“Mapping the Island”: Data Journalism Education through Coverage of Gentrification and Housing Issues in Montreal

Gwendolyn Roley

A Research-Creation Thesis
In the Department of
Journalism

Presented in Partial Fulfillment of the Requirements
For the Degree of
Master of Arts (Digital Innovation in Journalism)

at Concordia University
Montréal, Québec, Canada

August 2021

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By: Gwendolyn Roley

Entitled: "Mapping the Island": Data Journalism Education through Coverage of Gentrification and Housing

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Digital Innovation in Journalism)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. David Secko

_____ Thesis Supervisor
Dr. Elyse Amend

_____ Examiner
Dr. Andrea Hunter

Approved by:

_____
Dr. David Secko, Graduate Program Director

_____ , 2021

_____
Dr. Pascale Sciotte, Dean of Faculty of Arts and Sciences

Abstract

"Mapping the Island": Data Journalism Education Through Coverage of Gentrification and Housing Issues in Montreal

Gwendolyn Roley

The focus of this research-creation project is how data journalism can be learned and applied to the coverage of housing issues in the city of Montreal. In examining this subject, several questions are raised about the nature of data journalism as a subsection of the larger journalistic field. This research aims to investigate how data journalism practices can be learned, taught and incorporated into the existing work of journalism studies, through coverage of the city, housing and gentrification.

In applying a "learning-by-doing" model of research-creation, conducting this project did lead to higher competency in data journalism practices. The results of this experience indicate ways data journalism education could be applied differently in university and mid-career journalism training curricula, namely, using a specialized data journalism project to develop a specific set of data journalism skills and making those skills more accessible by focusing on a localized story. It was also determined that in creating a data journalism project that focused on a specific story—in this case, housing issues in Montreal—there was a need to narrow in on a particular set of skills—namely data analysis, programming and mapping proficiency—similar to how traditional reporters must develop certain skills to cover different beats in a newsroom. This research indicates that a more widespread application of this learning-by-doing model in data journalism training and education has the potential to allow students to more deeply develop specific skill sets that would lend themselves more effectively to the expansive and collaborative practice of data journalism in the industry.

Acknowledgements

I would like to thank my thesis supervisor, Dr. Elyse Amend, as well as my committee member, Dr. Andrea Hunter, for guiding me through this process with wisdom, understanding and humour. I'd also like to thank my family for their unwavering belief and support, and my friends, especially those based in Montreal, who provided me with moments of reprieve from my work but still let me lecture them about what my thesis was so it would seem more real before I had written it all down. Much of the coding I learned to create this project would not have been possible without the guidance of Clarissa Verish. A particular acknowledgment must also go out to my parents, Scott and Diana Roley, archivists and data journalists of the historic past.

Table of Contents

Introduction 1

Research Questions 6

Review of Existing Research 7

Methodology 12

Reflections on Website Creation Process 14

Data: Availability *16*

Data: Analysis *20*

Visualization: Generation *26*

Visualization: Interactivity *31*

Story: Traditional Reporting *36*

Story: Sources *40*

Further Exploration 44

Concluding Statements 45

References 49

Appendices 56

Introduction

There is a cliché one hears over and over again, in university journalism departments and in newsrooms, that journalists can't do math—or they went into the field so they wouldn't have to deal with numbers. While this assumption is difficult to qualify, it is also becoming an increasingly dangerous assertion to make, because journalists cannot afford to toss the values of data and numbers aside, as they can only augment their reporting in today's complex world. The past two years have been especially poignant in underlining the importance of data-driven reporting with the coverage of the COVID-19 pandemic. A number of journalists, such as the CBC NewsLab team, have been using data-driven reporting methods to show the spread of the outbreak within communities across the globe. Data journalism was also critical in news reports about the 2020 American presidential election, where the tracking of accurate numbers proved to be crucial, as the results came down to a thin margin of victory in swing states like Georgia and Pennsylvania (see, for example Nate Cohn and Charlie Smart's work at "The Upshot" for *The New York Times*). With an election in Canada having been called for September 2021, data and polling will be used as they have been in the past to make predictions and to point out inequalities in the election system, as did the reporting leading up to the 2019 election which highlighted the lack of gender parity in Canadian political campaigns (Ouellet & Shiab, 2019). Not only can the study of databases and sets of numbers break important stories, but they lend an accuracy to reporting that is sometimes unattainable when relying solely on human sources.

As Philip Meyer (1991) puts it, "The journalist who adopts the tools of the scientific method to his or her own trade can be in a position to make useful evaluations with the more positive objectivity of science" (p. 13). For example, if one were to go back to two of the biggest stories of 2020—COVID-19 and the America presidential election—journalists could not rely

solely on the epidemiologists' word in an interview on where the virus was spreading, or the public relations teams of American politicians and governments telling them how they think the election was going to shake out. To accurately report on these stories, journalists need data, math and the application of social science methods. In both cases, the use of maps as visual aids to clarify the data involved was also helpful.

Data journalism itself is currently a hot topic in journalism studies, as demonstrated by the growth in academic literature on the topic over the past decade or so (see Review of Existing Research). Suffice it to say, many authors have taken the time to study what it is that defines "data journalism" by observing working data journalists in their newsrooms and discussing what their work means for the profession (Usher, 2016; Bounegru, Chambers, & Gray (eds), 2012). However, despite this academic interest in the field, there has been limited academic research about, or support for, journalism students who go to universities and journalism schools to train to become data journalists. Several think-tank-like institutions specialized in data journalism, such as the KnightLab at Northwestern University, have been established in journalism departments at universities in the United States. There are almost 150 data journalism courses offered at American postsecondary institutions, according to Bahareh Heravi of University College, Dublin. There are only eight analogous courses in Canadian universities (Heravi, 2019). As Heravi (2018) also points out, "data journalism education is often referred to as teaching 'data' to 'journalists,'" (p. 353) meaning there is a lack of a holistic approach that teaches journalists how to perform data journalism by actually doing it. It is therefore worthwhile, especially in a Canadian context, to explore what it means to have a data journalism education, how one learns or teaches how to become a data journalist, and how a data journalism education differs from that of traditional forms of journalism.

On the practical side, reporting on urban issues has been an important part of the journalism industry in recent years. R. Alan Walks and Richard Maaranen (2008), in their book *The Timing, Patterning, & Forms of Gentrification & Neighbourhood Upgrading in Montreal, Toronto, & Vancouver, 1961 to 2001,* posit gentrification as a process which began in the late 1960s. It was characterized by sporadic projects to assemble or clear cheap land for residential use, evolving into the 1990s and onward by attracting young professionals and other higher-paying tenants to areas that were becoming more urbanized. This phenomenon pushed out long-time residents who traditionally lived in these neighbourhoods because of their reasonable rent rates. Much has been done to investigate the causes, effects and fallout of gentrification by journalists in urban hubs in the United States. For example, Sophie Kasakove (2019) investigated for *The Week* the swarming of young professionals to Brooklyn, which had historically been the home to much of New York City's working class who have now been pushed out by high rents. In Seattle, Robert McCartney (2019) for *The Washington Post* describes the effect the technology giant Amazon had in displacing some of the city's Black residents from a traditionally African-American neighbourhood to make way for state-of-the-art buildings and amenities for their employees. Canadian cities have also seen this process take shape in Vancouver and Toronto, with rich or corporate investors buying up residential properties, only to leave them sitting empty, a phenomenon that has been reported on by Joanne Lee-Young (2020) for *The Vancouver Sun* and Murray White (2020) in Toronto for *The Guardian*.

While Montreal has by no means been spared by the changing urban landscape, the effects of 21st-century cities have been slower to take hold here. For example, rents remain relatively low when compared to the other major cities in Canada and the United States. However, rising rents and the recent housing crisis (Thériault, 2021; Gagnon, 2021) indicate it is

likely Montreal will be heading down a similar path as other cities in North America. Residents and local news coverage have already raised alarms about gentrification pushing people out of accessible housing to appeal to young professionals who may be able to pay more in rent. For example, tenant advocacy groups in the traditionally working-class neighbourhood of Saint-Henri have had to fight for the rights of long-time tenants as the area has become urbanized and therefore more desirable to people who are willing to pay more to live there (McKenna, 2019). More recently, Parc-Extension, which is following trends like those in Saint-Henri, saw conflicts break out between tenants and landlords over evictions imposed after people couldn't pay their rent due to the economic stall of the COVID-19 pandemic (Cori-Manocchio, 2020). These examples all indicate that gentrification is a pertinent issue in Canada generally, and in Montreal specifically.

While there has been important coverage already done, there have been only a handful of stories covering this topic using the tools of data journalism, which can unveil issues not immediately visible without the aid of data analysis and social science methods. These data journalism projects covering housing and gentrification in Montreal include, but are not limited to, a report from *Le Devoir* which used data scraped from rental ads on the Canadian classified site Kijiji to show housing is getting more expensive for Montrealers (Gélinas, Léouzon, & Pavic, 2021), and a crowd-sourced map cataloguing rental prices in the city (n.d.). Gentrification is not a blind spot for reporters. But there are numerous data sources with information on occupancy rates, rent increases and landlord-tenant disputes—including open data easily available to the public online, government and corporate annual reports, studies released by activist groups, and data accessed through freedom of information requests—which could be more widely utilized in journalistic coverage on the issue. Using these datasets as a starting point

could help journalists augment their reporting on housing issues in the city to tell stories from angles that traditional reporting may not touch on.

The study of urban issues in Canada is not confined to journalistic endeavors. The collection *Toward Equity and Inclusion in Canadian Cities: Lessons from Critical Praxis-Oriented Research* (Andrew, Klodawsky, & Siltanen, 2017) presents a survey of different issues of inequality and injustice being faced by cities such as Toronto, Winnipeg, Vancouver and Ottawa. The studies detailed in this volume show how traditional power structures of capitalism, colonialism and heteropatriarchy contribute to the challenges of the Canadian city-dweller, particularly if they are from a historically marginalized community or group. The authors indicate that in order to solve these problems, one must turn away from these structures.

It should be noted that *Toward Equity and Inclusion in Canadian Cities* excluded studies about Canada's second-largest city, Montreal. That is not to say Montreal has been completely left out of the conversation about urbanization in 21st-century Canadian cities—indeed a book chronicling gentrification in Canada was mentioned earlier in this section which included Montreal as one of its case studies (Walks & Maaranen, 2008). But that book was published in 2008 and many of the urban issues in Montreal that are currently being raised by community organizations are acknowledged in the press, but beyond that no further action is taken by government officials. For example, Le Regroupement des comités logement et associations de locataires du Québec (RCLALQ) (2020) has often provided analysis of housing statistics in the province of Quebec and city of Montreal, showing that there is a housing crisis and makes recommendations for control measures, such as a government rent registry. This crisis is also acknowledged by advocates and in the press (Lauzon, 2020), yet the Quebec government has established no such registry (Karwacki, 2021). Montreal, therefore, is still experiencing these

urban housing problems, and there is data available to put together compelling data journalism stories, aided by maps and visualization.

