Project Report

Team members

- Ahmed Ashraf Mohamed
  - ID: 2022446758
- Abdelrhman Mohamed Abdelhady Hodid
  - ID : 2022513643
- Antonuose Gerges Nageh
  - ID : 20221903971

Project Description

Getting data set

importing the library we are going to use

```
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
```

```
library(ggplot2)
```

```
## dataPath <- readline("Enter the path to the data set : ")
grc <- as_tibble(read.csv("../dataset/grc.csv",stringsAsFactors = FALSE))
# displaying first 10 rows of our data
print(grc)
```

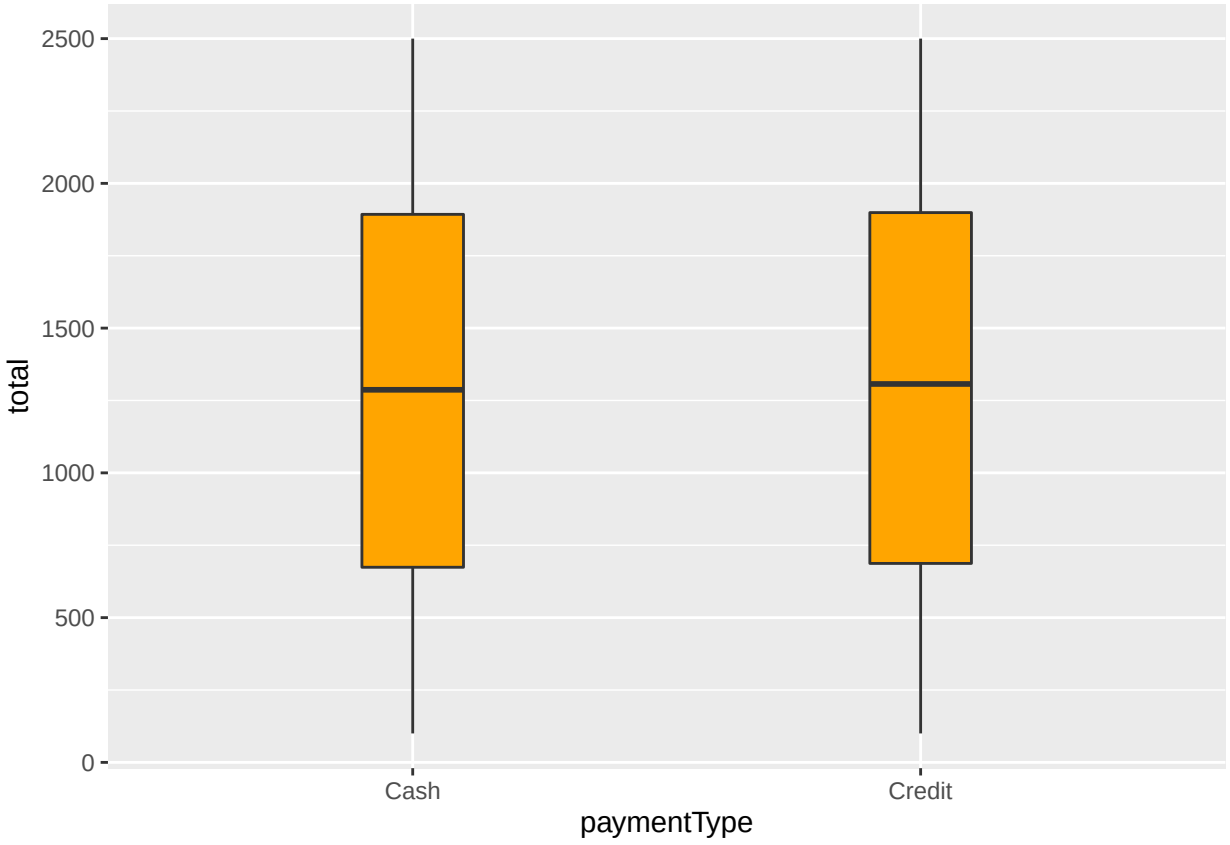
```
## # A tibble: 9,835 x 8
##    items                     count total   rnd customer   age city   paymentType
##    <chr>                     <int> <int> <int> <chr>    <int> <chr>  <chr>
##  1 citrus fruit,semi-finish~     4  1612     9 Maged       60 Hurgh~ Cash
##  2 tropical fruit,yogurt,co~     3   509    12 Eman        23 Aswan  Cash
##  3 whole milk                    1  2084     8 Rania       37 Dakah~ Cash
##  4 pip fruit,yogurt,cream c~     4   788     8 Rania       37 Dakah~ Cash
##  5 other vegetables,whole m~     4  1182    14 Magdy       36 Sohag  Cash
##  6 whole milk,butter,yogurt~     5  1771     3 Ahmed       30 Giza   Credit
##  7 rolls/buns                    1  2196     7 Huda        39 Gharb~ Cash
##  8 other vegetables,UHT-mil~     5  1657     6 Walaa       29 Cairo  Cash
##  9 pot plants                    1  2373     2 Mohamed     25 Alexa~ Credit
```

```
## 10 whole milk,cereals            2   343     5 Shimaa      55 Port ~ Cash
## # ... with 9,825 more rows
```

