

Appendix A: Descriptive statistics

Market characteristics and symbolic attributes:

| Market characteristics | N | Mean/ Proportion ¹ | Standard Deviation | Minimum | Maximum |
|--------------------------------|------|----------------------------------|-----------------------|---------|---------|
| Price (2010\$/lb) | 2889 | 5.95 | 7.07 | 1.37 | 97.40 |
| Quality score | 2889 | 87.71 | 1.90 | 84.00 | 95.18 |
| Auction lot size in 30kg boxes | 2889 | 27.69 | 15.37 | 2.00 | 145.00 |
| Number of coffees in auction | 2889 | 29.23 | 7.52 | 10.00 | 54.00 |
| Symbolic attributes | N | Mean/ Proportion ¹ | Standard Deviation | Minimum | Maximum |
| Altitude | 2889 | 1532.84 | 235.74 | 720.00 | 2300.00 |
| Number of descriptors | 2889 | 17.61 | 6.78 | 4.00 | 62.00 |
| Certifications | | | | | |
| Organic | 2889 | 0.02 | - | - | - |
| Rainforest Alliance | 2889 | 0.01 | - | - | - |
| Ranking in competition | | | | | |
| Rank 1 | 2889 | 0.05 | - | - | - |
| Rank 2 | 2889 | 0.05 | - | - | - |
| Rank 3 | 2889 | 0.04 | - | - | - |
| Rank 4 | 2889 | 0.04 | - | - | - |
| Year | | | | | |
| 2009 | 2889 | 0.08 | - | - | - |
| 2010 | 2889 | 0.05 | - | - | - |
| 2011 | 2889 | 0.08 | - | - | - |
| 2012 | 2889 | 0.09 | - | - | - |
| 2013 | 2889 | 0.07 | - | - | - |
| 2014 | 2889 | 0.10 | - | - | - |
| 2015 | 2889 | 0.12 | - | - | - |
| 2016 | 2889 | 0.04 | - | - | - |
| 2017 | 2889 | 0.12 | - | - | - |
| 2018 | 2889 | 0.10 | - | - | - |
| 2019 | 2889 | 0.08 | - | - | - |
| 2020 | 2889 | 0.05 | - | - | - |
| Country | | | | | |
| Brazil | 2889 | 0.17 | - | - | - |
| Bolivia | 2889 | 0.01 | - | - | - |
| Burundi | 2889 | 0.06 | - | - | - |
| Colombia | 2889 | 0.08 | - | - | - |
| Costa Rica | 2889 | 0.11 | - | - | - |
| El Salvador | 2889 | 0.12 | - | - | - |
| Ethiopia | 2889 | 0.01 | - | - | - |
| Guatemala | 2889 | 0.11 | - | - | - |
| Honduras | 2889 | 0.11 | - | - | - |
| Mexico | 2889 | 0.06 | - | - | - |
| Nicaragua | 2889 | 0.08 | - | - | - |
| Peru | 2889 | 0.02 | - | - | - |

¹ For binary variables only proportions are reported.

| | | | | | |
|------------------|------|------|---|---|---|
| Rwanda | 2889 | 0.07 | - | - | - |
| Processing | | | - | - | - |
| Wet processing | 2889 | 0.51 | - | - | - |
| Dry Processing | 2889 | 0.15 | - | - | - |
| Other Processing | 2889 | 0.34 | - | - | - |
| Variety | | | - | - | - |
| Bourbon | 2889 | 0.23 | - | - | - |
| Catuai | 2889 | 0.13 | - | - | - |
| Caturra | 2889 | 0.13 | - | - | - |
| Typica | 2889 | 0.03 | - | - | - |
| Pacamara | 2889 | 0.08 | - | - | - |
| Mixed variety | 2889 | 0.10 | - | - | - |
| Other variety | 2889 | 0.28 | - | - | - |

