

# STA 160 Final

Miriam Aguirre

2024-06-05

```
# 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")
```

Analyze four types of pitches: slider(SL), curveball(CU), changeup(CH), slurve(SL) using: Pfx\_z, pfx\_x (amount of vertical + horizontal movement by the time the ball reaches home plate) Velocity + acceleration of pitch Zone

## Data manipulation

```
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)))

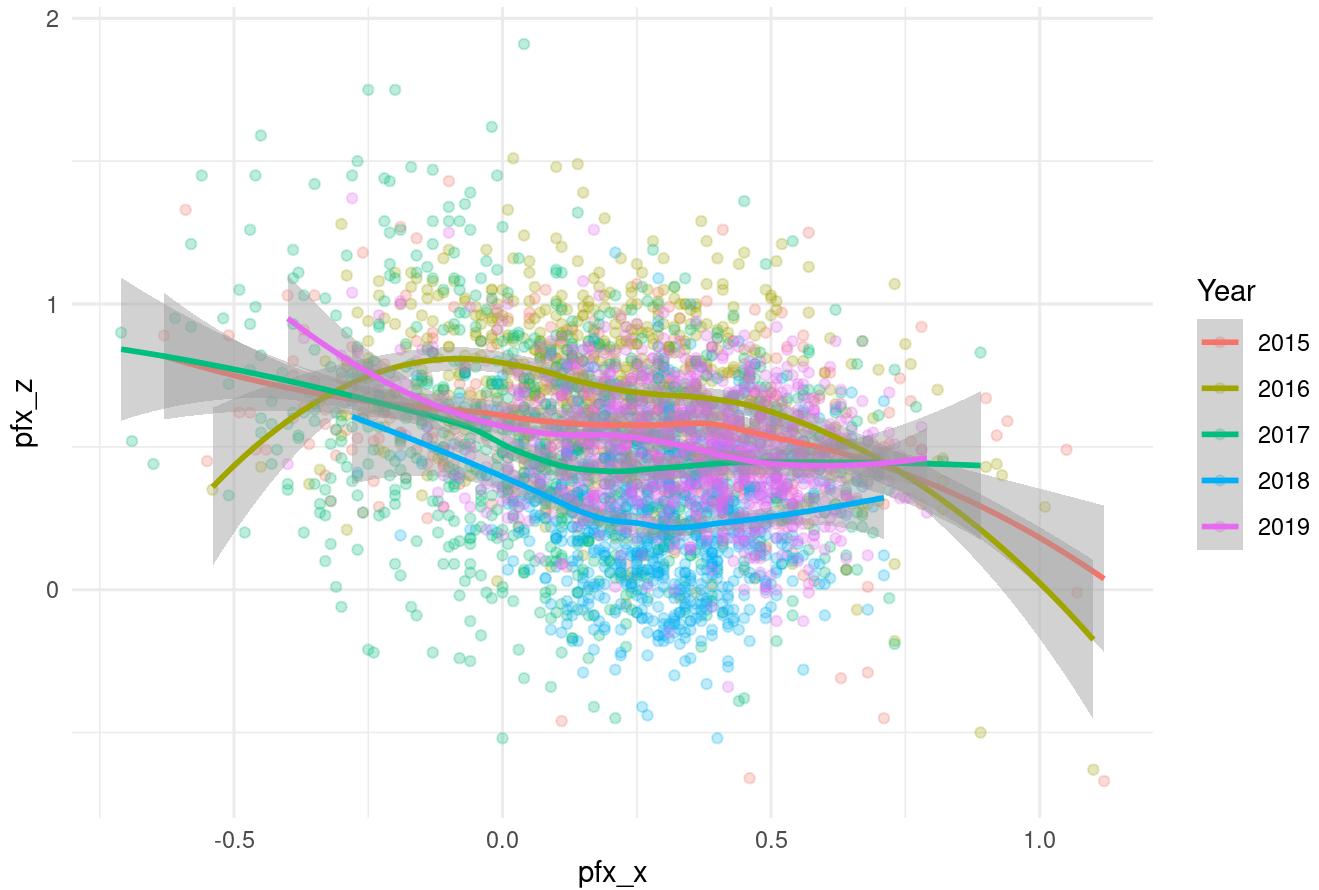
curveball <- df %>%
  filter(pitch_type == "CU") %>%
  select(pitch_type, game_date, pfx_z, pfx_x, zone, ax, ay, az, vx0, vy0, vz0) %>%
  mutate(game_year = as.factor(substr(game_date,1,4)))

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)))
```

## pfx\_x vs pfx\_z

```
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()
```

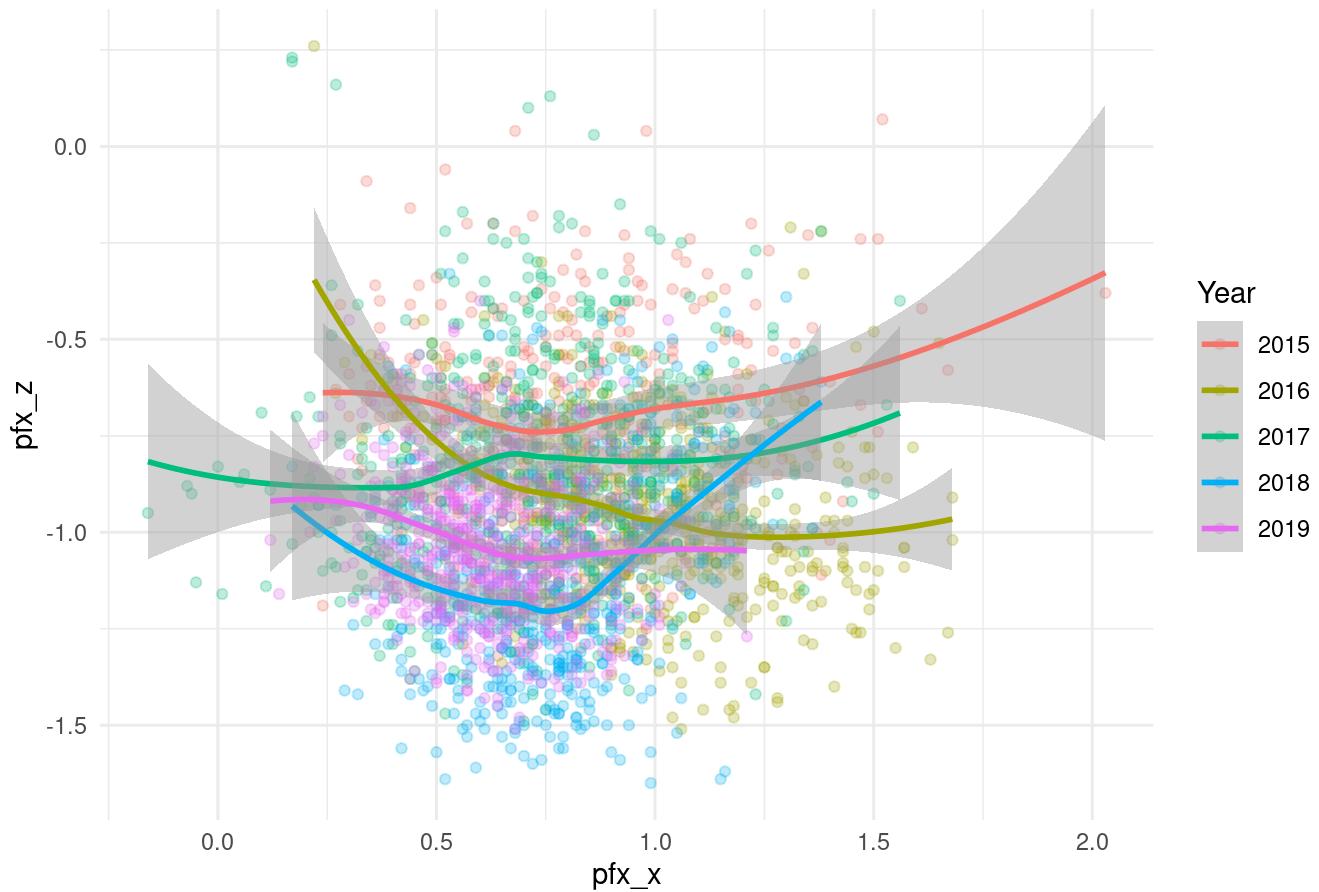
## Scatterplot of pfx\_x vs pfx\_z for Each Year for Slider pitch



