Business Tracking

Susie Nguyen

2022-04-19

I analyzed my small business transactions to look for long-term trends on outgoing expenses by using R. My sample data set can be obtained at my GitHub Repository. In this project,I used ggplot2 to find out:

- Comparison of Different Order Types
- Most Purchased Categories

Setting up my environment

Notes: setting up my environment by loading the `tidyverse` and `plyr` packages:

```
library(plyr)
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::arrange()   masks plyr::arrange()
## x purrr::compact()   masks plyr::compact()
## x dplyr::count()     masks plyr::count()
## x dplyr::failwith()  masks plyr::failwith()
## x dplyr::filter()    masks stats::filter()
## x dplyr::id()        masks plyr::id()
## x dplyr::lag()       masks stats::lag()
## x dplyr::mutate()    masks plyr::mutate()
## x dplyr::rename()    masks plyr::rename()
## x dplyr::summarise() masks plyr::summarise()
## x dplyr::summarize() masks plyr::summarize()
```

Importing data set

Here we will go through a series of visualizations

```
o <- read_csv('Orders.csv')

## Rows: 537 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (8): OrderNumber, Date, CustomerID, Website, Description, Type, Color, Size
## dbl (4): Quantity, Price, Tax, Total
##
```

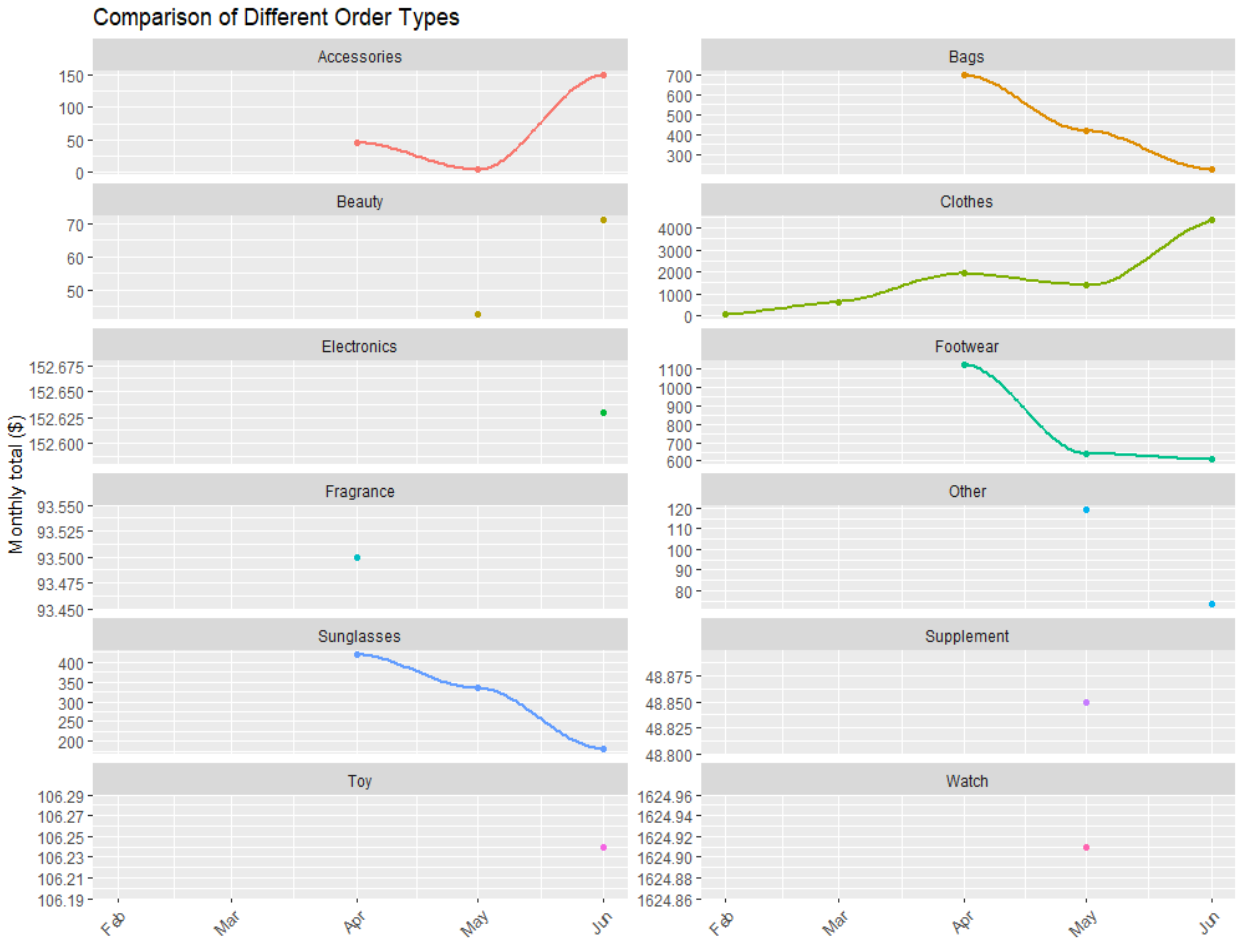
```
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

head(o)

## # A tibble: 6 x 12
##   OrderNumber Date     CustomerID Website Description Type  Color Size  Quantity
##   <chr>       <chr>    <chr>      <chr>   <chr>       <chr> <chr> <chr>    <dbl>
## 1 HM-A0241698 2/28/20~ A02        H&M     Satin Dress Clot~ Black 4            1
## 2 HM-A0241698 2/28/20~ A02        H&M     Boucle Car~ Clot~ Black S            1
## 3 HM-A0141698 2/28/20~ A01        H&M     Jerey Dress Clot~ Black M            1
## 4 HM-A0241698 2/28/20~ A02        H&M     Jersey Top  Clot~ Stri~ S            1
## 5 LC-A0141702 3/4/2014 A01        Lacoste Polo        Clot~ Cour~ M            1
## 6 LC-A0141702 3/4/2014 A01        Lacoste Shirt       Clot~ Cour~ M            1
## # ... with 3 more variables: Price <dbl>, Tax <dbl>, Total <dbl>

str(o)

## spec_tbl_df [537 x 12] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ OrderNumber: chr [1:537] "HM-A0241698" "HM-A0241698" "HM-A0141698" "HM-
A0241698" ...
