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 and those who do not, or who have restricted access[[1](#Jan061)]. 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

Abstract 2

Table of Figures 4

Acknowledgements 5

1 Introduction 6

1.1 Motivation 6

2 Background 6

2.1 Terms 6

2.2 Study Region 7

2.3 Data Sources & Limitations 7

2.4 Interface Design 10

2.4.1 User Goals 10

2.4.2 Prototype and Design Rational 10

3 User Guide 13

4 References 16

# Table of Figures

Figure 1: British Columbia NAD 83 Zones [17] 9

Figure 2: Scanned image of paper prototype 11

Figure 3: Screenshot of final interface design 13

Figure 4: Screenshot of initial map layout 14

Figure 5: Screenshot of map tools 15

Figure 6: Screenshot of polygon selection tool 15

Figure 7: Screenshot of properties list 16

Figure 8: Screenshot of Population Summary tab 16

Figure 9: Screenshot of Census Demographics tab 17

# Acknowledgements

# Introduction

## Motivation

Cities throughout British Columbia vary greatly in size and by their availability of resources to spend on Geographic Information Systems (GIS).

# Background

## Terms

**Census Subdivision** – “Area that is a municipality or an area that is deemed to be equivalent to a municipality for statistical reporting purposes” [[2](#Sta15)]

**Choropleth map** – thematic map where areas are shaded in proportion to a measured value [[3](#Esr)]

**CRS** – Coordinate Reference System, defines how georeferenced spatial data relates to real locations on the Earth’s surface [[4](#Geo15)]

**Dissemination Area (DA)** – “Small area composed of one or more neighbouring dissemination blocks, with a population of 400 to 700 persons.” [[5](#Sta151)]

**GET Request** – Hyper Text Transfer Protocol that requests data from a specified resource over the Internet [[6](#w3s16)]

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

**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 [[7](#Hew09)]

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

**Java** – Widely used programming language with emphasis on concurrent, class based, and object oriented design. [[8](#Gos15)]

**JSON** – JavaScript Object Notation, syntax used for storing and exchanging data in a human readable format. [[9](#w3s161)]

**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 [[10](#Pos16)]

**PostgreSQL** – open source relational database management

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

**WAR file** – Web application archive file used to distribute a collection of Java Servlets, Java Classes, XML files, static webpages, and related files [[11](#Hun93)]

**WMS** – Web map service, Hyper Text Transfer Protocol used to request geospatial map images from a GIS database [[12](#OGC16)].

**URL** – Uniform Resource Locater, reference to a location on the Internet

## 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 [[13](#Des15)]. Although Lumby is small it maintains autonomy from larger, neighboring cities with its own Mayor and City Council. Main sources of industry for the village include tourism, manufacturing, and logging [[13](#Des15)].

## 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 availably freely to the public and can be obtained from Statistics Canada. This project was interested specifically in the 2011 Census and NHS as they related to Lumby. 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” [[14](#Sta152)]. Although Canadian citizens are required by law to complete the Canadian Census between 5-10% of all residents of Lumby did not complete in the 2011 Census questionnaire [[15](#Sta12)]. The average response rate for all residents of British Columbia was 96.5% [[14](#Sta152)]. In 2011 in addition to the Census Canadians were asked to complete voluntary NHS. The NHS response rate for Lumby residents was 65.2% whereas the provincial average was 77.1% [[16](#Sta153)]. Due to the Lumby’s small population size, low response rates, and issues with data confidentiality data 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 GISlkp that acts 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/EPSG:26911. This projection is appropriate for large and medium scale mapping topographic mapping for regions located within the highlighted area in Figure 1 [[17](#Spa)]. 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. Once a more accurate parcel Shapefile is available BCAtlas Analytics will be updated to reflect the change.

