Files_Count

Edwin King'ori

2024-09-10

File_Counts Visual Analysis

The 'File_Count' table in the Data Summary sheet provides a complete set of data required for a comprehensive visual analysis to be conducted. For this reason, various visual graphs were generated to find the relationships of the given variables.

Loading the necessary Library Packages

```r
library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v ggplot2   3.5.1     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1

## -- 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 errors
```

```r
library(janitor)
```

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

```r
library(skimr)
library(ggplot2)
library(tidyr)
```

```r
library(readr)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

Importing the data-set

```r
files_count_df <- read_xlsx("Files_counts_summary.xlsx")
head(files_count_df)
```

```
## # A tibble: 6 x 6
##   Sector       `Client/Starter`    Gifts  Donors Promotions SustainerDonorCount
##   <chr>        <chr>               <dbl>   <dbl>      <dbl>               <dbl>
## 1 Environmental Envr            14608861 1987044  150998930                7915
## 2 social relief Social          11304354 1842730  101119876               84733
## 3 Veteran      Vet              50576096 3553488  296795118               33516
## 4 Advocacy     Advcy             1062540  275701   17048227                3377
## 5 Health       Health             907348  138062     231850                4052
## 6 Health2      Health2            227131   35306    6454338                 339
```

Inspecting the Data-Set

```r
summary(files_count_df)
```

```
##     Sector            Client/Starter         Gifts              Donors
##  Length:6           Length:6            Min.   :  227131    Min.   :  35306
##  Class :character   Class :character    1st Qu.:  946146    1st Qu.: 172472
##  Mode  :character   Mode  :character    Median : 6183447    Median :1059216
##                                         Mean   :13114388    Mean   :1305388
##                                         3rd Qu.:13782734    3rd Qu.:1950966
##                                         Max.   :50576096    Max.   :3553488
##    Promotions         SustainerDonorCount
##  Min.   :   231850    Min.   :  339
##  1st Qu.:  9102810    1st Qu.: 3546
##  Median : 59084052    Median : 5984
##  Mean   : 95441390    Mean   :22322
##  3rd Qu.:138529166    3rd Qu.:27116
##  Max.   :296795118    Max.   :84733
```

```r
str(files_count_df)
```

```
## tibble [6 x 6] (S3: tbl_df/tbl/data.frame)
##  $ Sector            : chr [1:6] "Environmental" "social relief" "Veteran" "Advocacy" ...
##  $ Client/Starter     : chr [1:6] "Envr" "Social" "Vet" "Advcy" ...
##  $ Gifts             : num [1:6] 14608861 11304354 50576096 1062540 907348 ...
##  $ Donors            : num [1:6] 1987044 1842730 3553488 275701 138062 ...
##  $ Promotions        : num [1:6] 1.51e+08 1.01e+08 2.97e+08 1.70e+07 2.32e+05 ...
##  $ SustainerDonorCount: num [1:6] 7915 84733 33516 3377 4052 ...
```

The data is clean and credible for the next visualization phase However, since the data is wide, it has to be transformed to a long format for it to be compatible with the desired plot.

In this case: Reshaping the data to long format for it to display the legend in the visualization.

```r
files_count_long_df <- files_count_df %>%
  pivot_longer(cols = c(Gifts, Promotions, Donors),
               names_to = "Category",
               values_to = "Count")
```

#1 Generating a horizontal bar plot to compare the gifts, donor and promotions across all sectors

