

# THE DIGITAL DIVIDE AND THE GOVERNMENT: DEVELOPING A TOOL FOR THE ANALYSIS OF GOVERNMENT DATA

by

Emily Millard

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF

BACHELOR OF SCIENCE HONOURS

in

The Irving K. Barber School of Arts and Sciences

(Honours Computer Science Major Computer Science Minor Data Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

April 2016

© Emily Millard, 2016

## **Abstract**

The digital divide is defined as the gap between groups, demographics, or regions that have access to information communication technologies (ICTs) and those who do not, or who have restricted access [1]. The digital divide impacts smaller cities within British Columbia in terms of their ability to represent local infrastructure, provincial, and federal datasets spatially. As a result these municipalities are unable to use spatial data to its fullest extent in their decision-making processes. In partnership with LandInfo Technologies an interactive, online analysis tool that allows local government officials to make use of municipal, provincial, and federal spatial data while engaging in decision-making process has been developed. The goal of this tool is to enable local government to store, maintain, and view infrastructure data, thus supporting more informed decisions and as a result reducing the overall cost of developing and maintaining municipal infrastructure.

# Table of Contents

<b>Abstract</b> .....	<b>2</b>
<b>Table of Figures</b> .....	<b>5</b>
<b>List of Tables</b> .....	<b>5</b>
<b>Acknowledgements</b> .....	<b>6</b>
<b>1 Introduction</b> .....	<b>7</b>
1.1 <i>Motivation</i> .....	7
1.2 <i>Goals and Objectives</i> .....	9
<b>2 Background</b> .....	<b>10</b>
2.1 <i>Terms</i> .....	10
2.2 <i>Study Region</i> .....	11
2.3 <i>Data Sources &amp; Limitations</i> .....	12
2.4 <i>Interface Design</i> .....	14
2.4.1 <i>User Goals</i> .....	14
2.4.2 <i>Prototype and Design Rational</i> .....	14
<b>3 BCAtlas Analytics – User Guide</b> .....	<b>17</b>
3.1 <i>Website Landing Page</i> .....	17
3.2 <i>Parcel Manipulation and Data Summary</i> .....	18
3.3 <i>Legend and Layer Manipulation</i> .....	22
3.4 <i>Data Analysis</i> .....	23
3.5 <i>Accessing BCAtlas Analytics’ Web Services</i> .....	24
3.5.1 <i>GET Request URLs</i> .....	24
3.5.2 <i>WMS Requests</i> .....	26
<b>4 Technical Details &amp; Implementation</b> .....	<b>27</b>
4.1 <i>Development Environment</i> .....	27
4.2 <i>Server Configuration</i> .....	27
4.2.1 <i>Database Configuration</i> .....	27

4.3	<i>Creating, Loading, and Querying the Map</i> .....	28
4.3.1	Geoserver .....	28
4.3.2	Loading Map Data .....	29
4.3.3	Styling Map Layers .....	30
4.3.4	Handling Map Interactions .....	33
4.4	<i>Displaying Information in the Attribute Panel</i> .....	34
4.4.1	Properties List Tab .....	34
4.4.2	Finding a Parcel's Census Subdivision Id.....	34
4.4.3	Population Summary Tab .....	35
4.4.4	Census and NHS Demographics Tabs.....	35
<b>5</b>	<b>Discussion and Conclusion</b> .....	<b>36</b>
5.1	<i>Discussion</i> .....	36
5.2	<i>Conclusion</i> .....	37
<b>6</b>	<b>References</b> .....	<b>38</b>

## Table of Figures

Figure 1: British Columbia NAD 83 Zones [28] .....	13
Figure 2: Scanned image of paper prototype .....	15
Figure 3: Screenshot of final interface design .....	17
Figure 4: Screenshot of initial map layout .....	18
Figure 5: Screenshot of map tools .....	18
Figure 6: Screenshot of polygon selection tool .....	19
Figure 7: Screenshot of properties list.....	20
Figure 8: Screenshot of Population Summary tab.....	21
Figure 9: Screenshot of Census Demographics tab.....	22
Figure 10: Screenshot of the legend .....	23
Figure 11: Screenshot of value per sq ft analysis layer.....	24
Figure 12: Database diagram for the BCAtlas Analytics PostgreSQL database .....	28
Figure 13: SQL statement used to create the Value per sq ft view .....	29
Figure 14: SQL statement used to create the Building value as a % of total value layer.....	29
Figure 15: Plot of Value per sq ft values .....	30
Figure 16: Plot of Building value as a % of total value values.....	31
Figure 17: Selected property added to the Highlight Overlay layer .....	33
Figure 18: Spatial query (with PreparedStatement values supplied) to find all intersections of a user drawn polygon and parcels .....	34

## List of Tables

Table 1: Layers for WMS requests.....	26
Table 2: Summary of layer data values obtained using the R summary function.....	31
Table 3: Value per sq ft classification values .....	32
Table 4: Building value as a % total value classification values.....	32

## **Acknowledgements**

I would like to thank Dr. Ramon Lawrence for his encouragement and support while completing my honours project and more generally throughout my undergraduate degree at the University of British Columbia Okanagan. Without Dr. Lawrence I would not have had the opportunity to explore my most prominent area of interest, the applications of GIS in municipal government. I would also like to thank Tjaart Van den Berg of LandInfo Technologies for putting forth this idea for a project and providing me with access to datasets from the Village of Lumby and BC Assessment.

# 1 Introduction

## 1.1 Motivation

A Geographic Information System (GIS) is computer software that is designed for the manipulation, analysis, and display of spatial data. GIS technology can “compare the locations of different things to discover how they relate to one another” [2]. GIS can also be used to combine spatial and non-spatial data. For example a GIS can be used to compare and contrast information about population, income level, locations of schools, and property value [2].

One of the main functions of municipal governments is to develop and maintain local infrastructure [3]. GIS technologies provide governments with a diverse set of tools to analyze geographic data along side corresponding aspatial data. Municipalities throughout British Columbia vary greatly in size and in their resources available to spend on GIS. Generally speaking, municipalities currently have two ways to display spatial data on the web. The first uses the industry standard ArcGIS suite produced by Esri. Alternatively municipalities can choose to use a smaller scale product such as MangoMap, GISCloud, MapBox, or CartoDB to name a few [4]. Most commonly, municipalities use one of Esri’s full service mapping products. However Esri’s products are very expensive, making them inaccessible to many potential users [5].

The various alternatives to the ArcGIS suite each offer different pros and cons and require various levels of programming knowledge in order to establish and maintain a GIS system [4]. One distinguishing feature for all of the solutions mentioned is that they require the municipality have at least one in-house GIS expert to build, manage, and maintain the implemented GIS solution as well as manage all of the municipal geospatial data. LandInfo Technologies, a geospatial management consulting company, aims to solve this problem. LandInfo Technologies builds, manages, and hosts web-mapping applications for small to medium municipalities in addition to updating and maintaining their geospatial data. However, the current product offered by LandInfo

Technologies does not make use of the open data provided by provincial and federal governments.

Through LandInfo Technologies' research it has been determined that small municipalities in British Columbia are most interested in the Canadian Census and National Household Survey data. This data is made available to users through the Canadian government's open data portal. Through access to these data sets municipalities are able incorporate population data into preexisting spatial datasets thus allowing them to make use of open data in their decision-making process. For example, if municipalities know the approximate population of a given subdivision block, city planners can estimate areas that may need new sanitary pipes installed to accommodate a larger volume of water flow resulting from an increase in population. This detailed population information can be provided using the Canadian Census dissemination area populations. Although all Canadian citizens have access to the data obtained through the Canadian Census, this data is not provided in a spatial format. This is a significant shortcoming as otherwise finding the population of a given area can be very time consuming.

There are two factors that impact the severity of the digital divide for an individual or municipality: access and skill. The access divide can be defined as lack of access to Information Communication Technologies (ICTs) through a computer, the Internet, or a software program. The skill divide is quantified by the level of "technical competence and information literacy" of a given user [6]. Small municipalities do not have access to software used by larger municipalities, like ArcGIS, that enables them to use spatial data to aid in the decision making process. Additionally, even if these smaller municipalities did have equal access to this software, there is a high probability that employees do not have the necessary skill set to take advantage of such access. It is clear that barriers of access and skill create a digital divide between smaller and larger municipalities in British Columbia. These barriers make it difficult, if not impossible for small municipalities to use spatial data when making important decisions. The resulting

lack of information can reduce the accuracy and increase the overall cost of the decision making process for smaller municipalities.

## **1.2 Goals and Objectives**

There are two goals for this project. First, this project will create an interactive web application that will allow small municipalities to make use of data from the 2011 Canadian Census, 2011 National Household Survey, and 2015 BC Assessment of parcels when making decisions. Interactive map layers, charts, and tables, will be used to display data in a simplistic format that will allow users with limited GIS experience to understand and interpret the information presented. Second, this project will make the data and layers used while creating the interactive web application available to 3<sup>rd</sup> party users via GET requests and WMS layers. This will allow the data to be repurposed and used when creating other projects and tools. The name of the software developed for this project is BCAtlas Analytics.

