ESSAI

Volume 3   Article 30

Spring 2005

Case Study: Sports - Baseball Stadiums

David Rojas
College of DuPage

Follow this and additional works at: http://dc.cod.edu/essai

Recommended Citation

Rojas, David (2005) "Case Study: Sports - Baseball Stadiums," *ESSAI*: Vol. 3, Article 30.
Available at: http://dc.cod.edu/essai/vol3/iss1/30

Case Study: Sports - Baseball Stadiums

by David Rojas

(Math 135)

The Assignment: Write an article about Major League ballparks and their effects on the success of the teams that play in them.

One aspect of Major League Baseball that is fascinating to study is the ballparks. For example, is there a correlation between the ballparks of Major League Baseball teams and the performance of the corresponding teams? An analysis of the attendance of the ballparks and the percentage of the games won by the teams, among other things, might prove to be informative in this matter.

The data for this project were all taken from the 1997 season.

Section 1: Capacity and Age

To begin, let us examine the capacities of various ballparks. Comparing the frequency distributions of the capacities of ballparks of the American and National Leagues, one comes to the conclusion that the ballparks of the National League are generally larger than those of the American League (see Figures 1.1 through 1.3).
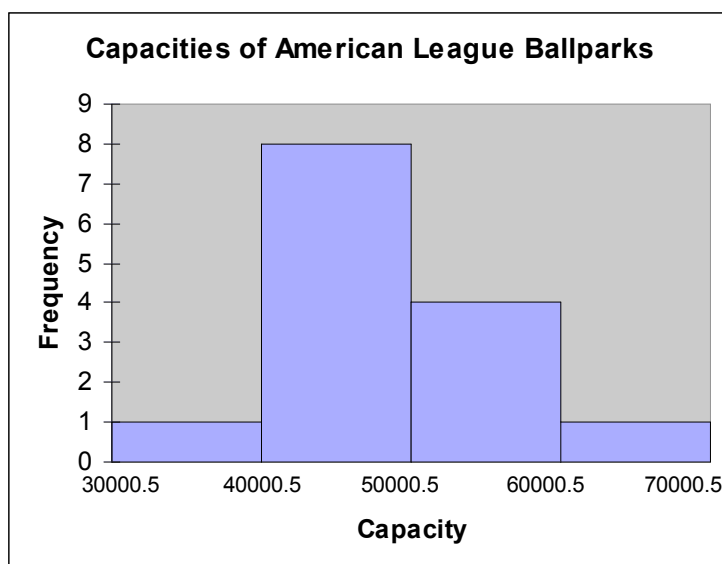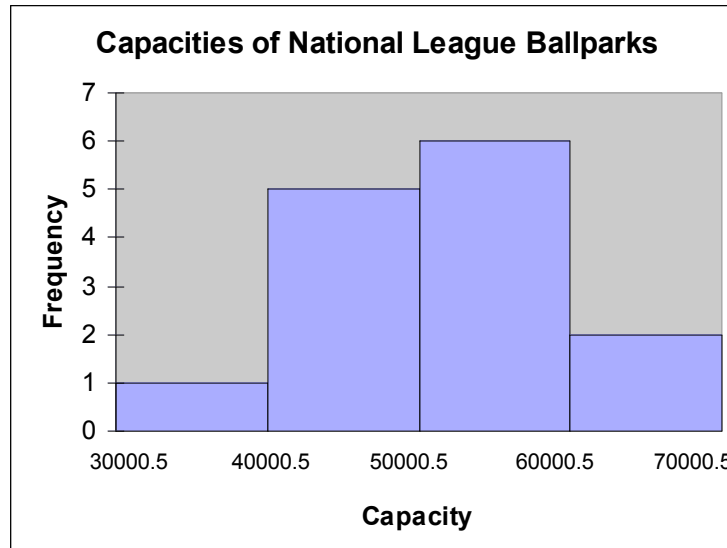

Figure 1.1

125

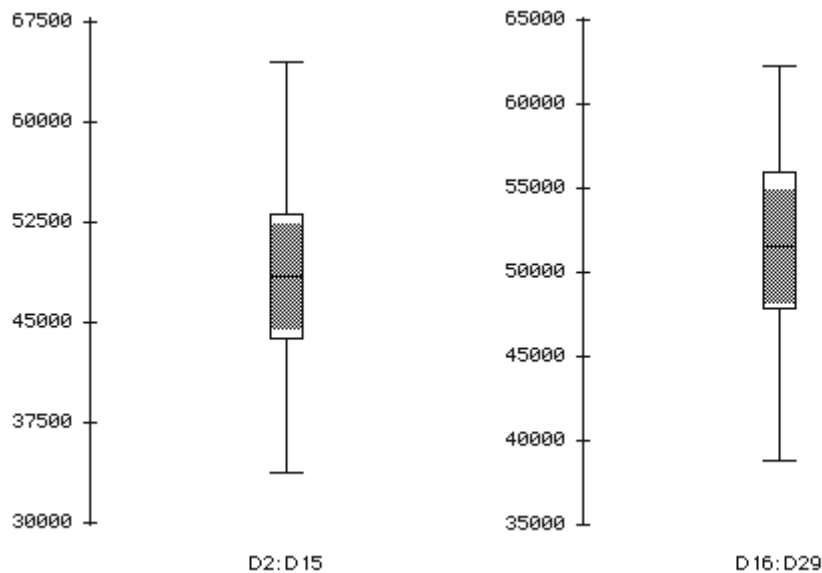**Capacities of National League Ballparks**

Figure 1.2

Figure 1.3

(The image on the left represents the American League ballpark capacities, while the one on the right represents the National League ballpark capacities.)

