In Course Assessment 1

Valentina Levi

2024-11-20

Investigation Question: Do regions in Northern Scotland, which experience shorter winter daylight hours, show higher seasonal increases in antidepressant prescriptions compared to Southern regions as a result of SAD?

Introduction Seasonal Affective Disorder (SAD) is a mood disorder characterised by depressive symptoms that occur at a specific time of year (usually autumn or winter) with full remission at other times of year (usually spring or summer) (Galima, Vogel and Kowalski, 2020). The key risk factors include family history, female sex and living at a more northern latitude. Although the etiology of SAD is still rather unclear, it is thought that the decreasing daylight/sunlight hours as a result of transitioning to winter triggers depressive episodes in individuals who are vulnerable to SAD (Praschak-Rieder and Willeit, 2003).

I want to assess whether the regions in Northern Scotland experience higher prescription rates of antidepressants than regions in Southern Scotland as a result of reduced daylight/sunlight causing SAD in individuals during the winter season.

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. Accessible at https://www.opendata.nhs.scot/dataset/prescriptions-in-the-community.

The Health Boards Area (2019) data from Scotland's Census will be used and each healthboard will be defined as either part of the northern or southern regions in Scotland. Accessible at https://www.scotlandscensus.gov.uk/webapi/jsf/tableView/tableView.xhtml.

```r
library(tidyverse) # Loads tidyverse packages (includes ggplot2, dplyr, tidyr)
library(janitor) # For cleaning the data
library(gt) # For tables
library(here) # For directory structure
library(readr) # For reading csv files
library(sf) # Handles spatial data for mapping
library(viridis) # For colour palettes
library(purrr) # Tools to tidy data (e.g. reshape data frames wide to long)

# 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"
  ))

# Define a pattern to match month files
month_pattern <- "january|february|march|april|may|june|july|august|september|october|november|december

# List only files matching the month names
monthly_files <- list.files(
  path = here("data"),
  pattern = month_pattern, # Using pattern
  full.names = TRUE,
  ignore.case = TRUE
)

# Define the processing function
process_monthly_data <- function(file_path) {
  read_csv(file_path) %>%
    clean_names() %>%
    select(-any_of("dmd_code")) %>%  # Get rid of `dmd_code` if it exists
    mutate(
      month = as.integer(substr(paid_date_month, 5, 6)),
      Season = case_when( #Assign season based on month
        month %in% c(12, 1, 2) ~ "Winter",
        month %in% c(3, 4, 5) ~ "Spring",
        month %in% c(6, 7, 8) ~ "Summer",
        month %in% c(9, 10, 11) ~ "Autumn"
      )
    )
}

# Read and combine the monthly files
prescription_monthly <- map_df(monthly_files, process_monthly_data)

# Join prescription data with Healthboard Data
prescription_hb <- full_join(healthboards, prescription_monthly, by = join_by(hb == hbt))
```

Loading in the data sets, creating seasons and joining healthboard data to prescription data

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 (Galima, Vogel and Kowalski, 2020).

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

```r
    str_detect(bnf_item_description, "SERTRALINE")
  )
```

