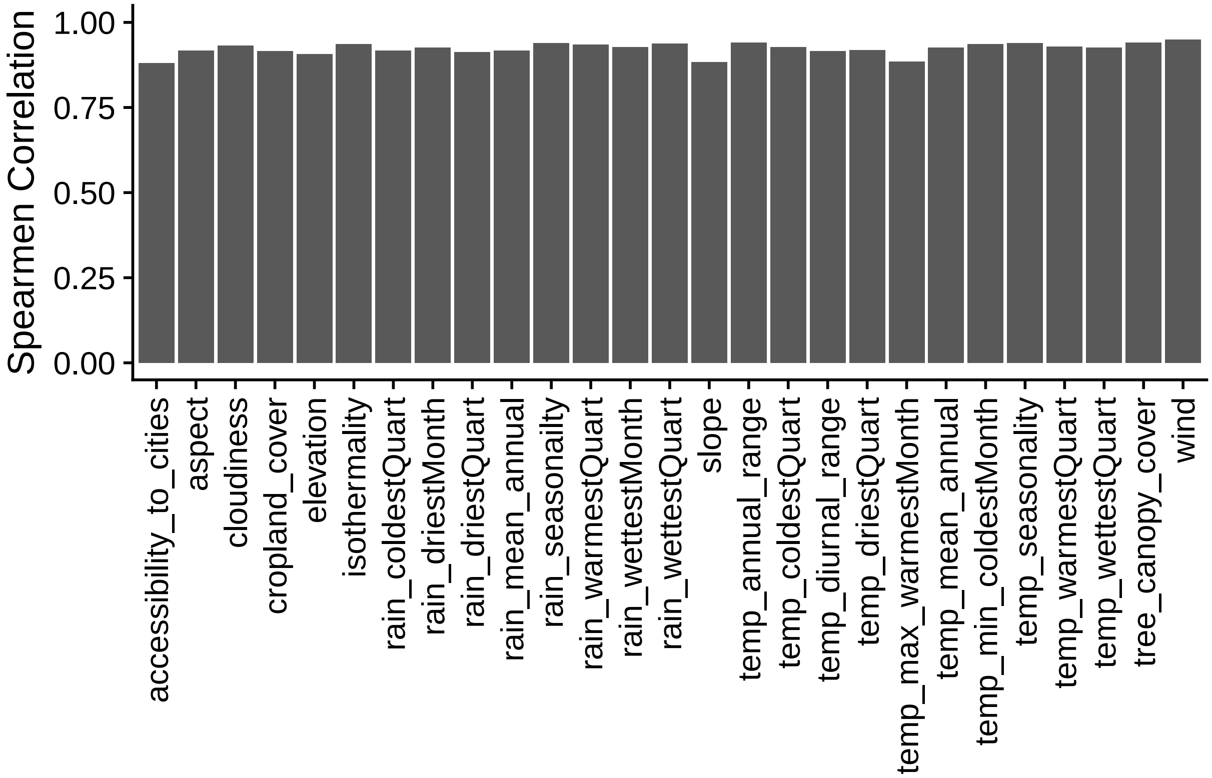
Structure of Data after binning based on above observation.

```
## 'data.frame':    137 obs. of  27 variables:
##  $ accessibility_to_cities: int  2 1 3 2 2 1 3 1 1 1 ...
##  $ elevation              : int  3 2 2 3 2 3 2 3 2 1 ...
##  $ aspect                 : int  3 3 3 2 1 3 3 2 1 2 ...
##  $ slope                  : int  3 3 1 1 1 3 1 2 2 1 ...
##  $ cropland_cover         : int  1 2 1 1 2 2 1 2 2 3 ...
##  $ tree_canopy_cover      : int  1 2 1 2 1 1 1 3 1 2 ...
##  $ isothermality          : int  1 1 2 2 2 1 2 1 1 2 ...
##  $ rain_coldestQuart      : int  1 3 1 1 1 1 1 2 1 1 ...
##  $ rain_driestMonth       : int  1 3 1 1 2 2 1 3 2 1 ...
##  $ rain_driestQuart       : int  1 2 1 1 1 1 1 3 1 1 ...
##  $ rain_mean_annual       : int  1 2 1 2 2 2 1 2 1 3 ...
##  $ rain_seasonailty       : int  3 1 2 3 1 1 2 1 1 3 ...
##  $ rain_warmestQuart      : int  1 2 1 3 2 2 2 3 1 3 ...
##  $ rain_wettestMonth      : int  1 2 1 2 1 1 1 2 1 3 ...
##  $ rain_wettestQuart      : int  1 2 1 2 1 1 1 2 1 3 ...
##  $ temp_annual_range      : int  3 2 3 2 2 3 2 2 3 2 ...
##  $ temp_coldestQuart      : int  1 2 2 3 2 1 2 1 2 3 ...
##  $ temp_diurnal_range     : int  3 1 3 2 2 2 2 1 1 1 ...
##  $ temp_driestQuart       : int  3 2 3 2 2 1 2 1 2 2 ...
##  $ temp_max_warmestMonth  : int  2 2 3 2 2 1 3 1 2 2 ...
##  $ temp_mean_annual       : int  1 1 2 2 2 1 2 1 1 3 ...
##  $ temp_min_coldestMonth  : int  1 1 2 2 2 1 2 1 1 3 ...
##  $ temp_seasonality       : int  3 2 3 1 2 3 2 2 3 2 ...
##  $ temp_warmestQuart      : int  2 1 3 2 2 1 3 1 2 3 ...
##  $ temp_wettestQuart      : int  1 1 2 2 2 1 2 1 1 3 ...
##  $ wind                   : int  3 2 4 2 4 1 4 2 2 2 ...
##  $ cloudiness             : int  1 2 1 2 2 2 1 3 2 2 ...
```

7.4 Spearman Correlation

Let's observe correlation between original data and binned data to make sure that neither the correlation ceofficient is too low or perfect.

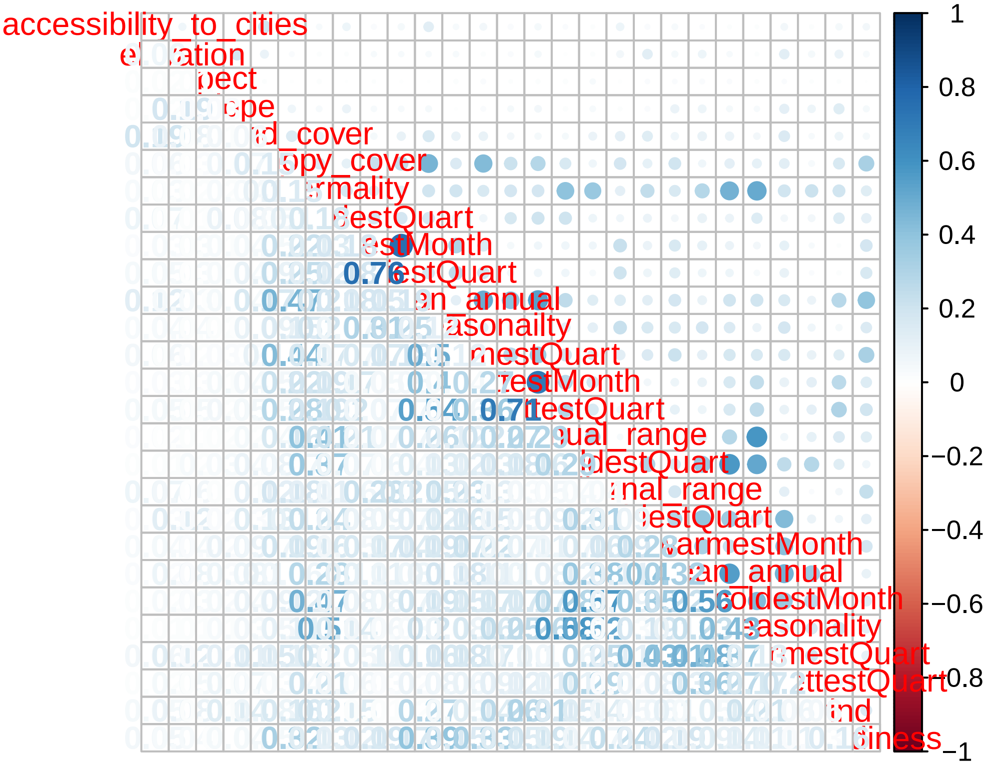
```
cor_spear <- mapply(function(x,y) cor(x, as.integer(y),method = "spearman"), country_env_df_for_pca, cou
#columns = colnames(country_env_df_for_pca)
#cor_df <- data.frame(col = columns, corr = cor_spear)
cor_p <- as.data.frame(cor_spear)

ggplot(data=cor_p, aes(x=rownames(cor_p), y=cor_p$cor_spear)) +
    geom_bar(stat="identity") + theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) +
  xlab("") + ylab("Spearmen Correlation") + ylim(0, 1)
```



7.5 Heatmap

- For binned data

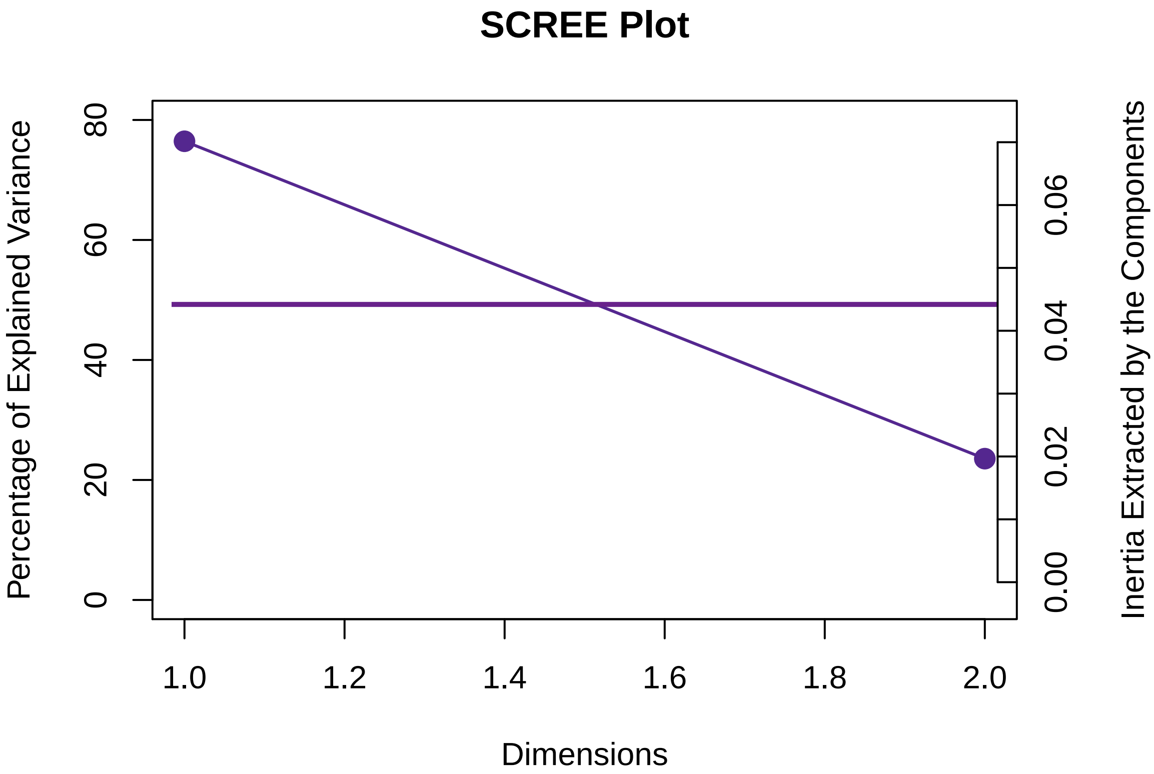
Visually analyze multicollinearity in the system of the original data



