

Coffee Ratings

Fernando Finotto Visani

7/11/2020

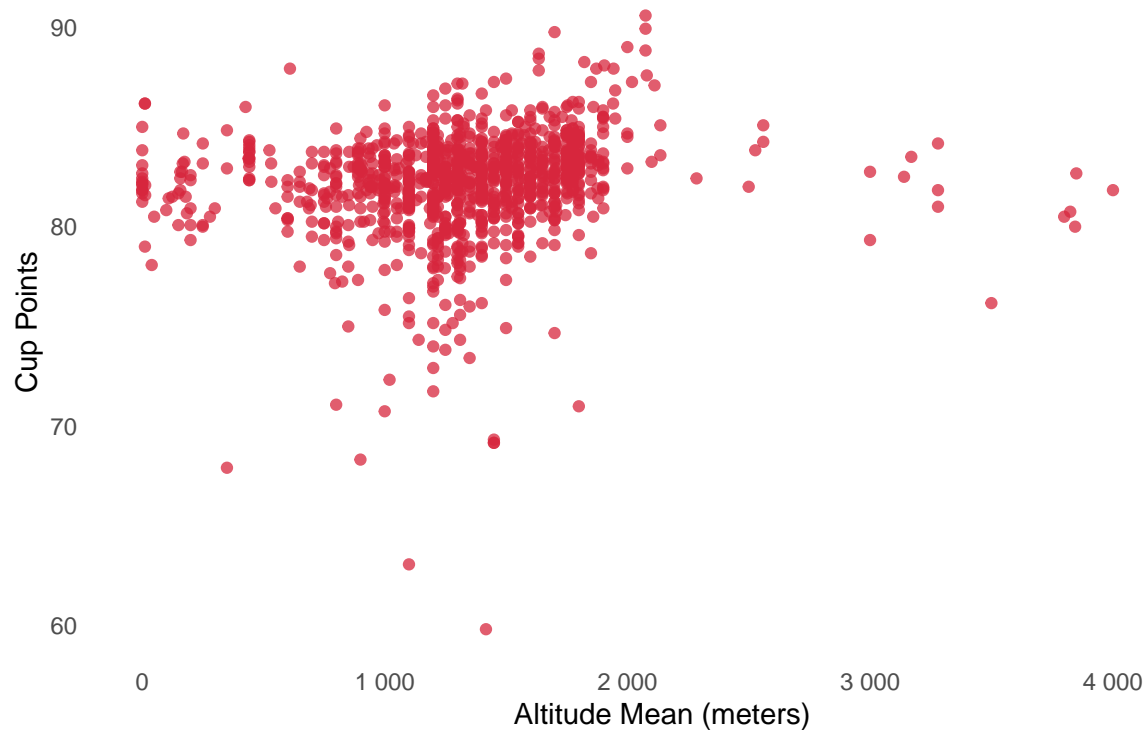
Does Altitude has a correlation with Coffee Ratings?

```
coffee_altitude <- coffee_ratings %>%
  filter(!is.na(altitude),
         total_cup_points > 0,
         altitude_mean_meters > 0) %>%
  select(total_cup_points,
         species,
         country_of_origin,
         harvest_year,
         variety:category_two_defects,
         altitude_low_meters:altitude_mean_meters)

coffee_altitude %>%
  ggplot(aes(altitude_mean_meters,
             total_cup_points)) +
  geom_point(alpha = 0.75, color = "#D7263D") +
  labs(title = "Does Altitude has a correlation with Coffee Ratings?",
       subtitle = "Answer: No",
       x = "Altitude Mean (meters)",
       y = "Cup Points") +
  theme(panel.grid = element_blank()) +
  scale_x_continuous(labels = scales::number_format())
```

Does Altitude has a correlation with Coffee Ratings?