Calculating the total antidepressant prescriptions for Fluoxetine and Sertraline by healthboard, season and region It is important to calculate the prescription (rates) per person because some healthboards have more people living in them than others.

```r
# Calculate seasonal totals for antidepressant prescriptions in 2023
seasonal_totals <- antidepressant_2023 %>%
  group_by(hb_name, Region, Season) %>%
  summarise(Total_Prescriptions = sum(paid_quantity, na.rm = TRUE)) %>%
  ungroup()

#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 prescri
rename(Spare = "...6",
       hb_name = "Health Board Area 2019",
       hb_population = Count) %>%
filter(Age == "All people" & Sex == "All people") %>% # Filter the data so that we get the population o
select(hb_name, hb_population) %>% # Select only the relevant columns
# Change health board names so they match the prescription data
mutate(hb_name = paste("NHS", hb_name))

#Joining population data and the total antidepressant prescriptions for Fluoxetine and Sertraline by he
antidepressant_population <- full_join(population,seasonal_totals, by = join_by(hb_name))

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

Creating Mapped Graph to present the data This plot gives us a more general overview of the differences in antidepressant prescription rates across the northern and southern healthboards in Scotland.

Contrary to expectations that winter might have higher antidepressant prescription rates due SAD, winter actually shows the lowest prescription rates overall.

The highest rates appear in autumn and spring, particularly in southern urban healthboards as indicated by the warmer (yellow) colors.Northern, more rural healthboards maintain consistently lower rates throughout all seasons, as represented by the cooler tones.

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

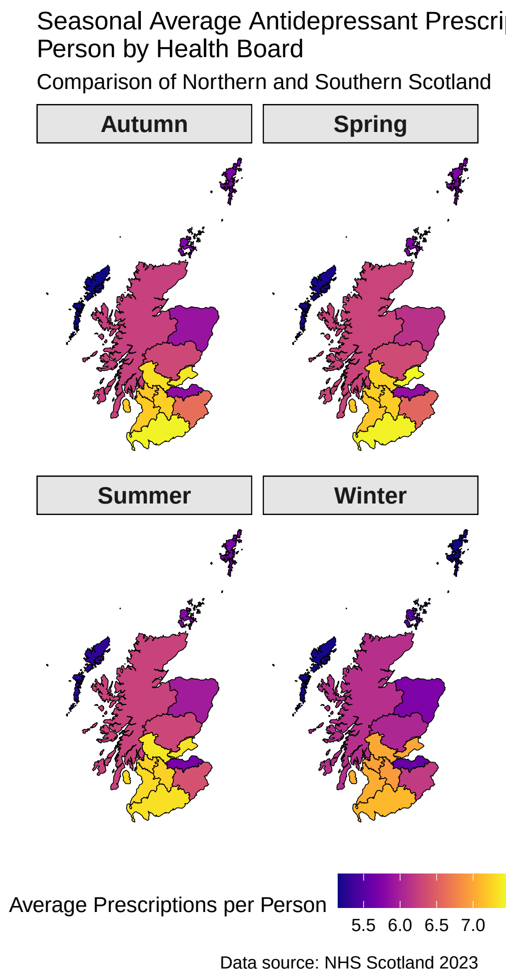
```
## Reading layer ‘NHS_healthboards_2019’ from data source
##   ‘/Users/valentinalevi/Desktop/data_science/B208015/data/NHS_healthboards_2019.shp’
##   using driver ‘ESRI Shapefile’
## Simple feature collection with 14 features and 4 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
```

```
## Bounding box:  xmin: 7564.996 ymin: 530635.8 xmax: 468754.8 ymax: 1218625
## Projected CRS: OSGB36 / British National Grid
```

```r
# Join prescriptions rates and healthboard data (with spatial data for creating map)
prescription_rates_map <- NHS_healthboards %>%
  left_join(prescription_rates, by = "hb_name")

# Creating map plot of healthboards in Scotland
map_plot <- ggplot(data = prescription_rates_map) +
  geom_sf(aes(fill = prescriptions_per_person), color = "black", linewidth = 0.2) +
  scale_fill_viridis(name = "Average Prescriptions per Person", option = "C", labels = scales::comma) +
  facet_wrap(~ Season, ncol = 2) + # Create facets for each season
  theme_minimal() +
  labs(
    title = "Seasonal Average Antidepressant Prescription Rates per\nPerson by Health Board",
    subtitle = "Comparison of Northern and Southern Scotland",
    caption = "Data source: NHS Scotland 2023"
  ) +
  theme(
    axis.text = element_blank(), # Remove axis text
    panel.grid = element_blank(), # Remove grid lines for a cleaner map
    strip.background = element_rect(fill = "grey90", color = "black"),
    strip.text = element_text(size = 12, face = "bold"), # Adjusting style of facet labels
    legend.position = "bottom",
  )

