CO2 Emission from Cereal Food Production

Nomvume Sandi

Introduction

The aim of the project is to compare the total amount of emissions produced by countries and the product contribution due to converting from Forest to agriculture land. The data extracted from Ritchie(2021), gives an overview of countries emission , agriculture land use which are rice , wheat and other cereals. Lastly the deforestation emission per capita. Various statistical analysis will be performed to show variation and impact and trends of the deforestation to CO2 emissions. The data comprises of countries (entity and code) , amount of emissions by products (emissions and per capital emission).

Figure 1: Figure1: Indian Rice farm

```
library(tidyverse)
library(skimr)
library(knitr)
```

```
url <- paste0("https://raw.githubusercontent.com/xrander/bootcamp-test/",
              "master/data/emission_data.csv")

emission_data <- read_csv(url)
```

```
Rows: 132 Columns: 5
```

```
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): entity, code, products
dbl (2): emission, per_capital_emission

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
emission_data <- emission_data |>
  mutate(kilo_emission =emission/1000)
```

Exploratory data analysis

In the following analysis it will give an overview of the data , what data comprises of. To view the top and last rows of the data which is a quick snapshot. Summary provides 5 number summary min, max,mean ,median and quartiles.

```
head(emission_data) |>
  kable()
```

entity	code	products	emission	per_capital_emission	kilo_emission
Australia	AUS	Rice	879389.07	0.5242982	879.38907
Australia	AUS	Wheat	41496.59	0.5242982	41.49659
Australia	AUS	Other Cereals	89034.21	0.5242982	89.03421
Austria	AUT	Rice	184118.42	0.3532005	184.11842
Austria	AUT	Wheat	15495.06	0.3532005	15.49506
Austria	AUT	Other Cereals	30146.20	0.3532005	30.14620

```
tail(emission_data) |>
  kable()
```

entity	code	products	emission	per_capital_emission	kilo_emission
United Kingdom	GBR	Rice	1279478.7	0.3912465	1279.4787
United Kingdom	GBR	Wheat	119004.5	0.3912465	119.0045
United Kingdom	GBR	Other Cereals	248065.5	0.3912465	248.0655

entity	code	products	emission	per_capital_emission	kilo_emission
United States	USA	Rice	4564029.4	0.3465427	4564.0294
United States	USA	Wheat	761894.5	0.3465427	761.8945
United States	USA	Other Cereals	1027923.6	0.3465427	1027.9236

```
str(emission_data)
```

```
tibble [132 x 6] (S3: tbl_df/tbl/data.frame)
 $ entity              : chr [1:132] "Australia" "Australia" "Australia" "Austria" ...
 $ code                : chr [1:132] "AUS" "AUS" "AUS" "AUT" ...
 $ products            : chr [1:132] "Rice" "Wheat" "Other Cereals" "Rice" ...
 $ emission            : num [1:132] 879389 41497 89034 184118 15495 ...
 $ per_capital_emission: num [1:132] 0.524 0.524 0.524 0.353 0.353 ...
 $ kilo_emission       : num [1:132] 879.4 41.5 89 184.1 15.5 ...
```

To compare the emissions regardless of the product factor comparing the minimum and maximum , the max is 47 times more than the min which shows variation of the emissions produced.

```
summary(emission_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 kilo_emission
 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
```

```
skim(emission_data)
```

Table 3: Data summary

Name	emission_data
Number of rows	132
Number of columns	6
_____	
Column type frequency:	
character	3
numeric	3
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
entity	0	1	5	14	0	44	0
code	0	1	3	3	0	44	0
products	0	1	4	13	0	3	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
emission	0	1	933254.92	2513554.87	1308.84	17532.86	79843.73	316456.06	12291319.14	
per_capital_emission	0	1	0.51	0.56	0.07	0.20	0.39	0.50	2.78	
kilo_emission	0	1	933.25	5513.55	1.31	17.53	79.84	316.46	62291.32	

