Taller #1 - Estadistica para la analítica de datos

Sergio Castañeda, Guillermo Castillo, Alexander Guecha

19/3/2022

knitr::opts\_chunk$set(echo = T)

# Actividad # 1

En la librería car podrá encontrar una base de datos llamada Chile, la cual proporciona parcialmente, información demográfica de Chile. La base de datos tiene 2700 filas y 8 columnas.Los datos provienen de una encuesta nacional de hogares llevada a cabo en abril y mayo de 1988 por FLACSO / Chile. Hay algunos datos que faltan .

library(readxl)  
Chile <- read\_excel("~/GitHub/estadistica/1. ChileTaller.xlsx",   
 col\_types = c("numeric", "text", "numeric",   
 "text", "numeric", "text", "numeric",   
 "numeric", "text"))

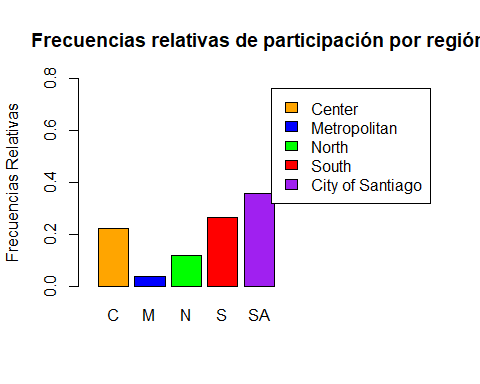
# a) Proporcione un resumen descriptivo univariado de esta información y b) Desarrolle gráficos pertinentes según el tipo de variable.

## Variable **Region**

Region\_= Chile$region  
Rtabla=data.frame(table(Region\_))  
porcentaje=prop.table(Rtabla[,2])  
Rtabla2= cbind(Rtabla, porcentaje)  
cum\_frequencia=cumsum(Rtabla2[,2])  
Rtabla3= cbind(Rtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Rtabla3[,3])  
Rtabla4= cbind(Rtabla3, cum\_porcentaje)  
Rtabla4

## Region\_ Freq porcentaje cum\_frequencia cum\_porcentaje  
## 1 C 600 0.22222222 600 0.2222222  
## 2 M 100 0.03703704 700 0.2592593  
## 3 N 322 0.11925926 1022 0.3785185  
## 4 S 718 0.26592593 1740 0.6444444  
## 5 SA 960 0.35555556 2700 1.0000000

BPregion <- barplot(prop.table(table(Chile$region)),col=c("orange","blue","green","red","purple"),legend.text=c("Center","Metropolitan","North","South","City of Santiago"),ylim=c(0,0.8),xlim=c(0,11),main="Frecuencias relativas de participación por región",ylab ="Frecuencias Relativas")



Para el caso de la variable *Region* que es cualitativa nominal, se destaca que la región de ***City of Santiago (n= 960)*** es la que tuvo mayor participación en las votaciones, seguida por las regiones ***Sur (n=718)* y *Center (n=600)***

## Variable **Population**

library(fBasics)

## Loading required package: timeDate

## Loading required package: timeSeries

basicStats(Chile$population)

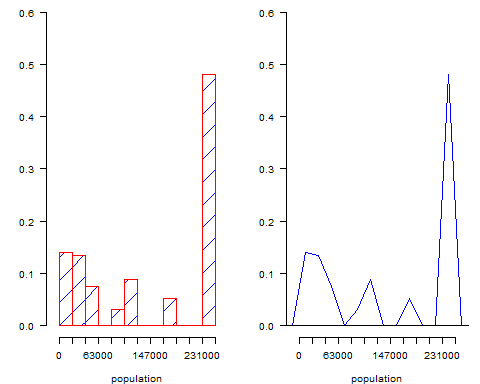
## X..Chile.population  
## nobs 2.700000e+03  
## NAs 0.000000e+00  
## Minimum 3.750000e+03  
## Maximum 2.500000e+05  
## 1. Quartile 2.500000e+04  
## 3. Quartile 2.500000e+05  
## Mean 1.522222e+05  
## Median 1.750000e+05  
## Sum 4.110000e+08  
## SE Mean 1.966802e+03  
## LCL Mean 1.483656e+05  
## UCL Mean 1.560788e+05  
## Variance 1.044444e+10  
## Stdev 1.021980e+05  
## Skewness -2.687220e-01  
## Kurtosis -1.719115e+00

