Project 2
Food vs. Driving Carbon Emission Comparison
Team: Kiersten Henderson and Micheline Casey
INFO W18 Kleeman, Thurs 4-5:30

Purpose and Project Overview

There is a nascent, but growing movement to understand the relationship between food production, carbon emissions, and global climate change. Several studies have been done over the last decade to understand these linkages, and the findings are quite clear that our diets impact carbon emissions. Globally, direct emissions from agriculture represent 10% to 12% of overall greenhouse gas (GHG) emissions; when including the impact of fertilizer and chemical production, fuel use, and agriculturally induced land-use change (which carries large uncertainty), the figure rises to 17% to 32% (Bellarby et al. 2008)[1].

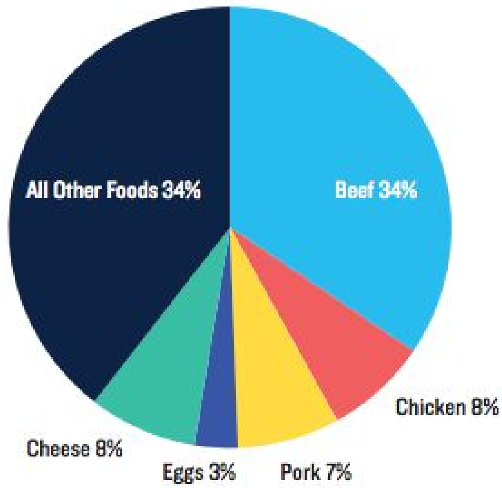
Transportation, housing, and food are the largest contributors to the carbon footprint of an individual U.S. household. Behavioral choices regarding food consumption and transportation useage are particularly useful to study given the potential feasibility and economics of trade-offs. This project will explore data sets from a variety of sources in order to answer some of our questions regarding food consumption, carbon emissions, and possible tradeoffs.

The Focus

We chose to analyze the daily consumption in the United States of four foods: beef, chicken, dairy products, and asparagus. Beef is well-known for producing the highest level of carbon dioxide of all foods, though it's not the most-consumed food. We chose chicken and dairy for their high frequency in U.S. consumers' diets. And we chose asparagus because it is a high carbon-intensity vegetable. The figure below is from the National Resource Defense Council's recent publication, "Less Beef, Less Carbon"[2] , and shows the relative contribution of top greenhouse gas (GHG) intensive foods to total per-capita food-related GHG emissions in 2014.

[1] Greenhouse gas emissions factors are based on a compilation of average of lifecycle analysis values. A life cycle analysis approximates the climate-warming pollution associated with the production of a food item—from fertilizer and pesticides used to grow the crops to transportation and refrigeration. Multiple life cycle analyses for the same food can yield different values if the calculation includes varying growing conditions or different assumptions are made for what to include the measurement. "Greenhouse Gas Emission Estimates of U.S. Dietary Choices and Food Loss", Heller, et.al., Journal of Industrial Ecology, 2014.
[2] "Less Beef, Less Carbon", NRDC, March 2017

FIGURE 4: RELATIVE CONTRIBUTIONS OF TOP 5 GHG-INTENSIVE FOODS AND ALL OTHER FOODS TO TOTAL PER CAPITA FOOD-RELATED GHG EMISSIONS IN 2014[20]



We then calculated the carbon emissions from the consumption of each of these foods per day for the average American, by gender, and for different age groups. We also will analyze miles driven by gender and age in the United States and identify the carbon emissions of each of these groups from driving. Finally, we will examine the relative daily contribution to carbon emissions of eating and driving habits based on these factors (four food groups, gender and age).

In our analyses, we were interested in the following questions:

1. How much do four commonly eaten foods (beef, chicken, dairy, and asparagus) influence one's carbon footprint?
    a. How much of each of these four food categories is consumed daily in the U.S. by gender and by age range;
    b. Are there differences in food consumption by age that may lead to different carbon footprints the older or younger we are?
2. How many auto miles do people commute each day in the U.S.? by gender[3] and age;
3. Are there comparable costs to the environment between consuming some of the highest food carbon producers and driving?

[3] For the study, the gender data only included male, female, and unknown categories.

Context and Data Sources
Our key source data files came from the United States Department of Agriculture (USDA) and the United States Department of Transportation (USDOT).

- FOOD DATA:
  National Health and Nutrition Examination Survey data (**NHANES data**) from a joint Centers for Disease Control and Prevention/United States Department of Agriculture survey.This data source consists of three relevant files.
        - Dietary Interview - Individual Foods, First Day (now called 2013_2014day1data.csv)
        - Dietary Interview Technical Support File - Food Codes (now called Foodcode_2013_2014.csv)
        - Demographic Variables and Sample Weights (now called Demographics_2013_2014.csv)
    - The data were originally available as SAS files and we converted them to CSVs using SAS Universal Viewer. There were 8,661 respondents to the survey as a whole; 7,397 people or 85% of respondents had at least one of our four food categories included in their food diaries. We only analyzed one of the two days in the survey.
- DRIVING DATA:
  We utilized the 1995 Transportation Survey from the United States Department of Transportation, Bureau of Transportation Statistics. This was the most recent U.S. individual survey we could find. Over 95,000 people across the U.S. participated in this year-long survey. (34074258_T_NPTS_PERSON_1995(1).csv)

Additionally, we used the Greenhouse Gas Equivalency Calculator from the Environmental Protection Agency (EPA).

Data Cleansing and Sanity Checks

Food Data
As previously mentioned the food data we used for our analyses came in three separate files, food consumption data, demographics data for the food consumption data, and related food codes. Each needed to be reviewed and checked for a variety of data quality components. We were able to convert the original SAS files from the CDC website into CSV files using a SAS viewer. From there, each file was imported into pandas and a variety of checks were run on the data.

The food consumption file and related demographics files were fairly straight forward review and clean up. Both files had numerous data fields (columns), but there were three key ones from each that were crucial to our analyses. From the food consumption file, we needed the unique person ID (SEQN), the USDA food code number (DR1IFDCD), and the food weight in grams (DR1IGRMS). The demographics data file also had a unique person identifier (SEQN), as well as gender (RIAGNDR) and age (RIDAGEYR). It was quite simple to cut the three fields from each of the two files as our basic building blocks for further analysis. The Gender field was then converted from an integer (0, 1, or 2) to a string ('unknown', 'male, 'female), and the age was mapped to an age range that was the same as the age ranges used in the driving data. These conversions made the readability, manipulation, and analyses of the data much easier.

The foodcode data (food code number, DRXFDCD, and food description,DRXFCSD) was a bit trickier to work with. While also a csv file, there are well over 8000 food codes that individually had to be culled through to identify the correct code ranges for the four food groups that we studied. It was fairly easy to write a function to be able to filter down the total amount of food in the food consumption file to the four food categories in this study. With the dairy food category, however, we decided that we did not care about infant formula, and so needed to do an additional slice to pull out the 200+ infant formula codes that exist within the broad range of dairy codes. We ended up with far more dairy codes that the other studies we reviewed, which could account for the high proportion of dairy results in our analyses. An additional set of manipulation was then done to change the food codes into food types (e.g., Dairy, Chicken, Beef, Asparagus), to simplify further filtering, aggregation, calculations, and analyses (rather than continuing to deal with several hundred food codes).

Driving Data

The driving data was fairly straightforward to clean up and sanity check. To analyze this data we needed to parse out values for gender, age, and yearly mileage. The gender data behaved as expected with only two sexes, and men and women being surveyed in equal numbers (~95, 000 people total).

In terms of the age data, the survey includes many parameters regarding transport, and therefore surveyed children who were not yet of driving age. It wasn't clear initially if they would have a yearly mileage reported (for miles they were driven places by adults), or if entries for these individuals would be NaN. This was resolved when examining the yearly mileage data.

A surprise came when we initially calculated the annual yearly mileage and got 320 000 miles per year per person. Upon inspection of the data, it became clear that for ~30, 000 people of ~95, 0000 people surveyed, yearly mileage was 999 999 miles (or a number in the 999 990's). Not only is this highly improbable, but the documentation for the data set indicates that yearly mileage is capped at 200 000 miles. Further probing revealed that when a value is not recorded in Department of Transportation datasets, it is replaced by a number beginning with 9. We therefore excluded individuals with spurious mileage values from our analysis and examined yearly mileage for the remaining ~66, 000 individuals. This lead to the more reasonable yearly annual mileage of 12 421, which is comparable to previous calculations. We were able to easily convert yearly mileage to daily mileage and from there, convert daily mileage to daily Kilograms of CO_2 emissions from driving by multiplying by the EPA's estimate of Kg CO_2/ mile driven by a typical passenger vehicle (0.4 Kg/mile).

