

# Trends, Disparities, and Care Pathways: An Analysis of Mental Health Inpatient Care in Scotland (1997-2023)

by Code & Care: Ruby, Daisy, Zihao, Suhana

2025-11-18

```
df <- read_csv("mental-health-inpatient-activity-age-sex (1).csv")
```

```
## Rows: 15390 Columns: 13
## └─ Column specification ──────────────────────────────────────────────────
## 
##   #> Delimiter: ","
##   #> chr (9): FinancialYear, HBR, HBRQF, AgeGroup, Sex, Dataset, DatasetQF, Disch...
##   #> dbl (4): DischargeCount, DischargeRates, PatientsCount, PatientsRates
## 
##   #> 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.
```

```
head(df)
```

```
## # A tibble: 6 × 13
##   FinancialYear HBR      HBRQF AgeGroup  Sex   Dataset DatasetQF DischargeCount
##   <chr>        <chr>    <chr>  <chr>   <chr> <chr>    <chr>          <dbl>
## 1 1997/98      S92000003 d     0-17 ye... Male  Total    d             780
## 2 1997/98      S92000003 d     0-17 ye... Fema... Total    d             715
## 3 1997/98      S92000003 d     18-24 ye... Male  Total    d            2020
## 4 1997/98      S92000003 d     18-24 ye... Fema... Total    d            1310
## 5 1997/98      S92000003 d     25-39 ye... Male  Total    d            6450
## 6 1997/98      S92000003 d     25-39 ye... Fema... Total    d            4895
## # ℹ 5 more variables: DischargeRates <dbl>, DischargeRatesQF <chr>,
## #   PatientsCount <dbl>, PatientsRates <dbl>, PatientRatesQF <chr>
```

```
df <- clean_names(df)
```

```
head(df)
```

```

## # A tibble: 6 × 13
##   financial_year hbr    hbrqf age_group sex   dataset dataset_qf discharge_count
##   <chr>          <chr>  <chr>  <chr>  <chr>  <chr>           <dbl>
## 1 1997/98        S9200... d     0-17 ye... Male   Total    d             780
## 2 1997/98        S9200... d     0-17 ye... Fema... Total    d             715
## 3 1997/98        S9200... d     18-24 ye... Male   Total    d            2020
## 4 1997/98        S9200... d     18-24 ye... Fema... Total    d            1310
## 5 1997/98        S9200... d     25-39 ye... Male   Total    d            6450
## 6 1997/98        S9200... d     25-39 ye... Fema... Total    d            4895
## # ℹ 5 more variables: discharge_rates <dbl>, discharge_rates_qf <chr>,
## #   patients_count <dbl>, patients_rates <dbl>, patient_rates_qf <chr>

```

```

df_scotland <- df %>%
  filter( hbrqf == "d", sex %in% c("Male", "Female"),
         !is.na(discharge_rates),
         !is.na(patients_rates),
         dataset == "Total"
       )

```

df\_scotland

```

## # A tibble: 270 × 13
##   financial_year hbr    hbrqf age_group sex   dataset dataset_qf discharge_count
##   <chr>          <chr>  <chr>  <chr>  <chr>  <chr>           <dbl>
## 1 1997/98        S9200... d     0-17 ye... Male   Total    d             780
## 2 1997/98        S9200... d     0-17 ye... Fema... Total    d             715
## 3 1997/98        S9200... d     18-24 ye... Male   Total    d            2020
## 4 1997/98        S9200... d     18-24 ye... Fema... Total    d            1310
## 5 1997/98        S9200... d     25-39 ye... Male   Total    d            6450
## 6 1997/98        S9200... d     25-39 ye... Fema... Total    d            4895
## 7 1997/98        S9200... d     40-64 ye... Male   Total    d            7280
## 8 1997/98        S9200... d     40-64 ye... Fema... Total    d            5395
## 9 1997/98        S9200... d     65plus y... Male   Total    d            5085
## 10 1997/98       S9200... d     65plus y... Fema... Total    d            7860
## # ℹ 260 more rows
## # ℹ 5 more variables: discharge_rates <dbl>, discharge_rates_qf <chr>,
## #   patients_count <dbl>, patients_rates <dbl>, patient_rates_qf <chr>

