Examinacion de datos 1

Melisa Garcia P y Jose Luis Osorio Q

May 2, 2017

Exploracion de informacion de los cuatro datasets

```
#Importacion de paquetes necesarios
#install.packages("foreign")
#install.packages("dplyr")
#install.packages("reshape2")
#install.packages("plotly")

#Librerias necsarias
library(foreign)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(plotly)
```

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

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

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

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library(reshape2)
```

CAPA DE APLICACION

CAPA DE APLICACION BENIGNA

```
matrizCAB<- read.csv("matriz_app_benigno.csv", sep=";", comment.char =
"#")
matrizCAB<- na.omit(matrizCAB)
head(matrizCAB)
```

```
##    URL LongitudURL CaracteristicasExtrannas    charset Servidor
## 1 B0_1          56                        8 iso-8859-1   Apache
## 2 B0_3          63                       12      UTF-8   Apache
## 3 B0_4          54                        9      UTF-8   Apache
## 4 B0_7          63                       10      UTF-8   Apache
## 5 B0_8          60                       11      utf-8    nginx
## 6 B0_9          35                        7 iso-8859-1   Apache
##                           CacheControl contentLength Pais Estado
## 1                                 None           257   US     CA
## 2                                 None          None None   None
## 3                                 None           193   US     WA
## 4 no-cache, must-revalidate, max-age=0         18235   US     WA
## 5                            no-cache          None   US     CA
## 6                                 None           330 None   None
##      Registro.date      Update.date             Dominio
## 1 05/11/1999 00:00 11/08/2016 00:00 startedbyamouse.com
## 2             None             None shortsweetpoems.com
## 3 27/06/2015 00:00 18/04/2016 00:00         findoha.com
## 4 10/10/2006 00:00 07/03/2017 00:00    lyricsanimal.com
## 5 24/09/2006 00:00 26/08/2015 00:00          scribd.com
## 6 10/11/2000 00:00 19/02/2016 00:00      beavertails.ca
```

```
#Se cambian los nombres de las columnas
names(matrizCAB) <- c("URL",
"URL_LENGTH",
"NUMBER_SPECIAL_CHARACTERS",
"CHARSET",
"SERVER",
"CACHE_CONTROL",
"CONTENT_LENGTH",
"WHOIS_COUNTRY",
"WHOIS_STATEPROV",
"WHOIS_REGDATE",
"UPDATE_DATE",
"WHITIN_DOMAIN")

#Se transforman a mayusculas, todas las letras de la matriz. Esto con el
fin, que tenga el mismo formato para estudiarlas. Si se tiene el mismo
nombre en mayusculas y minusculas, R lo toma como nombres diferentes.
matrizCAB <- mutate_each(matrizCAB, funs(toupper))
```

```r
#Se eliminan los campos "NA"
matrizCAB<- na.omit(matrizCAB)

head(matrizCAB)
```

```
##    URL URL_LENGTH NUMBER_SPECIAL_CHARACTERS    CHARSET SERVER
## 1 B0_1         56                          8 ISO-8859-1 APACHE
## 2 B0_3         63                         12      UTF-8 APACHE
## 3 B0_4         54                          9      UTF-8 APACHE
## 4 B0_7         63                         10      UTF-8 APACHE
## 5 B0_8         60                         11      UTF-8  NGINX
## 6 B0_9         35                          7 ISO-8859-1 APACHE
##                          CACHE_CONTROL CONTENT_LENGTH WHOIS_COUNTRY
## 1                                 NONE            257            US
## 2                                 NONE           NONE          NONE
## 3                                 NONE            193            US
## 4 NO-CACHE, MUST-REVALIDATE, MAX-AGE=0          18235            US
## 5                             NO-CACHE           NONE            US
## 6                                 NONE            330          NONE
##   WHOIS_STATEPROV    WHOIS_REGDATE      UPDATE_DATE
WHITIN_DOMAIN
## 1              CA 05/11/1999 00:00 11/08/2016 00:00
STARTEDBYAMOUSE.COM
## 2            NONE             NONE             NONE
SHORTSWEETPOEMS.COM
## 3              WA 27/06/2015 00:00 18/04/2016 00:00
FINDOHA.COM
## 4              WA 10/10/2006 00:00 07/03/2017 00:00
LYRICSANIMAL.COM
## 5              CA 24/09/2006 00:00 26/08/2015 00:00
SCRIBD.COM
## 6            NONE 10/11/2000 00:00 19/02/2016 00:00
BEAVERTAILS.CA
```

```r
#URL LENGTH BENIGNA

A1_B = mean(as.numeric(matrizCAB$URL_LENGTH))
A1_B
```

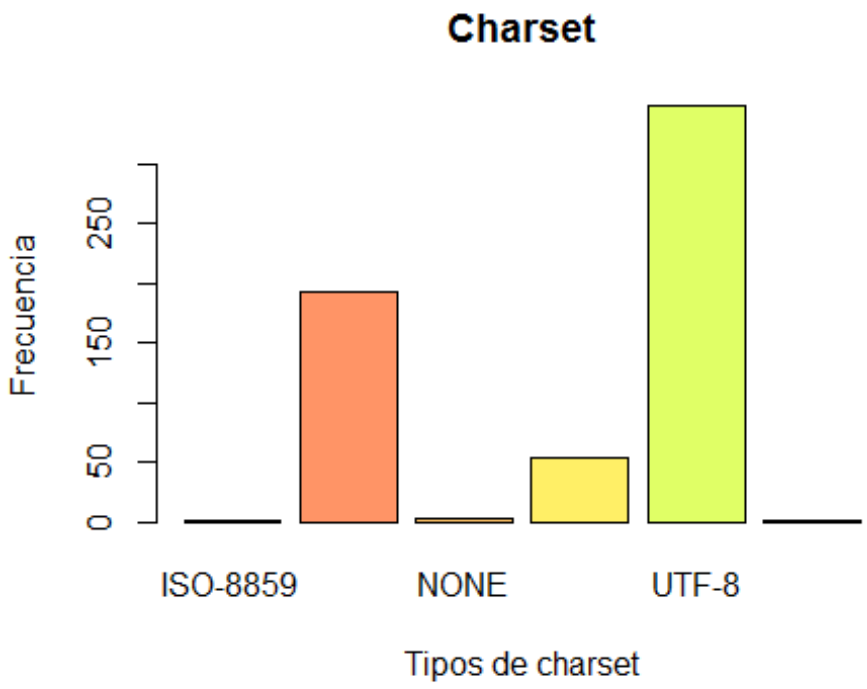
```
## [1] 56.29167
```

```r
#NUMBER SPECIAL CHARACTERS BENIGNA

A2_B = mean(as.numeric(matrizCAB$NUMBER_SPECIAL_CHARACTERS))
A2_B
```

```
## [1] 10.81
```

Frecuencia por cada valor caracteristico de los datasets de capa de aplicacion benigna:

```r
#CHARSET BENIGNA

A3_B<-table(matrizCAB$CHARSET)
head(A3_B)

##
##     ISO-8859   ISO-8859-1         NONE     US-ASCII        UTF-8
##            1          193            3           54          348
## WINDOWS-1252
##            1

barplot(table(matrizCAB$CHARSET),
        main="Charset",
        xlab="Tipos de charset",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```



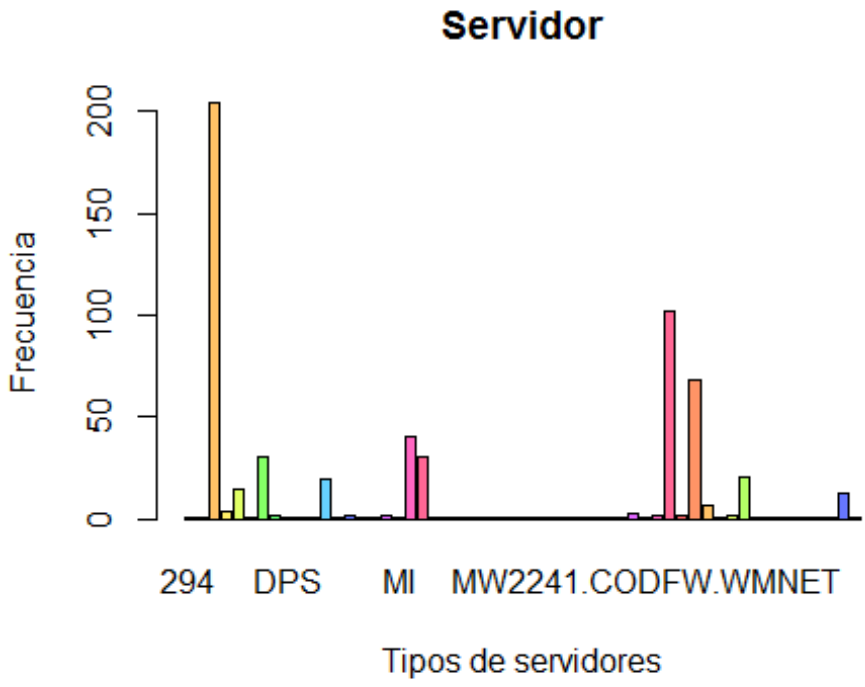
Se puede evidenciar que con los datos que se suministran, las URL's benignas suelen utilizar el charset UTF-8, mayor cantidad de veces.

```r
#SERVER BENIGNA

#SEPARACION DE SERVER (SOLO EL NOMBRE)
matrizCAB$SERVER <- sapply(strsplit(matrizCAB$SERVER,"/"), `[`, 1)
matrizCAB$OTHER_SERVER_ATTRIBUTES <-
sapply(strsplit(matrizCAB$SERVER,"/"), `[`, 2)
#matrizCAB <- matrizCAB[,-5]
```

```
A4_B<-table(matrizCAB$SERVER)
head(A4_B)
```

```
##
##           294      AMAZONS3        APACHE APACHE-COYOTE           ATS
##             1             1           204             4            14
##       BARISTA
##             1
```

```
barplot(table(matrizCAB$SERVER),
        main="Servidor",
        xlab="Tipos de servidores",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```



Por otro lado, los servidores mas utilizados son APACHE(203), NGINX (102), MICROSOFT-HTTPAPI (40), MICROSOFT-IIS (30), CLOUDFLARE-NGINX(30), que pertenece a la familia NGINX.

```
#CACHE CONTROL BENIGNA

A5_B<-table(matrizCAB$CACHE_CONTROL)
head(A5_B)
```

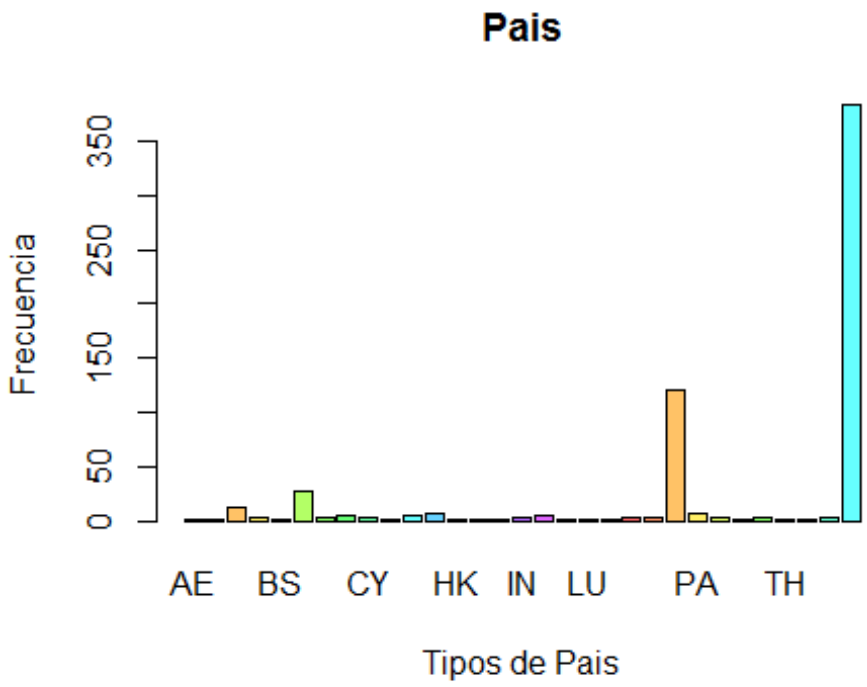
```
##
##
```

```
##                                                         1
##                                                 MAX-AGE=0
##                                                         1
##                          MAX-AGE=0, NO-CACHE, NO-STORE
##                                                         6
## MAX-AGE=0, NO-CACHE, NO-STORE, NO-TRANSFORM, PRIVATE
##                                                         1
##                   MAX-AGE=0, PRIVATE, MUST-REVALIDATE
##                                                         8
## MAX-AGE=0, PRIVATE, MUST-REVALIDATE, MAX-AGE=2592000
##                                                         1

#Esta no se le aplica split, porque se desea saber como vienen en la
forma original, para evidenciar las formas comunes de los cache-control.
```

La informacion presentada sobre los chache - control mas comunes en las URL's benignas de los datasets suministrados, son:

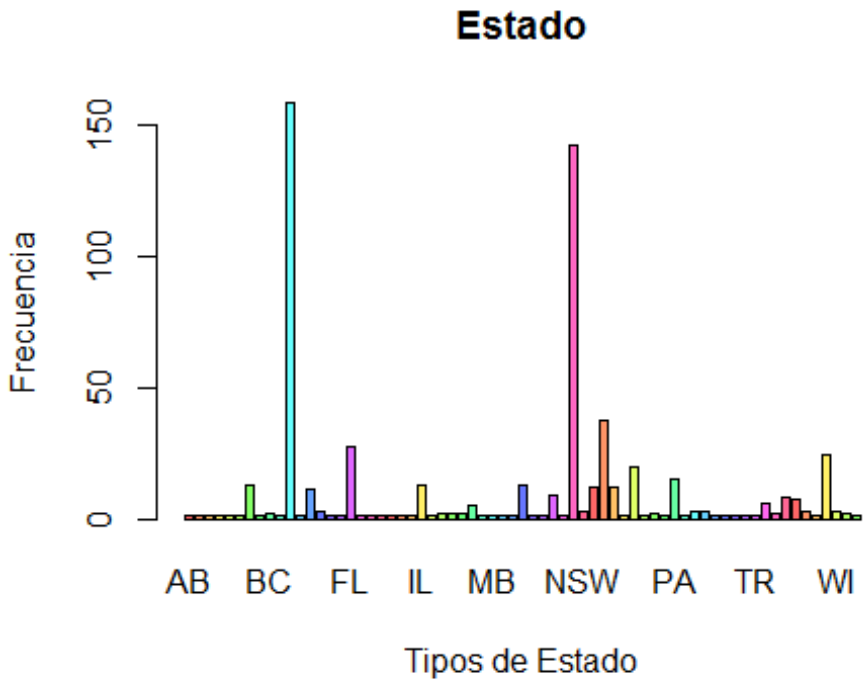
NO-CACHE (54) PRIVATE, S-MAXAGE=0, MAX-AGE=0, MUST-REVALIDATE (27) PRIVATE, NO-CACHE, NO-STORE, MUST-REVALIDATE (25) PRIVATE (23)

```
#WHOIS COUNTRY BENIGNA

A6_B<-table(matrizCAB$WHOIS_COUNTRY)
head(A6_B)

##
## AE AT AU BE BS CA
##  1  1 12  2  1 27

barplot(table(matrizCAB$WHOIS_COUNTRY),
        main="Pais",
        xlab="Tipos de Pais",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```



Para los paises, el mas destacado en donde se registraron los sitios web fue US, Estados Unidos, con 382, seguido de CA, Canada, con 27 resultados. Por ultimo, es importante resaltar que hay 120 URL's que no registran el pais.

```
#WHOIS STATEPROV BENIGNA

A7_B<-table(matrizCAB$WHOIS_STATEPROV)
head(A7_B)

##
##              AB              AK              AL ANDHRA PRADESH
ANTWERP
##               1               1               1               1
1
##              AT
##               1

barplot(table(matrizCAB$WHOIS_STATEPROV),
        main="Estado",
        xlab="Tipos de Estado",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```



Para los estados, el mas destacado en donde se registraron los sitios web fue CA con 158, seguido de NY, New York, con 37 resultados. Por ultimo, es importante resaltar que hay 142 URL's que no registran el estados

```
#WHOIS REGDATE  BENIGNA

A8_B<-table(matrizCAB$WHOIS_REGDATE)
head(A8_B)

##
## 01/02/1994 00:00 01/03/1994 00:00 01/03/2008 00:00 01/04/2008 22:47
##                1                1                1                1
## 01/05/1996 00:00 01/05/2009 00:00
##                1                1

#UPDATE DATE BENIGNA

A9_B<-table(matrizCAB$UPDATE_DATE)
head(A9_B)

##
## 01/01/2017 00:00 01/02/2013 00:00 01/02/2017 00:00 01/02/2017 05:28
##                1                1                1                1
## 01/03/2017 00:00 01/03/2017 19:31
##                1                1

#WHITIN DOMAIN BENIGNA
```

```
A10_B<-table(matrizCAB$WHITIN_DOMAIN)
head(A10_B)

##
## 123PEOPLE.COM       192.COM 247SPORTS.COM   4SHARED.COM  50STATES.COM
##             1             1             1             1             1
##   8ASIANS.COM
##             1
```

CAPA DE APLICACION MALIGNA

```
matrizCAM <- matrizCAB<- read.csv("matriz_app_maligno.csv", sep=";",
comment.char = "#")
matrizCAM<- na.omit(matrizCAM)
head(matrizCAM)

##     URL A1 A2         A4                    A5
## 1  M0_9 43 10 ISO-8859-1             DOSarrest
## 2 M0_15 26  8      utf-8                 nginx
## 3 M0_16 37  9 ISO-8859-1                 nginx
## 4 M0_21 28  9      UTF-8 Apache/2.2.15 (CentOS)
## 5 M0_39 40 10      utf-8                Apache
## 6 M0_40 42 10      utf-8          nginx/1.10.1
##                                   A6   A7   A8         A9
## 1                           no-cache None   us         WA
## 2                               None None None       None
## 3                               None  162   CN beijingshi
## 4                               None None   RU     MOSCOW
## 5                public, max-age=300 3016 None       None
## 6 no-cache, pre-check=0, post-check=0 None None       None
##                A10              A11               A12
## 1 23/06/2003 00:00 02/01/2017 00:00 realinnovation.com
## 2 09/03/2000 17:50             None         antalya.ru
## 3 09/02/2009 00:00 27/04/2017 00:00         img001.com
## 4 24/10/2007 00:00 22/10/2016 00:00         traff1.com
## 5             None             None           alice.it
## 6 07/04/2002 20:00             None          propan.ru

#Se convierten a mayus todos los valores de las columnas
matrizCAM <- mutate_each(matrizCAM, funs(toupper))

#Se cambian los nombres de las columnas
names(matrizCAM) <- c("URL",
"URL_LENGTH",
"NUMBER_SPECIAL_CHARACTERS",
"CHARSET",
"SERVER",
"CACHE_CONTROL",
"CONTENT_LENGTH",
"WHOIS_COUNTRY",
```

```
"WHOIS_STATEPROV",
"WHOIS_REGDATE",
"UPDATE_DATE",
"WHITIN_DOMAIN")

#Se quitan los valores "NA"
matrizCAM<- na.omit(matrizCAM)
head(matrizCAM)
```

```
##     URL URL_LENGTH NUMBER_SPECIAL_CHARACTERS    CHARSET
## 1  M0_9         43                        10 ISO-8859-1
## 2 M0_15         26                         8      UTF-8
## 3 M0_16         37                         9 ISO-8859-1
## 4 M0_21         28                         9      UTF-8
## 5 M0_39         40                        10      UTF-8
## 6 M0_40         42                        10      UTF-8
##                   SERVER                        CACHE_CONTROL
## 1              DOSARREST                              NO-CACHE
## 2                  NGINX                                  NONE
## 3                  NGINX                                  NONE
## 4 APACHE/2.2.15 (CENTOS)                                  NONE
## 5                 APACHE                 PUBLIC, MAX-AGE=300
## 6          NGINX/1.10.1 NO-CACHE, PRE-CHECK=0, POST-CHECK=0
##   CONTENT_LENGTH WHOIS_COUNTRY WHOIS_STATEPROV    WHOIS_REGDATE
## 1           NONE            US              WA 23/06/2003 00:00
## 2           NONE          NONE            NONE 09/03/2000 17:50
## 3            162            CN      BEIJINGSHI 09/02/2009 00:00
## 4           NONE            RU          MOSCOW 24/10/2007 00:00
## 5           3016          NONE            NONE             NONE
## 6           NONE          NONE            NONE 07/04/2002 20:00
##        UPDATE_DATE      WHITIN_DOMAIN
## 1 02/01/2017 00:00 REALINNOVATION.COM
## 2             NONE         ANTALYA.RU
## 3 27/04/2017 00:00         IMG001.COM
## 4 22/10/2016 00:00         TRAFF1.COM
## 5             NONE           ALICE.IT
## 6             NONE          PROPAN.RU
```

