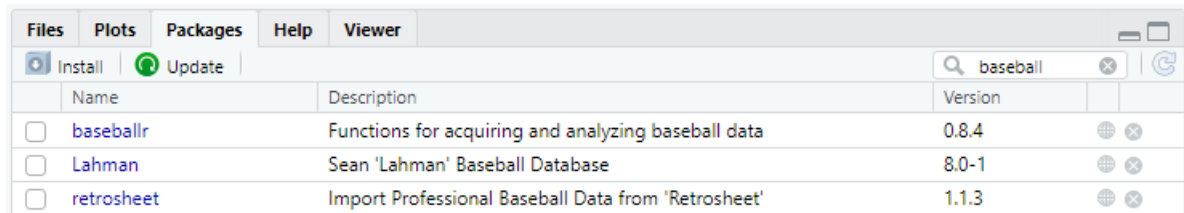


# Analysis of Sports Big Data Assignment 1

2021311175 Jae-Hyun Lee

## 1. Installing Lahman, retrosheet and baseballr packages

Installation completed.



Files	Plots	Packages	Help	Viewer
Install Update				
baseball				
	Name	Description	Version	
<input type="checkbox"/>	baseballr	Functions for acquiring and analyzing baseball data	0.8.4	
<input type="checkbox"/>	Lahman	Sean 'Lahman' Baseball Database	8.0-1	
<input type="checkbox"/>	retrosheet	Import Professional Baseball Data from 'Retrosheet'	1.1.3	

### Playing with Lahman: List of all Korean MLB players

```
> library(Lahman)
> People %>%
+   filter(birthCountry == "South Korea") %>%
+   select(debut, finalGame, birthDate, nameFirst, nameLast, birthCountry) %>%
+   arrange(debut)
```

	debut	finalGame	birthDate	nameFirst	nameLast	birthCountry
1	1994-04-08	2010-10-01	1973-06-30	Chan Ho	Park	South Korea
2	1998-07-04	1999-07-30	1975-08-16	Jin Ho	Cho	South Korea
3	1999-05-29	2007-09-28	1979-01-19	Byung-Hyun	Kim	South Korea
4	2000-06-29	2000-10-01	1971-03-11	Sang-Hoon	Lee	South Korea
5	2001-06-15	2006-09-17	1977-09-04	Sun-Woo	Kim	South Korea
6	2002-04-23	2004-06-20	1980-07-15	Jung	Bong	South Korea
7	2002-07-21	2007-05-29	1977-05-24	Jae Weong	Seo	South Korea
8	2002-09-03	2005-10-02	1979-03-16	Hee-Seop	Choi	South Korea
9	2003-03-31	2005-07-07	1974-03-04	Tommy	Phelps	South Korea
10	2004-08-08	2008-09-28	1980-05-29	Cha-Seung	Baek	South Korea
11	2005-04-04	2005-08-20	1969-08-02	Dae-Sung	Koo	South Korea
12	2005-04-21	2019-09-29	1982-07-13	Shin-Soo	Choo	South Korea
13	2006-05-14	2008-04-09	1983-05-30	Jae Kuk	Ryu	South Korea
14	2013-04-02	2019-09-28	1987-03-25	Hyun-Jin	Ryu	South Korea
15	2013-09-07	2013-09-27	1976-06-04	Chang-Yong	Lim	South Korea
16	2015-04-08	2019-07-29	1987-04-05	Jung Ho	Kang	South Korea
17	2015-07-11	2018-06-17	1991-03-26	Rob	Refsnyder	South Korea
18	2016-04-03	2019-05-30	1982-07-15	Seunghwan	Oh	South Korea
19	2016-04-04	2016-10-02	1982-06-21	Dae-ho	Lee	South Korea
20	2016-04-04	2016-06-28	1986-07-10	Byung Ho	Park	South Korea
21	2016-04-05	2019-09-29	1991-05-19	Ji-Man	Choi	South Korea
22	2016-04-10	2017-09-30	1988-01-12	Hyun Soo	Kim	South Korea
23	2017-06-28	2017-08-01	1987-07-28	Jae-gyun	Hwang	South Korea

- Not all players born in Korea are Korean. As you can tell from their names, Tommy Phelps is an American born in Seoul and Rob Refsnyder is an adopted Korean-American.
- The latest data: Kwang-hyun Kim playing for St.Louis Cardinals is missing.