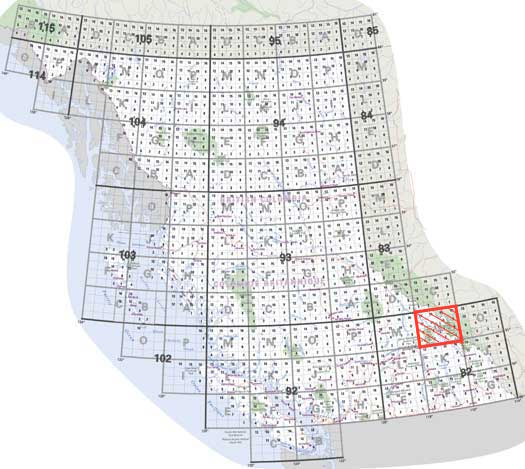


Figure : British Columbia NAD 83 Zones [[18](#Map01)]

Parcel assessment values were obtained from the 2015 assessment of parcels within the Village of Lumby completed by BC Assessment. This data included unique parcel identifiers that could be matched to parcel identifiers in the Lumby parcels Shapefile to obtain assessed property values for the land and building(s) located 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/EPSG:26911 CRS and obtained via WMS from LandInfo Technologies.

## Interface Design

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

### 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. Customize the interface to be able to maximize the area used to display the map

### 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 majority of the available screen space. To compensate for this problem the design of the this 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.

