
THE PERFECT CHOCOLATE CHIP COOKIE

*Predicting Recipe Ratings
using Linear Regression*

Winston Ma



OVERVIEW

SUGAR, SPICE, AND EVERYTHING NICE...

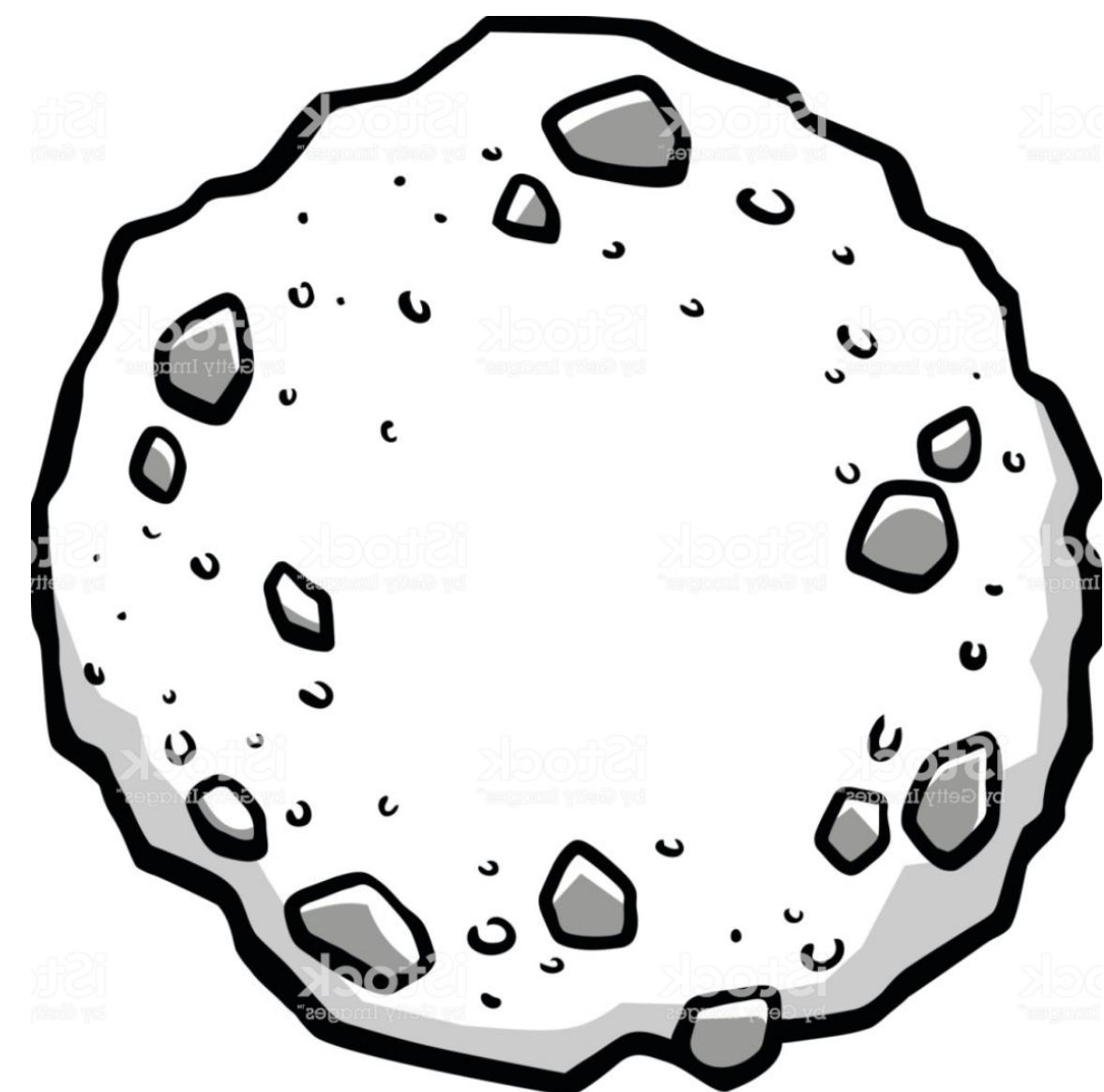
“...LIFE CHANGING!! ...the right balance of flavor and texture.”



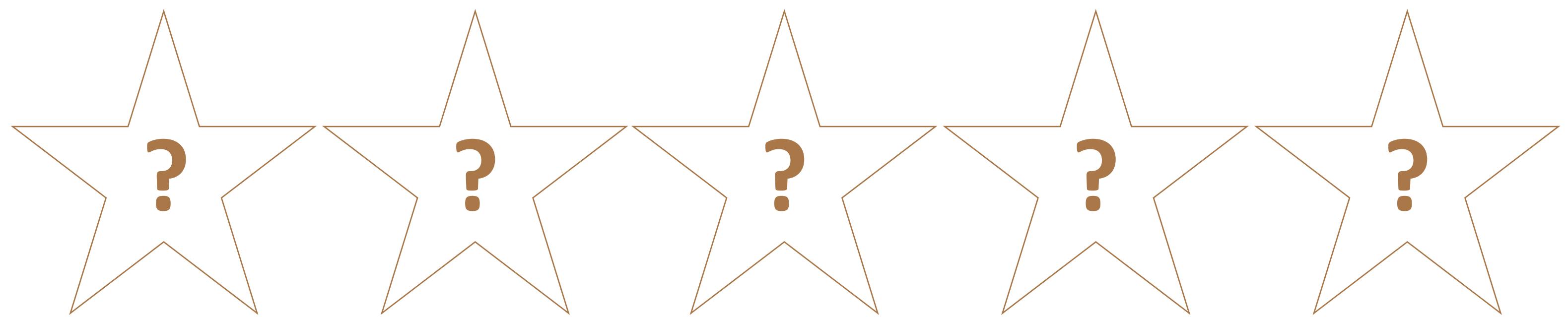
“...way too little flour for the amount of shortening and butter.”