```

```

ggplot(df_scotland,
       aes(x = financial_year, y = patients_rates, color = sex, group = sex)) +
  geom_line(size = 1.2) +
  geom_point(size = 1.5) +
  facet_wrap(~ age_group, scales = "free_y") +
  labs(title = "Long-term Trends in Mental Health Inpatient Rates by Age Group and Sex in Scotland (1997-2023)",
       x = "Year",
       y = "Discharge_rates",
       color = "Sex") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 3.5),
        strip.text = element_text(size = 10, face = "bold"))

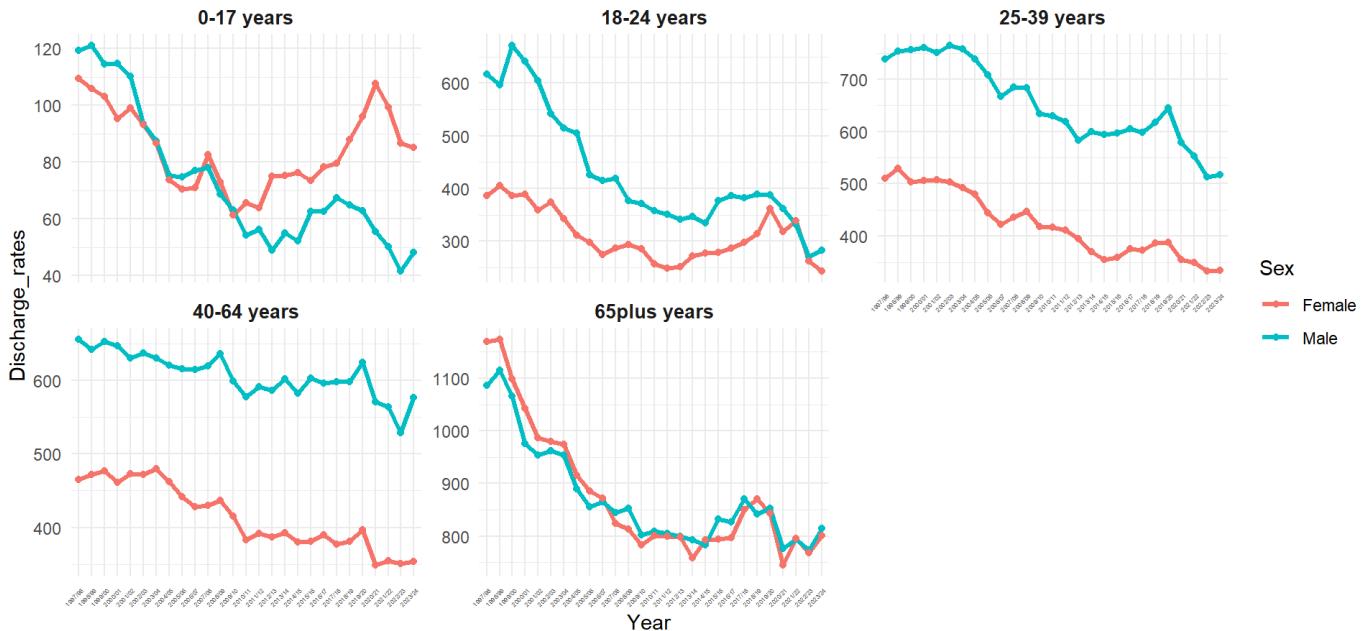
```

```

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

### Long-term Trends in Mental Health Inpatient Rates by Age Group and Sex in Scotland (1997-2023)



df\_scotland