(The class boundaries of Figures 1.1 and 1.2 were found by taking the upper and lower class limits of adjacent classes and finding the mean. For example, the upper class limit of the first class is 30000 and the lower class limit of the second class is 30001. The mean of this is 30000.5.)

Another comparison can be obtained by comparing the means of the ballparks of the two leagues. The mean capacity of the American League ballparks is 48935.4 persons (determined by adding all of the American League capacities and dividing by fourteen), while the mean capacity of the National League ballparks is 51996.5 persons. This

126

shows a definite difference in general size of the ballparks.

(In addition, the variation among the capacities of the American League ballparks is greater than that of the capacities of the National League ballparks: the standard deviation of the former is 8012.1 persons, while the standard deviation of the latter is 7008.8. The mean absolute deviation of the American League ballparks is 5893.7 (obtained from dividing the sum of the absolute value of all of the differences between the mean and the various values by the number of values), while the mean absolute deviation of the National League ballparks is 5596.4. Both measures of deviation show larger variety among the American League ballparks than among the National ones.)

One possible explanation for the size difference might come from the comparison of when the ballparks of the two leagues were built (see Figures 1.4 and 1.5).
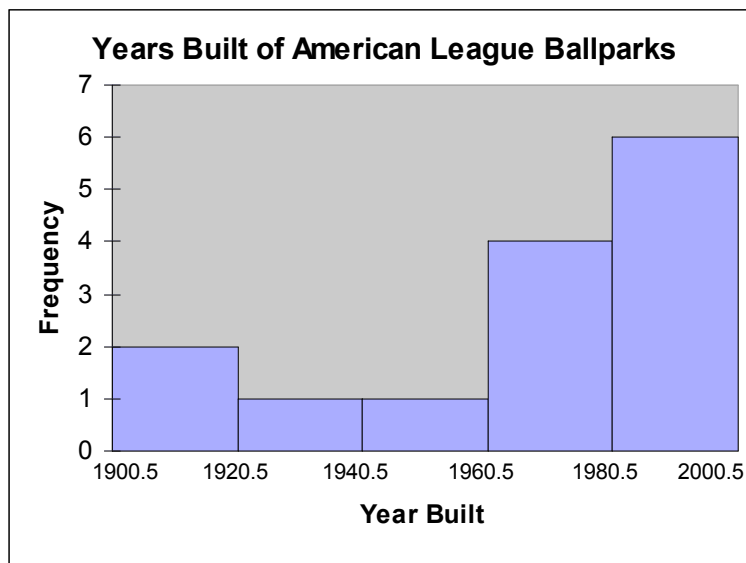

Figure 1.4
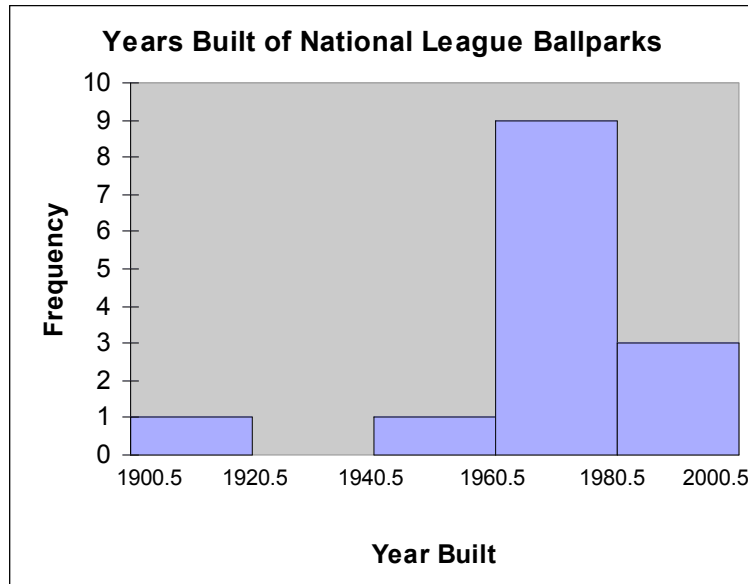
127

Years Built of National League Ballparks

Figure 1.5

The mean year built of the American League ballparks is 1965.9, while the mean year for the National League ballparks is 1968.8. This means that there were generally more recent ballparks built in the National League than in the American League. Since the efficiency and cost-effectiveness of building ballparks in more recent years is likely to be greater than that of earlier last century (in addition, the U.S. population and interest in baseball are only increasing – hence, why more ballparks are being built in the first place), this might explain why, in general, National League ballparks have greater capacities.

