

Contents

MAJOR CRIME INDICATORS IN TORONTO 2019	2
Introduction	2
The Data	2
Data Description	2
Research Questions	4
At what time most crimes happen in Toronto?	4
What is the relation between time of the day and crime type?	4
What are the top 10 Assault subgroups	4
Where in the city these crimes are most likely to happen?	5
Toronto's Crime Map	7
Conclusion	8
References	9

MAJOR CRIME INDICATORS IN TORONTO 2019

This document is an analysis of crimes occurred in Toronto in 2019 by the ‘Toronto Police Department’ (Public Safety Data Portal 2020) with a data description by ‘Toronto Open Data’ (Toronto, n.d.).

It was analyzed using R (R Core Team 2020), the ‘tidyverse’ package written by (Wickham et al. 2019), ‘dplyr’ package written by (Wickham et al. 2021), and ‘ggplot2’ package written by (Wickham 2016), and ‘ggmap’ written by (Kahle and Wickham 2013).

Introduction

This paper brings to light all Major Crime Indicators (MCI) from the original homonymous dataset with data between the years of 2014 and 2019 by reported date and related offences. However, the analysis will focus on the year of 2019 as the most recent related cases.

Major Crime Indicators (MCI) Dashboards and other information are available at Major Crime Indicators - Toronto 2014-2019.

The reported crime dataset is intended to provide communities with information regarding public safety and awareness. The data supplied to the Toronto Police Service by the reporting parties is preliminary and may not have been fully verified.

At the end, there is a showcase of clustering results over a map of the city of Toronto. The plot tries to represent the types of crimes that happened at different times of the day based on geographical locations.

The Data

Data analysis steps were performed to find the best temporal and spatial representation of crimes overtime. The results show a comparison against other regions and time frame for a better overview.

Data Description

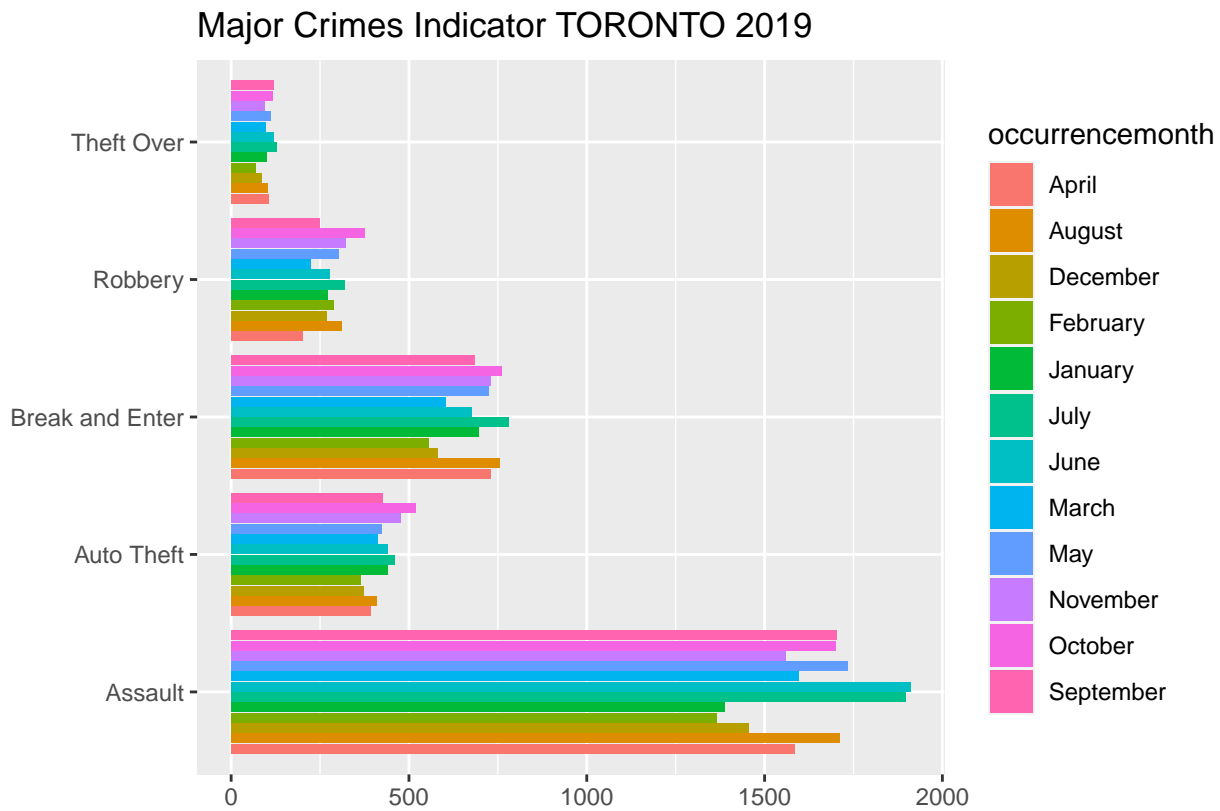
The Toronto Police Major Crime Indicators Data was collected between the years of 2014 and 2019 and, after the deletion of X, and Y features, it is distributed as follows:

