WEATHER AND BASEBALL

W200 - Project 2 - December 2019

Authored by: Alyssa Augsburger, Henry Bazakas, John Lee, Derek Topper

Introduction

Over the last few decades there has been a rise in academic study of the game of baseball. With the rise of data science, people are better able to understand the effect of different variables on the game. One such variable, that can have a significant impact on the sport, is the weather. As baseball is traditionally played outside, there tends to be a plethora of impacts on both the on and off field performance. Considering the effects on batting, pitching and attendance, certain weather conditions can have strong effects on how the players are able to perform on the field. For example, pitches can have lower spin rates and shorter breaks in colder weather, batters tend to score more runs in warmer weather and attendance tends to decline when it is raining. Ultimately, there are a number of effects on batting, pitching and the miscellaneous parts of the Major League Baseball game itself, when exposed to varying temperatures, wind speeds and precipitation effects.

For starters, baseball, also known as “America’s National Pastime” is a sport that has been played continuously in the United States since the mid 19th century. In the game’s modern version, two teams compete to score more points, known as runs, than the other team. The game consists of nine innings, which each consist of three outs where each team serves as the fielding team and as the batting team. On the fielding team, one player is required to throw a ball that one rotating player on the batting team must attempt to hit with a bat. If the batter is able to hit the ball into the field of play, then they are allowed to run the bases and if they are able to cross all four bases then they are awarded a run. Simply, the team that scores the most runs is declared the winner.

Research Question

Our main research question was to determine how weather affects baseball. We focused on three aspects of the game of baseball:

- Batting
- Pitching
- Miscellaneous (attendance, timing, and venue)

We also broke down weather into three aspects:

- Temperature
- Weather type (i.e. sunny, raining, cloudy)
- Wind speed

Data Cleaning

The dataset we used for this analysis contains detailed information about every pitch, at-bat, and game from the 2015-2018 regular seasons of Major League Baseball, the United States’ top baseball league.

For our analysis, we slimmed this information to exclude information from 2015-2017 and chose to focus solely on 2018. However, this means that this analysis could be easily broadened to include information from more seasons.

Each row of the data represents a single pitch, game or player, depending on the file. This data was initially obtained by Paul Schale, a Ph.D. candidate at the University of Oregon. Schale was able to use a web scraping technique to obtain four seasons of data from the website of Major League Baseball. The data is available at https://www.kaggle.com/pschale/mlb-pitch-data-20152018. The size of files range from 45 KB to 98 MB.

From this Kaggle site, we used the following datasets:

- atbats.csv (11 columns)
    - This lists general information about each at-bat that does not change during the at-bat.
    - This includes things like at-bat ID, IDs for each batter and pitcher, at-bat description, inning and how many outs there were.
- games.csv (17 columns)
    - This file contains general information about each game that does not change over the course of the game.
    - This includes attendance, which is the number of fans who attended the game, final scores, the teams in the game, the date of the game, the length of game, if there was a delay before the start of the game due to weather, the venue, the description of the wind and the weather.
- Pitches.csv (42 columns)
    - This contains the raw file with all of the pitches from 2015-2018 and their detailed information about each pitch.
    - Specifically, for example, this data includes information such as the pitch locations as they crossed the plate, pitch speeds, spin rate, break angle, break length, pitch type and many more characteristics about each pitch instance.

Our initial review of this data, by viewing with .head() and looking at each variable, acted as our sanity check and we definitely found a few rows of unexpected data such as games with attendance of 1, venues with 2 or less games, and a venue with two different names. These checks helped guide our initial data cleaning. As we continued our analysis, we would review our results and the visuals acted as continued sanity checks.

After verifying that our data made sense, in order to make our data easier to work with, we decided that we were only going to look at one year of data. However, since our data was compiled in multiple tables, we had to first join all of our tables together to make it possible to remove all of the pitches that were from 2015, 2016 or 2017. To do this, we first joined our table with our Pitches information onto our At-Bats table, using a unique identification number for each at-bat. We then joined this table onto our Games table, using a unique identifier for each game. We then used this table to filter out all of the non-2018 games.

While this would be the most important step to clean the data, we also did a number of other things to make the data work in an easy-to-palate format. For example, in order to look at this data, there were a few items we needed to clean and transform in the data. For starters, the variable that corresponded to each game's weather, was formatted as a string with a Fahrenheit temperature value, the word 'degrees', a comma and then a standardized weather description. In order to work with temperatures, weather types and precipitation, we needed to separate this into their columns. By splitting that column into two columns, we then possessed two new columns. One of those columns was called

“Temperature”, which only contained the recorded Fahrenheit temperature of the game and another called “Weather Type” that only contained the weather description, which was recorded as ‘partly cloudy’ or ‘sunny’, for example.Specifically, the “weather types” include dome, rain, overcast, cloudy, partly cloudy, clear, drizzle, sunny, and snow. In order to gather more data, these weather types are also binned into Precipitation and No Precipitation for a second analysis. Precipitation includes rain, drizzle, and snow. Any other weather types are under No Precipitation. Analysis of this “Weather Type” column alerted us to the fact that some of the data we had occurred indoors, and were labelled as either roof closed or dome, so we made sure to account for this when considering outdoor conditions.

The data contained a wind variable, which encapsulates both the wind speed and the direction of the wind. Similar to the temperature, we separated that larger wind variable, into 2 separate variables — one for wind speed and the other for wind direction. We also had to combine two runs columns, to make a column that could supply how many runs were scored by both teams throughout a game, as this was initially separated into home and away teams.

We also had to group and merge our dataframes, in order to depict certain trends. For example, we needed to group our data to obtain the number of home runs hit in a game, and had to merge this grouped column back into our larger pitch-level dataset so we could conduct further analyses. This process also had to be done to collect the number of errors in each game. We also had to clean various different things in our data like the venue names, as the name “Angel Stadium” and “Angel Stadium of Anaheim” referred to the same building and we needed to replace “Angel Stadium of Anaheim” with simply “Angel Stadium”, when we considered stadium names in our weather analyses.

It is worth noting, however, that our data is to be taken with a caveat, as we did notice that there was a decrease in sample size as we approached the extreme near-outlier values, for the weather variables, such as the fact that there was only one game in 2018 where it was snowing and were very few games with wind speeds of over 25 miles per hour. While we accounted for this, by grouping our data where possible, it is certainly worth noting and understanding the implications of this. For example, in some cases, we combined the temperatures into 20 degree wide bins, such as groups representing below 45 degrees, between 45 and 65 degrees, between 65 and 85 degrees and above 85 degrees. We also combined certain weather types, such as combining snow, rain and drizzling weather designations to create groups on whether it was raining or not. However, for most of our analysis we had a sufficient amount of data to consider the implications of the weather on each type of event.

Weather Versus Hitting

In order to investigate the impact of weather on hitting, we decided on three types of hitting data to look at - total runs scored, total home runs scored, and total errors (all for both teams in a game). We then looked at each of these factors and compared it to the weather factors - temperature, wind speed, and weather type.

How does weather affect total runs scored per game?

Temperature
In order to compare temperature and total runs per game, we first looked at all of the temperature and total runs per game data. The data set contains data for 2,431 games played in 2018 and for each game we have the temperature and the total number of runs scored in the game. Since this is a lot of data, creating both a scatter plot and bar chart of all of this data proved to be difficult to read and find any insight. In order to gain more insight, we created groupings for the temperature values. The

groupings were determined by looking at the minimum and maximum temperature values and the distribution of the temperature values. We assigned each game a temperature label based on the grouping that the temperature fell into. We created an additional column in the dataframe, “temp_label”, that contained this information. After grouping the temperature values we were then able to obtain the average total number of runs scored per game by temperature grouping.

Temperature Label	Average Total Runs Scored per Game
>65 and <=85 degrees	8.622785
>45 and <=65 degrees	8.820628
<=45 degrees	8.866667
>85 degrees	10.315152

The table of data above is sorted by the average total runs scored per game and shows that when the temperature is over 85 degrees the average total number of runs scored per game is greater than 10 while the average total number of runs scored per game in all other temperatures less than 85 degrees is in the 8-9 range. These results show that a game has a greater chance of having more total runs scored if the temperature is over 85 degrees.

