



From Bean to Cup: A Data Science

Perspective on Coffee Bean Quality Analysis and Optimization

Different Variables

Continuous

Quality Measures

Aroma

Flavor

Aftertaste

Acidity

Body

Balance

Uniformity

Cup Cleanliness

Sweetness

Moisture

Defects

Categorical

Bean Metadata

Processing Method

Color

Species (arabica / robusta)

Continuous

Farm Metadata

Owner

Country of Origin

Farm Name

Lot Number

Mill

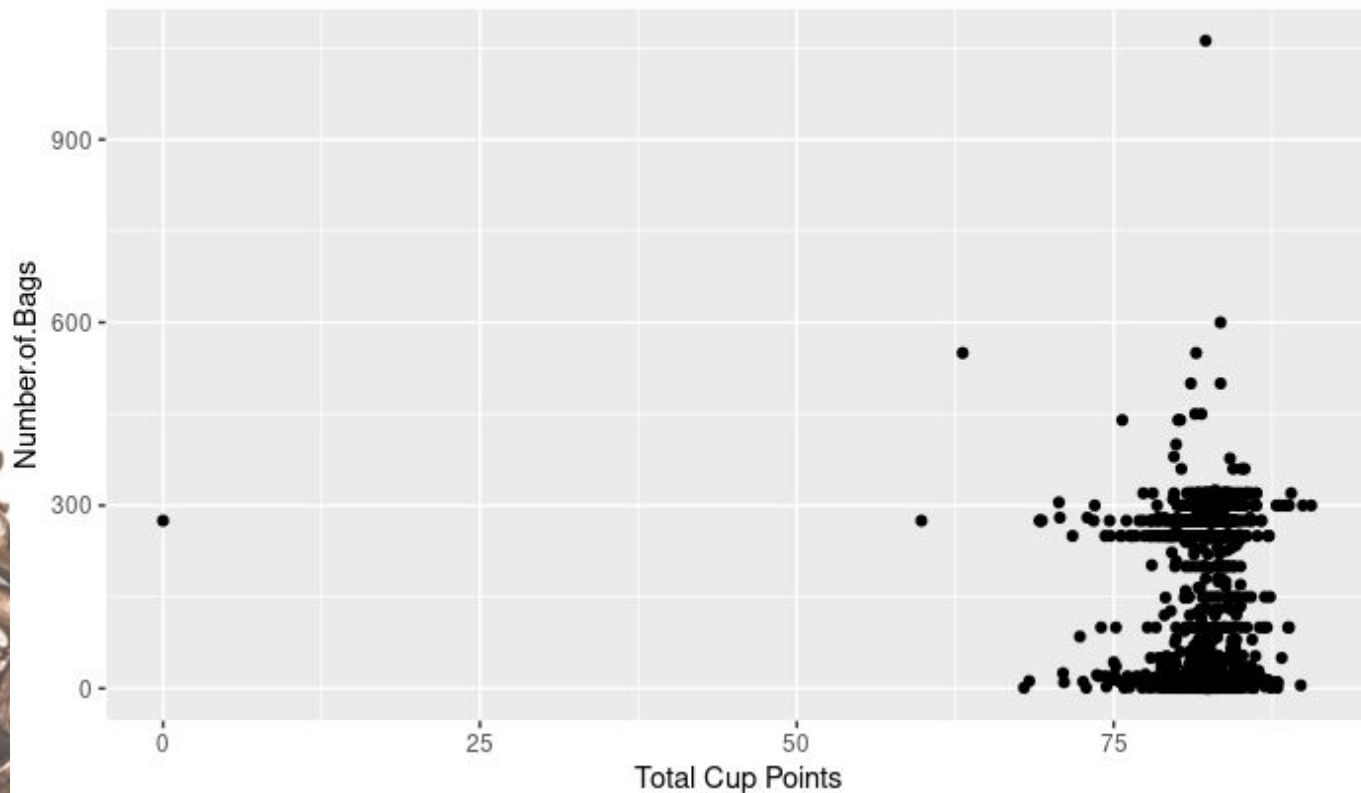
Company

Altitude

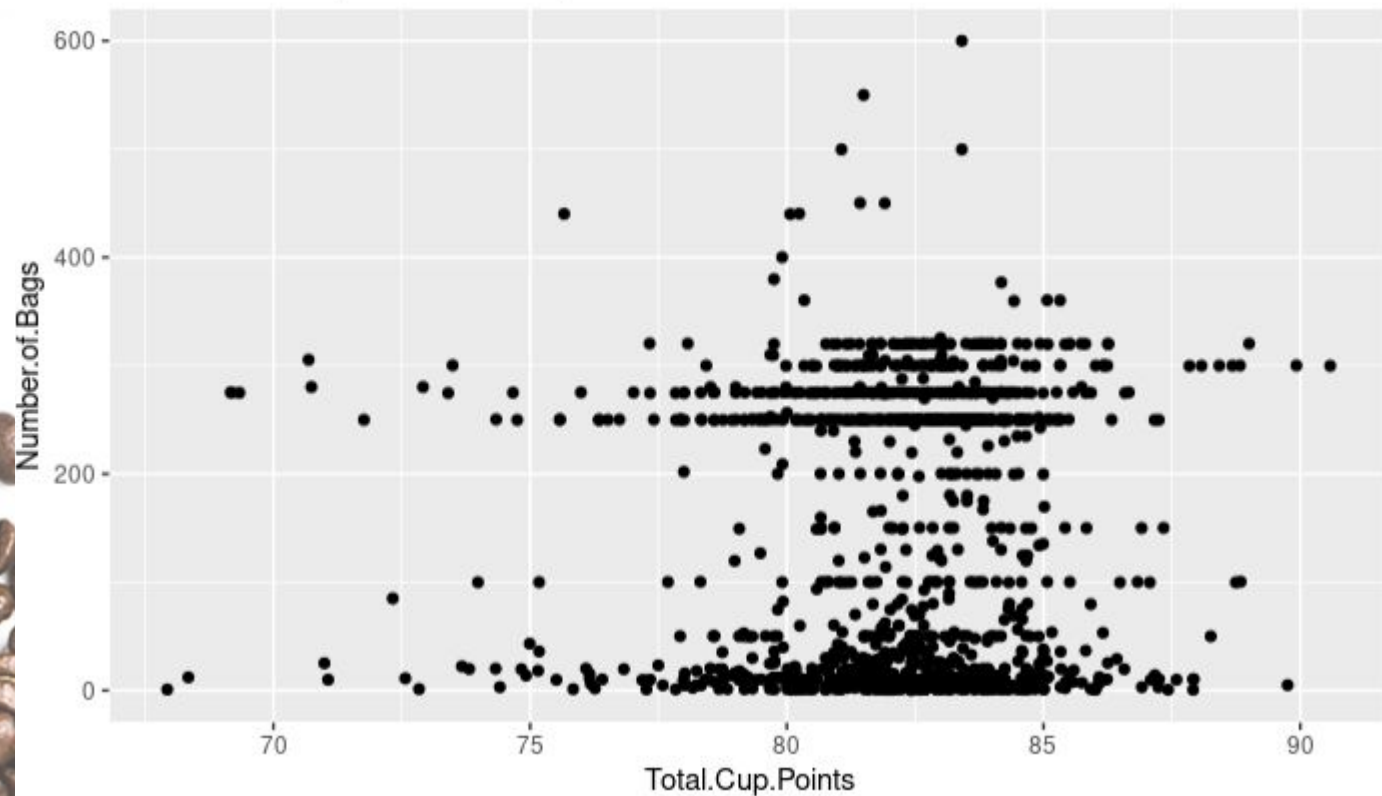
Region



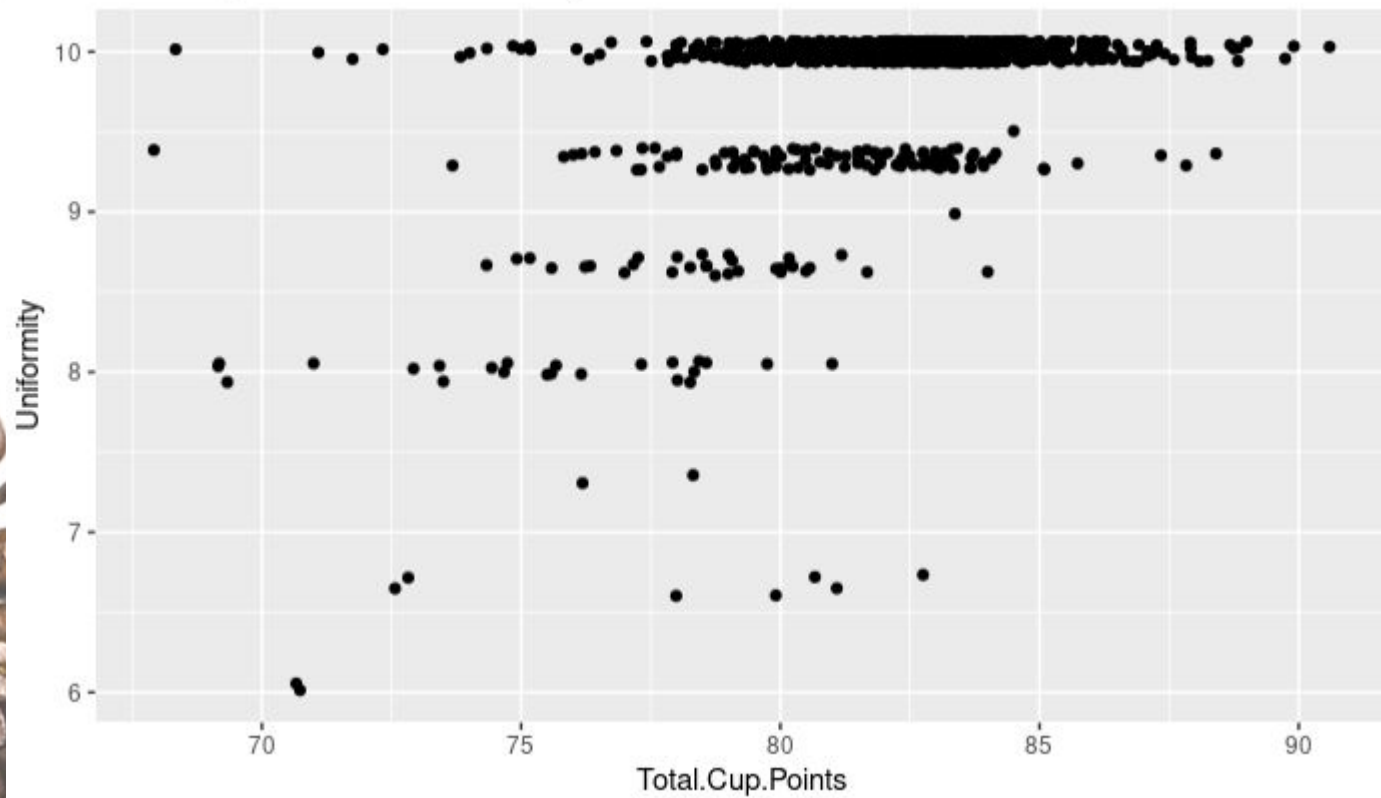
Number of Bags vs Total Cup Points



