

Wizualizacja analizy jakości kawy

Zbiór danych

Zbiór danych pochodzi z bazy danych [Coffee Quality Institute](#). Są to dane ze stycznia 2018 roku. Zbiór dzieli się na dwie części dotyczące dwóch gatunków kawy: arabiki i robusty.

Zmienne

Związane z jakością

Parametry wyznaczone przez tzw. Q Gradera.

Smakowe

- Aromat
- Smak
- Posmak
- Kwasowość
- *Body*
- Balans
- Jednorodność
- *Cup Cleanliness*
- Słodkość

Pozostałe

- Wilgotność
- Defekty pierwszej i drugiej kategorii

Związane z ziarnem

- Metoda obróbki
- Kolor

Związane z uprawą

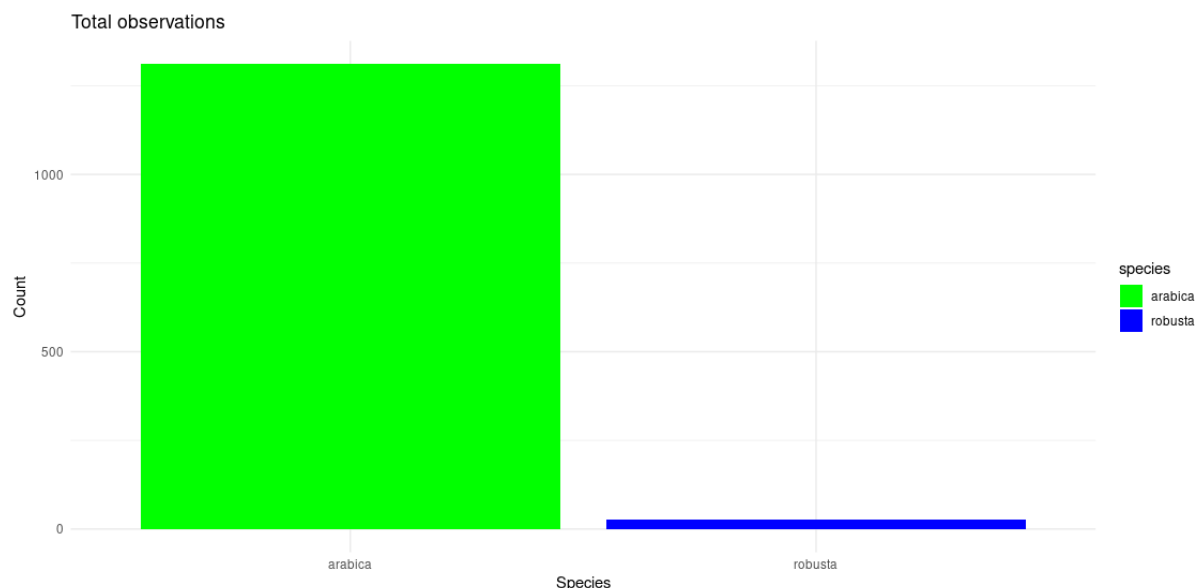
- Właściciel
- Kraj pochodzenia
- Nazwa farmy
- Numer partii
- Młyn
- Firma
- Wysokość uprawy
- Region

Analiza, wizualizacje

Ogólne statystyki

Liczebność obserwacji

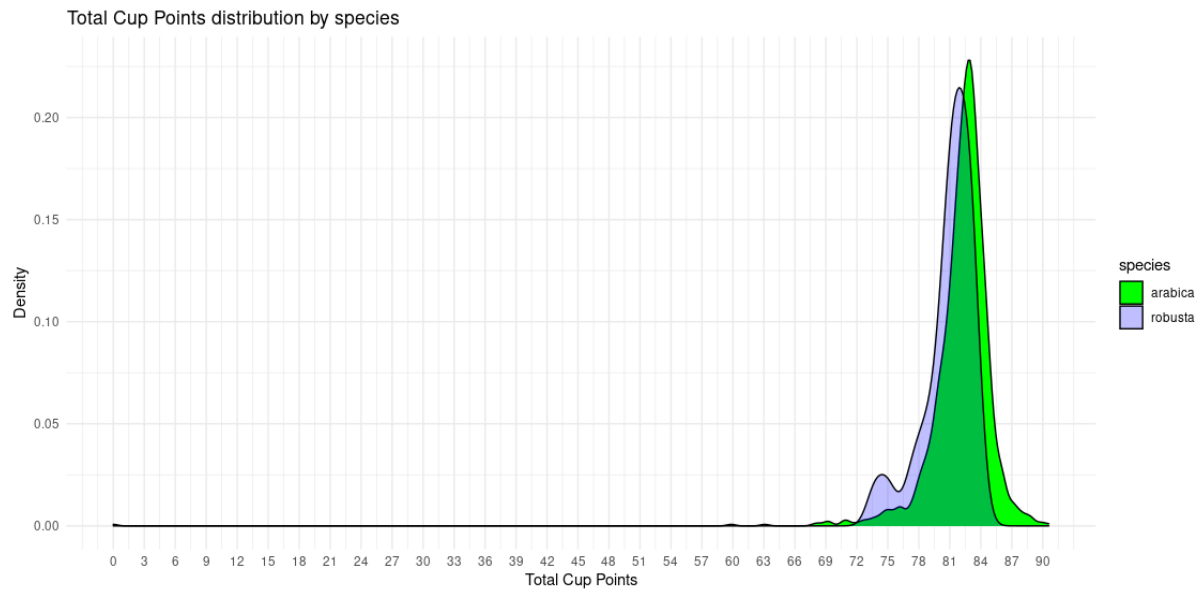
Wykres w wyższej rozdzielczości dostępny [tutaj](#).



Zbiór formalnych ocen arabiki jest niestety wielokrotnie większy niż robusty. W zbiorze jest odpowiednio 1311 i 28 ocen arabiki oraz robusty. Wynika to najpewniej z faktu, że robusta jest mniej popularną kawą w Stanach Zjednoczonych i Europie. Mała liczba ocen robusty będzie miała wpływ na jej analizę, więc te wizualizacje należy traktować z pewnym przymrużeniem oka.

Dystrybucja ocen obu gatunków kawy

Wykres w wyższej rozdzielczości dostępny [tutaj](#).

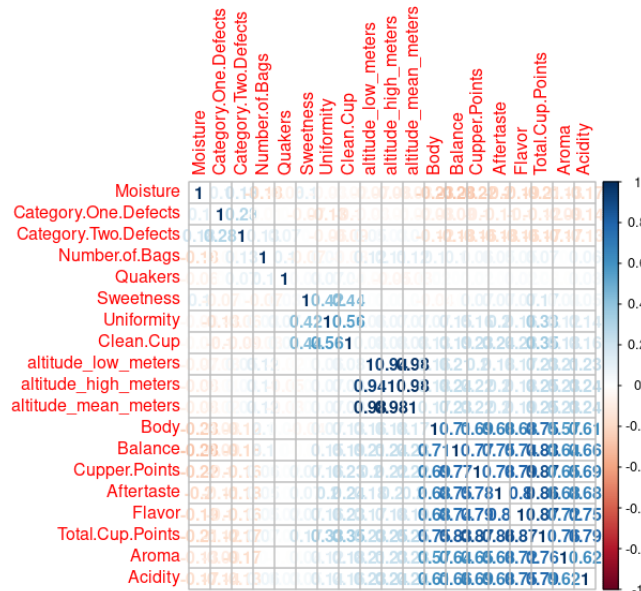


Tutaj możemy zauważyć, że w analizowanym zbiorze jest więcej lepiej ocenianych kaw w podzbiorze arabiki niż w podzbiorze robusty.