```

## # A tibble: 270 × 13
##   financial_year hbr    hbrqf age_group sex   dataset dataset_qf discharge_count
##   <chr>          <chr>  <chr>   <chr>   <chr>  <chr>   <dbl>
## 1 1997/98        S920... d     0-17 ye... Male   Total    d           780
## 2 1997/98        S920... d     0-17 ye... Fema... Total    d           715
## 3 1997/98        S920... d     18-24 ye... Male   Total    d          2020
## 4 1997/98        S920... d     18-24 ye... Fema... Total    d          1310
## 5 1997/98        S920... d     25-39 ye... Male   Total    d          6450
## 6 1997/98        S920... d     25-39 ye... Fema... Total    d          4895
## 7 1997/98        S920... d     40-64 ye... Male   Total    d          7280
## 8 1997/98        S920... d     40-64 ye... Fema... Total    d          5395
## 9 1997/98        S920... d     65plus y... Male   Total    d          5085
## 10 1997/98       S920... d     65plus y... Fema... Total    d          7860
## # ... with 260 more rows
## # ... with 5 more variables: discharge_rates <dbl>, discharge_rates_qf <chr>,
## #   patients_count <dbl>, patients_rates <dbl>, patient_rates_qf <chr>

```

```

df_long <- df_scotland %>%
  pivot_longer(cols = c(patients_rates, discharge_rates), names_to =
    "rate_type", values_to = "rate")

```

df\_long

```

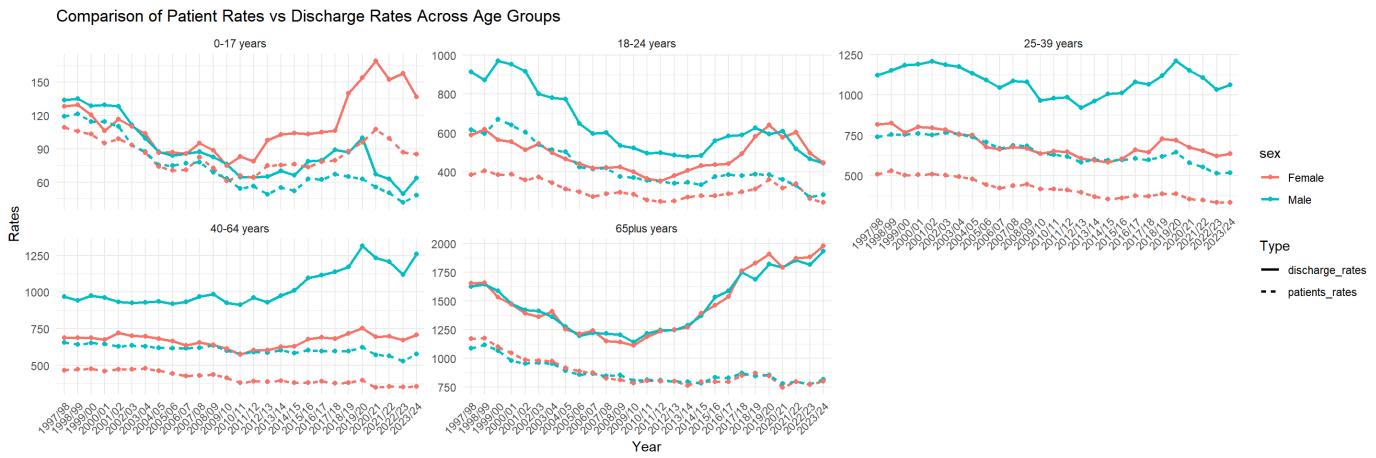
## # A tibble: 540 × 13
##   financial_year hbr    hbrqf age_group sex    dataset dataset_qf discharge_count
##   <chr>          <chr>  <chr>  <chr>  <chr>  <chr>           <dbl>
## 1 1997/98       S920... d     0-17 ye... Male   Total      d             780
## 2 1997/98       S920... d     0-17 ye... Male   Total      d             780
## 3 1997/98       S920... d     0-17 ye... Fema... Total      d             715
## 4 1997/98       S920... d     0-17 ye... Fema... Total      d             715
## 5 1997/98       S920... d     18-24 ye... Male   Total      d            2020
## 6 1997/98       S920... d     18-24 ye... Male   Total      d            2020
## 7 1997/98       S920... d     18-24 ye... Fema... Total      d            1310
## 8 1997/98       S920... d     18-24 ye... Fema... Total      d            1310
## 9 1997/98       S920... d     25-39 ye... Male   Total      d            6450
## 10 1997/98      S920... d     25-39 ye... Male   Total      d            6450
## # ... with 530 more rows
## # ... with 5 more variables: discharge_rates_qf <chr>, patients_count <dbl>,
## #   patient_rates_qf <chr>, rate_type <chr>, rate <dbl>

```

