

Paper1-Analyzation of Toronto Fire Incident Dataset

Yilin Lan

02/06/2022

Abstract

In this report, I introduce the Toronto public safety background and the major concerns about public safety hazard. The dataset I used collects fire incidents data in Toronto. Then, I introduce the observations, variables of the dataset as well as how the data is collected. The pros and cons are also mentioned in the report. I give some possible improvement on future data collection. Then, I analyze the relationship between estimated dollar loss and number of exposure fires, civilian casualties, building status. Finally, I give my data analyzation result based on the previse work.

Introduction

With the development of technology and society, citizens and government concern more about the public safety. Among different types of public safety hazard, fire accident one of the most common thing in people's daily life. In January 19, 2022, in Toronto, a fire broke out at a home on Gladstone Avenue, east of Dufferin Street at 7 p.m. A woman is dead and a man is in critical condition in this fire accident (Aguilar, n.d.). Fire accidents happened almost every day, people get injury from the accidents often. They even cause death. Statistically, there are about 24,000 house fires each year in Canada, resulting in an average of 377 deaths and 3,048 injuries per year. Fire causes huge casualties and property loss. Since then, due to the global warming and technology improvement, people use more electrical goods during their life so that there is an increasing trend in number of fire incidents globally every year.

Base on the previse background of public safety, government needs to collect the fire incidents information in order to improve the city's safe infrastructures as well as promote fire prevention. The dataset I used is provided by Ontario Fire Marshal (OFM) and share to the public on City of Toronto Open Data Portal. The OFM provides leadership and expertise on fire safety and it promote changes in laws, fire equipment in order to minimize the impact of fire and other public safety hazards on people, property and the environment in Ontario (Ontario.ca 2022). According to Fire Protection and Prevention Act (FPPA), the Fire Marshal is responsible for the investigation of the cause, origin and circumstances of any fire, explosion, etc (Ontario.ca 2022). Every year, the OFM investigates approximately 600 fire cases, including 80-100 death and some injury of the firefighters (Ontario.ca 2022). These investigations help experts to determine why fire accidents happened and how to prevent them in the future. They also help identify trends in fire emergencies and ensure fire protection services are maintained for every Ontarian (Ontario.ca 2022).

Given the importance of fire incidents statistics in characterizing the most common cause of fire incident and making improvement about fire protection infrastructure, it is important to understand how the fire incident dataset is reported and how it may be interpreted. For this report, I will use this dataset to analyze how the estimated dollar loss in a fire incident related to civilian casualties, number of exposure fires and different building status. Also, I summarize the major possible causes and the common origins of fire. Moreover, I will discuss the potential bias in how these data are collected and the impact on the final data analyzation. Some possible improvement when collect data in the future has also been given. The dataset will be processed and analyzed in R (R Core Team 2020) primarily using the tidyverse (Wickham et al. 2019) and dplyr (Wickham et al. 2021) packages. Figures and tables will be created with ggplot (Wickham 2016) and kableextra (Zhu 2020). The packages knitr (Xie 2021) are used to generate the R markdown report.¹

¹<https://github.com/YilinLan/STA304-Paper1>

Data Source

This dataset provides data about all the cases that sent to the Ontario Fire Marshal (OFM) relating to only fire incidents. The Dataset includes only Fire incidents as defined by the Ontario Fire Marshal and some incidents have been excluded pursuant to exemptions under Section 8 of Municipal Freedom of Information and Protection of Privacy Act (MFIPPA)(Ku 2022). The dataset was published and updated on City of Toronto Open Data Portal annually by Fire Services. The fire incidents analyzed in this report was obtained in csv format from the City of Toronto Open Data Portal using the R package opendatatoronto. It was last updated on Jan 31, 2022.