Arabika

Wizualizacja korelacji

Wykres w wyższej rozdzielczości dostępny [tutaj](#).

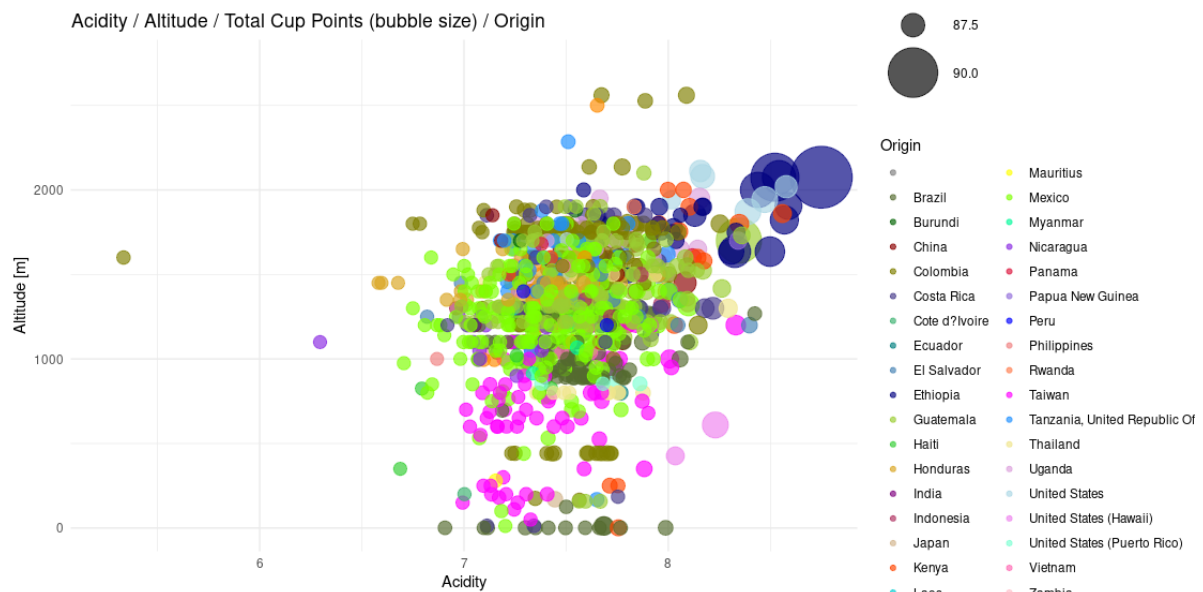


Korelacje, które są tutaj najbardziej interesujące to:

- Zależność kwasowości od wysokości uprawy
- Wpływ defektów na niektóre z parametrów formalnej oceny:
 - Wilgotność ziaren a balans (parametr podlegający formalnej ocenie)
 - Defekty drugiej kategorii a posmak (parametr podlegający formalnej ocenie)

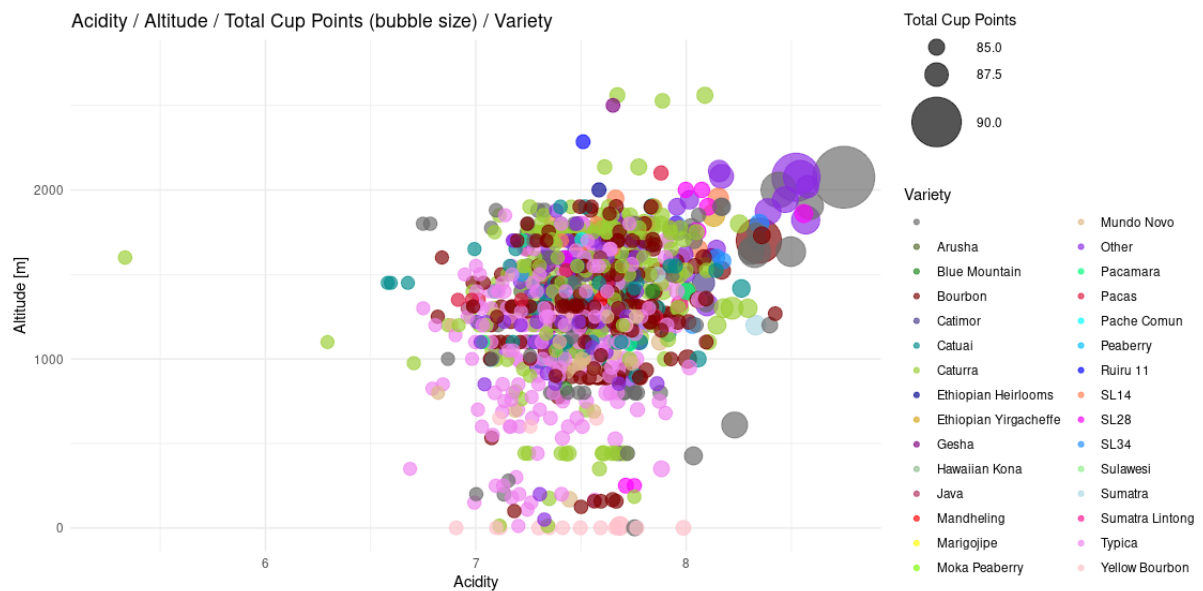
Kwasowość, wysokość uprawy, sumaryczna ocena (wielkość bąbelka) i kraj pochodzenia

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



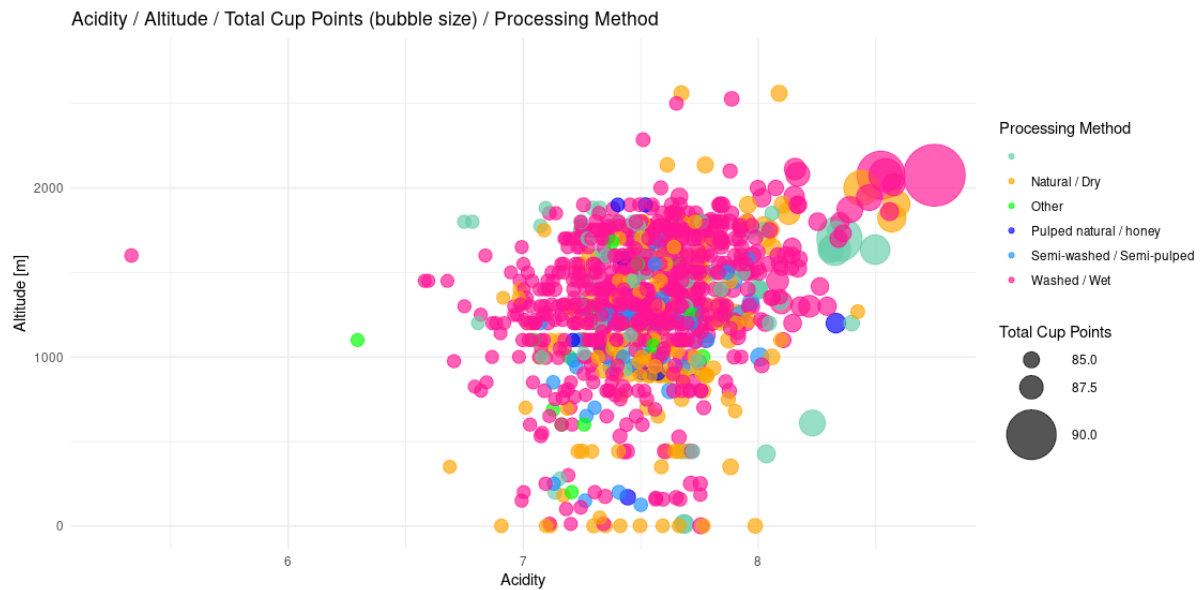
Kwasowość, wysokość uprawy, sumaryczna ocena (wielkość bąbelka) i odmiana arabiki

