Is the City of Toronto actually doing good with climate change?*

Yunkyung Park

February 6, 2022

Abstract

As climate change has been one of the most important global issues since the late 1990s, many countries have been determined to take action in achieving net-zero emissions by 2050. Meanwhile, the City of Toronto has reported being recognized as a global leader in environmental action and transparency. This report utilizes data on renewable energy installations provided by the City of Toronto to the public to verify the truth of the report from an individual perspective.

1 Introduction

It has been seven years since the Paris agreement was signed. Many countries have announced their goal to reach net-zero emissions by 2050 where net-zero emissions mean the equilibrium state where the amount of carbon emitted into the atmosphere is equal to the amount of carbon removed from it (United Nations 2015). In December 2019, the European Green Deal was announced, which aims to achieve at least 55% of net greenhouse gas emissions reduction from 1990 levels by 2030 and to reach a carbon-neutral by 2050 (Fleming and Mauger 2021). In 2021, President Biden announced a target for the United States to reduce net carbon emissions by 50% from 2005 levels by 2030 (House, n.d.). While other countries are active in reducing greenhouse gas emissions, the Government of Canada has announced a goal to reduce emissions by 40-45% by 2030 compared to 2005 levels and established the Canadian Net-Zero Emissions Accountability Act which commits to reaching net-zero emissions by 2050 (Canada Service, n.d.).

How the government acts on it is crucial to reaching the goal. The government needs to put more effort into addressing the issue, intervene to fight climate change, and launch new programs to reduce greenhouse gas emissions. However, a net-zero state cannot be reached solely by reducing carbon emissions. In fact, what matters more is to produce renewable energy and to find alternative fuel sources. Finding the alternatives is something that the researchers should work on, so this report focuses on producing renewable energy. More precisely, as the City of Toronto has reported being recognized as a global leader in environmental action and transparency (Cassells, n.d.), this report addresses how active the City of Toronto has been in producing renewable energy by observing the number of renewable energy systems installed in Toronto. Additionally, it checks whether the public would be able to recognize how the City of Toronto is doing with climate change by observing the quality of the open dataset.

Before beginning the report, it is worth mentioning that installing these devices produces carbon, but once it is installed, it lasts for at least twenty years, and it takes about three years to pay off the amount of carbon emitted in installation (Saini and Gidwani 2022). In addition, installing these systems will make you earn money without doing anything wrong, and part of the installation fee is sponsored by the government, which is the program called FIT and MicroFIT, and it will be further discussed in this report. They are the programs introduced in 2009 by Ontario to encourage installing renewable energy projects (Independent Electricity System Operator, n.d.a). However, the Independent Electricity System Operator (IESO) no longer accepts applications under the MicroFIT Program as of December 31, 2016 (Independent Electricity System Operator, n.d.b).

*Code and data are available at: https://github.com/parkyunk/OpenDataToronto.

2 Data

2.1 Data Source

This report utilizes data on the renewable energy installations on city-owned buildings in the City of Toronto (Toronto Open Data 2020) obtained from Toronto Open Data Portal using the R package 'opendatatoronto' (Gelfand 2020). The dataset was obtained in csv and geojson format. It includes the locations of the installed renewable energy systems as well as the year of the installation and the size of the system. Note that it only contains City divisions and not all agencies or corporations, and it was last updated on April 20, 2020. The dataset is processed and analyzed using 'R' (R Core Team 2021) mainly with the 'tidyverse' (Wickham et al. 2019) and 'dplyr' (Wickham et al. 2021) packages. The package 'janitor' (Firke 2021) is used to clean data, and the graphs and the tables are created in 'ggplot2' (Wickham 2016) and 'kableExtra' (Zhu 2021), respectively. The packages 'bookdown' (Xie 2016) and 'knitr' (Xie 2014) are used in generating the R Markdown report.

