

Covid-19 Analysis

José Armando Rosas Balderas

20 países con más casos reportados (actualizado Mayo 2022)

Estado Unidos: Epsilon India: Delta

Brasil: Gamma Francia: Alpha Alemania: Alpha Reino Unido: Alpha (B.1.1.7) Rusia: Alpha (B.1.1.31)

Corea del Sur: Omicron (BA.1.1.5) Italia: Alpha (A.2.5.2) Turquía: Wuhan (A.28) España: Alpha (C.21)

Vietnam: Omicron (BA.2.3.2) Argentina: Gamma(B.1.499) Países Bajos: Alpha (L.2) Japón: Wuhan Irán:

Alpha Australia: Wuhan Colombia: Gamma Indonesia: Alpha Polonia: Alpha

Obteniendo las secuencias de cada país

```
EUA_F = system.file("sequences/Sec_Pais/usa.fasta", package = "seqinr")
File_EUA = seqinr::read.fasta(file=EUA_F, as.string = TRUE,
                              seqtype = "AA")$OM418388.1[1]

India_F = system.file("sequences/Sec_Pais/india.fasta", package = "seqinr")
File_India = seqinr::read.fasta(file=India_F, as.string = TRUE,
                                seqtype = "AA")$MW828330.1[1]

Brasil_F = system.file("sequences/Sec_Pais/brazil.fasta", package = "seqinr")
File_Brasil = seqinr::read.fasta(file=Brasil_F, as.string = TRUE,
                                  seqtype = "AA")$MW520923.1[1]

Francia_F = system.file("sequences/Sec_Pais/france.fasta", package = "seqinr")
File_Francia = seqinr::read.fasta(file=Francia_F, as.string = TRUE,
                                   seqtype = "AA")$MZ056105.1[1]

Alemania_F = system.file("sequences/Sec_Pais/germany.fasta", package = "seqinr")
File_Alemania = seqinr::read.fasta(file=Alemania_F, as.string = TRUE,
                                    seqtype = "AA")$MW241372.1[1]

UK_F = system.file("sequences/Sec_Pais/uk.fasta",
                   package = "seqinr")
File_UK = seqinr::read.fasta(file=UK_F, as.string = TRUE,
                              seqtype = "AA")$MT731663.1[1]

Rusia_F = system.file("sequences/Sec_Pais/russia.fasta", package = "seqinr")
File_Rusia = seqinr::read.fasta(file=Rusia_F, as.string = TRUE,
                                 seqtype = "AA")$OU073031.1[1]

CS_F = system.file("sequences/Sec_Pais/southkorea.fasta", package = "seqinr")
File_CS = seqinr::read.fasta(file=CS_F, as.string = TRUE,
                              seqtype = "AA")$OM348680.1[1]
```

```

Italia_F = system.file("sequences/Sec_Pais/italy.fasta", package = "seqinr")
File_Italia = seqinr::read.fasta(file=Italia_F, as.string = TRUE,
                                seqtype = "AA")$ON302929.1[1]

Turquía_F = system.file("sequences/Sec_Pais/turkey.fasta", package = "seqinr")
File_Turquía = seqinr::read.fasta(file=Turquía_F, as.string = TRUE,
                                seqtype = "AA")$MW706337.1[1]

España_F = system.file("sequences/Sec_Pais/spain.fasta", package = "seqinr")
File_España = seqinr::read.fasta(file=España_F, as.string = TRUE,
                                seqtype = "AA")$MT873062.1[1]

Vietnam_F = system.file("sequences/Sec_Pais/vietnam.fasta",
                        package = "seqinr")
File_Vietnam = seqinr::read.fasta(file=Vietnam_F , as.string = TRUE,
                                seqtype = "AA")$ON414699.1[1]

Argentina_F = system.file("sequences/Sec_Pais/argentina.fasta", package = "seqinr")
File_Argentina = seqinr::read.fasta(file=Argentina_F, as.string = TRUE,
                                seqtype = "AA")$OK665087.1[1]

PB_F = system.file("sequences/Sec_Pais/netherlands.fasta", package = "seqinr")
File_PB = seqinr::read.fasta(file=PB_F, as.string = TRUE,
                              seqtype = "AA")$LR883305.1[1]

Japón_F = system.file("sequences/Sec_Pais/japan.fasta", package = "seqinr")
File_Japón = seqinr::read.fasta(file=Japón_F, as.string = TRUE,
                              seqtype = "AA")$OM982635.1[1]

Irán_F = system.file("sequences/Sec_Pais/iran.fasta", package = "seqinr")
File_Irán = seqinr::read.fasta(file=Irán_F, as.string = TRUE,
                              seqtype = "AA")$OL913332.1[1]

Australia_F = system.file("sequences/Sec_Pais/australia.fasta", package = "seqinr")
File_Australia = seqinr::read.fasta(file=Australia_F, as.string = TRUE,
                              seqtype = "AA")$OK426553.1[1]

Colombia_F = system.file("sequences/Sec_Pais/colombia.fasta",
                        package = "seqinr")
File_Colombia = seqinr::read.fasta(file=Colombia_F , as.string = TRUE,
                              seqtype = "AA")$OM898086.1[1]

Indonesia_F = system.file("sequences/Sec_Pais/indonesia.fasta", package = "seqinr")
File_Indonesia = seqinr::read.fasta(file=Indonesia_F, as.string = TRUE,
                              seqtype = "AA")$ON420280.1[1]

Polonia_F = system.file("sequences/Sec_Pais/poland.fasta", package = "seqinr")
File_Polonia = seqinr::read.fasta(file=Polonia_F, as.string = TRUE,
                              seqtype = "AA")$OM739138.1[1]

