New York cabs

Katerina Dimitrova, Jose Romero, Sergi Mucoz

March 18, 2018

Previous work

Load requiered packages

Select 5000 samples

```
#Load samples

#Use birthday of 1 member of the group
#set.seed(03101994)
#nrow(df)
#sam<-sample(1:nrow(df),5000)
#sam<-as.vector(sort(sam))

#df<-df[sam,]
#setwd("/Users/Sergi/Desktop/Sergi/CABS")
#df<-read.table("green_tripdata_2016-01.csv",header=T, sep=",")
#save.image("Taxi5000_raw.RData") # Dont execute again since it will create a new data and the followin
```

Load usefull functions

```
countX <- function(x,X) {
  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) }

countNA <- function(x) {
  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) }

man.dist.manual <- function(p1Lat, p1Lon, p2Lat, p2Lon) {
  #return(abs(pointDistance(c(p1$lon, p1$lat), c(p1$lon, p2$lat), longlat=TRUE)) + abs(pointDistance(c(
  R = 6371
  lat1 = degrees.to.radians(p1Lat)
  lon1 = degrees.to.radians(p1Lon)
```

```
  lat2 = degrees.to.radians(p2Lat)
  lon2 = degrees.to.radians(p2Lon)
  A_lat = lat2 - lat1
  A_lon = lon2 - lon1
  a = sin(A_lat/2)^2
  c = 2 * atan2(sqrt(a), sqrt(1-a))
  dist_lat = R * c
  a = sin(A_lon/2)^2
  c = 2 * atan2(sqrt(a), sqrt(1-a))
  dist_lon = R * c
  abs(dist_lat) + abs(dist_lon)
  return(abs(dist_lat) + abs(dist_lon))
}

degrees.to.radians<-function(value) {
  return(value*0.0174532925)
}
```

Delete unnecessary attributes

```
load("Taxi5000_raw2.RData")
table(df$Ehail_fee) ##Delete unnecessary row
```

```
## < table of extent 0 >
df$Ehail_fee<-NULL
```

Conversion of qualitative variables

Numeric variables corresponding to qualitative concepts are converted to factors. We also

VendorID

```
#No missing Data
missingData<-which(is.na(df$VendorID));length(missingData)
```

```
## [1] 0
#No errors
errors<-which(df$VendorID==0.0);length(errors)
```
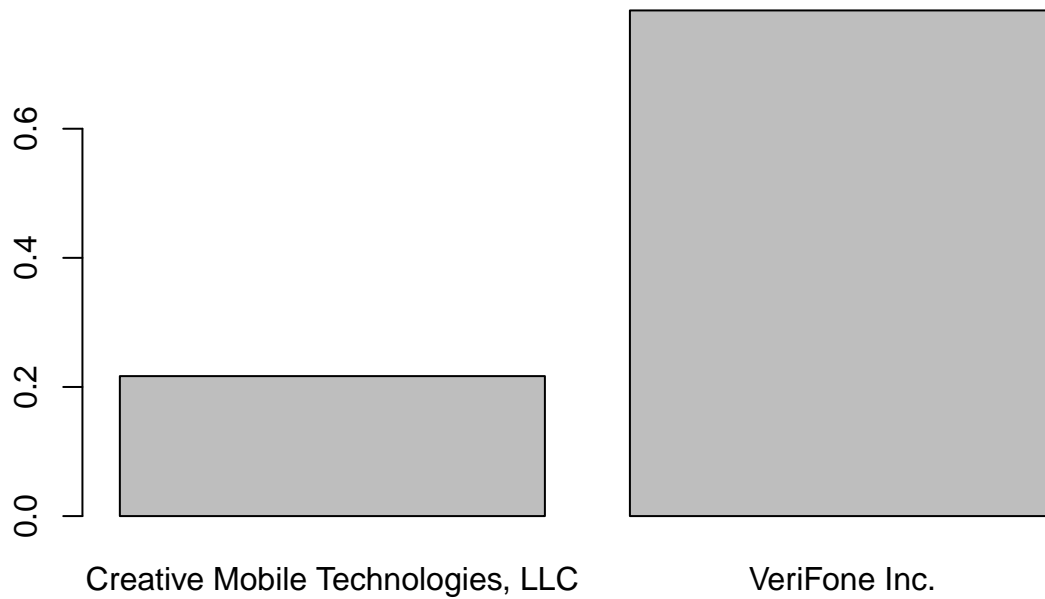
```
## [1] 0
df$VendorID<-factor(df$VendorID,labels=c("Creative Mobile Technologies, LLC","VeriFone Inc."))
table(df$VendorID)
```

```
##
## Creative Mobile Technologies, LLC                    VeriFone Inc.
##                              1084                             3916
barplot(prop.table(table(df$VendorID)))
```

RateCodeID

```
missingData<-which(is.na(df$RateCodeID));length(missingData) #No missing Data
```

```
## [1] 0
```

```
errors<-which(df$RateCodeID==0.0);length(errors) #No errors
```

```
## [1] 0
```

```
df$RateCodeID<-factor(df$RateCodeID,labels=c("Standard rate","JFK","Newark","Nassau or Westchester","Neg
table(df$RateCodeID)
```

```
##
##         Standard rate                 JFK              Newark
##                  4874                  14                   6
## Nassau or Westchester      Negotiated fare
##                     4                  102
```

```
barplot(prop.table(table(df$RateCodeID)))
```
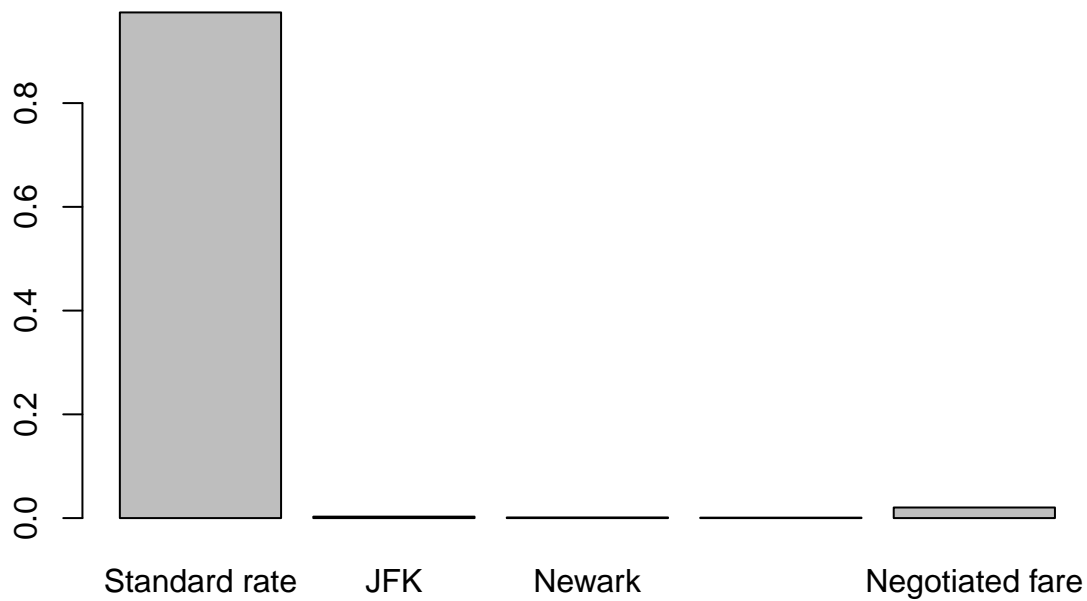
Store_and_fwd_flag

```
missingData<-which(is.na(df$Store_and_fwd_flag));length(missingData) #No missing Data
```

```
## [1] 0
```

```
erors<-which(df$Store_and_fwd_flag==0.0);length(erors) #No errors
```

```
## [1] 0
```

```
df$Store_and_fwd_flag<-factor(df$Store_and_fwd_flag,labels=c("not a store and forward trip","store and
table(df$Store_and_fwd_flag)
```

```
##
## not a store and forward trip        store and forward trip
##                         4982                            18
```

```
barplot(prop.table(table(df$Store_and_fwd_flag)))
```
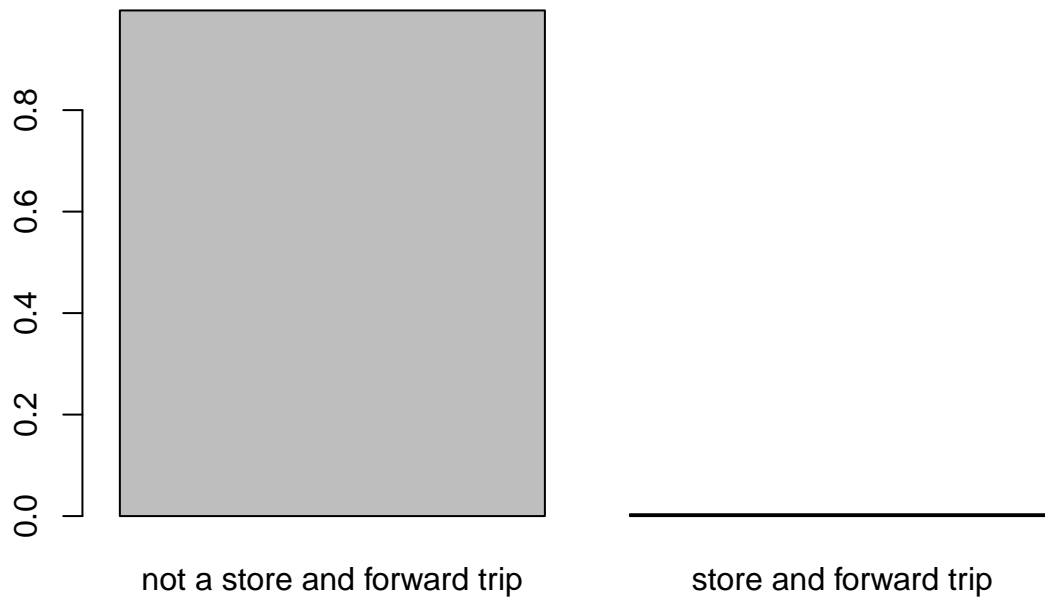
Payment_type

```
missingData<-which(is.na(df$Trip_type));length(missingData) #No missing Data
```

```
## [1] 0
```

```
errors<-which(df$Payment_type==0.0);length(errors) #No errors
```

```
## [1] 0
```

```
df$Payment_type<-factor(df$Payment_type,labels=c("Credit card","Cash", "No charge", "Dispute"))
table(df$Payment_type)
```

```
##
## Credit card        Cash   No charge     Dispute
##        2469        2485          23          23
```

```
barplot(prop.table(table(df$Payment_type)))
```
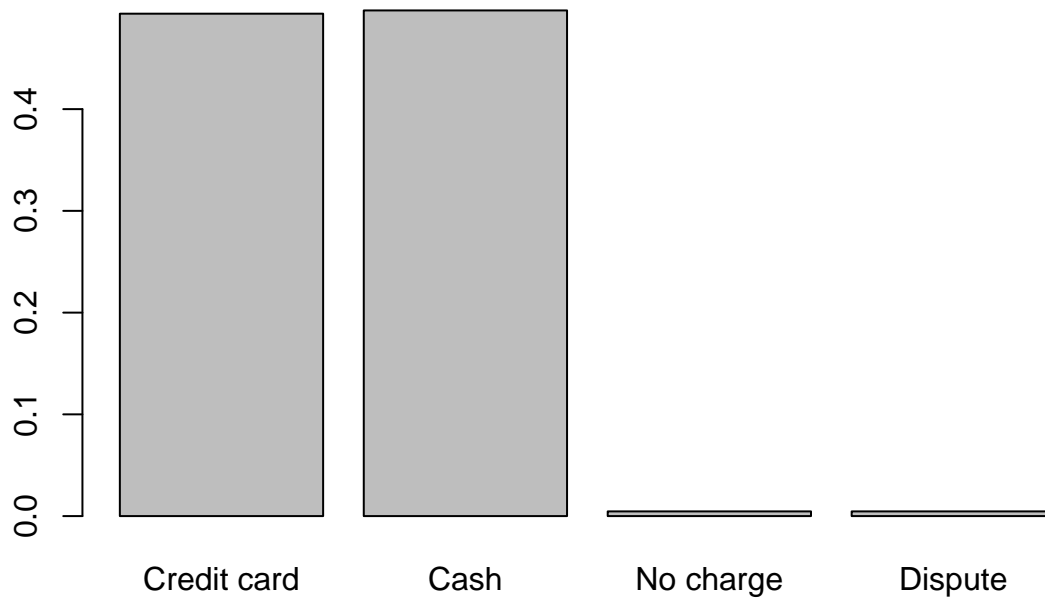
Trip_type

```
missingData<-which(is.na(df$Trip_type));length(missingData) #No missing Data
```

```
## [1] 0
```

```
errors<-which(df$Trip_type==0.0);length(errors) #No errors
```

```
## [1] 0
```

```
df$Trip_type<-factor(df$Trip_type,labels=c("Street-hail","Dispatch"))
table(df$Trip_type)
```

```
##
## Street-hail    Dispatch
##        4899         101
```

```
barplot(prop.table(table(df$Trip_type)))
```
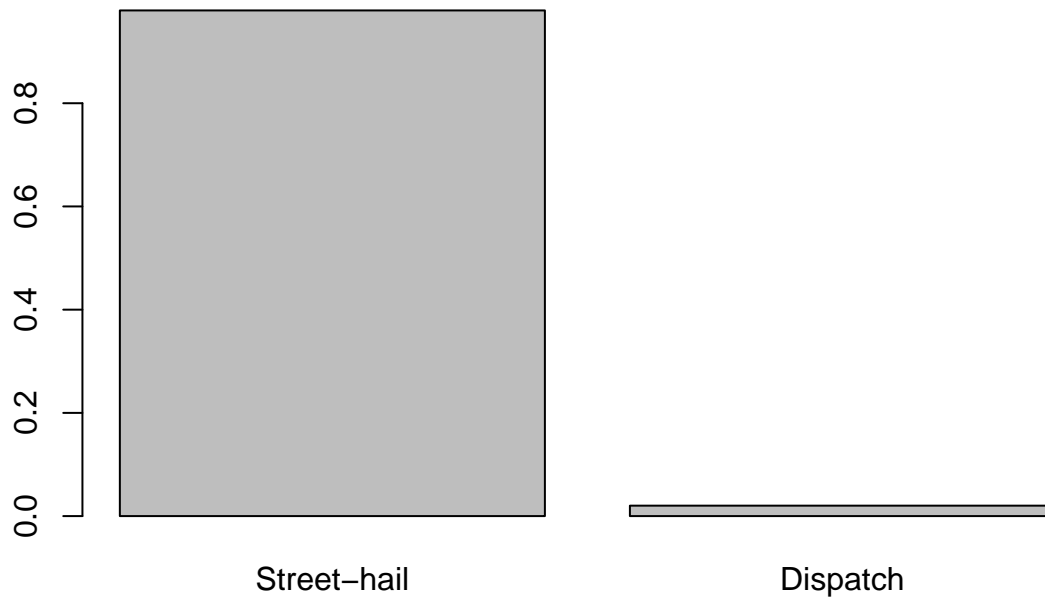
Univariant Descriptive Analysis

Passenger_count

```
## Number of missing values:
missingData<-which(is.na(df$Passenger_count));length(missingData) #No missing Data
```

```
## [1] 0
errors<-which(df$Passenger_count<=0.0);length(errors) #2 errors
```

```
## [1] 2
outliers<-which(df$Passenger_count>6.0);length(outliers) #0 outlier
```

```
## [1] 0
df[errors,"Passenger_count"]<-NA
df[outliers,"Passenger_count"]<-NA
boxplot(df$Passenger_count)
```
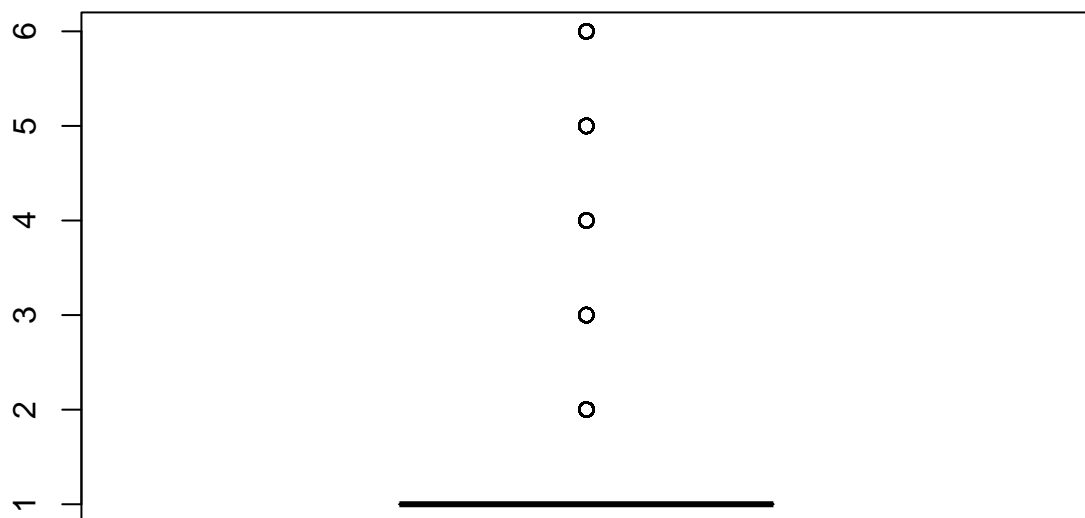
```r
hist(df$Passenger_count, col="pink")
```

Histogram of df$Passenger_count
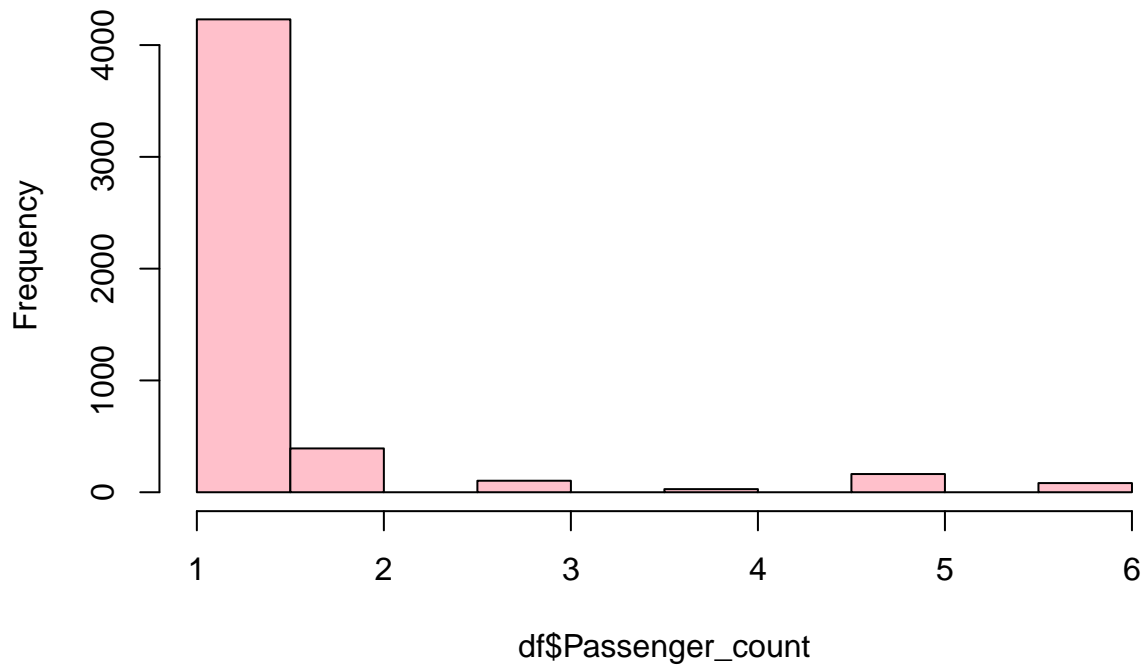
Trip_distance

```
missingData<-which(is.na(df$Trip_distance));length(missingData) #No missing Data
```

```
## [1] 0
```

```
errors<-which(df$Trip_distance<=0.0);length(errors) #59 errors
```

```
## [1] 59
```

```
dfaux<-df
ll<-which(is.na(df$Trip_distance));ll
```

```
## integer(0)
```

```
if(length(ll)>0){
  dfaux<-df[-ll,]
}
iqrvar<-IQR(dfaux$Trip_distance)
quantil3<-quantile(dfaux$Trip_distance, .75);quantil3 #get 3rd quartile
```

```
##    75%
## 3.4125
```

```
outliers<-which(df$Trip_distance>(iqrvar*3)+quantil3);length(outliers) #138 extreme outliers
```

```
## [1] 138
```

```
df[outliers,"Trip_distance"]<-NA
df[errors,"Trip_distance"]<-NA
boxplot(df$Trip_distance)
```
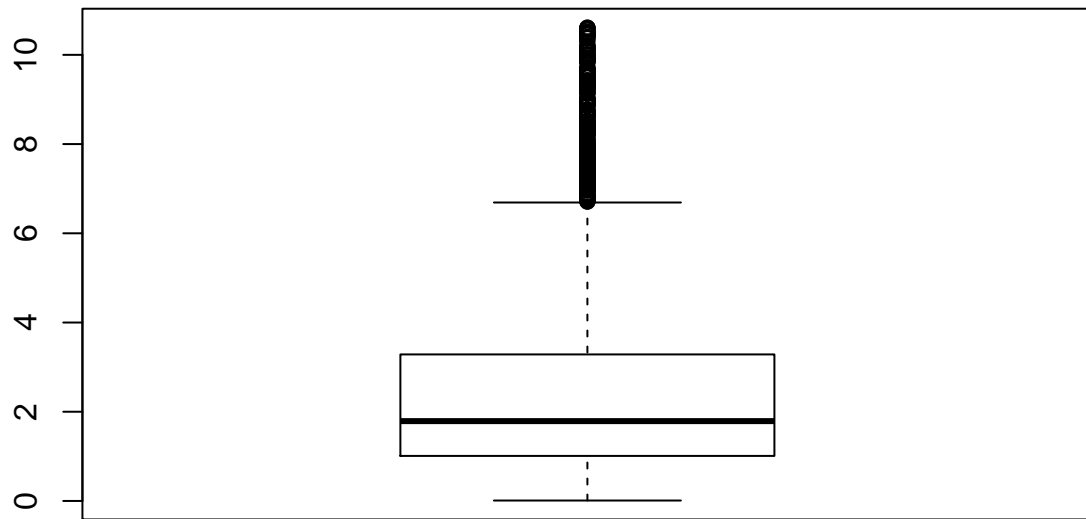


```
hist(df$Trip_distance, col="pink")
```