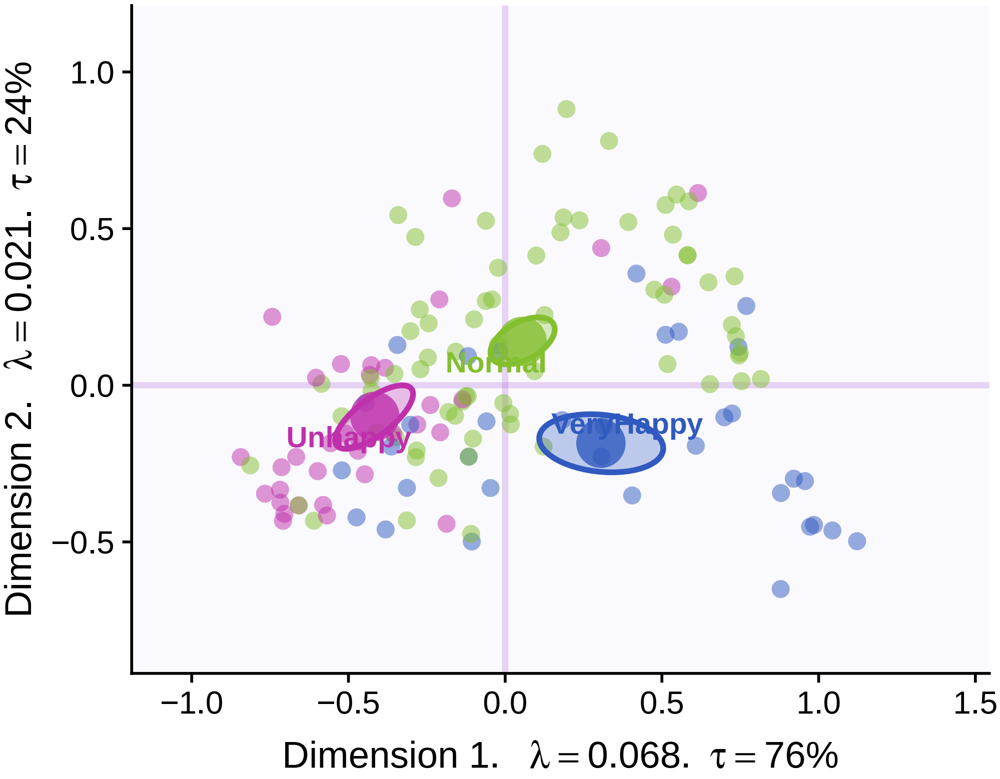
7.6 Scree Plot

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.

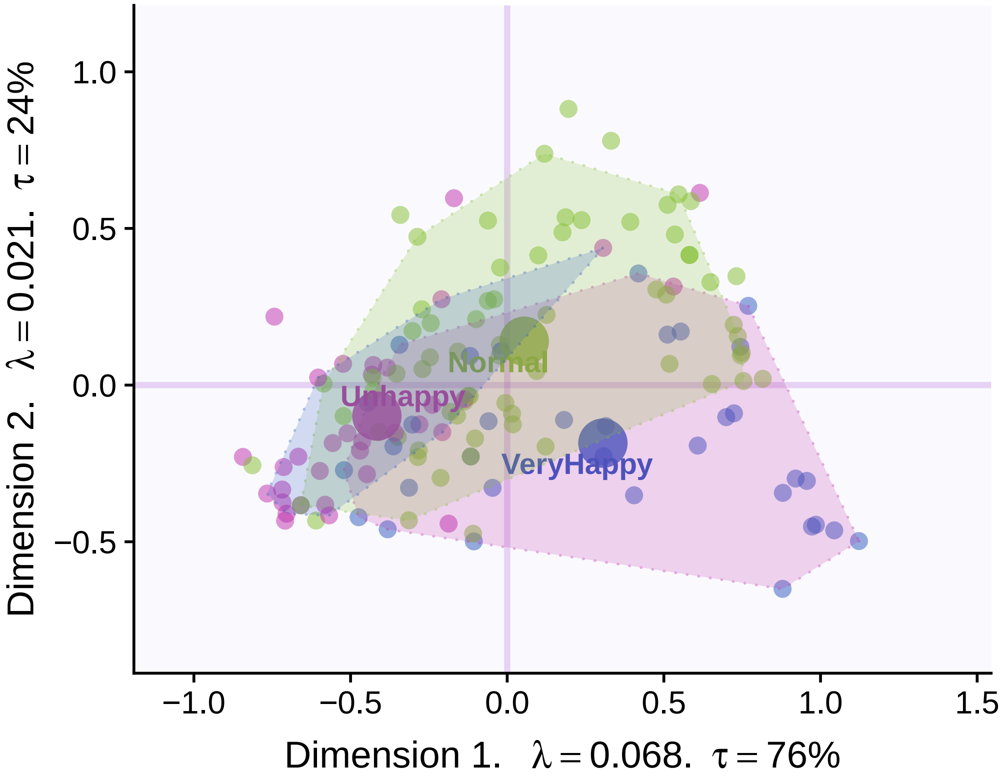
P.S. The most contribution component may not always be most useful for a given research question.



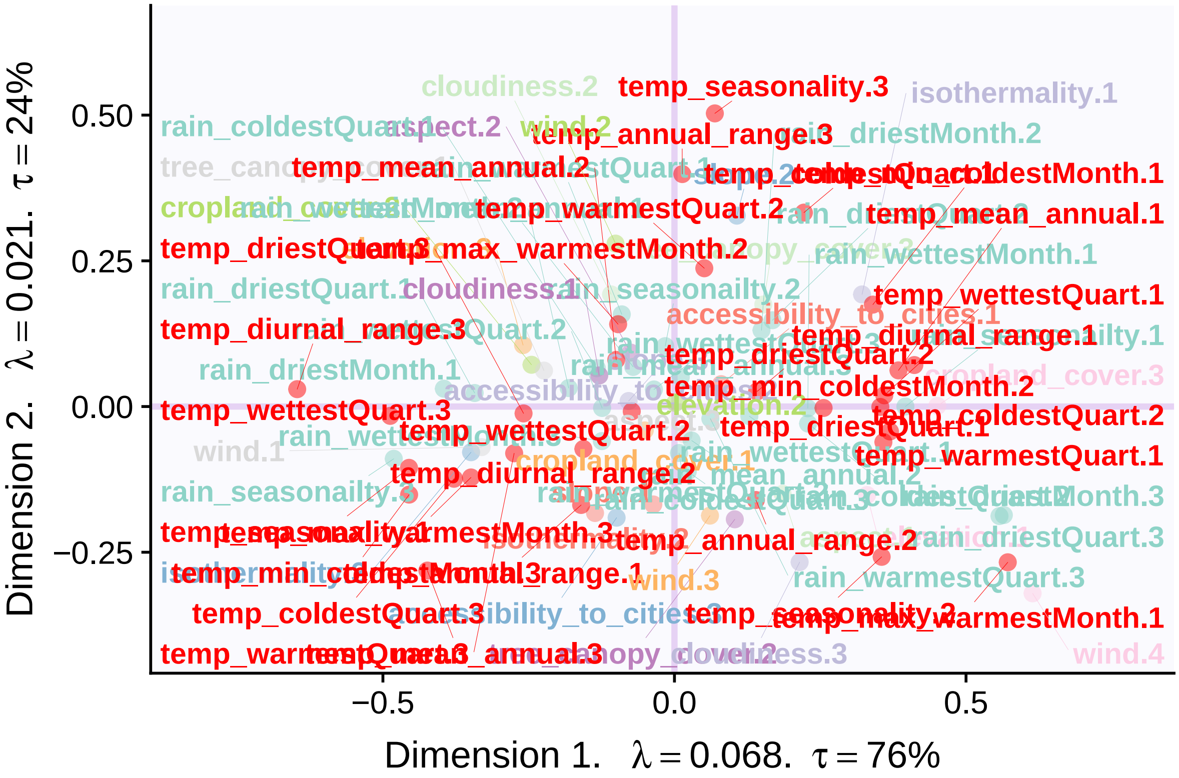
7.7 Factor Scores



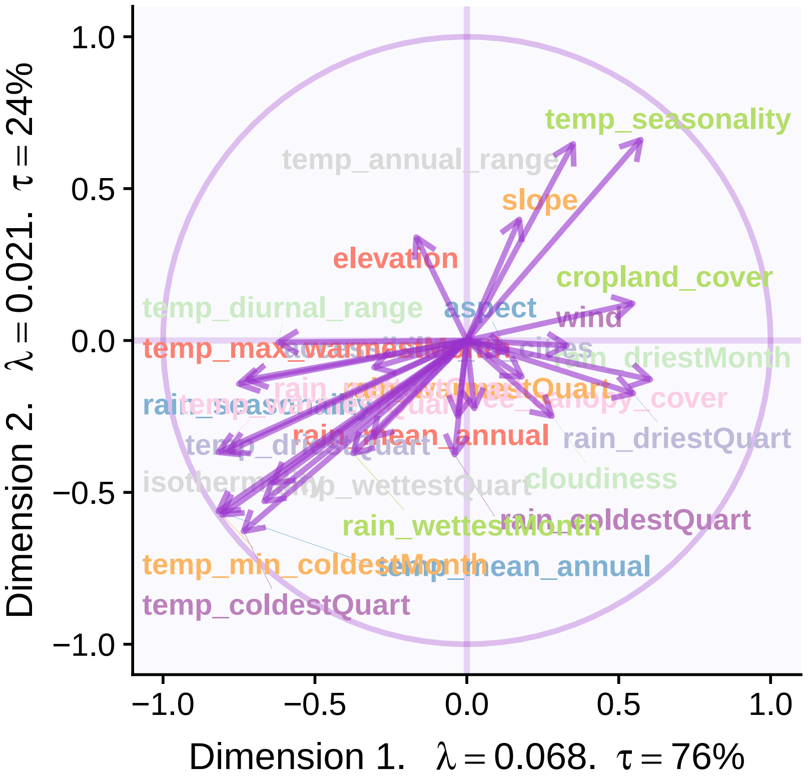
- With Confidence Interval
- With Tolerance Interval


Normal
Unhappy
VeryHappy
1.0
0.5
0.0
−0.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Dimension 2. λ = 0.021. τ = 24%
Dimension 1. λ = 0.068. τ = 76%

7.8 Loadings


cloudiness.2
temp_seasonality.3
isothermality.1
rain_coldestQuart.1
aspect.2
wind.2
temp_annual_range.3
rain_driestMonth.2
tree_canopy_cover.1
temp_mean_annual.2
temp_coldestQuart.1
temp_min_coldestMonth.1
cropland_cover.2
temp_warmestQuart.2
rain_driestQuart.2
temp_mean_annual.1
temp_driestQuart.3
temp_max_warmestMonth.2
rain_wettestMonth.1
rain_driestQuart.1
cloudiness.1
rain_seasonailty.2
temp_wettestQuart.1
accessibility_to_cities.1
temp_diurnal_range.3
temp_diurnal_range.1
rain_wettestQuart.2
temp_driestQuart.2
rain_driestMonth.1
rain_mean_annual.3
cropland_cover.3
temp_min_coldestMonth.2
accessibility_to_cities.2
elevation.2
temp_wettestQuart.3
temp_coldestQuart.2
temp_wettestQuart.2
temp_driestQuart.1
wind.1
rain_wettestMonth.3
temp_warmestQuart.1
temp_diurnal_range.2
cropland_cover.1
rain_mean_annual.2
rain_seasonailty.3
rain_coldestQuart.3
rain_coldestQuart.2
rain_driestMonth.3
temp_seasonality.1
temp_max_warmestMonth.3
isothermality.2
temp_annual_range.2
rain_driestQuart.3
isothermality.3
temp_min_coldestMonth.3
temp_annual_range.1
wind.3
rain_warmestQuart.3
temp_coldestQuart.3
accessibility_to_cities.3
temp_seasonality.2
temp_max_warmestMonth.1
temp_warmestQuart.3
temp_mean_annual.3
tree_canopy_cover.2
cloudiness.3
wind.4
0.50
0.25
0.00
−0.25
−0.5
0.0
0.5
Dimension 2. λ = 0.021. τ = 24%
Dimension 1. λ = 0.068. τ = 76%

