

Tarea 1

Marái José Meisel - Juan Pablo Dulcey

15/11/2021

Incendios en Idaho

Cree un nuevo dataframe que sea un subconjunto del dataframe original de dfFires. El subconjunto debe contener todos los incendios del Estado de Idaho y las columnas deben ser limitadas para que sólo estén presentes las columnas YEAR_, CAUSE y TOTALACRES. Cambie el nombre de las columnas. Agrupe los datos por CAUSE y YEAR_ y luego resuma por el total de acres quemados. Trazar los resultado

Descarga de base de datos

```
df = read.table("RDataSets/StudyArea.csv", header=TRUE, fill=TRUE, quote="", sep=",")
names(df)
```

```
## [1] "FID"      "ORGANIZATI" "UNIT"      "SUBUNIT"    "SUBUNIT2"
## [6] "FIRENAME" "CAUSE"      "YEAR_"     "STARTDATED" "CONTRDATED"
## [11] "OUTDATED" "STATE"     "STATE_FIPS" "TOTALACRES"
```

Flitramos la base de datos por Idaho y reducimos las variables a 3

```
dfFires <- filter(df, STATE %in% c("Idaho"))
dfFires <- select(dfFires, YEAR_, CAUSE, TOTALACRES)
```

Cambiamos el nombre de las columnas

```
dfFires <- rename(dfFires, Año=YEAR_ , Causa=CAUSE , Acres_Quemados=TOTALACRES)
head(dfFires)
```

```
##   Año   Causa Acres_Quemados
## 1 1987   Human             5
## 2 1991 Natural          150
## 3 1991   Human          800
## 4 1990 Natural             2
## 5 1985   Human          38
## 6 1988   Human             2
```

Detalles de los objetos

```
str(dfFires)
```

```
## 'data.frame':   36014 obs. of  3 variables:
##  $ Año          : chr  "1987" "1991" "1991" "1990" ...
##  $ Causa         : chr  "Human" "Natural" "Human" "Natural" ...
##  $ Acres_Quemados: num  5 150 800 2 38 2 0.2 150 15 30 ...
```

Comprobamos si existen vacios en Idaho.

```
sum(with(dfFires,Causa == " "))
```

```
## [1] 43
```

En fecto hay 43 datos, como no están determinados se sumaran a los “Unfetermined”

```
dfFires$Causa[dfFires$Causa == " "] <- "Undetermined"
```

Realizamos un descriptivo

```
summary(dfFires)
```

```
##      Año      Causa      Acres_Quemados
## Length:36014 Length:36014 Min.      :    0.0
## Class :character Class :character 1st Qu.:    0.1
## Mode  :character Mode  :character Median   :    0.2
##                                     Mean    :   481.3
##                                     3rd Qu.:    2.0
##                                     Max.    :590620.0
```

Agrupamos los datos en “Décadas”

```
dfFires <- mutate(dfFires, Decada = ifelse(Año %in% 1980:1989, "1980-1989",
                                           ifelse(Año %in% 1990:1999, "1990-1999",
                                           ifelse(Año %in% 2000:2009, "2000-2009",
                                           ifelse(Año %in% 2010:2016, "2010-2016", "-99")))))

knitr::kable(head(dfFires))
```

Año	Causa	Acres_Quemados	Decada
1987	Human	5	1980-1989
1991	Natural	150	1990-1999
1991	Human	800	1990-1999
1990	Natural	2	1990-1999
1985	Human	38	1980-1989
1988	Human	2	1980-1989

```
grp = group_by(dfFires, Decada, Causa)
sm = summarize(grp, mean(Acres_Quemados))
```

