miniessay2

1.1 Simulate

```
#### Workspace setup ####
install.packages("opendatatoronto")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
(as 'lib' is unspecified)
```

```
install.packages("knitr")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
(as 'lib' is unspecified)
```

```
install.packages("janitor")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
(as 'lib' is unspecified)
```

```
install.packages("lubridate")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
(as 'lib' is unspecified)
```

```
install.packages("tidyverse")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
(as 'lib' is unspecified)
```

```
library(knitr)
library(janitor)
```

```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```r
library(lubridate)
```

```

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```r
library(opendatatoronto)
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.4.4     ✔ tibble    3.2.1
✔ purrr     1.0.2     ✔ tidyr     1.3.0
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
#### Simulate ####
set.seed(853)

simulated_occupancy_data <-
  tibble(
    date = rep(x = as.Date("2023-01-01") + c(0:364), times = 3),
    # Based on Eddelbuettel: https://stackoverflow.com/a/21502386
    shelter = c(
      rep(x = "Shelter 1", times = 365),
      rep(x = "Shelter 2", times = 365),
      rep(x = "Shelter 3", times = 365)
    ),
    location_city = sample(
      x = c("Toronto","North York","Etobicoke","Scarborough","Vaughan"),
      size = 365*3,
      replace = TRUE),
    number_occupied =
      rpois(
        n = 365 * 3,
        lambda = 30
      ) # Draw 1,095 times from the Poisson distribution
  )

head(simulated_occupancy_data)
```

```
# A tibble: 6 × 4
  date       shelter   location_city number_occupied
  <date>     <chr>     <chr>                   <int>
1 2023-01-01 Shelter 1 Toronto                    31
2 2023-01-02 Shelter 1 North York                 31
3 2023-01-03 Shelter 1 Vaughan                    23
4 2023-01-04 Shelter 1 Toronto                    27
5 2023-01-05 Shelter 1 Vaughan                    31
6 2023-01-06 Shelter 1 Vaughan                    33
```

