개인과제1

최재혁

2020 5 24

### **주제 : 이글스 군단의 가성비 갑은 누구?**

## data 읽기

setwd("C:\\Workspace\\WorkspaceR")  
baseball <- read.csv( 'C:\\Workspace\\WorkspaceR\\batter\_stats\_2017.csv', header = T )

## data 구조 확인

class(baseball) # data.frame

## [1] "data.frame"

str(baseball) # 191 obs, 21 variables

## 'data.frame': 191 obs. of 21 variables:  
## $ 선수명 : chr "최정" "최형우" "박건우" "나성범" ...  
## $ 팀명 : chr "SK" "KIA" "두산" "NC" ...  
## $ 경기 : int 130 142 131 125 144 137 139 106 141 115 ...  
## $ 타석 : int 527 629 543 561 667 529 621 452 601 518 ...  
## $ 타수 : int 430 514 483 498 576 476 557 388 526 437 ...  
## $ 안타 : int 136 176 177 173 193 176 178 141 159 131 ...  
## $ 홈런 : int 46 26 20 24 20 5 27 3 23 35 ...  
## $ 득점 : int 89 98 91 103 113 84 118 84 90 91 ...  
## $ 타점 : int 113 120 78 99 80 64 111 47 114 111 ...  
## $ 볼넷 : int 70 96 41 48 83 39 41 46 58 65 ...  
## $ 삼진 : int 107 82 64 116 96 40 112 51 65 134 ...  
## $ 도루 : int 1 0 20 17 25 4 32 11 16 4 ...  
## $ BABIP : chr "0.316" "0.362" "0.39" "0.413" ...  
## $ 타율 : num 0.316 0.342 0.366 0.347 0.335 0.37 0.32 0.363 0.302 0.3 ...  
## $ 출루율 : num 0.427 0.45 0.424 0.415 0.42 0.42 0.373 0.441 0.376 0.402 ...  
## $ 장타율 : num 0.684 0.576 0.582 0.584 0.514 0.477 0.54 0.472 0.513 0.595 ...  
## $ OPS : num 1.111 1.026 1.006 0.999 0.934 ...  
## $ wOBA : num 0.442 0.43 0.424 0.416 0.398 0.391 0.38 0.404 0.375 0.411 ...  
## $ WAR : num 7.3 7.2 7.04 5.64 5.6 5.19 5.01 4.92 4.76 4.7 ...  
## $ 연봉.2018.: int 120000 150000 37000 43000 150000 28000 75000 32000 32000 70000 ...  
## $ 연봉.2017.: int 120000 150000 19500 35000 65000 8000 75000 25000 22000 80000 ...

dim(baseball)

## [1] 191 21

baseball$선수명 <- as.character( baseball$선수명 )  
baseball$팀명 <- as.character( baseball$팀명 )

## 팀명 나누기

team <- c()  
for ( i in 1:nrow( baseball ) ) {  
 if ( baseball[ i, "팀명" ] == "SK" ) {  
 team[ i ] <- 1  
 } else if ( baseball[ i, "팀명" ] == "KIA" ) {  
 team[ i ] <- 2  
 } else if ( baseball[ i, "팀명" ] == "두산" ) {  
 team[ i ] <- 3  
 } else if ( baseball[ i, "팀명" ] == "NC" ) {  
 team[ i ] <- 4  
 } else if ( baseball[ i, "팀명" ] == "롯데" ) {  
 team[ i ] <- 5  
 } else if ( baseball[ i, "팀명" ] == "넥센" ) {  
 team[ i ] <- 6  
 } else if ( baseball[ i, "팀명" ] == "삼성" ) {  
 team[ i ] <- 7  
 } else if ( baseball[ i, "팀명" ] == "한화" ) {  
 team[ i ] <- 8  
 } else if ( baseball[ i, "팀명" ] == "LG" ) {  
 team[ i ] <- 9  
 } else if ( baseball[ i, "팀명" ] == "KT" ) {  
 team[ i ] <- 10  
 }   
}  
team

## [1] 1 2 3 4 5 2 2 4 6 4 7 7 2 1 6 2 5 9 6 5 3 8 8 4 5  
## [26] 2 3 8 9 8 4 10 3 2 2 1 3 1 6 10 4 9 4 10 1 4 7 8 6 6  
## [51] 7 10 3 6 10 7 7 1 9 4 8 2 2 7 8 9 6 6 5 5 5 9 3 9 1  
## [76] 1 1 1 7 4 8 10 1 5 1 9 4 7 9 6 7 2 1 7 8 9 10 8 1 2  
## [101] 6 10 9 10 3 7 6 2 9 4 5 8 3 10 6 2 7 2 6 4 3 10 1 6 6  
## [126] 2 3 2 9 3 5 5 7 1 10 8 9 4 9 8 2 10 2 10 7 10 4 4 9 4  
## [151] 8 3 7 1 3 5 1 4 9 7 4 8 6 9 2 3 3 8 5 6 8 4 6 5 1  
## [176] 10 3 10 6 10 1 5 7 9 7 10 2 8 9 2 8

## 선수, 팀명 이름 분리

name <- unique( baseball[ , 1] )  
name

## [1] "최정" "최형우" "박건우" "나성범" "손아섭" "김선빈"   
## [7] "버나디나" "박민우" "김하성" "스크럭스" "러프" "구자욱"   
## [13] "안치홍" "한동민" "서건창" "나지완" "이대호" "박용택"   
## [19] "이정후" "전준우" "오재일" "정근우" "김태균" "손시헌"   
## [25] "번즈" "이범호" "양의지" "이성열" "유강남" "송광민"   
## [31] "권희동" "정현" "김재호" "김주찬" "이명기" "로맥"   
## [37] "최주환" "나주환" "김민성" "박경수" "박석민" "오지환"   
## [43] "모창민" "유한준" "정의윤" "이종욱" "조동찬" "하주석"   
## [49] "고종욱" "장영석" "이원석" "오정복" "박세혁" "박동원"   
## [55] "이해창" "박해민" "배영섭" "김동엽" "안익훈" "이상호"   
## [61] "오선진" "최원준" "서동욱" "김성훈" "양성우" "이형종"   
## [67] "김태완" "황진수" "정훈" "김문호" "김재율" "정진호"   
## [73] "이천웅" "이대수" "노수광" "최항" "조용호" "강한울"   
## [79] "강진성" "최재훈" "심우준" "이성우" "박헌도" "박정권"   
## [85] "양석환" "지석훈" "권정웅" "정상호" "이택근" "박한이"   
## [91] "노관현" "김성현" "김헌곤" "임익준" "김주성" "장성우"   
## [97] "김회성" "이재원" "최정민" "허정협" "김진곤" "윤진호"   
## [103] "이진영" "박유연" "최영진" "송성문" "김지성" "김용의"   
## [109] "강구성" "나경민" "이창열" "허경민" "정주후" "임병욱"   
## [115] "최병연" "안주형" "오준혁" "김규민" "조평호" "이우성"   
## [121] "김만수" "최승준" "김혜성" "홍성갑" "이인행" "류지혁"   
## [127] "백용환" "서상우" "조수행" "나종덕" "문규현" "김성윤"   
## [133] "박성한" "전민수" "이용규" "문선재" "김성욱" "장준원"   
## [139] "강상원" "이호신" "오태곤" "한승택" "홍현빈" "이현동"   
## [145] "남태혁" "박광열" "신진호" "조윤준" "이재율" "박준혁"   
## [151] "김민혁" "김민수" "김재현" "오재원" "김동한" "정진기"   
## [157] "김종민" "정주현" "정병곤" "도태훈" "최윤석" "박정음"   
## [163] "임훈" "신종길" "국해성" "김인태" "김태연" "신본기"   
## [169] "김지수" "강경학" "윤병호" "주효상" "김사훈" "김강민"   
## [175] "김동욱" "신성현" "박기혁" "하준호" "박승욱" "김상호"   
## [181] "백상원" "강승호" "이지영" "이대형" "김민식" "장민석"   
## [187] "채은성" "김주형" "이동훈"

team2 <- unique( baseball[ , 2 ] )  
team2

## [1] "SK" "KIA" "두산" "NC" "롯데" "넥센" "삼성" "LG" "한화" "KT"

## 한화 데이터 분리

hanwha <- baseball[ baseball$팀명 == team2[ 9 ],] # 숫자 바꾸면 해당 팀의 데이터 출력, 한화 = 9  
hanwha

