In [4]:
```r
library(psych)
```

In [1]:
```r
workingdirectory = "C:/Users/Administrator/Documents/Master/MSIS-5223-70250 -
 Programming for Data Sci - 8282017 - 159 PM/Homework"
setwd(workingdirectory)
Personnel = paste(workingdirectory,"\\CaliforniaHospitalData_Personnel.txt", s
ep="")
Hospital =paste(workingdirectory,"\\CaliforniaHospitalData.csv", sep="")
Personnel = read.table(Personnel,header=T, sep="\t" )
Hospital = read.csv(Hospital,header =T, sep =",")
df3 = merge(Personnel, Hospital,nomatch = 0)
```

```
In [2]: df3
        nrow(df3)
```

HospitalID	Work_ID	LastName	FirstName	Gender	PositionID	PositionTitle	Compens
10767	926378	Adamson	David	M	3	Acting Director	248904
11875	380668	Smith	Mary	F	1	Regional Representative	46978
12145	756481	Rodriguez	Jose	M	1	Regional Representative	46978
13722	625519	Brown	Marietta	F	4	Safety Inspection Member	23987
13738	229849	Milgrom	Patricia	F	4	Safety Inspection Member	23987
17700	704360	Frank	Vernon	M	3	Acting Director	248904
17718	811240	Charles	Kenneth	M	1	Regional Representative	46978
17736	811240	Charles	Kenneth	M	3	Acting Director	248904
17741	363672	Marlin	Bill	M	4	Safety Inspection Member	23987
17757	308150	Fourier	Craig	M	3	Acting Director	248904
19868	318165	Coulter	Tracy	F	1	Regional Representative	46978
20225	856720	Roberts	Melissa	F	4	Safety Inspection Member	23987
20231	966527	Flemming	Kris	F	2	State Board Representative	89473
20266	231581	Adams	Sandy	F	1	Regional Representative	46978
20277	398103	Johanson	Sandy	F	1	Regional Representative	46978
21053	318165	Coulter	Tracy	F	4	Safety Inspection Member	23987
22181	661286	Jones	Holly	F	2	State Board Representative	89473
22460	380668	Smith	Mary	F	4	Safety Inspection Member	23987

HospitalID	Work_ID	LastName	FirstName	Gender	PositionID	PositionTitle	Compens
22598	908658	Holmes	Holly	F	3	Acting Director	248904
25667	926378	Adamson	David	M	1	Regional Representative	46978
26206	308150	Fourier	Craig	M	3	Acting Director	248904
28283	548312	Smith	Frank	M	3	Acting Director	248904
28812	908658	Holmes	Holly	F	3	Acting Director	248904
29805	304224	Wood	Emily	F	2	State Board Representative	89473
29823	768057	Tanner	Patricia	F	3	Acting Director	248904
31032	310431	Dong	HongWei	F	4	Safety Inspection Member	23987
32930	411280	Manin	Emily	F	1	Regional Representative	46978
33192	308150	Fourier	Craig	M	1	Regional Representative	46978
33207	811240	Charles	Kenneth	M	2	State Board Representative	89473
33242	704360	Frank	Vernon	M	4	Safety Inspection Member	23987
...	...	...	...	...	...	...	...
33283	704360	Frank	Vernon	M	1	Regional Representative	46978
34454	721586	Web	David	M	3	Acting Director	248904
34959	398103	Johanson	Sandy	F	1	Regional Representative	46978
35512	625519	Brown	Marietta	F	1	Regional Representative	46978
36569	966527	Flemming	Kris	F	2	State Board Representative	89473
37043	664583	Paanua	Kaaluai	M	2	State Board Representative	89473
37285	304224	Wood	Emily	F	4	Safety Inspection Member	23987
37393	909709	Iwata	Akira	M	1	Regional Representative	46978

HospitalID	Work_ID	LastName	FirstName	Gender	PositionID	PositionTitle	Compens
37436	398103	Johanson	Sandy	F	2	State Board Representative	89473
38798	909709	Iwata	Akira	M	1	Regional Representative	46978
38802	310431	Dong	HongWei	F	1	Regional Representative	46978
38900	664583	Paanua	Kaaluai	M	1	Regional Representative	46978
39050	363672	Marlin	Bill	M	3	Acting Director	248904
39076	304224	Wood	Emily	F	1	Regional Representative	46978
39102	661286	Jones	Holly	F	2	State Board Representative	89473
39110	909709	Iwata	Akira	M	1	Regional Representative	46978
43353	756481	Rodriguez	Jose	M	4	Safety Inspection Member	23987
43437	768057	Tanner	Patricia	F	1	Regional Representative	46978
44817	363672	Marlin	Bill	M	2	State Board Representative	89473
44982	231581	Adams	Sandy	F	1	Regional Representative	46978
45067	926378	Adamson	David	M	4	Safety Inspection Member	23987
45736	721586	Web	David	M	2	State Board Representative	89473
45740	721586	Web	David	M	4	Safety Inspection Member	23987
45771	756481	Rodriguez	Jose	M	1	Regional Representative	46978
46342	768057	Tanner	Patricia	F	3	Acting Director	248904
46348	664583	Paanua	Kaaluai	M	1	Regional Representative	46978

HospitalID	Work_ID	LastName	FirstName	Gender	PositionID	PositionTitle	Compens
46838	229849	Milgrom	Patricia	F	1	Regional Representative	46978
46844	908658	Holmes	Holly	F	3	Acting Director	248904
46985	548312	Smith	Frank	M	1	Regional Representative	46978
46996	856720	Roberts	Melissa	F	4	Safety Inspection Member	23987

61

Subset random rows

```
In [52]: dat.sub <- df3[sample(nrow(df3), 20), ]
```

Select existing hospital

```
In [11]: Random = df3[sample(nrow(df3),1),]
         Random
```

	HospitalID	Work_ID	LastName	FirstName	Gender	PositionID	PositionTitle	Compe
19	22598	908658	Holmes	Holly	F	3	Acting Director	248904

Add New Position and my last and first name

In [12]:
```
Random$Work_ID <- as.integer(999999)
Random$LastName <- factor('Pham')
Random$FirstName <- factor('Khanh')
Random$PositionID <- as.integer(5)
Random$StartDate <- factor("9/9/2017")
Random
```

	HospitalID	Work_ID	LastName	FirstName	Gender	PositionID	PositionTitle	Compe
19	22598	999999	Pham	Khanh	F	5	Acting Director	248904

In [14]:
```
df3 <- merge(df3,Random,all=TRUE)
nrow(df3)
```

62

Covert Any date-tmime columns into a datetime datatype

In [15]:
```
df3$StartDate <- as.Date(df3$StartDate, format = "%m/%d/%Y")
str(df3$StartDate)
```

 Date[1:62], format: "2002-01-01" "2011-01-01" "2009-01-01" "2012-01-01" "201
1-01-01" ...

Remove the three primary keys

```
In [16]:  df3$HospitalID <- NULL
          df3$Work_ID <- NULL
          df3$PositionID <- NULL
          df3
```

LastName	FirstName	Gender	PositionTitle	Compensation	MaxTerm	StartDate	Name
Adamson	David	M	Acting Director	248904	8	2002-01-01	Kern Hospi
Smith	Mary	F	Regional Representative	46978	4	2011-01-01	San Gorg Memo Hospi
Rodriguez	Jose	M	Regional Representative	46978	4	2009-01-01	Victor Valley Comr Hospi
Brown	Marietta	F	Safety Inspection Member	23987	2	2012-01-01	Redw Memo Hospi
Milgrom	Patricia	F	Safety Inspection Member	23987	2	2011-01-01	St. Ma Medic Cente
Frank	Vernon	M	Acting Director	248904	8	2005-01-01	Mark St. Josep Hospi
Charles	Kenneth	M	Regional Representative	46978	4	2009-01-01	Mercy Medic Cente Moun Shast
Charles	Kenneth	M	Acting Director	248904	8	2006-01-01	Sierra Nevac Memo Hospi
Marlin	Bill	M	Safety Inspection Member	23987	2	2011-01-01	St. Elizat Comr Hospi
Fourier	Craig	M	Acting Director	248904	8	2005-01-01	St. Ma Medic Cente san franci
Coulter	Tracy	F	Regional Representative	46978	4	2010-01-01	Ojai \ Comr Hospi

LastName	FirstName	Gender	PositionTitle	Compensation	MaxTerm	StartDate	Name
Roberts	Melissa	F	Safety Inspection Member	23987	2	2011-01-01	Frank Howa Memo Hospi
Flemming	Kris	F	State Board Representative	89473	3	2012-01-01	Hanfo Comr Medic Cente
Adams	Sandy	F	Regional Representative	46978	4	2009-01-01	Sono Regio Medic Cente green
Johanson	Sandy	F	Regional Representative	46978	4	2012-01-01	Ukiah Valley Medic Cente
Coulter	Tracy	F	Safety Inspection Member	23987	2	2011-01-01	Santa Clara Valley Medic Cente
Jones	Holly	F	State Board Representative	89473	3	2012-01-01	Marsh Hospi
Smith	Mary	F	Safety Inspection Member	23987	2	2012-01-01	Stanfc Hospi Clinic
Holmes	Holly	F	Acting Director	248904	8	2012-01-01	Kern Medic Cente
Pham	Khanh	F	Acting Director	248904	8	2017-09-09	Kern Medic Cente
Adamson	David	M	Regional Representative	46978	4	2012-01-01	Pione Memo Hospi
Fourier	Craig	M	Acting Director	248904	8	2011-01-01	Arrow Regio Medic Cente

LastName	FirstName	Gender	PositionTitle	Compensation	MaxTerm	StartDate	Name
Smith	Frank	M	Acting Director	248904	8	2005-01-01	Hi-De Medic Cente
Holmes	Holly	F	Acting Director	248904	8	2003-01-01	Oak \ Distric Hospi
Wood	Emily	F	State Board Representative	89473	3	2011-01-01	Alame Hospi
Tanner	Patricia	F	Acting Director	248904	8	2009-01-01	Colus Regio Medic Cente
Dong	HongWei	F	Safety Inspection Member	23987	2	2012-01-01	Long Beacl Memo Miller Childi Hospi
Manin	Emily	F	Regional Representative	46978	4	2011-01-01	Scripp Mercy Hospi
Fourier	Craig	M	Regional Representative	46978	4	2011-01-01	Califo Pacifi Medic Cente Pacifi Camp
Charles	Kenneth	M	State Board Representative	89473	3	2010-01-01	Memo Hospi Los B
...	...	...	...	...	...	...	...
Frank	Vernon	M	Regional Representative	46978	4	2012-01-01	Sutter Lakes Hospi
Web	David	M	Acting Director	248904	8	2012-01-01	LAC L Medic Cente
Johanson	Sandy	F	Regional Representative	46978	4	2012-01-01	Biggs Gridle Memo Hospi

LastName	FirstName	Gender	PositionTitle	Compensation	MaxTerm	StartDate	Name
Brown	Marietta	F	Regional Representative	46978	4	2009-01-01	Comr Regic Medic Cente
Flemming	Kris	F	State Board Representative	89473	3	2010-01-01	River Coun Regic Medic Cente
Paanua	Kaaluai	M	State Board Representative	89473	3	2010-01-01	Color River Medic Cente
Wood	Emily	F	Safety Inspection Member	23987	2	2011-01-01	Twin Comr Hospi
Iwata	Akira	M	Regional Representative	46978	4	2011-01-01	Barst Comr Hospi
Johanson	Sandy	F	State Board Representative	89473	3	2012-01-01	Fallbr Hospi
Iwata	Akira	M	Regional Representative	46978	4	2011-01-01	Golet Valley Cotta Hospi
Dong	HongWei	F	Regional Representative	46978	4	2010-01-01	Santa Ynez Valley Cotta Hospi
Paanua	Kaaluai	M	Regional Representative	46978	4	2011-01-01	Ceda Sinai Medic Cente
Marlin	Bill	M	Acting Director	248904	8	2008-01-01	UC D Medic Cente
Wood	Emily	F	Regional Representative	46978	4	2010-01-01	UC Ir Medic Cente

LastName	FirstName	Gender	PositionTitle	Compensation	MaxTerm	StartDate	Name
Jones	Holly	F	State Board Representative	89473	3	2012-01-01	UCSD Medic Cente Hillcre
Iwata	Akira	M	Regional Representative	46978	4	2012-01-01	UCSF Medic Cente
Rodriguez	Jose	M	Safety Inspection Member	23987	2	2011-01-01	Catali Island Medic Cente
Tanner	Patricia	F	Regional Representative	46978	4	2010-01-01	South Inyo Hospi
Marlin	Bill	M	State Board Representative	89473	3	2010-01-01	Fairch Medic Cente
Adams	Sandy	F	Regional Representative	46978	4	2008-01-01	Heald Distric Hospi
Adamson	David	M	Safety Inspection Member	23987	2	2012-01-01	Glenr Medic Cente
Web	David	M	State Board Representative	89473	3	2011-01-01	Moun Comr Hospi
Web	David	M	Safety Inspection Member	23987	2	2012-01-01	Mamr Hospi
Rodriguez	Jose	M	Regional Representative	46978	4	2011-01-01	Easte Pluma Hospi Porto Camp
Tanner	Patricia	F	Acting Director	248904	8	2011-01-01	Coalir Regic Medic Cente
Paanua	Kaaluai	M	Regional Represeative	46978	4	2011-01-01	Barto Memo Hospi

LastName	FirstName	Gender	PositionTitle	Compensation	MaxTerm	StartDate	Name
Milgrom	Patricia	F	Regional Representative	46978	4	2009-01-01	Modo Medic Cente
Holmes	Holly	F	Acting Director	248904	8	2010-01-01	Palm Hospi
Smith	Frank	M	Regional Representative	46978	4	2010-01-01	Tahoe Fores Hospi Distric
Roberts	Melissa	F	Safety Inspection Member	23987	2	2009-01-01	Ridge Regic Hospi

Filter numeric variables and Provide a summary of the mean, median, minimum value, and maximum value

```
In [50]:  nums <- df3[sapply(df3, is.numeric)]
          summary(nums)

   Compensation        MaxTerm           NoFTE           NetPatRev
 Min.   : 23987   Min.   :2.000    Min.   :  55.0    Min.   :   14172
 1st Qu.: 46978   1st Qu.:3.000    1st Qu.: 216.4    1st Qu.:  72848
 Median : 46978   Median :4.000    Median : 385.9    Median : 172813
 Mean   : 95293   Mean   :4.306    Mean   :1358.3    Mean   : 666859
 3rd Qu.: 89473   3rd Qu.:4.000    3rd Qu.:1477.8    3rd Qu.: 531705
 Max.   :248904   Max.   :8.000    Max.   :8000.0    Max.   :4662582
    InOperExp         OutOperExp          OperRev
 Min.   :1.436e+06   Min.   :   3787718   Min.   :5.235e+06
 1st Qu.:1.383e+07   1st Qu.:  14293981   1st Qu.:2.686e+07
 Median :3.228e+07   Median :  35827740   Median :6.393e+07
 Mean   :1.664e+08   Mean   :  90204998   Mean   :2.551e+08
 3rd Qu.:1.737e+08   3rd Qu.:  93066706   3rd Qu.:1.982e+08
 Max.   :1.286e+09   Max.   :662155886    Max.   :1.912e+09
    OperInc            AvlBeds
 Min.   :-321163336   Min.   : 12.00
 1st Qu.:  -1198973   1st Qu.: 42.75
 Median :   1464124   Median :100.50
 Mean   :  -1459651   Mean   :194.79
 3rd Qu.:   6146421   3rd Qu.:228.25
 Max.   : 164872413   Max.   :909.00
```

```
In [17]:  write.table(df3, "Khanh_Pham_Export.txt", sep="\t")
```

