

Project1

Sri Seshadri

9/30/2017

1. Introduction

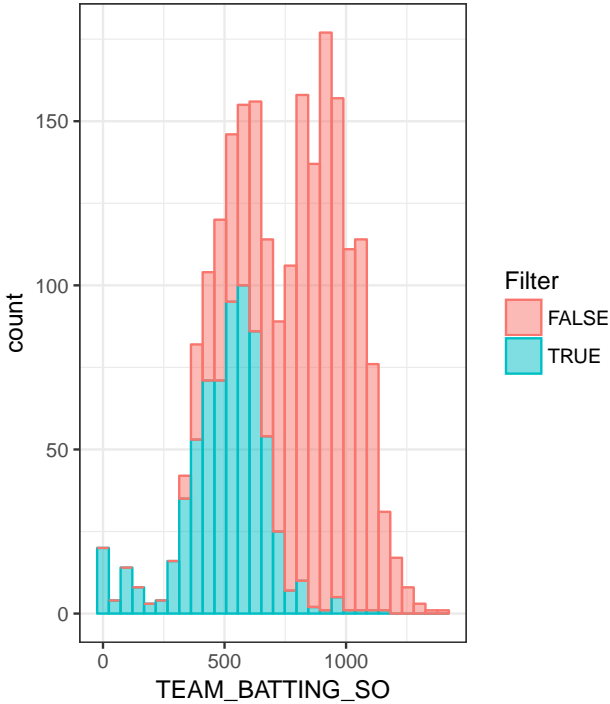
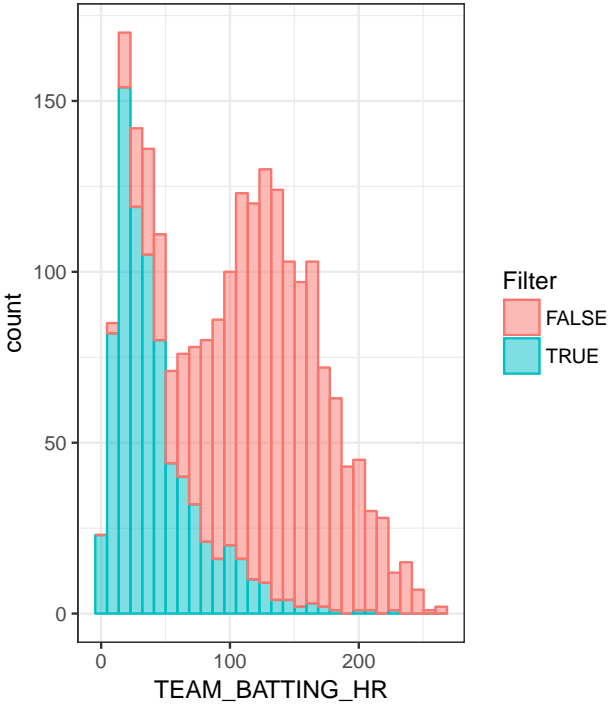
2. Exploratory Data Analysis (EDA)

2.1 About the data

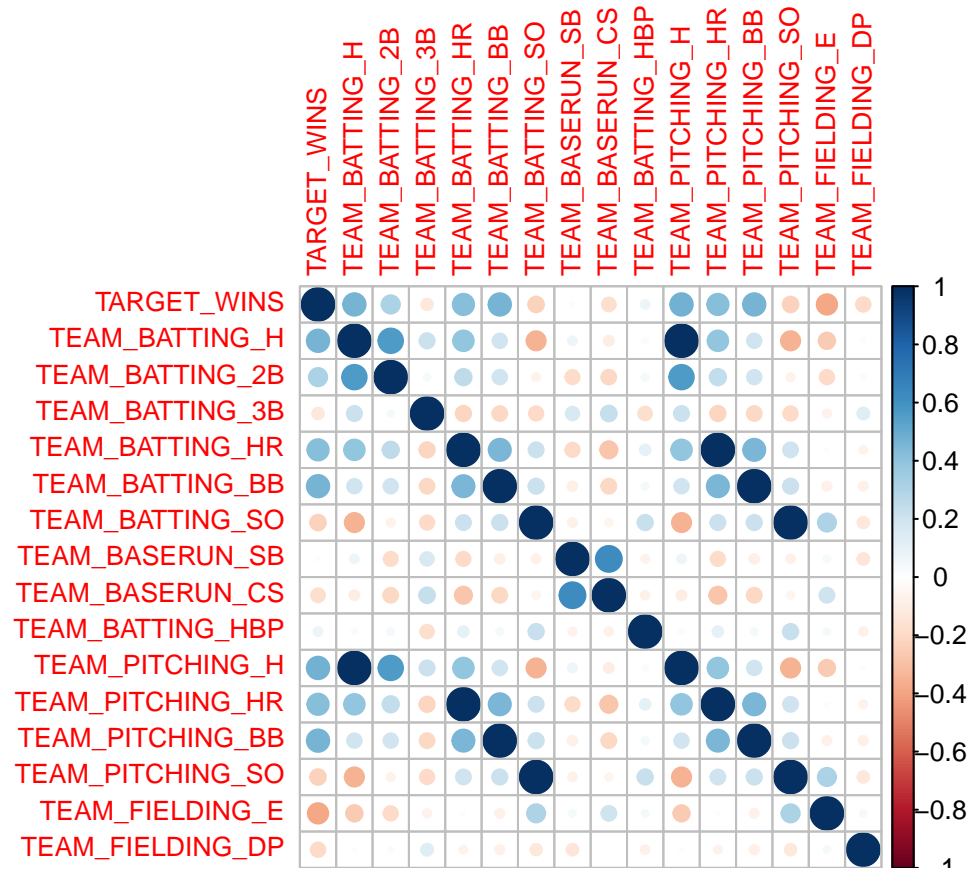
