# Decomposition - Class 1

Serena Vigezzi and José Manuel Aburto

# Why decompose?

- We use a lot of aggregate measures (TFR, life expectancy, Crude Rate of Net Migration,...)
- We compare these measures across populations and across time
- We want to explain why differences exist

# Why decompose?

- We use a lot of aggregate measures (TFR, life expectancy, Crude Migration Rate,...)
- We compare these measures across populations and across time
- We want to explain why differences exist
- BUT... many factors can influence them. How do we know which ones do what?

# Motivation to develop further methods of comparison: Decomposition

# Kitagawa decomposition

Consider the case of the crude death rate (CDR) for Japan (7.85) and Taiwan (5.82) in early 2000s. Does Taiwan really has lower mortality than the frontrunner of longevity?

# Kitagawa (1955)

$t_1$ is the initial period and $t_2$ is the final period

$D_x =$ number of deaths at age $x$

$M_x =$ death rate

$N_x =$ is the mid-year population

$N =$ total population over ages

Note that $D_x = M_x * N_x$

# Kitagawa (1955)

$$\Delta CDR = \sum_x M_x(t_2)\frac{N_x(t_2)}{N(t_2)} - \sum_x M_x(t_1)\frac{N_x(t_1)}{N(t_1)} \tag{1}$$

# Kitagawa (1955)

The difference between the crude rates can be expressed as

$$\Delta CDR = \underbrace{\sum_x M_x(t_1)\left[\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)}\right]}_{\text{Changes in x-composition}} + \underbrace{\sum_x \frac{N_x(t_1)}{N(t_1)}\left[M_x(t_2) - M_x(t_1)\right]}_{\text{Change in rates with pop 1 as standard}} + \underbrace{\sum_x (M_x(t_2) - M_x(t_1))\left[\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)}\right]}_{\text{Interaction of rates and compositions}} \tag{2}$$

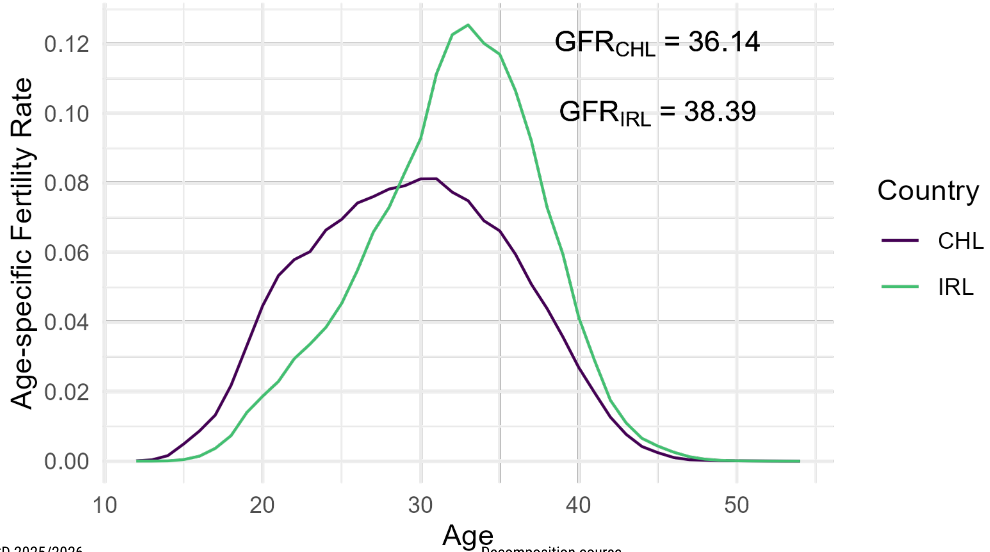
# Kitagawa (1955)

To avoid the interaction term, Kitagawa suggests

$$\Delta CDR = \underbrace{\sum_x \left( \frac{M_x(t_2) + M_x(t_1)}{2} \right) \left( \frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)} \right)}_{\text{Changes in x-composition}} + \underbrace{\sum_x \left( \frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2} \right) (M_x(t_2) - M_x(t_1))}_{\text{Changes in rates}} \tag{3}$$

