Assessment

Valentina Levi

2024-11-01

Investigation Question: Do regions in Northern Scotland, which experience longer winter nights, show higher seasonal increases in antidepressant prescriptions compared to Southern regions?

Introduction Seasonal Affective Disorder (SAD) is a type of depression that occurs during certain seasons, usually fall or winter.

I want to assess whether the regions in Northern Scotland, with more intense seasonal changes and longer winters, experience higher prescription rates of antidepressants than the during winter months as a result of SAD.

The data that will be used is "Prescriptions in the Community" from NHS Public Health Scotland. I have chosen to use the year 2023 because it is the most recent complete year with all 4 seasons.

The Health Boards Area (2019) data from Scotland's Census will be used to define northern and southern regions in Scotland.

```
library(tidyverse)
```

Loading in the data sets, creating columns and joining

```
## -- 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
```

```
library(janitor) # For cleaning the data
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```r
library(gt) # For tables
library(here) # For directory structure
```

```
## here() starts at /Users/valentinalevi/Desktop/data_science/B208015
```

```r
library(readr)
library(dplyr)
library(sf)
```

```
## Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE
```

```r
library(viridis)
```

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

```r
library(gt)

#Reading in healthboards data
healthboards <- read.csv(here("data/healthboards.csv")) %>%
  clean_names()

# Define northern and southern regions based on Health Boards
northern_boards <- c("NHS Grampian", "NHS Highland", "NHS Orkney", "NHS Shetland", "NHS Western Isles")
southern_boards <- c("NHS Ayrshire and Arran", "NHS Borders", "NHS Dumfries and Galloway",
                     "NHS Fife", "NHS Forth Valley", "NHS Greater Glasgow and Clyde",
                     "NHS Lanarkshire", "NHS Lothian", "NHS Tayside")

# Add a Region column to the Health Board data
healthboards <- healthboards %>%
  mutate(Region = case_when(
    hb_name %in% northern_boards ~ "Northern",
    hb_name %in% southern_boards ~ "Southern",
    TRUE ~ "Other"
  ))

#Reading in prescriptions data from each month in the year 2023.
january <- read.csv(here("data/january.csv")) %>%
  clean_names()

february <- read.csv(here("data/february.csv")) %>%
  clean_names()

march <- read.csv(here("data/march.csv")) %>%
  clean_names()

april <- read.csv(here("data/april.csv")) %>%
  clean_names()

may <- read.csv(here("data/may.csv")) %>%
  clean_names()
```

```r
june <- read.csv(here("data/june.csv")) %>%
  clean_names()

july <- read.csv(here("data/july.csv")) %>%
  clean_names()

august <- read.csv(here("data/august.csv")) %>%
  clean_names()

september <- read.csv(here("data/september.csv")) %>%
  clean_names()

october <- read.csv(here("data/october.csv")) %>%
  clean_names()

november <- read.csv(here("data/november.csv")) %>%
  clean_names()

december <- read.csv(here("data/december.csv")) %>%
  clean_names()

# List of monthly data frames
monthly_data <- list(
  january %>% mutate(Month = 1),
  february %>% mutate(Month = 2),
  march %>% mutate(Month = 3),
  april %>% mutate(Month = 4),
  may %>% mutate(Month = 5),
  june %>% mutate(Month = 6),
  july %>% mutate(Month = 7),
  august %>% mutate(Month = 8),
  september %>% mutate(Month = 9),
  october %>% mutate(Month = 10),
  november %>% mutate(Month = 11),
  december %>% mutate(Month = 12)
)


# Combine datasets for each season

# Winter: December, January, February
winter <- bind_rows(
  january %>% mutate(Month = 1),
  february %>% mutate(Month = 2),
  december %>% mutate(Month = 12)
) %>% mutate(Season = "Winter")

# Spring: March, April, May
spring <- bind_rows(
  march %>% mutate(Month = 3),
  april %>% mutate(Month = 4),
  may %>% mutate(Month = 5)
) %>% mutate(Season = "Spring")
```

```r
#Summer: June, July, August
summer <- bind_rows(
  june %>% mutate(Month = 6, dmd_code = as.character(dmd_code)),
  july %>% mutate(Month = 7, dmd_code = as.character(dmd_code)),
  august %>% mutate(Month = 8, dmd_code = as.character(dmd_code))
) %>% mutate(Season = "Summer")

# Autumn: September, October, November
autumn <- bind_rows(
  september %>% mutate(Month = 9),
  october %>% mutate(Month = 10),
  november %>% mutate(Month = 11)
) %>% mutate(Season = "Autumn")

# Removing the `dmd_code` column from each seasonal dataset
winter <- winter %>% select(-dmd_code)
spring <- spring %>% select(-dmd_code)
summer <- summer %>% select(-dmd_code)
autumn <- autumn %>% select(-dmd_code)

# Combine all seasonal data sets into one
combined_2023 <- bind_rows(winter, spring, summer, autumn)

#Joining the combined seasonal data to the healthboard data
joined_2023 <- full_join(healthboards,combined_2023, by = join_by(hb == hbt))
```

