Assignment Intro HDS - Infant Death Scotland

Sophie Charlotte Zeiz

2024-11-18

Infant mortality in Scotland

For this assignment, I wanted to look into the data from the **Scottish Puclic Health Observatory** regarding infant mortality. "The infant mortality rate is defined as the number of children who die before reaching their first birthday in a given year, expressed per 1 000 live births." OECD/World Health Organization (2018). Ending preventable deaths of newborns and children under the age of five is one of the World Health Organisations sustainable development goals 1.

The dataset used for this assignment was extracted from the ScotPHO Inline Profiles Tool and is freely available for the public.

```
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(here)
```

```
## here() starts at C:/Users/charl/OneDrive/Documentos/University of Aberdeen/Introduction to Health Da
```

```
library(readr)
infant_mortality_all_geo <- read_csv("ScotPHO_infant death all geo.csv")
```

```
## Rows: 1248 Columns: 11
## -- Column specification ---------------------------------------------------------
## Delimiter: ","
## chr (6): area_code, area_type, area_name, period, type_definition, indicator
## dbl (5): year, numerator, measure, upper_confidence_interval, lower_confiden...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(infant_mortality_all_geo)

mortality_allgeo <- infant_mortality_all_geo %>%
  select(`area_code`, `area_type`, `area_name`, `year`,`measure`) %>%
  rename(Year = `year`,
         Mortality_rate = `measure`
         )
head(mortality_allgeo)
```

```
## # A tibble: 6 x 5
##   area_code area_type area_name   Year Mortality_rate
##   <chr>     <chr>     <chr>      <dbl>          <dbl>
## 1 S00000001 Scotland  Scotland    2004            4.9
## 2 S00000001 Scotland  Scotland    2005            4.8
## 3 S00000001 Scotland  Scotland    2006            4.6
## 4 S00000001 Scotland  Scotland    2007            4.4
## 5 S00000001 Scotland  Scotland    2008            4.1
## 6 S00000001 Scotland  Scotland    2009            4.1
```

Evolution of the infant mortality rate in Scotland

```
mortality_scotland_total <- mortality_allgeo %>%
  select(`area_name`, `Year`,`Mortality_rate`) %>%
  filter(`area_name` == "Scotland")
```

```
library(ggplot2)
mortality_scotland_total %>%
  ggplot() +
  geom_line(aes(x = `Year`, y = `Mortality_rate`))
```
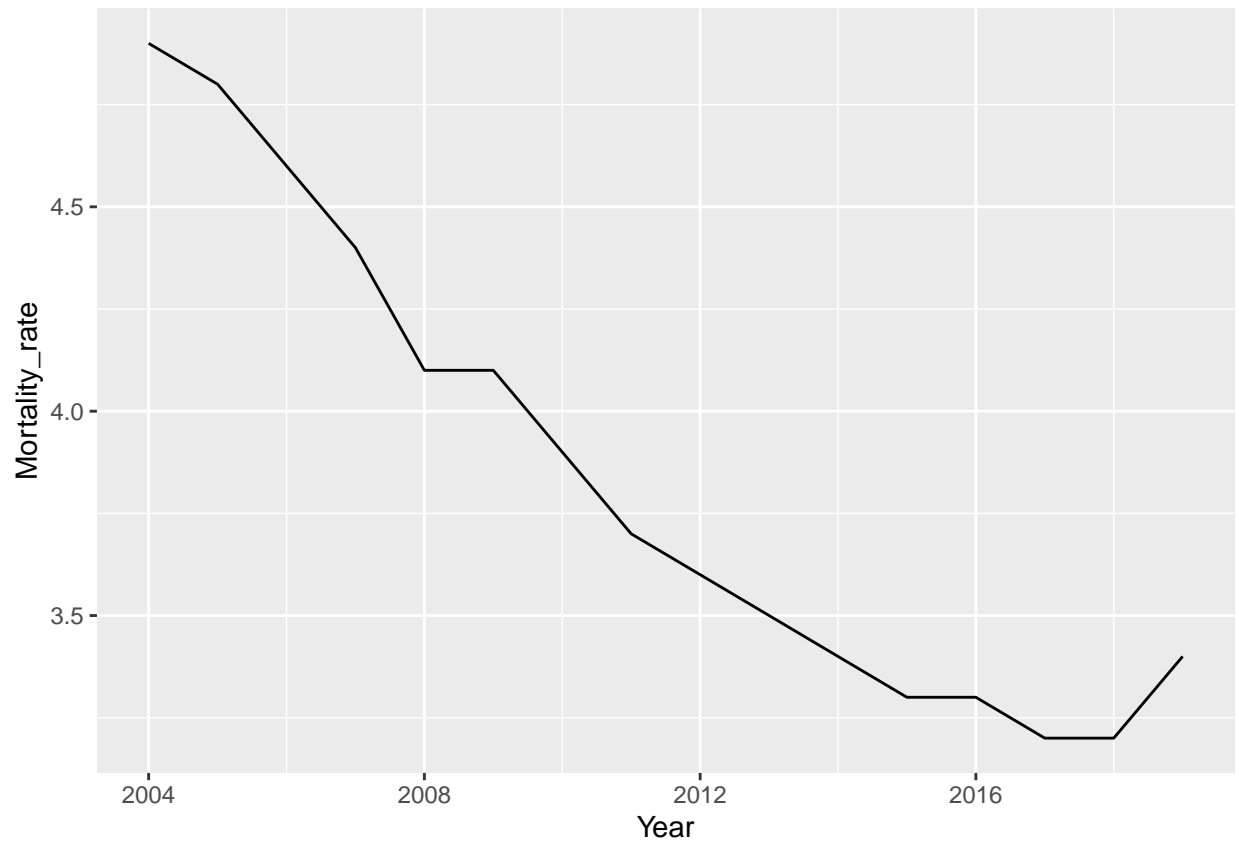
Our available data shows a downward trend in the mortality rate in Scotland between the years 2004 and 2019. In 2004 the mo