Challenge 1: show that (1) can be expressed as (3) (step by step)

# An example - Chile and Ireland in 2020



# An example - Chile and Ireland in 2020

# Standardisation

| | Chile | Ireland |
|---|---|---|
| GFR | 36.14 | 38.39 |
| DS | 32.91 | 40.48 |
| IS | 38.31 | 35.65 |

# Kitagawa decomposition

Let’s decompose the change in General Fertility Rate between two years, considering composition by age

# Kitagawa decomposition

$$\Delta GFR = \underbrace{\sum_x \left(\frac{ASFR_x(t_2)+ASFR_x(t_1)}{2}\right)\left(\frac{N_x(t2)}{N(t2)} - \frac{N_x(t_1)}{N(t_1)}\right)}_{\text{Difference in population composition}} +$$

$$\underbrace{\sum_x \left(\frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2}\right)(ASFR_x(t_2) - ASFR_x(t_1))}_{\text{Difference in rates}}$$

Where

- $t_1$ is the first population, $t_2$ is the second population
- $ASFR_x$ is the fertility rate for age $x$
- $N_x$ is the mid-year population of age $x$
- $N$ is the total mid-year population

# Kitagawa decomposition

- Composition effect: difference in age-composition, keeping rates at the average of the two populations
  $\rightarrow$ *how much of the difference in GFR is because the population is older/younger?*
- Rate effect: difference in fertility rates, keeping age-composition at the average of the two populations
  $\rightarrow$ *how much of the difference in GFR is because the fertility rates are different?*

# Kitagawa decomposition

NB: Effects can be positive (push GFR up) or negative (bring GFR down)

$\rightarrow$ The total change in GFR is the sum of the two effects

Let’s find out what happens between Chile and Ireland

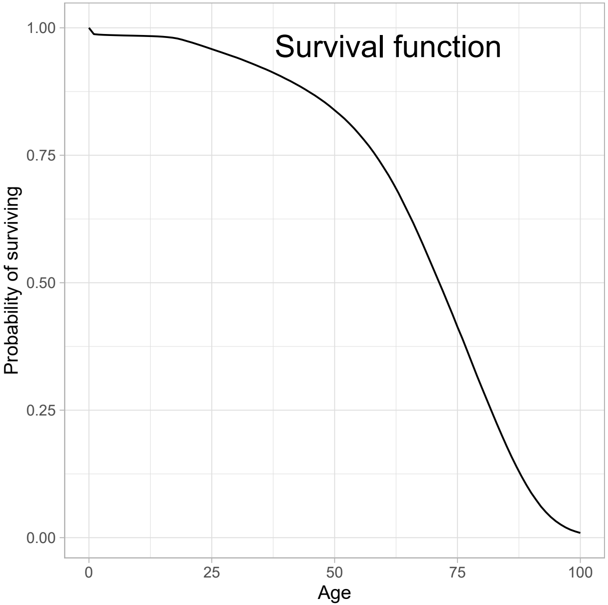
# Expansions of Kitagawa decomposition

- Basic method only handles one compositional factor (EITHER age OR marital status OR gender, etc.)
- Expanded to account for multiple compositional factors (age AND martial status AND gender, etc.)
  $\rightarrow$ more complex math
  $\rightarrow$ more factors do not necessarily explain differences better
- Similar approach in economics developed in 1970s: Oaxaca-Blinder decomposition
  $\rightarrow$ it could be considered a generalisation of the Kitagawa method (Oaxaca, 2025)

