Major League Baseball Pitcher Performance Data Analysis Report

Miriam Aguirre, Christina Li, Vinh Ton, Amir Yaacoobi

STA 160: Data Science

Professor Fushing Hsieh

Final Project Spring 2024

Background

Major League Baseball, also known as the MLB, is the professional baseball league of the United States and Canada and the teams consist of the highest rated baseball players. The MLB consists of 30 teams with two different leagues: National League and American League and three divisions: East, Central, and West. Baseball is played by two teams and needs at most nine players on each team. To win the game, there are nine total innings where one team attempts to score and the other team is the offense. Throughout the game, there are critical strategies that are played by the defense team to prevent the scoring team from scoring. An essential player is the pitcher, who throws the ball to the opponent team and is the first person to defend the team. Pitchers can throw more than 100 pitches per game, so there is a lot of data that can be utilized to determine what makes a pitcher successful.

Introduction

Justin Verlander, one of the most renowned pitchers in Major League Baseball (MLB), has had a star studded career marked by accolades and amazing performances. Although he has been in the MLB since 2004, in this exploration we want to analyze Verlander's pitching performance over a five-year period from 2015 to 2019. The objective is to examine the trends and variations in his pitches to understand how his career evolved during this time frame.

Major League Baseball pitchers use various types of pitches, which we will become acquainted with in this exploration. This project focuses on four of Verlander's primary pitch types: slider (SL), curveball (CU), changeup (CH), and fastball (FF). By analyzing metrics such as vertical and horizontal movement, velocity, acceleration, and pitch zones, we hope to find patterns in Verlander's performance and control over these pitches.

The data used for this analysis is sourced from game records from the 2015 - 2019 seasons. The scraped data sets contain information about each pitch, including the date, pitch type, movement characteristics, and pitch outcomes. With this holistic dataset, we hope to see how Verlander fared as a pitcher in the Major Leagues.

Data Description

pitch_type: The type of pitch thrown (e.g., SL for slider, CU for curveball, CH for changeup).

game_date: The date when the game was played.

release_speed: The speed of the pitch at the time of release, measured in miles per hour.

release_pos_x: The horizontal position of the pitch at the time of release.

release_pos_z: The vertical position of the pitch at the time of release.
player_name: The name of the player (Justin Verlander).
batter: The identifier for the batter facing the pitch.
pitcher: The identifier for the pitcher (Justin Verlander).
events: The result of the pitch (e.g., field_out, strikeout).
description: A detailed description of the pitch outcome (e.g., hit_into_play, swinging_strike).
pfx_x: The horizontal movement of the pitch by the time it reaches home plate.
pfx_z: The vertical movement of the pitch by the time it reaches home plate.
effective_speed: the perceived speed of the pitch by the batter
vx0: The velocity of the pitch in the x-direction at release.
vy0: The velocity of the pitch in the y-direction at release.
vz0: The velocity of the pitch in the z-direction at release.
ax: The acceleration of the pitch in the x-direction.
ay: The acceleration of the pitch in the y-direction.
az: The acceleration of the pitch in the z-direction.
zone: The zone location of the pitch when it crosses the plate.
pitch_name: the full name of the pitch
spin_axis: The axis of the pitch spin.
spin_rate_deprecated: The rate of spin of the pitch, measured in revolutions per minute (RPM).
release_spin_rate: The spin rate of the pitch at the time of release, measured in RPM.
plate_x: The horizontal position of the pitch as it crosses the plate.
plate_z: The vertical position of the pitch as it crosses the plate.
game_year: The year that the pitch was thrown

Data Cleaning and Summaries

The data for pitcher Justin Verlander is from Statcast. Statcast is an advanced technology to capture various variables. such as speed of the ball, location of the pitch, and pitch type throughout the game. In analyzing the performance of Verlander, we wanted to focus on a few key variables. We chose the variables pfx_x, pfx_z, vx0, vy0, vz0, ax, ay, az, and zone. We chose these variables because they depend entirely on the pitcher, without taking into account other positions, such as who the batter is, what the state of the field is, etc. We also kept game_year as a way to break up the dataset.

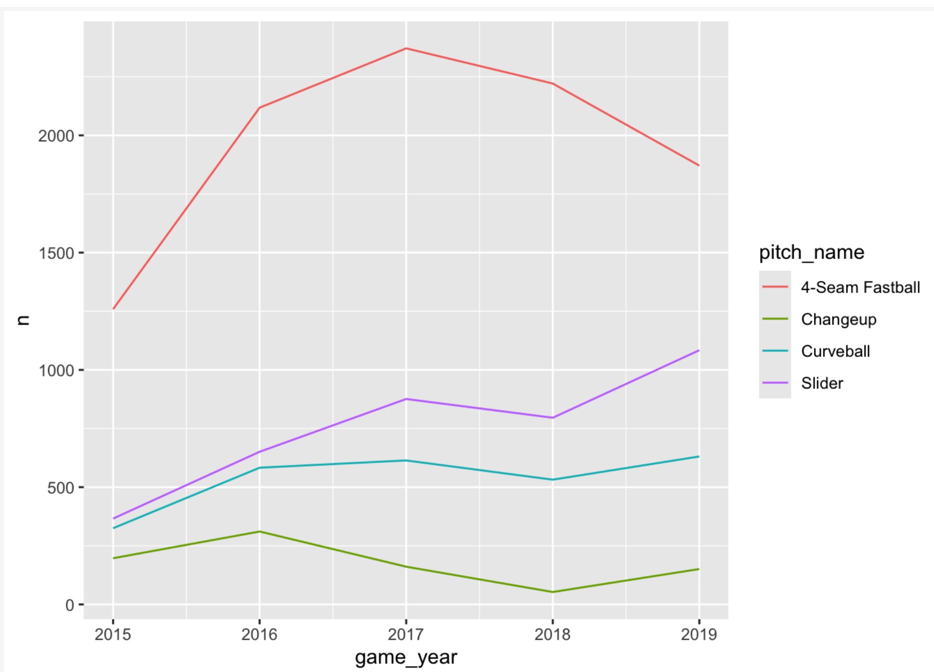
We also acknowledged that different pitch types would intrinsically be different; a fastball, for example, will naturally be much faster than a knuckleball, which is slow and unpredictable. Thus, we decided to analyze each of the aforementioned variables in relation to the pitch type.

In the years 2015 to 2019, Statcast listed nine different types of pitchers that Verlander used. The pitch types “intentional ball” and “pitch out” were excluded because they were used in specific game situations to allow for the best chance of success for the team. We also excluded cutters and sinkers due to their low sample size (n=48 and n=12, respectively), which would make analysis difficult. Finally, there were 214 unnamed pitches that were excluded because without a clear categorization, it makes it difficult to compare similar pitches

That left us with four pitch types to analyze over the five-year period: sliders, curve-balls, change-ups, and four-seam fastballs.

Exploratory Data Analysis

To understand how Justin Verlander’s game has evolved over the five year span we are analyzing, we must create visualizations and break up his pitching into four distinct groups: fast balls, curve balls, change up pitches, and sliders. We analyze different parts of the pitches, inspecting the velocity, acceleration, and accuracy based on the x, y, and z dimensions.