Wind Speed

Similar to when we looked at temperature and total runs scored per game, when we first plotted wind speed versus average total number of runs scored per game, there was a lot of data and it was difficult to read the chart and gain anything insightful from the data. To mitigate this, we grouped the wind speeds, similarly to how we grouped the temperatures. We looked at the minimum and maximum wind speed values and the distribution and created 5 groupings. We assigned each game a wind grouping and created an additional column in the data titled “wind_label” that contained this data. After creating these groupings, we then looked at the average number of runs scored per game for each wind speed grouping.

Wind Label	Avg Total Runs Scored per Game
>15 and <=20 mph	8.629630
<= 5 mph	8.704819
>5 and <=10 mph	8.928279
>10 and <=15 mph	9.202105
>20 mph	10.133333

This also showed something very interesting. We saw that the average total runs scored per game was actually the highest when the wind speed was greater than 20 miles per hour. This was counterintuitive to what we would have predicted.

Weather Type

In order to compare weather type and total runs we first looked at the average total number of runs scored per game for each weather type. This did not show anything very interesting and we didn’t see many differences between the 9 weather types in the total runs scored per game. Additionally, it was difficult to obtain much insight from this as there are 9 different types of weather in the data and several types could be considered very similar. In order to try and obtain something more useful out of the data, we created 2 groupings for the weather types - precipitation and no precipitation. After creating these

groupings, we then looked at the average total number of runs scored in games with precipitation and games without precipitation.

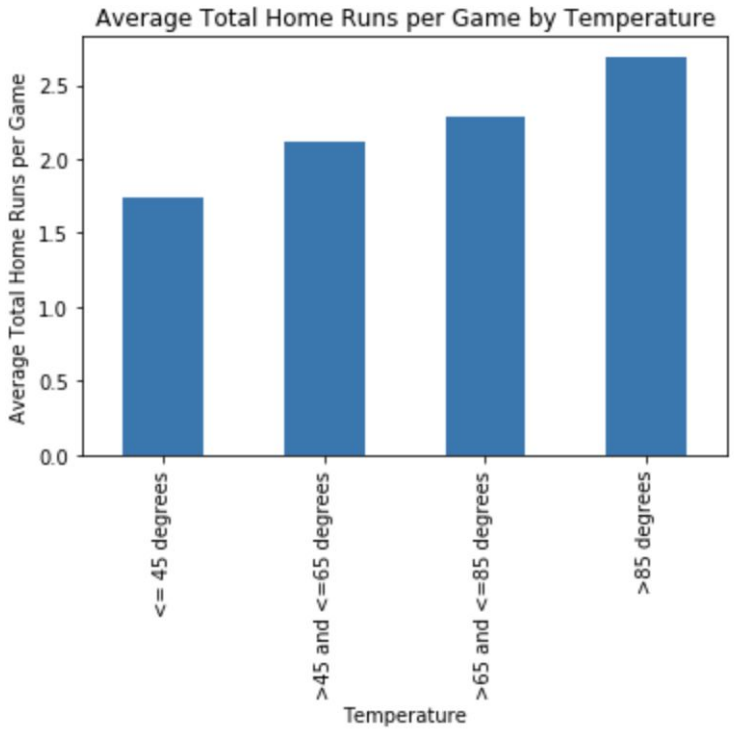
Weather Type Label	Average Total Runs Scored per Game
No precipitation	8.880903
Precipitation	9.868421

This shows that the average total number of runs scored per game is actually higher in games with precipitation as compared to games without precipitation. This is different than we expected and could potentially be due to our dataset having much fewer games that occurred with precipitation as compared to games without precipitation.

How does weather affect total home runs scored per game?

Temperature

After summing the number of home runs scored per game, we were able to work with the variable in order to look at how temperature affects the average total number of home runs scored per game. Using what we learned from looking at total runs scored per game, we looked at the temperature in groupings and calculated the average number of home runs scored per game for each temperature grouping.



This showed something very interesting. We see that as the temperature increases, so does the average total number of home runs scored per game. Air is less dense in hotter weather, so it makes sense that a ball would fly farther in the heat. Baseball teams should hope to play more games in hotter weather, or dome stadiums may want to consider increasing the temperature in order to increase the possibility of players scoring more home runs.

Wind Speed

We then looked at the final weather component, wind speed, to see how it affects the average total number of home runs scored per game. Looking at the wind speed in the same groupings as we created for analyzing total runs did not show anything very interesting as compared to the average total number of home runs scored per game. It appears that wind speed does not have much of an impact on home runs.

Weather Type
After briefly looking at weather type compared to the average total number of home runs scored per game, we did not see much of a difference among the 9 different weather types. We then looked at the weather type in the groupings of precipitation and no precipitation and calculated the average number of home runs scored per game for each weather type group.

Weather Type Label	Average Total Home Runs Scored per Game
No precipitation	2.295863
Precipitation	2.394737

This showed a very minimal difference in the average number of home runs scored per game in games with versus without precipitation. This was interesting because it was a much smaller difference than when we looked a total runs scored per game in games with and without precipitation. It appears that precipitation does not have an impact on scoring home runs.

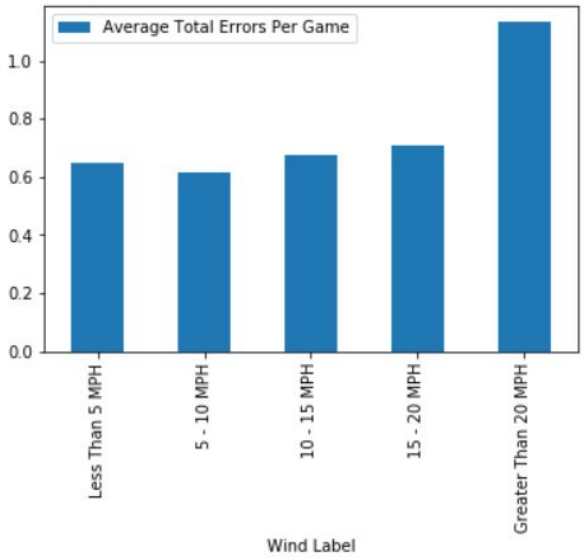
How does weather affect total errors per game?

Temperature
Using our temperature groupings, we looked at the total errors per game for each grouping. This showed something we did not expect.

Temperature Label	Average Total Errors per Game
<= 45 degrees	0.600000
>85 degrees	0.615152
>65 and <=85 degrees	0.630380
>45 and <=65 degrees	0.744395

The change in temperature does not follow a pattern as compared to a change in the average total errors per game. It actually shows that on average there are fewer errors per game in the more extreme cold or hot games. This perhaps could be due to not having as much data for games played in very hot or very cold conditions because this doesn't occur as frequently.

Wind Speed
We then looked at the average total number of errors per game by our wind speed groupings.



This showed that over the course of the 2018 season among all games played, as wind speed increases the average number of errors per game also increases. We see that when there are high winds (>20 mph) the average number of errors per game increases by 0.48 as compared to when there are low winds (<5 mph).

Weather Type

We then looked at the average total number of errors per game by the 9 different weather types. It showed that the weather type “rain” had the highest average number of errors per game, but the fewest average number of errors per game was in “drizzle.” Due to this contradiction, we then looked at the average total errors per game by only 2 weather type groupings - precipitation and no precipitation.

	Average Total Errors per Game
Weather Type Label	
Precipitation	0.631579
No precipitation	0.648558

This showed practically no difference in the average total number of errors per game in games that occurred with precipitation versus games that occurred without precipitation.

Summary of weather versus hitting

Through this data analysis it is clear that some aspects of weather have more of an impact on different batting factors than others. Baseball teams should want to play more games in hotter weather as it may result in scoring more total runs and home runs, based on our analysis. Baseball teams should also hope to play more games in less windy weather as high winds result in more errors. Additionally, a fan attending a game in the rain may have the benefit of seeing more total runs, as our analysis showed that on average more total runs are scored in games with precipitation.

