Deliverable 4

Final deliverable

Júlia Gasull i Claudia Sánchez

January 2, 2021

Contents

1 First setups

```
if(!is.null(dev.list())) dev.off()  # Clear plots
rm(list=ls())                       # Clean workspace
```

1.1 Some useful functions

```
calcQ <- function(x) { # Function to calculate the different quartiles
  s.x <- summary(x)
  iqr<-s.x[5]-s.x[2]
  list(souti=s.x[2]-3*iqr, mouti=s.x[2]-1.5*iqr, min=s.x[1], q1=s.x[2], q2=s.x[3],
       q3=s.x[5], max=s.x[6], mouts=s.x[5]+1.5*iqr, souts=s.x[5]+3*iqr )
}
```

```r
countNA <- function(x) { # Function to count the NA values
  mis_x <- NULL
  for (j in 1:ncol(x)) {mis_x[j] <- sum(is.na(x[,j])) }
  mis_x <- as.data.frame(mis_x)
  rownames(mis_x) <- names(x)
  mis_i <- rep(0,nrow(x))
  for (j in 1:ncol(x)) {mis_i <- mis_i + as.numeric(is.na(x[,j])) }
  list(mis_col=mis_x,mis_ind=mis_i)
}

countX <- function(x,X) { # Function to count a specific number of appearences
  n_x <- NULL
  for (j in 1:ncol(x)) {n_x[j] <- sum(x[,j]==X) }
  n_x <- as.data.frame(n_x)
  rownames(n_x) <- names(x)
  nx_i <- rep(0,nrow(x))
  for (j in 1:ncol(x)) {nx_i <- nx_i + as.numeric(x[,j]==X) }
  list(nx_col=n_x,nx_ind=nx_i)
}
```

2 Data description

- Description http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml
- Data Dictionary - SHL Trip Records -This data dictionary describes SHL trip data in visit http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

2.1 Variables

- VendorID
  - A code indicating the LPEP provider that provided the record.
  - Values:
    - 1= Creative Mobile Technologies, LLC
    - 2= VeriFone Inc.
- lpep_pickup_datetime
  - The date and time when the meter was engaged.
- lpep_dropoff_datetime
  - The date and time when the meter was disengaged.
- Passenger_count
  - The number of passengers in the vehicle.
  - This is a driver-entered value.
- Trip_distance
  - The elapsed trip distance in miles reported by the taximeter.
- Pickup_longitude
  - Longitude where the meter was engaged.
- Pickup_latitude
  - Latitude where the meter was engaged.
- RateCodeID
  - The final rate code in effect at the end of the trip.
  - Values:
    - 1=Standard rate
    - 2=JFK
    - 3=Newark
    - 4=Nassau or Westchester
    - 5=Negotiated fare

        * 6=Group ride

- Store_and_fwd_flag
  - This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka "store and forward," because the vehicle did not have a connection to the server:
  - Values
    * Y= store and forward trip

    * N= not a store and forward trip

- Dropoff_longitude
  - Longitude where the meter was timed off.

- Dropoff_latitude
  - Latitude where the meter was timed off.

- Payment_type
  - A numeric code signifying how the passenger paid for the trip.
  - Values:
    * 1= Credit card
    * 2= Cash
    * 3= No charge
    * 4= Dispute
- Fare_amount
  - The time-and-distance fare calculated by the meter.

- Extra
  - Miscellaneous extras and surcharges.

  - Currently, this only includes the $0.50 and $1 rush hour and overnight charges.
- MTA_tax
  - $0.50 MTA tax that is automatically triggered based on the metered rate in use.

- Improvement_surcharge
  - $0.30 improvement surcharge assessed on hailed trips at the flag drop.
  - The improvement surcharge began being levied in 2015.

- Tip_amount
  - This field is automatically populated for credit card tips.
  - Cash tips are not included.

- Tolls_amount
  - Total amount of all tolls paid in trip.

- Total_amount
  - The total amount charged to passengers.
  - Does not include cash tips.

- Trip_type
  - A code indicating whether the trip was a street-hail or a dispatch that is automatically assigned based on the metered rate in use but can be altered by the driver.
  - Values:
    * 1= Street-hail
    * 2= Dispatch

3 Load Required Packages for this deliverable

We load the necessary packages and set working directory

```
setwd("~/Documents/uni/FIB-ADEI-LAB/deliverable4")
filepath<-"~/Documents/uni/FIB-ADEI-LAB/deliverable4"

# Load Required Packages
```

```r
options(contrasts=c("contr.treatment","contr.treatment"))
requiredPackages <- c("missMDA","chemometrics","mvoutlier","effects","FactoMineR","car","lmtest","ggplot
missingPackages <- requiredPackages[!(requiredPackages %in% installed.packages()[,"Package"])]
if(length(missingPackages)) install.packages(missingPackages)
lapply(requiredPackages, require, character.only = TRUE)
```

4 Select a sample of 5000 records

From the proposed database, we need to select a sample of 5000 records randomly so we can start analyzing our data.