Tasting note variables:

| Material Attributes | N | Binary Encoding | Cumulative encoding | | | |
|-------------------------|------|--------------------|---------------------|-----------------------|---------|---------|
| | | Proportion | Mean | Standard Deviation | Minimum | Maximum |
| Level 1 of coffee wheel | | | | | | |
| Floral | 2889 | 0.53 | 0.76 | 0.89 | 0.00 | 6.00 |
| Fruity | 2889 | 0.98 | 4.14 | 2.27 | 0.00 | 14.00 |
| Green/vegetative | 2889 | 0.27 | 0.56 | 1.00 | 0.00 | 6.00 |
| Nutty/cocoa | 2889 | 0.60 | 1.13 | 1.27 | 0.00 | 8.00 |
| Sour/fermented | 2889 | 0.70 | 1.05 | 0.94 | 0.00 | 6.00 |
| Sweet | 2889 | 0.88 | 2.17 | 1.49 | 0.00 | 8.00 |
| Spices | 2889 | 0.31 | 0.40 | 0.68 | 0.00 | 4.00 |
| Roasted | 2889 | 0.12 | 0.13 | 0.40 | 0.00 | 4.00 |
| Other | 2889 | 0.01 | 0.01 | 0.12 | 0.00 | 2.00 |
| Non-flavour descriptors | | | | | | |
| Balanced cup | 2889 | 0.27 | 0.29 | 0.48 | 0.00 | 2.00 |
| Creamy body | 2889 | 0.37 | 0.39 | 0.52 | 0.00 | 2.00 |
| Long Aftertaste | 2889 | 0.22 | 0.24 | 0.46 | 0.00 | 3.00 |
| Buttery Mouthfeel | 2889 | 0.10 | 0.10 | 0.30 | 0.00 | 1.00 |
| Lingering Aftertaste | 2889 | 0.09 | 0.09 | 0.28 | 0.00 | 1.00 |
| Round body | 2889 | 0.19 | 0.19 | 0.40 | 0.00 | 2.00 |
| Juicy Mouthfeel | 2889 | 0.31 | 0.32 | 0.48 | 0.00 | 3.00 |
| Clean and clear cup | 2889 | 0.22 | 0.24 | 0.46 | 0.00 | 2.00 |
| Smooth Mouthfeel | 2889 | 0.21 | 0.22 | 0.43 | 0.00 | 2.00 |
| Transparent cup | 2889 | 0.13 | 0.13 | 0.34 | 0.00 | 2.00 |
| Big body | 2889 | 0.04 | 0.04 | 0.20 | 0.00 | 2.00 |

| Extended Material Attributes | N | Binary Encoding | Cumulative encoding | | | |
|---------------------------------|---|--------------------|---------------------|-----------------------|---------|---------|
| | | Proportion | Mean | Standard Deviation | Minimum | Maximum |

