

# week11

Lu Zhang

7/28/2019

## Principal Component Analysis

two functions:

`prcomp(x, center = TRUE, scale. = TRUE)` `princomp(x, cor = FALSE, scores = TRUE)`

### Compute the Principal Components

```
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
```

show 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"
```

## Plotting PCA

```
library(devtools)
```

```
## Loading required package: usethis
```

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

```
## Skipping install of 'ggbiplot' from a github remote, the SHA1 (7325e880) has not changed since last :  
## Use `force = TRUE` to force installation
```

call ggbiplot on PCA

```
library(ggbiplot)
```

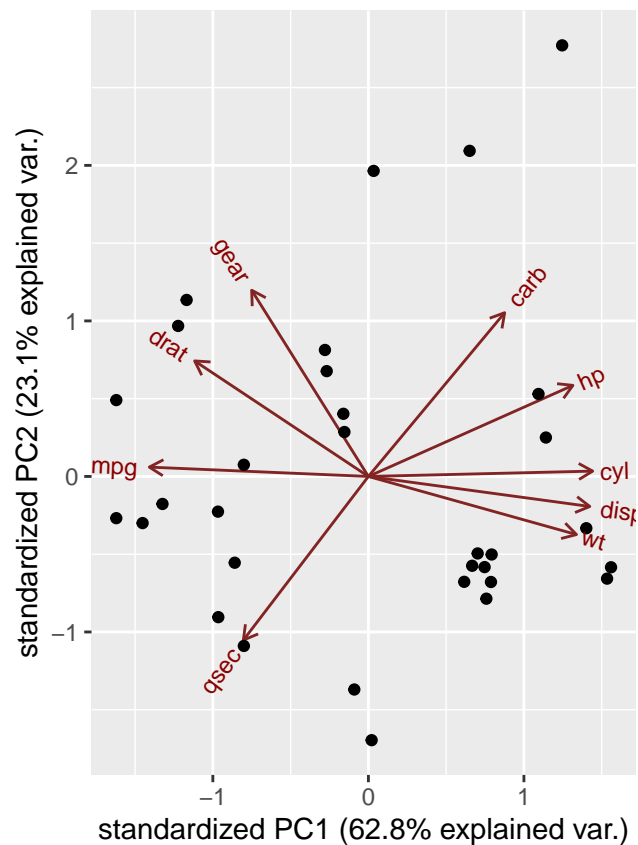
```
## Loading required package: ggplot2
```

```
## Loading required package: plyr
```

```
## Loading required package: scales
```

```
## Loading required package: grid
```

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

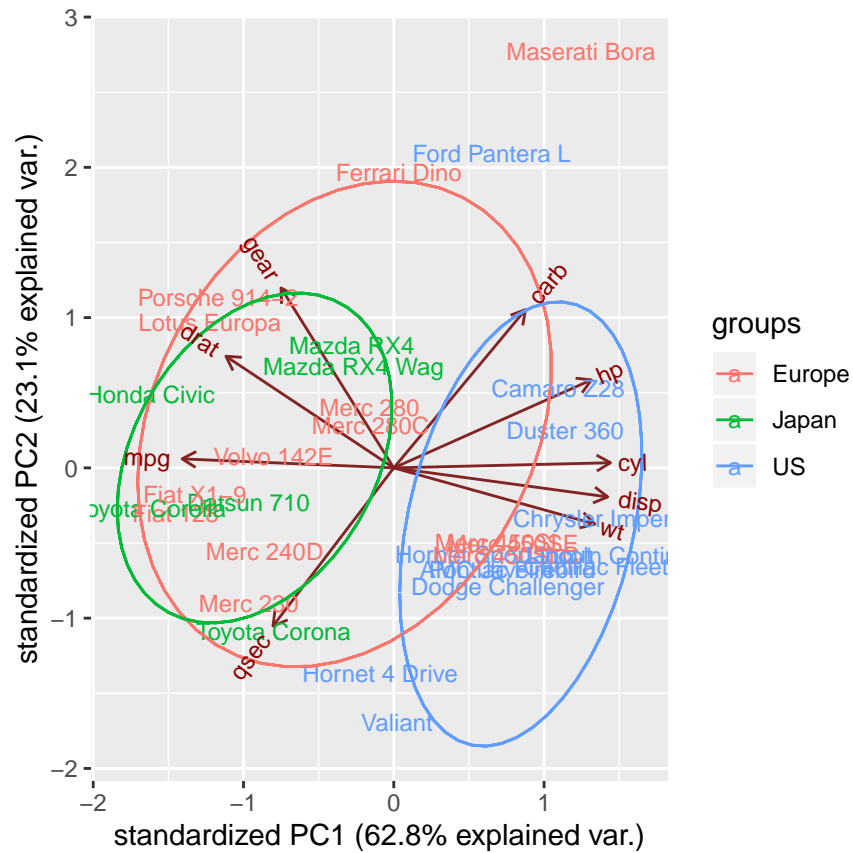