Answer: No

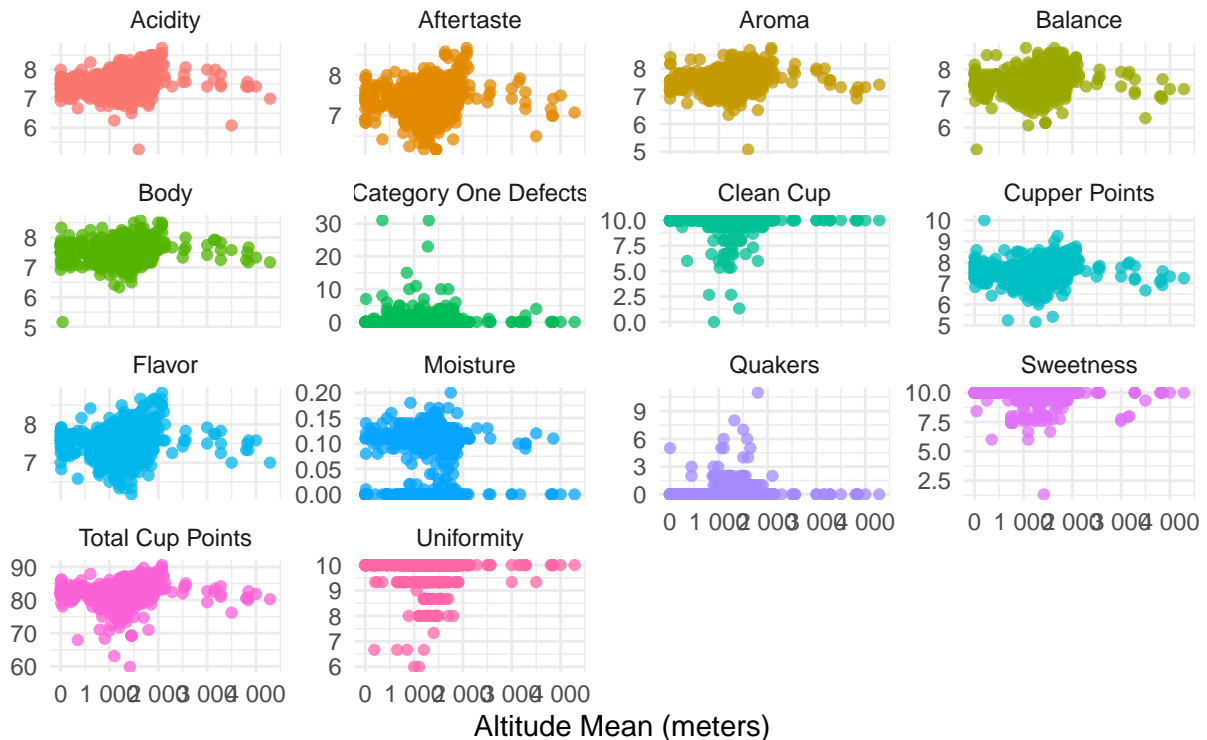


Does Altitude has a correlation with any metric?

```
coffee_altitude %>%
  select(altitude_mean_meters,
         total_cup_points,
         aroma:quakers) %>%
  pivot_longer(-altitude_mean_meters) %>%
  mutate(name = str_replace_all(name, "_", " "),
         name = str_to_title(name)) %>%
  ggplot(aes(altitude_mean_meters,
             value,
             color = name)) +
  geom_point(alpha = 0.75) +
  facet_wrap(~name, scales = "free_y") +
  theme(legend.position = "none") +
  scale_x_continuous(labels = scales::number_format()) +
  labs(title = "Does Altitude has a correlation with any metric?",
       subtitle = "Answer: No",
       x = "Altitude Mean (meters)",
       y = "")
```

Does Altitude has a correlation with any metric?