Table 1: Summary Stats and missing values

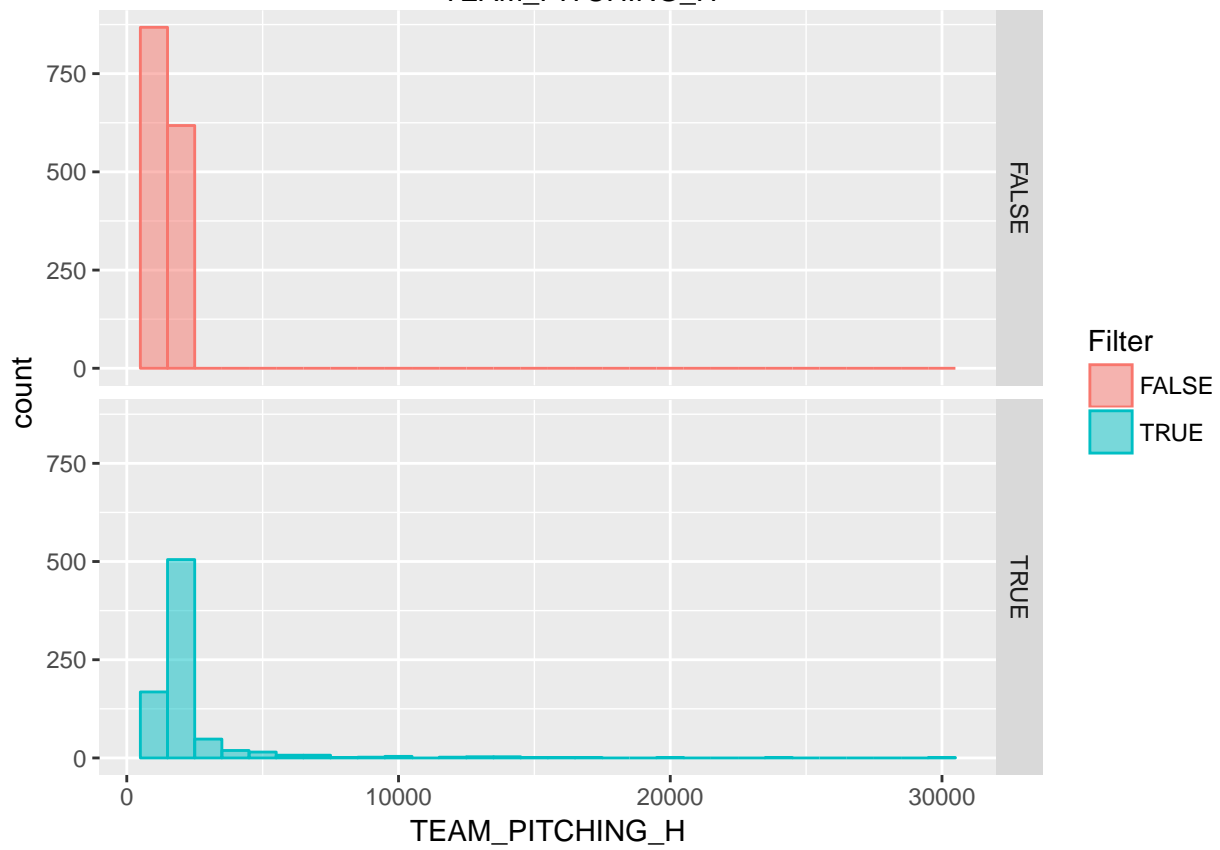
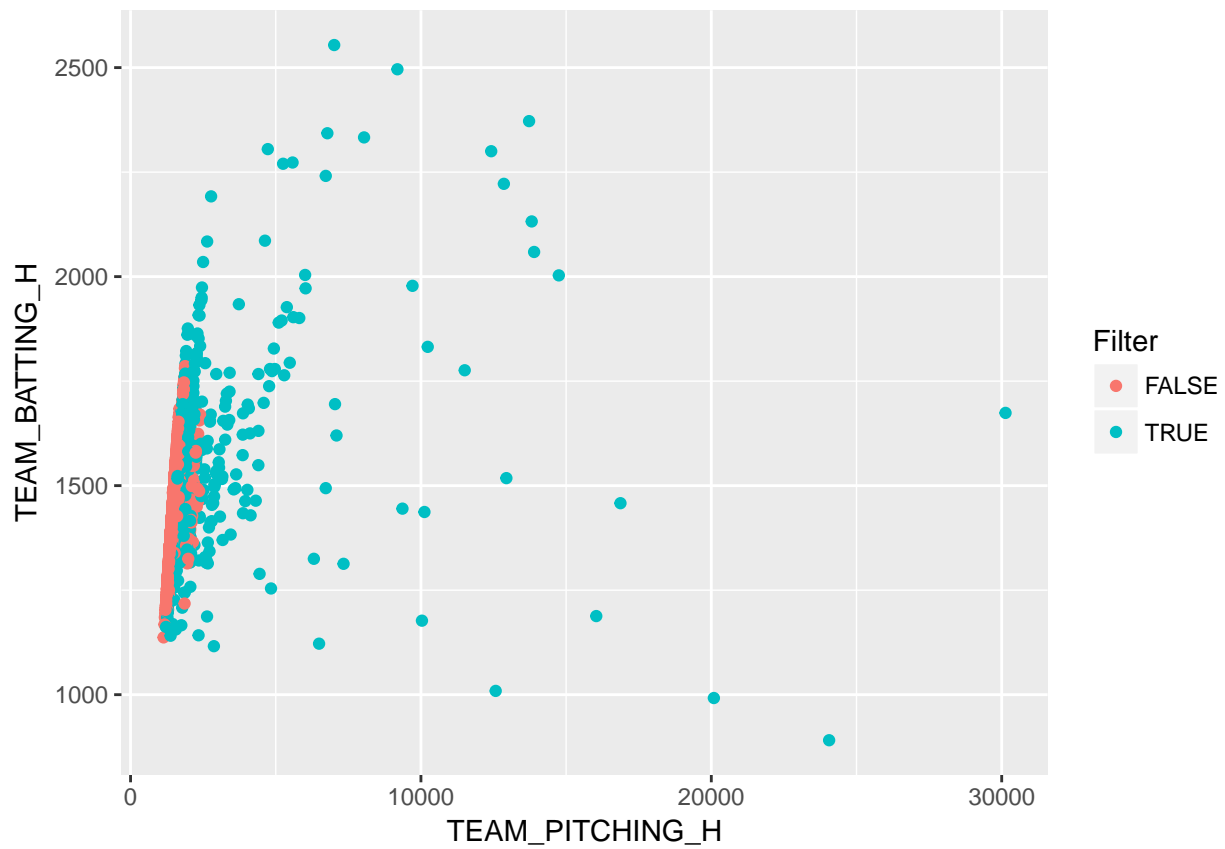
	min	Q1	median	Q3	max	mean	sd	n	missing	cv
INDEX	1	630.75	1270.5	1915.50	2535	1268.46	736.35	2276	0	0.58
