

# Informe Final Querétaro

Por Espinoza Carlos y Herrejon Kasandra

2025-11-11

## 00 Preprocesamiento

Este documento busca mediante un análisis el conocer cómo ha ido cambiando la esperanza de vida en la población correspondiente al Estado de Querétaro.

Durante el periodo comprendido entre 2010 y 2024, el estado de Querétaro experimentó una transformación profunda en su dinámica poblacional, impulsada por cambios socioeconómicos, migratorios y epidemiológicos. En 2010, el estado contaba con una población cercana a los 1.8 millones de habitantes, caracterizada por una estructura demográfica predominantemente joven y en crecimiento. A lo largo de la década, Querétaro se consolidó como uno de los polos industriales más importantes del país, atrayendo inversión nacional y extranjera, lo que reforzó su atractivo migratorio y aceleró su dinámica demográfica. Para 2024, la población se estima por encima de los 2.4 millones, reflejando un crecimiento continuo derivado tanto del aumento natural como, especialmente, del proceso de migración interna.

Uno de los rasgos más distintivos del periodo fue el fuerte flujo migratorio hacia el estado. Querétaro se consolidó como destino de trabajadores de entidades vecinas como Guanajuato, Estado de México, Hidalgo y Michoacán, así como de jóvenes profesionistas atraídos por los sectores automotriz, aeronáutico y de servicios. Este fenómeno produjo cambios visibles en la estructura poblacional: aumento en los grupos de edad económicamente activa, crecimiento acelerado de zonas urbanas como El Marqués, Corregidora y Querétaro capital, y presión sobre infraestructura urbana, vivienda y servicios públicos. A la par, la inmigración internacional, aunque menor, también contribuyó al incremento poblacional, especialmente proveniente de Sudamérica y Centroamérica entre 2016 y 2023.

En cuanto a mortalidad, el estado mostró tres etapas muy distintas. Entre 2010 y 2019, la tendencia general fue de mortalidad relativamente estable, con tasas moderadas y un perfil de causas dominado por enfermedades crónicas como diabetes mellitus, hipertensión arterial, cardiopatías y ciertos tipos de cáncer. En 2010, la epidemia de influenza H1N1 elevó la mortalidad en edades jóvenes, especialmente entre 5 y 20 años debido a la exposición escolar y la circulación del virus en espacios cerrados. Sin embargo, después de 2011, la mortalidad volvió a niveles regulares.

La ruptura más significativa ocurrió en 2020 y 2021 con la llegada del SARS-CoV-2 (COVID-19). En 2021, Querétaro registró la mayor mortalidad de su historia contemporánea: más de 17 mil decesos, con predominio masculino. Los grupos más afectados fueron los adultos mayores, especialmente aquellos mayores de 70 años, pero también se observaron incrementos en edades laborales (30 a 59 años), especialmente entre trabajadores expuestos en actividades presenciales. La pandemia tuvo repercusiones sociales directas, reduciendo temporalmente la esperanza de vida, alterando la estructura por edades y dejando secuelas económicas en miles de hogares.

A partir de 2022, la mortalidad regresó paulatinamente a niveles previos, aunque con una distribución etaria más envejecida. Esto se vincula con la transición demográfica del estado, ya que Querétaro muestra un proceso claro de envejecimiento paulatino: disminución relativa de nacimientos, incremento del adulto mayor y crecimiento de la esperanza de vida conforme avance la década. Entre 2010 y 2024, aunque la base de la pirámide poblacional sigue siendo amplia, se observa una tendencia hacia la contracción, mientras que los grupos mayores de 60 años incrementan su peso relativo.

Finalmente, el periodo también estuvo marcado por cambios sociales asociados al desarrollo económico del estado. El incremento del nivel educativo, la incorporación masiva de mujeres al mercado laboral y la urbanización acelerada modificaron la composición de hogares, patrones de fecundidad y dinámicas de movilidad interna. De esta manera, el Querétaro de 2024 es un estado con una economía más robusta, una población más diversa, mayor presión urbana y un perfil demográfico que avanza claramente hacia etapas más avanzadas de la transición demográfica.