OBJECTIVE

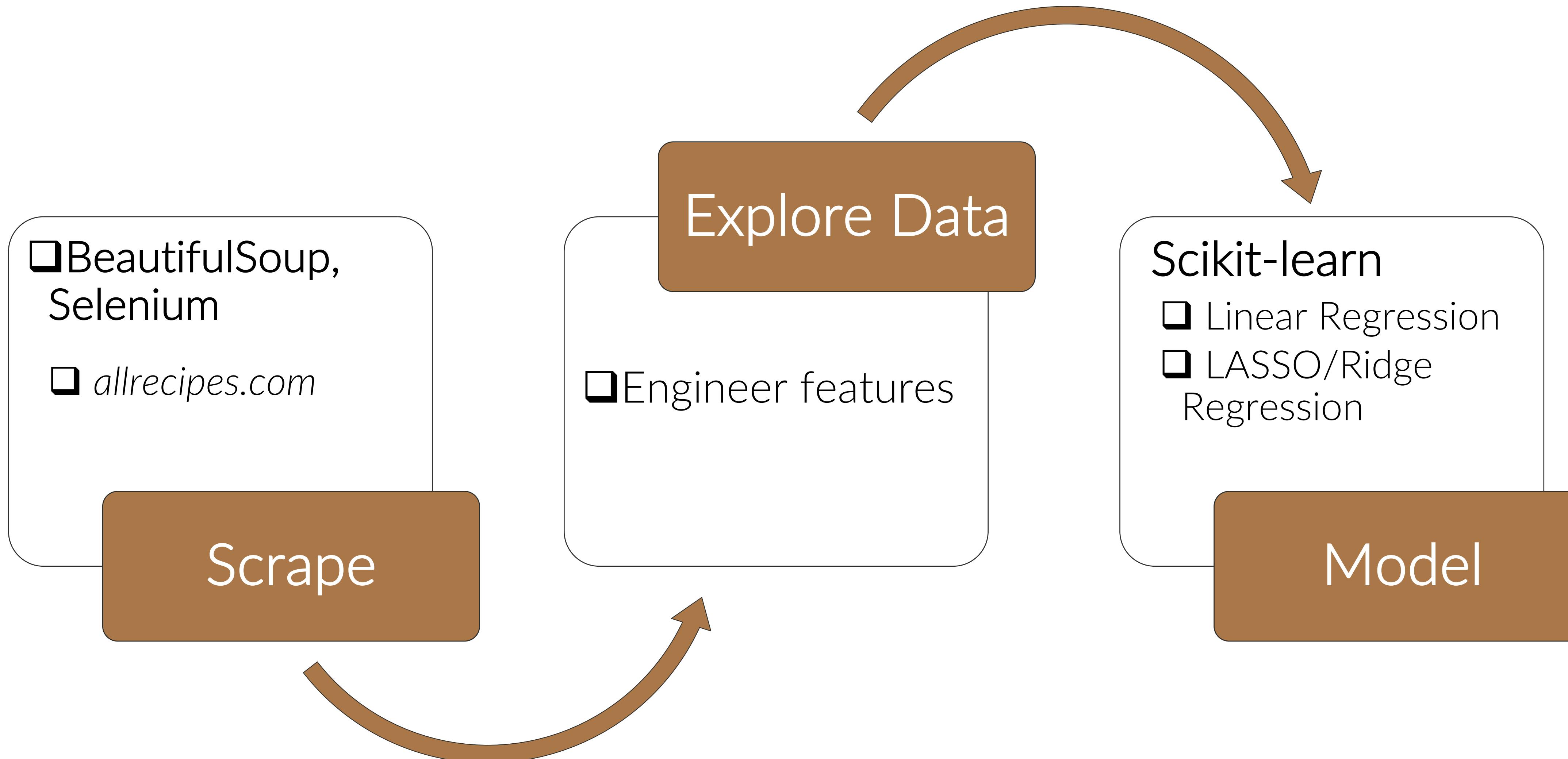


=



METHODOLOGY

METHODOLOGY



Outrageous Chocolate Chip Cookies



3k made it | 2k reviews | 482 photos

Recipe by: Joan

"A great combination of chocolate chips, oatmeal, and peanut butter."

Ingredients

1/2 cup butter
365 Everyday Value® Organic Butter
Everyday Savings: \$4.49 per 16 oz
[LEARN MORE](#)
ADVERTISEMENT

1 cup all-purpose flour
1 teaspoon baking soda
1/4 teaspoon salt
1/2 cup rolled oats

1/2 cup white sugar
1/3 cup packed brown sugar
1/2 cup peanut butter
1/2 teaspoon vanilla extract

1 cup semisweet chocolate chips
Ghirardelli Semi-Sweet Chocolate Chips
Make Cookies a Bite Better
[Buy Chips](#)
ADVERTISEMENT

Directions

Prep 15 m | Cook 10 m | Ready In 25 m
1 Preheat oven to 350 degrees F (175 degrees C).

2 In a medium bowl, cream together the butter, white sugar and brown sugar until smooth. Stir in the peanut butter, vanilla and egg until well blended. Combine the flour, baking soda and salt; stir into the batter just until moistened. Mix in the oats and chocolate chips until evenly distributed. Drop by tablespoonfuls on to lightly greased cookie sheets.



25 m 18 servings 207 cals

On Sale

What's on sale near you.

Whole Foods Market
270 Greenwich Street
NEW YORK, NY 10007
Sponsored

Add a note Print Watch

You might also like



Outrageous Chocolate Chip Cookies
Peanut butter, oats, and chocolate chips create an outrageous treat.

Get the magazine



Get a full year for just \$10!
Cook 5-star weekday dinners every time.

SCRAPED FEATURES

 TARGET: RatingPopularity:

- Number of ratings
- Number of reviews
- Number of photos

Ingredients (grams):

- Butter
- Flour
- White Sugar
- Brown Sugar
- Chocolate Chips
- Etc.

Recipe Description:

- Preparation time
- Servings per recipe
- Calories per serving

Recipe Instructions:

- Oven temperature

ENGINEERED FEATURES

EXAMPLES

- Cost per recipe (external effect)
- Total number of ingredients (complexity)
- Combined ingredient types:
 - Fats (butter, shortening)
 - Contains nuts (categorical)
- Created ingredient ratios:
 - White sugar : brown sugar
 - Fats : flours

“Fantastic texture and balance...I used **half butter and half shortening for the fat** and they came out perfect!”



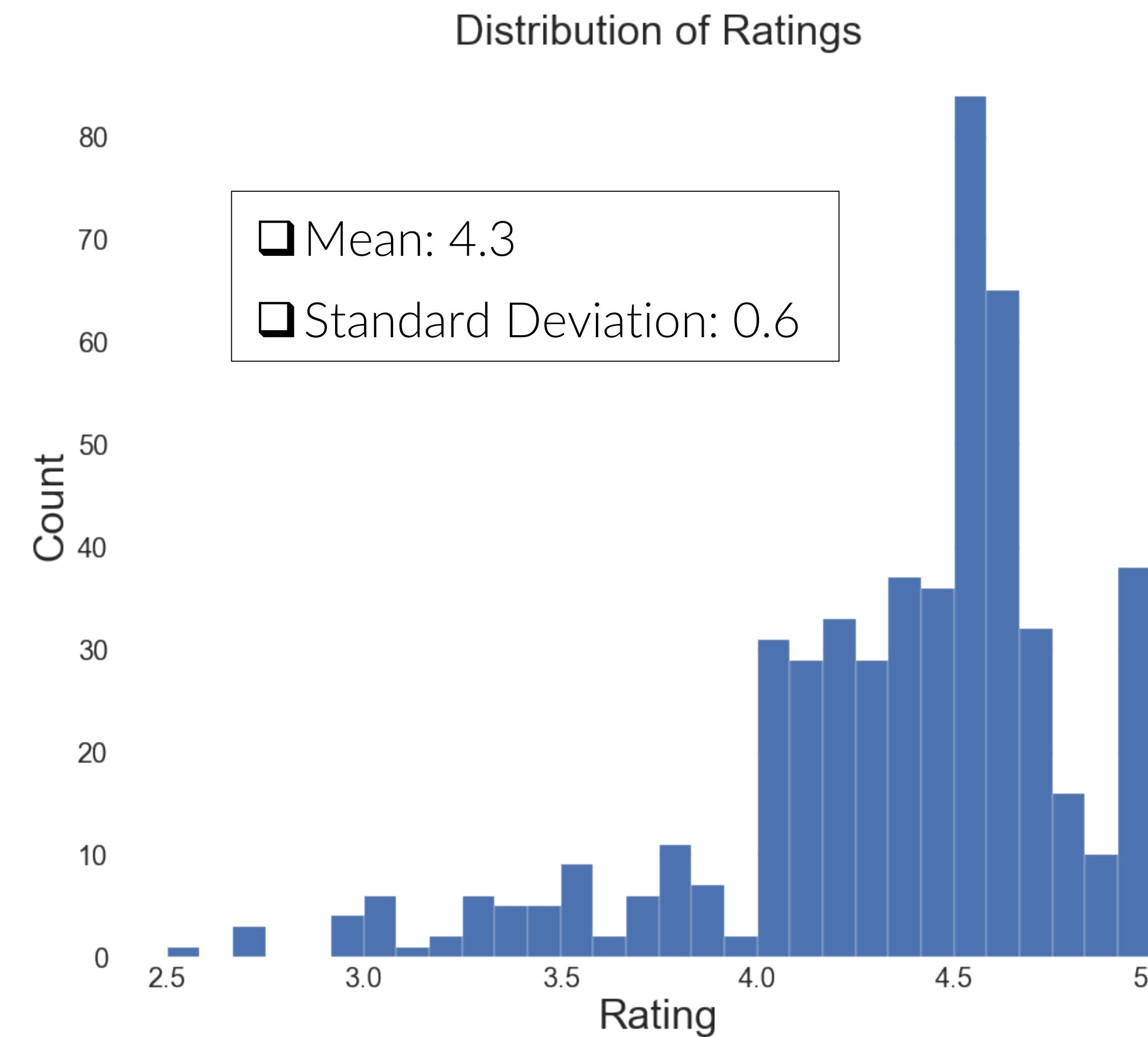
“waaaaayyyy too sweet!!!! ...**cut the sugar and replace with brown sugar!!**



