Edit New issue Jump to bottom

Code Review #1

Open julia-ve-kim opened this issue 49 minutes ago · 0 comments

julia-ve-kim commented 49 minutes ago • edited Collaborator

food_bank_usage_toronto

The name of the GitHub Repo is descriptive, immediately giving a sense of its contents & code. You might wish to make this repository *public*, so that the graders are able to examine the work and this GitHub Issue, and revert it to *private* once the grading is finalised. In the **About** section of the Repo, you may wish to put a brief description such as "Contains a reproducible Quarto file that downloads a dataset about food bank usage in Toronto, cleans it and makes a graph...".

README.md

The goals and structure of the repository were detailed well in the `README.md`. Though the structure of the GitHub Repo is sufficiently clear, you might wish to

- file `toronto_foodbank_paper.pdf` and `toronto_foodbank_paper.qmd` under `/outputs/paper`
- file `figures` under `/inputs`

I am not certain whether the `.gitignore` file is necessary for this repository.

toronto_foodbank_paper.pdf

It is helpful that your paper begins with an abstract, giving a high-level overview of the purpose of the research, its motives and its results. The date at the top is left "invalid"– you might wish to double check this.

> The project closely follows methods mentioned in *Telling Stories with Data* Alexander (2024).

You should revisit your `@drinking_from` in `bibliography.bib` to ensure that this citation is done as (Alexander 2023). What worked for me was:

```
@book{Rohan,
  title     = "Telling Stories with Data with Applications in R",
  author    = "Alexander, Peter",
  year      = 2023,
  publisher = "Chapman & Hall",
  isbn = "9781032134772",
  url = {https://tellingstorieswithdata.com/},
  }
```

> I have used R version 4.3.2. and a library called Tidyverse to create this simulation (R Core Team) (Wickham et al. 2019).

Proper in-text citation syntax would require "I have used R version 4.3.2. (R Core Team) and a library called Tidyverse (Wickham et al. 2019) to create this simulation."

> We downloaded the dataset from Toronto's data dashboard (Toronto 2023). ... It could be that Canada is growing in population at its highest rate since 1957 (Canada 2023).

Your **Share** summarises well the purpose, results and conclusions of the research, but again there are a few mistakes with citation. Toronto should really be "City of Toronto" and Canada "Statistics Canada." In `citations.bib`, you can avoid this mishap by putting `name = {{City of Toronto}}` in lieu of `name = {City of Toronto}` and `name = {{Statistics Canada}}` in lieu of `name = {Statistics Canada}`. Also, `n.d` (standing for `no date`) should not be the name of the author–– According to EasyBib, "when there isn't a known author, use the source's title in the in-text citation", so you should put instead `Daily Bread Food Bank`.

> Data was cleaned using programming language R (R Core Team 2023) and a package in R called `tidyverse` (Wickham et al. 2019).

You should also reference your use of the `janitor` and `lubridate` packages.

/scripts/00-simulate_data.R

```
#### Workspace setup ####
# Load packages
library(tidyverse)
```

You should install and load *all* packages you need to use in the Workspace setup. In this case, you should also include `library(janitor)`, `library(lubridate)`.

/scripts/01-clean_data.R

You might wish to separate this `.R` script into two chunks, one responsible for downloading the data and the other for cleaning the data, with each script ending by saving the data as a `.csv` file in the desired folder.

```
  clean_names(raw_data) |>
  mutate(period = recode(period,
   Jan = 1,
   Feb = 2,
   Mar = 3,
   Apr = 4,
   May = 5,
   Jun = 6,
   Jul = 7,
   Aug = 8,
   Sep = 9,
   Oct = 10,
   Nov = 11,
   Dec = 12
  )) |>
  mutate(date = make_date(year, period, 1)) |>
  select(date, visits = value) |>
  filter(as.Date("2013-09-01") <= date & date <= as.Date("2023-09-01"))
```

This code looks very good, but you might wish to consider making it more efficient. A possible suggestion might be:

```
  clean_names() |>
  mutate(
    period = match(tolower(period), tolower(month.abb)),
    date = make_date(year, period, 1)
  ) |>
  select(date, visits = value) |>
  filter(between(date, as.Date("2013-09-01"), as.Date("2023-09-01"))).
```

Here, I combined the two `mutate` calls into one and used `between` in the `filter` call for better readability. To avoid writing out each month individually, I used a combination of the `month.abb` vector and the `match` function to achieve the same result–– when re-run, it should produce a similar result!

> head(cleaned_data)

`head` is helpful as a sanity check, but may result in the formal paper being cluttered. You might wish to uncomment this line. If you are displaying the `cleaned_data` as a table for the readers, you can format it more nicely using `kable` (see, for example, Example 2.3 of *Telling Stories with Data*).

/scripts/02-explore_data.R

Something that *may* be useful before plotting your data as a graph is to handle possibly missing values in the dataset. You can do this as follows:

```
# Check for missing values
summary(cleaned_data)

# If needed, handle missing values
cleaned_data <- na.omit(cleaned_data)
```

Also, you appear to be plotting the monthly visits to foodbanks in Toronto.

```
cleaned_data |>
  ggplot(aes(x = date, y = visits)) +
  geom_point() +
  labs(x = "Date",
```

```
      y = "Monthly Visits to the Food Bank") +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma_format()) +
  expand_limits(y = 0)
```

However, when this graph is displayed, only *years* (2014, 2016, 2018, 2020, 2022, 2024) appear on the *x*-axis of the plot–– you need instead individual months (Jan, Feb, March...) to appear. You might wish to consider how individual months can appear on the *x*-axis for the years 2013-2023 without the graph becoming overcluttered.

Assignees

No one—assign yourself

Labels

None yet

Projects

None yet

Milestone

No milestone

Development

Create a branch for this issue or link a pull request.

1 participant

Pin issue