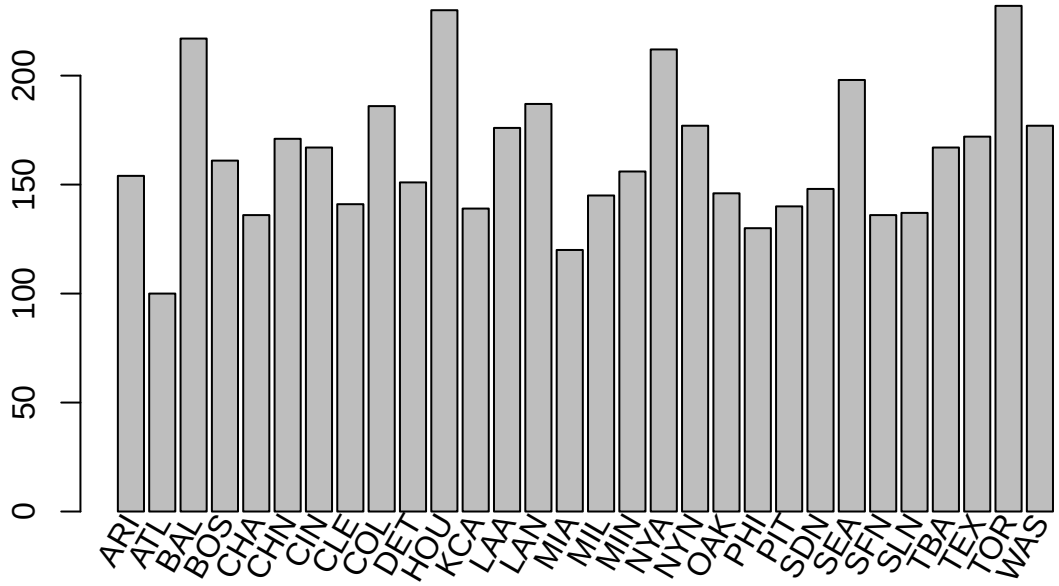
chap03. Histogram

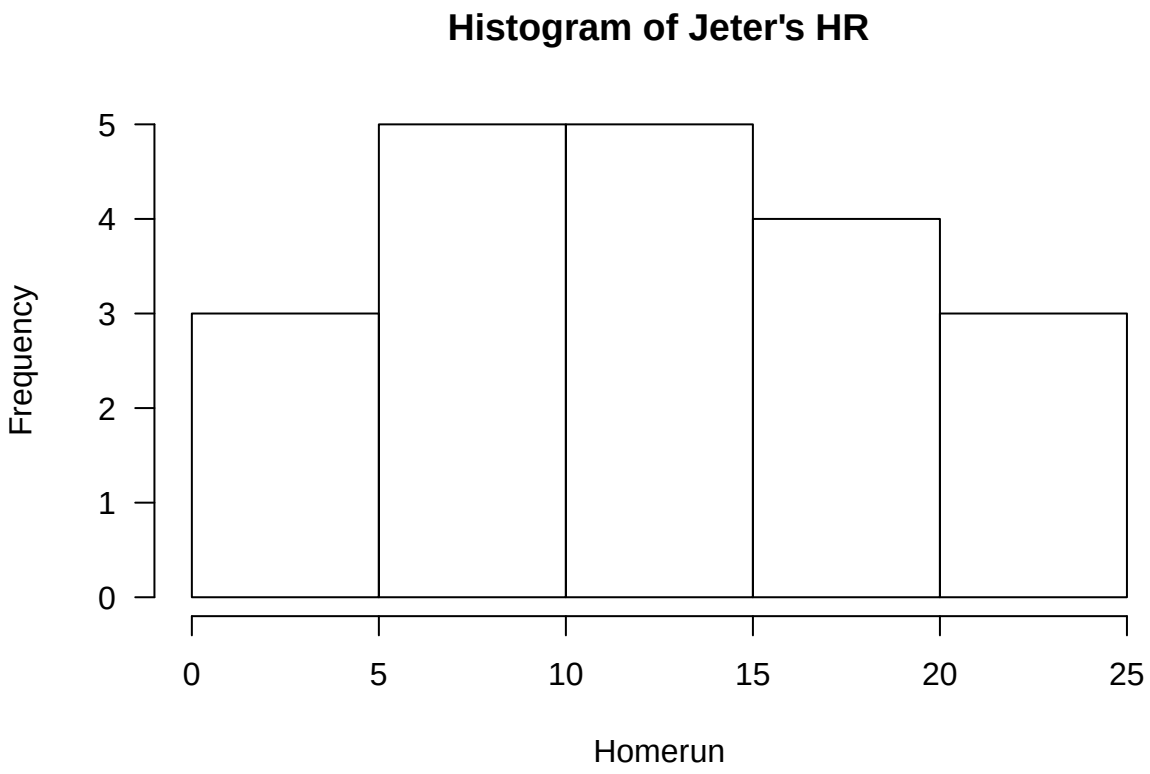
Major League 30 Teams HR bar graph.

```
library(Lahman)
a = subset(Teams, yearID==2015)
b = barplot(a$HR)
text(b, par('usr')[3], labels=a$teamID, srt=60, adj=c(1, 0.5), xpd=TRUE)
```

