# Proposal

## The group members (names and UNIs)

Jinghan Zhao - jz3902; Leyang Rui - lr3257; Yifei Yu - yy3563; Yuechu Hu - yh3822

## The tentative project title

Not your average joe: what makes the best coffee?

## The motivation for this project

Since the beginning of the 20th century, coffee has transcended national boundaries and catered to the tastes of diverse groups of people. As a group of coffee lovers, we are well-realized about how important a good cup of coffee can be for a fulfilling day. Therefore, we wish to analyze the quality of coffee from two aspects – the objective geographical factors and subjective costumer preferences.

## The intended final products

In general, we hope to obtain a comprehensive view on both the objective factors that influence the taste of the coffee (e.g. species, altitude) and how features of different populations affect their subjective preferences for coffee (e.g. gender, education level). There are also additional questions that could be explored, such as how much people would pay for a cup of coffee based on their quality, and does the public taste of coffee accords with the "official" level of certification.

## The anticipated data sources

The project utilizes the data from the Coffee Quality Institute (the original data can be found on James LeDoux's github) to rate coffee based on their acidity, sweetness, aroma, altitude, and region, offering a comprehensive look at coffee. By incorporating data on the "Great American Coffee Taste Test", our project also explores how consumer tastes align with professional ratings. This combination of objective and public data can provide insights into how coffee qualities influence consumer preferences, helping producers, coffee shops, and consumers make their favorite choices.

## The planned analyses / visualizations / coding challenges

Data cleaning, data binding, EDA, linear model, and correlation will be applied in this project. Potential coding challenges may include how to define and filter out categorical variables, how to quantify survey questions, how to visualize countries, regions, and altitudes, and how to discover and measure differences between regions.

## The planned timeline

- November 6 - submit proposal document
- November 11-15 - project review meeting, constructing outline and proposing research questions
- November 16 - data cleaning + data binding
- November 23 - EDA + visualizations
- November 30 - linear model + correlation test
- December 5 - complete written report, webpage and explanatory video
- December 12 - in-class discussion