

Projet : Super Market Sales Visualizations

Tek-Up School

2023-2024

Présenté par :

- Nour Baklouti
- Rania Dridi

Encadré par :

- Dr. Nadjib Mohamed Mehdi BENDAOUD

Supermarket Data Visualization

Ce rapport vise à visualiser et à analyser :

1. La relation des clients avec le supermarché.
2. Les méthodes de paiement utilisées dans le supermarché.
3. La relation entre les produits et les quantités.
4. Les produits et leurs évaluations.
5. Les types de produits et leurs ventes.

=> Pour plus d'informations, vous trouverez ci-joint le lien vers notre notebook Kaggle.

<https://www.kaggle.com/code/nourbaklouti/bi-supermarket>.

Résumé du DataFrame

```
library(ggplot2)
```

```
## Warning: le package 'ggplot2' a été compilé avec la version R 4.3.3
```

```
library(tidyverse)
```

```
## Warning: le package 'tidyverse' a été compilé avec la version R 4.3.3
```

```
## Warning: le package 'tidyr' a été compilé avec la version R 4.3.2
```

```
## Warning: le package 'readr' a été compilé avec la version R 4.3.2
```

```
## Warning: le package 'purrr' a été compilé avec la version R 4.3.2
```

```
## Warning: le package 'dplyr' a été compilé avec la version R 4.3.2
```

```
## Warning: le package 'stringr' a été compilé avec la version R 4.3.2
```

```
## Warning: le package 'forcats' a été compilé avec la version R 4.3.2
```

```
## Warning: le package 'lubridate' a été compilé avec la version R 4.3.3
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.3      v readr      2.1.4
```

```
## v forcats    1.0.0      v stringr    1.5.0
```

```
## v lubridate  1.9.3      v tibble     3.2.1
```

```
## v purrr      1.0.2      v tidyr      1.3.0
```

```
## -- 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
```

```
basesm<-read.csv("C:/Users/nourb/Desktop/SupermarketSales/bdsupermarket_sales.csv")
```

```
summary(basesm)
```

```
## Invoice.ID           Branch           City           Customer.type
## Length:1000         Length:1000         Length:1000         Length:1000
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##
##
##
## Gender             Product.line         Unit.price           Quantity
## Length:1000         Length:1000         Min. :10.08         Min. : 1.00
## Class :character    Class :character    1st Qu.:32.88       1st Qu.: 3.00
## Mode :character     Mode :character    Median :55.23       Median : 5.00
##                                     Mean :55.67       Mean : 5.51
##                                     3rd Qu.:77.94    3rd Qu.: 8.00
##                                     Max. :99.96       Max. :10.00
## Tax.5.             Total              Date                Time
## Min. : 0.5085       Min. : 10.68       Length:1000         Length:1000
## 1st Qu.: 5.9249     1st Qu.: 124.42    Class :character    Class :character
## Median :12.0880     Median : 253.85    Mode :character     Mode :character
## Mean :15.3794       Mean : 322.97
## 3rd Qu.:22.4453     3rd Qu.: 471.35
## Max. :49.6500       Max. :1042.65
## Payment            cogs              gross.margin.percentage gross.income
## Length:1000         Min. : 10.17       Min. :4.762         Min. : 0.5085
## Class :character    1st Qu.:118.50     1st Qu.:4.762         1st Qu.: 5.9249
## Mode :character     Median :241.76     Median :4.762         Median :12.0880
##                                     Mean :307.59       Mean :4.762         Mean :15.3794
##                                     3rd Qu.:448.90    3rd Qu.:4.762         3rd Qu.:22.4453
##                                     Max. :993.00       Max. :4.762         Max. :49.6500
##
## Rating
## Min. : 4.000
## 1st Qu.: 5.500
## Median : 7.000
## Mean : 6.973
## 3rd Qu.: 8.500
## Max. :10.000
```

Description des variables

Les variables présentes dans le jeu de données sont les suivantes :

- Invoice ID : Identifiant unique de la facture.
- Branch : Succursale du supermarché (A, B, C).
- City : Ville où se trouve le supermarché.

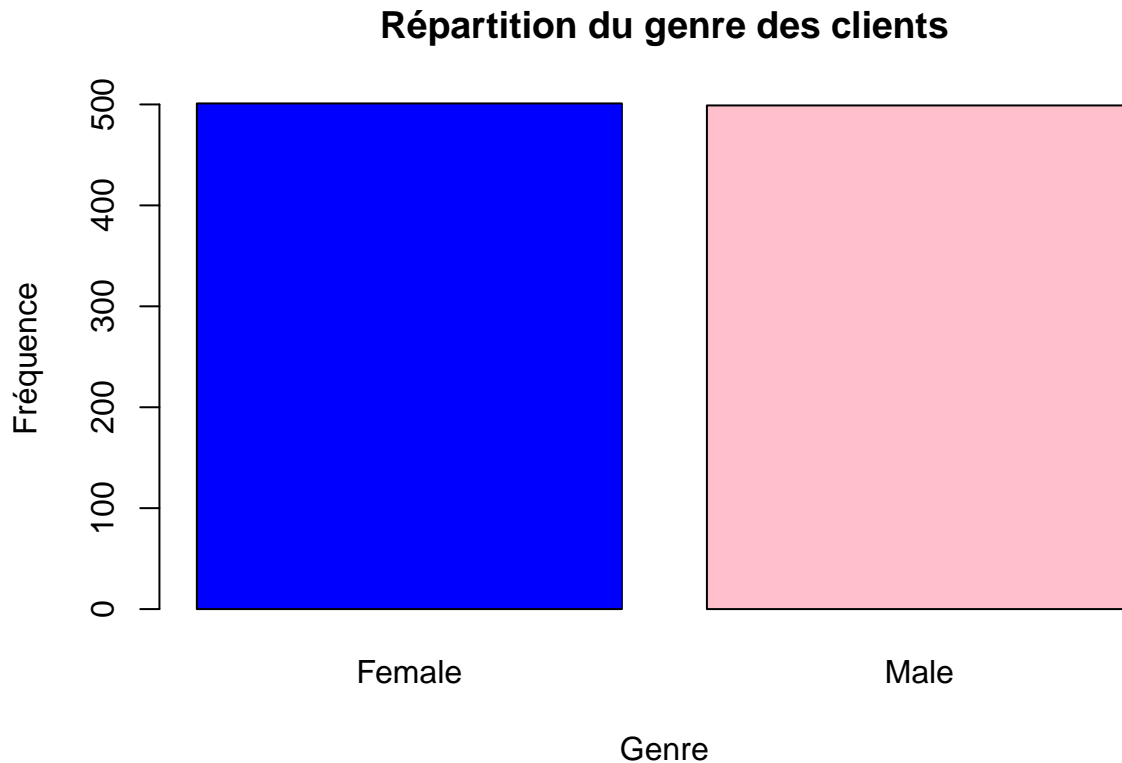
- Customer type :Type de client (Member pour membre, Normal pour non-membre).
- Gender : Genre du client (Male pour homme, Female pour femme).
- Product line :Catégorie du produit.
- Unit price : Prix unitaire du produit.
- Quantity :Quantité d'articles achetés.
- Tax 5% : Montant de la taxe (5%) sur l'achat.
- Total :Montant total de l'achat (y compris la taxe).
- Date : Date de l'achat .
- Time : Heure de l'achat .
- Payment : Mode de paiement (Ewallet, Cash, Credit card).
- cogs (Cost of Goods Sold) :Coût des biens vendus.
- gross margin percentage :Pourcentage de marge brute.
- gross income : Revenu brut
- Rating : Évaluation du client pour l'expérience d'achat