```
#URL LENGTH MALIGNA

A1_M = mean(as.numeric(matrizCAM$URL_LENGTH))
A1_M
```

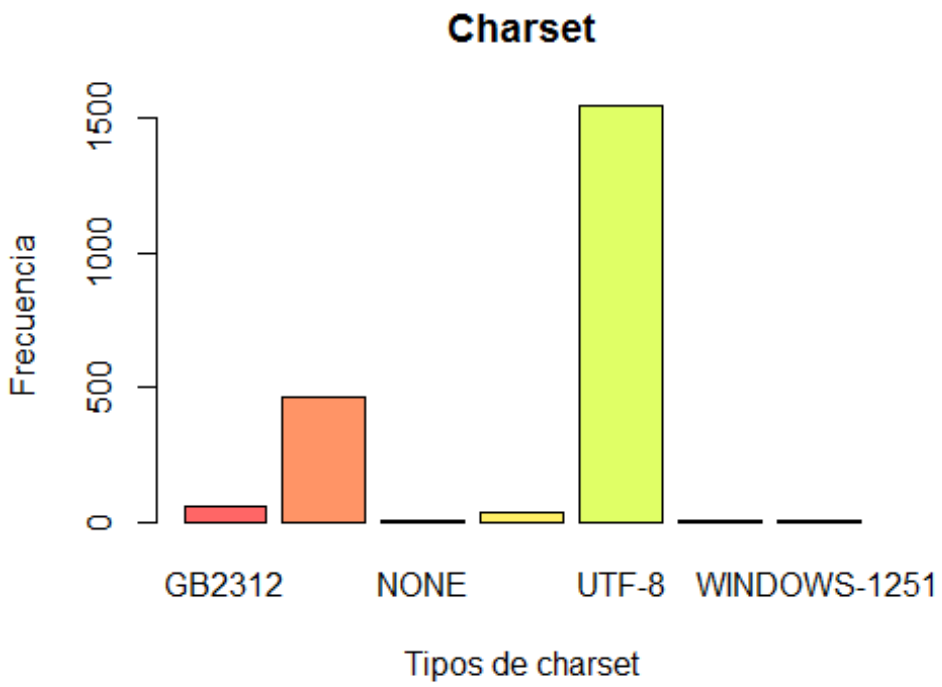
```
## [1] 85.45571
```

```
#NUMBER SPECIAL CHARACTERS MALIGNA

A2_M = mean(as.numeric(matrizCAM$NUMBER_SPECIAL_CHARACTERS))
A2_M
```

```
## [1] 17.20381
```

Frecuencia por cada valor caracteristico de los datasets de capa de aplicacion maligna:

```
#CHARSET MALIGNA

A3_M<-table(matrizCAM$CHARSET)
head(A3_M)

##
##       GB2312   ISO-8859-1         NONE     US-ASCII        UTF-8
##           56          461            1           33         1547
## WINDOWS-1250
##            1

barplot(table(matrizCAM$CHARSET),
        main="Charset",
        xlab="Tipos de charset",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```



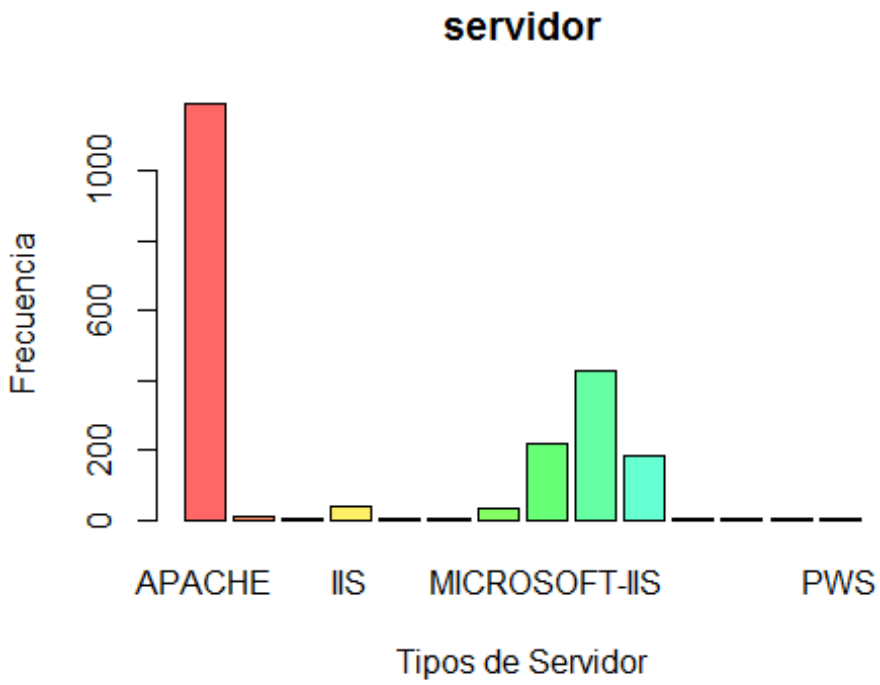
Pala los datos malignos, el charset mas utilizado, como en los datos benignos, es UTF-8 (1547), seguido por ISO-8859-1 (461)

```
#SERVER MALIGNA

matrizCAM$SERVER <- sapply(strsplit(matrizCAM$SERVER,"/"), `[`, 1)
```

```r
A4_M<-table(matrizCAM$SERVER)
head(A4_M)
```

```
##
##           APACHE CLOUDFLARE-NGINX        DOSARREST              IIS
##             1192                6                1               35
##         LIGHTTPD MARRAKESH 1.12.2
##                1                1
```

```r
barplot(table(matrizCAM$SERVER),
        main="servidor",
        xlab="Tipos de Servidor",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```

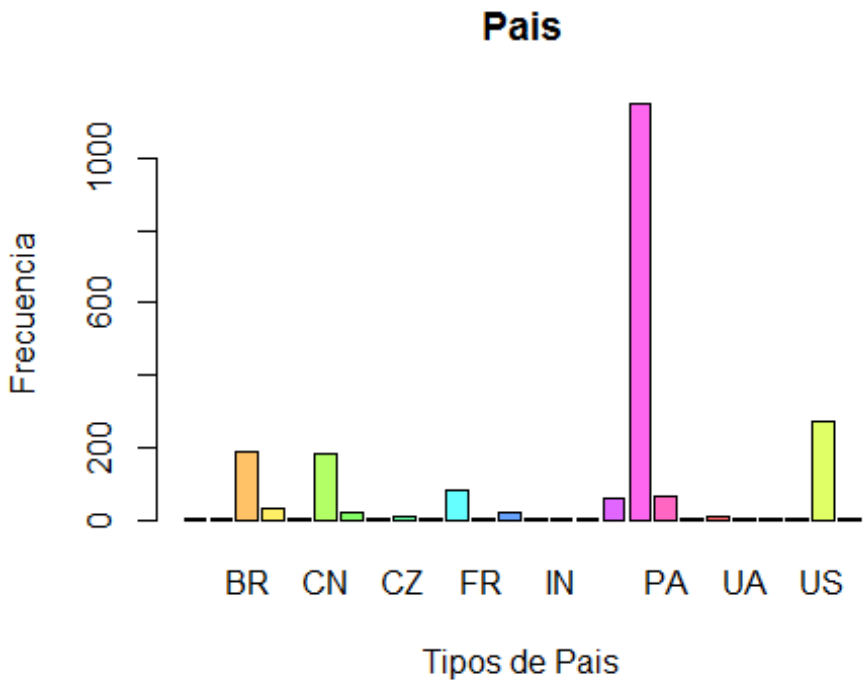


Los servidores mas comunes, son APACHE (1192) NGINX (425) MICROSOFT-IIS (220)

```r
#CACHE CONTROL MALIGNA

A5_M<-table(matrizCAM$CACHE_CONTROL)
head(A5_M)
```

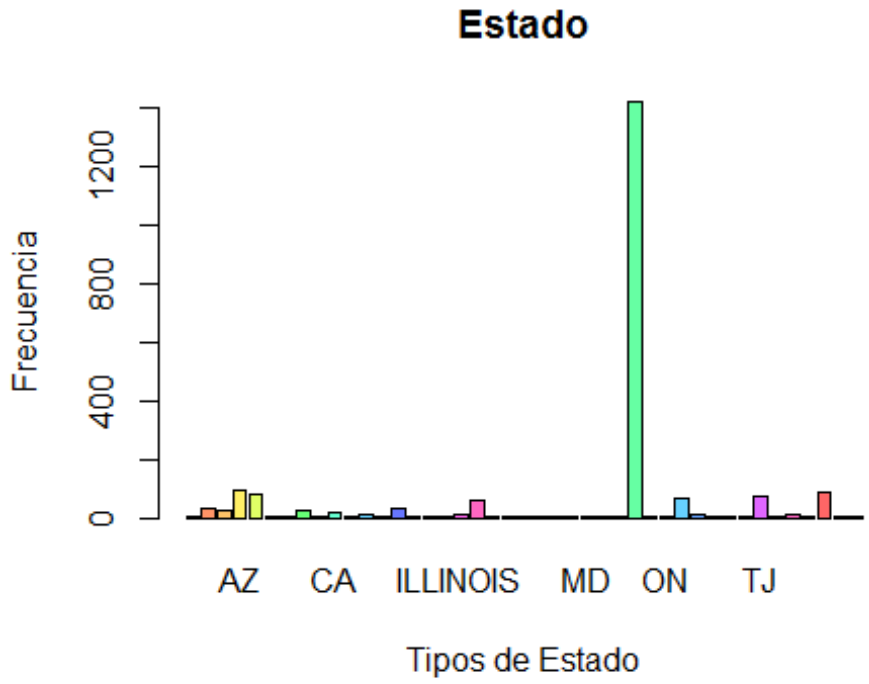
```
##
##
##                                  1
##                       MAX-AGE=172800
##                                 22
```

```
## MAX-AGE=3, MUST-REVALIDATE, MAX-AGE=0
##                                     1
## MAX-AGE=30, PRIVATE, PROXY-REVALIDATE
##                                     1
##                               MAX-AGE=300
##                                     1
##                              MAX-AGE=3600
##                                     1

#WHOIS COUNTRY MALIGNA

A6_M<-table(matrizCAM$WHOIS_COUNTRY)
head(A6_M)

##
##            AL     BR BRASIL     CA     CN
##      1      1    188     30      2    181

barplot(table(matrizCAM$WHOIS_COUNTRY),
        main="Pais",
        xlab="Tipos de Pais",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```



Los paises mas destacados son: Estados Unidos 272 Brasil 218 China 181 Sin registro 1151

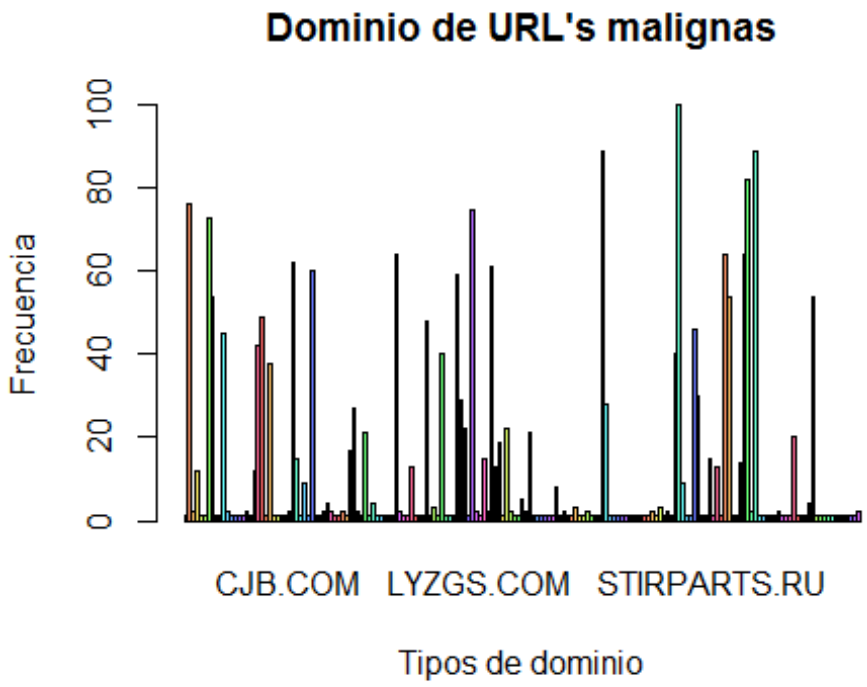
```
#WHOIS STATEPROV MALIGNA

A7_M<-table(matrizCAM$WHOIS_STATEPROV)
head(A7_M)

##
##                  AL     ANHUI        AZ BARCELONA   BEIJING
##         1        31        22        94        82         3

barplot(table(matrizCAM$WHOIS_STATEPROV),
        main="Estado",
        xlab="Tipos de Estado",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```



Los estados mas destacados son: Arizona - 92 Barcelona - 82 Jiangsu - 59 Sin registro 1422

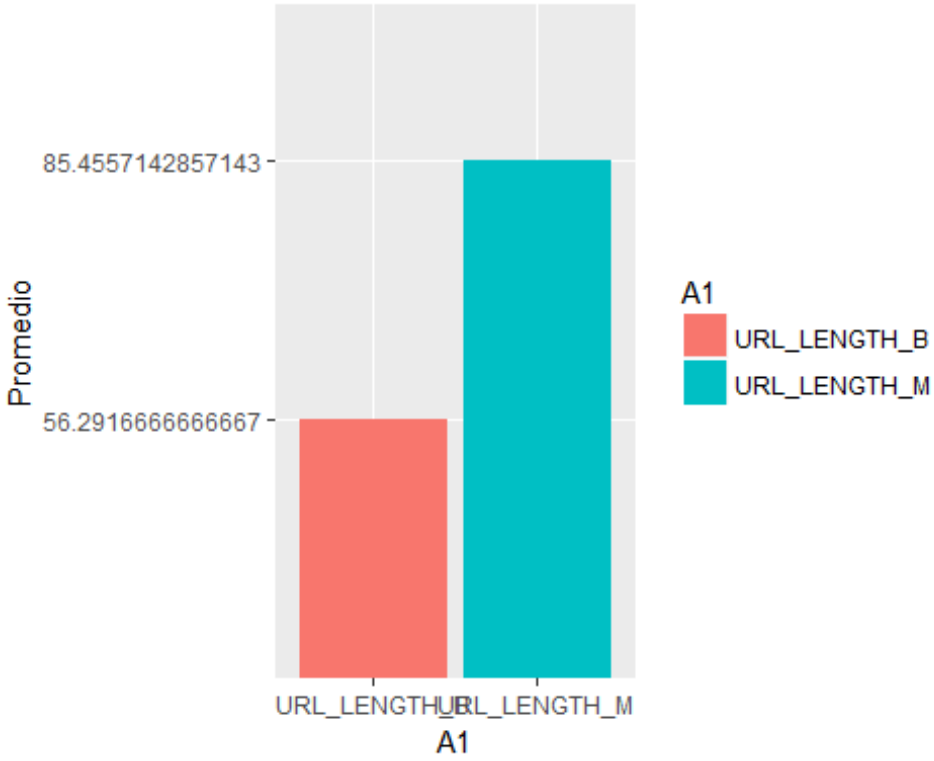
```
#WHOIS REGDATE MALIGNA

A8_M<-table(matrizCAM$WHOIS_REGDATE )
head(A8_M)

##
##                                0 01/03/2005 00:00 01/08/2006 00:00
##               1                1                1               59
```

```
## 01/08/2016 00:00 01/12/2003 00:00
##                1               42

#UPDATE DATE

A9_M<-table(matrizCAM$UPDATE_DATE)
head(A9_M)

##
##                  01/12/2016 00:00 02/01/2017 00:00 02/09/2016 00:00
##                1                1                1               62
## 03/01/2017 00:00 03/03/2017 00:00
##               76                1

#WHITIN DOMAIN

A10_M<-table(matrizCAM$WHITIN_DOMAIN)
head(A10_M)

##
##                      23283333.COM       4POWER.BIZ  88LOGISTICS.COM
##                1               76                2               12
##  ABESPREV.COM.BR AINTDOINSHIT.COM
##                1                1

barplot(table(matrizCAM$WHITIN_DOMAIN),
        main="Dominio de URL's malignas",
        xlab="Tipos de dominio",
        ylab="Frecuencia",
        col=rainbow(20, alpha = .6)
        )
```

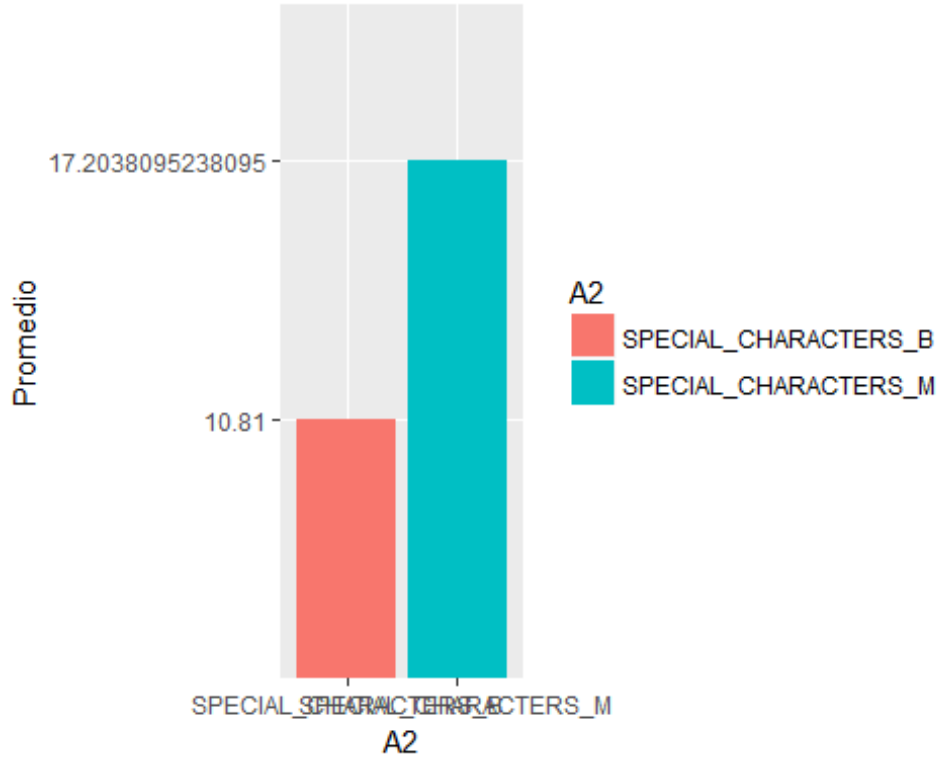


Comparacion entre la media de datos malignos y benginos de la capa aplicacion

```
#A1 que indica la caracteristica URL_LENGTH
NAMES_A1 = rbind("URL_LENGTH_M", "URL_LENGTH_B")
PROMEDIO_A1 = rbind(A1_M, A1_B)
URL_LENGTH_A1 = cbind(NAMES_A1,PROMEDIO_A1 )
URL_LENGTH_A1 <- data.frame(URL_LENGTH_A1)
names(URL_LENGTH_A1) <- c("A1","Promedio") #Se asignan los nombres a las
columas

URL_LENGTH_A1

##                 A1         Promedio
## A1_M URL_LENGTH_M 85.4557142857143
## A1_B URL_LENGTH_B 56.2916666666667

ggplot(data=URL_LENGTH_A1, aes(x=A1, y=Promedio, fill=A1)) +
    geom_bar(stat="identity", position=position_dodge())
```




```
#A2 que indica la caracteristica CHARACTERS
NAMES_A2 = rbind("SPECIAL_CHARACTERS_M", "SPECIAL_CHARACTERS_B")
PROMEDIO_A2 = rbind(A2_M, A2_B)
NUMBER_SPECIAL_CHARACTERS_A2= cbind(NAMES_A2, PROMEDIO_A2)
NUMBER_SPECIAL_CHARACTERS_A2 <- data.frame(NUMBER_SPECIAL_CHARACTERS_A2)
names(NUMBER_SPECIAL_CHARACTERS_A2) <- c("A2", "Promedio")

