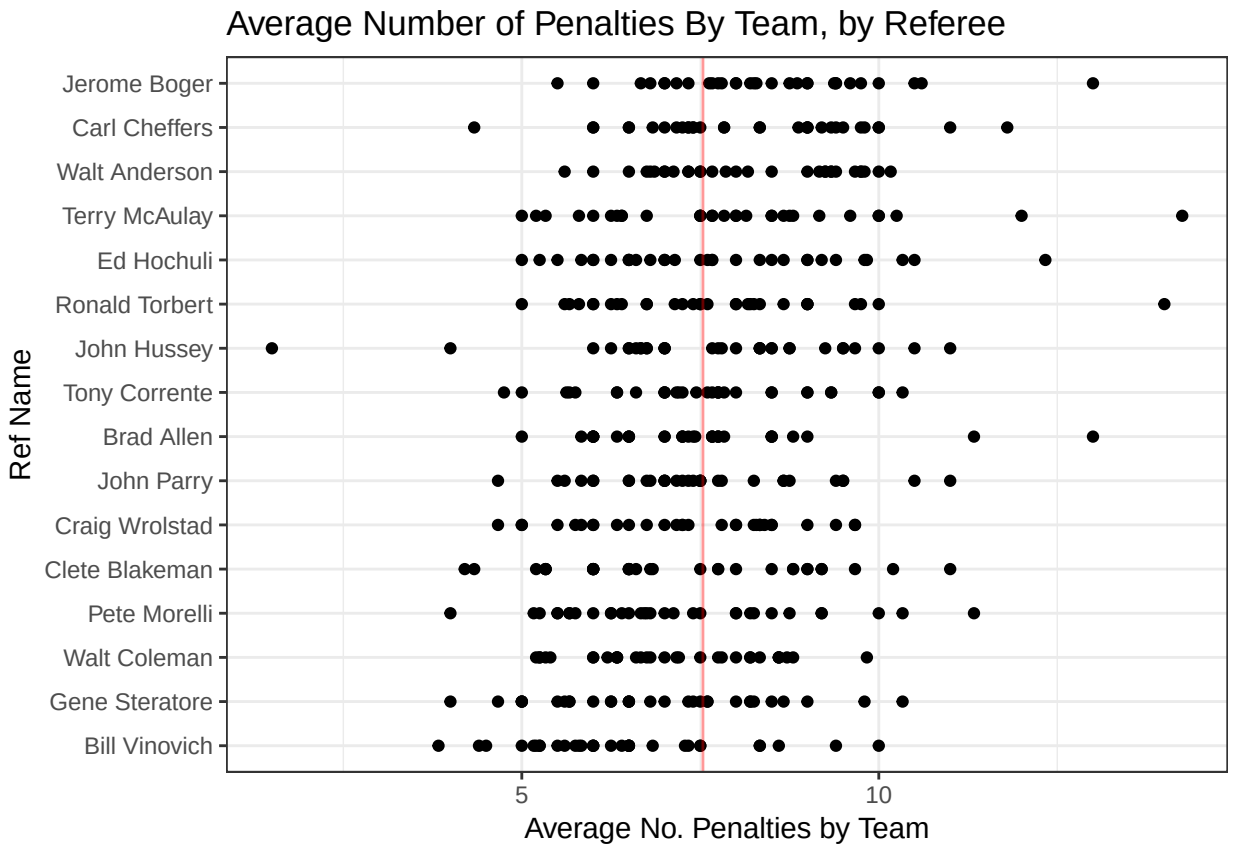
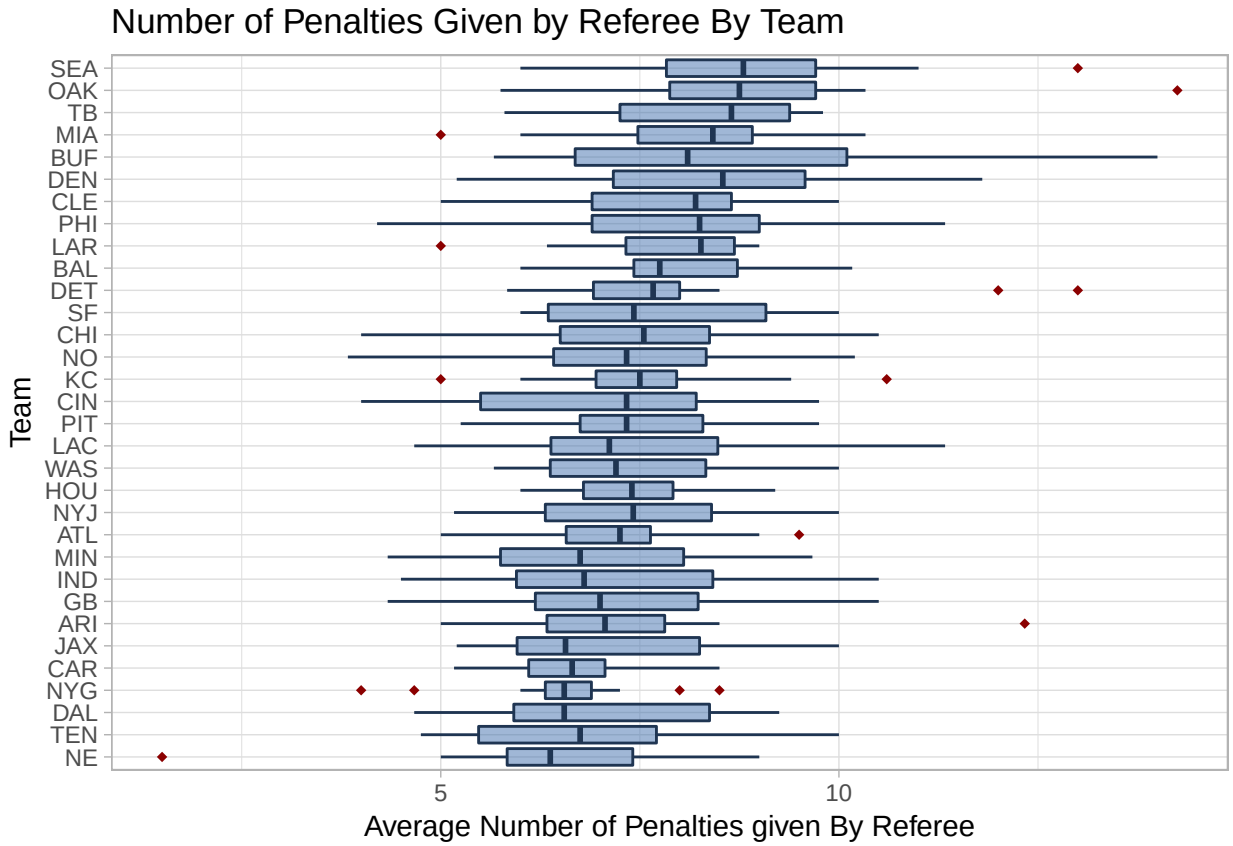
Penalties_byTeam_byReferee

```
#NOW WE LOOK BY REFEREE BY TEAM

penRef<-refData %>%
  group_by(Referee,PenaltyTeam) %>%
  summarize(n=n())

#penRef

penaltiesPerGame<-refData %>%
  group_by(GameDate,PenaltyTeam) %>%
  summarize(n=n())


colnames(penRef)<-c("Referee","PenaltyTeam","NPenalties")


#So we have the number of penalties by ref...now we just need the number of games each ref did for each
#The best way that we can do this is through using two DPYLR statements, one to get unique games by tea
#to summarize by ref by team

refDataTemp<-refData
NoGamesRef<- refDataTemp %>%
  filter(!is.na(PenaltyTeam)) %>%
  group_by(Referee,PenaltyTeam,GameDate) %>%
  summarize(Unique_Elements = n_distinct(GameDate))

NoGamesRef <- NoGamesRef%>%
  group_by(Referee,PenaltyTeam) %>%
  summarize(n=n())

colnames(NoGamesRef)<-c("Referee","PenaltyTeam","NGames")

penRef<-merge(x=penRef,y=NoGamesRef)
penRef$PenPerGame <- penRef$NPenalties/penRef$NGames


temP<-penRef%>%
  group_by(PenaltyTeam)%>%
  dplyr::summarize(Median = median(PenPerGame))

#temP
penRef<-merge(x=penRef,y=temP)

penRef<-penRef[complete.cases(penRef),]

ggplot(penRef, aes(x=reorder(Referee,PenPerGame),y=PenPerGame))+geom_point()+coord_flip()+ggtitle("Avera
```



The above is data prep.

```
fill <- "#4271AE"
line <- "#1F3552"

boxPlotTeam<-ggplot(penRef, aes(x=reorder(PenaltyTeam,Median),y=PenPerGame))+geom_boxplot(fill=fill,col
boxPlotTeam
```




```
#interactiveBoxTeam<-ggplotly(boxPlotTeam)
#interactiveBoxTeam
```

In the chart above, we are looking at the distribution of the average number of penalties per Referee. IE, for each team there will be 16 data points as we are looking at 16 referees. Each of the referees will have an average penalties per game for each time, which is what we are charting above. Note that in the absolute majority of cases, referee's call a similar number of penalties; however there are a few cases where referees call many more or many less penalties against a team. Note that some teams have a huge variance of penalties called against, such as Buffalo while others, such as the New York Giants, have a very small variance in the average number of penalties called. Regardless, what we are truly interested in within this chart are the outliers. Having referees make calls that are expected is not something that we are looking for, and as such a closer look at the outliers is below.