Methodology and Data Collection

The dataset contains all the fire incidents cases that have been reported to the OFM from year 2011 to 2022. Fire incidents are reported to the OFM through all Ontario fire departments. They are responsible for recording data on every emergency response they make (Ontario.ca 2022). Every fire incident is detailed reported, including the start and end time, location, estimated dollar loss, origin of fire, etc. Based on the background, it is always hard for experts to figure what happened in a fire incident. Crews usually need to be on scene for two to five days before conducting further analysis for weeks after (Westoll, n.d.). This situation has been reflected on the dataset. There are many missing values on number of exposure fires, Alarm Failure code, Alarm Type code and building status, etc. In fact, it takes a long time to investigate the origin of a fire incident. Manny Garcia, a training specialist with the OFM said, “Say the fire is on the interior of the building, then we commence our interior analysis so that we can then determine what’s our room of origin and where that fire started and then once we’ve determined the room of origin, we try to determine the area of origin within that room and see what ignition sources were there.” (Westoll, n.d.). Once the origin has been determined, people can start to investigate the building status and Alarm Type code. These are all hard to get an exact information because fire has already destroyed a lot of things.

The data is first collected by all Ontario fire departments and sent to the OFM. Then, the OFM is responsible for analyzes and transforms the collected data into valuable information (Ontario.ca 2022). After a fire incident, the specialists survey the scene in order to get the information about the fire incident. Data has been collected through the process. One of the advantages of the dataset is that it contains lots of information about each fire incidents, all possible information is detailed recorded in the dataset. When people need to review a fire incident, the dataset can provide enough information such as the origin of the fire, building status, start and end time, etc. It is convenient for the OFM and experts to summary the best way to prevent the fire incident in order to minimize the impact of public safety hazards.

However, due to some privacy concern, some data have been excluded in the dataset and that will lead to the incompleteness of the dataset. The data analyze result will not be completely accurate. There exist some people that don’t want to share the fire incident information to the public. This problem cannot be fixed because the law requires that local government institutions protect the privacy of an individual’s personal information existing in government records. Moreover, there are over 90% of observations contain missing value. Due to the limited technology, it is hard to get complete and detailed information in the fire incident. Although it is hard to collect all the information after the fire incidents, the missing value will increase the uncertainty of the data analyze result. This problem can be ameliorated by improving the survey methods. Overall, there will be some potential bias occurred during the data analyzation. Some fire incidents have not been recorded and there are some missing values in the dataset. People will get smaller than actual number of fire incidents and the mean of total dollar loss will be lower than average. Moreover, the result of the most common origin of fire will not be accurate.²

²<https://github.com/YilinLan/STA304-Paper1>

Data Characteristic

Data Overview

The fire incident data set contains aggregate data of all fire incidents reported to the OFM between year of 2011 and 2022 (Ku 2022). There are 17536 observations in the dataset and 43 attributes. Since the dataset includes comprehensive information about every fire incident, I will mainly focus on some important variables that closely related to fire incident and people's safety. The selected variable includes area of origin, building status, possible cause, civilian casualties, estimated Dollar loss and number of exposure fires. The variable types are either integer or character. As I mentioned in the previous part, the dataset contains lot of missing values and that will impact on the data analyzation process so that I remove all the missing data. The impact on the final result interpretation will be mentioned again below. A sample view of the dataset is displayed below.³

```
## # A tibble: 6 x 6
##   Estimated_Dollar_Lo~ Exposures Civilian_Casual~ Area_of_Origin Building_Status
##           <int>         <int>          <int> <chr>          <chr>
## 1           500000             2            0 50 - Basement~ 03 - Under Con~
## 2          1000000             2            0 11 - Lobby, E~ 02 - Under Ren~
## 3           100000             1            0 24 - Cooking ~ 01 - Normal (n~
## 4              7500             1            0 49 - Other St~ 01 - Normal (n~
## 5          1000000             3            0 79 - Other Ou~ 01 - Normal (n~
## 6          250000             1            0 42 - Garage    01 - Normal (n~
## # ... with 1 more variable: Possible_Cause <chr>
```

³<https://github.com/YilinLan/STA304-Paper1>

Count for Different Original On Fire Area and Building Status

According to the information provided by the OFM, it is necessary to summarize the number of different original on fire area. It is a useful information to help government promote fire safety requirement. When citizens see the result of the report, they will know which place in their home has higher possibility on fire. Later in their life, people will pay more attention when using electric in that place.

