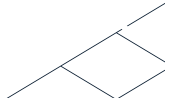**Get FREE Unlimited Access through April 30!**

Ends in   4 days 10 hrs 33 mins 55 secs

Log in   Create Free Account

Luke Hayden
August 9th, 2018

MUST READ   R PROGRAMMING

Principal Component Analysis in R

In this tutorial, you'll learn how to use PCA to extract data with many variables and create visualizations to display that data.

Principal Component Analysis (PCA) is a useful technique for exploratory data analysis, allowing you to better visualize the variation present in a dataset with many variables. It is particularly helpful in the case of "wide" datasets, where you have many variables for each sample. In this tutorial, you'll discover PCA in R.

More specifically, you'll tackle the following topics:

- You'll first go through an introduction to PCA: you'll learn about principal components and how they relate to eigenvalues and eigenvectors.

- Then, you'll try a simple PCA with a simple and easy-to-understand data set.

- Next, you'll use the results of the previous section to plot your first PCA - Visualization is very important!

- You'll also see how you can get started on interpreting the results of these visualizations and

- How to set the graphical parameters of your plots with the `ggbiplot` package!

- Of course, you want your visualizations to be as customized as possible, and that's why you'll also cover some ways of doing additional customizations to your plots!

- Wrap-up

Introduction to PCA

As you already read in the introduction, PCA is particularly handy when you're working with "wide" data sets. But why is that?

Well, in such cases, where many variables are present, you cannot easily plot the data in its raw format, making it difficult to get a sense of the trends present within. PCA allows you to see the overall "shape" of the data, identifying which samples are similar to one another and which are very different. This can enable us to identify groups of samples that are similar and work out which variables make one group different from another.

The mathematics underlying it are somewhat complex, so I won't go into too much detail, but the basics of PCA are as follows: you take a dataset with many variables, and you simplify that dataset by turning your original variables into a smaller number of "Principal Components".

But what are these exactly? Principal Components are the underlying structure in the data. They are the directions where there is the most variance, the directions where the data is most spread out. This means that we try to find the straight line that best spreads the data out when it is projected along it. This is the first principal component, the straight line that shows the most substantial variance in the data.

PCA is a type of linear transformation on a given data set that has values for a certain number of variables (coordinates) for a certain amount of spaces. This linear transformation fits this dataset to a new coordinate system in such a way that the most significant variance is found on the first coordinate, and each subsequent coordinate is orthogonal to the last and has a lesser variance. In this way, you transform a set of x correlated variables over y samples to a set of p uncorrelated principal components over the same samples.

Where many variables correlate with one another, they will all contribute strongly to the same principal component. Each principal component sums up a certain percentage of the total variation in the dataset. Where your initial variables are strongly correlated with one another, just a few principal components. As you add more principal components, you summarize more and more of the original dataset. Adding additional components makes your estimate of the total dataset more accurate, but also more unwieldy.

Eigenvalues and Eigenvectors

Just like many things in life, eigenvectors, and eigenvalues come in pairs: every eigenvector has a corresponding eigenvalue. Simply put, an eigenvector is a direction, such as "vertical" or "45 degrees", while an eigenvalue is a number telling you how much variance there is in the data in that direction. The eigenvector with the highest eigenvalue is, therefore, the first principal component.

So wait, there are possibly more eigenvalues and eigenvectors to be found in one data set?

That's correct! The number of eigenvalues and eigenvectors that exits is equal to the number of dimensions the data set has. In the example that you saw above, there were 2 variables, so the data set was two-dimensional. That means that there are two eigenvectors and eigenvalues. Similarly, you'd find three pairs in a three-dimensional data set.

We can reframe a dataset in terms of these eigenvectors and eigenvalues without changing the underlying information. Note that reframing a dataset regarding a set of eigenvalues and eigenvectors does not entail changing the data itself, you're just looking at it from a different angle, which should represent the data better.

Now that you've seen some of the theory behind PCA, you're ready to see all of it in action!

A Simple PCA

In this section, you will try a PCA using a simple and easy to understand dataset. You will use the `mtcars` dataset, which is built into R. This dataset consists of data on 32 models

of car, taken from an American motoring magazine (1974 Motor Trend magazine). For each car, you have 11 features, expressed in varying units (US units), They are as follows:

consume more fuel.

* `cyl` : Number of cylinders: more powerful cars often have more cylinders

* `disp` : Displacement (cu.in.): the combined volume of the engine's cylinders

* `hp` : Gross horsepower: this is a measure of the power generated by the car

* `drat` : Rear axle ratio: this describes how a turn of the drive shaft corresponds to a turn of the wheels. Higher values will decrease fuel efficiency.

* `wt` : Weight (1000 lbs): pretty self-explanatory!

* `qsec` : 1/4 mile time: the cars speed and acceleration

* `vs` : Engine block: this denotes whether the vehicle's engine is shaped like a "V", or is a more common straight shape.

* `am` : Transmission: this denotes whether the car's transmission is automatic (0) or manual (1).

* `gear` : Number of forward gears: sports cars tend to have more gears.

* `carb` : Number of carburetors: associated with more powerful engines

Note that the units used vary and occupy different scales.

Compute the Principal Components

Because PCA works best with numerical data, you'll exclude the two categorical variables (`vs` and `am`). You are left with a matrix of 9 columns and 32 rows, which you pass to the `prcomp()` function, assigning your output to `mtcars.pca` . You will also set two arguments, `center` and `scale` , to be `TRUE` . Then you can have a peek at your PCA object with `summary()` .

```
mtcars.pca <- prcomp(mtcars[,c(1:7,10,11)], center = TRUE,scale. = TRUE)

summary(mtcars.pca)
```

```
## Importance of components:
##                          PC1    PC2     PC3     PC4     PC5     PC6
## Standard deviation     2.3782 1.4429 0.71008 0.51481 0.42797 0.35184
## Proportion of Variance 0.6284 0.2313 0.05602 0.02945 0.02035 0.01375
## Cumulative Proportion  0.6284 0.8598 0.91581 0.94525 0.96560 0.97936
##                           PC7    PC8     PC9
## Standard deviation     0.32413 0.2419 0.14896
## Proportion of Variance 0.01167 0.0065 0.00247
## Cumulative Proportion  0.99103 0.9975 1.00000
```

You obtain 9 principal components, which you call PC1-9. Each of these explains a percentage of the total variation in the dataset. That is to say: PC1 explains 63% of the total variance, which means that nearly two-thirds of the information in the dataset (9 variables) can be encapsulated by just that one Principal Component. PC2 explains 23% of the variance. So, by knowing the position of a sample in relation to just PC1 and PC2, you can get a very accurate view on where it stands in relation to other samples, as just PC1 and PC2 can explain 86% of the variance.

Let's call `str()` to have a look at your PCA object.

```
str(mtcars.pca)
```

```
## List of 5
##  $ sdev    : num [1:9] 2.378 1.443 0.71 0.515 0.428 ...
##  $ rotation: num [1:9, 1:9] -0.393 0.403 0.397 0.367 -0.312 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:9] "mpg" "cyl" "disp" "hp" ...
##   .. ..$ : chr [1:9] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : Named num [1:9] 20.09 6.19 230.72 146.69 3.6 ...
##   ..- attr(*, "names")= chr [1:9] "mpg" "cyl" "disp" "hp" ...
##  $ scale   : Named num [1:9] 6.027 1.786 123.939 68.563 0.535 ...
##   ..- attr(*, "names")= chr [1:9] "mpg" "cyl" "disp" "hp" ...
##  $ x       : num [1:32, 1:9] -0.664 -0.637 -2.3 -0.215 1.587 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:32] "Mazda RX4" "Mazda RX4 Wag" "Datsun 710" "Hornet 4 Drive"
```

```
##   .. ..$ : chr [1:9] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```

information:

- The center point (`$center`), scaling (`$scale`), standard deviation(`sdev`) of each principal component

- The relationship (correlation or anticorrelation, etc) between the initial variables and the principal components (`$rotation`)

- The values of each sample in terms of the principal components (`$x`)

Plotting PCA

Now it's time to plot your PCA. You will make a biplot, which includes both the position of each sample in terms of PC1 and PC2 and also will show you how the initial variables map onto this. You will use the `ggbiplot` package, which offers a user-friendly and pretty function to plot biplots. A biplot is a type of plot that will allow you to visualize how the samples relate to one another in our PCA (which samples are similar and which are different) and will simultaneously reveal how each variable contributes to each principal component.

Before you can get started, don't forget to first install `ggbiplot` !