#hist(Chile$population)  
library(agricolae)

##   
## Attaching package: 'agricolae'

## The following objects are masked from 'package:timeDate':  
##   
## kurtosis, skewness

par(mfrow=c(1,2),mar=c(4,4,0,1),cex=0.6)  
h1<-graph.freq(Chile$population, density=6, col="blue",border="red",ylim=c(0,0.6), frequency=2,xlab="population")  
h2<-graph.freq(Chile$population, border=0,ylim=c(0,0.6), frequency=2,xlab="population")  
polygon.freq(h2,col="blue", frequency=2)

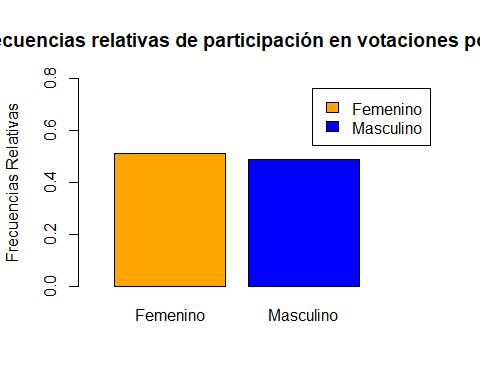


## Variable **Sex**

Sexo\_= Chile$sex  
Stabla=data.frame(table(Sexo\_))  
porcentaje=prop.table(Stabla[,2])  
Stabla2= cbind(Stabla, porcentaje)  
cum\_frequencia=cumsum(Stabla2[,2])  
Stabla3= cbind(Stabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Stabla3[,3])  
Stabla4= cbind(Stabla3, cum\_porcentaje)  
Stabla4

## Sexo\_ Freq porcentaje cum\_frequencia cum\_porcentaje  
## 1 F 1379 0.5107407 1379 0.5107407  
## 2 M 1321 0.4892593 2700 1.0000000

BPsex <- barplot(prop.table(table(Chile$sex)),col=c("orange","blue"),legend.text=c("Femenino","Masculino"),ylim=c(0,0.8),xlim=c(0,3),main="Frecuencias relativas de participación en votaciones por sexo",ylab ="Frecuencias Relativas",names.arg = c("Femenino","Masculino"))



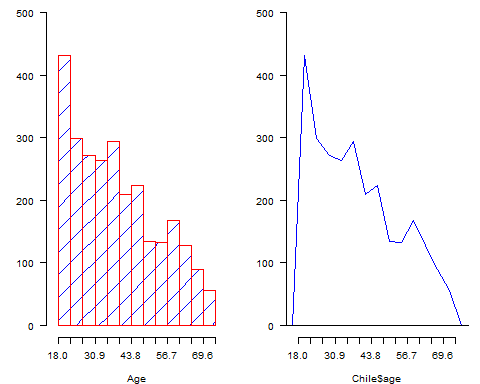
En relación con la variable ***Sexo*** que es cualitativa nominal, pudimos observar que la proporción de votantes hombres y mujeres es muy similar, con una leve diferencia de mayor participación por parte de las mujeres.

## Variable **Age**

basicStats(Chile$age)

## X..Chile.age  
## nobs 2700.000000  
## NAs 1.000000  
## Minimum 18.000000  
## Maximum 70.000000  
## 1. Quartile 26.000000  
## 3. Quartile 49.000000  
## Mean 38.548722  
## Median 36.000000  
## Sum 104043.000000  
## SE Mean 0.284040  
## LCL Mean 37.991764  
## UCL Mean 39.105680  
## Variance 217.751795  
## Stdev 14.756415  
## Skewness 0.472448  
## Kurtosis -0.862391

