Olakunle Capstone Project

Table of contents

1.0. Introduction

1.1. Background of the project

The conversion of forest land into agricultural products significantly contributes to greenhouse gas emissions, primarily through the release of carbon dioxide (CO2) stored in trees and soil. Deforestation often involves the removal of vegetation, which not only eliminates a carbon sink but also disrupts soil carbon stocks, leading to further emissions (Koga et al., 2020). Activities such as land clearing and burning biomass release large amounts of CO2 into the atmosphere (Garofalo et al., 2022). Additionally, land-use changes can exacerbate soil degradation, reducing its ability to sequester carbon, as highlighted in studies examining emissions from agricultural land conversion (Tubiello et al., 2021). The specific type of agricultural practice also influences emissions; for instance, converting peat swamp forests to oil palm plantations can lead to elevated greenhouse gas emissions due to the high carbon content of peat soils (Cooper et al., 2020). Overall, these factors underscore the environmental impact of agricultural expansion on carbon emissions.

1.2. Aim and Objectives of the project

This project is about the emission that is generated from converting forest land to rice, wheat, and other cereal products for different countries. The aim of the project is to find out if there is a significant difference in the emission between the product Rice, Wheat and Other Cereals) by performing relevant exploratory data analysis to understand the data and using relevant inferential statistics (ANOVA). This data for the project is an extract from Ritchie (2021), and it is available here (data)

Figure 1: image of the project

2.0. Importing and Loading the Data

2.1. Importing the Data

```
library(tidyverse)
library(readxl)
library(knitr)
project_data <- read_csv("emission_data.csv")
```

2.2. Previewing the data

```
head(project_data) %>% kable()
```

Table 1: emission data preview

entity	code	products	emission	per_capital_emission
Australia	AUS	Rice	879389.07	0.5242982

Table 1: emission data preview

entity	code	products	emission	per_capital_emission
Australia	AUS	Wheat	41496.59	0.5242982
Australia	AUS	Other Cereals	89034.21	0.5242982
Austria	AUT	Rice	184118.42	0.3532005
Austria	AUT	Wheat	15495.06	0.3532005
Austria	AUT	Other Cereals	30146.20	0.3532005

Table 1 reveals a preview of the first 6 entries of the emission data

3.0. Converting the emission from tonnes to kilotonnes

```
project_data$emission_kilotonne <- project_data$emission / 1000

head(project_data, 15) %>% kable()
```

Table 2: emission in kilotonnes created

entity	code	products	emission	per_capital_emission	emission_kilotonne
Australia	AUS	Rice	879389.071	0.5242982	879.389071
Australia	AUS	Wheat	41496.595	0.5242982	41.496595
Australia	AUS	Other Cereals	89034.207	0.5242982	89.034207
Austria	AUT	Rice	184118.422	0.3532005	184.118422
Austria	AUT	Wheat	15495.064	0.3532005	15.495064
Austria	AUT	Other Cereals	30146.200	0.3532005	30.146200
Belgium	BEL	Rice	458813.420	0.9707366	458.813420
Belgium	BEL	Wheat	50684.962	0.9707366	50.684962
Belgium	BEL	Other Cereals	86200.379	0.9707366	86.200379
Brazil	BRA	Rice	10277208.350	2.7096048	10277.208350
Brazil	BRA	Wheat	3112053.787	2.7096048	3112.053787
Brazil	BRA	Other Cereals	2329114.419	2.7096048	2329.114419
Bulgaria	BGR	Rice	32811.580	0.0889054	32.811580
Bulgaria	BGR	Wheat	2691.416	0.0889054	2.691416
Bulgaria	BGR	Other Cereals	5368.398	0.0889054	5.368398

Table 2 reveals a preview (the first 15 entries) of the project data with a new variable, "emission kilotonne" created. The new variable was created so as not to override the existing "emission" variable.