## 2 Background

### 2.1 Terms

**API** – Application Programming Interface, a set of building blocks containing routines, protocols, and tools that can be used to build software and web applications [7]

**Census Subdivision** – “Area that is a municipality or an area that is deemed to be equivalent to a municipality for statistical reporting purposes” [8]

**Choropleth map** – thematic map where areas are shaded in proportion to a measured value [9]

**CRS** – Coordinate Reference System, defines how georeferenced spatial data relates to real locations on the Earth’s surface [10]

**Dissemination Area (DA)** – “Small area composed of one or more neighbouring dissemination blocks, with a population of 400 to 700 persons.” [11]

**GeoJSON** – standard format used to encode geographic objects and their non-spatial attributes, based on JSON

**Geoserver** – open source server used for sharing and hosting geospatial data

**GET Request** – Hypertext transfer protocol that requests data from a specified resource over the Internet [12]

**GIS** – Geographical information system, designed for the manipulation, analysis, and display of spatial data

**HCI** – Human computer interaction focuses on the design, evaluation, and implementation of computing systems with a focus on usability. [13]

**IntelliJ** – Java integrated development environment used by this project to write source code and compile the project into a WAR file for deployment.

**Java** – Widely used programming language with emphasis on concurrent, class based, and object oriented design [14]

**JavaScript** – client side programming language used to create web applications

**JSON** – JavaScript object notation is a syntax used for storing and exchanging data in a human readable format. [15]

**Open data** – data that is machine readable and available to be used, built upon, and manipulated without restriction [16]

**OpenLayers** – “open source JavaScript library to load, display, and render maps from multiple sources on webpages” [17]

**Orthophoto** – geometrically corrected aerial photo

**Parcel** – single unit of land

**PostGIS** – spatial database extender for PostgreSQL adding support for the storage and querying of spatial objects [18]

**PostgreSQL** – open source relational database management

**RStudio** – development environment for R, a statistical programming language [19]

**Servlet** – Java program that runs on a webserver [20]

**Shapefile** – geospatial vector file format used to store points, lines, and polygons along with their associated attributes

**SLD** – Styled layer descriptor, XML style encoding that extends the WMS standards to allow user defined symbolization of geographic features [21]

**URL** – Uniform resource locator, reference to a location on the Internet

**Vector** – “coordinate-based data model that represents geographic features as points, lines, and polygons” [22]

**WAR file** – Web application archive file used to distribute a collection of Java Servlets, Java Classes, XML files, static webpages, and related files [23]

**WMS** – Web map service, Hyper Text Transfer Protocol used to request geospatial map images from a GIS database [24].

## **2.2 Study Region**

The data used and analysis performed in this project focus exclusively on the Village of Lumby. Lumby is a small community located in southwestern British Columbia with a population of approximately 1,700 people [25]. Although Lumby is small it maintains autonomy from larger neighboring cities through its own Mayor and City Council. Main sources of industry for Lumby include tourism, manufacturing, and logging [25].

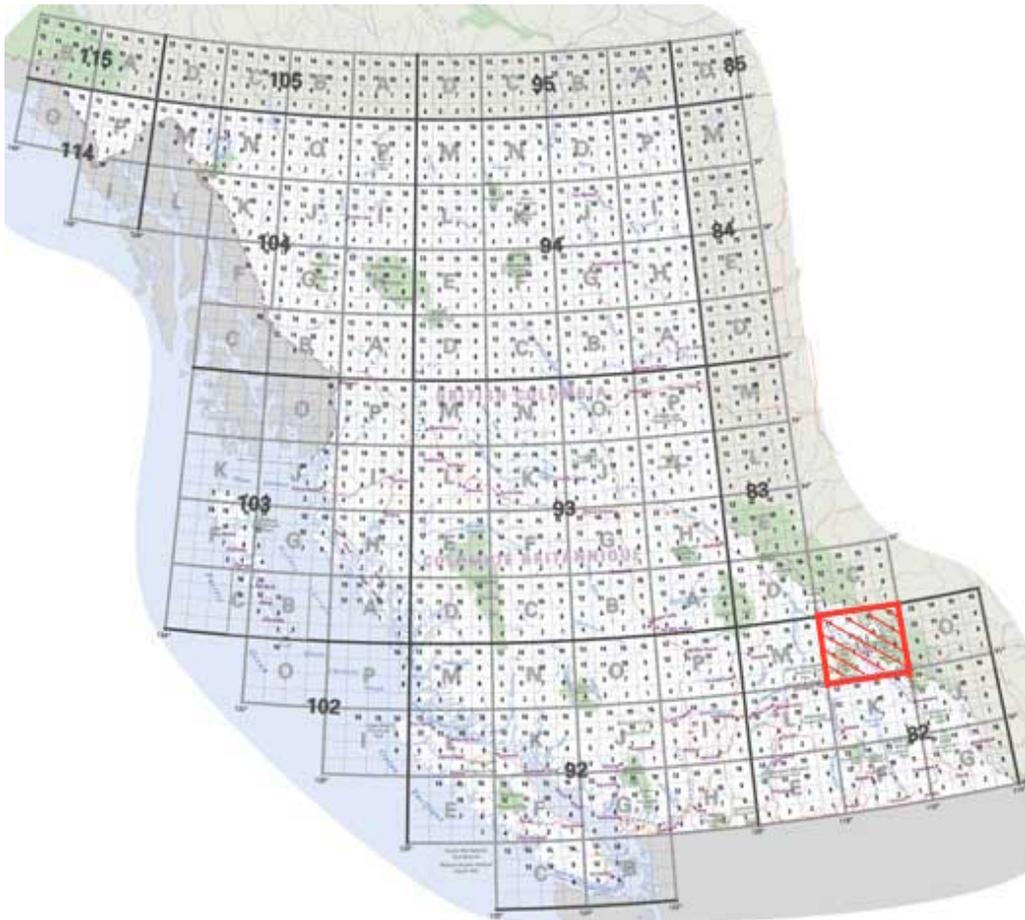
## 2.3 Data Sources & Limitations

The data used in this project was obtained from four sources: the 2011 Canadian Census, the 2011 National Household Survey (NHS), the Village of Lumby, and BC Assessment.

Data from the 2011 Census and NHS are available through the Government of Canada's open data portal. This project was interested specifically in the 2011 Census and NHS for Lumby at the DA level. However the response rates by Lumby residents to both the Census and NHS were flagged by Statistics Canada as being below the average response rate for the province of British Columbia. A response rate is defined by Statistics Canada to be "the number of private and collective dwellings that returned a completed questionnaire divided by the number of private and collective dwellings classified as occupied on the Census database" [26]. Although Canadian citizens are required by law to complete the Canadian Census between 5-10% of all residents of Lumby did not complete the 2011 Census questionnaire [27]. The average response rate for all residents of British Columbia was 96.5% [26]. In 2011 Canadians were asked to complete voluntary NHS in addition to the Census. The NHS response rate for Lumby residents was 65.2% whereas the provincial average was 77.1% [28]. Due to Lumby's small population size, low response rates, and issues with data confidentiality from both the 2011 Census and NHS, data is not available per Dissemination Area (DA) only per Census Subdivision, thus limiting the resolution of the data used in this project.

The Village of Lumby provided the locations of parcels within Lumby in Shapefile format. In addition to parcel locations the Shapefile included attributes describing the parcel's street address, total area, and GISIdp that acts as a unique parcel identifier for all parcels within British Columbia. The coordinate reference system associated with the geographic coordinates of the parcels in the Shapefile is NAD 83 UTM Zone 11N/EPSSG:26911. This projection is appropriate for large and medium scale topographic mapping for regions located within the highlighted area in Figure 1 [29]. While this data is projected in the correct CRS, there are significant issues with data

quality. The subdivision survey diagrams used to capture many of the parcel boundaries were completed before current GPS and LiDAR technologies were available for use while surveying properties. There is currently a province wide initiative in place by BC Land Title and Survey to create one accurate parcel mapping for the entire province but the process is on going and will take approximately two and a half years to complete [30]. Once a more accurate parcel Shapefile is available, BCAtlas Analytics will be updated to reflect the change.



**Figure 1: British Columbia NAD 83 Zones [31]**

Parcel assessment values were obtained from the 2015 assessment of parcels within the Village of Lumby performed by BC Assessment. This dataset included unique parcel identifiers (jur\_roll) that could be matched to the GISkp in the parcels Shapefile to retrieve the assessed values for land and building(s) within each parcel.

Additional geographic data was used when displaying land parcels to give users a frame of reference for parcel locations. These data sets included parcel labels, orthophotos, roads, road labels, and a municipal boundary. All of these data sets were in the NAD 83 UTM Zone 11N/EPSSG:26911 CRS and obtained via WMS from LandInfo Technologies.

## **2.4 Interface Design**

To ensure the usability of BCAtlas Analytics the user's goals while using the website, interface layouts deployed by similar mapping applications, and several HCI principles were considered throughout the development process.