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



Kwasowość, wysokość uprawy, sumaryczna ocena (wielkość bąbelka) i metoda obróbki

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



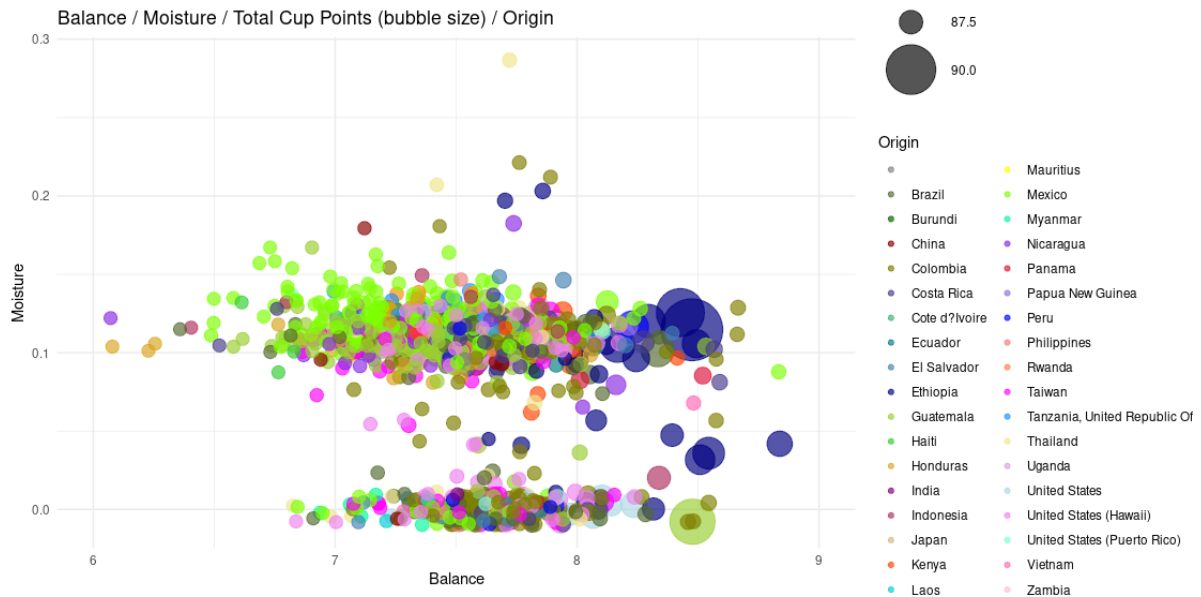
Notatki

Warto zauważyć, że najbardziej kwasowe kawy były uprawiane na wysokościach powyżej 1000 lub 1500 metrów nad poziomem morza. Są to głównie etiopskie ziarna.

Analiza wpływu defektów

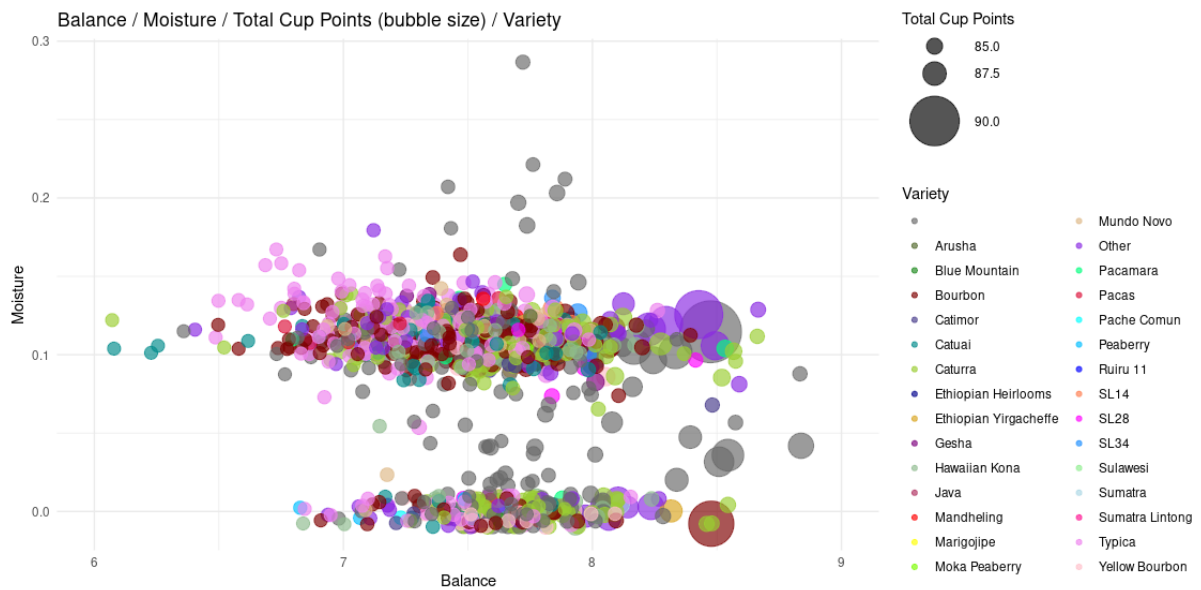
Balans, wilgotność, sumaryczna ocena (wielkość bąbelka) i kraj pochodzenia

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



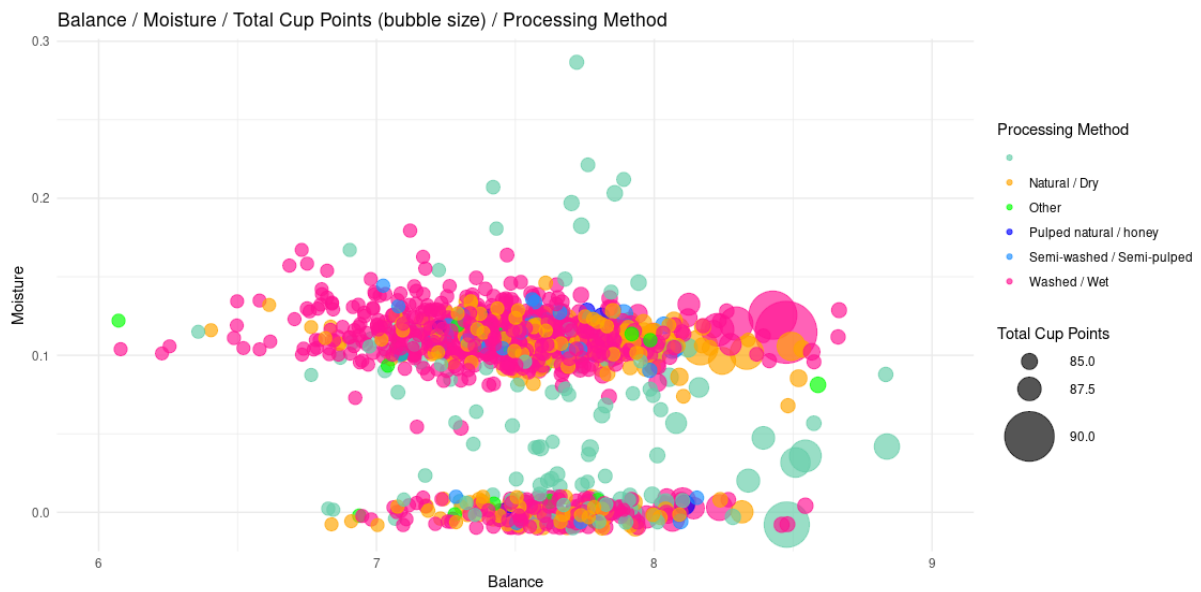
Balans, wilgotność, sumaryczna ocena (wielkość bąbelka) i odmiana arabiki

