A2_Q2

Chia-Ying Chao

15/02/2023

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.4.0      v purrr   0.3.4
## v tibble  3.1.4      v dplyr   1.0.10
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(ggpubr )
```

```
## Warning: package 'ggpubr' was built under R version 4.1.3
```

```
library(FactoMineR)
```

```
## Warning: package 'FactoMineR' was built under R version 4.1.3
```

```
breast_cancer<-read_csv("breast_cancer_data.csv",col_names=FALSE, col_types = cols())

names(breast_cancer)<-c("RecEv","AgeGrp","Meno","Size","InvNodes",
"NodeCaps","DegMal","Side","Quad","Irrad")

### Remove missing obs
breast_cancer<-breast_cancer %>% filter(Quad!="?",NodeCaps!="?")
breast_cancer<-breast_cancer %>% mutate(AgeGrp=paste("AG",AgeGrp,sep=""),
InvNodes=paste("IN",InvNodes,sep=""))
head(breast_cancer) %>% kable(.) %>% kable_styling()
```

RecEv	AgeGrp	Meno	Size	InvNodes	NodeCaps	DegMal	Side	Quad	Irrad
no-recurrence-events	AG30-39	premeno	30-34	IN0-2	no	3	left	left_low	no
no-recurrence-events	AG40-49	premeno	20-24	IN0-2	no	2	right	right_up	no
no-recurrence-events	AG40-49	premeno	20-24	IN0-2	no	2	left	left_low	no
no-recurrence-events	AG60-69	ge40	15-19	IN0-2	no	2	right	left_up	no
no-recurrence-events	AG40-49	premeno	0-4	IN0-2	no	2	right	right_low	no
no-recurrence-events	AG60-69	ge40	15-19	IN0-2	no	2	left	left_low	no

```
part_one<-breast_cancer %>% dplyr::select(Quad,DegMal)
head(part_one)
```

```
## # A tibble: 6 x 2
##   Quad      DegMal
##   <chr>      <dbl>
## 1 left_low       3
## 2 right_up       2
## 3 left_low       2
## 4 left_up        2
## 5 right_low      2
## 6 left_low       2
```

```
#i) two-way contingency table
contingency_table <- table(part_one$Quad, part_one$DegMal)
contingency_table
```

```
##
##              1  2  3
##   central    7 10  4
##   left_low  25 50 31
##   left_up   20 43 31
##   right_low  7 10  6
##   right_up   7 16 10
```

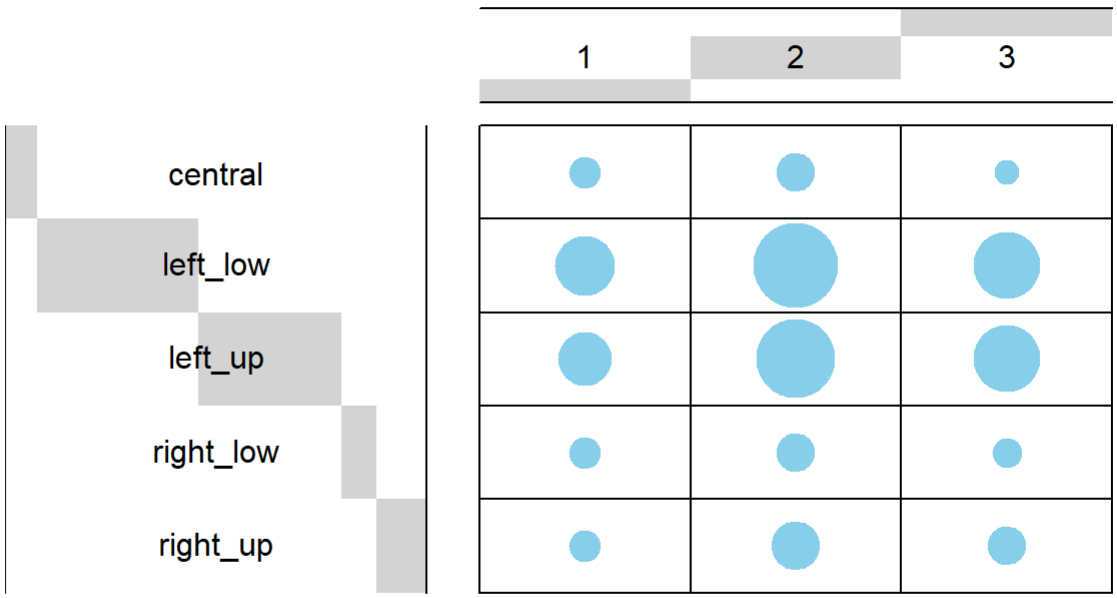
```
#ii) Balloon plots