```

Funciones

```
Str_to_Vector <- function(str){
  V <- c(strsplit(str, ""))[[1]]
  Z <-c()
  j = 1
  for (i in 1:length(V)){
    if (V[i]!="\n"){
      Z[j]<- V[i]
      j <- j+1
    }
  }
  return(Z)
}

Tamaño_ADN <- function(S) {
  return(length(S))
}

Porcentaje_ADN <- function(S) {
  a = (length(S[S=="A"])/length(S))*100
  g = (length(S[S=="G"])/length(S))*100
  c = (length(S[S=="C"])/length(S))*100
  t = (length(S[S=="T"])/length(S))*100
  n = (length(S[S=="N"])/length(S))*100
  o = (length(S[S=="-"])/length(S))*100
  porc <- sprintf("A=%f G=%f C=%f T=%f N=%f -=%f", a, g, c, t, n, o)
  return(porc)
}

Porcentaje_A_ADN <- function(S) {
  a = (length(S[S=="A"])/length(S))*100
  return(a)
}

Porcentaje_C_ADN <- function(S) {
  c = (length(S[S=="C"])/length(S))*100
  return(c)
}

Porcentaje_G_ADN <- function(S) {
  g = (length(S[S=="G"])/length(S))*100
  return(g)
}

Porcentaje_T_ADN <- function(S) {
  t = (length(S[S=="T"])/length(S))*100
  return(t)
}
```

Convertir str a vectores

```

EUA = Str_to_Vector(File_EUA)
India = Str_to_Vector(File_India)
Brasil = Str_to_Vector(File_Brasil)
Francia = Str_to_Vector(File_Francia)
Alemania = Str_to_Vector(File_Alemania)
UK = Str_to_Vector(File_UK)
Rusia = Str_to_Vector(File_Rusia)
CS = Str_to_Vector(File_CS)
Italia = Str_to_Vector(File_Italia)
Turquía = Str_to_Vector(File_Turquía)
España = Str_to_Vector(File_España)
Vietnam = Str_to_Vector(File_Vietnam)
Argentina = Str_to_Vector(File_Argentina)
PB = Str_to_Vector(File_PB)
Japon = Str_to_Vector(File_Japón)
Iran = Str_to_Vector(File_Irán)
Australia = Str_to_Vector(File_Australia)
Colombia = Str_to_Vector(File_Colombia)
Indonesia = Str_to_Vector(File_Indonesia)
Polonia = Str_to_Vector(File_Polonia)