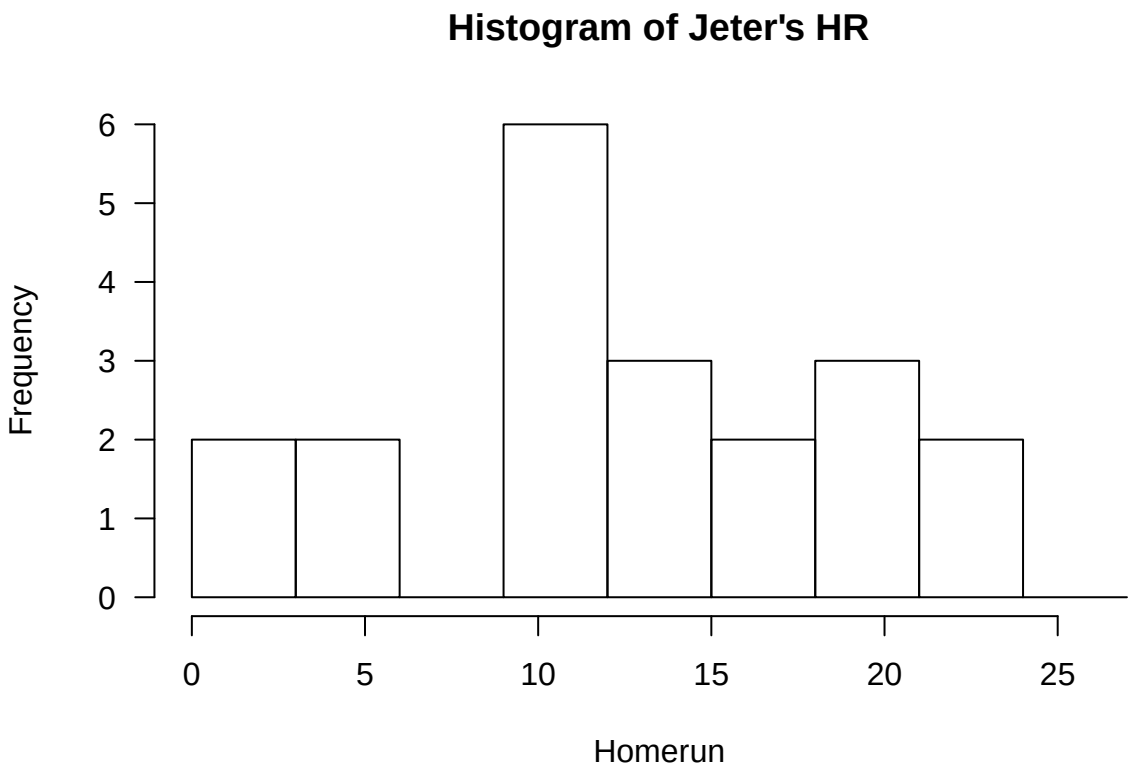


Historam can not only be a descriptive statistics to show probability of event accordances, but also be a basis for inferential statistics to infer the characteristics of major league Homerun population which is a set of events (Homerun).

```
a = subset(Batting, playerID=='jeterde01')
hist(a$HR, xlab='Homerun', main="Histogram of Jeter's HR", las = 1)
```