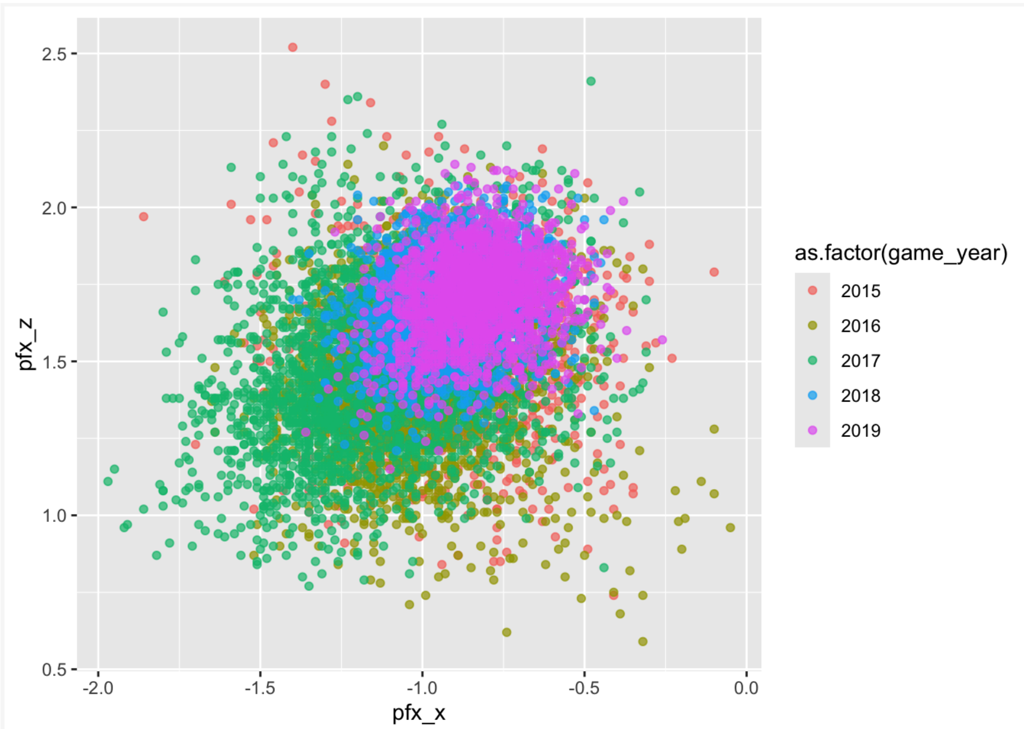


We can see that Verlander mostly pitches 4-seam fastballs for the entire 5 year period. From 2015 to 2017, the frequency with which he pitches these increases. The same pattern is true for sliders and curveballs. However, after 2016, the frequency with which he throws change-ups decreases drastically until 2018. From 2018 onwards, the amount of fastballs he pitches decreases harshly, while he throws more changeups, curveballs, and sliders.

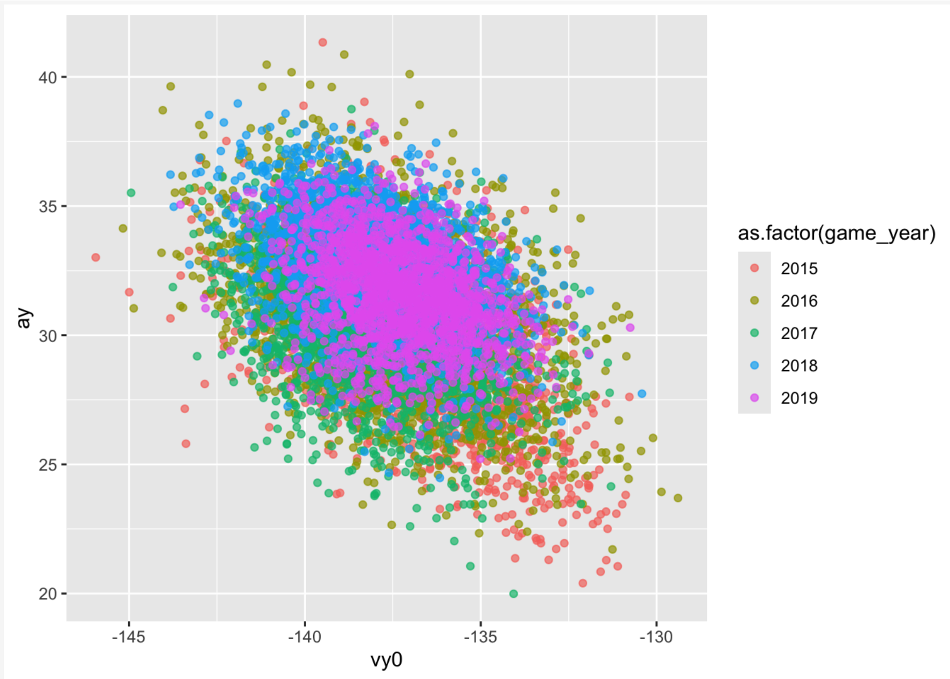
Four-Seam Fastballs

A fast ball, Verlander's most commonly thrown pitch over the course of the five years we are analyzing, is a pitch thrown at full speed, often reaching speeds of 95+ mph, that appears to rise as it approaches the batter. Fast ball pitches overall are the most commonly thrown pitch by all pitchers in the MLB, and seems to be Verlander's speciality.

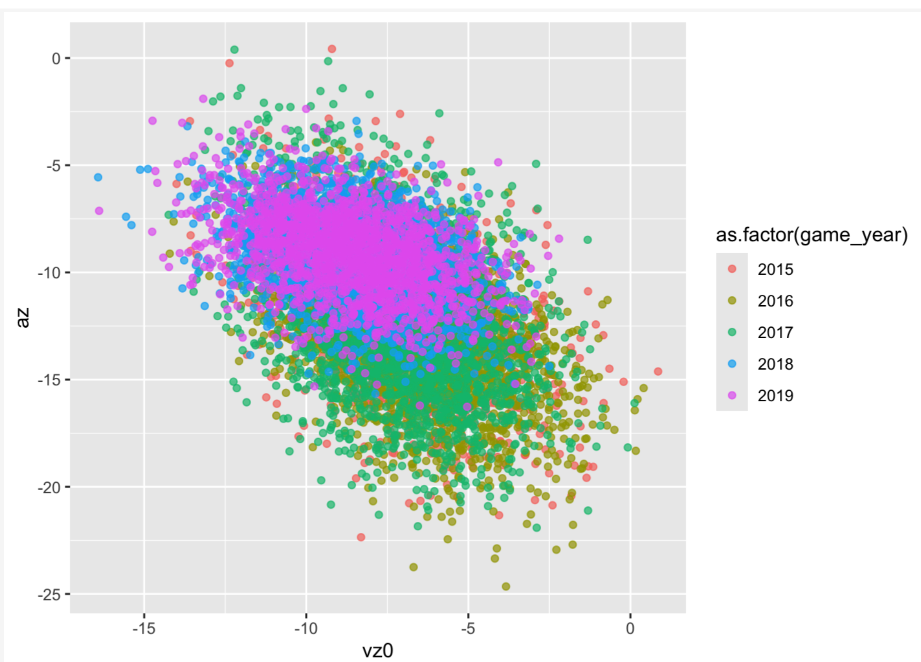
We can see that Verlander mostly pitches 4-seam fastballs for the entire 5 year period. From 2015 to 2017, the frequency with which he pitches these increases. The same pattern is true for sliders and curveballs. However, after 2016, the frequency with which he throws change-ups decreases drastically until 2018. From 2018 onwards, the amount of fastballs he pitches decreases harshly, while he throws more changeups, curveballs, and sliders.



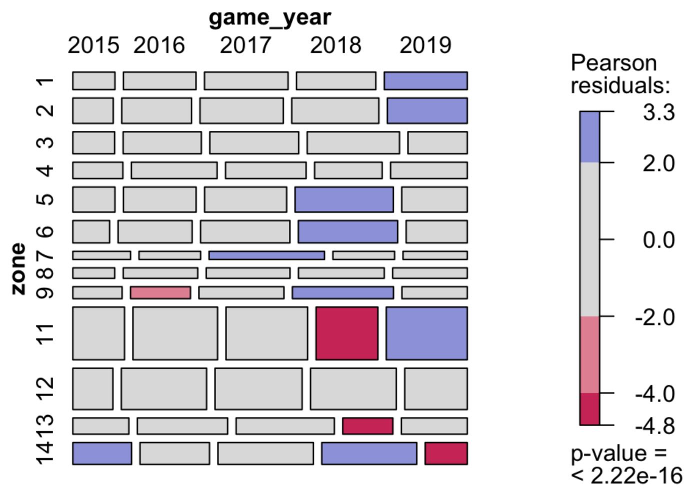
From a glimpse at this graph, we can see that over time, Verlander's pitches become much more condensed and accurate. His pitches in 2015 and 2016 tend to spread out the most in this scatter plot; however, his pitches in 2019 are all heavily concentrated in one area, in the middle right hand side with much less spread. This indicates an increase in control in his fastballs.