7.9 Loadings (correlation plot)

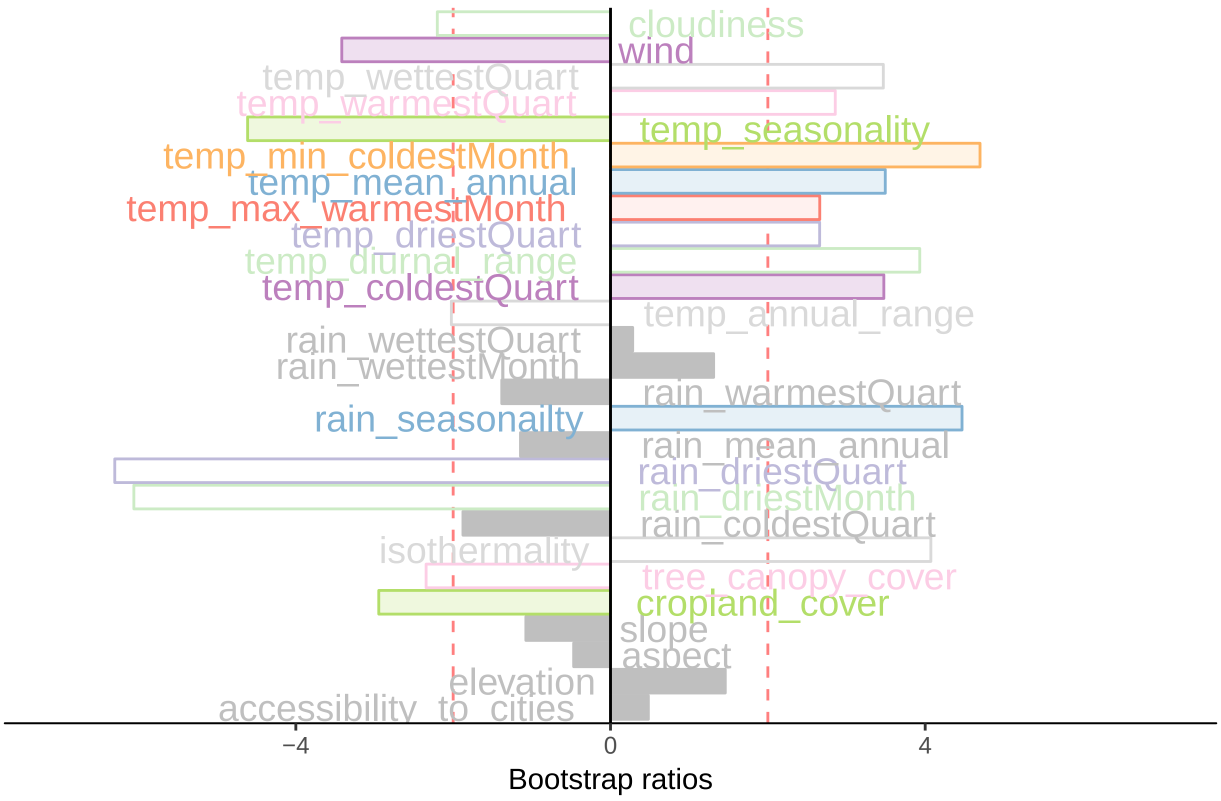


7.10 Most Contributing Variables (Inference)

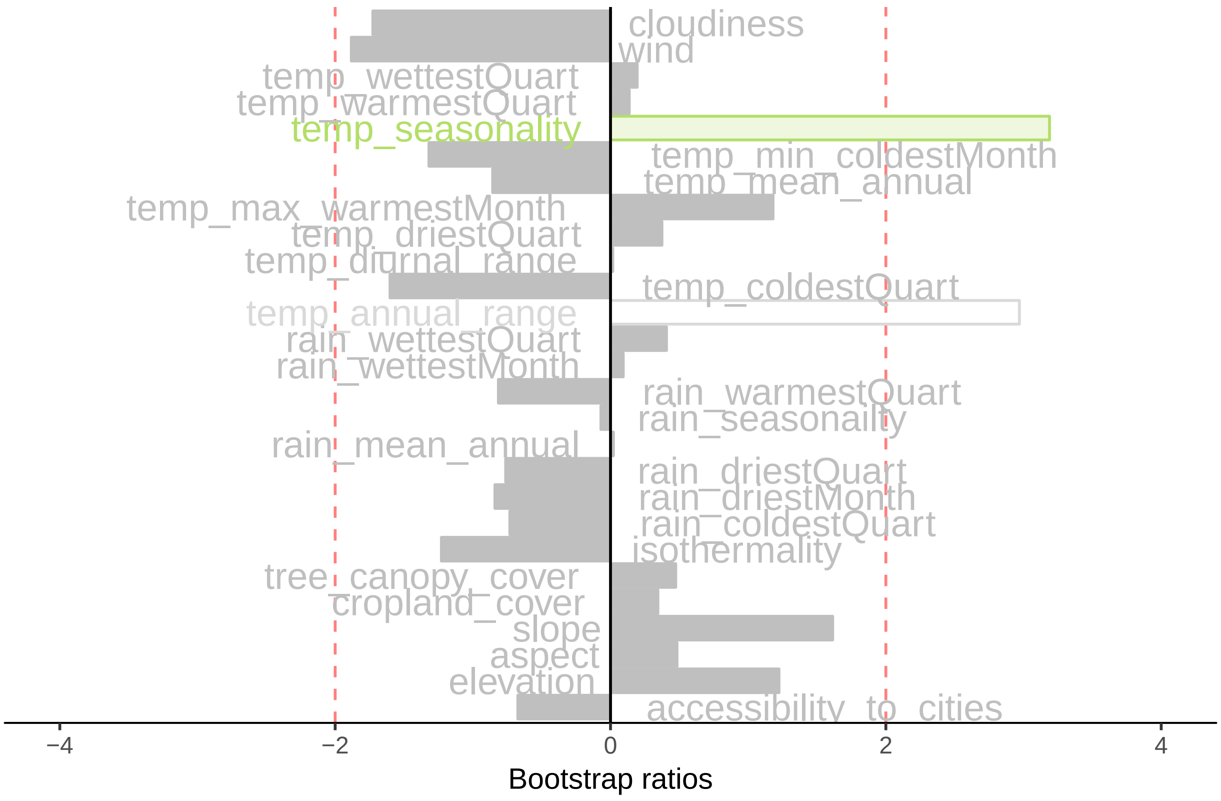
Let's plot variable contributions against each chosen components i.e. 1, 2.