#hist(Chile$age)  
par(mfrow=c(1,2),mar=c(4,4,0,1),cex=0.6)  
h1<-graph.freq(Chile$age, density=6, col="blue",border="red",ylim=c(0,500), frequency=1,xlab="Age")  
h2<-graph.freq(Chile$age, border=0,ylim=c(0,500), frequency=1)  
polygon.freq(h2,col="blue", frequency=1)

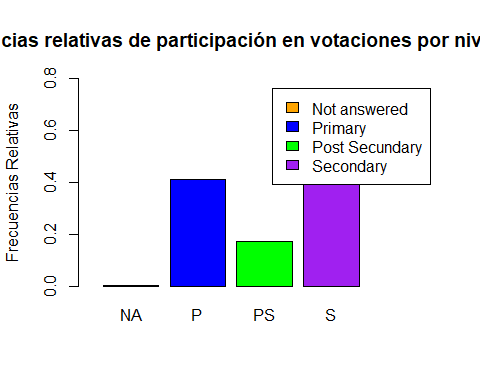


## Variable **Education**

Educacion\_= Chile$education  
Etabla=data.frame(table(Educacion\_))  
porcentaje=prop.table(Etabla[,2])  
Etabla2= cbind(Etabla, porcentaje)  
cum\_frequencia=cumsum(Etabla2[,2])  
Etabla3= cbind(Etabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Etabla3[,3])  
Etabla4= cbind(Etabla3, cum\_porcentaje)  
Etabla4

## Educacion\_ Freq porcentaje cum\_frequencia cum\_porcentaje  
## 1 NA 11 0.004074074 11 0.004074074  
## 2 P 1107 0.410000000 1118 0.414074074  
## 3 PS 462 0.171111111 1580 0.585185185  
## 4 S 1120 0.414814815 2700 1.000000000

BPEdu <- barplot(prop.table(table(Chile$education)),col=c("orange","blue","green","purple"),legend.text=c("Not answered","Primary","Post Secundary","Secondary"),ylim=c(0,0.8),xlim=c(0,6),main="Frecuencias relativas de participación en votaciones por nivel educativo",ylab ="Frecuencias Relativas")



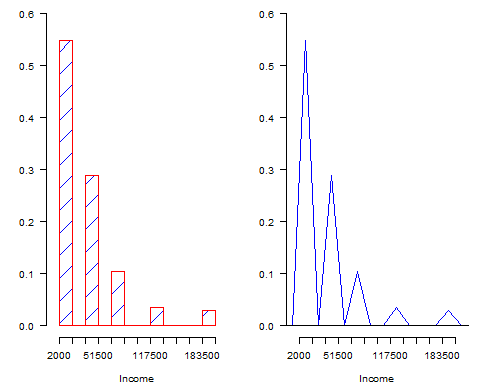
Frente a la variable ***educación*** que es cualitativa ordinal, se puede identificar que la proporción de votantes es mayor en personas con un nivel educativo de *“primaria” y “secundaria”*, por su parte *“post secundaria”* cuenta con una muy baja participación en las elecciones de Chile.

## Variable **Income**

basicStats(Chile$income)

## X..Chile.income  
## nobs 2.700000e+03  
## NAs 9.800000e+01  
## Minimum 2.500000e+03  
## Maximum 2.000000e+05  
## 1. Quartile 7.500000e+03  
## 3. Quartile 3.500000e+04  
## Mean 3.387586e+04  
## Median 1.500000e+04  
## Sum 8.814500e+07  
## SE Mean 7.744172e+02  
## LCL Mean 3.235733e+04  
## UCL Mean 3.539440e+04  
## Variance 1.560477e+09  
## Stdev 3.950287e+04  
## Skewness 2.584549e+00  
## Kurtosis 7.291944e+00

#hist(Chile$income)  
par(mfrow=c(1,2),mar=c(4,4,0,1),cex=0.6)  
h1<-graph.freq(Chile$income, density=6, col="blue",border="red", frequency=2,xlab="Income")  
h2<-graph.freq(Chile$income, border=0, frequency=2,xlab="Income")  
polygon.freq(h2,col="blue", frequency=2)