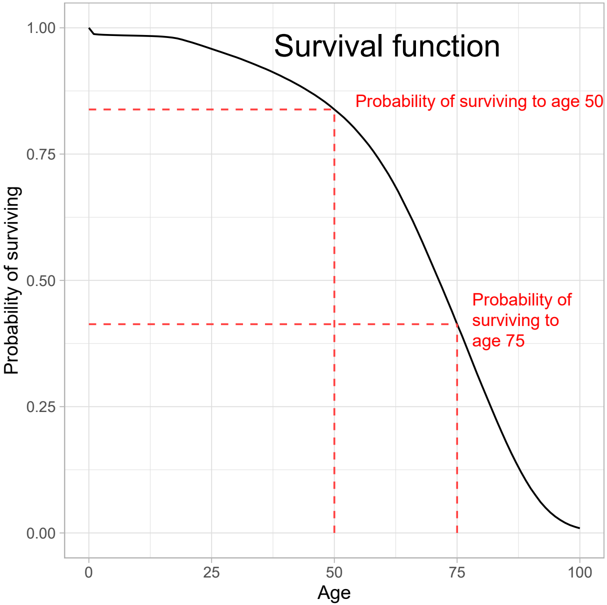
# More on compositional effects: Life expectancy, life years lost, lifespan inequality

