

PG1: S1 Exportaciones de cobre en el Perú en el 2021 y 2022

Grupo S1

2023-07-02

Contents

1. Estudio	2
1.1. Introducción	2
1.2. Relevancia	2
1.3. Objetivos	2
1.4. Planificación	3
1.5. Elaboración del informe	3
2. Datos	4
2.1. Recolección de datos	4
2.2. Reconocimiento de variables	4
3. Limpieza de datos	6
3.1. Decargamos los paquetes	6
3.2. Cargamos los paquetes	8
3.3. Leemos los archivos	8
3.4. Datos faltantes	8
a. Países Destino	9
b. Export	12
c. Minerales	14
d. Sector	16
e. Producto	19
4. Análisis descriptivo	22
4.1. Descriptores numéricos	22
4.2. Descriptores gráficos	27
4.3. Regresión lineal	47

5. Análisis Probabilístico	51
Probabilidades Empíricas	51
Probabilidades Condicionales	58
Distribuciones	70



Integrantes	Código	Porcentaje
Diego Orlando Bustamante Palomino(lider)	202120115	100%
Alejandra Pilco Rojas	202120443	100%
Brigitte Dayane Rojas Leon	202120474	100%
Yorman Cristian Taipe Flores	202120558	100%
Anayeli Mariel Quispe Quispe	202120744	100%

1. Estudio

1.1. Introducción

El Perú es uno de los principales países exportadores de minerales en el mundo, tuvo como años fructíferos **2021 y 2022** en la exportación de **cobre**, ese hecho fue comunicado por el **MINEN(Ministerio de Energía y Minas)**, por lo que vimos conveniente obtener datos acerca de estas exportaciones formando de esa manera los siguientes objetivos para este proyecto.

1.2.Relevancia

Este enfoque se centra en el análisis específico de las exportaciones de cobre, que es uno de los productos más importantes del sector minero peruano. Además, el enfoque en el análisis de las exportaciones permite explorar los patrones y tendencias en la venta de cobre al extranjero durante los dos años en cuestión.

1.3. Objetivos

1.3.1. Objetivo General

- Analizar las exportaciones de cobre del sector minero peruano en los años 2021 y 2022”

1.3.2. Objetivos específicos:

- Comparar el número de exportaciones del sector minero con respecto a otros sectores económicos
- Determinar a qué países de destino vendemos más cobre y qué cantidad.

- Comparar la cantidad de cobre que se exporta con respecto a otros minerales.
- Comparar la variación de números de registros de exportación del cobre durante los años 2021 y 2022 respectivamente.
- Identificar qué empresa exportadora de cobre genera mayores ingresos de divisas al Perú.
- Comparar el precio del cobre promedio mensual por toneladas entre los años 2021 y 2022.

1.4. Planificación

```
tinytex::tlmgr_update()
```

```
## tlmgr update --all --self
```

- Para poder cumplir con los objetivos del estudio es necesario organizar las actividades a través de algún gráfico, en este caso empleamos el gráfico de Gantt para poder organizar las actividades a desempeñar para la entrega PG1 del ciclo 2023-1.

1.5. Elaboración del informe

```
if (!require("ggplot2", character.only = TRUE)) {
  install.packages("ggplot2")
}
```

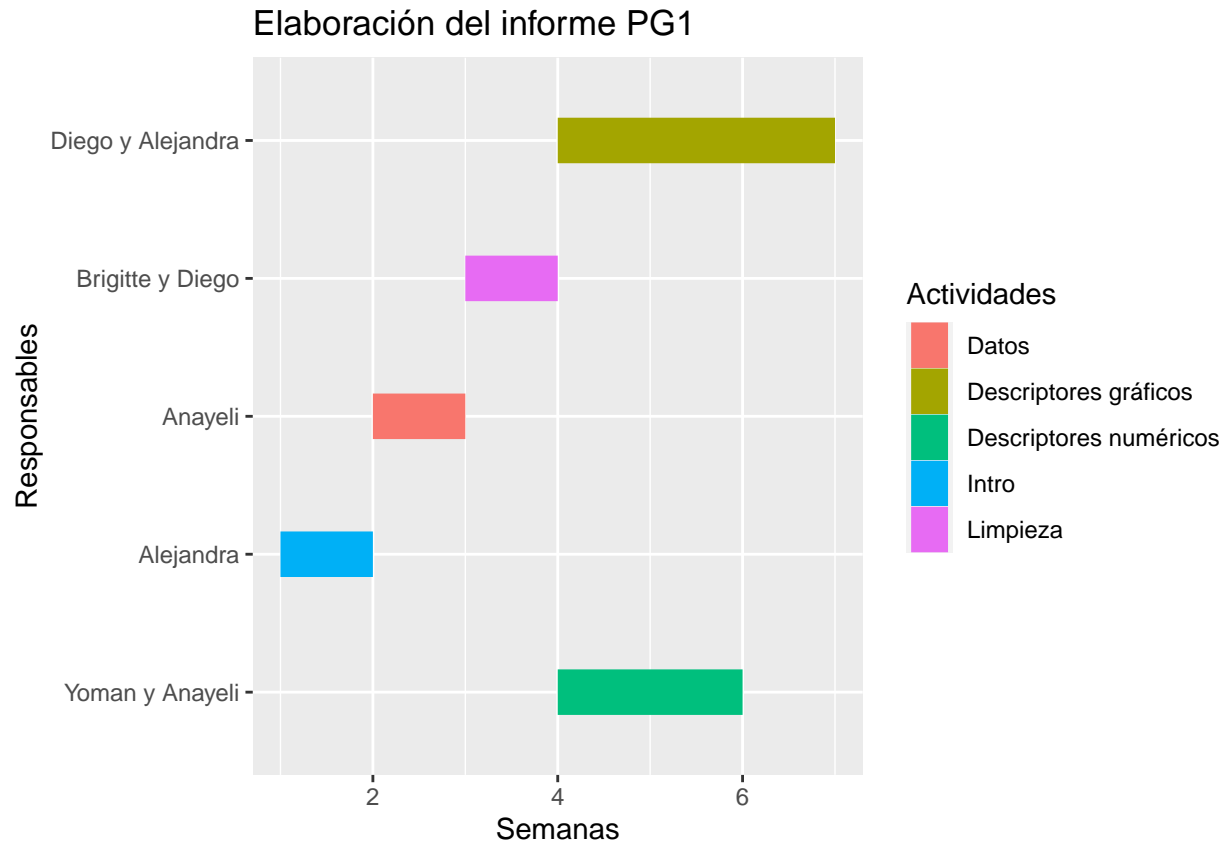
```
## Loading required package: ggplot2
```

```
if (!require("ggthemes", character.only = TRUE)) {
  install.packages("ggthemes")
}
```

```
## Loading required package: ggthemes
```

```
ggplot(data, aes(x=start, xend=end,
                  y=name, yend=name, color= Actividades)) +
  theme_bw() +
  geom_segment(size=8) +
  labs(title='Elaboración del informe PG1', x='Semanas', y='Responsables') +
  theme_economist() +
  theme_update(plot.title = element_text(hjust = 0.5), axis.title = element_text(hjust=0.5))
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



2. Datos

2.1. Recolección de datos

Utilizamos los datos de exportaciones de **cobre** tanto en el año 2021 y 2022, otra carpeta de **países destino** en el cual se encuentra la información de todos los países a las cuales exportamos cobre y finalmente se encuentra la carpeta de **sectores económicos** el cual nos servirá para cumplir uno de los objetivos planteados.

2.2. Reconocimiento de variables

Identificamos diferentes tipos de cada variable categóricas y numéricas, con sus respectivas restricciones.

2.2.1. Variables Cuantitativas

Variable	Tipo	Restricción(Característica)
Cantidad de minerales exportados		Es continua porque se puede medir en una escala de valores numéricos infinita

durante 2021-2022 en *kilogramos*	Continua	y no hay saltos o interrupciones en los valores posibles.
Cantidad de cobre exportado en *kilogramos*	Continua	Porque representa una cantidad numérica medible en una escala y la medida está en kilogramos.
Total de ingresos generados por la exportación de minerales por cada año en soles	Continua	Cuantitativa continua porque representa una cantidad numérica de ingresos en una escala.
Total de ingresos recaudados por la exportación de cobre en cada año en soles	Continua	Porque representa una cantidad numérica ingresos en una escala
Número total de empresas exportadoras de cobre	Discreta	Porque representa una cantidad numérica medible y la medida está en número de empresas exportadoras. Además, es discreta porque los valores posibles son números enteros y no hay valores fraccionarios o continuos.
Número total de registro de exportaciones	Discreta	Ya que son cantidades exactas en el que no incluye números fraccionarios.
Total de minerales exportados	Discreta	Es una variable cuantitativa discreta ya que son cantidades exactas.
Años 2021 y 2022.	Discreta	Representa a un tiempo determinado. Asimismo, el análisis del año 2021, permitiendo la comparación respecto al siguiente año.

2.2.2. Variables Cualitativas

Variable	Tipo	Restricción(Característica)
Países de destino	Nominal	Es una variable categórica nominal ya que los países se pueden clasificar en diferentes categorías sin una escala numérica que los ordene
Empresas exportadoras.	Nominal	Porque se puede analizar cuales fueron las empresas que exportaron mas y cuales no
Minerales.	Nominal	Es una variable categórica nominal porque se pueden clasificar en diferentes categorías, por lo que se pueden analizar y comparar las distintas cualidades que estos presentan..

Sectores económicos.	Nominal	Porque se puede clasificar en diferentes categorías y con ello podemos analizar la estructura económica del sector minero en el año 2021, y 2022.
+-----+-----+-----+		
Meses del año.	Ordinal	Es una variable categórica ordinal ya que tiene una escala numérica que permite ordenar las categorías en una jerarquía.
+-----+-----+-----+		
Empresas con más divisas.	Ordinal	ya que se pueden ordenar las empresas según la cantidad de divisas que generan o manejan, es decir, hay una jerarquía de por medio.
+-----+-----+-----+		

3. Limpieza de datos

Explicaremos de manera concisa todos los criterios utilizados para poder garantizar que nuestra base de datos final, sirve para cumplir con los objetivos del proyecto.

3.1. Decargamos los paquetes

```
if (!require("readr", character.only = TRUE)) {
  install.packages("readr")
}
```

```
## Loading required package: readr
```

```
if (!require("readxl", character.only = TRUE)) {
  install.packages("readxl")
}
```

```
## Loading required package: readxl
```

```
if (!require("dplyr", character.only = TRUE)) {
  install.packages("dplyr")
}
```

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

```
if (!require("stringr", character.only = TRUE)) {  
  install.packages("stringr")  
}
```

Loading required package: stringr

```
if (!require("plotrix", character.only = TRUE)) {  
  install.packages("plotrix")  
}
```

Loading required package: plotrix

```
if (!require("scales", character.only = TRUE)) {  
  install.packages("scales")  
}
```

Loading required package: scales

##

Attaching package: 'scales'

The following object is masked from 'package:plotrix':

##

rescale

The following object is masked from 'package:readr':

##

col_factor

```
if (!require("ggplot2", character.only = TRUE)) {  
  install.packages("ggplot2")  
}  
if (!require("ggthemes", character.only = TRUE)) {  
  install.packages("ggthemes")  
}  
if (!require("mvtnorm", character.only = TRUE)) {  
  install.packages("mvtnorm")  
}
```

Loading required package: mvtnorm

Warning: package 'mvtnorm' was built under R version 4.3.1

```
if (!require("rriskDistributions", character.only = TRUE)) {  
  install.packages("rriskDistributions")  
}
```

Loading required package: rriskDistributions

Warning: package 'rriskDistributions' was built under R version 4.3.1

3.2. Cargamos los paquetes

Es necesario que corra el siguiente fragmento de código para evitar errores en la ejecución de los gráficos y otros. Las librerías son **readr**, **dplyr**, **readxl**,***stringr**.

```
library(dplyr);
library(readr);
library(readxl);
library(stringr);
library(plotrix);
library(scales);
library(ggplot2);
library(ggthemes);
library(mvtnorm)
library(rriskDistributions)
```

```
library(rriskDistributions)
```

3.3. Leemos los archivos

Se asignó un nombre a cada base de datos que se usa en el estudio. Usamos el comando **read_csv2** y **read_excel** para poder manipular los datos del archivo.

```
#destinos <- read_excel("destino3.xlsx")
export <- read_excel("export.xlsx")
items <- read_excel("items.xlsx")
minerales <- read_excel("minerales.xlsx")
sector <- read_excel("sector.xlsx")
países<- read.csv2("destino31.csv",header =T,sep =';')
```

3.4. Datos faltantes

Primero, se verificó que no se cuente con datos o casos faltantes en nuestra base de datos. Para verificar, se usaron los comandos **is.na** y **!complete.cases**. Se verificó que no hay ni datos ni casos faltantes.

Para export:

```
datos_faltantes <- sum(is.na(export))
casos_faltantes <- sum(!complete.cases(export))
```

Para items:

```
datos_faltantes1 <- sum(is.na(items))
casos_faltantes1 <- sum(!complete.cases(items))
```

Para minerales:

```
datos_faltantes2 <- sum(is.na(minerales))
casos_faltantes2 <- sum(!complete.cases(minerales))
```

Hacemos un recorte a la fila NA(innecesaria de Sector):


```
sector = sector[-c(1), ] #Borro el NA que estobarba de sector
```

```
datos_faltantes3 <- sum(is.na(sector))
casos_faltantes3 <- sum(!complete.cases(sector))
```

a. Países Destino

1. Limpieza de NA's de columnas y filas innecesarias de Paises(Destinos)

```
países = países[-c(260:993), ] #Borro las filas que quedaron en NA

países = select(países,País.Destino,Año,Meses,Total.registros,Total.KG,US.KG,US)
##Veo los nombres de países
names(países)
```

```
## [1] "País.Destino"      "Año"               "Meses"             "Total.registros"
## [5] "Total.KG"         "US.KG"             "US"
```

2. Rename de Base de datos **Países Destinos** y segundo vuelvo a tomar las columnas que necesito (solo por seguridad)

```
países = rename(países,Pais_Destino = "País.Destino",Anio = "Año",Total_Registros = "Total.registros",T

países = select(países,Pais_Destino,Anio,Meses,Total_Registros,Total_KG,US_KG,US_Dolar)
names(países)
```

```
## [1] "Pais_Destino"      "Anio"              "Meses"             "Total_Registros"
## [5] "Total_KG"         "US_KG"             "US_Dolar"
```

3. Verifico los cambios

```
names(países)
```

```
## [1] "Pais_Destino"      "Anio"              "Meses"             "Total_Registros"
## [5] "Total_KG"         "US_KG"             "US_Dolar"
```

```
str(países)
```

```
## 'data.frame':   259 obs. of  7 variables:
## $ Pais_Destino   : chr  "CHINA" "JAPÓN" "ALEMANIA" "POLONIA" ...
## $ Anio           : int   2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
## $ Meses          : chr   " Enero " " Enero " " Enero " " Enero " ...
## $ Total_Registros: num   58 1 2 1 2 1 1 1 1 5 ...
## $ Total_KG       : chr   "429876939.00" "31501980.00" "20521552.00" "20002800.00" ...
## $ US_KG          : chr   "2.018" "1.557" "2.252" "2.302" ...
## $ US_Dolar       : chr   "867384807" "49037230" "46224769" "46047515" ...
```

4. Transformo las variables que necesitan ser convertidas a números o enteros.

- Total Registros -> integer
- Total_KG -> numeric
- US_KG -> numeric
- US_Dolar -> DOLAR

```
países$Total_Registros = as.integer(países$Total_Registros)
países$Total_KG = as.numeric(países$Total_KG)
países$US_KG = as.numeric(países$US_KG)
países$US_Dolar = as.numeric(países$US_Dolar)
```

```
str(países)
```

```
## 'data.frame': 259 obs. of 7 variables:
## $ Pais_Destino : chr "CHINA" "JAPÓN" "ALEMANIA" "POLONIA" ...
## $ Anio : int 2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
## $ Meses : chr " Enero " " Enero " " Enero " " Enero " ...
## $ Total_Registros: int 58 1 2 1 2 1 1 1 1 5 ...
## $ Total_KG : num 4.30e+08 3.15e+07 2.05e+07 2.00e+07 2.10e+07 ...
## $ US_KG : num 2.02 1.56 2.25 2.3 1.84 ...
## $ US_Dolar : num 8.67e+08 4.90e+07 4.62e+07 4.60e+07 3.85e+07 ...
```

5. Esta es otra forma de hacerlo para el cambio del tipo de variable de “Total de registros”, de “*double*” a “*int*”

```
países <- mutate(países, `Total_Registros` = as.integer(`Total_Registros`))
```

```
str(países)
```

```
## 'data.frame': 259 obs. of 7 variables:
## $ Pais_Destino : chr "CHINA" "JAPÓN" "ALEMANIA" "POLONIA" ...
## $ Anio : int 2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
## $ Meses : chr " Enero " " Enero " " Enero " " Enero " ...
## $ Total_Registros: int 58 1 2 1 2 1 1 1 1 5 ...
## $ Total_KG : num 4.30e+08 3.15e+07 2.05e+07 2.00e+07 2.10e+07 ...
## $ US_KG : num 2.02 1.56 2.25 2.3 1.84 ...
## $ US_Dolar : num 8.67e+08 4.90e+07 4.62e+07 4.60e+07 3.85e+07 ...
```

6. Verificamos cada una de las variables para detectar anomalías

```
summary(países$Meses)
```

```
## Length Class Mode
## 259 character character
```

```
print("Total Registros Range:")
```

```
## [1] "Total Registros Range:"
```

```
range(países$Total_Registros)
```

```
## [1] 1 124
```

```
print("Total KG Range:")
```

```
## [1] "Total KG Range:"
```

```
range(países$Total_KG)
```

```
## [1] 9.5 615426327.0
```

```
print("US_KG Range:")
```

```
## [1] "US_KG Range:"
```

```
range(países$US_KG)
```

```
## [1] 0.000 57.925
```

```
print("US_DOLAR Range:")
```

```
## [1] "US_DOLAR Range:"
```

```
range(países$US_Dolar)
```

```
## [1] 0 1304664438
```

7. Verificamos los tipos de variable que tiene nuestra base de datos

```
str(países)
```

```
## 'data.frame': 259 obs. of 7 variables:
## $ Pais_Destino : chr "CHINA" "JAPÓN" "ALEMANIA" "POLONIA" ...
## $ Anio : int 2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
## $ Meses : chr " Enero " " Enero " " Enero " " Enero " ...
## $ Total_Registros: int 58 1 2 1 2 1 1 1 5 ...
## $ Total_KG : num 4.30e+08 3.15e+07 2.05e+07 2.00e+07 2.10e+07 ...
## $ US_KG : num 2.02 1.56 2.25 2.3 1.84 ...
## $ US_Dolar : num 8.67e+08 4.90e+07 4.62e+07 4.60e+07 3.85e+07 ...
```

8. Verificamos si hay casos incompletos en Países Destinos:

```
sum(is.na(países))
```

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

```
sum(!complete.cases(paises))
```

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

9. Unique Paises

```
unique(paises$Anio)
```

```
## [1] 2021 2022
```

```
unique(paises$Meses)
```

```
## [1] " Enero "      " Febrero "     " Marzo "       " Abril "       " Mayo "
## [6] " Junio "      " Julio "       " Agosto "      " Septiembre "  " Octubre "
## [11] " Noviembre "   " Diciembre "
```

```
unique(paises$Total_Registros)
```

```
## [1] 58  1  2  5  3 75  4  6 47  7 17 82 12 92 15 78 11 95 10
## [20] 88 96 103 9 80 84 14 81 13 74 124 85 76
```

b. Export

```
names(export)
```

```
## [1] "Exportador"      "AÑO"           "Meses"         "Total registros"
## [5] "Total_KG"        "US$ / KG"      "US$"           "
```

1. Renombro en **export**

```
export = rename(export, Anio = "AÑO", Total_Registros = "Total registros", Total_KG = "Total KG", US_KG = "
names(export)
```

```
## [1] "Exportador"      "Anio"          "Meses"         "Total_Registros"
## [5] "Total_KG"        "US_KG"         "US_Dolar"      "
```

```
str(export)
```

```
## tibble [527 x 7] (S3: tbl_df/tbl/data.frame)
## $ Exportador      : chr [1:527] "SOCIEDAD MINERA CERRO VERDE S.A.A." "COMPAÑIA MINERA ANTAMINA S.A."
## $ Anio            : num [1:527] 2022 2022 2022 2022 2022 ...
## $ Meses           : chr [1:527] "Enero" "Enero" "Enero" "Enero" ...
## $ Total_Registros: num [1:527] 20 10 12 7 16 5 2 3 1 1 ...
## $ Total_KG        : num [1:527] 1.83e+08 1.02e+08 9.28e+07 5.50e+07 7.18e+07 ...
## $ US_KG           : num [1:527] 1.96 2.41 2.12 2.96 1.72 ...
## $ US_Dolar        : num [1:527] 3.59e+08 2.45e+08 1.97e+08 1.63e+08 1.23e+08 ...
```

2. Convierto Total_Registros y Anio a -> **Integer**

```
export <- mutate(export, `Total_Registros` = as.integer(`Total_Registros`), `Anio` = as.integer(`Anio`))
```

3. Verificamos cada una de las variables para detectar anomalías

```
summary(export$Exportador)
```

```
##      Length      Class      Mode  
##      527 character character
```

```
summary(export$Meses)
```

```
##      Length      Class      Mode  
##      527 character character
```

```
range(export$Total_Registros)
```

```
## [1]  1 34
```

```
range(export$Total_KG)
```

```
## [1]      15.57 222294770.10
```

```
range(export$US_KG)
```

```
## [1]  0.00000 57.92453
```

```
range(export$US_Dolar)
```

```
## [1]      0 509358565
```

```
str(export)
```

```
## tibble [527 x 7] (S3: tbl_df/tbl/data.frame)  
## $ Exportador      : chr [1:527] "SOCIEDAD MINERA CERRO VERDE S.A.A." "COMPAÑIA MINERA ANTAMINA S.A."  
## $ Anio            : int [1:527] 2022 2022 2022 2022 2022 2022 2022 2022 2022 2022 ...  
## $ Meses           : chr [1:527] "Enero" "Enero" "Enero" "Enero" ...  
## $ Total_Registros: int [1:527] 20 10 12 7 16 5 2 3 1 1 ...  
## $ Total_KG        : num [1:527] 1.83e+08 1.02e+08 9.28e+07 5.50e+07 7.18e+07 ...  
## $ US_KG           : num [1:527] 1.96 2.41 2.12 2.96 1.72 ...  
## $ US_Dolar        : num [1:527] 3.59e+08 2.45e+08 1.97e+08 1.63e+08 1.23e+08 ...
```

4. Verificamos si hay casos incompletos:

```
sum(is.na(export))
```

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

```
sum(!complete.cases(export))
```

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

c. Minerales

```
str(minerales)
```

```
## tibble [192 x 4] (S3: tbl_df/tbl/data.frame)
## $ Minerales: chr [1:192] "Cobre" "Hierro" "Plata refinada" "Plomo 1/" ...
## $ Año      : num [1:192] 2021 2021 2021 2021 2021 ...
## $ Meses    : chr [1:192] "Enero" "Enero" "Enero" "Enero" ...
## $ US $     : num [1:192] 1.27e+09 2.36e+08 9.54e+06 1.41e+08 1.72e+08 ...
```

```
names(minerales)
```

```
## [1] "Minerales" "Año"      "Meses"    "US $"
```

1. Rename a la data Minerales (solo los necesarios)

```
minerales = rename(minerales, US_Dolar = "US $", Anio = "Año");
```

2. Convertimos Año -> Integer

```
minerales$Anio = as.integer(minerales$Anio);
```

3. Verificamos cada una de las variables para detectar anomalías

```
summary(minerales$Minerales)
```

```
##      Length      Class      Mode
##      192 character character
```

```
summary(minerales$Meses)
```

```
##      Length      Class      Mode
##      192 character character
```

```
range(minerales$Anio)
```

```
## [1] 2021 2022
```

```
range(minerales$US_Dolar)
```

```
## [1] 4104611 2159793794
```

4. Verificamos si hay casos incompletos:

```
sum(is.na(minerales))
```

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

```
sum(!complete.cases(minerales))
```

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

```
unique(minerales$Minerales)
```

```
## [1] "Cobre"          "Hierro"          "Plata refinada" "Plomo 1/"
## [5] "Zinc"           "Oro"             "Estaño 2/"      "Resto 3/"
## [9] "Resto 3/"
```

```
unique(minerales$Anio)
```

```
## [1] 2021 2022
```

```
unique(minerales$Meses)
```

```
## [1] "Enero"      "Febrero"      "Marzo"      "Abril"      "Mayo"      "Junio"
## [7] "Julio"      "Agosto"      "Setiembre"  "Octubre"    "Noviembre" "Diciembre"
```

```
unique(minerales$US_Dolar)
```

```
## [1] 1271380805 236175153 9539442 141239456 171642856 621840631
## [7] 62725397 45856994 1532128038 187481955 6444505 132582957
## [13] 203554543 520716142 52828610 45093384 1569462442 202451987
## [19] 10997623 166127440 229918529 609249642 51484134 84367980
## [25] 1592925589 170130460 7306049 172190437 197690038 652911525
## [31] 57785586 58466059 1822248191 241613617 12140175 254343235
## [37] 231808151 642652947 32438411 58644748 1628898029 237568386
## [43] 10977475 127173664 230697240 728320701 50968253 91253839
## [49] 1627305592 124321362 11512510 218004166 196934990 689319745
## [55] 66731933 94848115 1911794012 243152887 9462459 158064442
## [61] 280719934 675073103 106745944 110934249 1787742720 124004919
## [67] 10395073 160685242 197551987 710651469 92301219 123979182
## [73] 1799002214 188751774 8509710 123079475 177595099 642168660
## [79] 120528254 82697720 2159793794 143948026 8382532 207042160
## [85] 291956681 689583261 60874254 92064808 1930152665 157182178
## [91] 11360981 168207851 318936784 632695609 123466298 145679050
## [97] 1663106862 155735271 7624868 111019889 203706168 596851825
## [103] 74987231 65218081 1785184300 187671310 7437963 133829859
## [109] 259094020 648286952 62099466 113551152 1258170465 180203931
## [115] 8507914 177409951 250344949 718317858 69333674 88613069
## [121] 1579964424 157364248 9386274 189461839 206358885 637564104
## [127] 67186519 101458961 1289462679 140906853 6284345 91061041
## [133] 213695757 599456093 57502926 74580840 1797520962 145781804
## [139] 7510809 173046703 193006123 685554349 61643140 84481392
```

```
## [145] 1249647188 117577782 6345854 136179200 220011053 575890960
## [151] 60042786 81703728 1516171080 121914349 6895019 154355544
## [157] 167136700 629462743 71422374 58929664 1307838034 118317837
## [163] 7565168 138832927 150042611 645975045 64481962 60425637
## [169] 1124375064 127389187 7184611 106054808 147727180 640586811
## [175] 45051923 55521028 853227557 103990499 6500327 111997604
## [181] 79867108 562733825 71471366 46358652 1052912315 4104611
## [187] 8986908 148197881 94348345 594630225 62070055 34964037
```

d. Sector

```
str(sector)
```

```
## tibble [28 x 14] (S3: tbl_df/tbl/data.frame)
## $ Sector : chr [1:28] "Minero" "Minero" "Minero" "Minero" ...
## $ Producto : chr [1:28] "Cobre" "Hierro" "Plata refinada" "Plomo" ...
## $ Enero : num [1:28] 1.27e+09 2.36e+08 9.54e+06 1.41e+08 1.72e+08 ...
## $ Febrero : num [1:28] 1.53e+09 1.87e+08 6.44e+06 1.33e+08 2.04e+08 ...
## $ Marzo : num [1:28] 1.57e+09 2.02e+08 1.10e+07 1.66e+08 2.30e+08 ...
## $ Abril : num [1:28] 1.59e+09 1.70e+08 7.31e+06 1.72e+08 1.98e+08 ...
## $ Mayo : num [1:28] 1.82e+09 2.42e+08 1.21e+07 2.54e+08 2.32e+08 ...
## $ Junio : num [1:28] 1.63e+09 2.38e+08 1.10e+07 1.27e+08 2.31e+08 ...
## $ Julio : num [1:28] 1.63e+09 1.24e+08 1.15e+07 2.18e+08 1.97e+08 ...
## $ Agosto : num [1:28] 1.91e+09 2.43e+08 9.46e+06 1.58e+08 2.81e+08 ...
## $ Setiembre: num [1:28] 1.79e+09 1.24e+08 1.04e+07 1.61e+08 1.98e+08 ...
## $ Octubre : num [1:28] 1.80e+09 1.89e+08 8.51e+06 1.23e+08 1.78e+08 ...
## $ Noviembre: num [1:28] 2.16e+09 1.44e+08 8.38e+06 2.07e+08 2.92e+08 ...
## $ Diciembre: num [1:28] 1.93e+09 1.57e+08 1.14e+07 1.68e+08 3.19e+08 ...
```

```
names(sector)
```

```
## [1] "Sector" "Producto" "Enero" "Febrero" "Marzo" "Abril"
## [7] "Mayo" "Junio" "Julio" "Agosto" "Setiembre" "Octubre"
## [13] "Noviembre" "Diciembre"
```

1. Verificamos cada una de las variables para detectara anomalías

```
summary(sector$Sector)
```

```
## Length Class Mode
## 28 character character
```

```
summary(sector$Producto)
```

```
## Length Class Mode
## 28 character character
```

2. Verificamos si hay casos incompletos y omite NA en caso haya quedado alguno:


```
sum(is.na(sector))# verificamos si hay datos incompletos y se obtuvo 14
```

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

```
sum(!complete.cases(sector))# Se encontro una fila incompleta
```

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

```
sector=na.omit(sector)#Eliminamos la fila
```

```
sector=data.frame(sector, row.names = NULL)# Reordenamos el orden numerico de los datos
```

3. Unique Sector

```
unique(sector$Sector)
```

```
## [1] "Minero"           "Pesquero"           "Petróleo y derivados"
## [4] "Agrícolas"         "Resto"
```

```
unique(sector$Producto)
```

```
## [1] "Cobre"           "Hierro"           "Plata refinada"
## [4] "Plomo"          "Zinc"             "Oro"
## [7] "Estaño"         "Resto"            "Harina de pescado"
## [10] "Aceite de pescado" "Crudo"            "Derivados"
## [13] "Gas Natural"    "Algodón"          "Azúcar"
## [16] "Café"           "Resto 4/"         "Agropecuario"
## [19] "Textil"         "Pesquero"         "Quimico"
## [22] "Metal-Mecánica" "Sidero-Metalúrgico" "Minería no Metálica"
## [25] "Artesanías"     "Madera y papeles" "Pielles y cueros"
## [28] "Varios (inc. joyería)"
```

```
unique(sector$Enero)
```

```
## [1] 1.271381e+09 2.361752e+08 9.539442e+06 1.412395e+08 1.716429e+08
## [6] 6.218406e+08 6.272540e+07 4.585699e+07 1.731875e+08 2.532194e+07
## [11] 2.238210e+07 1.113505e+08 1.792475e+08 1.643676e+04 5.890743e+06
## [16] 3.145639e+07 1.516751e+06 7.320657e+08 9.930468e+07 9.842755e+07
## [21] 1.207274e+08 3.569863e+07 9.188837e+07 4.701471e+07 4.140237e+04
## [26] 1.742633e+07 2.939934e+05 1.783210e+07
```

```
unique(sector$Febrero)
```

```
## [1] 1532128038.0 187481954.7 6444504.8 132582956.7 203554543.1
## [6] 520716142.4 52828609.9 45093384.2 228134860.1 105455474.3
## [11] 3765208.9 80122331.3 60763479.0 252561.8 1467439.8
## [16] 13301757.6 2060467.9 536642713.4 112624345.8 134024461.6
## [21] 121498193.6 38100639.4 102946203.1 51870631.7 169446.0
## [26] 21241319.7 375073.0 14403510.0
```

```
unique(sector$Marzo)
```

```
## [1] 1569462442.2 202451986.6 10997623.5 166127439.8 229918529.1
## [6] 609249641.7 51484134.2 84367979.8 185748846.6 57034032.7
## [11] 5033645.9 127604105.0 66211083.0 401258.9 1541253.9
## [16] 8977005.3 2454734.8 436684467.6 116359854.2 150920026.2
## [21] 131629956.5 45465104.5 114980077.4 50783542.5 170677.0
## [26] 21534647.6 282121.3 17088949.7
```

```
unique(sector$Abril)
```

```
## [1] 1.592926e+09 1.701305e+08 7.306049e+06 1.721904e+08 1.976900e+08
## [6] 6.529115e+08 5.778559e+07 5.846606e+07 1.745660e+08 3.734322e+07
## [11] 3.605068e+07 8.362428e+07 5.443010e+07 4.677141e+04 1.903533e+06
## [16] 7.940800e+06 2.750786e+06 4.781114e+08 1.040259e+08 1.588702e+08
## [21] 1.544785e+08 4.916637e+07 1.313856e+08 4.812725e+07 3.376399e+05
## [26] 2.739177e+07 3.814470e+05 1.571103e+07
```

```
unique(sector$Mayo)
```

```
## [1] 1822248191.1 241613617.2 12140175.2 254343234.9 231808151.1
## [6] 642652946.5 32438411.3 58644748.4 98937421.8 32390718.2
## [11] 24019241.4 128557851.9 0.0 402045.7 1740096.3
## [16] 11206409.4 2630388.8 499629344.1 124732040.4 148012498.2
## [21] 163203175.7 43770395.1 104712220.6 43752154.6 279824.3
## [26] 23600148.9 432549.2 22920481.5
```

```
unique(sector$Junio)
```

```
## [1] 1628898028.5 237568385.5 10977475.3 127173664.0 230697240.0
## [6] 728320700.5 50968253.4 91253838.8 182468659.4 23741468.6
## [11] 0.0 165557382.9 74747394.5 108283.0 1593244.7
## [16] 29107897.1 2069664.0 514027183.6 116316458.0 139344913.9
## [21] 128577007.8 44120340.3 99228417.9 48887318.9 216036.5
## [26] 20314130.8 392227.6 20359131.7
```

```
unique(sector$Julio)
```

```
## [1] 1.627306e+09 1.243214e+08 1.151251e+07 2.180042e+08 1.969350e+08
## [6] 6.893197e+08 6.673193e+07 9.484812e+07 1.964712e+08 8.274521e+07
## [11] 4.000000e+01 1.791185e+08 0.000000e+00 2.149411e+04 2.190334e+06
## [16] 5.558322e+07 2.410015e+06 6.086048e+08 1.411159e+08 1.339113e+08
## [21] 1.699321e+08 4.931201e+07 1.417423e+08 6.355065e+07 8.847098e+04
## [26] 2.397860e+07 7.506943e+05 2.408337e+07
```

```
unique(sector$Agosto)
```

```
## [1] 1911794012.0 243152887.0 9462458.8 158064442.0 280719934.3
## [6] 675073102.6 106745943.8 110934249.5 246555324.0 66863320.3
```

```
## [11] 11567756.9 127729566.8 0.0 4412292.1 68399257.7
## [16] 6745426.7 654454129.3 142910992.1 109608702.3 166898879.5
## [21] 48598239.6 114929075.4 61755088.7 351822.6 24132919.8
## [26] 703117.9 22192247.7
```

```
unique(sector$Setiembre)
```

```
## [1] 1.787743e+09 1.240049e+08 1.039507e+07 1.606852e+08 1.975520e+08
## [6] 7.106515e+08 9.230122e+07 1.239792e+08 1.354103e+08 3.683004e+07
## [11] 5.830355e+07 1.608303e+08 1.849887e+08 1.747572e+04 1.303174e+07
## [16] 1.268251e+08 3.047045e+06 8.226221e+08 1.384776e+08 1.231174e+08
## [21] 1.613964e+08 4.963245e+07 1.341826e+08 5.807016e+07 2.697764e+05
## [26] 2.441242e+07 7.496613e+05 2.398733e+07
```

```
unique(sector$Octubre)
```

```
## [1] 1.799002e+09 1.887518e+08 8.509710e+06 1.230795e+08 1.775951e+08
## [6] 6.421687e+08 1.205283e+08 8.269772e+07 1.034906e+08 2.450729e+07
## [11] 5.770745e+07 1.800292e+08 2.665758e+08 0.000000e+00 1.416264e+06
## [16] 1.358902e+08 2.210565e+06 7.875783e+08 1.589622e+08 1.037244e+08
## [21] 2.118247e+08 6.247342e+07 1.337990e+08 5.787554e+07 8.768293e+04
## [26] 2.519295e+07 7.093665e+05 2.750763e+07
```

```
unique(sector$Noviembre)
```

```
## [1] 2159793793.6 143948026.1 8382532.1 207042160.3 291956681.2
## [6] 689583260.9 60874254.2 92064807.8 9869143.9 20522127.3
## [11] 49949650.5 157791240.7 376932430.9 392791.2 1023949.1
## [16] 134510266.5 2106160.5 811747074.1 145723698.1 87539163.0
## [21] 169959714.9 48117870.6 143541828.1 73886501.5 130856.6
## [26] 22261024.4 682589.5 22615064.8
```

```
unique(sector$Diciembre)
```

```
## [1] 1.930153e+09 1.571822e+08 1.136098e+07 1.682079e+08 3.189368e+08
## [6] 6.326956e+08 1.234663e+08 1.456790e+08 7.114520e+07 1.652190e+07
## [11] 2.087813e+07 2.194199e+08 4.394056e+08 1.543880e+03 1.248031e+07
## [16] 1.458303e+08 7.628813e+06 9.858206e+08 1.643687e+08 1.285027e+08
## [21] 1.994552e+08 5.505365e+07 1.318359e+08 6.911241e+07 4.710833e+04
## [26] 2.841207e+07 5.198083e+05 2.690753e+07
```

e. Producto

```
str(items)
```

```
## tibble [24 x 7] (S3: tbl_df/tbl/data.frame)
## $ Descripción Partida Aduanera: chr [1:24] "2603000000 MINERALES DE COBRE Y SUS CONCENTRADOS" "2603
## $ Año : num [1:24] 2021 2021 2021 2021 2021 ...
```

```
## $ Meses : chr [1:24] "Enero" "Febrero" "Marzo" "Abril" ...
## $ Total registros : num [1:24] 79 96 86 116 124 109 106 129 119 124 ...
## $ Total KG : num [1:24] 5.90e+08 6.08e+08 6.18e+08 6.36e+08 6.61e+08 ...
## $ US$ / KG : num [1:24] 1.97 2.18 2.24 2.13 2.48 ...
## $ US$ : num [1:24] 1.16e+09 1.32e+09 1.39e+09 1.35e+09 1.64e+09 ...
```

1. Renombramos variables de Items

```
items = rename(items, Descrip_Part_Aduanera = "Descripción Partida Aduanera",
               Anio = "Año",
               Total_Registros = "Total registros",
               Total_KG = "Total KG",
               US_KG = "US$ / KG",
               US_Dolar = "US$")
```

2. Pasamos Anio a integer

```
items$Anio = as.integer(items$Anio)
items$Total_Registros = as.integer(items$Total_Registros)
```

- Reviso algunos datos importantes con summary y analisis

```
summary(items$Anio);
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2021   2021    2022     2022   2022     2022
```

```
summary(items$Meses);
```

```
##      Length      Class      Mode
##           24 character character
```

```
summary(items$Total_Registros)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      79.00   98.75  110.00  111.50  121.00   150.00
```

```
summary(items$Total_KG)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 405673369 593458243 630440338 631341214 690210594 747117647
```

```
summary(items$US_KG)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.843   1.998   2.191   2.172   2.263   2.489
```

```
summary(items$US_Dolar)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## 7.949e+08 1.254e+09 1.394e+09 1.375e+09 1.509e+09 1.857e+09
```

3. Reviso que no tenga datos faltantes

```
sum(is.na(items))# verificamos si hay datos incompletos y se obtuvo 14
```

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

```
sum(!complete.cases(items))
```

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

4. Unique items

```
unique(items$Anio)
```

```
## [1] 2021 2022
```

```
unique(items$Meses)
```

```
## [1] "Enero"      "Febrero"    "Marzo"      "Abril"      "Mayo"
## [6] "Junio"      "Julio"      "Agosto"    "Septiembre" "Octubre"
## [11] "Noviembre"  "Diciembre"
```

```
unique(items$Total_Registros)
```

```
## [1] 79 96 86 116 124 109 106 129 119 143 111 127 102 120 93 97 150 98 99
## [20] 104
```

5. Última vista a items

```
str(items)
```

```
## tibble [24 x 7] (S3: tbl_df/tbl/data.frame)
## $ Descrip_Part_Aduanera: chr [1:24] "2603000000 MINERALES DE COBRE Y SUS CONCENTRADOS" "2603000000 I
## $ Anio                  : int [1:24] 2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
## $ Meses                 : chr [1:24] "Enero" "Febrero" "Marzo" "Abril" ...
## $ Total_Registros      : int [1:24] 79 96 86 116 124 109 106 129 119 124 ...
## $ Total_KG              : num [1:24] 5.90e+08 6.08e+08 6.18e+08 6.36e+08 6.61e+08 ...
## $ US_KG                 : num [1:24] 1.97 2.18 2.24 2.13 2.48 ...
## $ US_Dolar              : num [1:24] 1.16e+09 1.32e+09 1.39e+09 1.35e+09 1.64e+09 ...
```

4. Análisis descriptivo

4.1. Descriptores numéricos

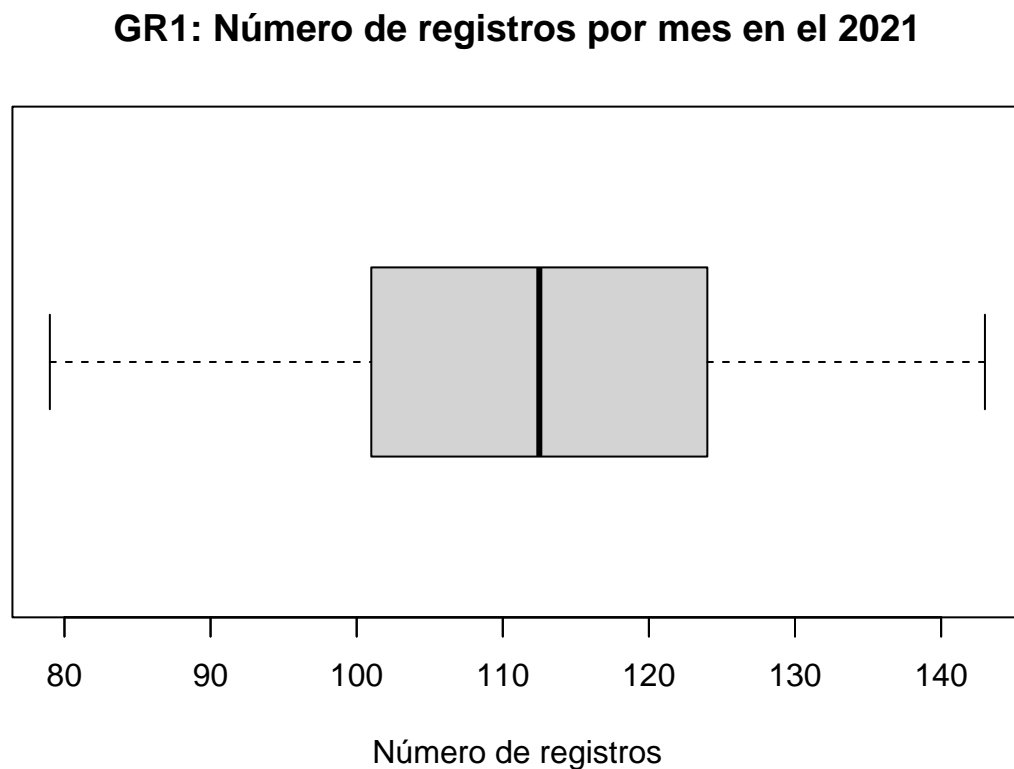
4.1.1 Caja de Bigotes de cada base de datos y sus respectivos descriptores

Exportadores

```
export %>% filter(Anio == 2021) %>% group_by(Meses) %>% summarise(sumTotalKGFull = sum(Total_KG),sumTotalKG= sum(Total_KG))
export %>% filter(Anio == 2022) %>% group_by(Meses) %>% summarise(sumTotalKGFull= sum(Total_KG),sumTotalKG= sum(Total_KG))
```

Exportadores 2021 1. Analizamos el comportamiento de nuestros datos en relacion al numero de registros de cada mes.