In our initial proposal, we planned to map average daily driving distance to location by state on a heat map of the US. Another surprise and disappointment came upon close examination of the Census Division and Census Region columns that we thought we'd use to derive location by state. We initially thought that combining the numbers listed in these two columns would allow us to uniquely identify the states of driver residence, however this was not the case. Instead, such an analysis would only narrow state to groups of states as large as 9 states and was not granular enough to use. However, if we had useable data, it would be very straightforward to plot Average Daily Miles by state using Open Heat Map.

Data Processing and Challenges

Food Data

As described in the previous section, there were some pretty straightforward conversions of items such as gender code and age to make the more advanced processing and analytics that we were doing easier for us. After this basic processing was done, the abridged versions of the food consumption and demographics data files were merged and the advanced calculations, manipulations, and analytics could be run. A variety of groupbys and sorts were done to slice the data into a variety of views based on food type, gender, and age.

When we started grouping by gender and age range , we ran into a problem. The double groupby generated multi-indexed data frames, with which we had no experience. While they looked great, we actually found it quite challenging to do calculations or add columns to these data frames. For example, to calculate the average grams of beef eaten by a male, none of the simple calculations that were previously done worked. Index errors kept being thrown.

```
GenderType   FoodType
Female       ASPARAGUS        3688.38
             BEEF            51243.92
             CHICKEN        105368.52
             DAIRY          879088.88
Male         ASPARAGUS        3200.12
             BEEF            84424.12
             CHICKEN        148757.72
             DAIRY         1035607.00
```

Having spent several hours researching and attempting different approaches, a workaround was created that added a new column, RowType, containing tuples of the first and second index (e.g., (male, beef) or (23 - 29, asparagus)).

		GramsEaten	**RowType**
GenderType	**FoodType**		
Female	**ASPARAGUS**	3688.38	(Female, ASPARAGUS)
	BEEF	51243.92	(Female, BEEF)
	CHICKEN	105368.52	(Female, CHICKEN)
	DAIRY	879088.88	(Female, DAIRY)
Male	**ASPARAGUS**	3200.12	(Male, ASPARAGUS)
	BEEF	84424.12	(Male, BEEF)
	CHICKEN	148757.72	(Male, CHICKEN)
	DAIRY	1035607.00	(Male, DAIRY)

This enabled us to then apply a lambda to either the first or second item in the tuple to generate the TotalPeople count column for that group, which then supported the calculation of the final calculations (e.g., AvgGrams, AvgCO2perKg).

		GramsEaten	RowType	TotalPeople	AvgGrams	AvgCO2perKg
GenderType	**FoodType**					
Female	**ASPARAGUS**	3688.38	(Female, ASPARAGUS)	4418	0.834853	94.120955
	BEEF	51243.92	(Female, BEEF)	4418	11.598895	438.521566
	CHICKEN	105368.52	(Female, CHICKEN)	4418	23.849823	4722.737317
	DAIRY	879088.88	(Female, DAIRY)	4418	198.978923	30102.711436
Male	**ASPARAGUS**	3200.12	(Male, ASPARAGUS)	4243	0.754212	85.029497
	BEEF	84424.12	(Male, BEEF)	4243	19.897271	752.259766
	CHICKEN	148757.72	(Male, CHICKEN)	4243	35.059562	6942.487452
	DAIRY	1035607.00	(Male, DAIRY)	4243	244.074240	36924.998476

It worked, but we felt like there may be a more elegant, and easier, way to arrive at these calculations. For instance, is there a way to pull from the tables we'd already created?

A dictionary of foods with their related carbon emissions was created for the final set of calculation of AvgCO2perKg. This data was pulled in from a separate source, "Less Beef, Less Carbon", National Resources Defence Council, March 2017. The Emissions Factor for beef, chicken, and asparagus was straightforward; an averaged amount for all the dairy categories was used.

TABLE 1: GREENHOUSE GAS EMISSIONS FACTORS OF TOP 10 FOODS FOR WHICH DROPS IN CONSUMPTION CONTRIBUTED TO GHG REDUCTIONS 2005–2014[18]

Food	Emissions Factor (kg CO_2 eq/kg)
Beef	26.45
Orange Juice	1.03
Pork	6.87
Plain Whole Milk	1.34
Chicken	5.05
High-Fructose Corn Syrup	0.96
Nonfat Dry Milk	10.40
Canned Tomatoes	1.10
Frozen Potatoes	1.44
Fresh Head Lettuce	1.08

TABLE 2: GREENHOUSE GAS EMISSIONS FACTORS OF TOP 10 FOODS THAT CONTRIBUTED MORE GHG EMISSIONS 2005–2014[22]

Food	Emissions Factor (kg CO_2 eq/kg)
Other American Cheese	9.78
Butter	11.92
Mozzarella Cheese	9.78
Other Italian Cheese	9.78
Yogurt	2.02
Evaporated Canned Skim Milk	3.10
1 Percent Milk	1.34
Cane and Beet Sugar	0.96
Fresh Asparagus	8.87
Fresh Leaf Lettuce	1.08

		Daily CO2
GenderType	FoodType	
Female	ASPARAGUS	0.007405
	BEEF	0.306791
	CHICKEN	0.120442
	DAIRY	1.315251
Male	ASPARAGUS	0.006690
	BEEF	0.526283
	CHICKEN	0.177051
	DAIRY	1.613331

```
CO2_day_foods_gender = df5.unstack(level=-1)
CO2_day_foods_gender
```

	Daily CO2			
FoodType	ASPARAGUS	BEEF	CHICKEN	DAIRY
GenderType				
Female	0.007405	0.306791	0.120442	1.315251
Male	0.006690	0.526283	0.177051	1.613331

There were no major challenges while analyzing the Driving data. The dataframe was parsed to keep only a few columns including yearly mileage, gender and age. To facilitate analysis and plotting, gender values or 1 or 2 were replaced with male or female using a function. The yearly mileage column was filtered to exclude spurious mileage numbers (999 999) meant to represent NaN values.
A daily mileage column was created from the yearly mileage column and from there, Kg of CO_2 were calculated per day based on multiplying the daily mileage column values by 0.4 Kg CO_2/mile.
Age range categories were created by using the “cut” method on the Age column continuous values.

While processing the Driving data wasn’t challenging, it was challenging to join the Food and Driving data because the food data was multi-indexed. We had to join multi-indexed subsets of the food dataframe with the single-indexed driving data to compare relative CO_2 emissions. To do this, subsets of the food dataframe had to be “unstacked” to create separate columns that contained the Asparagus, Beef, Chicken, and Dairy Daily CO_2 data rather than having them all “stacked” in the same column (see example left).