!!!!! PER DESCOMENTAR AL FINAL

```r
#df<-read.table(paste0(filepath,"/green_tripdata_2016-01.csv"),header=T, sep=",")
#set.seed(180998)
#sam<-as.vector(sort(sample(1:nrow(df),5000)))
#df<-df[sam,]
```

!!! ESBORRAR AL DFINAL

```r
load(paste0(filepath,"/Taxi5000_raw.RData"))
```

5 Rename variables and clean data

```r
summary(df)
```

```
##     VendorID     lpep_pickup_datetime Lpep_dropoff_datetime Store_and_fwd_flag
##  Min.   :1.000   Length:5000          Length:5000           Length:5000
##  1st Qu.:2.000   Class :character     Class :character      Class :character
##  Median :2.000   Mode  :character     Mode  :character      Mode  :character
##  Mean   :1.788
##  3rd Qu.:2.000
##  Max.   :2.000
##    RateCodeID  Pickup_longitude Pickup_latitude Dropoff_longitude
##  Min.   :1.0   Min.   :-75.39   Min.   : 0.00   Min.   :-75.31
##  1st Qu.:1.0   1st Qu.:-73.96   1st Qu.:40.70   1st Qu.:-73.97
##  Median :1.0   Median :-73.95   Median :40.75   Median :-73.94
##  Mean   :1.1   Mean   :-73.89   Mean   :40.72   Mean   :-73.80
##  3rd Qu.:1.0   3rd Qu.:-73.92   3rd Qu.:40.80   3rd Qu.:-73.91
##  Max.   :5.0   Max.   :  0.00   Max.   :41.04   Max.   :  0.00
##  Dropoff_latitude Passenger_count Trip_distance     Fare_amount
##  Min.   : 0.00    Min.   :0.000   Min.   : 0.000   Min.   :-52.0
##  1st Qu.:40.70    1st Qu.:1.000   1st Qu.: 1.020   1st Qu.:  6.0
##  Median :40.75    Median :1.000   Median : 1.800   Median :  9.0
##  Mean   :40.67    Mean   :1.375   Mean   : 2.765   Mean   : 11.9
##  3rd Qu.:40.79    3rd Qu.:1.000   3rd Qu.: 3.420   3rd Qu.: 14.5
##  Max.   :41.18    Max.   :6.000   Max.   :52.790   Max.   :200.0
##      Extra            MTA_tax          Tip_amount      Tolls_amount
##  Min.   :-1.0000   Min.   :-0.5000   Min.   : 0.000   Min.   : 0.00000
##  1st Qu.: 0.0000   1st Qu.: 0.5000   1st Qu.: 0.000   1st Qu.: 0.00000
##  Median : 0.5000   Median : 0.5000   Median : 0.000   Median : 0.00000
##  Mean   : 0.3517   Mean   : 0.4857   Mean   : 1.217   Mean   : 0.08369
##  3rd Qu.: 0.5000   3rd Qu.: 0.5000   3rd Qu.: 2.000   3rd Qu.: 0.00000
##  Max.   : 1.0000   Max.   : 0.5000   Max.   :96.000   Max.   :18.04000
##  Ehail_fee      improvement_surcharge  Total_amount     Payment_type
##  Mode:logical   Min.   :-0.3000       Min.   :-52.80   Min.   :1.00
##  NA's:5000      1st Qu.: 0.3000       1st Qu.:  7.80   1st Qu.:1.00
##                 Median : 0.3000       Median : 11.16   Median :2.00
##                 Mean   : 0.2914       Mean   : 14.33   Mean   :1.52
##                 3rd Qu.: 0.3000       3rd Qu.: 17.16   3rd Qu.:2.00
##                 Max.   : 0.3000       Max.   :260.00   Max.   :4.00
##    Trip_type
##  Min.   :1.000
```

```
##  1st Qu.:1.000
##  Median :1.000
##  Mean   :1.023
##  3rd Qu.:1.000
##  Max.   :2.000
```

```
names(df)[names(df) == "VendorID"] <- "q.vendor_id"
names(df)[names(df) == "lpep_pickup_datetime"] <- "qual.lpep_pickup_datetime"
names(df)[names(df) == "Lpep_dropoff_datetime"] <- "qual.lpep_dropoff_datetime"
names(df)[names(df) == "Store_and_fwd_flag"] <- "qual.store_and_fwd_flag"
names(df)[names(df) == "RateCodeID"] <- "q.rate_code_id"
names(df)[names(df) == "Pickup_longitude"] <- "q.pickup_longitude"
names(df)[names(df) == "Pickup_latitude"] <- "q.pickup_latitude"
names(df)[names(df) == "Dropoff_longitude"] <- "q.dropoff_longitude"
names(df)[names(df) == "Dropoff_latitude"] <- "q.dropoff_latitude"
names(df)[names(df) == "Passenger_count"] <- "q.passenger_count"
names(df)[names(df) == "Trip_distance"] <- "q.trip_distance"
names(df)[names(df) == "Fare_amount"] <- "q.fare_amount"
names(df)[names(df) == "Extra"] <- "q.extra"
names(df)[names(df) == "MTA_tax"] <- "q.mta_tax"
names(df)[names(df) == "Tip_amount"] <- "q.tip_amount"
names(df)[names(df) == "Tolls_amount"] <- "q.tolls_amount"
