Proposal

John-Tianyu-Fabian

2022-04-05

Modeling

As mentioned from the previous section of exploratory data analysis, we found a positive relationship between the altitude measure and the total cup score. Therefore, altitude are being chosen as the key variable in our model.

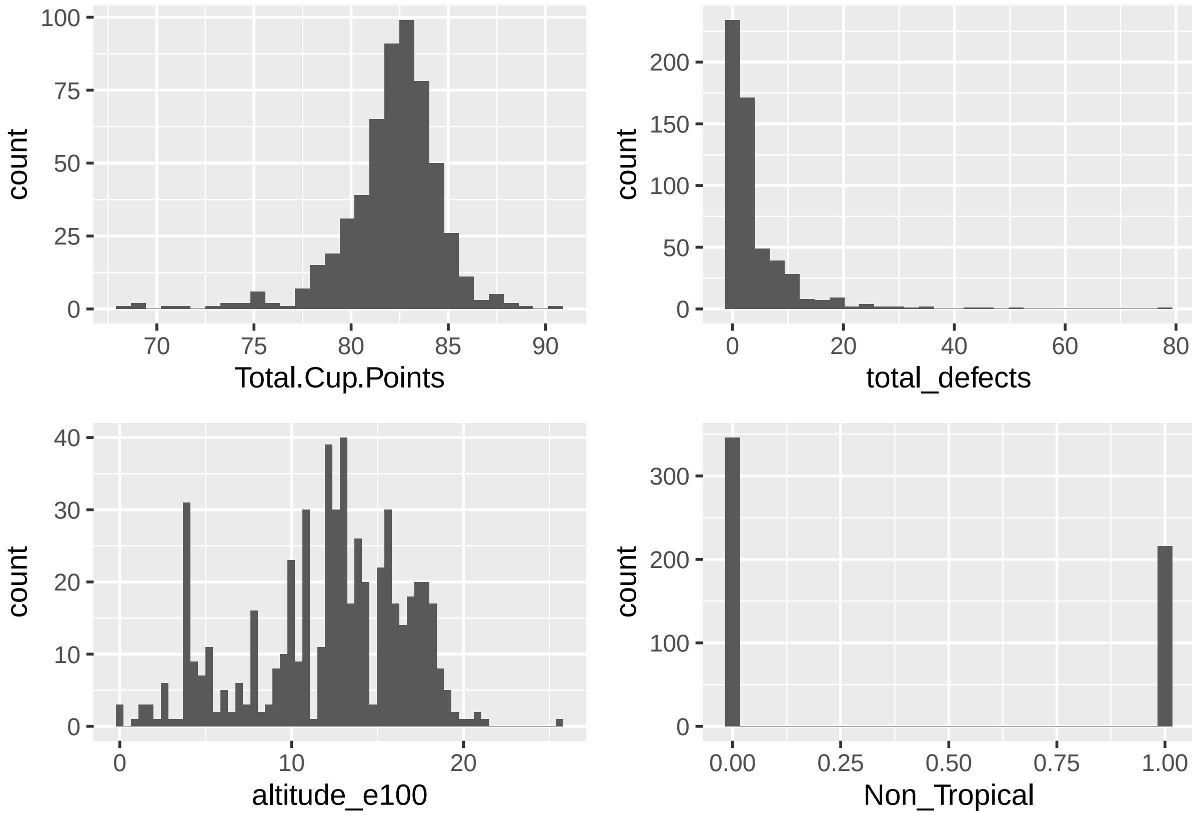
On top of the altitude, the climate and environment which coffee beans are growing in is considered as an additional factor which can impact the quality. Since coffee plants are generated well grown between the tropic of Cancer and the tropic of Capricorn, usually termed the bean belt or the coffee belt[1], we will mainly investigate the climate difference between tropical and non-tropical region. In this study, the latitude information of each country is integrated from an extra data set which allow us to define the region of each country.