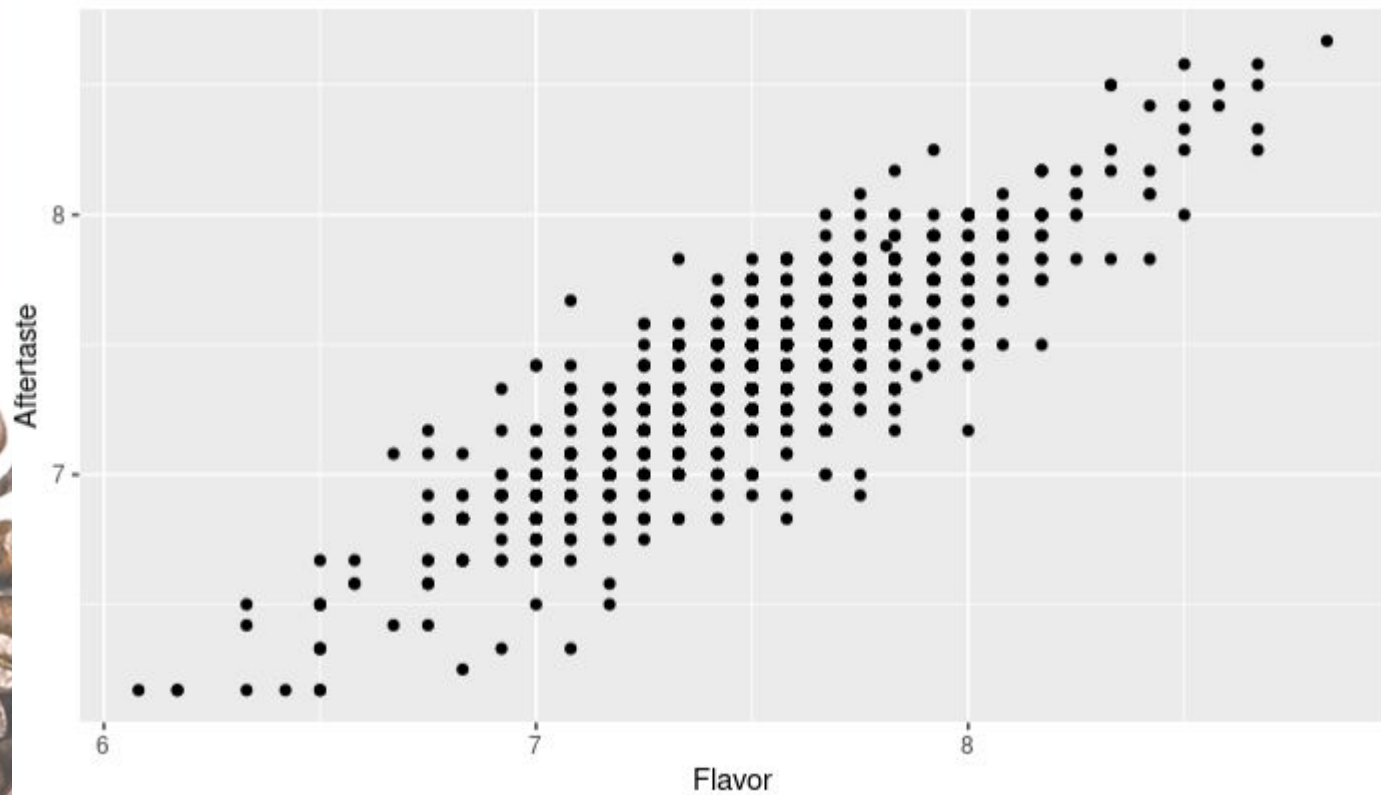
Number of Bags vs Total Cup Points



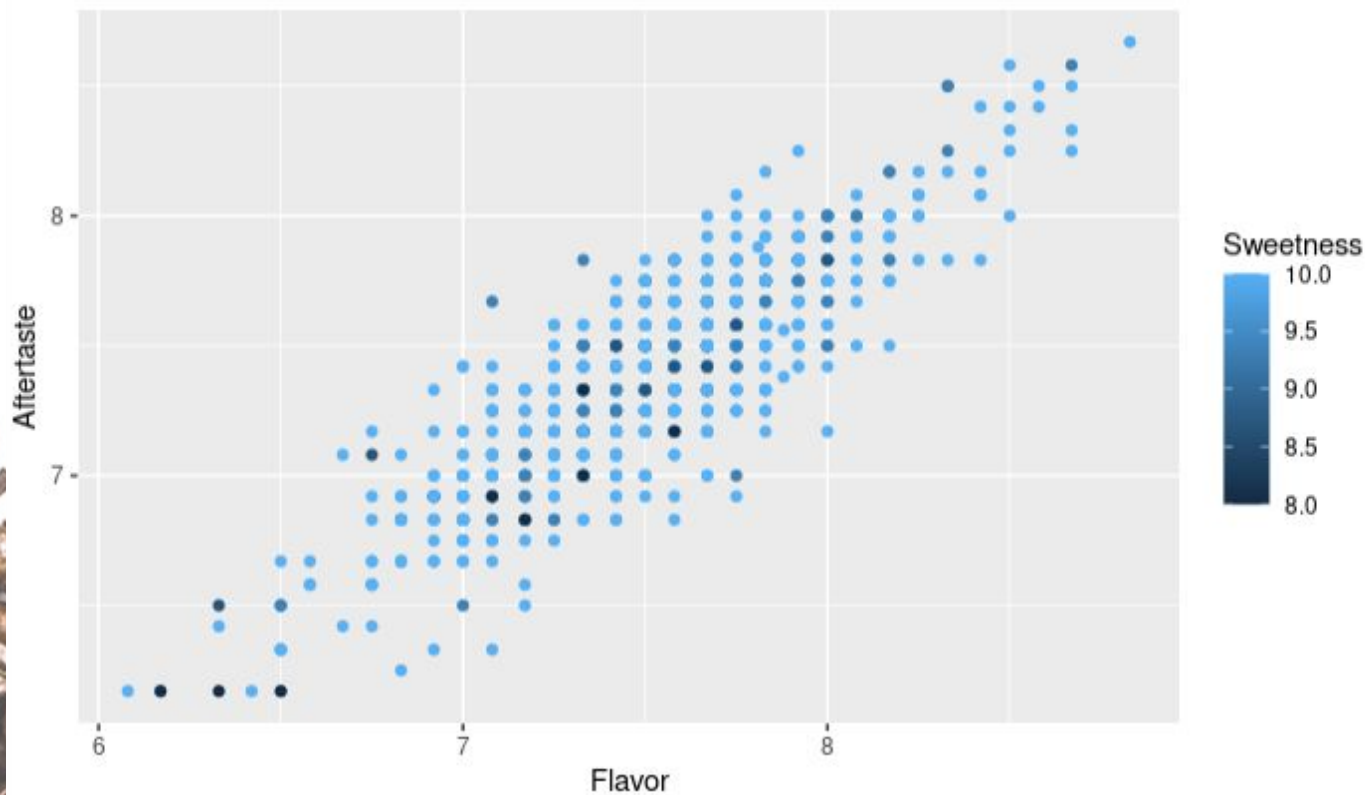
Total Cup Points vs. Uniformity



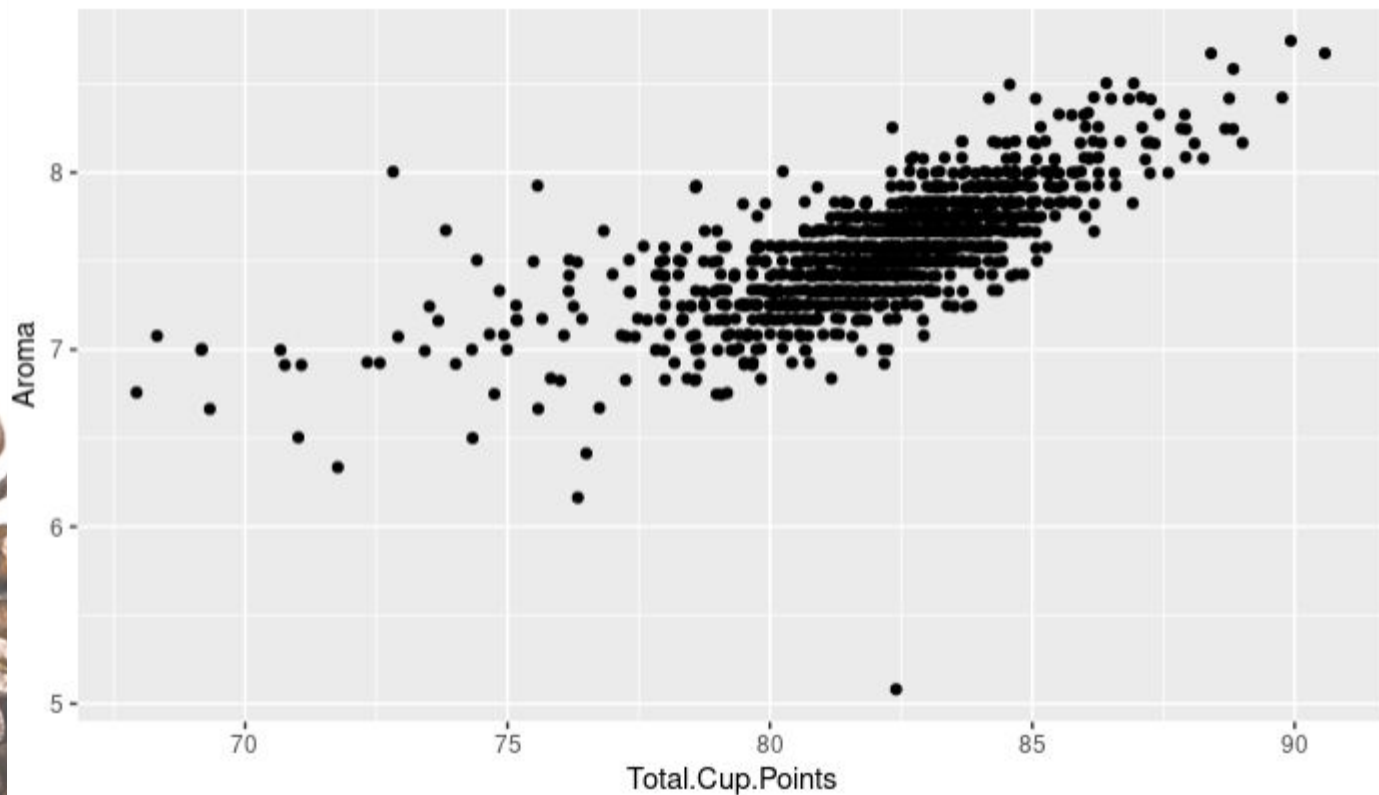
Flavor vs. Aftertaste



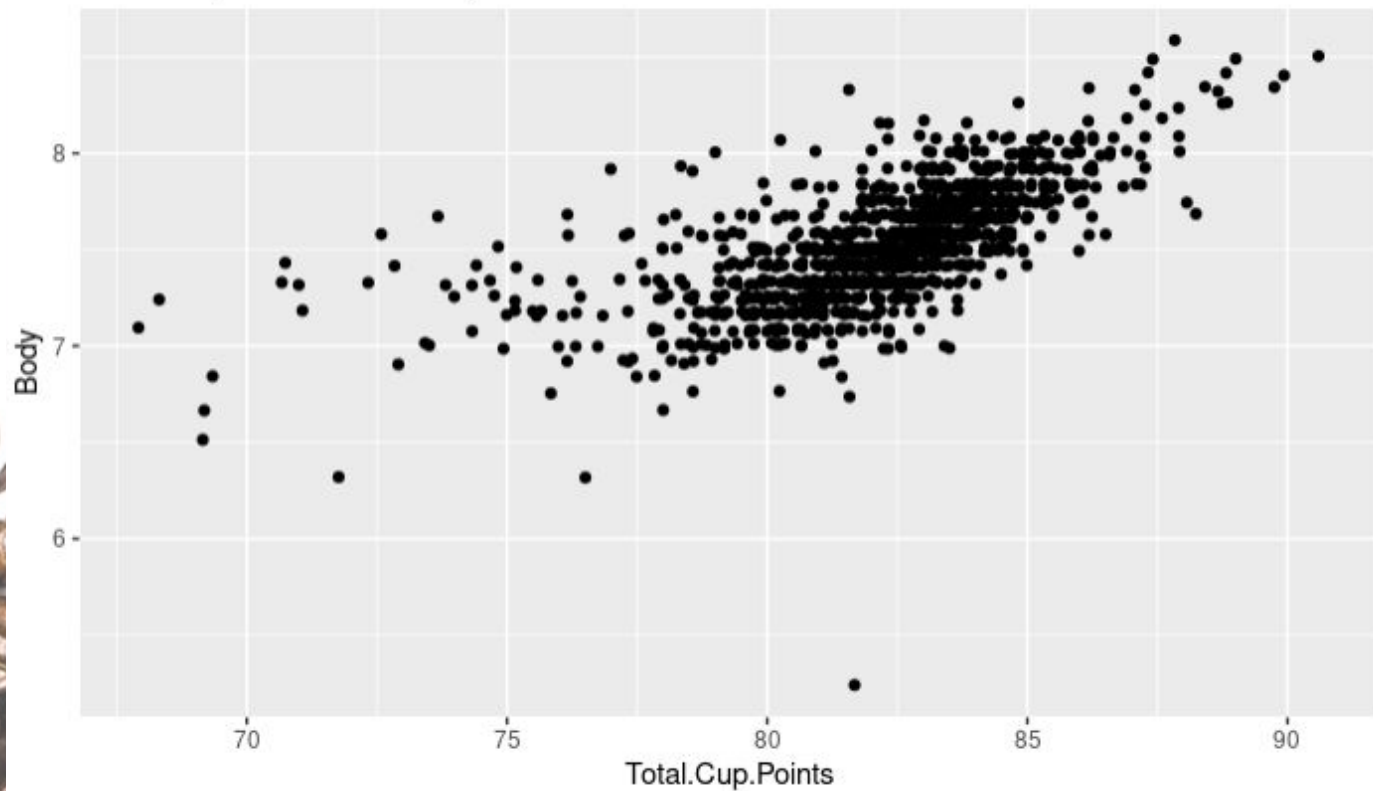
Flavor vs. Aftertaste with Sweetness



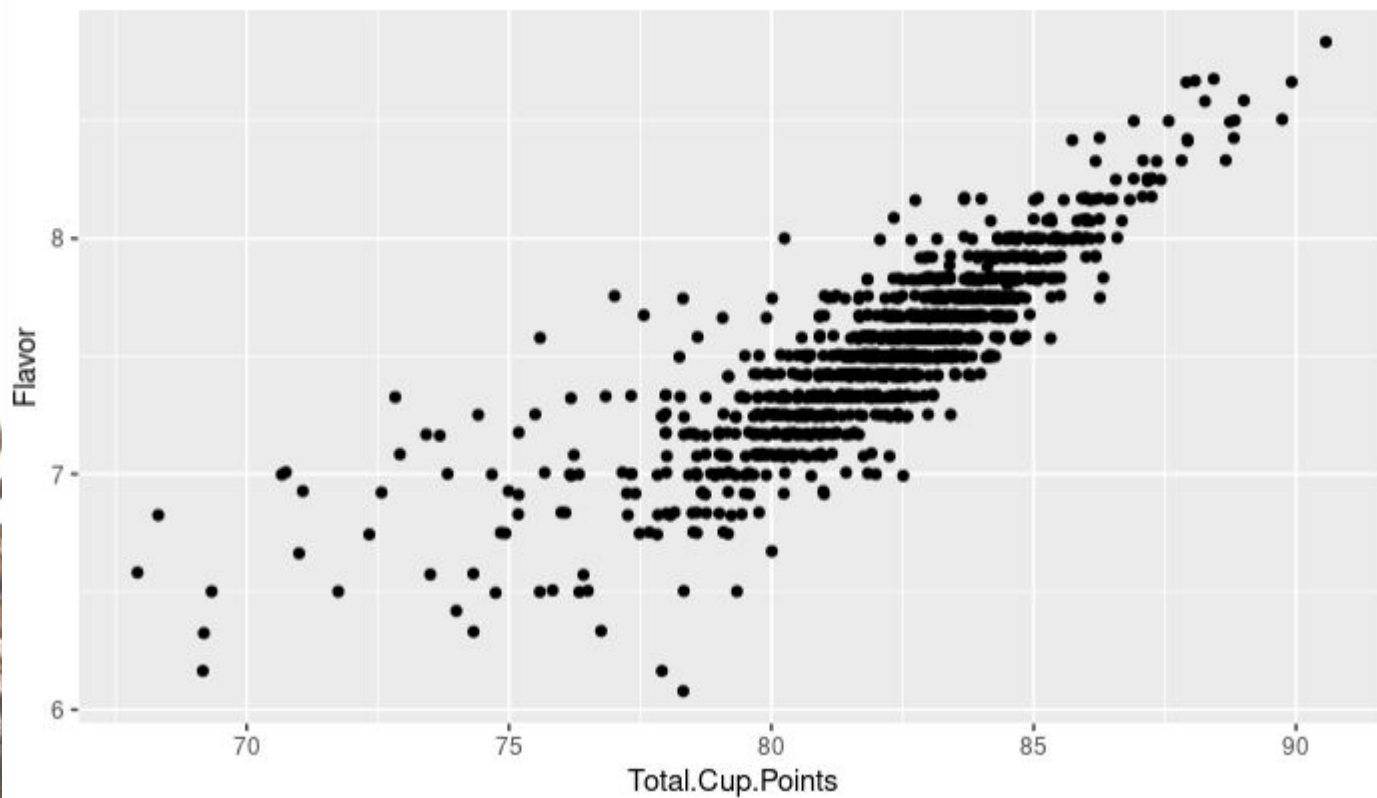
Total.Cup.Points vs. Aroma



Total.Cup.Points vs. Body



Total.Cup.Points vs. Flavor



Multiple Regression

	Aroma	Body	Flavor	Total.Cup.Points
Aroma	1.0000000	0.6956131	0.8135973	0.7966370
Body	0.6956131	1.0000000	0.7612108	0.7763467
Flavor	0.8135973	0.7612108	1.0000000	0.8776274