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



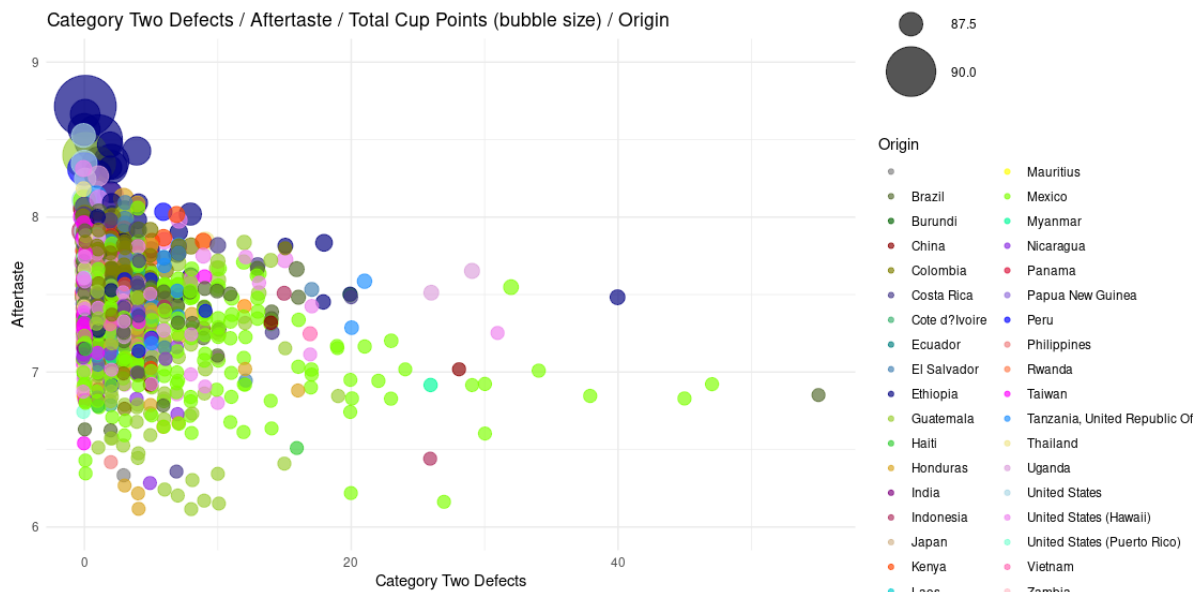
Balans, wilgotność, sumaryczna ocena (wielkość bąbelka) i metoda obróbki

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



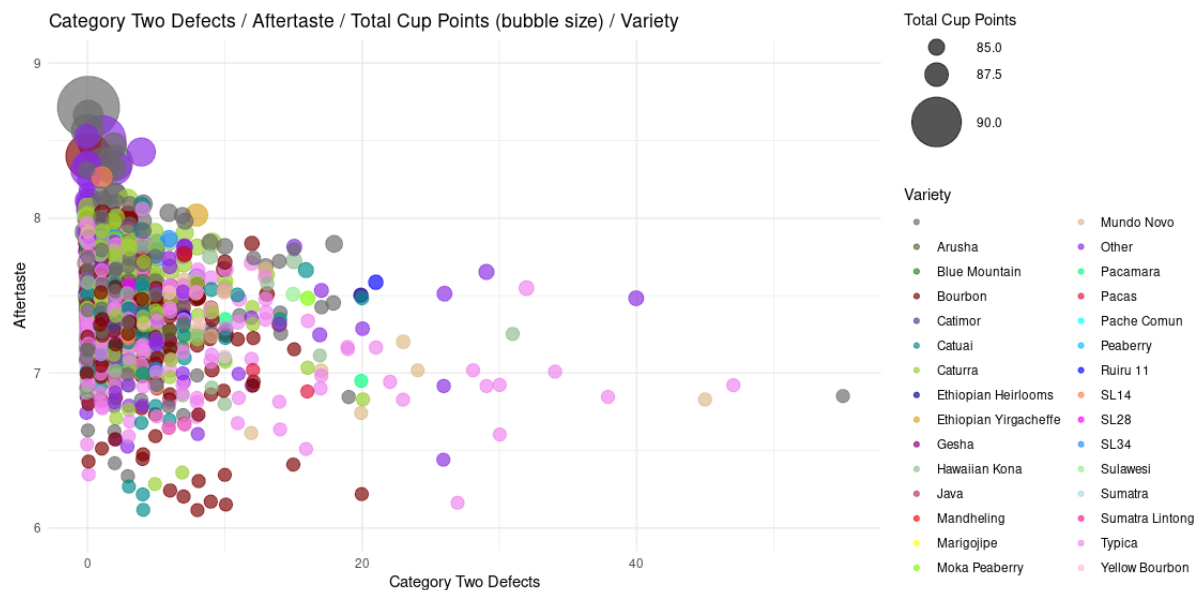
Defekty drugiej kategorii, posmak, sumaryczna ocena (wielkość bąbelka) i kraj pochodzenia

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



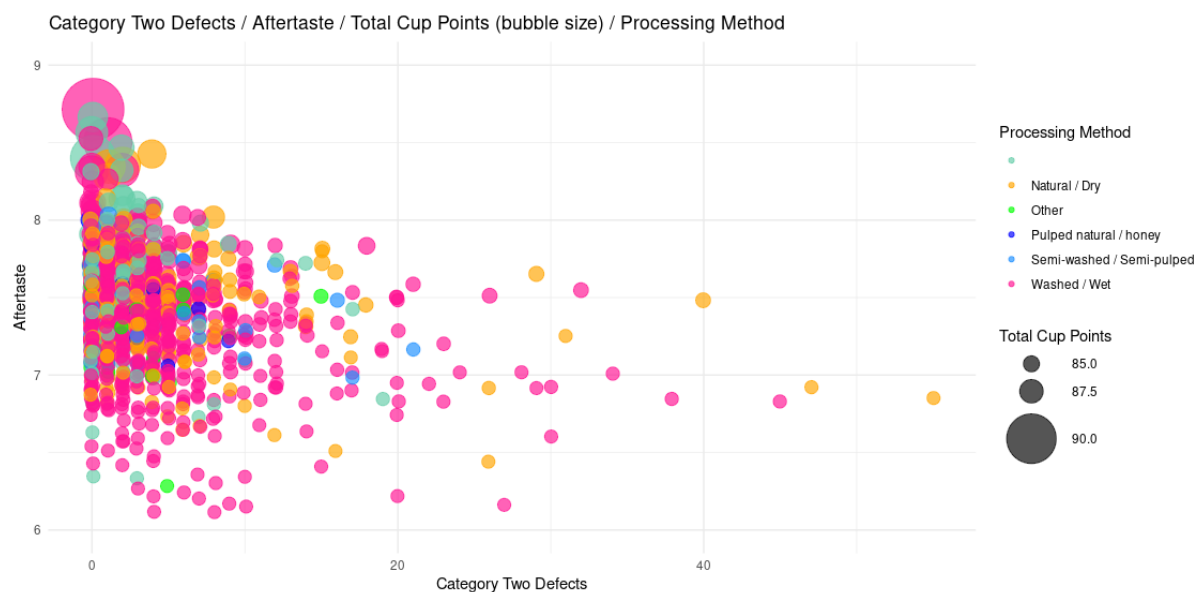
Defekty drugiej kategorii, posmak, sumaryczna ocena (wielkość bąbelka) i odmiana arabiki