Histogram of df$Trip_distance

Frequency (y-axis), df$Trip_distance (x-axis)
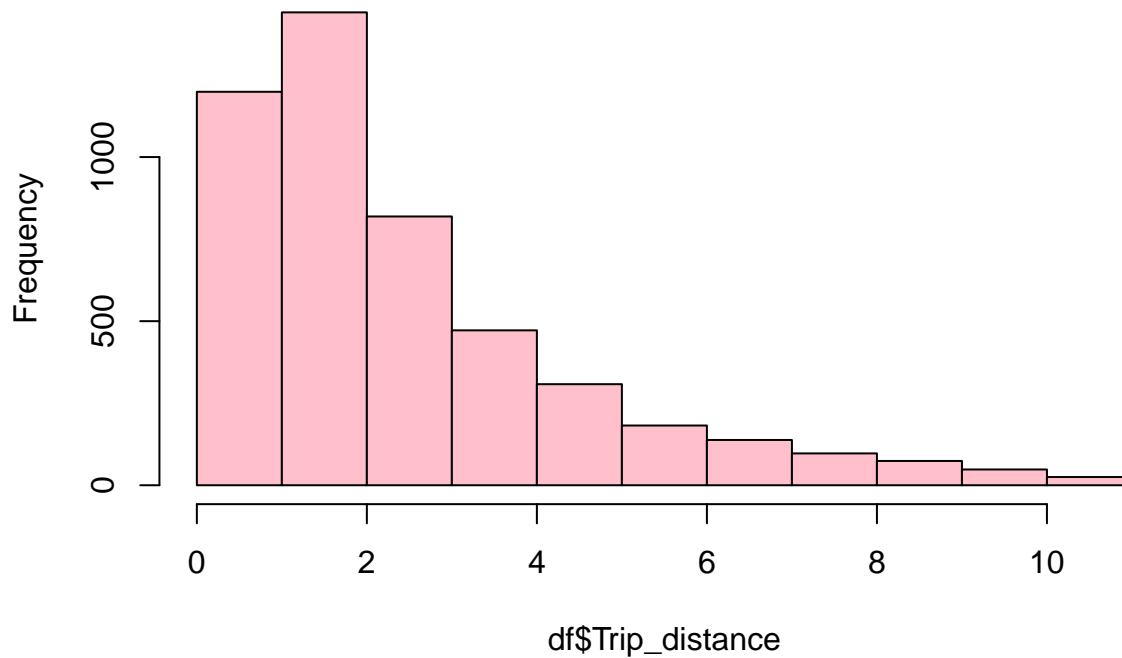
Pickup_longitude

```
missingData<-which(is.na(df$Trip_distance));length(missingData) #No missing Data
```

```
## [1] 197
```

```
#min and max longitudes for New York city boundaries
min_long <- -74.15
max_long <- -73.7004

errors<-which(df$Pickup_longitude< min_long);length(errors)
```

```
## [1] 1
```

```
errors<-c(errors,which(df$Pickup_longitude> max_long));length(errors)
```

```
## [1] 7
```

```
errors<-c(errors,which(df$Pickup_longitude==0.0));length(errors)
```

```
## [1] 12
```

```
df[errors,"Pickup_longitude"]<-NA #12 errors

ll<-which(is.na(df$Pickup_longitude));ll
```

```
## [1] 1580 1652 2639 3197 3221 4305 4639
```

```
if(length(ll)>0){
  dfaux<-df[-ll,]
}

iqrvar<-IQR(dfaux$Pickup_longitude)
quantil3<-quantile(dfaux$Pickup_longitude, .75);quantil3 #get 3rd quartile
```

```
##      75%
## -73.91782
```

```
quantil1<-quantile(dfaux$Pickup_longitude, .25);quantil1 #get 1st quartile
```

```
##      25%
## -73.96023
```

```
UpperOutlier<-which(df$Pickup_longitude>quantil3+(iqrvar*3));length(UpperOutlier) #14 extreme UpperOutl
```

```
## [1] 14
```

```
LowerOutlier<-which(df$Pickup_longitude<quantil1-(iqrvar*3));length(LowerOutlier) #1 extreme LowerOutli
```

```
## [1] 1
```

```
df[UpperOutlier,"Pickup_longitude"]<-NA
df[LowerOutlier,"Pickup_longitude"]<-NA
boxplot(df$Pickup_longitude)
```
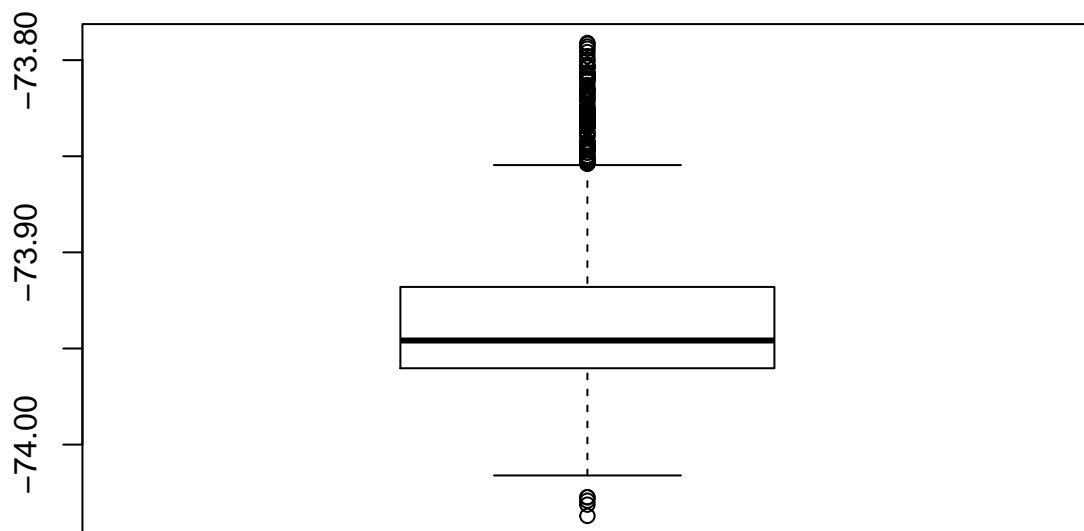


```
summary(df$Pickup_longitude)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
## -74.04  -73.96  -73.95  -73.94  -73.92  -73.79       22
```

Pickup_latitude

```
missingData<-which(is.na(df$Pickup_latitude));length(missingData) #No missing Data
```

```
## [1] 0
```

```
#we need to add here error control (what if longitude is out of scope?) and outlier management

summary(df$Pickup_latitude)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   40.69   40.75   40.70   40.80   40.92
#min and max latitudes for New York city boundaries
min_lat <- 40.5774
max_lat <- 40.9176

errors<-which(df$Pickup_latitude< min_lat);length(errors)
```

```
## [1] 11
```

```
errors<-c(errors,which(df$Pickup_latitude> max_lat));length(errors)
```

```
## [1] 12
```

```
errors<-c(errors,which(df$Pickup_latitude==0.0));length(errors)
```

```
## [1] 17
```

```
df[errors,"Pickup_latitude"]<-NA #17 errors

ll<-which(is.na(df$Pickup_latitude));ll
```

```
##  [1]  179 1580 2110 2241 2354 2639 2971 3197 3221 4305 4635 4639
```

```
if(length(ll)>0){
  dfaux<-df[-ll,]
}

iqrvar<-IQR(dfaux$Pickup_latitude)
quantil3<-quantile(dfaux$Pickup_latitude, .75);quantil3 #get 3rd quartile
```

```
##      75%
## 40.79892
```

```
quantil1<-quantile(dfaux$Pickup_latitude, .25);quantil1 #get 1st quartile
```

```
##      25%
## 40.69458
```

```
UpperOutlier<-which(df$Pickup_latitude>quantil3+(iqrvar*3));length(UpperOutlier) #0 extreme UpperOutlier
```

```
## [1] 0
```

```
LowerOutlier<-which(df$Pickup_latitude<quantil1-(iqrvar*3));length(LowerOutlier) #0 extreme LowerOutlie
```

```
## [1] 0
```

```r
df[UpperOutlier,"Pickup_latitude"]<-NA
df[LowerOutlier,"Pickup_latitude"]<-NA
boxplot(df$Pickup_latitude)
```



```r
summary(df$Pickup_latitude)
```
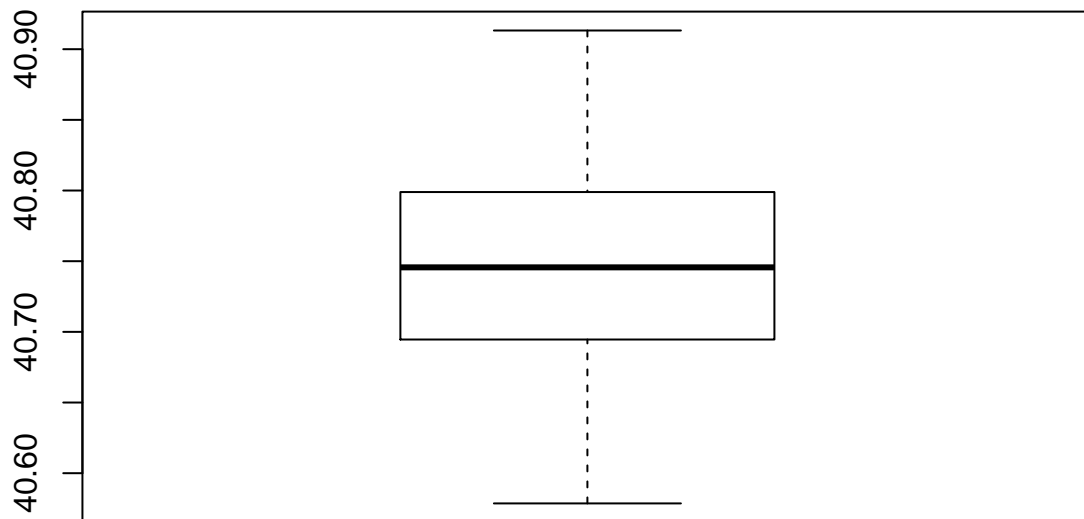
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   40.58   40.69   40.75   40.75   40.80   40.91      12
```

Dropoff_longitude

```r
missingData<-which(is.na(df$Dropoff_longitude));length(missingData) #No missing Data
```

```
## [1] 0
```

```r
errors<-c(errors,which(df$Dropoff_longitude==0.0));length(errors) #26 errors
```

```
## [1] 26
```

```r
df[errors,"Dropoff_longitude"]<-NA
```

```r
ll<-which(is.na(df$Dropoff_longitude));ll
```

```
##  [1]  179  638 1580 1713 1986 2026 2110 2241 2354 2639 2698 2971 3109 3197
## [15] 3221 4097 4285 4305 4635 4639
```

```
if(length(ll)>0){
  dfaux<-df[-ll,]
}

iqrvar<-IQR(dfaux$Dropoff_longitude)
quantil3<-quantile(dfaux$Dropoff_longitude, .75);quantil3 #get 3rd quartile
```

```
##      75%
## -73.91151
```

```
quantil1<-quantile(dfaux$Dropoff_longitude, .25);quantil1 #get 1st quartile
```

```
##      25%
## -73.9675
```

```
UpperOutlier<-which(df$Dropoff_longitude>quantil3+(iqrvar*3));length(UpperOutlier) #0 extreme UpperOutl
```

```
## [1] 7
```

```
LowerOutlier<-which(df$Dropoff_longitude<quantil1-(iqrvar*3));length(LowerOutlier) #0 extreme LowerOutl
```

```
## [1] 5
```

```
df[UpperOutlier,"Dropoff_longitude"]<-NA
df[LowerOutlier,"Dropoff_longitude"]<-NA

boxplot(df$Dropoff_longitude)
```
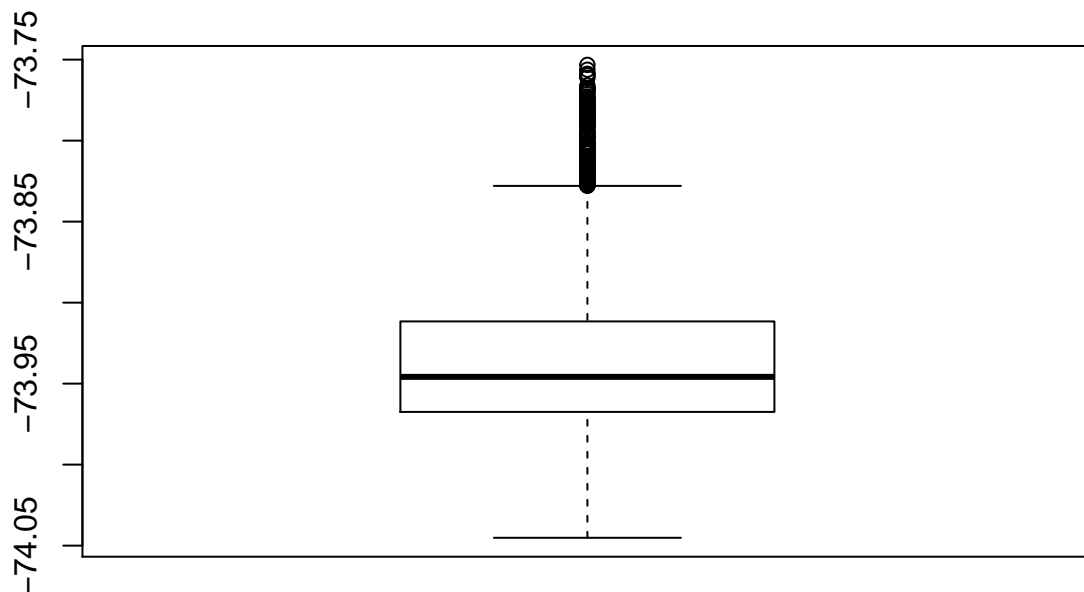
Dropoff_latitude

```
missingData<-which(is.na(df$Dropoff_latitude));length(missingData) #No missing Data
```

```
## [1] 0
```

```
errors<-c(errors,which(df$Dropoff_latitude==0.0));length(errors) #35 errors
```

```
## [1] 35
```

```
df[errors,"Dropoff_latitude"]<-NA

ll<-which(is.na(df$Dropoff_latitude));ll
```

```
##  [1]  179  638 1580 1713 1986 2026 2110 2241 2354 2639 2698 2971 3109 3197
## [15] 3221 4097 4285 4305 4635 4639
```

```
if(length(ll)>0){
  dfaux<-df[-ll,]
}

iqrvar<-IQR(dfaux$Dropoff_latitude)
quantil3<-quantile(dfaux$Dropoff_latitude, .75);quantil3 #get 3rd quartile
```

```
##      75%
## 40.78581
```

```
quantil1<-quantile(dfaux$Dropoff_latitude, .25);quantil1 #get 1st quartile
```

```
##      25%
## 40.69629
```

```
UpperOutlier<-which(df$Dropoff_latitude>quantil3+(iqrvar*3));length(UpperOutlier) #0 extreme UpperOutli
```

```
## [1] 0
```

```
LowerOutlier<-which(df$Dropoff_latitude<quantil1-(iqrvar*3));length(LowerOutlier) #0 extreme LowerOutli
```

```
## [1] 0
```

```
df[UpperOutlier,"Dropoff_latitude"]<-NA
df[LowerOutlier,"Dropoff_latitude"]<-NA

boxplot(df$Dropoff_latitude)
```
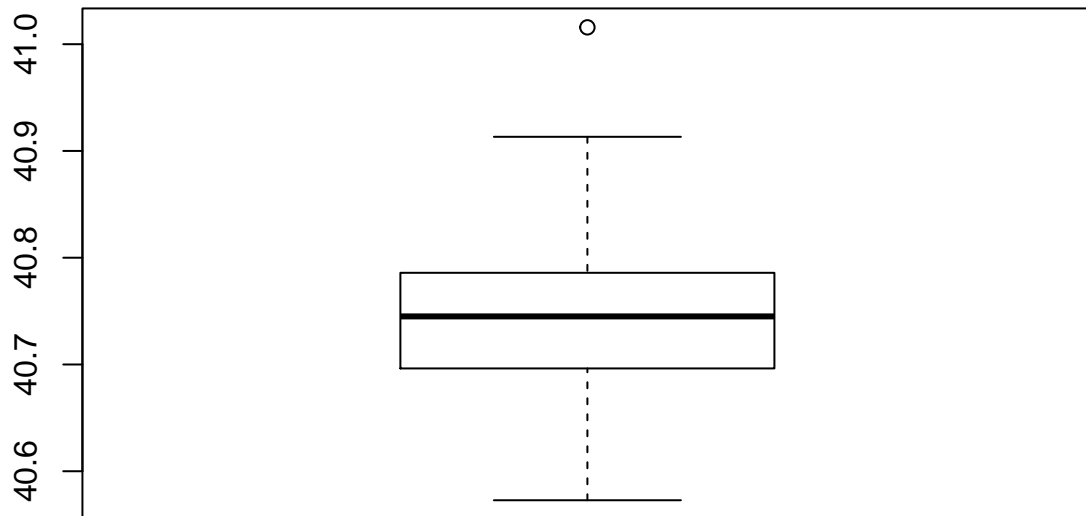
Fare_amount

```r
missingData<-which(is.na(df$Fare_amount));length(missingData) #No missing Data
```