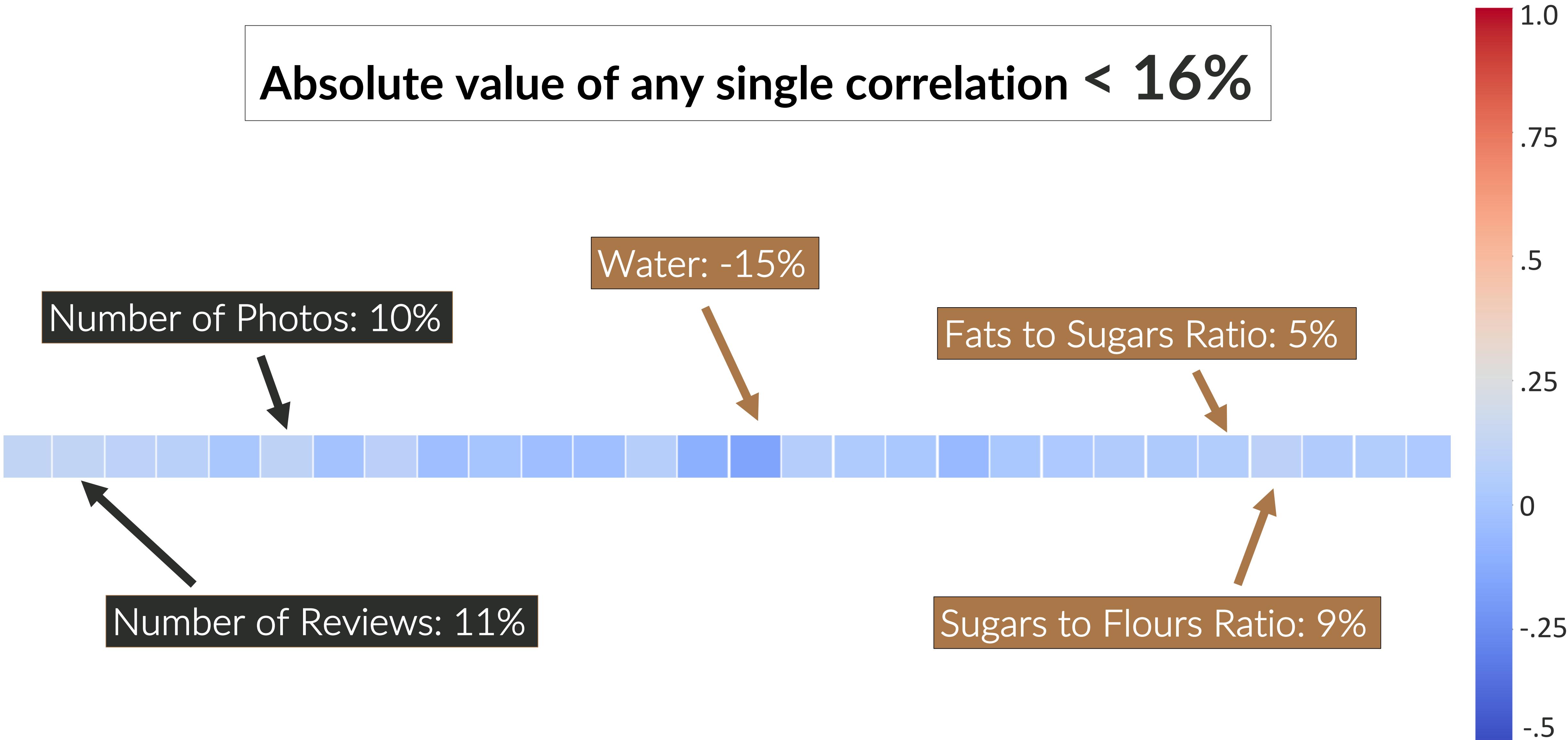
“...way **too little flour for the amount of shortening and butter.**”