```

Calcular longitud

```

T_EUA = Tamaño_ADN(EUA)
T_EUA

```

```
## [1] 29725
```

```

T_India = Tamaño_ADN(India)
T_India

```

```
## [1] 29807
```

```

T_Brasil = Tamaño_ADN(Brasil)
T_Brasil

```

```
## [1] 29819
```

```

T_Francia = Tamaño_ADN(Francia)
T_Francia

```

```
## [1] 29848
```

```

T_Alemania = Tamaño_ADN(Alemania)
T_Alemania

```

```
## [1] 29853
```

```
T_UK = Tamaño_ADN(UK)
T_UK
```

```
## [1] 29903
```

```
T_Rusia = Tamaño_ADN(Rusia)
T_Rusia
```

```
## [1] 29682
```

```
T_CS = Tamaño_ADN(CS)
T_CS
```

```
## [1] 29747
```

```
T_Italia = Tamaño_ADN(Italia)
T_Italia
```

```
## [1] 29669
```

```
T_Turquía = Tamaño_ADN(Turquía)
T_Turquía
```

```
## [1] 29707
```

```
T_España = Tamaño_ADN(España)
T_España
```

```
## [1] 29900
```

```
T_Vietnam = Tamaño_ADN(Vietnam)
T_Vietnam
```

```
## [1] 29505
```

```
T_Argentina = Tamaño_ADN(Argentina)
T_Argentina
```

```
## [1] 29764
```

```
T_PB = Tamaño_ADN(PB)
T_PB
```

```
## [1] 29779
```

```
T_Japón = Tamaño_ADN(Japon)
T_Japón
```

```
## [1] 29839
```

```
T_Irán = Tamaño_ADN(Iran)
T_Irán
```

```
## [1] 29014
```

```
T_Australia = Tamaño_ADN(Australia)
T_Australia
```

```
## [1] 29782
```

```
T_Colombia = Tamaño_ADN(Colombia)
T_Colombia
```

```
## [1] 29540
```

```
T_Indonesia = Tamaño_ADN(Indonesia)
T_Indonesia
```

```
## [1] 29746
```

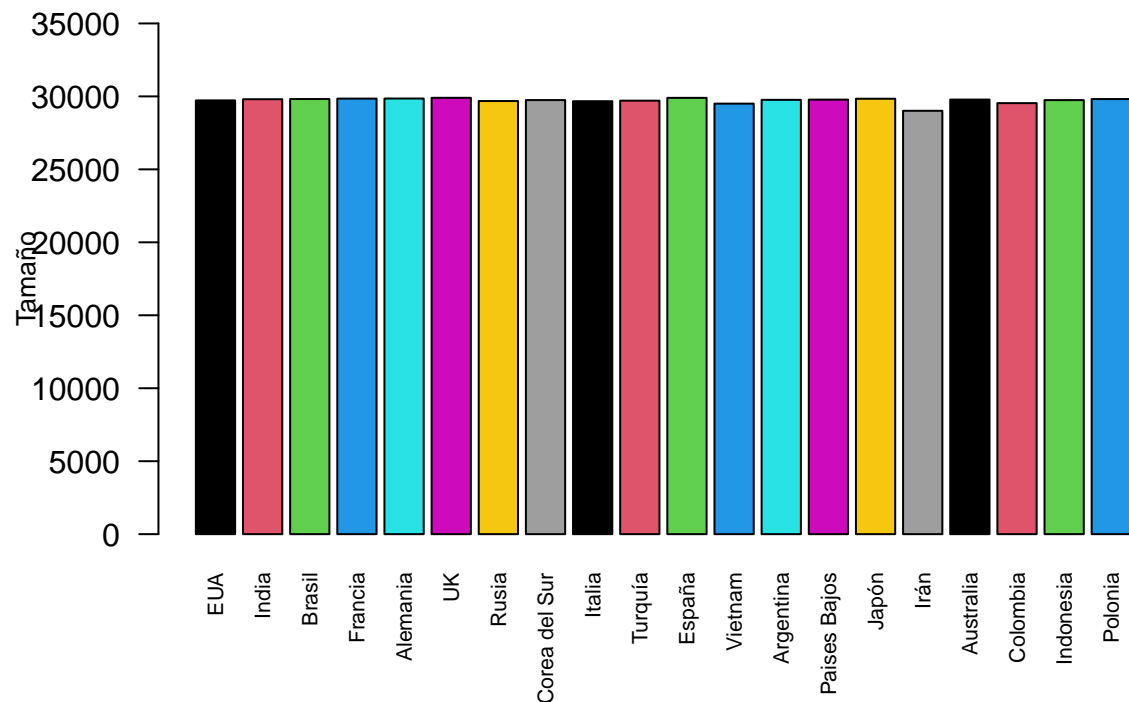
```
T_Polonia = Tamaño_ADN(Polonia)
T_Polonia
```

```
## [1] 29819
```