1.2 Acquire

```
#### Acquire ####
toronto_shelters <-
  # Each package is associated with a unique id  found in the "For
  # Developers" tab of the relevant page from Open Data Toronto
  # https://open.toronto.ca/dataset/daily-shelter-overnight-service-occupancy-capacity/
  list_package_resources("21c83b32-d5a8-4106-a54f-010dbe49f6f2") |>
  # Within that package, we are interested in the 2021 dataset
  filter(name ==
    "daily-shelter-overnight-service-occupancy-capacity-2023.csv") |>
  # Having reduced the dataset to one row we can get the resource
  get_resource()

write_csv(
  x = toronto_shelters,
  file = "toronto_shelters.csv"
)
toronto_shelters$LOCATION_CITY |>
  unique()
```

```
[1] "Toronto"     "North York"  "Etobicoke"   "Scarborough" "Vaughan"
[6] ""
```

```
head(toronto_shelters)
```

```
# A tibble: 6 × 32
   X_id OCCUPANCY_DATE      ORGANIZATION_ID ORGANIZATION_NAME         SHELTER_ID
  <int> <chr>                         <int> <chr>                          <int>
1     1 2023-01-01T00:00:00              24 COSTI Immigrant Services          40
2     2 2023-01-01T00:00:00              24 COSTI Immigrant Services          40
3     3 2023-01-01T00:00:00              24 COSTI Immigrant Services          40
4     4 2023-01-01T00:00:00              24 COSTI Immigrant Services          40
5     5 2023-01-01T00:00:00              24 COSTI Immigrant Services          40
6     6 2023-01-01T00:00:00              14 Christie Ossington Neigh…         22
# ℹ 27 more variables: SHELTER_GROUP <chr>, LOCATION_ID <int>,
#   LOCATION_NAME <chr>, LOCATION_ADDRESS <chr>, LOCATION_POSTAL_CODE <chr>,
#   LOCATION_CITY <chr>, LOCATION_PROVINCE <chr>, PROGRAM_ID <int>,
#   PROGRAM_NAME <chr>, SECTOR <chr>, PROGRAM_MODEL <chr>,
#   OVERNIGHT_SERVICE_TYPE <chr>, PROGRAM_AREA <chr>, SERVICE_USER_COUNT <int>,
#   CAPACITY_TYPE <chr>, CAPACITY_ACTUAL_BED <int>, CAPACITY_FUNDING_BED <int>,
#   OCCUPIED_BEDS <int>, UNOCCUPIED_BEDS <int>, UNAVAILABLE_BEDS <int>, …
```

```
# List all column names in the dataframe
names(toronto_shelters)
```

```
 [1] "X_id"                  "OCCUPANCY_DATE"         "ORGANIZATION_ID"
 [4] "ORGANIZATION_NAME"     "SHELTER_ID"             "SHELTER_GROUP"
 [7] "LOCATION_ID"           "LOCATION_NAME"          "LOCATION_ADDRESS"
[10] "LOCATION_POSTAL_CODE"  "LOCATION_CITY"          "LOCATION_PROVINCE"
[13] "PROGRAM_ID"            "PROGRAM_NAME"           "SECTOR"
[16] "PROGRAM_MODEL"         "OVERNIGHT_SERVICE_TYPE" "PROGRAM_AREA"
[19] "SERVICE_USER_COUNT"    "CAPACITY_TYPE"          "CAPACITY_ACTUAL_BED"
[22] "CAPACITY_FUNDING_BED"  "OCCUPIED_BEDS"          "UNOCCUPIED_BEDS"
[25] "UNAVAILABLE_BEDS"      "CAPACITY_ACTUAL_ROOM"   "CAPACITY_FUNDING_ROOM"
[28] "OCCUPIED_ROOMS"        "UNOCCUPIED_ROOMS"       "UNAVAILABLE_ROOMS"
[31] "OCCUPANCY_RATE_BEDS"   "OCCUPANCY_RATE_ROOMS"
```

```
toronto_shelters_clean <- toronto_shelters %>%
  select(OCCUPANCY_DATE, LOCATION_CITY, OCCUPIED_BEDS)

head(toronto_shelters_clean)
```

```
# A tibble: 6 × 3
  OCCUPANCY_DATE      LOCATION_CITY OCCUPIED_BEDS
  <chr>               <chr>                 <int>
1 2023-01-01T00:00:00 Toronto                  NA
2 2023-01-01T00:00:00 Toronto                  NA
3 2023-01-01T00:00:00 Toronto                   8
4 2023-01-01T00:00:00 North York               NA
5 2023-01-01T00:00:00 North York               NA
6 2023-01-01T00:00:00 Etobicoke                NA
```

1.3 Explore

```
write_csv(
  x = toronto_shelters_clean,
  file = "cleaned_toronto_shelters.csv"
)
```

```
#### Explore ####
toronto_shelters_clean <-
```

