Sustainable Consumption In Martin Luther King High School Households: A Correlation Study Between Dietary Choices/Carbon Footprint and Physical Activity

Ishan Gonehal, Jaslyn Gill

We aim to explore the food preferences of the student population in the context of their environmental impact. It is important to understand how students' dietary choices align with the environmental impact of specific foods. This project is based on the findings from the Environmental Working Group and CleanMetrics Corp.'s "Meat Eater's Guide to climate change + health," which identified the top 9 least environmentally friendly foods based on their carbon footprints.

The project addresses the growing concern of climate change and the role that individual food choices play in contributing to or mitigating its effects. The carbon footprint of food refers to the greenhouse gas emissions associated with its production, transportation, and consumption. Food production, especially in the case of animal products, can significantly contribute to greenhouse gas emissions. The top 9 highest carbon footprint foods, as mentioned in the report, include:

1. Lamb: 39.2 kg CO_2 per kg of food
2. Beef: 27.0 kg CO_2 per kg of food
3. Cheese: 13.5 kg CO_2 per kg of food
4. Pork: 12.1 kg CO_2 per kg of food
5. Farmed Salmon: 11.9 kg CO_2 per kg of food
6. Turkey: 10.9 kg CO_2 per kg of food
7. Chicken: 6.9 kg CO_2 per kg of food
8. Canned Tuna: 6.1 kg CO_2 per kg of food
9. Eggs: 4.8 kg CO_2 per kg of food

Understanding how students' dietary choices align with the environmental impact of specific foods is crucial, so we aimed to explore the food preferences of the student population in relation to their environmental impact. To investigate this, we focused on analyzing carbon footprint metrics associated with the dietary habits and weekly active hours of the Martin Luther King High School student population.

Our research questions that form our investigation are:

1. Within the student population of Martin Luther King Jr. High School, what foods do households consume on a weekly basis among the top 9 foods with the highest carbon footprint?
2. To add, how many pounds of these foods are consumed on a weekly basis?
3. Furthermore, based upon this data, which of these foods consumed on a weekly basis generate the highest carbon footprint?
4. Also, what is the distribution of household carbon footprints regarding dietary choices in Martin Luther King Jr. High School?
5. Lastly, what is the correlation between the hours a household is active weekly and total carbon emission per household?

For our study, we employed an observational approach and conducted an online survey among a randomly selected group of students, as directed by their teachers. We requested that teachers distribute the survey to all their classes.

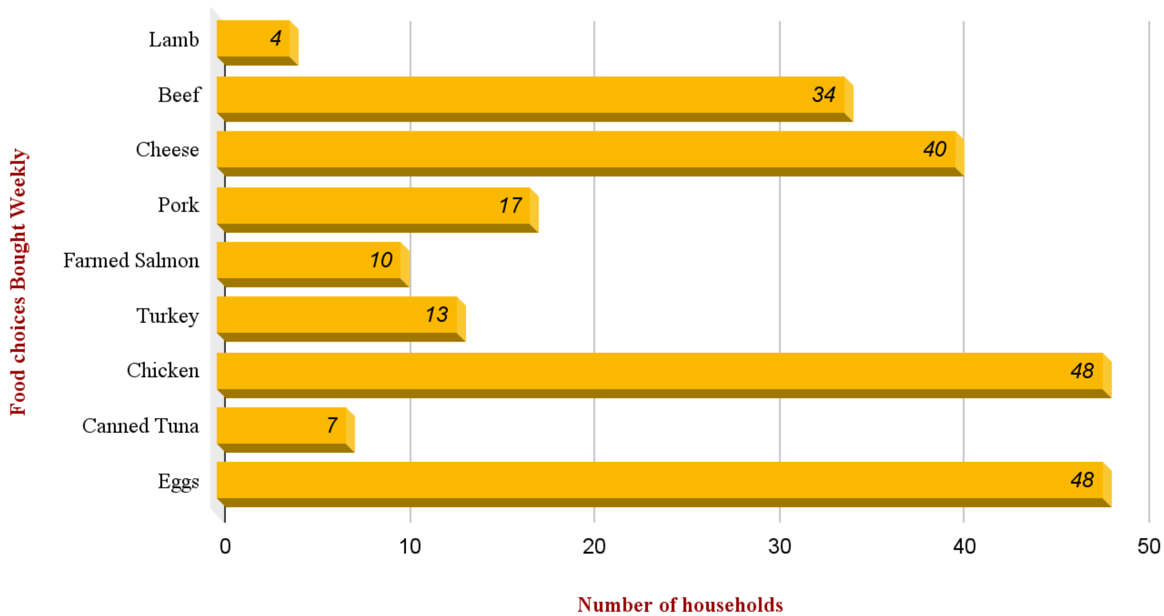
To mitigate accidental or intentional biases in our sampling process, we cataloged all the teachers on campus based on the alphabetical order of their last names and assigned each teacher a number between 000 and 110. Next, we used the rand function on our calculator with values between 000 and 110, to randomly select which classroom we will analyze data from (cluster sampling). Thirdly, we will email those selected teachers asking them to distribute the survey to all the classes they instruct. We then repeat these steps for 18 more times. However, we were only able to get responses from 5 teachers. Ultimately, we were able to collect 52 responses from students in those teachers' classes who had filled out the survey.