df$Ehail_fee <- NULL # deleting it --> only NA's
names(df)[names(df) == "improvement_surcharge"] <- "q.improvement_surcharge"
names(df)[names(df) == "Total_amount"] <- "q.target.total_amount"
names(df)[names(df) == "Payment_type"] <- "q.payment_type"
names(df)[names(df) == "Trip_type"] <- "q.trip_type"
summary(df); names(df)
```

```
##   q.vendor_id    qual.lpep_pickup_datetime qual.lpep_dropoff_datetime
##  Min.   :1.000   Length:5000               Length:5000
##  1st Qu.:2.000   Class :character          Class :character
##  Median :2.000   Mode  :character          Mode  :character
##  Mean   :1.788
##  3rd Qu.:2.000
##  Max.   :2.000
##  qual.store_and_fwd_flag q.rate_code_id q.pickup_longitude q.pickup_latitude
##  Length:5000             Min.   :1.0    Min.   :-75.39     Min.   : 0.00
##  Class :character        1st Qu.:1.0    1st Qu.:-73.96     1st Qu.:40.70
##  Mode  :character        Median :1.0    Median :-73.95     Median :40.75
##                          Mean   :1.1    Mean   :-73.89     Mean   :40.72
##                          3rd Qu.:1.0    3rd Qu.:-73.92     3rd Qu.:40.80
##                          Max.   :5.0    Max.   :  0.00     Max.   :41.04
##  q.dropoff_longitude q.dropoff_latitude q.passenger_count q.trip_distance
##  Min.   :-75.31      Min.   : 0.00      Min.   :0.000     Min.   : 0.000
##  1st Qu.:-73.97      1st Qu.:40.70      1st Qu.:1.000     1st Qu.: 1.020
##  Median :-73.94      Median :40.75      Median :1.000     Median : 1.800
##  Mean   :-73.80      Mean   :40.67      Mean   :1.375     Mean   : 2.765
##  3rd Qu.:-73.91      3rd Qu.:40.79      3rd Qu.:1.000     3rd Qu.: 3.420
##  Max.   :  0.00      Max.   :41.18      Max.   :6.000     Max.   :52.790
##  q.fare_amount      q.extra          q.mta_tax        q.tip_amount
##  Min.   :-52.0   Min.   :-1.0000   Min.   :-0.5000   Min.   : 0.000
##  1st Qu.:  6.0   1st Qu.: 0.0000   1st Qu.: 0.5000   1st Qu.: 0.000
##  Median :  9.0   Median : 0.5000   Median : 0.5000   Median : 0.000
##  Mean   : 11.9   Mean   : 0.3517   Mean   : 0.4857   Mean   : 1.217
##  3rd Qu.: 14.5   3rd Qu.: 0.5000   3rd Qu.: 0.5000   3rd Qu.: 2.000
##  Max.   :200.0   Max.   : 1.0000   Max.   : 0.5000   Max.   :96.000
##  q.tolls_amount     q.improvement_surcharge q.target.total_amount
##  Min.   : 0.00000   Min.   :-0.3000         Min.   :-52.80
##  1st Qu.: 0.00000   1st Qu.: 0.3000         1st Qu.:  7.80
##  Median : 0.00000   Median : 0.3000         Median : 11.16
##  Mean   : 0.08369   Mean   : 0.2914         Mean   : 14.33
##  3rd Qu.: 0.00000   3rd Qu.: 0.3000         3rd Qu.: 17.16
##  Max.   :18.04000   Max.   : 0.3000         Max.   :260.00
```

```
##  q.payment_type  q.trip_type
##  Min.   :1.00   Min.   :1.000
##  1st Qu.:1.00   1st Qu.:1.000
##  Median :2.00   Median :1.000
##  Mean   :1.52   Mean   :1.023
##  3rd Qu.:2.00   3rd Qu.:1.000
##  Max.   :4.00   Max.   :2.000
```

```
##  [1] "q.vendor_id"                "qual.lpep_pickup_datetime"
##  [3] "qual.lpep_dropoff_datetime" "qual.store_and_fwd_flag"
##  [5] "q.rate_code_id"             "q.pickup_longitude"
##  [7] "q.pickup_latitude"          "q.dropoff_longitude"
##  [9] "q.dropoff_latitude"         "q.passenger_count"
## [11] "q.trip_distance"            "q.fare_amount"
## [13] "q.extra"                    "q.mta_tax"
## [15] "q.tip_amount"               "q.tolls_amount"
## [17] "q.improvement_surcharge"    "q.target.total_amount"
## [19] "q.payment_type"             "q.trip_type"
```

6 Creating factors

6.1 Vendor ID

This variable expresses the Creative Mobile Technologies, LLC as 1 and Verifone Inc as 2, so we create a factor to make it more readable. With the initial summary we see that this variable does not have any missing value, so we proceed to factor it.

