

# Project 2

Jorge Aranda

Friday, May 26

1)

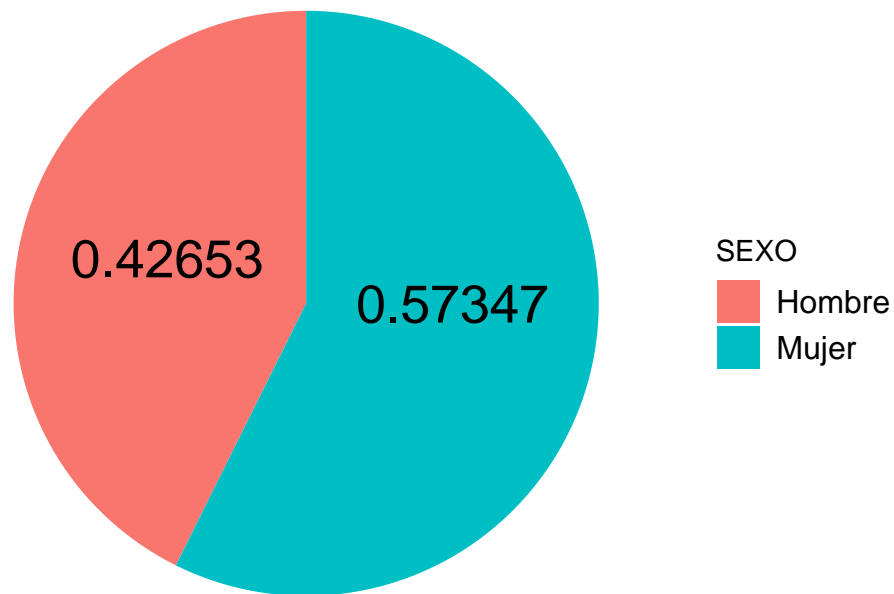
Compara las proporciones en cada sexo de individuos infectados

```
Sex_Total <- covid_19 %>% mutate(SEX0= case_when(
  .$SEX0==1 ~ "Mujer",
  .$SEX0==2 ~ "Hombre",
  .$SEX0==99 ~ "No Especificado")) %>% group_by(SEX0) %>% summarise(Total=n()) %>%
  pull(Total)
prop.test(Sex_Total,c(100000,100000))
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data: Sex_Total out of c(1e+05, 1e+05)
## X-squared = 4317.7, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.151285 -0.142595
## sample estimates:
## prop 1 prop 2
## 0.42653 0.57347
```

Con un 95% de confianza, hay evidencia significativa de que la proporción de mujeres infectadas por COVID-19 es mayor a la de hombres. Así, es factible que siendo mujer, haya una mayor probabilidad de contraer la enfermedad.

```
covid_19 %>% mutate(SEX0= case_when(
  .$SEX0==1 ~ "Mujer",
  .$SEX0==2 ~ "Hombre",
  .$SEX0==99 ~ "No Especificado")) %>% group_by(SEX0) %>% summarise(Total=n()) %>%
  cbind(data.frame(position=c(0.8,0.25))) %>%
  ggplot(aes(1,Total,fill=SEX0)) +
  geom_bar(stat="identity",position = "fill") +
  geom_text(aes(1,position,label=Total/100000),size=7) +
  coord_polar(theta="y") + theme_classic()+
  theme(axis.title = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        axis.line = element_blank(),
        legend.text = element_text(size=12))
```



2)

