Coffee Quality Database

Source: https://github.com/jldbc/coffee-quality-database

Data

These data contain reviews of 1312 arabica and 28 robusta coffee beans from the Coffee Quality Institute's trained reviewers. The features include:

Quality Measures

- Aroma
- Flavor
- Aftertaste
- Acidity
- Body
- Balance
- Uniformity
- Cup Cleanliness
- Sweetness
- Moisture
- Defects

Bean Metadata

- Processing Method
- Color
- Species (arabica / robusta)

Farm Metadata

- Owner
- Country of Origin
- Farm Name
- Lot Number
- Mill
- Company
- Altitude
- Region

The data folder contains both raw and cleaned data. The raw data is exactly as it was found on the CQI site. Since these human-recorded data use a variety of different encodings, abbreviations, and units of measurement for their farm names, altitude, region, and other fields, I recommend using the cleaned data as a starting point.

The site was scraped using a Selenium headless browser and Beautiful Soup. To replicate this or collect updated data, create a login for the CQI site and enter your credentials in the [scraper](#)

Source

These data were collected from the Coffee Quality Institute's [review pages](#) in January 2018.