

## One Stop Center for principal Component Analysis (PCA)

This shiny application takes a CSV file of clean data, allows you to inspect the data and compute a Principal Components Analysis, and will return several diagnostic plots and tables. The plots include a correlation matrix, a scree plot, and a biplot of Principal Components.

This PCA dashboard has Six (6) tabs namely, Introduction, Data input, Corelation plots, Compute PCA, PC plots, PCA output.

The introduction tab of this shiny application gives an introduction to what Principal Component Analysis is and what is the main idea behind PCA.

Here is a screenshot of the introduction tab of the shiny app One Stop Center for principal Component Analysis (PCA).

One Stop Center for Principal Component Analysis(PCA)

Introduction Data input Correlation Plots Compute PCA PC Plots PCA output

## WHAT IS PRINCIPAL COMPONENT ANALYSIS?

Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. Reducing the number of variables of a data set naturally comes at the expense of accuracy, but the trick in dimensionality reduction is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process. So to sum up, the idea of PCA is simple – reduce the number of variables of a data set, while preserving as much information as possible.

### Main Idea behind PCA:

Variance synonymous with amount of information, more variance ==> more info. PCA determines principal components (or high variance components) and projects data to a new subspace which is in the direction of these components.

Eigen vectors of the covariance matrix represent the directions in which data varies the most. Eigen values tell us the amount of variance in each of these directions. A d-dimensional dataset corresponds to at most d eigen values/vectors (some may repeat).

The second tab of the shiny application is the Data input tab, which is the first step in computing PCA for the given dataset. We can upload the dataset into the application using this tab.

Here is a screenshot of the Data input tab of the shiny app One Stop Center for principal Component Analysis (PCA).

# One Stop Center for Principal Component Analysis(PCA)

[Introduction](#) [Data input](#) [Correlation Plots](#) [Compute PCA](#) [PC Plots](#) [PCA output](#)

Before uploading your data, check that it is clean, especially ensure that the the numeric variables contain only the digits 0-9 or NA (to indicate missing data).

Rows that contain one or more NAs will be excluded from the PCA.

Columns that contain a mixture of numbers and text will not be included in the computation of the PCA results.

Have a look at the [iris.csv](#) file included with this app to see what a clean CSV file looks like.

Select the options that match your CSV file, then upload your file:

Header

- ☐ Columns have headers  
☒ Columns do not have headers

Separator

- ☐ Comma  
☒ Semicolon  
☐ Tab

Quote

- ☒ None  
☐ Double Quote  
☐ Single Quote

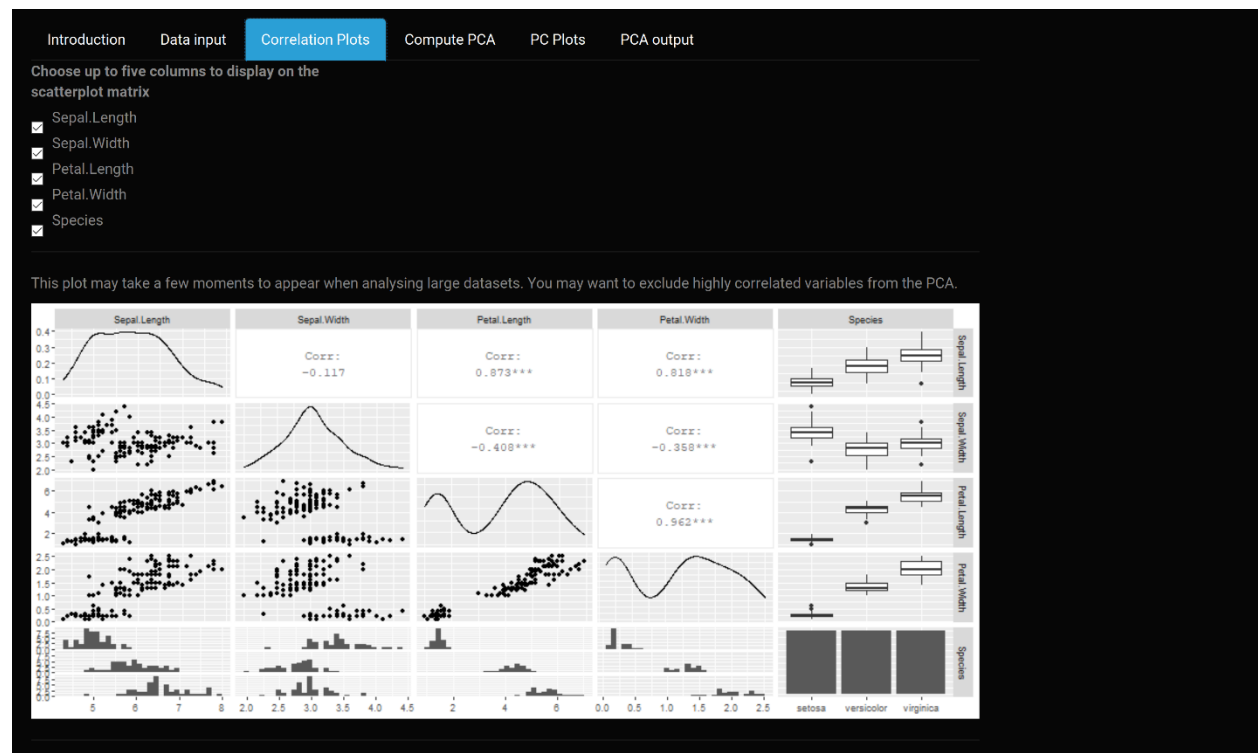
Choose a CSV file to upload:

Browse...

No file selected

The third tab of the shiny application is the Correlation plot tab, Here in this tab we can select the features of the dataset that we want to have the plot for by ticking or selecting the boxes corresponding to the features.

Here is a screenshot of the Correlations plot tab of the shiny app One Stop Center for principal Component Analysis (PCA).

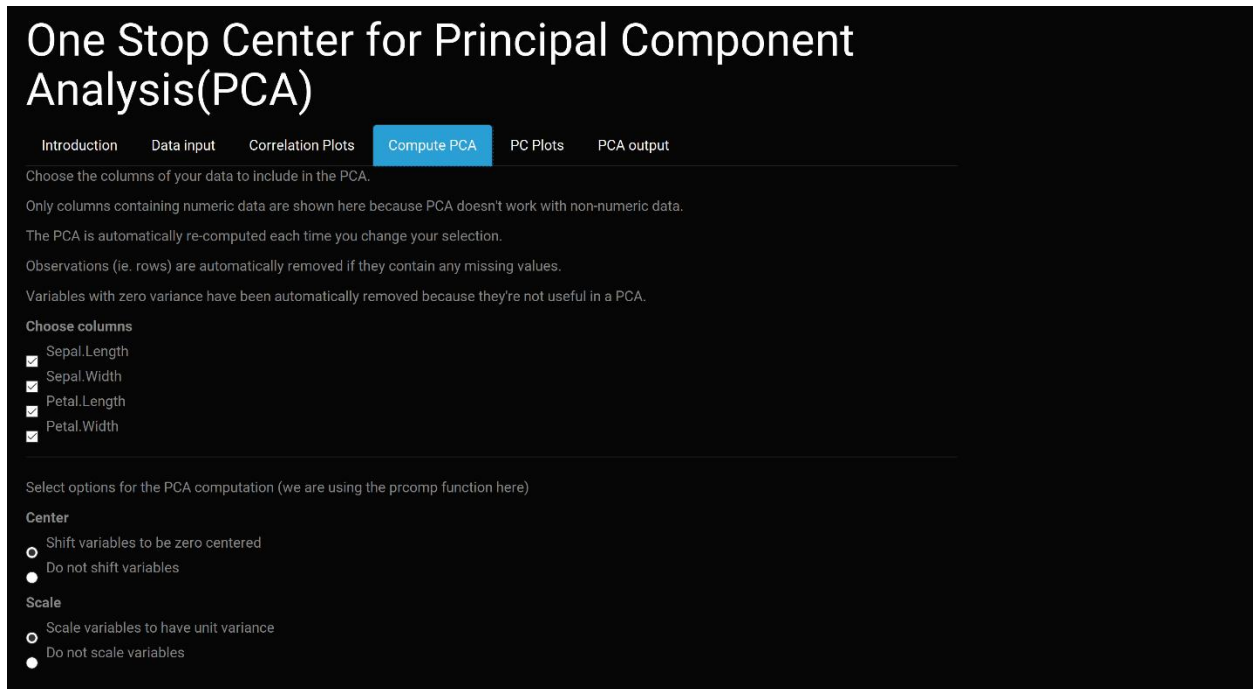


The fourth tab of the shiny application is the Compute PCA tab which enables us to compute the principal components by giving us various options to chose from such as

Center – Shift variables to zero centered or don't shift

And also scaling features.

Here is a screenshot of the fourth tab.



The screenshot shows the 'Compute PCA' tab of a Shiny application. The title is 'One Stop Center for Principal Component Analysis(PCA)'. The navigation bar includes 'Introduction', 'Data input', 'Correlation Plots', 'Compute PCA' (active), 'PC Plots', and 'PCA output'. The main content area has the following text:

Choose the columns of your data to include in the PCA.  
Only columns containing numeric data are shown here because PCA doesn't work with non-numeric data.  
The PCA is automatically re-computed each time you change your selection.  
Observations (ie. rows) are automatically removed if they contain any missing values.  
Variables with zero variance have been automatically removed because they're not useful in a PCA.

**Choose columns**

- ☒ Sepal.Length
- ☒ Sepal.Width
- ☒ Petal.Length
