Sustainability in Fashion

Reese Putnam

2025-01-24

In this project we will find the countrys with the best sustanbile fashion brands and if they have recycling prgram efforts to sustain their production.

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
install.packages("tidyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4      ✓ readr     2.1.5
## ✓ forcats   1.0.0      ✓ stringr   1.5.1
## ✓ ggplot2   3.5.1      ✓ tibble    3.2.1
## ✓ lubridate 1.9.4      ✓ tidyr     1.3.1
## ✓ purrr     1.0.2
## ── Conflicts ───────────────────────────────── tidyverse_conflicts() ──
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
library(ggplot2)
library(dplyr)
library(tidyr)
```

```
data <- read_csv("sustainable_fashion_trends.csv")
```

```
## Rows: 5000 Columns: 15
## — Column specification ————————————————————————————————————
## Delimiter: ","
## chr (9): Brand_ID, Brand_Name, Country, Sustainability_Rating, Material_Type...
## dbl (6): Year, Carbon_Footprint_MT, Water_Usage_Liters, Waste_Production_KG,...
##
## 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.
```

The data we are using is the sustainable fashion trends data set found for free on Kaggle. This data set has 5000 rows and 15 columns showing various insights on brands and their usuage of resources to create their clothing. We will use this data set to see which cpuntry's have the most sustainable brands and what resources they use to creat their clothing that contributes to their sustainability.

```
str(data)
```

```
## spc_tbl_ [5,000 × 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Brand_ID               : chr [1:5000] "BRAND-0001" "BRAND-0002" "BRAND-0003" "BRAND-000
4" ...
## $ Brand_Name             : chr [1:5000] "Brand_1" "Brand_2" "Brand_3" "Brand_4" ...
## $ Country                : chr [1:5000] "Australia" "Japan" "USA" "Italy" ...
## $ Year                   : num [1:5000] 2018 2015 2024 2023 2016 ...
## $ Sustainability_Rating  : chr [1:5000] "D" "D" "A" "D" ...
## $ Material_Type          : chr [1:5000] "Tencel" "Vegan Leather" "Vegan Leather" "Bamboo
Fabric" ...
## $ Eco_Friendly_Manufacturing: chr [1:5000] "No" "Yes" "No" "No" ...
## $ Carbon_Footprint_MT    : num [1:5000] 1.75 124.39 336.66 152.04 415.63 ...
## $ Water_Usage_Liters     : num [1:5000] 4511153 1951566 467455 899577 1809220 ...
## $ Waste_Production_KG    : num [1:5000] 97844 37268 38386 32665 37295 ...
## $ Recycling_Programs     : chr [1:5000] "No" "No" "No" "No" ...
## $ Product_Lines          : num [1:5000] 2 15 2 13 19 10 17 18 11 9 ...
## $ Average_Price_USD      : num [1:5000] 38.3 250.1 146.2 165.5 211.6 ...
## $ Market_Trend           : chr [1:5000] "Growing" "Growing" "Growing" "Stable" ...
## $ Certifications         : chr [1:5000] "GOTS" "GOTS" "B Corp" "OEKO-TEX" ...
## - attr(*, "spec")=
##   .. cols(
##   ..   Brand_ID = col_character(),
##   ..   Brand_Name = col_character(),
##   ..   Country = col_character(),
##   ..   Year = col_double(),
##   ..   Sustainability_Rating = col_character(),
##   ..   Material_Type = col_character(),
##   ..   Eco_Friendly_Manufacturing = col_character(),
##   ..   Carbon_Footprint_MT = col_double(),
##   ..   Water_Usage_Liters = col_double(),
##   ..   Waste_Production_KG = col_double(),
##   ..   Recycling_Programs = col_character(),
##   ..   Product_Lines = col_double(),
##   ..   Average_Price_USD = col_double(),
##   ..   Market_Trend = col_character(),
##   ..   Certifications = col_character()
##   .. )
## - attr(*, "problems")=<externalptr>
```

