

Diary Entry

Lew Bo Cong

2023-10-23

Diary Entry (Week 10)

(1) What is the question that you are going to answer? (Answer: One sentence that ends with a question mark that could act like the title of your data story)

What contributes to the difference in wait times for admission to wards in local public hospitals?

(2) Why is this an important question? (Answer: 3 sentences, each of which has some evidence, e.g., “According to the United Nations...” to justify why the question you have chosen is important).

The CNA had reported in April 2023 that the median waiting time for ward admissions had increased from about 5 hours to 7.2 hours during that period¹. The health ministry had explained that it was due to the rise in number of older patients with more complex medical needs requiring longer hospitalisation¹. Due to the ageing population in Singapore, long waiting time for admission to wards has been a long-standing issue over the years, and it was further exacerbated by the Covid-19 pandemic^{2,3}.

(3) Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you plan to use)

The main dataset will be the wait times in each hospital across different dates. Supplementary dataset includes the bed occupancy rate in each hospital across different dates.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## 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
```

```
wait_times <- read_csv("../Proj dataset/wt-for-admission-to-ward_week39y2023.csv")
```

```
## Rows: 266 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (1): Date
## dbl (8): AH, CGH, KTPH, NTFGH, NUH(A), SGH, SKH, TTSH
##
## 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.
```

```
glimpse(wait_times)
```

```
## Rows: 266
## Columns: 9
## $ Date      <chr> "Sun, 01/01/23", "Mon, 02/01/23", "Tue, 03/01/23", "Wed, 04/0~
## $ AH        <dbl> 1.6, 2.0, 2.1, 2.5, 2.4, 2.8, 1.3, 1.2, 3.3, 9.0, 2.2, 2.4, 1~
## $ CGH        <dbl> 7.7, 12.7, 16.0, 16.3, 21.5, 18.3, 11.5, 5.4, 11.1, 17.8, 18.~
## $ KTPH       <dbl> 2.9, 12.2, 12.9, 21.4, 20.7, 20.0, 6.8, 2.8, 17.9, 6.2, 6.7, ~
## $ NTFGH      <dbl> 7.9, 4.2, 22.8, 14.7, 18.0, 8.2, 2.4, 2.6, 6.7, 9.7, 12.0, 9.~
## $ 'NUH(A)'   <dbl> 2.1, 2.5, 4.0, 8.2, 5.2, 6.4, 3.7, 6.1, 5.4, 5.3, 4.8, 5.4, 3~
## $ SGH        <dbl> 1.3, 2.5, 9.3, 11.6, 11.7, 6.8, 1.7, 1.7, 7.6, 6.6, 6.1, 3.1,~
## $ SKH        <dbl> 3.0, 2.7, 5.0, 10.3, 7.9, 4.8, 7.7, 5.0, 13.9, 15.0, 16.3, 5.~
## $ TTSH       <dbl> 3.8, 4.4, 4.9, 6.9, 5.0, 4.9, 5.1, 4.8, 5.4, 6.7, 6.8, 4.6, 5~
```

```
bed_occ <- read_csv("../Proj dataset/bed-occupancy-rate_week39y2023.csv")
```

```
## Rows: 2092 Columns: 10
## -- Column specification -----
## Delimiter: ","
## chr (9): Date, AH, CGH, KTPH, NTFGH, NUH(A), SGH, SKH, TTSH
## dbl (1): Years
##
## 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.
```

```
glimpse(bed_occ)
```

```
## Rows: 2,092
## Columns: 10
## $ Years      <dbl> 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2~
## $ Date       <chr> "1/1/2018", "2/1/2018", "3/1/2018", "4/1/2018", "5/1/2018", "~
## $ AH         <chr> "39.90%", "38.20%", "39.90%", "42.40%", "42.90%", "43.30%", "~
## $ CGH        <chr> "81.30%", "85.80%", "86.40%", "85.50%", "84.40%", "83.00%", "~
## $ KTPH       <chr> "96.80%", "100.00%", "99.90%", "100.00%", "100.00%", "97.20%"~
## $ NTFGH      <chr> "78.90%", "84.70%", "87.50%", "87.10%", "84.30%", "84.50%", "~
## $ 'NUH(A)'   <chr> "73.50%", "78.40%", "83.10%", "82.60%", "79.80%", "76.30%", "~
## $ SGH        <chr> "74.30%", "83.80%", "88.10%", "87.50%", "84.80%", "82.90%", "~
## $ SKH        <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ TTSH       <chr> "93.60%", "93.40%", "92.50%", "91.90%", "90.50%", "88.70%", "~
```

(4) Challenges and errors that I faced and how I overcame them.

I faced some errors when I tried to read my dataset in R. It was originally in ".xlsx" format, which resulted in a tibble of 1 x 1 and unable to display any data. After saving my excel sheet in ".csv" format, I could

read my data but its display was erroneous, because there were unnecessary table heading and spaces which disrupted the data structure crucial for a .csv file. After tidying the file by removing unnecessary portions of the table, I managed to produce the expected dataset with correct number of columns and rows.

Diary Entry (Week 9)

(1) What is the topic that you have finalized? (Answer in 1 or 2 sentences)

Due to the ageing population, the healthcare sector in Singapore is facing various challenges, i.e. bed crunch. To gain a deeper insight on this specific issue, we analyse datasets made available by the Ministry of Health.

(2) What are the data sources that you have curated so far? (Answer 1 or 2 sentences).

Two data sources have been curated.

- a) Beds Occupancy Rate. (**Link**)
- b) Waiting Time for Admission to Ward. (**Link**)