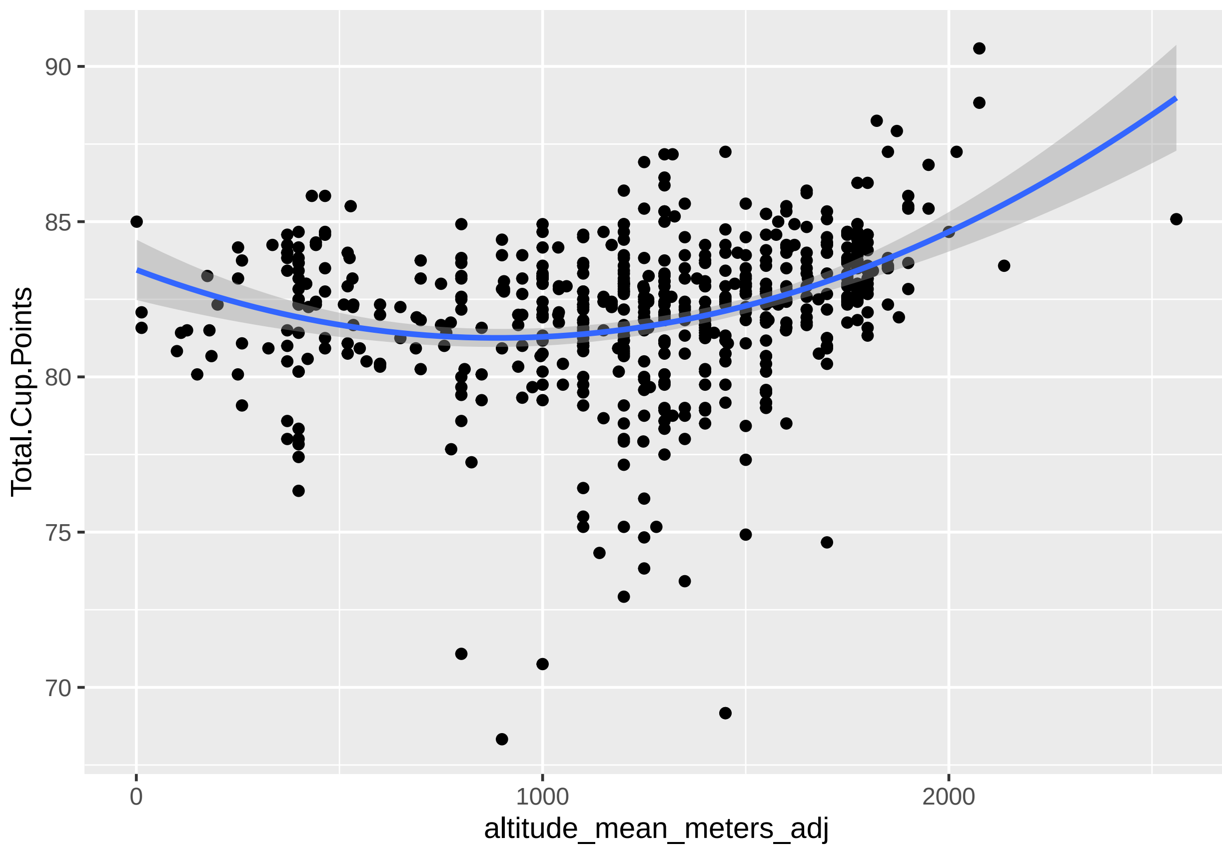
Beyond the two geographical factors mentioned above, we are considering to include the properties and processes that are associated with the beans as control variables in our model:

- Processing Method
- Variety
- Total Defects

While conducting the analysis and modeling, we found that the Total score increment described by the effect size will be too small if the variable unit increases by every meter. Thus, `altitude_e100` column is generated from the original altitude information devided by 100.

[1]Consonni, Roberto, Laura Ruth Cagliani, and Clelia Cogliati. "NMR Based Geographical Characterization of Roasted Coffee." Talanta (Oxford) 88 (2012): 420–426. Web.


count
Total.Cup.Points
total_defects
altitude_e100
Non_Tropical



Base Model

In the base model, only the key explanatory variable `altitude_e100` is included:

$$Total.Cup.Points = \beta_0 + \beta_1 \cdot altitude\_e100 + \beta_2 \cdot altitude\_e100^2$$

We can determine that altitude variable has a statistically significant impact to the total cup points. In the base model, the coefficient of determination(R^2) is 0.106, indicating that 10.6% of the variance in the total cup point is being described by the base model. The effect size of altitude per 100 meter is -0.493, indicating that a 100-meter increase of the altitude decrease the total cup score by 0.493 units at 0 meter altitude. The effect size of altitude squared per 100 meter is 0.028, indicating that a 100-meter increase of the altitude will slow down the total cup score decreament by 0.028 units.

Second Model

As mentioned above, we include non_tropical variable in the second model to explore another geographical elements in the total cup point evaluation.