```
## [1] 0
```

```r
sel<-which(df$Fare_amount<=0.0);length(sel) #10 missings
```

```
## [1] 23
```

```r
outlier<-which(df$Fare_amount>100);length(outlier) #1 outlier
```

```
## [1] 3
```

```r
df[sel,"Fare_amount"]<-NA
df[outlier,"Fare_amount"]<-NA
boxplot(df$Fare_amount)
```
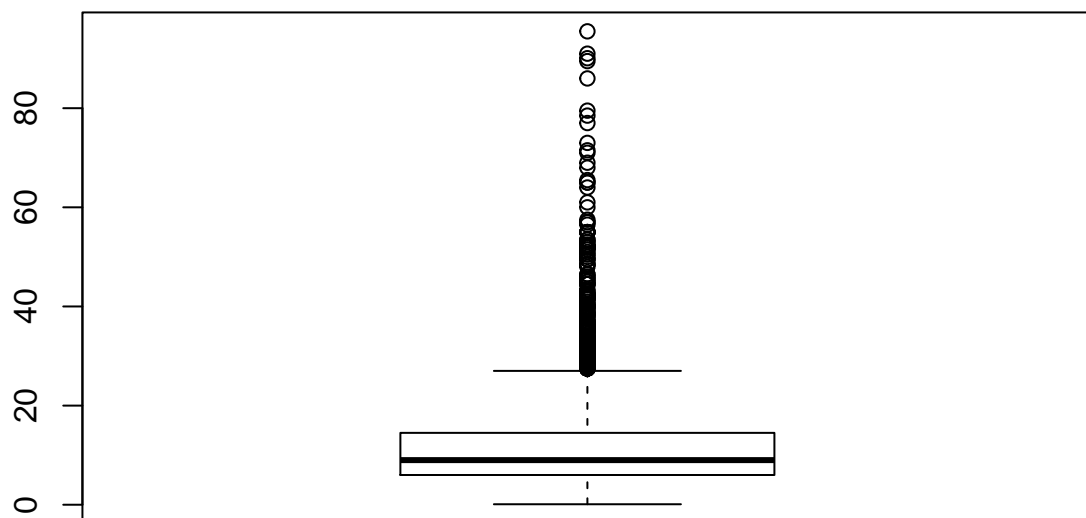
```r
hist(df$Fare_amount, col="pink")
```

Histogram of df$Fare_amount
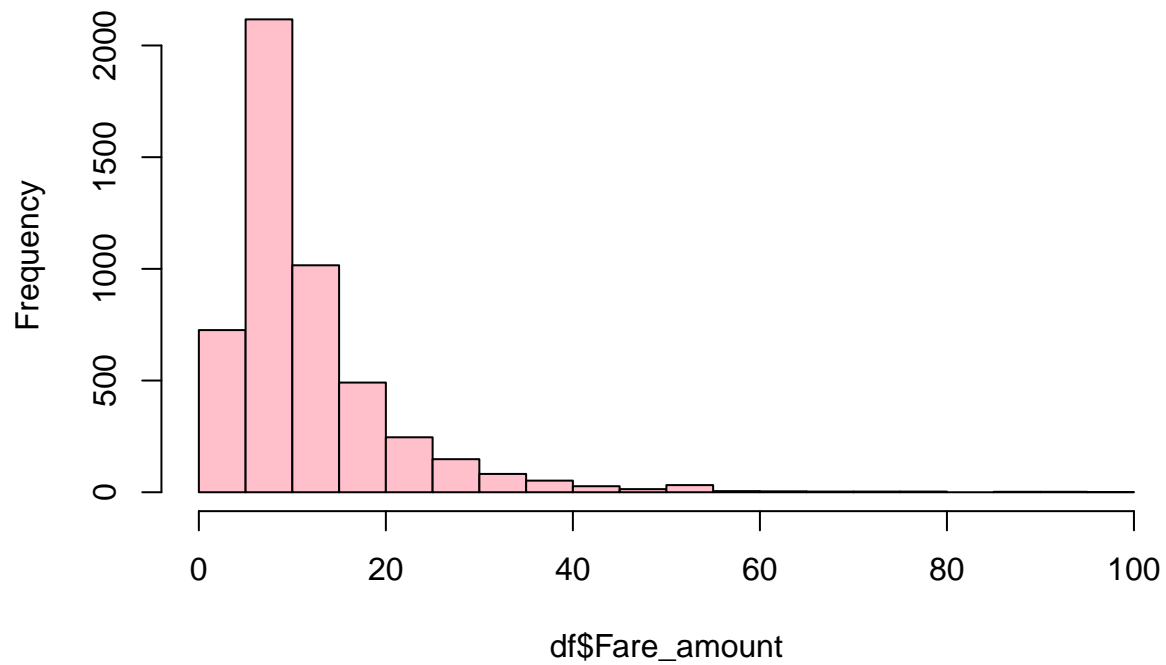

```r
summary(df$Fare_amount)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     0.1     6.0     9.0    12.0    14.5    95.5      26
```

Extra

```r
missingData<-which(is.na(df$Extra));length(missingData) #No missing Data
```

```
## [1] 0
```

```r
sel<-which(df$Extra<0.0);length(sel) #10 missings
```

```
## [1] 4
```

```r
df[sel,"Extra"]<-NA
boxplot(df$Extra)
```
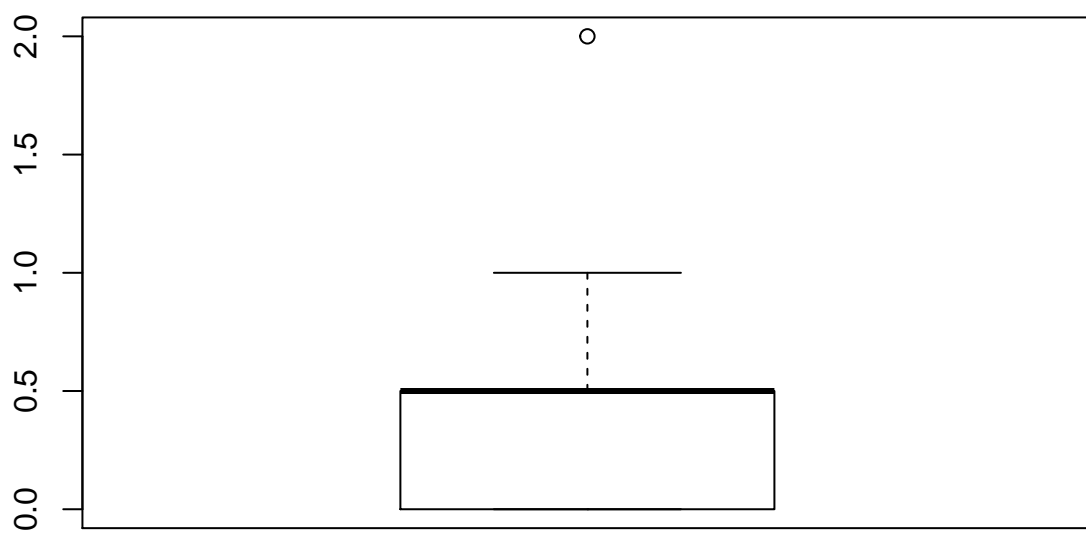
```r
hist(df$Extra, col="pink")
```

Histogram of df$Extra
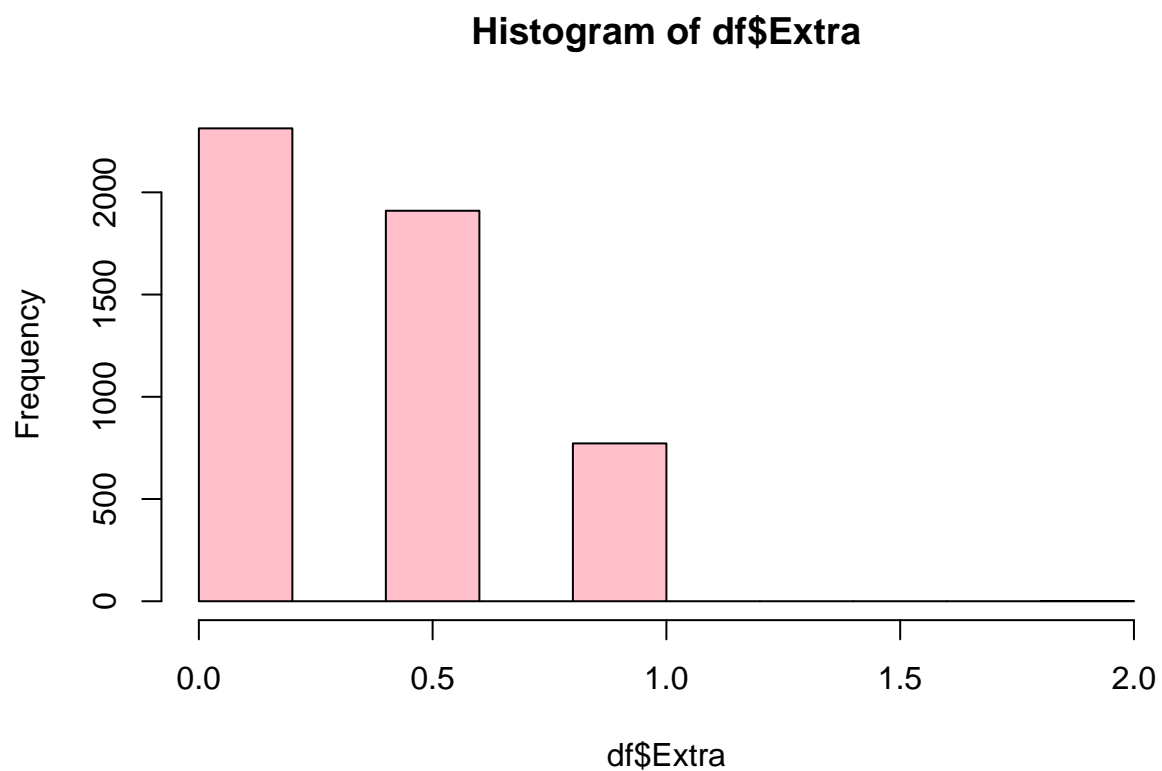
MTA__tax

```
missingData<-which(is.na(df$MTA_tax));length(missingData) #No missing Data
```

```
## [1] 0
```

```
sel<-which(df$MTA_tax<0.0);length(sel) #103 missings
```

```
## [1] 10
```

```
df[sel,"MTA_tax"]<-NA
boxplot(df$MTA_tax)
```
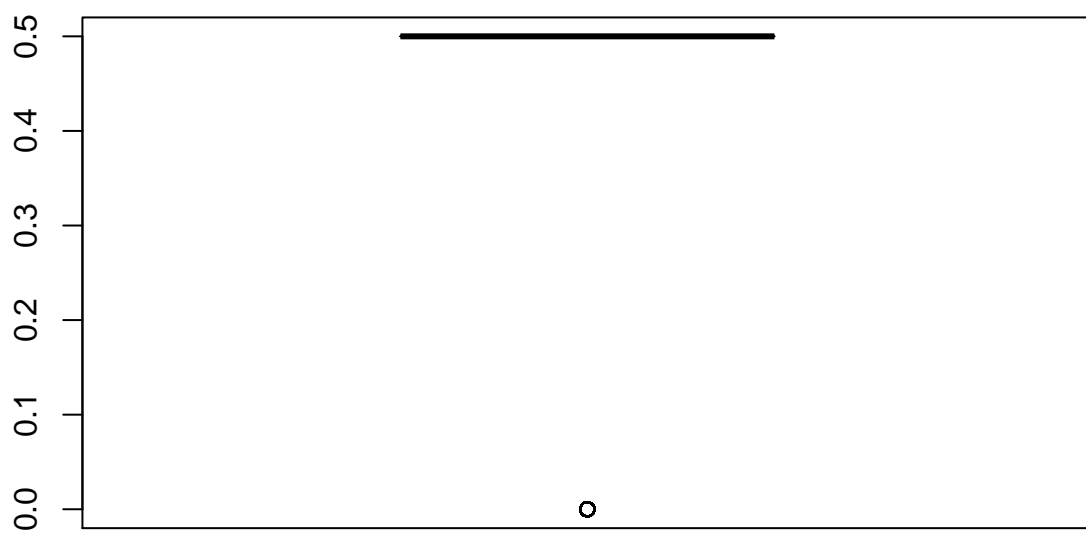
```r
hist(df$MTA_tax, col="pink")
```

Histogram of df$MTA_tax
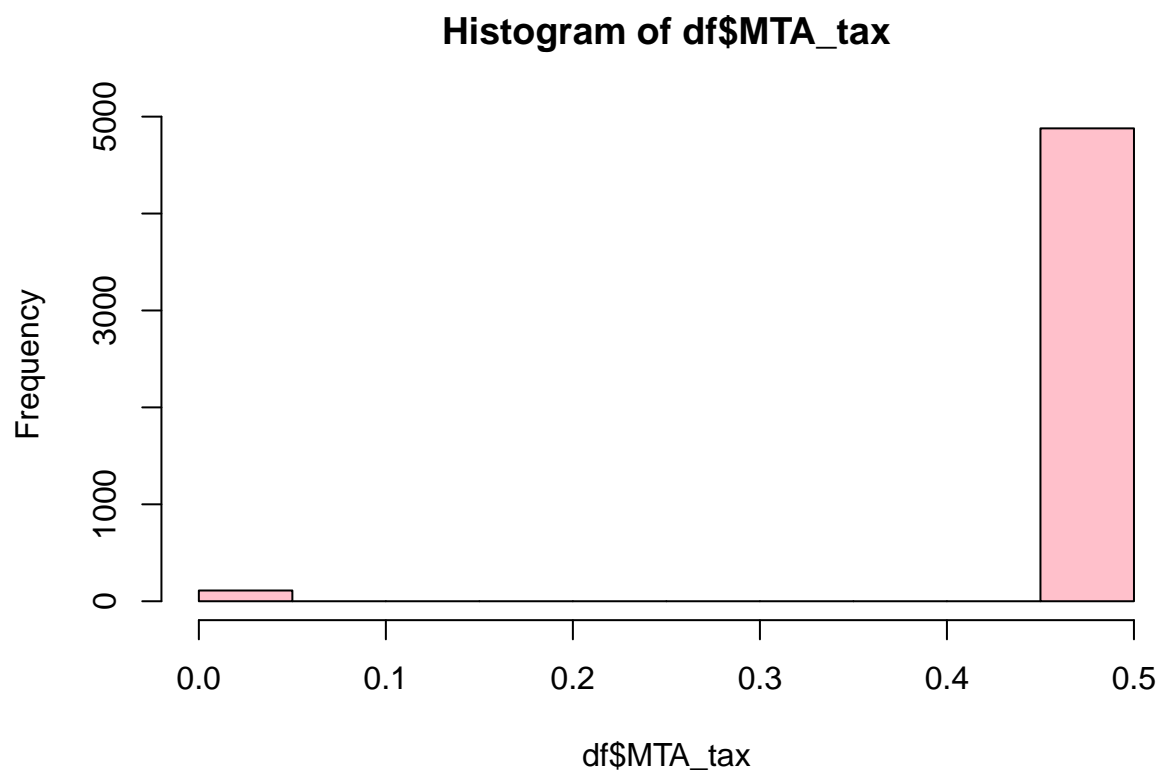
Improvement_surcharge

```
missingData<-which(is.na(df$improvement_surcharge));length(missingData) #No missing Data
```

```
## [1] 0
```

```
sel<-which(df$improvement_surcharge<0.0);length(sel)
```

```
## [1] 10
```

```
df[sel,"improvement_surcharge"]<-NA
boxplot(df$improvement_surcharge)
```
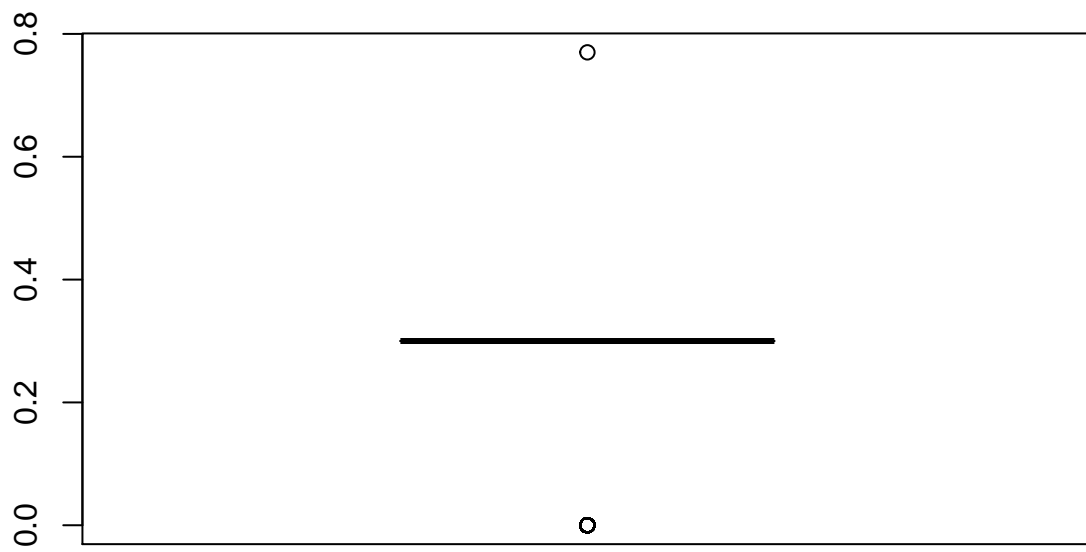
```
hist(df$improvement_surcharge, col="pink")
```

Histogram of df$improvement_surcharge

Tip_amount

```
missingData<-which(is.na(df$Tip_amount));length(missingData) #No missing Data
```

```
## [1] 0
```

```
sel<-which(df$Tip_amount<0.0);length(sel) #107 missings
```

```
## [1] 1
```

```
outlier<-which(df$Tip_amount>60.0);length(outlier) #1 missings
```

```
## [1] 3
```

```
df[outlier,"Tip_amount"]<-NA
df[sel,"Tip_amount"]<-NA
boxplot(df$Tip_amount)
```
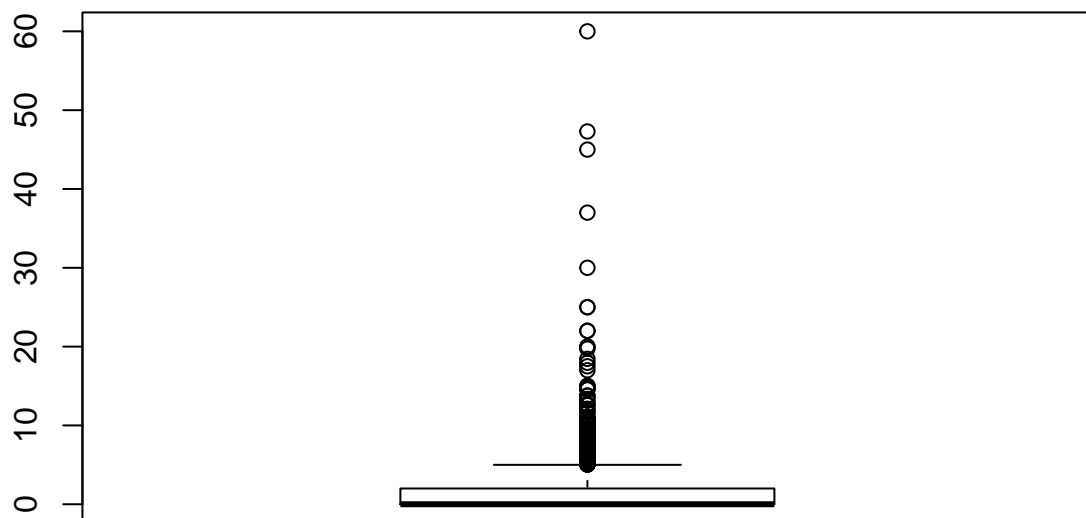
```r
hist(df$Tip_amount, col="pink")
```

Histogram of df$Tip_amount
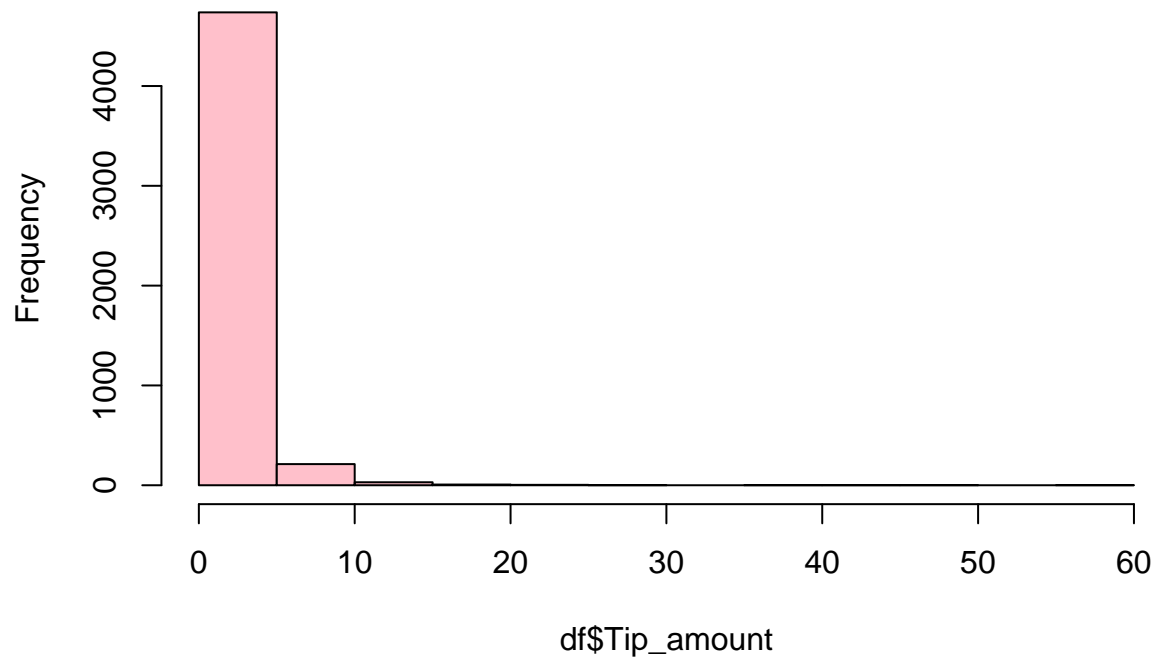
Tolls_amount