Filtering the data to focus on antidepressants. From research online I have discovered that the typical antidepressants prescribed for people diagnosed with SAD are selective serotonin reuptake inhibitors (SSRIs). In particular, sertraline and fluoxetine.

```r
# Filtering out only the data with fluoxetine and sertaline in the description of the drug
antidepressant_2023 <- joined_2023 %>%
  filter(
    str_detect(bnf_item_description, "FLUOXETINE") |
    str_detect(bnf_item_description, "SERTRALINE")
  )
```

```r
seasonal_totals <- antidepressant_2023 %>%
  group_by(hb_name, Region, Season) %>%
  summarise(Total_Prescriptions = sum(paid_quantity, na.rm = TRUE)) %>%
  ungroup()
```

Calculating the total antidepressant prescriptions for Fluoxetine and Sertraline by healthbaord, season and region

```
## `summarise()` has grouped output by 'hb_name', 'Region'. You can override using
## the `.groups` argument.
```

```r
#Reading in population data
population <- read_csv(here("data/UV103_age_health_board_census.csv"), skip = 10) %>%
  # Rename the last column to avoid the messy name in column 6
  # and to match column names with the prescription dataset
  rename(Spare = "...6",
         hb_name = "Health Board Area 2019",
         hb_population = Count) %>%
  # filter the data so that we get the population of the entire health board
  filter(Age == "All people" & Sex == "All people") %>%
  # select only the relevant columns
  select(hb_name, hb_population) %>%
  # change health board names so they match the prescription data
  mutate(hb_name = paste("NHS", hb_name))
```

```
## New names:
## * `` -> `...6`
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 3657 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): Counting, Health Board Area 2019, Age, Sex
## dbl (1): Count
## lgl (1): ...6
##
## 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
#Joining population data and the total antidepressant prescriptions for Fluoxetine and Sertraline by he
joined_data <- full_join(population,seasonal_totals, by = join_by(hb_name))


# Calculate the prescription rate per person by healthboard, season and region.
prescription_rates <- joined_data %>%
  mutate(prescriptions_per_person = Total_Prescriptions / hb_population)
```

```r
# Load spatial data and standardize the column
NHS_healthboards <- st_read("~/Desktop/data_science/B208015/data/NHS_healthboards_2019.shp") %>%
  rename(hb_name = HBName) %>%
  mutate(hb_name = paste("NHS", hb_name))
```

Creating Mapped Graph to present the data

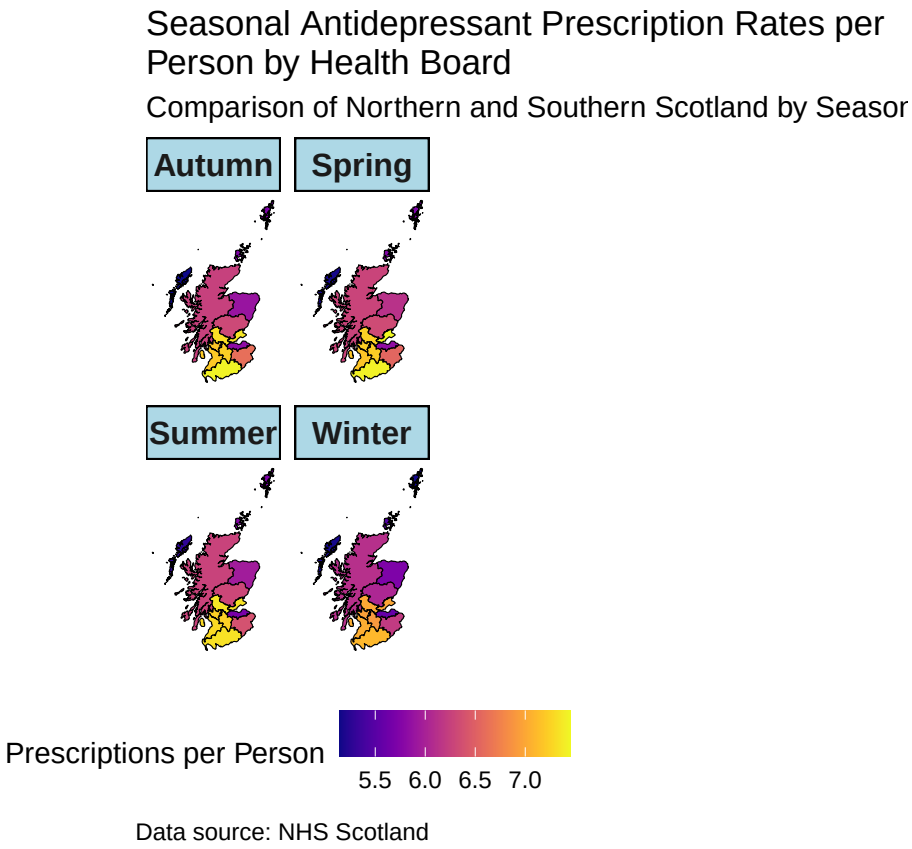
```
## Reading layer `NHS_healthboards_2019' from data source
##   `/Users/valentinalevi/Desktop/data_science/B208015/data/NHS_healthboards_2019.shp'
```

```
##   using driver ‘ESRI Shapefile’
## Simple feature collection with 14 features and 4 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 7564.996 ymin: 530635.8 xmax: 468754.8 ymax: 1218625
## Projected CRS: OSGB36 / British National Grid
```

```r
# Join data
prescription_rates_map <- NHS_healthboards %>%
  left_join(prescription_rates, by = "hb_name")

# Plot
map_plot <- ggplot(data = prescription_rates_map) +
  geom_sf(aes(fill = prescriptions_per_person), color = "black", linewidth = 0.2) +
  scale_fill_viridis(name = "Prescriptions per Person", option = "C", labels = scales::comma) +
  facet_wrap(~ Season, ncol = 2) +
  theme_minimal() +
  labs(
    title = "Seasonal Antidepressant Prescription Rates per\nPerson by Health Board",
    subtitle = "Comparison of Northern and Southern Scotland by Season",
    caption = "Data source: NHS Scotland"
  ) +
  theme(
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank(),
    strip.background = element_rect(fill = "lightblue", color = "black"),
    strip.text = element_text(size = 12, face = "bold"),
    legend.position = "bottom",
  )

# Display the plot
print(map_plot)
```




```r
# Prepare the data for the North and South tables
north_table <- prescription_rates %>%
  filter(Region == "Northern") %>%
  select(hb_name, Season, prescriptions_per_person) %>%
  pivot_wider(names_from = hb_name, values_from = prescriptions_per_person) %>%
  arrange(Season)

south_table <- prescription_rates %>%
  filter(Region == "Southern") %>%
  select(hb_name, Season, prescriptions_per_person) %>%
  pivot_wider(names_from = hb_name, values_from = prescriptions_per_person) %>%
  arrange(Season)

# Create the Northern table with gt
north_table_gt <- north_table %>%
  gt() %>%
  tab_header(
    title = "Northern Scotland NHS Healthboards Antidepressant Prescription Data per person"
  ) %>%
  cols_label(
    Season = "Season"
  ) %>%
  fmt_number(
```

