MLB New Ball

Setup

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(ggthemes)
library(knitr)
tinytex::install_tinytex()
```

```
## The directory /usr/local/bin is not writable. I recommend that you make it writable. See https://gitl
```

```
new_ball <- read_csv("new_ball_data.csv")
```

```
## New names:
## Rows: 8449 Columns: 92
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (16): pitch_type, player_name, events, description, des, game_type, sta... dbl
## (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,... lgl
## (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len... date
## (1): game_date
```

```
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `pitcher` -> `pitcher...8`
## * `fielder_2` -> `fielder_2...42`
## * `pitcher` -> `pitcher...60`
## * `fielder_2` -> `fielder_2...61`
```

```
new_ball_clean <- new_ball %>% select(outcome = events, bb_type, distance = hit_distance_sc, exit_veloc
```

Filter data into outcomes

```
new_ball_clean %>% group_by(outcome) %>%
  summarize(number =n())
```

```
## # A tibble: 8 x 2
##   outcome             number
##   <chr>                <int>
## 1 double                1100
## 2 double_play              8
## 3 field_out             1913
## 4 home_run              5078
## 5 sac_fly                110
## 6 sac_fly_double_play      1
## 7 single                  58
## 8 triple                 181
```

Importing

```
fb_2015 <- read_csv("2015_fb_375.csv")
```

```
## New names:
## Rows: 3776 Columns: 92
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (17): pitch_type, player_name, events, description, des, game_type, sta... dbl
## (66): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,... lgl
## (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len... date
## (1): game_date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `pitcher` -> `pitcher...8`
## * `fielder_2` -> `fielder_2...42`
## * `pitcher` -> `pitcher...60`
## * `fielder_2` -> `fielder_2...61`
```

```
fb_2016 <- read_csv("2016_fb_375.csv")
```

```
## New names:
## Rows: 4935 Columns: 92
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (17): pitch_type, player_name, events, description, des, game_type, sta... dbl
## (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,... lgl
## (7): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len... date
## (1): game_date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `pitcher` -> `pitcher...8`
## * `fielder_2` -> `fielder_2...42`
## * `pitcher` -> `pitcher...60`
## * `fielder_2` -> `fielder_2...61`
```

```
fb_2017 <- read_csv("2017_fb_375.csv")
```

```
## New names:
## Rows: 6171 Columns: 92
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (16): pitch_type, player_name, events, description, des, game_type, sta... dbl
## (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,... lgl
## (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len... date
## (1): game_date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `pitcher` -> `pitcher...8`
## * `fielder_2` -> `fielder_2...42`
## * `pitcher` -> `pitcher...60`
## * `fielder_2` -> `fielder_2...61`
```

```
fb_2018 <- read_csv("2018_fb_375.csv")
```

```
## New names:
## Rows: 5987 Columns: 92
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (16): pitch_type, player_name, events, description, des, game_type, sta... dbl
## (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,... lgl
## (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len... date
## (1): game_date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `pitcher` -> `pitcher...8`
## * `fielder_2` -> `fielder_2...42`
## * `pitcher` -> `pitcher...60`
## * `fielder_2` -> `fielder_2...61`
```

```
fb_2019 <- read_csv("2019_fb_375.csv")
```

```
## New names:
## Rows: 7013 Columns: 92
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (16): pitch_type, player_name, events, description, des, game_type, sta... dbl
## (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,... lgl
## (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len... date
## (1): game_date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `pitcher` -> `pitcher...8`
## * `fielder_2` -> `fielder_2...42`
## * `pitcher` -> `pitcher...60`
## * `fielder_2` -> `fielder_2...61`
```

```
fb_2020 <- read_csv("2020_fb_375.csv")
```

```
## New names:
## Rows: 2368 Columns: 92
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (16): pitch_type, player_name, events, description, des, game_type, sta... dbl
## (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,... lgl
## (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len... date
## (1): game_date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `pitcher` -> `pitcher...8`
## * `fielder_2` -> `fielder_2...42`
## * `pitcher` -> `pitcher...60`
## * `fielder_2` -> `fielder_2...61`
```

```
fb_2021 <- read_csv("2021_fb_375.csv")
```

```
## New names:
## Rows: 6902 Columns: 92
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (16): pitch_type, player_name, events, description, des, game_type, sta... dbl
## (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,... lgl
## (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len... date
## (1): game_date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `pitcher` -> `pitcher...8`
## * `fielder_2` -> `fielder_2...42`
## * `pitcher` -> `pitcher...60`
## * `fielder_2` -> `fielder_2...61`
```

Cleaning