TARGET_WINS	0	71.00	82.0	92.00	146	80.79	15.75	2276	0	0.19
TEAM_BATTING_H	891	1383.00	1454.0	1537.25	2554	1469.27	144.59	2276	0	0.10
TEAM_BATTING_2B	69	208.00	238.0	273.00	458	241.25	46.80	2276	0	0.19
TEAM_BATTING_3B	0	34.00	47.0	72.00	223	55.25	27.94	2276	0	0.51
TEAM_BATTING_HR	0	42.00	102.0	147.00	264	99.61	60.55	2276	0	0.61
TEAM_BATTING_BB	0	451.00	512.0	580.00	878	501.56	122.67	2276	0	0.24
TEAM_BATTING_SO	0	548.00	750.0	930.00	1399	735.61	248.53	2174	102	0.34
TEAM_BASERUN_SB	0	66.00	101.0	156.00	697	124.76	87.79	2145	131	0.70
TEAM_BASERUN_CS	0	38.00	49.0	62.00	201	52.80	22.96	1504	772	0.43
TEAM_BATTING_HBP	29	50.50	58.0	67.00	95	59.36	12.97	191	2085	0.22
TEAM_PITCHING_H	1137	1419.00	1518.0	1682.50	30132	1779.21	1406.84	2276	0	0.79
TEAM_PITCHING_HR	0	50.00	107.0	150.00	343	105.70	61.30	2276	0	0.58
TEAM_PITCHING_BB	0	476.00	536.5	611.00	3645	553.01	166.36	2276	0	0.30
TEAM_PITCHING_SO	0	615.00	813.5	968.00	19278	817.73	553.09	2174	102	0.68