2.2 Data Process

```
## Simple feature collection with 6 features and 40 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: -79.50348 ymin: 43.60508 xmax: -79.3279 ymax: 43.69098
## Geodetic CRS:  WGS 84
## # A tibble: 6 x 41
##      id address_point_id address_point_id_link address_number linear_name_full
##   <int>            <int>                 <int> <chr>          <chr>
## 1     1         20006001                    NA 9              Hanna Ave
## 2     2          4257357                    NA 850            Coxwell Ave
## 3     3          7985900                    NA 462            Runnymede Rd
## 4     4         14069050                    NA 339            Queens Quay W
## 5     5          7348693                    NA 31             Drummond St
## 6     6           987966                    NA 70             Birmingham St
## # ... with 36 more variables: address_full <chr>, postal_code <chr>,
## #   municipality <chr>, city <chr>, ward <chr>, place_name <chr>,
## #   general_use_code <int>, general_use <chr>, centreline_id <int>,
## #   centreline_side <chr>, centreline_measure <dbl>, address_class_desc <chr>,
## #   lo_num <int>, lo_num_suf <lgl>, hi_num <lgl>, hi_num_suf <lgl>,
## #   linear_name_id <int>, x <lgl>, y <lgl>, longitude <lgl>, latitude <lgl>,
## #   id_2 <int>, building_name <chr>, client_address <chr>, ...
```

The raw data contains 100 observations with 41 attributes. However, as observed from the sample view of the dataset above, it contains so many redundant variables, and there are variables with no values. Therefore, only the 9 variables will be used in this report. Based on the needs, some of them will be mutated, and there might be the ones that will be dropped later. However, these are the variables that have remained so far: address_full, municipality, general_use_code, centreline_measure, building_name, type_install, year_install, and size_install.

2.2.1 Ownership of the locations

Since this report focuses on the government action on renewable energy, I will only consider whether it is owned by the government or not. By using the general_use_code, I could determine whether the place is publicly-owned or privately-owned. The list of codes for the privately-owned locations was formed by Justin Rai using the mapping software ArcGIS (Rai 2019). Therefore, I created a new feature named owned and the feature general_use_code has been removed as it will no longer be used.

2.2.2 FIT/MicroFIT program

The MicroFIT program is intended to support the development of "micro" renewable electricity generation projects of 10 kilowatts (kW) or less, and the FIT program is for projects with greater than 10kW (Independent Electricity System Operator, n.d.a). In the dataset, the programs have been more specified such as MicroFIT A, B, and C and FIT 1, 3, and 4. However, it did not provide any details on the differences among these programs, and I could not find any information about it anywhere.

Instead, there is other data on renewable energy installation (Toronto Open Data 2014) provided in xlsx format from the City of Toronto Open Data Portal using the R package opendatatoronto (Gelfand 2020). It was last updated on October 24, 2014, and it will be used only in this part of the report. This dataset contains a variable named r_type which indicates the exact type of the installed renewable energy systems such as solar photovoltaic, solar pool heating, and geothermal instead of naming its program type such as MicroFIT A, B, and so on as in the dataset that we have been working on where these types are mentioned under the feature type_install. Due to their analogous feature names and values, I assumed that they might be related to each other. Thus, a statistical test will be conducted to determine if they have any relationship.

One thing to note is that this new dataset contains only 59 observations, and since it was last updated in 2014, it does not contain most data that are present in the one we are working with, which is why it will be used only in this section of the report. Only the samples that are present in both the new dataset and the main dataset will be used to conduct the testing. The others will be discarded in the merged dataset.

Now, here is the general idea of how the test will be conducted. There will be two tests: one for MicroFIT program and the other for FIT program; the tests compare two categorical variables, so the Chi-square test would be common to use in this case. However, due to the small sample size, Fisher's exact test will be used instead. This is because the Chi-square test is sensitive to the sample size (Kim 2017). Also, note that there are two data with the type 'Storage + NM', but both were installed in 2018 and 2019, which were after the newly-imported dataset was last updated, so we will only take FIT/MicroFIT programs into consideration.