```

ggplot(df_long, aes(x = financial_year, y= rate, color = sex, linetype =
rate_type, group = interaction(sex,rate_type)))+
  geom_line(size =1)+ 
  geom_point(size=1.5)+ 
  facet_wrap(~age_group, scales = "free_y")+
  labs(
    title="Comparison of Patient Rates vs Discharge Rates Across Age Groups",
    x= "Year" , y="Rates",
    color="sex",linetype ="Type"
  ) +
  theme_minimal() +
  theme(axis.text.x=element_text(angle =45,hjust =1))

```



```
df_scotland
```

```

## # A tibble: 270 × 13
##   financial_year hbr    hbrqf age_group sex    dataset dataset_qf discharge_count
##   <chr>          <chr>  <chr>  <chr>  <chr>  <chr>           <dbl>
## 1 1997/98       S920... d     0-17 ye... Male   Total    d      780
## 2 1997/98       S920... d     0-17 ye... Fema... Total    d      715
## 3 1997/98       S920... d     18-24 ye... Male   Total    d     2020
## 4 1997/98       S920... d     18-24 ye... Fema... Total    d     1310
## 5 1997/98       S920... d     25-39 ye... Male   Total    d     6450
## 6 1997/98       S920... d     25-39 ye... Fema... Total    d     4895
## 7 1997/98       S920... d     40-64 ye... Male   Total    d     7280
## 8 1997/98       S920... d     40-64 ye... Fema... Total    d     5395
## 9 1997/98       S920... d     65plus y... Male   Total    d     5085
## 10 1997/98      S920... d     65plus y... Fema... Total    d     7860
## # ℹ 260 more rows
## # ℹ 5 more variables: discharge_rates <dbl>, discharge_rates_qf <chr>,
## #   patients_count <dbl>, patients_rates <dbl>, patient_rates_qf <chr>

```

```

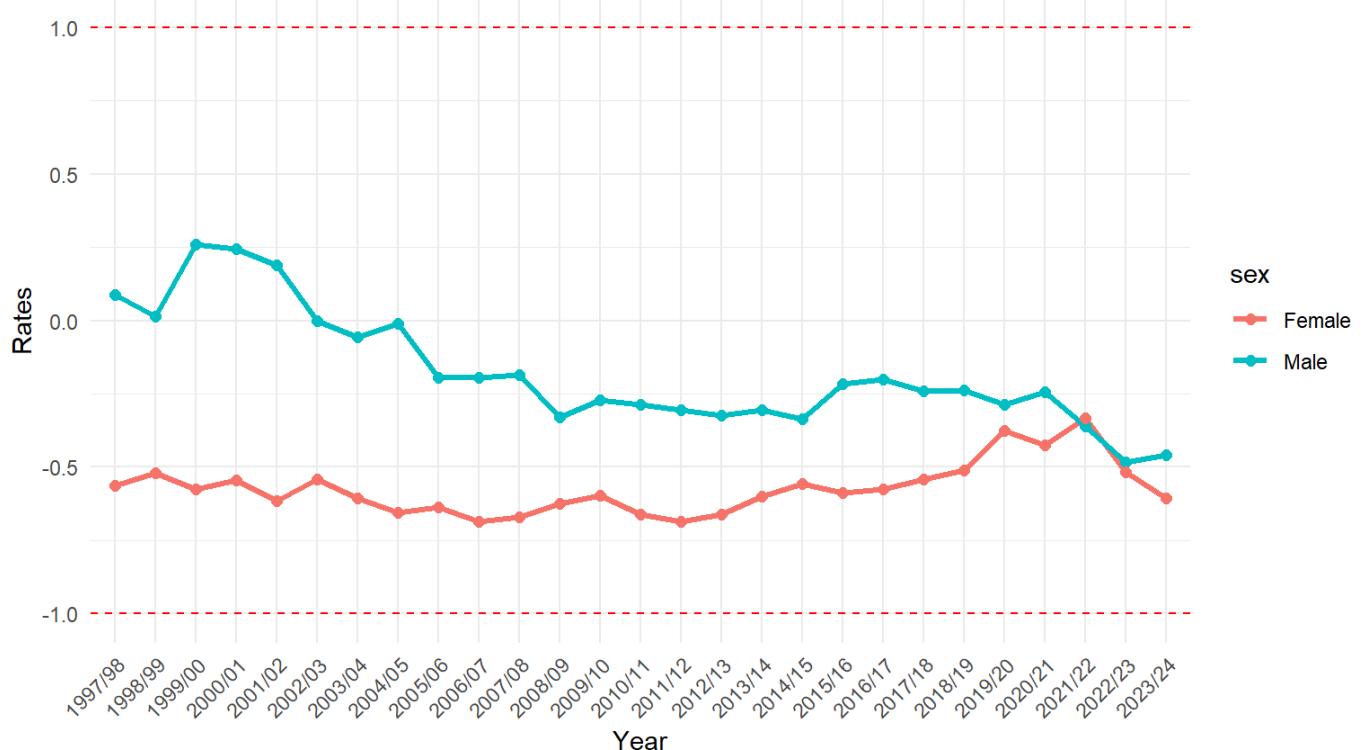
df_z <- df_scotland %>%
  group_by(financial_year) %>%
  mutate(
    z_score = (patients_rates - mean(patients_rates)) / sd(patients_rates)
  ) %>%
  filter(age_group == "18-24 years")

