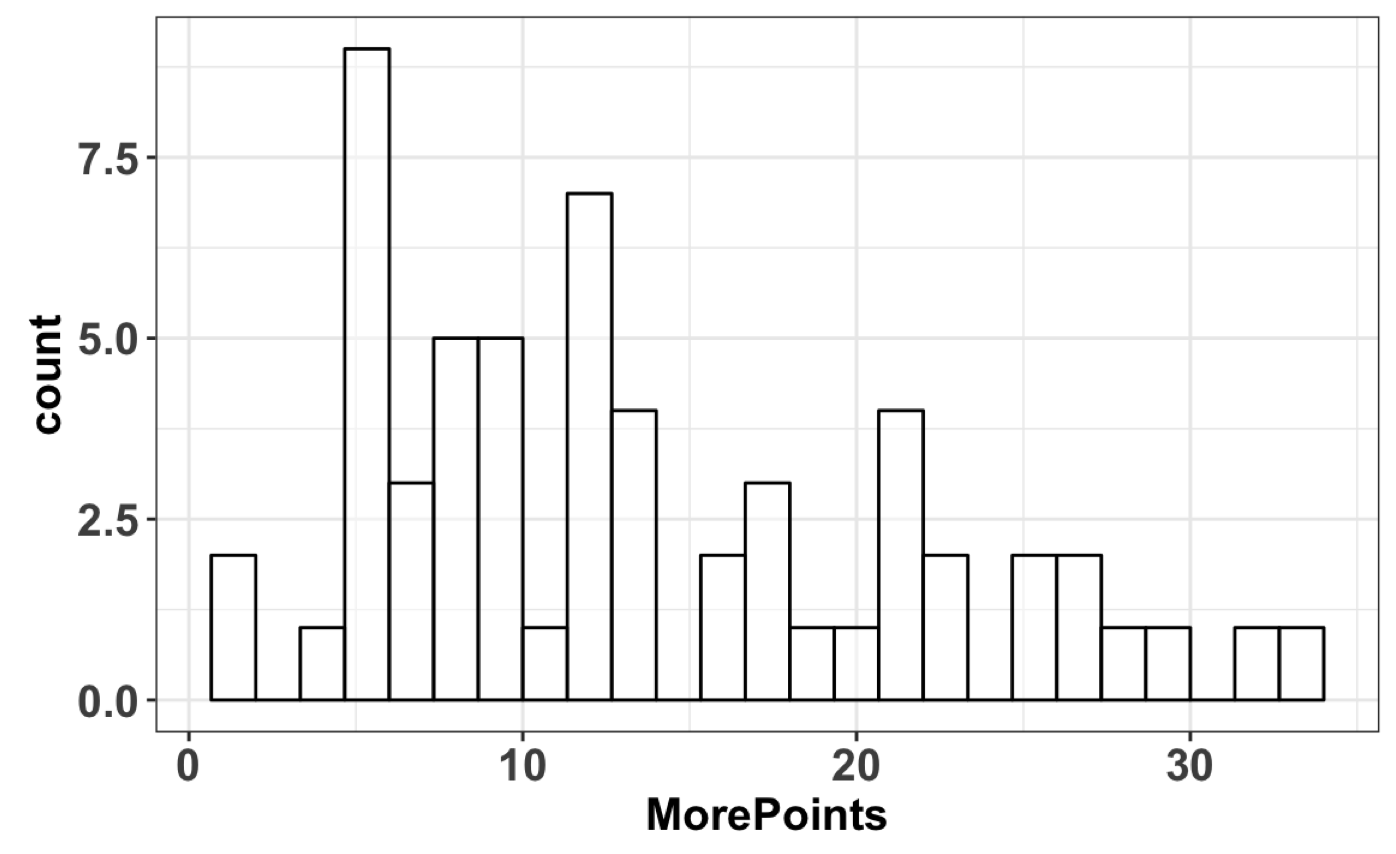
**Name:**

**Classwork 10**

1. Now let’s think like a bookie. If you had to predict whether Miami Heat was *going to win* their next game based on this data set, would it be helpful to use the empty model of **Points**? Why or why not?
2. Now let’s say you knew a little information about the upcoming game. Let’s say it’s against TOR (Toronto Raptors). Would you predict win or lose? What do you think all the people making bets would predict? Why is that a problem for your business as a bookie?
3. Now let’s say you (and unfortunately everyone else) thinks that Miami Heat will win the next game. We’ll have to learn a little bit about why bookie’s use point spreads [mini-lecture]. What is a point spread and how does a point spread help you? (Note: This isn’t EXACTLY point spread – the real way point spread is calculated is a bit more involved – we need to have way more data! And a different kind of data set would be very helpful. But here we will write the basic concept… caution -- I’m not really a bookie! If you want to know more about point spreads, look here: <https://www.thesportsgeek.com/sports-betting/strategy/point-spread/>)
4. Now that online sports betting is legal, there are a ton of different bets you can make. Above or below point spread is one kind of bet. Key: In a good point spread (good from the bookie’s point of view), bets on points should roughly equally fall above and below the point spread. Does that idea sound kind of like the mean or median? How?
5. What would be the reason for using the mean instead of the median for the point spread?
6. As baby bookies (we’re developing!), we want to get data on how many MORE points the Miami Heat score when they win. How would we get this data out of this data set? (Save the data frame as **MiamiWins** and save the variable you create as **MorePoints**.)
7. If we had predicted the mean point difference as our point spread for the games in our **MiamiWins** data set, would we have been right most of the time? How many times?
8. If we had predicted the mean as the point spread for the games in our **MiamiWins** data set, on average, how off would we have been? Draw it and have a sense of how big it *should* be before you use R.



1. Why is this distribution skewed in the way that it is?
2. Let’s say we want to know, what is the likelihood that the point difference will be 0 or greater? Maybe we can just use our data as a model. What would this likelihood be if we used the **MiamiWins** data as our model? What if we used **MiamiHeat**?
3. We know that the next season of games will not be exactly like this season of games. What might be similar about the distribution? What might be dissimilar? (Think about shape, center, spread.)
4. If we wanted to know the likelihood that the point difference will be 0 or greater, should we use the data as the model or the normal distribution as a model? Will the normal distribution be a better model for all of Miami Heat’s games in general (**MiamiHeat**) or for the games when they win (**MiamiWins**)? Is it possible that even the MiamiWins data were generated by a normal data generating process (DGP)?
5. Based on your discussion and thinking in the previous question, use the normal distribution to estimate the likelihood that they will have a point spread of 0 or more.
6. As an additional exercise, use R to depict the best fitting normal distribution, mean, and 0 on top of your histogram. Does that picture match the one you got from xpnorm()?