Visualization des données

La relation des clients avec le supermarché:

la distribution par genre

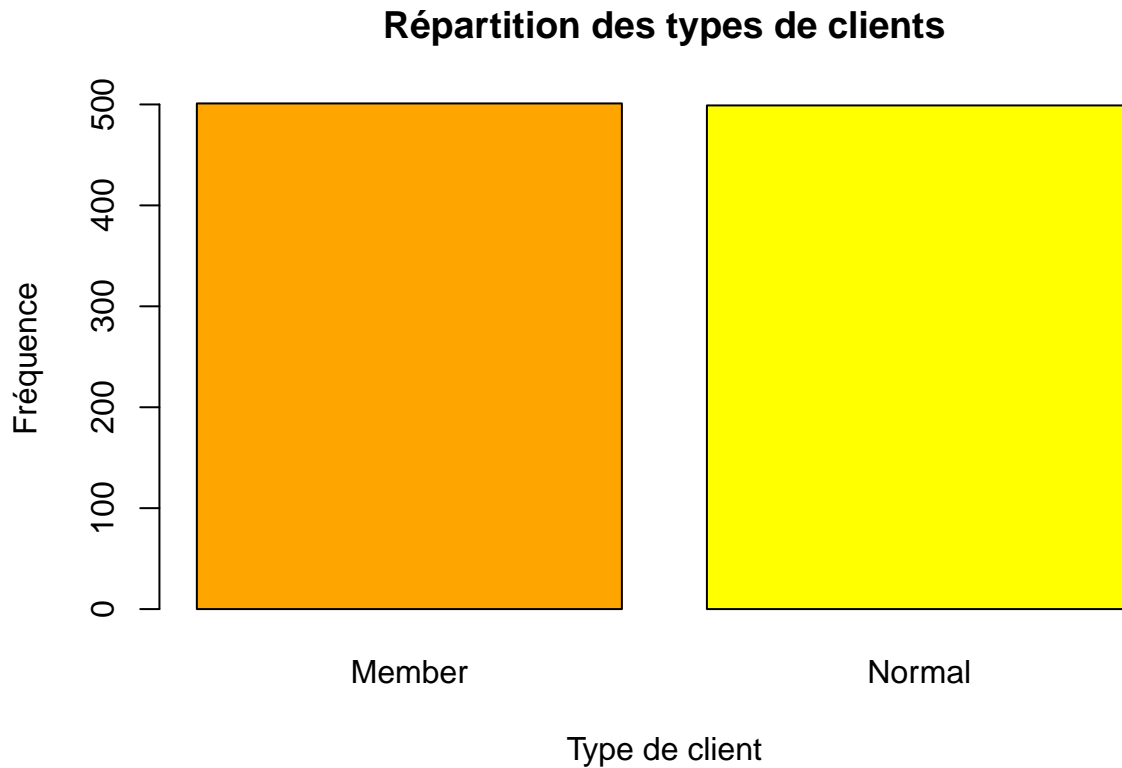
```
barplot(table(basesm$Gender), col = c("blue", "pink"),
        main = "Répartition du genre des clients",
        xlab = "Genre", ylab = "Fréquence")
```



=> Ce graphique présente la répartition du genre des clients dans le supermarché, avec les couleurs bleue pour les clients masculins et rose pour les clients féminins. On peut constater que la différence entre le nombre des deux sexes des clients n'est pas vraiment significative.