## Playing with retrosheet: Ryu's Second Complete Game Shutout in MLB

```
library(retrosheet)
gamelog_2019 <- getRetrosheet("game", 2019)
gamelog_2019 %>%
  filter(Date == 20190507 & HmTm == "LAN")
```

	Date	DblHd	Day	VisTm	VisTmLg	VisTmGNum	HmTm	HmTmLg	HmTmGNum	VisRuns	HmRuns	NumOuts	DayNight					
1	20190507	0	Tue	ATL	NL	36	LAN	NL	38	0	9	51	N					
	Completion	Forfeit	Protest	ParkID	Attendance	Duration	VisLine	HmLine	VisAB	VisH	VisD	VisT						
1		NA	NA	LOS03	47337	162	000000000	31001004x	30	4	1	0						
	VisHR	VisRBI	VisSH	VisSF	VisHBP	VisBB	VisIBB	VisK	VisSB	VisCS	VisGDP	VisCI	VisLOB	VisPs	VisER			
1	0	0	0	0	0	0	0	6	0	0	1	0	3	5	9			
	VisTER	VisWP	VisBalks	VisPO	VisA	VisE	VisPassed	VisDB	VisTP	HmAB	HmH	HmD	HmT	HmHR	HmRBI	HmSH	HmSF	
1	9	2	0	24	5	0	0	0	0	33	11	1	0	3	9	1	1	
	HmHBP	HmBB	HmIBB	HmK	HmSB	HmCS	HmGDP	HmCI	HmLOB	HmPs	HmER	HmTER	HmWP	HmBalks	HmPO	HmA	HmE	HmPass
1	1	3	0	5	0	0	0	0	6	1	0	0	0	0	27	11	0	0
	HmDB	HmTP	UmpHID	UmpPHNm	Ump1BID	Ump1BNm	Ump2BID	Ump2BNm	Ump3BID	Ump3BNm								
1	1	0	fleta001	Andy Fletcher	littw901	Will Little	westj901	Joe West	coope901	Eric Cooper								
	UmpLFID	UmpLFNm	UmpRFID	UmpRFNm	VisMgrID	VisMgrNm	HmMgrID	HmMgrNm	WinPID									
1		NA	(none)	NA	(none)	snitb801	Brian Snitker	robed001	Dave Roberts	ryu-h001								
	WinPNm	PID	PName	SavePID	SavePNm	GwinRBIID	GwinRBINm	VisStPchID	VisStPchNm									
1	Hyun-Jin Ryu	friem001	Max Fried	(none)	turnj001	Justin Turner	friem001	Max Fried										
	HmStPchID	HmStPchNm	VisBat1ID	VisBat1Nm	VisBat1Pos	VisBat2ID	VisBat2Nm	VisBat2Pos										
1	ryu-h001	Hyun-Jin Ryu	albio001	Ozzie Albies	4	donaj001	Josh Donaldson	5										
	VisBat3ID	VisBat3Nm	VisBat3Pos	VisBat4ID	VisBat4Nm	VisBat4Pos	VisBat5ID	VisBat5Nm										
1	freef001	Freddie Freeman	3	acunr001	Ronald Acuna	7	markn001	Nick Markakis										
	VisBat5Pos	VisBat6ID	VisBat6Nm	VisBat6Pos	VisBat7ID	VisBat7Nm	VisBat7Pos	VisBat8ID										
1	9	swand001	Dansby Swanson	6	flowt001	Tyler Flowers	2	incie001										
	VisBat8Nm	VisBat8Pos	VisBat9ID	VisBat9Nm	VisBat9Pos	HmBat1ID	HmBat1Nm	HmBat1Pos										
1	Ender Inciarte	8	friem001	Max Fried	1	herne001	Enrique Hernandez	8										
	HmBat2ID	HmBat2Nm	HmBat2Pos	HmBat3ID	HmBat3Nm	HmBat3Pos	HmBat4ID	HmBat4Nm										
1	turnj001	Justin Turner	5	bellc002	Cody Bellinger	9	freed001	David Freese										
	HmBat4Pos	HmBat5ID	HmBat5Nm	HmBat5Pos	HmBat6ID	HmBat6Nm	HmBat6Pos	HmBat7ID	HmBat7Nm									
1	3	taylc001	Chris Taylor	6	muncm001	Max Muncy	4	martr004	Russell Martin									
	HmBat7Pos	HmBat8ID	HmBat8Nm	HmBat8Pos	HmBat9ID	HmBat9Nm	HmBat9Pos	Additional Acquisition										
1	2	verda001	Alex Verdugo	7	ryu-h001	Hyun-Jin Ryu	1											Y