Visualizing our Data

Comparison between cash and creadit total spending using box plot

```
ggplot(
  grc,
  aes(x = paymentType, y = total)
  ) +
  geom_boxplot(width = .2,
               fill = "orange",
               outlier.color = "orange",
               outlier.size = 2)
```



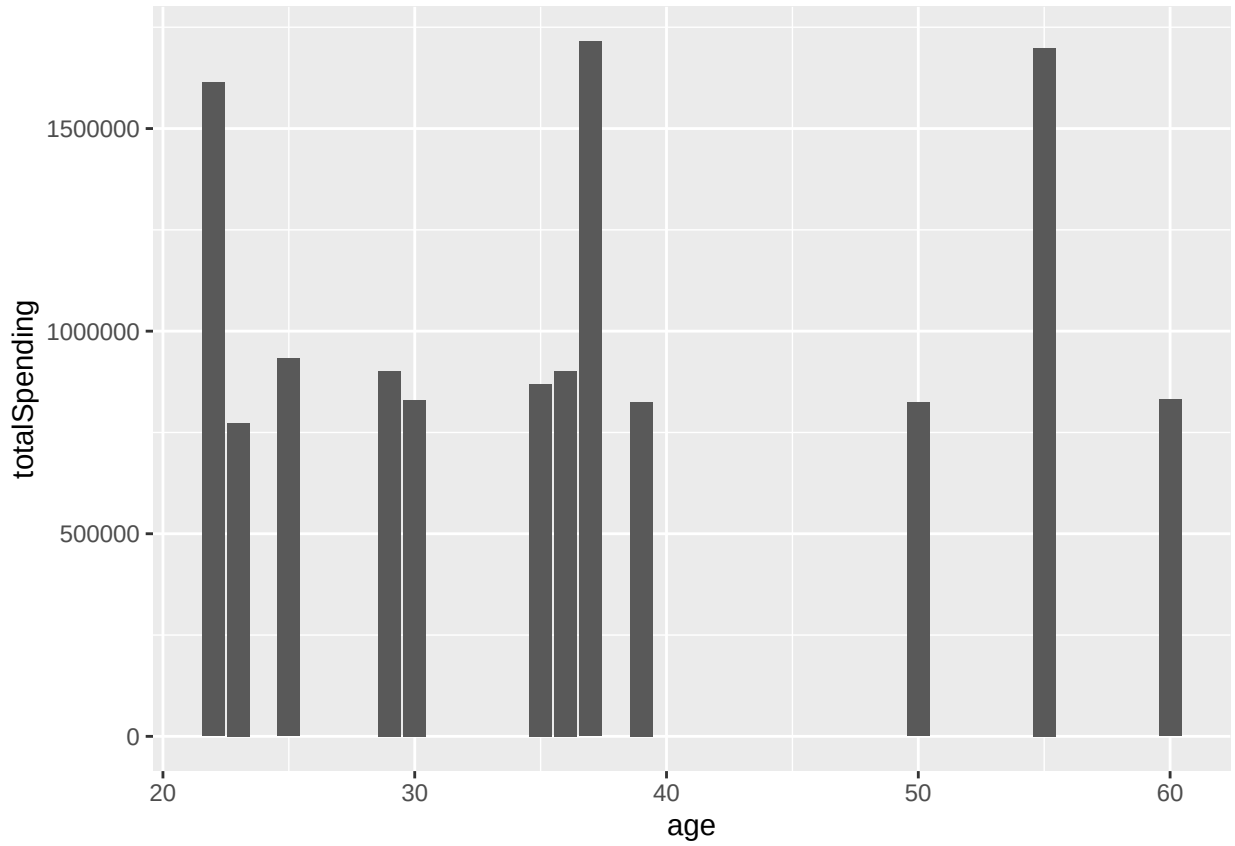
Compare each age and sum of total spending.

```
grc_age <- select(grc,age,total)
grc_age <- group_by(grc_age, age)
grc_age <- summarise(grc_age,totalSpending = sum(total))
print(grc_age)
```