Table 1: Contingency table for specified types of the programs

	Geothermal	Solar Hot Water	Solar Photovoltaic (PV)
FIT 1	1	1	12
FIT 4	2	3	3
MicroFIT	0	0	2
MicroFIT A	1	1	6
MicroFIT B	2	0	0
MicroFIT C	0	1	0

Based on Table 1, I have first created two tables where one is for the MicroFIT program and the other is for the FIT program, so one is of the size 3x2 and the other is of the size 3x3. Now, we are ready to run a Fisher's exact test. The null hypothesis for both two cases is the following: the two categorical variables (type_install and r_type) are independent.

Table 2: Contingency table for the FIT programs

	FIT.1	FIT.4
Geothermal	1	2
Solar Hot Water	1	3
Solar Photovoltaic (PV)	12	3

Table 3: Contingency table for the MicroFIT programs

	MicroFIT.A	MicroFIT.B	MicroFIT.C
Geothermal	1	2	0
Solar Hot Water	1	0	1
Solar Photovoltaic (PV)	6	0	0

Table 4: Fisher's Exact Test for Count Data

	p.values
FIT	0.06296
MicroFIT	0.04848

The p-value obtained from Fisher's exact test for the FIT program is $0.06 > 0.05$, so we failed to reject the null hypothesis. The p-value is the answer to the question - the null hypothesis is true (Sheather 2009). Smaller p-values indicate that the result is statistically significant, and the data supports the alternative hypothesis whereas larger p-values indicate that the data is consistent with the null hypothesis (Sheather 2009). In particular, if it is greater than 0.05, then it implies that the evidence against the null hypothesis is weak, and therefore we failed to reject the null hypothesis (Sheather 2009). On the other hand, the p-value obtained from Fisher's exact test for the MicroFIT program is $0.048 < 0.05$, and if the p-value is in between 0.01 and 0.05, we say that there is moderate certainty of the evidence against the null hypothesis (Sheather 2009). However, since we failed to show that they are related with the FIT program types, and the p-value does not imply that there is a strong relationship between the two, I will not consider the specified type of FIT and MicroFIT and only consider FIT and MicroFIT program in this report.

2.2.3 Centreline

Next, there exists a feature centreline_measure in the dataset where it is unclear what it indicates. What it means is not specified. According to the Oxford Dictionary, the centreline is a real or imaginary line that passes through the centre of something ("Centreline," n.d.). When it comes to measuring a centreline of the building, it usually means the item's mean length, and it is one of the most important techniques in constructing the building (Cunningham 2015). However, it is unclear how the centrelines are measured in this dataset. Also, there are negative values, so it becomes more ambiguous what is meant by the feature centreline. Therefore, I will first conduct statistical testing to make the terminology clear. If it meant the one defined by the dictionary, then the larger the centreline is, the greater the size of the installation is. The Pearson correlation coefficient will be used to determine if there is any correlation between the measure of the centreline and the size of the installation. The reason no other explanatory variables are used here is that it is evident that we can install the larger size of the product when the area is large, and the centreline is the only variable in the given dataset that seems to be related to the area of the place.

```
## `geom_smooth()` using formula 'y ~ x'
```