Open New data

```
In [18]:   dfnew = paste(workingdirectory,"\\Khanh_Pham_Export.txt", sep="")
           dfnew = read.table(dfnew, header = T,sep = "\t")
```

Histogram

```
In [19]:   attach(dfnew)
```

In [20]:
```
hist(NoFTE)
hist(InOperExp)
hist(OutOperExp)
hist(OperRev)
hist(OperInc)
hist(AvlBeds)
hist(NetPatRev)
```

Histogram of NoFTE
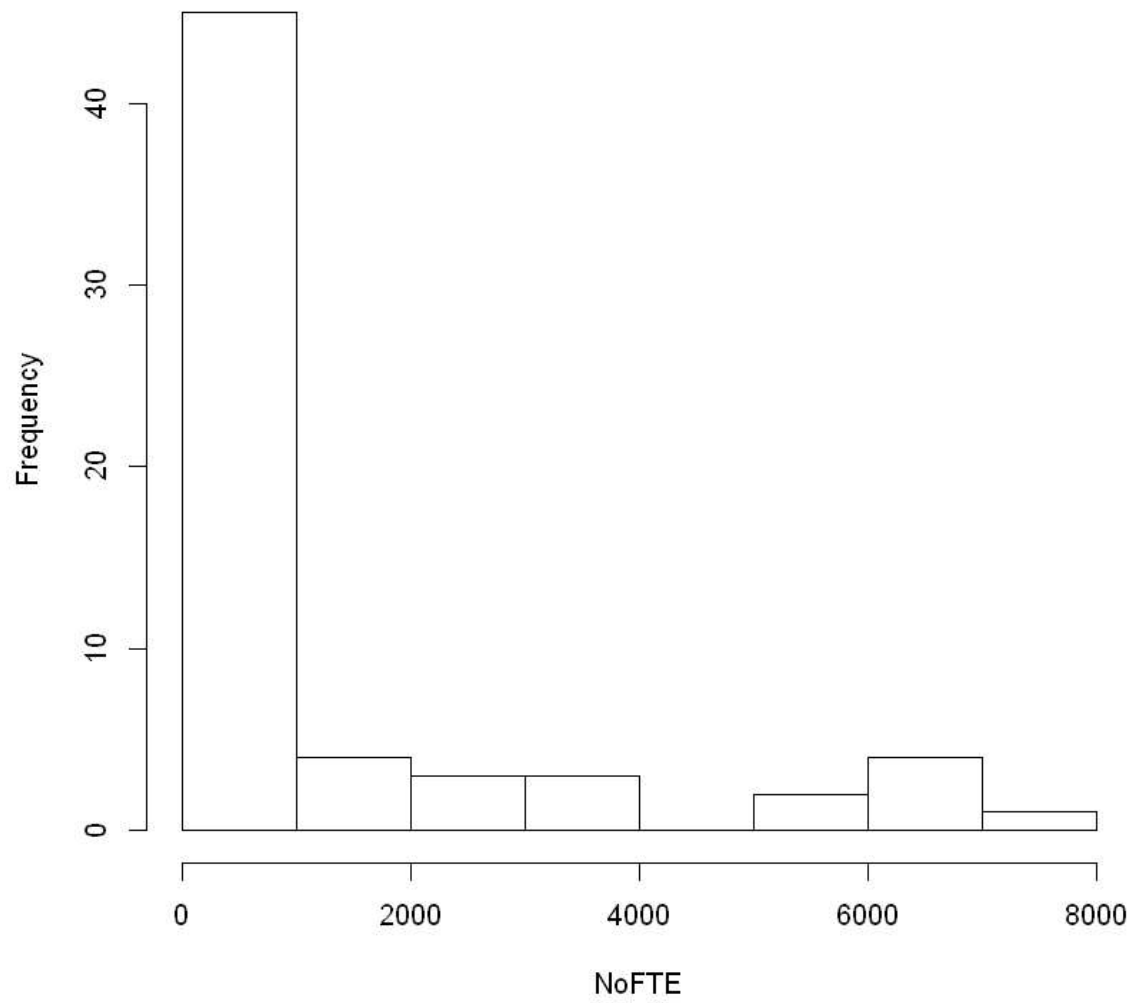