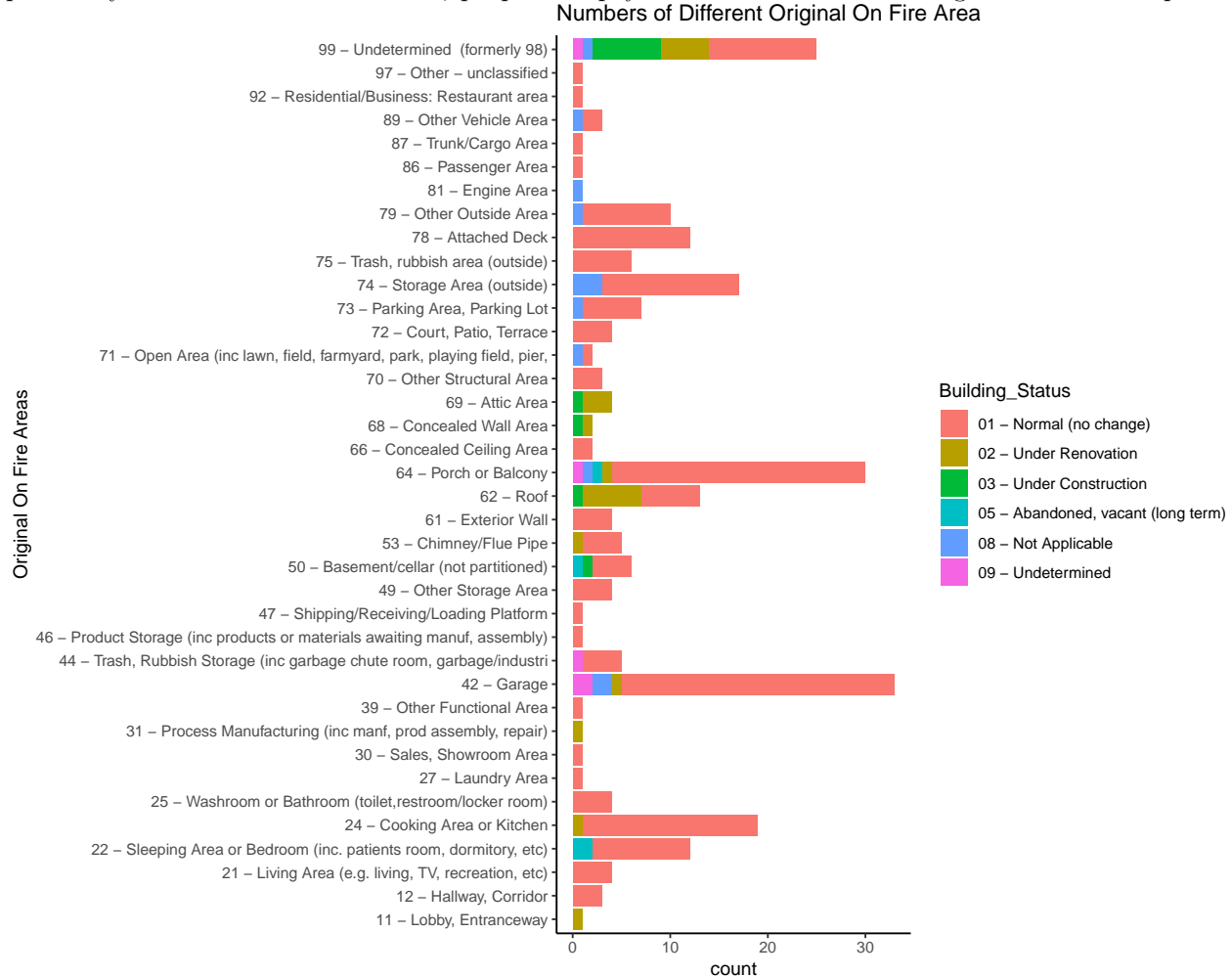


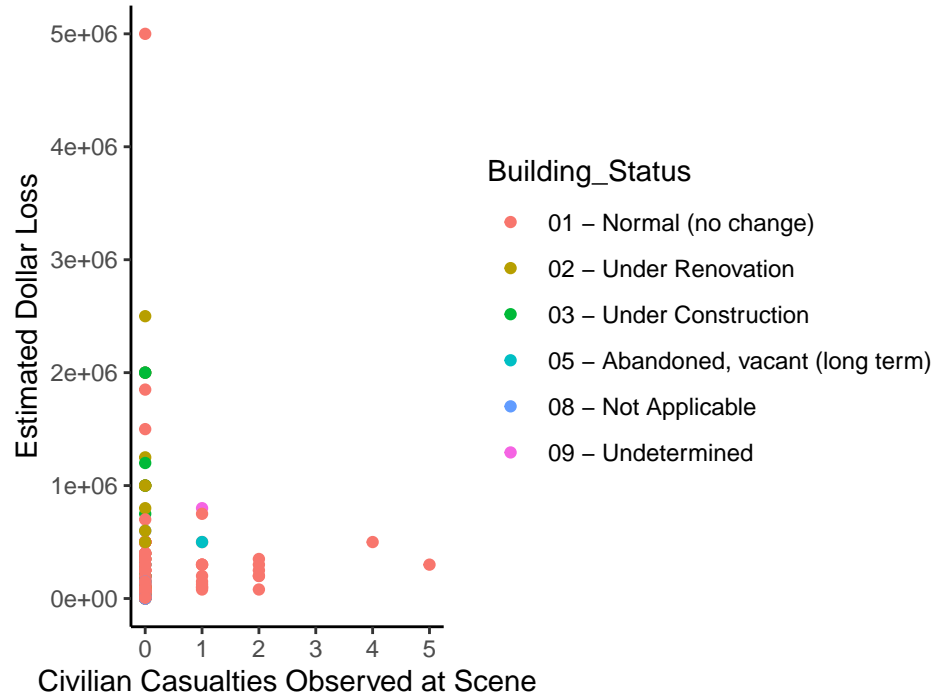
Figure 1 illustrates that garage, porch or balcony and kitchen are the most common origin of fire. The count are about 35, 30 and 28 respectively. Garage and balcony are usually filled with sundries and people seldom spend time in there while people use fire to cook in the kitchen. Due to the wrong display of items and lack of management, these places become the most common fire origin. As the article state, kitchen is the heart of home and it is the most common spot in the house for fires to start (Tholen 2021). Moreover, other than normal building status, the under construction and renovation buildings are also easy to occur a fire incident. To prevent the fire incident, people should always keep an eye on the stove, oven, and other appliances like toaster ovens, and move combustible items away from the stove. Government should improve the infrastructure and reduce the number of under renovation and under construction buildings.⁴

⁴<https://github.com/YilinLan/STA304-Paper1>

Relationship between Estimated Dollar Loss and Civilian Casualties

People care about the property loss and civilian casualties observed at scene in a fire incident because these two variable directly reflect how serious a fire incident is. Usually, more civilian casualties means more serious of a fire incident. To investigate whether the graph follow our common knowledge, I use Figure 2 illustrates the rela-

Under Renovation and Construction Buildings Have Hi



tionship between them.