1. *Index_*: Record Unique Identifier
2. *event_unique_id*: Event Unique Identifier
3. *occurrencedate*: Date of occurrence
4. *reporteddate*: Date occurrence was reported
5. *premisetype*: Premise where occurrence took place
6. *ucr_code*: URC Code
7. *ucr_ext*: URC Code Extension
8. *offence*: Offence related to the occurrence
9. *reportedyear*: Year occurrence was reported
10. *reportedmonth*: Month occurrence was reported
11. *reportedday*: Day occurrence was reported
12. *reporteddayofyear*: Day of year Occurrence was reported.
13. *reporteddayofweek*: Day of week occurrence was reported
14. *reportedhour*: Hour occurrence was reported
15. *occurrenceyear*: Occurrence year
16. *occurrencemonth*: Occurrence month
17. *occurrenceday*: Occurrence day

18. *occurrencedayofyear*: Occurrence day of year
19. *occurrencedayofweek*: Occurrence day of week
20. *occurrencehour*: Occurrence hour
21. ****MCI**: Major Crime Indicator related to the offence
22. *Division*: Division where event occurred
23. *Hood_ID*: Neighbourhood Name
24. *Neighborhood*: Neighborhood Identifier
25. *Long*: Longitude of point extracted after offsetting X and Y Coordinates to nearest intersection node
26. *Lat*: Latitude of point extracted after offsetting X and Y Coordinates to nearest intersection node
27. *ObjectId*: Id of the report

***MCI** categories included in this analysis are: **Assault, Break and Enter, Auto Theft, Robbery and Theft Over.**

This study does not rely on the analysis of any specific time frame and to speed up the process, we will be analyzing the data of occurrences throughout the year of 2019. The **reporteddate** variable might present either the same period or sometime in the future. However, it does not compromise the accuracy of the results.