- With Bootstrap Ratio

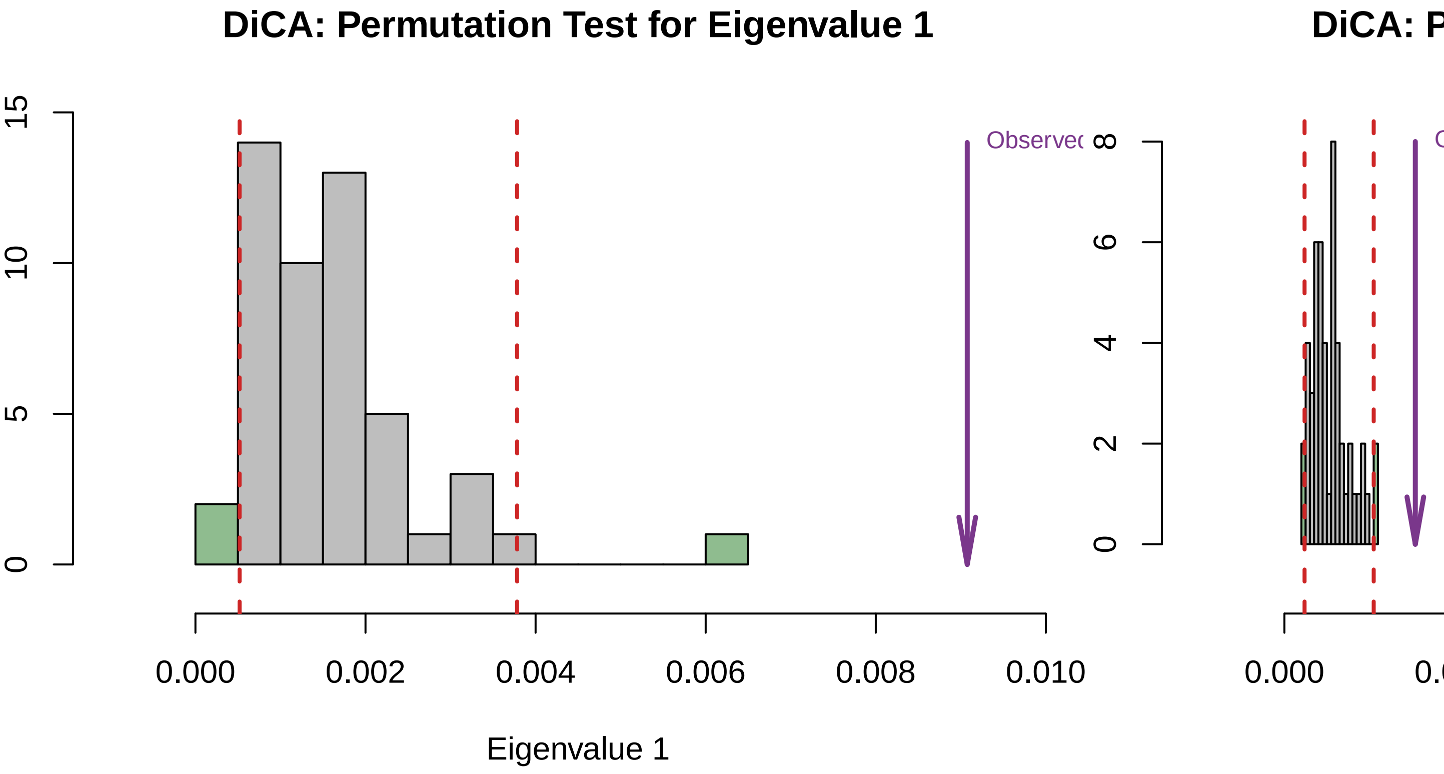
DiCA: Bootstrap ratio 1



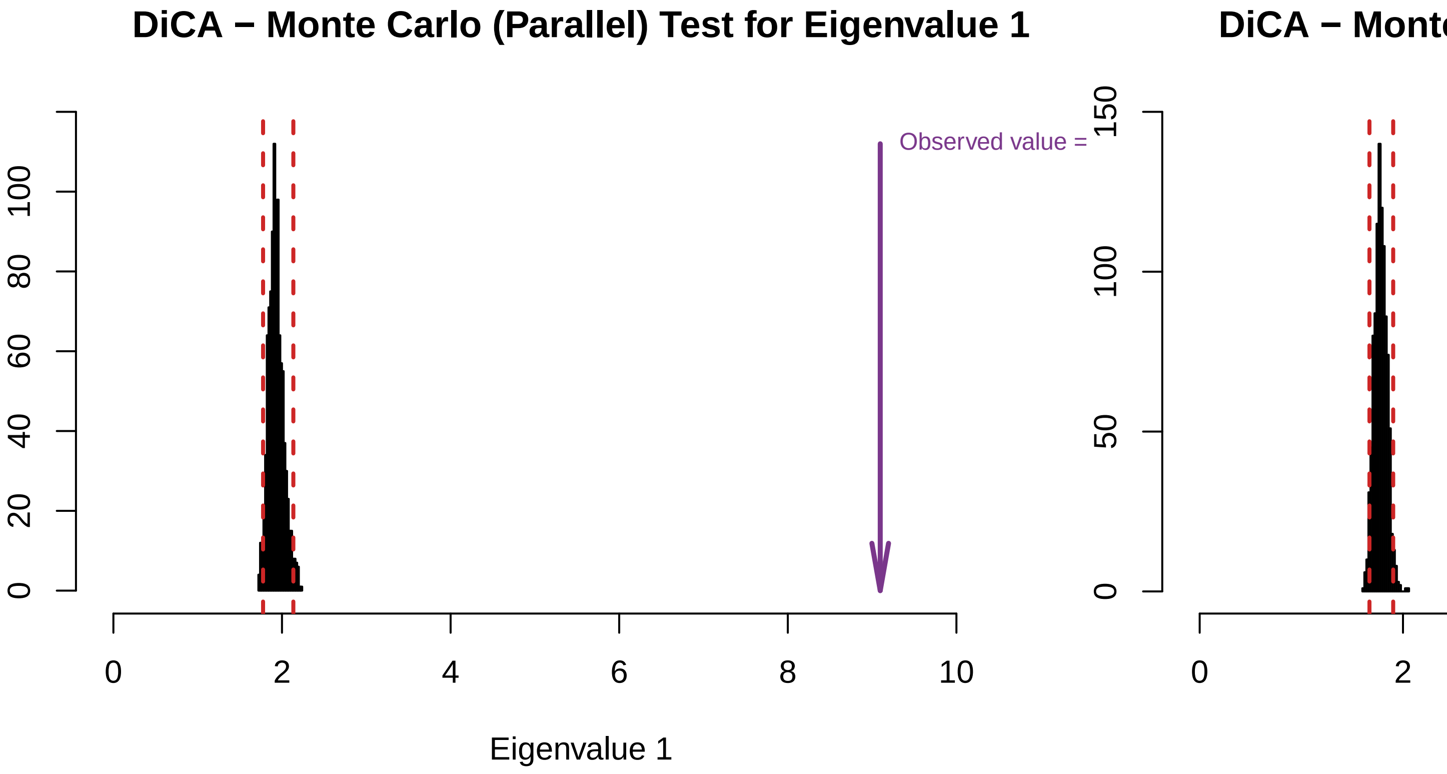
DiCA: Bootstrap ratio 2



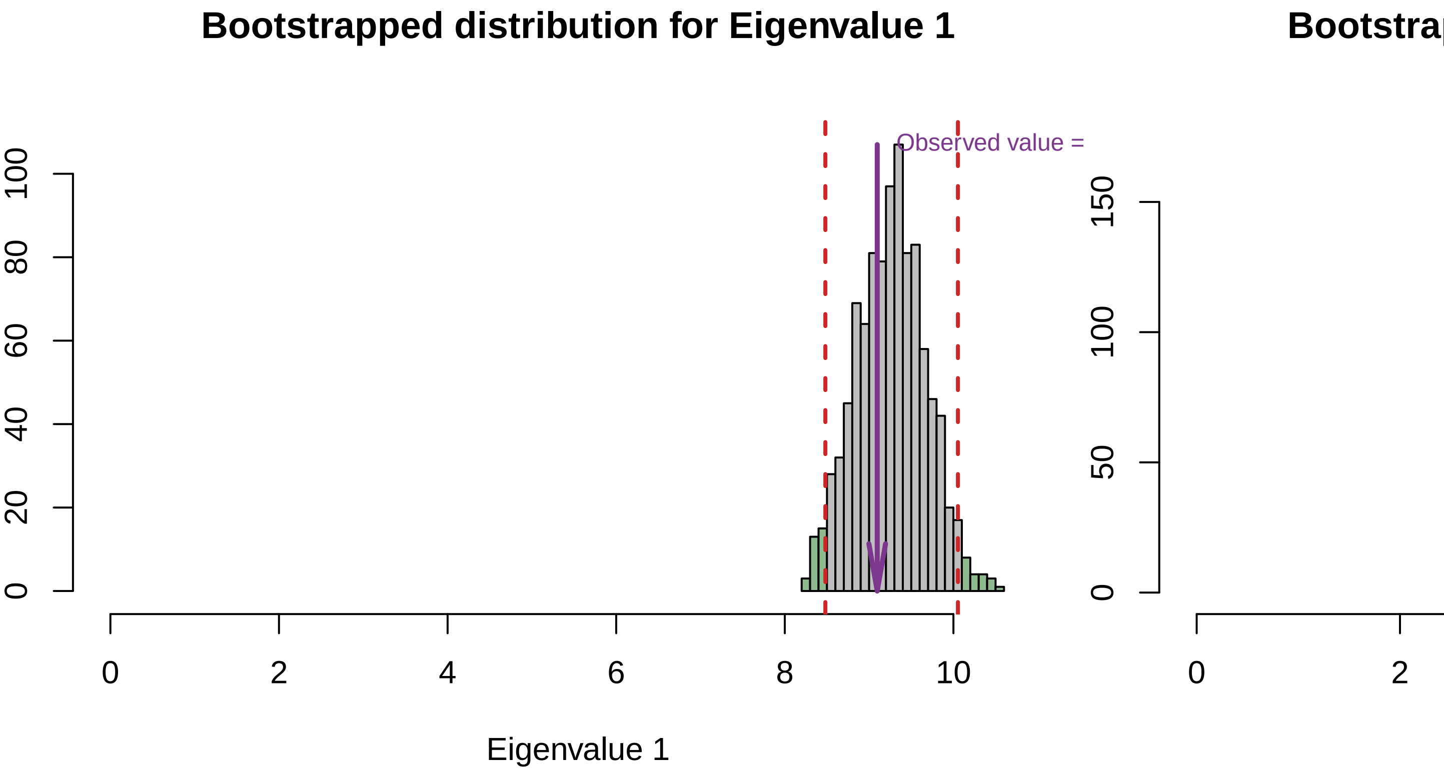
7.11 Permutation Test



7.12 Parallel Test



7.13 Bootstrap Test



7.14 Conclusion

Among PCA, MCA, BADA and DiCA, DiCA is able to best seperate the data based on levels of happiness. It seems increase in temperature is correlated with increase in happiness and increase in rain decreases happiness.

Chapter 8

Multiple Factor Analyis

8.1 Method: MFA

8.2 MFA

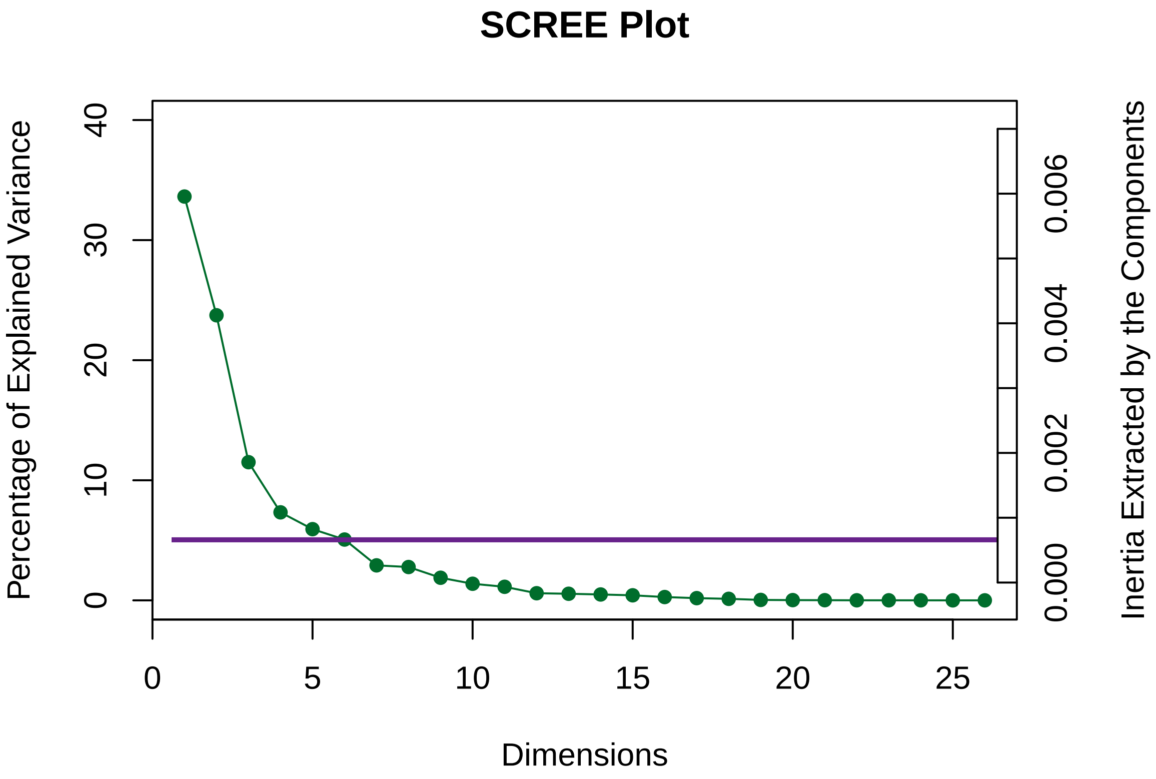
We have divided the data into 3 tables, separate tables for rain and temperature related columns and 3rd table for rest of the columns.

```
## [1] "Preprocessed the Rows of the data matrix using:  None"
## [1] "Preprocessed the Columns of the data matrix using:  Center_1Norm"
## [1] "Preprocessed the Tables of the data matrix using:  MFA_Normalization"
## [1] "Preprocessing Completed"
## [1] "Optimizing using:  None"
## [1] "Processing Complete"
```

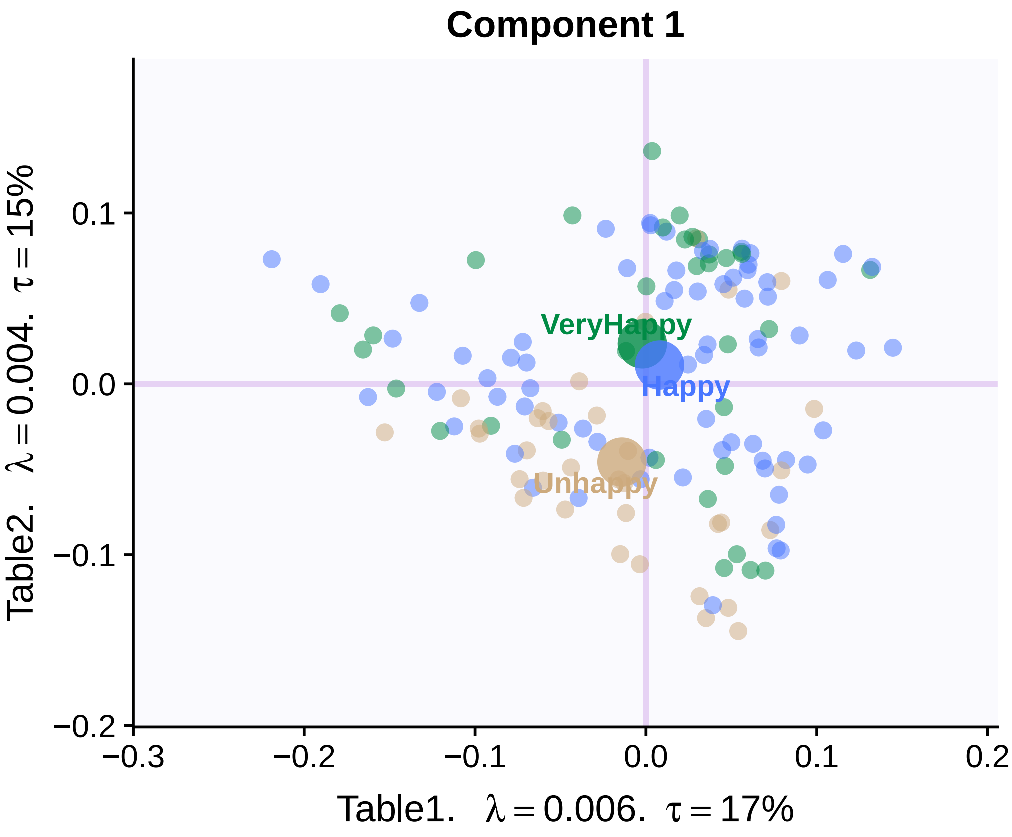
8.3 Scree Plot

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.