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.

```
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()
```

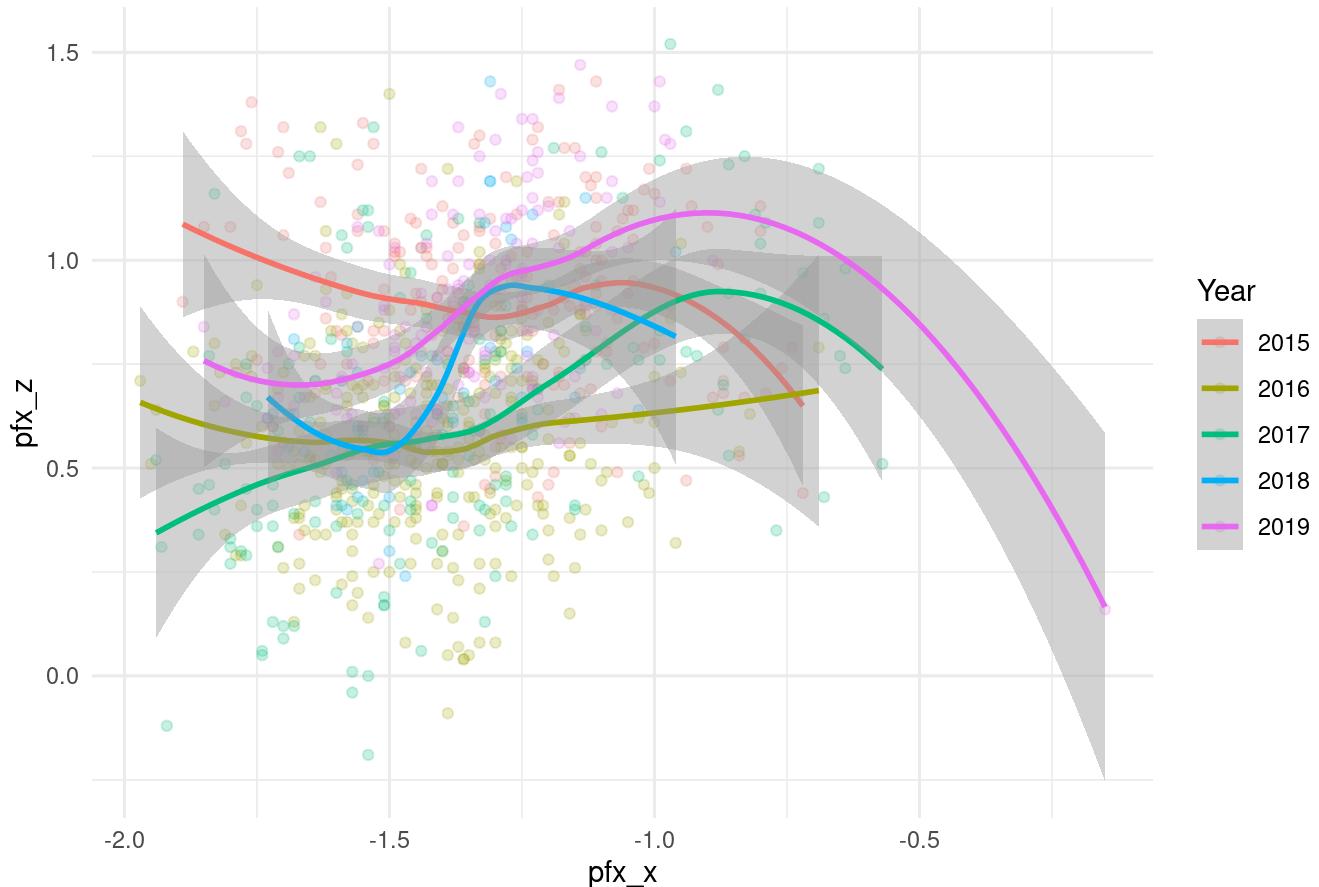
### Scatterplot of pfx\_x vs pfx\_z for Each Year for Curveball pitch



Over time, Verlander's curveball pitches have shifted more to the center of the

```
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()
```