Gráfica del tamaño de cada secuencia

```
barplot(c(T_EUA, T_India, T_Brasil, T_Francia, T_Alemania, T_UK, T_Rusia, T_CS,
  T_Italia, T_Turquía, T_España, T_Vietnam, T_Argentina, T_PB, T_Japón, T_Irán,
  T_Australia, T_Colombia, T_Indonesia, T_Polonia),
names.arg = c("EUA", "India", "Brasil", "Francia", "Alemania", "UK", "Rusia",
  "Corea del Sur", "Italia", "Turquía", "España", "Vietnam", "Argentina",
  "Países Bajos",
  "Japón", "Irán", "Australia", "Colombia", "Indonesia", "Polonia" ), las=0,
  col = palette("Polychrome 36"), ylab="Tamaño", cex.lab=0.8,
ylim = c(0, 35000),
  main = "Tamaño de las secuencias de las variantes", cex.names = 0.7,
las=2)
```

Tamaño de las secuencias de las variantes



Esta gráfica representa la longitud de las secuencias de la variante del SARS COV-2 de los 20 países con mas casos de Covid-19.

```
Por_EUA = Porcentaje_ADN(EUA)
Por_EUA
```

```
## [1] "A=29.880572 G=19.636669 C=18.351556 T=32.131203 N=0.000000 -=0.000000"
```

```
Por_India = Porcentaje_ADN(India)
Por_India
```

```
## [1] "A=29.868823 G=19.619552 C=18.378233 T=32.133391 N=0.000000 -=0.000000"
```

```
Por_Brasil = Porcentaje_ADN(Brasil)
Por_Brasil
```

```
## [1] "A=29.866863 G=19.604950 C=18.354070 T=32.174117 N=0.000000 -=0.000000"
```

```
Por_Francia = Porcentaje_ADN(Francia)
Por_Francia
```

```
## [1] "A=29.878049 G=19.609354 C=18.349638 T=32.146207 N=0.000000 -=0.000000"
```

```
Por_Alemania = Porcentaje_ADN(Alemania)
Por_Alemania
```

```
## [1] "A=29.655311 G=19.445282 C=18.192476 T=31.812548 N=0.894382 ==0.000000"
```

```
Por_UK = Porcentaje_ADN(UK)
Por_UK
```

```
## [1] "A=29.946828 G=19.596696 C=18.349329 T=32.107146 N=0.000000 ==0.000000"
```

```
Por_Rusia = Porcentaje_ADN(Rusia)
Por_Rusia
```

```
## [1] "A=29.657705 G=19.439391 C=18.246749 T=31.945287 N=0.710869 ==0.000000"
```

```
Por_CS = Porcentaje_ADN(CS)
Por_CS
```

```
## [1] "A=29.256732 G=19.201936 C=17.904326 T=31.421656 N=2.215349 ==0.000000"
```

```
Por_Italia = Porcentaje_ADN(Italia)
Por_Italia
```

```
## [1] "A=29.667330 G=19.488355 C=18.254744 T=31.864910 N=0.724662 ==0.000000"
```

```
Por_Turquía = Porcentaje_ADN(Turquía)
Por_Turquía
```

```
## [1] "A=28.838321 G=18.948396 C=17.733194 T=31.056653 N=3.423436 ==0.000000"
```

```
Por_España = Porcentaje_ADN(España)
Por_España
```

```
## [1] "A=29.936455 G=19.605351 C=18.347826 T=32.110368 N=0.000000 ==0.000000"
```

```
Por_Vietnam = Porcentaje_ADN(Vietnam)
Por_Vietnam
```

```
## [1] "A=29.872903 G=19.606846 C=18.268090 T=32.252161 N=0.000000 ==0.000000"
```

```
Por_Argentina = Porcentaje_ADN(Argentina)
Por_Argentina
```

```
## [1] "A=29.854858 G=19.637817 C=18.317430 T=32.189894 N=0.000000 ==0.000000"
```



```
Por_PB = Porcentaje_ADN(PB)
Por_PB
```

```
## [1] "A=29.856610 G=19.638000 C=18.365291 T=32.140099 N=0.000000 -=0.000000"
```

```
Por_Japón = Porcentaje_ADN(Japon)
Por_Japón
```

```
## [1] "A=29.866953 G=19.625323 C=18.371929 T=32.135795 N=0.000000 -=0.000000"
```

```
Por_Irán = Porcentaje_ADN(Iran)
Por_Irán
```

```
## [1] "A=26.945612 G=17.560488 C=16.471359 T=28.689598 N=10.332943 -=0.000000"
```

```
Por_Australia = Porcentaje_ADN(Australia)
Por_Australia
```

```
## [1] "A=29.309650 G=19.296891 C=18.044456 T=31.602982 N=1.746021 -=0.000000"
```

```
Por_Colombia = Porcentaje_ADN(Colombia)
Por_Colombia
```

```
## [1] "A=29.888287 G=19.603927 C=18.341232 T=32.166554 N=0.000000 -=0.000000"
```

```
Por_Indonesia = Porcentaje_ADN(Indonesia)
Por_Indonesia
```

```
## [1] "A=29.876286 G=19.632892 C=18.331877 T=32.155584 N=0.003362 -=0.000000"
```

```
Por_Polonia = Porcentaje_ADN(Polonia)
Por_Polonia
```

```
## [1] "A=29.394011 G=19.333311 C=18.045541 T=31.510111 N=1.717026 -=0.000000"
```