```
is_outlier <- function(x) {
  return(x < quantile(x, 0.25) - 1.5 * IQR(x) | x > quantile(x, 0.75) + 1.5 * IQR(x))
}

#penRef %>%
#  group_by(PenaltyTeam) %>%
#  mutate(outlier = ifelse(is_outlier(PenPerGame), Referee,NA)) %>%
#  ggplot(., aes(x = factor(PenaltyTeam), y = PenPerGame)) +
#    geom_boxplot() +
#    geom_text(aes(label = outlier), na.rm = TRUE, hjust = -0.3)
```

```
#Buffalo (BUF), Oakland (OAK), Detroit (DET), NYG seem like interesting ones
#SEA,KC, ATL,ARI,NE
```

```
BUFPenRef<-penRef[penRef[,1]=='BUF',]

OAKPenRef<-penRef[penRef[,1]=='OAK',]

NYGPenRef<-penRef[penRef[,1]=='NYG',]

DETPenRef<-penRef[penRef[,1]=='DET',]

NEPenRef<-penRef[penRef[,1]=='NE',]

KCPenRef<-penRef[penRef[,1]=='KC',]

ATLPenRef<-penRef[penRef[,1]=='ATL',]

ARIPenRef<-penRef[penRef[,1]=='ARI',]

SEAPenRef<-penRef[penRef[,1]=='SEA',]
SEAPenRef
```

```
##     PenaltyTeam        Referee NPenalties NGames PenPerGame Median
## 452         SEA   Walt Coleman         36      5   7.200000    8.9
## 453         SEA Gene Steratore         68      8   8.500000    8.9
## 454         SEA  Walt Anderson         58      6   9.666667    8.9
## 455         SEA  Bill Vinovich         30      5   6.000000    8.9
## 456         SEA  Carl Cheffers         47      6   7.833333    8.9
## 457         SEA     Brad Allen         44      5   8.800000    8.9
## 458         SEA Ronald Torbert         39      4   9.750000    8.9
## 459         SEA    John Hussey         37      4   9.250000    8.9
## 460         SEA   Jerome Boger         13      1  13.000000    8.9
## 461         SEA     John Parry         29      4   7.250000    8.9
## 462         SEA  Terry McAulay         48      5   9.600000    8.9
## 463         SEA  Tony Corrente         47      6   7.833333    8.9
## 464         SEA Clete Blakeman         22      2  11.000000    8.9
## 465         SEA   Pete Morelli         32      4   8.000000    8.9
## 466         SEA     Ed Hochuli         59      6   9.833333    8.9
```

```
subsetTeams <- rbind(KCPenRef,OAKPenRef)
subsetTeams <- rbind(subsetTeams,NYGPenRef)
subsetTeams <- rbind(subsetTeams,DETPenRef)
subsetTeams <- rbind(subsetTeams,NEPenRef)
subsetTeams <- rbind(subsetTeams,ARIPenRef)
subsetTeams <- rbind(subsetTeams,SEAPenRef)
subsetTeams <- rbind(subsetTeams,ATLPenRef)

subsetTeams
```

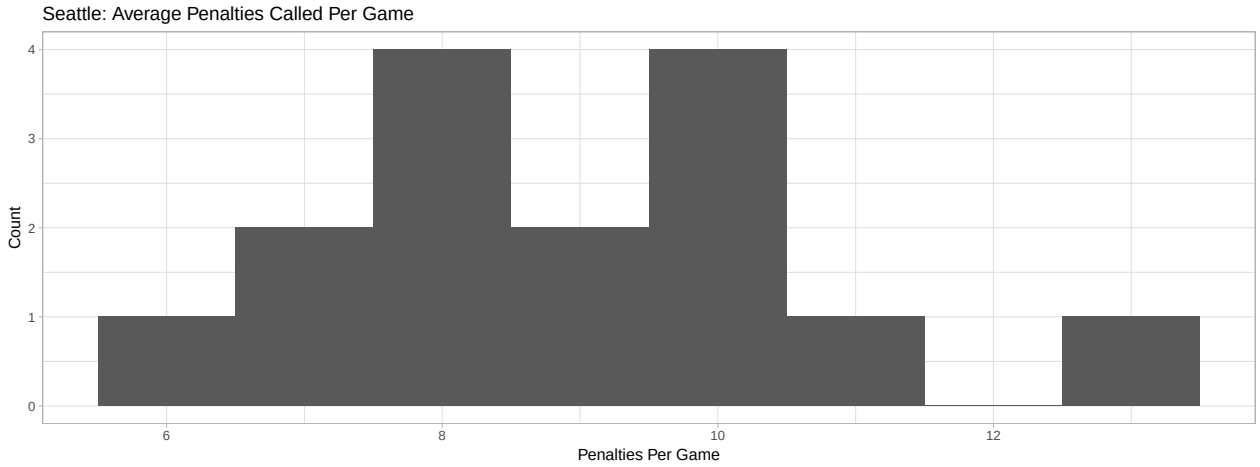
```