Weather Versus Pitching

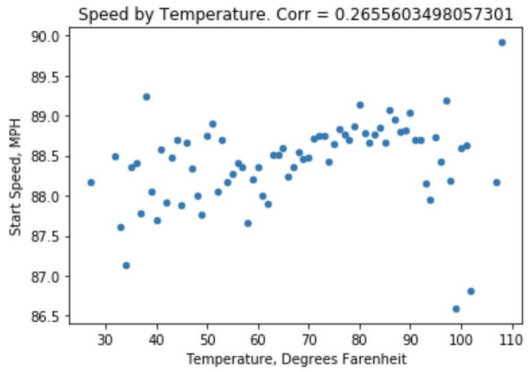
In this section, the effect of weather on various aspects of pitching is analyzed. We looked at how temperature, wind speed, and weather type impact pitches with regards to velocity, change in velocity, break, and spin rate. The following section will include scatter plots of average pitches in these various conditions and our interpretations of the relationships we've identified.

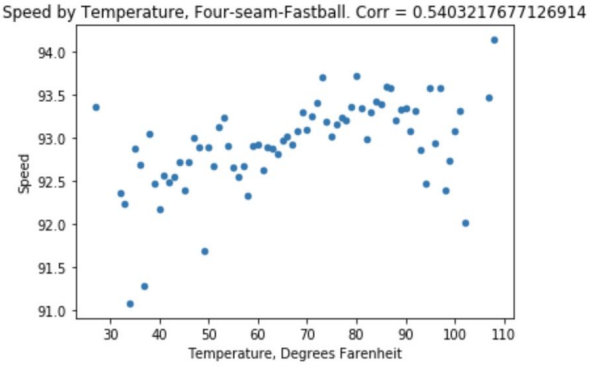
How does weather affect pitch velocity?

We looked at the effect of temperature, wind speed, and weather type on velocity of pitches. We used the column "start_speed" to look at velocity. We were seeking to answer the question of whether weather affects how hard pitchers are able to throw, rather than whether weather causes pitches to slow down after they are thrown. This is answered in a later section. Our results were as follows.

Temperature

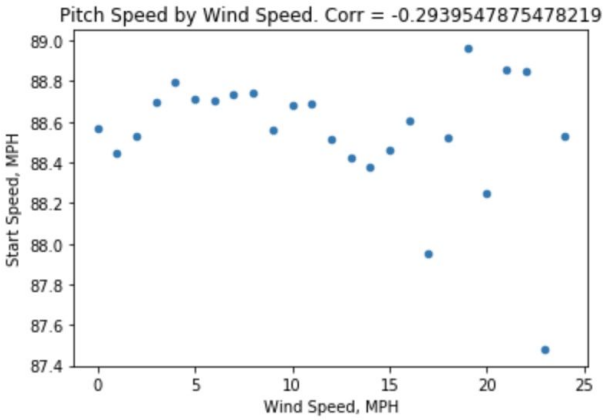
In order to evaluate the effect of temperature on pitch speed, we first plotted the average speed of pitches thrown at each temperature. This indicated that there was a possible relationship. We then extrapolated this to look at the average speeds of each pitch thrown at each temperature. This showed a stronger relationship for specific pitches, most notably four-seam fastballs. It appears that higher temperatures are related to faster pitches, especially fastballs.



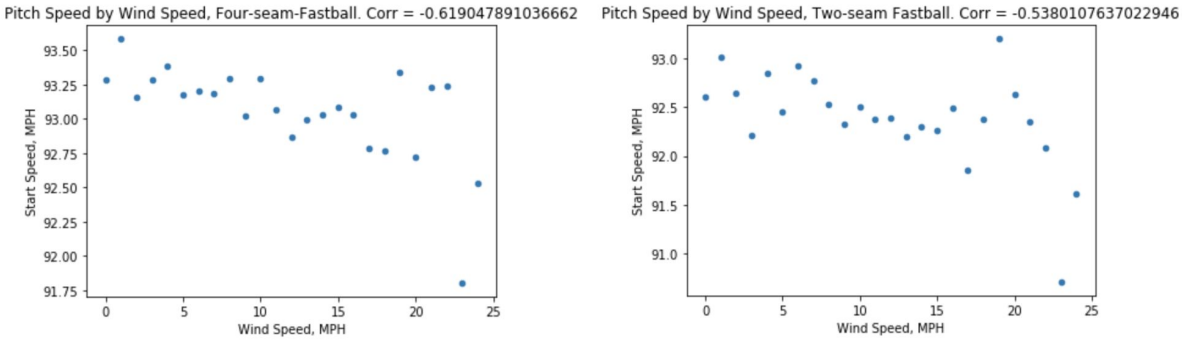


Wind Speed

Going through a similar approach with wind speed, we found no strong relationship between wind speed and pitch speed when all pitches were looked at together. However, when we looked at the relationship for pitches, we saw some evidence of a relationship with four-seam and two-seam fastballs. This relationship was present to a lesser extent in other pitches as well. It appears that higher winds decrease pitch velocity, especially fastballs.



:



Weather Type

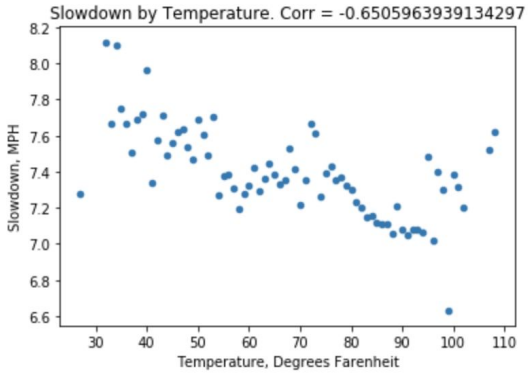
We were unable to find any consistent relationship between weather type and pitch speed. This could be because the other components of weather have a bigger impact on pitch speed, so they dominate any effects of weather type. This could be worth further inquiry, but as far as we could tell the effects of weather type on pitch speed are minimal.

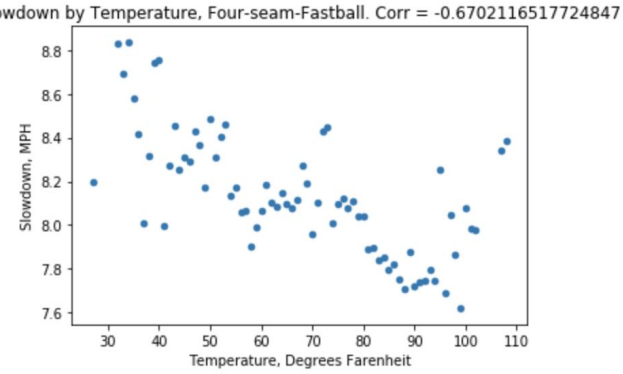
How does weather affect change in pitch velocity?

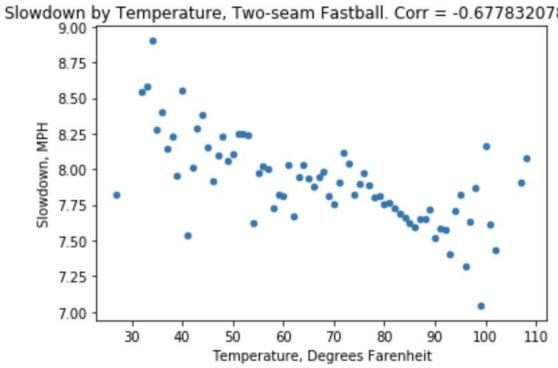
In addition to looking at start speed of pitches, we looked into how much pitches slow down between the time they are thrown and the time they are caught. We called this “slowdown”, and calculated it by simply subtracting “end_speed” from “start_speed”. The following are the results of our investigations into the effects of weather on “slowdown”.