```r
ggplot(files_count_long_df, aes(x = Sector,y = Count, fill = Category))+
  geom_bar(stat = "identity", position = "dodge")+
  scale_y_continuous(labels = scales::comma) +
  scale_fill_manual(values = c("Gifts" = "blue", "Donors" = "red", "Promotions" = "green")) +
  theme_minimal()+
  labs(title = "Comparing the gifts, donor and  promotions across all sectors",
       x = "Sector", y = "Count")+
  theme(axis.text.x = element_text(angle = 90, hjust = 0.1))+

  coord_flip()
```

# Comparing the gifts, donor and promotions across all sectors
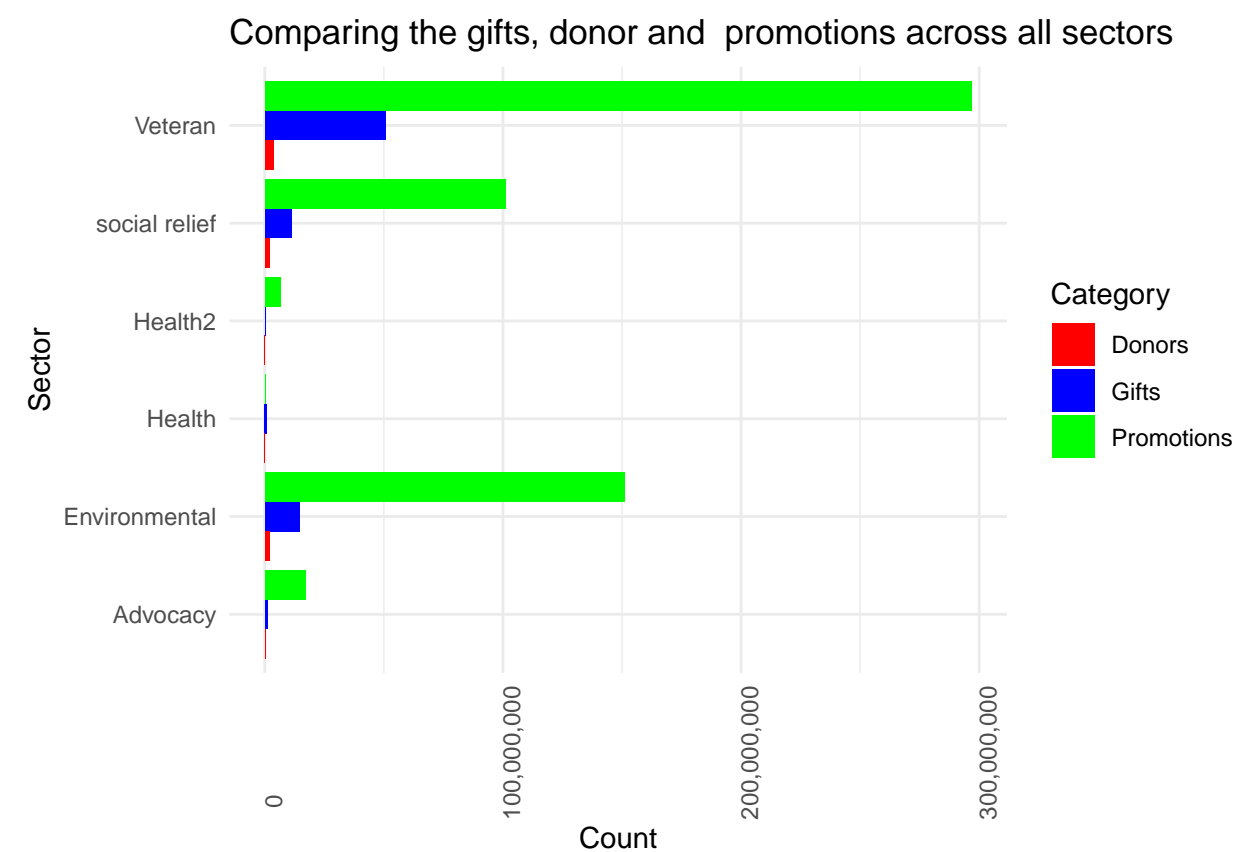
2 Visualizing the Correlation of Gifts, Donors, Promotions and Sustainable Donor Count by Sector using a Stacked bar plot