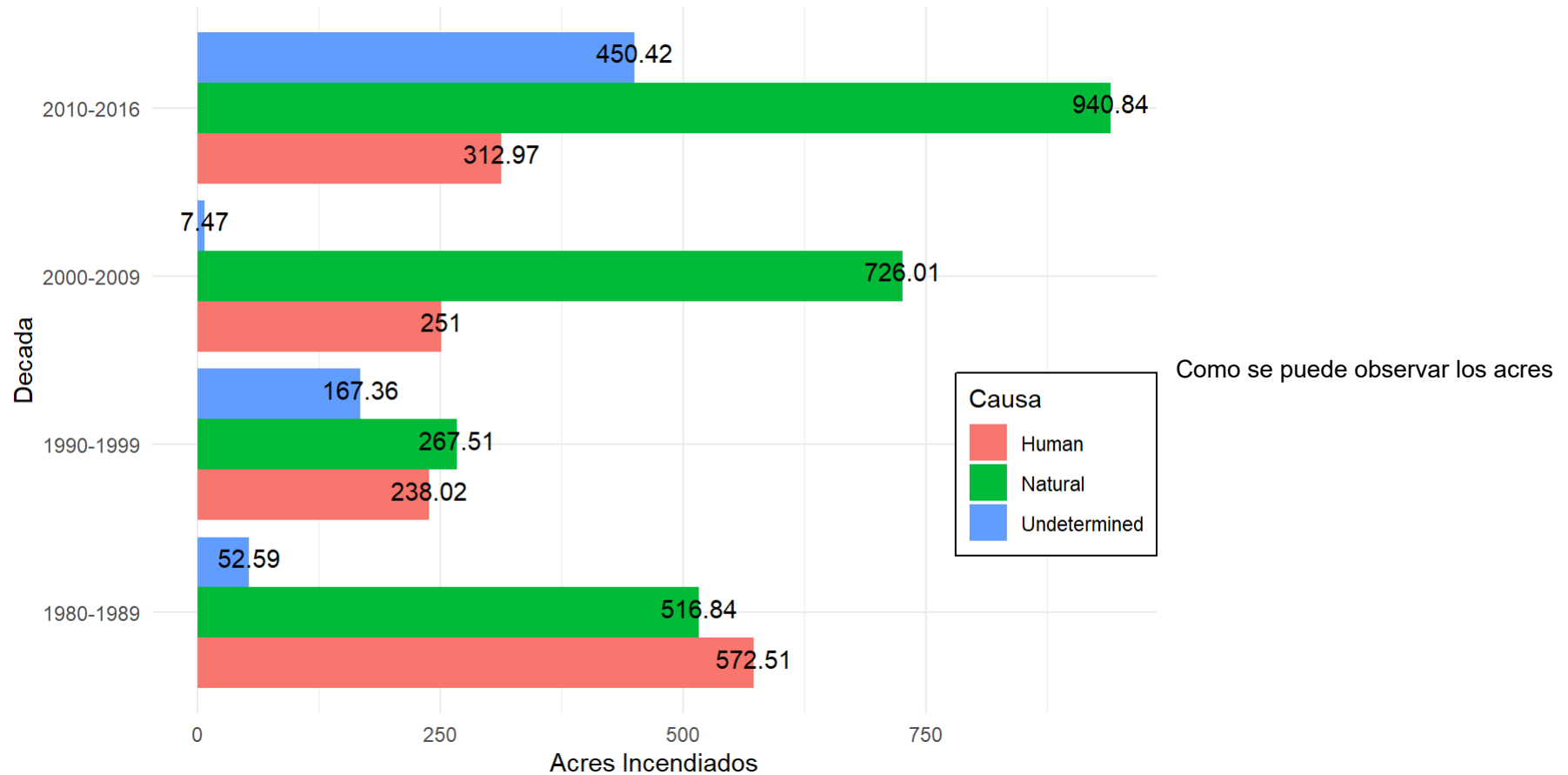
```
## `summarise()` has grouped output by 'Decada'. You can override using the `.groups` argument.
```

```
sm <- rename(sm, Decada=Decada , Causa=Causa , Acres_Quemados="mean(Acres_Quemados)")
```

Visualización gráfico

```
ggplot(data=sm,aes(x=Decada, y=Acres_Quemados,fill=
Causa))+ theme_minimal() +
  geom_bar(stat='identity', position='dodge') +
  geom_text(aes(label=round(Acres_Quemados,2)), position=position_dodge(width=0.9), vjust=0.25) +
  coord_flip() + scale_y_continuous(name="Acres Incendiados", labels = comma) +
  ggtitle ("Media de Acres Incendiados por Década según su Causa") +
  theme(
    legend.position = c(1, 1),
    legend.justification = c(1, 3),
    legend.background = element_rect(fill = "white", colour = "black"),
    plot.title = element_text(
      size = rel(1.2), lineheight = .9
    )
  )
```

Media de Acres Incendizados por Década según su Causa



incendiados por causas naturales han aumentado de forma significativa desde que se lleva el registro. El cambio climático está impactando de forma importante doblar la media de acres cremados desde 1980-1989 hasta 2010-2016.