### **2.4.1 User Goals**

The following were established as a user's top priorities when visiting BCAtlas Analytics:

1. Identify and view information about a single parcel
2. Identify and view information about multiple parcels
3. View charts and graphs regarding the demographics of an area surrounding a given parcel
4. Manipulate and view layers using a legend
5. Be able to customize the interface to maximize the area used to display the map

### **2.4.2 Prototype and Design Rational**

Web mapping applications produced by Google Maps, the City of Kelowna, Geocortex, and LandInfo Technologies were tested and compared for usability. After performing these comparisons it was determined that many of these applications were plagued by the problem that the area available to display the given map was very small. Panels for map tools, a legend, and attribute display took up the majority of the available screen space. To compensate for this problem the design of the BCAtlas Analytics web application opted to only use a panel for attribute display that defaulted to take up 25% of the screen. Additionally, the panel can be collapsed allowing the map to occupy the entire width of the screen. To reduce the amount of space required by the legend a small legend tile was added to the top left side of the map. When the user clicks the legend tile it expands to show layers that can be manipulated. Once the user clicks off

the legend it collapses to its original size. After establishing the user's main goals when visiting BCAtlas Analytics and performing a design rationale the paper prototype shown in Figure 2 was established.

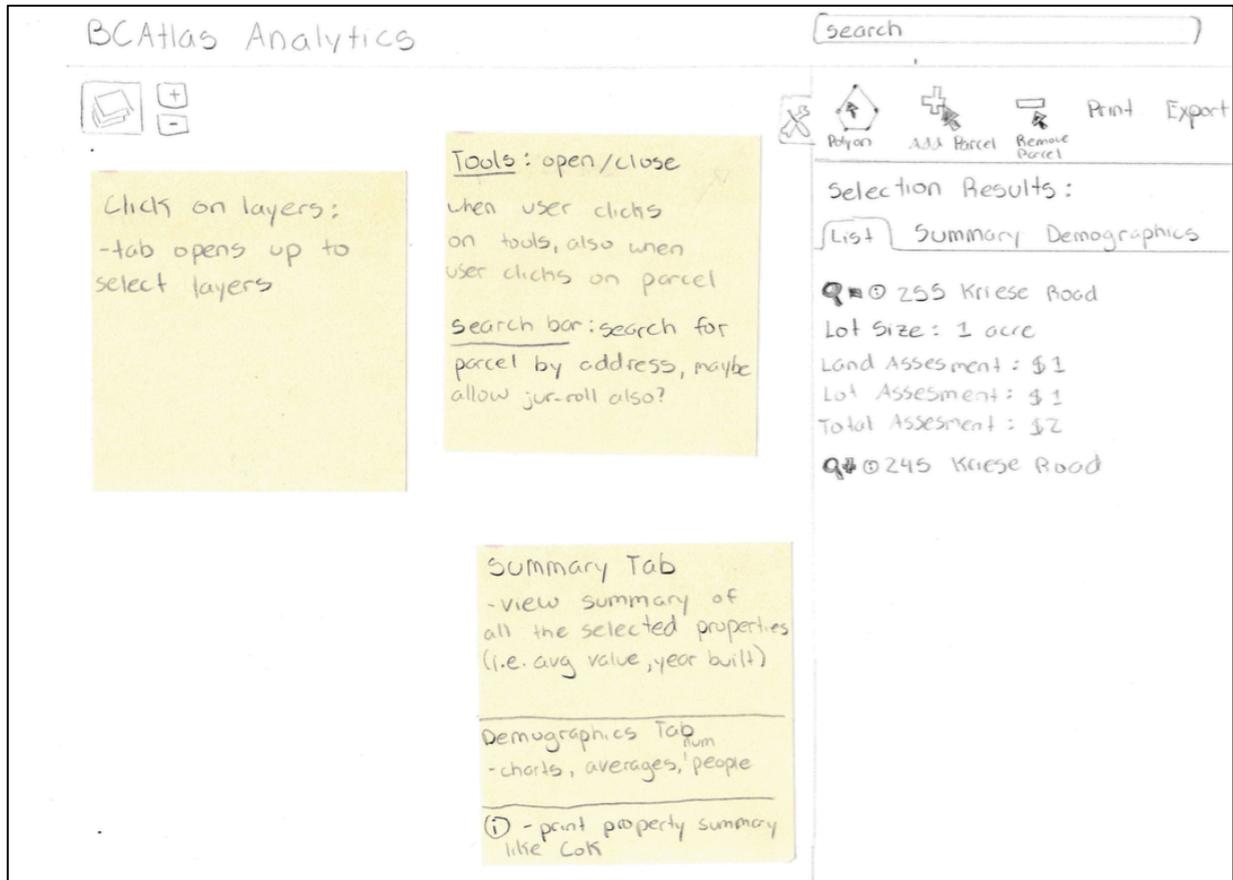


Figure 2: Scanned image of paper prototype

Throughout the design and development process Shneiderman's Eight Golden Rules of Interface Design were considered. These rules are a collection of principles derived heuristically from experience and are used by designers with the goal of creating a well designed and useable user interface [32]. During the design process, principles one, five, and eight were used to either justify or modify the components outlined in Figure 2. Shneiderman's first principle states designers should strive for consistency. A consistent sequence of actions should be required in similar situations and identical terminology should be used [32]. The consistent use of terms, icons and colors throughout the interface support this principle (i.e. the term "property" is always used to refer to a parcel, property and parcel are not used interchangeably). Additionally, the use of icons

that are commonly used by other mapping interfaces to represent various components such as the legend and map tools were used to further support Shneiderman's first principle. Shneiderman's fifth design principle requires designers and developers to offer simple error handling and as much as possible to design the system so the user cannot make a serious error [32]. One significant problem encountered while designing user interactions for the web application was not all parcels on the map were associated with BC Assessment data. As a result when users clicked on a parcel that was not related to any BC Assessment data the application would crash. To prevent this from occurring proper error handling was put in place to account for this problem and a message is displayed to the user if this error occurs. Additional exceptions related to database connection issues and GIS errors are also handled. Finally, Shneiderman's eighth design principle states that user interfaces should reduce short-term memory load by keeping displays simplistic, consolidating pages, and reducing window-motion frequency [32]. This principle was very difficult to accomplish when designing the user interface for BCAtlas Analytics because one of the main design goals was to maximize the space available to display the map. As a result of this goal some features had to be placed inside collapsible containers (i.e. legend, map tools, etc). This design choice requires users to recognize and recall which icons represent which features. However, since the icons used are common to many mapping applications, a smaller load is placed on the user's short-term memory. After engaging in iterative design and development processes while considering Shneiderman's Eight Golden Rules a final interface was created and can be seen in Figure 3.

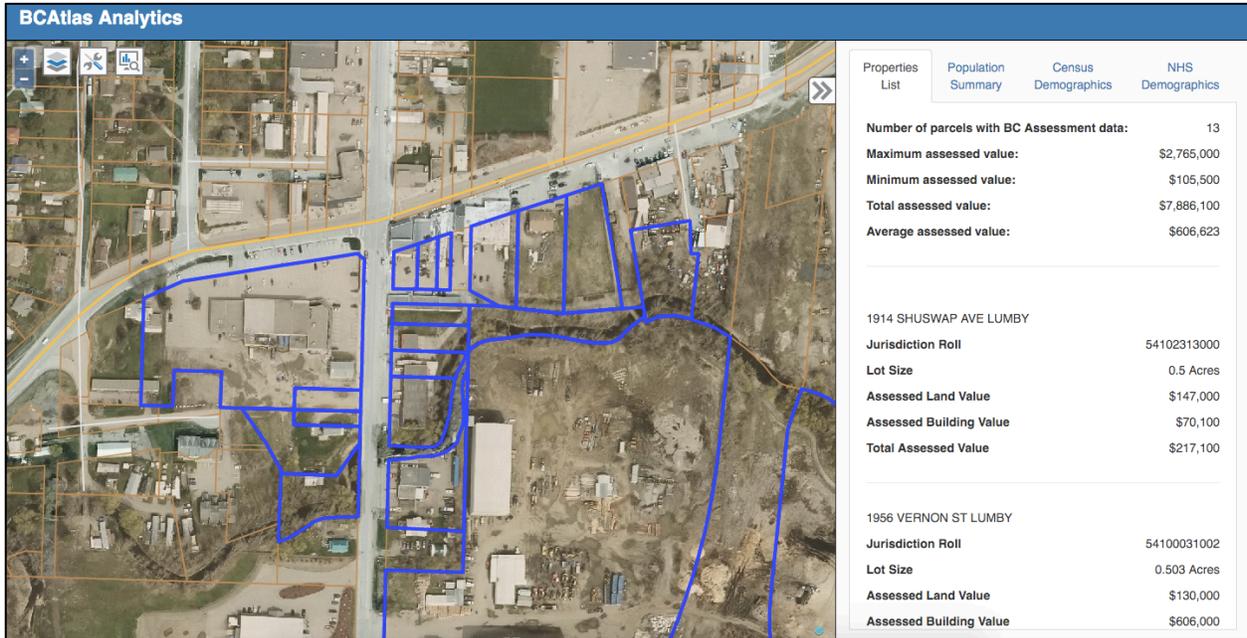


Figure 3: Screenshot of final interface design

### 3 BCAtlas Analytics – User Guide

BCAtlas Analytics can be found at <http://159.203.2.8:8080/>. Data from BCAtlas Analytics can be accessed in two ways. Users can view the data presented at the above URL or they can access the data using a combination of web services. For users only interested in accessing BCAtlas Analytics using web services refer to Section 3.5 Accessing BCAtlas Analytics’ Web Services.