# Display the plot
print(map_plot)
```



Creating two tables (northern region and southern region) to show the average prescription per person for each healthboard across the 4 seasons. Here I have included the average sunlight hours from the MET office data because a theory of SAD is that a reduction in sunlight stops the hypothalamus in the brain from working properly, therefore affecting the production of serotonin (a hormone that affects your mood - low levels leads to depression) (NHS, 2021). I have used the North Scotland MET data for the northern region and combined the East and West of Southern Scotland MET data for the southern region.

Sunlight hours follow an expected pattern, peaking in summer and hitting their lowest point in winter. However, antidepressant prescription rates display the opposite pattern to what is expected, increasing as the average sunlight hours increase. For both tables highlighted in yellow is the healthboard with the average highest antidepressant prescription rates across seasons and highlighted in blue is the lowest.

```
# Function to process Met Office sunshine data
process_sunshine <- function(url, year_filter = 2023) {
  read_table(url, skip = 5) %>% # Read the table from the URL, skipping the first 5 rows (headings - no
    filter(year == year_filter) %>% # Filter data for year specified (2023)
    pivot_longer( # Reshape the data from wide to long
      cols = jan:dec, # The columns represent monthly data
      names_to = "month", # Create a new column called "month" from the column names
      values_to = "sunshine_hours" # Create a new column called "sunshine_hours"
    ) %>%
    mutate(
```

```
      month = factor(month, levels = c("jan", "feb", "mar", "apr", "may", "jun", "jul", "aug", "sep", "
      season = case_when( # Create a new "season" column based on the month (specifying which month is
        month %in% c("dec", "jan", "feb") ~ "Winter",
        month %in% c("mar", "apr", "may") ~ "Spring",
        month %in% c("jun", "jul", "aug") ~ "Summer",
        month %in% c("sep", "oct", "nov") ~ "Autumn",
      )
    )
}

# Function to generate a GT table
gt_table <- function(prescription_data, title) {
  data_wide <- prescription_data %>% # Process data into wide format
    select(hb_name, Season, prescriptions_per_person, avg_sunlight_hours) %>%
    pivot_wider(names_from = hb_name, values_from = prescriptions_per_person) %>%
    arrange(Season)

 numeric_columns <- data_wide %>%
    select(-Season, -avg_sunlight_hours) # Exclude columns that are not average precription rates

# Find the column names for the highest and lowest average prescription rates
  max_column <- names(numeric_columns)[which.max(colSums(numeric_columns, na.rm = TRUE))]
  min_column <- names(numeric_columns)[which.min(colSums(numeric_columns, na.rm = TRUE))]

# Function to continue generating and styling gt table
data_wide %>%
  gt() %>%
    tab_header(
      title = title,
      subtitle = "Including Average Sunlight Hours per Season"
    ) %>%
    cols_label(
      Season = "Season",
      avg_sunlight_hours = "Avg Sunlight (Hours)"
    ) %>%
    fmt_number(columns = everything(), decimals = 2) %>% # Round everything to 2 decimal places
    tab_style(
      style = cell_fill(color = "yellow"),
      locations = cells_body(columns = all_of(max_column)) # Highlight the healthboard with the average
    ) %>%
    tab_style(
      style = cell_fill(color = "lightblue"),
      locations = cells_body(columns = all_of(min_column)) # Highlight the heathboard with the lowest a
    ) %>%
    tab_style(
      style = cell_text(weight = "bold"), # Make column headers bold to stand out
      locations = cells_column_labels(everything())
    ) %>%
    tab_style(
      style = cell_text(weight = "bold"), # Make Season and Average Sunlight hours colums bold too
      locations = cells_body(columns = c(Season, avg_sunlight_hours))
    ) %>%
    opt_row_striping() %>% # Add row striping to make the table more readable
```

```r
    tab_options(
      table.font.size = px(14), # Adjust the font size and align headings
      heading.align = "center"
    )
}

