Domestic abuse in Scotland

Alifya Mukadam

Aim

To create data science workflow to study the trend of domestic abuse incidents and crimes in Scotland from 2003 to 2021, as well as the prevalence of domestic abuse across NHS health boards in 2021.

Load packages

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.1      v tibble    3.2.1
## v lubridate 1.9.3      v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(viridis)
```

```
## Loading required package: viridisLite
```

```r
library(patchwork)
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
```

Data acquisition

Domestic abuse data sets contain information on the number of domestic abuse incidents and crimes recorded by Scotland police, and the prevalence (Crude rate per 10,000 population) recorded from 2003 to 2021 across Scotland and NHS Scotland health boards. The data sets were downloaded from the Scottish Government (Scottish Crime Statistics) available on Scottish Public Health Observatory

```
library(here)
```

```
## here() starts at /Users/alifyamukadam/Documents/GitHub/R_Project/R_report
```

```
DA_Scot_data <- read_csv(here("inputs/Domestic-abuse-data_Scotland.csv"))
```

```
## Rows: 19 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.
```

```
DA_Scot_HB_data <- read_csv(here("inputs/Scoland-HB-data.csv"))
```

```
## Rows: 266 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.
```

Prepare the data

```
glimpse(DA_Scot_data) %>%
summary()
```

```
## Rows: 19
## Columns: 11
## $ area_code                <chr> "S00000001", "S00000001", "S00000001", "S000~
## $ area_type                <chr> "Scotland", "Scotland", "Scotland", "Scotlan~
## $ area_name                <chr> "Scotland", "Scotland", "Scotland", "Scotlan~
## $ year                     <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 20~
## $ period                   <chr> "2003/04 financial year", "2004/05 financial~
## $ type_definition          <chr> "Crude rate per 10,000 population", "Crude r~
## $ indicator                <chr> "Domestic abuse", "Domestic abuse", "Domesti~
## $ numerator                <dbl> 41235, 43633, 45331, 48884, 49949, 53931, 51~
## $ measure                  <dbl> 81.4, 85.8, 88.7, 95.2, 96.6, 103.7, 99.2, 1~
## $ upper_confidence_interval <dbl> 82.1, 86.6, 89.5, 96.1, 97.5, 104.5, 100.1, ~
## $ lower_confidence_interval <dbl> 80.6, 85.0, 87.9, 94.4, 95.8, 102.8, 98.4, 1~
```

```
##   area_code          area_type          area_name              year
## Length:19          Length:19          Length:19          Min.   :2003
## Class :character   Class :character   Class :character   1st Qu.:2008
## Mode  :character   Mode  :character   Mode  :character   Median :2012
##                                                          Mean   :2012
##                                                          3rd Qu.:2016
##                                                          Max.   :2021
##     period          type_definition     indicator          numerator
## Length:19          Length:19          Length:19          Min.   :41235
## Class :character   Class :character   Class :character   1st Qu.:50938
## Mode  :character   Mode  :character   Mode  :character   Median :58439
##                                                          Mean   :55731
##                                                          3rd Qu.:59981
##                                                          Max.   :65251
##     measure        upper_confidence_interval lower_confidence_interval
## Min.   : 81.4   Min.   : 82.1             Min.   : 80.6
## 1st Qu.: 97.9   1st Qu.: 98.8             1st Qu.: 97.1
## Median :108.8   Median :109.7             Median :107.9
## Mean   :105.0   Mean   :105.9             Mean   :104.1
## 3rd Qu.:112.5   3rd Qu.:113.3             3rd Qu.:111.5
## Max.   :119.4   Max.   :120.3             Max.   :118.5
```