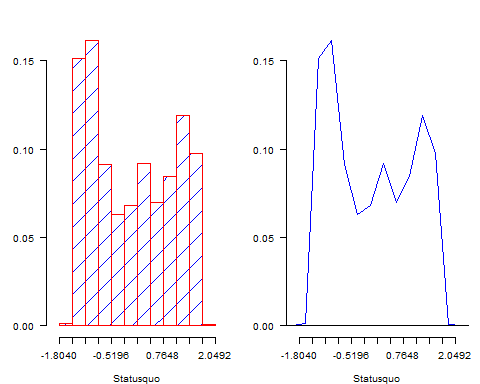


## Variable **Statusquo**

library(fBasics)  
basicStats(Chile$statusquo)

## X..Chile.statusquo  
## nobs 2700.000000  
## NAs 17.000000  
## Minimum -1.803010  
## Maximum 2.048590  
## 1. Quartile -1.002235  
## 3. Quartile 0.968575  
## Mean 0.000000  
## Median -0.045580  
## Sum -0.000030  
## SE Mean 0.019309  
## LCL Mean -0.037863  
## UCL Mean 0.037863  
## Variance 1.000373  
## Stdev 1.000186  
## Skewness 0.161683  
## Kurtosis -1.454072

#hist(Chile$statusquo)  
par(mfrow=c(1,2),mar=c(4,4,0,1),cex=0.6)  
h1<-graph.freq(Chile$statusquo, density=6, col="blue",border="red", frequency=2,xlab="Statusquo")  
h2<-graph.freq(Chile$statusquo, border=0, frequency=2,xlab="Statusquo")  
polygon.freq(h2,col="blue", frequency=2)

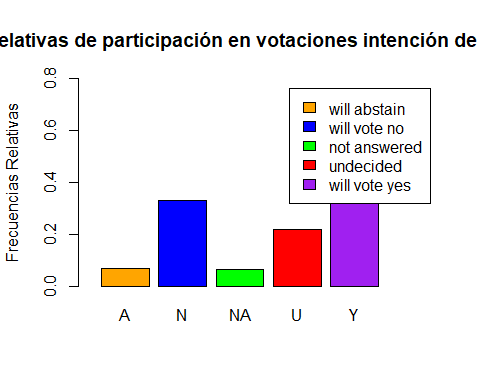


## Variable **Vote**

Vote\_= Chile$vote  
Vtabla=data.frame(table(Vote\_))  
porcentaje=prop.table(Vtabla[,2])  
Vtabla2= cbind(Vtabla, porcentaje)  
cum\_frequencia=cumsum(Vtabla2[,2])  
Vtabla3= cbind(Vtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Vtabla3[,3])  
Vtabla4= cbind(Vtabla3, cum\_porcentaje)  
Vtabla4

## Vote\_ Freq porcentaje cum\_frequencia cum\_porcentaje  
## 1 A 187 0.06925926 187 0.06925926  
## 2 N 889 0.32925926 1076 0.39851852  
## 3 NA 168 0.06222222 1244 0.46074074  
## 4 U 588 0.21777778 1832 0.67851852  
## 5 Y 868 0.32148148 2700 1.00000000

BPvote <- barplot(prop.table(table(Chile$vote)),col=c("orange","blue","green","red","purple"),legend.text=c("will abstain","will vote no","not answered","undecided","will vote yes"),ylim=c(0,0.8),xlim=c(0,7),main="Frecuencias relativas de participación en votaciones intención de voto por Pinochet",ylab ="Frecuencias Relativas")



Respecto a la variable ***Vote*** que es cualitativa nominal y está relacionada con la intención de voto por el presidente “Pinochet”, se encuentra que las personas que votaron en SI (33%) son muy cercanas a las personas que votaron en NO (32%)

1. Desagregando esta base de datos, solo para hogares del Norte y del Sur, proporcione un análisis bivariado (tabulación cruzada y un diagrama de barras comparativas) de las variables Educación y Voto.