##Trends in infant mortality rate per NHS health board in Scotland

```
mortality_healthboards <- mortality_allgeo %>%
  select(`area_type`,`area_name`, `Year`,`Mortality_rate`) %>%
  filter(`area_type` == "Health board")
```

```
library(ggplot2)
mortality_healthboards %>%
  select(`area_name`, `Year`,`Mortality_rate`) %>%
  rename( NHS_board = `area_name`) %>%
  group_by(`NHS_board`) %>%
  ggplot() +
  geom_line(aes(x = `Year`, y = `Mortality_rate`, colour = `NHS_board`))
```
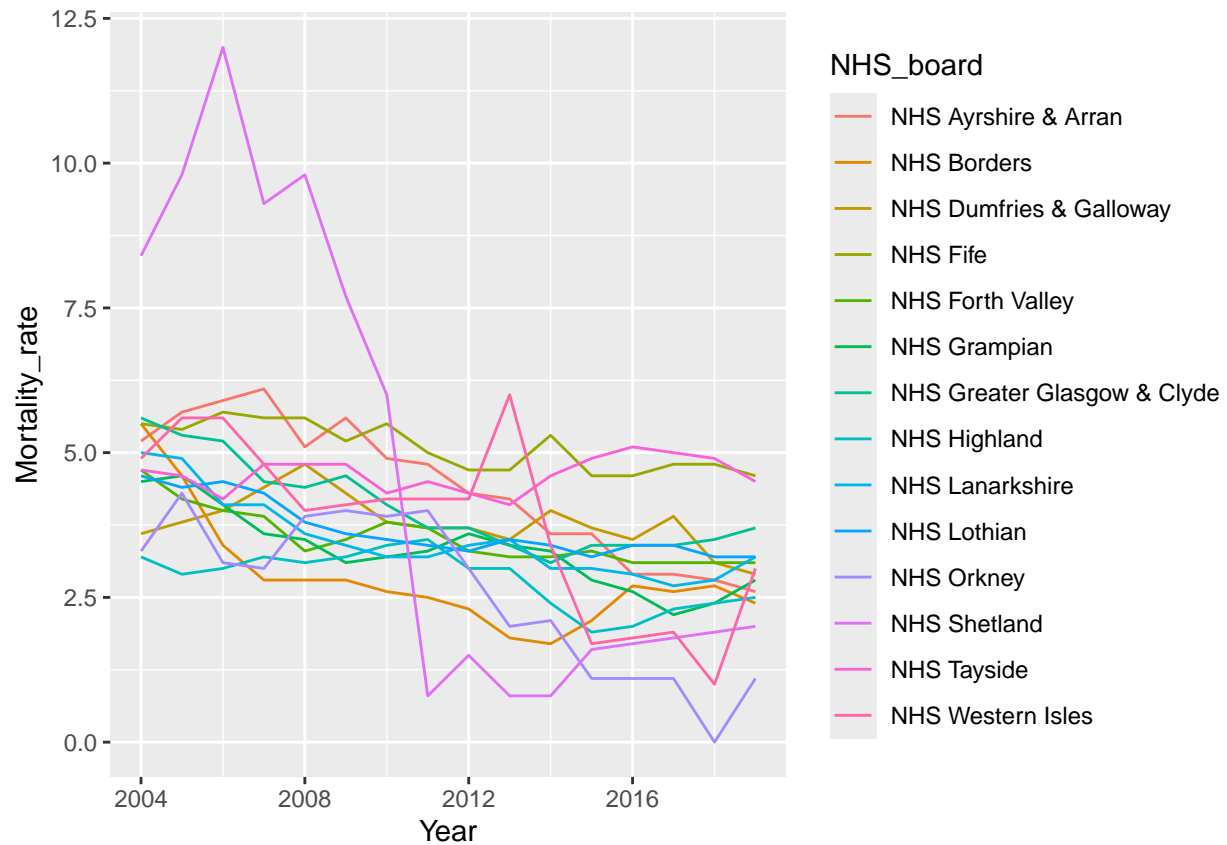
```r
big4_healthboards <- mortality_healthboards %>%
  select(`area_name`, `Year`, `Mortality_rate`) %>%
  filter(`area_name` %in% c("NHS Greater Glasgow & Clyde", "NHS Grampian", "NHS Lothian", "NHS Tayside")
```

```r
library(ggplot2)
big4_healthboards %>%
