Getting Hitched: Torontonians Prefer to Get Married During Warm Weather*

Diana Liu

2024-01-25

We examine marriage licenses obtained in the city of Toronto in the year 2023 by months. By dividing the year into warm months (April, May, June, July, August, September) and cold months (October, November, December, January, Febuary, March), the number of marriage licenses obtained in Toronto increase significantly during warm months, supporting the hypothesis that Torontonians prefer to get married during warm weather.

1 Introduction

You can and should cross-reference sections and sub-sections.

The remainder of this paper is structured as follows. Section 2….

2 Data

Our data is Marriage Licence Statistics from Open Data Toronto (Gelfand (2022)) which contains data sets maintained by the City of Toronto. The Marriage Licence Statistics dataset consists of monthly number of marriage licenses obtained from 2011 to 2023 in the Greater Toronto Area, it was last updated 2024-01-14. No similar data sets could have been used because this dataset is the only one of marriage licences. The Marriage Licence Statistics dataset contains only four variables: x_id, civic centres, marriage licenses, and date, but only civic centres, marriage licenses, and date were used in our analysis. The data was cleaned and analysed using the statistical software R (R Core Team (2022)) with the assistance of tidyverse (Wickham et al. (2019)), dyplr (Wickham et al. (2023)), and stringr (Wickham (2023)).

*Code and data are available at: https://github.com/Diana-Guanzhi-Liu/Term-Paper-1.

2.1 Civic Centres

CIVIC_CENTRE is a column a two letter abbreviation of the city in which the marriage licenses are issued, ET for Etobicoke NY for North York, SC for Scarborough, and TO for Toronto. Each civic centre has its own row of corresponding date, number of marriage licenses, and id. During the data cleaning process, we filter out ET, NY, and SC so we can focus on Toronto. After filtering, we will not need the variable anymore.

```
# A tibble: 6 x 4
   X_id CIVIC_CENTRE MARRIAGE_LICENSES TIME_PERIOD
  <dbl> <chr>                    <dbl> <chr>
1 11101 ET                          80 2011-01
2 11102 NY                         136 2011-01
3 11103 SC                         159 2011-01
4 11104 TO                         367 2011-01
5 11105 ET                         109 2011-02
6 11106 NY                         150 2011-02
```

```
spc_tbl_ [528 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ X_id             : num [1:528] 11101 11102 11103 11104 11105 ...
 $ CIVIC_CENTRE     : chr [1:528] "ET" "NY" "SC" "TO" ...
 $ MARRIAGE_LICENSES: num [1:528] 80 136 159 367 109 150 154 383 177 231 ...
 $ TIME_PERIOD      : chr [1:528] "2011-01" "2011-01" "2011-01" "2011-01" ...
 - attr(*, "spec")=
  .. cols(
  ..   X_id = col_double(),
  ..   CIVIC_CENTRE = col_character(),
  ..   MARRIAGE_LICENSES = col_double(),
  ..   TIME_PERIOD = col_character()
  .. )
 - attr(*, "problems")=<externalptr>
```

2.2 Marriage Licenses

MARRIAGE_LICENSES is the variable denoting how many licenses were obtained each month, with a max of 1649 certificates obtained, min of 563, and average of 1058.8 in Toronto in 2023. Marriage licenses are obtained in the province of Ontario via an online application process followed by an in-person appointment at the Toronto City Hall. The actual wedding ceremony is irrelevant for obtaining a marriage license.

```
# A tibble: 6 x 2
  Month      Marriages
  <date>         <dbl>
1 2023-01-01       563
2 2023-02-01       617
3 2023-03-01       929
4 2023-04-01      1086
5 2023-05-01      1507
6 2023-06-01      1507
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  563.0   639.5  1007.5  1058.8  1507.0  1649.0
```

2.3 Date (Month)

TIME_PERIOD is the month in which the marriage licenses were obtained in a yyyy-mm date format. When we clean the data, the date is filtered by year to retain only 2023 information, then it is reformatted so that it is the name of the month (e.g. 'January')

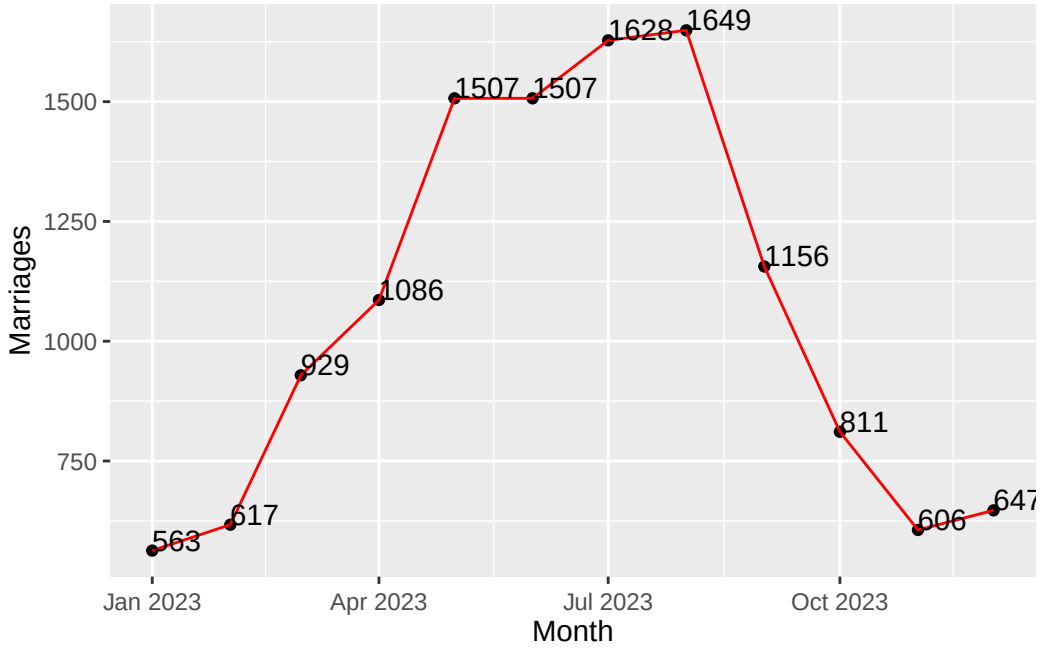
3 Results

```
# A tibble: 12 x 2
   Month      Marriages
   <date>         <dbl>
 1 2023-01-01       563
 2 2023-02-01       617
 3 2023-03-01       929
 4 2023-04-01      1086
 5 2023-05-01      1507
 6 2023-06-01      1507
 7 2023-07-01      1628
 8 2023-08-01      1649
 9 2023-09-01      1156
10 2023-10-01       811
11 2023-11-01       606
12 2023-12-01       647
```

```
Rows: 12 Columns: 2
-- Column specification --------------------------------------------------------
Delimiter: ","
```

```
dbl  (1): Marriages
date (1): Month

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.
```



4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

References

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2023. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://stringr.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.