Sustainable Fashion Survey Analysis

Ju Kim (4156998)

2024-11-06

- import libraries

```
rm(list=ls())
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(lubridate)
library(dplyr)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(readr)
library(ggplot2)
library(tidyr)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(readxl)
library(reshape)
```

```
##
## Attaching package: 'reshape'
##
## The following object is masked from 'package:lubridate':
##
##     stamp
##
## The following object is masked from 'package:dplyr':
##
##     rename
##
## The following objects are masked from 'package:tidyr':
##
##     expand, smiths
```

- setting working directory and viewing dataset

```
setwd("~/Desktop/projects/sustainable fashion project")
fashion <- read_excel("fashion_survey.xlsx")
```

```
## New names:
## * `` -> `...1`
## * `Eco-friendly actions` -> `Eco-friendly actions...13`
## * `Eco-friendly actions` -> `Eco-friendly actions...143`
## * `` -> `...248`
## * `` -> `...249`
```

```
#View(fashion)
```

- cleaning data

```
fashion <- fashion %>%
  clean_names()

#glimpse(fashion)
unique(fashion$age)
```

```
## [1] "55-64" "25-34" "45-54" "18-24" "35-44"
```

```
unique(fashion$gender)
```

```
## [1] "NB" "M"  "F"
```

```
#changing data type as factor for gender column
fashion$gender <- as.factor(fashion$gender)
class(fashion$gender)
```

```
## [1] "factor"
```

```
#change 'NB' in gender column to 'NA'
fashion$gender[fashion$gender == "NB"] <- "NA"
```

```
## Warning in `[<-.factor`(`*tmp*`, fashion$gender == "NB", value =
## structure(c(NA, : invalid factor level, NA generated
```

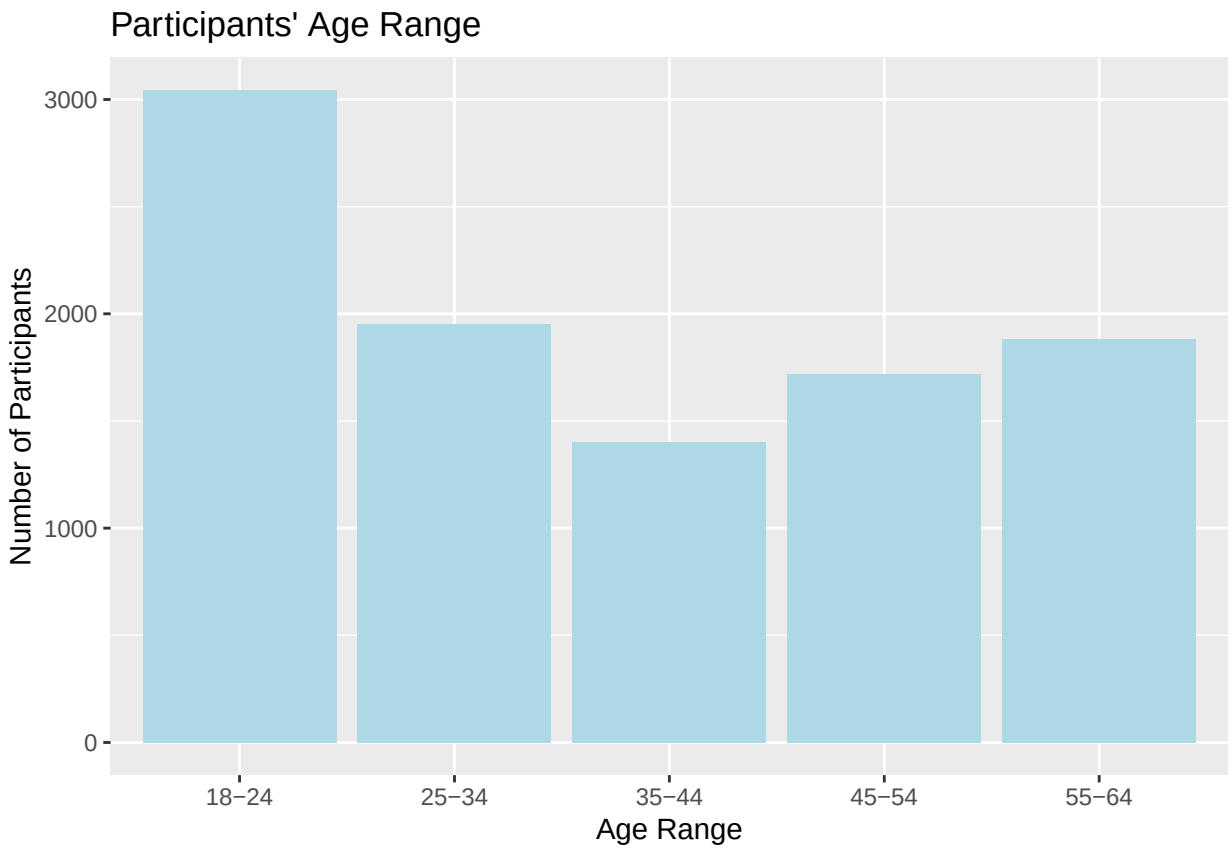
```
nrow(fashion)
```

```
## [1] 10000
```

```
#####

#####
# age range frequency
age <- data.frame(fashion$age)
age_separated <- age %>%
  mutate(row_id = row_number()) %>%
  clean_names()

#age_separated

age_frequency <- table(age_separated$fashion_age)
#print(age_frequency)
```