Histogram of InOperExp
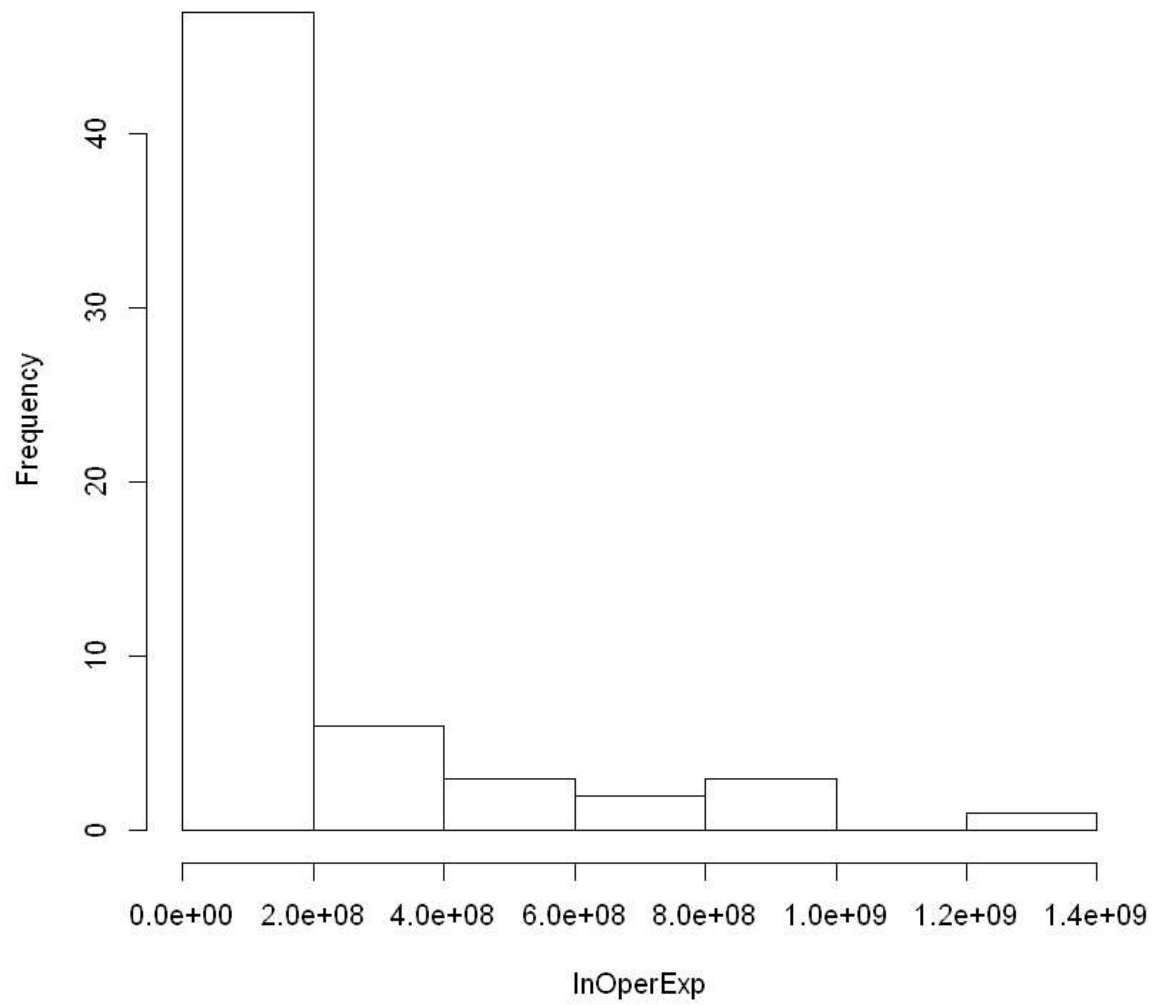
Histogram of OutOperExp
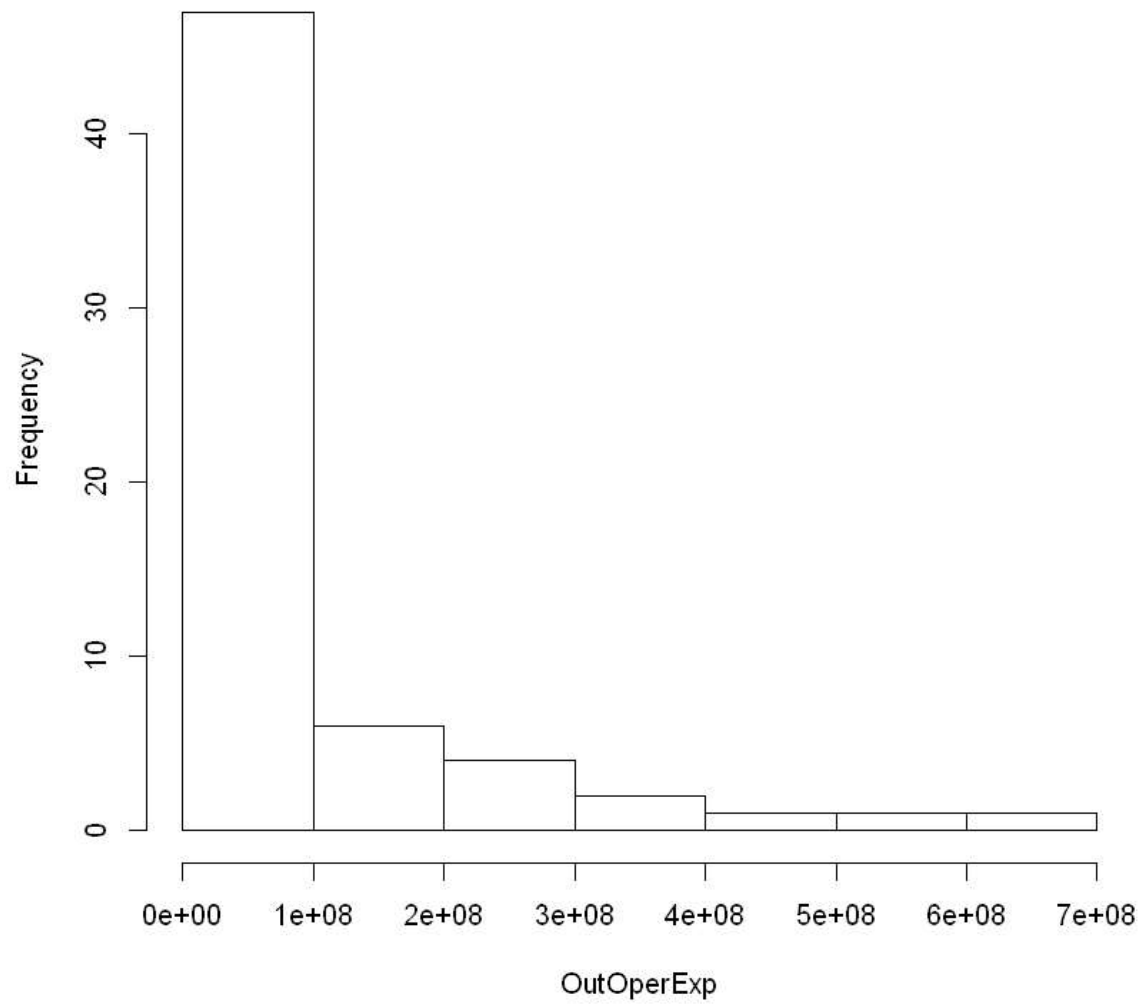
Histogram of OperRev
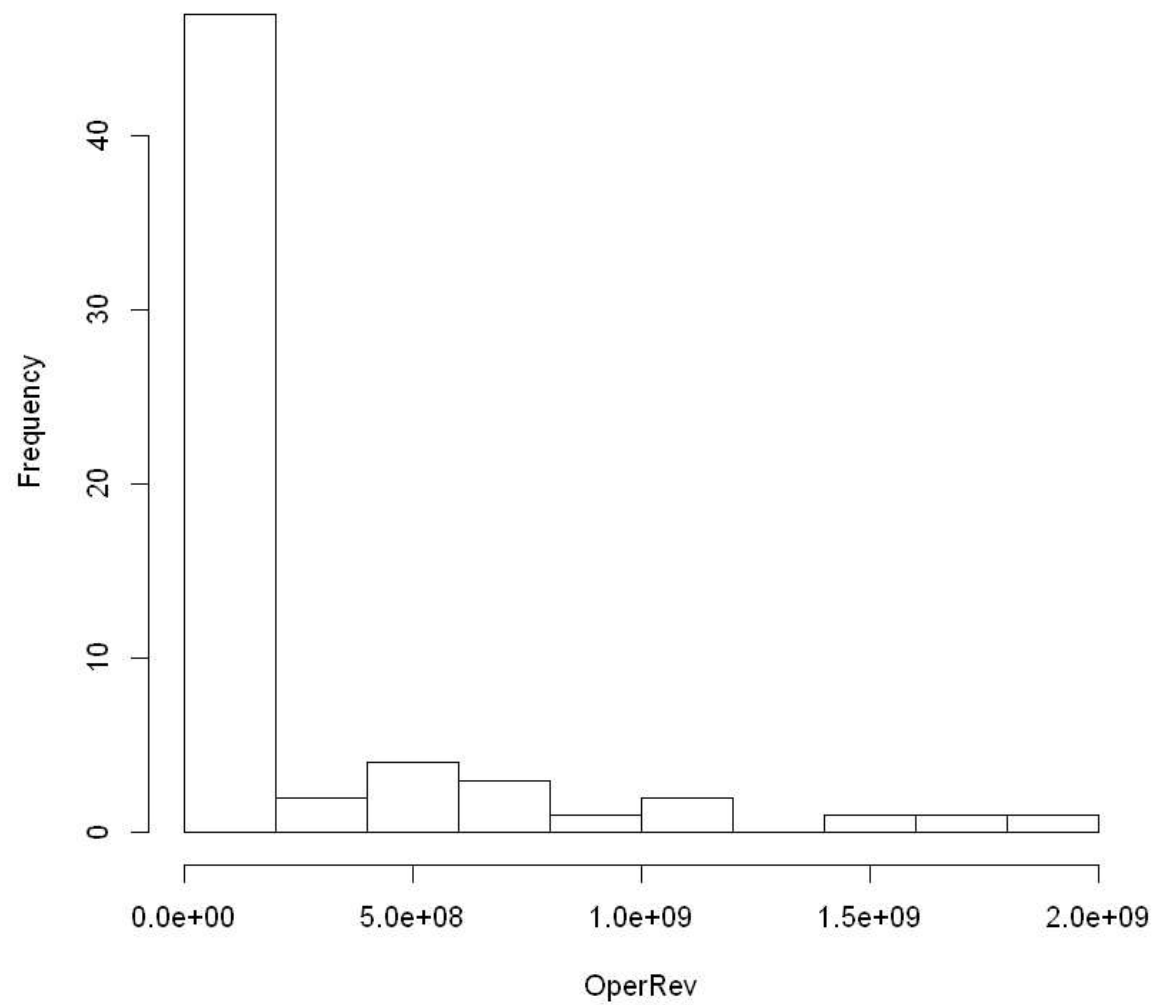
Histogram of OperInc
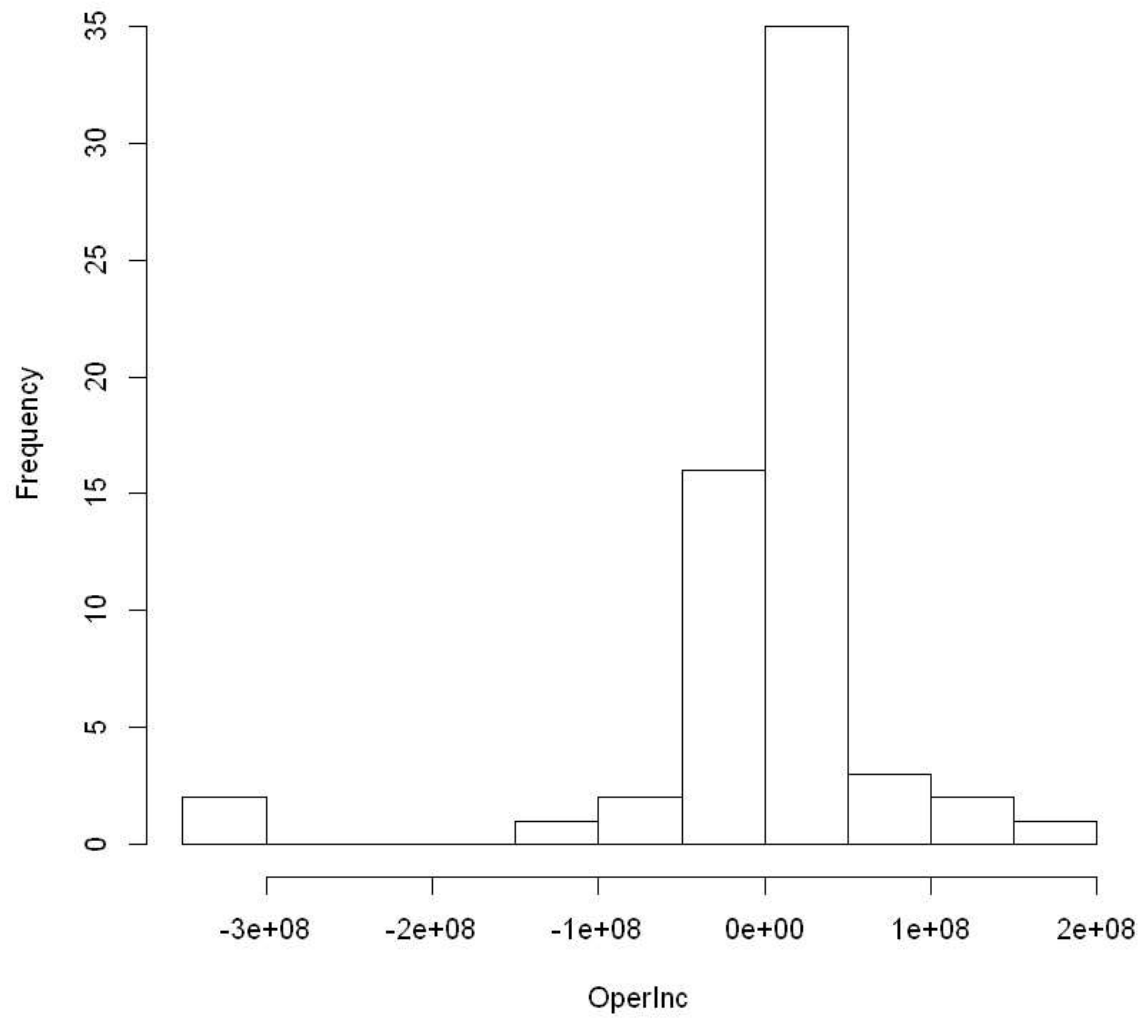
Histogram of AvlBeds
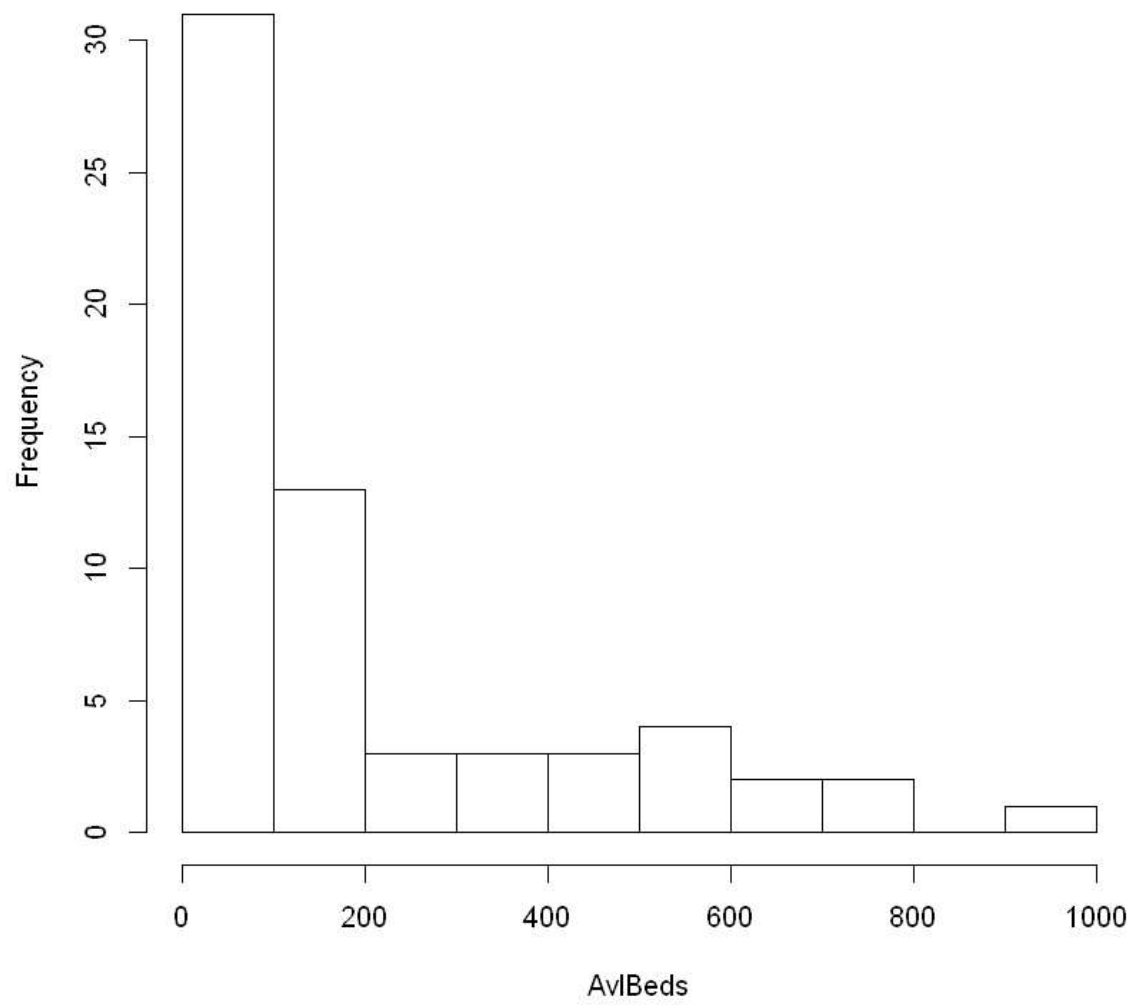
Histogram of NetPatRev
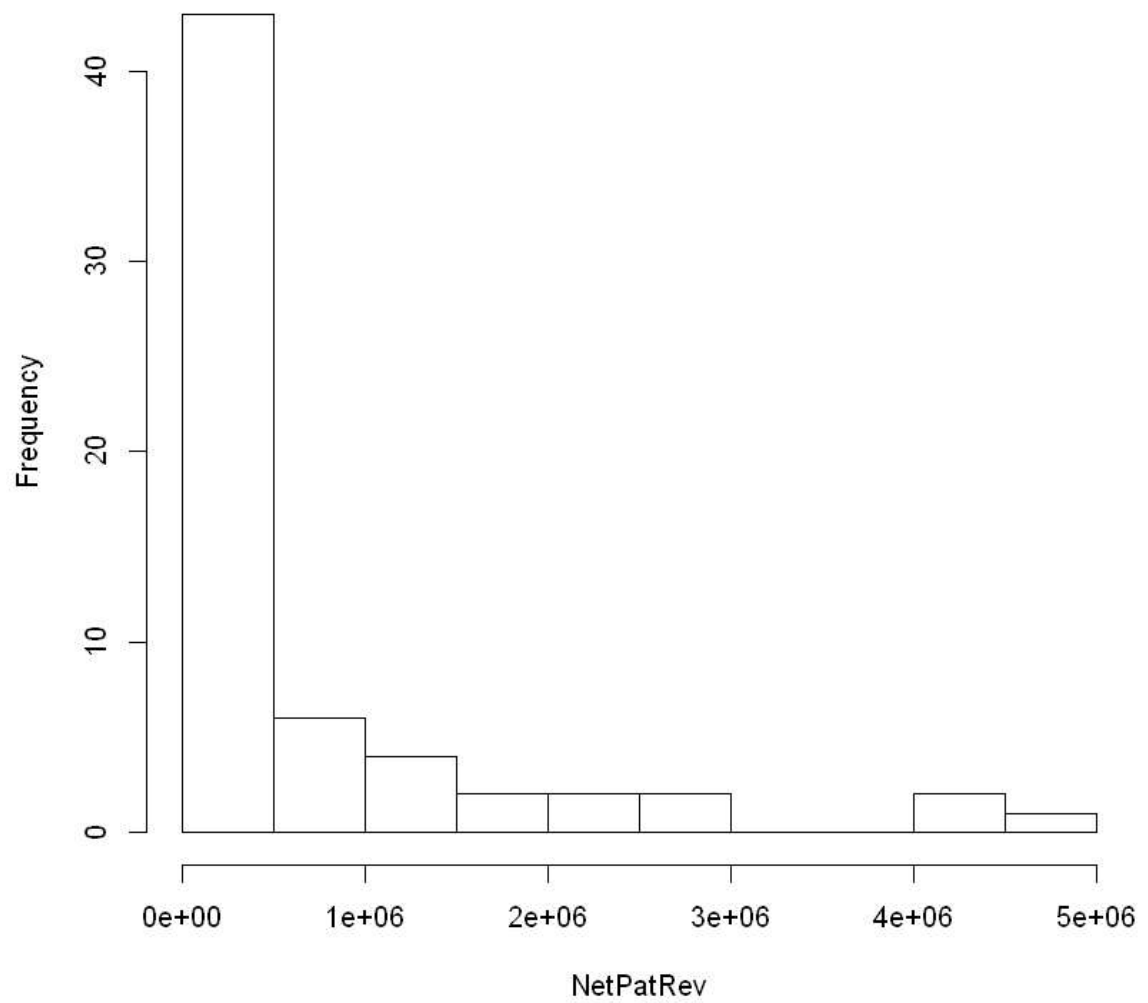