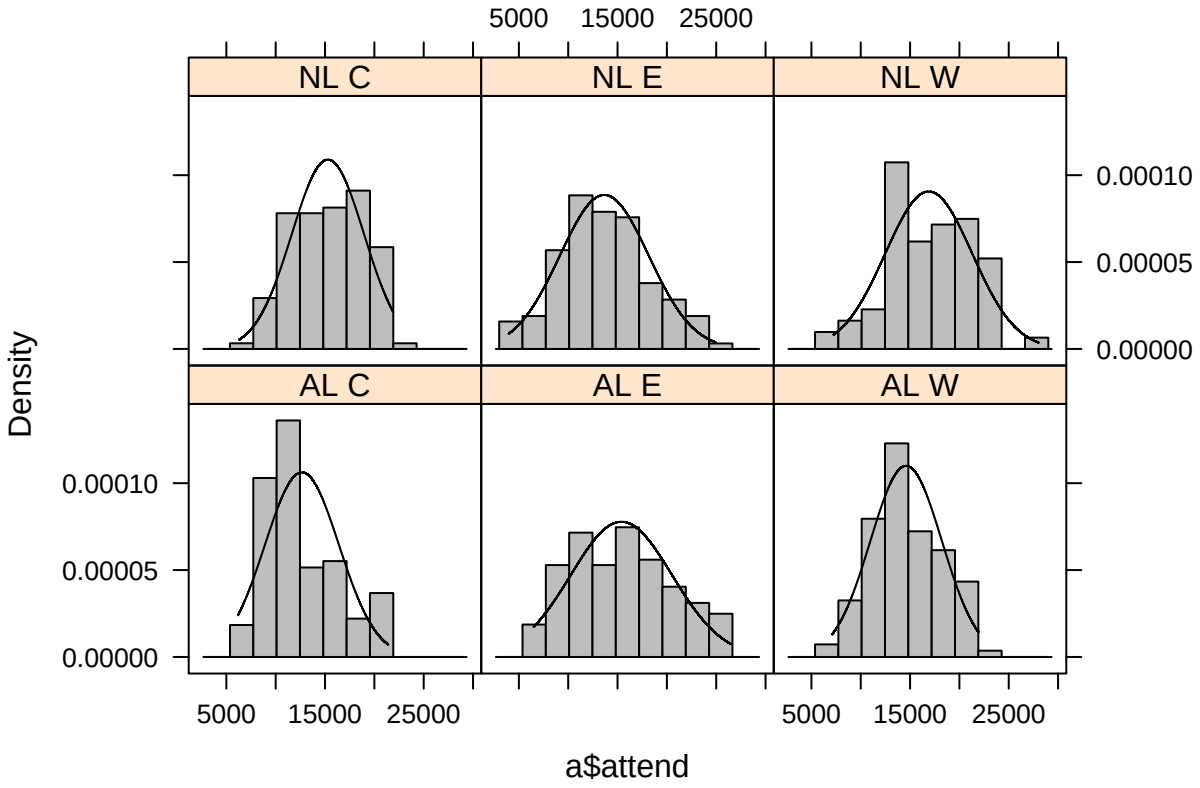

```
hist(a$HR, xlab='Homerun', main="Histogram of Jeter's HR", las = 1,
     breaks=seq(from=0, to=27, by=3))
```



The divisions belonged a team may affect the attendance counts. We can plot lattice panel graphs for each divisions.