- EDA

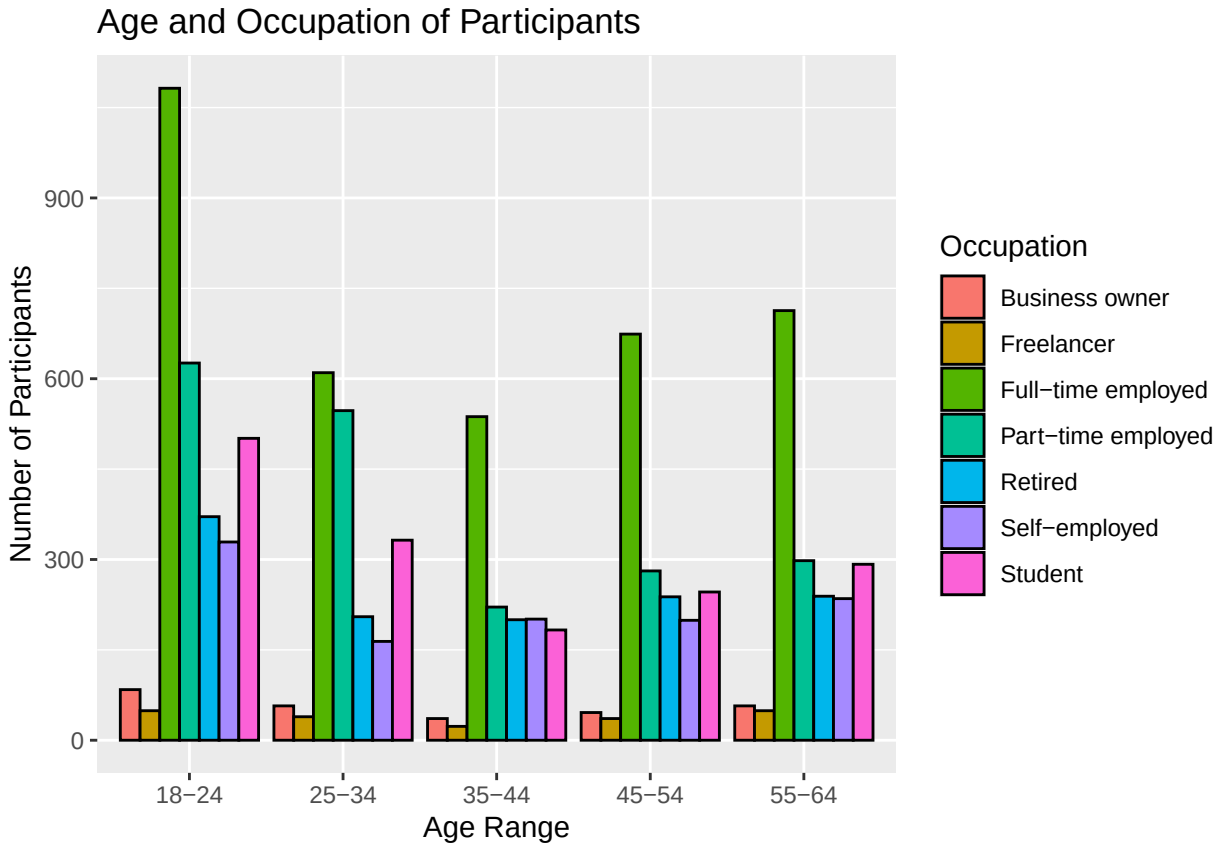
```
# plotting age range of participants
age_freq_df <- as.data.frame(age_frequency)

ggplot(age_freq_df, aes(x=Var1, y=Freq)) +
  geom_bar(stat = "identity", fill = "lightblue") +
  labs(title = "Participants' Age Range", x = "Age Range", y = "Number of Participants")
```