```
fb_2015_clean <- fb_2015 %>% select(player_name, outcome = events, bb_type, distance = hit_distance_sc,
fb_2016_clean <- fb_2016 %>% select(player_name, outcome = events, bb_type, distance = hit_distance_sc,
fb_2017_clean <- fb_2017 %>% select(player_name, outcome = events, bb_type, distance = hit_distance_sc,
fb_2018_clean <- fb_2018 %>% select(player_name, outcome = events, bb_type, distance = hit_distance_sc,
fb_2019_clean <- fb_2019 %>% select(player_name, outcome = events, bb_type, distance = hit_distance_sc,
fb_2020_clean <- fb_2020 %>% select(player_name, outcome = events, bb_type, distance = hit_distance_sc,
fb_2021_clean <- fb_2021 %>% select(player_name, outcome = events, bb_type, distance = hit_distance_sc,
```

Mean dataframes

```
fb_2015_mean <- fb_2015 %>% select(distance = hit_distance_sc, exit_velocity = launch_speed, launch_angl
fb_2016_mean <- fb_2016 %>% select(distance = hit_distance_sc, exit_velocity = launch_speed, launch_angl
fb_2017_mean <- fb_2017 %>% select(distance = hit_distance_sc, exit_velocity = launch_speed, launch_angl
fb_2018_mean <- fb_2018 %>% select(distance = hit_distance_sc, exit_velocity = launch_speed, launch_angl
fb_2019_mean <- fb_2019 %>% select(distance = hit_distance_sc, exit_velocity = launch_speed, launch_angl
fb_2020_mean <- fb_2020 %>% select(distance = hit_distance_sc, exit_velocity = launch_speed, launch_angl
fb_2021_mean <- fb_2021 %>% select(distance = hit_distance_sc, exit_velocity = launch_speed, launch_angl
```

Finding averages

```
mean(fb_2015_clean$distance)
```

```
## [1] 399.3729
```

```
mean(fb_2016_clean$distance)
```

```
## [1] 401.2898
```

```
mean(fb_2017_clean$distance)
```

```
## [1] 402.8848
```

```
mean(fb_2018_clean$distance)
```

```
## [1] 400.556
```

```
mean(fb_2019_clean$distance)
```

```
## [1] 403.2912
```

```
mean(fb_2020_clean$distance)
```

```
## [1] 403.272
```

```
mean(fb_2021_clean$distance)
```

```
## [1] 402.0804
```

Putting means into columns

```
colMeans(fb_2015_mean)
```

```
##      distance exit_velocity  launch_angle           xBA
##   399.3728814   102.1717426    29.6295021     0.6539465
```

```
colMeans(fb_2016_mean)
```

```
##      distance exit_velocity  launch_angle           xBA
##   401.2897670   102.5002837    29.3922999     0.6971422
```

```
colMeans(fb_2017_mean)
```

```
##      distance exit_velocity  launch_angle           xBA
##   402.8847837   102.5373683    28.6608329     0.7260269
```

```
colMeans(fb_2018_mean)
```

```
##      distance exit_velocity  launch_angle           xBA
##    400.556038    103.054084     28.396860      0.707097
```

```
colMeans(fb_2019_mean)
```

```
##      distance exit_velocity  launch_angle           xBA
##   403.2911735   103.1214316    28.6164266     0.7415557
```

```
colMeans(fb_2020_mean)
```

```
##      distance exit_velocity  launch_angle           xBA
##   403.2719595   102.9128378    30.0819257     0.7157061
```

```
colMeans(fb_2021_mean)
```

```
##      distance exit_velocity  launch_angle           xBA
##    402.080411    103.615575     29.123877      0.713156
```

Adding new mean columns to datasets

```
year <- c(2015)

fb_2015_clean$year <- year

fb_2015_avg <- fb_2015_clean %>% group_by(year) %>% summarise(mean_distance = mean(fb_2015_clean$distanc
```

```
year <- c(2016)

fb_2016_clean$year <- year

fb_2016_avg <- fb_2016_clean %>% group_by(year) %>% summarise(mean_distance = mean(fb_2016_clean$distanc
```

```
year <- c(2017)

fb_2017_clean$year <- year

fb_2017_avg <- fb_2017_clean %>% group_by(year) %>% summarise(mean_distance = mean(fb_2017_clean$distanc
```

```
year <- c(2018)

fb_2018_clean$year <- year

fb_2018_avg <- fb_2018_clean %>% group_by(year) %>% summarise(mean_distance = mean(fb_2018_clean$distanc
```

```
year <- c(2019)

fb_2019_clean$year <- year

fb_2019_avg <- fb_2019_clean %>% group_by(year) %>% summarise(mean_distance = mean(fb_2019_clean$distanc
```

```
year <- c(2020)

fb_2020_clean$year <- year

fb_2020_avg <- fb_2020_clean %>% group_by(year) %>% summarise(mean_distance = mean(fb_2020_clean$distanc
```