  select(`area_name`, `Year`,`Mortality_rate`) %>%
  rename( NHS_board = `area_name`) %>%
  group_by(`NHS_board`) %>%
  ggplot() +
  geom_line(aes(x = `Year`, y = `Mortality_rate`, colour = `NHS_board`))
```
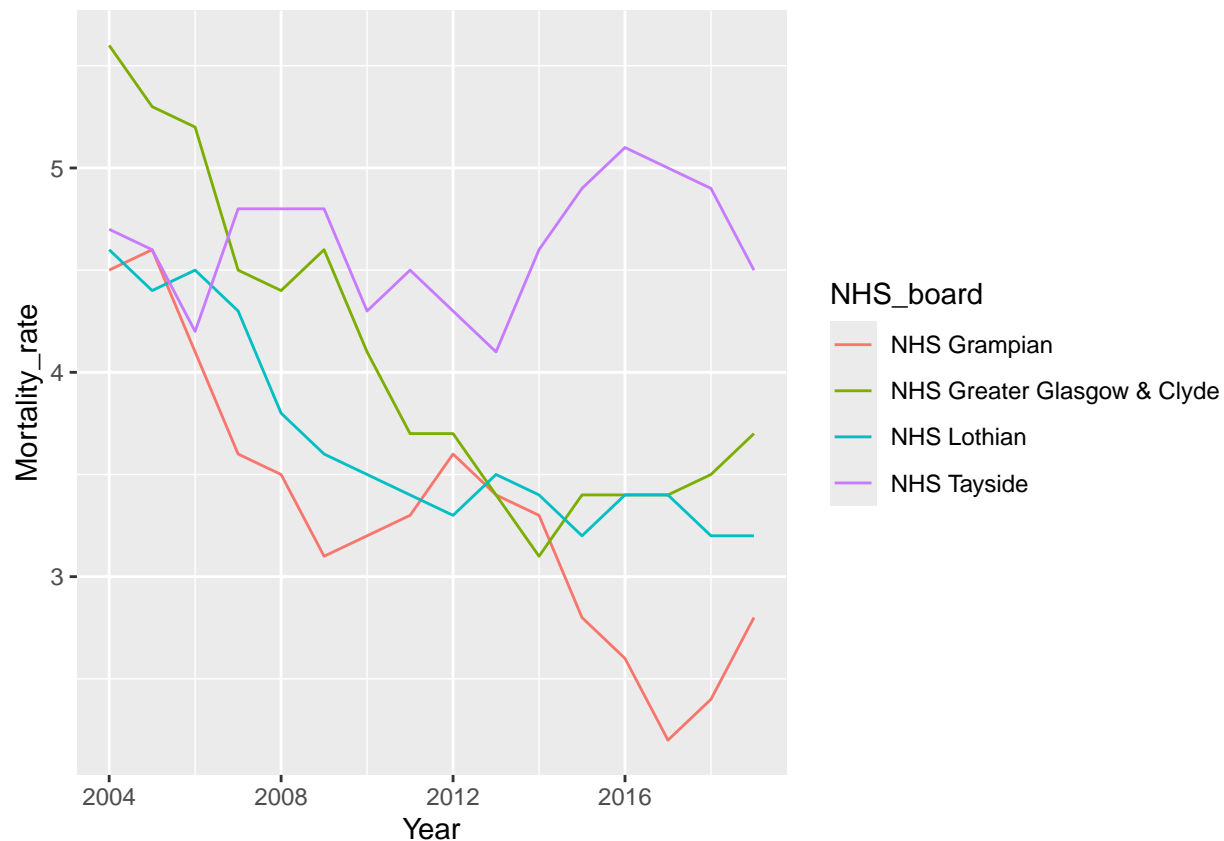
Enter urban rural data from 2011

```
library(readr)
rural_class_complete <- read_csv("scottish-government-urban-rural-classification-2020-data-zone-2011-loc
```

```
## Rows: 6976 Columns: 5
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (1): DataZone
## dbl (4): UR2FOLD, UR3FOLD, UR6FOLD, UR8FOLD
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
View(rural_class_complete)
```

```
rural_3fold <- rural_class_complete %>%
  select(`DataZone`, `UR6FOLD`) %>%
  rename(area_code = `DataZone`,
         Classification = `UR6FOLD`
         )
```

