Coding for Sustainability: Food Self-Reliance in the South West British Columbia

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

Introduction

My keystone project turned an established model of Food Self-Reliance in Southwest British Columbia 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 local food systems might play in increasing regional food security.

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 uses national and provincial data sets to model for Food Self Reliance (Mullinix et al., 2016), which is “defined as the ability to satisfy local food need with food grown locally”(Mullinix et al., 2016).



Figure 1: Map of Southwest British Columbia bioregion, comprising of the Metro Vancouver, Fraser Valley, Squamish-Lillooet, Powell River, and Sunshine Coast sub regions. Source: (Mullinix et al., 2016)

The model was built to serve as a policy tool as well as a quantitative basis for advocating for the economic development of resilient local food systems. 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”(Dorward, 2015). 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. 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 Excel Workbooks that were complex and unwieldy. Modifying, comparing, and updating data sets was tedious, and could only be accomplished by staff members intimately familiar with the model. The modification of a single cell can break the model. While the report was recognized and well received, the model itself is not usable for further research and diverse applications.

My Keystone

Keystone 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.

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.

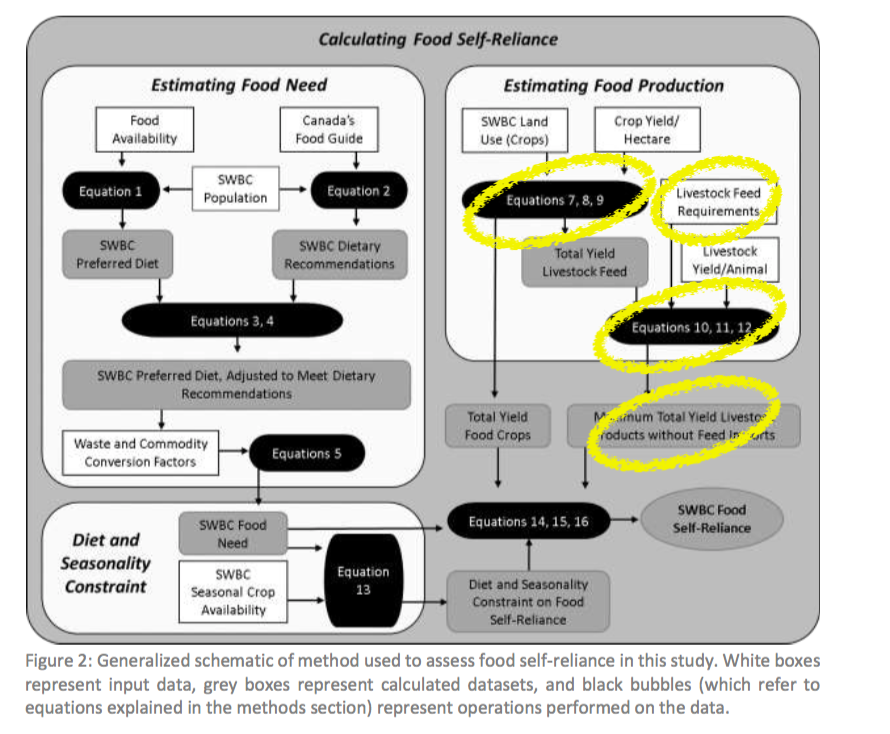
While designing my keystone, I looked for a project that would allow me to practice computer and data science skills while working on a sustainability project. I wanted my finished product to be something needed and useful, but also something that addressed real world issues.

“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” (Dorward, 2015).

The purpose of this project is to generate the science that will inform this debate, and the subsequent evolution of our food system.

**2. Methods**

The following methods, figures, and equations rely heavily on work presented in Caitlin Dorward’s thesis (Dorward, 2015) that formed the foundation of the Southwest BC Bio-regional Food System Design Project.



