Chapter 2

Nick Lauerman

Contents

Libraries used

```
library(Lahman)
```

Data Sets

Vectors

Career of Wareen Spshn

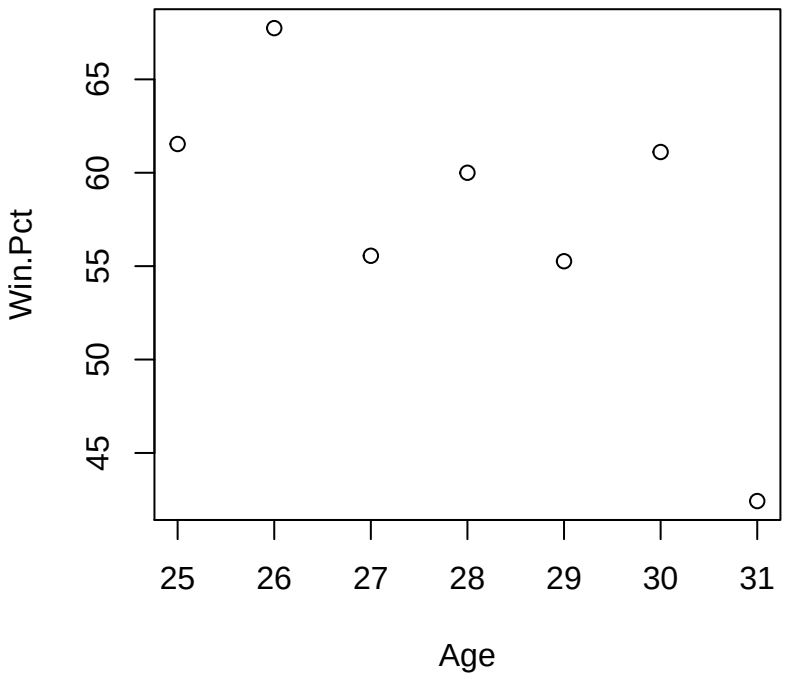
```
W <- c(8, 21, 15, 21, 21, 22, 14)
L <- c(5, 10, 12, 14, 17, 14, 19)
Win.Pct <- 100* W / (W +L)
Win.Pct
```

```
## [1] 61.53846 67.74194 55.55556 60.00000 55.26316 61.11111 42.42424
```

```
Year <- seq(1946, 1952)
Year
```

```
## [1] 1946 1947 1948 1949 1950 1951 1952
```

```
Year <- 1946:1952
Age <- Year - 1921
plot(Age, Win.Pct)
```



Vector Functions

```
mean(Win.Pct)
```

```
## [1] 57.66207
```

```
100 * sum(W) / (sum(W) + sum(L))
```

```
## [1] 57.277
```

```
sort(W)
```

```
## [1]  8 14 15 21 21 21 22
```

```
cumsum(W)
```

```
## [1]   8  29  44  65  86 108 122
```

```
summary(Win.Pct)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##   42.42   55.41   60.00   57.66   61.32   67.74
```

Vector index and logical variables

```
W[c(1,2,5)]
```

```
## [1]  8 21 21
```

```
W[1:4]
```

```
## [1]  8 21 15 21
```

```
W[-c(1, 6)]
```

```
## [1] 21 15 21 21 14
```

```
Win.Pct > 60
```

```
## [1]  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE
```

```
(W > 20) & (Win.Pct > 60)
```

```
## [1] FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE
```

```
Win.Pct == max(Win.Pct)
```

```
## [1] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
```

```
Year[Win.Pct == max(Win.Pct)]
```

```
## [1] 1947
```

```
Year[W + L > 30]
```

```
## [1] 1947 1949 1950 1951 1952
```

Object and Container in R

Character data nd matrices

```
NL <- c("FLA",
        "STL",
        "HOU",
        "STL",
        "COL",
        "PHI",
        "PHI",
        "SFG",
        "STL",
        "SFG")
AL <- c("NYY",
        "BOS",
        "CHW",
        "DET",
        "BOS",
        "TBR",
        "NYY",
        "TEX",
        "TEX",
```

```
        "DET")
Winner <- c("NL",
            "AL",
            "AL",
            "NL",
            "NL",
            "NL",
            "AL",
            "NL",
            "NL",
            "NL")
N.Games <- c(6,
             4,
             4,
             5,
             4,
             5,
             6,
             5,
             7,
             4)
Year <- 2003:2012
results <- matrix(c(NL,AL), 10, 2)
results
```

```
##       [,1]  [,2]
##  [1,] "FLA" "NYY"
##  [2,] "STL" "BOS"
##  [3,] "HOU" "CHW"
##  [4,] "STL" "DET"
##  [5,] "COL" "BOS"
##  [6,] "PHI" "TBR"
##  [7,] "PHI" "NYY"
##  [8,] "SFG" "TEX"
##  [9,] "STL" "TEX"
## [10,] "SFG" "DET"
```

```
dimnames(results)[[1]] <- Year
dimnames(results)[[2]] <- c("NL Teams", "AL Teams")
results
```