Procentaje por base nitrogenada

```
EUA_a = Porcentaje_A_ADN(EUA)
EUA_c = Porcentaje_C_ADN(EUA)
EUA_t = Porcentaje_T_ADN(EUA)
EUA_g = Porcentaje_G_ADN(EUA)
```

```
India_a = Porcentaje_A_ADN(India)
India_c = Porcentaje_C_ADN(India)
```

```

India_t = Porcentaje_T_ADN(India)
India_g = Porcentaje_G_ADN(India)

Brasil_a = Porcentaje_A_ADN(Brasil)
Brasil_c = Porcentaje_C_ADN(Brasil)
Brasil_t = Porcentaje_T_ADN(Brasil)
Brasil_g = Porcentaje_G_ADN(Brasil)

Francia_a = Porcentaje_A_ADN(Francia)
Francia_c = Porcentaje_C_ADN(Francia)
Francia_t = Porcentaje_T_ADN(Francia)
Francia_g = Porcentaje_G_ADN(Francia)

Alemania_a = Porcentaje_A_ADN(Alemania)
Alemania_c = Porcentaje_C_ADN(Alemania)
Alemania_t = Porcentaje_T_ADN(Alemania)
Alemania_g = Porcentaje_G_ADN(Alemania)

UK_a = Porcentaje_A_ADN(UK)
UK_c = Porcentaje_C_ADN(UK)
UK_t = Porcentaje_T_ADN(UK)
UK_g = Porcentaje_G_ADN(UK)

Rusia_a = Porcentaje_A_ADN(Rusia)
Rusia_c = Porcentaje_C_ADN(Rusia)
Rusia_t = Porcentaje_T_ADN(Rusia)
Rusia_g = Porcentaje_G_ADN(Rusia)

CS_a = Porcentaje_A_ADN(CS)
CS_c = Porcentaje_C_ADN(CS)
CS_t = Porcentaje_T_ADN(CS)
CS_g = Porcentaje_G_ADN(CS)

Italia_a = Porcentaje_A_ADN(Italia)
Italia_c = Porcentaje_C_ADN(Italia)
Italia_t = Porcentaje_T_ADN(Italia)
Italia_g = Porcentaje_G_ADN(Italia)

Turquia_a = Porcentaje_A_ADN(Turquía)
Turquia_c = Porcentaje_C_ADN(Turquía)
Turquia_t = Porcentaje_T_ADN(Turquía)
Turquia_g = Porcentaje_G_ADN(Turquía)

España_a = Porcentaje_A_ADN(España)
España_c = Porcentaje_C_ADN(España)
España_t = Porcentaje_T_ADN(España)
España_g = Porcentaje_G_ADN(España)

Vietnam_a = Porcentaje_A_ADN(Vietnam)

