Coding for Sustainability: Optimizing for Food Self-Reliance in the South West BC

by

Rachel Moizeau Surtshin

A KEYSTONE PROJECT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

Bachelors of Arts and Science

at

Quest University Canada

(Squamish, British Columbia)

Question: *How Can Big Data Inform Sustainability?*

April, 2018

**1.** **Introduction**

Preface

In the age of information we find ourselves with abundance data and an abundance of problems to solve. My question, *how can big data inform sustainability,* asks how we might create useful tools for meaningfully interpreting data in order to ameliorate 21st century dilemmas. The dilemmas I am interested in are those that pose threats to the life of future humans, problems like climate change, limited resource availability, and threats to food security. In the sense that my goal is the minimization of future human suffering, I aim to work towards the development of sustainable solutions to modern problems. Sustainability has a myriad of definitions. The definition widely accepted by WHO is defined as DEFINE.

While designing my keystone, I looked for a project that would allow me to practice computer and data science skills while working on a sustainability project. I wanted my finished product to be something needed and useful. **My keystone project turned an established optimization model of the local food system for increased food self-reliance into a functional computer program.** The project required the use and development of computer programming and data management skills in order to investigate the role that resilient local food systems might play in increasing regional food security.

Food Security and Food Self-Reliance

* What is the problem?
* How does food self-reliance connect to sustainability

“There is growing awareness that climate change, economic instability, resource limitations and population growth are profoundly impacting the capacity of the contemporary global food system to meet human nutrition needs. Although there is widespread recognition that food systems must evolve in the face of these issues, a polarized debate has emerged around the merit of global-verses-local approaches to this evolution.” (Caitlin’s Thesis pg ii)

Southwest British Columbia Bio-regional Food System Design Project

In 2011 a team at the Institute for Sustainable Food Systems at Kwantlen Polytechnic University built a mathematical model of the food system in South West British Columbia, called the Southwest British Columbia Bio-regional Food System Design Project. The Southwest British Columbia Bio-regional Food System Design Project used national and provincial data sets to model and optimize for food-self reliance SOURCE. The aim of the project was to develop “methods to assess current (2011) status and model future (2050) capacity for land based **food self-reliance** in a diet satisfying nutritional recommendations and food preferences that accounts for seasonality of crop production, and comparing self-reliance in livestock raised with and without locally produced feedstocks.”[[1]](#footnote-1) Food self-reliance is defined as the ability of an area to satisfy food needs with food grown locally SOURCE.

The model was built to serve as a tool to inform policy decision as well as a quantitative basis for advocating for the economic development of resilient local food systems. The model was developed into a report called *The Future of Our Food System*, that was well received by on both the regional and municipal level. MORE ABOUT WHAT IT IS. There has been interest from other regions, namely the Okanagan Region of British Columbia, to apply the model in other locales.

The model was built using Microsoft Excel and Open Solver. While Microsoft Excel is accessible to a wide variety of users the sheer depth of the model made a conglomeration of Workbooks that were complex and unwieldy. Modifying, comparing, and updating data sets is tedious, if not impossible. The modification of a single cell can break the model. While the report was recognized and well received, the model itself was not usable for further research and diverse applications.

My Keystone

For my Keystone I endeavored to transform the model into a usable tool. I undertook a translation of the model from Microsoft Excel and Open Solver into a computer program written in the Python programming language. My goal was to transform the model into something more functional and user friendly. The first step was to collect all the data and write a Python code that cleaned and organized the data before executing the functions of the model. Download instructions for the data sets were created so that data sets could be retrieved in a reproducible manner. The second step was to streamline, simplify, and debug all the code, verifying the accuracy of the program. The third step was to add automated data acquisition features and build a graphical user interface (GUI) with integrated options for selecting different data sets.

The Process

My program was written in the Python programming language using Spyder Integrated Development Environment on the Anaconda Platform. My code made use of Numpy, Scipy, and Pandas libraries as well as the fuzzywuzzy library for fuzzy string matching. Much of my time was spent becoming familiar with the Python language, getting to know the data sets, and understanding the idiosyncrasies of the model.