```
ggbiplot(mtcars.pca, labels=rownames(mtcars))
```



```
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)
                  )

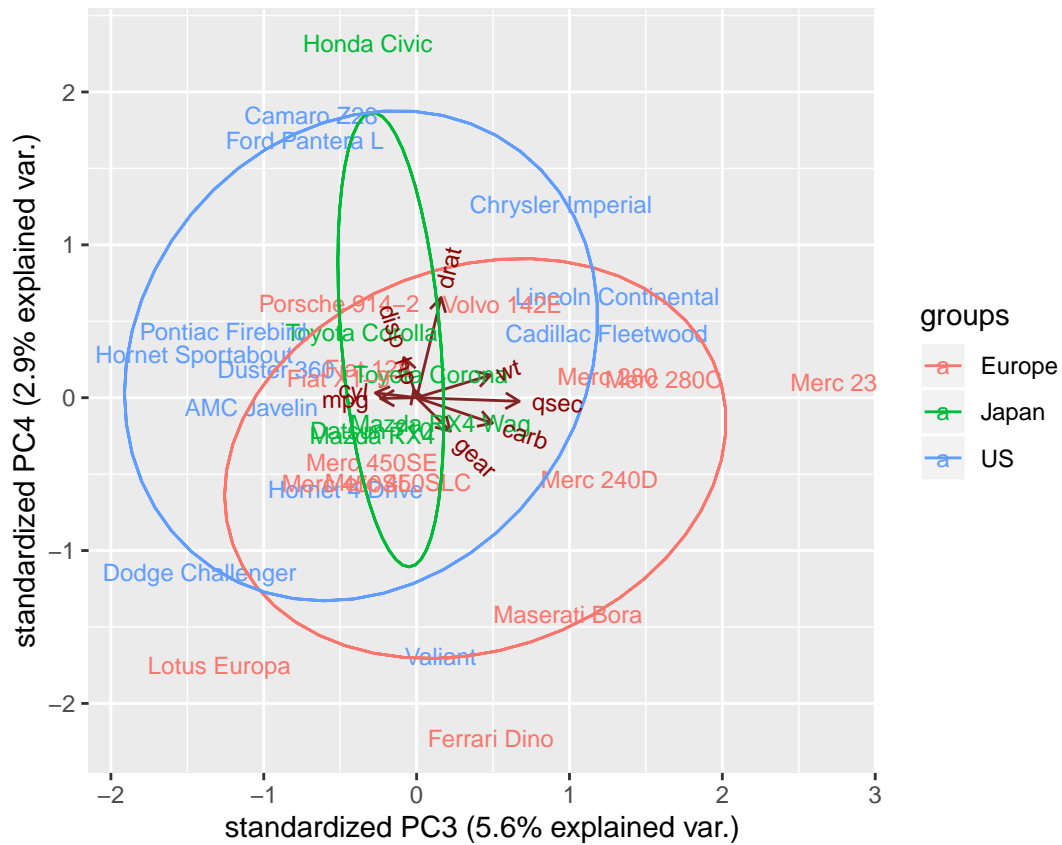
ggbiplot(mtcars.pca, ellipse=TRUE, labels=rownames(mtcars), groups=mtcars.country)