Enter council areas to classification data

```r
library(readr)
area_to_hsc <- read_csv("dz2011_codes_and_labels_21042020.csv") %>%
  rename(area_code = `DataZone`)
```

```
## Rows: 6976 Columns: 11
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (11): DataZone, DataZoneName, IntZone, IntZoneName, CA, CAName, HSCP, HS...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
rural3fold_CA <- rural_3fold %>%
  inner_join(`area_to_hsc`, by = "area_code") %>%
  select(area_code, Classification, CA, HSCP, HB)
```

Combine council area, mortality rates and rural-urban classification

```r
combined_mortalityallgeo_rural3fold <- mortality_allgeo %>%
  inner_join(`rural3fold_CA`, by = c("area_code" = "CA")) %>%
  select(area_type, area_name, Year, Mortality_rate, Classification)
```

```
## Warning in inner_join(., rural3fold_CA, by = c(area_code = "CA")): Detected an unexpected many-to-mar
## i Row 241 of `x` matches multiple rows in `y`.
## i Row 904 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```r
library(dplyr)
get_mode <- function(x) {
  uniq_x <- unique(x)
  uniq_x[which.max(tabulate(match(x, uniq_x)))]
}
```

```r
AVG_mort_class_percouncil_peryear <- combined_mortalityallgeo_rural3fold %>%
  group_by(area_name, Year) %>%
  summarise(
    avg_mortality_rate = mean(Mortality_rate, na.rm = TRUE),
    mode_classification = get_mode(Classification)
  )
```
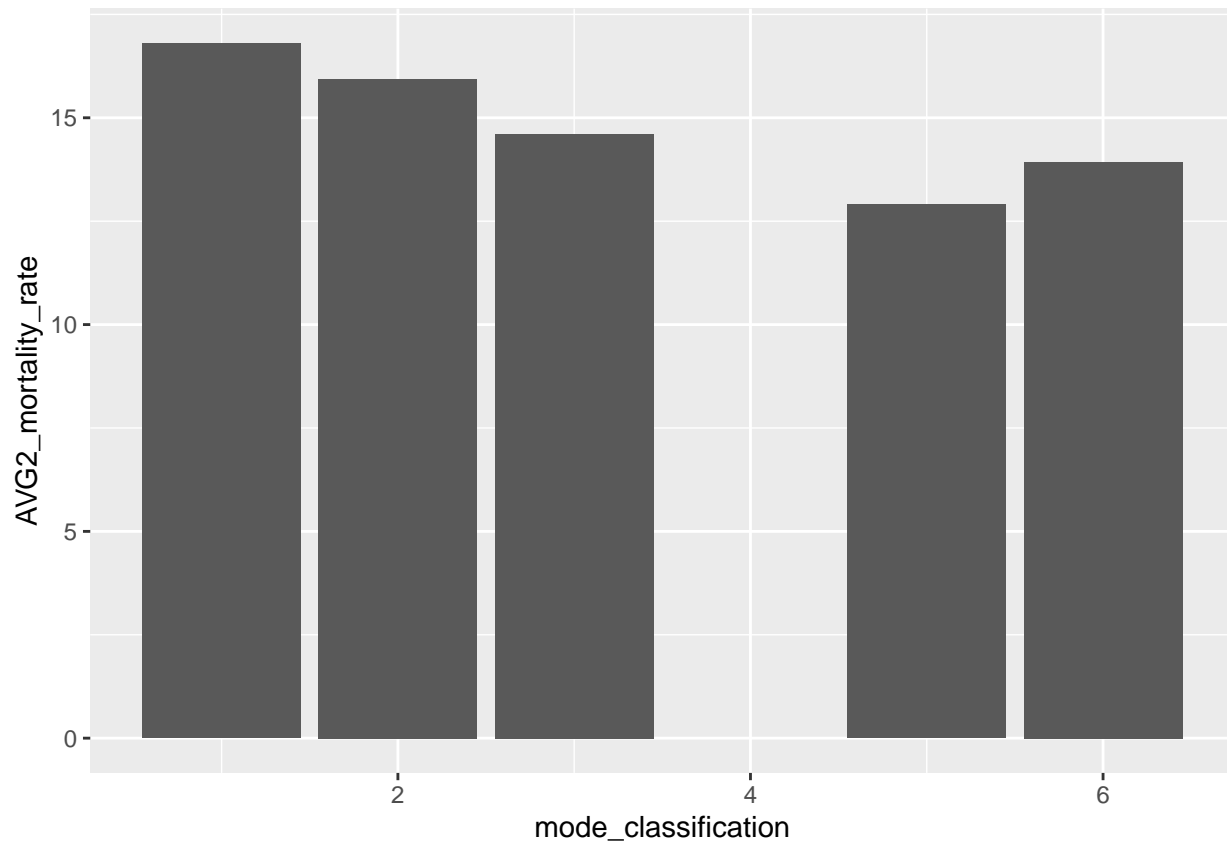
```
## `summarise()` has grouped output by 'area_name'. You can override using the
## `.groups` argument.
```

```r
group_classification <- AVG_mort_class_percouncil_peryear %>%
  mutate(mode_classification = case_when(
    mode_classification == 1 ~ "Large Urban Areas",