library("gplots")
```

```
## Warning: package 'gplots' was built under R version 4.1.3
```

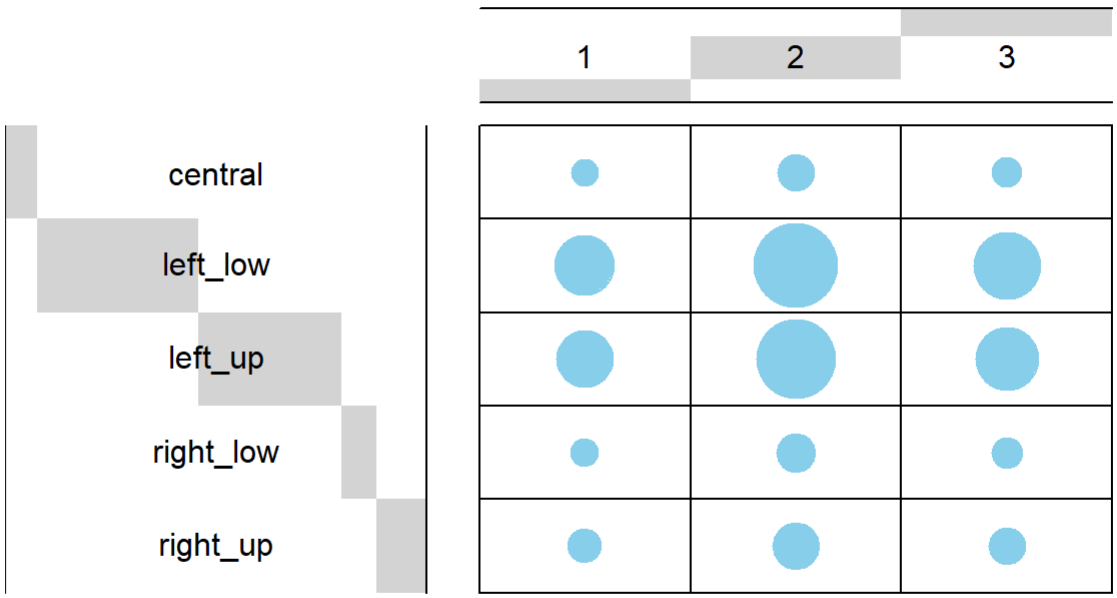
```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
balloonplot(t(contingency_table), main="Observed cell proportions", xlab="", ylab="", label = FA
LSE, show.margins=FALSE)
```




```
chisq <- chisq.test(contingency_table)

expected <- chisq$expected
exp2<-as.table(expected)
balloonplot(t(exp2), main="Expected cell proportions", xlab="", ylab="", label = FALSE, show.mar
gins=FALSE)
```



Since there is no huge difference between the size of the dots of the balloon plots of observed cell proportions and expected proportions, this tells us that the relationship between the quadrant and degree of malignancy is most likely independent.

```
#iii)
#a) Table of eigenvalues
table_eigen<- CA(contingency_table,ncp=2, graph=FALSE)
table_eigen$eig
```