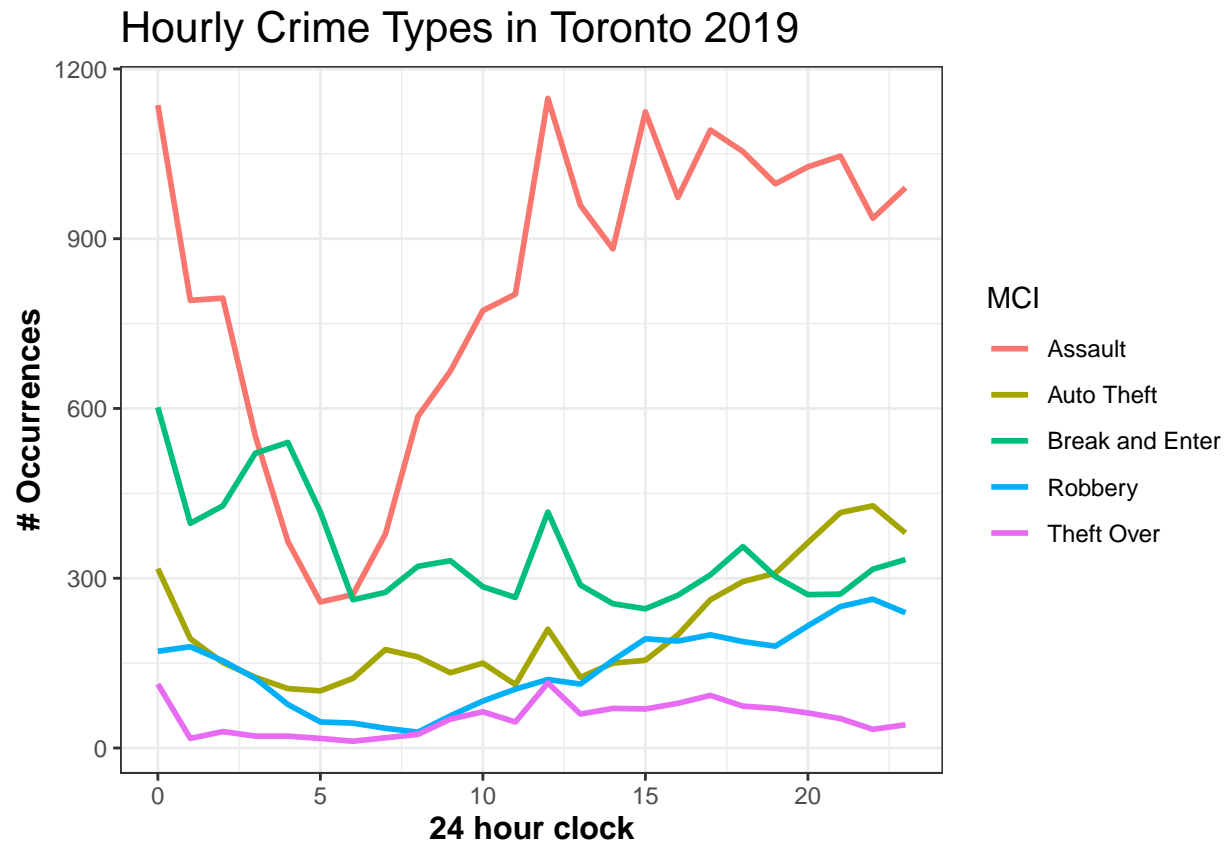
Assault:19600; Auto-theft:5137; Break and Enter:8278; Robbery:3408; Theft-over:1251

Assault is undoubtedly the most prevalent form of violent crime throughout the year in Toronto, with no exception. It appears to have a few more assault incidents in June and July, with January and February holding the lowest number of occurrences.

It is not clear if seasons of the year are related to the number of occurrences and this analysis only considers one-year worth data which, in that case, might not be enough to draw further conclusions about seasonality correlation.

Research Questions

At what time most crimes happen in Toronto?



We are analyzing the variable **occurrencehour** and **MCI** to check the correlation between hours of the day and the number of crimes. It seems that the worst hours are at noon and the hours between 11 pm and 3 am. Hence, supporting this statement we can see above crimes that are most frequent and the respective hour of the day.