Answer: No



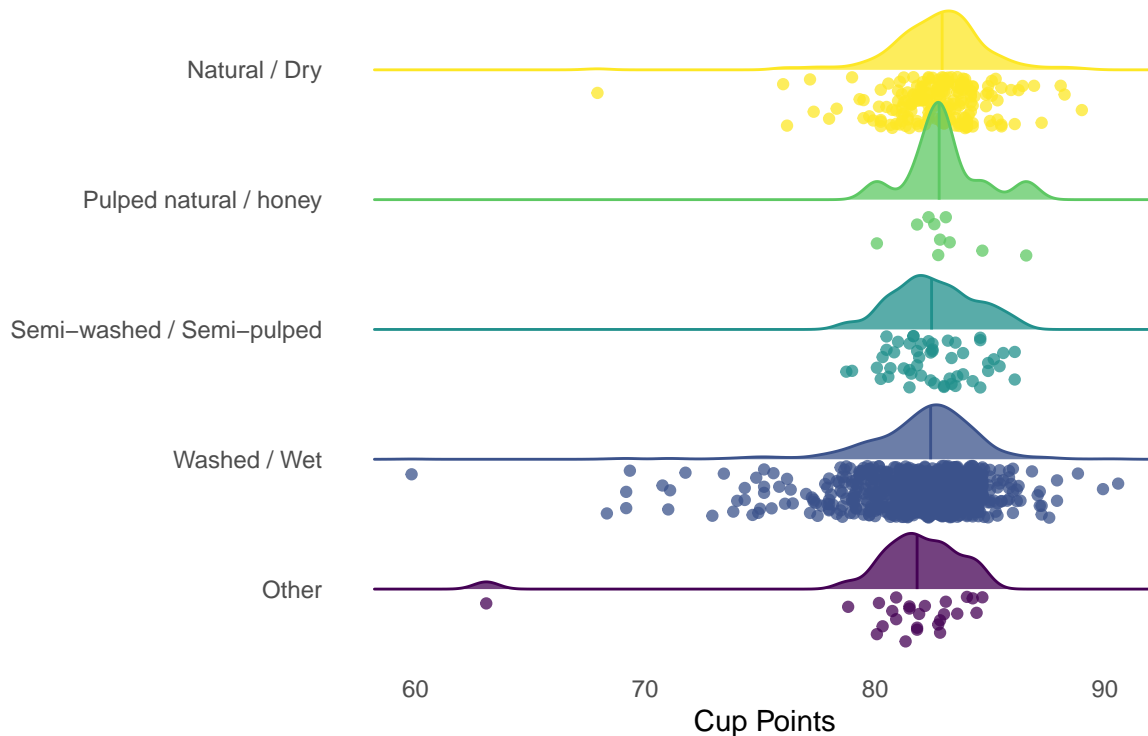
What's the Cup Point distribution for each processing method?

```
coffee_ratings %>%
  filter(total_cup_points > 0,
         !is.na(processing_method)) %>%
  select(processing_method,
         total_cup_points) %>%
  mutate(processing_method = fct_reorder(processing_method, total_cup_points)) %>%
  ggplot(aes(total_cup_points,
             processing_method,
             color = processing_method,
             fill = processing_method)) +
  geom_density_ridges(alpha = 0.75,
                    scale = 0.75,
                    quantile_lines = TRUE,
                    quantiles = 0.5,
                    jittered_points = TRUE,
                    position = "raincloud") +
  theme(panel.grid = element_blank(),
        legend.position = "none") +
  labs(title = "What's the Cup Point distribution for each Processing Method?",
       subtitle = "Mean lines are indicated",
       x = "Cup Points",
```

```
y = "" +
  scale_fill_viridis_d() +
  scale_color_viridis_d()
```

```
## Picking joint bandwidth of 0.539
```

What's the Cup Point distribution for each Processing Method? Mean lines are indicated