##  $ Date       : chr [1:537] "2/28/2014" "2/28/2014" "2/28/2014" "2/28/2014" ...
##  $ CustomerID : chr [1:537] "A02" "A02" "A01" "A02" ...
##  $ Website    : chr [1:537] "H&M" "H&M" "H&M" "H&M" ...
##  $ Description: chr [1:537] "Satin Dress" "Boucle Cardigan" "Jerey Dress" "Jersey
Top" ...
##  $ Type       : chr [1:537] "Clothes" "Clothes" "Clothes" "Clothes" ...
##  $ Color      : chr [1:537] "Black" "Black" "Black" "Stripe" ...
##  $ Size       : chr [1:537] "4" "S" "M" "S" ...
##  $ Quantity   : num [1:537] 1 1 1 1 1 1 1 1 1 1 ...
##  $ Price      : num [1:537] 7 20 10 9.95 67.99 ...
##  $ Tax        : num [1:537] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Total      : num [1:537] 7 20 10 9.95 67.99 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   OrderNumber = col_character(),
##   ..   Date = col_character(),
##   ..   CustomerID = col_character(),
##   ..   Website = col_character(),
##   ..   Description = col_character(),
##   ..   Type = col_character(),
##   ..   Color = col_character(),
##   ..   Size = col_character(),
##   ..   Quantity = col_double(),
##   ..   Price = col_double(),
##   ..   Tax = col_double(),
##   ..   Total = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>

o$date <- as.Date(o$Date, format = "%m/%d/%Y")
o$month <- as.Date(cut(o$date, breaks = "month"))

smr <- ddply(o, .(month, Type), summarise, cost = sum(Total), count = sum(Quantity))
smr <- smr[1:26,]
```

Visualizing

Comparison of Different Order Types

Here, we plot monthly expense of different order types

```
ggplot(smr, aes(month, cost, col=Type)) +
  facet_wrap(~Type, ncol=2, scale="free_y") +
  geom_smooth(method="loess", se=F) + geom_point() +
  theme(axis.text.x=element_text(angle=45, hjust=1),
        legend.position="none") +
  labs(title="Comparison of Different Order Types",
       x="", y="Monthly total ($)")
```



Most Purchased Categories

Here, we plot flipper length against body mass and look at the breakdown by species and sex

```
yl <- ddply(smr, .(Type), summarise, m=mean(count))

ggplot(yl, aes(x=Type, y=m, fill = Type)) +
  geom_bar(stat="identity") +
  labs(y="Average monthly expense ($)", x="",
       title="Most Purchased Categories")
```