Total.Cup.Points	0.7966370	0.7763467	0.8776274	1.0000000

Call:
lm(formula = Total.Cup.Points ~ Aroma + Body + Flavor, data = coffee)

Standardized Coefficients::

(Intercept)	Aroma	Body	Flavor
NA	0.1941813	0.2222581	0.5504568

Call:
lm(formula = Total.Cup.Points ~ Aroma + Body + Flavor, data = coffee)

Residuals:

Min	1Q	Median	3Q	Max
-17.6591	-0.3437	0.2386	0.7233	4.3959

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	15.7495	0.9444	16.676	<2e-16 ***
Aroma	1.8034	0.1968	9.164	<2e-16 ***
Body	2.1740	0.1858	11.700	<2e-16 ***
Flavor	4.8393	0.2064	23.451	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.532 on 1303 degrees of freedom
(12 observations deleted due to missingness)

Multiple R-squared: 0.8103, Adjusted R-squared: 0.8099



F-statistic: 1856 on 3 and 1303 DF, p-value: < 2.2e-16

Total.Cup.Points = 15.75 + (1.80 * Aroma) + (2.17 * Body) + (4.84 * Flavor)

```
## {r}
Total.Cup.Points <- 15.75 + (1.80 * 9.2) + (2.17 * 8.7) + (4.84 * 9.9)
Total.Cup.Points
## [1] 99.105
```

[1] 99.105

unique(coffee\$Country.of.Origin)

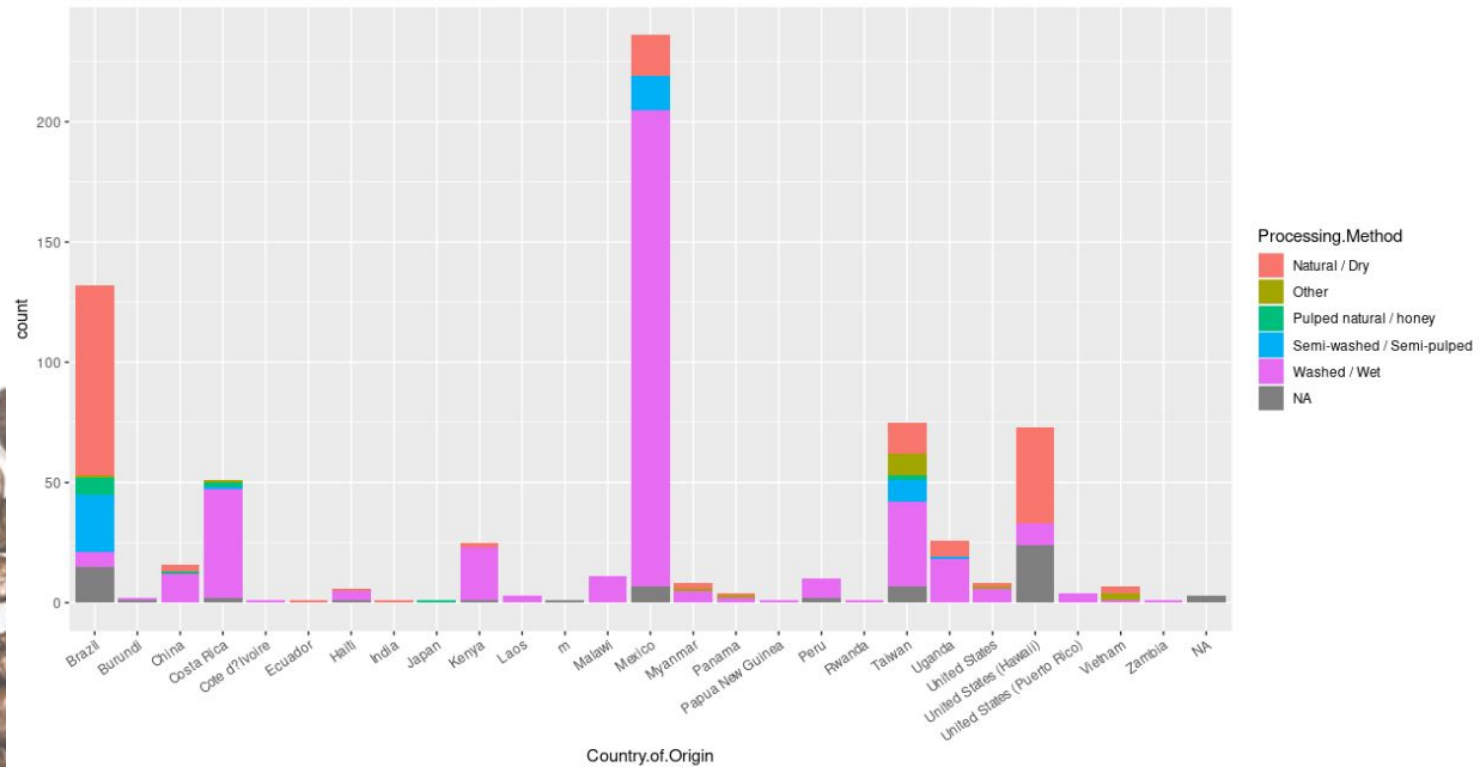


```
[1] "Ethiopia"  
[3] "Brazil"  
[5] "United States"  
[7] "Indonesia"  
[9] "Costa Rica"  
[11] "Uganda"  
[13] "Taiwan"  
[15] "Tanzania, United Republic Of"  
[17] "Thailand"  
[19] "Panama"  
[21] "El Salvador"  
[23] "Ecuador"  
[25] "Haiti"  
[27] "Vietnam"  
[29] "Rwanda"  
[31] "Laos"  
[33] "Myanmar"  
[35] "Natural / Dry"  
[37] "Mauritius"  
[39] "7.25"  
[41] "Cote d'Ivoire"  
[43] "oriente"  
"Guatemala"  
"Peru"  
"United States (Hawaii)"  
"China"  
"Mexico"  
"Honduras"  
"Nicaragua"  
"Kenya"  
"Colombia"  
"Papua New Guinea"  
"Japan"  
"United States (Puerto Rico)"  
"Burundi"  
"Philippines"  
"Malawi"  
"Zambia"  
"7.42"  
"0878a7d4b9d35ddbfe0fe2ce69a2062cceb45a660"  
NA  
"m"  
"India"
```

