**I - Introduction**

In the US, Major League Baseball (MLB) teams have more home games (81) than the National Basketball Association (41) and the National Football League (8) combined. Result? Fans are much less likely to attend every game. There are so many options throughout the seasons. Given the current market situation, with so many kinds of new entertainment as a baseball game’s competitor, finding ways to sell tickets is even more important.

Because ticket sales are the main source of revenue for many teams, understanding which promotion can persuade fans to come is crucial to a team’s sustainability. What kinds of promotions will lead to the most increase in attendance at Major League Baseball (MLB) games? The answer is likely to be bobblehead giveaways. Since 1999, promotional departments of MLB teams have been using these little men to squeeze extra attendance as fans love it.

However, there’re so many factors, both on the field and off, affecting an individual’s decision whether or not to have an MLB game. In addition to promotions, fans also consider the timing, weather, the winning percentage of the games, and whether or not the game is a marquee game. Therefore, sometimes baseball marketers fail to understand the true effect of the bobblehead promotion on attendance. Our analysis aims to answer one central question: offering bobblehead giveaways can increase how much sales, by how much amount and how this result varies among MLB teams.

Understanding this can help us to make a sophisticated recommendation for promotional baseball department when it comes to bobblehead promotion.

Our report contains the following parts:

II — Data Characteristics

III — Model Selection & Interpretation

IV — Recommendation & Conclusion

V — Appendix

**II - Data Characteristics**

The dataset provides attendance information for all MLB teams for 2012 season. Each row illustrates attributes of one game, which can be divided into 4 buckets: Time (Month, Date, Day of Week, Day/Night), Weather (Temperature, Sky), teams (home\_team, opponent) and promotion (Cap, Shirt, Fireworks, Bobblehead). All of these variables can have more or less effect on the attendance of the game.

|  |  |  |
| --- | --- | --- |
| **Variables** | **Type** | **Meaning** |
| Month | Categorical | the month when the game takes place |
| Date | Integer | the date of the month when the game takes place |
| Attend  (independent var) | Integer | Number of attendance during that game |
| Home\_team | Categorical | 30 MLB teams (to name a few: New York Yankee, Boston Red Sock, |
| Day of week | categorical | the day of the week when the game takes place |
| Opponent | categorical | who is the component in the game |
| Temperature | Integer | temperature |
| Sky | categorical | whether the sky is cloudy, rainy, dome or |
| Day/Night | categorical | whether the game takes place at daytime or night time |
| Cap | categorical | whether the cap are offered to the fans during the game |
| Shirt | categorical | whether the shirts are offered to the fans during the game |
| Fireworks | categorical | whether firework is displayed at the beginning of the game |
| Bobblehead | categorical | whether the bobbleheads are offered to the fans during the games |

Table 1: Dependent and Independent variables of the dataset

What we really care about is the effect of bobblehead on attendance. We want to know how offering bobblehead affected the attendance and whether the effect varied by teams or by the time when it was offered, other conditions remain equal.

This is the panel data, which includes both cross-sectional and time series data. In this case, the dataset provides observations of bobblehead promotion at different times of the season and across different teams. Therefore, the effect of bobblehead might vary not only by time but also among teams. Let’s confirm it through our visualization.

**a)** **Home\_team on attendance**

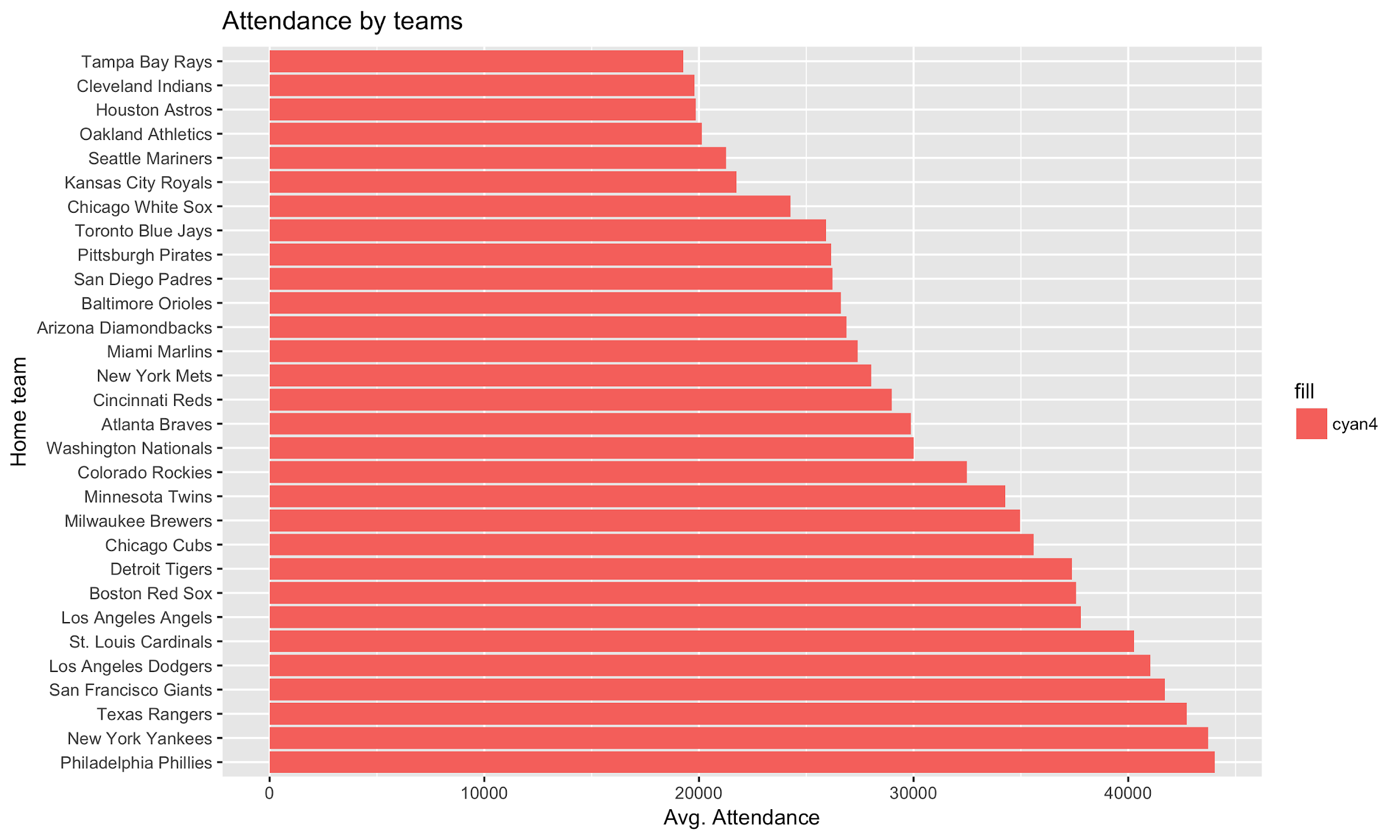


Figure 1: Avg. attendance by MLB teams

There’re rich teams and not-so-rich teams in MLB. There’s a difference between the popularity between the big-market teams like New York Yankees, LA Dodgers, Philadelphia Phillies, San Francisco Giant, etc. and the small-market teams like Tampa Bay Ray’s, Oakland Athletics, Seattle Mariners, etc. Those big-market teams usually have higher revenues, more budget, more star-players, thus attracting more attendance than small-market teams. In this market, the brand name was a crucial factor in determining the amount of attendance of each game.

