**DATA 557**

**Project Proposal**

**Winter 2020**

**Group members**

1. Amol Dhaygude
2. Carl Parker
3. Dhruv Arora
4. Jesse Freitas
5. Sandhya Tharanian
6. Vikrant Bhosale

**Project description**

We will analyze heart-disease mortality data for the United States. The response variable will be mean mortality rate. The analysis will use a factorial design that compares mean mortality rates using factors for state, county, gender, and race/ethnicity.

**Analysis questions**

**Global (across US) questions**

* Across the United States, is the mean mortality rate the same between genders?
* Across the United States, is the mean mortality rate the same between different races/ethnicities?
* Is the mean mortality rate same between different states in the United States? If not, which states show the greatest differences?

**Nested (within state) questions**

* Within each state, is the mean mortality rate the same between genders? If not, which states show the greatest differences.
* Within each state, is the mean mortality rate the same between different races/ethnicities? If not, which states show the greatest differences.
* Within each state, is the mean mortality rate the same between counties? If not, which states show the greatest differences?

**Interaction questions**

* Across the United States, is there an interaction between state and race/ethnicity on mortality rates?
* Across the United States, is there an interaction between gender and race/ethnicity on mortality rates?

**Data description**

We are using the following dataset from the Centers for Disease Control (CDC).

# [Heart Disease Mortality Data Among US Adults (35+) by State/Territory and County – 2015-2017](https://catalog.data.gov/dataset/heart-disease-mortality-data-among-us-adults-35-by-state-territory-and-county-2015-2017-d0c95)

The following table describes the subset of columns from the dataset that we will use for our analysis.