Research Questions

The main question driving this research-creation project is, how can data journalism be applied to the coverage of urban housing issues in the city of Montreal in order to inform robust ways of teaching and learning data journalism? Stories of urban housing issues are often driven by numbers, such as rental prices, property values, the number of landlord-tenant disputes, and so on. This type of data can be represented in the form of a map, across the scope of a city's boundaries. But, in examining this subject, several questions are raised about the nature of data journalism as a subsection of the larger journalistic field. The purpose of this project was to investigate how its practices can be learned, taught and incorporated into the existing work of journalism studies, and how the coverage of the city, housing and gentrification can be approached using the tools of data journalism.

Thus, this research-creation project seeks to address the following two research questions:

RQ1: How can data journalism methods be applied in a local news context to tell stories about housing and gentrification in Montreal? How can data journalism methods help find and tell important stories about housing and gentrification in Montreal and help fill gaps about change over time which traditional forms of reporting leave out?

RQ2: How can taking a learning-by-doing approach to a data journalism research-creation project help inform ways of teaching data journalism and training novice journalists in data journalism methods?

Review of Existing Research

The definition of data journalism itself appears to be aqueous within the journalistic profession and academia, as many authors and journalists agree that it involves a combination of computer and analytical skills apart from the traditional journalistic research methods of interviewing and observation used for newspaper and magazine stories (Meyer, 1991; Vallance-Jones & McKie, 2017). One of the professional roles Nikki Usher (2016) examines in her book, *Interactive Journalism: Hackers, Data, and Code,* is the data journalist who analyzes numbers to find and generate stories. "Data journalists identify as working closely with the analysis of data and see themselves as concerned with the presentation of this data for the public," (Usher, 2016, p. 97). She notes to be a well-rounded data journalist or put together a well-rounded data journalism team, one must draw on skills that use data, programming and interactivity. Data journalism does not always necessarily involve the skills of programming and interactivity, which people often lump together, as data journalism can be as simple as the inclusion of some data or statistical analysis as a supporting point in a news story. However, Usher (2016) does argue that these skills should not be mutually exclusive, and that any journalist practicing these forms of digital journalism should understand the others, as their combination often makes for the most compelling presentation of journalistic projects. One thing Usher—and indeed, most authors writing on the practice of data journalism—emphasize is the role that classic "shoe-leather" reporting skills (such as observing events in-person and collecting interviews) must continue to take to create a compelling data journalism project (Usher, 2016; Meyer, 2002; Vallance-Jones & McKie, 2017)

According to C.W. Anderson (2018) in *Apostles of Certainty: Data Journalism and the Politics of Doubt*, the foundational practices of data journalism began in the early 20th century,

with a culture of surveys present in social action groups during the Progressive Era.[1] In essence, social activists used industrial journalism techniques of interviewing people who were experiencing problems to try and discover a deeper meaning or cause. As Anderson (2018) explains, they applied social science methods that condensed participants' responses into massive datasets. The results of their efforts were widely accepted at the time, because at the beginning of the 20th century the quantitative approach, which is part of the traditional scientific method, was seen as more valid than qualitative methods in the humanities—plus data could just be more easily presented visually (Anderson, 2018). To promote their findings of wealth inequality and poor working conditions in American cities like Chicago and Pittsburgh, they shared their methods with the journalism industry at the time; soon some journalists were also using social science approaches and survey methods to find stories.

Philip Meyer (2002) in his book *Precision Journalism* also cites social science methods journalists should employ to accurately produce number-driven stories. *Precision Journalism* has appeared in multiple editions over its thirty-year publication history. In it, Meyer goes into detail about how a journalist can use social science methods, particularly accurate survey sampling construction, as well as how computer-assisted reporting (CAR) techniques, such as coding or even humble spreadsheet functions, can be used to parse through huge sets of data to find the most prescient numbers that will make for a compelling news story. During the last half of the 20th century, more and more journalists were starting to adapt the techniques described by Meyer to collect surveys and break stories on demographic participation in social movements and to predict election outcomes (Meyer, 2002).

[1] The Progressive Era took place between the 1890s and 1920s in the United States and was defined by social activists and politicians studying social issues, such as poverty and industrial working conditions, and calling for reform to improve the quality of life of Americans, particularly those in the working class.

Despite numerous updates over the years, Meyer's book lacks a review of data journalism tools that have gained popularity in the 21st century. Since his heyday, there have been a variety of data journalism textbooks and handbooks authored by former and current journalists which chronicle steps one can follow to practice savvy data journalism by understanding data collection methods, accurate data interpretation and the creation of compelling information graphics. For the university classroom setting, Fred Vallance-Jones and David McKie (2017) put together a textbook titled *The Data Journalist: Getting the Story*. This text teaches the practices of "traditional" or "industrial" journalism *alongside* newer data journalism and CAR techniques, as opposed to *adapting* older practices to a new landscape where one needs to be more technologically literate. The textbook's authors sum up this approach to teaching data journalism by writing: "As soon as you know a little bit about formulas and some basic functions, spreadsheets are an incredible tool to see trends and anomalies" (Vallance-Jones & McKie, 2017 p. 7). This attitude can be applied not only to spreadsheets, but to coding in languages such as JavaScript, which can be used to design data visualizations, and Python, which can scrape websites automatically for data and analyze data locally on a desktop.

Another example is *The Data Journalism Handbook* (Bounegru, Chambers, & Gray, 2012), a collaborative effort created by mostly European journalists to sketch out what their individual data journalism techniques are. The resource is an attempt to paint a collective portrait of data journalism at the current moment. The editors of this handbook underscore data journalism's importance in the 21st century because of the astronomical growth of the amount of data constantly being generated and stored about our society. "Gathering, filtering and visualizing what is happening beyond what the eye can see has a growing value" (Bounegru, Chambers, & Gray, 2012).

Meyer and Anderson will tell you that the way forward is with numbers, but without excluding the human element. For many of the stories that Meyer (2002) outlines in *Precision Journalism*—including stories about the demographics of riot participants in the 1960s and predicting elections—he begins with statistics and uses social science methods to parse out not only the underlying truth, but what would end up being the major headline. He would then go on to seek out the human sources, to add colour and detail to stories. When one looks at a spreadsheet full of Montreal housing data, it will not look engaging or exciting to most. It is thus the task of the data journalist to discover and then present the most compelling or useful information in a way that is comprehensible to the consumer. The presentation of stories focused on housing issues could be done using text, but words are not the only way in which the brain processes information. As Alberto Cairo (2013) notes in *The Functional Art: An Introduction to Information Graphics and Visualization:* "The brain doesn't just process information that comes through the eyes. It also creates mental visual images that allow us to reason and plan actions that facilitate survival" (p. 2). As Anderson (2018) underscores in his early history of data journalism in the Progressive Era, journalists started adapting social science methods from the social activists who were trying to advocate their causes. They did this because they noted how the visual presentations of activists' survey results at exhibitions and in publications caused the public to react to the social issues they were trying to shed a light on. Thus, the choice to supplement a story, which will eventually be told through words, with the help of intriguing visuals and interactive features becomes even more appealing.

The use of data visualization to tell the story of an urban problem with the aid of a map is a practice that is arguably older than the introduction of social science methods into the journalistic profession. Edward Tufte (1997) details in *Visual Explanations* the story of the

British doctor John Snow who collected data from around London to pinpoint the source of a cholera outbreak in 1854. He presented his findings on a map which showed a cluster of cholera cases around a particular public water pump which was proven to be the source of the outbreak. In the 175 years since Snow's synthesis of data into a map visualization, data journalists used and developed similar techniques to prove a theory or tell a story within a confined geographical area because of the ease with which the findings can be communicated visually.

While there is indeed academic research and professional interest in data journalism, research specifically about data journalism education has found the curricula being provided to data journalism students and trainees leaves much to be desired. Heravi (2018) points out that while many institutions are starting to offer courses in data journalism, they are usually introductory-level classes and do not offer training in advanced methods like scraping websites for data using Python or analyzing data sets using structured query language (SQL). This is a gap in data journalism education that needs to be filled because, as James Foust and Katherine Bradshaw (2020) conclude, "[…] coding skills are required in the production of online content and journalistic content is increasingly delivered and consumed online, coding skills must be present at some point in the journalistic production process."

Norman P. Lewis (2020) posits that data journalism has not maintained apace with practices in the journalism profession because there is a lack of qualified instructors and hesitancy over which skills to emphasize. The journalists who come into newsrooms and use data journalism techniques are generally not part of the "old guard," but are instead usually younger and maybe have a background in a computer science field, and therefore not many of them are clamouring after teaching jobs in university journalism departments (Lewis, 2020). This often means the professors in journalism departments who may lack experience in data

journalism are left to decide which relevant data skills to incorporate in their curricula. Add to this complication the fact that the technology and tools of data journalism are continuously evolving and changing, with many eventually becoming obsolete. An example of this would be the use of edge-notched cards, which Meyer (2002) discusses as a method to manually record data collected from surveys which has now been replaced with digital ways of storing participants' responses. As Heravi (2018) notes: "As a general observation, journalists and journalism graduates lack sufficient data skills" (p. 364). But she also points out that many journalism students show interest in learning about these types of courses and would sign up for them if they were offered (Heravi, 2018). The history of data journalism practices has seen a long evolution which has shaped its practice in newsrooms, however, there has been a disconnect between this evolution and that of journalism education, leaving journalism students without sufficient space to explore data journalism tools in their curricula

Methodology

This research-creation project seeks to use the practical tools of data journalism to create an interactive website that uses a map of Montreal populated with data visualizations, text-based articles, and photography to tell stories about gentrification and housing inequality in the city. The processes of creating the website and engaging in data journalism to report on housing issues in the city were used to create knowledge about the practice of data journalism to help inform more robust approaches to data journalism education.

This research-creation project mobilizes a learning-by-doing approach: I taught myself the coding skills and designed the website using a trial-and-error system to assess which tools and aesthetic choices—learned from the study of data journalism theory and criticism, outlined above—best communicate the story of housing and urban issues in Montreal visually. I used

journaling throughout every step of the project production to track my learning process and chronicle which methods were the most effective in communicating the story of Montreal's evolving cityscape. A central focus of my notetaking involved keeping track of how easy or difficult it was for a humanities-trained journalist like myself to adapt and learn the techniques of data journalism and visualization. Using myself as the test-subject is beneficial because, as Usher (2016) notes, many journalists who practice data journalism come from math or computer science backgrounds and so it is worthwhile to explore how efficiently someone without that training picks those skills up. It is also valuable to observe how this reversal (a humanities-trained person learning computer and digital skills, instead of a computer scientist adapting those skills to journalist storytelling) might result in different methods of website generation and storytelling.

Following the model of research-as-creation or research-through-creation as posited by Owen Chapman and Karen Sawchuk (2012), my aim in collecting data around how one becomes a data journalist or teaches and learns data journalism was to *reveal* things about data journalism practice that can be discerned through the act, as opposed to scholarly observation. The objective here was similar to what is described by Mia Lindgren (2014) in her article on radio production as "autoethnography" for scholarly research, a process of generating a self-reflective account of the production process to illustrate how the practice emulates the theory.