**b) Bobblehead on attendance**

* Some teams had a clear difference in term of attendance between the game with bobblehead and game without bobblehead such as San Francisco Giant, Philadelphia Phillies.
* Some teams didn’t experience clear effect or didn’t have enough data to have a clear effect such as Boston Red Sox, San Diego Padres, Colorado Rockies, Minnesota Twins.
* The majority of teams saw an increase in the attendance: Houston Astros, Seattle Mariners, Cincinnati Reds, Toronto Blue Jay etc.

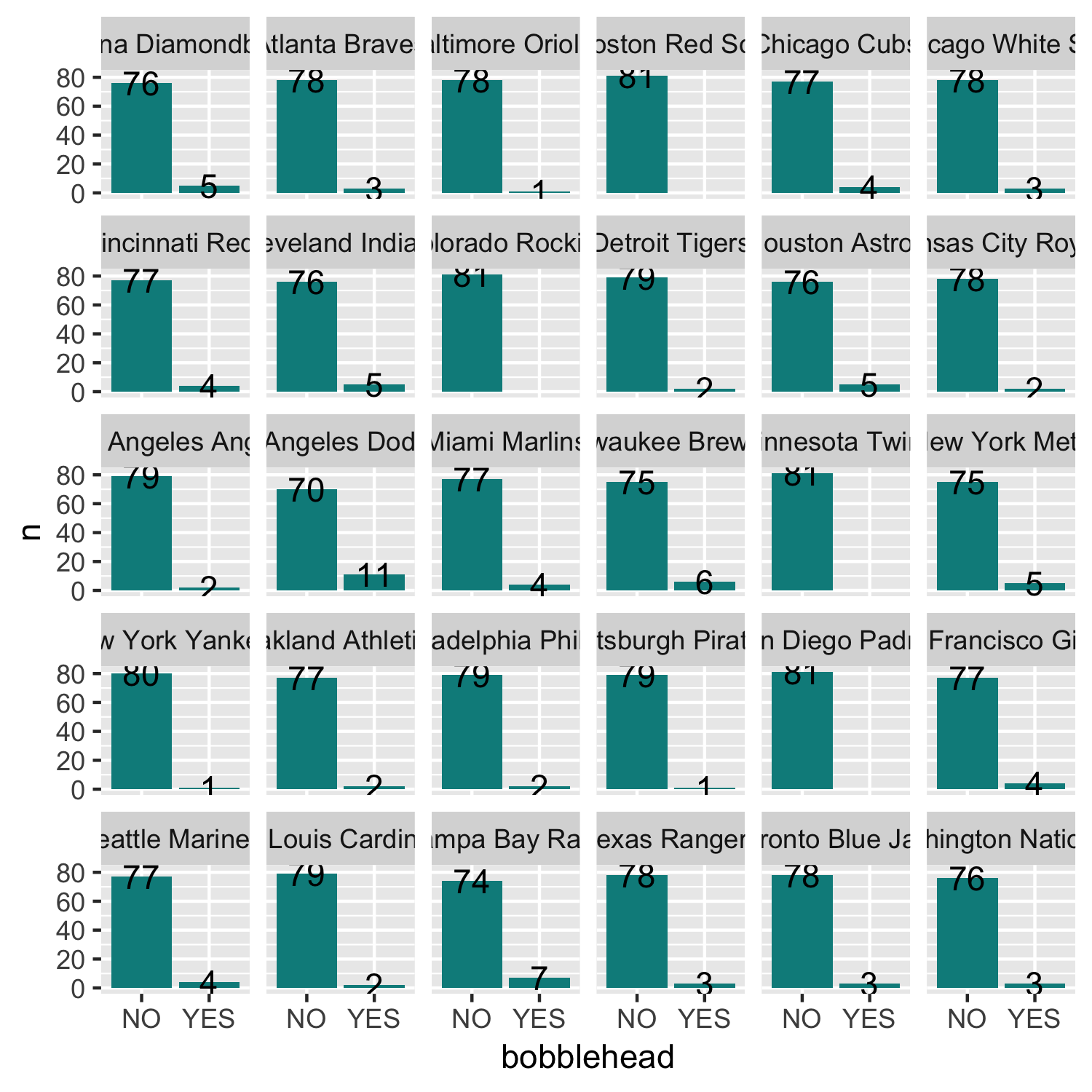
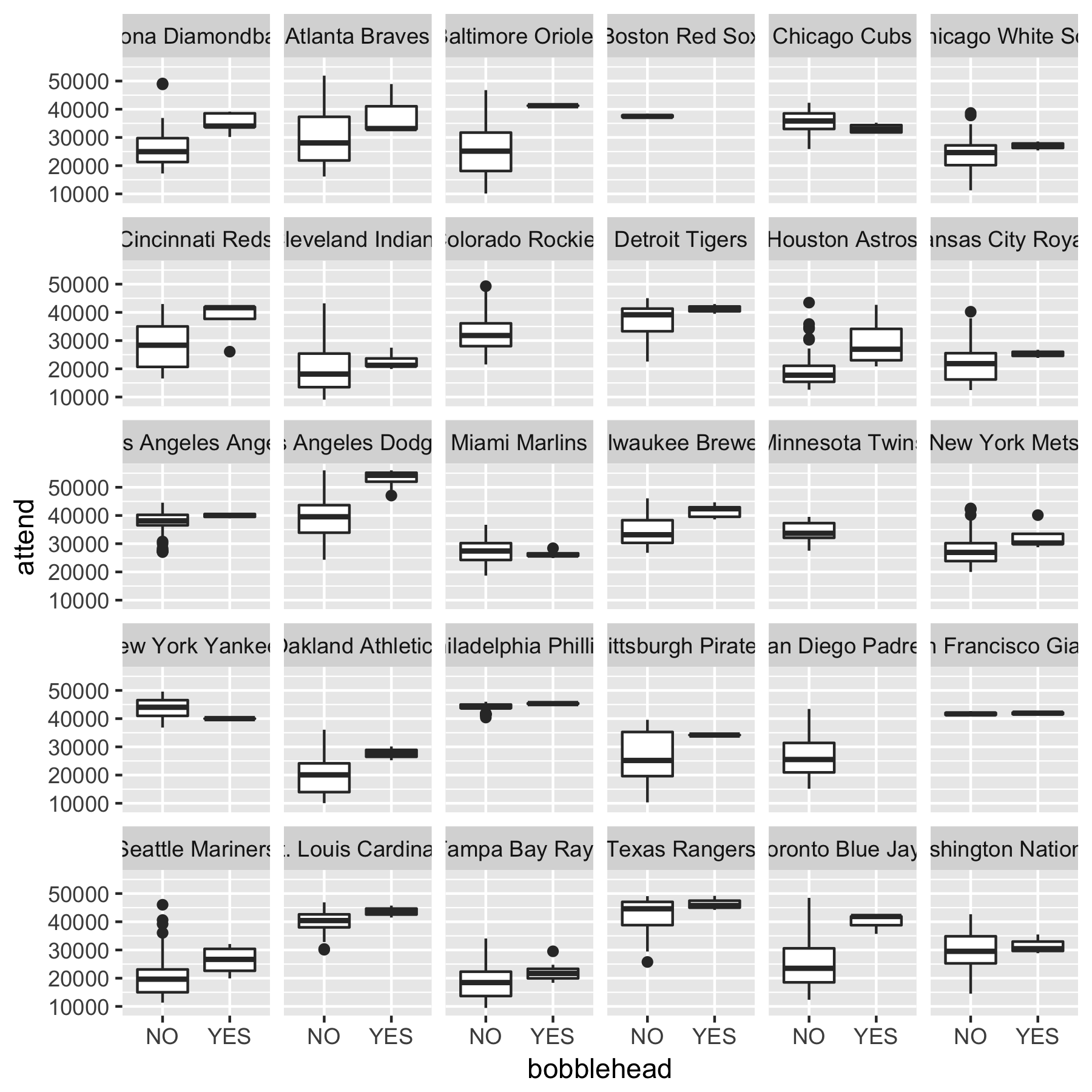


Figure 2: Bobblehead on attendance

However, was the increase the consequence of the bobblehead promotion? Not really. From the following figure, you can see that a significant portion of bobblehead was offered on Saturday and Sunday, which were usually the most crowded days of the week. It’s possible that people will go to the games because on the weekend even without the bobblehead. So we’re motivated to determine the true increase in attendance as the consequence of bobblehead. Also, was there any difference in term of an increase in attendance by bobblehead in weekdays and weekend?