01	Ali	
02	Allbee	
03	Alvarez	<>
04	Amarin-Cruz	
05	Anguiano	
06	Applegate	<>
07	Arellano	
08	Barry	<>
09	Beronich	
10	Brewer	
11	Brinkle	
12	Brooks	
13	Brown	
14	Camacho	<>
15	Case	
16	Chavez	
17	Christensen	
18	Christine	<>
19	Cleary D	
20	Cleary M	

Spreadsheet where we cataloged all teachers of Martin Luther King High School (20 out of 111 teachers in the master class list spreadsheet displayed here).

From our survey results, when we examine the graph 'Food Choice Distribution Among Households,' chicken and eggs emerge as the two most popular food choices, appearing in the weekly meals of 48 out of 52 participants from King High School. To add, cheese was also a very popular choice with 40 households stating that they consume this on a weekly basis.

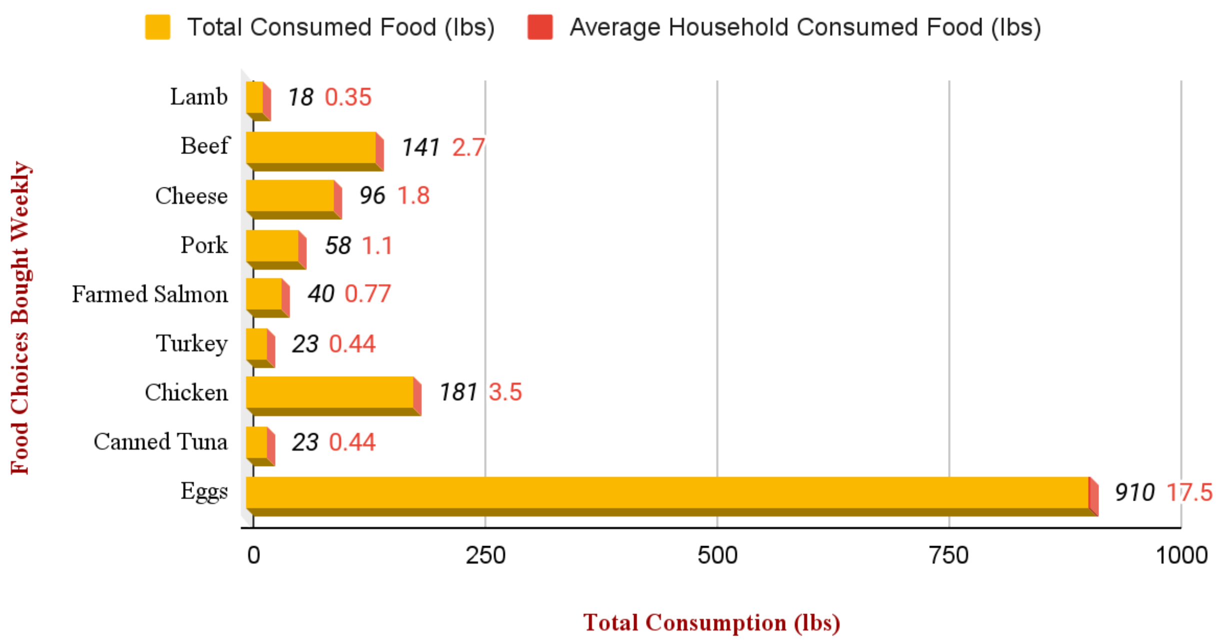
Fig. A: Food Choice Distribution Among Households



