Brandon Zink
Assignment 3

1. The average R/9 (runs per nine innings) across the MLB for this data set is 8.43 R/9. The minimum is 6.96 R/9 at the San Francisco Giants stadium. The maximum is 10.83 R/9 at the Colorado Rockies stadium.

This looks to have to do a lot with park size and elevation. Rockies stadium is at high elevation so balls carry with medium depth corner outfield walls and short left field walls, giving balls a fair chance to get out of the park. The number two park is Boston, which has a very short (but tall) LF wall and has Pesky Pole in RF. This allows a lot of balls to also get out the park, boosting the runs scored. San Francisco, on the other hand, is at close to sea level and has fairly deep and tall walls with a large center field, keeping balls in the park. Most of the lower R/9 parks are near sea level: San Francisco, Seattle, San Diego, Los Angeles, Tampa Bay, Miami, and Oakland.

2. The RE (run expectancy) for the bases loaded with 1 out at Coors Stadium is 0.83 runs. The RE for bases loaded with 1 out at Petco Park (Padres Stadium) is 0.74 runs.

This follows a similar line of reasoning to question 1, high elevation parks have more balls carry out of the stadium, while low elevation parks have less carry. San Diego has the 3rd lowest elevation ballpark with a big outfield, so it's hard to hit grand slams in this situation.

3. Being clutch isn't just about hitting well in high pressure situation, but hitting better in high pressure situations. If a player A is a lifetime .320 hitter, but his .310 in high pressure situations, he's not particularly clutch. But player B might be a lifetime .260 hitter, but hits .310 in high pressure situations, meaning he's extremely clutch.

For this, I am going to look at a players slash line (BA/OBP/SLG) in normal situations versus in clutch situations (later than the 6th inning, with runners in scoring position in a game which is within 2 runs in either direction). This gives us a 'Clutch Value' of:

$$\textit{Clutch Value} = (\textit{Clutch BA} - \textit{BA}) + (\textit{Clutch OBP} - \textit{OBP}) + (\textit{Clutch SLG} - \textit{SLG})$$

Where the higher the clutch value, the better a player does in high pressure situations. The clutch value is weighted to represent the impact of the inning, where the 7th inning is weighted to 0.8, 8th inning to 1, 9th+ inning to 1.2 (still averages to ~1).

To do this, I computer the general BA, OBP, and SLG for each player throughout the data set using the following query creating a temporary table:

```
#BA, OBP, SLG for all players over all PA
drop table if exists batting_stats;
create table batting_stats as
        select
```

```
                BAT_ID,
        round((sum(case when EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0
end)),3) as 'BA',
        round(((sum(case when EVENT_CD > 19 then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0
end)+sum(case EVENT_CD when 15 then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end))/(sum(case
AB_FL when 'T' then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15
then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end)+sum(case SF_FL when 'T' then 1 else 0
end))),3) as 'OBP',
        round(((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end))),3) as 'SLG',
                sum(case when EVENT_CD > 19 then 1 else 0 end) as H,
                sum(case EVENT_CD when 20 then 1 else 0 end) as '1B',
                sum(case EVENT_CD when 21 then 1 else 0 end) as '2B',
                sum(case EVENT_CD when 22 then 1 else 0 end) as '3B',
                sum(case EVENT_CD when 23 then 1 else 0 end) as 'HR',
                sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15 then 1 else 0 end)
as 'BB',
                sum(case EVENT_CD when 16 then 1 else 0 end) as 'HBP',
                sum(case SF_FL when 'T' then 1 else 0 end) as 'SF',
                sum(case AB_FL when 'T' then 1 else 0 end) as 'AB'
                from events
        group by BAT_ID;
```

Then I did the same calculations for clutch situations (later than the 6th inning, RISP (runners in scoring position), game within two runs) using the following queries with their appropriate weights for their inning:

#BA, OBP, SLG for clutch appearances in the 7th inning

```
drop table if exists clutch_batting_stats7;
create table clutch_batting_stats7 as
        select
                BAT_ID,
        ((sum(case when EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0 end))*0.8)
as 'C_BA7',
        (((sum(case when EVENT_CD > 19 then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0
end)+sum(case EVENT_CD when 15 then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end))/(sum(case
AB_FL when 'T' then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15
then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end)+sum(case SF_FL when 'T' then 1 else 0
end)))*0.8) as 'C_OBP7',
                (((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end)))*0.8) as 'C_SLG7',
        sum(case AB_FL when 'T' then 1 else 0 end) as 'C_AB7'
                from events
        where INN_CT = 7 and
    (BASE2_RUN_ID <> '' or BASE3_RUN_ID <> '') and
    AWAY_SCORE_CT - HOME_SCORE_CT < 3 and AWAY_SCORE_CT - HOME_SCORE_CT > -3
        group by BAT_ID;
```