How The Project Connects to My Question

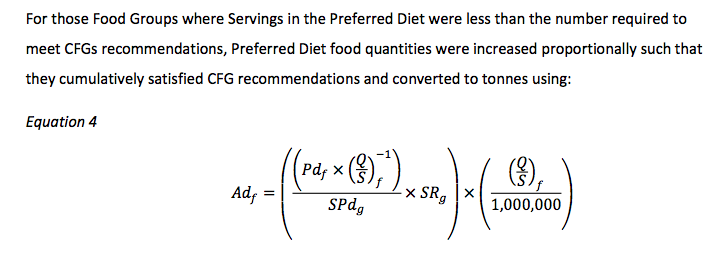
* Why is the project worth doing?
* How does it connect to my academic work at Quest?

Creating a usable tool from a novel academic methodology

**2. Calculating Food Need – Balancing with Food Availability with the Dietary Recommendation**

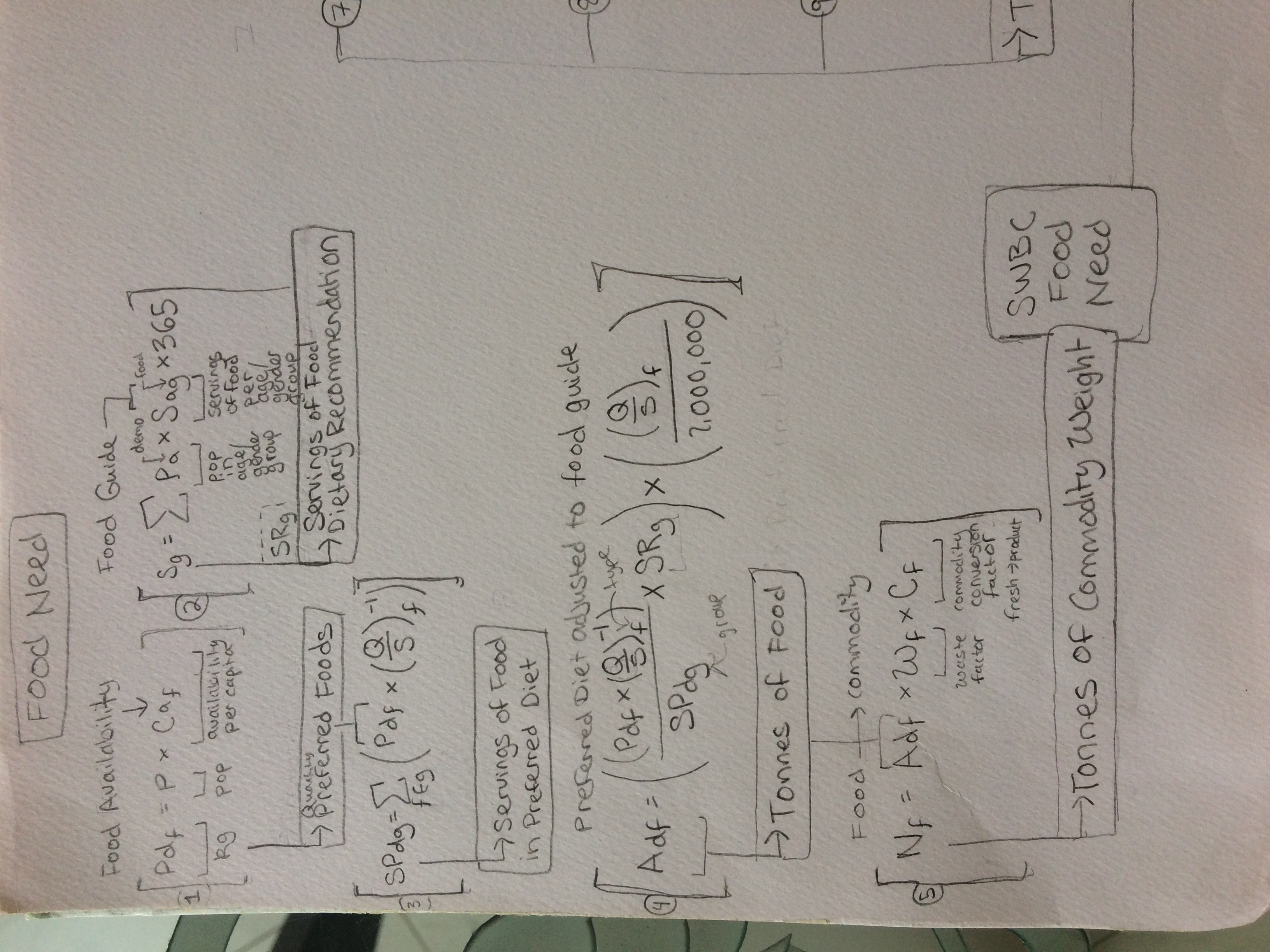
Explain: What is food need, why is it the first step, what data sets were used?