En la primera década se presentó un registro importante de media de acres quemados y apartir de ahí empezaron a dismunir, sin embargo, el último registro presentó un aumento importante.

Atletas

```
athe <- athe %>%  
  filter(!is.na(Medal)) %>% filter( !is.na(Sex)) %>% filter( !is.na(Age)) %>%  
  filter(!is.na(Height)) %>% filter(!is.na(Weight)) %>% filter(!is.na(Team)) %>%  
  filter(!is.na(NOC)) %>% filter(!is.na(Year)) %>% filter(!is.na(Season)) %>%  
  filter(!is.na(Sport))
```

Filtra el DataFrame para incluir solo las filas correspondientes a los ganadores de medallas de 2016.

```
athe<- athe %>% filter(Year == 2016)
```

Descubre las medallas concedidas en 2016 en cada deporte

```
sort(table(athe$Sport), decreasing = TRUE)
```

```
##
##      Swimming      Athletics      Rowing
##      190          188          144
##      Football      Hockey      Handball
##      105          99          89
##      Cycling      Canoeing      Water Polo
##      84          82          78
##      Rugby Sevens      Basketball      Volleyball
##      74          72          72
##      Wrestling      Gymnastics      Fencing
##      72          66          65
##      Judo          Boxing      Equestrianism
##      56          51          45
##      Sailing      Shooting      Weightlifting
##      45          45          45
##      Diving Synchronized Swimming      Taekwondo
##      36          32          32
##      Archery      Table Tennis      Tennis
##      24          24          24
##      Badminton      Rhythmic Gymnastics      Beach Volleyball
##      23          18          12
##      Modern Pentathlon      Trampoline      Triathlon
##      6          6          6
##      Golf
##      4
```

Enumera los cinco deportes más importantes en función del mayor número de medallas concedidas. Filtra el DataFrame una vez más para incluir solo los registros de los cinco deportes principales en 2016.

```
atdeport5 <- athe %>%
  filter(Sport %in% c("Swimming" , "Athletics", "Rowing", "Football", "Hockey"))
```

```
table(atdeport5$Medal,atdeport5$Sport)
```

```
##
##           Athletics Football Hockey Rowing Swimming
## Bronze      61         35      33      48         57
## Gold        65         35      34      48         70
## Silver      62         35      32      48         63
```