```
summary(data)
```

```
##    Brand_ID           Brand_Name          Country              Year
## Length:5000       Length:5000        Length:5000       Min.    :2010
## Class :character   Class :character   Class :character   1st Qu.:2013
## Mode  :character   Mode  :character   Mode  :character   Median :2017
##                                                          Mean    :2017
##                                                          3rd Qu.:2021
##                                                          Max.    :2024
##  Sustainability_Rating Material_Type       Eco_Friendly_Manufacturing
## Length:5000            Length:5000        Length:5000
## Class :character       Class :character   Class :character
## Mode  :character       Mode  :character   Mode  :character
##
##
##
##  Carbon_Footprint_MT Water_Usage_Liters Waste_Production_KG Recycling_Programs
## Min.    :  1.04       Min.    :  50106   Min.    : 1026      Length:5000
## 1st Qu.:126.61        1st Qu.:1293087    1st Qu.:25341       Class :character
## Median :250.65        Median :2499096    Median :50466       Mode  :character
## Mean    :250.32       Mean    :2517862   Mean    :50107
## 3rd Qu.:372.25        3rd Qu.:3763860    3rd Qu.:74985
## Max.    :499.93       Max.    :4999597   Max.    :99948
##  Product_Lines    Average_Price_USD Market_Trend       Certifications
## Min.    : 1.00    Min.    : 20.02   Length:5000        Length:5000
## 1st Qu.: 5.00    1st Qu.:142.87     Class :character   Class :character
## Median :10.00    Median :258.62     Mode  :character   Mode  :character
## Mean    :10.43    Mean    :259.35
## 3rd Qu.:15.00    3rd Qu.:378.60
## Max.    :20.00    Max.    :499.94
```

Now that we have seen the structre of our data set we first need to check if there is any missing values in the data set and remove if needed.

```
colSums(is.na(data))
```

```
##                   Brand_ID                Brand_Name
##                          0                         0
##                    Country                      Year
##                          0                         0
##      Sustainability_Rating             Material_Type
##                          0                         0
## Eco_Friendly_Manufacturing       Carbon_Footprint_MT
##                          0                         0
##         Water_Usage_Liters       Waste_Production_KG
##                          0                         0
##         Recycling_Programs             Product_Lines
##                          0                         0
##          Average_Price_USD              Market_Trend
##                          0                         0
##             Certifications
##                          0
```
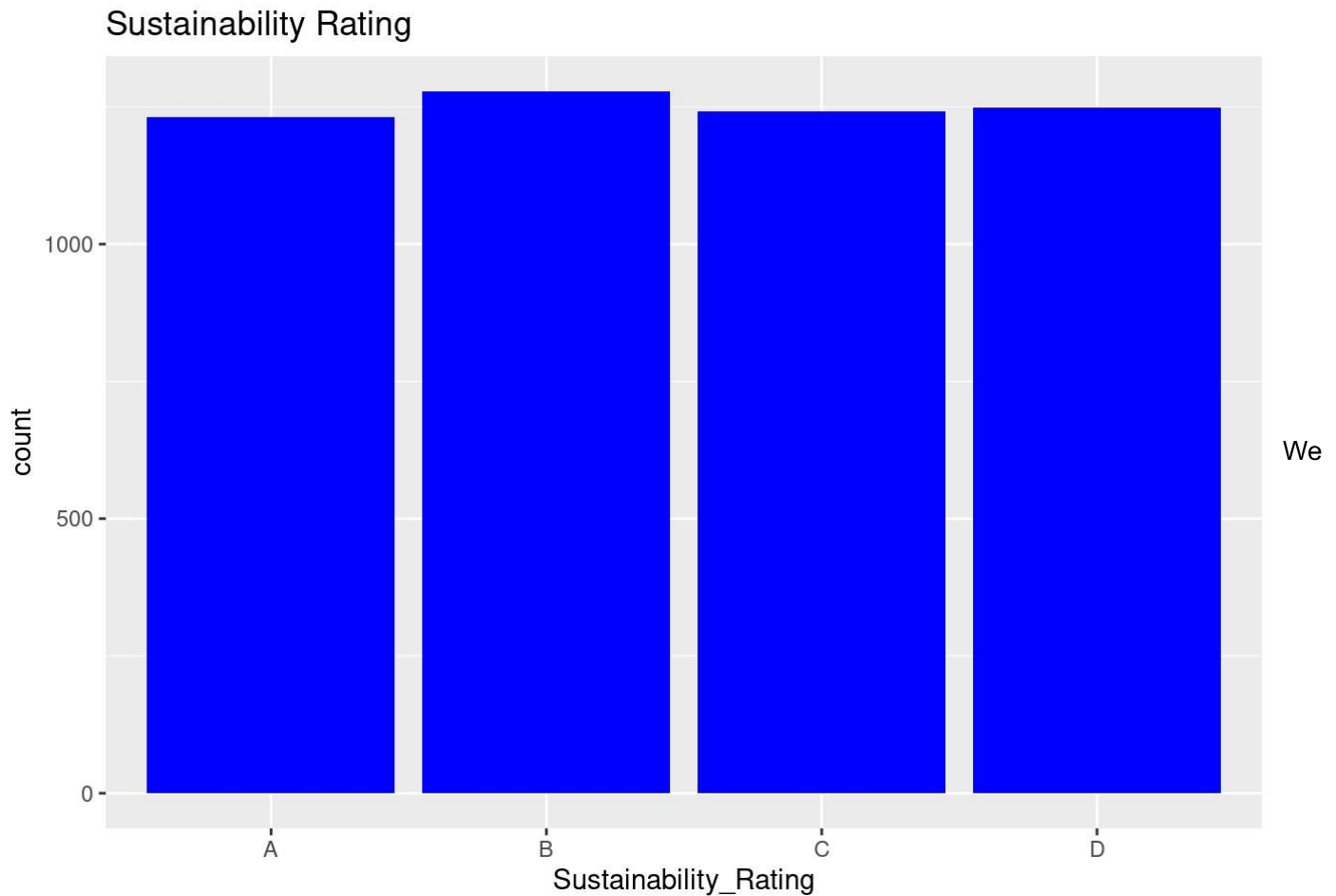
As shown above we see there is no data missing in each column so we can begin analzing and visualizing the data.