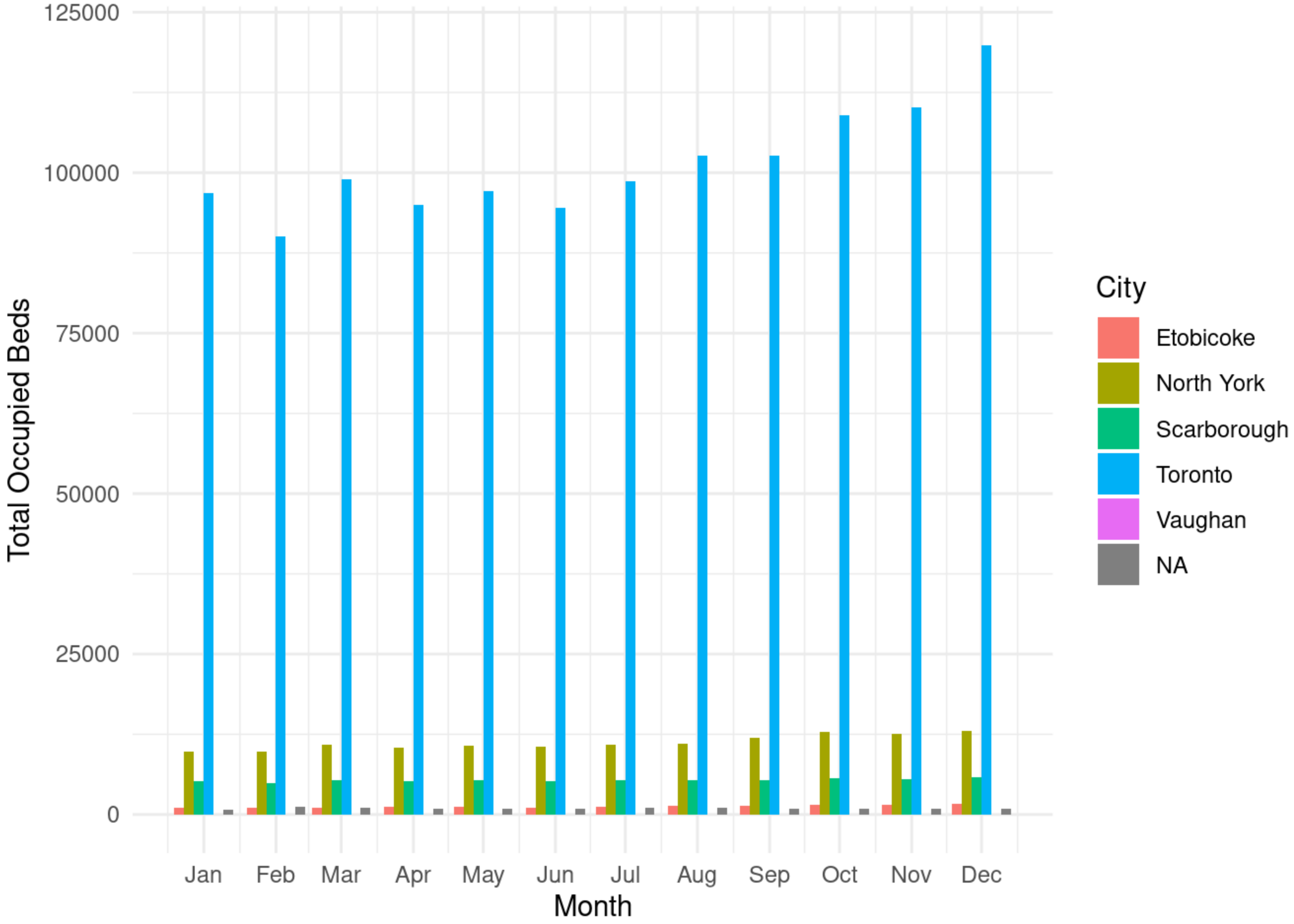
```r
  read_csv(
    "cleaned_toronto_shelters.csv",
    show_col_types = FALSE
  )
```

```r
toronto_shelters_clean <- toronto_shelters_clean %>%
  mutate(Month = floor_date(as.Date(OCCUPANCY_DATE), "month"))

monthly_beds <- toronto_shelters_clean %>%
  group_by(Month, LOCATION_CITY) %>%
  summarize(Total_Beds = sum(OCCUPIED_BEDS, na.rm = TRUE))
```

```
`summarise()` has grouped output by 'Month'. You can override using the
`.groups` argument.
```

```r
ggplot(monthly_beds, aes(x = Month, y = Total_Beds, fill = LOCATION_CITY)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(x = "Month", y = "Total Occupied Beds", fill = "City") +
  scale_x_date(date_breaks = "1 month", date_labels = "%b")
```



1.4 Share

Toronto has a sizable homeless population. Because of the harsh winters, it is vital that shelters have ample space. We want to know how the use of shelters differs in the colder months against the warmer months, and we want to know which city has the most demand for shelters. We use statistics on Toronto shelter bed occupancy given by the City of Toronto. Each night at 4 a.m., a count of the occupied beds is made. We're curious in the overall amount of these across the month and city. We cleaned, tidied, and analyzed the dataset using R (R Core Team 2023) as well as the `tidyverse` (Wickham 2017), `janitor` (Firke 2023), `opendatatoronto` (Gelfand 2022), `lubridate` (Grolemund and Wickham 2011), and `knitr` (Xie 2023).

In analyzing the shelter bed occupancy data for Toronto, a clear pattern emerges: Toronto consistently shows the highest number of occupied shelter beds, with North York ranking second. As the months progress from January to December, there is a noticeable increase in shelter bed occupancy. This trend is congruent with the intuitive notion that colder months drive a higher demand for shelter services. The data underscores the importance of ensuring adequate shelter capacity to meet the rising needs as winter approaches, and it highlights the need for targeted strategies to manage the seasonal influx in shelter usage.