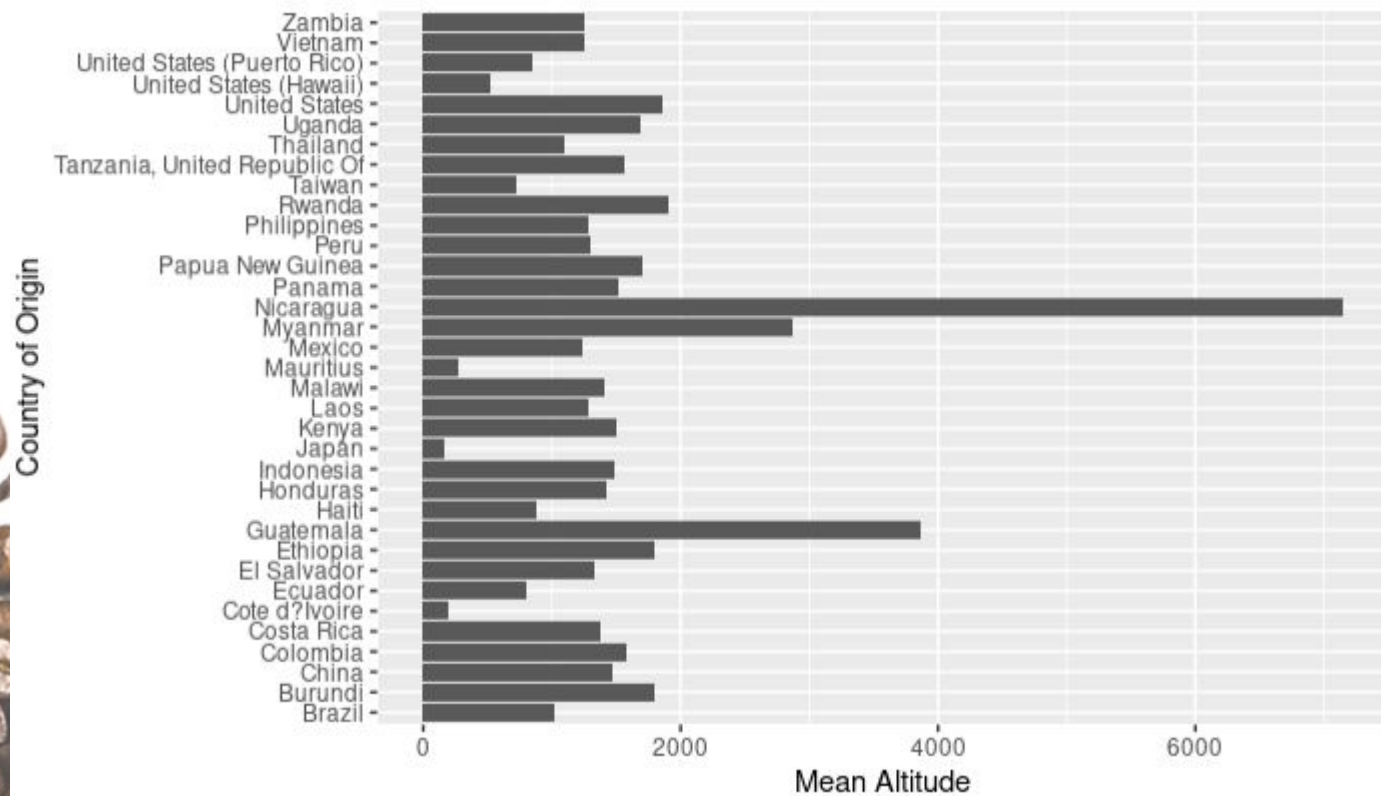

unique(coffee\$Altitude)

```
[1] 2075.0000 1700.0000 2000.0000      NA 1635.0000 1822.5000 1905.0000 1872.0000 1943.0000
[10] 609.6000 2080.0000 1500.0000 1450.0000 1850.0000 2019.0000 1300.0000 1320.0000 2112.0000
[19] 1250.0000 1950.0000 1400.0000 1200.0000 1775.0000 1800.0000 1941.0000 12.0000 1000.0000
[28] 1754.0000 1860.0000 1650.0000 1750.0000 426.7200 1600.0000 1900.0000 1524.0000 1417.3200
[37] 1350.0000 1680.0000 1731.2640 1770.0000 1550.0000 1325.0000 2560.0000 2136.0000 1.0000
[46] 1580.0000 1100.0000 1325.8800 800.0000 1620.0000 350.0000 925.0000 170.0000 1150.0000
[55] 1310.6400 1219.2000 1575.0000 900.0000 442.0000 1275.0000 1050.0000 1493.5200 1386.8400
[64] 1099.4136 1170.0000 1116.7872 1565.0000 3280.0000 1606.5000 1530.0000 250.5000 1038.5000
[73] 968.0000 1480.0000 1706.8800 1260.0000 890.0000 934.0000 950.0000 2527.0000 525.0000
[82] 1676.5000 1218.0000 1499.9208 750.0000 853.4400 700.0000 1625.0000 1356.6648 1128.3696
[91] 1799.0000 940.0000 1402.0800 439.0000 1330.0000 1813.0000 1227.0000 1280.1600 1880.0000
[100] 2100.0000 1261.0000 1483.0000 175.0000 1901.0000 168.0000 1653.0000 532.0000 1089.0000
[109] 1380.0000 905.0000 894.0000 872.0000 1473.0000 1645.0000 1371.6000 775.0000 1059.0000
[118] 11000.0000 1248.0000 1040.0000 157.8864 650.0000 1442.0000 165.0000 1296.0000 1240.0000
[127] 3850.0000 200.0000 1317.0000 975.0000 1679.0000 2285.0000 1560.0000 1396.0000 982.0000
[136] 1268.0000 1676.4000 600.0000 1752.6000 441.0000 1180.0000 1338.0000 533.0000 1390.8024
[145] 1144.0000 13.0000 944.0000 2500.0000 1020.0000 1877.0000 981.0000 690.0000 973.0000
[154] 4001.0000 1425.0000 1556.0000 774.0000 150.0000 190164.0000 850.0000 1600.2000 180.0000
[163] 1599.0000 825.0000 125.0000 110.0000 1422.0000 763.0000 1130.0000 680.0000 995.0000
[172] 695.0000 1456.0000 1642.0000 758.0000 1525.0000 1066.8000 550.0000 688.0000 300.0000
[181] 1186.0000 1210.0000 100.0000 3825.0000 110000.0000 185.0000 1383.7920 50.0000 280.0000
[190] 3800.0000 1859.2800 914.4000 4287.0000 808.0000 1187.5000 250.0000 3845.0000 1022.0000
[199] 1264.0000 3500.0000 1280.0000 1140.0000
```