```



choices argument plot other components

eg. PC3 and PC4

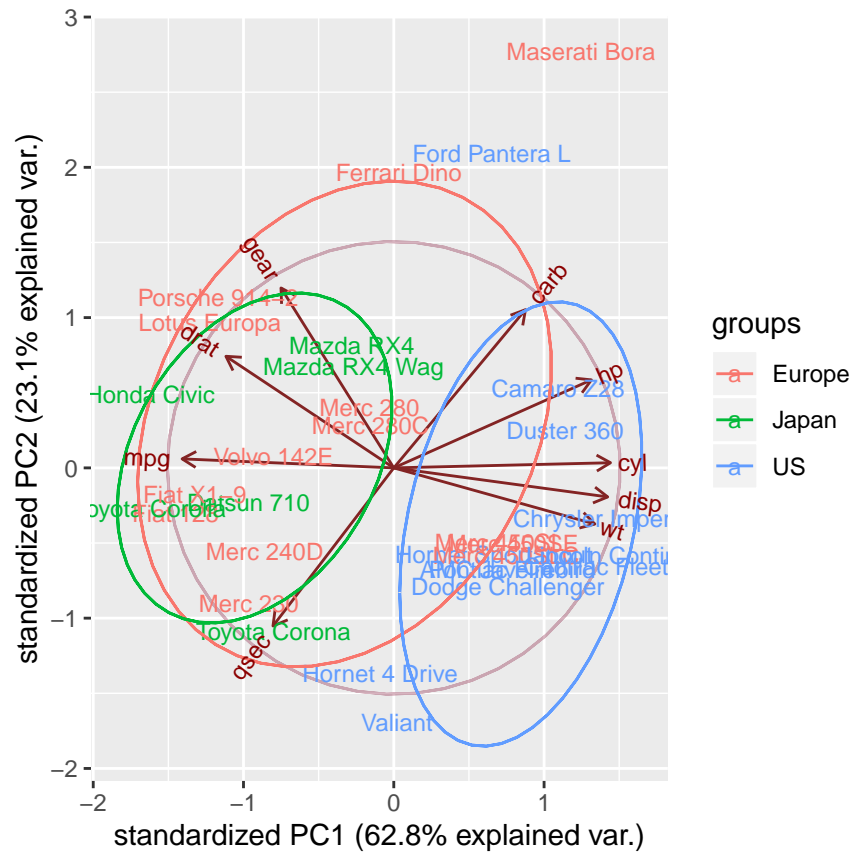
```
ggbiplot(mtcars.pca,
         ellipse=TRUE,
         choices=c(3,4),
         labels=rownames(mtcars),
         groups=mtcars.country)
```



