*Tarea # 2*

Seleccione una base datos y realice un análisis exploratorio

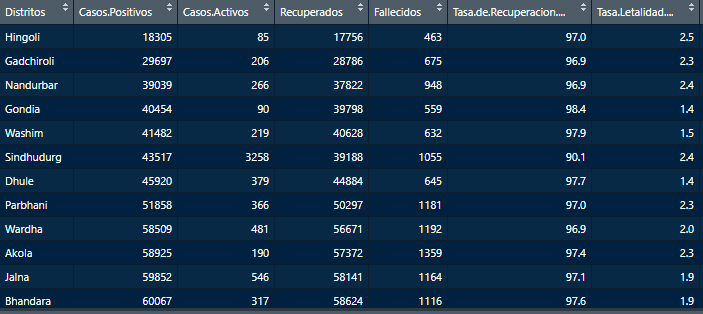
a. Use conexiones a SQL: Se utilizó "DB kaggle", se adjunta archivo

b. Utilice mínimo 4 funciones de dplyr para manipular datos:

\*\*Casos\_Covid\_19\_India= data.frame(Casos\_Covid\_19\_India)

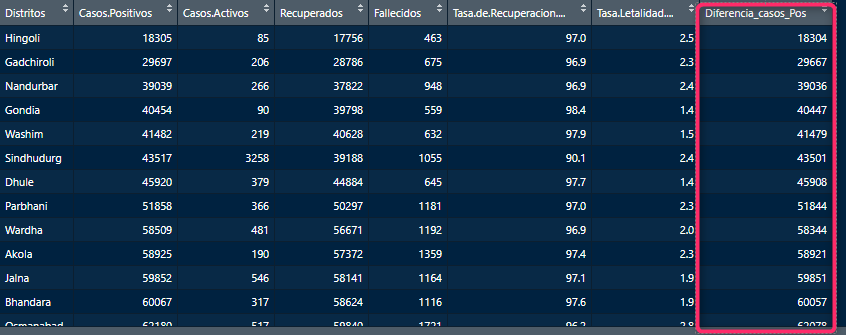
1.Casos\_Covid\_19\_India = arrange(Casos\_Covid\_19\_India, (Casos.Positivos))

View(Casos\_Covid\_19\_India)



2. Casos\_Covid\_19\_India = mutate (Casos\_Covid\_19\_India,

Diferencia\_casos\_Pos=(Casos.Activos+Recuperados+Fallecidos ))

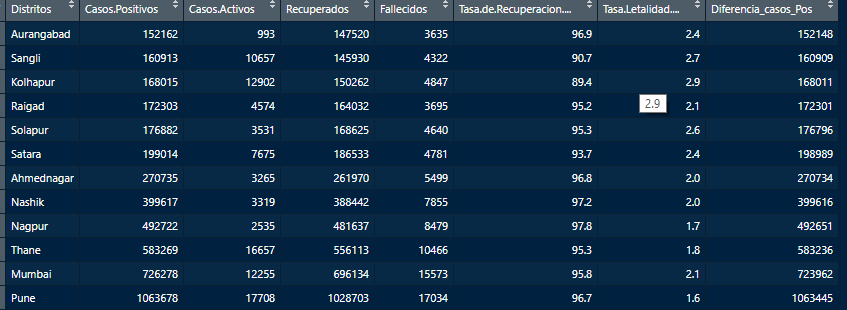


3. Orden\_Tasas\_Lel\_Rec=Casos\_Covid\_19\_India %>% select(Distritos,Casos.Positivos,Tasa.de.Recuperacion....,

Tasa. Letalidad....)%>%arrange(desc(Tasa.de.Recuperacion....))