Visualization

The analysis will allow a visualization of countries with the highest amount of the emissions will only show the most 10 countries with highest emissions and the countries with lowest emission. Lastly a summary of products contributing to the emissions.

```
total_emissions <- emission_data|>
  group_by(code) |>
  summarize(total_emission = sum(kilo_emission, na.rm = TRUE))

total_emissions |>
  kable()
```

code	total_emission
AUS	1009.91987
AUT	229.75968
BEL	595.69876
BGR	40.87139
BRA	15718.37656
CAN	882.78948
CHE	393.39554
CHN	6109.72339
CYP	25.70653
CZE	158.37562
DEU	1988.00595
DNK	199.78931
ESP	1196.69436
EST	33.11000
FIN	147.27767
FRA	1885.50176
GBR	1646.54864
GRC	129.02292
HRV	36.09888
HUN	80.17478
IDN	64161.32172
IND	3471.96060
IRL	224.79825
ITA	1462.99509
JPN	2932.99163
KOR	1473.68790
LTU	33.60716
LUX	79.63687
LVA	19.24015
MEX	5021.20101
MLT	14.47665
NLD	799.34081
NOR	244.03889
POL	398.85900
PRT	222.55816
ROU	106.11337
RUS	1128.60563
SVK	98.91165
SVN	38.10742
SWE	270.59773
TUR	656.27742

code	total_emission
TWN	901.87440
USA	6353.84751
ZAF	567.75907

```
top_20_emissions <- total_emissions |>
  slice_max(order_by = total_emissions, n = 20)

top_20_emissions |>
  kable()
```

code	total_emission
ZAF	567.75907
USA	6353.84751
TWN	901.87440
TUR	656.27742
SWE	270.59773
SVN	38.10742
SVK	98.91165
RUS	1128.60563
ROU	106.11337
PRT	222.55816
POL	398.85900
NOR	244.03889
NLD	799.34081
MLT	14.47665
MEX	5021.20101
LVA	19.24015
LUX	79.63687
LTU	33.60716
KOR	1473.68790
JPN	2932.99163

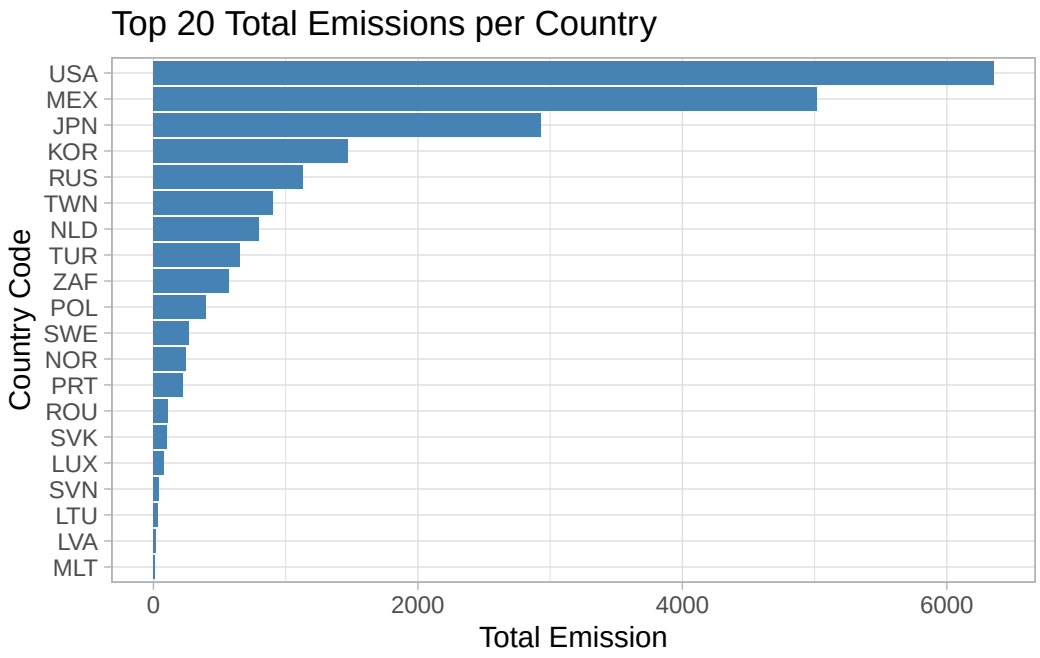
```
Bottom_20_emission <- total_emissions |>
  slice_min(order_by = total_emissions, n = 20)

Bottom_20_emission |>
  kable()
```

