

Causes of Death in the United Kingdom (2021)

AUTHOR

ID 2941737

Summary

Understanding the leading causes of mortality is essential for guiding public health policy, allocating healthcare resources effectively, and assessing the impact of medical interventions. This project aims to answer the research question: *What are the top causes of death in the United Kingdom, and how do they differ by sex?* Specifically, it explores mortality patterns across males and females using official death statistics.

This study is of personal and academic interest because it provides insight into public health priorities and highlights disparities in health outcomes between population groups. The COVID-19 pandemic has further underscored the importance of understanding mortality trends, making this analysis especially relevant. Examining sex-based differences in leading causes of death can help inform targeted health interventions, improve disease prevention strategies, and contribute to ongoing discussions in public health research.

The dataset used in this project was obtained from the World Health Organization (WHO) webpage (<https://data.who.int/countries/826>), which compiles mortality statistics from national vital registration systems. The data, disaggregated by country, year, sex, age group, and cause of death, was selected for the year 2021 to provide the most recent and complete information for the United Kingdom, reflecting the ongoing impact of COVID-19. This publicly available dataset, follows international standards such as the International Classification of Diseases (ICD), ensuring accuracy and standardization.

Key steps in the analysis

The analysis was conducted entirely in R, with a strong focus on creating clear, compelling, and reproducible visualizations following the principles from *Storytelling with Data* by Cole Nussbaumer Knaflic. This book emphasizes selecting visuals that best match the story we want to tell, eliminating unnecessary clutter, and guiding the audience's attention to the most important insights.

The primary visual tool used was the bar chart, which served both as an introduction and as a precise means of comparing mortality causes. Bar charts were chosen because they allow for a clear and immediate understanding of rank and magnitude, making them ideal for guiding the audience through the key findings. At the start of the analysis, special emphasis was placed on highlighting the leading cause of death—Alzheimer's disease and other dementias—which stood out as the main cause of mortality in 2021, even surpassing COVID-19, which ranked second. This deliberate choice to open with a strong, attention-grabbing message aligns with Knaflic's recommendations to focus the audience's attention and set the narrative tone.

As the presentation progressed, the focus shifted to highlighting the top three causes of death, which were visualized using a card-style graphic format. These "cards" worked as visual summaries, helping to distill the main findings into an accessible and engaging format, and allowing the audience to grasp key messages at a glance. To examine differences by sex, separate datasets were used for females and males, drawn from the same WHO source and year. This approach allowed for the creation of targeted visualizations that clearly presented the top causes of death for each sex, before culminating in a comparative analysis. The comparative visualizations were designed to emphasize contrasts and similarities between the sexes, helping to reveal patterns that might not be visible when looking at each group separately.

Data manipulation involved filtering by country and year, grouping by sex and cause, and summarisation. Throughout the visual design, I prioritized simplicity, clear labeling, and strategic use of color to maintain the audience's focus and support the overall storytelling approach.

To ensure full reproducibility of this analysis, the following R packages must be installed and loaded

```
options(repos = c(CRAN = "https://cran.rstudio.com"))  
install.packages("here")
```

Installing package into '/Users/aitanaburgos/Library/R/arm64/4.4/library'
(as 'lib' is unspecified)

The downloaded binary packages are in
/var/folders/qg/kmf3xg_10y7fkww_05qjnvxm0000gn/T//Rtmpvbz0Em/downloaded_packages

```
install.packages("ggplot2")
```

Installing package into '/Users/aitanaburgos/Library/R/arm64/4.4/library'
(as 'lib' is unspecified)

The downloaded binary packages are in
/var/folders/qg/kmf3xg_10y7fkww_05qjnvxm0000gn/T//Rtmpvbz0Em/downloaded_packages

```
install.packages("dplyr")
```

Installing package into '/Users/aitanaburgos/Library/R/arm64/4.4/library'
(as 'lib' is unspecified)

The downloaded binary packages are in
/var/folders/qg/kmf3xg_10y7fkww_05qjnvxm0000gn/T//Rtmpvbz0Em/downloaded_packages

```
install.packages("plotly")
```

Installing package into '/Users/aitanaburgos/Library/R/arm64/4.4/library'
(as 'lib' is unspecified)

The downloaded binary packages are in
/var/folders/qg/kmf3xg_10y7fkww_05qjnvxm0000gn/T//Rtmpvbz0Em/downloaded_packages

```
install.packages("tibble")
```

Installing package into '/Users/aitanaburgos/Library/R/arm64/4.4/library'
(as 'lib' is unspecified)

The downloaded binary packages are in
/var/folders/qg/kmf3xg_10y7fkww_05qjnvxm0000gn/T//Rtmpvbz0Em/downloaded_packages

Required libraries

```
library(here)
```

here() starts at /Users/aitanaburgos/Downloads/2941737_files

```
library(tidyverse)
```

— Attaching core tidyverse packages — tidyverse 2.0.0 —

| | | | |
|-------------|-------|-----------|-------|
| ✓ dplyr | 1.1.4 | ✓ readr | 2.1.5 |
| ✓ forcats | 1.0.0 | ✓ stringr | 1.5.1 |
| ✓ ggplot2 | 3.5.2 | ✓ tibble | 3.2.1 |
| ✓ lubridate | 1.9.4 | ✓ tidyr | 1.3.1 |
| ✓ purrr | 1.0.2 | | |

— Conflicts — tidyverse_conflicts() —

✖ dplyr::filter() masks stats::filter()

✖ dplyr::lag() masks stats::lag()

ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

```
library(ggplot2)
```

```
library(dplyr)
```

```
library(plotly)
```

Attaching package: 'plotly'

The following object is masked from 'package:ggplot2':

last_plot

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

filter

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

layout

```
library(tibble)
```

Start of the analysis

The mortality dataset used includes age-standardized death rates (per 100,000 population) for all age groups, all causes classified as rankable, and for both sexes combined, filtered for available countries and the year 2021.

```
here("Data")
```

