Poor Psychosocial Health across the UK

Jake Wilson

29/03/2022

Introduction This page will show how I used data from the Office of National Statistics to create a map of poor psychosocial health in the UK. This is for the module PSY6422 Data Management and Visualisation as part of an MSc in Cognitive Neuroscience and Human Neuroimaging. Please see all relevant data in my github repository here. My idea for this project was to project **psychosocial health data** on a map of the **United Kingdom** and compare the scores across different regions using a colour scale to represent higher and lower symptoms of psychosocial health.

Data Origins The data for psychosocial health across the UK can be found on the ONS website. Poor psychosocial health is measured using the General Health Questionnaire which asks participants 12 questions about their recent feelings. Attaining a score of 4 or more indicates the individual has symptoms of mild to moderate illness such as depression or anxiety, as explained in the document. This is collected by the ONS and the prevalence of a GHQ score of 4 or above across the UK is calculated. The prevalence is used in this project.

I didn't want my data to be affected by the SARS-CoV-2 pandemic and wanted it to represent *normal* conditions in the UK. Unfortunately the ONS does not have data past 2016 but before the SARS-CoV-2 pandemic but the data is suitable regardless. I did find data from Public Health England however couldn't find a suitable shapefile to match this data that wouldn't require too much work like the one I found here.

Both map data files are also produced by the ONS on their open geography website. The map data files containing England's regions can be found here and the data containing countries can be found here. These essentially give coordinates for r to use to construct a map of the UK.

Research Question What is the prevalence of poor psychosocial health across the United Kingdom? How does this change between regions?

My data attempts to educate people on poor psychosocial health across the UK. Before I did this data visualisation, I could only guess at places with lower/higher levels of poor psychosocial health, and my guesses were wrong. Looking at the ONS data gave some indication, however, mapping the data made it much more clear.

Below are the libraries needed for the project to run correctly.

```
library(readxl)
library(here)
library(rgdal)
library(broom)
library(tidyverse)
library(plotly)
```

```
#find ONS data file
file <- here(
  "data",
  "domainsandmeasuresautumn2019.xls"
  )

#displays the sheets in the file
excel_sheets(
  file
  )
```

Data Preparation

```
##  [1] "All domains overview"            "Further information"
##  [3] "Useful links"                    "Assessment of change"
##  [5] "1.1 Satisfaction"                "1.2 Worthwhile"
##  [7] "1.3 Happy"                       "1.4 Anxious"
##  [9] "1.5 Pop'n mental well-being"     "2.1 Prop in unhappy rel"
## [11] "2.2 Feelings of loneliness"      "2.3 People to rely on "
## [13] "3.1 Healthy life expectancy"     "3.2 People with a disability"
## [15] "3.3 Satis with general health"   "3.4 Psychosocial health"
## [17] "4.1 ILO unemployment"            "4.2 Satisfaction with job"
## [19] "4.3 Satis amount of leisure"     "4.4 Voluntary work"
## [21] "4.5 Arts participation"          "4.6 Sports participation"
## [23] "5.1 Crime rates"                 "5.2 Walking alone after dark"
## [25] "5.3 Visiting natural envir'ment" "5.4 Belonging to neighbourhood"
## [27] "5.5 Minimum travel time"         "5.6 Satisf'n with accommodation"
## [29] "6.1 Thresholds of income"        "6.2 Household wealth"
## [31] "6.3 Median household income"     "6.4 Satisfaction with income"
## [33] "6.5 Managing financially"        "7.1 Real net disposable income "
## [35] "7.2 Public sector debt"          "7.3 CPIH"
## [37] "8.1 Human capital"               "8.2 NEETs"
## [39] "8.3 Level of qualifications"     "9.1 Voter turnout"
## [41] "9.2 Trust in Government"         "10.1 Greenhouse gas emissons"
## [43] "10.2 Extent of protected areas"  "10.3 Energy consumption"
## [45] "10.4 Household recycling "
```

It's clear a lot of work is needed!