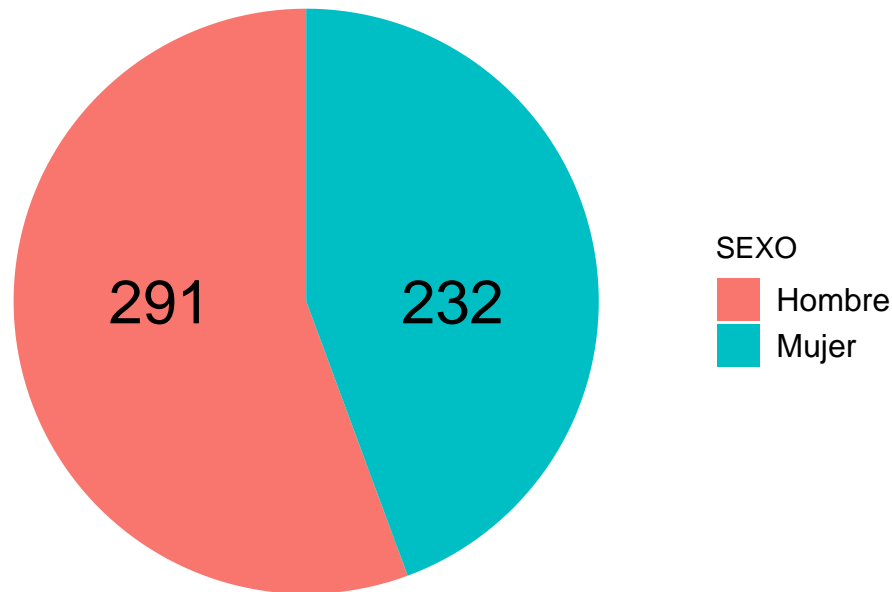
Compara las proporciones en cada sexo de individuos fallecidos .

```
Sex_muertes <- covid_19 %>% filter(!is.na(FECHA_DEF)) %>%
  mutate(SEXO= case_when(
    .$SEXO==1 ~ "Mujer",
    .$SEXO==2 ~ "Hombre",
    .$SEXO==99 ~ "No Especificado")) %>%
  group_by(SEXO) %>% summarise(Total=n())
muertes <- Sex_muertes %>% pull(Total)
prop.test(muertes,c(rep(sum(muertes),2)))
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data:  muertes out of c(rep(sum(muertes), 2))
## X-squared = 12.864, df = 1, p-value = 0.0003349
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.05068417 0.17493725
## sample estimates:
##      prop 1      prop 2
## 0.5564054 0.4435946
```

```
Sex_muertes %>% cbind(data.frame(position=c(0.75,0.25))) %>%
  ggplot(aes(1,Total,fill=SEXO)) + geom_bar(stat="identity",position = "fill") +
```

```
geom_text(aes(1,position,label=Total),size=8) + coord_polar(theta="y") +
theme_classic()+
theme(axis.title = element_blank(),
      axis.text = element_blank(),
      axis.ticks = element_blank(),
      axis.line = element_blank(),
      legend.text = element_text(size=12))
```

