Untitled

Jon\_Minton

Thursday, July 31, 2014

print("HELLO!")

## [1] "HELLO!"

setwd("E:/repos/Population\_Age\_Residuals/")  
source("Script.R")

## PACKAGE INSTALLATION REPORT:  
## The package xlsx was already installed correctly.  
## The package plyr was already installed correctly.  
## The package reshape2 was already installed correctly.  
## The package lattice was already installed correctly.  
## The package repmis was already installed correctly.  
## The package RCurl was already installed correctly.  
## The package devtools was already installed correctly.  
## The package httr was already installed correctly.  
## The package digest was already installed correctly.  
## The package ggplot2 was already installed correctly.  
## The package stringr was already installed correctly.  
## The package car was already installed correctly.

The first suggestion was:

Maybe draw a single overview graph, with lines again, and with the Y axis still being death rates - but the X axis being year 1990-2008. An a separate line for each single year of age and sex.

In other words just one line for women aged 25 on the graph with the height of points being death rates and the X position be year. It should work with so many lines as they tend to be parallel. lines will bend quickly down towards 2008 as mortality rates of the youngest age groups drop quickly in recent years - quicker than makes much sense unless there is a problem with the data.

To do this we first have to join exp\_eu\_all with rates\_eu\_all

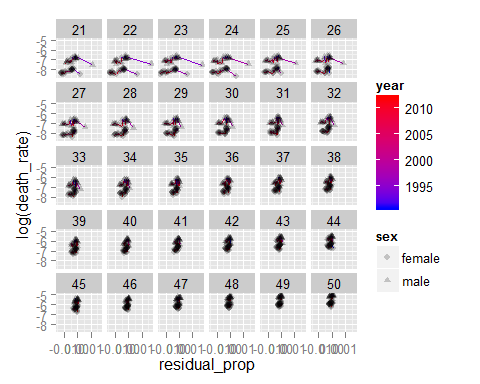
mrate\_resp <- join(  
rates\_eu\_all,  
exp\_eu\_all,   
by=c("year", "age", "sex"),  
type="inner"  
)  
  
mrate\_resp <- arrange(mrate\_resp, age, year, sex)

We want to group the data into discrete age groups, as we assume age will affect the amount of residual.

mrate\_resp$age\_group <- recode(  
 mrate\_resp$age,  
 recodes="  
 1:4='4 or under';  
 5:9 = '5-9';  
 10:14 = '10-14';  
 15:19 = '15-19';  
 20:24 = '20-24';  
 25:29 = '25-29';  
 30:34 = '30-34';  
 35:39 = '35-39';  
 40:44 = '40-44';  
 45:49 = '45-49';  
 else = '50 or older'  
 ",  
 as.factor.result=T,  
 levels=c(  
 '4 or under',  
 '5-9',  
 '10-14',  
 '15-19',  
 '20-24',  
 '25-29',  
 '30-34',  
 '35-39',  
 '40-44',  
 '45-49',  
 '50 or older'  
 )   
)  
  
# cut ages into groups   
  
mrate\_resp$age\_group <- ordered(  
 mrate\_resp$age\_group,   
 levels=c(  
 '4 or under',  
 '5-9',  
 '10-14',  
 '15-19',  
 '20-24',  
 '25-29',  
 '30-34',  
 '35-39',  
 '40-44',  
 '45-49',  
 '50 or older'  
 )  
)

Our first graph. This shows, for males and females aged between 20 and 50 years, the relationship between log death rate and the size of the residual over time, since 1990.

The year is indicated by colour, with the earliest year represented as a strong blue and the latest year represented as a strong red.



In this graph each age is shown in a different tile. There are too many age groups. Instead let's now look at the same but with fewer distinct age group categories.