# Análisis Bivariado Educación y voto

library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

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

sel<-c("N","S")  
Chile\_NteySur <- Chile %>% filter(region %in% sel)  
library(gmodels)  
CrossTable(Chile\_NteySur$education,Chile\_NteySur$vote)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1040   
##   
##   
## | Chile\_NteySur$vote   
## Chile\_NteySur$education | A | N | NA | U | Y | Row Total |   
## ------------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## NA | 0 | 1 | 0 | 0 | 1 | 2 |   
## | 0.138 | 0.253 | 0.092 | 0.373 | 0.057 | |   
## | 0.000 | 0.500 | 0.000 | 0.000 | 0.500 | 0.002 |   
## | 0.000 | 0.003 | 0.000 | 0.000 | 0.002 | |   
## | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 | |   
## ------------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## P | 23 | 99 | 23 | 107 | 227 | 479 |   
## | 3.114 | 14.884 | 0.036 | 3.486 | 7.713 | |   
## | 0.048 | 0.207 | 0.048 | 0.223 | 0.474 | 0.461 |   
## | 0.319 | 0.313 | 0.479 | 0.552 | 0.554 | |   
## | 0.022 | 0.095 | 0.022 | 0.103 | 0.218 | |   
## ------------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## PS | 14 | 80 | 8 | 20 | 51 | 173 |   
## | 0.342 | 14.319 | 0.000 | 4.666 | 4.339 | |   
## | 0.081 | 0.462 | 0.046 | 0.116 | 0.295 | 0.166 |   
## | 0.194 | 0.253 | 0.167 | 0.103 | 0.124 | |   
## | 0.013 | 0.077 | 0.008 | 0.019 | 0.049 | |   
## ------------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## S | 35 | 136 | 17 | 67 | 131 | 386 |   
## | 2.564 | 2.986 | 0.037 | 0.348 | 2.946 | |   
## | 0.091 | 0.352 | 0.044 | 0.174 | 0.339 | 0.371 |   
## | 0.486 | 0.430 | 0.354 | 0.345 | 0.320 | |   
## | 0.034 | 0.131 | 0.016 | 0.064 | 0.126 | |   
## ------------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
## Column Total | 72 | 316 | 48 | 194 | 410 | 1040 |   
## | 0.069 | 0.304 | 0.046 | 0.187 | 0.394 | |   
## ------------------------|-----------|-----------|-----------|-----------|-----------|-----------|  
##   
##

**Conclusiones**

De acuerdo con el resultado de la anterior tabla de contingencia, encontramos que el mayor valor ***Chi Cuadrado*** se ubica entre un nivel de estudio de primaria (P) y la intención de votar en **NO** por Pinochet como presidente. De otra parte, se evidencia que dicha relación entre la intención de votar en **NO** por Pinochet como presidente es alta en personas con un nivel educativo de PostSecundaria (PS).

En otro de los cruces de variable en donde encontramos un ***Chi Cuadrado*** superior es en la intención de voto de respaldar a Pinochet con el SI en personas con un nivel educativo de primaria (P)

# Actividad # 2

Una muestra de 226 personas mayores que viven en Burdeos (Gironde, suroeste de Francia) fueron entrevistados en 2000 para un estudio nutricional (base de datos: nutrition\_elderly ). La siguiente tabla presenta la descripción de las variables de estudio.

Proporcione un resumen estadístico descriptivo completo de dos variables cualitativas y dos cuantitativas, solo con aquellas personas mayores de 79 de género femenino.

En cada uno de los casos, proporcione los análisis, conclusiones y recomendaciones analíticas.

library(readxl)  
Nutrition <- read\_excel("~/GitHub/estadistica/2. nutrition\_elderly.xlsx",   
 col\_types = c("text", "text", "numeric",   
 "numeric", "numeric", "numeric",   
 "numeric", "text", "text", "text",   
 "text", "text", "text"))

**Selección de personas mayores de 79 de género femenino**

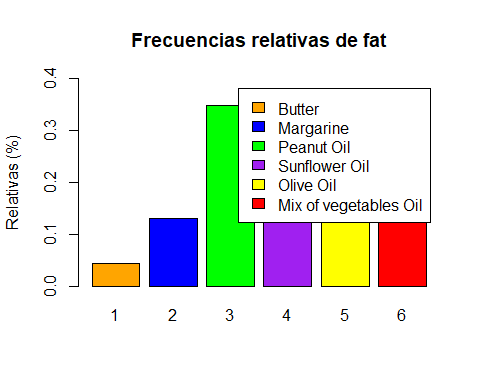
Nutrition\_2<-subset(Nutrition,age>79 & gender=="2")