```
# transforming the data from a wide data to a long data using the melt function
files_count_long <- melt(files_count_df, id.vars = "Sector", measure.vars = c("Gifts", "Donors", "Promot

ggplot(files_count_long, aes(x = Sector, y = value, fill = variable)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Breakdown of Gifts, Donors, Promotions, and Sustainer Donor Count by Sector", x = "Secto
  scale_y_continuous(labels = comma)+
  scale_fill_manual(values = c("Gifts" = "blue", "Donors" = "red", "Promotions" = "green", "SustainerDo
```

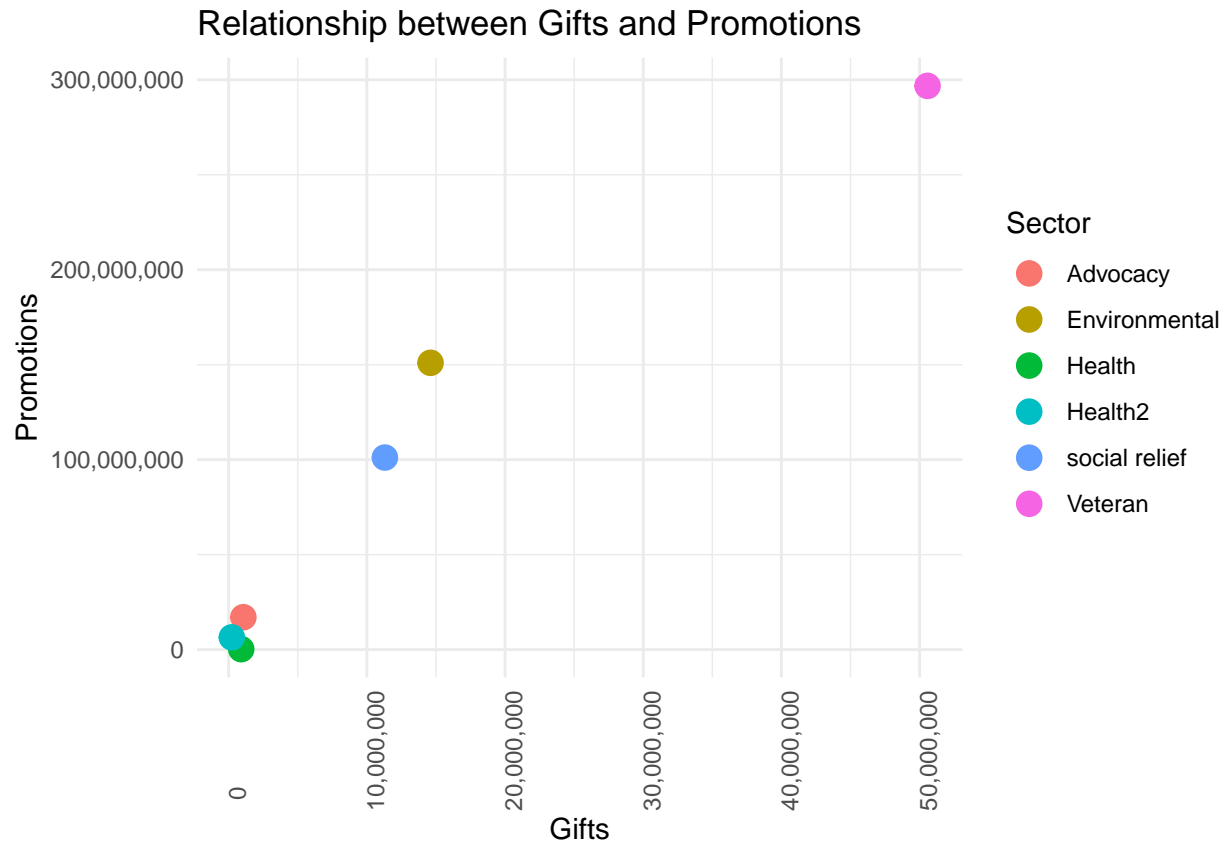# Breakdown of Gifts, Donors, Promotions, and Sustainer Donor Count
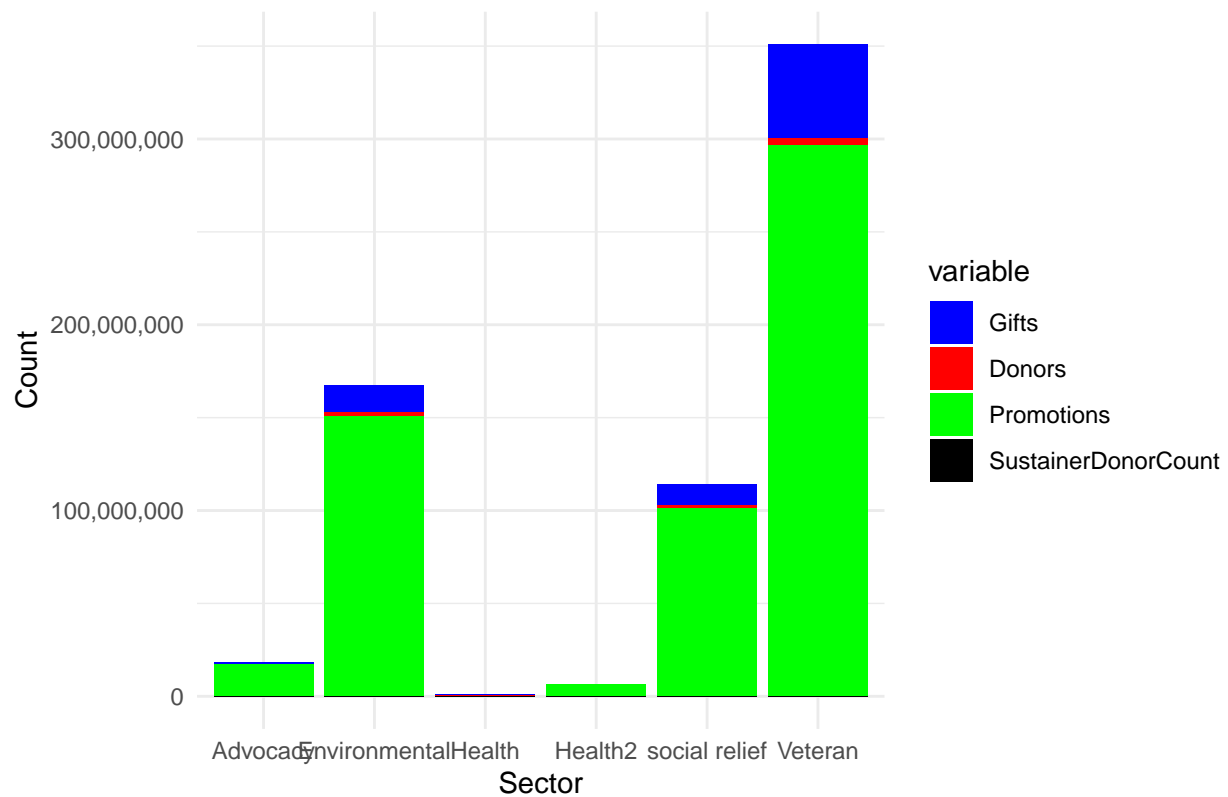
3 Visualizing the Relationship between Gifts and Donors

