HW 1 solutions

Stats and sports class

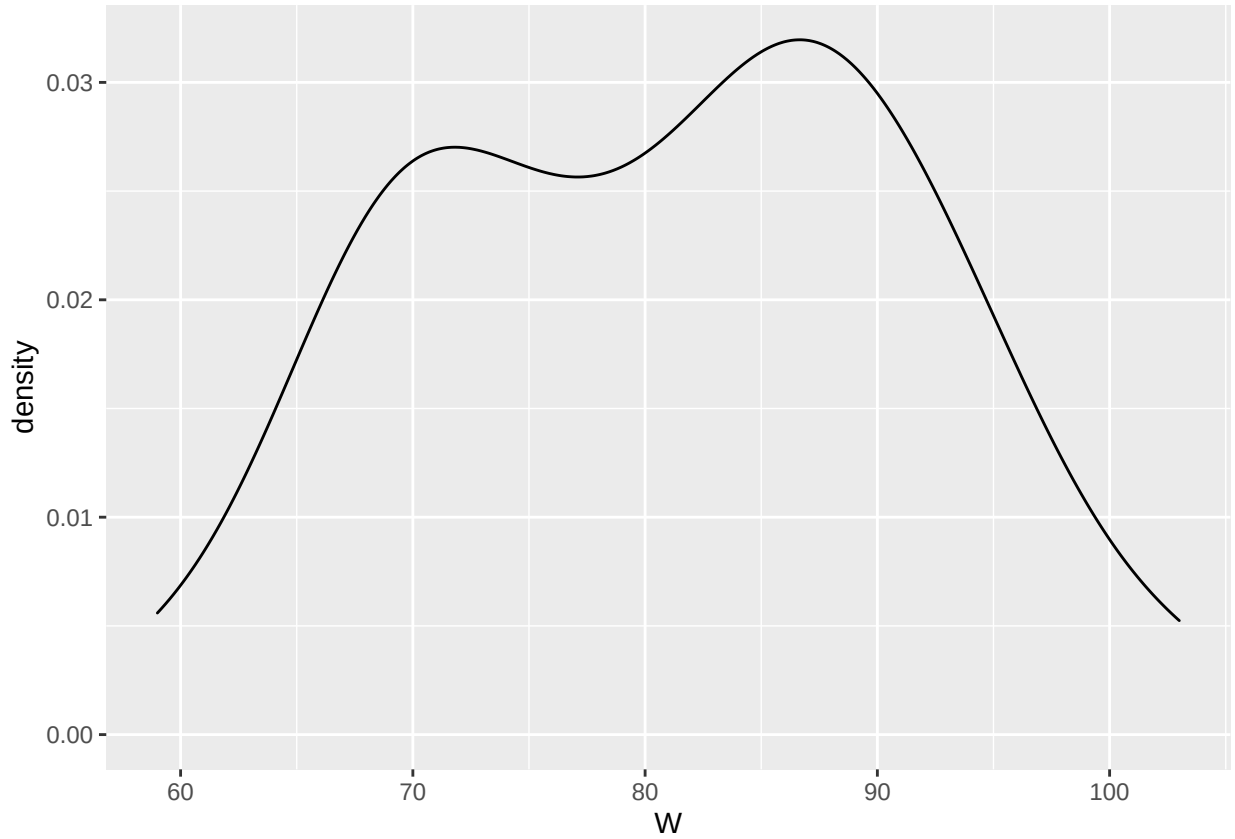
```
library(tidyverse)
library(Lahman)
teams_2016 <- Teams %>% filter(yearID == 2016)
teams_2016_batting <- teams_2016 %>% select (yearID:teamID, R:SF)
```

1. Make an appropriate graph of team wins during this season. Is the distribution of wins skewed left, right, or symmetric?