One might argue that a difference in the means of about three years is insignificant. However, if one looks at the distributions in even larger segments, one can see that there were twelve National ballparks (twelve out of fourteen is 85.7%) built since 1960, but only ten American ones (71.4%). Yet there were three American ballparks (21.4%) built before 1940, but only one National one (7.1%). The shape of the graphs indicate that, while neither league had particularly many ballparks built earlier in the century, the American League had more, while the National League had a stronger presence in the last half of the century.

One reason there might be a smaller difference between the two means than one might expect is because the concentration of American ballparks since 1960 is more heavily weighted in the 1980-2000 segment, while the concentration of National ballparks is more heavily weighted in the 1960-1980 segment. Simply looking at the shape of the graphs, with the emphasis of the American League histogram on the most recent segment and the emphasis of the National League histogram on the second-most recent segment, demonstrates this.

(In addition, the variation among the years built of the American League ballparks is greater than that of the years built of the National League ballparks (much like that of the

128

capacities): the standard deviation of the former is 29.9 years, while the standard deviation of the latter is 19.6. The mean absolute deviation of the American League ballparks is 23.4, while the mean absolute deviation of the National League ballparks is 11.9. A comparison of both measures of deviation show larger variety among the years built of American League ballparks than among the National ones.)

This may or may not be a valid explanation (that, *since* the American League ballparks were built earlier in general, their capacities are smaller), as the numbers *themselves* do not make such a conclusion and this is only speculation regarding the discrepancy of the capacities of the two leagues. The data somewhat support correlation between the capacity and age of the two leagues, but not necessarily causality.

Section 2: **Attendance and Performance**

The next aspect of Major League ballparks that we will examine is the comparison of game attendance and the percentage of games won by the corresponding teams. We will begin by studying the American League, then the National League, and finally juxtaposing the two results. The way we will evaluate any correlation between game attendance and team performance is by comparing the graphs and seeing if there is any similarities between them (see Figures 2.1 and 2.2).
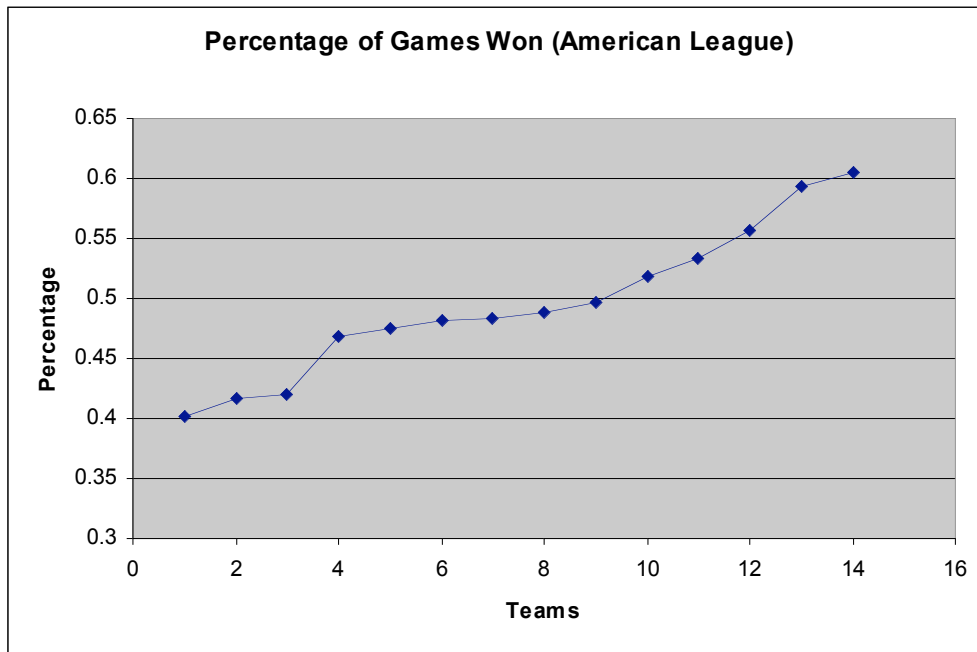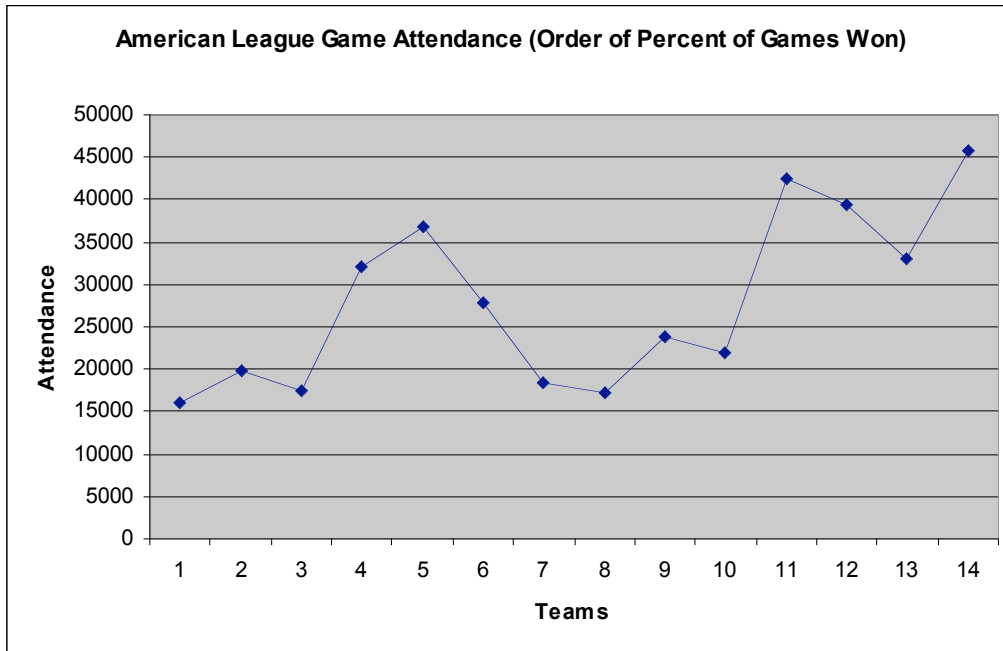