#### 3.1 Website Landing Page

When website users first visit BCAtlas Analytics they will be presented with a basic map of the Village of Lumby with the layers Parcels, Roads, and Municipal Boundary enabled. As seen in Figure 4 no parcels are selected and no analysis tools are enabled.

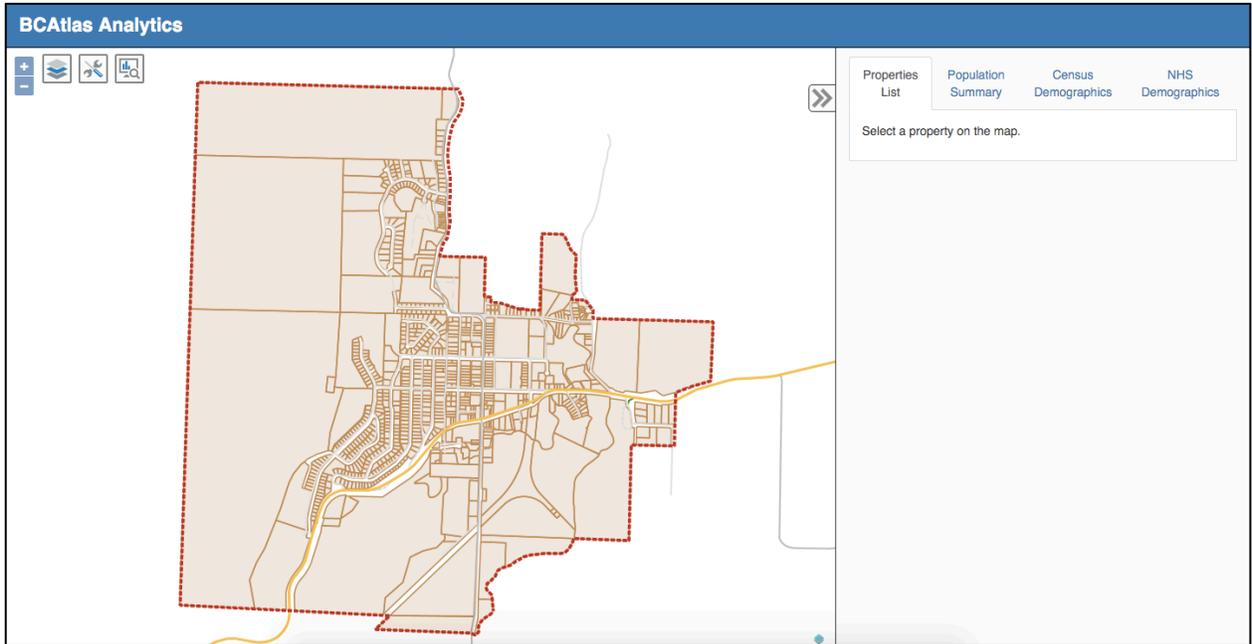


Figure 4: Screenshot of initial map layout

### 3.2 Parcel Manipulation and Data Summary

Parcels may be selected individually or as a group. To select an individual parcel click the desired parcel. To select a group of parcels click the map tools icon and select the polygon selection tool (Figure 5).

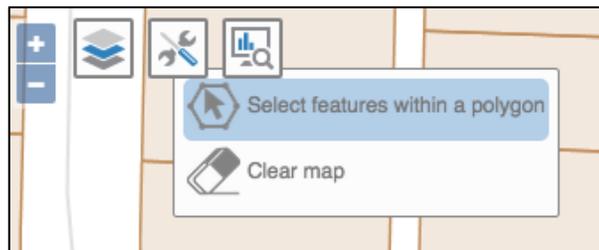
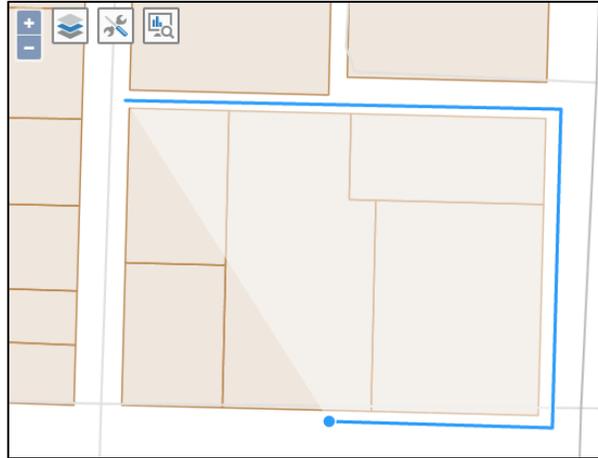


Figure 5: Screenshot of map tools

Once the polygon selection tool has been enabled click a location on the map to add a vertex. Continue this process until the desired shape has been drawn (Figure 6). Double click to finish drawing. All properties that fall within or intersect the drawn polygon will be highlighted.



**Figure 6: Screenshot of polygon selection tool**

After a property or group of properties has been selected the attributes panel will populate with information relating to the properties. The Properties List tab contains a list of all of the selected properties along with an aggregated list of summarized property details if more than one property is selected. Only properties that have associated BC Assessment data are included in the property list and summary (Figure 7). The information contained in the property list summary includes:

- Number of selected parcels with BC Assessment data
- Maximum total assessed value of all selected parcels BC Assessment data
- Minimum total assessed value of all selected parcels BC Assessment data
- Sum of total assessed value of all selected parcels with BC Assessment data
- Average total assessed value of all selected parcels with BC Assessment data

The information displayed for each selected parcel with BC Assessment data contains:

- Parcel address
- Jurisdiction roll number (unique property identifier)
- Lot size
- Assessed land value
- Assessed building value
- Total assessed value

Properties List	Population Summary	Census Demographics	NHS Demographics
<b>Number of parcels with BC Assessment data:</b>			5
<b>Maximum assessed value:</b>			\$322,400
<b>Minimum assessed value:</b>			\$162,000
<b>Total assessed value:</b>			\$1,124,100
<b>Average assessed value:</b>			\$224,820
<hr/>			
2230 SHUSWAP AVE LUMBY			
<b>Jurisdiction Roll</b>			54102321010
<b>Lot Size</b>			0.402 Acres
<b>Assessed Land Value</b>			\$74,300
<b>Assessed Building Value</b>			\$163,000
<b>Total Assessed Value</b>			\$237,300
<hr/>			
2059 GLENCAIRD ST LUMBY			
<b>Jurisdiction Roll</b>			54102318000
<b>Lot Size</b>			0.77 Acres
<b>Assessed Land Value</b>			\$89,400
<b>Assessed Building Value</b>			\$233,000

**Figure 7: Screenshot of properties list**

The Population Summary tab provides a summary of the population in the Census subdivision of the selected properties. This summary is based upon the 2011 Canadian Census (Figure 8). The information included for a given Census subdivision in the Population Summary tab is consists of:

- Census subdivision population
- Median age of the Census subdivision
- Number of households within the Census subdivision
- Average number of people per family within the Census subdivision
- Average number of children at home within the Census subdivision

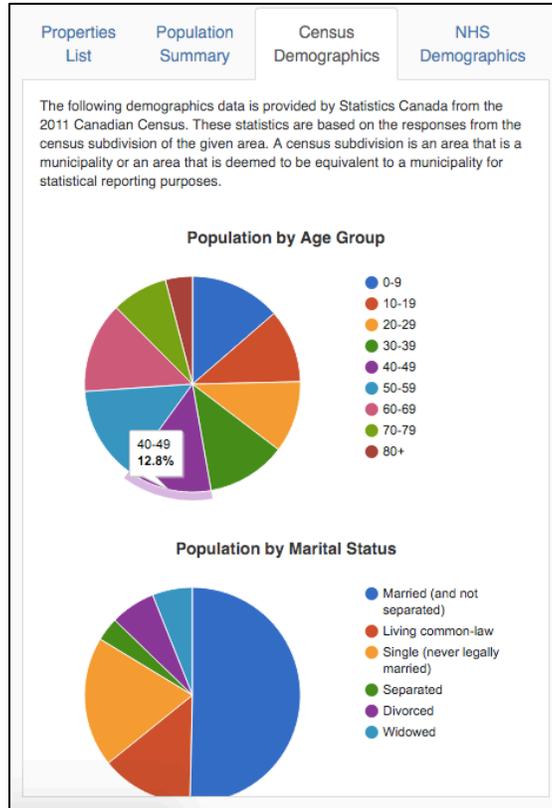


**Figure 8: Screenshot of Population Summary tab**

The Census Demographics tab provides a detailed demographic breakdown, using pie charts, of the selected property or properties' surrounding Census subdivision (Figure 9). The data used in this detailed breakdown is based upon the 2011 Canadian Census. The charts displayed within the Census Demographics tab include the following demographic breakdowns for a given Census subdivision:

- Population by age group
- Population by language spoken at home
- Population by dwelling type
- Population by dwelling size

To view the percentage breakdown for a specific data item, hover over the chart area for the data item of interest and a tooltip containing the name of the data item along with its percentage of the total chart area will appear.



**Figure 9: Screenshot of Census Demographics tab**

The NHS Demographics tab also presents information using pie charts. The NHS Demographics tabs provide the following demographic breakdowns for the given Census subdivision based upon the 2011 NHS:

- Population by highest level of education
- Population by occupation type
- Population by gross income

### **3.3 Legend and Layer Manipulation**

The legend can be toggled using the layers icon located in the top right corner of the map (Figure 10). Checking the checkbox associated with a given layer will turn on the layer. Uncheck to checkbox to turn off the layer. Note: the more layers you have turned on, the greater the time required to load the map.