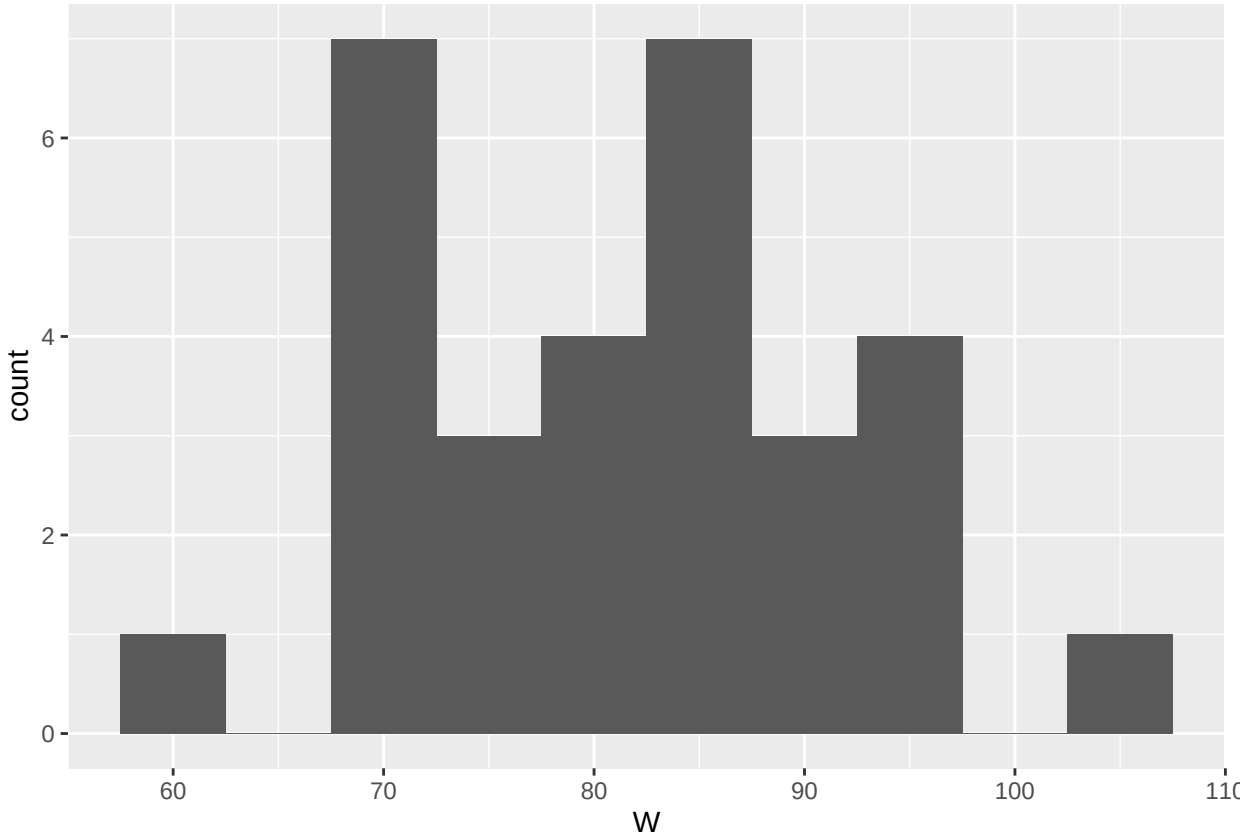
Answers

A histogram or density plot are likely the most appropriate plots.

```
ggplot(teams_2016, aes(W)) + geom_density()
```



```
ggplot(teams_2016, aes(W)) + geom_histogram(binwidth = 5)
```



Wins is roughly symmetric – answering bimodal is reasonable. There does not seem to be any skewness.

10. *Estimate* the correlation between (i) slugging percentage and runs, (ii) on base percentage and runs and (iii) batting average and runs. Which would you prioritize as a coach using these results? Why?