I chose the game log of LA Dodgers on May 7, 2019. This is a memorable game for Korean baseball fans when Hyun-Jin Ryu recorded his second complete game shutout in MLB against Atlanta Braves.

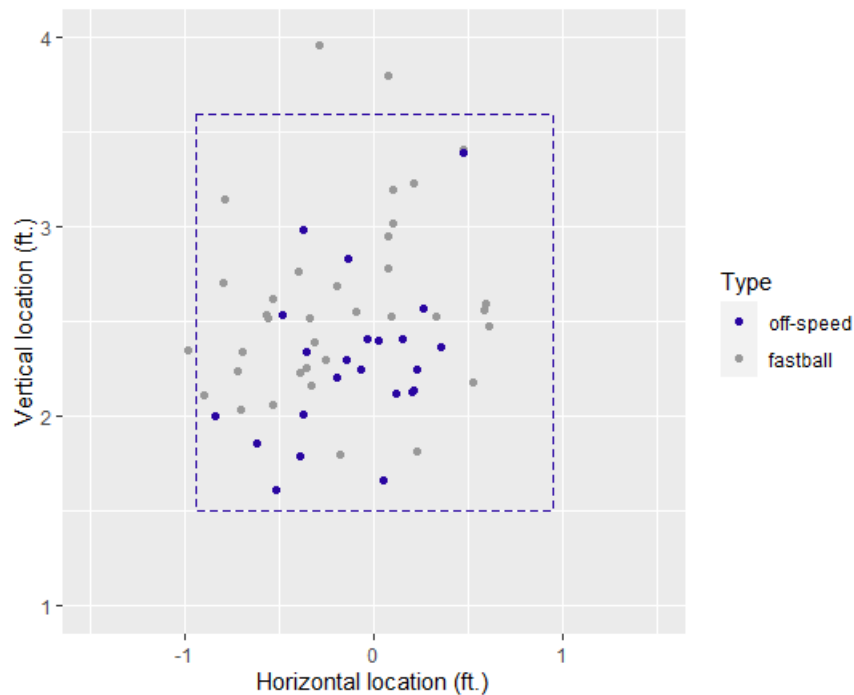
- This game took place on Tuesday (Day = Tue)
- in front of 47,337 people (Attendance = 47337)
- in Los Angeles (HmTm = LAN)
- which lasted 2.7 hours (Duration = 162 minutes).

## 2. 3. 4. Visiting Relevant Websites

- <http://www.seanlahman.com/>
  - Read [A Bill James Primer](#), showing some interesting insights of statistics in baseball.
  - Read [The Sabermetric Manifesto](#), which underlines the key concepts of arguments based on statistics.
    1. Does the statistic measure an important contribution to a goal?
    2. How well does the statistic measure the player's own contribution?
    3. Is there a better way to measure the same thing?
- <https://www.retrosheet.org/#>
  - Reviewed [Games/People/Parks](#).
- <https://baseballsavant.mlb.com/>
  - Reviewed the elements and obtained insights for visualization.
- [https://github.com/beanumber/baseball\\_R](https://github.com/beanumber/baseball_R)

- Reviewed the elements.
- <https://www.mlb.com/news/sabermetrics-in-baseball-a-casual-fans-guide>
  - Read `what is Sabermetrics?`, a brief introduction to sabermetrics.
- <https://library.fangraphs.com/>
  - Checked the `intro` tab, `offense` tab and `pitching` tab.
  - Read articles on `OBP`, `OBP` and `OPS+`, `wOBA`, `wRC` and `wRC+` and `wRAA` in the `offense` tab.
  - Read articles on `ERA`, `WHIP` and `FIP` in the `pitching` tab.