```
library(devtools)
install_github("vqv/ggbiplot")
```

Next, you can call `ggbiplot` on your PCA:

```
library(ggbiplot)

ggbiplot(mtcars.pca)
```
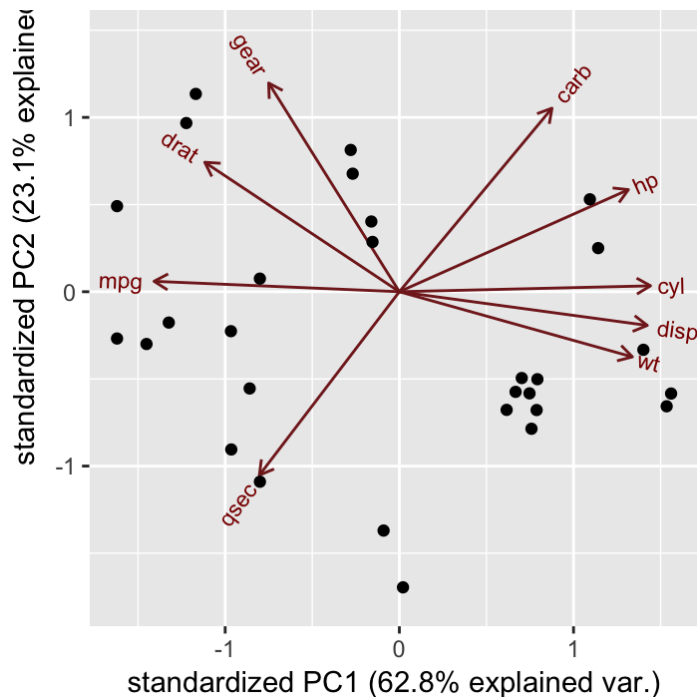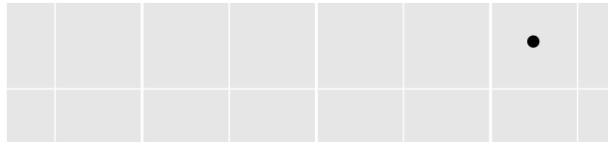
The axes are seen as arrows originating from the center point. Here, you see that the variables `hp`, `cyl`, and `disp` all contribute to PC1, with higher values in those variables moving the samples to the right on this plot. This lets you see how the data points relate to the axes, but it's not very informative without knowing which point corresponds to which sample (car).

You'll provide an argument to `ggbiplot`: let's give it the `rownames` of `mtcars` as `labels`. This will name each point with the name of the car in question:

```
ggbiplot(mtcars.pca, labels=rownames(mtcars))
```
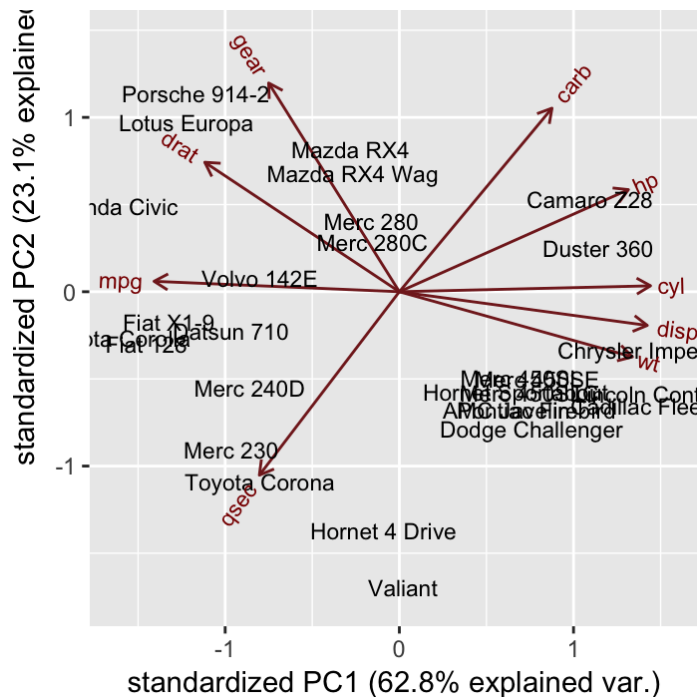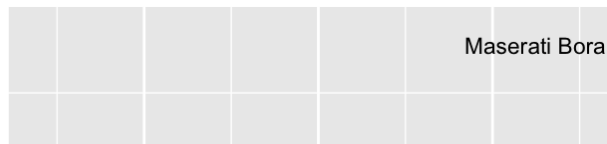
Now you can see which cars are similar to one another. For example, the Maserati Bora, Ferrari Dino and Ford Pantera L all cluster together at the top. This makes sense, as all of these are sports cars.

How else can you try to better understand your data?

Interpreting the results

Maybe if you look at the origin of each of the cars. You'll put them into one of three categories (cartegories?), one each for the US, Japanese and European cars. You make a list for this info, then pass it to the `groups` argument of ggbiplot. You'll also set the `ellipse` argument to be `TRUE`, which will draw an ellipse around each group.

```
mtcars.country <- c(rep("Japan", 3), rep("US",4), rep("Europe", 7),rep("US",3), "Eu

ggbiplot(mtcars.pca,ellipse=TRUE,  labels=rownames(mtcars), groups=mtcars.country)
```
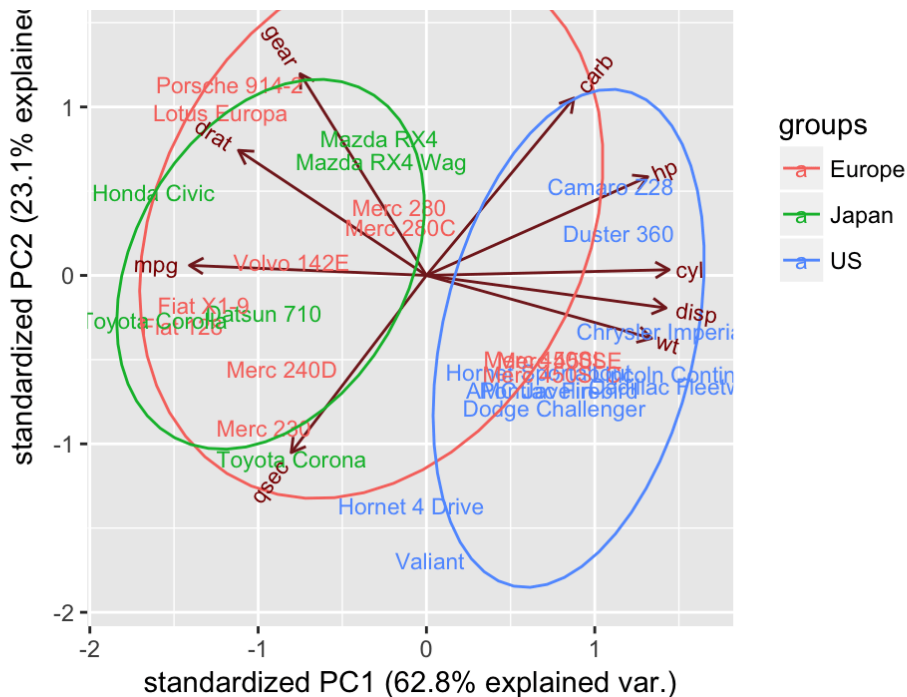
Now you see something interesting: the American cars form a distinct cluster to the right. Looking at the axes, you see that the American cars are characterized by high values for `cyl`, `disp`, and `wt`. Japanese cars, on the other hand, are characterized by high `mpg`. European cars are somewhat in the middle and less tightly clustered than either group.

Of course, you have many principal components available, each of which map differently to the original variables. You can also ask `ggbiplot` to plot these other components, by using the `choices` argument.

Let's have a look at PC3 and PC4:

```
ggbiplot(mtcars.pca,ellipse=TRUE,choices=c(3,4),   labels=rownames(mtcars), groups=
```
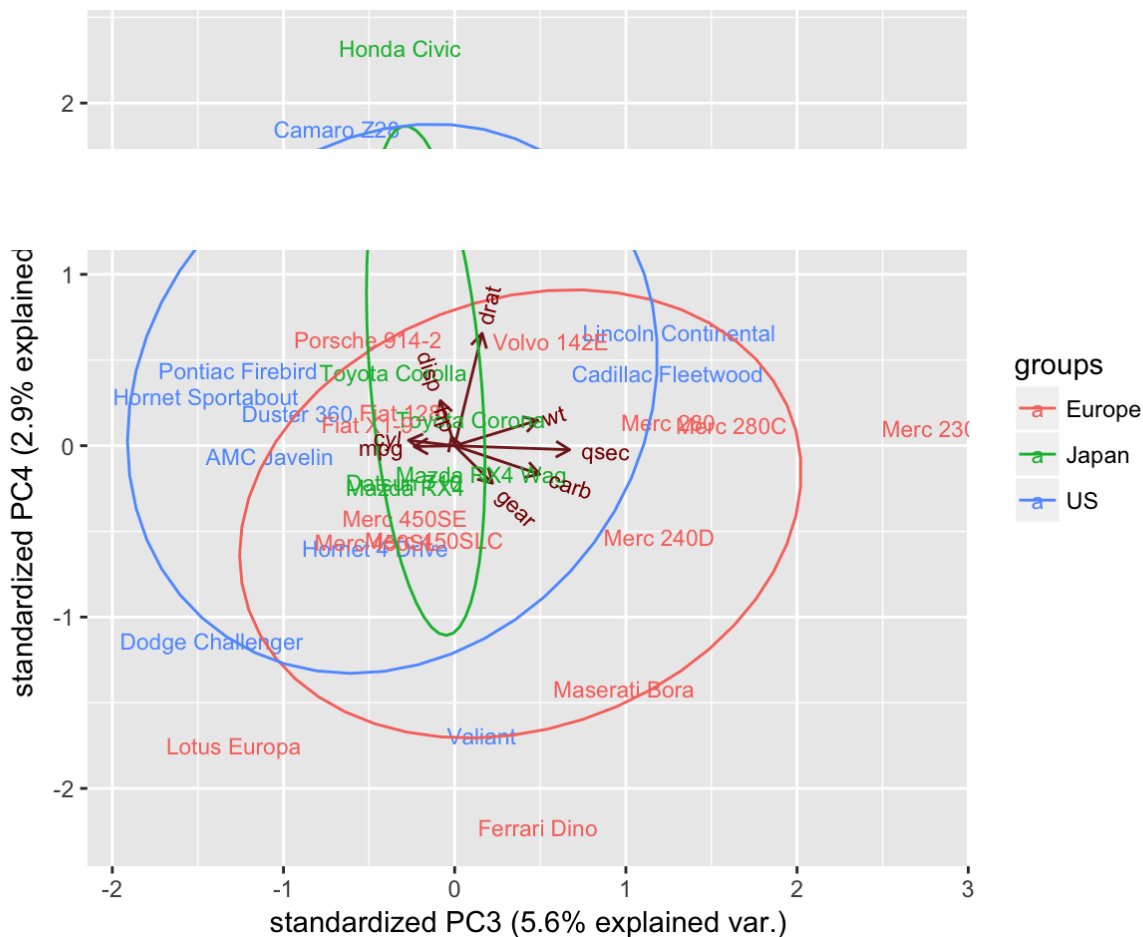
You don't see much here, but this isn't too surprising. PC3 and PC4 explain very small percentages of the total variation, so it would be surprising if you found that they were very informative and separated the groups or revealed apparent patterns.

Let's take a moment to recap: having performed a PCA using the `mtcars` dataset, we can see a clear separation between American and Japanese cars along a principal component that is closely correlated to `cyl`, `disp`, `wt`, and `mpg`. This provides us with some clues for future analyses; if we were to try to build a classification model to identify the origin of a car, these variables might be useful.

Graphical parameters with `ggbiplot`

There are also some other variables you can play with to alter your biplots. You can add a circle to the center of the dataset (`circle` argument):

```
ggbiplot(mtcars.pca,ellipse=TRUE,circle=TRUE, labels=rownames(mtcars), groups=mtcar
```
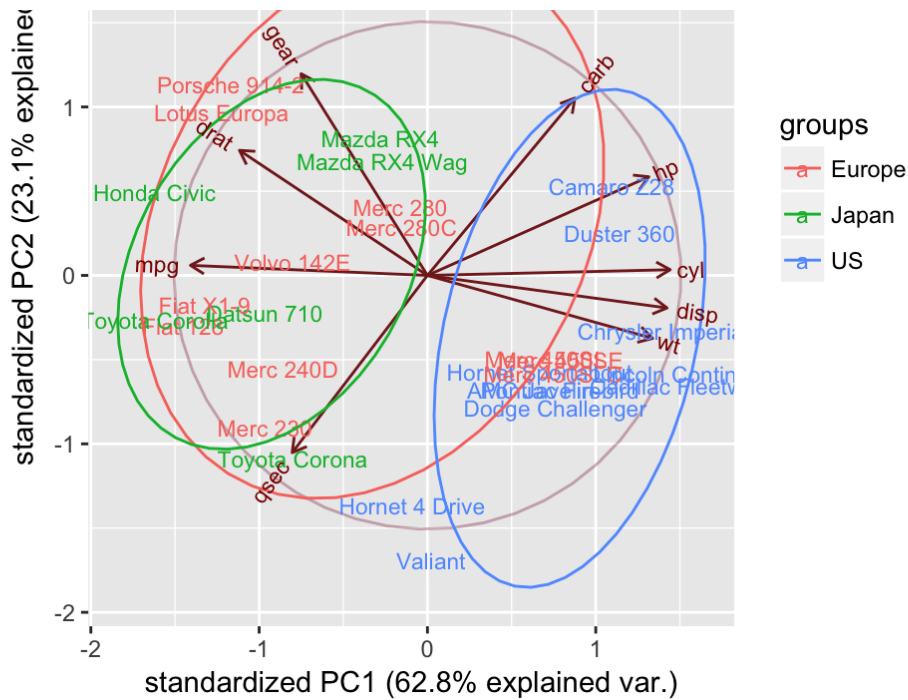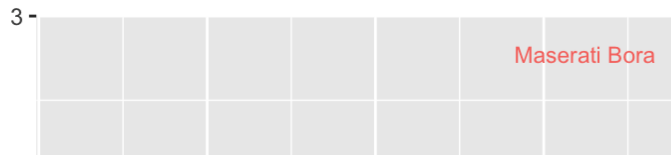
You can also scale the samples (`obs.scale`) and the variables (`var.scale`):

```
ggbiplot(mtcars.pca,ellipse=TRUE,obs.scale = 1, var.scale = 1,  labels=rownames(mtc
```
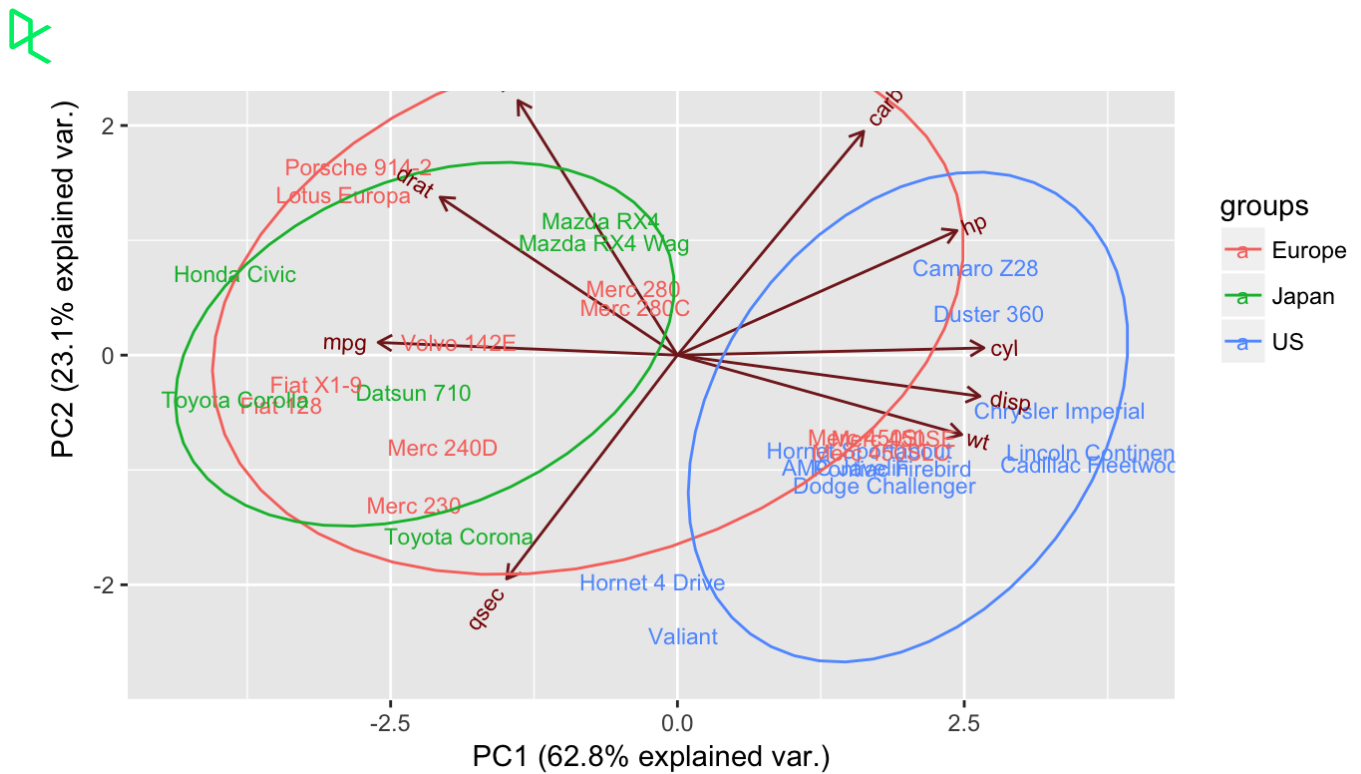
You can also remove the arrows altogether, using `var.axes`.