To document these observations, I divided the key elements of the project into categories and kept a record of how my experience coincided with practices of data journalism education defined in the literature summarized earlier and what journalists going through data training understand about the use of data journalism in the newsroom. In tracking my progress, I focused on the three categories McKie and Vallance-Jones' (2017) highlight as the major tenets of

successful data journalism projects, namely "data," "visualization" and "story," with a number of subcategories under each of them to add detail. Within the "**data**" category I observed data *availability* and the techniques used for data *analysis*. Within the "**visualization**" category I recorded what tools I had to learn and use for the website *generation,* as well as what had to be considered when designing the *interactive* elements of the website. Finally, for the "**story**" category I analyzed what methods of *traditional reporting* (i.e., non-data-driven reporting) were used in addition to data journalism tools and techniques and observing whether the *sources* I was able to find and include using these traditional reporting techniques were actually representative of the data. The goal with recording the process of producing data journalism from scratch in this way goes back to the project's dual goals of "doing" data journalism to report on housing issues in Montreal and learning how to teach data journalism in the process.

Reflections on Website Creation Process

This project was admittedly ambitious. Teaching myself to code with essentially no prior programming knowledge is something whole bootcamps are dedicated to. After learning the coding skills necessary for a project like this (only attaining intermediate coding literacy), there was the matter of understanding the dataset I would be using, generating the site and then sourcing the actual stories. With learning and teaching data journalism in mind, the overarching conclusion I came to while producing this project is that data journalism's nature as a sector of the industry that overlaps with so many other specializations (such as programming and traditional storytelling) make it a massive subject to teach and learn. Compounding the acquirement of data skills with trying to adapt these newfound skills to traditional storytelling techniques makes this learning process even more difficult to achieve independently. It was also determined that in creating a data journalism project that focused on a specific story, there was a

need to narrow in on a specific set of skills, instead of trying to teach myself everything. One of the things I found to be beneficial when balancing what I *wanted* to produce with what I was *able* to produce was relying on publicly available files and data, so I could focus on developing my own skill set, instead of generating the materials I would need to build a site like this via web scraping or survey collection.

Because of the project's scope at the start, I had to regularly re-evaluate my goals and readjust my plan to coincide with the level of competency I was able to gain with these skills over the course of six months. I began actually learning the coding skills I would need to build the site, researching what design techniques I would be able to utilize and analyzing the data I planned to use at the end of January 2021 and began to program and design the site and reach out to sources in April 2021 to have the project completed by the end of July 2021. Having the ultimate goals of creating a map visualization which told the story of housing issues in Montreal and understanding I wanted to chase stories which supplemented that visualization gave me direction in my work which allowed me to mark my progress (understand website structure by this date, begin data analysis by this date, start building website by this date, etc.) and motivated me on to the next step.

Having a directed and focused project such as this as an assignment or curricula topic could improve upon many of the strategies used to teach data journalism right now. As Heravi (2018) points out, much of data journalism education is taught by giving journalists basic data skills, as opposed to teaching them how to use data through practice. Asking journalism students or mid-career trainees to create a project such as this to acquire data journalism skills aligns with Heravi's (2018) further ideas that data journalism education is most effective when the curriculum is short and targeted. Understanding how to read a spreadsheet and how to build a

data visualization are still important skills but learning them through the creation of a specific project felt more engaging than having to sit through a semester's-worth of lectures on the topic.

In the following sections, I will discuss how my research-creation work on this data journalism project about housing issues in Montreal led to the conclusions that acquiring data journalism skills in a learning-by-doing model made it difficult to chases stories with a specific data-driven angle, but still provided space to develop robust stories about housing while including data. The process also led to the development of a specific set of data skills. This personalized set of data skills could in turn be applied to later data journalism projects, as it has been indicated that the collaborative spirit of data journalism—suggested by Usher (2016) as a crossover from the practice's relationship with the team aspect present in programming—was beneficial for the areas of the research-creation project itself which received guidance from people and learning programs.

Data: Availability

The main data set that was chosen for this project came from the Canadian Mortgage and Housing Corporation (CMHC), specifically their Rental Market Survey Reliability tables for the centre of Montreal. Each year the CMHC collects housing data through surveys with owners of multi-unit buildings across the country to glean information about housing through rental prices (how much housing costs), vacancy rates (how much housing is available) and turnover rates (how often people moved), broken down by neighbourhood. This data set was chosen because it is made public every year. In addition to current data, there is also a wealth of downloadable archival data from 2014 to 2019 that shows changes and trends over time. This type of data was preferable for the project I wanted to create, because I was aiming to show how the housing situation in Montreal had evolved over time and how certain neighbourhoods had been touched

differently by specific housing trends. As far as journalism education goes, a public set of data like this is a beneficial starting point for someone trying to learn data journalism skills like data analysis and data visualization, because the numbers were available to be used and analyzed, allowing me time and energy to focus on developing my own skills (Vallance-Jones & McKie, 2017).

As I was able to discover during my experience with asking for information using Access to Information (ATI) requests (which will be outlined below), public bodies often restrict the data they release. Trying to get a complete dataset can often take months and sometimes even years (Vallance-Jones & McKie, 2017), so starting off with an ATI request often means your data project is a deep enterprise piece that doesn't have a strict deadline. The piece I was trying to produce was trying to go deep, but not in a way that interrogates systems the way most stories fueled by ATI pieces do. Rather, this was done to present a landing page where the available information was already organized in a way that visually made sense and could be a jumping off point for further exploration on that topic. Taking a different route and putting together a social science survey would mean finding a sample and collecting answers from respondents, which would have required an immense amount of time finding people to talk to, organizing their responses and then analyzing that data. Similarly, while scraping advertised rental price data from a site like Kijiji may have presented a more accurate picture of the current situation by drawing all the advertised housing prices on that site at a given time, the skills I would need to acquire for writing that type of script would have been too time-consuming, given the dedication I had to give to acquiring other data skills for this project. Both the web scraping technique and survey data rely on third-party reported numbers, i.e., advertisers' asking price or respondents answers to the survey questions. I did end up using survey data that I had not generated, but this

was because the CMHC data is already generated by trained social scientists and economists and I was careful to include information about how the collection of this data can create different results from other collection types (i.e., web scraping) in the stories and explainers later when I designed my website.

As previously stated, outside of publicly available data, many working data journalists also employ access to information/freedom of information requests to obtain data from government and other public bodies that have not yet been made public. I thus sent ATI requests to the Tribunal administratif de logement (TAL) for data on the number of landlord-tenant disputes between 2018 and 2020, broken down by neighbourhood, and to the Office municipal d'habitation de Montréal (OMHM) for data on the growth of social housing units between 2017 and 2020, broken down by neighbourhood. The response from the TAL did render some results, but the information released to me was not broken down by neighbourhood as requested, apparently because the TAL decided this would be an invasion of privacy. The OMHM was not able to release any data to me, saying they had a new system which did not allow them to extract data from before March 2021 and the only project done in the recent years was concerning the Saint-Michel-Nord housing project (Hénault, 2020), which mostly consisted of rebuilding the same number of dwellings.

In the profession of data journalism, a vast amount of the data used in the production of data-driven projects is gained through ATI requests. One good example of this is "The Implant Files" from CBC News (2018), which used data obtained through ATI requests filed with Health Canada to report on the number of adverse incidents with medical devices occur in Canada. Even though time did not allow me to go deeper with the ATI requests to find the data I was looking for, this project did shine a light on the on the importance of teaching novice journalists how to

use ATI requests for data journalism specifically. I was able to practice how to word requests to get the data I was after and with the small amount of data I was able to get from the TAL, I got a chance to analyze data-heavy records that are not always publicly available. The lack of ease of access to government data is negative setback data journalists and data journalism students should be aware of, as Tabary et al. (2016) point out in their study on data journalism in Quebec. "This situation is not conducive to journalists conducting original analyses of statistical data," (Tabary et al., 2016, p. 81).

A mention should be made about the use of data that is obtained from scraping websites. Scraping, the process of writing a code script that pulls data from websites and organizes it into a document like a CSV, is frequently used by data journalists, such as in the *Le Devoir* piece on housing generated from a website scrape of Kijiji (Gélinas, Léouzon, & Pavic, 2021). Website scrapes are not the most complicated codes to write. There are templates you can use and there is even a pedagogical activity in McKie and Vallance-Jones' (2017) *The Data Journalist* which walked me through how to perform a practice scrape on a stock website during my research before I began work on building my own project. As with the situation with ATI requests, when deciding if I could write a scrape script, the ultimate conclusion I came to was that adding data that came from a scrape would just be too complicated and time-consuming with the skill-level I had and would overburden the content of the project. Again, the goal of this project was to provide a visual representation of publicly available data. Generating a set of data from website scraping would have most likely produced an entirely different dataset than that of the CMHC, and while it would have an interesting concept to compare these two data sets, it ultimately would have distracted from the intention of visually representing the story of housing issues in Montreal. Plus, there had already been several projects using web scraped data (RCLALQ 2020;

RCLALQ 2021; Gélinas, Léouzon, & Pavic, 2021) produced to tell stories and discuss issues about housing in the same city.

More data does not necessarily always equate to better data journalism. With the CMHC data set as my main focus, I knew I needed to create a website and produce actual stories about housing in Montreal. I had to consider how many overlapping sets of data a reader would be able process when trying to learn about housing from the site. Vallance-Jones and McKie (2017) and *The Data Journalism Handbook* (Bounegru, Chambers, & Gray, 2012) also suggest those starting out in data journalism focus on one or two data sets to develop their skills before moving on to more complex projects. In this learning-by-doing model, it was therefore more efficient to focus on one form of data access, but it was still beneficial for me to have investigated these other methods of accessing data to understand their process. A discussion of the pros and cons of using the publicly available CMHC data follows in the section on data analysis.

Data: Analysis

Having data is all well and good, but if you do not understand how and why it was collected, you could end up interpreting it in the wrong way and producing stories that are inaccurate. Data journalists should understand how data is compiled, so they may be able to explain any inconsistencies or anomalies (Vallance-Jones & McKie, 2017; Bounegru, Chambers, & Gray, 2012). Data journalism literature points to a number of ways journalists can become more data literate and better understand the data they work with, for example the previously mentioned *The Data Journalist* by Fred Vallance-Jones and David McKie (2017), *Data Journalism Handbook* edited by Liliana Bounegru, Lucy Chambers and Jonathan Gray (2012) and *Precision Journalism* by Philip Meyer (2002). I also found Joel Best's (2013) *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians and Activists* to be particularly

helpful when it came to comprehending how and why data is collected and what motivations or inconsistencies I should consider during my own analysis.