```
##         eigenvalue percentage of variance cumulative percentage of variance
## dim 1 0.0099029594              93.612145                           93.61214
## dim 2 0.0006757528               6.387855                          100.00000
```

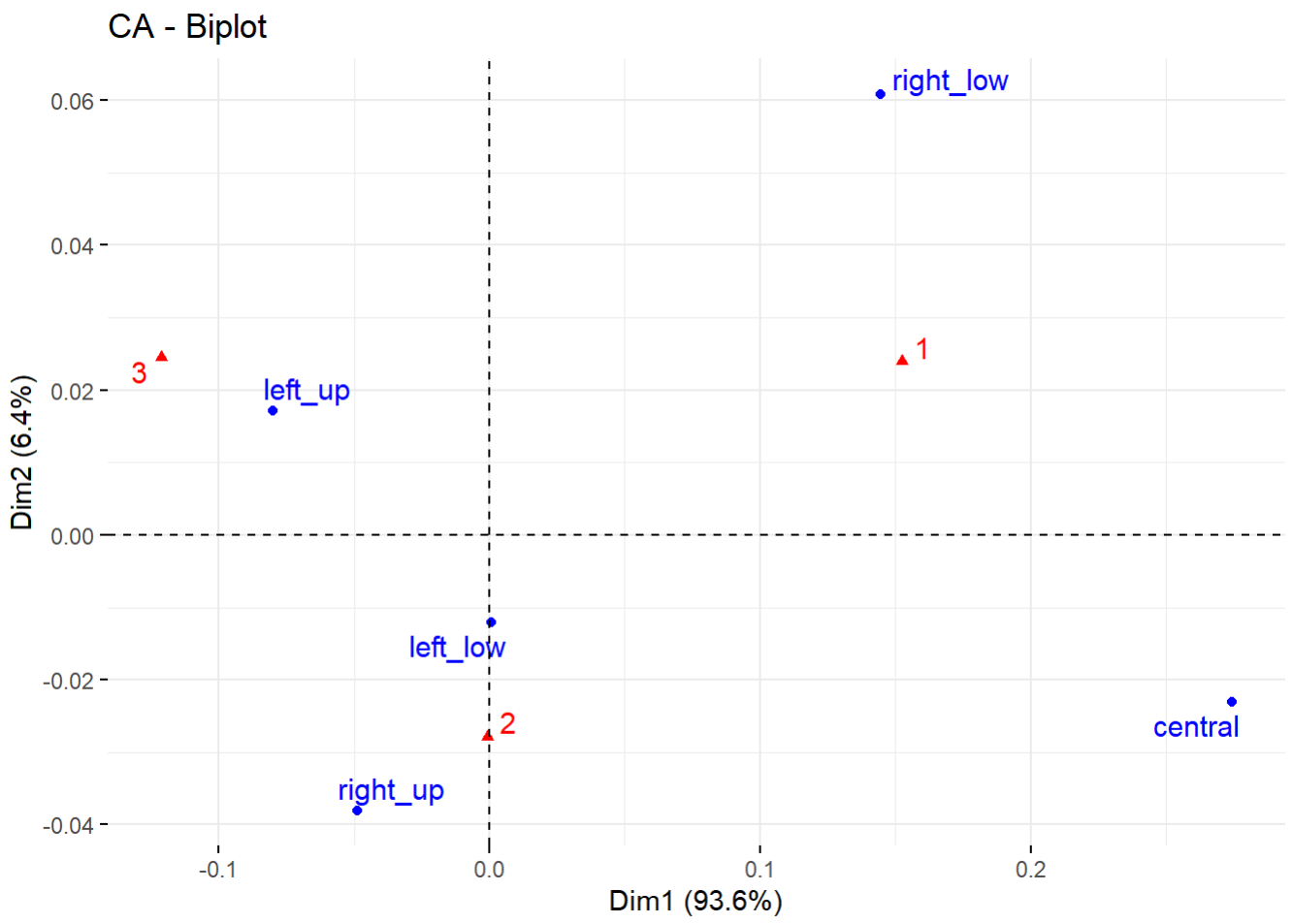
It is sufficient to just have the first dimension to analyze the data since it explains 93.6% of the variability in the table.