threshold <- 1

ggplot(df_z, aes(x = financial_year, y = z_score, group = sex, color = sex)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  geom_hline(yintercept = c(-threshold, threshold), color = "red", linetype = "dashed") +
  labs(
    title = "Z-Score Analysis of Mental Health Inpatient Rates for Young Adults (18-24 Years) Relative to All Age Groups",
    subtitle = "No Significant Difference from Other Age Groups",
    x= "Year" , y="Rates",
  ) +
  theme_minimal() +
  theme(axis.text.x=element_text(angle =45,hjust =1))

```

Z-Score Analysis of Mental Health Inpatient Rates for Young Adults (18-24 Years) Relative to A  
No Significant Difference from Other Age Groups



```
df_gender <- df_scotland %>%
  filter(age_group == "18-24 years") %>%
  pivot_wider(
    id_cols = financial_year, names_from = sex, values_from = patients_rates) %>%
  mutate(
    abs_diff = abs(Male - Female),
    rel_diff = Male / Female * 100
  )

df_gender_scaled <- df_gender %>%
  mutate(
    rel_diff_scaled = rel_diff / max(rel_diff) * max(abs_diff)
  ) %>%
  select(financial_year, abs_diff, rel_diff, rel_diff_scaled)

df_gender
```

```
## # A tibble: 27 × 5
##   financial_year Male Female abs_diff rel_diff
##   <chr>        <dbl>  <dbl>    <dbl>    <dbl>
## 1 1997/98       617.   387.    230.    160.
## 2 1998/99       596.   406.    191.    147.
## 3 1999/00       671.   386.    285.    174.
## 4 2000/01       641.   389.    252.    165.
## 5 2001/02       605.   360.    246.    168.
## 6 2002/03       542.   374.    168.    145.
## 7 2003/04       515.   343.    171.    150.
## 8 2004/05       504.   312.    192.    162.
## 9 2005/06       426.   298.    128.    143.
## 10 2006/07      415.   274.    141.    151.
## # ℹ 17 more rows
```

df\_gender\_scaled