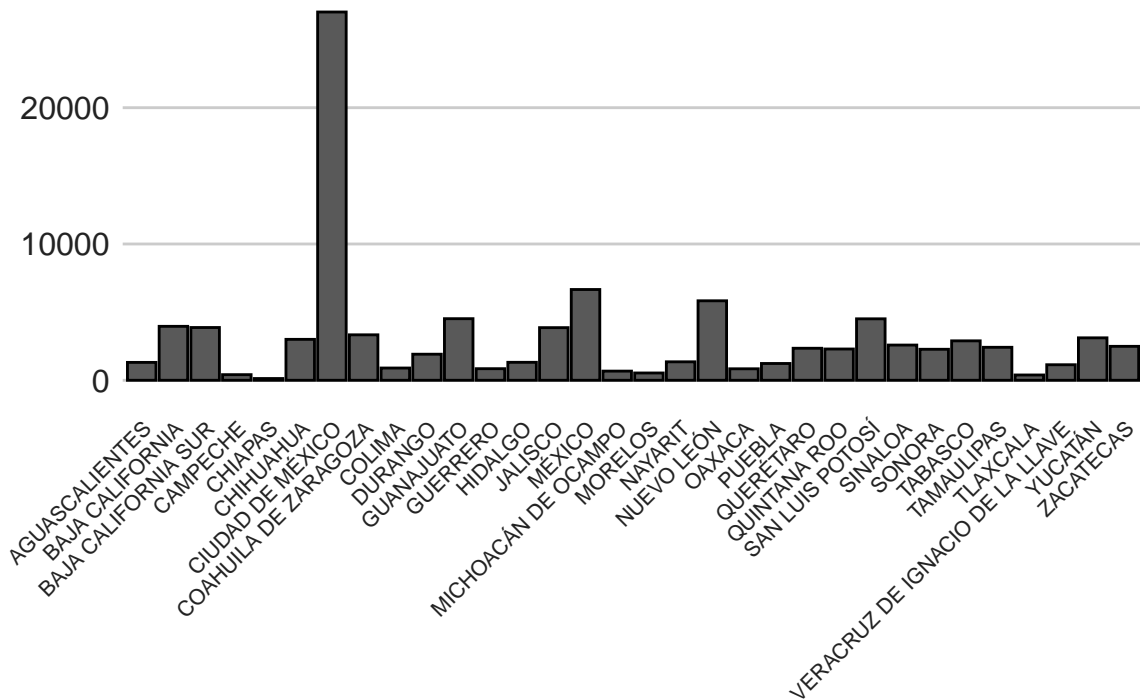


Con un 95% de confianza, hay evidencia significativa de que la proporción de fallecimientos en hombres por COVID-19 es mayor a la de mujeres. Así, es factible que siendo hombre, haya una mayor probabilidad de morir a causa de la infección.

3).

Compara las proporciones de número de casos por entidad (respecto a la población de cada entidad).

```
Catalogos<- loadWorkbook("C:/Users/jorge/Downloads/Catalogos_071020.xlsx")
Entidad_RED <- readWorksheet(Catalogos,sheet=9)
Entidad_RES <- as_tibble(Entidad_RED)
Entidades <- covid_19 %>%
  left_join(Entidad_RES,by=c('ENTIDAD_RES'='CLAVE_ENTIDAD')) %>%
  group_by(ENTIDAD_FEDERATIVA) %>% summarise(Total=n())
ggplot(Entidades,aes(ENTIDAD_FEDERATIVA,Total)) +
  geom_col(col="black") +
  bbc_style() +
  theme(axis.text.x = element_text(angle=45,hjust=1,size=8),
        axis.text.y = element_text(size=12))
```



```
chisq.test(Entidades$Total)
```

```
##
## Chi-squared test for given probabilities
##
## data: Entidades$Total
## X-squared = 214458, df = 31, p-value < 2.2e-16
```

Con un 95% de confianza, hay evidencia significativa de que la proporción de casos positivos es diferente para al menos una de las entidades.

4).