Respondents' ages ranged from 18-64, with the 18-24 age category reporting the highest number of respondents.

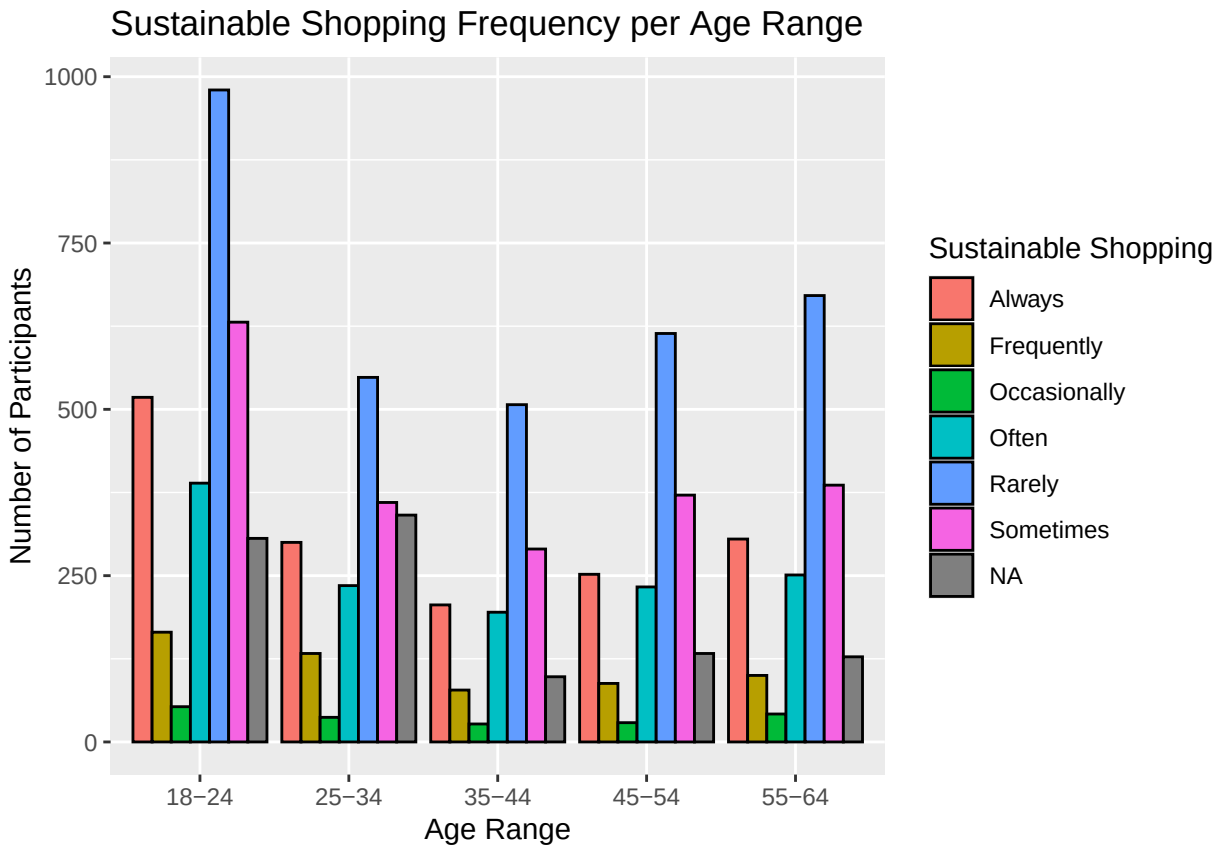
```
#plotting participant frequency per age range and occupation
fashion %>%
  mutate(age = factor(age, ordered = TRUE, levels = c("18-24", "25-34", "35-44", "45-54", "55-64"))) %>%
```

```
  ggplot(aes(x=age, fill = occupation)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "Age and Occupation of Participants", x = "Age Range", y = "Number of Participants", fil
```