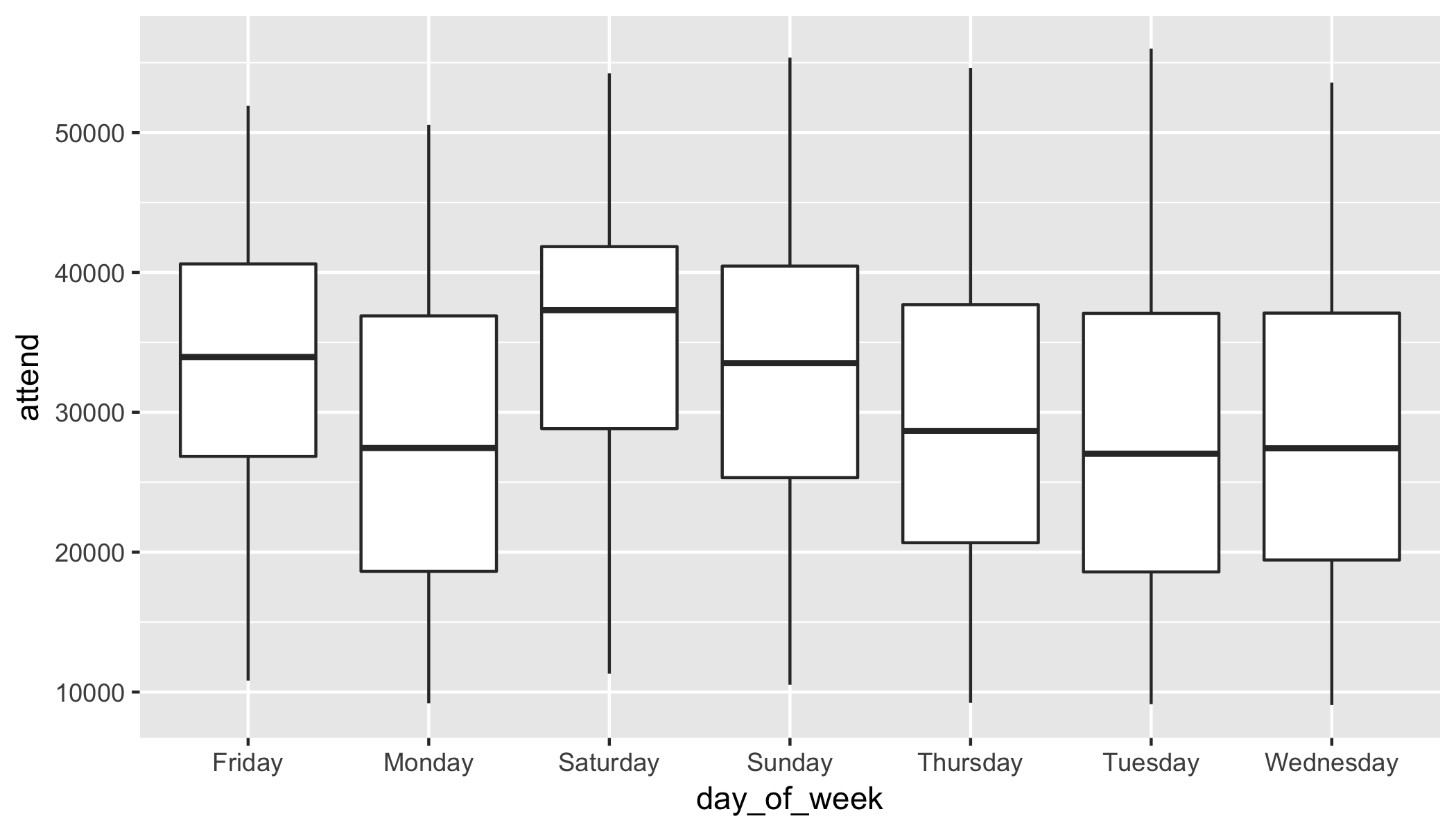
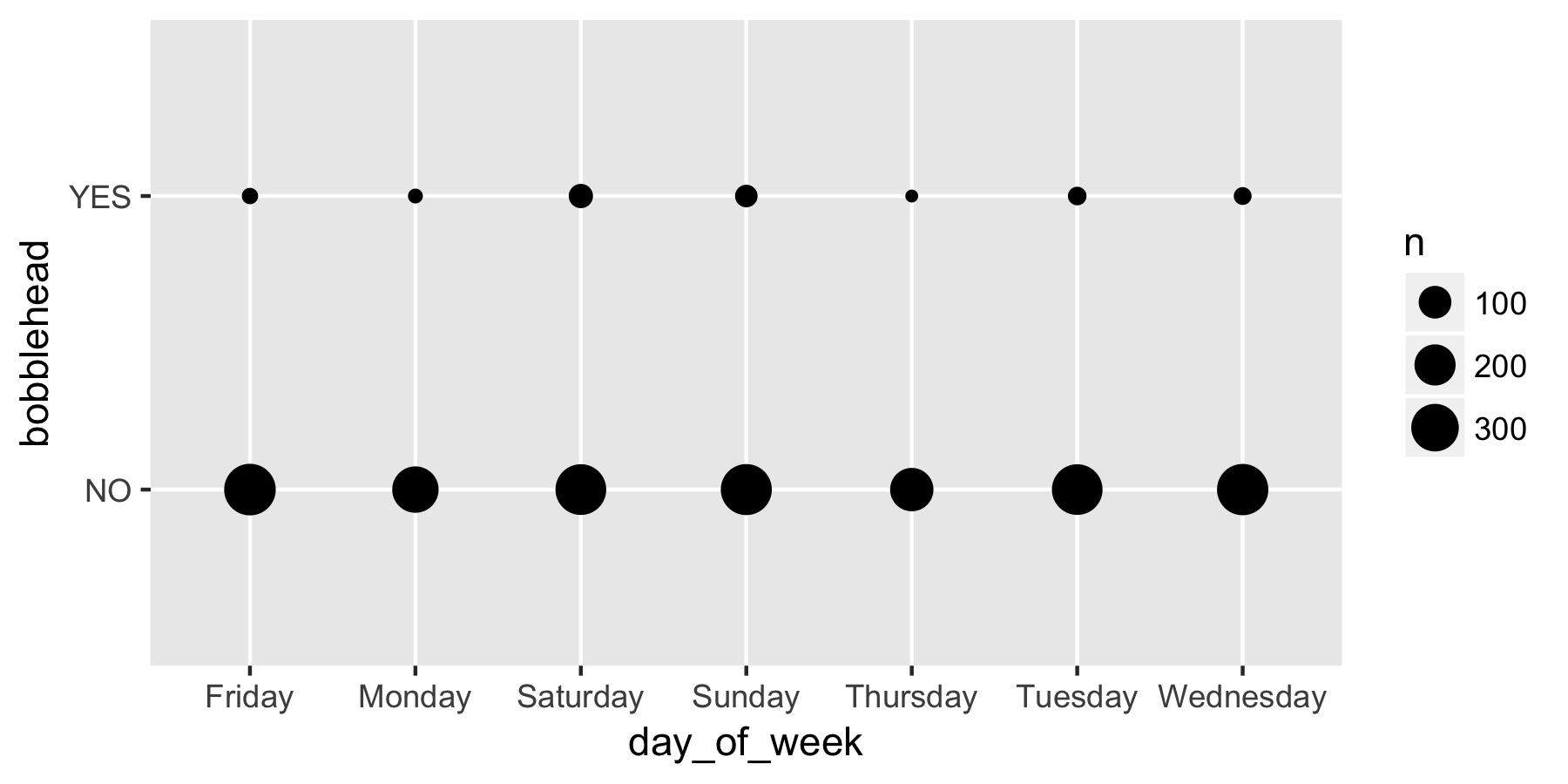