4.0. Exploratory Data Analysis

4.1. Summary of the project data

```
summary(project_data)
```

```
    entity               code             products            emission       
 Length:132         Length:132         Length:132         Min.   :    1309  
 Class :character   Class :character   Class :character   1st Qu.:   17533  
 Mode  :character   Mode  :character   Mode  :character   Median :   79844  
                                                          Mean   :  933255  
                                                          3rd Qu.:  316456  
                                                          Max.   :62291319  
 per_capital_emission emission_kilotonne
 Min.   :0.06882      Min.   :    1.31  
 1st Qu.:0.20271      1st Qu.:   17.53  
 Median :0.38503      Median :   79.84  
 Mean   :0.50759      Mean   :  933.25  
 3rd Qu.:0.50162      3rd Qu.:  316.46  
 Max.   :2.77977      Max.   :62291.32  
```

The summary output shows that the project data set consist of 132 observations across five variables namely, "entity", "code", "products", "emission", "per_capital_emission", and "emission_kilotonne". The first three variables (entity, code, and products) are categorical, each recorded as character types. The remaining three variables—emission, per_capital_emission, and emission_kilotonne—are numeric, and their summary statistics show a wide range in values. Emissions range from 1,309 to 62,291,319, while per capita emissions vary from 0.0688 to 2.7798, and emission in kilotonnes ranges from 1.31 to 62,291.32. The distribution indicates significant variability in emissions across different entities and products.

4.2. Univariate Analysis

4.2.1. Visualizing the emission variable

```
ggplot(project_data, aes(x = emission_kilotonne)) +
  geom_density(fill = "lightcoral", alpha = 0.5) +
  theme_minimal() +
  labs(title = "Density Plot of Emission
```

```
       (in Kilotonnes)"
       , x = "Emission (kilotonnes)", y = "Density")
```

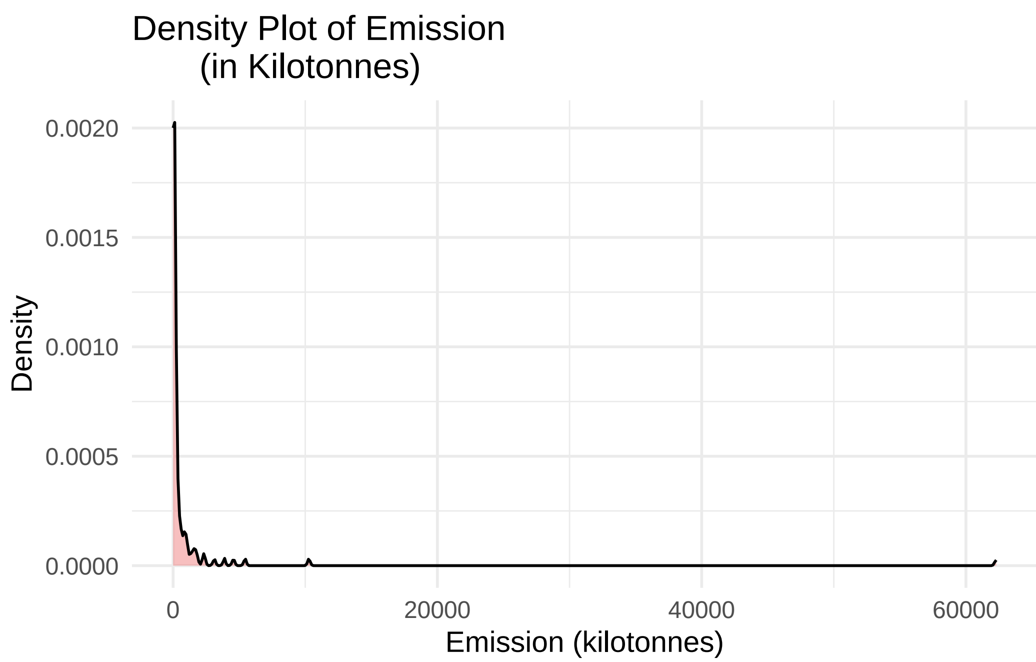

Figure 2

@ fig-count-emission shows the distribution of emission (in kilotonnes) values. The x-axis represents the emission values in kilotonnes, and the y-axis represents the density of the emission values. The shape of the curve reveals the underlying distribution of the data. It can be observed that the emission data is heavily skewed to the right, with a single peak suggesting a unimodal distribution. This indicates that most of the emission values are relatively low, but there are a few very high emission values that are pulling the mean to the right.