### Scatterplot of pfx\_x vs pfx\_z for Each Year for Changeup pitch



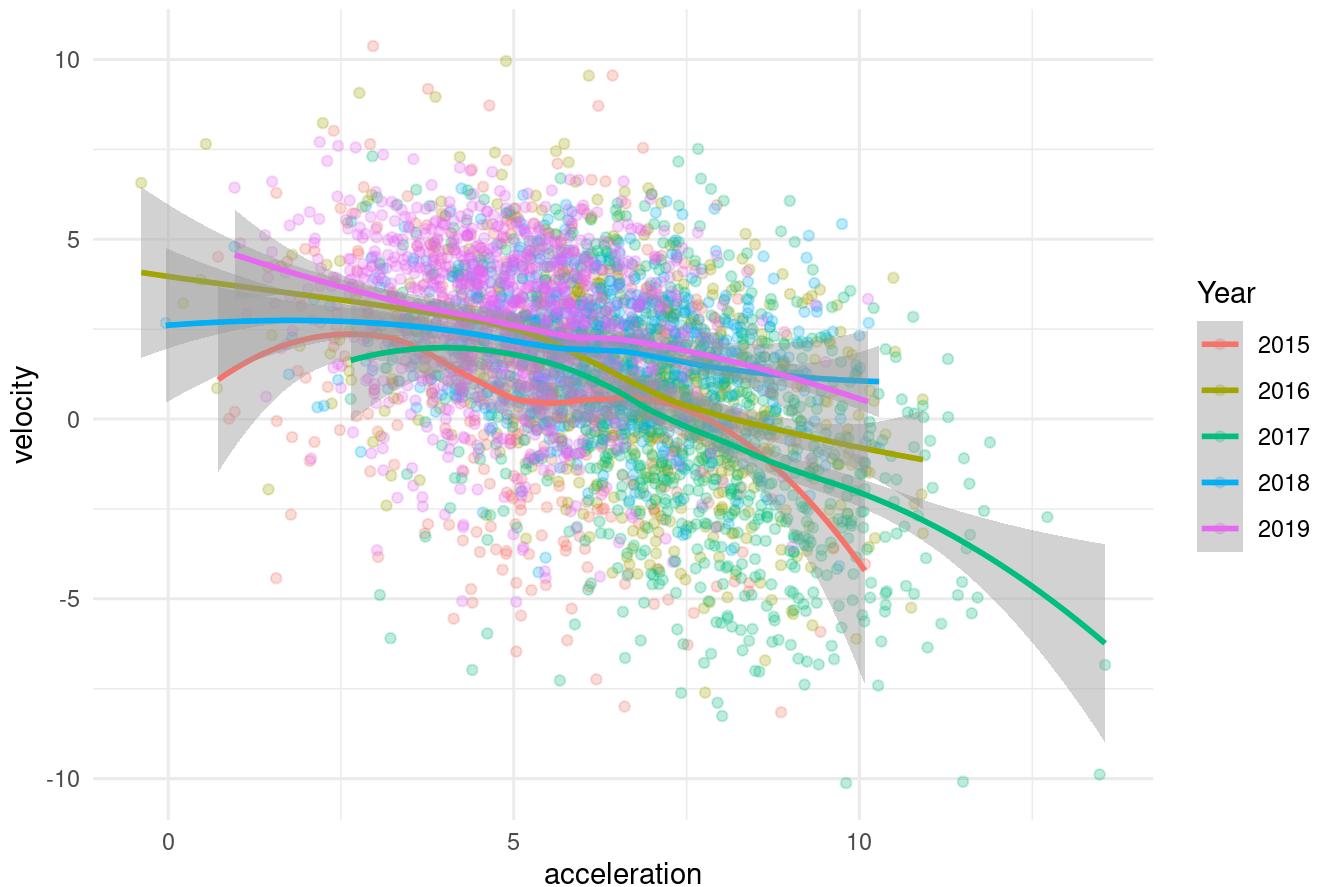
## Velocity vs acceleration

Acceleration and velocity are measured in three dimensions: x, y, and z. These scatter plots compare velocity and acceleration in the three dimensions for the three pitch types.

