An R Markdown for environmental-footprints

Context:

As the world’s population has expanded and gotten richer, the demand for food, energy and water has seen a rapid increase. Not only has demand for all three increased, but they are also strongly interlinked: food production requires water and energy; traditional energy production demands water resources; agriculture provides a potential energy source. This project focuses on the environmental impacts of food. Ensuring everyone in the world has access to a nutritious diet in a sustainable way is one of the greatest challenges we face.

Questions:

- Which types of food have more negative impact on the environment?
- What types of food production should be encouraged to consume nutritious diet in a sustainable way?
- Which stage of food production contributes more to the greenhouse gas emission?
- Compare carbon footprint of plant-based foods?
- Compare carbon footprint of animal-based foods?
- Compare carbon footprint of protein rich foods?

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.6     ✓ purrr   0.3.4
## ✓ tibble  3.1.7     ✓ dplyr   1.0.9
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

1. Loading data from a CSV file

```
data<-read_csv("./input/environment-impact-of-food-production/Food_Production.csv")
```

```
## Rows: 43 Columns: 23
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (1): Food product
## dbl (22): Land use change, Animal Feed, Farm, Processing, Transport, Packgin...
##
## 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.
```

```
head(data)
```

```
## # A tibble: 6 × 23
##   `Food product`      `Land use change` `Animal Feed`  Farm Processing Transport
##   <chr>                           <dbl>         <dbl> <dbl>      <dbl>     <dbl>
## 1 Wheat & Rye (Bread)               0.1             0   0.8        0.2       0.1
## 2 Maize (Meal)                      0.3             0   0.5        0.1       0.1
## 3 Barley (Beer)                     0               0   0.2        0.1       0  
## 4 Oatmeal                           0               0   1.4        0         0.1
## 5 Rice                              0               0   3.6        0.1       0.1
## 6 Potatoes                          0               0   0.2        0         0.1
## # … with 17 more variables: Packging <dbl>, Retail <dbl>,
## #   Total_emissions <dbl>,
## #   `Eutrophying emissions per 1000kcal (gPO₄eq per 1000kcal)` <dbl>,
## #   `Eutrophying emissions per kilogram (gPO₄eq per kilogram)` <dbl>,
## #   `Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)` <dbl>,
## #   `Freshwater withdrawals per 1000kcal (liters per 1000kcal)` <dbl>,
## #   `Freshwater withdrawals per 100g protein (liters per 100g protein)` <dbl>, …
```

2. Cleaning the data

Investigating the structure of data

```
str(data)
```

```
## spec_tbl_df [43 × 23] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Food product                                                              : chr [1:43] "Wheat
& Rye (Bread)" "Maize (Meal)" "Barley (Beer)" "Oatmeal" ...
##  $ Land use change                                                           : num [1:43] 0.1 0.
3 0 0 0 0 0.6 1.2 0 0 ...
##  $ Animal Feed                                                               : num [1:43] 0 0 0
0 0 0 0 0 0 0 ...
##  $ Farm                                                                      : num [1:43] 0.8 0.
5 0.2 1.4 3.6 0.2 0.2 0.5 0.5 1.1 ...
##  $ Processing                                                                : num [1:43] 0.2 0.
1 0.1 0 0.1 0 0 0 0.2 0 ...
##  $ Transport                                                                 : num [1:43] 0.1 0.
1 0 0.1 0.1 0.1 0.1 0.8 0.6 0.1 ...
##  $ Packging                                                                  : num [1:43] 0.1 0.
1 0.5 0.1 0.1 0 0 0.1 0.1 0.4 ...
##  $ Retail                                                                    : num [1:43] 0.1 0
0.3 0 0.1 0 0 0 0 0 ...
##  $ Total_emissions                                                           : num [1:43] 1.4 1.
1 1.1 1.6 4 0.3 0.9 2.6 1.4 1.6 ...
##  $ Eutrophying emissions per 1000kcal (gPO₄eq per 1000kcal)                  : num [1:43] NA NA
NA 4.28 9.51 ...
##  $ Eutrophying emissions per kilogram (gPO₄eq per kilogram)                  : num [1:43] NA NA
NA 11.2 35.1 ...
##  $ Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)  : num [1:43] NA NA
NA 8.64 49.39 ...
##  $ Freshwater withdrawals per 1000kcal (liters per 1000kcal)                 : num [1:43] NA NA
NA 184 610 ...
##  $ Freshwater withdrawals per 100g protein (liters per 100g protein)         : num [1:43] NA NA
NA 371 3167 ...
##  $ Freshwater withdrawals per kilogram (liters per kilogram)                 : num [1:43] NA NA
NA 482 2248 ...
##  $ Greenhouse gas emissions per 1000kcal (kgCO₂eq per 1000kcal)              : num [1:43] NA NA
NA 0.945 1.207 ...
##  $ Greenhouse gas emissions per 100g protein (kgCO₂eq per 100g protein)      : num [1:43] NA NA
NA 1.91 6.27 ...
##  $ Land use per 1000kcal (m² per 1000kcal)                                   : num [1:43] NA NA
NA 2.9 0.76 ...
##  $ Land use per kilogram (m² per kilogram)                                   : num [1:43] NA NA
NA 7.6 2.8 ...
##  $ Land use per 100g protein (m² per 100g protein)                           : num [1:43] NA NA
NA 5.85 3.94 ...
##  $ Scarcity-weighted water use per kilogram (liters per kilogram)            : num [1:43] NA NA
NA 18786 49576 ...
##  $ Scarcity-weighted water use per 100g protein (liters per 100g protein) : num [1:43] NA NA
NA 14451 69826 ...
##  $ Scarcity-weighted water use per 1000kcal (liters per 1000 kilocalories): num [1:43] NA NA
NA 7162 13450 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   `Food product` = col_character(),
##   ..   `Land use change` = col_double(),
##   ..   `Animal Feed` = col_double(),
```

```
##   ..   Farm = col_double(),
##   ..   Processing = col_double(),
##   ..   Transport = col_double(),
##   ..   Packging = col_double(),
##   ..   Retail = col_double(),
##   ..   Total_emissions = col_double(),
##   ..   `Eutrophying emissions per 1000kcal (gPO₄eq per 1000kcal)` = col_double(),
##   ..   `Eutrophying emissions per kilogram (gPO₄eq per kilogram)` = col_double(),
##   ..   `Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)` = col_double
(),
##   ..   `Freshwater withdrawals per 1000kcal (liters per 1000kcal)` = col_double(),
##   ..   `Freshwater withdrawals per 100g protein (liters per 100g protein)` = col_double(),
##   ..   `Freshwater withdrawals per kilogram (liters per kilogram)` = col_double(),
##   ..   `Greenhouse gas emissions per 1000kcal (kgCO₂eq per 1000kcal)` = col_double(),
##   ..   `Greenhouse gas emissions per 100g protein (kgCO₂eq per 100g protein)` = col_double(),
##   ..   `Land use per 1000kcal (m² per 1000kcal)` = col_double(),
##   ..   `Land use per kilogram (m² per kilogram)` = col_double(),
##   ..   `Land use per 100g protein (m² per 100g protein)` = col_double(),
##   ..   `Scarcity-weighted water use per kilogram (liters per kilogram)` = col_double(),
##   ..   `Scarcity-weighted water use per 100g protein (liters per 100g protein)` = col_double
(),
##   ..   `Scarcity-weighted water use per 1000kcal (liters per 1000 kilocalories)` = col_double
()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

replacing spaces in column names with '_'