```r
glimpse(DA_Scot_HB_data) %>%
  summary()
```

```
## Rows: 266
## Columns: 11
## $ area_code              <chr> "S08000015", "S08000015", "S08000015", "S080~
## $ area_type              <chr> "Health board", "Health board", "Health boar~
## $ area_name              <chr> "NHS Ayrshire & Arran", "NHS Ayrshire & Arra~
## $ year                   <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 20~
## $ period                 <chr> "2003/04 financial year", "2004/05 financial~
## $ type_definition        <chr> "Crude rate per 10,000 population", "Crude r~
## $ indicator              <chr> "Domestic abuse", "Domestic abuse", "Domesti~
## $ numerator              <dbl> 2589, 3213, 3171, 3679, 3868, 3996, 4251, 44~
## $ measure                <dbl> 70.4, 87.2, 85.9, 99.6, 104.3, 107.4, 114.1,~
## $ upper_confidence_interval <dbl> 73.2, 90.2, 89.0, 102.9, 107.7, 110.8, 117.6~
## $ lower_confidence_interval <dbl> 67.7, 84.2, 83.0, 96.4, 101.1, 104.1, 110.7,~
```

```
##   area_code          area_type          area_name              year
## Length:266         Length:266         Length:266         Min.   :2003
## Class :character   Class :character   Class :character   1st Qu.:2007
## Mode  :character   Mode  :character   Mode  :character   Median :2012
##                                                          Mean   :2012
##                                                          3rd Qu.:2017
##                                                          Max.   :2021
##     period          type_definition     indicator          numerator
## Length:266         Length:266         Length:266         Min.   :   21.0
## Class :character   Class :character   Class :character   1st Qu.:  858.8
## Mode  :character   Mode  :character   Mode  :character   Median : 3485.0
##                                                          Mean   : 3980.8
##                                                          3rd Qu.: 4982.5
##                                                          Max.   :17412.0
```

```
##      measure       upper_confidence_interval lower_confidence_interval
## Min.   : 10.10    Min.   : 15.50             Min.   :  6.30
## 1st Qu.: 60.38    1st Qu.: 64.97             1st Qu.: 54.70
## Median : 91.20    Median : 94.20             Median : 88.20
## Mean   : 86.82    Mean   : 91.28             Mean   : 82.72
## 3rd Qu.:114.40    3rd Qu.:118.08             3rd Qu.:111.65
## Max.   :153.40    Max.   :155.70             Max.   :151.10
```

Data cleaning

Select and tidy the data to plot domestic abuse trend in Scotland

```
DA_trend_data <- DA_Scot_data %>%
  select('area_name','year','measure') %>%
  rename('Area'= 'area_name',
         'Year' = 'year',
         'Prevalence' = 'measure'
         )

head(DA_trend_data)
```

```
## # A tibble: 6 x 3
##   Area       Year Prevalence
##   <chr>     <dbl>      <dbl>
## 1 Scotland   2003       81.4
## 2 Scotland   2004       85.8
## 3 Scotland   2005       88.7
## 4 Scotland   2006       95.2
## 5 Scotland   2007       96.6
## 6 Scotland   2008      104.
```

Select and tidy the data to plot domestic abuse across Scotland NHS health boards in 2021

```
DA_2021_HB_data <- DA_Scot_HB_data %>%
  filter(year == "2021") %>%
  select('area_name','measure') %>%
  mutate(area_name = gsub("NHS", "", area_name)) %>%
  rename('NHS_Health_Board' = 'area_name',
         'Prevalence' = 'measure') %>%
  arrange(desc(Prevalence))

head(DA_2021_HB_data)
```

```
## # A tibble: 6 x 2
##   NHS_Health_Board       Prevalence
##   <chr>                       <dbl>
## 1 " Fife"                      144.
## 2 " Forth Valley"              130.
## 3 " Lanarkshire"               129.
```

```
## 4 " Greater Glasgow & Clyde"        126.
## 5 " Tayside"                         123.
## 6 " Ayrshire & Arran"                122.
```

```
summary(DA_2021_HB_data)
```

```
##  NHS_Health_Board      Prevalence
##  Length:14          Min.    : 40.40
##  Class :character   1st Qu.: 89.95
##  Mode  :character   Median :116.00
##                     Mean    :103.73
##                     3rd Qu.:124.97
##                     Max.    :143.50
```

Categorisation

