Task 5

Sidneva Anna

23 11 2020

I use data about a single payment document in Moscow districts. For this work, I took a sample of values for March 2016.

Next, I will calculate the median and 5-number summary (quantile) for the two selected indicators and build a plots.

library(readxl)

## Warning: package 'readxl' was built under R version 4.0.3

payment<-read\_xlsx('E:/payment\_msk.xlsx',sheet=1)  
str(payment)

## tibble [7,049 x 9] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:7049] "1" "2" "3" "4" ...  
## $ global\_id : chr [1:7049] "171563455" "171563456" "171563457" "171563458" ...  
## $ District : chr [1:7049] "район Арбат" "Басманный район" "район Замоскворечье" "Красносельский район" ...  
## $ Month : chr [1:7049] "Март" "Март" "Март" "Март" ...  
## $ Year : chr [1:7049] "2016" "2016" "2016" "2016" ...  
## $ CurrentQuantityUPD : chr [1:7049] "8340" "35768" "17553" "13052" ...  
## $ CurrentAveragePaymentUPD: chr [1:7049] "4421.98" "4159.37" "3695.76" "4099.45" ...  
## $ DebtQuantityUPD : chr [1:7049] "1444" "3760" "3850" "1390" ...  
## $ DebtAveragePaymentUPD : chr [1:7049] "4396.35" "2710.66" "1711.61" "3830" ...

debt\_q<-as.numeric(payment$DebtQuantityUPD)  
pay<-as.numeric(payment$CurrentQuantityUPD)  
median(debt\_q)

## [1] 6701

quantile(debt\_q)

## 0% 25% 50% 75% 100%   
## 0 2774 6701 11806 101950

median(pay)

## [1] 27330

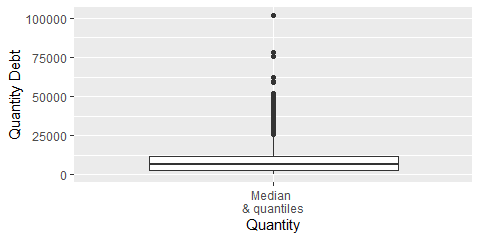
quantile(pay)

## 0% 25% 50% 75% 100%   
## 1 17271 27330 36498 114134

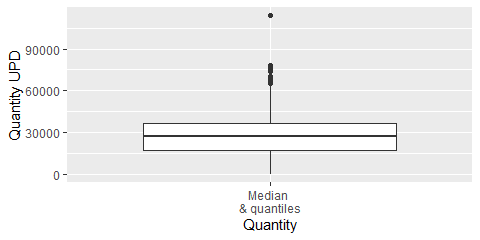
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.3

data\_pay<-data.frame(y=debt\_q)  
ggplot(data\_pay)+geom\_boxplot(aes(x = "Median \n& quantiles",y = debt\_q))+labs(x = "Quantity",y = "Quantity Debt")



data\_pay1<-data.frame(y=pay)  
ggplot(data\_pay1)+geom\_boxplot(aes(x = "Median \n& quantiles",y = pay))+labs(x = "Quantity",y = "Quantity UPD")

 The next step is to calculate the mean and standard deviation for both indicators and build a plots.

mean(debt\_q)

## [1] 8140.192

sd(debt\_q)

## [1] 7036.44

mean(pay)

## [1] 27090.85

sd(pay)

## [1] 15519.13

library(Hmisc)

## Warning: package 'Hmisc' was built under R version 4.0.3

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Warning: package 'Formula' was built under R version 4.0.3

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, units

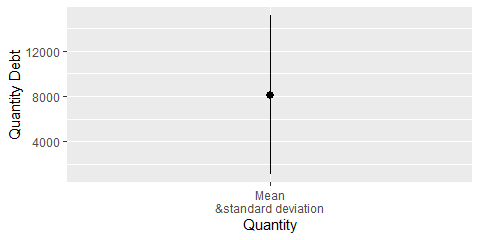
mean\_sdl(debt\_q)

## y ymin ymax  
## 1 8140.192 -5932.687 22213.07

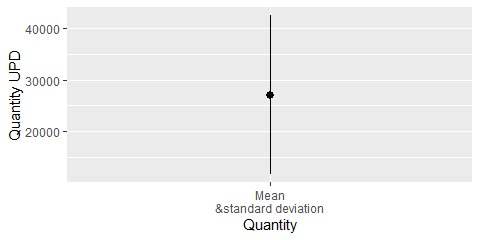
mean\_sdl(pay)

## y ymin ymax  
## 1 27090.85 -3947.409 58129.1

library(ggplot2)  
data\_pay<-data.frame(y=debt\_q)  
ggplot(data=data\_pay)+stat\_summary(geom='pointrange',fun.data=mean\_sdl,fun.args=list(mult=1),aes(x='Mean\n&standard deviation',y=debt\_q))+labs(x='Quantity',y='Quantity Debt')



data\_pay1<-data.frame(y=pay)  
ggplot(data=data\_pay1)+stat\_summary(geom='pointrange',fun.data=mean\_sdl,fun.args=list(mult=1),aes(x='Mean\n&standard deviation',y=pay))+labs(x='Quantity',y='Quantity UPD')

 Next, I will conduct a correlation analysis of two indicators using two methods.

cor(debt\_q,pay,method="pearson")

## [1] 0.6293186

cor(debt\_q,pay,method="spearman")

## [1] 0.6911058

There is a correlation between indicators.