

DOT Diagram 개요

개체의 값(관측값, 요약 통계량)을 이차원 그래프에 표현

개체를 x-y 축 한 축에 지정, 그에 대응하는 관측값은 y 혹은 x 축에 지정된다.

1) 데이터 관측치가 있는 경우

미국 MLB 타자 연봉 데이터 (http://wolfpack.hnu.ac.kr/Stat_Notes/example_data/baseball.csv)

첫 타자 30명 데이터 활용

```
baseball<-read.csv('http://wolfpack.hnu.ac.kr/Stat_Notes/example_data/
baseball.csv')
names(baseball);dim(baseball)
baseball.subset<-baseball[c(1:30),]
```

```
> names(baseball);dim(baseball)
```

| | | | |
|-------------------|--------------------|---------------|--------------|
| [1] "Player_Name" | "Team" | "TimesatBat" | "Hits" |
| [5] "HomeRuns" | "Runs" | "RBIs" | "Walks" |
| [9] "YearsinMLB" | "CareerTimesatBat" | "CareerHits" | "CareerHome" |
| [13] "CareerRuns" | "CareerRBIs" | "CareerWalks" | "League" |
| [17] "Division" | "Position" | "PutOuts" | "Assists" |
| [21] "Errors" | "Salary" | "League2" | |
| [1] 322 | 23 | | |