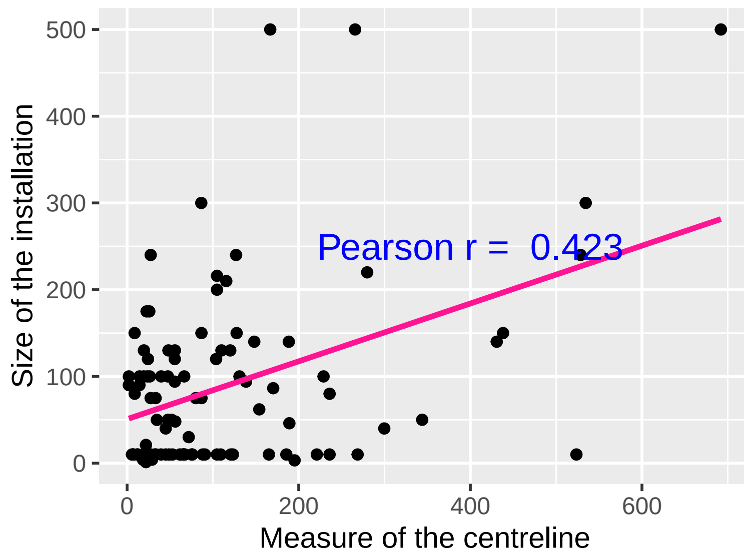

Figure 1: Relationship between the measure of the centreline and size installed

First, note that I have used the absolute value of centreline_measure when plotting a graph. This is because I am assuming that the magnitude of the centreline is related to the size of the installation. Second, Figure 1 tells that the Pearson correlation coefficient is 0.423. The Pearson correlation coefficient r lies between -1 and 1, and if the value r is far from 0, it tells us that there exists a stronger linear relationship between the two variables (Sheather 2009). Also, if the value is positive, then it means when one increases, then the other increases as well whereas if it is negative, then they change in the opposite direction. It is generally considered to have a strong relationship when $|r|>0.7$ (Sheather 2009), so the coefficient that I have here is not high. Therefore, in this report, I will assume that the centreline and the installed size have no relationship, and I will no longer take centreline_measure into the consideration since it is hard to understand what the centreline is whereas the installation size is clear.

To substantiate that the feature size_install is clear, I have plotted a jitter plot Figure 2. The jitter plot is better in visualizing when there are many overlapping data points (Sheather 2009). It can be observed from the figure that what the feature size_install indicates is perspicuous. Therefore, we will only use the size of the installation in this report.

2.2.4 Year

Lastly, only the year will be taken into the consideration, the values in feature year_install are mutated to only contain the year, not the exact date included.

2.3 Analysis

Now, we are ready to analyze the data. To briefly mention what features will be focused on, I will use whether the location is public-owned or private-owned, what type of program has been used more, and the trend of installations over the years.

2.3.1 Percentage of the ownership of the location

Note that there are 100 observations, so when it says 81% in Figure 3, it means that there are 81 observations of that feature. From the figure, it can be said that the number of installations in the public-owned property significantly exceeds the number of installations in the private-owned property. However, we should bear in mind that this database is incomplete. It includes City divisions but not all Agencies or Corporations.

