Final report

S Piggott; student ID: 50095266

2024-12-03

Summary

The aim of this report is to answer the question: in Scotland, how do rates of depression symptoms in adults vary by sex and by deprivation status? The target audience is policymakers.

Data were acquired from the Scottish Public Health Observatory (ScotPHO) website and downloaded on 28 November 2024 from: https://scotland.shinyapps.io/ScotPHO_profiles_tool/. Data selected for download were: dataset = inequalities dataset; region = Scotland; indicator = depression symptoms in past week; time period = latest available year (2021).

Data were visualised in a diverging bar chart with female data on the left and male data on the right, % of people with depression symptoms in the past week on the x-axis and Scottish Index of Multiple Deprivation (SIMD) quintile on the y-axis. The aim of this visualisation approach was to provide an accessible and quick means for policymakers to see similarities and differences in the data for each sex.

Key limitations of the dataset are:

- The latest time period available was 2021, which is 3 years ago. Therefore, policymakers may question the validity of the data for future healthcare decision-making as we soon move into 2025.
- Data are presented overall for Scotland, but trends may vary across different health boards.
- The ScotPHO data source does not appear to provide detail about the type of confidence intervals e.g. if they are 95% confidence intervals.
- The data were acquired from the Scottish Health Survey, so survey errors must be considered. Information about the key sampling and non-sampling errors can be found here: https://www.gov.scot/publications/scottish-household-survey-2021-methodology-fieldwork-outcomes/pages/8/.

Key strengths of the approach taken for the data visualisation are:

- A relatively small number of data-points were selected and a simple graph format has been used, meaning the information can be digested quickly by the policymaker.
- Error bars are included, so the policymaker can see the degree of uncertainty in the data.
- Additional information about the methodology was included in the caption section of the graph. This helps to provide important context to the policymaker as they interpret the results.

Key limitations of the approach taken for the data visualisation are:

- The visualisation approach required manipulation of the data (transforming female % values to negatives) in order to create the diverging bars. This has the potential to cause confusion when someone is trying to reproduce the analysis and visualisation. To mitigate this, comments were included in the code chunks to assist with understanding the methodology.
- The x-axis limits are 40%, whereas the true limit of the x-axis variable is 100%. If a reader were to glance at the graph very quickly without reading the x-axis labels or the data labels, they could misinterpret this health problem to be more extreme than the reality.
- Some readers may find the caption section to be too information-dense.

Set-up file

Load packages

The tidyverse family of packages has been used.

```
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
```

Read in and inspect the data

```
data <- read.csv("Inputs/ScotPHO_data_extract_2024-11-28.csv")

glimpse(data)
```

```
## Rows: 27
## Columns: 12
## $ area_code                <chr> "S00000001", "S00000001", "S00000001", "S000~
## $ area_type                <chr> "Scotland", "Scotland", "Scotland", "Scotlan~
## $ area_name                <chr> "Scotland", "Scotland", "Scotland", "Scotlan~
## $ year                     <int> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 20~
## $ period                   <chr> "Survey year (2021)", "Survey year (2021)", ~
## $ indicator                <chr> "Depression symptoms in past week", "Depress~
## $ quintile                 <chr> "1 - most deprived", "2", "3", "4", "5 - lea~
## $ measure                  <chr> "Percentage", "Percentage", "Percentage", "P~
## $ value                    <dbl> 18.7, 15.4, 8.2, 7.2, 6.4, 10.2, 10.1, 14.1,~
## $ upper_confidence_interval <dbl> 29.5, 25.4, 13.2, 14.0, 14.3, 13.0, 20.8, 24~
## $ lower_confidence_interval <dbl> 11.2, 8.8, 5.0, 3.6, 2.7, 7.9, 4.6, 7.9, 8.6~
## $ label_inequality         <chr> "Adults", "Adults", "Adults", "Adults", "Adu~
```