4.2.2. Visualizing the per_capital_emission variable

```
ggplot(project_data, aes(x = per_capital_emission)) +
  geom_density(fill = "lightcoral", alpha = 0.5) +
  theme_minimal() +
  labs(title =
         "Density Plot of the per_capital_emission",
       x = "per_capital_emission", y = "Density")
```

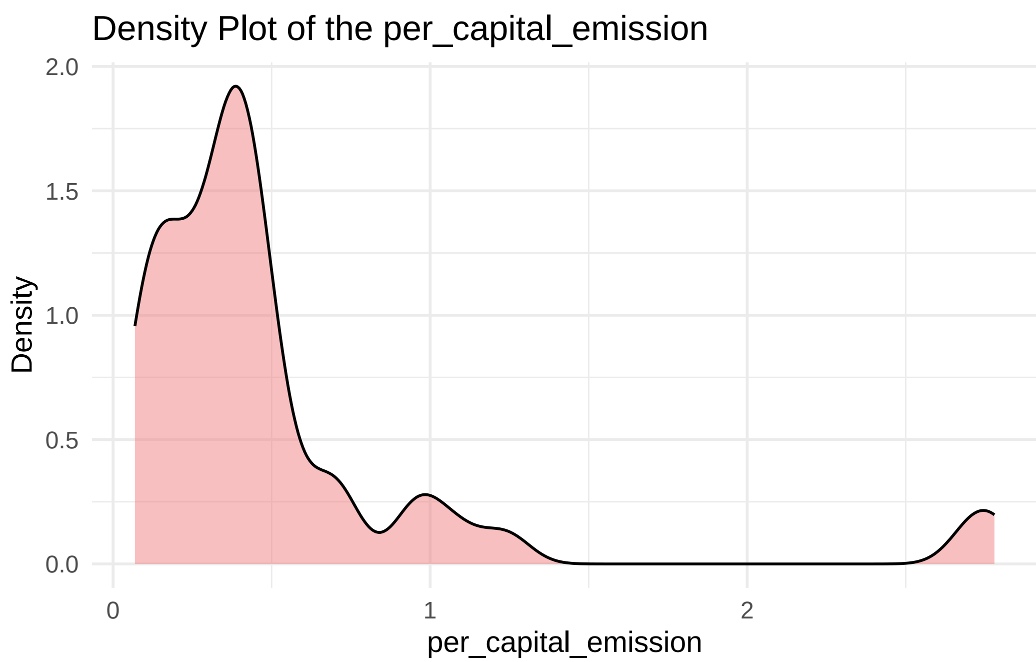

Figure 3: univariate analysis showing the distribution of the per_capital_emission

@ fig-count-percapitalemission shows the distribution of per capital emission values. The x-axis represents the per capital emission values, and the y-axis represents the density of the per capital emission values. The shape of the curve reveals the underlying distribution of the data. As revealed by the density plot, the per capital emission data is multimodal, with evidence of distinct groups or clusters. The slight right skew indicates that there are a few higher per capital emission values, but the majority of the values are concentrated around the peaks of the distribution. This suggests that there are different patterns or trends in per capital emissions among the groups represented in the data.