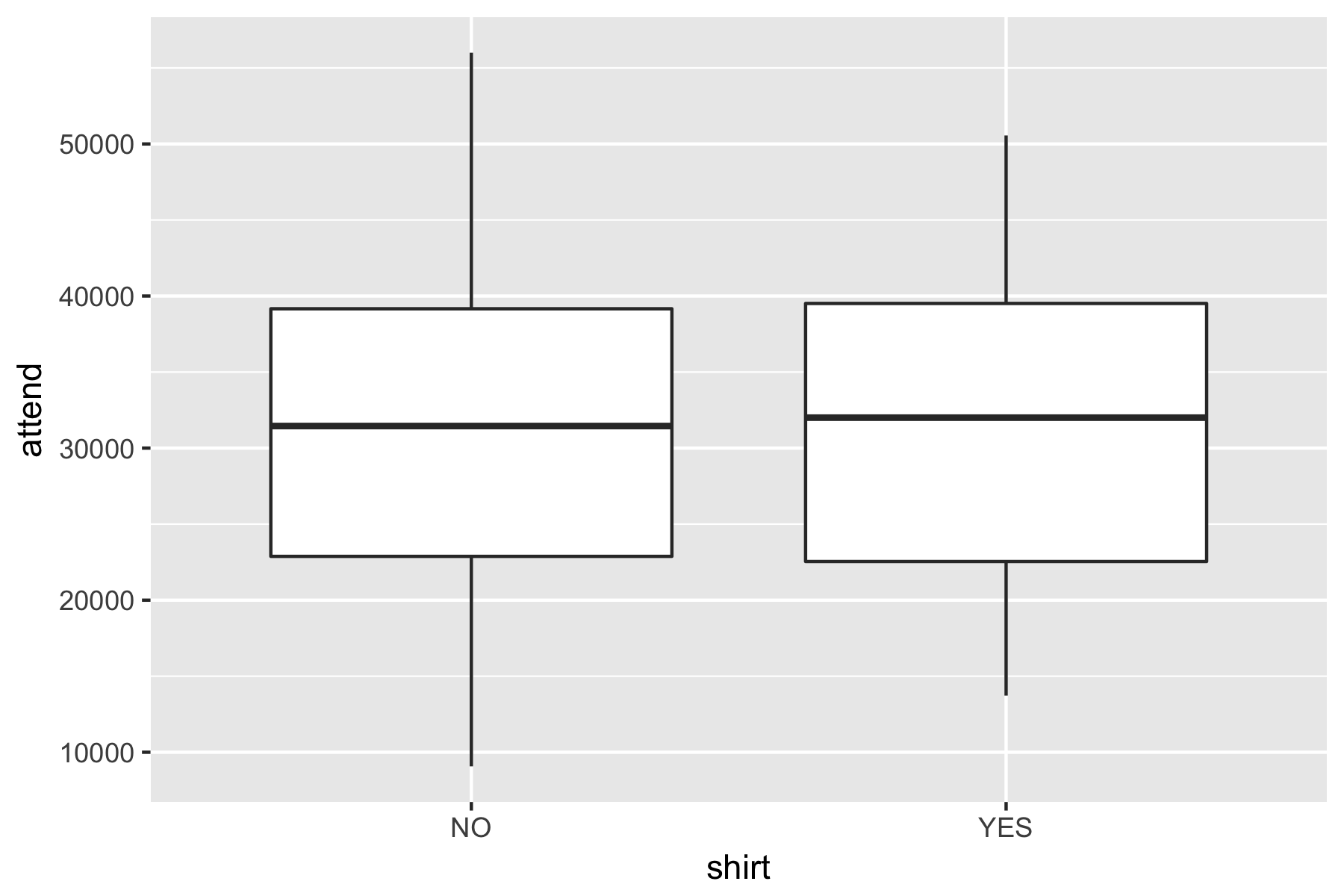
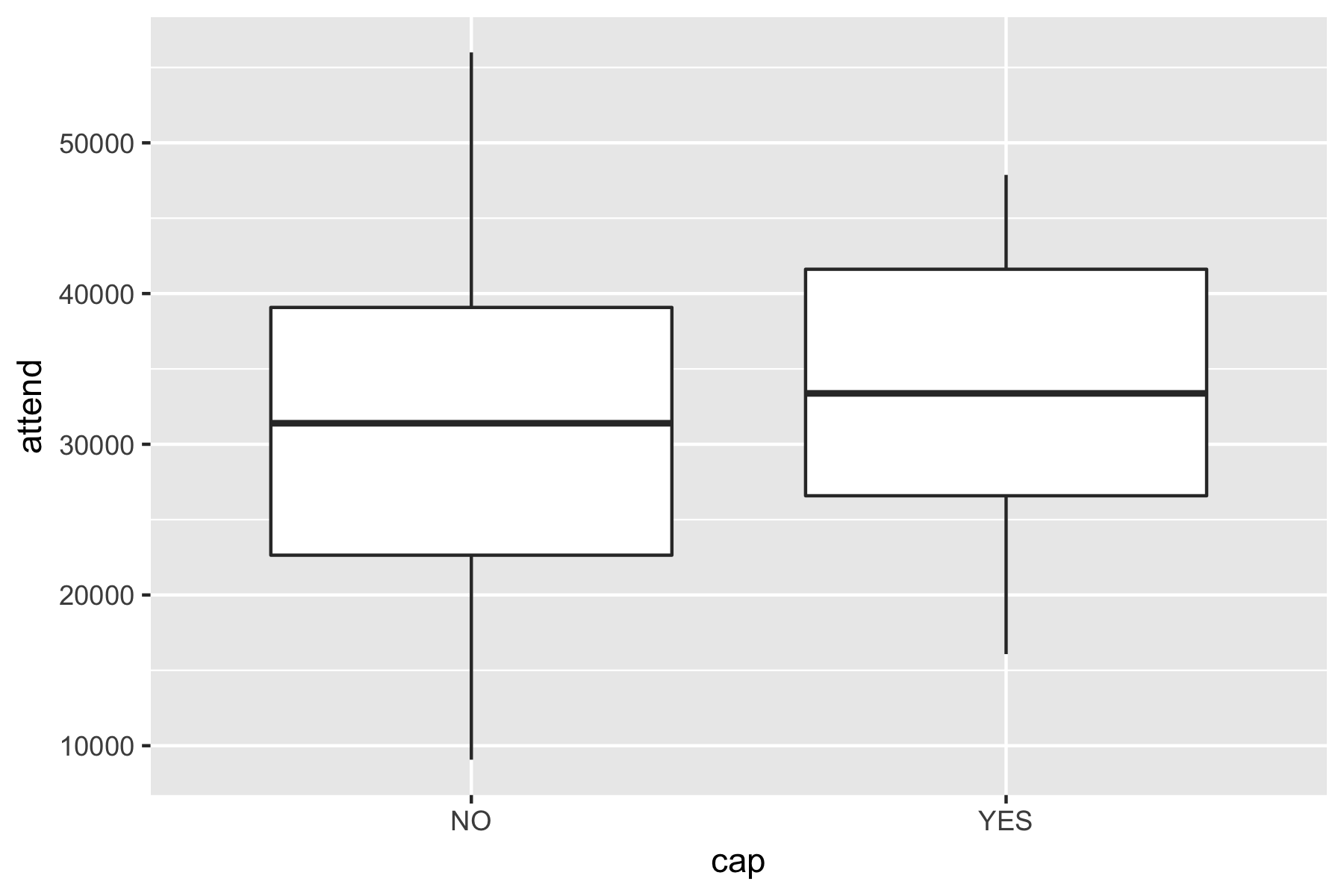