Due to how this graph's data was collected, finding the average amount of food choices amongst households is unnecessary to the investigation and impossible to meaningfully compute in conventional manners, thus it is not calculated here.

Furthermore, regarding total pounds consumed, eggs were by far the most consumed type of food among the King student participants, with approximately 910 pounds consumed weekly among 52 participating households. Chicken came in second at 181 pounds, followed by Beef in third place with a total weekly consumption of 141 pounds. Among the top 9 foods with the highest carbon footprints, turkey and canned tuna were the least consumed types of food in households of King students at only around 23 pounds.

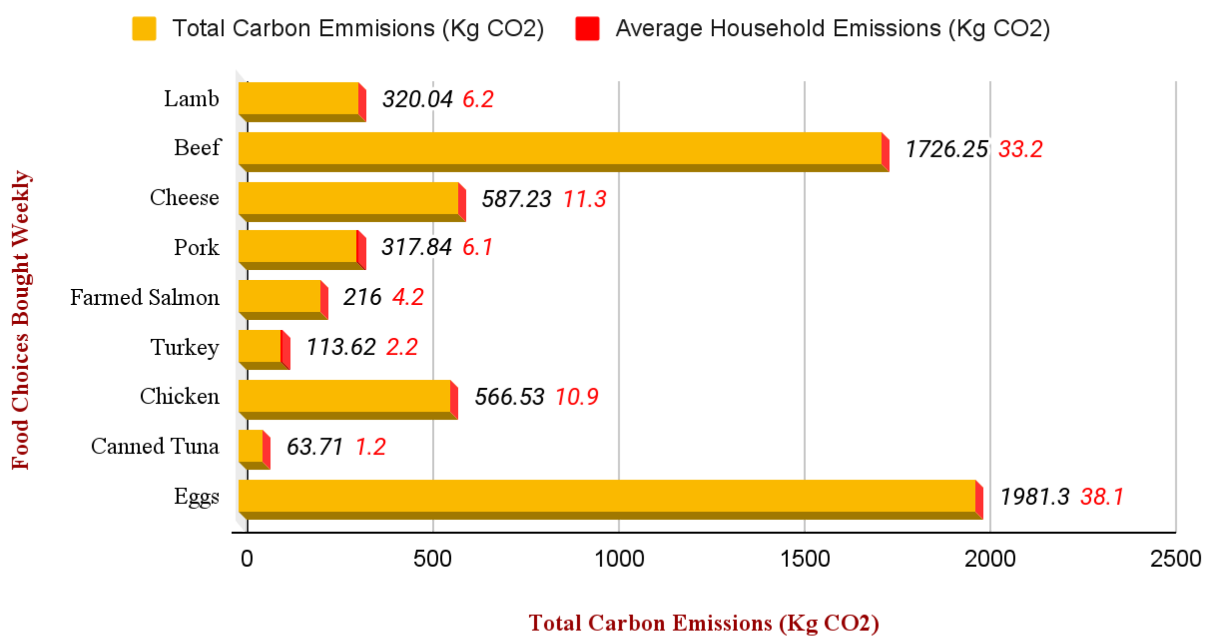
Fig. B: The Total Pounds Consumed Weekly by Food Choice



To find average household consumption by food choice, we divide each value by 52, (as there were 52 participating households) to get the average pounds of consumption regarding each food choice. Thus, the red colored numerical values represent this calculated average of each food choice.

