# Hyperiondev

# Exploratory Data Analysis on the Total Population Death Rates Data Set for England and Wales

Visit our website

# Introduction

The data set[1] gives the number of deaths per person in England and Wales, broken down by year, sex and age group. The first 23 rows of the data set are shown in **Figure 1** for illustrative purposes. The time series runs from 1841 (as shown) until 2020, inclusive. The figures shown in this report were produced in the file 'death_rates.ipynb'.

| | Year | Age | Female | Male | Total |
|---|---|---|---|---|---|
| 0 | 1841 | 0 | 0.136067 | 0.169189 | 0.152777 |
| 1 | 1841 | 1-4 | 0.035751 | 0.037354 | 0.036548 |
| 2 | 1841 | 5-9 | 0.009277 | 0.009614 | 0.009446 |
| 3 | 1841 | 10-14 | 0.005401 | 0.005107 | 0.005252 |
| 4 | 1841 | 15-19 | 0.007967 | 0.007168 | 0.007574 |
| 5 | 1841 | 20-24 | 0.008955 | 0.009230 | 0.009084 |
| 6 | 1841 | 25-29 | 0.009988 | 0.009843 | 0.009919 |
| 7 | 1841 | 30-34 | 0.010247 | 0.009740 | 0.010002 |
| 8 | 1841 | 35-39 | 0.012524 | 0.012056 | 0.012294 |
| 9 | 1841 | 40-44 | 0.012005 | 0.012407 | 0.012202 |
| 10 | 1841 | 45-49 | 0.015165 | 0.016955 | 0.016043 |
| 11 | 1841 | 50-54 | 0.015729 | 0.018582 | 0.017111 |
| 12 | 1841 | 55-59 | 0.026112 | 0.030241 | 0.028107 |
| 13 | 1841 | 60-64 | 0.028219 | 0.032153 | 0.030089 |
| 14 | 1841 | 65-69 | 0.050663 | 0.056947 | 0.053591 |
| 15 | 1841 | 70-74 | 0.066053 | 0.071833 | 0.068734 |
| 16 | 1841 | 75-79 | 0.113454 | 0.120098 | 0.116518 |
| 17 | 1841 | 80-84 | 0.144390 | 0.162246 | 0.152207 |
| 18 | 1841 | 85-89 | 0.207801 | 0.227472 | 0.215942 |
| 19 | 1841 | 90-94 | 0.286633 | 0.315919 | 0.297634 |
| 20 | 1841 | 95-99 | 0.372440 | 0.423774 | 0.389173 |
| 21 | 1841 | 100-104 | 0.461807 | 0.579146 | 0.492271 |
| 22 | 1841 | 105-109 | 0.666953 | 2.098266 | 0.765881 |
| 23 | 1841 | 110+ | . | . | . |

Figure 1: The beginning of the data set

## DATA CLEANING AND MISSING DATA

Two things immediately become apparent on inspecting the data shown in Figure 1.

Firstly, there are no data for death rates in the age group 110+: these fields are populated with '.'. In total, this was the case for 444 fields in the data set where a death rate would be (out of 12 960 total), with the following split:

---

[1] Source: *HMD*, The Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org. Data downloaded on 5 January 2023.

| Age | Female | Male | Total |
|---|---|---|---|
| 105-109 | 0 | 26 | 0 |
| 110+ | 130 | 160 | 128 |

It seems reasonable to assume that there were no data in these fields because, in that year and for that sex, there was no-one in that age category to begin with, and so it would be impossible to produce a death rate because the attempt would involve division by zero. Apart from this, there were no other missing data in the data set.

The second striking aspect of the data shown in
Figure 1 is that the death rate shown for males aged 105-109 is, implausibly, over 1. Taken at face value, that seems to be saying that more people in this category died than there were to begin with, which is incoherent. Death rates over 1 appeared 271 times in the data set, with the following distribution:

| Age | Female | Male | Total |
|---|---|---|---|
| 105-109 | 61 | 82 | 69 |
| 110+ | 22 | 14 | 23 |

I understand that the statistical methods used to calculate these death rates[2] are quite sophisticated and so these results may not be as incoherent as they seem. Nevertheless, the doubts that these figures raised, along with the null values in the same age groups, made me decide to simply drop all the rows for the age categories 105-109 and 110+. It seems reasonable to assume that whatever trends emerge for the age categories up to 100-104 should carry over to even older age categories.