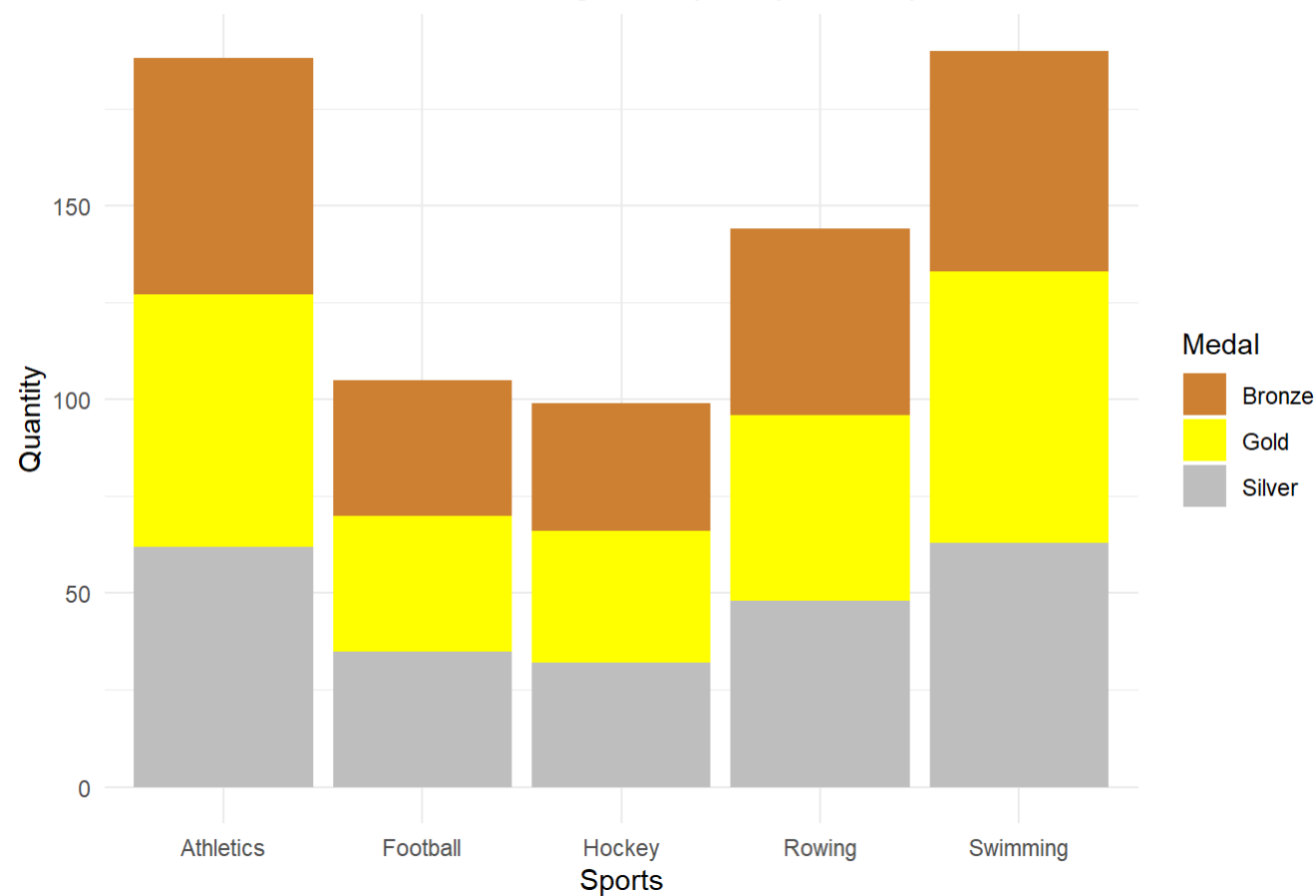
```
Principal_sports = data.frame(table(atdeport5$Medal,atdeport5$Sport))
Principal_sports <- rename(Principal_sports, Medal=Var1 , Sports=Var2 ,
Quantity=Freq)
head(Principal_sports)
```

```
##   Medal   Sports Quantity
## 1 Bronze Athletics      61
## 2  Gold Athletics      65
## 3 Silver Athletics      62
## 4 Bronze  Football      35
## 5  Gold  Football      35
## 6 Silver  Football      35
```