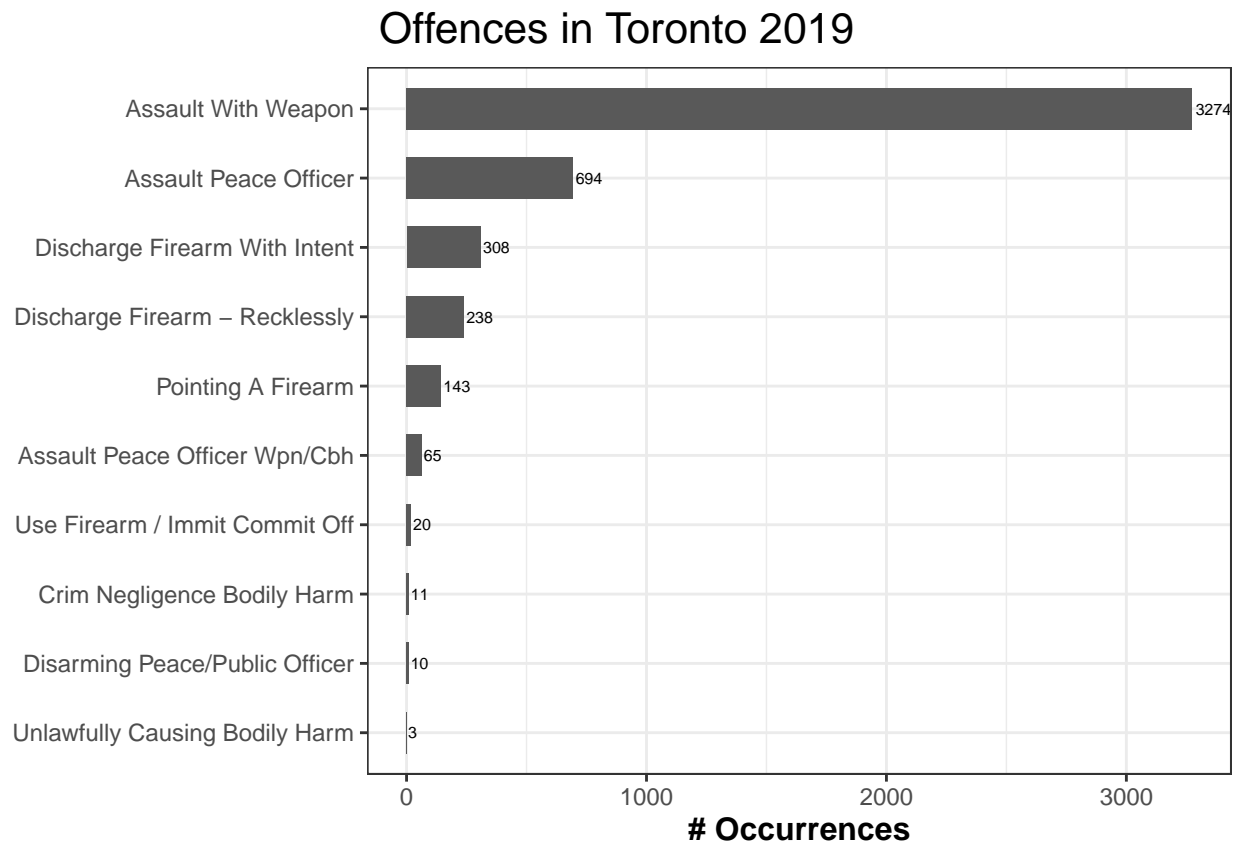
What is the relation between time of the day and crime type?

Assaults are the most common crimes at all times, happening more frequently at early afternoon and late at night. Conversely, Break and Enter get the top position in the morning when, technically, most people are sleeping followed by another peak at noon when people are out picking their kids at school or doing groceries for lunch. Auto theft is another crime type that reaches a peak, but not before eight at night when many cars have just parked, and drivers are distracted by family and unpacking the trunk. All crime patterns and its timeline make sense.