## Graphical parameters with ggbiplot

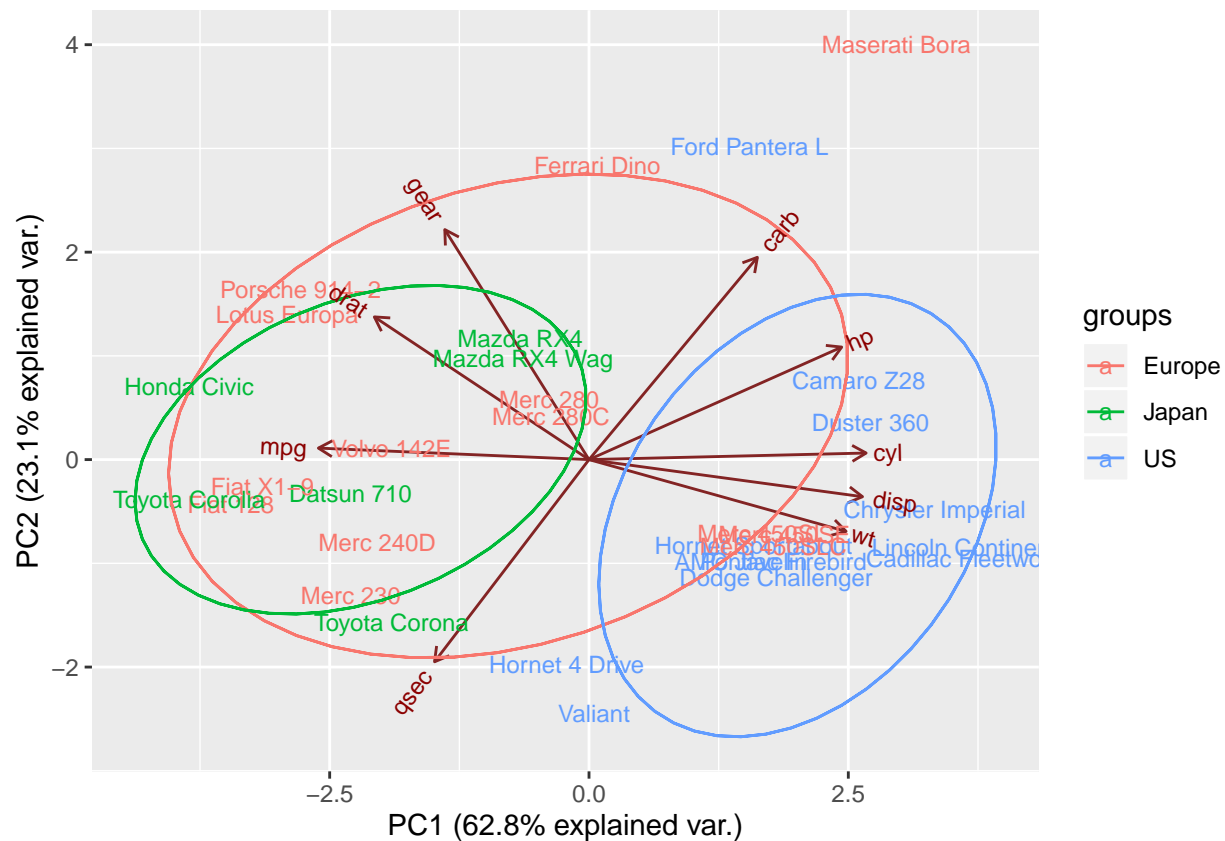
circle argument

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



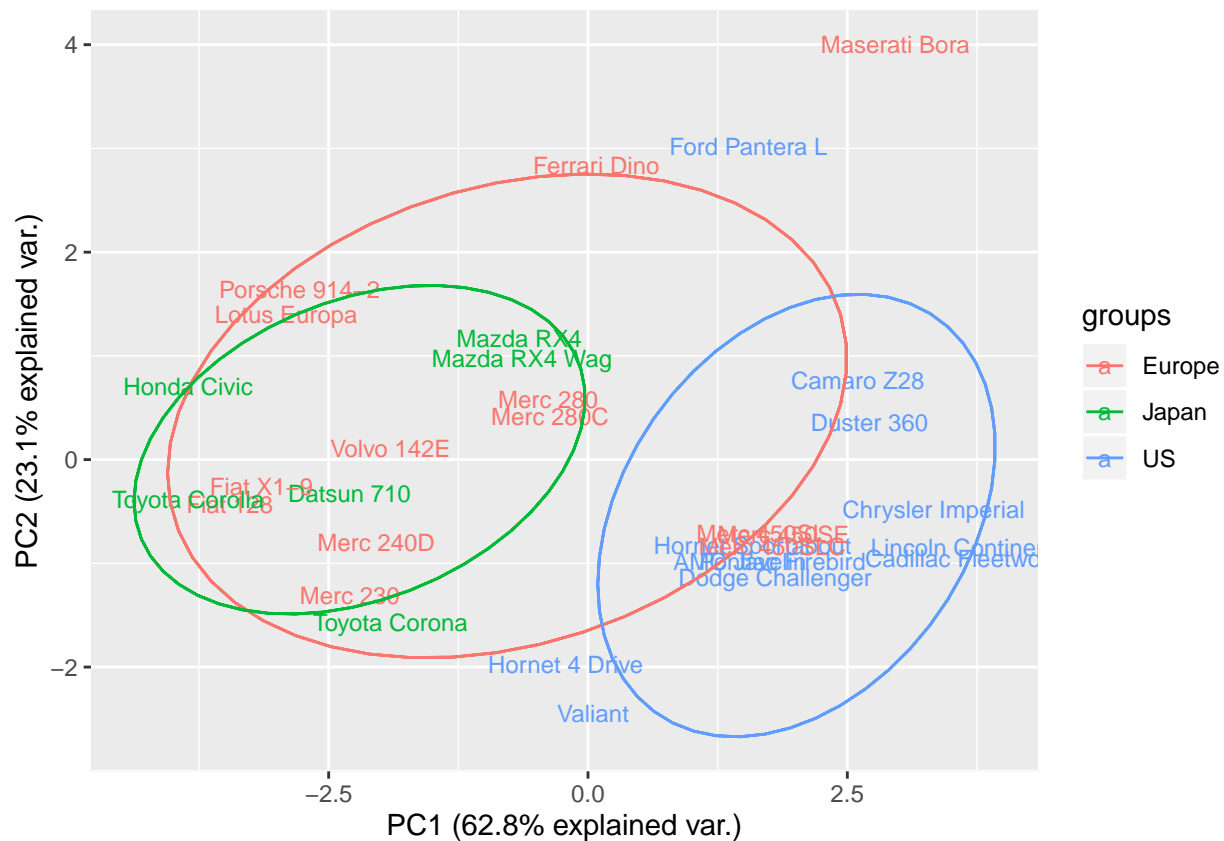
scale the samples (obs.scale) and the variables (var.scale)