The most common occupation for respondents consisted of a stable full-time job, with the next most common occupation being part-time employees and students for the younger respondents.

```
#plotting participant frequency per age range and sustainable shopping frequency
fashion %>%
  mutate(age = factor(age, ordered = TRUE, levels = c("18-24", "25-34", "35-44", "45-54", "55-64"))) %>%
  ggplot(aes(x=age, fill = sustainable_shopping_frequency)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "Sustainable Shopping Frequency per Age Range", x = "Age Range", y = "Number of Particip
```



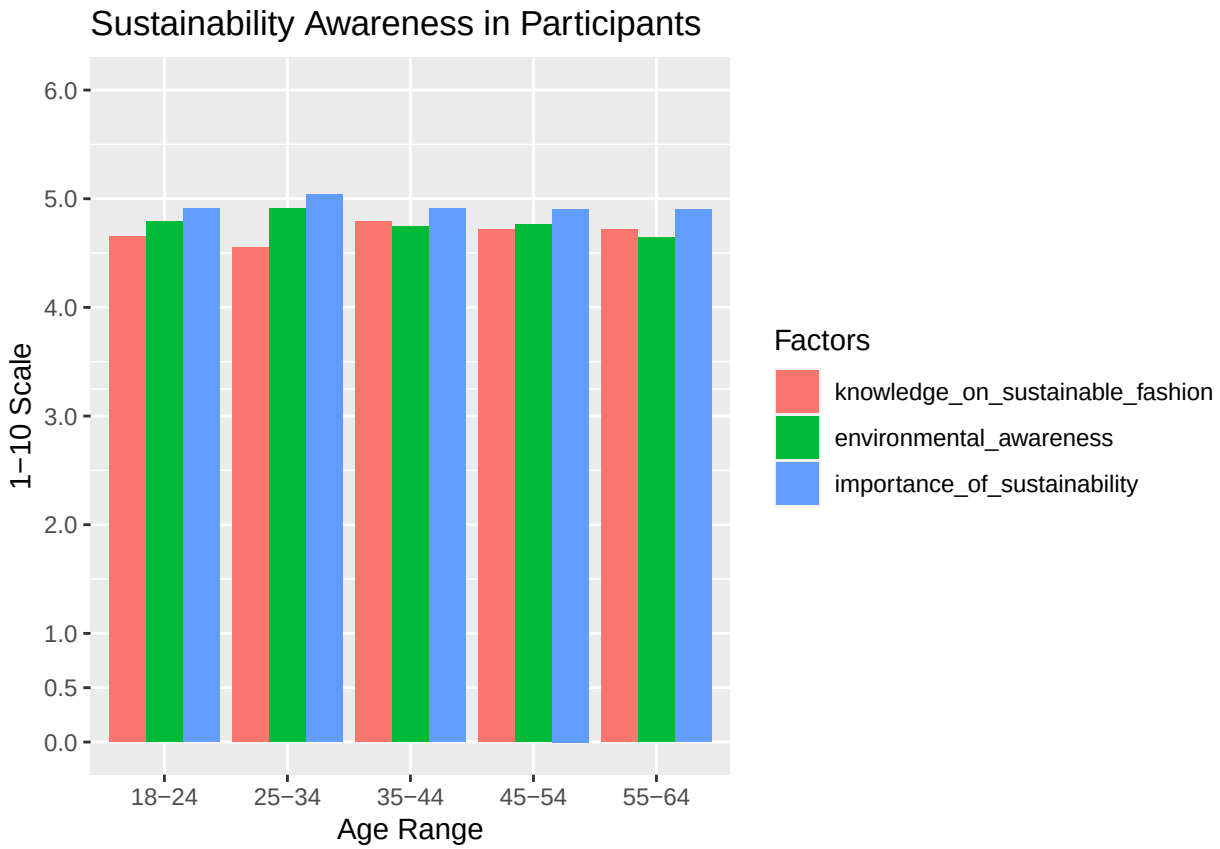
From this graph, an overwhelming number of respondents 'rarely' shop sustainably with 'sometimes' being the next most frequent answer to the "Do you shop sustainbly?" question.

```
#group by age,then show mean knowledge, environmental awareness, importance,

factors <- data.frame(fashion %>%
  select(age, knowledge_on_sustainable_fashion, environmental_awareness, importance_of_sustainability))
#factors

factors_melt <- factors %>%
  melt(id.vars = "age")
names(factors_melt) <- c("age", "factor", "scale")

head(factors_melt)
```

```
##     age                              factor scale
## 1 55-64 knowledge_on_sustainable_fashion     4
## 2 25-34 knowledge_on_sustainable_fashion     4
## 3 25-34 knowledge_on_sustainable_fashion     4
## 4 45-54 knowledge_on_sustainable_fashion     6
## 5 55-64 knowledge_on_sustainable_fashion     6
## 6 25-34 knowledge_on_sustainable_fashion     4
```

```
factors_melt %>%
  mutate(Age = factor(age, ordered = TRUE, levels = c("18-24", "25-34", "35-44", "45-54", "55-64"))) %>%
  ggplot(aes(x=age, y= scale, fill = factor)) +
  scale_y_continuous(limits = c(0,6), breaks = c(0:6, 0.5)) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean") +
  labs(title = "Sustainability Awareness in Participants", x = "Age Range", y = "1-10 Scale", fill = "Fa
```

```
## Warning: Removed 3681 rows containing non-finite values (`stat_summary()`).
```