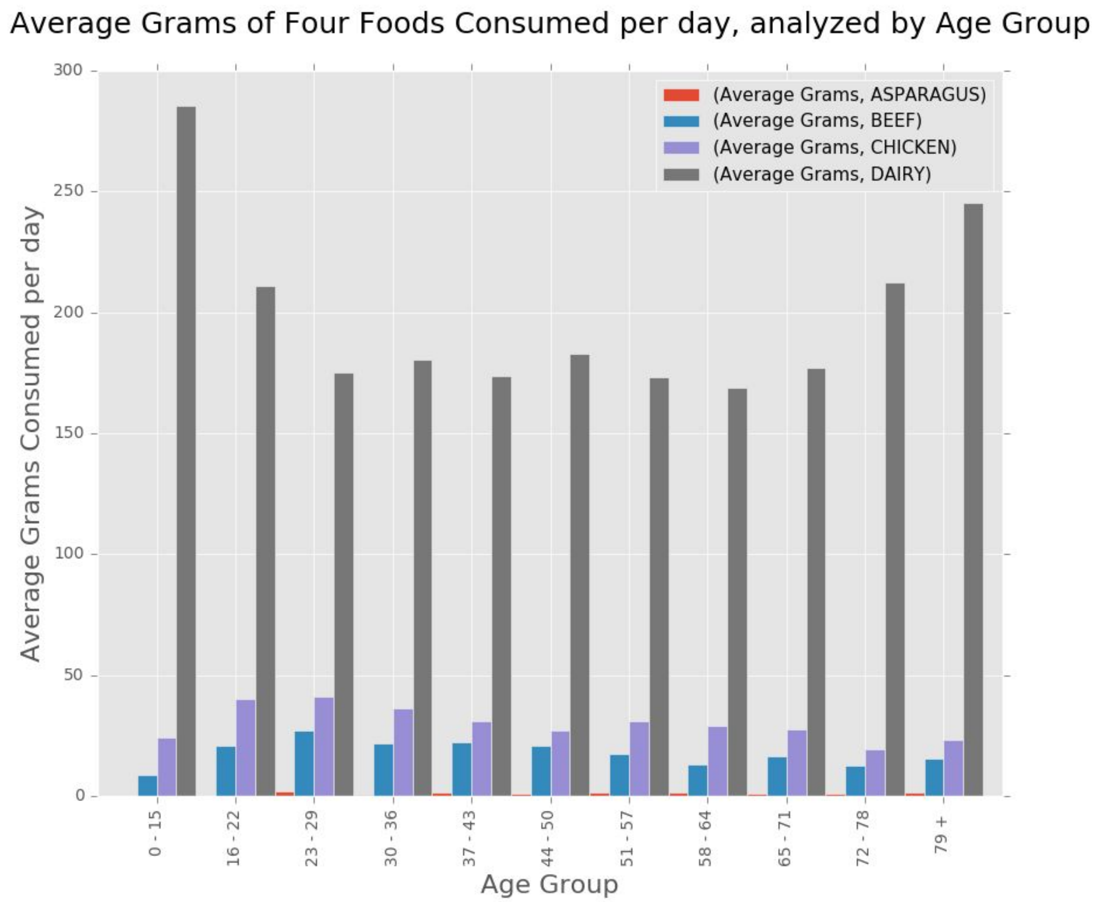
The unstacked food dataframes could then be joined to driving data to make the appropriate comparisons and charts. We capitalized on our newfound ability to “unstack” data in order to make multi-dimensional plots of the driving data that grouped daily mileage or daily Kg CO2 equivalents by both age and gender.

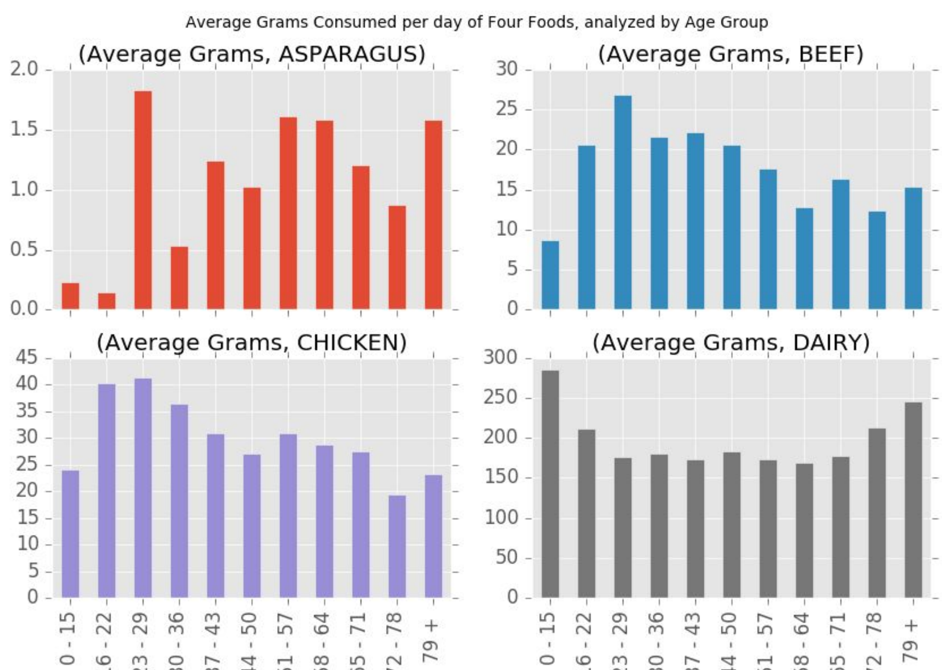
The Story

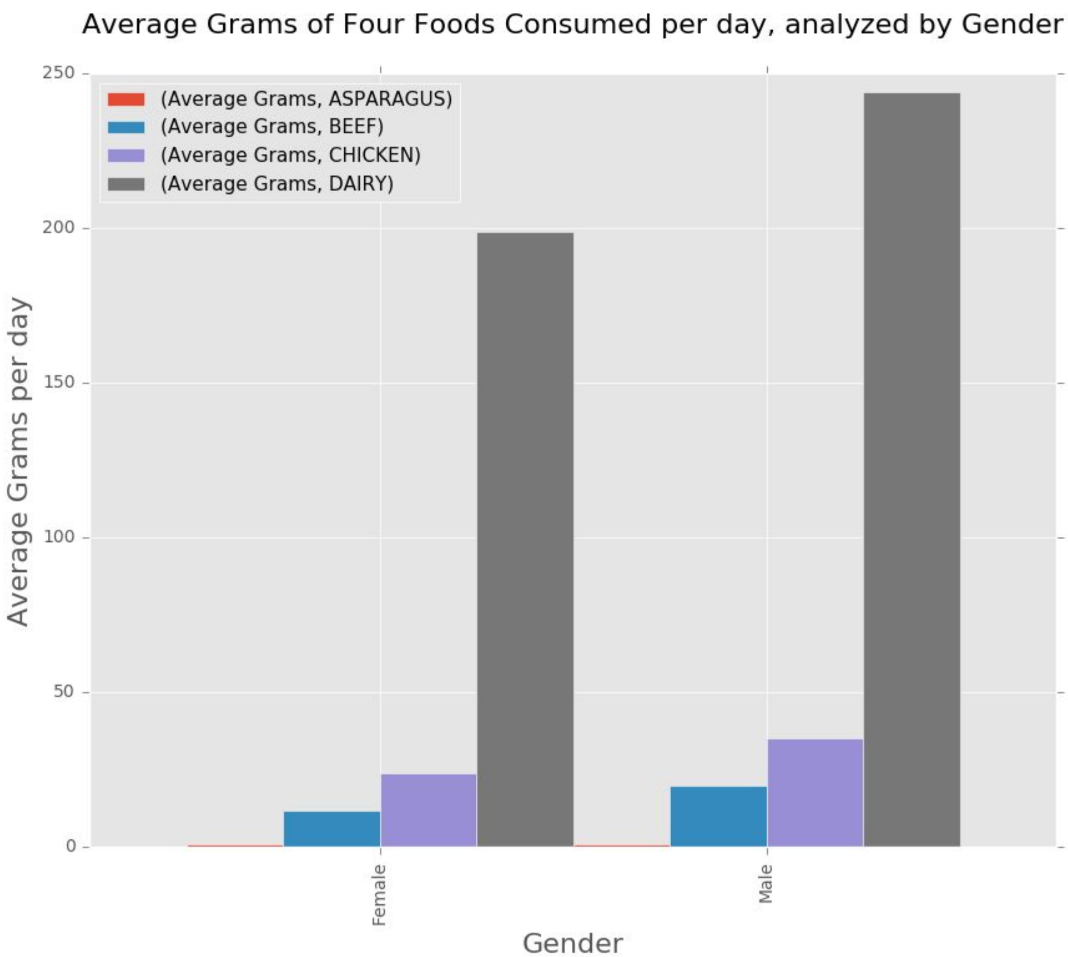
Based on news stories and journal articles, we anticipated that Kg CO_2 emission equivalents from beef would be the highest for any food category consumed and that the Kg CO_2 emission equivalents from beef (or all high-emissions foods) might be comparable to Kg CO_2 emission due to driving.

We chose four commonly-consumed, high-emissions foods to analyze patterns of their consumption by gender and age in the US population (Beef, Asparagus, Dairy, Chicken).
In order to calculate Kg CO_2 equivalents due to consumption of these foods, we calculated the average grams of each of the four foods consumed by age and gender.

While analyzing the average grams consumed per day of each food type by Age group (Figure below: “Average Grams of Four Foods Consumed per day, analyzed by Age Group), it became apparent that for all age, dairy products were the most heavily consumed (by at least 3-fold more than any other food across all age groups). Notably, children and the elderly consumed the most dairy.