Caitlyn’s thesis addresses Food Need in pages 9-12. On page 11 she balances food availability with food availability with dietary recommendation and coverts servings to tonnes as following:



* Pdf  - **Kg** of food in the preferred diet
* SPdg – **Servings** of food in the preferred diet
* (Q/S) – **kg/serving** (the value is available as kg/serving but we need it as serving/kg thus we put it to the power of -1 to get the inverse)
* SRg – **Servings** of food in the preferred diet

I’ve visualized this equation in the context of the overall calculation of food need as such:

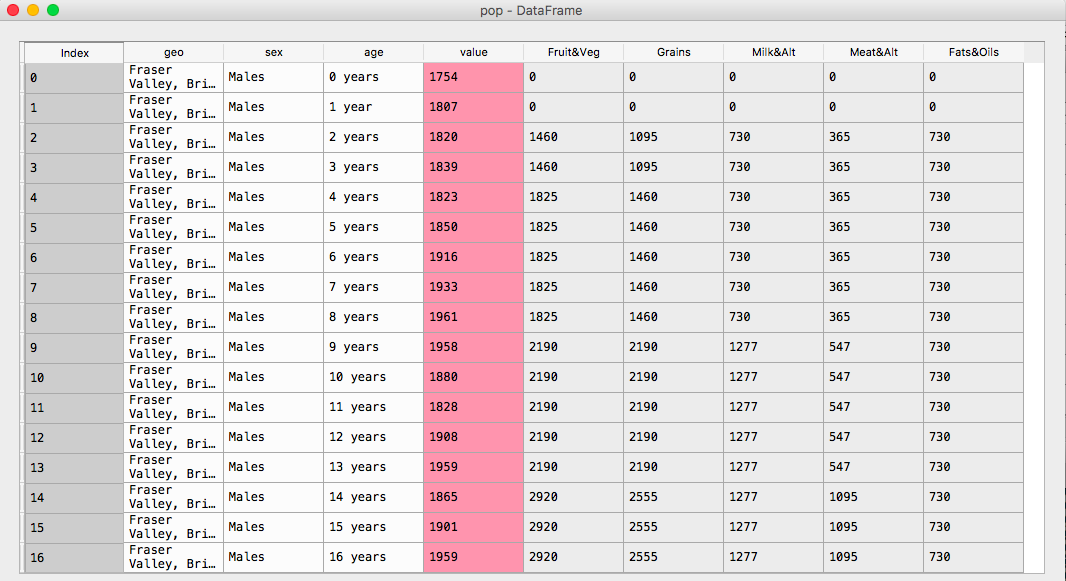
****

In this step we want to:

1. Find out how much more the dietary recommendation per food group per age group differs from the actual food availability.
2. When the recommendation exceeds the availability we want to adjust the food availability values up proportionally to this difference.
3. Availability of foods in kg must be converted to servings to compared to the recommendation, converted back to kg, then to tonnes.