NUMBER_SPECIAL_CHARACTERS_A2

##                         A2         Promedio
## A2_M SPECIAL_CHARACTERS_M 17.2038095238095
## A2_B SPECIAL_CHARACTERS_B            10.81

ggplot(data=NUMBER_SPECIAL_CHARACTERS_A2, aes(x=A2, y=Promedio, fill=A2))
+
    geom_bar(stat="identity", position=position_dodge())
```

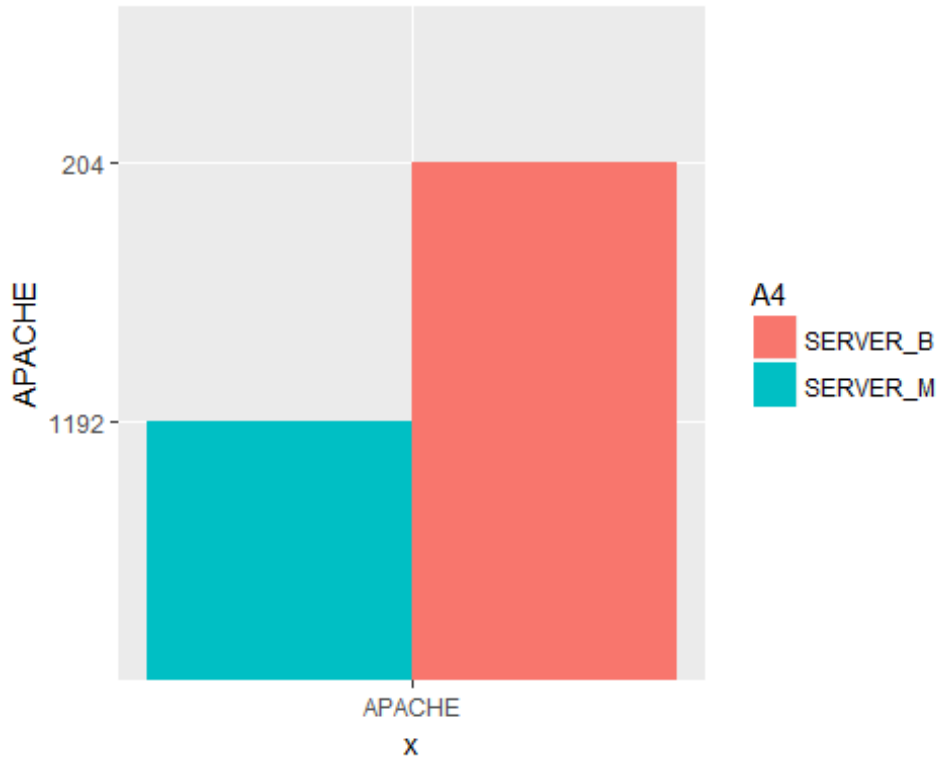


A3 que indica la caracteristica CHARACTERS

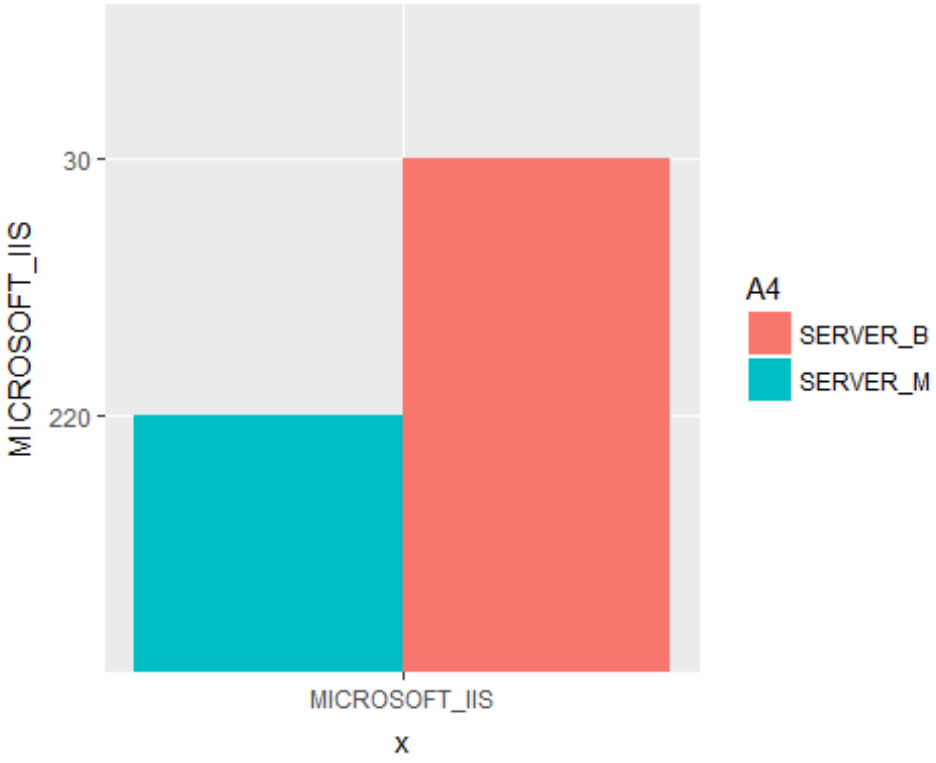
Este no tiene ningun valor en comun, por ende, no se compara

```
#A4 que indica la caracteristica SERVER


A4_M1 <- rbind(A4_M)
A4_M1 <- data.frame(A4_M1)
APACHE <- A4_M1$APACHE
MICROSOFT_IIS <- A4_M1$MICROSOFT.IIS
CLOUDFLARE_NGINX <- A4_M1$CLOUDFLARE.NGINX
MICROSOFT_HTTPAPI <- A4_M1$MICROSOFT.HTTPAPI
NGINX <- A4_M1$NGINX
NONE <- A4_M1$NONE

A4_M2 = cbind(APACHE, MICROSOFT_IIS,CLOUDFLARE_NGINX ,MICROSOFT_HTTPAPI
,NGINX ,NONE )


A4_B1 <- rbind(A4_B)
A4_B1 <- data.frame(A4_B1)
APACHE <- A4_B1$APACHE
MICROSOFT_IIS <- A4_B1$MICROSOFT.IIS
CLOUDFLARE_NGINX<- A4_B1$CLOUDFLARE.NGINX
MICROSOFT_HTTPAPI <- A4_B1$MICROSOFT.HTTPAPI
NGINX <- A4_B1$NGINX
```

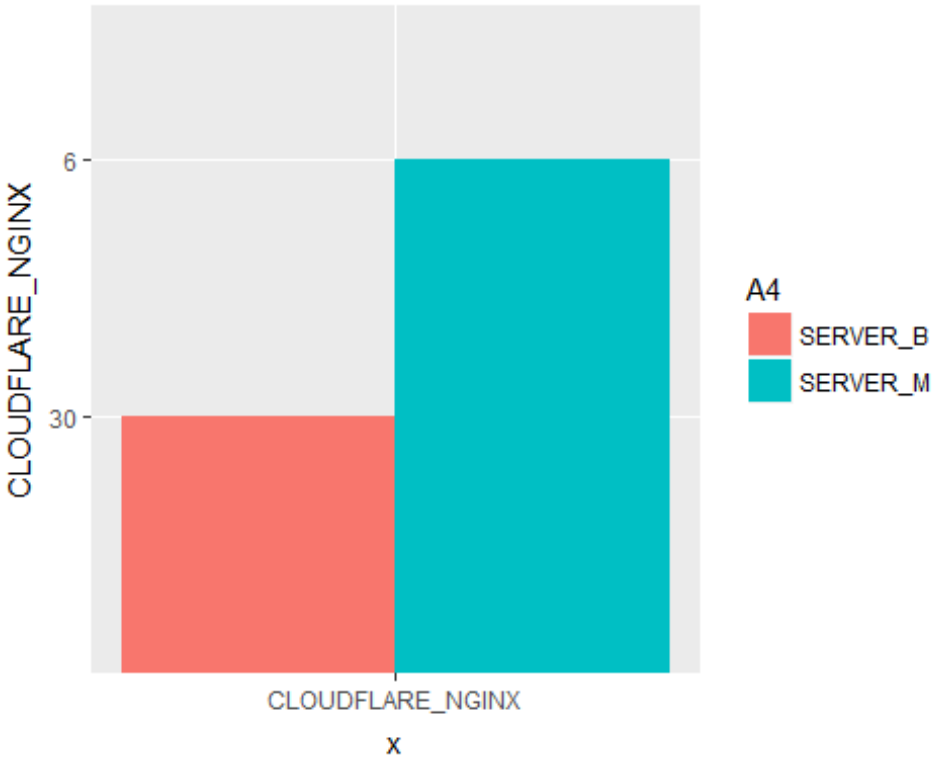
```
NONE <- A4_B1$NONE

A4_B2 = cbind(APACHE, MICROSOFT_IIS,CLOUDFLARE_NGINX ,MICROSOFT_HTTPAPI
,NGINX ,NONE )


NAMES_A4 = rbind("SERVER_B","SERVER_M")
PROMEDIO_A4 = rbind(A4_B2,A4_M2)
SERVER_A4= cbind(NAMES_A4, PROMEDIO_A4)
SERVER_A4 <- data.frame(SERVER_A4)
names(SERVER_A4) <- c("A4", "APACHE","MICROSOFT_IIS" ,"CLOUDFLARE_NGINX"
,"MICROSOFT_HTTPAPI" ,"NGINX","NONE" )

ggplot(data=SERVER_A4, aes(x="APACHE", y=APACHE, fill=A4)) +
    geom_bar(stat="identity", position=position_dodge())
```



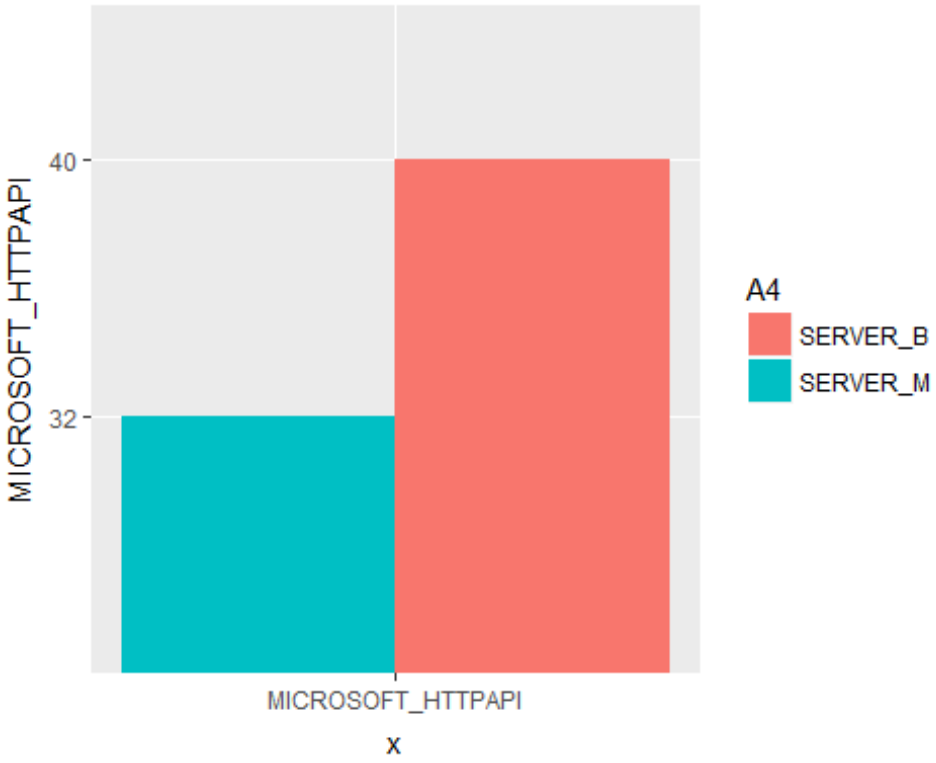

```
ggplot(data=SERVER_A4, aes(x="MICROSOFT_IIS", y=MICROSOFT_IIS, fill=A4))
+
    geom_bar(stat="identity", position=position_dodge())
```



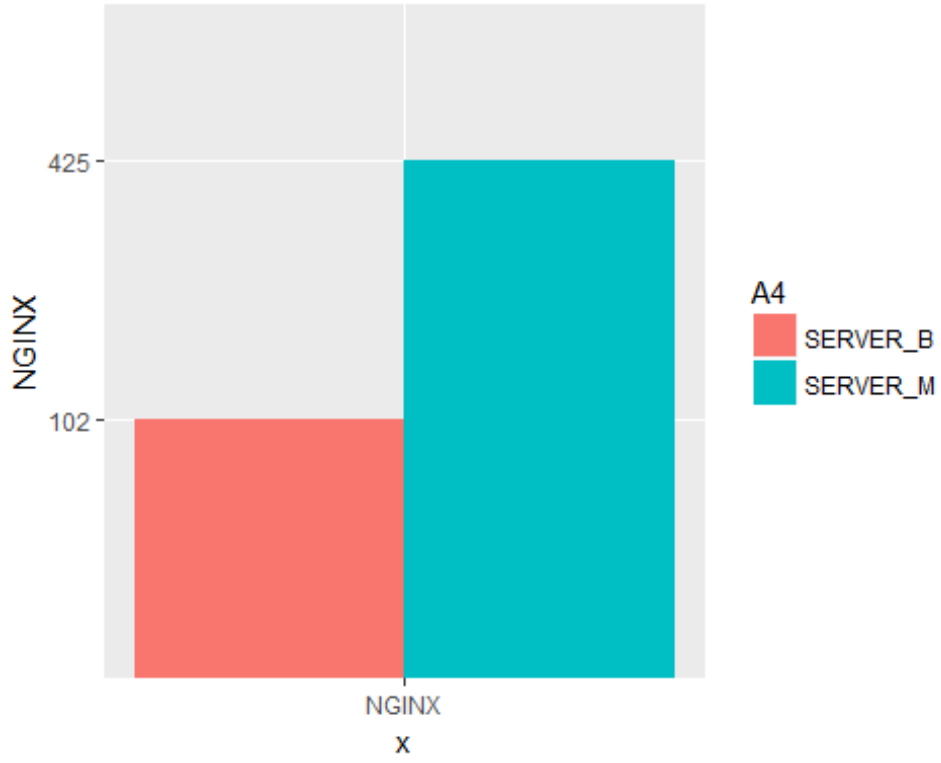

```
ggplot(data=SERVER_A4, aes(x="CLOUDFLARE_NGINX", y=CLOUDFLARE_NGINX,
fill=A4)) +
    geom_bar(stat="identity", position=position_dodge())
```



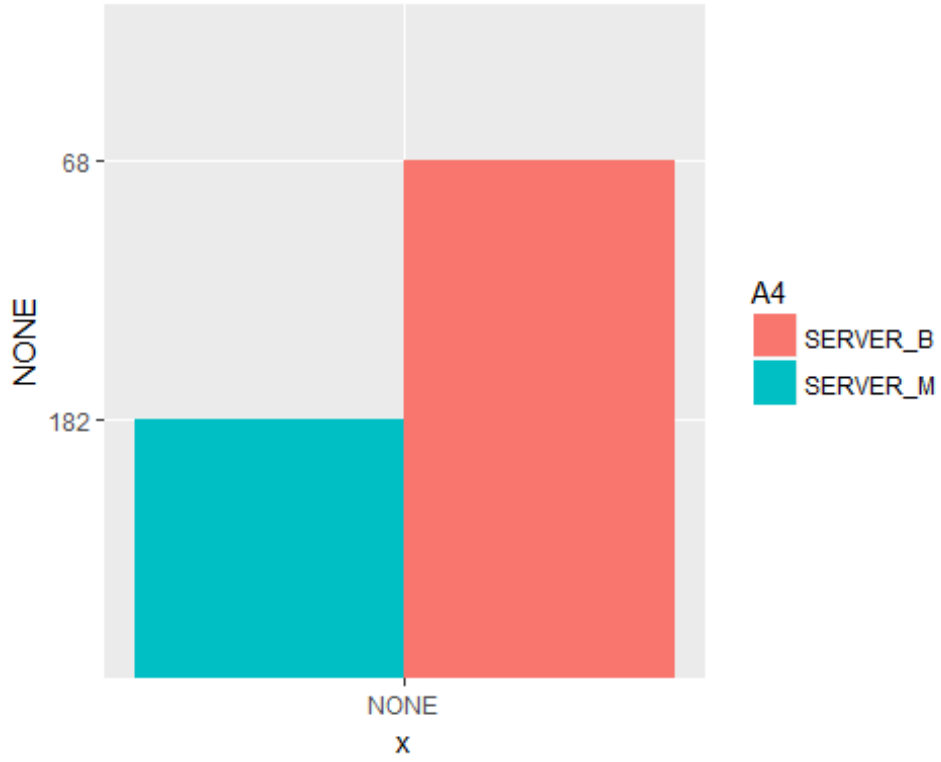
```
ggplot(data=SERVER_A4, aes(x="MICROSOFT_HTTPAPI", y=MICROSOFT_HTTPAPI,
fill=A4)) +
    geom_bar(stat="identity", position=position_dodge())
```




```
ggplot(data=SERVER_A4, aes(x="NGINX", y=NGINX, fill=A4)) +
    geom_bar(stat="identity", position=position_dodge())
```



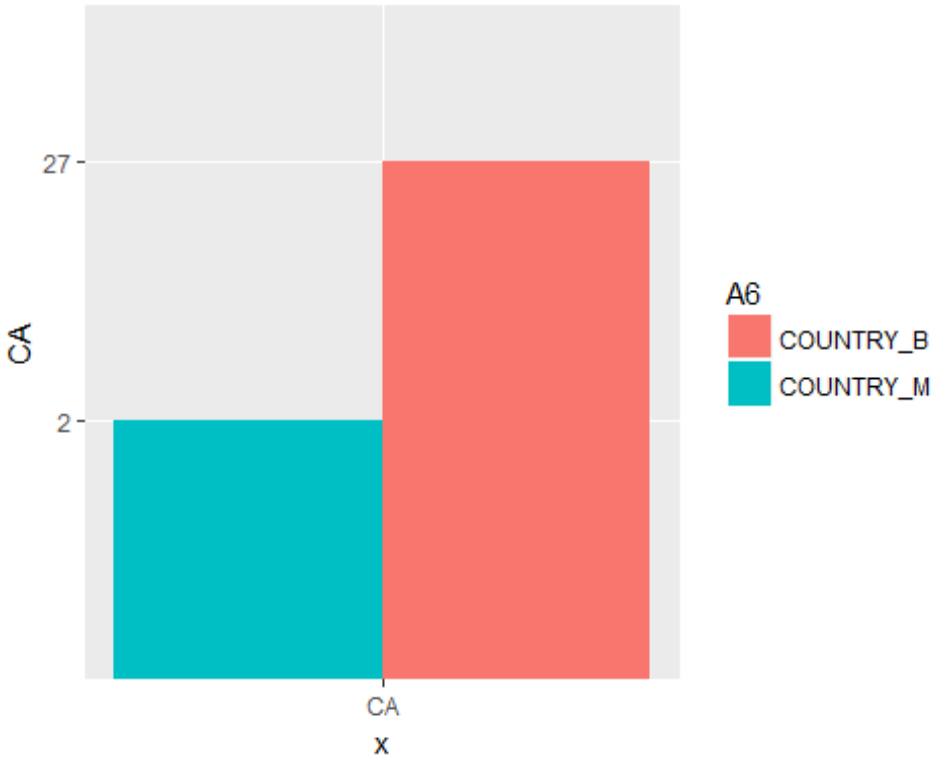

```
ggplot(data=SERVER_A4, aes(x="NONE", y=NONE, fill=A4)) +
    geom_bar(stat="identity", position=position_dodge())
```



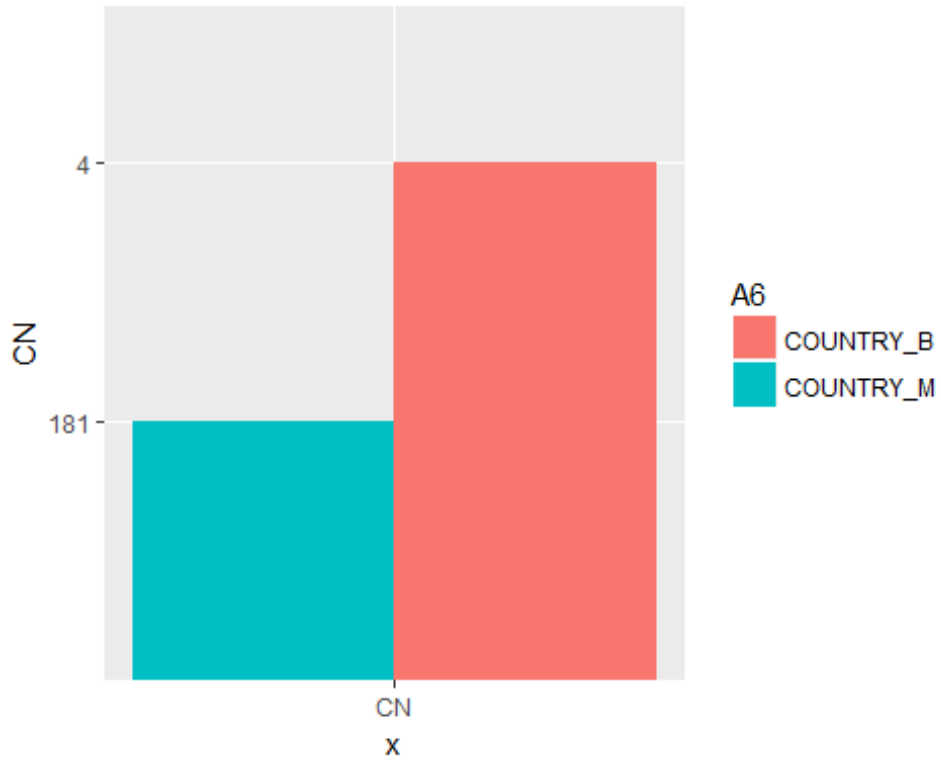
```r
#A6 que indica la caracteristica COUNTRY


A6_M1 <- rbind(A6_M)
A6_M1 <- data.frame(A6_M1)
CA <- A6_M1$CA
CN <- A6_M1$CN
CZ <- A6_M1$CZ
FR <- A6_M1$FR
GB <- A6_M1$GB
HK <- A6_M1$HK
IN <- A6_M1$IN
NONE <- A6_M1$NONE
PA <- A6_M1$PA
UK <- A6_M1$UK
US <- A6_M1$US

A6_M2 = cbind(CA, CN,CZ ,FR ,GB ,HK, IN,NONE,PA, UK,US)


A6_B1 <- rbind(A6_B)
A6_B1 <- data.frame(A6_B1)
CA <- A6_B1$CA
CN <- A6_B1$CN
CZ <- A6_B1$CZ
FR <- A6_B1$FR
GB <- A6_B1$GB
HK <- A6_B1$HK
IN <- A6_B1$IN
NONE <- A6_B1$NONE
PA <- A6_B1$PA
UK <- A6_B1$UK
US <- A6_B1$US

A6_B2 = cbind(CA, CN,CZ ,FR ,GB ,HK, IN,NONE,PA, UK,US)