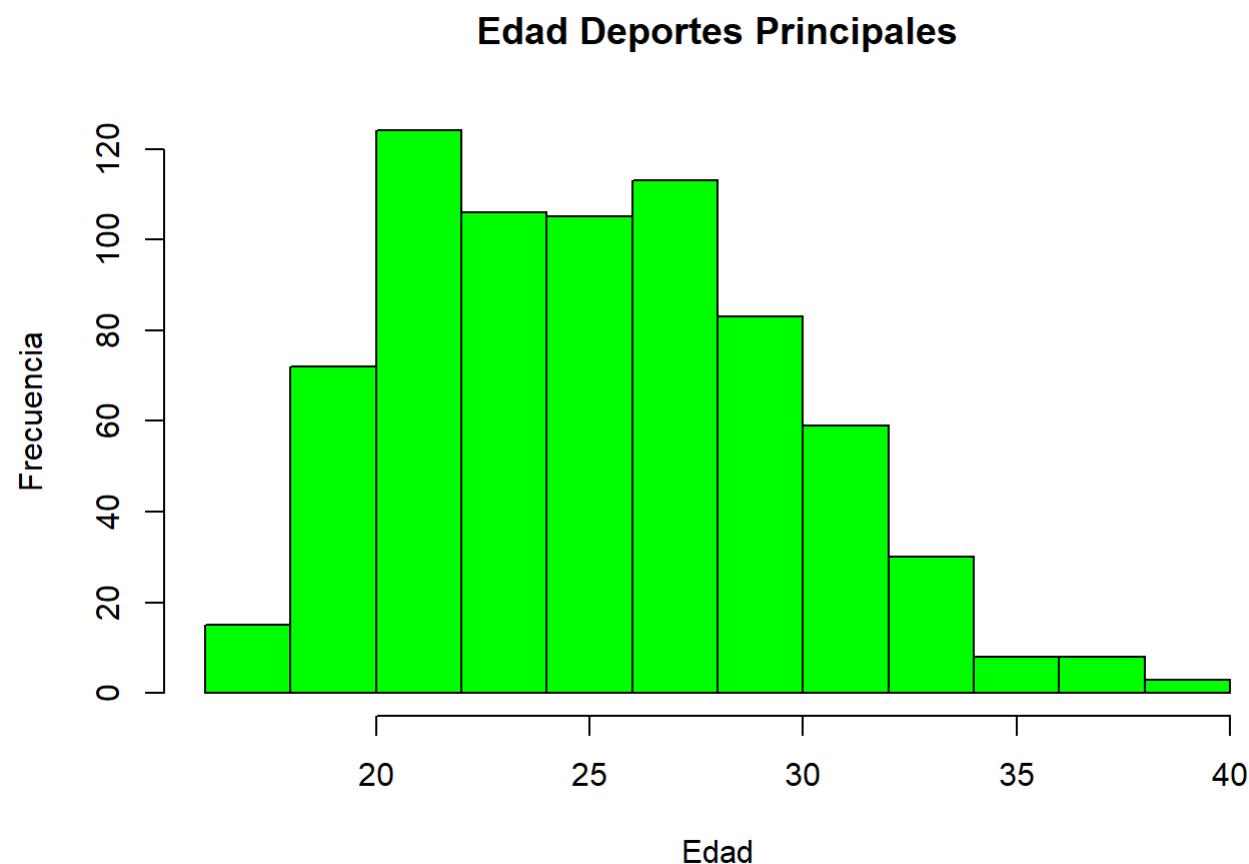
Genere un gráfico de barras con los recuentos de registros correspondientes a cada uno de los cinco deportes principales.

```
ggplot(data=Principal_sports, aes(x=Sports, y=Quantity, fill=Medal)) +
  geom_bar(stat="identity") +
  scale_fill_manual(values=c("#cd7f32", "yellow","gray"))+ theme_minimal() +
  ggtitle ("Distribución de las Medallas según los principales Deportes")
```


Distribución de las Medallas según los principales Deportes



Generar un histograma para la característica Edad de todos los ganadores de medallas en los cinco deportes principales (2016).