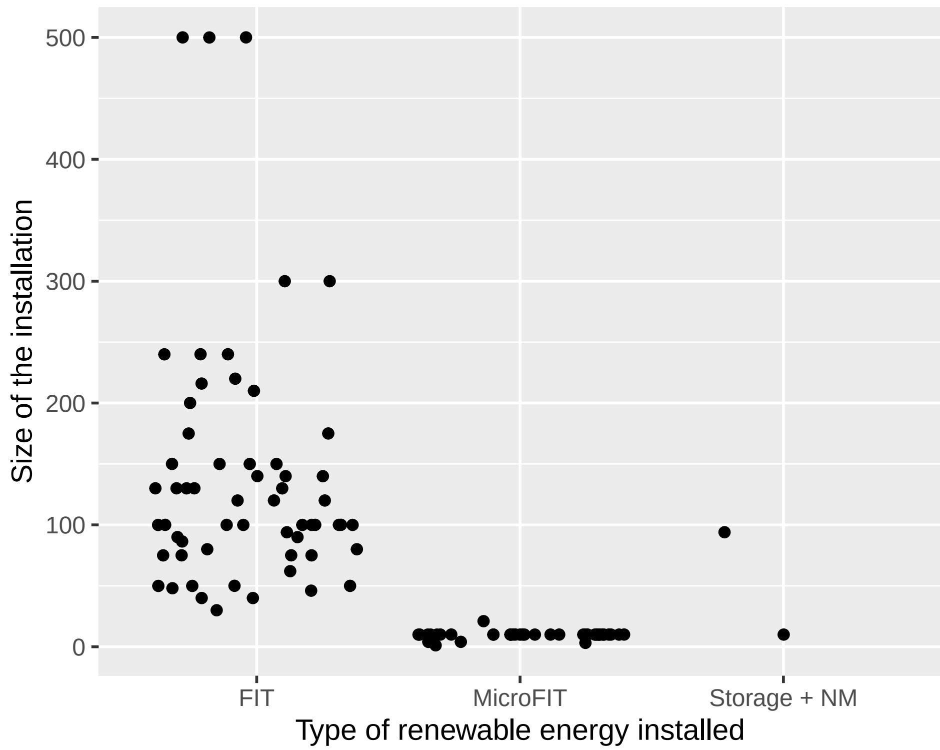

Figure 2: Relationship between the program type and the size installed

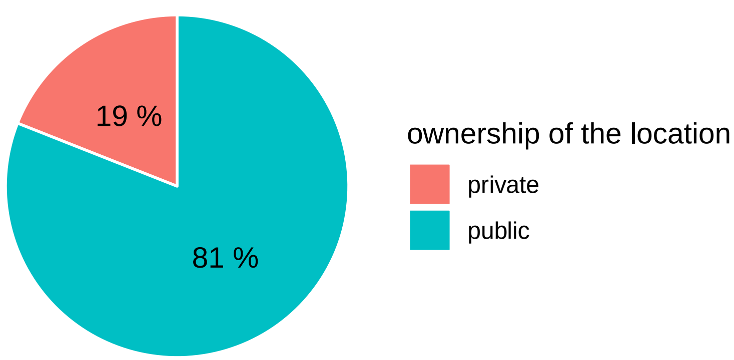

Figure 3: The ownership of the locations that have renewable energy installed

2.3.2 Type of renewable energy installed

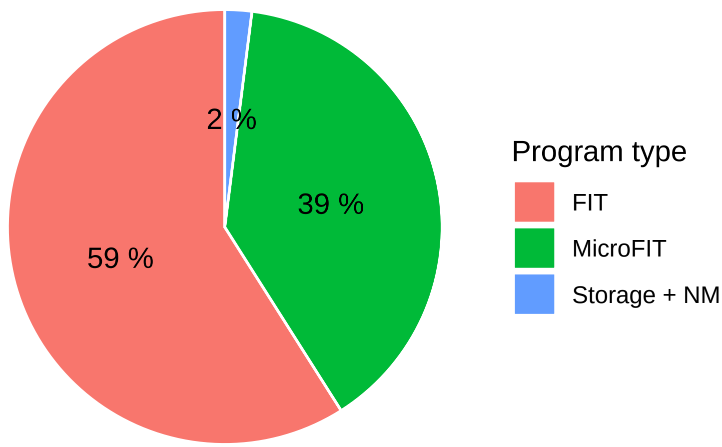

Figure 4: Type of renewable energy installed

Figure 4 shows the number of contracts that were made in each program. Remember that the FIT program is for the locations where it can generate over 10kW, and the MicroFIT program is for the locations where it generates 10W or less (Independent Electricity System Operator, n.d.a). On that account, we can say that more renewable energy was produced. For type 'Storage + NM', it consists 2% of the total, but there is no additional information provided. To see the trend of installation, we will sketch a bar plot where it shows the number of installations over the years.

2.3.3 Number of installations over the years

First, it can be observed that the number of installations has increased over the years, and besides FIT and MicroFIT program, which started in 2006, we can observe that the new program named Storage + NM has begun as well. However, there is not enough information provided for that program, so we will not take it into consideration. Also, the figure shows that there is no new MicroFIT contract made in 2017. As mentioned in the introduction, this is because IESO is no longer accepting applications as they have achieved the target in 2017 (Independent Electricity System Operator, n.d.b). This is good news to hear but bad at the same point. Climate change is an issue that cannot be solved within a decade. It is a long-term goal, which is why the Paris agreement has set 2050 to achieve a net-zero emission. The Paris agreement was signed on December 12, 2015, and they have set the several decades to reach it (United Nations 2015). The main purpose of this agreement is to slow down the rate of global warming so that eventually we will be able to stop by 2050 once it reaches an equilibrium state. Therefore, it would have been better if they updated their goal and continued the program after reaching the initial goal.