(Dorward, 2015)

**2.1 Calculating Food Need**

The first step in calculating Food Self-Reliance is calculating the Food Need, the food requirements of the population living in Southwest British Columbia. To do this we used CANSIM Food Availability Data (Government of Canada, 2017g) for the food requirements (kg of food per person) for the average Canadian. We omitting missing data and ‘foods’ that were deemed inessential diet, such as alcoholic beverages, or lacked the necessary conversion data. The quantity of food available per capita was obtained using:

*Equation 1*

𝑃𝑑𝑓 =𝑃×𝐶𝑎𝑓

Where 𝑃𝑑𝑓 is the Preferred Diet (kilograms); 𝑃 is the population of SWBC; 𝐶𝑎𝑓 is the quantity of foods (kilograms per capita).

The project designers at Kwantlen designed the model so that Food Self-Reliance was not only calculated based on the current dietary patterns of the population but also based on the dietary recommendation laid out in Canada’s Food Guide (Canada & Canada, 2007). We balanced the food availability data with the dietary recommendation. In food groups where the dietary recommendation exceeded food consumption we increased the consumption data relative to that difference. In food groups where the dietary recommendation did not exceed food consumption, we left the food consumption as it was.

Using population data from the census of districts in Southwest BC (Government of Canada, 2017m) we binned the actual population data by the age groups used in the dietary recommendation. Then we multiplied the yearly dietary recommendation by the number of people in each age group then dividing the sum of the recommendations by the total number of people in Southwest BC. Thus, we found the Dietary Recommendations (servings) for the average person in Southwest BC:

*Equation 2*

Where 𝑆𝑔 is the annual quantity (servings) from Food group g required by the Dietary Recommendation; a is every age group in the dietary recommendation; is the SWBC population in every age group; 𝑆𝑎𝑔 is the recommended quantity of food (servings) required per age group and food group given in the Dietary Recommendation; is the total population of Southwest BC.

In order to compare the dietary recommendation to the availability data, we needed to express the availability data in (servings). Because the availability data was given in (kg/person) we multiplied by the number of (servings/kg) factor to get (servings/person). By binning the food availability data by food group we found the total food availability by group in servings. We obtained the Preferred Diet in servings as such:

*Equation 3*

𝑆𝑃𝑑𝑔=

Where 𝑆𝑃𝑑𝑔 is the quantity of food in the preferred diet (servings) by food group; is the quantity of food in the preferred diet (kilograms), (Q/S) is the quantity of food (kilograms) per serving for each food in the preferred diet.

We expressed how much of the dietary recommendation was met by dividing the recommendation by the food availability. To balanced the availability to the recommendation we multiplied the amount of each food (kilograms) by the percent of the recommendation met by that food’s group (when that percent of recommendation met was >1).

*Equation 4*

Where is the quantity of food (kilograms) in the Preferred Diet, balanced to the Dietary Recommendation. To calculate Food Need, this value was then converted from kg to tonnes, multiplied by the waste factor, and the commodity conversion factor.

*Equation 5*

Where is the Food Need (tonnes); is the waste factor for each food; is the commodity conversion factor for each food. Thus, we reached the final result for Food Need (tonnes of commodity).

Proof of Methods, Own Questions

In keeping with the methods of the Institute for Sustainable Food Systems, I balanced the food availability data with the dietary recommendation. This was done to prove that the methods my program employed produced the same results as the model created by the Institute for Sustainable Food Systems.

Once I was satisfied that my methods produced results consistent with the original model I started to ask some of my own questions. I calculated Food Need without balancing the availability to the dietary recommendation. This was done to explore the different results obtained when we include “what Canadian’s should eat” into the equation. By balancing to the dietary recommendation we create a result for Food Need that has a net excess, since we have only increased the amount of foods in groups which Canadian’s consume less than the recommendation and not decreased the amount of foods in groups which Canadian’s consume less than the recommendation. For this reason I was curious to calculate Food Self-Reliance as it was in 2011, based on what people in Southwest BC were *actually* eating.