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



Defekty drugiej kategorii, posmak, sumaryczna ocena (wielkość bąbelka) i metoda obróbki

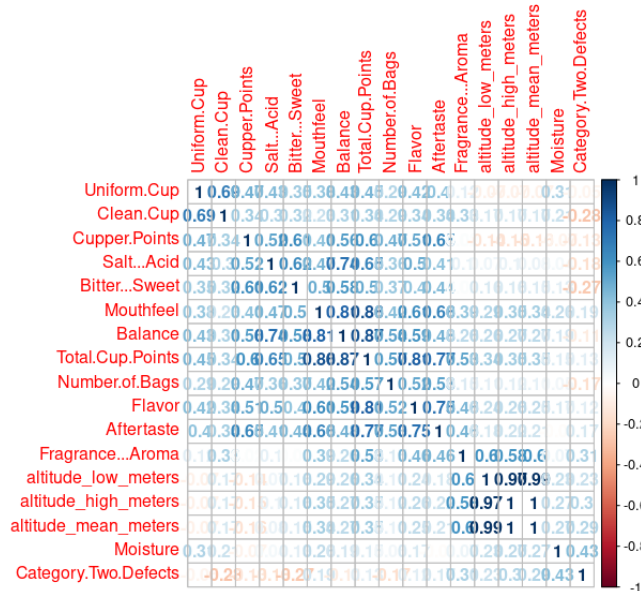
Wykres w wyższej rozdzielczości dostępny [tutaj](#).



Robusta

Wizualizacja korelacji

Wykres w wyższej rozdzielczości dostępny [tutaj](#).

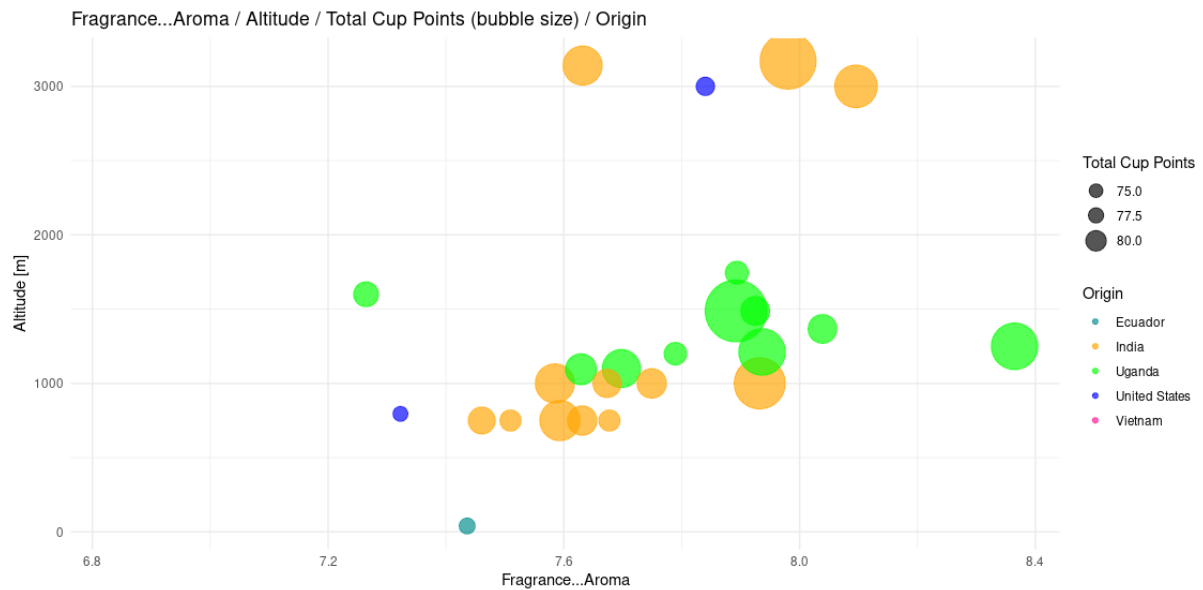


Korelacje, które są tutaj najbardziej interesujące to:

- Zależność aromatu od wysokości uprawy
- Wpływ defektów:
 - Wilgotność ziaren a defekty drugiej kategorii
 - Defekty drugiej kategorii a *Cup Cleanliness* (parametr formalnej oceny)

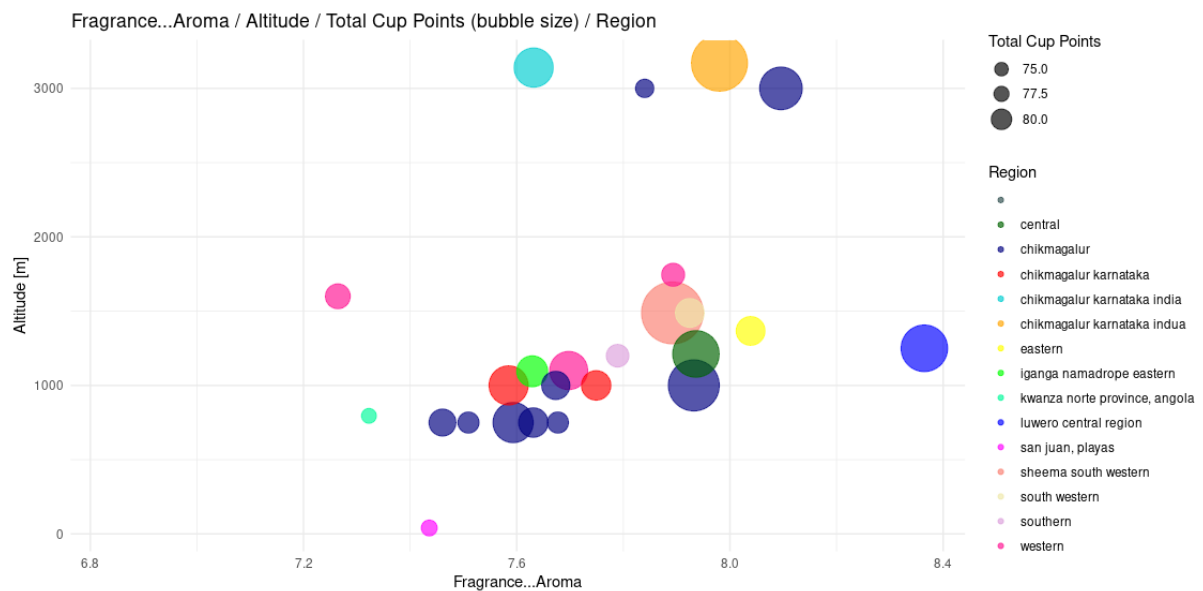
Aromat, wysokość uprawy, sumaryczna ocena (wielkość bąbelka) i kraj pochodzenia