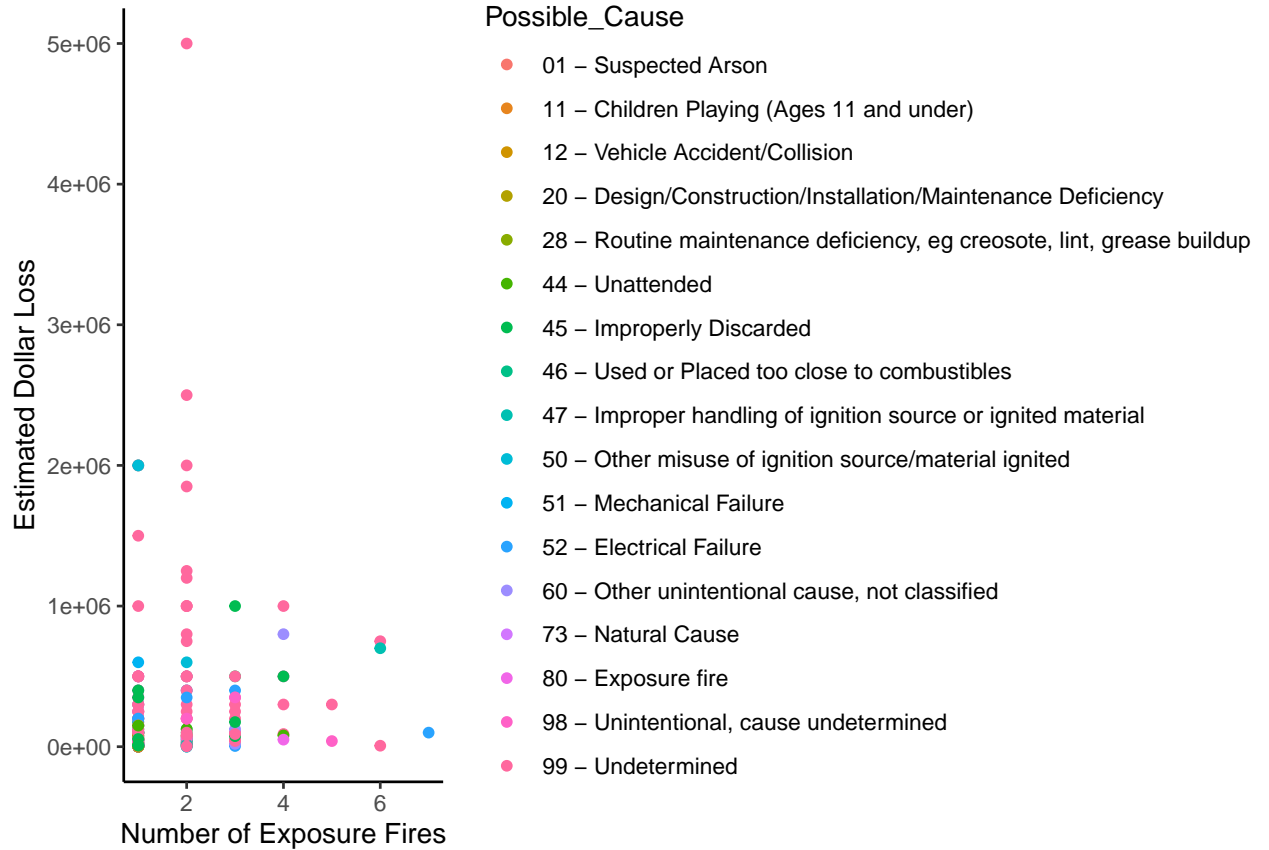
I choose variables Estimated_Dollar_Loss, Civilian_Casualties to draw a scatterplot and categorize the plot using variable building status. From the graph, the estimated property loss spreads around 5000-500000 dollars and there is no obvious relationship between civilian casualties. The estimated property loss tends to be lower when there are more civilian casualties. Moreover, there are three outliers in the plot. One has extremely high estimated dollar loss and the other two have high civilian casualties. The three outliers will generally rise the average of both two variables. The dark yellow and green dots are located at the upper side of the scatterplot. That means when the building status is under renovation or under construction, there will be more property loss. Besides, when the property loss is high, there likely to be no civilian casualties.⁵

⁵<https://github.com/YilinLan/STA304-Paper1>

Relationship between Estimated Dollar Loss and Number of Exposure Fires

The estimated dollar loss can also associated with the number of exposure fires. Government can use this result to see whether the firefighters use in an effective way to put out a fire. They can even do some training for the

The Common Possible Causes are Unattended, Improperly Discarded



firefighters.