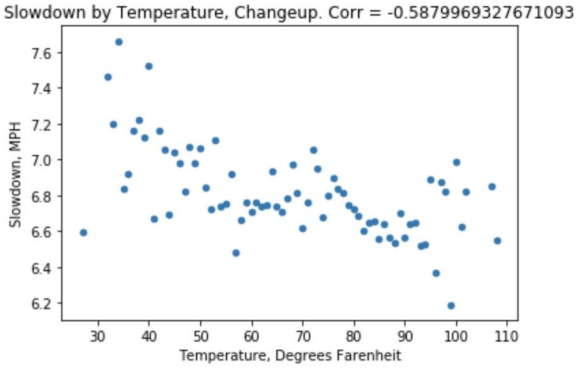
Temperature

It appears that pitches slow down less at higher temperatures. This trend was true for all pitches, but the correlation was strongest for four-seam fastballs, two-seam fastballs, and changeups. This is one of the most clear relationships we were able to discern. Below are charts relating slowdown to temperature for all pitches and for the three pitches with the most compelling relationships.



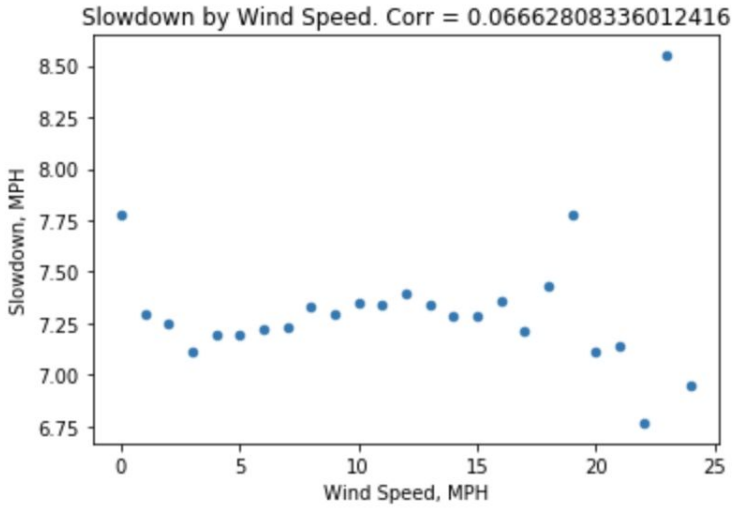






Wind Speed

Our investigation into the relationship between wind speed and slowdown did not discern any relationship. When looked at on the level of individual pitches there were also no interesting trends. The only consistent conclusion we came to was that slowdown became erratic at high wind speeds. We believe however that this is possibly due to small sample sizes, as relatively few games were played in high winds. Nonetheless it may warrant further investigation.

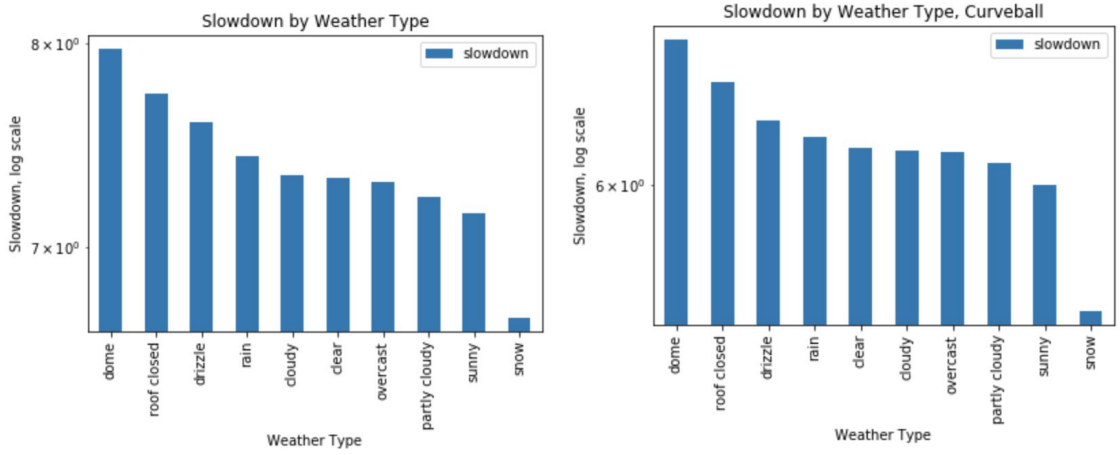


Weather Type

While the difference was small, it appears that slowdown is greatest in domed stadiums or stadiums with the roof closed. This trend held true across nearly all pitches. This could be due to the temperatures that indoor stadiums play at, since it is regulated at a reasonably cold temperature compared to many baseball games. This would align with what we observed earlier in this report.

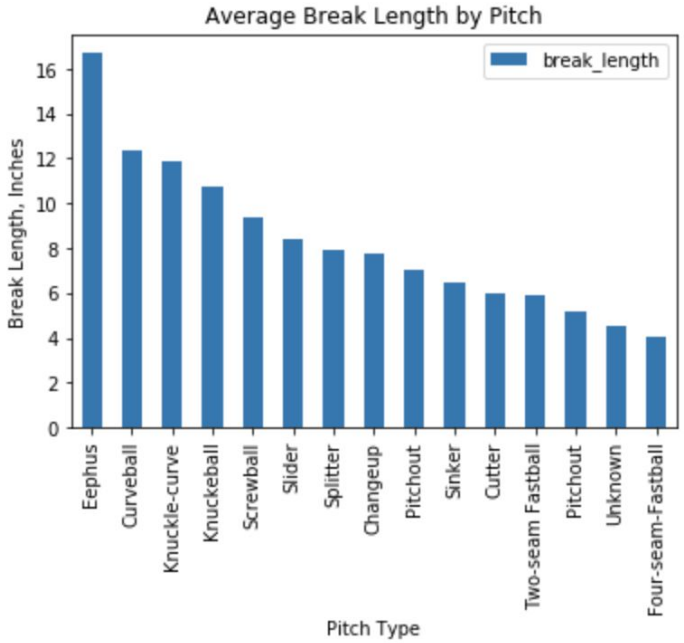
It also appears that rain and drizzle coincide with greater slowdown than other weather types. This too may have to do with the correlation between precipitation and cold temperatures. It is hard to say what the low result for snow means, given that only one game in our dataset was played in the snow.

These results were consistent across pitch types. We've included charts of average slowdown by weather type for all pitches as well as for curveballs as an example.



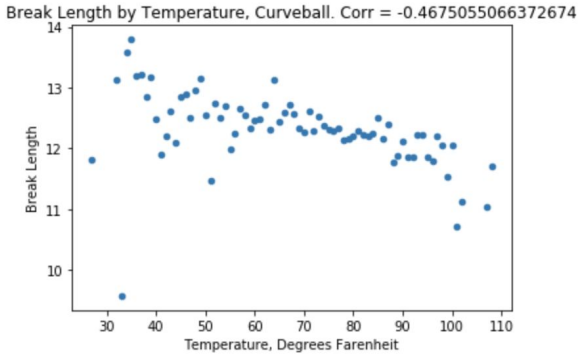
How does weather affect how much pitches break?

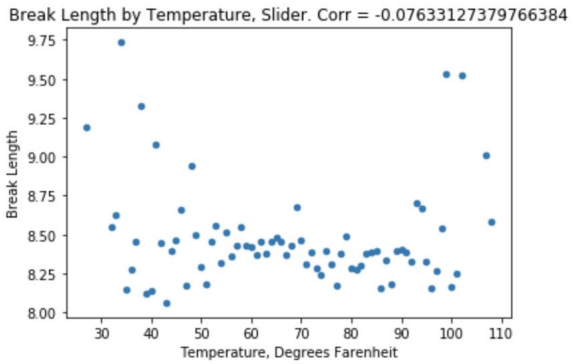
Break is a measure of how much a pitch moves. In our dataset it was "break_length". For some pitches, such as fastballs, break is not terribly important, but for some, such as curveballs and sliders, it is critical to effectiveness. The following chart shows how much pitches break on average for reference.



Temperature

The relationship between temperature and break length is unclear. For most pitches, it appears to be trivial, however for curveballs it was notable. As far as we can tell, this relationship is unique to the curveball, with no other pitches exhibiting nearly as strong of a pattern. Thus, the data suggests that curveballs break less in hotter weather, but other pitches are unaffected.





