Can Plant-based Diets Help Solve Climate Change?

Hanna Dettki* **Davide Mornatta***
Department of Computer Science
University of Tübingen
{hanna.dettki, davide.mornatta}@student.uni-tuebingen.de

Abstract

Food is responsible for one quarter of the world's greenhouse gas (GHG) emissions [1], one of the main drivers of climate change. Providing a nutritious diet for the ever-growing world population while combatting global warming at the same time therefore poses a major challenge. This yields the question whether there are diets that are both sustainable for the climate and meet nutritional requirements. In this study, we investigate if plant-based diets emit significantly less GHG than animal based products. We found that plant-based food groups have significant lower GHG-emissions not only with respect to total emissions but also when controlling for nutritional value. For people wanting to modify their lifestyle towards a more sustainable one both for the planet and their health, as a growing body of research suggests [2], an easy rule to guide their daily food-decisions derived from this study would be to eat more plant-based.

1 Introduction

Climate change is a serious global problem. To prevent disaster, big changes are needed - both in government policies and personal choices.[3] Food, energy, and water contribute to the problem, particularly due to population growth.[4] However, people who want to live sustainably often face obstacles. Food systems alone are responsible for a third of greenhouse gas emissions[5], yet the emissions of different food options are often not widely known, making it difficult for individuals to make climate-friendly choices. This highlights the importance and lack of education in addressing climate change.

Research Question Therefore, we investigate the environmental impact of the most common food categories, which is a first step towards providing more information in the jungle of food choices that everybody is presented with. In particular, we hypothesize that plant-based diets have a significantly smaller CO_2-footprint than diets including animal-derived products.

2 Data

Data generating process The data used in this analysis is the largest meta-analysis of food systems to date and was published in the journal *Science* by Poore and Nemecek [1]. 570 studies met the eleven criteria[2] Poore and Nemecek [1] were applying to the initially 1530 studies. The final dataset covers approximately 38,700 farms across 119 countries and spans a total of 40 food products, which represent approximately 90% of global protein and calorie intake. The breadth of this analysis ensures that results are only minimally biased towards any geographic region, any particular faming method or income level.

Greenhouse gases (GHG) and how they are quantified Since there are many different greenhouse gases (GHG), researchers commonly aggregate them into a measurement that is easy to use for comparisons. Poore and Nemecek [1] use the metric "carbon dioxide-equivalents (CO_2eq)", which is the most common measurement and is also used as a target-setting metric in official reportings like the Paris Agreement. CO_2eq express the "global warming potential" by aggregating the impact of all GHG into a single metric and aims to represent the amount of warming that each specific gas generates relative to CO_2. The "global warming potential over a 100-year timescale (GWP_{100})" expresses the mid- to longterm period for climate-policies. To calculate the CO_2eq-value, the amount of each GHG needs to factorized by its GWP_{100}-value. For example, the Intergovernmental Panel on Climate Change (IPCC) sets the GWP_{100}-value of the GHG methane to 28, which

*Equal contribution.
[2]Such as methodologies used were standardized, e.g., that all stages of the supply chain were considered.

Preprint. Under review.

is based on the rationale that methane's global warming impact over a period of 100 years, will be 28 times higher than that of one kilogram of CO_2 [3].

Dataset modifications and handling of missing data We substituted missing data by taking the mean of the existing values of the respecting feature. To facilitate data queries, we added a column "Plant-based", which holds binary values $[0, 1]$ and indicates whether a product is plant-based (1) or animal-derived (0).

3 Analysis

To get an overview as to which features of the dataset provide most information, we calculated the pairwise correlation of a subset of features. Given a pair of random variables (X, Y), the population Pearson correlation coefficient is:

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \tag{1}$$

where $cov(X, Y)$ is the covariance, σ_X is the standard deviation of X and σ_Y is the standard deviation of Y. Equation (1) applied to the features is depicted as a heatmap in Figure 1. Figure 1 illustrates that plant-based products are strongly anti-correlated with total emissions ($\rho = -0.600367$). This is indicative of our hypothesis that plant-based products emit significantly less than animal-derived products as stated in Section 2.