## 선수명 팀명 경기 타석 타수 안타 홈런 득점 타점 볼넷 삼진 도루 BABIP 타율  
## 22 정근우 한화 105 442 391 129 11 73 46 32 48 6 0.352 0.330  
## 23 김태균 한화 94 407 356 121 17 51 76 43 56 0 0.362 0.340  
## 28 이성열 한화 81 318 280 86 21 41 65 19 80 3 0.353 0.307  
## 30 송광민 한화 117 473 437 143 13 71 75 21 88 2 0.378 0.327  
## 48 하주석 한화 111 470 432 123 11 69 52 23 83 7 0.33 0.285  
## 61 오선진 한화 65 209 184 57 2 23 21 15 22 3 0.34 0.310  
## 65 양성우 한화 118 417 358 99 2 42 40 38 51 4 0.316 0.277  
## 81 최재훈 한화 98 301 262 66 1 21 15 25 36 0 0.289 0.252  
## 95 임익준 한화 30 47 36 9 0 7 2 7 8 1 0.321 0.250  
## 98 김회성 한화 54 103 85 20 3 10 13 13 24 0 0.283 0.235  
## 112 이창열 한화 3 1 1 0 0 1 0 0 0 0 0 0.000  
## 136 이용규 한화 57 200 179 47 0 31 12 17 20 10 0.294 0.263  
## 140 강상원 한화 33 14 14 3 0 8 1 0 5 0 0.333 0.214  
## 151 박준혁 한화 8 10 10 1 0 1 0 0 4 0 0.167 0.100  
## 162 최윤석 한화 21 33 32 7 0 3 2 0 6 0 0.269 0.219  
## 168 김태연 한화 12 22 21 1 1 1 2 1 11 0 0 0.048  
## 171 강경학 한화 59 93 84 18 0 17 4 8 19 1 0.277 0.214  
## 188 장민석 한화 88 327 299 83 1 38 21 15 58 8 0.34 0.278  
## 191 이동훈 한화 67 145 131 29 0 23 2 5 15 4 0.25 0.221  
## 출루율 장타율 OPS wOBA WAR 연봉.2018. 연봉.2017.  
## 22 0.392 0.471 0.863 0.374 3.01 70000 70000  
## 23 0.413 0.545 0.958 0.401 2.96 160000 160000  
## 28 0.364 0.596 0.960 0.393 2.53 25000 23000  
## 30 0.356 0.476 0.832 0.354 2.29 24000 20000  
## 48 0.328 0.440 0.768 0.328 1.32 12000 9000  
## 61 0.371 0.402 0.773 0.344 0.65 8600 7600  
## 65 0.363 0.372 0.735 0.333 0.57 10000 8000  
## 81 0.336 0.302 0.638 0.297 0.29 8000 5800  
## 95 0.372 0.278 0.650 0.312 0.07 3800 3500  
## 98 0.343 0.388 0.731 0.323 0.05 5800 5500  
## 112 0.000 0.000 0.000 0.000 -0.04 3200 3200  
## 136 0.332 0.318 0.650 0.297 -0.14 40000 90000  
## 140 0.214 0.286 0.500 0.216 -0.16 3000 2700  
## 151 0.100 0.100 0.200 0.090 -0.24 3000 2900  
## 162 0.219 0.219 0.438 0.196 -0.35 4800 5200  
## 168 0.091 0.190 0.281 0.115 -0.43 3000 2700  
## 171 0.290 0.262 0.552 0.256 -0.49 5800 6200  
## 188 0.311 0.318 0.629 0.283 -1.08 11000 10000  
## 191 0.255 0.221 0.476 0.222 -1.32 3300 2800

dim(hanwha)

## [1] 19 21

str(hanwha)

## 'data.frame': 19 obs. of 21 variables:  
## $ 선수명 : chr "정근우" "김태균" "이성열" "송광민" ...  
## $ 팀명 : chr "한화" "한화" "한화" "한화" ...  
## $ 경기 : int 105 94 81 117 111 65 118 98 30 54 ...  
## $ 타석 : int 442 407 318 473 470 209 417 301 47 103 ...  
## $ 타수 : int 391 356 280 437 432 184 358 262 36 85 ...  
## $ 안타 : int 129 121 86 143 123 57 99 66 9 20 ...  
## $ 홈런 : int 11 17 21 13 11 2 2 1 0 3 ...  
## $ 득점 : int 73 51 41 71 69 23 42 21 7 10 ...  
## $ 타점 : int 46 76 65 75 52 21 40 15 2 13 ...  
## $ 볼넷 : int 32 43 19 21 23 15 38 25 7 13 ...  
## $ 삼진 : int 48 56 80 88 83 22 51 36 8 24 ...  
## $ 도루 : int 6 0 3 2 7 3 4 0 1 0 ...  
## $ BABIP : chr "0.352" "0.362" "0.353" "0.378" ...  
## $ 타율 : num 0.33 0.34 0.307 0.327 0.285 0.31 0.277 0.252 0.25 0.235 ...  
## $ 출루율 : num 0.392 0.413 0.364 0.356 0.328 0.371 0.363 0.336 0.372 0.343 ...  
## $ 장타율 : num 0.471 0.545 0.596 0.476 0.44 0.402 0.372 0.302 0.278 0.388 ...  
## $ OPS : num 0.863 0.958 0.96 0.832 0.768 0.773 0.735 0.638 0.65 0.731 ...  
## $ wOBA : num 0.374 0.401 0.393 0.354 0.328 0.344 0.333 0.297 0.312 0.323 ...  
## $ WAR : num 3.01 2.96 2.53 2.29 1.32 0.65 0.57 0.29 0.07 0.05 ...  
## $ 연봉.2018.: int 70000 160000 25000 24000 12000 8600 10000 8000 3800 5800 ...  
## $ 연봉.2017.: int 70000 160000 23000 20000 9000 7600 8000 5800 3500 5500 ...

head(hanwha)

## 선수명 팀명 경기 타석 타수 안타 홈런 득점 타점 볼넷 삼진 도루 BABIP 타율  
## 22 정근우 한화 105 442 391 129 11 73 46 32 48 6 0.352 0.330  
## 23 김태균 한화 94 407 356 121 17 51 76 43 56 0 0.362 0.340  
## 28 이성열 한화 81 318 280 86 21 41 65 19 80 3 0.353 0.307  
## 30 송광민 한화 117 473 437 143 13 71 75 21 88 2 0.378 0.327  
## 48 하주석 한화 111 470 432 123 11 69 52 23 83 7 0.33 0.285  
## 61 오선진 한화 65 209 184 57 2 23 21 15 22 3 0.34 0.310  
## 출루율 장타율 OPS wOBA WAR 연봉.2018. 연봉.2017.  
## 22 0.392 0.471 0.863 0.374 3.01 70000 70000  
## 23 0.413 0.545 0.958 0.401 2.96 160000 160000  
## 28 0.364 0.596 0.960 0.393 2.53 25000 23000  
## 30 0.356 0.476 0.832 0.354 2.29 24000 20000  
## 48 0.328 0.440 0.768 0.328 1.32 12000 9000  
## 61 0.371 0.402 0.773 0.344 0.65 8600 7600

tail(hanwha)