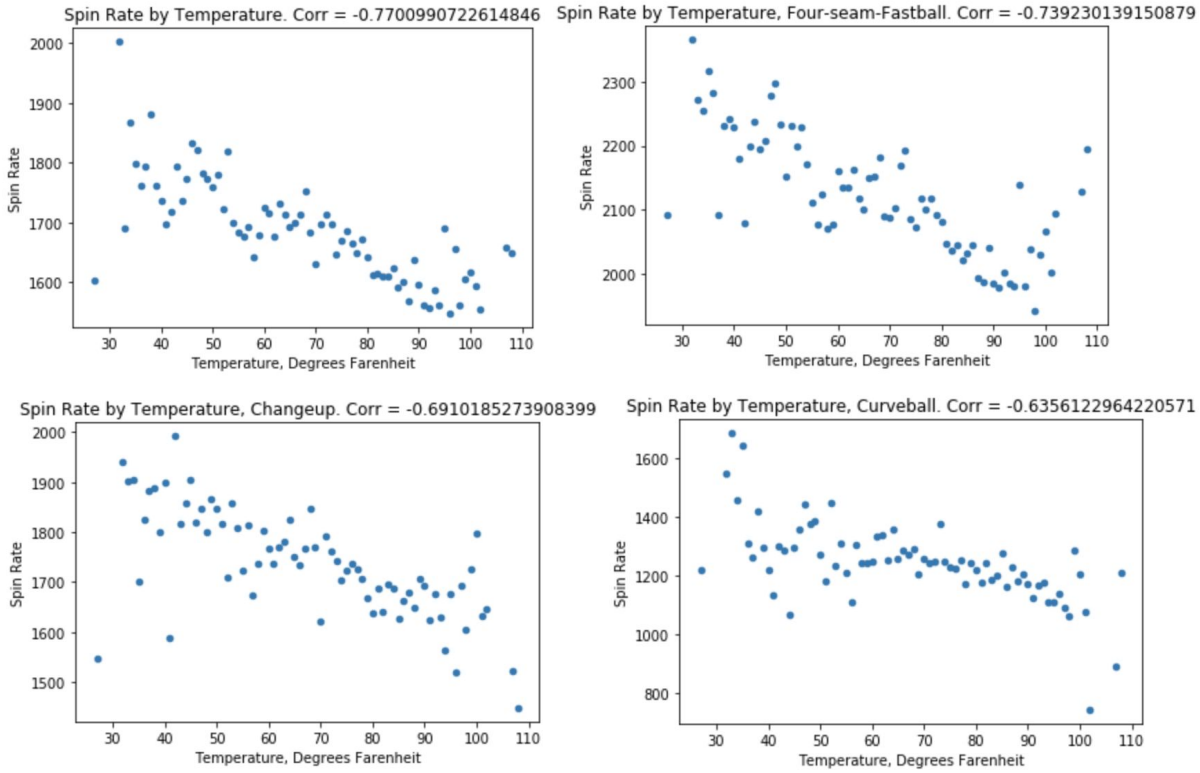
Wind Speed
Similarly with what we observed regarding the relationship between wind speed and slowdown, wind speed seems to have very little impact on pitch break. No pitch exhibits any evidence that its break is impacted by wind speed. As with slowdown, erratic behavior is observed in high winds, likely due to small sample sizes.

Weather Type
Differences in break length by weather type are minimal and likely trivial. As far as we can tell weather type has no meaningful effect on how much pitches break.

How does weather affect spin rate?

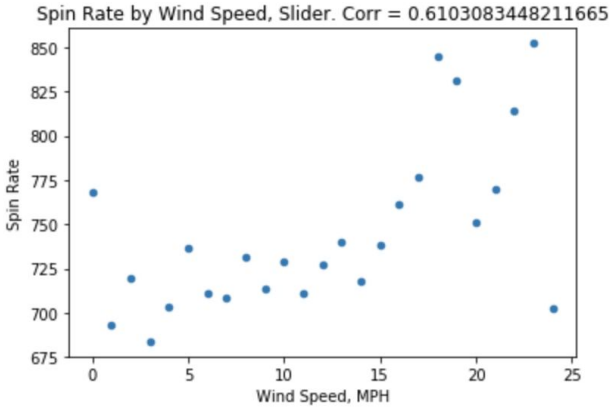
Spin rate is a metric to describe how fast (in revolutions per minute) a pitch is spinning as it approaches the plate. Generally speaking, pitches with higher spin rates move more and are thus more difficult to hit. Exceptions to this rule are changeups and knuckleballs, which are more effective at lower spin rates. We have found evidence that weather has an impact on spin rate, described below.

Temperature
We observed that pitches exhibit lower spin rates at higher temperatures. This is true across all pitch types. This relationship may impact how effective different pitches are in different temperatures. Below are charts showing this relationship for all pitches, as well as for three specific pitches that strongly exemplify it.



Wind Speed

The relationship between wind speed and spin rate is inconclusive. For most pitches it seems trivial, however for sliders it seems noteworthy. This warrants further investigation, but as far as we can tell it seems that sliders have higher spin rates in windier conditions, which would make them more effective. Below is a chart of the relationship between spin rate and wind speed for sliders.



Weather Type

The relationship between weather type and spin rate appears to be minimal. We are not convinced that weather type has any effect on spin rate.

Summary of weather versus pitching

We found a variety of relationships between weather and pitching. The most immediately interesting among them were the correlations between temperature and spin rate and between temperature and slowdown. The relationship between temperature and fastball speed was also promising, as was the effect of wind speed on slider movement. We think that of the three aspects of weather we have

identified temperature has the greatest effect on pitching, but all show some interesting relationships that could be of use to professional baseball teams.

Weather Versus Miscellaneous Factors

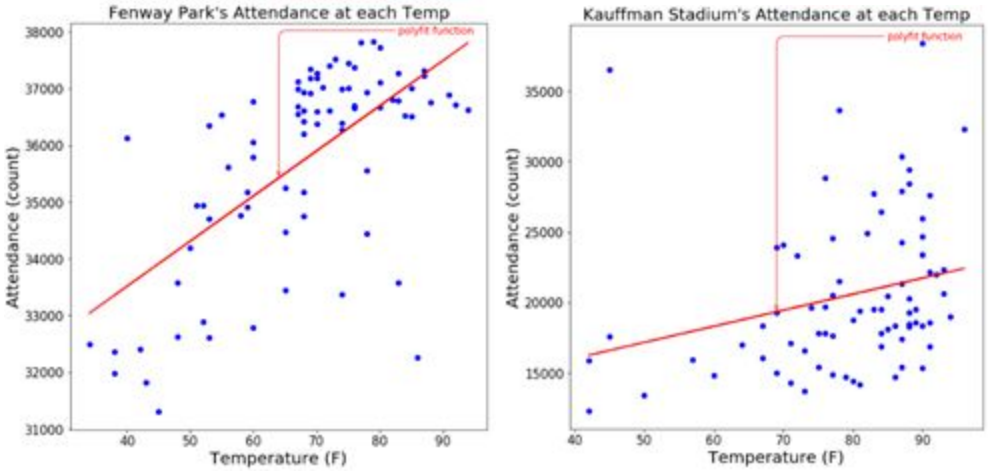
In this section, the effect of weather is analyzed against miscellaneous factors of attendance, timing, and the venue and teams of the 2018 baseball games. Weather effects include temperature, wind speed, and weather type. The assumptions in the introduction apply to this section as well. An explanation of each analysis is provided prior to each figure.

How does weather affect attendance of a game?

An assumption for this subsection is that games in the dataset with 1 attendance are removed due to fallacious data.

Temperature
Scatter plots of the temperature versus attendance were plotted for each of the 30 venues to mitigate any bias due to seating capacity of each venue. In general, most venues did not show a clear pattern except for two venues: Fenway Park and Kauffman Stadium.

The two scatter plots below show each game as a blue dot describing their attendance and temperature. The red line is a “poly-fit” function of 1-degree used to help our analysts recognize a pattern more easily even though this function generates a “poorly-conditioned” error. It is not to be mistaken that any in-depth statistics have been applied to this analysis. Fenway Park and Kauffman Stadium are the two venues that show a higher concentration of blue dots at the top (higher attendance) and to the right (higher temperature). These venues may benefit with temperature control such as a heating/cooling or a roof/dome such that at colder outdoor weathers, the stadium would still be warm and host more attendees.