I carried out two other pieces of data cleaning. First, I cast the columns 'Female', 'Male' and 'Total' to `float`. They had been of type `str`, presumably because of the presence of instances of '.' originally. Second, I multiplied all the death rates by 1 000. This was so as to be able to refer to deaths per 1 000 population, which is the more commonly-used way of referring to death rates, and also because I judged that this would make the visualisations easier for viewers to get a handle on quickly.

## DATA STORIES AND VISUALISATIONS

### Infant and Childhood Mortality

The plots in Figure 2 show infant (aged under 1) and childhood (aged 1-4 inclusive) mortality over time.

---

[2] 'Methods Protocol for the Human Mortality Database', Version 6. Downloaded from https://www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf, 6 January 2023.
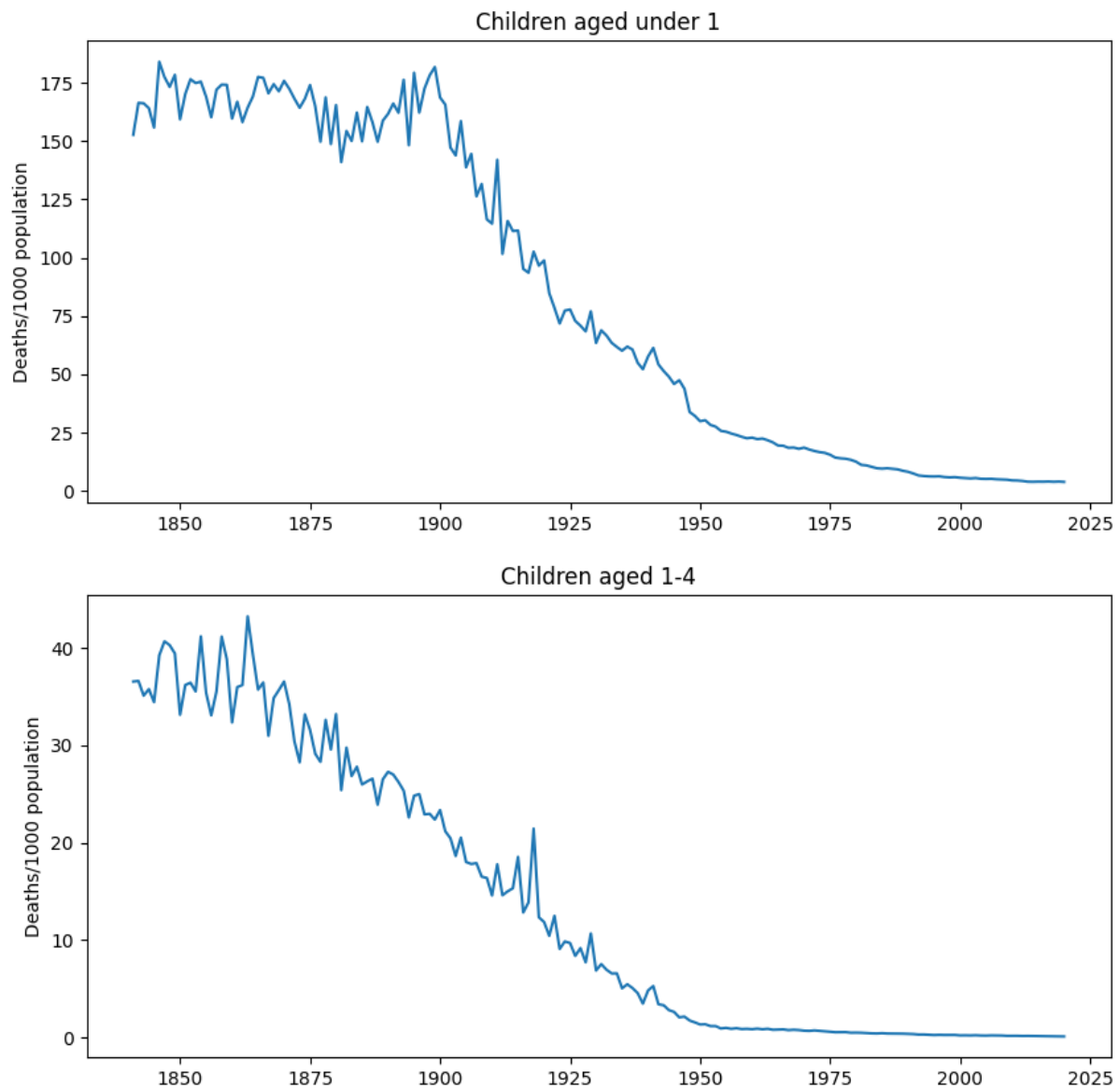
*Figure 2: Infant and childhood mortality*

For both age groups, mortality has been on a downward curve for some time, which may be flattening off now. However, the downward curve for the age group 1-4 began much earlier – possibly even earlier than the beginning of the data set. On the other hand, the mortality rate for infants was flat for some time, and the downward curve seems to have begun some time around the year 1900. That should prompt an investigation into what social, economic or health changes took place around that time that could have begun this period of improvement.

