Assignment 1

Please use the Pandas and Numpy libraries to answer the question I posed in the document. You will create a markdown cell with the question in it and then answer the question in the next cell.

1. Load in the appropriate csv file as a pandas dataframe (batting.csv)
2. Print out the dimensions and info about the dataframe you just created

Info() and describe() methods

For next part I am going to ask you a series of question, for some I am going to want to simply see a number. For others I will want to see the entire dataframe. I will make it very clear which is which.

1. How many players have hit 40 or more HRs in one single season? (Number only)
2. How many players have hit more than 600 HRs for their career? (Dataframe with name and player first and last name only and amount of homers)
3. How many players have hit 40 2Bs, 10 3Bs, 200 Hits, and 30 HRs (inclusive) in one season? (Number Only)
4. How many players have had 100 or more SBs in a season? (Dataframe)
5. How many players in the 1960s have hit more than 200 HRs? (Dataframe)

1960 to 1969. Hint use aggregation function. i.e. sum.

1. Who has hit the most HRs in history? (Dataframe)
2. Who had the most hits in the 1970s? (Dataframe)
3. Top 5 highest OBP (on base percentage) with at least 500 PAs in 1977? (Dataframe)

OBP = (hits + walks + hit by pitch) / (at bats + walks + hit by pitch + sacrifice flies)

1. Top 8 highest averages in 2013 with at least 300 PAs? (Dataframe)
2. Leaders in hits from 1940 up to and including 1949. (Dataframe)
3. Who led MLB with the most hits the most times? And how many times? (Dataframe, Number)
4. Which players have played the most games for their careers?  Top 5, descending by games played presented as a dataframe
5. How many players have had more than 3000 hits for their careers while also hitting 500 or more HRs?  Just a number is okay here
6. How many HRs were hit during the entire 1988 season?  Just a number is okay here
7. Please filter out and show me the top 3 average seasons by Wade Boggs during his career in seasons in which he had at least 500 ABs.  I would like a dataframe sorted by average.
8. Please filter out the top OBPs for the 1995 season with at least 400 PAs, sorted by OBP.  I would like a dataframe for this
9. Who had the most 3Bs (in total) in 1922, 1925, 1926, and 1928?  I would like a dataframe with just the leader
10. How many players have hit 30 or more HRs in season while also stealing (SB) 30 more or bases?  A number is okay here
11. Who had the highest OBP is 1986 with at least 400 PAs? (Dataframe)
12. Same question but for 1997 and only in the NL (check league ID)? (Dataframe)
13. Who had more than the league average HRs in 2012 (filter out all players with less 500 PAs)? (Dataframe)
14. Who is the youngest player to hit 50 or more HRs in a single season? (Dataframe)
15. Who are the five youngest players to hit 300 or more HRs for their career? (Dataframe)

Grading: Each task will be worth 3.5 pts (there will be 25 total).  Naming convention and github link will be worth 6.25 pts apiece.  I will include the two bonus questions which will be worth 1 pt apiece (they will be simple viz questions)

BONUS:  Graph total HRs per season using bar graph

Using a line graph please graph the average HRs per AB (think about this) per season