Sean McLean

ALY 6000

Project 2

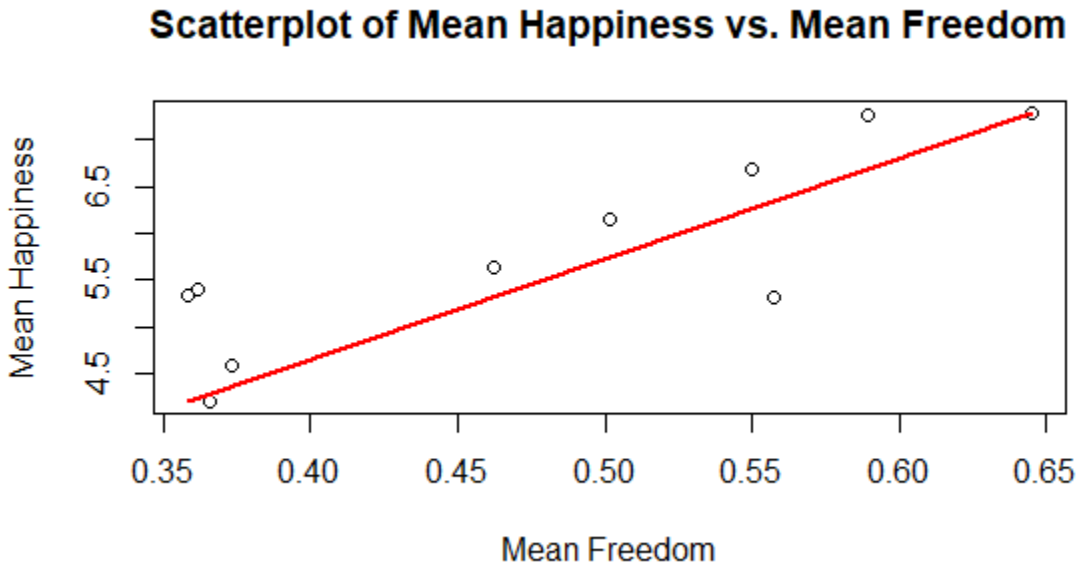
10/01/2023

Introduction

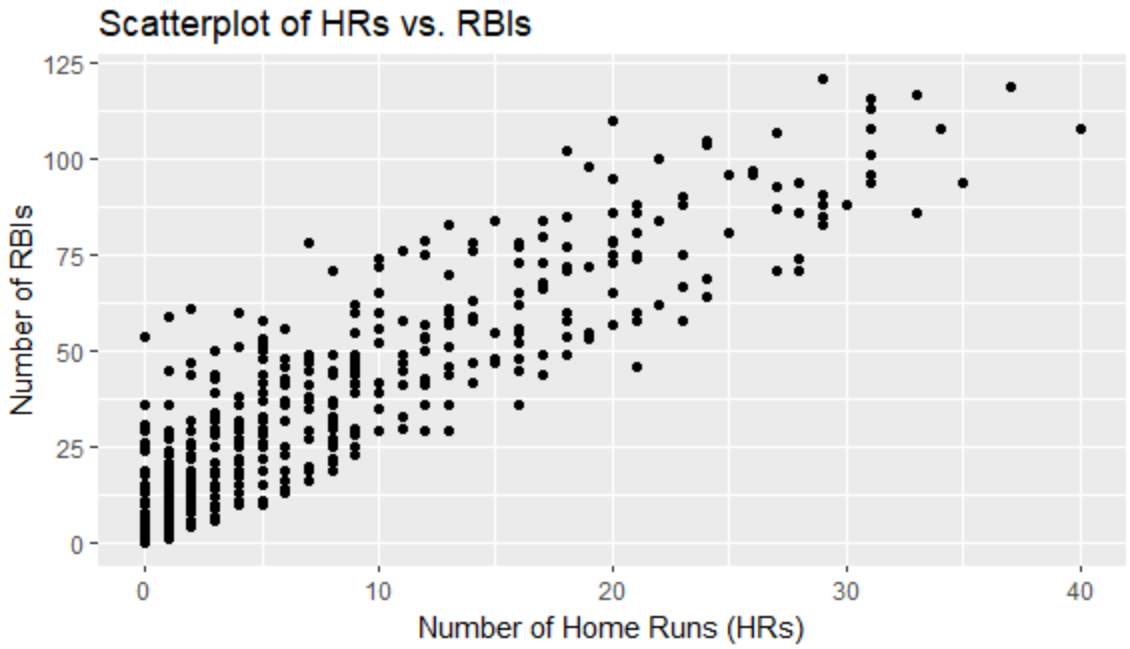
The second project consists of two datasets where data cleaning and analysis are two of the focal points toward the completion. The first part of the assignment evaluated a dataset on countries gross domestic product and several aspects of human life including happiness and freedom. The second part of the project entailed the dissection of all hitters that played in the 1986 Major League Baseball season.