In [21]:
```
plot(NoFTE)
plot(InOperExp)
plot(OutOperExp)
plot(OperRev)
plot(OperInc)
plot(AvlBeds)
```
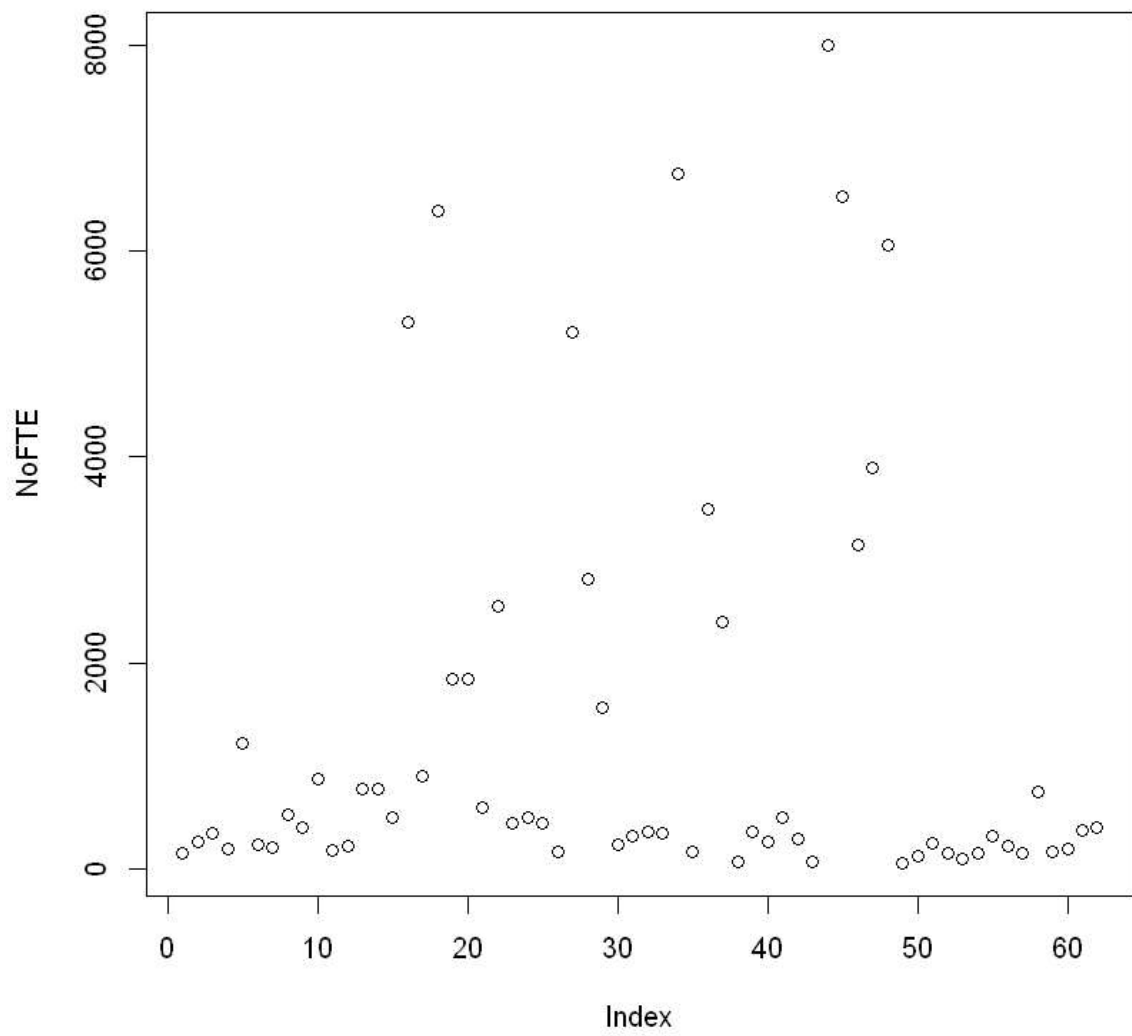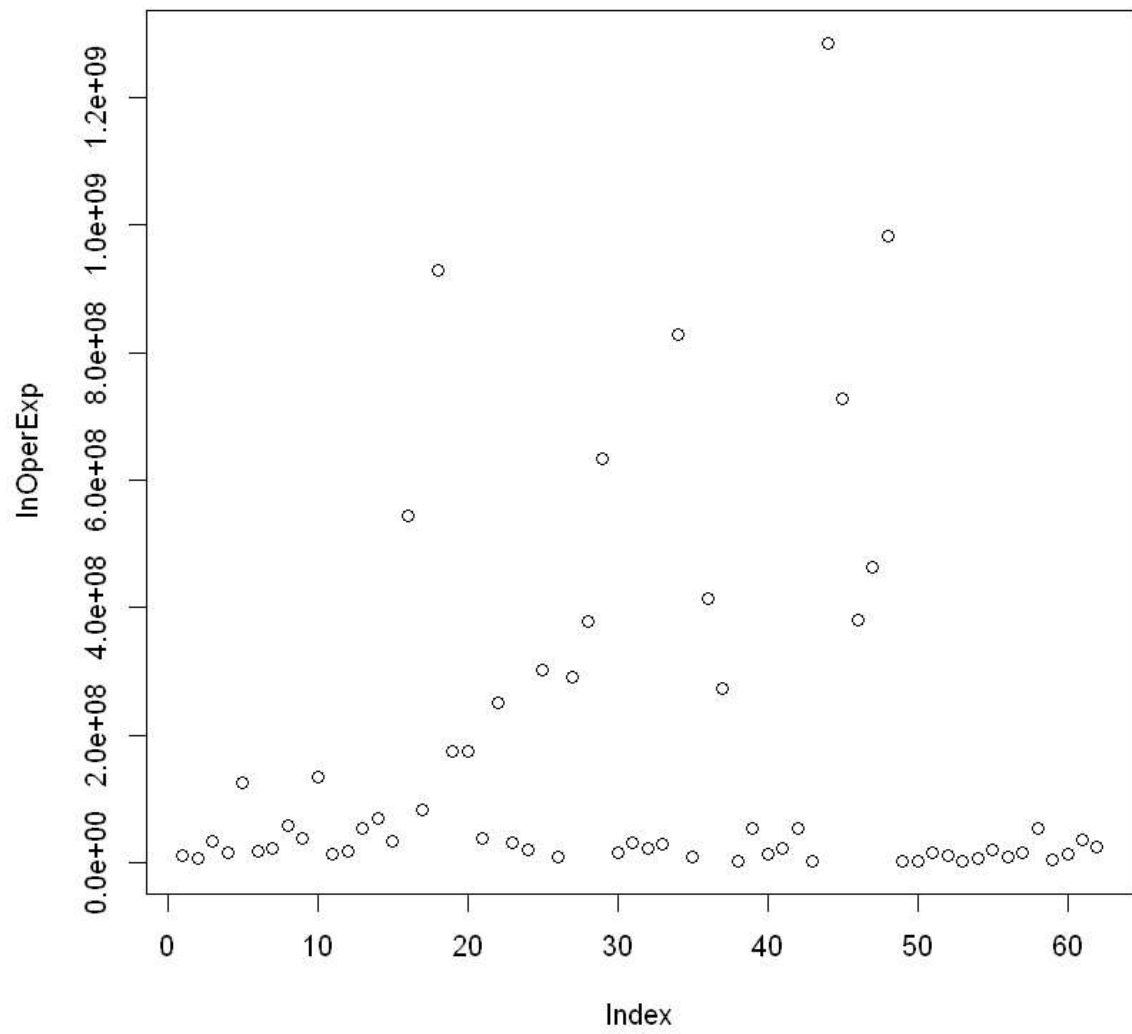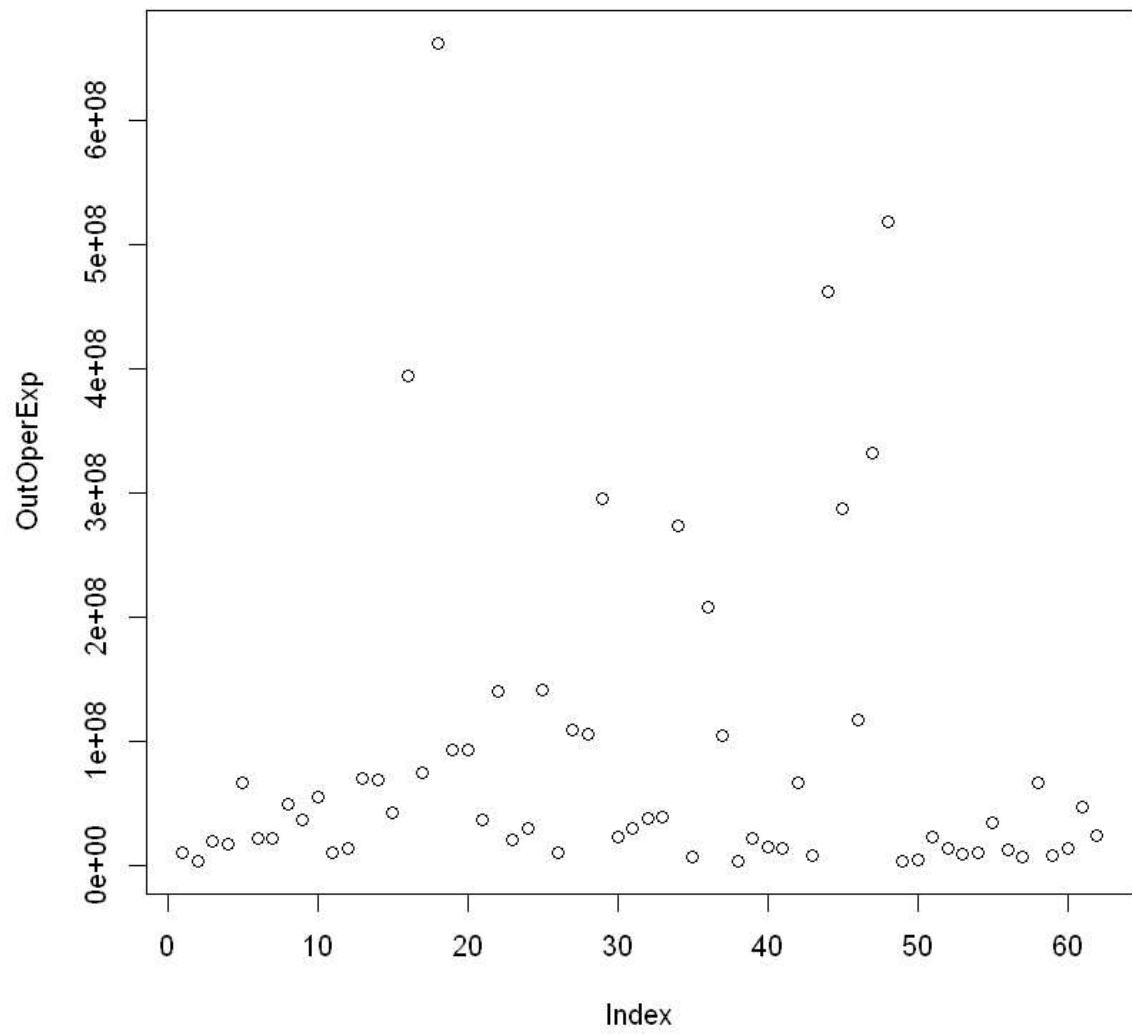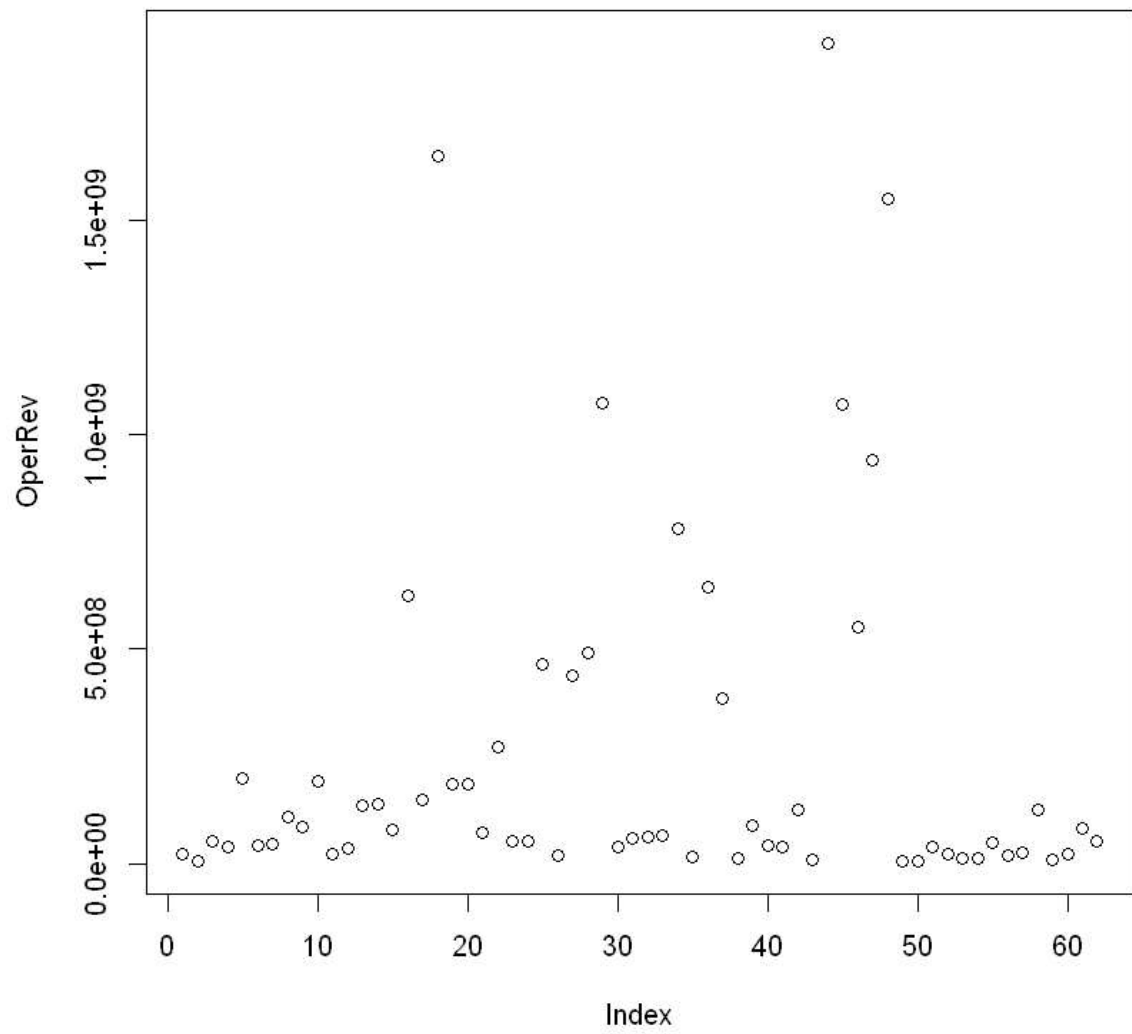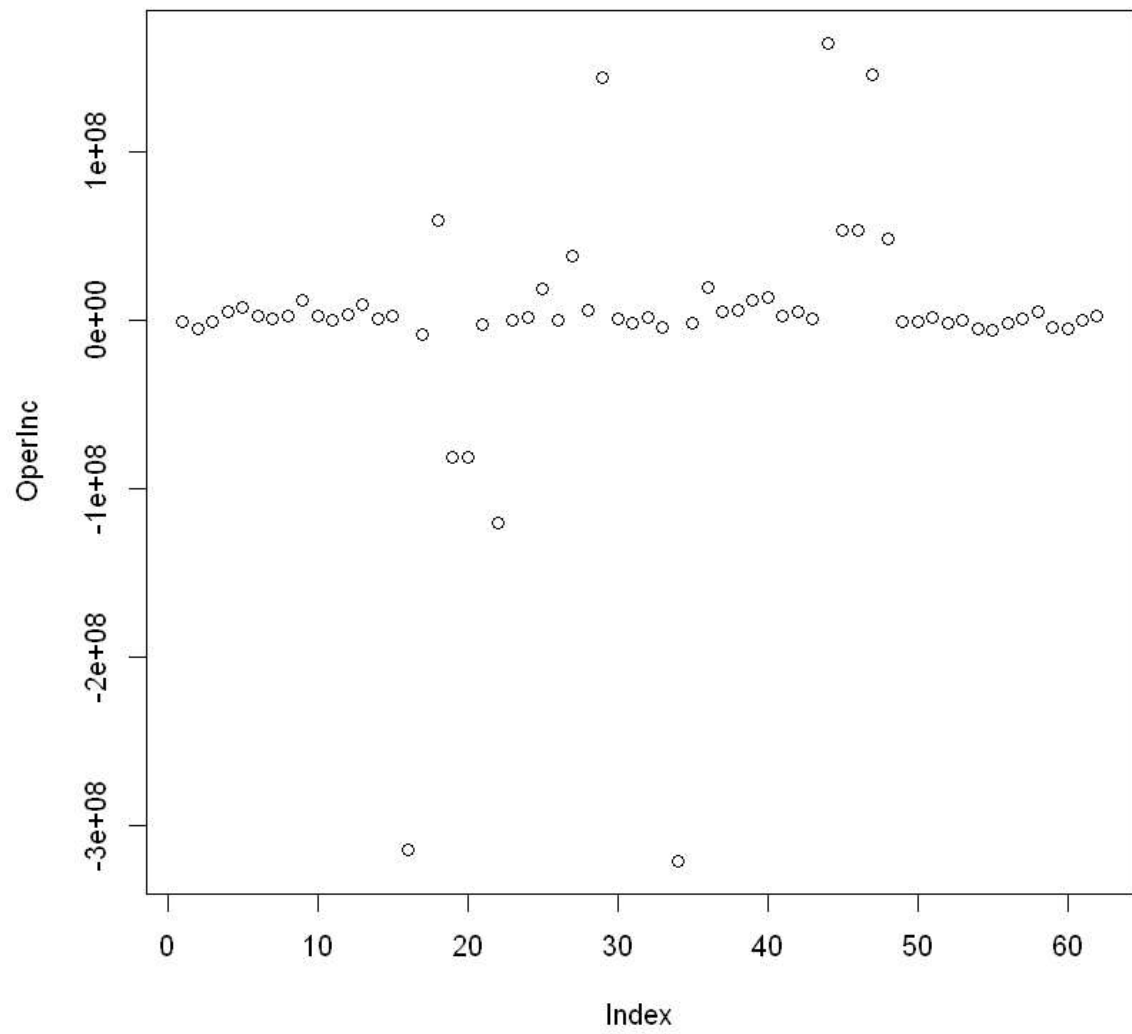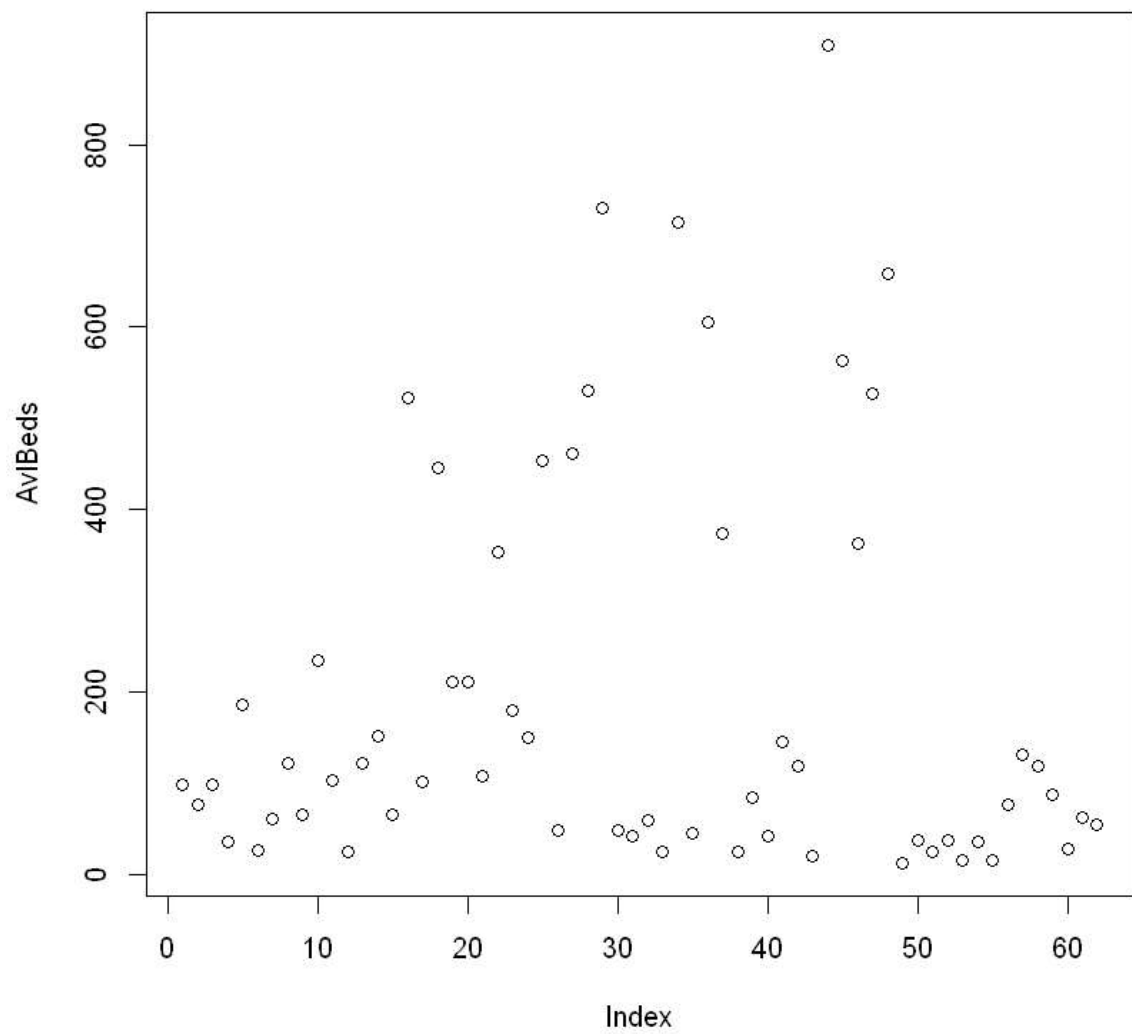
Create boxplot