```
#b) Create Factor Map

library(factoextra)
```

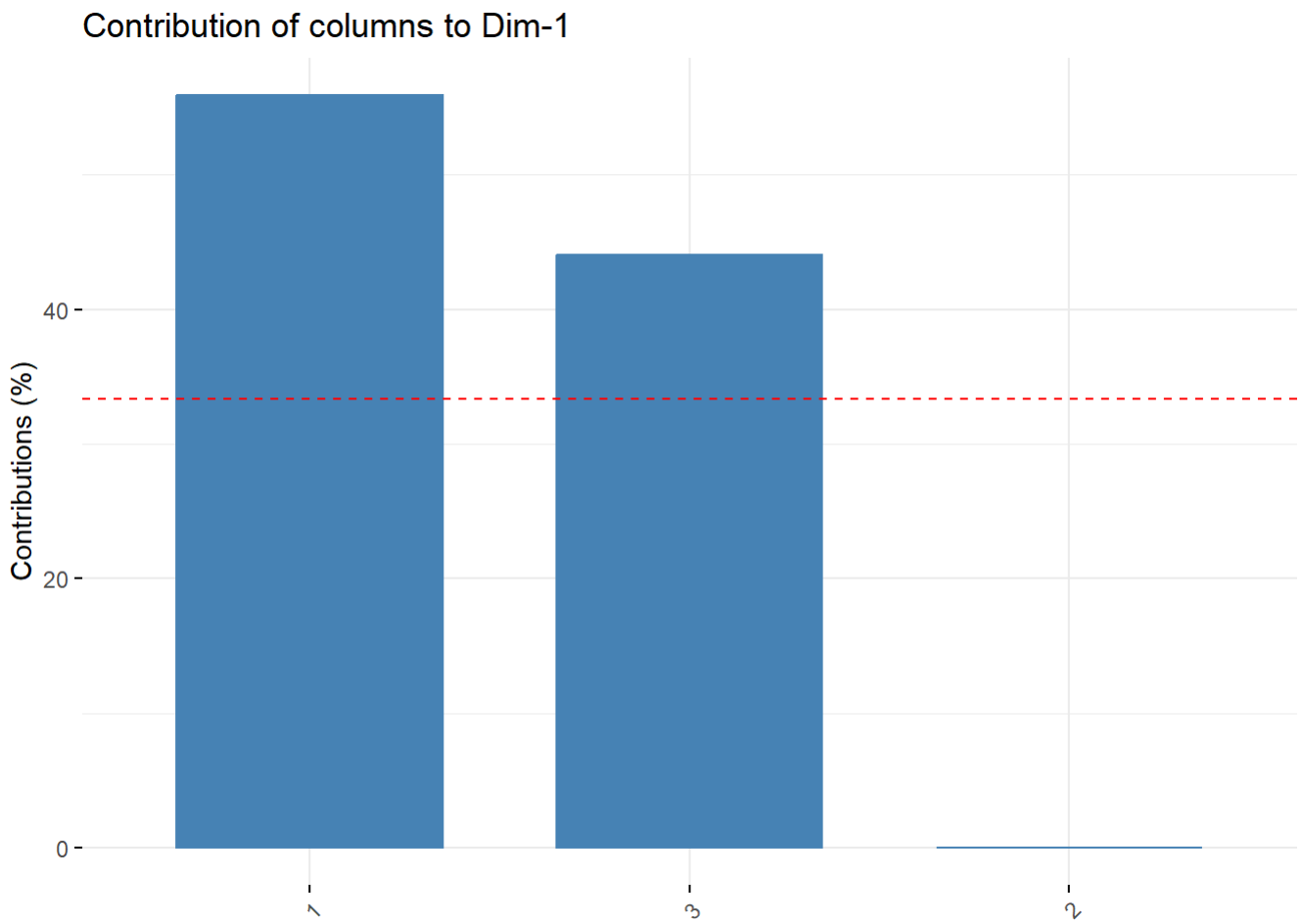
```
## Warning: package 'factoextra' was built under R version 4.1.3
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