```

```
Vietnam_c = Porcentaje_C_ADN(Vietnam)
Vietnam_t = Porcentaje_T_ADN(Vietnam)
Vietnam_g = Porcentaje_G_ADN(Vietnam)
```

```
Argentina_a = Porcentaje_A_ADN(Argentina)
Argentina_c = Porcentaje_C_ADN(Argentina)
Argentina_t = Porcentaje_T_ADN(Argentina)
Argentina_g = Porcentaje_G_ADN(Argentina)
```

```
PB_a = Porcentaje_A_ADN(PB)
PB_c = Porcentaje_C_ADN(PB)
PB_t = Porcentaje_T_ADN(PB)
PB_g = Porcentaje_G_ADN(PB)
```

```
Japon_a = Porcentaje_A_ADN(Japon)
Japon_c = Porcentaje_C_ADN(Japon)
Japon_t = Porcentaje_T_ADN(Japon)
Japon_g = Porcentaje_G_ADN(Japon)
```

```
Iran_a = Porcentaje_A_ADN(Iran)
Iran_c = Porcentaje_C_ADN(Iran)
Iran_t = Porcentaje_T_ADN(Iran)
Iran_g = Porcentaje_G_ADN(Iran)
```

```
Australia_a = Porcentaje_A_ADN(Australia)
Australia_c = Porcentaje_C_ADN(Australia)
Australia_t = Porcentaje_T_ADN(Australia)
Australia_g = Porcentaje_G_ADN(Australia)
```

```
Colombia_a = Porcentaje_A_ADN(Colombia)
Colombia_c = Porcentaje_C_ADN(Colombia)
Colombia_t = Porcentaje_T_ADN(Colombia)
Colombia_g = Porcentaje_G_ADN(Colombia)
```

```
Colombia_a = Porcentaje_A_ADN(Colombia)
Colombia_c = Porcentaje_C_ADN(Colombia)
Colombia_t = Porcentaje_T_ADN(Colombia)
Colombia_g = Porcentaje_G_ADN(Colombia)
```

```
Indonesia_a = Porcentaje_A_ADN(Indonesia)
Indonesia_c = Porcentaje_C_ADN(Indonesia)
Indonesia_t = Porcentaje_T_ADN(Indonesia)
Indonesia_g = Porcentaje_G_ADN(Indonesia)
```

```
Polonia_a = Porcentaje_A_ADN(Polonia)
Polonia_c = Porcentaje_C_ADN(Polonia)
Polonia_t = Porcentaje_T_ADN(Polonia)
Polonia_g = Porcentaje_G_ADN(Polonia)
```

```
#par(mai = c(.1, .1,1 ,1))
Entries = c(EUA_a, EUA_c, EUA_t, EUA_g, India_a, India_c, India_t, India_g,
            Brasil_a, Brasil_c, Brasil_t, Brasil_g, Francia_a, Francia_c,
            Francia_t, Francia_g,
```