My calculation of Food Need, without balancing to the Dietary Recommendation was done as such:

This is the equivalent of skipping Equations 2-4 and calculating Food Need directly from , the Preferred Diet. Thus, my final result expresses the different results for Food Need, and subsequently Food Self-Reliance, with and without balancing the dietary recommendation.

**2.2 Calculating Food Production**

The next step in calculating Food Self-Reliance was to calculate Food Production in Southwest

BC. This was accomplished separately for crops and for livestock commodities. For crops this

was calculated by multiplying the British Columbia crop yields (Government of Canada, 2017a, 2017b, 2017c, 2017d, 2017e, 2017f, 2017g) by census land use data for Southwest BC (Government of Canada, 2017i, p. 004, 2017j, 2017k, 2017l). BC crop yield data was obtained by dividing the amount of production (tonnes) by the area planted (hectares). Conversions were necessary to obtain these units. This data was specific to BC and so certain yields (such as fruit crops, whose high yields are due to the drier warmer inland climate of the Okanagan) were modified[[1]](#footnote-1) based on consultation with agricultural scientists. Southwest BC Food Production for

*Equation 7*

𝑇𝑌𝑟𝑐= × 𝐴𝑟𝑐

Where 𝑇𝑌𝑟𝑐 is the total SWBC production of each food commodity; is the provincial production (tonnes) of commodity produced, is the total provincial area of commodity planted (hectares), 𝐴𝑟𝑐 is the SWBC regional area of the commodity planted (hectares).

*Equation 8*

𝑇𝑌𝑟𝑐= ×

Since greenhouse area in SWBC was not given by specific greenhouse commodity we divided the total SWBC greenhouse area bythe number of greenhouse crops. We assumed the greenhouse area was equally portioned for these, in this case three, commodities.

*Equation 9*

𝑇𝑌𝑟𝑐= ×

Since canola oil is made from crushed canola seed we multiplied by the yield of oil per crushed seed, where is the amount (tonnes) of oil produced and is the amount (tonnes) of seed crushed.

Proof of Methods, Own Questions

In keeping with the methods employed by the Institute for Sustainable Food Systems I used 2011 data for the crop yields. This was done to prove that the methods my program employed produced the same results as the model created by the Institute for Sustainable Food Systems. Once I was satisfied that my methods produced results consistent with the original model I started to ask some of my own questions. I used both 2011 Baseline Yields and 10 Year Average yields to see how much fluctuations in yearly yield affected Food Self-Reliance. To calculate 10 year average crop yield:

Results using 10 year average yields are presented alongside results using baseline 2011 yield.

Livestock Food Production

To calculate livestock production I used census data about the number of livestock in Southwest BC (Government of Canada, 2017h, p. 004). I multiplied the number of livestock by static data about the tonnes of commodity produced per animal to get the final amount of livestock commodity (tonnes):

*Equation 10*

Where is the tonnes of livestock commodity produced in Southwest BC; is the tonnes of commodity produced per animal;is the number of head of each type of livestock in Southwest BC. The head of livestock data was unspecific about which types of cows and chickens the numbers of breeding stock contributed to. For this reason I added in half of the possible breeding stock to the number of head of the livestock types in question.

*Equation 11*

Where is the total SWBC yield of livestock commodity; is the yield of livestock product per animal adjusted to take into account breeding stock; is the number of head of livestock in SWBC used for animal commodities; is the number of head of breeding livestock in SWBC used for animal commodities.

Merging the yield data with the land area data was challenging, since small differences in the crop names (expressed as strings in the computer program) made a simple merge of the tables impossible. For this reason I used fuzzy string matching to match the strings before merging the tables.

**2.4 Calculating Diet and Seasonality Constraint**

**2.5 Calculating Food Self Reliance**

Diet and seasonality constraint was calculated for each commodity based on the number of months of the year that the commodity could be produced in Southwest BC. The diet and seasonality constraint was calculated by dividing the Food Need of the commodity (tonnes) by 12 and then multiplying by the number of months out of the year that the food could be produced in Southwest BC.