```
ggbiplot(mtcars.pca,
  ellipse=TRUE,
  obs.scale = 1,
  var.scale = 1,
  labels=rownames(mtcars),
  groups=mtcars.country)
```



var.axes=FALSE remove the arrows altogether

```
ggbiplot(mtcars.pca,
  ellipse=TRUE,
  obs.scale = 1,
  var.scale = 1,
  var.axes=FALSE,
  labels=rownames(mtcars),
  groups=mtcars.country)
```



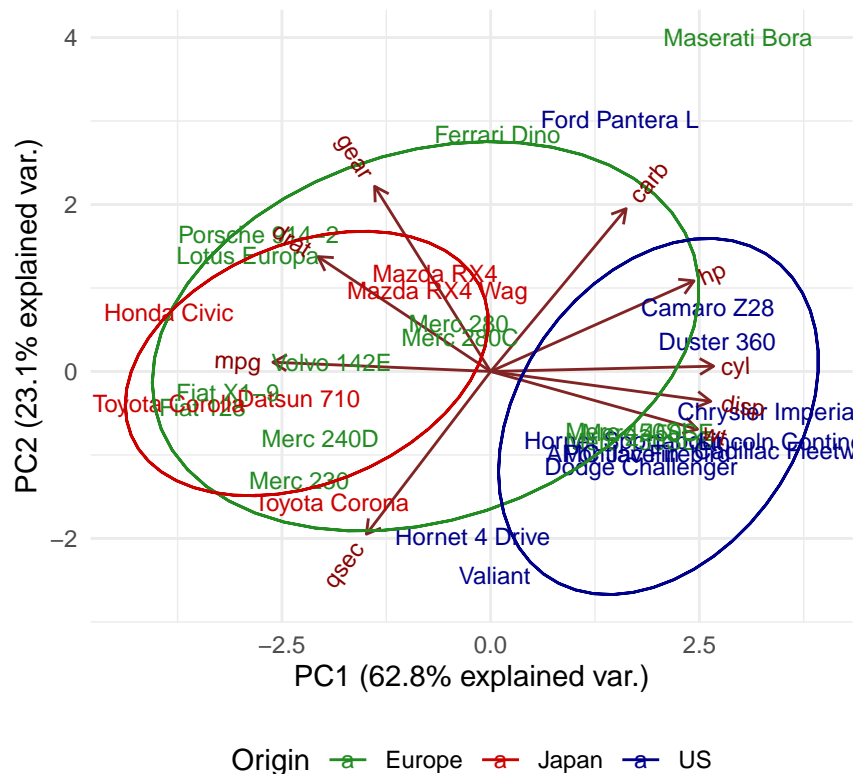
## Customize ggbiplot

`scale_colour_manual()` Specify the colours for the groups `ggtitle()` Add a title `minimal()` Specify the minimal() theme `theme()` Move the legend position