##     PenaltyTeam        Referee NPenalties NGames PenPerGame   Median
## 251          KC  Walt Anderson         45      6   7.500000 7.500000
## 252          KC  Carl Cheffers         29      4   7.250000 7.500000
## 253          KC     John Parry         30      4   7.500000 7.500000
## 254          KC  Bill Vinovich         43      5   8.600000 7.500000
## 255          KC Craig Wrolstad         47      5   9.400000 7.500000
## 256          KC Ronald Torbert         45      6   7.500000 7.500000
## 257          KC Gene Steratore         27      3   9.000000 7.500000
## 258          KC   Pete Morelli         47      7   6.714286 7.500000
```

```
## 259          KC     Ed Hochuli          5      1   5.000000 7.500000
## 261          KC   Jerome Boger         53      5  10.600000 7.500000
## 262          KC   Walt Coleman         34      5   6.800000 7.500000
## 263          KC     Brad Allen         31      4   7.750000 7.500000
## 264          KC  Tony Corrente         56      8   7.000000 7.500000
## 265          KC    John Hussey         14      2   7.000000 7.500000
## 266          KC  Terry McAulay         23      3   7.666667 7.500000
## 267          KC Clete Blakeman         18      3   6.000000 7.500000
## 403         OAK Craig Wrolstad         23      4   5.750000 8.825000
## 404         OAK  Terry McAulay         57      4  14.250000 8.825000
## 405         OAK   Jerome Boger         39      4   9.750000 8.825000
## 406         OAK     Brad Allen         31      4   7.750000 8.825000
## 407         OAK Gene Steratore         31      3  10.333333 8.825000
## 408         OAK  Walt Anderson         47      5   9.400000 8.825000
## 409         OAK   Pete Morelli         41      5   8.200000 8.825000
## 410         OAK Ronald Torbert         29      3   9.666667 8.825000
## 412         OAK     Ed Hochuli         50      6   8.333333 8.825000
## 413         OAK  Tony Corrente         48      6   8.000000 8.825000
## 414         OAK    John Hussey         20      3   6.666667 8.825000
## 415         OAK     John Parry         35      4   8.750000 8.825000
## 416         OAK  Bill Vinovich         30      4   7.500000 8.825000
## 417         OAK Clete Blakeman         63      7   9.000000 8.825000
## 418         OAK  Carl Cheffers         39      4   9.750000 8.825000
## 369         NYG Craig Wrolstad         28      6   4.666667 6.600000
## 370         NYG     Brad Allen         29      4   7.250000 6.600000
## 372         NYG   Jerome Boger         34      5   6.800000 6.600000
## 373         NYG  Tony Corrente         33      5   6.600000 6.600000
## 374         NYG     Ed Hochuli         26      4   6.500000 6.600000
## 375         NYG  Carl Cheffers         43      6   