```
boxplot(mesescomparacion2021$sumRegistros, main="GR1: Número de registros por mes en el 2021",horizont
axis(side = 1, labels = FALSE)
title(xlab = "Número de registros")
```



- Después de realizar el diagrama de caja y bigotes realizaremos el análisis mediante el uso de algunos descriptores numéricos:

```
round(mean(mesescomparacion2021$sumRegistros),0)
```

Descriptores numéricos

```
## [1] 111
```

```
round(median(mesescomparacion2021$sumRegistros),2)
```

```
## [1] 112.5
```

Interpretación 1

- Dado que nuestra variable presenta valores no atípicos, podemos tomar al promedio como análisis.
- Al tratarse de una variable cuantitativa discreta podemos decir que el promedio de cada mes durante el 2021 es de 111 registros de exportaciones de cobre.
- La mediana obtenida nos permite corroborar la no presencia de datos atípicos puesto que si este dato excede notoriamente al promedio no podríamos hacer un análisis mediante el uso del promedio.

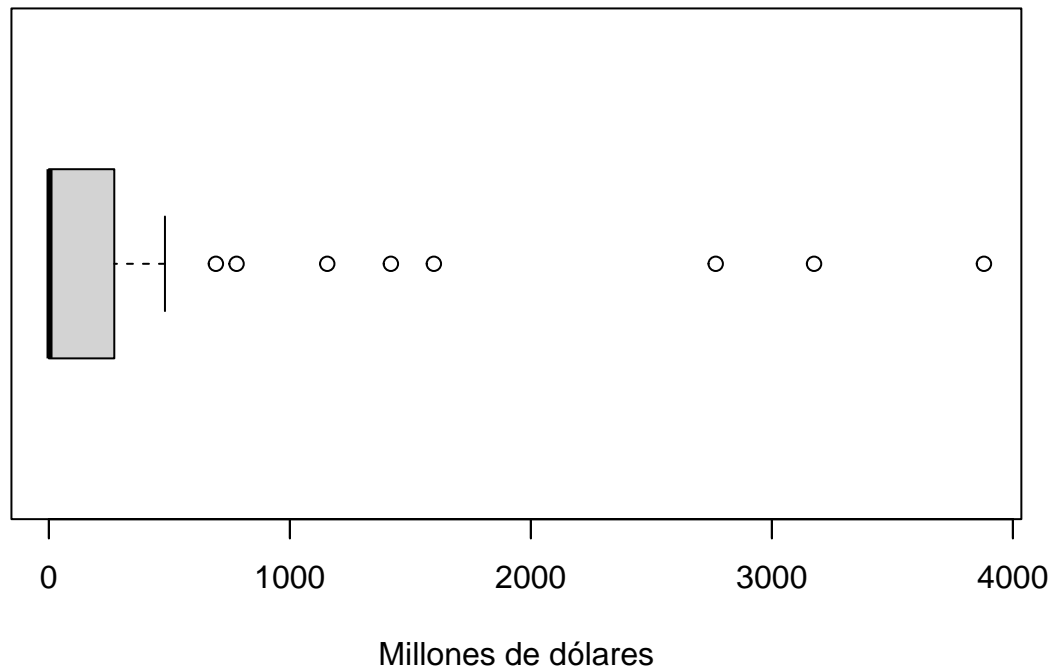
2. Analizamos el comportamiento de nuestros datos en relación al ingreso de divisas por empresa.

```
#2021
export %>% filter(Anio == 2021) %>% group_by(Exportador) %>% summarise(TotalDolaresFull = sum(US_Dolar))
#2022
export %>% filter(Anio == 2022) %>% group_by(Exportador) %>% summarise(TotalDolaresFull = sum(US_Dolar))
```

```
comparacionexport2021$newDollarsMillion = round(comparacionexport2021$TotalDolaresFull/1000000 , 4)
comparacionexport2022$newDollarsMillion = round(comparacionexport2022$TotalDolaresFull/1000000 , 4)
```

```
boxplot(comparacionexport2021$newDollarsMillion, main= "GR2: Ingreso de divisas por empresa en el 2021"
axis(side = 1, labels = FALSE)
title(xlab = "Millones de dólares")
```

GR2: Ingreso de divisas por empresa en el 2021



- Una vez realizada el diagrama de caja y bigotes, podemos ver claramente la presencia de datos atípicos por ello analizaremos los siguientes descriptores:

```
diff(range(comparacionexport2021$newDollarsMillion))
```

```
## [1] 3879.96
```

```
round(median(comparacionexport2021$newDollarsMillion),2)
```

```
## [1] 2.79
```

```
round(quantile(comparacionexport2021$newDollarsMillion, c(0.25, 0.75)),2)
```

```
##    25%    75%
```

```
##  0.12 271.60
```

Interpretación 2

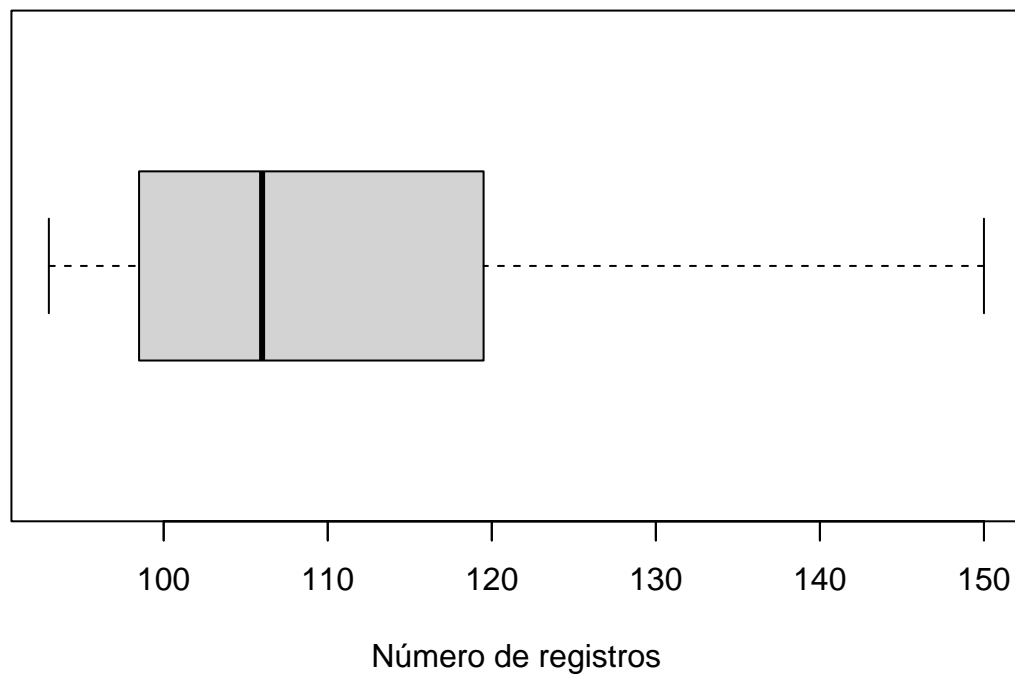
- El rango de nuestros datos es 3879.96 ,lo que nos permite tener una idea de la amplitud de nuestros datos, en este caso es un numero demasiado grande casi igual al máximo entonces hay presencia de datos atípicos, de esta manera corroboramos lo obtenido por el boxplot.

- La media nos indica que el 50% de las empresas exportadoras generan un ingreso menor o igual 2.79 millones de dolares respecto a las importaciones de cobre.
- A partir del percentil calculado solo un 25% de las empresas exportadoras de cobre genera un ingreso mayor a 271.60 millones de dolares

Exportadores 2022 1. Analizamos el comportamiento de nuestros datos en relacion al numero de registros de cada mes.

```
boxplot(mesescomparacion2022$sumRegistros, main="GR3: Número de registros por mes en el 2022", horizontal=TRUE,
axis(side = 1, labels = FALSE)
title(xlab = "Número de registros")
```

GR3: Número de registros por mes en el 2022



- Después de realizar el diagrama de caja y bigotes realizaremos el análisis mediante el uso de algunos descriptores numéricos:

```
round(mean(mesescomparacion2022$sumRegistros),0)
```

Descriptores numéricos

```
## [1] 111
```

```
round(median(mesescomparacion2022$sumRegistros),2)
```

```
## [1] 106
```

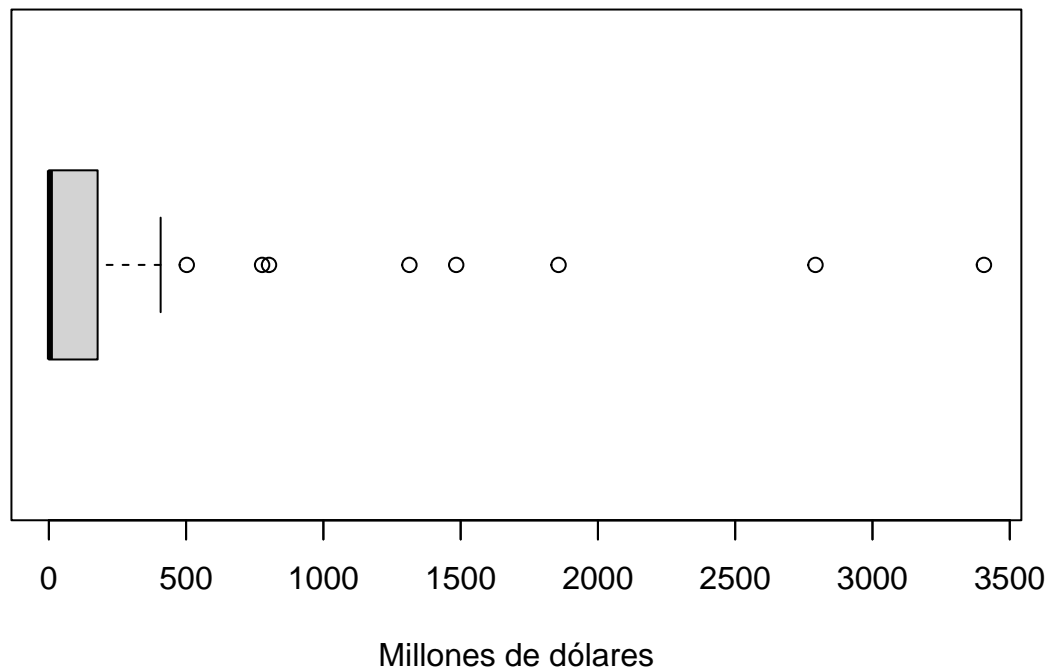
Interpretación 1 “ - Dado que nuestra variable presenta valores no atípicos, podemos tomar al promedio como análisis.

- Al tratarse de una variable cuantitativa discreta podemos decir que el promedio de cada mes durante el 2022 es de 111 registros de exportaciones de cobre si comparamos respecto al año anterior, podemos afirmar que no hubo cambios significativos respecto al numero de exportaciones por cada mes en los dos años.
- La mediana obtenida nos permite corroborar la no presencia de datos atípicos puesto que si este dato excede notoriamente al promedio no podríamos hacer un análisis mediante el uso del promedio.

2. Analizamos el comportamiento de nuestros datos en relacion al ingresos de divesas por empresa.

```
boxplot(comparacionexport2022$newDollarsMillion,main= "GR4: Ingreso de divisas por empresa en el 2022",  
axis(side = 1, labels = FALSE)  
title(xlab = "Millones de dólares"))
```

GR4: Ingreso de divisas por empresa en el 2022



- Una vez realizada el diagrama de caja y bigotes, podemos ver claramente la presencia de datos atípicos por ello analizaremos los siguientes descriptores:

```
round(diff(range(comparacionexport2022$newDollarsMillion)),2)
```

```
## [1] 3406.05
```

```
round(median(comparacionexport2022$newDollarsMillion),2)
```

```
## [1] 3.82
```

```
round(quantile(comparacionexport2022$newDollarsMillion, c(0.25, 0.75)),2)
```

```
##      25%      75%  
##    0.32 171.72
```

Interpretación 2 El rango de nuestros datos es 3406.05 ,lo que nos permite tener una idea de la amplitud de nuestros datos, en este caso es un numero demasiado grande casi igual al máximo entonces hay presencia de datos atípicos, de esta manera corroboramos lo obtenido por el boxplot.

La media nos indica que el 50% de las empresas exportadoras generan un ingreso menor o igual 3.82 millones de dolares respecto a las importaciones de cobre.

A partir del percentil calculado solo un 25% de las empresas exportadoras de cobre genera un ingreso mayor a 171.72 millones de dolares

4.2. Descriptores gráficos

4.2.1. Análisis gráficos de acuerdo a los objetivos propuestos

Objetivo 1

- Comparar el numero de dolares del sector Minero con los otros sectores
- Gráfico Utilizado: Gráficos de barras
- ¿Por qué? Utilizar un diagrama de barras nos permitirá comparar fácilmente la cantidad de millones de dólares obtenidas en cada uno de los sectores como el agrícola,minero,pesquero,petróleo y resto o demás. Cada barra nos estaría representando un sector y la altura de la barra es proporcional a la cantidad de dólares que han obtenido.

```
sector %>% group_by(Sector) %>% summarise(sumi = sum(Enero + Febrero + Marzo + Abril + Junio + Julio + Agosto + Septiembre + Octubre + Noviembre + Diciembre))
```

```
#Entre 1000 toneladas  
comparacionSectores2021$newToneladas = round(comparacionSectores2021$sumi/1000000,2)
```

```
barplot(comparacionSectores2021$newToneladas,names = c("Agricolas","Minero","Pesquero","Petroleo","Resto"))
```

```
comparacionSectors2021 %>% slice_max(sumi) %>% pull(Sector)
max(comparacionSectors2021$sumi)/1000000
```

Gráfico 5: Exportaciones de sectores económicos 2021

Interpretación_GR 5:

- En este gráfico observamos como el sector minero sobrepasa con creces a los demás sectores como el agrícola , pesquero , petroleo y resto con \$34,196.45 millones de dólares. De acuerdo a ello, podemos concluir que el sector minero es el que genera mayores ingresos de divisas, en consecuencia, reactiva más la economía del Perú ,debido a que se necesitará mayor mano de obra para poder satisfacer la demanda de cobre.

Objetivo 2

- Determinar que países de destino vendemos más cobre y que cantidad.
- Gráficos Usados: BoxPlot y Gráficos circulares
- ¿Por qué? Los boxplot nos permiten determinar la distribución de un conjunto de datos e identificar los valores atípicos y los gráficos circulares porque nos dan una mejor visualización de la proporción y porcentajes , además de que son fáciles de interpretar.

```
países %>% filter(Anio == 2021) %>% group_by(Pais_Destino) %>% summarise(TotalesRegistrosFull = Total_RegistrosFull)
```

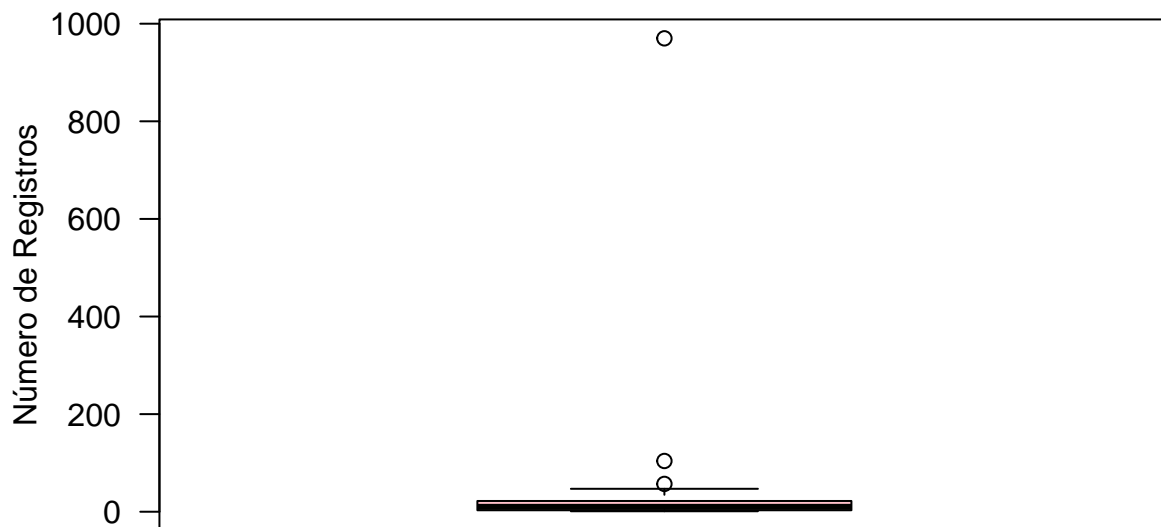
Gráfico 6: Cantidad de registros en cada país en 2021(GRÁFICO DE CAJA Y BIGOTES DE PAÍSES DESTINO 2021)

```
## Warning: Returning more (or less) than 1 row per 'summarise()' group was deprecated in
## dplyr 1.1.0.
## i Please use 'reframe()' instead.
## i When switching from 'summarise()' to 'reframe()', remember that 'reframe()'
## always returns an ungrouped data frame and adjust accordingly.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

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

```
# Crear el gráfico de caja y bigotes
boxplot(paísesxregistrosfull2021$CantidadRegistrosPorAño ,
        col = "pink",
        ylab = "Número de Registros",
        xlab = ".",
        main = "GR6: Cantidad de registros según cada país en 2021",
        las=2)
```

GR6: Cantidad de registros según cada país en 2021



Interpretación Grafico 6 :

- Aquí utilizamos diagramas de caja y bigotes porque proporcionan una representación más adecuada de la distribución y los patrones en los datos del numero de registros de 2021.El valor a atipicó más grande obtenido es China con 970 numeros de registros

```
paísesxregistrosfull2021 = paísesxregistrosfull2021[order(paísesxregistrosfull2021$CantidadRegistrosPorAño)]
otros<-data.frame(Pais_Destino = "Otros",CantidadRegistrosPorAño = sum(paísesxregistrosfull2021[6:nrow(paísesxregistrosfull2021V2),]))
paísesxregistrosfull2021V2 = rbind(paísesxregistrosfull2021[1:5,],otros)
```

```
# Pie Chart with Percentages
slices <- c(paísesxregistrosfull2021V2$CantidadRegistrosPorAño)
lbls <- c(unique(paísesxregistrosfull2021V2$Pais_Destino))
pct <- round(slices/sum(slices)*100,2)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices,labels = lbls, col=c("blue", "green", "yellow", "red","gray","skyblue"),
    main="GR7:Destinos de exportación 2021")
```

GR7:Destinos de exportación

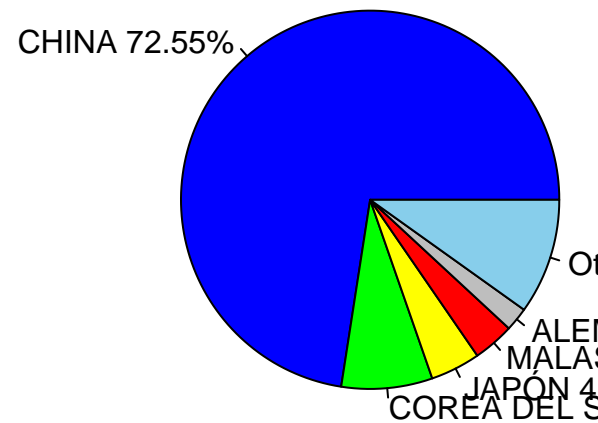


Gráfico 7: Destinos de exportación 2021

```
paísesxregistrosfull2021 %>% slice_max(CantidadRegistrosPorAño) %>% pull(Pais_Destino)
```

```
## [1] "CHINA"
```

```
max(paísesxregistrosfull2021$CantidadRegistrosPorAño)
```

```
## [1] 970
```

Interpretación Grafico 7:

- Como podemos ver en el siguiente gráfico en el 2021 , China es el país destino al que más exportamos cobre según el numero de registros , con 970 registros y con un porcentaje de participación en el mercado internacional peruano de cobre de 72.55%.

```
países %>% filter(Anio == 2022) %>% group_by(Pais_Destino) %>% summarise(TotalesRegistrosFull = Total_RegistrosFull)
```

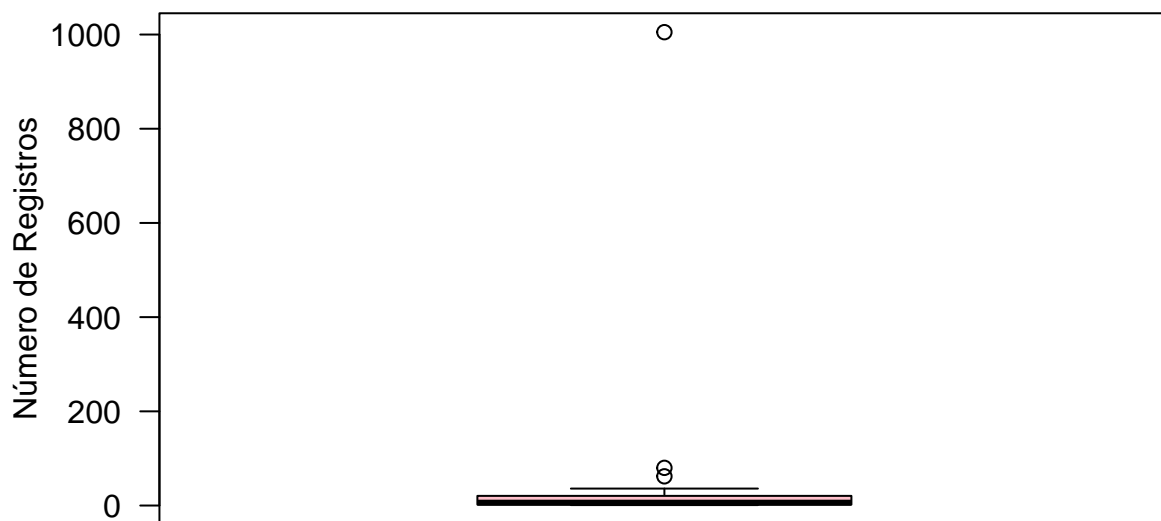
Gráfico 8: Cantidad de registros según cada país en 2022(GRÁFICO DE CAJA Y BIGOTES DE PAÍSES DESTINO 2022)

```
## Warning: Returning more (or less) than 1 row per 'summarise()' group was deprecated in
## dplyr 1.1.0.
## i Please use 'reframe()' instead.
## i When switching from 'summarise()' to 'reframe()', remember that 'reframe()'
## always returns an ungrouped data frame and adjust accordingly.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

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

```
# Crear el gráfico de caja y bigotes
boxplot(paísesxregistrosfull2022$CantidadRegistrosPorAño,
        col = "pink",
        ylab = "Número de Registros",
        xlab = ".",
        main = "GR8:Cantidad de registros según cada país en 2022",
        las=2)
```

GR8:Cantidad de registros según cada país en 2022



Interpretación Gráfico 8:

- Llegamos a la conclusión según este gráfico de cajas y bigotes que nuestro dato más predominante, y nuestro dato atípico sería China con 1005 registros con respecto a los otros países en el año 2022.