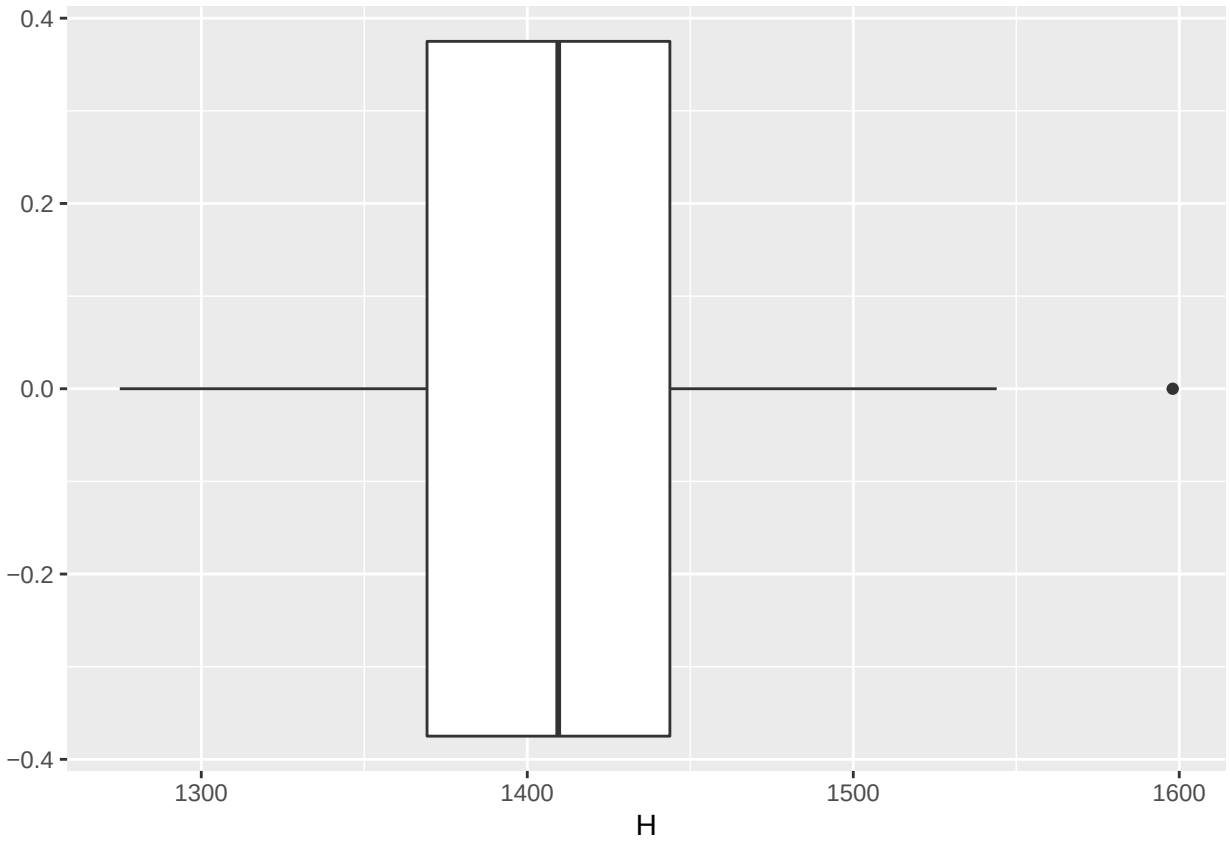
Answers

These are rough estimates only, but somewhere in the neighborhood of 0.5 or 0.7 are reasonable. The correlation between slugging and on base with runs is stronger than the correlation between batting average and runs. As a result, prioritizing the two metrics that more closely link with runs score (slugging or on base) would make sense as a coach.