```
ggplot(data, aes(x= Sustainability_Rating)) +
  geom_bar(fill= "blue") +
  ggtitle( "Sustainability Rating")
```


We

can see a pretty even rating excpet for a majority being rated B in sustainability

```
country_ratings <- data %>%
  group_by( Country, Sustainability_Rating) %>%
  summarise(Count = n())
```

```
## `summarise()` has grouped output by 'Country'. You can override using the
## `.groups` argument.
```

```
ggplot(country_ratings, aes(x= Country,y= Count, fill= Sustainability_Rating))+
  geom_bar(stat="identity", position = "dodge")+
  ggtitle("Sustainability Rating by Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Sustainability Rating by Country
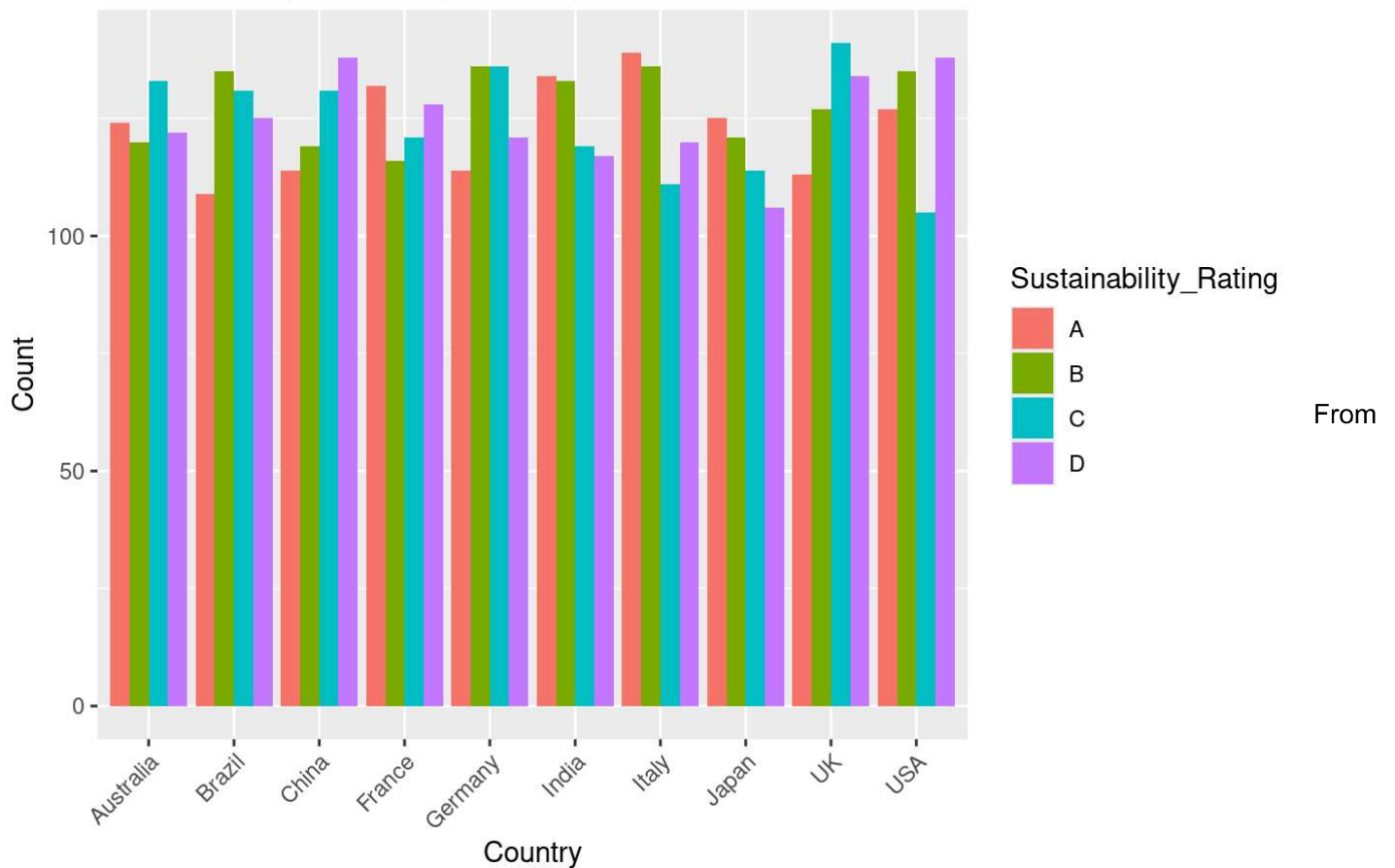
From the graph above we can observe that France, India, and Italy have the highest A sustainability rating compared to the other countrys.

```
ggplot(data, aes(x= Country, fill= Material_Type))+
  geom_bar(stat="Count", position= "dodge") +
  ggtitle("Country's Most Popular Material") +
  xlab("Country")+
  ylab("Count") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Country's Most Popular Material
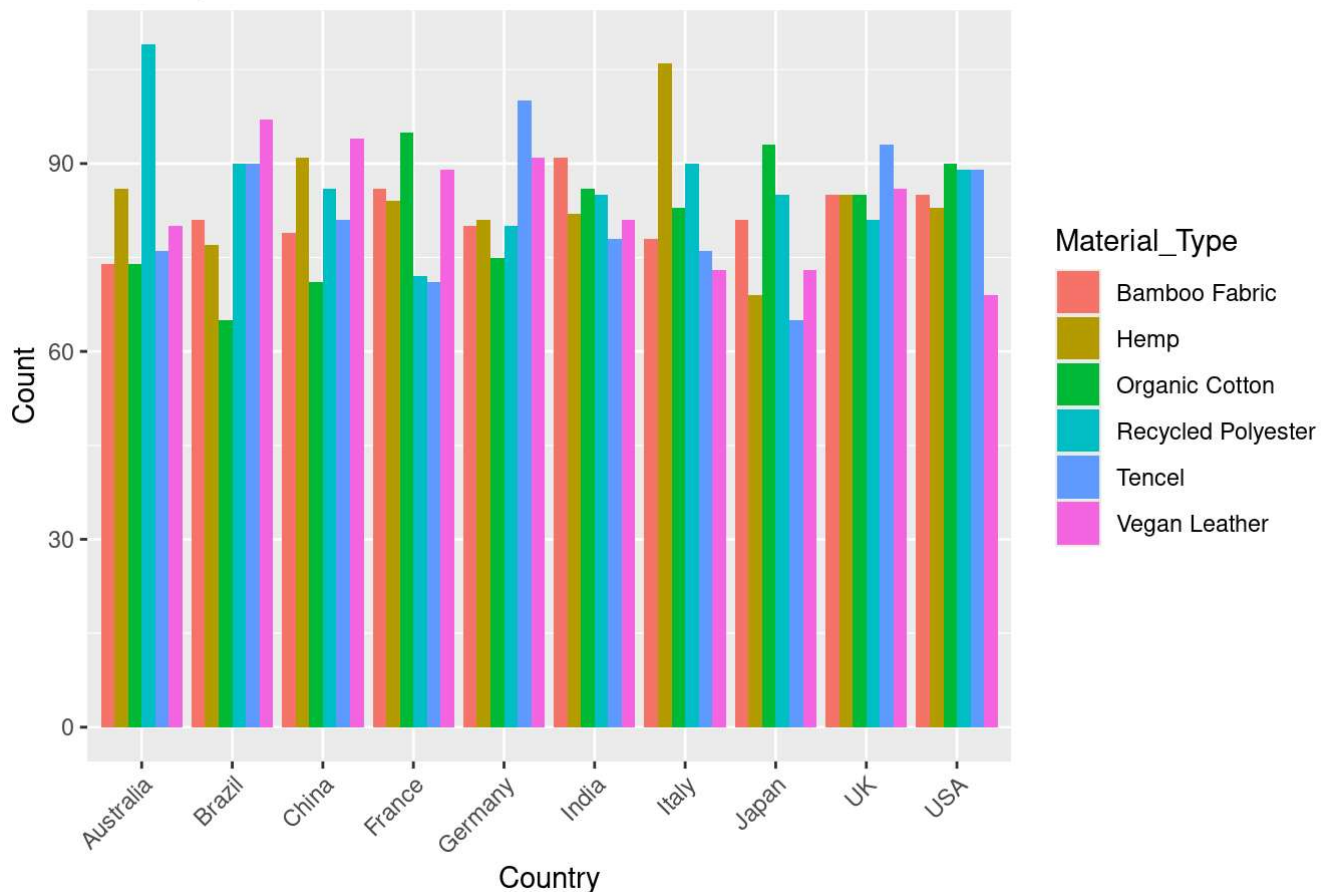
There shows to be a consistent use of all material expect for Australia having the most use of recycled polyester and Italy using Hemp.