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



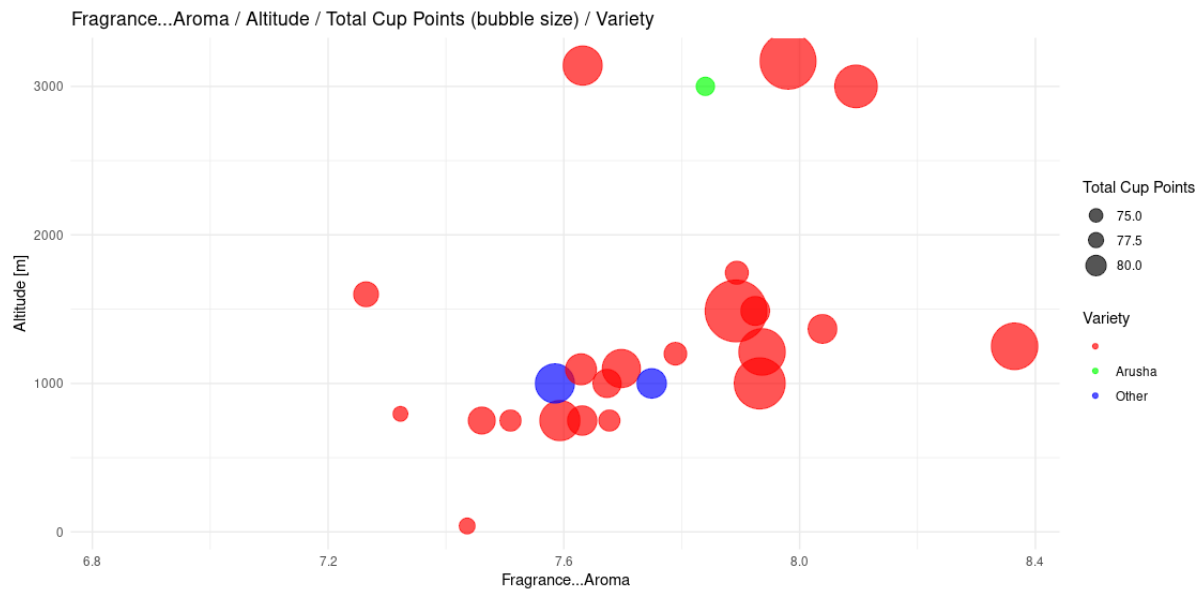
Aromat, wysokość uprawy, sumaryczna ocena (wielkość bąbelka) i region

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



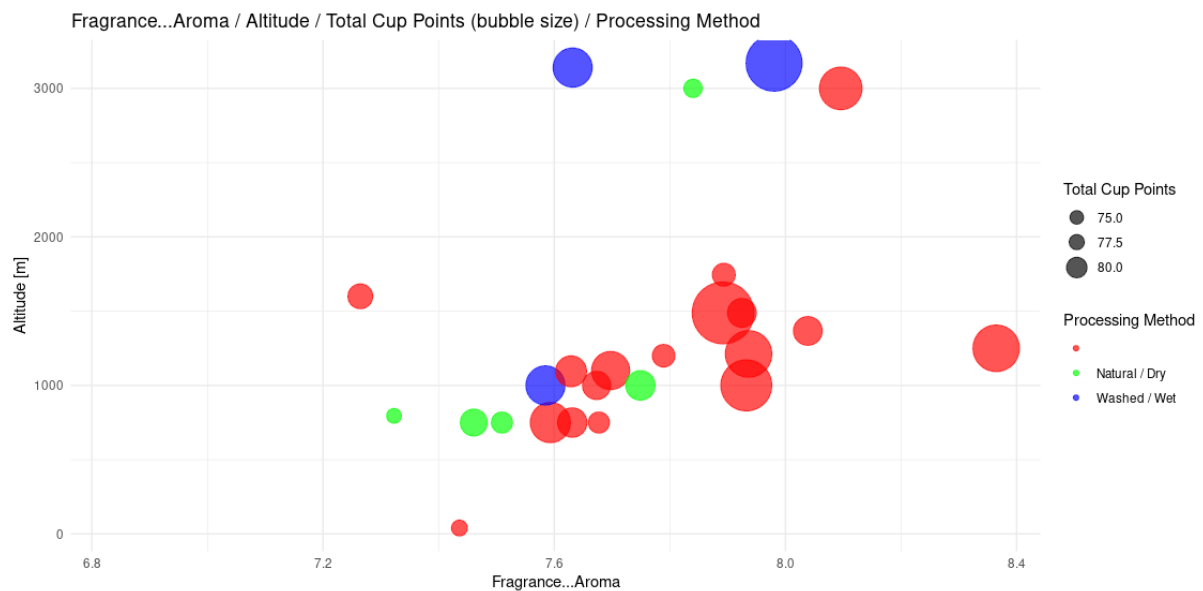
Aromat, wysokość uprawy, sumaryczna ocena (wielkość bąbelka) i odmiana robusty

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



Aromat, wysokość uprawy, sumaryczna ocena (wielkość bąbelka) i metoda obróbki

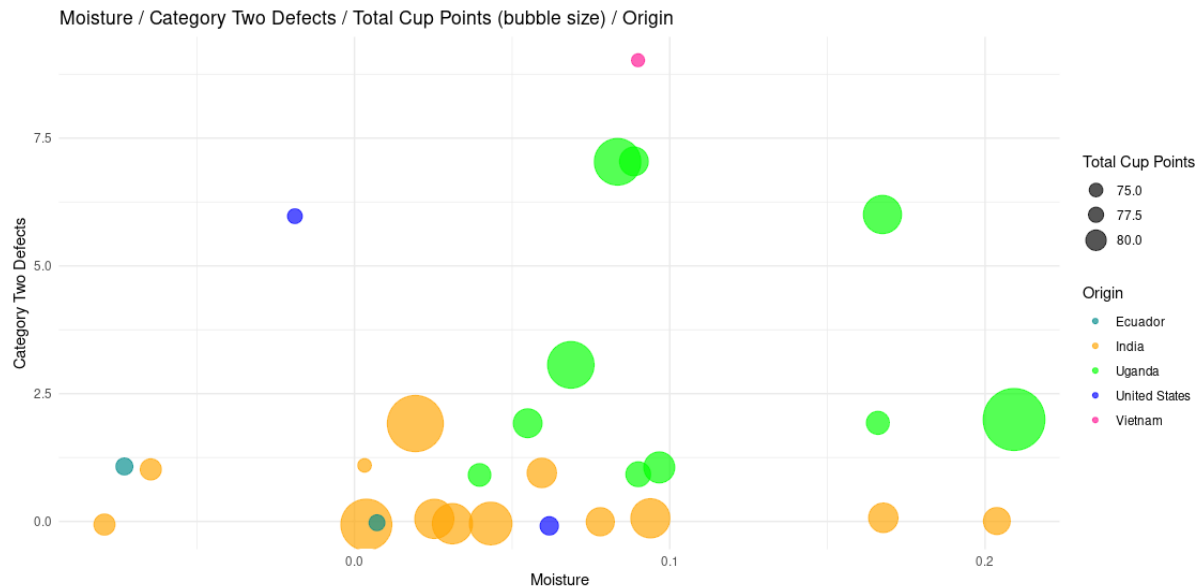
Wykres w wyższej rozdzielczości dostępny [tutaj](#).