La distribution du type de client

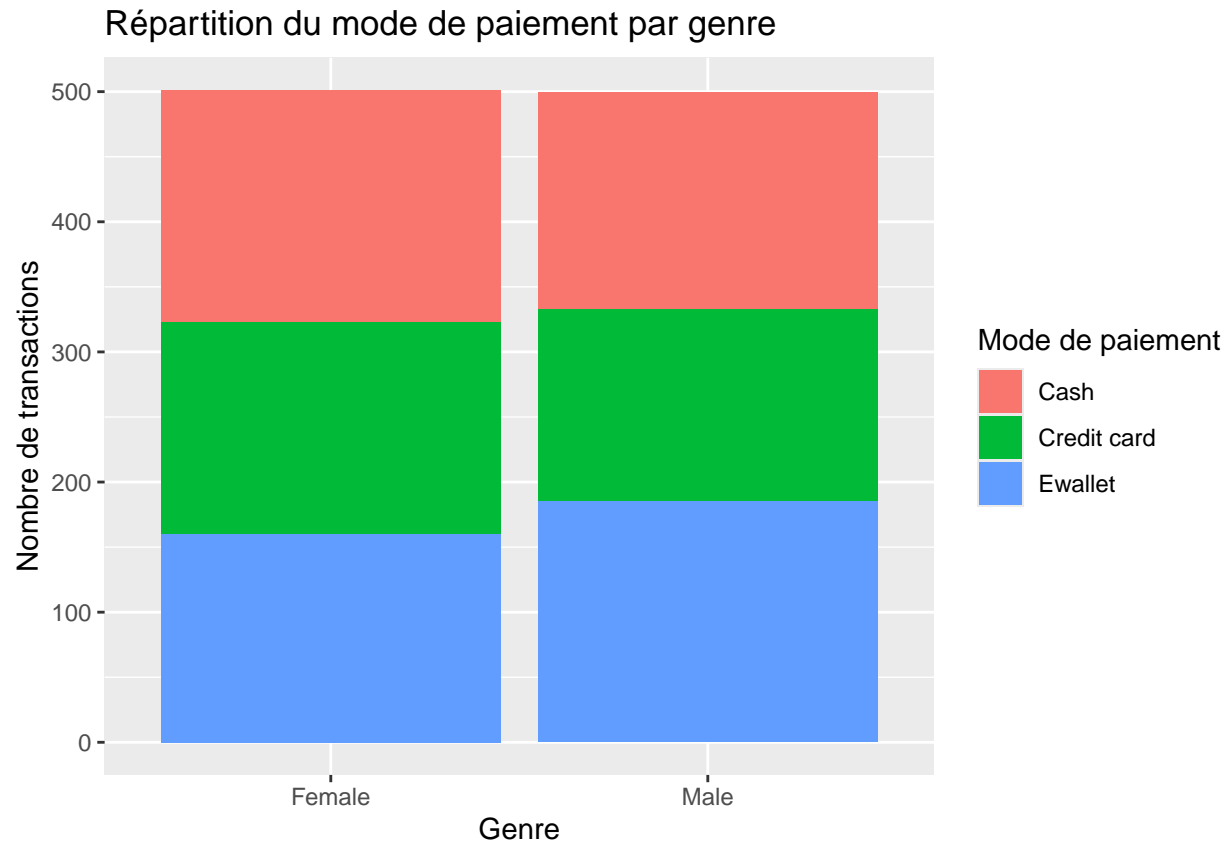
```
barplot(table(basesm$Customer.type), col = c("orange", "yellow", "green"),  
        main = "Répartition des types de clients",  
        xlab = "Type de client", ylab = "Fréquence")
```



=> Ce graphique illustre la répartition des types de clients dans le supermarché, distinguant entre les membres (Member) et les clients non-membres (Normal). On conclut qu'il y a autant des membres que des clients normaux.

La distribution du mode de paiement par genre

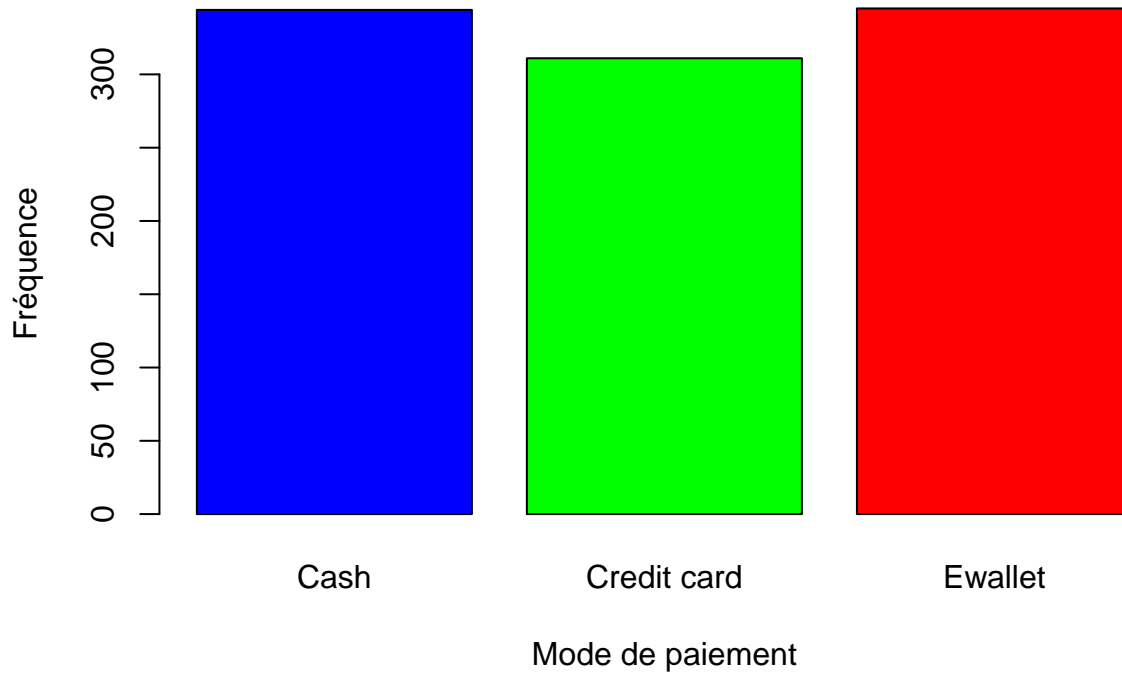
```
ggplot(basesm, aes(x = Gender, fill = Payment)) +  
  geom_bar(position = "stack") +  
  labs(title = "Répartition du mode de paiement par genre",  
        x = "Genre",  
        y = "Nombre de transactions",  
        fill = "Mode de paiement")
```



=> Ce graphique présente la répartition du mode de paiement utilisé par genre de client, empilant les différents modes de paiement pour chaque genre. On peut visualiser que les hommes paient un peu plus que les femmes avec leurs Ewallet pendant que les femmes utilisent plus du cash. ## Les méthodes de paiement utilisées dans le supermarché: # La distribution des modes de paiement.

```
barplot(table(basesm$Payment), col = c("blue", "green", "red", "yellow"),
        main = "Répartition des types de paiements",
        xlab = "Mode de paiement", ylab = "Fréquence")
```