## 선수명 팀명 경기 타석 타수 안타 홈런 득점 타점 볼넷 삼진 도루 BABIP 타율  
## 151 박준혁 한화 8 10 10 1 0 1 0 0 4 0 0.167 0.100  
## 162 최윤석 한화 21 33 32 7 0 3 2 0 6 0 0.269 0.219  
## 168 김태연 한화 12 22 21 1 1 1 2 1 11 0 0 0.048  
## 171 강경학 한화 59 93 84 18 0 17 4 8 19 1 0.277 0.214  
## 188 장민석 한화 88 327 299 83 1 38 21 15 58 8 0.34 0.278  
## 191 이동훈 한화 67 145 131 29 0 23 2 5 15 4 0.25 0.221  
## 출루율 장타율 OPS wOBA WAR 연봉.2018. 연봉.2017.  
## 151 0.100 0.100 0.200 0.090 -0.24 3000 2900  
## 162 0.219 0.219 0.438 0.196 -0.35 4800 5200  
## 168 0.091 0.190 0.281 0.115 -0.43 3000 2700  
## 171 0.290 0.262 0.552 0.256 -0.49 5800 6200  
## 188 0.311 0.318 0.629 0.283 -1.08 11000 10000  
## 191 0.255 0.221 0.476 0.222 -1.32 3300 2800

## 결측치 확인

is.na(hanwha)

## 선수명 팀명 경기 타석 타수 안타 홈런 득점 타점 볼넷 삼진 도루  
## 22 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 23 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 28 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 30 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 48 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 61 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 65 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 81 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 95 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 98 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 112 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 136 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 140 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 151 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 162 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 168 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 171 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 188 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 191 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## BABIP 타율 출루율 장타율 OPS wOBA WAR 연봉.2018. 연봉.2017.  
## 22 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 23 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 28 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 30 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 48 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 61 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 65 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 81 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 95 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 98 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 112 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 136 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 140 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 151 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 162 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 168 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 171 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 188 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## 191 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

sum(is.na(hanwha)) # 결측치 0개

## [1] 0

## 특이값 찾기

boxplot.stats( hanwha$홈런 )$out # 특이값 21(이성열)

## [1] 21

boxplot.stats( hanwha$OPS )$out # 특이값 0

## [1] 0

boxplot.stats( hanwha$WAR )$out # 특이값 3.01(정근우), 2.96(김태균), 2.53(이성열)

## [1] 3.01 2.96

boxplot.stats( hanwha$연봉.2018.)$out # 특이값 70000(정근우), 160000(김태균), 25000(이성열)

## [1] 70000 160000 40000

## 한화 선수 분리

library( dplyr )

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

hanwha.player <- hanwha %>% select(선수명)  
hanwha.player

## 선수명  
## 22 정근우  
## 23 김태균  
## 28 이성열  
## 30 송광민  
## 48 하주석  
## 61 오선진  
## 65 양성우  
## 81 최재훈  
## 95 임익준  
## 98 김회성  
## 112 이창열  
## 136 이용규  
## 140 강상원  
## 151 박준혁  
## 162 최윤석  
## 168 김태연  
## 171 강경학  
## 188 장민석  
## 191 이동훈

length(hanwha.player$선수명) # 한화 선수 19명

## [1] 19

## 한화 선수별 홈런 정리

select <- dplyr::select   
  
homerun <- hanwha %>% select(선수명, 홈런) # 선수명 & 홈런  
homerun

## 선수명 홈런  
## 22 정근우 11  
## 23 김태균 17  
## 28 이성열 21  
## 30 송광민 13  
## 48 하주석 11  
## 61 오선진 2  
## 65 양성우 2  
## 81 최재훈 1  
## 95 임익준 0  
## 98 김회성 3  
## 112 이창열 0  
## 136 이용규 0  
## 140 강상원 0  
## 151 박준혁 0  
## 162 최윤석 0  
## 168 김태연 1  
## 171 강경학 0  
## 188 장민석 1  
## 191 이동훈 0

homerun %>% arrange(desc(홈런)) %>% head

## 선수명 홈런  
## 1 이성열 21  
## 2 김태균 17  
## 3 송광민 13  
## 4 정근우 11  
## 5 하주석 11  
## 6 김회성 3

mean(homerun$홈런)

## [1] 4.368421

quantile(homerun$홈런)

## 0% 25% 50% 75% 100%   
## 0 0 1 7 21

homerun2 <- homerun[,2] # 벡터형 변환  
homerun2

## [1] 11 17 21 13 11 2 2 1 0 3 0 0 0 0 0 1 0 1 0

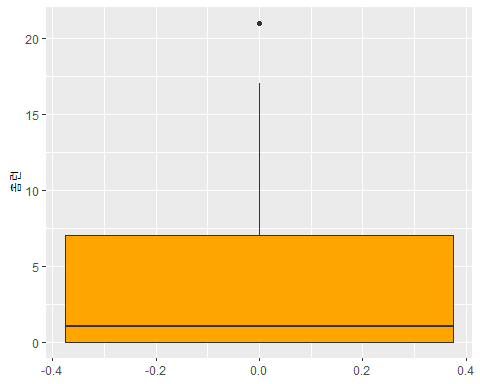
table( homerun2 ) # 도수분포

## homerun2  
## 0 1 2 3 11 13 17 21   
## 8 3 2 1 2 1 1 1

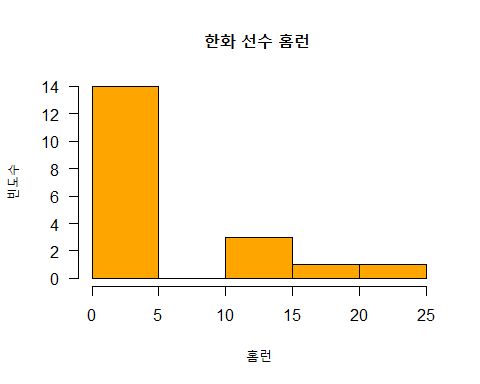
boxplot.stats(homerun2)

## $stats  
## [1] 0 0 1 7 17  
##   
## $n  
## [1] 19  
##   
## $conf  
## [1] -1.537338 3.537338  
##   
## $out  
## [1] 21

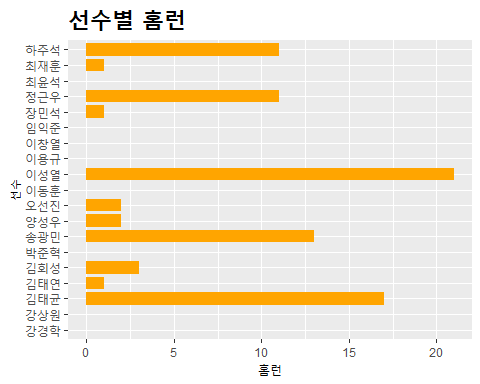
library(ggplot2)  
  
ggplot( homerun, aes( y = 홈런 ) ) + # 상자그림  
 geom\_boxplot( fill = "orange" )



hist( homerun2, main = "한화 선수 홈런 ", # 히스토그램  
 xlab = "홈런", ylab = "빈도수",  
 border = "black", col = "orange",  
 las = 1, breaks = 5 )

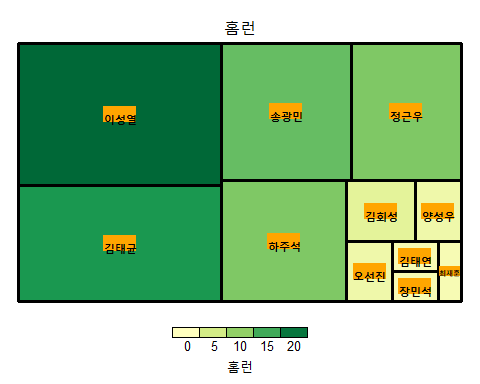


ggplot( homerun, aes( x = 선수명, y = 홈런 ) ) + # 막대그래프  
 geom\_bar( stat = "identity",   
 width = 0.8,   
 fill = "orange" ) +   
 ggtitle( "선수별 홈런" ) +   
 theme( plot.title = element\_text( size = 20,  
 face = "bold",  
 colour = "black" ) ) +  
 labs( x = "선수", y = "홈런" ) +  
 coord\_flip()



library( treemap )  
  
treemap( homerun, index = c( "선수명", "홈런" ), # 나무 지도  
 vSize = "홈런",  
 vColor = "홈런",  
 type = "value",  
 bg.labels = "orange" )

## Warning in if (class(try(col2rgb(bg.labels), silent = TRUE)) == "try-error")  
## stop("Invalid bg.labels"): length > 1 이라는 조건이 있고, 첫번째 요소만이 사용될  
## 것입니다