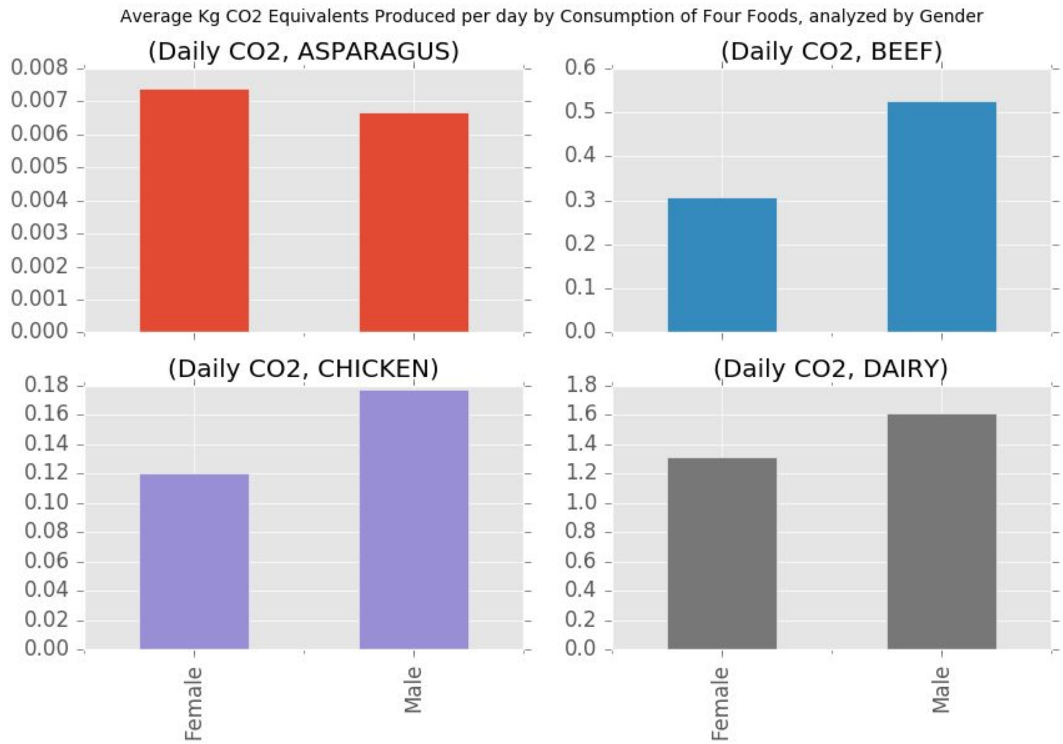


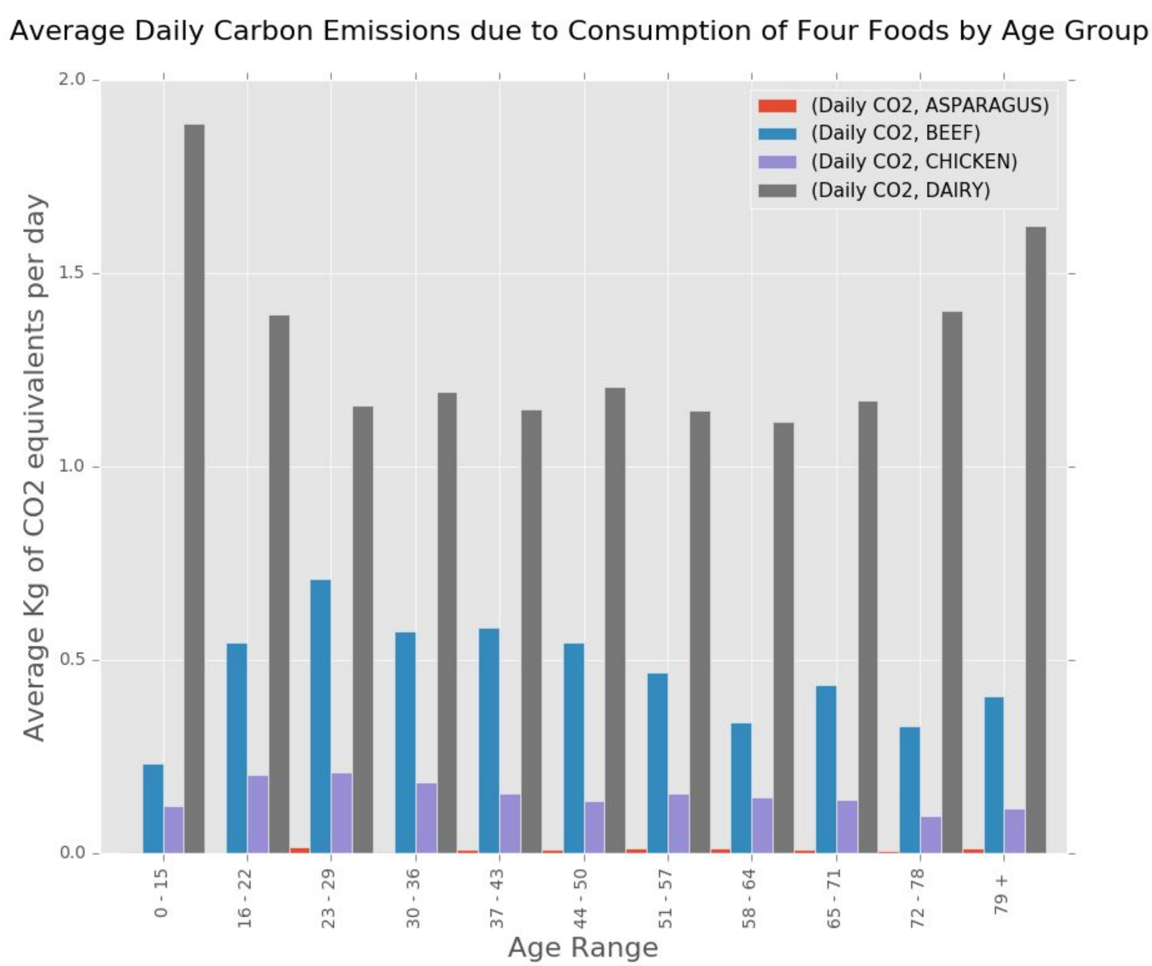


Looking across ages at grams of each food consumed individually (Figure Above:“Average Grams Consumed per day of Four Foods, analyzed by Age Group”), it is apparent that children don’t eat much asparagus (but very little asparagus is consumed on average across all age groups), and young adults eat the most beef and chicken.

Looking by gender at grams of each of the four foods consumed (Figure Above: “Average Grams of Four Foods Consumed per day, analyzed by Gender), it is apparent that men eat more beef, chicken, and dairy than women.

Looking by gender at grams of each of the four foods consumed on their own axes (Figure Below: “Average Grams Consumed per day of Four Foods, analyzed by Gender), it is apparent that women eat more asparagus than men and that men eat almost twice as much beef as women.

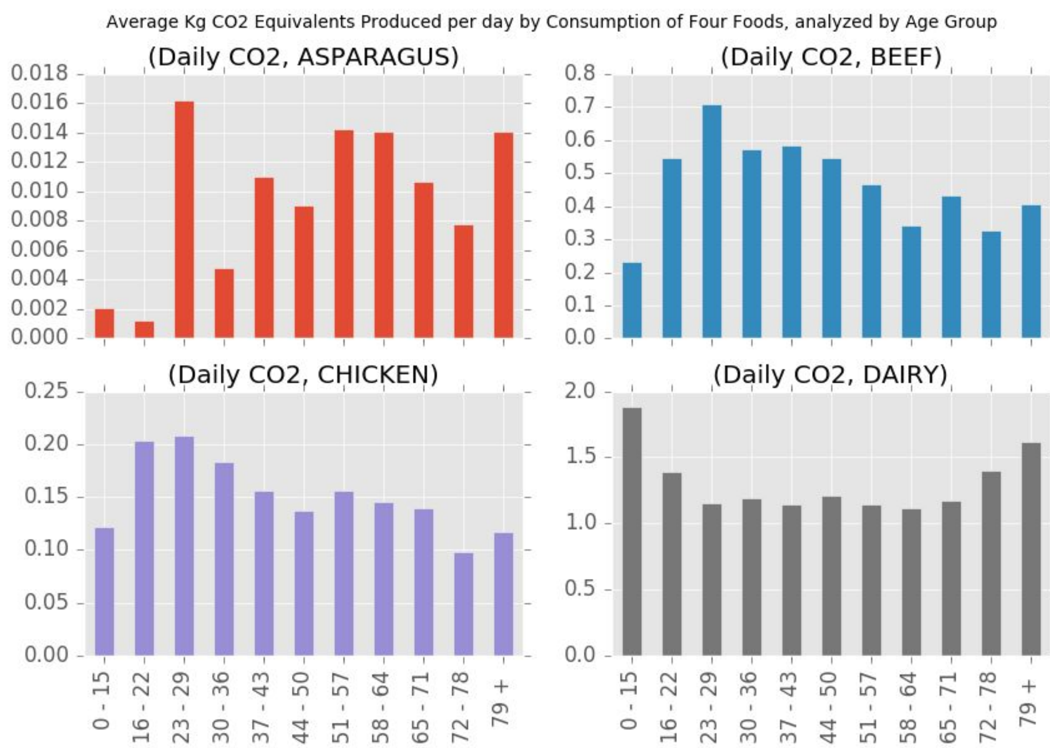

Average Kg CO2 Equivalents Produced per day by Consumption of Four Foods, analyzed by Gender
(Daily CO2, ASPARAGUS)
(Daily CO2, BEEF)
(Daily CO2, CHICKEN)
(Daily CO2, DAIRY)
Female
Male
Female
Male


