DATA557 - Team Project

Will Wright

February 24, 2019

Exploration

```
# search for NAs
naSummer <- function(column_input) sum(is.na(column_input))
apply(crimeData, 2, naSummer)
```

```
##                Report.Number              Occurred.Date
##                            0                          2
##                Occurred.Time              Reported.Date
##                            2                          0
##                Reported.Time           Crime.Subcategory
##                            2                          0
## Primary.Offense.Description                   Precinct
##                            0                          0
##                       Sector                       Beat
##                            0                          0
##                 Neighborhood               reportedYear
##                            0                          0
##                 occurredYear              reportedMonth
##                            2                          0
##               occurredMonth
##                            2
```

```
# show data with NAs
kable(rbind(
  crimeData[which(is.na(crimeData$Reported.Date)),],
  crimeData[which(is.na(crimeData$Occurred.Date)),])
  )
```

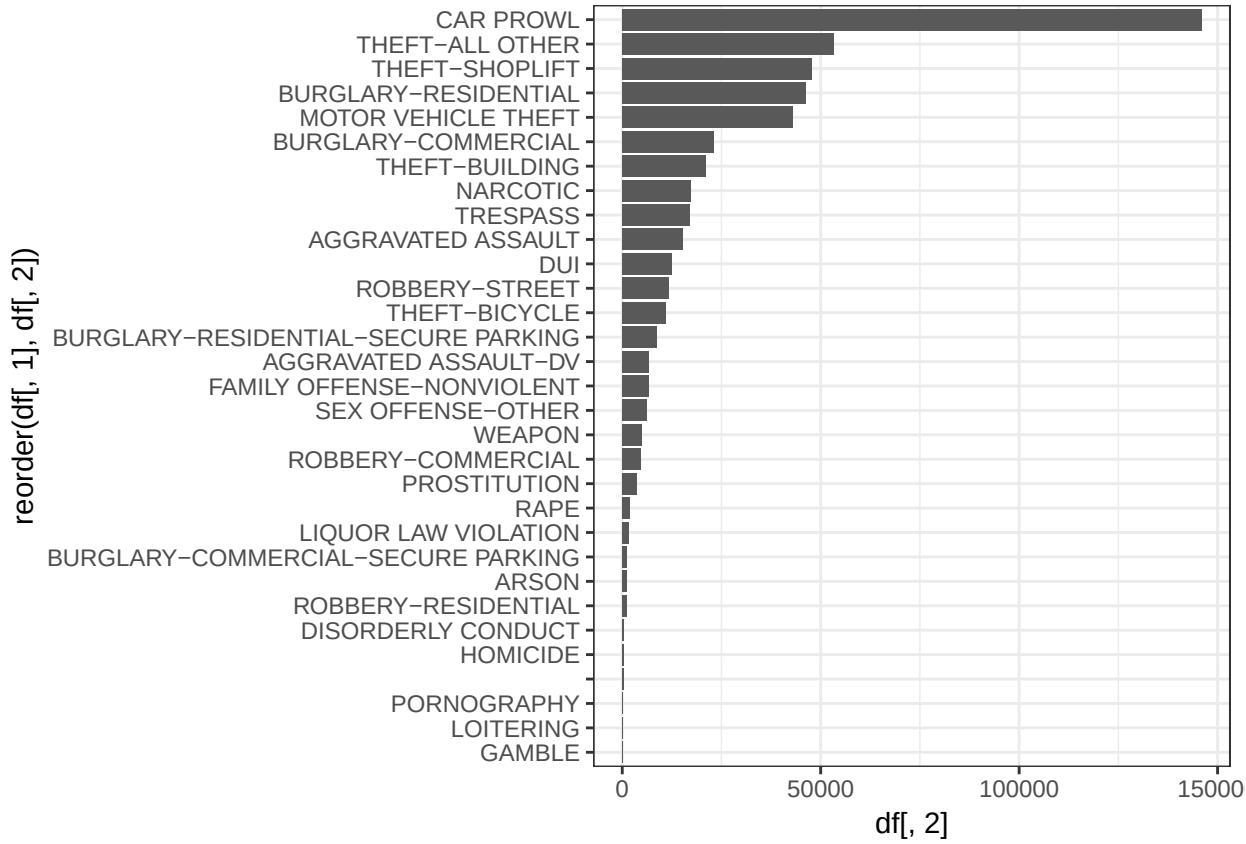
	Report.Number	Occurred.Date	Occurred.Time	Reported.Date	Reported.Time	Crime.Subcategory	Pri
10	1.999e+13	NA	NA	1999-01-01	NA	THEFT-SHOPLIFT	TH
123	2.007e+13	NA	NA	2007-01-01	NA	RAPE	RA

Looks like there are two NA values for dates. I recommend excluding for date-based analyses.