NAMES_A6 = rbind("COUNTRY_M", "COUNTRY_B")
PROMEDIO_A6 = rbind(A6_M2, A6_B2)
COUNTRY_A6= cbind(NAMES_A6, PROMEDIO_A6)
COUNTRY_A6 <- data.frame(COUNTRY_A6)
names(COUNTRY_A6) <- c("A6", "CA", "CN","CZ" ,"FR" ,"GB" ,"HK",
"IN","NONE","PA", "UK","US")

COUNTRY_A6
```

```
##          A6 CA  CN CZ FR GB HK IN NONE PA UK  US
## 1 COUNTRY_M  2 181  8  2 18  2  1 1150 63  1 272
## 2 COUNTRY_B 27   4  1  4  7  1  3  120  6  2 383
```

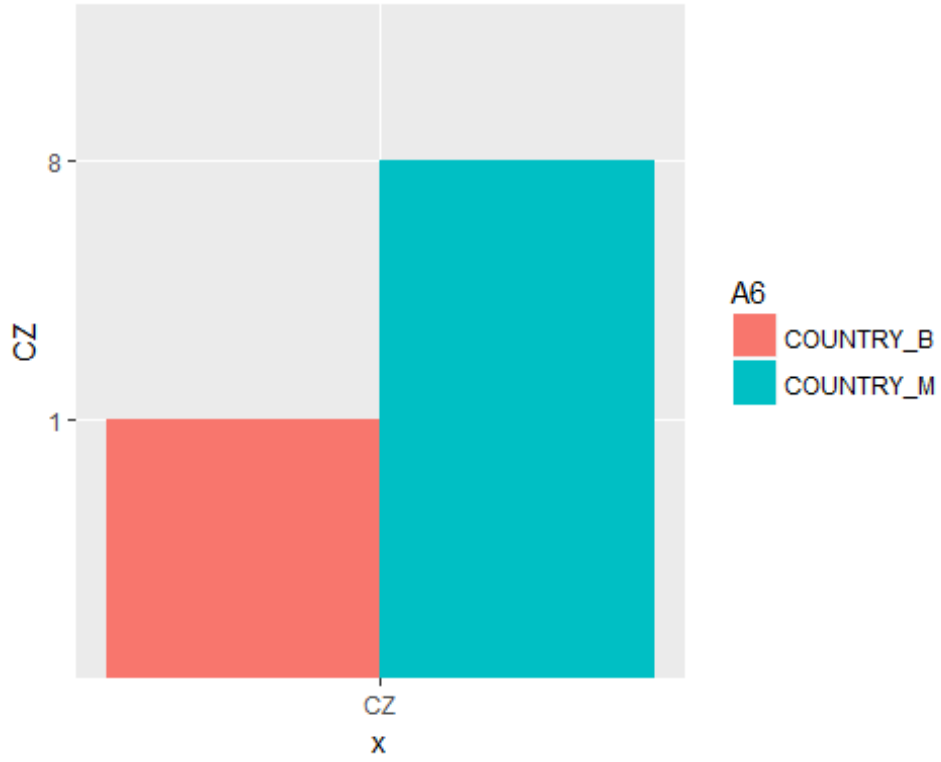
```
ggplot(data=COUNTRY_A6, aes(x="CA", y=CA, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



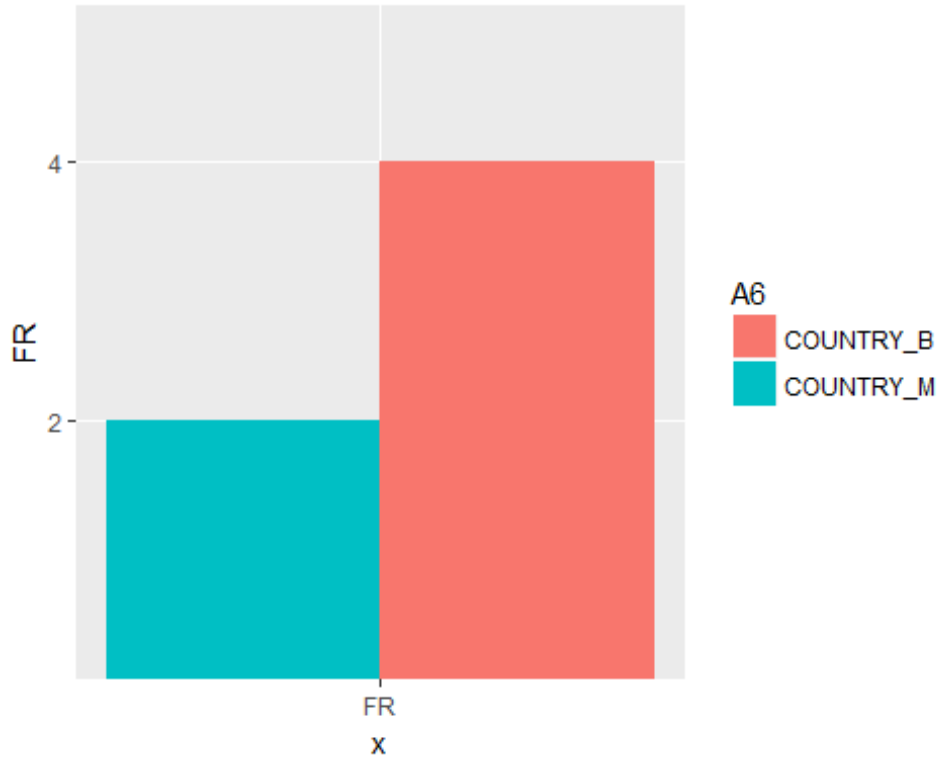

```
ggplot(data=COUNTRY_A6, aes(x="CN", y=CN, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



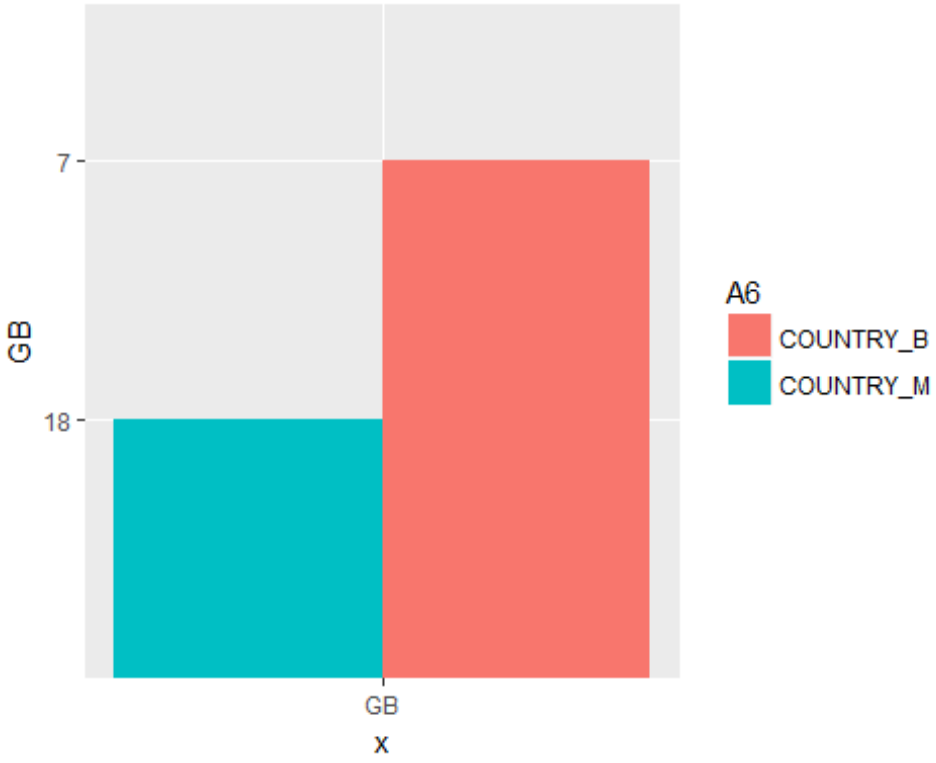

```
ggplot(data=COUNTRY_A6, aes(x="CZ", y=CZ, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



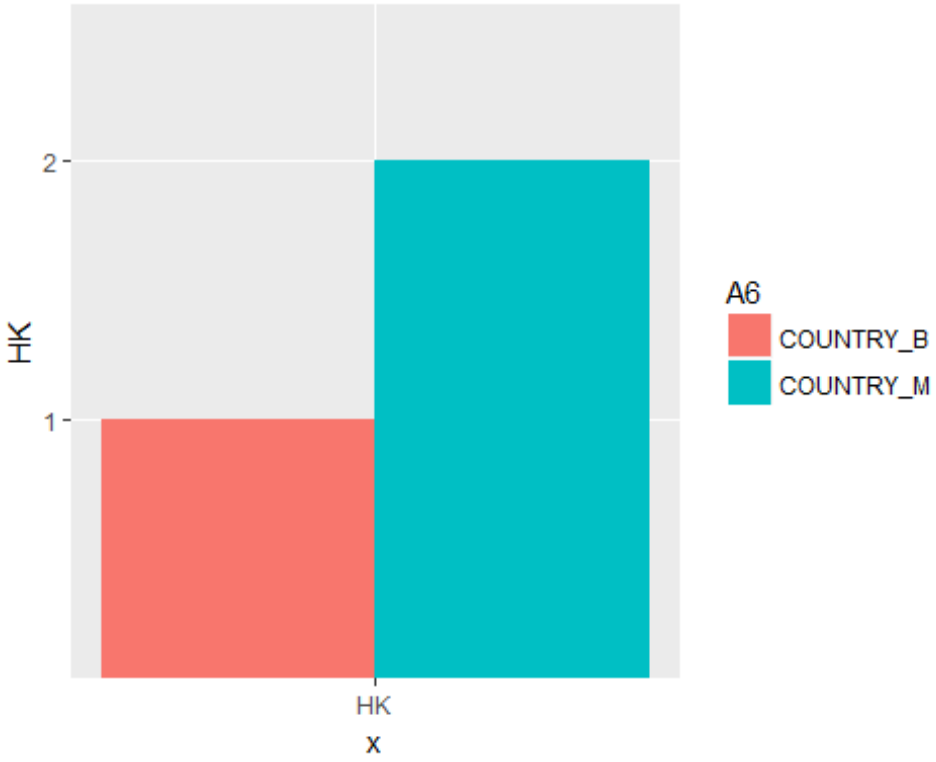
```
ggplot(data=COUNTRY_A6, aes(x="FR", y=FR, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



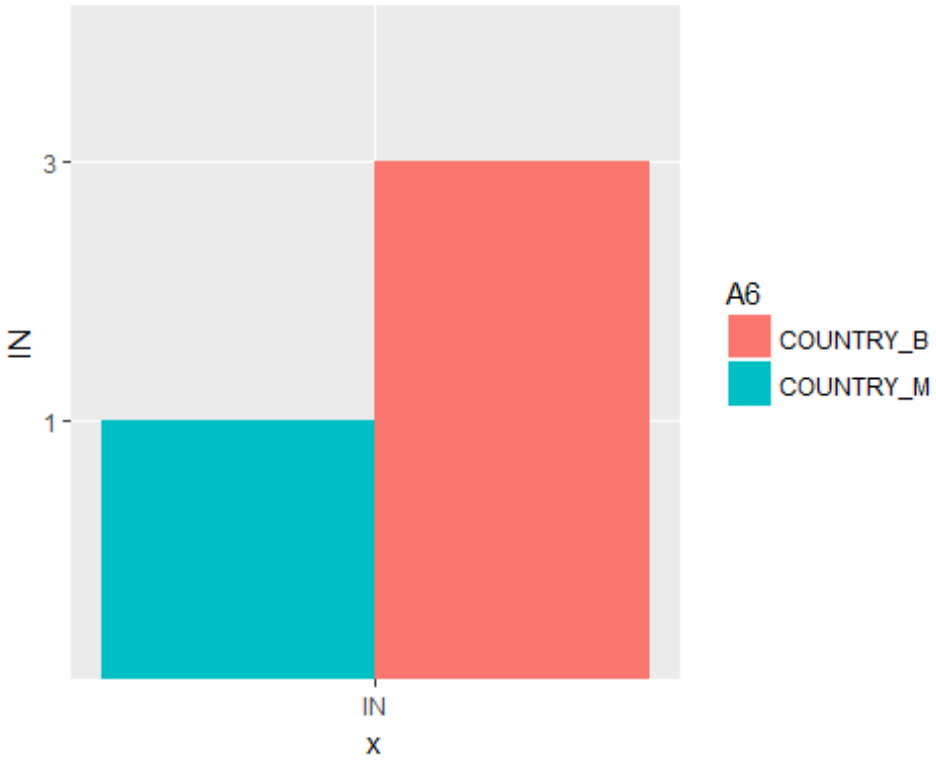

```
ggplot(data=COUNTRY_A6, aes(x="GB", y=GB, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



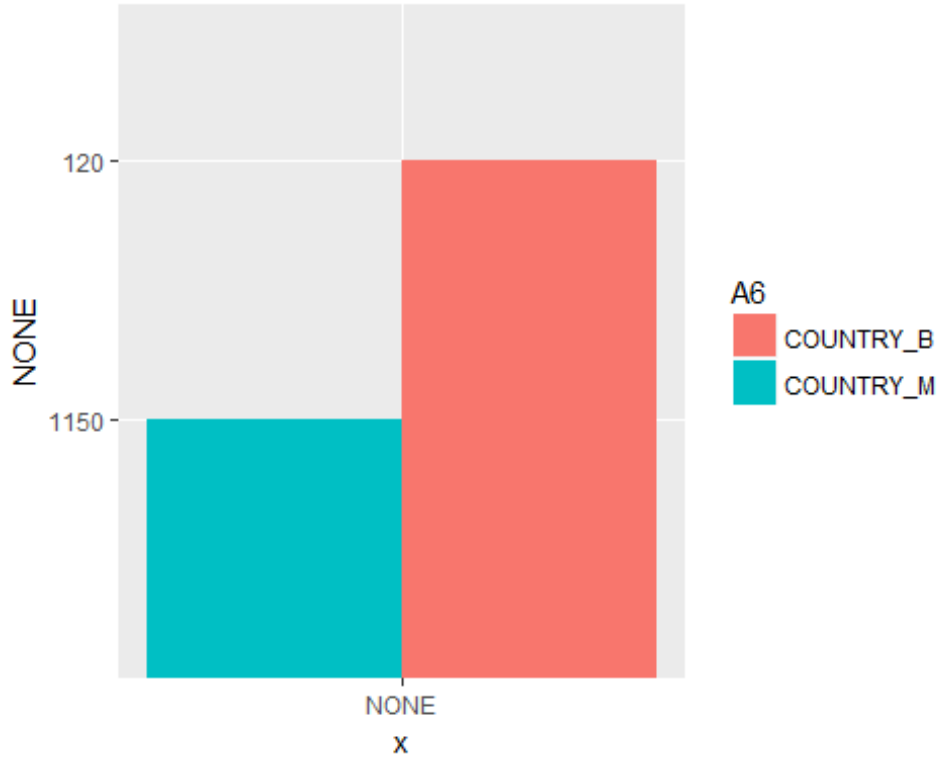

```
ggplot(data=COUNTRY_A6, aes(x="HK", y=HK, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



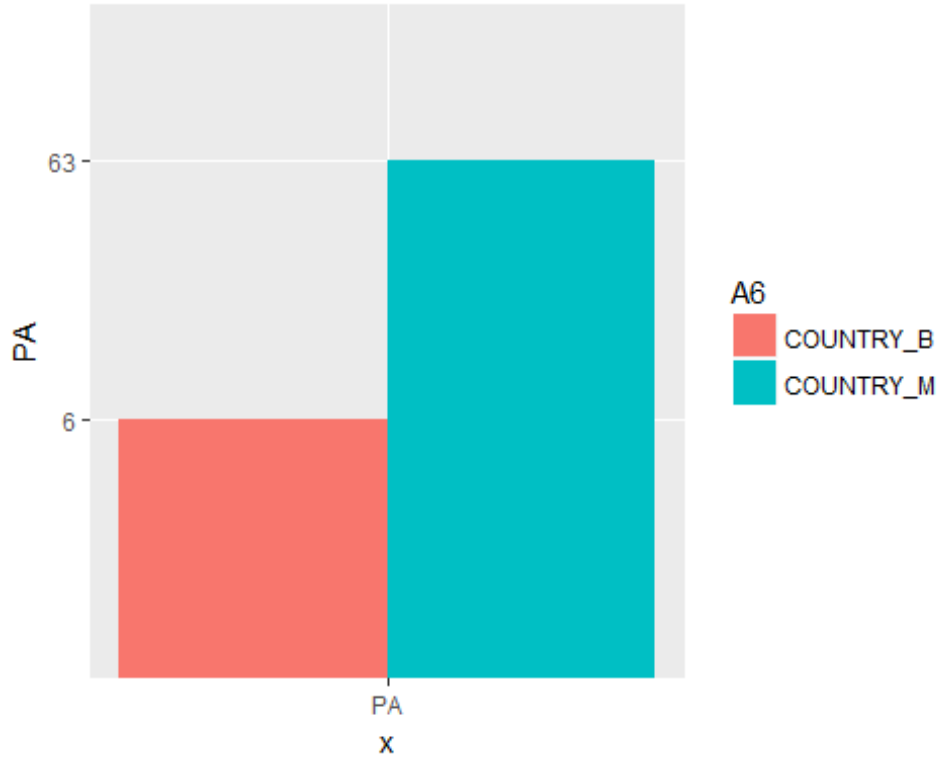
```
ggplot(data=COUNTRY_A6, aes(x="IN", y=IN, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



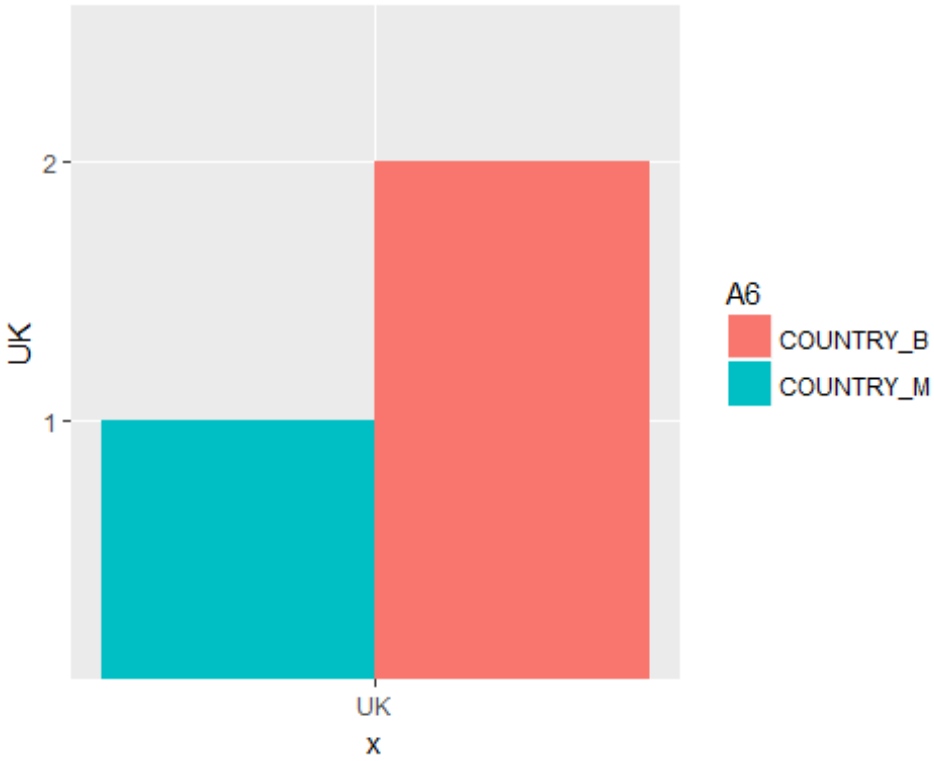

```
ggplot(data=COUNTRY_A6, aes(x="NONE", y=NONE, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



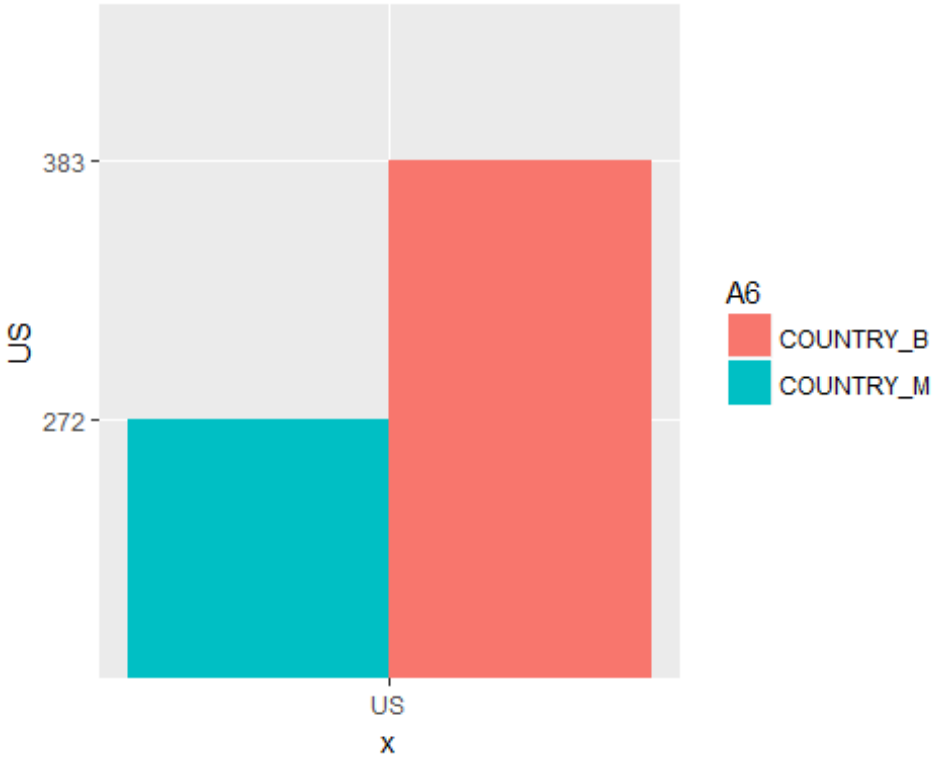