Figure 3: Bobble offering and attendance by days of week

**c) Other variables on attendance**

Shirt, cap, firework (on the same graph): figure 4 shows that promotion of cap and shirt had slightly effect on the attendance. Fireworks seemed to have the biggest effect on attendance among these three. On average, games that have fireworks had 34,400 attendance while games without fireworks only attracted 30,600 fans.



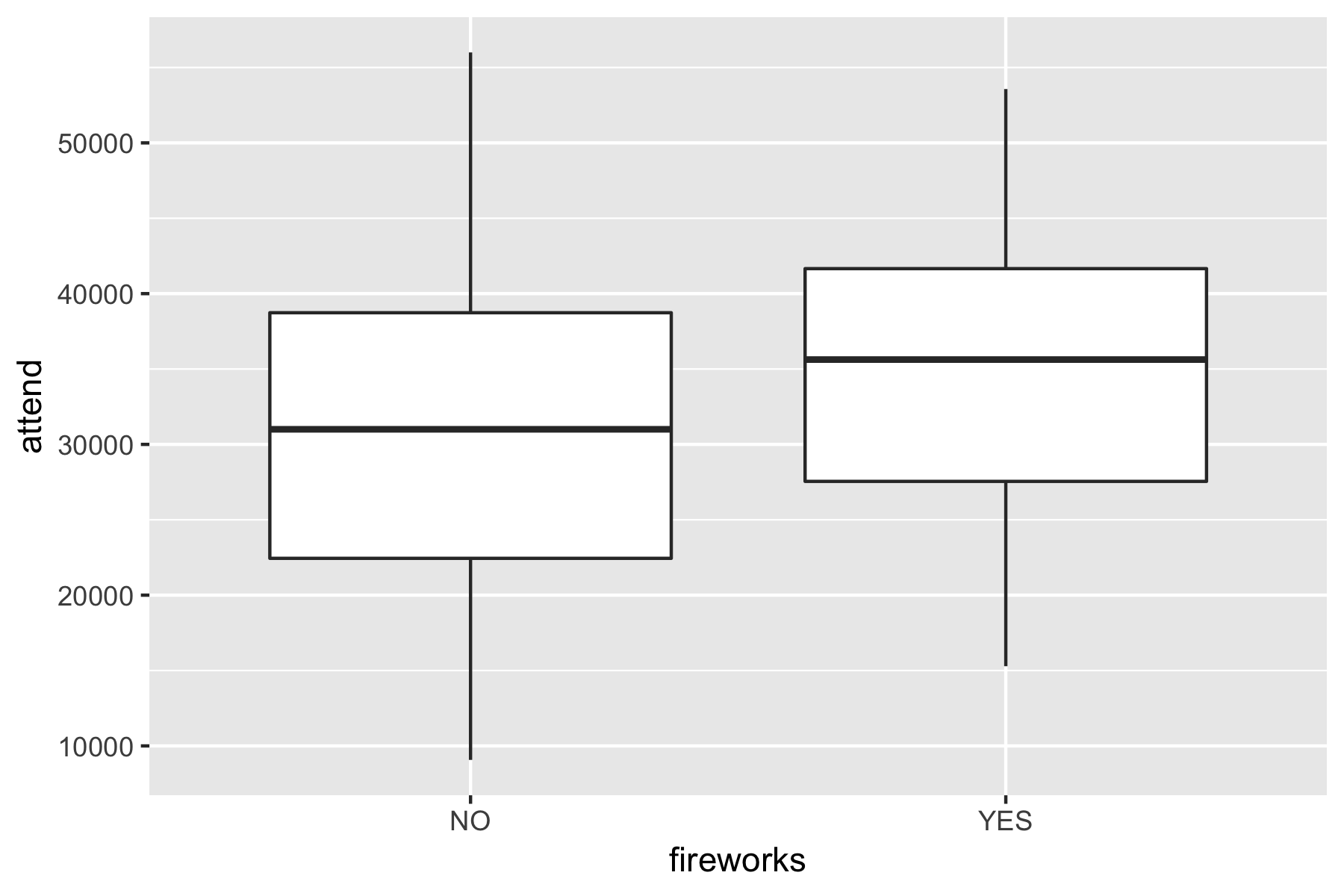


Figure 4: Variation in attendance by Cap, Shirt and Fireworks

**d) Seasonal effect**

Except for some teams (Boston Red Sox, San Francisco Giant, Philadelphia Phillies, Milwaukee Brewers, St. Louis Cardinal) where the attendance was stable across months, other teams seemed to have some seasonal effect with the highest attendance on June and July and going down in April, August, and September.