```
missingData<-which(is.na(df$Tolls_amount));length(missingData) #No missing Data
```

```
## [1] 0
```

```
sel<-which(df$Tolls_amount<0.0);length(sel) #0 missings
```

```
## [1] 0
```

```
df[sel,"Tolls_amount"]<-NA
boxplot(df$Tolls_amount)
```

```r
hist(df$Tolls_amount, col="pink")
```
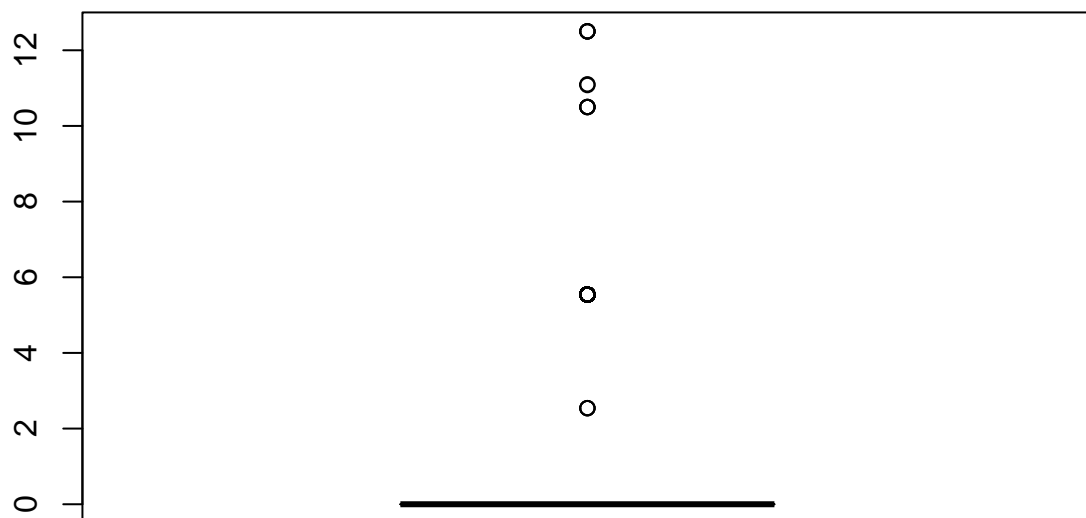
Histogram of df$Tolls_amount
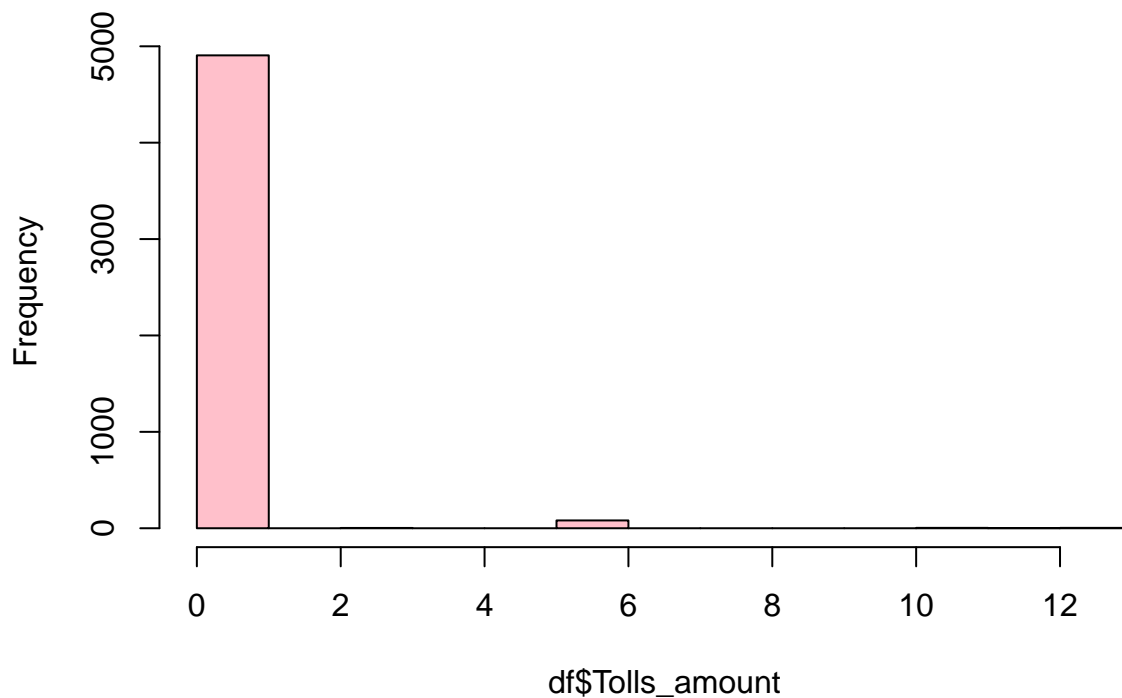
Total_amount (Target)

```
missingData<-which(is.na(df$Total_amount));length(missingData) #No missing Data
```

```
## [1] 0
```

```
ll<-which(is.na(df$Total_amount));ll
```

```
## integer(0)
```

```
if(length(ll)>0){
  dfaux<-df[-ll,]
}
iqrvar<-IQR(dfaux$Total_amount)
quantil3<-quantile(dfaux$Total_amount, .75) #get 3rd quartile
sel<-which(df$Total_amount<=0.0);length(sel) #22 errors
```

```
## [1] 23
```

```
df[sel,"Total_amount"]<-NA
outlier<-which(df$Total_amount>(iqrvar*3)+quantil3);length(outlier) #72 extreme outliers
```

```
## [1] 111
```

```
df[outlier,"Total_amount"]<-NA
boxplot(df$Total_amount)
```
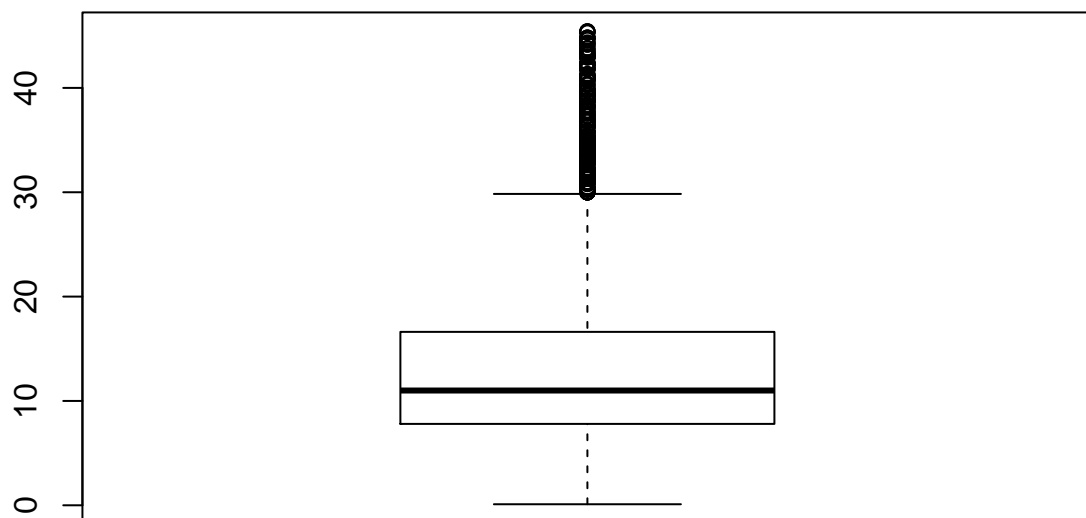
```r
hist(df$Total_amount, col="pink")
```

Histogram of df$Total_amount
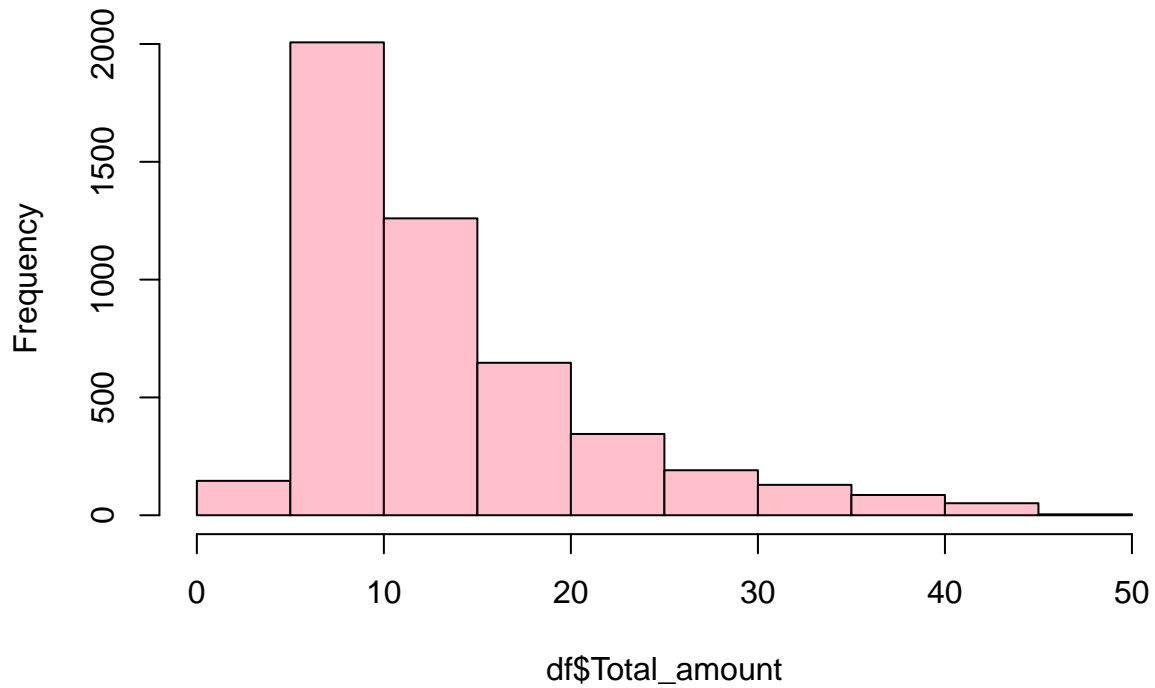
Number of missing values:

```
mis1<-countNA(df)
attributes(mis1)
```

```
## $names
## [1] "mis_col" "mis_ind"
```

```
rank(mis1$mis_col)
```

```
## mis_x  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>
##  4.5   4.5   4.5   4.5   4.5  16.0  14.0  18.0  15.0   9.0  20.0  17.0
## <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>
## 10.5  12.5  10.5   4.5  12.5  19.0   4.5   4.5
```

```
df$mis_ind <- mis1$mis_ind # new attribute missing values
mis1$mis_col
```

```
##                        mis_x
## VendorID                   0
## lpep_pickup_datetime       0
## Lpep_dropoff_datetime      0
## Store_and_fwd_flag         0
## RateCodeID                 0
## Pickup_longitude          22
```

```
## Pickup_latitude           12
## Dropoff_longitude         32
## Dropoff_latitude          20
## Passenger_count            2
## Trip_distance            197
## Fare_amount               26
## Extra                      4
## MTA_tax                   10
## Tip_amount                 4
## Tolls_amount               0
## improvement_surcharge     10
## Total_amount             134
## Payment_type               0
## Trip_type                  0
```

Creating AnyTip

```
df$AnyTip<-ifelse(df$Tip_amount<0.0001,0,1)
df$AnyTip<-factor(df$AnyTip,labels=paste("AnyTip",c("No","Yes")))
```

Creating auxiliar variables and doing their analysis

Trip length

```
for (i in 1:nrow(df)){
  df$trip_length[i] <-man.dist.manual(df$Pickup_latitude[i],df$Pickup_longitude[i],df$Dropoff_latitude[
}

summary(df$trip_length)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.000   1.790   3.151   4.544   5.684  38.866      43
```

```
boxplot(df$trip_length)
```
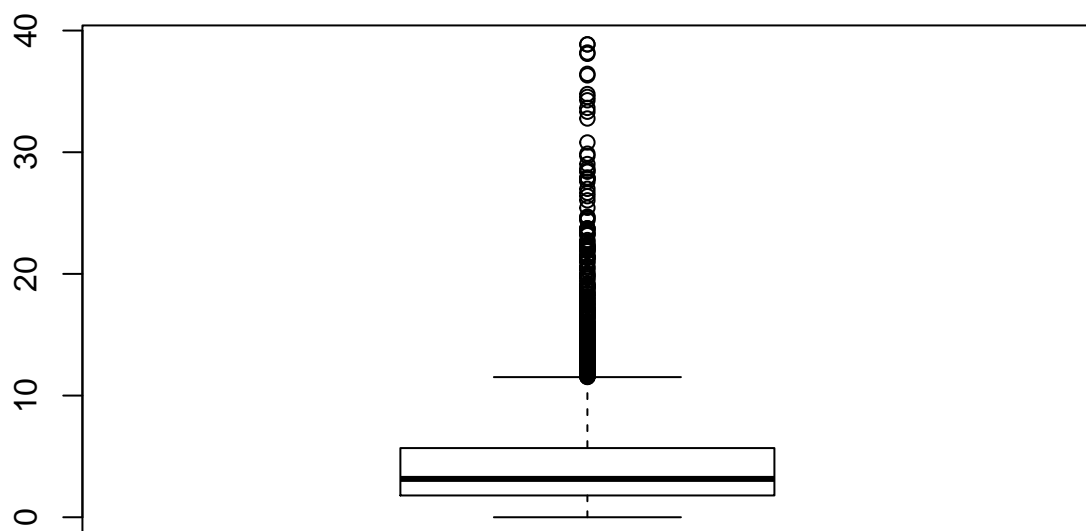
Trip distance in km

```
df$trip_distance_km<-df$Trip_distance*1.609344 # Miles to km

summary(df$trip_distance_km)

##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.01609 1.62544 2.88073 3.99502 5.28670 17.07514     197
boxplot(df$trip_distance_km)
```
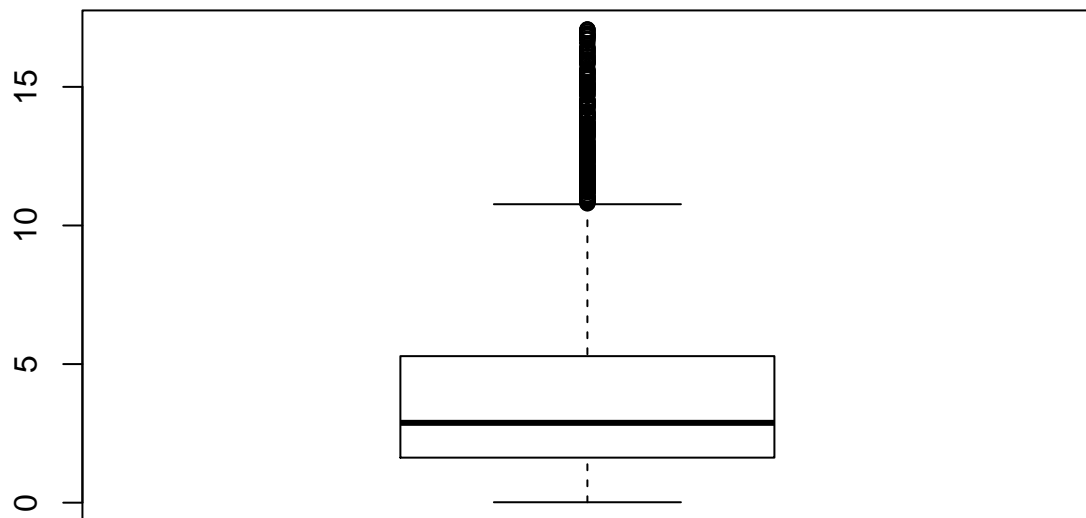
Travel time in minutes

```
b1<-as.POSIXlt(df$lpep_pickup_datetime)
b2<-as.POSIXlt(df$Lpep_dropoff_datetime)
df$travel_time<-as.double(difftime(b2,b1,units='min'))
error<-which(df$travel_time== 0.0);length(error) #No missing Data
```

```
## [1] 6
```

```
df[error,"travel_time"]<-NA
summary(df$travel_time)
```

```
##      Min.  1st Qu.   Median      Mean  3rd Qu.      Max.     NA's
##    0.0167   5.9208   9.8833   20.5748  16.2792 1438.3167        6
```

```
boxplot(df$travel_time)
```
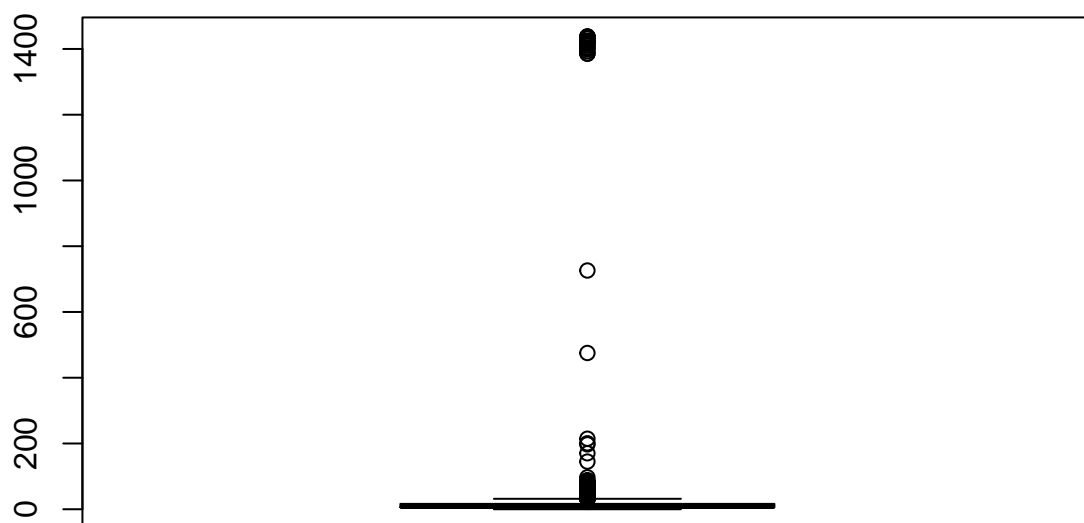
Pick_up_hour

```
mydate <- as.POSIXlt(df$lpep_pickup_datetime)
df$pick_up_hour <- mydate$hour
```

```
summary(df$pick_up_hour)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    9.00   15.00   13.47   19.00   23.00
```

```
boxplot(df$pick_up_hour)
```

Pick_up_period
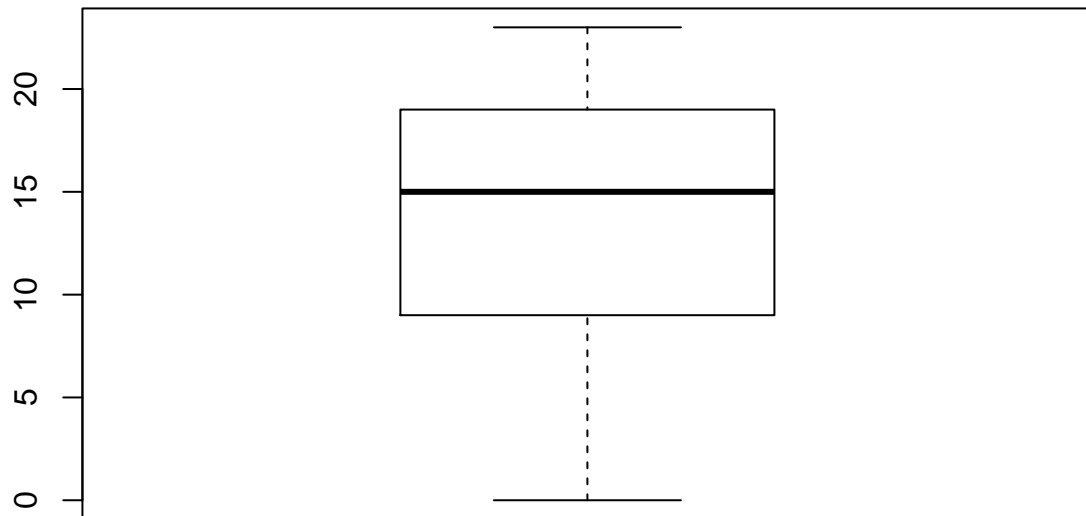
```r
# night, morning, valley and afternoon

df$pick_up_period= cut(df$pick_up_hour, breaks = c(-1, 5, 11, 17, 23), labels= c("night", "morning", "va

summary(df$pick_up_period)
```

```
##    night  morning   valley afternoon
##      807     1011     1411      1771
```

Declaring vectors of data