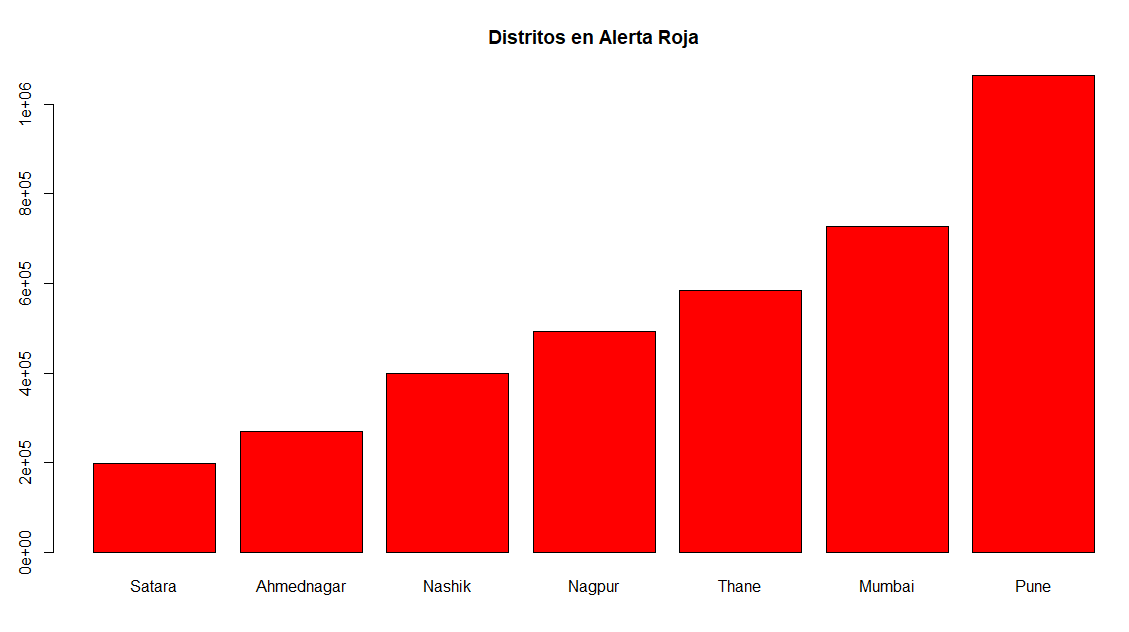
4. Distritos\_alerta\_Roja = filter(Casos\_Covid\_19\_India , Casos.Positivos > 100000)



C. Generar plots explicando el análisis que sacó de cada uno:

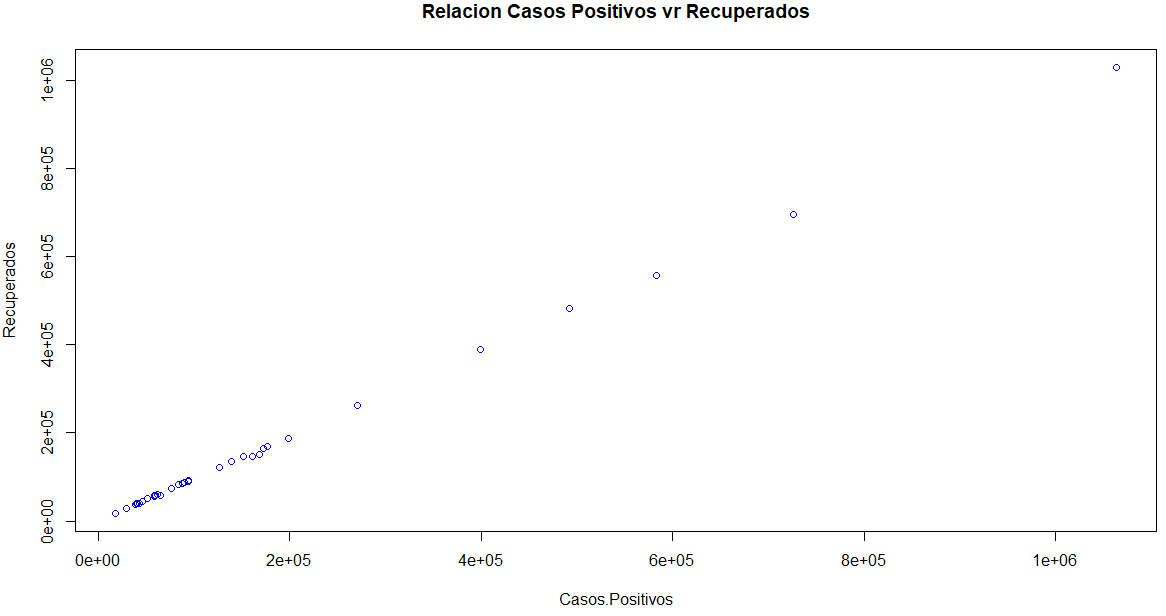
1. barplot (Distritos\_alerta\_Roja$Casos.Positivos, names.arg=as.factor(Distritos\_alerta\_Roja