Similarly, for y, we see that Verlander's y velocity actually saw a tighter spread, which meant his velocity became more consistent, at around -137 ft/s/s. A similar story can be said of acceleration.



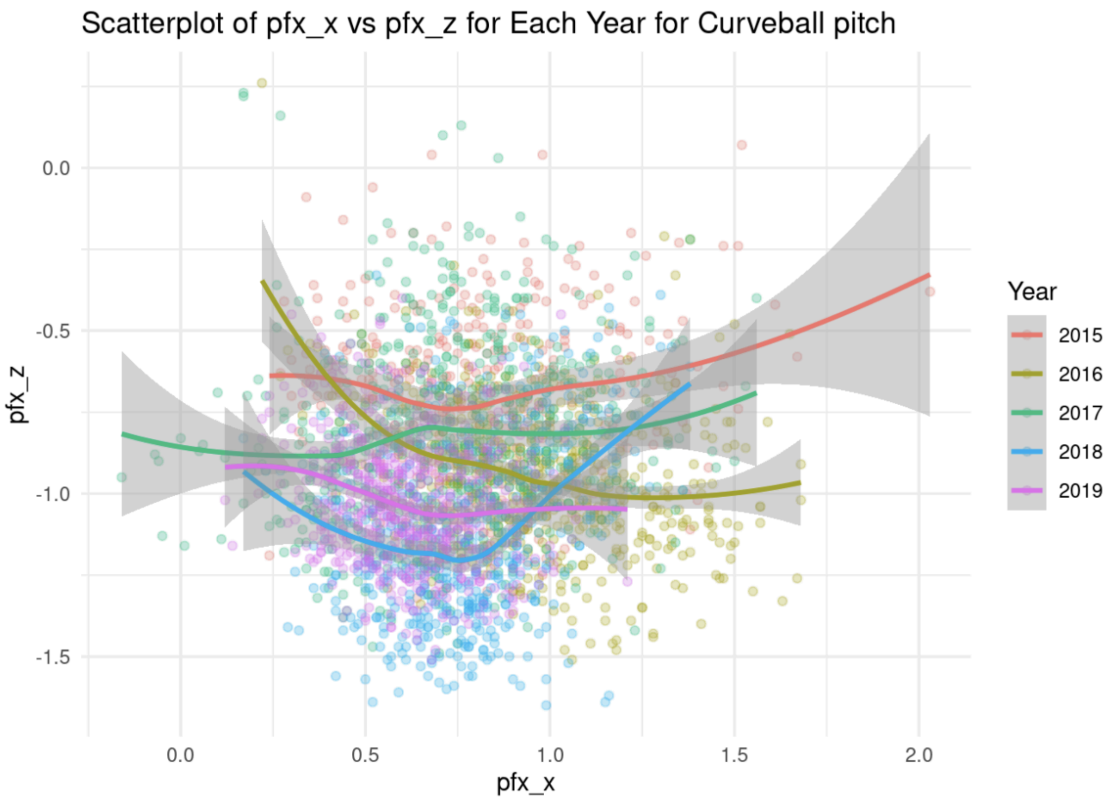
In terms of z velocity, Verlander's fastball pitch speed became slower but again more consistent. However, his z acceleration in 2019 was consistently higher and more condensed than his pitches in previous years, hovering at around -7 ft/s/s.



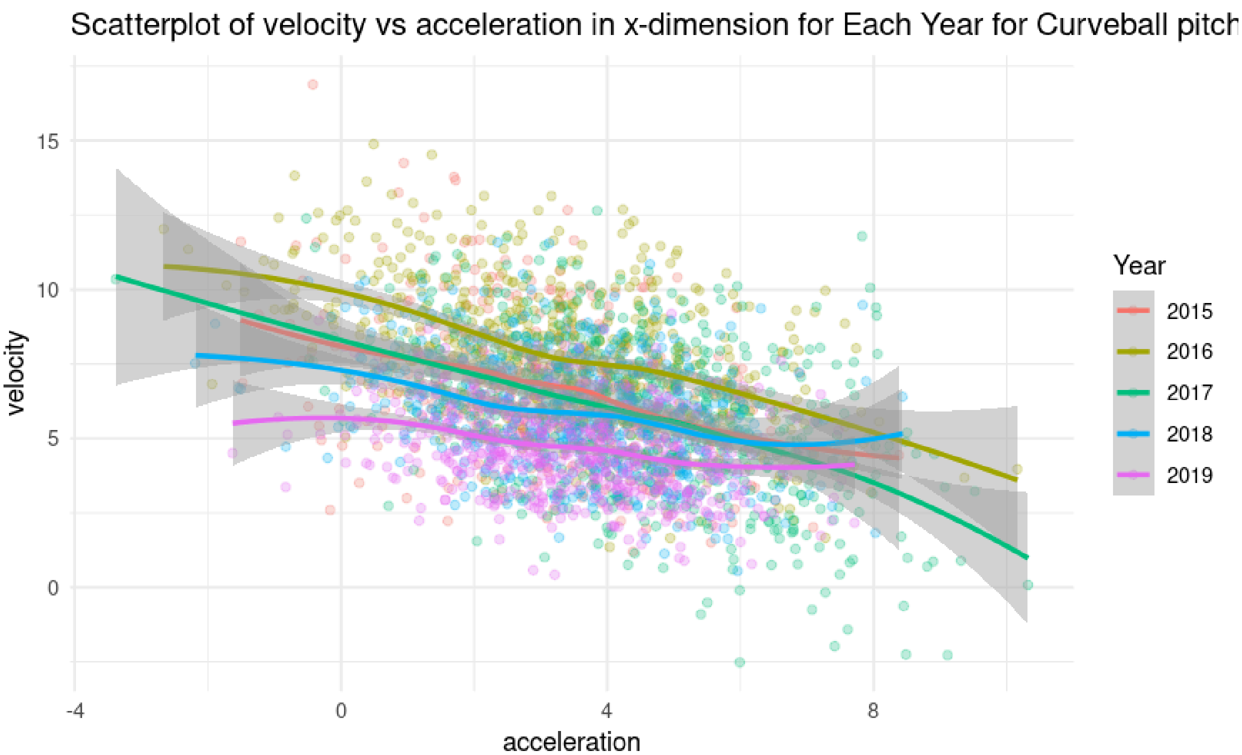
Here, the colors of the mosaic plot indicate values of note. Specifically, we can see that in 2015, Verlander pitched in zone 14 more often than you would expect; however, in 2019, he pitched more frequently in zones 1 and 2 and much less frequently than in zone 14. This would indicate, for example, that Verlander's fastball pitches landed consistently higher in 2019 than he did in 2015.