## 01 Diagrama de Flujo

## 03 Algoritmos

Crecimiento exponencial

$$P_t = P_0 e^{rt}$$

**Tablas de vida**

Probabilidad de morir

$$q_x = \frac{m_x}{1 + (1 - a_x)m_x}$$

Probabilidad de sobrevivir

$$p_x = 1 - q_x$$

Supervivencia

$$l_{x+n} = l_x p_x$$

Número de Muertes

$$d_x = l_x - l_{x+n}$$

APV

$$L_x = l_{x+n} + a_x d_x$$

Descomposición por método de Arriga

$$C_x = C_x^{(D)} + C_x^{(I)}$$

$$e_0^{(x)} - e_0^{(x)} = \sum_{y=x} C_x$$

Causa eliminada

$$l_{x+n}^{(-i)} = l_x^{(-i)} (1 - q_x^{(-i)})$$

Total acumulado APV

$$T_x = \sum_{y=x}^{\omega} L_y$$

Esperanza de vida

$$e_x = \frac{T_x}{l_x}$$

Tasa de mortalidad

$$m_x = \frac{D_x}{N_x}$$

Suavizamiento

$$MA_t = \frac{X_t + X_{t-1} + \dots + X_{t-k+1}}{k}$$

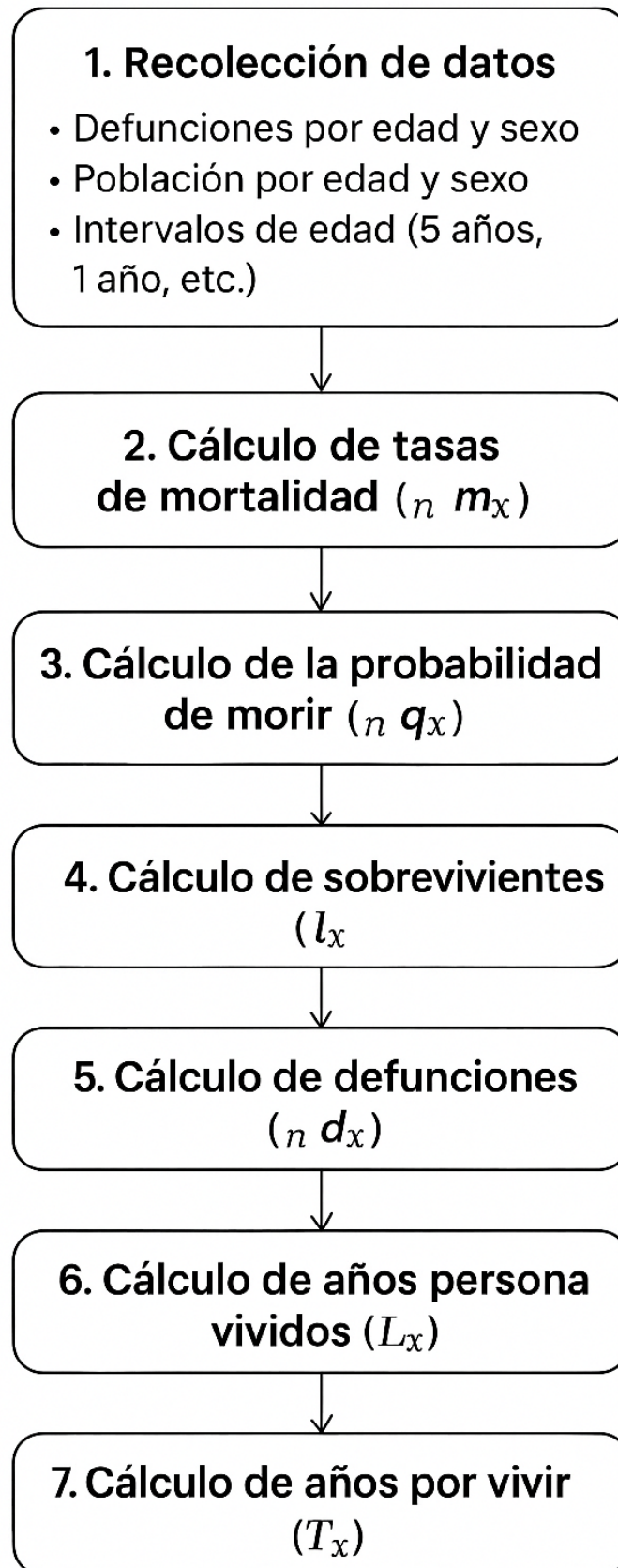


Figure 1: Diagrama de Flujo sobre obtención de datos

## 04 Códigos

En primera instancia, debemos señalar que el realizar la recolección de datos en “00\_preprocesamiento”, fue a través de la pagina del INEGI, tanto para 2010, como para 2020, tomando los datos de población total, sexo, edad, Entidad y Municipio, donde una vez obtenidos los en formato de excel, nos enfocamos en unicamente los datos de Querétaro, una vez que se han obtenido los datos filtramos por entidad y borramos el resto de información que no es relevante, mismo caso para las defunciones, tomamos los datos generales y nos quedamos unicamente con los de Querétaro.

En este apartado nos enfocaremos en los algoritmos que utilizamos para obtener los APV, las defunciones asi como las tablas de vida. Tablas de población.

```
names(c2010) <- c("age", "tot", "male", "female")
setDT(c2010)

c2010 <- c2010[-1, ]

c2010[, age := gsub("De ", "", age)]
c2010[, age := substr(age, 1, 2)]
c2010[age=="No", age:=NA]
c2010[, age:=as.numeric(age)]

c2010[, tot := as.numeric(gsub(",", "", tot))]
c2010[, male := as.numeric(gsub(",", "", male))]
c2010[, female := as.numeric(gsub(",", "", female))]

c2010 <- c2010[, age := ifelse(age %in% 1:4, 1, age)] %>%
  .[, .(tot = sum(tot),
        male = sum(male),
        female = sum(female)), .(age)]

c2010 <- melt.data.table(c2010,
                        id.vars = "age",
                        measure.vars = c("male", "female"),
                        variable.name = "sex",
                        value.name = "pop")

sum(c2010$pop)
c2010[, year:=2010]

# Preprocesamiento censo 2020 ----
names(c2020) <- c("age", "tot", "male", "female")
setDT(c2020)

c2020 <- c2020[-1, ]

c2020[, age := gsub("De ", "", age)]
c2020[, age := substr(age, 1, 2)]
c2020[age=="No", age:=NA]
c2020[, age:=as.numeric(age)]

c2020[, tot := as.numeric(gsub(",", "", tot))]
```

```

c2020[ , male := as.numeric(gsub(",", "", male))]
c2020[ , female := as.numeric(gsub(",", "", female))]

c2020 <- c2020[ , age := ifelse(age %in% 1:4, 1, age)] %>%
  .[ , .(tot = sum(tot),
    male = sum(male),
    female = sum(female)), .(age)]

c2020 <- melt.data.table(c2020,
  id.vars = "age",
  measure.vars = c("male", "female"),
  variable.name = "sex",
  value.name = "pop")

sum(c2020$pop)
c2020[ , year:=2020]

# Unir los dos censos

censos <- rbind(c2010, c2020)

# Prorrrateo de los valores perdidos (missing)

censos_pro <- censos[ !is.na(age) ] %>%
  .[ , p_pop := pop / sum(pop), .(year, sex)] %>%
  merge( censos[ is.na(age),
    .(sex, year, na_pop=pop)],
    by = c("sex", "year")) %>%
  .[ , pop_adj := pop + na_pop * p_pop] %>%
  .[ , .(year, sex, age, pop = pop_adj) ]

# Comprobacion de prorrrateo
censos_pro[ , sum(pop), .(year, sex)]
censos[ , sum(pop), .(year, sex)]

## Guardar tabla de censos ----
write.csv(censos_pro, "data/censos_pro.csv")

# -----