TARGET VARIABLE DISTRIBUTION



CORRELATIONS WITH TARGET VARIABLE



LINEAR REGRESSION RESULTS

LINEAR REGRESSION RESULTS

25 Features

510 Recipes

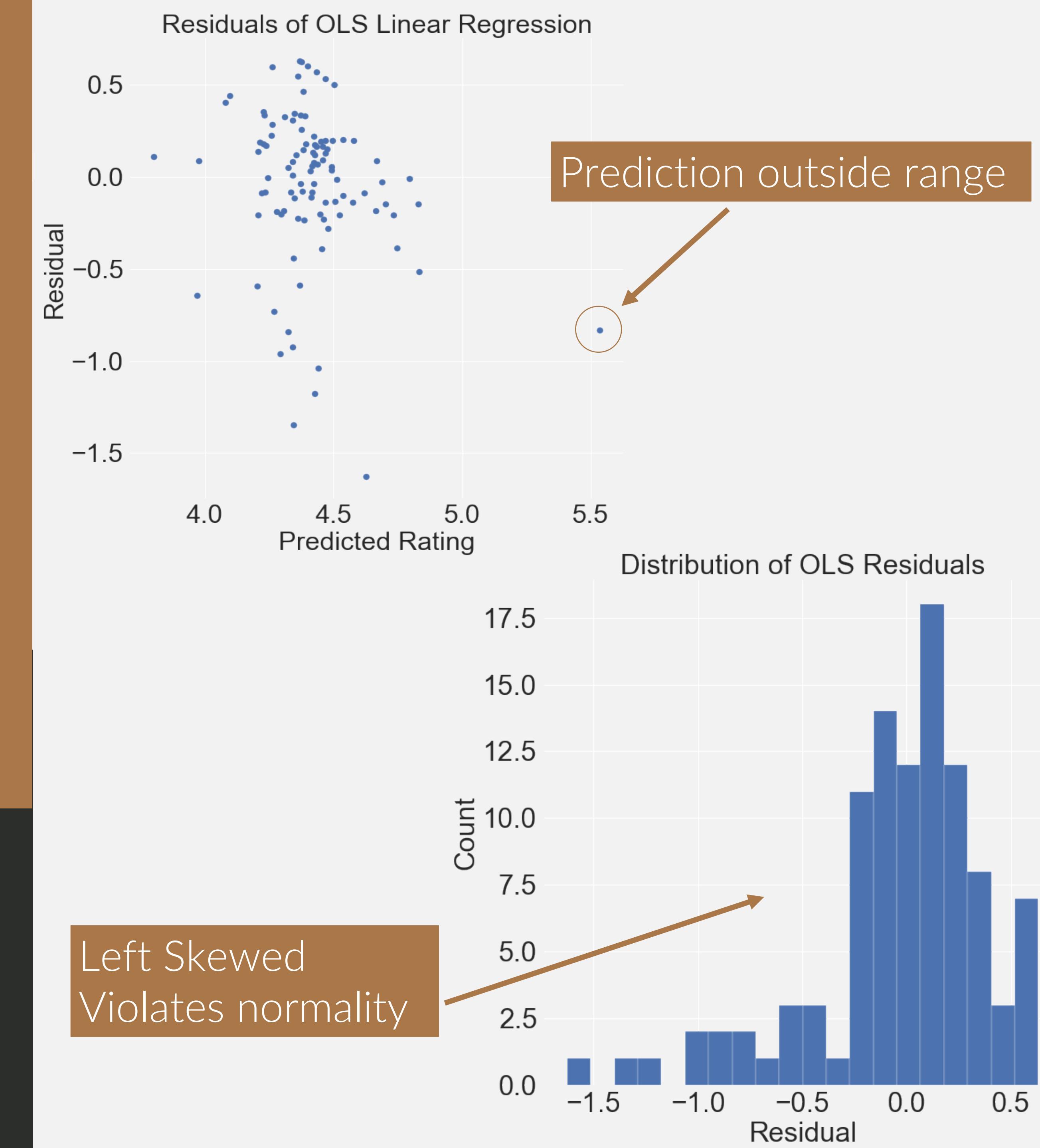
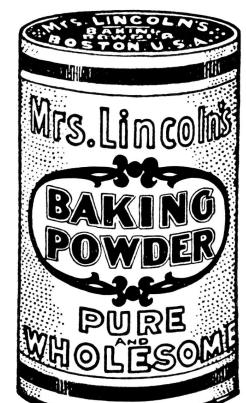
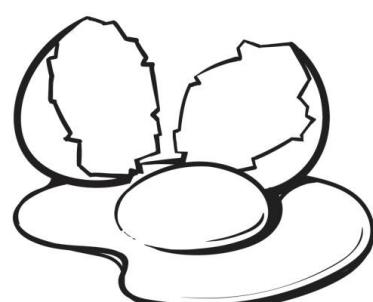
$$R^2 = 0.062$$

RESULTS ANALYSIS

Cross-Validated Results

- ❑ Ordinary Least Squares $R^2 = 0.025$
- ❑ Polynomial Features added more noise
- ❑ LASSO Regression $R^2 = 0.062$
- ❑ Ridge Regression $R^2 = 0.058$

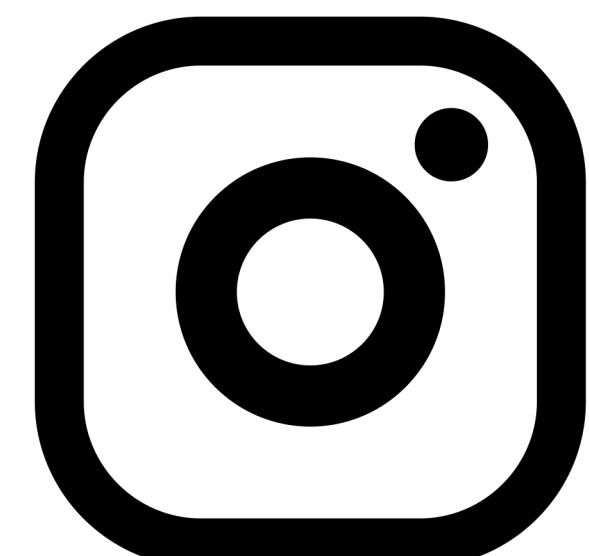
Feature	OLS Coefficient	p-value
Eggs	-1.159	0.081
Vanilla Extract	0.251	0.067
Baking Powder	-0.348	0.10



NON-LINEAR?

