

FIFA WORLD CUP 2019 EXTENSIVE Data Analysis By KATTA SRI SAI PRAVEEN

This is a simple web app to analyze the FIFA World Cup 2019 data (assumed FIFA 19 game data).

Upload FIFA 2019 Image



FIFA 2019

IMPORT THE DATA SET

	Unnamed: 0	ID	Name	Age	Photo	Nation
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium

COLUMN NAMES

0
Unnamed: 0
ID
Name
Age
Photo
Nationality
Flag
Overall
Potential
Club

VIEW THE SUMMARY OF THE DATA SET

	Unnamed: 0	ID	Age	Overall	Potential	Special	International Reputation	Weak F
count	18207	18207	18207	18207	18207	18207	18159	18
mean	9103	214298.3386	25.1222	66.2387	71.3073	1597.8099	1.1132	2.9
std	5256.0525	29965.2442	4.6699	6.9089	6.1365	272.586	0.394	0.6
min	0	16	16	46	48	731	1	
25%	4551.5	200315.5	21	62	67	1457	1	
50%	9103	221759	25	66	71	1635	1	
75%	13654.5	236529.5	28	71	75	1787	1	
max	18206	246620	45	94	95	2346	5	

- This dataset contains 89 variables.
- Out of the 89 variables, 44 are numerical variables: 38 are float64, 6 are int64.
- The remaining 45 variables are of character data type.

DELETING A COLUMN

	ID	Name	Age	Photo	Nationality	Flag
0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/argentina.png
1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/portugal.png
2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/brazil.png
3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/spain.png
4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/belgium.png

Body Type Distribution

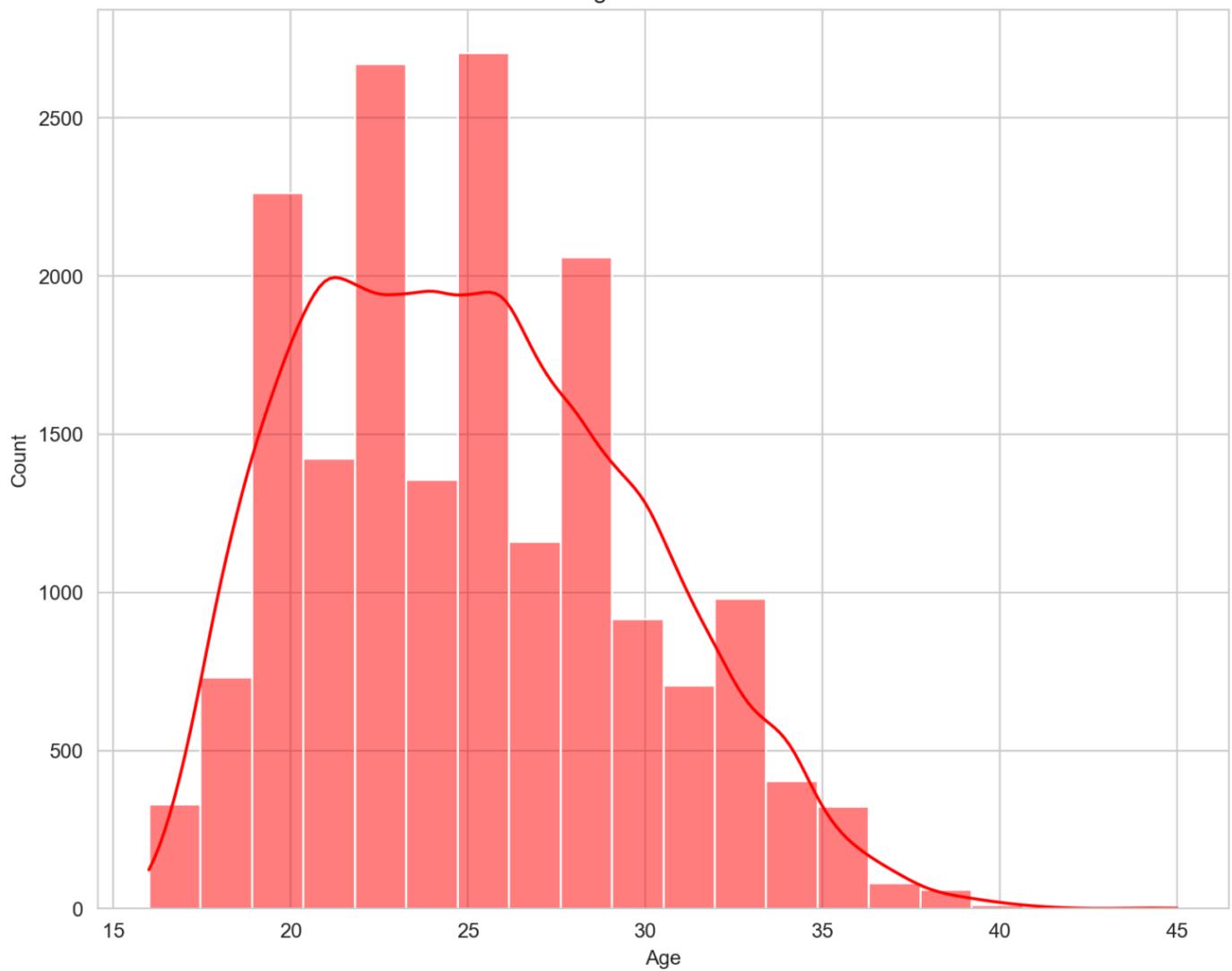
Body Type	count
Normal	10595
Lean	6417
Stocky	1140
Messi	1
C. Ronaldo	1
Neymar	1
Courtois	1
PLAYER_BOI	1
Shaqiri	1
Akinfenwa	1

Explore `Age` variable

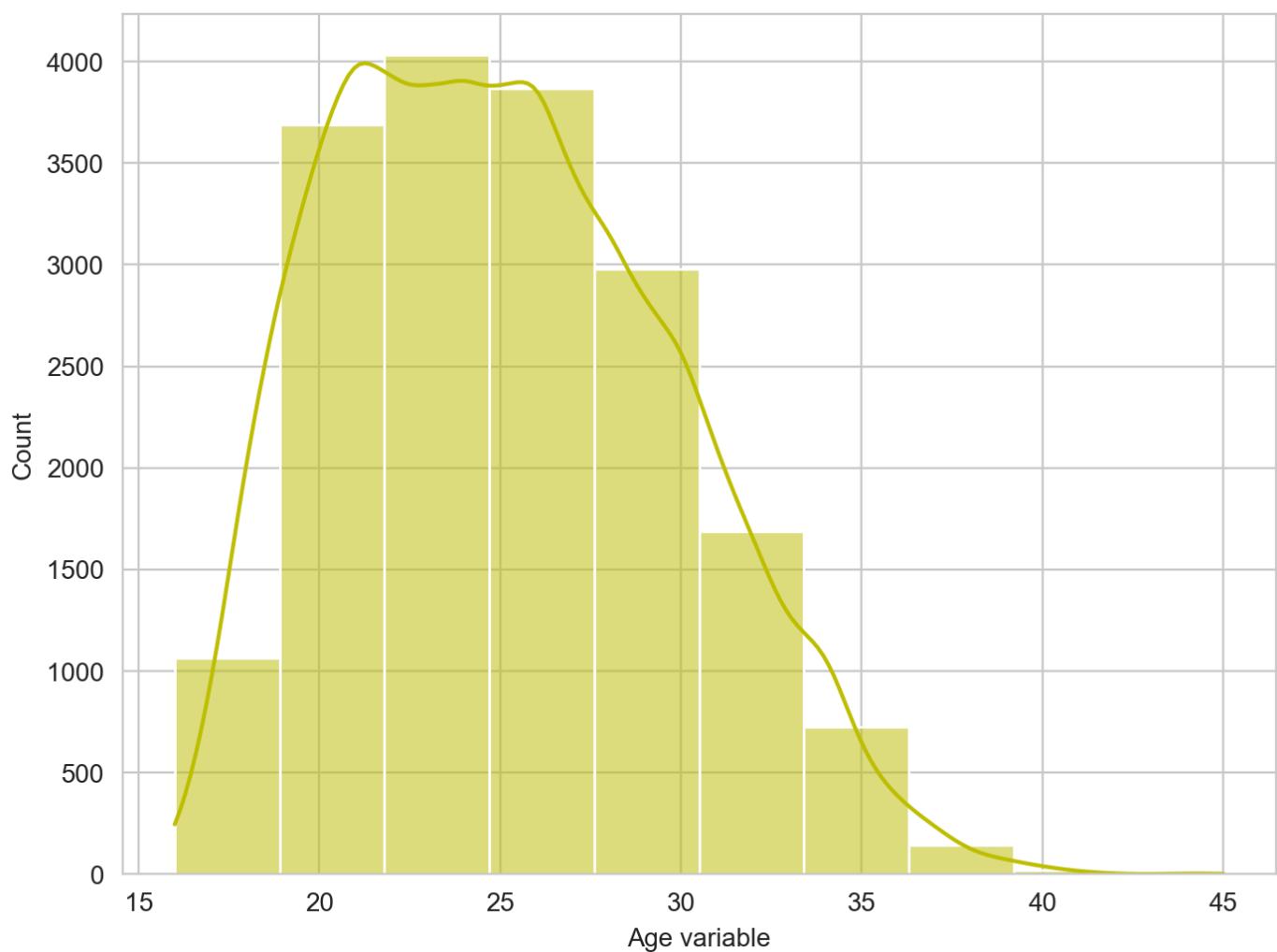
Visualize distribution of `Age` with Seaborn `histplot()`

- Seaborn `histplot()` plots a univariate distribution of observations.
- It combines Matplotlib's histogram with kernel density estimation (KDE).

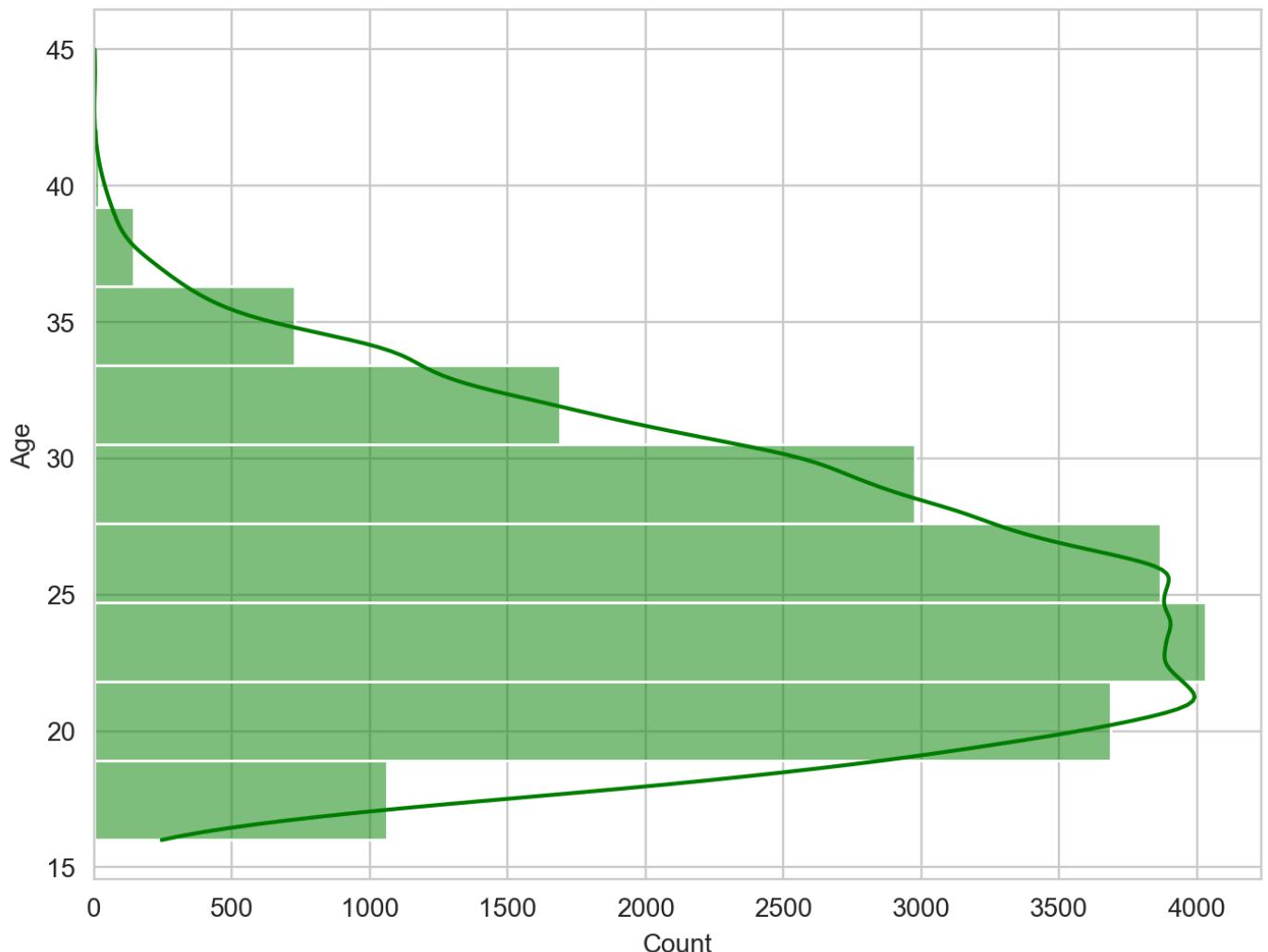
Age Distribution



- The `Age` variable is slightly positively skewed.



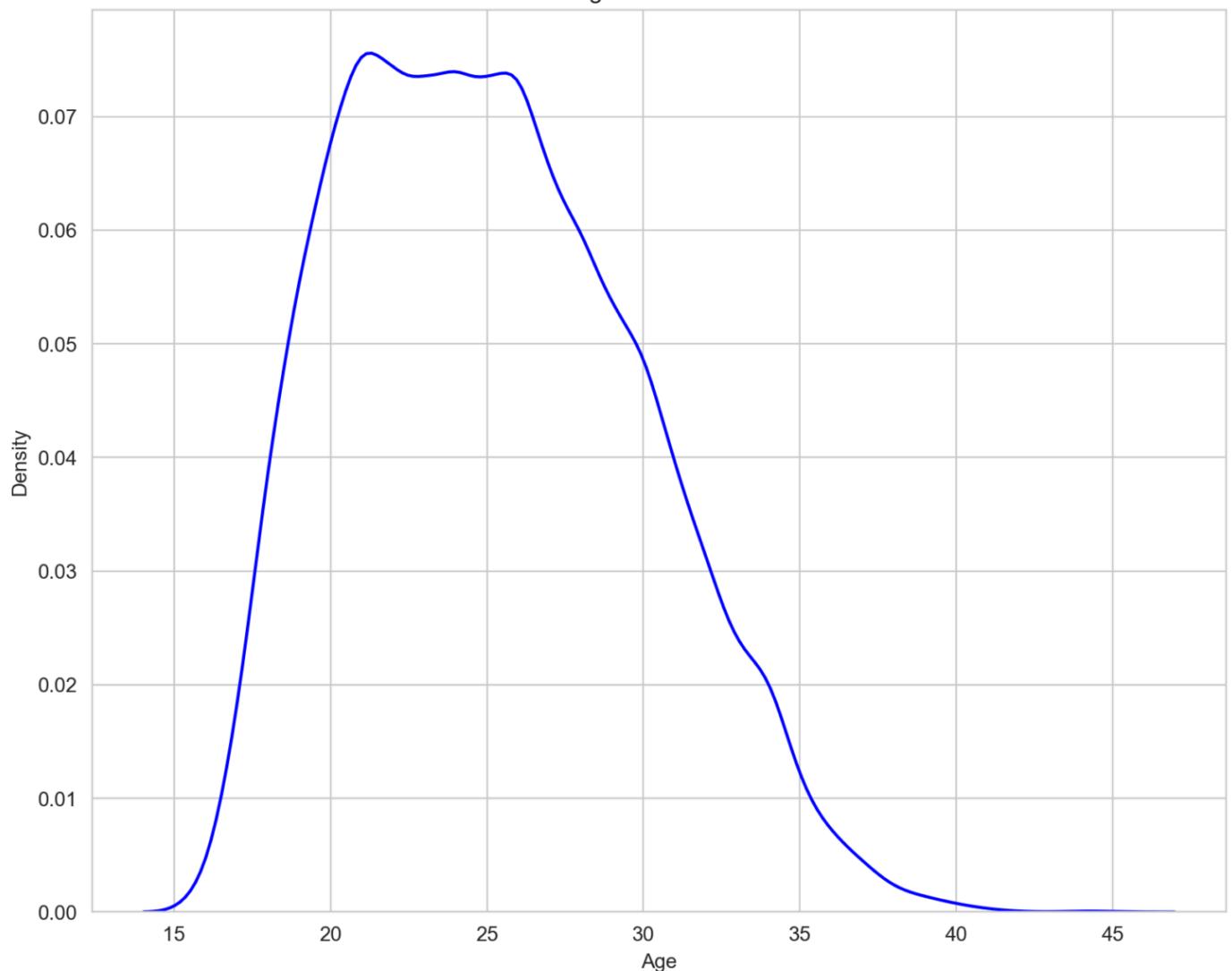
- Plotting distribution on the vertical axis



Seaborn Kernel Density Estimation (KDE) Plot

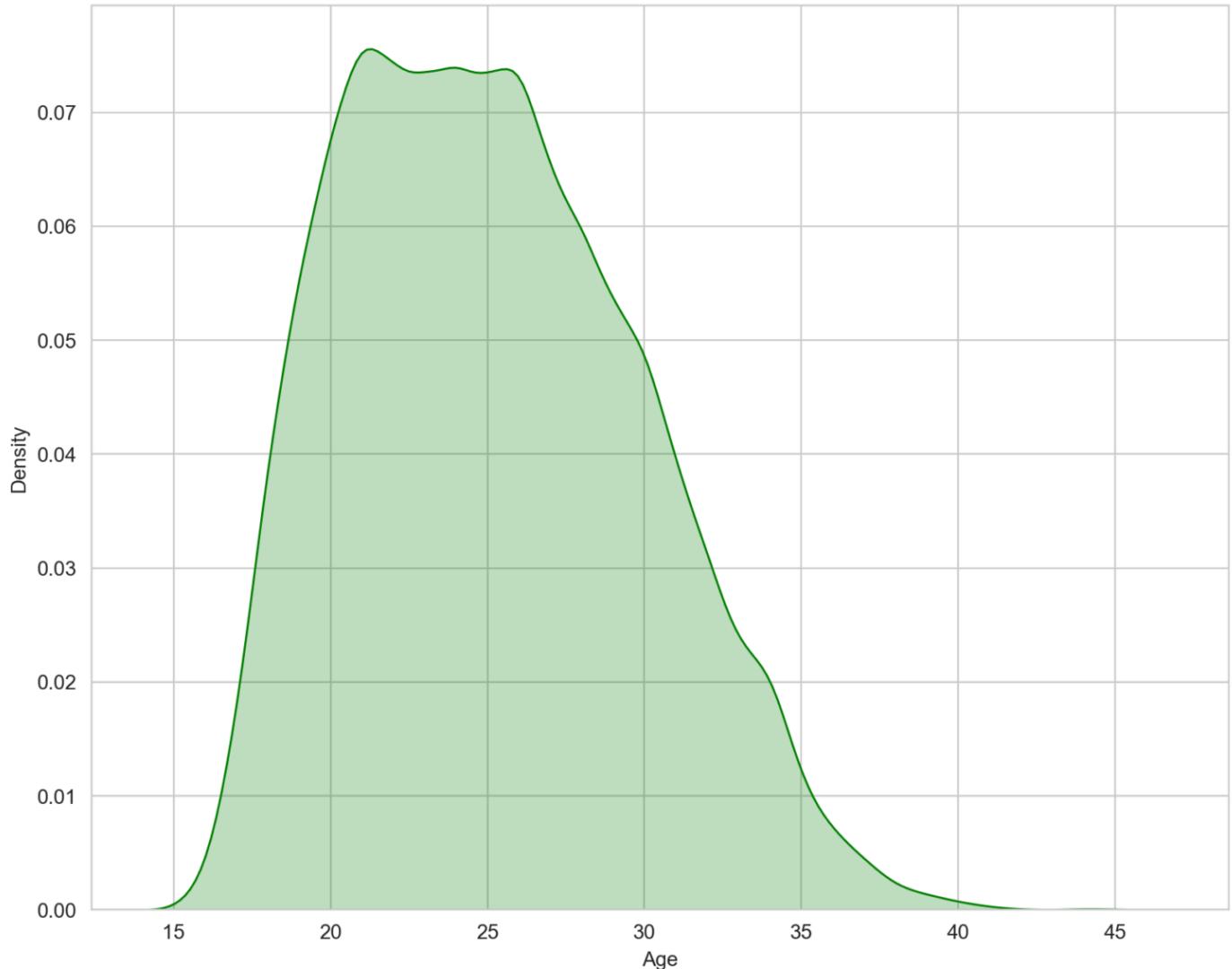
- The KDE plot visualizes the shape of a distribution.
- Seaborn `kdeplot` fits and plots a univariate or bivariate kernel density estimate.

Age Distribution



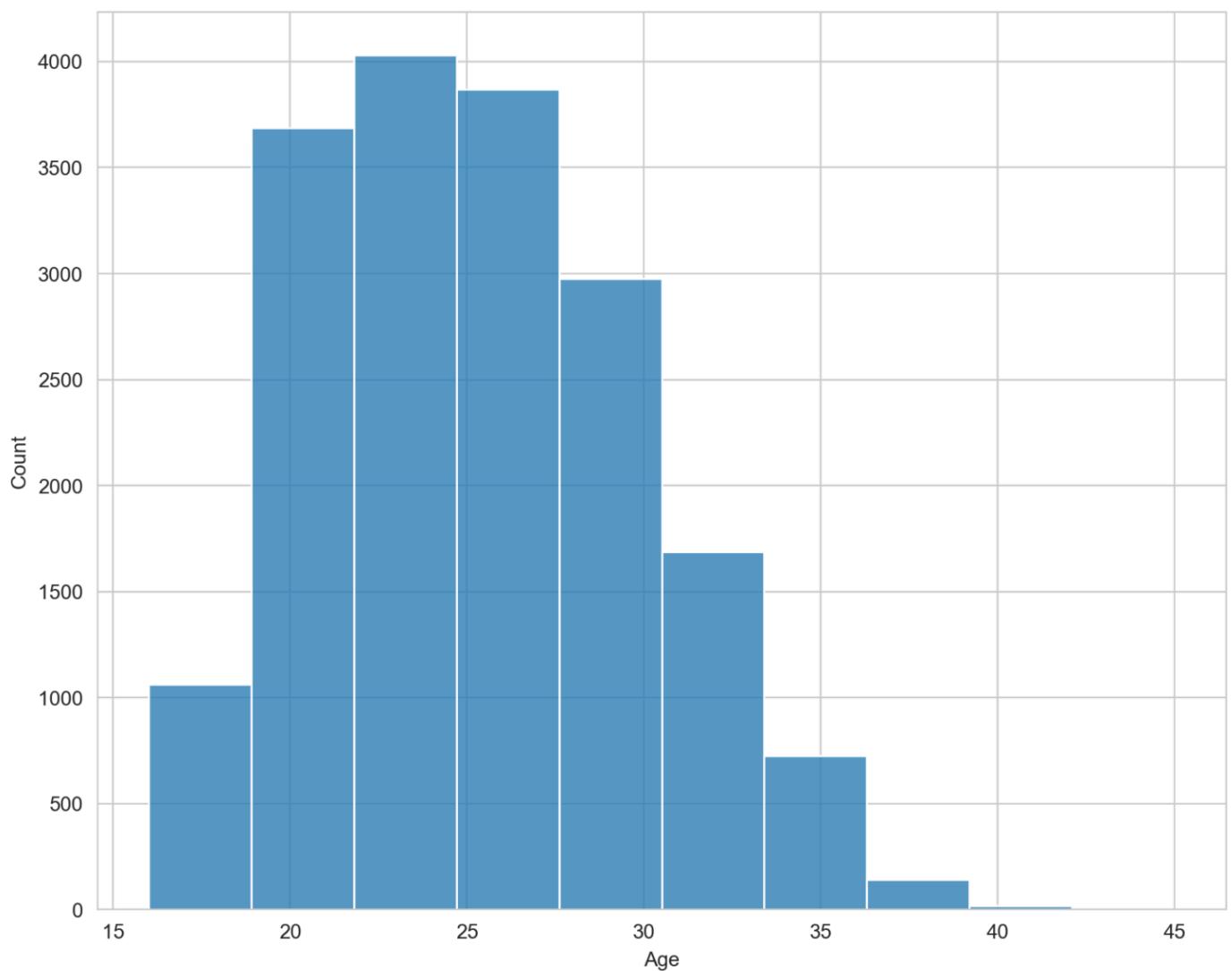
- Shading under the density curve with a different color:

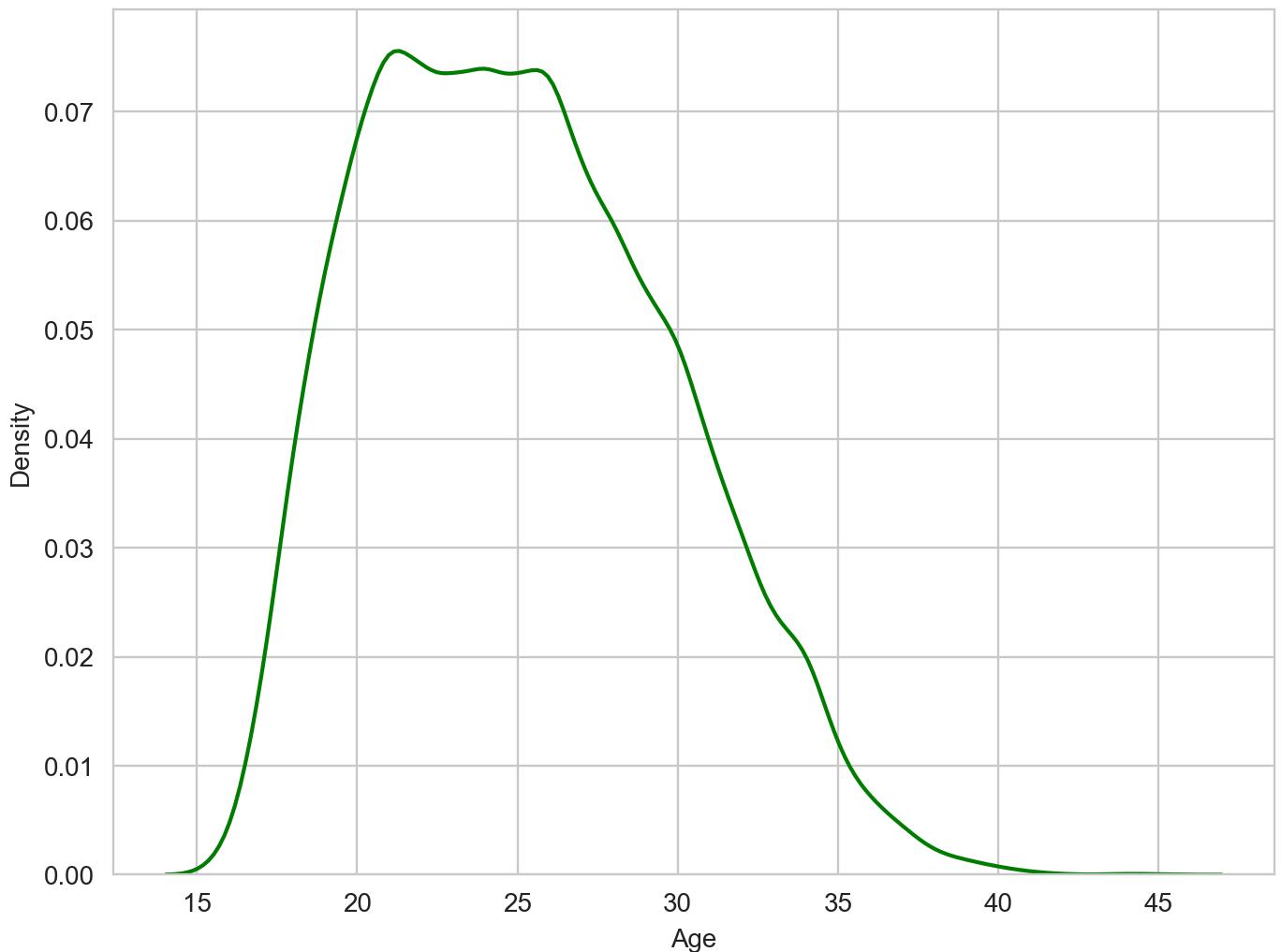
Age Distribution



Histograms

- A histogram shows the distribution of data by forming bins and drawing bars for observation counts.
- Seaborn enhances Matplotlib's `hist()` functionality.





Explore Preferred Foot variable

Number of unique values in Preferred Foot

2

- There are two unique values in Preferred Foot .