Figure 2.1

129

Figure 2.2

The first figure shows the percentage of games won by each team (lined up in order of smallest percentage to largest). The second figure graphs the game attendance of the ballparks of each team; the teams are ordered from least percentage of games won to most. While the second graph varies greatly especially near the middle teams, both show a general increase in size. The first team, the Oakland Athletics (with the worst record of 0.401), has the smallest game attendance (15965 persons); the last team, the Baltimore Orioles (with the best record of 0.605), has the largest game attendance (45816 persons).

Let us examine another way of determining that there is a correlation between the two. The mean game attendance of the ballparks of the American League teams is 28013 persons. If you divide the teams in half, and start by examining the worst seven teams, you find that their ballparks' mean game attendance was 24031.1 persons. The best seven teams' ballparks' mean game attendance, on the other hand, was 31994.9 persons. We can see that the smallest value is the worst teams' game attendance mean, followed by overall mean, followed by the best teams' game attendance mean. Small attendance to large attendance corresponds with bad record to good record.

Now we will take a look at the game attendance and team performance of the National League in the same fashion as that of the American League (see Figures 2.3 and 2.4).
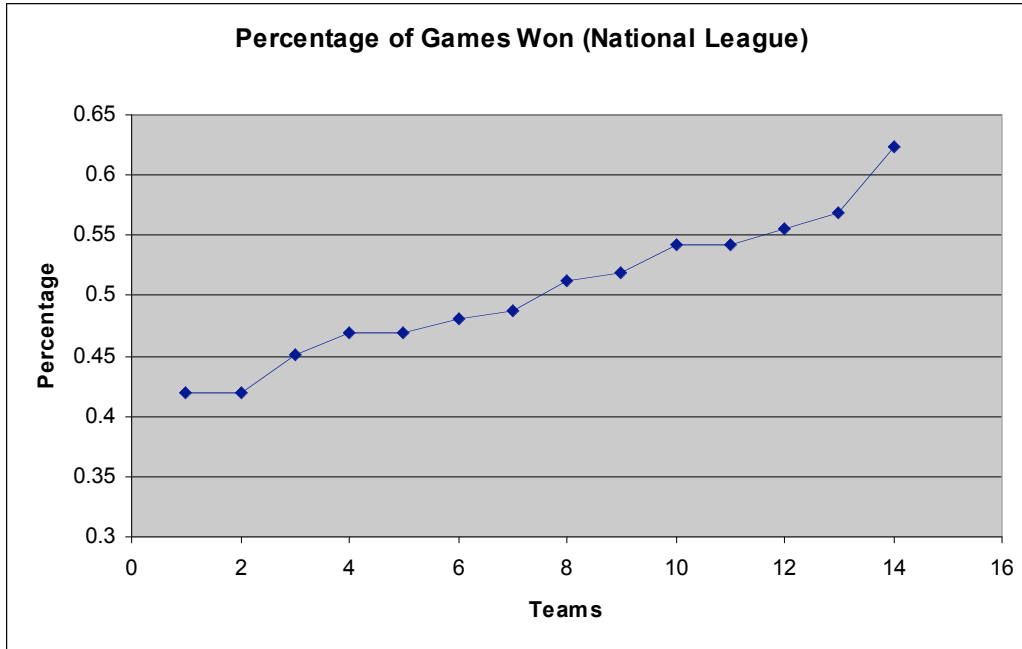
130

Percentage of Games Won (National League)