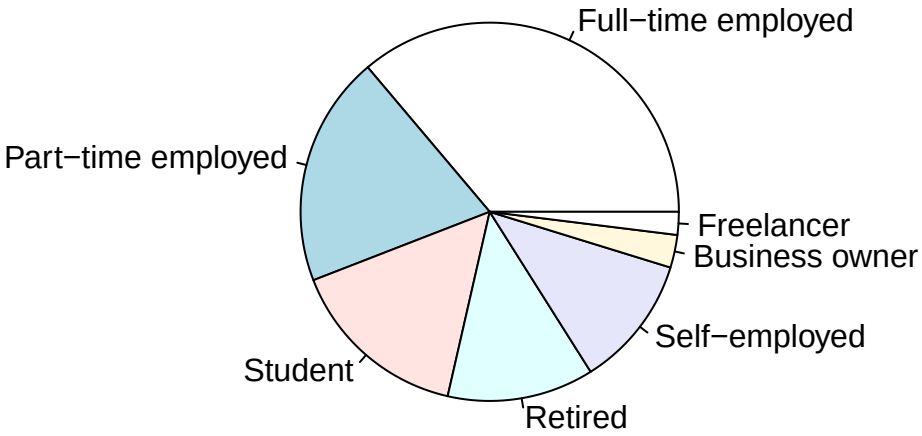


There is no clear difference in scale for the different age demographics on their knowledge on sustainable fashion, environmental awareness, or the importance of sustainaibility. Each age group averaged around 5 or below for each factor, revealing neutrality for these variables.

```
#group by occupation

occupation <- data.frame(fashion$occupation) %>%
  mutate(row_id = row_number()) %>%
  clean_names()

occupation_freq <- table(occupation$fashion_occupation)
occupation_freq
```

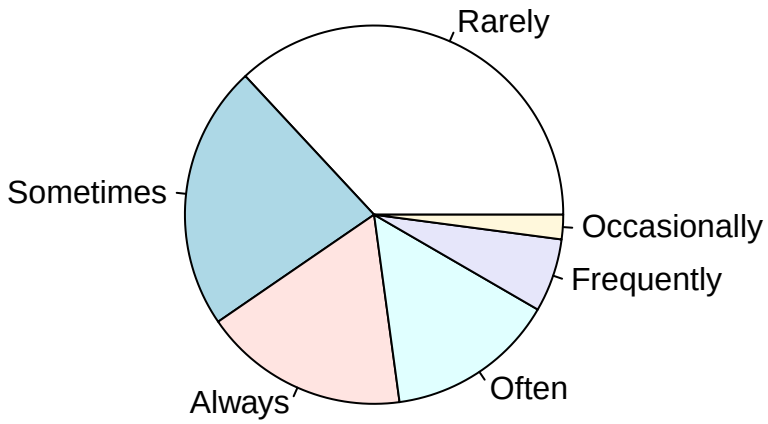
```
##
##     Business owner         Freelancer Full-time employed Part-time employed
##                280                196               3616               1973
##            Retired      Self-employed            Student
##               1253               1128               1554
```

```
occupation_df <- as.data.frame(occupation_freq)
names(occupation_df) <- c("Occupation", "Frequency")

occupation_df <- occupation_df %>% arrange(desc(Frequency))

library(RColorBrewer)
pie(occupation_df$Frequency, labels = occupation_df$Occupation)
```




```
#group by sustainable shopping frequency
ssf <- data.frame(fashion$sustainable_shopping_frequency) %>%
  mutate(row_id = row_number()) %>%
  clean_names()

sustainable_shopping_freq <- table(ssf$fashion_sustainable_shopping_frequency)
sustainable_shopping_df <- as.data.frame(sustainable_shopping_freq)
sustainable_shopping_df <- sustainable_shopping_df %>%
  arrange(desc(Freq))
names(sustainable_shopping_df) <- c("Sustainable Shopping", "Frequency")

pie(sustainable_shopping_df$Frequency, labels = sustainable_shopping_df$`Sustainable Shopping`,
    main = "Sustainable Shopping Frequency of Respondents")
```