[1] "/Users/aitanaburgos/Downloads/2941737_files/Data"

```
Death <- read.csv("https://xmart-api-public.who.int/DEX_CMS/GHE_FULLL_DD?$filter=FLAG_RANKABLE%20in%
```

Check the data

```
glimpse(Death)
```

```
Rows: 24,522
Columns: 5
$ DIM_COUNTRY_CODE      <chr> "AFG", "AFG", "AFG", "AFG", "AFG", "AFG", "A...
$ DIM_YEAR_CODE         <int> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 20...
$ DIM_GHECAUSE_TITLE    <chr> "Syphilis", "Genital herpes", "Diphtheria", ...
$ DIM_SEX_CODE          <chr> "BTSX", "BTSX", "BTSX", "BTSX", "BTSX", "BTS...
$ VAL_DTHS_RATE100K_NUMERIC <dbl> 1.46, 0.00, 0.11, 0.00, 0.00, 0.17, 33.11, 0...
```

This code filters the full WHO mortality dataset to extract the top 10 leading causes of death in the United Kingdom in 2021, based on the age-standardized death rate per 100,000 population:

```
Top10_uk <- Death %>%
  filter(DIM_COUNTRY_CODE == "GBR", DIM_SEX_CODE == "BTSX", DIM_YEAR_CODE == 2021) %>%
  arrange(desc(VAL_DTHS_RATE100K_NUMERIC)) %>%
  select(DIM_GHECAUSE_TITLE, VAL_DTHS_RATE100K_NUMERIC) %>%
  head(10)
```

Select the top 10 causes of death in the UK

```
Top10_uk
```

| | DIM_GHECAUSE_TITLE | VAL_DTHS_RATE100K_NUMERIC |
|-----------------|---------------------------------------|---------------------------|
| | <chr> | <dbl> |
| 1 | Alzheimer disease and other dementias | 128.04 |
| 2 | COVID-19 | 103.85 |
| 3 | Ischaemic heart disease | 103.69 |
| 4 | Lower respiratory infections | 56.92 |
| 5 | Stroke | 53.91 |
| 6 | Chronic obstructive pulmonary disease | 53.80 |
| 7 | Trachea, bronchus, lung cancers | 51.99 |
| 8 | Colon and rectum cancers | 29.22 |
| 9 | Prostate cancer | 19.84 |
| 10 | Breast cancer | 19.84 |
| 1-10 of 10 rows | | |

Load WHO mortality dataset

For simplicity and consistency with WHO reporting, this analysis uses standardized mortality rates (per 100,000 population) directly from the source dataset, without attempting to estimate absolute death counts.

```
Death <- read.csv ("https://xmart-api-public.who.int/DEX_CMS/GHE_FULL_DD?$filter=FLAG_RANKABLE%20in
```

Filter

```
uk_data <- Death %>%
  filter(DIM_COUNTRY_CODE == "GBR", DIM_YEAR_CODE == 2021)
```

Select the top 10 causes of death and their mortality rates

```
top10_uk <- uk_data %>%
  arrange(desc (VAL_DTHS_RATE100K_NUMERIC)) %>%
  slice(1:10)
```

Display

```
top10_uk %>%
  select (DIM_GHECAUSE_TITLE, VAL_DTHS_RATE100K_NUMERIC)
```

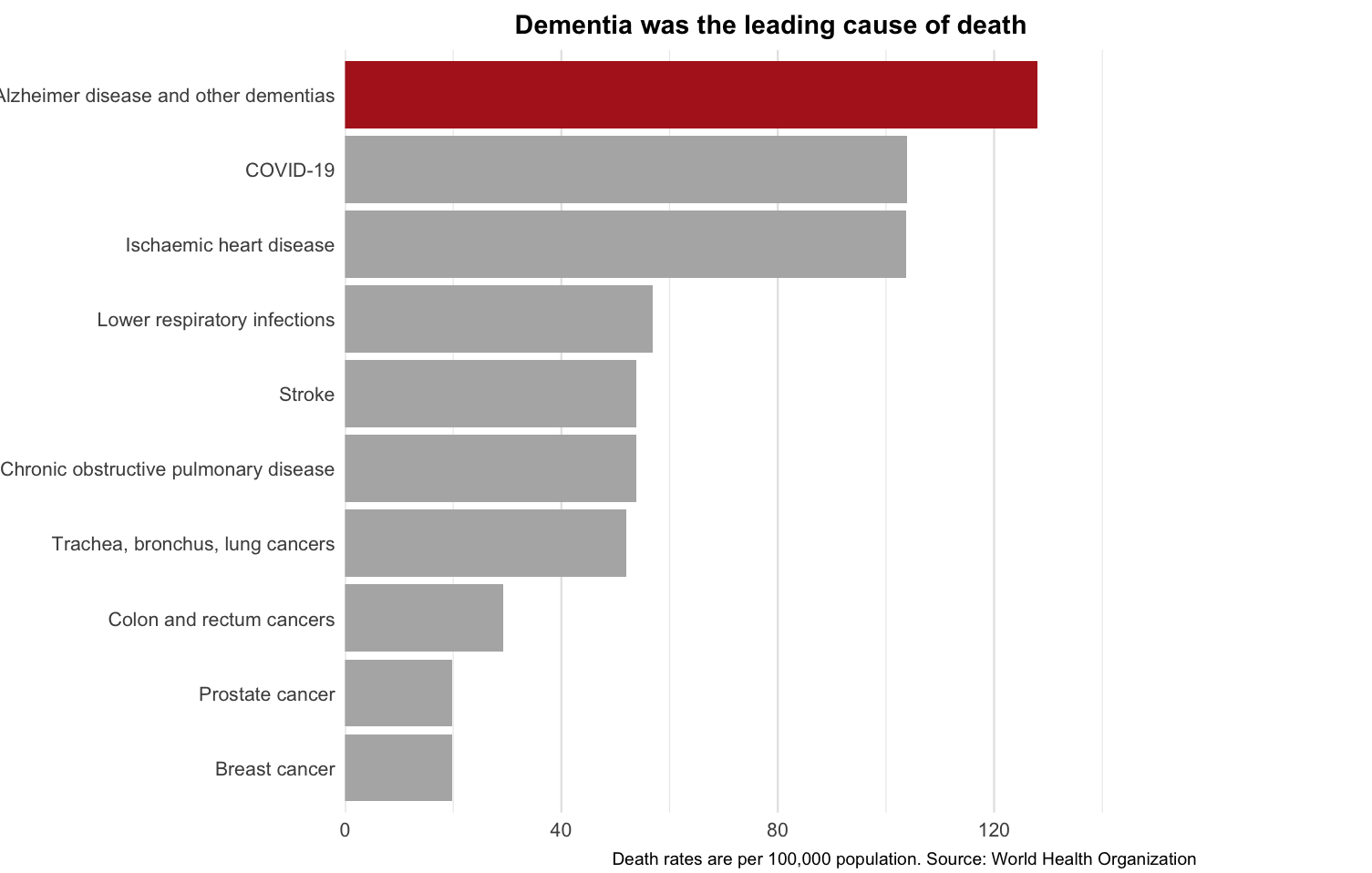
| DIM_GHECAUSE_TITLE | VAL_DTHS_RATE100K_NUMERIC |
|---------------------------------------|---------------------------|
| <chr> | <dbl> |
| Alzheimer disease and other dementias | 128.04 |
| COVID-19 | 103.85 |
| Ischaemic heart disease | 103.69 |
| Lower respiratory infections | 56.92 |
| Stroke | 53.91 |
| Chronic obstructive pulmonary disease | 53.80 |
| Trachea, bronchus, lung cancers | 51.99 |
| Colon and rectum cancers | 29.22 |
| Prostate cancer | 19.84 |
| Breast cancer | 19.84 |
| 1-10 of 10 rows | |