In [22]:
```
boxplot(NoFTE)
boxplot(InOperExp)
boxplot(OutOperExp)
boxplot(OperRev)
boxplot(OperInc)
boxplot(AvlBeds)
boxplot(NetPavRev)
```
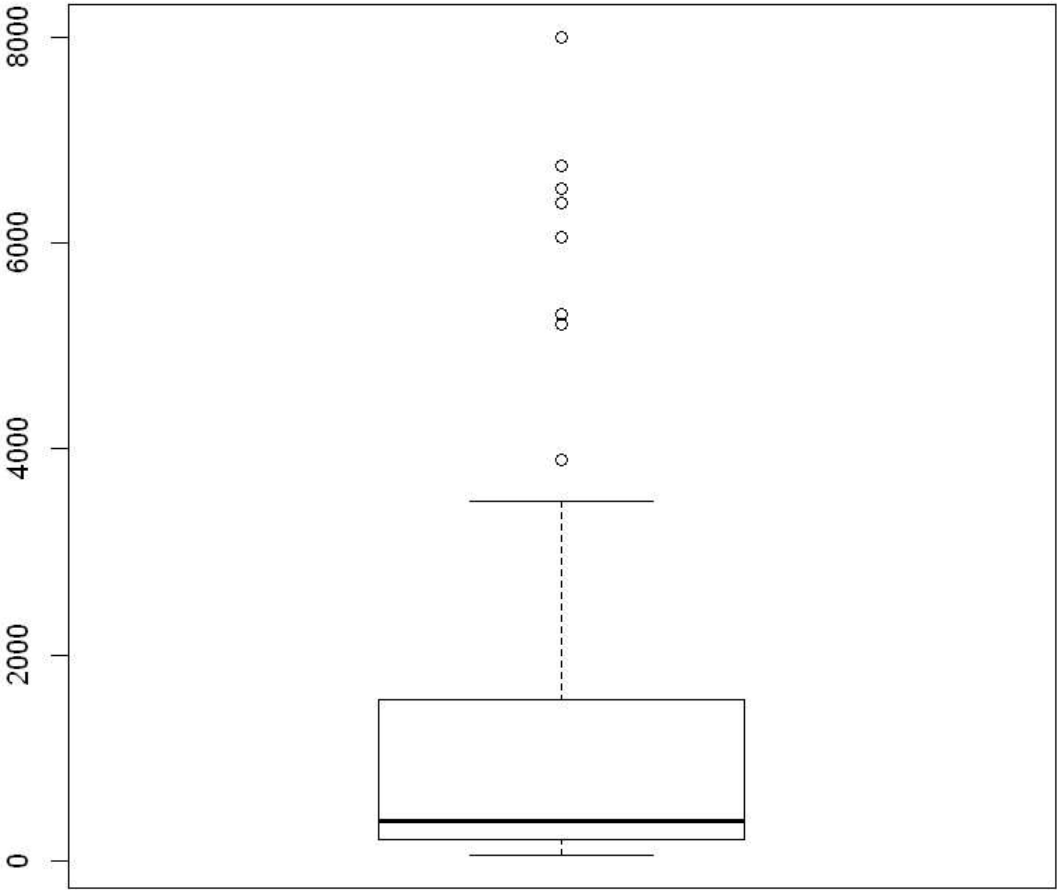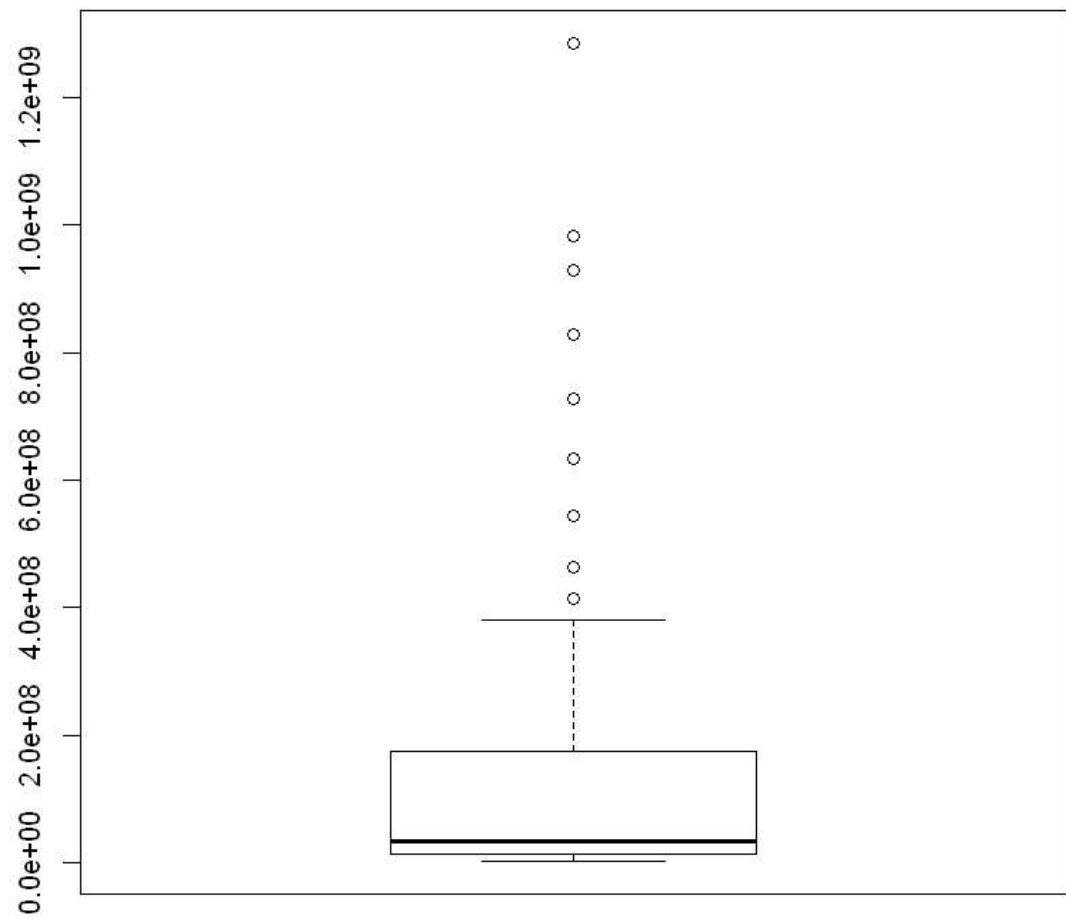
```
Warning message in x[floor(d)] + x[ceiling(d)]:
"NAs produced by integer overflow"
```
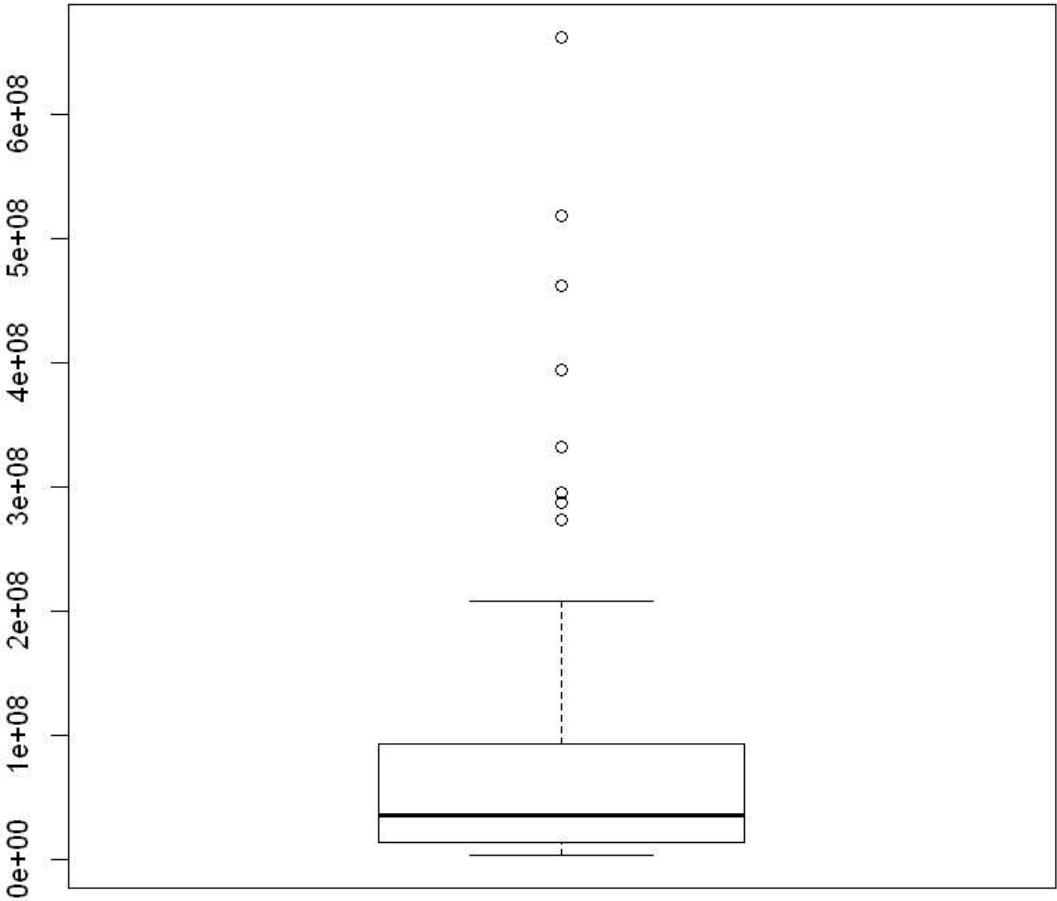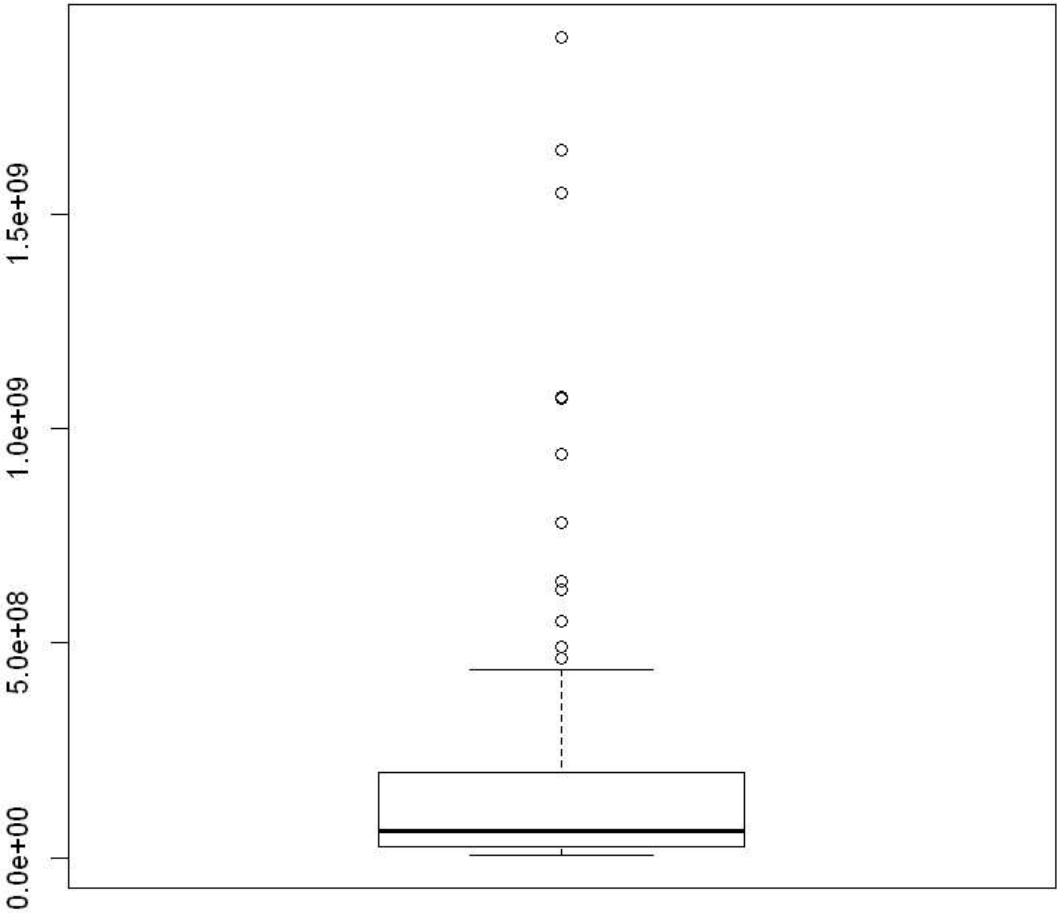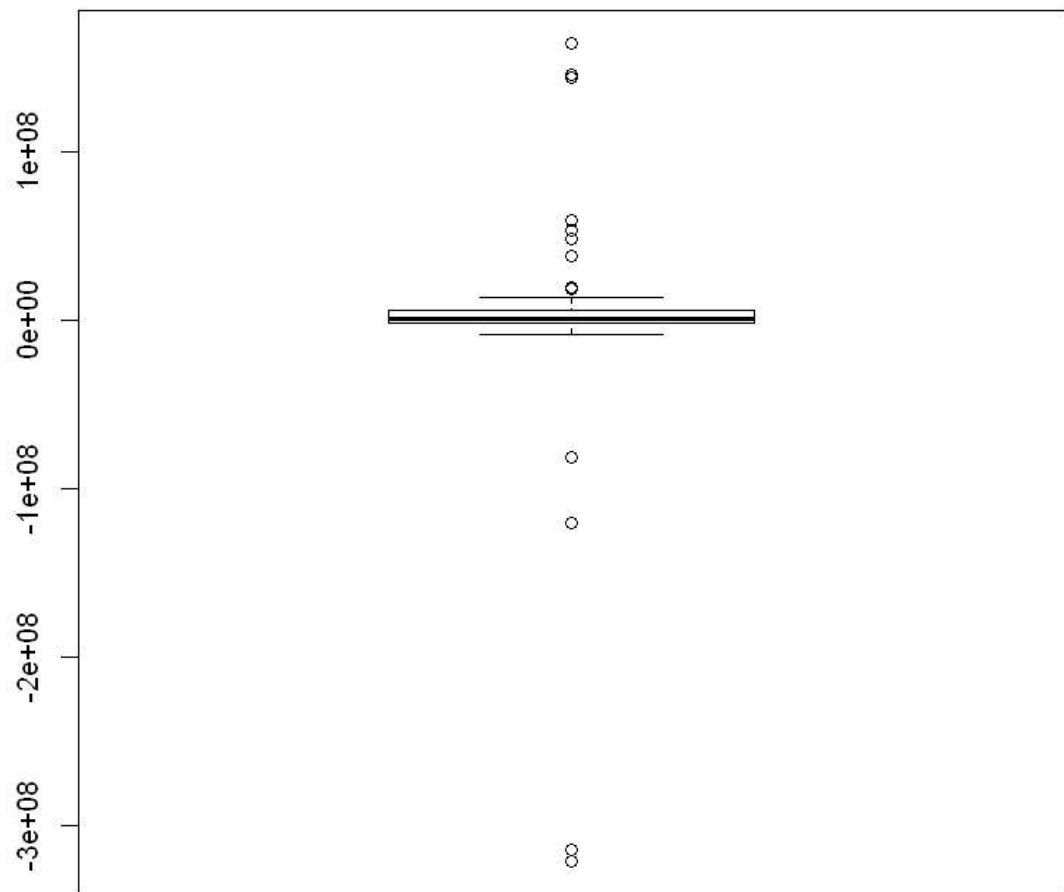
```
Error in boxplot(NetPavRev): object 'NetPavRev' not found
Traceback:

1. boxplot(NetPavRev)
```
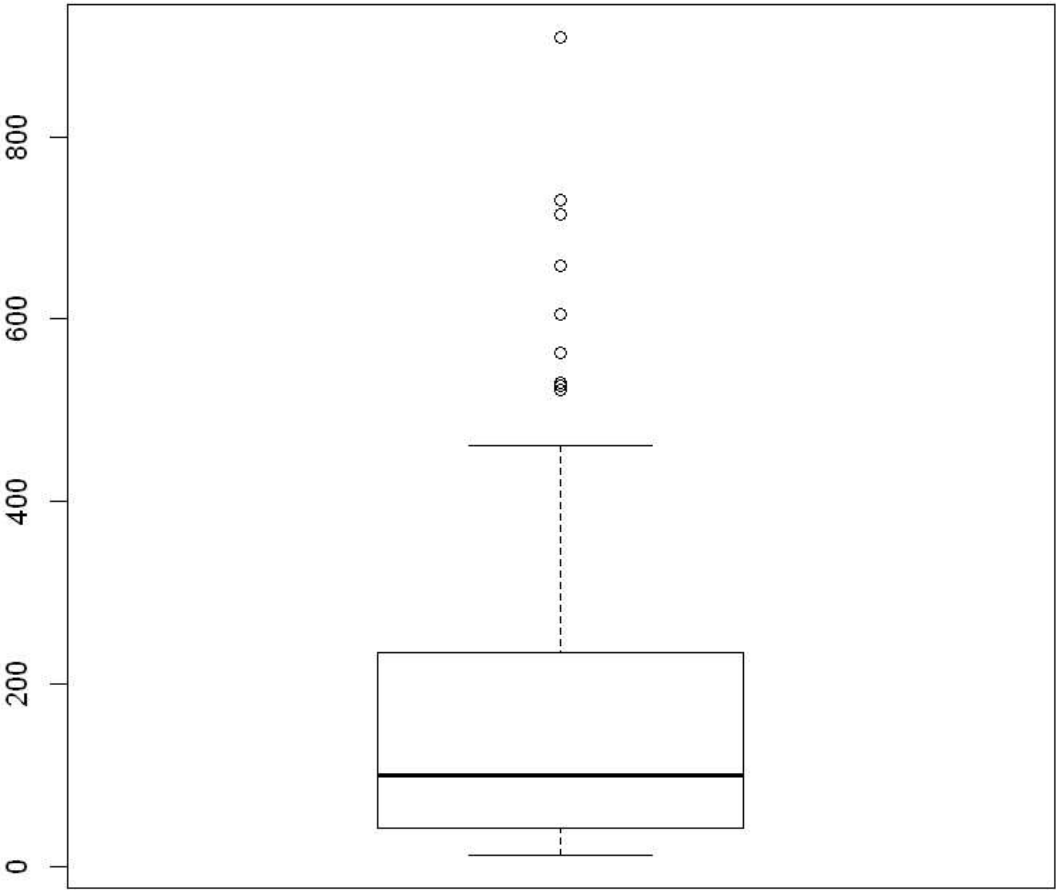
In [23]:
```
qqnorm(NoFTE, main = 'NoFTE')
qqline(NoFTE, col ='red')
qqnorm(InOperExp, main = 'InOperExp')
qqline(InOperExp, col ='red')
qqnorm(OutOperExp, main = 'OutOperExp')
qqline(OutOperExp, col ='red')
qqnorm(OperRev, main = 'OperRev')
qqline(OperRev, col ='red')
qqnorm(OperInc, main = 'OperInc')
qqline(OperInc, col ='red')
qqnorm(AvlBeds, main = 'AvlBeds')
qqline(AvlBeds, col ='red')
```