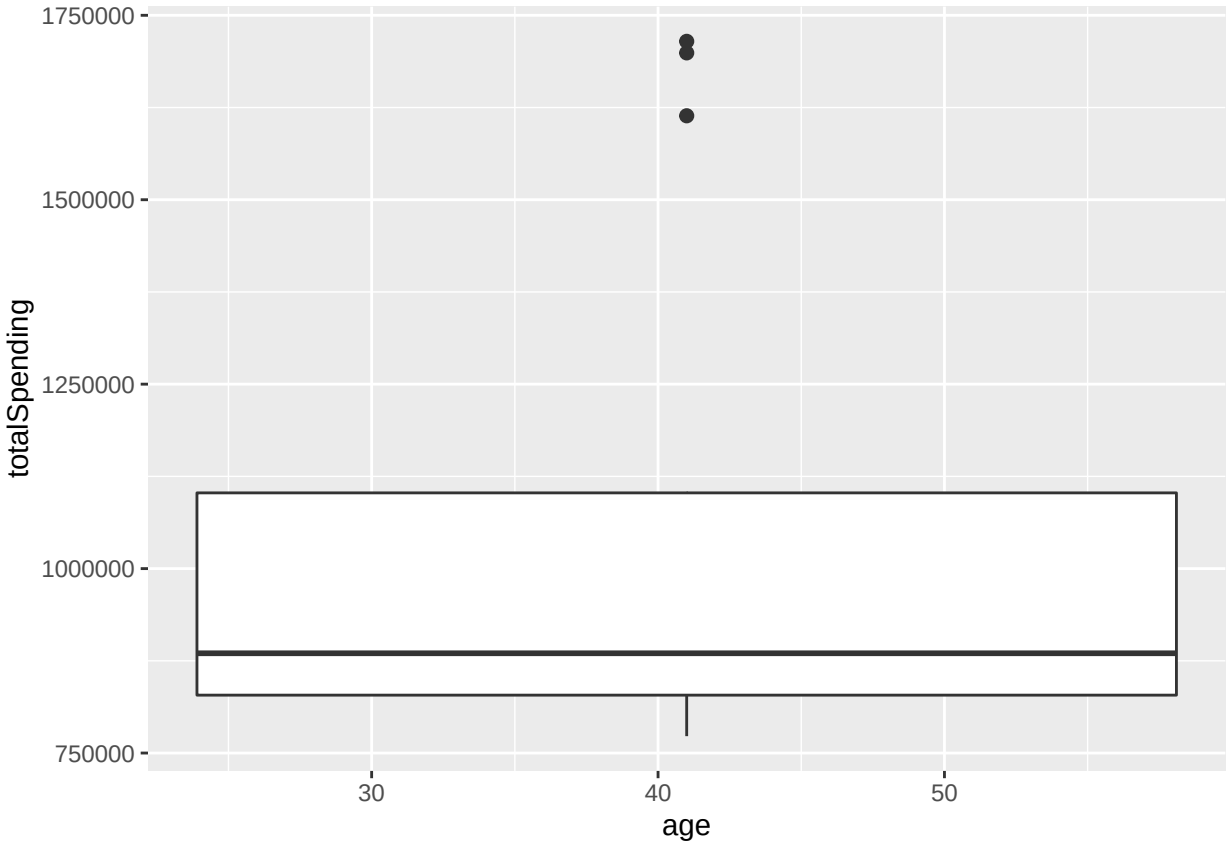
```
## # A tibble: 12 x 2
##      age totalSpending
##    <int>         <int>
##  1    22       1613801
```

```
##  2     23         772871
##  3     25         932250
##  4     29         900797
##  5     30         829587
##  6     35         869668
##  7     36         901010
##  8     37        1714689
##  9     39         825147
## 10     50         824064
## 11     55        1699068
## 12     60         831272
```

```
ggplot(grc_age,aes(x = age, y = totalSpending)) + geom_col()
```




```
ggplot(grc_age, aes(x= age, y = totalSpending, group =1)) + geom_boxplot(outlier.size = 2)
```


1750000
1500000
1250000
1000000
750000
totalSpending
30
40
50
age