7.166667 6.600000
## 376         NYG Ronald Torbert         27      4   6.750000 6.600000
## 377         NYG     John Parry         36      6   6.000000 6.600000
## 378         NYG   Walt Coleman         19      3   6.333333 6.600000
## 379         NYG  Bill Vinovich         39      6   6.500000 6.600000
## 380         NYG   Pete Morelli         25      4   6.250000 6.600000
## 381         NYG  Terry McAulay         27      4   6.750000 6.600000
## 382         NYG  Walt Anderson         24      3   8.000000 6.600000
## 383         NYG Gene Steratore         16      4   4.000000 6.600000
## 384         NYG Clete Blakeman         26      4   6.500000 6.600000
## 385         NYG    John Hussey         34      4   8.500000 6.600000
## 168         DET Craig Wrolstad         16      2   8.000000 7.708333
## 169         DET   Walt Coleman         40      5   8.000000 7.708333
## 170         DET     John Parry         31      4   7.750000 7.708333
## 171         DET  Bill Vinovich         35      6   5.833333 7.708333
## 172         DET     Brad Allen         26      2  13.000000 7.708333
## 173         DET Gene Steratore         32      4   8.000000 7.708333
## 174         DET  Terry McAulay         12      1  12.000000 7.708333
## 175         DET  Walt Anderson         34      4   8.500000 7.708333
## 176         DET   Jerome Boger         46      6   7.666667 7.708333
## 177         DET   Pete Morelli         45      6   7.500000 7.708333
## 179         DET     Ed Hochuli         28      4   7.000000 7.708333
## 180         DET Clete Blakeman         36      6   6.000000 7.708333
## 181         DET  Carl Cheffers         41      6   6.833333 7.708333
## 182         DET  Tony Corrente         43      6   7.166667 7.708333
## 183         DET    John Hussey         36      6   6.000000 7.708333
```

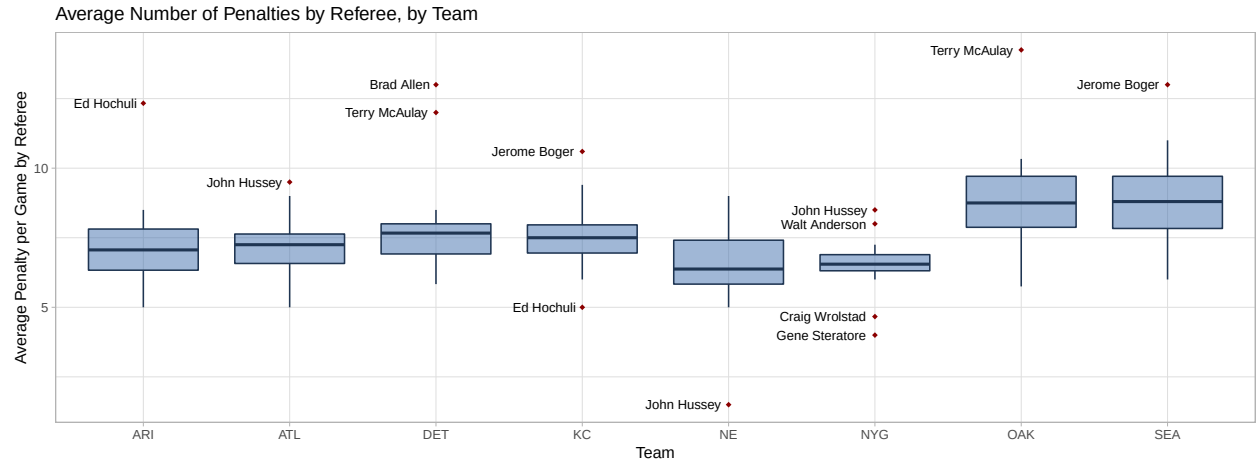
```
## 336          NE Craig Wrolstad        24      4   6.000000 6.500000
## 337          NE   Pete Morelli        57      8   7.125000 6.500000
## 338          NE  Bill Vinovich        25      4   6.250000 6.500000
## 339          NE     Brad Allen        25      5   5.000000 6.500000
## 340          NE     John Parry        26      4   6.500000 6.500000
## 341          NE  Tony Corrente        67      9   7.444444 6.500000
## 342          NE   Walt Coleman        26      5   5.200000 6.500000
## 343          NE Clete Blakeman        16      3   5.333333 6.500000
## 344          NE  Walt Anderson        30      4   7.500000 6.500000
## 345          NE     Ed Hochuli        