4.3. Bivariate Analysis

4.3.1. Visualizing emission across the entities (countries)

```
ggplot(project_data, aes(x = entity, y = emission_kilotonne)) +
  geom_bar(stat="identity") +
  theme_minimal() +
  labs(title = "Comparison of Emission by Entity",
       x = "Entity",
```

```
      y = "Emission (Kilotonnes)") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

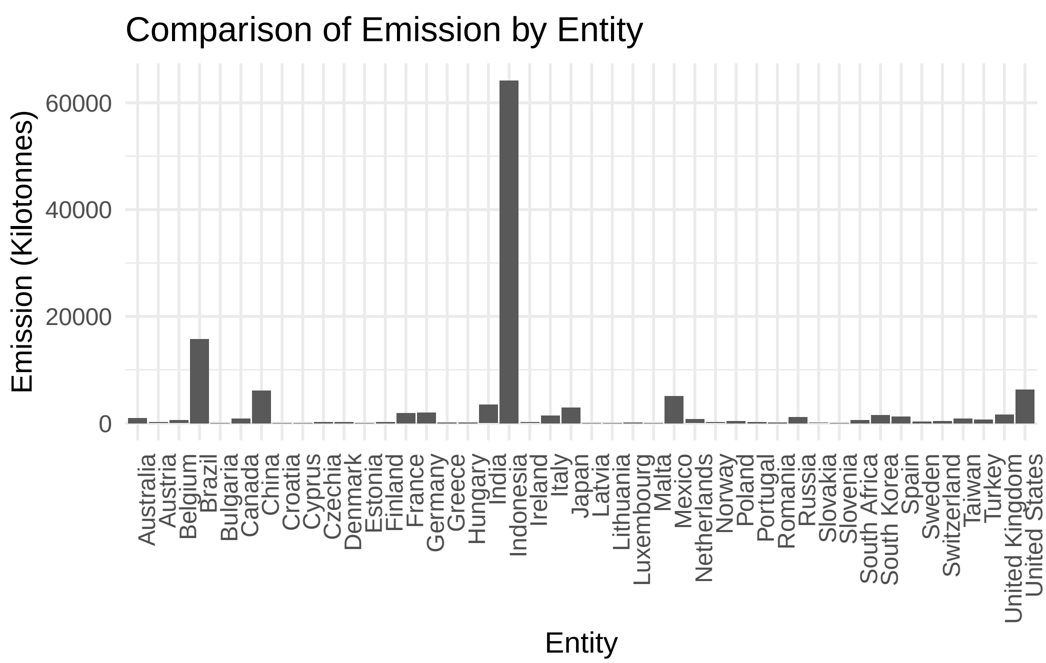

Figure 4: bivariate analysis showing the comparison of the emission values across the enities

Figure 4 shows the comparison of emission values (in kilotonnes) for different countries (entities). The x-axis represents the countries, and the y-axis represents the emission values. The height of each bar corresponds to the emission value for the respective country. The bar chart clearly shows that China and the United States are the major contributors to emissions among the listed countries. The majority of the other countries have significantly lower emissions. The wide range of emission values and the clustering of some countries suggest that there are different factors influencing emissions in different regions.

4.3.2. Visualizing emission across products

```
ggplot(project_data, aes(x = products, y = emission_kilotonne)) +
  geom_bar(stat="identity") +
  theme_minimal() +
  labs(title = "Comparison of Emission by Product",
      x = "Product",
```

```
      y = "Emission (Kilotonnes)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