```

Años Persona Vivos

```

# Cálculo de APV 2010 ----
# Cálculo de años persona vividos (población a mitad de año)
N <- expo(censos_pro[year==2010] %>% .$pop,
  censos_pro[year==2020] %>% .$pop,
  t_0 = "2010-06-25", t_T = "2020-03-15", t = 2010.5)

apv2010 <- censos_pro[year==2010, .(age, sex, N)]

```

```

apv2010[, year:=2010]

ggplot(apv2010, aes(x = factor(age), y = ifelse(sex == "male", -N/1e6, N/1e6), fill = sex)) +
  geom_col(width = 0.7, alpha = 0.8) +
  coord_flip() +
  scale_y_continuous(
    labels = function(x) paste0(abs(x), "M"),
    breaks = scales::pretty_breaks(n = 8)
  ) +
  scale_fill_manual(
    values = c("male" = "#1f77b4", "female" = "#d62728"),
    labels = c("male" = "Hombres", "female" = "Mujeres")
  ) +
  labs(
    title = "Pirámide Poblacional 2010",
    subtitle = "Distribución por edad y sexo",
    x = "Grupo de edad",
    y = "Población mitad de año (millones)",
    fill = "Sexo",
    caption = "Fuente: INEGI"
  ) +
  theme_minimal() +
  theme(
    legend.position = "bottom",
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    plot.subtitle = element_text(hjust = 0.5, color = "gray40"),
    axis.text.y = element_text(size = 8),
    panel.grid.major.y = element_blank()
  )

# Cálculo de APV 2019 ----
# Cálculo de años persona vividos (población a mitad de año)
N <- expo(censos_pro[year==2010] %>% .$pop,
  censos_pro[year==2020] %>% .$pop,
  t_0 = "2010-06-25", t_T = "2020-03-15", t = 2019.5)

apv2019 <- censos_pro[year==2020, .(age, sex, N)]
apv2019[, year:=2019]

# Cálculo de APV 2020 ----
# Cálculo de años persona vividos (población a mitad de año)
N <- expo(censos_pro[year==2010] %>% .$pop,
  censos_pro[year==2020] %>% .$pop,
  t_0 = "2010-06-25", t_T = "2020-03-15", t = 2020.5)

apv2020 <- censos_pro[year==2020, .(age, sex, N)]
apv2020[, year:=2020]

ggplot(apv2020, aes(x = factor(age), y = ifelse(sex == "male", -N/1e6, N/1e6), fill = sex)) +
  geom_col(width = 0.7, alpha = 0.8) +
  coord_flip() +

```

```

scale_y_continuous(
  labels = function(x) paste0(abs(x), "M"),
  breaks = scales::pretty_breaks(n = 8)
) +
scale_fill_manual(
  values = c("male" = "#1f77b4", "female" = "#d62728"),
  labels = c("male" = "Hombres", "female" = "Mujeres")
) +
labs(
  title = "Pirámide Poblacional 2020",
  subtitle = "Distribución por edad y sexo",
  x = "Grupo de edad",
  y = "Población mitad de año (millones)",
  fill = "Sexo",
  caption = "Fuente: INEGI"
) +
theme_minimal() +
theme(
  legend.position = "bottom",
  plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
  plot.subtitle = element_text(hjust = 0.5, color = "gray40"),
  axis.text.y = element_text(size = 8),
  panel.grid.major.y = element_blank()
)

# Cálculo de APV 2021 ----
# Cálculo de años persona vividos (población a mitad de año)
N <- expo(censos_pro[year==2010] %>% .$pop,
          censos_pro[year==2020] %>% .$pop,
          t_0 = "2010-06-25", t_T = "2020-03-15", t = 2021.5)

apv2021 <- censos_pro[year==2020, .(age, sex, N)]
apv2021[, year:=2021]

```

Defunciones.

```

# Defunciones 1990-2024 Preprocesamiento----
def <- read_xlsx("Data/INEGI_deaths.xls", sheet = 1,
               range = "A6:G20560")

names(def) <- c("age", "year", "reg",
               "tot", "male", "female", "ns")
setDT(def)

## Filtro ----
def <- def[age!="Total" & year!="Total" & year>=1990]
def[, .N, .(age)] # Como viene la variable edad

## Limpieza de la edad ----
def[, age := gsub("Menores de ", "", age)]
def[, age := substr(age, 1, 2)]
def[age=="1 ", age:=0]

```

```

def[age=="1-", age:=1]
def[age=="5-", age:=5]
def[age=="No", age:=NA] # prorrateo
def[ , age:=as.numeric(age)]

## Convertir a numero las defunciones ----
def[ , tot := as.numeric(gsub(",", "", tot))]
def[ , male := as.numeric(gsub(",", "", male))]
def[ , female := as.numeric(gsub(",", "", female))]
def[ , ns := as.numeric(gsub(",", "", ns))]

# Tabla de defunciones - comprobación con Excel
def_comp <- def[ , .(tot=sum(tot, na.rm = T),
                        male=sum(male, na.rm = T),
                        female=sum(female, na.rm = T),
                        ns=sum(ns, na.rm = T)),
                .(year)]

# Imputación
def[year=="No especificado", year:=reg]
def[ , year:=as.numeric(year)]
def_comp[ , sum(tot)]

# Tabla final de defunciones
def_pro <- def[ , .(male=sum(male, na.rm = T),
                        female=sum(female, na.rm = T),
                        ns=sum(ns, na.rm = T)),
                .(year, age)]

# Prorrateo de los valores perdidos (missing)
### Sexo ----
def_pro[ , tot:=male+female][ , `:=`(p_male=male/tot, p_female=female/tot)]
def_pro[ , `:=`(male_adj=male+p_male*ns, female_adj=female+p_female*ns)]
def_pro <- def_pro[ , .(year, age, male=male_adj, female=female_adj)]
sum(def_pro$male)+sum(def_pro$female)

#Formato long
def_pro <- melt.data.table(def_pro,
                           id.vars = c("year", "age"),
                           measure.vars = c("male", "female"),
                           variable.name = "sex",
                           value.name = "deaths")

# Extra, formato wide
dcast(def_pro, formula = year +age ~ sex)