Additionally, regarding total carbon emissions, we took our total pounds data, converted it into kilograms, and then multiplied the values with their corresponding conversion factors (Lamb: 39.2 kg CO_2 per kg of food, Beef: 27.0 kg CO_2 per kg of food, etc.). From our observations, Eggs and Beef had the highest total carbon emissions generated by our 52 participants by far at 1981.3 Kg of CO_2 for Eggs and 1726.5 Kg for beef. Cheese and chicken followed at 587.23 Kg of CO_2 and 566.53 Kg of CO_2 respectively. Among the top 10 foods with the highest carbon footprints, turkey and canned tuna were the least impactful regarding carbon footprint within our surveyed households of King students at only around 113.62 Kg of CO_2 and 63.71 Kg of CO_2 respectively.

Fig. C: The Total Carbon Emissions Generated Weekly by Food Choice



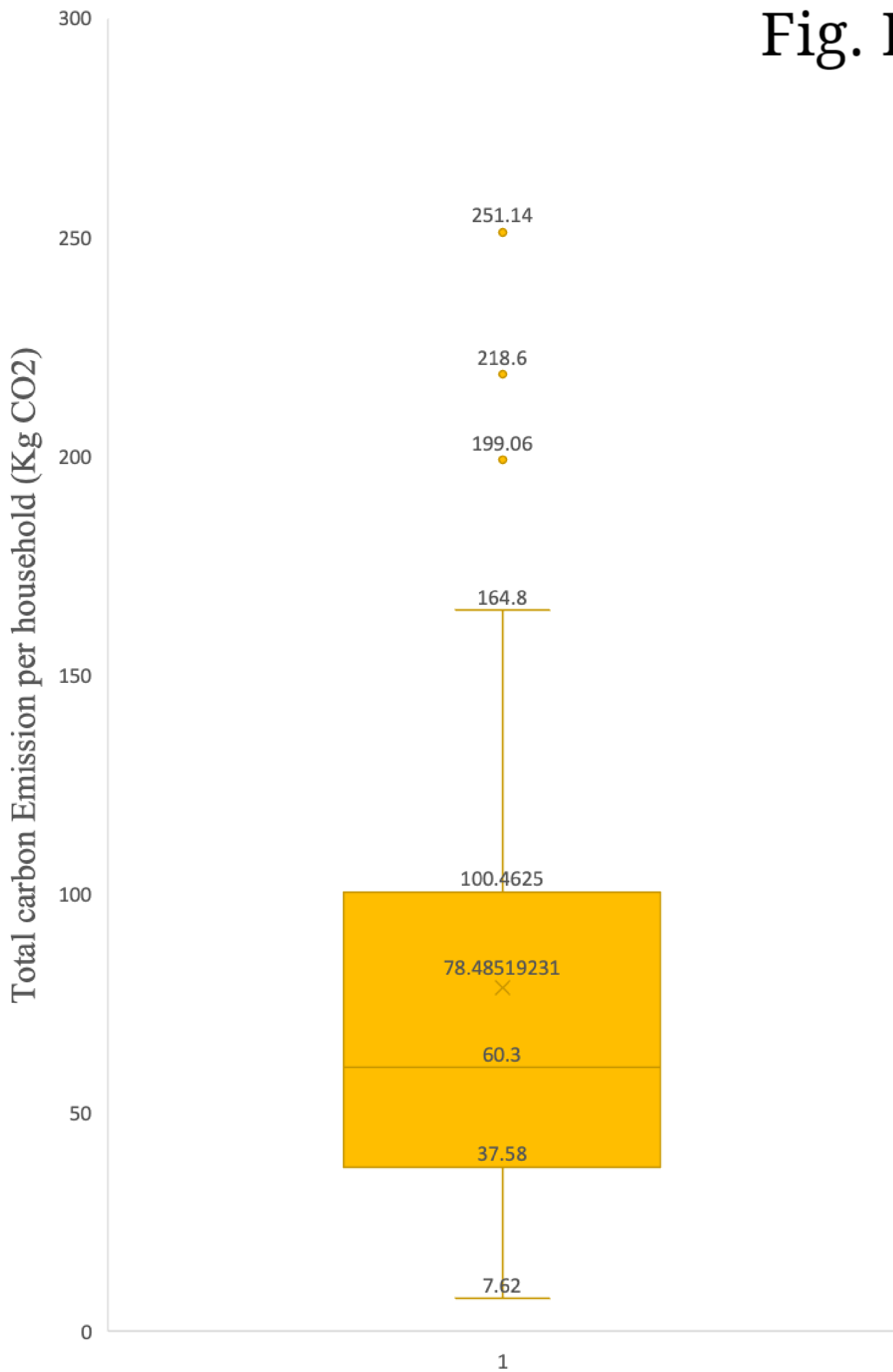
To find average household consumption by food choice, we do the same process as the one in Fig. B while also noting the same information about the visibility of statistical trends.