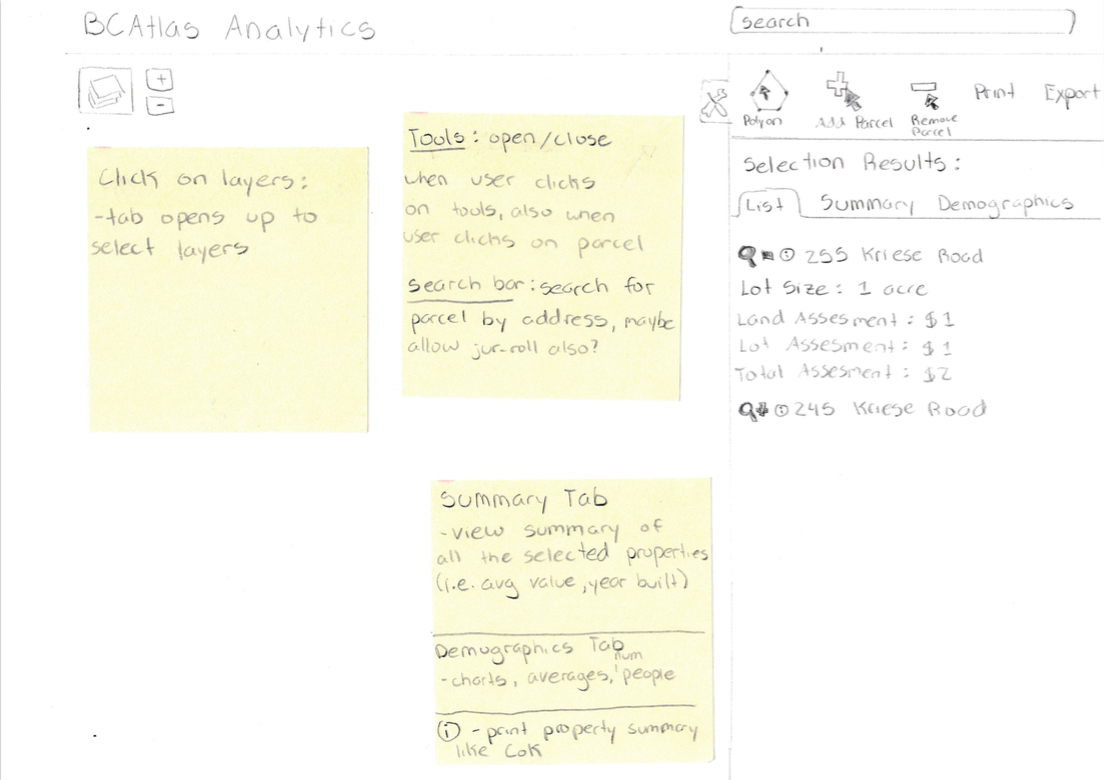


Figure : Scanned image of paper prototype

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 application and used by designers with the goal of creating a well designed and useable user interface [[19](#Des13)]. During this 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 [[19](#Des13)]. 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 or 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 [[19](#Des13)]. 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 should this error occur. 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 displays and pages, and reducing window-motion frequency [[19](#Des13)]. This principle was very difficult to accomplish when designing the user interface for this web application 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 was placed on the 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.