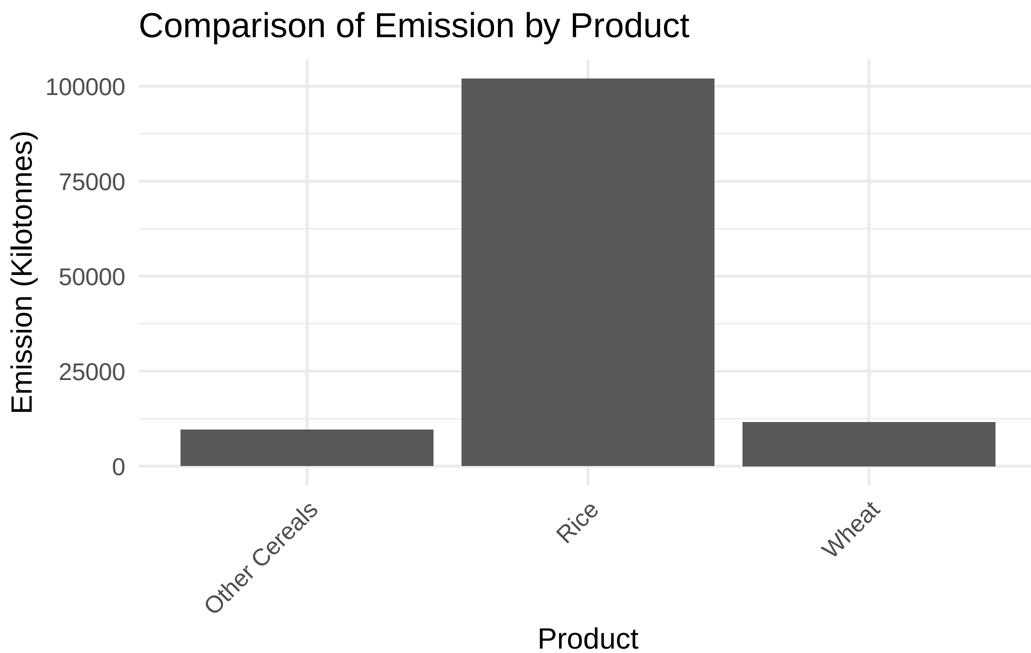

Figure 5: bivariate analysis showing the comparison of the emission values across the products

Figure 5 shows a comparison of emission (in kilotonnes) by the product. The x-axis represents the different products (Other Cereals, Rice, and Wheat), and the y-axis represents the emission. The height of each bar corresponds to the emission value for the respective product. The bar chart clearly shows that Rice is the product with the highest emissions among the three categories. Other Cereals and Wheat have considerably lower emission levels. This suggests that the production of Rice contributes significantly more to greenhouse gas emissions compared to the production of Other Cereals and Wheat.

4.4. Multivariate Analysis

```
ggplot(project_data, aes(x=entity, y=emission_kilotonne, fill=products)) +
  geom_bar(stat="identity") +
  theme_minimal() +
  labs(title="Emissions Breakdown by Products within Entities",
```

```
    y="Emissions", x="Entity") +
  theme(axis.text.x = element_text(angle=90, hjust=1, size=6))
```

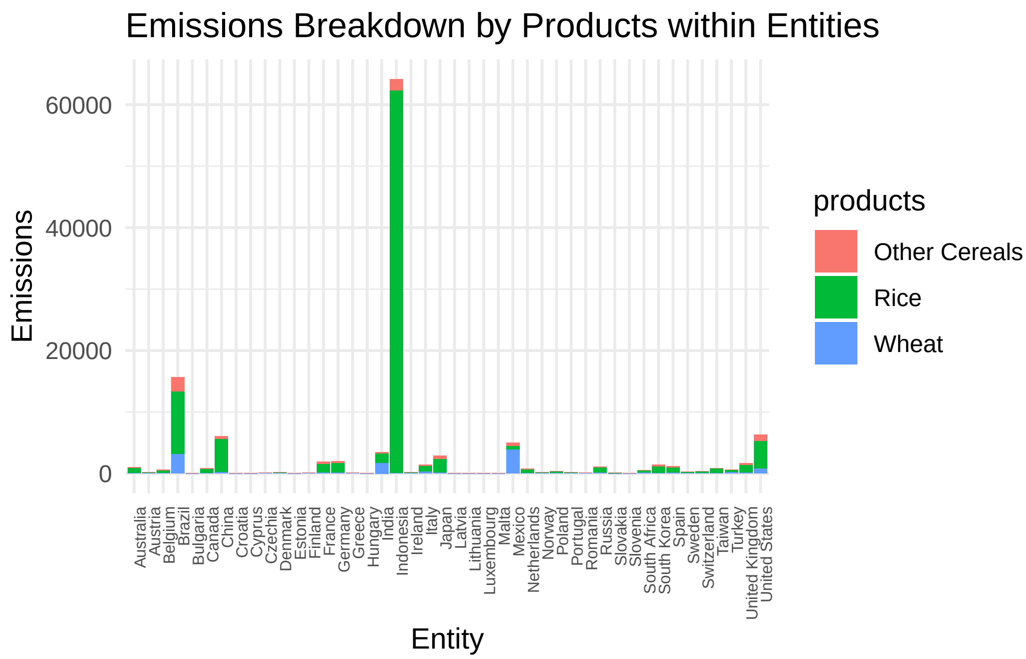

Figure 6: multivariate analysis showing the emissions of the entities (countries) by their products