Bar plot

The use of horizontal bars supports easy reading of long cause-of-death labels, while the data is sorted in descending order to immediately highlight the most significant values. To focus the audience’s attention, the top-ranked cause (dementias) is visually distinguished in red, drawing the eye without overwhelming the rest of the chart.

```
#Highlight the top cause to focus attention
top_cause <- top10_uk$DIM_GHECAUSE_TITLE[1]
```

```
ggplot(top10_uk, aes(x = reorder(DIM_GHECAUSE_TITLE, VAL_DTHS_RATE100K_NUMERIC),
                                y = VAL_DTHS_RATE100K_NUMERIC,
                                fill = DIM_GHECAUSE_TITLE == top_cause)) +
  geom_col(show.legend = FALSE) +
  coord_flip() +
  scale_fill_manual(values = c("TRUE" = "firebrick", "FALSE" = "grey70")) +
  scale_y_continuous(limits = c(0, 150), expand = expansion(mult = c(0, 0.05))) +
  labs(
    title = "Dementia was the leading cause of death",
    caption = "Death rates are per 100,000 population. Source: World Health Organization"
  ) +
  theme_minimal(base_size = 9) +
  theme(
    axis.title.x = element_blank(), # Remove x-axis label
    axis.title.y = element_blank(), # Remove y-axis label
    axis.text = element_text(size = 8),
    plot.title = element_text(face = "bold", hjust = 0.5),
    panel.grid.major.y = element_blank(), # Remove horizontal gridlines for clarity
    panel.grid.major.x = element_line(color = "grey90") # Subtle reference lines
  )
```



Top 3 graphic

To present the top 3 causes of death, I used a clean card-style layout to prioritize clarity and comparability. This format highlights key values without visual clutter. I used ChatGPT for technical guidance—e.g., aligning text elements and refining