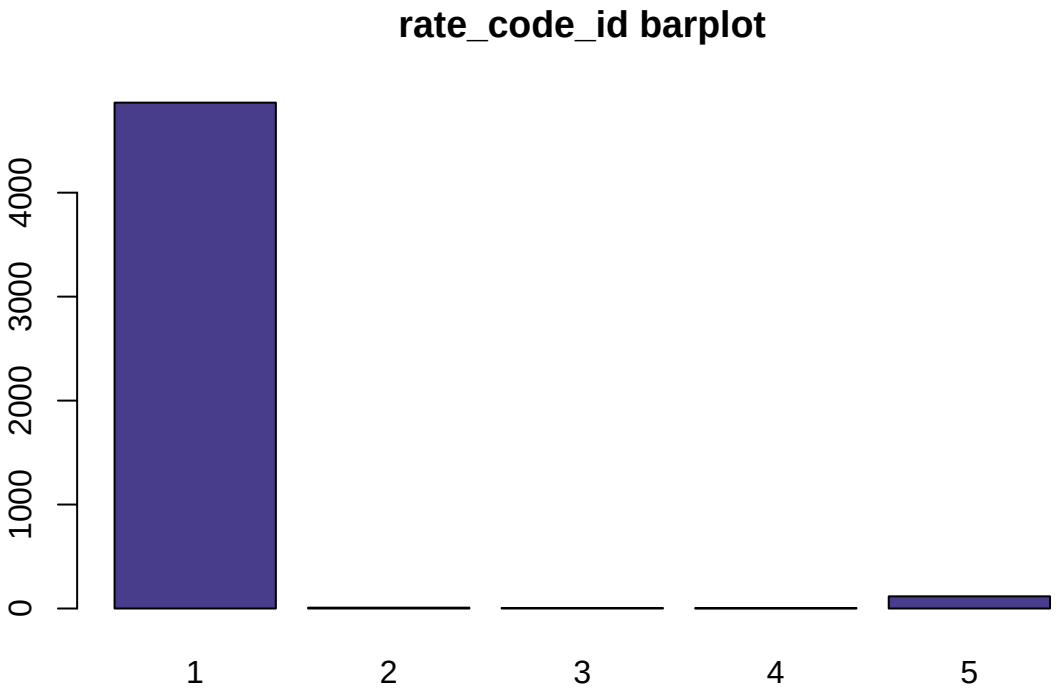
```
names(df)[names(df) == "q.vendor_id"] <- "f.vendor_id"
df$f.vendor_id<-factor(df$f.vendor_id,labels=c("mobile","verifone"))
levels(df$f.vendor_id)<-paste0("vendor_",levels(df$f.vendor_id))
summary(df$f.vendor_id)
```

```
##   vendor_mobile vendor_verifone
##            1062            3938
```

6.2 Rate Code ID

This variable expresses the different RateCodeIDs that we can have as numerical values, so we need to categorize them in order to be able to work with them.

```
names(df)[names(df) == "q.rate_code_id"] <- "f.rate_code_id"
df$f.rate_code_id<-factor(df$f.rate_code_id)
barplot(summary(df$f.rate_code_id),main="rate_code_id barplot",col="darkslateblue")
```