Figure 6 shows the emissions breakdown by products within different entities. Rice appears to be the product with the highest emissions in most entities, as evidenced by the tall green bars. China has the highest overall emissions, with tall bars for all three products. Other Cereals and Wheat have lower emissions overall, but there are some entities where they contribute significantly to the total emissions. Overall, the chart provides a visual representation of how emissions are distributed among different products within various entities. It highlights the dominant role of Rice in emissions and the significant contribution of China to overall emissions. By analyzing the chart, you can identify specific products and entities that are major contributors to emissions and explore potential areas for reducing emissions.

5.0. Inferential Statistics

5.1. ANOVA Test

```
anova_results <- aov(emission_kilotonne ~ products, data = project_data)

anova_results
```

```
Call:
   aov(formula = emission_kilotonne ~ products, data = project_data)

Terms:
                 products  Residuals
Sum of Squares  126420729 3855885915
Deg. of Freedom         2        129

Residual standard error: 5467.229
Estimated effects may be unbalanced
```

```
summary(anova_results)
```

```
             Df    Sum Sq  Mean Sq F value Pr(>F)
products      2 1.264e+08 63210365   2.115  0.125
Residuals   129 3.856e+09 29890588
```

5.2. Interpretation of the result

The ANOVA test conducted to see if there is a difference in the emission between the product shows that since this p-value is greater than 0.05, the null hypothesis is accepted meaning that there is no statistically significant difference in emissions between the products at the 5% significance level.

6.0. Summary

This project analyzes the emissions generated from converting forest land into agricultural products, specifically rice, wheat, and other cereals, across various countries. Utilizing data sourced from Ritchie (2021), the aim is to determine whether there are significant differences in emissions among these product groups. The project involves several stages: importing and preparing the data, conducting exploratory data analysis (EDA) to understand the distribution of emissions, and applying inferential statistics, particularly ANOVA, to test for significant differences in emissions between the product categories. Through visualizations and statistical analysis, insights into the variability and contributing factors of emissions are explored.

7.0. Conclusion

The analysis revealed that while emissions varied widely across different entities and products, the ANOVA test indicated no statistically significant difference in emissions among rice, wheat, and other cereals at the 5% significance level. This suggests that, despite differences in individual emission values, the overall emissions from these agricultural products are not significantly different from one another. This outcome emphasizes the need for further research.

8.0. References

Garofalo, D. F. T., Novaes, R. M. L., Pazianotto, R. A., Maciel, V. G., Brandão, M., Shimbo, J. Z., & Folegatti-Matsuura, M. I. (2022). Land-use change CO2 emissions associated with agricultural products at municipal level in Brazil. *Journal of Cleaner Production*, *364*, 132549.

Koga, N., Shimoda, S., Shirato, Y., Kusaba, T., Shima, T., Niimi, H., … & Atsumi, K. (2020). Assessing changes in soil carbon stocks after land use conversion from forest land to agricultural land in Japan. *Geoderma*, *377*, 114487.

Tubiello, F. N., Conchedda, G., Wanner, N., Federici, S., Rossi, S., & Grassi, G. (2021). Carbon emissions and removals from forests: new estimates, 1990–2020. *Earth System Science Data*, *13*(4), 1681-1691.

Cooper, H. V., Evers, S., Aplin, P., Crout, N., Dahalan, M. P. B., & Sjogersten, S. (2020). Greenhouse gas emissions resulting from conversion of peat swamp forest to oil palm plantation. *Nature communications*, *11*(1), 407.