34      5   6.800000 6.500000
## 346          NE   Jerome Boger        36      4   9.000000 6.500000
## 347          NE Gene Steratore        37      5   7.400000 6.500000
## 349          NE  Terry McAulay        12      2   6.000000 6.500000
## 350          NE  Carl Cheffers        24      4   6.000000 6.500000
## 351          NE Ronald Torbert        26      3   8.666667 6.500000
## 352          NE    John Hussey         3      2   1.500000 6.500000
## 1           ARI  Bill Vinovich        15      3   5.000000 7.000000
## 2           ARI  Walt Anderson        57      8   7.125000 7.000000
## 3           ARI     Brad Allen        31      4   7.750000 7.000000
## 4           ARI     John Parry        27      4   6.750000 7.000000
## 5           ARI Ronald Torbert        28      5   5.600000 7.000000
## 6           ARI   Pete Morelli        24      3   8.000000 7.000000
## 7           ARI  Tony Corrente        23      4   5.750000 7.000000
## 8           ARI Craig Wrolstad        35      6   5.833333 7.000000
## 9           ARI  Terry McAulay        34      4   8.500000 7.000000
## 10          ARI Gene Steratore        35      5   7.000000 7.000000
## 12          ARI    John Hussey        14      2   7.000000 7.000000
## 13          ARI     Ed Hochuli        37      3  12.333333 7.000000
## 14          ARI   Jerome Boger        43      6   7.166667 7.000000
## 15          ARI Clete Blakeman        40      5   8.000000 7.000000
## 16          ARI   Walt Coleman        45      6   7.500000 7.000000
## 17          ARI  Carl Cheffers        26      4   6.500000 7.000000
## 452         SEA   Walt Coleman        36      5   7.200000 8.900000
## 453         SEA Gene Steratore        68      8   8.500000 8.900000
## 454         SEA  Walt Anderson        58      6   9.666667 8.900000
## 455         SEA  Bill Vinovich        30      5   6.000000 8.900000
## 456         SEA  Carl Cheffers        47      6   7.833333 8.900000
## 457         SEA     Brad Allen        44      5   8.800000 8.900000
## 458         SEA Ronald Torbert        39      4   9.750000 8.900000
## 459         SEA    John Hussey        37      4   9.250000 8.900000
## 460         SEA   Jerome Boger        13      1  13.000000 8.900000
## 461         SEA     John Parry        29      4   7.250000 8.900000
## 462         SEA  Terry McAulay        48      5   9.600000 8.900000
## 463         SEA  Tony Corrente        47      6   7.833333 8.900000
## 464         SEA Clete Blakeman        22      2  11.000000 8.900000
## 465         SEA   