```r
year <- c(2021)

fb_2021_clean$year <- year

fb_2021_avg <- fb_2021_clean %>% group_by(year) %>% summarise(mean_distance = mean(fb_2021_clean$distanc
```

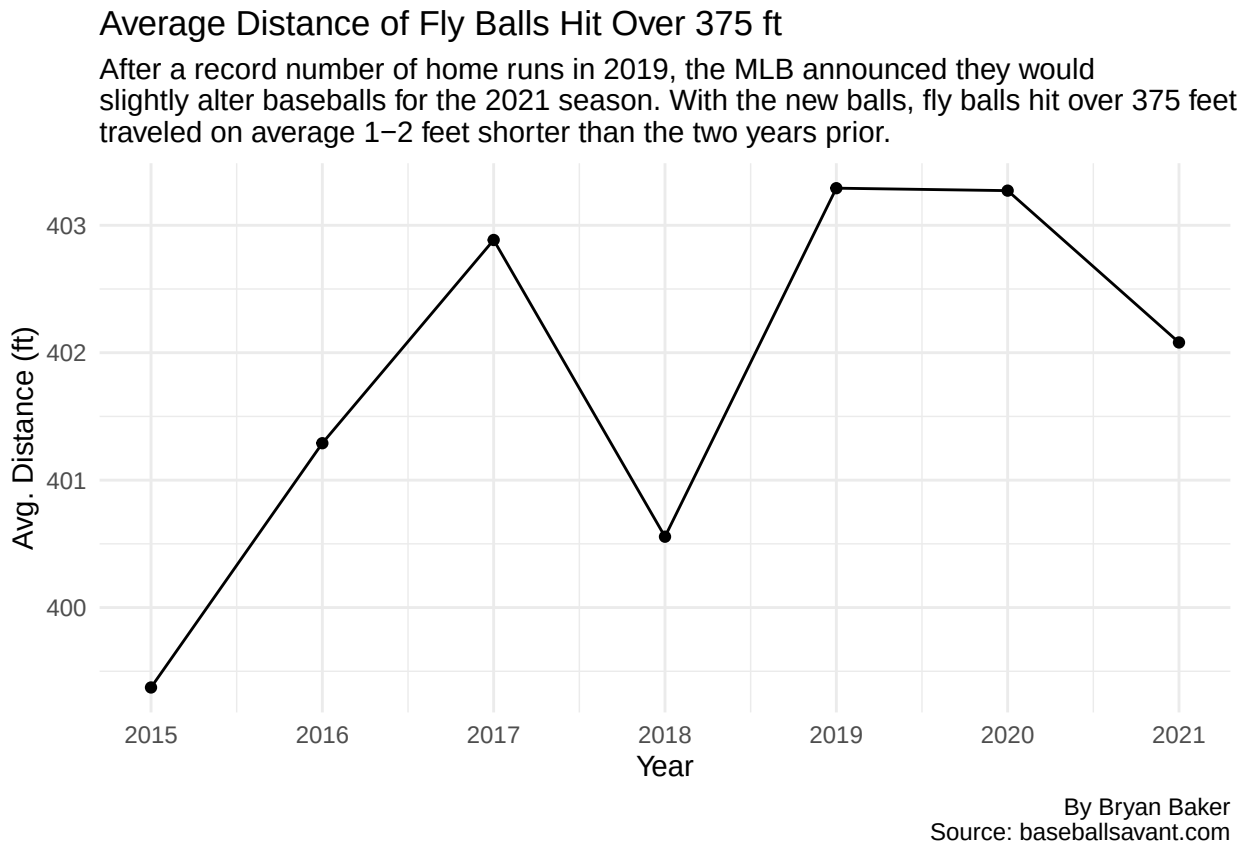
Merging the yearly datasets

```r
fb_yearly_merged <- fb_2015_avg %>% bind_rows(fb_2016_avg) %>%
  bind_rows(fb_2017_avg) %>% bind_rows(fb_2018_avg) %>% bind_rows(fb_2019_avg) %>%
  bind_rows(fb_2020_avg) %>% bind_rows(fb_2021_avg)

fb_yearly_merged
```

```
## # A tibble: 7 x 4
##    year mean_distance mean_ev mean_la
##   <dbl>         <dbl>   <dbl>   <dbl>
## 1  2015          399.    102.    29.6
## 2  2016          401.    103.    29.4
## 3  2017          403.    103.    28.7
## 4  2018          401.    103.    28.4
## 5  2019          403.    103.    28.6
## 6  2020          403.    103.    30.1
## 7  2021          402.    104.    29.1
```

Plotting merged data