layout with `geom_tile()`.

```
# Reordering the top 3
top3 <- Death %>%
  filter(DIM_COUNTRY_CODE == "GBR", DIM_YEAR_CODE == 2021) %>%
  filter(DIM_GHECAUSE_TITLE %in% c("Alzheimer disease and other dementias", "COVID-19", "Ischaemic
mutate(
  ShortName = case_when(
    DIM_GHECAUSE_TITLE == "Alzheimer disease and other dementias" ~ "Dementia",
    DIM_GHECAUSE_TITLE == "COVID-19" ~ "COVID-19",
    DIM_GHECAUSE_TITLE == "Ischaemic heart disease" ~ "Heart disease"
  ),
  RankText = case_when(
    ShortName == "Dementia" ~ "Highest rate",
    ShortName == "COVID-19" ~ "2nd highest rate",
    ShortName == "Heart disease" ~ "3rd highest rate"
  ),
  ShortName = factor(ShortName, levels = c("Dementia", "COVID-19", "Heart disease")),
  Rate = round(VAL_DTHS_RATE100K_NUMERIC, 1)
)

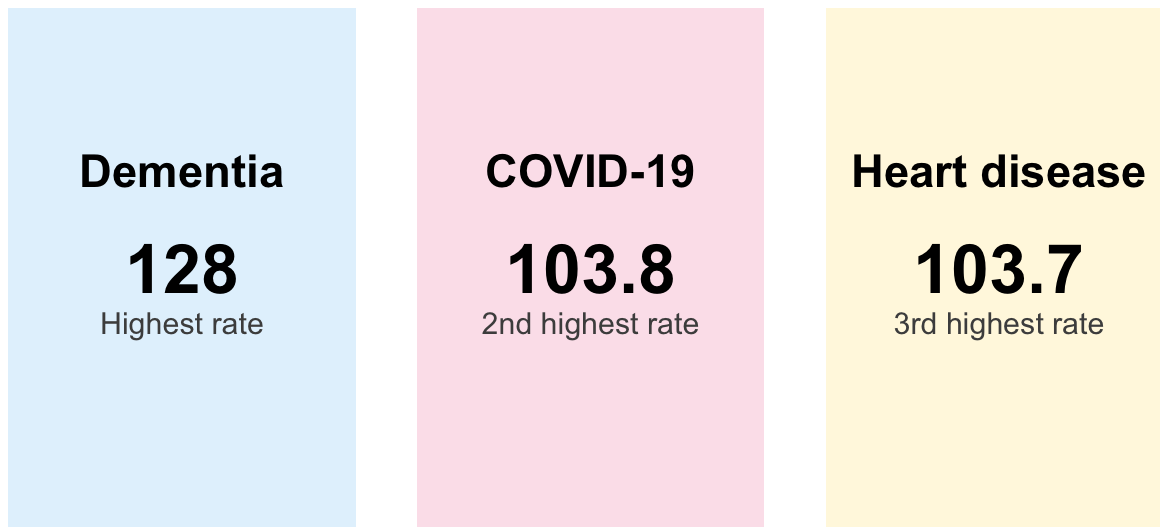
# Colors
bg_colors <- c("Dementia" = "#E3F2FD", "COVID-19" = "#FCE4EC", "Heart disease" = "#FFF8E1")

# Plot
ggplot(top3, aes(x = ShortName, y = 1, fill = ShortName)) +
  geom_tile(width = 0.85, height = 0.85, color = NA) + # card-like background
  geom_text(aes(label = ShortName), vjust = -2.5, size = 6, fontface = "bold") + # title
  geom_text(aes(label = Rate), vjust = 0.5, size = 9, fontface = "bold", color = "black") + # valu
  geom_text(aes(label = RankText), vjust = 3, size = 4, color = "gray30") + # description

scale_fill_manual(values = bg_colors) +
scale_x_discrete(position = "bottom") +
coord_fixed(ratio = 1.5) +
labs(
  title = "Top 3 Causes of Death in the UK (2021)",
  subtitle = "Deaths per 100,000 population",
  x = NULL, y = NULL
) +
theme_minimal() +
theme(
  axis.text = element_blank(),
  axis.ticks = element_blank(),
  panel.grid = element_blank(),
  plot.title = element_text(hjust = 0.5, face = "bold", size = 15),
  plot.subtitle = element_text(hjust = 0.5, size = 11),
  legend.position = "none"
)
```

Top 3 Causes of Death in the UK (2021)

Deaths per 100,000 population



Top 5 bar chart

Following the visual logic of the previous graphic, the next chart expands the view to the top five while maintaining the same color scheme for consistency. This continuation helps build a cohesive narrative while introducing new insights and specific numbers.

```
# Filtering top 5 causes
top5 <- Death %>%
  filter(DIM_COUNTRY_CODE == "GBR", DIM_YEAR_CODE == 2021) %>%
  arrange(desc(VAL_DTHS_RATE100K_NUMERIC)) %>%
  slice(1:5) %>%
  mutate(
    DIM_GHECAUSE_TITLE = case_when(
      DIM_GHECAUSE_TITLE == "Alzheimer disease and other dementias" ~ "Dementia",
      DIM_GHECAUSE_TITLE == "Ischaemic heart disease" ~ "Heart disease",
      TRUE ~ DIM_GHECAUSE_TITLE
    ),
    cause_order = factor(DIM_GHECAUSE_TITLE, levels = c(
      "Dementia", "COVID-19", "Heart disease", "Stroke", "Lower respiratory infections"
    )),
    label = paste0(round(VAL_DTHS_RATE100K_NUMERIC, 1))
  )

# Consistent color palette
color_palette <- c(
  "Dementia" = "#1E88E5",
```