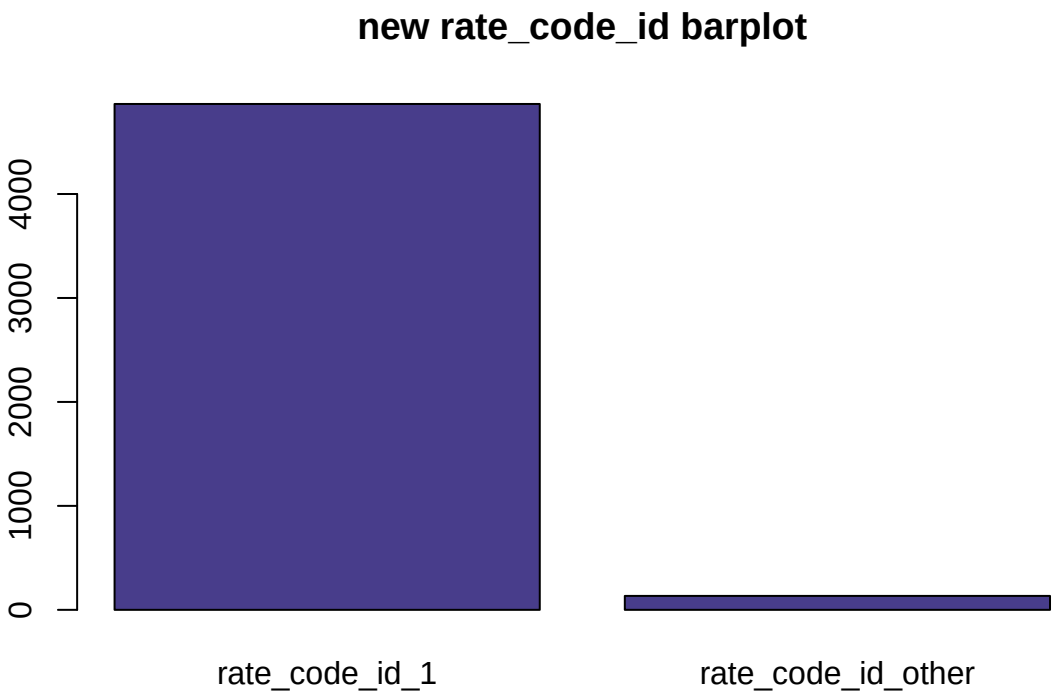


We see that most samples are in rate_code_id=1, which is what we are interested in. Therefore, we factorize and create only two groups, the one with rate_code_id=1 and the rest.

```
df$f.rate_code_id[df$f.rate_code_id != 1] = 2
summary(df$f.rate_code_id)
```

```
##    1    2    3    4    5
## 4866  134    0    0    0
```

```
df$f.rate_code_id <- factor(df$f.rate_code_id, labels=c("rate_code_id_1","rate_code_id_other"))
barplot(summary(df$f.rate_code_id),main="new rate_code_id barplot",col="darkslateblue")
```



Now is more balanced.

6.3 Store and fwd flag

This is a categorical variable with the values Y and N, so we need to factor it.

```
summary(df$qual.store_and_fwd_flag)
```

```
##    Length     Class      Mode
##      5000 character character
```

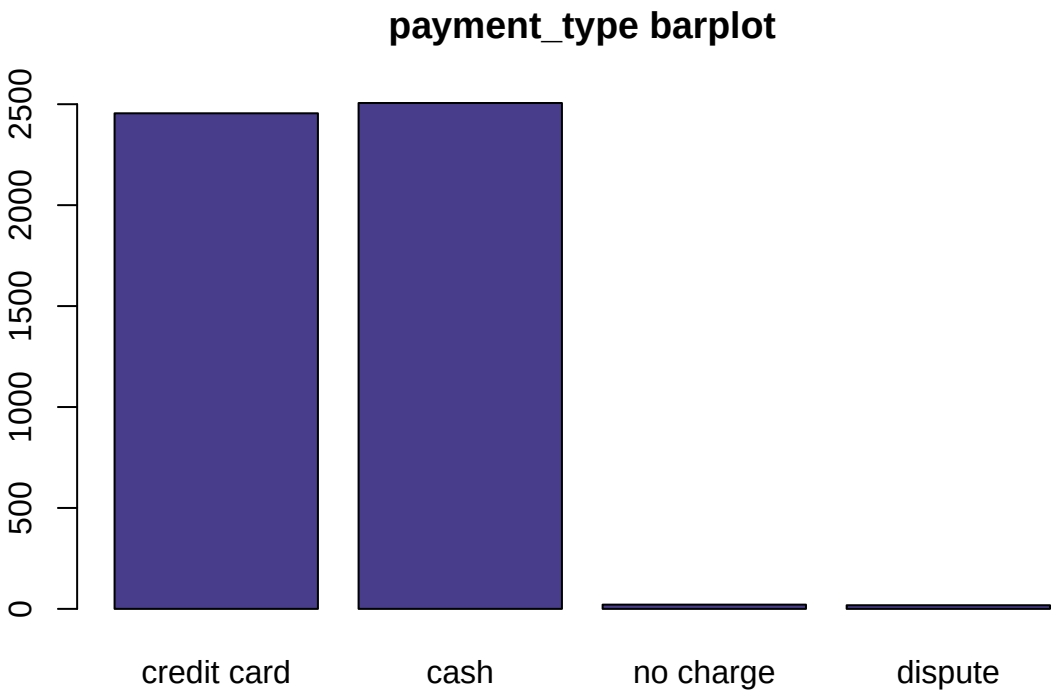
```
names(df)[names(df) == "qual.store_and_fwd_flag"] <- "f.store_and_fwd_flag"
df$f.store_and_fwd_flag<-factor(df$f.store_and_fwd_flag, labels=c("no","yes"))
summary(df$f.store_and_fwd_flag)
```

```
##   no  yes
## 4982   18
```