**3. Results**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Fruits & Vegetables | Grains | Meat & Alts | Milk & Alts | Fats & Oils | Total Food Self Reliance |
| ISFS | 21% | 1% | 49% | 86% | 1% | 40% |
| Rachel – balanced | 22.40% | 2.70% | 46.40% | 83% | 0.22% | 34.70% |
|  |  |  |  |  |
| Rachel – without balancing | 41.00% | 69.90% | 100% | 36.94% |

**Figure 1:**

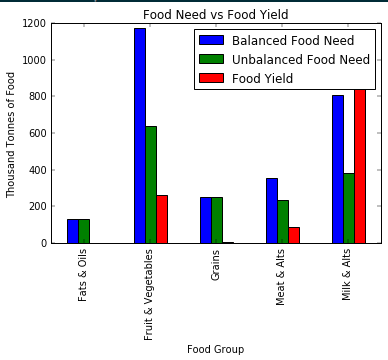
****

Figure 1: Above we see the comparison of Southwest BC Food Need with the Southwest BC Food Yield. Food Need is shown both Unbalanced (not taking into account the dietary recommendation) and Balanced (taking into account the Dietary Recommendation).

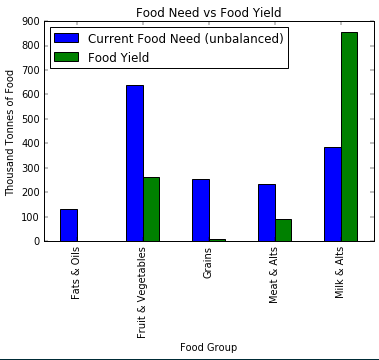
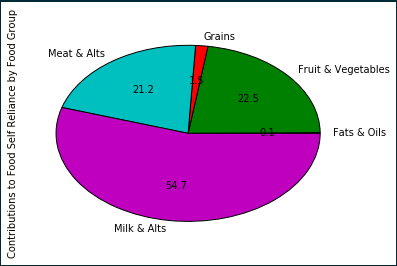


Figure 2: Above we see the comparison of Southwest BC Food Need (based on current Canada Food Availability data and unbalanced to the dietary recommendation) with the Southwest BC Food Yield.

  
Figure 3: Above we see the contributions to Food Self Reliance made by each Food Group

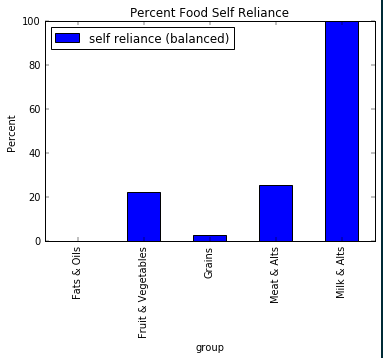


Figure 4: Above we see the percent Food Self-Reliance of each food group. Add ‘Total Self Reliance’ column.

**4. Sources of Uncertainty**

Sources of uncertainty were introduced during multiple phases of the model. In the Food Need section of the model error was introduced when foods from the CANSIM Food Availability Data (DATA SOURCE) were omitted because of lack of information.

|  |
| --- |
| **Foods Omitted from Availability Data** |
| Fresh and frozen sea fish edible weight, Freshwater fish edible weight, Processed sea fish edible weight, Total sea fish edible weight, Sugar, Honey, Shellfish, Dry beans, Tea, Coffee, Cocoa, Ale beer stout, Distilled Wines, Soft Bottled Cider, Distilled Spirits, Sweetened concentrated skim milk, Milkshake, cream, Sherbet, Ice milk, Chocolate drink, Other fresh berries, Melons musk and cantaloupes fresh, Other fresh melons, Watermelons, Wintermelons, Quinces, Fruits not specified, Oranges fresh, Lemons fresh, Grapefruits fresh, Limes fresh, Mandarins fresh, Other citrus fresh, Chinese cabbage, Other edible roots fresh, leguminous vegetables, Olives fresh, Potatoes chips, Potatoes total, Potatoes white fresh and processed, Vegetables not specified, Guavas, Tree nuts, Melons total fresh, Kiwis, Nectarines, Artichokes, Eggplant, Kohlrabi, Garlic, Leeks, Okra, Parsley, Parsnips, Rappini |