```
DA_2021_HB_categorised_data <- DA_2021_HB_data %>%
  mutate(Category = case_when(
    Prevalence <= 89.95 ~ "Low",
    Prevalence > 89.95 & Prevalence <= 124.97 ~ "Medium",
    Prevalence > 124.97 ~ "High"))

head(DA_2021_HB_categorised_data)
```

```
## # A tibble: 6 x 3
##   NHS_Health_Board          Prevalence Category
##   <chr>                          <dbl> <chr>
## 1 " Fife"                         144. High
## 2 " Forth Valley"                 130. High
## 3 " Lanarkshire"                  129. High
## 4 " Greater Glasgow & Clyde"      126. High
## 5 " Tayside"                      123. Medium
## 6 " Ayrshire & Arran"             122. Medium
```
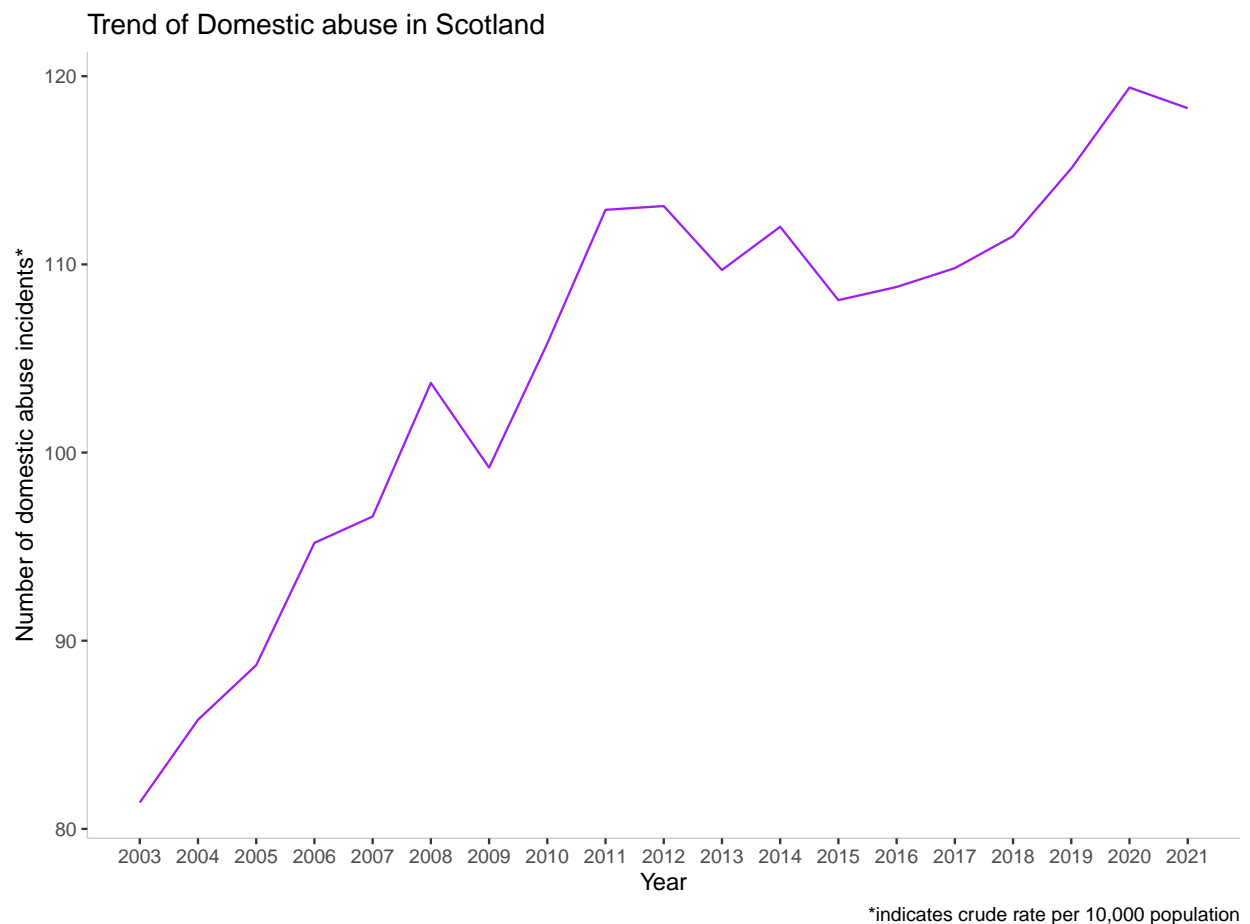
Data visualisation
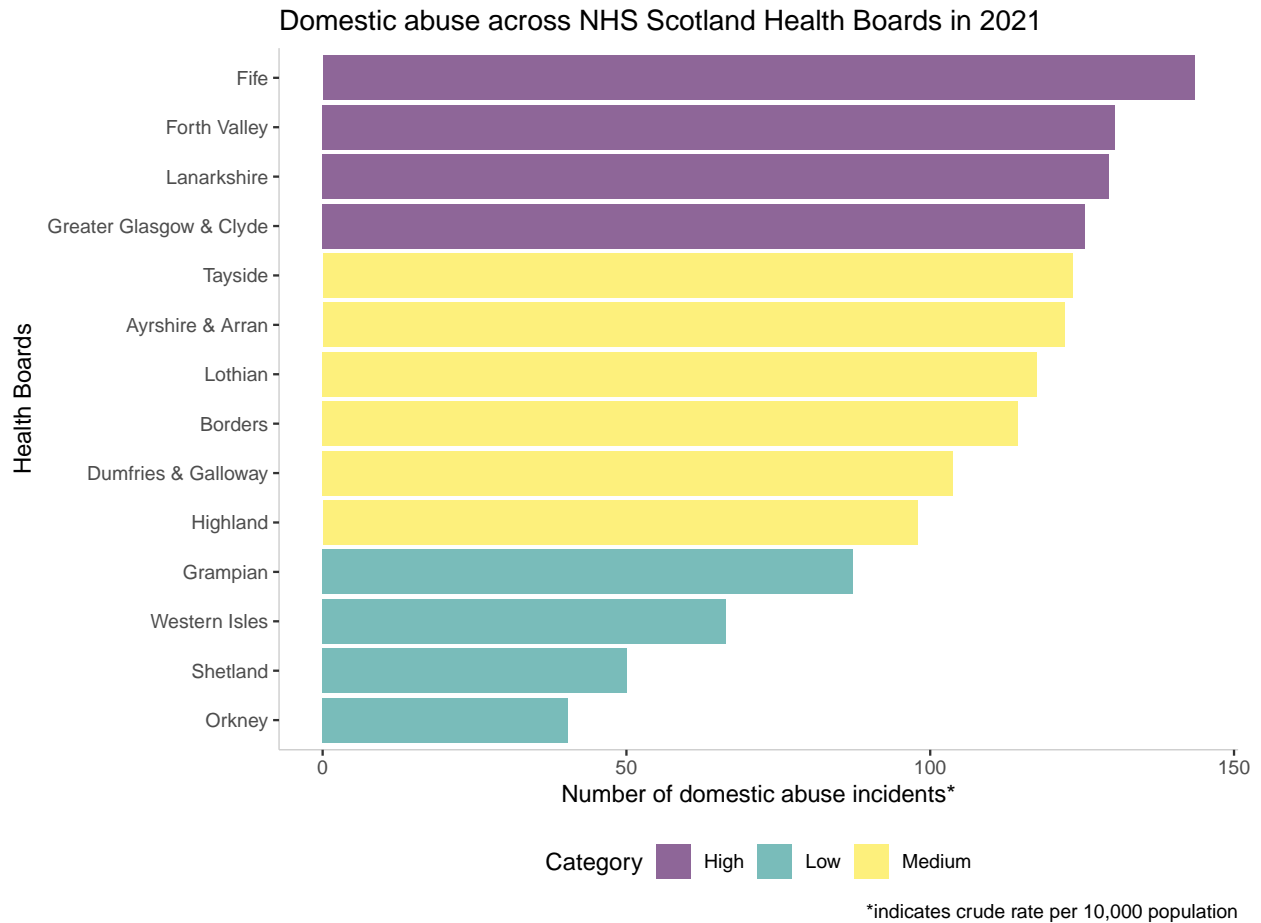
```
p1 <- DA_trend_data %>%
  ggplot(aes(x = Year,
             y = Prevalence,
             group = 1)) +
  geom_line(colour = "purple") +
  ggtitle("Trend of Domestic abuse in Scotland") +
  xlab("Year")+
  ylab("Number of domestic abuse incidents*") +
  labs(caption = "*indicates crude rate per 10,000 population") +
  scale_x_continuous (breaks = seq(2003, 2021, by=1) ) +
  theme(panel.background = element_rect(fill = NA),
        axis.line = element_line(linewidth = 0.3, colour = "grey80"))
```