What are the top 10 Assault subgroups

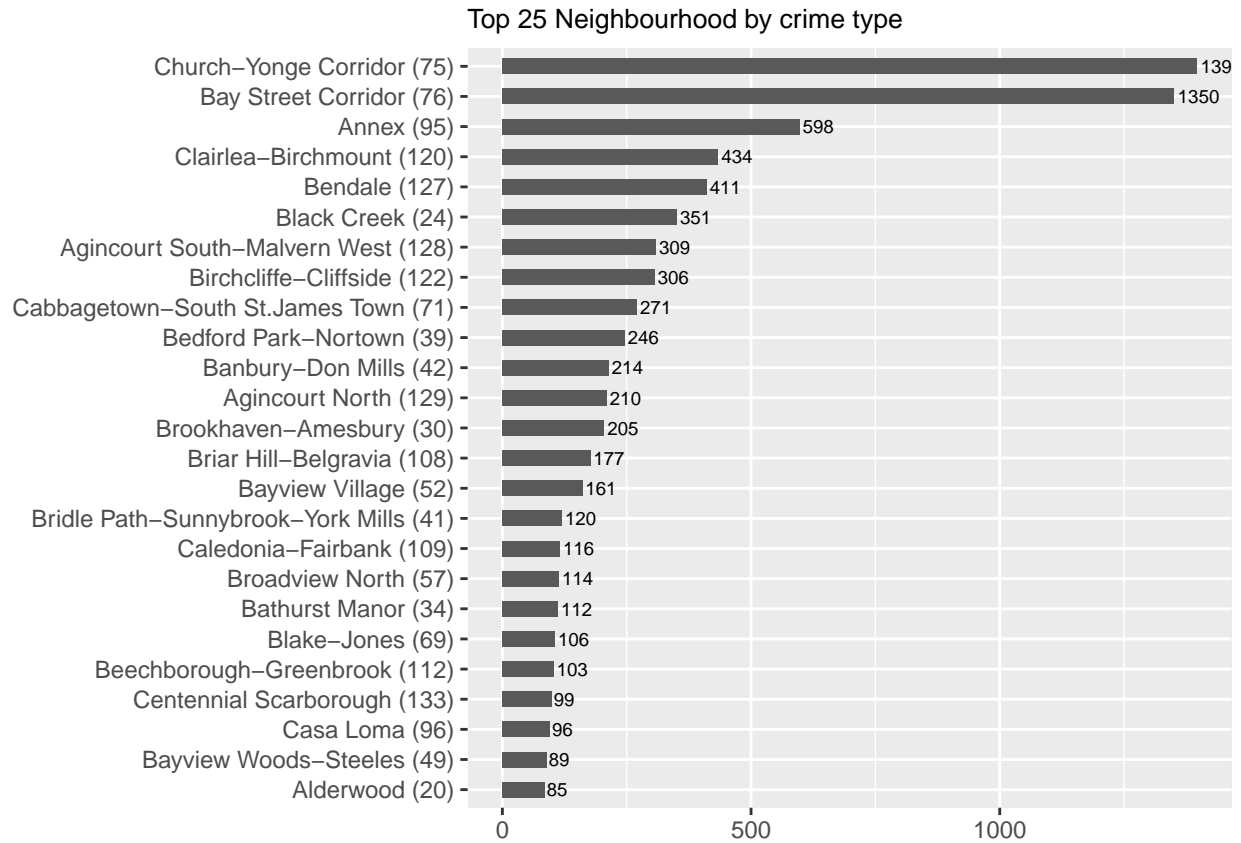
In criminal and Civil law: "Assault is an attempt to initiate harmful or offensive contact with a person or a threat to do so" and it was the most prevalent form of violent crime in Toronto in 2019.

Assault is a broad category and most of the crimes are classified as such. The plot below shows the isolation of the Crime type: Assault from the variable **MCI** and the subgroup variable **offence** that was related to the occurrence.



Where in the city these crimes are most likely to happen?