```



```

sum(def_pro$deaths)
def_pro[, sum(deaths), .(year, sex)]

### Edad ----
def_pro <- def_pro[ !is.na(age) ] %>%
  .[, p_deaths := deaths / sum(deaths), .(year, sex)] %>%
  merge( def_pro[ is.na(age),
                .(sex, year, na_deaths=deaths)],
        by = c("sex", "year")) %>%
  .[, deaths_adj := deaths + na_deaths * p_deaths] %>%
  .[, .(year, sex, age, deaths = deaths_adj) ]

```

Descomposición por metodo de arriaga

```

lt <- fread("data/lt_mex.csv")

# Confirmar columnas requeridas
if(!all(c("year","age","mx") %in% names(lt))){
  stop("El archivo debe tener columnas: year, age, mx")
}

# Asegurar orden por edad
setorder(lt, year, age)

# Calculo de funciones

calc_life_table <- function(mx){

  qx <- mx / (1 + 0.5*mx)
  qx[qx > 1] <- 1

  lx <- numeric(length(mx)); Lx <- lx; dx <- lx; Tx <- lx
  lx[1] <- 100000

  for(i in seq_along(mx)){
    dx[i] <- lx[i] * qx[i]
    Lx[i] <- lx[i] - 0.5 * dx[i]
    if(i < length(mx)) lx[i+1] <- lx[i] - dx[i]
  }

  Tx[length(mx)] <- Lx[length(mx)]
  for(i in (length(mx)-1):1){
    Tx[i] <- Tx[i+1] + Lx[i]
  }

  ex <- Tx / lx
  return(list(ex=ex, lx=lx, Lx=Lx))

# Evaluación por año
years_to_eval <- c(2010, 2019, 2021)
lt <- lt[year %in% years_to_eval]

# Calcular tablas de vida por año

```

```

lt_full <- lt[, {
  tb <- calc_life_table(mx)
  .(age = age, mx = mx, lx = tb$lx, Lx = tb$Lx, ex = tb$ex)
}, by = year]

# Comparación de datos
arriaga_decomp <- function(yr1, yr2){
  d1 <- lt_full[year == yr1]
  d2 <- lt_full[year == yr2]

  # Fórmula clásica Arriaga
  Cx <- (d2$mx - d1$mx) * (d1$Lx / d1$lx[1])

  contrib <- 0.5 * Cx

  data.frame(
    age = d1$age,
    contrib = contrib,
    period = paste0(yr1, "→", yr2)
  )
}

# 2010 vs 2019 y 2019 vs 2021
A_10_19 <- arriaga_decomp(2010, 2019)
A_19_21 <- arriaga_decomp(2019, 2021)

# Juntar
A <- rbind(A_10_19, A_19_21)

# Ordenar de mayor contribución a menor
A_sorted <- A %>% arrange(period, desc(contrib))

```

## 05 Esperanza de vida

Tablas de Vida.

!-Tablas de vida

## 07 Análisis sobre 2021

La comparación de las esperanzas de vida entre 2019 y 2021 utilizando el método de Arriaga permite identificar la contribución de cada grupo de edad al cambio total observado, distinguiendo entre efectos directos e indirectos en la mortalidad. En 2019, Querétaro presentaba una estructura de mortalidad relativamente estable, con predominio de enfermedades crónicas y un patrón de riesgo concentrado en adultos mayores. Sin embargo, en 2021, en plena segunda ola de COVID-19, se registró un incremento abrupto de la mortalidad total, generando una caída significativa en la esperanza de vida al nacer. El método de Arriaga permite cuantificar este impacto y revela que la mayor parte de la reducción provino de los grupos en edades avanzadas, aunque también se observaron contribuciones importantes en edades laborales.

En términos directos, los grupos de edad mayores de 60, 70 y 80 años aportaron la contribución negativa más alta a la diferencia en esperanza de vida. Esto se debe a que el COVID-19 afectó de manera desproporcionada

a los adultos mayores, elevando notablemente sus tasas de mortalidad específicas, lo que redujo los valores de  $L$  y  $T$  entre 2021 y 2019. De acuerdo con la descomposición, la contribución directa de estos grupos es la que mayor peso tiene en la disminución total, reflejando el aumento abrupto de defunciones en pacientes con comorbilidades como diabetes, hipertensión y enfermedad cardiovascular. Este comportamiento coincide con los patrones documentados a nivel nacional durante la pandemia.

Por otro lado, los efectos indirectos revelan el impacto acumulado que tiene la mortalidad temprana sobre los años de vida perdidos en edades posteriores. En 2021 se observó un incremento relevante en la mortalidad de adultos de 40 a 59 años, muchos de ellos pertenecientes a la población económicamente activa que no pudo realizar confinamiento estricto por sus condiciones laborales. La mortalidad en estas edades genera efectos indirectos más amplios porque reduce el número de sobrevivientes que alcanzan edades avanzadas, lo cual disminuye los valores de  $L$  en toda la tabla de vida. Esto contribuye a amplificar la reducción total de la esperanza de vida más allá del impacto inmediato del COVID-19.

En conjunto, la descomposición de Arriaga muestra que la diferencia entre 2019 y 2021 no se explica únicamente por un aumento puntual en las defunciones, sino por una reconfiguración profunda de la mortalidad por edad, encabezada por un incremento drástico en las edades mayores pero acompañado también por un aumento relevante en la mortalidad de adultos en edad laboral. Así, el método permite concluir que la pandemia provocó tanto efectos directos —a través del aumento en las tasas de mortalidad específicas— como indirectos —por la disminución del número de sobrevivientes a edades avanzadas—, lo cual se tradujo en una de las mayores caídas temporales de la esperanza de vida en el estado en tiempos recientes.