Curve Ball

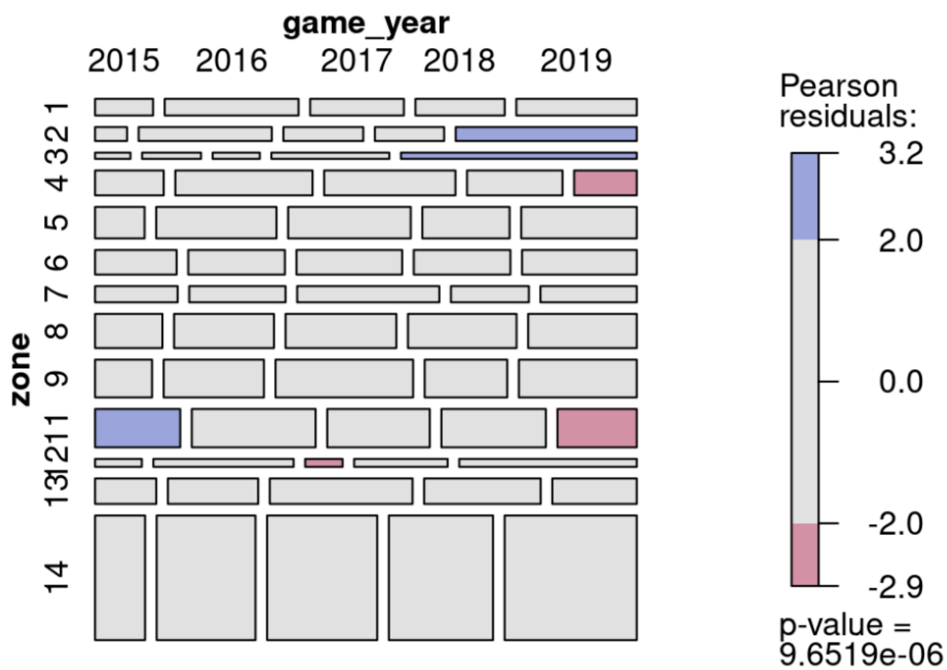
A curveball pitch in baseball is a type of pitch that is characterized by its significant downward movement as it closes in on the batter. The trick to throwing this type of pitch is the topspin applied at the throwing point. The curveball's trajectory can also include lateral movement, depending on the pitcher's grip and release technique. When compared to a fastball, curveball pitches are purposefully thrown at a much slower speed, anywhere between 65- 80 mph, in hopes of throwing off the batter with the change and direction and speed.



Over time, Verlander's curveball pitches have shifted more to the center of the plate in the horizontal direction. They also seem to have shifted lower down in the vertical direction. It seems that Verlander's curve balls have transitioned more from horizontal movement to vertical movement. This could eba change in his pitching style to have more of a drop curve ball instead of a side curveball.



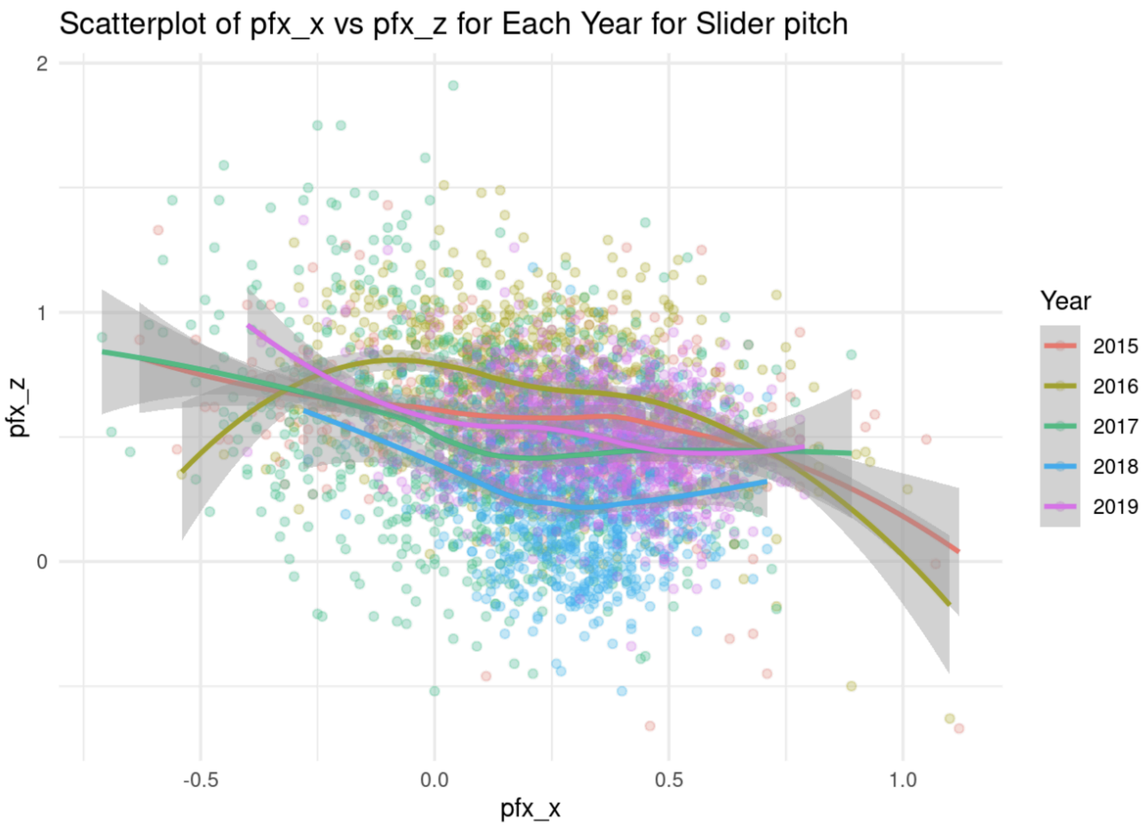
On average, it seems velocity in the x-direction has had some slight changes throughout the years for curveballs. Compared to 2015, we see significant differences in 2017, 2018, and 2019. Comparing 2019 to 2018, we see little significant difference. This shows that Verlander may have become more comfortable with that specific type of velocity in 2018 and continued to apply it in the next year. As for acceleration, there are significant differences in all the years except for 2017 vs 2018. However, there is a significant difference between 2019 and 2018, which shows that Verlander may have changed his acceleration techniques after that year. Due to the differences between 2019 and all the other years, we may conclude that Verlander's curveball pitches have changed significantly in terms of x-dimension acceleration. Due to the unpredictable nature of curveballs, it is important to have variation in acceleration to confuse opponents.



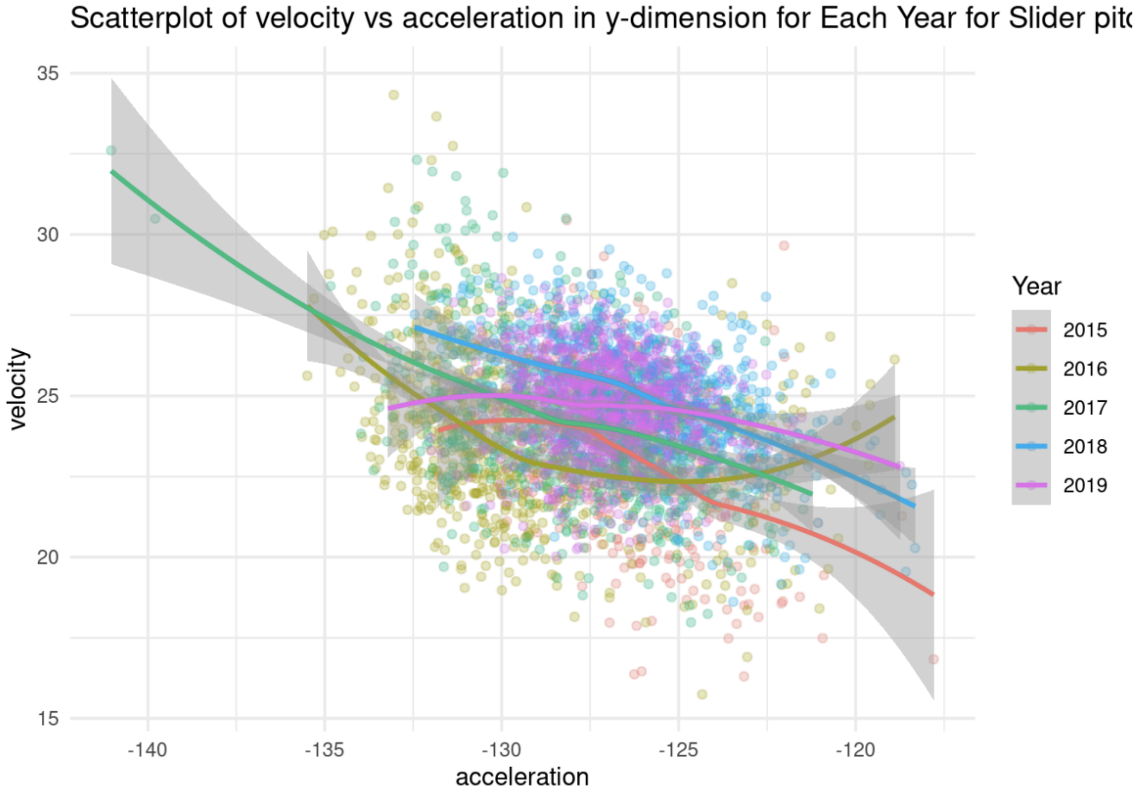
Verlander's curveballs seemed to be largely consistent, with few noticeable differences in which zones he pitched in. However, in 2015 he did pitch more frequently in zone 11; by 2019, this shifted so that he was pitching more in zones 2 and 3.

