Mr. Salemi's Lame Statistics

This report will use the `TeamHitting.csv` data to analyze three statistics: Hits per game, Walks per game, and Strikeouts per game. First we read our hitting data.

```
hitting = readtable("TeamHitting.csv")
```

```
Warning: Column headers from the file were modified to make them valid MATLAB identifiers
before creating variable names for the table. The original column headers are saved in the
VariableDescriptions property.
Set 'VariableNamingRule' to 'preserve' to use the original column headers as table variable names.
hitting = 30×20 table
```

...

	Team	G	PA	AB	R	H	x2B	x3B
1	'Diamondbacks'	138	5288	4674	680	1169	236	33
2	'Athletics'	138	5245	4723	625	1193	250	16
3	'Braves'	137	5242	4663	604	1138	207	14
4	'Orioles'	137	5099	4608	595	1113	225	17
5	'Red Sox'	138	5290	4729	678	1192	274	21
6	'Cubs'	137	5224	4664	676	1164	229	25
7	'White Sox'	137	5064	4547	536	1060	206	9
8	'Reds'	137	5159	4613	611	1137	217	20
9	'Guardians'	135	4947	4413	519	982	192	14
10	'Rockies'	137	5018	4590	516	1098	222	35
11	'Tigers'	138	5182	4643	661	1159	211	32
12	'Astros'	137	5141	4628	572	1167	206	11
13	'Royals'	137	5064	4582	529	1129	242	19
14	'Angels'	136	5073	4534	581	1037	187	14
15	'Dodgers'	137	5234	4625	700	1166	217	17
16	'Marlins'	137	5192	4685	599	1178	230	26
17	'Brewers'	138	5296	4694	703	1217	224	15
18	'Twins'	136	5081	4552	573	1083	207	16
19	'Mets'	137	5227	4608	650	1152	226	18
20	'Yankees'	137	5292	4646	719	1162	220	19
21	'Phillies'	137	5209	4663	650	1197	225	20
22	'Pirates'	138	5119	4577	500	1071	206	18
23	'Padres'	137	5152	4576	582	1154	224	16
24	'Mariners'	137	5232	4652	627	1124	186	7
25	'Giants'	137	5133	4550	580	1072	205	25

	Team	G	PA	AB	R	H	x2B	x3B
26	'Cardinals'	138	5183	4641	603	1146	227	6
27	'Rays'	136	5081	4596	611	1148	204	15
28	'Rangers'	138	5154	4635	596	1104	214	12
29	'Blue Jays'	137	5253	4676	675	1253	247	9
30	'Nationals'	136	5046	4551	571	1107	223	21

Hits per game

I imagine that getting more hits per game leads to more runs per game. I expect to see a positive coorelation of this statistic:

$$\text{hpg} = \frac{H}{G}$$

```
hitting.hpg = hitting.H ./ hitting.G;
```

Correlation to Runs and Scatter Diagram

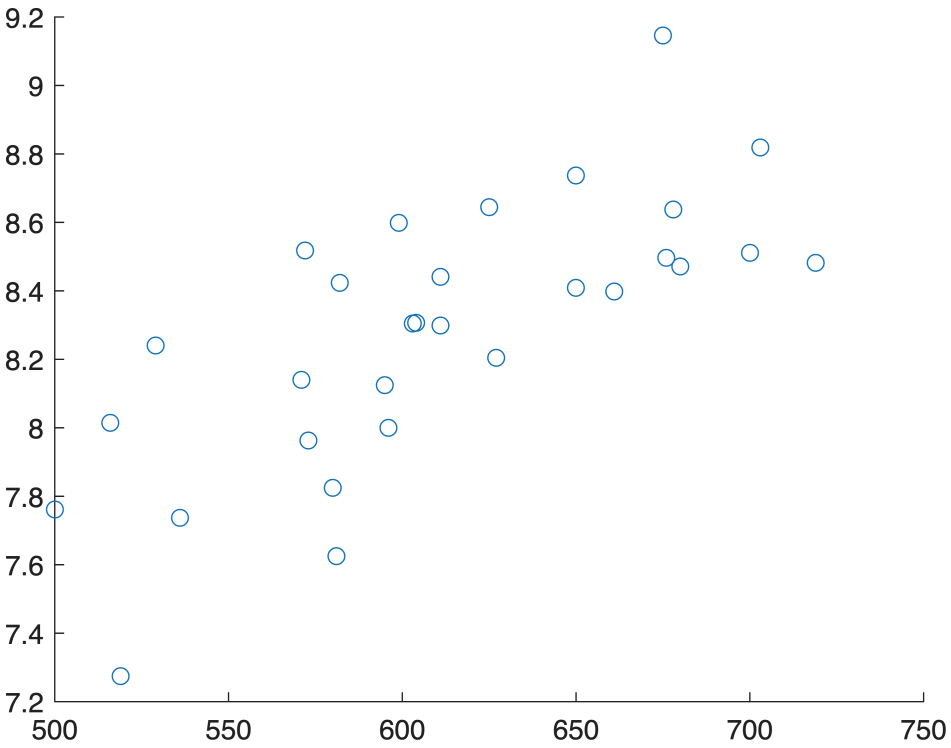
Now I'll find the correlation of *hits per game* to *runs*:

```
hpr_cor = corrcoef([hitting.hpg  hitting.R])
```

```
hpr_cor = 2×2
    1.0000    0.7307
    0.7307    1.0000
```