```
ggbiplot(mtcars.pca,ellipse=TRUE,obs.scale = 1, var.scale = 1,var.axes=FALSE,    lab
```

Customize `ggbiplot`

As `ggbiplot` is based on the `ggplot` function, you can use the same set of graphical parameters to alter your biplots as you would for any `ggplot`. Here, you're going to:

- Specify the colours to use for the groups with `scale_colour_manual()`

- Add a title with `ggtitle()`

- Specify the `minimal()` theme

- Move the legend with `theme()`

```
ggbiplot(mtcars.pca,ellipse=TRUE,obs.scale = 1, var.scale = 1,  labels=rownames(mtc
  scale_colour_manual(name="Origin", values= c("forest green", "red3", "dark blue")
  ggtitle("PCA of mtcars dataset")+
  theme_minimal()+
  theme(legend.position = "bottom")
```

PCA of mtcars dataset

Adding a new sample

Okay, so let's say you want to add a new sample to your dataset. This is a very special car, with stats unlike any other. It's super-powerful, has a 60-cylinder engine, amazing fuel economy, no gears and is very light. It's a "spacecar", from Jupiter.

Can you add it to your existing dataset and see where it places in relation to the other cars?

You will add it to `mtcars`, creating `mtcarsplus`, then repeat your analysis. You might expect to be able to see which region's cars it is most like.

```
spacecar <- c(1000,60,50,500,0,0.5,2.5,0,1,0,0)

mtcarsplus <- rbind(mtcars, spacecar)
mtcars.countryplus <- c(mtcars.country, "Jupiter")

mtcarsplus.pca <- prcomp(mtcarsplus[,c(1:7,10,11)], center = TRUE,scale. = TRUE)

ggbiplot(mtcarsplus.pca, obs.scale = 1, var.scale = 1, ellipse = TRUE, circle = FAL
```
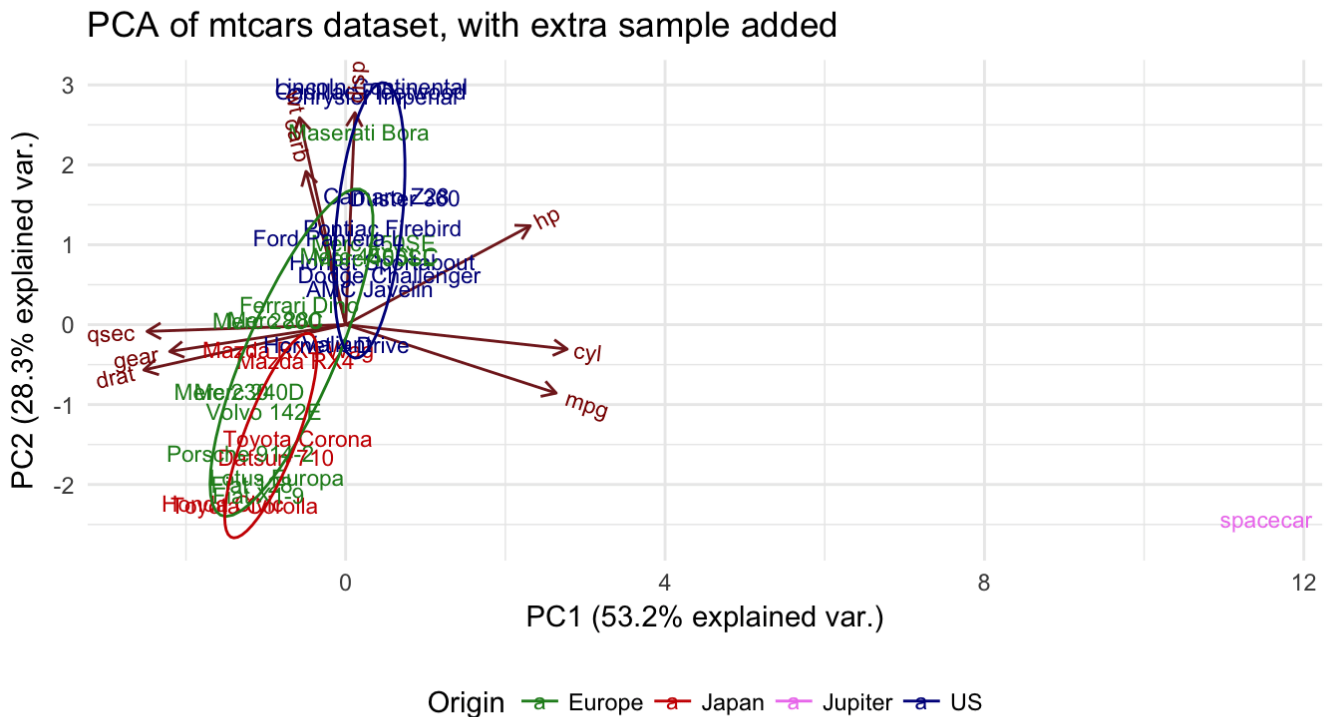
```
scale_colour_manual(name="Origin", values= c("forest green", "red3", "violet", "d
ggtitle("PCA of mtcars dataset, with extra sample added")+
theme_minimal()+
```


But that would be a naive assumption! The shape of the PCA has changed drastically, with the addition of this sample. When you consider this result in a bit more detail, it actually makes perfect sense. In the original dataset, you had strong correlations between certain variables (for example, `cyl` and `mpg`), which contributed to PC1, separating your groups from one another along this axis. However, when you perform the PCA with the extra sample, the same correlations are not present, which warps the whole dataset. In this case, the effect is particularly strong because your extra sample is an extreme outlier in multiple respects.

If you want to see how the new sample compares to the groups produced by the initial PCA, you need to *project* it onto that PCA.

Project a new sample onto the original PCA

What this means is that the principal components are defined without relation to your
`spacecar` sample, then you compute where `spacecar` is placed in relation to the other
samples by applying the transformations that your PCA has produced. You can think of

mean, you get the mean of the rest of the samples and look at `spacecar` in relation to
this.

What this means is that you simply scale the values for `spacecar` in relation to the PCA's
center (`mtcars.pca$center`). Then you apply the rotation of the PCA matrix to the
`spacecar` sample. Then you can `rbind()` the projected values for `spacecar` to the rest
of the `pca$x` matrix and pass this to `ggbiplot` as before:

```
s.sc <- scale(t(spacecar[c(1:7,10,11)]), center= mtcars.pca$center)
s.pred <- s.sc %*% mtcars.pca$rotation


mtcars.plusproj.pca <- mtcars.pca
mtcars.plusproj.pca$x <- rbind(mtcars.plusproj.pca$x, s.pred)