To understand how the data for the CMHC sets were collected, I scheduled an interview with an economist from the CMHC, who explained the method used to collect the data in these annual reports, that was outlined in the previous section. The purpose of this interview was to more thoroughly understand how that data was collected. The economist explained to me how they made a random sample of building owners and collected survey results from them and how those results could be reflective of the situation or, since data was collected from a random sample and not a universal survey (where every building owner would be surveyed), the results could be unknowingly skewed. I also asked about the empty spaces found in some of the CMHC's spreadsheets and the economist explained that when there weren't enough respondents or when data was spread over too far of a range, the results were considered inconclusive, and the category was left blank. These idiosyncrasies of social science survey methods align with those discussed by Meyer (2002) and Anderson (2018) in their observations on survey practice in data journalism. Both authors agree that journalists using data from these types of collections or collecting them themselves should be aware of how numbers drawn from these techniques can be presented and used.

The process the economist described and my use of the data from the CMCH was indeed very reminiscent of the symbiotic relationship between social scientists and proto-data journalists in the early 20th century, as Anderson (2018) describes in *Apostles of Certainty*. One of the things that Anderson does highlight, and which I was trying to keep in mind as I was using the insights gained through my interview with the economist as I sifted through the data, is that these 20th century social scientists had a certain agenda driving their work, namely to advocate for

particular social causes. The CMHC is by no means an advocacy organization, but the idea I was trying to carry with me when analyzing these numbers is that no data is compiled without purpose—purposes that may be different than what a data journalist is trying to achieve—or without human error and should therefore be approached critically. As Best (2013) underlines, data is often released to the public with a certain message in mind from whichever person or body published them. At the very least, data is collected and complied by humans and therefore includes at least some human errors. For this reason, I knew I had to be cognizant about who collected the data and for what purpose, the motivations behind releasing that data publicly, the collection methods and mistakes that could have been made and how all of this could lead to elements missing from the data.

The CMHC economist did admit when we spoke that the numbers which appeared in the reports may not completely reflect the situation on the ground, because of their reliance on landlord/owners self-reporting their rental prices, vacancy and turnover rates. This did lead me to question whether the use of this data was even viable. Indeed, the most recent data set the CMHC had published for the centre of Montreal, broken down by neighbourhood, was only from 2019. This is where the possibility of a scrape of rental ads from a site like Kijiji seemed to look more appealing, as a data journalist can write a script themselves and create a snapshot of a certain moment in time. The information from a scrape like this would not only be able to provide more recent data, but would also be able to cover the bases missed in survey sampling that led to inconclusive results. Later on, when I would interview an advocate from the Regroupement des comités logement et associations de locataires du Québec (RCLALQ), a tenant advocacy group for the province, she would explain to me that they had performed a

scrape on Kijiji ads for rental units in Montreal in 2020 and in 2021 and found drastically different rates than the what was represented in the CMHC data.

Using data from either of these sources would be seen as "procedural" by Meyer (2002) and Anderson (2018), because the data would have either been collected using social science survey methods, or from an advocacy group trying to champion their cause. Both authors describe in-depth how data journalism methods have their roots in social advocates sharing survey data with journalists, and journalists adapting those methods to chase stories. The CMHC data was chosen because it was able to show change over a longer period of time, but the comparison between different types of data collection should be brought up in discussion of data journalism, alongside the articles where data is being used, and in the classroom where data journalism is being taught. It is important for data journalists to explain the methods behind different data sets to their audiences and this why many of the data journalism pieces you see also include a "Methodology" section, such as the piece on Airbnbs in Montreal (Shiab, 2019). For the purposes of reflection on data journalism education, examining how different forms of data collection can lead to different results is indeed vital when it comes to training journalists, as they need to understand how public bodies and advocacy groups draw their conclusions from the data they collect.

The RCLALQ studies and the previously mentioned story from *Le Devoir* on housing generated from a website scrape of Kijiji (Gélinas, Léouzon, & Pavic, 2021) captured compelling pictures of the Montreal rental market at the current moment. In contrast, with this present project I aim to show change over time. This speaks to how already-published data journalism and advocate-produced studies shaped the goals of my own data journalism project. Obviously, these other projects produced by *Le Devoir* and RCLALQ focused on rental prices and when I

interviewed advocates from RCLALQ, as well as from the neighbourhoods of Rosemont–La-Petite-Patrie[2] and Hochelaga-Maisonneuve[3], they spoke about rising rents as their big proof. For these reasons, I wanted to include the CMHC's data on rents in my own work. There was also a lot of talk about "available housing" when I began talking to housing advocates (specifically one from the Hochelaga-Maisonneuve borough, who focused on affordable housing for those in precarious and low-income situations), so I decided to grab the numbers the CMHC had for vacancy as well. Turnover rate is a more complicated data set, which talks about how frequently apartments change hands. Since no one I interviewed for my stories seemed as focused on that piece of data as prices and availability, I decided to not include it. Just as the other data projects covering housing issues in Montreal influenced the decision of what data I would focus on, the external interest advocacy sources seemed to show in certain topics motivated me to use data that aligned with those interests to generate my site.

I began work on this project in the winter 2021, almost exactly a year after the CMHC had published their data for 2019. I made the false assumption data for 2020 would be available during this project's timeframe, however it was not. Thus, I moved forward with the data that was available and grabbed the rental market survey reliability tables for the years 2015, 2016, 2017, 2018 and 2019. This began the process of cleaning and selecting the data I wanted to use. These were spreadsheets with over 15 pages of information for each year. For each category the CMHC provided a statistic for each neighbourhood it defined,[4] which was then further specified by bedroom-types and a category marked "total." This "total" category was selected as the main

[2] Rosemont–La-Petite-Patrie is a borough designated by the City of Montreal with the same bounds and name as the zone defined by the Canadian Mortgage and Housing Corporation.
[3] Hochelaga-Maisonneuve neighbourhood defined by the Canadian Mortgage and Housing Corporation is part of the borough Mercier–Hochelaga-Maisonneuve designated by the City of Montreal.
[4] It should be noted that the Canadian Mortgage and Housing Corporation defines different boundaries for neighbourhoods than the municipal boundaries drawn by the City of Montreal.

category of interest as it was the easiest one to compare from year to year to show change over time without complicating the analysis with extra numbers from different bedroom-types.

I then began a process called "cleaning the data," where I would scan the data for any human errors or inconsistencies like misspellings, missing entries and other inconsistencies (Vallance-Jones & McKie, 2017). I decided to do this manually as opposed to with a program, like data cleaning tool OpenRefine, because there weren't many actual words in the data set (as one might find in, say, a list of addresses), so there wasn't room for spelling errors. The missing pieces of data were explained by the interview with the economist from the CMHC. Each spreadsheet for each year also provided the numbers for each category for the previous year, so I spent a decent amount of time comparing the numbers between years. As was explained above, some of these categories were left empty when the results of the survey proved inconclusive for a given year. Cleaning data and understanding how it was collected are vital skills that any journalist dealing with data needs to learn. "[…] unlike a document that might contain one killer phrase, a dataset contains several possibilities, making it worth the effort to find, clean and maintain as many of them as possible," (Vallance-Jones & McKie, 2017, p. 76).

After this, I started something which is described by Vallance-Jones and McKie (2017) as "interviewing the data," as I was parsing through the numbers I had in front of me, asking questions about what they meant and which pieces of data could lead to the most pertinent stories. I transferred the numbers I knew I wanted to use from the "total" category for each year and neighbourhood for rent price and vacancy to their own spreadsheet so I wouldn't be overwhelmed by all the other data in the market survey reliability tables. I performed my own calculations to see which neighbourhoods saw the greatest change over time. I used a variety of calculations, including simple differences, average change per year and percent change over

time. The goal here was to see which neighbourhoods yielded the most outstanding changes and if there were any that had significant changes over multiple categories. I knew if a certain neighbourhood showed significant change in, say, the simple difference category and the percent change over time category, there were large statistical changes happening in that neighbourhood that could lead to a big story, or at least a compelling hook that used the data to indicate to readers why they should be interested in the issue.

This would initially guide me to which neighbourhoods I wanted to follow as main topics for the stories that would populate the site. I saw the neighbourhood of Le Sud-Ouest; Verdun[5] had seen the greatest percent increase over time for rent and the greatest increase in simple dollars, so I wanted to look into that. I was even curious about how the neighbourhood of Notre-Dame-de-Grâce; Côte-Saint-Luc; Hampstead; Westmount; Montréal-Ouest[6] had only seen a 0.5 per cent increase of only $5 over five years, while the average increase for the city over that same time period was 11 per cent. As I will discuss in the section covering traditional reporting, I was not able to pursue each of these stories in the way I would have preferred. Based on the sources who got back to me, I ended up pursuing stories in Rosemont–La-Petite-Patrie and Hochelaga-Maisonneuve, which each saw an average increase in rent of 12.94 per cent and 9.16 per cent over five years, respectively.

Visualization: Generation

Coding, computer science and programming languages are huge topics which people dedicate whole degrees and careers to. If you want to create a website, it is possible to do so

[5] Le-Sud-Ouest and Verdun are two different boroughs within the City of Montreal. The Canadian Mortgage and Housing Corporation's classification combines the two boroughs as one district, "Le-Sud-Ouest; Verdun."
[6] The CMHC classification of Notre-Dame-de-Grâce; Côte-Saint-Luc; Hampstead; Westmount; Montréal-Ouest is made up of the independent municipalities Côte-Saint-Luc, Hampstead, Westmount and Montréal-Ouest on the Island of Montreal and the neighbourhood Notre-Dame-de-Grâce, which is part of the City of Montreal borough Côte-des-Neiges–Notre-Dame-de-Grâce.

without a knowledge of coding—there are tools available to do this through online design templates that have most of the design and coding already done for you and that you can personalize with your own content. Many of the publications I have worked for in the past have used template styles, like WordPress, to host their online versions, but if you want to add even the smallest amount of personalization (for example, an embedded link of a Google Maps element), a basic understanding of the coding language HTML comes in handy. But the project I wanted to create required much more personalization and interactivity. It was thus worthwhile to me, and would be worthwhile to any journalist interested in producing interactive online data journalism, to learn how to code in more than one programming language.

To build my site I had to decide which programming languages would be most beneficial to learn. The conclusions I initially came to when I was deciding what languages to learn were based on the definitions provided by McKie and Vallance-Jones (2017) in *The Data Journalist.* HTML, or Hypertext Markup Language, is one of the simpler languages you can learn and is the foundation for most sites on the internet. Cascading Style Sheets, or CSS, deals more with colors, placement and design of web pages and is important to learn if you don't want your site to look like a Word document set in Times New Roman. JavaScript is a functional language which causes reactions to a user's actions on the page—for example, a pop-up appearing when you hover over a certain item. It is also the language that is used for the data visualization library D3.js, which I knew I wanted to learn because it is also used by large publications, such as *The New York Times,* for their data visualizations. I was also initially interested in learning Python, as this is a language that can be used to scrape websites for data and quickly analyze data locally on your own computer.