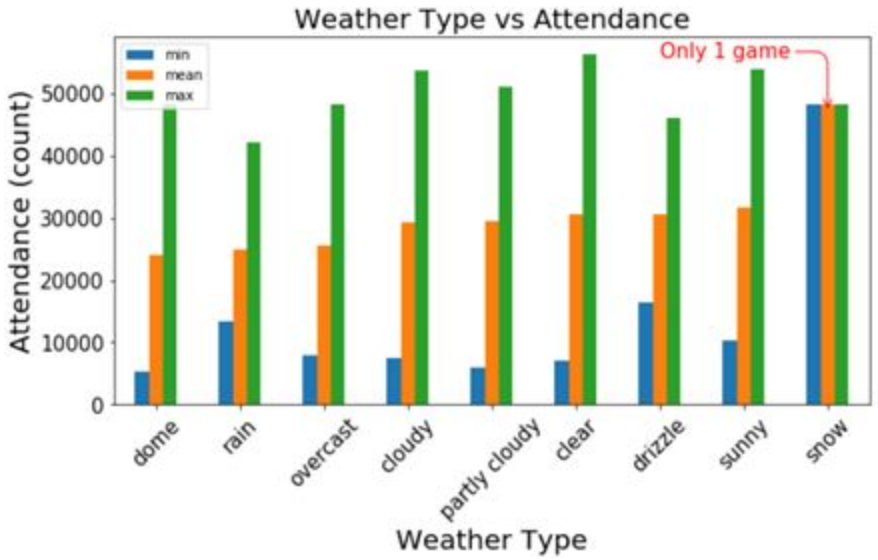
Wind Speed
A similar analysis was created for the attendance versus the wind speed. However, due to lack of insight and limitation of report length, plots were not displayed. Regardless, lack of insight may show that wind does not have an effect on attendance of each game. Therefore, it is not recommended to

provide wind control (such as a domed-stadium) to any of the 30 venues in order to regulate attendance at each game.

Weather Type
Due to the lack of varying weather types at each venue and that each venue can likely experience most of the different weather types, the following figure includes all 2018 baseball games plotted against each weather type.

The “Weather Type vs Attendance” plot shows the min, mean, and max attendance at each weather type. It is important to note that there was only one game with snow which means that it can be safe to ignore this weather type in the plot due to lack of data. Otherwise, the pattern is as expected: when it is sunny, clear, and drizzling, venues experience higher average attendance than when it is raining, overcast, or cloudy. It can also be expected that “dome” would have the lowest attendance since building a stadium with a dome may be smaller and hold less attendees due to the higher cost of construction compared to a stadium without a dome. Also, the 2018 domed stadiums tend to be older and reside in cities that are less into baseball with home teams having worse performance (another reason why the attendance may be lower than all other weather types). This figure shows that while a dome eliminates different weather types, a smaller stadium could potentially limit the seating capacity, hence lowering the total attendance compared to every other weather type.

We have not included a figure showing precipitation’s effect on attendance as the data analysis showed very little difference between attendance with and without precipitation. Precipitation may not affect attendances at each game.



How does weather affect timing of a game (elapsed time and delay)?

An assumption for this subsection is that a single game of 1860 minutes was removed (compared to an average delay of 68 minutes). This outlier affected the averages and plots and exaggerated the effects of weather on delay.

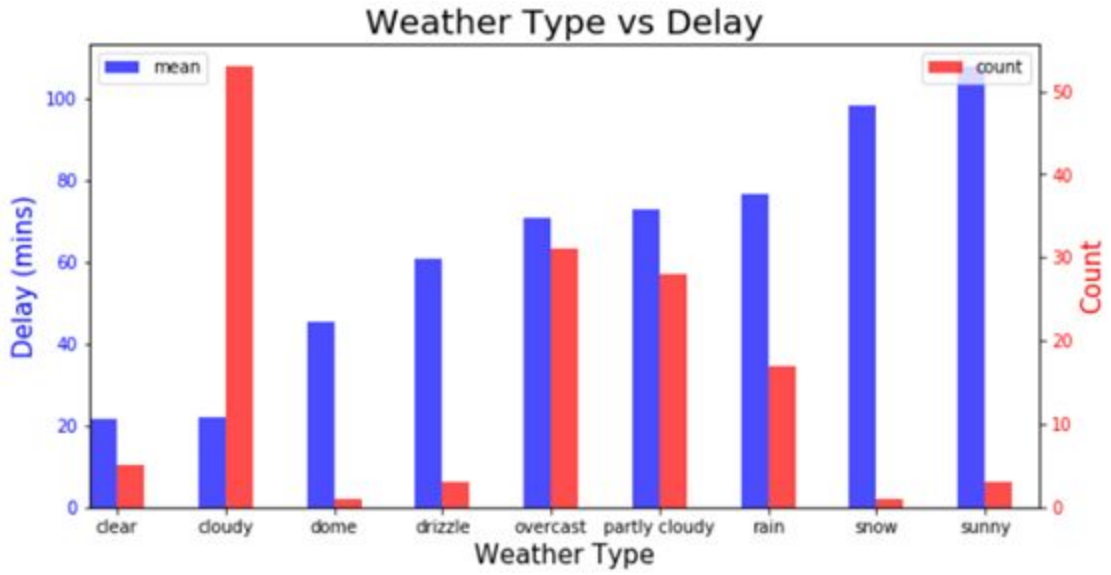
Temperature and Wind Speed
There did not appear to be any relationship between temperature and timing of a game nor any relationship between wind speed and timing of a game. Therefore, the non-insightful figures are not included.

Weather Type
The bar plot below shows some patterns regarding weather type versus delay. This figure plots the average delay of the baseball games at each weather type (in blue and the left y-axis). It also plots the frequency of delays at each weather type (in red and the right y-axis). Note that a precipitation plot was disregarded due to lack of information.

Let's explain the patterns:

1. When it is clear or cloudy, the delay isn't as long as when it is raining, snowing, or sunny.
2. The most delays occur when it is cloudy.
3. As noted before, there was only one snow game. The one snow game in 2018 led to the 2nd highest average delay of a game, as expected.
4. There are not very many delays when it is a dome! Yay dome!
5. There are also not many delays when it is drizzling or sunny or clear, as expected.

A dome stadium may help mitigate the number of delays.



How does the weather affect the venue and team?

Temperature and Wind Speed
A summary of the Top 5 Hottest Average, Coldest Average, and Highest Average Wind Speed venues are shown below. Based on these tables, further analysis was conducted on the top venue of each table.

Top 5 Hottest Avg. Temps

venue_name	
Globe Life Park in Arlington	86.51
SunTrust Park	80.37
Chase Field	80.13
Kauffman Stadium	79.45
Busch Stadium	78.34

Top 5 Coldest Avg. Temps

venue_name	
AT&T Park	62.69
Oakland Coliseum	65.81
Safeco Field	66.76
Fenway Park	67.84
Guaranteed Rate Field	69.05