```
names(data)
```

```
##  [1] "Food product"
##  [2] "Land use change"
##  [3] "Animal Feed"
##  [4] "Farm"
##  [5] "Processing"
##  [6] "Transport"
##  [7] "Packging"
##  [8] "Retail"
##  [9] "Total_emissions"
## [10] "Eutrophying emissions per 1000kcal (gPO₄eq per 1000kcal)"
## [11] "Eutrophying emissions per kilogram (gPO₄eq per kilogram)"
## [12] "Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"
## [13] "Freshwater withdrawals per 1000kcal (liters per 1000kcal)"
## [14] "Freshwater withdrawals per 100g protein (liters per 100g protein)"
## [15] "Freshwater withdrawals per kilogram (liters per kilogram)"
## [16] "Greenhouse gas emissions per 1000kcal (kgCO₂eq per 1000kcal)"
## [17] "Greenhouse gas emissions per 100g protein (kgCO₂eq per 100g protein)"
## [18] "Land use per 1000kcal (m² per 1000kcal)"
## [19] "Land use per kilogram (m² per kilogram)"
## [20] "Land use per 100g protein (m² per 100g protein)"
## [21] "Scarcity-weighted water use per kilogram (liters per kilogram)"
## [22] "Scarcity-weighted water use per 100g protein (liters per 100g protein)"
## [23] "Scarcity-weighted water use per 1000kcal (liters per 1000 kilocalories)"
```

```
names(data)<- gsub("[[:space:]+]", "_", names(data))
names(data)
```

```
##  [1] "Food_product"
##  [2] "Land_use_change"
##  [3] "Animal_Feed"
##  [4] "Farm"
##  [5] "Processing"
##  [6] "Transport"
##  [7] "Packging"
##  [8] "Retail"
##  [9] "Total_emissions"
## [10] "Eutrophying_emissions_per_1000kcal_(gPO₄eq_per_1000kcal)"
## [11] "Eutrophying_emissions_per_kilogram_(gPO₄eq_per_kilogram)"
## [12] "Eutrophying_emissions_per_100g_protein_(gPO₄eq_per_100_grams_protein)"
## [13] "Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)"
## [14] "Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)"
## [15] "Freshwater_withdrawals_per_kilogram_(liters_per_kilogram)"
## [16] "Greenhouse_gas_emissions_per_1000kcal_(kgCO₂eq_per_1000kcal)"
## [17] "Greenhouse_gas_emissions_per_100g_protein_(kgCO₂eq_per_100g_protein)"
## [18] "Land_use_per_1000kcal_(m²_per_1000kcal)"
## [19] "Land_use_per_kilogram_(m²_per_kilogram)"
## [20] "Land_use_per_100g_protein_(m²_per_100g_protein)"
## [21] "Scarcity-weighted_water_use_per_kilogram_(liters_per_kilogram)"
## [22] "Scarcity-weighted_water_use_per_100g_protein_(liters_per_100g_protein)"
## [23] "Scarcity-weighted_water_use_per_1000kcal_(liters_per_1000_kilocalories)"
```

check the data types inferred and convert them if necessary

```
data <- type_convert(data)
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   Food_product = col_character()
## )
```

```
#chocolateData$Cocoa_Percent <- sapply(chocolateData$Cocoa_Percent, function(x) gsub("%", "",
 x))
```

3. Exploring the data