```
ggplot(data=COUNTRY_A6, aes(x="PA", y=PA, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```



```
ggplot(data=COUNTRY_A6, aes(x="UK", y=UK, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```

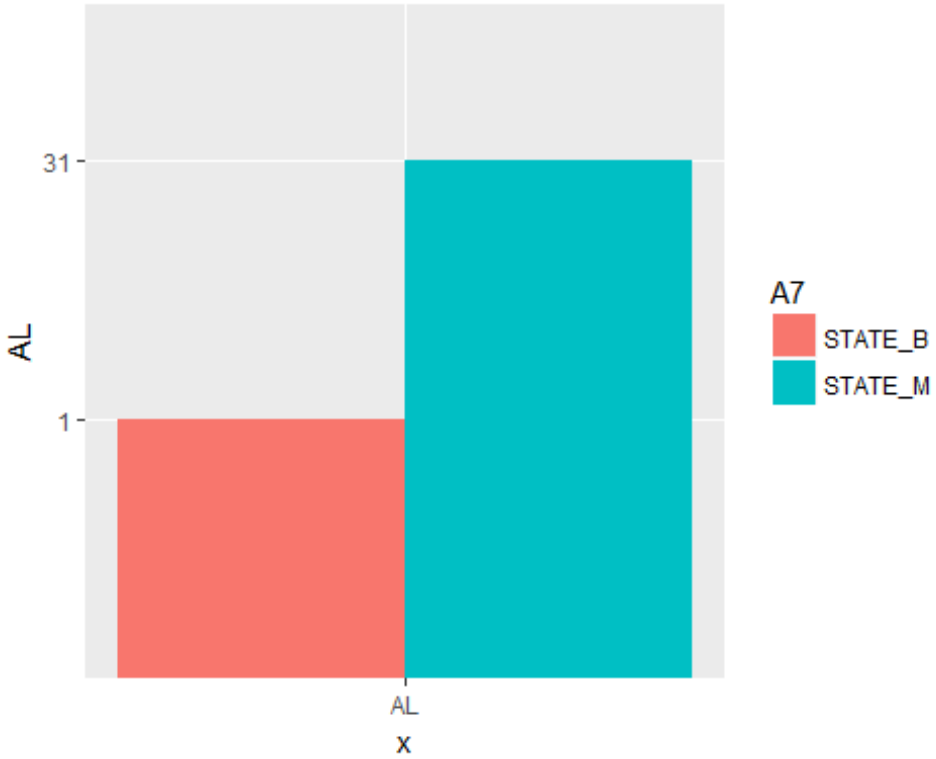



```
ggplot(data=COUNTRY_A6, aes(x="US", y=US, fill=A6)) +
    geom_bar(stat="identity", position=position_dodge())
```

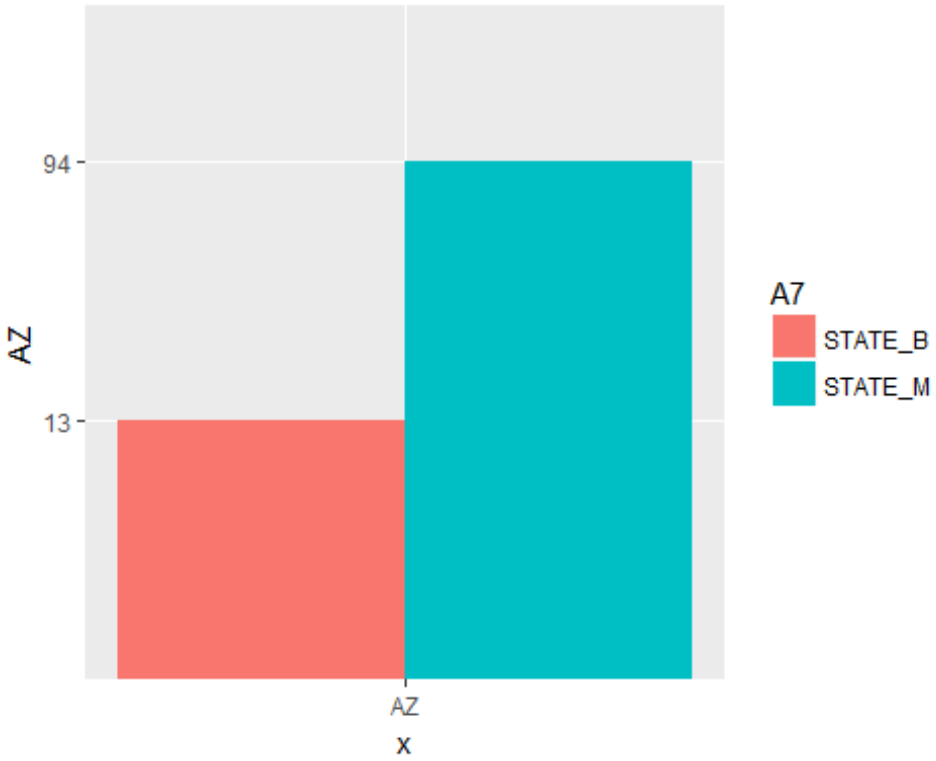



```
#A7 que indica la caracteristica WHOIS_STATEPROV


A7_M1 <- rbind(A7_M)
A7_M1 <- data.frame(A7_M1)
AL <- A7_M1$AL
AZ <- A7_M1$AZ
BEIJINGSHI <- A7_M1$BEIJINGSHI
CA <- A7_M1$CA
CO <- A7_M1$CO
FL <- A7_M1$FL
LA <- A7_M1$LA
MA <- A7_M1$MA
PA <- A7_M1$PA
TX <- A7_M1$TX
WA <- A7_M1$WA

A7_M2 = cbind(AL, AZ,BEIJINGSHI ,CA ,CO ,FL, LA,MA,PA, TX,WA)

A7_B1 <- rbind(A7_B)
A7_B1 <- data.frame(A7_B1)
AL <- A7_B1$AL
AZ <- A7_B1$AZ
BEIJINGSHI <- A7_B1$BEIJINGSHI
CA <- A7_B1$CA
CO <- A7_B1$CO
```

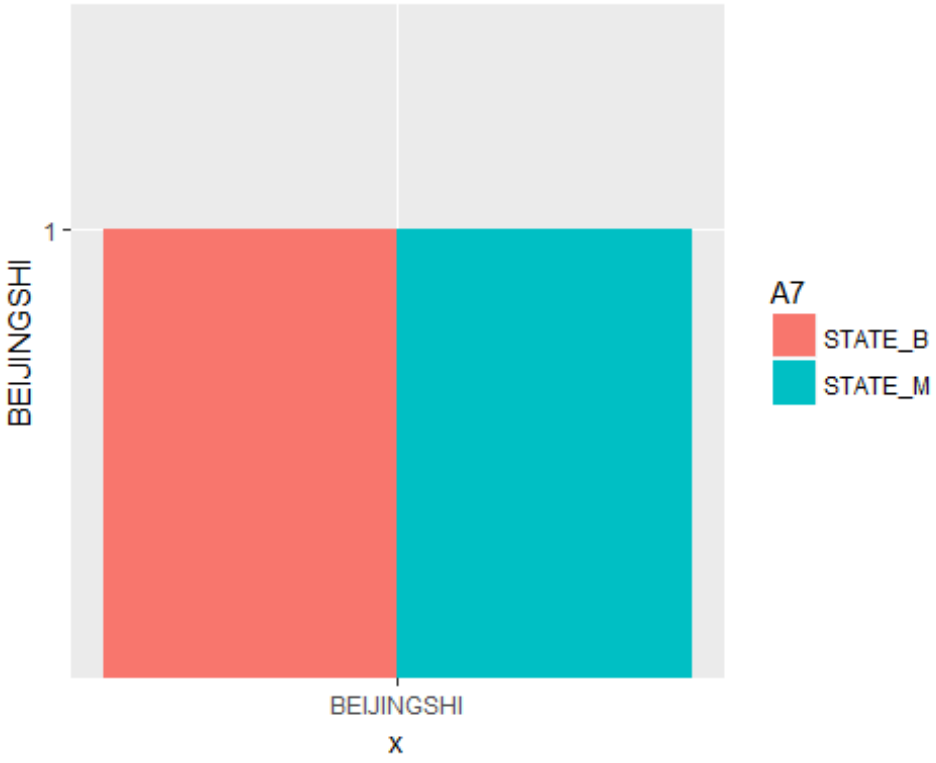
```
FL <- A7_B1$FL
LA <- A7_B1$LA
MA <- A7_B1$MA
PA <- A7_B1$PA
TX <- A7_B1$TX
WA <- A7_B1$WA

A7_B2 = cbind(AL, AZ,BEIJINGSHI ,CA ,CO ,FL, LA,MA,PA, TX,WA)


NAMES_A7 = rbind("STATE_M", "STATE_B")
PROMEDIO_A7 = rbind(A7_M2, A7_B2)
STATE_A7= cbind(NAMES_A7, PROMEDIO_A7)
STATE_A7 <- data.frame(STATE_A7)
names(STATE_A7) <- c("A7", "AL", "AZ","BEIJINGSHI" ,"CA" ,"CO" ,"FL",
"LA","MA","PA", "TX","WA")

STATE_A7

##        A7 AL AZ BEIJINGSHI  CA CO FL LA MA PA TX WA
## 1 STATE_M 31 94          1  16  2 32  1  1 66  1 89
## 2 STATE_B  1 13          1 158 11 27  2  5 15  6 24

ggplot(data=STATE_A7, aes(x="AL", y=AL, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



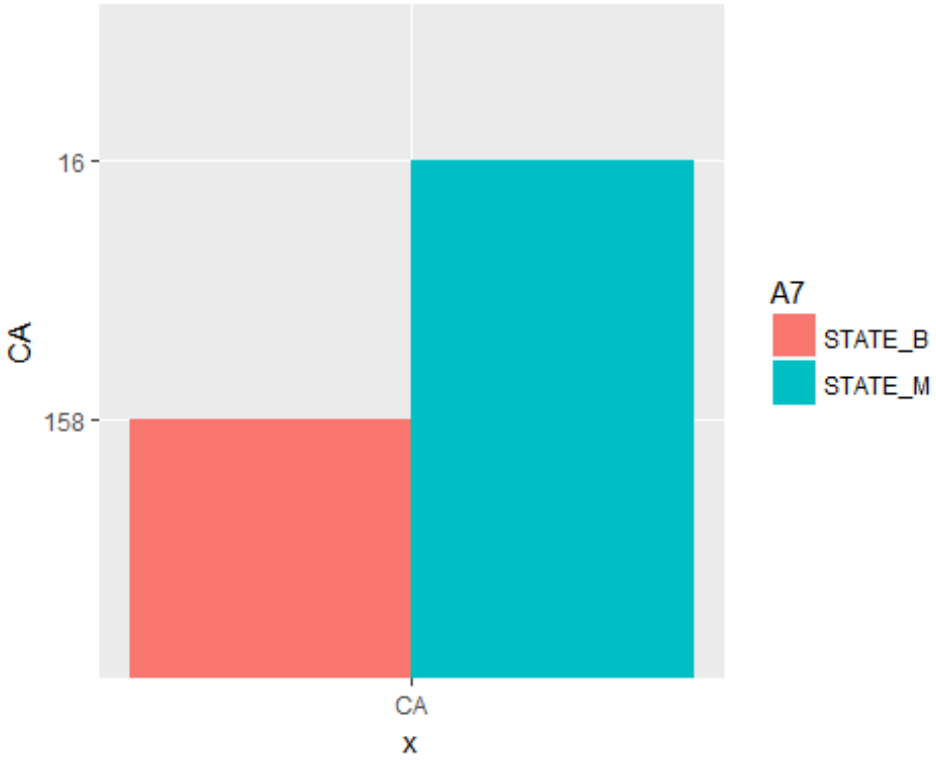
```
ggplot(data=STATE_A7, aes(x="AZ", y=AZ, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



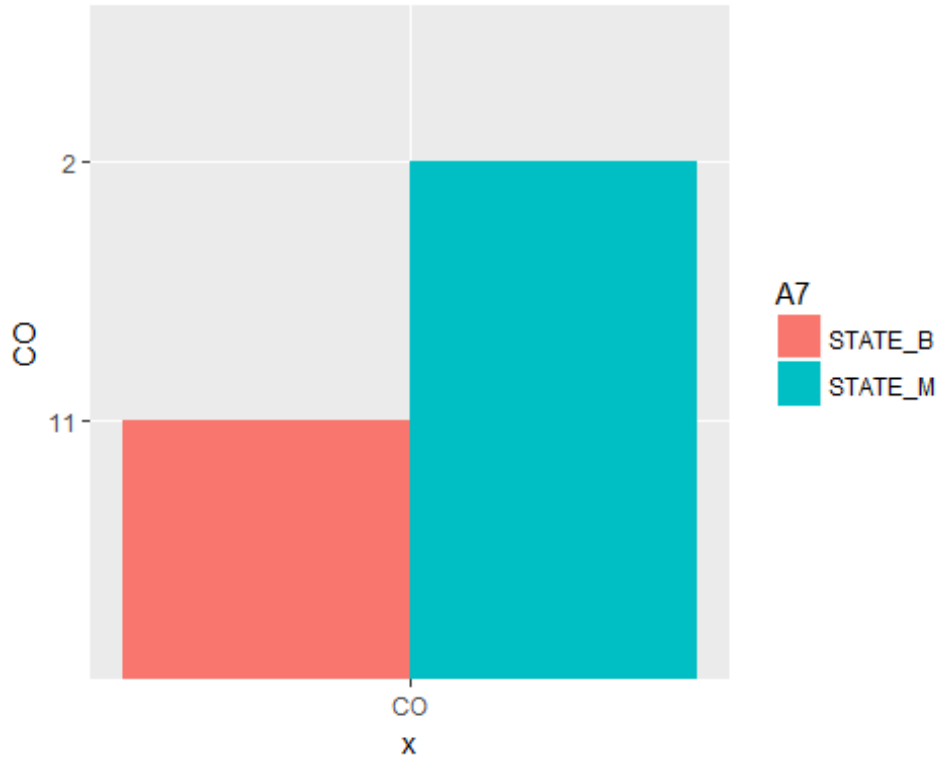

```
ggplot(data=STATE_A7, aes(x="BEIJINGSHI", y=BEIJINGSHI, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



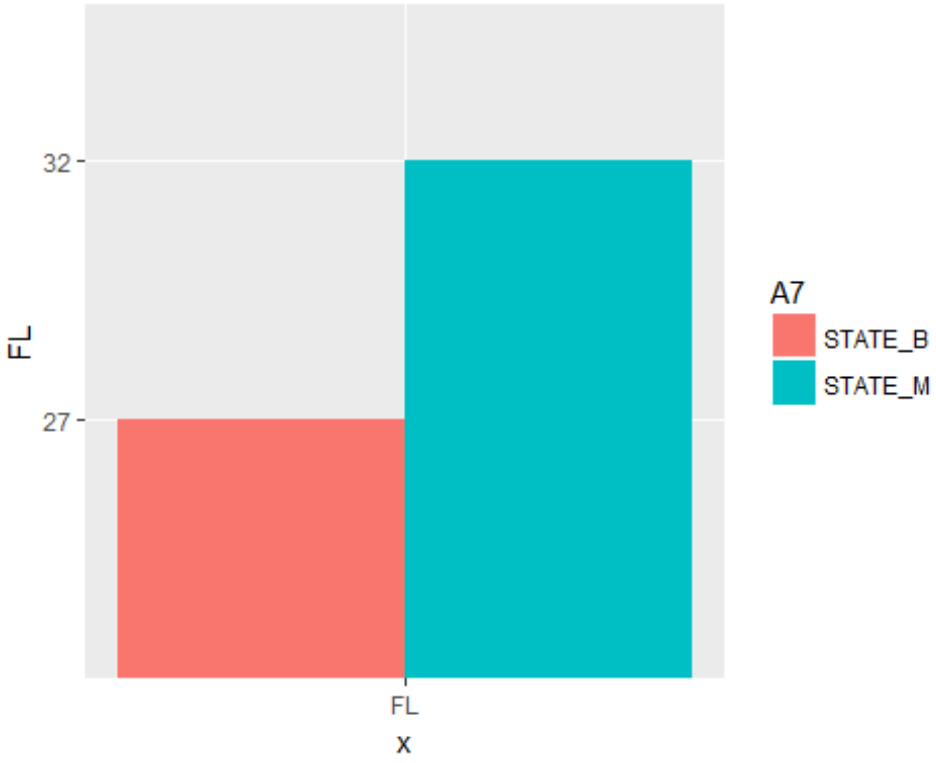

```
ggplot(data=STATE_A7, aes(x="CA", y=CA, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



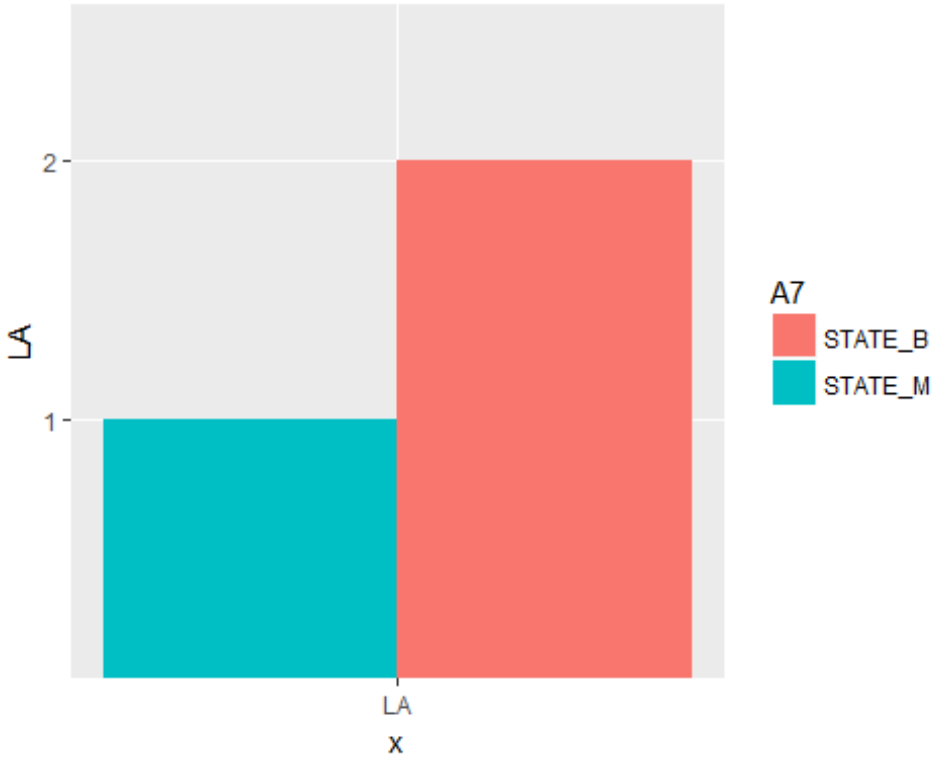
```
ggplot(data=STATE_A7, aes(x="CO", y=CO, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