Répartition des types de paiements

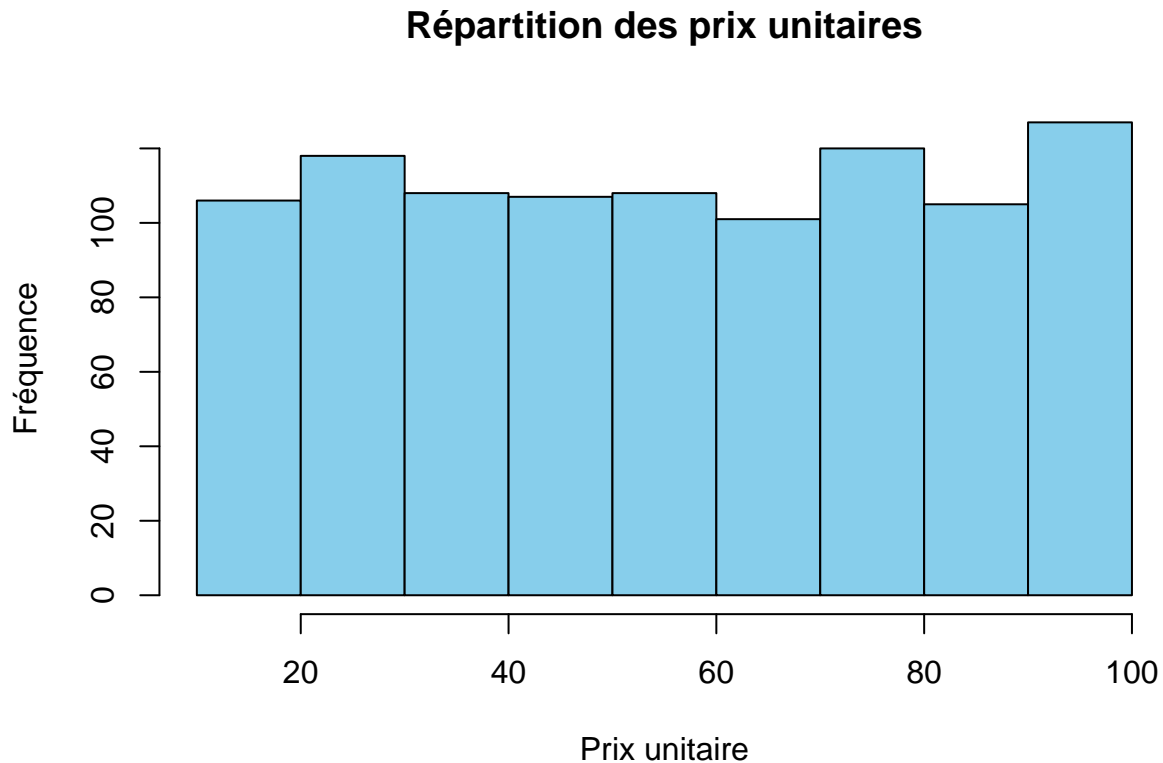


```
fill = "Mode de paiement"
```

=> Ce graphique montre la répartition des différents modes de paiement utilisés dans le supermarché. L'utilisation de la Ewallet et Cash sont les plus courants et la credit card est un peu moins utilisée pour le paiement.

L'histogramme des prix unitaires

```
hist(basesm$Unit.price, col = "skyblue",  
     main = "Répartition des prix unitaires",  
     xlab = "Prix unitaire",  
     ylab = "Fréquence")
```

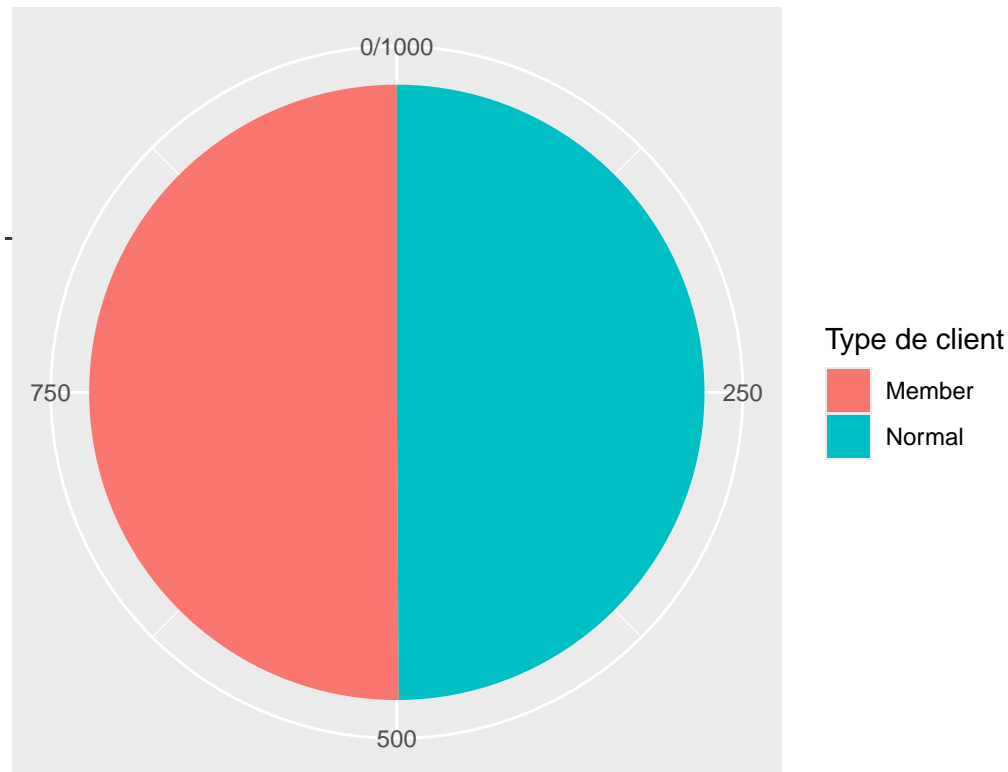



=> Cet histogramme présente la distribution des prix unitaires des produits vendus dans le supermarché. Les Produits les plus vendus sont les produit les plus chères , sinon tout les produits sont vendus d'une fréquence entre 100 et 120.

Plot the distribution of Payment modes by Customer type

```
ggplot(basesm, aes(x = "", fill = `Customer.type`)) +
  geom_bar(width = 1) +
  coord_polar(theta = "y") +
  labs(title = "Répartition du mode de paiement par type de client",
       fill = "Type de client",
       x = NULL,
       y = NULL)
```