NoFTE

lnOperExp

OutOperExp

OperRev
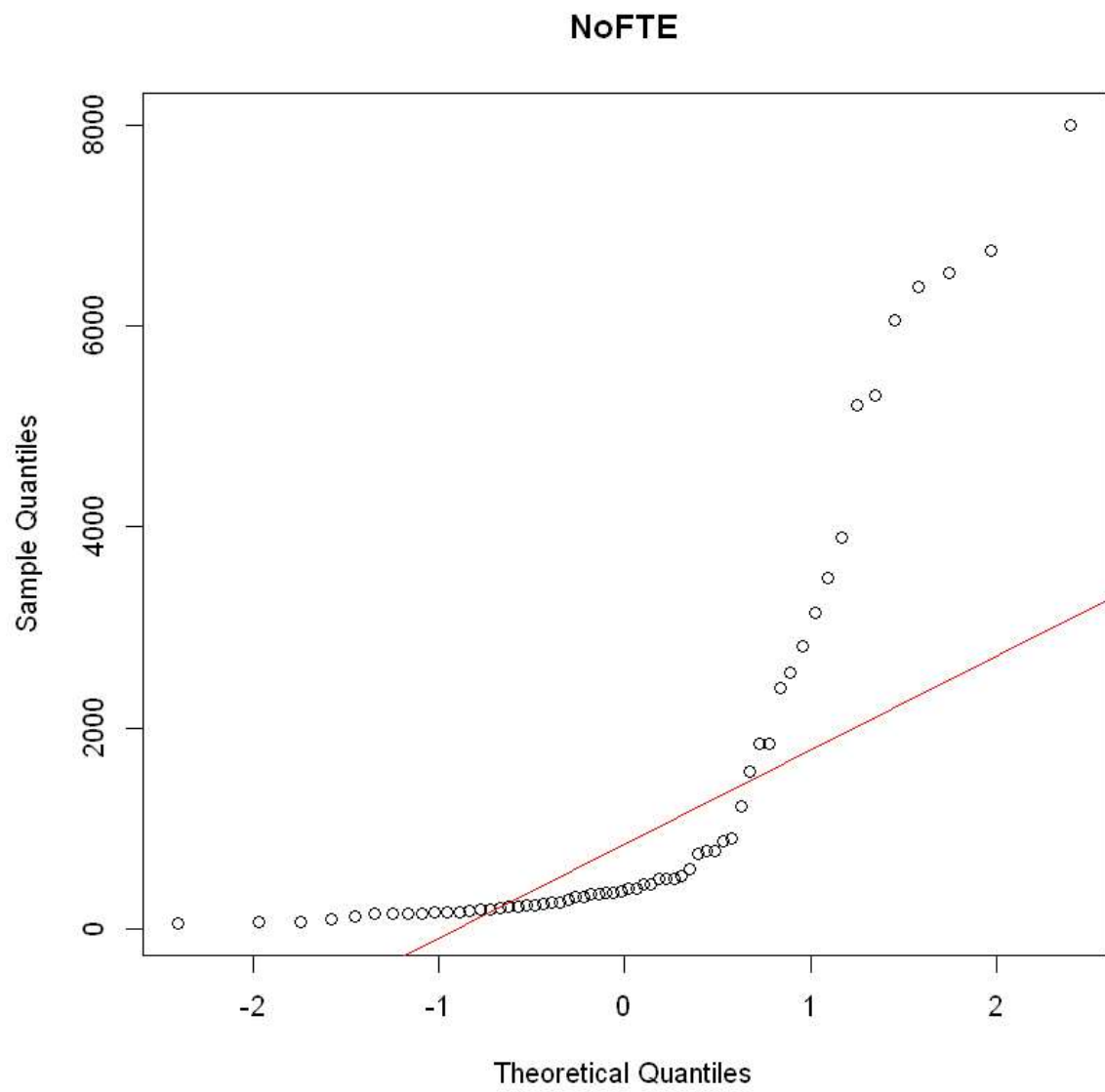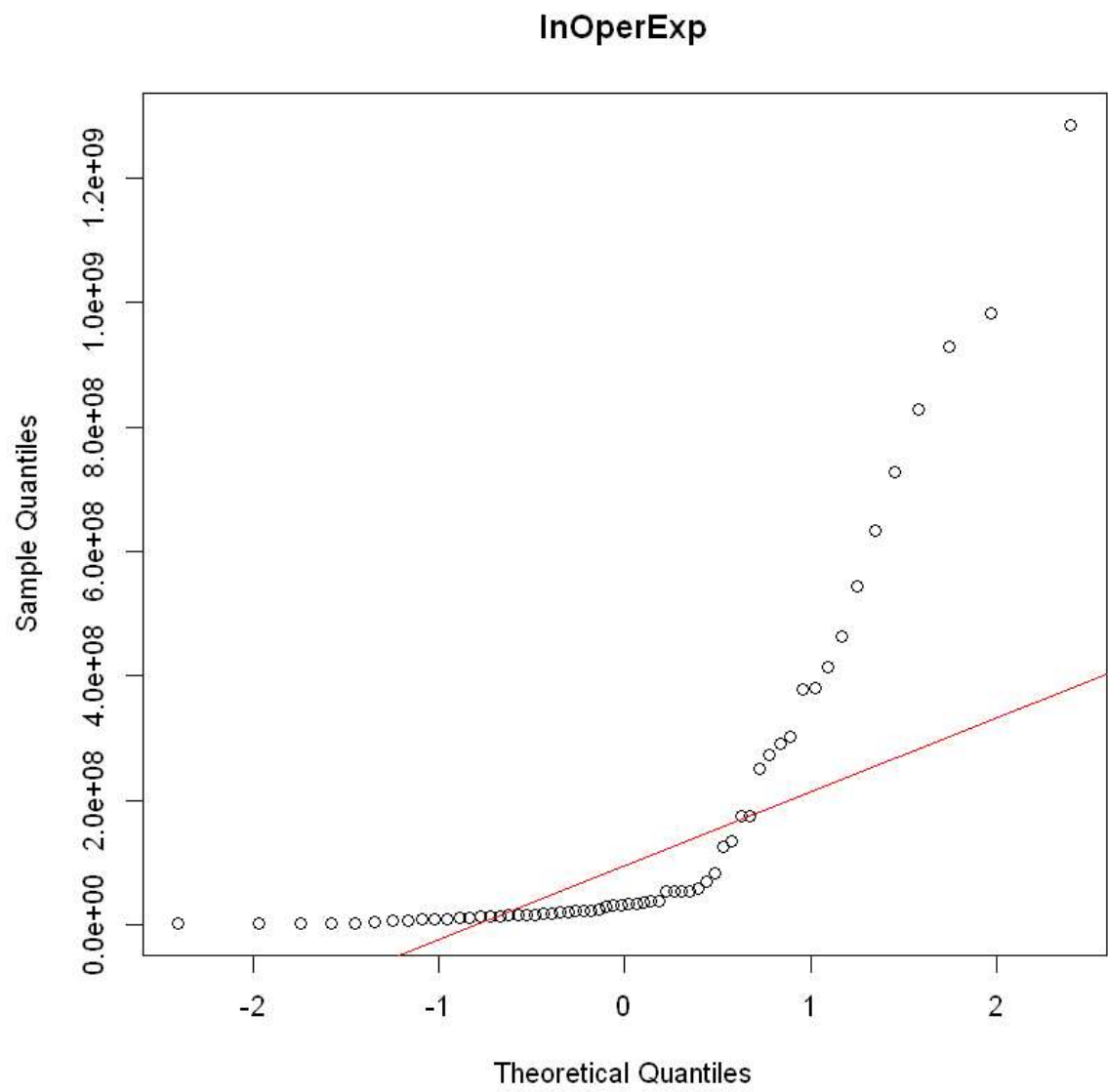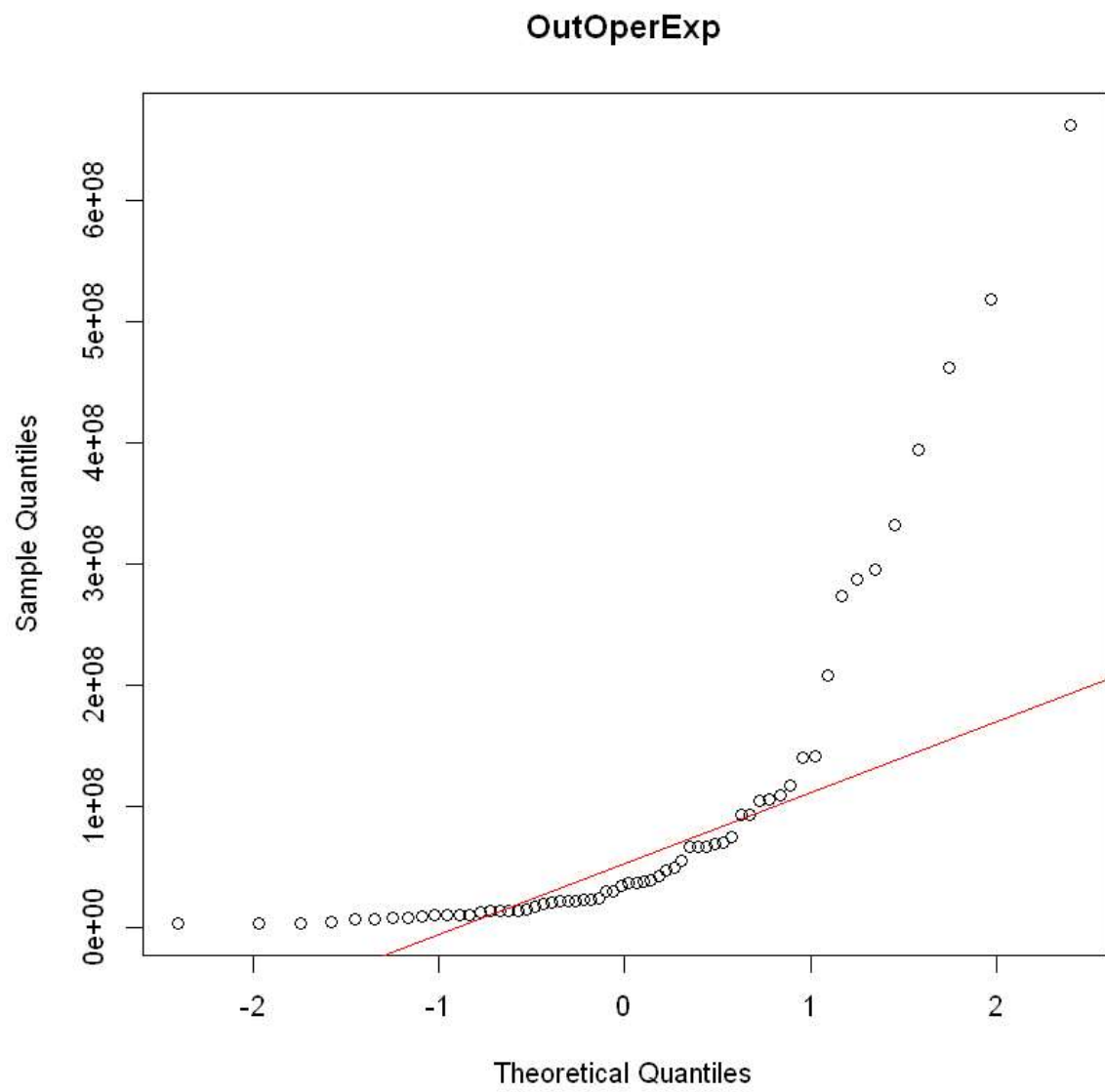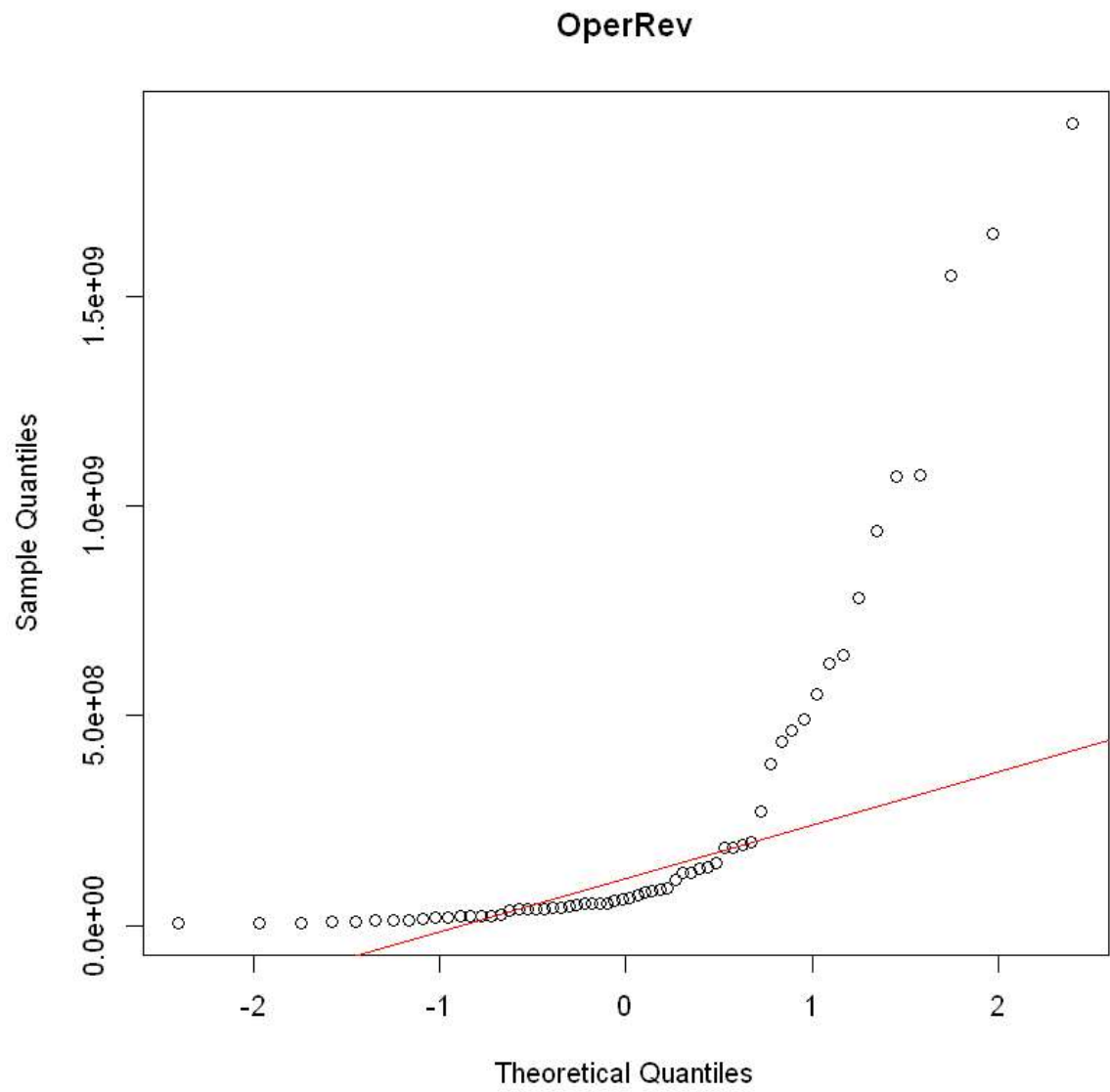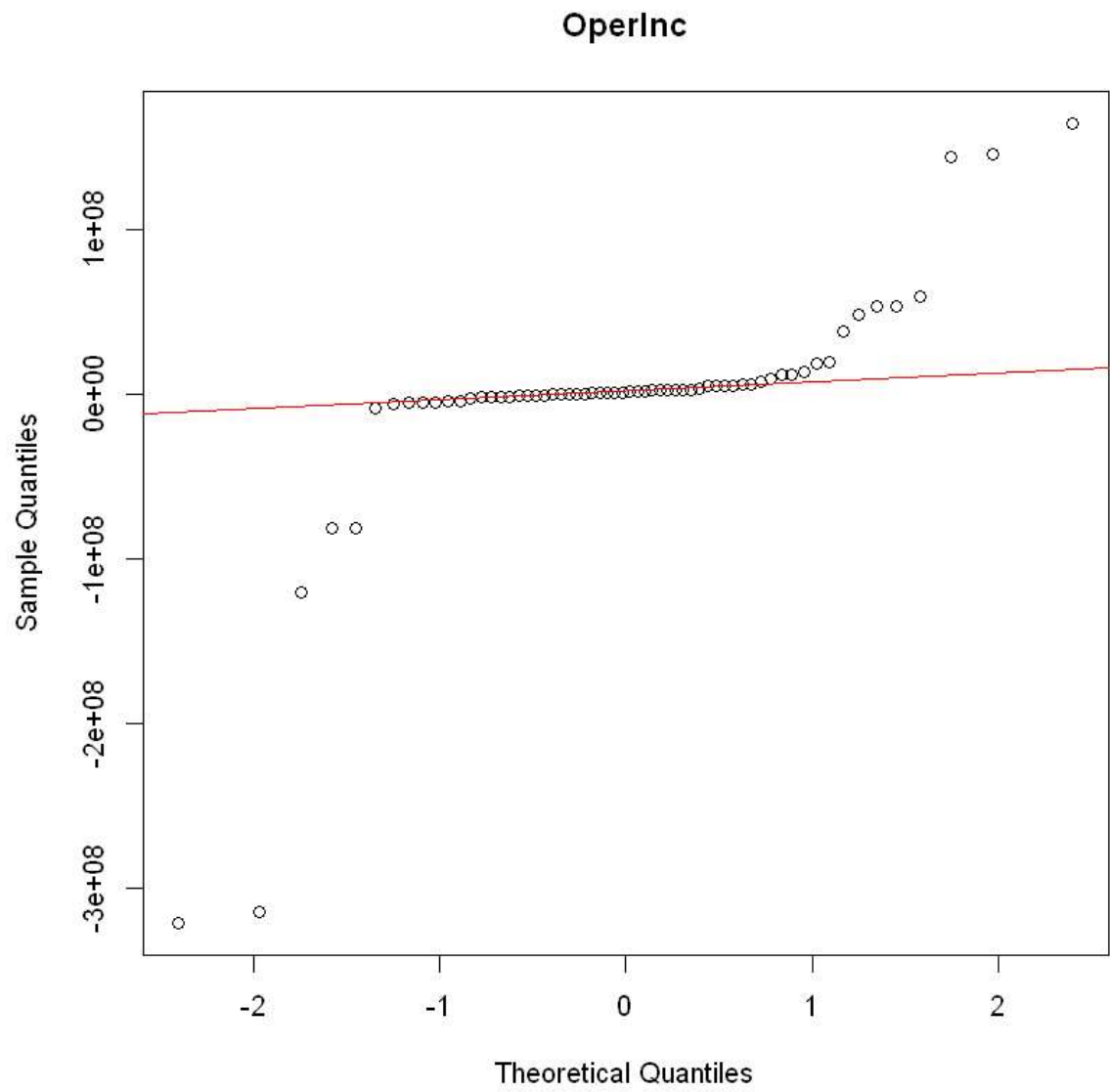
OperInc