Prepare data

Select the rows and columns of interest for this report

```
Scotland_2021_depression_data <- data %>%
  select(indicator, quintile, measure, value,
```

```
          upper_confidence_interval, lower_confidence_interval) %>%
#Select columns of interest
  slice(1:5, 7:11)
#Selecting the rows for male and female data.
#Sex did not appear to be included in the downloadable csv file,
#so the only option seemed to be to manually check this
#against the bar charts on the ScotPHO profiles tool
```

Add female and male labels

```
Scotland_2021_depression_data <- Scotland_2021_depression_data %>%
  mutate(Sex = c('Female','Female','Female','Female','Female',
                 'Male','Male','Male','Male','Male'))
#Sex did not appear to be included in the downloadable csv file,
#so the only option seemed to be to manually check this
#against the bar charts on the ScotPHO profiles tool.
```

Rename column headings for clarity

```
Scotland_2021_depression_data <- Scotland_2021_depression_data %>%
  rename('SIMD_quintile' = 'quintile')
```

Reorder columns for clarity

```
Scotland_2021_depression_data <- Scotland_2021_depression_data %>%
  relocate('Sex', 'SIMD_quintile', .before = 'indicator')

tibble(Scotland_2021_depression_data)
```

```
## # A tibble: 10 x 7
##     Sex    SIMD_quintile      indicator      measure value upper_confidence_int~1
##     <chr>  <chr>              <chr>          <chr>   <dbl>                  <dbl>
##  1 Female 1 - most deprived  Depression sy~ Percen~  18.7                   29.5
##  2 Female 2                  Depression sy~ Percen~  15.4                   25.4
##  3 Female 3                  Depression sy~ Percen~   8.2                   13.2
##  4 Female 4                  Depression sy~ Percen~   7.2                   14
##  5 Female 5 - least deprived Depression sy~ Percen~   6.4                   14.3
##  6 Male   1 - most deprived  Depression sy~ Percen~  10.1                   20.8
##  7 Male   2                  Depression sy~ Percen~  14.1                   24
##  8 Male   3                  Depression sy~ Percen~  19.6                   38.7
##  9 Male   4                  Depression sy~ Percen~   9.7                   18.6
## 10 Male   5 - least deprived Depression sy~ Percen~   8.8                   14.8
## # i abbreviated name: 1: upper_confidence_interval
## # i 1 more variable: lower_confidence_interval <dbl>
```

Save new CSV file

```
write_csv(Scotland_2021_depression_data, "Scotland_2021_depression_data.csv")
```

Manipulate the data

To produce the diverging bar chart, one group (in this case, females) will need to have the data values transformed to negatives.

```
Scotland_2021_depression_data_female_negative  <- Scotland_2021_depression_data %>%
  mutate(value = ifelse(Sex == 'Male',
                                value,
                                -1*value)) %>%
  #For female rows, multiply value by -1
  mutate(upper_confidence_interval = ifelse(Sex == 'Male',
                                upper_confidence_interval,
                                -1*upper_confidence_interval)) %>%
  #For female rows, multiply upper CI by -1
  mutate(lower_confidence_interval = ifelse(Sex == 'Male',
                                lower_confidence_interval,
                                -1*lower_confidence_interval))
  #For female rows, multiply lower CI by -1
```

Plot the data

```
breaks_values = c(seq(-40, 40, by = 10)) #Manually specifying breaks
#for the y-axis, which will be flipped to become the x-axis;
#limits of -40 and 40 and increments of 10

Scotland_2021_depression_data_female_negative %>%
  ggplot(aes(x = SIMD_quintile, y = value, fill = Sex)) +
#Plot % of people with depression patients in past week on y-axis
#by SIMD quintile on x-axis and differentiate sex with different fill colour
  geom_bar(stat = "identity") + #Plot a bar chart
  coord_flip() + #Flip the x-axis and y-axis
  geom_hline(yintercept = 0) + #Add a solid line through y=0
  theme_classic() +
#Use a classic theme with a white background and solid lines on the axes
  theme(legend.position = "top") + #Position the legend above the plot
  scale_y_continuous(limits = c(-40,40), breaks = breaks_values,
                     labels = abs(breaks_values)) +
#Make the y-axis run from -40 to 40,
#with increments of 10, as specified in breaks_values,
#and use absolute values for the labels, rather than negatives for females
  xlab("SIMD quintile") + #Remove underscore from x-axis label
  ylab("Adults aged 16+ years with depression symptoms in past week, %") +
#Add more detail to y-axis label
  ggtitle("Adults with depression symptoms in Scotland in 2021",
```

```
            subtitle = "Data from ScotPHO, presented by sex and SIMD quintile") +
#Add title and subtitle
  labs(caption = "ScotPHO = Scottish Public Health Observatory, SIMD = Scottish Index of Multiple Depriv
#Add caption to define abbreviations
#and include extra notes from ScotPHO website regarding the data
  geom_errorbar(aes(ymin = lower_confidence_interval,
                    ymax= upper_confidence_interval, width=0.1)) +
#Add error bars using confidence interval data and specify size of bars
  geom_text(aes(label = abs(value)), nudge_x = 0.2, size = 3)
```