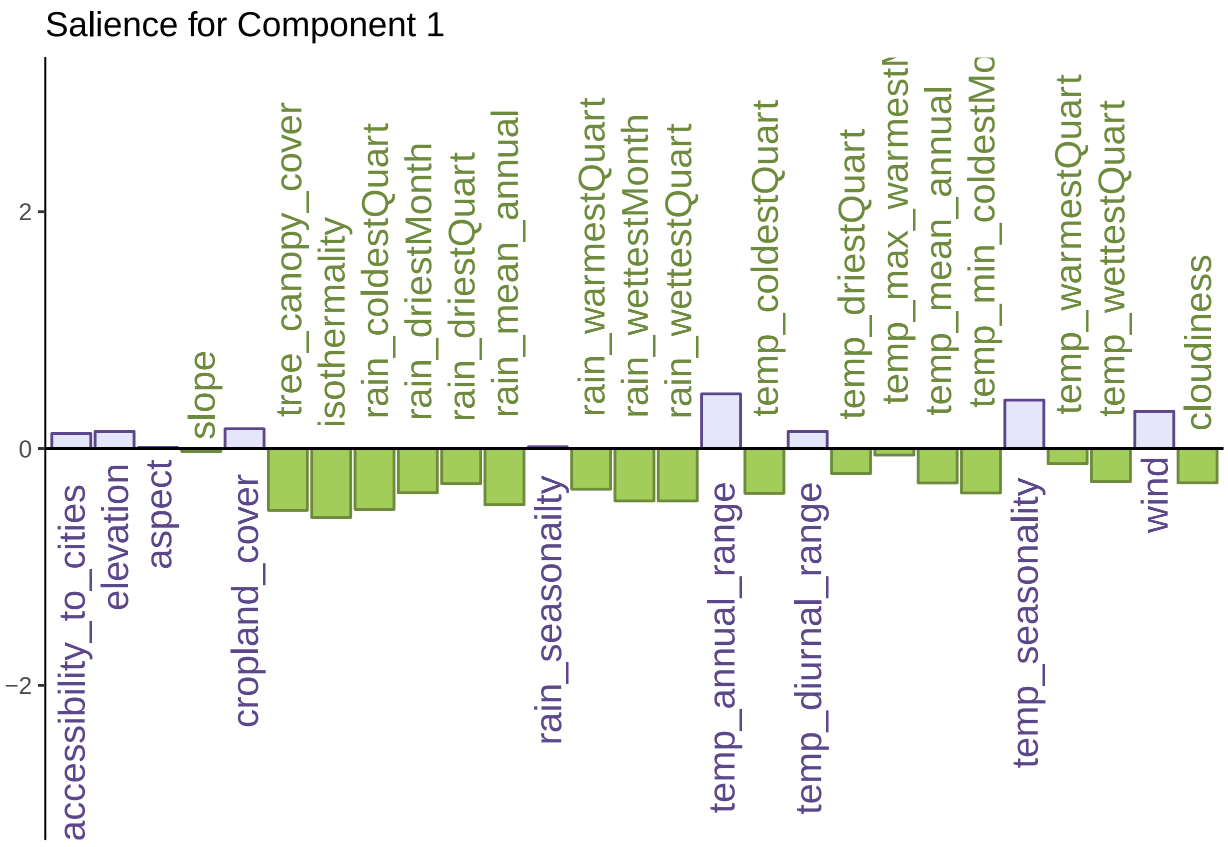
P.S. The most contribution component may not always be most useful for a given research question.

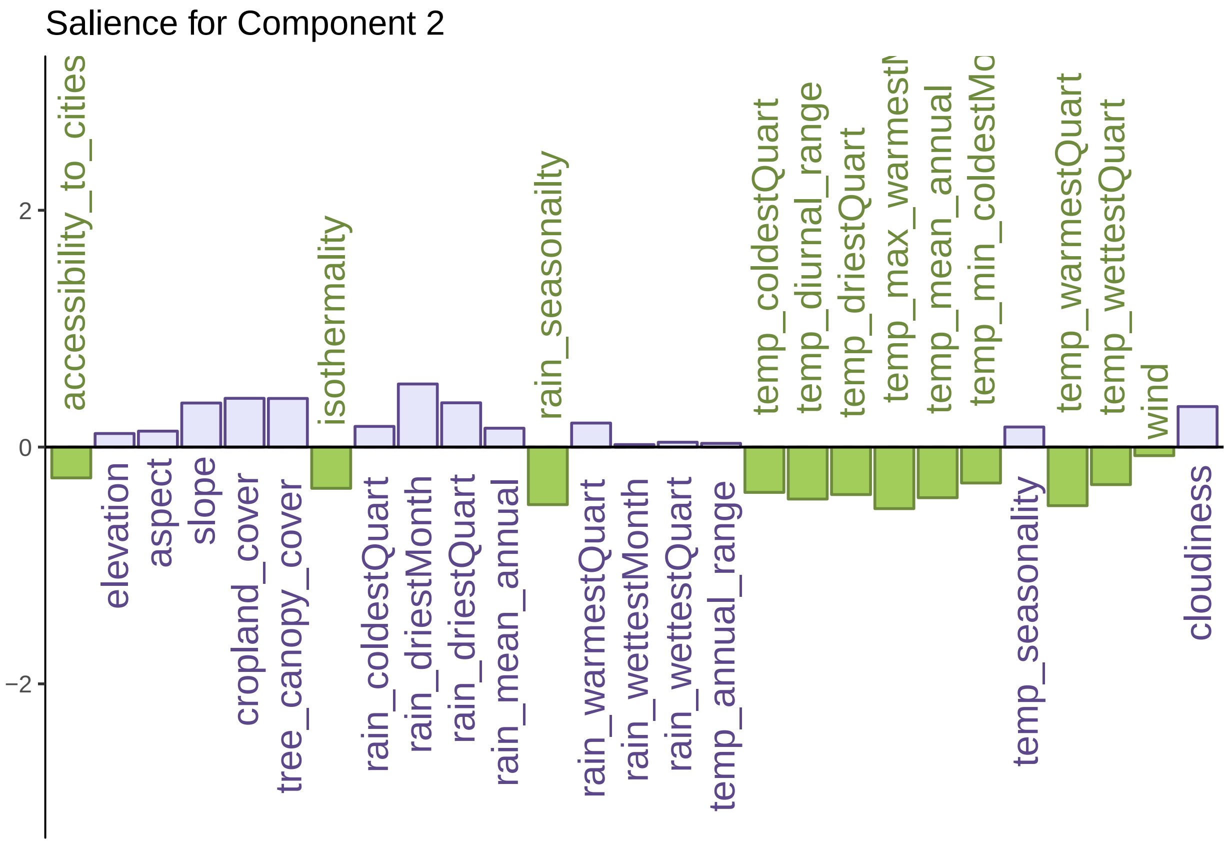


8.4 Factor Scores

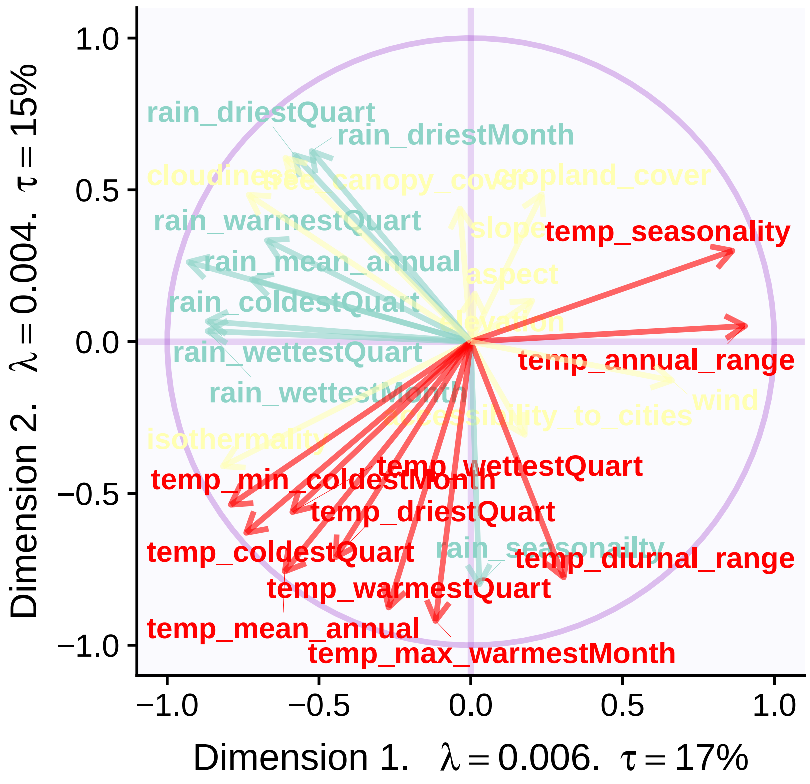


8.5 Loadings


Salience for Component 1
2
0
−2
accessibility_to_cities
elevation
aspect
slope
cropland_cover
tree_canopy_cover
isothermality
rain_coldestQuart
rain_driestMonth
rain_driestQuart
rain_mean_annual
rain_seasonailty
rain_warmestQuart
rain_wettestMonth
rain_wettestQuart
temp_annual_range
temp_coldestQuart
temp_diurnal_range
temp_driestQuart
temp_max_warmestM
temp_mean_annual
temp_min_coldestMo
temp_seasonality
temp_warmestQuart
temp_wettestQuart
wind
cloudiness


Salience for Component 2
2
0
−2
accessibility_to_cities
elevation
aspect
slope
cropland_cover
tree_canopy_cover
isothermality
rain_coldestQuart
rain_driestMonth
rain_driestQuart
rain_mean_annual
rain_seasonailty
rain_warmestQuart
rain_wettestMonth
rain_wettestQuart
temp_annual_range
temp_coldestQuart
temp_diurnal_range
temp_driestQuart
temp_max_warmestM
temp_mean_annual
temp_min_coldestMo
temp_seasonality
temp_warmestQuart
temp_wettestQuart
wind
cloudiness

8.6 Correlation Circle



8.7 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.

Chapter 9

Other Methods

9.1 Method: CA

Correspondence Analysis (CA) is a multivariate graphical technique designed to explore relationships among categorical variables. The outcome from correspondence analysis is a graphical display of the rows and columns of a contingency table that is designed to permit visualization of the salient relationships among the variable responses in a low-dimensional space. Such a representation reveals a more global picture of the relationships among row-column pairs which would otherwise not be detected through a pairwise analysis.

Calculate CA:

- Step 1: Compute row and column averages
- Step 2: Compute the expected values
- Step 3: Compute the residuals
- Step 4: Plotting labels with similar residuals close together
- Step 5: Interpreting the relationship between row and column labels

How to Interpret Correspondence Analysis Plots

Correspondence analysis does not show us which rows have the highest numbers, nor which columns have the highest numbers. It instead shows us the relativities.

- The further things are from the origin, the more discriminating they are.
- Look at the length of the line connecting the row label to the origin. Longer lines indicate that the row label is highly associated with some of the column labels (i.e., it has at least one high residual).
- Look at the length of the label connecting the column label to the origin. Longer lines again indicate a high association between the column label and one or more row labels.
- Look at the angle formed between these two lines. Really small angles indicate association. 90 degree angles indicate no relationship. Angles near 180 degrees indicate negative associations.