The code to the data preprocessing and analysis can be found at: `https://github.com/hannma/data-literacy-project.git`

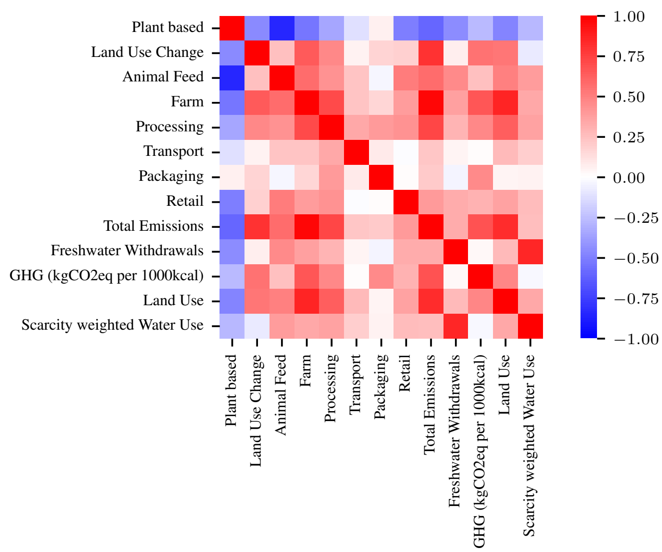


Figure 1: Correlation of a subset of features depicted in a heatmap. Plant-based products are anti-correlated ($\rho \in [-1.0, -0.5]$) with every other feature, in particular *Total Emissions*, with the exception of *Transport* and *Packaging*.

Hypothesis Testing

In order to decide as to whether the data supports our hypothesis, we will run a T-test, which tests, if the two distributions have equal means (Null-hypothesis). Since the samples are neither of same size, nor of same variance and are assumed to be independent, we opt for the independent unequal variance T-Test, also known as *Welch's Test* [6].

The test-statistic is computed as follows:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \tag{2}$$

where $\bar{X}$ and $\bar{Y}$ are the sample mean with respective standard deviations σ_X and σ_Y and we will set the significance level to $\alpha = 0.01$.

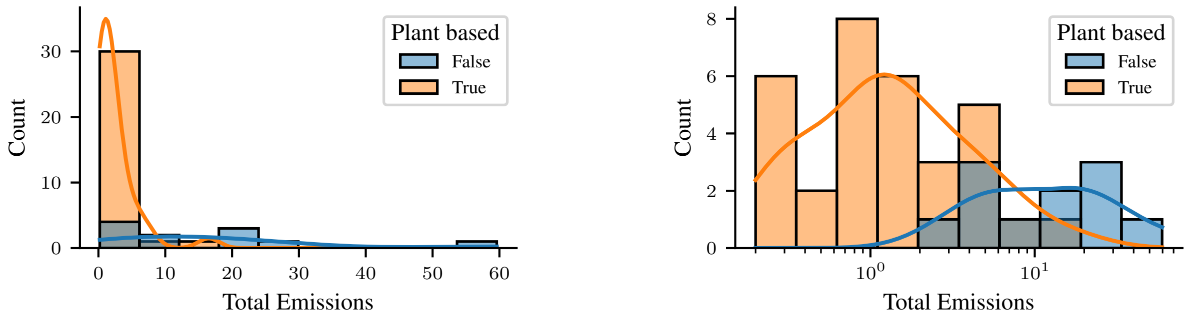
Do plant- and non-plant-based food products significantly differ with respect to their total emissions?

We will consider *plant-based* and *not plant-based* with respect to the two features *Total emissions* and *GHG (kgCO2eq per 1000kcal)* respectively.

Kernel density estimation to check for distribution family Since *Welch's Test* assumes implicitly Gaussian-distributed data, we have to ensure that our assumption of Gaussian distributed data is reasonable. We therefore generate a histogram and estimate the kernel density with the Python package Seaborn [7]. Plant-based and non-plant-based food categories are depicted with respect to their total emissions in Figure 2.

Analyzing Figure 2, two things can be concluded: First, our hypothesis, that the plant-based and animal-based food products have two different underlying distributions can be confirmed. Second, the distribution over total emissions of plant-based products clearly does not follow a Gaussian distribution.

Data transformation We apply the logarithm with basis 10 to the data, which is non-negative. Applying the logarithm, we obtain the kernel density estimation depicted in Figure 2b. The transformed data now seem to follow a Gaussian distribution and we therefore apply the Welch's Test on the transformed data with its results reported in Section 4.