```
summary(data)
```

```
##  Food_product       Land_use_change  Animal_Feed          Farm
##  Length:43          Min.   :-2.10   Min.   :0.0000   Min.   : 0.10
##  Class :character   1st Qu.: 0.00   1st Qu.:0.0000   1st Qu.: 0.35
##  Mode  :character   Median : 0.20   Median :0.0000   Median : 0.80
##                     Mean   : 1.26   Mean   :0.4535   Mean   : 3.47
##                     3rd Qu.: 0.80   3rd Qu.:0.0000   3rd Qu.: 2.20
##                     Max.   :16.30   Max.   :2.9000   Max.   :39.40
##
##    Processing       Transport         Packging          Retail
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.1000   1st Qu.:0.1000   1st Qu.:0.00000
##  Median :0.1000   Median :0.1000   Median :0.1000   Median :0.00000
##  Mean   :0.2535   Mean   :0.1953   Mean   :0.2698   Mean   :0.06977
##  3rd Qu.:0.3000   3rd Qu.:0.2000   3rd Qu.:0.3000   3rd Qu.:0.15000
##  Max.   :1.3000   Max.   :0.8000   Max.   :1.6000   Max.   :0.30000
##
##  Total_emissions  Eutrophying_emissions_per_1000kcal_(gPO₄eq_per_1000kcal)
##  Min.   : 0.200   Min.   :  0.7084
##  1st Qu.: 0.850   1st Qu.:  4.2149
##  Median : 1.600   Median :  7.0000
##  Mean   : 5.972   Mean   : 27.1816
##  3rd Qu.: 6.000   3rd Qu.: 26.3243
##  Max.   :59.600   Max.   :197.3571
##                   NA's   :10
##  Eutrophying_emissions_per_kilogram_(gPO₄eq_per_kilogram)
##  Min.   :  0.690
##  1st Qu.:  3.752
##  Median : 11.460
##  Mean   : 46.141
##  3rd Qu.: 45.840
##  Max.   :365.290
##  NA's   :5
##  Eutrophying_emissions_per_100g_protein_(gPO₄eq_per_100_grams_protein)
##  Min.   :  3.384
##  1st Qu.: 17.855
##  Median : 37.333
##  Mean   : 52.772
##  3rd Qu.: 55.297
##  Max.   :185.051
##  NA's   :16
##  Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)
##  Min.   :   0.724
##  1st Qu.: 106.928
##  Median : 338.059
##  Mean   : 504.189
##  3rd Qu.: 694.805
##  Max.   :2062.179
##  NA's   :13
##  Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)
##  Min.   :  32.38
##  1st Qu.: 373.57
##  Median :1083.33
```

```
##  Mean   :1437.97
##  3rd Qu.:1832.39
##  Max.   :6003.33
##  NA's   :17
##  Freshwater_withdrawals_per_kilogram_(liters_per_kilogram)
##  Min.   :   0.0
##  1st Qu.: 105.5
##  Median : 417.1
##  Mean   : 932.6
##  3rd Qu.:1340.4
##  Max.   :5605.2
##  NA's   :5
##  Greenhouse_gas_emissions_per_1000kcal_(kgCO₂eq_per_1000kcal)
##  Min.   : 0.06992
##  1st Qu.: 0.62842
##  Median : 1.35135
##  Mean   : 5.63394
##  3rd Qu.: 5.33514
##  Max.   :50.94643
##  NA's   :10
##  Greenhouse_gas_emissions_per_100g_protein_(kgCO₂eq_per_100g_protein)
##  Min.   : 0.2633
##  1st Qu.: 4.0274
##  Median : 6.5000
##  Mean   :13.5249
##  3rd Qu.:14.9833
##  Max.   :93.3000
##  NA's   :16
##  Land_use_per_1000kcal_(m²_per_1000kcal)
##  Min.   :  0.2738
##  1st Qu.:  1.3125
##  Median :  2.9762
##  Mean   : 12.4232
##  3rd Qu.:  6.6054
##  Max.   :119.4908
##  NA's   :10
##  Land_use_per_kilogram_(m²_per_kilogram)
##  Min.   :  0.330
##  1st Qu.:  1.113
##  Median :  6.865
##  Mean   : 29.265
##  3rd Qu.: 14.918
##  Max.   :369.810
##  NA's   :5
##  Land_use_per_100g_protein_(m²_per_100g_protein)
##  Min.   :  3.000
##  1st Qu.:  5.088
##  Median :  7.936
##  Mean   : 29.105
##  3rd Qu.: 23.002
##  Max.   :184.813
##  NA's   :16
```

```
##  Scarcity-weighted_water_use_per_kilogram_(liters_per_kilogram)
##  Min.   :     0
##  1st Qu.:  3325
##  Median : 14533
##  Mean   : 36607
##  3rd Qu.: 35960
##  Max.   :229890
##  NA's   :5
##  Scarcity-weighted_water_use_per_100g_protein_(liters_per_100g_protein)
##  Min.   :   421.2
##  1st Qu.: 11018.4
##  Median : 20917.2
##  Mean   : 59196.4
##  3rd Qu.: 70651.7
##  Max.   :431620.0
##  NA's   :17
##  Scarcity-weighted_water_use_per_1000kcal_(liters_per_1000_kilocalories)
##  Min.   :    4.1
##  1st Qu.: 2969.1
##  Median :12605.3
##  Mean   :17380.6
##  3rd Qu.:28056.5
##  Max.   :49735.9
##  NA's   :13
```

```
summarise_all(data, funs(mean))
```

```
## Warning: `funs()` was deprecated in dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
## Warning in mean.default(Food_product): argument is not numeric or logical:
## returning NA
```

```
## # A tibble: 1 × 23
##   Food_product Land_use_change Animal_Feed  Farm Processing Transport Packging
##          <dbl>           <dbl>       <dbl> <dbl>      <dbl>     <dbl>    <dbl>
## 1           NA            1.26       0.453  3.47      0.253     0.195    0.270
## # … with 16 more variables: Retail <dbl>, Total_emissions <dbl>,
## #   `Eutrophying_emissions_per_1000kcal_(gPO₄eq_per_1000kcal)` <dbl>,
## #   `Eutrophying_emissions_per_kilogram_(gPO₄eq_per_kilogram)` <dbl>,
## #   `Eutrophying_emissions_per_100g_protein_(gPO₄eq_per_100_grams_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)` <dbl>,
## #   `Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_kilogram_(liters_per_kilogram)` <dbl>, …
```