Next, regarding the distribution of carbon footprints, we calculated the household carbon footprints from our 52 participants by taking the total pounds of each food each household consumed, converting each value into Kg of CO_2, and taking the sum of the total Kg of CO_2 of each household. From this, we can describe the distribution as such:

1. Shape: Unimodal Right Skewed
2. Center:
   a. Mean Carbon Emissions per household: 78.5 Kg CO_2
   b. Median Carbon Emissions per household: 60.3 Kg CO_2
3. Spread:
   a. First Quartile: 37.6 Kg CO_2
   b. Third Quartile: 100.5 Kg CO_2
   c. Interquartile Range: 62.9 Kg CO_2
   d. Upper Fence: 164.8 Kg CO_2
   e. Lower Fence: 7.62 Kg CO_2
   f. Outliers: Three households -
      i. 199.06 Kg CO_2
      ii. 218.6 Kg CO_2
      iii. 251.14 Kg CO_2

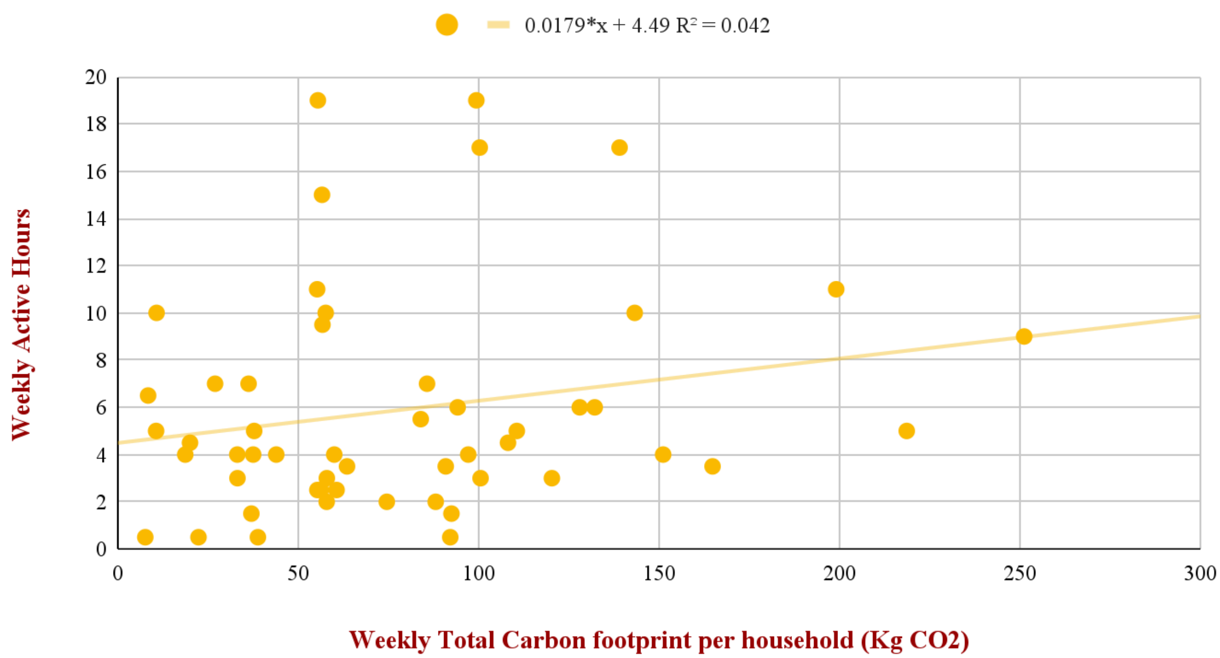
Since we didn't consider the size of each household nor the lifestyles of each household (lifestyles regarding ethnicity, religion, activity, consumption, etc.), these outliers likely result from larger-than-average household sizes and/or more extreme lifestyles resulting from the diversity of American/Californian households among the participants.