9.1.1 Dataset

- Data: Measurements of Weekly Earnings per Race
- Rows: There are 6 observations representing Asian/White/Black, Men/Woman.
- Columns: Total 6 variables grouping people based on Decile and Quartile ranges of their weekly income.

```
##                            White.men White.women Black.men Black.Women
## 1st decile                       412         374       361         331
## 1st quartile                     594         506       483         423
## 2nd quartile                     920         743       680         615
```

Table 9.1: 4 Pianist for each of 3 Composers

	1stQ	2ndQ	3rdQ
White.men	594	326	547
White.women	506	237	397
Black.men	483	197	366
Black.Women	423	192	320
Asian.Men	648	481	731
Asian.Women	551	326	534
Hispanic.Men	451	180	348
Hispanic.Women	404	162	264

```
## 3rd quartile                     1467        1140      1046         935
## 9th decile                       2278        1726      1551        1453
## Total people (in thousands)     48746       36698      6445        7142
##                             Asian.Men Asian.Women Hispanic.Men
## 1st decile                        420         385          358
## 1st quartile                      648         551          451
## 2nd quartile                     1129         877          631
## 3rd quartile                     1860        1411          979
## 9th decile                       2699        2024         1498
## Total people (in thousands)      3684        2954        11142
##                             Hispanic.Women
## 1st decile                             320
## 1st quartile                           404
## 2nd quartile                           566
## 3rd quartile                           830
## 9th decile                            1266
## Total people (in thousands)           7168
```

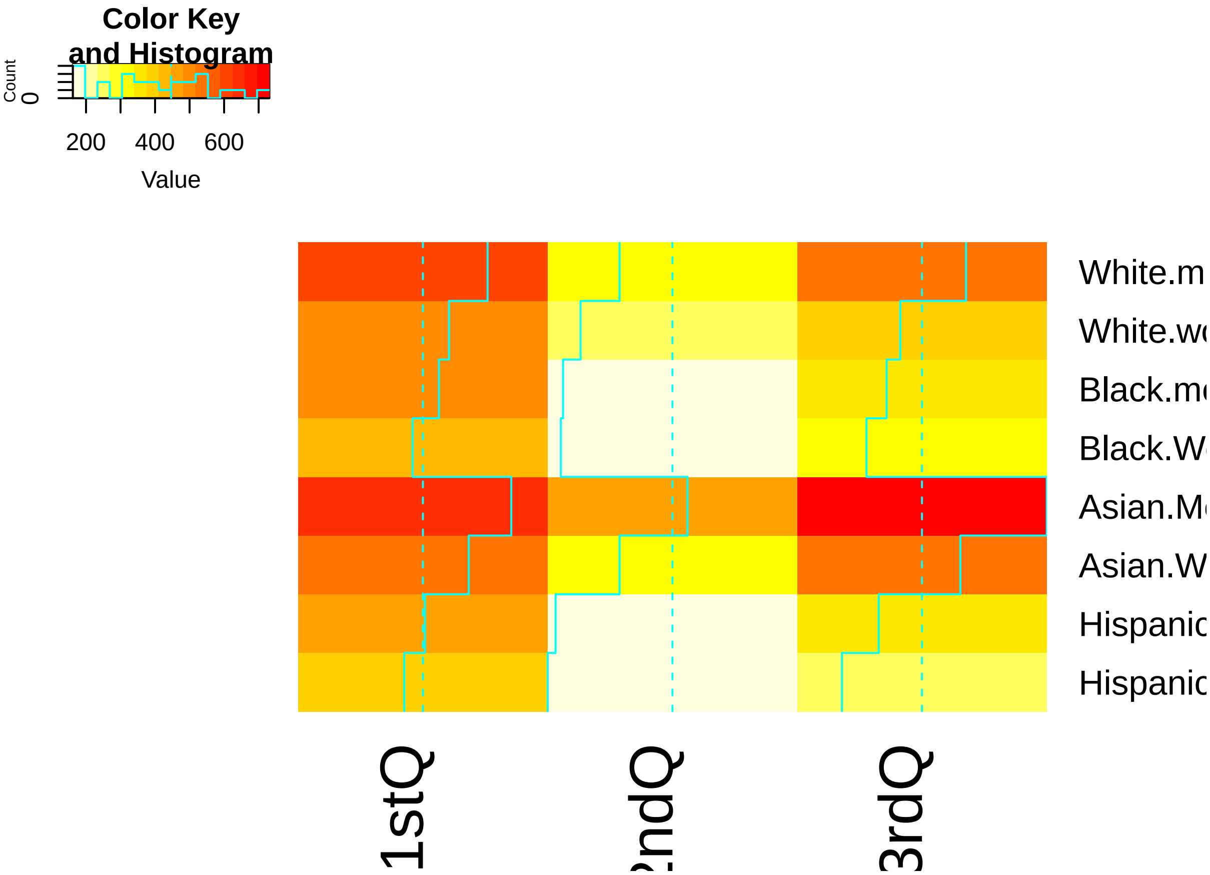
However, here we can see that it may not be advisable to include Quartile and Decile intervals in the same analysis. Hence, we go ahead with Quartile Ranges only.

- Research Question
  - Does total earning of different races differ.
  - Which race get less than median salary (2nd Quartile)

9.1.2 Heatmap

```
## Warning in heatmap.2(WE_data, Colv = FALSE, Rowv = FALSE, col =
## rev(heat.colors(16))): Discrepancy: Rowv is FALSE, while dendrogram is
## `both'. Omitting row dendogram.
```

```
## Warning in heatmap.2(WE_data, Colv = FALSE, Rowv = FALSE, col =
## rev(heat.colors(16))): Discrepancy: Colv is FALSE, while dendrogram is
## `column'. Omitting column dendogram.
```

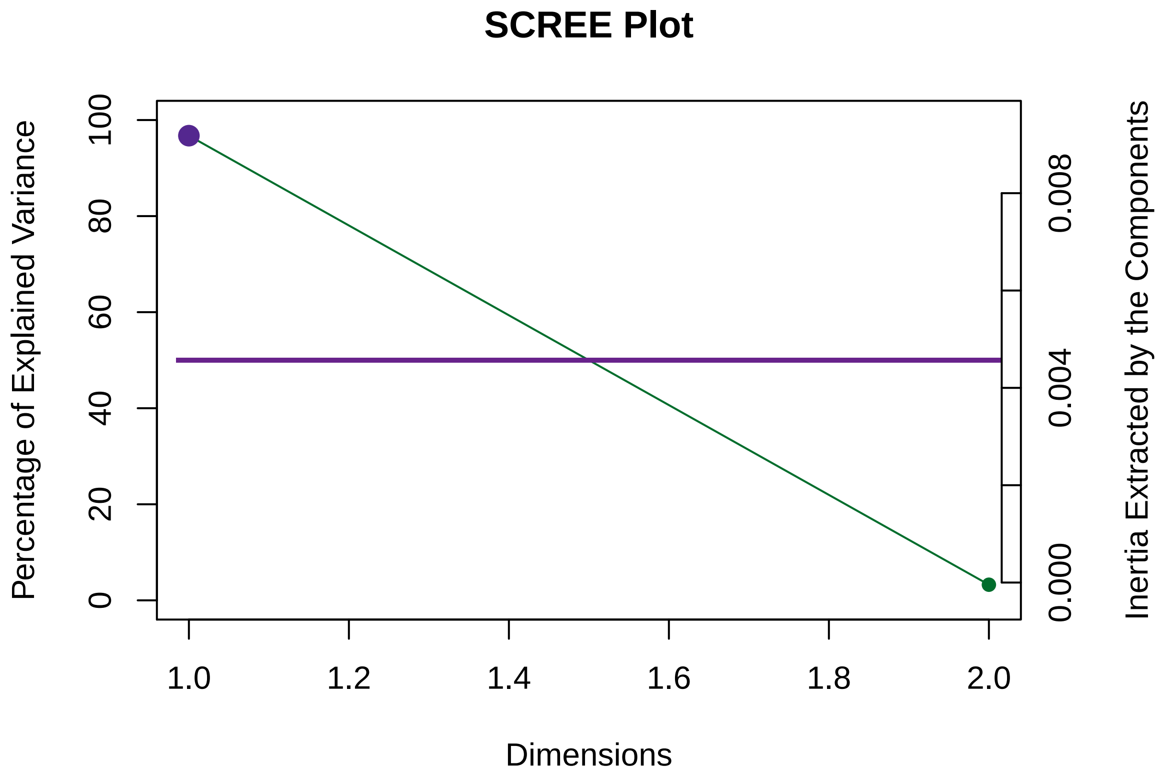


9.1.3 Scree Plot

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.

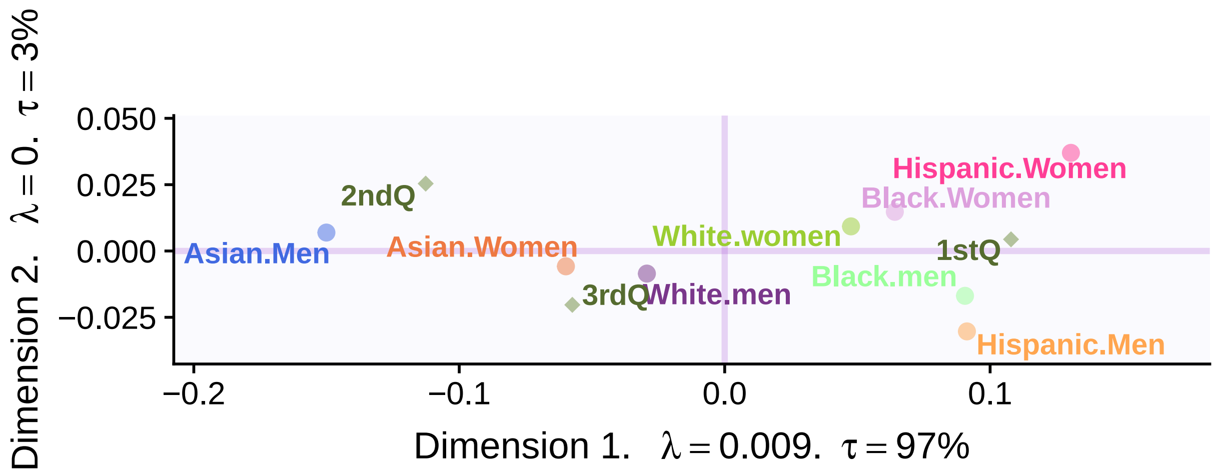
```
PTCA4CATA::PlotScree(ev = resCA.sym$ExPosition.Data$eigs,
                     p.ev =  we_data_inf$Inference.Data$components$p.vals,
                     title = 'SCREE Plot',
                     plotKaiser = TRUE
)
```