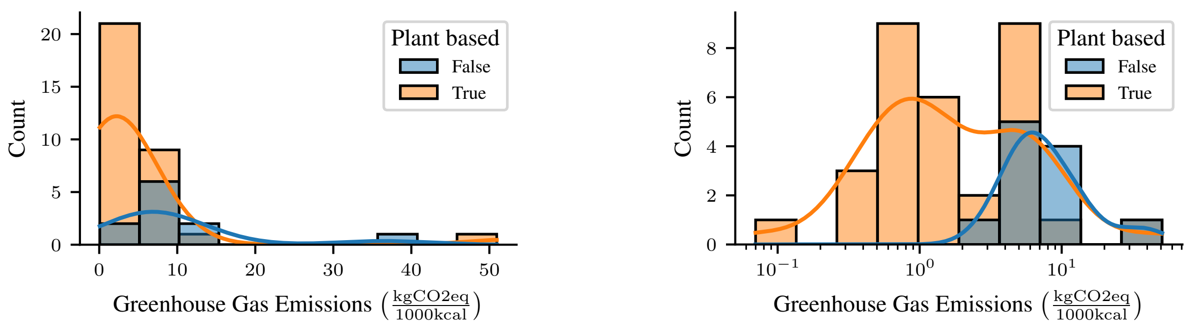
(a) Raw data.

(b) Log-transformed data.

Figure 2: Total emissions for plant-based and animal-based food categories.

Accounting for Nutritional Value: Do plant- and non-plant-based food products significantly differ with respect to their GHG per 1000kcal?

Since it is very important that everybody has access to diets that meet all the nutritional requirements, we will now consider *plant-based* and *not plant-based* with respect to their nutritional value (feature *GHG in kgCO2eq per 1000kcal*). We follow the the same analysis process as described above and see the histogram and kernel density estimation on the original and log-transformed data respectively in Figure 3.



(a) Raw data.

(b) Log-transformed data.

Figure 3: GHG-Emissions (kgCO2eq per 1000kcal) for plant-based and animal-based food categories.

3.1 Comparing Example Diets

We now take an even more granular view at the nutritional value of (non) plant-based diets by considering two typical diets with respect to recommended macro-nutrient make-up (carbohydrates, fat, protein) according to [8].

Diet calories and nutrients composition The consumer of these diets is an average woman, whose daily calories intake is around 2000kcal [9]. The division amongst macro-nutrients was taken from Walle [8] and is depicted in Table 1.

Table 1: Average woman's daily nutrients intake

Macronutrient	Suggested percentage of calories	Daily grams for the average woman
Carbohydrates	$45\% - 65\%$	225g-325g
Fats	$20\% - 35\%$	45g-78g
Proteins	$10\% - 35\%$	50g-175g

What the example diets contain The following diets shown in Table 2 and Table 3, are described by the amount of food category consumed in a day in order to reach the suggested intake of nutrients, thus not considering specific recipes or meals. Note that food categories are named the same as they are named in the dataset used in this research, while nutritional facts of the different foods can be found on various health websites such as [8].

Emissions calculation The total emissions of a daily diet is computed by the sum of the single contributions of every food category in kgCO2eq per kg, each calculated in proportion of the amount consumed in the diet.

Table 2: Exemplary Plant-based Diet

Food Category	Grams
Wheat & Rye (Bread)	400
Tofu	300
Bananas	300
Peas	200
Nuts	50
Soymilk	150

Table 3: Exemplary Omnivor Diet

Food Category	Grams
Wheat & Rye (Bread)	400
Pig Meat	200
Bananas	300
Cheese	150

4 Results

Table 4: Results of Welch's Hypothesis Test

Feature	Test-Statistic	p-value
Tot.Emissions	≈ 6.49	$2.2 \cdot 10^{-6}$
GHG	≈ 5.03	$1.4 \cdot 10^{-5}$

Based on statistical analysis we found that not only total emissions but also GHG emissions when accounting for nutritional value (in kgCO2eq per 1000kcal) have a p-value of $p = 2.2 \cdot 10^{-6}$ and $p = 1.4 \cdot 10^{-5}$ when conducting the unequal variance test with $\alpha = 0.01$ (see Table 4). With regards to the diet examples comparison, as seen in Table 5, the plant-based diet emits 2.02 kgCO2eq while the omnivor one emits 5.42 kgCO2eq, i.e. 268% of the first one.