Average Daily Carbon Emissions due to Consumption of Four Foods by Age Group
(Daily CO2, ASPARAGUS)
(Daily CO2, BEEF)
(Daily CO2, CHICKEN)
(Daily CO2, DAIRY)
Average Kg of CO2 equivalents per day
0 - 15
16 - 22
23 - 29
30 - 36
37 - 43
44 - 50
51 - 57
58 - 64
65 - 71
72 - 78
79 +
Age Range

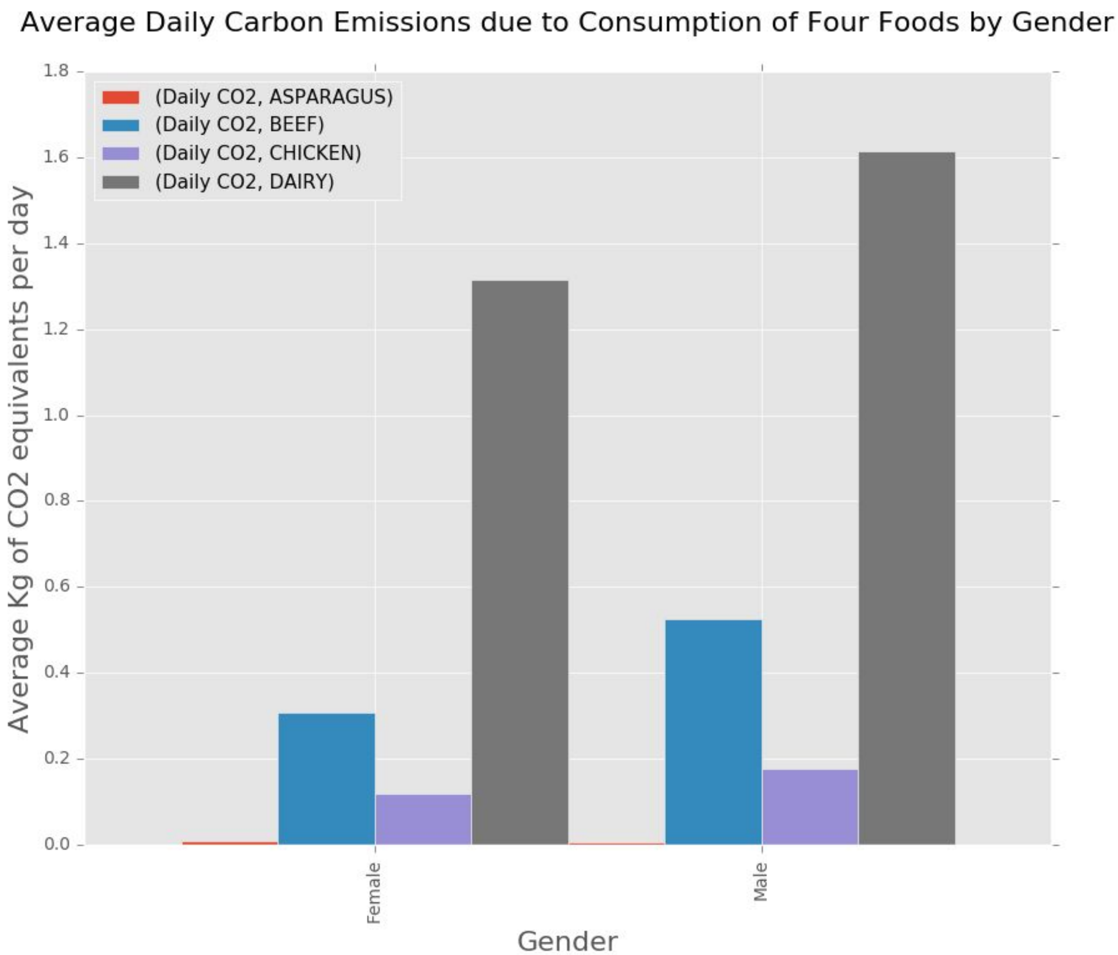
As mentioned above, we had heard from several sources that the Kg CO_2 equivalents from beef was the largest dietary source of CO_2 equivalents. Thus, we were very surprised to find that across age groups, the average Kg of CO_2 equivalents per day were largest for dairy (Figure above: "Average Daily CO_2 equivalents due to consumption of four foods by age group). Thus it follows that the individuals who consumed the most dairy (children and the elderly) produced the highest dairy related Kg of CO_2 equivalents. Moreover, it follows from our analysis that children and the elderly most likely have the highest diet-related carbon footprint. Again, we were surprised by this because we anticipated that high beef consumers (adult men) would have the largest diet-related carbon footprint.

Of the foods we analyzed, beef consumption generates the second largest dietary source of Kg CO_2 equivalents per day, followed by chicken. While asparagus is a high intensity CO_2 equivalents food, so little of it is consumed that asparagus contributions to carbon emissions are nearly negligible across all age groups.

Looking at the average Kg of CO_2 equivalents produced per day due to consumption of each of the four foods on its their own axis (Figure Below, "Average Kg of CO_2 equivalents produced per day by Consumption of Four Foods, analyzed by age group") further emphasizes how negligible the CO_2 contributions from asparagus are compared to other dietary sources. In addition, the average Kg of CO_2 equivalents produced per day due to consumption of beef and chicken is highest from young adults.