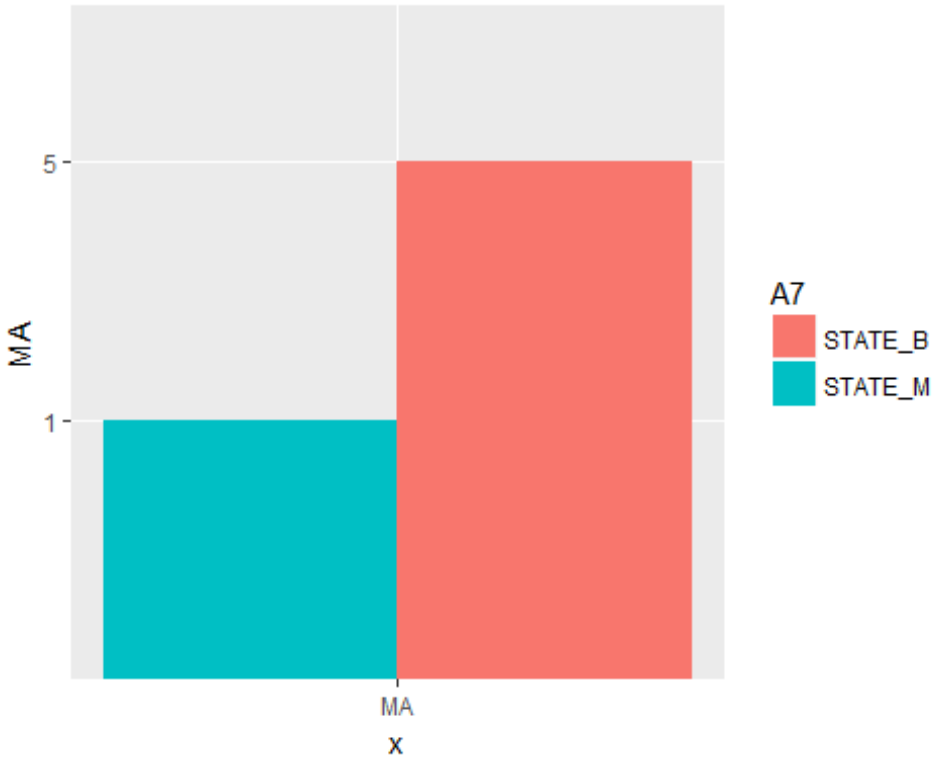

```
ggplot(data=STATE_A7, aes(x="FL", y=FL, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



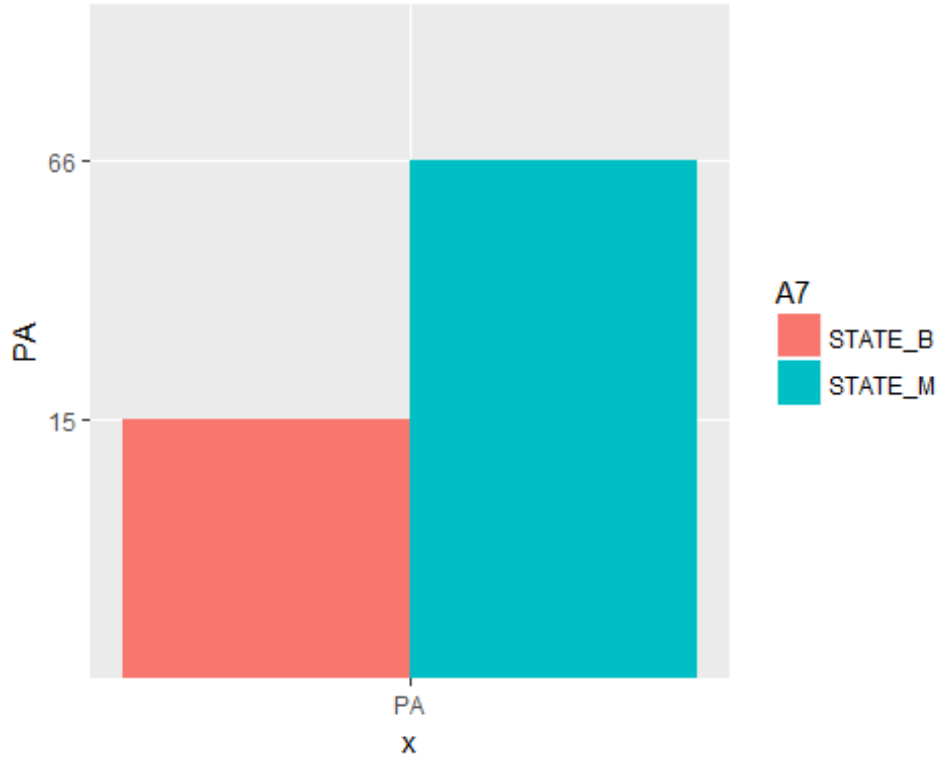

```
ggplot(data=STATE_A7, aes(x="LA", y=LA, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



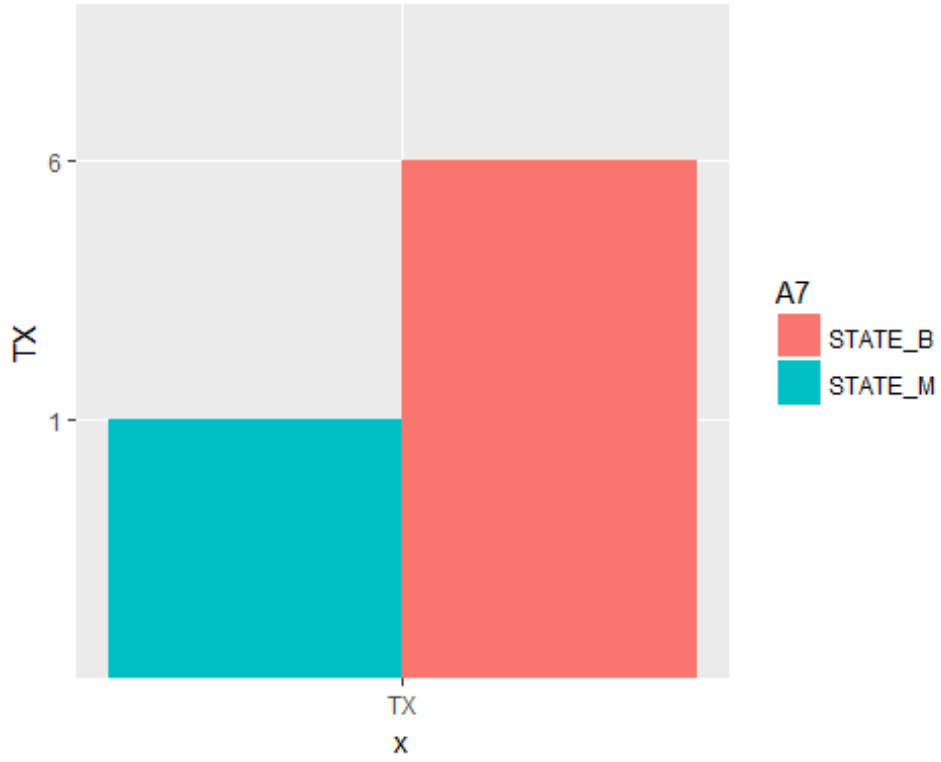
```
ggplot(data=STATE_A7, aes(x="MA", y=MA, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



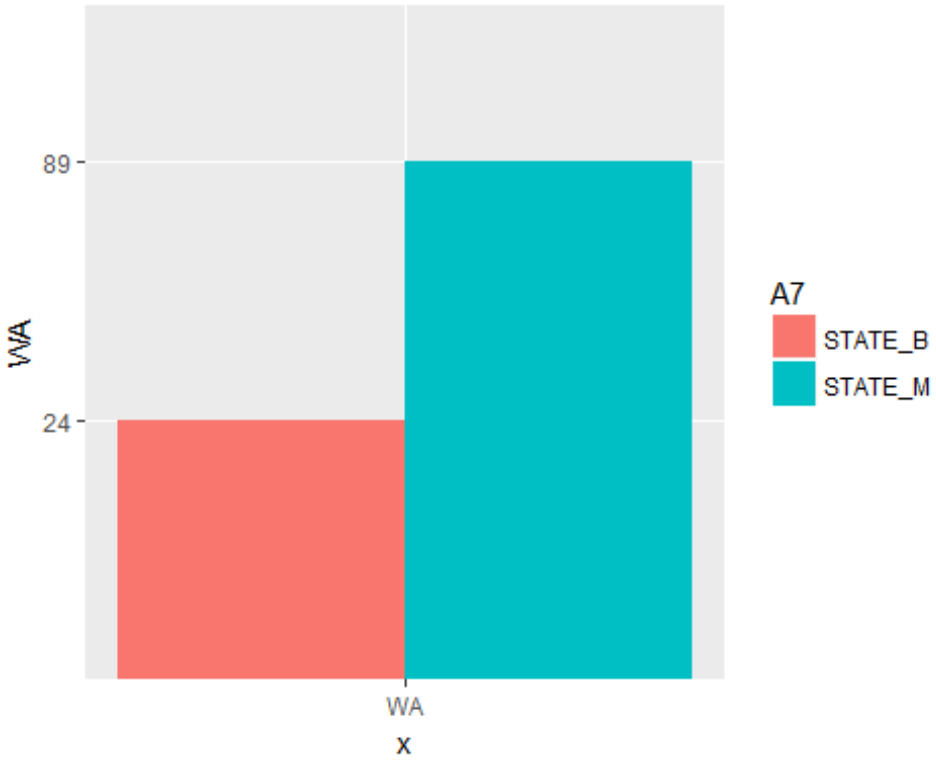

```
ggplot(data=STATE_A7, aes(x="PA", y=PA, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```




```
ggplot(data=STATE_A7, aes(x="TX", y=TX, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



```
ggplot(data=STATE_A7, aes(x="WA", y=WA, fill=A7)) +
    geom_bar(stat="identity", position=position_dodge())
```



CAPA DE RED

Capa de Red Benigna

```
matrizCRB <- read.csv("matriz_red_benigno.csv", sep=";", comment.char =
"#")
matrizCRB<- na.omit(matrizCRB)
names(matrizCRB) <- c("URL",
"TCP_CONVERSATION_EXCHANGE",
"DIST_REMOTE_TCP_PORT",
"REMOTE_IPS",
"APP_BYTES",
"UDP_PACKETS",
"TCP_URG_PACKETS",
"SOURCE_APP_PACKETS",
"REMOTE_APP_PACKETS",
"SOURCE_APP_BYTES",
"REMOTE_APP_BYTES",
"DURATION",
"AVG_LOCAL_PKT_RATE",
"AVG_REMOTE_PKT_RATE",
"APP_PACKETS",
"DNS_QUERY_TIMES")
```

```r
head(matrizCRB)
```

```
##      URL TCP_CONVERSATION_EXCHANGE DIST_REMOTE_TCP_PORT REMOTE_IPS
## 3   D0_3                       48                    0          1
## 7   D0_7                        4                    1          2
## 8   D0_8                       14                    9          2
## 9   D0_9                       16                    9          8
## 10 D0_10                        7                    0          2
## 12 D0_12                       15                   10          1
##    APP_BYTES UDP_PACKETS TCP_URG_PACKETS SOURCE_APP_PACKETS
## 3       3840           0               0                52
## 7        583           0               0                 6
## 8       2531           0               0                18
## 9       1275           0               0                18
## 10      1088           0               0                11
## 12      2421           0               0                21
##    REMOTE_APP_PACKETS SOURCE_APP_BYTES REMOTE_APP_BYTES DURATION
## 3                  51             52729             4156   680743
## 7                   8               737              735   716694
## 8                  18              8177             2819   718203
## 9                  23              2306             1423   727022
## 10                 10              7650             1440   727538
## 12                 17              7924             2861   729069
##    AVG_LOCAL_PKT_RATE AVG_REMOTE_PKT_RATE APP_PACKETS DNS_QUERY_TIMES
## 3             7640000             7490000          52               4
## 7               83700             1120000           6               2
## 8             2510000             2510000          18               4
## 9             2480000             3160000          18               2
## 10            1510000             1370000          11               4
## 12             288000             2330000          21               6
```

```r
#TCP CONVERSATION EXCHANGE BENIGNA

N1_B = mean(matrizCRB$TCP_CONVERSATION_EXCHANGE)
N1_B
```

```
## [1] 32.79622
```

```r
#DIST REMOTE TCP PORT BENIGNA

N2_B =mean(matrizCRB$DIST_REMOTE_TCP_PORT)
N2_B
```

```
## [1] 10.63866
```

```r
#REMOTE IPS BENIGNA

N3_B = mean(matrizCRB$REMOTE_IPS)
N3_B
```

```
## [1] 5.560924
```

```
#APP BYTES BENIGNA

N4_B = mean(matrizCRB$APP_BYTES)
N4_B
```

```
## [1] 3562.863
```

```
#UDP PACKETS BENIGNA

N5_B = mean(matrizCRB$UDP_PACKETS)
N5_B
```

```
## [1] 0
```

```
#TCP URG PACKETS BENIGNA

N6_B = mean(matrizCRB$TCP_URG_PACKETS)
N6_B
```

```
## [1] 0
```

```
#SOURCE APP PACKETS BENIGNA

N7_B = mean(matrizCRB$SOURCE_APP_PACKETS)
N7_B
```

```
## [1] 37.9958
```

```
#REMOTE APP PACKETS BENIGNA

N8_B = mean(matrizCRB$REMOTE_APP_PACKETS)
N8_B
```

```
## [1] 37.89916
```

```
#SOURCE APP BYTES BENIGNA

N9_B = mean(matrizCRB$SOURCE_APP_BYTES)
N9_B
```

```
## [1] 35038.54
```

```
#REMOTE APP BYTES BENIGNA

N10_B = mean(matrizCRB$REMOTE_APP_BYTES)
N10_B
```

```
## [1] 3961.748
```

```
#DURATION BENIGNA
```

```
N11_B = mean(as.numeric((matrizCRB$DURATION)))
N11_B

## [1] 8785775

#AVG LOCAL PKT RATE BENIGNA

N12_B = mean(matrizCRB$AVG_LOCAL_PKT_RATE)
N12_B

## [1] 643211.5

#AVG REMOTE PKT RATE BENIGNA

N13_B = mean(matrizCRB$AVG_REMOTE_PKT_RATE)
N13_B

## [1] 609425.4

#APP PACKETS BENIGNA

N14_B = mean(matrizCRB$APP_PACKETS)
N14_B

## [1] 37.9958

#DNS QUERY TIMES BENIGNA

N15_B = mean(matrizCRB$DNS_QUERY_TIMES)
N15_B

## [1] 5.17437

DatosTotalRedBenigno = cbind(N1_B, N2_B, N3_B, N4_B, N5_B, N6_B, N7_B,
N8_B, N9_B, N10_B, N11_B, N12_B, N13_B, N14_B, N15_B)
DatosRedBenigno <- data.frame(DatosTotalRedBenigno)

DatosRedBenigno

##       N1_B     N2_B     N3_B     N4_B N5_B N6_B    N7_B     N8_B
N9_B
## 1 32.79622 10.63866 5.560924 3562.863    0    0 37.9958 37.89916
35038.54
##      N10_B   N11_B    N12_B    N13_B   N14_B   N15_B
## 1 3961.748 8785775 643211.5 609425.4 37.9958 5.17437
```

CAPA DE RED MALIGNA

```
matrizCRM <- read.csv("matriz_red_maligno.csv", sep=";", comment.char =
"#")
matrizCRM<- na.omit(matrizCRM)
```

```r
names(matrizCRM) <- c("URL",
"TCP_CONVERSATION_EXCHANGE",
"DIST_REMOTE_TCP_PORT",
"REMOTE_IPS",
"APP_BYTES",
"UDP_PACKETS",
"TCP_URG_PACKETS",
"SOURCE_APP_PACKETS",
"REMOTE_APP_PACKETS",
"SOURCE_APP_BYTES",
"REMOTE_APP_BYTES",
"DURATION",
"AVG_LOCAL_PKT_RATE",
"AVG_REMOTE_PKT_RATE",
"APP_PACKETS",
"DNS_QUERY_TIMES")

head(matrizCRM)
```

```
##      URL TCP_CONVERSATION_EXCHANGE DIST_REMOTE_TCP_PORT REMOTE_IPS
## 4   M0_5                         5                    1          2
## 5   M0_6                        16                   12          3
## 14 M0_15                         9                    1          2
## 19 M0_20                         3                    0          1
## 20 M0_21                         5                    0          1
## 38 M0_39                        11                    0          2
##    APP_BYTES UDP_PACKETS TCP_URG_PACKETS SOURCE_APP_PACKETS
## 4        636           0               0                 11
## 5       2409           0               0                 20
## 14       878           0               0                 11
## 19       460           0               0                  9
## 20       592           0               0                  7
## 38      1285           0               0                 17
##    REMOTE_APP_PACKETS SOURCE_APP_BYTES REMOTE_APP_BYTES   DURATION
## 4                  13             2095             1092 4294930697
## 5                  20            13375             2717 4294930697
## 14                 12             2242             1018 4294935825
## 19                  9             1026              916 4294940333
## 20                  9             1062              732 4294940333
## 38                 18             2462             1741 4294946623
##    AVG_LOCAL_PKT_RATE AVG_REMOTE_PKT_RATE APP_PACKETS DNS_QUERY_TIMES
## 4                 256                 303          11               6
## 5                 466                 466          20               4
## 14                256                 279          11               2
## 19                210                 210           9               6
## 20                163                 210           7               2
## 38                396                 419          17               6
```

```r
#TCP CONVERSATION EXCHANGE MALIGNA
```

```r
N1_M = mean(matrizCRM$TCP_CONVERSATION_EXCHANGE)
N1_M
```

```
## [1] 22.47028
```

```r
#DIST REMOTE TCP PORT MALIGNA

N2_M=mean(matrizCRM$DIST_REMOTE_TCP_PORT)
N2_M
```

```
## [1] 2.478036
```

```r
#REMOTE IPS MALIGNA

N3_M = mean(matrizCRM$REMOTE_IPS)
N3_M
```

```
## [1] 4.093023
```

```r
#APP BYTES MALIGNA

N4_M = mean(matrizCRM$APP_BYTES)
N4_M
```

```
## [1] 2691.38
```

```r
#UDP PACKETS MALIGNA

N5_M = mean(matrizCRM$UDP_PACKETS)
N5_M
```

```
## [1] 0
```

```r
#TCP_URG_PACKETS MALIGNA

N6_M = mean(matrizCRM$TCP_URG_PACKETS)
N6_M
```

```
## [1] 0
```

```r
#SOURCE APP PACKETS MALIGNA

N7_M = mean(matrizCRM$SOURCE_APP_PACKETS)
N7_M
```

```
## [1] 27.66408
```

```r
#REMOTE APP PACKETS MALIGNA

N8_M = mean(matrizCRM$REMOTE_APP_PACKETS)
N8_M
```

```
## [1] 25.42636
```

```r
#SOURCE APP BYTES MALIGNA

N9_M = mean(matrizCRM$SOURCE_APP_BYTES)
N9_M
```

```
## [1] 11117.65
```

```r
#REMOTE APP BYTES MALIGNA

N10_M = mean(matrizCRM$REMOTE_APP_BYTES)
N10_M
```

```
## [1] 3088.233
```

```r
#DURATION MALIGNA

N11_M = mean(matrizCRM$DURATION)
N11_M
```

```
## [1] 190370459
```

```r
#AVG LOCAL PKT RATE MALIGNA

N12_M = mean(as.numeric(matrizCRM$AVG_LOCAL_PKT_RATE))
N12_M
```

```
## [1] 139.1886
```

```r
#AVG REMOTE PKT RATE MALIGNA

N13_M = mean(as.numeric(matrizCRM$AVG_REMOTE_PKT_RATE))
N13_M
```

```
## [1] 130.5891
```

```r
#APP PACKETS MALIGNA

N14_M = mean(matrizCRM$APP_PACKETS)
N14_M
```

```
## [1] 27.66408
```

```r
#DNS QUERY TIMES MALIGNA

N15_M = mean(matrizCRM$DNS_QUERY_TIMES)
N15_M
```

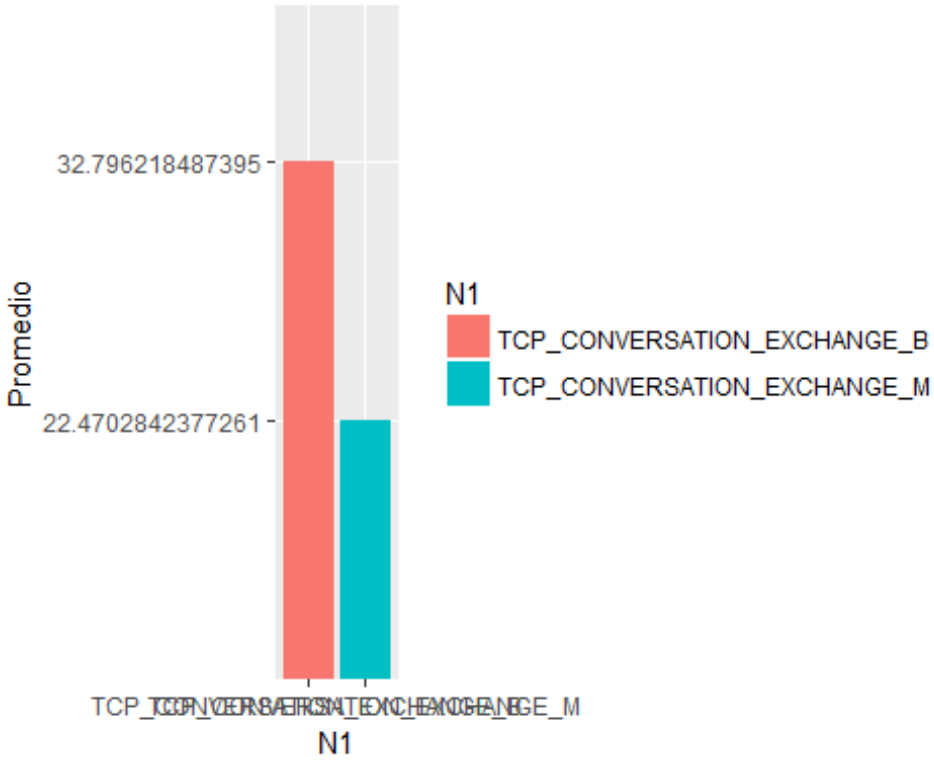
```
## [1] 5.193798
```

Comparacion entre la media de datos malignos y benginos de la capa red

```
#N1 que indica la caracteristica TCP_CONVERSATION_EXCHANGE

NAMES_N1 = rbind("TCP_CONVERSATION_EXCHANGE_M",
"TCP_CONVERSATION_EXCHANGE_B")
PROMEDIO_N1 = rbind(N1_M, N1_B)
TCP_CONVERSATION_EXCHANGE_N1= cbind(NAMES_N1, PROMEDIO_N1)
TCP_CONVERSATION_EXCHANGE_N1 <- data.frame(TCP_CONVERSATION_EXCHANGE_N1)
names(TCP_CONVERSATION_EXCHANGE_N1) <- c("N1", "Promedio")

TCP_CONVERSATION_EXCHANGE_N1
```

```
##                                   N1         Promedio
## N1_M TCP_CONVERSATION_EXCHANGE_M 22.4702842377261
## N1_B TCP_CONVERSATION_EXCHANGE_B  32.796218487395
```