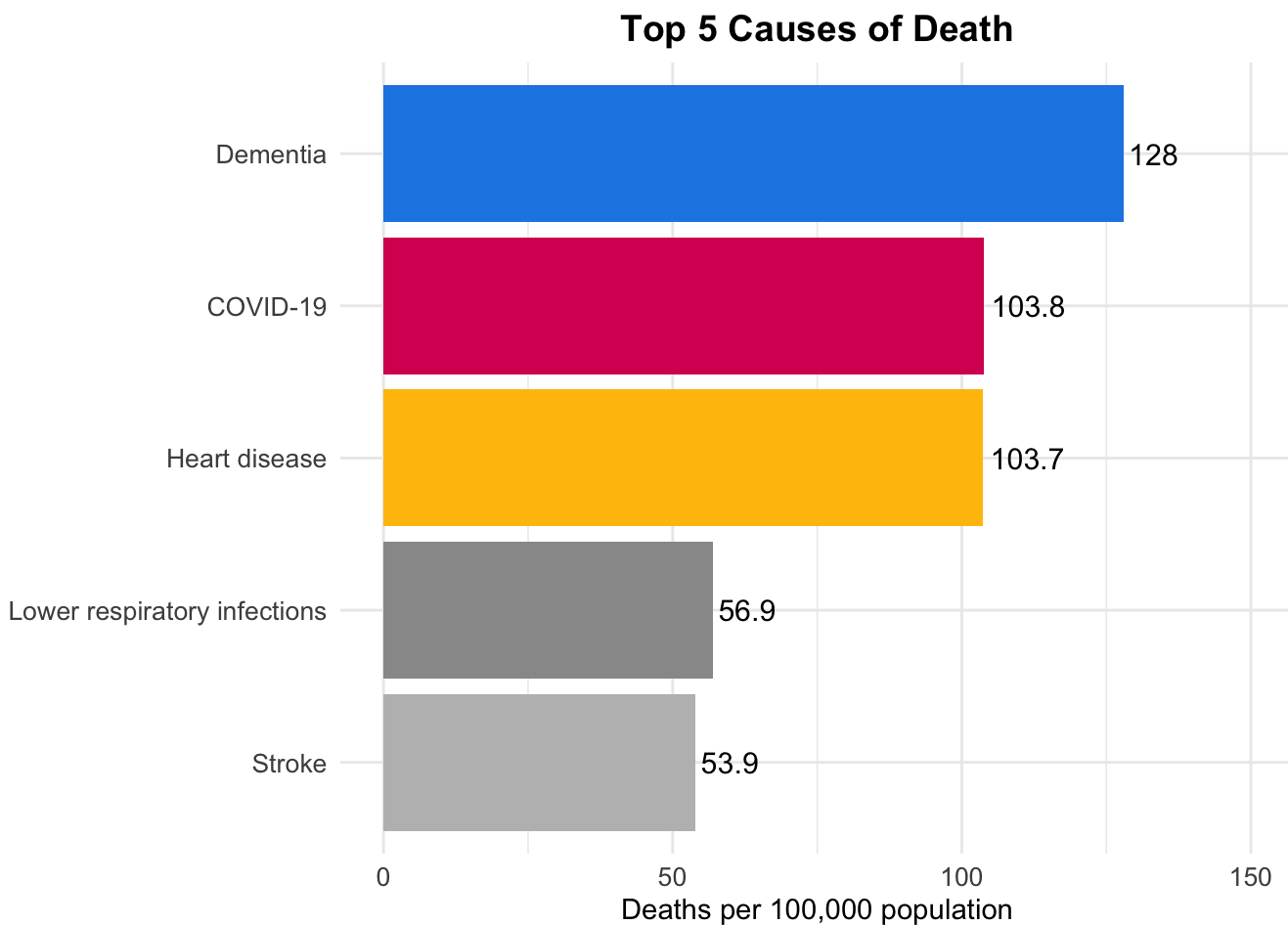
```

"COVID-19" = "#D81B60",
"Heart disease" = "#FFC107",
"Stroke" = "#BDBDBD",
"Lower respiratory infections" = "#999999"
)

# Bar chart
ggplot(top5, aes(x = reorder(DIM_GHECAUSE_TITLE, VAL_DTHS_RATE100K_NUMERIC),
                          y = VAL_DTHS_RATE100K_NUMERIC,
                          fill = cause_order)) +

  geom_col() +
  coord_flip() +
  geom_text(aes(label = label), hjust = -0.1, size = 4) +
  scale_fill_manual(values = color_palette, guide = "none") +
  labs(
    title = "Top 5 Causes of Death",
    y = "Deaths per 100,000 population",
    x = NULL
  ) +
  ylim(0, 150) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    axis.text = element_text(size = 10)
  )

```



Top causes of death (females)

A dedicated dataset filtered by sex (DIM_SEX_CODE == 'FMLE') is employed to focus on the leading causes of death among females. This approach allows for a targeted analysis of gender-specific mortality patterns and draws attention to diseases that disproportionately affect women.

```
Females <- read.csv("https://xmart-api-public.who.int/DEX_CMS/GHE_FULL_DD?$filter=FLAG_RANKABLE%20i
```

Check the data

```
glimpse(Females)
```

Rows: 24,522
Columns: 5
\$ DIM_COUNTRY_CODE <chr> "AFG", "AFG", "AFG", "AFG", "AFG", "AFG", "A...
\$ DIM_YEAR_CODE <int> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 20...
\$ DIM_GHECAUSE_TITLE <chr> "Syphilis", "Genital herpes", "Diphtheria", ...
\$ DIM_SEX_CODE <chr> "FMLE", "FMLE", "FMLE", "FMLE", "FMLE", "FML...
\$ VAL_DTHS_RATE100K_NUMERIC <dbl> 1.40, 0.00, 0.10, 0.00, 0.00, 0.05, 35.80, 0...

Donut plot 1

The donut plot was selected to clearly visualize proportional differences. Bright colors emphasize the leading causes of death, while less significant causes were shown in gray to avoid distracting from the main analysis. Basic interactivity allows users to click on segments to view specific death counts and causes.

```
# Top 5 causes of death for women in the UK (2021)
top5_female_causes <- Females %>%
  filter(DIM_COUNTRY_CODE == "GBR", DIM_YEAR_CODE == 2021) %>%
  arrange(desc(VAL_DTHS_RATE100K_NUMERIC)) %>%
  slice(1:5) %>%
  mutate(
    DIM_GHECAUSE_TITLE = case_when(
      DIM_GHECAUSE_TITLE == "Alzheimer disease and other dementias" ~ "Dementia",
      DIM_GHECAUSE_TITLE == "Ischaemic heart disease" ~ "Heart disease",
      TRUE ~ DIM_GHECAUSE_TITLE
    ),
    hovertext = case_when(
      DIM_GHECAUSE_TITLE == "Dementia" ~ "Nearly 128 out of every 100,000 women in the UK<br>died f
      DIM_GHECAUSE_TITLE == "COVID-19" ~ "Approximately 103 out of every 100,000 women<br>died from
      DIM_GHECAUSE_TITLE == "Heart disease" ~ "Around 104 per 100,000 women<br>died from heart dise
      DIM_GHECAUSE_TITLE == "Stroke" ~ "Roughly 76 per 100,000 women<br>died from stroke in 2021.",
      DIM_GHECAUSE_TITLE == "Lower respiratory infections" ~ "About 46 per 100,000 women<br>died fr
    )
  )

# Color palette matching earlier graphics
color_palette <- c(
```

```

"Dementia" = "#1E88E5",
"COVID-19" = "#D81B60",
"Heart disease" = "#FFC107",
"Stroke" = "#BDBDBD",
"Lower respiratory infections" = "#999999"
)

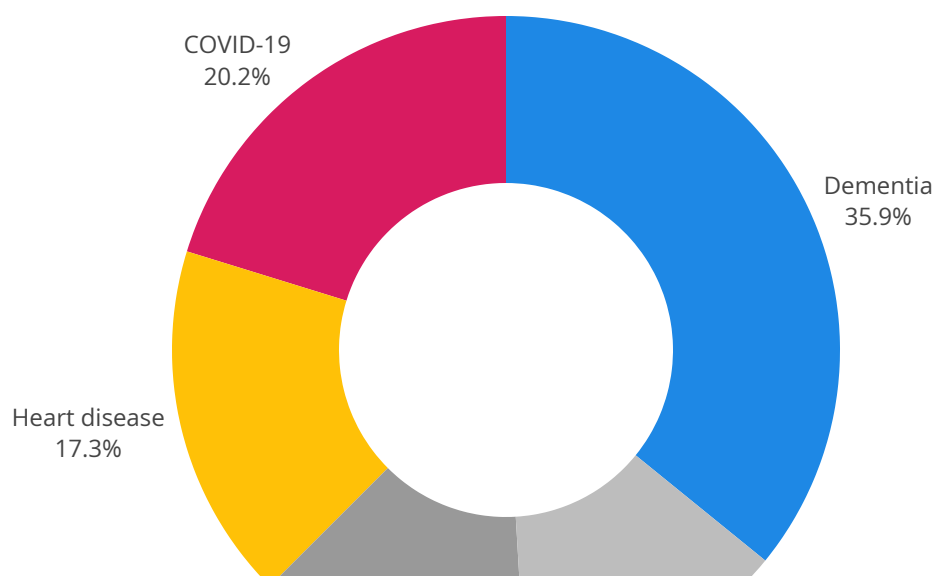
colors <- unname(color_palette[top5_female_causes$DIM_GHECAUSE_TITLE])

# Create donut plot
donut <- plot_ly(
  top5_female_causes,
  labels = ~DIM_GHECAUSE_TITLE,
  values = ~VAL_DTHS_RATE100K_NUMERIC,
  type = 'pie',
  hole = 0.5,
  text = ~hovertext,
  hoverinfo = 'text',
  textinfo = 'label+percent',
  textposition = 'outside',
  marker = list(colors = colors),
  sort = FALSE
)

# Layout
donut %>%
  layout(
    title = list(
      text = "Top 5 Causes of Death in UK Women (2021)",
      x = 0.5,
      y = 0.95
    ),
    showlegend = FALSE,
    margin = list(t = 80, b = 50)
  )