## The Impact of Major Events

The bar plot in Figure 3 has been chosen to illustrate just how starkly the effect of the two World Wars (especially the First World War) can be seen in the data.
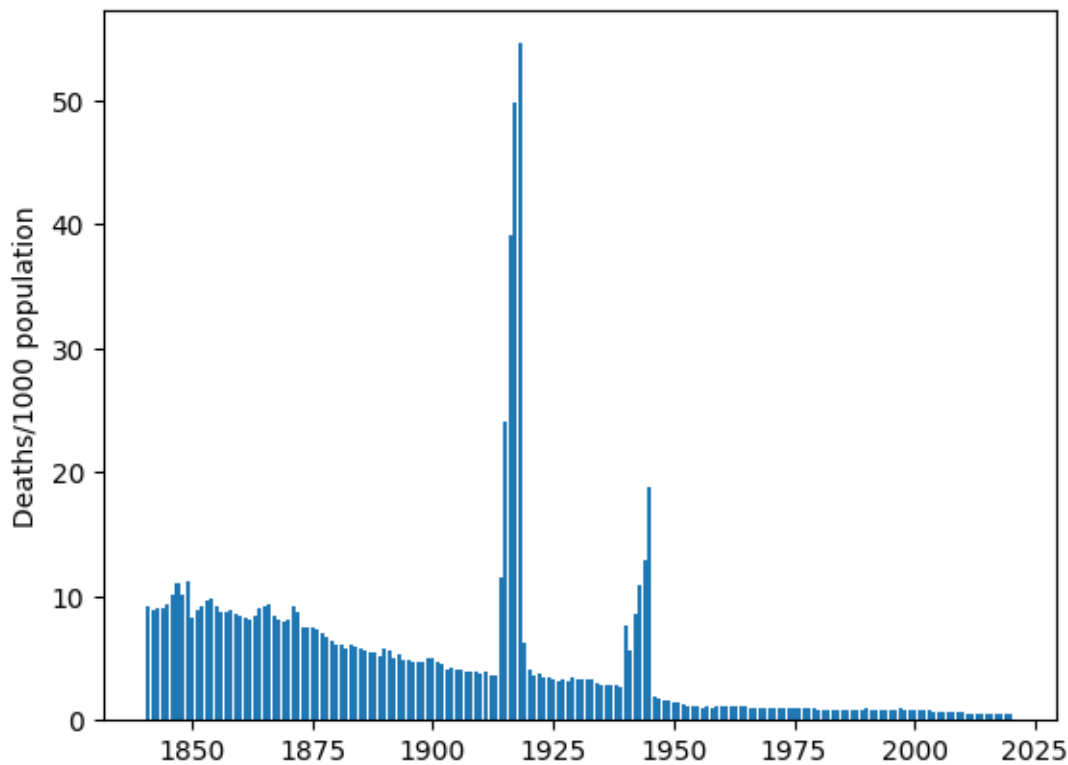
*Figure 3: Mortality rates of men aged 20-24*

The aim of this exercise is to determine how significantly the different age/sex groups were affected by the following four major events that fall within the period covered by the data set:

1. The First World War (1914-1918).
2. The Spanish Flu pandemic (1918-1920).
3. The Second World War (1939-1945).
4. The COVID-19 pandemic (2020).