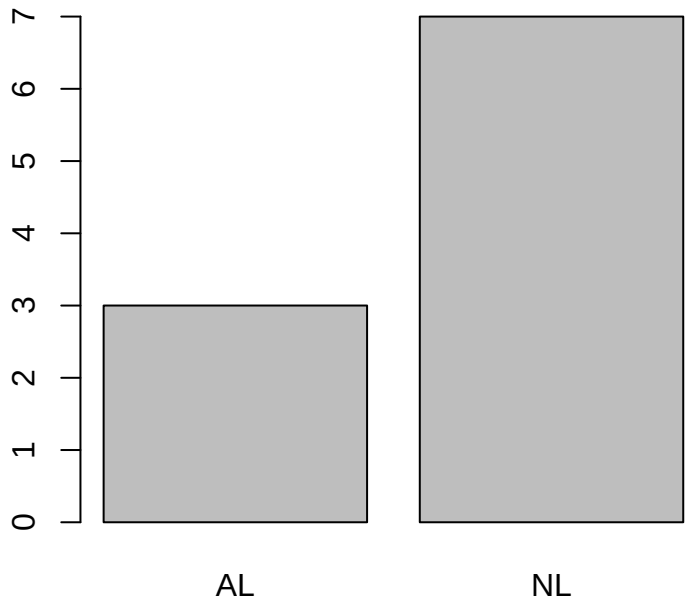
```
##      NL Teams AL Teams
## 2003 "FLA"    "NYY"
## 2004 "STL"    "BOS"
## 2005 "HOU"    "CHW"
## 2006 "STL"    "DET"
## 2007 "COL"    "BOS"
## 2008 "PHI"    "TBR"
## 2009 "PHI"    "NYY"
## 2010 "SFG"    "TEX"
## 2011 "STL"    "TEX"
## 2012 "SFG"    "DET"
```

```
table(Winner)
```

```
## Winner
```

```
## AL NL
##  3  7
```

```
barplot(table(Winner))
```



Factors

```
table(NL)
```

```
## NL
## COL FLA HOU PHI SFG STL
##   1   1   1   2   2   3
```

```
NL2 <- factor(NL, levels = c("FLA",
                             "PHI",
                             "HOU",
                             "STL",
                             "COL",
                             "SFG"))
str(NL2)
```

```
##  Factor w/ 6 levels "FLA","PHI","HOU",..: 1 4 3 4 5 2 2 6 4 6
```

```
table(NL2)
```

```
## NL2
## FLA PHI HOU STL COL SFG
##   1   2   1   3   1   2
```

Lists

```
World.Series <- list(Winner = Winner,
                     Number.Games = N.Games,
                     Seasons = "2003 to 2012")
World.Series$Number.Games
```

```
##  [1] 6 4 4 5 4 5 6 5 7 4
```

```
World.Series[[2]]
```

```
##  [1] 6 4 4 5 4 5 6 5 7 4
```

```
World.Series["Number.Games"]
```

```
## $Number.Games
##  [1] 6 4 4 5 4 5 6 5 7 4
```