AvlBeds
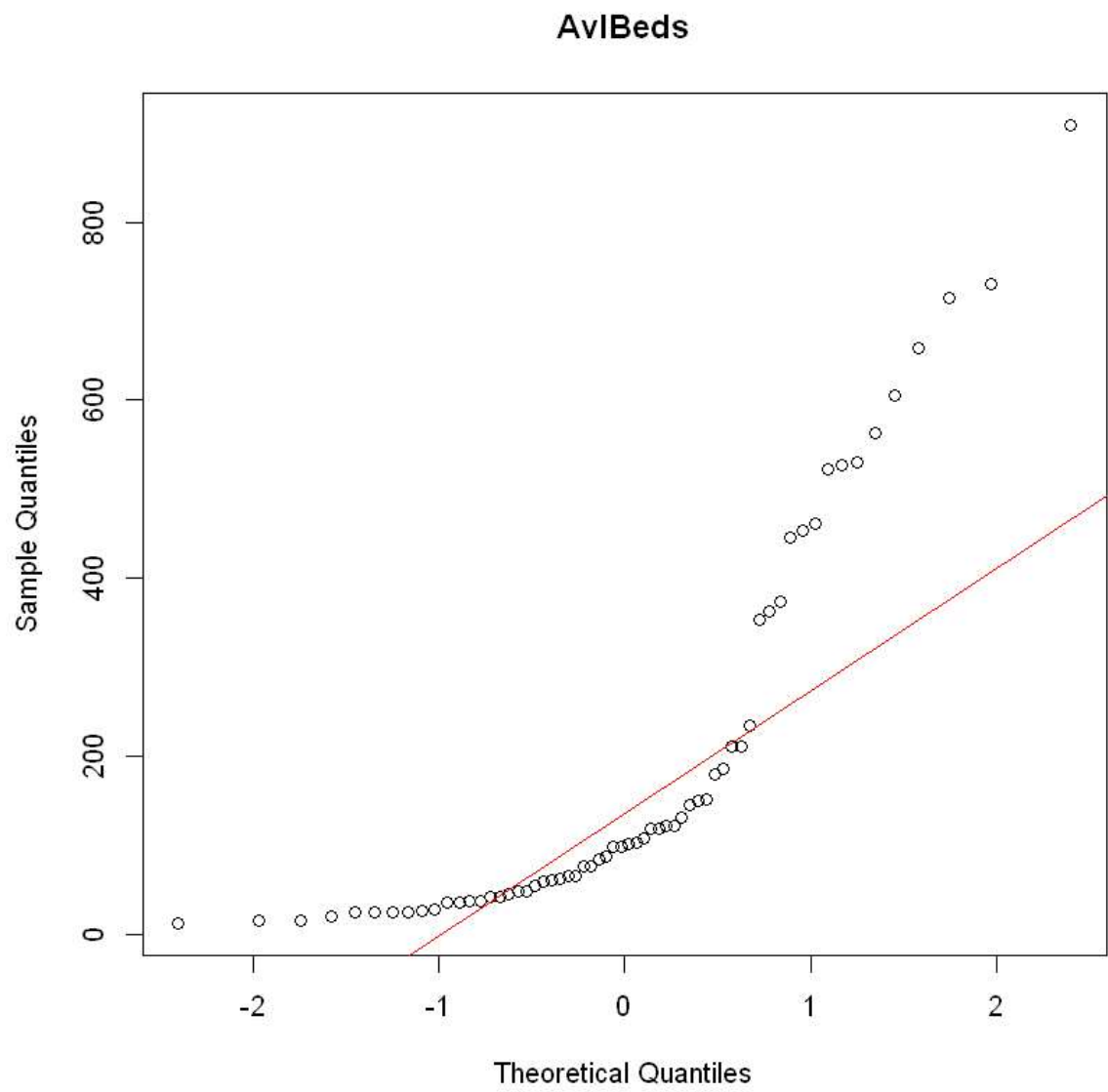
```
In [24]:  shapiro.test(NoFTE)
          shapiro.test(InOperExp)
          shapiro.test(OutOperExp)
          shapiro.test(OperRev)
          shapiro.test(OperInc)
          shapiro.test(AvlBeds)
```