Northern Scotland NHS Healthboards Antidepressant Prescription Data per person

Season	NHS Grampian	NHS Highland	NHS Orkney	NHS Shetland	NHS Western Isles
Autumn	5.89	6.25	5.74	5.72	5.15
Spring	6.13	6.28	5.74	5.73	5.18
Summer	5.95	6.26	5.78	5.72	5.29
Winter	5.73	6.11	5.48	5.17	5.17

Southern Scotland NHS Healthboards Antidepressant Prescription Data per person

Season	NHS Ayrshire and Arran	NHS Borders	NHS Dumfries and Galloway	NHS Forth Valley	N
Autumn	7.20	6.61	7.42	7.30	
Spring	7.17	6.54	7.41	7.23	
Summer	7.28	6.38	7.32	7.35	
Winter	7.03	6.20	7.09	7.00	

```
    columns = everything(),
    decimals = 2
  )

# Create the Southern table with gt
south_table_gt <- south_table %>%
  gt() %>%
  tab_header(
    title = "Southern Scotland NHS Healthboards Antidepressant Prescription Data per person"
  ) %>%
  cols_label(
    Season = "Season"
  ) %>%
  fmt_number(
    columns = everything(),
    decimals = 2
  )

# Display the Northern and Southern tables
north_table_gt
```

```
south_table_gt
```

Creating tables to present data This analysis demonstrates that antidepressant prescription rates actually tend to be higher in Southern Scotland and are actually lowest in the winter season. This reveals that although Seasonal Affective Disorder (SAD) is associated with reduced daylight during winter months, its prevalence and the prescription of antidepressants do not consistently increase with latitide. This highlights that factors beyond daylight duration (as a result of seasons), such as genetic predispositions, environmental conditions, and individual sensitivity to light changes, play significant roles in the occurrence of SAD.

Additionally, some people experience "reverse SAD," where they feel low during the summer months instead of winter. This could be linked to disruptions in their circadian rhythm due to extended daylight hours, higher

temperatures, or changes in routine. It's a complex interplay of biological, psychological, and environmental factors that could explain why SAD affects people differently.

The reasons for higher antidepressant prescription rates in southern Scotland, compared to northern regions, could be linked to several factors instead of seasonal affective disorder. These include:

- Population Density and Accessibility of Services: Southern Scotland has more urban areas, particularly around cities like Glasgow and Edinburgh, where people might have greater access to mental health services and are therefore more likely to receive antidepressant prescriptions. Rural regions often have less access to mental health services, potentially lowering prescription rates.
- Environmental and Lifestyle Differences: Although northern Scotland experiences longer winter nights, factors like weather patterns, lifestyle, and occupational stress in the southern areas might contribute to a different type of mental health stress, unrelated to Seasonal Affective Disorder but still resulting in a need for antidepressants.

Another mental