Sliders

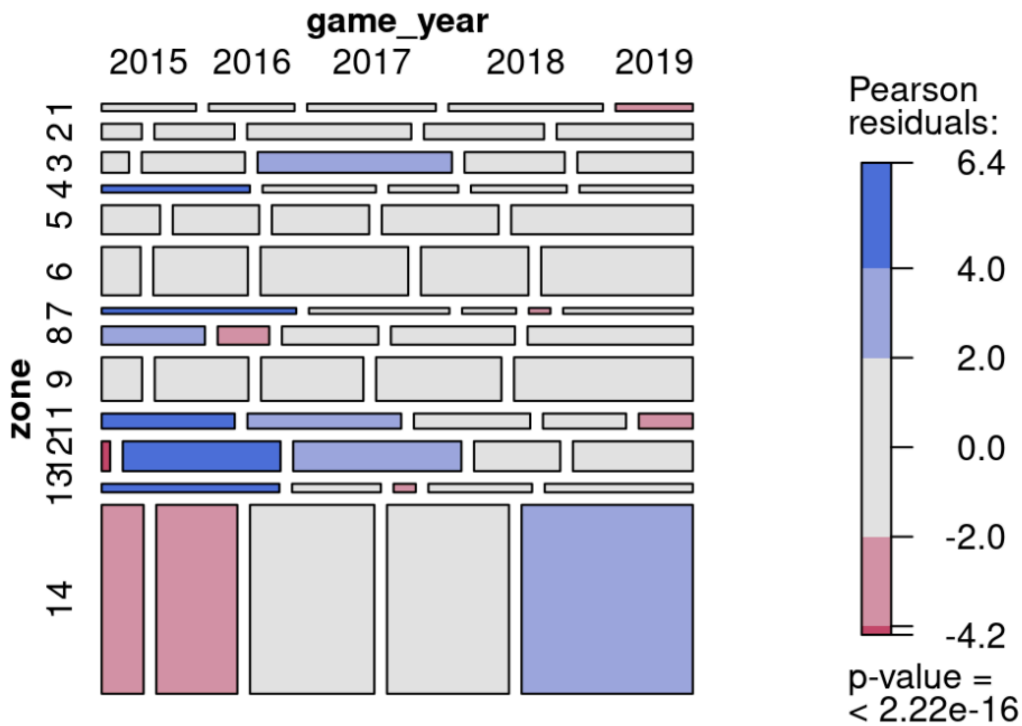
A slider pitch, Verlander's second most commonly thrown pitch, is typically thrown faster than a curveball but slower than a fastball, usually ranging between 80 to 90 miles per hour. The goal of a slider pitch is to be disguised as a fast ball pitch, but having a late breaking, in either the downward or lateral direction to deceive the batter. The following visualizations we will be analyzing encompass Verlander's slider pitches during the time frame discussed.



It seems that overtime, Justin Verlander's slider pitches have become much more condensed and accurate. In the years 2015-2017, we can see the most variation in both axes whereas 2018 and 2019 show the least variation. Further, it seems that variability in 2019 was higher than 2018. This indicates that Verlander's control in slider pitches has increased overall though there was a slight decrease in performance from 2018 to 2019. Additionally, for this type of pitch it seems that Verlander has a lot more variation in horizontal movement than in vertical movement.



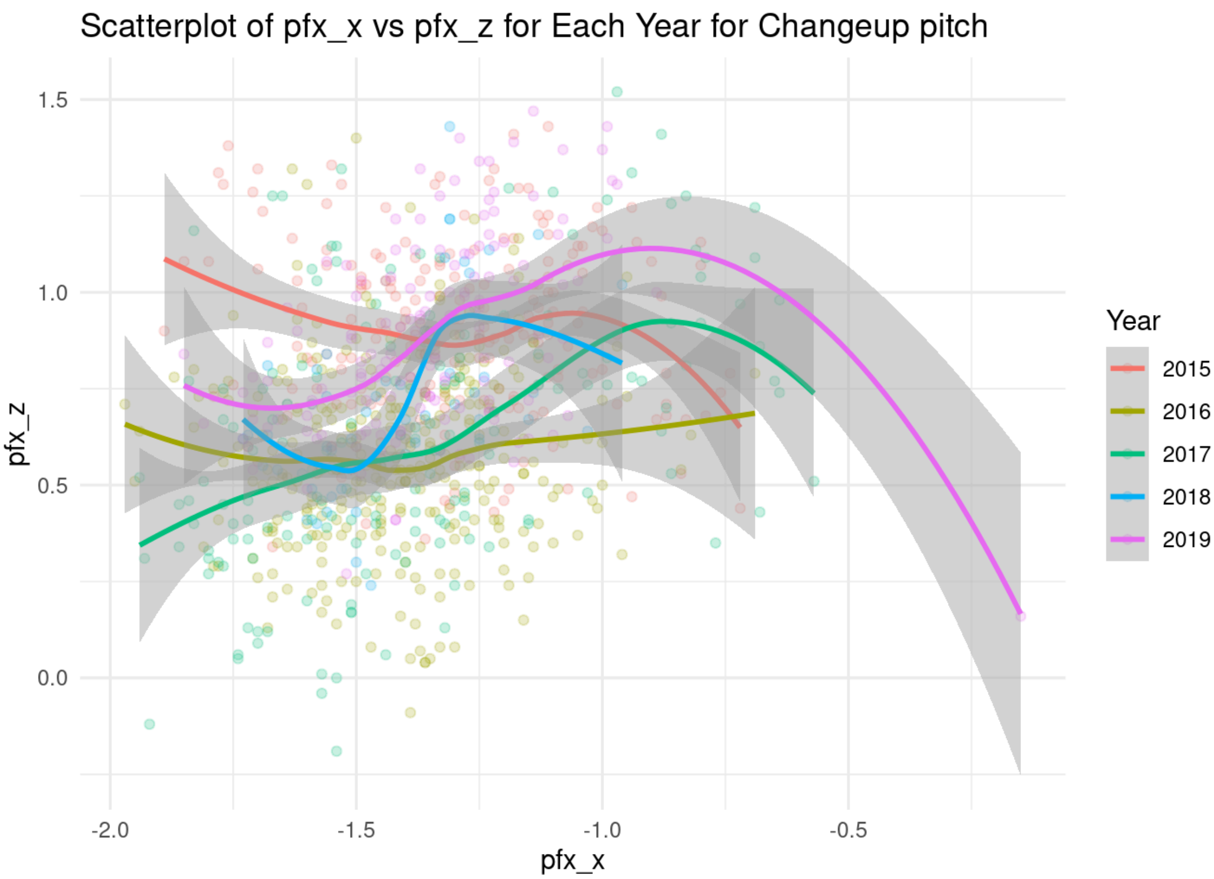
It seems that velocity in the y-dimension has changed drastically throughout the years for the slider pitch. There is a lot of variation in how fast Verlander throws the slider pitch throughout each year. Similarly, acceleration in the y-dimension has changed significantly for the slider pitch. The inconsistency in this dimension is understandable since each slider pitch may be thrown at a different speed from each other.