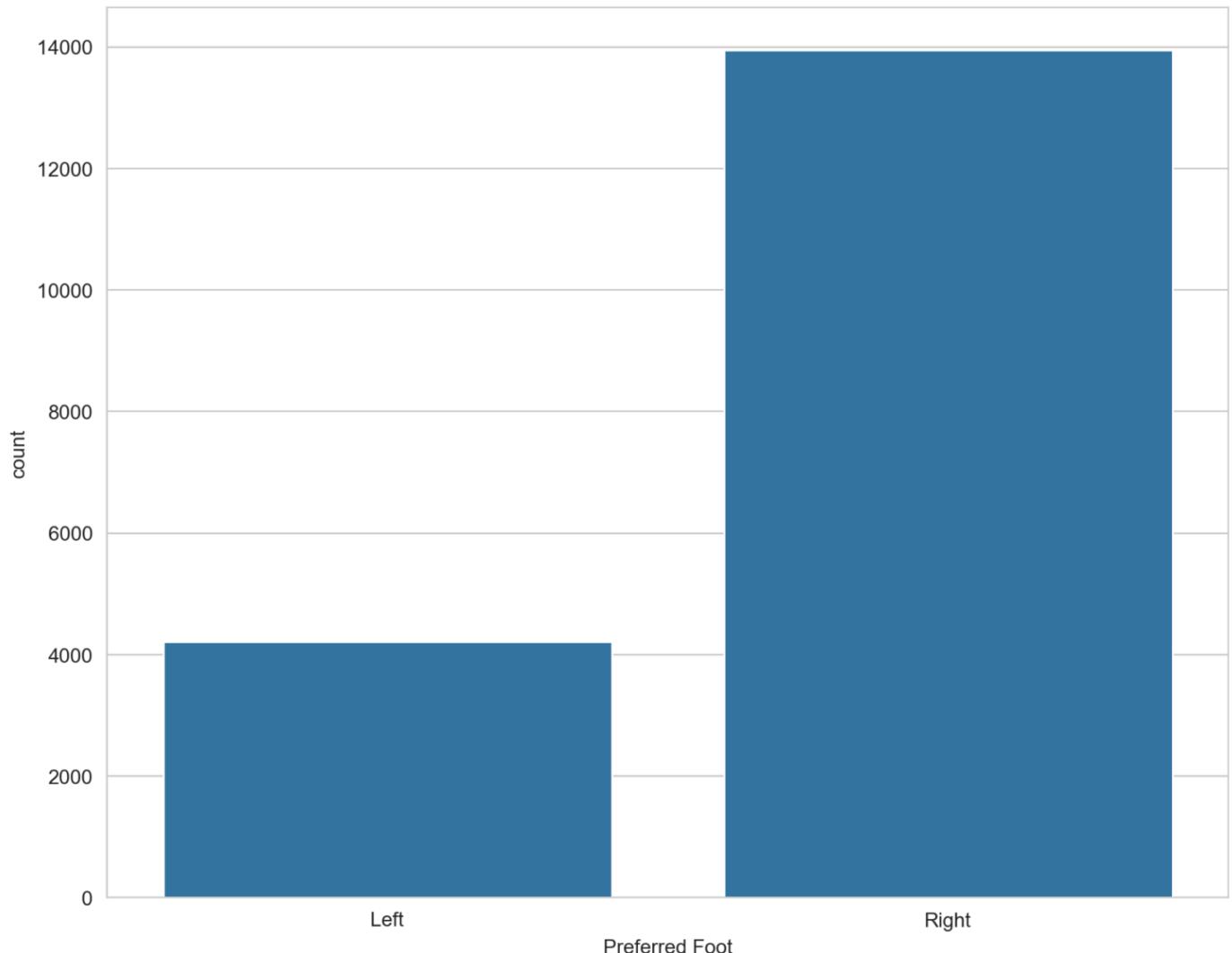
Frequency distribution of Preferred Foot

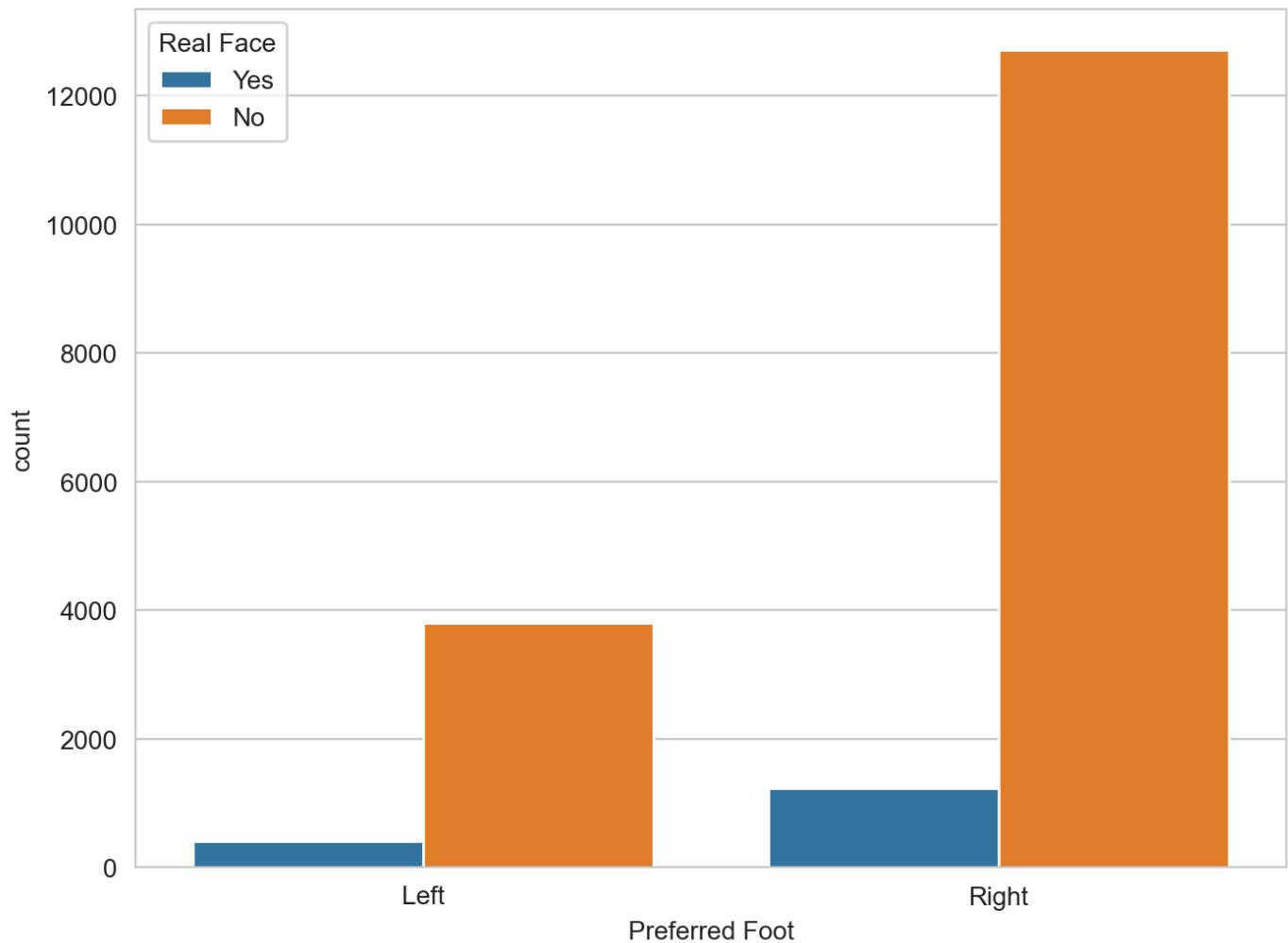
Preferred Foot	count
Right	13948
Left	4211

- Values are Right and Left .

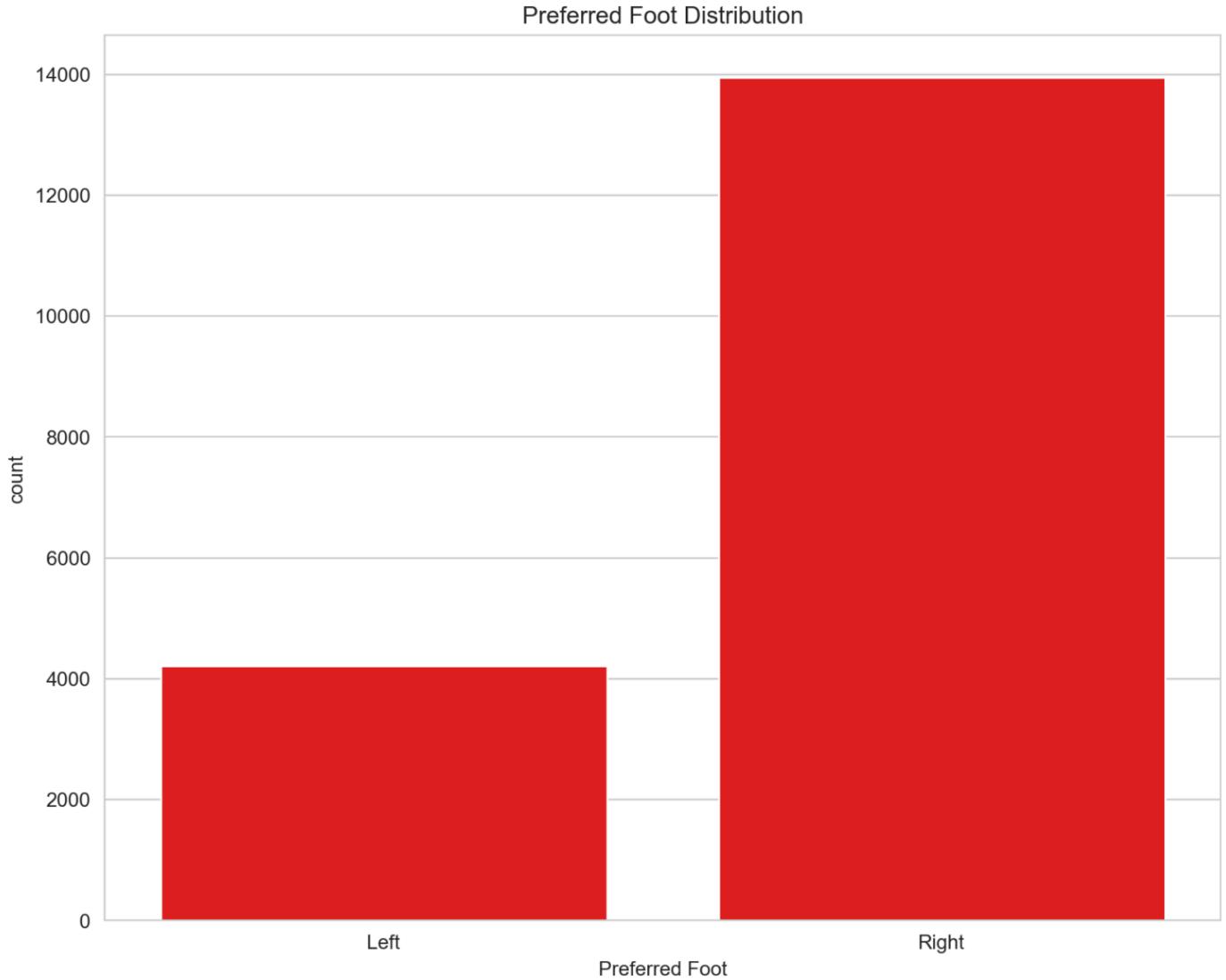
Visualize with Seaborn countplot()

- A `countplot` shows observation counts in categorical bins using bars.
- It's like a histogram for categorical variables.



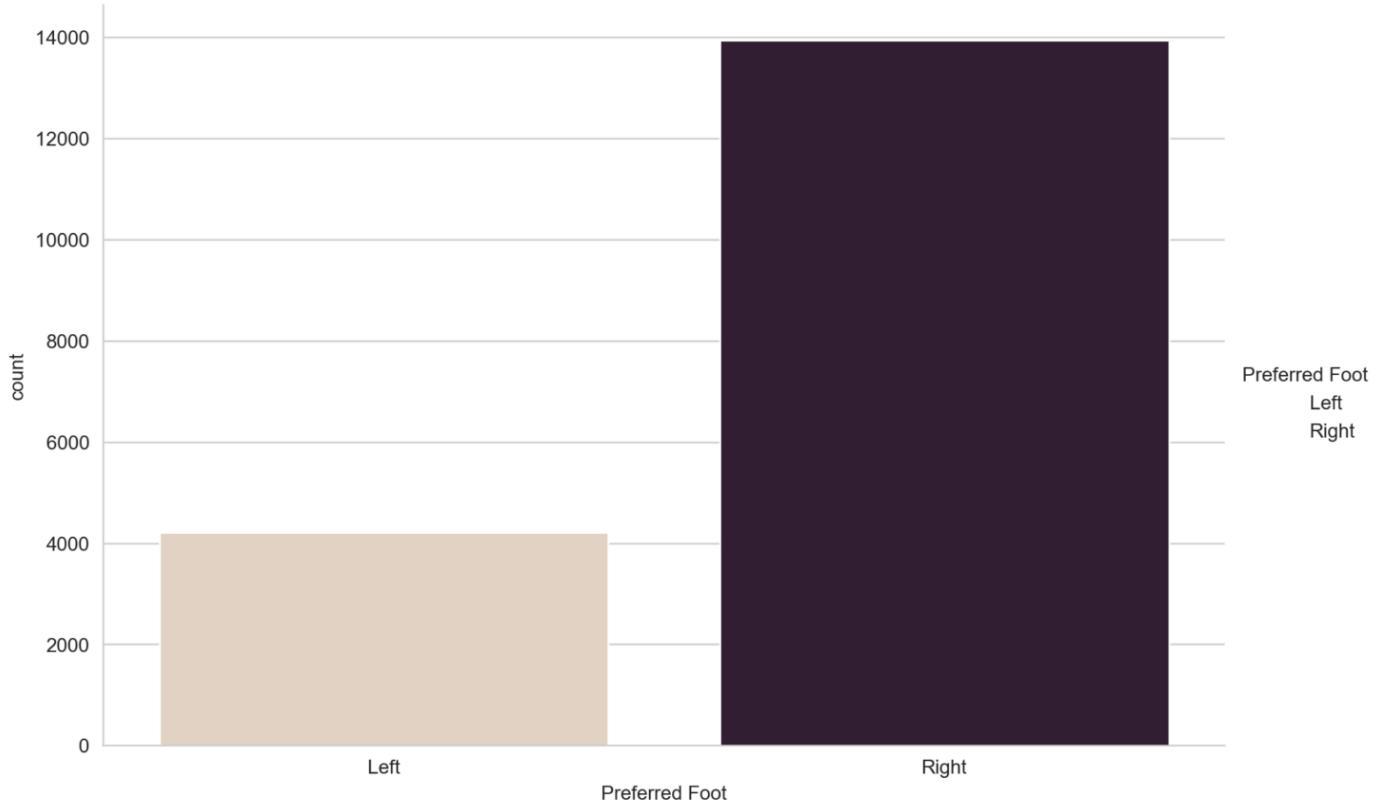


Vertical countplot:



Seaborn `catplot()`

- `catplot()` draws categorical plots onto a FacetGrid.
- Default is a scatterplot; `kind='count'` makes it a count plot.



Explore `International Reputation` variable

Unique values in `International Reputation`

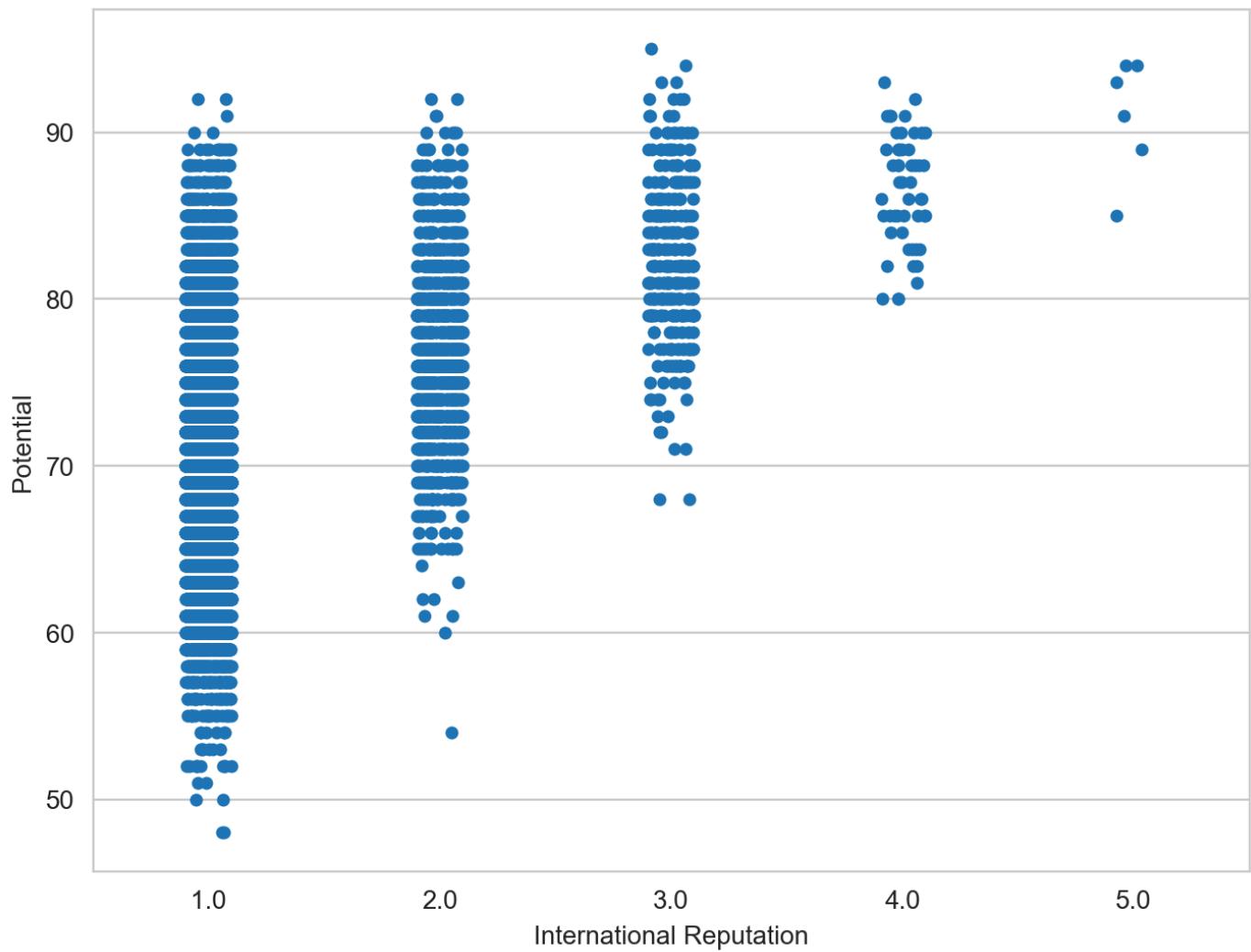
5

Distribution of `International Reputation`

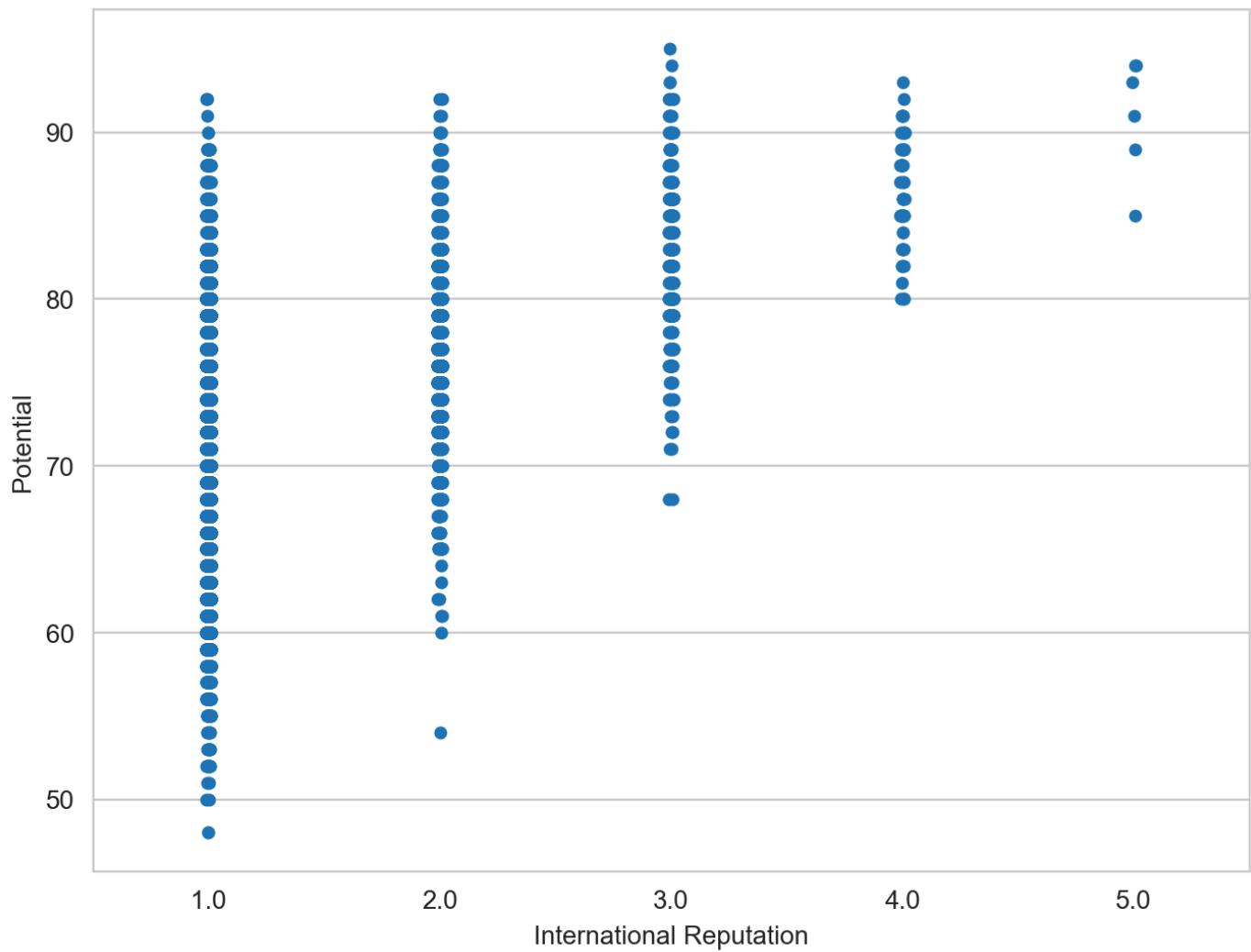
International Reputation	count
1	16532
2	1261
3	309
4	51
5	6

Seaborn `stripplot()`

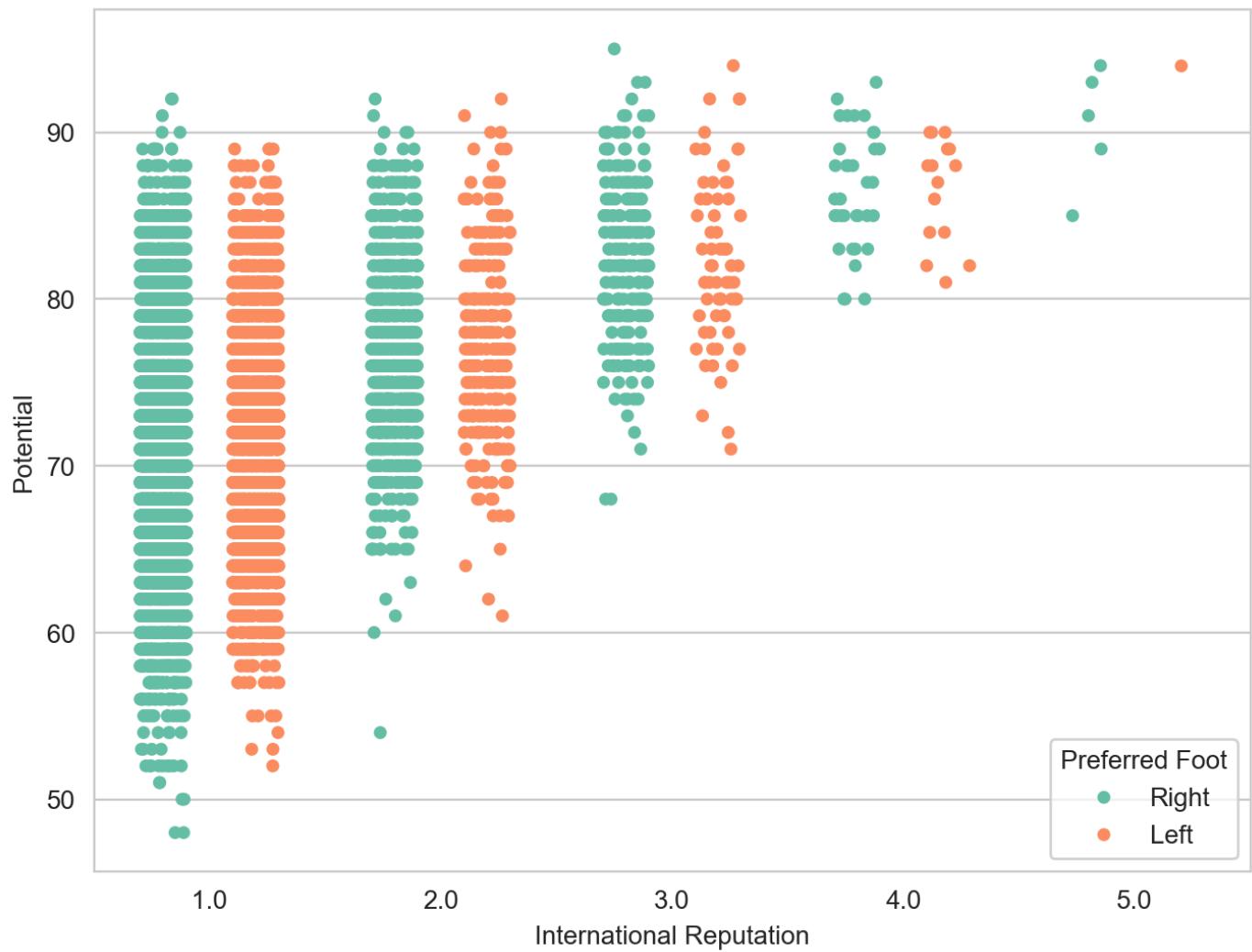
- `stripplot` draws a scatterplot with a categorical variable.
- Useful with box or violin plots to show all observations.



With jitter:

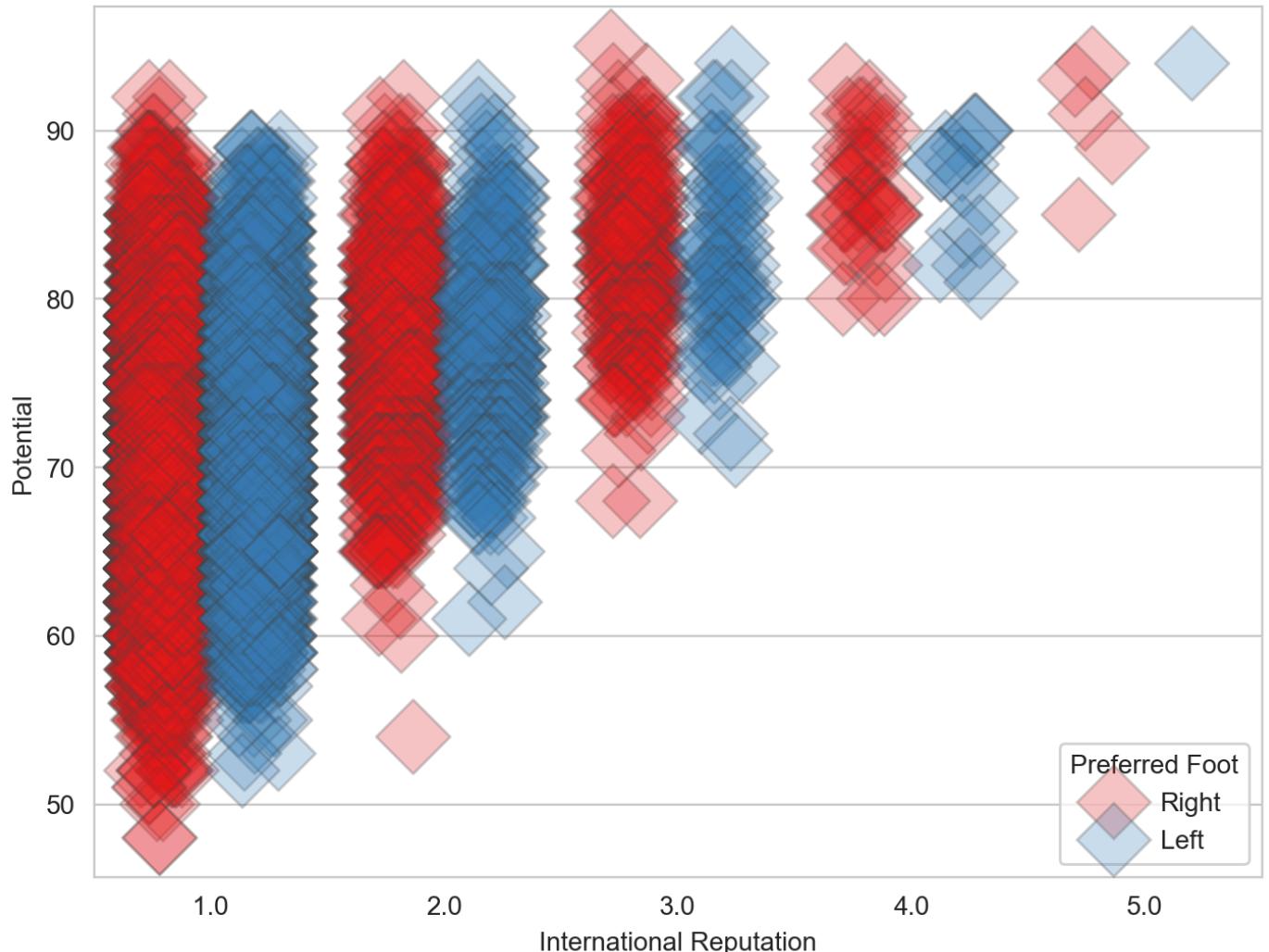


Nested with `Preferred_Foot` :



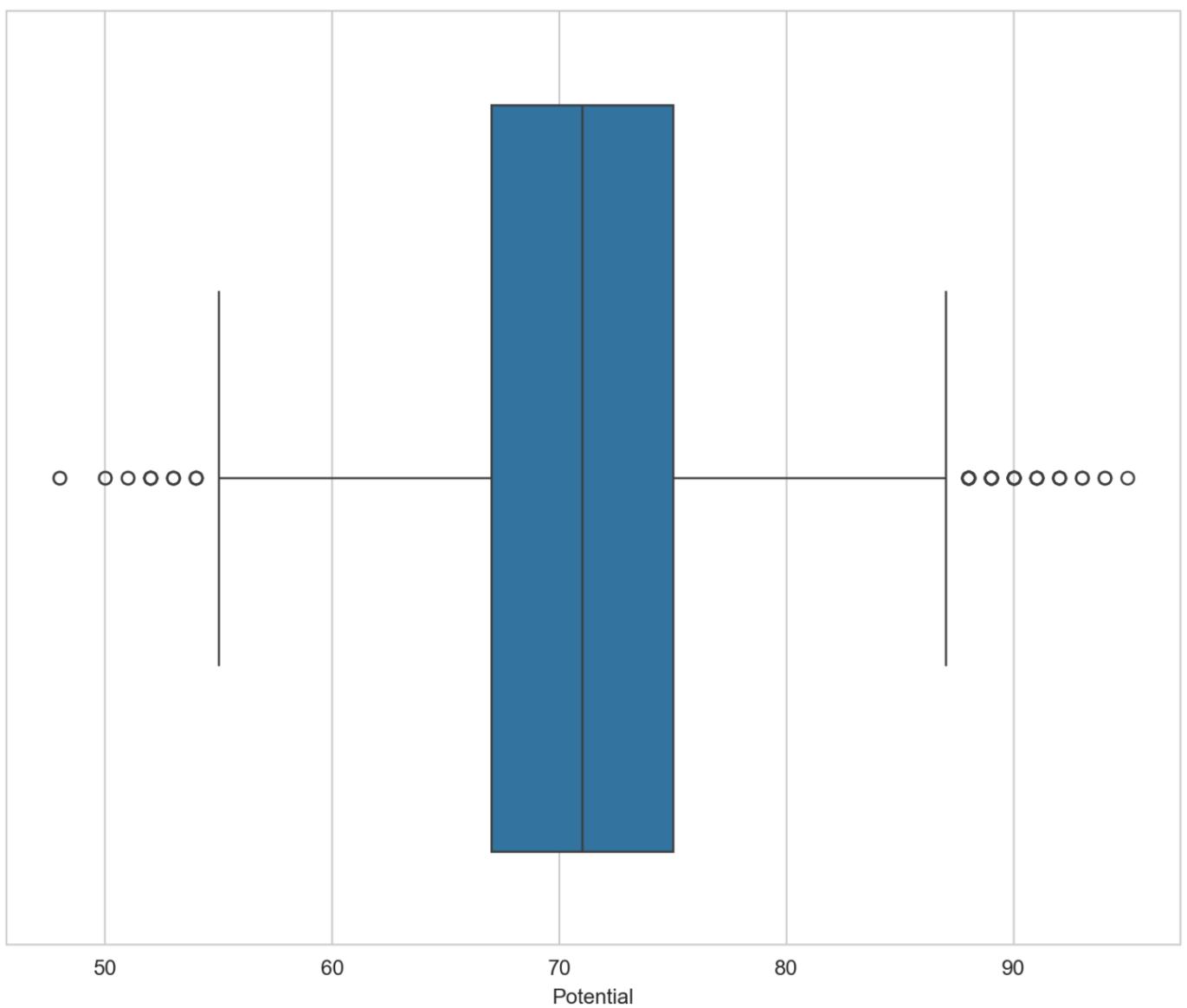
Large points with different aesthetics:

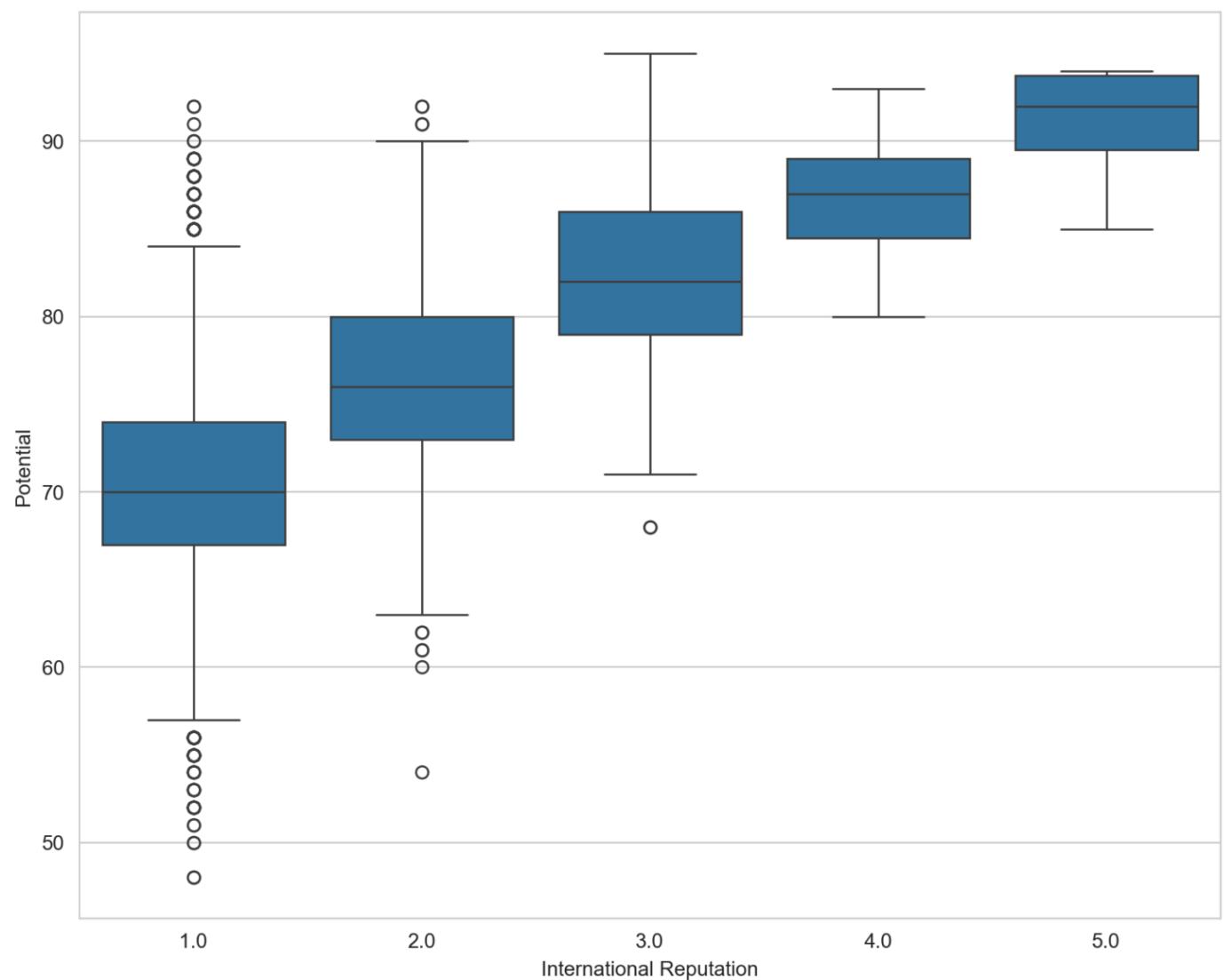
International Reputation vs Potential



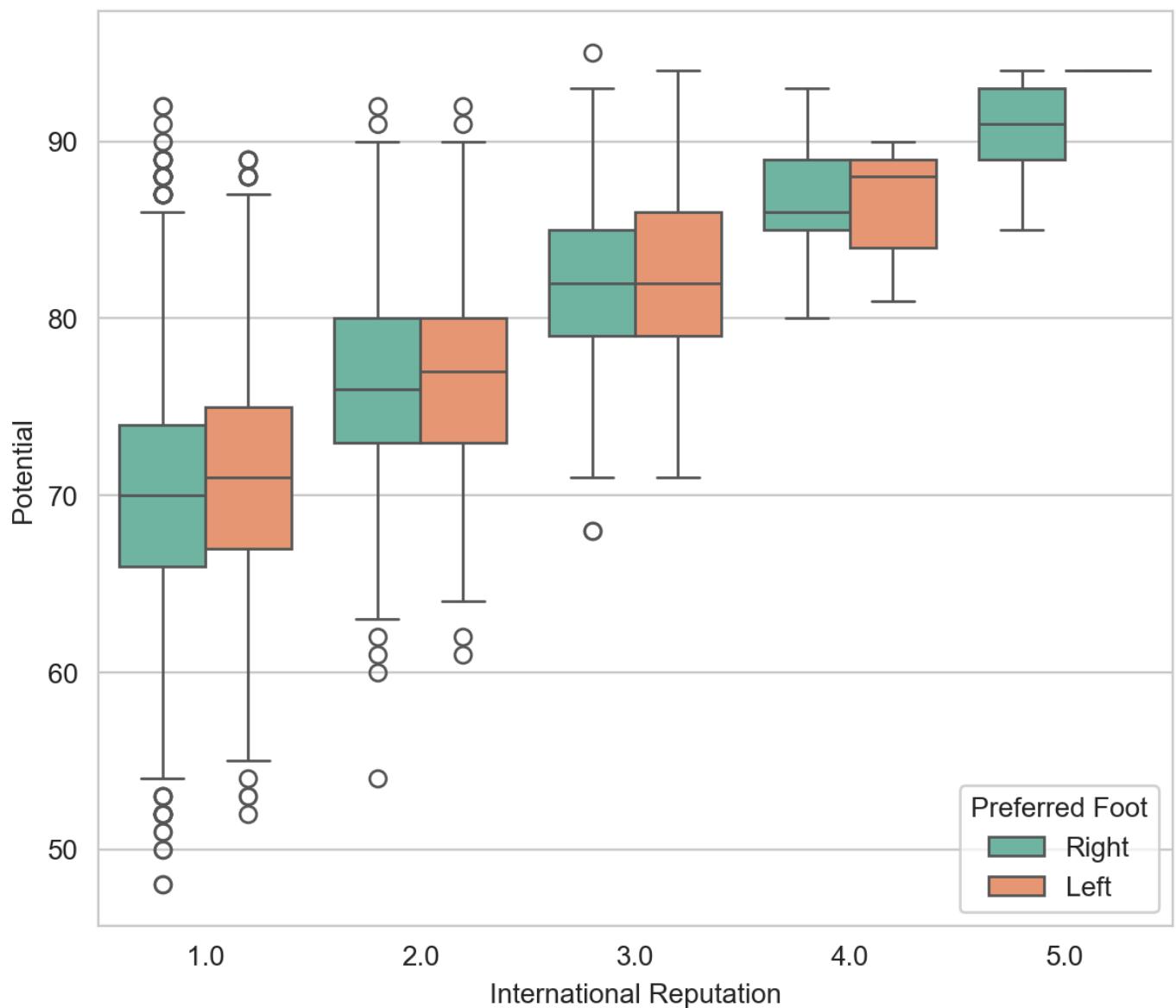
Seaborn `boxplot()`

- `boxplot` shows distributions with respect to categories using quartiles and whiskers.



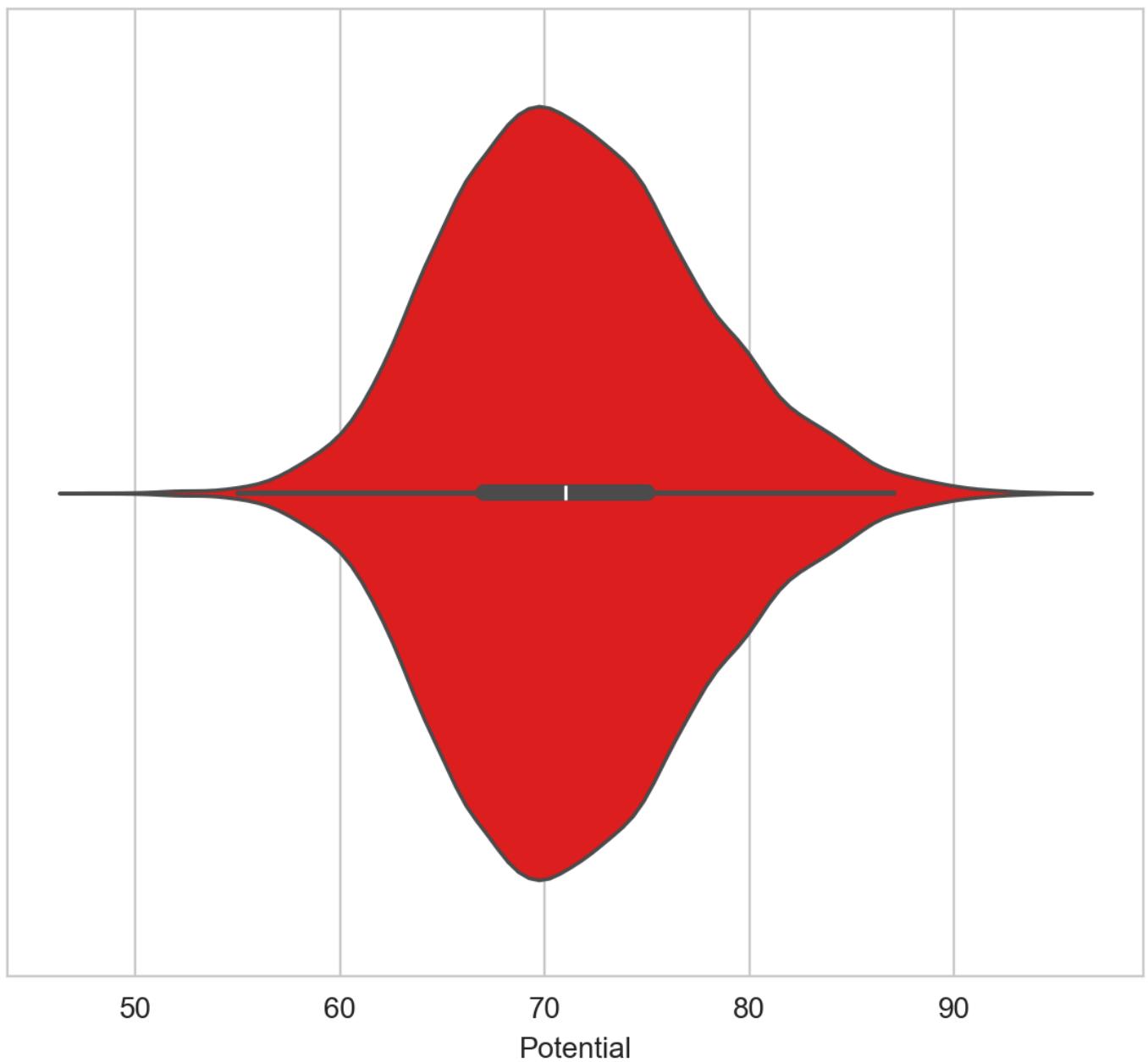


Nested grouping by two variables:

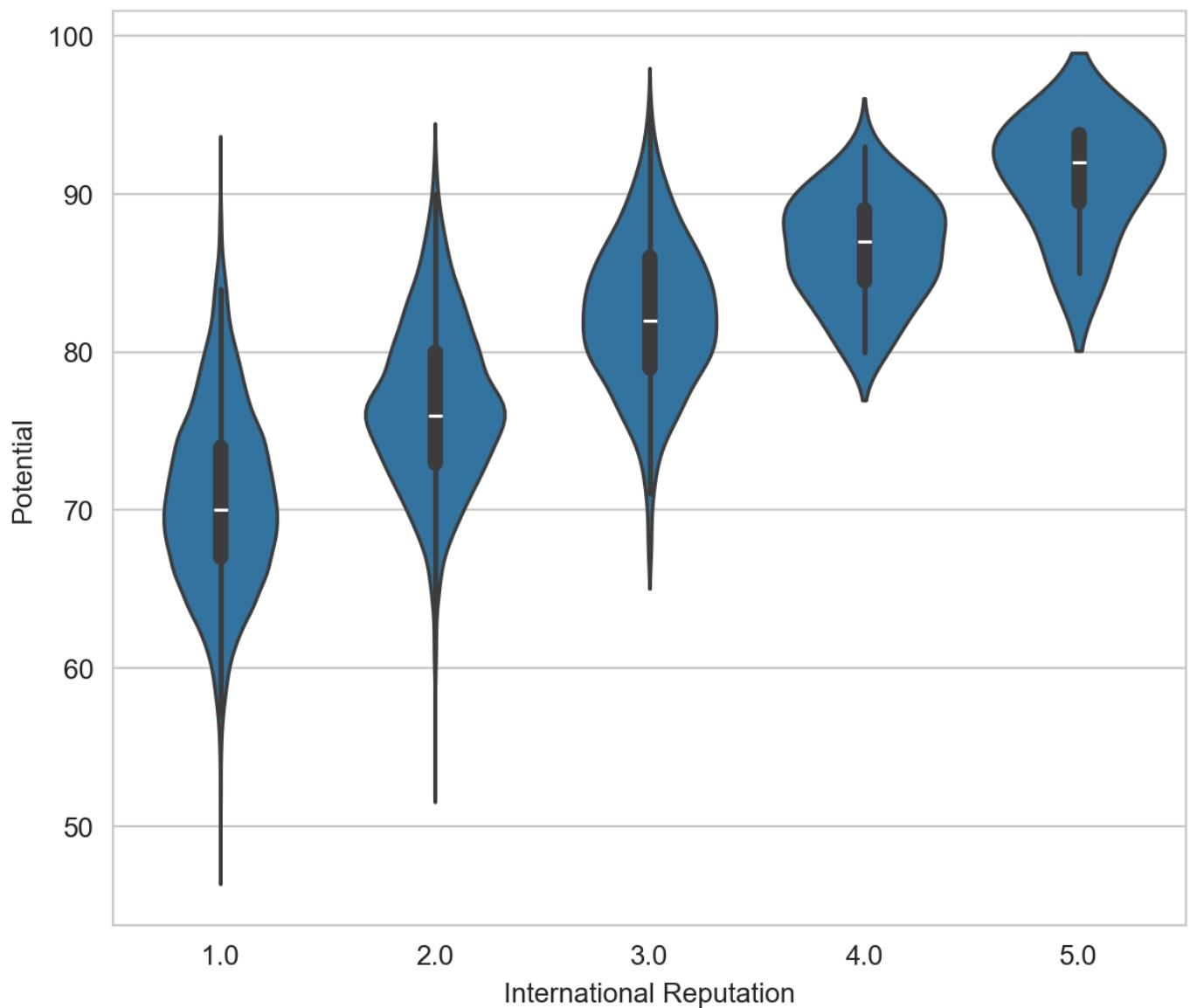


Seaborn `violinplot()`

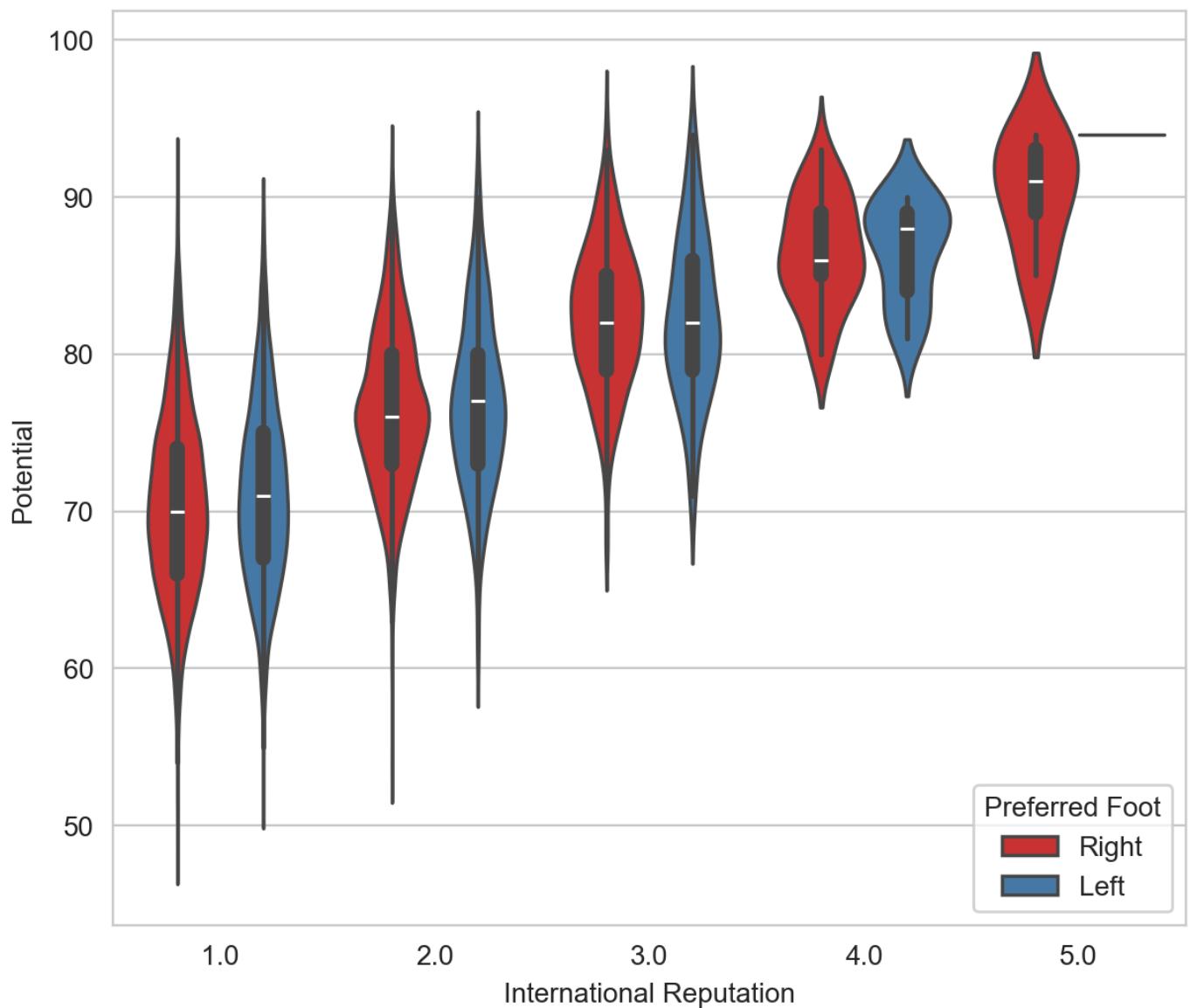
- `violinplot` combines boxplot and KDE to show distribution across categorical variables.



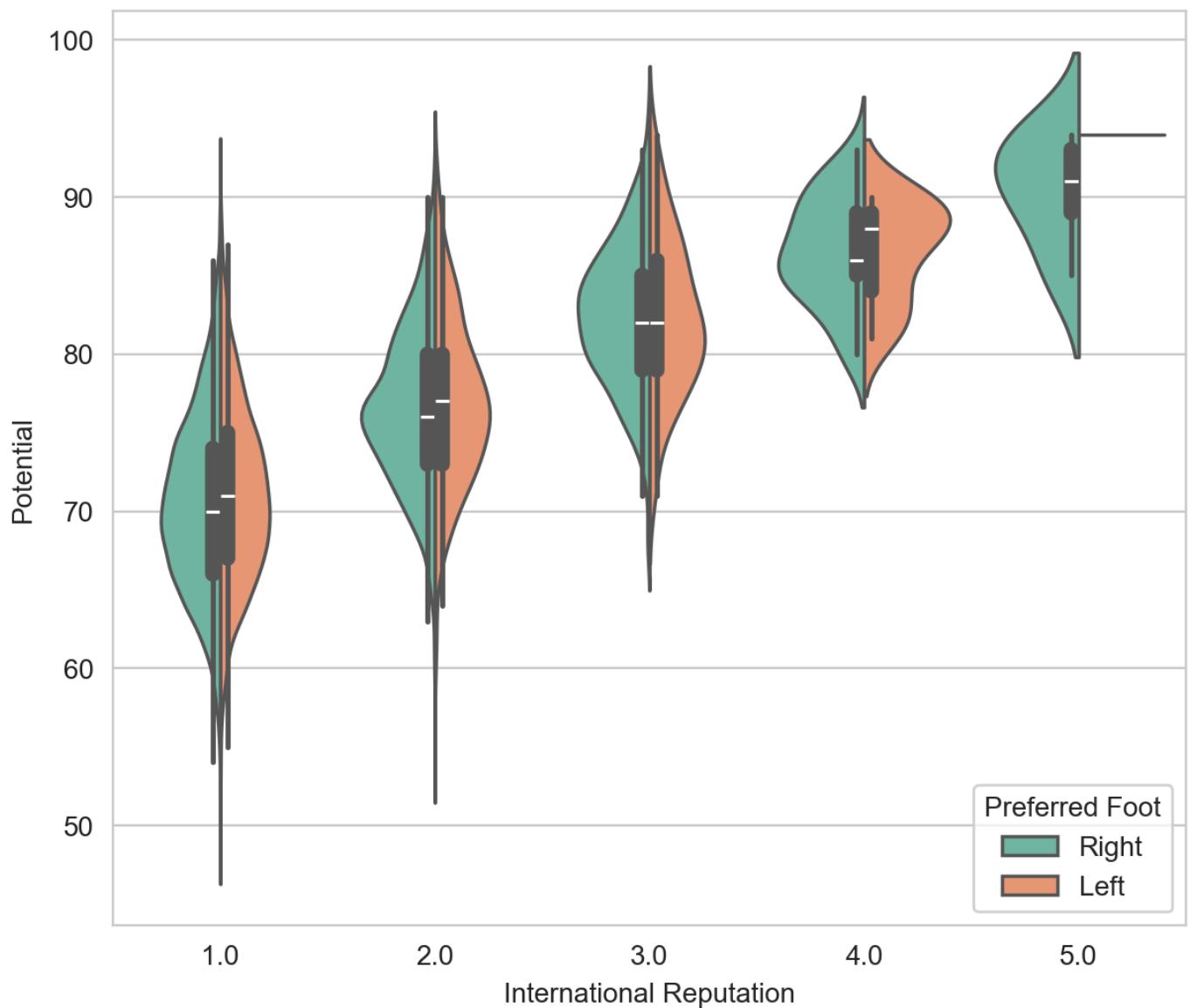
Grouped by `International Reputation`:



Nested grouping by two variables:

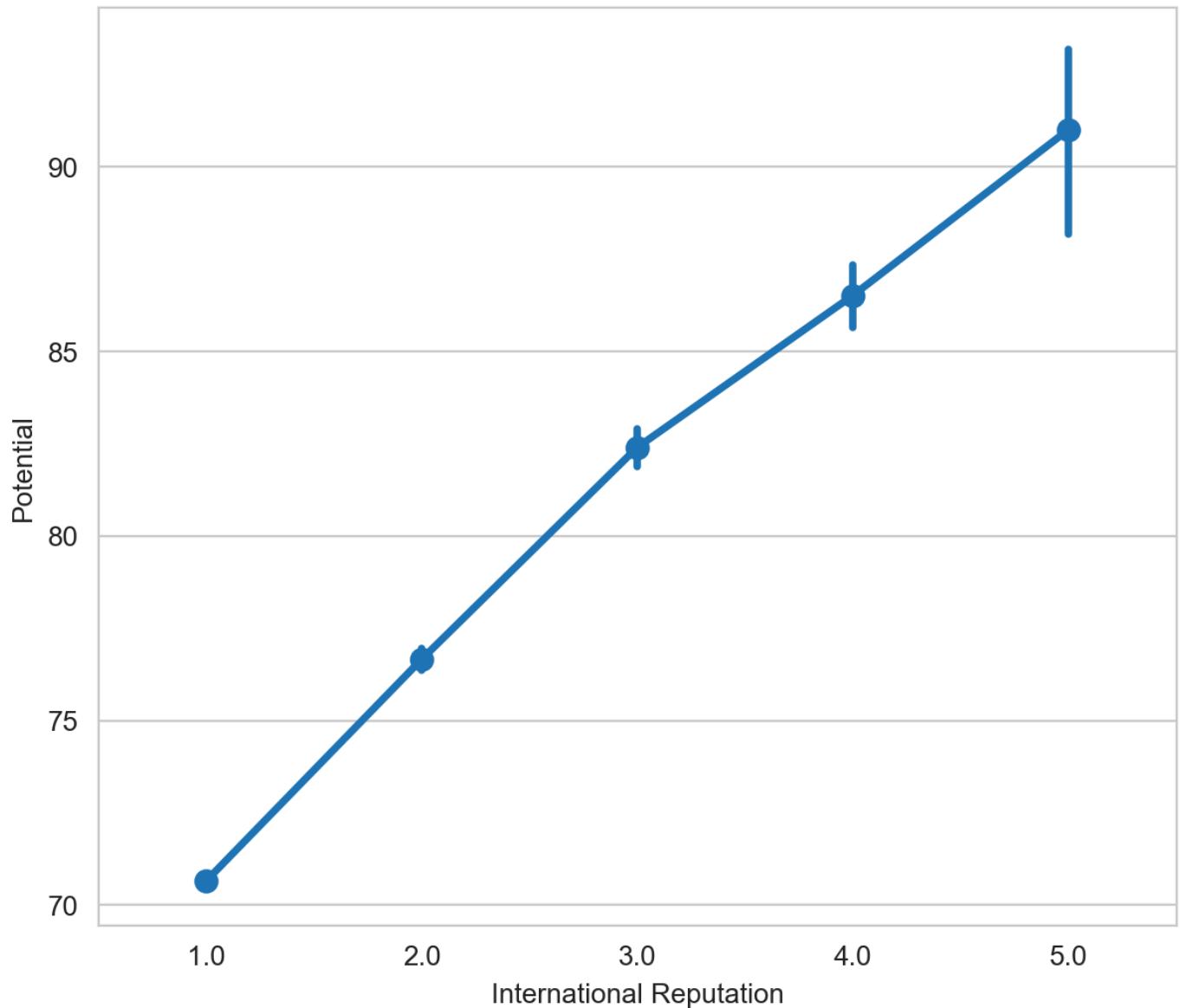


Split violins:

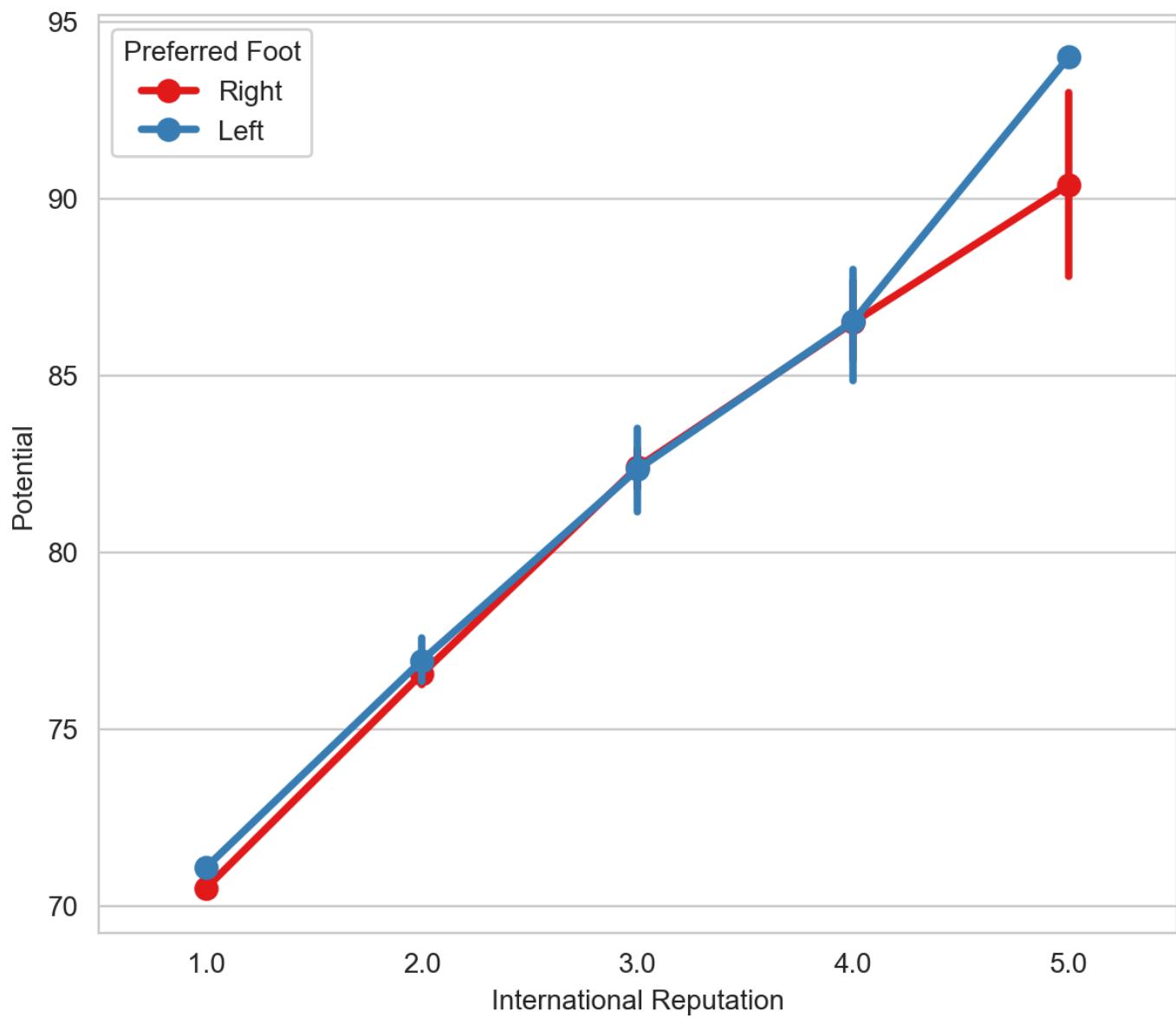


Seaborn `pointplot()`

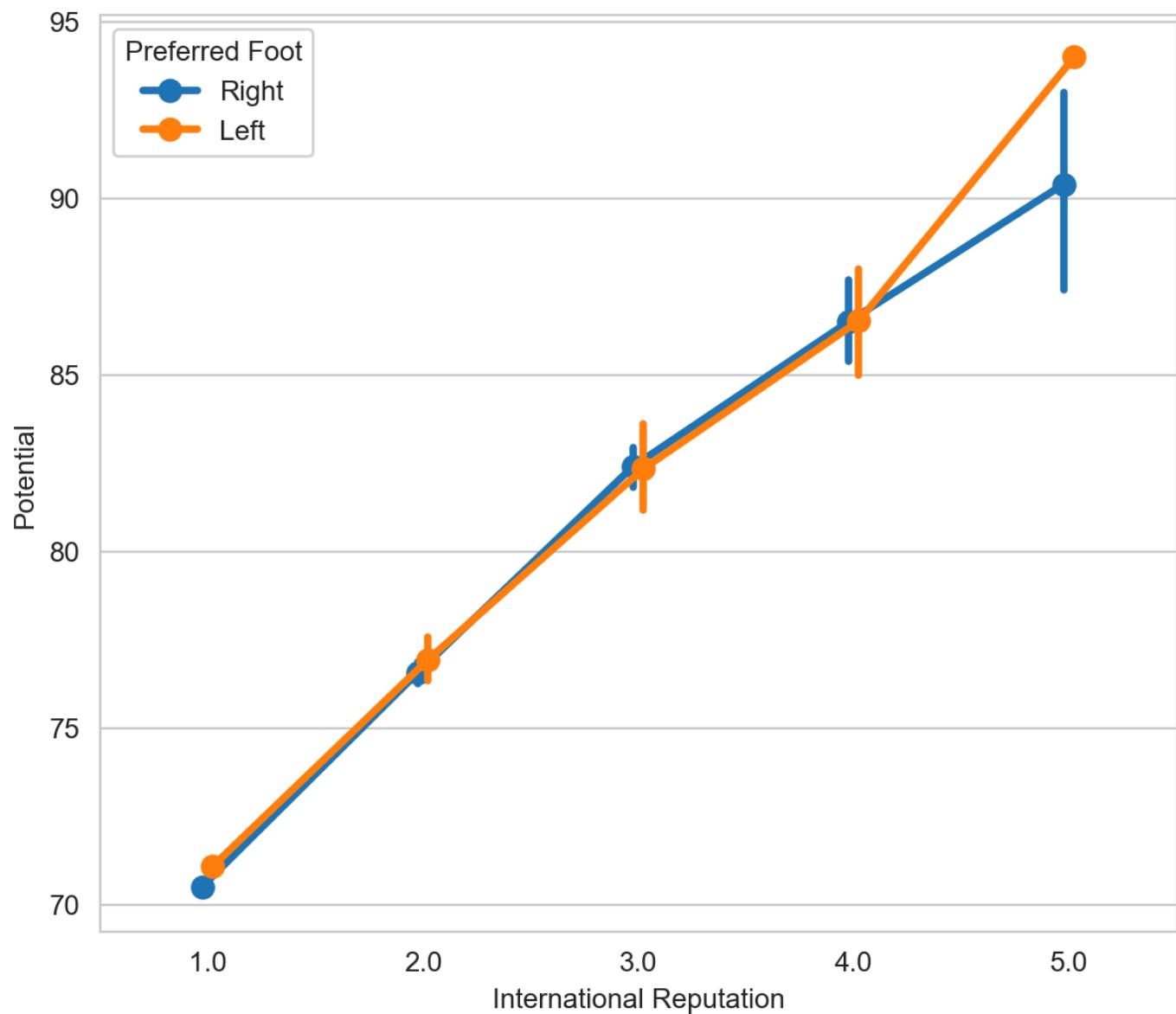
- `pointplot` shows point estimates and confidence intervals with scatter glyphs.