5 Limitations

Limitations start in the data collection process and stretch into the analysis part, both of which strongly influence the study's plausibility and impact. Some critical points will be briefly discussed for both aspects, followed by a brief outlook on possible future work in Section 6.

Table 5: Results of diet emissions calculation

Diet	Emissions (kgCO2eq)
Plant-based	2.02
Omnivor	5.42

Limitations in the data A wide variety of fruits are available year-round in Western grocery stores irrespective as to whether they are in season or are native to the Northern hemisphere.[3] Seasonality is not considered in the data collection process but clearly affects both energy for cooling and potentially transport (when imported in off-seasons) and hence affects total emissions. A potential solution to control for this confounding variable *seasonality* could be, to first introduce this information in the data-collection process and second, to condition on this variable in the analysis part. Furthermore, it is arguably up for discussion as to how different greenhouse gases are weighted when calculating the "carbon dioxide-equivalents (CO_2eq)" as described in Section 2. For instance, methane is a much shorter-lived but more powerful greenhouse gas than CO_2.

Limitations in the analysis First of all, the assumption of Gaussian-distributed data may be criticized and subsequently, the T-Test used, which implicitly assumes Gaussian distributed data. In particular, the data of plant-based food depicted in Figure 3b seems to follow a bimodal distribution as opposed to a Gaussian distribution. Second, a more granular analysis regarding the variance within each food-group would be interesting. This could elude to the differences of producers regarding greenhouse gas emissions and may therefore provide the opportunity to further minimize emissions during the food-production process by optimizing towards the lowest-emitting producers.

6 Conclusion

Our results demonstrate a significant difference between plant- and animal-based products, not only with respect to their total emissions but also when accounting for nutritional value. Thus, plant-based diets not only seem to satisfy nutritional requirements [2] but importantly, also would contribute significantly to solving climate change when consumed by many. Importantly, eating more plant-based is a clear rule that would help individuals willing to contribute to solving climate change, to make more informed decisions in their daily life.

Outlook and future work A more detailed analysis of emissions by country would reveal inefficiencies and aid in reducing global warming. Vegan diets, while potentially lacking in certain micronutrients, usually meet protein intake recommendations and account for 90% of global calorie and protein intake.[2]. However, people on vegan diets should be aware of potential micronutrient deficiencies and further research is needed in this area.

[3]Note that this study's authors are Europeans and hence everything is written though the perspective of such but likely also applies to North Americans.

References

[1] J. Poore and T. Nemecek. Reducing food's environmental impacts through producers and consumers. *Science*, 360(6392):987–992, 2018. doi: 10.1126/science.aaq0216. URL https://www.science.org/doi/abs/10.1126/science.aaq0216.

[2] WHO European Office. Plant-based diets and their impact on health, sustainability and the environment: a review of the evidence. Online, 2021. URL https://apps.who.int/iris/bitstream/handle/10665/349086/WHO-EURO-2021-4007-43766-61591-eng.pdf?sequence=1&isAllowed=y. Retrieved Jan, 18 2023.

[3] Debra C. Roberts. Climate change 2022: Impacts, adaptation and vulnerability, 2022. URL https://report.ipcc.ch/ar6/wg2/IPCC_AR6_WGII_FullReport.pdf.

[4] Hannah Ritchie and Max Roser. Environmental impacts of food production. *Our World in Data*, 2020. URL https://ourworldindata.org/environmental-impacts-of-food.

[5] M Crippa, E Solazzo, D Guizzardi, F Monforti-Ferrario, FN Tubiello, and A Leip. Food systems are responsible for a third of global anthropogenic ghg emissions. *Nature Food*, 2(3):198–209, 2021.

[6] Bernard L Welch. The generalization of "student's" problem when several different population variances are involved. *Biometrika*, 34 (1-2):28–35, 1947.

[7] Seaborn: statistical data visualization. Online. URL https://seaborn.pydata.org/.

[8] Gavin Van De Walle. The best macronutrient ratio, 2018. URL https://www.healthline.com/nutrition/best-macronutrient-ratio.

[9] NHS. Understanding calories, 2019. URL https://www.nhs.uk/live-well/healthy-weight/managing-your-weight/understanding-calories.