```
# filtering and exploring some columns and rows in the data

#data %>%
    #group_by() %>%
#    summarise(averageFeed = mean(Land_use_change),
#             FeedSD = sd(Land_use_change))
#ncol(data)
#data%>%select(Animal_Feed) %>% summary()
#data[,c("Animal_Feed","Farm")] %>%summary()
#data.filter(Product="Rice")

#data%>%filter(data['Food product']=='Rice')
#head(data['Processing']) # or data[5]
#data[1,1] #first row first column
#data[1,] #first row
#data[-1,] #everything except first row
```

4. Answering questions

Which types of food have more negative impact on the environment?

```
data %>% arrange(desc(Total_emissions)) %>% head()
```

```
## # A tibble: 6 × 23
##   Food_product   Land_use_change Animal_Feed  Farm Processing Transport Packging
##   <chr>                    <dbl>       <dbl> <dbl>      <dbl>     <dbl>    <dbl>
## 1 Beef (beef he…            16.3         1.9  39.4        1.3       0.3      0.2
## 2 Lamb & Mutton              0.5         2.4  19.5        1.1       0.5      0.3
## 3 Cheese                     4.5         2.3  13.1        0.7       0.1      0.2
## 4 Beef (dairy h…             0.9         2.5  15.7        1.1       0.4      0.3
## 5 Dark Chocolate            14.3         0     3.7        0.2       0.1      0.4
## 6 Coffee                     3.7         0    10.4        0.6       0.1      1.6
## # … with 16 more variables: Retail <dbl>, Total_emissions <dbl>,
## #   `Eutrophying_emissions_per_1000kcal_(gPO₄eq_per_1000kcal)` <dbl>,
## #   `Eutrophying_emissions_per_kilogram_(gPO₄eq_per_kilogram)` <dbl>,
## #   `Eutrophying_emissions_per_100g_protein_(gPO₄eq_per_100_grams_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)` <dbl>,
## #   `Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_kilogram_(liters_per_kilogram)` <dbl>, …
```

coffee and chocolate are in the top ten products that produce the highest amount of emission , however, they do not provide any calories but their production use so much resources, so measurements such as `Scarcity-weighted_water_use_per_1000kcal_(liters_per_1000_kilocalories)` is very high for such products. Thus, the following question comes up:

What types of food production should be encouraged to consume nutritious diet in a sustainable way?

```
data %>% arrange(Total_emissions) %>% head(10)
```

```
## # A tibble: 10 × 23
##    Food_product  Land_use_change Animal_Feed  Farm Processing Transport Packging
##    <chr>                   <dbl>       <dbl> <dbl>      <dbl>     <dbl>    <dbl>
##  1 Nuts                     -2.1           0   2.1        0         0.1      0.1
##  2 Citrus Fruit             -0.1           0   0.3        0         0.1      0
##  3 Potatoes                  0             0   0.2        0         0.1      0
##  4 Onions & Lee…             0             0   0.2        0         0.1      0
##  5 Root Vegetab…             0             0   0.2        0         0.1      0
##  6 Apples                    0             0   0.2        0         0.1      0
##  7 Brassicas                 0             0   0.3        0         0.1      0
##  8 Other Vegeta…             0             0   0.2        0.1       0.2      0
##  9 Other Fruit               0.1           0   0.4        0         0.2      0
## 10 Peas                      0             0   0.7        0         0.1      0
## # … with 16 more variables: Retail <dbl>, Total_emissions <dbl>,
## #   `Eutrophying_emissions_per_1000kcal_(gPO₄eq_per_1000kcal)` <dbl>,
## #   `Eutrophying_emissions_per_kilogram_(gPO₄eq_per_kilogram)` <dbl>,
## #   `Eutrophying_emissions_per_100g_protein_(gPO₄eq_per_100_grams_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)` <dbl>,
## #   `Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_kilogram_(liters_per_kilogram)` <dbl>, …
```

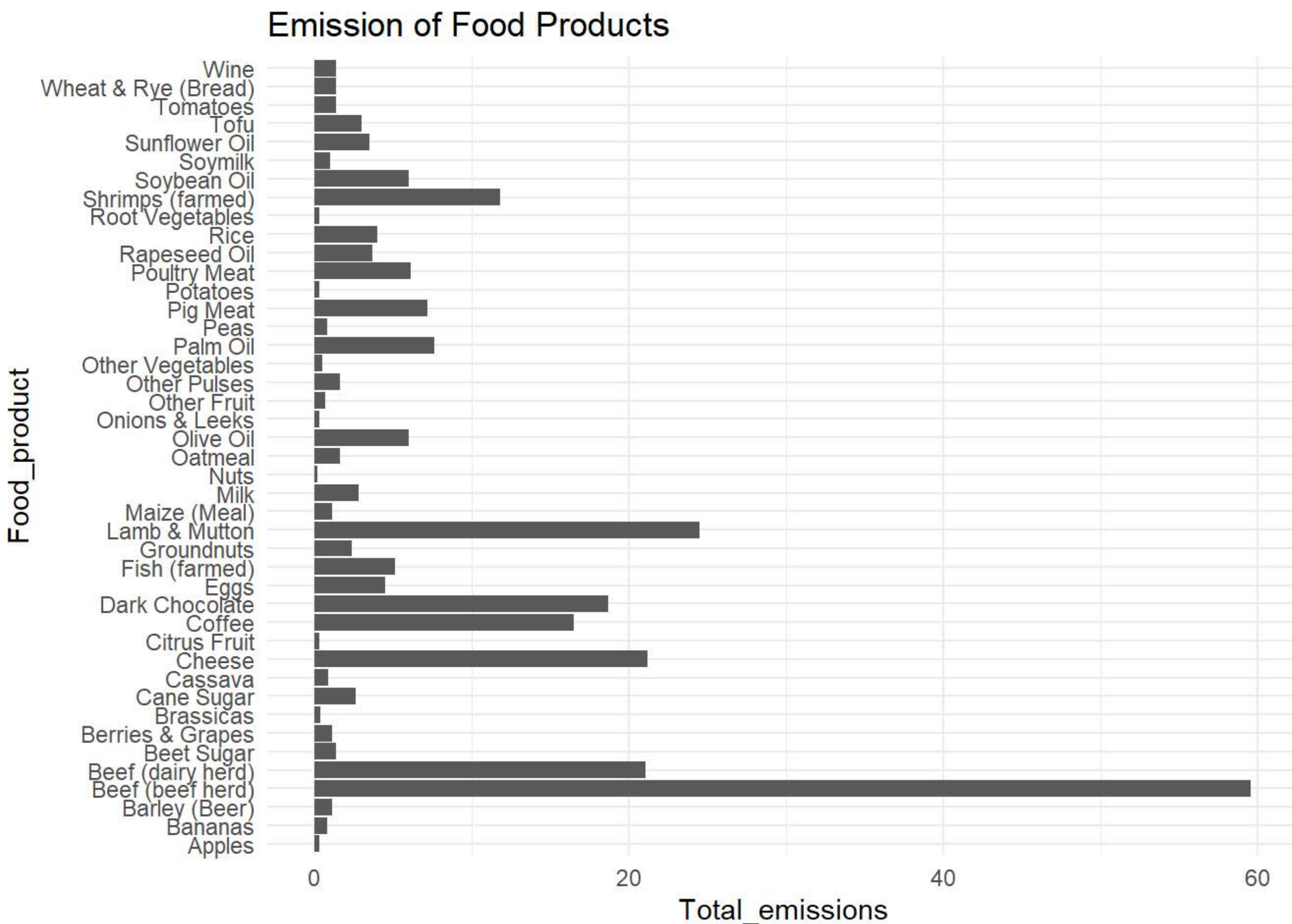
```
water<- "Scarcity-weighted_water_use_per_1000kcal_(liters_per_1000_kilocalories)"
co2<-"Greenhouse_gas_emissions_per_1000kcal_(kgCO₂eq_per_1000kcal)"
land<-"Land_use_per_1000kcal_(m²_per_1000kcal)"
waste<-"Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)"
data %>% arrange(desc(water), desc(co2),desc(land), desc(waste)) %>% head(10)
```