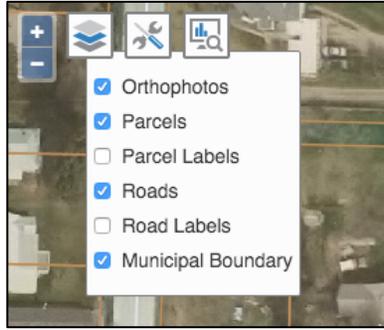


Figure 10: Screenshot of the legend

### 3.4 Data Analysis

BCAtlas Analytics offers two types of geographic analysis in the form of map layers based on the 2015 BC Assessment data. To access the map's analysis layers click the analysis icon. The first analysis layer shows the value per square foot for each property. Each value is calculated by dividing the total assessed value of each property by the calculated square footage. The total assessed value of a property is determined by totaling the assessed building and land values for each property. The values for each property were ranked on a scale and presented on a choropleth map (Figure 11). The second analysis layer shows the property's building value as a percentage of the property's total value. To calculate this value for each property the property's building value was divided by the property's total value and multiplied by 100%. These values were also displayed using a choropleth map. Properties that are not colored on either of the choropleth maps do not have any associated BC Assessment data.

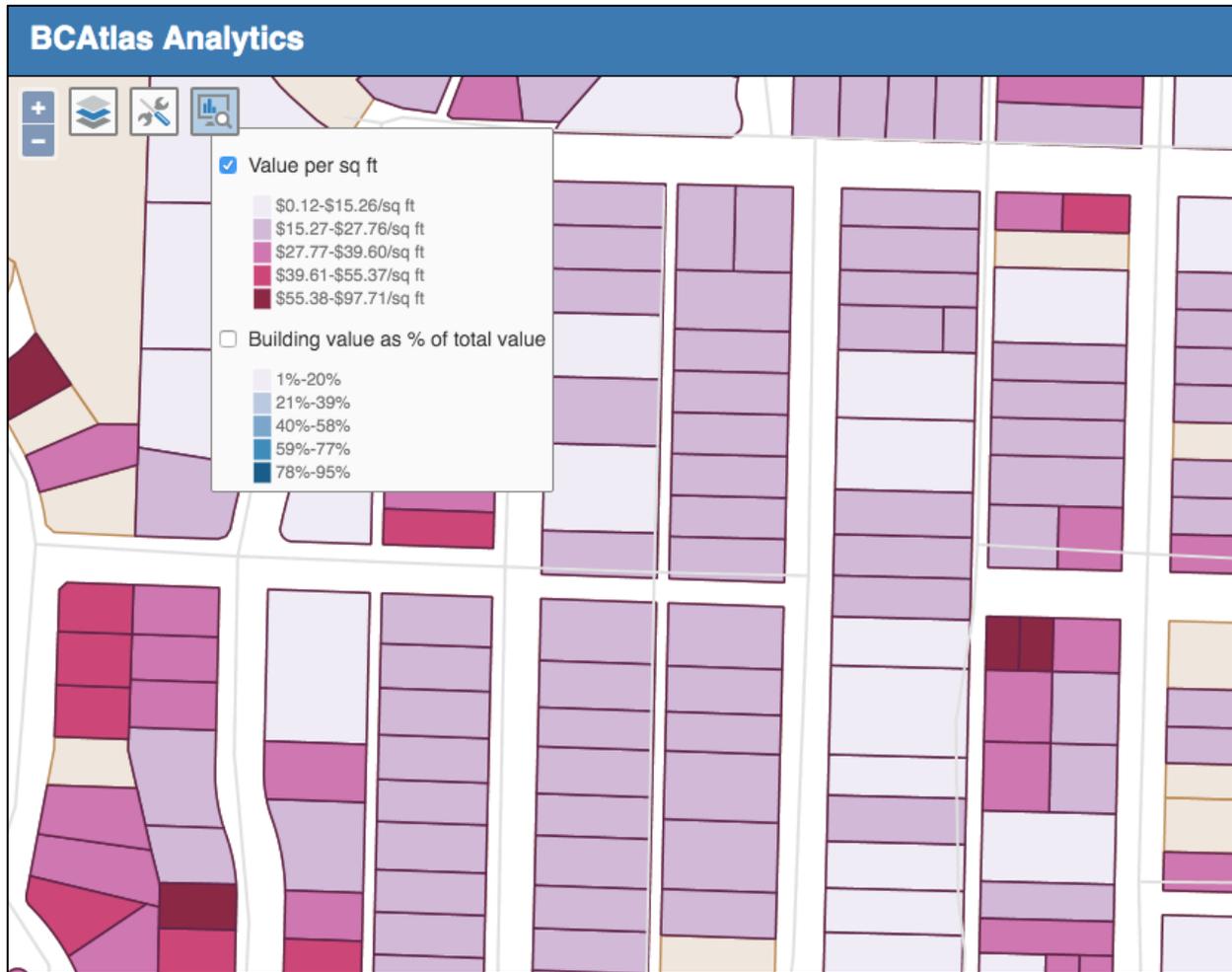


Figure 11: Screenshot of value per sq. ft. analysis layer

### 3.5 Accessing BCAtlas Analytics' Web Services

The attribute tables containing information regarding the 2011 Canadian Census and the 2011 NHS can be accessed using GET requests. The two data analysis layers and the Village of Lumby Parcels layer can be accessed using a WMS request.

#### 3.5.1 GET Request URLs

BCAtlas Analytics supports GET requests for users to obtain data from the 2011 Canadian Census and the 2011 NHS. Request responses are returned in JSON format. The following URLs can be used to perform a GET request to BCAtlas Analytics where 54101100000 is a valid parcel id and 5937005 is a valid Census subdivision id.

1. 2011 Canadian Census summary of a Census subdivision

- a. By parcel id:  
<http://159.203.2.8:8080/getsummaryinfo?type=pid&pid=54101100000>
  - b. By Census subdivision id:  
<http://159.203.2.8:8080/getsummaryinfo?type=csduid&csduid=5937005>
2. 2011 Canadian Census demographics
- a. By parcel id:  
<http://159.203.2.8:8080/getdemoinfo?type=pid&pid=54101100000&source=Census>
  - b. By Census subdivision id:  
<http://159.203.2.8:8080/getdemoinfo?type=csduid&csduid=5937005&source=Census>
3. 2011 Canadian NHS demographics
- a. By parcel id:  
<http://159.203.2.8:8080/getdemoinfo?type=pid&pid=54101100000&source=nhs>
  - b. By Census subdivision id:  
<http://159.203.2.8:8080/getdemoinfo?type=csduid&csduid=5937005&source=nhs>

Responses from a query using URLs 1a and 2a are returned in the format:

```
{
  "Property": "value",
  ...
}
```

Where a “Property” is a String data attribute (i.e. “Population in 2011”) and value is a String data value (i.e. “1731”). Responses from a query using URLs 2a, 2b, 3a, and 3b are returned in the format:

```
{
  "Group": {
    "Property": {
      "total": x,
      "male": x,
```

```

        "female": x
    },
    ...
},
...
}

```

Where “Group” is a common category for a group of attributes (i.e. “Education”), “Property” is a String data attribute (i.e. “Total population aged 15 years and over by highest certificate, diploma or degree”), “total” is the identifier to obtain the numeric count of the number of individuals that meet the property’s criteria, “male” is the identifier to obtain the numeric count of the number of males that meet the property’s criteria, and “female” is the identifier to obtain the numeric count of the number of females that meet the property’s criteria.

### 3.5.2 WMS Requests

WMS requests can be made to use any of the map layers that are generated by BCAtlas Analytics. The WMS URL for BCAtlas Analytics is <http://159.203.2.8:8080/geoserver/bcatlas/wms>. The layers that are available for WMS requests are listed in Table 1.

Table 1: Layers for WMS requests

Map Layer Name	WMS Layer Name
Parcels	BCAtlas:parcel_postgis
Value per sq. ft.	BCAtlas:parcelvalue
Building value as a % of total value	BCAtlas:buildingpct

## **4 Technical Details and Implementation**

### **4.1 Development Environment**

All of the development for this application was completed in the Java IDE IntelliJ. IntelliJ's Web development module was enabled to compile source code into a WAR file for application deployment on a web server.

### **4.2 Server Configuration**

BCAtlas Analytics is hosted on a Digital Ocean Droplet that is located in a server warehouse in Toronto. During the Droplet's setup process Ubuntu, Apache Tomcat, Geoserver, and a PostgreSQL database with a PostGIS extension were installed.

#### **4.2.1 Database Configuration**

A PostgreSQL database was selected for this project because of the database's spatial database extension PostGIS. PostGIS adds support for geographic objects allowing SQL spatial queries to be performed [18]. BCAtlas Analytics uses spatial objects and queries to store Shapefiles, find the intersection of polygons, and to create the map's data analysis layers.