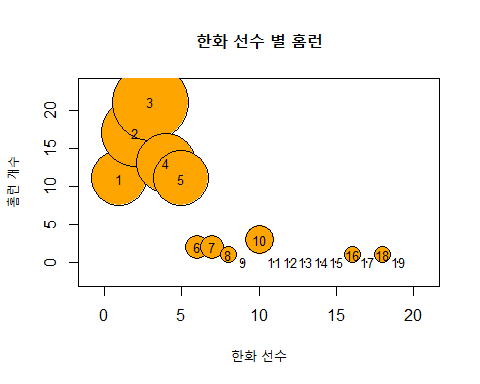
library( MASS )

##   
## Attaching package: 'MASS'

## The following object is masked \_by\_ '.GlobalEnv':  
##   
## select

## The following object is masked from 'package:dplyr':  
##   
## select

radius <- sqrt( homerun$홈런 ) # 원의 반지름  
symbols( homerun$홈런, homerun$y, # x, y 좌표 위치  
 circle = radius, # 원의 반지름  
 inches = 0.4, # 원의 크기 조절값  
 fg = "black", bg = "orange", # 원 테두리, 바탕색  
 lwd = 1.5, # 원의 테두리선 두께  
 xlab = "한화 선수", # x축 설명  
 ylab = "홈런 개수", # y축 설명  
 main = "한화 선수 별 홈런" ) # 제목  
text( homerun$홈런, homerun$y, # 문자 출력 x, y 위치  
 1:nrow( homerun ), # 문자로 출력할 값  
 cex = 0.8, # 글자 크기  
 col = "black" ) # 글자 색



## 한화 선수별 OPS 정리 ( OPS : 출루율과 장타율의 합, 타자의 성적 요약 지표 )

OPS <- hanwha %>% select(선수명, OPS) # 선수명 & OPS  
OPS

## 선수명 OPS  
## 22 정근우 0.863  
## 23 김태균 0.958  
## 28 이성열 0.960  
## 30 송광민 0.832  
## 48 하주석 0.768  
## 61 오선진 0.773  
## 65 양성우 0.735  
## 81 최재훈 0.638  
## 95 임익준 0.650  
## 98 김회성 0.731  
## 112 이창열 0.000  
## 136 이용규 0.650  
## 140 강상원 0.500  
## 151 박준혁 0.200  
## 162 최윤석 0.438  
## 168 김태연 0.281  
## 171 강경학 0.552  
## 188 장민석 0.629  
## 191 이동훈 0.476

OPS %>% arrange(desc(OPS)) %>% head

## 선수명 OPS  
## 1 이성열 0.960  
## 2 김태균 0.958  
## 3 정근우 0.863  
## 4 송광민 0.832  
## 5 오선진 0.773  
## 6 하주석 0.768

mean(OPS$OPS)

## [1] 0.6123158

quantile(OPS$OPS)

## 0% 25% 50% 75% 100%   
## 0.0000 0.4880 0.6500 0.7705 0.9600

OPS2<- OPS[,2] # 벡터형 변환  
OPS2

## [1] 0.863 0.958 0.960 0.832 0.768 0.773 0.735 0.638 0.650 0.731 0.000 0.650  
## [13] 0.500 0.200 0.438 0.281 0.552 0.629 0.476

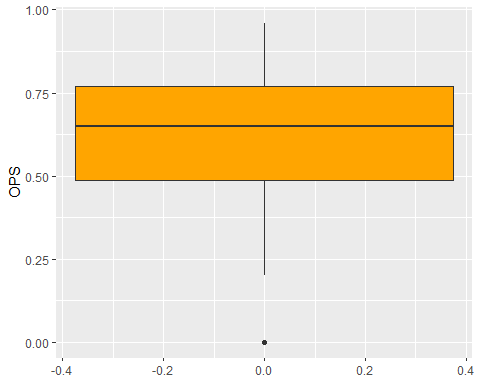
table( OPS2 ) # 도수분포

## OPS2  
## 0 0.2 0.281 0.438 0.476 0.5 0.552 0.629 0.638 0.65 0.731 0.735 0.768   
## 1 1 1 1 1 1 1 1 1 2 1 1 1   
## 0.773 0.832 0.863 0.958 0.96   
## 1 1 1 1 1

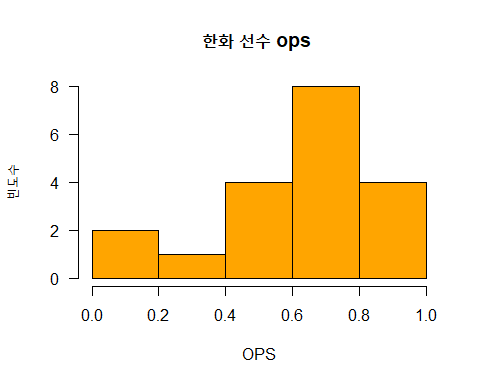
boxplot.stats(OPS2)

## $stats  
## [1] 0.2000 0.4880 0.6500 0.7705 0.9600  
##   
## $n  
## [1] 19  
##   
## $conf  
## [1] 0.5476003 0.7523997  
##   
## $out  
## [1] 0

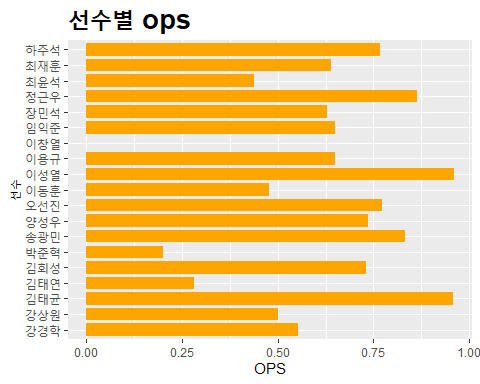
ggplot( OPS, aes( y = OPS ) ) + # 상자그림  
 geom\_boxplot( fill = "orange" )



hist( OPS2, main = "한화 선수 ops ", # 히스토그램  
 xlab = "OPS", ylab = "빈도수",  
 border = "black", col = "orange",  
 las = 1, breaks = 5 )

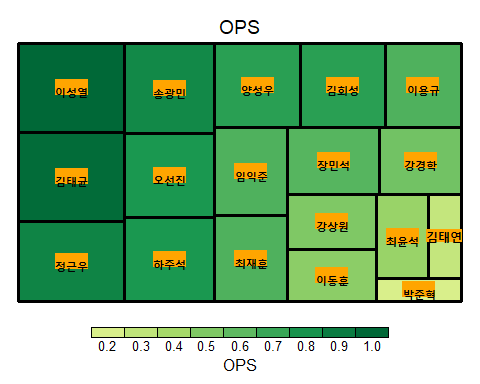


ggplot( OPS, aes( x = 선수명, y = OPS ) ) + # 막대그래프  
 geom\_bar( stat = "identity",   
 width = 0.8,   
 fill = "orange" ) +   
 ggtitle( "선수별 ops" ) +   
 theme( plot.title = element\_text( size = 20,  
 face = "bold",  
 colour = "black" ) ) +  
 labs( x = "선수", y = "OPS" ) +  
 coord\_flip()

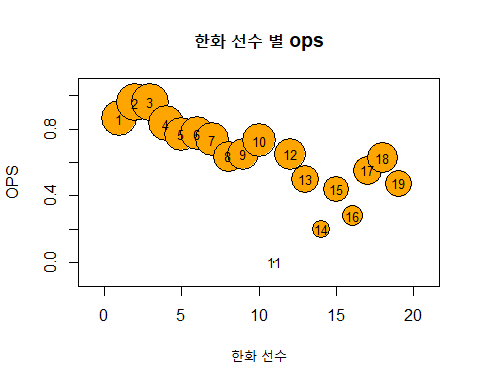


treemap( OPS, index = c( "선수명", "OPS" ), # 나무 지도  
 vSize = "OPS",  
 vColor = "OPS",  
 type = "value",  
 bg.labels = "orange" )