    mode_classification == 2 ~ "Other Urban Areas",
    mode_classification == 3 ~ "Accessible Small Towns",
    mode_classification == 4 ~ "Remote Small Towns",
    mode_classification == 5 ~ "Accessible Rural",
    mode_classification == 6 ~ "Remote Rural",
    TRUE ~ "Unknown"  # In case there are other values that don't fit the categories
  ))
```

```r
avg_class_mort_year <- AVG_mort_class_percouncil_peryear %>%
  select(`Year`,`avg_mortality_rate`, `mode_classification`) %>%
  group_by(mode_classification, Year) %>%
  summarise(AVG2_mortality_rate = mean(avg_mortality_rate, na.rm = TRUE))
```

```
## Adding missing grouping variables: 'area_name'
## 'summarise()' has grouped output by 'mode_classification'. You can override
## using the '.groups' argument.
```

```r
library(ggplot2)
avg_class_mort_year %>%
  select(`Year`, `AVG2_mortality_rate`, `mode_classification`) %>%
  filter(`Year` %in% c("2004", "2008", "2012", "2016")) %>%
  ggplot(aes(x = `mode_classification`, y = `AVG2_mortality_rate`, group = `Year`)) +
  geom_col()
```

```
avg_class_mort_year <- group_classification %>%
  select(`Year`,`avg_mortality_rate`, `mode_classification`) %>%
  group_by(mode_classification, Year) %>%
  summarise(avg_mortalityrate = mean(avg_mortality_rate, na.rm = TRUE))
```

```
## Adding missing grouping variables: 'area_name'
## 'summarise()' has grouped output by 'mode_classification'. You can override
## using the '.groups' argument.
```

```
library(ggplot2)
avg_class_mort_year %>%
  select(`Year`,`avg_mortalityrate`, `mode_classification`) %>%
  group_by(mode_classification) %>%
  ggplot() +
  geom_line(aes(x= Year, y= avg_mortalityrate, colour = mode_classification))
```