After settling on these four programming languages—HTML, CSS, JavaScript, and Python—as a baseline, I enrolled myself in online courses on the programming education site, Codecademy. The learning process here was by no means simple—it's safe to say it was an arduous process for me. I had to commit several hours every day to walking myself through tutorials that sometimes struggled to hold my attention. A lot of the time I didn't even feel like I was actually retaining the information. However, this learning process was not impossible to get through, and there were a few entertaining activities which made some of my programming learning sessions a little less painful, such as being placed in a simulation as a grocer who needed to keep track of all their inventory using a Python script that organized that data. Once I had made it through the introductory courses for these languages I had to figure out where and how to write my own code so I could start producing my website.

Luckily, I was able to draw on the help of a friend of mine who is a programmer. She taught me how to use the command line in my operating system (a local way of executing functions manually on your desktop), got me set up in Microsoft Visual Studio Code (a local code editor that allows you to edit and preview a programming project without it going live online) and helped me register a GitHub repository online, which allows you to share your project publicly or privately with registered collaborators and can even host your project so it can be eventually published. This contribution to my project merits discussion, not only because she deserves credit for teaching me these things, but I feel as though her assistance also speaks to a general spirit of friendly collaboration within the programming community which data journalists should be aware of (Usher, 2016). If there is a bug in your code, there is usually someone who has run into the same problem and is willing to talk about it either in person, or on forums such as Stack Overflow (n.d.). Authors like Vallance-Jones and McKie (2017) also talk

about how data journalism is best done as a collaborative effort, so that data journalists don't have to do everything on their own or be specialists in all the tools and technologies. Once I was set up with all these tools, I had the coding knowledge in HTML and CSS to produce some skeletons of the different pages I wanted for my website. Now it was time to move on to the main feature, which I wanted to be the landing page for my project: the map.

As previously mentioned, D3.js is a popular data visualization library and I took inspiration for the site I wanted to create by browsing the options they had. Through this, I learned the map I wanted to create was called a "choropleth," which is a type of thematic map where different areas of it are shaded differently to represent different values of data. This is a type of data visualization I had seen so many times and was anxiously trying to emulate, and I didn't even know the name of it. But I knew it would be the best way to represent the different rent and vacancy values across different neighbourhoods because, according to *The Data Visualization Catalogue*, it "provides a way to visualise values over a geographical area, which can show variation or patterns across the displayed location," (Ribecca, n.d.). The *Catalogue* also points out that one of the pitfalls of choropleth maps is that audiences can't accurately compare or read values across different regions, so I knew I would have to include a feature that would display the value for each region.

After working my way through a D3 tutorial on how to create a choropleth map of the United States that represented data on different levels of education broken down by county (Ganesh H, 2020), I learned choropleth maps are powered by a file called a GeoJSON, which stores geographical information, which will appear as polygons representing any set geographical boundary in your visualization, alongside with whatever data you want to represent. Thankfully, the City of Montreal's Open Data Portal (2020) has publicly available GeoJSON

files which have boundaries already drawn for the city's different boroughs, so I didn't have to generate my own from scratch. But, as previously mentioned, the boundaries for Montreal's boroughs are different from the boundaries the CMHC uses to define the neighbourhoods it collected its data on. This meant I would have to edit this GeoJSON file from the City of Montreal to include the CMHC data driving my project.

I had wanted to use D3.js to make my map, because most of the code is already written, and you just need to customize it by adding in your own data with something like a GeoJSON. But I wasn't able to find a D3 method which would allow me to customize my own GeoJSON, so I had to look somewhere else. The online tool I eventually used was Mapbox (n.d), which allows you to upload your GeoJSON, redraw the lines of the polygons, add your own data and even create custom choropleth maps. Mapbox also provides a collection of customizable interactivity tools which will be discussed in the following section. Once I had designed the choropleth map representing rental prices for 2019 by neighbourhood, which I knew I wanted to use as the homepage of my website because it seemed like the piece of information people would be most interested in, I was able to follow Mapbox's tutorials on how to use JavaScript to embed this design into your own code. I went back to the webpage skeletons in Microsoft Visual Studio and added the map to what I had already programmed. When I opened the file preview in my browser, the map was there, showing how rent prices varied from neighbourhood to neighbourhood. The foundation of my website was complete and now it was time to customize it by adding different features so consumers could interact with the different pieces of data and stories I wanted to present.

I gained most of the knowledge I have about coding from courses on the site Codecademy, discussions on the Stack Overflow site, a supportive friend who has a background

in programming and was able to give me advice on my code and—in all honesty—a readily available supply of YouTube tutorial videos. The information to become moderately proficient in these skills is available (some of it for a subscription price), but there was no organization to guide me through which skills I should be trying to acquire to produce a project like this. I just knew what elements I wanted to create, and I would go after them and see what I was able to learn. I would now be able to give you a syllabus of what you would need to cover to create a site like the one produced for this project (HTML, CSS, JavaScript, GeoJSONs and Mapbox), but for another project, I might not be able to tell you where to start. Even my endeavour to learn Python turned out to be a bit of a misdirection, as I didn't end up using it for this website. But that doesn't mean it couldn't be helpful for another project in the future. There is such a large universe of website generation tools and programming languages that data journalists can learn—but to learn all of them would be impossible. I thus found it was most useful to focus on the story I was trying to tell and adopt the skills I needed along the way to best tell that story. As Vallance-Jones and McKie (2017) put it: "Being conversant in the broader structures of the web, and understanding how different applications speak to each other, is an increasingly important skillset for the modern data journalist" (p. 221), but being conversant is all that it takes. The two go on to say that it is good to have one programming language solidly under your belt (Vallance-Jones and McKie, 2017), but this is not nearly as efficient as having a mastery of a basic set of programming skills that can be adapted and added to in the future.

Visualization: Interactivity

The foundational platform of a map seemed like it would be the most efficient visual to construct this web application around. But, the use of a map, a geographical representation of the city of Montreal, served a critical and journalistic purpose. In *The Functional Art: An*

Introduction to Information Graphics and Visualization, Cairo (2013) outlines which data visualization tools journalists should use because of their evolutionary appeal to the human senses—in other words, how to make the most important information jump out the fastest in a visualization. When putting together a visual representation of the story of housing in Montreal in the form of a map, the hope was to immediately present the story of the inequalities, the injustices of an urban space by having the landing page comparing housing statistics for different categories and different neighbourhoods be the first thing to catch the audience's eye.

When deciding which interactive elements I wanted to implement on the site, I was trying to balance this idea of communicating a story visually with what would be the easiest way for a consumer to navigate and take in that story. This began with the landing page I spoke about in the previous section. Based off my interviews with an advocate from RCLALQ and the economist from the CMHC, I knew the best place to start would be the piece of data people are most interested in—rent prices—from 2019, the most recent year data was available for. The map showed higher rates of rents by neighbourhood in increasingly darker shades of red, with more expensive neighborhoods in darker shades of red, and less expensive neighbourhood in lighter shades of red. I wanted readers to be able to explore the story of change over time, so I added a legend feature that would link them to different maps for different years showing how rent prices have gone up by neighborhood. I created these maps using the Mapbox tool again.

I then repeated the process for the data I had for vacancy rates and made sure you could distinguish between the two by colour – red for rent prices and green for vacancy rates. I was left with a bit of a dilemma when designing the vacancy map, because traditionally larger numbers are represented with darker colours (Wong, 2013, p. 46). This made sense for the rental map, as the darker areas meant rent going up and less access to housing. But with the vacancy rates,

higher numbers meant more vacancy and therefore more access to housing. Matching the scale of lighter to darker to correlate with growing numbers was kept for consistency when going between the two categories, but I still sometimes get stuck on this as the two maps could be implying different things. To make it so you could navigate between the rent price and vacancy rate data sets, I added another feature that would link you to the 2019 map for each category because, again, the most recent data was what would most likely be the most interesting piece of information for someone entering the site.

Finally, to provide a more synthesized representation of the idea of change over time, I added two more maps which presented rent per cent increase over the five years available on the site—represented in blue—and vacancy rate change over those same five years—represented in yellow. These were numbers I had already calculated in my analysis. I added the links to these change-over-time maps to the legend feature on each page which contained the links to the 2019 rent map and the 2019 vacancy rate map as this would differentiate between the four categories of data you could explore. On the rent prices and vacancy rate map pages, there is always the other legend feature which links you to different maps representing those categories over five years. But when you go to the rent per cent increase over time map and the vacancy rate change over time, this legend feature for years disappears, to signify that there weren't multiple years-worth of data for these categories.

The site I was left with at this point adequately communicated the story of rent increasing and vacancy rates changing visually because it was representing data over five years from different angles by including multiple years' worth of data from both categories. I then set out with how to fill in the further questions the audience might have about the map and the housing situation with text because as Ribecca (n.d.) notes, the choropleth maps I was using only

represent colours across a scale without giving you specific information about what the complete number is for each neighbourhood in each category. Firstly, the scales of colours on each map needed to be given definitions, so a key was added to each page to show what the colours represented. I then coded a hover feature in JavaScript that would allow you to hold your mouse over a neighbourhood on the map and see the exact number which was being represented for that neighbourhood in that category. I also knew I wanted to embed links to stories on the map, over the neighbourhoods they would be discussing. Again, using JavaScript, I added points to the map and so they wouldn't conflict with the hover feature; the points reacted with a click which would trigger a pop-up giving you a blurb of the story associated with the neighbourhood in question, and providing a link to the full story. This JavaScript feature is preferable because I can easily add more of them to the map as more stories get added to the site. I had wanted the article pages to be a bit more engaging with maybe more graphs and interactive data elements, but I had to come to terms with my coding skills and included a basic scrolling article page, designed with HTML and CSS, which only included text, photos and static graphs I designed externally.

I also felt like the map landing pages needed to have some explanation of what the project was and why it was created. So, as not to overwhelm the large amount of visual information already on the page, two small info boxes were added with short texts on the housing situation in Montreal and the methodology of how the data for the site was collected and presented. These small texts were provided to give just enough information without overburdening the landing pages, but they did not include all the information I wanted the reader to know about housing situation and methodology. To balance this, underneath these texts are "Learn more" buttons which open pop-ups that contain an original article on the housing situation in Montreal and a deeper explanation on methodology. A further discussion of the

contents of the housing situation article and the methodology box will follow in the section on traditional reporting.

As I was designing the site, the things which were motivating my decisions were ease of understanding for the audience. If something didn't look quite right in the design, I would change it, but I was mostly trying to look at the site like I was reporting and writing a story and designing elements that would lead the reader through that story. I didn't want any of these elements to conflict with each other and I wanted the audience to be able to easily navigate through the site to the information that would be most important to them. Therefore, I tried to present as much information as I could on the main landing pages without overwhelming the site. The article pages could have been more interactive, but—as will be discussed in the final two sections—the hope with the articles was to highlight the stories that existed outside of the data.