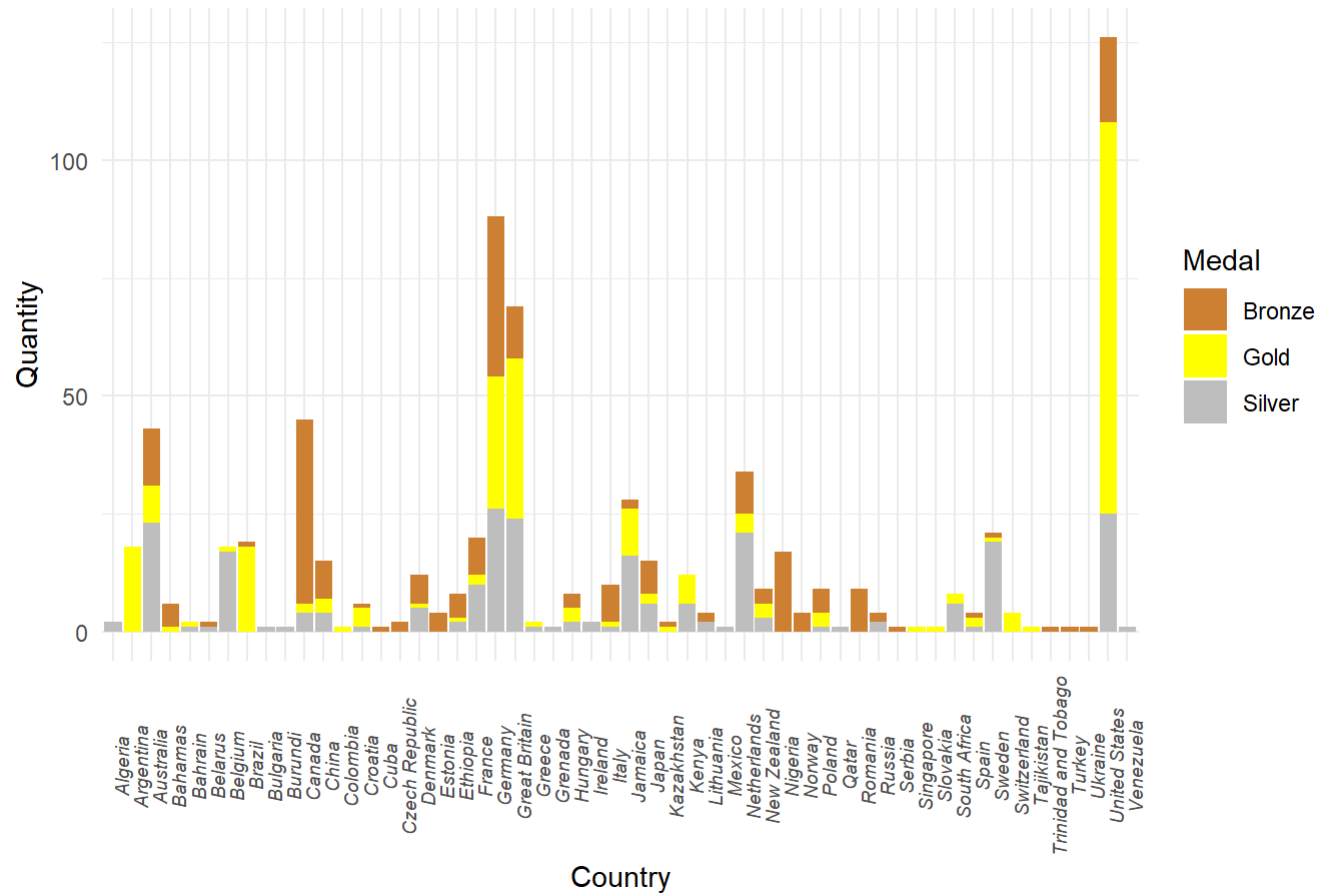
Genera un gráfico de barras que indique cuántas medallas ganó el equipo de cada país en los cinco deportes principales en 2016.

```
Principal_sports2 = data.frame(table(atdeport5$Medal, atdeport5$Sport, atdeport5$Team))
Principal_sports2 <- rename(Principal_sports2, Medal=Var1 , Sports=Var2 ,
Quantity=Freq, Country = Var3 )
head(Principal_sports2)
```

```
##      Medal   Sports Country Quantity
## 1 Bronze Athletics Algeria         0
## 2   Gold Athletics Algeria         0
## 3 Silver Athletics Algeria         2
## 4 Bronze  Football Algeria         0
## 5   Gold  Football Algeria         0
## 6 Silver  Football Algeria         0
```

```
ggplot(data=Principal_sports2, aes(x=Country, y=Quantity, fill=Medal)) +  
  geom_bar(stat="identity") +  
  scale_fill_manual(values=c("#cd7f32", "yellow", "gray"))+ theme_minimal() +  
  ggtitle ("Distribución de las Medallas según los principales Deportes") +  
  theme (axis.text.x = element_text(face="italic", size=rel(0.8), angle = 90))
```

Distribución de las Medallas según los principales Deportes



Genere un gráfico de barras que indique el peso medio de los jugadores, clasificados en función del género, que ganaron en los cinco principales deportes en 2016.

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 4.0.3
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)
```

```
## -----
```

```
##  
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':  
##  
##   arrange, count, desc, failwith, id, mutate, rename, summarise,  
##   summarize
```

```
sexmed<-ddply(atdeport5,.(Sex), summarize, mean=mean(Weight))
```

```
barplot(sexmed$mean, axes=TRUE,axisname=TRUE,  
        col=c('pink', 'blue'),main="Peso medio de los Deportistas",  
        xlab="Sexo", ylab="Kilos")  
legend(x = "right", legend = c("Femenino", "Masculino"), fill = c("pink", "blue"),  
       title = "Sexo")
```