Problems:

- There are several sheets but I only need one.
- The sheet contains a lot of useles information for my plot.
- The columns are inapproriately named, especially to work with the map data.
- One region name doesn't match the corresponding region in the map data.

```
#read excel sheet
df_uk <- read_excel(
  file,
  "3.4 Psychosocial health"
  ) %>%
```

```
#extracts the data from df of regions in the UK
  slice(
    36:48
    ) %>%

#renames columns to a meaningful value and also to work with the map data (see below)
  rename(
    "id" = "National Well-being Measures, October 2019 release",
    "people" = "...2"
    ) %>%

#renames East to East of England to match the shapefile data (see below)
  mutate(
    id=recode(
      id,
      East="East of England"
      ))

#displays the top of the dataframe
head(df_uk)
```

```
## # A tibble: 6 x 4
##   id                        people             ...3               ...4
##   <chr>                     <chr>              <chr>              <chr>
## 1 England                   19.199999999999999 18.5               19.80000000000~
## 2 North East                18.100000000000001 15.199999999999999 21.5
## 3 North West                20.699999999999999 18.800000000000001 22.69999999999~
## 4 Yorkshire and The Humber 18.699999999999999 16.899999999999999 20.60000000000~
## 5 East Midlands             18.600000000000001 16.699999999999999 20.60000000000~
## 6 West Midlands             20.800000000000001 19                 22.80000000000~
```

The ONS GHQ data is now ready to be used! Next the map data needs to be manipulated. *Note*: map data of regions of England as well as countries in the UK are included as the ONS GHQ data file contains both. There is not one file with both on the ONS website so this has to be dealt with.

```
#load shapefile of England's regions
shapefile_regions <- readOGR(
  here(
    "data",
    "RGN_DEC_2021_EN_BFC.shp"
    ))

#load shapefile of the rest of the UK
shapefile_uk <- readOGR(
  here(
  "data",
  "CTRY_DEC_2021_UK_BFC.shp"
  ))
```

```
#reshape England shapefile so ggplot can interpret it
mapdata_regions <- tidy(
  shapefile_regions,
```

```
  region = "RGN21NM"
  )

#displays 10 random rows of the dataframe
sample_n(mapdata_regions, 10)
```

```
## # A tibble: 10 x 7
##       long     lat   order hole  piece group             id
##      <dbl>   <dbl>   <int> <lgl> <fct> <fct>             <chr>
##  1 546665. 181117.  506402 FALSE 1     London.1          London
##  2 537815. 338791.   47945 FALSE 1     East Midlands.1   East Midlands
##  3 646322. 302026.  167807 FALSE 1     East of England.1 East of England
##  4 242292   54235. 1125869 FALSE 1     South West.1      South West
##  5 381674. 219959. 1291566 FALSE 1     South West.1      South West
##  6 340425. 425897.  756013 FALSE 5     North West.5      North West
##  7 347144. 480541.  691866 FALSE 1     North West.1      North West
##  8 519965. 176900.  499058 FALSE 1     London.1          London
##  9 597371. 210895.  266371 FALSE 1     East of England.1 East of England
## 10 286806.  52190. 1078114 FALSE 1     South West.1      South West
```

```
#reshape UK shapefile into a tidy data frame so ggplot can interpret it
mapdata_uk <- tidy(
  shapefile_uk,
  region = "CTRY21NM"
  )

#displays 10 random rows of the dataframe
sample_n(mapdata_uk, 10)
```

```
## # A tibble: 10 x 7
##       long      lat   order hole  piece group             id
##      <dbl>    <dbl>   <int> <lgl> <fct> <fct>             <chr>
##  1 298422.  529000   993741 FALSE 1     England.1         England
##  2 304998.  864756. 1535589 FALSE 1     Scotland.1        Scotland
##  3  54152.  575313. 1429148 FALSE 1     Northern Ireland.1 Northern Ireland
##  4 368738. 1039243. 2699499 FALSE 22    Scotland.22       Scotland
##  5 186448.  885091  2018916 FALSE 1     Scotland.1        Scotland
##  6 367912.  758001. 1585319 FALSE 1     Scotland.1        Scotland
##  7 162757.  550764. 1280174 FALSE 1     Northern Ireland.1 Northern Ireland
##  8 162549.  620443. 1858008 FALSE 1     Scotland.1        Scotland
##  9 115444   756226. 2671439 FALSE 19    Scotland.19       Scotland
## 10  76016.  811742. 3035711 FALSE 649   Scotland.649      Scotland
```