$Distritos), las=2, col ="red”, main = "Distritos en Alerta Roja")



Análisis: En el gráfico anterior se puede visualizar los distritos que se encuentran en alerta rojo por el aumento de casos positivos, en donde se muestran los que están por encima de 200.000 mil casos y el que más llama la atención es “Pune”. Es interesante realizar otro análisis si esto se debe a la cantidad de población o que hay un foco importante de contagio en esa zona.

2. plot(select(Casos\_Covid\_19\_India,2,4), main="Relación Casos Positivos vr Recuperados”, col="blue")

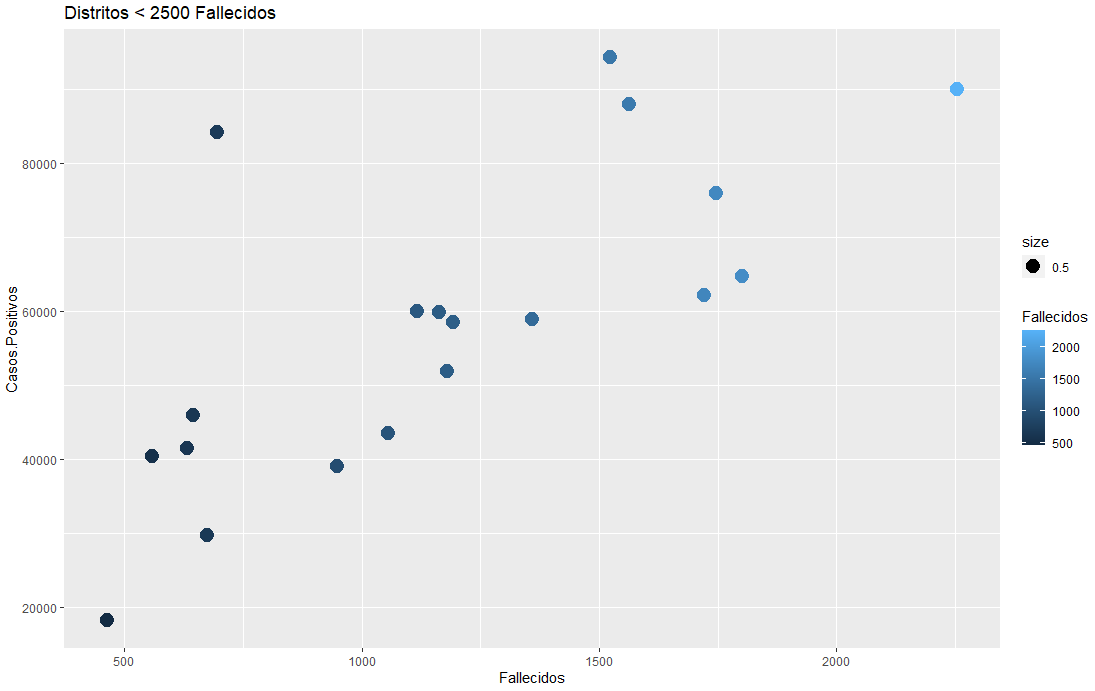


**Análisis:** Con el Grafico anterior se puede determinar que hay un comportamiento creciente según los casos positivos con su tasa de recuperación. Sin embargo, a menor cantidad de casos más rápido es la recuperación de las personas. Esto se puede deber a que el virus no se dispersa entre ellos ya que hay menos movilidad y contacto entre las personas.