```

Top 5 Causes of Death in UK Women (2021)





Similarly, we analyze mortality data specific to males using the same methodology, filtering for records where DIM_SEX_CODE == "MLE". This subset reveals the most significant causes of death among men in the UK in 2021. By isolating male-specific data, we can identify patterns and risk factors that may differ substantially from those affecting women, informing gender-sensitive policy and healthcare planning.

Top causes of death (male)

```
Male <- read.csv("https://xmart-api-public.who.int/DEX_CMS/GHE_FULL_DD?$filter=FLAG_RANKABLE%20in%20(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000,1001,1002,1003,1004,1005,1006,1007,1008,1009,1010,1011,1012,1013,1014,1015,1016,1017,1018,1019,1020,1021,1022,1023,1024,1025,1026,1027,1028,1029,1030,1031,1032,103
```

Check the data

glimpse (Male)

Rows: 24,522

Columns: 5

```
$ DIM_COUNTRY_CODE      <chr> "AFG", "AFG", "AFG", "AFG", "AFG", "AFG", "A...
```

```
$ DIM_YEAR_CODE      <int> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 20...
```

```
$ DIM_GHECAUSE_TITLE      <chr> "Syphilis", "Genital herpes", "Diphtheria", ...
```

```
$ DIM_SEX_CODE      <chr> "MLE", "MLE", "MLE", "MLE", "MLE", "MLE", "M...
```

```
$ VAL_DTHS_RATE100K_NUMERIC <dbl> 1.53, 0.00, 0.11, 0.00, 0.00, 0.29, 30.51, 0...
```