```
ggplot(files_count_df, aes(x = Gifts, y = Donors, color = Sector)) +
  geom_point(size = 4) +
  theme_minimal() +
  labs(title = "Relationship between Gifts and Donors", x = "Gifts", y = "Donors")+
  scale_y_continuous(labels = comma)+
  scale_x_continuous(labels = comma)+
  theme(axis.text.x = element_text(angle =90 , hjust = 0.1))
```

Relationship between Gifts and Donors
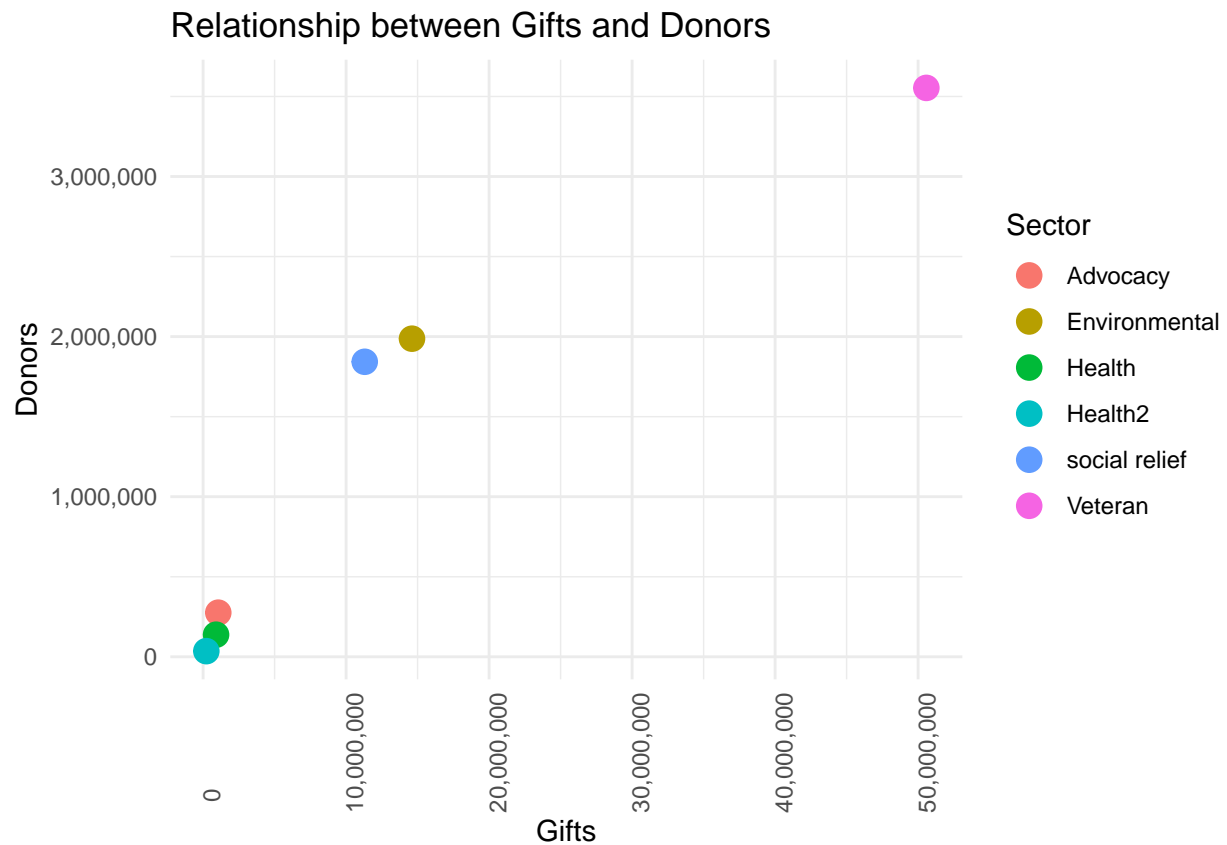
4 Visualizing the Relationship between Gifts and Promotions