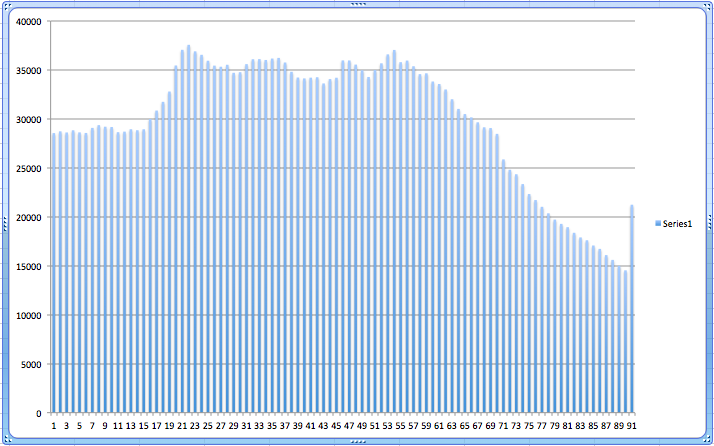
Here is how I did it. First I converted all the availability data from **kg to servings** by multiplying by the reference value stored in a units table.

Then I downloaded the population data and made an excel file containing the dietary recommendation for each age group. I manually multiplied the recommendations by 365 to get the years recommendation per age group. Then I populated my population table with the yearly recommendations for each age year like so:



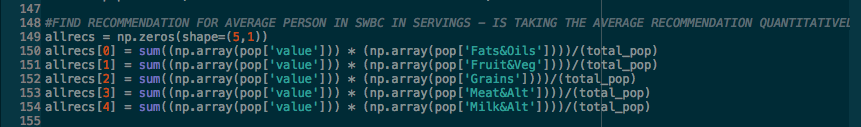
* Value –number of people in the age/region/gender group
* All other columns are the recommended yearly servings for each food group people in that age/region/gender group.

But we need to compare to the food availability, which is for the **average individual** in Canada. I **assumed a normal distribution** for the population data (I will need to verify this quickly using R). To find the total number of servings of each food group recommended in SWBC, for each food group I multiplied the value column by the recommended servings column and divided by total population of SWBC (the sum of the values column). This gave me the dietary recommendation by food group for the **average person in SWBC.**



I made a histogram of the population of SWBC and the distribution looked normal enough for me to feel ok taking the average. A secondary check would be to run a Chi-square test. DO THIS, BETTER.

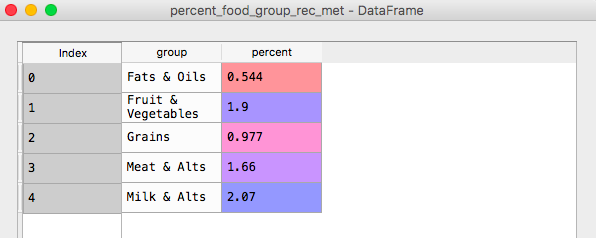
In the code this looks like:



The result was then stored in a table called **allrecs,** a list of the dietary recommendations for the five food groups for the **average person on SWBC.**

Then I needed to find out what percent of the dietary recommendation was met by the food availability. I summed the values of available food in **servings** by group, giving me five values for the availability of the food in each group.

Then I divided the recommendation/availability to get the percent of the food group recommendation met by the food availability:



We can see that the recommendation exceeds availability for Fruits & Vegetables, Meat & Alt, Milk & Alt. All foods in these categories will need to be adjusted upwards. So I just multiplied the **food availability in kg** by any percentage >1 that corresponds to that foods group. So the availability of all Fruits and Vegetables was multiplied by 1.9, etc.

Then, I multiplied the balanced food in **kg** by 1000 to get **tonnes**. Then, I multiplied this by the **waste factor** and **commodity conversion** factor to get **food need** as a new value stored next to original availability in a table of all foods in SWBC.