### x-dimension

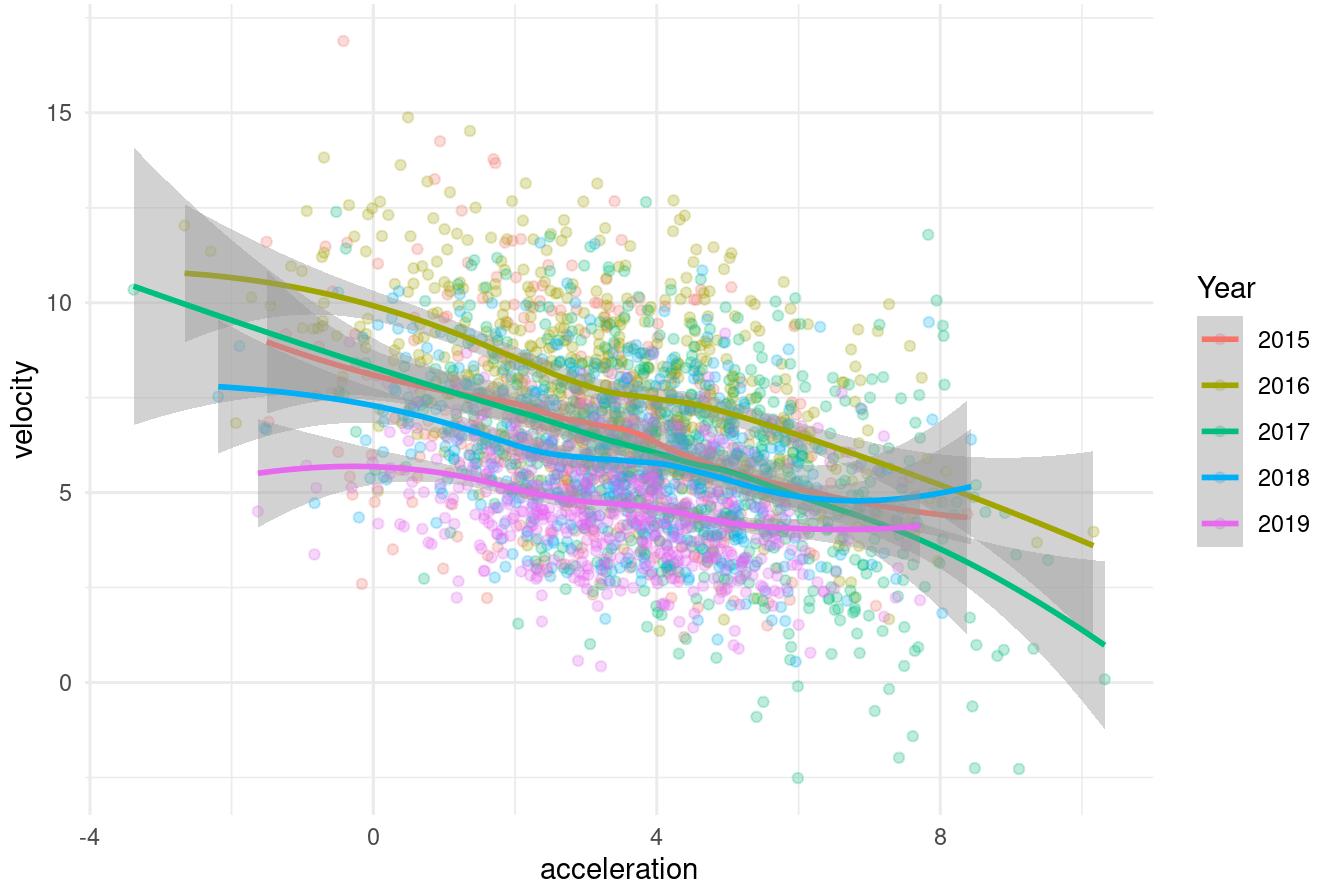
```
ggplot(sliders, aes(x = vx0, y = ax, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year), formula = y~x, method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in x-dimension for Each Year for Slider pitch",
       x =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

## Scatterplot of velocity vs acceleration in x-dimension for Each Year for Slider pitch



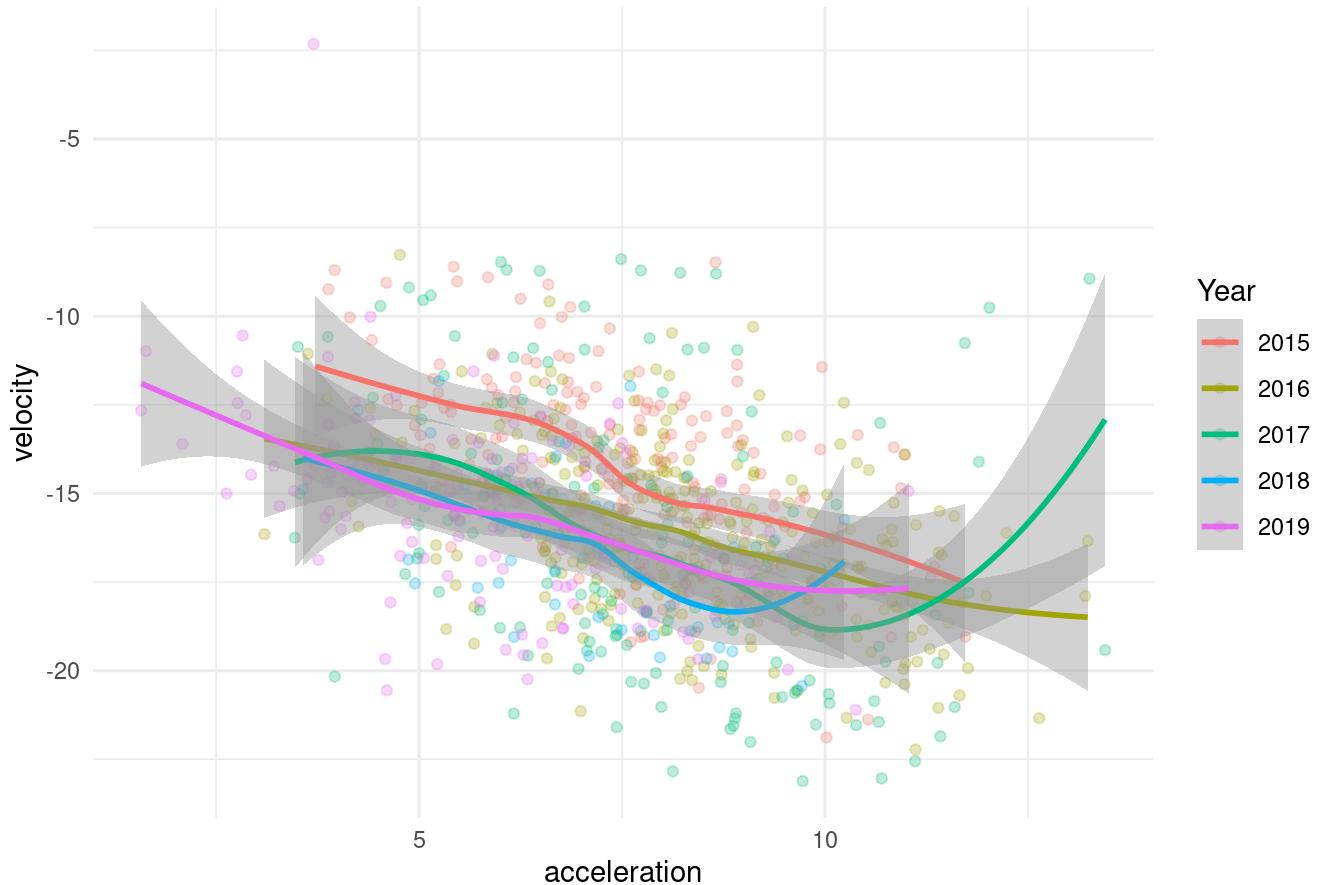
```
ggplot(curveball, aes(x = vx0, y = ax, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year), formula = y~x, method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in x-dimension for Each Year for Slider pitch",
       x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

## Scatterplot of velocity vs acceleration in x-dimension for Each Year for Slider pitch



```
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()
```

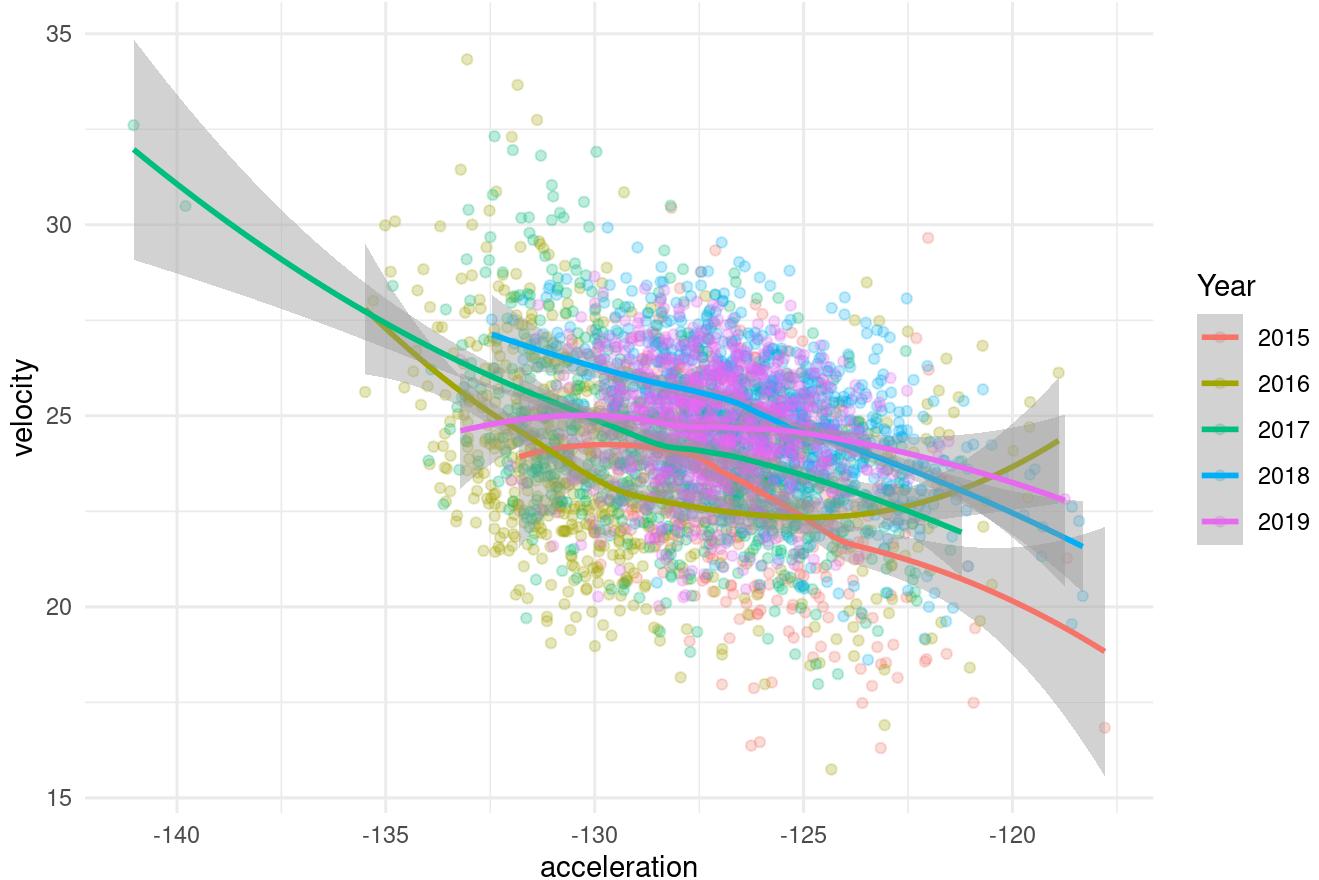
## Scatterplot of velocity vs acceleration in x-dimension for Each Year for Changeup