```r
names(df)
```

```
##  [1] "VendorID"              "lpep_pickup_datetime"
##  [3] "Lpep_dropoff_datetime" "Store_and_fwd_flag"
##  [5] "RateCodeID"            "Pickup_longitude"
##  [7] "Pickup_latitude"       "Dropoff_longitude"
##  [9] "Dropoff_latitude"      "Passenger_count"
## [11] "Trip_distance"         "Fare_amount"
## [13] "Extra"                 "MTA_tax"
## [15] "Tip_amount"            "Tolls_amount"
```

```
## [17] "improvement_surcharge" "Total_amount"
## [19] "Payment_type"          "Trip_type"
## [21] "mis_ind"               "AnyTip"
## [23] "trip_length"           "trip_distance_km"
## [25] "travel_time"           "pick_up_hour"
## [27] "pick_up_period"
```

```
vars_con<-names(df)[c(6,7,8,9,10,11,12,13,14,15,16,17,23,24,25,26)]
vars_con
```

```
##  [1] "Pickup_longitude"      "Pickup_latitude"
##  [3] "Dropoff_longitude"     "Dropoff_latitude"
##  [5] "Passenger_count"       "Trip_distance"
##  [7] "Fare_amount"           "Extra"
##  [9] "MTA_tax"               "Tip_amount"
## [11] "Tolls_amount"          "improvement_surcharge"
## [13] "trip_length"           "trip_distance_km"
## [15] "travel_time"           "pick_up_hour"
```

```
vars_dis<-names(df)[c(1,4,5,10,19,20,27,28)]
vars_dis
```

```
## [1] "VendorID"          "Store_and_fwd_flag" "RateCodeID"
## [4] "Passenger_count"   "Payment_type"       "Trip_type"
## [7] "pick_up_period"    NA
```

```
vars_res<-names(df)[c(18,22)]
vars_res
```

```
## [1] "Total_amount" "AnyTip"
```

Multivariant Outlier Detection

We haven't sorted some problems during implementation to achieve a logic result of aq.plot in order to identify multivariate outliers.

Correlations error variable

```
library(FactoMineR)
names (df)
```

```
##  [1] "VendorID"              "lpep_pickup_datetime"
##  [3] "Lpep_dropoff_datetime" "Store_and_fwd_flag"
##  [5] "RateCodeID"            "Pickup_longitude"
##  [7] "Pickup_latitude"       "Dropoff_longitude"
##  [9] "Dropoff_latitude"      "Passenger_count"
## [11] "Trip_distance"         "Fare_amount"
## [13] "Extra"                 "MTA_tax"
## [15] "Tip_amount"            "Tolls_amount"
## [17] "improvement_surcharge" "Total_amount"
## [19] "Payment_type"          "Trip_type"
## [21] "mis_ind"               "AnyTip"
## [23] "trip_length"           "trip_distance_km"
## [25] "travel_time"           "pick_up_hour"
```

```
## [27] "pick_up_period"
```

```
summary(df$mis_ind)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.0946  0.0000 10.0000
```

```
corV <- cor(df[,vars_con], df$mis_ind,  use = "complete.obs")
corV
```

```
##                              [,1]
## Pickup_longitude       0.003635284
## Pickup_latitude       -0.019736299
## Dropoff_longitude     -0.022786865
## Dropoff_latitude      -0.024234651
## Passenger_count       -0.017596988
## Trip_distance          0.128988183
## Fare_amount            0.340164553
## Extra                 -0.030636173
## MTA_tax               -0.157197287
## Tip_amount             0.410628927
## Tolls_amount           0.076495470
## improvement_surcharge -0.111629536
## trip_length            0.083428922
## trip_distance_km       0.128988183
## travel_time            0.055980499
## pick_up_hour           0.001477716
```

```
 # rank
rank(corV)
```

```
##  [1]  9.0  6.0  5.0  4.0  7.0 13.5 15.0  3.0  1.0 16.0 11.0  2.0 12.0 13.5
## [15] 10.0  8.0
```

Imputation

Remove observations with NA at targets

```
ll<-which(is.na(df$Total_amount));ll
```

```
##   [1]   57   74   82  145  176  247  323  333  351  404  454  460  468  472
##  [15]  526  553  609  637  690  734  745  825  831  883  907 1001 1022 1059
##  [29] 1062 1078 1082 1100 1105 1130 1159 1331 1361 1367 1368 1395 1421 1657
##  [43] 1689 1697 1723 1759 1761 1780 1854 1867 1905 2004 2069 2106 2140 2187
##  [57] 2249 2257 2334 2335 2411 2413 2428 2490 2506 2575 2634 2698 2722 2744
##  [71] 2840 2842 2845 2866 2874 2919 2971 2981 3005 3054 3067 3101 3181 3197
##  [85] 3293 3295 3346 3412 3484 3541 3705 3742 3759 3787 3788 3802 3803 3813
##  [99] 3874 3894 3933 3936 3947 3953 3988 4063 4075 4164 4206 4222 4252 4294
## [113] 4328 4348 4370 4391 4418 4431 4524 4574 4576 4597 4605 4659 4687 4700
## [127] 4714 4733 4778 4817 4890 4920 4923 4968
```

```
if(length(ll)>0){
  df<-df[-ll,]
}
```

```
ll<-which(is.na(df$AnyTip));ll
```

```
## integer(0)
```

```
if(length(ll)>0){
  df<-df[-ll,]
}
```

Imputation of numeric variables

```
library(missMDA)
names(df)
```

```
##  [1] "VendorID"              "lpep_pickup_datetime"
##  [3] "Lpep_dropoff_datetime" "Store_and_fwd_flag"
##  [5] "RateCodeID"            "Pickup_longitude"
##  [7] "Pickup_latitude"       "Dropoff_longitude"
##  [9] "Dropoff_latitude"      "Passenger_count"
## [11] "Trip_distance"         "Fare_amount"
## [13] "Extra"                 "MTA_tax"
## [15] "Tip_amount"            "Tolls_amount"
## [17] "improvement_surcharge" "Total_amount"
## [19] "Payment_type"          "Trip_type"
## [21] "mis_ind"               "AnyTip"
## [23] "trip_length"           "trip_distance_km"
## [25] "travel_time"           "pick_up_hour"
## [27] "pick_up_period"
```

```
res.comp <- imputePCA(df[,vars_con], ncp=4)
attributes(res.comp$completeObs)
```

```
## $dim
## [1] 4866   16
##
## $dimnames
## $dimnames[[1]]
##     [1] "285"   "307"   "401"   "593"   "636"   "886"   "904"
##     [8] "978"   "1135"  "1282"  "1409"  "1475"  "1495"  "1905"
##    [15] "2126"  "2151"  "2201"  "2271"  "2747"  "3065"  "3089"
##    [22] "3130"  "3221"  "3420"  "3679"  "4310"  "4754"  "5241"
##    [29] "5277"  "5649"  "6353"  "6364"  "6755"  "6869"  "7079"
##    [36] "7211"  "7342"  "7802"  "8138"  "8443"  "8619"  "8891"
##    [43] "8960"  "9207"  "9503"  "9747"  "9765"  "9984"  "10034"
##    [50] "10199" "10951" "10955" "10974" "11189" "11506" "11713"
##    [57] "12492" "12792" "13043" "13274" "13332" "13875" "13927"
##    [64] "14874" "14916" "15407" "15830" "16080" "16166" "16345"
##    [71] "16391" "17136" "17355" "18278" "18596" "18734" "19101"
##    [78] "19344" "19408" "19991" "20004" "20009" "20044" "20077"
##    [85] "20271" "20342" "20361" "20543" "20621" "20733" "20917"
##    [92] "21425" "21439" "21539" "21559" "21735" "22197" "22332"
##    [99] "22825" "22946" "23091" "23132" "23811" "24338" "24863"
##   [106] "25262" "25356" "26062" "26832" "27216" "27482" "27495"
##   [113] "27594" "27984" "28083" "28512" "29375" "29522" "30659"
```

```
## [120] "30856"  "31236"  "31456"  "31571"  "31583"  "31617"  "31726"
## [127] "32873"  "32952"  "33882"  "34250"  "34280"  "34374"  "34390"
## [134] "34922"  "35039"  "35207"  "35386"  "36076"  "36428"  "36540"
## [141] "36696"  "36863"  "36933"  "37035"  "37273"  "37506"  "37517"
## [148] "37561"  "37764"  "37821"  "37877"  "38445"  "38480"  "39213"
## [155] "39623"  "39723"  "39943"  "40226"  "40245"  "40497"  "40560"
## [162] "40802"  "40941"  "40943"  "40953"  "40969"  "42048"  "42779"
## [169] "43577"  "43958"  "44992"  "46311"  "46572"  "46653"  "46790"
## [176] "47428"  "47471"  "48166"  "48518"  "48796"  "48903"  "48915"
## [183] "49242"  "49244"  "49383"  "49421"  "49783"  "49849"  "50027"
## [190] "50328"  "50542"  "50979"  "50996"  "51868"  "51965"  "52728"
## [197] "52825"  "52931"  "53452"  "53536"  "53680"  "54025"  "54342"
## [204] "54359"  "54794"  "54843"  "54958"  "54994"  "55030"  "55082"
## [211] "55144"  "55353"  "55479"  "55718"  "56090"  "56696"  "56726"
## [218] "56914"  "56920"  "57200"  "57278"  "57590"  "58422"  "58631"
## [225] "59389"  "59449"  "59578"  "59938"  "60074"  "60146"  "60728"
## [232] "61110"  "61236"  "61265"  "61370"  "61424"  "61547"  "61650"
## [239] "61948"  "61959"  "62009"  "62273"  "62544"  "62605"  "62911"
## [246] "62949"  "63123"  "63251"  "63256"  "63814"  "64004"  "64740"
## [253] "64772"  "64773"  "65262"  "65285"  "65688"  "65815"  "65878"
## [260] "66075"  "66344"  "66764"  "66777"  "66868"  "66995"  "67087"
## [267] "67205"  "67465"  "67583"  "67752"  "67849"  "68112"  "69030"
## [274] "69045"  "69361"  "69625"  "69718"  "70070"  "70657"  "71033"
## [281] "71191"  "71590"  "71898"  "72223"  "72871"  "73135"  "73254"
## [288] "73256"  "73529"  "73585"  "73593"  "73666"  "73710"  "74043"
## [295] "74216"  "74784"  "75165"  "75448"  "75715"  "75730"  "76378"
## [302] "76595"  "77764"  "77948"  "77969"  "78118"  "78368"  "78598"
## [309] "78646"  "79148"  "79658"  "79861"  "80074"  "80093"  "80326"
## [316] "80599"  "80754"  "81034"  "81302"  "81813"  "82015"  "82045"
## [323] "82439"  "83356"  "83371"  "84398"  "84735"  "84843"  "85100"
## [330] "85254"  "85340"  "85766"  "86980"  "87246"  "87458"  "87900"
## [337] "89079"  "89243"  "89853"  "90225"  "90238"  "90794"  "91078"
## [344] "91083"  "91243"  "91511"  "91550"  "92387"  "92587"  "92598"
## [351] "93699"  "93809"  "94247"  "94305"  "94495"  "95014"  "95226"
## [358] "95241"  "95530"  "96119"  "96194"  "96298"  "96544"  "96980"
## [365] "97571"  "98401"  "98688"  "98906"  "98945"  "98956"  "98966"
## [372] "98981"  "99572"  "99893"  "99988"  "100096" "100198" "100571"
## [379] "100795" "101048" "101193" "101216" "101421" "101664" "101790"
## [386] "102224" "102368" "102932" "103000" "103682" "103858" "104389"
## [393] "104792" "105704" "106216" "106842" "106937" "107062" "107417"
## [400] "107932" "108185" "108201" "108206" "108304" "108334" "108479"
## [407] "108515" "108606" "108839" "108929" "109040" "109260" "109333"
## [414] "109734" "110047" "110199" "110565" "110673" "110873" "110913"
## [421] "110937" "111064" "111150" "111223" "111543" "112353" "112893"
## [428] "112901" "113083" "113597" "113753" "113964" "114436" "115174"
## [435] "115978" "116095" "116206" "116369" "116640" "117031" "117517"
## [442] "118785" "119035" "119554" "120535" "120802" "121010" "121184"
## [449] "121442" "121485" "121530" "122037" "122299" "122910" "123241"
## [456] "124259" "124620" "124715" "126173" "126470" "126481" "126587"
## [463] "126592" "126715" "127134" "127652" "127683" "127966" "128224"
## [470] "128390" "128587" "128926" "128937" "129264" "129479" "129660"
## [477] "129793" "129938" "129958" "129974" "130639" "131180" "131369"
## [484] "131482" "131592" "131736" "132350" "132383" "132433" "132534"
## [491] "132670" "132761" "133262" "133422" "133475" "134419" "135413"
```

```
## [498] "135495" "135800" "135935" "136039" "136229" "136265" "136309"
## [505] "136416" "136544" "136888" "137150" "137172" "137527" "138047"
## [512] "139374" "139883" "140233" "140567" "141027" "141534" "141835"
## [519] "141905" "141983" "142214" "142290" "142398" "142615" "142824"
## [526] "143558" "144638" "144756" "145268" "145938" "147177" "147246"
## [533] "147283" "147486" "147510" "147625" "147678" "148258" "148591"
## [540] "149155" "149218" "149842" "149887" "149898" "150136" "151342"
## [547] "151412" "151661" "151767" "151963" "152346" "152470" "152513"
## [554] "152725" "152828" "153560" "153673" "153796" "153878" "154253"
## [561] "154323" "154578" "154581" "154751" "155031" "155155" "155371"
## [568] "155384" "155441" "155551" "155566" "155950" "156244" "156250"
## [575] "156393" "156707" "156954" "157064" "157195" "157324" "158033"
## [582] "158490" "158962" "159223" "159480" "159725" "159831" "160040"
## [589] "160092" "160337" "160606" "160748" "161313" "161421" "161481"
## [596] "161512" "162070" "162131" "162954" "163032" "163237" "163255"
## [603] "163666" "163693" "164248" "164387" "165000" "165535" "165920"
## [610] "166238" "167445" "167876" "169257" "169467" "169526" "169710"
## [617] "169822" "169826" "169833" "170195" "170331" "170391" "171098"
## [624] "171131" "171189" "171559" "171702" "172037" "172208" "172245"
## [631] "172772" "172876" "172903" "173364" "173415" "173788" "173824"
## [638] "173854" "174720" "174998" "175544" "175695" "175809" "175857"
## [645] "176099" "176276" "176434" "176666" "177236" "177344" "177354"
## [652] "177374" "177448" "177526" "177870" "177916" "178368" "178672"
## [659] "179677" "179782" "180180" "180797" "180853" "181044" "181235"
## [666] "182265" "182609" "182666" "183517" "183591" "183684" "184157"
## [673] "185409" "185603" "186082" "186192" "186363" "186546" "186713"
## [680] "186748" "187874" "188151" "188163" "188195" "188514" "188904"
## [687] "189301" "189474" "189845" "190000" "190068" "190073" "190658"
## [694] "190709" "190844" "191673" "191838" "192112" "192276" "192478"
## [701] "192654" "193397" "193682" "193690" "194697" "195428" "195550"
## [708] "195748" "196375" "196873" "197836" "198562" "198780" "198895"
## [715] "199161" "199326" "199973" "200506" "200522" "201276" "201411"
## [722] "201548" "201646" "201970" "202639" "203359" "204377" "205293"
## [729] "205522" "205842" "205885" "206032" "206062" "206324" "206917"
## [736] "207260" "207320" "207328" "207378" "207802" "207851" "208109"
## [743] "208269" "208378" "208517" "209135" "209271" "209395" "209436"
## [750] "209639" "209705" "210246" "210669" "210976" "211159" "211184"
## [757] "211461" "211659" "211709" "211761" "211788" "211829" "211844"
## [764] "211869" "211981" "212174" "212803" "212885" "213114" "213363"
## [771] "214466" "214549" "214603" "214791" "215091" "215122" "215378"
## [778] "215943" "216301" "216346" "216844" "217437" "217502" "217888"
## [785] "218240" "218849" "218901" "219337" "219716" "219899" "219915"
## [792] "220165" "220482" "220771" "221147" "221841" "222117" "222326"
## [799] "222512" "222532" "223057" "223380" "223816" "223895" "223913"
## [806] "224258" "226368" "226489" "226956" "227217" "227913" "227943"
## [813] "229968" "230460" "231340" "231370" "231515" "231737" "232196"
## [820] "232252" "232483" "233616" "233632" "233723" "233970" "234369"
## [827] "234503" "234526" "234673" "234872" "235081" "235882" "236757"
## [834] "236865" "237072" "237279" "237409" "237435" "237814" "237965"
## [841] "238052" "238235" "239055" "239215" "239225" "239528" "239602"
## [848] "239736" "239876" "240252" "240675" "241255" "241276" "241388"
## [855] "241584" "242154" "242161" "242239" "242437" "243189" "243433"
## [862] "243661" "244399" "244494" "244739" "244877" "245087" "245281"
## [869] "246099" "246217" "246340" "246419" "246460" "246634" "246786"
```

```
##   [876] "246945" "247100" "247552" "247708" "247838" "247994" "248421"
##   [883] "249032" "249075" "249129" "249205" "250064" "250526" "250637"
##   [890] "251118" "251311" "251334" "251357" "251616" "251851" "251999"
##   [897] "252342" "253448" "253799" "253862" "254014" "254346" "254467"
##   [904] "254627" "254712" "254740" "254986" "255175" "255492" "256140"
##   [911] "256536" "256597" "256631" "256760" "256818" "257408" "258499"
##   [918] "259005" "260317" "260670" "260784" "261100" "261448" "261686"
##   [925] "261907" "262144" "262320" "262413" "262434" "262746" "262772"
##   [932] "263319" "263342" "263388" "263485" "264214" "264624" "264743"
##   [939] "264770" "264845" "265453" "265881" "266621" "266722" "266875"
##   [946] "267414" "267415" "267517" "267728" "267868" "268622" "269130"
##   [953] "269201" "269391" "269793" "270412" "270431" "270465" "270792"
##   [960] "270897" "270951" "271545" "271635" "272072" "272448" "272451"
##   [967] "272766" "272929" "273093" "274245" "274842" "274998" "275072"
##   [974] "275525" "275582" "276163" "276264" "276876" "277459" "277578"
##   [981] "277871" "278060" "278068" "279353" "279553" "279563" "280282"
##   [988] "280326" "280521" "280727" "280735" "281116" "281558" "281956"
##   [995] "282382" "283662" "283914" "283974" "284313" "284717"
##   [ reached getOption("max.print") -- omitted 3866 entries ]
##
## $dimnames[[2]]
##  [1] "Pickup_longitude"       "Pickup_latitude"
##  [3] "Dropoff_longitude"      "Dropoff_latitude"
##  [5] "Passenger_count"        "Trip_distance"
##  [7] "Fare_amount"            "Extra"
##  [9] "MTA_tax"                "Tip_amount"
## [11] "Tolls_amount"           "improvement_surcharge"
## [13] "trip_length"            "trip_distance_km"
## [15] "travel_time"            "pick_up_hour"
```