'Significance' in the sense of this exercise was measured by the mortality rate, for each age/sex group, in comparison with the distribution of mortality rates for the same group over the five years preceding the event in question. However, since the First World War and the Spanish Flu pandemic overlapped, the comparison years for the First World War and Spanish Flu were the same. Overall, then, the comparisons made were as shown in the table below.

| Event | Year | Compared to |
|---|---|---|
| First World War | 1914 | 1909-1913 |
| First World War | 1915 | 1909-1913 |
| First World War | 1916 | 1909-1913 |
| First World War | 1917 | 1909-1913 |
| First World War / Spanish Flu | 1918 | 1909-1913 |
| Spanish Flu | 1919 | 1909-1913 |
| Spanish Flu | 1920 | 1909-1913 |

| | | |
|---|---|---|
| Second World War | 1939 | 1934-1938 |
| Second World War | 1940 | 1934-1938 |
| Second World War | 1941 | 1934-1938 |
| Second World War | 1942 | 1934-1938 |
| Second World War | 1943 | 1934-1938 |
| Second World War | 1944 | 1934-1938 |
| Second World War | 1945 | 1934-1938 |
| COVID-19 | 2020 | 2014-2019 |

The respective comparisons are shown in the following figures. In each case, each cell in the crosstab shows, for that age/sex group, the Z-score for the death rate when compared to the distribution for the comparison years. So, for example, by this measure the deadliest cell is 1918 for males aged 20-24 (Figure 5). The death rate for men aged 20-24 in 1918 was almost 350 standard deviations above the mean death rate for men aged 20-24 between 1909 and 1913. The cells have been colour-coded using Pandas.Styler: from a Z-score of 2 the cells get progressively darker shades of grey; from a Z-score of 50 they get a progressively darker shade of red.

### The First World War and Spanish Flu

Figure 4 and Figure 5 show the results for the WW1 and Spanish Flu years (combined), for females and males, respectively. We can see that WW1 was incredibly deadly for boys and men in the age brackets from 15 up to 39, and got ever-deadlier as the war progressed. Some of the effect in 1918 is undoubtedly due to the Spanish Flu as well; 1918 was also very deadly for women in these same age brackets, and also for younger boys and girls. This effect persists into 1919, and by 1920 things seem to be back to the pre-war normal.

| Year | 1914 | 1915 | 1916 | 1917 | 1918 | 1919 | 1920 |
|---|---|---|---|---|---|---|---|
| **Age** | | | | | | | |
| **a_0** | -0.508798 | -0.609534 | -1.771049 | -1.828635 | -1.168957 | -1.714826 | -1.549846 |
| **b_1-4** | -0.244858 | 2.028706 | -2.261143 | -1.472372 | 5.071684 | -2.665207 | -3.032978 |
| **c_5-9** | 0.927526 | 3.553937 | -0.314696 | -0.196389 | 15.603488 | 1.921303 | 0.016563 |
| **d_10-14** | 2.226914 | 4.456841 | 2.784396 | 2.965201 | 29.438056 | 4.396573 | -0.003013 |
| **e_15-19** | 0.346448 | 4.514246 | 4.071417 | 6.741412 | 43.965053 | 9.528627 | 0.385521 |
| **f_20-24** | 0.029555 | 2.824952 | 2.541718 | 2.849581 | 58.104849 | 13.353870 | 1.421096 |
| **g_25-29** | 0.003070 | 1.269597 | 0.786014 | -0.012281 | 49.405289 | 12.299894 | 2.006485 |
| **h_30-34** | -0.833899 | -0.316554 | -1.540767 | -2.319346 | 24.971897 | 4.821043 | -1.371733 |
| **i_35-39** | -0.341240 | 0.520935 | -2.964068 | -5.695803 | 22.873969 | -0.522751 | -7.819478 |
| **j_40-44** | -0.497611 | -0.321582 | -1.627352 | -2.171204 | 3.314607 | -2.307824 | -4.112782 |
| **k_45-49** | -0.277674 | 0.351372 | -3.184685 | -4.543275 | 3.191385 | -4.267835 | -8.239385 |
| **l_50-54** | 0.341440 | 0.967069 | -1.166928 | -2.503890 | 1.488142 | -2.519317 | -4.859000 |
| **m_55-59** | -1.096070 | 0.563930 | -2.156374 | -3.379247 | -1.874590 | -3.514720 | -7.261010 |
| **n_60-64** | -0.270542 | 0.840780 | -1.014098 | -1.801116 | -1.451405 | -1.953572 | -3.809787 |
| **o_65-69** | -0.500394 | 0.713147 | -0.091784 | -0.814916 | -0.517291 | -0.779969 | -2.544596 |
| **p_70-74** | 0.242845 | 4.773751 | 0.736022 | -2.260483 | -4.464288 | -2.952480 | -7.884900 |
| **q_75-79** | -0.150244 | 3.124918 | 1.605774 | 1.098619 | -1.112121 | 1.165899 | -2.462369 |
| **r_80-84** | -0.474838 | 2.429936 | 1.547696 | 0.154309 | -2.210093 | 0.404231 | -2.196997 |
| **s_85-89** | -0.016183 | 6.252476 | 5.197927 | 3.458465 | -2.377111 | 3.582330 | -2.761717 |
| **t_90-94** | 0.431181 | 2.384720 | 2.033369 | 1.358006 | -0.887915 | 1.733404 | -1.736073 |
| **u_95-99** | 0.108184 | 1.884634 | 1.738192 | 1.443643 | -0.286875 | 1.589891 | -0.699756 |
| **v_100-104** | 0.616496 | -0.522339 | 0.867368 | -0.078549 | -1.093473 | 0.469042 | -1.461196 |