## y-dimension

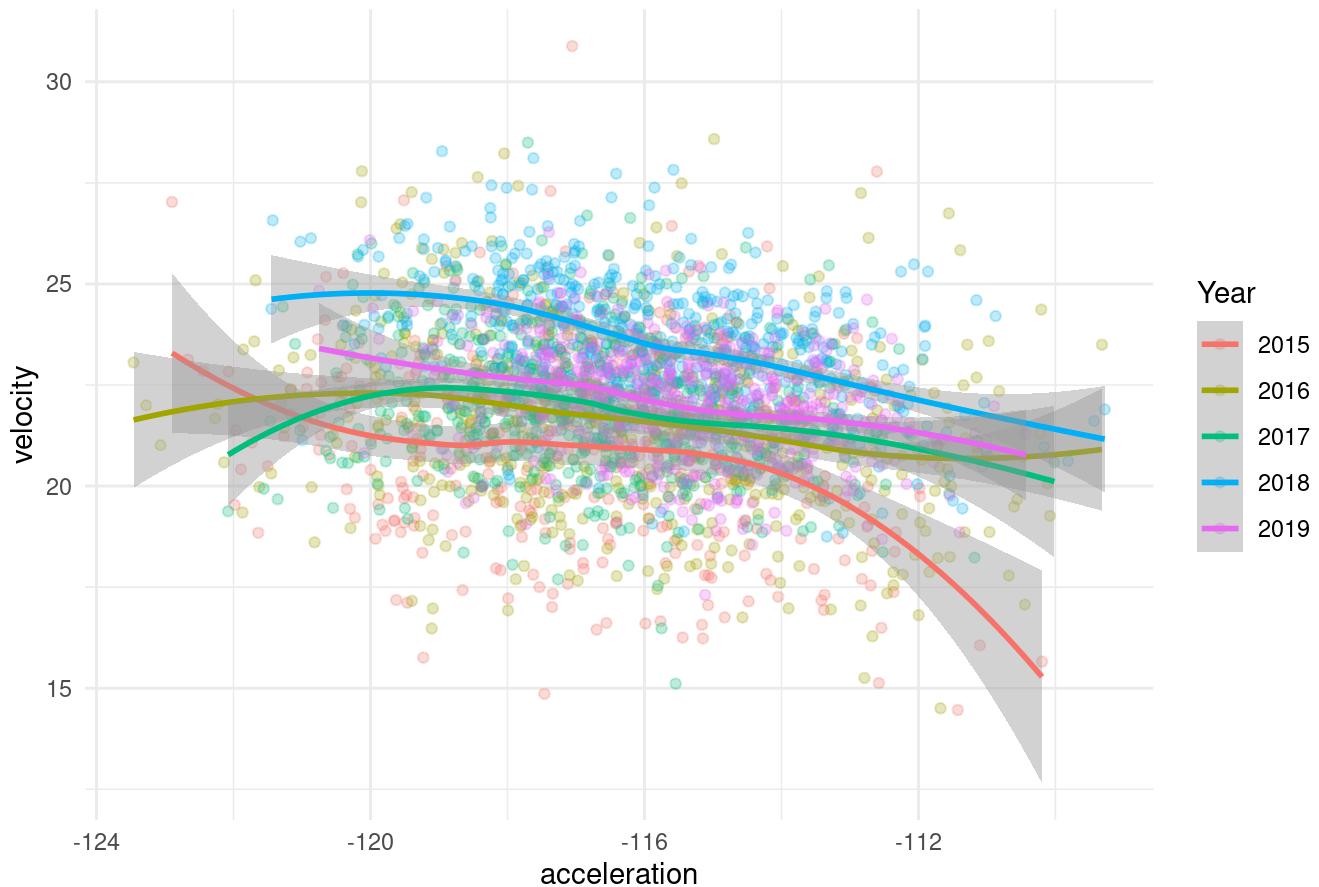
```
ggplot(sliders, aes(x = vy0, y = ay, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year), formula = y~x, method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in y-dimension for Each Year for Slider pitch",
       x =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

# Scatterplot of velocity vs acceleration in y-dimension for Each Year for Slider pitch



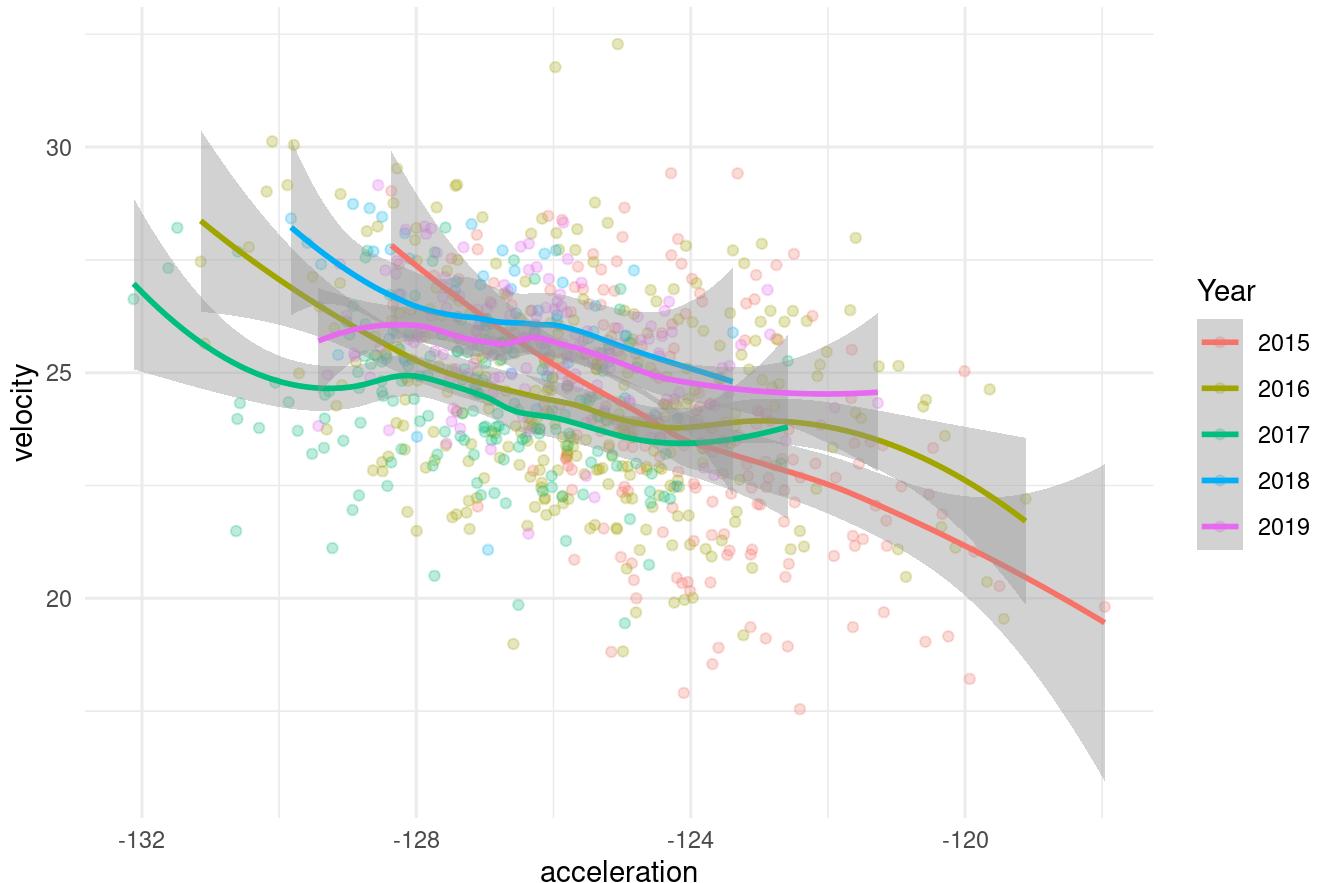
```
ggplot(curveball, aes(x = vy0, y = ay, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year), formula = y~x, method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in y-dimension for Each Year for Curveball pitch",
     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