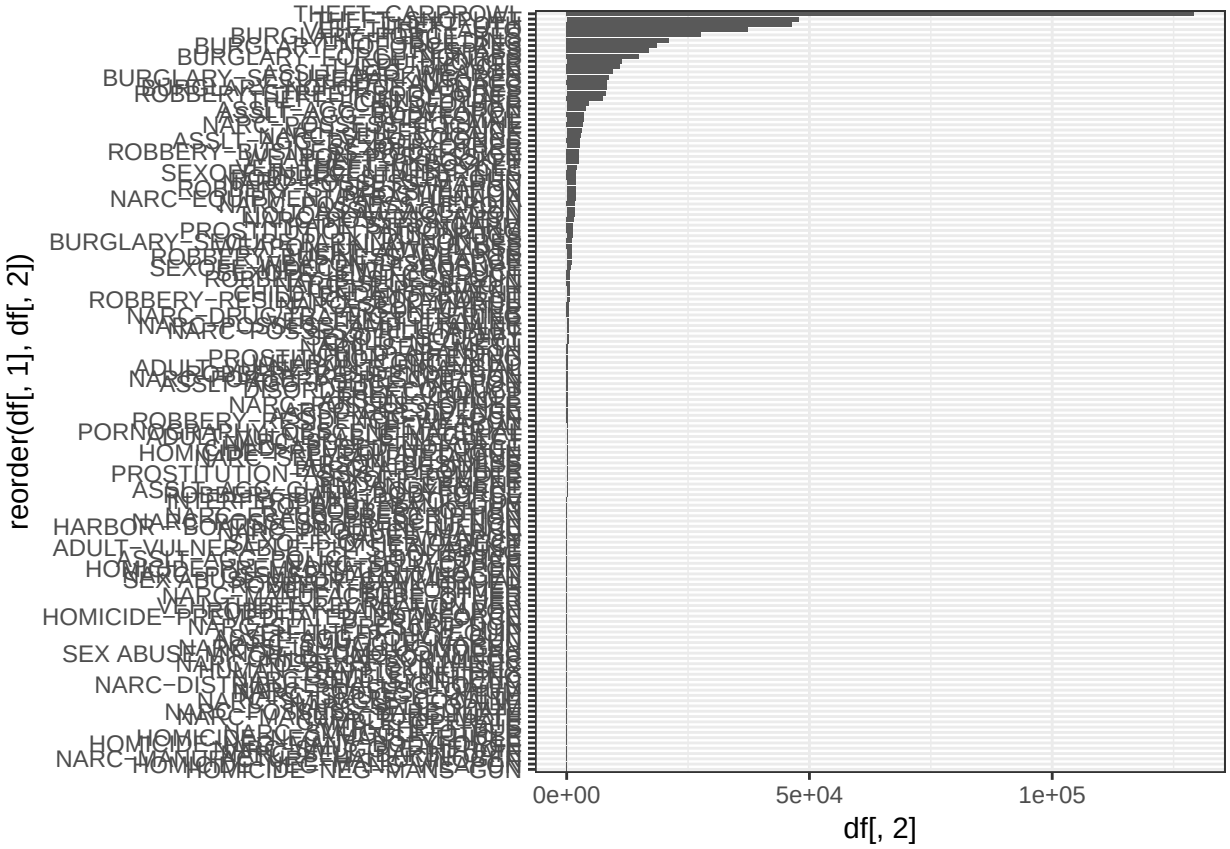
```
discrete_distVizer <- function(column_name){

  df <- as.data.frame(table(column_name))
  df[,2] <- as.integer(df[,2])
  df <- arrange(df, desc(df[,2]))
  g <- ggplot(df, aes(x = reorder(df[,1], df[,2]), y = df[,2]))
  g+ geom_bar(stat = "identity") +
    coord_flip() +
    theme_bw()
}

apply(crimeData[,which(names(crimeData) %in% c("Crime.Subcategory",
                                               "Primary.Offense.Description",
                                               "Precinct",
```

```
                                                "Sector",
                                                "Beat",
                                                "Neighborhood",
                                                "reportedYear",
                                                "occurredYear",
                                                "reportedMonth",
                                                "occurredMonth"))],
     2, discrete_distVizer)
```

