Research Milestone #3

Your submission should include the following **two** files:

1. PDF file that includes a 500-1000 word writeup with the following:

Objectives and goals.

- What do you want to learn and accomplish?
  - Try to make your objective/hypothesis as specific as possible ("I think republicans earn more money than democrats because they are more business oriented" vs. "I want to examine why some people make more money that others")
- Outline the relationships you will explore and how they will help you achieve your objective

Five Visualizations that will include the following elements:

a) Bar plot
b) Density plot or histogram
c) Box plot
d) line or scatter plot
e) Color mapping
f) Faceting

Additional Requirements:

- Visualizations are appropriate for the variable types
- Variables have been cleaned (levels collapsed)
- Visualizations use at least six variables from your dataset
- Visualizations include at least two variables in each visualization

Discussion

For each visualization, address the following:

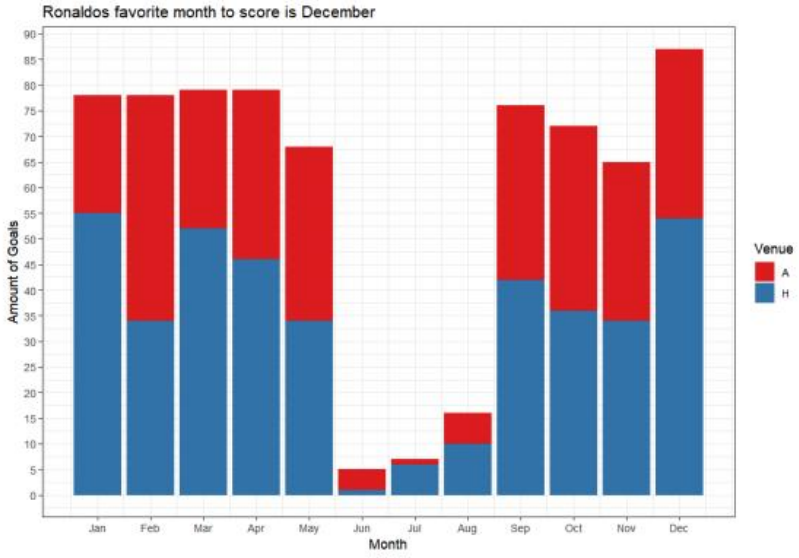
1. What relationship are you exploring with the visualization?
2. What is the hypothesized relationship between the variables in your visualization?
3. Describe your visualizations accurately
4. Identify interesting patterns or trends
5. Provide a possible explanation for the observed pattern, ideally linking to external information
6. Explain how the observed relationships address your overall research question/hypothesis

Generative AI acknowledgement

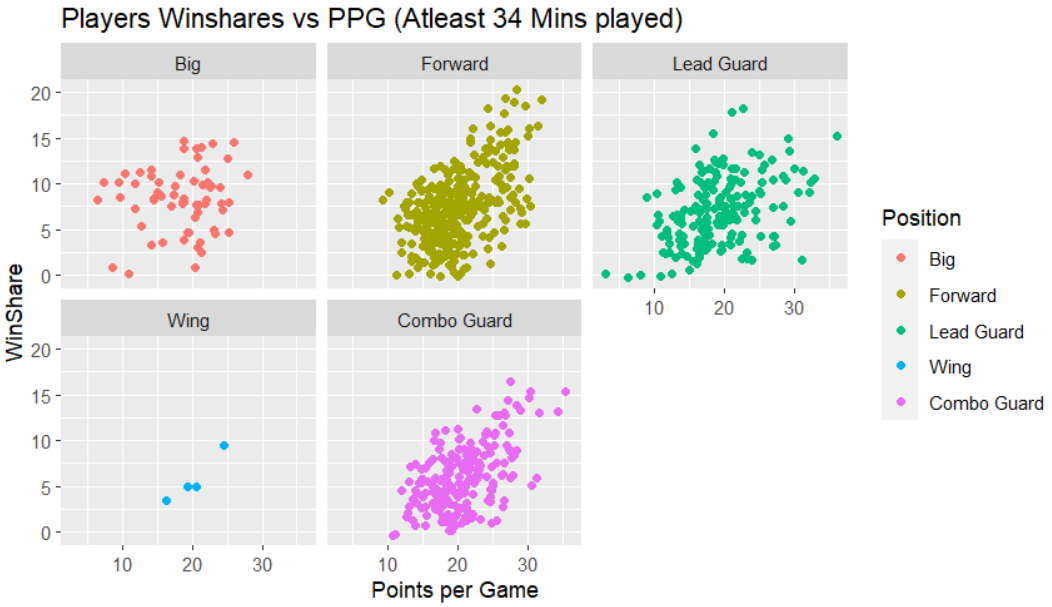
a. If you used generative AI, please explain what you used it for in a few sentences.
b. If you did not use generative AI, please include the following statement: *I did not use generative AI for any portion of this assignment. I understand that not disclosing the use of generative AI is a violation of the academic integrity policy.*

2. R script:
    a. Reads in your data (MS # 2)
    b. Cleans your data (MS #2)
    c. Creates your 5 new visualizations

Examples of student work:



I made this bar chart in order to see if I can spot any trend between Ronaldo's goal scoring capabilities between venue type and the month of a year. To a non Soccer viewer this may be confusing since the chart clearly shows that Ronaldo cannot seem to score during the summer. This is explained by summer being the time most seasons and competitions come to an end. Players are usually granted vacation and family time during the three months of June, July, and August leaving little room for games. The small exception to these would be Super Cup games which are mini tournaments played between two teams who won a competition. Overall there appears to be a little bit more blue then then red which allows us to determine that Ronaldo is a better goal scorer when it comes to home games.



The figures above show the relationship between Win Shares per player and how many points per game were scored that season. The scatterplot shows a small positive correlation between scoring more leading to contributing more. When breaking down by position the correlation, Bigs had the most volatility with relation to PPG, while every other position apart from small outliers in the Lead Guard position share the same relationship as the first figure (combined positions). The possible explanation for the Bigs outcome could be due to the fact the Big position usually carries more defensive responsibilities than any other position's offensive duties. There are many cases of dominant Bigs however who do have "high" points per game averages. The outliers in the Lead Guard position could be explained by the tendency for that position leading to more ball handling duties which in return lead to more point making opportunities regardless of win or loss outcome. This is an example of minutes handling the ball correlating with efficiency in output, as it takes more skill to operate with the ball.