| | | | | | | |
|-------------------------|------|------|------|------|------|------|
| Level 2 of Coffee wheel | | | | | | |
| Alcohol/fermented | 2889 | 0.16 | 0.18 | 0.45 | 0.00 | 4.00 |
| Berry | 2889 | 0.43 | 0.59 | 0.81 | 0.00 | 6.00 |
| Brown sugar | 2889 | 0.76 | 1.37 | 1.09 | 0.00 | 6.00 |
| Citrus fruit | 2889 | 0.70 | 1.19 | 1.10 | 0.00 | 6.00 |
| Cocoa | 2889 | 0.55 | 0.83 | 0.94 | 0.00 | 5.00 |
| Floral | 2889 | 0.29 | 0.34 | 0.58 | 0.00 | 4.00 |
| Green/vegetative | 2889 | 0.27 | 0.30 | 0.53 | 0.00 | 3.00 |
| Other fruit | 2889 | 0.81 | 1.64 | 1.25 | 0.00 | 7.00 |
| Brown spice | 2889 | 0.18 | 0.20 | 0.46 | 0.00 | 3.00 |
| Dried fruit | 2889 | 0.13 | 0.14 | 0.38 | 0.00 | 3.00 |
| Sour | 2889 | 0.64 | 0.87 | 0.82 | 0.00 | 6.00 |
| Vanilla | 2889 | 0.24 | 0.24 | 0.44 | 0.00 | 2.00 |
| Pipe tobacco | 2889 | 0.01 | 0.01 | 0.07 | 0.00 | 1.00 |
| Tobacco | 2889 | 0.07 | 0.07 | 0.26 | 0.00 | 2.00 |
| Cereal | 2889 | 0.03 | 0.03 | 0.17 | 0.00 | 2.00 |
| Nutty | 2889 | 0.17 | 0.18 | 0.42 | 0.00 | 3.00 |
| Chemical | 2889 | 0.01 | 0.01 | 0.10 | 0.00 | 2.00 |
| Pepper | 2889 | 0.03 | 0.03 | 0.17 | 0.00 | 2.00 |
| Black tea | 2889 | 0.08 | 0.08 | 0.28 | 0.00 | 2.00 |
| Pungent | 2889 | 0.00 | 0.00 | 0.03 | 0.00 | 1.00 |
| Raw | 2889 | 0.00 | 0.00 | 0.04 | 0.00 | 1.00 |
| Papery/musty | 2889 | 0.00 | 0.00 | 0.05 | 0.00 | 1.00 |
| Burnt | 2889 | 0.01 | 0.01 | 0.08 | 0.00 | 1.00 |
| Level 3 of Coffee wheel | | | | | | |
| Cherry | 2889 | 0.28 | 0.31 | 0.52 | 0.00 | 3.00 |
| Chocolate | 2889 | 0.51 | 0.61 | 0.68 | 0.00 | 4.00 |
| Fresh | 2889 | 0.04 | 0.04 | 0.21 | 0.00 | 1.00 |
| Honey | 2889 | 0.41 | 0.42 | 0.53 | 0.00 | 3.00 |
| Jasmine | 2889 | 0.22 | 0.22 | 0.43 | 0.00 | 2.00 |
| Lemon | 2889 | 0.26 | 0.30 | 0.53 | 0.00 | 3.00 |
| Lime | 2889 | 0.18 | 0.20 | 0.43 | 0.00 | 3.00 |
| Raspberry | 2889 | 0.11 | 0.11 | 0.32 | 0.00 | 2.00 |
| Winey | 2889 | 0.16 | 0.18 | 0.44 | 0.00 | 4.00 |
| Cinnamon | 2889 | 0.14 | 0.14 | 0.36 | 0.00 | 2.00 |
| Citric acid | 2889 | 0.49 | 0.52 | 0.56 | 0.00 | 4.00 |
| Grape | 2889 | 0.28 | 0.32 | 0.54 | 0.00 | 3.00 |
| Grapefruit | 2889 | 0.12 | 0.12 | 0.33 | 0.00 | 2.00 |
| Orange | 2889 | 0.41 | 0.46 | 0.60 | 0.00 | 4.00 |
| Peach | 2889 | 0.29 | 0.31 | 0.49 | 0.00 | 3.00 |
| Pineapple | 2889 | 0.12 | 0.12 | 0.33 | 0.00 | 2.00 |
| Blueberry | 2889 | 0.08 | 0.08 | 0.28 | 0.00 | 2.00 |
| Molasses | 2889 | 0.11 | 0.11 | 0.32 | 0.00 | 2.00 |
| Pear | 2889 | 0.07 | 0.07 | 0.25 | 0.00 | 2.00 |
| Strawberry | 2889 | 0.13 | 0.13 | 0.35 | 0.00 | 2.00 |
| Apple | 2889 | 0.40 | 0.48 | 0.65 | 0.00 | 4.00 |
| Rose | 2889 | 0.10 | 0.11 | 0.33 | 0.00 | 3.00 |
| Caramelized | 2889 | 0.41 | 0.43 | 0.53 | 0.00 | 3.00 |
| Dark chocolate | 2889 | 0.13 | 0.14 | 0.35 | 0.00 | 2.00 |
| Malt | 2889 | 0.03 | 0.03 | 0.16 | 0.00 | 1.00 |