## Warning in if (class(try(col2rgb(bg.labels), silent = TRUE)) == "try-error")  
## stop("Invalid bg.labels"): length > 1 이라는 조건이 있고, 첫번째 요소만이 사용될  
## 것입니다



radius <- sqrt( OPS$OPS ) # 원의 반지름  
symbols( OPS$OPS, OPS$y, # x, y 좌표 위치  
 circle = radius, # 원의 반지름  
 inches = 0.2, # 원의 크기 조절값  
 fg = "black", bg = "orange", # 원 테두리, 바탕색  
 lwd = 1.5, # 원의 테두리선 두께  
 xlab = "한화 선수", # x축 설명  
 ylab = "OPS",  
 main = "한화 선수 별 ops") # y축 설명) # 제목  
text( OPS$OPS, OPS$y, # 문자 출력 x, y 위치  
 1:nrow( OPS ), # 문자로 출력할 값  
 cex = 0.8, # 글자 크기  
 col = "black" ) # 글자 색



## 한화 선수별 WAR 정리 ( WAR : 승리기여도 )

WAR <- hanwha %>% select(선수명, WAR) # 선수명 & WAR  
WAR

## 선수명 WAR  
## 22 정근우 3.01  
## 23 김태균 2.96  
## 28 이성열 2.53  
## 30 송광민 2.29  
## 48 하주석 1.32  
## 61 오선진 0.65  
## 65 양성우 0.57  
## 81 최재훈 0.29  
## 95 임익준 0.07  
## 98 김회성 0.05  
## 112 이창열 -0.04  
## 136 이용규 -0.14  
## 140 강상원 -0.16  
## 151 박준혁 -0.24  
## 162 최윤석 -0.35  
## 168 김태연 -0.43  
## 171 강경학 -0.49  
## 188 장민석 -1.08  
## 191 이동훈 -1.32

WAR %>% arrange(desc(WAR)) %>% head

## 선수명 WAR  
## 1 정근우 3.01  
## 2 김태균 2.96  
## 3 이성열 2.53  
## 4 송광민 2.29  
## 5 하주석 1.32  
## 6 오선진 0.65

mean(WAR$WAR)

## [1] 0.4994737

quantile(WAR$WAR)

## 0% 25% 50% 75% 100%   
## -1.320 -0.295 0.050 0.985 3.010

WAR2 <- WAR[,2] # 벡터형 변환  
WAR2

## [1] 3.01 2.96 2.53 2.29 1.32 0.65 0.57 0.29 0.07 0.05 -0.04 -0.14  
## [13] -0.16 -0.24 -0.35 -0.43 -0.49 -1.08 -1.32

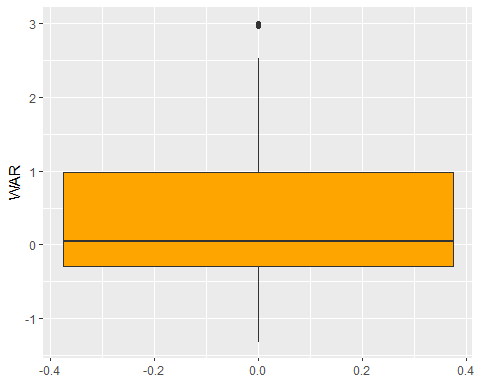
table( WAR2 ) # 도수분포

## WAR2  
## -1.32 -1.08 -0.49 -0.43 -0.35 -0.24 -0.16 -0.14 -0.04 0.05 0.07 0.29 0.57   
## 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 0.65 1.32 2.29 2.53 2.96 3.01   
## 1 1 1 1 1 1

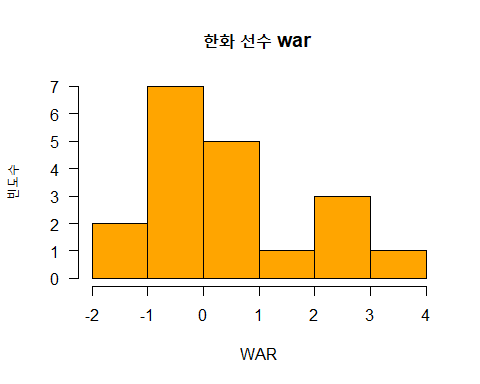
boxplot.stats(WAR2)

## $stats  
## [1] -1.320 -0.295 0.050 0.985 2.530  
##   
## $n  
## [1] 19  
##   
## $conf  
## [1] -0.4139704 0.5139704  
##   
## $out  
## [1] 3.01 2.96

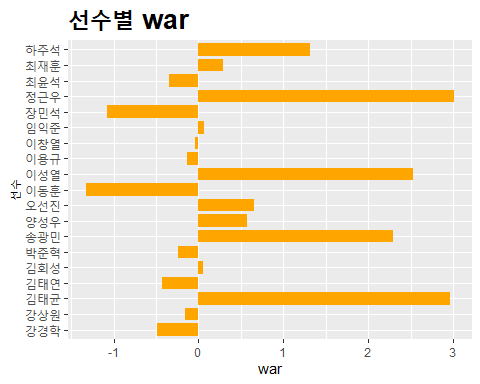
ggplot( WAR, aes( y = WAR ) ) + # 상자그림  
 geom\_boxplot( fill = "orange" )



hist( WAR2, main = "한화 선수 war ", # 히스토그램  
 xlab = "WAR", ylab = "빈도수",  
 border = "black", col = "orange",  
 las = 1, breaks = 5 )



ggplot( WAR, aes( x = 선수명, y = WAR ) ) + # 막대그래프  
 geom\_bar( stat = "identity",   
 width = 0.8,   
 fill = "orange" ) +   
 ggtitle( "선수별 war" ) +   
 theme( plot.title = element\_text( size = 20,  
 face = "bold",  
 colour = "black" ) ) +  
 labs( x = "선수", y = "war" ) +  
 coord\_flip()



## 한화 선수별 연봉.2018.정리

annual.income <- hanwha %>% select(선수명,연봉.2018. ) # 선수명 & 홈런  
annual.income

## 선수명 연봉.2018.  
## 22 정근우 70000  
## 23 김태균 160000  
## 28 이성열 25000  
## 30 송광민 24000  
## 48 하주석 12000  
## 61 오선진 8600  
## 65 양성우 10000  
## 81 최재훈 8000  
## 95 임익준 3800  
## 98 김회성 5800  
## 112 이창열 3200  
## 136 이용규 40000  
## 140 강상원 3000  
## 151 박준혁 3000  
## 162 최윤석 4800  
## 168 김태연 3000  
## 171 강경학 5800  
## 188 장민석 11000  
## 191 이동훈 3300

annual.income %>% arrange(desc(연봉.2018.)) %>% head

## 선수명 연봉.2018.  
## 1 김태균 160000  
## 2 정근우 70000  
## 3 이용규 40000  
## 4 이성열 25000  
## 5 송광민 24000  
## 6 하주석 12000

mean(annual.income$연봉.2018.)

## [1] 21278.95

quantile(annual.income$연봉.2018)

## 0% 25% 50% 75% 100%   
## 3000 3550 8000 18000 160000

annual.income2 <- annual.income[,2] # 벡터형 변환  
annual.income2

## [1] 70000 160000 25000 24000 12000 8600 10000 8000 3800 5800  
## [11] 3200 40000 3000 3000 4800 3000 5800 11000 3300

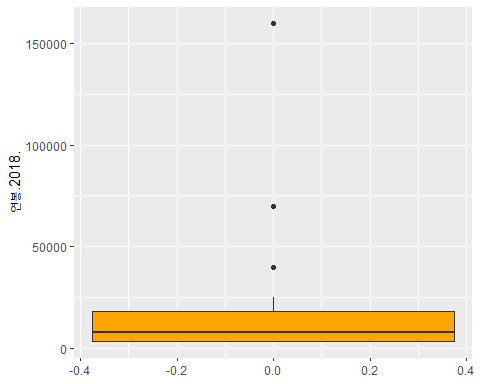
table( annual.income2 ) # 도수분포

## annual.income2  
## 3000 3200 3300 3800 4800 5800 8000 8600 10000 11000 12000   
## 3 1 1 1 1 2 1 1 1 1 1   
## 24000 25000 40000 70000 160000   
## 1 1 1 1 1