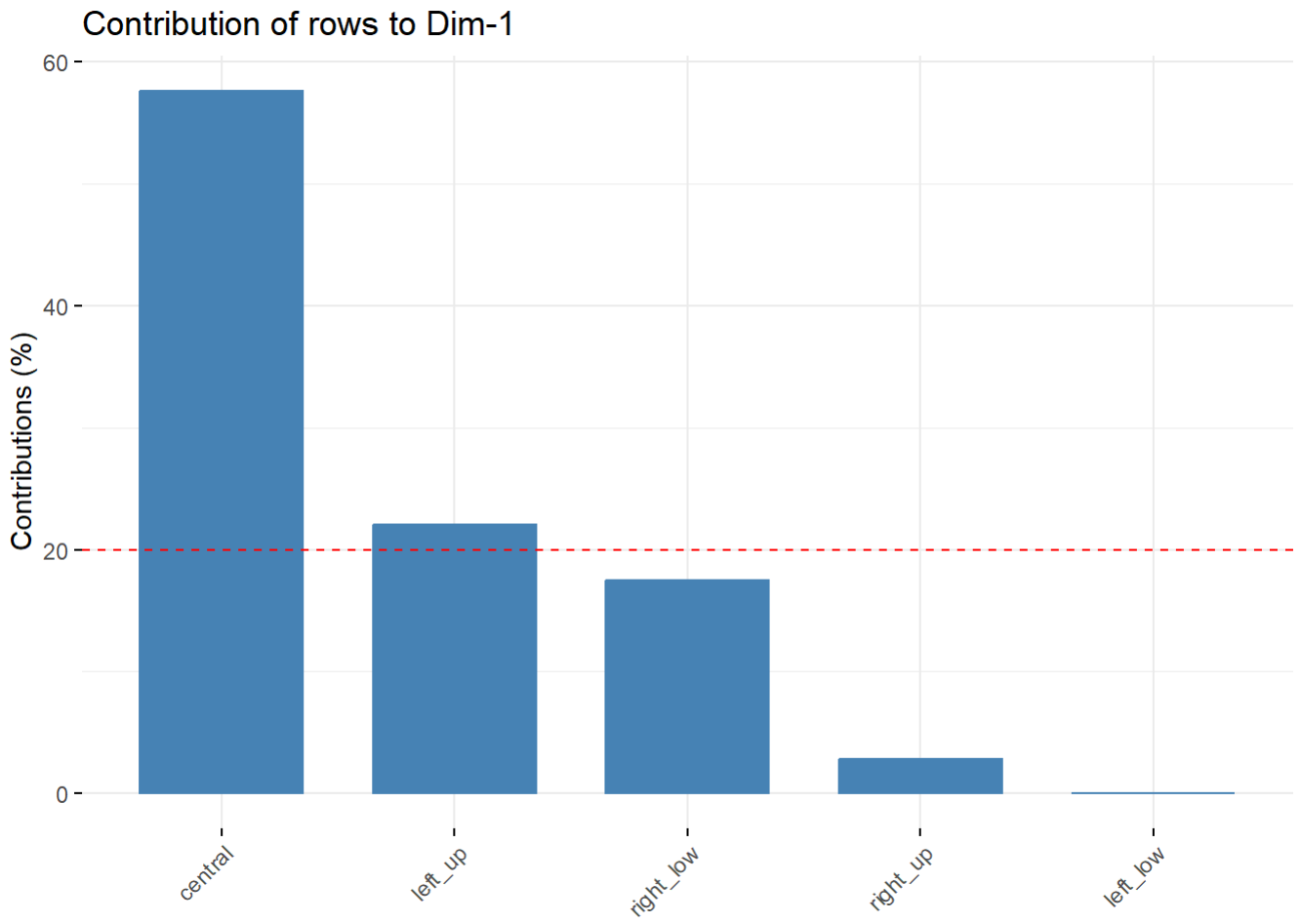
```
fviz_ca_biplot(table_eigen, repel = TRUE)
```