The omitted data contained values in the availability data set (kg/person) but the lack of information about the waste factors, commodity conversion factors, and seasonal availability associated with these foods meant that they could not be used to calculate Food Need. For these reason we are unable to know what portion of the SWBC Food Need these omitted crops contributed to. We are able to see that they come to a total of (584 kg/person) compared to the (336.86 kg/person) of food included in the food need. This doesn’t tell us much. The astronomical difference is easily comprehended by the large amounts of liquid ‘foods’ omitted (for example 88 kg/person of coffee). The large differential between the final weight of the product (drinkable coffee) and the weight of the agricultural products required to produce the food (several tablespoons of ground beans) explains why the omitted foods comprise such a large portion of the dataset.

Omitting liquid ‘foods’ gives us a total of (96 kg/person) of omitted foods, a much more reasonable number. The issue brings into question a decision made on the part of the model designers about what constitutes ‘food’. Products like alcoholic beverages and soft drinks might make up a calorically substantial part of the Canadian diet, but that might not qualify them for inclusion in the model.

The calculation of 2011 SWBC Food Self Reliance was based of 2011 British Columbia crop yields (DATA SOURCE). My final result employed this method, however it is useful to know how much the result is based off of the specific growing conditions of 2011[[2]](#footnote-2). To test this, I ran the model using 10 year average crop yields. Though data was available through 2017, I used 2002-2013 data in keeping with the methodology developed by IFSFS.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **10 year average yields** | Fruits & Vegetables | Grains | Meat & Alts | Milk & Alts | Fats & Oils | Total Food Self Reliance |
| ISFS | 21% | 1% | 49% | 86% | 1% | 40% |
| Rachel – balanced | 23.77% | 4.57% |  |  | 0.22% | 34.78% |
|  |  |  |  |  |
| Rachel – without balancing | 43.63% |  |  | 36.98% |

Using the ten year average affected the Food Self Reliance for the Fruits & Vegetables and Grains category, but had little effect on the overall calculation of Food Self Reliance.

To test the affect of varying crop yields on Food Self Reliance I ran the model using the ten year average crop yields plus and minus one standard error.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Average Yield**  **-1 Standard Error** | Fruits & Vegetables | Grains | Meat & Alts | Milk & Alts | Fats & Oils | Total Food Self Reliance |
| ISFS | 21% | 1% | 49% | 86% | 1% | 40% |
| Rachel – balanced | 22% | 3.34% |  |  | 0.22% | 34.60% |
|  |  |  |  |  |
| Rachel – without balancing | 41.00% |  |  | 36.80% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Average Yield**  **+1 Standard Error** | Fruits & Vegetables | Grains | Meat & Alts | Milk & Alts | Fats & Oils | Total Food Self Reliance |
| ISFS | 21% | 1% | 49% | 86% | 1% | 40% |
| Rachel – balanced | 25% | 5.79% |  |  | 0.22% | 34.90% |
|  |  |  |  |  |
| Rachel – without balancing | 46.00% |  |  | 37.10% |

While the change was somewhat pronounced in the Fruits & Vegetables and Grains categories, there was minimal change to the final calculation of food Self Reliance. It would have been interesting to include this same methodology for the data about the area of each planted crop in SWBC, but unfortunately this district specific data comes from the census and the data sets were only available for 2011 and 2016.