```
paísesxregistrosfull2022 = paísesxregistrosfull2022[order(paísesxregistrosfull2022$CantidadRegistrosPorAño),]
otros2<-data.frame(Pais_Destino = "Otros",CantidadRegistrosPorAño = sum(paísesxregistrosfull2022[6:nrow(paísesxregistrosfull2022V2),]))
paísesxregistrosfull2022V2 = rbind(paísesxregistrosfull2022[1:5,],otros2)
```

```
# Pie Chart with Percentages
slices2 <- c(paísesxregistrosfull2022V2$CantidadRegistrosPorAño)
lbls2 <- c(unique(paísesxregistrosfull2022V2$Pais_Destino))
pct2 <- round(slices2/sum(slices2)*100,2)
lbls2 <- paste(lbls2, pct2) # add percents to labels
lbls2 <- paste(lbls2,"%",sep="") # ad % to labels
pie(slices2,labels = lbls2, col=c("blue", "green", "yellow", "red","gray","skyblue"),
    main="GR9 :Destinos de exportación 2022")
```

GR9 :Destinos de exportación

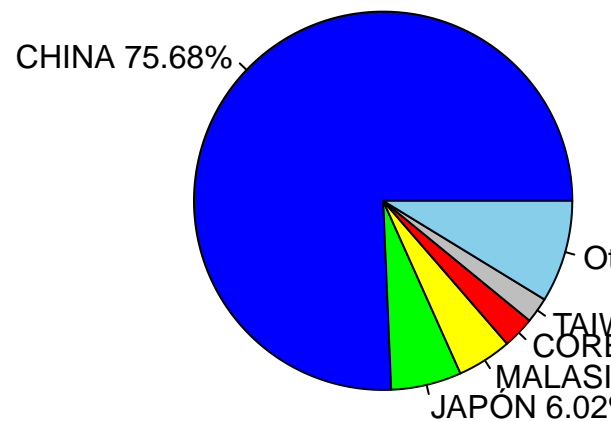


Gráfico 9: Destinos de exportación 2022

```
paísesxregistrosfull2022 %>% slice_max(CantidadRegistrosPorAño) %>% pull(Pais_Destino)
```

```
## [1] "CHINA"
```

```
max(paísesxregistrosfull2022$CantidadRegistrosPorAño)
```

```
## [1] 1005
```


Intepretación Grafico 9:

- Como podemos ver en el siguiente gráfico en el 2022, China lidera en el país importador de cobre en Perú y representa el 75.68% de los países que compran cobre en Perú. Hay que hacer énfasis de como China influye en la economía peruana con la compra de minerales como es el caso del cobre.

Objetivo 3

- Comparar la cantidad de dolares\$ de cobre que se consigue al exportar respecto a otros minerales
- Graficos Usados: Graficos de barras
- ¿Por qué? Al ser nuestras cantidades expresadas en millones de dólares, un gráfico de barras nos proporciona una mejor facilidad de comparación y de lectura, pero también podemos interpretar magnitudes como por ejemplo si existe una gran diferencia de dolares obtenido en cada mineral mediante el grafico.

```
minerales %>% filter(Anio==2021)%>% group_by(Minerales)%>% summarise(cantDolarsFull = sum(US_Dolar)) ->  
minerales %>% filter(Anio==2022)%>% group_by(Minerales)%>% summarise(cantDolarsFull = sum(US_Dolar)) ->
```

```
comparDolarMine2021$newDolarsMillion = round(comparDolarMine2021$cantDolarsFull/1000000 , 2)  
comparDolarMine2022$newDolarsMillion = round(comparDolarMine2022$cantDolarsFull/1000000 , 2)
```

```
barplot(comparDolarMine2021$newDolarsMillion,names =c("Cobre","Estaño","Hierro","Oro","Plata","Plom","R
```

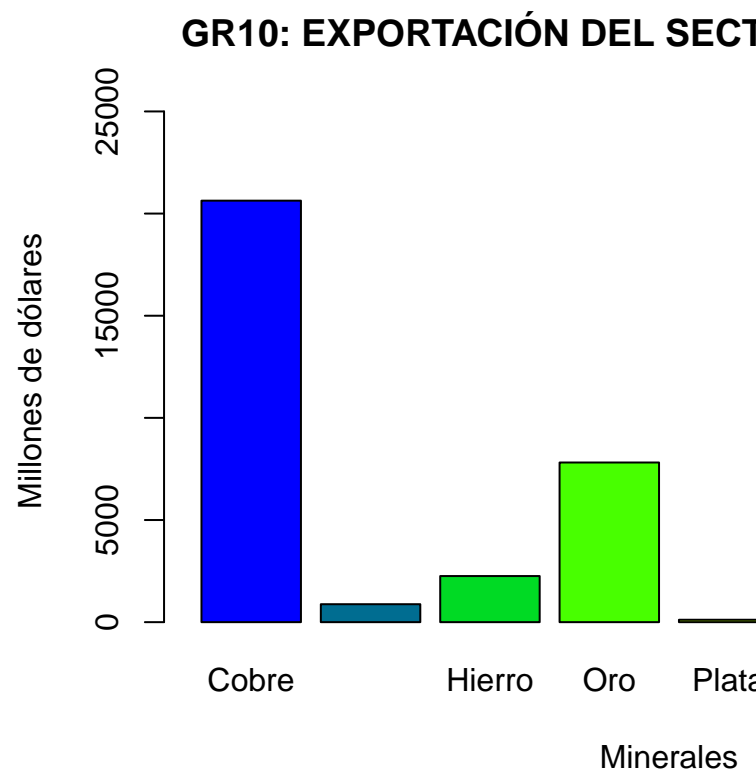


Gráfico 10: Exportación del sector minero del 2021

```
length(unique(comparDolarMine2021$Minerales))
```

```
## [1] 8
```

```
comparDolarMine2021 %>% slice_max(cantDolarsFull) %>% pull(Minerales)
```

```
## [1] "Cobre"
```

```
max(comparDolarMine2021$cantDolarsFull)
```

```
## [1] 20632834090
```

Interpretación Gráfico 10

- El gráfico de barras nos da una idea de clara de la participación de los minerales en los mercados internacionales, podemos llegar a la conclusión que el cobre que mayor se exporta con respecto a otros minerales en el año 2021. Por otro lado, el mineral que menos se exporta es plata.
- Como segundo punto la venta de cobre genera un ingreso de divisas de un monto de \$20,663.83 millones de dólares. En segundo lugar, tenemos al oro que asciende a un valor de \$7,815.18 millones de dólares.

Objetivo 4

- Comparar que empresa exportadora de cobre genera mayores ingresos de divisas en el Perú
- Graficos Usados: Graficos circulares
- ¿Por qué? Usamos los gráficos circulares debido a que cada empresa genera una cantidad de divisas diferente y conforman un todo, por lo que al dividirlo en proporciones y porcentajes de ese total de divisas de cada 2021 y 2022 nos será más sencillo y simple identificar que empresa a generado el mayor ingreso de divisas.

```
comparacionexport2021= comparacionexport2021[order(comparacionexport2021$TotalDolaresFull,decreasing = '
otrosexport1<-data.frame(Exportador = "Otros",TotalDolaresFull = sum(comparacionexport2021[9:nrow(comparacionexport2021V2 = rbind(comparacionexport2021[1:8,],otrosexport1)
```

```
# Pie Chart with Percentages
slices3 <- c(comparacionexport2021V2$newDollarsMillion)
lbls3 <- c("C.Mine.Antamina","Soc.Mine.Cerro Verde","Mine.Las Bambas","Comp. Mine. Antapaccay","Trafigu
pct3 <- round(slices3/sum(slices3)*100,2)
lbls3 <- paste(lbls3, pct3) # add percents to labels
lbls3 <- paste(lbls3,"%",sep="") # ad % to labels
pie(slices3,labels = lbls3, col=colorRampPalette(c("blue", "green", "yellow", "red"))(length(unique(comparacionexport2021V2$Exportador))))
main="GR11:Sociedades Exportadoras en el 2021")
```

GR11:Soc

Soc.Mine.Cerro Verde

Mine.Las Bambas 15.62%

Comp. Mine. Antapaccay 9.02

Trafigura Peru S.A.

Gráfico 11: Proporción de ingresos de sociedades exportadoras en el 2021

```
max(comparacionexport2021$TotalDolaresFull)
```

```
## [1] 3879960410
```

```
max(comparacionexport2021$newDollarsMillion)
```

```
## [1] 3879.96
```

Interpretación Gráfico 11:

- De acuerdo a la tabla con respecto al año *2021* podemos llegar a la conclusión, que la empresa exportadora que genero mayores ingresos de divisas al Perú fue COMPAÑÍA MINERA ANTAMINA S.A. y la segunda empresa es SOCIEDAD MINERA CERRO VERDE S.A.A.
- La COMPAÑÍA MINERA ANTAMINA S.A. es el líder con respecto a las sociedades exportadoras de cobre peruanas en los mercados internacionales ha generado en el año 2021 un monto de \$3879.96 millones de dólares.

```
comparacionexport2022 %>% arrange(desc(TotalDolaresFull))
```

Gráfico 12: Proporción de ingresos de empresas exportadoras en el 2022

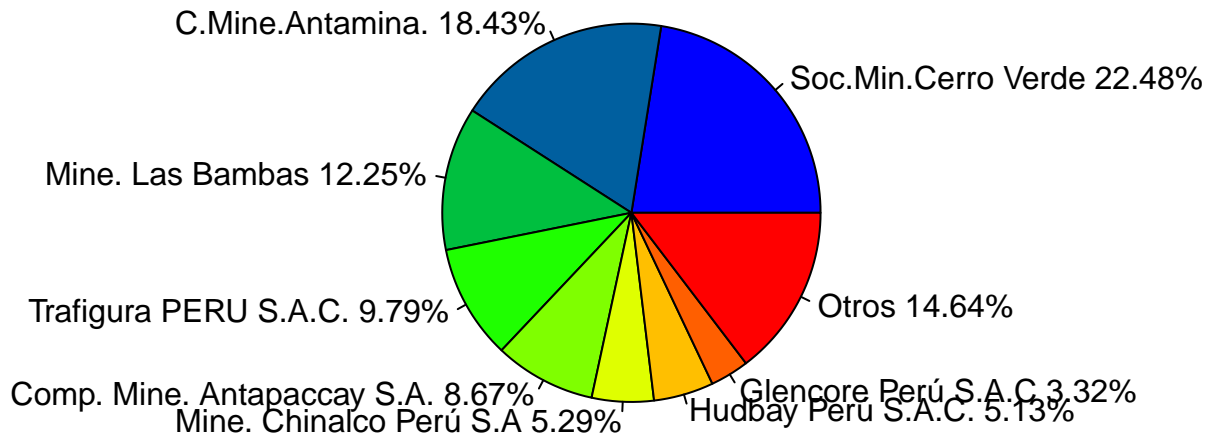
```
## # A tibble: 50 x 3
##   Exportador                                TotalDolaresFull newDollarsMillion
##   <chr>                                <dbl>          <dbl>
## 1 SOCIEDAD MINERA CERRO VERDE S.A.A.      3406053041.      3406.
## 2 COMPAÑIA MINERA ANTAMINA S.A.          2792431018.      2792.
## 3 MINERA LAS BAMBAS S.A.                  1856569488.      1857.
## 4 TRAFIGURA PERU SOCIEDAD ANONIMA CERRADA -- 1483849772      1484.
## 5 COMPAÑIA MINERA ANTAPACCAY S.A.          1313348485      1313.
## 6 MINERA CHINALCO PERÚ S.A.                801914648.       802.
## 7 HUDBAY PERU S.A.C.                      776759691        777.
## 8 GLENCORE PERU S.A.C.                     502389919.       502.
## 9 MARCOBRE S.A.C.                         407206593        407.
## 10 GOLD FIELDS LA CIMA S.A.                362434095        362.
## # i 40 more rows
```

```
comparacionexport2022= comparacionexport2022[order(comparacionexport2022$TotalDolaresFull,decreasing = TRUE),]
otrosexport2<-data.frame(Exportador = "Otros",TotalDolaresFull = sum(comparacionexport2022[9:nrow(comparacionexport2022)$TotalDolaresFull]),newDollarsMillion = sum(comparacionexport2022[9:nrow(comparacionexport2022)$newDollarsMillion]))
comparacionexport2022V2 = rbind(comparacionexport2022[1:8,],otrosexport2)
```

```
# Pie Chart with Percentages
```

```
slices4 <- c(comparacionexport2022V2$newDollarsMillion)
lbls4 <- c("Soc.Min.Cerro Verde","C.Mine.Antamina.","Mine. Las Bambas","Trafigura PERU S.A.C.","Comp. M. Antamina")
pct4 <- round(slices4/sum(slices4)*100,2)
lbls4 <- paste(lbls4, pct4) # add percents to labels
lbls4 <- paste(lbls4,"%",sep="") # ad % to labels
pie(slices4,labels = lbls4, col=colorRampPalette(c("blue", "green", "yellow", "red"))(length(unique(comparacionexport2022V2$Exportador))))
main="GR12: Proporción de ingresos de empresas exportadoras en el 2022")
```

GR12: Proporción de ingresos de empresas exportadoras en el 202



```
max(comparacionexport2022$TotalDolaresFull)
```

```
## [1] 3406053041
```

```
max(comparacionexport2022$newDollarsMillion)
```

```
## [1] 3406.053
```

Interpretación Gráfico 12:

- Entre las sociedades la que generó mayores ventas por la exportación de cobre en el año 2022 fue SOCIEDAD MINERA CERRO VERDE S.A.A. con un monto de \$3406.05 millones de dólares. Tanto la empresa CMine Antamina y Soc.,Min.Cerro Verde son las sociedades más solicitadas por los ofertantes y juntan representa casi el 41% del mercado exportador de cobre.
- Obs: En esta tabla se refleja la cantidad de dolares generadas por cada empresa exportadora de cobre. De acuerdo, podemos decir que la segunda empresa que generó mayores ventas en el año 2022 fue COMPAÑIA MINERA ANTAMINA S.A. con \$2792.43 millones de dólares.

```
comparacionexport2022 %>% arrange(desc(TotalDolaresFull))
```

```
## # A tibble: 50 x 3
```

```
## Exportador
```

```
TotalDolaresFull newDollarsMillion
```

```
##      <chr>                                <dbl>          <dbl>
## 1 SOCIEDAD MINERA CERRO VERDE S.A.A.      3406053041.      3406.
## 2 COMPAÑIA MINERA ANTAMINA S.A.           2792431018.      2792.
## 3 MINERA LAS BAMBAS S.A.                  1856569488.      1857.
## 4 TRAFIGURA PERU SOCIEDAD ANONIMA CERRADA -- 1483849772      1484.
## 5 COMPAÑIA MINERA ANTAPACCAY S.A.         1313348485      1313.
## 6 MINERA CHINALCO PERÚ S.A.               801914648.       802.
## 7 HUDBAY PERU S.A.C.                      776759691        777.
## 8 GLENCORE PERU S.A.C.                    502389919.       502.
## 9 MARCOBRE S.A.C.                         407206593        407.
## 10 GOLD FIELDS LA CIMA S.A.               362434095        362.
## # i 40 more rows
```

Objetivo 5

- Comparar la cantidad de cobre por mes en kilo toneladas metricas(kTM) entre los años 2021 y 2022
- Gráficos Usados: Graficos de barras y grafico de barras de doble comparación
- ¿Por qué? Usamos los graficos de barras de doble comparación porque nos permite ver la tendencia o patrón que pueden estar siguiendo la cantidad de kTM de cobre que está siendo demandada en cada mes del año 2021 y 2022 , incluyendo los 12 meses de Enero a Diciembre

```
mesescomparacion2021 <- mesescomparacion2021[order(match(mesescomparacion2021$Meses,c("Enero","Febrero"
mesescomparacion2022 <- mesescomparacion2022[order(match(mesescomparacion2022$Meses,c("Enero","Febrero"
```

```
mesescomparacion2021$newKgTon = round(mesescomparacion2021$sumTotalKGFull/1000000 , 4)
mesescomparacion2022$newKgTon= round(mesescomparacion2022$sumTotalKGFull/1000000 , 4)
mesescomparacion2021$newDolarMillion = round(mesescomparacion2021$sumTotalDolar/1000000,4)
mesescomparacion2022$newDolarMillion = round(mesescomparacion2022$sumTotalDolar/1000000,4)
```

```
barplot(mesescomparacion2021$newKgTon,names = c("En.", "Feb.", "Mar.", "Abr.", "May.", "Jun.", "Jul.", "Ag.", "S.", "Oct.", "Nov.", "Dic."))
```

GR 13: DEMANDA DEL COBRE

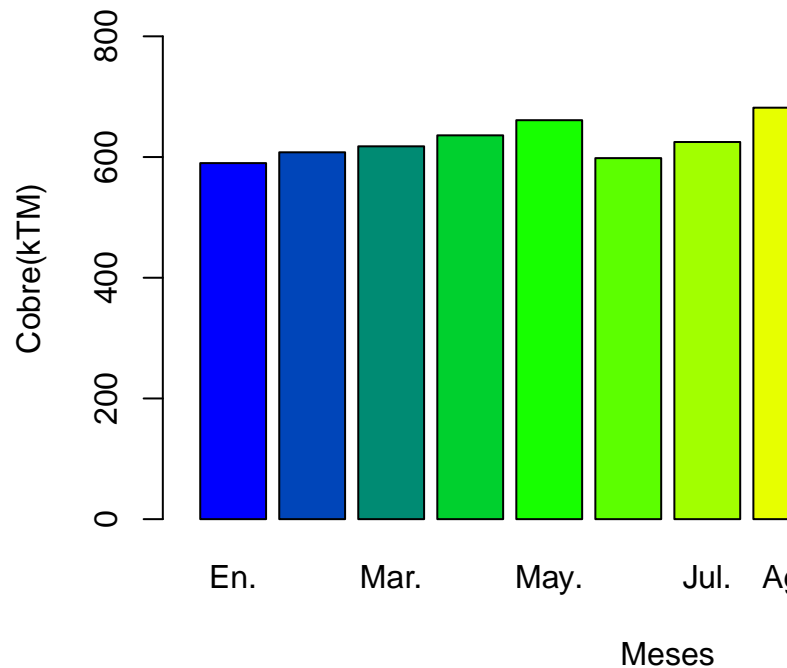


Gráfico 13: Demanda de cobre del 2021 Mensual

```
length(unique(mesescomparacion2021$Meses))
```

```
## [1] 12
```

```
mesescomparacion2021 %>% slice_max(sumTotalKGFull) %>% pull(Meses)
```

```
## [1] "Noviembre"
```

```
max(mesescomparacion2021$sumTotalKGFull)
```

```
## [1] 746239657
```

```
max(mesescomparacion2021$newKgTon)
```

```
## [1] 746.2397
```

Interpretación Gráfico 13:

- Como podemos observar la cantidad de kilogramos alcanza su punto más alto en Noviembre y Diciembre, y la menor cantidad de cobre(kTM) importada en 2021 fue en Octubre una de las posibles respuestas es que el tipo de cambio que es el dólar valía menos con respecto a su moneda nacional, esto significa que la empresa importadora necesita más dólares para costear el cobre importado de Perú, una de las consecuencias es que reduce su capacidad adquisitiva en la compra de la mercancía.

```
barplot(mesescomparacion2022$newKgTon,names = c("En.", "Feb.", "Mar.", "Abr.", "May.", "Jun.", "Jul.", "Ag.", "S.", "Oct.", "Nov.", "Dic."))
```

GR14: DEMANDA DEL COBRE

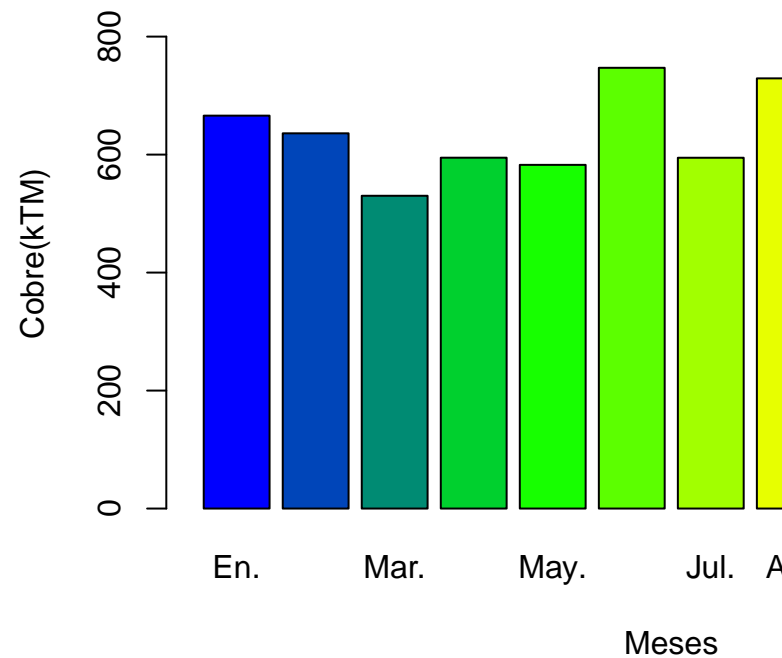


Gráfico 14: Demanda del cobre del 2022 Mensual

```
length(unique(mesescomparacion2022$Meses))
```

```
## [1] 12
```

```
mesescomparacion2022 %>% slice_max(sumTotalKGFull) %>% pull(Meses)
```

```
## [1] "Septiembre"
```

```
max(mesescomparacion2022$sumTotalKGFull)
```

```
## [1] 765726601
```

Interpretación Gráfico 14:

- Como podemos observar la demanda de cobre cantidad alcanzó su punto más alto en Junio y Septiembre, y la menor cantidad de kg importada en 2022 fue en Noviembre, interpretando la grafica que esto se debió al bajo precio de cobre por kilogramo que incentivo a que las empresas importadoras demanden más cobre.