The second model assessed is thus as follows:

$$Total.Cup.Points = \beta_0 + \beta_1 \cdot altitude\_e100 + \beta_2 \cdot altitude\_e100^2 + \beta_3 \cdot Non\_Tropical$$

Third Model

In the third model, additional control variables are accounted for aspects of the coffee bean's properties and processing methods.

The third model assessed is thus as follows:

$$Total.Cup.Points = \beta_0 + \beta_1 \cdot altitude\_e100 + \beta_2 \cdot altitude\_e100^2$$
$$+\beta_3 \cdot Non\_Tropical + \beta_4 \cdot variety + \beta_5 \cdot Processing.Method + \beta_6 \cdot Total.Defect$$

Results

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Apr 11, 2022 - 02:03:05 AM

Table 1:

	Dependent variable:		
	Total.Cup.Points		
	(1)	(2)	(3)
altitude_e100	-0.493^{***}	-0.325^{***}	-0.281^{***}
	(0.093)	(0.096)	(0.095)
I(altitude_e100ˆ2)	0.028^{***}	0.019^{***}	0.018^{***}
	(0.004)	(0.004)	(0.004)
Non_Tropical		-1.224^{***}	-0.617^{**}
		(0.219)	(0.289)
total_defects			-0.096^{***}
			(0.014)
variety_adj_Bourbon			-0.081
			(0.310)
variety_adj_Catuai			-0.981^{**}
			(0.447)
variety_adj_Caturra			-0.200
			(0.273)
variety_adj_Typica			-0.597^{*}
			(0.330)
variety_adj_Yellow.Bourbon			-1.065
			(0.769)
Processing.Method_Washed...Wet			-0.931^{***}
			(0.283)
Processing.Method_Semi.washed...Semi.pulped			0.190
			(0.474)
Processing.Method_Pulped.natural...honey			0.483
			(1.269)
Processing.Method_Other			-0.399
			(0.658)
Constant	83.451^{***}	83.285^{***}	84.193^{***}
	(0.495)	(0.483)	(0.556)
Observations	562	562	562
R^2	0.109	0.157	0.254
Adjusted R^2	0.106	0.152	0.236
Akaike Inf. Crit.	2,590.734	2,562.044	2,513.205
Residual Std. Error	2.415 (df = 559)	2.352 (df = 558)	2.232 (df = 548)
F Statistic	34.286^{***} (df = 2; 559)	34.537^{***} (df = 3; 558)	14.340^{***} (df = 13; 548)
Note:			*p<0.1; **p<0.05; ***p<0.01

The base model indicates a effect size of -0.493 with a coefficient of determination of 0.106. The effect size of altitude per 100 meter indicates that a 100-meter increase of the altitude decrease the total cup score by

0.493 units at 0 meter altitude. The effect size of altitude squared per 100 meter is 0.028, indicating that a 100-meter increase of the altitude will slow down the total cup score dropping by 0.028 units.

The second model includes additional explanatory variables `Non_tropical`, which are expected to explain meaningful aspects of the data. With the addition of the tropical information, the second model indicates a effect size of -0.325 with a coefficient of determination of 0.152. It indicates that if the growing region of one coffee bean being moved out from tropical region, the total cup score will have a 1.224 unit drop. This shows that the tropical information contributes a big part to the second model and dilutes the effect size of altitude variable.

The third model, which includes the explanatory variables from the second model while also adding in control variables, shows an increase in explanatory power (0.236 vs. 0.152). The addition of these control variables results in a reduction of effect size on an absolute value basis across all of the explanatory variables; however, the altitude variable effect size remains relatively stable from a practical perspective (-0.281 vs. 0.325).

The final model selected is thus as follows:

$$
\begin{aligned}
Total.Cup.Points = \beta_0 + \beta_1 \cdot altitude\_e100 + \beta_2 \cdot altitude\_e100^2 \\
+\beta_3 \cdot Non\_Tropical + \beta_4 \cdot variety + \beta_5 \cdot Processing.Method + \beta_6 \cdot Total.Defect
\end{aligned}
$$

The total cup score of coffee beans is impacted by it's growing altitude. At the beginning of 0 meter altitude, an increase in the altitude by 100 meter decreases the total cup score by 0.281 units, and is highly statistically significant. However, the decreasing trend get slower with a rate of 0.018 for every 100 meters of altitude increment. The score will hit the bottom at about 1561 meter altitude, after which, the score will go up again. As a guidence for coffee investors, it seems better to grow coffee either on the lower altitude area close to 0 meter or on the higher altitude area close to 3000 meter.