- ☒ Petal.Width

Select options for the PCA computation (we are using the prcomp function here)

**Center**

- ☐ Shift variables to be zero centered
- ☒ Do not shift variables

**Scale**

- ☐ Scale variables to have unit variance
- ☒ Do not scale variables

The fifth tab namely, PC plots gives us the plot of the principal components.

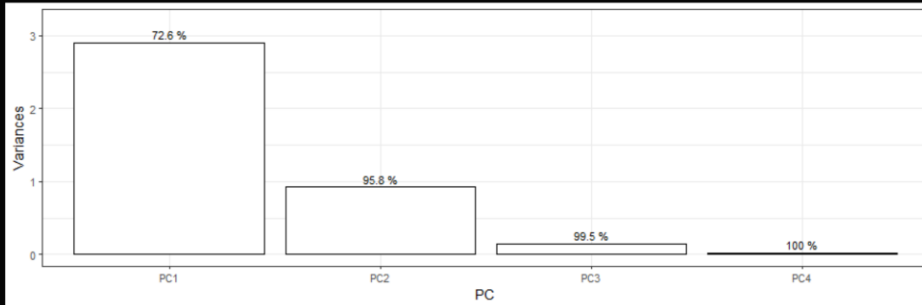
Here is screenshot of the fifth tab.

# One Stop Center for Principal Component Analysis(PCA)

[Introduction](#) [Data input](#) [Correlation Plots](#) [Compute PCA](#) **[PC Plots](#)** [PCA output](#)

## Scree plot

The scree plot shows the variances of each PC, and the cumulative variance explained by each PC (in %)



## PC plot: zoom and select points

Select the grouping variable.

And the final tab that is the PCA output tab gives us the output PCA values for the given features of the dataset.

# One Stop Center for Principal Component Analysis(PCA)

[Introduction](#) [Data input](#) [Correlation Plots](#) [Compute PCA](#) [PC Plots](#) **[PCA output](#)**

```
PC1    PC2    PC3    PC4
Sepal.Length  0.5253803 -0.35849347  0.7241684  0.2665298
Sepal.Width   -0.2558932 -0.93089991 -0.2301368 -0.1223971
Petal.Length  0.5816560 -0.02059017 -0.1371824 -0.8015194
Petal.Width   0.5658362 -0.06696136 -0.6354512  0.5211021
Importance of components:
PC1    PC2    PC3    PC4
Standard deviation  1.7047 0.9618 0.38447 0.14607
Proportion of Variance 0.7265 0.2313 0.03696 0.00533
Cumulative Proportion 0.7265 0.9577 0.99467 1.00000
```