Répartition du mode de paiement par type de client

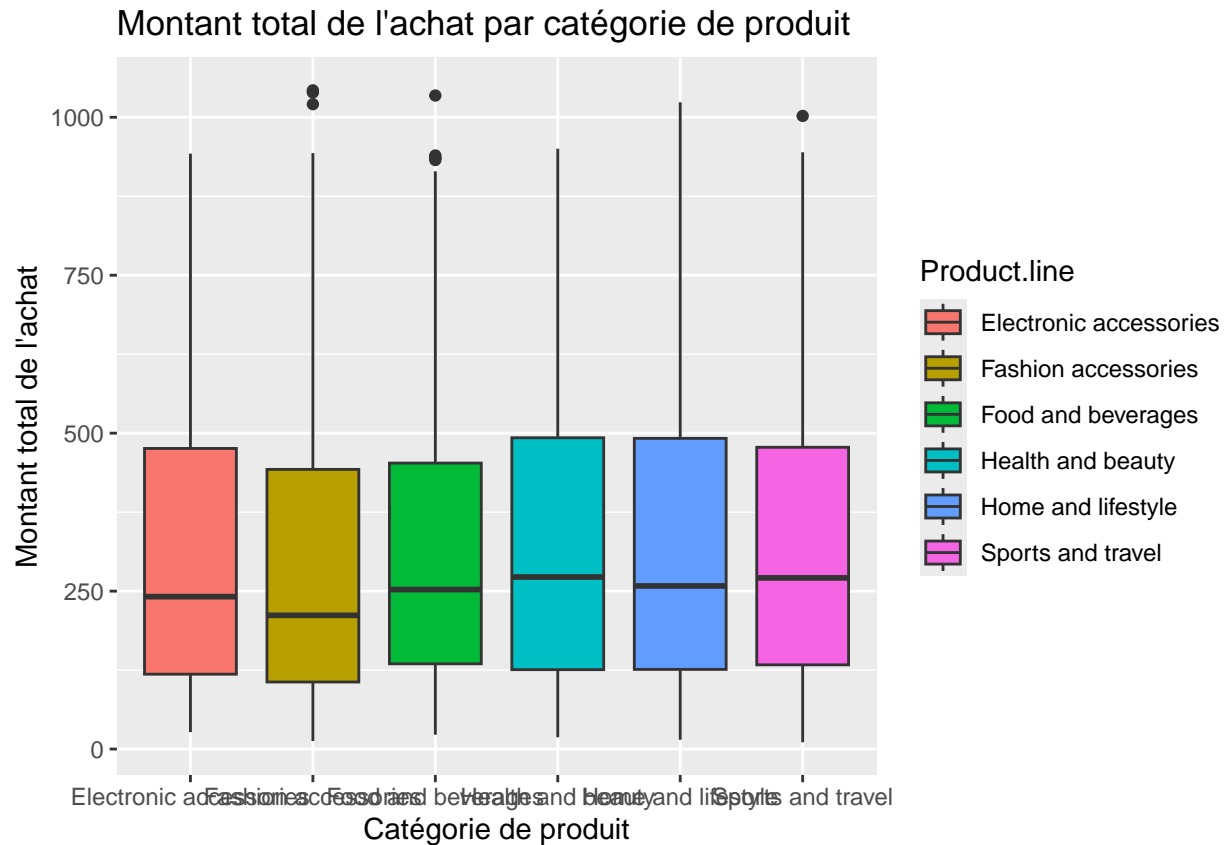


=> Ce graphique montre la répartition du mode de paiement par type de client, en affichant les modes de paiement sous forme de diagramme en secteurs, séparés par type de client. N'importe qu'il soit le type du clients , les méthodes de paiements restent identiques.

La relation entre les produits et les quantités:

Plot the boxplot of Total purchase amount by Product line

```
ggplot(basesm, aes(x = Product.line, y = Total, fill = Product.line)) +  
  geom_boxplot() +  
  labs(title = "Montant total de l'achat par catégorie de produit",  
        x = "Catégorie de produit",  
        y = "Montant total de l'achat")
```

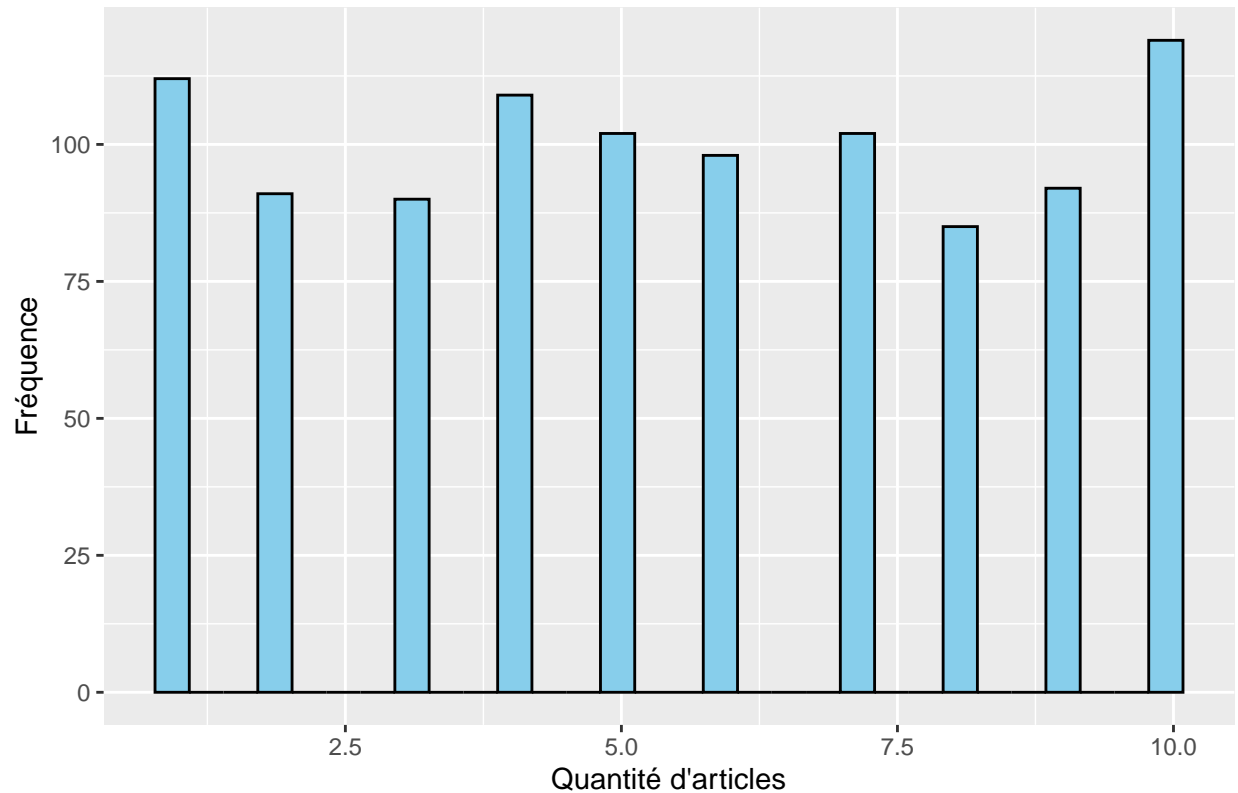


=> Ce graphique présente la répartition du montant total de l'achat par catégorie de produit sous forme de boîtes à moustaches, permettant de visualiser les écarts entre les différentes catégories. Toutes les categories sont ont des montant d'achat proches , mais Home and LifeStyle et Health and Beauty sont les catégories les plus vendus suivi par Electronic accessories.