**3. Calculating Yield Data**

* Fuzzy string matching - explanation

**4. Calculating Livestock Yield**

**5. Optimizing for Food Self-Reliance**

**6. Future Steps**

**7. Conclusion**

**8. Appendix**

* Parts 1: Food Need

Food Need Data Download Instructions

Download the Folder ‘Model Data’ and save it to an accessible location. Open the folder. Inside you should see files called ‘units.csv’ and ‘dietaryrec.csv’

Follow instructions here for setting your Global/Current Working Directory in Spyder. Set your working directory to the ‘Model Data’ folder.

Food Availability Data

1. Go to <http://www5.statcan.gc.ca/cansim/a01?lang=eng>
2. Type **0020011** into Search CANSIM, click Search
3. Go to Add/Remove Data tab
   1. Step 1- Select: Geography – check **All**
   2. Step 2- Select: Food categories – check **Food available adjusted for losses**
   3. Step 3- Select: Commodity – check **All**
   4. Step 4- Select: Time Frame – choose 1 year for both from and to categories
   5. Step 5- Select: Select the Screen output format – choose **HTML table, time as columns – normal retrieval**
   6. Step 6- Click **Apply**
4. Navigate to Download tab
   1. Language: **English**
   2. Data output Format Type: **for database loading**
   3. Select the file format: **CSV**
   4. Series details: **normal retrieval**
   5. Click **Download data**
5. When prompted by ‘Alternative format - CSV document’ page click the first link to **Download file from Cansim (CSV Version, …kb)**
6. Navigate to the downloaded file in your downloads folder. Open in, click Save As, save to the ‘m1.data’ folder and change file name to **cansim0020011.csv**

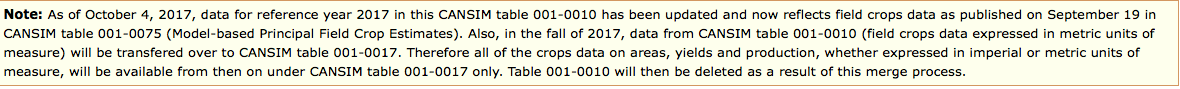
Population Data

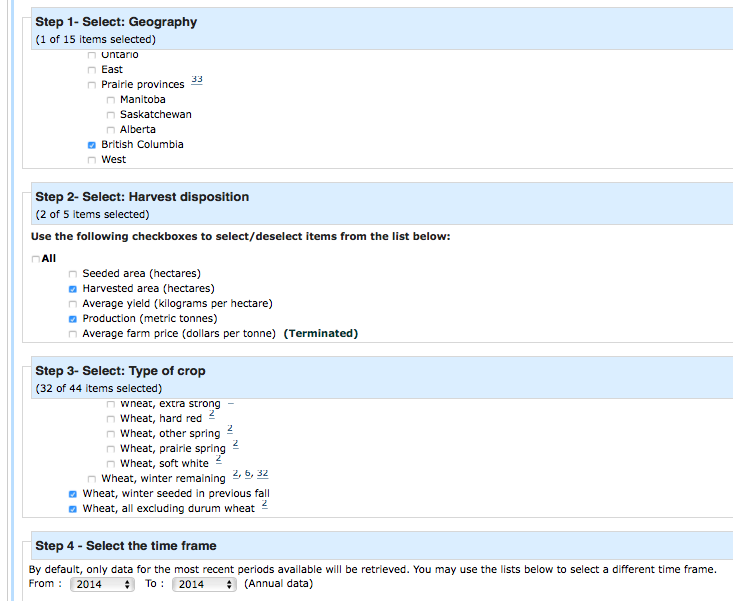
1. Go to <http://www5.statcan.gc.ca/cansim/a01?lang=eng>
2. Type **0510062** into Search CANSIM, click Search
3. Go to Add/Remove Data tab
   1. Step 1- Select: Geography – uncheck everything then scroll down to the British Columbia section. Check **Fraser Valley, British Columbia [**[**5909**](http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVDPage1&db=imdb&dis=2&adm=8&TVD=116940)**], Greater Vancouver, British Columbia [**[**5915**](http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVDPage1&db=imdb&dis=2&adm=8&TVD=116940)**] Powell River, British Columbia [**[**5927**](http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVDPage1&db=imdb&dis=2&adm=8&TVD=116940)**], Sunshine Coast, British Columbia [**[**5929**](http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVDPage1&db=imdb&dis=2&adm=8&TVD=116940)**], Squamish-Lillooet, British Columbia [**[**5931**](http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVDPage1&db=imdb&dis=2&adm=8&TVD=116940)**]**
   2. Step 2- Select: Sex – unselect All and Both sexes and check the boxes in front of  **Males and Females**
   3. Step 3- Select: Unselect All and All Ages. Click the check box in the very upper right corner to select **every age**. None of the brackets such as 0 to 4 should be selected **only the individual years**.
   4. Step 4- Select: Time Frame – choose 1 year for both from and to categories
   5. Step 5- Select: Select the Screen output format – choose **HTML table, time as columns – normal retrieval**
   6. Step 6- Click **Apply**
4. Navigate to Download tab
   1. Language: **English**
   2. Data output Format Type: **for database loading**
   3. Select the file format: **CSV**
   4. Series details: **normal retrieval**
   5. Click **Download data**
5. When prompted by ‘Alternative format - CSV document’ page click the first link to **Download file from Cansim (CSV Version, …kb)**
6. Navigate to the downloaded file in your downloads folder. Open in, click Save As, save to the ‘m1.data’ folder and change file name to **cansim0510062.XXXX.csv** where XXXX is the digits of the year the population data is from.
7. Go to line \_ in the code and change the ‘cansim0510062.XXXX.2.csv’ to match the exact name of the file in your working directory.
   * Code