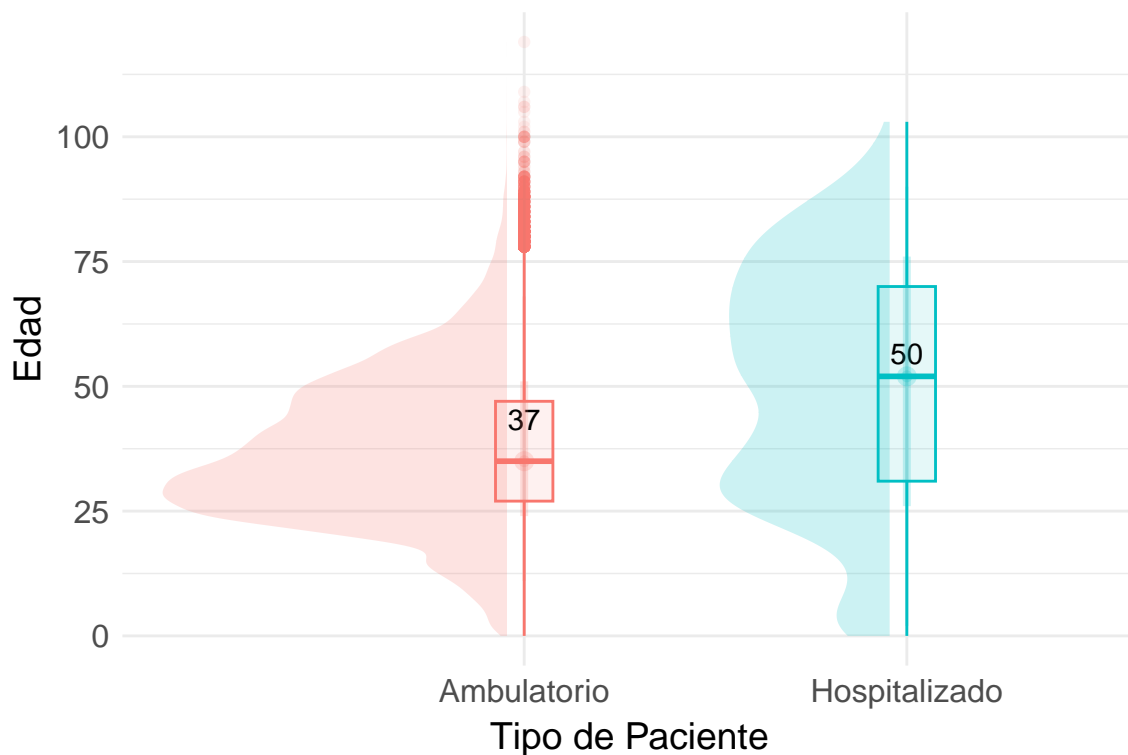
Compara la media de edad del grupo de casos ambulatorios con el grupo de casos hospitalizados

```
paciente <- covid_19 %>% mutate(TIPO_PACIENTE=case_when(
  .$TIPO_PACIENTE==1 ~ "Ambulatorio",
  .$TIPO_PACIENTE==2 ~ "Hospitalizado",
  .$TIPO_PACIENTE==99 ~ "No especificado"
)) %>% group_by(TIPO_PACIENTE) %>% reframe(Edad=EDAD)

Analisis <- paciente %>% mutate(ID=1:100000) %>%
  pivot_wider(names_from=TIPO_PACIENTE,values_from=Edad) %>% select(-ID)
Ambulatorio = Analisis %>% filter(!is.na(Ambulatorio)) %>% pull(Ambulatorio)
Hospitalizado = Analisis %>% filter(!is.na(Hospitalizado)) %>% pull(Hospitalizado)
t.test(Ambulatorio,Hospitalizado)
```

```
##
## Welch Two Sample t-test
##
## data: Ambulatorio and Hospitalizado
## t = -25.924, df = 2193.4, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -14.27477 -12.26701
## sample estimates:
## mean of x mean of y
## 37.00436 50.27525

text <- paciente %>% group_by(TIPO_PACIENTE) %>% summarise(mean=round(mean(Edad,na.rm=T)))
paciente %>%
  ggplot(aes(TIPO_PACIENTE,Edad,col=TIPO_PACIENTE,fill=TIPO_PACIENTE)) +
  geom_boxplot(alpha=0.1,width=0.15) +
  annotate("text",
    x=1:length(table(paciente$TIPO_PACIENTE)),
    y=aggregate(Edad ~ TIPO_PACIENTE, paciente,mean)[,2],
    label=round(aggregate(Edad ~ TIPO_PACIENTE, paciente,mean)[,2],digits = 0),
    vjust=-1)+
  stat_halfeye(side="left",alpha=0.2, justification=1.05) +
  theme_minimal() + xlab("Tipo de Paciente") + ylab("Edad") +
  theme(axis.title.x=element_text(size=14,vjust=.2),
    axis.title.y=element_text(size=14),
    axis.text = element_text(size=12),
    legend.position = "none")
```



```
aggregate(Edad ~ TIPO_PACIENTE, paciente, mean)
```

```
##   TIPO_PACIENTE   Edad
## 1 Ambulatorio 37.00436
## 2 Hospitalizado 50.27525
```

Con un 95% de confianza, hay evidencia significativa de que el numero de casos ambulatorios es mayor al de casos que acabaron hospitalizados por covid-19