## 08 Analisis de Causa eliminada

Las defunciones generales reflejan el conjunto total de muertes ocurridas en la entidad federativa, producto de una amplia gama de causas que incluyen enfermedades crónicas, accidentes, padecimientos respiratorios, enfermedades metabólicas, muertes perinatales, suicidios, homicidios y otras causas externas. En contraste, las defunciones por homicidio representan únicamente la mortalidad atribuida a eventos violentos intencionales. Comparar ambos conjuntos de datos permite dimensionar el peso real de la violencia letal dentro de la mortalidad total y entender cómo afecta la estructura por edad y por sexo de los fallecimientos.

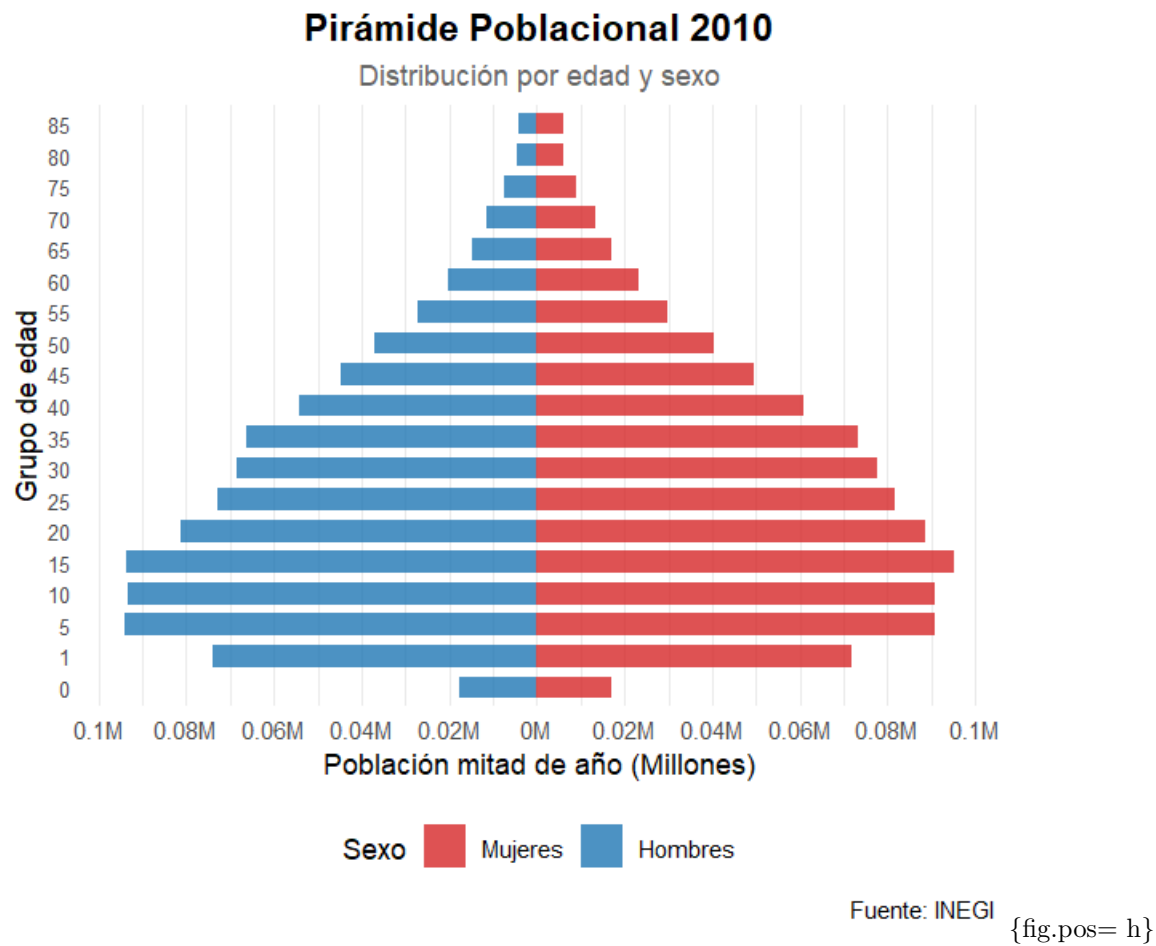
Al comparar los archivos proporcionados, se observa que los homicidios constituyen solo una fracción del total de defunciones, pero una fracción que se concentra de manera notable en edades muy específicas. Mientras que las defunciones generales siguen el patrón demográfico tradicional —bajas en edades infantiles, aumentan gradualmente durante la adultez y alcanzan su punto más alto en adultos mayores—, las defunciones por homicidio muestran una distribución completamente distinta. La mayoría de las muertes por homicidio se concentran en edades jóvenes-adultas, especialmente entre los 15 y 39 años, un rango que normalmente presenta niveles muy bajos de mortalidad por causas naturales. Esto contrasta con la mortalidad general, donde la mayor carga se ubica en edades avanzadas, principalmente después de los 60 años.

Otro punto importante es la marcada diferencia por sexo. Mientras que en la mortalidad general la distribución por sexo tiende a ser más equilibrada —aunque con cierta sobremortalidad masculina en edades adultas—, la mortalidad por homicidios muestra una desproporción extrema: la gran mayoría de víctimas son hombres, lo cual refleja el perfil de riesgo asociado a la violencia letal en México. Esta diferencia hace que los homicidios impacten significativamente más en la esperanza de vida masculina que en la femenina, lo que se confirma al comparar la tabla de vida estándar con la tabla de vida de causa eliminada. Al eliminar los homicidios, la esperanza de vida masculina aumenta de manera visible, mientras que en las mujeres el aumento es mucho menor debido a que ellas presentan una incidencia mucho más baja de homicidios.