* Part 2: Yield Data

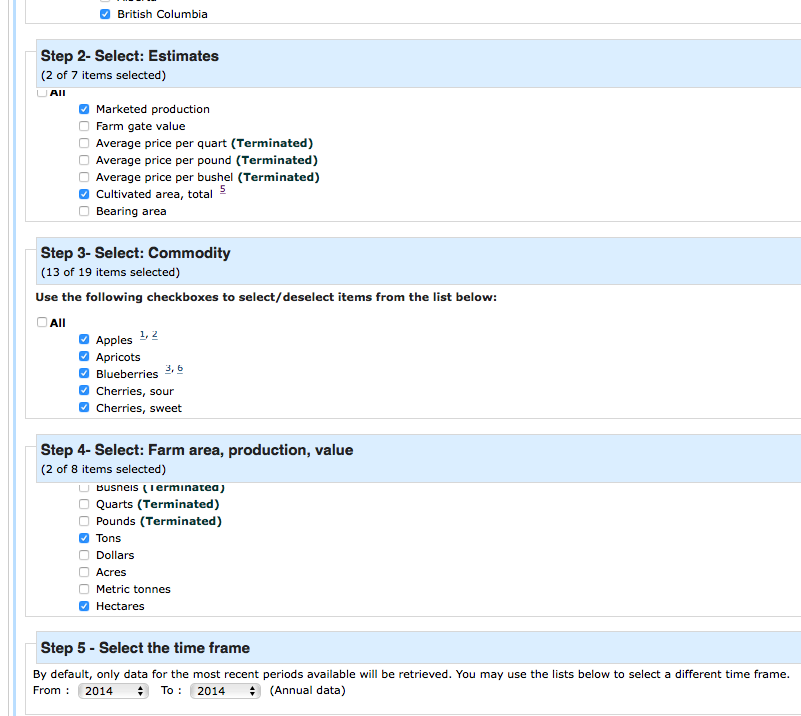
**Yield Data Acquisition**

1. **CANSIM Table 001-0010 ("production"/"seeded area") (Statistics Canada, 2014)**