ggbiplot(mtcars.plusproj.pca, obs.scale = 1, var.scale = 1, ellipse = TRUE, circle
  scale_colour_manual(name="Origin", values= c("forest green", "red3", "violet", "d
  ggtitle("PCA of mtcars dataset, with extra sample projected")+
  theme_minimal()+
  theme(legend.position = "bottom")
```

PCA of mtcars dataset, with extra sample projected
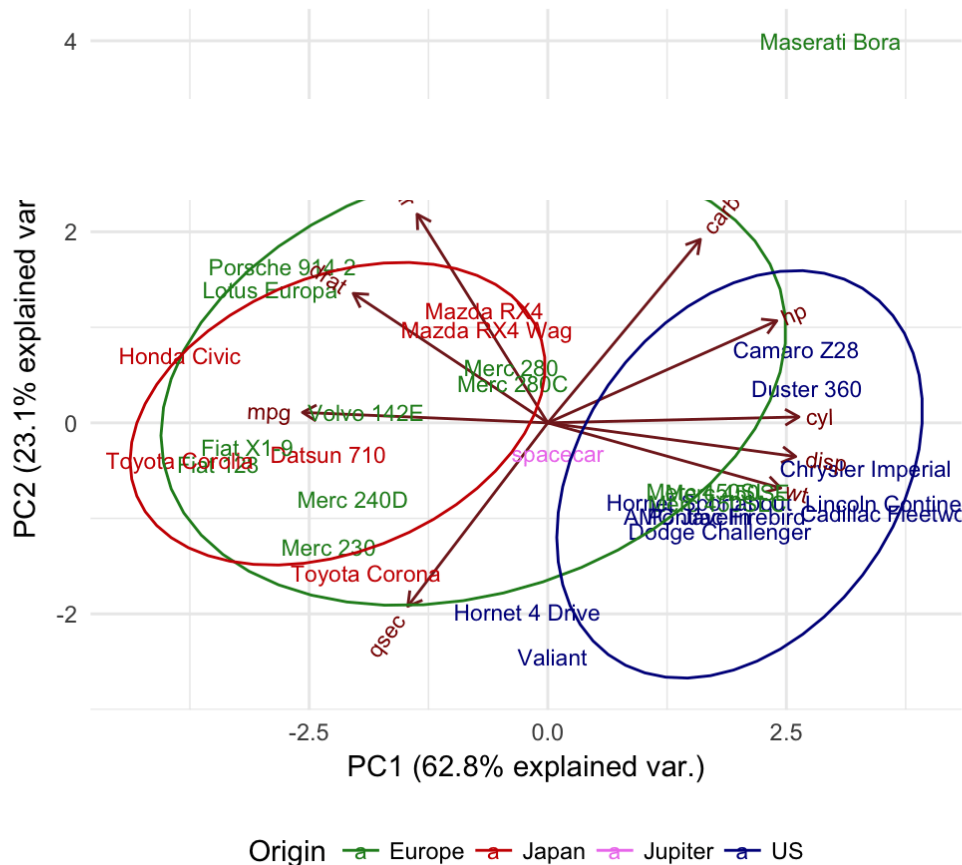
This result is drastically different. Note that all the other samples are back in their initial positions, while `spacecar` is placed somewhat near the middle. Your extra sample is no longer skewing the overall distribution, but it can't be assigned to a particular group.

But which is better, the projection or the recomputation of the PCA?

It depends somewhat on the question that you are trying to answer; the recomputation shows that `spacecar` is an outlier, the projection tells you that you can't place it in one of the existing groups. Performing both approaches is often useful when doing exploratory data analysis by PCA. This type of exploratory analysis is often a good starting point before you dive more deeply into a dataset. Your PCAs tell you which variables separate American cars from others and that `spacecar` is an outlier in our dataset. A possible next step would be to see if these relationships hold true for other cars or to see how cars cluster by marque or by type (sports cars, 4WDs, etc).

Wrap-up

So, there you have it!

You have learned the principles of PCA, how to create a biplot, how to fine-tune that plot and have seen two different methods for adding samples to a PCA analysis. Thanks for reading!

If you would like to learn more about R, take DataCamp's free Introduction to R course.

▲
200

💬
88

f   twitter   in

COMMENTS

David Passmore
14/08/2018 03:37 PM

Superb explanation. Posted for my students to see.

▲ **13**

Wishmore Stanley
20/08/2018 04:24 PM

very insightful and practical was having difficulty finding an application for pca. thanks.

▲ **5**

Leandro Marx
12/09/2018 09:30 PM

Is there a way to do this with ggplot2 tool? ggbiplot isn't available for 3.5.1 :(

▲ **8**

Albert de Roos
23/09/2018 03:42 PM

Use the install_github from devtools that attempts to install a package directly from GitHub. Did the trick for me:

install.packages("devtools")

library(devtools)

install_github("vqv/ggbiplot")

library(ggbiplot)

▲ 8

Simone Rabeling
15/10/2018 10:05 PM

Sohil Gala
14/11/2018 11:01 PM

i get this "ERROR: lazy loading failed for package 'ggbiplot'" any ideas on what to do here

▲ 9

Jyothsna Harithsa
22/04/2019 11:23 PM

install.packages("devtools")

install.packages("fs")

library(devtools)

install_github("vqv/ggbiplot")

library(ggbiplot)

▲ 2

RichaKr Kumari
13/09/2018 09:43 AM

Best ever Tutorial for PCA. Thank you for sharing your knowledge.

▲ 3

Kevin Pan
18/09/2018 06:48 PM

Thank you. It helps me a lot.

▲ 2

Uli Wellner
30/09/2018 01:36 PM

perfect shortcut intro. thanks !

▲ 2

Dr Pankaj Kumar Medhi
04/10/2018 04:47 PM

Great explanation! new insights about the projection of new sample

Miguel Barbosa
16/10/2018 11:11 PM

Thanks for the great tutorial!

▲ 1

Sourav Mandal
18/10/2018 09:45 AM

Thanks very much for touching upon a much awaited topic! Are you planning to do a tutorial on things like partial least square regression (PLSR) etc?

▲ 2

Harshali Chaudhari
23/10/2018 12:47 PM

Very useful information... very well explained.. Thank you for sharing it..

▲ 2

Lucas Ishikawa
09/11/2018 07:32 PM

Amazing! So rich and still simple in every single detail
Thanks for sharing

▲ 1

Dexter Pante
10/11/2018 12:32 PM

Just a question, in the graph where you plotted PC1 and PC2, you said that "hp, cyl and disp all contribute to PC1". Does this mean that wt and carb do not contribute to PC1? And what was the explanation for this. I'd like also to know of all the variables you mentioned that contribute to PC1 which one has the most contribution?

Thank you for taking time to read my comment. I hope you could enlighten me on this.

▲ 7

Luke Hayden
12/11/2018 07:37 PM

hp, cyl and disp contribute most strongly to PC1, while wt and carb also contribute, but less
strongly. Strictly speaking, every variable will contribute to each PC. The degree of

▲ 2

Mark Vermeersch
18/01/2019 02:36 PM

Actually cyl, disp and wt contribute more to PC1. cyl 0.402; disp 0.397; hp 0.367; wt
0.3734. Maybe this could be updated in the tutorial? Great work!

▲ 1

Philip Doyle
14/11/2018 01:54 PM

Great tutorial! Exceptionally clear writing (which I find is rare for R tutorials)

▲ 1

Philip Doyle
15/11/2018 11:12 AM

For some reason I can't utilise ggbiplot after installing. After entering - library(ggbiplot) - I get
the response:

Loading required package: ggplot2

Error: package or namespace load failed for 'ggplot2' in loadNamespace(j <- i[[1L]], c(lib.loc,
.libPaths()), versionCheck = vI[[j]]):

 namespace 'rlang' 0.2.0 is already loaded, but >= 0.2.1 is required

Error: package 'ggplot2' could not be loaded

In addition: Warning message:

package 'ggplot2' was built under R version 3.4.4

▲ 3

alvaro barroqueiro
22/03/2019 10:32 AM

Philipe, did you try to install and run a older version of R?

▲ 0

Daniel Gonzalez
24/07/2019 08:16 PM

Apparently what you need is to update rlang

Try quit the session and then go back and run **install.packages("rlang")** (You may try this without quiting R but it will probably ask you to finish the session anyway).

~~Doing that will update your package (don't try **update.packages()** as it will update all your~~

▲ 1

JEREMIAH KABISSA
15/11/2018 07:47 PM
Dear Luke

How are you. My name is Jeremiah Joe Kabissa a student of KU Leuven, I real want to learn different tactics of using R, for various analytical tools. However for now I real like to learn PCA in R, FA in R, PLS in R and the like in relation to the former three.

Sincerely

Jeremiah

▲ 2

Simone Schütz
20/11/2018 02:52 PM
Thanks. How to adapted the PCA if my data includes missing values?

▲ 7

Tram Ly
21/11/2018 07:22 PM
Dear Luke ,

Thanks you so much for your lesson

I don't understand about interpreter result with " the Japanese cars, on the other hand, are characterized by high `mpg` " `because I think low mpg is right because the` arrow of mpg is on negative side.

I am so appreciate if you explain

Thanks you so much

▲ 5

Abdellatif EL MSAYRYB
11/12/2018 09:16 AM

Hello, i'm trying to install the ggbiplot package but doesn't work for me, is there a difference between the ggplot package and ggbiplot in the python 3 ?

▲ 1

15/12/2018 11:22 AM

Both ggplot and ggbiplot are packages for R, not for Python. ggbiplot depends on ggplot, however.

▲ 1

Elijah Juma
24/12/2018 05:23 AM

Hi Luke,

I'd like to understand the intuition behind how the grouping has been done. Would you be able to offer additional explanation?

▲ 5

Izzy Bizzy
06/01/2019 07:18 PM

Hello,

I have heard that the output obtained through PCA analysis can be used as predictors in regression. For example, if you have two data sets: A contains the results of an experiment with each row representing a participant response, B contains (highly variable) social information about each participant in the study. I would like to perform a PCA analysis on data set B and use the resulting vectors as predictor variables in a mixed effects regression performed on data set A. In other words, I am trying to test how can social factors account for participant behavior. How would I go about including those principal components in my regression? I am familiar with mixed effects regression in R but if anybody can provide some sample code to achieve this, that would be great.

Thanks!

▲ 4

David C
12/01/2019 10:09 PM

Like many others, ggbiplot is not installing. I get the error.

Error in install_github("vqv/ggbiplot") :

could not find function "install_github"

▲ 1

Waite Cheung
13/01/2019 07:34 AM

Hi, David. First, you should install devtools, then you can use install_github function.

install.packages("devtools")

library(devtools)

install_github("vqv/ggbiplot")

Good luck!

any problem, contact me: admin@ncrna.net

▲ 1

Jyothsna Harithsa
22/04/2019 11:22 PM

install.packages("devtools")

install.packages("fs")

library(devtools)

install_github("vqv/ggbiplot")

library(ggbiplot)

▲ 2

Isabela Pichardo
27/11/2019 07:06 PM

Great! Thanks!

▲ 1

Sobia Ahmed
21/02/2019 07:22 AM

A great tutorial on PCA. Thanks for this

▲ 1

alvaro barroqueiro
21/03/2019 07:53 PM
> library(ggbiplot)

does not exist a package with a name 'ggbiplot' ...

🔺 1

alvaro barroqueiro
21/03/2019 07:54 PM
ou I saw tge answer up there :)

🔺 1

alvaro barroqueiro
21/03/2019 08:05 PM
however ...

> library(ggbiplot)

Error in library(ggbiplot) : there is no package called 'ggbiplot'

🔺 1

alvaro barroqueiro
23/03/2019 02:09 PM
Warning in install.packages :

  package 'ggbiplot' is not available (for R version 3.5.3)

can I install older versions of R?

🔺 1

Jyothsna Harithsa
22/04/2019 11:21 PM
install.packages("devtools")

install.packages("fs")

library(devtools)

install_github("vqv/ggbiplot")

library(ggbiplot)

▲ 1

27/03/2019 03:53 AM

how to download dataset "mtcars"? Thank you

▲ 1

hellozah
02/04/2019 01:26 AM
mtcars is installed with R, that is, it's built in. If you already have R installed you can view the mtcars data set by typing mtcars at the prompt.

▲ 1

Mohan Arthanari
05/04/2019 06:32 PM

Warning in install.packages :

 package 'ggbiplot' is not available (for R version 3.5.0)

i cant install this package

▲ 1

ellen plantsoil
07/04/2019 09:52 PM
Hi Mohan,

I had the same problem, but there is away to fix .

library(devtools)

Warning message:

package 'devtools' was built under R version 3.4.4

> install_github("vqv/ggbiplot")

Downloading GitHub repo vqv/ggbiplot@master

These packages have more recent versions available.

Which would you like to update?

> 1: ggplot2 (3.1.0 -> 3.1.1) [CRAN] 2: gtable (0.2.0 -> 0.3.0) [CRAN]
>
> 3: lazyeval (0.2.1 -> 0.2.2) [CRAN] 4: munsell (0.4.3 -> 0.5.0) [CRAN]
>
> 9: stringr (1.3.1 -> 1.4.0) [CRAN] 10: tibble (1.4.2 -> 2.1.1) [CRAN]
>
> 11: CRAN packages only   12: All
>
> 13: None
>
> Enter one or more numbers separated by spaces*, or an empty line to cancel*
>
> Note that the "devtools" was built in R 3.4.4 version, we may have R 3.5 or somethings, so that's why it is not available.
>
> If you skip all the update, as **" an empty line to cancel", then you will get the ggbiplot!**
>
> I tried and it totally worked.
>
> Good luck.

▲ 2

Zamzam Al-Rawahi
11/04/2019 05:56 AM

How to group samples if I have 930 samples instead of 32?

This would be very complicated for 930 samples.