Donut plot 2

To ensure comparability, the male chart was designed to mirror the female version, with the same colors applied to shared causes of death. The male data required additional preprocessing, utilizing label mapping techniques to shorten lengthy cause names, such as “Trachea, bronchus, lung cancers,” thereby enhancing readability.

```
# Names
label_map <- tibble::tibble(
  original = c(
    "Alzheimer disease and other dementias",
    "COVID-19",
    "Ischaemic heart disease",
    "Trachea, bronchus, lung cancers",
    "Stroke",
    "Chronic obstructive pulmonary disease",
    "Lower respiratory infections"
  ),
  short_label = c(
    "Dementia",
    "COVID-19",
    "Heart disease",
    "Cancer",
  )
)
```

```

    "Stroke",
    "Chronic lung disease",
    "Respiratory infections"
  ),
  hovertext = c(
    "Nearly 88 out of every 100,000 men in the UK<br>died from Alzheimer disease and other dementia",
    "Approximately 122 out of every 100,000 men<br>died from COVID-19 in 2021.",
    "Around 163 per 100,000 men<br>died from ischaemic heart disease in 2021.",
    "Nearly 70 per 100,000 men<br>died from Trachea, bronchus, lung cancers in 2021.",
    "Roughly 93 per 100,000 men<br>died from stroke in 2021.",
    "Roughly 52 per 100,000 men<br>died from chronic obstructive pulmonary disease in 2021.",
    "About 46 per 100,000 men<br>died from lower respiratory infections in 2021."
  )
)

# Filtering top 5
top5_male_causes <- Male %>%
  filter(DIM_COUNTRY_CODE == "GBR", DIM_YEAR_CODE == 2021, DIM_SEX_CODE == "MLE") %>%
  arrange(desc(VAL_DTHS_RATE100K_NUMERIC)) %>%
  slice(1:5) %>%
  left_join(label_map, by = c("DIM_GHECAUSE_TITLE" = "original")) %>%
  mutate(
    label = ifelse(!is.na(short_label), short_label, DIM_GHECAUSE_TITLE),
    hover = ifelse(!is.na(hovertext), hovertext, DIM_GHECAUSE_TITLE)
  )

# Colors
color_palette <- c(
  "Dementia" = "#1E88E5",
  "COVID-19" = "#D81B60",
  "Heart disease" = "#FFC107",
  "Cancer" = "#999999",
  "Stroke" = "#F4511E",
  "Chronic lung disease" = "#BDBDBD",
  "Respiratory infections" = "#3949AB"
)

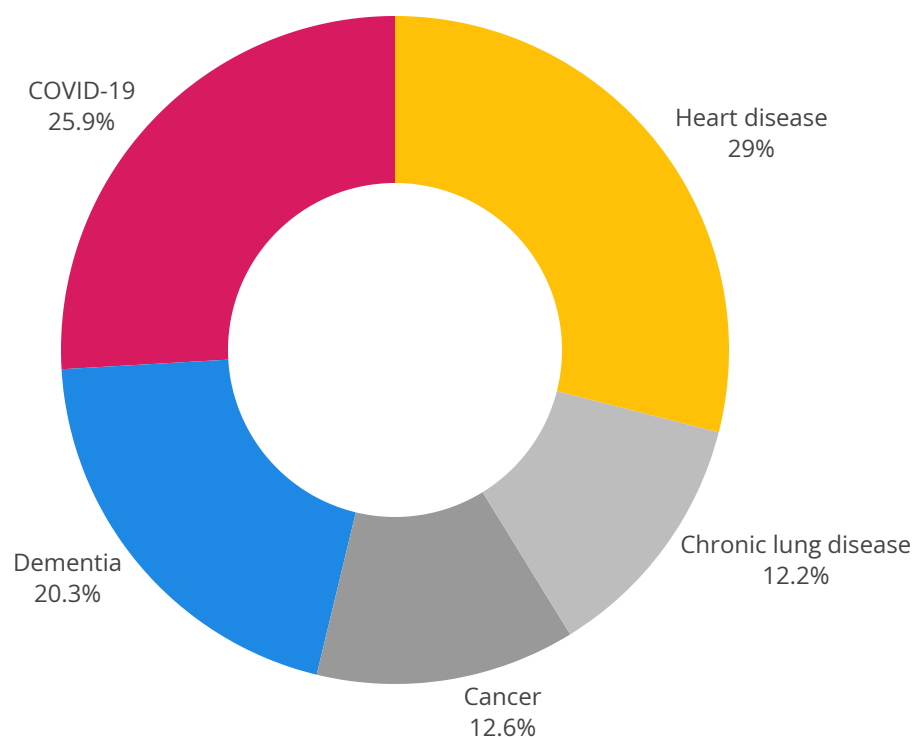
colors <- unname(color_palette[top5_male_causes$label])

# Donut plot
donut_male <- plot_ly(
  top5_male_causes,
  labels = ~label,
  values = ~VAL_DTHS_RATE100K_NUMERIC,
  type = 'pie',
  hole = 0.5,
  text = ~hover,
  hoverinfo = 'text',
  textinfo = 'label+percent',
  textposition = 'outside',
  marker = list(colors = colors),
  sort = FALSE
)

```

```
# Layout
donut_male %>%
  layout(
    title = list(
      text = "Top 5 Causes of Death in UK Men (2021)",
      x = 0.5,
      y = 0.95
    ),
    showlegend = FALSE,
    margin = list(t = 80, b = 50)
  )
```

Top 5 Causes of Death in UK Men (2021)



Combined chart

To provide a comprehensive overview, comparative visualisations highlight differences in the leading causes of death between men and women. The use of pink and blue, traditional gender-associated colours, aids intuitive interpretation. This side-by-side analysis is essential for identifying disparities, prioritising interventions, and allocating health resources equitably.

```
# Female data
top_females <- Females %>%
  filter(DIM_COUNTRY_CODE == "GBR", DIM_YEAR_CODE == 2021, DIM_SEX_CODE == "FMLE") %>%
  arrange(desc(VAL_DTHS_RATE100K_NUMERIC)) %>%
  slice(1:10) %>%
  mutate(
    deaths_per_100k = VAL_DTHS_RATE100K_NUMERIC,
    sex = "Female"
  )
```

```

# Male data
top_males <- Male %>%
  filter(DIM_COUNTRY_CODE == "GBR", DIM_YEAR_CODE == 2021, DIM_SEX_CODE == "MLE") %>%
  filter(DIM_GHECAUSE_TITLE %in% top_females$DIM_GHECAUSE_TITLE) %>%
  mutate(
    deaths_per_100k = VAL_DTHS_RATE100K_NUMERIC,
    sex = "Male"
  )

# Combining datasets
combined <- bind_rows(top_females, top_males)

# Ordering the causes of death from highest to lowest
combined <- combined %>%
  group_by(DIM_GHECAUSE_TITLE) %>%
  mutate(total_abs_deaths = sum(deaths_per_100k)) %>%
  ungroup() %>%
  mutate(
    DIM_GHECAUSE_TITLE = factor(
      DIM_GHECAUSE_TITLE,
      levels = rev(unique(DIM_GHECAUSE_TITLE[order(total_abs_deaths)]))
    ),
    sex = factor(sex, levels = c("Male", "Female"))
  )

# Scale
max_val <- 100

# Graphic
plot_ly(
  combined,
  y = ~DIM_GHECAUSE_TITLE,
  x = ~ifelse(sex == "Male", -deaths_per_100k, deaths_per_100k),
  color = ~sex,
  colors = c("Male" = "royalblue", "Female" = "deeppink"),
  type = 'bar',
  orientation = 'h',
  text = ~paste0(sex, "<br>", DIM_GHECAUSE_TITLE, "<br>", round(deaths_per_100k, 1), " per 100k"),
  hoverinfo = 'text',
  textposition = 'none'
) %>%
  layout(
    title = "Causes of Death by Gender",
    barmode = 'relative',
    xaxis = list(
      title = "",
      range = c(-max_val, max_val),
      tickvals = seq(-100, 100, by = 25),
      tickformat = ",d",
      zeroline = TRUE,
      showgrid = FALSE
    ),
    yaxis = list(
      title = "",