| | | | | | | |
|--------------|------|------|------|------|------|------|
| Coconut | 2889 | 0.02 | 0.02 | 0.15 | 0.00 | 1.00 |
| Almond | 2889 | 0.11 | 0.11 | 0.31 | 0.00 | 2.00 |
| Malic acid | 2889 | 0.32 | 0.33 | 0.48 | 0.00 | 2.00 |
| Blackberry | 2889 | 0.08 | 0.08 | 0.27 | 0.00 | 1.00 |
| Anise | 2889 | 0.02 | 0.02 | 0.13 | 0.00 | 1.00 |
| Raisin | 2889 | 0.09 | 0.09 | 0.30 | 0.00 | 2.00 |
| Bitter | 2889 | 0.01 | 0.01 | 0.10 | 0.00 | 2.00 |
| Clove | 2889 | 0.03 | 0.03 | 0.17 | 0.00 | 1.00 |
| Hazelnut | 2889 | 0.05 | 0.05 | 0.21 | 0.00 | 2.00 |
| Prune | 2889 | 0.03 | 0.03 | 0.16 | 0.00 | 1.00 |
| Maple syrup | 2889 | 0.02 | 0.02 | 0.15 | 0.00 | 1.00 |
| Musty/earthy | 2889 | 0.00 | 0.00 | 0.03 | 0.00 | 1.00 |
| Smoky | 2889 | 0.00 | 0.00 | 0.06 | 0.00 | 1.00 |
| Whiskey | 2889 | 0.00 | 0.00 | 0.05 | 0.00 | 1.00 |
| Nutmeg | 2889 | 0.01 | 0.01 | 0.11 | 0.00 | 1.00 |
| Peanuts | 2889 | 0.00 | 0.00 | 0.06 | 0.00 | 1.00 |
| Pomegranate | 2889 | 0.02 | 0.02 | 0.12 | 0.00 | 1.00 |
| Grain | 2889 | 0.00 | 0.00 | 0.04 | 0.00 | 1.00 |
| Chamomile | 2889 | 0.01 | 0.01 | 0.10 | 0.00 | 1.00 |
| Woody | 2889 | 0.00 | 0.00 | 0.04 | 0.00 | 1.00 |
| Musty/dusty | 2889 | 0.00 | 0.00 | 0.02 | 0.00 | 1.00 |
| Salty | 2889 | 0.00 | 0.00 | 0.03 | 0.00 | 1.00 |
| Acetic acid | 2889 | 0.00 | 0.00 | 0.02 | 0.00 | 1.00 |

Appendix B: “SCAA Coffee Taster's Flavor wheel”



SCAA Coffee Taster's Flavor wheel (Spencer et al., 2016)

Appendix C: Software used

| R packages | Version | Reference | Link |
|------------|---------|--------------------------|---|
| corrplot | 0.84 | (Wei et al., 2013) | https://github.com/taiyun/corrplot |
| naniar | 0.6.0 | (Tierney et al., 2020) | https://github.com/njtierney/naniar |
| qdapTools | 1.3.2 | (Rinker, 2015) | https://github.com/trinker/qdapTools |
| R Core | 4.0.3 | (R Core Team, 2020) | https://www.r-project.org/ |
| Rcrawler | 0.1.9-1 | (Khalil and Fakir, 2017) | https://github.com/salimk/Rcrawler |
| tictoc | 1.0 | (Izrailev, 2014) | https://github.com/collectivemedia/tictoc |
| Tidyverse | 1.3.0 | (Wickham et al., 2019) | https://www.tidyverse.org/ |
| viridis | 0.5.1 | (Garnier, 2018) | https://github.com/sjmgarnier/viridis |
| visdat | 0.5.3 | (Tierney, 2017) | https://github.com/ropensci/visdat |