```
mtcars.country <- c(rep("Japan", 3), rep("US",4), rep("Europe", 7),rep("US",
```

▲ 6

Luke Hayden
26/04/2019 12:41 PM

What does your data look like? How do you identify groups? In this case, I identify country via part of the model name. You could make a simple case_when approach if you have simple rules to assign your samples to groups.

▲ 3

Rafik Margaryan

23/04/2019 07:24 PM

Very nice, what I was searching for.

miriama vuiyasawa

24/04/2019 09:43 PM

How did you label the categories to get ellipses? I can't follow how you labelled them.

▲ 3

Abdullah Al Mahmud

25/04/2019 05:55 PM

Thanks for the great insight. However, I did not understand one aspect. How did you select and order the country names? I mean, they weren't in the original data.

▲ 2

Yifan Feng

29/04/2019 06:10 AM

Very useful! Can you explain how to come up with the below codes? Thank you.

mtcars.country <- c(rep("Japan", 3), rep("US",4), rep("Europe", 7),rep("US",3), "Europe", rep("Japan", 3), rep("US",4), rep("Europe", 3), "US", rep("Europe", 3))

▲ 3

marco arena

08/05/2019 06:29 AM

excellent tutorial

▲ 1

Pete De Jager

13/05/2019 06:42 PM

Exceptional article, covering PCA in depth and clarifying all of the mysteries. All the examples work flawlessly and the explanations are lucid. Many thanks!

▲ 1

Keiana Dunn

15/05/2019 07:44 AM

I installed install.packages("fs") and them received an error: Warning in install.packages :
package 'library(devtools)' is not available (for R version 3.4.4). Also Warning in
install.packages :

▲ 1

Graeme Dean
17/05/2019 06:08 PM

Hi everyone, I am following the tutorial and so far am delighted with the explanation, but now I
have hit a wall, the package is not available. I have tired several times to download it but
always returns the same message: package 'ggbiplot' is not available (for R version 3.5.2)

▲ 1

> **Luke Hayden**
> 21/05/2019 09:56 PM
>
> Did you install it via the install_github() function from devtools?
>
> ▲ 1

Tasos Baltadakis
03/06/2019 07:53 PM

Hello there, its a very helpful tutorial considering my level.

I wanted to ask if there is a command through ggbiplot to change the arrows from red to
black.

Cheers

▲ 1

> **Luke Hayden**
> 10/06/2019 08:15 PM
>
> I don't know of any easy way to do so from within ggbiplot. You might be able to pass a
> colour for the arrows via ggplot' theming functions.
>
> ▲ 3

Salahuddin Khan
19/06/2019 11:58 PM

Wow! Loads of plain information. Just need to understand the following codes:

```
mtcars.country <- c(rep("Japan", 3), rep("US",4), rep("Europe", 7),rep("US",

ggbiplot(mtcars.pca,ellipse=TRUE,  labels=rownames(mtcars), groups=mtcars.cc
```

Where did the numbers 3, 4, 7 etc.. after each country name come from?

▲ 2

Jiapeng He
25/06/2019 07:01 PM

The first number "3" represent the first 3 cars are from Japan, and the "4" represent the following 4 cars are from US. All the numbers add up to 32, which means 32 type of cars in the mtcars data set

▲ 2

huiping zhou
19/07/2019 10:46 PM

I cannot install ggbiplot

> library(devtools)

> install_github("vqv/ggbiplot")

Downloading GitHub repo vqv/ggbiplot@master

Error in utils::download.file(url, path, method = method, quiet = quiet, :

  cannot open URL 'https://api.github.com/repos/vqv/ggbiplot/tarball/master'

in this URL , I didn't find the ggbiplot zip file...

please help!

▲ 1

Debomitra Dey
05/08/2019 08:49 PM

Great tutorial. Thanks! I am unable to download the package ggbiplot.

▲ 2

Jack Dundas
14/08/2019 11:51 PM
Thanks for the awesome write up!

position, which indicates that PC1 and PC2 don't do a great job of "explaining" the space car.
Correspondingly, the Valiant is summarized nicely by PC2, but not PC1.

Further, the contsituents of PC1 are primarily ctl, disp and wt. Does this mean that cyl, disp
and wt are all highly colinear?
Can the same be said about gear and carb due to their high PC2 values?

▲ 1

> **Luke Hayden**
> 21/08/2019 12:13 PM
> "In the last chart, with the spacecar, my interpretation is that it is placed right near the 0,0
> position, which indicates that PC1 and PC2 don't do a great job of "explaining" the space
> car.
> Correspondingly, the Valiant is summarized nicely by PC2, but not PC1. "
>
> Not quite. You can't think of a PC as "explaining" a sample, but as summarising the
> differences between samples. The Valiant is different from the Dino according to PC2, but
> not according to PC1. So, PC1 doesn't capture the difference between the Dino and the
> Valiant, but PC2 does.
>
> "Further, the contsituents of PC1 are primarily ctl, disp and wt. Does this mean that cyl,
> disp and wt are all highly colinear?"
>
> Generally speaking, yes. Variables that contribute strongly to a PC will tend to be
> correlated. Pairwise plotting of variables is the best way to tease these relationships out.
>
> "Can the same be said about gear and carb due to their high PC2 values?"
>
> Probably, but see above regarding pairwise plotting.
>
> ▲ 1

anwesha Saha
19/08/2019 02:55 PM
Hi Luke,

It an excellent tutorial. Thank you so much for this. I am constantly having trouble with
installing ggbiplot. Do you have any suggestions. I tried

> options(timeout=20000)

> install_github("vqv/ggbiplot")

it still did not resolve the issue. Do you have any other options to prepare similar graphs other than ggbiplot package.

▲ 2

Hector Alvaro Rojas
24/08/2019 04:48 PM
 Hi Luke:

Congratulations man!

You have made a great article on this topic, especially the biplot explanation.

This is a great tutorial on PCA.

Superb explanation. Very insightful and practical at the same time.

I think this is one of the best articles I have read about it.

Now, please let me know by the time you create a similar article but on the "Multidimensional scaling" topic.

My best regards to you!

▲ 1

Kaitlin DeAeth
26/08/2019 05:02 PM
Question. How were you able to label with data by car names if they weren't in the original data? How do you incorporate the characters of your variables if all the data in a PCA needs to be numerical?

▲ 1

Xin Sun
30/09/2019 03:25 AM
Informative article. Thanks!

▲ 1

Claudia Silva

10/10/2019 03:25 AM