```r
summary(res.comp$completeObs)
```

```
##  Pickup_longitude Pickup_latitude Dropoff_longitude Dropoff_latitude
##  Min.   :-74.03   Min.   :40.58   Min.   :-74.03    Min.   :40.57
##  1st Qu.:-73.96   1st Qu.:40.69   1st Qu.:-73.97    1st Qu.:40.70
##  Median :-73.95   Median :40.75   Median :-73.95    Median :40.75
##  Mean   :-73.94   Mean   :40.75   Mean   :-73.94    Mean   :40.74
##  3rd Qu.:-73.92   3rd Qu.:40.80   3rd Qu.:-73.91    3rd Qu.:40.79
##  Max.   :-73.79   Max.   :40.91   Max.   :-73.75    Max.   :40.91
##  Passenger_count Trip_distance    Fare_amount        Extra
##  Min.   :1.000   Min.   : 0.010   Min.   : 0.10   Min.   :0.0000
##  1st Qu.:1.000   1st Qu.: 1.010   1st Qu.: 6.00   1st Qu.:0.0000
##  Median :1.000   Median : 1.800   Median : 9.00   Median :0.5000
##  Mean   :1.349   Mean   : 2.525   Mean   :11.15   Mean   :0.3497
##  3rd Qu.:1.000   3rd Qu.: 3.310   3rd Qu.:14.00   3rd Qu.:0.5000
##  Max.   :6.000   Max.   :11.085   Max.   :42.50   Max.   :2.0000
##     MTA_tax        Tip_amount      Tolls_amount
##  Min.   :0.0000   Min.   : 0.000   Min.   : 0.00000
##  1st Qu.:0.5000   1st Qu.: 0.000   1st Qu.: 0.00000
##  Median :0.5000   Median : 0.000   Median : 0.00000
##  Mean   :0.4915   Mean   : 1.124   Mean   : 0.07864
##  3rd Qu.:0.5000   3rd Qu.: 2.000   3rd Qu.: 0.00000
##  Max.   :0.5000   Max.   :22.000   Max.   :12.50000
##  improvement_surcharge trip_length     trip_distance_km
##  Min.   :0.000         Min.   : 0.000   Min.   : 0.01609
```

```
##  1st Qu.:0.300          1st Qu.: 1.792    1st Qu.: 1.62544
##  Median :0.300          Median : 3.132    Median : 2.89682
##  Mean   :0.295          Mean   : 4.306    Mean   : 4.06433
##  3rd Qu.:0.300          3rd Qu.: 5.541    3rd Qu.: 5.32693
##  Max.   :0.770          Max.   :29.880    Max.   :17.83972
##   travel_time           pick_up_hour
##  Min.   :   0.0167  Min.   : 0.00
##  1st Qu.:   5.9000  1st Qu.: 9.00
##  Median :   9.7500  Median :15.00
##  Mean   :  18.9308  Mean   :13.48
##  3rd Qu.:  15.7500  3rd Qu.:19.00
##  Max.   :1438.3167  Max.   :23.00
```

```r
df[,"Pickup_longitude"]<-res.comp$completeObs[,"Pickup_longitude"]
df[,"Pickup_latitude"]<-res.comp$completeObs[,"Pickup_latitude"]
df[,"Dropoff_longitude"]<-res.comp$completeObs[,"Dropoff_longitude"]
df[,"Dropoff_latitude"]<-res.comp$completeObs[,"Dropoff_latitude"]
df[,"Trip_distance"]<-res.comp$completeObs[,"Trip_distance"]
df[,"Passenger_count"]<-res.comp$completeObs[,"Passenger_count"]
df[,"Fare_amount"]<-res.comp$completeObs[,"Fare_amount"]
df[,"Extra"]<-res.comp$completeObs[,"Extra"]
df[,"travel_time"]<- res.comp$completeObs[,"travel_time"]
df[,"MTA_tax"]<-res.comp$completeObs[,"MTA_tax"]
df[,"Tip_amount"]<-res.comp$completeObs[,"Tip_amount"]
df[,"Tolls_amount"]<-res.comp$completeObs[,"Tolls_amount"]
df[,"improvement_surcharge"]<-res.comp$completeObs[,"improvement_surcharge"]
```

Espeed (km/h)

```r
#efective speed : trigonometric distance between pickup point and dropoff point divided by travel time

summary(df$travel_time)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.       Max.
##    0.0167    5.9000    9.7500   18.9308   15.7500  1438.3167
```

```r
for (i in 1:nrow(df)){
  df$espeed[i] <- df$trip_length[i]/(df$travel_time[i]/60)
}
names(df)
```

```
##  [1] "VendorID"              "lpep_pickup_datetime"
##  [3] "Lpep_dropoff_datetime" "Store_and_fwd_flag"
##  [5] "RateCodeID"            "Pickup_longitude"
##  [7] "Pickup_latitude"       "Dropoff_longitude"
##  [9] "Dropoff_latitude"      "Passenger_count"
## [11] "Trip_distance"         "Fare_amount"
## [13] "Extra"                 "MTA_tax"
## [15] "Tip_amount"            "Tolls_amount"
## [17] "improvement_surcharge" "Total_amount"
## [19] "Payment_type"          "Trip_type"
## [21] "mis_ind"               "AnyTip"
## [23] "trip_length"           "trip_distance_km"
## [25] "travel_time"           "pick_up_hour"
```

```
## [27] "pick_up_period"         "espeed"
```
```r
vars_con<-names(df)[c(6,7,8,9,10,11,12,13,14,15,16,17,23,24,25,26,28)]
vars_con
```
```
##  [1] "Pickup_longitude"      "Pickup_latitude"
##  [3] "Dropoff_longitude"     "Dropoff_latitude"
##  [5] "Passenger_count"       "Trip_distance"
##  [7] "Fare_amount"           "Extra"
##  [9] "MTA_tax"               "Tip_amount"
## [11] "Tolls_amount"          "improvement_surcharge"
## [13] "trip_length"           "trip_distance_km"
## [15] "travel_time"           "pick_up_hour"
## [17] "espeed"
```
```r
res.comp <- imputePCA(df[,vars_con], ncp=4)
attributes(res.comp$completeObs)
```
```
## $dim
## [1] 4866    17
##
## $dimnames
## $dimnames[[1]]
##     [1] "285"    "307"    "401"    "593"    "636"    "886"    "904"
##     [8] "978"    "1135"   "1282"   "1409"   "1475"   "1495"   "1905"
##    [15] "2126"   "2151"   "2201"   "2271"   "2747"   "3065"   "3089"
##    [22] "3130"   "3221"   "3420"   "3679"   "4310"   "4754"   "5241"
##    [29] "5277"   "5649"   "6353"   "6364"   "6755"   "6869"   "7079"
##    [36] "7211"   "7342"   "7802"   "8138"   "8443"   "8619"   "8891"
##    [43] "8960"   "9207"   "9503"   "9747"   "9765"   "9984"   "10034"
##    [50] "10199"  "10951"  "10955"  "10974"  "11189"  "11506"  "11713"
##    [57] "12492"  "12792"  "13043"  "13274"  "13332"  "13875"  "13927"
##    [64] "14874"  "14916"  "15407"  "15830"  "16080"  "16166"  "16345"
##    [71] "16391"  "17136"  "17355"  "18278"  "18596"  "18734"  "19101"
##    [78] "19344"  "19408"  "19991"  "20004"  "20009"  "20044"  "20077"
##    [85] "20271"  "20342"  "20361"  "20543"  "20621"  "20733"  "20917"
##    [92] "21425"  "21439"  "21539"  "21559"  "21735"  "22197"  "22332"
##    [99] "22825"  "22946"  "23091"  "23132"  "23811"  "24338"  "24863"
##   [106] "25262"  "25356"  "26062"  "26832"  "27216"  "27482"  "27495"
##   [113] "27594"  "27984"  "28083"  "28512"  "29375"  "29522"  "30659"
##   [120] "30856"  "31236"  "31456"  "31571"  "31583"  "31617"  "31726"
##   [127] "32873"  "32952"  "33882"  "34250"  "34280"  "34374"  "34390"
##   [134] "34922"  "35039"  "35207"  "35386"  "36076"  "36428"  "36540"
##   [141] "36696"  "36863"  "36933"  "37035"  "37273"  "37506"  "37517"
##   [148] "37561"  "37764"  "37821"  "37877"  "38445"  "38480"  "39213"
##   [155] "39623"  "39723"  "39943"  "40226"  "40245"  "40497"  "40560"
##   [162] "40802"  "40941"  "40943"  "40953"  "40969"  "42048"  "42779"
##   [169] "43577"  "43958"  "44992"  "46311"  "46572"  "46653"  "46790"
##   [176] "47428"  "47471"  "48166"  "48518"  "48796"  "48903"  "48915"
##   [183] "49242"  "49244"  "49383"  "49421"  "49783"  "49849"  "50027"
##   [190] "50328"  "50542"  "50979"  "50996"  "51868"  "51965"  "52728"
##   [197] "52825"  "52931"  "53452"  "53536"  "53680"  "54025"  "54342"
##   [204] "54359"  "54794"  "54843"  "54958"  "54994"  "55030"  "55082"
##   [211] "55144"  "55353"  "55479"  "55718"  "56090"  "56696"  "56726"
##   [218] "56914"  "56920"  "57200"  "57278"  "57590"  "58422"  "58631"
```

```
## [225] "59389"  "59449"  "59578"  "59938"  "60074"  "60146"  "60728"
## [232] "61110"  "61236"  "61265"  "61370"  "61424"  "61547"  "61650"
## [239] "61948"  "61959"  "62009"  "62273"  "62544"  "62605"  "62911"
## [246] "62949"  "63123"  "63251"  "63256"  "63814"  "64004"  "64740"
## [253] "64772"  "64773"  "65262"  "65285"  "65688"  "65815"  "65878"
## [260] "66075"  "66344"  "66764"  "66777"  "66868"  "66995"  "67087"
## [267] "67205"  "67465"  "67583"  "67752"  "67849"  "68112"  "69030"
## [274] "69045"  "69361"  "69625"  "69718"  "70070"  "70657"  "71033"
## [281] "71191"  "71590"  "71898"  "72223"  "72871"  "73135"  "73254"
## [288] "73256"  "73529"  "73585"  "73593"  "73666"  "73710"  "74043"
## [295] "74216"  "74784"  "75165"  "75448"  "75715"  "75730"  "76378"
## [302] "76595"  "77764"  "77948"  "77969"  "78118"  "78368"  "78598"
## [309] "78646"  "79148"  "79658"  "79861"  "80074"  "80093"  "80326"
## [316] "80599"  "80754"  "81034"  "81302"  "81813"  "82015"  "82045"
## [323] "82439"  "83356"  "83371"  "84398"  "84735"  "84843"  "85100"
## [330] "85254"  "85340"  "85766"  "86980"  "87246"  "87458"  "87900"
## [337] "89079"  "89243"  "89853"  "90225"  "90238"  "90794"  "91078"
## [344] "91083"  "91243"  "91511"  "91550"  "92387"  "92587"  "92598"
## [351] "93699"  "93809"  "94247"  "94305"  "94495"  "95014"  "95226"
## [358] "95241"  "95530"  "96119"  "96194"  "96298"  "96544"  "96980"
## [365] "97571"  "98401"  "98688"  "98906"  "98945"  "98956"  "98966"
## [372] "98981"  "99572"  "99893"  "99988"  "100096" "100198" "100571"
## [379] "100795" "101048" "101193" "101216" "101421" "101664" "101790"
## [386] "102224" "102368" "102932" "103000" "103682" "103858" "104389"
## [393] "104792" "105704" "106216" "106842" "106937" "107062" "107417"
## [400] "107932" "108185" "108201" "108206" "108304" "108334" "108479"
## [407] "108515" "108606" "108839" "108929" "109040" "109260" "109333"
## [414] "109734" "110047" "110199" "110565" "110673" "110873" "110913"
## [421] "110937" "111064" "111150" "111223" "111543" "112353" "112893"
## [428] "112901" "113083" "113597" "113753" "113964" "114436" "115174"
## [435] "115978" "116095" "116206" "116369" "116640" "117031" "117517"
## [442] "118785" "119035" "119554" "120535" "120802" "121010" "121184"
## [449] "121442" "121485" "121530" "122037" "122299" "122910" "123241"
## [456] "124259" "124620" "124715" "126173" "126470" "126481" "126587"
## [463] "126592" "126715" "127134" "127652" "127683" "127966" "128224"
## [470] "128390" "128587" "128926" "128937" "129264" "129479" "129660"
## [477] "129793" "129938" "129958" "129974" "130639" "131180" "131369"
## [484] "131482" "131592" "131736" "132350" "132383" "132433" "132534"
## [491] "132670" "132761" "133262" "133422" "133475" "134419" "135413"
## [498] "135495" "135800" "135935" "136039" "136229" "136265" "136309"
## [505] "136416" "136544" "136888" "137150" "137172" "137527" "138047"
## [512] "139374" "139883" "140233" "140567" "141027" "141534" "141835"
## [519] "141905" "141983" "142214" "142290" "142398" "142615" "142824"
## [526] "143558" "144638" "144756" "145268" "145938" "147177" "147246"
## [533] "147283" "147486" "147510" "147625" "147678" "148258" "148591"
## [540] "149155" "149218" "149842" "149887" "149898" "150136" "151342"
## [547] "151412" "151661" "151767" "151963" "152346" "152470" "152513"
## [554] "152725" "152828" "153560" "153673" "153796" "153878" "154253"
## [561] "154323" "154578" "154581" "154751" "155031" "155155" "155371"
## [568] "155384" "155441" "155551" "155566" "155950" "156244" "156250"
## [575] "156393" "156707" "156954" "157064" "157195" "157324" "158033"
## [582] "158490" "158962" "159223" "159480" "159725" "159831" "160040"
## [589] "160092" "160337" "160606" "160748" "161313" "161421" "161481"
## [596] "161512" "162070" "162131" "162954" "163032" "163237" "163255"
```

```
## [603] "163666" "163693" "164248" "164387" "165000" "165535" "165920"
## [610] "166238" "167445" "167876" "169257" "169467" "169526" "169710"
## [617] "169822" "169826" "169833" "170195" "170331" "170391" "171098"
## [624] "171131" "171189" "171559" "171702" "172037" "172208" "172245"
## [631] "172772" "172876" "172903" "173364" "173415" "173788" "173824"
## [638] "173854" "174720" "174998" "175544" "175695" "175809" "175857"
## [645] "176099" "176276" "176434" "176666" "177236" "177344" "177354"
## [652] "177374" "177448" "177526" "177870" "177916" "178368" "178672"
## [659] "179677" "179782" "180180" "180797" "180853" "181044" "181235"
## [666] "182265" "182609" "182666" "183517" "183591" "183684" "184157"
## [673] "185409" "185603" "186082" "186192" "186363" "186546" "186713"
## [680] "186748" "187874" "188151" "188163" "188195" "188514" "188904"
## [687] "189301" "189474" "189845" "190000" "190068" "190073" "190658"
## [694] "190709" "190844" "191673" "191838" "192112" "192276" "192478"
## [701] "192654" "193397" "193682" "193690" "194697" "195428" "195550"
## [708] "195748" "196375" "196873" "197836" "198562" "198780" "198895"
## [715] "199161" "199326" "199973" "200506" "200522" "201276" "201411"
## [722] "201548" "201646" "201970" "202639" "203359" "204377" "205293"
## [729] "205522" "205842" "205885" "206032" "206062" "206324" "206917"
## [736] "207260" "207320" "207328" "207378" "207802" "207851" "208109"
## [743] "208269" "208378" "208517" "209135" "209271" "209395" "209436"
## [750] "209639" "209705" "210246" "210669" "210976" "211159" "211184"
## [757] "211461" "211659" "211709" "211761" "211788" "211829" "211844"
## [764] "211869" "211981" "212174" "212803" "212885" "213114" "213363"
## [771] "214466" "214549" "214603" "214791" "215091" "215122" "215378"
## [778] "215943" "216301" "216346" "216844" "217437" "217502" "217888"
## [785] "218240" "218849" "218901" "219337" "219716" "219899" "219915"
## [792] "220165" "220482" "220771" "221147" "221841" "222117" "222326"
## [799] "222512" "222532" "223057" "223380" "223816" "223895" "223913"
## [806] "224258" "226368" "226489" "226956" "227217" "227913" "227943"
## [813] "229968" "230460" "231340" "231370" "231515" "231737" "232196"
## [820] "232252" "232483" "233616" "233632" "233723" "233970" "234369"
## [827] "234503" "234526" "234673" "234872" "235081" "235882" "236757"
## [834] "236865" "237072" "237279" "237409" "237435" "237814" "237965"
## [841] "238052" "238235" "239055" "239215" "239225" "239528" "239602"
## [848] "239736" "239876" "240252" "240675" "241255" "241276" "241388"
## [855] "241584" "242154" "242161" "242239" "242437" "243189" "243433"
## [862] "243661" "244399" "244494" "244739" "244877" "245087" "245281"
## [869] "246099" "246217" "246340" "246419" "246460" "246634" "246786"
## [876] "246945" "247100" "247552" "247708" "247838" "247994" "248421"
## [883] "249032" "249075" "249129" "249205" "250064" "250526" "250637"
## [890] "251118" "251311" "251334" "251357" "251616" "251851" "251999"
## [897] "252342" "253448" "253799" "253862" "254014" "254346" "254467"
## [904] "254627" "254712" "254740" "254986" "255175" "255492" "256140"
## [911] "256536" "256597" "256631" "256760" "256818" "257408" "258499"
## [918] "259005" "260317" "260670" "260784" "261100" "261448" "261686"
## [925] "261907" "262144" "262320" "262413" "262434" "262746" "262772"
## [932] "263319" "263342" "263388" "263485" "264214" "264624" "264743"
## [939] "264770" "264845" "265453" "265881" "266621" "266722" "266875"
## [946] "267414" "267415" "267517" "267728" "267868" "268622" "269130"
## [953] "269201" "269391" "269793" "270412" "270431" "270465" "270792"
## [960] "270897" "270951" "271545" "271635" "272072" "272448" "272451"
## [967] "272766" "272929" "273093" "274245" "274842" "274998" "275072"
## [974] "275525" "275582" "276163" "276264" "276876" "277459" "277578"
```

```
##    [981] "277871" "278060" "278068" "279353" "279553" "279563" "280282"
##    [988] "280326" "280521" "280727" "280735" "281116" "281558" "281956"
##    [995] "282382" "283662" "283914" "283974" "284313" "284717"
##    [ reached getOption("max.print") -- omitted 3866 entries ]
##
## $dimnames[[2]]
##  [1] "Pickup_longitude"     "Pickup_latitude"
##  [3] "Dropoff_longitude"    "Dropoff_latitude"
##  [5] "Passenger_count"      "Trip_distance"
##  [7] "Fare_amount"          "Extra"
##  [9] "MTA_tax"              "Tip_amount"
## [11] "Tolls_amount"         "improvement_surcharge"
## [13] "trip_length"          "trip_distance_km"
## [15] "travel_time"          "pick_up_hour"
## [17] "espeed"
```

```r
summary(res.comp$completeObs)
```

```
##   Pickup_longitude Pickup_latitude Dropoff_longitude Dropoff_latitude
##   Min.   :-74.03   Min.   :40.58   Min.   :-74.03    Min.   :40.57
##   1st Qu.:-73.96   1st Qu.:40.69   1st Qu.:-73.97    1st Qu.:40.70
##   Median :-73.95   Median :40.75   Median :-73.95    Median :40.75
##   Mean   :-73.94   Mean   :40.75   Mean   :-73.94    Mean   :40.74
##   3rd Qu.:-73.92   3rd Qu.:40.80   3rd Qu.:-73.91    3rd Qu.:40.79
##   Max.   :-73.79   Max.   :40.91   Max.   :-73.75    Max.   :40.91
##   Passenger_count Trip_distance     Fare_amount        Extra
##   Min.   :1.000   Min.   : 0.010   Min.   : 0.10   Min.   :0.0000
##   1st Qu.:1.000   1st Qu.: 1.010   1st Qu.: 6.00   1st Qu.:0.0000
##   Median :1.000   Median : 1.800   Median : 9.00   Median :0.5000
##   Mean   :1.349   Mean   : 2.525   Mean   :11.15   Mean   :0.3497
##   3rd Qu.:1.000   3rd Qu.: 3.310   3rd Qu.:14.00   3rd Qu.:0.5000
##   Max.   :6.000   Max.   :11.085   Max.   :42.50   Max.   :2.0000
##      MTA_tax        Tip_amount       Tolls_amount
##   Min.   :0.0000   Min.   : 0.000   Min.   : 0.00000
##   1st Qu.:0.5000   1st Qu.: 0.000   1st Qu.: 0.00000
##   Median :0.5000   Median : 0.000   Median : 0.00000
##   Mean   :0.4915   Mean   : 1.124   Mean   : 0.07864
##   3rd Qu.:0.5000   3rd Qu.: 2.000   3rd Qu.: 0.00000
##   Max.   :0.5000   Max.   :22.000   Max.   :12.50000
##   improvement_surcharge  trip_length     trip_distance_km
##   Min.   :0.000          Min.   : 0.000   Min.   : 0.01609
##   1st Qu.:0.300          1st Qu.: 1.792   1st Qu.: 1.62544
##   Median :0.300          Median : 3.132   Median : 2.89682
##   Mean   :0.295          Mean   : 4.307   Mean   : 4.06209
##   3rd Qu.:0.300          3rd Qu.: 5.543   3rd Qu.: 5.32693
##   Max.   :0.770          Max.   :29.880   Max.   :18.12329
##    travel_time        pick_up_hour       espeed
##   Min.   :   0.0167   Min.   : 0.00   Min.   :  0.00
##   1st Qu.:   5.9000   1st Qu.: 9.00   1st Qu.: 15.26
##   Median :   9.7500   Median :15.00   Median : 20.36
##   Mean   :  18.9308   Mean   :13.48   Mean   : 22.46
##   3rd Qu.:  15.7500   3rd Qu.:19.00   3rd Qu.: 27.14
##   Max.   :1438.3167   Max.   :23.00   Max.   :678.00
```