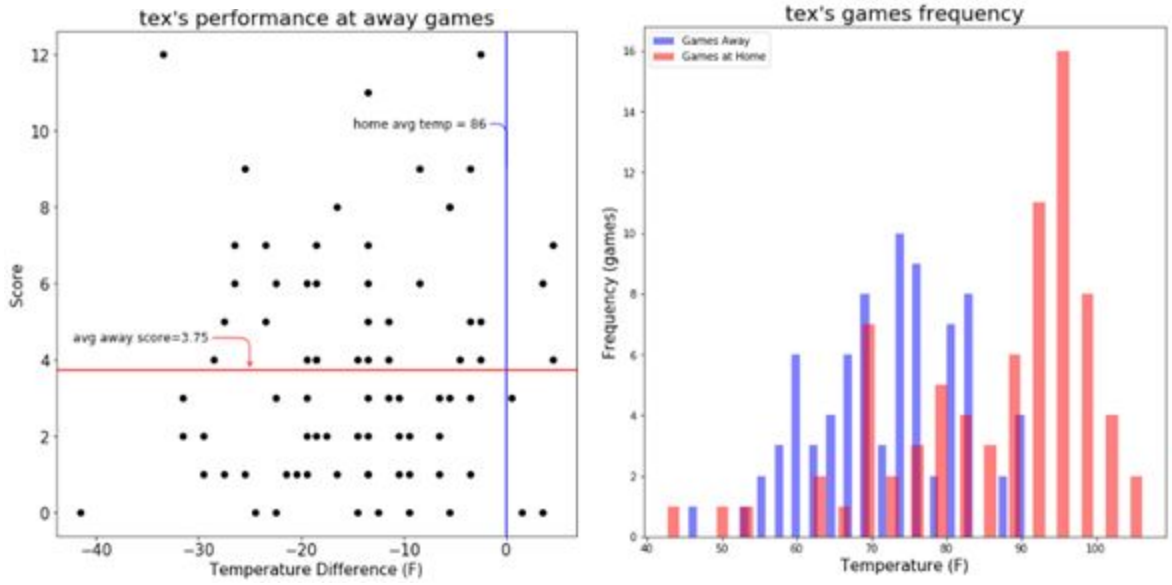
Top 5 Avg. Wind Speed (mph)

venue_name	
AT&T Park	14.12
Oakland Coliseum	11.20
Fenway Park	11.19
Citi Field	10.68
Globe Life Park in Arlington	10.26

Globe Life Park in Arlington is the home venue for the Rangers with an abbreviation of ‘tex’. This venue has the hottest average temperature. The following two plots explain the Rangers’ performance and game frequency relative to temperature.

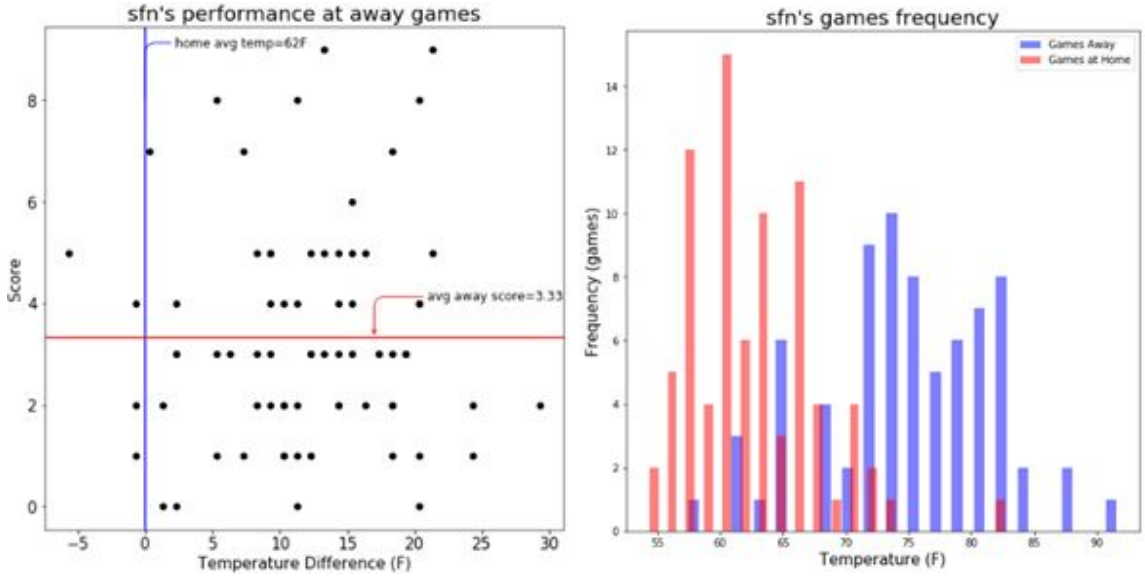
The plot to the left shows the Rangers’ score when they play away games (not at their home venue) versus the temperature difference of their away game compared to their home venue’s average temperature. Its intent is to answer, how do the Rangers’ perform when they are not “comfortable” (or playing at their home field conditions)? The blue line shows their home venue’s temperature (0F temp difference) of 86F. The red line shows their average away score of 3.75. Overall, this scatter plot doesn’t show any pattern in scoring, which means that the Rangers’ performance at temperatures different from their home venue does not change much. But if there was a recommendation to improve their performance, what would it be?

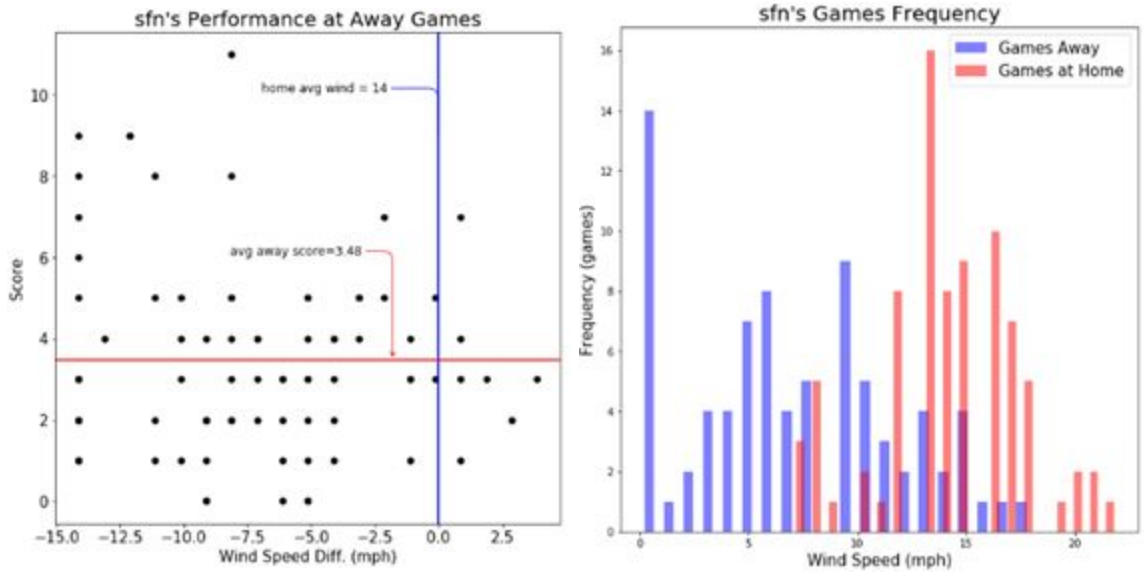
The plot to the right shows the Rangers’ number of games away from home (blue) and games at home (red) plotted against the temperature of each game. The Rangers’ typically play at colder temperatures when they are away than when they are at home. The Rangers’ may perform even better if they prepared for the colder temperatures when playing away games.



AT&T Park is the home venue for the Giants with an abbreviation of ‘sfn’. This venue has the coldest average temperature and highest average wind speed. The same two plots from the previous analysis are created for the Giants for temperature and wind speed.

The (x2) plots on the left also show that the Giants do not perform differently at their away games based on varying temperatures and wind speeds. The (x2) plots on the right show that the Giants frequently play at hotter temperatures and lower wind speeds at away games than at their home games. The Giants may perform better if they prepared for the hotter temperatures and lower wind speeds when playing away games.