Analiza wpływu defektów

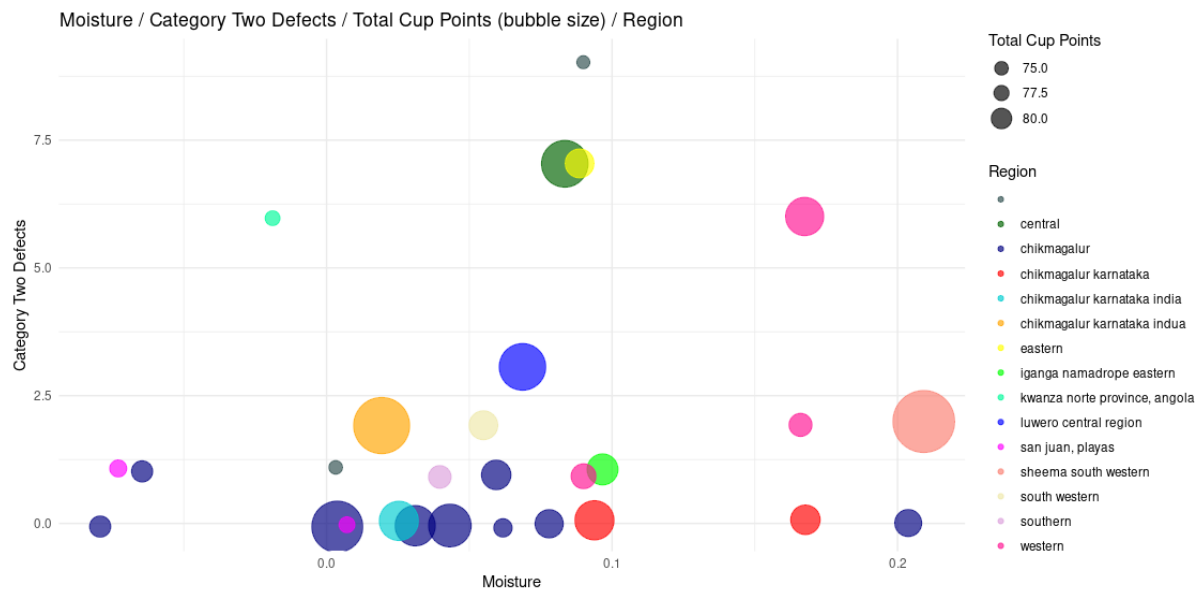
Wilgotność ziaren, defekty drugiej kategorii, sumaryczna ocena (wielkość bąbelka) i kraj pochodzenia

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



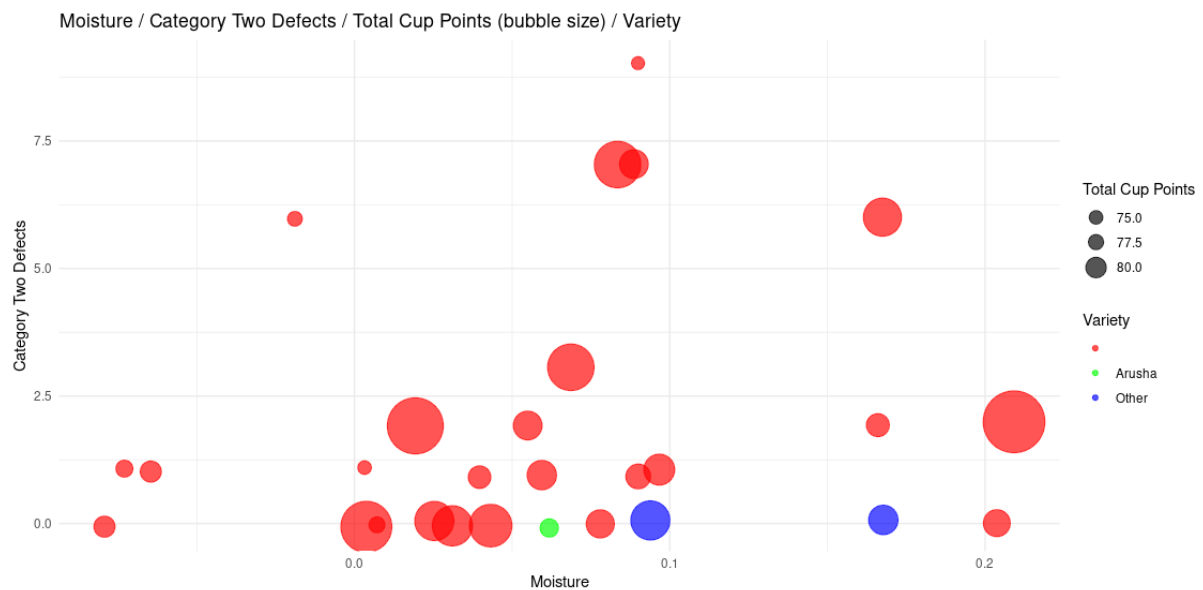
Wilgotność ziaren, defekty drugiej kategorii, sumaryczna ocena (wielkość bąbelka) i region

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



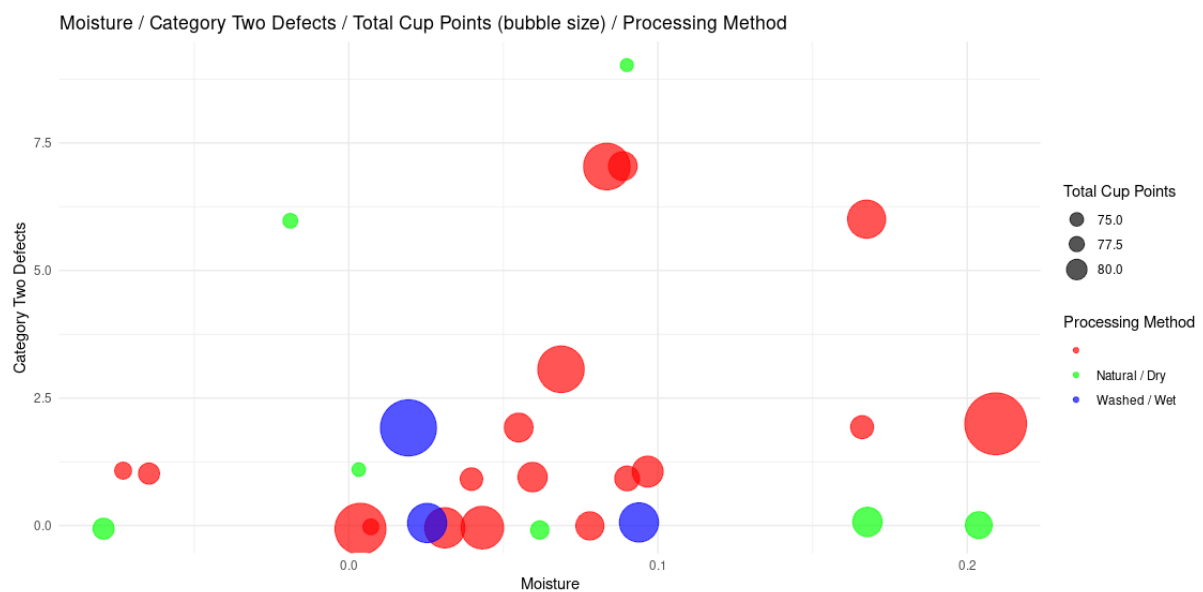
Wilgotność ziaren, defekty drugiej kategorii, sumaryczna ocena (wielkość bąbelka) i odmiana robusty