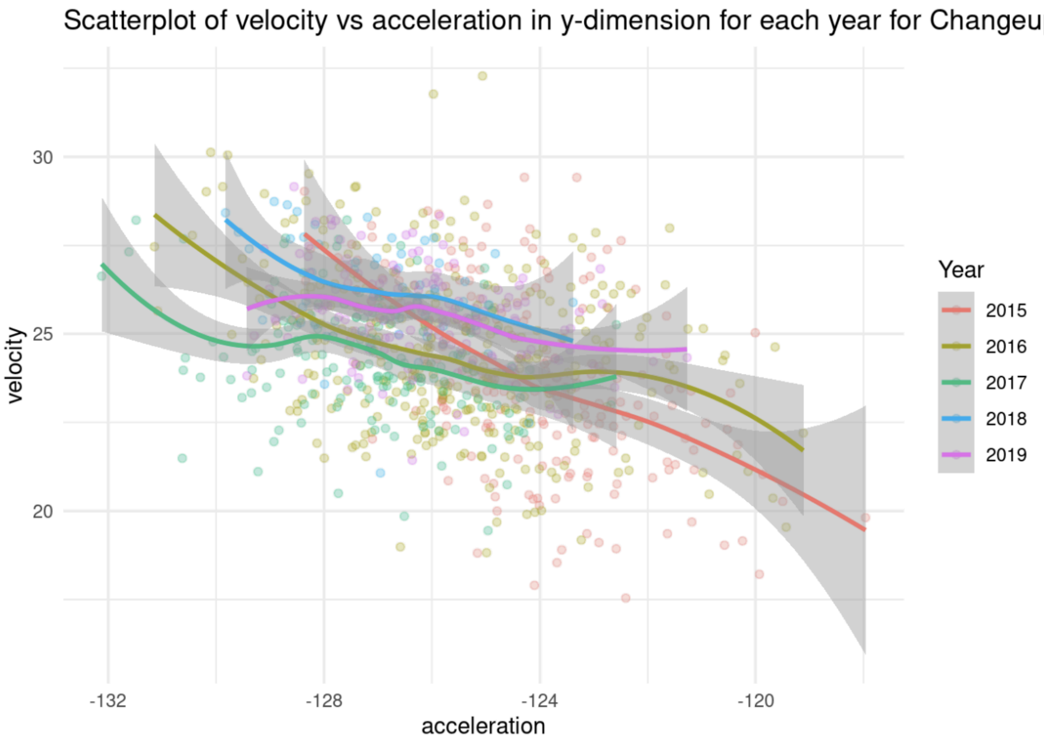
Verlander's sliders landed more in zones 4, 7, 11, 12, and 13, while falling less frequently in zone 14. By 2019, however, Verlander's sliders landed more evenly throughout every zone, except for landing a little more frequently in zone 14 and less frequently in zone 1. In general, it seems that sliders in 2015 are likely to be in the left-most zones (4, 7, 11, 13) which is understandable since Verlander is a right-handed pitcher. However, in 2019, sliders were less likely to fall in zones 1 and 11, and slightly more likely to fall in zone 14. This shows a difference in his sliders as they shifted more to the right as time progressed.

Changeup Pitches

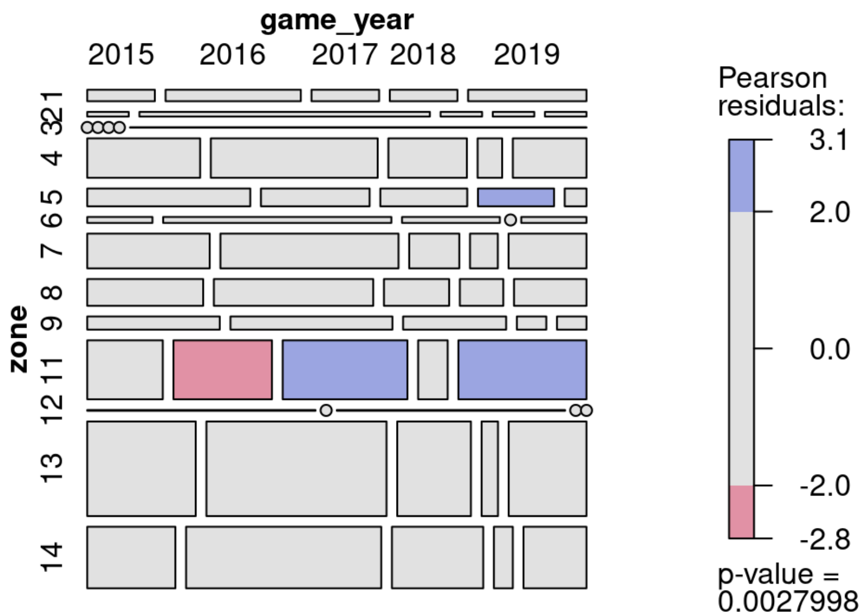
A changeup is a type of pitch in baseball that is thrown to look like a fastball but arrives at the batter much slower. This difference in speed is intended to deceive the batter, making them swing early. The changeup often has additional movement, like sinking or fading away from the hitter, which adds to its effectiveness. It is called a changeup pitch due to the fact that it is used in place of a fastball pitch, as a way to throw the batter off.



In the first three years, Verlander's change up pitches have had similar spread in the horizontal direction. In 2019, it seems that there is a lot more spread in the pitches, but that is due to outliers. In the vertical direction, however, there seems to be much more changes. In 2016-2018, it seems he had more variation in vertical movement, which was on average lower than 2015 and 2019. In 2019, he seemed to have vertical movement closer to 2015; this may show that he tried to change his change up pitch style, but resorted to a similar style that he started with.



Velocity in the y-direction seems to have become more consistent in 2017 - 2019 due to the smaller slope compared to earlier years. There are significant differences in velocity for 2015 vs all the other years, which shows that Verlander's change up pitch has changed significantly from the beginning. However, it seems that there is no significant difference between 2016 vs 2019, and 2017 vs 2018. It seems that overall, there has not been a vast variation in the velocity of pitches from year to year. This may show that Verlander has found a good velocity in the y-dimension for change up pitches since he seems to have a similar style on average.



Finally, Verlander's change-ups didn't see many noticeable changes. However, he did pitch less in zone 11 in 2016, while pitching more frequently in zone 11 in 2017 and 2019. His changeups also fell in zone 5 more often in 2019.

Methodology

Wilcox rank sum test is a nonparametric test that ranks the absolute difference between the two groups. The smallest difference between the two groups is ranked the lowest and if the difference number is negative, apply the negative sign back to the rank. If there are duplicates, you take the mean of the rank numbers. From the ranks, we will add the negative numbers to obtain a negative rank sum and do the same for the positive numbers too. To determine the test statistic, using a reference table for the critical value. Based on the calculated total rank for negative and positive, if it is larger than the critical value, then we fail to reject the null hypothesis.

Kruskal Wallis is a nonparametric test that determines whether at least two groups differ from each other. The test is based on ranking the entire dataset between the different variables from 2015 to 2019. Once numbers are ranked, sum up the ranks based on each year, take the mean of the rank rum. Find the expected value of the rankings which is (n+1)/2, then test statistics to determine what results are significant. For the test statistics is also the chi square statistics, so compute this value and compare it to the chi square distribution tables. Compared to a significance level of 5%, if chi square < critical value, then we fail to reject the null hypothesis and it is not significant.

The Pearson Chi-Squared test is a goodness-of-fit test used to evaluate how likely it is to obtain the observed data assuming independence. It follows the null hypothesis that the data is independent. The test is conducted by first computing the expected counts for each category. Then, take the squared difference of the observed counts and the expected counts and divide by the expected count. The sum of this quantity for each category becomes the chi-squared statistic. This means large Pearson residuals correlate with low p-values.

Results and Discussion

For fastballs, the Kruskal-Wallis p-values indicated that for acceleration and velocity across all axes, as well as the horizontal and vertical displacements, there were differences across the five-year periods. The Wilcox rank-sum test further concludes that for almost every pair across these variables, the pairwise differences of fastball velocity, acceleration, and displacement were significant. Finally, the pairwise differences were most significant when comparing the end of the period (2019) to the beginning few years of the period (2015, 2016).

Based on the Kruskal-Wallis p-values, we can conclude that there is indeed a difference in both horizontal and vertical movement in his curve balls across the five-year period. Specifically, 2015 vs 2017 and 2018 have non-significant differences in the horizontal direction based on the Wilcox rank sum test. In the vertical direction, however, all pairwise comparisons are significant.