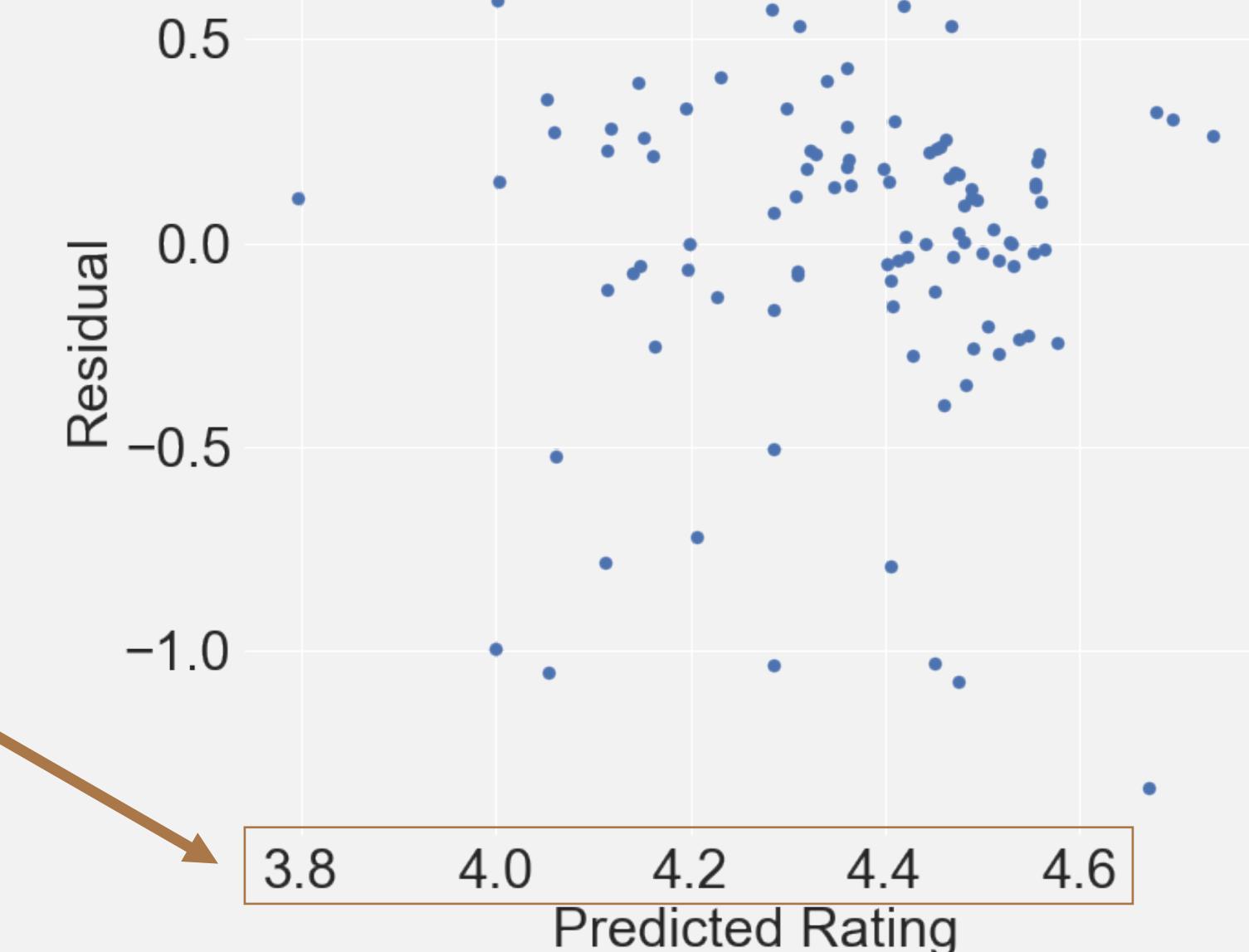
Random Forest Regression $R^2 = 0.146$

Feature	Importance
# of Photos	0.184
Prep. Time	0.153
Oven Temp.	0.055
Baking Powder	0.047
Calories	0.046

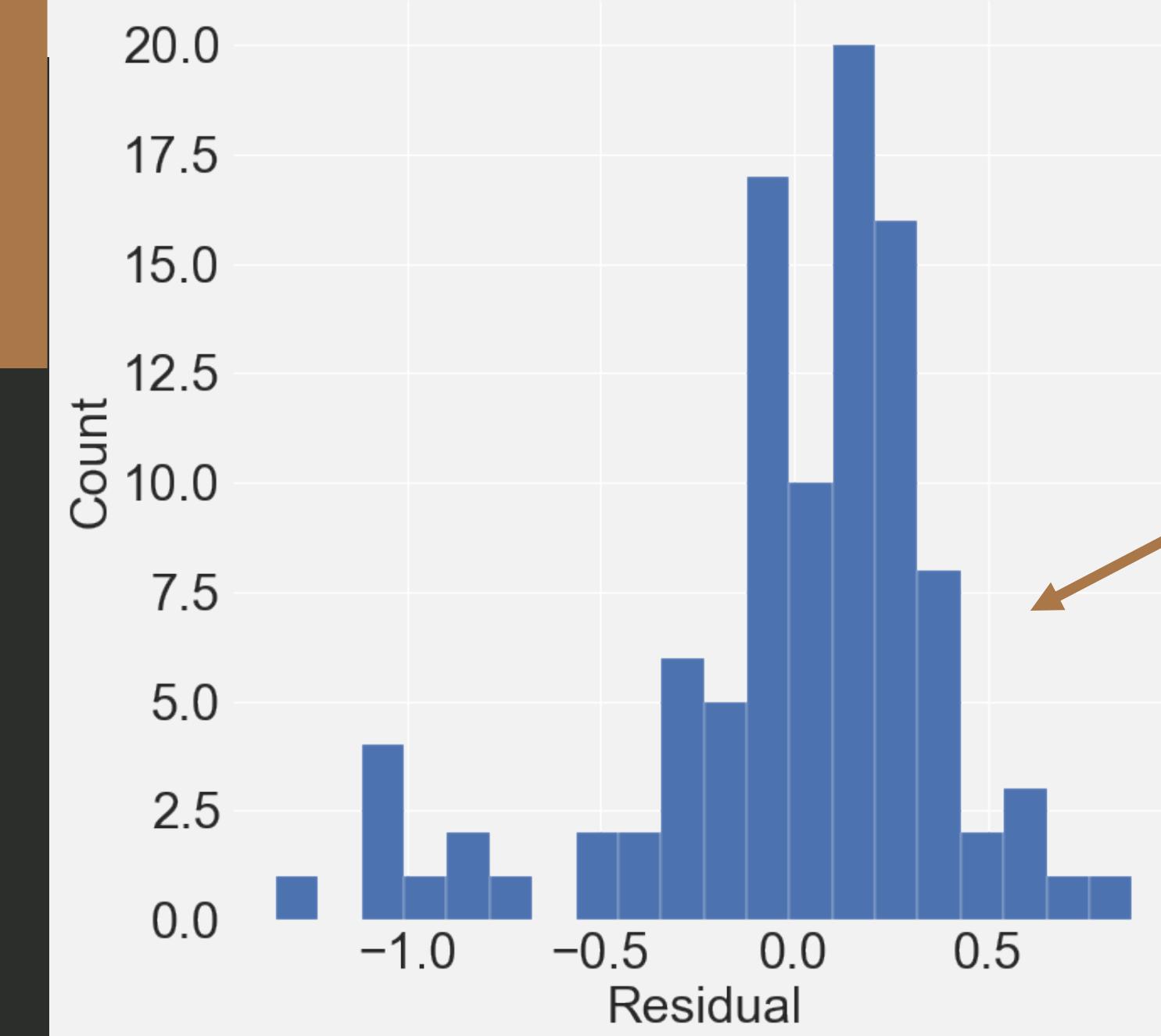


Predictions
within range

Residuals of Random Forest Regression



Distribution of Random Forest Residuals



Residuals more
normally distributed

TAKEAWAYS

LESSON || FUTURE DIRECTIONS

Linear Regression
is **not a good**
model for
predicting ratings
on **chocolate chip**
cookie recipes

- 1-5 scale is not continuous
- Should not be calculated using simple average
- Subjective in nature (user error)
- Classification or other models more appropriate
- Sentiment analysis/NLP



QUESTIONS?

APPENDIX

OLS:

Training R^2 = 0.153

Val R^2 = 0.0218

Test R^2 = 0.025

MAE = 0.298

MSE = 0.18

RMSE = 0.42

LASSO:

Training R^2 = 0.0833

Val R^2 = 0.0637

Test R^2 = 0.0613

MAE = 0.294

MSE = 0.17

RMSE = 0.41

Ridge:

Training R^2 = 0.116

Val R^2 = 0.0224

Test R^2 = 0.058

MAE = 0.294

MSE = 0.17

RMSE = 0.41

Random Forest:

Training R^2 = 0.574

Val R^2 = 0.159

Test R^2 = 0.147

MAE = 0.28

MSE = 0.16

RMSE = 0.394

APPENDIX

LASSO:

- alpha = 0.02479
- Features selected:
 - Baking powder -0.0495
 - Photos -0.0207
 - Reviews -0.016
 - Salt -0.01
 - Fats/Flours ratio 0.01
 - Egg -0.01
 - Made it -0.01
 - Flours -0.01
 - Prep time 0.01
 - Oven Temp 0.009