```r
df[,"espeed"]<-res.comp$completeObs[,"espeed"]
```

```r
summary(df$espeed)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   15.26   20.36   22.46   27.14  678.00
```

```r
boxplot(df$espeed)
```
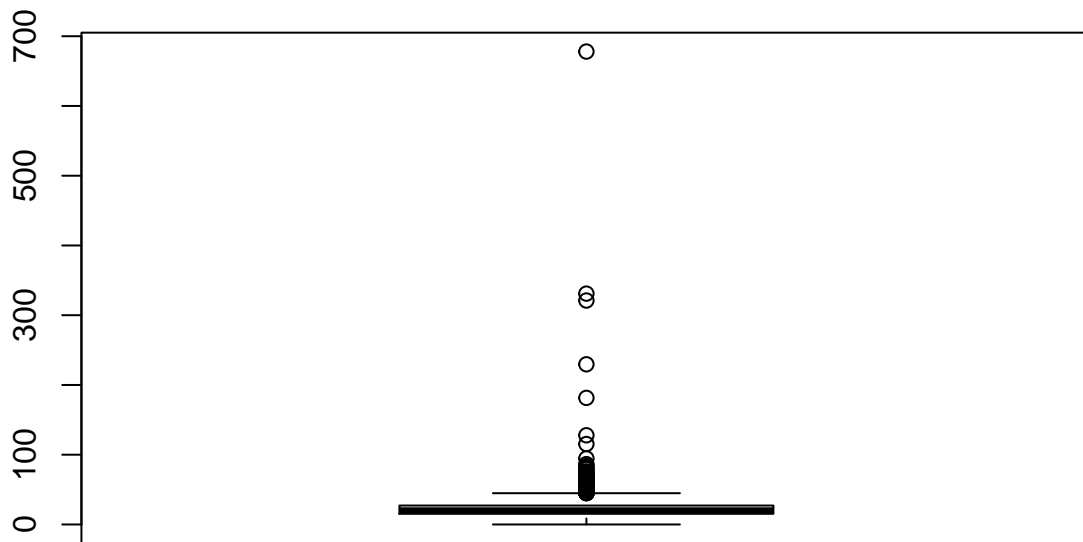

Imputation of qualitative variables

```r
library(missMDA)
vars_dis_qual <-vars_dis[c(1,2,3,5,6)]
summary(df$Passenger_count)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.349   1.000   6.000
```

```r
res.impute = imputeMCA(df[,vars_dis_qual],ncp = 3)
res.impute
```

```
## $tab.disj
##         Creative Mobile Technologies, LLC VeriFone Inc.
## 285                                     0             1
## 307                                     0             1
## 401                                     0             1
```

48

```
## 593                                    0                    1
## 636                                    0                    1
## 886                                    0                    1
## 904                                    0                    1
## 978                                    0                    1
## 1135                                   0                    1
## 1282                                   0                    1
## 1409                                   0                    1
## 1475                                   0                    1
## 1495                                   0                    1
## 1905                                   0                    1
## 2126                                   0                    1
## 2151                                   0                    1
## 2201                                   0                    1
## 2271                                   0                    1
## 2747                                   0                    1
## 3065                                   0                    1
## 3089                                   0                    1
## 3130                                   0                    1
## 3221                                   0                    1
## 3420                                   0                    1
## 3679                                   0                    1
## 4310                                   0                    1
## 4754                                   0                    1
## 5241                                   1                    0
## 5277                                   1                    0
## 5649                                   1                    0
## 6353                                   1                    0
## 6364                                   1                    0
## 6755                                   0                    1
## 6869                                   0                    1
## 7079                                   0                    1
## 7211                                   0                    1
## 7342                                   0                    1
## 7802                                   0                    1
## 8138                                   0                    1
## 8443                                   0                    1
## 8619                                   0                    1
## 8891                                   0                    1
## 8960                                   0                    1
## 9207                                   0                    1
## 9503                                   0                    1
## 9747                                   0                    1
## 9765                                   0                    1
## 9984                                   0                    1
## 10034                                  0                    1
## 10199                                  0                    1
## 10951                                  0                    1
## 10955                                  0                    1
## 10974                                  0                    1
## 11189                                  0                    1
## 11506                                  0                    1
## 11713                                  0                    1
## 12492                                  1                    0
```

```
## 12792                                              1            0
## 13043                                              1            0
## 13274                                              1            0
## 13332                                              0            1
## 13875                                              0            1
## 13927                                              0            1
## 14874                                              0            1
## 14916                                              0            1
## 15407                                              0            1
## 15830                                              0            1
## 16080                                              0            1
## 16166                                              0            1
## 16345                                              0            1
## 16391                                              0            1
##          not a store and forward trip store and forward trip Standard rate
## 285                                1                        0             1
## 307                                1                        0             1
## 401                                1                        0             1
## 593                                1                        0             1
## 636                                1                        0             1
## 886                                1                        0             1
## 904                                1                        0             1
## 978                                1                        0             1
## 1135                               1                        0             1
## 1282                               1                        0             1
## 1409                               1                        0             1
## 1475                               1                        0             1
## 1495                               1                        0             1
## 1905                               1                        0             1
## 2126                               1                        0             1
## 2151                               1                        0             1
## 2201                               1                        0             1
## 2271                               1                        0             1
## 2747                               1                        0             1
## 3065                               1                        0             1
## 3089                               1                        0             1
## 3130                               1                        0             1
## 3221                               1                        0             1
## 3420                               1                        0             1
## 3679                               1                        0             1
## 4310                               1                        0             1
## 4754                               1                        0             1
## 5241                               1                        0             1
## 5277                               1                        0             1
## 5649                               1                        0             1
## 6353                               1                        0             1
## 6364                               1                        0             1
## 6755                               1                        0             1
## 6869                               1                        0             1
## 7079                               1                        0             1
## 7211                               1                        0             1
## 7342                               1                        0             1
## 7802                               1                        0             1
## 8138                               1                        0             1
```

```
## 8443                                                       1                         0                    1
## 8619                                                       1                         0                    1
## 8891                                                       1                         0                    1
## 8960                                                       1                         0                    1
## 9207                                                       1                         0                    1
## 9503                                                       1                         0                    1
## 9747                                                       1                         0                    1
## 9765                                                       1                         0                    0
## 9984                                                       1                         0                    1
## 10034                                                      1                         0                    1
## 10199                                                      1                         0                    1
## 10951                                                      1                         0                    1
## 10955                                                      1                         0                    1
## 10974                                                      1                         0                    1
## 11189                                                      1                         0                    1
## 11506                                                      1                         0                    1
## 11713                                                      1                         0                    1
## 12492                                                      1                         0                    1
## 12792                                                      1                         0                    1
## 13043                                                      1                         0                    1
## 13274                                                      1                         0                    1
## 13332                                                      1                         0                    1
## 13875                                                      1                         0                    1
## 13927                                                      1                         0                    1
## 14874                                                      1                         0                    1
## 14916                                                      1                         0                    1
## 15407                                                      1                         0                    1
## 15830                                                      1                         0                    1
## 16080                                                      1                         0                    1
## 16166                                                      1                         0                    1
## 16345                                                      1                         0                    1
## 16391                                                      1                         0                    1
##           Newark Nassau or Westchester Negotiated fare Credit card Cash
## 285            0                     0               0           0    1
## 307            0                     0               0           0    1
## 401            0                     0               0           0    1
## 593            0                     0               0           0    1
## 636            0                     0               0           1    0
## 886            0                     0               0           1    0
## 904            0                     0               0           1    0
## 978            0                     0               0           0    1
## 1135           0                     0               0           0    1
## 1282           0                     0               0           1    0
## 1409           0                     0               0           0    1
## 1475           0                     0               0           1    0
## 1495           0                     0               0           1    0
## 1905           0                     0               0           0    1
## 2126           0                     0               0           0    1
## 2151           0                     0               0           0    1
## 2201           0                     0               0           1    0
## 2271           0                     0               0           1    0
## 2747           0                     0               0           0    1
## 3065           0                     0               0           1    0
## 3089           0                     0               0           0    1
```

```
## 3130           0                0              0        0  1
## 3221           0                0              0        0  1
## 3420           0                0              0        0  1
## 3679           0                0              0        0  1
## 4310           0                0              0        1  0
## 4754           0                0              0        1  0
## 5241           0                0              0        1  0
## 5277           0                0              0        0  1
## 5649           0                0              0        1  0
## 6353           0                0              0        0  1
## 6364           0                0              0        0  1
## 6755           0                0              0        0  1
## 6869           0                0              0        1  0
## 7079           0                0              0        0  1
## 7211           0                0              0        0  1
## 7342           0                0              0        0  1
## 7802           0                0              0        0  1
## 8138           0                0              0        1  0
## 8443           0                0              0        1  0
## 8619           0                0              0        0  1
## 8891           0                0              0        0  1
## 8960           0                0              0        1  0
## 9207           0                0              0        0  1
## 9503           0                0              0        0  1
## 9747           0                0              0        0  1
## 9765           0                0              1        1  0
## 9984           0                0              0        1  0
## 10034          0                0              0        1  0
## 10199          0                0              0        0  1
## 10951          0                0              0        0  1
## 10955          0                0              0        0  1
## 10974          0                0              0        1  0
## 11189          0                0              0        0  1
## 11506          0                0              0        0  1
## 11713          0                0              0        1  0
## 12492          0                0              0        0  1
## 12792          0                0              0        0  1
## 13043          0                0              0        0  1
## 13274          0                0              0        0  1
## 13332          0                0              0        1  0
## 13875          0                0              0        1  0
## 13927          0                0              0        0  1
## 14874          0                0              0        0  1
## 14916          0                0              0        0  1
## 15407          0                0              0        1  0
## 15830          0                0              0        0  1
## 16080          0                0              0        0  1
## 16166          0                0              0        1  0
## 16345          0                0              0        1  0
## 16391          0                0              0        1  0
##        No charge Dispute Street-hail Dispatch
## 285            0       0            1        0
## 307            0       0            1        0
## 401            0       0            1        0
```

```
## 593             0           0           1           0
## 636             0           0           1           0
## 886             0           0           1           0
## 904             0           0           1           0
## 978             0           0           1           0
## 1135            0           0           1           0
## 1282            0           0           1           0
## 1409            0           0           1           0
## 1475            0           0           1           0
## 1495            0           0           1           0
## 1905            0           0           1           0
## 2126            0           0           1           0
## 2151            0           0           1           0
## 2201            0           0           1           0
## 2271            0           0           1           0
## 2747            0           0           1           0
## 3065            0           0           1           0
## 3089            0           0           1           0
## 3130            0           0           1           0
## 3221            0           0           1           0
## 3420            0           0           1           0
## 3679            0           0           1           0
## 4310            0           0           1           0
## 4754            0           0           1           0
## 5241            0           0           1           0
## 5277            0           0           1           0
## 5649            0           0           1           0
## 6353            0           0           1           0
## 6364            0           0           1           0
## 6755            0           0           1           0
## 6869            0           0           1           0
## 7079            0           0           1           0
## 7211            0           0           1           0
## 7342            0           0           1           0
## 7802            0           0           1           0
## 8138            0           0           1           0
## 8443            0           0           1           0
## 8619            0           0           1           0
## 8891            0           0           1           0
## 8960            0           0           1           0
## 9207            0           0           1           0
## 9503            0           0           1           0
## 9747            0           0           1           0
## 9765            0           0           0           1
## 9984            0           0           1           0
## 10034           0           0           1           0
## 10199           0           0           1           0
## 10951           0           0           1           0
## 10955           0           0           1           0
## 10974           0           0           1           0
## 11189           0           0           1           0
## 11506           0           0           1           0
## 11713           0           0           1           0
## 12492           0           0           1           0
```

```
## 12792                  0         0         1         0
## 13043                  0         0         1         0
## 13274                  0         0         1         0
## 13332                  0         0         1         0
## 13875                  0         0         1         0
## 13927                  0         0         1         0
## 14874                  0         0         1         0
## 14916                  0         0         1         0
## 15407                  0         0         1         0
## 15830                  0         0         1         0
## 16080                  0         0         1         0
## 16166                  0         0         1         0
## 16345                  0         0         1         0
## 16391                  0         0         1         0
##  [ reached getOption("max.print") -- omitted 4795 rows ]
##
## $completeObs
##                                   VendorID        Store_and_fwd_flag
## 285                          VeriFone Inc. not a store and forward trip
## 307                          VeriFone Inc. not a store and forward trip
## 401                          VeriFone Inc. not a store and forward trip
## 593                          VeriFone Inc. not a store and forward trip
## 636                          VeriFone Inc. not a store and forward trip
## 886                          VeriFone Inc. not a store and forward trip
## 904                          VeriFone Inc. not a store and forward trip
## 978                          VeriFone Inc. not a store and forward trip
## 1135                         VeriFone Inc. not a store and forward trip
## 1282                         VeriFone Inc. not a store and forward trip
## 1409                         VeriFone Inc. not a store and forward trip
## 1475                         VeriFone Inc. not a store and forward trip
## 1495                         VeriFone Inc. not a store and forward trip
## 1905                         VeriFone Inc. not a store and forward trip
## 2126                         VeriFone Inc. not a store and forward trip
## 2151                         VeriFone Inc. not a store and forward trip
## 2201                         VeriFone Inc. not a store and forward trip
## 2271                         VeriFone Inc. not a store and forward trip
## 2747                         VeriFone Inc. not a store and forward trip
## 3065                         VeriFone Inc. not a store and forward trip
## 3089                         VeriFone Inc. not a store and forward trip
## 3130                         VeriFone Inc. not a store and forward trip
## 3221                         VeriFone Inc. not a store and forward trip
## 3420                         VeriFone Inc. not a store and forward trip
## 3679                         VeriFone Inc. not a store and forward trip
## 4310                         VeriFone Inc. not a store and forward trip
## 4754                         VeriFone Inc. not a store and forward trip
## 5241   Creative Mobile Technologies, LLC not a store and forward trip
## 5277   Creative Mobile Technologies, LLC not a store and forward trip
## 5649   Creative Mobile Technologies, LLC not a store and forward trip
## 6353   Creative Mobile Technologies, LLC not a store and forward trip
## 6364   Creative Mobile Technologies, LLC not a store and forward trip
## 6755                         VeriFone Inc. not a store and forward trip
## 6869                         VeriFone Inc. not a store and forward trip
## 7079                         VeriFone Inc. not a store and forward trip
## 7211                         VeriFone Inc. not a store and forward trip
```

```
## 7342                            VeriFone Inc. not a store and forward trip
## 7802                            VeriFone Inc. not a store and forward trip
## 8138                            VeriFone Inc. not a store and forward trip
## 8443                            VeriFone Inc. not a store and forward trip
## 8619                            VeriFone Inc. not a store and forward trip
## 8891                            VeriFone Inc. not a store and forward trip
## 8960                            VeriFone Inc. not a store and forward trip
## 9207                            VeriFone Inc. not a store and forward trip
## 9503                            VeriFone Inc. not a store and forward trip
## 9747                            VeriFone Inc. not a store and forward trip
## 9765                            VeriFone Inc. not a store and forward trip
## 9984                            VeriFone Inc. not a store and forward trip
## 10034                           VeriFone Inc. not a store and forward trip
## 10199                           VeriFone Inc. not a store and forward trip
## 10951                           VeriFone Inc. not a store and forward trip
## 10955                           VeriFone Inc. not a store and forward trip
## 10974                           VeriFone Inc. not a store and forward trip
## 11189                           VeriFone Inc. not a store and forward trip
## 11506                           VeriFone Inc. not a store and forward trip
## 11713                           VeriFone Inc. not a store and forward trip
## 12492   Creative Mobile Technologies, LLC not a store and forward trip
## 12792   Creative Mobile Technologies, LLC not a store and forward trip
## 13043   Creative Mobile Technologies, LLC not a store and forward trip
## 13274   Creative Mobile Technologies, LLC not a store and forward trip
## 13332                           VeriFone Inc. not a store and forward trip
## 13875                           VeriFone Inc. not a store and forward trip
## 13927                           VeriFone Inc. not a store and forward trip
## 14874                           VeriFone Inc. not a store and forward trip
## 14916                           VeriFone Inc. not a store and forward trip
## 15407                           VeriFone Inc. not a store and forward trip
## 15830                           VeriFone Inc. not a store and forward trip
## 16080                           VeriFone Inc. not a store and forward trip
## 16166                           VeriFone Inc. not a store and forward trip
## 16345                           VeriFone Inc. not a store and forward trip
## 16391                           VeriFone Inc. not a store and forward trip
## 17136                           VeriFone Inc. not a store and forward trip
## 17355                           VeriFone Inc. not a store and forward trip
## 18278                           VeriFone Inc. not a store and forward trip
## 18596   Creative Mobile Technologies, LLC not a store and forward trip
## 18734   Creative Mobile Technologies, LLC not a store and forward trip
## 19101   Creative Mobile Technologies, LLC     store and forward trip
## 19344   Creative Mobile Technologies, LLC not a store and forward trip
## 19408   Creative Mobile Technologies, LLC not a store and forward trip
## 19991                           VeriFone Inc. not a store and forward trip
## 20004                           VeriFone Inc. not a store and forward trip
## 20009                           VeriFone Inc. not a store and forward trip
## 20044                           VeriFone Inc. not a store and forward trip
## 20077                           VeriFone Inc. not a store and forward trip
## 20271                           VeriFone Inc. not a store and forward trip
## 20342                           VeriFone Inc. not a store and forward trip
## 20361                           VeriFone Inc. not a store and forward trip
## 20543                           VeriFone Inc. not a store and forward trip
## 20621                           VeriFone Inc. not a store and forward trip
## 20733                           VeriFone Inc. not a store and forward trip
```

```
## 20917                        VeriFone Inc. not a store and forward trip
## 21425                        VeriFone Inc. not a store and forward trip
## 21439                        VeriFone Inc. not a store and forward trip
## 21539                        VeriFone Inc. not a store and forward trip
## 21559                        VeriFone Inc. not a store and forward trip
## 21735                        VeriFone Inc. not a store and forward trip
## 22197                        VeriFone Inc. not a store and forward trip
## 22332                        VeriFone Inc. not a store and forward trip
## 22825                        VeriFone Inc. not a store and forward trip
## 22946                        VeriFone Inc. not a store and forward trip
## 23091                        VeriFone Inc. not a store and forward trip
## 23132                        VeriFone Inc. not a store and forward trip
## 23811                        VeriFone Inc. not a store and forward trip
## 24338 Creative Mobile Technologies, LLC not a store and forward trip
## 24863 Creative Mobile Technologies, LLC not a store and forward trip
## 25262 Creative Mobile Technologies, LLC not a store and forward trip
## 25356                        VeriFone Inc. not a store and forward trip
## 26062                        VeriFone Inc. not a store and forward trip
## 26832                        VeriFone Inc. not a store and forward trip
## 27216                        VeriFone Inc. not a store and forward trip
## 27482                        VeriFone Inc. not a store and forward trip
## 27495                        VeriFone Inc. not a store and forward trip
## 27594                        VeriFone Inc. not a store and forward trip
## 27984                        VeriFone Inc. not a store and forward trip
## 28083                        VeriFone Inc. not a store and forward trip
## 28512                        VeriFone Inc. not a store and forward trip
## 29375 Creative Mobile Technologies, LLC not a store and forward trip
## 29522 Creative Mobile Technologies, LLC not a store and forward trip
## 30659                        VeriFone Inc. not a store and forward trip
## 30856                        VeriFone Inc. not a store and forward trip
## 31236                        VeriFone Inc. not a store and forward trip
## 31456                        VeriFone Inc. not a store and forward trip
## 31571                        VeriFone Inc. not a store and forward trip
## 31583                        VeriFone Inc. not a store and forward trip
## 31617                        VeriFone Inc. not a store and forward trip
## 31726                        VeriFone Inc. not a store and forward trip
## 32873 Creative Mobile Technologies, LLC not a store and forward trip
## 32952 Creative Mobile Technologies, LLC not a store and forward trip
## 33882                        VeriFone Inc. not a store and forward trip
## 34250                        VeriFone Inc. not a store and forward trip
## 34280                        VeriFone Inc. not a store and forward trip
## 34374                        VeriFone Inc. not a store and forward trip
## 34390                        VeriFone Inc. not a store and forward trip
## 34922                        VeriFone Inc. not a store and forward trip
## 35039                        VeriFone Inc. not a store and forward trip
## 35207                        VeriFone Inc. not a store and forward trip
## 35386                        VeriFone Inc. not a store and forward trip
## 36076 Creative Mobile Technologies, LLC not a store and forward trip
## 36428                        VeriFone Inc. not a store and forward trip
## 36540                        VeriFone Inc. not a store and forward trip
## 36696                        VeriFone Inc. not a store and forward trip
## 36863                        VeriFone Inc. not a store and forward trip
## 36933 Creative Mobile Technologies, LLC not a store and forward trip
## 37035 Creative Mobile Technologies, LLC not a store and forward trip
```