Aburto et al 2022, PNAS

**Black men USA 2020**

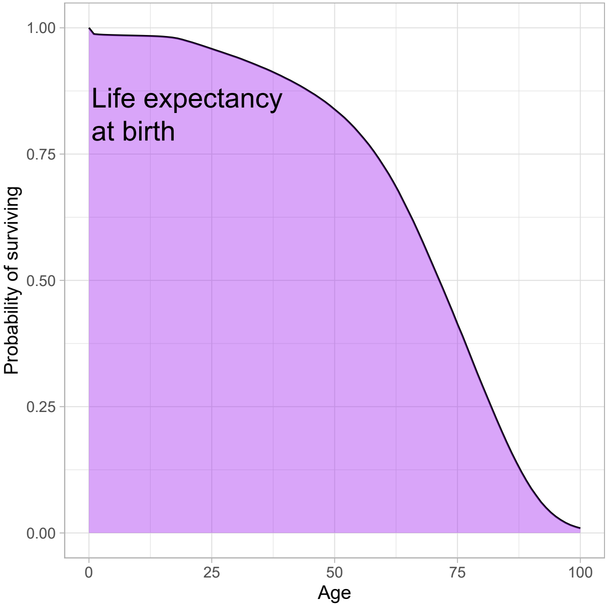


# Black men USA 2020



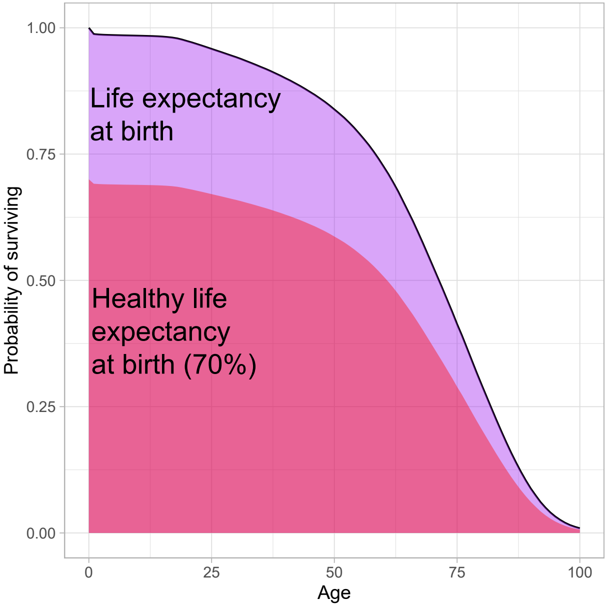
**Black men USA 2020**

Life expectancy = 67.8y



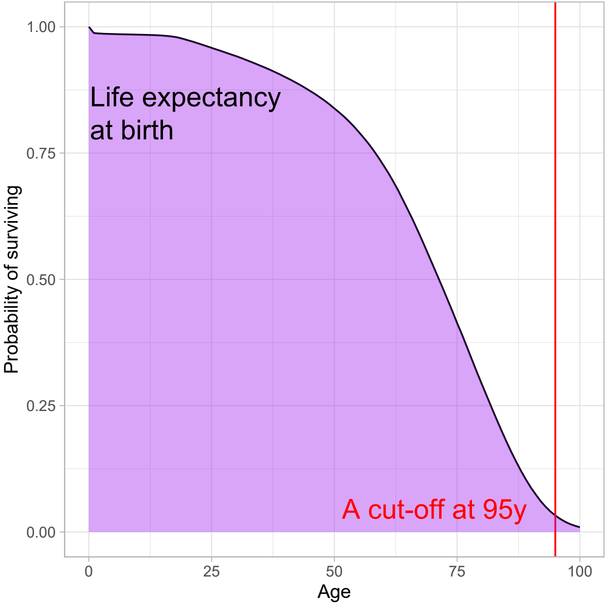
**Black men USA 2020**

Life expectancy = 67.8y



**Black men USA 2020**

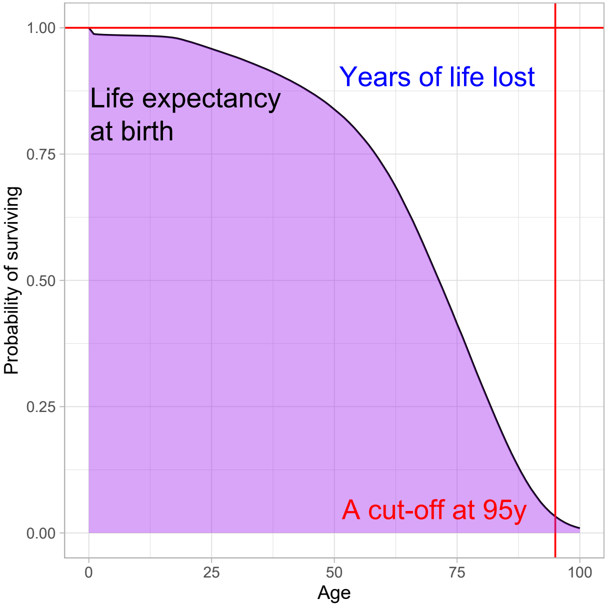
Life expectancy = 67.8y



**Black men USA 2020**

Life expectancy = 67.8y

Years lost = 28.3y

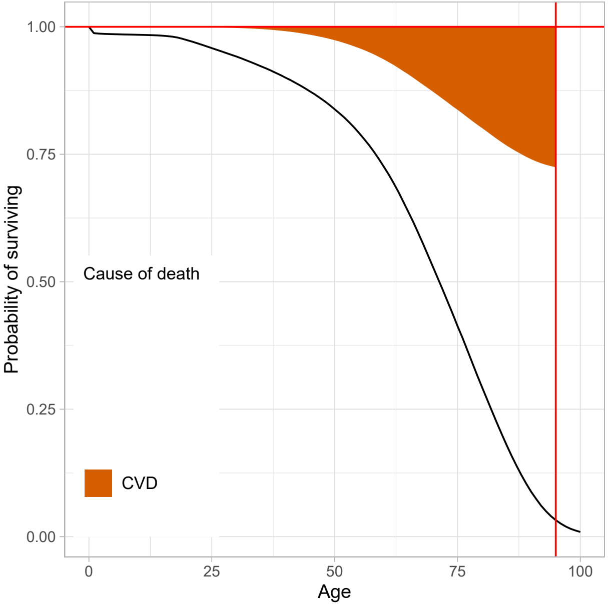


**Black men USA 2020**

Life expectancy = 67.8y

Years lost = 28.3y

CVD 7y ⟶ 24.4%



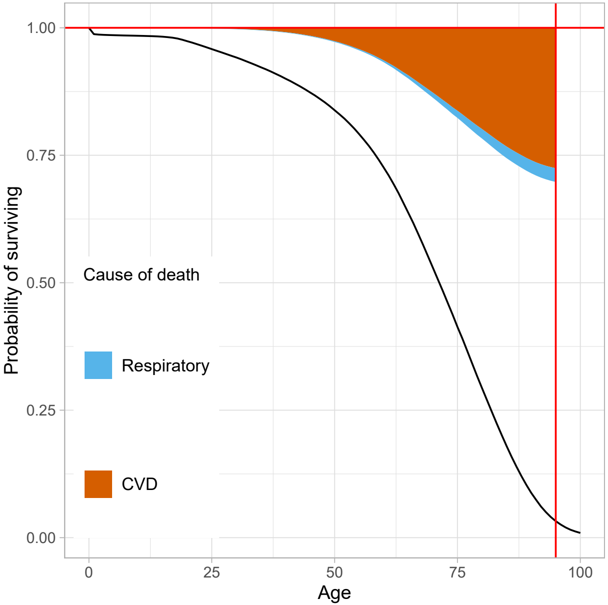
**Black men USA 2020**

Life expectancy = 67.8y