```
ggbiplot(mtcars.pca,
  ellipse=TRUE,
  obs.scale = 1,
  var.scale = 1,
  labels=rownames(mtcars),
  groups=mtcars.country) +
  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

add a new sample to mtcars, creating mtcarsplus, then repeat the analysis

```
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")

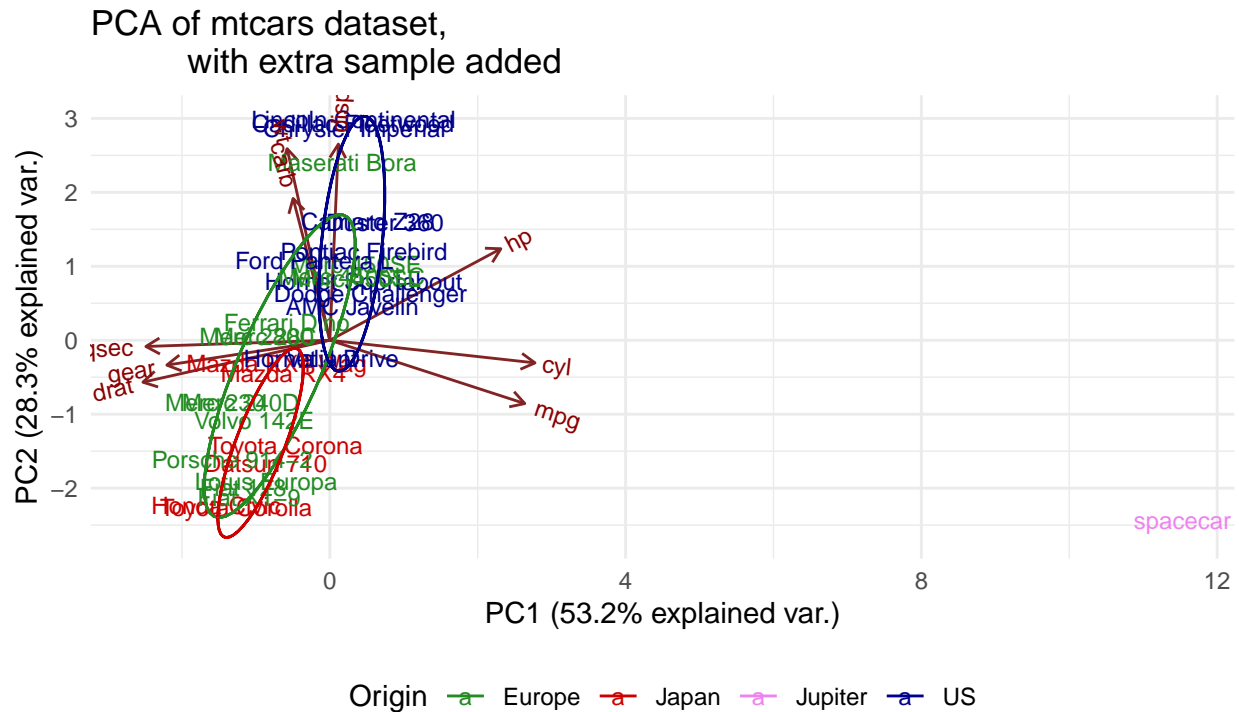
mtcarsplus.pca <- prcomp(mtcarsplus[,c(1:7,10,11)],
                        center = TRUE,
                        scale. = TRUE)

ggbiplot(mtcarsplus.pca,
         obs.scale = 1,
         var.scale = 1,
         ellipse = TRUE,
         circle = FALSE,
         var.axes=TRUE,
         labels=c(rownames(mtcars), "spacecar"),
         groups=mtcars.countryplus)+
  scale_colour_manual(name="Origin",
                     values= c("forest green",
```

```

      "red3",
      "violet",
      "dark blue")))+
ggtitle("PCA of mtcars dataset,
        with extra sample added")+
theme_minimal()+
theme(legend.position = "bottom")

```



## Project a new sample onto the original PCA

1. scale the values for spacecar in relation to the PCA's center (mtcars.pca\$center)
2. apply the rotation of the PCA matrix to the spacecar sample
3. rbind() the projected values for spacecar to the rest of the pca\$x matrix

```

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 = FALSE,
var.axes=TRUE,
labels=c(rownames(mtcars),
         "spacecar"),
groups=mtcars.countryplus)+
scale_colour_manual(name="Origin",
                    values= c("forest green", "red3", "violet", "dark blue"))+
ggtitle("PCA of mtcars dataset,
        with extra sample projected")+
theme_minimal()+
theme(legend.position = "bottom")

```