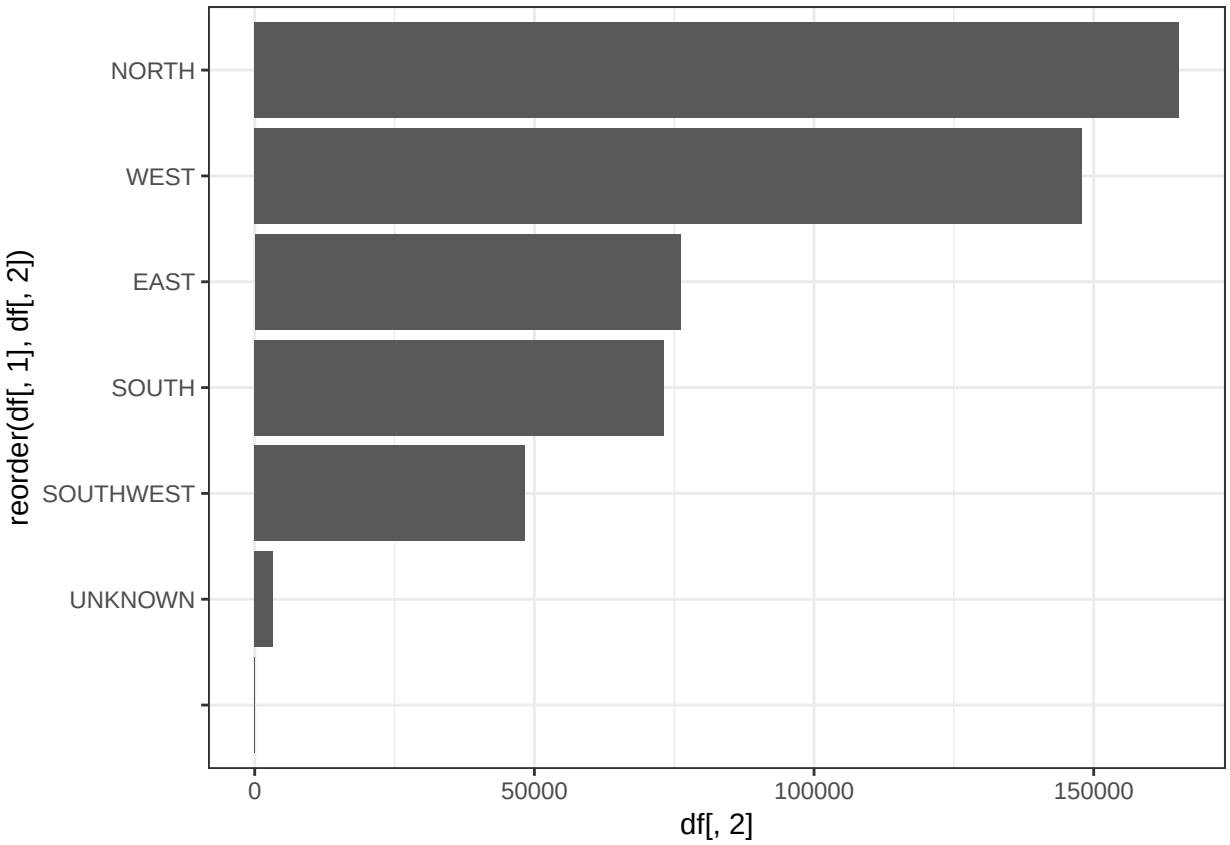
```
## $Crime.Subcategory
```



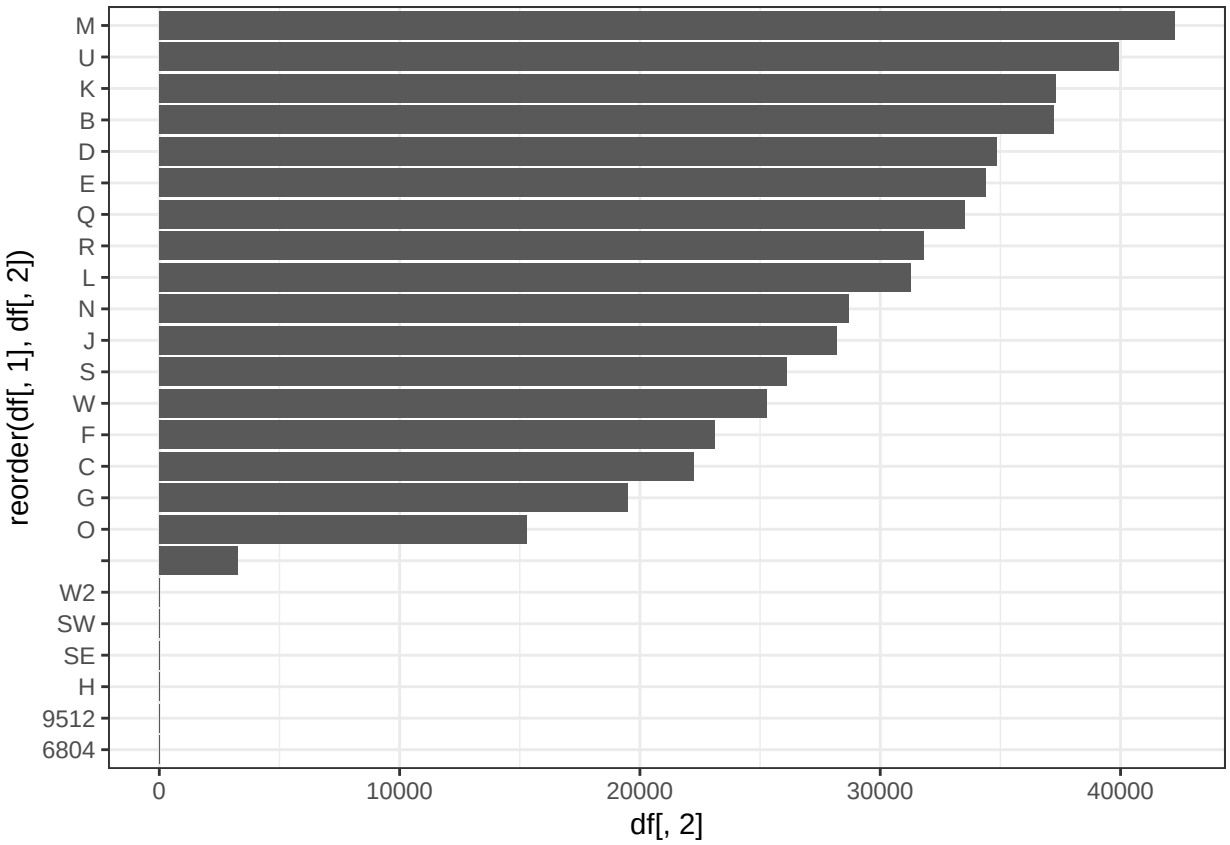

```
##
## $Primary.Offense.Description
```