```
library(ggbiplot) is not available for R 3.6
```

I cannot upload here the output when I tried to install from GitHub. But it says that it can't do install the package ggbiplot.

is there any other library to do the same?

▲ 3

Yang Wang

10/10/2019 11:04 PM

Hello, thank you for the explanation! I have a question about the ggbiplot. So I looked around and found there seems no way to modify the width of the ellipse, in base r plots, we just use lwd = 1 or whatever values, but in ggbiplot, is there also a way to change this? Also, is it possible to add a polygon of the ellipse? Thank you, the two questions really confuse me a some time!

▲ 1

Lina Gao

28/10/2019 07:31 PM

Dear Luke

How about the contribution of mpg and drat to PC1? Can we say that the original variables with negative loading contribute less than the ones with positive loading? Another question is whether we can use predict function to get the scores of the new sample and draw it in the plot

▲ 1

Debbie Jenkins

04/11/2019 05:51 PM

Hello - like so many others I found this tutorial very helpful. Unfortunately, I can not, after trying everything that I've found on-line, download ggbiplot.

package 'digest' successfully unpacked and MD5 sums checked

Error: Failed to install 'ggbiplot' from GitHub:

 (converted from warning) cannot remove prior installation of package 'digest'

In addition: Warning messages:

1: In untar2(tarfile, files, list, exdir) :

2: In untar2(tarfile, files, list, exdir) :

   skipping pax global extended headers

I hope you can help me. The package is exactly what I need :)

ps I tried   install_github("vqv/ggbiplot", force = TRUE)   but got the same message !

▲ 1

Aaron Soderstrom
13/11/2019 05:18 AM
Great introduction to plotting a PCA. Thank you for taking the time and putting this together.
Its funny over the years of using R for data science, I know a lot about cars now.. Cheers, Aaron

▲ 1

Frederico Faleiro
29/11/2019 02:58 PM
Dear @Luke Hayden, thanks for the very helpful tutorial. However I think there is an error in
your prediction part. As you use the option center and scale in the PCA, you must use both in
the prediction, but you only center the prediction. I think for this reason the spacecar are not
an outlier in your graph. Check a discussion about it here: https://stat.ethz.ch/pipermail/r-
help/2008-April/160033.html.

I tryed put an example code but the system return the following error: Purify checking
crashed :(

▲ 1

Waseem Ashfaq
03/12/2019 06:56 AM
Superb and very informative . . My question is how to make PCA representation bold (Whole
PCA graphs).

▲ 1

Angela Marcela Suarez Mayorga
03/12/2019 01:47 PM

Thanks a lot for your detailed and easy-to-follow explanation. Very useful! (Also the comments, thank you all).

▲ 1

16/01/2020 05:16 PM

Humraz Cin Tarikatı Filmi izle | Humraz Cin Tarikatı Filmi Full HD Tek Parça izle

Humraz Cin Tarikatı Humraz Cin Tarikatı, eski bir tarikat ile ilgili araştırma yapan genç bir kadının yaşadıklarını konu ediyor. Tuğba, yıllar önce anne ve babasını faili meçhul bir şekilde kaybeden genç bir kadındır. Humraz Arkeoloji bölümünde yüksek lisans yapan Tuğba, araştırma yaptığı sırada Derinsuyluk köyü çevresinde Humruz adında eski bir tarikatın izine rastlar.

Humraz Cin Tarikatı Filmi izle Bu konuyu araştırmaya karar veren Tuğba, ağabeyi Selim ve danışman hocasının da yol göstermesi ile bölgeye doğru yola koyulur. Gece geç saatlerde bölgeye giden Tuğba, burasa kan dondurucu gerçeklerle yüzleşmek zorunda kalır.

Dizi Fragmanları - Vizyondaki En İyi Filmler -

Humraz Cin Tarikatı izle, Humraz Cin Tarikatı filmi full hd izle, Humraz Cin Tarikatı filmi Tek parça izle, Humraz Cin Tarikatı Filmi Oyuncuları, Humraz Cin Tarikatı Filmi Yönetmeni, Humraz Cin Tarikatı Filmi Konusu

****** Dark Waters Filmi izle | Dark Waters Filmi Full HD Tek Parça izle

Dark Waters Dark Waters, dünyanın en büyük şirketlerinden biri ile ilgili karanlık bir sırrı açığa çıkaran bir avukatın hikayesini konu ediyor. Robert Bilott, işinde oldukça başarılı olan bir avukattır. Genç adamın ele aldığı yeni dava, toplumu yakından ilgilendirir.

Dark Waters Filmi izle Giderek artan ve nedeni açıklanamayan ölüm vakalarını araştıran Robert Bilott, yaşananların arkasında dünyanın en büyük şirketlerinden birinin olduğunu keşfeder. Olayı iyice araştırmaya başlayan Robert, gerçekleri ortaya çıkarmak için geleceğini, ailesini ve kendi yaşamını riske atar.

Dizi Fragmanları - Vizyondaki En İyi Filmler -

Dark Waters izle, Dark Waters filmi full hd izle, Dark Waters filmi Tek parça izle, Dark Waters Filmi Oyuncuları, Dark Waters Filmi Yönetmeni, Dark Waters Filmi Konusu

********* Şeker Çocuk - Honey Boy Filmi izle | Şeker Çocuk - Honey Boy Filmi Full HD Tek Parça izle

Şeker Çocuk - Honey Boy Alkol bağımlısı babasıyla ilişkisini düzeltmeye çalışan bir çocuk yıldızın hikayesini anlatacak film, genel hatlarıyla Shia LaBeouf'ın hayatından uyarlanıyor.

Şeker Çocuk - Honey Boy Filmi izle

Dizi Fragmanları - Vizyondaki En İyi Filmler -

Şeker Çocuk - Honey Boy izle, Şeker Çocuk - Honey Boy filmi full hd izle, Şeker Çocuk - Honey Boy filmi Tek parça izle, Şeker Çocuk - Honey Boy Filmi Oyuncuları, Şeker Çocuk - Honey Boy Filmi Yönetmeni, Şeker Çocuk - Honey Boy Filmi Konusu

▲ 1

sportzulu
16/01/2020 05:17 PM
Afili Aşk 30. Bölüm Fragmanı Özeti Full HD Tek Parça izle

Afili Aşk 30. Bölüm izle Afili Aşk 30. Bölüm Fragmanı yayında! Aileler arasındaki kriz, Kerem ve Ayşe'nin yollarını sonsuza dek ayırıyor! Afili Aşk yeni bölüm fragmanında Kerem ve Ayşe aldıkları boşanma kararıyla herkesi şok ediyor! Aileler arasındaki gerginlik onların yollarını ayırırken, Muhsin bu duruma sebep olmak istemiyor.

Afili Aşk 30. Bölüm Fragmanı Ailelerin sorunu yüzünden onların ayrılmasını istemeyen Muhsin bir plan yapıyor. Ayşe ise eşyalarını toplamak için eve geliyor ancak burada beklenmedik bir durumla karşılaşıyor. Muhsin, Kerem ve Ayşe'yi bir odaya kapatıyor. Baş başa kalan ikili boşanma kararından vazgeçecek mi?

Dizi Fragmanları - Vizyondaki En İyi Filmler -

Afili Aşk 30. Bölüm izle, Afili Aşk 30. Bölüm full izle, Afili Aşk 30. Bölüm hd izle, Afili Aşk 30. Bölüm tek parça izle, Afili Aşk 30. Bölüm Son Bölüm izle, Afili Aşk 30. Bölüm Yeni Bölüm izle, Afili Aşk 30. Bölüm Özeti, Afili Aşk 30. Bölüm Fragmanı