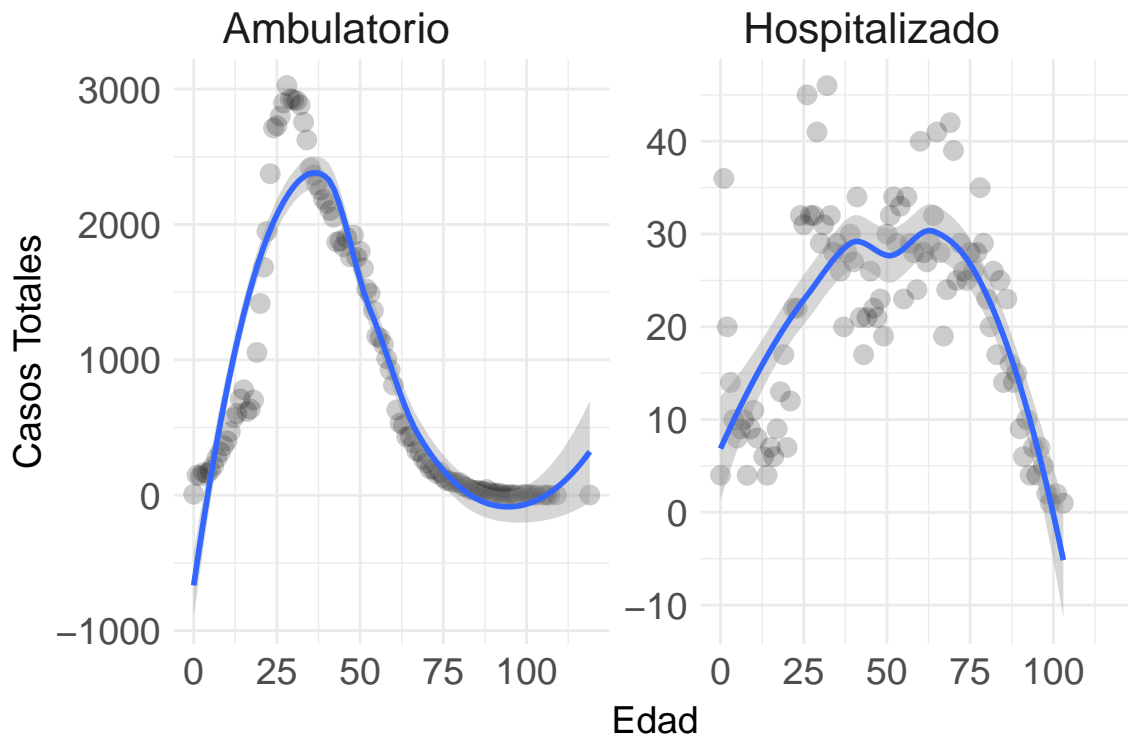
5).