```
carbon_footprint <- data %>%
  group_by( Country, Carbon_Footprint_MT) %>%
  summarise(count= n())
```

```
## `summarise()` has grouped output by 'Country'. You can override using the
## `.groups` argument.
```

```
ggplot(carbon_footprint, aes(x= Country, y= count, fill= Carbon_Footprint_MT))+
  geom_bar(stat="identity")+
  ggtitle("Carbon Footprint by Country")+
  xlab("Country")+
  ylab("Count")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
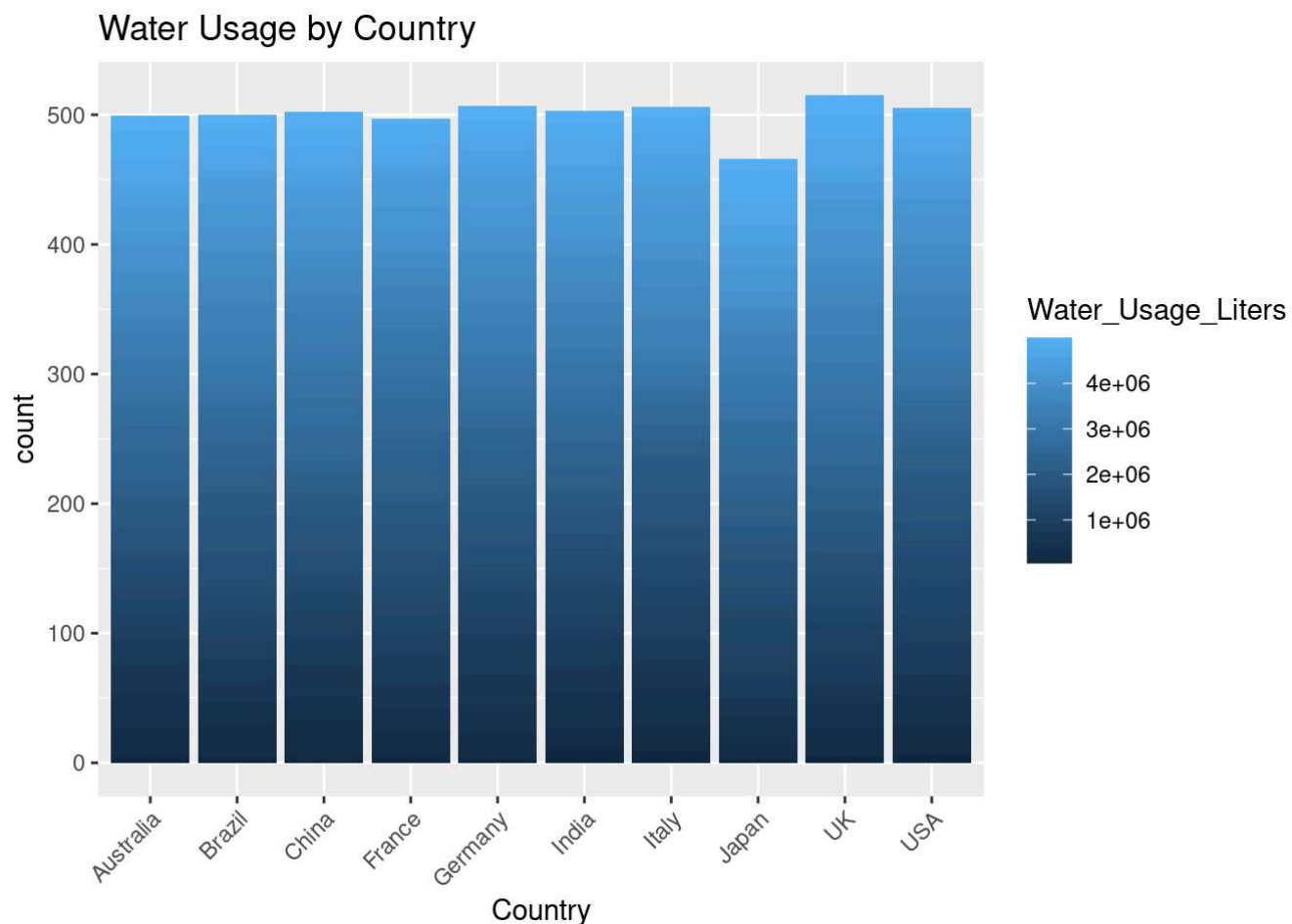
Carbon Footprint by Country
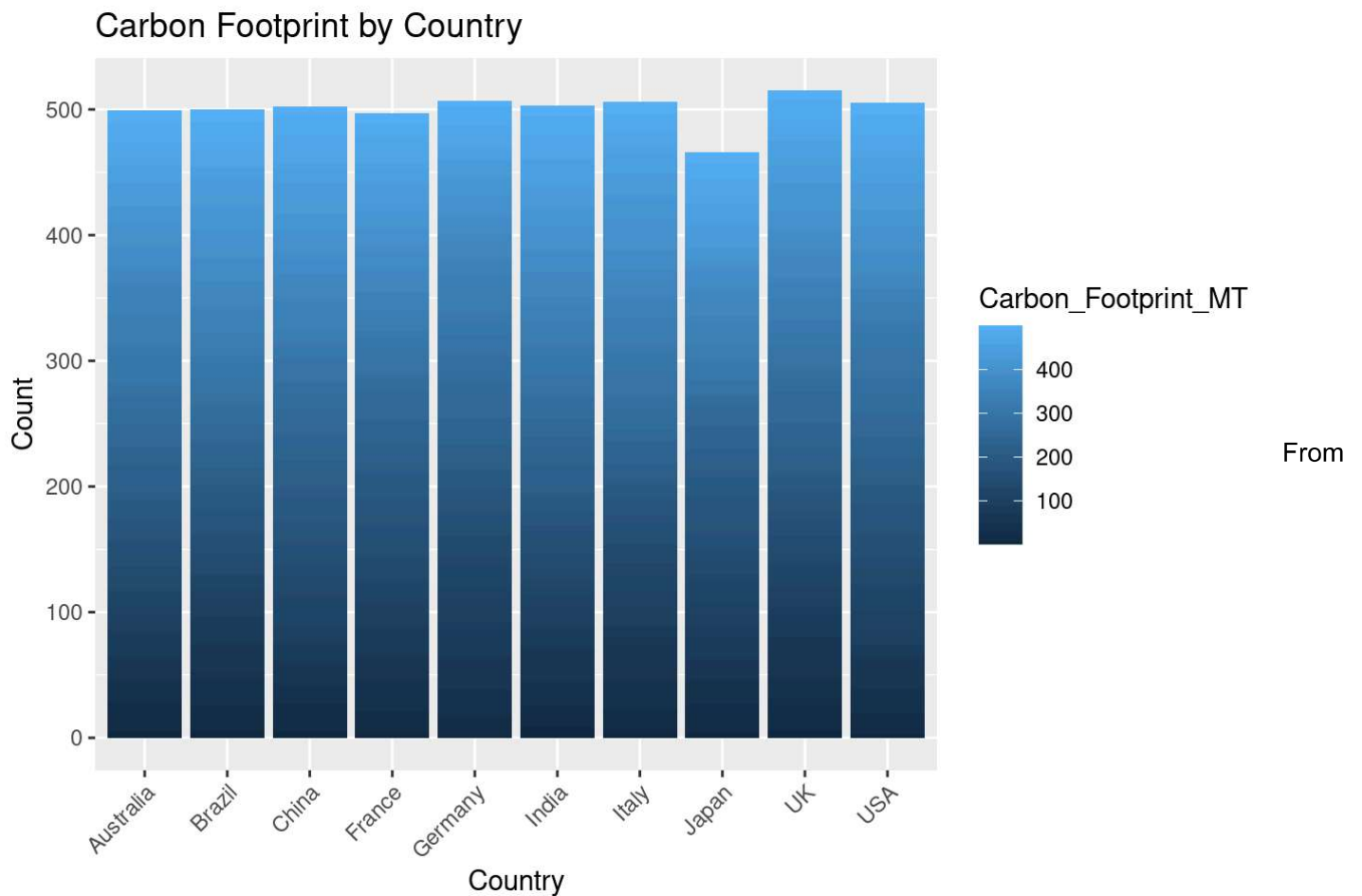
the chart above it shows a correlation between all country's having a similar carbon footprint except for Japan showing to have a noticablly smaller cabon footprint than the rest of the countrys.

```
water_usage <- data %>%
  group_by(Country, Water_Usage_Liters) %>%
  summarise(count= n())
```

```
## `summarise()` has grouped output by 'Country'. You can override using the
## `.groups` argument.
```

```
ggplot(water_usage, aes(x= Country, y= count, fill= Water_Usage_Liters)) +
  geom_bar(stat= "identity", postion= "dodge")+
  ggtitle("Water Usage by Country")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning in geom_bar(stat = "identity", postion = "dodge"): Ignoring unknown
## parameters: `postion`
```