Figure 2.3

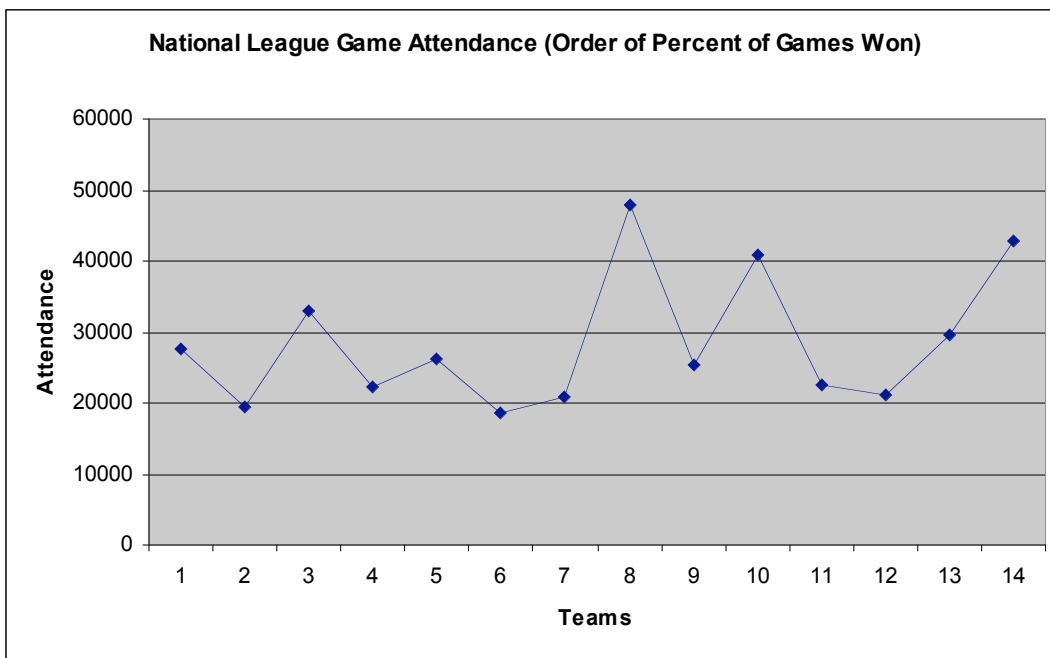**National League Game Attendance (Order of Percent of Games Won)**

Figure 2.4

Once again, there is a general upward trend, although not as pronounced as that of the American League. The mean of the lower seven teams is 23934.9, the overall mean is 28420.8, and the mean of the better seven teams is 32906.7. Again, this shows an upward trend consistent with the idea that there is a correlation between overall performance and ballpark game attendance.

131

One possible explanation for a correlation of this sort is that, as a team does progressively well, there is a much greater incentive for people to go to its games. However, for those teams that do not do as well (especially if they have a historical record of not doing well), their ballparks' game attendance will likely not be as great.

The variation among the ballpark game attendances of the American League teams is greater than that of the National League. The standard deviation for American League ballpark game attendance is 10251.1, while that of the National League is 9393.5. The mean absolute deviation for American League ballpark game attendance is 8808, while the mean absolute deviation for National League ballpark game attendance is 7314.2.
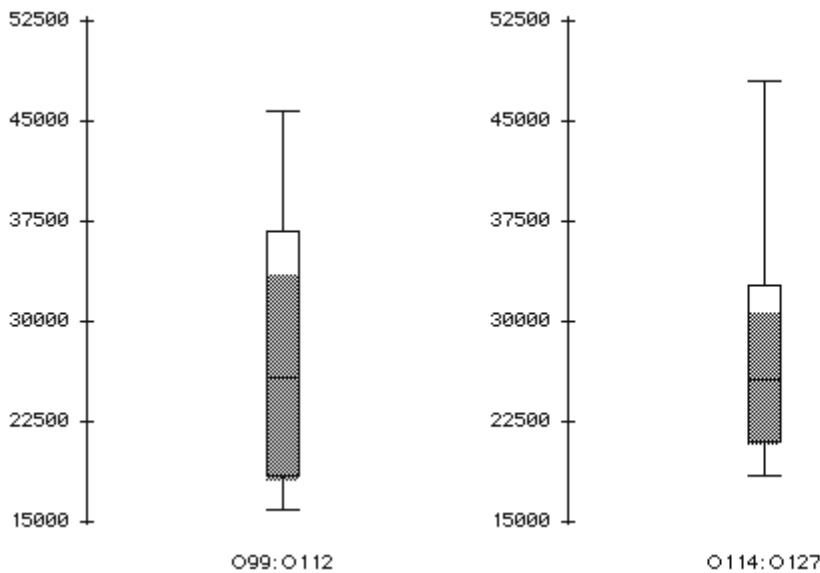

Figure 2.5
(The image on the left represents the American League game attendances, while the one on the right represents the National League game attendances.)

Figure 2.5 shows that, although the minimum and maximum of the National League are greater than that of the American League, one can see that the general centers of the distributions are similar. In addition, the variations are made more obvious with the boxplots (the National League concentration segment is much smaller than that of the American League).