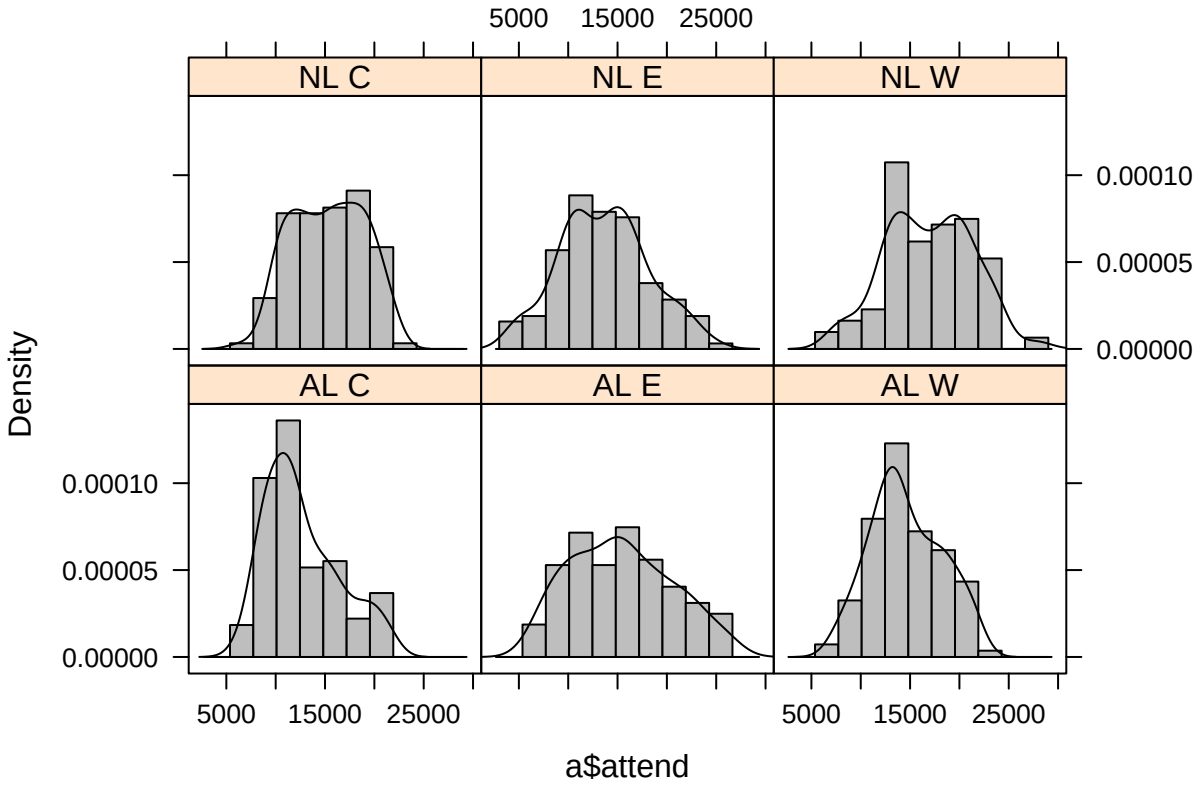
```
library(Lahman)
a = subset(Teams, yearID>1990)
a$attend = a$attendance / a$G
a$affiliation = paste(a$lgID, a$divID)
```

```
library(lattice)
histogram(~a$attend | a$affiliation, type='density', panel = function(x, ...){
  panel.histogram(x, ..., col='gray')
  apg = seq(min(x), max(x))
  density = dnorm(apg, mean(x), sd(x))
  panel.lines(apg, density, col='black')})
```



Without assuming populations' parameters, we only add nonparametric p.d.f. with the histogram.

```
histogram(
  ~a$attend | a$affiliation, type='density', panel = function(x, ...){
  panel.histogram(x, ..., col='gray')
  apg = seq(min(x), max(x))
  density = dnorm(apg, mean(x), sd(x))
  panel.densityplot(x, col='black', plot.points=FALSE)
  })
```


NL C
NL E
NL W
AL C
AL E
AL W
Density
0.00000
0.00005
0.00010
5000
15000
25000
a$attend