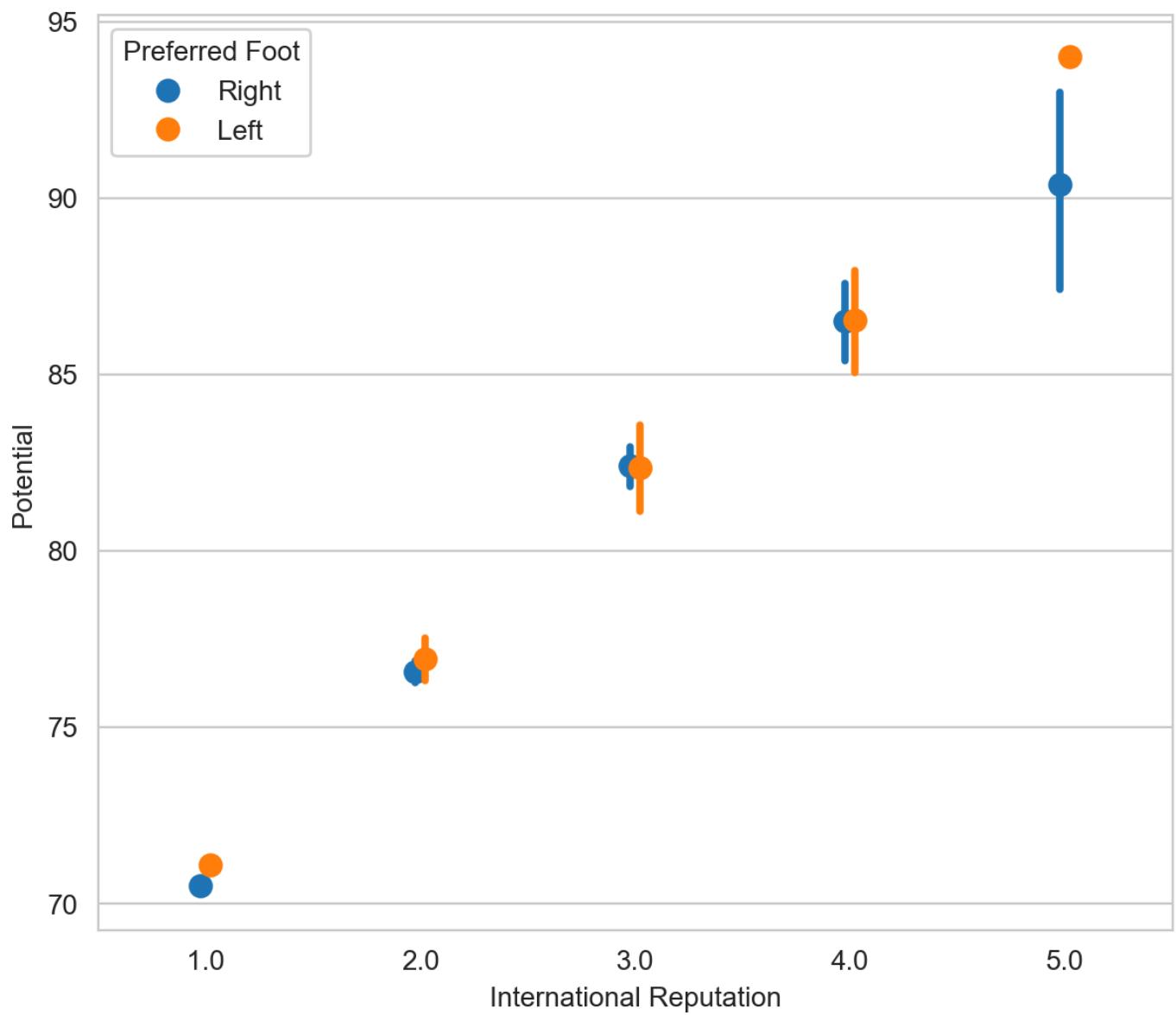
Nested grouping by two variables:



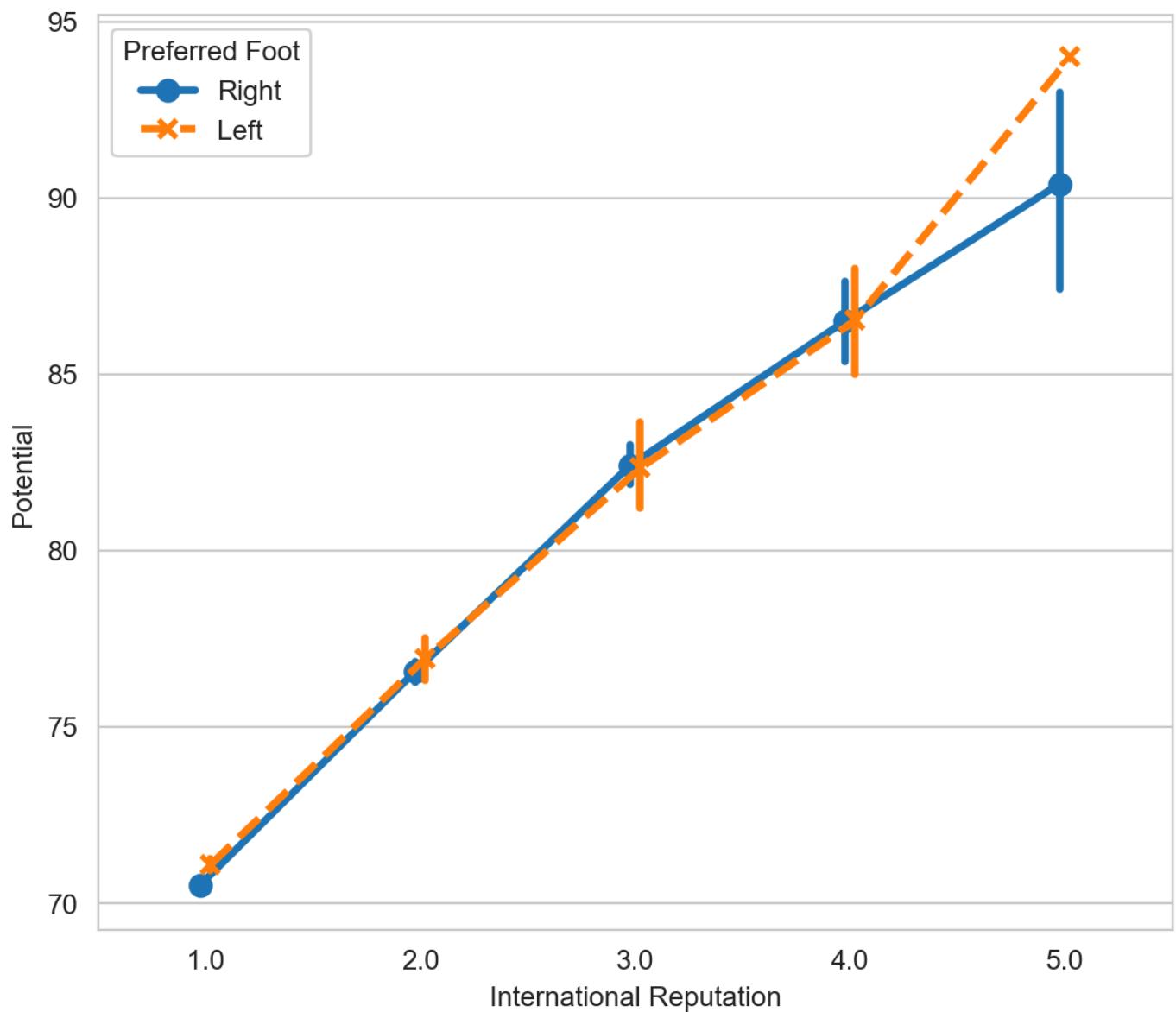
Separated points:



No linking:



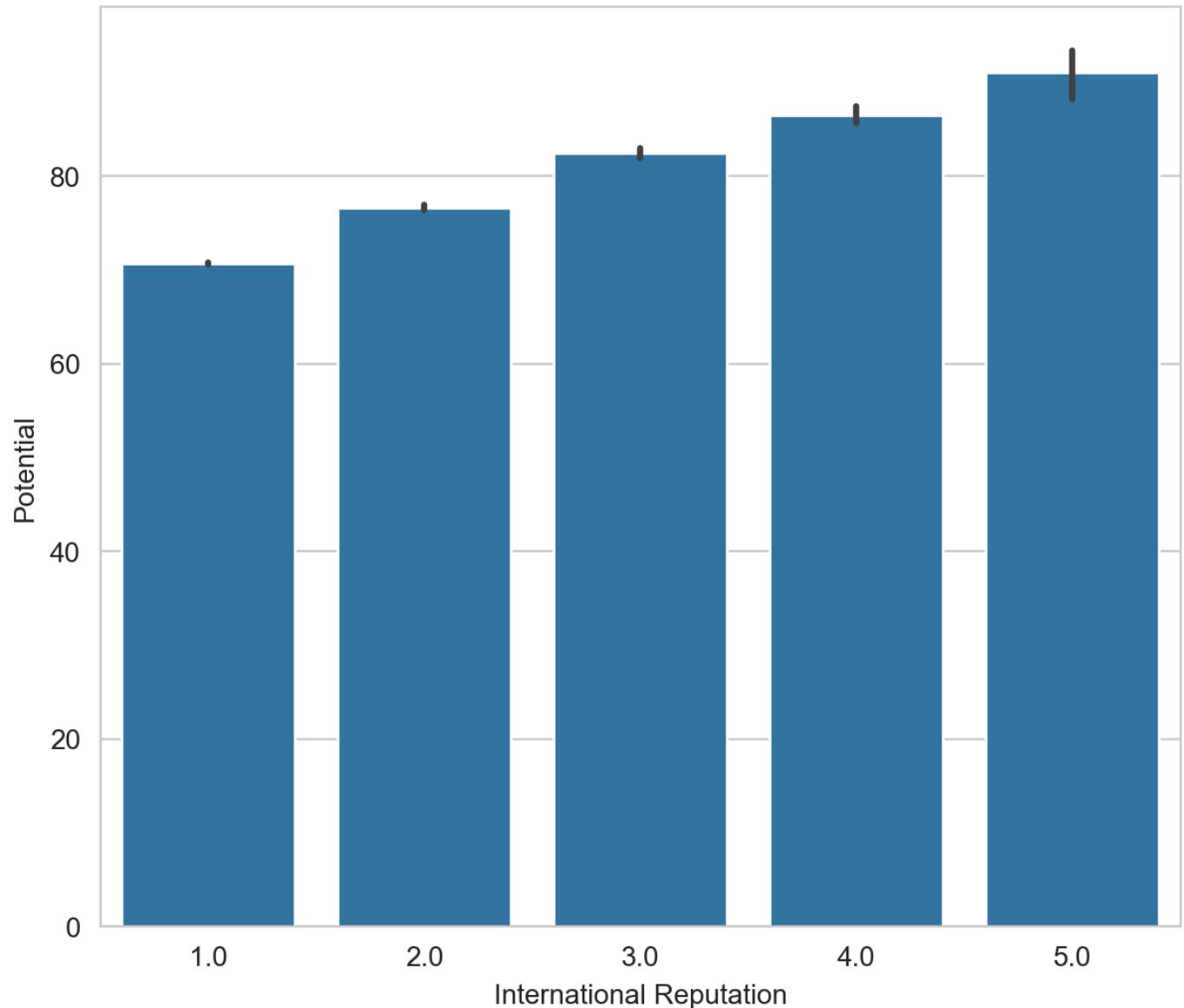
Different markers and line styles:



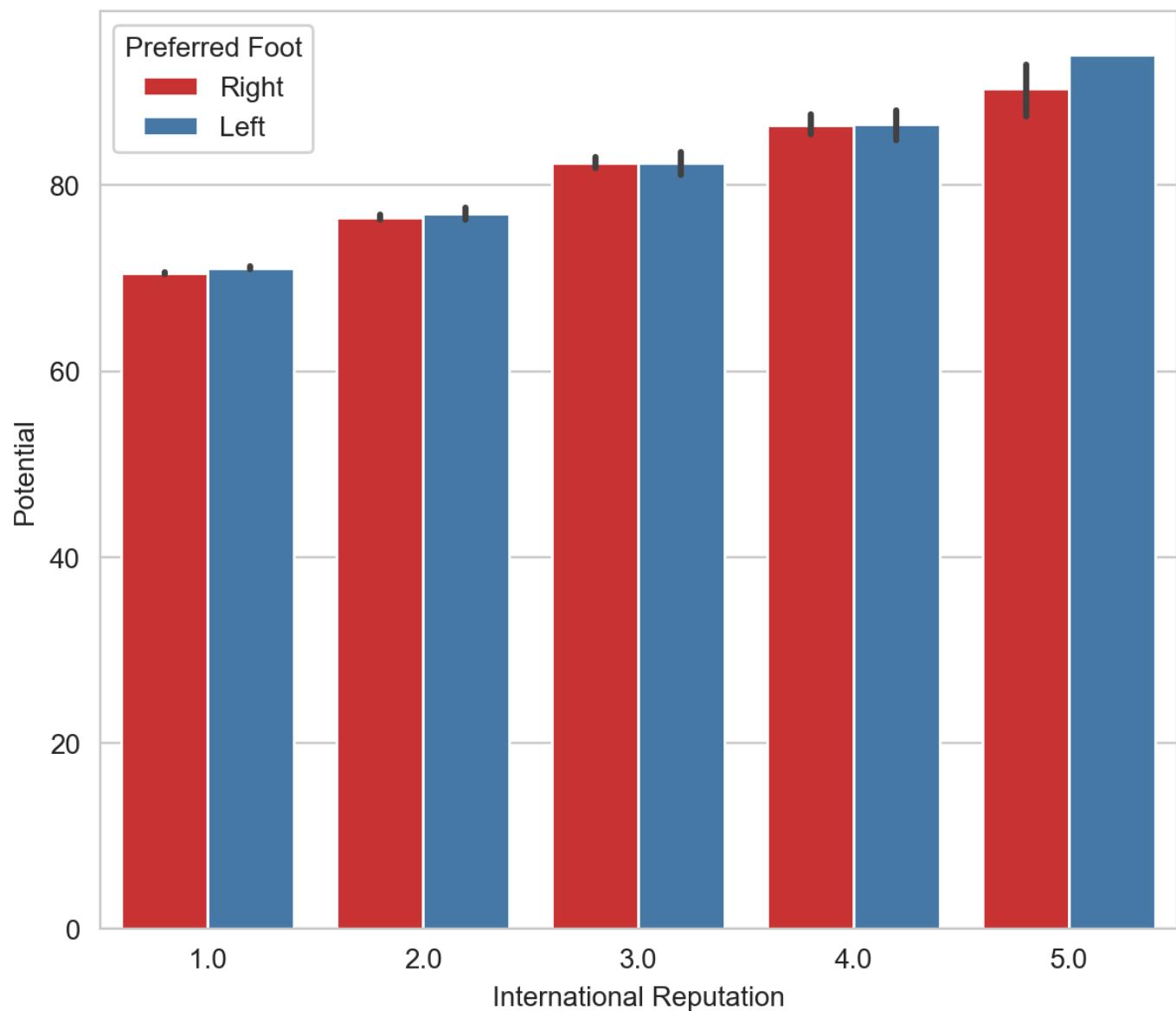
Seaborn `barplot()`

- `barplot` shows point estimates and confidence intervals as bars.

International Reputation vs Potential

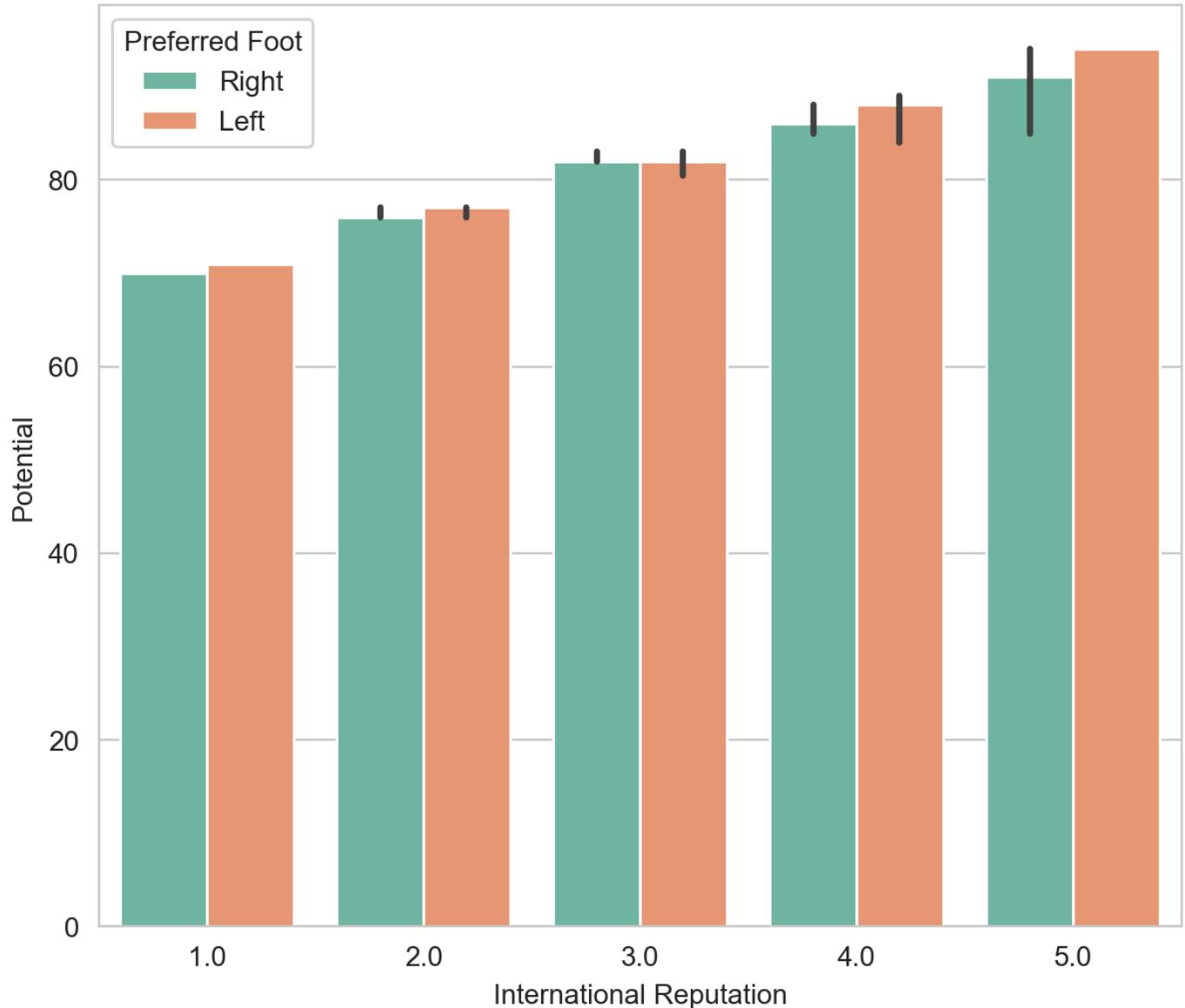


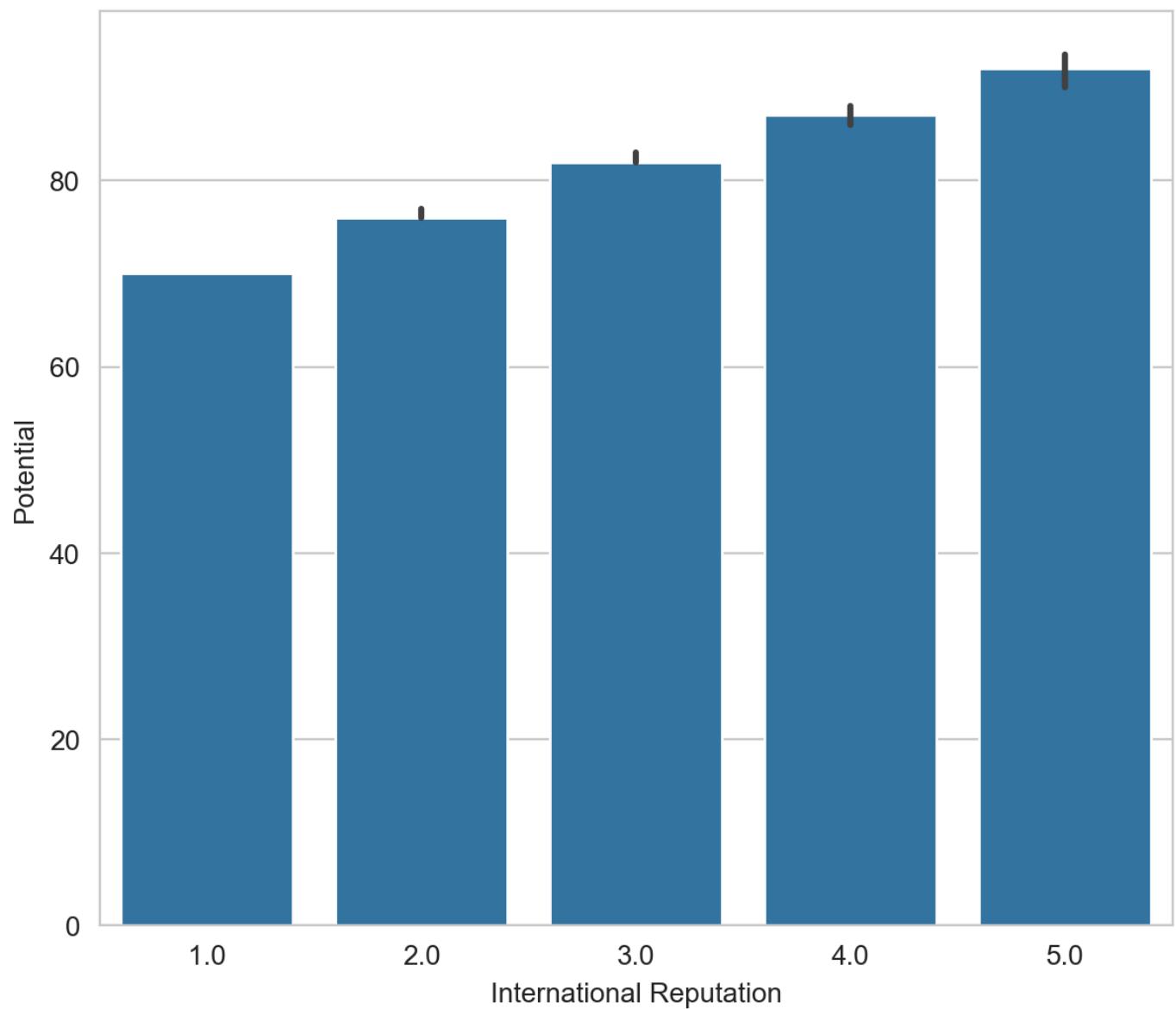
Nested grouping by two variables:



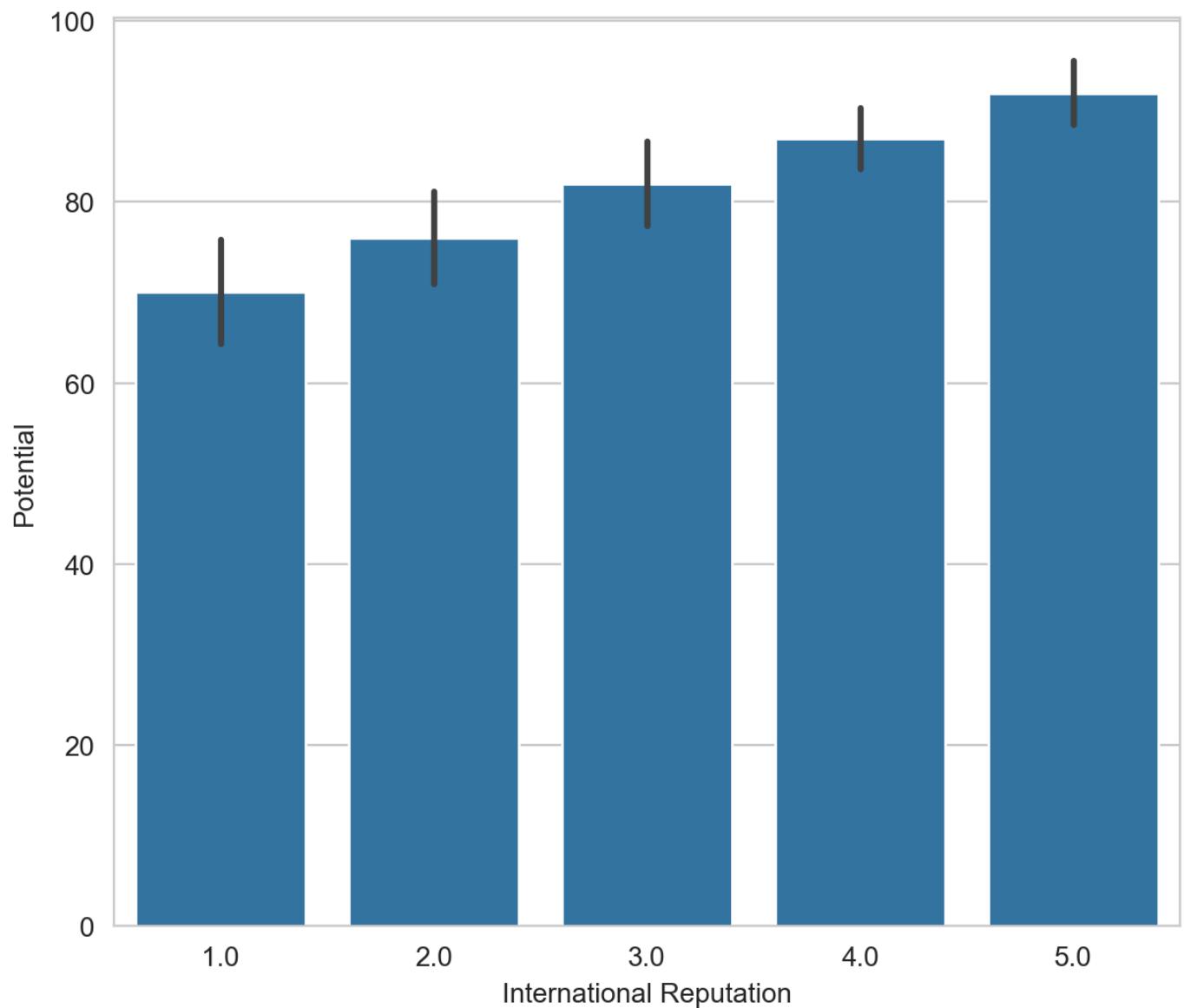
Using median:

International Reputation vs Potential

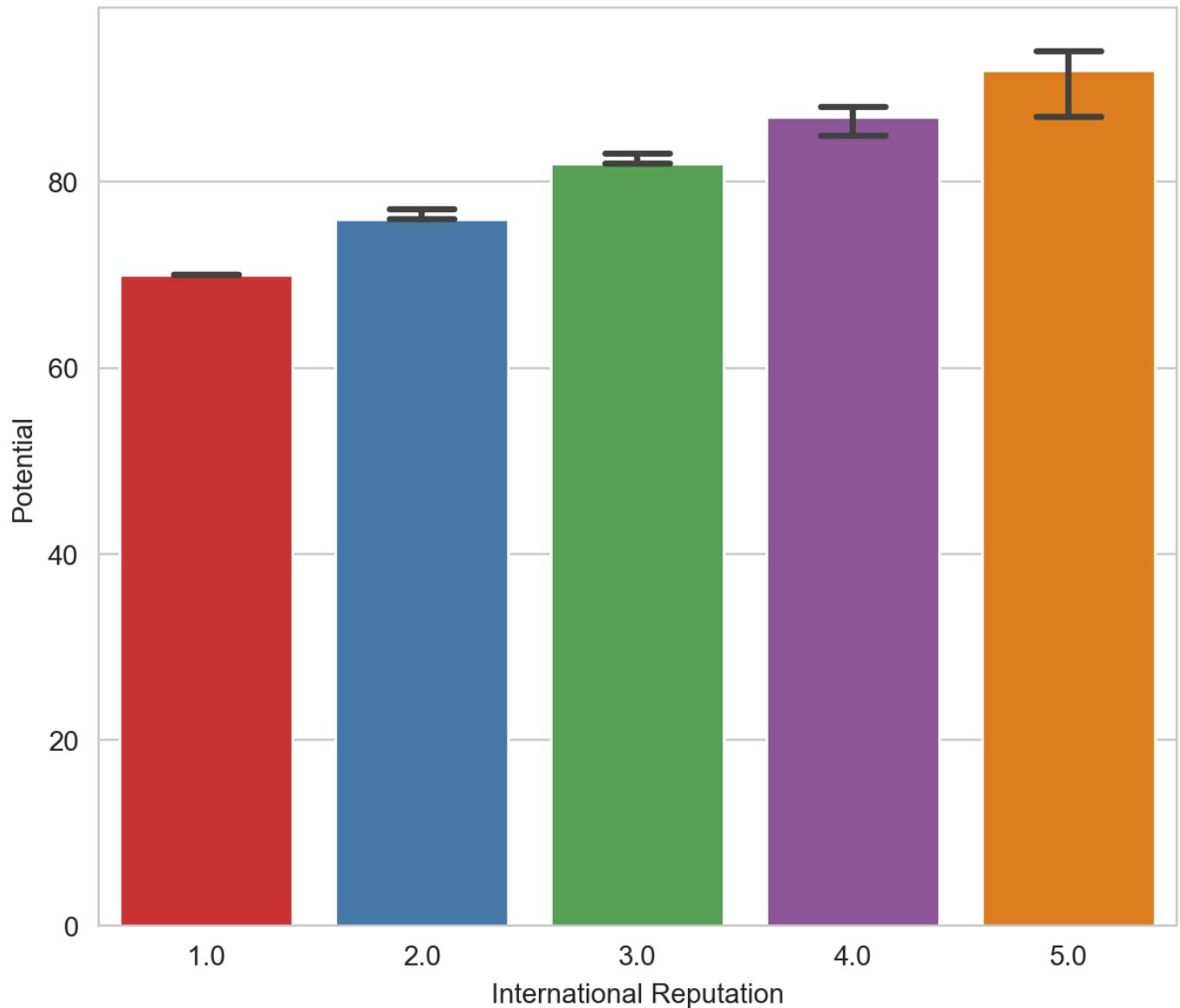




Standard deviation instead of CI:



With caps on error bars:

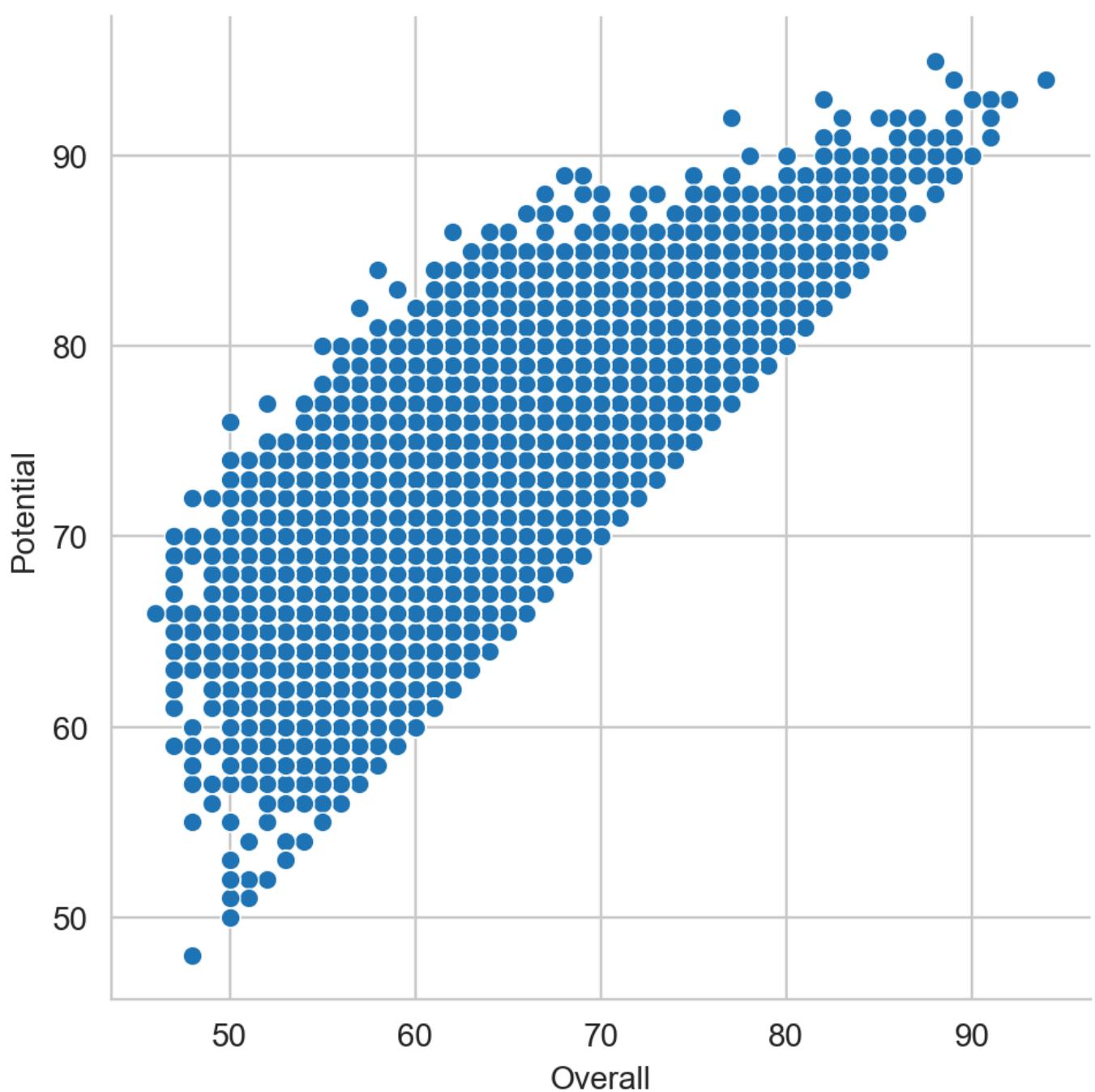


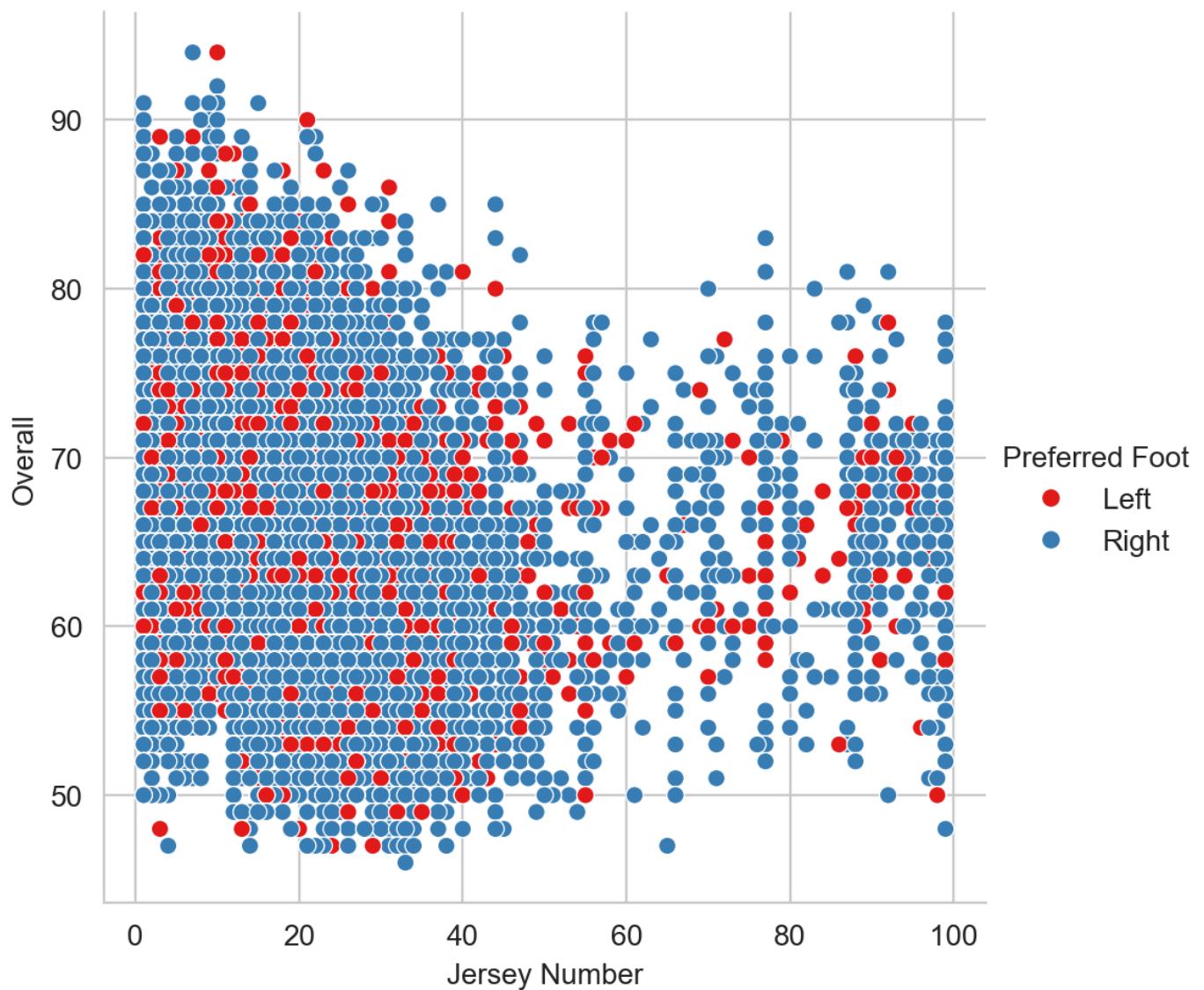
Visualizing statistical relationship with Seaborn

`relplot()`

Seaborn `relplot()`

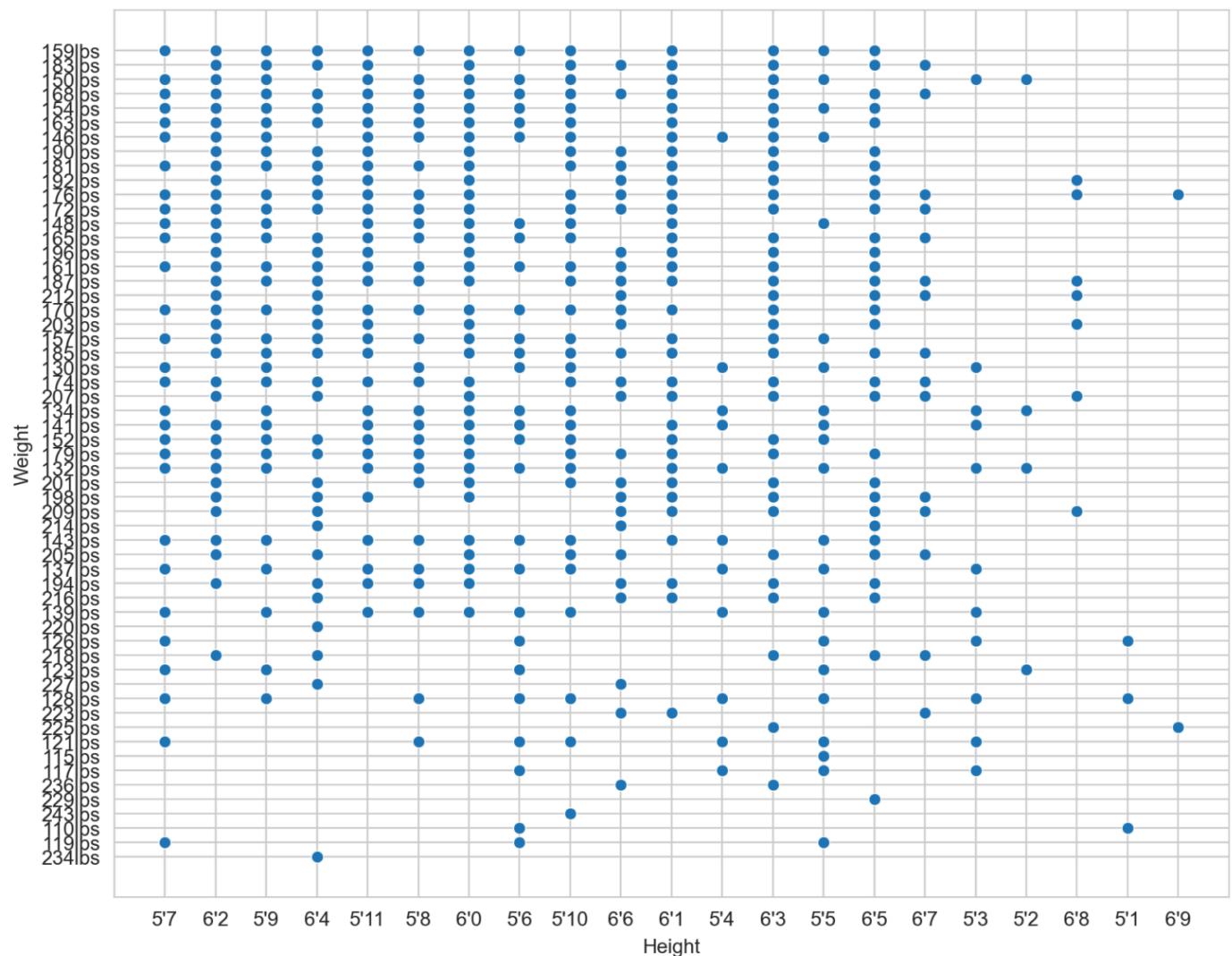
- `relplot()` draws relational plots onto a FacetGrid with `kind='scatter'` or `kind='line'`.





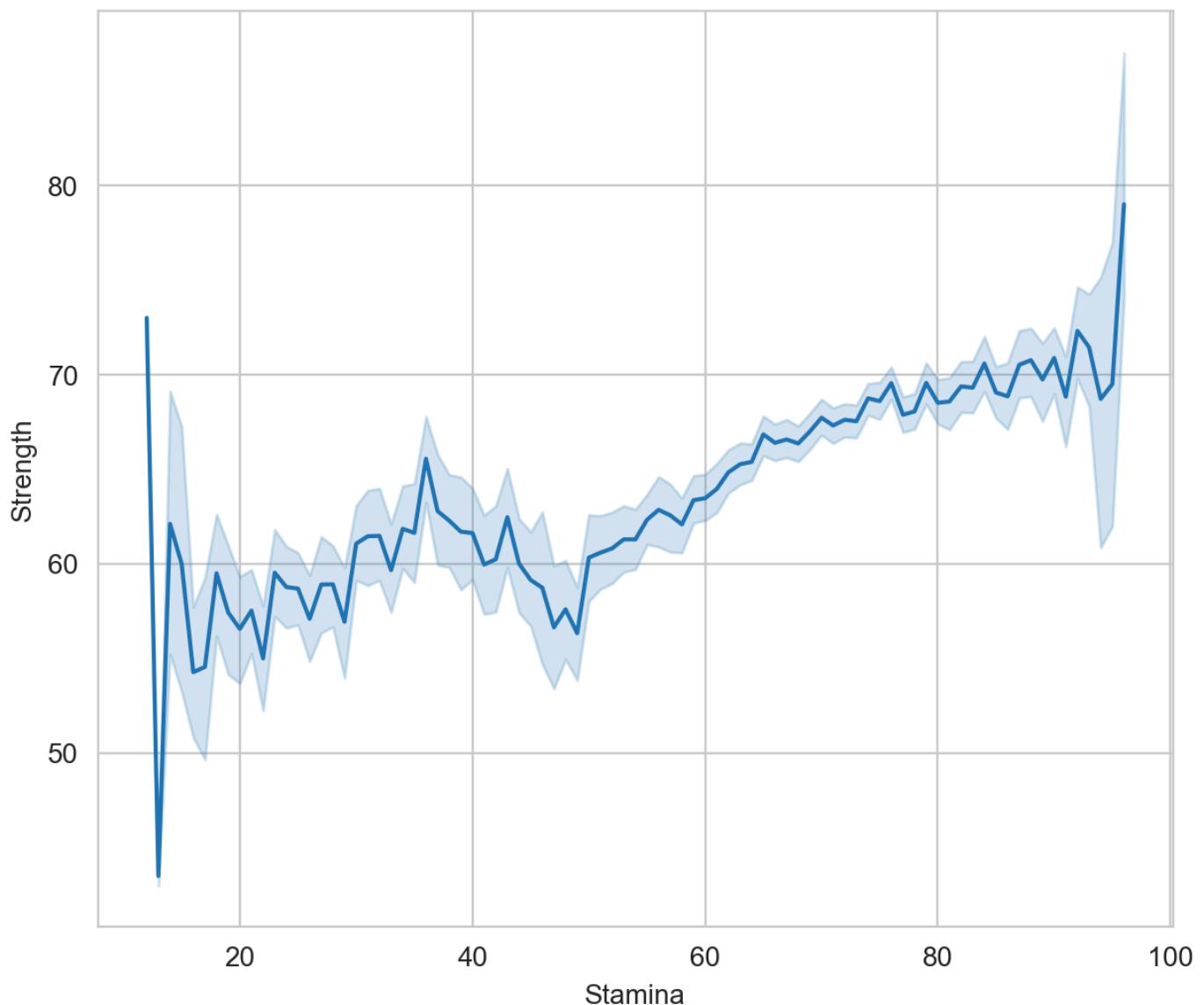
Seaborn `scatterplot()`

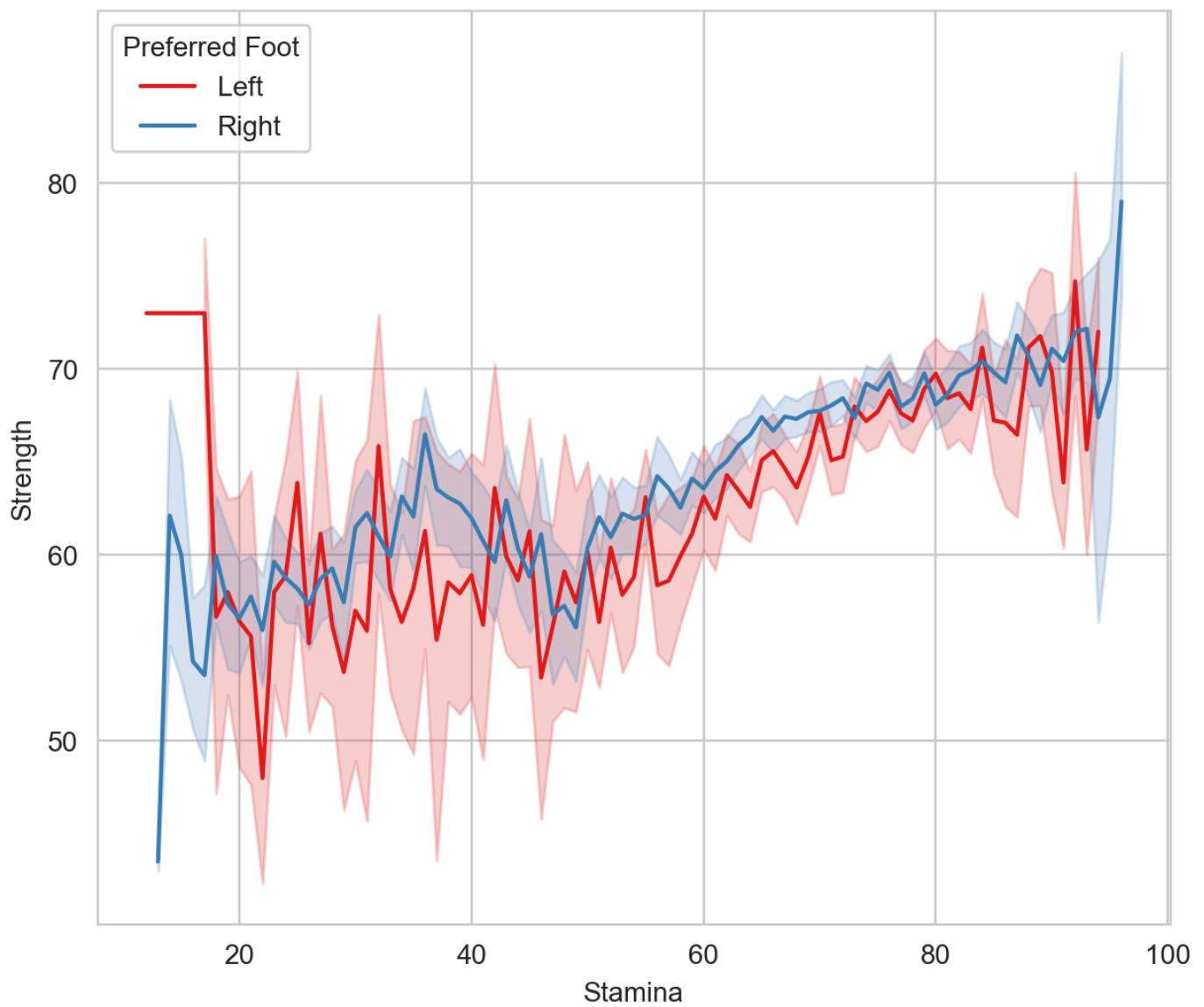
- `scatterplot` draws a scatter plot with semantic groupings via `hue`, `size`, and `style`.



Seaborn `lineplot()`

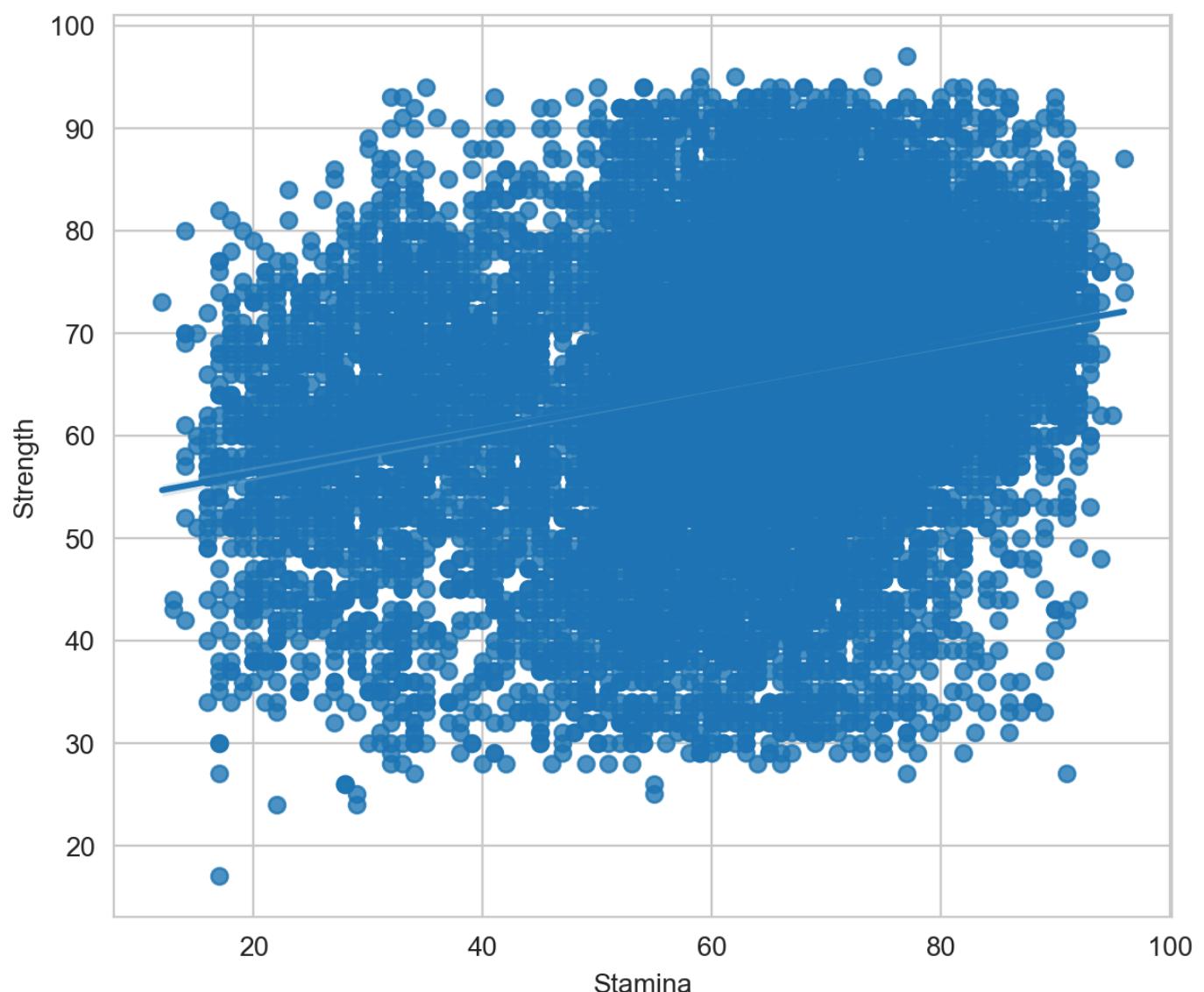
- `lineplot` draws a line plot with semantic groupings.

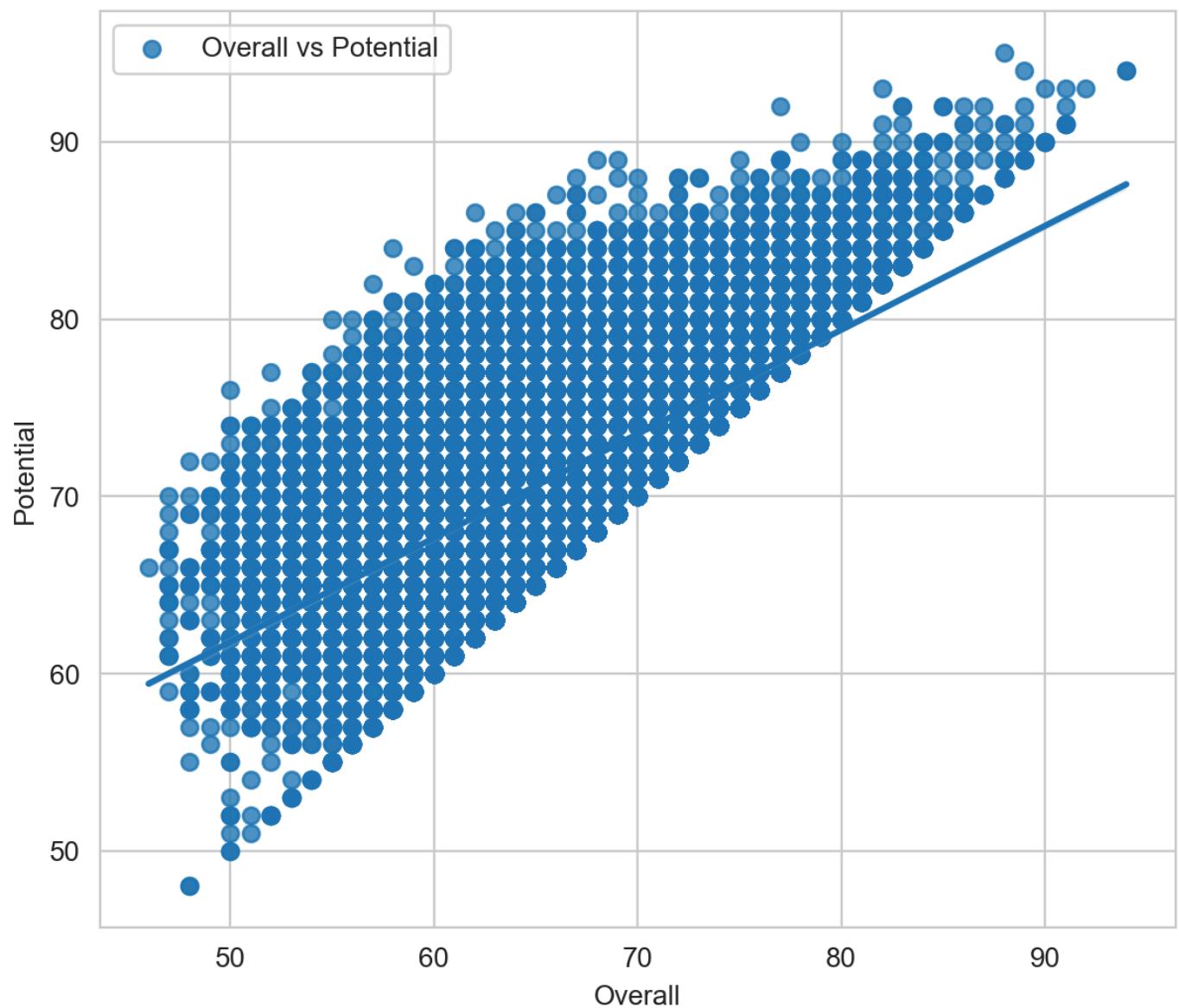


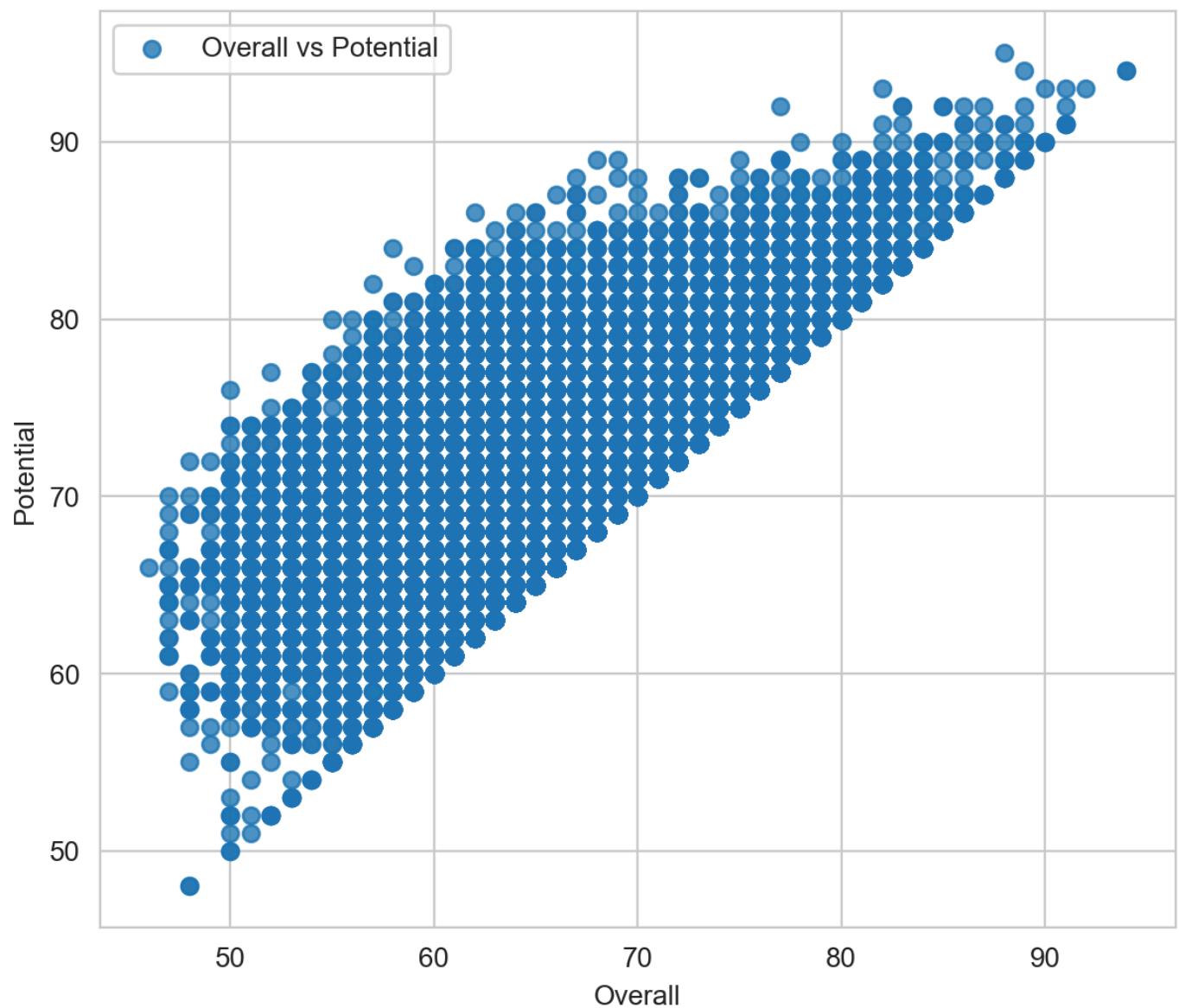


Visualize linear relationship with Seaborn `regplot()`

- `regplot()` plots data and a linear regression model fit.