```r
ggplot(files_count_df, aes(x = Gifts, y = Promotions, color = Sector)) +
  geom_point(size = 4) +
  theme_minimal() +
  labs(title = "Relationship between Gifts and Promotions", x = "Gifts", y = "Promotions")+
  scale_y_continuous(labels = comma)+
  scale_x_continuous(labels = comma)+
  theme(axis.text.x = element_text(angle =90 , hjust = 0.1))
```

Relationship between Gifts and Promotions

5. Visualizing the Relationship between Donors and Promotions

```
ggplot(files_count_df, aes(x = Donors, y = Promotions, color = Sector)) +
  geom_point(size = 4) +
  theme_minimal() +
  labs(title = "Relationship between Donors and Promotions", x = "Donors", y = "Promotions")+
  scale_y_continuous(labels = comma)+
  scale_x_continuous(labels = comma)+
  theme(axis.text.x = element_text(angle =90 , hjust = 0.1))
```

Relationship between Donors and Promotions
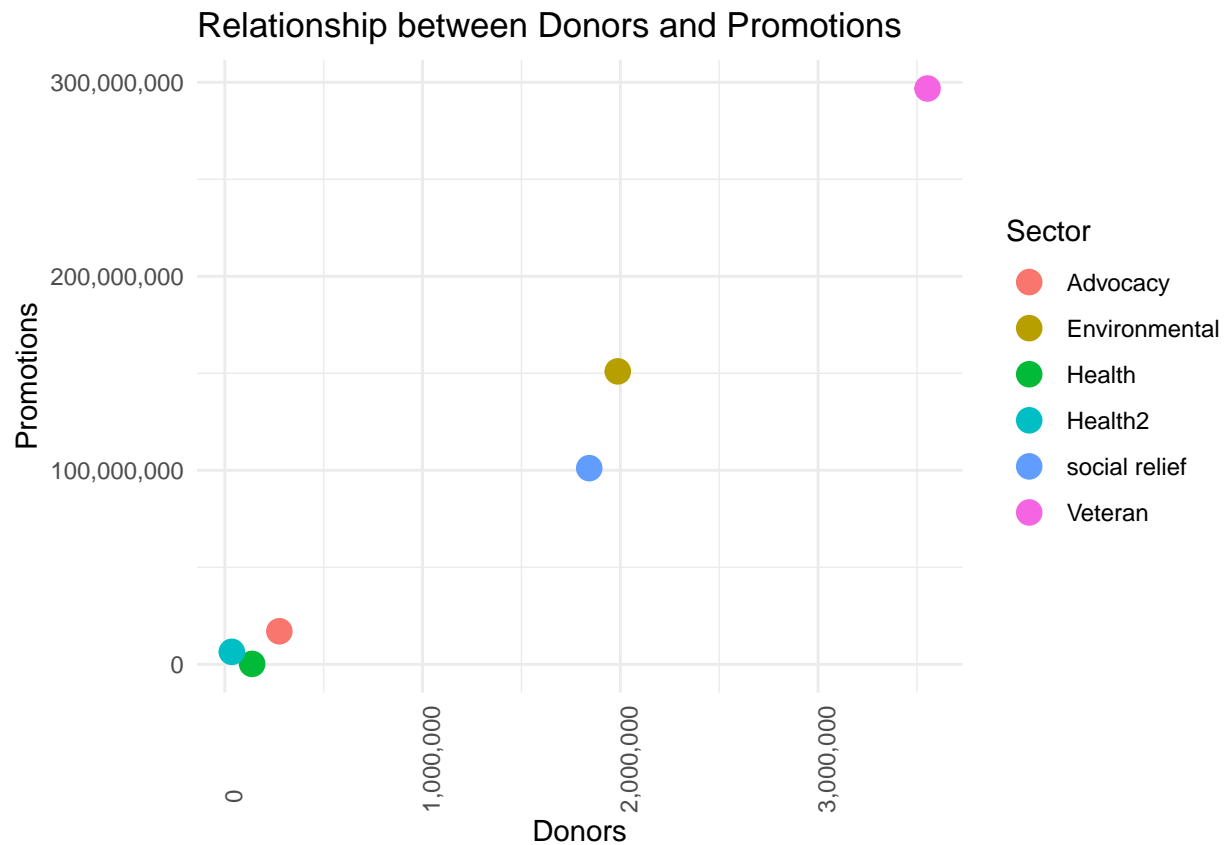
6. Heatmap: Correlation of matrix