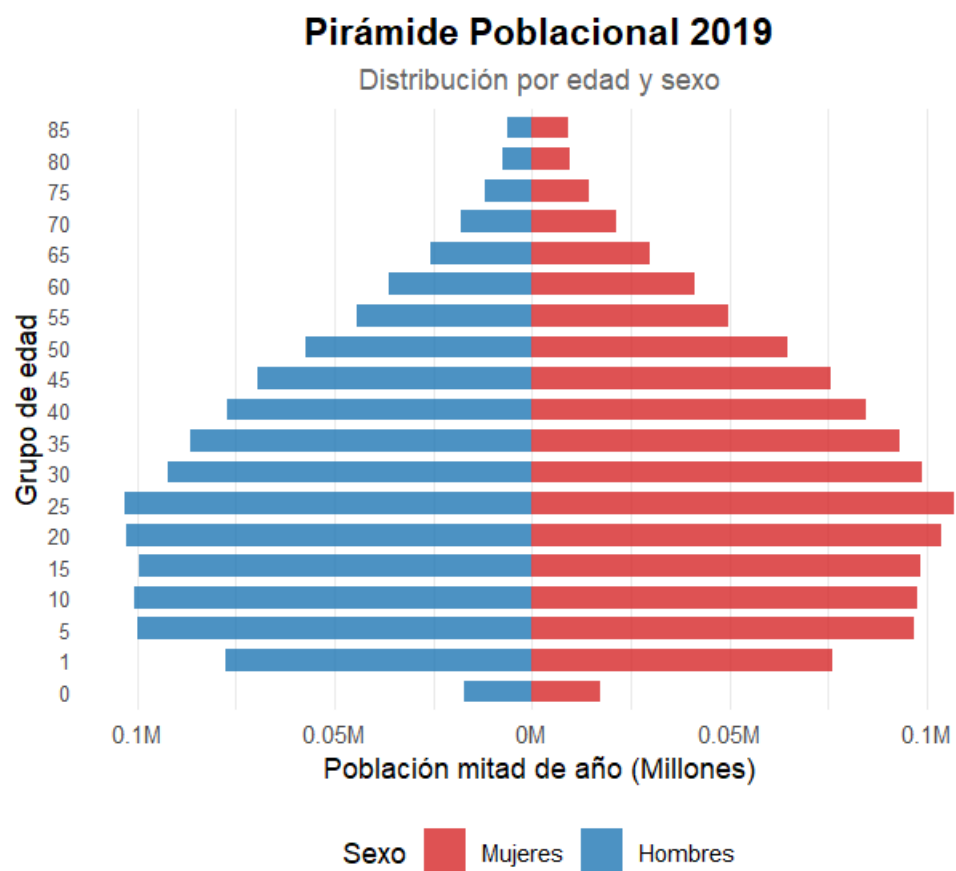
La comparación entre defunciones generales y defunciones por homicidio permite ver con claridad que, aunque los homicidios representan una proporción relativamente pequeña del total de muertes, su efecto demográfico es profundo debido a que ocurren en edades donde normalmente la mortalidad es baja. Por ello, pequeñas variaciones en el número de homicidios tienen un efecto notable sobre la esperanza de vida al nacer, especialmente en los hombres. Este contraste entre el patrón por edad y el impacto sobre la esperanza de vida

demuestra que la violencia letal no solo constituye un problema de seguridad, sino un factor determinante en la estructura de la mortalidad y en el nivel general de salud poblacional.