*Figure 4: Female death rates by age group compared to the 1909-1913 average (standardised scores)*

| Year | 1914 | 1915 | 1916 | 1917 | 1918 | 1919 | 1920 |
|---|---|---|---|---|---|---|---|
| **Age** | | | | | | | |
| **a_0** | -0.496684 | -0.383195 | -1.720637 | -1.904277 | -1.192110 | -1.575551 | -1.405318 |
| **b_1-4** | -0.316323 | 2.574515 | -2.291742 | -1.421855 | 4.198465 | -2.682353 | -3.087342 |
| **c_5-9** | 1.220335 | 3.859383 | 0.679300 | 0.504966 | 13.880554 | 1.857554 | 0.631208 |
| **d_10-14** | 1.592311 | 5.252056 | 3.364398 | 2.940638 | 19.955246 | 4.096347 | -0.243983 |
| **e_15-19** | 36.635927 | 72.557168 | 112.186407 | 140.624367 | 142.186811 | 7.839129 | 0.871679 |
| **f_20-24** | 53.361389 | 138.737697 | 241.673573 | 315.311579 | 347.865529 | 16.846789 | 2.556980 |
| **g_25-29** | 30.370537 | 85.638388 | 159.976575 | 224.742225 | 284.108143 | 23.307911 | 0.208017 |
| **h_30-34** | 12.785503 | 37.999237 | 71.912510 | 107.175235 | 153.974668 | 15.130452 | -1.068500 |
| **i_35-39** | 11.377844 | 33.561006 | 57.100150 | 86.125262 | 124.837228 | 9.552142 | -5.868833 |
| **j_40-44** | 0.658371 | 3.299248 | 4.566295 | 7.367350 | 13.293922 | -0.021357 | -3.091428 |
| **k_45-49** | -0.077623 | 2.519476 | 0.009431 | 0.578906 | 9.077512 | -4.306976 | -8.909934 |
| **l_50-54** | 0.538968 | 1.612827 | 0.002822 | -0.922108 | 1.902847 | -2.817430 | -5.125051 |
| **m_55-59** | 0.047316 | 3.853889 | -0.555963 | -1.495186 | -0.574889 | -5.273369 | -10.674491 |
| **n_60-64** | -0.063470 | 3.433563 | -0.021088 | -0.422166 | -0.855290 | -3.018840 | -6.273985 |
| **o_65-69** | -0.402437 | 1.929229 | 1.327529 | 1.076464 | 0.768589 | -0.166644 | -3.354130 |
| **p_70-74** | 0.421534 | 5.180574 | 2.979559 | 2.743736 | 0.396423 | -0.655501 | -5.946781 |
| **q_75-79** | -0.920806 | 5.041483 | 4.844103 | 4.293184 | 0.306446 | 3.753095 | -3.258281 |
| **r_80-84** | -0.140505 | 2.969379 | 2.208778 | 2.021136 | -1.116092 | 1.269566 | -2.087037 |
| **s_85-89** | 0.858789 | 8.904931 | 7.736921 | 6.436199 | -1.541504 | 6.239415 | -3.932673 |
| **t_90-94** | 0.354465 | 2.029598 | 2.137355 | 2.148925 | -1.817917 | 1.512410 | -0.526876 |
| **u_95-99** | -0.410017 | 1.171468 | 0.331084 | 0.782586 | -0.282215 | 0.153828 | -0.596448 |
| **v_100-104** | -0.634324 | -0.486394 | -1.355570 | 1.499150 | 0.430891 | -1.833266 | -1.934602 |

*Figure 5: Male death rates by age group compared to the 1909-1913 average (standardised scores)*