```
## # A tibble: 27 × 4
##   financial_year abs_diff rel_diff rel_diff_scaled
##   <chr>        <dbl>    <dbl>          <dbl>
## 1 1997/98       230.    160.        262.
## 2 1998/99       191.    147.        241.
## 3 1999/00       285.    174.        285.
## 4 2000/01       252.    165.        271.
## 5 2001/02       246.    168.        276.
## 6 2002/03       168.    145.        238.
## 7 2003/04       171.    150.        246.
## 8 2004/05       192.    162.        265.
## 9 2005/06       128.    143.        234.
## 10 2006/07      141.    151.        248.
## # ℹ 17 more rows
```

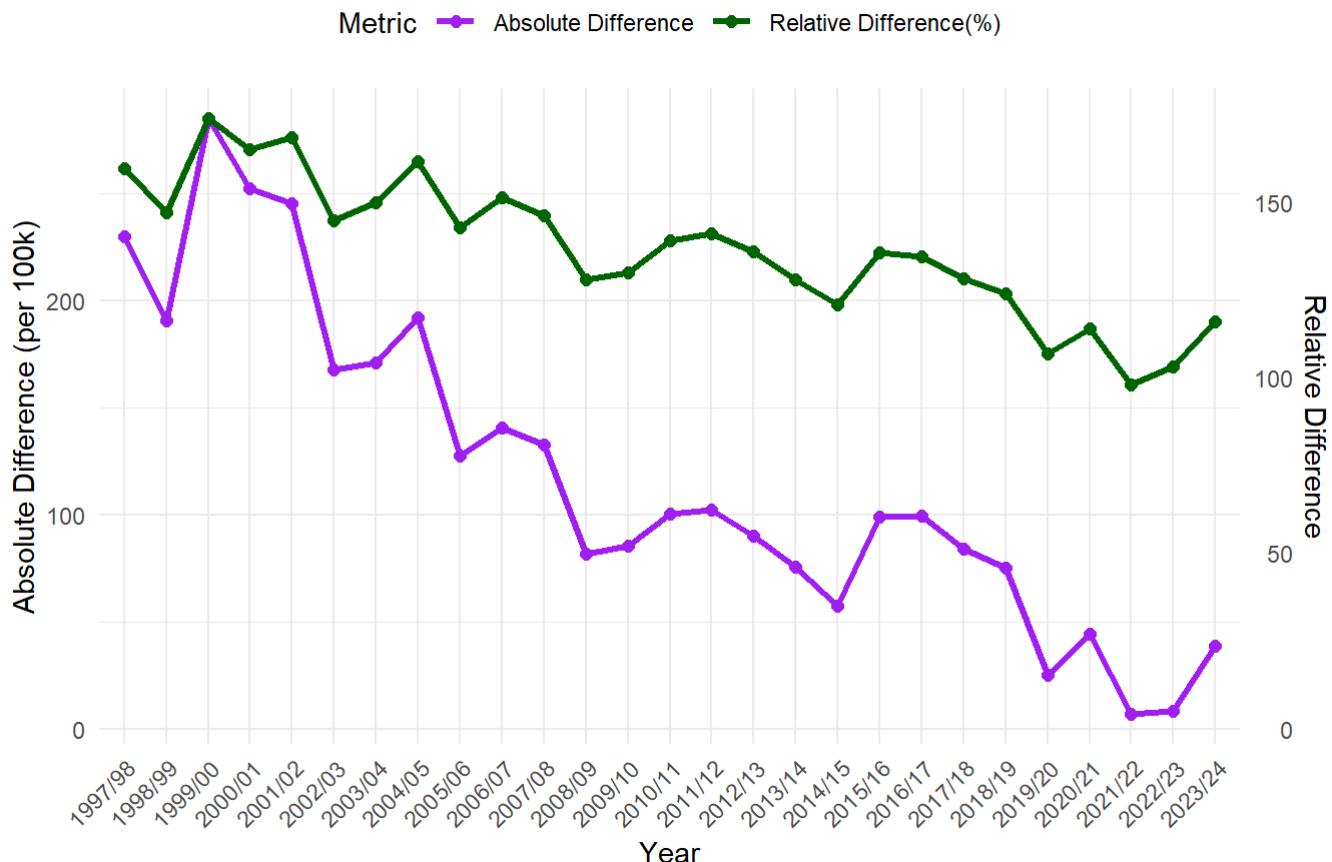
```

df_plot <- df_gender_scaled %>%
  pivot_longer(cols = c(abs_diff, rel_diff_scaled),
               names_to = "metric",
               values_to = "value"
  )