```
# Crear el gráfico de barras de doble barras
```

```
barplot(rbind(mesescomparacion2022$newKgTon,mesescomparacion2021$newKgTon), beside = TRUE, names.arg = c
```

GR15: DEMANDA D

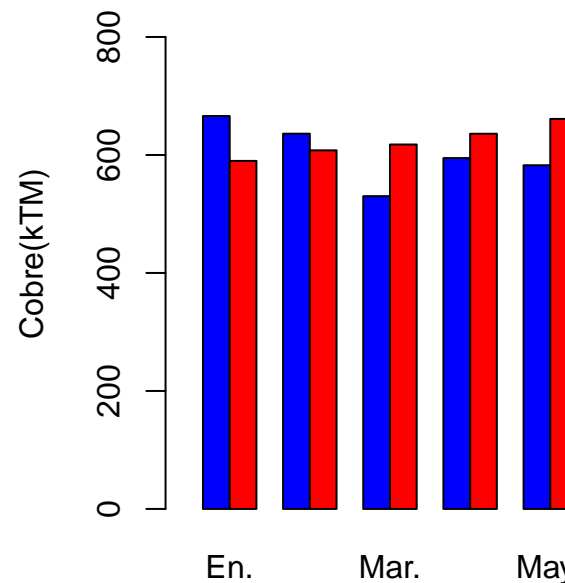


Gráfico 15: Comparación de la demanda del cobre del 2021 y 2022

Objetivo 6

- Comparar el precio del cobre promedio mensual por toneladas entre los años 2021 y 2022.
- Graficos Usados: Graficos de doble comparación y grafico de función de variación
- ¿Por que? En este objetivo tambien recurrimos al uso de grafico de barras de doble comparación para poder visualizar de forma más sencilla que que tanto vario el promedio del precio de un mes tanto del 2021 y 2022 , la funciones de variación expresadas en graficos fue empleada por que nos permiten visualizar que tanto se encontraba variando el precio y la cantidad de millones de dolares en cada mes del año

```
items %>% filter(Anio == 2021) %>% group_by(Meses) %>% summarise(Promedio = round(mean(US_KG),3)) ->Kilo
items %>% filter(Anio == 2022) %>% group_by(Meses) %>% summarise(Promedio = round(mean(US_KG),3)) ->Kilo
```

```
Kilo_DolarComparacion2021 <- Kilo_DolarComparacion2021[order(match(Kilo_DolarComparacion2021$Meses,c("E",
Kilo_DolarComparacion2022 <- Kilo_DolarComparacion2022[order(match(Kilo_DolarComparacion2022$Meses,c("E
```

```
# Crear el gráfico de barras de doble barras
```

```
barplot(rbind(Kilo_DolarComparacion2021$Promedio,Kilo_DolarComparacion2022$Promedio), beside = TRUE, nar
```

GR16: PRECIO PROMEDIO DEL

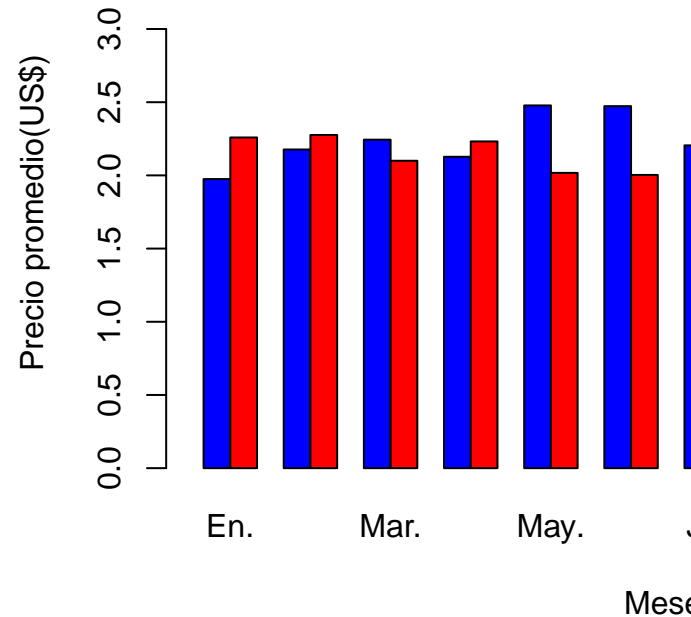


Gráfico 16: Precio promedio del cobre entre 2021 y 2022

Interpretación Grafico 16:

- De acuerdo al diagrama de barras podemos decir que en el año 2021 el precio promedio fue mayor con respecto al 2022 en la mayoría de meses, podemos llegar a una conclusión, si bien en el año 2022 hubo un mayor ventas de cobre, pero no fue acompañado con el precio del cobre, ya que se vendió a menor precio con respecto al 2021. Además que en el año 2022 los precios de venta era menor con respecto al 2021 en esta gráfica se observa que en el año 2022 hubo mayor demanda y producción de cobre a diferencia del 2021.

```
plot(c(1:length(Kilo_DolarComparacion2021$Meses)),Kilo_DolarComparacion2021$Promedio ,type = "b",ylab =  
text(x = c(1:length(Kilo_DolarComparacion2021$Meses)),y =Kilo_DolarComparacion2021$Promedio,round(unique
```

GR17: OSCILACIÓN DEL PRECIO

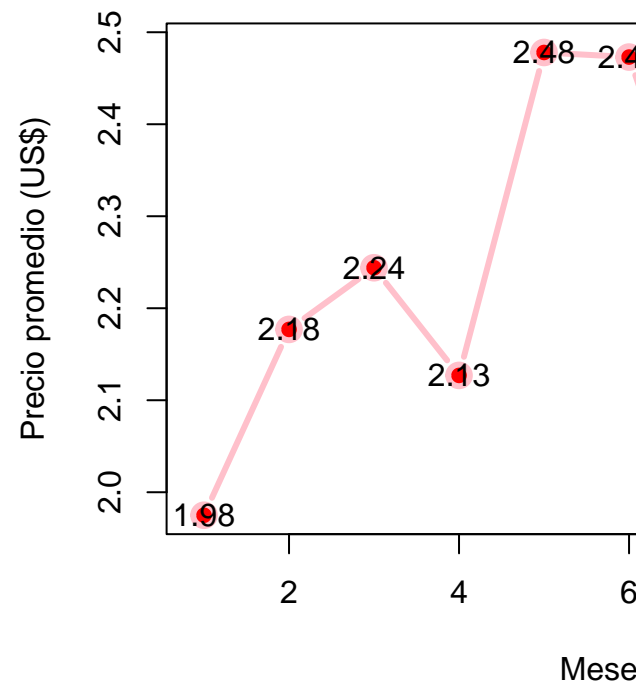


Gráfico 17: Oscilación del precio promedio del cobre en 2021

Interpretación Grafico 17:

- De acuerdo al gráfico de línea que me da una idea clara de como ha fluctuado el precio promedio del cobre durante todo el año 2021, donde alcanzo su pico alto en el mes de octubre con un precio de \$2.49 dólares por 1 kg de cobre. Sin embargo, con un pico bajo en el mes de enero \$1.98 dolares por kg de cobre. Observamos que en el año 2021, el precio del cobre no ha es estable ha ido fluctuando durante todos los meses.

```
plot(c(1:length(Kilo_DolarComparacion2022$Meses)),Kilo_DolarComparacion2022$Promedio ,type = "b",ylab =  
text(x = c(1:length(Kilo_DolarComparacion2022$Meses)),y =Kilo_DolarComparacion2022$Promedio,round(unique
```

GR18: FLUCTUACIÓN DEL

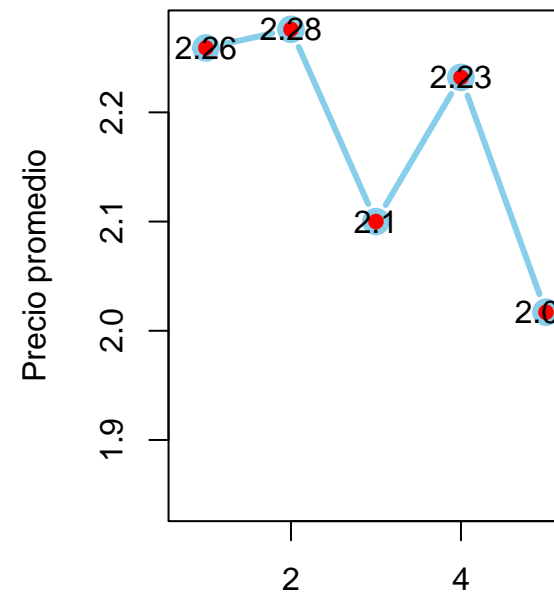


Gráfico 18: Fluctuación del precio promedio del cobre del año 2022

Interpretación Gráfico 18:

- De acuerdo al gráfico de línea que me da una idea clara de como ha fluctuado el precio promedio del cobre durante todo el año 2022, donde alcanzo su pico alto en el mes de febrero con un precio de \$2.28 dólares por 1 kg de cobre. Sin embargo, con un pico bajo en el mes de setiembre \$1.84 dólares por kg de cobre. Observamos que en el año 2022, el precio del cobre no ha estado estable ha ido fluctuando durante todos los meses y ha tenido un precio más bajo que en el 2021.

```
items %>% filter(Anio == 2021) %>% select(Meses,US_Dolar) ->comparacionDolarTotalItems2021
items %>% filter(Anio == 2022) %>% select(Meses,US_Dolar) ->comparacionDolarTotalItems2022
```

```
comparacionDolarTotalItems2021$US_Dolar = round(comparacionDolarTotalItems2021$US_Dolar/1000000,2)
comparacionDolarTotalItems2022$US_Dolar = round(comparacionDolarTotalItems2022$US_Dolar/1000000,2)
```

Total de dolares promedio en meses 2021 y 2022

```
# Crear el gráfico de barras de doble barras
```

```
barplot(rbind(comparacionDolarTotalItems2021$US_Dolar,comparacionDolarTotalItems2022$US_Dolar), beside = TRUE)
```

Gráfico 19: Comparación de millones de dolares mensual entre año 2021 y 2022

¡R19:Comparación de millones de dólares mensual entre el año 2021 y

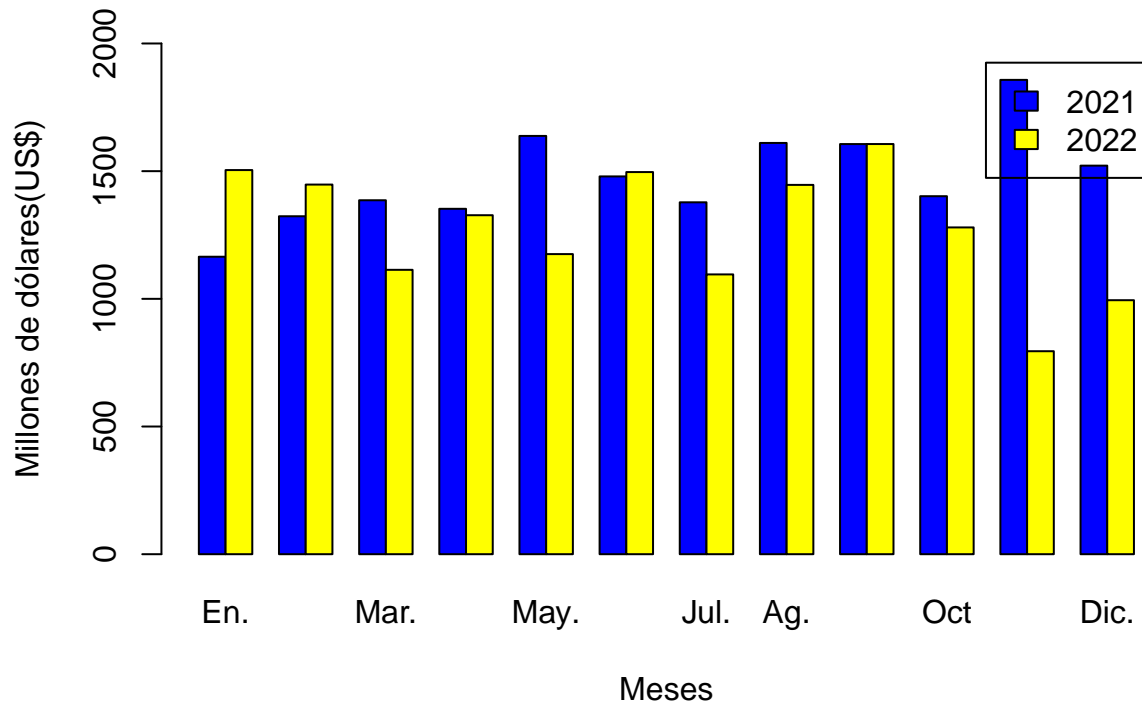


Gráfico 20: Variación de Millones de dólares mensuales del año 2021

- Aquí creamos un gráfico para poder ver la variación de millones en dolares por meses evaluado en el año 2021

```
plot(c(1:length(comparacionDolarTotalItems2021$Meses)),comparacionDolarTotalItems2021$US_Dolar ,type = 'b')
text(x = c(1:length(comparacionDolarTotalItems2021$Meses)),y =comparacionDolarTotalItems2021$US_Dolar,u
```

GR20: Variación de Millones de dólares del año 2021

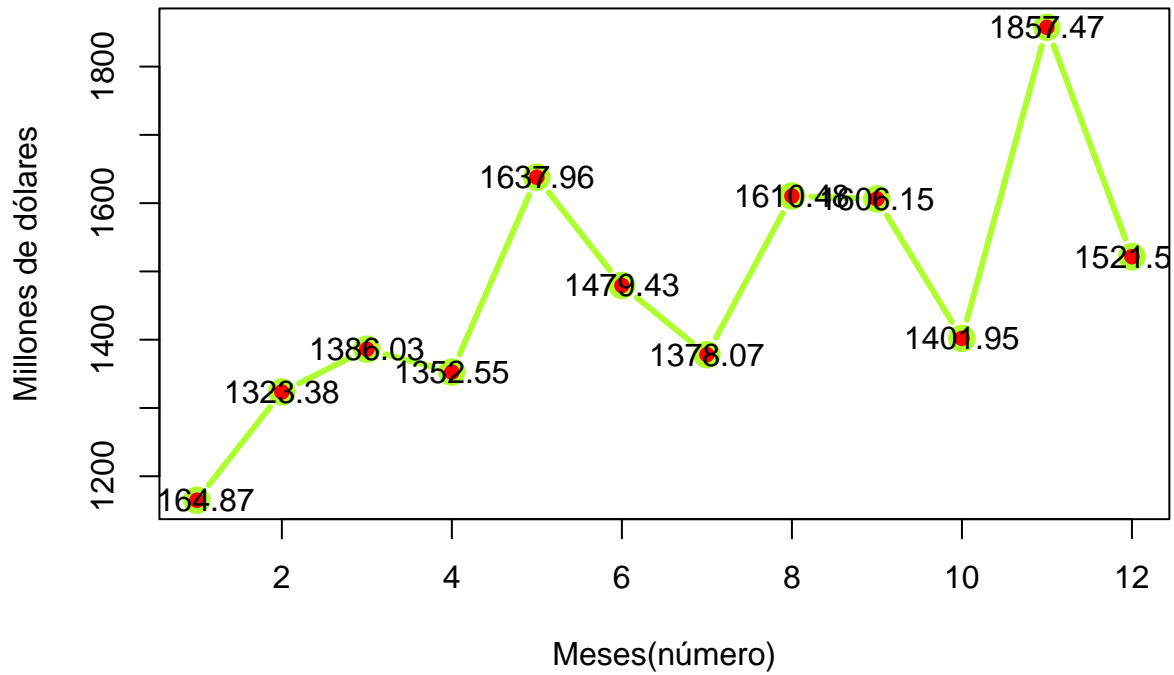
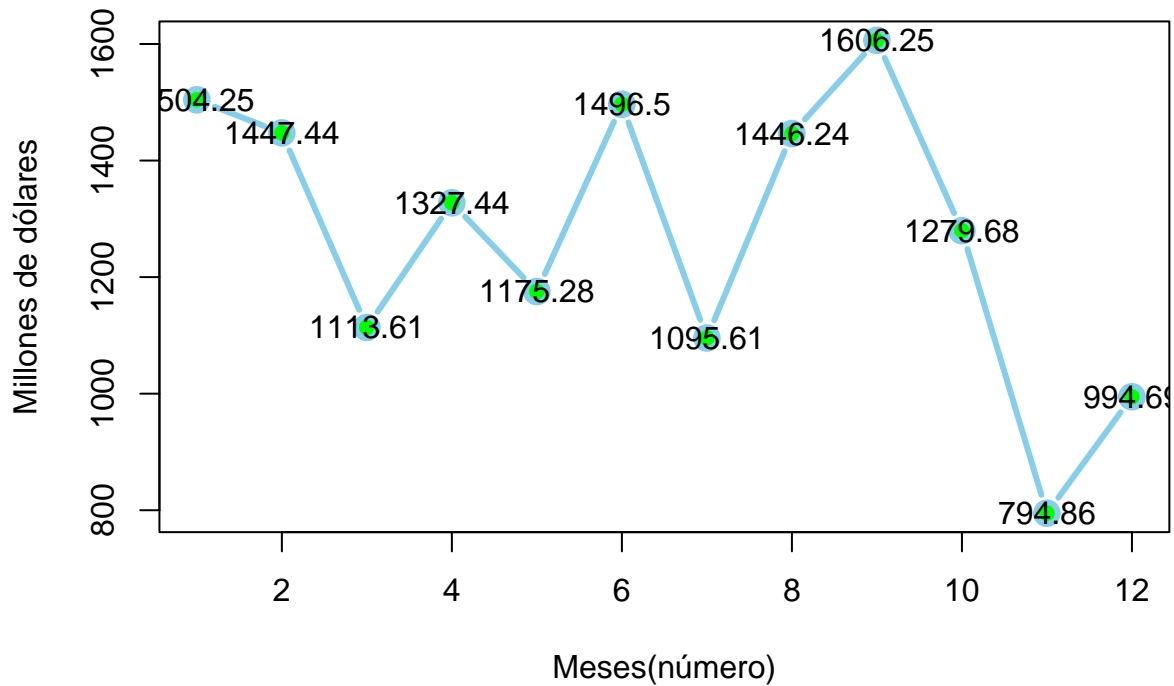


Gráfico 21: Variación de Millones de dólares mensuales del año 2022

- Aquí creamos un gráfico para poder ver la variación de millones en dolares por meses evaluado en el año 2022

```
plot(c(1:length(comparacionDolarTotalItems2022$Meses)),comparacionDolarTotalItems2022$US_Dolar ,type = 'l')
text(x = c(1:length(comparacionDolarTotalItems2022$Meses)),y =comparacionDolarTotalItems2022$US_Dolar,labels = comparacionDolarTotalItems2022$US_Dolar)
```

GR21: Variación de Millones de dólares del año 2022



4.3. Regresión lineal

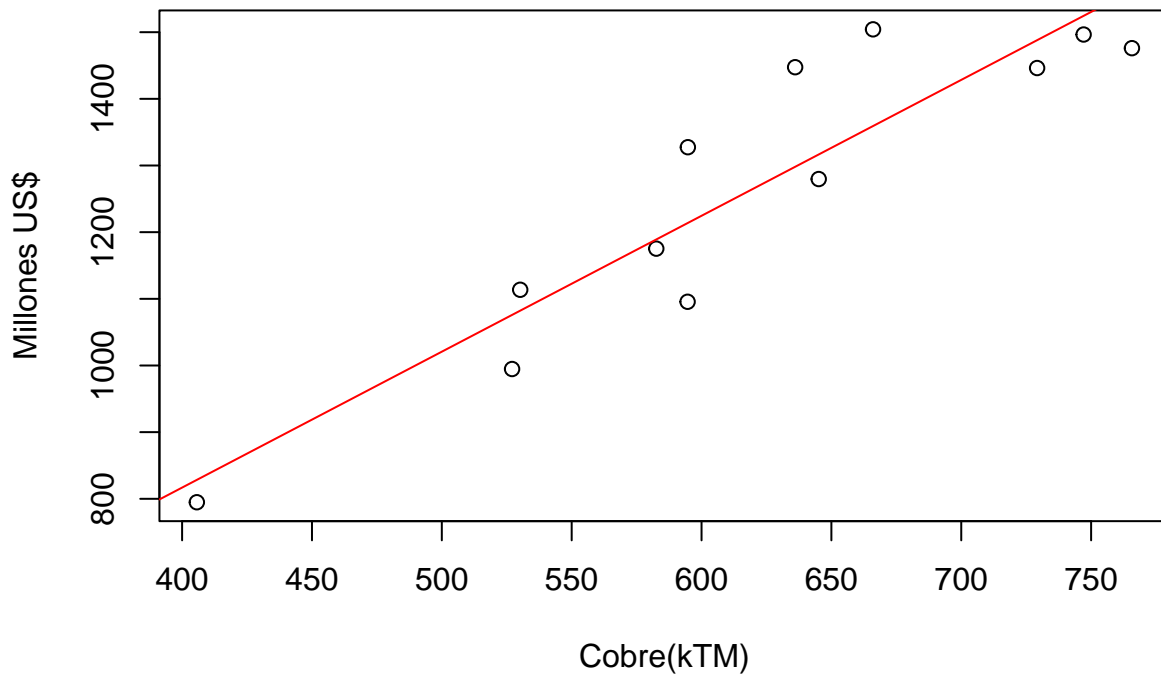
Relación de total de cobre (kTM) y Millones US

1. Creación del modelo de regresión lineal (Total de cobre (kTM) y Millones US\$)