Verifica la relación de dependencia entre el “TIPO\_PACIENTE” y el grupo de edad (con grupos de 10 años, iniciando con 0-9 años)

```
covid_19_edades <- covid_19 %>%
  mutate(EDAD= case_when(
    .$EDAD<10 ~ "0 a 9 años",
    .$EDAD>=10 & .$EDAD<20 ~ "10 a 19 años",
    .$EDAD>=20 & .$EDAD<30 ~ "20 a 29 años",
    .$EDAD>=30 & .$EDAD<40 ~ "30 a 39 años",
    .$EDAD>=40 & .$EDAD<50 ~ "40 a 49 años",
    .$EDAD>=50 & .$EDAD<60 ~ "50 a 59 años",
    .$EDAD>=60 & .$EDAD<70 ~ "60 a 69 años",
    .$EDAD>=70 & .$EDAD<80 ~ "70 a 79 años",
    .$EDAD>=80 & .$EDAD<90 ~ "80 a 89 años",
    .$EDAD>=90 & .$EDAD<100 ~ "mas de 90 años",
    !is.na(.$EDAD) ~ "No Especificado"))

relation <- covid_19_edades %>% mutate(TIPO_PACIENTE=case_when(
  .$TIPO_PACIENTE==1 ~ "Ambulatorio",
  .$TIPO_PACIENTE==2 ~ "Hospitalizado",
  .$TIPO_PACIENTE==99 ~ "No especificado"
))

table1<- relation %>%
  group_by(TIPO_PACIENTE,EDAD) %>% summarise(Total=n()) %>%
  pivot_wider(names_from=EDAD,values_from = Total) %>% select(-`No Especificado`)
covid_19 %>% mutate(TIPO_PACIENTE=case_when(
  .$TIPO_PACIENTE==1 ~ "Ambulatorio",
  .$TIPO_PACIENTE==2 ~ "Hospitalizado",
  .$TIPO_PACIENTE==99 ~ "No especificado"
)) %>%
  group_by(TIPO_PACIENTE,EDAD) %>% summarise(Total=n()) %>%
  ggplot(aes(EDAD,Total)) + geom_point(size=3,alpha=0.2) +
  geom_smooth() + facet_wrap(.~TIPO_PACIENTE,scales="free_y")+
  theme_minimal() + xlab("Edad") + ylab("Casos Totales") +
  theme(axis.title.x=element_text(size=14,vjust=0.2),
        axis.title.y=element_text(size=14),
        axis.text = element_text(size=14),
        strip.text = element_text(size=16,hjust=0.2))
```



```
matrix <- as.matrix(table1[,-1])
matrix
```

```
##      0 a 9 años 10 a 19 años 20 a 29 años 30 a 39 años 40 a 49 años
## [1,]      1982       6552      24518      25619      19240
## [2,]       124         81        276        299        231
##      50 a 59 años 60 a 69 años 70 a 79 años 80 a 89 años mas de 90 años
## [1,]      13242       4657       1489        463         63
## [2,]       296        310        290        193         55
```

```
chisq.test(matrix)
```

```
## Warning in chisq.test(matrix): Chi-squared approximation may be incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  matrix
## X-squared = 6003.6, df = 9, p-value < 2.2e-16
```

Existe, con un 95% de confianza, una dependencia entre la edad del paciente y si sera tratado de forma ambulatoria y requerira hospitalización.

6).

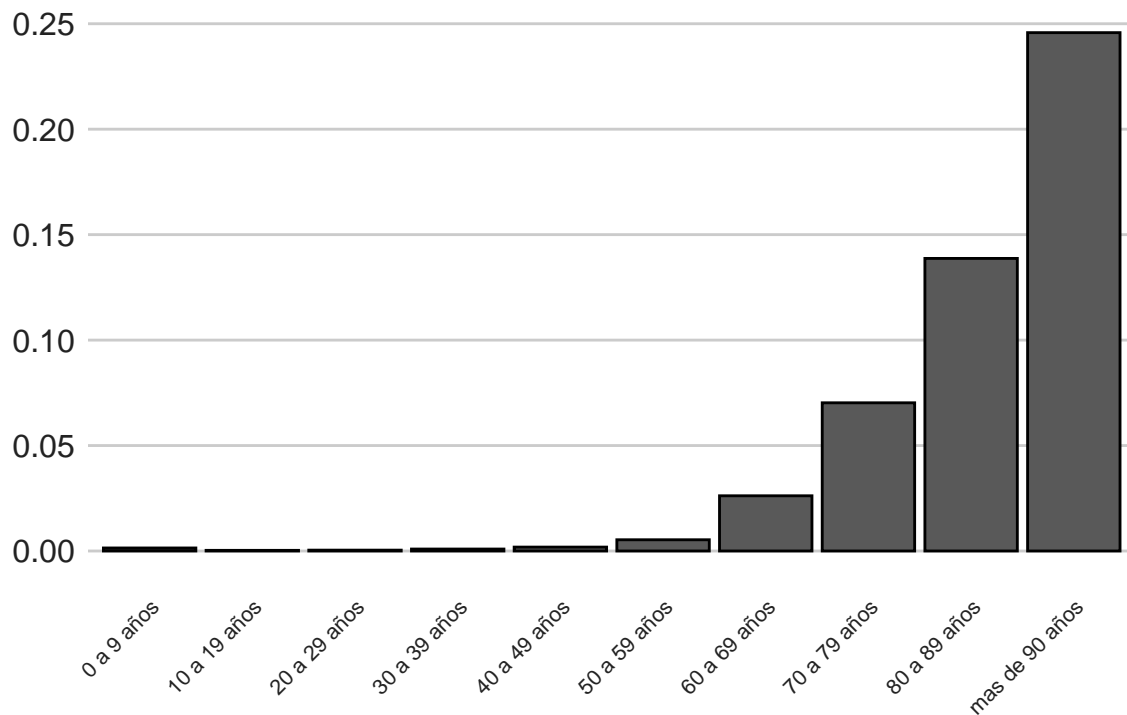
Compara la tasa de letalidad (i.e. la probabilidad de muerte por infección) en cada grupo de edad (con grupos de 10 años, iniciando con 0-9 años).

```

muertes_por_grupo <- covid_19_edades %>%
  filter(!is.na(FECHA_DEF)) %>% group_by(EDAD) %>%
  summarise(Total=n())
Letalidad <- covid_19_edades %>%
  group_by(EDAD) %>% summarise(Total=n()) %>%
  inner_join(muertes_por_grupo,by="EDAD",suffix=c("_casos","_muertes")) %>%
  mutate(Ind_letalidad=Total_muertes/Total_casos) %>%
  filter(EDAD != "No Especificado")

ggplot(Letalidad,aes(EDAD,Ind_letalidad)) +
  geom_col(col="black") +
  bbc_style() +
  theme(axis.text.x = element_text(angle=45,hjust=1,size=8),
        axis.text.y = element_text(size=12))