## Resumen estadístico descriptivo Variable **Fat**

Fat\_= Nutrition\_2$fat  
Fat\_tabla=data.frame(table(Fat\_))  
porcentaje=prop.table(Fat\_tabla[,2])  
Fat\_tabla2= cbind(Fat\_tabla, porcentaje)  
cum\_frequencia=cumsum(Fat\_tabla2[,2])  
Fat\_tabla3= cbind(Fat\_tabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Fat\_tabla3[,3])  
Fat\_tabla4= cbind(Fat\_tabla3, cum\_porcentaje)  
Fat\_tabla4

## Fat\_ Freq porcentaje cum\_frequencia cum\_porcentaje  
## 1 1 1 0.04347826 1 0.04347826  
## 2 2 3 0.13043478 4 0.17391304  
## 3 3 8 0.34782609 12 0.52173913  
## 4 4 5 0.21739130 17 0.73913043  
## 5 5 3 0.13043478 20 0.86956522  
## 6 6 3 0.13043478 23 1.00000000

BPFat <- barplot(prop.table(table(Nutrition\_2$fat)),col=c("orange","blue","green","purple","yellow","red"),   
 ylim=c(0,0.4),main="Frecuencias relativas de fat",ylab ="Relativas (%)", legend.text = c("Butter","Margarine","Peanut Oil","Sunflower Oil","Olive Oil","Mix of vegetables Oil"))



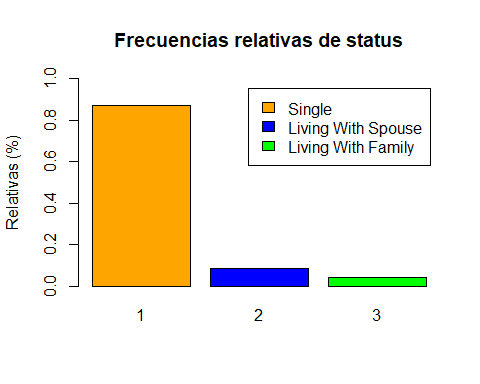
En relación con el resultado del análisi estadístico de la variable cualitativa ***Fat*** en personas mayores de 79 de género femenino que viven en la ciudad de Burdeos, se observa que en primer lugar es preferido el aceite de cacahuete con un 35%, seguidamente el aceite de girasol con un 22% de preferencia

## Resumen estadístico descriptivo Variable **Status**

Status\_= Nutrition\_2$status  
Status\_tabla=data.frame(table(Status\_))  
porcentaje=prop.table(Status\_tabla[,2])  
Status\_tabla2= cbind(Status\_tabla, porcentaje)  
cum\_frequencia=cumsum(Status\_tabla2[,2])  
Status\_tabla3= cbind(Status\_tabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Status\_tabla3[,3])  
Status\_tabla4= cbind(Status\_tabla3, cum\_porcentaje)  
Status\_tabla4

## Status\_ Freq porcentaje cum\_frequencia cum\_porcentaje  
## 1 1 20 0.86956522 20 0.8695652  
## 2 2 2 0.08695652 22 0.9565217  
## 3 3 1 0.04347826 23 1.0000000

BPStatus <- barplot(prop.table(table(Nutrition\_2$status)),col=c("orange","blue","green"), ylim=c(0,1),main="Frecuencias relativas de status",ylab ="Relativas (%)",legend.text = c("Single","Living With Spouse","Living With Family"))



En relación con el resultado del análisi estadístico de la variable cualitativa ***Status*** en personas mayores de 79 de género femenino que viven en la ciudad de Burdeos, se observa que el 85% viven solas y el 15% restante conviven con la familia o con su esposo.