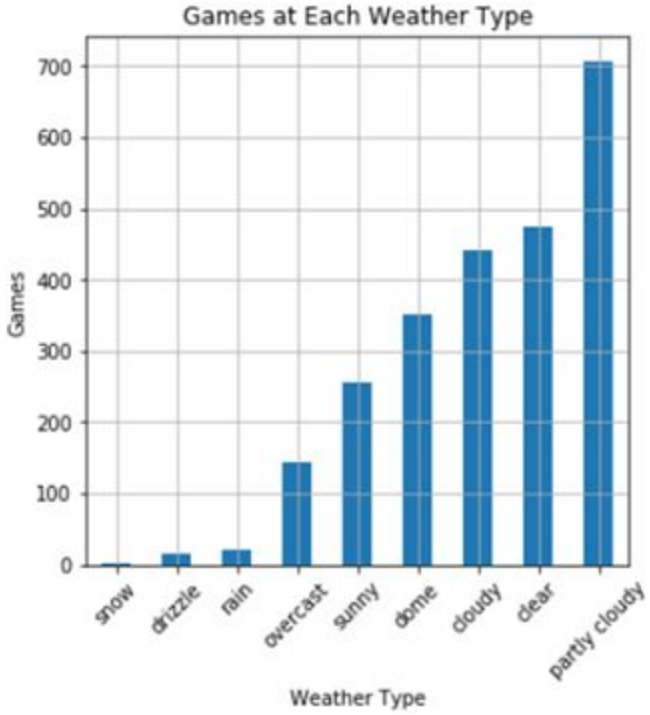
Weather Type

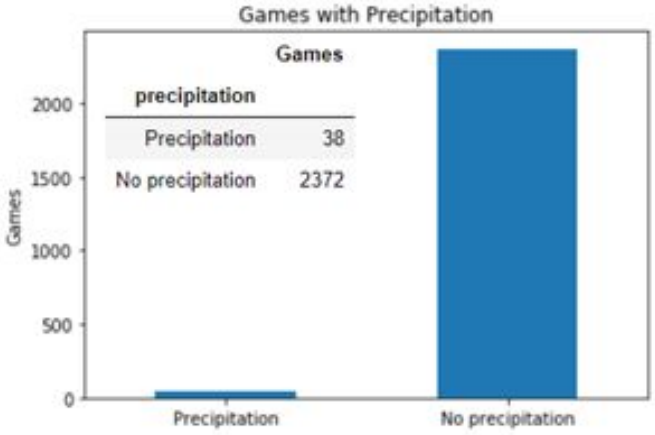
A summary of weather types versus the games is provided in the charts below.

Firstly, the bar graph on the left below shows the number of games at each venue. The graph of the weather types shows that baseball teams most frequently play at partly or cloudy weather. Sunny

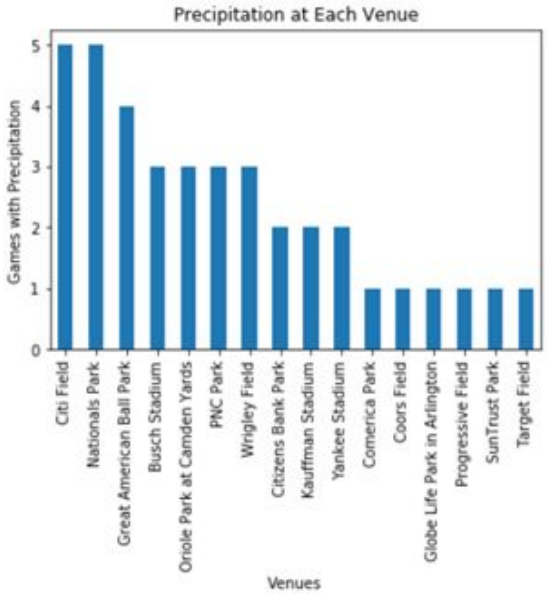
games do not occur frequently meaning that most of the time, players may not need to wear sunglasses.

The bar graph on the right below shows the number of games with and without precipitation. The table within the graph shows the actual values. This plot shows how few games there are with precipitation.





To find more helpful data, the bar graph below shows the number of games with precipitation at each venue. The table to the right shows the rarity of precipitation at these games. The top 5 venues with most frequent precipitation only had single digit percentages of games that had precipitation. These figures show that these games do not typically have precipitation but if an attendee would like to bring an umbrella, bring it when going to Citi Field or Nationals Park.



venue_name	Games w/o Precip.	Games w/ Precip.	% of Games w/ Precip
Citi Field	74	5	6.33
Nationals Park	75	5	6.25
Great American Ball Park	77	4	4.94
Oriole Park at Camden Yards	75	3	3.85
PNC Park	75	3	3.85

Summary of weather versus miscellaneous factors

Attendance of a game:

- Many venues may not increase in attendance if a form of temperature control is provided such as a heating/cooling unit or a roof/dome. However, venues that may benefit include Fenway Park and Kauffman Stadium.
- Implementing wind speed control at venues does not appear that it would affect attendance of a game.
- It may not be a good idea to provide a dome to mitigate negative effects of weather on attendance. The extra cost in constructing a dome may lower the maximum seating capacity at a venue, hence, the data may show that a dome has a lower average attendance than venues without domes at all other weather types (sunny, drizzling, etc…).

Timing of a game:

- Temperature and wind speed may not affect the length or the delay of a game.
- A dome stadium may help mitigate the number of delays at a game.

Venue and Team's Performance:

- The Rangers may perform better if they are prepared for the colder temperatures at away games.
- The Giants may perform better if they are prepared for the hotter temperatures and lower wind speeds at away games.
- Players may not need to wear sunglasses often since sunny weather is not frequent compared to other weather types.
- Attendees may not need to worry about precipitation when attending a baseball game but may consider bringing an umbrella when going to Citi Field or Nationals Park as those are the venues with the most precipitation (at 5 games each).

Conclusion

There were a variety of different trends related to weather and baseball. As we have demonstrated, the outside air temperature, wind speed and direction and precipitation have numerous effects on hitting, pitching and the game itself.

In regards to hitting, it is apparent that higher temperatures tend to lead to nearly 20% more runs being scored in the game. Thus, it seems like there is generally more offensive output in games where the temperature is above 85 degrees. We also noticed that there were more home runs in higher temperatures, so it is possible that these are related. It is also notable that there tended to be more runs scored when there was precipitation. There was an average increase of more than a full run, when it was raining during a game in 2018. However, this tended to occur as a result of increased hits, as there was not a noticeable effect between errors and precipitation nor a strong effect between home runs and precipitation. In addition to the effect of precipitation on runs, there also tended to be more runs scored when there were winds of greater than twenty miles per hour. This tended to cause a distinct increase in errors, as the ball tends to move more in the air, and thus possibly led to more runs.

On the pitching side, it is notable that there was a relationship between pitching speed and temperature increases. It appeared that the pitch speed increased, particularly for pitches, like the four-seam

fastball, as the pitches were thrown in higher temperatures. There also tended to be a greater change in ‘pitch slowdown’ as the weather temperature tended to increase. It seemed that there tended to be a greater slow down in speed as the pitch left the pitcher’s hand and crossed the plate, in colder weather games, particularly for two-seam fastballs and changeup pitches. Additionally, we found that there was a negative trend between spin rate and temperature, as the spin rate tended to be lower in higher temperatures. We also noticed a slight trend related to pitch break, as it tended to less break on pitches in lower temperatures and more break in higher temperatures. Generally, these tended to be our strongest effects, however we noticed weaker trends between variables such as the wind velocity and pitch speed. However, our other pitching variable trends such as pitch slowdown and wind speed, spin rate and weather type, or weather type and pitch speed, had little variation between them.

In the other aspects of baseball, there were certainly a variety of effects as well. Venues may see an increase in attendance in higher temperatures, and may benefit from some sort of temperature control unit to attract more fans, in those colder weathers. However, those venues need not worry about the effect of wind speed on attendance, as wind speed at venues does not appear to affect attendance of a game. However, the weather can have other consequences, as well. For example, the weather can often cause games to be delayed. For example, the Denver-based Colorado Rockies experienced a long delay in April due to snow falling on the field at their outdoor stadium. Alternatively, there are also not many delays when it is drizzling or sunny or clear, as expected. There is also almost never a delay in a domed stadium, as this helps to mitigate the number of delays and having delays can certainly affect a team, particularly since each team plays half of their games in one building. However, this can have an effect as well, as a team that is used to playing in one area has to adapt to the changing climate when playing in another. For example, teams will play games in cities with higher or lower average temperatures than their own, as well as both rainier and sunnier cities, all of which causes different effects on players.

After considering the problem from multiple angles, it is obvious that there is a demonstrable effect of the weather on the game of baseball. Temperatures, weather types, wind speeds and directions and precipitation levels can dramatically affect how the game is played. In terms of hitting, batters may experience better or worse performance depending on the conditions. In terms of pitching, pitchers may have to alter their strategy for how they choose to throw. In terms of attendance, teams may need to lower ticket prices to increase attendance when the weather is bad. Ultimately, weather can have dramatic effects on baseball games.