```
## # A tibble: 10 × 23
##    Food_product  Land_use_change Animal_Feed  Farm Processing Transport Packging
##    <chr>                   <dbl>       <dbl> <dbl>      <dbl>     <dbl>    <dbl>
##  1 Wheat & Rye …             0.1           0   0.8        0.2       0.1      0.1
##  2 Maize (Meal)              0.3           0   0.5        0.1       0.1      0.1
##  3 Barley (Beer)             0             0   0.2        0.1       0        0.5
##  4 Oatmeal                   0             0   1.4        0         0.1      0.1
##  5 Rice                      0             0   3.6        0.1       0.1      0.1
##  6 Potatoes                  0             0   0.2        0         0.1      0
##  7 Cassava                   0.6           0   0.2        0         0.1      0
##  8 Cane Sugar                1.2           0   0.5        0         0.8      0.1
##  9 Beet Sugar                0             0   0.5        0.2       0.6      0.1
## 10 Other Pulses              0             0   1.1        0         0.1      0.4
## # … with 16 more variables: Retail <dbl>, Total_emissions <dbl>,
## #   `Eutrophying_emissions_per_1000kcal_(gPO₄eq_per_1000kcal)` <dbl>,
## #   `Eutrophying_emissions_per_kilogram_(gPO₄eq_per_kilogram)` <dbl>,
## #   `Eutrophying_emissions_per_100g_protein_(gPO₄eq_per_100_grams_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)` <dbl>,
## #   `Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_kilogram_(liters_per_kilogram)` <dbl>, …
```