Fig. D


Total carbon Emission per household (Kg CO2)
300
250
200
150
100
50
0
251.14
218.6
199.06
164.8
100.4625
78.48519231
60.3
37.58
7.62
1

To try explaining an expected distribution pattern, we asked in our survey the hours a week a household is active for, as we believed that how active a person is could change their eating habits, causing them to consume more food that could possibly result in high carbon emissions.

Fig. E: Weekly Active Hours Vs. Weekly Total Carbon Emission Per household



From comparing these values, we calculated and analyzed the data regarding correlation, linear regression, and regression analysis in order to investigate a possible cause and effect relationship between weekly active hours per household and the weekly total carbon footprint per household. The investigation resulted in such values and interpretations:

1. Correlation Coefficient (r): 0.2047
    a. The positive value of 0.2047 suggests a weak positive correlation between the hours someone is active per week and their total carbon emissions regarding dietary choices. However, the strength of the correlation is considered weak because the absolute value is relatively close to zero.
        i. The vertical outliers decrease the correlation since their high magnitude pulls the line of best fit vertically, thus decreasing the model's accuracy. However, the decrease is negligible regarding their influence, as there are only 3 outliers of 52 data points. To add, the largest of these outliers is above the upper fence by only 100 Kg of CO_2, therefore further proving its negligibility on correlation.

2. Coefficient of Determination (r^2): 0.0419
    a. This value indicates that approximately 4.19% of the variability in total carbon emissions can be explained by the variability in the hours of activity per week. This suggests that the relationship between these two variables is relatively weak, and other factors may contribute to carbon emissions not accounted for in this analysis.

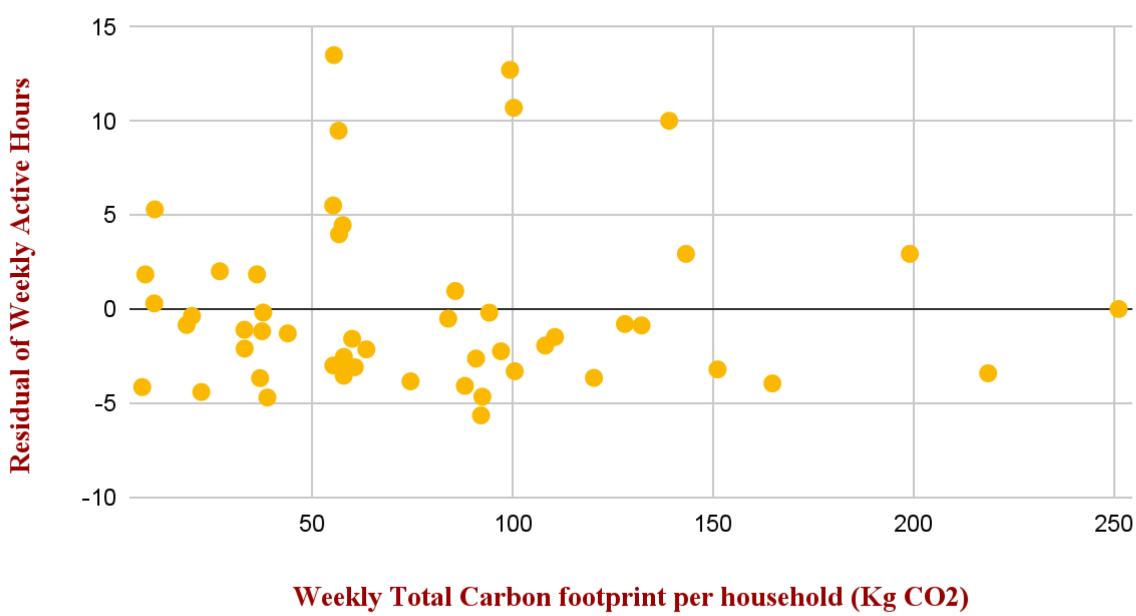
3. Line of Best Fit: $\hat{y} = 0.01786X + 4.49224$
    a. The coefficient 0.01786 (the slope) indicates the change in the predicted total carbon emissions for a one-unit increase in hours of activity per week. In this