Pete Morelli        32      4   8.000000 8.900000
## 466         SEA     Ed Hochuli        59      6   9.833333 8.900000
## 18          ATL   Pete Morelli        32      5   6.400000 7.267857
## 19          ATL  Walt Anderson        48      7   6.857143 7.267857
## 20          ATL  Bill Vinovich        51      7   7.285714 7.267857
## 21          ATL Ronald Torbert        36      4   9.000000 7.267857
## 22          ATL  Tony Corrente        38      5   7.600000 7.267857
## 23          ATL  Terry McAulay        30      4   7.500000 7.267857
## 25          ATL Gene Steratore        20      4   5.000000 7.267857
```

```
## 26          ATL    John Hussey         38      4   9.500000 7.267857
## 27          ATL     John Parry         28      5   5.600000 7.267857
## 28          ATL Craig Wrolstad         27      4   6.750000 7.267857
## 29          ATL     Ed Hochuli         23      3   7.666667 7.267857
## 30          ATL     Brad Allen         29      4   7.250000 7.267857
## 31          ATL  Carl Cheffers         27      3   9.000000 7.267857
## 32          ATL Clete Blakeman         32      6   5.333333 7.267857
## 33          ATL   Walt Coleman         27      4   6.750000 7.267857
```

```
ggplot(SEAPenRef,aes(x=PenPerGame))+geom_histogram(binwidth =1)+theme_light()+ylab('Count')+xlab('Penal
```



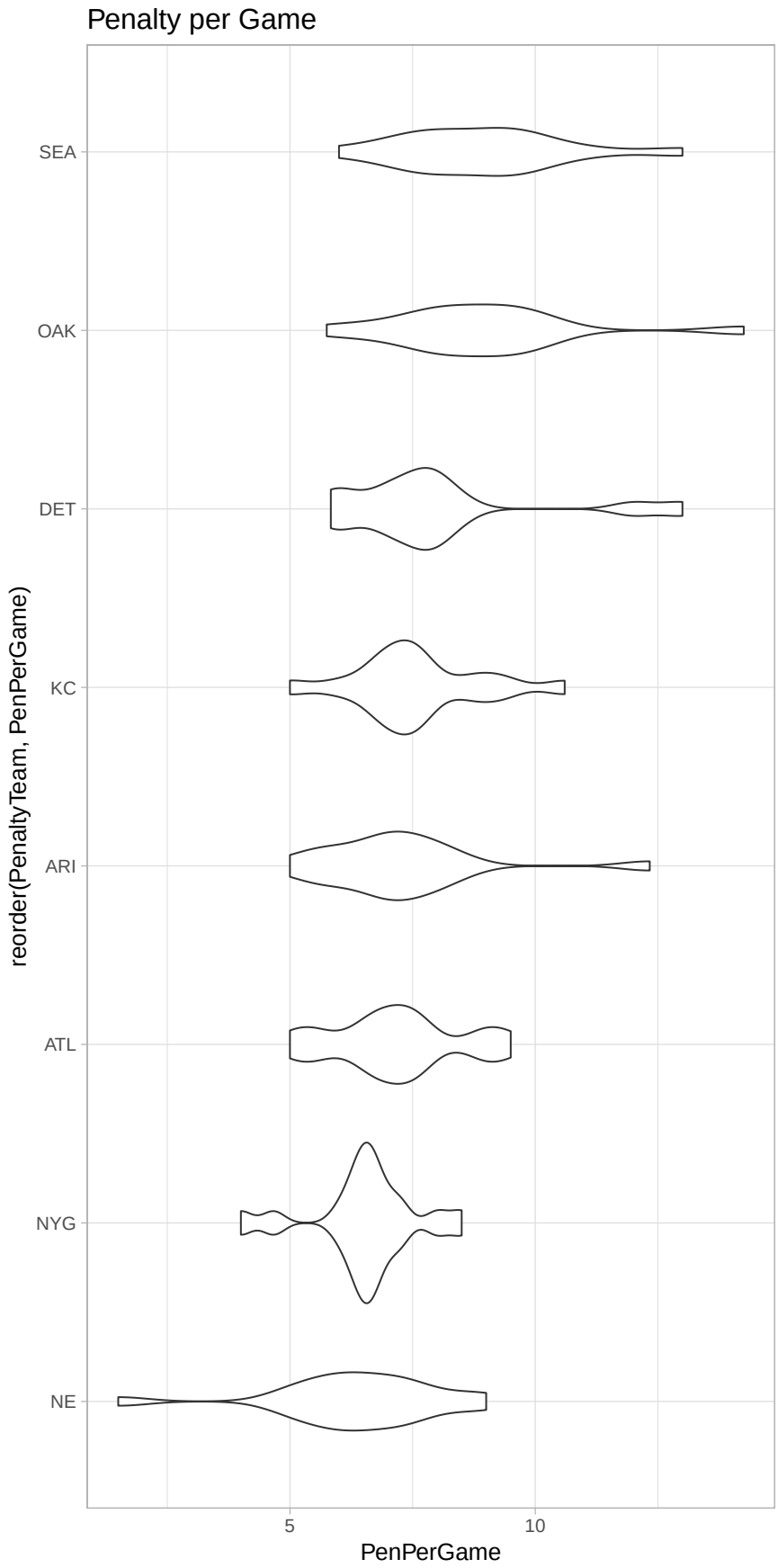

```
fill <- "#4271AE"
line <- "#1F3552"

subsetTeams %>%
  group_by(PenaltyTeam) %>%
  mutate(outlier = ifelse(is_outlier(PenPerGame), Referee,NA)) %>%
  ggplot(., aes(x = factor(PenaltyTeam), y = PenPerGame)) +
    geom_boxplot(fill=fill,colour=line,alpha = 0.5,outlier.colour = "red4", outlier.alpha=1,outlier.shap
    geom_text(aes(label = outlier), na.rm = TRUE,hjust = 1.1,size=3)+
    theme_light()+xlab('Team')+ylab('Average Penalty per Game by Referee')+ggtitle('Average Number of Pe
```



To recap what this chart is showing, we essentially took the previous chart and removed the teams that did not have outliers. In this way, we are focusing on the points that we really care about. Note the recurring Referees. We see Terry McAulay's crew appears to call significantly more penalties against Detroit and Oakland than the other referees. Jerome Boger's crew also calls significantly more calls against Kansas City

and Seattle than other crews. Note the deviation in some of these cases. Perhaps more interestingly, we also see John Hussey's crew calls much fewer penalties against New England, while calling significantly more penalties against the New York Giants.