```
ggplot(data=TCP_CONVERSATION_EXCHANGE_N1, aes(x=N1, y=Promedio, fill=N1))
+
    geom_bar(stat="identity", position=position_dodge())
```

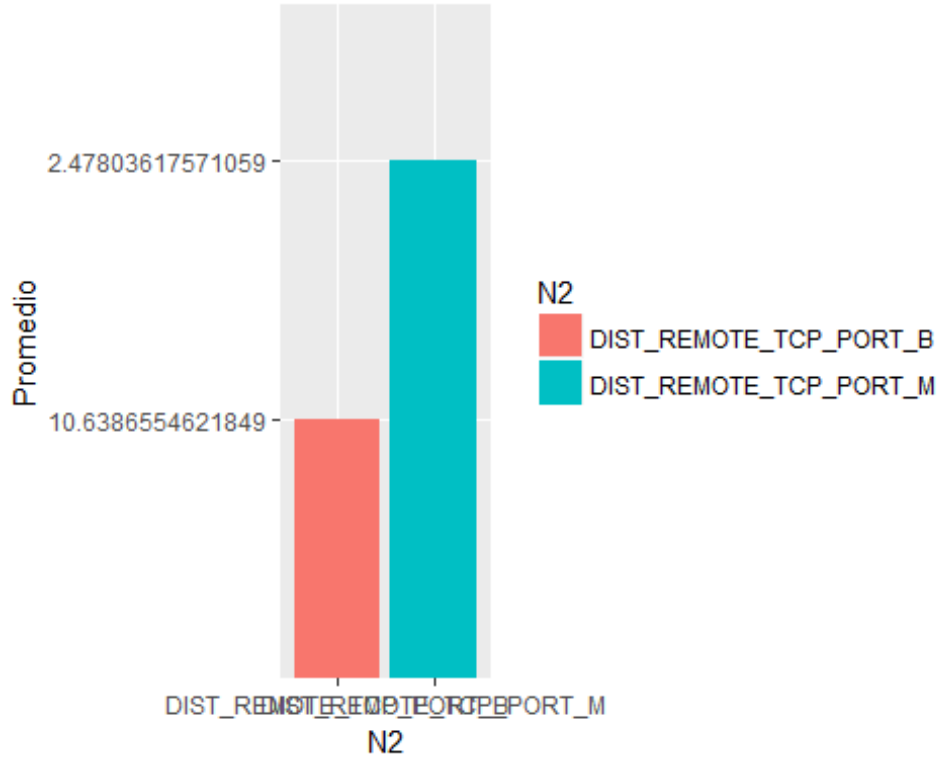



```
#N2 que indica la caracteristica DIST_REMOTE_TCP_PORT

NAMES_N2 = rbind("DIST_REMOTE_TCP_PORT_M", "DIST_REMOTE_TCP_PORT_B")
PROMEDIO_N2 = rbind(N2_M, N2_B)
DIST_REMOTE_TCP_PORT_N2= cbind(NAMES_N2, PROMEDIO_N2)
```

```
DIST_REMOTE_TCP_PORT_N2 <- data.frame(DIST_REMOTE_TCP_PORT_N2)
names(DIST_REMOTE_TCP_PORT_N2) <- c("N2", "Promedio")

DIST_REMOTE_TCP_PORT_N2
```

```
##                            N2         Promedio
## N2_M DIST_REMOTE_TCP_PORT_M 2.47803617571059
## N2_B DIST_REMOTE_TCP_PORT_B 10.6386554621849
```

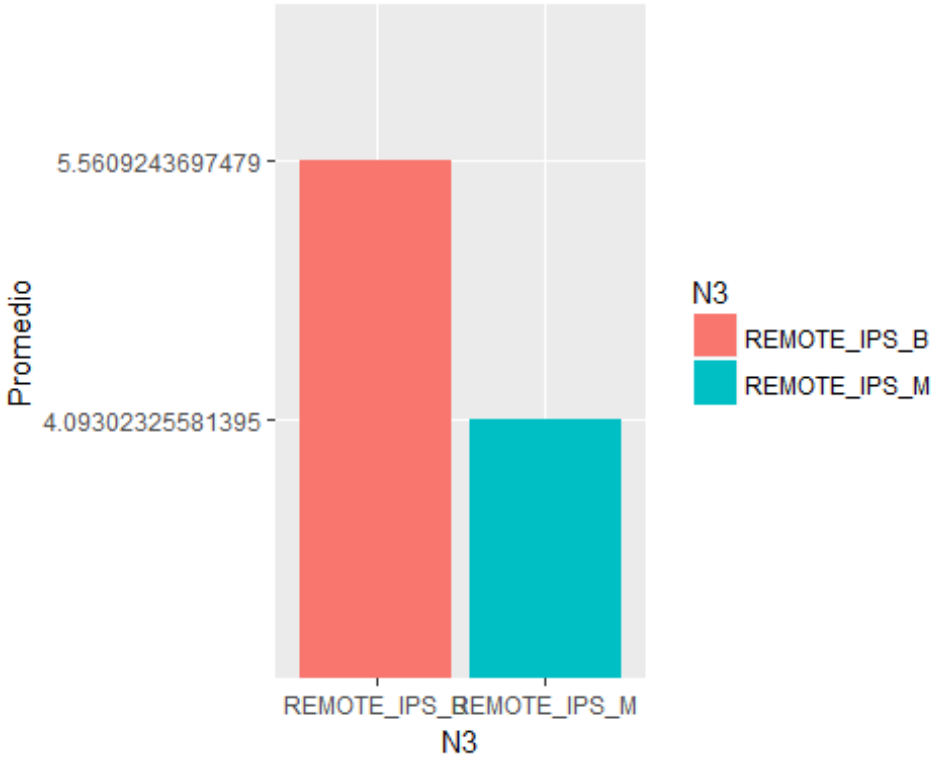
```
ggplot(data=DIST_REMOTE_TCP_PORT_N2, aes(x=N2, y=Promedio, fill=N2)) +
    geom_bar(stat="identity", position=position_dodge())
```



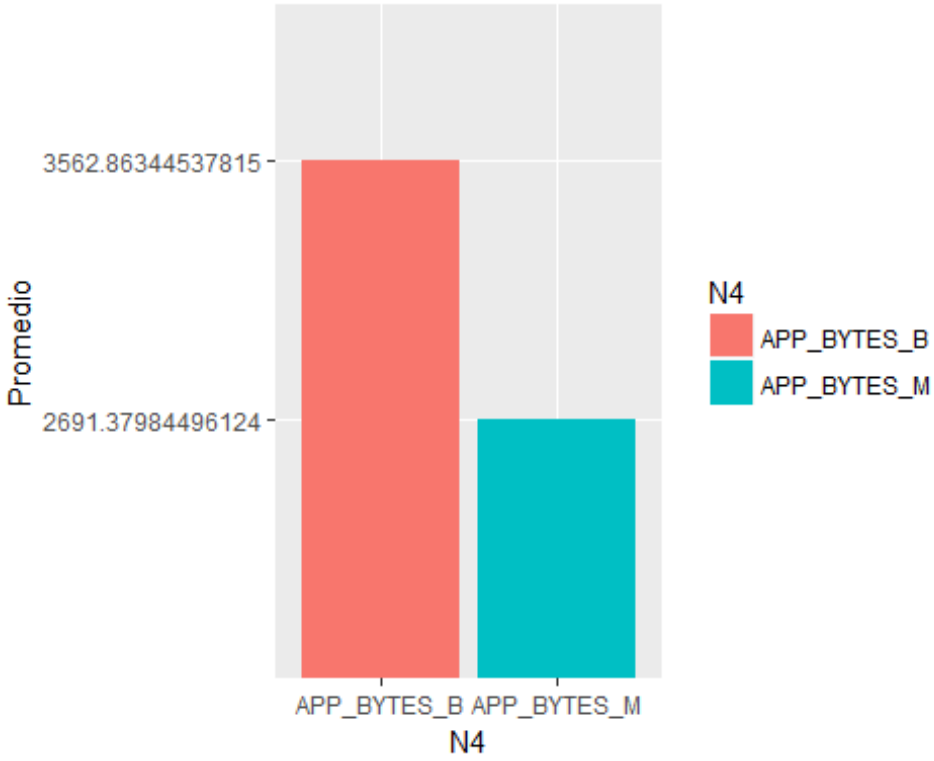

```
#N3 que indica la caracteristica REMOTE_IPS

NAMES_N3 = rbind("REMOTE_IPS_M", "REMOTE_IPS_B")
PROMEDIO_N3 = rbind(N3_M, N3_B)
REMOTE_IPS_N3= cbind(NAMES_N3, PROMEDIO_N3)
REMOTE_IPS_N3 <- data.frame(REMOTE_IPS_N3)
names(REMOTE_IPS_N3) <- c("N3", "Promedio")

REMOTE_IPS_N3
```

```
##                 N3         Promedio
## N3_M REMOTE_IPS_M 4.09302325581395
## N3_B REMOTE_IPS_B  5.5609243697479
```

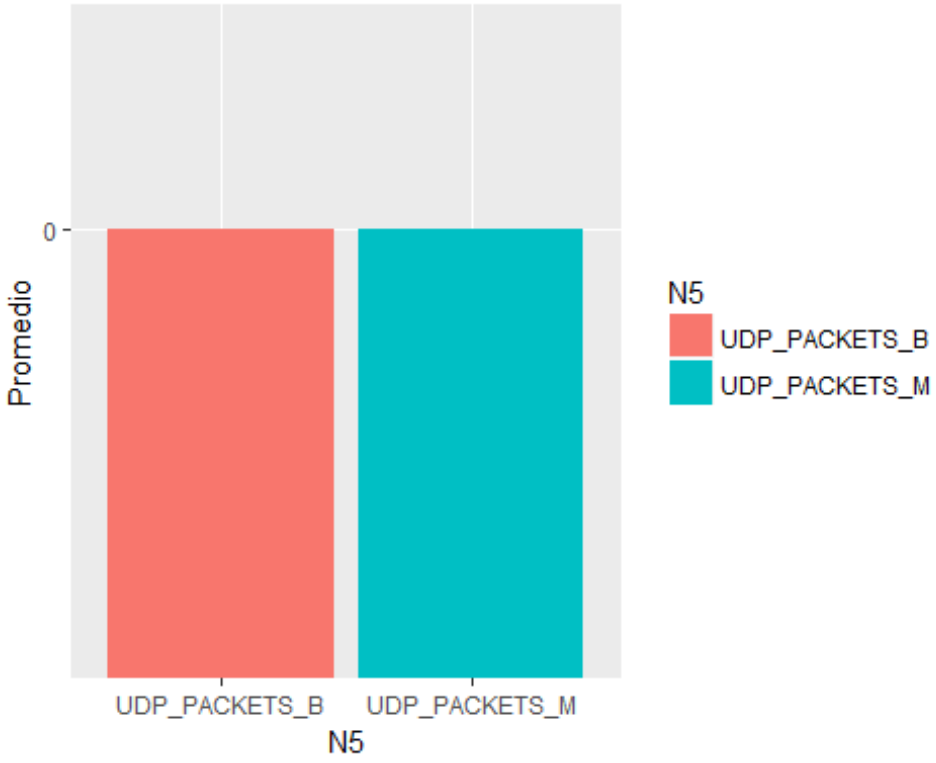
```
ggplot(data=REMOTE_IPS_N3, aes(x=N3, y=Promedio, fill=N3)) +
    geom_bar(stat="identity", position=position_dodge())
```



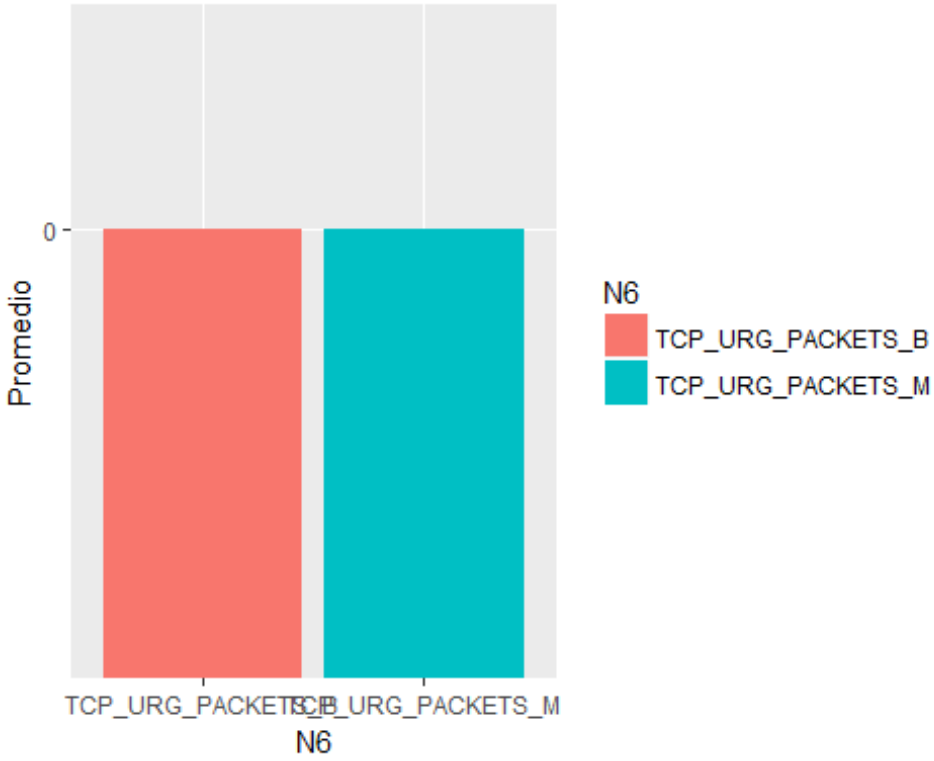

```
#N4 que indica la caracteristica APP_BYTES

NAMES_N4 = rbind("APP_BYTES_M", "APP_BYTES_B")
PROMEDIO_N4 = rbind(N4_M, N4_B)
APP_BYTES_N4= cbind(NAMES_N4, PROMEDIO_N4)
APP_BYTES_N4 <- data.frame(APP_BYTES_N4)
names(APP_BYTES_N4) <- c("N4", "Promedio")

APP_BYTES_N4

##                N4         Promedio
## N4_M APP_BYTES_M 2691.37984496124
## N4_B APP_BYTES_B 3562.86344537815

ggplot(data=APP_BYTES_N4, aes(x=N4, y=Promedio, fill=N4)) +
    geom_bar(stat="identity", position=position_dodge())
```



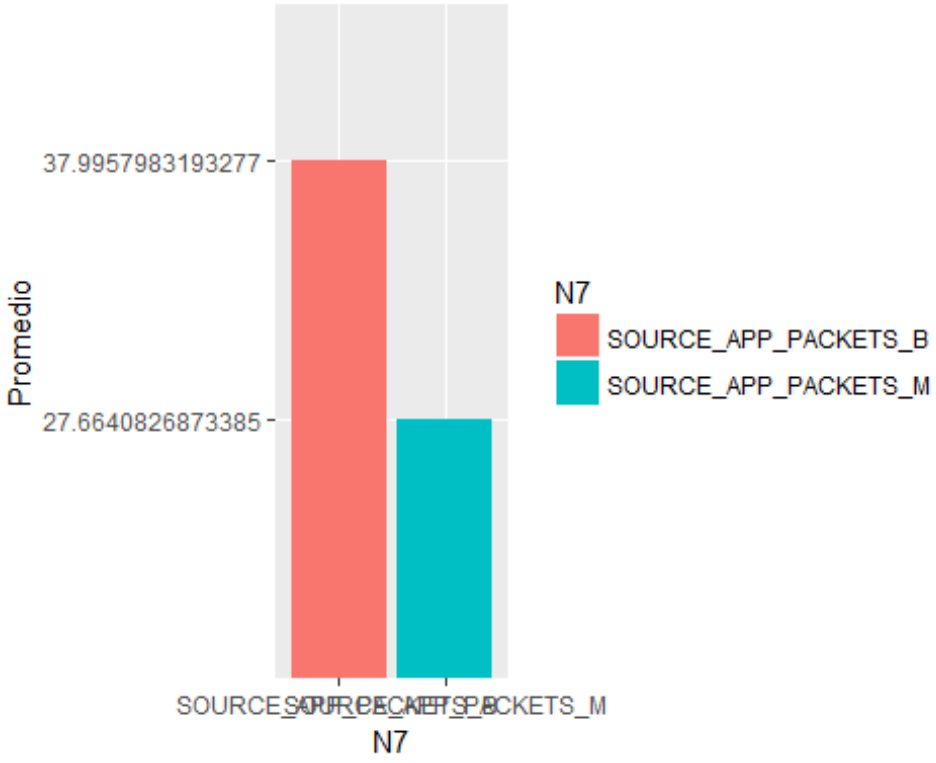

```
#N5 que indica la caracteristica UDP_PACKETS

NAMES_N5 = rbind("UDP_PACKETS_M", "UDP_PACKETS_B")
PROMEDIO_N5 = rbind(N5_M, N5_B)
UDP_PACKETS_N5= cbind(NAMES_N5, PROMEDIO_N5)
UDP_PACKETS_N5 <- data.frame(UDP_PACKETS_N5)
names(UDP_PACKETS_N5) <- c("N5", "Promedio")

UDP_PACKETS_N5

##                   N5 Promedio
## N5_M UDP_PACKETS_M        0
## N5_B UDP_PACKETS_B        0

ggplot(data=UDP_PACKETS_N5, aes(x=N5, y=Promedio, fill=N5)) +
    geom_bar(stat="identity", position=position_dodge())
```



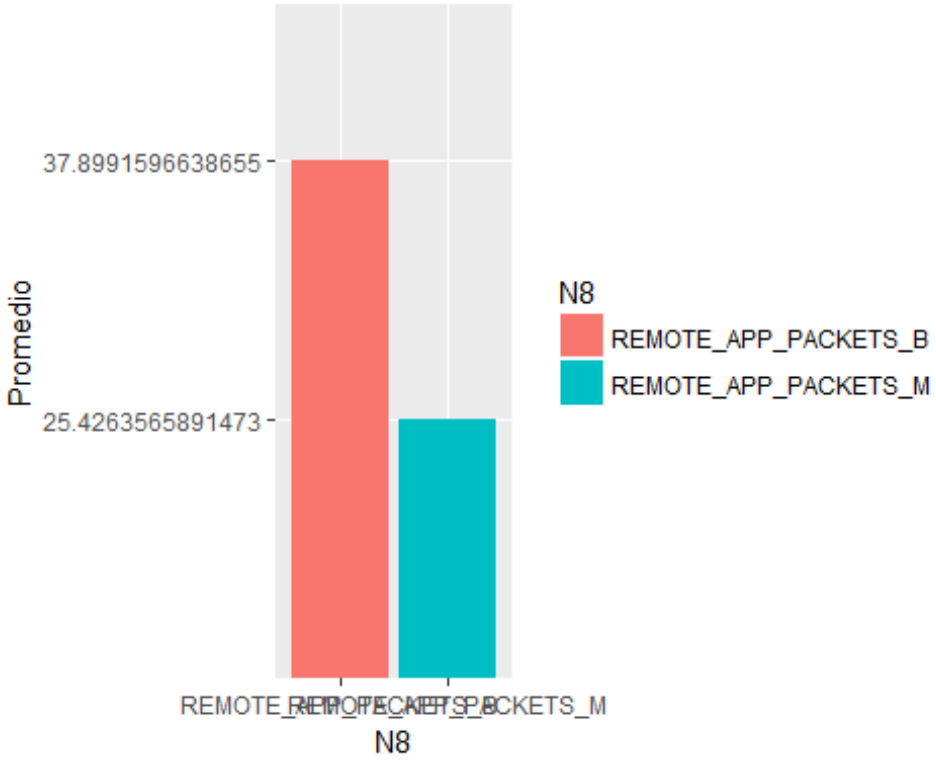

```
#N6 que indica la caracteristica TCP_URG_PACKETS

NAMES_N6 = rbind("TCP_URG_PACKETS_M", "TCP_URG_PACKETS_B")
PROMEDIO_N6 = rbind(N6_M, N6_B)
TCP_URG_PACKETS_N6= cbind(NAMES_N6, PROMEDIO_N6)
TCP_URG_PACKETS_N6 <- data.frame(TCP_URG_PACKETS_N6)
names(TCP_URG_PACKETS_N6) <- c("N6", "Promedio")

TCP_URG_PACKETS_N6

##                       N6 Promedio
## N6_M TCP_URG_PACKETS_M        0
## N6_B TCP_URG_PACKETS_B        0

ggplot(data=TCP_URG_PACKETS_N6, aes(x=N6, y=Promedio, fill=N6)) +
    geom_bar(stat="identity", position=position_dodge())
```




```
#N7 que indica la caracteristica SOURCE_APP_PACKETS

NAMES_N7 = rbind("SOURCE_APP_PACKETS_M", "SOURCE_APP_PACKETS_B")
PROMEDIO_N7 = rbind(N7_M, N7_B)
SOURCE_APP_PACKETS_N7= cbind(NAMES_N7, PROMEDIO_N7)
SOURCE_APP_PACKETS_N7 <- data.frame(SOURCE_APP_PACKETS_N7)
names(SOURCE_APP_PACKETS_N7) <- c("N7", "Promedio")