Adults with depression symptoms in Scotland in 2021

Data from ScotPHO, presented by sex and SIMD quintile
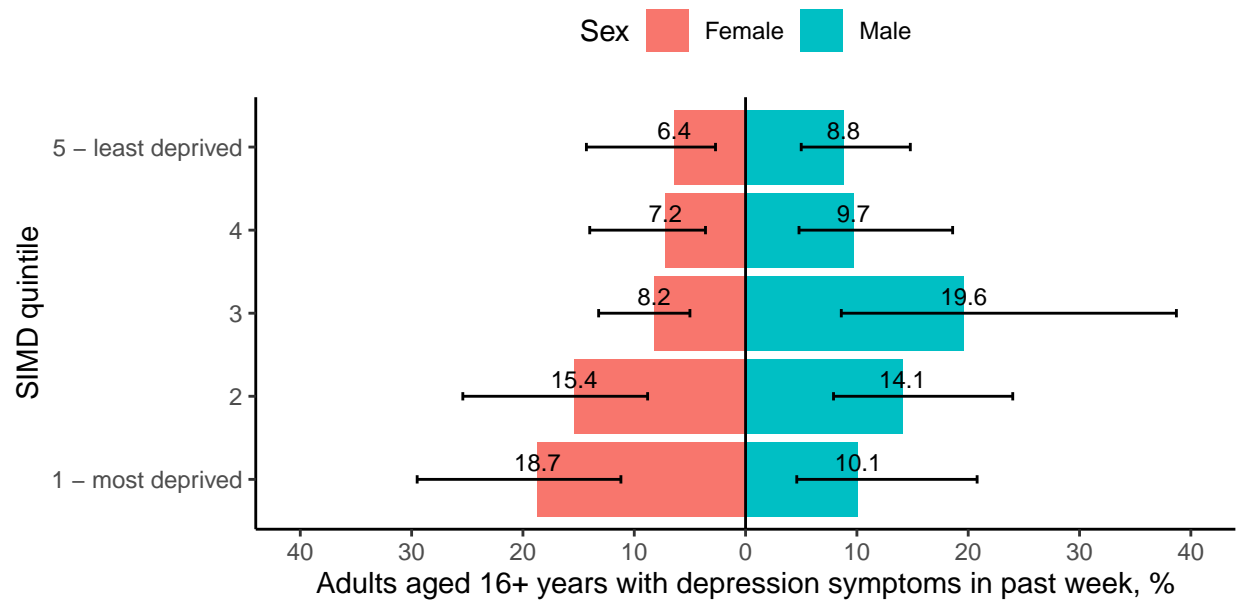
ScotPHO = Scottish Public Health Observatory, SIMD = Scottish Index of Multiple Deprivation.
A person was considered to have depression symptoms in the past week if they had a symptom score
of two or more on the depression section of the Revised Clinical Interview Schedule, indicating symptoms
of moderate to high severity. The SIMD figures were age–standardised to the population of Scotland
to aid comparison between the quintiles. Error bars are confidence intervals.

```
#Label the text, using absolute values, rather than negatives for females,
#and specify position and text size
```