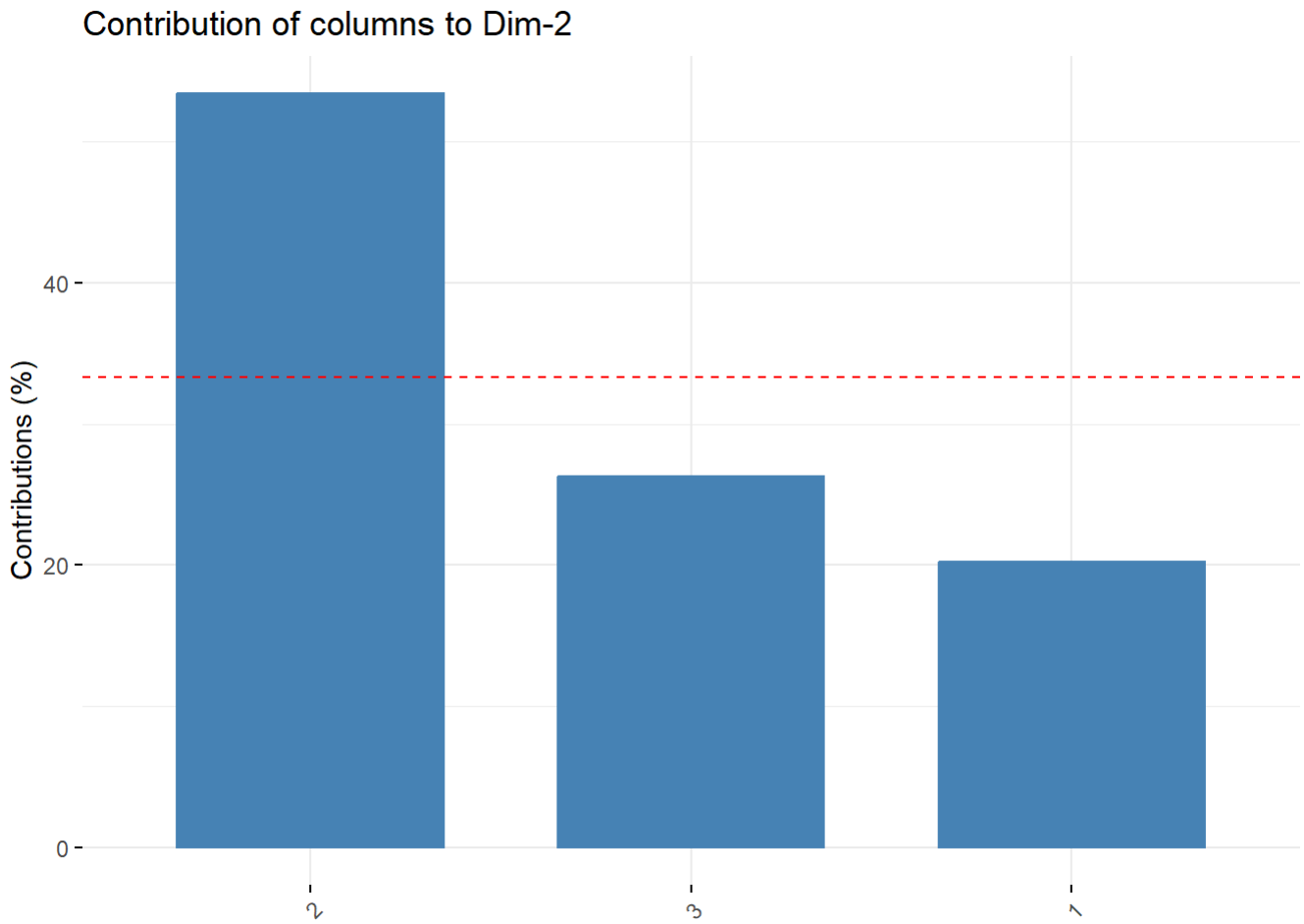

```
fviz_contrib(table_eigen, choice="col")
```