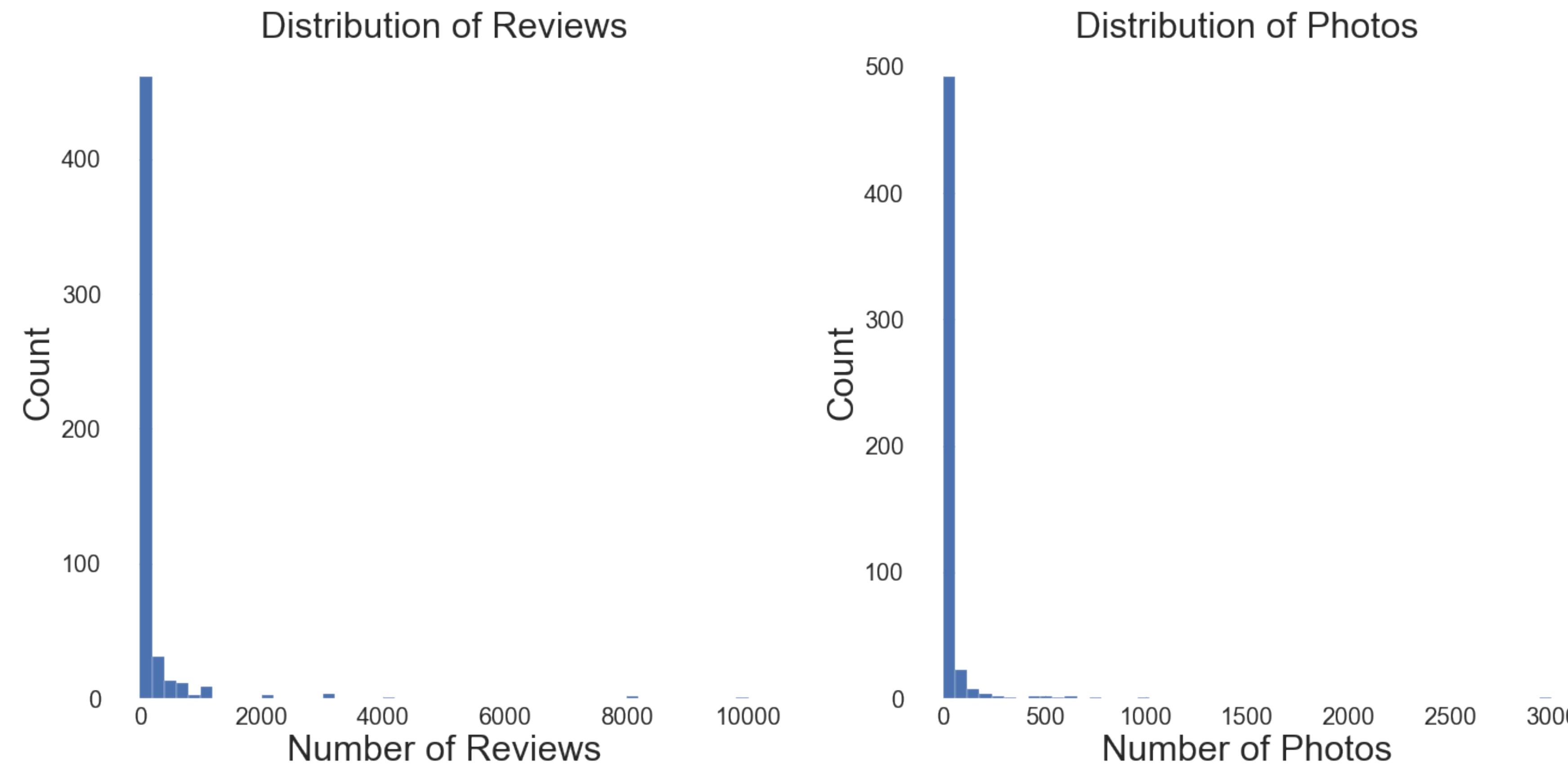
Ridge:

- alpha = 255.50971
- Features selected:
 - Baking powder -0.035
 - Flours -0.0279
 - Salt -0.022
 - Egg -0.02
 - Prep time 0.0197
 - Servings 0.0178
 - Photos 0.0177
 - Made it 0.0166
 - Vanilla extract 0.0163
 - Oven temp 0.015

Random Forest:

- n_estimators = 30
- min_sample_split = 20
- Features selected:
 - Photos 0.184
 - Prep time 0.129
 - Ratings 0.082
 - Reviews 0.075
 - Baking powder 0.047
 - Calories 0.046
 - Oven temp 0.044
 - Salt 0.036
 - Flours 0.027
 - Sugars/Flowers ratio 0.021