# MET Northern Scotland Sunshine Data URL
northern_sunshine_url <- "https://www.metoffice.gov.uk/pub/data/weather/uk/climate/datasets/Sunshine/da

# Process Northern sunshine data
northern_sunshine <- process_sunshine(northern_sunshine_url)

# Calculate average sunshine for each season in Northern Scotland
northern_seasonal_sunshine <- northern_sunshine %>%
  group_by(season) %>%
  summarise(avg_sunlight_hours = mean(sunshine_hours, na.rm = TRUE))

# Filter Northern prescription rates
northern_prescription_rates <- prescription_rates %>%
  filter(Region == "Northern") %>%
  left_join(northern_seasonal_sunshine, by = c("Season" = "season"))

# GT table for Northern Scotland
northern_gt_table <- gt_table(
  northern_prescription_rates,
  "Northern Scotland NHS Healthboards Average Antidepressant Prescription Data per Person in 2023"
)

# MET Southern Scotland Sunshine Data URLs
east_sunshine_url <- "https://www.metoffice.gov.uk/pub/data/weather/uk/climate/datasets/Sunshine/date/S
west_sunshine_url <- "https://www.metoffice.gov.uk/pub/data/weather/uk/climate/datasets/Sunshine/date/S

# Process East and West sunshine data
east_sunshine <- process_sunshine(east_sunshine_url)
west_sunshine <- process_sunshine(west_sunshine_url)

# Combine East and West sunshine data and calculate seasonal averages
southern_sunshine <- east_sunshine %>%
  inner_join(west_sunshine, by = c("year", "month", "season")) %>%
  mutate(
    avg_sunshine_hours = (sunshine_hours.x + sunshine_hours.y) / 2
  ) %>%
  group_by(season) %>%
  summarise(avg_sunlight_hours = mean(avg_sunshine_hours, na.rm = TRUE))

# Filter Southern prescription rates
southern_prescription_rates <- prescription_rates %>%
  filter(Region == "Southern") %>%
  left_join(southern_sunshine, by = c("Season" = "season"))

# Generate GT table for Southern Scotland
southern_gt_table <- gt_table(
  southern_prescription_rates,
```

Northern Scotland NHS Healthboards Average Antidepressant Prescription Data per Person in 2023
Including Average Sunlight Hours per Season

Season	Avg Sunlight (Hours)	NHS Grampian	NHS Highland	NHS Orkney	NHS Shetland	NHS
Autumn	**76.73**	5.89	6.25	5.74	5.72	
Spring	**140.67**	6.13	6.28	5.74	5.73	
Summer	**155.43**	5.95	6.26	5.78	5.72	
Winter	**32.50**	5.73	6.11	5.48	5.17	

Southern Scotland NHS Healthboards Antidepressant Prescription Data per Person in 2023
Including Average Sunlight Hours per Season

Season	Avg Sunlight (Hours)	NHS Ayrshire and Arran	NHS Borders	NHS Dumfries and Gallowa
Autumn	**84.15**	7.20	6.61	7.4
Spring	**144.23**	7.17	6.54	7.4
Summer	**167.88**	7.28	6.38	7.3
Winter	**43.75**	7.03	6.20	7.0

```
  "Southern Scotland NHS Healthboards Antidepressant Prescription Data per Person in 2023"
)