```

```

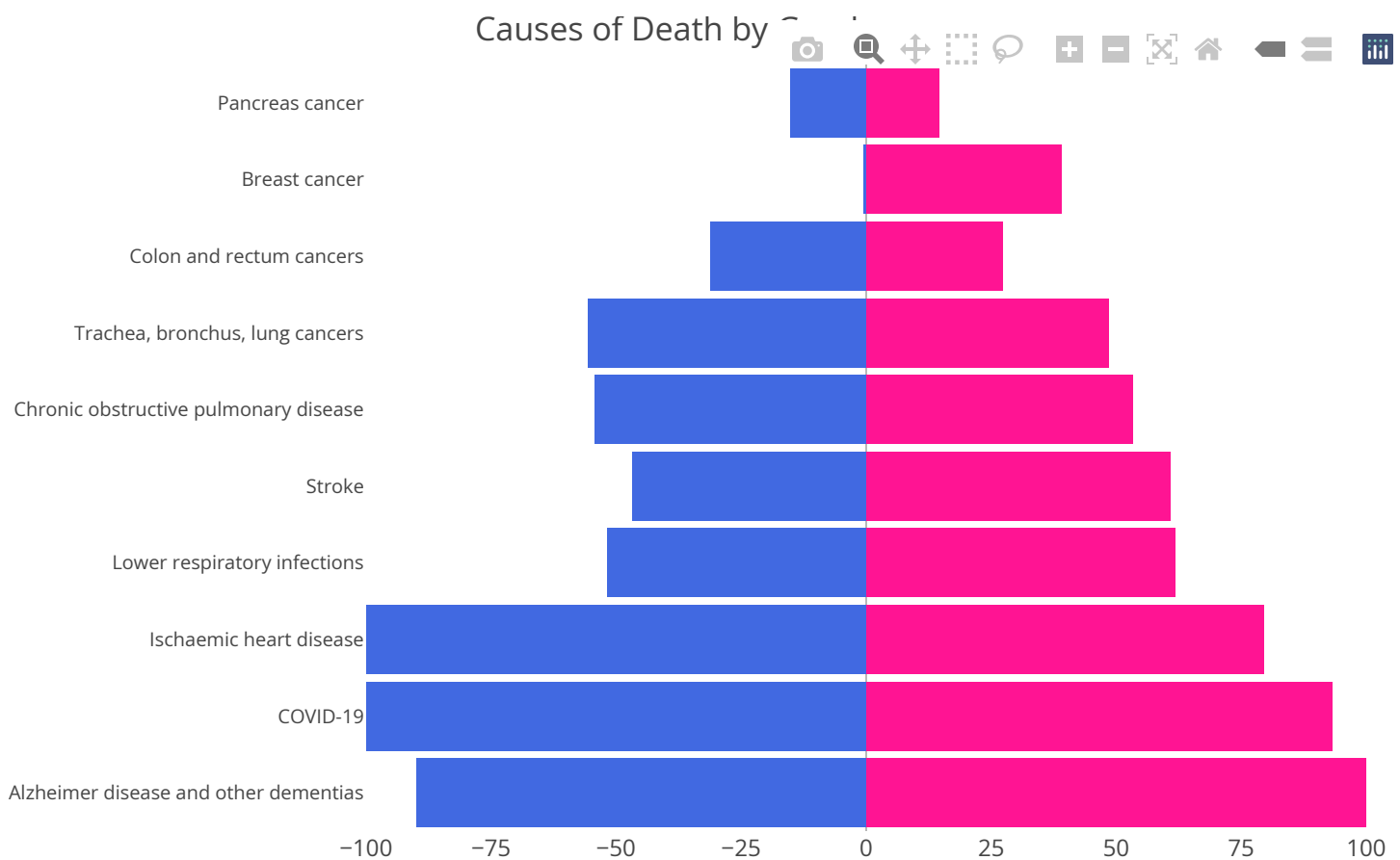
    tickfont = list(size = 10),
    showgrid = FALSE
  ),
  margin = list(l = 150),
  automargin = TRUE,
  showlegend = FALSE,
  bargap = 0.1
)

```

Warning: 'layout' objects don't have these attributes: 'automargin'

Valid attributes include:

'_deprecated', 'activeshape', 'annotations', 'autosize', 'autotypenumbers', 'calendar', 'clickmode', 'coloraxis', 'colorscale', 'colorway', 'computed', 'datarevision', 'dragmode', 'editrevision', 'editType', 'font', 'geo', 'grid', 'height', 'hidesources', 'hoverdistance', 'hoverlabel', 'hovermode', 'images', 'legend', 'mapbox', 'margin', 'meta', 'metasrc', 'modebar', 'newshape', 'paper_bgcolor', 'plot_bgcolor', 'polar', 'scene', 'selectdirection', 'selectionrevision', 'separators', 'shapes', 'showlegend', 'sliders', 'smith', 'spikedistance', 'template', 'ternary', 'title', 'transition', 'uirevision', 'uniformtext', 'updatemenus', 'width', 'xaxis', 'yaxis', 'barmode', 'bargap', 'mapType'



Conclusion

This project addressed the research question *What are the top causes of death in the United Kingdom, and how do they differ by sex?* by analyzing official mortality statistics. The findings reveal that dementia disproportionately affected women, with a 15% higher mortality rate compared to men, while men were 12% more likely to die from heart disease and 6% more affected by COVID-19. Notably, despite the impact of the pandemic, dementia emerged as the leading cause of death overall. These results

underscore the importance of recognizing sex-specific mortality patterns to inform targeted public health interventions and policies.

References

- World Health Organization (2023). *Mortality Database*. Retrieved from <https://data.who.int/countries/826>
- Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.