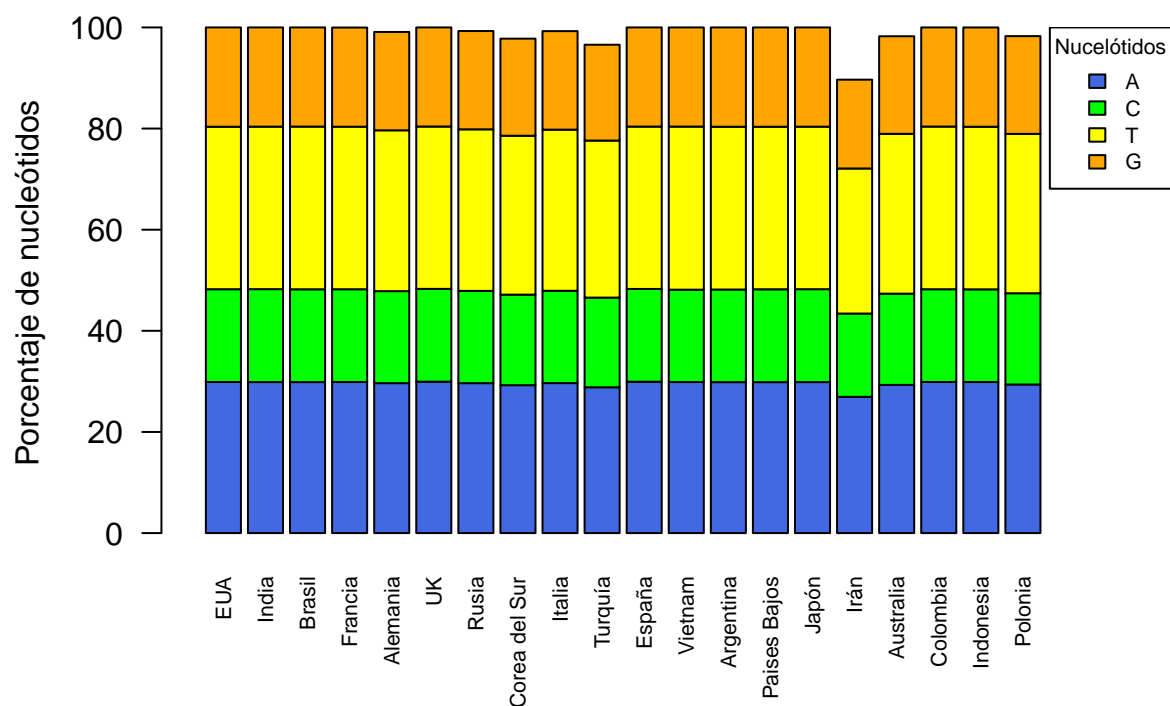
```

Alemania_a, Alemania_c, Alemania_t, Alemania_g,
UK_a, UK_c, UK_t, UK_g, Rusia_a, Rusia_c, Rusia_t, Rusia_g,
CS_a, CS_c, CS_t, CS_g,
Italia_a, Italia_c, Italia_t, Italia_g,
Turquia_a, Turquia_c, Turquia_t, Turquia_g, España_a, España_c,
España_t, España_g, Vietnam_a, Vietnam_c, Vietnam_t, Vietnam_g,
Argentina_a, Argentina_c, Argentina_t, Argentina_g, PB_a, PB_c, PB_t,
PB_g, Japon_a, Japon_c, Japon_t, Japon_g, Iran_a, Iran_c, Iran_t,
Iran_g, Australia_a, Australia_c, Australia_t, Australia_g,
Colombia_a, Colombia_c, Colombia_t, Colombia_g, Indonesia_a,
Indonesia_c, Indonesia_t, Indonesia_g, Polonia_a, Polonia_c,
Polonia_t, Polonia_g)
M = matrix(Entries, nrow = 4, byrow = FALSE)
barplot(M, width=c(0.3,0.3,0.3,0.3),
names.arg = c("EUA", "India", "Brasil", "Francia", "Alemania", "UK", "Rusia",
"Corea del Sur", "Italia", "Turquía", "España", "Vietnam", "Argentina",
"Países Bajos",
"Japón", "Irán", "Australia", "Colombia", "Indonesia", "Polonia" ),
col = c("royalblue", "green", "yellow", "Orange"), ylab=
"Porcentaje de nucleótidos",
main = "Porcentaje de nucleótidos en las secuencias de las variantes",
xlim=c(0,8), cex.names = 0.7, las=2)

legend(x ="topright", legend = c("A", "C", "T", "G" ), cex=0.7,
fill = c("royalblue", "green", "yellow", "Orange"),
title = "Nucleótidos")

```

Porcentaje de nucleótidos en las secuencias de las variantes



Este gráfico compara los porcentajes de bases nitrogenadas de la variante del SARS COV-2 de los 20 países con mas casos de Covid-19.