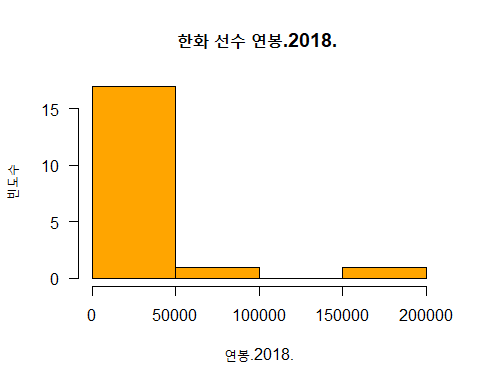
boxplot.stats(annual.income2)

## $stats  
## [1] 3000 3550 8000 18000 25000  
##   
## $n  
## [1] 19  
##   
## $conf  
## [1] 2762.209 13237.791  
##   
## $out  
## [1] 70000 160000 40000

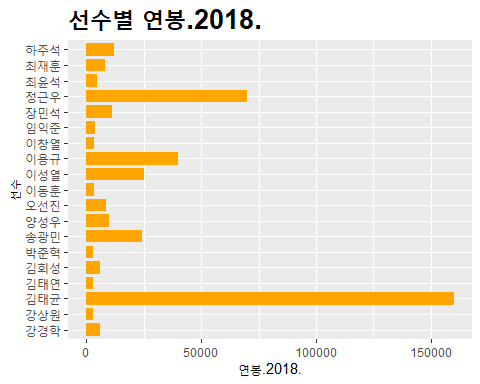
library(ggplot2)  
  
ggplot( annual.income, aes( y = 연봉.2018. ) ) + # 상자그림  
 geom\_boxplot( fill = "orange" )



hist( annual.income2, main = "한화 선수 연봉.2018. ", # 히스토그램  
 xlab = "연봉.2018.", ylab = "빈도수",  
 border = "black", col = "orange",  
 las = 1, breaks = 5 )

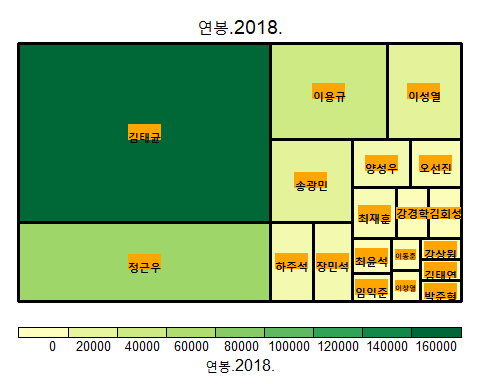


ggplot( annual.income, aes( x = 선수명, y = 연봉.2018. ) ) + # 막대그래프  
 geom\_bar( stat = "identity",   
 width = 0.8,   
 fill = "orange" ) +   
 ggtitle( "선수별 연봉.2018." ) +   
 theme( plot.title = element\_text( size = 20,  
 face = "bold",  
 colour = "black" ) ) +  
 labs( x = "선수", y = "연봉.2018." ) +  
 coord\_flip()

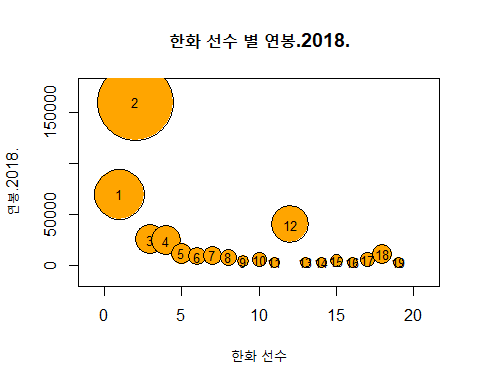


treemap( annual.income, index = c( "선수명", "연봉.2018." ), # 나무 지도  
 vSize = "연봉.2018.",  
 vColor = "연봉.2018.",  
 type = "value",  
 bg.labels = "orange" )

## Warning in if (class(try(col2rgb(bg.labels), silent = TRUE)) == "try-error")  
## stop("Invalid bg.labels"): length > 1 이라는 조건이 있고, 첫번째 요소만이 사용될  
## 것입니다



radius <- sqrt( annual.income$연봉.2018. ) # 원의 반지름  
symbols( annual.income$연봉.2018., annual.income$y, # x, y 좌표 위치  
 circle = radius, # 원의 반지름  
 inches = 0.4, # 원의 크기 조절값  
 fg = "black", bg = "orange", # 원 테두리, 바탕색  
 lwd = 1.5, # 원의 테두리선 두께  
 xlab = "한화 선수", # x축 설명  
 ylab = "연봉.2018.", # y축 설명  
 main = "한화 선수 별 연봉.2018." ) # 제목  
text( annual.income$연봉.2018., annual.income$y, # 문자 출력 x, y 위치  
 1:nrow( annual.income ), # 문자로 출력할 값  
 cex = 0.8, # 글자 크기  
 col = "black" ) # 글자 색



## 변수들 간의 상관관계 분석( 홈런, OPS, WAR, 연봉.2018. ) : 그룹별 관측값 분포 확인, 다중 산점도

library( corrplot )

## corrplot 0.84 loaded

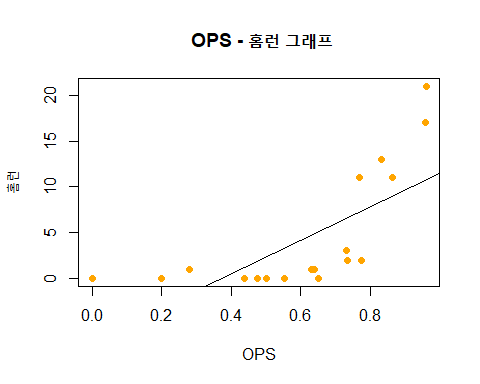
H.O <- data.frame( homerun, OPS)  
H.O

## 선수명 홈런 선수명.1 OPS  
## 22 정근우 11 정근우 0.863  
## 23 김태균 17 김태균 0.958  
## 28 이성열 21 이성열 0.960  
## 30 송광민 13 송광민 0.832  
## 48 하주석 11 하주석 0.768  
## 61 오선진 2 오선진 0.773  
## 65 양성우 2 양성우 0.735  
## 81 최재훈 1 최재훈 0.638  
## 95 임익준 0 임익준 0.650  
## 98 김회성 3 김회성 0.731  
## 112 이창열 0 이창열 0.000  
## 136 이용규 0 이용규 0.650  
## 140 강상원 0 강상원 0.500  
## 151 박준혁 0 박준혁 0.200  
## 162 최윤석 0 최윤석 0.438  
## 168 김태연 1 김태연 0.281  
## 171 강경학 0 강경학 0.552  
## 188 장민석 1 장민석 0.629  
## 191 이동훈 0 이동훈 0.476

H.O.D <- H.O[,-c(1,3)] # 복수 열(선수명) 제거  
H.O.D

## 홈런 OPS  
## 22 11 0.863  
## 23 17 0.958  
## 28 21 0.960  
## 30 13 0.832  
## 48 11 0.768  
## 61 2 0.773  
## 65 2 0.735  
## 81 1 0.638  
## 95 0 0.650  
## 98 3 0.731  
## 112 0 0.000  
## 136 0 0.650  
## 140 0 0.500  
## 151 0 0.200  
## 162 0 0.438  
## 168 1 0.281  
## 171 0 0.552  
## 188 1 0.629  
## 191 0 0.476

plot( 홈런~OPS, data= H.O, # OPS - 홈런 상관도  
 main = " OPS - 홈런 그래프",  
 xlab = "OPS",  
 ylab = "홈런",  
 col = "orange",  
 pch = 19)  
  