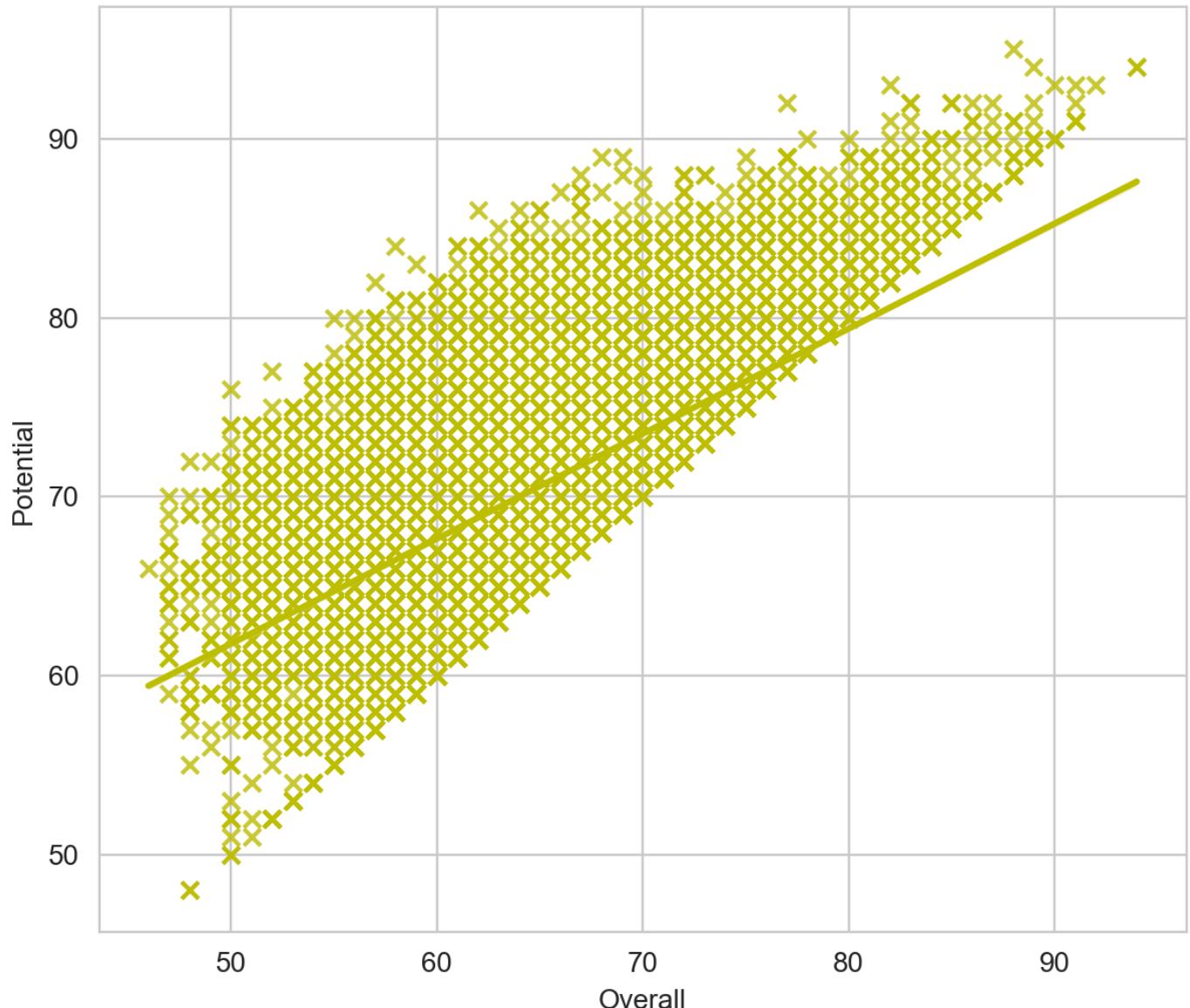






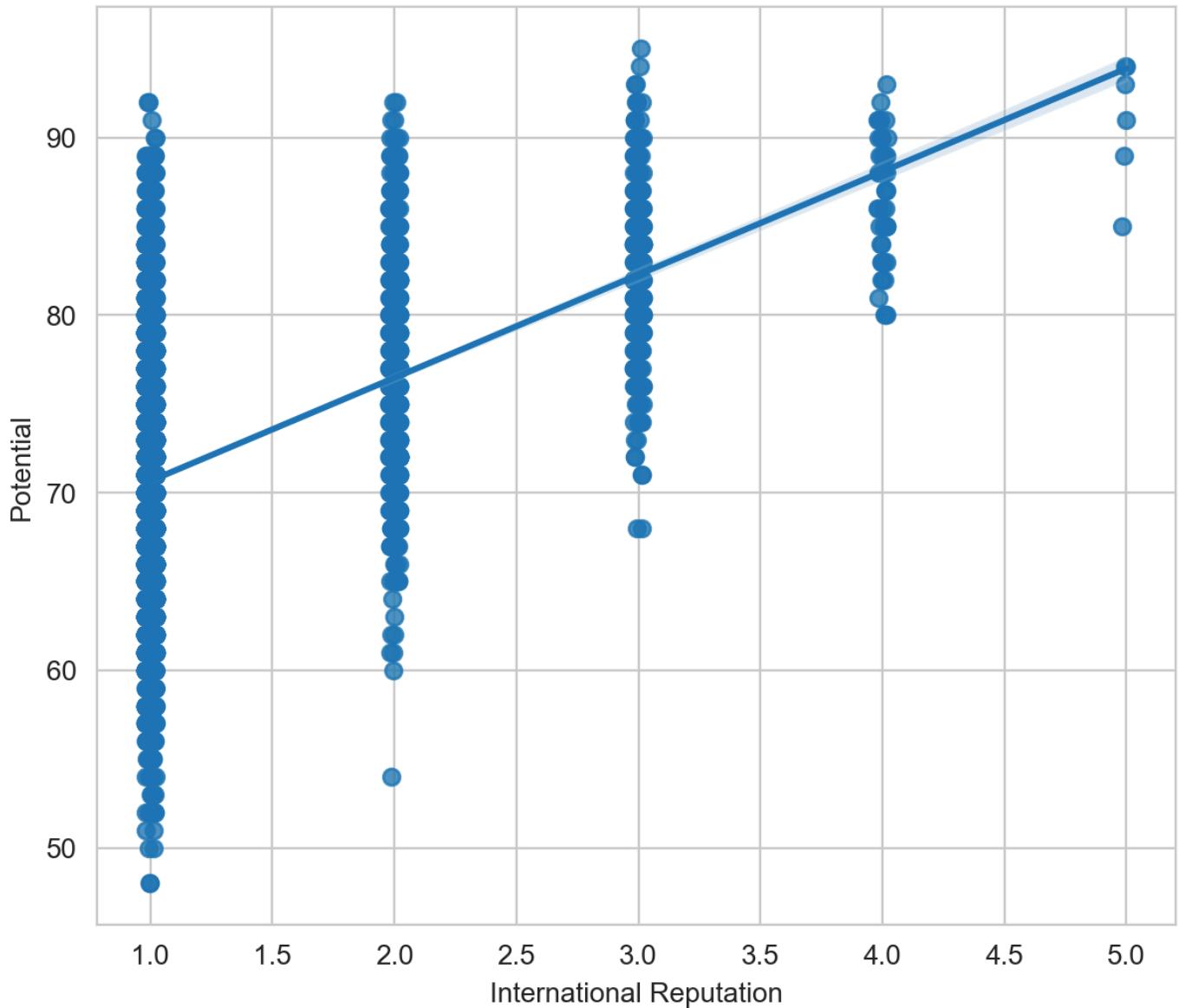
Different color and marker:

Overall vs Potential



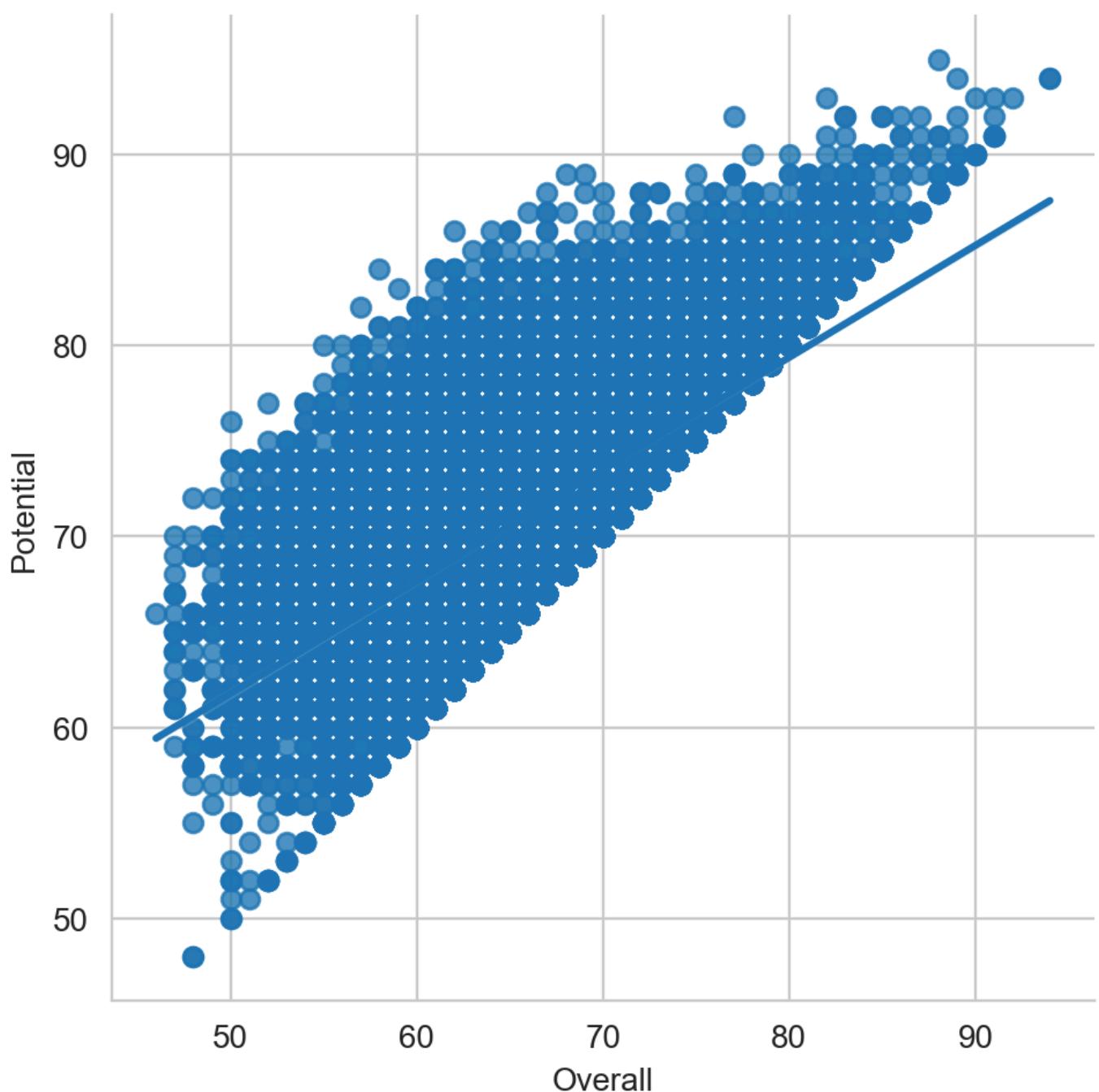
Discrete variable with jitter:

International Reputation vs Potential

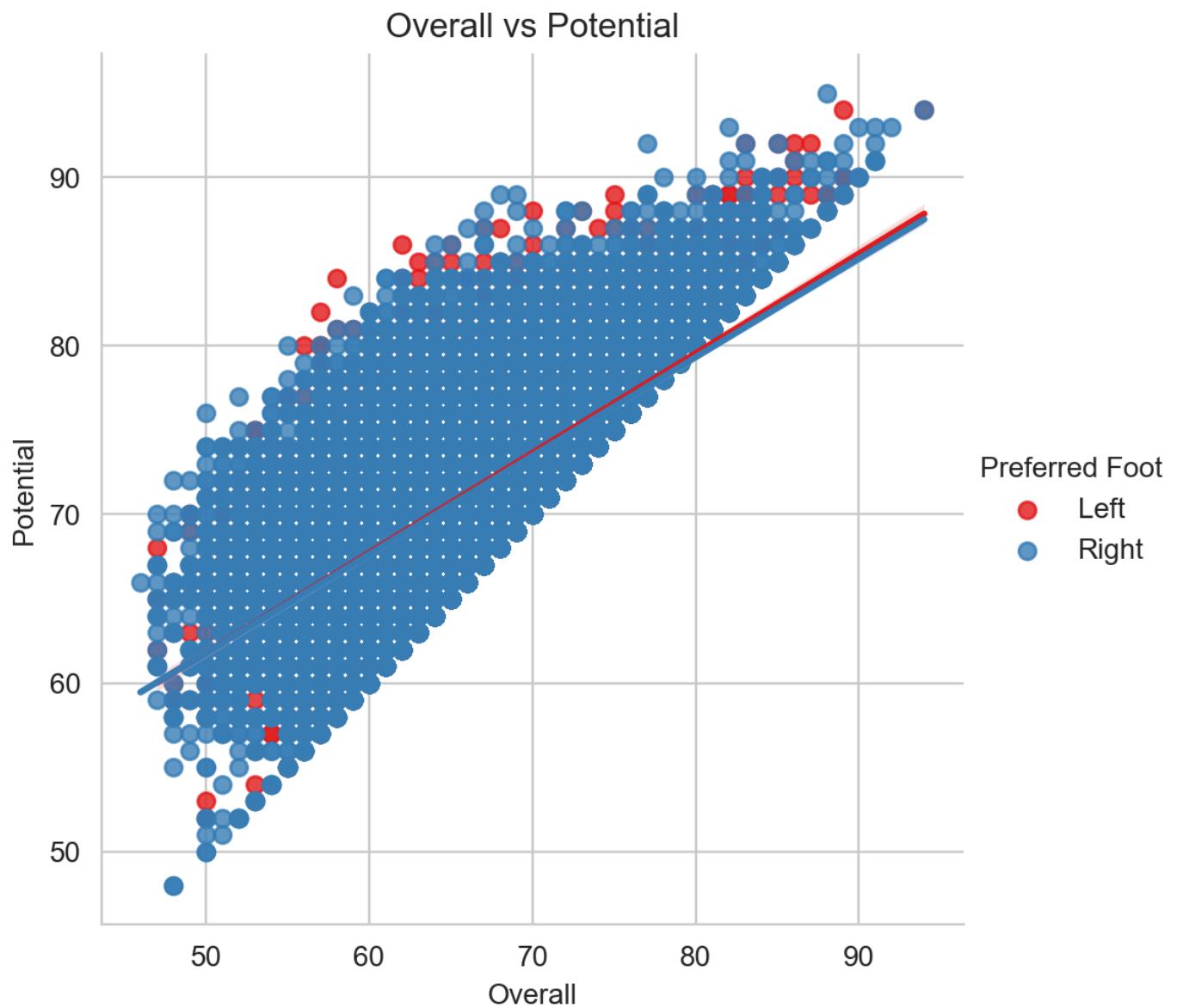


Seaborn lmplot()

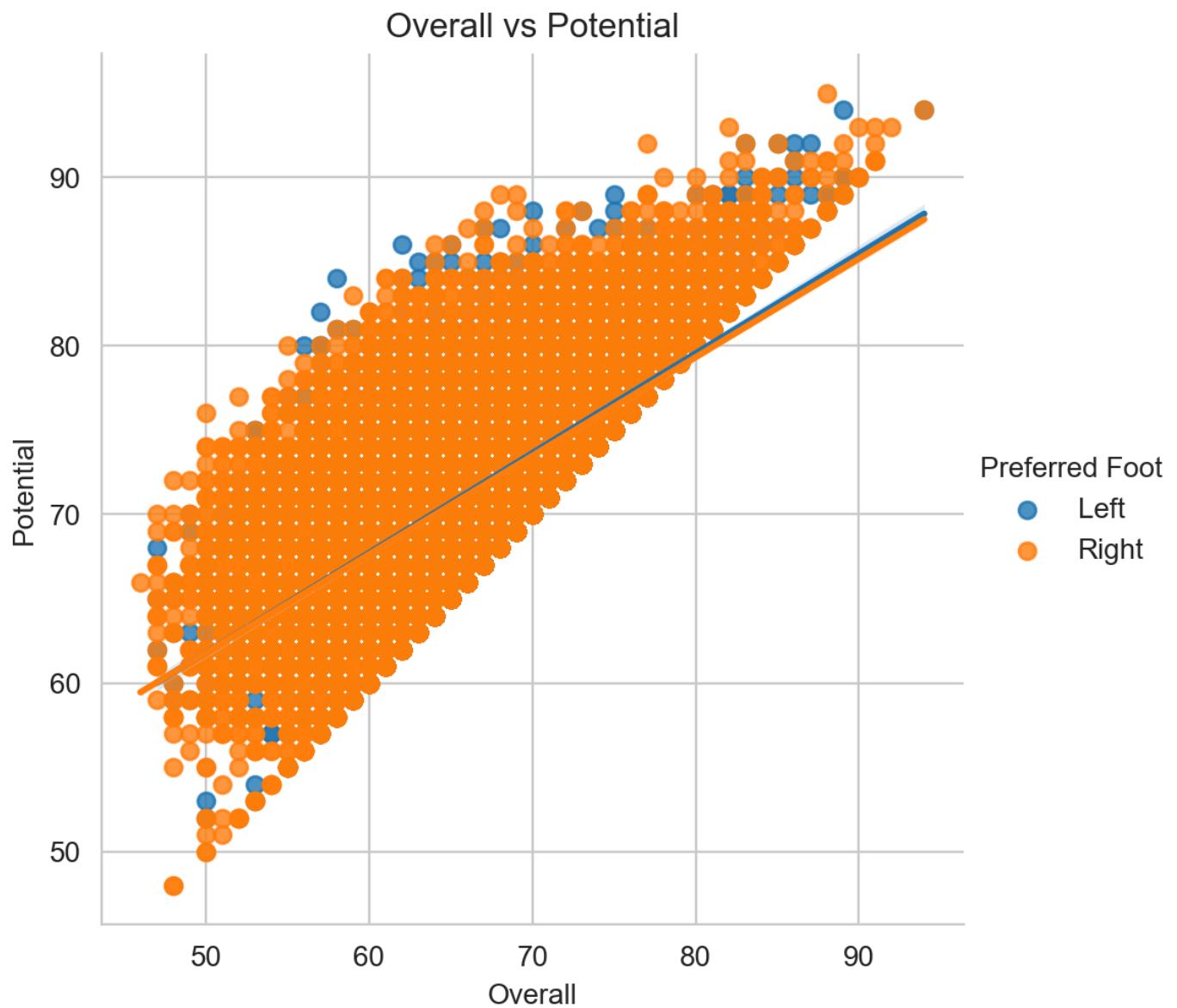
- `lmplot()` combines `regplot()` and `FacetGrid` for regression across subsets.



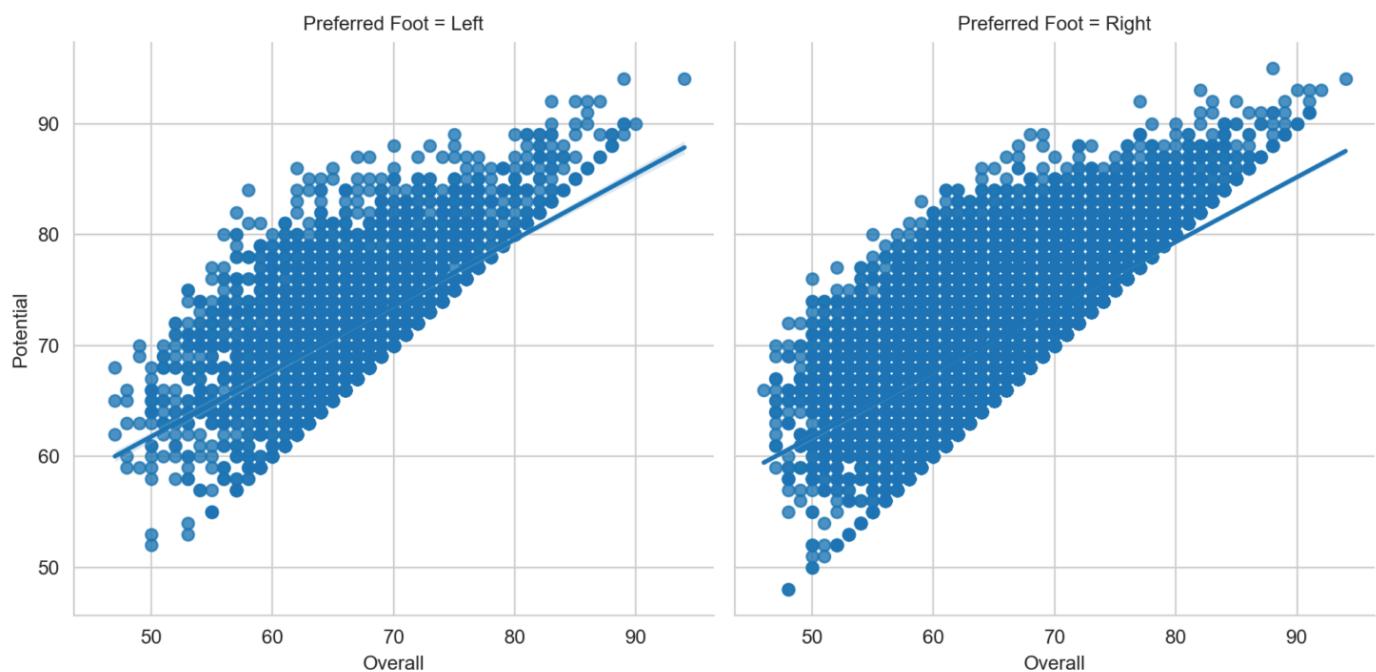
Different color palette:



Condition on third variable:



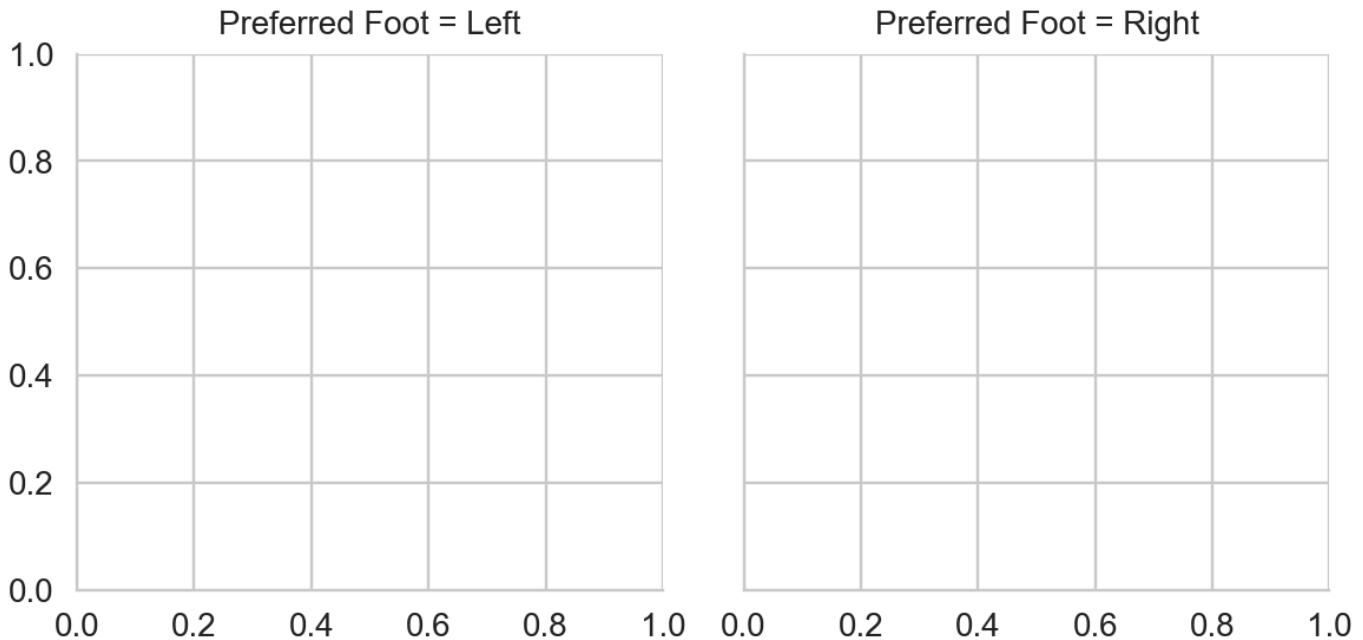
Across different columns:



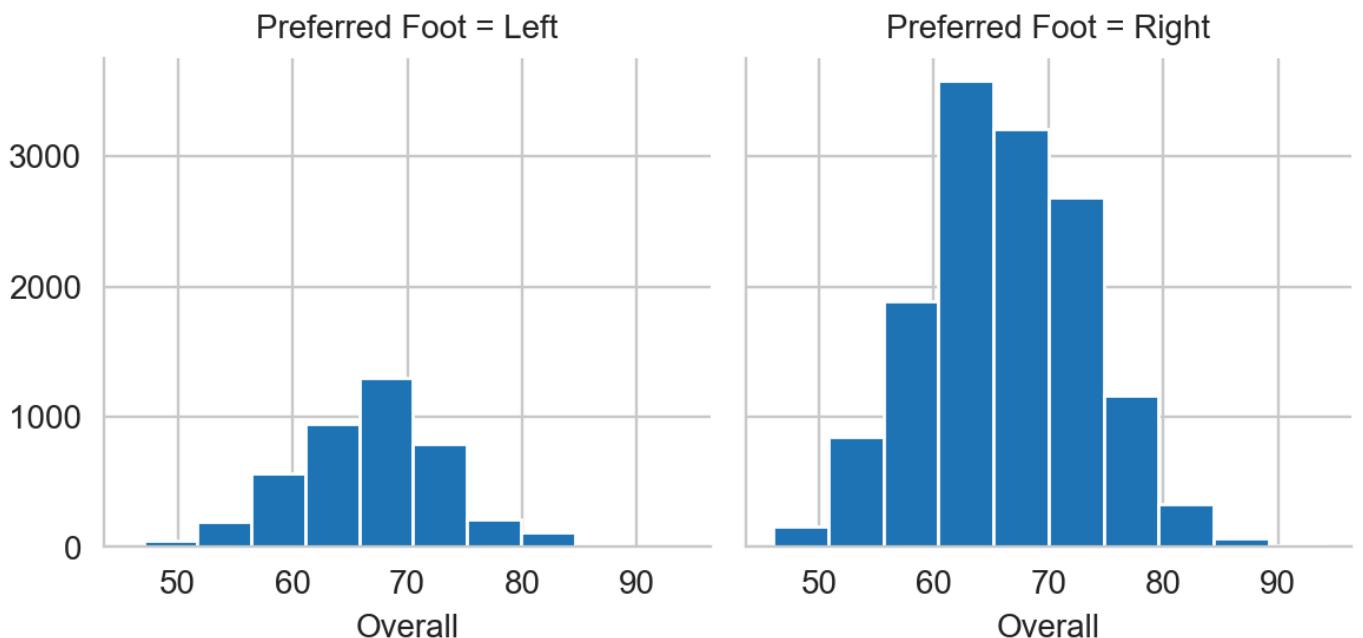
MULTI PLOT GRIDS

Seaborn `FacetGrid()`

- `FacetGrid` visualizes variable distributions within dataset subsets.