TEAM_FIELDING_E	65	127.00	159.0	249.25	1898	246.48	227.77	2276	0	0.92
TEAM_FIELDING_DP	52	131.00	149.0	164.00	228	146.39	26.23	1990	286	0.18

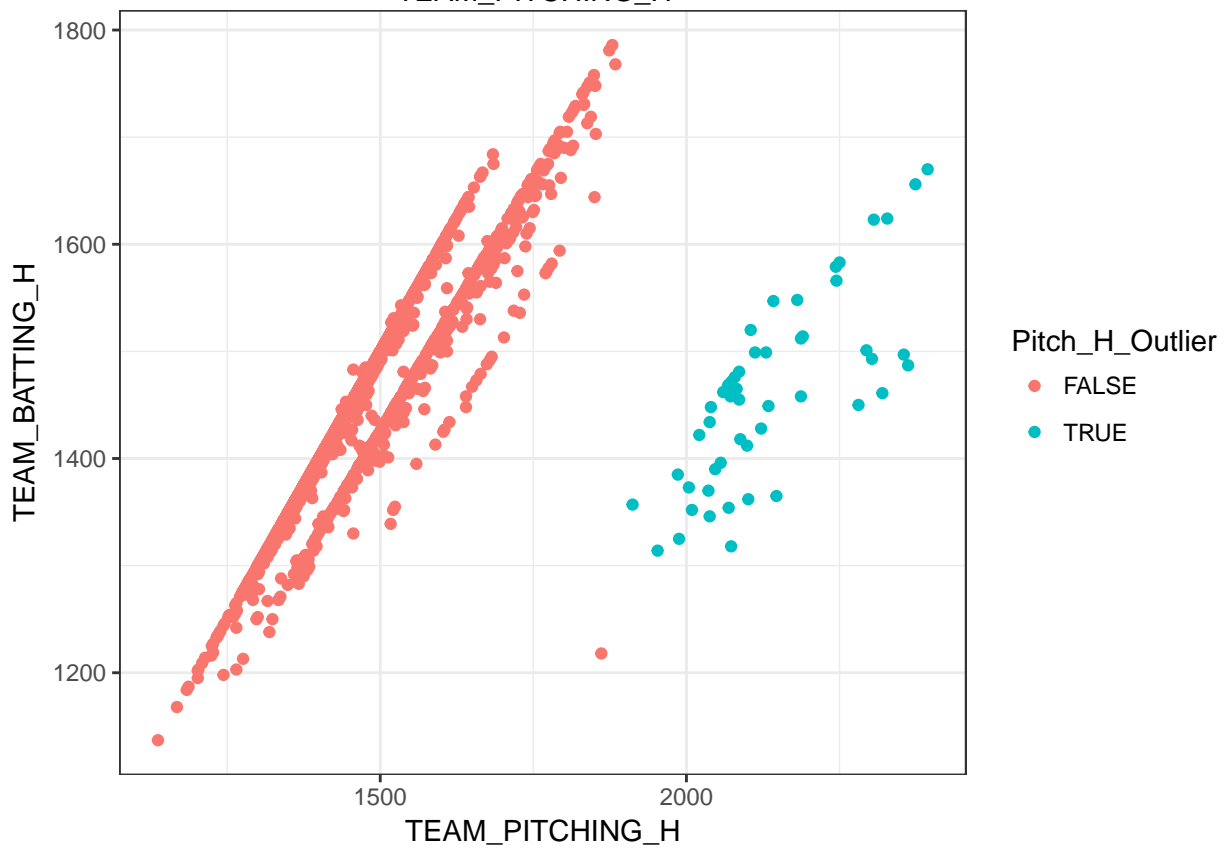
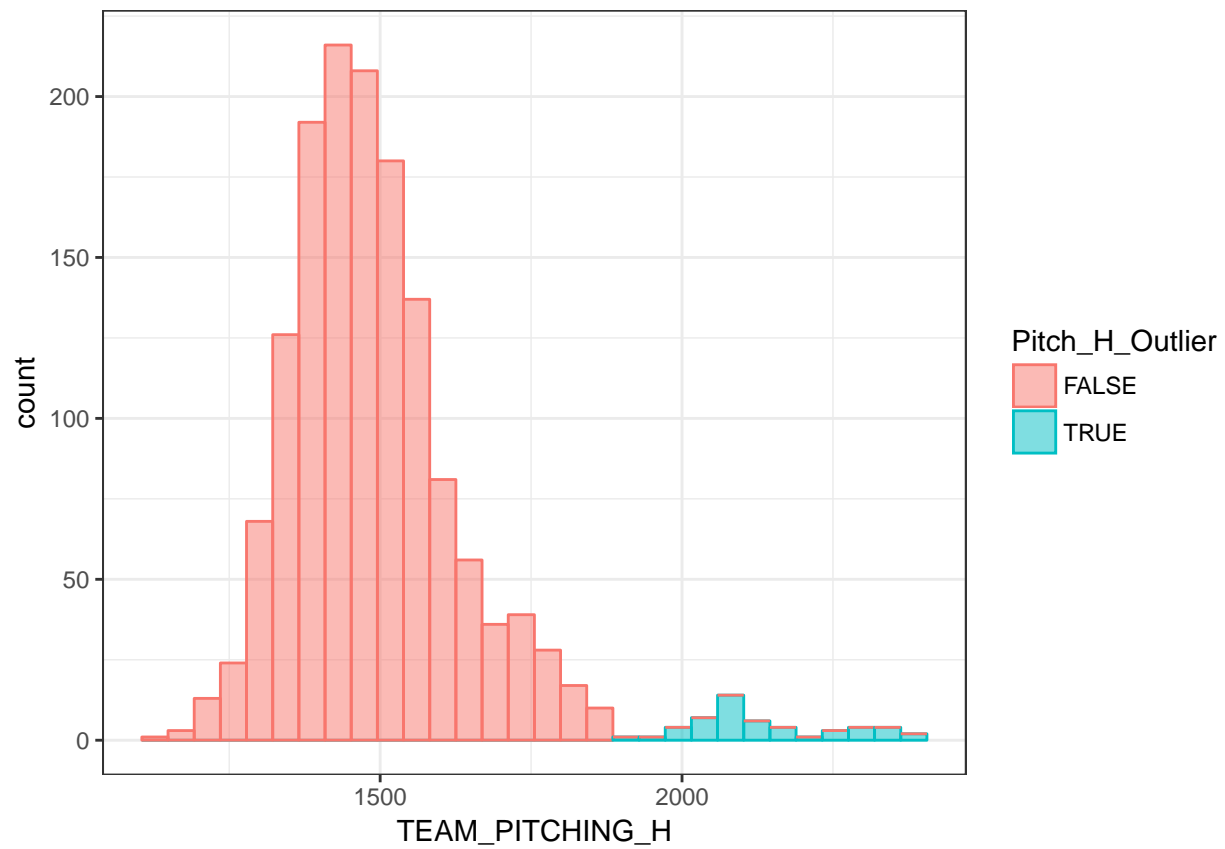