The absolute devastation of WW1 on male mortality can also be starkly illustrated by the plot shown in Figure 6.
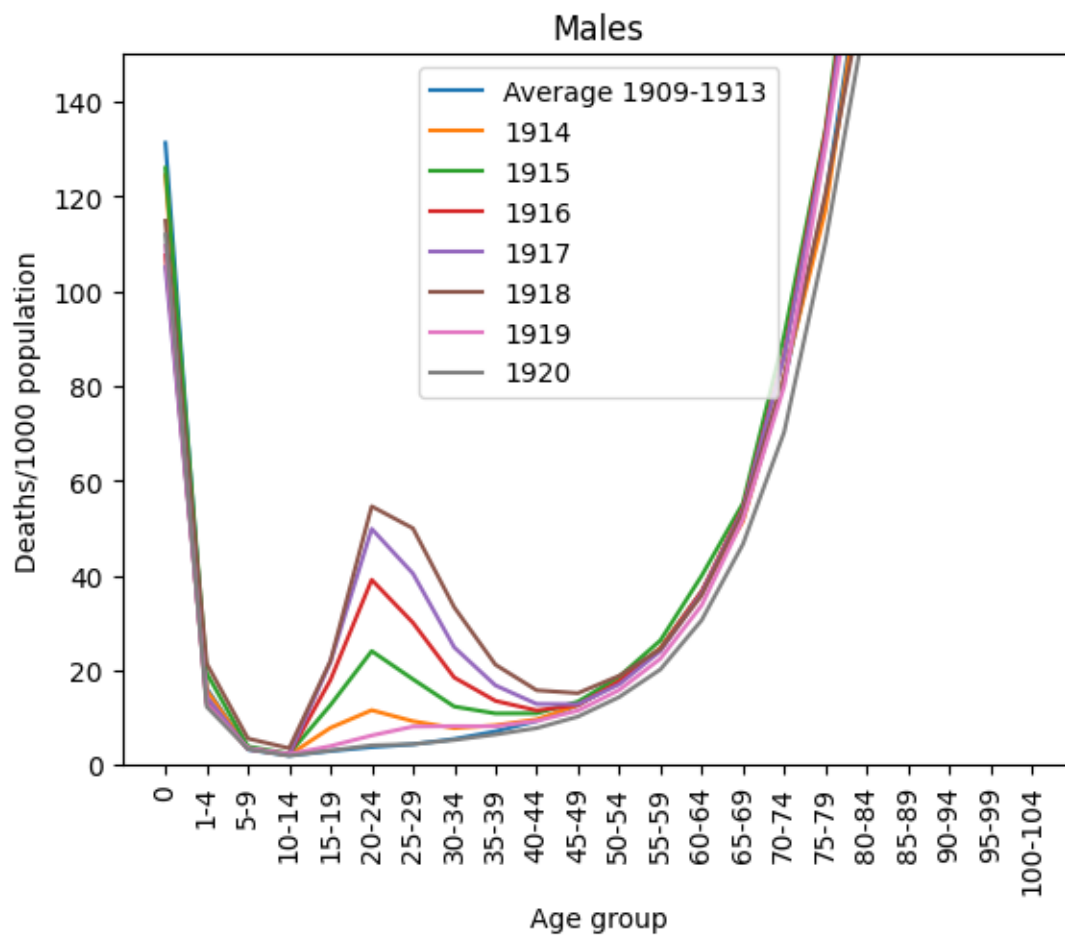


*Figure 6: Comparison of male death rates, 1909-1913 vs. 1914-1920 (raw scores)*

### The Second World War

The corresponding crosstabs for the Second World War years, this time compared to 1934-1938, are shown in Figure 7 and Figure 8 for females and males respectively.