SOURCE_APP_PACKETS_N7

##                          N7         Promedio
## N7_M SOURCE_APP_PACKETS_M 27.6640826873385
## N7_B SOURCE_APP_PACKETS_B 37.9957983193277

ggplot(data=SOURCE_APP_PACKETS_N7, aes(x=N7, y=Promedio, fill=N7)) +
    geom_bar(stat="identity", position=position_dodge())
```

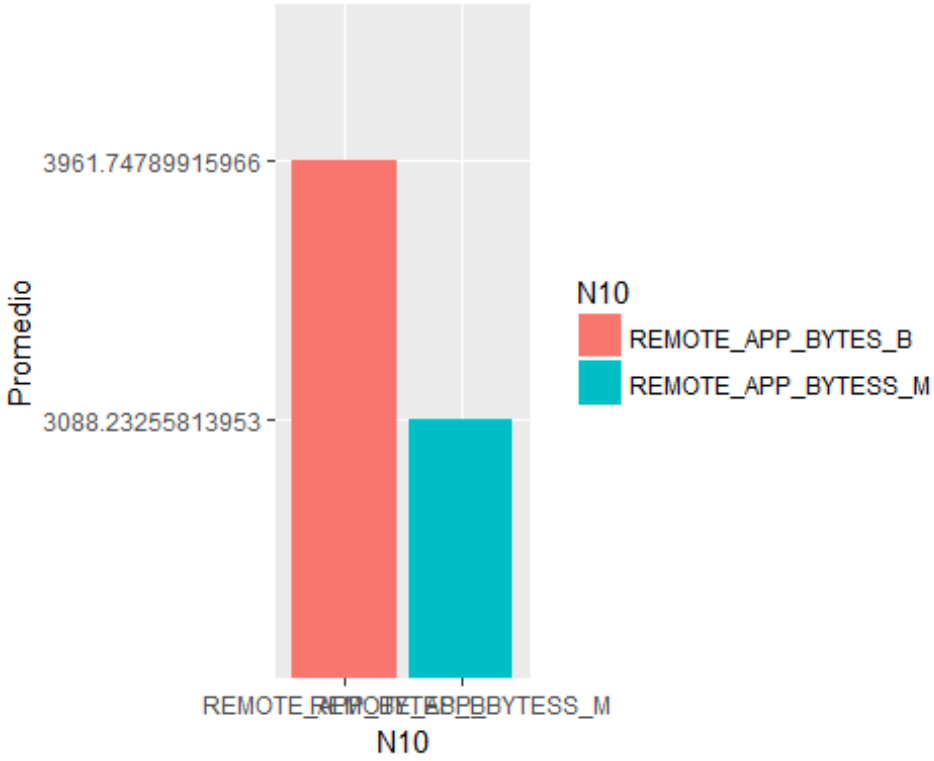



```
#N8 que indica la caracteristica REMOTE_APP_PACKETS

NAMES_N8 = rbind("REMOTE_APP_PACKETS_M", "REMOTE_APP_PACKETS_B")
PROMEDIO_N8 = rbind(N8_M, N8_B)
REMOTE_APP_PACKETS_N8= cbind(NAMES_N8, PROMEDIO_N8)
REMOTE_APP_PACKETS_N8 <- data.frame(REMOTE_APP_PACKETS_N8)
names(REMOTE_APP_PACKETS_N8) <- c("N8", "Promedio")

REMOTE_APP_PACKETS_N8

##                          N8         Promedio
## N8_M REMOTE_APP_PACKETS_M 25.4263565891473
## N8_B REMOTE_APP_PACKETS_B 37.8991596638655

ggplot(data=REMOTE_APP_PACKETS_N8, aes(x=N8, y=Promedio, fill=N8)) +
    geom_bar(stat="identity", position=position_dodge())
```

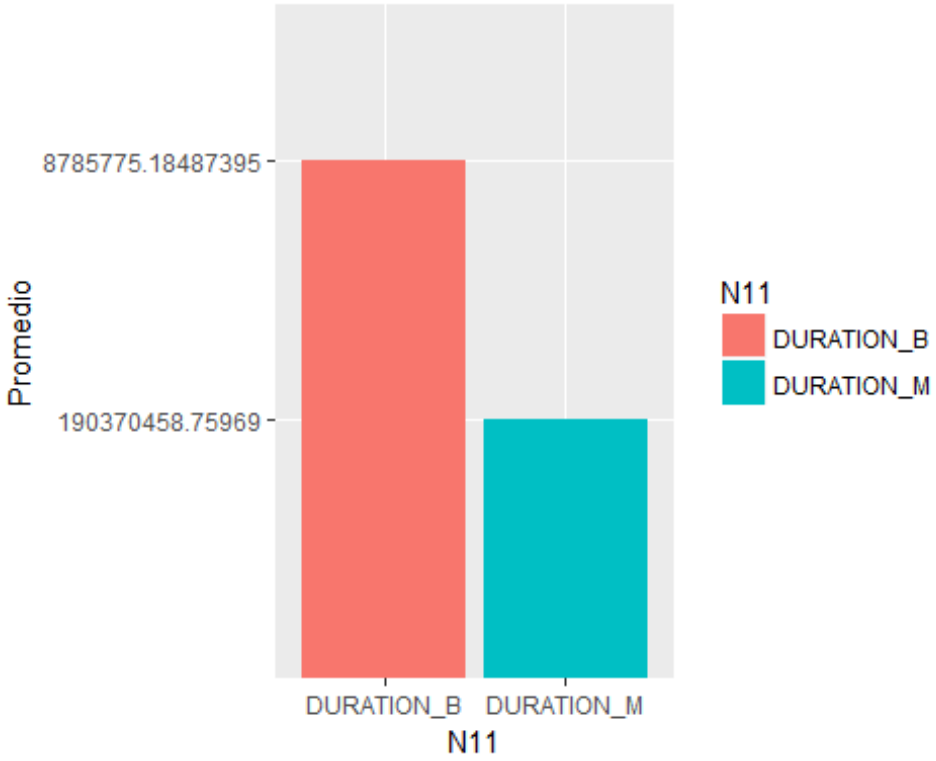



```
#N9 que indica la caracteristica SOURCE_APP_BYTES

NAMES_N9 = rbind("SOURCE_APP_BYTES_M", "SOURCE_APP_BYTES_B")
PROMEDIO_N9 = rbind(N9_M, N9_B)
SOURCE_APP_BYTES_N9= cbind(NAMES_N9, PROMEDIO_N9)
SOURCE_APP_BYTES_N9 <- data.frame(SOURCE_APP_BYTES_N9)
names(SOURCE_APP_BYTES_N9) <- c("N9", "Promedio")

SOURCE_APP_BYTES_N9

##                        N9         Promedio
## N9_M SOURCE_APP_BYTES_M   11117.65374677
## N9_B SOURCE_APP_BYTES_B 35038.5441176471

ggplot(data=SOURCE_APP_BYTES_N9, aes(x=N9, y=Promedio, fill=N9)) +
    geom_bar(stat="identity", position=position_dodge())
```

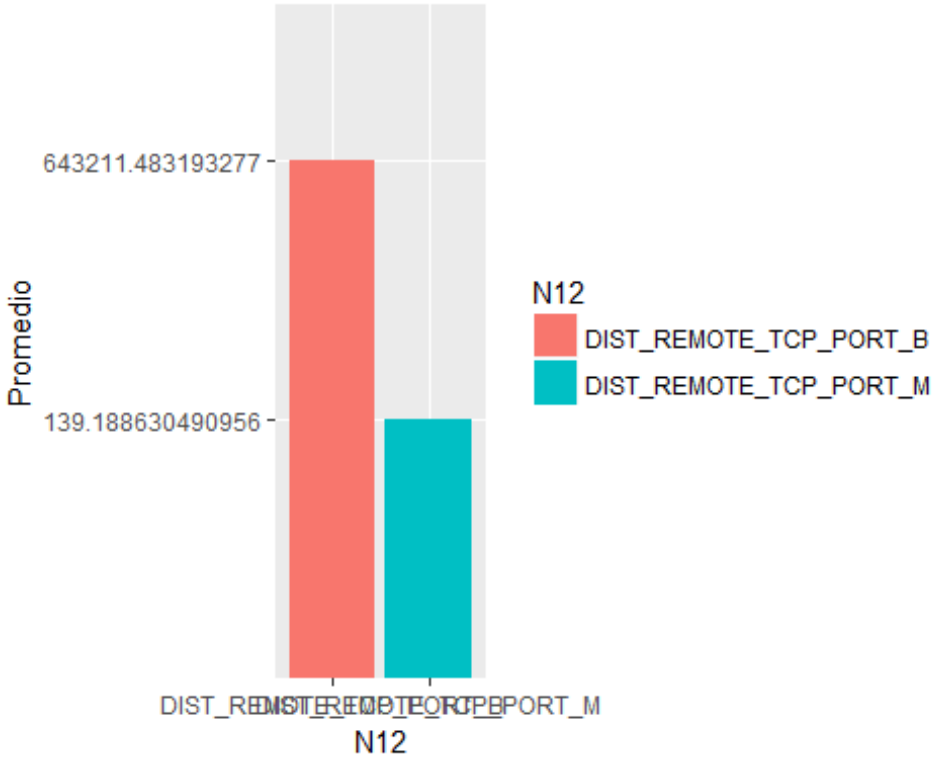

```
#N10 que indica la caracteristica REMOTE_APP_BYTES

NAMES_N10 = rbind("REMOTE_APP_BYTESS_M", "REMOTE_APP_BYTES_B")
PROMEDIO_N10 = rbind(N10_M, N10_B)
REMOTE_APP_BYTES_N10= cbind(NAMES_N10, PROMEDIO_N10)
REMOTE_APP_BYTES_N10 <- data.frame(REMOTE_APP_BYTES_N10)
names(REMOTE_APP_BYTES_N10) <- c("N10", "Promedio")

REMOTE_APP_BYTES_N10

##                         N10         Promedio
## N10_M REMOTE_APP_BYTESS_M 3088.23255813953
## N10_B  REMOTE_APP_BYTES_B 3961.74789915966

ggplot(data=REMOTE_APP_BYTES_N10, aes(x=N10, y=Promedio, fill=N10)) +
    geom_bar(stat="identity", position=position_dodge())
```




```
#N11 que indica la caracteristica DURATION

NAMES_N11 = rbind("DURATION_M", "DURATION_B")
PROMEDIO_N11 = rbind(N11_M, N11_B)
DURATION_N11= cbind(NAMES_N11, PROMEDIO_N11)
DURATION_N11 <- data.frame(DURATION_N11)
names(DURATION_N11) <- c("N11", "Promedio")

DURATION_N11

##              N11         Promedio
## N11_M DURATION_M  190370458.75969
## N11_B DURATION_B 8785775.18487395

ggplot(data=DURATION_N11, aes(x=N11, y=Promedio, fill=N11)) +
    geom_bar(stat="identity", position=position_dodge())
```



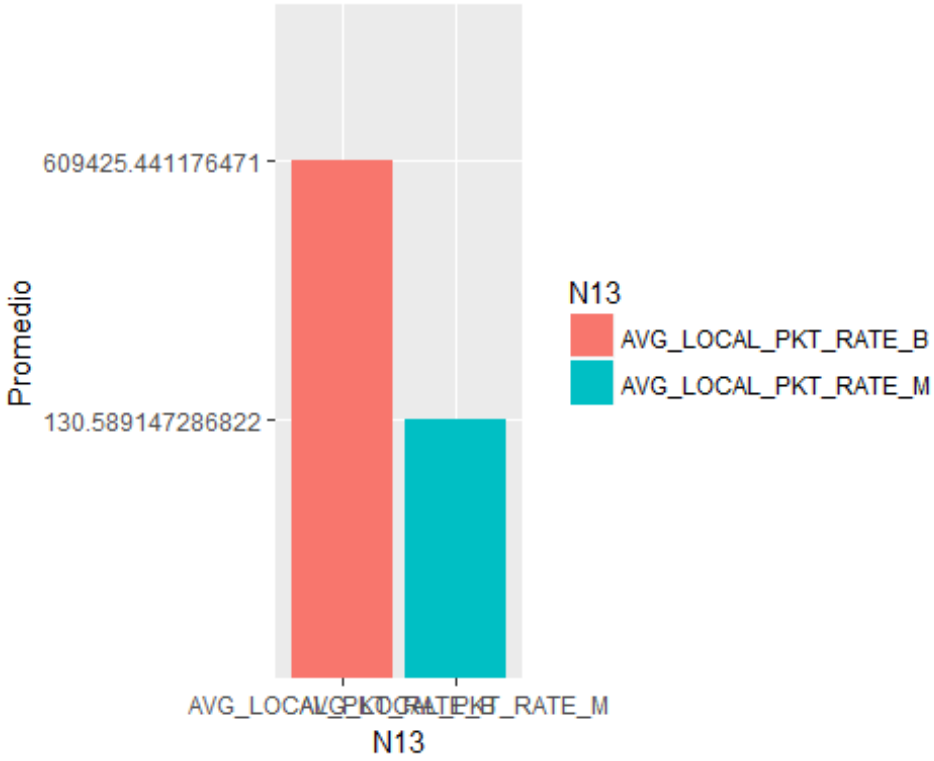

```
#N12 que indica la caracteristica DIST REMOTE TCP PORT

NAMES_N12 = rbind("DIST_REMOTE_TCP_PORT_M", "DIST_REMOTE_TCP_PORT_B")
PROMEDIO_N12 = rbind(N12_M, N12_B)
DIST_REMOTE_TCP_PORT_N12= cbind(NAMES_N12, PROMEDIO_N12)
DIST_REMOTE_TCP_PORT_N12 <- data.frame(DIST_REMOTE_TCP_PORT_N12)
names(DIST_REMOTE_TCP_PORT_N12) <- c("N12", "Promedio")

DIST_REMOTE_TCP_PORT_N12

##                            N12         Promedio
## N12_M DIST_REMOTE_TCP_PORT_M 139.188630490956
## N12_B DIST_REMOTE_TCP_PORT_B 643211.483193277

ggplot(data=DIST_REMOTE_TCP_PORT_N12, aes(x=N12, y=Promedio, fill=N12)) +
    geom_bar(stat="identity", position=position_dodge())
```



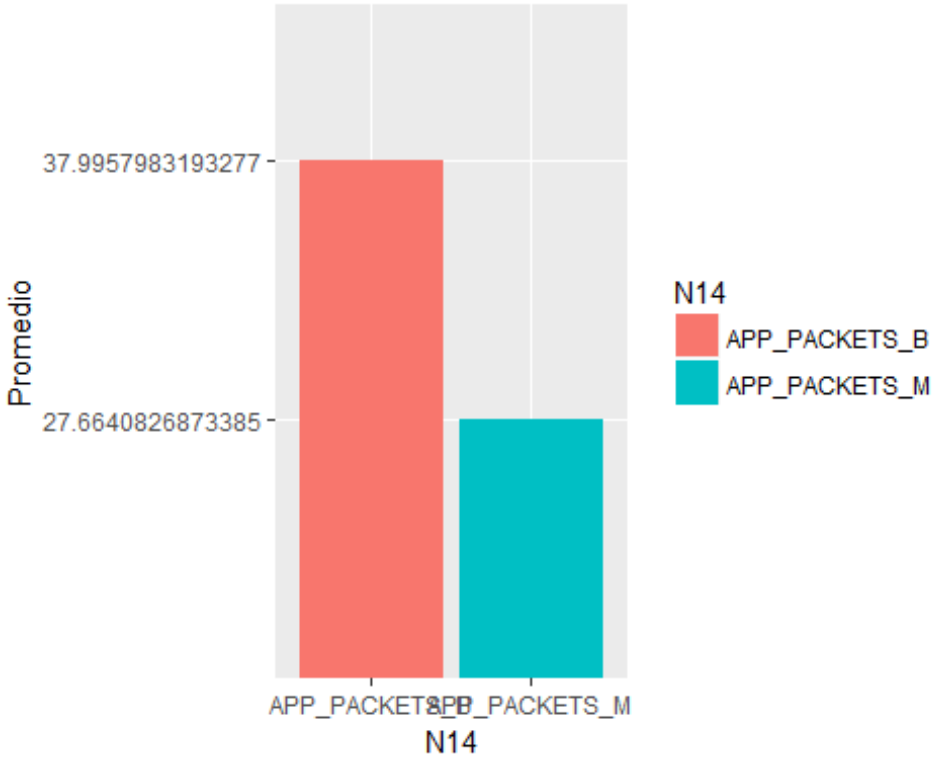

```
#N13 que indica la caracteristica AVG_LOCAL_PKT_RATE


NAMES_N13 = rbind("AVG_LOCAL_PKT_RATE_M", "AVG_LOCAL_PKT_RATE_B")
PROMEDIO_N13 = rbind(N13_M, N13_B)
AVG_LOCAL_PKT_RATE_N13= cbind(NAMES_N13, PROMEDIO_N13)
AVG_LOCAL_PKT_RATE_N13 <- data.frame(AVG_LOCAL_PKT_RATE_N13)
names(AVG_LOCAL_PKT_RATE_N13) <- c("N13", "Promedio")

AVG_LOCAL_PKT_RATE_N13
```

```
##                         N13         Promedio
## N13_M AVG_LOCAL_PKT_RATE_M 130.589147286822
## N13_B AVG_LOCAL_PKT_RATE_B 609425.441176471
```

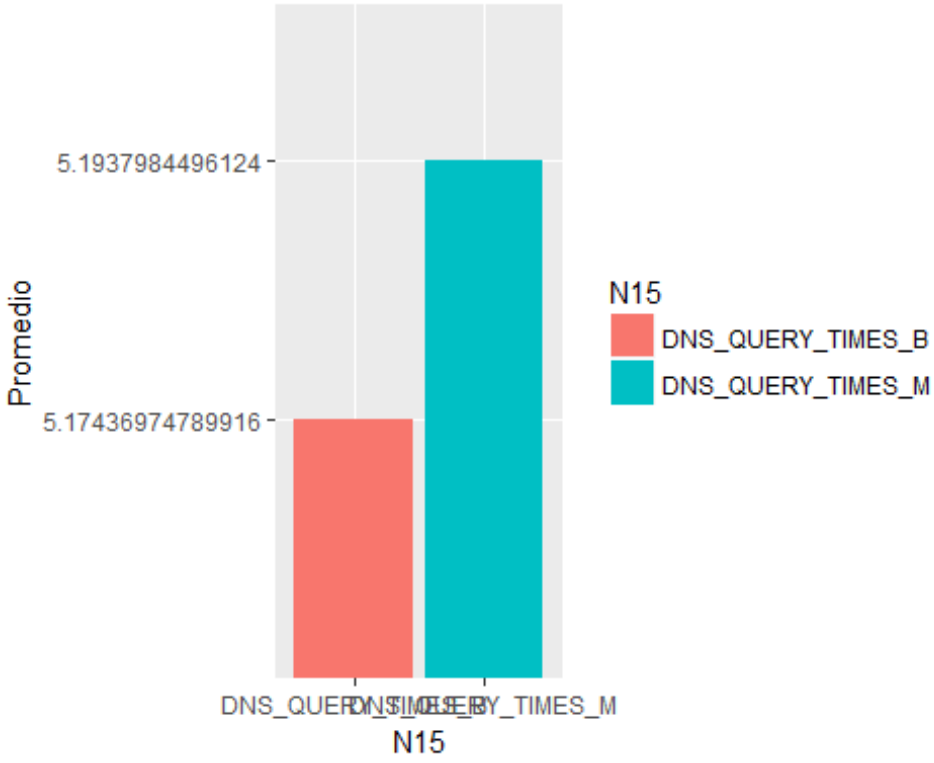
```
ggplot(data=AVG_LOCAL_PKT_RATE_N13, aes(x=N13, y=Promedio, fill=N13)) +
    geom_bar(stat="identity", position=position_dodge())
```




```
#N14 que indica la caracteristica APP_PACKETS

NAMES_N14 = rbind("APP_PACKETS_M", "APP_PACKETS_B")
PROMEDIO_N14 = rbind(N14_M, N14_B)
APP_PACKETS_N14= cbind(NAMES_N14, PROMEDIO_N14)
APP_PACKETS_N14 <- data.frame(APP_PACKETS_N14)
names(APP_PACKETS_N14) <- c("N14", "Promedio")

APP_PACKETS_N14

##                 N14         Promedio
## N14_M APP_PACKETS_M 27.6640826873385
## N14_B APP_PACKETS_B 37.9957983193277

ggplot(data=APP_PACKETS_N14, aes(x=N14, y=Promedio, fill=N14)) +
    geom_bar(stat="identity", position=position_dodge())
```




```
#N15 que indica la caracteristica DNS_QUERY_TIMES


NAMES_N15 = rbind("DNS_QUERY_TIMES_M", "DNS_QUERY_TIMES_B")
PROMEDIO_N15 = rbind(N15_M, N15_B)
DNS_QUERY_TIMES_N15= cbind(NAMES_N15, PROMEDIO_N15)
DNS_QUERY_TIMES_N15 <- data.frame(DNS_QUERY_TIMES_N15)
names(DNS_QUERY_TIMES_N15) <- c("N15", "Promedio")

DNS_QUERY_TIMES_N15

##                       N15         Promedio
## N15_M DNS_QUERY_TIMES_M  5.1937984496124
## N15_B DNS_QUERY_TIMES_B 5.17436974789916

ggplot(data=DNS_QUERY_TIMES_N15, aes(x=N15, y=Promedio, fill=N15)) +
    geom_bar(stat="identity", position=position_dodge())
```


5.1937984496124
5.17436974789916
Promedio
N15
DNS_QUERY_TIMES_B
DNS_QUERY_TIMES_M
N15