Water Usage by Country

Similar to the carbon footprint graph above Japan seemingly has a lower water usuage in their production.

```
waste_production <- data %>%
  group_by(Country, Waste_Production_KG) %>%
  summarise(count = n())
```

```
## `summarise()` has grouped output by 'Country'. You can override using the
## `.groups` argument.
```

```
ggplot(waste_production, aes(x=Country, y= Waste_Production_KG, fill= Waste_Production_KG))+
  geom_bar(stat= "identity", postion= "dodge")+
  ggtitle("Waste Prodction by Country")+
  xlab("Country")+
  ylab("Waste Production")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning in geom_bar(stat = "identity", postion = "dodge"): Ignoring unknown
## parameters: `postion`
```

Waste Prodction by Country
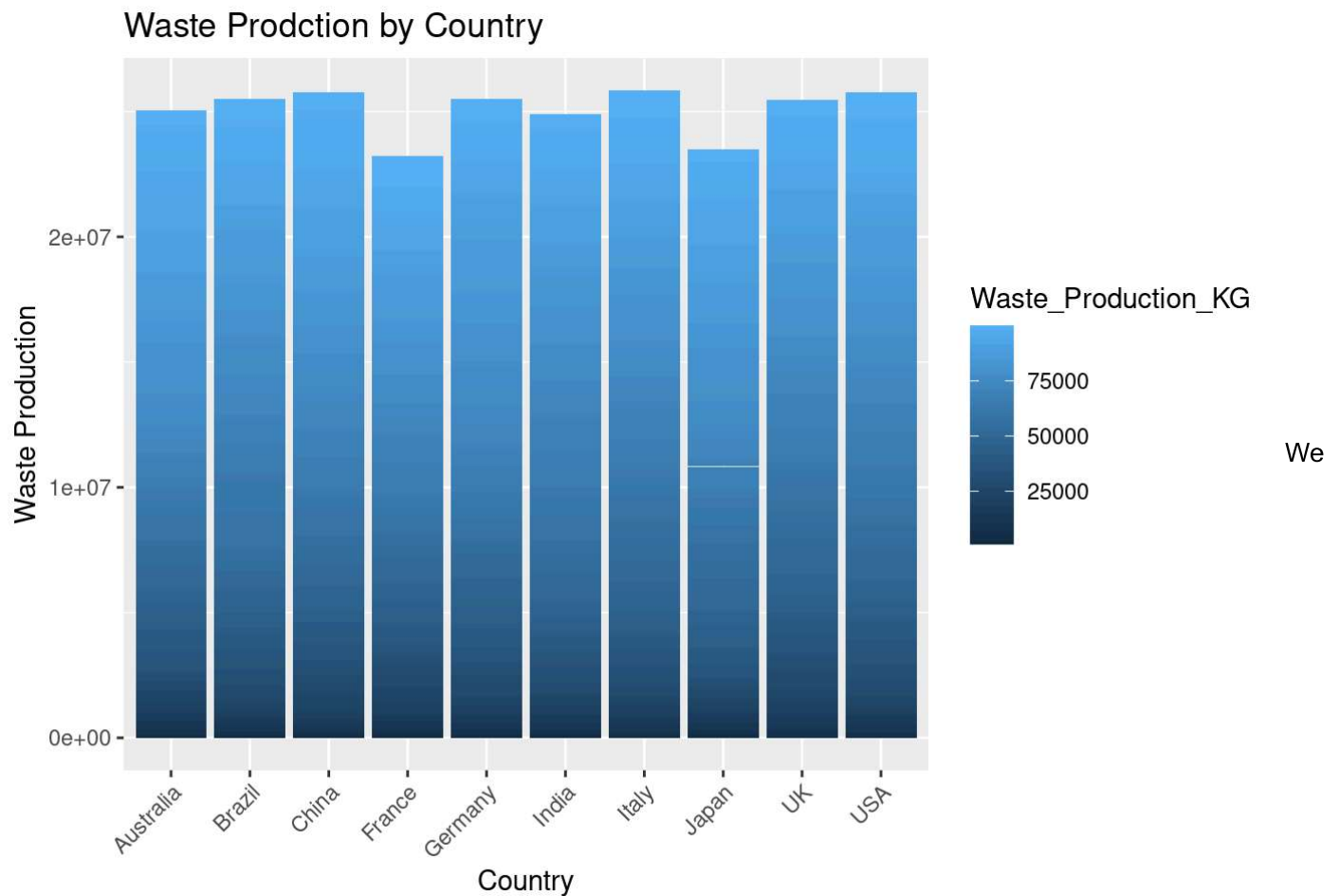
can start to see a trend in Japan having a lower amount of emmisions when creating clothing but France is showing in this particular graph to have a significally less waste production compared to the rest of the countries.