```
file_matrix <- cor(files_count_df[, c("Gifts", "Donors", "Promotions", "SustainerDonorCount")])
ggplot(melt(file_matrix), aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0, limit = c(-1, 1), space
  theme_minimal() +
  labs(title = "Heatmap of Correlation Matrix", x = "", y = "")
```

Heatmap of Correlation Matrix
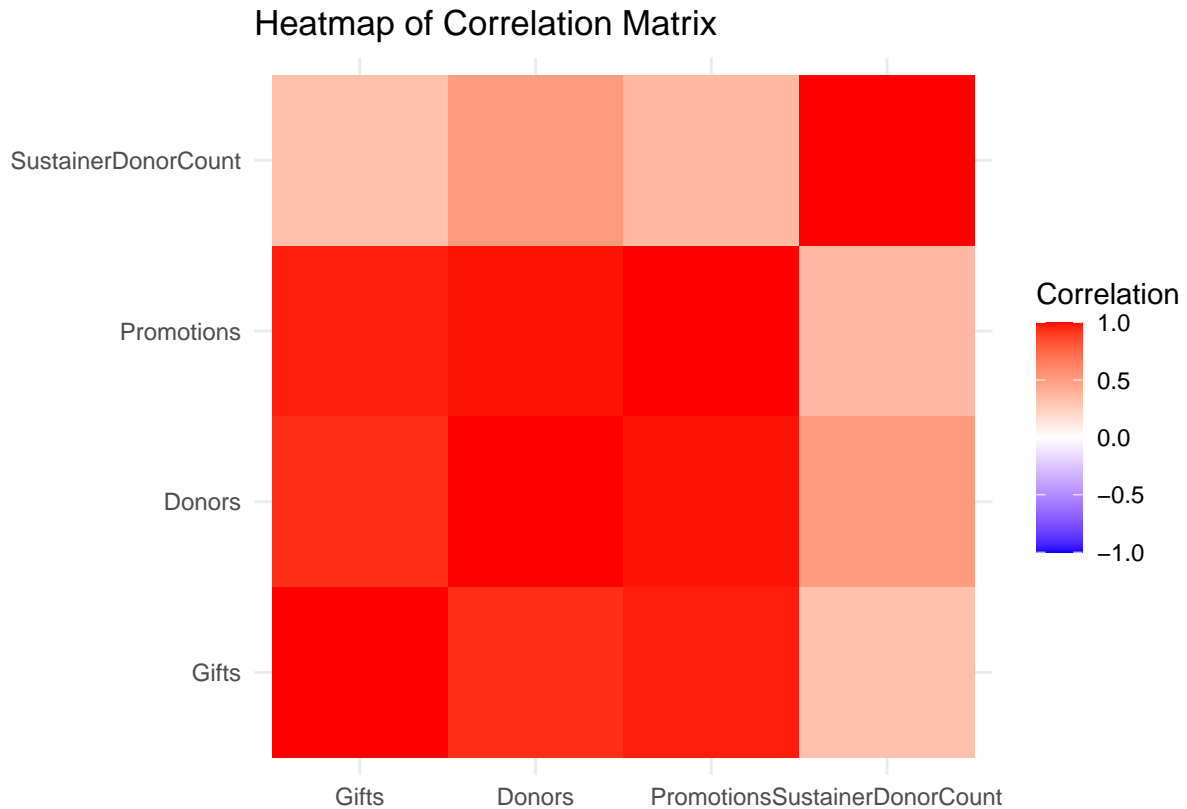
Visual Translation

In all the plots, the variables Gifts, Donors and Promotions exhibit strong positive relationship, which is indicated by the higher values in these categories correlating with each other. However, there are noticeable differences of counts across the different sectors. For instance, the veterans receive a higher rate of promotions compared to sectors like health, which receive a higher number of gifts. . Despite the huge count differences between the variables, the intense red coloring exhibited in the heatmap confirms a positive correlation of the three variables. The sustainerDonorCount variable shows a moderate correlation with other variables but a much stronger relationship with the Donors variable.