| Year | 1939 | 1940 | 1941 | 1942 | 1943 | 1944 | 1945 |
|------|------|------|------|------|------|------|------|
| **Age** | | | | | | | |
| a_0 | -2.692817 | -1.011614 | 0.642854 | -2.008678 | -2.962353 | -3.812597 | -5.110752 |
| b_1-4 | -2.704884 | -0.857515 | -0.092518 | -2.875549 | -2.836626 | -3.692943 | -3.926484 |
| c_5-9 | -2.814687 | -0.637972 | -0.407913 | -2.827960 | -3.389835 | -3.425229 | -4.155225 |
| d_10-14 | -1.884516 | 0.778726 | 0.202469 | -2.305028 | -1.830005 | -2.297240 | -3.294009 |
| e_15-19 | -1.631887 | 2.951587 | 2.469116 | -1.631887 | -1.944074 | -2.724542 | -3.228298 |
| f_20-24 | -2.183866 | 2.769999 | 1.795590 | -2.049978 | -3.016949 | -2.749172 | -3.009510 |
| g_25-29 | -1.423106 | 1.375669 | 0.895371 | -1.440895 | -2.152448 | -2.644606 | -3.415455 |
| h_30-34 | -1.873515 | 0.671282 | -0.129858 | -2.852686 | -2.067255 | -2.056782 | -3.695715 |
| i_35-39 | -2.442601 | 0.874309 | -1.234962 | -3.639312 | -3.305981 | -4.360617 | -5.070993 |
| j_40-44 | -2.092332 | 1.435560 | -0.670565 | -4.145405 | -3.402691 | -4.559203 | -5.779376 |
| k_45-49 | -1.816437 | 1.920771 | -0.388259 | -3.471242 | -3.608074 | -4.518858 | -5.694753 |
| l_50-54 | -1.408415 | 2.292100 | -1.241989 | -4.694498 | -4.116904 | -5.046928 | -5.797473 |
| m_55-59 | -1.494751 | 2.682777 | -1.579952 | -4.561999 | -4.663176 | -6.122249 | -6.479030 |
| n_60-64 | -1.246591 | 2.431091 | -2.572785 | -5.693243 | -5.395526 | -6.250467 | -6.303006 |
| o_65-69 | -0.626804 | 2.974030 | -1.361958 | -4.582052 | -4.054483 | -5.427579 | -5.697946 |
| p_70-74 | -0.377342 | 2.608496 | -1.456862 | -4.347929 | -3.277124 | -4.776579 | -4.442701 |
| q_75-79 | 0.702037 | 4.090541 | -1.220521 | -4.837538 | -3.567932 | -5.810573 | -5.321374 |
| r_80-84 | 1.002400 | 3.243167 | -0.401960 | -3.009056 | -1.262599 | -3.008299 | -2.922159 |
| s_85-89 | 1.190757 | 3.690296 | 0.906333 | -2.225016 | -0.513219 | -2.828101 | -2.676400 |
| t_90-94 | 0.829323 | 2.846735 | 0.732331 | -1.895685 | -0.143717 | -1.836676 | -1.537314 |
| u_95-99 | 1.660683 | 3.241705 | 1.373398 | -0.606772 | 0.118509 | -0.714800 | -1.217296 |
| v_100-104 | 0.765754 | 2.638822 | 1.272347 | -0.220734 | 1.404898 | 1.057109 | -2.219613 |

*Figure 7: Female death rates by age group compared to the 1934-1938 average (standardised scores)*