Section 3: **Attendance Percentage and Performance**

In this section, we will follow a somewhat similar process as that of Section 2. However, instead of dealing with the average game attendance, we will look at how full the ballparks were. We will compare the percentage of the ballpark that was filled (in other words, game attendance divided by ballpark capacity) to the performance of the various teams (again, the percentage of games won). To begin, we will look at the American League data (see Figure 3.1).
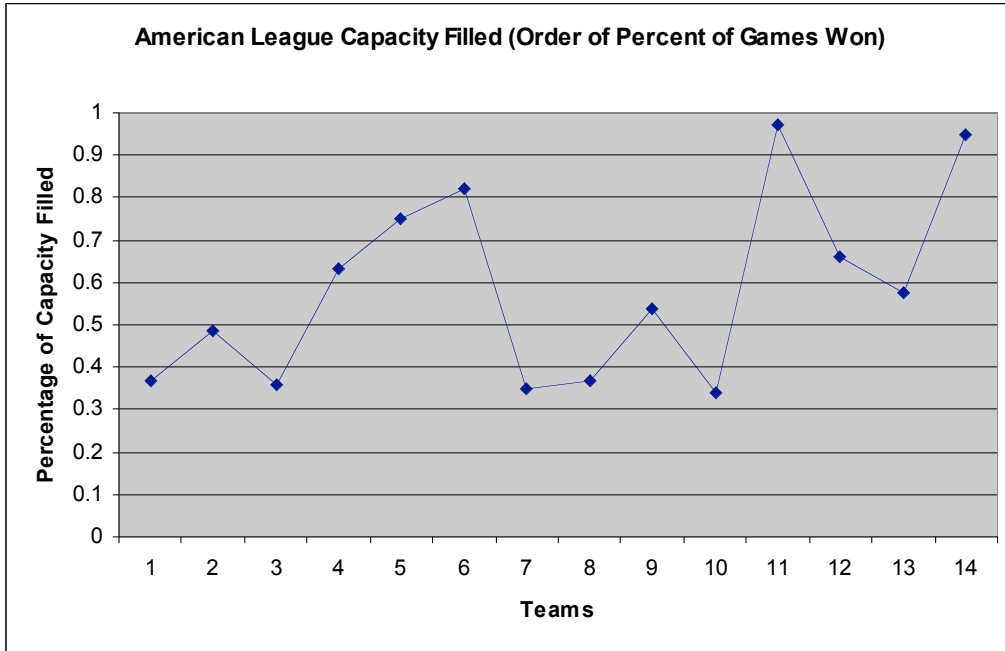
132

Figure 3.1

The general upward trend is noticeable. This subjective conclusion is evinced by using the same method used in the last section, taking the mean of the worst seven teams and comparing it to the overall mean and the mean of the best seven teams. The mean for the first seven teams is 0.5371, the mean for all seven teams is 0.5828, and the mean for the final seven teams is 0.6285. This helps support the conclusion that attendance (or, in this case, how much of the ballpark is filled) is correlated to games won.

Let us examine the National League data, and see whether or not the results are similar (see Figure 3.2).
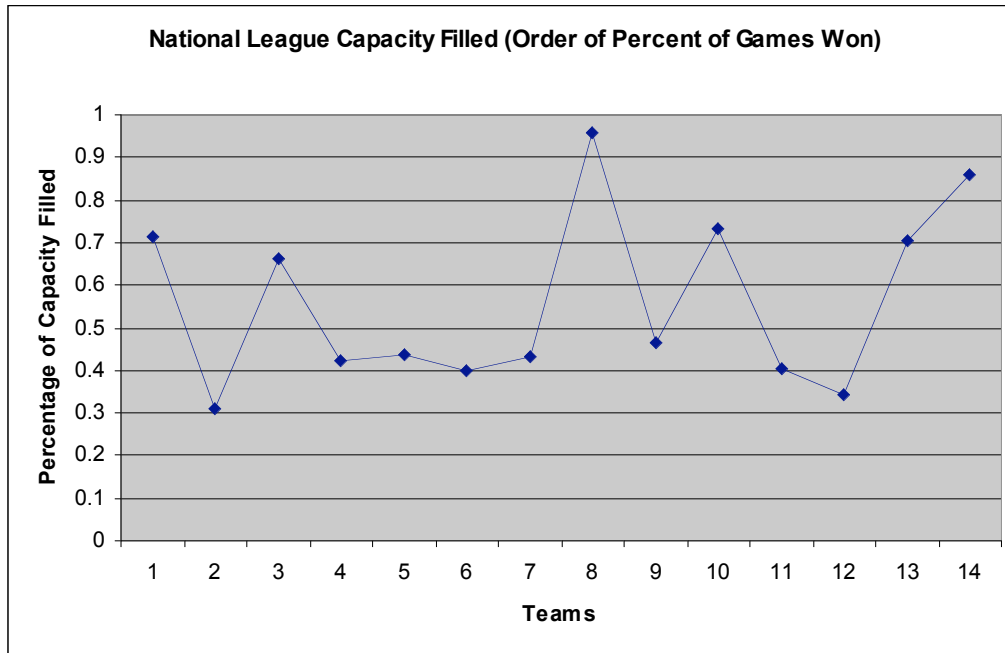
133

Figure 3.2

The trend is less obviously noticeable in this graph. However, if one follows the same procedure as when evaluating the American League data, one can see there is a general increase: the mean of the capacity filled of the lower seven teams' ballparks is 0.4818, the mean of all fourteen teams is 0.5599, and the mean of the upper seven teams is 0.6380.