## Scatterplot of velocity vs acceleration in y-dimension for Each Year for Curveball



```
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()
```

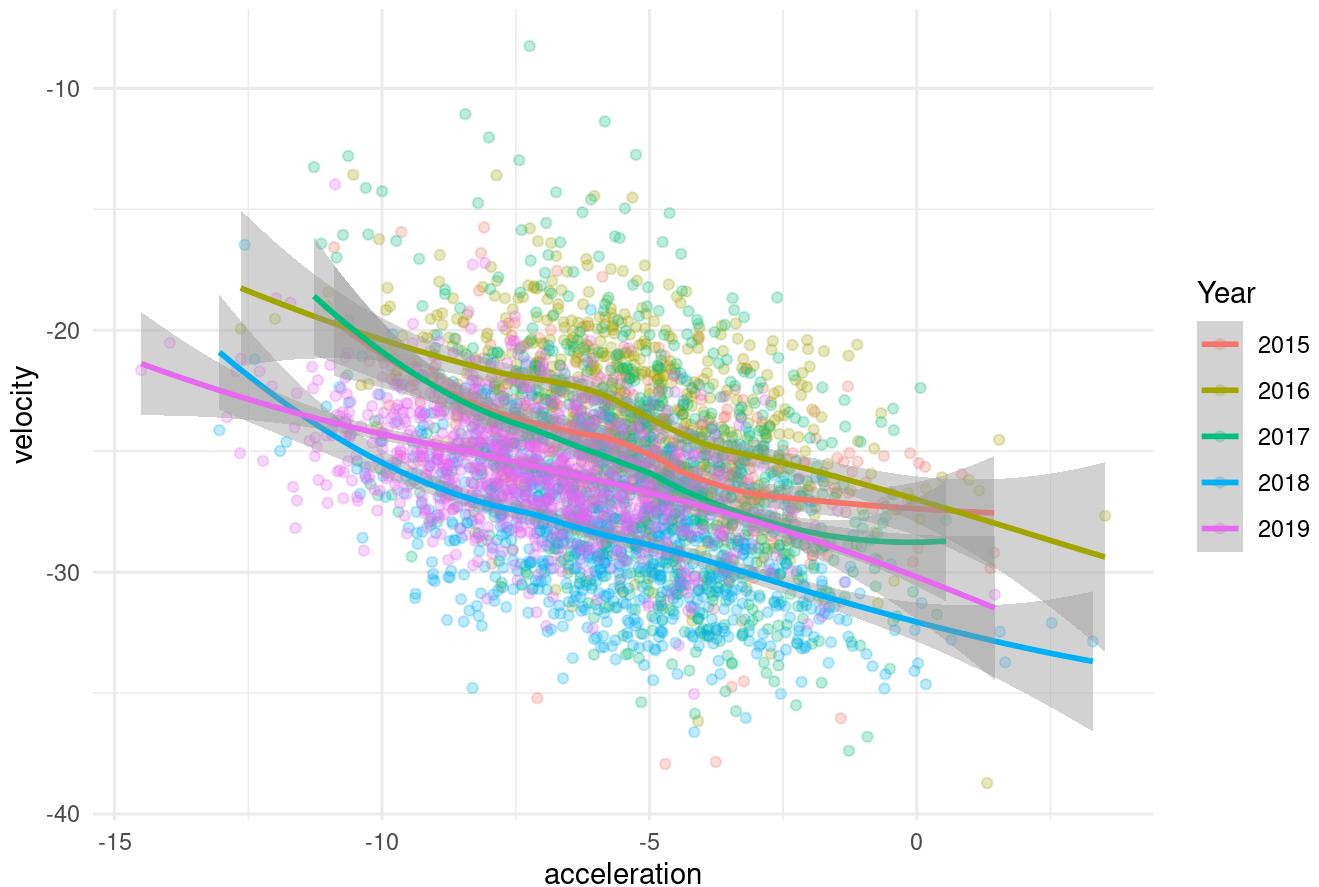
### Scatterplot of velocity vs acceleration in y-dimension for each year for Changeup