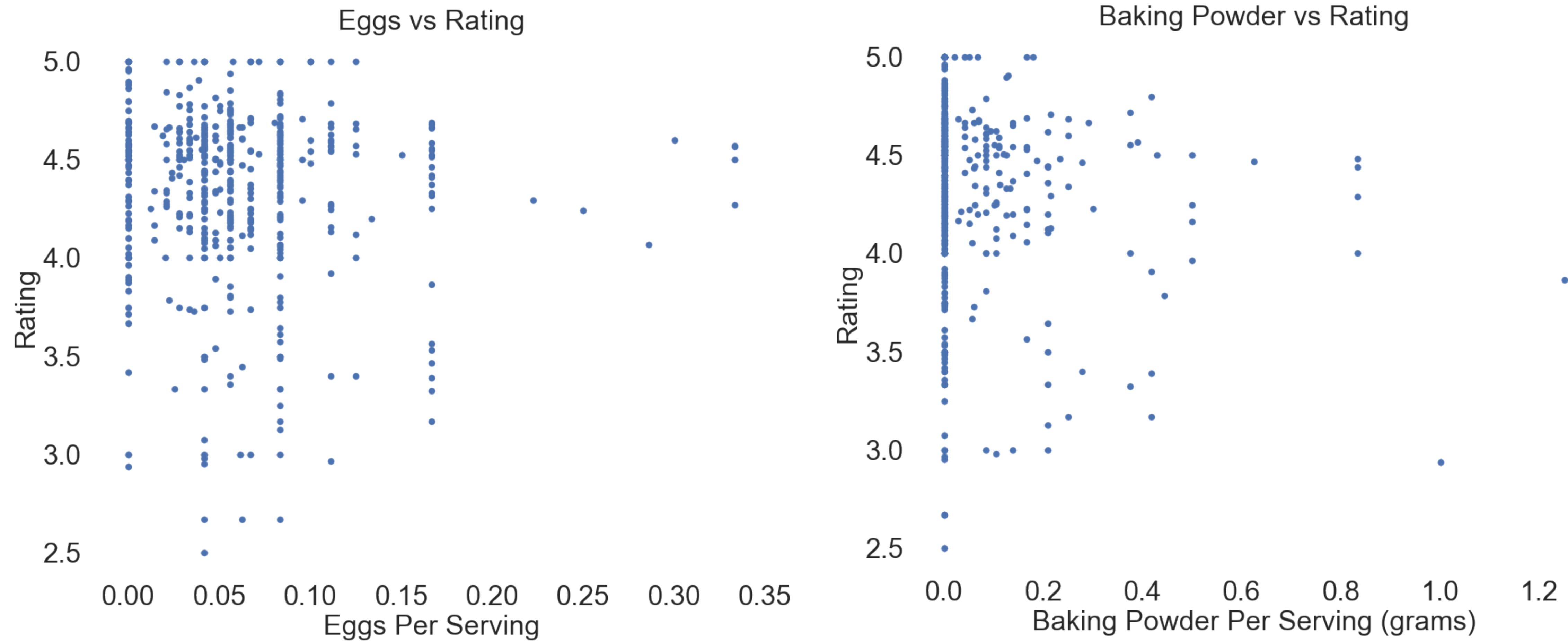
APPENDIX



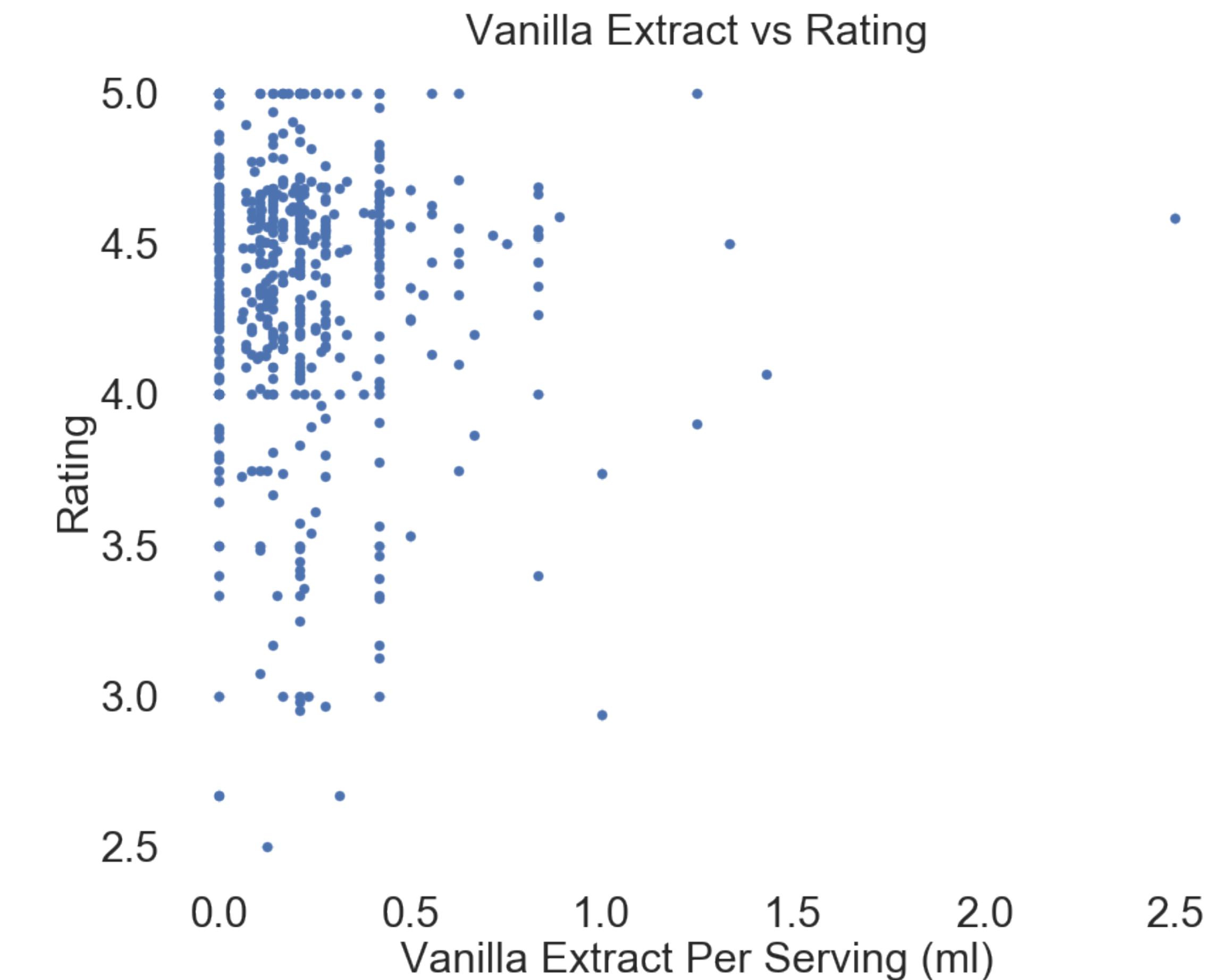
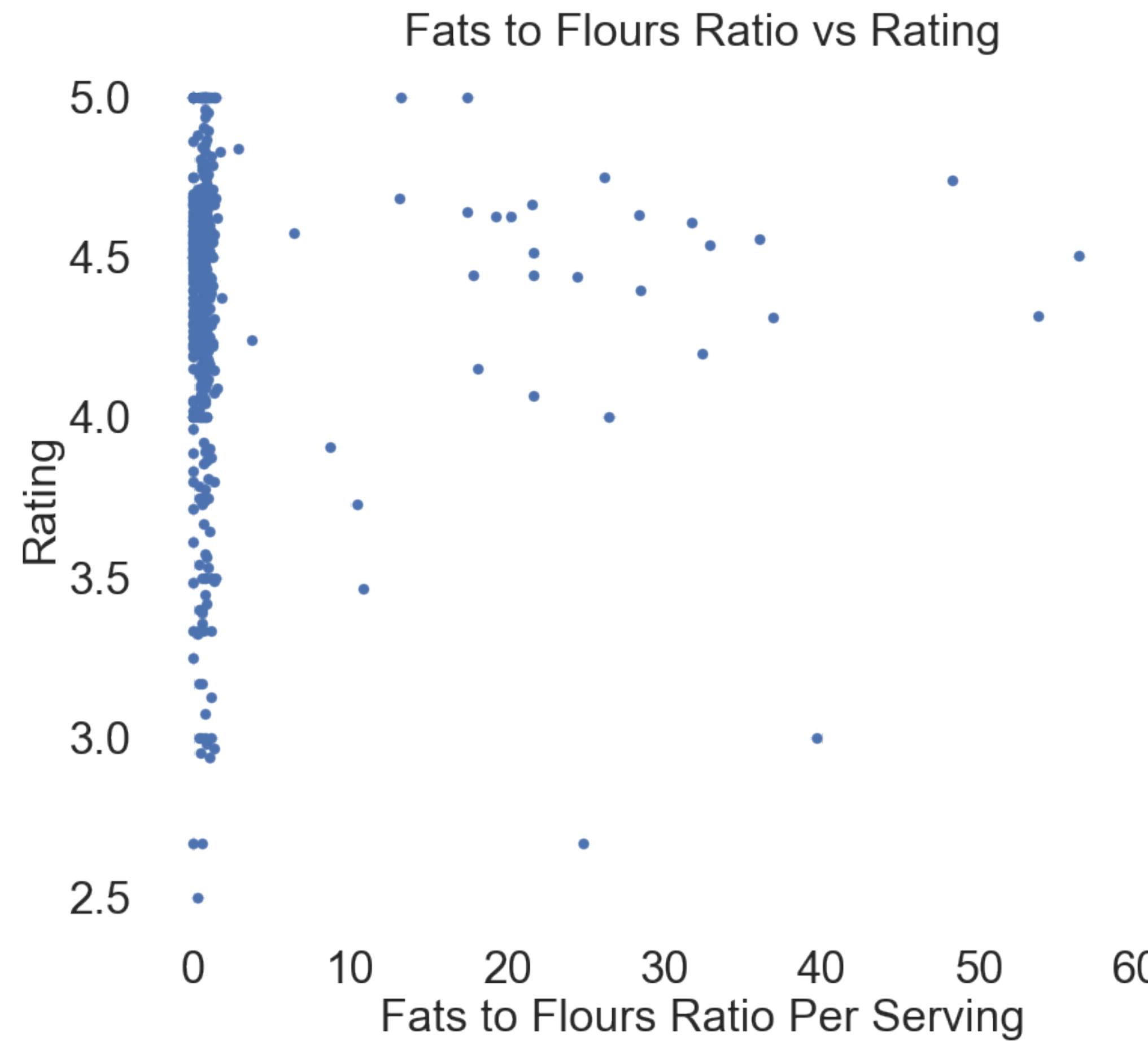
APPENDIX