09 Graficas APV

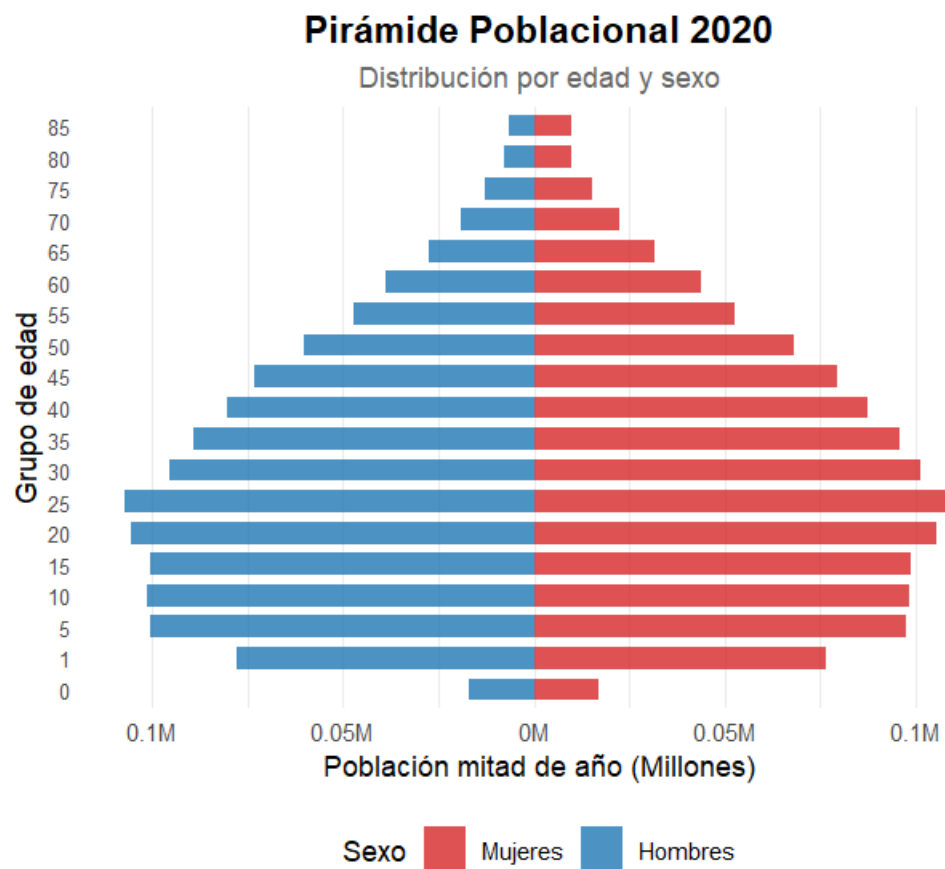


Mortalidad por año  
Mortalidad General  
APV



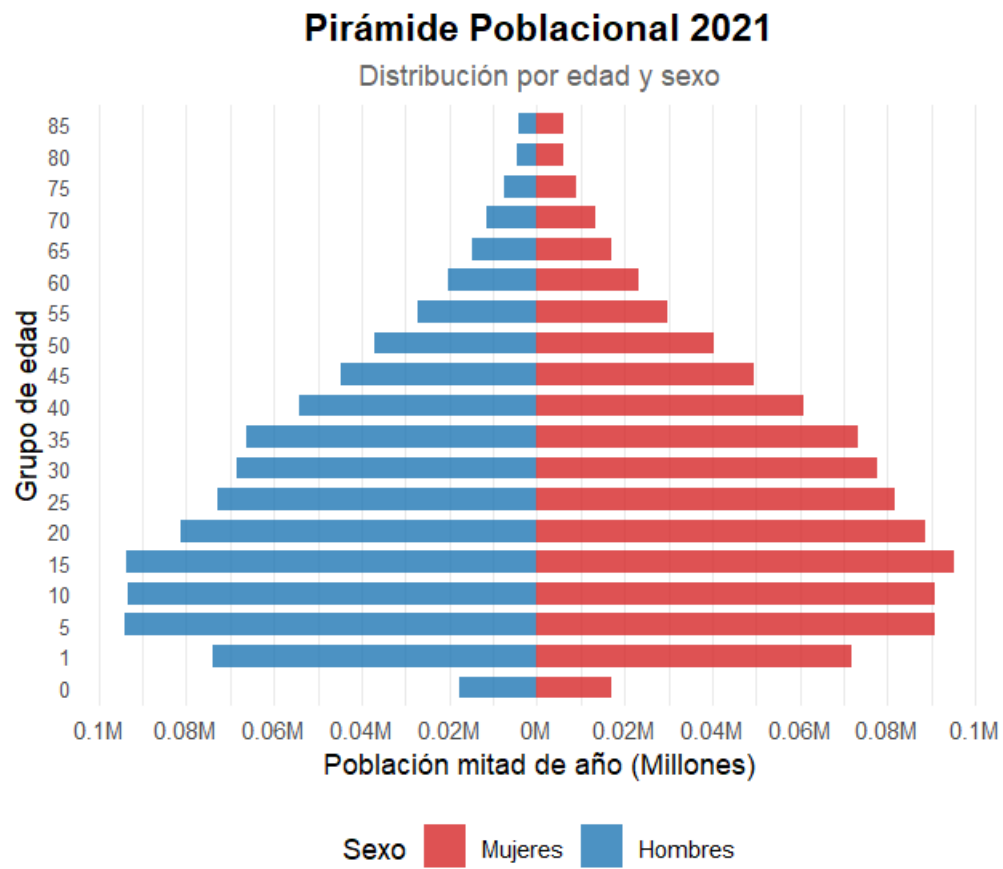
Fuente: INEGI

Figure 2: 2019



Fuente: INEGI

Figure 3: 2020



Fuente: INEGI