northern_gt_table
```

```
southern_gt_table
```

Line graph showing how the average prescription per person changes across the 4 seasons for each healthboard Here we can more specifically see which healthboards have the highest and lowest rates of antidepressant prescriptions across the seasons.

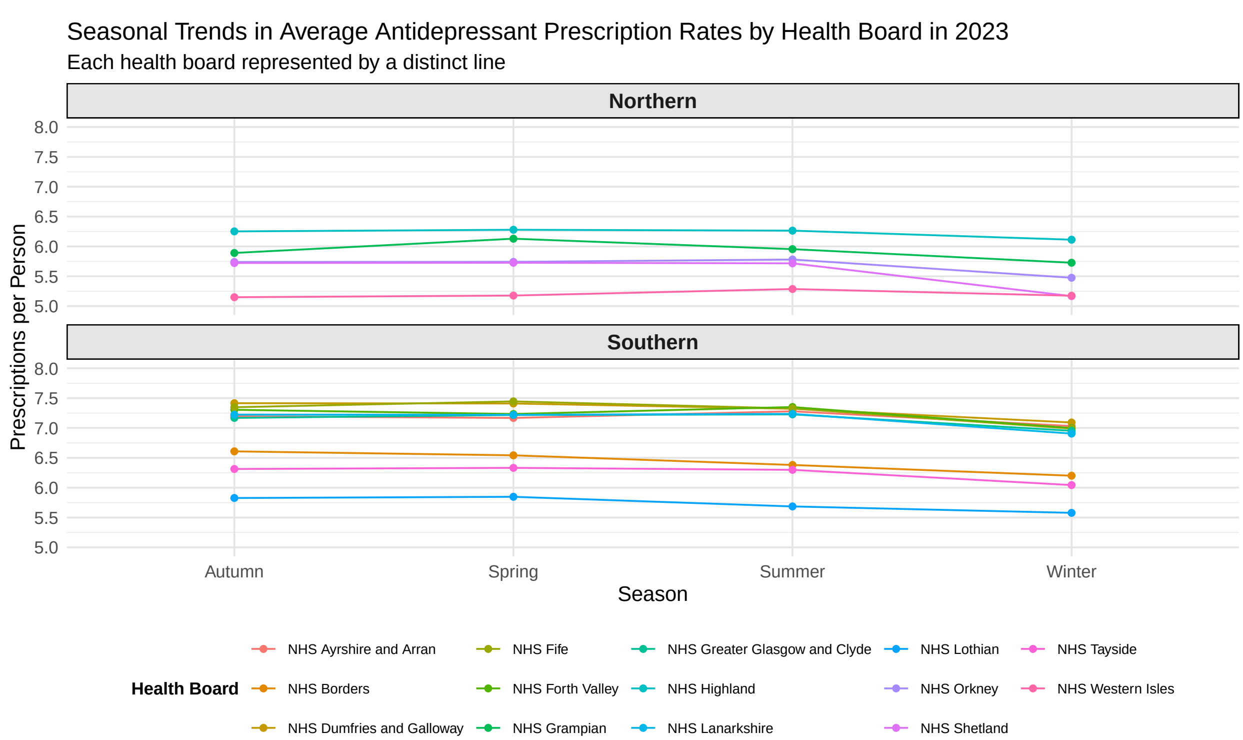
In Northern Scotland, prescription rates show subtle variation across seasons, with slightly higher rates observed in spring and summer, and the lowest rates in winter. Among the northern boards, NHS Highland consistently reports the highest antidepressant prescription rates, whereas NHS Western Isles consistently reports the lowest. In Southern Scotland, the antidepressant prescription rates are generally higher than in the north, with autumn and spring showing peaks, and again in winter slightly lower rates. NHS Dumfries and Galloway stands out with consistently high rates across all seasons, whereas NHS Lothian reports relatively lower rates compared to the other southern healthboards.

```
# Generating bright colour for each healthboard
bright_colors <- scales::hue_pal()(length(unique(prescription_rates$hb_name)))

# Creating line plot of prescription rates across seasons - showing each healthboard
line_plot <- prescription_rates %>%
  ggplot(aes(x = Season, y = prescriptions_per_person, group = hb_name, color = hb_name)) +
  geom_line(linewidth = 0.5) +
  geom_point(size = 1.5) +
  scale_color_manual(values = bright_colors, name = "Health Board") + # Using the custom colour palette
  scale_y_continuous( # Customising the y-axis
    name = "Prescriptions per Person",
```