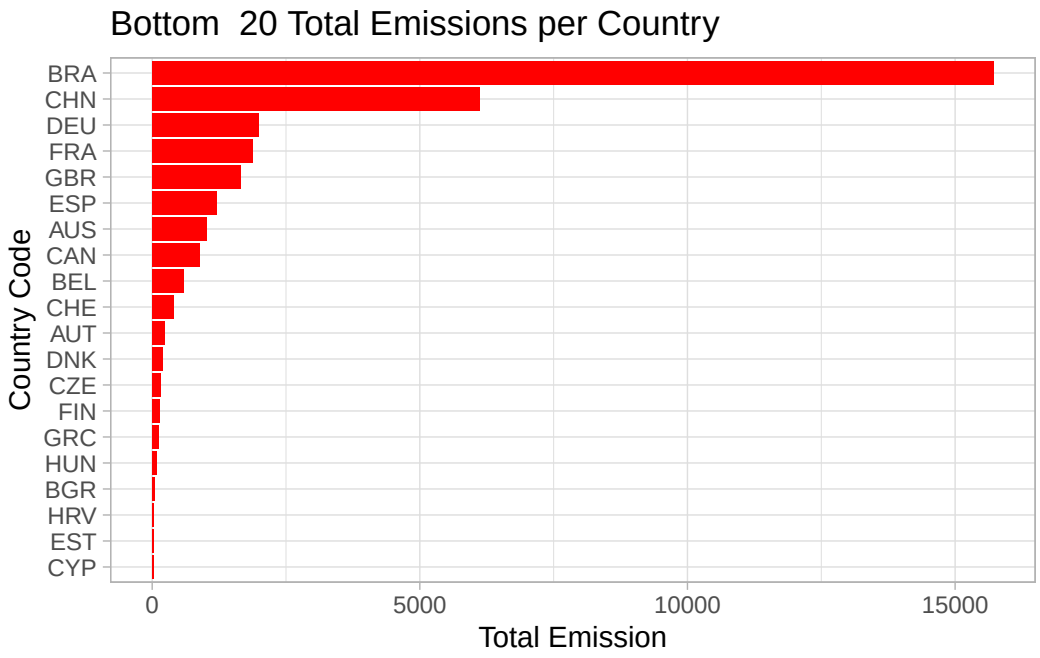
code	total_emission
AUS	1009.91987
AUT	229.75968
BEL	595.69876
BGR	40.87139
BRA	15718.37656
CAN	882.78948
CHE	393.39554
CHN	6109.72339
CYP	25.70653
CZE	158.37562
DEU	1988.00595
DNK	199.78931
ESP	1196.69436
EST	33.11000
FIN	147.27767
FRA	1885.50176
GBR	1646.54864
GRC	129.02292
HRV	36.09888
HUN	80.17478

The results shows America followed by Mexico and Japan in emission production regardless of the product they emit with have highest production of emissions. In the bottom 20 with the least amount of emission is Cyprus, Estonia and Croatia respectively. With the Rice conversion contributing the most to the emissions.

```
ggplot(top_20_emissions, aes(x = reorder(code, total_emission), y = total_emission)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(
    x = "Country Code",
    y = "Total Emission",
    title = "Top 20 Total Emissions per Country") +
  coord_flip() +
  theme_light()
```




```
ggplot(Bottom_20_emission, aes(x = reorder(code, total_emission), y = total_emission)) +
  geom_bar(stat = "identity", fill = "red") +
  labs(
    x = "Country Code",
    y = "Total Emission",
    title = "Bottom  20 Total Emissions per Country"
  ) +
  coord_flip() +
  theme_light()
```