Sustainable Shopping Frequency of Respondents




```
#group by occupation AND sustainable shopping frequency

occupation_ssf <- fashion %>%
  select(occupation, sustainable_shopping_frequency)

head(occupation_ssf) #missing data
```

```
## # A tibble: 6 x 2
##   occupation         sustainable_shopping_frequency
```

```
##   <chr>              <chr>
## 1 Full-time employed <NA>
## 2 Full-time employed Rarely
## 3 Full-time employed <NA>
## 4 Full-time employed Frequently
## 5 Full-time employed Sometimes
## 6 Student            Always
```

```
occupation_ssf <- drop_na(occupation_ssf) #drop na values

head(occupation_ssf)
```

```
## # A tibble: 6 x 2
##   occupation         sustainable_shopping_frequency
##   <chr>              <chr>
## 1 Full-time employed Rarely
## 2 Full-time employed Frequently
## 3 Full-time employed Sometimes
## 4 Student            Always
## 5 Full-time employed Rarely
## 6 Student            Rarely
```

```
occupation_ssf_freq <- occupation_ssf %>%
  count(occupation, sustainable_shopping_frequency) #count per subgroup

occupation_ssf_freq <- occupation_ssf_freq %>% arrange(desc(n)) #arrange by desc order

head(occupation_ssf_freq)
```

```
## # A tibble: 6 x 3
##   occupation         sustainable_shopping_frequency     n
##   <chr>              <chr>                          <int>
## 1 Full-time employed Rarely                          1238
## 2 Full-time employed Sometimes                        746
## 3 Full-time employed Always                           569
## 4 Part-time employed Rarely                           560
## 5 Full-time employed Often                            498
## 6 Student            Rarely                           495
```