```
modelo <- lm(mesescomparacion2022$newDolarMillion ~ mesescomparacion2022$newKgTon)
```

```
plot(mesescomparacion2022$newKgTon, mesescomparacion2022$newDolarMillion, xlab="Cobre(kTM)", ylab="Millones de dólares",  
abline(modelo, col="red"))
```

GR22:Relacion del total de cobre (kTM) y millones US\$



2. Hallamos la correlación de Total de cobre (kTM) y Millones US

```
cor(mesescomparacion2022$newKgTon,mesescomparacion2022$newDolarMillion,use = "pairwise.complete.obs")
```

```
## [1] 0.9186374
```

3. Resumen del modelo

```
summary(modelo)
```

```
##
## Call:
## lm(formula = mesescomparacion2022$newDolarMillion ~ mesescomparacion2022$newKgTon)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -118.01  -51.59  -30.76   51.93  149.41
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.4011    173.7208   0.008   0.994
## mesescomparacion2022$newKgTon  2.0386     0.2773   7.353 2.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
##
## Residual standard error: 95.1 on 10 degrees of freedom
## Multiple R-squared:  0.8439, Adjusted R-squared:  0.8283
## F-statistic: 54.06 on 1 and 10 DF,  p-value: 2.447e-05
```

Interpretación 1

- Analizando la correlación entre “**Total de cobre (kTM)**” VS “**Millones US**” , tenemos un valor **0.9186374**, lo que me indicaría que hay una relación positiva fuerte entre “Total de cobre (kTM)” y “Millones US”.

Escenario positivo

- Esto implica que a medida que aumente “Total de cobre (kTM)” es probable que también aumente “los Millones US”.
- Si tiene sentido porque si aumenta la cantidad de cobre que se exporta a los países destino, esto implicará un aumento de el número de divisas al Perú.

Resumen del modelo

- De acuerdo al resumen del modelo, podemos indicar con respecto a las estimaciones que por cada 1(kTM) de cobre que se exporta se recibe 2.039 millones de dólares.
- R-cuadrado múltiple de 0.8439 indica que aproximadamente el 84.39% de la variabilidad de la variable “Millones US”. Por otro, el Multiple R-squared o el R al cuadrado es **0.8439** es cercano a 1 por lo que el modelo tiene un buen ajuste a los datos y explica la mayoría de las variaciones en la variable dependiente que es “Total de cobre (kTM)” a través de la relación lineal con la variable independiente que es “Millones US”.

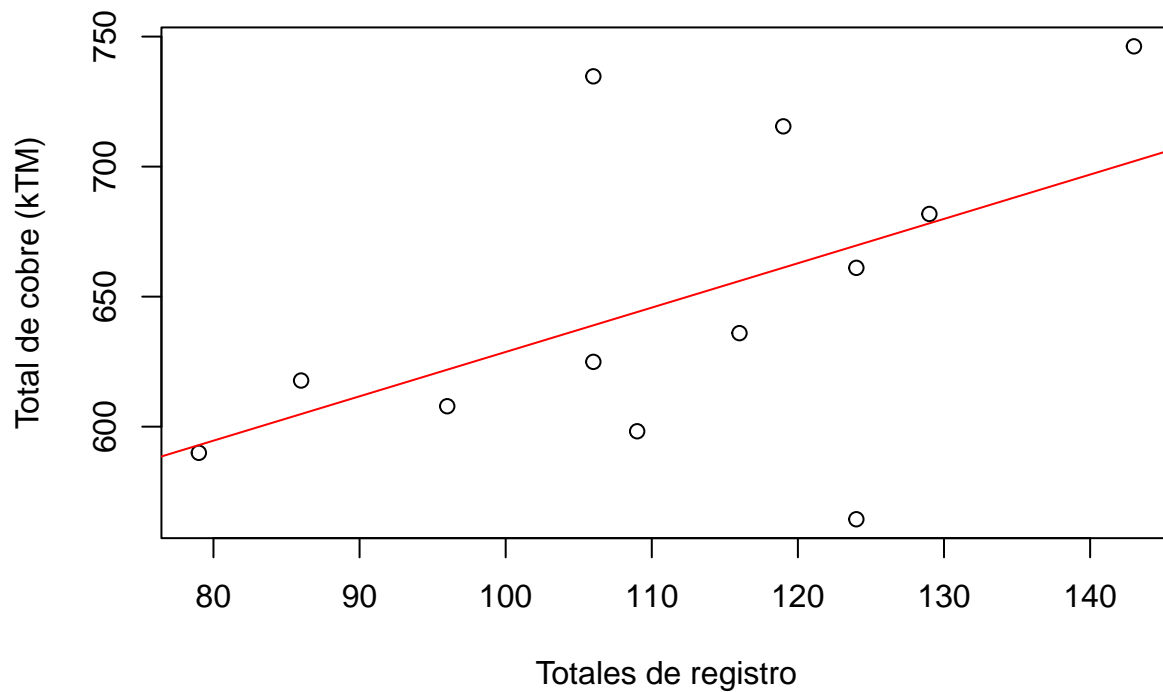
Total de registros vs Total de cobre (kTM)

- Creación del modelo de regresión lineal

```
modelo2<- lm(mesescomparacion2021$newKgTon ~ mesescomparacion2021$sumRegistros)
```

```
plot(mesescomparacion2021$sumRegistros, mesescomparacion2021$newKgTon,xlab="Totales de registro",ylab="
abline(modelo2,col="red")
```

GR 23: Relación de total de registro y cobre (kTM)



2. Hallamos la correlación de Total de cobre (kTM) y Millones US

```
cor(mesescomparacion2021$sumRegistros,mesescomparacion2021$newKgTon,use = "pairwise.complete.obs")
```

```
## [1] 0.5248361
```

3. Resumen del modelo

```
summary(modelo2)
```

```
##
## Call:
## lm(formula = mesescomparacion2021$newKgTon ~ mesescomparacion2021$sumRegistros)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -105.254  -15.552   -5.774   20.713   95.754
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      458.1814     98.6555   4.644 0.000916 ***
## mesescomparacion2021$sumRegistros  1.7054      0.8747   1.950 0.079775 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 53.21 on 10 degrees of freedom
## Multiple R-squared:  0.2755, Adjusted R-squared:  0.203
## F-statistic: 3.802 on 1 and 10 DF,  p-value: 0.07977
```

Interpretación 2

- La correlación hallada determina que existe una correlación positiva moderada entre las variable **“Total de registro VS”** **“Total de cobre (kTM)”**, con un valor de **0.5248361** lo que sugiere que tienden a variar en la misma dirección de manera muy estrecha. Esto implica que cuando los valores de **“Total de kg”** aumentan, es altamente probable que los valores de **“US”** también aumenten, y viceversa.

Resumen del modelo

- De acuerdo al resumen del modelo podemos decir que por 1 registro de exportación se vende 1.705(kTM) de cobre **“Multiple R-squared”** es 0.2755, indica que aproximadamente el 27.55% de la variabilidad de **“Total de cobre (kTM)”** puede ser explicada por el modelo de regresión múltiple y no por una regresión lineal. Esto se debe que el R-cuadrado es relativamente bajo, lo que sugiere que el modelo de regresión múltiple, ya que tiene un ajuste limitado de los datos y no puede explicar una gran parte de la variabilidad observada en la variable **“Total de cobre (kTM)”**.

5. Análisis Probabilístico

Una variable interesante el número de empresas exportadoras peruanas de cobre y sus ventas en el mercado internacional. De acuerdo a la data obtenido de la Sunat y Veritream, se observó el incremento del año 2022 y 2021 de empresas dedicadas a este rubro . Por lo que podemos colegir que este sector genera rentables ganancias. Variable: Empresas exportadoras 2021 y 2022.

Probabilidades Empíricas

Por intersección:

Problema 1:

Algunas empresas exportadoras se mantienen dentro del mercado internacional durante los años 2021 y 2022, debido a los beneficios que la da la Sunat, por ejemplo, están exentas al pago de aranceles e IGV y tienen derecho al saldo a favor del exportador. Por ello, se desea saber, ¿cual es la probabilidad que las empresas ofertaron el cobre 2021 esten presentes en el 2022?

X: El número de empresas exportadoras del año 2021 Y: El número de empresas exportadoras del año 2022

$$A = \{ \text{Empresas exportadoras del año 2021 pero que esten en 2022} \}$$

$$N(A) = 21$$

```
(length(comparacionexport2021$Exportador) - length(intersect(comparacionexport2021$Exportador, comparacionexport2022$Exportador))) / length(comparacionexport2021$Exportador)
```

```
## [1] 21
```

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{21}{71} = 0.2957746$$

```
(length(comparacionexport2021$Exportador) - length(intersect(comparacionexport2021$Exportador, comparacionexport2022$Exportador))) / length(union(comparacionexport2021$Exportador, comparacionexport2022$Exportador))
```

```
## [1] 0.2957746
```

$$B = \{\text{Empresas exportadoras del año 2022 pero que no estan en 2021}\}$$

$$N(B) = 24$$

```
(length(comparacionexport2022$Exportador) - length(intersect(comparacionexport2021$Exportador, comparacionexport2022$Exportador))) / length(union(comparacionexport2021$Exportador, comparacionexport2022$Exportador))
```

```
## [1] 24
```

$$P(B) = \frac{N(B)}{N(\Omega)} = \frac{24}{71} = 0.3380282$$

```
(length(comparacionexport2022$Exportador) - length(intersect(comparacionexport2021$Exportador, comparacionexport2022$Exportador))) / length(union(comparacionexport2021$Exportador, comparacionexport2022$Exportador))
```

```
## [1] 0.3380282
```

$$A \cap B = \{\text{Empresas presentes en el año 2021 y 2022}\}$$

$$N(A \cap B) = 26$$

```
length(intersect(comparacionexport2021$Exportador, comparacionexport2022$Exportador))
```

```
## [1] 26
```

$$P(A \cap B) = \frac{N(A \cap B)}{N(\Omega)} = \frac{26}{71} = 0.3661972$$

```
length(intersect(comparacionexport2021$Exportador, comparacionexport2022$Exportador)) / length(union(comparacionexport2021$Exportador, comparacionexport2022$Exportador))
```

```
## [1] 0.3661972
```

$$\Omega = \{\text{Empresas del año 2021 y 2022}\}$$

$$N(\Omega) = 71$$

```
length(union(comparacionexport2021$Exportador, comparacionexport2022$Exportador))
```

```
## [1] 71
```

$$P(\Omega) = P(A \cap B) + P(A) + P(B) = 1 = 0.3661972 + 0.2957746 + 0.3380282$$

```
0.3661972 + 0.2957746 + 0.3380282
```

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

Analizando :

Número de empresas del 2021 y 2022 : 71

Número de empresas presentes en el 2021 y 2022 : 26

La probabilidad que representa la empresas presentes en el año 2021 y 2022 : 0.37

```
print("La probabilidad que representa el numero de empresas exportadoras esten presentes en 2021 y 2022")
```

```
## [1] "La probabilidad que representa el numero de empresas exportadoras esten presentes en 2021 y 2022"
```

```
round(length(intersect(comparacionexport2021$Exportador,comparacionexport2022$Exportador))/length(union(comparacionexport2021$Exportador,comparacionexport2022$Exportador)),2)
```

```
## [1] 0.37
```

$$P(A \cap B) = \frac{N(A \cap B)}{N(\Omega)} = \frac{26}{71} = 0.37$$

- Por intersección

Problema 2:

¿Cuál es la probabilidad que las empresas exportadoras tengan mayor de un registro durante el año 2021 y 2022 respectivamente?

```
export %>% filter(Anio == 2021) %>% group_by(Exportador) %>% summarise(TotalRegistrosFull = sum(Total_Registros))
```

Análisis 2021 Resolución:

$$\Omega = \{\text{Empresas del 2021}\}$$

$$N(\Omega) = 47$$

```
nrow(exportRegistrosFull2021)
```

```
## [1] 47
```

$$A = \{\text{Empresas que su numero de registros es mayor a 1}\}$$

$$N(A) = 32$$

```
nrow(exportRegistrosFull2021 %>% filter(TotalRegistrosFull > 1))
```

```
## [1] 32
```

$$A^c = \{Empresas \text{ que su numero de registros es menor o igual a } 1\}$$

$$N(A^c) = 15$$

```
nrow(exportRegistrosFull2021 %>% filter(TotalRegistrosFull <= 1))
```

```
## [1] 15
```

$$P(A^c) = \frac{N(A^c)}{N(\Omega)} = \frac{15}{47} =$$

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{32}{47} = 0.6808511$$

```
32/47
```

```
## [1] 0.6808511
```

La probabilidad de que las empresas tengan un numero mayor a 1 es 0.6808511

```
round(nrow(exportRegistrosFull2021 %>% filter(TotalRegistrosFull > 1))/nrow(exportRegistrosFull2021),2)
```

```
## [1] 0.68
```

Análisis 2022 Resolución: $\Omega = \{Empresas \text{ del } 2022\}$ $A = \{Cantidad \text{ de empresas que su número de registro es mayor a } 1\}$
 $P(A) = \frac{N(A)}{N(\Omega)}$

```
export %>% filter(Anio == 2022) %>% group_by(Exportador) %>% summarise(TotalRegistrosFull = sum(Total_R
```

Resolución:

$$\Omega = \{Empresas \text{ del } 2022\}$$

$$N(\Omega) = 50$$

```
nrow(exportRegistrosFull2022)
```

```
## [1] 50
```

$$A = \{Empresas \text{ que su numero de registros es mayor a } 1\}$$

$$N(A) = 35$$

```
nrow(exportRegistrosFull2022 %>% filter(TotalRegistrosFull > 1))
```

```
## [1] 35
```

$$A^c = \{\text{Empresas que su numero de registros es menor o igual a 1}\}$$
$$N(A^c) = 15$$

```
nrow(exportRegistrosFull2022 %>% filter(TotalRegistrosFull <= 1))
```

```
## [1] 15
```

$$P(A^c) = \frac{N(A^c)}{N(\Omega)} = \frac{15}{50} = 0.3$$

```
15/50
```

```
## [1] 0.3
```

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{35}{50} = 0.7$$

```
35/50
```

```
## [1] 0.7
```

La probabilidad de que las empresas tengan un numero mayor a 1 es 0.7

```
round(nrow(exportRegistrosFull2022 %>% filter(TotalRegistrosFull > 1))/nrow(exportRegistrosFull2022),2)
```

```
## [1] 0.7
```

Problema 3

Analizar los datos financieros de los años 2021 y 2022 de la cantidad de ingresos recaudados por las empresas exportadoras es muy interesante, sobre todo debemos estar observando a las empresas que tienen una cantidad de ingresos menor a la media para ver su progreso durante los siguientes años. *Analisis 2021 ##### Analisis 2021* ¿Cuál es la probabilidad de que las empresas exportadoras hallan generado una cantidad de dinero menor a la media de la cantidad de millones de dolares recaudados en 2021?

Resolución:

$$\text{Media de ingresos en millones de dolares} = 377.0193$$

```
mean(comparacionexport2021$newDollarsMillion)
```

```
## [1] 377.0193
```

$$A = \{\text{Empresas que tuvieron ingreso menor a la media en 2021}\}$$

$$N(A) = 37$$

```
nrow(comparacionexport2021 %>% filter(newDollarsMillion < mean(newDollarsMillion)))
```

```
## [1] 37
```

$$N(A^c) = 11$$

```
abs(nrow(comparacionexport2021 %>% filter(newDollarsMillion < mean(newDollarsMillion))) - nrow(comparacionexport2021))
```

```
## [1] 10
```

$$P(A^c) = \frac{10}{47} = \frac{N(A^c)}{N(\Omega)} = 0.212766$$

```
abs(nrow(comparacionexport2021 %>% filter(newDollarsMillion < mean(newDollarsMillion))) - nrow(comparacionexport2021)) / nrow(comparacionexport2021)
```

```
## [1] 0.212766
```

$$\Omega = \{\text{Empresas de 2021}\}$$

$$N(\Omega) = 47$$

```
nrow(comparacionexport2021)
```

```
## [1] 47
```

$$P(A) = \frac{37}{47} = \frac{N(A)}{N(\Omega)} = 0.787234$$

```
nrow(comparacionexport2021 %>% filter(newDollarsMillion < mean(newDollarsMillion))) / nrow(comparacionexport2021)
```

```
## [1] 0.787234
```

```
round(nrow(comparacionexport2021 %>% filter(newDollarsMillion < mean(newDollarsMillion))) / nrow(comparacionexport2021), 2)
```

```
## [1] 0.79
```

La probabilidad del evento atómico A que las empresas del año hayan tenido ingresos menores a la media es 0.787234

Análisis 2022 ¿Cuál es la probabilidad de que las empresas exportadoras hallan generado una cantidad de dinero menor a la media de la cantidad de millones de dolares recaudados en 2022?

Resolución:

$$\text{Media de ingresos en millones de dolares} = 303.0325$$

```
mean(comparacionexport2022$newDollarsMillion)
```

```
## [1] 303.0325
```

$$A = \{\text{Empresas que tuvieron ingreso menor a la media en 2022}\}$$

$$N(A) = 39$$

```
nrow(comparacionexport2022 %>% filter(newDollarsMillion < mean(newDollarsMillion)))
```

```
## [1] 39
```

$$N(A^c) = 11$$

```
abs(nrow(comparacionexport2022 %>% filter(newDollarsMillion < mean(newDollarsMillion))) - nrow(comparacionexport2022))
```

```
## [1] 11
```

$$P(A^c) = \frac{11}{50} = \frac{N(A^c)}{N(\Omega)} = 0.22$$

```
abs(nrow(comparacionexport2022 %>% filter(newDollarsMillion < mean(newDollarsMillion))) - nrow(comparacionexport2022)) / nrow(comparacionexport2022)
```

```
## [1] 0.22
```

$$\Omega = \{\text{Empresas de 2022}\}$$

$$N(\Omega) = 50$$

```
nrow(comparacionexport2022)
```

```
## [1] 50
```

$$P(A) = \frac{39}{50} = \frac{N(A)}{N(\Omega)} = 0.78$$

```
nrow(comparacionexport2022 %>% filter(newDollarsMillion < mean(newDollarsMillion))) / nrow(comparacionexport2022)
```

```
## [1] 0.78
```

```
round(nrow(comparacionexport2022 %>% filter(newDollarsMillion<mean(newDollarsMillion)))/nrow(comparacionexport2022))

## [1] 0.78
```

La probabilidad del evento atómico A que las empresas del año hayan tenido ingresos menores a la media es 0.78

Probabilidades Condicionales

```
export %>% filter(Anio == 2021) %>% group_by(Exportador) %>% summarise(registerFull = sum(Total_Registros))
```

Problema 1:

Las empresas exportadoras suelen recibir solicitudes por parte de empresas importadoras por lo que alguna de ellas tienen un número de registros elevado al compararse con otra por lo general podemos considerar un número de registro elevado a partir del 90, además de ello es muy interesante saber cuáles de ellas superan por encima de la media la cantidad de millones de dólares recaudados. Entonces, si sabe que las empresas que tienen una cantidad de dólares por encima de la media ¿Cuál es la probabilidad de que las empresas hayan tenido un número mayor o igual de 90 registros?

Análisis 2021

$$A = \{\text{Empresas que tienen un número de registro mayor o igual a 90}\}$$

$$B = \{\text{Empresas que tienen una cantidad de millones de dólares mayor a la media}\}$$

Los eventos que hemos usado son dependientes y a continuación la demostración:

$$A = \{\text{Empresas que tienen un número de registro mayor o igual a 90}\}$$

$$B = \{\text{Empresas que tienen una cantidad de millones de dólares mayor a la media}\}$$

$$P(A) = 0.1489362$$

```
nrow(exportDolarRegisFull2021 %>% filter(registerFull >= 90))/nrow(exportDolarRegisFull2021)
```

```
## [1] 0.1489362
```

$$P(B) = 0.212766$$

```
nrow(exportDolarRegisFull2021 %>% filter(DolaresMillion > mean(DolaresMillion)))/nrow(exportDolarRegisFull2021)
```

```
## [1] 0.212766
```

Si los eventos fueran independientes cumplirían esto:

$$P(A/B) = P(A)$$

$$P(B/A) = P(B)$$

$$P(A \cap B) = P(A) \times P(B)$$

Sabiendo que:

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(B/A) \times P(B)}{P(A)} = \frac{P(B \cap A)}{P(A)}$$

$$P(A \cap B) = 0.06382979 = P(B \cap A)$$

```
nrow(exportDolarRegisFull2021%>%filter(registerFull>=90)%>% filter(DolaresMillion > mean(DolaresMillion
```

```
## [1] 0.06382979
```

$$P(A/B) = P(A)$$

$$\frac{0.06382979}{0.212766} \neq 0.1489362$$

$$0.3 \neq 0.1489362$$

```
0.06382979/0.212766 == 0.1489362
```

```
## [1] FALSE
```

$$P(B/A) = P(B)$$

$$\frac{0.06382979}{0.1489362} \neq 0.212766$$

$$0.4285714 \neq 0.212766$$

```
0.06382979/0.1489362
```

```
## [1] 0.4285714
```

$$P(A \cap B) = P(B) \times P(A)$$

$$0.06382979 \neq 0.212766 \times 0.1489362$$

$$0.06382979 \neq 0.03168856$$

```
0.1489362*0.212766 == 0.06382979
```

```
## [1] FALSE
```

Evento Atomico A y B son eventos dependientes

Resolviendo la pregunta de nuestro problema: 1. Analizamos cual es la media de millones de dolares

```
mean(exportDolarRegisFull2021$DolaresMillion)
```

```
## [1] 3770.193
```

2. Analizamos cual es la probabilidad total de que las empresas tenga un numero de dolares recaudados mayor a la media

2.1 Hallamos la probabilidad de $P(B/A)$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = 0.4285714 = \frac{3}{7}$$

```
nrow(exportDolarRegisFull2021 %>% filter(registerFull >= 90) %>% filter(DolaresMillion > mean(DolaresMil
```