```r
nb_dist_plot <- ggplot(fb_yearly_merged, aes(x = year, y = mean_distance)) + geom_point() +
  geom_line() +
  scale_y_continuous(breaks = c(395,396,397,398,399,400,401,402,403,404,405)) +
  scale_x_continuous(breaks = c(2015,2016,2017,2018,2019,2020,2021)) +
  labs(title = "Average Distance of Fly Balls Hit Over 375 ft",
       subtitle = "After a record number of home runs in 2019, the MLB announced they would \nslightly a
       x = "Year", y = "Avg. Distance (ft)", caption = "By Bryan Baker\nSource: baseballsavant.com") + t

nb_dist_plot
```

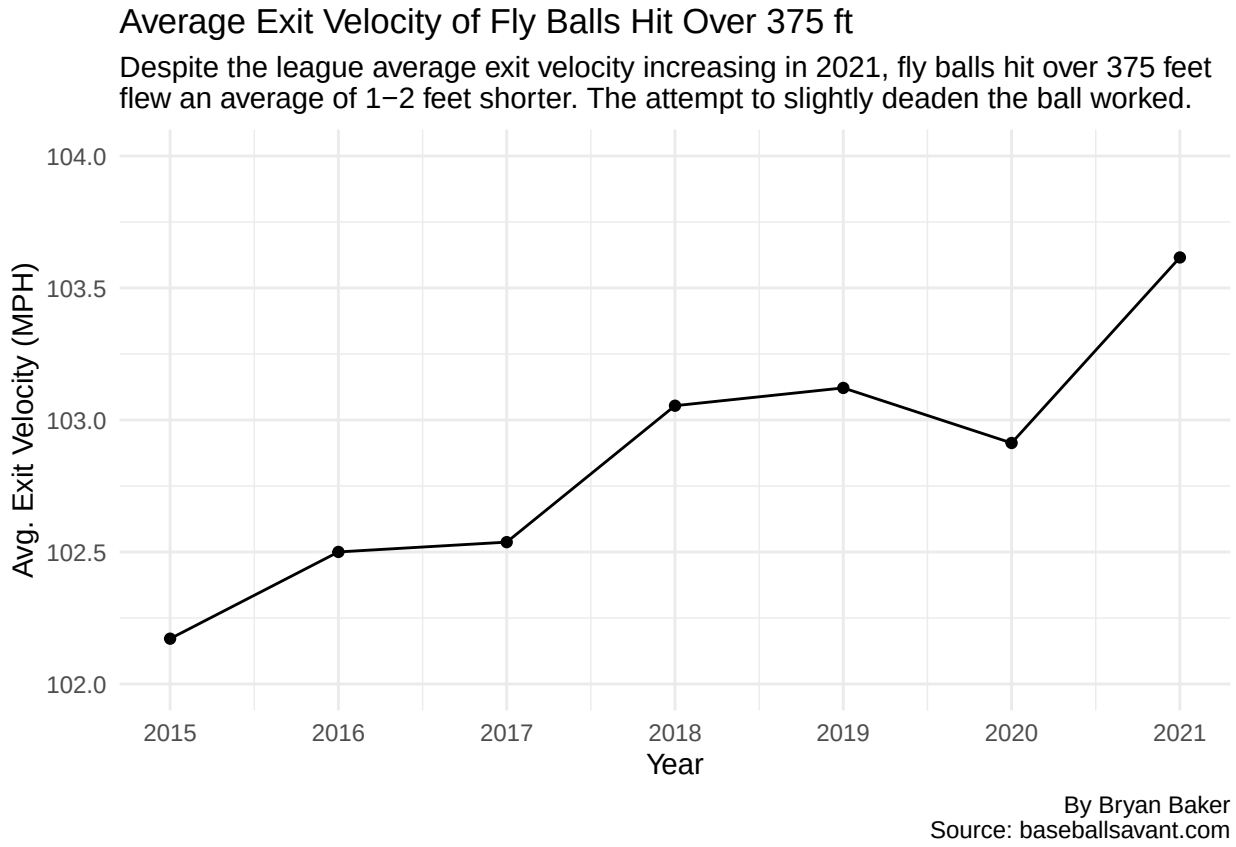



```
ggsave("images/nb_dist_plot.png", plot = nb_dist_plot)
```

```
## Saving 6.5 x 4.5 in image
```

Exit Velocity Averages Plot

```
ev_plot <- ggplot(fb_yearly_merged, aes(x = year, y = mean_ev)) + geom_point() +
  geom_line() + ylim(102,104) +
  scale_x_continuous(breaks = c(2015,2016,2017,2018,2019,2020,2021)) +
  labs(title = "Average Exit Velocity of Fly Balls Hit Over 375 ft", subtitle = "Despite the league ave

ev_plot
```




```
ggsave("images/ev_plot.png", plot = ev_plot)
```

```
## Saving 6.5 x 4.5 in image
```