2.2 Missing data

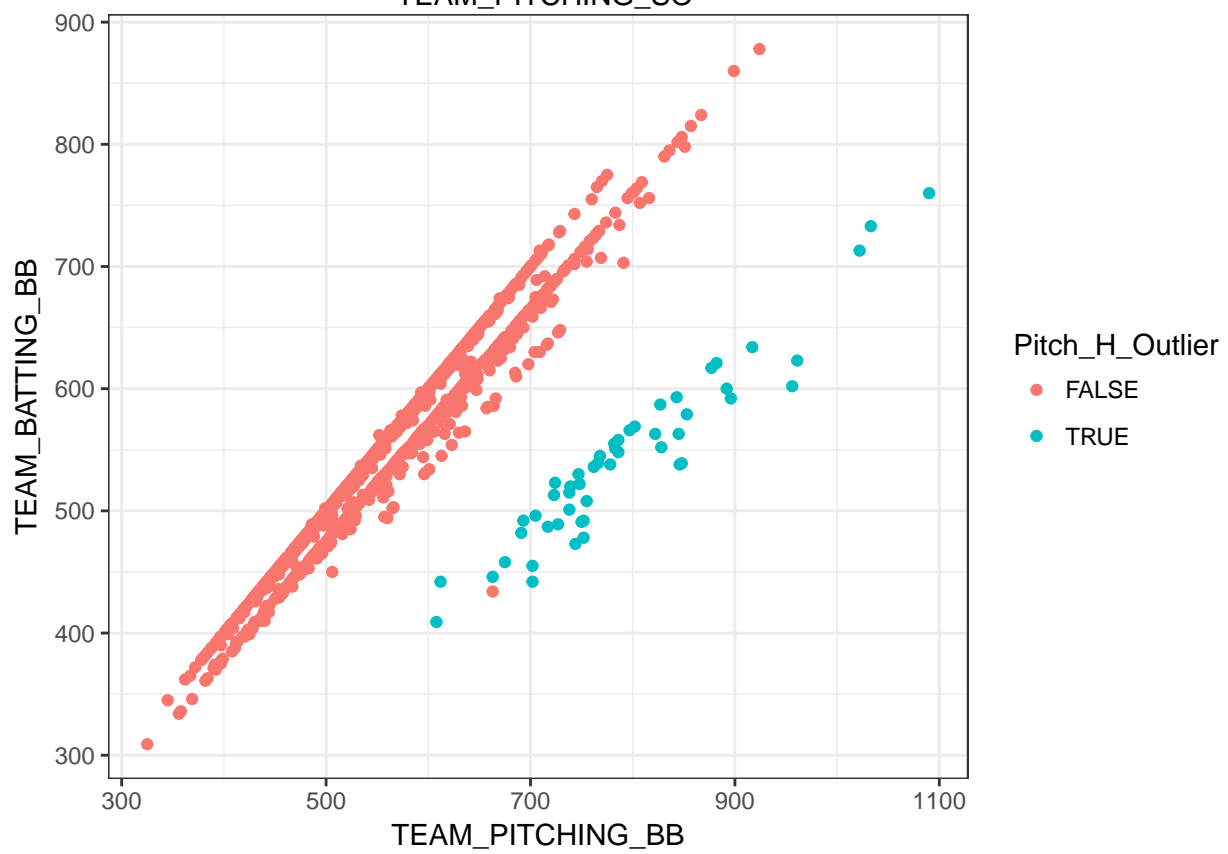
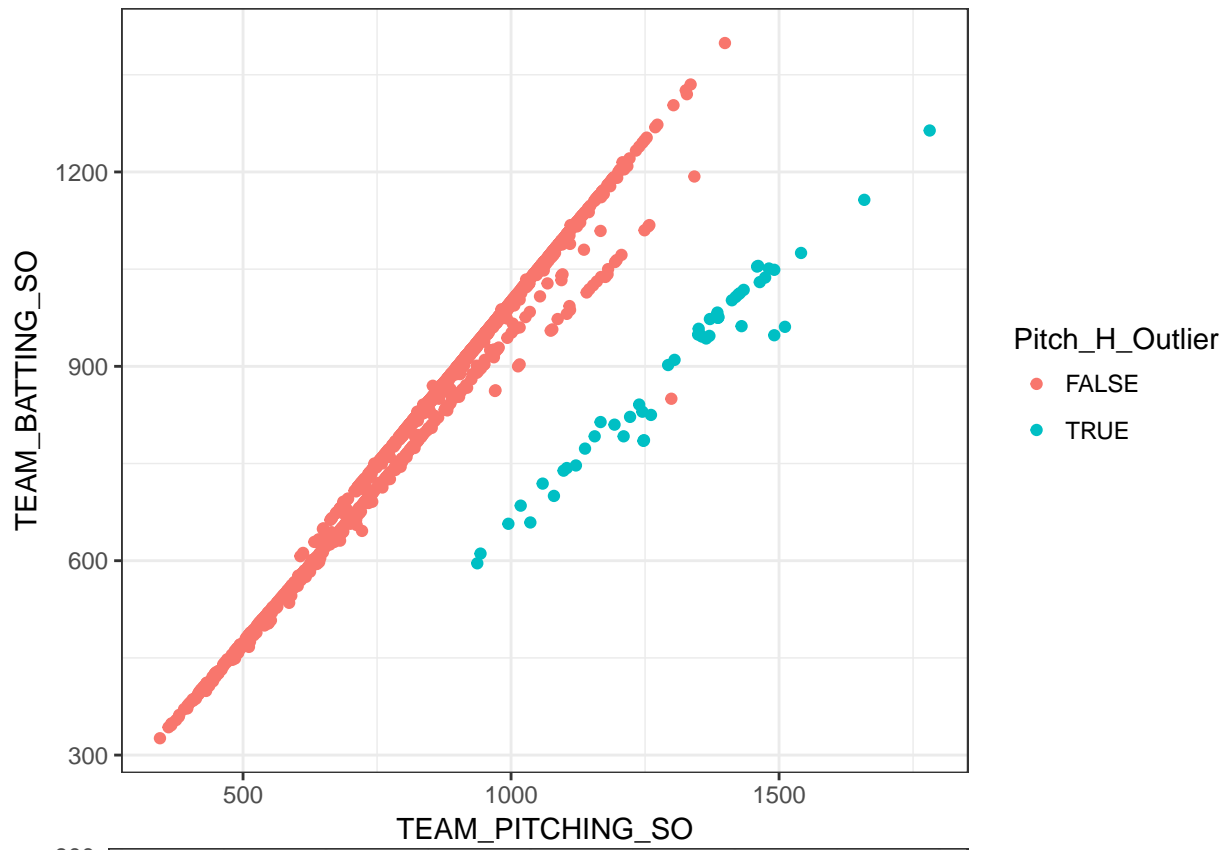


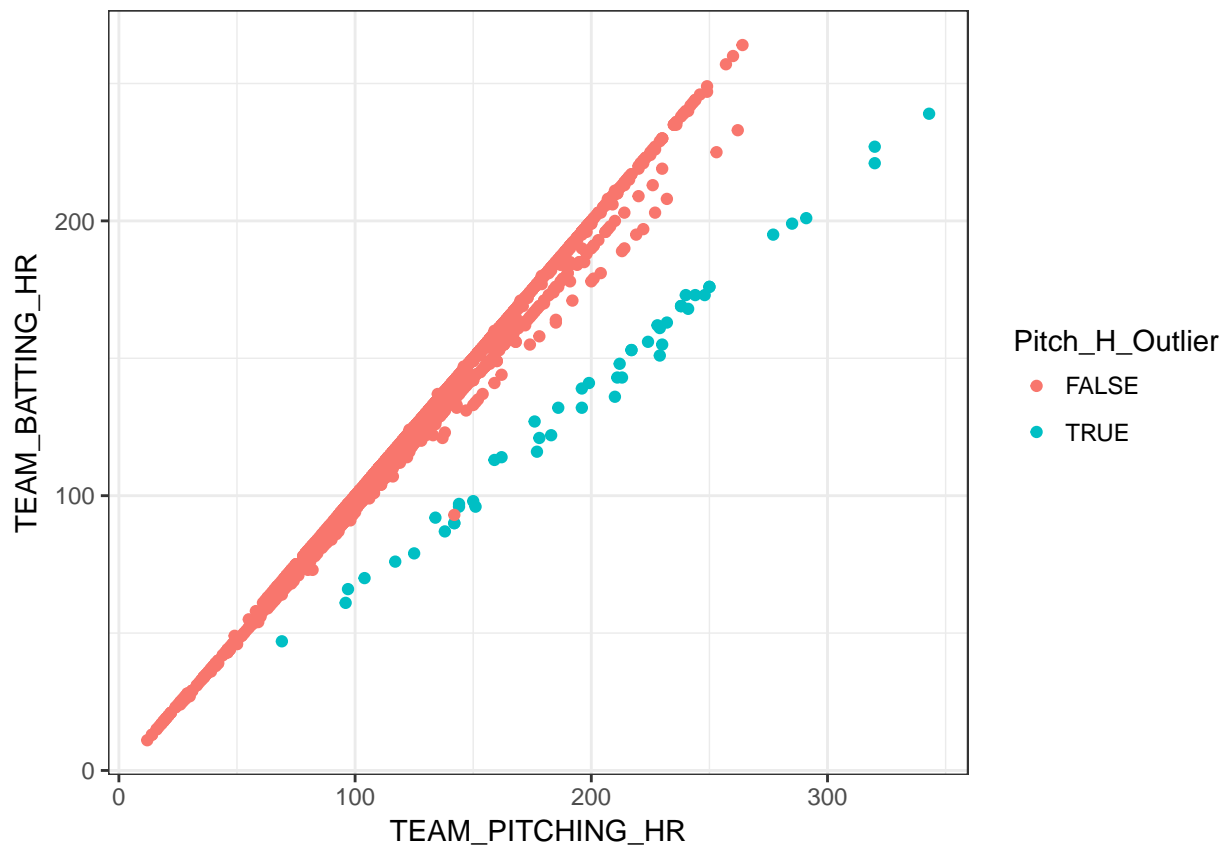
2.3 Relationships among variables











2.2.1 Do we throw away a lot of data?