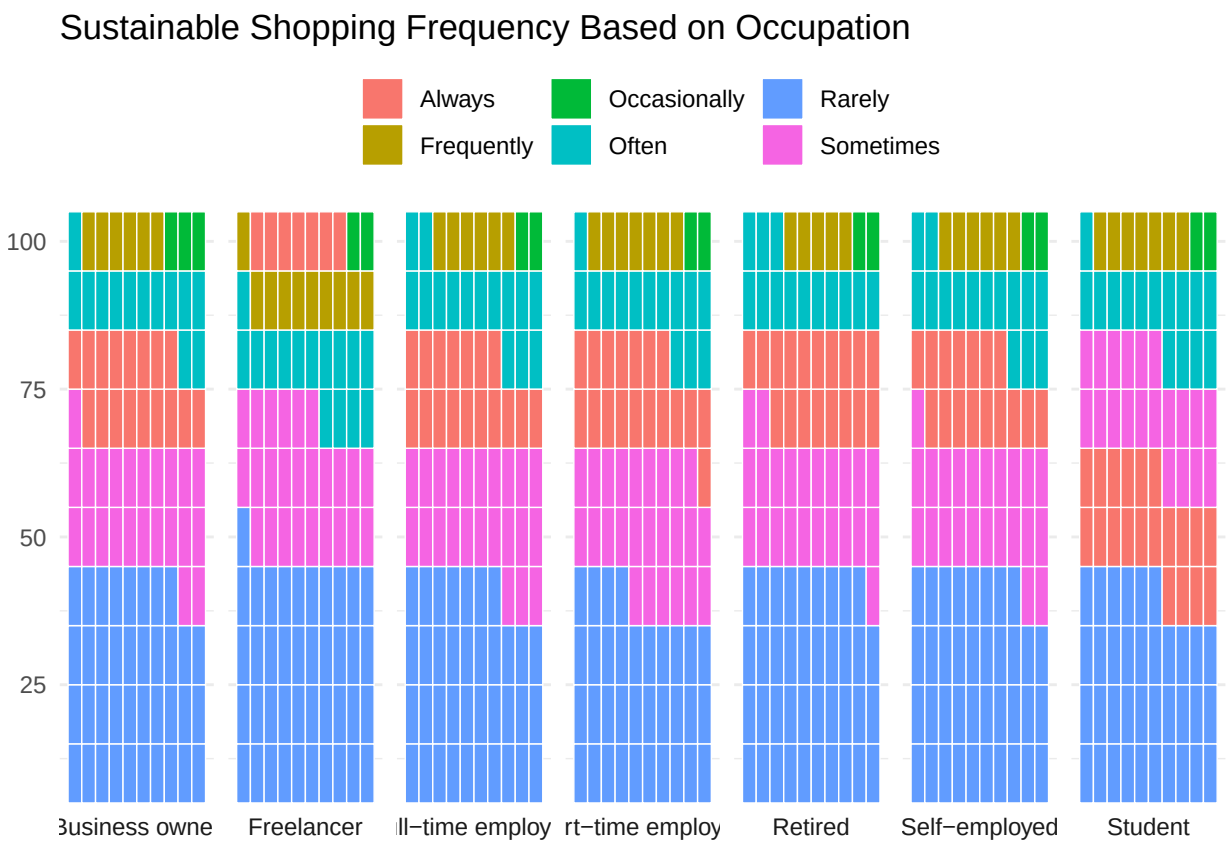
```
#install.packages("waffle")
#install.packages("ggtext")
#install.packages("showtext")
library(showtext)
```

```
## Loading required package: sysfonts
```

```
## Loading required package: showtextdb
```

```
library(ggtext)
library(waffle)

waf <- ggplot(occupation_ssf_freq, aes(fill=sustainable_shopping_frequency, values = n)) +
  geom_waffle(color = "white", size = .25, n_rows = 10, flip = TRUE, make_proportional = TRUE) +
  facet_wrap(~occupation, nrow = 1, strip.position = "bottom") +
  scale_x_discrete() +
  scale_y_continuous(labels = function(x) x * 10,
                     expand = c(0,0)) +
```

```
  labs(title = "Sustainable Shopping Frequency Based on Occupation") +
  theme_minimal() +
  theme(
    legend.position = "top",
    legend.title = element_blank()
  )

waf
```



The most notable difference from this waffle chart would be the smaller frequency of shoppers who "always" shop sustainably from the 'Freelancer' category. It's important to note that the biggest sustainable shopper demographic - who responded "always" - is from the 'Student' category.

```
###### mapping
#install.packages("sf")
library("sf")
```

```
## Warning: package 'sf' was built under R version 4.3.3
```

```
## Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE
```