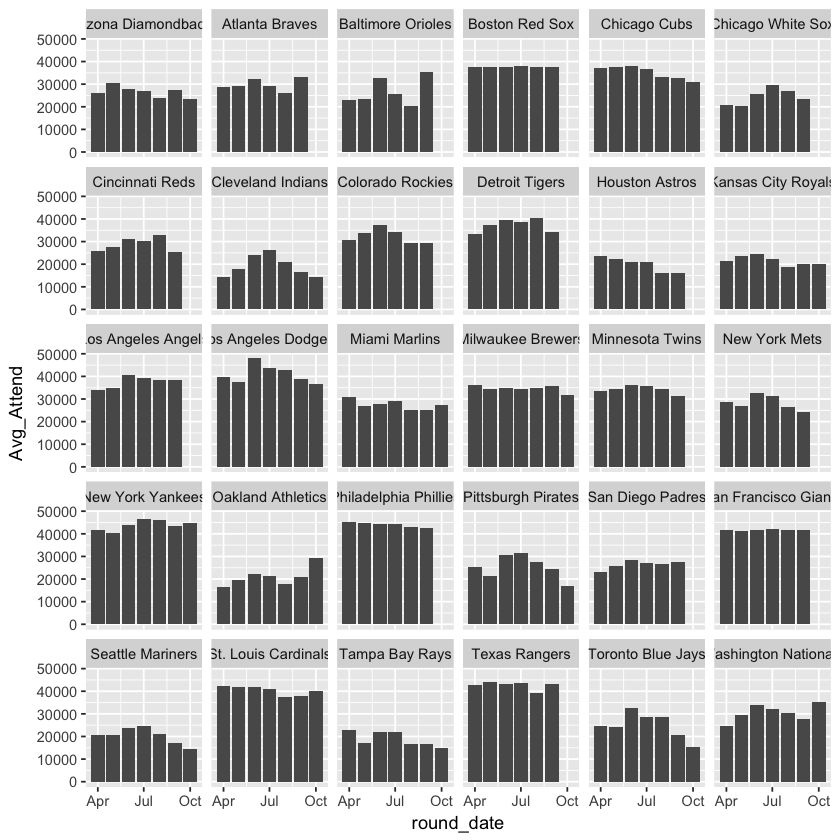


Figure 5: monthly average attendance by teams

**e) Weather (temperature, sky) on attendance**

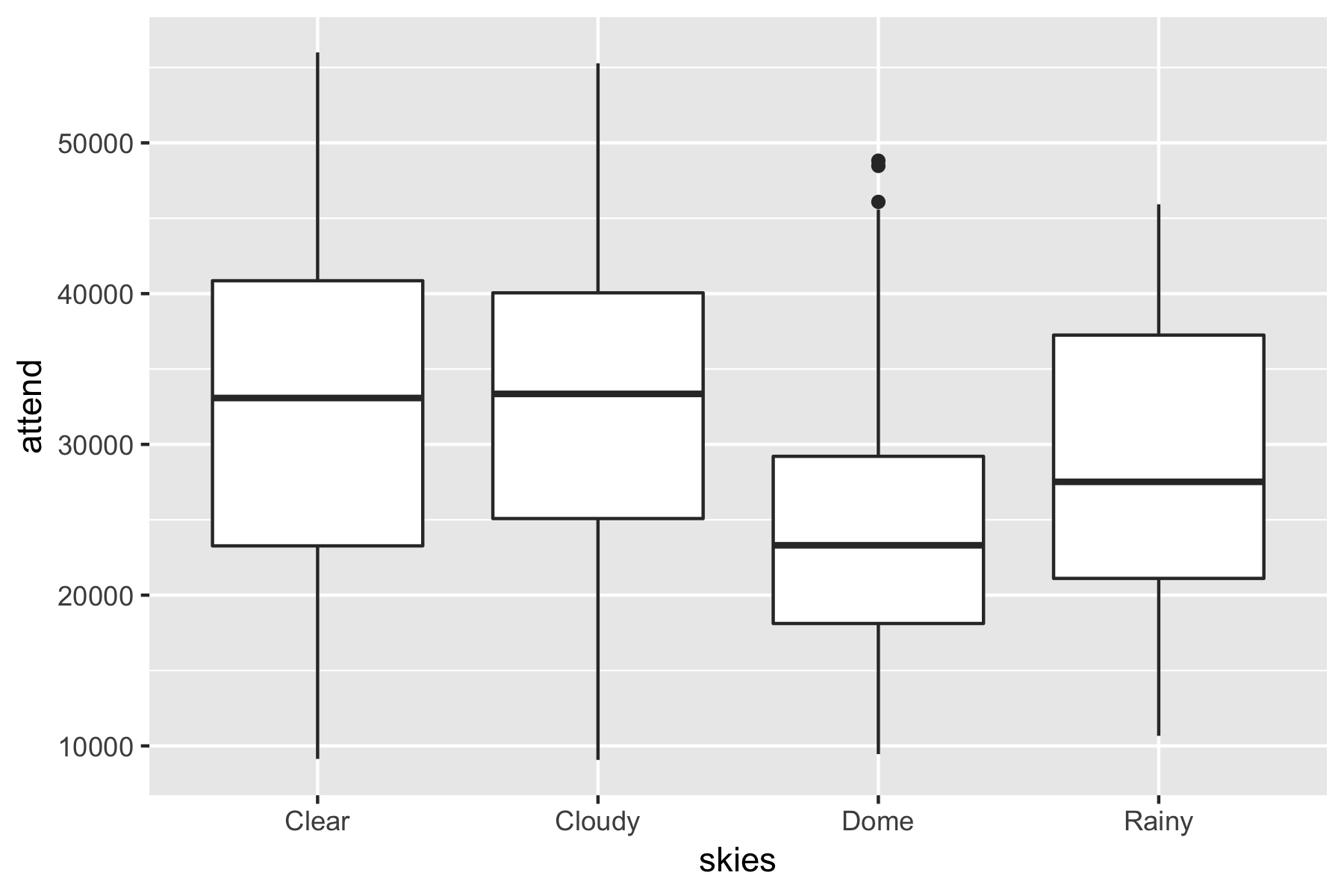
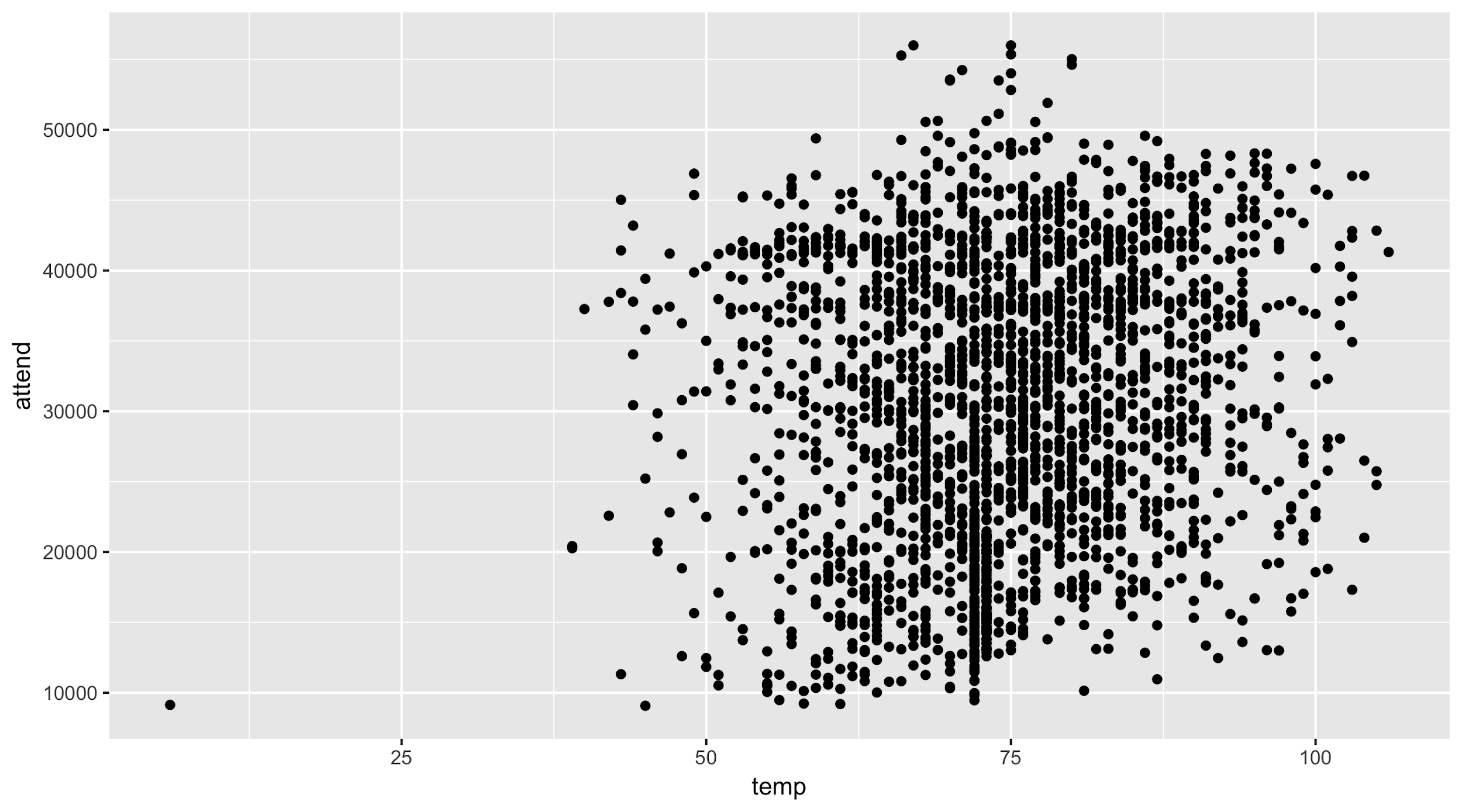


Figure 6: The relationship between temperature and sky conditions with the game attendance

From the graph above, we don’t see a clear relationship between temperature and game attendance. Also, sky condition seemed to be a significant factor. Fans tend to go to the ballparks on the days when there was a cloudy or clear sky. They don’t want to go out on rainy days or when there was dome.