```



```
prop.test(Letalidad$Ind_letalidad,Letalidad$Total_casos)
```

```

## Warning in prop.test(Letalidad$Ind_letalidad, Letalidad$Total_casos):
## Chi-squared approximation may be incorrect

##
## 10-sample test for equality of proportions without continuity
## correction
##
## data:  Letalidad$Ind_letalidad out of Letalidad$Total_casos
## X-squared = 110.27, df = 9, p-value < 2.2e-16
## alternative hypothesis: two.sided

```



```
## sample estimates:
##      prop 1      prop 2      prop 3      prop 4      prop 5      prop 6
## 6.764014e-07 4.545797e-08 1.626698e-08 3.572796e-08 9.495678e-08 3.928470e-07
##      prop 7      prop 8      prop 9      prop 10
## 5.269326e-06 3.949645e-05 2.114627e-04 2.082735e-03
```

Con un 95% de confianza, hay evidencia suficiente que indica diferencias significativas en el índice de letalidad de los pacientes en cada grupo de edad

7).

Verifica la relación de dependencia entre el “TIPO\_PACIENTE” y el número de co-morbilidades (crea 3 grupos: 0, 1 y 2 o más comorbilidades)

```
tabla <- covid_19 %>%
  select(TIPO_PACIENTE,DIABETES,EPOC,ASMA,INMUSUPR,HIPERTENSION,OTRA_COM) %>%
  mutate(across(.cols=c(DIABETES,EPOC,ASMA,HIPERTENSION,INMUSUPR,OTRA_COM),
    .fns= \(x) ifelse(x %in% c(98,2),0,1)),
    TIPO_PACIENTE=case_when(
      .$TIPO_PACIENTE==1 ~ "Ambulatorio",
      .$TIPO_PACIENTE==2 ~ "Hospitalizado",
      .$TIPO_PACIENTE==99 ~ "No especificado"),
    Comorbilidades=rowSums(across(where(is.numeric))),
    Comorbilidades= ifelse(Comorbilidades >=2,2,Comorbilidades)) %>%
  group_by(TIPO_PACIENTE,Comorbilidades) %>%
  summarize(Total=n())
Totales <- tabla %>% pull(Total)
matriz <- matrix(Totales, byrow=T, nrow=2,
  dimnames=list(c("Ambulatorio","Hospitalizado"),
    c("0","1","2 o mas")))
names(dimnames(matriz)) <- list(c("Tipo de Paciente"),
  c("Numero de comorbilidades"))
matriz
```