case, as the hours of activity per week increase, the predicted total carbon emissions also increase, but the relationship is relatively weak due to the small magnitude of the slope.

b. The intercept 4.49224 represents the predicted total carbon emissions when the hours of activity per week (X) are zero. However, in the context of this study, it is not necessarily a practical interpretation of the data since having zero hours of activity per week might be an unlikely scenario.

To clarify, the accuracy and reliability of predictions from this model depend on the strength and nature of the underlying relationship in your data. Since the correlation coefficient (r) is very weak at a value of 0.2047, it suggests that the linear model may not fully capture all values regarding the comparison between hours of activity and total carbon emissions in a necessarily accurate manner. This clarification on linear accuracy and reliability can be further visualized by a residual plot.

Fig. F: Residual Plot of Fig. E



The information from the residuals plot can give insights into the performance of the linear regression model and the overall correlation between the variables. Here are some interpretations:

1. Negative Residuals Clustered between 0 and -5:
    a. The negative residuals suggest that, on average, the actual values (observed total carbon emissions) tend to be lower than the values predicted by the linear regression model.
    b. A consistent negative bias in residuals could indicate that the model tends to overestimate total carbon emissions based on the hours of activity.

2. Rarely Positive Residuals with Highs of 14 and Lows of <1:
    a. The fact that positive residuals are less common and tend to have larger magnitudes (as high as 14) indicates that the model sometimes underestimates total carbon emissions but sometimes also grossly overestimates.
    b. This suggests that there are instances where the linear model fails to capture the variability in total carbon emissions adequately due to stout outliers.

Overall, the very weak positive correlation coefficient (r) and the limited explanatory power of the model (as indicated by the low r^2 value) are consistent with the findings from the residuals plot. Additionally, the consistent negative bias and occasional large positive residuals imply either:

1. The linear regression model may not fully capture the underlying patterns/relationships between hours of activity and total carbon emissions.

2. The overall correlation between both variables might simply just be very weak meaning no other existing model could unearth a hidden variable that would increase the correlation between the two, as any such variable is nonexistent.

In essence, the information from the residuals plot suggests that the linear regression model may have limitations in accurately predicting total carbon emissions based on hours of activity, and there may be room for improvement or exploration of alternative modeling approaches.

As a result, we examined the data in nonlinear models in an attempt to observe what the linear model may not account for. The models we investigated included exponential regression, logarithmic regression, power regression, and reciprocal regression, all of which failed to re-express the data in an explainable linear pattern, thus indicating that the second implication (which described how both variables could simply just have a weak correlation) was likely the correct one.

To summarize, the investigation into the food preferences and environmental impact of Martin Luther King High School students revealed several key findings. Chicken, eggs, and cheese emerged as the most popular food choices among surveyed students, with eggs being the most consumed, amounting to approximately 910 pounds per week across 52 households. Conversely, turkey and canned tuna, known for having lower carbon footprints, were among the least consumed foods. In terms of carbon emissions, eggs and beef contributed significantly to the total, generating 1981.3 Kg and 1726.5 Kg of CO_2, respectively. The distribution of carbon footprints per household exhibited a right-skewed pattern, with outliers likely influenced by larger household sizes and diverse lifestyles.