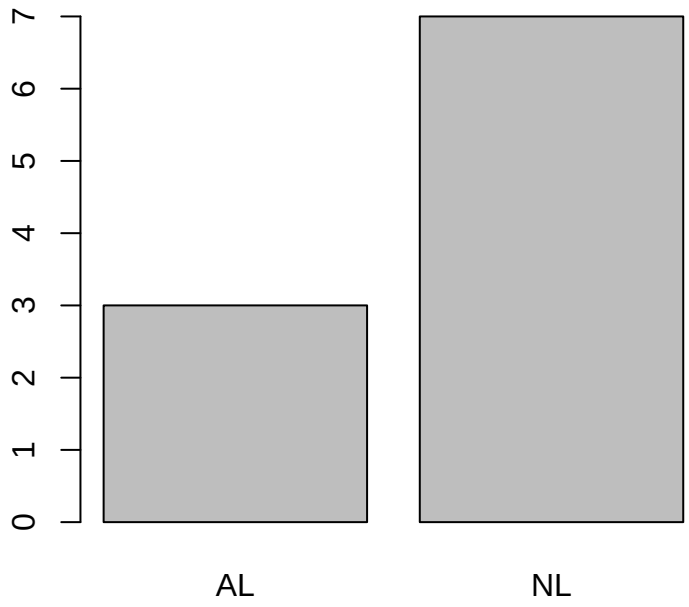
Collection of R Commands

r scripts

```
table(Winner)
```

```
## Winner
## AL NL
##  3  7
```

```
barplot(table(Winner))
```




```
by(N.Games, Winner, summary)
```

```
## Winner: AL
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.000   4.000   4.000   4.667   5.000   6.000
## ------------------------------------------------------------
## Winner: NL
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.000   4.500   5.000   5.143   5.500   7.000
```

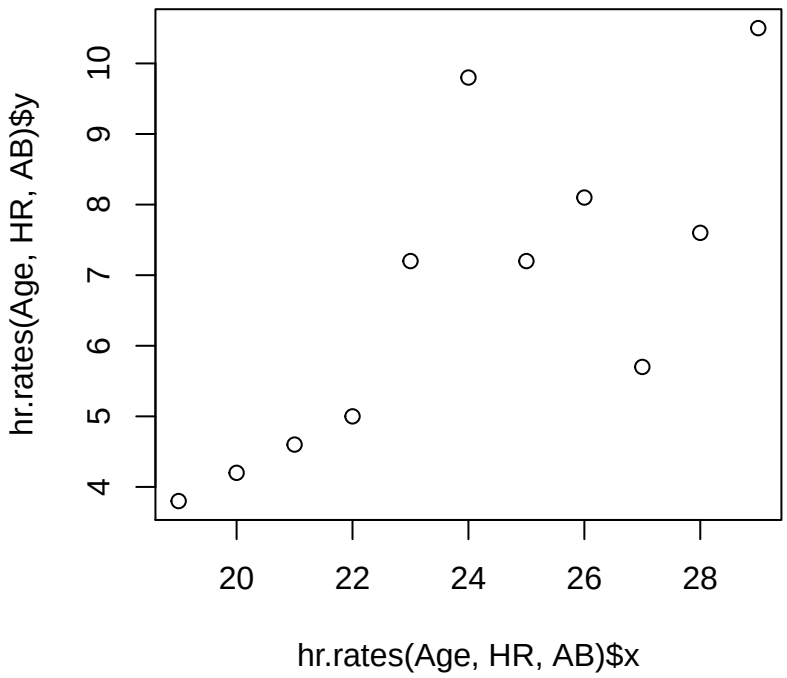
R Functions

```
# also in scripts/Chapter2.R
hr.rates <- function(age, hr, ab){
   rates <- round(100 * hr / ab, 1)
   list(x=age, y=rates)
}
source(file = "scripts/Chapter2.R")
HR <- c(13,
        23,
        21,
        27,
        37,
        52,
        34,
        42,
```

```
        31,
        40,
        54)
AB <- c(341,
        549,
        461,
        543,
        517,
        533,
        474,
        519,
        541,
        527,
        514)
Age <- 19:29
hr.rates(Age, HR, AB)
```

```
## $x
##  [1] 19 20 21 22 23 24 25 26 27 28 29
##
## $y
##  [1]  3.8  4.2  4.6  5.0  7.2  9.8  7.2  8.1  5.7  7.6 10.5
```

```
plot(hr.rates(Age, HR, AB))
```



Reading and Writing Data in R

Importing data from a file

```
spahn <- read.csv("./BookData/spahn.csv")
```

Saving datasets

```
HR
```

```
##  [1] 13 23 21 27 37 52 34 42 31 40 54
```

```
HR.Rates <- hr.rates(Age, HR, AB)
Mantle <- cbind(Age, HR, AB, Rates=HR.Rates$y)
write.csv(Mantle, "Mantle.csv", row.names = FALSE)
```