```
water<- "Scarcity-weighted_water_use_per_100g_protein_(liters_per_100g_protein)"
co2<-"Greenhouse_gas_emissions_per_100g_protein_(kgCO₂eq_per_100g_protein)"
land<-"Land_use_per_100g_protein_(m²_per_100g_protein)"
waste<-"Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)"
data %>% arrange(desc(water), desc(co2), desc(land), desc(waste)) %>% head(10)
```

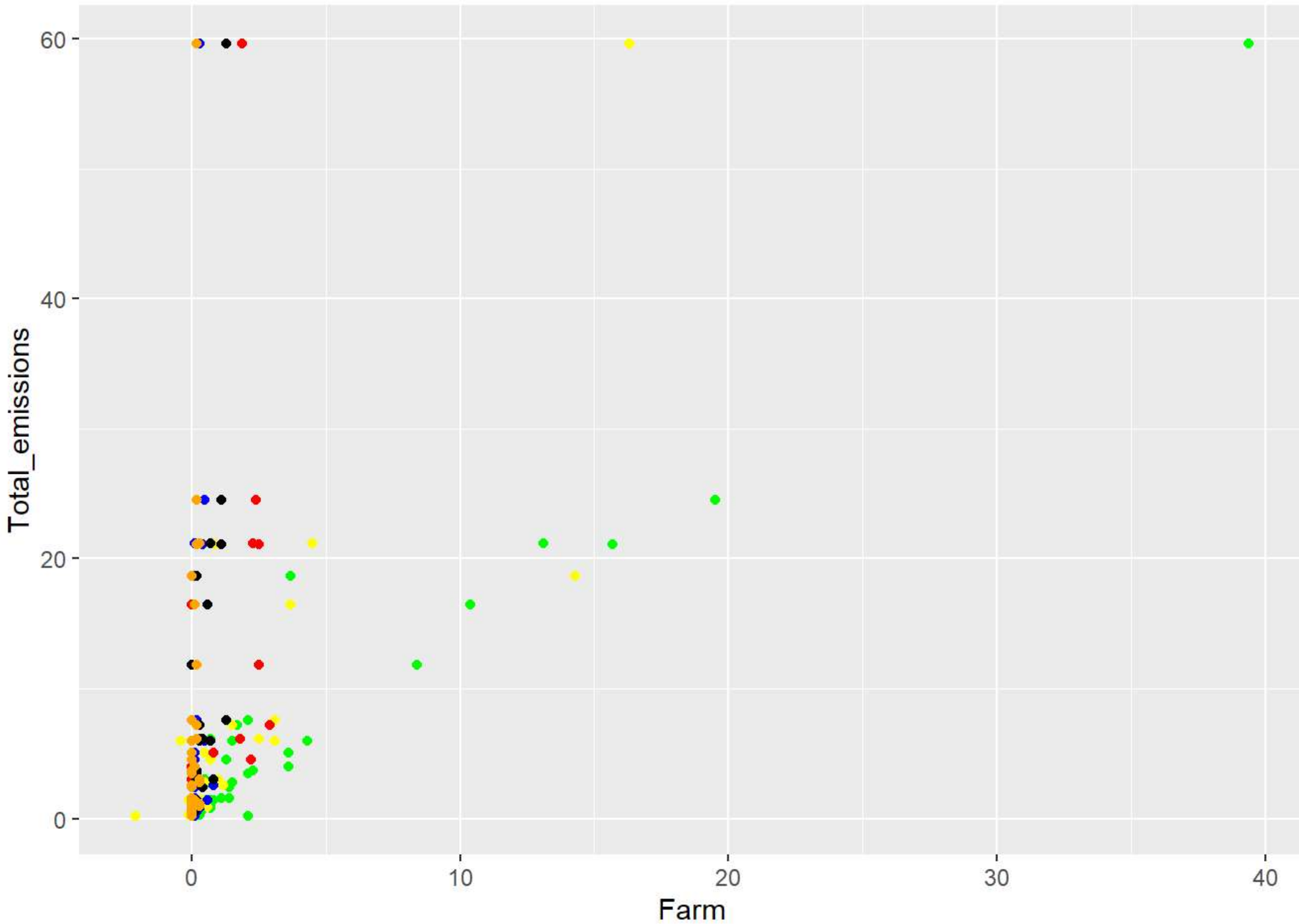
```
## # A tibble: 10 × 23
##    Food_product  Land_use_change Animal_Feed  Farm Processing Transport Packging
##    <chr>                   <dbl>       <dbl> <dbl>      <dbl>     <dbl>    <dbl>
##  1 Wheat & Rye …             0.1           0   0.8        0.2       0.1      0.1
##  2 Maize (Meal)              0.3           0   0.5        0.1       0.1      0.1
##  3 Barley (Beer)             0             0   0.2        0.1       0        0.5
##  4 Oatmeal                   0             0   1.4        0         0.1      0.1
##  5 Rice                      0             0   3.6        0.1       0.1      0.1
##  6 Potatoes                  0             0   0.2        0         0.1      0
##  7 Cassava                   0.6           0   0.2        0         0.1      0
##  8 Cane Sugar                1.2           0   0.5        0         0.8      0.1
##  9 Beet Sugar                0             0   0.5        0.2       0.6      0.1
## 10 Other Pulses              0             0   1.1        0         0.1      0.4
## # … with 16 more variables: Retail <dbl>, Total_emissions <dbl>,
## #   `Eutrophying_emissions_per_1000kcal_(gPO₄eq_per_1000kcal)` <dbl>,
## #   `Eutrophying_emissions_per_kilogram_(gPO₄eq_per_kilogram)` <dbl>,
## #   `Eutrophying_emissions_per_100g_protein_(gPO₄eq_per_100_grams_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)` <dbl>,
## #   `Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_kilogram_(liters_per_kilogram)` <dbl>, …
```

```
plt<-ggplot(data=data, mapping=aes(y=Food_product, x=Total_emissions))+
geom_col()+
theme_minimal()+
ggtitle("Emission of Food Products")

# save our plot
ggsave("product_emmisions.png", # the name of the file where it will be save
       plot = plt, # what plot to save
       height=6, width=10, units="in")
plt
```



Which stage of food production contributes more to the greenhouse gas emission?

```
ggplot(data=data)+
geom_point(mapping=aes(x=Farm, y=Total_emissions), color="green") +
geom_point(mapping=aes(x=Land_use_change, y=Total_emissions), color="yellow") +
geom_point(mapping=aes(x=Animal_Feed, y=Total_emissions), color="red") +
geom_point(mapping=aes(x=Transport, y=Total_emissions), color="blue") +
geom_point(mapping=aes(x=Processing, y=Total_emissions), color="black") +
geom_point(mapping=aes(x=Retail, y=Total_emissions), color="orange")
```