reorder(df[, 1], df[, 2])
THEFT-CARPROWL
BURGLARY-RESIDENCE
BURGLARY-FORCIBLE
BURGLARY-SAGEBRUSH
ROBBERY-STREET
ASSLT-AGG-BODYFORCE
ASSLT-AGG-WEAPON
ROBBERY-BUSINESS
SEXOFF-RAPE
NARC-EQUIPMENT
BURGLARY-SECURE
PROSTITUTION-PATRON
ROBBERY-RESIDENCE
NARC-DRUGPARAPHERNALIA
NARC-POSSESS
ADULT-VULNERABLE
PORNOGRAPHY-PRODUCE
HOMICIDE-ISOLATED
PROSTITUTION-PANDER
ASSLT-AGG-GUN
HARBOR-BORDER
ADULT-VULNERABLE
HOMICIDE-PREMEDITATED
SEX ABUSE-CHILD
NARC-DISTRIBUTE
HOMICIDE-NEG MANS-GUN
0e+00
5e+04
1e+05
df[, 2]

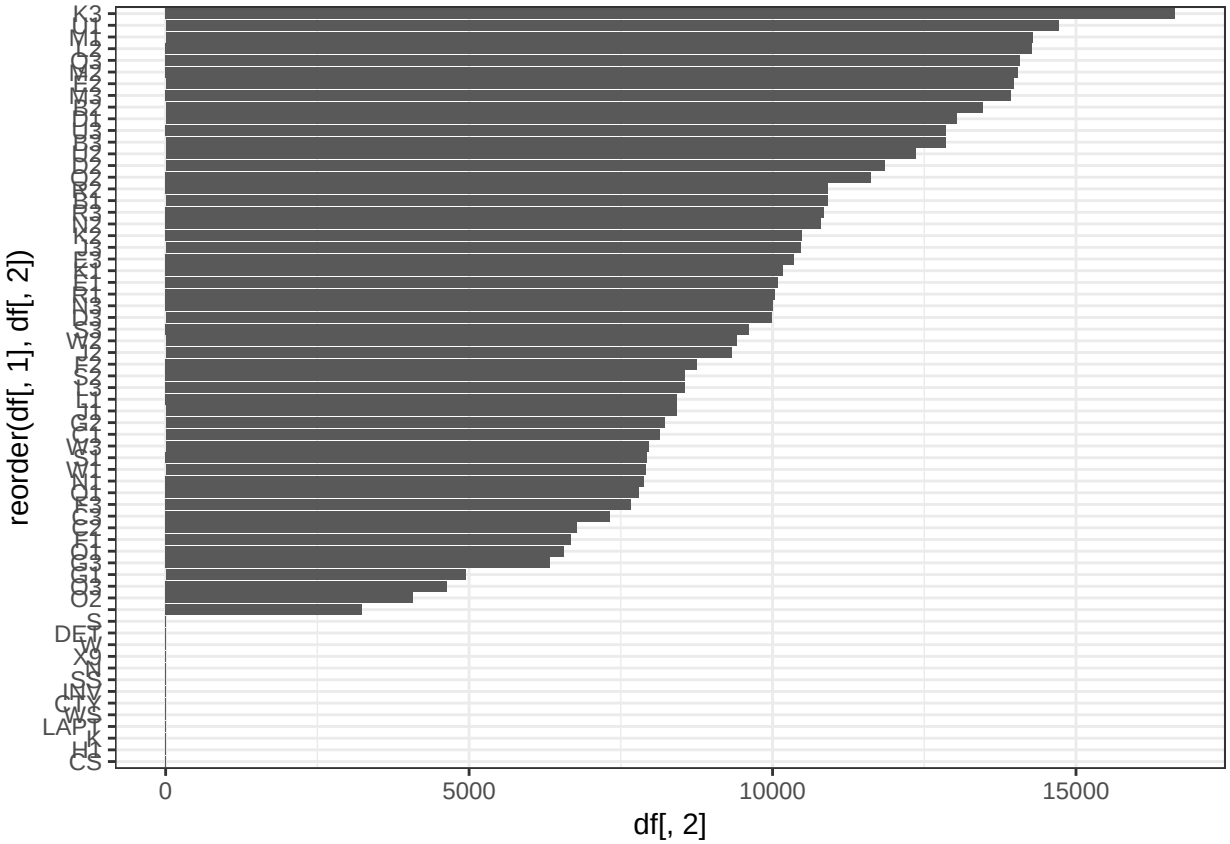
```
##
## $Precinct
```


NORTH
WEST
EAST
SOUTH
SOUTHWEST
UNKNOWN
reorder(df[, 1], df[, 2])
0
50000
100000
150000
df[, 2]

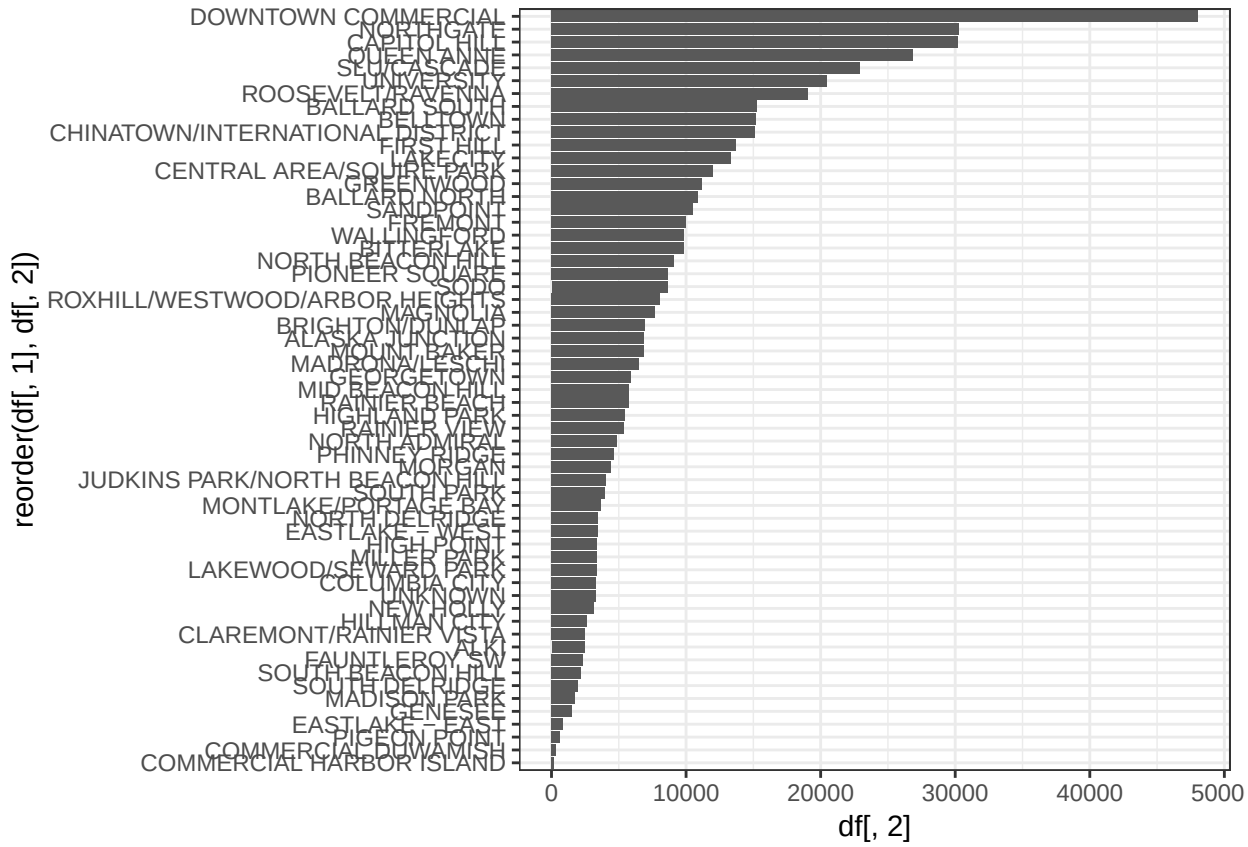
```
##
## $Sector
```


M
U
K
B
D
E
Q
R
L
N
J
S
W
F
C
G
O
W2
SW
SE
H
9512
6804
reorder(df[, 1], df[, 2])
0
10000
20000
30000
40000
df[, 2]