While analyzing the data and then using the results of that analysis to create the map site, the main goal was always ease of communication of the story of the housing situation in Montreal using the tools of data journalism and visualization. In trying to achieve this goal I was considering how data was collected and by whom and what pieces of data would most succinctly communicate the point of the story of change over time. After I had decided on those pieces, I was able to focus on what programming and design skills I needed to prioritize and acquire to make sure this story was presented in a way that made sense to the audience. This whole process was a practice of extracting pieces of data from the larger dataset to simplify it and reduce it to its most prescient points and whittling down the wide range of coding and visualization techniques suggested to me to focus on the ones which would serve the purposes of the story best.

Story: Traditional Reporting

The goal for this site had always been for it to be an ongoing project which would host different types of stories about housing in Montreal and be updated as new data was published. The initial hope, as mentioned in the *Data: Analysis* section, was to find the most outstanding pieces of data, such as the highest rent increase or the lowest vacancy rate, and chase stories in those neighbourhoods, searching for the "why" of these superlatives by talking to the people, such as advocates, tenants and property owners, who understand how these large numbers or drastic changes have occurred over the past five years. The traditional reporting techniques I used to chase these stories did not lead to entirely successful results, as I was not able to report on the neighbourhoods I originally set out to cover. The story about Le Sud-Ouest; Verdun still seemed appealing because that neighbourhood had seen the highest rent increases in the city, and I was also looking into chasing a story in Villeray–St-Michel–Parc Extension[7] because it had a low vacancy rate that had been decreasing over the past five years. With both these neighbourhoods, I reached out to advocates asking them if they could talk to me more about what was going on their neighbourhoods, but they said they were too busy to speak given the upcoming moving day on July 1st. I posted in groups against gentrification that were specific to these neighbourhoods, and no one responded. Since I couldn't get a firm enough foothold in these neighbourhoods that would give me a knowledgeable starting point for my reporting, I didn't feel like I could pursue stories in these neighbourhoods at this point. Given these setbacks, I had to pursue only stories where the initial sources got back to me and gave me an idea of the overall picture in their neighbourhoods, which ended up being housing advocates from Rosemont–La Petite-Patrie and Hochelaga-Maisonneuve. With these sources giving me a good

[7] Villeray–St-Michel–Parc Extension is a borough designated by the City of Montreal with the same bounds and name as the zone defined by the Canadian Mortgage and Housing Corporation.

base of information to start with, I was still able to generate some coverage which gave voice to housing issues people were experiencing in Montreal that have not already been covered by the news media in this city.

The tactic I was taught to follow in my undergraduate journalism training from Dr. Carolyn Nielsen at Western Washington University was to contact sources who not only studied the issue and advocated the issue (i.e., experts and community organizations) but also, perhaps even more importantly, those who lived the issue (Nielsen, 2010). This is what I attempted to do for these stories covering housing issues in Montreal. As previously discussed, I started by contacting the economist at the CMHC to see what he could teach me about how the data I was using was collected. It bears repeating in this section on "traditional reporting" that this is an extra step that deviates from the research journalists do for stories that don't involve data. After talking to the economist, I cast my net wide and reached out to housing advocacy groups for every neighbourhood I wanted to cover, plus a few extra in case they didn't get back to me. This is also when I reached out to RCLALQ and Le Front d'action populaire en réaménagement urbain (FRAPRU), a citizen action group for housing rights.

The hope here was also that the housing advocates could lead me to the sources who lived the issue and had compelling stories to share. This is the point when I had my interview with RCLALQ where they explained the overall housing situation in Montreal to me. The picture RCLALQ painted was quite dire. I was also starting to reach out to people in late April and early May 2021, right when most housing conflicts occur leading up to the traditional July 1 moving date in Quebec, so they explained that many of the housing advocates I reached out to would be inundated with work and might not be responsive. This proved to be the case as FRAPRU declined to be interviewed and only housing advocates from the neighbourhoods of Rosemont–

La Petite-Patrie and Hochelaga-Maisonneuve got back to me. Thus, these were the places I decided to pursue stories on.

I decided to use the quotes RCLALQ gave me as part of the housing situation explainer article on the landing page because this overarching information would be a good entry point for readers coming to the site. The information from the RCLALQ data scrape of Kijiji and how it compares with the CMHC data acted as a pathway that could increase the audience's data literacy. It felt like this would be an important piece of the project, as it would not only support data literacy education, but it would also increase transparency (along with the methodology section) about how data was used for this project. Transparency about methods is one of the core tenets of traditional journalism, as posited by The Canadian Association of Professional Journalists Ethical Guidelines (2011), and there is no reason why this principle cannot be translated to data journalism.

Even though I was not going to be able to pursue the stories I had initially been most interested in, the stories in Rosemont–La Petite-Patrie and Hochelaga-Maisonneuve still proved to be worthwhile. I had wanted to possibly do a story about Verdun, as the data indicated it had the highest rent increase in dollars and in percent change over time. However, as my source at RCLALQ told me, Verdun is already widely recognized as one of the most gentrified neighbourhoods in the city. While that still merits coverage, by reporting on the neighbourhoods of Rosemont–La-Petite-Patrie and Hochelaga-Maisonneuve and using data-driven techniques, I was able to cover places that were receiving less attention in the press but may very soon be in a similar situation to Verdun. The housing advocate source from Rosemont–La-Petite-Patrie even referenced a piece of data analysis in her interview. She said Rosemont–La-Petite-Patrie now has nearly equal rental rates to Plateau-Mont-Royal, where the rent has been rising from where

Rosemont's rent was five years ago for decades, meaning Rosemont was becoming more expensive at a faster rate. One of the main points of this project was to use the predictive power of data combined with social science methods and traditional reporting to see what stories about gentrification could be revealed. The data shows rising rent and lowering vacancy rates in the neighbourhoods of Rosemont–La-Petite-Patrie and Hochelaga-Maisonneuve, indicating housing issues were becoming more prevalent in those two neighbourhoods. Once I had information from housing advocates, I needed to find other sources that could make for a compelling story.

The sources I was able to reach after this initial round of research influenced the types of stories I decided to produce. As previously mentioned, a story on the overall housing situation was presented to give space to the information provided by RCLALQ and with the goal of increasing data literacy by explaining how different pieces of data about housing are collected. The other two pieces on Rosemont–La-Petite-Patrie and Hochelaga-Maisonneuve were dictated by how far my reach was able to go with my sources. With the Rosemont story, the first housing advocate I spoke to connected me with many people, so I was able to collect six interviews with renters in the neighbourhood (even though I only ended up including three) and write a feature-length story. This provided space to explore the topic and incorporate data in a way that was seamless and would feel familiar to readers. For the story on Hochelaga-Maisonneuve, I had a great interview with a housing advocate who had a lot to say specifically about the situation in her neighbourhood, but unfortunately she was not able to connect me with any renters. However, I still had all these quotes and a set of data about how the housing situation was changing in Hochelaga-Maisonneuve. To put this research to use, I went out and took photos in Hochelaga on moving day and created a piece that incorporated the results of the interview and the data points I had been able to extract from the CMHC tables about the neighbourhood.

Although the initial goal of publishing stories by following the juiciest data angle was not achieved with the stories I was able to produce for this website, what I was able to accomplish gave me space to develop "traditional" stories with supplementary points about data used to support those stories. The idea that data should be a part of the story, but not necessarily drive the story is something journalists should keep in mind as they are learning to incorporate data into their stories (Vallance-Jones and McKie, 2017; Meyer, 2002; Anderson, 2018). Data-driven methods can add depth and context, and a different entry point for readers and reporters when coming to story, but that story would be lacking if it highlighted just the statistical elements without engaging the audience with ideas behind how those numbers appeared.

Story: Sources

As was already alluded to in the previous section, finding sources for these stories proved to be a challenge. I had started out by viewing the data as a source unto itself, but when I was mentally putting together the hierarchy of which sources would lead to what, "data" was at the top of that pyramid. However, I had to rearrange my process so I would actually be able to produce stories that news audiences would want to read. As Vallance-Jones & McKie (2017) and multiple case studies in *The Data Journalism Handbook* (Bounegru, Chambers, & Gray, 2012) note, a good news story does not end at a compelling data hook. The embellishment and support which comes from the inclusion of further research and the experience of actual human beings are what gives a story depth and the audience its importance. As Vallance-Jones & McKie (2017) put it: "[…] the data tells you what, and the people tell you why" (p. 249). The direction I ended up going led to a collection of articles that balanced human stories with points about data that correlated with these sources' experiences. While collecting information from these sources, there were important issues about story structure and ethics that also came up.

I reached out to all the housing advocates in the city I could and scheduled interviews with those who were available and interested. There was only one housing advocate who passed my information along to a group of people she had worked with who lived the issue and then those people reached out to me letting me know they were interested in being interviewed. During the interviews, I did not ask leading questions about the data, as this was not what I was looking for from these sources—I already had the story the numbers provided through my analysis of the data set. Instead, I wanted to see how their testimonies lined up with what appeared in the data and if there were stories they presented which did not appear in the data. I asked them about the specifics of their personal situation, how they had seen their neighbourhood changing over the past five years and how they felt their situation fit into the patterns they saw in their neighbourhood. Even though many of them naturally brought up the data trends of rising rent and less affordable housing, more of the interview was dedicated to them explaining why and even how they felt like their rent was going up. For example, in the story on Rosemont–La-Petite-Patrie, multiple sources acknowledged that rent was indeed going up and linked this to a trend toward gentrification in that neighbourhood which they saw as being harmful not only because it meant it would be more expensive to live there but since the area was becoming more "trendy" (with more speciality cafes and bars moving in), what had traditionally been a neighbourhood for families was now inadvertently pushing those families out. This was the type of information I was trying to gain from these interviews. I knew from the data set I had that rent was already going up, I needed the sources to provide me with information that would complement that point and make for a more robust story.

At this point I had also been trying to schedule interviews with the Tribunal administratif du logement and the Quebec Landlords Association to include their perspectives in any of these

stories, but the TAL declined, and I didn't hear back from the Quebec Landlords Association before the deadline for this project. With a topic like this, it would be important to include comments from both organizations, but if they are not willing to respond, there is only so much you can do. I found this to even be the case with some of the housing organizations I reached out to. Many of them cited being too busy, because this project was being produced during the spring and summer, when housing advocates are most inundated because of the upcoming moving day in Quebec on July 1. But I also had one person tell me they were less willing to be interviewed because of the nature of the project being for a Master's degree and not being attached to a known news media outlet or publication.