Moreover, one can see that the second worst team (which, actually, had the same record as the worst team; therefore, the Phillies and the Cubs both were the worst teams) had the lowest capacity filled, and the best team had the second highest capacity filled.

The same explanation as for Section 2 can be used here. As a team does better, more people wish to see its games, so the percentage of capacity filled for ballparks of teams that do well will naturally be higher.

Another attribute that we can compare is the variation of the percentage of capacity filled. The variation was a little greater for the American League than for the National League. The standard deviation of the percentage of capacity filled for the American League ballparks is 0.04984, while the standard deviation for that of the National League ballparks is 0.04243. The mean absolute deviation for the American League was 0.1837, while the mean absolute deviation for the National League was 0.1812.
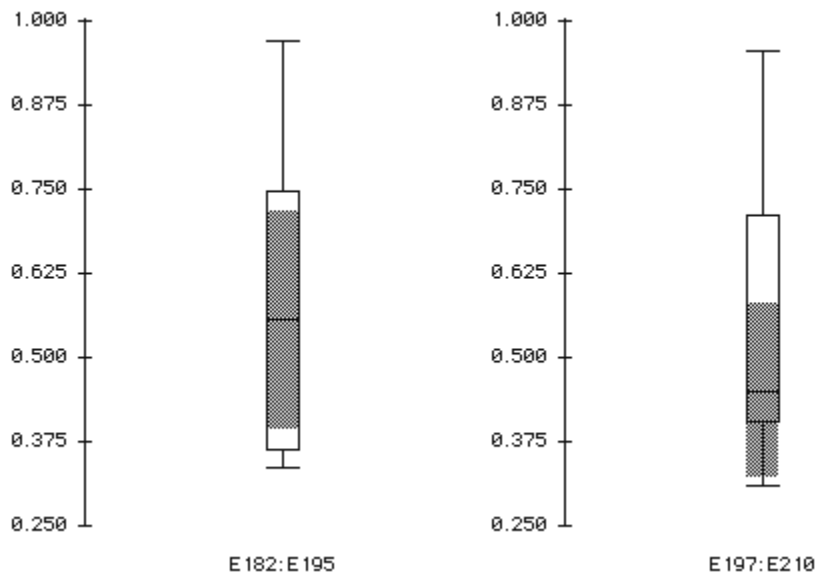
134

Figure 3.3

The small difference in variation is shown through the boxplots in Figure 3.3, as the focus of values for the National League was a little more concentrated together.