Based on the Kruskal Wallis test p-values, we can conclude that indeed there is a difference in both horizontal and vertical movement between the five years for the slider pitch. Specifically, only 2015 vs 2016 and 2017 have non-significant differences in the horizontal direction based on the Wilcox rank sum test. Only 2017 and 2019 have non-significant differences in the vertical direction.

The Kruskal-Wallis results show that there were differences in both horizontal and vertical movement throughout the years. The p-value for horizontal movement was smaller than that for vertical movement, which supports the idea that there was less change in the horizontal movement for this pitch type.

According to the Kruskal-Wallis test, there is no significant difference in zone by game year for the changeup or curveball pitches; each year shows a similar zone frequency for this pitch type.

Zone type seems to have changed for slider pitches over the five-year period; however, this was a slow progression as there is little significant difference in zone frequency between subsequent years. In fact, the most significant difference was 2015 vs 2019 which shows that there were slight changes in zone frequency for the slider pitch over five years.

Wilcox rank sum is to determine what value is significant to his pitching between the different years. For FASTBALLS, compared to the p value of 0.05, the most notable change is regarding zones, Verlander does not focus much on location of the pitch when it crosses the plate.

Conclusion

We can see that across these five years, Verlander's fastball pitches have changed the most. The displacement of his fastballs become much more consistent and condensed over time. Additionally, although the exact trends of his y and z velocity and acceleration may differ, they both have in common that in 2015 his pitch velocity and acceleration were very spread out. By 2019, they were much more consistent, with a tighter spread of both velocity and acceleration for both y and z axes. Verlander's fastballs also moved from the bottom right corner to further center and left.

Across the five years, there has been a change to a drop curveball instead of a sideways curveball based on position. In the x-dimension, there has been significant changes in acceleration which shows the shift away from a horizontal-movement ball. As for zones, there was very little difference in zone frequency over the five year period.

There is more variation in the horizontal movement than vertical movement over the five-year period, which shows Verlander may favor lateral breaking. According to both Pearson residuals and Kruskal-Wallis test, sliders had significant change in zone frequency over the five year period. Specifically, sliders were frequently in the left-most zones in 2015, but in 2019 sliders were less likely to fall in zones 1 and 11, and slightly more likely to fall in zone 14. This shows a difference in his sliders as they shifted more to the right as time progressed.

Due to the sporadic nature of the changeup pitch, it is hard to analyze trends. From year to year, we can see various changes in position and acceleration. However, there are no significant differences in zone frequency for this pitch type due to the large Kruskal-Wallis p-value.

Appendix

Code Appendix 1

```
library(baseballr)
library(dplyr)
library(ggplot2)
library(reshape2)
library(zoo)

#get data
player_id <- playerid_lookup("Verlander") %>%
   dplyr::filter(first_name == "Justin") %>%
   dplyr::select(mlbam_id, first_name, last_name)

player_id[1,1]
verData1 = statcast_search(start_date="2015-03-31", end_date="2015-10-31", playerid
= 434378, player_type='pitcher')

verData2 = statcast_search(start_date="2016-03-31", end_date="2016-10-31", playerid
= 434378, player_type='pitcher')

verData3 = statcast_search(start_date="2017-03-31", end_date="2017-10-31", playerid
= 434378, player_type='pitcher')

verData4 = statcast_search(start_date="2018-03-31", end_date="2018-10-31", playerid
= 434378, player_type='pitcher')

verData5 = statcast_search(start_date="2019-03-31", end_date="2019-10-31", playerid
= 434378, player_type='pitcher')
verDataFinal = rbind(verData1, verData2, verData3, verData4, verData5)

#write to csv for ease of access in future
write.table(verData1, file="verData1.csv", sep= ',')
write.table(verData2, file="verData2.csv", sep= ',')
write.table(verData3, file="verData3.csv", sep= ',')
write.table(verData4, file="verData4.csv", sep= ',')
write.table(verData5, file="verData5.csv", sep= ',')

#keep relevant variables
cols = c("pitch_type", "game_year", "pitch_name", "release_speed", "release_pos_x",
"release_pos_z", "pfx_x", "pfx_z", "vx0", "vy0", "vz0", "ax", "ay", "az",
"launch_speed", "launch_angle", "release_spin_rate", "spin_axis", "zone",
"plate_x", "plate_z")
verDataAdjusted = verDataFinal %>% filter(pitch_type!="", pitch_type!="PO",
pitch_type!="IN",
```

```
                                          pitch_type!="FC", pitch_type!="SI") #drop
pitch ypes
verDataAdjusted = verDataAdjusted %>% select(all_of(cols))

#overall pitch frequency by type
verDataCounts = verDataAdjusted %>% group_by(pitch_name, game_year) %>% count()
verDataCounts %>% ggplot(aes(x=game_year, y=n, color=pitch_name)) + geom_line()

#only get fastballs
fastballs = verDataAdjusted %>% filter(pitch_type=="FF")

#pfx_x and pfx_z
fastballs %>% ggplot(aes(x=pfx_x, y=pfx_z, color=as.factor(game_year))) +
geom_point(alpha=0.7)
kruskal.test(pfx_x~game_year, data=fastballs)
pairwise.wilcox.test(fastballs$pfx_x, fastballs$game_year,
p.adjust.method="bonferroni")

kruskal.test(pfx_z~game_year, data=fastballs)
pairwise.wilcox.test(fastballs$pfx_z, fastballs$game_year,
p.adjust.method="bonferroni")


#x velocity + acceleration
fastballs %>% ggplot(aes(x=vx0, y=ax, color=as.factor(game_year))) +
geom_point(alpha=0.7)

pairwise.wilcox.test(fastballs$vx0, fastballs$game_year,
p.adjust.method="bonferroni")
kruskal.test(vx0~game_year, data=fastballs)
pairwise.wilcox.test(fastballs$ax, fastballs$game_year,
p.adjust.method="bonferroni")
kruskal.test(ax~game_year, data=fastballs)
pairwise.wilcox.test(fastballs$ay, fastballs$game_year,
p.adjust.method="bonferroni")
kruskal.test(ay~game_year, data=fastballs)

# y velocity + acceleration
fastballs %>% ggplot(aes(x=vy0, y=ay, color=as.factor(game_year))) +
geom_point(alpha=0.7)
pairwise.wilcox.test(fastballs$vy0, fastballs$game_year,
p.adjust.method="bonferroni")
kruskal.test(vy0~game_year, data=fastballs)

#z velocity + acceleration
fastballs %>% ggplot(aes(x=vz0, y=az, color=as.factor(game_year))) +
geom_point(alpha=0.7)
```

```r
pairwise.wilcox.test(fastballs$vz0, fastballs$game_year,
p.adjust.method="bonferroni")
kruskal.test(vz0~game_year, data=fastballs)
pairwise.wilcox.test(fastballs$az, fastballs$game_year,
p.adjust.method="bonferroni")
kruskal.test(az~game_year, data=fastballs)

library(vcd)
zonesYear = xtabs(~zone+game_year, data=fastballs)
zonesYear %>% mosaic(shade=TRUE, direction="h")
kruskal.test(zone~game_year, data=fastballs)
pairwise.wilcox.test(fastballs$zone, fastballs$game_year,
p.adjust.method="bonferroni")
```