Plot the histogram of Quantity

```
ggplot(basesm, aes(x = Quantity)) +
  geom_histogram(fill = "skyblue", color = "black", bins = 30) +
  labs(title = "Distribution de la quantité d'articles achetés",
        x = "Quantité d'articles",
        y = "Fréquence")
```

Distribution de la quantité d'articles achetés

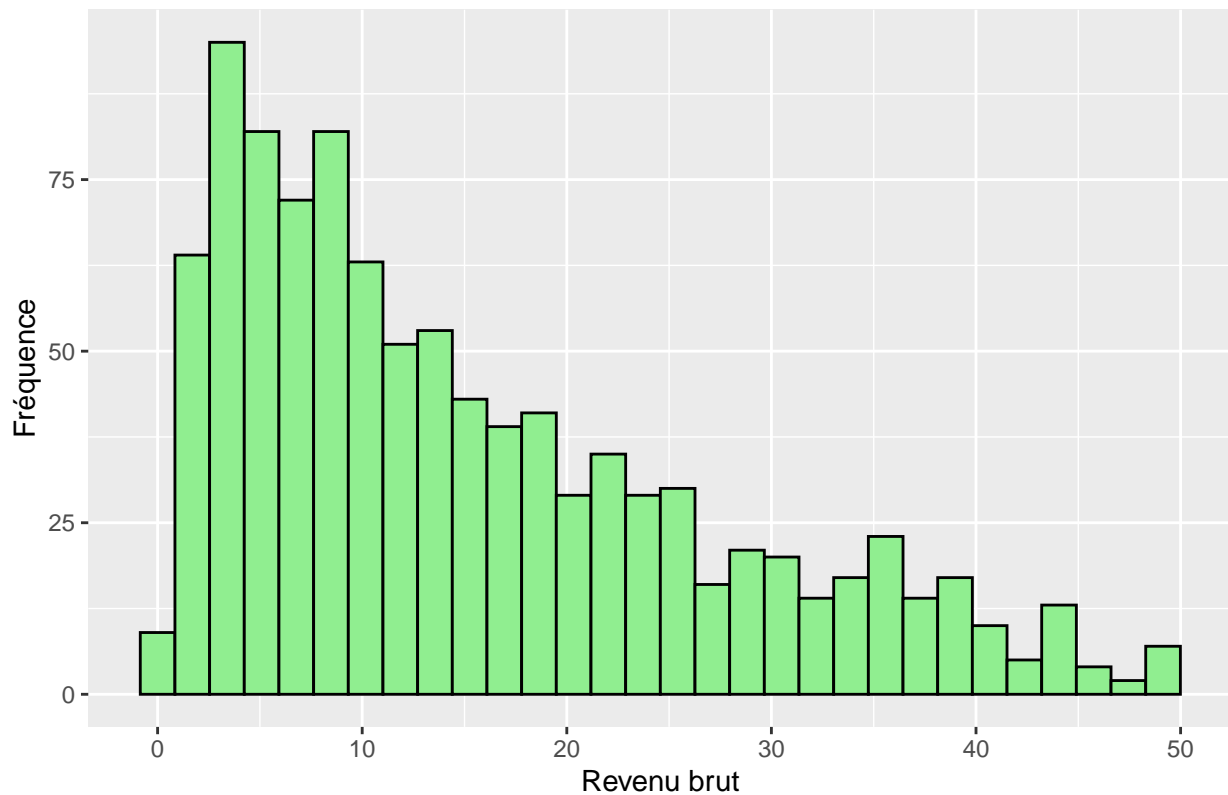


=> Cet histogramme illustre la distribution de la quantité d'articles achetés dans le supermarché. La fréquence des quantité d'articles égale à 10 est la plus supérieur pendant que la la quantité d'article égale à 8 a la plus faible fréquence.

Plot the histogram of Gross income

```
ggplot(basesm, aes(x = gross.income)) +
  geom_histogram(fill = "lightgreen", color = "black", bins = 30) +
  labs(title = "Distribution des revenus bruts",
       x = "Revenu brut",
       y = "Fréquence")
```

Distribution des revenus bruts



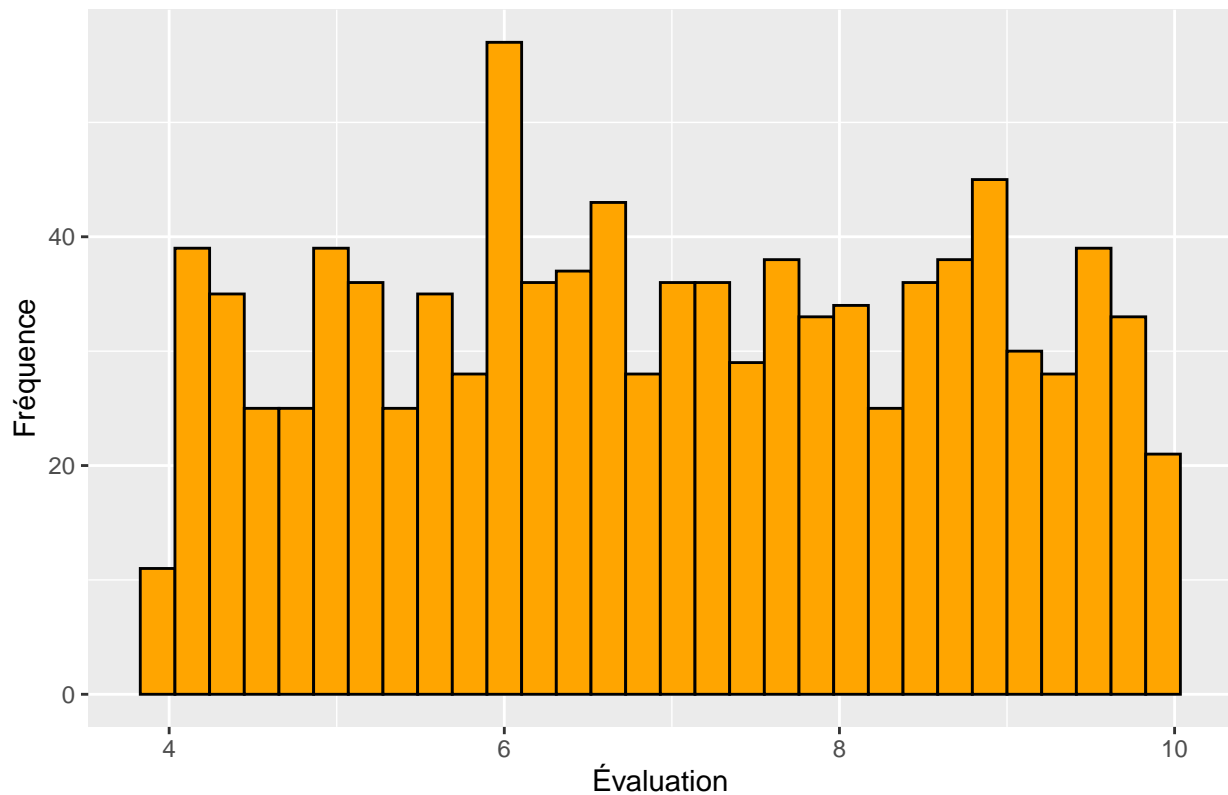
=> Cet histogramme présente la distribution des revenus bruts générés par les ventes de produits dans le supermarché. Plus le revenue brute croit plus la frequence decroit .