A note should be made on the ethical considerations that had to go into sourcing a story about housing. Many of the sources asked to remain anonymous because of the precarity of the situations they were discussing, which often involved conflicts between landlords and tenants. Discussing their situations publicly, with their names attached to their comments, risks causing further conflicts and ruptures in their living situations. To further protect these sources from situations that might endanger their housing situation, the landlords and companies that operated their buildings were not contacted to directly discuss their tenants' situations. To be sure, this presents a problematic situation when it comes to fair and balanced reporting, as the "other side" of a story should be allowed to respond. But in the interest of doing the least amount of harm, as outlined in the CAJ Ethical Guidelines and the Society of Professional Journalists' Code of Ethics (2014), the decision was made not to discuss specific tenant situations with their landlords in order to protect vulnerable sources. The stories of the people I spoke to were also in-line with other stories about housing in Montreal and were therefore not completely unbelievable. The tenants whose stories I used didn't call out their landlords or building owners by name and their

experiences were included because they related to much bigger problems and patterns in the city, rather just to specific people or companies. To obtain a more general explanation of the position and roles of landlords and building owners in Montreal's current housing situation, I contacted the Quebec Landlords Association but, again, they did not respond by the deadlines for this project[8]. To be transparent, a disclaimer was placed at the top of the story that involved landlord-tenant disputes, to explain the need for anonymous sources, why the landlords weren't included and the attempt that was made to include the Quebec Landlords Association.

These sourcing issues and ethical dilemmas are not something that is exclusive to data journalism, but are issues that journalists come across in all forms of reporting. Reaching out to sources outside of the data was vital in the production of these stories, because it provided the context needed to bring to the data I was using into a more compelling light. Stories driven by data can grab people's attention to an extent, but without human interest there is no way for the audience to connect to what they are reading. Throughout these reflections I have alluded to the use of data as a "source" and while carrying this mindset and translating traditional reporting methods to data journalism methods, it became clear that—much like with traditional journalism—other sources outside of data are needed to compliment and balance each other to create a compelling story. As an example, the fact that vacancy was going down in Hochelaga-Maisonneuve would be an interesting enough angle for a story, but to include the testimonies of a housing advocate explaining her observations of seeing this happen over the past decade and why she believes it is happening make it all the more intriguing. This combination of traditional sourcing of people involved in the issue with the facts about the data is what I found to be one of

[8] The Quebec Landlords Association did respond to requests for an interview after the research-creation thesis project was submitted for review, which meant there wasn't enough time to schedule interviews or include additional information from them. Should an interview be secured, the stories will be updated with their perspective and responses.

the most important building blocks of data journalism while creating this project. It is not that either of these pieces felt incomplete or untrue when they were standing on their own but using them as support for one another only reenforces the urgency of the story of the housing situation in Montreal which I was trying to tell.

Further Exploration

The housing and gentrification story has been widely reported on in many cities and is also an understandably emotional story that involves high stakes for the individuals involved on either side of the displacement of communities. It is also a story that has been framed almost statically, even though it is a continually-evolving process in many North American cities. Even when previous attempts have been made to report on this story using data journalism and interactivity, such as the Toronto Gentrification ArcGIS site (n.d.), it is often framed as a phenomenon that is finished. The site I produced provides a look at how gentrification has taken hold in the city with data visualization and text illustrating the previous patterns of urban evolution. Even though one can see where those patterns may be heading next, the future of the process is not addressed in the project.

One of the strengths of data journalism is that it can be used, not necessarily to predict the future, but to see trends (Anderson, 2018). Furthermore, its value is in assessing those trends in a way that can illuminate what the future of that problem may look like (Meyer, 2002). The hope for this site is that it can be updated as more data is published and further trends that appear can be analyzed and reported on. But given that this project was being created as a part of a research-creation thesis, I was bound by a limited timeline that did not give me enough time to explore how this site could have evolved as more data became available.

Concluding Statements

The focus of this research-creation project was to examine how data journalism can be learned and applied to the coverage of housing issues in the city of Montreal. The project itself was designed around a process that entailed learning-by-doing many of the data journalism tools and techniques discussed by Vallance-Jones & McKie (2017), Usher (2016), Anderson (2018), Meyer (2002) and others in their texts on data journalism history and practice. In examining this subject, several questions were raised about the nature of data journalism as a subsection of the larger journalistic field. This research aimed to investigate how data journalism practices can be learned, taught and incorporated into the existing work of journalism studies, through coverage of the city, housing and gentrification.

Over the course of six months, I was able to learn the data analysis skills required for a project such as this, become proficient enough in coding to design my own site and use the data I had access to as a guiding element in the stories I produced about housing. Throughout the production of the stories for this project, issues with sources' availability and timing made it difficult to produce the specifically data-driven stories I had originally anticipated. However, the production of this project did provide different avenues for learning about the story of housing issues in Montreal and important space for coverage of specific housing stories for certain neighbourhoods, while also incorporating some supporting data elements, which embellished on each story's main points.

The route I chose to take with my coverage added to the current focus on the city's housing issues by filling a gap and telling a story of change over time, rather than a static picture of a certain moment, as reporting on this subject using data-driven methods had already done (Gélinas, Léouzon, & Pavic, 2021). Using data over a five-year period added context about how Montreal has arrived at this point in the housing situation. The maps I was able to produce

showed that not only was rent getting higher and vacancy getting lower in Montreal, but they had also been doing so over the course of several years and were affecting different neighbourhoods in the city at different rates. After I was able to illustrate this point using data visualization, I was able to reach out to sources and have them further confirm this with their experiences in housing advocacy or personal housing issues in their neighbourhoods. The combination of testimonies from human sources and presentation of data analysis made for more robust reporting—particularly in the two neighbourhood-specific stories I was able to produce in Rosemont—La Petite-Patrie and Hochelaga-Maisonneuve—with these elements backed up by not just one piece of data, but data across multiple years and categories being compared to data from other neighbourhoods. I was not, however, able to chase the stories with what I had judged to be the most compelling data angle—highest rent, greatest rent increase, lowest vacancy, etc.—because of issues with scheduling time with sources that were caused by timing of reporting close to Quebec's traditional moving day on July 1 and a hesitancy from the people I wanted to talk to over the fact that the project was not attached to a publication. These are setbacks that may not have come up if this were a project produced in a traditional newsroom, but did affect the learning-by-doing approach taken in this research creation project.

There were benefits to using the learning-by-doing model which were present in each of the three categories I reflected on during the project's creation. When acquiring and analyzing the data I would be using I was able to apply the tools of numbers and mathematics in my toolkit (Meyer, 1991) to understand how the data was collected and presented to show the story of how housing was getting more expensive and less accessible because of rising rents and lowering vacancy in Montreal. While learning coding skills and generating the website, I was able to more accurately understand which overlapping skills of programming and interactivity I should focus

in on to produce my specific project and how the collaborative spirit that exists in data journalism, outlined by Usher (2016), can be used to one's benefit when creating a site such as this. And, finally, while applying traditional reporting techniques to these stories driven by data, I was able to use the information from sources who understood these housing issues to provide context to the numbers I was using in my reporting, as suggested should be the case by Vallance-Jones and McKie (2017).

Assembling this project and the learning-by-doing model did make me more competent in data journalism practices and is a way that data journalism education could be applied in university and mid-career training curricula. This is not to say necessarily that students should just be left to their own devices, but rather that if students were given more opportunities to explore and acquire data skills through the process of creating a specific project with set goals. This may be a more effective way of teaching and learning data journalism skills than just teaching basic skills like spreadsheets and coding without application to a specific story or project, as Heravi (2018) suggests is often the case. If there is something the creation of this website did underline for me, it is the assertion that data journalism is a collection of a broad range of skills that never lead to the same skill set for each journalist. This goes back to the idea from Usher (2016) that there is no one definition for what makes a data journalist, and the practice of data journalism in the field is often accomplished by an overlapping of skills in data, interactivity and programming from a team of people.

While it was possible for me to create this project individually and there were programs and people I relied on to support me as I was acquiring the skills I needed to create this website, it became overwhelmingly clear as the process went on that data journalism skills should not be learned in a vacuum. The section on website generation covered the importance of collaboration

in the coding community (Usher, 2016) and this spirit needs to be implemented in whatever culture of data (and hacker) journalists is going to appear over the course of the 21st century. Not only that, but the need for a certain level of mentorship in data journalism and education is clear. As Lewis (2020) notes, there is a generational gap in the passing down of knowledge from the current cohort of data journalists, which means that the upcoming cohort is not connecting with data journalists working in newsrooms in a way that will develop their own skills and allow them to innovate on current data journalism practices. Again, the individual acquisition of data skills and knowledge for the purpose of this project was moderately successful, but having the opportunity to learn these skills, for example, in a newsroom that is already innovating data journalism practice would have ameliorated the process and indeed given more space to explore ideas about the best way to present the information in the story.

This research indicates that a more widespread application of this learning-by-doing model in data journalism training and education has the potential to allow students to more deeply develop specific skill sets that would lend themselves more effectively to the expansive and collaborative practice of data journalism in the industry. Acquiring data journalism skills in a learning-by-doing model did not lead to the initially desired data-driven stories, however there was still an opportunity to develop stories embellished with prescient pieces of data and the goals set out for the project provided direction for acquiring the programming and design skills which would create a map site which communicates the story of the housing situation in Montreal.

References

Anderson, C.W. (2018). *Apostles of certainty: Data journalism and the politics of doubt.* Oxford.

Andrew, C., Klodawsky, F., & Siltanen, J. (eds., 2017). *Toward equity and inclusion in Canadian cities: Lessons from critical praxis-oriented research*. McGill-Queen's University Press.

Best, J. (2012). *Damned lies and statistics: untangling numbers from the media, politicians, and activists*. University of California Press.

Bounegru, L., Chambers, L., & Gray, J. (eds., 2012). *The data journalism handbook 1*. Retrieved October 1, 2020, from: https://datajournalism.com/read/handbook/one.

Cairo, A. (2013). *The functional art: An introduction to information graphics and visualization*. New Riders.

The Canadian Association of Professional Journalists. (2011). *Ethics guidelines*. Retrieved from: https://caj.ca/ethics-guidelines.

The Canadian Mortgage and Housing Corporation. (n.d.). *Rental market survey reliability tables – centre level data.* Retrieved from: https://www.cmhc-schl.gc.ca/en/professionals/housing-markets-data-and-research/housing-data/data-tables/rental-market/centre-level-data-tables.

CBC News. (2018, Nov. 25). The implant files: Inside Health Canada's medical device records. *Canadian Broadcasting Corporation.* Retrieved from: https://newsinteractives.cbc.ca/medical-devices-11-2018/.

Chapman, O., & Sawchuk, K. (2012). "Research-creation: Intervention, analysis and 'family resemblances.'" *Canadian Journal of Communication, 37*(2012), pp. 5-26.

Chung, E., et al. (2020, March 20). Tracking the coronavirus. *Canadian Broadcasting Corporation*. Retrieved from: https://newsinteractives.cbc.ca/coronavirustracker/.

Codecademy. (n.d.). Retrieved from: https://www.codecademy.com/.

Cohn, N., & Smart, C. (2021, January 8). Why Warnock and Ossoff won in Georgia. *The Upshot, The New York Times*. Retrieved from: https://www.nytimes.com/2021/01/07/upshot/warnock-ossoff-georgia-victories.html.

Cori-Manocchio, V. (2020, June 14). Housing crisis amid a pandemic: Parc-Ex residents call for pause on evictions. *Canadian Broadcasting Corporation*. Retrieved from: https://www.cbc.ca/news/canada/montreal/parc-extension-evictions-housing-covid-19-1.5611460.