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



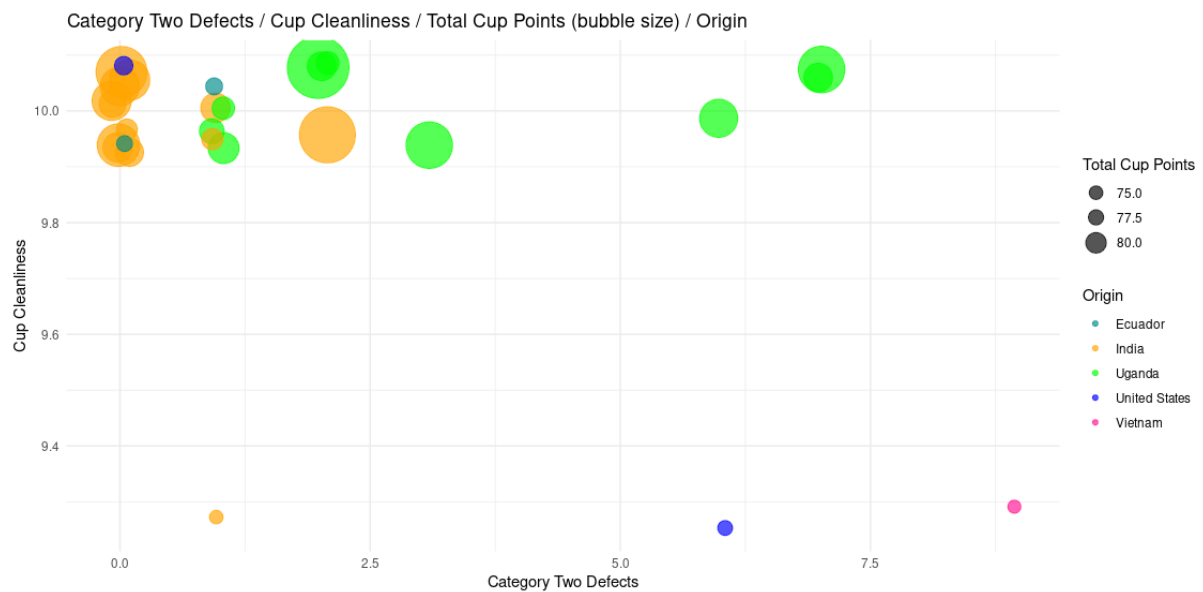
Wilgotność ziaren, defekty drugiej kategorii, sumaryczna ocena (wielkość bąbelka) i metoda obróbki

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



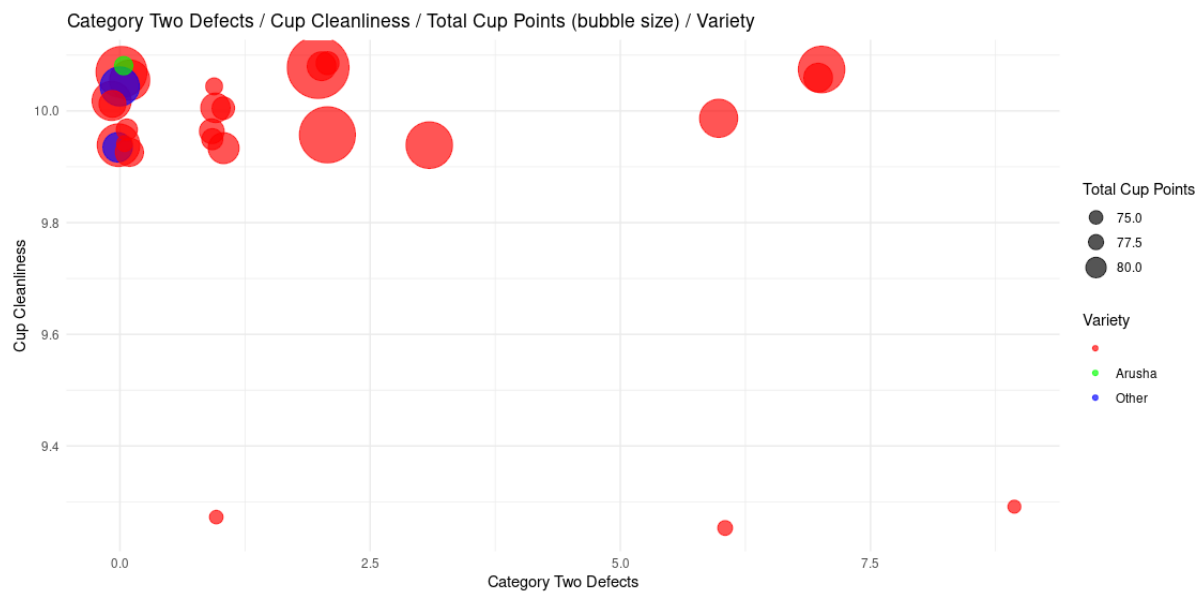
Defekty drugiej kategorii, *Cup Cleanliness*, sumaryczna ocena (wielkość bąbelka) i kraj pochodzenia

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



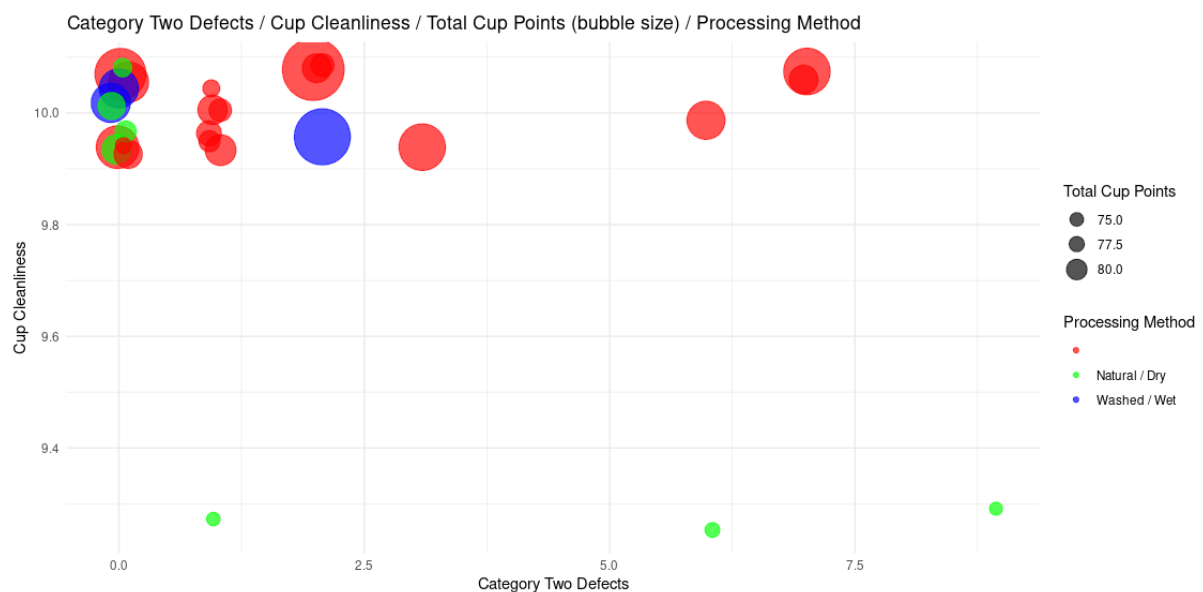
Defekty drugiej kategorii, *Cup Cleanliness*, sumaryczna ocena (wielkość bąbelka) i odmiana robusty

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



Defekty drugiej kategorii, *Cup Cleanliness*, sumaryczna ocena (wielkość bąbelka) i metoda obróbki

Wykres w wyższej rozdzielczości dostępny [tutaj](#).



Podsumowanie

Korelacje czynników zewnętrznych z parametrami formalnej oceny są relatywnie intensywne w przypadku ziaren robusty. Wynika to niestety najprawdopodobniej z małej liczności obserwacji w tym zbiorze. Gdyby ten zbiór był tak duży, jak ten dotyczący arabiki to najprawdopodobniej zależności nie byłyby tak silnie widoczne (choć możliwe, że korelowałyby te same parametry). Taka sytuacja ma miejsce w przypadku ziaren arabiki, ale i tutaj wymienione wcześniej zależności zasługują na uwagę. Wysokość uprawy zdaje się w jakiś sposób wpływać na parametry smakowe takie jak kwasowość i aromat (także na inne w mniejszym stopniu). Wiedza, na jakie parametry smakowe i w jakiej liczności wpływają defekty danej kategorii lub jakie warunki środowiskowe wpływają na tworzenie się defektów, może być szczególnie istotna. Pokazane wizualizacje wydają się potwierdzać niektóre informacje powtarzane przez ludzi profesjonalnie związanych z kawą.

Kod

Kod, który posłużył do wygenerowania wizualizacji: [link](#).