Key Findings

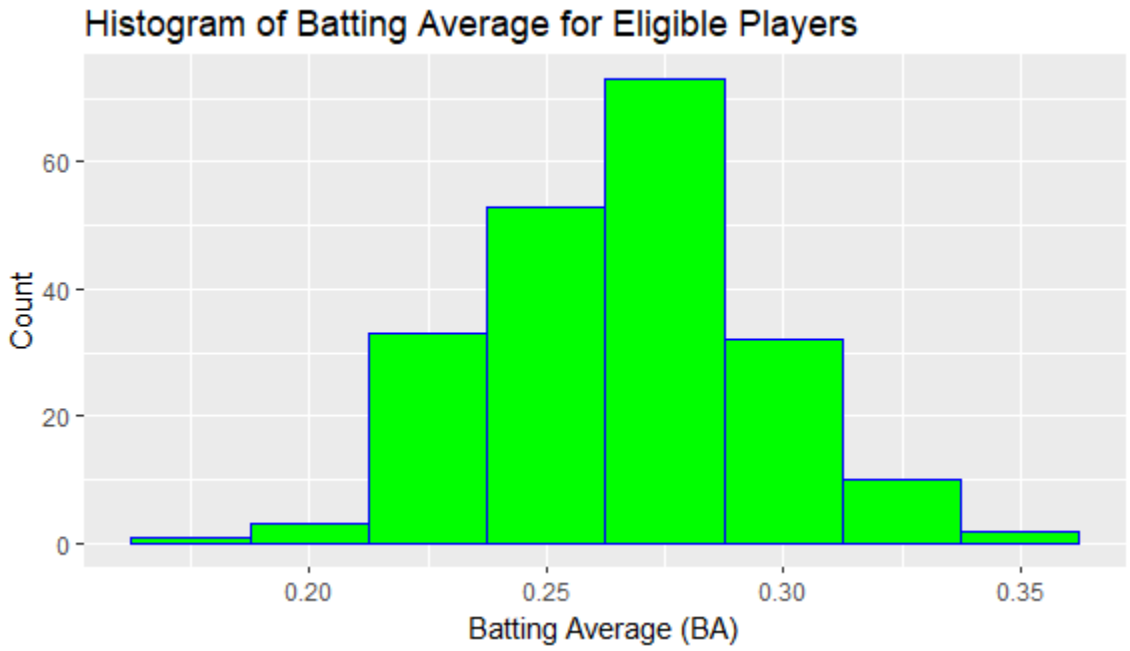
The main takeaway from part one of the assignment is the strong correlation between mean happiness and mean freedom that is evidenced in the scatterplot below. From the countries that were included in the dataset and were analyzed, the evidence points to the fact that the more freedom a nation has, the more happiness they will have overall. When narrowed down to the top 10 countries on the planet, most of the countries were in Western Europe, with Australia and New Zealand as well as Canada in North America also appearing on the list.



The second part of the assignment delves into the debate to who is the best hitter in Major League Baseball in 1986. Its breaks down the dataset into several key statistics including home runs and hits and creates several new categories like batting average and on-base percentage to show who was the most valuable. It also filters down the dataset to hitters who would only qualify statistically for several of the major categories including batting average. The following plot shows how many runs batted in and home runs were hit for each hitter in the database.



The plot shows a direct correlation of more home runs a player will hit usually will mean more runs batted in. The top home run hitter had 40 home runs and is also among the top leaders in runs batted in. The next plot below shows the total batting average among all the hitters in Major League Baseball in 1986. The histogram is a bell-shaped curve with the most hitters being around .275 with a lot fewer hitters above that average and more below.



Recommendations

Looking at the overall picture I would make the recommendation of giving Don Mattingly the most valuable player award for 1986. What stands out the most looking at his statistics is that he is the only hitter that is in the top 10 in home runs, runs batted in, and on-base percentage on the list. He is also first in hits and doubles and second in runs scored and home runs. While he is surprisingly only 16th of the top 20 in the top rankings for walks which is surprising to me for having a great on-base percentage, he strikes out the least by far. This would indicate to me that he puts the ball in play more than other hitters.

Conclusion

The analysis and data visualizations from the first part of the project would indicate that people are happier when they have more freedoms and basic human rights. This would make nations more attractive in other ways like its economy and government that would be beneficial. The second part of the assignment shows who were the best hitters for one season and Don Mattingly would across the board be the most valuable because he gets on base and has higher quality at-bats because he doesn't strike out very much.

Works Cited

Data Manipulation in R. (2018, October 19). *Select Data Frame Columns in R.* Datanovia. https://www.datanovia.com/en/lessons/select-data-frame-columns-in-r/

Naveen. (2022, July 2). *R Filter Dataframe by Column Value.* Spark By {Examples}. https://sparkbyexamples.com/r-programming/r-filter-dataframe-by-column-value/#:~:text=You%20can%20do%20this%20by,frame%20and%20return%20data.

Chat GPT. (2023, October 5th). *Default (GPT-3.5).* https://chat.openai.com/c/95a2395f-ef31-4c10-a835-7d8c0550b49a

R Scatter Plot. (2023, October 5th). *Scatter Plots.* W3 Schools. https://www.w3schools.com/r/r_graph_scatterplot.asp

Zach. (2021, July 29th). *How to Round Numbers in R (5 Examples).* Statology. https://www.statology.org/round-in-r/#:~:text=You%20can%20use%20the%20following,values%20up%20to%20nearest%20integer.

Learn R. (2023, February). *How to Make a Histogram in Base R: 6 Steps with Examples.* Datacamp. https://www.datacamp.com/tutorial/make-histogram-basic-r