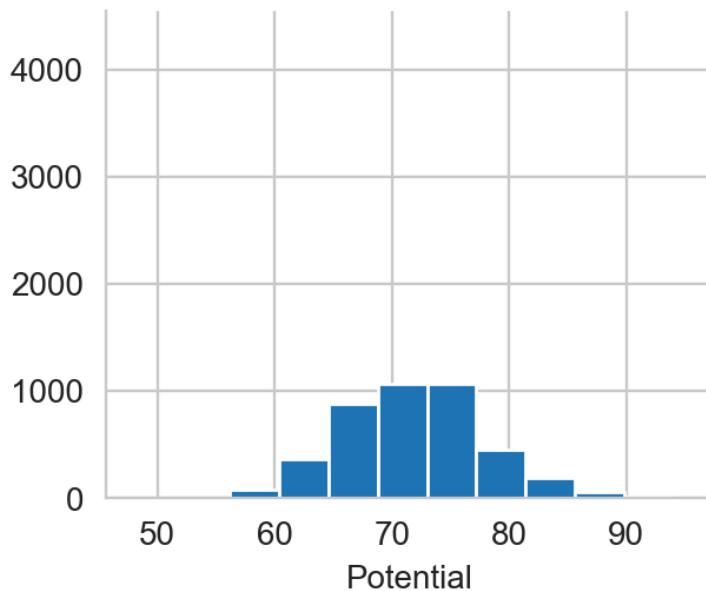


Univariate plot of `Overall`:

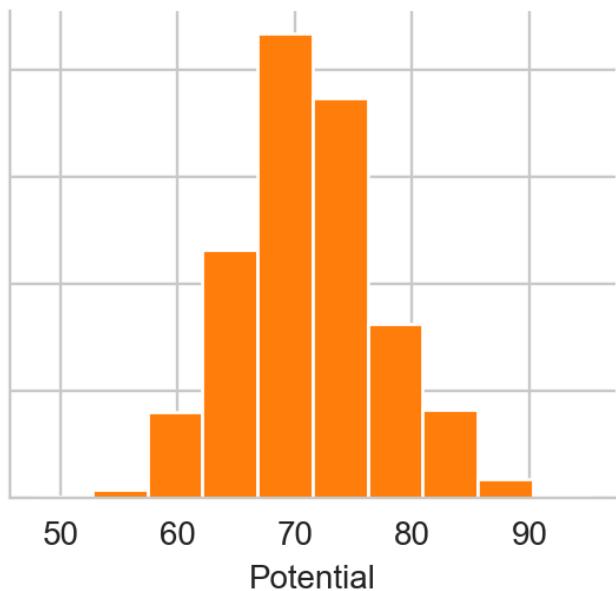


Univariate plot of `Potential`:

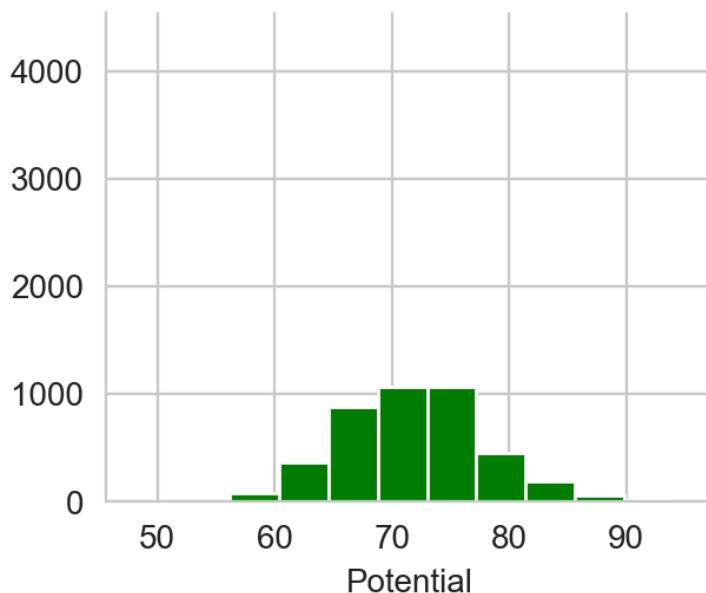
Preferred Foot = Left



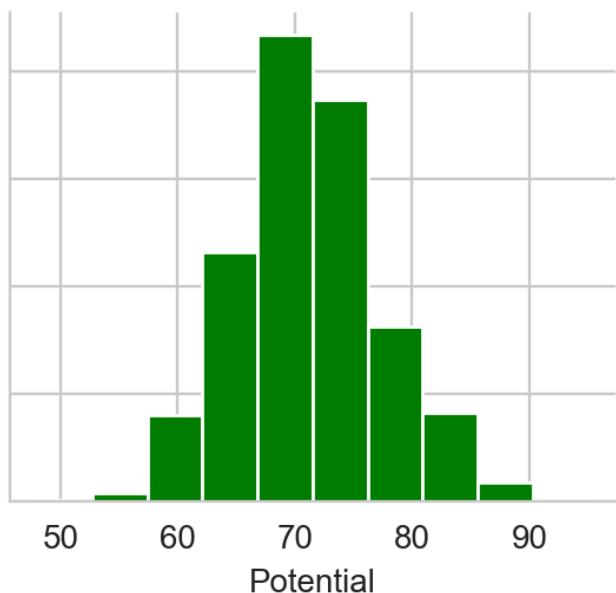
Preferred Foot = Right



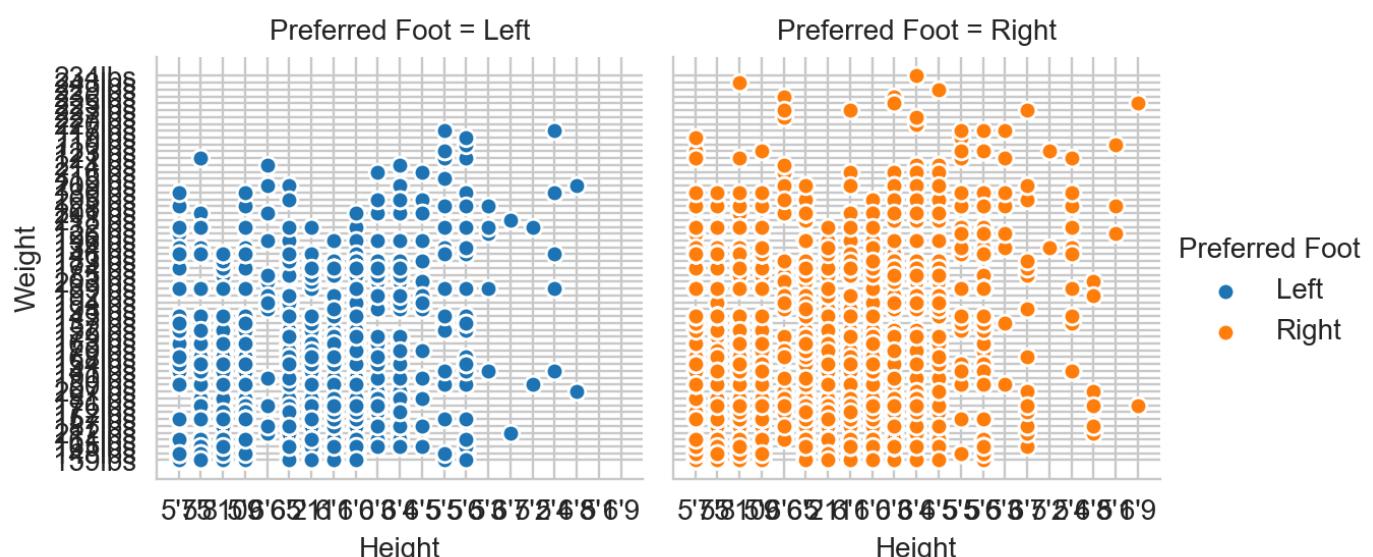
Preferred Foot = Left



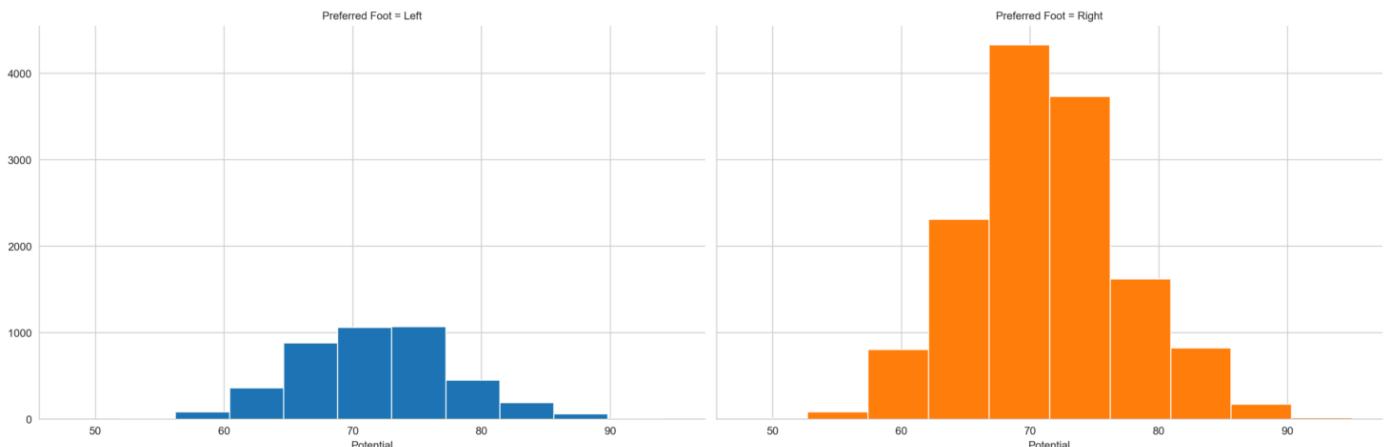
Preferred Foot = Right



Bivariate plot:

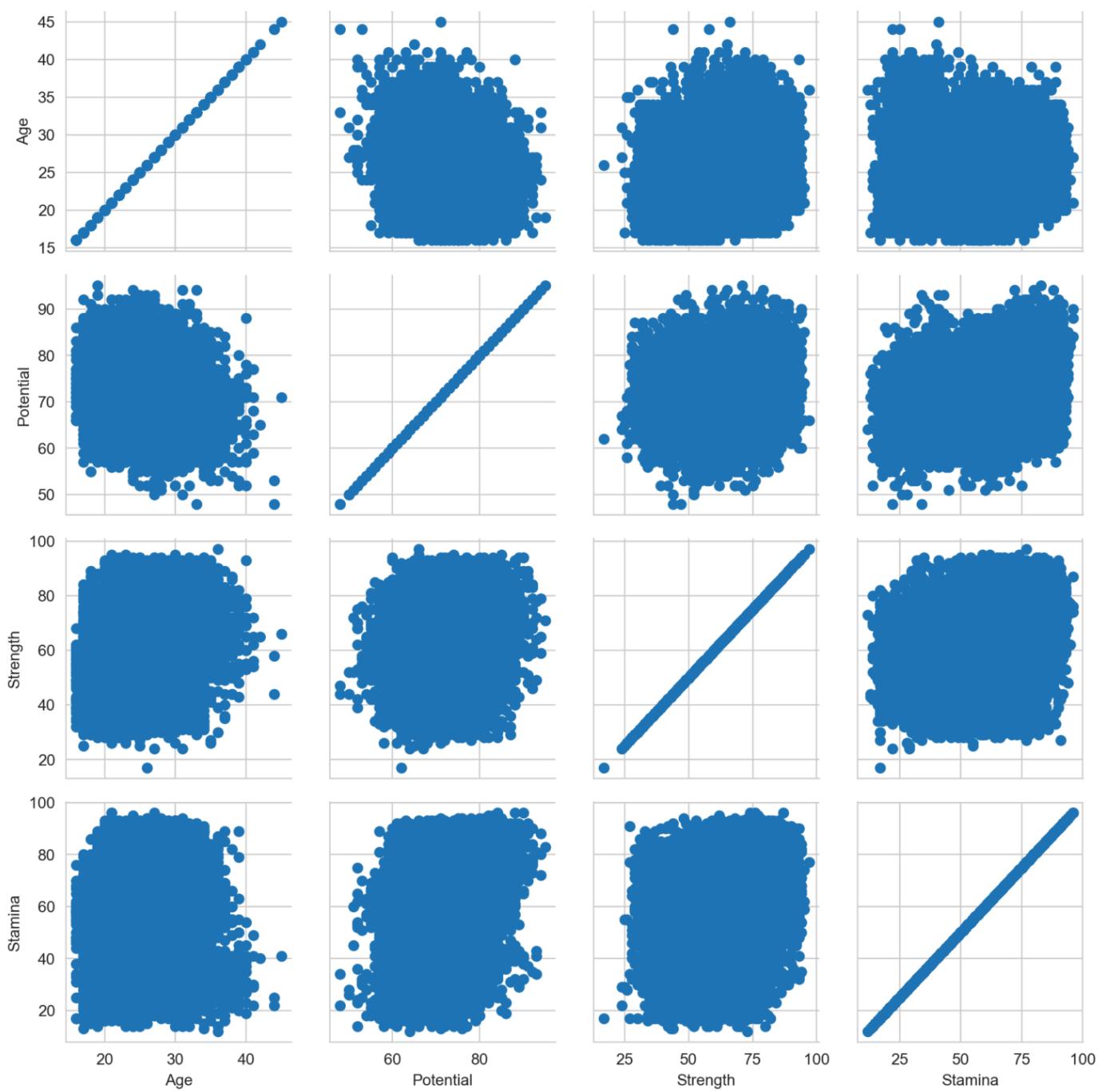


Custom size:

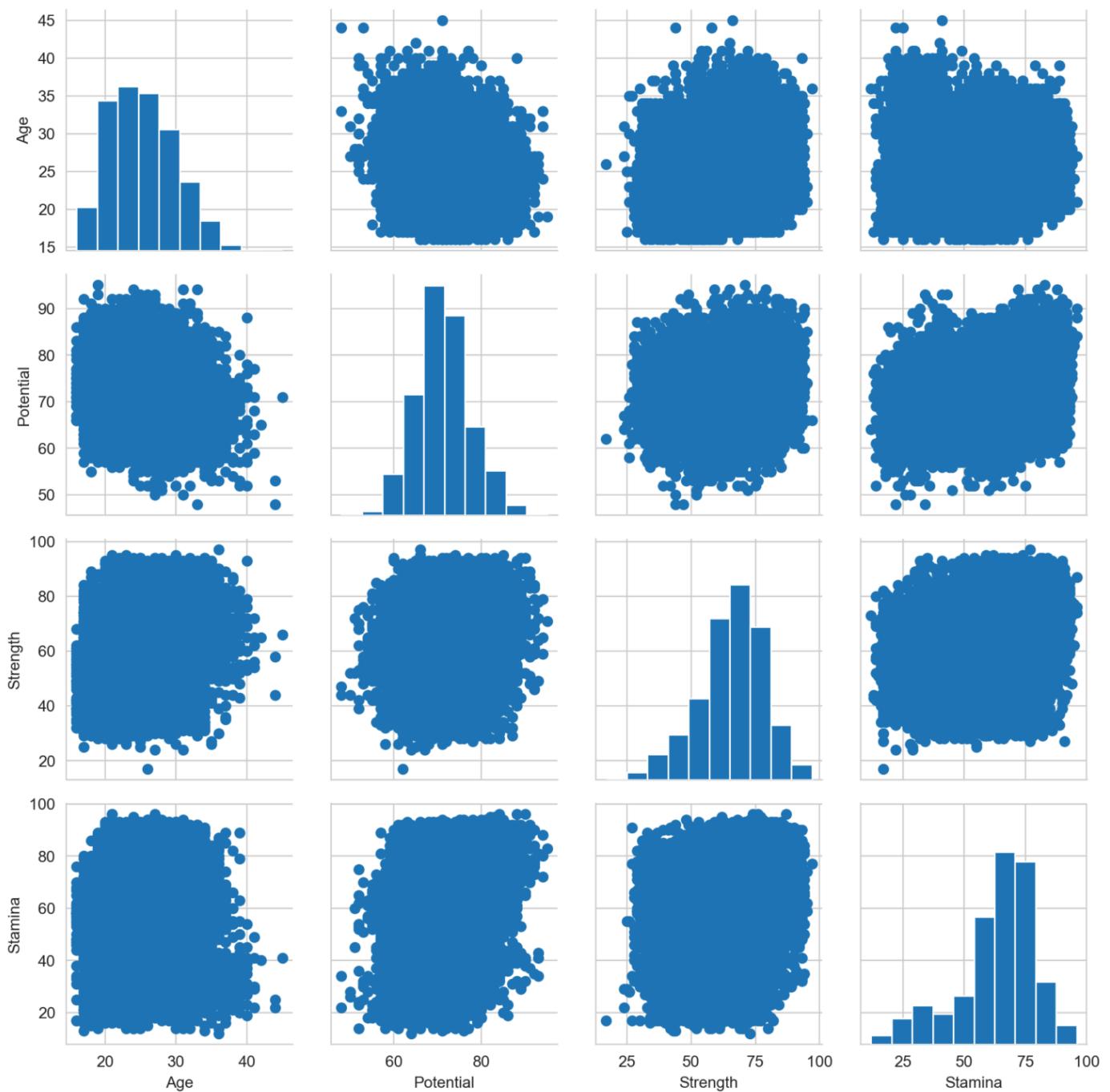


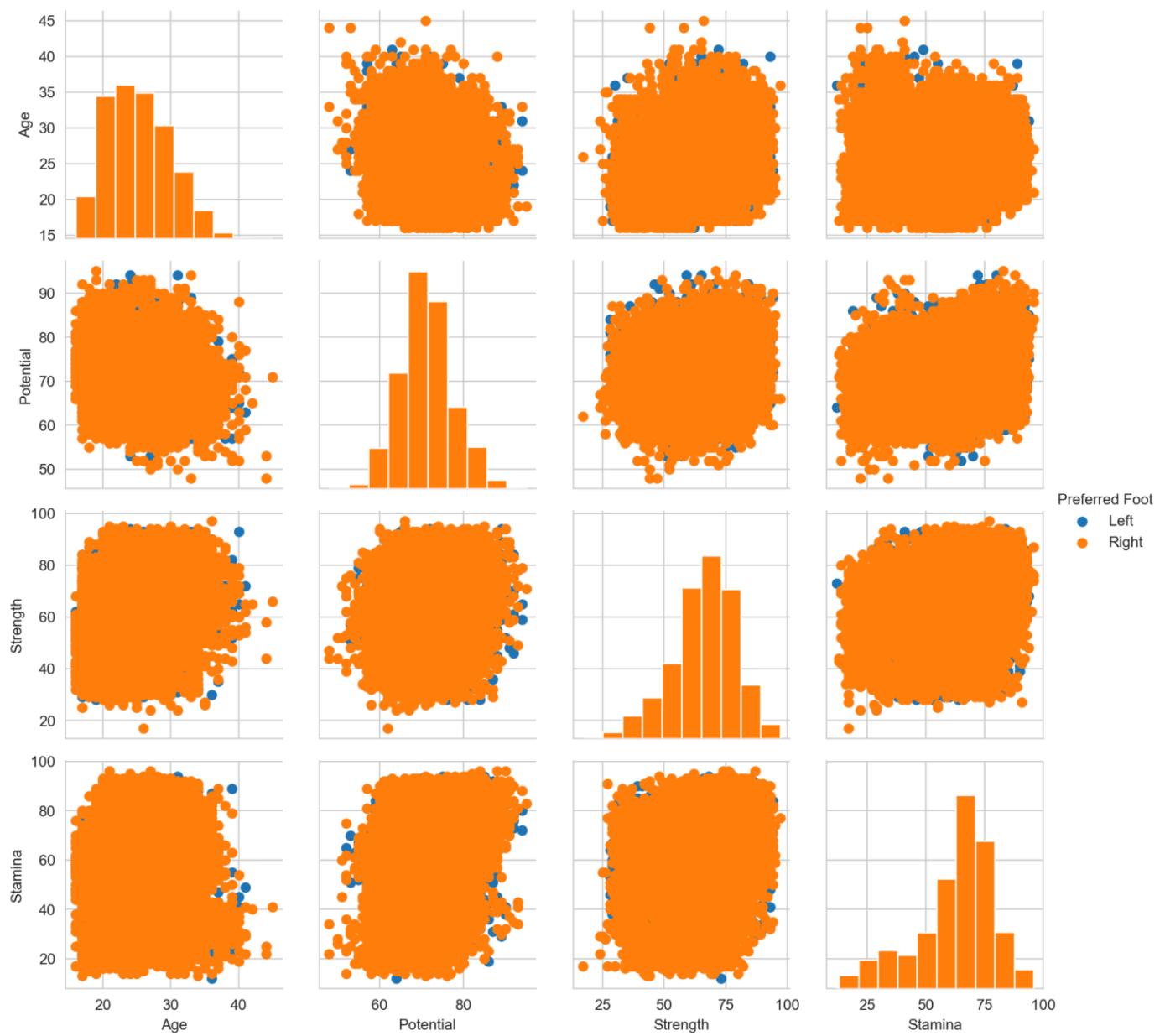
Seaborn `PairGrid()`

- `PairGrid` plots pairwise relationships in a grid.

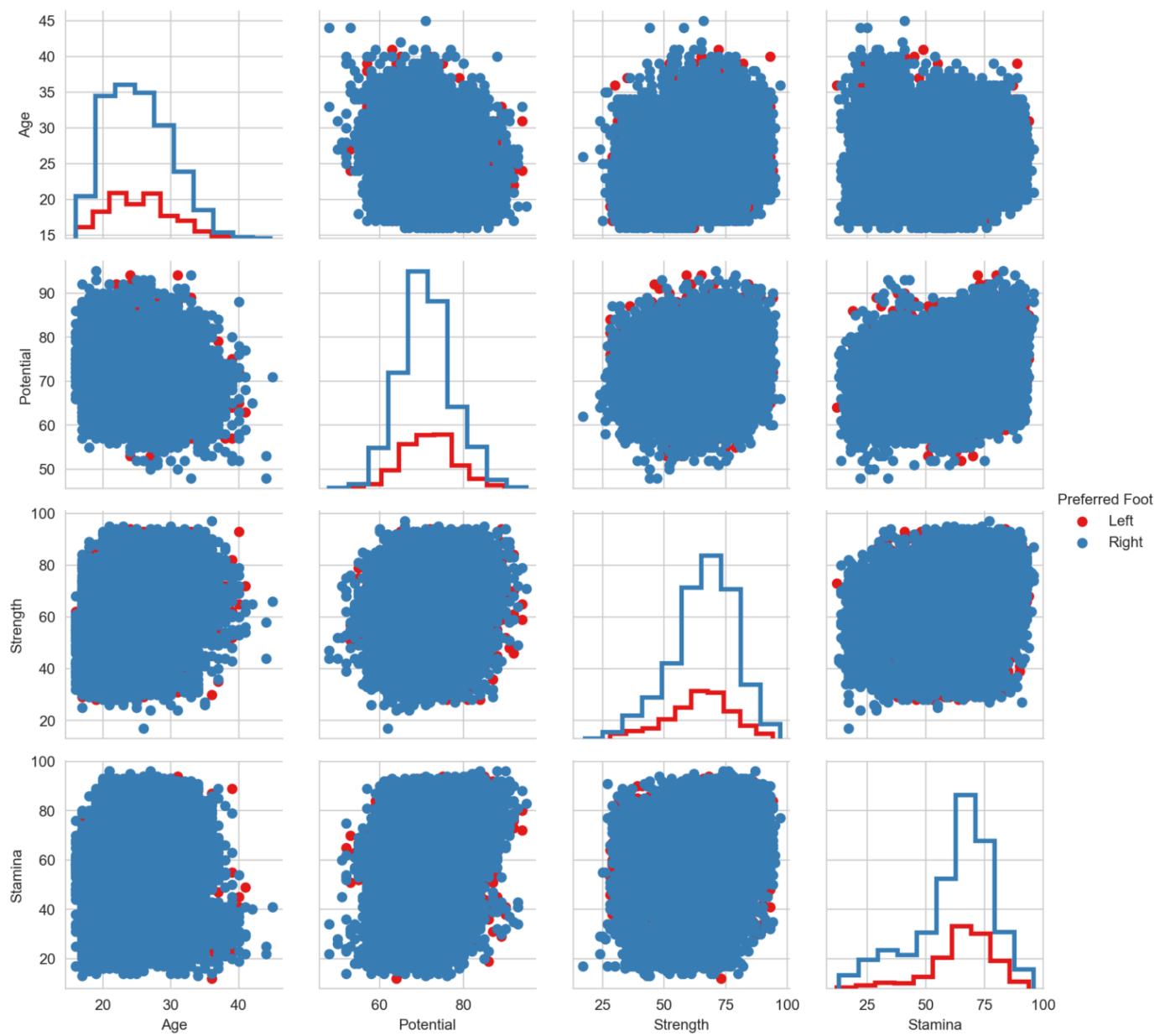


Univariate distribution on diagonal:

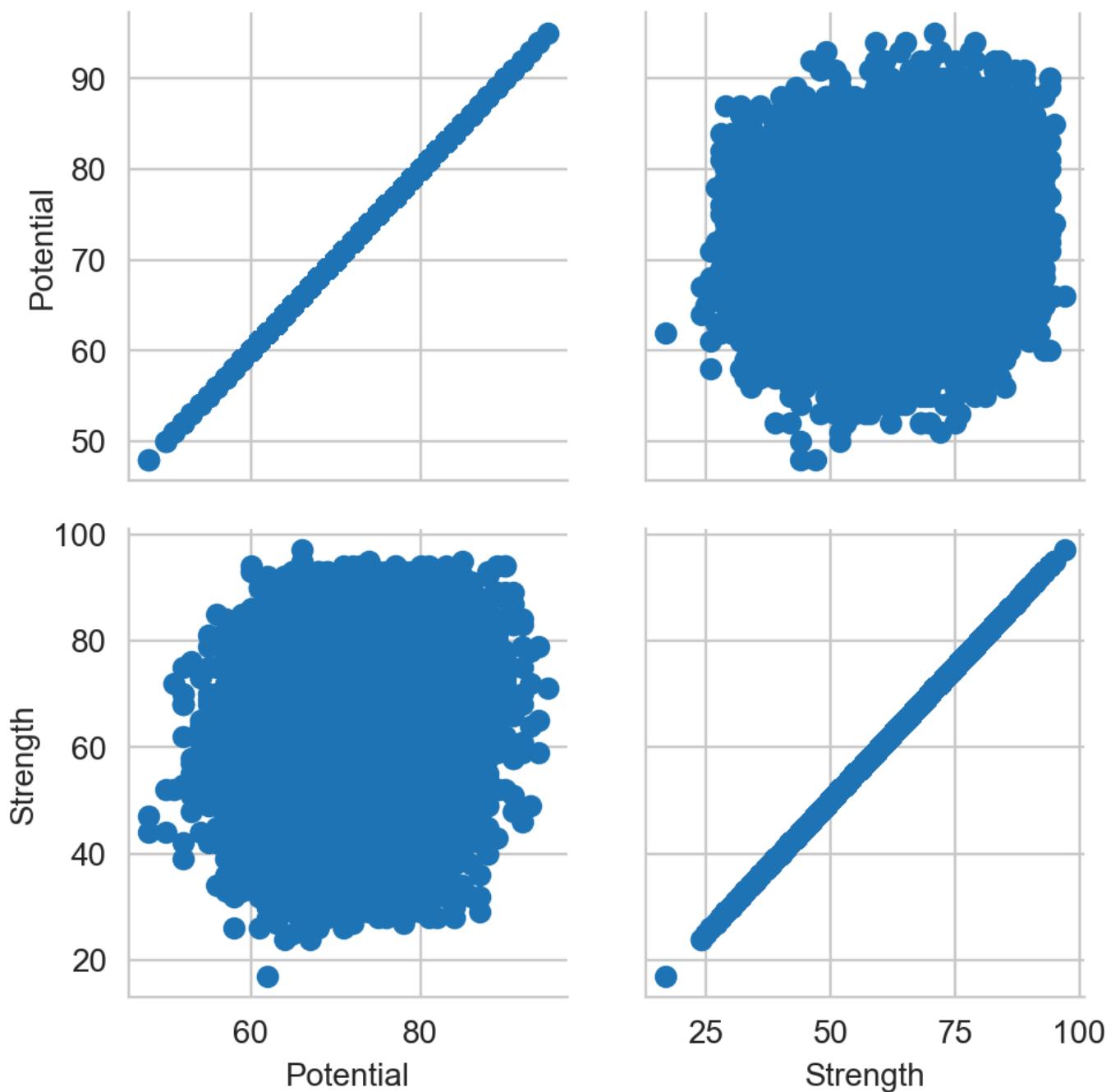


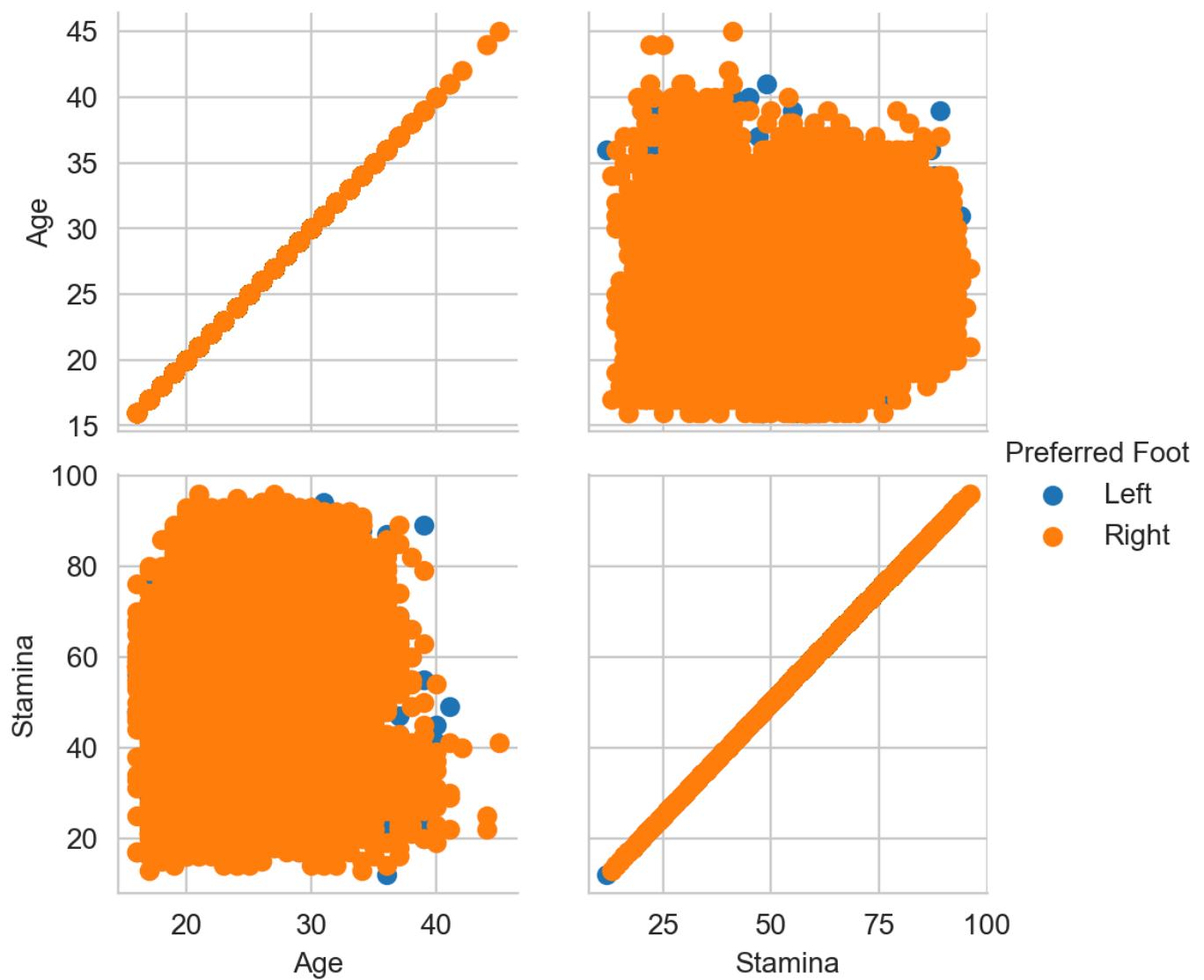


Different histogram style:

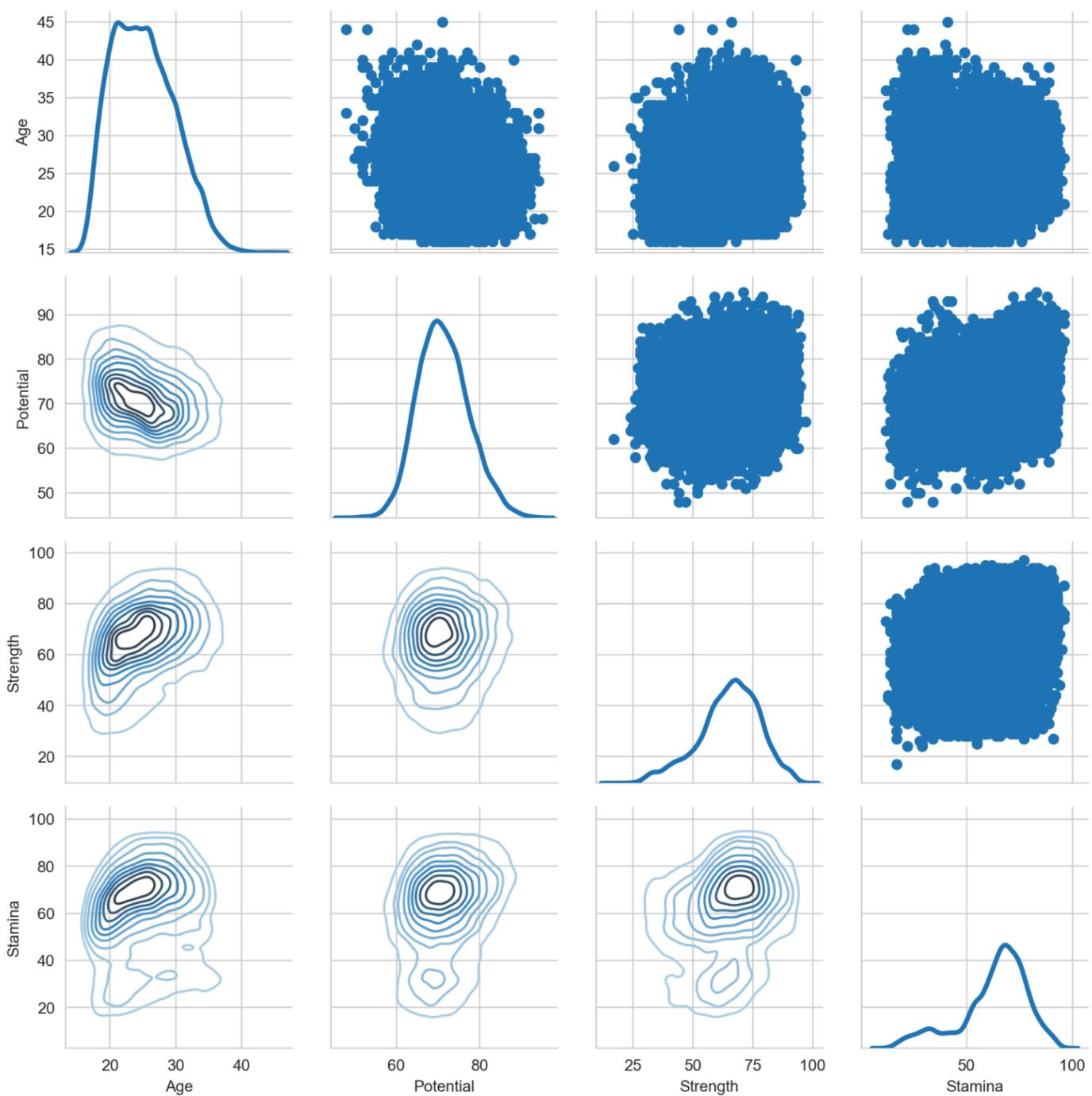


Subset of variables:



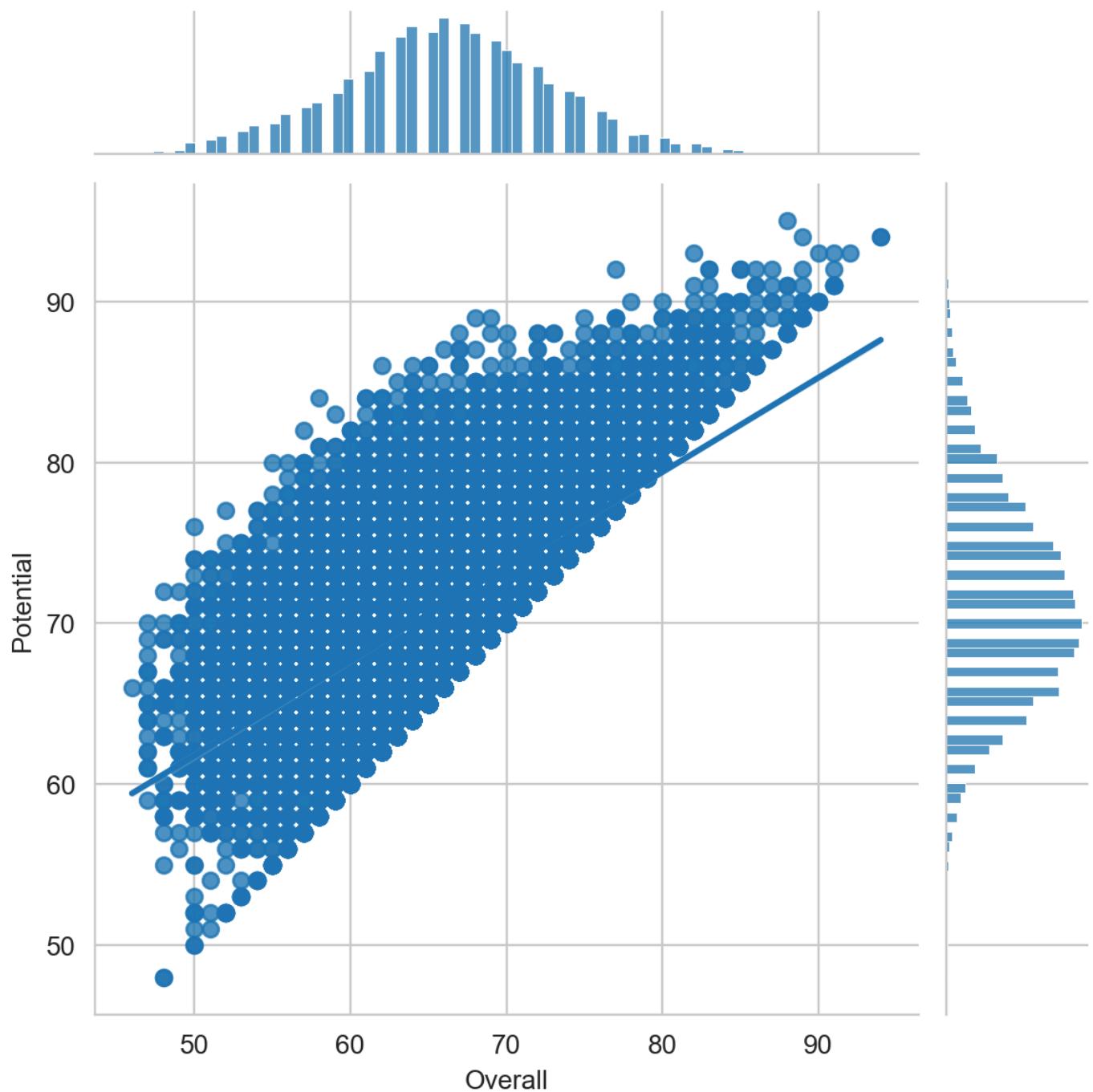


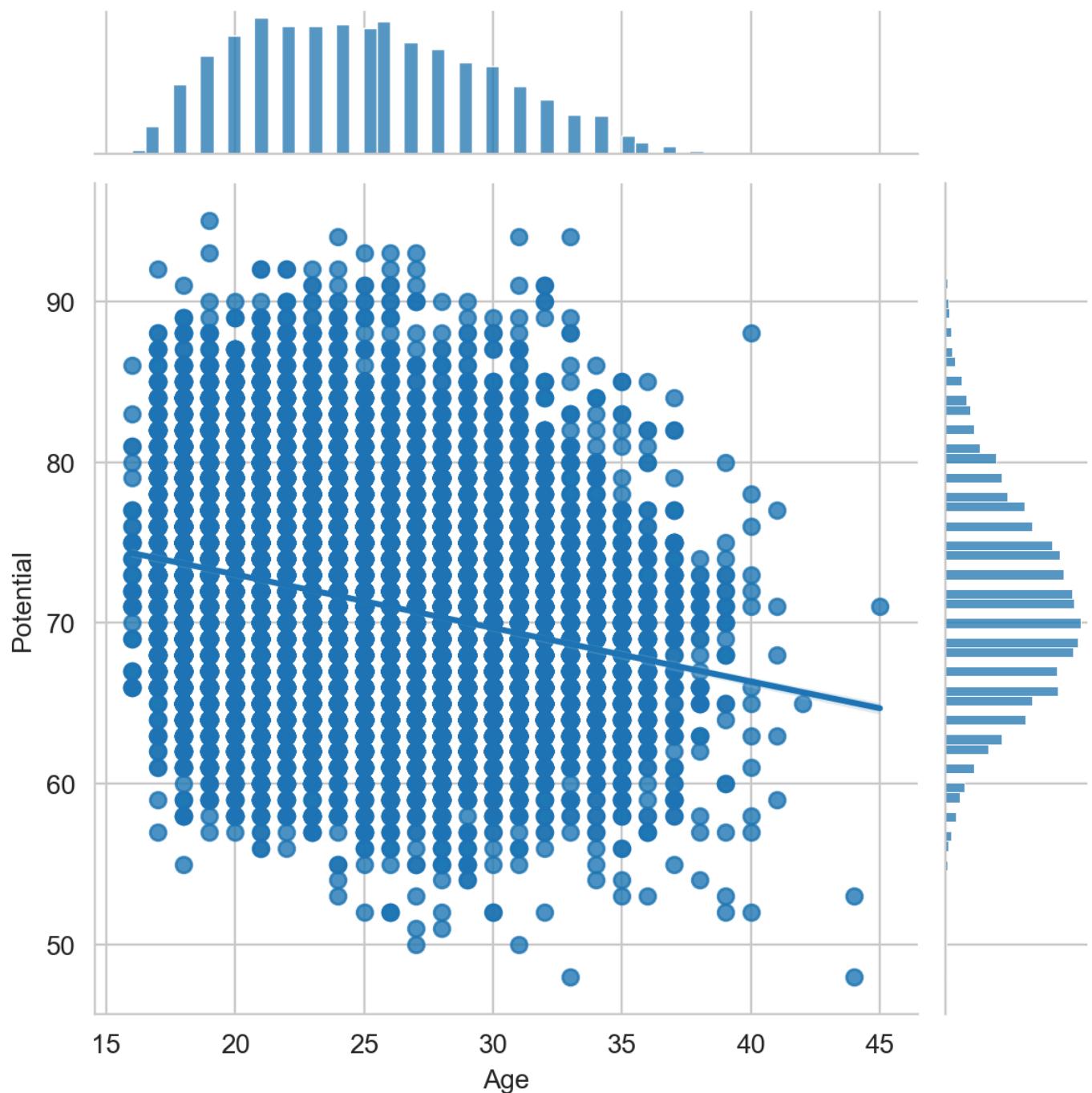
Different functions on triangles:



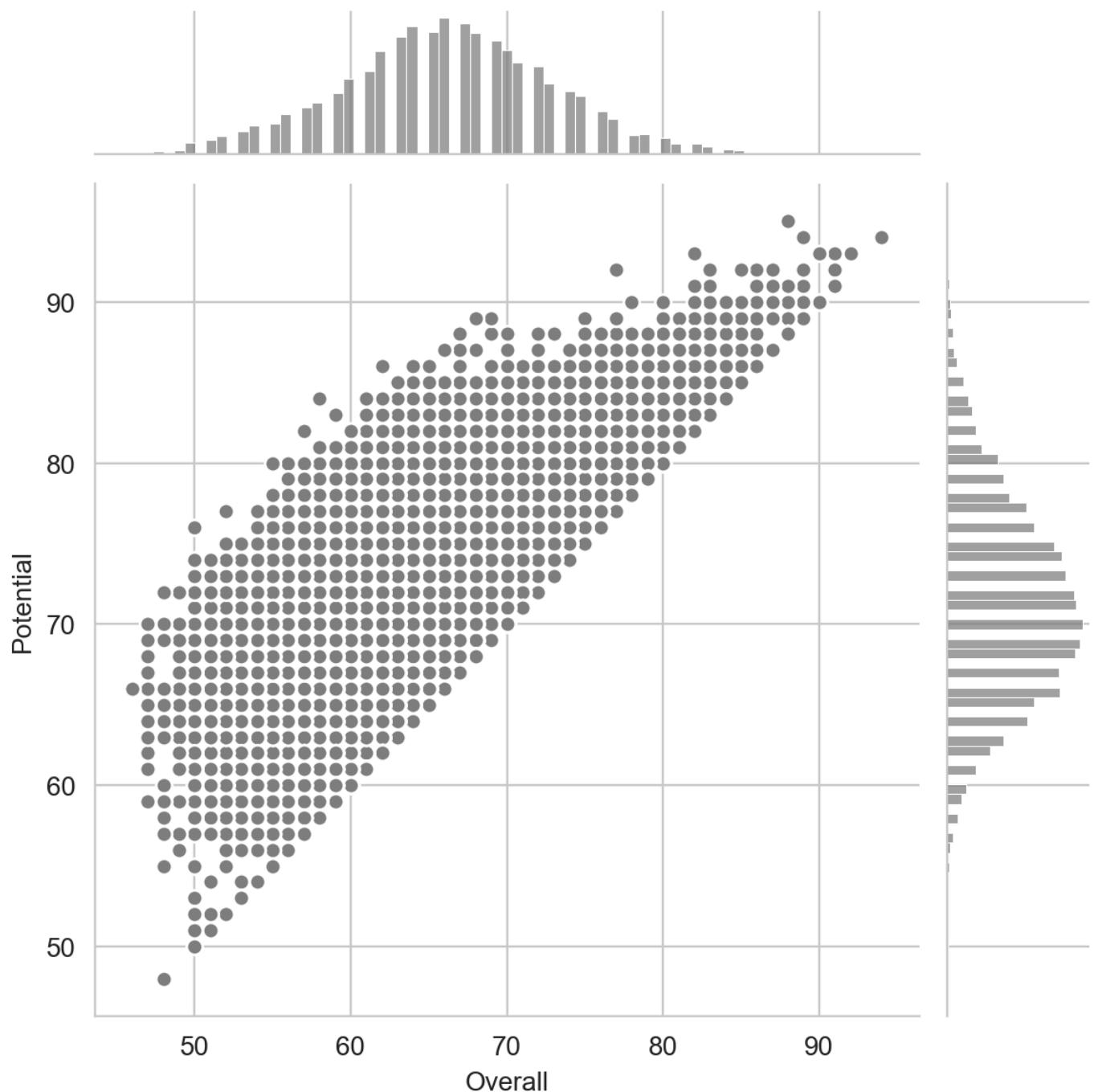
Seaborn `JointGrid()`

- `JointGrid` draws bivariate plots with marginal univariate plots.

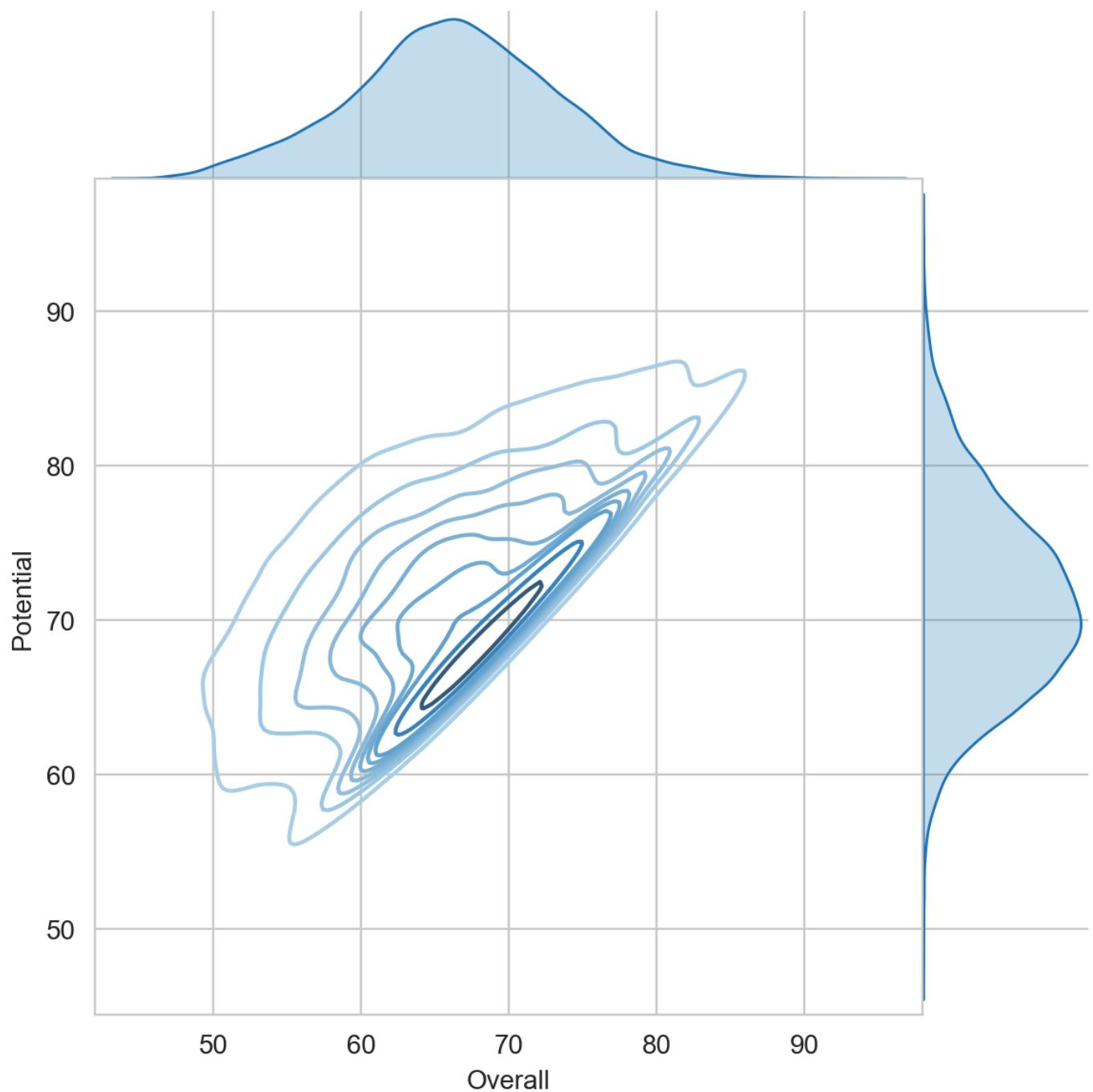




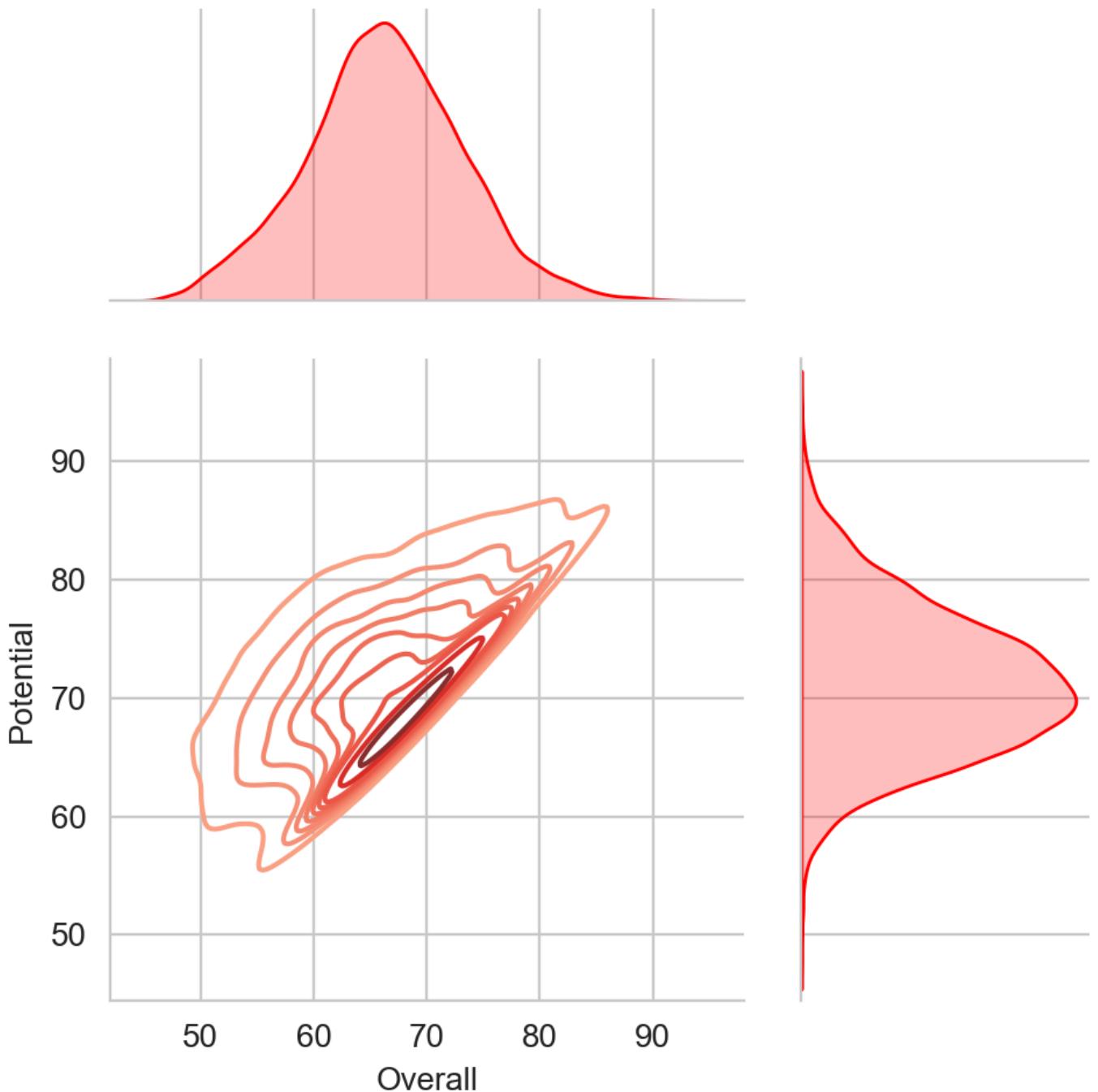
Separate joint and marginal plots:



No space between axes:

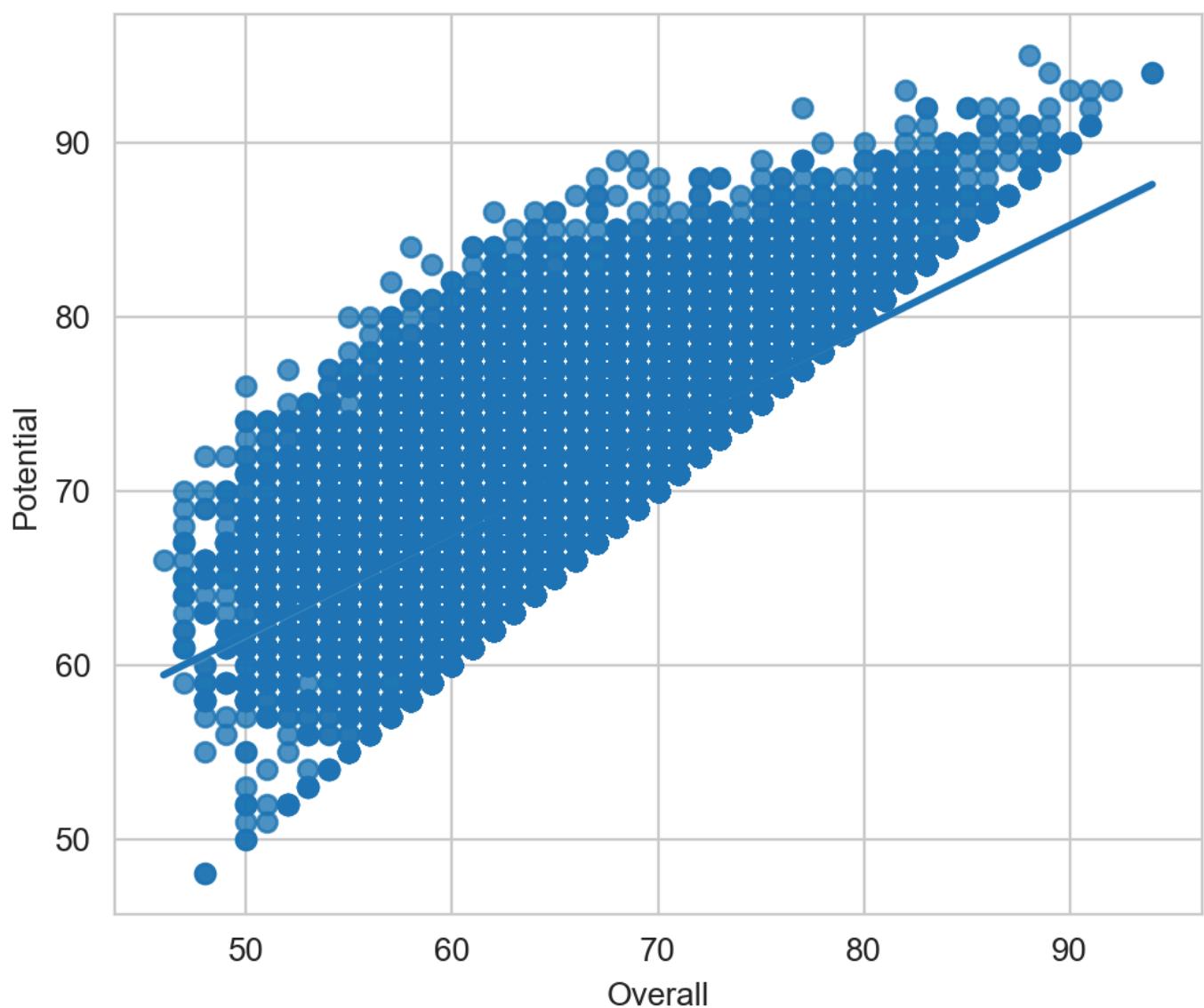


Smaller plot with larger marginals:

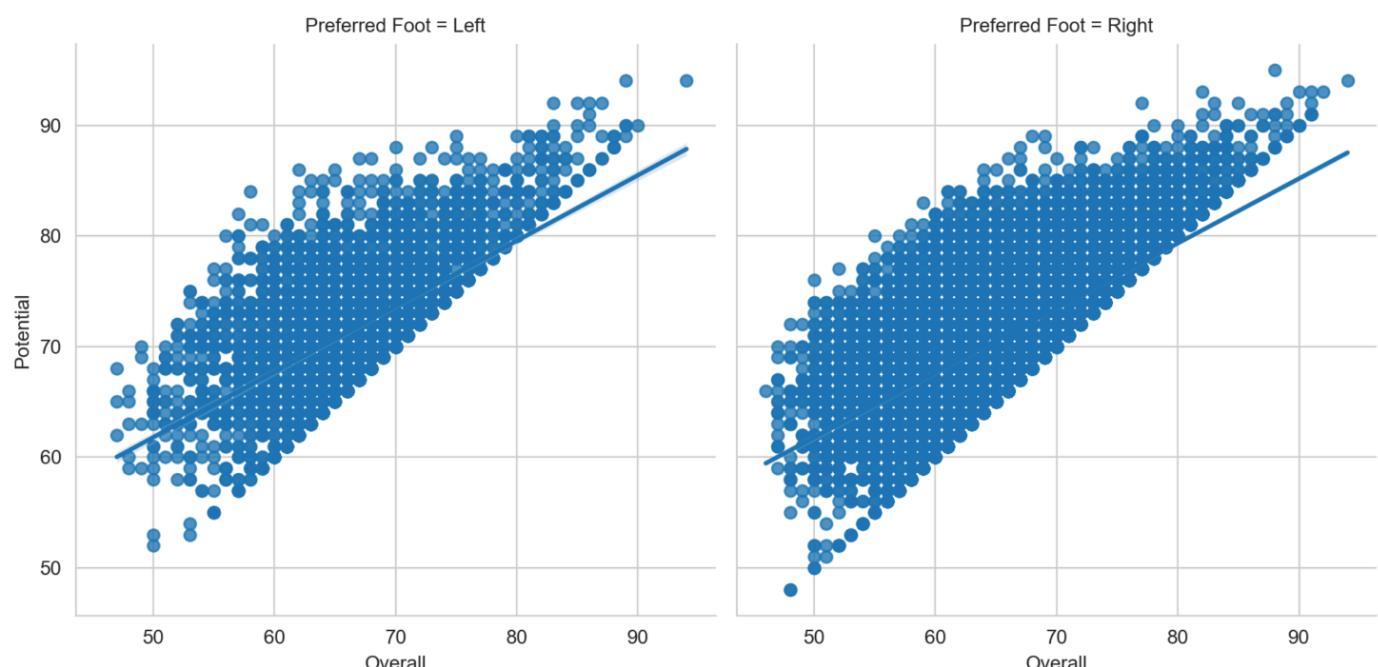


Controlling the size and shape of the plot

- `regplot()` size is controlled via Matplotlib figure.



`lmplot()` size is controlled via FacetGrid:





FIFA 2019