Tabla=filter(Casos\_Covid\_19\_India,Casos\_Covid\_19\_India$Fallecidos<2500)

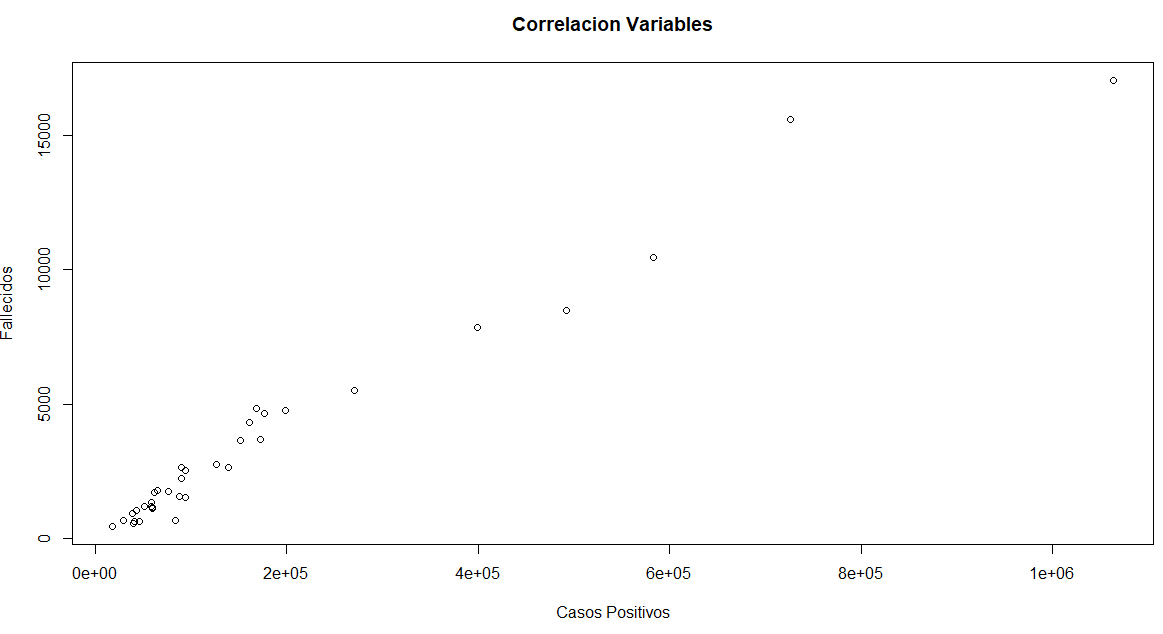
qplot (Fallecidos,Casos.Positivos,data =Tabla, color =Fallecidos,size=0.5,

main= "Distritos < 2500 Fallecidos")



D. Haga la búsqueda de correlación de variables y explíquelas (matriz)

plot(Casos\_Covid\_19\_India$Casos.Positivos, Casos\_Covid\_19\_India$Fallecidos, main="Correlacion Variables", xlab ="Casos Positivos", ylab="Fallecidos")



Análisis: En el Gráfico se muestra una correlación directa entre la proporción de los contagios de casos positivos y la cantidad de fallecidos, en los distritos con menos casos positivos se muestra que hay concentración, sin embargo, hay un comportamiento que se desvía en algunos casos.

F. Cree al menos 2 funciones.

1.

Diferencia\_Act\_y\_Rec= function (Casos.Positivos, Diferencia\_casos\_Pos) {Casos.Positivos-Diferencia\_casos\_Pos}

Diferencia\_Act\_y\_Rec(Casos\_Covid\_19\_India$Casos.Positivos,Casos\_Covid\_19\_India$Diferencia\_casos\_Pos)



2.

Tasa\_Caso\_Activos= function(Casos.Activos,Casos.Positivos){ Casos.Activos/Casos.Positivos\*100

}

Tasa\_Caso\_Activos (Casos\_Covid\_19\_India$Casos.Activos,Casos\_Covid\_19\_India$

Casos.Positivos)