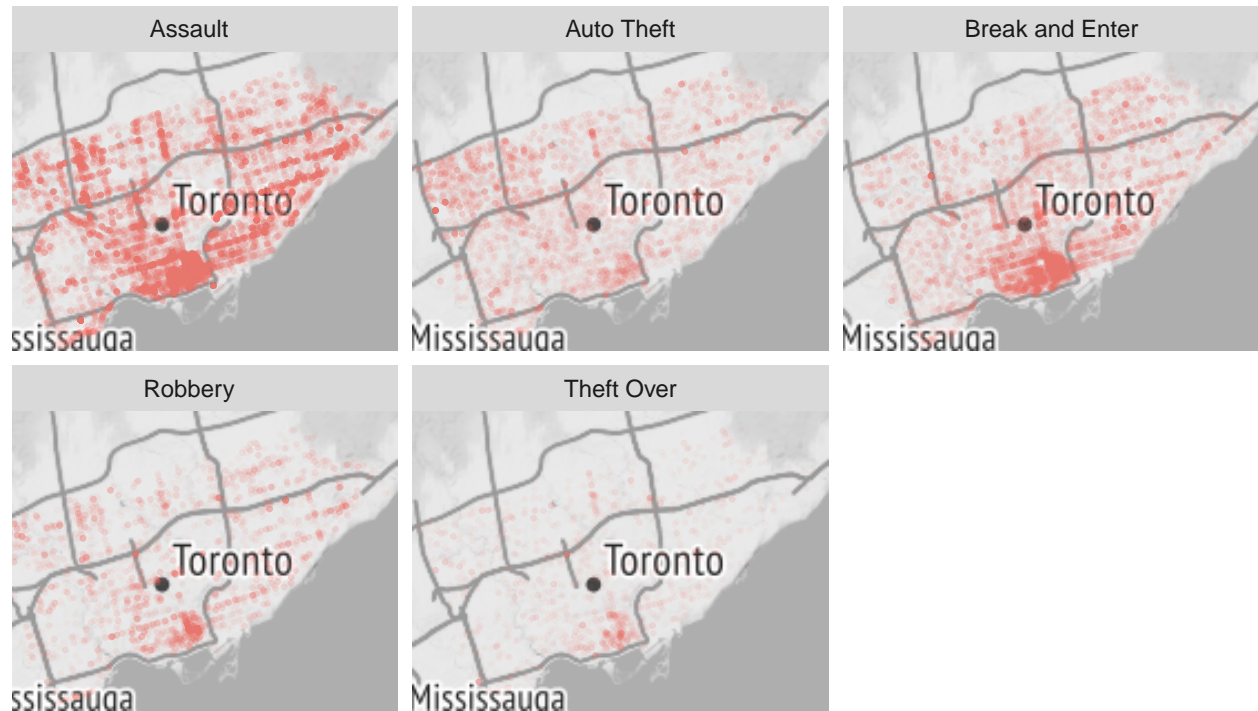
After isolating **MCI** and **Neighbourhood** from the dataset (below plot), we see that Church-Yonge and Bay-Street corridor are the top 2 most dangerous neighborhoods. The former by being very close to the Ryerson University and home of the Toronto's Gay Village, but both being the heart of downtown Toronto holding the highest number of high-rise condominiums, office towers, apartments, rich area with restaurants, stores and bars. Therefore, the Crime rate is pronounced very high indeed with different rates of crime.



Toronto's Crime Map

The map below gathers three variables: **MCI**, **Latitude** and **Longitude** to show where are the major occurrences of crimes in the city of Toronto. As before mentioned, there is a large concentration of assaults happening downtown at Church-Yonge and Bay Street Corridor. East and North directions are more peaceful than other areas. Hence, we can compare high- and low-density areas, and optimize the visualization with the following map.

Toronto's Crime Map 2019



It is no surprise that downtown holds the major concentration of crimes of all types. However, it is noticeable that Auto theft has a higher volume in the north-west region in comparison with other areas. Conversely to Break and Enter that has its highest volume of occurrences, after downtown, in the north-east region (Scarborough and such).

Conclusion

This analysis enlightened readers the correlation between Toronto neighborhood, crime rate, type and propensity to occur across the variable time. Clustering methods were applied to showcase the subcategory of offence in Assault crimes to provide the audience with a more detailed overview of the city. Several visual representations were used to illustrate what have been talked on this preliminary analysis. However, other steps seem necessary to cross-relate them with other aspects of the crime compound (seasonality, ethnicity of offenders, educational and employment levels, etc.) in an investigative and even preventive approach.

Results of the analysis points out that certain neighborhoods characteristics are related to higher or lower crime rates. There is a spatial correspondence between crimes and the area of occurrences, but this study does not provide a description of the complex mechanisms underlying this correlation. To address such questions, it would be useful to analyze the data over a larger period of time and the collection of economic and demographic features to better understand the mechanisms that are related to higher or lower crime rates across the studied area.

Undoubtedly, Toronto repeats what is seen around the globe: the majority of crimes happen in commercial areas which in Toronto's case has also the major residential density Furthermore, this study identified dangerous hot spots at certain times and places and, I hope, it can help with a better planning and safety measures to prevent crime rates to further escalate.

References

- Kahle, David, and Hadley Wickham. 2013. “Ggmap: Spatial Visualization with Ggplot2.” *The R Journal* 5 (1): 144–61. <https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>.
- Public Safety Data Portal, Toronto Police Service -. 2020. *Major Crime Indicators*. <https://data.torontopolice.on.ca/pages/major-crime-indicators>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto, Open Data. n.d. *Major Crime Indicators*. <https://open.toronto.ca/dataset/major-crime-indicators/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, 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, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.