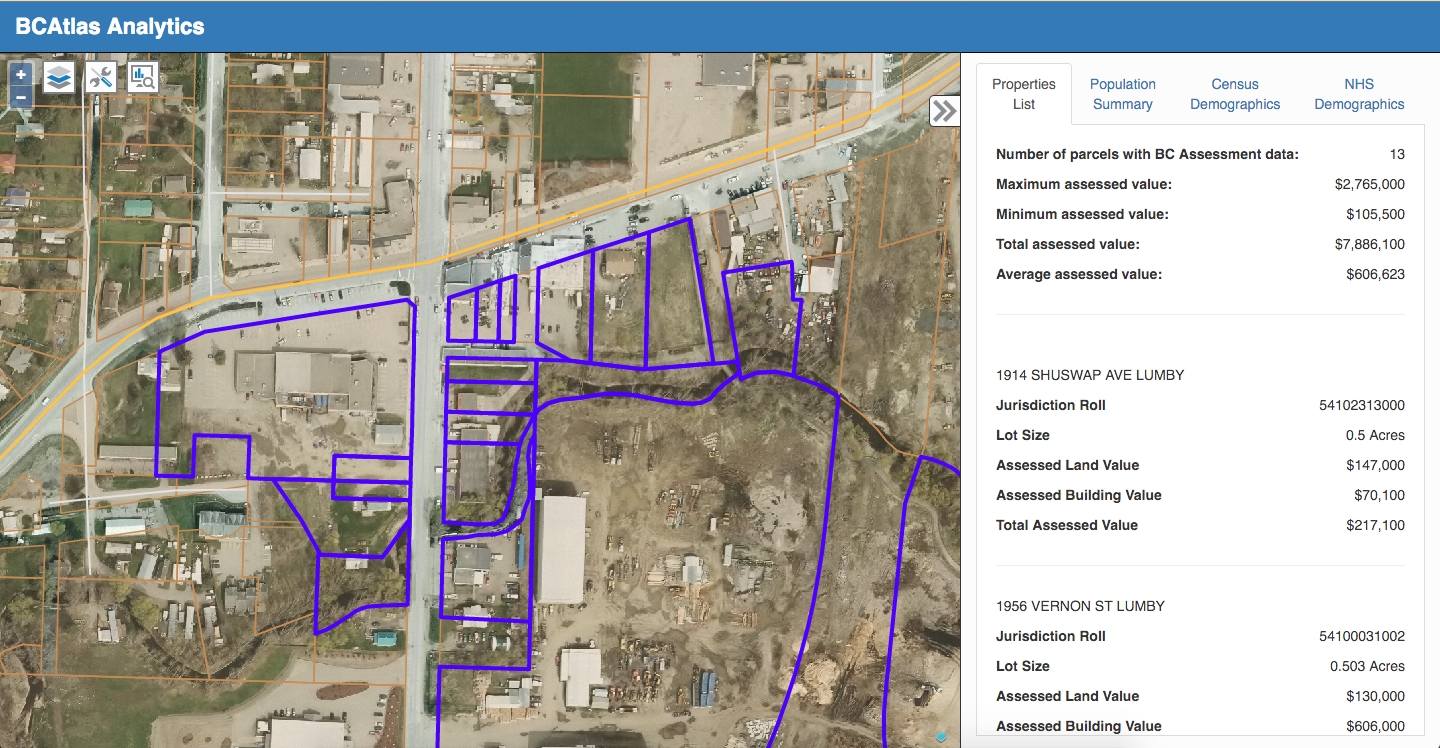


Figure : Screenshot of final interface design

# 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 either 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.

## Website Landing Page

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

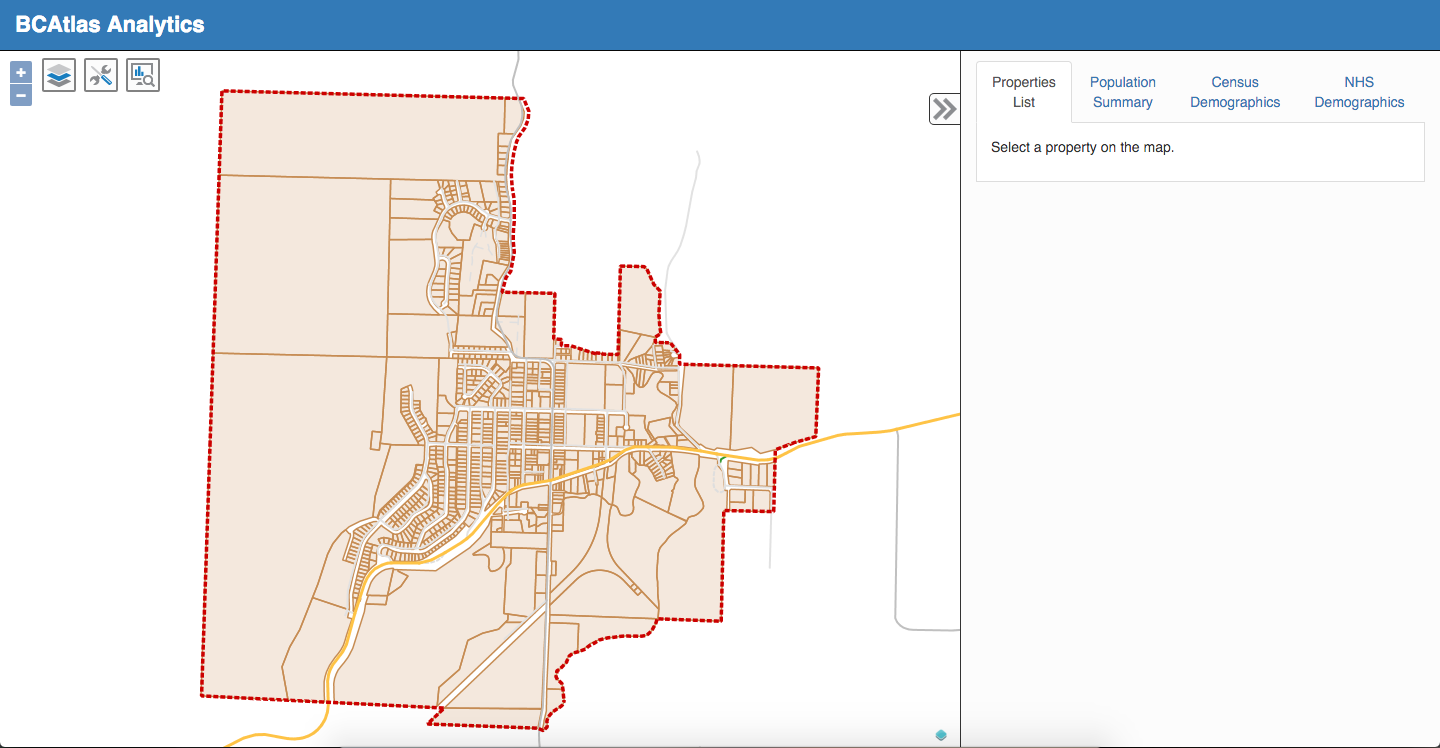


Figure : Screenshot of initial map layout

## Parcel Manipulation and Data Summary

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

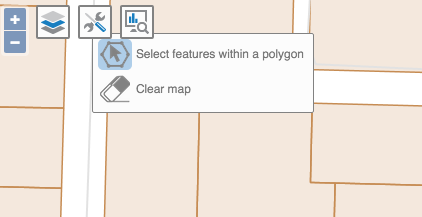


Figure : Screenshot of map tools

Once the polygon selection tool has been enabled click a location on the map to draw 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.