#BA, OBP, SLG for clutch appearances in the 8th inning

```
drop table if exists clutch_batting_stats8;
create table clutch_batting_stats8 as
        select
                BAT_ID,
        (sum(case when EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0 end)) as
'C_BA8',
```

```
        ((sum(case when EVENT_CD > 19 then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0
end)+sum(case EVENT_CD when 15 then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end))/(sum(case
AB_FL when 'T' then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15
then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end)+sum(case SF_FL when 'T' then 1 else 0
end)))*0.8 as 'C_OBP8',
                ((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end))) as 'C_SLG8',
        sum(case AB_FL when 'T' then 1 else 0 end) as 'C_AB8'
                from events
        where INN_CT = 8 and
    (BASE2_RUN_ID <> '' or BASE3_RUN_ID <> '') and
    AWAY_SCORE_CT - HOME_SCORE_CT < 3 and AWAY_SCORE_CT - HOME_SCORE_CT > -3
        group by BAT_ID;
```

#BA, OBP, SLG for clutch appearances in innings after the 8th

```
drop table if exists clutch_batting_stats9;
create table clutch_batting_stats9 as
        select
                BAT_ID,
        ((sum(case when EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0 end))*1.2)
as 'C_BA9',
        (((sum(case when EVENT_CD > 19 then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0
end)+sum(case EVENT_CD when 15 then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end))/(sum(case
AB_FL when 'T' then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15
then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end)+sum(case SF_FL when 'T' then 1 else 0
end)))*1.2) as 'C_OBP9',
                (((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end)))*1.2) as 'C_SLG9',
        sum(case AB_FL when 'T' then 1 else 0 end) as 'C_AB9'
                from events
        where INN_CT > 8 and
    (BASE2_RUN_ID <> '' or BASE3_RUN_ID <> '') and
    AWAY_SCORE_CT - HOME_SCORE_CT < 3 and AWAY_SCORE_CT - HOME_SCORE_CT > -3
        group by BAT_ID;
```

Then I combine all of the clutch tables into a singular clutch table:

#Compiling the clutch stats with their appropriate weights

```
drop table if exists clutch_batting_stats;
create table clutch_batting_stats as
        select distinct
                cbs7.BAT_ID,
        round((cbs7.C_BA7 + cbs8.C_BA8 + cbs9.C_BA9)/3,3) as 'C_BA',
        round((cbs7.C_OBP7 + cbs8.C_OBP8 + cbs9.C_OBP9)/3,3) as 'C_OBP',
        round((cbs7.C_SLG7 + cbs8.C_SLG8 + cbs9.C_SLG9)/3,3) as 'C_SLG',
        cbs7.C_AB7 + cbs8.C_AB8 + cbs9.C_AB9 as 'C_AB'
        from clutch_batting_stats7 as cbs7
        join clutch_batting_stats8 cbs8 on cbs8.BAT_ID = cbs7.BAT_ID
    join clutch_batting_stats9 cbs9 on cbs9.BAT_ID = cbs7.BAT_ID;
```

Then I calculated the *Clutch Value* score as described above with the conditions that the player has at least 250 AB in non-clutch situations and 50 AB in clutch situations (any more and the number of players that qualified was too small) using the following query:

#Calculating the Clutch Value and Normalized Clutch Value for all players with 250 normal PA and 50 'Clutch' PAs

```
select distinct r.FIRST_NAME_TX as 'First Name', r.LAST_NAME_TX as 'Last Name', bs.BA, cbs.C_BA as 'Clutch
BA', bs.OBP, cbs.C_OBP as 'Clutch OBP', bs.SLG, cbs.C_SLG as 'Clutch SLG',
round(((cbs.C_BA-bs.BA)+(cbs.C_OBP-bs.OBP)+(cbs.C_SLG-bs.SLG)),3) as 'Clutch Value',
```

```
round(((((cbs.C_BA-bs.BA)+(cbs.C_OBP-bs.OBP)+(cbs.C_SLG-bs.SLG))-(-0.015420001))/(0.419383332+0.015420001)
),3) as 'Normalized Clutch Value'
from batting_stats bs
join clutch_batting_stats cbs on cbs.BAT_ID = bs.BAT_ID
join rosters r on r.PLAYER_ID = bs.BAT_ID and r.PLAYER_ID = cbs.BAT_ID
where bs.AB > 250 and cbs.C_AB > 50
order by ((cbs.C_BA-bs.BA)+(cbs.C_OBP-bs.OBP)+(cbs.C_SLG-bs.SLG)) DESC;
```

The output results in the following players:

First Name	Last Name	BA	Clutch BA	OBP	Clutch OBP	SLG	Clutch SLG	Clutch Value	Normalized Clutch Value
Angel	Pagan	0.278	0.340	0.330	0.390	0.413	0.711	0.420	1.001
Michael	Bourn	0.275	0.396	0.339	0.464	0.371	0.513	0.388	0.928
Joey	Votto	0.316	0.391	0.433	0.585	0.540	0.699	0.386	0.923
Coco	Crisp	0.265	0.408	0.327	0.404	0.418	0.549	0.351	0.843
Ryan	Zimmerman	0.288	0.335	0.358	0.494	0.478	0.572	0.277	0.673
Ryan	Raburn	0.253	0.321	0.313	0.383	0.440	0.574	0.272	0.661
Greg	Dobbs	0.257	0.315	0.301	0.397	0.361	0.474	0.267	0.650
Raul	Ibanez	0.252	0.328	0.315	0.380	0.450	0.568	0.259	0.631
Jimmy	Rollins	0.255	0.310	0.324	0.419	0.389	0.490	0.251	0.613
Carlos	Lee	0.261	0.263	0.321	0.374	0.410	0.601	0.246	0.601
Danny	Valencia	0.263	0.386	0.302	0.376	0.411	0.459	0.245	0.599
Jason	Kipnis	0.269	0.309	0.348	0.351	0.423	0.622	0.242	0.592
Prince	Fielder	0.284	0.340	0.392	0.506	0.496	0.565	0.239	0.585
Carlos	Ruiz	0.291	0.360	0.371	0.406	0.430	0.563	0.237	0.581
Travis	Hafner	0.253	0.308	0.350	0.445	0.432	0.517	0.235	0.576
Brandon	Inge	0.224	0.247	0.290	0.330	0.356	0.515	0.222	0.546
James	Loney	0.278	0.319	0.330	0.446	0.398	0.462	0.221	0.544
Andres	Torres	0.246	0.315	0.325	0.385	0.385	0.476	0.220	0.541
Adrian	Gonzalez	0.308	0.318	0.373	0.482	0.498	0.594	0.215	0.530
Brandon	Belt	0.269	0.316	0.348	0.396	0.443	0.562	0.214	0.528
Casey	Kotchman	0.250	0.335	0.312	0.407	0.360	0.394	0.214	0.528
Chris	Davis	0.268	0.297	0.337	0.416	0.519	0.621	0.210	0.518
Miguel	Cabrera	0.334	0.373	0.424	0.509	0.608	0.694	0.210	0.518
Jason	Heyward	0.257	0.317	0.348	0.452	0.439	0.483	0.208	0.514
Matthew	Joyce	0.248	0.265	0.339	0.382	0.445	0.589	0.204	0.505
Ryan	Howard	0.257	0.321	0.335	0.481	0.476	0.470	0.204	0.505
Nate	Schierholtz	0.257	0.320	0.312	0.424	0.426	0.453	0.202	0.500

And the bottom few players:

First Name	Last Name	BA	Clutch BA	OBP	Clutch OBP	SLG	Clutch SLG	Clutch Value	Normalized Clutch Value
Elvis	Andrus	0.275	0.302	0.341	0.322	0.341	0.317	-0.016	-0.001
Jamey	Carroll	0.271	0.277	0.345	0.358	0.321	0.287	-0.015	0.001
John	Mayberry	0.246	0.232	0.306	0.343	0.426	0.389	-0.014	0.003
Ty	Wigginton	0.239	0.218	0.310	0.319	0.397	0.395	-0.014	0.003
Alex	Gordon	0.279	0.250	0.352	0.369	0.447	0.448	-0.011	0.010
Lance	Berkman	0.271	0.240	0.382	0.400	0.459	0.462	-0.010	0.012
Albert	Pujols	0.293	0.282	0.370	0.509	0.535	0.402	-0.005	0.024

There are a few things that we can take away from these calculations:

- Angel Pagan (CF) is not necessarily the most “overall” clutch player. Instead, he is the player that improves the most in clutch situations. He’s far past his prime anymore at 36 years old, but in the early 2010s he was an alright player with a 4.5 oWAR in 2012 (All-Star level offensive production), but was often a liability in the field, especially as he got older (a lot of negative dWAR seasons).

- Many of the top Clutch Value players are known for their excellent offensive and batting approach. Joey Votto, Ryan Zimmerman, and Jason Kipnis are all excellent hitters and apparently get even better in clutch situations.
- At the bottom of the list, we have relatively no-name players, although Albert Pujols makes an appearance which is fairly surprising considering he was still an offensive powerhouse in the early 2010s.
- There are more positive Clutch Value players than negative, which makes sense because of the fact that you don't' want bad clutch hitters to get more ABs, and so they probably do not qualify for this metric in a PA sense.

The next question I would ask is if there is a relationship between past clutch performance and future clutch performance? Is clutchness a fairly random stat that waxes and wanes throughout a players career or are some hitter just absolutely more clutch than others? You could also perform the same analysis with pitchers, but SPs wouldn't get a fair representation considering they do not get to the 7th inning very often with the emphasis on pitch count and inning limitations.