The BCAtlas Analytics database consists of five interconnected relations: Parcel, ParcelGIS, ParcelMappings, CensusSubdivision, and BCNHS (Figure 12). The Parcel relation contains non-spatial information related to a parcel and was obtained from the 2015 BC Assessment of the Village of Lumby. The ParcelGIS relation contains spatial information related to a parcel. This relation was populated using the PostGIS `shp2pgsql` command that converted the parcels from the Shapefile provided by the Village of Lumby to a spatial relation. The CensusSubdivision relation is made up of Census profile information from the 2011 Census for all of Canada's Census subdivisions. The BCNHS relation contains the 2011 NHS profile information for all Census subdivisions in British Columbia. All parcels are uniquely identified by parcel id that is called a `jur_roll` or `gislcp` depending on the data source and census subdivisions

are uniquely identified by a Census subdivision id (csduid). There are multiple parcels in every Census subdivision.

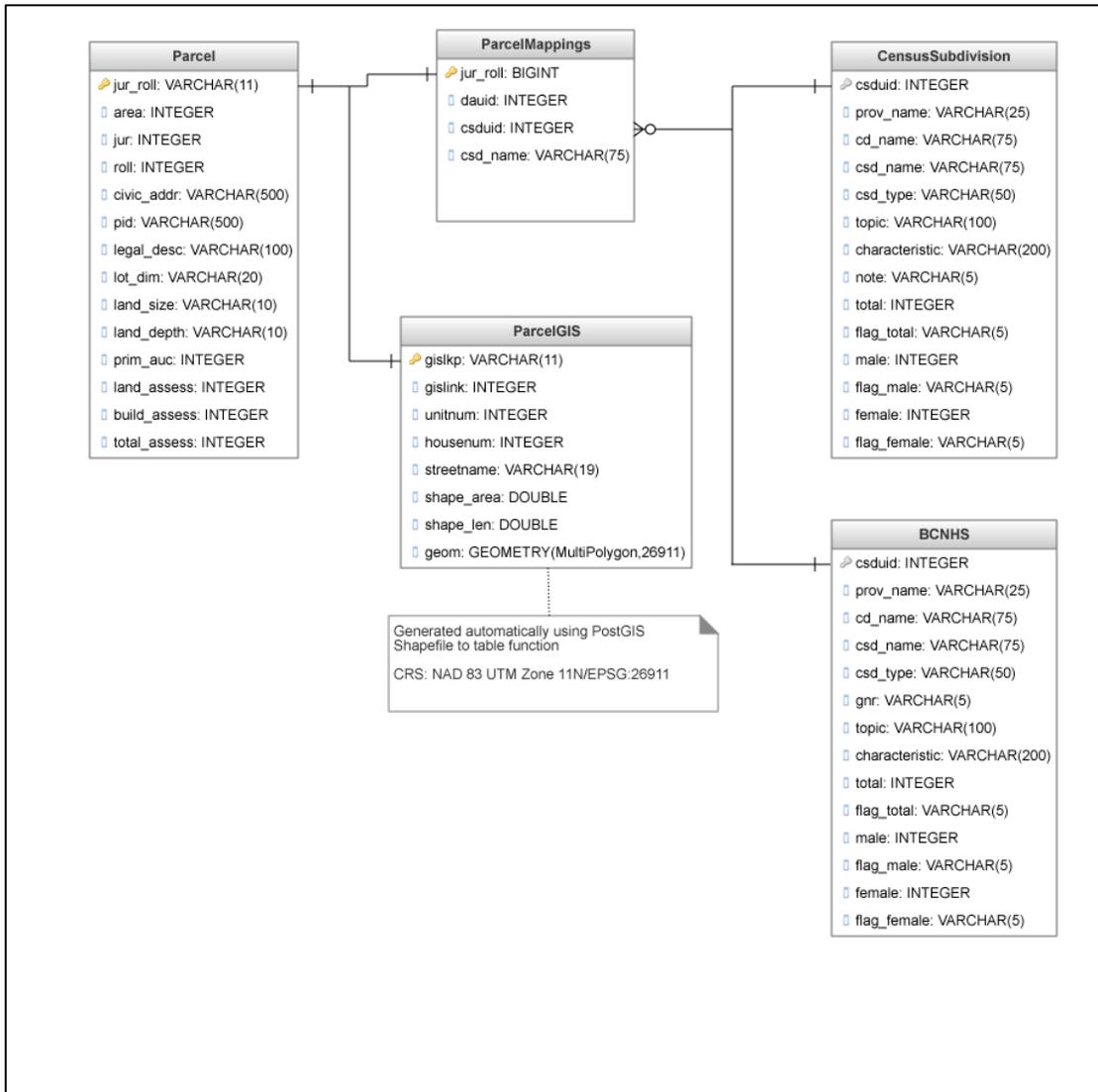


Figure 12: Database diagram for the BCAtlas Analytics PostgreSQL database

### 4.3 Creating, Loading, and Querying the Map

#### 4.3.1 Geoserver

Geoserver was installed on the BCAtlas Analytics server to enable the web application to publish its own spatial data. Geoserver is used to publish the Parcels, Value per sq. ft., and Building value as a % of total value layers. Geoserver obtains the data for the Parcels layer from the ParcelGIS relation in the BCAtlas Analytics PostgreSQL

database. The other two layers are SQL Views created in Geoserver. Joining the Parcel and ParcelGIS relations created the view for the Value per sq. ft. layer (Figure 13).

```
SELECT (total_assess/shape_area)*0.09290304 as sqft_val, gislkp,
shape_area, land_assess, build_assess, total_assess, shape_len,
geom
FROM ParcelGIS, Parcel
WHERE Parcel.jur_roll::varchar = ParcelGIS.gislkp
ORDER BY sqft_val
```

Figure 13: SQL statement used to create the Value per sq. ft. view

The default unit of measurement for shape areas in the NAD 83 UTM Zone 11N/EPSSG:26911 CRS is meters squared so the initial units of the value calculation were \$/m<sup>2</sup>. This value was multiplied by 0.09290304 to convert the units to \$/sq. ft. The Parcel and ParcelGIS relations were joined again to create the Building value as a % of total value layer (Figure 14). The geometry and shape area fields were included in both views to maintain the spatiality of the data. Additional attributes such as gislkp, land\_assess, build\_assess, etc. were included so users who accessed this layer via WMS would also have access to these attribute fields.

```
SELECT (build_assess/total_assess::float)*100 as building_pct,
gislkp, shape_area, land_assess, build_assess, total_assess,
shape_len, geom
FROM ParcelGIS, Parcel
WHERE Parcel.jur_roll::varchar = ParcelGIS.gislkp
ORDER BY building_pct
```

Figure 14: SQL statement used to create the Building value as a % of total value layer

### 4.3.2 Loading Map Data

OpenLayers, an open source JavaScript library, was used to render and manipulate layers on BCAtlas Analytics. A new instance of the BCAtlas Analytics map is loaded every time the user visits or refreshes the web application. There are two steps to loading the map. First, the map is initialized to the NAD 83 UTM Zone 11N/EPSSG:26911 and then layers are added. All layers with the exception of the Highlight Overlay and Drawing layers are preexisting on either the BCAtlas Analytics or LandInfo Technologies servers and are loaded using the OpenLayers WMS function. The Highlight Overlay and Drawing layers are vector layers that are created on page load and persist throughout a user's session. Both of these layers are used to handle a user's interactions with the map.

### 4.3.3 Styling Map Layers

Geoserver does not have a default styling or graphical styling interface so all layer styling had to be completed manually using a text editor and the WMS SLD specification. Basic line and polygon styles were applied to the Parcels, Parcel Labels, Roads, Road Labels, and Municipal Boundary layers as described by the Geoserver SLD Cookbook [33]. Prior to completing the choropleth styling for the Value per sq. ft. and Building value as a % of total value layers a statistical analysis was applied to determine where breaks should be placed to create classed data. The placement of breaks within the data is important because incorrectly placed breaks can skew data values and alter a viewer's interpretation of the map [34]. In order to correctly place class breaks, the values from the Value per sq. ft. and Building value as a % of total value layers were loaded into RStudio, plotted, and summarized to view general trends (Figure 15, Figure 17, and Table 2).