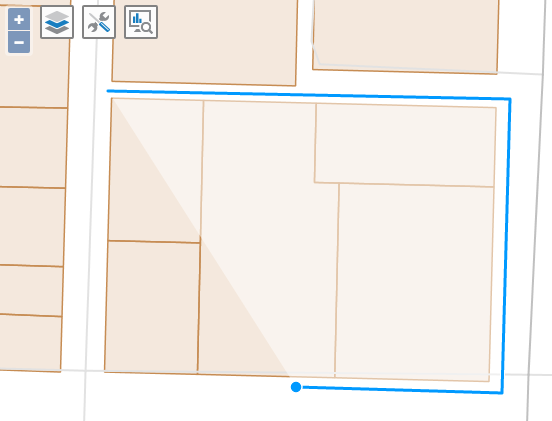


Figure : Screenshot of polygon selection tool

After a property or group of properties has been selected the attributes panel will populate with information relating to the property(s). The Properties List tab contains a list of all of the selected property(s) 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
* Summation 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 consists of:

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

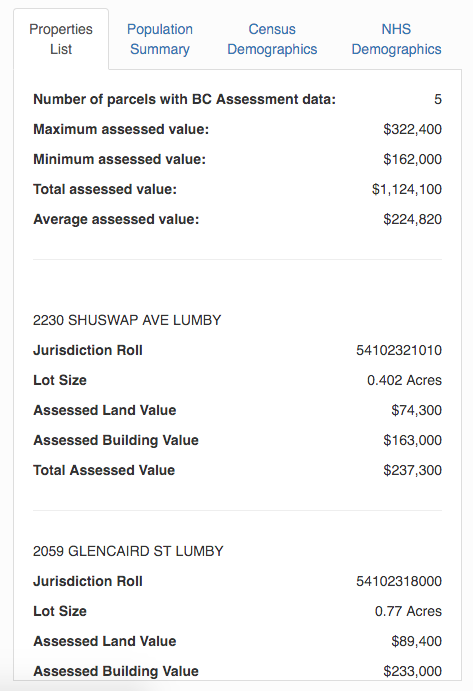


Figure : Screenshot of properties list

The Population Summary tab provides a summary of the population in the census subdivision of the selected property(s). 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 made up 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

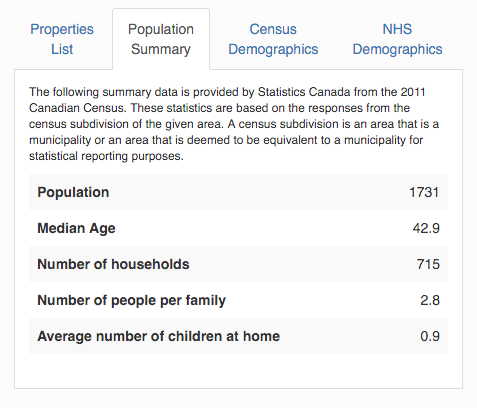


Figure : Screenshot of Population Summary tab

The Census Demographics tab provides a detailed demographic breakdown of the selected property or properties’ surrounding census subdivision. The data used in this detailed breakdown is based upon the 2011 Canadian Census. 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). 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