Now that we have compared the carbon footprint, water usuage, and waste prodution as factors that contribute to each country's sustainability let's see if price will show any contribution to that score as well.

```
ggplot(data, aes(x= Sustainability_Rating, y= Average_Price_USD, fill= Sustainability_Rating))+
  geom_boxplot()+
  ggtitle("Avergae Price by Sustainability Rating")+
  xlab("Sustainability Rating")+
  ylab("Average Price (USD)")
```

Avergae Price by Sustainability Rating

There isn't any trend shown in the graph above assuming that price effects the sustainability of the country. So we will see if the certifications show any trend in sustainability

```
ggplot(data, aes(x= Country, fill= Certifications))+
  geom_bar(stat= "Count", position= "dodge")+
  ggtitle("Certifications by Country")+
  xlab("Country")+
  ylab("Count")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Certifications by Country
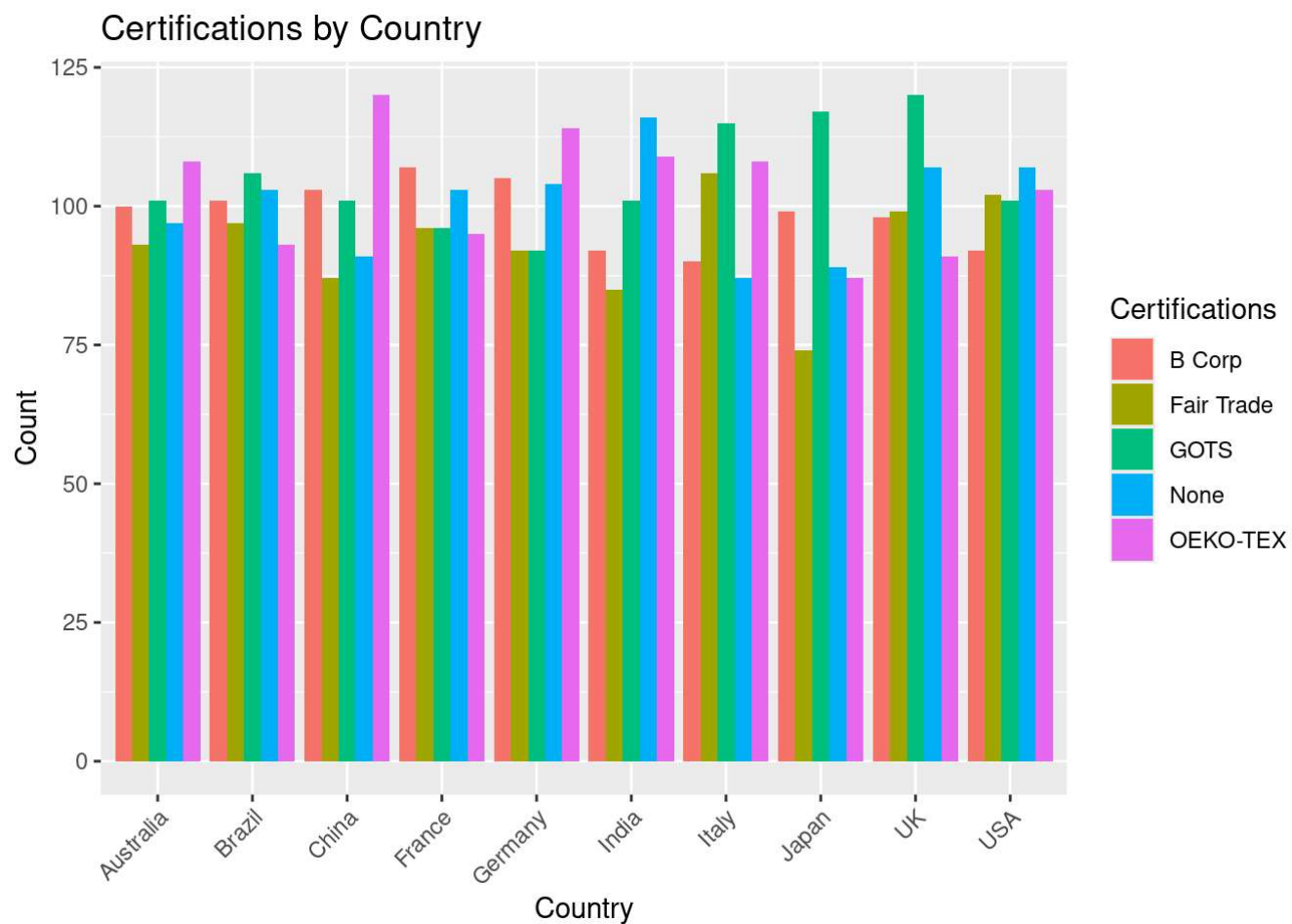
Finally we will see which country's have recycling programs in place.

```
ggplot(data, aes(x= Country, fill= Recycling_Programs))+
  geom_bar(stat="Count", position= "dodge")+
  ggtitle("Recycling Programs by Country")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Recycling Programs by Country