Years lost = 28.3y

CVD 7y ⟶ 24.4%

Respiratory .6y ⟶ 2.2%



**Black men USA 2020**

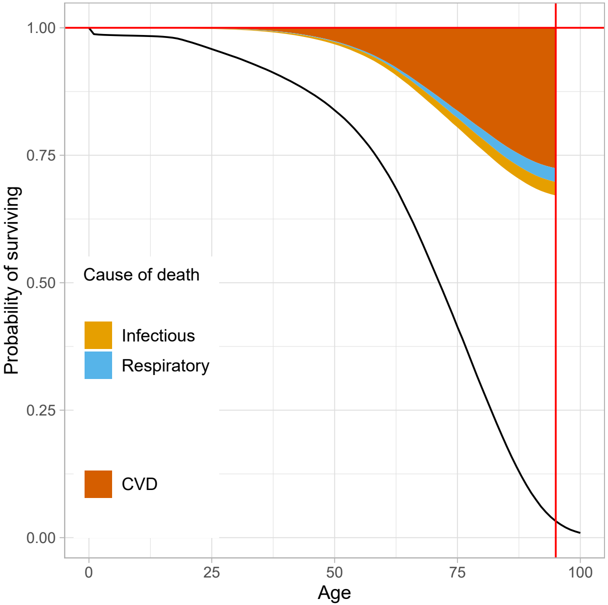
Life expectancy = 67.8y

Years lost = 28.3y

CVD 7y ⟶ 24.4%

Respiratory .6y ⟶ 2.2%

Infectious .8y ⟶ 2.8%



**Black men USA 2020**

Life expectancy = 67.8y

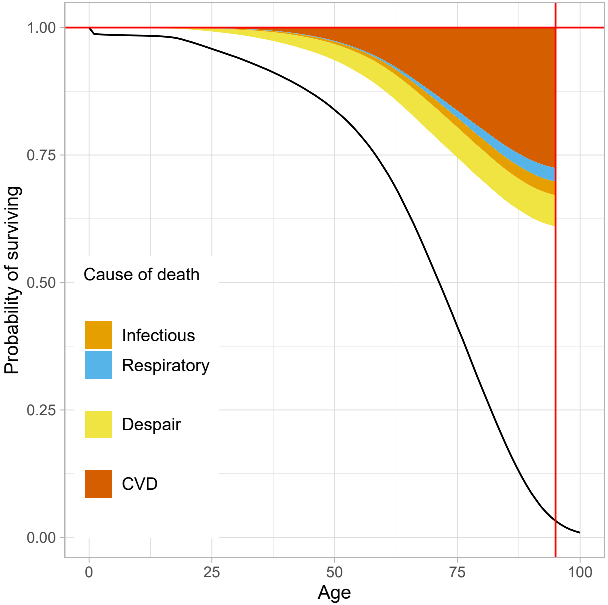
Years lost = 28.3y

CVD 7y ⟶ 24.4%

Respiratory .6y ⟶ 2.2%

Infectious .8y ⟶ 2.8%

Despair 2.9y ⟶ 10.2%



**Black men USA 2020**

Life expectancy = 67.8y

Years lost = 28.3y

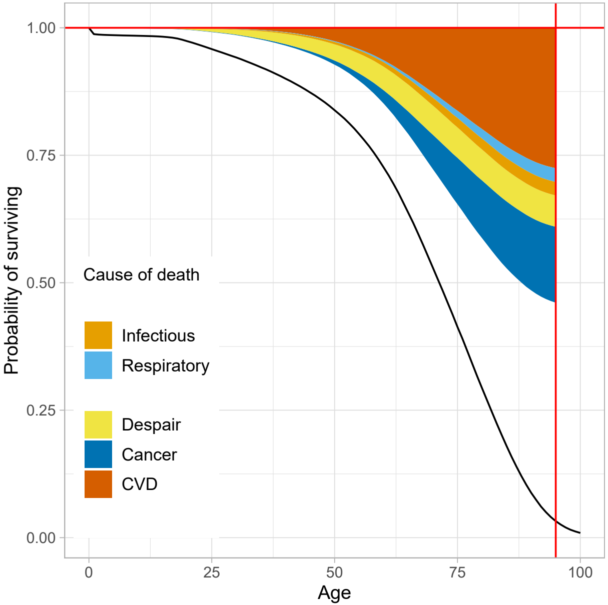
CVD 7y ⟶ 24.4%

Respiratory .6y ⟶ 2.2%

Infectious .8y ⟶ 2.8%

Despair 2.9y ⟶ 10.2%