| Python packages | Version | Reference | Link |
|-----------------|---------|---|---|
| Joblib | 1.0.1 | (Joblib Development Team, 2020) | https://joblib.readthedocs.io/ |
| Matplotlib | 3.3.4 | (Hunter, 2007) | https://matplotlib.org/ |
| NumPy | 1.19.2 | (Harris et al., 2020) | https://numpy.org/ |
| pandas | 1.2.2 | (The pandas development team, 2020; McKinney, 2010) | https://pandas.pydata.org/ |
| PyALE | 1.1.0 | (Jomar, n.d.) | https://pypi.org/project/PyALE/ |
| Python | 3.0.8 | (Van Rossum and Drake, 2009) | https://www.python.org/ |
| Scikit-learn | 0.24.1 | (Pedregosa et al., 2011) | https://scikit-learn.org/ |
| Scikit-Optimize | 0.24.1 | (Head et al., 2020) | https://scikit-optimize.github.io/stable/ |
| Seaborn | 0.11.1 | (Waskom, 2021) | https://seaborn.pydata.org/ |
| XGBoost | 1.3.3 | (Chen and Guestrin, 2016) | https://xgboost.readthedocs.io/ |

Appendix D: Definitions of dataset permutations

Tasting note variables

| Dataset permutations | Full variables | | | | Principal components | | | | | |
|----------------------|------------------|------------------|------------------|-------------------------|----------------------|------------------|------------------|-------------------------|------------|--------------------------------------|
| | Level 1 flavours | Level 2 flavours | Level 3 flavours | Non-flavour descriptors | Level 1 flavours | Level 2 flavours | Level 3 flavours | Non-flavour descriptors | All levels | All levels & non-flavour descriptors |
| Baseline | X | | | X | | | | | | |
| lvl2 | | X | | X | | | | | | |
| lvl3 | | | X | X | | | | | | |
| full | X | X | X | X | | | | | | |
| lvl1_noext | X | | | | | | | | | |
| lvl2_noext | | X | | | | | | | | |
| lvl3_noext | | | X | | | | | | | |
| lvl1+lvl2pca | X | | | X | | X | | | | |
| lvl3+lvl2pca | | | X | X | | X | | | | |
| lvl1+lvl3pca | X | | | X | | | X | | | |
| lvl2+lvl3pca | | X | | X | | | X | | | |
| lvl1pca | | | | X | X | | | | | |
| lvl2pca | | | | X | | X | | | | |
| lvl3pca | | | | X | | | X | | | |
| lvl1pca_noext | | | | | X | | | | | |
| lvl2pca_noext | | | | | | X | | | | |
| lvl3pca_noext | | | | | | | X | | | |
| fulllvlpca | | | | X | | | | | X | |
| fulllvlpca_noext | | | | | | | | | X | |
| extrapca-nolevel | | | | | | | | X | | |
| extrapca-lvl1 | X | | | X | | | | X | | |
| extrapca-lvl2 | | X | | X | | | | X | | |
| extrapca-lvl3 | | | X | X | | | | X | | |
| fullpca | | | | | | | | | | X |
| lvl1pca+lvl2pca | | | | | X | X | | | | |
| lvl3pca+lvl2pca | | | | | | X | X | | | |
| lvl1pca+lvl3pca | | | | | X | | X | | | |
| allseplvlpca | | | | X | X | X | X | | | |
| allsepfllpca | | | | | X | X | X | X | | |
| lvl1pca+extpca | | | | | X | | | X | | |
| lvl2pca+extpca | | | | | | X | | X | | |
| lvl3pca+extpca | | | | | | | X | X | | |