            Shapiro-Wilk normality test

     data: NoFTE
     W = 0.65295, p-value = 7.768e-11

            Shapiro-Wilk normality test

     data: InOperExp
     W = 0.63354, p-value = 3.574e-11

            Shapiro-Wilk normality test

     data: OutOperExp
     W = 0.63921, p-value = 4.47e-11

            Shapiro-Wilk normality test

     data: OperRev
     W = 0.62106, p-value = 2.201e-11

            Shapiro-Wilk normality test

     data: OperInc
     W = 0.58291, p-value = 5.34e-12

   Shapiro-Wilk normality test

     data: AvlBeds
     W = 0.76667, p-value = 1.489e-08

```
In [32]:  names(dfnew)
          unique(PositionTitle)
```

     'LastName' 'FirstName' 'Gender' 'PositionTitle' 'Compensation' 'MaxTerm'
     'StartDate' 'Name' 'Zip' 'Website' 'TypeControl' 'Teaching' 'DonorType'
     'NoFTE' 'NetPatRev' 'InOperExp' 'OutOperExp' 'OperRev' 'OperInc'
   'AvlBeds'

     Acting Director   Regional Representative   Safety Inspection Member
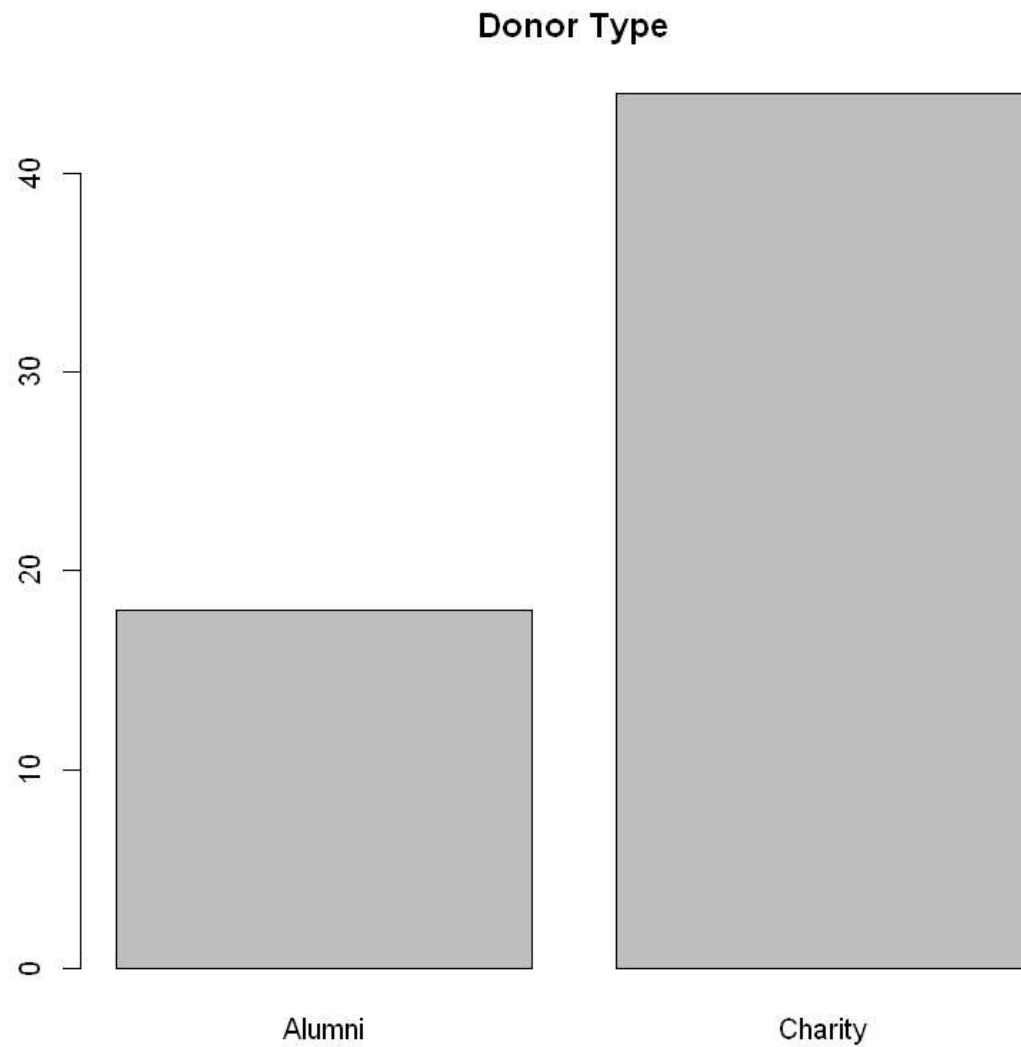   State Board Representative

In [39]:
```
barplot(table(PositionTitle),cex.names = 0.5,horiz =TRUE, beside=TRUE,main =
'Postition Tittle')
barplot(table(TypeControl),horiz =TRUE, beside=TRUE, main = ' Control Type')
barplot(table(Teaching), beside=TRUE, main= 'Teaching Type')
barplot(table(DonorType), beside=TRUE, main = 'Donor Type')
```

Postition Tittle

Control Type

Teaching Type

Donor Type
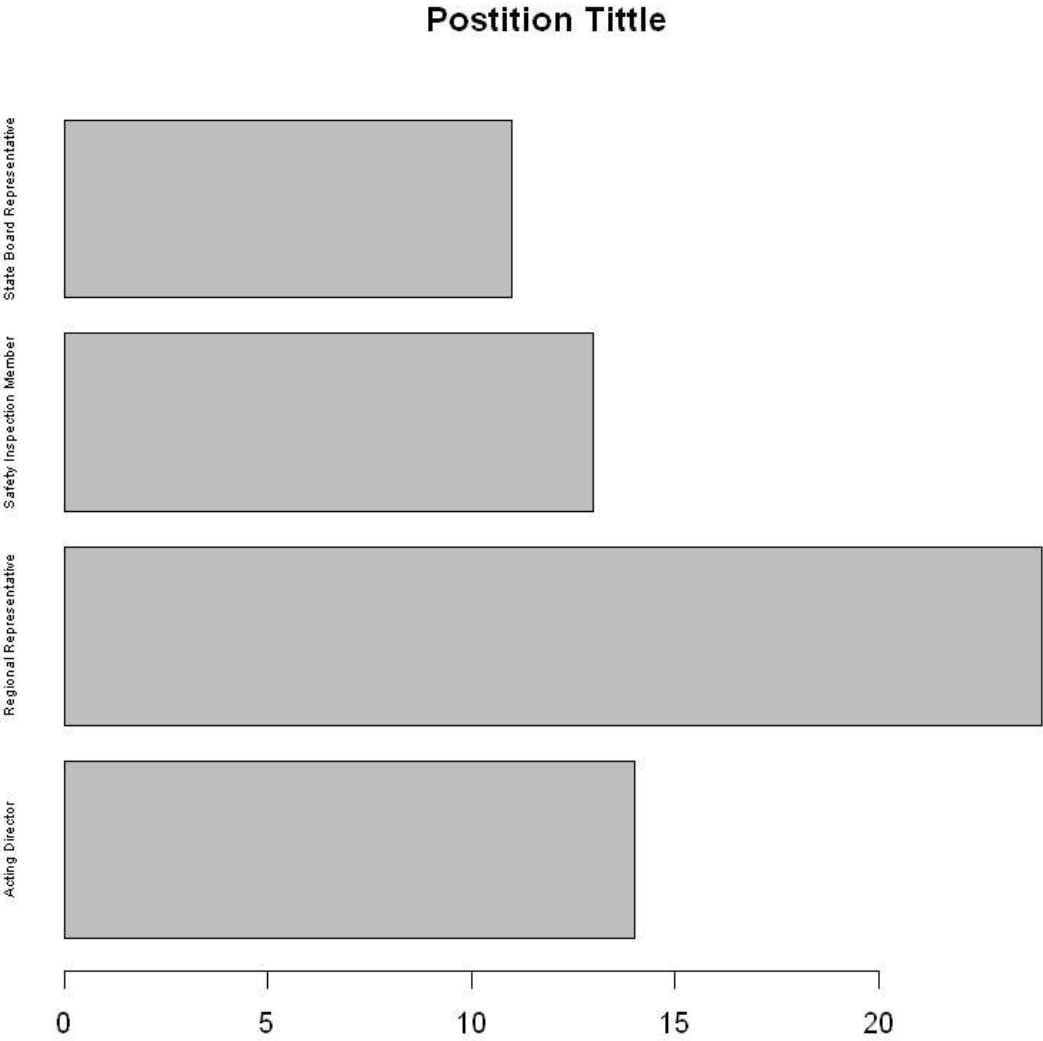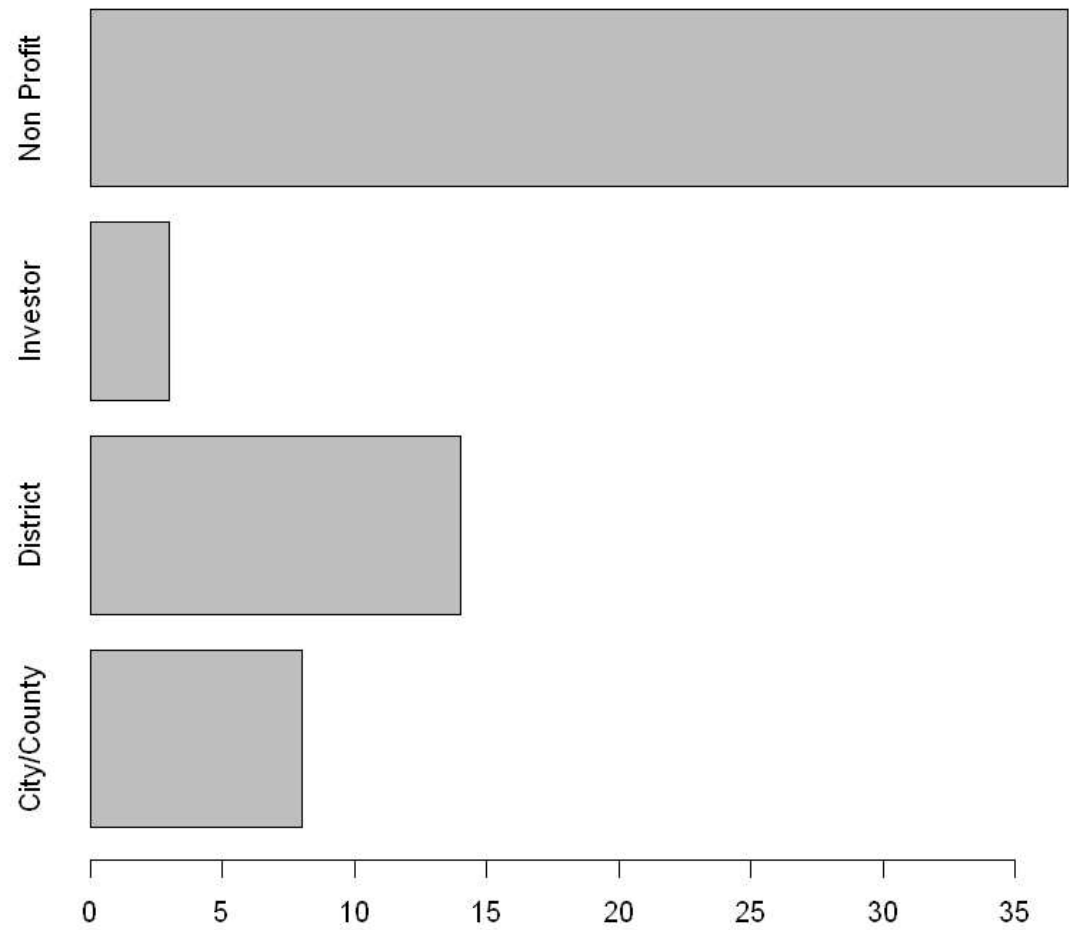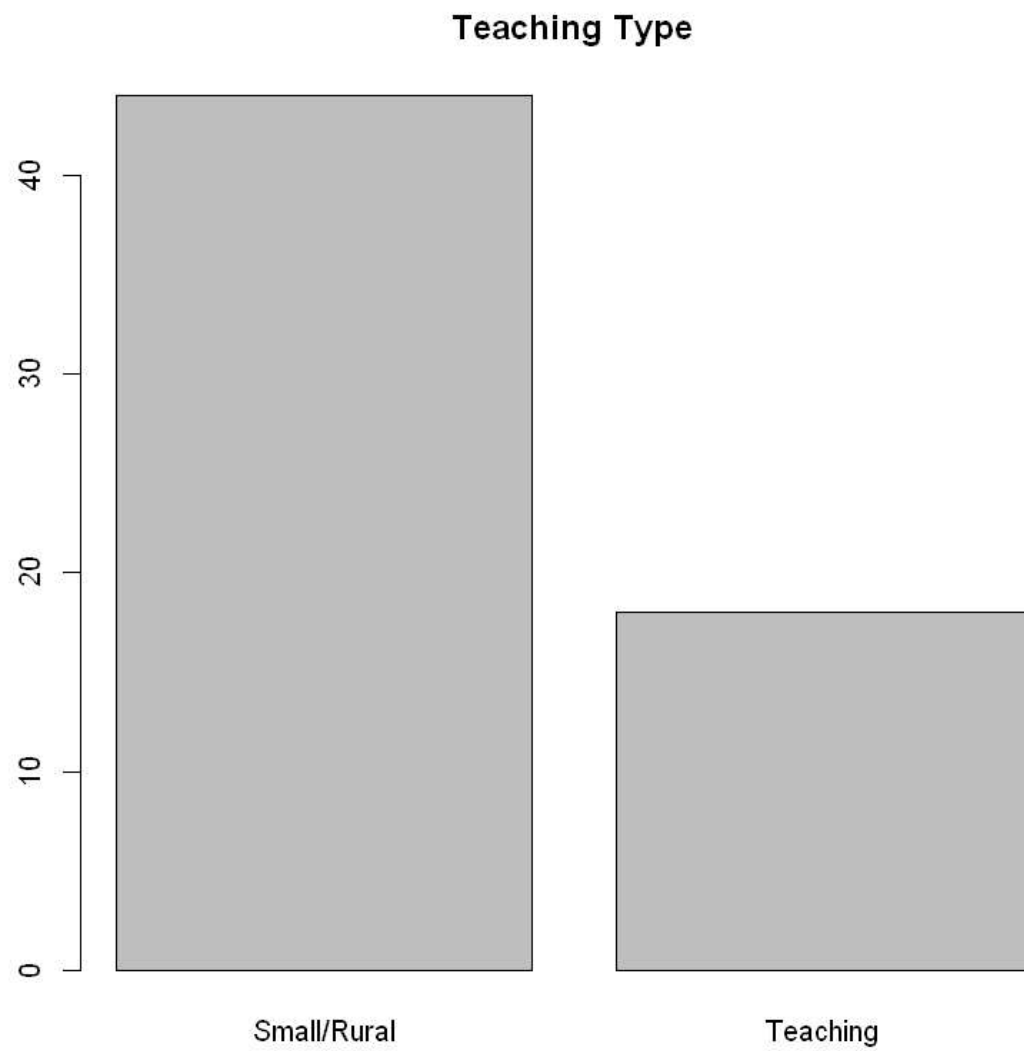
Only operinc variable is likely normaly distribution. Others is not so if non-Normality is an issue my manager may need to use a non-pamatric test