ggplot(df_plot, aes(x= financial_year, y= value, color = metric, group = metric)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  scale_color_manual(values = c("abs_diff" = "purple", "rel_diff_scaled"="darkgreen"),
                     labels = c ("Absolute Difference", "Relative Difference(%)" ),
                     name = "Metric"
  ) +
  scale_y_continuous(
    name = "Absolute Difference (per 100k)",
    sec.axis = sec_axis(
      ~ . * max(df_gender$rel_diff) /max(df_gender$abs_diff),
      name = "Relative Difference"))
  +
  labs(
    title = "absolute difference vs relative difference",
    x = "Year"
  ) + theme_minimal() + theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "top")

```

## absolute difference vs relative difference



```

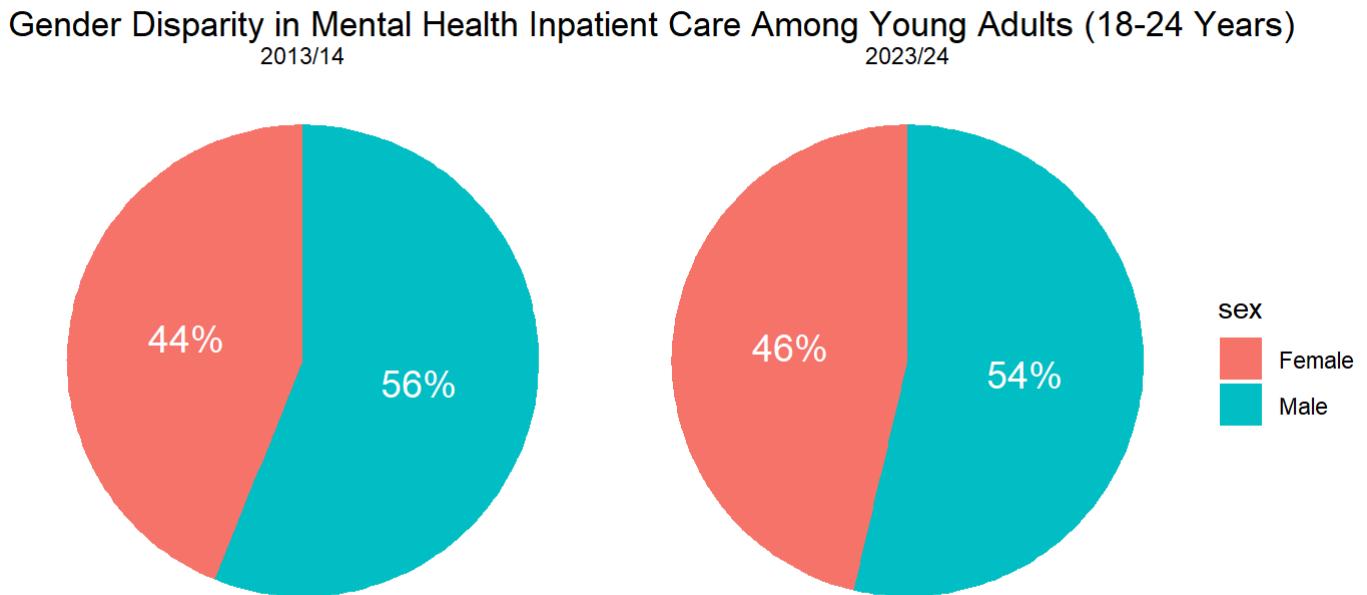
df_gender_end <- df_gender %>%
  filter(financial_year %in% c("2013/14", "2023/24")) %>%
  pivot_longer(cols = c(Male, Female), names_to = "sex", values_to = "patients_rates")

df_gender_pie <- df_gender_end %>%
  group_by(financial_year) %>%
  mutate(percentage = patients_rates / sum(patients_rates),
         label = scales::percent(percentage, accuracy = 1))

ggplot(df_gender_pie, aes(x = "", y = percentage, fill = sex)) +
  geom_col(width = 1) +
  coord_polar("y") +
  geom_text(aes(label = label), position = position_stack(vjust = 0.5), size = 5, color = "white") +
  facet_wrap(~ financial_year) +
  labs(
    title = "Gender Disparity in Mental Health Inpatient Care Among Young Adults (18-24 Years)", fill = "sex")
  ) + theme_void() + theme(stripp.text = element_text(size = 12, face = "bold"))

```