**f) Inference**

This visualization provides us the hint on what should be included in our model as the independent variables. For most of the variables, we see a correlation between them and attendance. Therefore, we will incorporate home\_team, month, weekday, weather (except temperature) in our modeling step.

**III — Model Selection**

1. **Aggregation**

An aggregation approach helps us find out the average effect of bobblehead on attendance across team and time. Using the linear regression model, we examine the relationship between whether a bobblehead promotion was offered in a game and the attendance for that game. We have the coefficient of bobblehead is 4,494, meaning that offering bobblehead would increase the attendance of a baseball game by 4,494, on average.

The problem of the aggregation approach is that it fails to demonstrate the heterogeneity between MLB teams regarding the impact of bobblehead promotion. In the next sections, we will use 2 other approaches that help us identify this variation.

**2. Grouping**

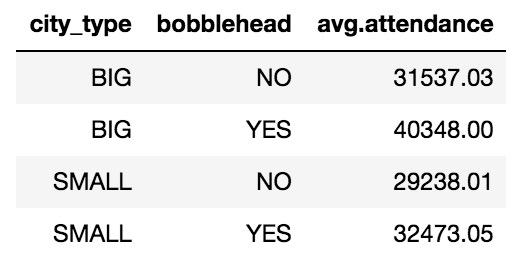
There are 2 central questions of grouping approach:

* Was the timing of bobblehead promotion an important factor? As mentioned above, a significant portion of bobblehead was offered on the weekend. Was there any difference in effect on attendance if the bobblehead is offered on the weekday?
* Was the size of the city an important factor? We all know that MLB attendance varied by cities since each city has a different living standard, fan base and the number of substitutes. All these heterogeneities will affect the way bobblehead impacts attendance. To simplify, we only based on city population as the first heterogeneity factor since it indicates how big fan base is.

To answer the first question, we split our dataset into 2 groups: games that happen on weekdays and games that happen on the weekend. After that, we apply linear regression for each subset of the data. The results show that there was a difference between the effect of bobblehead offered on weekend and the effect of bobblehead offered on the weekday. On average, bobblehead on weekday lifted game attendance by 6,628 (24.7% on average) while bobblehead on weekend increased 2,938 attendance only (8.9% on average). This can be explained by the fact that on weekend, many fans go to watch baseball no matter whether there are bobblehead giveaways; while on weekdays, they need more incentive to attend the game.

Regarding the second question, we acquired a new dataset about city population and based on that, we split the dataset into 2 groups: 1 group with the home team comes from small city (meaning that the population is less than 1 million) and the other group comes from big city (meaning that the population is bigger than 1 million). There are 9 teams basing in big cities and 21 teams in small cities.

Again, linear regression result shows that there was a difference between the effect of bobblehead for teams from the big city and the effect of bobblehead for teams from small cities. On average, bobblehead in the big city increases attendance by 7,638 (24.2%), while bobblehead in the small city increases attendance by 2,485 only (8.5%). Though differing in town population, the average attendance might not vary that much between the teams when there is no bobblehead (31.5K vs 29.2K on average). When bobblehead is offered, attendance in big cities was improved much more significant than in small cities. It might be the case that big cities have a bigger fan base and there are more incremental fans attracted to the bobblehead giveaway than in small cities.



Using the grouping approach give us more insights about the diversity in baseball when it comes to bobblehead promotion. However, there’s still some level of aggregation in this approach, which might clear out helpful information that we can offer to baseball marketers. We want to find an approach that can demonstrate the heterogeneity at the team level, and multi-level model is our next attempt.

**3. Multi-level model**

While the aggregate model provides a fixed intercept and fixed slope of bobblehead across all games in the season, the multi-level model assume that the effect will vary across teams, as we treat it as the first level in this case. There is still a fixed effect of slope and intercept, which are 3997 and 32004, respectively. This can be interpreted that there’s an average increase of 3997 in game attendance by teams. (Note that this fixed slope is slightly different from the slope in aggregate approach (4494) because the slope in aggregate approach is the average increase by games).

The varying slopes and intercepts as the result of the multi-level model are shown via the following graph.

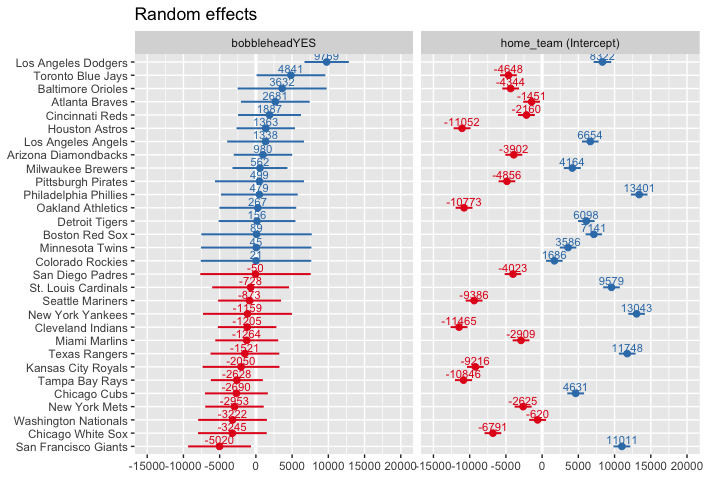


Figure 7: Random effects of bobblehead and intercepts among teams

Here, the random effect of intercept demonstrates the different level of attendance among MLB teams. Also, the random effects of bobblehead are the deviation of each team from the fixed effect. Therefore, we can come up with the final average effect for each team by adding the fixed effect to the random effects. Here is our final result:

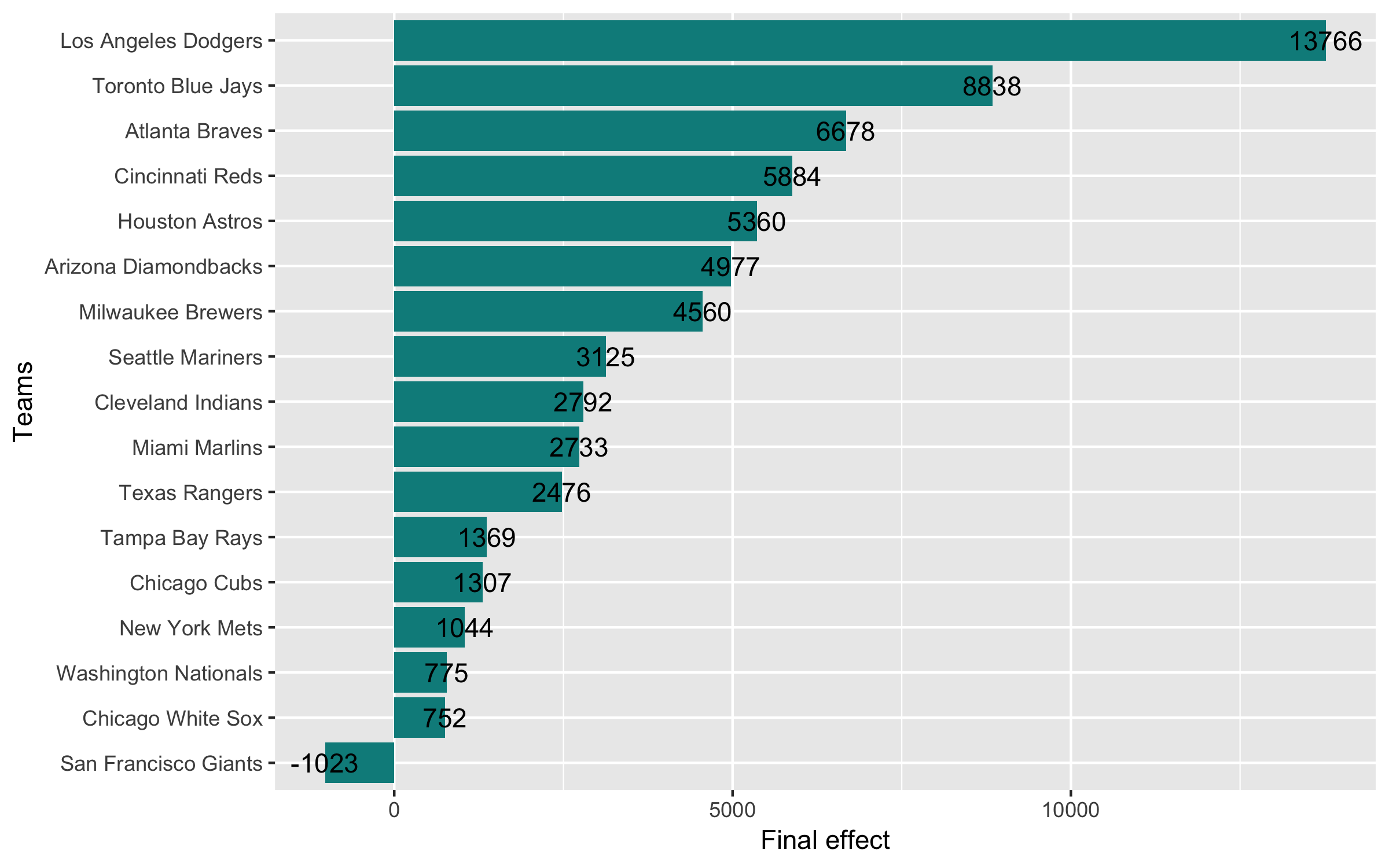


Figure 8: The verying final average effects of bobblehead on MLT teams

As we can see from the graph above:

● LA Dodgers enjoyed the biggest positive effect of bobblehead on attendance (on average offering bobblehead increased an additional 13,766 attendance), followed by Toronto Blue Jay (~8,838), and Atlanta Braves (6,678).

● On the other hand, San Francisco Giants was the only team to bear the negative effect of bobblehead on attendance (on average, offering bobblehead decrease 1023 attendance to their games). Also, Chicago White Sox, Washington Nationals, New York Mets and Chicago Cubs were the teams that see a slight increase in ticket sold with a bobblehead giveaway.

● 13/30 teams such as San Diego Padres, Colorado Rockies, Minnesota Twins, Boston Red Sox, Detroit Tigers didn’t see a clear effect of a bobblehead (with the random effect ~ 0) because we don’t have enough observation with bobblehead (the minimum is 3 observations) from these teams.

● The random effects in intercepts simply illustrate the discrepancy in the average level of attendance for each team.

**IV — Recommendation & Conclusion**

Bobbleheadpromotion was proved to be effective in increasing baseball attendance. Promotional department of MLB teams can expect to increase the attendance by an average of 4428 per game. However, since there is a huge heterogeneity between teams due to many factors, this effect of bobblehead also varied all the way from about 13,700 down to a negative number among teams. Los Angeles Dodgers was the team seeing the biggest impact of bobblehead while San Francisco Giant slightly suffered from the promotion.

The analysis also gives us some insights about when and where a bobblehead giveaway campaign should be implemented. While the majority of games that have bobblehead were on the weekend, our analysis shows that the impact of bobblehead on the weekend was inferior to on a weekday. Bobblehead on weekend increased the attendance by 8.9% while the number for a weekday was 24.2%. Also, teams in big cities should strongly consider offering bobblehead promotion more than teams from small cities because they have a much bigger fan base and therefore they can attract more fan to go to the ballparks with the bobblehead.

For further research, we suggest that a bigger data sample with more observation of bobblehead promotion implemented should be used to have a more significant result. In the above analysis, there 13/30 teams we cannot infer the final effect of bobblehead since we only have 0–2 observations with bobblehead offered for each team. We suggest that at least 3 observations should be gathered.

**V – Appendix**

**# Aggregation approach**

>agg.model <- lm(attend ~ home\_team + month + day\_of\_week + temp + skies + day\_night + cap + shirt + fireworks + bobblehead, data)

>summary(agg.model)

**#Grouping approach**

**#Timing**

>data$weekend <- ifelse(data$day\_of\_week == "Friday", "YES", ifelse(data$day\_of\_week == "Saturday", "YES", ifelse(data$day\_of\_week == "Sunday", "YES", "NO")))

>data.weekend <- data %>% filter(weekend == "YES")

>data.weekday <- data%>% filter(weekend == "NO")

>weekend\_BH <- lm(attend ~ home\_team + month + day\_of\_week + temp + skies + day\_night + cap + shirt + fireworks + bobblehead, data.weekend)

>weekday\_BH <- lm(attend ~ home\_team + month + day\_of\_week + temp + skies + day\_night + cap + shirt + fireworks + bobblehead, data.weekday)

>summary(weekday\_BH)

>summary(weekend\_BH)

**#City**

>city <- read\_excel("/Users/a1/Desktop/MBA study/Courses/Winter 2019/Predictive analytic project/Bobblehead/MDS\_Chapter\_8/city.xlsx", sheet =1)

>city2 <- read\_excel("/Users/a1/Desktop/MBA study/Courses/Winter 2019/Predictive analytic project/Bobblehead/MDS\_Chapter\_8/city.xlsx", sheet =2)

>population = merge(city, city2, by.x = "City", by.y = "Name", all.x = TRUE, )

>population$city\_type <- ifelse(population$POP2019 > 1000000, "BIG", "SMALL")

>population <- population[-c(3:8)]

**#Joining 2 datasets**

>data <- merge(data, population, by.x = "home\_team", by.y = "Teams", all.x = TRUE)

>data\_big <- data %>% filter(city\_type == "BIG")

>data\_small <- data %>% filter(city\_type == "SMALL")

>BH\_big <- lm(attend ~ home\_team + month + day\_of\_week + temp + skies + day\_night + cap + shirt + fireworks + bobblehead, data\_big)

>BH\_small <- lm(attend ~ home\_team + month + day\_of\_week + temp + skies + day\_night + cap + shirt + fireworks + bobblehead, data\_small)

>summary(BH\_big)

>summary(BH\_small)

**#Multi-level model**

>ml.model <- lmer(attend ~ month + day\_of\_week + temp + skies + day\_night + cap + shirt + fireworks + bobblehead + (1 + bobblehead | home\_team), data = data)

>summary(ml.model)

**#Visualizing the multi-level model**

>gg <- plot\_model(ml.model, type = "re", sort.est = "bobbleheadYES", show.intercept = TRUE,

free.scale = FALSE, show.values= TRUE,value.size =3, digits =0, value.offset = 0.5, dot.size = 2)

>bobblehead\_ranef <- as.data.frame(ranef(ml.model), stringsAsFactors = default.stringsAsFactors()) %>% filter(term == "bobbleheadYES")

> bobblehead\_ranef %>% ggplot(aes(x = reorder(Team,Final\_avg\_effect), y = Final\_avg\_effect)) + geom\_col(fill = "cyan4") + coord\_flip() + geom\_text(aes(label = round(Final\_avg\_effect, digits = 0)), position = "identity", digits =0) + labs(x = "Teams", y = "Final effect")