## 5. Reproducing Figure 2.3



Added lines on top to avoid errors.

```
library(ggplot2)
library(tidyverse)
crcblue <- "#2905a1" #The reason of an unknown error
```

## 6. Constructing a similar figure for Aaron Judge

Using the same data (statcast2017.csv), simply replace Giancarlo Stanton with Aaron Judge.

```
judge_hr <- statcast2017 %>%
  filter(player_name == "Judge, Aaron", events == "home_run") %>%
  mutate(is_fb = pitch_type %in% c("FF", "FT", "SI", "FC"))

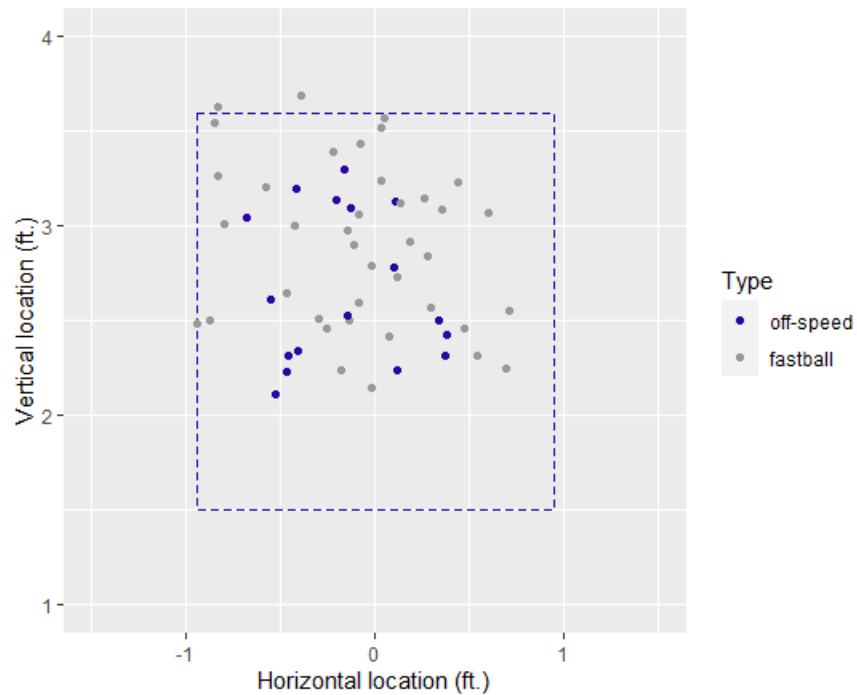
plate_width <- 17 + 2 * (9/pi)
k_zone_plot <- ggplot(NULL, aes(x = plate_x, y = plate_z)) +
  geom_rect(xmin = -(plate_width/2)/12,
            xmax = (plate_width/2)/12,
            ymin = 1.5,
            ymax = 3.6, color=crcblue, fill = "lightgray",
```

```

linetype = 2, alpha = 0.01) +
coord_equal() +
scale_x_continuous("Horizontal location (ft.)",
                    limits = c(-1.5, 1.5)) +
scale_y_continuous("Vertical location (ft.)",
                    limits = c(1, 4))

k_zone_plot %>% judge_hr +
  aes(color = is_fb) +
  geom_point() +
  scale_color_manual("Type", values = c(crcblue,"gray60"),
                    labels = c("off-speed", "fastball")
  )

```



### Comments on Judge's plot

- Most of the balls he hit was in the strike zone. (Just like any other batters)
- The *off-speed* balls and *fastballs* are roughly the same. (Just like Stanton)
- More upper-side of the strike zone than lower-side of the zone.