Code Appendix 2

```
knitr::opts_chunk$set(echo = TRUE)

library(dplyr)
library(ggplot2)
# Read in all of the scraped data

df2015 <- read.csv("verData1.csv")
df2016 <- read.csv("verData2.csv")
df2017 <- read.csv("verData3.csv")
df2018 <- read.csv("verData4.csv")
df2019 <- read.csv("verData5.csv")

df <- rbind(df2015, df2016, df2017, df2018, df2019)

slider <- df %>%
  filter(pitch_type == "SL") %>%
  select(pitch_type, game_date, pfx_z, pfx_x, zone, ax, ay, az, vx0, vy0, vz0) %>%
  mutate(game_year = as.factor(substr(game_date,1,4)))

curveball <- df %>%
  filter(pitch_type == "CU") %>%
  select(pitch_type, game_date, pfx_z, pfx_x, zone, ax, ay, az, vx0, vy0, vz0) %>%
  mutate(game_year = as.factor(substr(game_date,1,4)))

changeup <- df %>%
  filter(pitch_type == "CH") %>%
  select(pitch_type, game_date, pfx_z, pfx_x, zone, ax, ay, az, vx0, vy0, vz0) %>%
  mutate(game_year = as.factor(substr(game_date,1,4)))

fastball <- df %>%
  filter(pitch_type == "FF") %>%
  select(pitch_type, game_date, pfx_z, pfx_x, zone, ax, ay, az, vx0, vy0, vz0) %>%
  mutate(game_year = as.factor(substr(game_date,1,4)))
ggplot(slider, aes(x = pfx_x, y = pfx_z, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of pfx_x vs pfx_z for Each Year for Slider pitch",
       x = "pfx_x",
       y = "pfx_z",
       color = "Year") +
  theme_minimal()
kruskal.test(pfx_x~game_year, data=slider)
kruskal.test(pfx_z~game_year, data=slider)
```

```
#post hoc testing
pairwise.wilcox.test(slider$pfx_x, slider$game_year, p.adjust.method =
"bonferroni")
pairwise.wilcox.test(slider$pfx_z, slider$game_year, p.adjust.method =
"bonferroni")
ggplot(curveball, aes(x = pfx_x, y = pfx_z, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of pfx_x vs pfx_z for Each Year for Curveball pitch",
       x = "pfx_x",
       y = "pfx_z",
       color = "Year") +
  theme_minimal()
kruskal.test(pfx_x~game_year, data=curveball)
kruskal.test(pfx_z~game_year, data=curveball)

ggplot(curveball, aes(x=pfx_x, color=game_year)) +
  geom_boxplot()

ggplot(curveball, aes(x=pfx_z, color=game_year)) +
  geom_boxplot()

pairwise.wilcox.test(curveball$pfx_x, curveball$game_year, p.adjust.method =
"bonferroni")
pairwise.wilcox.test(curveball$pfx_z, curveball$game_year, p.adjust.method =
"bonferroni")
ggplot(changeup, aes(x = pfx_x, y = pfx_z, color = factor(game_year))) +
  geom_point(alpha=.2) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of pfx_x vs pfx_z for Each Year for Changeup pitch",
       x = "pfx_x",
       y = "pfx_z",
       color = "Year") +
  theme_minimal()
kruskal.test(pfx_x~game_year, data=changeup)
kruskal.test(pfx_z~game_year, data=changeup)


ggplot(changeup, aes(x=pfx_x, color=game_year)) +
  geom_boxplot()

ggplot(changeup, aes(x=pfx_z, color=game_year)) +
  geom_boxplot()

pairwise.wilcox.test(changeup$pfx_x, changeup$game_year, p.adjust.method =
```

```
"bonferroni")
pairwise.wilcox.test(changeup$pfx_z, changeup$game_year, p.adjust.method =
"bonferroni")

ggplot(slider, aes(x = vx0, y = ax, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in x-dimension for Each
Year for Slider pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vx0~game_year, data=slider)
kruskal.test(ax~game_year, data=slider)

pairwise.wilcox.test(slider$vx0, slider$game_year, p.adjust.method = "bonferroni")
pairwise.wilcox.test(slider$ax, slider$game_year, p.adjust.method = "bonferroni")
ggplot(curveball, aes(x = vx0, y = ax, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in x-dimension for Each
Year for Slider pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vx0~game_year, data=curveball)
kruskal.test(ax~game_year, data=curveball)

pairwise.wilcox.test(curveball$vx0, curveball$game_year, p.adjust.method =
"bonferroni")
pairwise.wilcox.test(curveball$ax, curveball$game_year, p.adjust.method =
"bonferroni")
ggplot(changeup, aes(x = vx0, y = ax, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in x-dimension for Each
Year for Changeup pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vx0~game_year, data=changeup)
```

```
kruskal.test(ax~game_year, data=changeup)

pairwise.wilcox.test(changeup$vx0, changeup$game_year, p.adjust.method =
"bonferroni")
pairwise.wilcox.test(changeup$ax, changeup$game_year, p.adjust.method =
"bonferroni")
ggplot(changeup, aes(x=vx0, color=game_year)) +
  geom_boxplot()

ggplot(changeup, aes(x=ax, color=game_year)) +
  geom_boxplot()
ggplot(slider, aes(x = vy0, y = ay, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in y-dimension for Each
Year for Slider pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vy0~game_year, data=slider)
kruskal.test(ay~game_year, data=slider)

pairwise.wilcox.test(slider$vy0, slider$game_year, p.adjust.method = "bonferroni")
pairwise.wilcox.test(slider$ay, slider$game_year, p.adjust.method = "bonferroni")
ggplot(slider, aes(x=vy0, color=game_year)) +
  geom_boxplot()

ggplot(slider, aes(x=ay, color=game_year)) +
  geom_boxplot()
ggplot(curveball, aes(x = vy0, y = ay, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in y-dimension for Each
Year for Curveball pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vy0~game_year, data=curveball)
kruskal.test(ay~game_year, data=curveball)

pairwise.wilcox.test(curveball$vy0, curveball$game_year, p.adjust.method =
"bonferroni")
pairwise.wilcox.test(curveball$ay, curveball$game_year, p.adjust.method =
```

```
"bonferroni")
ggplot(changeup, aes(x = vy0, y = ay, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in y-dimension for each
year for Changeup pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vy0~game_year, data=changeup)
kruskal.test(ay~game_year, data=changeup)

pairwise.wilcox.test(changeup$vy0, changeup$game_year, p.adjust.method =
"bonferroni")
pairwise.wilcox.test(changeup$ay, changeup$game_year, p.adjust.method =
"bonferroni")
ggplot(slider, aes(x = vz0, y = az, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in z-dimension for Each
Year for Slider pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vz0~game_year, data=slider)
kruskal.test(az~game_year, data=slider)

pairwise.wilcox.test(slider$vz0, slider$game_year, p.adjust.method = "bonferroni")
pairwise.wilcox.test(slider$az, slider$game_year, p.adjust.method = "bonferroni")
ggplot(curveball, aes(x = vz0, y = az, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in z-dimension for Each
Year for Curveball pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vz0~game_year, data=curveball)
kruskal.test(az~game_year, data=curveball)

pairwise.wilcox.test(curveball$vz0, curveball$game_year, p.adjust.method =
```

```
"bonferroni")
pairwise.wilcox.test(curveball$az, curveball$game_year, p.adjust.method =
"bonferroni")
ggplot(changeup, aes(x = vz0, y = az, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year),formula = y~x,
method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in z-dimension for Each
Year for Changeup pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
kruskal.test(vz0~game_year, data=changeup)
kruskal.test(az~game_year, data=changeup)

pairwise.wilcox.test(changeup$vz0, changeup$game_year, p.adjust.method =
"bonferroni")
pairwise.wilcox.test(changeup$az, changeup$game_year, p.adjust.method =
"bonferroni")
library(vcd)
zonesYear = xtabs(~zone+game_year, data=fastball)

zonesYear %>% mosaic(shade=TRUE, direction="h")
#library(vcd)
zonesYear = xtabs(~zone+game_year, data=slider)

zonesYear %>% mosaic(shade=TRUE, direction="h")

kruskal.test(zone~game_year, data=slider)

pairwise.wilcox.test(slider$zone, slider$game_year, p.adjust.method = "bonferroni")
zonesYear = xtabs(~zone+game_year, data=curveball)

zonesYear %>% mosaic(shade=TRUE, direction="h")
kruskal.test(zone~game_year, data=curveball)

pairwise.wilcox.test(curveball$zone, curveball$game_year, p.adjust.method =
"bonferroni")
zonesYear = xtabs(~zone+game_year, data=changeup)

zonesYear %>% mosaic(shade=TRUE, direction="h")
kruskal.test(zone~game_year, data=changeup)

pairwise.wilcox.test(changeup$zone, changeup$game_year, p.adjust.method =
"bonferroni")
```