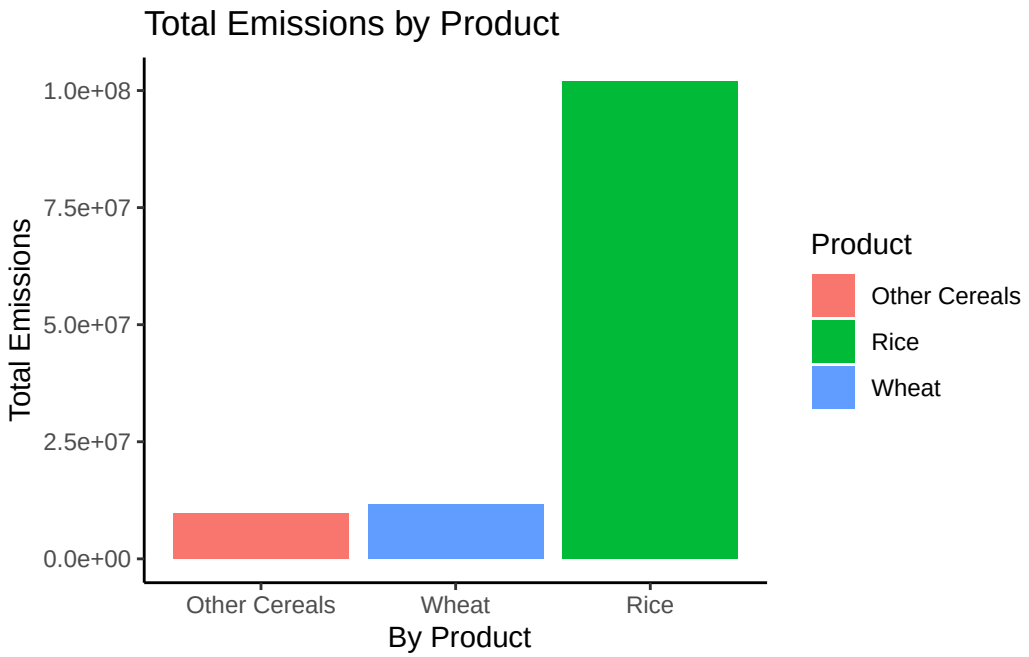

```
product_emissions <- emission_data |>
  group_by(products) |>
  summarize(total_emission = sum(emission, na.rm = TRUE))

product_emissions |>
  kable()
```

products	total_emission
Other Cereals	9637471
Rice	101948790
Wheat	11603388

```
ggplot(
  product_emissions,
  aes(
    x = fct_reorder(products, total_emission),
    y = total_emission,
    fill = products)
) +
  geom_bar(stat = "identity") +
  labs(
    title = "Total Emissions by Product",
```

```
    x = " By Product",
    y = "Total Emissions",
    fill = "Product"
  ) +
  theme_classic()
```



Inferential Statistical analysis

The Anova analysis conducted to test difference between products and emissions.

- **Null Hypothesis (H0)**: There is no significant difference in mean emissions between product and emission produced
- **Alternative Hypothesis (H1)**: There is a significant difference in mean emissions between product and emission produced.

```
anova_result <- aov(kilo_emission ~ products, data = emission_data)

summary(anova_result)
```

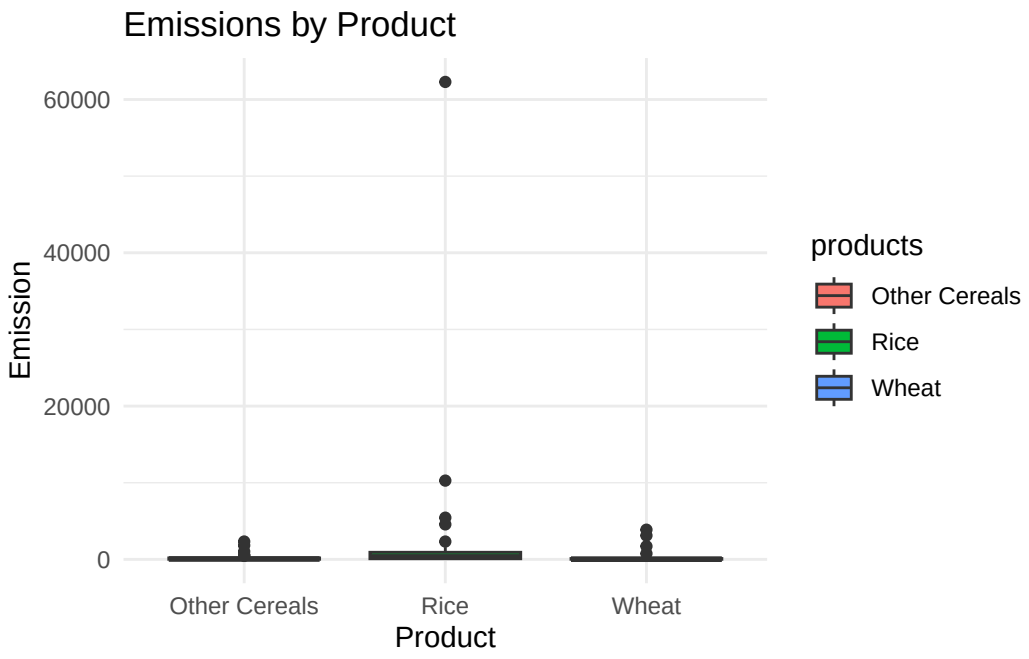
```
             Df    Sum Sq  Mean Sq F value Pr(>F)
products      2 1.264e+08 63210365   2.115  0.125
Residuals   129 3.856e+09 29890588
```