D3.js – Data-driven documents. (n.d.). Retrieved from: https://d3js.org/.

Find my landlord MTL. (n.d.) Retrieved December 14, 2020 from: https://findmylandlord.netlify.app/?lng=en&fbclid=IwAR1XY1zcDEd0oNg7Ojm_VzRnqLcmdi5Ha7AbfT8MkndMvutYnlr75IHq0BM.

Foust, J. C., & Bradshaw, K.A. (2020). Pushing boundaries: how coding is (and isn't) taught in accredited journalism programs. *Journalism and Mass Communication Educator*, 75(3), 335–347.

Gagnon, K. (2021, April 24). Loyers en folie. *La Presse*. Retrieved from: https://www.lapresse.ca/actualites/enquetes/2021-04-24/crise-du-logement/loyers-en-folie.php.

Ganesh H. (2020, June 20). *Visualize Data with a Choropleth Map - freeCodeCamp Data Visualization Project Tutorial* [Video]. YouTube. https://www.youtube.com/watch?v=ha1toFtBfF8.

Gélinas, O., Léouzon, R., & Pavic, C. (2021, April 8). Les loyers grimpent et s'affolent, à

Montréal comme ailleurs. *Le Devoir*. Retrieved from: https://www.ledevoir.com/documents/special/04-21-loyers-quebec/index.html.

Gentrification in Toronto. (n.d.). Retrieved from: https://www.arcgis.com/apps/Cascade/index.html?appid=781c6bd5af8248238f2adda89b244533.

Hénault, O. (2021, April 1). Goodbye tristesse! Habitations Saint-Michel Nord, Montreal, Quebec. *Canadian Architect.* Retrieved from: https://www.canadianarchitect.com/goodbye-tristesse-habitations-saint-michel-nord-montreal-quebec/.

Heravi, B. R. (2018). 3ws of data journalism education: what, where and who? *Journalism Practice*, pp. 349–366.

Heravi, B. (2019, April 5). *Where in the world can I study data journalism?* Retrieved from: https://medium.com/ucd-ischool/where-in-the-world-can-i-study-data-journalism-44c006e55ea5.

Kasakove, S. (2019, April 27). The fight to save Brooklyn from gentrification. *The Week.* Retrieved from: https://theweek.com/articles/834372/fight-save-brooklyn-from-gentrification.

Karwacki, J. (2021, March 24). Montreal's landlords are bumping up prices to whittle down applicants even as vacancy rate increases. *Montreal Gazette.* Retrieved from: https://montrealgazette.com/news/local-news/montreals-landlords-are-bumping-up-prices-to-whittle-down-applicants-even-as-vacancy-rate-increases.

Lauzon, V. (2020, June 28). Crise du logement: «Le coût du loyer flambe». *La Presse.* Retrieved

from: https://www.lapresse.ca/actualites/2020-06-28/crise-du-logement-le-cout-du-loyer-flambe.php.

Lee-Young, J. (2020, December 5). Vancouver real estate: Shift on to big corporate landlords of apartment buildings. *The Vancouver Sun*. Retrieved from: https://vancouversun.com/business/real-estate/vancouver-real-estate-shift-on-to-big-corporate-landlords-of-apartment-buildings.

Lewis, N. P. (2020). Defining and teaching data journalism: a typology. *Journalism & Mass Communication Educator*, (20200629).

Limite administrative de l'agglomération de Montréal (Arrondissements et Villes liées) - Jeu de données. (n.d.). *La Ville de Montréal*. Retrieved from: https://donnees.montreal.ca/ville-de-montreal/polygones-arrondissements.

Lindgren, M. (2014). "Radio journalism as research — a P.h.D. model." *The Radio Journal International Studies in Broadcast & Audio Media, 12*(½), pp. 169-182.

Mapbox. (n.d.). Retrieved from: https://www.mapbox.com/.

Meyer, P. (1991). *The new precision journalism*. Bloomington, IN: Indiana University Press.

Meyer, P. (2002). *Precision journalism: A reporter's introduction to social science methods, 4th ed*. Rowman & Littlefield.

McCartney, R. (2019, April 8). Amazon in Seattle: Economic godsend or self-centered behemoth? *The Washington Post*. Retrieved from: https://www.washingtonpost.com/local/trafficandcommuting/amazon-in-seattle-economic-godsend-or-self-centered-behemoth/2019/04/08/7d29999a-4ce3-11e9-93d0-64dbcf38ba41_story.html.

McKenna, K. (2019, May 3). Latest victim of gentrification in Saint-Henri? An

anti-gentrification group. *Canadian Broadcasting Corporation*. Retrieved from: https://www.cbc.ca/news/canada/montreal/popir-rent-saint-henri-gentrification-1.5121624.

Nielsen, C. (2010). Conceptualizing diversity in applied learning. *Western Washington University Center for Instructional Innovation and Assessment.* Retrieved from: https://cii.wwu.edu/showcase2010/nielsen/.

Ouellet, V., & Shiab, N. (2019, Sept. 4). Set up to fail: Why women still don't win elections as often as men in Canada. *Canadian Broadcasting Corporation*. Retrieved from: https://ici.radio-canada.ca/info/2019/elections-federales/femmes-hommes-probabilites-vote-egalite-chateaux-forts/index-en.html.

OpenRefine. (n.d.). Retrieved from: https://openrefine.org/.

Ribecca, S. (n.d.). Choropleth map. *The data visualization catalogue.* Retrieved from: https://datavizcatalogue.com/methods/choropleth.html.

Registre des loyers citoyen. (n.d.). Retrieved from: https://registredesloyers.quebec/.

Le Regroupement des comités logement et associations de locataires du Québec. (2020, June). La flambée des loyers: Enquête sur le prix des logements à louer au Québec. *Le Regroupement des comités logement et associations de locataires du Québec.* Retrieved from: https://rclalq.qc.ca/wp-content/uploads/2020/06/La-flamb%C3%A9e-des-loyersVF.pdf.

Le Regroupement des comités logement et associations de locataires du Québec. (2021, June). Les loyers explosent: Enquête sur le prix des logements à louer au Québec. *Le Regroupement des comités logement et associations de locataires du Québec.* Retrieved from: https://rclalq.qc.ca/wp-content/uploads/2021/06/Les-loyers-explosent.pdf.

Shiab, N. (2019, April 30). How Airbnb's biggest players invaded Montreal's hottest neighbourhoods. *Canadian Broadcasting Corporation*. Retrieved from: https://ici.radio-canada.ca/info/2019/04/airbnb-annonces-location-logements-plateforme-montreal-canada/index-en.html.

The Society of Professional Journalists (2014). *Code of ethics*. Retrieved from: https://www.spj.org/ethicscode.asp.

Stack Overflow. (n.d.). Retrieved from: https://stackoverflow.com/.

Tabary, C., Provost, A., & Trottier, A. (2016). "Data journalism's actors' practices and skills: A case study from Quebec." *Journalism, 17*(1), pp. 66-84.

Thériault, J. (2021, July 30). Presque à la rue, un Montréalais raconte comment il vit la crise du logement. *Radio-Canada.* Retrieved from: https://ici.radio-canada.ca/nouvelle/1812842/eviction-itinerance-sauvetage-logement-office-habitation.

Tufte, E. (1997). *Visual Explanations: images and quantities, evidence and narrative*. Graphics Press.

Usher, N. (2016). *Interactive journalism: Hackers, data, and code.* Urbana, IL: University of Illinois Press.

Vallance-Jones, F., & McKie, D. (2017). *The data journalist: Getting the story*. Oxford.

La Ville de Montréal. (2020, November). *Limite administrative de l'agglomération de Montréal (Arrondissements et Villes liées).* Retrieved from: https://donnees.montreal.ca/ville-de-montreal/polygones-arrondissements.

Walks, R. A., & Maaranen, R. (2008). *The timing, patterning, & forms of gentrification & neighbourhood upgrading in Montreal, Toronto, & Vancouver, 1961 to 2001*. Centre for Urban and Community Studies, Cities Centre, University of Toronto.

White, M. (2020, January 14). 'My Parkdale is gone': how gentrification reached the one place that seemed immune. *The Guardian*. Retrieved from: https://www.theguardian.com/cities/2020/jan/14/my-parkdale-is-gone-how-gentrification-reached-the-one-place-that-seemed-immune.

Wong, D. (2013). *The Wall Street Journal guide to information graphics.* W. W. Norton.

Appendices

Appendix A: Data journalism project resource guide

While creating this project, I focused on how to combine traditional and data-driven reporting techniques to tell housing stories and how to use data journalism to tell localized stories in the city of Montreal. What follows below is a resource list of the data materials, learning tools and integral texts—broken down by the three major categories of data, visualization and story—used to build this specific project, accompanied by short explanations of what role they played in the creation of the map site and accompanying articles.

Data

- The Canadian Mortgage and Housing Corporation Rental Market Survey Reliability tables for the centre of Montreal for 2015 to 2019 were the main sets of data analyzed and later used to generate the map and story ideas for the site.
- *Precision Journalism* by Philip Meyer (2002) to understand the historical practices of data analysis in journalism and how they relate to social science methods.
- *The Data Journalist* by Fred Vallance-Jones and David McKie (2017) to understand modern applications of data analysis in journalism.
- *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists* by Joel Best (2013) to understand how and why certain bodies collect data and present it to the public.
- An interview with an economist from the CMHC to understand how the data for the set was collected and compiled.

Visualization

- Complete introductory courses on Codecademy for HTML, CSS and JavaScript to understand the foundations of website building.
- Draw inspiration from D3.js and study tutorial on how to build choropleth maps.
- Download GeoJSON of Montreal neighbourhoods from the City of Montreal to further edit in Mapbox Studio to fit the specifications of the boundaries defined by the CMHC in their data collection.
- Mapbox Studio used to also generate the styles for different maps that represent different categories of the data being used to generate the site.
- Microsoft Visual Studio Code application to begin locally coding the project on a desktop without publishing immediately.
- Mapbox GL JS guides for how to embed features created on Mapbox onto the site and Stack Overflow forums to consult community responses to problems which occurred while programming the site.
- *The Functional Art: An Introduction to Information Graphics and Visualization* by Alberto Cairo (2013) to understand how to design information graphics that are seamlessly comprehensible for the way the human brain processes visual information.
- *The Wall Street Journal Guide to Information Graphics* by Dona M. Wong (2013) to understand recent standards and practices for infographic generation in the journalism industry.

Story

- At this point it will be important to revisit Meyer (2002) and Vallance Jones & McKie (2017) for their sections on incorporating data journalism into traditional reporting.

- *The Data Journalism Handbook* edited by Liliana Bounegru, Lucy Chambers and Jonathan Gray (2012) to further understand recent approaches to finding human sources to tell data-driven stories
- The Canadian Association of Journalists' Ethical Guidelines and The Society of Professional Journalists' Code of Ethics for guidance on ethical dilemmas which may crop up during reporting.