

```

#Tabla de frecuencias para cualitativas (tfc)
tfc<-function(variable){
  fi<-table(variable)
  hi<-fi/length(variable)
  pi<-hi*100
  atributos<-cbind(fi,hi,pi)
  totales<-c(sum(fi),sum(hi),sum(pi))
  pie(table((variable)))
  r<-rbind(atributos,totales)
  View(r)
}

#Tabla de frecuencias para discretas (tfd)
tfd<-function(variable){
  fi<-table(variable)
  hi<-fi/length(variable)
  pi<-hi*100
  atributos<-cbind(fi,hi,pi)
  totales<-c(sum(fi),sum(hi),sum(pi))
  barplot(table(variable))
  plot(table(variable))
  r<-rbind(atributos,totales)
  View(r)
}

#Tabla de frecuencias para variables continuas (tfcont)
#calculo del ancho de clase
calW<-function(v){
  r=max(v)-min(v)
  k=1+3.32*log10(length(v))
  k<-round(k)
  w=r/k
  w
}

#Tabla de frecuencia y grafico de histograma (tfCont)
tfCont<-function(v,w,nombre){
  k=1+3.32*log10(length(v))
  k=round(k)
  LS<-c()
  LI<-c()
  LS[1]=min(v)+w
  LI[1]=min(v)
  for (i in 2:k){
    LS[i]=LS[i-1]+w
    LI[i]=LI[i-1]+w
  }
  h<-hist(v,seq(min(v),max(LS),w),plot=F)
  fi<-h$counts
  hi<-fi/sum(fi)
  pi<-hi*100
  Fi<-cumsum(fi)
  Hi<-cumsum(hi)
  Pi<-cumsum(pi)

hist(v,seq(min(v),max(LS),w),freq=FALSE,col='darkolivegreen1',plot=T,main
= paste('Histograma de la Variable:',nombre))
  lines(density(v),col="red",lwd=2)
}

```

```

    abline(v=mean(v),col='red')
    abline(v=median(v),col='blue')
    curve(dnorm(x,mean=mean(v),sd=sd(v)), from=800,to=1600,add=TRUE,
col="blue", lwd=2)
    K<-1:k
    atributos<-cbind(K,LI,LS,fi,hi,pi,Fi,Hi,Pi)
    totales<-c(' ',' ','TOTALES',sum(fi),sum(hi),sum(pi),' ',' ',' ')
    r<-rbind(atributos,totales)
    View(r)
}

```

#Funcion para medidas de resumen

```
GrafC<-function(x){
```

```

  #media
  me=mean(x)
  #mediana
  md=median(x)
  #Cuantiles
  q1=as.numeric(quantile(x,0.25))
  q2=as.numeric(quantile(x,0.50))
  q3=as.numeric(quantile(x,0.75))
  #MEDIDAS DE DISPERSION
  #rango
  r=max(x)-min(x)
  #Rango intercuartico
  iqr=q3-q1
  #GRAFICO DE CAJAS Y BIGOTES
  #par(mfrow=c(1,2))
  boxplot(x,horizontal = TRUE)
  #plot(density(x, adjust = 5),col = 'black', lwd = 3)
  LI=q1-1.5*iqr
  LS=q3+1.5*iqr
  as=3*(mean(x)-median(x))/sd(x)
  d9<-as.numeric(quantile(x,0.9))
  d1<-as.numeric(quantile(x,0.1))
  k=(0.5*(q3-q1))/(d9-d1)
  ss=var(x)
  s=sd(x)
  cv=s/me*100
  numero=c(me,md,q1,q2,q3,r,iqr,LI,LS,round(as,3),k,ss,s,cv)

```

```

nombre=c('Media','Mediana','Q1','Q2','Q3','r','iqr','LI','LS','as','k','v
ar','sd','cv')

```

```

  View(cbind(nombre,numero))
}

```

#grafico de la varaianza

```

grafVar<-function(x){
  n=length(x)
  plot(x,1:n)
  abline(v=mean(x),col='red')
  abline(v=median(x),col='blue')
}

```

```
nube_de_palabras<-function(texto){
```

```

  texto = iconv(texto, to="ASCII//TRANSLIT")

```

```

texto = Corpus(VectorSource(texto))

##### LIMPIAMOS NUESTRO TEXTO CON EL COMANDO tm_map

#ponemos todos los datos a minuscula (A!=a)
discurso=tm_map(texto, tolower)
#quitamos los espacios en blanco
discurso =tm_map(discurso, stripWhitespace)
#quitamos la puntuacion
discurso = tm_map(discurso, removePunctuation)
#quitamos los numeros
discurso = tm_map(discurso, removeNumbers)

#mostramos palabras vacias y genericas
stopwords("spanish")
#quitamos palabras genericas
discurso=tm_map(discurso, removeWords,stopwords("spanish"))

#tambien podemos tener nuestra propia lista de palabras a quitar

##### DATA FRAME DE PALABRAS CON SU FRECUENCIA

#Creamos matriz de letras
letras= TermDocumentMatrix(discurso)
findFreqTerms(letras, lowfreq=5)
matrix=as.matrix(letras)

#lo ordenamos y sumamos las letras de nuestra matriz
vector <- sort(rowSums(matrix),decreasing=TRUE)
#creamos la data con las palabras y su frecuencia
dataletras <- data.frame(word= names(vector),frecuencia=vector)
# lo nombra y le da formato de data.frame

##### GRAFICAMOS FRECUENCIA DE LAS PALABRAS

barplot(dataletras[1:10,]$freq, las = 2, names.arg =
dataletras[1:10,]$word,col ="blue", main ="PALABRAS MÁS FRECUENTES",
ylab = "Frecuencia de palabras")

##### GRAFICAMOS LA NUBE DE PALABRAS
wordcloud(words = dataletras$word, freq = dataletras$freq, min.freq =
2,
          max.words=70)
#en el centro la palabra mas importante,
wordcloud(words = dataletras$word, freq = dataletras$freq, min.freq =
2,
          max.words=70, random.order=FALSE, rot.per=0.2,
          colors=brewer.pal(8, "Dark2"))

}

#Calcular el promedio de las publicaciones por empresa
#aggregate(love,by=list(empresa),mean)

```