Les produits et leurs évaluations:

Plot the histogram of Rating

```
ggplot(basesm, aes(x = Rating)) +
  geom_histogram(fill = "orange", color = "black", bins = 30) +
  labs(title = "Distribution des évaluations des clients",
        x = "Évaluation",
        y = "Fréquence")
```

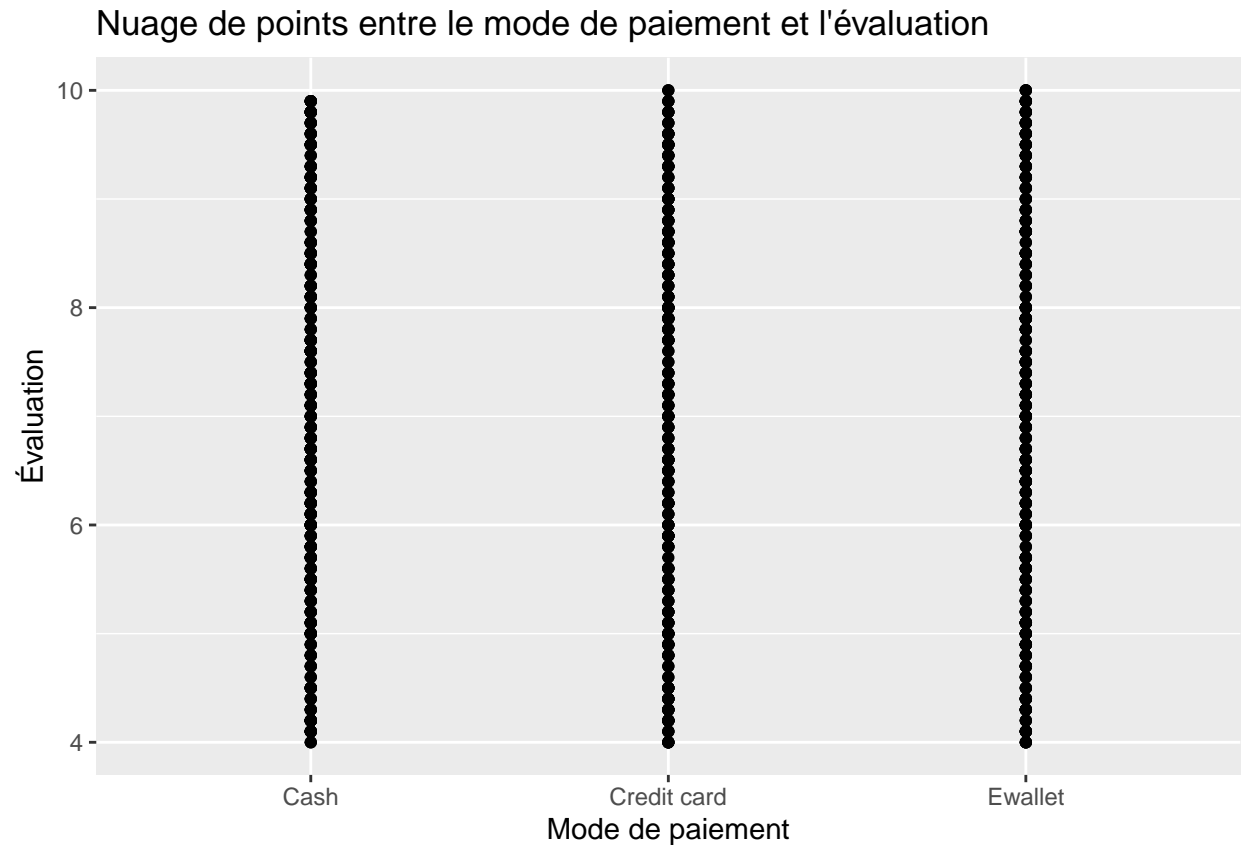
Distribution des évaluations des clients



=> Cet histogramme illustre la distribution des évaluations données par les clients pour leur expérience d'achat dans le supermarché. la note la plus donnée est égale a 6 . c'est un peut rare pour un client qu'il donne la note 4 ou 10.

La dispersion entre le mode de paiement et la notation.

```
ggplot(basesm) +
  geom_point(aes(x = Payment, y = Rating)) +
  labs(title = "Nuage de points entre le mode de paiement et l'évaluation",
        x = "Mode de paiement",
        y = "Évaluation")
```