2.3.4 Number of installations among the municipalities

As mentioned in the introduction, these programs were introduced by Ontario, so the community involvement depends a lot on the municipal action. Therefore, we will compare the number of installations made per

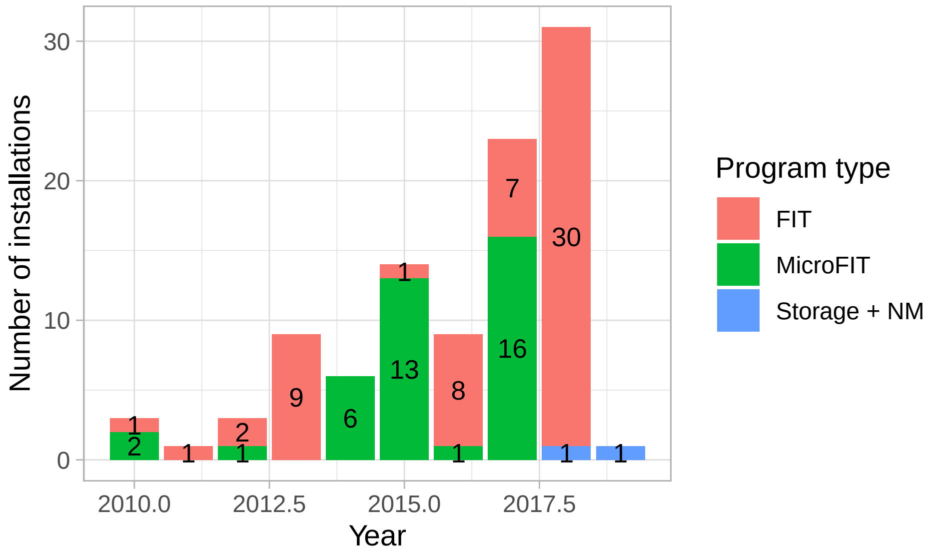

Figure 5: Number of renewable energy installations over years

municipality. Note that the municipal council changes every four years, and the election was held in 2010, 2014, 2018 (Association of Municipalities Ontario, n.d.), so I have grouped 2010-2014 and 2015-2019 to see if there is a significant change between those two periods.

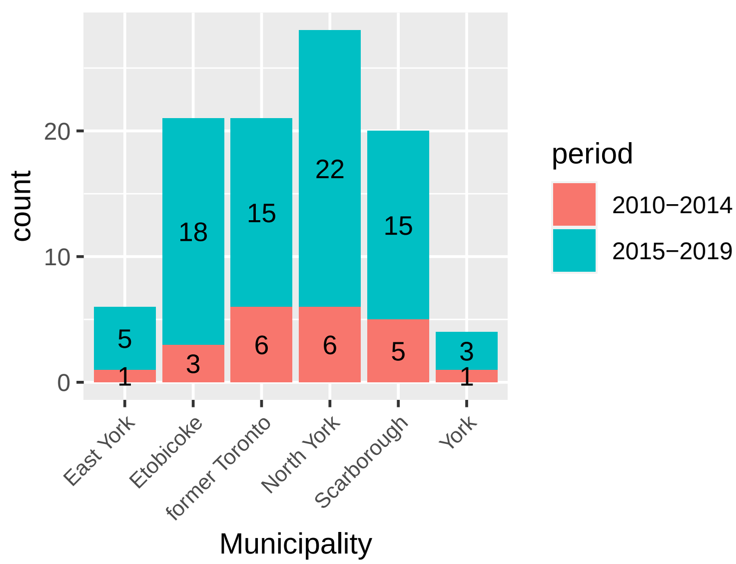

Figure 6: Number of renewable energy installations among the municipalities

First, as it can be observed from Figure 5, Figure 6 also shows that the interest has increasingly increased as time went by in all municipalities. Next, when we compare the number of installations among the municipalities, there was not much difference in the period 2010-2014 since it was just when the program launched. However, in the period 2015-2019, the number of installations made in North York greatly exceeded other municipalities. There can be further research on the election pledges on climate change for the ward councilors, and how they executed. However, that would be beyond the scope of this report.