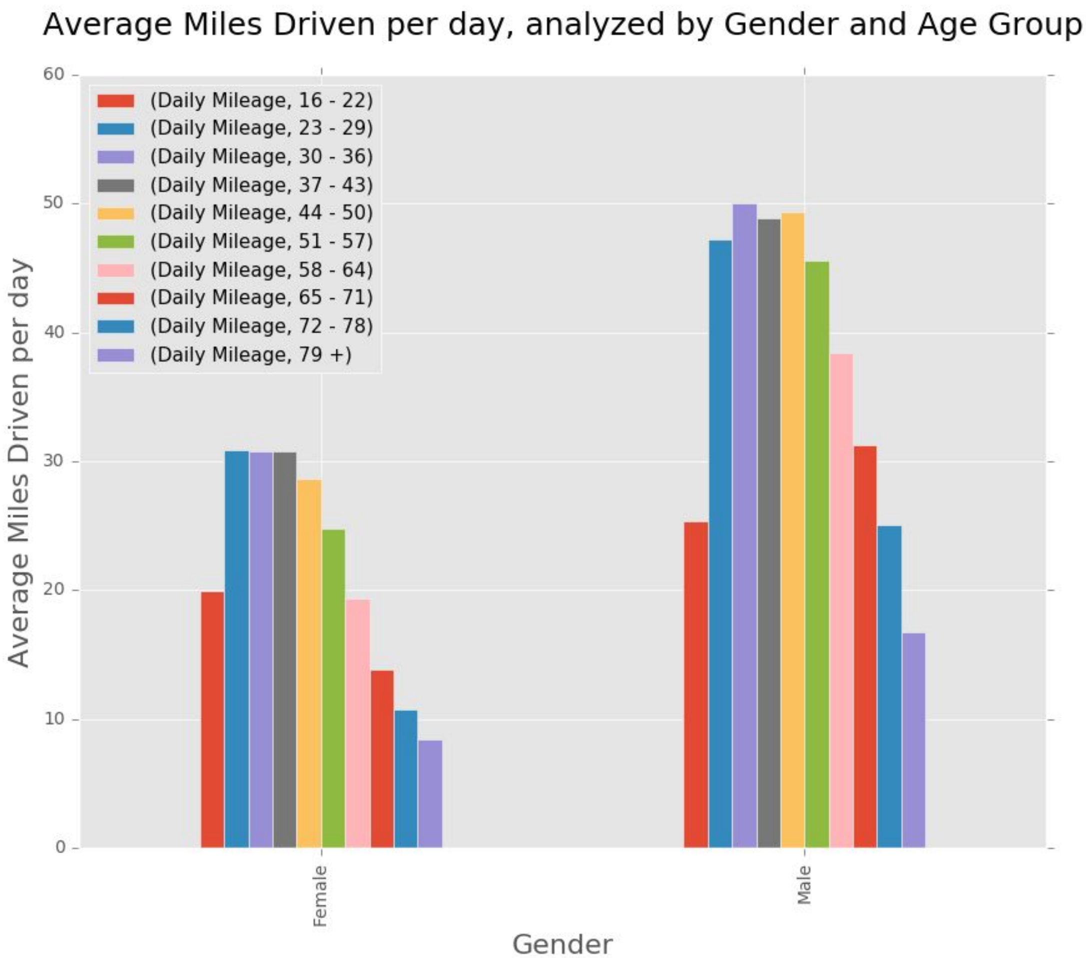


Because men consume more of every food category than women (except CO_2-negligible asparagus!), we anticipated that men would have higher average Kg of CO_2 equivalents produced per day due to food consumption for all foods. That is indeed what we found (Figure below: “Average Daily Carbon Emissions due to Consumption of Four Foods by Gender).

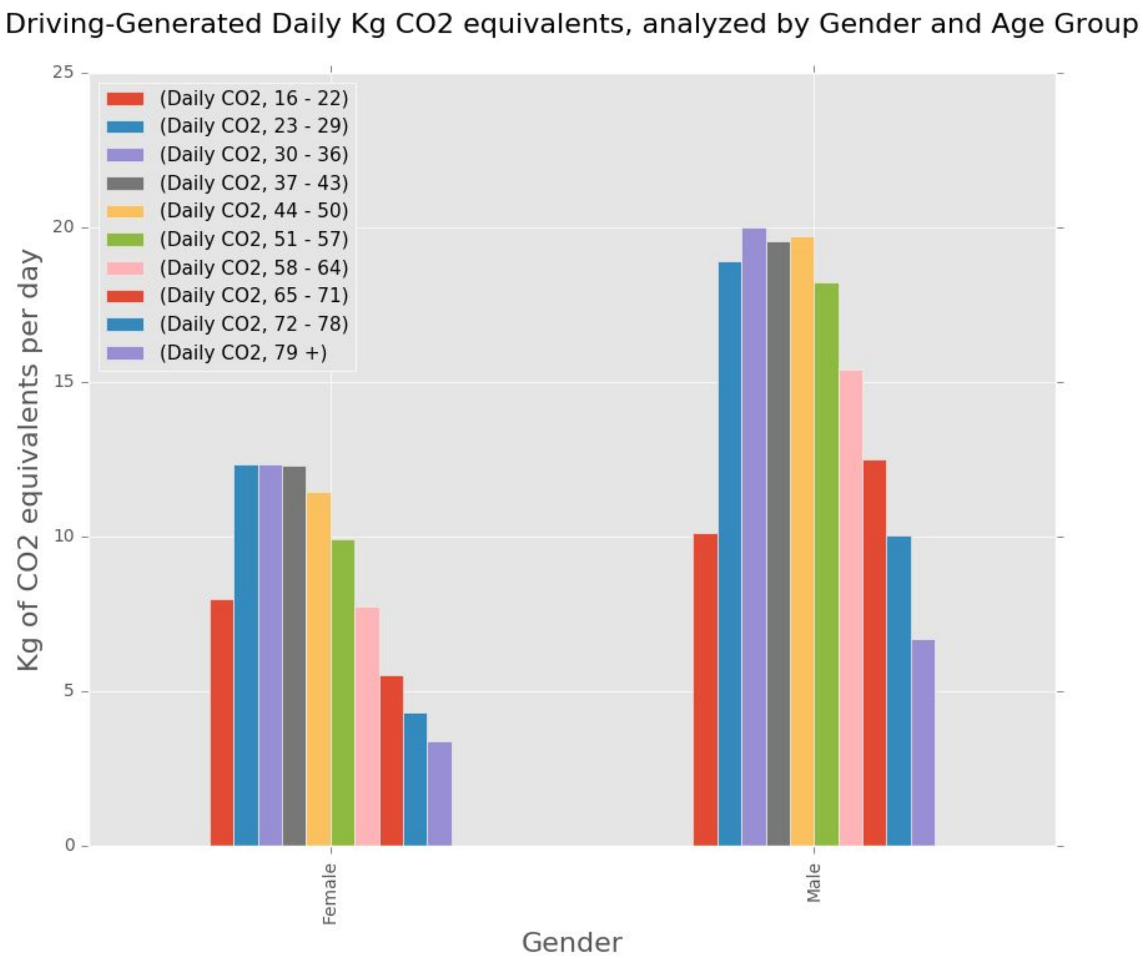


In order to compare diet-derived Kg CO_2 equivalents to Kg CO_2 equivalents due to driving, we analyzed the average miles driven per day by gender and age group.

The average driving distance per year is 12 421 miles. That equates to an average distance of 34 miles per day. In the dataset we analyzed (from 1995, US DOT), men drive nearly 1.7 times as many miles per day as women (26 miles for women, 43 miles for men). This trend is very obvious in the bar chart below ("Average Miles Driven per day, analyzed by Gender and Age Group"). Another obvious trend from our analysis is that the youngest and oldest driving-aged adults drive fewer miles/day than do people aged 23 to ~60. This trend is obvious for both men and women.