6.4 Payment type

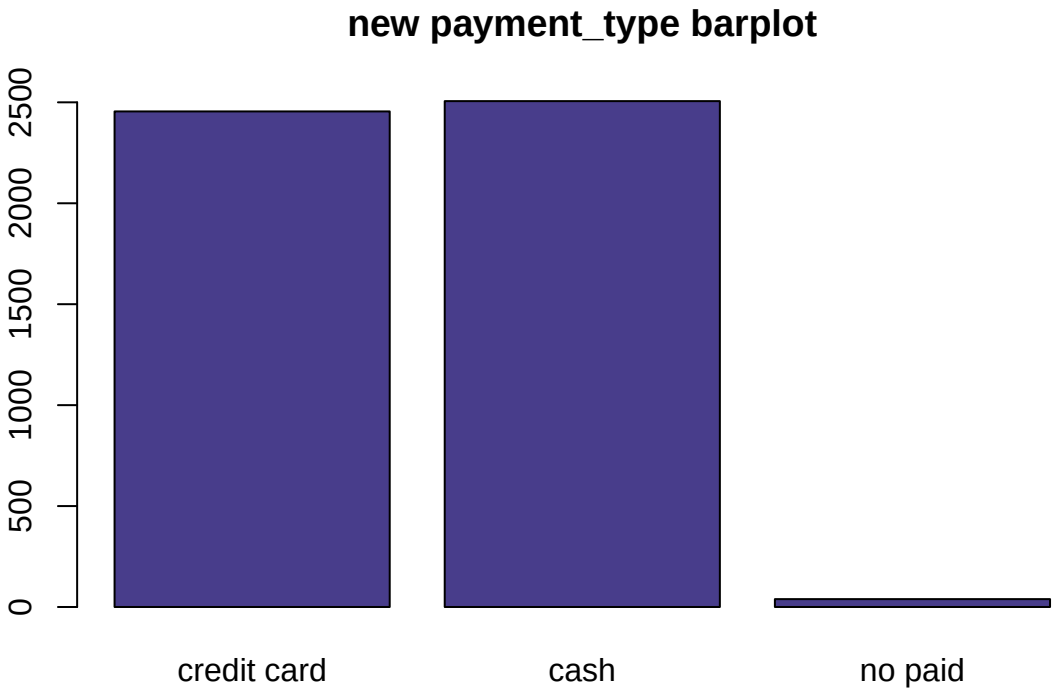
This variable is categorical but it is expressed as numerical, so we need to factor it in order to be able to work with it.

```
names(df)[names(df) == "q.payment_type"] <- "f.payment_type"
df$f.payment_type<-factor(df$f.payment_type,labels=c("credit card","cash","no charge","dispute"))
barplot(summary(df$f.payment_type),main="payment_type barplot",col="darkslateblue")
```



As we can see, there are few values with "No charge" or "Dispute" category, so we decided to categorize it into a new category ("No paid").

```
levels(df$f.payment_type) <- c("credit card","cash","no paid","no paid")
barplot(summary(df$f.payment_type),main="new payment_type barplot",col="darkslateblue")
```



Now is more balanced.

6.5 Trip_type

This variable is categorical but it is expressed as numerical, so we need to factor it in order to be able to work with it.

```
names(df)[names(df) == "q.trip_type"] <- "f.trip_type"
df$f.trip_type<-factor(df$f.trip_type,labels=c("street_hail","dispatch"))
```

7 Clean some variables

7.1 lpep pickup datetime

We just keep the hours

```
df$f.pickup<-substr(strptime(df$qual.lpep_pickup_datetime, "%Y-%m-%d %H:%M:%S"), 12, 13)
df$f.pickup<-factor(df$f.pickup)
summary(df$f.pickup)
```

```
##  00  01  02  03  04  05  06  07  08  09  10  11  12  13  14  15  16  17  18  19
## 242 173 144 125  92  53  64 127 182 204 210 187 191 182 250 240 310 275 339 329
##  20  21  22  23
## 323 251 277 230
```

7.2 lpep dropoff datetime

We just keep the hours

```
df$f.dropoff<-substr(strptime(df$qual.lpep_dropoff_datetime, "%Y-%m-%d %H:%M:%S"), 12, 13)
df$f.dropoff<-factor(df$f.dropoff)
summary(df$f.dropoff)
```

```
##  00  01  02  03  04  05  06  07  08  09  10  11  12  13  14  15  16  17  18  19
## 246 183 140 127 111  53  47 113 169 202 206 212 178 190 237 235 306 279 334 335
##  20  21  22  23
## 308 280 270 239
```