F-value: The F-value of 2.115 is the ratio of the variance between the groups (products) to the variance within the groups (residuals).

p-value: The p-value of 0.125 is greater than the commonly used significance level of 0.05. This suggests that the observed differences among group means are not statistically significant.

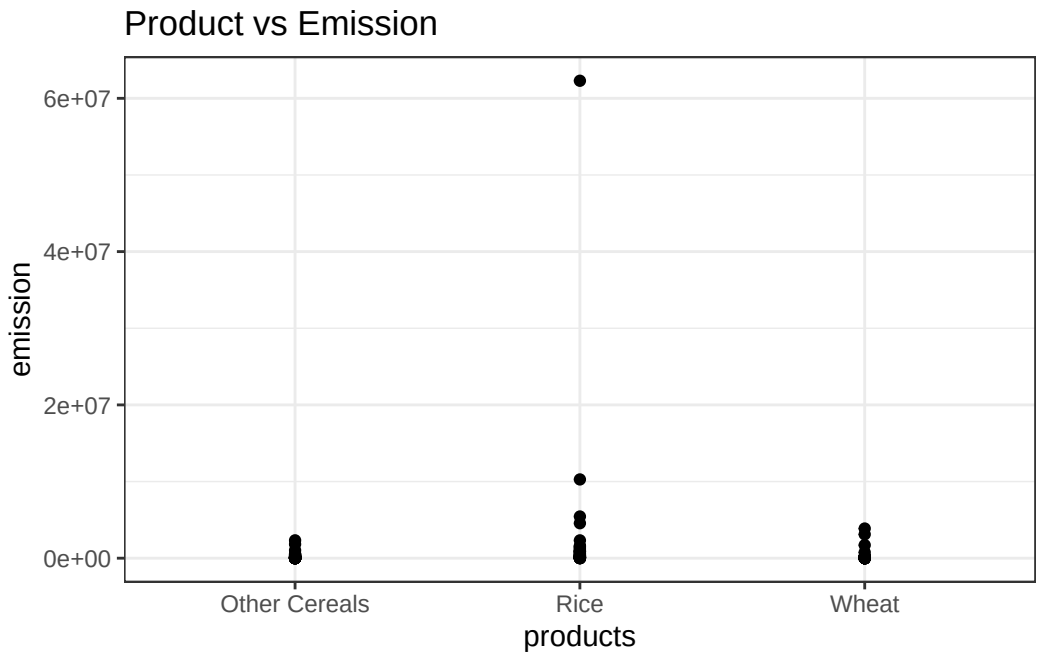
Since the p-value (0.125) is greater than 0.05 we **fail to reject the null hypothesis**,this mean emissions do not differ significantly between the products.

```
ggplot(emission_data, aes(x = products, y = kilo_emission, fill = products)) +
  geom_boxplot() +
  labs(
    title = "Emissions by Product",
    x = "Product",
    y = "Emission"
  ) +
  theme_minimal()
```




```
emission_data |>
  ggplot(aes(products, emission)) +
  geom_point() +
  geom_smooth( se = FALSE, method = "lm" ) +
  ggtitle("Product vs Emission") +
  theme_bw()
```

```
`geom_smooth()` using formula = 'y ~ x'
```



Conclusion

The amount of emission are influenced by the products individually the is no link in terms of products. Each products have its own contribution to the emission in general. Also the increase in per_capita_emission is the contribution of the deforestation.