| Year | 1939 | 1940 | 1941 | 1942 | 1943 | 1944 | 1945 |
|------|------|------|------|------|------|------|------|
| **Age** | | | | | | | |
| a_0 | -3.252506 | -0.774827 | 0.514783 | -2.385336 | -3.589680 | -4.616061 | -5.776533 |
| b_1-4 | -2.764156 | -0.676919 | -0.099201 | -2.808596 | -3.109640 | -3.736098 | -4.034274 |
| c_5-9 | -2.547216 | -0.067115 | 0.577816 | -2.400403 | -2.751707 | -2.914250 | -4.078272 |
| d_10-14 | -2.759440 | 1.816915 | 1.987251 | -1.794204 | -2.543681 | -1.601157 | -3.259092 |
| e_15-19 | -1.863467 | 16.682465 | 13.142864 | 27.439302 | 45.521831 | 59.581639 | 66.798875 |
| f_20-24 | -1.647754 | 47.200704 | 27.711680 | 57.222191 | 79.974493 | 101.286522 | 159.995317 |
| g_25-29 | -1.591087 | 23.900816 | 15.164367 | 32.773487 | 50.981989 | 60.735571 | 100.757949 |
| h_30-34 | -2.042625 | 11.682310 | 8.788256 | 15.697727 | 21.922294 | 22.224457 | 38.453992 |
| i_35-39 | -2.570170 | 6.327860 | 4.198131 | 4.092202 | 5.541755 | 4.359812 | 11.089086 |
| j_40-44 | -2.296980 | 5.183739 | 2.644141 | -0.679527 | -0.220920 | -1.103616 | -0.477346 |
| k_45-49 | -1.950498 | 3.991259 | 0.468961 | -2.836162 | -2.568087 | -3.467325 | -4.122242 |
| l_50-54 | -0.468330 | 6.286958 | 0.624073 | -3.949161 | -2.248644 | -3.938154 | -4.648079 |
| m_55-59 | 0.650876 | 7.844974 | 2.649450 | -1.778424 | -0.983972 | -2.119177 | -2.735595 |
| n_60-64 | 0.547104 | 5.243898 | 1.213527 | -1.919381 | -1.620569 | -1.290951 | -1.604139 |
| o_65-69 | -0.131450 | 7.267435 | 1.133979 | -2.735909 | -2.805636 | -3.368623 | -3.932901 |
| p_70-74 | -0.000113 | 3.074745 | -1.731174 | -4.822476 | -4.389286 | -5.351490 | -5.459788 |
| q_75-79 | 1.352374 | 4.209758 | -0.189731 | -3.776351 | -3.213762 | -5.124343 | -5.253032 |
| r_80-84 | 1.542973 | 4.772629 | 0.283471 | -3.552227 | -2.508396 | -3.881869 | -3.816564 |
| s_85-89 | 1.183233 | 4.127366 | -0.036169 | -2.799917 | -0.396133 | -1.893633 | -2.864791 |
| t_90-94 | 1.471013 | 2.151764 | -0.452604 | -1.591825 | -0.095721 | -1.270609 | -1.305247 |
| u_95-99 | 1.138239 | 3.124521 | 0.417138 | -0.444185 | 0.257684 | 0.790840 | -0.408780 |
| v_100-104 | 1.050526 | 0.763146 | 0.496710 | -0.777205 | 0.003358 | 0.143913 | -0.083682 |

*Figure 8: Male death rates by age group compared to the 1934-1938 average (standardised scores)*

We see that WW2 was deadly for men in the same age bracket as WW1 (i.e., fighting age). We also see that, just like WW1, it got deadlier as it went on.

One unexpected finding of this section is that there were excess deaths in almost all adult age categories, for both men and women, in 1940 specifically. The cause of this needs investigation.

## COVID-19

The corresponding comparison of 2020 to the 2015-2019 average is shown in Figure 9 and Figure 10 for females and males respectively.

| Year Age | 2020 | | Year Age | 2020 |
|---|---|---|---|---|
| a_0 | -0.815313 | | a_0 | -1.687005 |
| b_1-4 | -2.368749 | | b_1-4 | -1.410711 |
| c_5-9 | -4.407413 | | c_5-9 | -0.162221 |
| d_10-14 | -0.503740 | | d_10-14 | -2.235879 |
| e_15-19 | -2.113067 | | e_15-19 | -1.892445 |
| f_20-24 | -1.399769 | | f_20-24 | -1.445274 |
| g_25-29 | -0.573288 | | g_25-29 | -1.252502 |
| h_30-34 | 1.520360 | | h_30-34 | 0.243109 |
| i_35-39 | 4.781023 | | i_35-39 | 5.959585 |
| j_40-44 | 7.240596 | | j_40-44 | 2.350742 |
| k_45-49 | 5.288306 | | k_45-49 | 6.647341 |
| l_50-54 | 7.032352 | | l_50-54 | 8.645110 |
| m_55-59 | 1.835121 | | m_55-59 | 19.425635 |
| n_60-64 | 4.284064 | | n_60-64 | 3.950455 |
| o_65-69 | 3.125926 | | o_65-69 | 9.936493 |
| p_70-74 | 2.341903 | | p_70-74 | 3.687102 |
| q_75-79 | 2.656917 | | q_75-79 | 3.399905 |
| r_80-84 | 2.298083 | | r_80-84 | 4.243291 |
| s_85-89 | 2.454991 | | s_85-89 | 4.755565 |
| t_90-94 | 1.657361 | | t_90-94 | 1.558748 |
| u_95-99 | 2.252201 | | u_95-99 | 2.305190 |
| v_100-104 | 0.771755 | | v_100-104 | 0.617599 |

*Figure 9: Female death rates by age group compared to the 2015-2019 average (standardised scores)*

*Figure 10: Male death rates by age group compared to the 2015-2019 average (standardised scores)*

We see the effect of COVID-19 in excess deaths in almost every age from 35-39 up.

**THIS REPORT WAS WRITTEN BY: MATTHEW GOTHAM**