```r
    breaks = seq(floor(min(prescription_rates$prescriptions_per_person)),  # Defining y axis breaks
                 ceiling(max(prescription_rates$prescriptions_per_person)),
                 by = 0.5),
    limits = c(floor(min(prescription_rates$prescriptions_per_person)), # Setting axis limits
               ceiling(max(prescription_rates$prescriptions_per_person)))
  ) +
  labs(
    title = "Seasonal Trends in Average Antidepressant Prescription Rates by Health Board in 2023",
    subtitle = "Each health board represented by a distinct line",
    x = "Season",
    y = "Average Prescriptions per Person",
    color = "Health Board"
  ) +
  facet_wrap(~ Region, ncol = 1, scales = "fixed") + # Separate panel for North and South regions
  theme_minimal() +
  theme( # Styling the text
    legend.position = "bottom",
    legend.text = element_text(size = 8),
    legend.title = element_text(size = 10, face = "bold"),
    axis.text.x = element_text(size = 10),
    axis.text.y = element_text(size = 10),
    strip.background = element_rect(fill = "grey90", color = "black"),
    strip.text = element_text(size = 12, face = "bold"),
    panel.grid.major = element_line(color = "grey90"),
    text = element_text(size = 12)
  )

print(line_plot)
```



Analysis of Results: Interestingly the results 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 the occurence of Seasonal Affective Disorder (SAD) is typically 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.

Further research suggests that some people actually 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. Therefore, its a complex combination of biological, psychological, and environmental factors that cause SAD to affect people differently.

Limitations: A key limitation of this research question is that it assumes that SSRI antidepressant prescriptions are all a result of SAD when in fact it is likely to be only a small proportion. It is difficult to therefore establish causation rather than just a correlation.

The reasons for higher antidepressant prescription rates could be linked to several factors rather than SAD. For example, Southern Scotland has more urban areas, particularly around cities like Glasgow, where people might have better access to mental health services and are therefore are more likely to receive diagnosis of SAD and antidepressant prescriptions. Rural regions often have less access to mental health services, potentially lowering prescription rates. Additionally, although northern Scotland experiences longer winter nights, factors like lifestyle, and occupational stress in the southern areas might contribute to a different type of depression, resulting in a need for antidepressants.

Another key limitation is that there are alternative therapies used to treat SAD other than just antidepressant prescriptions such as light therapy and cognitive behavioral therapy (CBT) (Galima, Vogel and Kowalski, 2020). Finally, it is important to take into consideration that some people will collect prescriptions to last them the next few months so they could be collected antidepressants for SAD in autumn to last them through the winter. This limitation could affect the results making the prescription rate low for winter when in fact more people could be taking antidepressants at this time.

Next Steps: Analysing the other potential factors that influence antidepressant prescriptions. This includes exploring datasets on sex (SAD is more prevelant in females) , weather patterns (e.g. rainfall), population genetics, levels of stress, access to mental health services and therefore receiving diagnosis / prescription. It would also be interesting to analyse data that is exclusively the antidepressant presciptions for SAD.

References: Galima, S.V., Vogel, S.R. and Kowalski, A.W. (2020). Seasonal Affective Disorder: Common Questions and Answers. American Family Physician, [online] 102(11), pp.668–672. Available at: https://pubmed.ncbi.nlm.nih.gov/33252911/.NHS (2021).

Overview - Seasonal affective disorder (SAD). [online] nhs.uk. Available at: https://www.nhs.uk/mental-health/conditions/seasonal-affective-disorder-sad/overview/.

Praschak-Rieder, N. and Willeit, M. (2003). Treatment of seasonal affective disorders. Chronobiology and Mood Disorders, 5(4), pp.389–398. doi:https://doi.org/10.31887/dcns.2003.5.4/npraschakrieder.

Use of Generative AI: I used ChatGPT to help me with error messages but only when I could not find the error myself. I also used it to help me scale my plots so they fit and were fully visible when I knitted my code.