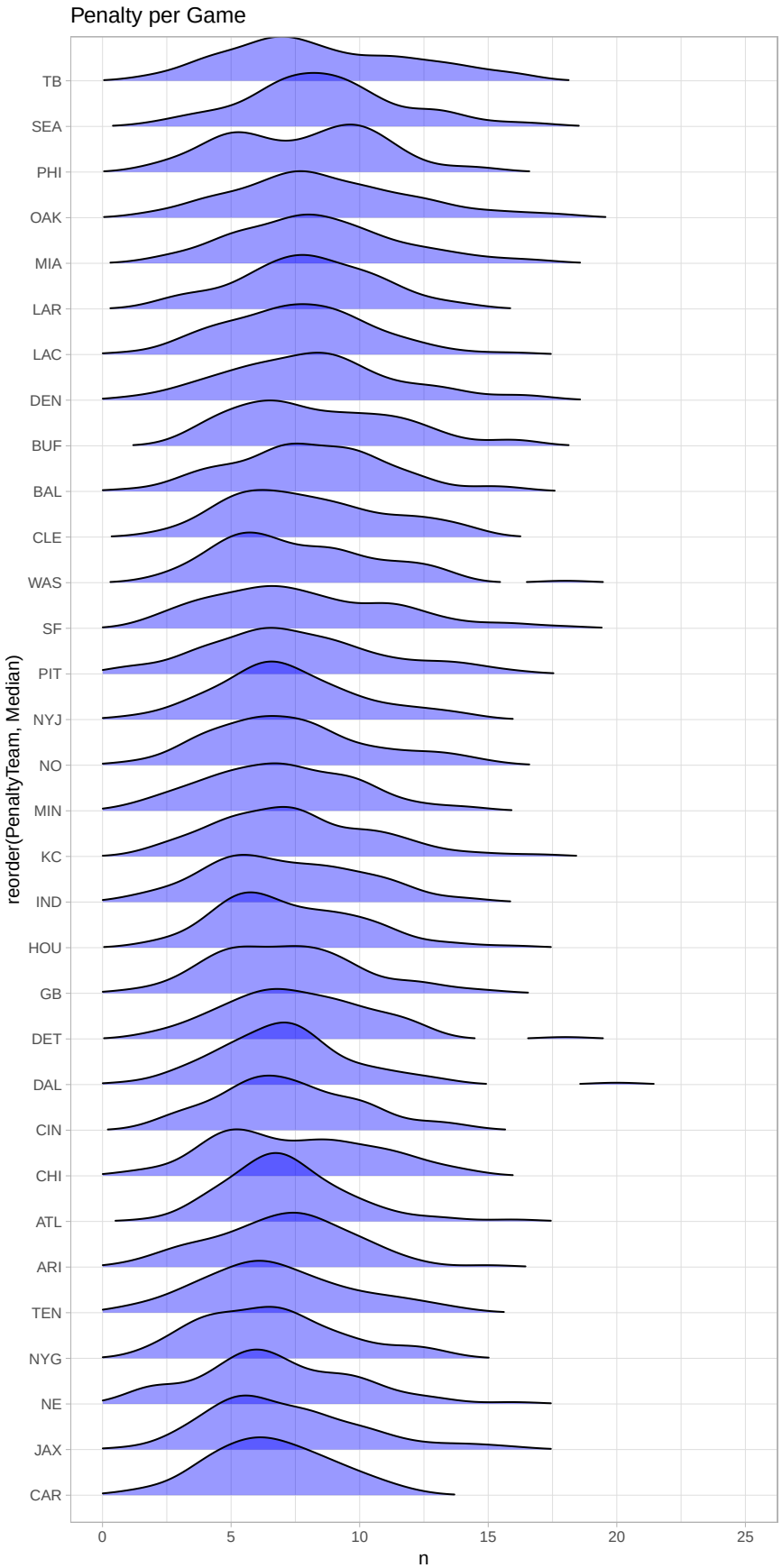
```
#Some Ridgeline Plots and Violin plots in case they are useful...but probably dont want to include all

ggplot(subsetTeams, aes(x = reorder(PenaltyTeam, PenPerGame),y = PenPerGame)) +geom_violin() +
    ggtitle("Penalty per Game") +
    coord_flip() + theme_light(14)
```


Penalty per Game
SEA
OAK
DET
KC
ARI
ATL
NYG
NE
reorder(PenaltyTeam, PenPerGame)
5
10
PenPerGame

```r
#ggplot(subsetTeams, aes(x = PenPerGame,y = reorder(PenaltyTeam, Median))) +geom_density_ridges(fill='b

Temp1<-penaltiesPerGame%>%
  group_by(PenaltyTeam)%>%
  dplyr::summarize(Median = median(n))

penaltiesPerGame<-merge(x=penaltiesPerGame,y=Temp1)
#penaltiesPerGame

ggplot(penaltiesPerGame, aes(x = n,y =reorder(PenaltyTeam,Median))) +geom_density_ridges(fill = 'blue',a
    ggtitle("Penalty per Game") + theme_light()+xlim(0,25)
```

```
## Picking joint bandwidth of 1.04
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density_ridges).
```


Penalty per Game
reorder(PenaltyTeam, Median)
TB
SEA
PHI
OAK
MIA
LAR
LAC
DEN
BUF
BAL
CLE
WAS
SF
PIT
NYJ
NO
MIN
KC
IND
HOU
GB
DET
DAL
CIN
CHI
ATL
ARI
TEN
NYG
NE
JAX
CAR
0
5
10
15
20
25
n

```
#ggplot(penRef, aes(x = PenPerGame,y = reorder(Referee, Median))) +geom_density_ridges(fill = 'blue',al
#    ggtitle("Distribution of Average # Of Penalties by Team") + theme_light()+ylab('Referee')+xlab('Di
```