Pretty good at .73. Let's look at the scatter diagram.

```
scatter(hitting.R, hitting.hpg)
```



This was sort of up and to the right, but not clearly connected. Let's try the next statistic.

Walks (BB = Bases on Balls) Per Game

A team that walks more should score more runs. First we define wpg (walks per game)

$$\text{wpg} = \frac{\text{BB}}{G}$$

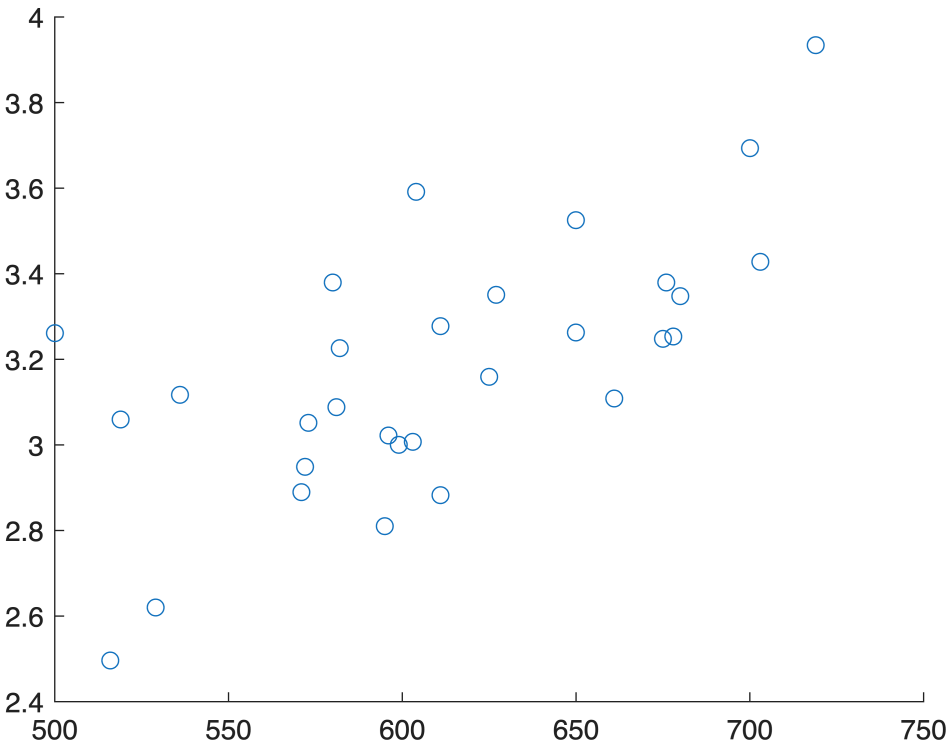
```
hitting.wpg = hitting.BB ./ hitting.G;
```

Correlating wpg to Runs and creating a scatter diagram

```
corrcoef([hitting.wpg hitting.R])
```

```
ans = 2×2
    1.0000    0.6697
    0.6697    1.0000
```

```
scatter(hitting.R, hitting.wpg)
```



The a correlation but not as good as the one for hits.

Strikeouts (SO) per game (kpg)

The team that strikes out less should score more runs.

$$\text{kpg} = \frac{\text{SO}}{G}$$

```
hitting.kpg = hitting.SO ./ hitting.G;
```