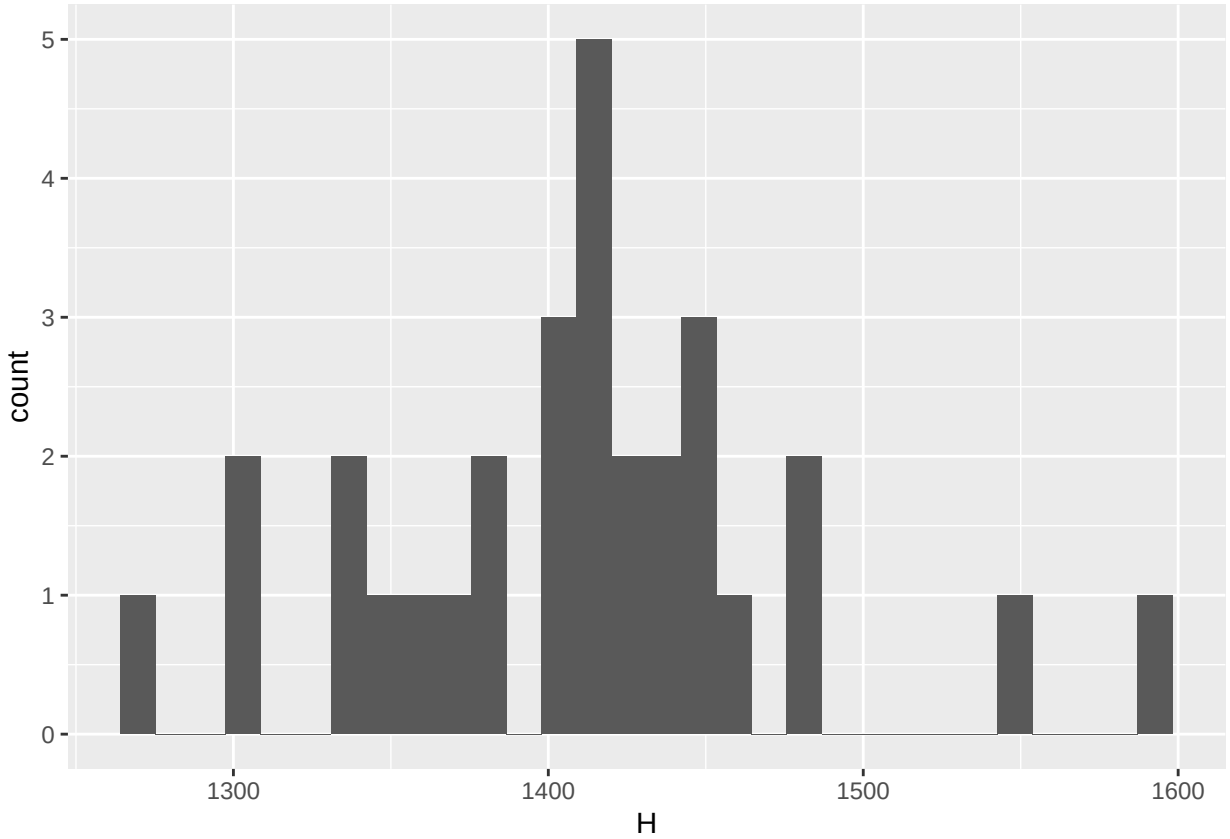
11. Make both a histogram and a boxplot of `hits`. What features are apparent in the histogram that aren't apparent in the boxplot? What features are apparent in the boxplot that aren't apparent in the histogram?

Answers Boxplots explicitly show the median, as well as any extreme outliers. Histograms can make it a bit easier to discover shape/skewness.

```
ggplot(teams_2016, aes(H)) + geom_boxplot()
```




```
ggplot(teams_2016, aes(H)) + geom_histogram()
```



Part II

Read Voros McCracken's "Pitching and Defense: How Much Control Do Hurlers Have?", provided here and also on the reading page.

1. What is McCracken's primary finding?

Answer McCracken's main finding is this – "There is little if any difference among major-league pitchers in their ability to prevent hits on balls hit in the field of play." That is, pitchers control walks, home runs, strikeouts, but for balls in play, most of it is luck whether or not the ball ends up as a hit.

2. Why would traditional baseball followers feel surprised with this result?

Answer Primarily, leaving "luck" as the main driver of hits on balls in play reduces the amount of skill that is generally assumed to exist in baseball.

3. What might one consider to supplement McCracken's analysis?

Answers will vary