res.a <- lm( 홈런~OPS, data = H.O.D ) # 회귀식 도출  
  
abline( res.a ) # 회귀선 그리기



cor( H.O.D$홈런, H.O.D$OPS ) # 상관계수 : 0.6962551

## [1] 0.6962551

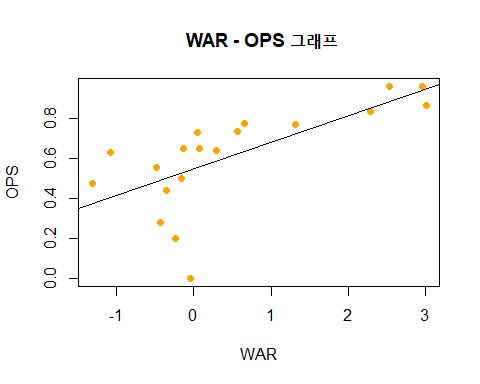
O.W <- data.frame( OPS, WAR)  
O.W

## 선수명 OPS 선수명.1 WAR  
## 22 정근우 0.863 정근우 3.01  
## 23 김태균 0.958 김태균 2.96  
## 28 이성열 0.960 이성열 2.53  
## 30 송광민 0.832 송광민 2.29  
## 48 하주석 0.768 하주석 1.32  
## 61 오선진 0.773 오선진 0.65  
## 65 양성우 0.735 양성우 0.57  
## 81 최재훈 0.638 최재훈 0.29  
## 95 임익준 0.650 임익준 0.07  
## 98 김회성 0.731 김회성 0.05  
## 112 이창열 0.000 이창열 -0.04  
## 136 이용규 0.650 이용규 -0.14  
## 140 강상원 0.500 강상원 -0.16  
## 151 박준혁 0.200 박준혁 -0.24  
## 162 최윤석 0.438 최윤석 -0.35  
## 168 김태연 0.281 김태연 -0.43  
## 171 강경학 0.552 강경학 -0.49  
## 188 장민석 0.629 장민석 -1.08  
## 191 이동훈 0.476 이동훈 -1.32

O.W.D <- O.W[,-c(1,3)] # 복수 열(선수명) 제거  
O.W.D

## OPS WAR  
## 22 0.863 3.01  
## 23 0.958 2.96  
## 28 0.960 2.53  
## 30 0.832 2.29  
## 48 0.768 1.32  
## 61 0.773 0.65  
## 65 0.735 0.57  
## 81 0.638 0.29  
## 95 0.650 0.07  
## 98 0.731 0.05  
## 112 0.000 -0.04  
## 136 0.650 -0.14  
## 140 0.500 -0.16  
## 151 0.200 -0.24  
## 162 0.438 -0.35  
## 168 0.281 -0.43  
## 171 0.552 -0.49  
## 188 0.629 -1.08  
## 191 0.476 -1.32

plot( OPS~WAR, data= O.W, # OPS - WAR 상관도  
 main = " WAR - OPS 그래프",  
 xlab = "WAR",  
 ylab = "OPS",  
 col = "orange",  
 pch = 19)  
  
res.b <- lm( OPS~WAR, data = O.W.D ) # 회귀식 도출  
  
abline( res.b ) # 회귀선 그리기



cor( O.W.D$WAR, O.W.D$OPS ) # 상관계수 : 0.6847353

## [1] 0.6847353

hanwha

## 선수명 팀명 경기 타석 타수 안타 홈런 득점 타점 볼넷 삼진 도루 BABIP 타율  
## 22 정근우 한화 105 442 391 129 11 73 46 32 48 6 0.352 0.330  
## 23 김태균 한화 94 407 356 121 17 51 76 43 56 0 0.362 0.340  
## 28 이성열 한화 81 318 280 86 21 41 65 19 80 3 0.353 0.307  
## 30 송광민 한화 117 473 437 143 13 71 75 21 88 2 0.378 0.327  
## 48 하주석 한화 111 470 432 123 11 69 52 23 83 7 0.33 0.285  
## 61 오선진 한화 65 209 184 57 2 23 21 15 22 3 0.34 0.310  
## 65 양성우 한화 118 417 358 99 2 42 40 38 51 4 0.316 0.277  
## 81 최재훈 한화 98 301 262 66 1 21 15 25 36 0 0.289 0.252  
## 95 임익준 한화 30 47 36 9 0 7 2 7 8 1 0.321 0.250  
## 98 김회성 한화 54 103 85 20 3 10 13 13 24 0 0.283 0.235  
## 112 이창열 한화 3 1 1 0 0 1 0 0 0 0 0 0.000  
## 136 이용규 한화 57 200 179 47 0 31 12 17 20 10 0.294 0.263  
## 140 강상원 한화 33 14 14 3 0 8 1 0 5 0 0.333 0.214  
## 151 박준혁 한화 8 10 10 1 0 1 0 0 4 0 0.167 0.100  
## 162 최윤석 한화 21 33 32 7 0 3 2 0 6 0 0.269 0.219  
## 168 김태연 한화 12 22 21 1 1 1 2 1 11 0 0 0.048  
## 171 강경학 한화 59 93 84 18 0 17 4 8 19 1 0.277 0.214  
## 188 장민석 한화 88 327 299 83 1 38 21 15 58 8 0.34 0.278  
## 191 이동훈 한화 67 145 131 29 0 23 2 5 15 4 0.25 0.221  
## 출루율 장타율 OPS wOBA WAR 연봉.2018. 연봉.2017.  
## 22 0.392 0.471 0.863 0.374 3.01 70000 70000  
## 23 0.413 0.545 0.958 0.401 2.96 160000 160000  
## 28 0.364 0.596 0.960 0.393 2.53 25000 23000  
## 30 0.356 0.476 0.832 0.354 2.29 24000 20000  
## 48 0.328 0.440 0.768 0.328 1.32 12000 9000  
## 61 0.371 0.402 0.773 0.344 0.65 8600 7600  
## 65 0.363 0.372 0.735 0.333 0.57 10000 8000  
## 81 0.336 0.302 0.638 0.297 0.29 8000 5800  
## 95 0.372 0.278 0.650 0.312 0.07 3800 3500  
## 98 0.343 0.388 0.731 0.323 0.05 5800 5500  
## 112 0.000 0.000 0.000 0.000 -0.04 3200 3200  
## 136 0.332 0.318 0.650 0.297 -0.14 40000 90000  
## 140 0.214 0.286 0.500 0.216 -0.16 3000 2700  
## 151 0.100 0.100 0.200 0.090 -0.24 3000 2900  
## 162 0.219 0.219 0.438 0.196 -0.35 4800 5200  
## 168 0.091 0.190 0.281 0.115 -0.43 3000 2700  
## 171 0.290 0.262 0.552 0.256 -0.49 5800 6200  
## 188 0.311 0.318 0.629 0.283 -1.08 11000 10000  
## 191 0.255 0.221 0.476 0.222 -1.32 3300 2800

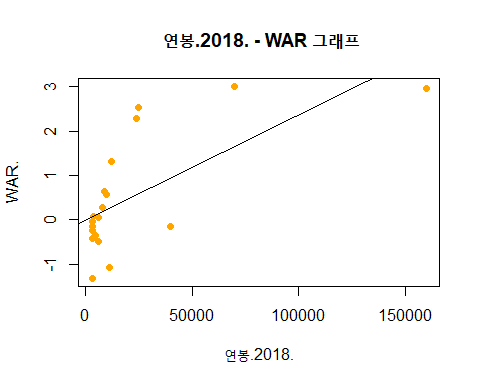
W.2018 <- data.frame( WAR, annual.income )  
W.2018

## 선수명 WAR 선수명.1 연봉.2018.  
## 22 정근우 3.01 정근우 70000  
## 23 김태균 2.96 김태균 160000  
## 28 이성열 2.53 이성열 25000  
## 30 송광민 2.29 송광민 24000  
## 48 하주석 1.32 하주석 12000  
## 61 오선진 0.65 오선진 8600  
## 65 양성우 0.57 양성우 10000  
## 81 최재훈 0.29 최재훈 8000  
## 95 임익준 0.07 임익준 3800  
## 98 김회성 0.05 김회성 5800  
## 112 이창열 -0.04 이창열 3200  
## 136 이용규 -0.14 이용규 40000  
## 140 강상원 -0.16 강상원 3000  
## 151 박준혁 -0.24 박준혁 3000  
## 162 최윤석 -0.35 최윤석 4800  
## 168 김태연 -0.43 김태연 3000  
## 171 강경학 -0.49 강경학 5800  
## 188 장민석 -1.08 장민석 11000  
## 191 이동훈 -1.32 이동훈 3300