9.1.4 Factor Scores

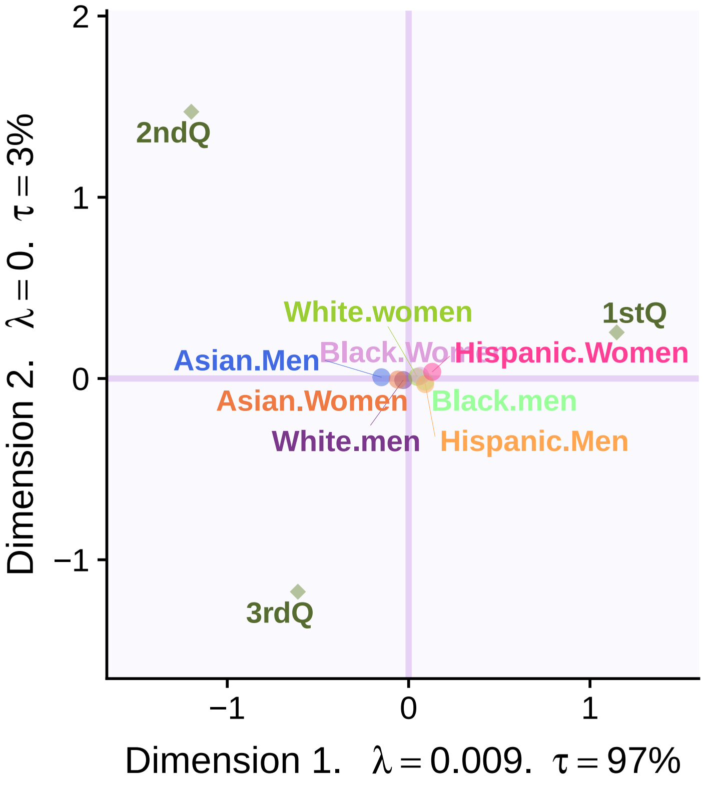
9.1.4.1 Symmetric Plot

```
map.IJ.sym <- symMap$baseMap + symMap$I_labels + symMap$I_points +
  symMap$J_labels + symMap$J_points + labels4CA
print(map.IJ.sym)
```



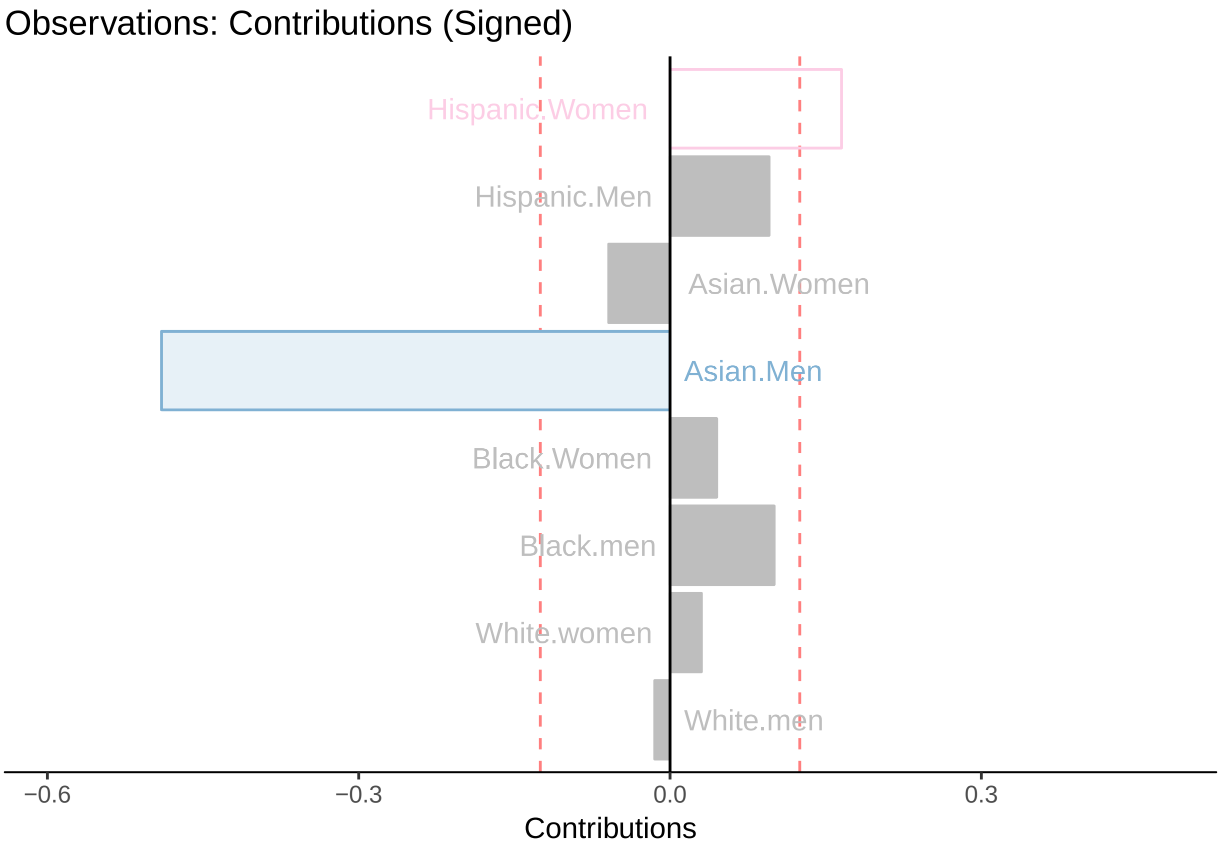
9.1.4.2 Asymmetric Plot

```
map.IJ.asym <- asymMap$baseMap + asymMap$I_labels +
  asymMap$I_points + asymMap$J_labels +
  asymMap$J_points + labels4CA
print(map.IJ.asym)
```

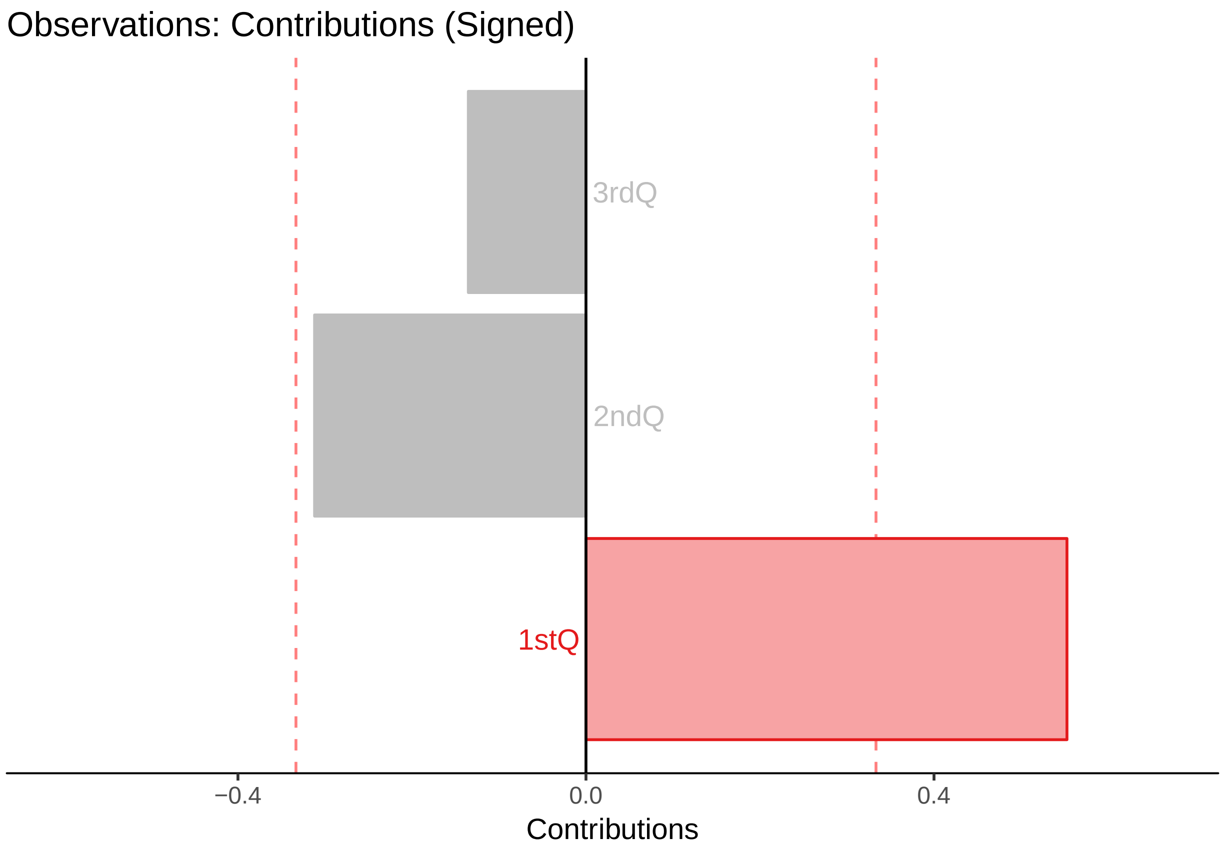


9.1.5 Most Contributing Variables

```
PTCA4CATA::PrettyBarPlot2(ctr.I[,1],
                          threshold = 1 / NROW(ctr.I),
                          font.size = 4,
                          color4bar = gplots::col2hex(color4I),
                          color4ns = 'grey',
                          main = 'Observations: Contributions (Signed)',
                          ylab = 'Contributions', ylim = c(1.2*min(ctr.I),
                           1.2*max(ctr.I) ),
                          horizontal = FALSE )
```

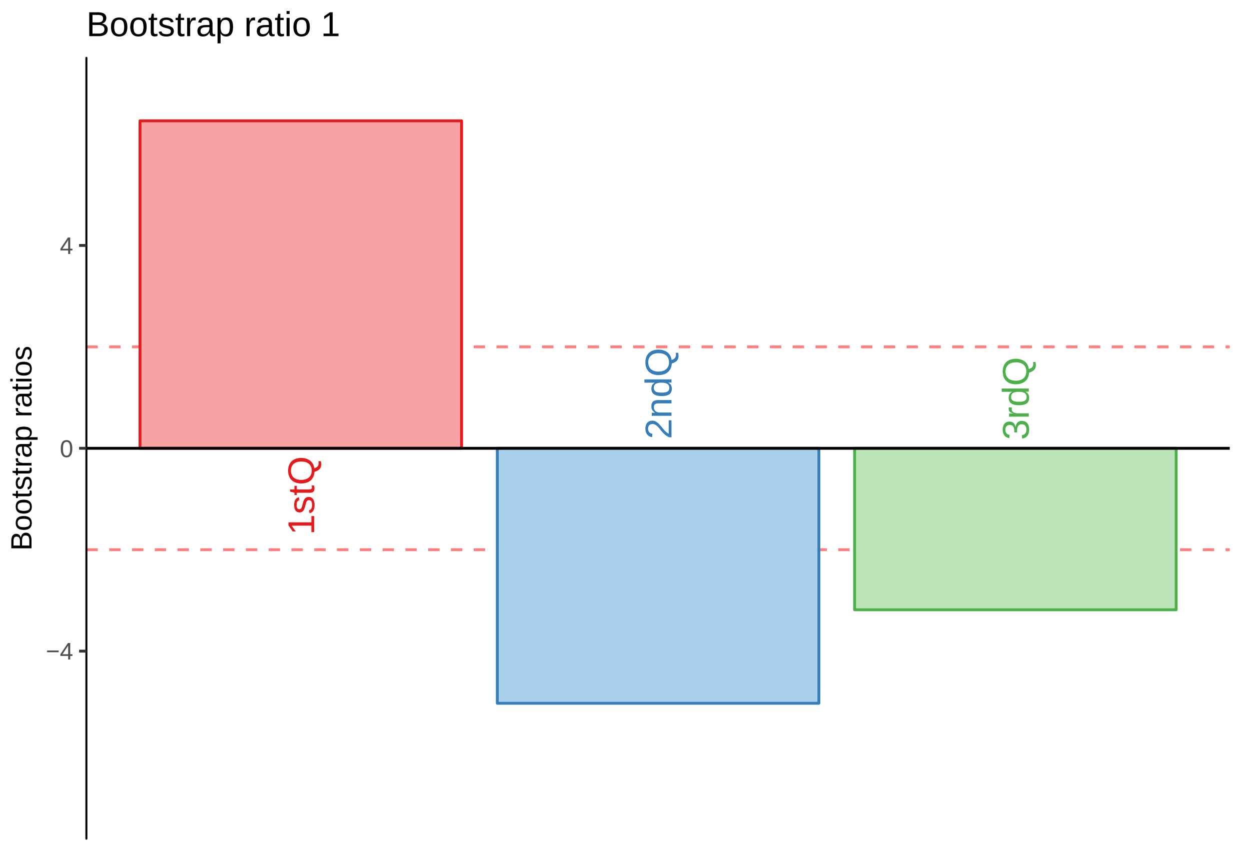


```
PTCA4CATA::PrettyBarPlot2(ctr.J[,1],
                        threshold = 1 / NROW(ctr.J),
                        font.size = 4,
                        color4bar = color4J,
                        color4ns = 'grey',
                        main = 'Observations: Contributions (Signed)',
                        ylab = 'Contributions', ylim = c(1.2*min(ctr.J),
                         1.2*max(ctr.J) ),
                        horizontal = FALSE )
```