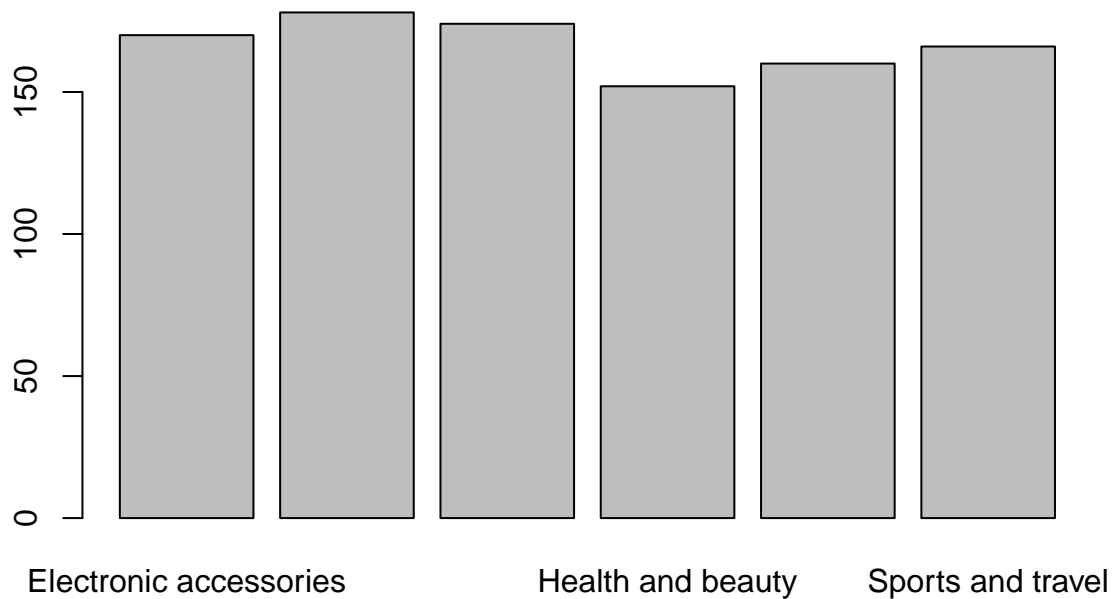


=> Ce nuage de points montre la relation entre le mode de paiement utilisé et l'évaluation donnée par les clients pour leur expérience d'achat. Les clients qui paient avec la credit card ou Ewallet donnent les meilleures évaluations.

Les types de produits et leurs ventes:

La distribution des lignes de produits

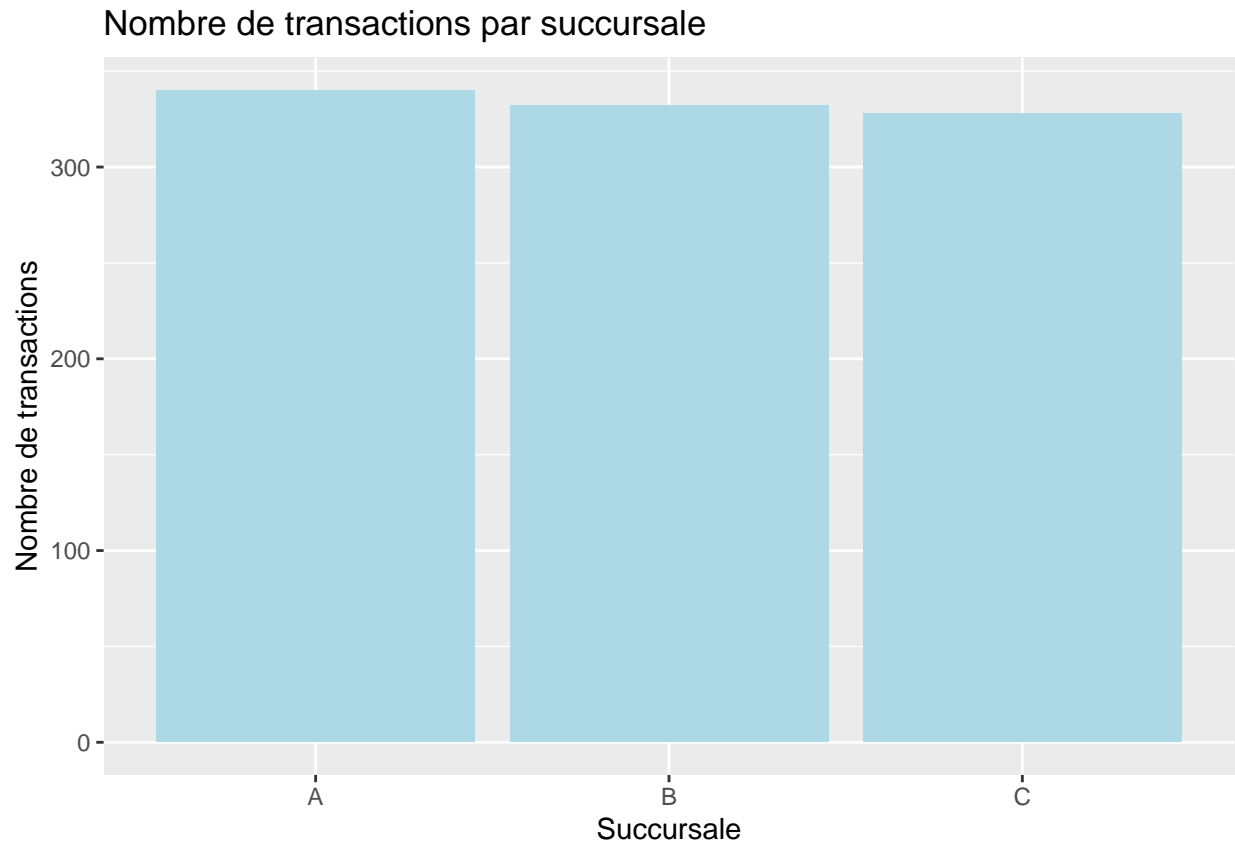
```
barplot(table(basesm$Product.line))
```



=> Ce barplot présente la répartition des différents types de produits vendus dans le supermarché. L'existence des produits du Fashion Accessoires est la plus remarquable pendant qu'il n'y a pas trop de produit de Health and Beauty.

la distribution des transactions par branche.

```
ggplot(basesm, aes(x = Branch)) +
  geom_bar(fill = "lightblue") +
  labs(title = "Nombre de transactions par succursale",
        x = "Succursale",
        y = "Nombre de transactions")
```

=> Ce graphique montre le nombre de transactions effectuées dans chaque succursale du supermarché. La succursale A a le plus nombre de transaction pendant que C a le nombre de transaction le plus inférieure.

Conclusion

En conclusion, l'analyse des ventes pour ce supermarché révèle une croissance constante dans les grandes villes, avec une concurrence accrue sur le marché. Les données historiques sur les ventes dans les trois succursales pendant trois mois fournissent une base solide pour l'application de méthodes d'analyse prédictive, facilitant ainsi la prise de décisions éclairées pour optimiser les performances et rester compétitif sur le marché des supermarchés.