How each region compares to each other for the cup points?

```
wikipedia_url <- "https://meta.wikimedia.org/wiki/List_of_countries_by_regional_classification"

countries <- read_html(wikipedia_url) %>%
  html_node("table") %>%
  html_table() %>%
  as_tibble() %>%
  clean_names() %>%
  mutate(region = str_remove(region, "South/"),
         country = case_when(
           str_detect(country, "^Tanzania") ~ "Tanzania",
           str_detect(country, "Laos") ~ "Laos",
           TRUE ~ country
         ))
```

```

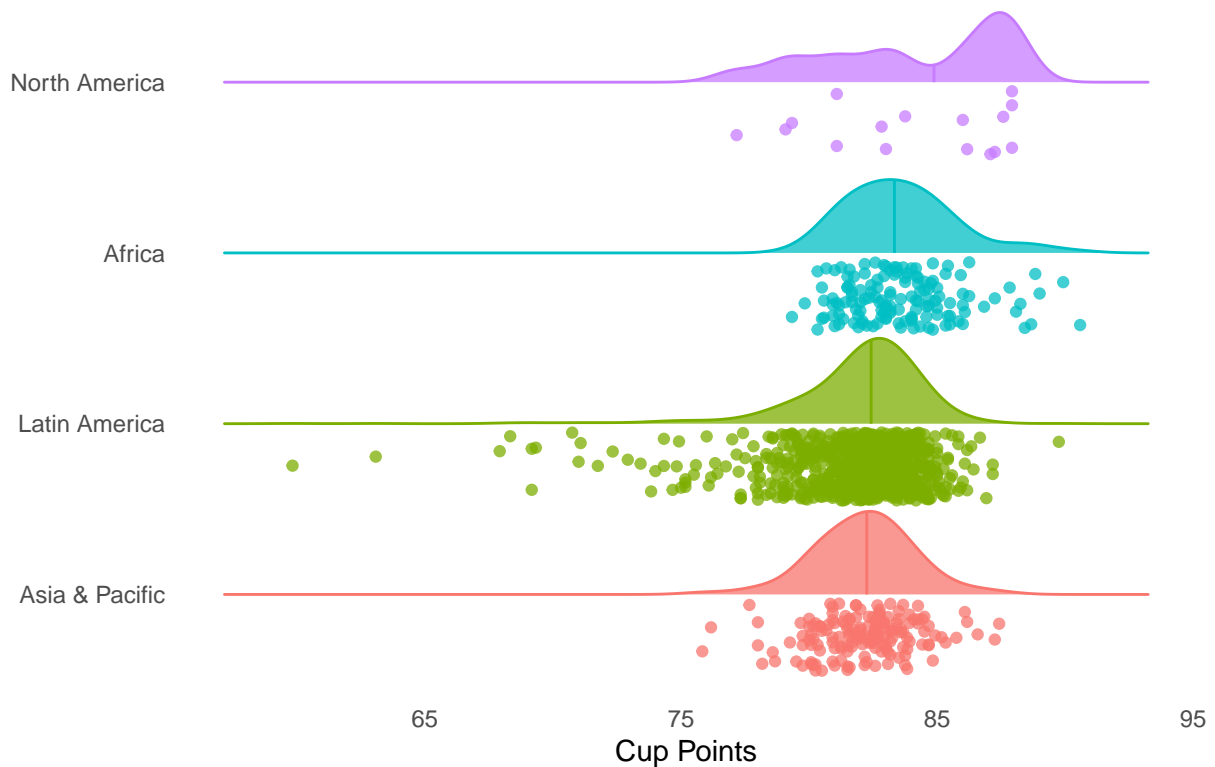
coffee_regions <- coffee_ratings %>%
  mutate(country_of_origin = case_when(
    str_detect(country_of_origin, "Tanzania") ~ "Tanzania",
    str_detect(country_of_origin, "United States") ~ "United States",
    str_detect(country_of_origin, "Laos") ~ "Lao People's Democratic Republic",
    str_detect(country_of_origin, "^Cote") ~ "Côte D'Ivoire",
    TRUE ~ country_of_origin
  )) %>%
  left_join(countries, by = c("country_of_origin" = "country")) %>%
  select(region = region.y,
    species,
    processing_method,
    country_of_origin,
    total_cup_points)

coffee_regions %>%
  filter(total_cup_points > 0) %>%
  mutate(region = fct_reorder(region, total_cup_points)) %>%
  ggplot(aes(total_cup_points,
    region,
    fill = region,
    color = region)) +
  geom_density_ridges(alpha = 0.75,
    scale = 0.5,
    quantile_lines = TRUE,
    quantiles = 0.5,
    jittered_points = TRUE,
    position = "raincloud") +
  theme(panel.grid = element_blank(),
    legend.position = "none") +
  labs(title = "How each region compares to each other for Cup Points?",
    x = "Cup Points",
    y = "")

```

```
## Picking joint bandwidth of 0.886
```

How each region compares to each other for Cup Points?



How each country rank by region?

```
countries_to_keep <- coffee_regions %>%
  group_by(country_of_origin) %>%
  summarise(mean = mean(total_cup_points, na.rm = TRUE)) %>%
  arrange(desc(mean)) %>%
  head(20) %>%
  pull(country_of_origin)
```

'summarise()' ungrouping output (override with '.groups' argument)

```
coffee_regions %>%
  group_by(country_of_origin) %>%
  filter(n() > 2) %>%
  ungroup() %>%
  filter(country_of_origin %in% countries_to_keep) %>%
  mutate(country_of_origin = fct_reorder(country_of_origin, total_cup_points)) %>%
  ggplot(aes(total_cup_points,
             country_of_origin,
             fill = region)) +
  geom_density_ridges() +
  theme(panel.grid = element_blank(),
```

```

legend.title = element_blank(),
legend.position = "top",
plot.title = element_text(size = 12, hjust = 0.5),
legend.text = element_text(size = 8),
legend.key.size = unit(0.5, "cm"),
axis.text = element_text(size = 8),
axis.title.x = element_text(size = 8)) +
labs(title = "How the Top 20 countries rank by Cup Points mean?",
x = "Cup Points",
y = "") +
scale_fill_manual(values = c("#495867", "#84A98C", "#FE5F55", "#BDD5EA"))

```

Picking joint bandwidth of 0.753