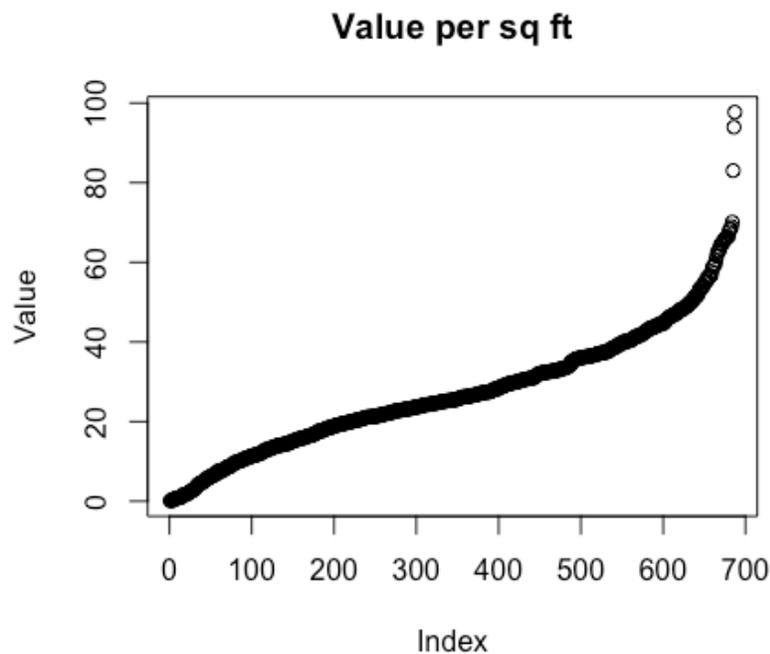


Figure 15: Plot of Value per sq. ft. values

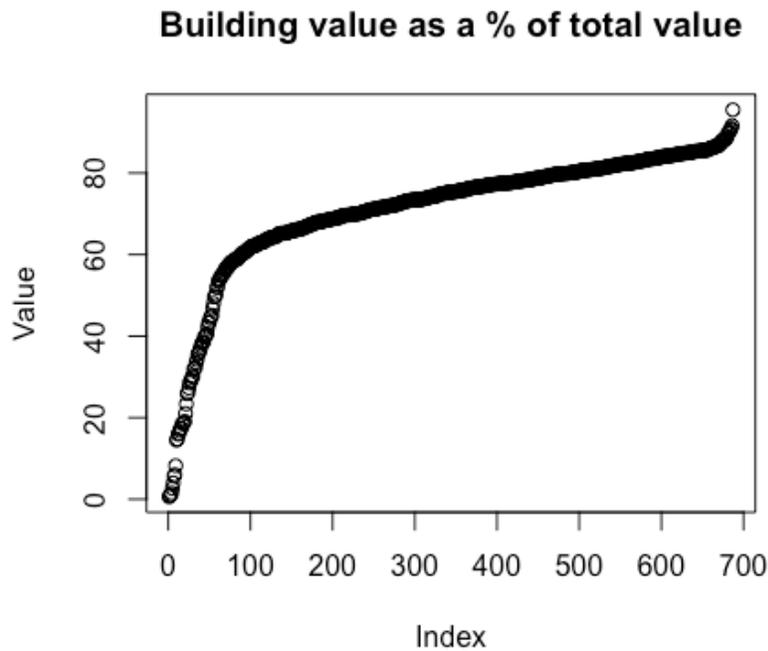


Figure 16: Plot of Building value as a % of total value values

Table 2: Summary of layer data values obtained using the R summary function

	Value per sq. ft.	Building values as % total value
<b>Minimum</b>	0.1248	0.6061
<b>1<sup>st</sup> Quartile</b>	16.5866	67.2725
<b>Median</b>	25.3735	75.2515
<b>Mean</b>	27.4384	71.1111
<b>3<sup>rd</sup> Quartile</b>	36.6626	80.8614
<b>Maximum</b>	97.7131	95.4977

The three classification methods that could be used for data of this nature are equal interval, quantile, or natural breaks [34]. The equal interval classification method divides the data into a specified number of equal-sized classes. This classification method should be applied to familiar data ranges such as temperature and percentages [34]. The quantile classification method divides data into classes that contain an equal number of features and is well suited for linear data. Finally, the natural breaks method is based upon the Jenks Natural Breaks algorithm and classifies data based upon inherent groupings in the data. This algorithm identifies class breaks “that best group similar values that maximize the differences between the classes” [34]. Considering the

data presented in Figures 16, 17, and Table 2 and the data requirements for each of the classification methods mentioned above the natural breaks and interval classification methods were selected for the Value per sq. ft. data and Building value as a % of total value layers respectively. The data in the Value per sq. ft. layer is not a familiar data range, nor is it linearly distributed therefore the natural breaks classification method is the only valid method of classification. The Building value as a % of total value is not linearly distributed but since the data values are represented as a percentage the equal interval classification method was selected to classify the data. Each selected classification method was applied to the data using R and the resulting classifications are shown in Table 3 and Table 4.

**Table 3: Value per sq. ft. classification values**

<b>Value per sq. ft.</b>		
<b>Class Number</b>	<b>Data Range</b>	<b>Number of Data Points</b>
1	[0.1248175,15.25848]	153
2	(15.25848,27.75906]	239
3	(27.75906,39.60249]	158
4	(39.60249,55.36778]	103
5	(55.36778,97.71308]	34

**Table 4: Building value as a % total value classification values**

<b>Building value as a % total value</b>		
<b>Class Number</b>	<b>Data Range</b>	<b>Number of Data Points</b>
1	[0.6060606,19.58438)	20
2	[19.58438,38.56271)	22
3	[38.56271,57.54103)	32
4	[57.54103,76.51935)	299
5	[76.51935,95.49767]	314

Majority of the data points in the Building value as a % of total value classification fall into classes four or five. While this limits the separation it provides an accurate representation of the data by showing the user that majority of buildings are worth between 58-95% of a parcel's total value. Using another classification method, like natural breaks, would create a more even distribution of data points amongst the

classes but it would suggest to the user that the different building values between parcels is greater than the data indicates.

After the appropriate classifications had been determined, colors were selected to show an increase in value magnitude with each increase in class. Finally, the selected colors and class breaks were formatted using the SLD standard and applied to the two layers.

#### 4.3.4 Handling Map Interactions

A user can interact with the BCAtlas Analytics map via parcel selection. There are two ways a user can select a parcel: clicking on single parcel or using the polygon select tool to select a group of parcels. When a user clicks on a single parcel the coordinates of the user's click event are logged and sent with the map's spatial information to Geoserver to perform a feature lookup on the Parcels layer. Geoserver returns a response in GeoJSON. The coordinates of the parcel are then extracted from the GeoJSON object and added to the Highlight Overlay layer (Figure 17).

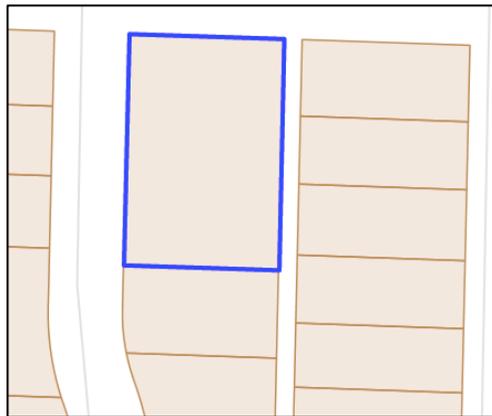


Figure 17: Selected property added to the Highlight Overlay layer

When a user uses the polygon select tool to select a group of parcels, the process to determine which parcels have been selected is more difficult. When a user finishes drawing his polygon the coordinates of the polygon are collected and sent via GET request to the GetShapeIntersection Java servlet. This servlet makes a spatial query to the ParcelGIS relation to find all parcels that intersect the polygon drawn by the user (Figure 18). For all parcels that intersect the drawn polygon, this query returns the giskp (unique identifier), x-coordinate of the geometric center of the parcel, and y-coordinate

of the geometric center. These values are then parsed into a JSON object and returned to the requesting JavaScript function. After this initial request has returned the x-y coordinates of each parcel are extracted, the same process required to highlight a single parcel is followed.

```
SELECT gislkp, ST_X(ST_Centroid(geom)), ST_Y(ST_Centroid(geom))
FROM ParcelGIS AS p, (
  SELECT ST_TRANSFORM(
    ST_Polygon(
      ST_GeomFromText(
        'LINESTRING(359001.7155040242 5568481.755225071
1,360189.9351145816 5568028.418425863 1,358524.5188732782
5567503.502132042 1,359001.7155040242 5568481.755225071 1)'
      ), 26911),26911) as geometry) AS poly WHERE
ST_Intersects(p.geom, poly.geometry) = true
```

Figure 18: Spatial query (with PreparedStatement values supplied) to find all intersections of a user drawn polygon and parcels

## 4.4 Displaying Information in the Attribute Panel

After a parcel has been selected and highlighted a series of requests are made to populate the Properties, Population Summary, Census Demographics, and NHS Demographics Tabs. All of these requests are made asynchronously.

### 4.4.1 Properties List Tab

After the selected parcel(s) has been highlighted the parcel id(s) is sent via GET request to the GetParcelInfo servlet. This servlet makes a query to the Parcel relation to obtain the BC Assessment data related to the given parcel(s) and returns a JSON object that it is parsed and displayed in the Properties List tab using JavaScript. If more than one parcel is returned an aggregated summary of all parcels will also be displayed within the tab.

### 4.4.2 Finding a Parcel's Census Subdivision Id

To populate the remaining tabs a parcel's Census subdivision id must be known. A GET request that facilitates a query to the ParcelMappings table provides the Census subdivision id(s) for the provided parcel(s). The response to this query is then formatted as a JSON object and returned.

#### **4.4.3 Population Summary Tab**

Using the Census subdivision id obtained from the process outlined in Section 4.4.2, a GET request is made to the GetSummaryInfo servlet. This servlet takes in either a Census subdivision id or a parcel id and makes a query to the CensusSubdivision relation for the information required to populate the Population Summary tab. The results from this query are formatted as a JSON object and returned. Upon return, a table of values is created using the contents of the JSON object and displayed to the user.