**Tabla de contingencia cruzada entre Fat y Status**

CrossTable(Nutrition\_2$fat,Nutrition\_2$status)

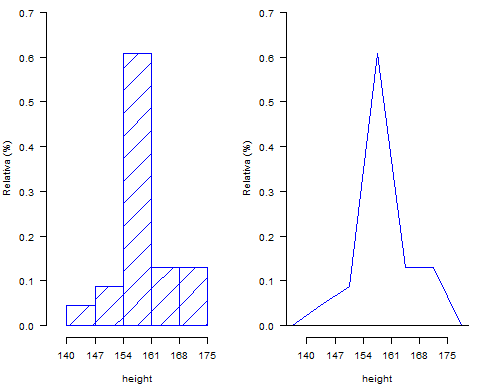
##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 23   
##   
##   
## | Nutrition\_2$status   
## Nutrition\_2$fat | 1 | 2 | 3 | Row Total |   
## ----------------|-----------|-----------|-----------|-----------|  
## 1 | 0 | 0 | 1 | 1 |   
## | 0.870 | 0.087 | 21.043 | |   
## | 0.000 | 0.000 | 1.000 | 0.043 |   
## | 0.000 | 0.000 | 1.000 | |   
## | 0.000 | 0.000 | 0.043 | |   
## ----------------|-----------|-----------|-----------|-----------|  
## 2 | 3 | 0 | 0 | 3 |   
## | 0.059 | 0.261 | 0.130 | |   
## | 1.000 | 0.000 | 0.000 | 0.130 |   
## | 0.150 | 0.000 | 0.000 | |   
## | 0.130 | 0.000 | 0.000 | |   
## ----------------|-----------|-----------|-----------|-----------|  
## 3 | 8 | 0 | 0 | 8 |   
## | 0.157 | 0.696 | 0.348 | |   
## | 1.000 | 0.000 | 0.000 | 0.348 |   
## | 0.400 | 0.000 | 0.000 | |   
## | 0.348 | 0.000 | 0.000 | |   
## ----------------|-----------|-----------|-----------|-----------|  
## 4 | 3 | 2 | 0 | 5 |   
## | 0.418 | 5.635 | 0.217 | |   
## | 0.600 | 0.400 | 0.000 | 0.217 |   
## | 0.150 | 1.000 | 0.000 | |   
## | 0.130 | 0.087 | 0.000 | |   
## ----------------|-----------|-----------|-----------|-----------|  
## 5 | 3 | 0 | 0 | 3 |   
## | 0.059 | 0.261 | 0.130 | |   
## | 1.000 | 0.000 | 0.000 | 0.130 |   
## | 0.150 | 0.000 | 0.000 | |   
## | 0.130 | 0.000 | 0.000 | |   
## ----------------|-----------|-----------|-----------|-----------|  
## 6 | 3 | 0 | 0 | 3 |   
## | 0.059 | 0.261 | 0.130 | |   
## | 1.000 | 0.000 | 0.000 | 0.130 |   
## | 0.150 | 0.000 | 0.000 | |   
## | 0.130 | 0.000 | 0.000 | |   
## ----------------|-----------|-----------|-----------|-----------|  
## Column Total | 20 | 2 | 1 | 23 |   
## | 0.870 | 0.087 | 0.043 | |   
## ----------------|-----------|-----------|-----------|-----------|  
##   
##

## Resumen estadístico descriptivo Variable **height**

library(fBasics)  
basicStats(Nutrition\_2$height)

## X..Nutrition\_2.height  
## nobs 23.000000  
## NAs 0.000000  
## Minimum 140.000000  
## Maximum 175.000000  
## 1. Quartile 154.500000  
## 3. Quartile 161.500000  
## Mean 158.695652  
## Median 159.000000  
## Sum 3650.000000  
## SE Mean 1.530242  
## LCL Mean 155.522125  
## UCL Mean 161.869179  
## Variance 53.857708  
## Stdev 7.338781  
## Skewness -0.131839  
## Kurtosis 0.540451

par(mfrow=c(1,2),mar=c(4,4,0,1),cex=0.6)  
h1<-graph.freq(Nutrition\_2$height, density=6, col="blue", frequency=2,xlab="height",ylab="Relativa (%)",ylim=c(0,0.7))  
h2<-graph.freq(Nutrition\_2$height, border=0, frequency=2,ylim=c(0,0.7),xlab="height",ylab="Relativa (%)")  
polygon.freq(h2,col="blue", frequency=2)