W.2018.D <- W.2018[,-c(1,3)] # 복수 열(선수명) 제거  
W.2018.D

## WAR 연봉.2018.  
## 22 3.01 70000  
## 23 2.96 160000  
## 28 2.53 25000  
## 30 2.29 24000  
## 48 1.32 12000  
## 61 0.65 8600  
## 65 0.57 10000  
## 81 0.29 8000  
## 95 0.07 3800  
## 98 0.05 5800  
## 112 -0.04 3200  
## 136 -0.14 40000  
## 140 -0.16 3000  
## 151 -0.24 3000  
## 162 -0.35 4800  
## 168 -0.43 3000  
## 171 -0.49 5800  
## 188 -1.08 11000  
## 191 -1.32 3300

plot( WAR~연봉.2018., data = W.2018, # WAR - 연봉.2018. 상관도  
 main = " 연봉.2018. - WAR 그래프",  
 xlab = "연봉.2018.",  
 ylab = "WAR.",  
 col = "orange",  
 pch = 19)  
  
res.c <- lm( WAR~연봉.2018., data = W.2018.D ) # 회귀식 도출  
  
abline( res.c ) # 회귀선 그리기



cor( W.2018.D$WAR, W.2018.D$연봉.2018. ) # 0.6800321

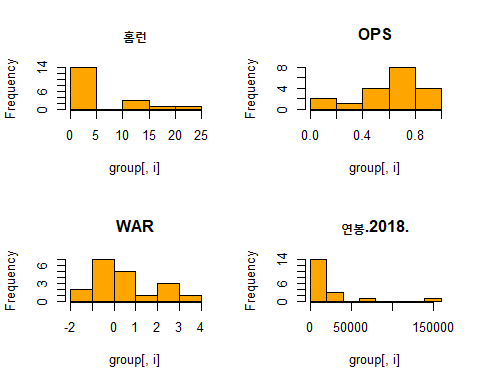
## [1] 0.6800321

## Data.frame 변수 그룹 묶기

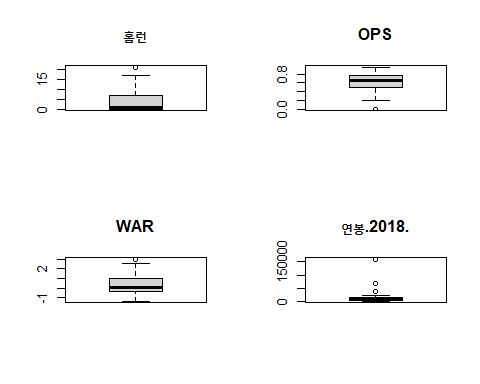
group <- hanwha[,c("홈런","OPS","WAR","연봉.2018.")] # 사용할 변수 그룹 지정  
group

## 홈런 OPS WAR 연봉.2018.  
## 22 11 0.863 3.01 70000  
## 23 17 0.958 2.96 160000  
## 28 21 0.960 2.53 25000  
## 30 13 0.832 2.29 24000  
## 48 11 0.768 1.32 12000  
## 61 2 0.773 0.65 8600  
## 65 2 0.735 0.57 10000  
## 81 1 0.638 0.29 8000  
## 95 0 0.650 0.07 3800  
## 98 3 0.731 0.05 5800  
## 112 0 0.000 -0.04 3200  
## 136 0 0.650 -0.14 40000  
## 140 0 0.500 -0.16 3000  
## 151 0 0.200 -0.24 3000  
## 162 0 0.438 -0.35 4800  
## 168 1 0.281 -0.43 3000  
## 171 0 0.552 -0.49 5800  
## 188 1 0.629 -1.08 11000  
## 191 0 0.476 -1.32 3300

par( mfrow = c( 2, 2 ) ) # 변수별 히스토그램  
for ( i in 1:4 ) {  
 hist( group[ , i ], main = colnames( group )[ i ],  
 col = "orange" )}



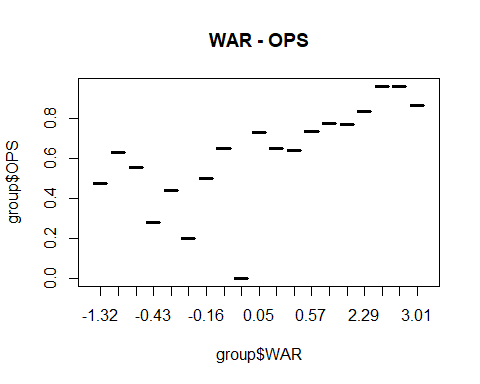
for ( i in 1:4 ) { # 상자 그래프에 의한 관측값 분포 확인  
 boxplot( group[ , i ], main = colnames( group )[ i ] )  
}



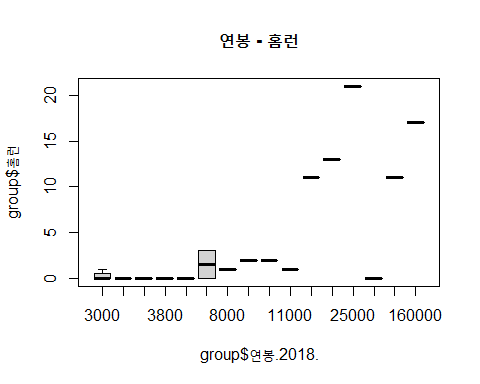
par( mfrow = c( 1, 1 ) )

## 그룹별 관측값 분포 확인

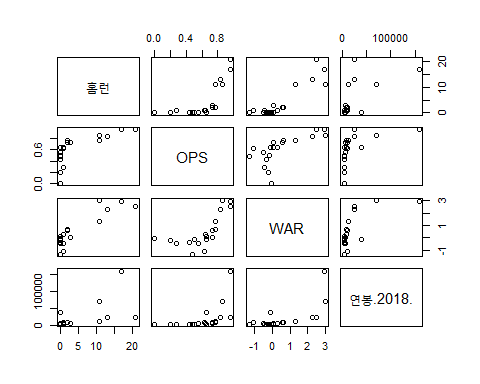
boxplot( group$OPS~group$WAR, main = "WAR - OPS" ) # OPS : 1.0 이하



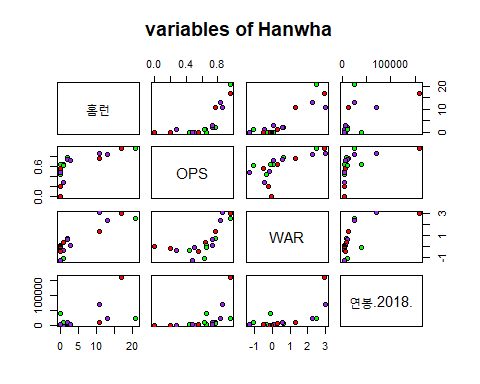
boxplot( group$홈런~group$연봉.2018., main = "연봉 - 홈런" )



pairs( group[ , -6 ] ) # 다중 산점도(변수 간 상관관계)



pairs(~홈런+OPS+WAR+연봉.2018., main = "variables of Hanwha", pch = 21, bg = c("Purple","Red","green"), data = group) # 다중 산점도 색 입히기



cor( group[ , -6 ] ) # 변수들 간의 상관계수 종합

## 홈런 OPS WAR 연봉.2018.  
## 홈런 1.0000000 0.6962551 0.8966346 0.6155394  
## OPS 0.6962551 1.0000000 0.6847353 0.5135825  
## WAR 0.8966346 0.6847353 1.0000000 0.6800321  
## 연봉.2018. 0.6155394 0.5135825 0.6800321 1.0000000

## 결론

한화 이글스 선수들을 4가지 변수(홈런, OPS(출루율과 장타율의 합, 타자의 성적 요약 지표), WAR(승리 기여도), 연봉.2018.)으로 분석해 본 결과 홈런(21개, 1등), OPS(0.960, 1등), WAR(2.53, 3등)은 구단 내 최상위권 이고 연봉(25000)은 적절한 **이성열**선수가 가장 가성비 좋은 선수라고 볼 수 있다.