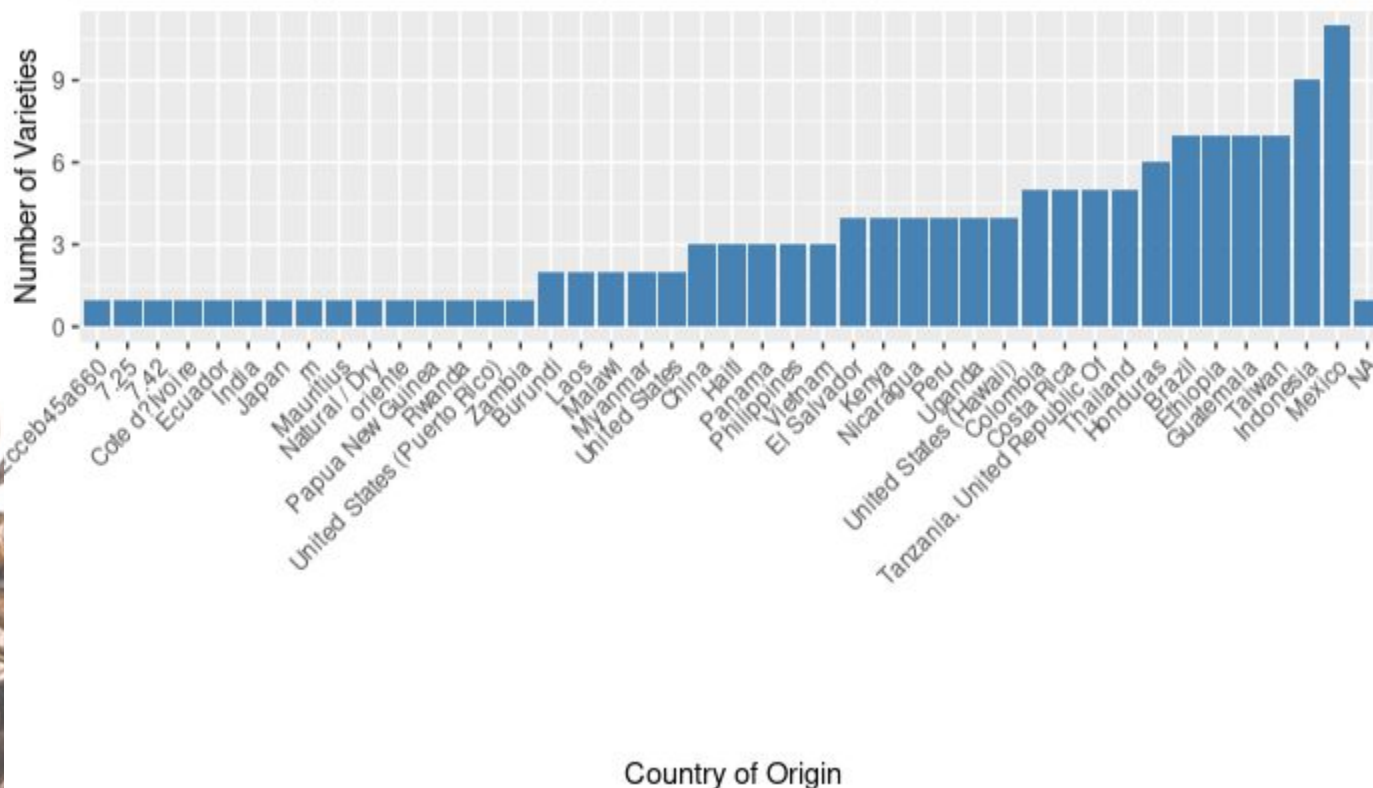
Processing Methods Used By Each Country



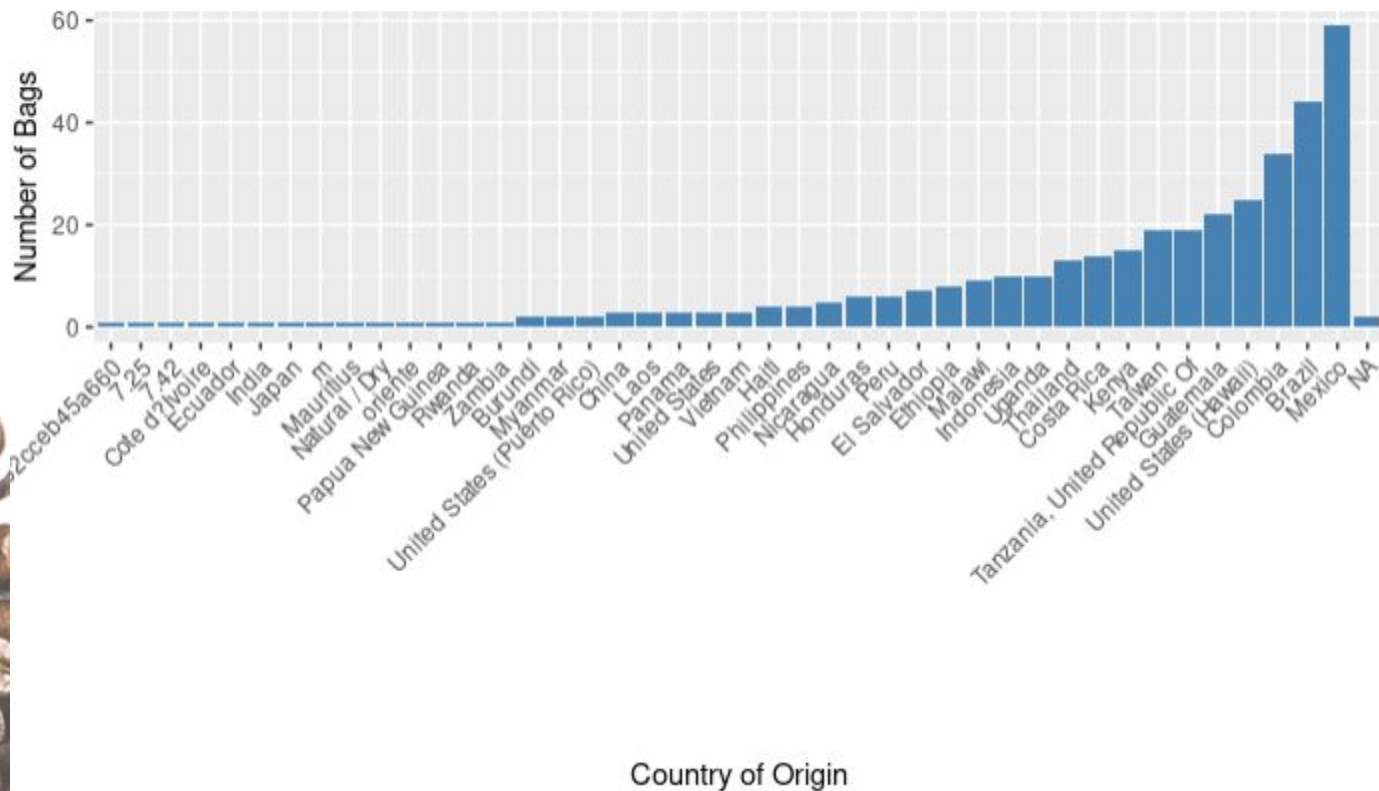
Mean Altitude of Coffee Beans by Country