```
p2 <- DA_2021_HB_categorised_data %>%
  ggplot(aes(x = reorder(NHS_Health_Board, Prevalence),
             y = Prevalence,
             fill = Category,
             group = 1)) +
  geom_col(alpha = 0.6) +
  coord_flip() +
  ggtitle("Domestic abuse across NHS Scotland Health Boards in 2021") +
  xlab("Health Boards")+
  ylab("Number of domestic abuse incidents*") +
  labs(caption = "*indicates crude rate per 10,000 population") +
  scale_fill_viridis(discrete=TRUE,) +
  theme(panel.background = element_rect(fill = NA),
        axis.line = element_line(linewidth = 0.3, colour = "grey80"),
        legend.position = "bottom")

p1
```
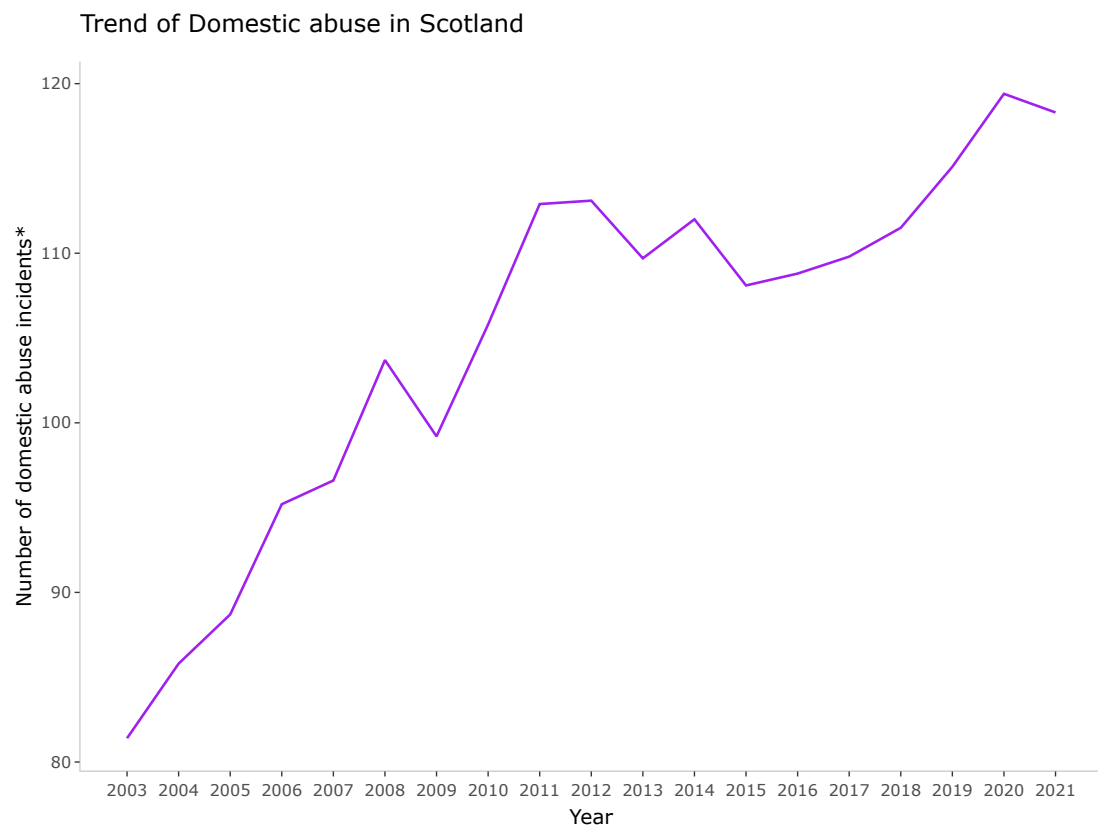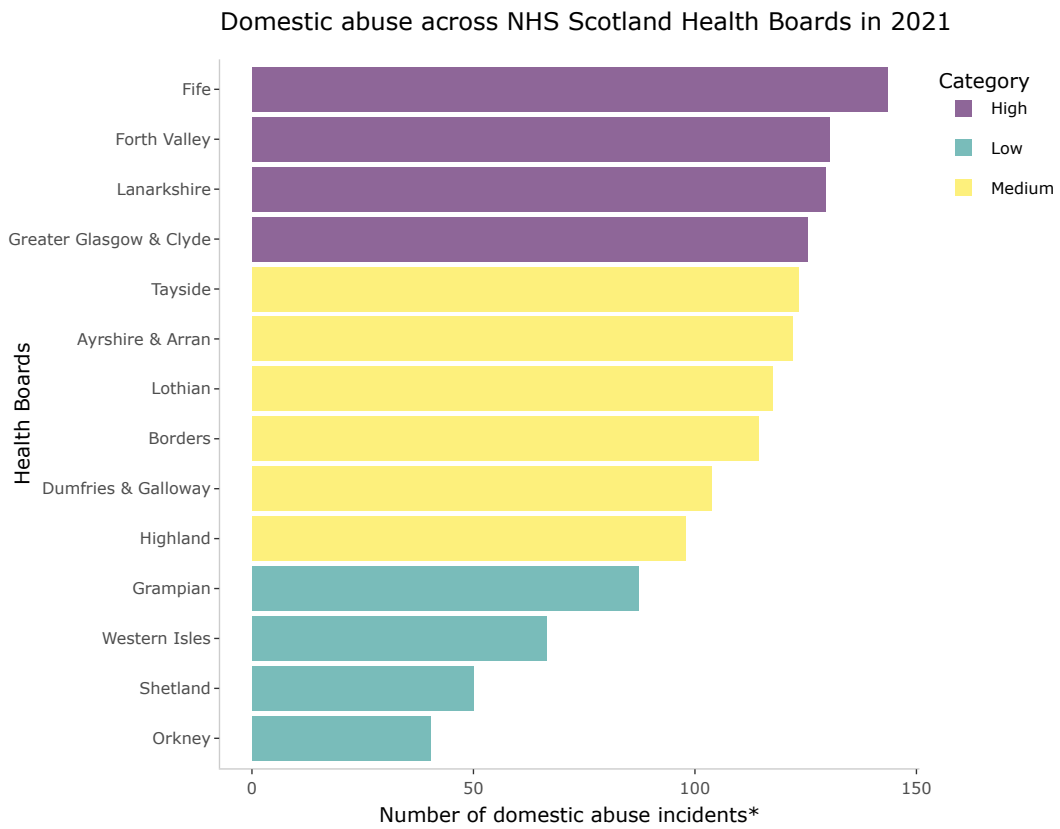

Trend of Domestic abuse in Scotland

*indicates crude rate per 10,000 population

p2

Domestic abuse across NHS Scotland Health Boards in 2021

Fife
Forth Valley
Lanarkshire
Greater Glasgow & Clyde
Tayside
Ayrshire & Arran
Lothian
Borders
Dumfries & Galloway
Highland
Grampian
Western Isles
Shetland
Orkney

Health Boards

0   50   100   150

Number of domestic abuse incidents*

Category   High   Low   Medium

*indicates crude rate per 10,000 population

Interactive graphs for html

```
ggplotly(p1, tooltip = c("x", "y"), width = 800, height = 600)
```

Trend of Domestic abuse in Scotland

```
ggplotly(p2, tooltip = c("x", "y"), width = 800, height = 600)
```

Domestic abuse across NHS Scotland Health Boards in 2021

Summary