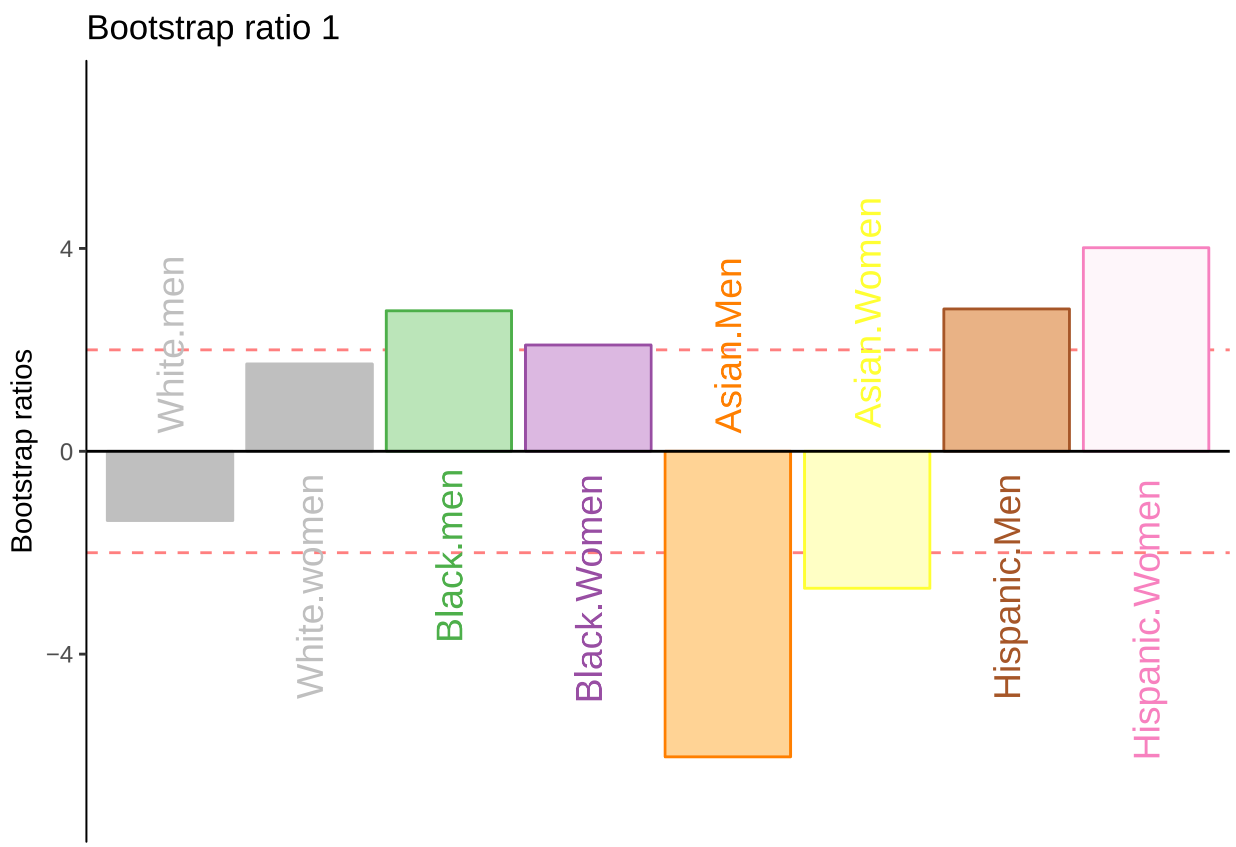


9.1.6 Inference CA

```
ba001.BR1 <- PrettyBarPlot2(BR[,laDim],
                          threshold = 2,
                          font.size = 5,
                          color4bar = gplots::col2hex(col4J), # we need hex code
                          main = paste0('Bootstrap ratio ',laDim),
                          ylab = 'Bootstrap ratios'
                          #ylim = c(1.2*min(BR[,laDim]), 1.2*max(BR[,laDim]))
)
print(ba001.BR1)
```



```r
wedata.BR1 <- PrettyBarPlot2(BR[,laDim],
                             threshold = 2,
                             font.size = 5,
                             color4bar = gplots::col2hex(col4J), # we need hex code
                             main = paste0('Bootstrap ratio ',laDim),
                             ylab = 'Bootstrap ratios'
                             #ylim = c(1.2*min(BR[,laDim]), 1.2*max(BR[,laDim]))
)
print(wedata.BR1)
```



```r
rm(list = ls())

devtools::install_github('HerveAbdi/PTCA4CATA')
```

```
## Skipping install of 'PTCA4CATA' from a github remote, the SHA1 (0a982b85) has not changed since last
##   Use `force = TRUE` to force installation
```

```r
suppressMessages(library(PTCA4CATA))
# PTCA4CATA should first to avoid conflict with TInPosition
suppressMessages(library(ExPosition))
#suppressMessages(library(InPosition))
#suppressMessages(library(TInPosition))
suppressMessages(library(ggplot2))
suppressMessages(library(dplyr))
suppressMessages(library(officer))
suppressMessages(library(flextable))
suppressMessages(library(rvg))
suppressMessages(library(useful))
suppressMessages(library(DistatisR))
library(RColorBrewer)


# Source the function file:
#
# install.packages('TExPosition')
# install.packages('MExPosition')
library(MExPosition)
rm(list = ls())
```

Table 9.2: 4 Pianist for each of 3 Composers

	bc001	bc002	bc003	bc004	bc005	bc006	bc007	bc008	bc009	bc010	bc011	bc012	bc0
Bach.Arrau.1	1	2	1	2	3	1	1	3	1	1	2	2	
Bach.Arrau.2	2	1	1	2	3	1	1	2	1	1	2	3	
Bach.Arrau.3	1	3	2	1	1	1	2	2	3	3	3	2	
Bach.Baren.4	1	3	2	2	2	1	2	3	1	1	3	1	
Bach.Baren.5	1	1	2	3	3	3	3	3	1	2	2	2	
Bach.Baren.6	1	3	3	2	3	1	2	1	3	1	3	2	
Bach.Pires.7	1	2	3	3	2	3	1	3	3	1	2	1	
Bach.Pires.8	2	3	1	2	3	1	2	2	1	3	1	3	
Bach.Pires.9	1	3	2	1	3	3	1	2	2	2	1	2	
Bach.Richt.10	2	1	1	3	2	2	2	2	2	3	1	3	
Bach.Richt.11	3	3	3	2	1	3	1	3	3	1	3	3	
Bach.Richt.12	2	1	1	1	1	2	3	3	3	3	1	3	
Beet.Arrau.13	1	2	2	2	3	1	2	1	1	2	2	3	
Beet.Arrau.14	3	2	3	3	1	3	1	1	2	1	3	1	
Beet.Arrau.15	3	2	3	1	2	2	1	1	2	3	3	2	
Beet.Baren.16	1	3	3	3	2	3	2	1	2	1	3	1	
Beet.Baren.17	1	3	2	2	2	1	1	1	1	1	2	1	
Beet.Baren.18	3	1	3	2	3	1	1	3	2	1	3	2	
Beet.Pires.19	1	3	3	3	2	2	3	1	2	1	3	3	
Beet.Pires.20	3	3	3	2	1	1	2	1	2	2	2	3	

9.2 DiSTATIS

To be filled

9.2.1 Dataset

9.3 [1] Bootstrap On Factor Scores. Iterations #:

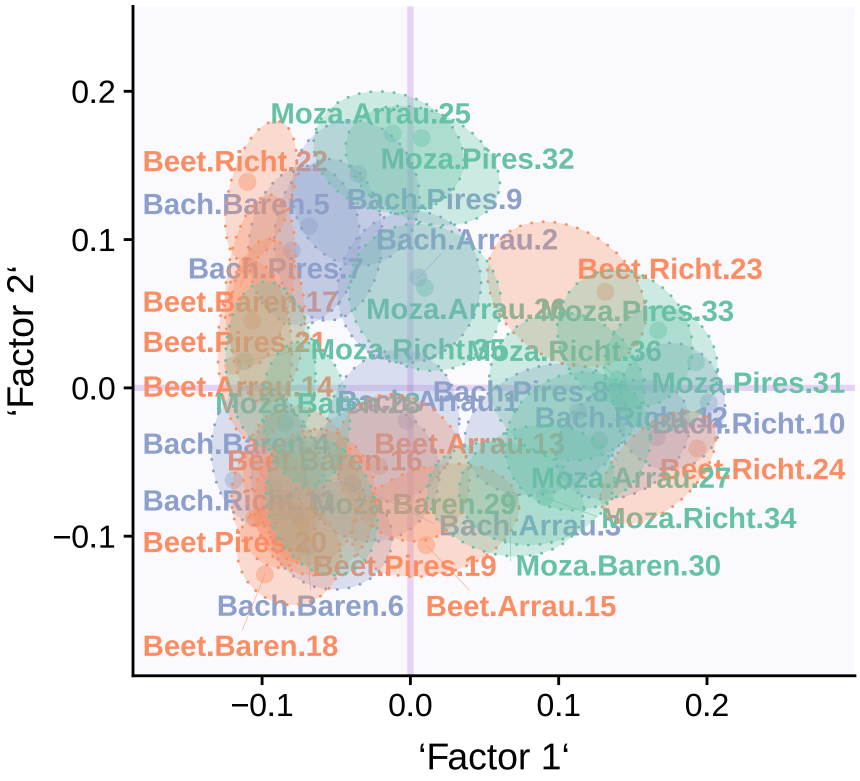
9.4 [2] 1000

```
### SCREE Plot

![](Country_Environment_conditions_corelated_with_Happiness_files/figure-latex/unnamed-chunk-115-1.pdf)

### Plotting Assessor Matrix

![](Country_Environment_conditions_corelated_with_Happiness_files/figure-latex/unnamed-chunk-116-1.pdf)
```

```
### ConvexHull

![](Country_Environment_conditions_corelated_with_Happiness_files/figure-latex/unnamed-chunk-117-1.pdf)


![](Country_Environment_conditions_corelated_with_Happiness_files/figure-latex/unnamed-chunk-118-1.pdf)

### I Set

![](Country_Environment_conditions_corelated_with_Happiness_files/figure-latex/unnamed-chunk-119-1.pdf)
```

9.5 Warning: Removed 3 rows containing missing values (geom_text_repel).

‘‘


Moza.Arrau.25
Beet.Richt.22
Moza.Pires.32
Bach.Baren.5
Bach.Pires.9
Bach.Arrau.2
Bach.Pires.7
Beet.Richt.23
Beet.Baren.17
Moza.Arrau.26
Moza.Pires.33
Beet.Pires.21
Moza.Richt.35
Moza.Richt.36
Beet.Arrau.14
Bach.Pires.8
Moza.Pires.31
Moza.Baren.28
Bach.Arrau.1
Bach.Richt.12
Bach.Richt.10
Bach.Baren.4
Beet.Arrau.13
Beet.Baren.16
Beet.Richt.24
Moza.Arrau.27
Bach.Richt.11
Moza.Baren.29
Bach.Arrau.3
Moza.Richt.34
Beet.Pires.20
Beet.Pires.19
Moza.Baren.30
Bach.Baren.6
Beet.Arrau.15
Beet.Baren.18
‘Factor 2‘
‘Factor 1‘