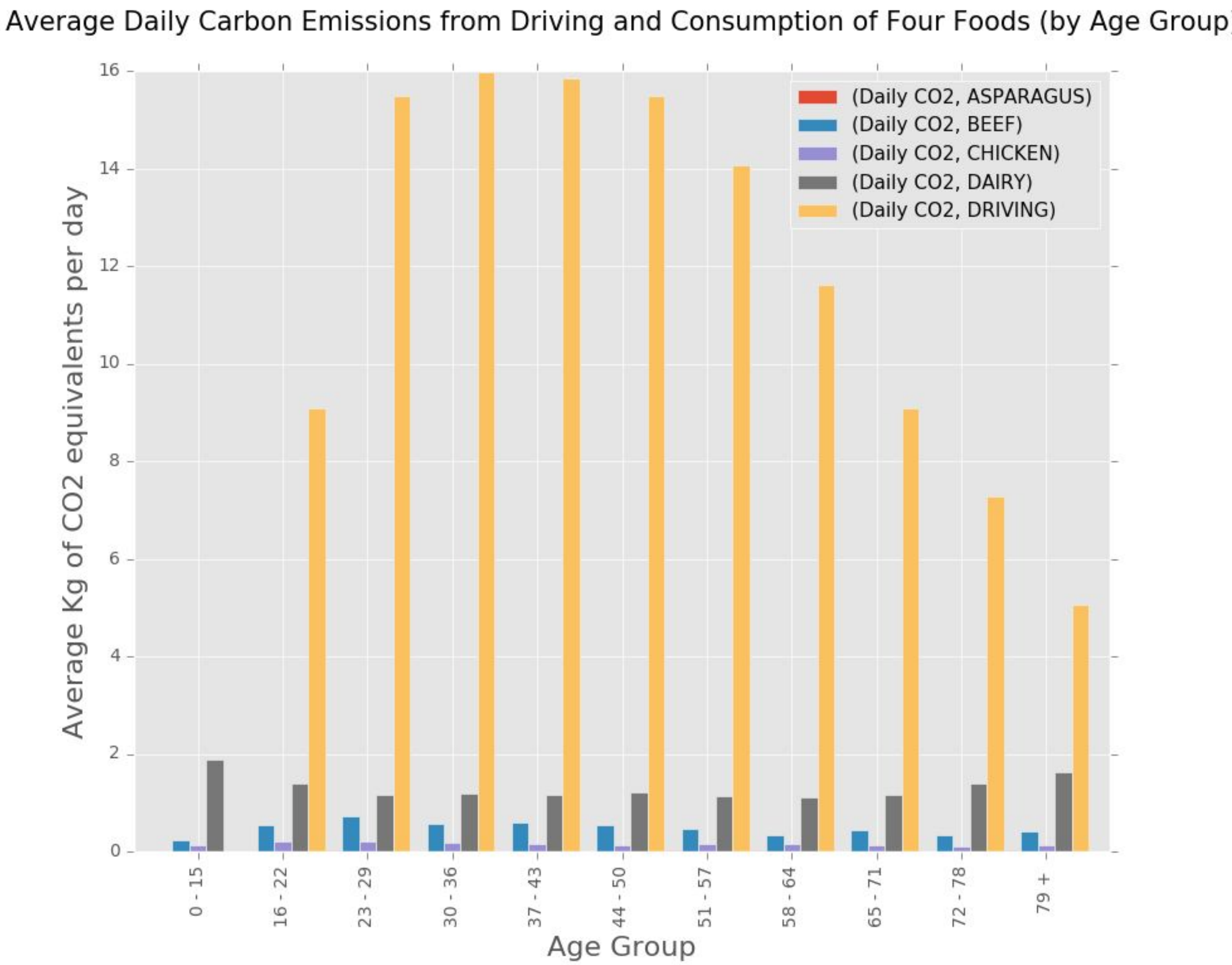


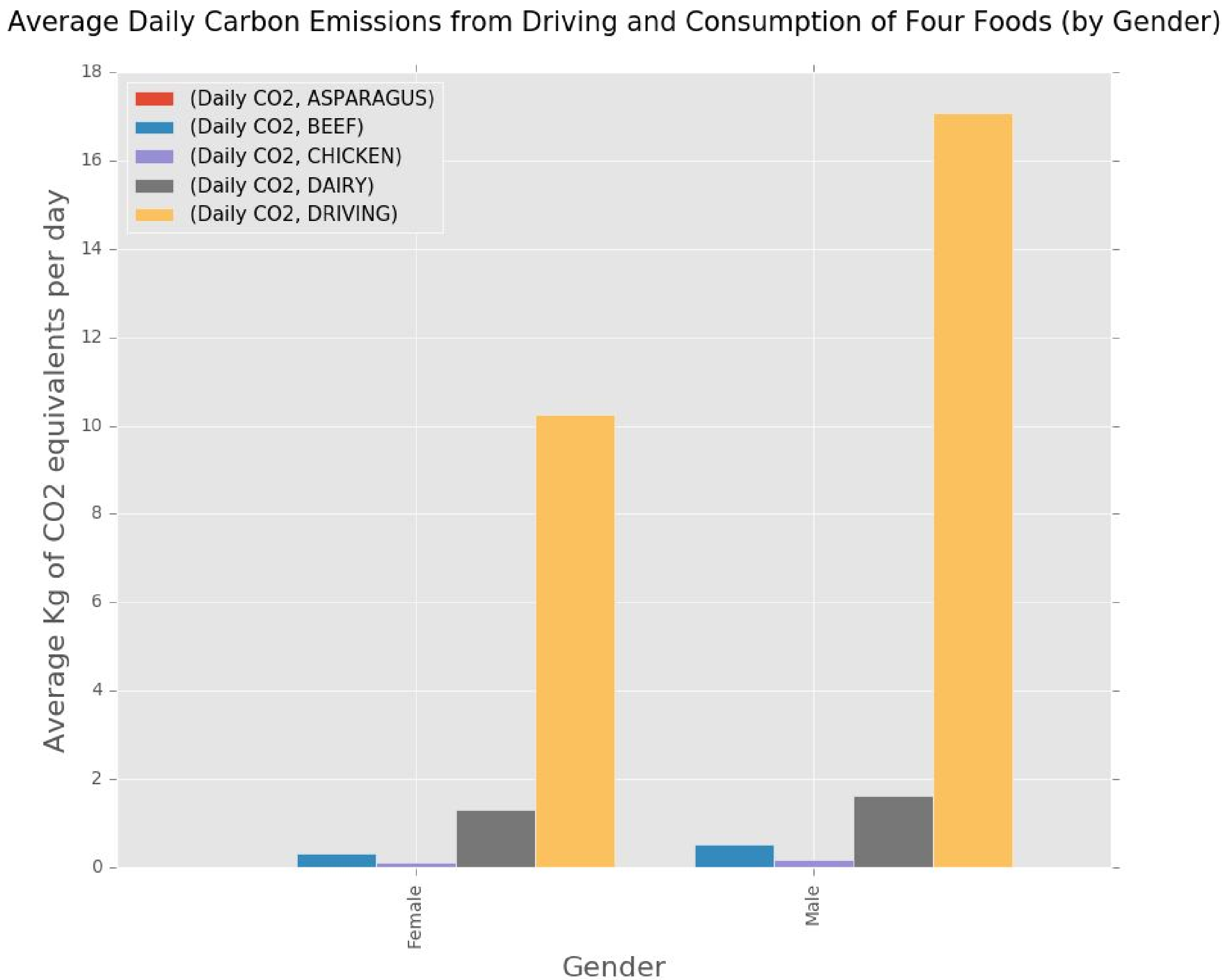
Because daily mileage scales proportionally to daily Kg CO_2 equivalents generated by driving (by a factor of 0.4 Kg CO_2 /mile), the same trends that we gleaned for average daily mileage were apparent for driving-generated daily Kg CO_2 equivalents. Men produce more daily Kg CO_2 equivalents than women and working-aged adults produce more daily Kg CO_2 equivalents from driving that do the youngest and oldest adults.



A surprising insight came when we compared daily Kg CO_2 equivalents from driving to those from consuming highly consumed carbon-intensive foods. Our interest in pursuing the topic of this project came because we had heard from several sources that carbon emissions due to the daily consumption of carbon-intensive foods (like beef) was comparable to - and potentially equivalent to - the carbon emissions that result from daily driving.

When plotted together, it is very clear that daily Kg CO_2 emissions from driving are much higher than the daily Kg CO_2 equivalent emissions that result from consuming any one (and most likely a combination of all) carbon-intensive food across all age groups and for both genders (two charts below: Average Daily Carbon Emissions from Driving and Consumption of Four Foods (by Age group) and (by gender)) .





Average Daily Carbon Emissions from Driving and Consumption of Four Foods (by Gender)
(Daily CO2, ASPARAGUS)
(Daily CO2, BEEF)
(Daily CO2, CHICKEN)
(Daily CO2, DAIRY)
(Daily CO2, DRIVING)
Average Kg of CO2 equivalents per day
0
2
4
6
8
10
12
14
16
18
Female
Male
Gender

CONCLUSION:

In total, we analyzed three food-related and one driving-related files to identify and understand carbon emission patterns of food consumption and driving. In the U.S., the average carbon footprint of a household is most importantly impacted by three factors: house, transportation, and food consumption. While house and transportation choices are fairly expensive to make, food consumption behaviors can be inexpensively (relatively) changed to support reduction in individual contributions to carbon emissions. We went into this project expecting, based on recent news stories and journal publications, that there would be near equivalency in reducing food consumption of high CO_2 foods with reduced driving behavior. However, what the data indicates is that driving is a proportionately higher contributor to a person's overall CO_2 footprint than food.