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

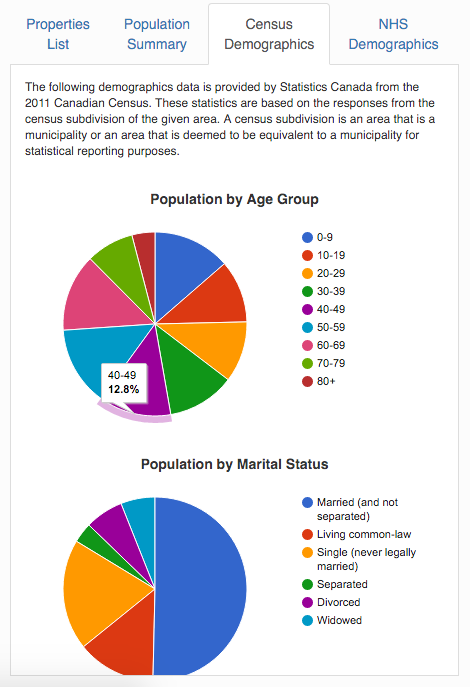


Figure : Screenshot of Census Demographics tab

## 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.

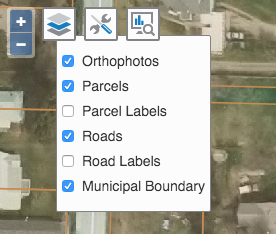


Figure : Screenshot of the legend

## 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 divided 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 choroplethmap (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. Property’s that do not receive a color value on either of the cholorpleth maps do not have any associated BC Assesment data.

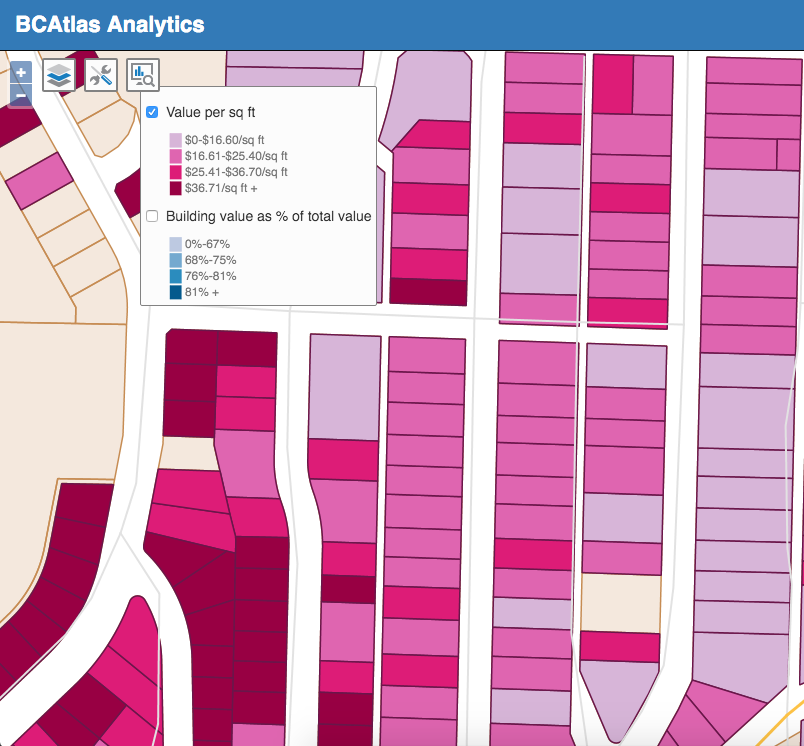


Figure : Screenshot of value per sq ft analysis layer

## 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 data analysis layers can be accessing using a WMS request.

### GET Request URLs

BCAtlas Analytics supports requests GET requests for users to obtain data the 2011 Canadian Census and the 2011 NHS. Request responses are returned in JSON format.

