**First playing with demography R package**

**First lessons**

* The package requires very specific data inputs.
  + It is designed to take the .txt files from <https://www.mortality.org/Home/Index> and conduct life expectancy calculations.
  + Consequently, with our data, step one will be to manipulate the data into a) death rates by age, sex and year and b) ‘exposure-to-risk’ by age, sex and year.
    - Exposure-to-risk, as I understand it – and going by Matthew’s paper too, is simply the number of people alive and resident in that age, sex and year. But this needs more checking. Just one of those painful academic things where people feel a need to replace a word like ‘population size’ with ‘exposures’…
  + But this should all be quite achievable given DoB + DoD.
    - There is going to be some mess with people moving and so on but think a rough first go is do-able.
* Matthew then did this for a) everyone b) migrants and c) native-born.
  + The outputs in his paper are then seemingly quite achievable.
  + Just need to double check the differences in his figures are simply (everyone minus native-born)
* Matthew uses life expectancy at age 1 not age 0 because of something about migrant vs native children not being comparable for first 12 months.
* The default plots from the package are so so ugly – I will extract the numbers and plot in ggplot so they don’t look so gross.

**Questions I have**

* Does our data allow for annual estimates? Or are we simply doing 5 snapshots 71,81,91,01,11?
  + I think yearly estimates can be done given year of death and year of residency.
* What additional complications are we interested in testing with migration status?
  + Anything other than binary var? Anything about age and sex interesting to us?
* It seems from google that LS deaths were updated in 2017 – I have my fingers crossed that means we have deaths updated until then so we can zoom in on post-austerity.
  + But not so convinced based on the data info I was given.

**First efforts**

* I downloaded UK death rates from <https://www.mortality.org/Home/Index>
* I read them into the R package Matthew uses.
* The package lets us plot the data – where the colour represents year (red=1922, purple = 2020)
  + A picture containing screenshot, plot, diagram, line

    Description automatically generated
* Or we can plot the death rate by year – (red = age 1, purple = age 110+)
  + A picture containing text, screenshot, line, plot

    Description automatically generated
  + You can see WWII in there and death rate at aged 0 is similar to that at aged 50ish
* Then it is very simple to calculate life expectancy
  + A picture containing text, line, diagram, plot

    Description automatically generated
* And plot it over time
  + A picture containing diagram, text, line, plot

    Description automatically generated
* Next, I wanted to practice comparing two populations…
  + So, I took the GBR data and divided the death rate by a random number (to simulate a random second population which dies less often than the GBR average.) This is obviously scuffy but works to give us two populations – one dying more than another.
  + Then I can plot them using the same package options.
  + So I made these two datasets – one is slightly healthier than the other
  + A picture containing text, line, diagram, plot

    Description automatically generatedA picture containing text, line, diagram, plot

    Description automatically generated
* The last thing I wanted to practice was extracting the data and plotting my own pretty graphs. To compare life expectancy over time of two different populations.
  + Not too shabby
  + A picture containing text, line, diagram, plot

    Description automatically generated

**Feelings I have**

* This is good fun, new to me so useful and exciting, but not too hard.
* Seems almost too easy – need to keep feet on the ground, I will have to read about the stats at some point.
* My mind is stimulated by what kind of policy stuff can be answered with this data and methods – Brexit/experiences of racism/NHS charges to migrants(!)/austerity etc. etc.
  + Could be an amazing paper looking at diff-in-diff of LE between migrant and native population before and after NHS charges increased/introduced in 2015?
    - Tories killed people by charging for nhs… not certain it works but if it does could be an amazing second paper…