Figure 4: 2020

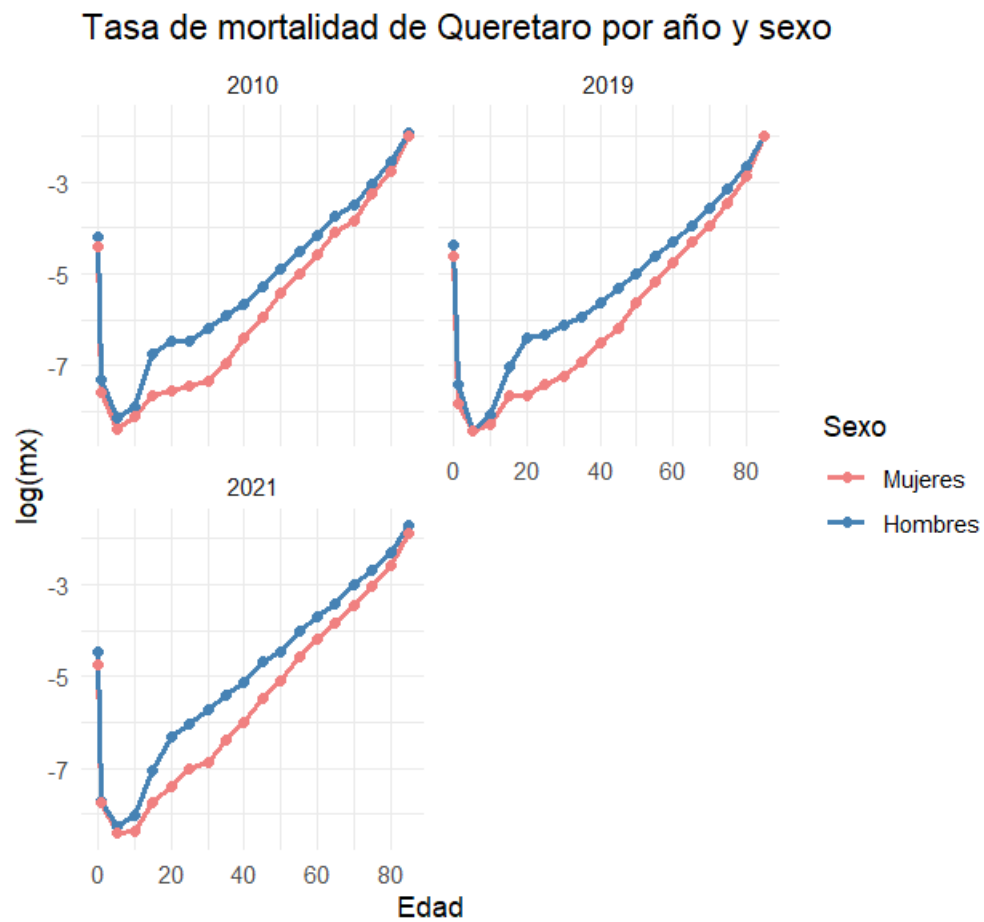


Figure 5: Mortadlida por año



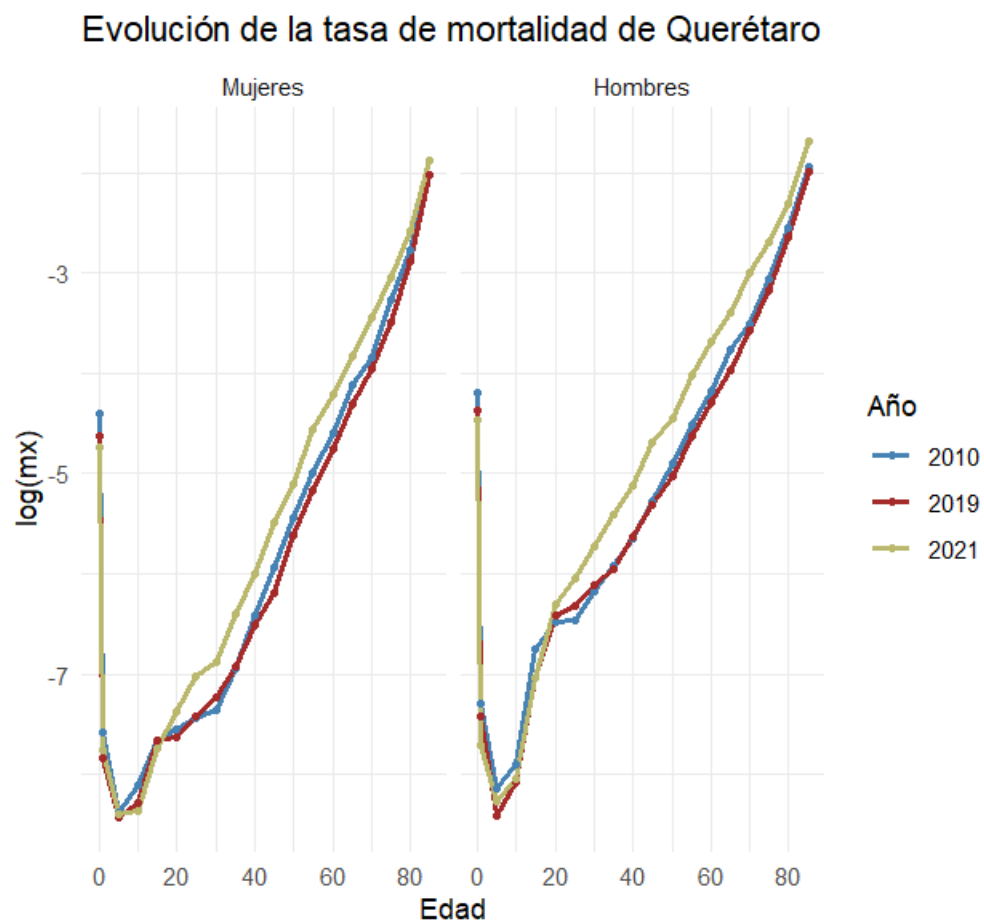


Figure 6: Mortalidad General

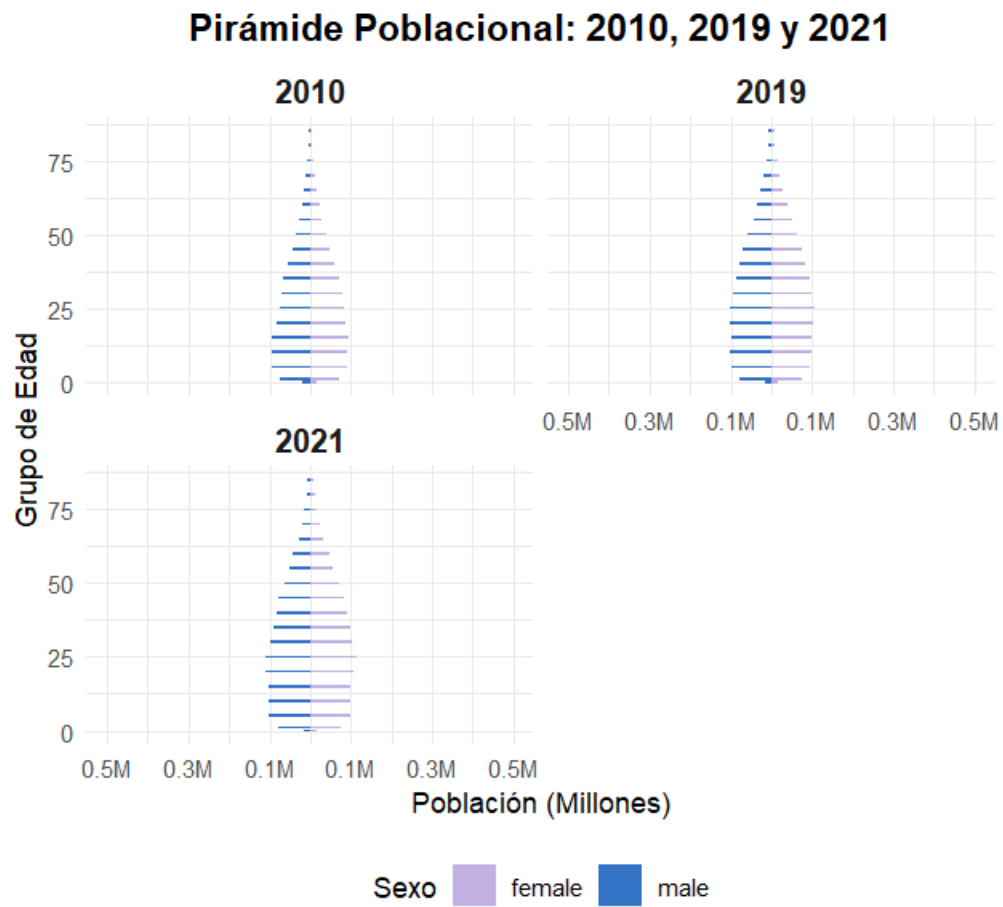


Figure 7: APV