```
## 37273                           VeriFone Inc. not a store and forward trip
## 37506                           VeriFone Inc. not a store and forward trip
## 37517                           VeriFone Inc. not a store and forward trip
## 37561                           VeriFone Inc. not a store and forward trip
## 37764 Creative Mobile Technologies, LLC not a store and forward trip
## 37821 Creative Mobile Technologies, LLC not a store and forward trip
## 37877 Creative Mobile Technologies, LLC not a store and forward trip
## 38445                           VeriFone Inc. not a store and forward trip
## 38480                           VeriFone Inc. not a store and forward trip
## 39213                           VeriFone Inc. not a store and forward trip
## 39623 Creative Mobile Technologies, LLC not a store and forward trip
## 39723 Creative Mobile Technologies, LLC not a store and forward trip
## 39943                           VeriFone Inc. not a store and forward trip
## 40226                           VeriFone Inc. not a store and forward trip
## 40245                           VeriFone Inc. not a store and forward trip
## 40497                           VeriFone Inc. not a store and forward trip
## 40560 Creative Mobile Technologies, LLC         store and forward trip
## 40802 Creative Mobile Technologies, LLC not a store and forward trip
## 40941                           VeriFone Inc. not a store and forward trip
## 40943                           VeriFone Inc. not a store and forward trip
## 40953                           VeriFone Inc. not a store and forward trip
## 40969                           VeriFone Inc. not a store and forward trip
## 42048 Creative Mobile Technologies, LLC not a store and forward trip
## 42779                           VeriFone Inc. not a store and forward trip
## 43577 Creative Mobile Technologies, LLC not a store and forward trip
## 43958                           VeriFone Inc. not a store and forward trip
## 44992                           VeriFone Inc. not a store and forward trip
## 46311                           VeriFone Inc. not a store and forward trip
## 46572                           VeriFone Inc. not a store and forward trip
## 46653                           VeriFone Inc. not a store and forward trip
## 46790                           VeriFone Inc. not a store and forward trip
## 47428 Creative Mobile Technologies, LLC not a store and forward trip
## 47471 Creative Mobile Technologies, LLC not a store and forward trip
## 48166                           VeriFone Inc. not a store and forward trip
## 48518                           VeriFone Inc. not a store and forward trip
## 48796                           VeriFone Inc. not a store and forward trip
## 48903                           VeriFone Inc. not a store and forward trip
## 48915                           VeriFone Inc. not a store and forward trip
## 49242                           VeriFone Inc. not a store and forward trip
## 49244                           VeriFone Inc. not a store and forward trip
## 49383                           VeriFone Inc. not a store and forward trip
## 49421                           VeriFone Inc. not a store and forward trip
## 49783 Creative Mobile Technologies, LLC not a store and forward trip
## 49849 Creative Mobile Technologies, LLC not a store and forward trip
## 50027                           VeriFone Inc. not a store and forward trip
## 50328                           VeriFone Inc. not a store and forward trip
## 50542                           VeriFone Inc. not a store and forward trip
## 50979                           VeriFone Inc. not a store and forward trip
## 50996                           VeriFone Inc. not a store and forward trip
## 51868                           VeriFone Inc. not a store and forward trip
## 51965 Creative Mobile Technologies, LLC not a store and forward trip
## 52728                           VeriFone Inc. not a store and forward trip
## 52825                           VeriFone Inc. not a store and forward trip
## 52931                           VeriFone Inc. not a store and forward trip
```

```
## 53452                           VeriFone Inc. not a store and forward trip
## 53536                           VeriFone Inc. not a store and forward trip
##                   RateCodeID Payment_type   Trip_type
## 285            Standard rate         Cash Street-hail
## 307            Standard rate         Cash Street-hail
## 401            Standard rate         Cash Street-hail
## 593            Standard rate         Cash Street-hail
## 636            Standard rate  Credit card Street-hail
## 886            Standard rate  Credit card Street-hail
## 904            Standard rate  Credit card Street-hail
## 978            Standard rate         Cash Street-hail
## 1135           Standard rate         Cash Street-hail
## 1282           Standard rate  Credit card Street-hail
## 1409           Standard rate         Cash Street-hail
## 1475           Standard rate  Credit card Street-hail
## 1495           Standard rate  Credit card Street-hail
## 1905           Standard rate         Cash Street-hail
## 2126           Standard rate         Cash Street-hail
## 2151           Standard rate         Cash Street-hail
## 2201           Standard rate  Credit card Street-hail
## 2271           Standard rate  Credit card Street-hail
## 2747           Standard rate         Cash Street-hail
## 3065           Standard rate  Credit card Street-hail
## 3089           Standard rate         Cash Street-hail
## 3130           Standard rate         Cash Street-hail
## 3221           Standard rate         Cash Street-hail
## 3420           Standard rate         Cash Street-hail
## 3679           Standard rate         Cash Street-hail
## 4310           Standard rate  Credit card Street-hail
## 4754           Standard rate  Credit card Street-hail
## 5241           Standard rate  Credit card Street-hail
## 5277           Standard rate         Cash Street-hail
## 5649           Standard rate  Credit card Street-hail
## 6353           Standard rate         Cash Street-hail
## 6364           Standard rate         Cash Street-hail
## 6755           Standard rate         Cash Street-hail
## 6869           Standard rate  Credit card Street-hail
## 7079           Standard rate         Cash Street-hail
## 7211           Standard rate         Cash Street-hail
## 7342           Standard rate         Cash Street-hail
## 7802           Standard rate         Cash Street-hail
## 8138           Standard rate  Credit card Street-hail
## 8443           Standard rate  Credit card Street-hail
## 8619           Standard rate         Cash Street-hail
## 8891           Standard rate         Cash Street-hail
## 8960           Standard rate  Credit card Street-hail
## 9207           Standard rate         Cash Street-hail
## 9503           Standard rate         Cash Street-hail
## 9747           Standard rate         Cash Street-hail
## 9765         Negotiated fare  Credit card     Dispatch
## 9984           Standard rate  Credit card Street-hail
## 10034          Standard rate  Credit card Street-hail
## 10199          Standard rate         Cash Street-hail
## 10951          Standard rate         Cash Street-hail
```

```
## 10955          Standard rate          Cash Street-hail
## 10974          Standard rate   Credit card Street-hail
## 11189          Standard rate          Cash Street-hail
## 11506          Standard rate          Cash Street-hail
## 11713          Standard rate   Credit card Street-hail
## 12492          Standard rate          Cash Street-hail
## 12792          Standard rate          Cash Street-hail
## 13043          Standard rate          Cash Street-hail
## 13274          Standard rate          Cash Street-hail
## 13332          Standard rate   Credit card Street-hail
## 13875          Standard rate   Credit card Street-hail
## 13927          Standard rate          Cash Street-hail
## 14874          Standard rate          Cash Street-hail
## 14916          Standard rate          Cash Street-hail
## 15407          Standard rate   Credit card Street-hail
## 15830          Standard rate          Cash Street-hail
## 16080          Standard rate          Cash Street-hail
## 16166          Standard rate   Credit card Street-hail
## 16345          Standard rate   Credit card Street-hail
## 16391          Standard rate   Credit card Street-hail
## 17136          Standard rate          Cash Street-hail
## 17355          Standard rate          Cash Street-hail
## 18278          Standard rate          Cash Street-hail
## 18596          Standard rate       Dispute Street-hail
## 18734          Standard rate          Cash Street-hail
## 19101          Standard rate   Credit card Street-hail
## 19344          Standard rate   Credit card Street-hail
## 19408          Standard rate   Credit card Street-hail
## 19991          Standard rate          Cash Street-hail
## 20004          Standard rate          Cash Street-hail
## 20009          Standard rate   Credit card Street-hail
## 20044          Standard rate   Credit card Street-hail
## 20077          Standard rate   Credit card Street-hail
## 20271          Standard rate   Credit card Street-hail
## 20342          Standard rate   Credit card Street-hail
## 20361          Standard rate   Credit card Street-hail
## 20543          Standard rate   Credit card Street-hail
## 20621          Standard rate          Cash Street-hail
## 20733          Standard rate          Cash Street-hail
## 20917          Standard rate   Credit card Street-hail
## 21425          Standard rate   Credit card Street-hail
## 21439          Standard rate   Credit card Street-hail
## 21539          Standard rate          Cash Street-hail
## 21559          Standard rate   Credit card Street-hail
## 21735          Standard rate   Credit card Street-hail
## 22197          Standard rate          Cash Street-hail
## 22332          Standard rate          Cash Street-hail
## 22825          Standard rate          Cash Street-hail
## 22946          Standard rate          Cash Street-hail
## 23091          Standard rate          Cash Street-hail
## 23132          Standard rate          Cash Street-hail
## 23811          Standard rate   Credit card Street-hail
## 24338          Standard rate          Cash Street-hail
## 24863          Standard rate          Cash Street-hail
```

```
## 25262       Standard rate         Cash Street-hail
## 25356       Standard rate         Cash Street-hail
## 26062       Standard rate         Cash Street-hail
## 26832       Standard rate         Cash Street-hail
## 27216       Standard rate  Credit card Street-hail
## 27482       Standard rate         Cash Street-hail
## 27495       Standard rate         Cash Street-hail
## 27594       Standard rate         Cash Street-hail
## 27984       Standard rate  Credit card Street-hail
## 28083       Standard rate         Cash Street-hail
## 28512       Standard rate  Credit card Street-hail
## 29375       Standard rate         Cash Street-hail
## 29522       Standard rate         Cash Street-hail
## 30659       Standard rate         Cash Street-hail
## 30856       Standard rate         Cash Street-hail
## 31236       Standard rate         Cash Street-hail
## 31456       Standard rate         Cash Street-hail
## 31571       Standard rate  Credit card Street-hail
## 31583       Standard rate         Cash Street-hail
## 31617       Standard rate  Credit card Street-hail
## 31726       Standard rate         Cash Street-hail
## 32873       Standard rate         Cash Street-hail
## 32952       Standard rate         Cash Street-hail
## 33882       Standard rate         Cash Street-hail
## 34250       Standard rate         Cash Street-hail
## 34280       Standard rate         Cash Street-hail
## 34374       Standard rate         Cash Street-hail
## 34390       Standard rate         Cash Street-hail
## 34922       Standard rate         Cash Street-hail
## 35039       Standard rate  Credit card Street-hail
## 35207       Standard rate  Credit card Street-hail
## 35386       Standard rate         Cash Street-hail
## 36076       Standard rate         Cash Street-hail
## 36428       Standard rate         Cash Street-hail
## 36540       Standard rate         Cash Street-hail
## 36696       Standard rate  Credit card Street-hail
## 36863       Standard rate         Cash Street-hail
## 36933       Standard rate    No charge Street-hail
## 37035       Standard rate         Cash Street-hail
## 37273       Standard rate  Credit card Street-hail
## 37506       Standard rate         Cash Street-hail
## 37517       Standard rate  Credit card Street-hail
## 37561       Standard rate         Cash Street-hail
## 37764       Standard rate         Cash Street-hail
## 37821       Standard rate         Cash Street-hail
## 37877       Standard rate         Cash Street-hail
## 38445       Standard rate  Credit card Street-hail
## 38480       Standard rate  Credit card Street-hail
## 39213       Standard rate         Cash Street-hail
## 39623       Standard rate         Cash Street-hail
## 39723       Standard rate         Cash Street-hail
## 39943       Standard rate  Credit card Street-hail
## 40226       Standard rate  Credit card Street-hail
## 40245       Standard rate         Cash Street-hail
```

```
## 40497        Standard rate  Credit card Street-hail
## 40560        Standard rate         Cash Street-hail
## 40802        Standard rate    No charge Street-hail
## 40941        Standard rate         Cash Street-hail
## 40943        Standard rate  Credit card Street-hail
## 40953        Standard rate  Credit card Street-hail
## 40969        Standard rate         Cash Street-hail
## 42048        Standard rate         Cash Street-hail
## 42779        Standard rate         Cash Street-hail
## 43577        Standard rate  Credit card Street-hail
## 43958        Standard rate  Credit card Street-hail
## 44992        Standard rate  Credit card Street-hail
## 46311        Standard rate  Credit card Street-hail
## 46572        Standard rate  Credit card Street-hail
## 46653        Standard rate         Cash Street-hail
## 46790        Standard rate         Cash Street-hail
## 47428        Standard rate  Credit card Street-hail
## 47471        Standard rate  Credit card Street-hail
## 48166        Standard rate         Cash Street-hail
## 48518        Standard rate  Credit card Street-hail
## 48796        Standard rate  Credit card Street-hail
## 48903        Standard rate  Credit card Street-hail
## 48915        Standard rate         Cash Street-hail
## 49242        Standard rate         Cash Street-hail
## 49244       Negotiated fare         Cash     Dispatch
## 49383        Standard rate  Credit card Street-hail
## 49421        Standard rate         Cash Street-hail
## 49783        Standard rate         Cash Street-hail
## 49849        Standard rate  Credit card Street-hail
## 50027        Standard rate         Cash Street-hail
## 50328        Standard rate  Credit card Street-hail
## 50542        Standard rate  Credit card Street-hail
## 50979        Standard rate         Cash Street-hail
## 50996        Standard rate         Cash Street-hail
## 51868        Standard rate         Cash Street-hail
## 51965        Standard rate  Credit card Street-hail
## 52728       Negotiated fare         Cash     Dispatch
## 52825        Standard rate  Credit card Street-hail
## 52931        Standard rate         Cash Street-hail
## 53452        Standard rate         Cash Street-hail
## 53536        Standard rate  Credit card Street-hail
##  [ reached getOption("max.print") -- omitted 4666 rows ]
```

```r
df[,"VendorID"]<-res.impute$completeObs[,"VendorID"]
df[,"RateCodeID"]<-res.impute$completeObs[,"RateCodeID"]
df[,"Payment_type"]<-res.impute$completeObs[,"Payment_type"]
df[,"Trip_type"]<-res.impute$completeObs[,"Trip_type"]
```

Creating factors

Factorize function:

```
factorize<- function(x) {
  quantile(x,seq(0,1,0.1))
  pp<-quantile(x);pp
  breaks<-c(unique(pp))
  f.x<-factor(cut(x,breaks))
  return(f.x);
}
```

f.passenger

```
df$f.passenger<-factor(cut(df$Passenger_count,breaks=c(0,1,6)))
summary(df$f.passenger)
```

```
## (0,1] (1,6]
##  4122   744
```

f.distance

```
df$f.distance<-factorize(df$Trip_distance) # NO VA be?
summary(df$f.distance)
```

```
## (0.01,1.01]  (1.01,1.8]   (1.8,3.31] (3.31,11.1]         NA's
##        1222        1222         1205        1216            1
```

f.pickup_longitude

```
df$f.pickup_longitude<-factorize(df$Pickup_longitude)
summary(df$f.pickup_longitude)
```

```
## (-74.03,-73.96] (-73.96,-73.95] (-73.95,-73.92] (-73.92,-73.79]
##            1216            1216            1217            1216
##            NA's
##               1
```

f.pickup_latitude

```
df$f.pickup_latitude<-factorize(df$Pickup_latitude)
summary(df$f.pickup_latitude)
```

```
## (40.58,40.69] (40.69,40.75]  (40.75,40.8]  (40.8,40.91]            NA's
##          1216          1216          1216          1217               1
```

f.dropoff_longitude

```
df$f.dropoff_longitude<-factorize(df$Dropoff_longitude)
summary(df$f.dropoff_longitude)
```

```
## (-74.03,-73.97] (-73.97,-73.95] (-73.95,-73.91] (-73.91,-73.75]
##            1217            1215            1216            1217
##             NA's
##                1
```

f.dropoff_latitude

```
df$f.dropoff_latitude<-factorize(df$Pickup_latitude)
summary(df$f.dropoff_latitude)
```

```
## (40.58,40.69] (40.69,40.75]  (40.75,40.8]  (40.8,40.91]          NA's
##          1216          1216          1216          1217             1
```

f.fare_amount

```
df$f.fare_amount<-factorize(df$Fare_amount)
summary(df$f.fare_amount)
```

```
##    (0.1,6]       (6,9]     (9,14] (14,42.5]       NA's
##       1250        1254       1203      1158          1
```

f.extra

```
df$f.extra<-factorize(df$Extra)
summary(df$f.extra)
```

```
## (0,0.5] (0.5,2]    NA's
##    1879     761    2226
```

f.MTA_tax

```
df$f.MTA_tax<-factorize(df$MTA_tax)
summary(df$f.MTA_tax) #11 NA's -> values of -0.5 => Outliers?
```

```
## (0,0.5]    NA's
##    4783      83
```

f.Improvement_surcharge

```
df$f.Improvement_surcharge<-factorize(df$improvement_surcharge)
summary(df$f.Improvement_surcharge) #11 NA's -> values of -0.3 => Outliers?
```

```
##    (0,0.3] (0.3,0.77]       NA's
##       4783          1        82
```

f.tip_amount

```
df$f.tip_amount<-factor(df$Tip_amount)
summary(df$f.tip_amount) #2869 NA's
```

```
##        0        1        2     1.46     1.56        3     1.36     1.66     1.96
##     2839      152      148       45       43       42       38       38       35
##      1.7     1.76     2.16     2.26     2.06      1.5     2.36     1.86      2.7
##       31       31       30       29       28       26       26       25       25
##      1.2     1.26     2.46     2.66     1.16     1.45     1.95        4     1.06
##       22       22       21       21       20       20       20       20       19
##      2.2     2.45        5     2.86     1.55     2.96     1.85     2.76     2.32
##       19       19       19       18       17       17       16       16       15
##     2.56     2.95     3.06     3.36     3.46     1.82      3.2     2.08     2.58
##       15       15       15       15       15       14       14       13       13
##     3.16     1.25     2.05     3.32      0.7     3.56     3.86     4.06     1.32
##       13       12       12       12       11       11       11       11       10
##     3.26     3.58      3.7     3.96      0.5     0.96     1.35     1.58     1.65
##       10       10       10       10        9        9        9        9        9
##     1.75     2.19     1.15     1.89     2.15      2.5     3.76     4.26     1.74
##        9        9        8        8        8        8        8        8        7
##     3.05     3.15     3.95     4.16     2.25     2.34     2.55     3.45     3.66
##        7        7        7        7        6        6        6        6        6
##     4.08      5.2     5.66     5.76     1.59     2.04     2.35     2.75      3.8
##        6        6        6        6        5        5        5        5        5
##     4.32     4.36     4.45     4.55     4.56     4.76     4.86     6.46     0.86
##        5        5        5        5        5        5        5        5        4
## (Other)
##      353
```

f.tolls_amount

```
df$f.toll<-factor(cut(df$Tolls_amount,breaks=c(-1,1,50)))
summary(df$f.toll)
```

```
## (-1,1] (1,50]
##   4799     67
```

f.total_amount

```
df$f.total<-factorize(df$Total_amount)
summary(df$f.total)
```

```
##   (0.1,7.8]    (7.8,11]  (11,16.6] (16.6,45.4]        NA's
##        1252        1187       1216        1210           1
```

Profiling

```r
library(FactoMineR)
# Numeric Target Total_Amount
#vars_con;vars_dis
#names(df)

#condes(df,1)
#condes(df, num.var=18)

# Binary Target AnyTip
#vars_con;vars_dis
#names(df)

#catdes(df[,c(vars_dis,vars_con)],5)
```