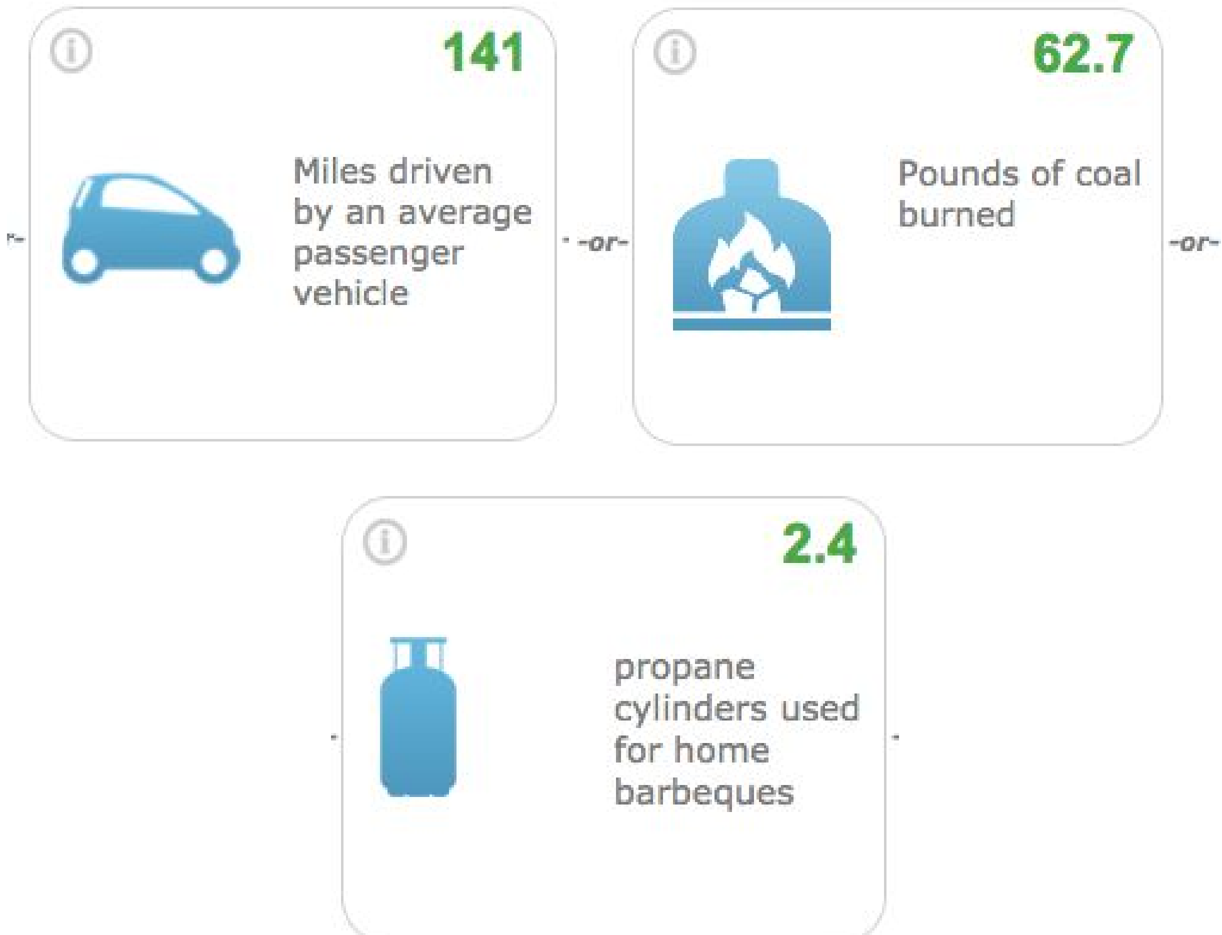
The driving data set in our study is dated: it's over twenty years old. We can make some assumptions that women may have actually closed the driving gap with men, as more women are in the workplace today, and the unemployment rates of men continue to rise. In a related survey we found from the AAA Foundation for Traffic Safety, their "The American Driving Survey" (2014-2015 for which we could not obtain survey-level data) reports that men and women currently drive more similar average miles per day (2015: men = 33, women = 27 versus 1995: men = x, women = 26). So, women are unfortunately also closing the CO_2 gap with men (perhaps faster than closing the wage gap), at least in terms of driving.

Importantly, we've seen that the amount of dairy consumed by people in our study is significantly higher than the amount of beef, chicken, or asparagus. Whether we're talking about drinking milk or eating ice cream, there is a lot of dairy consumed in the U.S., and it is the highest contributor to CO_2 emissions of the foods we looked at. Men have an overall higher contribution to CO_2 than women, based on both food consumption and driving patterns.

Our findings of dairy being the highest contributor to CO_2 emissions of the foods we looked at are at odds with several studies. This could be for several reasons. First, our analysis makes use of the most comprehensive survey of what Americans eat and may therefore be more detailed than other studies. Several published studies do not use estimates of dietary consumption, but instead use loss-adjusted food availability data that takes food wastage into account. Secondly, dairy food codes are very highly represented in the USDA food survey and this might, therefore, allow a more accurate survey of dairy consumption than beef consumption that could skew the results. Third, we made judgement calls on whether to include "mixed foods" (like a steak salad) in our analysis. We had no idea as to what percentage of the total mass in grams resulted from beef and therefore excluded "mixed foods". This could have caused an underestimate of beef consumption in our analysis.

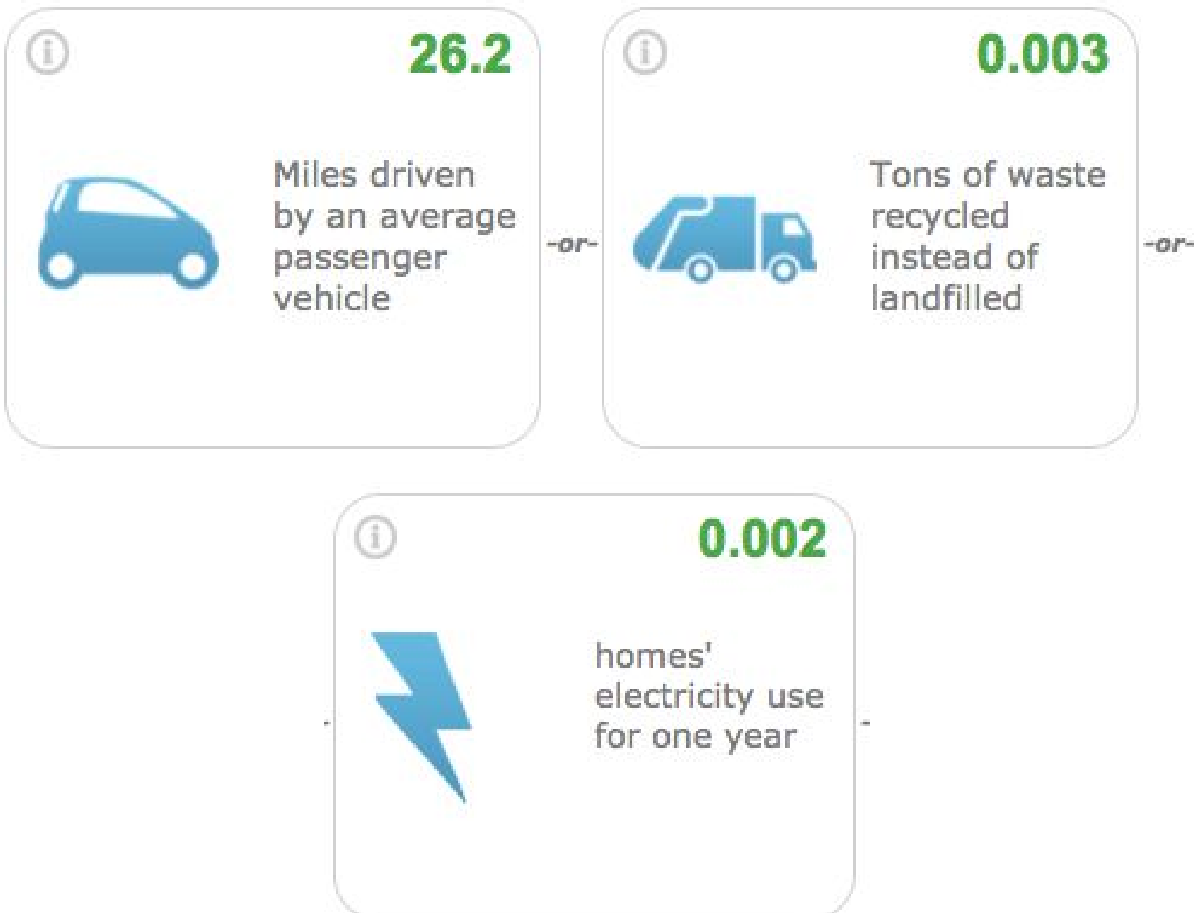
In addition, our multiplier for Kg CO_2 per Kg of dairy was an average of the dairy items and could be an overestimate depending on how much of the total dairy grams consumed were higher emissions dairy items like cheese versus the lower emissions dairy items like milk.

Taking our findings at face value, we are surprised to see that a 10% reduction of dairy (daily total = 244.07 grams per day) by a man for a year (1.61 Kg CO_2 x 0.1 x 365) is equivalent to just 141 miles driven by an average passenger vehicle. This is the equivalent of just three driving days for the average man (based on the 1995 DOT data, men drive an average of 43 miles per day).



Source: Greenhouse Gas Equivalencies Calculator, Environmental Protection Agency

As another example, if a woman eliminated 10% of her beef consumption over the course of a year, it would equate to 26.2 miles driven. This is the same as a single day of driving for a women (based on the 1995 DOT data, women drive an average of 26 miles per day).



Source: Greenhouse Gas Equivalencies Calculator, Environmental Protection Agency