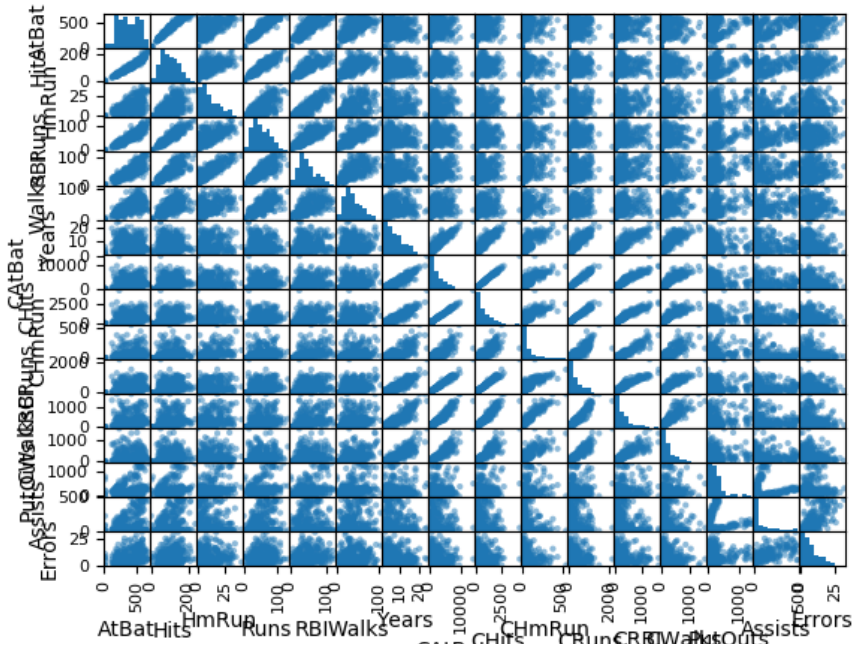
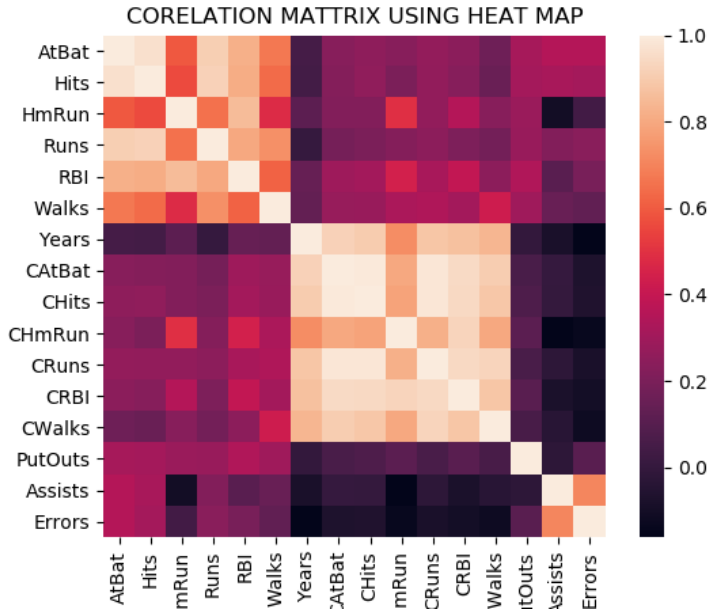
Craig Huff
CS 161 - 01

Homework 3: Exploratory Data Analysis using scikit-learn

2. Code included in homework3.py file

3. Correlation and scatter plot images.





4. The descriptive statistics describe the data and how they appear in different graphs. The statistics are the mean, count, standard deviation, minimum and maximum, and percentiles of the data increasing quarterly.

Index	AtBat	Hits	HmRun	Runs	RBI	Walks	Years	CAtBat	CHits	CHmRun
count	322	322	322	322	322	322	322	322	322	322
mean	380.929	101.025	10.7702	50.9099	48.028	38.7422	7.4441	2648.68	717.571	69.4907
std	153.405	46.4547	8.70904	26.0241	26.1669	21.6393	4.92609	2324.21	654.473	86.2661
min	16	1	0	0	0	0	1	19	4	0
25%	255.25	64	4	30.25	28	22	4	816.75	209	14
50%	379.5	96	8	48	44	35	6	1928	508	37.5
75%	512	137	16	69	64.75	53	11	3924.25	1059.25	90
max	687	238	40	130	121	105	24	14053	4256	548

5. Career statistics have a higher correlation of the other career statistics with a similar exponential distribution, and a similar distribution with salary. Whereas the current season statistics are more random and less related to the players salaries. The strongest correlation that exists is between the number of years the player has played and their multiple career statistics such as hits, runs, RBIs, etc.