```
#install.packages("reshape2")
library(reshape2)
```

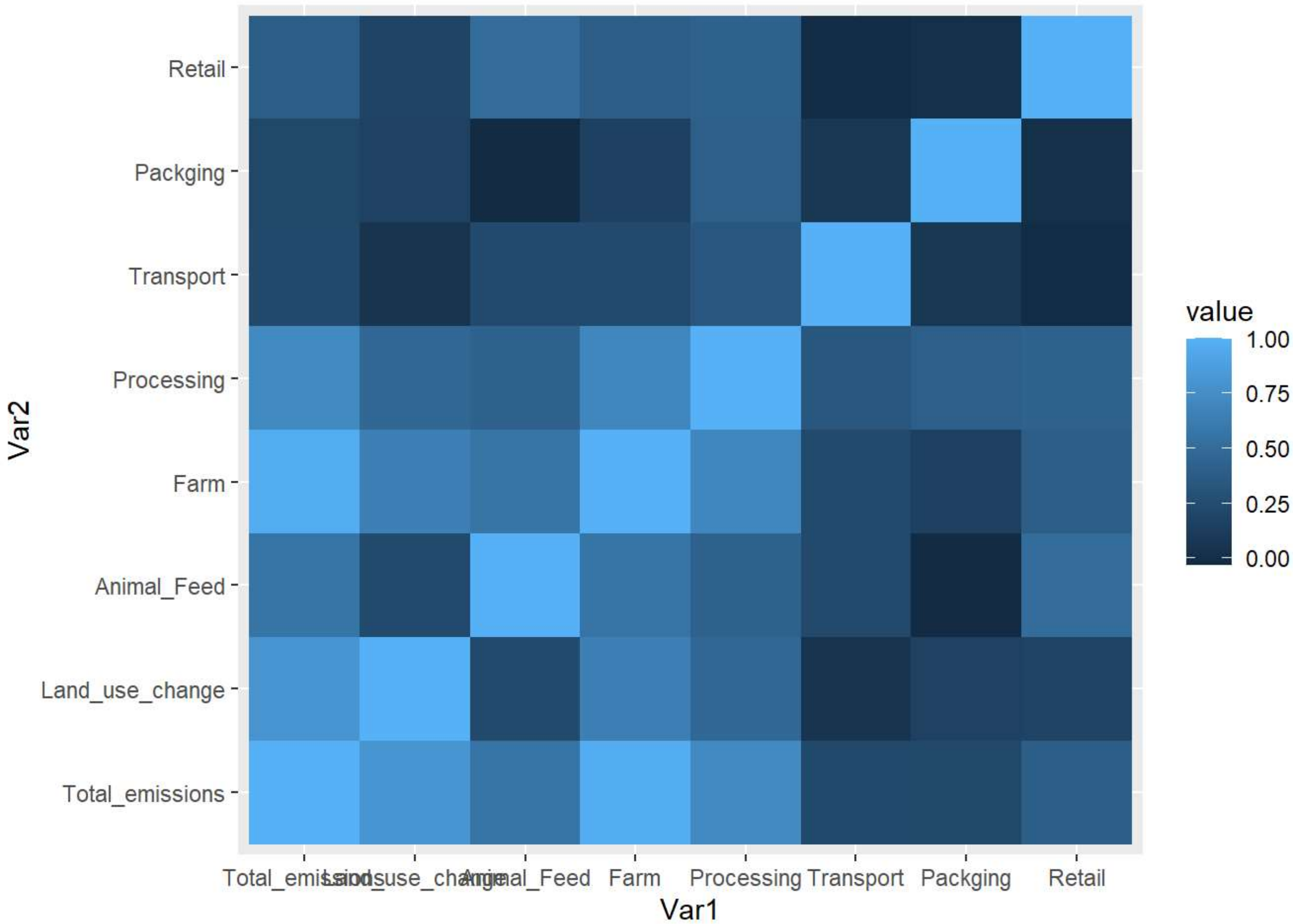
```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
cols<-c("Total_emissions", "Land_use_change","Animal_Feed","Farm","Processing","Transport","Packaging","Retail")
corr_mat <- round(cor(data[cols]),2)
corr_mat
```