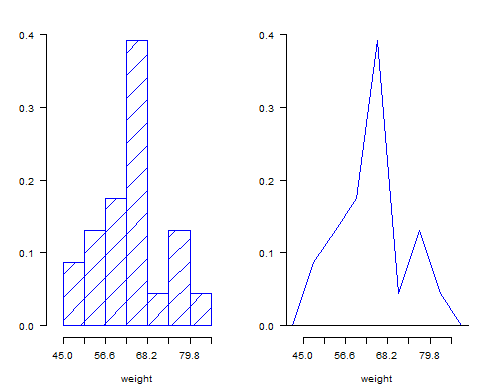
En relación con el resultado del análisis estadístico de la variable cuantitativa ***Height*** en personas mayores de 79 de género femenino que viven en la ciudad de Burdeos, se observa que al rededor del 60% tienen una estatura entre 154 y 161 centímetros.

## Resumen estadístico descriptivo Variable **Weight**

library(fBasics)  
basicStats(Nutrition\_2$weight)

## X..Nutrition\_2.weight  
## nobs 23.000000  
## NAs 0.000000  
## Minimum 45.000000  
## Maximum 80.000000  
## 1. Quartile 60.000000  
## 3. Quartile 67.500000  
## Mean 63.826087  
## Median 64.000000  
## Sum 1468.000000  
## SE Mean 1.786701  
## LCL Mean 60.120696  
## UCL Mean 67.531478  
## Variance 73.422925  
## Stdev 8.568718  
## Skewness -0.144788  
## Kurtosis -0.427002

par(mfrow=c(1,2),mar=c(4,4,0,1),cex=0.6)  
h1<-graph.freq(Nutrition\_2$weight, density=6, col="blue", frequency=2,xlab="weight")  
h2<-graph.freq(Nutrition\_2$weight, border=0, frequency=2,xlab="weight")  
polygon.freq(h2,col="blue", frequency=2)



En relación con el resultado del análisis estadístico de la variable cuantitativa ***Weight*** en personas mayores de 79 de género femenino que viven en la ciudad de Burdeos, se observa que al rededor del 40% tienen tienen un peso corporal que oscila entre 62 y 68 kilogramos.

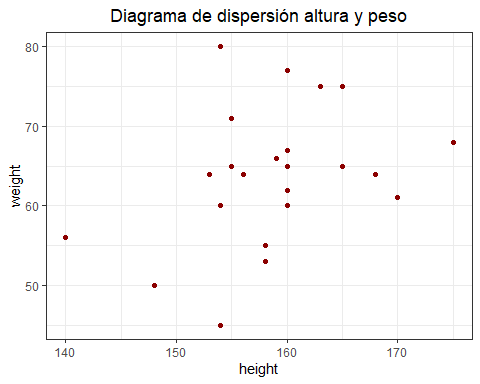
**Coeficiente de Correlación de Pearson entre las variables cuantitativas height y weight:**

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library(ggplot2)  
  
  
# Gráfico Simple (X,Y)  
ggplot(data = Nutrition\_2, aes(x = height, y = weight)) +   
 geom\_point(colour = "red4") +  
 ggtitle("Diagrama de dispersión altura y peso") +  
 theme\_bw() +  
 theme(plot.title = element\_text(hjust = 0.5))



# Midiendo el nivel de correlación:  
#cor(x = log10(Nutrition\_2$height), y = Nutrition\_2$weight)  
cor(x = Nutrition\_2$height, y = Nutrition\_2$weight, method = "pearson")

## [1] 0.3511394

De acuerdo con el resultado del análisis de correlación entre las variables ***Height*** y ***Weight*** en personas mayores de 79 de género femenino que viven en la ciudad de Burdeos, se observa que el nivel de correlación es del 35%, lo cual es muy bajo, no obstante se sugiere analizar otras variables exógenas que puedan afectar esta correlación como son el clima o el nivel social y económico de las muejeres entrevistadas, entre otras.