```
## [1] 0.4285714
```

2.2 Hallamos la probabilidad de que las empresas tenga un numero menor a 90 de registros y su cantidad de dolares sea mayor a la media

Probabilidad $P(B/A^c)$:

$$P(B/A^c) = \frac{P(B \cap A^c)}{P(A^c)} = 0.175 = \frac{7}{40}$$

```
nrow(exportDolarRegisFull2021 %>% filter(registerFull < 90) %>% filter(DolaresMillion > mean(DolaresMil
```

```
## [1] 0.175
```

2.3 Sumamos esas 2 probabilidad anteriores y multiplicadas por sus correspondientes eventos A $P(B/A^c) * P(A^c) + P(B/A) * P(A)$

Calculando la probabilidad de A ($P(A)$)

$$P(A) = \frac{N(A)}{N(\Omega)} = 0.1489362 = \frac{7}{47}$$

```
nrow(exportDolarRegisFull2021 %>% filter(registerFull >= 90))/nrow(exportDolarRegisFull2021)
```

```
## [1] 0.1489362
```

Calculando la probabilidad de A^c ($P(A^c)$)

$$P(A^c) = \frac{N(A^c)}{N(\Omega)} = 0.8510638 = \frac{40}{47}$$

```
nrow(exportDolarRegisFull2021 %>% filter(registerFull < 90))/nrow(exportDolarRegisFull2021)
```

```
## [1] 0.8510638
```

```
nrow(exportDolarRegisFull2021 %>% filter(registerFull < 90))
```

```
## [1] 40
```

La probabilidad total de que las empresas tengan una cantidad de dolares por encima de la media $P(B)$

```
(0.4285714*0.1489362 + 0.175 *0.8510638)
```

```
## [1] 0.212766
```

$A = \{\text{Empresas que tienen un numero de registro mayor o igual a 90}\}$

$B = \{\text{Empresas que tienen una cantidad de millones de dolares mayor a la media}\}$

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$

2.5 Aplicamos el teorema de Bayes hallamos la probabilidad de que las empresas tenga un numero mayor o igual a 90 registros sabiendo que superaban a la media d es:

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} = \frac{0.4285714 * 0.1489362}{0.212766} = 0.3$$

```
(0.4285714*0.1489362)/0.212766
```

```
## [1] 0.3
```

La probabilidad de que las empresas tengan un numero de registros mayor o igual a 90 dado que se sabia que superaban la media es:

Analisis 2022

$A = \{\text{Empresas que tienen un numero de registro mayor o igual a 90}\}$

$B = \{\text{Empresas que tienen una cantidad de millones de dolares mayor a la media}\}$

```
export %>% filter(Anio == 2022) %>% group_by(Exportador) %>% summarise(registerFull = sum(Total_Registrados))
```

Resolviendo la pregunta de nuestro problema: 1. Analizamos cual es la media de millones de dolares

```
mean(exportDolarRegisFull2022$DolaresMillion)
```

```
## [1] 3030.325
```

2. Analizamos cual es la probabilidad total de que las empresas tenga un numero de dolares recaudados mayor a la media

2.1 Hallamos la probabilidad de $P(B/A)$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = 0.4 = \frac{2}{5}$$

```
nrow(exportDolarRegisFull12022 %>% filter(registerFull >= 90) %>% filter(DolaresMillion > mean(DolaresMil
```

```
## [1] 0.4
```

2.2 Hallamos la probabilidad de $P(B/A^c)$

Probabilidad $P(B/A^c)$:

$$P(B/A^c) = \frac{P(A^c \cap B)}{P(A^c)} = 0.2222222 = \frac{10}{45}$$

```
nrow(exportDolarRegisFull12022 %>% filter(registerFull < 90) %>% filter(DolaresMillion > mean(DolaresMil
```

```
## [1] 0.2222222
```

2.3 Sumamos esas 2 probabilidad anteriores y multiplicadas por sus correspondientes eventos A $P(B/A^c) * P(A^c) + P(B/A) * P(A)$

Calculando la probabilidad de A ($P(A)$)

$$P(A) = 0.1 = \frac{5}{50}$$

```
nrow(exportDolarRegisFull12022 %>% filter(registerFull >= 90))/nrow(exportDolarRegisFull12022)
```

```
## [1] 0.1
```

```
nrow(exportDolarRegisFull12022 %>% filter(registerFull >= 90))
```

```
## [1] 5
```

Calculando la probabilidad de A^c $P(A^c)$

$$P(A^c) = 0.9 = \frac{45}{50}$$

```
nrow(exportDolarRegisFull12022 %>% filter(registerFull < 90))/nrow(exportDolarRegisFull12022)
```

```
## [1] 0.9
```

```
nrow(exportDolarRegisFull12022 %>% filter(registerFull < 90))
```

```
## [1] 45
```

La probabilidad total de que las empresas tengan una cantidad de dolares por encima de la media $P(B) = P(B/A) * P(A) + P(B/A^c) * P(A^c)$

$$P(B) = P(B/A) * P(A) + P(B/A^c) * P(A^c)$$

\$

```
(0.1*0.4)+(0.9*0.2222222)
```

```
## [1] 0.24
```

2.4 Analizando el problema

$A = \{\text{Empresas que tienen un numero de registro mayor o igual a 90}\}$

$B = \{\text{Empresas que tienen una cantidad de millones de dolares mayor a la media}\}$

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$

2.5 Aplicamos el teorema de Bayes hallamos la probabilidad de que las empresas tenga un numero mayor o igual a 90 registros sabiendo que superaban a la media es:

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$

```
(0.1*0.4)/(0.24)
```

```
## [1] 0.1666667
```

```
###Problema 2 ##### Analisis 2022
```

```
subset(export,Meses == "Diciembre" & Anio == 2022)-> empresas_diciembre2022
```

$\Omega = \text{Empresas del mes de diciembre de 2022}$

$N(\Omega) = \text{Numero de empresas del mes de diciembre de 2022} = 25$

```
nrow(empresas_diciembre2022)
```

```
## [1] 25
```

$A = \{\text{Empresas con más de 2 registros en el mes de Diciembre}\}$

$B = \{\text{Empresas que exportaron más de 10433357.50kg de cobre}\}$

$$P(A) = \frac{11}{25} = \frac{N(A)}{N(\Omega)} = 0.44$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros > 2))/nrow(empresas_diciembre2022)
```

```
## [1] 0.44
```

$$P(B) = \frac{10}{25} = \frac{N(B)}{N(\Omega)} = 0.4$$

```
nrow(empresas_diciembre2022 %>% filter(Total_KG > 10433357.50))/nrow(empresas_diciembre2022)
```

```
## [1] 0.4
```

Vamos a verificar que nuestros eventos son dependientes:

Si los eventos fueran independientes cumplirían esto:

$$P(A/B) = P(A)P(B/A) = P(B)P(A \cap B) = P(A) \times P(B)$$

Sabiendo que:

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(B/A) \times P(B)}{P(A)} = \frac{P(B \cap A)}{P(A)}$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = 0.7$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros > 2 & Total_KG >10433357.50))/nrow(empresas_diciembre2022)
```

```
## [1] 0.7
```

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = 0.6363636$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros > 2 & Total_KG >10433357.50))/nrow(empresas_diciembre2022)
```

```
## [1] 0.6363636
```

$$P(A) = \frac{N(A)}{N(\Omega)} = 0.44$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros > 2))/nrow(empresas_diciembre2022)
```

```
## [1] 0.44
```

$$P(B) = \frac{N(B)}{N(\Omega)} = 0.4$$

```
nrow(empresas_diciembre2022 %>% filter(Total_KG >10433357.50))/nrow(empresas_diciembre2022)
```

```
## [1] 0.4
```

$$P(A \cap B) = \frac{N(A \cap B)}{N(\Omega)} = 0.28$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros > 2 & Total_KG >10433357.50))/nrow(empresas_diciembre2022)
```

```
## [1] 0.28
```

Comprobando:

$$P(A/B) = P(A)0.7 \neq 0.44$$


```
0.7 == 0.44
```

```
## [1] FALSE
```

$$P(B/A) = P(B)0.6363636 \neq 0.4$$

```
0.6363636 == 0.4
```

```
## [1] FALSE
```

$$P(A \cap B) = P(B) \times P(A)$$

$$0.28 \neq 0.4 \times 0.44$$

$$0.28 \neq 0.176$$

```
0.28 == 0.176
```

```
## [1] FALSE
```

Evento Atomico A y B son dependientes porque no cumplen las condiciones para ser independientes

Planteando el problema: Es muy interesante saber que tal fue la exportación de cobre durante el ultimo mes del año 2022 , sobre todo la cantidad de cobre que fue exportada por arriba de los 10433357.50Kg y verificar que su numero de registro es mayor a 2. Por lo que debemos preguntarnos ahora ¿Cual es la probabilidad de que las empresas tengan un numero mayor a 2 registros ,si sabemos que esta tiene una cantidad de kilogramos por encima de los 10433357.50Kg?

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros > 2))
```

```
## [1] 11
```

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros <= 2))
```

```
## [1] 14
```

```
nrow(empresas_diciembre2022)
```

```
## [1] 25
```

$A = \{\text{Empresas con más de 2 registros en el mes de Diciembre}\}$

$B = \{\text{Empresas que exportaron más de 10433357.50kg de cobre}\}$

Hallamos la probabilidad total de que las empresas exportaron más de 10433357.50kg de cobre

$$P(A^c) = \frac{N(A^c)}{N(\Omega)} = 0.56$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros <= 2))/nrow(empresas_diciembre2022)
```

```
## [1] 0.56
```

$$P(A^c \cap B) = 0.12$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros <= 2 & Total_KG > 10433357.50))/nrow(empresas_diciembre2022)
```

```
## [1] 0.12
```

$$P(A \cap B) = 0.28$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros > 2 & Total_KG > 10433357.50))/nrow(empresas_diciembre2022)
```

```
## [1] 0.28
```

Teorema de la probabilidad total :

$$P(B) = P(B/A) * P(A) + P(B/A^c) * P(A^c)$$

$$P(B/A^c) = \frac{P(A^c \cap B)}{P(A^c)} = \frac{0.12}{0.56} = 0.2142857$$

```
nrow(empresas_diciembre2022 %>% filter(Total_Registros <= 2 & Total_KG > 10433357.50))/nrow(empresas_diciembre2022)
```

```
## [1] 0.12
```

```
0.12/0.56
```

```
## [1] 0.2142857
```

$$P(B) = (0.6363636 * 0.44) + (0.2142857 * 0.56) = 0.4$$

```
(0.6363636 * 0.44) + (0.2142857*0.56)
```

```
## [1] 0.4
```

Teorema de Bayes

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.28}{0.4} = 0.7$$

```
0.28/0.4
```

```
## [1] 0.7
```

La probabilidad de que las empresas sean de un número de registro mayor a 2 dado que la cantidad que exporto en el mes de

###Problema 3 ###Análisis 2021

```
países %>% filter(Anio == 2021) %>% group_by(Pais_Destino) %>% summarise(newKgTon = round(sum(Total_KG), 1))
```

```
países_KgTon2021
```

```
## # A tibble: 21 x 2
##   Pais_Destino newKgTon
##   <chr>         <dbl>
## 1 CHINA         5623.
## 2 JAPÓN         599.
## 3 COREA DEL SUR  424.
## 4 ALEMANIA      284.
## 5 ESPAÑA        150.
## 6 INDIA         141.
## 7 BULGARIA      96.0
## 8 FILIPINAS     84.5
## 9 TAIWÁN        75.0
## 10 FINLANDIA    70.9
## # i 11 more rows
```

```
mean(países_KgTon2021$newKgTon)
```

```
## [1] 370.3966
```

$A = \{\text{Empresas que pertenecen a Asia}\}$

$B = \{\text{Empresas con cantidad de cobre exportado en países mayor a la media de 370.3966 Tm}\}$

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{7}{21} = 0.3333333$$

```
nrow(países_KgTon2021 %>% filter(Pais_Destino %in% c("TAIWÁN", "INDIA", "FILIPINAS", "CHINA", "JAPÓN", "COREA DEL SUR"))
```

```
## [1] 0.3333333
```

$$P(B) = \frac{N(B)}{N(\Omega)} = \frac{3}{21} = 0.1428571$$

```
nrow(países_KgTon2021 %>% filter(newKgTon > mean(newKgTon)))/nrow(países_KgTon2021)
```

```
## [1] 0.1428571
```

Si los eventos fueran independientes cumplirían esto:

$$P(A/B) = P(A)$$

$$P(B/A) = P(B)$$

$$P(A \cap B) = P(A) \times P(B)$$

Sabiendo que:

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(B/A) \times P(B)}{P(A)} = \frac{P(B \cap A)}{P(A)}$$

$$P(A \cap B) = \frac{3}{21} = \frac{N(A \cap B)}{N(\Omega)} = 0.1428571 = P(B \cap A)$$

```
nrow(países_KgTon2021 %>% filter(newKgTon > mean(newKgTon) & País_Destino %in% c("TAIWÁN", "INDIA", "FILIPINAS"))
```

```
## [1] 0.1428571
```

$$P(A/B) = \frac{0.1428571}{0.1428571} = 1$$

$$P(A/B) = P(A)1 \neq 0.3333333$$

```
1 == 0.3333333
```

```
## [1] FALSE
```

$$P(B/A) = \frac{0.1428571}{0.3333333} = \frac{P(B \cap A)}{P(A)} = 0.4285713$$

```
0.1428571/0.3333333
```

```
## [1] 0.4285713
```

$$P(B/A) = P(B)0.4285713 \neq 0.1428571$$

```
0.4285713 == 0.1428571
```

```
## [1] FALSE
```

$$P(A \cap B) = P(B) * P(A)$$

$$0.1428571 \neq 0.1428571 * 0.3333333$$

$$0.1428571 \neq 0.04761903$$

```
0.1428571 * 0.3333333
```

```
## [1] 0.04761903
```

Los atomicos A y B son dependientes porque no cumple ninguna de las propiedades para ser independientes

Planteando el problema: La mayoría de las exportaciones de cobre son gracias a los países asiáticos debido a su gran afluencia en el mercado del comercio minero, se tiene un reporte sobre los países del año 2021 y la cantidad de Tm de cobre que importaron con el fin de saber quienes superaron la media de ese año. ¿Cuál es la probabilidad de que la empresa pertenezca a un país asiático si la cantidad que importó durante ese año es mayor a la media?

$$A = \{\text{Países destino de Asia a los que se exportó cobre}\}$$

$$B = \{\text{Países que exportaron una cantidad mayor a la media de Tm de cobre}\}$$

$$\text{Media de Tm de cobre} = 370.3966$$

```
mean(países_KgTon2021$newKgTon)
```

```
## [1] 370.3966
```

$$P(B/A^c) = \frac{P(A^c \cap B)}{P(A^c)} = \frac{0}{0.6666667} = 0$$
$$P(A^c \cap B) = \frac{0}{21} = \frac{N(A^c \cap B)}{N(\Omega)} = 0$$

```
nrow(países_KgTon2021 %>% filter(newKgTon > mean(newKgTon) & !(Pais_Destino %in% c("TAIWÁN", "INDIA", "F
```

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

$$P(A^c) = \frac{14}{21} = 0.6666667$$

```
nrow(países_KgTon2021 %>% filter(!(Pais_Destino %in% c("TAIWÁN", "INDIA", "FILIPINAS", "CHINA", "JAPÓN", "C
```

```
## [1] 0.6666667
```

Teorema de la probabilidad total

$$P(B) = P(B/A) * P(A) + P(B/A^c) * P(A^c)$$

$$P(B) = 0.4285713 * 0.3333333 + 0 * 0.6666667$$

$$P(B) = 0.1428571$$

```
0.4285713*0.3333333
```

```
## [1] 0.1428571
```

Teorema de Bayes

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1428571}{0.1428571} = 1$$

```
0.1428571/0.1428571
```

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

La probabilidad de que el país sea asiático dado que su Tm importadas superan a la media es 1 es decir un evento seguro

Distribuciones

Distribución de Variables Discretas

Distribución de Bernoulli ¿Cuál es la probabilidad que al país destino que se exporte no sea china?

X : Número de países destino que no son China

```
fi=table(países$Pais_Destino[países$Anio == 2021])
pi=round(prop.table(fi),4)
cbind(fi,pi)
```

```
##           fi      pi
## ALEMANIA    10 0.0758
## AUSTRALIA     1 0.0076
## BRASIL       3 0.0227
## BULGARIA      5 0.0379
## CANADÁ        9 0.0682
## CHILE         7 0.0530
## CHINA        12 0.0909
## COREA DEL SUR 12 0.0909
## ESPAÑA       11 0.0833
## FILIPINAS     6 0.0455
## FINLANDIA     5 0.0379
## INDIA         6 0.0455
## JAPÓN        11 0.0833
## MALASIA      12 0.0909
## MÉXICO        2 0.0152
## NAMIBIA       2 0.0152
## POLONIA       3 0.0227
## SINGAPUR      1 0.0076
## SUECIA        3 0.0227
## TAIWÁN       10 0.0758
## ZONA FRANCA PERÚ 1 0.0076
```

```
sum(pi)
```

```
## [1] 1.0002
```

```
data.frame(table(paises$Pais_Destino[paises$Anio == 2021 & (paises$Pais_Destino != "China" | paises$Pais_Destino == "China"])
```

```
##           Var1 Freq
## 1      ALEMANIA    10
## 2    AUSTRALIA     1
## 3      BRASIL     3
## 4    BULGARIA     5
## 5    CANADÁ       9
## 6      CHILE       7
## 7      CHINA     12
## 8  COREA DEL SUR    12
## 9     ESPAÑA     11
## 10   FILIPINAS     6
## 11   FINLANDIA     5
## 12      INDIA       6
## 13     JAPÓN      11
## 14    MALASIA     12
## 15     MÉXICO       2
## 16    NAMIBIA       2
## 17    POLONIA       3
## 18   SINGAPUR       1
## 19     SUECIA       3
## 20    TAIWÁN      10
## 21 ZONA FRANCA PERÚ    1
```

```
print("Cantidad de Frecuencia de No aparición de China")
```

```
## [1] "Cantidad de Frecuencia de No aparición de China"
```

```
sum(data.frame(table(paises$Pais_Destino[paises$Anio == 2021 & (paises$Pais_Destino != "China" | paises$Pais_Destino == "China"])
```

```
## [1] 132
```

```
print("Cantidad de Frecuencia de aparición de China")
```

```
## [1] "Cantidad de Frecuencia de aparición de China"
```

```
sum(data.frame(table(paises$Pais_Destino[paises$Anio == 2021 & (paises$Pais_Destino == "China" | paises$Pais_Destino != "China"])
```

```
## [1] 12
```

```
A=round(12/(132+12),4)
```

```
print("Por complementos de conjuntos")
```

```
## [1] "Por complementos de conjuntos"
```

```
CA=1-A
CA
```

```
## [1] 0.9167
```

$A^C = 0.9167$

RPTA: La probabilidad que al país destino que se exporte no sea china es 0.9167

Distribución de Bernoulli $P(X = x) = \begin{cases} p^x (1 - p)^{1-x}, & x = 0, 1 \\ 0, & \text{otro caso} \end{cases}$

```
fi=table(paises$Pais_Destino[paises$Anio == 2021])
pi=round(prop.table(fi),4)
cbind(fi,pi)
```

```
##           fi      pi
## ALEMANIA    10 0.0758
## AUSTRALIA     1 0.0076
## BRASIL      3 0.0227
## BULGARIA     5 0.0379
## CANADÁ       9 0.0682
## CHILE        7 0.0530
## CHINA       12 0.0909
## COREA DEL SUR 12 0.0909
## ESPAÑA      11 0.0833
## FILIPINAS    6 0.0455
## FINLANDIA    5 0.0379
## INDIA        6 0.0455
## JAPÓN       11 0.0833
## MALASIA     12 0.0909
## MÉXICO      2 0.0152
## NAMIBIA     2 0.0152
## POLONIA     3 0.0227
## SINGAPUR    1 0.0076
## SUECIA      3 0.0227
## TAIWÁN     10 0.0758
## ZONA FRANCA PERÚ 1 0.0076
```

Solucion :

$X = \begin{cases} 1 & \text{si el importador no es China} \\ 0 & \text{si el importador si es China} \end{cases} \quad P(X = x) = \begin{cases} 0.9167^x (1 - 0.9167)^{1-x}, & x = 0, 1 \\ 0, & \text{otro caso} \end{cases}$

Donde : $E(X) = 0.9167$ y $V(X) = 0.9167(0.0833) = 0.07636111$