```
##                 Total_emissions Land_use_change Animal_Feed Farm Processing
## Total_emissions            1.00            0.80        0.57 0.97       0.72
## Land_use_change            0.80            1.00        0.24 0.65       0.47
## Animal_Feed                0.57            0.24        1.00 0.57       0.43
## Farm                       0.97            0.65        0.57 1.00       0.70
## Processing                 0.72            0.47        0.43 0.70       1.00
## Transport                  0.22            0.05        0.23 0.23       0.34
## Packging                   0.21            0.17       -0.03 0.16       0.40
## Retail                     0.39            0.18        0.51 0.39       0.43
##                 Transport Packging Retail
## Total_emissions      0.22     0.21   0.39
## Land_use_change      0.05     0.17   0.18
## Animal_Feed          0.23    -0.03   0.51
## Farm                 0.23     0.16   0.39
## Processing           0.34     0.40   0.43
## Transport            1.00     0.08  -0.01
## Packging             0.08     1.00   0.01
## Retail              -0.01     0.01   1.00
```

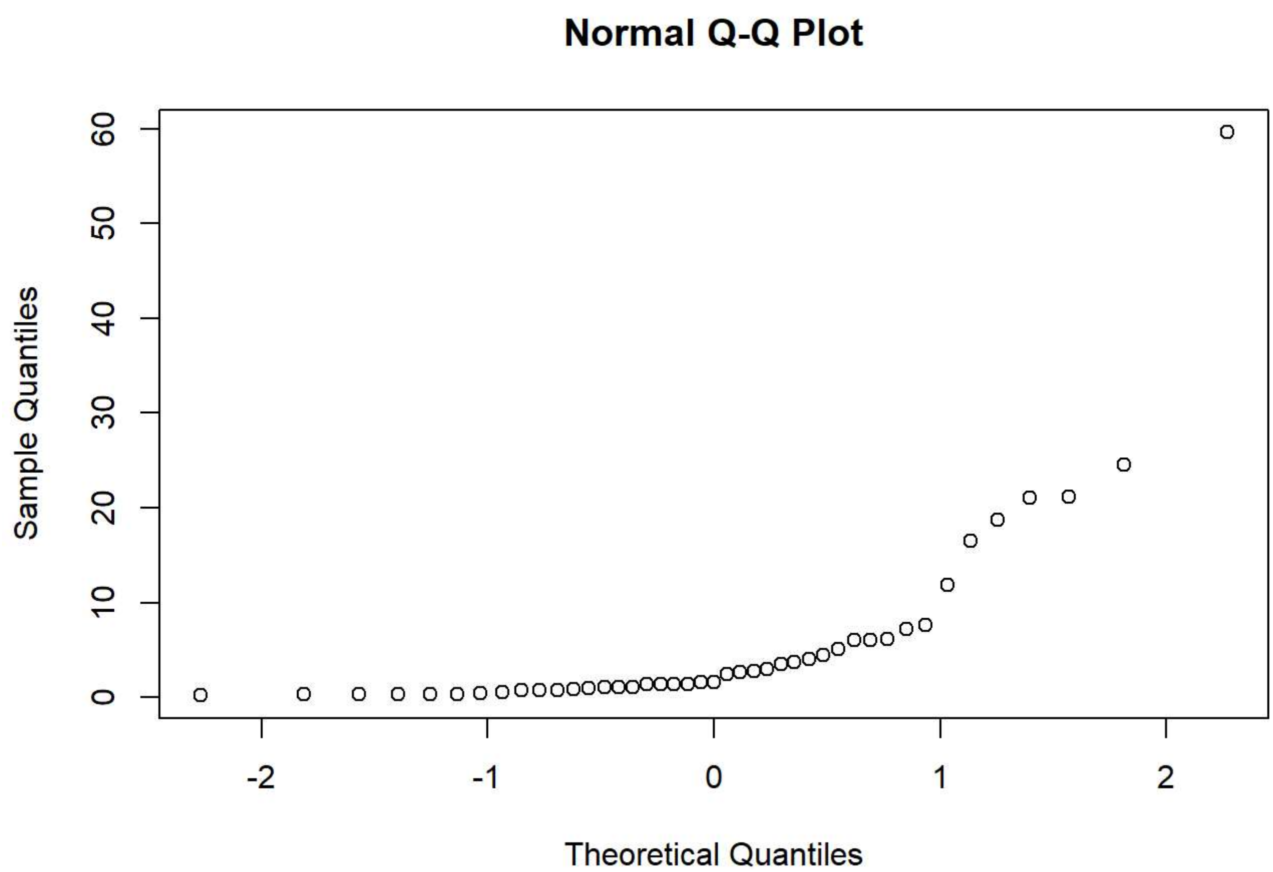
```
# reduce the corr matrix
melted_corr_mat <- melt(corr_mat)

# plotting the corr heatmap
ggplot(data = melted_corr_mat, aes(x=Var1,
                                   y=Var2,
                                   fill=value)) +
  geom_tile()
```



Farming, Processing and Land use change has the highest emission contribution

Hypothesis : the plant based food produce less emission

```
qqnorm(data$Total_emissions)
```




```
new_data<-data%>% mutate(Plant_based=Animal_Feed<=0)
new_data
```

```
## # A tibble: 43 × 24
##    Food_product  Land_use_change Animal_Feed  Farm Processing Transport Packging
##    <chr>                   <dbl>       <dbl> <dbl>      <dbl>     <dbl>    <dbl>
##  1 Wheat & Rye …             0.1           0   0.8        0.2       0.1      0.1
##  2 Maize (Meal)              0.3           0   0.5        0.1       0.1      0.1
##  3 Barley (Beer)             0             0   0.2        0.1       0        0.5
##  4 Oatmeal                   0             0   1.4        0         0.1      0.1
##  5 Rice                      0             0   3.6        0.1       0.1      0.1
##  6 Potatoes                  0             0   0.2        0         0.1      0
##  7 Cassava                   0.6           0   0.2        0         0.1      0
##  8 Cane Sugar                1.2           0   0.5        0         0.8      0.1
##  9 Beet Sugar                0             0   0.5        0.2       0.6      0.1
## 10 Other Pulses              0             0   1.1        0         0.1      0.4
## # … with 33 more rows, and 17 more variables: Retail <dbl>,
## #   Total_emissions <dbl>,
## #   `Eutrophying_emissions_per_1000kcal_(gPO₄eq_per_1000kcal)` <dbl>,
## #   `Eutrophying_emissions_per_kilogram_(gPO₄eq_per_kilogram)` <dbl>,
## #   `Eutrophying_emissions_per_100g_protein_(gPO₄eq_per_100_grams_protein)` <dbl>,
## #   `Freshwater_withdrawals_per_1000kcal_(liters_per_1000kcal)` <dbl>,
## #   `Freshwater_withdrawals_per_100g_protein_(liters_per_100g_protein)` <dbl>, …
```

```
t.test(new_data$Total_emissions ~ new_data$Plant_based)
```

```
##
##  Welch Two Sample t-test
##
## data:  new_data$Total_emissions by new_data$Plant_based
## t = 2.482, df = 9.3358, p-value = 0.03399
## alternative hypothesis: true difference in means between group FALSE and group TRUE is not eq
ual to 0
## 95 percent confidence interval:
##   1.270118 25.879579
## sample estimates:
## mean in group FALSE  mean in group TRUE
##           16.390000            2.815152
```

```
ggplot(data = new_data, aes(x = Total_emissions, fill = Plant_based)) +
geom_histogram(binwidth=10)
```