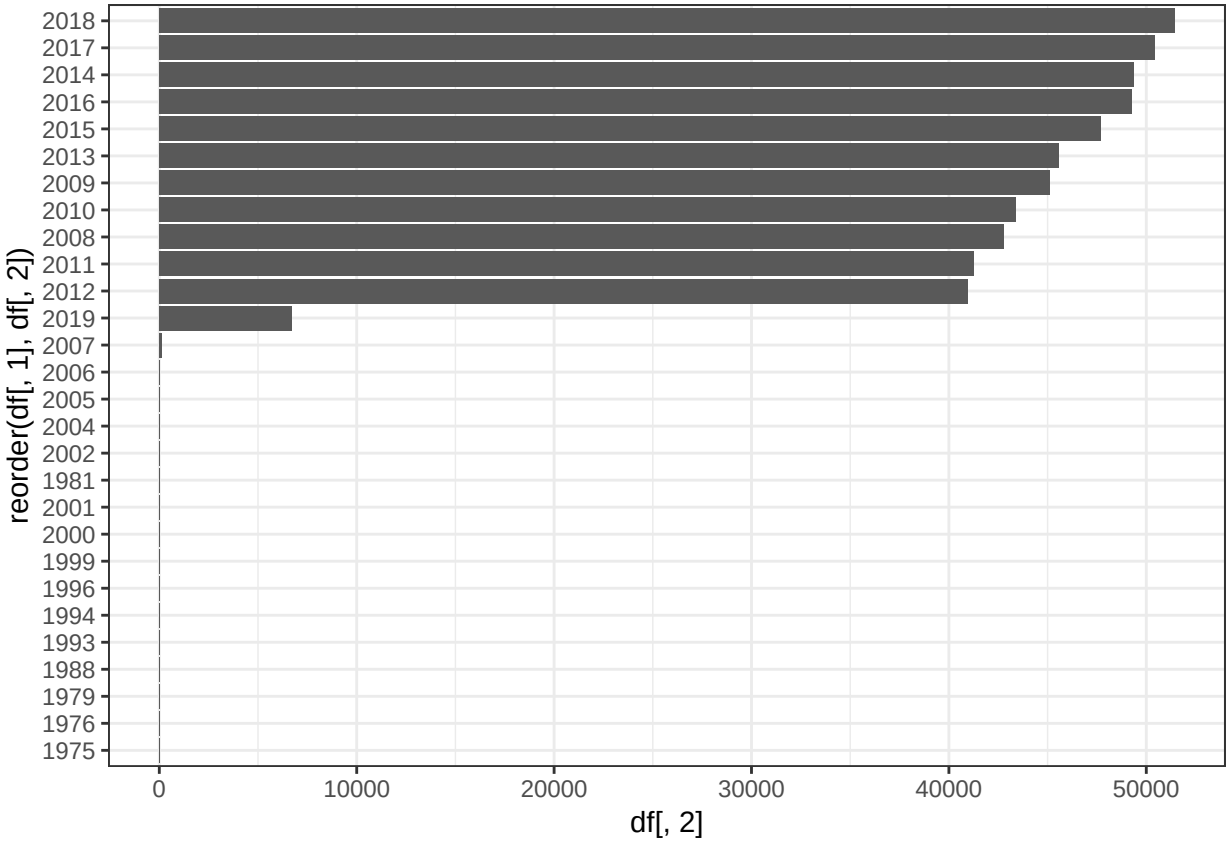
```
##
## $Beat
```



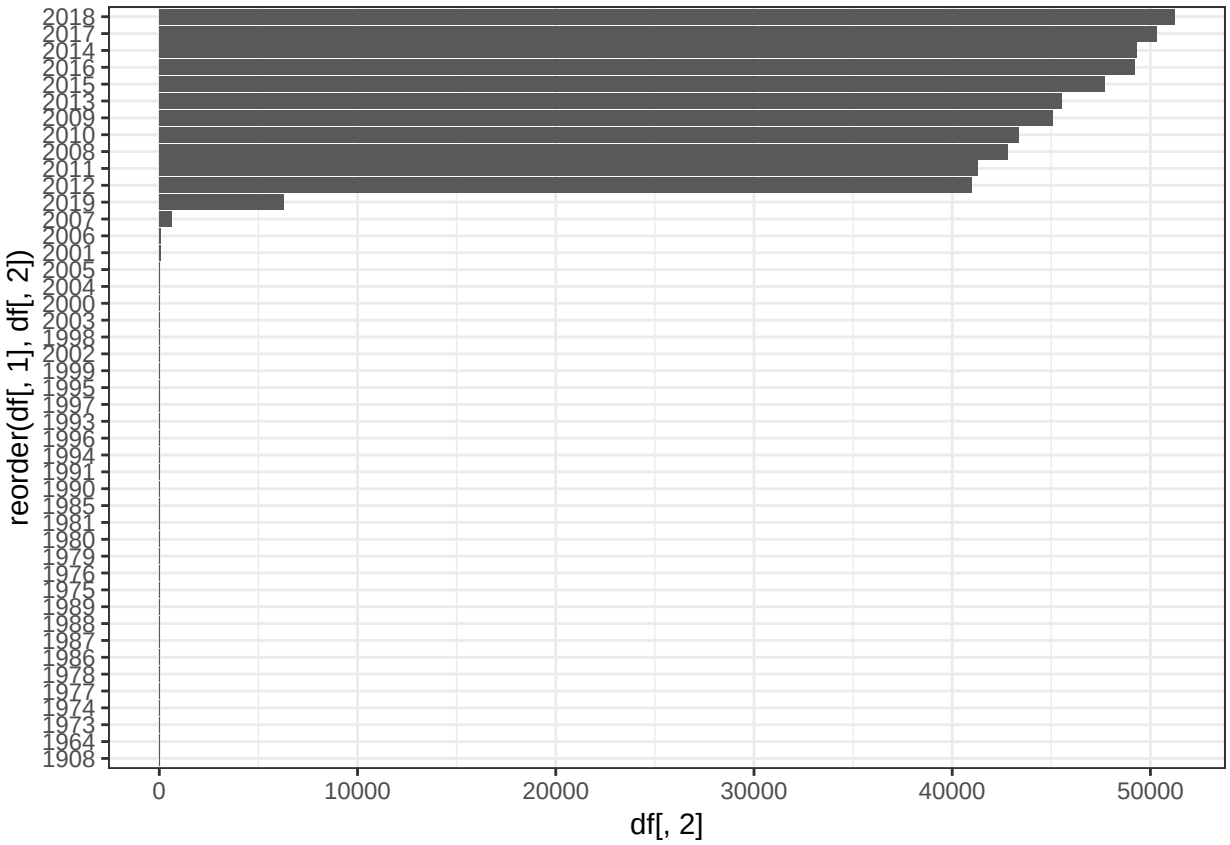

```
##
## $Neighborhood
```