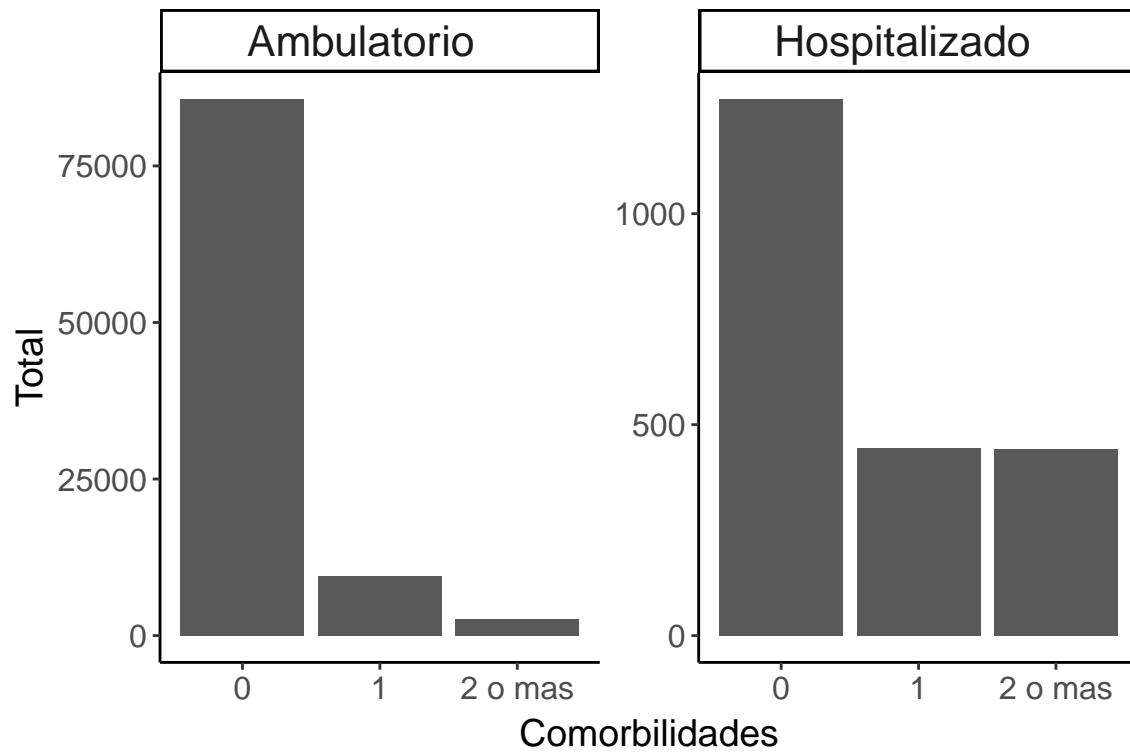
```
##              Numero de comorbilidades
## Tipo de Paciente      0      1 2 o mas
##   Ambulatorio   85640 9577    2625
##   Hospitalizado  1271  445    442
```

```
chisq.test(matriz)
```

```
##
## Pearson's Chi-squared test
##
## data:  matriz
## X-squared = 2627.4, df = 2, p-value < 2.2e-16
```

```
tabla %>% mutate(Comorbilidades=as.character(Comorbilidades)) %>%
  mutate(Comorbilidades=ifelse(Comorbilidades == 2,"2 o mas",Comorbilidades)) %>%
  ggplot(aes(Comorbilidades>Total)) +
  geom_col(position = "dodge") + facet_wrap(.~TIPO_PACIENTE,
```

```
scales = "free_y") +
theme_classic() + theme(axis.title.x=element_text(size=14,vjust=-2.8),
axis.title.y=element_text(size=14),
axis.title.x.bottom = element_text(vjust=.2),
axis.text = element_text(size=12),
strip.text = element_text(size=16,hjust=0.4))
```



Existe una dependencia entre el tipo de paciente y el numero de comorbilidades que este presente.