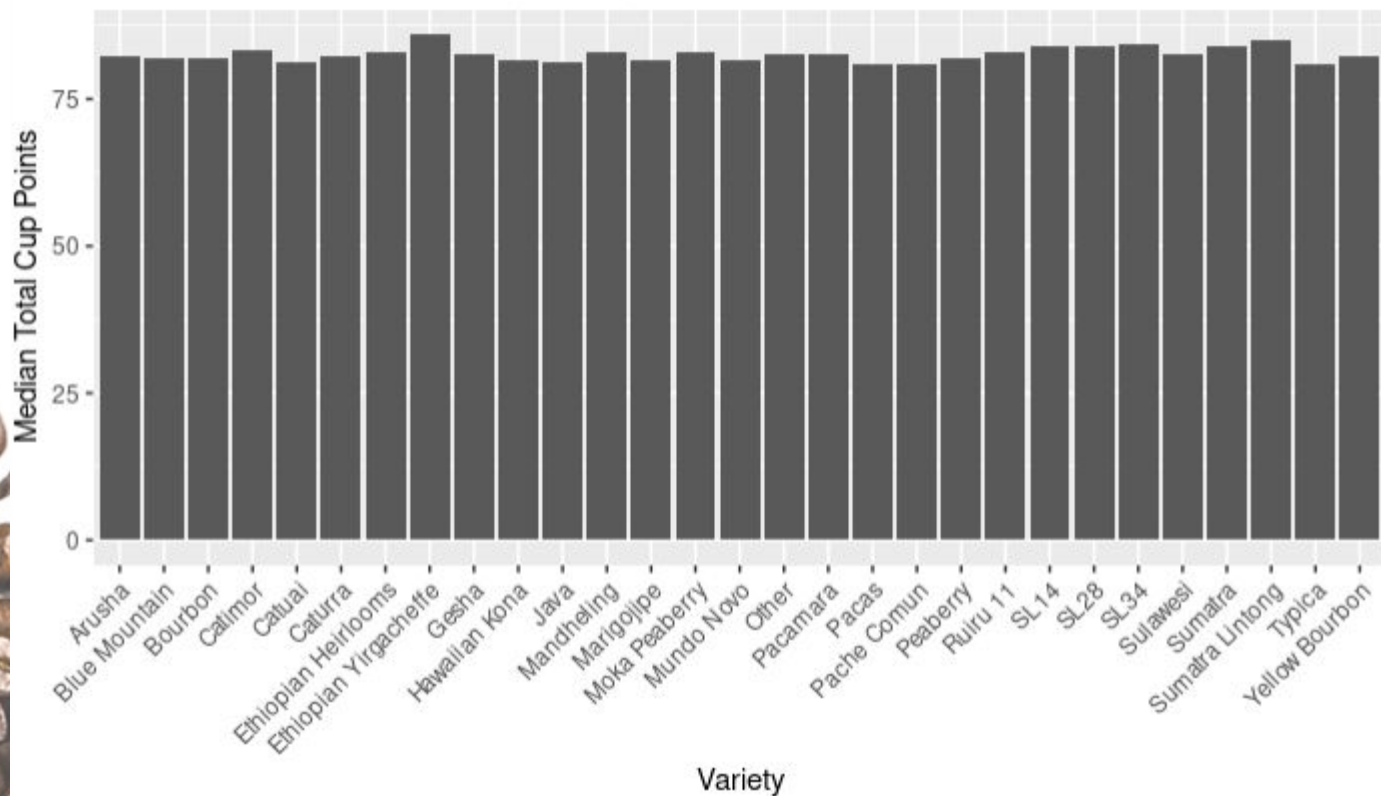
Number of Coffee Bean Varieties per Country of Origin



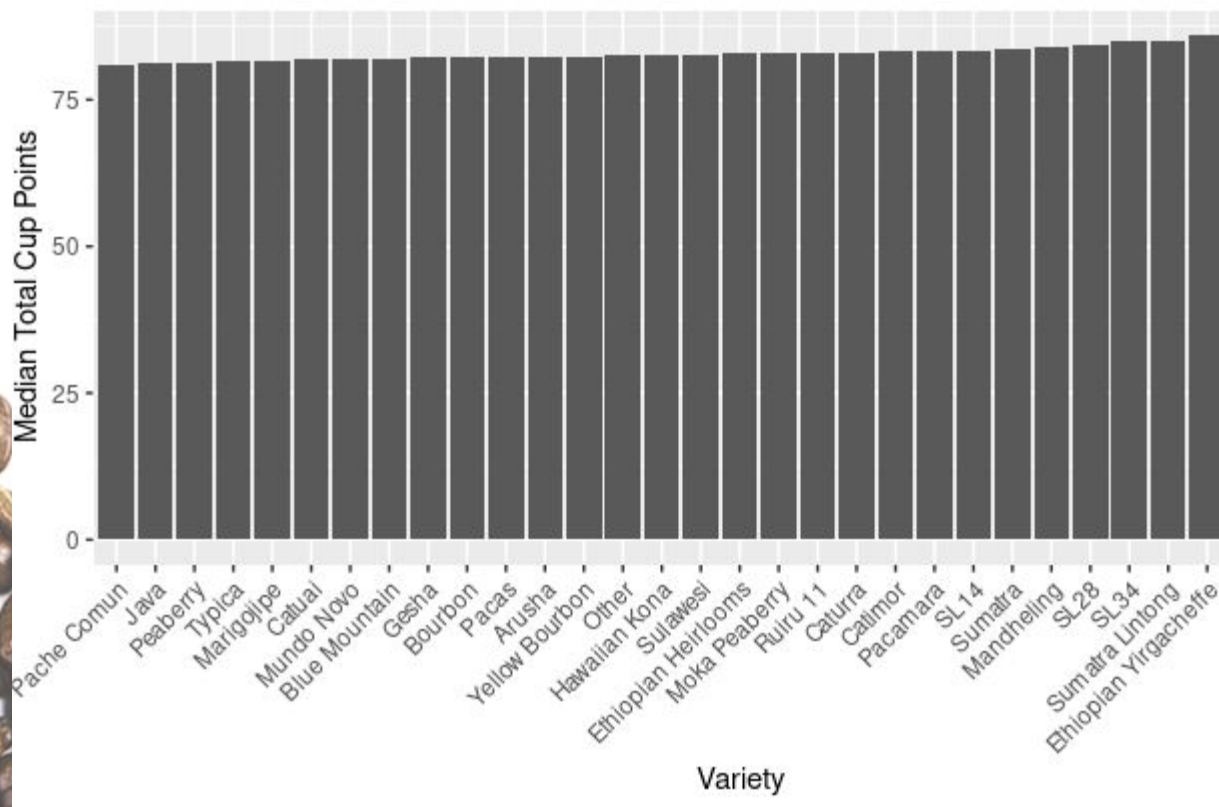
Number of Coffee Bean Bags per Country of Origin



Median Total Cup Points by Variety



Median Total Cup Points by Variety



Sources

<https://github.com/jldbc/coffee-quality-database>