An exploration of the correlation between students' weekly activity hours and their households' total carbon emissions indicated a weak positive correlation. The coefficient of determination suggested that only about 4.19% of the variability in carbon emissions could be explained by weekly activity hours. The linear regression model, while revealing a weak positive correlation, had limited accuracy, as indicated by a residuals analysis. Consistent negative bias in residuals suggested the model tended to overestimate carbon emissions, while occasional large positive residuals hinted at instances where the model failed to capture variability, possibly due to outliers.

This study also contained a possibility of bias that could have altered results. To start, there could have been a nonresponse bias in the survey regarding the people chosen. Despite the teachers being chosen at random, the people who choose to take the survey become a flaw, because it does not include everyone in that randomly chosen teachers class. As a result, said students could have a hidden variable influencing them to respond to the teacher and take the survey while others did not or vice versa.

To elaborate, further nonresponse bias could occur with the time of day the survey was given. Each teacher that gave out the survey released it at a random time, which could affect the people available at that specified time, creating a bias to people who couldn't make it to their class, had connectivity issues, etc.

Another bias in the form of a sampling bias can be observed by how the teachers chosen by the random number generator possibly are not representing the entire group of students residing in Martin Luther King High. Each teacher's class can hold a different student population, where a few of them could choose a non-popular response that could create greater variation and affect the data.

Lastly, there's also a limited amount of food listed on the survey. Although the chances of it being very slim, there could be a possibility that a student does not intake any of the popular choices due to other variables in their life, thus resulting in a slight response bias.

Taking from the studies results and possible biases, to enhance the robustness of future analyses, we proposed several recommendations. Firstly, additional factors such as household size, cultural preferences, and lifestyle variations should be considered to refine the model and improve its predictive accuracy. Addressing the sources of consistent negative bias in residuals and outliers is crucial for a more precise understanding of the correlation between weekly activity hours and carbon emissions. Further exploration of nonlinear models may unveil patterns not captured by linear models, providing a more nuanced understanding of the relationship.

To elaborate, we could have expanded our consideration beyond food choices for meat eaters, accommodating individuals with vegetarian diets and those who abstain from the listed foods. Including the top 9 foods known for releasing the highest carbon emissions in vegan diets would have provided a more comprehensive analysis. Additionally, we could have factored in the participants' household sizes during the survey. This consideration is crucial in determining what factors were influencing the outliers.

Secondly, expanding the sample size would strengthen the statistical power of the analysis, yielding more representative and reliable results. Although, among our 52 participants, it is possible to extrapolate the data to the 3000 households of Martin Luther King Jr. High School, and possibly even further to a district/county level (for example, the beef emissions of 1726.25 Kg of CO2 among the 52 participants would be extrapolated by multiplying Kg of CO2 via population scaling) a larger sample size would solidify these findings as extrapolation can often be flawed and incorrect.

Thirdly, incorporating qualitative data through interviews or open-ended survey questions could provide insights into individual motivations and decision-making processes related to food choices and environmental awareness. A longitudinal study would allow for the observation of changes in food preferences and carbon emissions over time, offering a more comprehensive understanding of trends.

Lastly, educational interventions within the school community could raise awareness about the environmental impact of food choices, potentially influencing future dietary behaviors.

In conclusion, the investigation into the food preferences and environmental impact of Martin Luther King High School students has provided valuable insights into the complex interplay between dietary choices, carbon emissions, and lifestyle factors. The study illuminated the popularity of certain foods among students and their corresponding carbon footprints, revealing notable disparities in consumption patterns. However, the findings also underscored potential biases in the survey methodology, emphasizing the need for a more comprehensive and inclusive approach in future analyses. Recommendations for refining the model, expanding the sample size, and incorporating qualitative data aim to enhance the robustness of future research efforts. Ultimately, this study contributes to the broader conversation on sustainable practices and highlights the importance of considering diverse factors in understanding the intricate relationships between food choices, lifestyle, and environmental impact within American/Californian settings.