STAT 428 Final

Ziqin Xiong, zxiong8

21 April, 2019

1. Data Generation

```
setwd('~/courserel/STAT 428/final')
gl18 <- read.csv("GL2018.TXT",header = F)
#head(gl18)
unique(gl18$V4)
```

```
##  [1] COL PHI SFN CHN SLN MIL MIN CHA ANA CLE BOS HOU NYA WAS PIT LAN KCA BAL TEX
## [20] SEA TBA MIA CIN ARI NYN DET ATL OAK SDN TOR
## 30 Levels: ANA ARI ATL BAL BOS CHA CHN CIN CLE COL DET HOU KCA LAN MIA ... WAS
```

```
unique(gl18[,c(5,8)])
```

```
##    V5 V8
## 1  NL NL
## 7  AL AL
## 20 NL AL
## 54 AL NL
```

```
nrow(subset(gl18,V5=='AL'&V8=='NL'))
```

```
## [1] 150
```

```
nrow(subset(gl18,V161!='Y'))
```

```
## [1] 0
```

```
# data files were downloaded from
# https://www.retrosheet.org/gamelogs/
# The information used here was obtained free of
# charge from and is copyrighted by Retrosheet.  Interested
# parties may contact Retrosheet at "www.retrosheet.org".
gl16 <- read.csv("GL2016.TXT",header = F)[,c(4,5,10,7,8,11)]
gl17 <- read.csv("GL2017.TXT",header = F)[,c(4,5,10,7,8,11)]
gl18 <- read.csv("GL2018.TXT",header = F)[,c(4,5,10,7,8,11)]
gl <- rbind(gl16,gl17,gl18)
colnames(gl) <- c('vteam','vleague','vscore','hteam','hleague','hscore')

win <- lose <- draw <- matrix(0,30,30,dimnames =
         list(paste0(levels(gl$vteam),'v'),paste0(levels(gl$vteam),'h')))
for(i in 1:nrow(gl)){
  if(gl[i,3]>gl[i,6])
    win[gl$vteam[i],gl$hteam[i]] = win[gl$vteam[i],gl$hteam[i]]+1
  else if(gl[i,3]<gl[i,6])
    lose[gl$vteam[i],gl$hteam[i]] = lose[gl$vteam[i],gl$hteam[i]]+1
  else
    draw[gl$vteam[i],gl$hteam[i]] = draw[gl$vteam[i],gl$hteam[i]]+1
}
total <- win+lose+draw
```

```r
#winning probability of visiting teams, where rows are visiting teams and
#cols are home teams.
#For example (ANAv,CHAh) means the avg probability of ANA winning CHA as a
#visiting team. This also means the avg probability of CHA losing or drawing
#ANA as a home team (P(lose|draw) = 1-P(win))
#NaN means there's no game records for 2 teams
winprob.v <- win/total
winprob.v[1:5,1:5]
```

```
##         ANAh    ARIh    ATLh    BALh    BOSh
## ANAv     NaN 0.00000     NaN 0.55556 0.33333
## ARIv 0.50000     NaN 0.55556 0.00000 0.00000
## ATLv 0.33333 0.54545     NaN     NaN 0.40000
## BALv 0.33333     NaN 0.66667     NaN 0.50000
## BOSv 0.60000     NaN 1.00000 0.75862     NaN
```