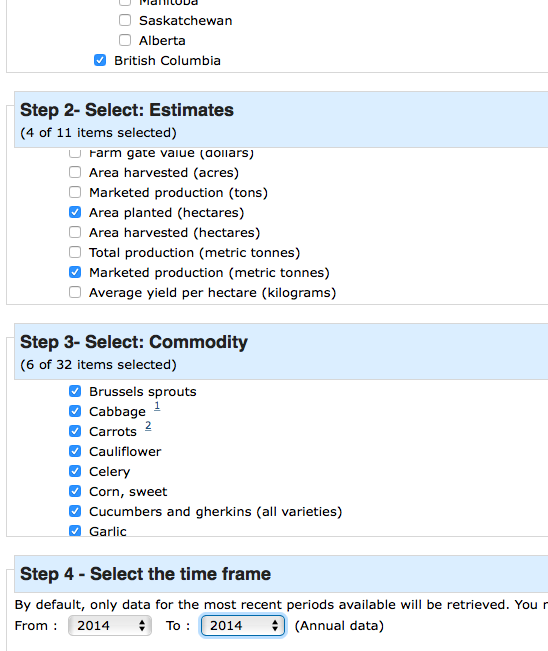


****

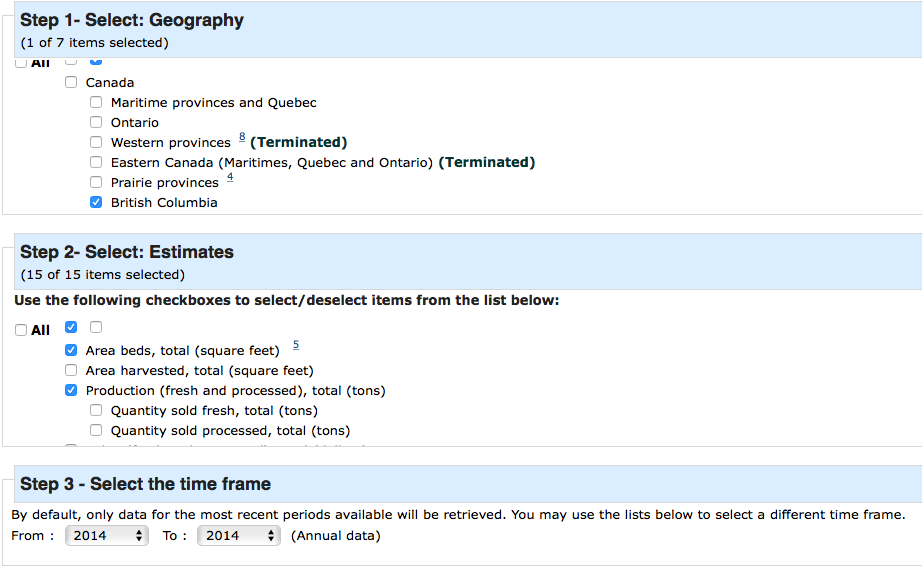
1. Download the file for database loading (as always)
2. Navigate to the downloaded file in your downloads folder. Open in, click Save As, save to the ‘m1.data’ folder and change file name to **cansim0010010.XXXX.csv** where XXXX is the digits of the year the population data is from.
3. Go to line \_ in the code and change the ‘cansim00010010.XXXX.csv’ to match the exact name of the file in your working directory.
4. **CANSIM Table 001-0009 ("marketed production"/"cultivated area"), (Statistics Canada, 2014)**