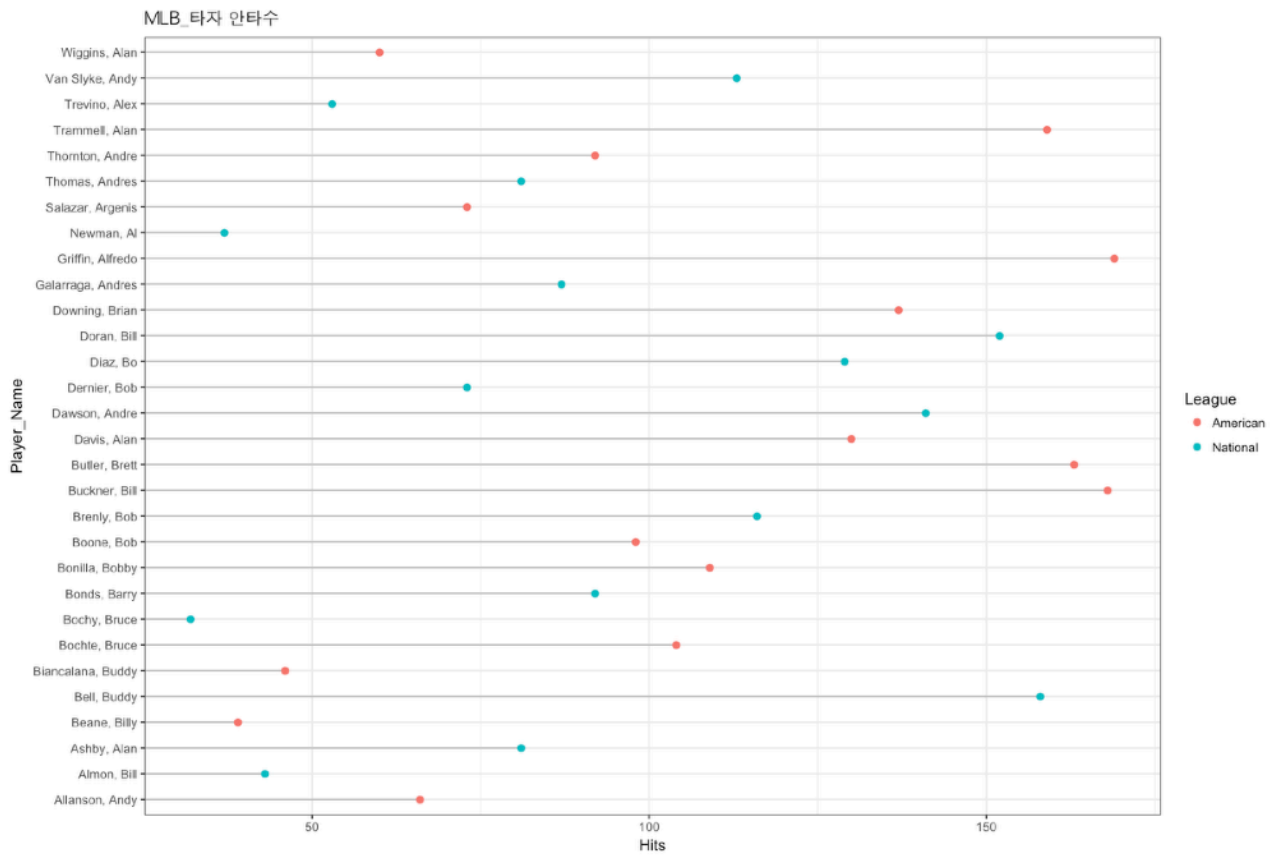
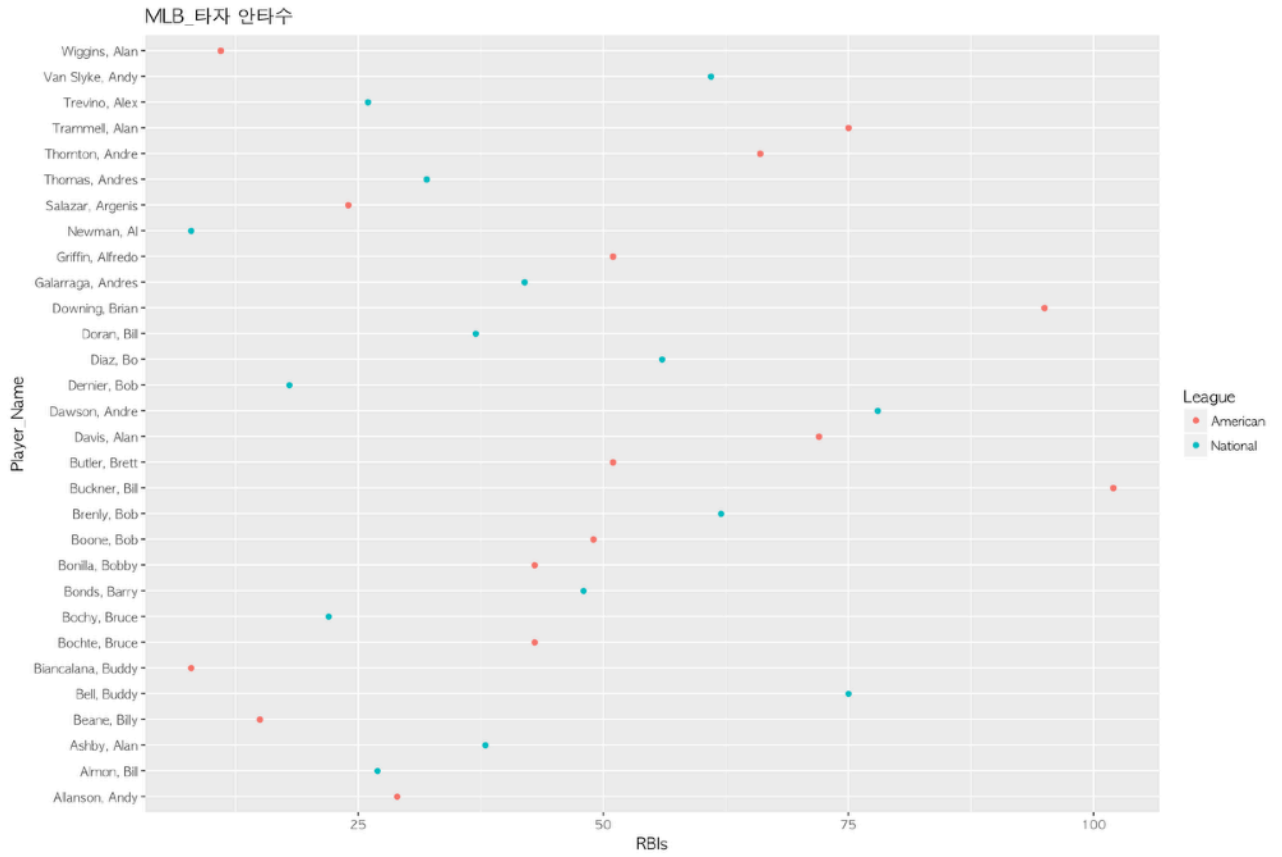
선수이름-y축, 안타수 - x-축, 그리고 점 구별 - 리그

ggplot() : x-축, y-축 변수 설정, 그리고 GROUP변수 기본적 사항을 지정한다. 모든 ggplot()은 이 스크립터로 시작한다.

나머지 설정은 + geom_???() 을 활용한다.

```
require(ggplot2)
ggplot(baseball.subset,aes(x=Hits,y=Player_Name,col=League)) +
  geom_point() +
  ggtitle('MLB_타자 안타수')
```

```
ggplot(baseball.subset,aes(x=Hits,y=Player_Name)) +
  geom_segment(aes(yend=Player_Name,xend=0,color='grey')) +
  geom_point(size=2,aes(color=League)) +
  ggtitle('MLB_타자 안타수') +
  theme_bw()
```



2) 데이터 - 요약 통계량

MLB 선수 포지션별 타자 평균

```
baseball.by<-aggregate(Hits~League+Position,baseball,FUN=mean,na.rm=T)
library(ggplot2)
ggplot(baseball.by,aes(x=Hits,y=Position)) +
  geom_point(size=2,aes(color=League)) +
  ggtitle('MLB_타자 안타수') +
  theme_bw()
```