### z-dimension

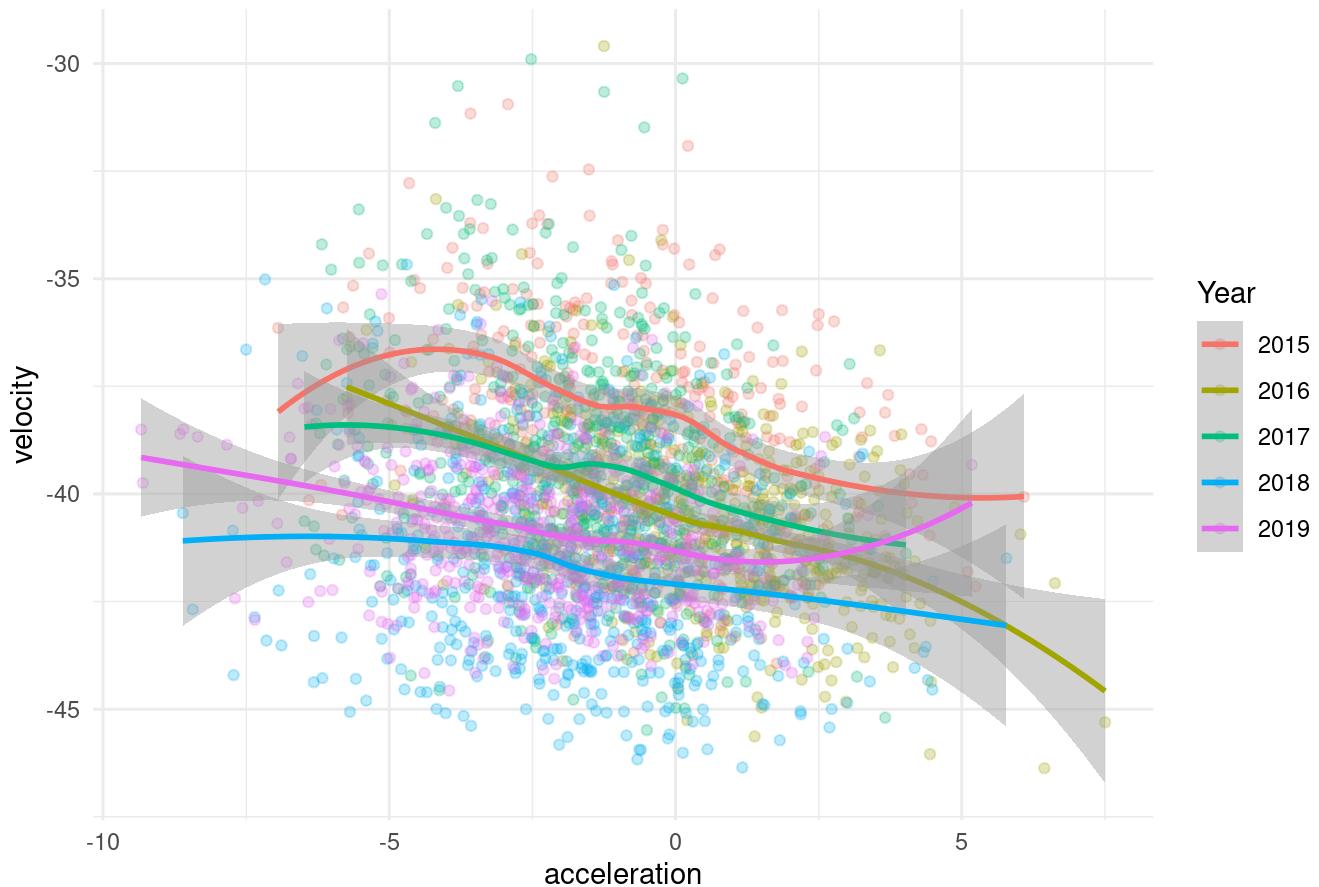
```
ggplot(sliders, aes(x = vz0, y = az, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year), formula = y~x, method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in z-dimension for Each Year for Slider pitch",
       x = "acceleration",
       y = "velocity",
       color = "Year") +
  theme_minimal()
```