8 Creating new variables

8.1 Period (as factor)

```
df$q.hour<-as.numeric(substr(strptime(df$qual.lpep_pickup_datetime, "%Y-%m-%d %H:%M:%S"),12,13))
df$f.period<-1
df$f.period[df$q.hour>19]<-1
df$f.period[df$q.hour>7]<-2
df$f.period[df$q.hour>10]<-3
df$f.period[df$q.hour>16]<-4
df$f.period<-factor(df$f.period,labels=paste("period",c("night","morning","valley","afternoon")))
```

8.2 Trip length in km

```
df$q.tlenkm<-df$q.trip_distance*1.609344 # Miles to km
```

8.3 Travel time in min

```
df$q.traveltime<-(as.numeric(as.POSIXct(df$qual.lpep_dropoff_datetime))-as.numeric(as.POSIXct(df$qual.lp
```

8.4 Effective speed in km/h

```
df$q.espeed<-(df$q.tlenkm/(df$q.traveltime))*60
```

8.5 Caterogial variable for trip distance

We are going to set a categorical variable for the Trip_distance range.

We decided to create 3 levels: “Short_dist”, “Medium_dist” and“Long_dist”.

- Short_dist <= 2.5
- Medium_dist 2.5 < Trip_distance <= 5
- Long_dist > 5

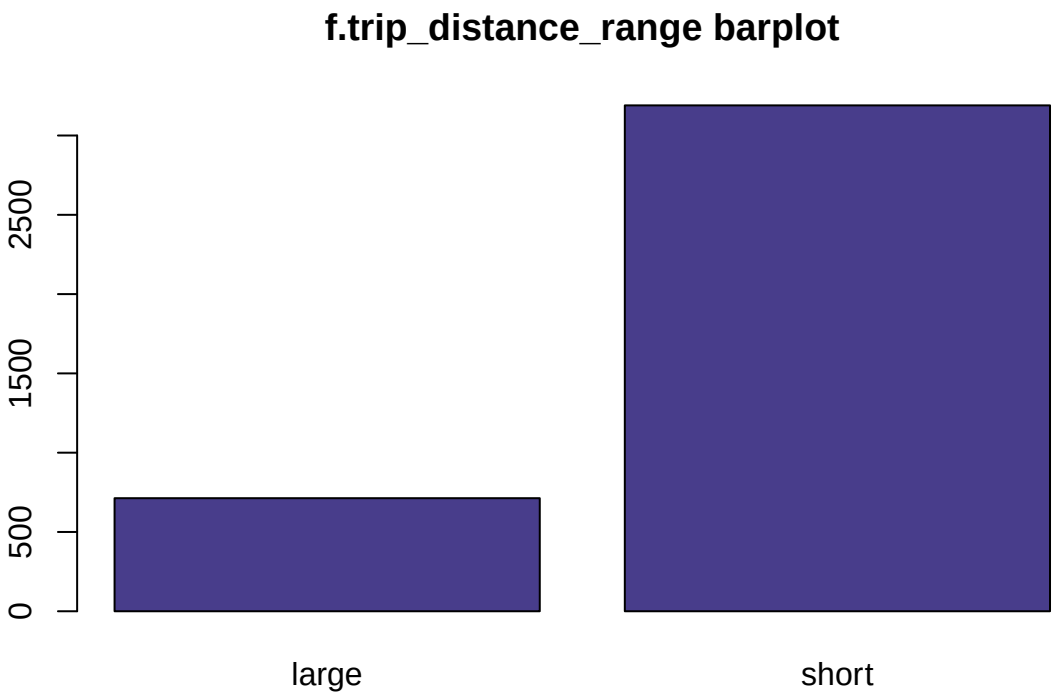
```
df$f.trip_distance_range[df$q.trip_distance <= 2.5] = "short"
df$f.trip_distance_range[(df$q.trip_distance > 2.5) & (df$Trip_distance <= 5)] = "medium"
df$f.trip_distance_range[df$q.trip_distance > 5] = "large"
```

We see, though, that it is not a factor yet, so we factor it.

```
df$f.trip_distance_range <- factor(df$f.trip_distance_range)
```

We see a barplot for the factor we created.

```
barplot(table(df$f.trip_distance_range),main="f.trip_distance_range barplot",col="darkslateblue")
```