********** Benim Adım Melek 17. Bölüm Fragmanı Özeti Full HD Tek Parça izle

Benim Adım Melek 17. Bölüm izle Benim Adım Melek dizisi 16. son bölümüyle TRT 1 ekranlarında yayınlandı. Başrollerini Nehir Erdoğan, Kutsi, Rabia Soytürk, Ulvi Kahyaoğlu, Şerif Sezer, Mustafa Mert Koç'un paylaştığı Benim Adım Melek dizisinin son bölümünde; çocuklarının düzeni ve mutluluğu için sağlığını geri plana iten Melek için artık yolun sonu mu? Benim Adım Melek 17. yeni bölüm fragmanı yayınlandı ve 17. bölüme dair detaylar ortaya çıktı.

Benim Adım Melek 17. Bölüm Fragmanı Benim Adım Melek dizisinin son bölümünde; Meryem, duyduklarından dolayı şaşırmış ve ne yapacağını bilemez haldedir. Halil, Melek'in sağlığından endişe duyarken yüreğinin sesini duyabilecek midir? Gelişmeler Halil'in Meryem'le çizdiği yolda değişikliğe gitmesine sebep olacak mıdır? İşte 17. yeni bölüm fragmanı ve detaylar…

Dizi Fragmanları - Vizyondaki En İyi Filmler -

Benim Adım Melek 17. Bölüm izle, Benim Adım Melek 17. Bölüm full izle, Benim Adım Melek 17. Bölüm hd izle, Benim Adım Melek 17. Bölüm tek parça izle, Benim Adım Melek 17. Bölüm Son Bölüm izle, Benim Adım Melek 17. Bölüm Yeni Bölüm izle, Benim Adım Melek 17. Bölüm Özeti, Benim Adım Melek 17. Bölüm Fragmanı
********* Doğduğun Ev Kaderindir 3. Bölüm Fragmanı Özeti Full HD Tek Parça izle

Doğduğun Ev Kaderindir 3. Bölüm izle Gerçek bir hayat hikayesinden uyarlanan Doğduğun Ev Kaderindir dizisinin başrollerini Demet Özdemir ve İbrahim Çelikkol paylaşıyor,

yönetmenliğini Çağrı Bayrak'ın yaptığı dizinin senaryosunu ise Eylem Canpolat yazıyor. Diziseverler merak ediyor: Doğduğun Ev Kaderindir yeni bölüm fragmanı yayınlandı mı?

Doğduğun Ev Kaderindir 3. Bölüm Fragmanı Doğduğun Ev Kaderindir 2. bölümü ile TV 8

Dizi Fragmanları - Vizyondaki En İyi Filmler -

Doğduğun Ev Kaderindir 3. Bölüm izle, Doğduğun Ev Kaderindir 3. Bölüm full izle, Doğduğun Ev Kaderindir 3. Bölüm hd izle, Doğduğun Ev Kaderindir 3. Bölüm tek parça izle, Doğduğun Ev Kaderindir 3. Bölüm Son Bölüm izle, Doğduğun Ev Kaderindir 3. Bölüm Yeni Bölüm izle, Doğduğun Ev Kaderindir 3. Bölüm Özeti, Doğduğun Ev Kaderindir 3. Bölüm Fragmanı
 ******** Zemheri 2. Bölüm Fragmanı Özeti Full HD Tek Parça izle

Zemheri 2. Bölüm izle Zemheri dizisi bu akşam ilk bölümüyle ekranlara geldi. Show Tv'nin Çarşamba günleri yayınlamaya başladığı dizinin ilk bölümüyle Firuze ve Ayaz'ın da aşk hikayesi başlamış oldu. Konusu ve oyuncularıyla dikkat çeken Zemheri dizisinin 2. bölüm fragmanı da yayınlandı. İşte Zemheri yeni bölüm fragmanı ve dizi ilgili merak edilenler...

Zemheri 2. Bölüm Fragmanı Show Tv ekranlarının yeni dizisi Zemheri ilk bölümü ile bu akşam izleyici karşısına çıktı. Firuze ve Ayaz... İki güzel yürek, iki aile yükü taşıyan çocuk. Kaderin bir araya getirdiği yollarını saf, tertemiz aşklarıyla döşemektir tek dertleri. İşte Zemheri 2. yeni bölüm fragmanı ve ilk bölüm izleme ekranı...

Dizi Fragmanları - Vizyondaki En İyi Filmler -

Zemheri 2. Bölüm izle, Zemheri 2. Bölüm full izle, Zemheri 2. Bölüm hd izle, Zemheri 2. Bölüm tek parça izle, Zemheri 2. Bölüm Son Bölüm izle, Zemheri 2. Bölüm Yeni Bölüm izle, Zemheri 2. Bölüm Özeti, Zemheri 2. Bölüm Fragmanı
 ******** Kuruluş Osman 7. Bölüm Fragmanı Özeti Full HD Tek Parça izle

Kuruluş Osman 7. Bölüm izle Atv'nin sevilen dizisi Kuruluş Osman 7. bölüm fragmanı yayınlandı! Yılbaşı nedeniyle tatile giren diziler nedeniyle 'Kuruluş Osman bu akşam var mı?' sorusu diziyi yakından takip eden kişiler tarafından aratılmaya başladı.

Kuruluş Osman 7. Bölüm Fragmanı Başrollerini Burak Özçivit, Nurettin Sönmez, Ragıp Savaş ve Özge Törer gibi isimlerin paylaştığı dizi Çarşamba günü izlenen diziler arasında 1. sırada yer alıyor. Peki, Kuruluş Osman yeni bölüm ne zaman, hangi gün yayınlanacak?

Dizi Fragmanları - Vizyondaki En İyi Filmler -

Kuruluş Osman 7. Bölüm izle, Kuruluş Osman 7. Bölüm full izle, Kuruluş Osman 7. Bölüm hd izle, Kuruluş Osman 7. Bölüm tek parça izle, Kuruluş Osman 7. Bölüm Son Bölüm izle, Kuruluş Osman 7. Bölüm Yeni Bölüm izle, Kuruluş Osman 7. Bölüm Özeti, Kuruluş Osman 7. Bölüm Fragmanı

▲ 1

lignalerde

18/01/2020 06:02 AM

▲ 1

Stephana Müller

20/01/2020 08:53 PM

Great addition to the Unsupervised Machine Learning Course I took on DataCamp. Thanks!

▲ 1

lignalerde

26/01/2020 11:45 AM

Wales vs Ireland Rugby online

Six Nation Rugby is knocking at the door. In this Six Nation Rugby Wales vs Ireland match will air on. We know Wales vs Ireland match is always very enjoyable.

▲ 1

sportzulu

27/01/2020 01:02 AM

Aile Filmleri - Aksiyon Filmleri - Macera Filmleri - Animasyon Filmleri - Bilim Kurgu Filmleri - Biyografi Filmleri - Dram Filmleri - Fantastik Filmleri - Gerilim Filmleri - Korku Filmleri - Komedi Filmleri - Polisiye Filmleri - Romantik Filmleri - Savaş Filmleri - Tarih Filmleri - Western Filmleri - Dizi Fragmanları - Dizi Haberleri - Vizyondaki Filmler -

▲ 1

Infocampus Logics

03/02/2020 05:45 AM

I was more than happy to uncover this great site. I need to to thank you for your time due to this fantastic read!! I definitely enjoyed every bit of it and I have you bookmarked to see new information on your blog. Java Training in Bangalore

▲ 1

webroot safe

03/02/2020 06:46 AM

Enter Product Key | www.webroot.com/safe is the great antivirus software tool that protect your system from kind of Trojans,malware ,viruses,etc. Keep your pc smooth and moving with use of this antivirus tool .Software make your system safe and great .Webroot Login

Jonathon Silva
11/02/2020 02:00 AM

PCA is a transformation method that helps reduce a large number of variables ... Before going into the algorithm to determine the number of main components in PCA, ...

strike force heroes 3

▲ 1

Benjamin Malunda
08/02/2020 07:55 PM

greetings

when i try to run

```
install_github("vqv/ggbiplot")
```

it keeps giving me the following error:ERROR: failed to lock directory 'C:/Users/bkaso/Documents/R/win-library/3.6' for modifying

```
Try removing 'C:/Users/bkaso/Documents/R/win-library/3.6/00LOCK-
stringi'

Error: Failed to install 'ggbiplot' from GitHub:

  (converted from warning) installation of package 'stringi' had
non-zero exit status
```

```
how can i solve this
```

▲ 1

lignalerde
10/02/2020 05:51 AM

Thank you Carlton v Geelong Cats

▲ 1

Paul Cotter

library(devtools) or library(ggbiplot) where else might these be found?

```
ggbiplot(mtcars.pca)
```

▲ 2

thevapekart
18/02/2020 01:56 PM
Could you please brief more on **Feature elimination**?

Smart watches for women

▲ 1

Kamden Glade
10/03/2020 03:33 PM
Is there a way to use categorical variables to group the data from the original data set, rather than adding the grouping as you did for country?

▲ 2

🔊 Subscribe to RSS

About   Terms   Privacy