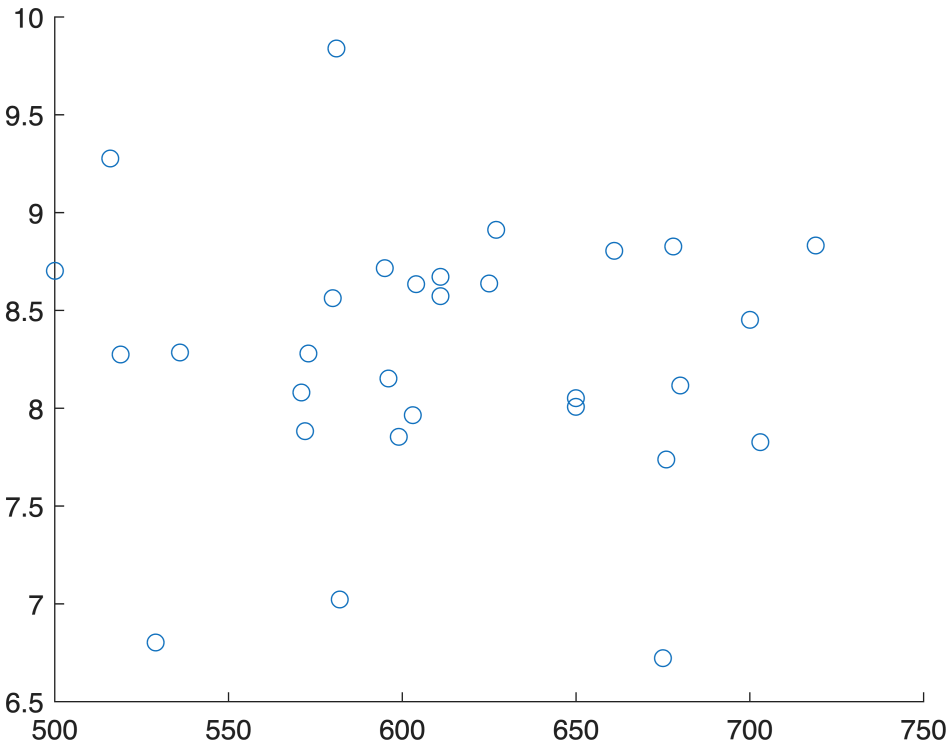
Correleating KPG to Runs

We should see a negative correlation between strikeouts per game and runs.

```
corrcoef([hitting.kpg, hitting.R])
```

```
ans = 2×2
    1.0000   -0.0612
   -0.0612    1.0000
```

```
scatter(hitting.R, hitting.kpg)
```



Wow. There is pretty much no correlation at all.

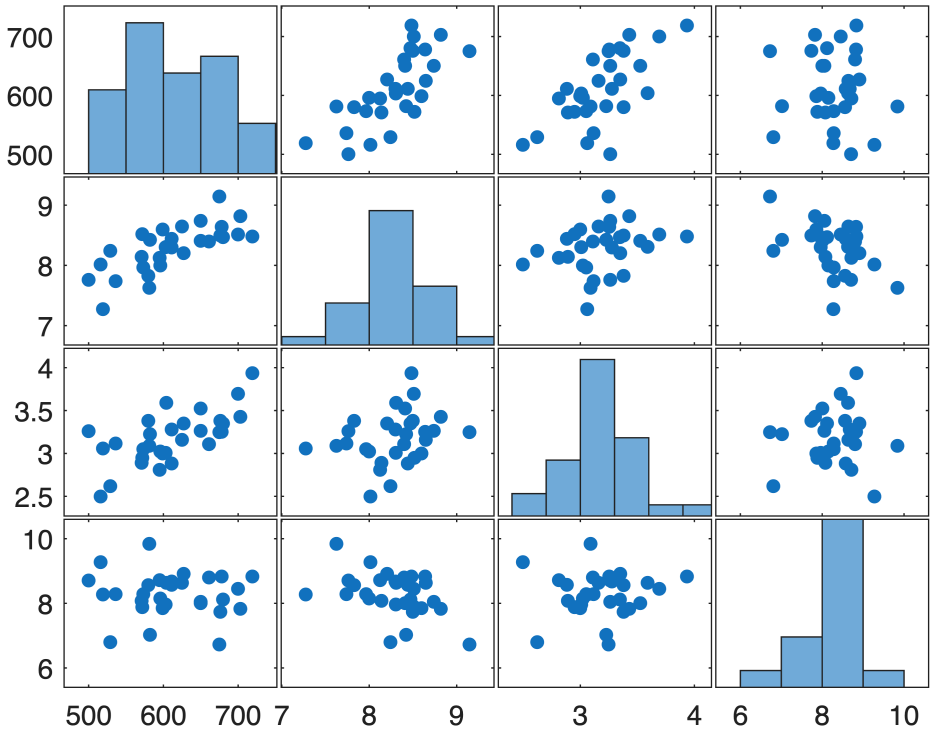
Comparing the statistics to each other

Now we'll compare the statistics to see which is best. First we make a correlation table.

```
corrcoef([hitting.R hitting.hpg hitting.wpg hitting.kpg])
```

```
ans = 4×4
    1.0000    0.7307    0.6697   -0.0612
    0.7307    1.0000    0.2758   -0.4371
    0.6697    0.2758    1.0000    0.0704
   -0.0612   -0.4371    0.0704    1.0000
```

```
plotmatrix([hitting.R, hitting.hpg, hitting.wpg, hitting.kpg])
```



In summary, *hits per game* was the best statistic for predicting which team would score more runs. However, none of these statistics was great.