				coef	std err	t	P> t	[0.025	0.975]	
Dep. Variable:	rating	R-squared:	0.153	const	3.9865	0.157	25.458	0.000	3.678	4.295
Model:	OLS	Adj. R-squared:	0.068	num_ratings	-0.0004	0.001	-0.493	0.623	-0.002	0.001
Method:	Least Squares	F-statistic:	1.792	num_reviews	0.0006	0.001	0.594	0.553	-0.001	0.003
Date:	Thu, 18 Apr 2019	Prob (F- statistic):	0.0101	made_it	6.305e-05	0.000	0.207	0.836	-0.001	0.001
Time:	20:15:59	Log-Likelihood:	-157.25	servings	0.0023	0.002	1.030	0.304	-0.002	0.007
No. Observations:	306	AIC:	372.5	calories	0.0007	0.000	1.610	0.108	-0.000	0.002
Df Residuals:	277	BIC:	480.5	num_photos	0.0020	0.003	0.781	0.436	-0.003	0.007
Df Model:	28			oven_temp	0.0004	0.000	1.198	0.232	-0.000	0.001
Covariance Type:	nonrobust			prep_time_minute s	0.0007	0.000	1.574	0.117	-0.000	0.002
Omnibus:	60.507	Durbin- Watson:	1.884	egg_count	-1.1581	0.661	-1.753	0.081	-2.459	0.142
Prob(Om nibus):	0.000	Jarque- Bera (JB):	108.754	vanilla_extract_ml	0.2500	0.136	1.841	0.067	-0.017	0.517
Skew:	-1.078	Prob(JB):	2.42e-24	salt_g	-0.1156	0.095	-1.219	0.224	-0.302	0.071
Kurtosis:	4.970	Cond. No.	1.83e+04	cinnamon_g	-0.1212	0.779	-0.156	0.876	-1.654	1.412
				baking_soda_g	0.3488	0.364	0.958	0.339	-0.368	1.065
				baking_powder_g	-0.3481	0.211	-1.647	0.101	-0.764	0.068
				water_ml	-0.0430	0.052	-0.830	0.407	-0.145	0.059
				cocoa_powder_g	0.0228	0.032	0.703	0.483	-0.041	0.087
				fats	-0.0004	0.001	-0.436	0.663	-0.002	0.001
				sugars	-0.0048	0.007	-0.710	0.478	-0.018	0.008
				flours	-0.0167	0.007	-2.243	0.026	-0.031	-0.002
				nuts	-0.0060	0.014	-0.444	0.658	-0.033	0.021
				chocolate_chips	-0.0043	0.008	-0.519	0.604	-0.021	0.012
				other_chocolate_c hips	0.0096	0.009	1.122	0.263	-0.007	0.027
				num_ingredients	0.0086	0.012	0.721	0.472	-0.015	0.032
				fats_flours_ratio	0.0050	0.007	0.738	0.461	-0.008	0.019
				fats_sugars_ratio	0.0010	0.003	0.301	0.764	-0.006	0.008
				sugars_flours_ratio	0.0353	0.047	0.752	0.453	-0.057	0.128
				chips_nuts_ratio	0.0009	0.011	0.081	0.935	-0.021	0.023
				white_brown_ratio	-0.0624	0.051	-1.215	0.226	-0.164	0.03

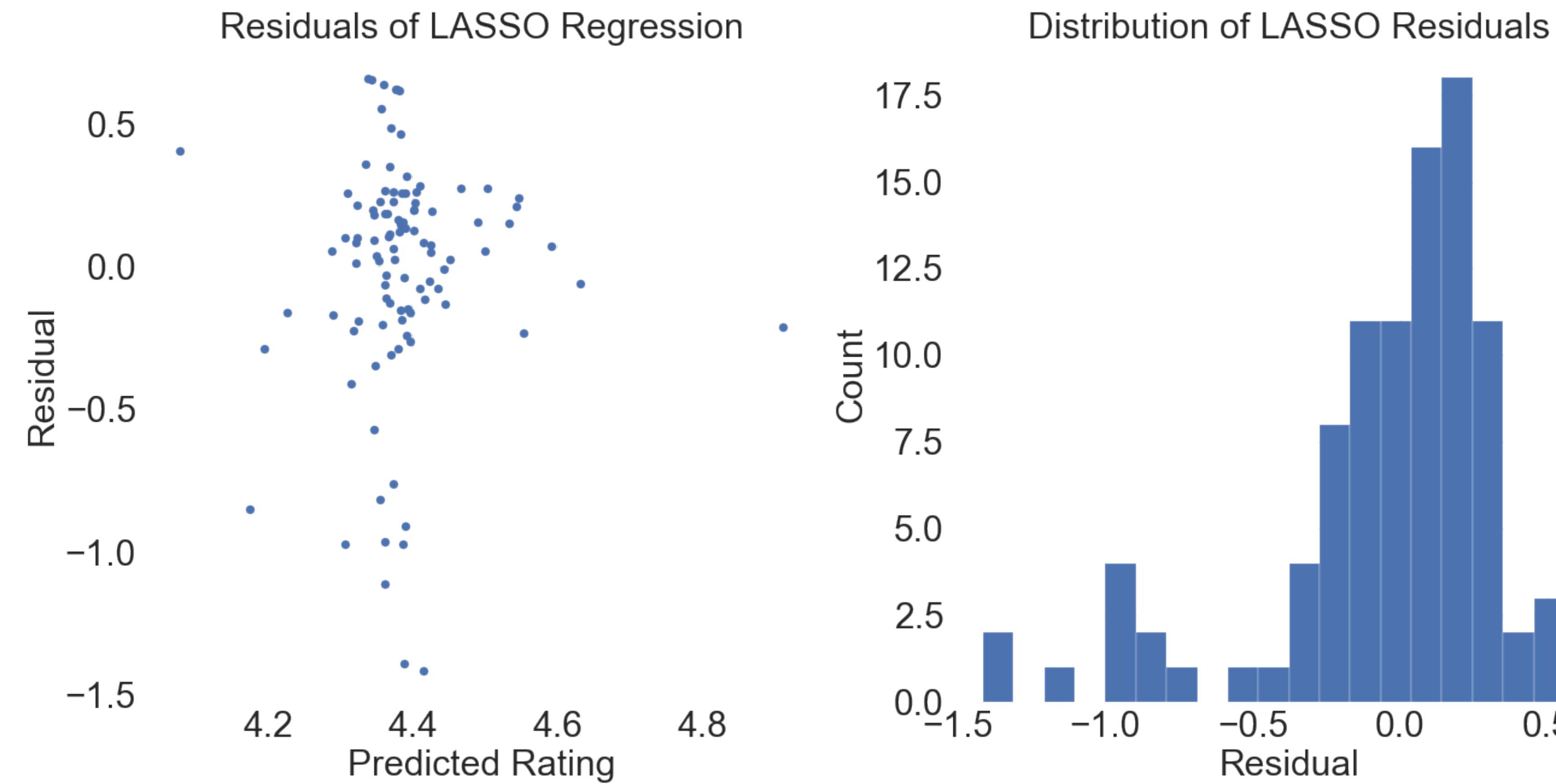
APPENDIX



APPENDIX



APPENDIX



APPENDIX