```
## Warning in plot_theme(plot): The `stripp.text` theme element is not defined in
## the element hierarchy.
```



## ## Research Question

This study examines patterns of mental health inpatient care in Scotland from 1997/98 to 2021/22, with a focused investigation into the 18–24 age group—a formative life stage marked by increased vulnerability. The research is structured around four interrelated questions that progress from broad trends to specific disparities.

The first question establishes a baseline by analysing how overall inpatient and discharge rates have changed across all age groups over the past two decades. This population-level overview reveals long-term shifts in service utilisation, contextualising broader social, economic, and policy developments affecting mental health provision.

The second question identifies periods when inpatient rates for 18–24-year-olds significantly diverged from other age groups. Such divergence may indicate times when this cohort's mental health needs or access to care were disproportionately affected by societal pressures, economic conditions, or policy changes, highlighting moments of heightened vulnerability.

The third question investigates gender disparities within the 18–24 group, comparing male and female patient rates over time. This analysis explores whether gender influences help-seeking behaviour, diagnosis, or hospitalization patterns, contributing to the development of more equitable mental health strategies.

Finally, the fourth question analyses the proportion of inpatient episodes for young adults originating from general (SMR01) versus psychiatric (SMR04) records. This distinction offers insights into care pathways, indicating whether young people more often access care through emergency/general services rather than specialist psychiatric routes, and thus revealing the functioning of early intervention systems.

Together, these questions build a comprehensive, layered analysis of how demographic, systemic, and social factors intersect to influence mental health inpatient care for young adults in Scotland.

## Data

Briefly describe your dataset here - include a description of the key variables you will be investigating. Include a citation for your data. See <http://libraryguides.vu.edu.au/c.php?g=386501&p=4347840> (<http://libraryguides.vu.edu.au/c.php?g=386501&p=4347840>) for guidance on proper citation for datasets. If you got your data off the web, make sure to state the retrieval date.

This study utilizes the “Age and Sex” subset of Scotland’s Mental Health Inpatient Activity data, published by Public Health Scotland. This dataset covers a 25-year period (1997/98–2021/22) and records individuals with mental health conditions who received inpatient or day-case care in Scottish hospitals.

The data integrates two sources: general hospital admissions (SMR01) and specialist psychiatric admissions (SMR04). Key variables include Financial Year, Age Group, and Sex. The dataset provides two primary metrics: the number of unique patients discharged annually and the total number of discharges (including readmissions), both presented as crude rates per 100,000 population for comparison.

Data is provided under the Open Government Licence. Its longitudinal and demographic detail makes it valuable for analysing service use disparities. Data completeness for the latest year (2021/22) exceeded 90% nationally, and statistical measures are applied to protect patient confidentiality.

## Findings

The write-up of your project and findings go here.

Think of this as the text of your presentation with some extra detail to cover what there was not time to discuss in the presentation. This might include any assumptions you made when doing your analysis, any limitations of the work you have done or any ideas you have for future work. Feel free to split this section into subsections to make it easier to read.

The length should be roughly 1,500 words. If you want to use a word count addin, you can install this by copying and pasting the following into RStudio:

```
devtools::install_github("benmarwick/wordcountaddin", type = "source", dependencies = TRUE)
```

You will then need to restart RStudio. Once you have done that, select the text you want to count the words of, go to Addins, and select the `Word count addin`. This addin counts words using two different algorithms, but the results should be similar and as long as you're in the ballpark of 1,500 words, you're good! The addin will ignore code chunks and only count the words in prose. If you don't want to use the addin you can always copy and paste the text into Microsoft Word to do a Word Count!

You can also load your data here and present any analysis results / plots. Make sure to hide your code with `echo = FALSE` unless the point you are trying to make is about the code itself, in which case you should show your code.

## References

List any references here. You should, at a minimum, list your data source.