Data Frames

Introduction

```
spahn[1:3, 1:10]
```

```
##   Year Age  Tm Lg  W  L   W.L  ERA  G GS
## 1 1942  21 BSN NL  0  0    NA 5.74  4  2
## 2 1946  25 BSN NL  8  5 0.615 2.94 24 16
## 3 1947  26 BSN NL 21 10 0.677 2.33 40 35
```

```
spahn[1,]
```

```
##   Year Age  Tm Lg W L W.L  ERA G GS GF CG SHO SV   IP  H  R ER HR BB IBB
## 1 1942  21 BSN NL 0 0  NA 5.74 4  2  0  1   0  0 15.2 25 15 10  0 11  NA
##   SO HBP BK WP BF ERA.  WHIP  H.9 HR.9 BB.9 SO.9 SO.BB Awards
## 1  7   0  0  0 79   59 2.298 14.4    0  6.3    4  0.64
```

```
spahn[1 : 10, c("Age",
                "W",
                "L",
                "ERA")]
```

```
##    Age  W  L  ERA
## 1   21  0  0 5.74
## 2   25  8  5 2.94
## 3   26 21 10 2.33
## 4   27 15 12 3.71
## 5   28 21 14 3.07
## 6   29 21 17 3.16
## 7   30 22 14 2.98
## 8   31 14 19 2.98
## 9   32 23  7 2.10
## 10  33 21 12 3.14
```

```
summary(spahn$ERA)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.100   2.940   3.040   3.256   3.260   5.740
```

```
spahn$Age[spahn$ERA == min(spahn$ERA)]
```

```
## [1] 32
```

Manipulation with data frames

```
spahn$FIP <- with(spahn, (13 * HR + 3 * BB - 2 * SO) / IP)
pos <- order(spahn$FIP)
head(spahn[pos, c("Year",
                  "Age",
                  "W",
                  "L",
                  "ERA",
                  "FIP")])
```

```
##    Year Age  W  L  ERA       FIP
## 8  1952  31 14 19 2.98 0.3448276
## 9  1953  32 23  7 2.10 0.3619910
## 2  1946  25  8  5 2.94 0.4153355
## 15 1959  38 21 15 2.96 0.6746575
## 3  1947  26 21 10 2.33 0.6950207
## 12 1956  35 20 11 2.78 0.8004269
```

```
spahn1 <- subset(spahn,
                 Tm == "BSN" | Tm == "MLN")
by(spahn1[, c("W.L",
              "ERA",
              "WHIP",
              "FIP")],
   spahn1$Tm,
   summary)
```

```
## spahn1$Tm: BSN
##       W.L              ERA             WHIP            FIP
##  Min.   :0.4240   Min.   :2.330   Min.   :1.136   Min.   :0.3448
##  1st Qu.:0.5545   1st Qu.:2.970   1st Qu.:1.154   1st Qu.:0.6251
##  Median :0.6000   Median :3.025   Median :1.222   Median :0.8219
##  Mean   :0.5766   Mean   :3.364   Mean   :1.331   Mean   :0.7922
##  3rd Qu.:0.6130   3rd Qu.:3.297   3rd Qu.:1.230   3rd Qu.:0.9836
##  Max.   :0.6770   Max.   :5.740   Max.   :2.298   Max.   :1.2500
##  NA's   :1
## ------------------------------------------------------------
## spahn1$Tm: MLN
##       W.L              ERA             WHIP            FIP
##  Min.   :0.3160   Min.   :2.100   Min.   :1.058   Min.   :0.3620
##  1st Qu.:0.5780   1st Qu.:2.757   1st Qu.:1.123   1st Qu.:0.8345
##  Median :0.6405   Median :3.030   Median :1.163   Median :0.9944
##  Mean   :0.6202   Mean   :3.121   Mean   :1.187   Mean   :0.9839
##  3rd Qu.:0.6695   3rd Qu.:3.170   3rd Qu.:1.226   3rd Qu.:1.0764
##  Max.   :0.7670   Max.   :5.290   Max.   :1.474   Max.   :1.7263
## ------------------------------------------------------------
## spahn1$Tm: TOT
## NULL
```

Merging and selectin from data frames

```
NLBatting <- read.csv("BookData/NLbatting.csv")
ALBatting <- read.csv("BookData/ALbatting.csv")
batting <- rbind(NLBatting, ALBatting)
NLpitching <- read.csv("BookData/NLpitching.csv")
NL <- merge(NLBatting, NLpitching,
            by = "Tm")
NL.150 <- subset(NLBatting, HR > 150)
```

Packages

```
library(Lahman)
#?Batting
```