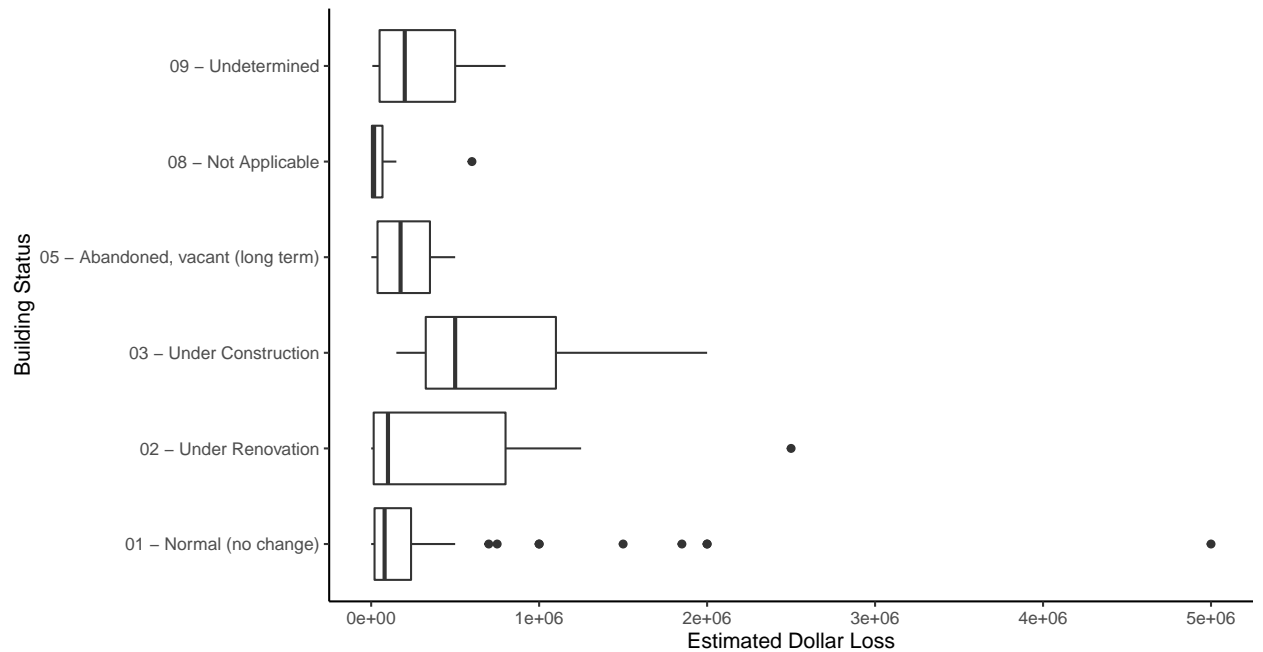
Then, I want to investigate the relationship between estimated dollar loss and number of exposure fires. I choose these two variables and categorize the plot using variable possible cause. In Figure 3, there is a slightly negative relationship between estimated dollar loss and number of exposure fires. When the number of exposure fires increases, the estimated dollar loss decreases. This result is different from the news and common knowledge. Usually, the estimated dollar loss is positively related to the number of exposures fires. This may due to the incompleteness of the data. Moreover, other than the undetermined cause, there are many green and blue dots in the scatterplot and that means the major possible causes are unattended, improperly discarded and mechanical failure.⁶

⁶<https://github.com/YilinLan/STA304-Paper1>

Estimated Dollar Loss in Different Building Status

The boxplot help people to see the number of property loss in different building status in a clearly way. Government is encouraged to build more normal buildings and reduce the under renovation or under construction

Under Construction Building Has the Largest Medium of Dollar Loss



buildings.

Building Status	min	median	mean	max
01 - Normal (no change)	20	80000	203841.51	5000000
02 - Under Renovation	250	100000	446916.67	2500000
03 - Under Construction	150000	500000	804545.45	2000000
05 - Abandoned, vacant (long term)	1000	175000	212750.00	500000
08 - Not Applicable	100	17500	83716.67	600000
09 - Undetermined	7000	200000	311400.00	800000

Last, in figure 4, I illustrate the medium of estimated property loss under different building status. It can be seen that if the fire happened in an under-construction building, the medium property loss will be high at 500000 dollars. I also create a summary table to summarise the estimated dollar loss in different building status.⁷

⁷<https://github.com/YilinLan/STA304-Paper1>

Reference

- Aguilar, Bryann. n.d. “One Person Dead, Another Injured After Toronto House Fire.” <https://doi.org/https://toronto.ctvnews.ca/one-person-dead-another-injured-after-toronto-house-fire-1.5747111>.
- Ku, Kevin. 2022. “Open Data Dataset.” <https://doi.org/https://open.toronto.ca/dataset/fire-incidents/>.
- Ontario.ca. 2022. “The Office of the Fire Marshal.” <https://doi.org/https://www.ontario.ca/page/office-fire-marshal#section-0>.
- Tholen, Celeste. 2021. “The Most Common Places That Fires Occur in the Home.” <https://doi.org/https://www.safewise.com/blog/the-most-common-places-that-fires-occur-in-the-home/>.
- Westoll, Nick. n.d. “An Inside Look at How Ontario’s Office of the Fire Marshal Investigators Probe Blazes, Explosions.” <https://doi.org/https://globalnews.ca/news/7999856/ofm-ontario-office-of-the-fire-marshal-fires-explosions-investigations/>.
- 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*.