```
barplot(c(0.9167, 1 - 0.9167), names.arg = nombres <- c("1", "0"), xlab = "Exito")
```


GR24: Gráfico Bernoulli de los países destinos excepto China

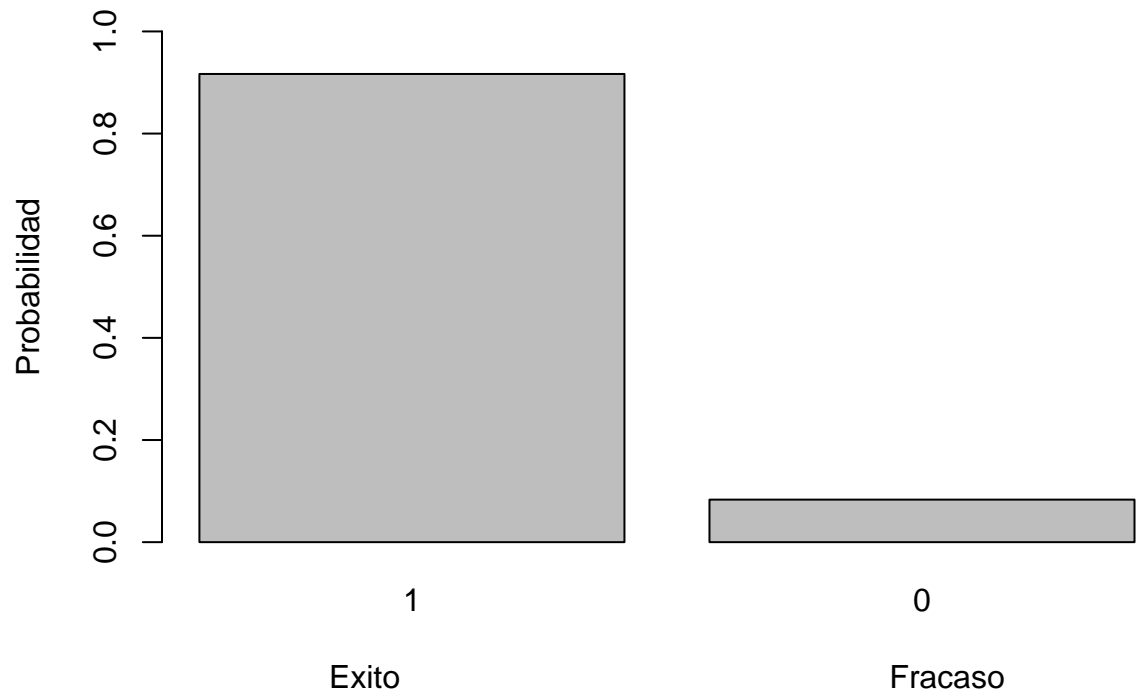


Grafico 24: Bernoulli

Interpretación

En la gráfica, se puede observar que la barra correspondiente al valor 1 (importador no es China) es significativamente más alta que la barra correspondiente al valor 0 (importador es China). Esto se debe a que el porcentaje de veces que el importador no es China es del 91.67% (0.9167), mientras que el porcentaje de veces que el importador es China es del 8.33% (1 - 0.9167).

La interpretación de esta gráfica es que hay una alta probabilidad de que el importador no sea China en comparación con la probabilidad de que el importador sea China. El hecho de que la barra correspondiente al valor 1 sea mucho más alta indica que la mayoría de las veces el importador no es China.

Distribución Binomial Si el 31.91% de empresas tienen 1 registro de exportación en el 2021 ¿Cuál es la probabilidad de que entre 10 empresas elegidas al azar, a lo más halla 3 empresas que tengan solo un registro? X : Número de empresas que tienen 1 registro

```
sum(data.frame(table(exportDolarRegisFull2021$registerFull))$Freq)
```

```
## [1] 47
```

```
data.frame(table(exportDolarRegisFull2021$registerFull))
```

```
##      Var1 Freq
## 1      1    15
## 2      2     4
## 3      3     4
## 4      4     1
```

```
## 5      5      1
## 6      8      1
## 7      9      1
## 8     11      2
## 9     12      1
## 10     14      1
## 11     15      1
## 12     16      1
## 13     18      1
## 14     28      1
## 15     31      1
## 16     35      1
## 17     36      1
## 18     48      1
## 19     60      1
## 20     90      1
## 21    102      1
## 22    116      1
## 23    140      1
## 24    149      1
## 25    160      1
## 26    184      1
```

```
round(15/sum(data.frame(table(exportDolarRegisFull2021$registerFull))$Freq),2)
```

```
## [1] 0.32
```

```
round((15/sum(data.frame(table(exportDolarRegisFull2021$registerFull))$Freq))*100,2)
```

```
## [1] 31.91
```

```
round(pbinom(3,10,0.3191),2)
```

```
## [1] 0.6
```

RPTA: La probabilidad de encontrar 3 empresas que solo tienen un registro dentro de 10 empresas es 0.6

```
barplot(pbinom(c(0:3),10,0.3191), names.arg = c(0:3), xlab = "Número de éxitos", ylab = "Probabilidad a
```

GR25: Distribución Binomial de empresas con 1 registro

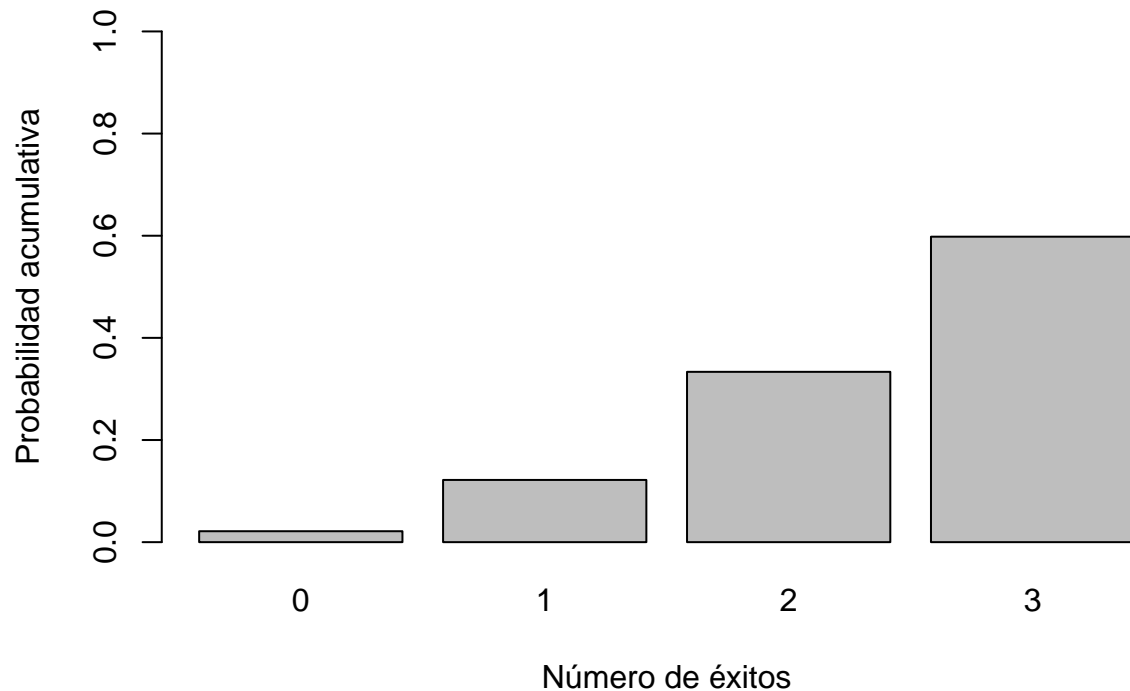


Grafico 25: Binomial

Interpretación

La gráfica binomial muestra la distribución de probabilidad acumulada de la variable aleatoria X, que representa el número de empresas que tienen solo un registro de exportación entre 10 empresas elegidas al azar.

Distribución de Variables Continua

Problema 1

```
export %>% filter(Anio == 2021) %>% group_by(Exportador) %>% summarise(new_Ton = sum(Total_KG)) -> exportEmpresasTonFull2021
exportEmpresasTonFull2021$new_Ton = round(exportEmpresasTonFull2021$new_Ton/1000000,4)
exportEmpresasTonFull2021 = exportEmpresasTonFull2021[order(exportEmpresasTonFull2021$new_Ton, decreasing=TRUE)]
```

```
fit.cont(exportEmpresasTonFull2021$new_Ton)
```

Análisis de tipo de distribución de la variable cobre(TM)

```
##
## Begin fitting distributions -----
## * fitting normal distribution ... OK
```

```

## * fitting Cauchy distribution ... OK

## * fitting logistic distribution ... OK

## * fitting beta distribution ... failed

## * fitting exponential distribution ... OK

## * fitting chi-square distribution ... OK

## * fitting uniform distribution ... OK

## * fitting gamma distribution ... failed

## * fitting lognormal distribution ... OK

## * fitting Weibull distribution ... OK

## * fitting F-distribution ... OK

## * fitting Student's t-distribution ... OK

## * fitting Gompertz distribution ... failed

## * fitting triangular distribution ... failed

## End fitting distributions -----

##          logL      AIC      BIC Chisq(value) Chisq(p) AD(value)
## Normal      -345.68  695.36  699.06      5546.33      0.00      9.52
## Cauchy      -256.85  517.7   521.4       680.40      0.00     25.23
## Logistic    -334.37  672.74  676.44      3292.85      0.00      8.41
## Exponential -287.12  576.24  578.09     51569.53      0.00     76.79
## Chi-square  -3918.05 7838.11 7839.96          Inf      0.00      Inf
## Uniform      NULL     NULL     NULL      6978.68      0.00      Inf
## Lognormal   -166.83  337.67  341.37        6.76      0.24      0.81
## Weibull     -164.6   333.2   336.9         5.14      0.40      0.52
## F           -169.11  342.22  345.92         8.45      0.13      1.18
## Student     -227.17  456.34  458.19        38.83      0.00     15.16
##          H(AD) KS(value)          H(KS)
## Normal      rejected      0.35    rejected
## Cauchy      rejected      0.40    rejected
## Logistic    rejected      0.37    rejected
## Exponential rejected      0.60    rejected
## Chi-square      NULL      0.39    rejected
## Uniform      NULL      0.37    rejected
## Lognormal    rejected      0.10 not rejected
## Weibull      not rejected      0.10 not rejected
## F           NULL      0.14 not rejected
## Student      NULL      0.50    rejected

##
## Chosen continuous distribution is: Log-normal (lnorm)
## Fitted parameters are:
##   meanlog      sdlog
## 0.5880079 4.6771341

```

Distribución Log Normal Parámetros: Media y desviación estándar - Cálculo de la media y la desviación estándar

```
sd(exportEmpresasTonFull12021$new_Ton)
```

```
## [1] 382.4758
```

```
mean(exportEmpresasTonFull12021$new_Ton)
```

```
## [1] 165.4964
```

La cantidad de kg exportados por empresa al año es una variable aleatoria continua X (expresada en TM), donde su media es 165.4964 y desviación estándar 382.4758. ¿Cuál es la probabilidad de que la cantidad de cobre exportado supere las 500 TM? X : cantidad de cobre exportados al año por cada empresa expresado TM $P(X > 500) = 1 - P(X \leq 500) = 1 - F(500)$

```
1-plnorm(500,meanlog= log(165.4964),sdlog = log(1 + (382.4758^2 / 165.4964^2)))
```

```
## [1] 0.2747175
```

RPTA: La probabilidad de que la cantidad de cobre exportado supere las 500 TM es 0.27

```
# Parámetros de la distribución lognormal
media <- 165.4964
desviacion <- 382.4758

# Valor límite
limite <- 500

# Cálculo de la probabilidad
probabilidad <- 1 - plnorm(limite, meanlog = log(media), sdlog = log(desviacion))

# Gráfica de la distribución lognormal
curve(dlnorm(x, meanlog = log(media), sdlog = log(desviacion)),
      xlim = c(0, 1000),
      ylim = c(0, 0.02),
      xlab = "Cobre (TM)",
      ylab = "Densidad",
      main = "GR26: Distribución Lognormal del cobre(TM)")

# Área sombreada correspondiente a la probabilidad
x <- seq(limite, 1000, length.out = 100)
y <- dlnorm(x, meanlog = log(media), sdlog = log(desviacion))
polygon(c(limite, x, 1000), c(0, y, 0), col = "skyblue")

# Texto de la probabilidad
text(600, 0.015, paste("Probabilidad =", round(probabilidad, 4)))
```

```
# Línea vertical en el límite
abline(v = limite, lty = 2)

# Leyenda
legend("topright", legend = c("Área de probabilidad", "Límite"),
      fill = c("skyblue", "white"), lty = c(0, 2))
```

GR26: Distribución Lognormal del cobre

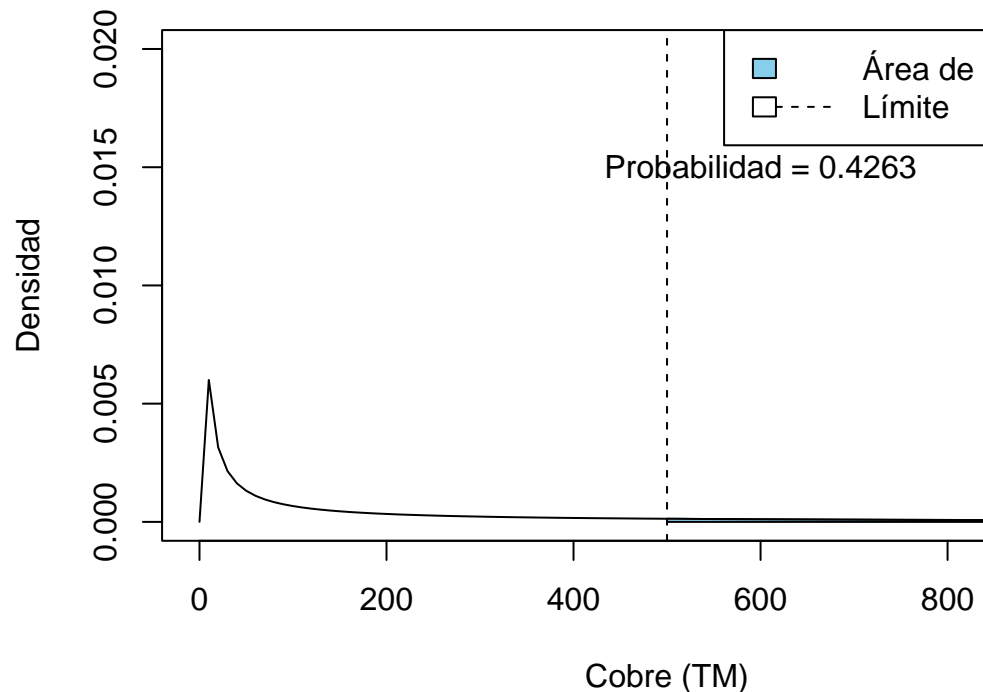


Gráfico 26: Densidad Acumulada

Interpretación

“Probabilidad = 0.27”, lo que significa que hay un 27% de probabilidad de que una empresa exporte más de 500 TM de cobre en un año. En resumen, la gráfica representa la distribución lognormal del cobre exportado y muestra visualmente la probabilidad de que la cantidad de cobre supere un determinado valor, en este caso, 500 TM. ### Problema 2

```
fit.cont(comparacionexport2021$newDollarsMillion)
```

Análisis de la distribución de la variable continua millones de dólares

```
##
## Begin fitting distributions -----
## * fitting normal distribution ... OK
```

```

## * fitting Cauchy distribution ... OK

## * fitting logistic distribution ... OK

## * fitting beta distribution ... failed

## * fitting exponential distribution ... OK

## * fitting chi-square distribution ... OK

## * fitting uniform distribution ... OK

## * fitting gamma distribution ... failed

## * fitting lognormal distribution ... OK

## * fitting Weibull distribution ... OK

## * fitting F-distribution ... OK

## * fitting Student's t-distribution ... OK

## * fitting Gompertz distribution ... failed

## * fitting triangular distribution ... failed

## End fitting distributions -----

##          logL      AIC      BIC Chisq(value) Chisq(p) AD(value)
## Normal      -383.64  771.27  774.97   22916.58    0.00    9.45
## Cauchy      -295.97  595.93  599.63   1408.06    0.00   34.08
## Logistic    -372.69  749.38  753.08  13701.98    0.00    8.16
## Exponential -325.82  653.64  655.49  93900.39    0.00   91.09
## Chi-square  -8884.73 17771.46 17773.31      Inf    0.00    Inf
## Uniform      NULL      NULL      NULL   28706.64    0.00    Inf
## Lognormal   -180.65   365.3    369      6.32    0.28    0.76
## Weibull     -179.24   362.49   366.19    4.57    0.47    0.62
## F           -185.06   374.11   377.81   11.28    0.05    1.21
## Student     -250.63   503.26   505.11   66.40    0.00   14.69
##          H(AD) KS(value)      H(KS)
## Normal      rejected    0.35  rejected
## Cauchy      rejected    0.43  rejected
## Logistic    rejected    0.36  rejected
## Exponential  rejected    0.57  rejected
## Chi-square      NULL    0.44  rejected
## Uniform      NULL    0.37  rejected
## Lognormal     rejected    0.10 not rejected
## Weibull      not rejected    0.11 not rejected
## F            NULL    0.15 not rejected
## Student      NULL    0.50  rejected

##
## Chosen continuous distribution is: Log-normal (lnorm)
## Fitted parameters are:
## meanlog      sdlog
## 0.780739 5.175300

```

Distribución Log Normal Parámetros: Media y desviación estándar - Cálculo de la media y la desviación estándar

```
round(sd(comparacionexport2021$newDollarsMillion),2)
```

```
## [1] 857.74
```

```
round(mean(comparacionexport2021$newDollarsMillion),2)
```

```
## [1] 377.02
```

Las cantidades de millones dólares que generan anualmente las empresas exportadoras en el año 2021 siguen una distribución Log normal. Si la media es 377.02 y la desviación estándar es 857.74 ¿Cuál es la probabilidad que la cantidad de millones de dólares sea mayor a 400, pero menor a 800 ?

X : Número de millones de dólares generados por las empresas exportadoras en el año 2021

```
round(plnorm(800,meanlog= log(377.02),sdlog = log(857.74)) - plnorm(400,meanlog= log(377.02),sdlog = log(857.74)),2)
```

```
## [1] 0.04
```

RPTA: La probabilidad de que la cantidad de millones dólares sean mayor a 400, pero menor a 800 es 0.04

```
media <- 377.02
desviacion <- 857.74
```

```
# Definir la función de densidad de probabilidad (PDF) de la distribución lognormal
pdf_lognormal <- function(x) {
  plnorm(x, meanlog = log(media), sdlog = log(desviacion))
}
```

```
# Graficar la distribución lognormal
curve(pdf_lognormal, from = 0, to = 1500, xlab = "Millones de dólares", ylab = "Probabilidad", main = "Distribución Lognormal")
```


GR27: Distribución Lognormal de millones

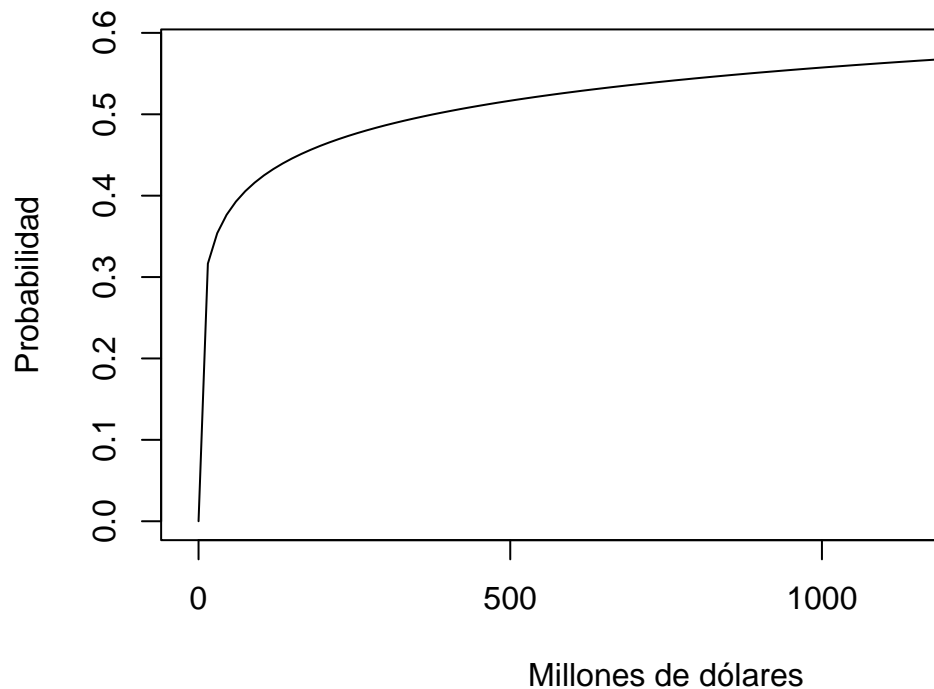


Grafico 27: Probabilidad Acumulada

Interpretación:

La gráfica representa la distribución lognormal de las cantidades de millones de dólares generadas anualmente por las empresas exportadoras en el año 2021. La distribución lognormal es una distribución de probabilidad continua que se caracteriza por tener una forma de campana asimétrica, donde la mayoría de los valores se concentran alrededor de la media y la dispersión de los valores aumenta a medida que nos alejamos de la media.

```
media <- 377.02
desviacion <- 857.74

# Definir la función de densidad de probabilidad (PDF) de la distribución lognormal
pdf_lognormal <- function(x) {
  dlnorm(x, meanlog = log(media), sdlog = log(desviacion))
}

# Graficar la distribución lognormal
curve(pdf_lognormal, from = 0, to = 1500, xlab = "Millones de dólares", ylab = "Densidad de probabilidad")
```

GR28: Distribución Lognormal de millones d

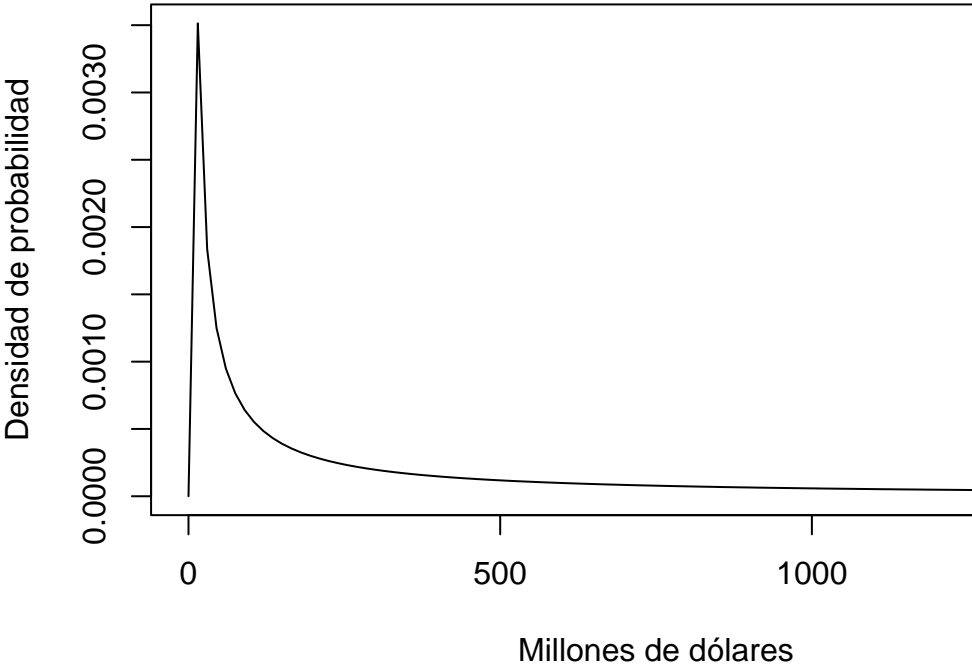


Grafico 28: Densidad Acumulada