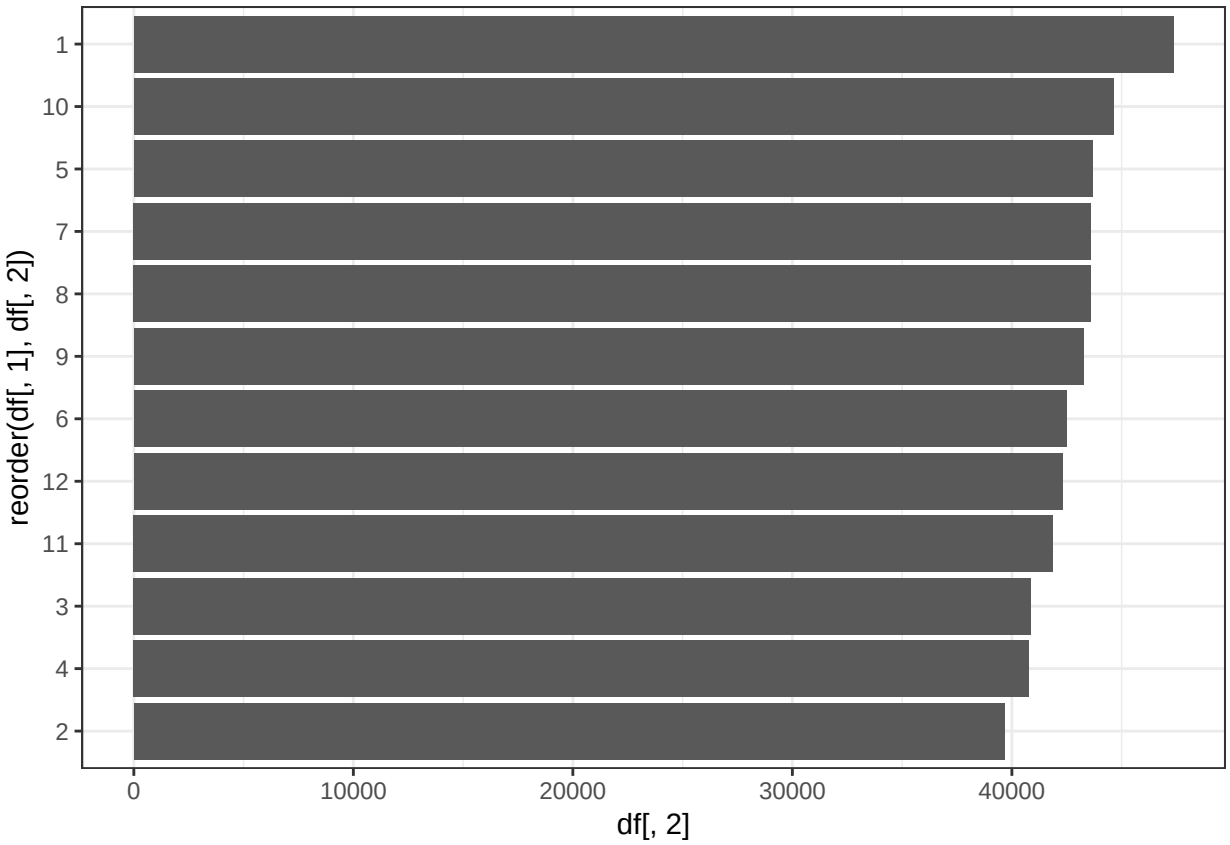

```
##
## $reportedYear
```



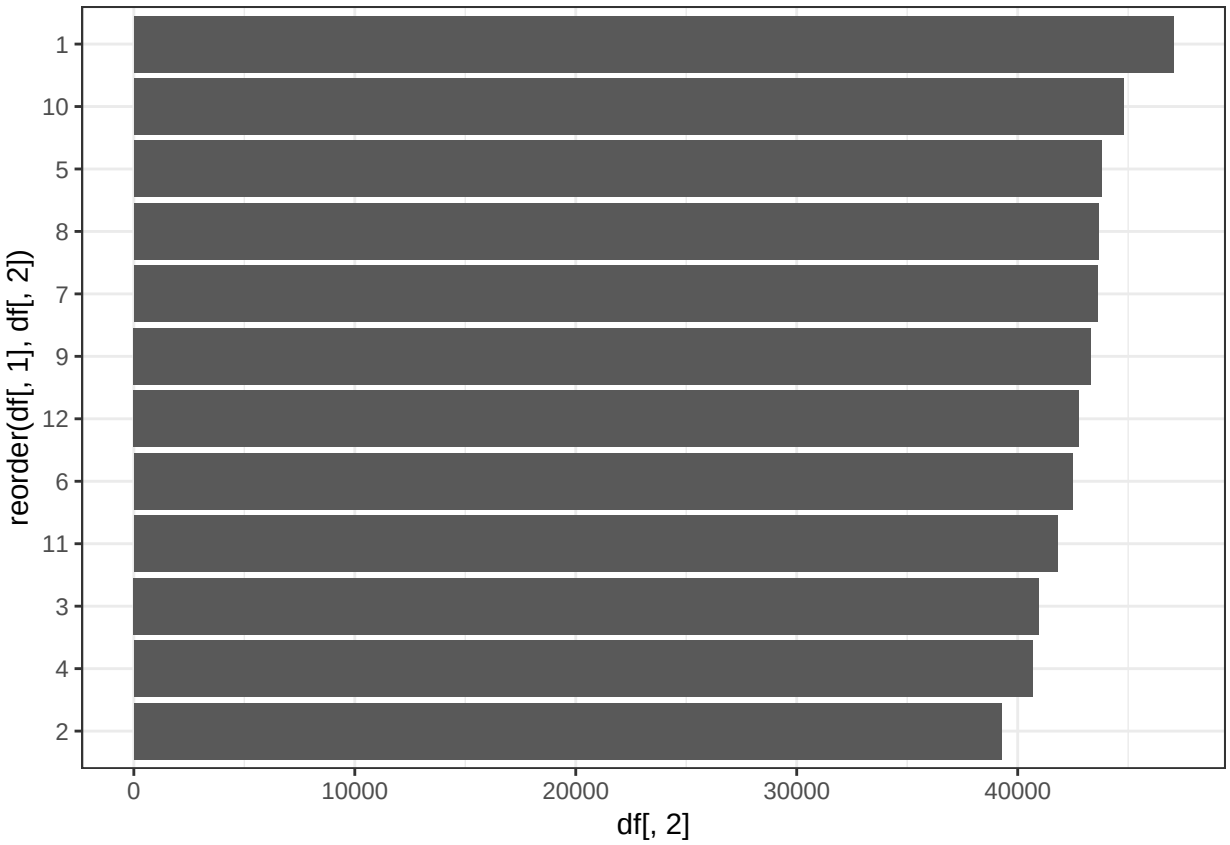

```
##
## $occurredYear
```




```
##
## $reportedMonth
```




```
##
## $occurredMonth
```


reorder(df[, 1], df[, 2])
1
10
5
8
7
9
12
6
11
3
4
2
0
10000
20000
30000
40000
df[, 2]