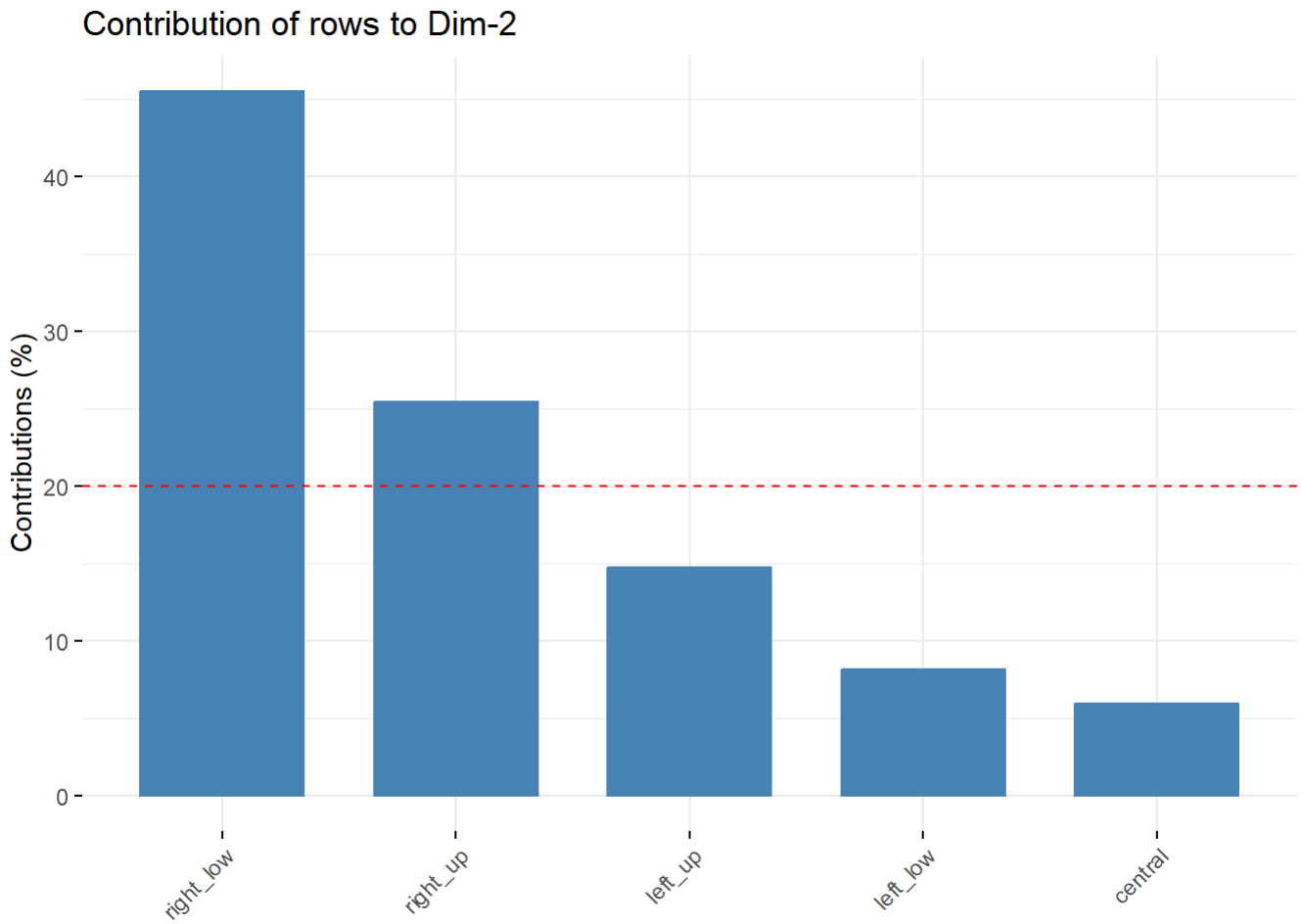

```
fviz_contrib(table_eigen, choice="row")
```




```
fviz_contrib(table_eigen, choice="col", axes=2)
```




```
fviz_contrib(table_eigen, choice="row", axes=2)
```



Columns are presented by the red triangles and rows are presented by the blue dots in the factor map. We can see that left_low and right_up and degree of malignancy 2 are associated to one and another. Same concept applies to degree of malignancy 3 and left_up, and right_low and degree of malignancy 1.

By looking at the contribution of rows to Dim-1, we can observe that central quadrant of the tumor and the left up quadrant of the tumor contributes have larger value which means that they contribute the most to the first dimension. As for the contribution of the column profile to dimension 1, first and third degree of malignancy contribute the most to first dimension.

However, in dimension 2, only the second degree of malignancy contribute the most in the contribution of column, and right low and right up quadrant of the tumors have the most contribution instead in the contribution of rows plot.