Another source of uncertainty came in the delination of breeding stock to different types of livestock. The SWBC census data on the number of livestock gives the numbers of breeding stock without specifying if they are for beef or dairy cows, meat chickens or laying chickens.

|  |  |
| --- | --- |
| **Beef and Dairy Breeders** | **Chicken and Laying Hen Breeders** |
| Bulls, 1 year and over (890)  Calves, under 1 year (27,373)  Total heifers, 1 year and over (28,373)  **Total: 56,636 head of livestock** | Layer and broiler breeders (pullet and hens) (876,370)  Other poultry (447,756)  **Total: 1,324,126 head of livestock** |

The delineation of the breeding stock between the shared categories of livestock became a large source of uncertainty. I manipulated the delineation of breeding stock by extremes to measure the uncertainty in my calculations. By manipulating the attribution of breeding stock from a 50-50 baseline I found that the attribution of breeder stock to Beef and Dairy had a large impact on the final calculation of Food Self Reliance. Attributing 100% of the breeding stock to Beef instead of Dairy greatly reduced the Food Self Reliance for the Milks and Alternatives category and for the Total Food Self Reliance. Attributing 100% of the breeding stock to Dairy greatly increased the Food Self Reliance for the Milks and Alternatives category and for the Total Food Self Reliance. Changing the delineation between the breeding stock of meat chickens and laying chickens had little to no effect on the calculation of Food Self Reliance in the Meats and Alternatives category or on Total Food Self-Reliance.

To create the upper and lower bounds of uncertainty I compounded both the uncertainty from the standard error of the ten year average yield data and the uncertainty in the delineation of breeding stock to different types of livestock.

For my lower bound I used the average of the ten year yield data minus one standard error and I attributed 100% of the breeding stock to Beef instead of Dairy. Because there was no discernable difference between the delineation of meat chicken and laying hen breeding stock I left that delineation at 50% for each type of livestock.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Lower Bound + Beef(100%)** | Fruits & Vegetables | Grains | Meat & Alts | Milk & Alts | Fats & Oils | Total Food Self Reliance |
| ISFS | 21% | 1% | 49% | 86% | 1% | 40% |
| Rachel – balanced | 22.34% | 3.34% | 46.70% | 59.70% | 0.22% | 28.10% |
|  |  |  |  |  |
| Rachel – without balancing | 41.00% | 70.47% | 100% | 36.80% |

For my upper bound I used the average of the ten year yield data plus one standard error and I attributed 100% of the breeding stock to Dairy instead of Beef.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Upper Bound +Dairy(100%)** | Fruits & Vegetables | Grains | Meat & Alts | Milk & Alts | Fats & Oils | Total Food Self Reliance |
| ISFS | 21% | 1% | 49% | 86% | 1% | 40% |
| Rachel – balanced | 25.10% | 5.79% | 40.40% | 100.00% | 0.22% | 39.49% |
|  |  |  |  |  |
| Rachel – without balancing | 46.20% | 60.40% | 100% | 36.40% |

Thus, I found that my measurement of uncertainty for Total Food Self Reliance fell in the range between 28.1 and 39.49 percent food self-reliance, a range of 11.5%.

There are prob other uncertainties that have not yet been evaluated.

**5. Discussion**

**6. Looking Forward**

To access trends in agricultural land use, a simple linear model could be created to produce year specific data between the available 4 year census data.

**7. Appendix**

**A. All Results By Food**

**B. All Data Sets – data source and acquisition info**

Data Set 1:

CANSIM Table 002-0011, Food available in Canada annual (kilograms per person, per year)

**C. Code**

1. What crops yields were modified for SWBC? [↑](#footnote-ref-1)
2. Besides mushrooms, dry peas, mixed grains, and sour cherries which lacked a yield value for 2011. In these cases I used 10 year averages instead of 2011 values for the final calculation. In the case of mushrooms, there was a substantial amount of missing data so a 1996-2005 average was used. [↑](#footnote-ref-2)