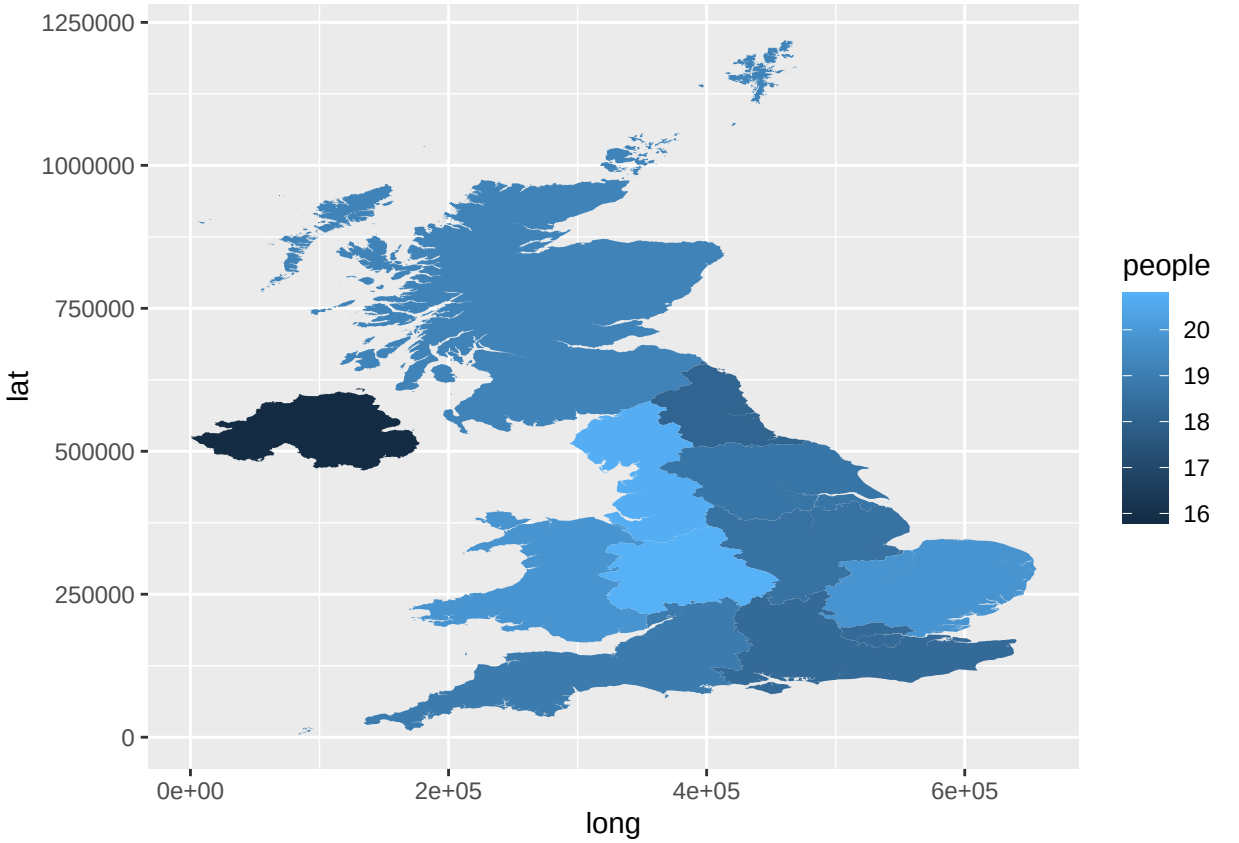
There are again problems:

- Remove England from the country data.
- Combine mapdata to remove the England gap.
- Add ONS GHQ data to plot map data with it.
- Plot map.

```
#removes England from mapdata_uk as we already have the regions of the UK
mapdata_uk_noeng <- subset(mapdata_uk, mapdata_uk$id != "England")

#combines both mapdata dfs
combined_mapdata <- bind_rows(
  mapdata_uk_noeng,
  mapdata_regions
  ) %>%

#adds ONS GHQ data to shapefiles
inner_join(
  df_uk,
  by="id"
  ) %>%

#defines the people column as a numeric vector
mutate(people = as.numeric(
  people
  ))

#displays the top of the dataframe
sample_n(combined_mapdata, 10)
```

```
## # A tibble: 10 x 10
##       long     lat   order hole  piece group            id    people ...3  ...4
##      <dbl>   <dbl>   <int> <lgl> <fct> <fct>            <chr>  <dbl> <chr> <chr>
##  1 603424. 205423.  280780 FALSE 1     East of England~ East~   19.8 17.8~ 21.8~
##  2 180900. 203119. 3295135 FALSE 1     Wales.1          Wales   19.8 17.8~ 22
##  3 335054. 473746.  695074 FALSE 1     North West.1     Nort~   20.7 18.8~ 22.6~
##  4 220946. 553152. 1754454 FALSE 1     Scotland.1       Scot~   19.2 17.1~ 21.3~
##  5  87218. 856330. 2830141 FALSE 57    Scotland.57      Scot~   19.2 17.1~ 21.3~
##  6 442799.  91720.  954332 FALSE 2     South East.2     Sout~   18.3 16.8~ 19.8~
##  7 168598. 743772. 1920102 FALSE 1     Scotland.1       Scot~   19.2 17.1~ 21.3~
##  8 219246  958324. 3038661 FALSE 678   Scotland.678     Scot~   19.2 17.1~ 21.3~
##  9 166024. 636209. 1860340 FALSE 1     Scotland.1       Scot~   19.2 17.1~ 21.3~
## 10  97490. 885795. 2895932 FALSE 108   Scotland.108     Scot~   19.2 17.1~ 21.3~
```

```
#creates the initial map
map_test <- ggplot() +
  geom_polygon(
  data = combined_mapdata,
  aes(
    x = long,
    y = lat,
    group = group,
    fill = people
    ))

#plots the initial map
map_test
```



The problems don't stop there!

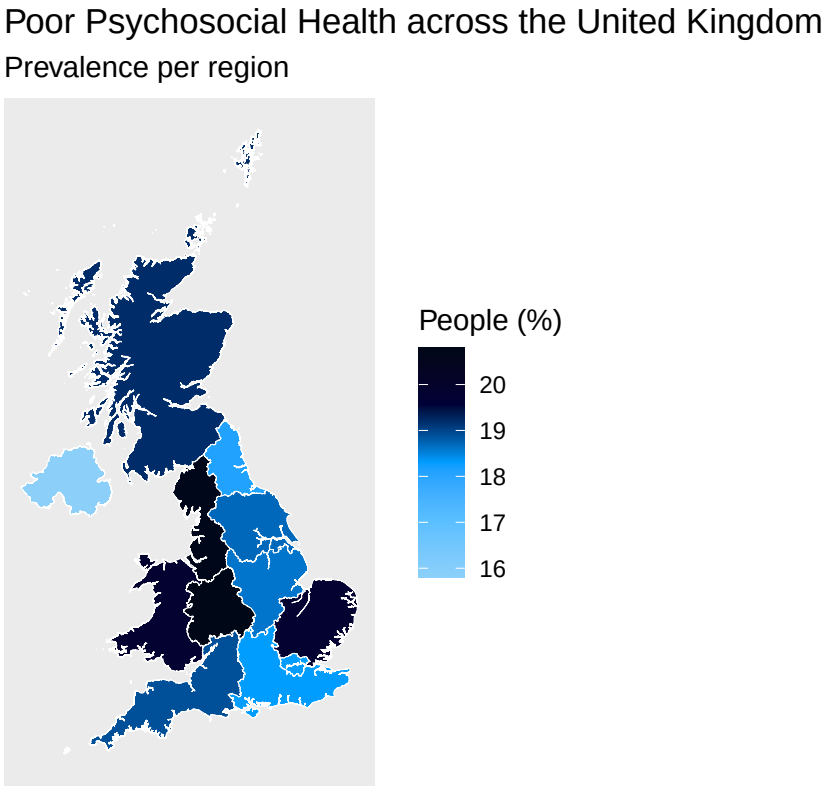
Problems:

- The map plots on a graph, has labels on its axis and ticks on the axis that need removing.
- An appropriate title needs to be given to the graph and to the legend.
- The legend shows darker colours as a lower prevalence of GHQ which is counter intuitive.
- It's hard to see some boundaries with some of the regions in England and the borders between these bits are unclear.
- The map is squashed

Data Visualisation Below is the code to plot the final graph.

```
#create plot, removes grid and axis and adds labels and adds appropriate colours
map <- combined_mapdata %>%
  ggplot(
    aes
         (
           long,
           lat,
           group= group,
           fill= people,
           text = paste
           (
             id, "<br>","People:",people,"%"
           ))) +
```

```
  labs(
    title = "Poor Psychosocial Health across the United Kingdom",
    subtitle = "Prevalence per region",
    fill = "People (%)",
    caption = "Source: Office for National Statistics") +
  geom_polygon(color = "white",
               size = 0.01) +
  theme(
    axis.ticks = element_blank(),
    axis.text = element_blank(),
    panel.grid = element_blank()
    ) +
  labs(
    x=element_blank(),
    y=element_blank()
    ) +
  coord_fixed(1) +
    scale_fill_gradientn(
    colours = c(
      "#8cd0fa",
      "#59bfff",
      "#009dff",
      "#000137",
      "#000817"
      ))

#displays final map but without interactivity
map
```



Source: Office for National Statistics

```
#saves map to directory
ggsave(filename = "GHQ_uk.png")

#makes the map interactive
map_interact <- ggplotly(
  map, tooltip = "text"
  ) %>%
  layout(
    title = list(
    text = paste0(
      "Poor Psychosocial Health across the United Kingdom",
                                    "<br>",
                                    "<sup>",
                                    "Prevalence per region",
                                    "</sup>"
                                   )))

#displays the interactive map
map_interact
```

Summary In summary, I have used publicly available data to create a an interactive map of the prevalence of mild to moderate depression or anxiety across the UK.

From the map it is clear that:

- Northern Ireland has the lowest prevalence of mild to moderate depression or anxiety.

- The West Midlands have the highest prevalence.

Caveats The data does not say anything about severity, gender differences, individual county differences, age differences etc. In fact, the data represents an average and doesn't account for these differences on the individual level. For example, within the North West, it is more likely that across the whole region that someone will be depressed, yet if we account for gender, it may be that women would be more depressed than men. Therefore, the data assumes that the above variables don't affect it.

Lastly, if you averaged the data for across regions in England, it would show that England has a prevalence of mild to moderate depression or anxiety of 19.2% which is similar to that of Wales and Scotland which may be unclear from the data in individual regions in England.

Future Directions If I had more time on this project I would include counties as opposed to regions. This would enable readers to see the distribution of GHQ scores across the UK in more detail. I did find data for Clinical Commissioning Groups (CCG) which is essentially areas the NHS covers over counties however I couldn't find a suitable shapefile that matched, despite finding a CCG shapefile. To use that shapefile I would have had to amend over approximately 100 CCG rows or find GHQ data for each county, which I couldn't.

I would have also compared this data across multiple time-points and so produce multiple maps. It may have also been interesting to investigate age, gender and income against GHQ scores, to name a few.