9 Clean a little

```
names(df)
```

```
##  [1] "f.vendor_id"                "qual.lpep_pickup_datetime"
##  [3] "qual.lpep_dropoff_datetime" "f.store_and_fwd_flag"
##  [5] "f.rate_code_id"             "q.pickup_longitude"
##  [7] "q.pickup_latitude"          "q.dropoff_longitude"
##  [9] "q.dropoff_latitude"         "q.passenger_count"
## [11] "q.trip_distance"            "q.fare_amount"
## [13] "q.extra"                    "q.mta_tax"
## [15] "q.tip_amount"               "q.tolls_amount"
## [17] "q.improvement_surcharge"    "q.target.total_amount"
## [19] "f.payment_type"             "f.trip_type"
## [21] "f.pickup"                   "f.dropoff"
## [23] "q.hour"                     "f.period"
## [25] "q.tlenkm"                   "q.traveltime"
## [27] "q.espeed"                   "f.trip_distance_range"
```

```
summary(df)
```

```
##           f.vendor_id   qual.lpep_pickup_datetime qual.lpep_dropoff_datetime
##  vendor_mobile  :1062   Length:5000               Length:5000
##  vendor_verifone:3938   Class :character          Class :character
##                         Mode  :character          Mode  :character
##
##
##
##
##  f.store_and_fwd_flag            f.rate_code_id q.pickup_longitude
##  no :4982             rate_code_id_1    :4866   Min.   :-75.39
##  yes:  18             rate_code_id_other: 134   1st Qu.:-73.96
##                                                Median :-73.95
```

```
##                                                 Mean   :-73.89
##                                                 3rd Qu.:-73.92
##                                                 Max.   :  0.00
##
##  q.pickup_latitude q.dropoff_longitude q.dropoff_latitude q.passenger_count
##  Min.   : 0.00     Min.   :-75.31      Min.   : 0.00      Min.   :0.000
##  1st Qu.:40.70     1st Qu.:-73.97      1st Qu.:40.70      1st Qu.:1.000
##  Median :40.75     Median :-73.94      Median :40.75      Median :1.000
##  Mean   :40.72     Mean   :-73.80      Mean   :40.67      Mean   :1.375
##  3rd Qu.:40.80     3rd Qu.:-73.91      3rd Qu.:40.79      3rd Qu.:1.000
##  Max.   :41.04     Max.   :  0.00      Max.   :41.18      Max.   :6.000
##
##  q.trip_distance  q.fare_amount      q.extra          q.mta_tax
##  Min.   : 0.000   Min.   :-52.0   Min.   :-1.0000   Min.   :-0.5000
##  1st Qu.: 1.020   1st Qu.:  6.0   1st Qu.: 0.0000   1st Qu.: 0.5000
##  Median : 1.800   Median :  9.0   Median : 0.5000   Median : 0.5000
##  Mean   : 2.765   Mean   : 11.9   Mean   : 0.3517   Mean   : 0.4857
##  3rd Qu.: 3.420   3rd Qu.: 14.5   3rd Qu.: 0.5000   3rd Qu.: 0.5000
##  Max.   :52.790   Max.   :200.0   Max.   : 1.0000   Max.   : 0.5000
##
##   q.tip_amount    q.tolls_amount     q.improvement_surcharge
##  Min.   : 0.000   Min.   : 0.00000   Min.   :-0.3000
##  1st Qu.: 0.000   1st Qu.: 0.00000   1st Qu.: 0.3000
##  Median : 0.000   Median : 0.00000   Median : 0.3000
##  Mean   : 1.217   Mean   : 0.08369   Mean   : 0.2914
##  3rd Qu.: 2.000   3rd Qu.: 0.00000   3rd Qu.: 0.3000
##  Max.   :96.000   Max.   :18.04000   Max.   : 0.3000
##
##  q.target.total_amount     f.payment_type      f.trip_type      f.pickup
##  Min.   :-52.80        credit card:2455   street_hail:4885   18     : 339
##  1st Qu.:  7.80        cash       :2506   dispatch   : 115   19     : 329
##  Median : 11.16        no paid    :  39                      20     : 323
##  Mean   : 14.33                                              16     : 310
##  3rd Qu.: 17.16                                              22     : 277
##  Max.   :260.00                                              17     : 275
##                                                              (Other):3147
##    f.dropoff        q.hour                  f.period       q.tlenkm
##  19     : 335   Min.   : 0.00   period night    :1020   Min.   : 0.000
##  18     : 334   1st Qu.: 9.00   period morning  : 596   1st Qu.: 1.642
##  20     : 308   Median :15.00   period valley   :1360   Median : 2.897
##  16     : 306   Mean   :13.41   period afternoon:2024   Mean   : 4.450
##  21     : 280   3rd Qu.:19.00                          3rd Qu.: 5.504
##  17     : 279   Max.   :23.00                          Max.   :84.957
##  (Other):3158
##   q.traveltime         q.espeed       f.trip_distance_range
##  Min.   :   0.000   Min.   :   0.00   large: 713
##  1st Qu.:   5.917   1st Qu.:  14.60   short:3190
##  Median :   9.833   Median :  18.58   NA's :1097
##  Mean   :  20.059   Mean   :  23.07
##  3rd Qu.:  16.246   3rd Qu.:  23.70
##  Max.   :1438.183   Max.   :3881.74
##                     NA's   :2
```

```
df$qual.lpep_dropoff_datetime <- NULL
df$qual.lpep_pickup_datetime <- NULL
```

10

!!!!! fins aquí