#### **4.4.4 Census and NHS Demographics Tabs**

The process to population the Census and NHS Demographics tabs is very similar. Using the Census subdivision id obtained from the process outlined in Section 4.4.2, a request is made to the GetDemolInfo servlet with the Census subdivision id as a parameter and a source parameter value of either “census” or “nhs” depending on the desired data source. Requests to this servlet can also be made using a parcel id in place of a Census subdivision id. Based upon the source parameter a query to either the CensusSubdivision or BCNHS relations is made and the results are parsed and returned as a JSON object. Once the result has been returned the data is extracted and formatted into pie charts using the Google Charts JavaScript charting library.

## 5 Discussion and Conclusion

### 5.1 Discussion

This project set out to accomplish two goals. First, it looked to develop an interactive web application that would allow small municipalities to make use of government data to aid in the decision making process. Second, it aimed to act as a feature service allowing 3<sup>rd</sup> party users to access the data and layers used by the BCAtlas Analytics web interface for use in other web applications. Both of these goals were accomplished but not without challenges. The most significant difficulty faced while completing this project was the availability of data at high-resolution levels. Due to a small population size, low response rates, and issues with data confidentiality the greatest level of data detail for Census and NHS data in the study region were Census subdivisions. This limited the types of analysis that could be offered because the entire study region encompassed one Census subdivision. Projecting spatial data and performing spatial queries in the correct CRS also proved to be especially difficult. During initial development the map was not correctly projected into the NAD 83 UTM Zone 11N/EPG:26911. This created a problem when developing user interactions because x-y coordinate click locations were logged in the default WGS 84/EPG:4326. As a result no parcels were returned from feature lookup requests sent to Geoserver, however an error was not generated because the requests were valid. This made the problem especially difficult to debug. Finally, learning how to perform spatial queries in PostGIS was very challenging. Geoserver does not have a built in function that finds the intersection of two features so this functionality had to be created by building a Java servlet that generates a PostGIS polygon object and queries the database to find all intersecting parcels.

This project only focuses on providing a solution to the digital divide faced by the Village of Lumby however the system was built so it can be generalized to other municipalities within British Columbia. The BCAtlas Analytics database contains 2011 Canadian Census for all of Canada and NHS data for the entire province of British Columbia. To extend the system to work with data from another municipality a Shapefile of the

municipality's parcels would have to be obtained in addition to BC Assessment data for the area. Depending on the location of the municipality a new map projection may have to be considered. Referring to the map in Figure 1 if the new municipality is outside of the shaded red area a different projection will have to be used to ensure data quality.

After the 2016 Canadian Census is completed the data in the BCAtlas Analytics database should be updated. Unlike the 2011 Census the 2016 Census will be a long form Census, making the NHS obsolete. This will reduce the number of sources required to generate demographics data for the BCAtlas Analytics web application. Additionally, unlike the NHS, the Census is mandatory to complete so the issues with data quality discussed in Section 2.3 will be minimized. Currently, the locations of parcels within the Village of Lumby are outdated and largely inaccurate. In approximately 1-2 years time when the BC Land and Title Survey has completed an accurate parcel map of British Columbia the parcels within the Village of Lumby should be extracted from this map and used in place of the current parcels on the BCAtlas Analytics map.

## **5.2 Conclusion**

As the availability of open data and online resources that can be used to aid decision making increase, the digital divide among municipalities will continue to increase unless open source or inexpensive solutions are put forth. BCAtlas Analytics aims to help reduce the digital divide between municipalities within British Columbia. Through BCAtlas Analytics, municipalities can use their own spatial datasets along with provincial and federal data to make more accurate and informed decisions. Additionally, to encourage the development of new applications that will help reduce the digital divide the data and layers used by BCAtlas Analytics are available to 3<sup>rd</sup> party users to access via GET or WMS requests.

## 6 References

- [1] Jan A.G.M. van Dijk, "Digital divide research, achievements and shortcomings," *Poetics*, vol. 34, no. 4-5, pp. 221-235, 2006.
- [2] National Geographic. National Geographic - GIS. [Online].  
<http://education.nationalgeographic.org/encyclopedia/geographic-information-system-gis/>
- [3] Esri. Martha's Vineyard Commission. [Online].  
[http://mvcommission.actwin.com/doc.php/GIS\\_for\\_government.pdf?id=57](http://mvcommission.actwin.com/doc.php/GIS_for_government.pdf?id=57)
- [4] Christopher Brown. (2016, March) GIS Stackexchange. [Online].  
<http://gis.stackexchange.com/a/55264>
- [5] Jeremy Nicoletti, Michael Rivera, Steven Whitmeyer, and De Paror Declan, "Creating visual aids and interactive geological maps using virtual globes," *Geological Society of America*, vol. 41, no. 1, p. 11, March 2009.
- [6] France Bélanger and Lemuria Carter, "The Effects of the Digital Divide on E-government: An Empirical Evaluation," in *Proceedings of the 39th Hawaii International Conference on System Sciences - 2006*, Kauia, HI, 2006, pp. 81c - 81c.
- [7] Vangie Beal. Webopedia. [Online]. <http://www.webopedia.com/TERM/A/API.html>
- [8] Statistics Canada. (2015, November) Statistics Canada. [Online].  
<https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo012-eng.cfm>
- [9] Esri. GIS Dictionary - Choropleth Map. [Online].  
<http://support.esri.com/en/knowledgebase/GISDictionary/term/choropleth%20map>
- [10] GeoServer. (2015, April) GeoServer 2.8.x User Manual. [Online].  
<http://docs.geoserver.org/2.8.x/en/user/webadmin/data/layers.html>
- [11] Statistics Canada. (2015, November) Statistics Canada. [Online].  
<https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo021-eng.cfm>

- [12] w3schools.com. HTTP Methods: GET vs. POST. [Online].  
[http://www.w3schools.com/tags/ref\\_httpmethods.asp](http://www.w3schools.com/tags/ref_httpmethods.asp)
- [13] Baecker, Card Hewett. (2009, July) ACM SIGCHI Curricula for Human-Computer Interaction. [Online]. [http://old.sigchi.org/cdg/cdg2.html#2\\_1](http://old.sigchi.org/cdg/cdg2.html#2_1)
- [14] James Gosling, Bill Joy, and Guy Steele, "The Java® Language Specification," p. 1, March 2015.
- [15] w3schools.com. JSON Tutorial. [Online]. <http://www.w3schools.com/json/>
- [16] Government of Canada. (2014, November) Open Data 101. [Online].  
<http://open.canada.ca/en/open-data-principles#toc94>
- [17] <http://openlayers.org/>. Openlayers 3. [Online]. <http://openlayers.org/>
- [18] PostGIS. About PostGIS. [Online]. <http://postgis.net/>
- [19] RStudio. RStudio. [Online]. <https://www.rstudio.com/about/>
- [20] Tomcat Apache. (2012) Java Servlet API. [Online].  
<https://tomcat.apache.org/tomcat-5.5-doc/servletapi/javase/servlet/Servlet.html>
- [21] OpenGIS. OGC - Styled Layer Descriptor. [Online].  
<http://www.opengeospatial.org/standards/sld>
- [22] Esri. Esri GIS Dictionary - Vector. [Online].  
<http://support.esri.com/en/knowledgebase/GISDictionary/term/vector>
- [23] Jason Hunter. (1993, October) Oracle. [Online].  
<http://www.oracle.com/technetwork/articles/javase/servletapi-137835.html>
- [24] OGC. (2016) Web Map Service. [Online].  
<http://www.opengeospatial.org/standards/wms>
- [25] Destination BC. (2015) Super, Natural British Columbia. [Online].  
<http://www.hellobc.com/lumby.aspx>
- [26] Statistics Canada. (2015, December) Statistics Canada. [Online].  
<https://www12.statcan.gc.ca/census-recensement/2011/ref/about-a-propos/rates-taux-eng.cfm>

- [27] Statistics Canada. (2012, October) Statistics Canada. [Online].  
<http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=5937005&Geo2=PR&Code2=01&Data=Count&SearchText=lumby&SearchType=Begins&SearchPR=01&B1=All&Custom=&TABID=1>
- [28] Statistics Canada. (2015, December) National Household Survey Final Response Rates. [Online]. [https://www12.statcan.gc.ca/nhs-enm/2011/ref/about-apropos/nhs-enm\\_r013.cfm?PRCODE=59&Lang=E](https://www12.statcan.gc.ca/nhs-enm/2011/ref/about-apropos/nhs-enm_r013.cfm?PRCODE=59&Lang=E)
- [29] Spatial Reference. EPSG:26911. [Online].  
<http://spatialreference.org/ref/epsg/nad83-utm-zone-11n/>
- [30] BC Land and Title Survey. ParcelMap BC. [Online]. <https://ltsa.ca/online-services/parcelmap-bc>
- [31] Map Town. (2001) Canadian Topographic Maps - British Columbia. [Online].  
<http://www.maptown.com/canadiantopographical/bcntsindex.html>
- [32] Design Principles FTW. (2013, September) Shneiderman's "Eight Golden Rules of Interface Design". [Online].  
<http://www.designprinciplesftw.com/collections/shneidermans-eight-golden-rules-of-interface-design>
- [33] Geoserver. (2016, April) SLD Cookbook. [Online].  
<http://docs.geoserver.org/latest/en/user/styling/sld-cookbook/>
- [34] ArcGIS Pro. ArcGIS for Desktop - Data Classification Methods. [Online].  
<http://pro.arcgis.com/en/pro-app/help/mapping/symbols-and-styles/data-classification-methods.htm>