3 Conclusion

Although the City of Toronto has been recognized as a global leader in environmental action and transparency (Cassells, n.d.), it has been constantly mentioned throughout this report that the dataset is incomplete. On the IESO website, it says that over 300,000 MicroFIT contracts have been made (Independent Electricity System Operator, n.d.b), but most of them are not reflected in this open data. On top of it, we could see that this dataset is very poorly given. There are many values that are missing and there exists a feature that is obscure (e.g. centreline_measure). In addition, it was hard to notice what is meant by microFIT A, B, etc. Despite its poor quality, it is still worth stating that the awareness of renewable energy has increased over the years. To raise more awareness of the issue, these data should be more open to the public and updated frequently so that people can keep a key eye on the current status of how the City of Toronto is doing against climate change.

References

Association of Municipalities Ontario. n.d. *Municipal Elections*. Accessed Feb. 6, 2022 [Online].

Canada Service. n.d. “Net-Zero Emissions by 2050.” https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/net-zero-emissions-2050.html.

Cassells, Valerie. n.d. *Toronto Scores an “a” for Climate Leadership. City of Toronto*. https://www.toronto.ca/news/toronto-scores-an-a-for-climate-leadership/#:~:text=The%20City%20of%20Toronto%20is,for%20the%20second%20consecutive%20year.&text=In%202019%2C%20City%20Council%20voted,zero%20by%202050%20or%20sooner.

“Centreline.” n.d. https://www.collinsdictionary.com/dictionary/english/centerline.

Cunningham, Tony. 2015. “Measuring Building Perimeters and Centrelines - Worked Examples.” https://doi.org/10.21427/b3mf-pw56.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Fleming, Ruven C., and Romain Mauger. 2021. “Green and Just? An Update on the ’European Green Deal’.” *Journal for European Environmental & Planning Law* 18 (1): 164–80. https://doi.org/10.1163/18760104-18010010.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.

House, The White. n.d. *President biden sets 2030 greenhouse gas pollution reduction target aimed at creating good-paying union jobs and securing u. S. Leadership on clean energy technologies.*, 2021. https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/22/fact-sheet-president-biden-sets-2030-greenhouse-gas-pollution-reduction-target-aimed-at-creating-good-paying-union-jobs-and-securing-u-s-leadership-on-clean-energy-technologies/.

Independent Electricity System Operator. n.d.a. “Feed-in Tariff Program.” Accessed Feb. 6, 2022 [Online]. https://www.ieso.ca/en/Sector-Participants/Feed-in-Tariff-Program/Overview.

———. n.d.b. “MicroFIT Program.” Accessed Feb. 6, 2022 [Online]. https://www.ieso.ca/en/Get-Involved/microfit/Program-Documents.

Kim, Hae-Young. 2017. “Statistical Notes for Clinical Researchers:Chi-Squared Test and Fisher’s Exact Test.” *Restorative Dentistry & Endodontics* 42 (2): 152–55. https://doi.org/10.5395/rde.2017.42.2.152.

Rai, Justin. 2019. *Privately Owned Property Within Ch658 Protected Areas*.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Saini, Pawan, and Lata Gidwani. 2022. “An Investigation for Battery Energy Storage System Installation with Renewable Energy Resources in Distribution System by Considering Residential, Commercial and Industrial Load Models.” *Journal of Energy Storage* 45 (103493). https://doi.org/10.1016/j.est.2021.103493.

Sheather, Simon J. 2009. *A Modern Approach to Regression with R*. New York: Springer.

Toronto Open Data. 2014. “Renewable Energy Installations.” https://open.toronto.ca/dataset/renewable-energy-installations/.

———. 2020. “Renewable Energy Installations.” https://open.toronto.ca/dataset/renewable-energy-installations/.

United Nations. 2015. *Paris Agreement*.

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 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, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

———. 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/bookdown.

Zhu, Hao. 2021. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.