****

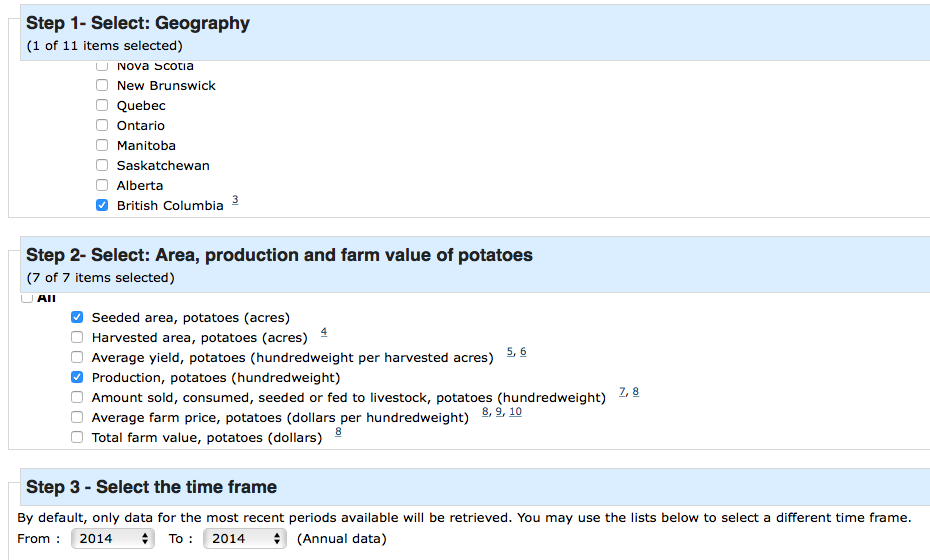
1. Download the file for database loading (as always)
2. Navigate to the downloaded file in your downloads folder. Open in, click Save As, save to the ‘m1.data’ folder and change file name to **cansim0010009.XXXX.csv** where XXXX is the digits of the year the population data is from.
3. Go to line \_ in the code and change the ‘cansim00010009.XXXX.csv’ to match the exact name of the file in your working directory.
4. **CANSIM Table 001-0013 ("marketed production"/"seeded area"), (Statistics Canada, 2014)**

****

1. Download the file for database loading (as always)
2. Navigate to the downloaded file in your downloads folder. Open in, click Save As, save to the ‘m1.data’ folder and change file name to **cansim0010013.XXXX.csv** where XXXX is the digits of the year the population data is from.
3. Go to line \_ in the code and change the ‘cansim00010013.XXXX.csv’ to match the exact name of the file in your working directory.
4. **CANSIM Table 001-0012 ("production, fresh and processed"/"area beds total"), (Statistics Canada, 2014)**

****

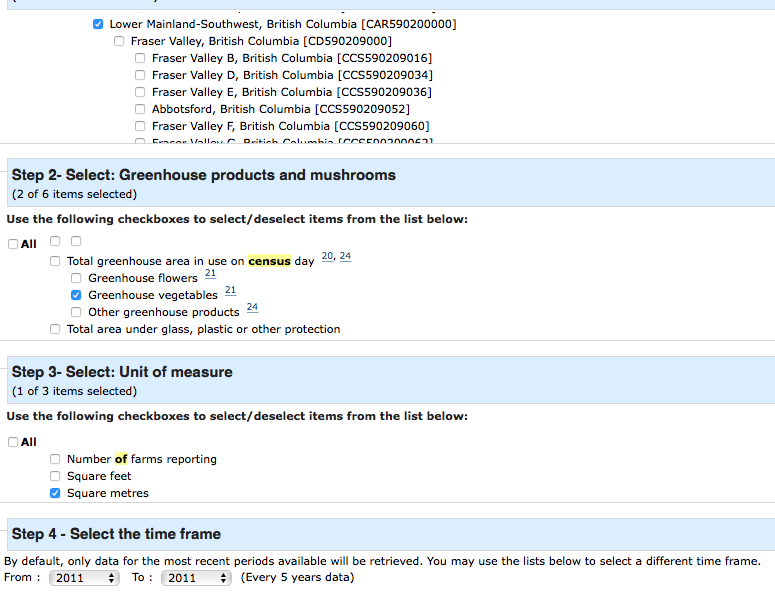
1. **CANSIM Table 001-0014 ("marketed production"/"seeded area"), (Statistics Canada, 2014)**

****

1. **CANSIM Table 001-0006, (Statistics Canada, 2014)**

**SWBC Area**

**CANSIM Table 004-0217 Census of Agriculture, greenhouse products and mushrooms**

****

* + Code
* Part 3: Livestock Data
  + Download Instructions
  + Code
* Full tutorial
* Notes

1. (Caitlin’s Thesis pg ii) [↑](#footnote-ref-1)