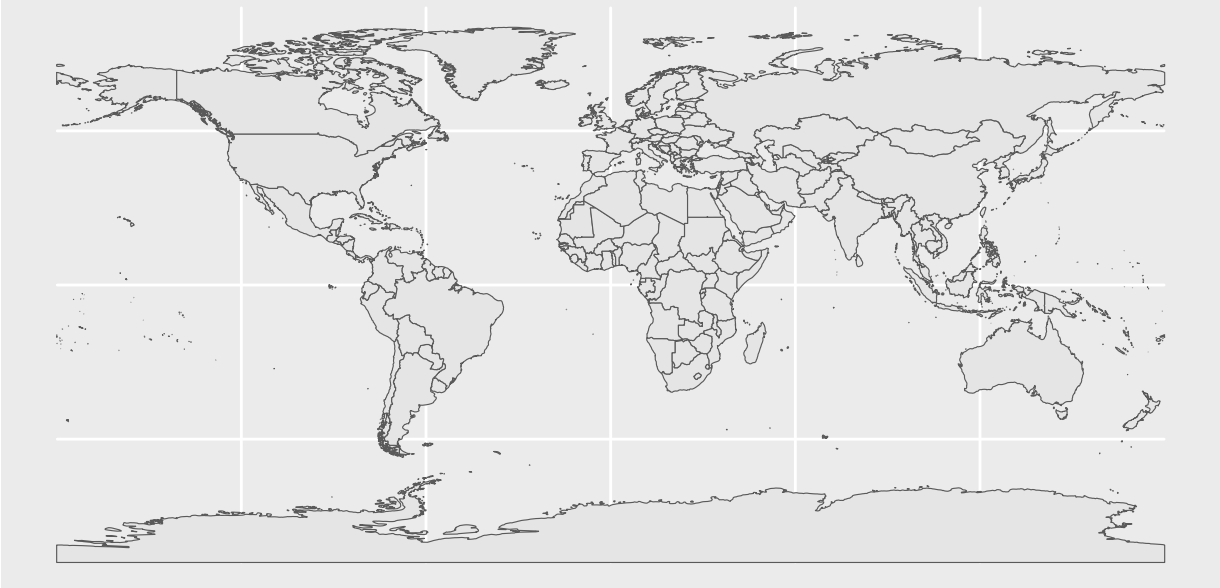
```
#install.packages("rnaturalearth")
library(rnaturalearth)
#install.packages("rnaturalearthdata")
library(rnaturalearthdata)
```

```
##
## Attaching package: 'rnaturalearthdata'
```

```
## The following object is masked from 'package:rnaturalearth':
```

```
##
##     countries110
```

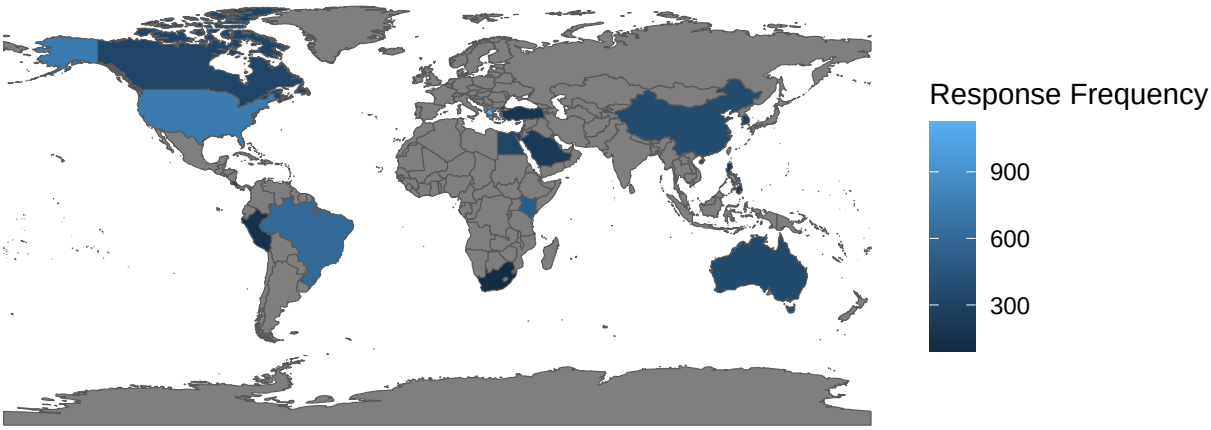
```
world <- ne_countries(scale="medium", returnclass = "sf")

ggplot(data=world) + geom_sf()
```



```
view(world)

country <- data.frame(fashion$country)
country_list <- country %>%
  mutate(row_id = row_number())
country_count <- table(country_list$fashion.country)
print(country_count)
```

```
##
##   AU  AUS   BR   CA   CN   CR   DE   EG   ES   FR   GR   IE   IN   IT   JP   KE
##  350  175  599  316  368  199  469  310  580  633 1125  179  346  520  295  502
##   KR   NO   PE   PH   PT   RU   SA   TR   UK   US   ZA
##  274  591  143  180   89  353  224  171  184  735   90
```

```
country_df <- data.frame(country_count)
names(country_df) <- c("postal", "frequency")

fashion_world <- merge(x = world, y=country_df, by="postal",
                       all.x = TRUE)
view(fashion_world)

ggplot(data=fashion_world) + geom_sf(aes(fill=frequency)) +
  labs(fill = "Response Frequency") +
  theme(legend.position = "bottom") +
  theme_void()
```



This map visualizes where the responses were collected from, with the highest number of respondents being from the U.S.