Section 4: **Attendance Percentage Comparisons**

Finally, we can divide the teams into four categories indicating how full the ballparks were throughout the season. The first category is for those teams whose ballparks were less than 40% full, the second category is for those between 40% and 59% full, the third is for those between 60% and 79% full, and the final one is for those at least 80% full. We will first examine the American League, then we will look at the National League, and finally we will look at all of the data. In order to distribute these numbers, we will use histograms (see Figure 4.1 through 4.3).
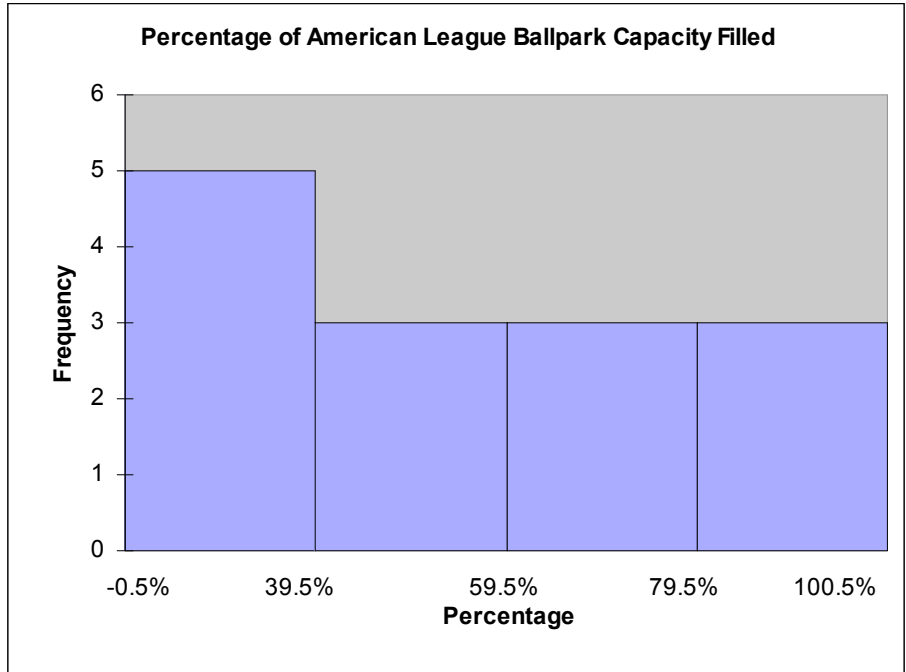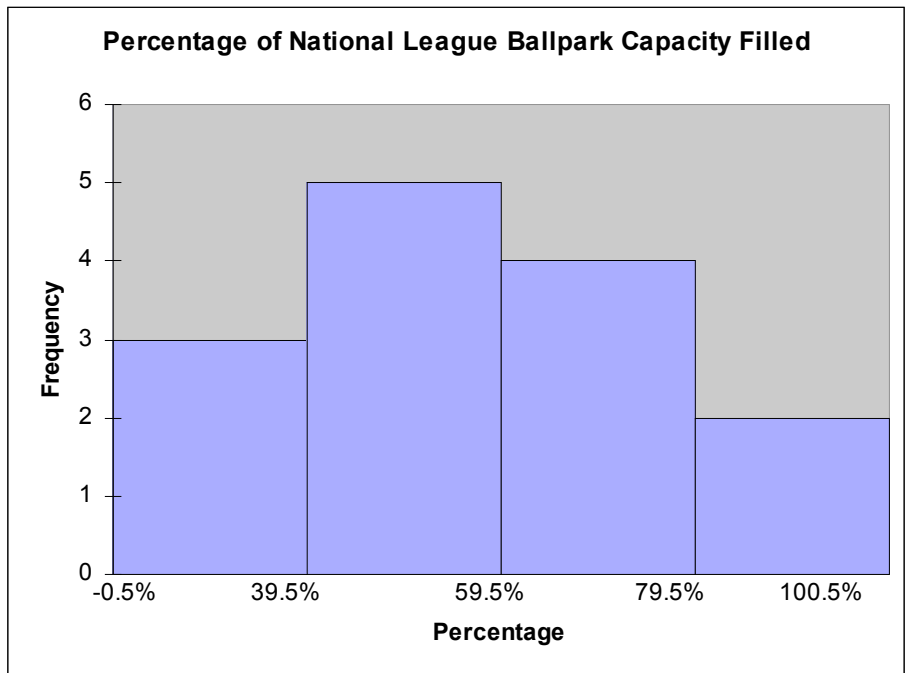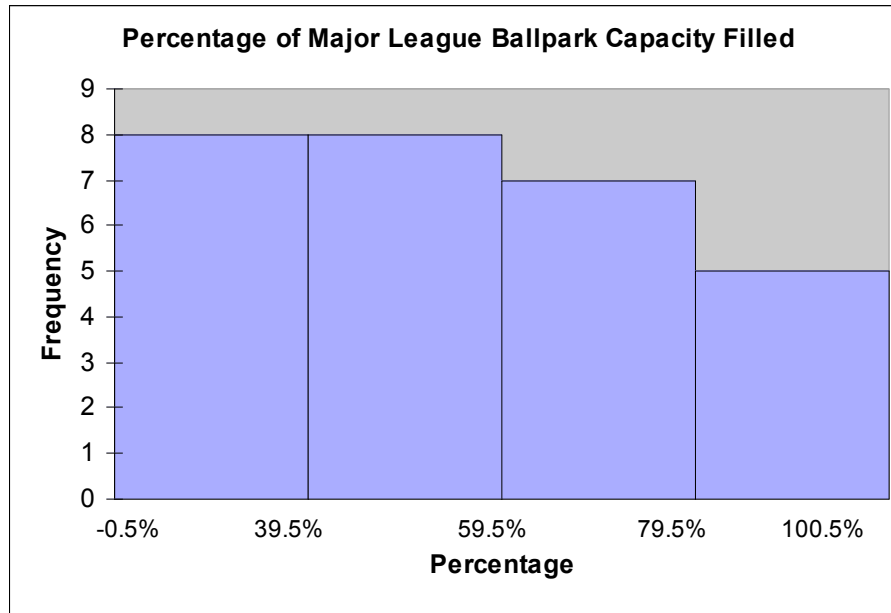
135

Figure 4.1

Figure 4.2

136

Figure 4.3

An interesting note about the first two histograms is that, while it appears that the attendance percentage of the National League ballparks were generally larger than that of the American League, the means determined (the mean for the American ballparks was 0.5828, while the mean for the National ones was 0.5599) that the American League ballparks usually had a generally larger attendance percentage.

Section 5: **Summary and Conclusions**

The overarching purpose of this project (besides simply analyzing ballpark data) was to determine the relationship between game attendance and team performance. The conclusion from Sections 2 and 3 (where we first analyzed game attendance and then the percentage of ballpark capacity filled and compared them to team performance) was that there was a correlation between game attendance and team performance.

One possible explanation for this apparent correlation that was mentioned in those sections was that the better a team does, the more individuals would want to watch the team play. This seems to be a plausible rationalization for part of the relationship.

However, one might also make the argument that the more people who watch a game, the better a team will play, as the pressure to do well or the encouragement from fans increase. This seems to be a much less plausible explanation than the other one presented.

137

We also examined the capacities of the ballparks (without respect to the game attendance) and compared them to their respective ages. There also seemed to be a correlation between these two attributes. The explanation that was given in Section 1 was that, since National League ballparks were generally built more recently than American League ones, their capacities would be larger (due to increased building efficiency, increased population, and increased interest in baseball).

The final aspect of the ballpark that we took a look at was placing the various levels of attendance percentage of capacity into different categories, so one can see the distribution of this attribute of Major League ballparks.