## Scatterplot of velocity vs acceleration in z-dimension for Each Year for Slider pit



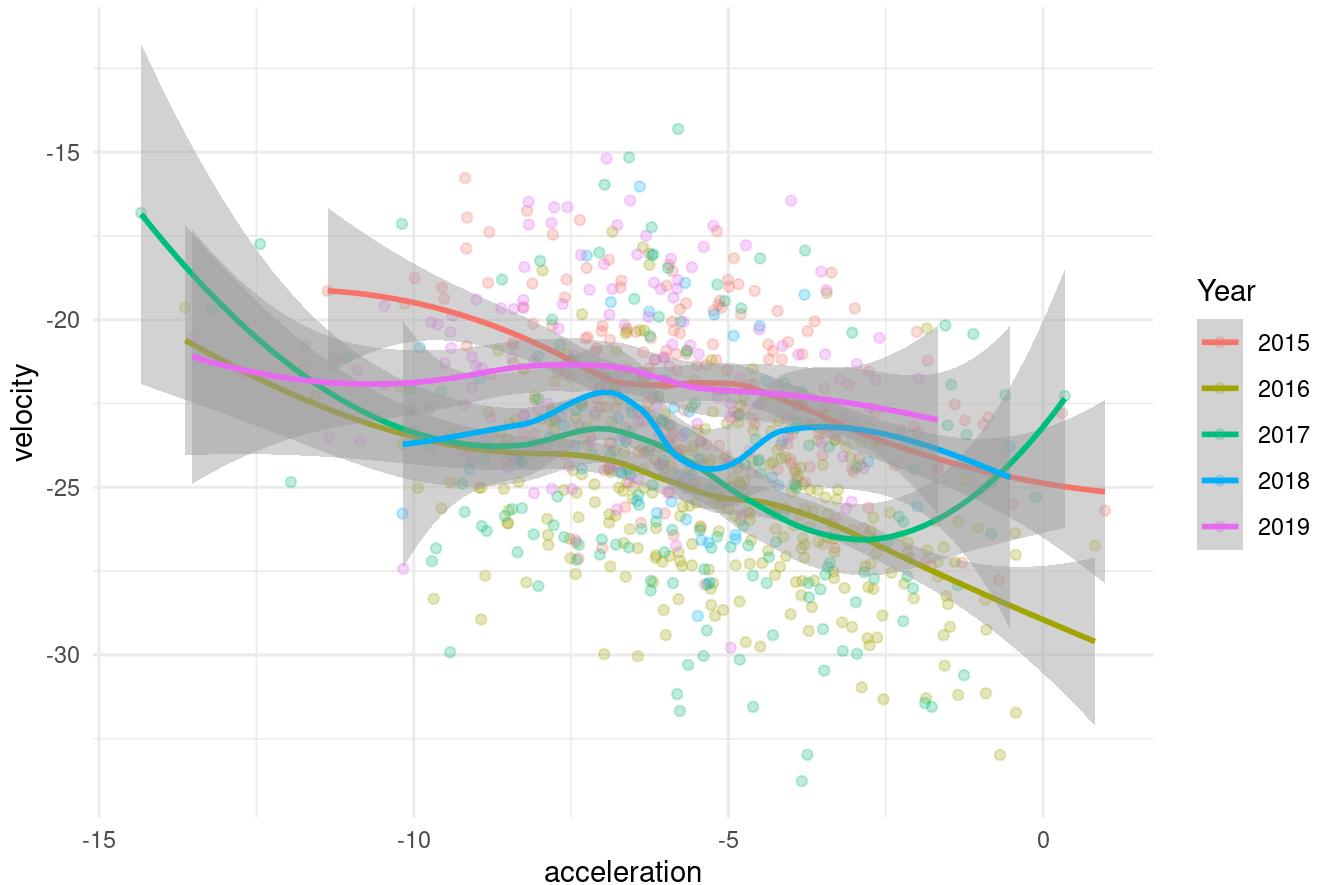
```
ggplot(curveball, aes(x = vz0, y = az, color = factor(game_year))) +
  geom_point(alpha=.25) +
  stat_smooth(aes(group = game_year, color = game_year), formula = y~x, method='loess')+
  labs(title = "Scatterplot of velocity vs acceleration in z-dimension for Each Year for Curveball pitch",
     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

### Scatterplot of velocity vs acceleration in z-dimension for Each Year for Curveball



```
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()
```

## Scatterplot of velocity vs acceleration in z-dimension for Each Year for Changeup



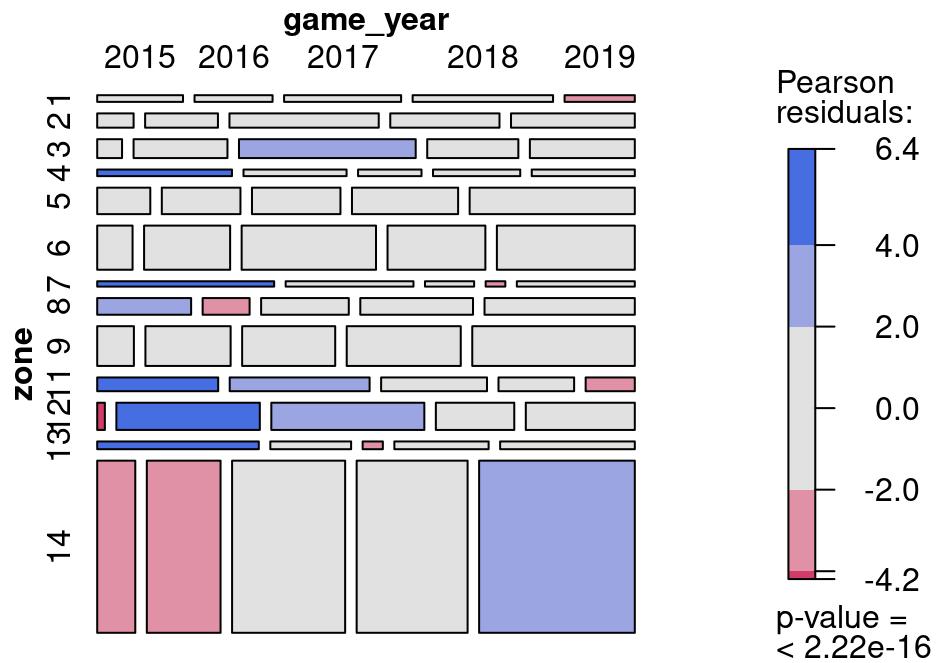
## Zone

```
library(vcd)
```

```
## Loading required package: grid
```

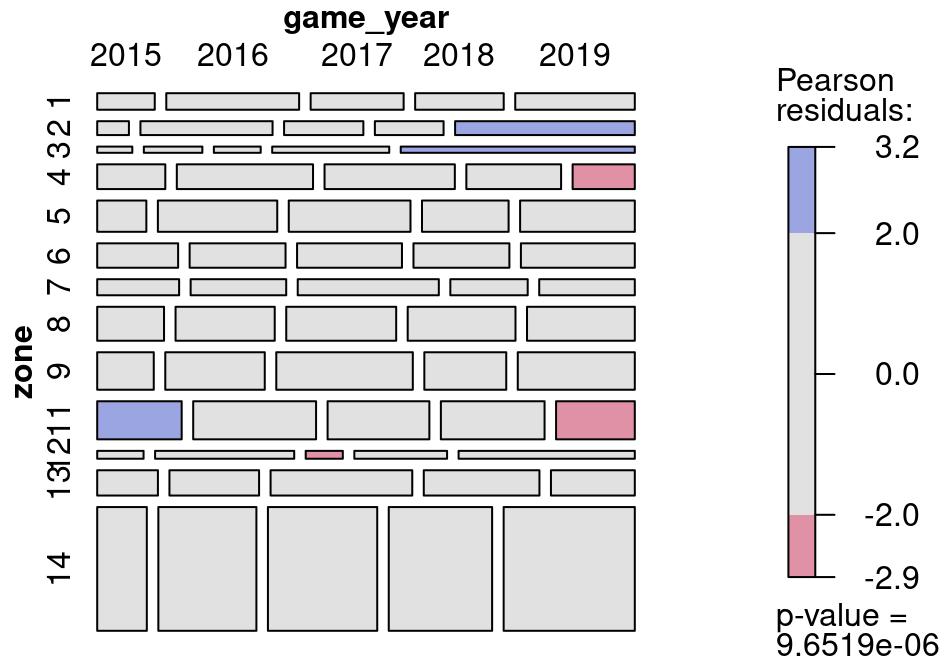
```
zonesYear = xtabs(~zone+game_year, data=slider)
```

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



```
zonesYear = xtabs(~zone+game_year, data=curveball)
```

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



```
zonesYear = xtabs(~zone+game_year, data=slider)
```

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