Cancer 3.7y ⟶ 12.9%



**Black men USA 2020**

Life expectancy = 67.8y

Years lost = 28.3y

CVD 7y ⟶ 24.4%

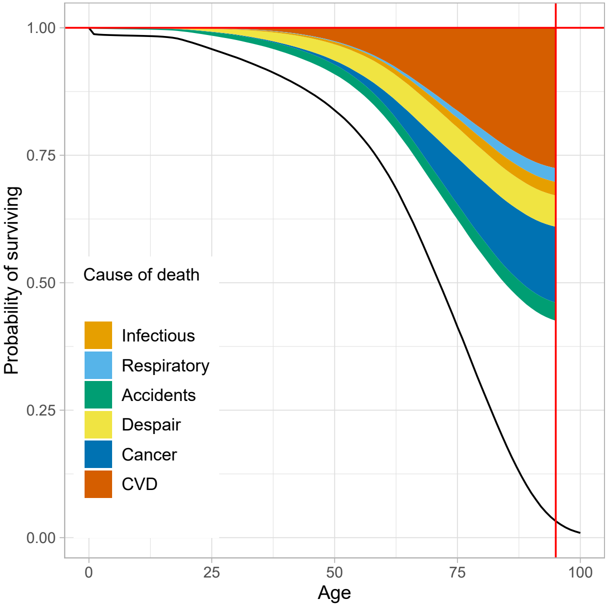
Respiratory .6y ⟶ 2.2%

Infectious .8y ⟶ 2.8%

Despair 2.9y ⟶ 10.2%

Cancer 3.7y ⟶ 12.9%

Accidents 1.7y ⟶ 5.9%



**Black men USA 2020**

Life expectancy = 67.8y

Years lost = 28.3y

CVD 7y ⟶ 24.4%

Respiratory .6y ⟶ 2.2%

Infectious .8y ⟶ 2.8%

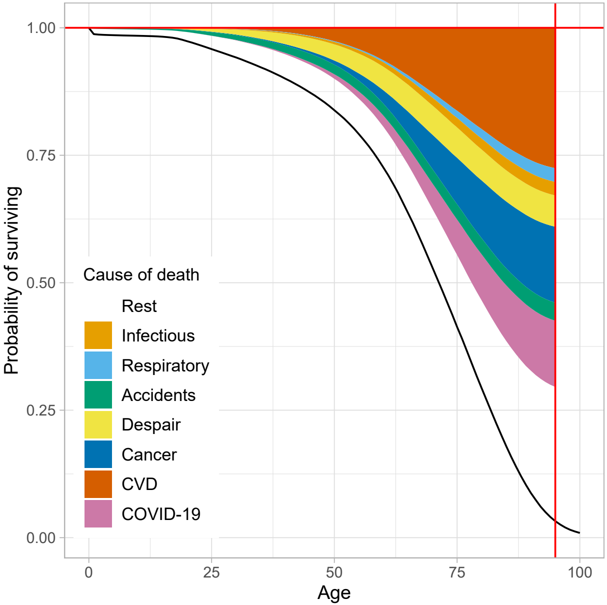
Despair 2.9y ⟶ 10.2%

Cancer 3.7y ⟶ 12.9%

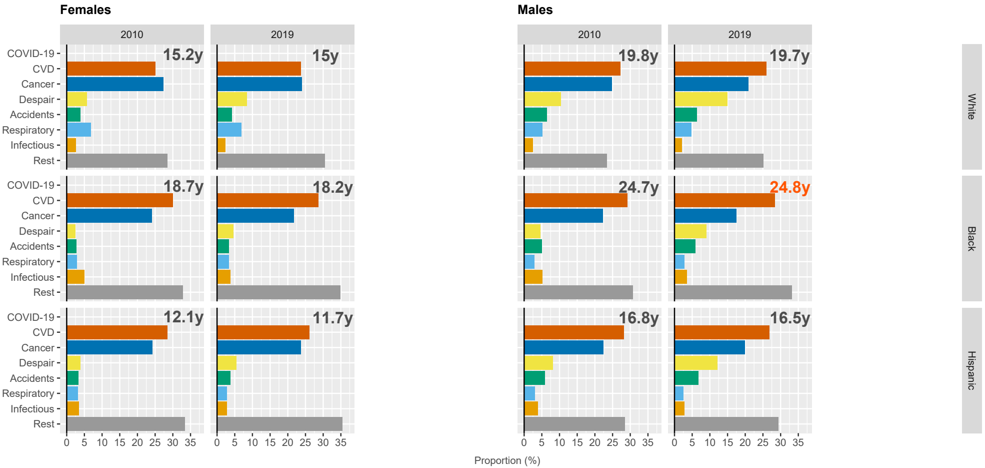
Accidents 1.7y ⟶ 5.9%

COVID-19 3y ⟶ 10.6%

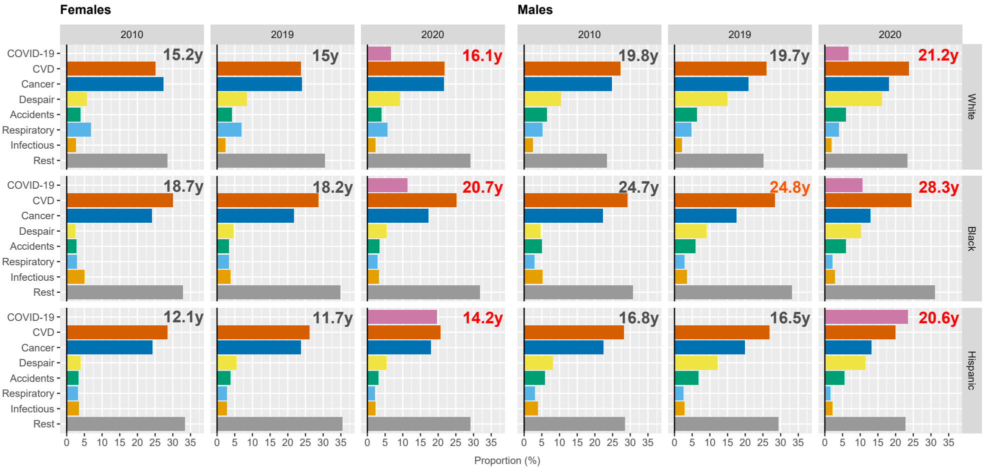
Rest 8.7y ⟶ 31%



# Average years of life lost



# Average years of life lost



# References

J.N. Tetens (1786). *Einleitung zur Berechnung der Leibrenten und Antwartschaften. Zweyter Theil. Versuche über einige bey Versorgungs-Anstalten erhebliche Puncte*, Weidmanns Erben und Reich: Leipzig.

F.G.P. Neison (1844). "On a method recently proposed for conducting inquiries into the comparative sanatory condition of various districts, with illustrations, derived from numerous places in Great Britain at the period of the last census", *Journal of the Statistical Society of London*, 7(1): 40—68.

R.L. Oaxaca and E. Sierminska (2025). "Oaxaca-Blinder meets Kitagawa: What is the link?", *PloS One*, 20(5): e0321874.