The following URLs are 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
   1. By parcel id: [http://159.203.2.8:8080/getsummaryinfo?type=pid&pid=54101100000](http://159.203.2.8:8080/getsummaryinfo?type=pid&pid=54101100000" \t "_blank)
   2. By census subdivision id: [http://159.203.2.8:8080/getsummaryinfo?type=csduid&csduid=5937005](http://159.203.2.8:8080/getsummaryinfo?type=csduid&csduid=5937005" \t "_blank)
2. 2011 Canadian Census demographics
   1. By parcel id: [http://159.203.2.8:8080/getdemoinfo?type=pid&pid=54101100000&source=census](http://159.203.2.8:8080/getdemoinfo?type=pid&pid=54101100000&source=census" \t "_blank)
   2. By census subdivision id: [http://159.203.2.8:8080/getdemoinfo?type=csduid&csduid=5937005&source=census](http://159.203.2.8:8080/getdemoinfo?type=csduid&csduid=5937005&source=census" \t "_blank)
3. 2011 Canadian NHS demographics
   1. By parcel id: [http://159.203.2.8:8080/getdemoinfo?type=pid&pid=54101100000&source=nhs](http://159.203.2.8:8080/getdemoinfo?type=pid&pid=54101100000&source=nhs" \t "_blank)
   2. By census subdivision id: [http://159.203.2.8:8080/getdemoinfo?type=csduid&csduid=5937005&source=nhs](http://159.203.2.8:8080/getdemoinfo?type=csduid&csduid=5937005&source=nhs" \t "_blank)

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.

### 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 their attributes that are available for WMS requests are

# References

x

|  |  |
| --- | --- |
| [1] | Jan A.G.M. van Dijk, "Digital divide research, achievements and shortcomings," *Poetics*, vol. 34, no. 4-5, pp. 221-235, 2006. |
| [2] | Statistics Canada. (2015, November) Statistics Canada. [Online]. <https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo012-eng.cfm> |
| [3] | Esri. GIS Dictionary. [Online]. <http://support.esri.com/en/knowledgebase/GISDictionary/term/choropleth%20map> |
| [4] | GeoServer. (2015, April) GeoServer 2.8.x User Manual. [Online]. <http://docs.geoserver.org/2.8.x/en/user/webadmin/data/layers.html> |
| [5] | Statistics Canada. (2015, November) Statistics Canada. [Online]. <https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo021-eng.cfm> |
| [6] | w3schools.com. HTTP Methods: GET vs. POST. [Online]. <http://www.w3schools.com/tags/ref_httpmethods.asp> |
| [7] | Baecker, Card Hewett. (2009, July) ACM SIGCHI Curricula for Human-Computer Interaction. [Online]. <http://old.sigchi.org/cdg/cdg2.html#2_1> |
| [8] | James Gosling, Bill Joy, and Guy Steele, "The Java® Language Specification," p. 1, March 2015. |
| [9] | w3schools.com. JSON Tutorial. [Online]. <http://www.w3schools.com/json/> |
| [10] | PostGIS. About PostGIS. [Online]. <http://postgis.net/> |
| [11] | Jason Hunter. (1993, October) Oracle. [Online]. <http://www.oracle.com/technetwork/articles/javase/servletapi-137835.html> |
| [12] | OGC. (2016) Web Map Service. [Online]. <http://www.opengeospatial.org/standards/wms> |
| [13] | Destination BC. (2015) Super, Natural British Columbia. [Online]. <http://www.hellobc.com/lumby.aspx> |
| [14] | Statistics Canada. (2015, December) Statistics Canada. [Online]. <https://www12.statcan.gc.ca/census-recensement/2011/ref/about-apropos/rates-taux-eng.cfm> |
| [15] | 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> |
| [16] | 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> |
| [17] | Spatial Reference. ESPG:26911. [Online]. <http://spatialreference.org/ref/epsg/nad83-utm-zone-11n/> |
| [18] | Map Town. (2001) Canadian Topographic Maps - British Columbia. [Online]. <http://www.maptown.com/canadiantopographical/bcntsindex.html> |
| [19] | 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> |

x